Rough Volatility: Pushing the boundaries of quantitative modelling past the Markovian era

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

by

Aitor Muguruza Gonzalez

Department of Mathematics

Imperial College London

180 Queen's Gate

London SW7 2BZ

September 2020

I, Aitor Muguruza Gonzalez, certify that this thesis and the research to which it refers are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed: _____

Copyright

Rough Volatility: Pushing the boundaries of quantitative modelling past the Markovian era

Abstract

Rough volatility models have brought a breeze of fresh air into financial modelling, which historically has been mostly tied to Markovian models. Alas, this novelty comes with a number of challenges and questions that require new techniques to be answered. The present thesis first explores the numerical implementation of rough volatility models, both to provide convergence results and to ensure a competitive implementation. Our findings conclude that fractional calculus and Hölder spaces provide the right framework to set the theoretical grounds of numerical schemes related to rough volatility models and fractional diffusions in general. In addition, we explore the use of deep learning methods to leverage the speed of numerical schemes and provide state-of-the-art calibration techniques that make rough volatility models as simple and efficient to implement as the Black-Scholes formula. The second part of the thesis explores the theoretical properties of rough volatility models regarding volatility products such as VIX and realized variance options. We explore short-time asymptotics using two different techniques: Malliavin Calculus and Large Deviations Principle. Remarkably, we find that both approaches complement each other and provide a very accurate description of the short-time implied volatility smile. Building upon these short-time asymptotics, we further provide approximation formulas for the at-the-money implied volatility and its skew. Finally, in the third part of the thesis we cross the bridge between the physical $\mathbb{P}$ and risk-neutral $\mathbb{Q}$ measures. We provide sufficient conditions for such change of measure to be well defined and explore the numerical estimation of the market risk premium process in rough volatility models.

To my parents.

Acknowledgements

I am deeply indebted to my supervisors Antoine Jacquier and Blanka Horvath, who not only have guided me throughout this 4 year academic path, but have introduced me to an amazing network of researchers and allowed me to pursue my own projects, ideas and industry internships. I really hope this to be just the beginning of a long lasting collaboration (and friendship!).

A number of thanks is also due to my co-authors: Mehdi, Christian, and Benjamin you get my "deepest" gratitude. David and specially Elisa for introducing me to Malliavin calculus and hosting me in Barcelona, I am profoundly grateful. I would like to also thank Claude; words probably can't describe the inspiration and eagerness to share knowledge I have received in every single visit to Paris. I am also grateful to Ismail, Jacopo and Pierre at Zeliade for hearing me out every time I visited.

A special mention goes here to Henry and Chloe who are both more than co-authors to me. We have shared almost all the PhD with its ups and downs, and the moto "Let's go rough vol" will probably stick around for a while...

There are many important people in my academic path that led me towards Quantitative Finance and to pursue a PhD. In particular, I would like to thank Professor Thaleia Zariphopoulou for introducing me in the first place to mathematical finance and always believing on my capacities no matter my academic background and to Milica Cudina at UT.

My colleagues at Natixis, Sylvain and Luc also deserve a special mention for many hours of fruitful discussions and willingness to test many new modelling approaches. I am also grateful to Adil for giving me the opportunity to join the team in the first place.

Perhaps, one of the key factors in the development of my PhD has been the ability to attend several conferences and meet an amazing network or researchers in mathematical finance. I will specially have good memories of the Oberwolfach conference and Jim Gatheral's birthday conference for which I am grateful to Jim for hosting me in Baruch and the EPSRC for supporting my stay. My office colleague Ryan also deserves a mention here; we have definitely shared quite a few hours of discussions and brainstorming.

I am specially thankful to my parents for their unconditional support and kindness. They educated me this way and always encouraged me to pursue my curiosity. Without them it would have been impossible for me to accomplish many academic achievements and be able to continue this path.

Finally and most importantly, I would like to thank my wife, Andrea. We both have fought together our downs during our PhD and enjoyed each others successes. Thank you for bringing the best out of me and for being there whenever I needed you. You always encouraged and supported me to start the PhD journey in the first place and that's a decision I shall never regret!

List of Figures

CHAPTER 2. FUNCTIONAL CENTRAL LIMIT THEOREMS FOR ROUGH VOLATILITY

2.1 fractional Binomial trees 18

2.2 Rough Bergomi Call options price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$. 22

2.3 Rough Bergomi Call options price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$. 23

2.4 Rough Bergomi Call options price convergence using rDonsker method with moment-matching and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$. 24

2.5 Rough Bergomi Call options price convergence using rDonsker method with left point Euler and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$. 25

2.6 Rough Bergomi Call options price comparison with $H = 0.01, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 26

2.7 Rough Bergomi Call options price comparison with $H = 0.05, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 27

2.8 Rough Bergomi Call options price comparison with $H = 0.10, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 28

2.9 Rough Bergomi Call options price comparison with $H = 0.2, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 29

2.10 Rough Bergomi Call options price comparison with $H = 0.3, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 30

2.11 Rough Bergomi Call options price comparison with $H = 0.4, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes. 31

2.12 Computational time benchmark using Hybrid scheme, rDonsker and Markovian (forward Euler) for different grid sizes n. 31

2.13 Rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 5 time steps. 34

2.14 rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, 1, 0.04)$ with 5 time steps. 35

2.15 rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps. 36

2.16 Error analysis for the rDonsker moment-match tree for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps. 36

2.17 American and European Put prices in the rough Bergomi model for different values of H and $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 26 time steps. 37

CHAPTER 3. DEEP LEARNING VOLATILITY: A DEEP NEURAL NETWORK APPROACH TO PRICING AND CALIBRATION IN (ROUGH) VOLATILITY MODELS

3.1 In detail neuron behaviour 50

3.2 Our neural network architecture with 4 hidden layers and 30 neurons on each hidden layer, with the model parameters of the respective model on the input layer and with the 8×11 implied volatility grid on the output layer. 51

3.3 Monte-Carlo rough Bergomi errors 53

3.4 Monte-Carlo 1 Factor Bergomi errors 54

3.5 Neural Network performance rough Bergomi model 55

3.6 Neural Network performance 2 Factor Bergomi model 55

3.7 Average calibrations times for all models using a range of optimizers. 56

3.8 Cumulative Distribution Function (CDF) of Rough Bergomi parameter relative errors (left) and RMSE (right) after Levengerg-Marquardt calibration across test set random parameter combinations. 57

3.9 Cumulative Distribution Function (CDF) of 1 Factor Bergomi parameter relative errors (left) and RMSE (right) after Levengerg-Marquardt calibration across test set random parameter combinations. 57

3.10 Historical parameters in the rough Bergomi model fitted to SPX 59

3.11 Neural network historical calibration RMSE on SPX 59

3.12 Down-and-Out and Down-and-In error analysis 60

CHAPTER 4. NOT SO PARTICULAR ABOUT CALIBRATION: SMILE PROBLEM RESOLVED

4.1 Graphical description of the forward in time LSV calibration algorithm. 63

4.2 SX5E (3-Sept-2019) Implied volatility surface and Rough Bergomi LSV fit with $\nu = 380\%$ 68

4.3 SX5E (3-Sept-2019) Implied volatility surface and Rough Bergomi LSV fit with $\nu = 800\%$ 68

4.4 Computational times of the method. 70

4.5 RSME as a function of number of simulations. 70

4.6 SX5E (3-Sept-2019) Implied volatility surface LSV fit with Rough Bergomi and Vasicek 71

4.7 SX5E (3-Sept-2019) Implied volatility surface LSV fit with 2FBergomi and Vasicek 73

4.8 Nikkei (11-Sept-2018) Implied volatility surface LSV fit with 2FBergomi and Vasicek 74

4.9 SPX (09-Jun-2011) Implied volatility surface LSV fit with 2FBergomi and Vasicek 75

CHAPTER 5. ON VIX FUTURES IN THE ROUGH BERGOMI MODEL

5.1 Bounds vs. Monte Carlo (Truncated Cholesky) in all three scenarios. 80

5.2 Monte-Carlo benchmarking for all three methods with 10^6 simulations and the corresponding 95% confidence level standard deviations 83

5.3 Monte-Carlo standard deviations and computational time ratios for a fixed maturity ($T = 2$ years) for all methods using efficient parallel computing. 83

5.4 Log-normal approximations vs. Monte Carlo (Truncated Cholesky) in all three scenarios. 86

5.5 Log-normal approximation vs. Monte Carlo (Truncated Cholesky) with 10^6 simulations 86

5.6 VIX Futures prices following Approximation 5.3.15 and Approximation 5.3.16. 87

5.7 Comparison of the variance following Approximation 5.3.15 (σ^2) and Approximation 5.3.16 ($\tilde{\sigma}^2$). . . 87

5.8 eSSVI calibration results on 4/12/2015 using traded SPX options. 90

5.9 VIX Futures calibration on 4/12/2015. Optimal parameters: $(H, \nu) = (0.09237, 1.004)$. 90

5.10 VIX Futures calibration on 22/2/2016. Optimal parameters: $(H, \nu) = (0.10093, 1.00282)$. 91

5.11 VIX Futures calibration on 4/1/2016. Optimal parameters: $(H, \nu) = (0.0509, 1.2937)$. 91

5.12 Calibration of SPX smiles on 4/12/2015. Calibrated parameters: $(\nu, \rho) = (1.19, -0.999)$. 93

CHAPTER 6. ON SMILE PROPERTIES OF VOLATILITY PRODUCTS

6.1 VIX options implied volatilities on 08/04/2016. Solid lines represent the VIX future prices corresponding to each maturity. Source: OptionMetrics. 101

6.2 VIX ATMI Short time limit in an exponential $\mathcal{TBSS}$ with $\nu = 2$ 103

6.3 VIX ATMI in a Generalized rough volatility model with $\nu = 2$ 104

6.4 VIX ATMI in a Mixed Generalized rough volatility model with $(\delta, \nu, \eta) = (1/2, 2, 1)$ 104

6.5 Generalized rough volatility model ATMI skew on VIX options with $\nu = 7.5$ 105

6.6 Generalized rough volatility model ATMI skew with $\nu = 2.5$ 106

6.7 Mixed generalized rough volatility model ATMI skew with $(\delta, \nu, \eta) = (1/2, 3, 1)$ 106

6.8 VIX options implied volatilities on 28/01/2015. Calibrated parameters 107

6.9 VIX options implied volatilities on05/02/2015. Calibrated parameters 108

6.10 Realized variance ATMI for S&P 500 on February 27, 2018 (from OTC market quotes obtained through different brokers). 108

6.11 Realized variance option ATMI with $\nu = 2$ 110

6.12 Realized variance option ATMI with $\delta = 0.5, \nu = 2$ and $\eta = 1$ 110

6.13 ATMI skew short time limit for RV options in a rough Bergomi model with $\nu = 2$. 111

6.14 ATMI skew short time limit for RV options in a mixed rough Bergomi model with $(\delta, \nu, \eta) = (1/2, 2, 1)$. 111

CHAPTER 7. ASYMPTOTICS FOR VOLATILITY DERIVATIVES IN MULTI-FACTOR ROUGH VOLATILITY MODELS

7.1 Implied volatility smiles for Call options on VIX for small maturities, close to the money. Data provided by OptionMetrics. 114

7.2 Comparison of Monte Carlo vs LDP estimates. 122

7.3 Rate function $\hat{\Lambda}^V$ for different values of α. 122

7.4 Comparison of Chapter 6 at-the-money implied volatility asymptotics and LDP based implied volatilities for different values of α, in the rough Bergomi model, with $\eta = 1.5$ and $V_0 = 0.04$. 123

7.5 Comparison of LDP based implied volatilities for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$ in the mixed rough Bergomi process (7.6) such that $\gamma_1\nu_1 + \gamma_2\nu_2 = 2$, with $\alpha = -0.4$. 123

7.6 Rate function and corresponding implied volatilities for the model (7.18), with $(\alpha, \nu, \eta) = (-0.4, 0.75, 0.75)$. 124

7.7 Rate function and corresponding implied volatilities for the model (7.19) with $(\alpha, \nu, \eta) = (-0.4, 1.0, 3.0)$. 125

7.8 Implied volatilities and superimposed linear smiles for the model (7.19), with $(\alpha, \nu, \eta) = (-0.4, 1.0, 3.0)$. 125

CHAPTER 8. RISK PREMIUM, NON-MARKOVIANITY AND ROUGH VOLATILITY; A MEASURE CHANGE POINT OF VIEW

8.1 Variance Swap volatility daily quotes on SX5E 138

8.2 Forward variances extracted from variance swap quotes on SX5E 138

8.3 Annualised daily realized volatility on SX5E 139

8.4 Estimated $\hat{H}$ and $\hat{\nu}$ on SX5E 139

8.5 Daily correlation estimate on SX5E returns and realized volatility. 140

8.6 Daily risk premium dynamics on SX5E; dashed lines represent the mean value of each process with the corresponding color. 141

8.7 Daily 1-day ahead forecasts vs. actual 1 month Variance Swap prices on SX5E 142

8.8 Daily 1-day ahead forecasts vs. actual 3 month Variance Swap prices on SX5E 143

8.9 Daily 1-day ahead forecasts vs. actual 6 month Variance Swap prices on SX5E 143

APPENDIX A.

B.1 Out of sample relative errors per parameter calibration 150

B.2 Stars represent the out of sample RMSE via neural network (NN) in the one-step approach and brute force Monte Carlo (MC). Dashed black line represents the identity function. 151

B.3 Stars represent the out of sample RMSE via neural network (NN) using the two-step approach and brute force Monte Carlo (MC). Dashed black line represents the identity function. 151

APPENDIX E.

E.1 Volatility Swap Monte Carlo price estimates (straight lines) and LDP based approximation (stars) for $\eta = 1.5$ and $V_0 = 0.04$; for Monte Carlo we use $200,000$ simulations and $\Delta t = \frac{1}{1008}$. 165

E.2 Absolute difference of subsequent approximating polynomials $|\hat{\Lambda}^V_{n+1}(\mathrm{y}) - \hat{\Lambda}^V_n(\mathrm{y})|$ 172

E.3 Absolute error of the rate function. We consider the truncated basis approach against the standard polynomial basis with $\eta = 1.5$, $V_0 = 0.04$ and different values of α. 173

List of Tables

CHAPTER 1. INTRODUCTION

CHAPTER 2. FUNCTIONAL CENTRAL LIMIT THEOREMS FOR ROUGH VOLATILITY

CHAPTER 3. DEEP LEARNING VOLATILITY: A DEEP NEURAL NETWORK APPROACH TO PRICING AND CALIBRATION IN (ROUGH) VOLATILITY MODELS

3.1 Comparison of Gradient vs. Gradient-free methods. 42

3.2 Computational time of pricing map (entire implied volatility surface) and gradients via Neural Network approximation and Monte Carlo (MC). If the forward variance curve is a constant value, then the speed-up is even more pronounced 53

Contents

Chapter 1. Introduction 1

1.1 Markovian models and Quantitative Finance . . . 1

1.1.1 Markovian models and option pricing . . . 2

1.1.2 The Markovian assumption in question . . . 3

1.2 Rough volatility: a new breed of non-Markovian models . . . 3

1.3 Thesis Overview . . . 5

Chapter 2. Functional Central Limit Theorems for Rough Volatility 7

2.1 Introduction . . . 7

2.2 Weak convergence of rough volatility models . . . 8

2.2.1 Hölder spaces and fractional operators . . . 8

2.2.2 Examples . . . 10

2.2.3 The approximation scheme . . . 11

2.3 Functional Central limit theorems for a family of Hölder continuous processes . . . 11

2.3.1 Weak convergence of Brownian motion in Hölder spaces . . . 11

2.3.2 Weak convergence of Itô diffusions in Hölder spaces . . . 13

2.3.3 Invariance principle for rough processes . . . 14

2.3.4 Extending the weak convergence to the Skorokhod space and discussion of Conjecture 2.2.11 16

2.4 Applications . . . 17

2.4.1 Weak convergence of the Hybrid scheme . . . 17

2.4.2 Application to fractional binomial trees . . . 18

2.4.3 Monte-Carlo . . . 18

2.4.4 Numerical example: Rough Bergomi model . . . 21

2.4.5 Speed benchmark against Markovian stochastic volatility models . . . 31

2.4.6 Implementation guidelines and conclusion . . . 32

2.4.7 Bushy trees and binomial markets . . . 32

2.4.8 American options in rough volatility models . . . 33

2.5 Summary . . . 35

Chapter 3. Deep Learning Volatility: A deep neural network approach to pricing and calibration in (rough) volatility models 38

3.1 Introduction . . . 38
3.2 Model calibration . . . 40
3.2.1 The numerical optimisation . . . 41
3.2.2 A showcase of (rough) Stochastic Volatility models . . . 42
3.3 Deep calibration . . . 43
3.3.1 One-step approach: Deep calibration by the inverse map . . . 43
3.3.2 Two-step approach: Learning the implied volatility map of models . . . 44
3.3.3 Some relevant properties of deep neural networks as functional approximators . . . 49
3.3.4 Network architecture and training . . . 50
3.4 Numerical experiments . . . 52
3.4.1 Numerical accuracy and speed of the price approximation for vanillas . . . 52
3.4.2 Calibration speed and accuracy for implied volatility surfaces . . . 55
3.4.3 Numerical experiments with barrier options in the rough Bergomi model . . . 59
3.5 Summary . . . 60

Chapter 4. Not so particular about calibration: Smile problem Resolved 61

4.1 Introduction . . . 61
4.2 LSV framework and mathematical setting . . . 61
4.3 The calibration of the leverage function in a nutshell . . . 62
4.3.1 The original particle method . . . 63
4.3.2 Limitations of kernel regression: Bias, Variance and Convergence . . . 64
4.4 The exact formula . . . 64
4.4.1 Further exploiting the discrete nature of numerical algorithms: lognormal SV case . . . 65
4.5 The new LSV calibration algorithm . . . 66
4.6 Overture to local correlation and basket smiles . . . 69
4.7 Numerical tests with hybrid models . . . 69
4.7.1 Vasicek and rough local stochastic volatility . . . 69
4.7.2 Vasicek and local 2F Bergomi . . . 70
4.8 Summary . . . 72

Chapter 5. On VIX Futures in the rough Bergomi model 76

5.1 Introduction . . . 76
5.2 Rough volatility and the rough Bergomi model . . . 76
5.3 Rough Bergomi and VIX . . . 78
5.3.1 VIX Futures in the rough Bergomi model. . . . 78

5.3.2 Upper and lower bounds for VIX Futures . . . 79
5.3.3 Numerical implementation of VIX process . . . 80
5.3.4 The VIX process and log-normal approximations . . . 83
5.3.5 VIX Futures calibration in rBergormi . . . 87
5.4 From VIX Futures to SPX options . . . 91
5.4.1 Pricing in the rough Bergomi model . . . 91
5.4.2 Calibration of SPX options via VIX Futures . . . 92
5.4.3 Results . . . 92
5.5 Summary . . . 93

Chapter 6. On Smile properties of volatility products 94

6.1 Introduction . . . 94
6.2 Malliavin calculus . . . 95
6.3 Transforming variance options into options on the spot process of a stochastic volatility model . . . 96
6.4 Implied volatility for variance options . . . 97
6.4.1 The ATMI short-time limit . . . 97
6.4.2 The short-time limit of the ATM implied volatility skew . . . 99
6.4.3 The sign of the VIX skew in some popular models . . . 100
6.5 VIX options . . . 101
6.5.1 ATMI of VIX options . . . 101
6.5.2 ATMI skew of VIX options . . . 104
6.5.3 MGRV fit to empirical VIX options data . . . 107
6.6 Realized variance options . . . 108
6.6.1 ATMI on realized variance options . . . 108
6.6.2 ATMI skew for realized variance options . . . 110
6.7 Summary . . . 112

Chapter 7. Asymptotics for volatility derivatives in multi-factor rough volatility models 113

7.1 Introduction . . . 113
7.2 A showcase of rough volatility models . . . 115
7.3 Small-time results for options on integrated variance . . . 116
7.3.1 Small-time results for the rough Bergomi model . . . 117
7.3.2 Small-time results for the mixed rough Bergomi model . . . 118
7.3.3 Small-time results for the multi-factor rough Bergomi model . . . 120

7.4 Numerical results . . . 121
7.4.1 RV smiles for rough Bergomi . . . 121
7.4.2 RV smiles for mixed rough Bergomi . . . 121
7.4.3 RV smiles for mixed multi-factor rough Bergomi . . . 124
7.5 Options on VIX . . . 125
7.6 Summary . . . 127

Chapter 8. Risk premium, Non-Markovianity and Rough Volatility; a measure change point of view 129

8.1 Introduction . . . 129
8.2 Rough volatility models and change of measure . . . 130
8.2.1 Characterisation of rough volatility models via Generalised Fractional Operators . . . 133
8.3 Modelling the risk premium process: A practical approach . . . 134
8.3.1 Risk premiums driven by Itô diffusion . . . 135
8.3.2 Gaussian risk premiums . . . 135
8.3.3 A risk premium driven by a CIR process . . . 136
8.4 Roughly extracting the Risk Premium from the Market . . . 136
8.4.1 Numerical experiment with the rough Bergomi model I: estimating the risk premium . . . 137
8.4.2 Numerical experiment with the rough Bergomi model II: forecasting Variance Swaps . . . 141
8.5 Summary . . . 143

Appendix A. 145

A.1 Discrete convolution . . . 145
A.2 Riemann-Liouville operators . . . 146
A.3 Additional Proofs . . . 148
A.3.1 Proof of Proposition 2.2.5 . . . 148

Appendix B. 149

B.1 A numerical experiment with the inverse map . . . 149

Appendix C. 152

C.1 Proof . . . 152

Appendix D. 153

D.1 Proofs of Theorems 6.4.4 and 6.4.6 . . . 153

D.1.1 Proof of Theorem 6.4.4 . . . 154
D.1.2 Proof of Theorem 6.4.6 . . . 155
D.2 Proof of Proposition 6.5.4 . . . 156
D.3 Computations of the ATMI skew sign in the Heston model . . . 159
D.4 Proof of Proposition 6.6.6 . . . 160

Appendix E. 163

E.1 Approximating the density of realised variance in the mixed rough Bergomi model . . . 163
E.2 Proof of Main Results . . . 166
E.2.1 Proof of Theorem 7.3.3 . . . 166
E.2.2 Proof of Corollary 7.3.7 . . . 166
E.2.3 Proof of Theorem 7.3.10 . . . 167
E.2.4 Proof of Theorem 7.3.14 . . . 168
E.2.5 Proof of Proposition 7.5.2 . . . 169
E.2.6 Proof of Theorem 7.5.6 . . . 169
E.3 Numerical recipes . . . 170
E.3.1 Convergence Analysis of the numerical scheme . . . 171
E.3.2 A tailor-made polynomial basis for rough volatility . . . 172
E.3.3 Multi-Factor case . . . 173
E.4 Exponential Equivalence and Contraction Principle . . . 174
E.5 Proof of $V_0\phi(d_1(x)) = x\phi(d_2(x))$ . . . 174

Appendix F. 176

F.1 Proof of Proposition 8.3.6 . . . 176

1 Introduction

Quantitative analytics and financial modelling in particular, have experienced a number of groundbreaking technological advances over the last decade; arguably, the most disruptive one being the ability to store and efficiently ingest large amounts of financial data. This fact, has created a number of novel challenges, but most importantly has created the data-centric vs. model-driven debate on classical problems, such as option pricing or risk management.

Historically, quantitative research in finance has been model-driven starting from the Black-Scholes [BS73] option pricing model in 1973. It is well known that market implied volatilities used to be flat (in line with the Black-Scholes constant volatility assumption) until 1987. After the 1987 crash, traders realised that markets had a significant skew with extreme downward risk, which naturally led to the so-called implied volatility smile or skew, markedly deviating from the Black-Scholes model. This new behaviour in turn, led researchers to seek alternative modelling choices that would reproduce such behaviour and shifted the premise of research into explainability of market phenomena through intuitive, yet realistic models.

In contrast, most of the data-centric approaches based on Machine Learning and Artificial Intelligence advocate against the use of fixed mathematical models and offer flexibility to solve classical problems at the cost of interpretability and intuition. Data-centric approaches take the advantage of directly interacting with the market data and are (generally speaking) model agnostic, whilst being consistent with the market behaviour as it evolves in time.

Within this dichotomy, reconciling the best of each approach and constructing relevant and realistic models is of crucial importance. The aim of the present work, is to analyse a new breed of financial models termed rough volatility models and explore their properties along with their non-Markovian nature, which gives them the ability to influence future behaviour based on the past. In a field where the Markovian property is the norm, rough volatility models bring a number of exciting features and challenge the robustness of classical (Markovian) models.

1.1 Markovian models and Quantitative Finance

The Markov property [Mar54] is a fundamental property in classical stochastic analysis. In plain words, the Markov property implies that the future behaviour of a process only depends on the current state, and is also informally known as the memoryless property. Mathematically, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an $\mathcal{F}$ adapted process X, we say that X satisfies the (weak) Markov property if

$$\mathbb{E}[f(X_t) \mid \mathcal{F}_s] = \mathbb{E}[f(X_t) \mid \sigma(X_s)]$$

for all $t \geq s \geq 0$ and $f : S \to \mathbb{R}$ bounded and measurable.

Mathematically speaking, the Markov property is convenient as it allows to compress all the past information of a process into a single state, without the loss of any information. On top of that, most of the well known stochastic analysis results (such as Feynman-Kac formula or Itô's lemma) naturally apply to Markovian models. This in turn, historically led the financial modelling framework to concentrate in the Markovian family of processes.

1.1.1 Markovian models and option pricing

The pricing of financial derivatives products is the quintessential problem in Quantitative Finance. It all started with the Black-Scholes model [BS73], who postulated the following model on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for the logarithm of the stock process:

$$\mathrm{d}X_t = \mu \mathrm{d}t - \frac{1}{2}\sigma^2 \mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 = \log(S_0) \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R} \tag{1.1}$$

where W is a $\mathbb{P}$-Brownian motion. Black and Scholes [BS73] further showed that in a economy with risk-free rate, $r \in \mathbb{R}$ the fair price of a financial security with payoff $F(X_T)$ is given by:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-rT} F(\widetilde{X}_T)\right], \tag{1.2}$$

where

$$\mathrm{d}\widetilde{X}_t = r\mathrm{d}t - \frac{1}{2}\sigma^2 \mathrm{d}t + \sigma \mathrm{d}\widetilde{W}_t, \quad \widetilde{X}_0 = \log(S_0) \in \mathbb{R}, \tag{1.3}$$

where $\widetilde{W}$ is a $\mathbb{Q}$-Brownian motion and $\mathbb{Q}$ is termed as the Risk Neutral measure. The probability measures $\mathbb{P}$ and $\mathbb{Q}$ are related through the Girsanov [Gir60] change of measure and is further discussed in Chapter 8. As previously discussed, the constant volatility σ assumption turned out to be too strong after the 1987 crisis and this led to the so-called Stochastic Volatility (SV) models, which in view of the pricing formula (1.2) were directly presented in the Risk Neutral measure $\mathbb{Q}$, rather than the Physical measure $\mathbb{P}$. The following general form of SV model is usually found in the literature:

$$\begin{aligned} \mathrm{d}X_t &= r_t \mathrm{d}t - \frac{1}{2} V_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}W_t, \quad X_0 = \log(S_0) \in \mathbb{R}, \\ \mathrm{d}V_t &= \alpha(S_t, V_t, t)\mathrm{d}t + \beta(S_t, V_t, t)\mathrm{d}Z_t \\ \mathrm{d}[W, Z]_t &= \rho \mathrm{d}t, \quad \rho \in [-1, 1], \end{aligned} \tag{1.4}$$

where W and Z are $\mathbb{Q}$-Brownian motions and for some functions $\alpha(\cdot)$ and $\beta(\cdot)$ such that the above Stochastic Differential Equation (SDE) is well defined. Examples of such models in the literature are bast, arguably a subset of the most popular ones being Hull and White [HW87], Stein and Stein [SS91], Heston [Hes93], Dupire [Dup94], Ball and Roma [BR94] and Bergomi [Ber16]. The SABR model [Hag+02] also deserves a mention here, even though it falls into a slightly more general setting. The reasons for the popularity of such models and setting (1.4) are (at least) twofold: Inmediate PDE representations through Feynman-Kac type of arguments and simple Monte Carlo simulation frameworks. The former one performs fast calibrations to market data through PDE solvers and the later one allows to price any security, even path dependent ones, using simple simulations. In addition, a bast literature in asymptotic and expansion methods have grown in the past decades in order to approximate derivative prices and further speed up calibration tasks. Examples of such methods can be found in [Hag+02; MN11; Deu+14; Gat+12; Fuk11b; Fuk11a] and references therein. A further extension to (1.4) whilst mantaining the Markov property is the addition of Levy-jump processes (see [Tan03]), however the same two premises apply: fast calibration and simple simulation.

1.1.2 The Markovian assumption in question

In the previous section we have discovered the benefits and convenience of a Markovian framework in financial modelling. A legimitimate question, however, is whether this setting accurately represents the behaviour of markets or not. Among several stylised facts for financial markets (see [MFE15] or [Con01] for an excellent summary), it is well known that volatility exhibits persistence or memory. The seminal works by Engle [Eng82; Eng01] are clear examples of the autoregressive features of the volatility in financial markets, modelled in a discrete time setting under the physical measure $\mathbb{P}$ by

$$V_{t_i} = \sum_{k=1}^{q} \beta_k V_{t_{i-k}} + \epsilon_t, \quad q \in \mathbb{N}, \beta_k \in \mathbb{R}$$

where ϵ_t is an iid white noise process and $\{t_i\}_{i=1,\dots,n}$ is a given time partition for $n \in \mathbb{N}$. In this simple setting, it is clear that the process V does not satisfy the Markov condition. Perhaps surprisingly, for the past decades this feature has clearly been outweighed by the analytical tractability of Markovian models when considering a derivatives pricing model. Therefore, a natural approach is to explore models beyond the Markovian family with the goal of reproducing additional stylised facts of financial markets.

1.2 Rough volatility: a new breed of non-Markovian models

In the previous section we have made clear that the Markovian assumption is too restrictive in order to reproduce the behaviour of markets in all its complexity. Nonetheless, it has not been trivial to find a non-Markovian extension of (1.4). Perhaps, the most natural path to extend a (continuous-time) Markovian model to a non-Markovian counterpart is to start from the basic driving process, i.e. the Brownian motion.

The first notion of fractional Brownian motion or Wiener Helix (as was first named) is due to Kolmogorov in 1940 [Kol40] from a functional analytic point of view. Nevertheless, the modern notion of fractional Brownian motion (fBm) and its stochastic integral representation were first introduced by Mandelbrot and Van Ness [MV68] as a natural extension of the classical Brownian motion.

Definition 1.2.1. A fractional Brownian motion $\left(W_t^H\right)_{t\geq 0}$ with Hurst index $H \in (0,1)$ is a continuous Gaussian process such that $W_0^H = 0$ and covariance structure,

$$\mathbb{E}[W_t^H W_s^H] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

Definition 1.2.2 (Mandelbrot-Van Ness representation)**.** A fractional Brownian motion $\left(W_t^H\right)_{t\geq 0}$ with Hurst index $H \in (0,1)$ such that $W_0^H = 0$ admits the following stochastic integral representation:

$$W_t^H = C_H \left\{ \int_{-\infty}^{0} \left[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right] dW_s + \int_0^t (t-s)^{H-\frac{1}{2}} dW_s \right\} = C_H \int_{\mathbb{R}} \left[(t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}\right] dW_s$$

where

$$C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}} = \left(\int_0^\infty \left[(1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}}\right]^2 ds + \frac{1}{2H}\right)^{-1/2},$$

$\Gamma(\cdot)$ is the gamma function, $(W_t)_{t\in\mathbb{R}}$ a standard two-sided Brownian motion.

From the integral representaion of the fractional Brownian motion in Definition 1.2.1 it is obvious that the process is non-Markovian (except for $H = \frac{1}{2}$). The first attempt to fit a fBm into (1.4) is due to Comte and Renault in 1996 [CR96], who considered an extension of the Heston model driven by a fBm with $H \in (\frac{1}{2}, 1)$ motivated by autoregressive models of Engle type [Eng82; Eng01]. Remarkably, Comte and Renault considered the fBm at the

level of the instantaneous variance process V, but kept a standard Brownian motion as the driver for the stock process which prevents arbitrage. It was later proved by Rogers [Rog97] that if the stock process in (1.4) is driven by an fBm with $H \neq \frac{1}{2}$, it leads to arbitrage as the stock is not a semimartingale.

Alòs, León and Vives in 2007 [ALV07] discovered that in order to reproduce the empirical power law decay of the at-the-money implied volatility skew a fractional Brownian motion with $H \in (0, \frac{1}{2})$ was needed in the volatility process. Interestingly, in [ALV07] was also proven that not even Levy-jump Markovian processes generate such power law as was conjectured at the time. 4 years later, Fukasawa [Fuk11b] found similar asymptotic results supporting the need of a fBm with $H \in (0, \frac{1}{2})$ in the volatility process. We note that both approaches found evidence under the $\mathbb{Q}$ measure. Later on, Gatheral, Jaisson and Rosenbaum [GJR18] found empirical evidence under the $\mathbb{P}$ measure supporting the use of a fBm with $H \in (0, \frac{1}{2})$. The concept of "rough volatility" was also introduced in [GJR18] to refer to a model with volatility driven by a fBm with $H \in (0, \frac{1}{2})$. In particular, they introduced the so-called Rough Fractional Stochastic Volatility (RFSV) model, a lognormal type of SV model:

$$\begin{aligned} \mathrm{d}X_t^{\mathbb{P}} &= \mu_t \mathrm{d}t - \frac{1}{2} V_t \mathrm{d}t + \sqrt{V_t}\, \mathrm{d}W_t^{\mathbb{P}}, \\ V_t^{\mathbb{P}} &= V_0 \exp\left(\nu C_H \int_{\mathbb{R}} \left[(t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}} \right] dZ_s^{\mathbb{P}} \right), \quad \nu > 0, H \in \left(0, \frac{1}{2}\right) \\ \mathrm{d}[W^{\mathbb{P}}, Z^{\mathbb{P}}]_t &= \rho \mathrm{d}t, \quad \rho \in (-1, 1). \end{aligned} \tag{1.5}$$

This model, was later extended to the $\mathbb{Q}$ measure in Bayer, Friz and Gatheral [BFG16] supported by previous findings in [ALV07; Fuk11b] and led to the so-called rough Bergomi model:

$$\begin{aligned} \mathrm{d}X_t^{\mathbb{Q}} &= r_t \mathrm{d}t - \frac{1}{2} V_t \mathrm{d}t + \sqrt{V_t}\, \mathrm{d}W_t^{\mathbb{Q}}, \\ V_t^{\mathbb{Q}} &= \xi_0(t) \exp\left(\nu C_H \int_0^t (t-s)^{H-\frac{1}{2}} dZ_s^{\mathbb{Q}} \right), \quad \nu > 0, H \in \left(0, \frac{1}{2}\right) \\ \mathrm{d}[W^{\mathbb{Q}}, Z^{\mathbb{Q}}]_t &= \rho \mathrm{d}t, \quad \rho \in (-1, 1). \end{aligned} \tag{1.6}$$

The series of papers [ALV07; Fuk11b; GJR18; BFG16] perhaps brought more questions than answers to a field where the Markovian assumption is customary. Fundamental questions, such as the martingale property of the stock process in the rough Bergomi model have only recently been answered (see Gassiat [Gas19]) and the efficient implementation and asymptotic analysis of such models has opened the door to new methodologies that move away from the classical Feynman-Kac or Itô arguments that simply do not work for fBm as it is not a semimartingale. The seed of [ALV07; Fuk11b; GJR18; BFG16] has flourished into a vast literature in many different directions [Bay+19b; Bay+19a; BLP16; ER19; FZ17; FTW19; Gue+18; Gul18; HJL19; JPS18; NR18] to name a few, and the interested reader is referred to Rough volatility network for an exhaustive bibliographical archive.

The interplay between non-Markovianity and market microstructure

We have argued that the fact that market dynamics inherently exhibit a non-Markovian structure is a well stablished fact. One could question, whether such behaviour is also observed in the market at a microscopic level. Recently, several self-exciting processes of Hawkes [Haw71] type have been proposed to model the Limit Order Book (LOB) yielding very realistic dynamics and market fits (see [Bac+13; MP18b; MP18c]. Remarkably, the series of papers [JR16; ER19; TR19] have proven that such self-exciting processes converge to fractional diffusions, where the driving fBm is found to be $H \in (0, \frac{1}{2})$. This in turn, provides an additional argument under the $\mathbb{P}$ measure to favour rough volatility models over their Markovian counterpart.

1.3 Thesis Overview

The aim of the present thesis is to analyse rough volatility models of type (1.5)-(1.6) and explore their full potential to solve some classical problems in derivatives pricing. The reader should keep in mind that all results presented also apply to Markovian models, as setting the value $H = \frac{1}{2}$ allows us to recover the Markovian property.

The first part of the thesis tackles the issue of numerically implementing rough volatility models. Concretely, Chapters 2, 3 and 4 aim at making rough volatility models competitive in practice; in Chapter 2 we explore efficient Monte-Carlo methods and prove some weak convergence (although the convergence to the log-stock process still remains a conjecture in this work), whereas in Chapter 3 we apply deep learning methods to efficiently perform the calibration task. These results combined, ensure a sound numerical framework in which rough volatility models do not impose additional complexity when compared to Markovian models. To conclude, in Chapter 4 we explore a novel numerical scheme to calibrate Local Stochastic Volatility (LSV) models that applies to all SV models (including the rough volatility family) and substantially improves the reliability and performance compared to state-of-the-art methods in the field.

The second part of the thesis focuses on volatility derivatives such as VIX futures, VIX options and Realized Variance options. In Chapter 5 we characterise the dynamics of VIX in the rough Bergomi model and explore whether such model is able to jointly fit S&P 500 options and VIX futures. Following, in Chapters 6 and 7 we explore short-time asymptotics of volatility products in rough volatility models and their multi-factor extensions. On the one hand, in Chapter 6 we follow a Malliavin calculus approach, whereas on the other hand, in Chapter 7 we follow of Large Deviations Principle (LDP) viewpoint. Interestingly, we find that both approached combined yield some fascinating results.

The third and last part of the thesis addresses the foundations of the change of measure between the physical measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$ for rough volatility models. In Chapter 8 we investigate sufficient conditions for the change of measure to be well defined and present a first attempt to numerically estimate the risk premium process.

The following table summarises the stage of the preprints/publications that this thesis builds upon.

Chapter	Preprint/ Publication	Stage
2	[HJM17] B. Horvath, A. Jacquier and A. Muguruza. Functional central limit theorems for rough volatility. arXiv preprint: arXiv:1711.03078, 2017	2nd Revision
3	[Bay+19c] C. Bayer, B. Horvath, A. Muguruza, B. Stemper and M. Tomas. On deep calibration of (rough) stochastic volatility models. arXiv preprint: arXiv:1908.08806, 2019.	1st Revision
3	[HMT21] B. Horvath, A. Muguruza and M. Tomas. Deep Learning Volatility. *Quantitative Finance*, 21(1): 11-27, 2021.	Published
4	[Mug19] A. Muguruza. Not so particular about calibration: smile problem resolved. arXiv preprint: arXiv:1909.13366, 2019.	2nd Revision
5	[JMM18] A. Jacquier, C. Martini and A. Muguruza. On VIX futures in the rough Bergomi model. *Quantitative Finance*, 18(1): 45-61, 2018.	Published
6	[AGM18] E. Alòs, D. García-Lorite and A. Muguruza. On smile properties of volatility derivatives and exotic products: understanding the VIX skew. arXiv:1808.03610, 2018.	Accepted in SIAM
7	[LMS21] C. Lacombe, A. Muguruza and H. Stone. Asymptotics for volatility derivatives in multi-factor rough volatility models. *Mathematics and Financial Economics*, 15: 545-577, 2021.	Published

2

Functional Central Limit Theorems for Rough Volatility

2.1 Introduction

As discussed in the introduction, rough volatility models have led to theoretical and practical challenges to understand and implement them. One of the main issues, at least from a practical point of view, is on the numerical side: absence of Markovianity rules out PDE-based schemes, and simulation is the only possibility. However, classical simulation methods for fractional Brownian motion (based on Cholesky decomposition or circulant matrices) are notoriously slow, and faster techniques are needed. The state of the art, so far, is the recent hybrid scheme developed by Bennedsen, Pakkanen and Lunde [BLP17], and its turbocharged version [MP18a]. We rise here to this challenge, and propose an alternative tree-based approach, mathematically rooted in an extension of Donsker's theorem to rough volatility.

Donsker [Don51] (and later Lamperti [Lam62b]) proved a functional central limit for Brownian motion, thereby providing a theoretical justification of its random walk approximation. Many extensions have been studied in the literature, and we refer the interested reader to [Dud99] for an overview. In the fractional case, Sottinen [Sot01] and Nieminen [Nie04] constructed–following Donsker's ideas of using iid sequences of random variables– an approximating sequence converging to the fractional Brownian motion, with Hurst parameter $H > 1/2$. In order to deal with the non-Markovian behaviour of fractional Brownian motion, Taqqu [Taq75] considered sequences of non-iid random variables, again with the restriction $H > 1/2$. Unfortunately, neither methodologies seem to carry over to the 'rough' case $H < 1/2$, mainly because of the topologies involved. The recent development of rough paths theory [FV10; Lyo98] provided an appropriate framework to extend Donsker's results to processes with sample paths of Hölder regularity strictly smaller than $1/2$. For $H \in (1/3, 1/2)$, Bardina, Nourdin, Rovira and Tindel [Bar+10] used rough paths to show that functional central limit theorems (in the spirit of Donsker) apply. This in particular suggests that the natural topology at work for rough fractional Brownian motion is the topology induced by the Hölder norm of the sample paths. Indeed, switching the topology from the Skorokhod one used by Donsker to the (stronger) Hölder topology is the right setting for rough central limit theorems, as we outline in this Chapter. Recent results [BP12; Par17; Par14] provide convergence for (geometric) fractional Brownian motions with general $H \in (0, 1)$ using Wick calculus, assuming that the approximating sequences are Bernoulli random variables. We conjecture 2.2.11 this to an universal functional central limit theorem, involving general (discrete or continuous) random variables as approximating sequences, only requiring finiteness of moments.

We consider a general class of continuous processes with any Hölder regularity, including fractional Brownian motion with $H \in (0, 1)$, truncated Brownian semi-stationary processes, Gaussian Volterra processes, as well as rough volatility models recently proposed in the financial literature. The fundamental novelty here is an approximating

sequence capable of simultaneously keeping track of the approximated rough volatility process (fractional Brownian motion, Brownian semistationary process, or any continuous path functional thereof) and of the underlying Brownian motion. This is crucial in order to take into account the correlation of the two processes, the so-called leverage effect in financial modelling. While approximations of two-dimensional (correlated) semimartingales are well understood in the standard case, the rough case is so far an open problem. Our analysis easily generalises beyond Brownian drivers to more general semimartingales, emphasising that the subtle, yet essential, difficulties lie in the passage from the semimartingale setup to the rough case. This is the first Monte-Carlo method available in the literature, specifically tailored to two-dimensional rough systems, based on an approximating sequence for which we conjecture a Donsker-Lamperti-type functional central limit theorem (FCLT). This conjecture further provides a pathwise justification of the hybrid scheme by Bennedsen, Lunde and Pakkanen [BLP17], and to develop tree-based schemes, opening the doors to pricing early-exercise options such as American options. In Section 2.2, we present the class of models we are considering and state our main results and conjectures that are left for future research. The proof of the main conjecture is developed in Section 2.3 in several steps.In spite of the full convergence being conjectured in this thesis, several steps are formally proven. We reserve Section 2.4 to applications of the main conjecture, namely weak convergence of the hybrid scheme, binomial trees as well as numerical examples. We present simple numerical recipes, providing a pedestrian alternative to the advanced hybrid schemes in [BLP17; MP18a], and develop a simple Monte-Carlo with low implementation complexity, for which we provide comparison charts against [BLP17] in terms of accuracy and against [MP18a] in terms of speed. Reminders on Riemann-Liouville operators and additional technical proofs are postponed to the appendix.

Notations: On the interval $\mathbb{I} := [0,1]$, $\mathcal{C}_{\mathbb{R}}(\mathbb{I})$ and $\mathcal{C}^{\alpha}_{\mathbb{R}}(\mathbb{I})$ denote the spaces of real-valued continuous and α-Hölder continuous functions on $\mathbb{I}$ with Hölder regularity $\alpha \in (0,1)$; $\mathcal{C}^{\infty}_{\mathbb{R}^+}(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R}^+$ such that all derivatives of f exist and are continuous on$\mathbb{I}\}$, $\mathcal{C}^{1}_{\mathbb{R}}(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R}$ such that f' exists and is continuous on $\mathbb{I}\}$ and $\mathcal{C}^{1}_{\mathbb{R}^+}(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R}^+$ such that f' exists and is continuous on $\mathbb{I}\}$. All definitions imply bounded first order derivatives on $\mathbb{I}$. We use $C, \widetilde{C}, \hat{C}, C_1, C_2, \overline{C}, \underline{C}$ as strictly positive real constants which may change from line to line, the exact values of which do not matter.

2.2 Weak convergence of rough volatility models

Donsker's invariance principle [Don51] (also termed 'functional central limit theorem') ensures the weak convergence of an approximating sequence to a Brownian motion in the Skorokhod space. As opposed to the central limit theorem, Donsker's theorem is a pathwise statement which ensures that convergence takes place for all times. This result is particularly important for Monte-Carlo methods, which aim to approximate pathwise functionals of a given process (essential requirement in order to price path-dependent financial securities for example). We prove here a version of Donsker's result, not only in the Skorokhod topology, but also in the stronger Hölder topology, for a general class of continuous stochastic processes.

2.2.1 Hölder spaces and fractional operators

For $\beta \in (0,1]$, the β-Hölder space $\mathcal{C}^{\beta}_{\mathbb{R}}(\mathbb{I})$, with the norm

$$\|f\|_{\beta} := |f|_{\beta} + \|f\|_{\infty} = \sup_{\substack{t,s\in\mathbb{I}\\ t\neq s}} \frac{|f(t)-f(s)|}{|t-s|^{\beta}} + \max_{t\in\mathbb{I}} |f(t)|,$$

is a non-separable Banach space [Kry96, Chapter 3]. Following the spirit of Riemann-Liouville fractional operators recalled in Appendix A.2, we introduce the class of Generalised Fractional Operators (GFO). For $\lambda \in (0,1)$ we introduce the family of intervals

$$\left\{\mathfrak{R}^{\lambda}\right\}_{\lambda\in(0,1)} := \left\{(-\lambda, 1-\lambda)\right\}_{\lambda\in(0,1)},$$

$\mathfrak{R}^\lambda_+ := \mathfrak{R}^\lambda \cap (0,1)$, $\mathfrak{R}^\lambda_- := \mathfrak{R}^\lambda \cap (-1,0)$, and the space $\mathcal{L}^\alpha := \left\{u \mapsto u^\alpha L(u) : L \in \mathcal{C}^\infty_{\mathbb{R}^+}(\mathbb{I}) \text{ and } L'(u) \leq 0\right\}$, for any $\alpha \in \mathfrak{R}^\lambda$.

Definition 2.2.1. For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, the GFO associated to $g \in \mathcal{L}^\alpha$ is defined on $\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I})$ as

$$(\mathcal{G}^\alpha f)(t) := \begin{cases} \displaystyle\int_0^t (f(s) - f(0)) \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s, & \text{if } \alpha \in (0, 1-\lambda), \\ \displaystyle\frac{\mathrm{d}}{\mathrm{d}t} \int_0^t (f(s) - f(0)) g(t-s)\mathrm{d}s, & \text{if } \alpha \in (-\lambda, 0). \end{cases} \tag{2.1}$$

We shall further use the notation $G(t) := \int_0^t g(u)\mathrm{d}u$, for any $t \in \mathbb{I}$. Of particular interest in mathematical finance are the following kernels and operators:

$$\begin{array}{lll} \text{Riemann-Liouville:} & g(u) = u^\alpha, & \text{for } \alpha \in (-1,1); \\ \text{Gamma fractional:} & g(u) = u^\alpha \mathrm{e}^{\beta u}, & \text{for } \alpha \in (-1,1), \beta < 0; \\ \text{Power-law:} & g(u) = u^\alpha (1+u)^{\beta-\alpha}, & \text{for } \alpha \in (-1,1), \beta < -1. \end{array} \tag{2.2}$$

Remark 2.2.2. We note that the following equivalent representation (Lemma 13.1 in Samko, Kilbas and Marichev [SKM+93] for details) holds for $\alpha \in (-\lambda, 0)$

$$(\mathcal{G}^\alpha f)(t) := (f(t) - f(0))\, g(t) - \int_0^t (f(t) - f(s))\, g'(t-s)\mathrm{d}s$$

The following theorem generalises the classical mapping properties of Riemann-Liouville fractional operators first proved by Hardy and Littlewood [HL28], and will be of fundamental importance in the rest of our analysis.

Proposition 2.2.3. *For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, the operator $\mathcal{G}^\alpha : \mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I}) \to \mathcal{C}^{\lambda+\alpha}_{\mathbb{R}}(\mathbb{I})$ is continuous.*

Proof. The proof directly follows from Friz and Hairer [FH20, Theorem 14.17]. □

We develop here an approximation scheme for the following system, generalising the concept of rough volatility introduced in [ALV07; Fuk11b; GJR18] in the context of mathematical finance, where the process X represents the dynamics of the logarithm of a stock price process:

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2} V_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t, \quad X_0 = 0, \\ V_t &= \Phi\left(\mathcal{G}^\alpha Y\right)(t), \end{aligned} \tag{2.3}$$

with $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, and Y the (strong) solution to the stochastic differential equation

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + a(Y_t)\mathrm{d}W_t, \quad Y_0 \in \mathcal{D}_Y, \tag{2.4}$$

where $\mathcal{D}_Y$ denotes the state space of the process Y, usually $\mathbb{R}$ or $\mathbb{R}_+$. The two Brownian motions B and W, defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{I}}, \mathbb{P})$, are correlated by the parameter $\rho \in [-1,1]$, and $\Phi : \mathcal{C}^1(\mathbb{I}) \to \mathcal{C}^{1+}(\mathbb{I})$. This in particular, implies that whenever Y is in $C^\lambda(\mathbb{I})$ then V is in $C^{\alpha+\lambda}(\mathbb{I})$ and non-negative. It remains to formulate the precise definition for $\mathcal{G}^\alpha Y$ (Proposition 2.2.5) to fully specify the system (2.3) and clarify the existence of solutions.

Assumption 2.2.4. There exist $C_b, C_a > 0$ such that, for all $x, y \in \mathcal{D}_Y$,

$$|b(x)| \leq C_b(1+|x|) \qquad \text{and} \qquad |a(x)| \leq C_a(1+|x|).$$

Existence of solutions to (2.4) along with Assumption 2.2.4 need to be checked on a case by case basis beyond standard Lipschitz and linear growth conditions (we provide a showcase of models in Examples 2.2.6-2.2.9 below satisfying existence and pathwise uniqueness conditions). General conditions for stochastic invariance can be found in [AD03; BJ02; DF07] for diffusions, and in [JLP19] for affine Volterra processes. Not only is the solution

to (2.4) continuous, but $(\frac{1}{2}-\varepsilon)$-Hölder continuous for any $\varepsilon \in (0, \frac{1}{2})$ as a consequence of the Kolmogorov-Čentsov theorem [Čen56]. Existence and precise meaning of $\mathcal{G}^\alpha Y$ is delicate, and is treated below.

2.2.2 Examples

Before constructing our approximation scheme, let us discuss a few examples of processes within our framework. As a first useful application, these generalised fractional operators render a (continuous) mapping between a standard Brownian motion and its fractional counterpart:

Proposition 2.2.5. *For any $\alpha \in \mathfrak{R}^{1/2}$, the equality $(\mathcal{G}^\alpha W)(t) = \int_0^t g(t-s)\mathrm{d}W_s$ holds almost surely for $t \in \mathbb{I}$.*

We refer to Appendix A.3.1 for the proof. Modulo a constant multiplicative factor C_α, the (left) fractional Riemann-Liouville operator (Appendix A.2) is identical to the GFO in (2.2), so that the Riemann-Liouville (or Type-II) fractional Brownian motion can be written as $C_\alpha \mathcal{G}^\alpha W$. Furthermore, Proposition 2.2.3 yields that the Riemann-Liouville operator is continuous from $\mathcal{C}^{1/2}_{\mathbb{R}}(\mathbb{I})$ to $\mathcal{C}^{1/2+\alpha}_{\mathbb{R}}(\mathbb{I})$ for $\alpha \in \mathfrak{R}^{1/2}$. Each kernel in (2.2) gives rise to processes proposed in turbulence modelling and in mathematical finance by Barndorff-Nielsen and Schmiegel [BS07].

Example 2.2.6. The rough Bergomi model introduced by Bayer, Friz and Gatheral [BFG16] reads

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha \mathrm{d}W_s\right),$$

with $V_0, \nu, \xi_0(\cdot) > 0$, $\alpha \in \mathfrak{R}^{1/2}$ and $\mathcal{E}(\cdot)$ is the Wick stochastic exponential. This corresponds exactly to (2.3) with $g(u) \equiv u^\alpha$, $Y = W$ and

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left\{-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha}\mathrm{d}s\right\}.$$

Example 2.2.7. A truncated Brownian semistationary ($\mathfrak{t}\mathcal{BSS}$) process is defined as $\int_0^t g(t-s)\sigma(s)\mathrm{d}W_s$, for $t \in \mathbb{I}$, where σ is $(\mathcal{F}_t)_{t\in\mathbb{I}}$-predictable with locally bounded trajectories and finite second moments, and $g : \mathbb{I}\setminus\{0\} \to \mathbb{I}$ is Borel measurable and square integrable. If $\sigma \in \mathcal{C}^1_{\mathbb{R}}(\mathbb{I})$, this class falls within the GFO framework.

Example 2.2.8. Bennedsen, Lunde and Pakkanen [BLP16] considered adding a Gamma kernel to the volatility process, which yields the Truncated Brownian semi-stationary (Bergomi-type) model:

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha \mathrm{e}^{-\beta(t-s)}\mathrm{d}W_s\right),$$

with $\beta > 0$, $\alpha \in \mathfrak{R}^{1/2}$. This corresponds to (2.3) with $Y = W$, Gamma fractional kernel $g(u) \equiv u^\alpha \mathrm{e}^{-\beta u}$ in (2.2),

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left\{-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha}\mathrm{e}^{-2\beta(t-s)}\mathrm{d}s\right\}.$$

Example 2.2.9. The rough Heston model introduced by Guennoun, Jacquier, Roome and Shi [Gue+18] reads

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \kappa(\theta - Y_s)\mathrm{d}t + \int_0^t \xi\sqrt{Y_s}\mathrm{d}W_s, \\ V_t &= \eta + \int_0^t (t-s)^\alpha \mathrm{d}Y_s, \end{aligned}$$

with $Y_0, \kappa, \xi, \theta > 0$, $2\kappa\theta > \xi^2$ and $\eta > 0$, $\alpha \in \mathfrak{R}^{1/2}$. This corresponds exactly to (2.3) with $g(u) \equiv u^\alpha$, $\Phi(\varphi)(t) := \eta + \varphi(t)$, and the coefficients of (2.4) read $b(y) \equiv \kappa(\theta - y)$ and $a(y) \equiv \xi\sqrt{y}$. This model is markedly different from the rough Heston introduced by El Euch and Rosenbaum [ER19] (for which the characteristic function is known in semi-closed form). Unfortunately, this second version is out of the scope of our invariance principle.

2.2.3 The approximation scheme

We now move on to the core of the project, namely an approximation scheme for the system (2.3). The basic ingredient to construct approximating sequences is a family of iid random variables, which satisfies the following assumption:

Assumption 2.2.10. The family $(\xi_i)_{i\geq 1}$ forms an iid sequence of centered random variables with finite moments of all orders and $\mathbb{E}[\xi_1^2] = \sigma^2 > 0$.

Following Donsker [Don51] and Lamperti [Lam62b], we first define, for any $\omega \in \Omega$, $n \geq 1$, $t \in \mathbb{I}$, the approximating sequence for the driving Brownian motion B as

$$B_n(t) := \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}. \tag{2.5}$$

As will be explained later, a similar construction holds to approximate the process Y:

$$Y_n(t) := \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} b\left(Y_n^{k-1}\right) + \frac{nt - \lfloor nt \rfloor}{n} b\left(Y_n^{\lfloor nt \rfloor}\right) + \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} a\left(Y_n^{k-1}\right) \zeta_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} a\left(Y_n^{\lfloor nt \rfloor}\right) \zeta_{\lfloor nt \rfloor + 1}, \tag{2.6}$$

where $Y_n(0) = Y_0$, $Y_n^k := Y_n(t_k)$ and $\mathcal{T}_n := \{t_k = \frac{k}{n}\}_{k=0,\dots,n}$. Here $\{\xi\}_{i=1}^{\lfloor nt \rfloor}$ and $\{\zeta\}_{i=1}^{\lfloor nt \rfloor}$ satisfy Assumption 2.2.10, with appropriate correlation structure between the pairs $\{(\zeta_i, \xi_i)\}_{i=1}^{\lfloor nt \rfloor}$ that will be made precise later. We shall always use (ξ_i) to denote the sequence generating B and (ζ_i) the one generating W. Consequently, we deduce an approximating scheme (up to the interpolating term which decays to zero by Chebychev's inequality) for X as

$$X_n(t) \quad := -\frac{1}{2n} \sum_{k=1}^{\lfloor nt \rfloor} \Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k) + \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(t_k)} \xi_k. \tag{2.7}$$

All the approximations above, should be understood pathwise, but we omit the ω dependence in the notations for clarity. The following result, discussed in Section 2.3.4, conjectures convergence of the approximating sequence $(X_n)_{n\geq 1}$. As usual in weak convergence analysis [Bil68], convergence is stated in the Skorokhod space $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$ of real-valued càdlàg processes equipped with the Skorokhod topology.

Conjecture 2.2.11. The sequence $(X_n)_{n\geq 1}$ converges weakly to X in $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$.

In principle, extensions to the case where Y is a d-dimensional diffusion should not impose additional work given that X is a 2-dimensional system itself. Although the result is not proven in this work, we do present a number of auxiliary results which we believe are necessary to prove the conjecture in a future work: we start with the convergence of the approximation (Y_n) in some Hölder space, which we translate, first into convergence of the stochastic integral in (2.3), then, by continuity of the mapping Φ, into convergence of the sequence $(\Phi(\mathcal{G}^\alpha Y_n))$. All these ingredients are detailed in Section 2.3 below. Once this is achieved, we discuss a potential approach to proving conjecture 2.2.11 which is left for future research.

2.3 Functional Central limit theorems for a family of Hölder continuous processes

2.3.1 Weak convergence of Brownian motion in Hölder spaces

Donsker's classical convergence result was proven under the Skorokhod topology. We concentrate here on convergence in the Hölder topology, due to Lamperti [Lam62a]. The standard convergence result for Brownian motion

can be stated as follows:

Theorem 2.3.1. *For $\lambda < \frac{1}{2}$, the sequence (B_n) in (2.5) converges weakly to a Brownian motion in $\left(\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_\lambda\right)$.*

The proof relies on finite-dimensional convergence and tightness of the approximating sequence. Not surprisingly, the tightness criterion [Bil68] in the Skorokhod space $\mathcal{D}_{\mathbb{R}}(\mathbb{I})$ and in a Hölder space setting are different. In fact, the tightness criterion in Hölder space is strictly related to Kolmogorov-Čentsov's continuity [Čen56]. Note in passing, that the approximating sequence (2.5) is piecewise differentiable in time for each $n \geq 1$ even though its limit is obviously not. The proof of Theorem 2.3.1 follows from Theorem 2.3.3 under Assumption 2.2.10.

Theorem 2.3.2 (Sufficient conditions for weak convergence in Hölder spaces (Račkauskas-Suquet [RS04]))**.** *Let $Z \in \mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I})$ and $(Z_n)_{n\geq 1}$ an approximating sequence in the sense that for any sequence $(\tau_k)_k$ in $\mathbb{I}$, $(Z_n(\tau_k))_k$ converges in distribution to $(Z(\tau_k))_k$ as n tends to infinity. If*

$$\mathbb{E}\left[|Z_n(t) - Z_n(s)|^\gamma\right] \leq C|t-s|^{1+\beta} \tag{2.8}$$

holds for all $n \geq 1$, $t, s \in \mathbb{I}$, and some $C, \gamma, \beta > 0$. Then $(Z_n)_{n\geq 1}$ converges weakly to Z in $\mathcal{C}^\mu_{\mathbb{R}}(\mathbb{I})$ for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

The proof of this theorem relies on results of Račkauskas and Suquet [RS04], who prove the convergence in the Hölder space $C^\lambda_0(\mathbb{I})$ endowed with the norm $\|f\|^0_\lambda := |f|_\lambda + |f(0)|$, for all functions that satisfy

$$\lim_{\delta \downarrow 0} \sup_{\substack{0<t-s<\delta \\ t,s\in\mathbb{I}}} \frac{|f(t)-f(s)|}{(t-s)^\gamma} = 0.$$

From here the proof of Theorem 2.3.2 is a straightforward consequence, since $\left(C^\lambda_0(\mathbb{I}), \|\cdot\|^0_\lambda\right)$ is a separable closed subspace of $\left(\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_\lambda\right)$ (see [Ham00; RS04] for details), and one can then use the simple tightness criterion introduced above to conclude the proof of Theorem 2.3.2. Moreover, as the identity map from $C^\lambda_0(\mathbb{I})$ into $\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I})$ is continuous, weak convergence in the former implies weak convergence in the latter. To conclude our review of weak convergence in Hölder spaces, the following theorem, due to Račkauskas and Suquet [RS04] provides necessary and sufficient conditions ensuring convergence in Hölder space:

Theorem 2.3.3. *[Račkauskas-Suquet [RS04]] For any $\lambda \in (0, \frac{1}{2})$, the sequence $(B_n)_{n\geq 1}$ in (2.5) converges weakly to a Brownian motion in $\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I})$ if and only if $\mathbb{E}[\xi_1] = 0$ and $\lim_{t\uparrow\infty} t^{\frac{1}{1-2\lambda}} \mathbb{P}(|\xi_1| \geq t) = 0$.*

Assumption 2.2.10 ensures the conditions in Theorem 2.3.3. The following statement allows us to apply Theorem 2.3.2 on $\mathbb{I}$ and extend the Hölder convergence result via linear interpolation to a continuous sequence.

Theorem 2.3.4. *Let $Z \in \mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I})$ and $(Z_n)_{n\geq 1}$ an approximation sequence such that finite dimensional convergence holds. Moreover, if*

$$\mathbb{E}\left[|Z_n(t_i) - Z_n(t_j)|^\gamma\right] \leq C\,|t_i - t_j|^{1+\beta}, \tag{2.9}$$

for any $t_i, t_j \in \mathcal{T}_n$ and some $\beta, \gamma, C > 0$, then the linear interpolating sequence

$$\overline{Z}_n(t) := Z_n\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(Z_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - Z_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right)$$

converges weakly to Z in $\mathcal{C}^\mu_{\mathbb{R}}(\mathbb{I})$ for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

Proof. For any $t, s \in \mathbb{I}$, we can write, letting $Z^k_n := Z_n(t_k)$ and $\overline{Z}^k_n := \overline{Z}_n(t_k)$,

$$\begin{aligned}
\mathbb{E}\left[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma\right] &= \mathbb{E}\left[\left|Z^{\lfloor nt \rfloor}_n + (nt - \lfloor nt \rfloor)\left(Z^{\lfloor nt \rfloor + 1}_n - Z^{\lfloor nt \rfloor}_n\right) - Z^{\lfloor ns \rfloor}_n - (ns - \lfloor ns \rfloor)\left(Z^{\lfloor ns \rfloor + 1}_n - Z^{\lfloor ns \rfloor}_n\right)\right|^\gamma\right] \\
&\leq 3^{\gamma-1}\mathbb{E}\left[\left|Z^{\lfloor nt \rfloor}_n - Z^{\lfloor ns \rfloor}_n\right|^\gamma + (nt - \lfloor nt \rfloor)^\gamma \left|Z^{\lfloor nt \rfloor + 1}_n - Z^{\lfloor nt \rfloor}_n\right|^\gamma + (ns - \lfloor ns \rfloor)^\gamma \left|Z^{\lfloor ns \rfloor + 1}_n - Z^{\lfloor ns \rfloor}_n\right|^\gamma\right] \\
&\leq C\left(\left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n}\right)^{1+\beta} + \frac{(nt - \lfloor nt \rfloor)^\gamma}{n^{1+\beta}} + \frac{(ns - \lfloor ns \rfloor)^\gamma}{n^{1+\beta}}\right) \leq C(t-s)^{1+\beta},
\end{aligned}$$

where we used (2.9) and the fact that $\frac{\lfloor nt\rfloor-\lfloor ns\rfloor}{n}\leq 2(t-s)$, $nt-\lfloor nt\rfloor\leq 1$ for $t\geq 0$ and $\frac{1}{n}\leq(t-s)$. Finally, it is left to prove the case $\frac{1}{n}>(t-s)$. There are two possible scenarios here:

- if $\lfloor nt\rfloor=\lfloor ns\rfloor$, then

$$\mathbb{E}[|\overline{Z}_n(t)-\overline{Z}_n(s)|^\gamma]=\mathbb{E}\left[\left|(nt-ns)\left(Z_n^{\lfloor nt\rfloor+1}-Z_n^{\lfloor nt\rfloor}\right)\right|^\gamma\right]\leq\frac{C(t-s)^\gamma}{n^{1+\beta-\gamma}}\leq C(t-s)^{1+\beta};$$

- if $\lfloor nt\rfloor\neq\lfloor ns\rfloor$, then either $\lfloor nt\rfloor+1=\lfloor ns\rfloor$ or $\lfloor nt\rfloor=\lfloor ns\rfloor+1$. Without loss of generality consider the second case. Then

$$\begin{aligned}\mathbb{E}\left[|\overline{Z}_n(t)-\overline{Z}_n(s)|^\gamma\right]&=\mathbb{E}\left[\left|\overline{Z}_n(t)-Z_n^{\lfloor nt\rfloor}+Z_n^{\lfloor nt\rfloor}-\overline{Z}_n(s)\right|^\gamma\right]\leq 2^{\gamma-1}\mathbb{E}\left[\left|\overline{Z}_n(t)-Z_n^{\lfloor nt\rfloor}\right|^\gamma+\left|Z_n^{\lfloor nt\rfloor}-\overline{Z}_n(s)\right|^\gamma\right]\\&\leq C\left((t-s)^{1+\beta}+\mathbb{E}\left[\left|(\lfloor nt\rfloor-ns)\left(Z_n^{\lfloor nt\rfloor}-Z_n^{\lfloor nt\rfloor-1}\right)\right|^\gamma\right]\right),\end{aligned}$$

  and the result follows as before since $t-\frac{\lfloor nt\rfloor}{n}<|t-s|$ and $|s-\frac{\lfloor nt\rfloor}{n}|\leq|t-s|$.

□

2.3.2 Weak convergence of Itô diffusions in Hölder spaces

The first important step in our analysis is to extend Donsker-Lamperti's weak convergence from Brownian motion to the Itô diffusion Y in (2.4).

Theorem 2.3.5. *The sequence $(Y_n)_{n\geq 1}$ in (2.6) converges weakly to Y in (2.4) in $\left(\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I}),\|\cdot\|_\lambda\right)$ for all $\lambda<\frac{1}{2}$.*

Proof. Finite-dimensional convergence is a classical result by Kushner [Kus74], so only tightness needs to be checked. In particular, using Theorem 2.3.4 we need only consider the partition $\mathcal{T}_n$. Thus, we get

$$\begin{aligned}\mathbb{E}\left[|Y_n^j-Y_n^i|^{2p}\right]&=\mathbb{E}\left[\left|\sum_{k=i+1}^j\frac{1}{n}b\left(Y_n^{k-1}\right)+\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k-1}\right)\zeta_k\right|^{2p}\right]\\&\leq 2^{2p-1}\left\{\mathbb{E}\left[\left|\sum_{k=i+1}^j\frac{1}{n}b\left(Y_n^{k-1}\right)\right|^{2p}\right]+\mathbb{E}\left[\left|\sum_{k=i+1}^j\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k-1}\right)\zeta_k\right|^{2p}\right]\right\}\\&\leq 2^{2p-1}\left\{\mathbb{E}\left[\left|\sum_{k=i+1}^j\frac{1}{n}b\left(Y_n^{k-1}\right)\right|^{2p}\right]+C(p)\mathbb{E}\left[\left|\sum_{k=i+1}^j\frac{1}{\sigma^2 n}\left(a\left(Y_n^{k-1}\right)\zeta_k\right)^2\right|^p\right]\right\}\\&\leq 2^{2p-1}\left\{\frac{(j-i)^{2p-1}}{n^{2p}}\sum_{k=i+1}^j\mathbb{E}\left[\left|b\left(Y_n^{k-1}\right)\right|^{2p}\right]+\frac{(j-i)^{p-1}}{n^p}C(p)\frac{\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\sum_{k=i+1}^j\mathbb{E}\left[\left(a\left(Y_n^{k-1}\right)\right)^{2p}\right]\right\}\\&\leq 2^{2p-1}\frac{(j-i)^{p-1}}{n^p}\left\{\sum_{k=i+1}^j\left(C_b^{2p}\mathbb{E}\left[(1+|Y_n^{k-1}|)^{2p}\right]+C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\mathbb{E}\left[\left(1+|Y_n^{k-1}|\right)^{2p}\right]\right)\right\}\\&\leq\max\left(C_b^{2p},C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\frac{(j-i)^{p-1}}{n^p}\left\{(j-i)+\sum_{k=i+1}^j\left(\mathbb{E}\left[|Y_n^{k-1}|)^{2p}\right]\right)\right\}\\&\leq\max\left(C_b^{2p},C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp\left(\sum_{k=i+1}^j 2^{2p}\frac{(j-i)^{p-1}}{n^p}\right)(t_j-t_i)^p\\&\leq\max\left(C_b^{2p},C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp\left(2^{2p}\right)(t_j-t_i)^p:=\mathfrak{C}(p)(t_j-t_i)^p,\end{aligned}$$

where we have used the discrete version of the BDG inequality [BS15, Theorem 6.3] in the martingale term $\sum_{k=i+1}^j\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k-1}\right)\zeta_k$ with $C(p):=6^p(p-1)^{p-1}$. We note that if we define the discrete time martingale process $(x_n^{i,j})_u:=\sum_{k=1}^u\frac{1}{\sigma\sqrt{n}}a\left(Y_n^{k+i-1}\right)\zeta_{i+k}$ for $u\in\{1,...,j-i\}$, we have that $|(x_n^{i,j})_{j-i}|\leq\max_{u\in\{1,..,j-i\}}|(x_n^{i,j})_u|$ and the

BDG inequality clearly also applies to $|x_{j-i}^{i,j}|$. Then, we have used independence of ζ_k with respect to Y_{k-1} and the linear growth of $b(\cdot)$ and $a(\cdot)$. Next, we used Hölder inequality and in the last step the discrete version of Gronwall's lemma [Cla87]. We note that $\mathbb{E}[\zeta_k^{2p}]$ is bounded by Assumption 2.2.10 and the constant $\mathfrak{C}(p)$ only depends on p, but not on n. Consequently, the tightness criterion (2.8) holds for $p > 1$ with $\gamma = 2p$ and $\beta = p - 1$. □

2.3.3 Invariance principle for rough processes

We have set the ground to extend our results to processes that are not necessarily $(1/2 - \varepsilon)$-Hölder continuous, Markovian nor semimartingales. More precisely, we are interested in α-Hölder continuous paths with $\alpha \in (0, 1)$, such as Riemann-Liouville fractional Brownian motion or some $t\mathcal{BSS}$ processes. A key tool is the Continuous Mapping Theorem, first proved by Mann and Wald [MW43], which establishes the preservation of weak convergence under continuous operators.

Theorem 2.3.6 (Continuous Mapping Theorem)**.** *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be two normed spaces and assume that $g : \mathcal{X} \to \mathcal{Y}$ is a continuous operator. If the sequence of random variables $(Z_n)_{n\geq 1}$ converges weakly to Z in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, then $(g(Z_n))_{n\geq 1}$ also converges weakly to $g(Z)$ in $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$.*

Many authors have exploited the combination of Theorems 2.3.1 and 2.3.6 to prove weak convergence [Pol84, Chapter IV]. This path avoids the lengthy computations of tightness and finite-dimensional convergence in classical proofs [Bil68]. In fact, Hamadouche [Ham00] already realised that Riemann-Liouville fractional operators are continuous, hence Theorem 2.3.6 holds under mapping by Hölder continuous functions. In contrast, the novelty here is to consider the family of GFO applied to Brownian motion together with the extension of Brownian motion to Itô diffusions. In fact, minimal changes to the proof of Proposition 2.2.5 yield the following:

Corollary 2.3.7. *If Y solves (2.4), then $(\mathcal{G}^\alpha Y)(t) = \int_0^t g(t-s)\mathrm{d}Y_s$ almost surely for all $t \in \mathbb{I}$ and $\alpha \in \mathfrak{R}^{\frac{1}{2}}$.*

The analogue of Theorem 2.3.5 for $\mathcal{G}^\alpha Y$ follows by continuous mapping along with the fact that $\mathcal{G}^\alpha$ is a continuous operator from $\left(\mathcal{C}_{\mathbb{R}}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $\left(\mathcal{C}_{\mathbb{R}}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha}\right)$ for all $\lambda \in (0, 1)$ and $\alpha \in \mathfrak{R}^\lambda$.

Theorem 2.3.8 (Generalised rough Donsker)**.** *For (Y_n) in (2.6), Y its weak limit in $\left(\mathcal{C}_{\mathbb{R}}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ for $\lambda < \frac{1}{2}$,*

$$\left(\mathcal{G}^\alpha Y_n\right)(t) = \sum_{i=1}^{\lfloor nt\rfloor} n\left[G\left(t - t_{i-1}\right) - G\left(t - t_i\right)\right]\left(Y_n^i - Y_n^{i-1}\right) + nG\left(t - t_{\lfloor nt\rfloor}\right)\left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right), \quad t \in \mathbb{I}, \tag{2.10}$$

converges weakly to $\mathcal{G}^\alpha Y$ in $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$ for any $\alpha \in \mathfrak{R}^\lambda$.

Proof. Recall that the sequence (2.6) is piecewise differentiable in time. For $\alpha \in \mathfrak{R}_+^\lambda$, integration by parts [Wei74, Section 2.4] (where Y_n is piecewise differentiable) yields, for $n \geq 1$ and $t \in \mathbb{I}$,

$$\begin{aligned}
\left(\mathcal{G}^\alpha Y_n\right)(t) &= \int_0^t g'(t-s)Y_n(s)\mathrm{d}s = \int_0^t g(t-s)\frac{\mathrm{d}Y_n(s)}{\mathrm{d}s}\mathrm{d}s \\
&= \frac{1}{\sigma\sqrt{n}}\left[\sum_{i=1}^{\lfloor nt\rfloor} n\int_{t_{i-1}}^{t_i} g(t-s)a\left(Y_n^{i-1}\right)\zeta_i\mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^{t} g(t-s)a\left(Y_n^{\lfloor nt\rfloor}\right)\zeta_{\lfloor nt\rfloor+1}\mathrm{d}s\right] \\
&+ \frac{1}{n}\left[n\sum_{i=1}^{\lfloor nt\rfloor}\int_{t_{i-1}}^{t_i} g(t-s)b\left(Y_n^{i-1}\right)\mathrm{d}s + n\int_{t_{\lfloor nt\rfloor}}^{t} g(t-s)b\left(Y_n^{\lfloor nt\rfloor}\right)\mathrm{d}s\right] \\
&= \sum_{i=1}^{\lfloor nt\rfloor} n\left[G(t-t_{i-1}) - G(t-t_i)\right]\left(Y_n^i - Y_n^{i-1}\right) + nG(t-t_{\lfloor nt\rfloor})\left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right),
\end{aligned}$$

since $G(0) = g(0) = 0$. When $\alpha \in \mathfrak{R}_-^\lambda$, similar steps imply

$$\begin{aligned}(\mathcal{G}^\alpha Y_n)(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t g(t-s)Y_n(s)\mathrm{d}s = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t G(t-s)\frac{\mathrm{d}Y_n(s)}{\mathrm{d}s}\mathrm{d}s \\ &= \sum_{i=1}^{\lfloor nt\rfloor} n\left[G(t-t_{i-1}) - G(t-t_i)\right]\left(Y_n^i - Y_n^{i-1}\right) + nG(t-t_{\lfloor nt\rfloor})\left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right);\end{aligned}$$

when $\frac{\lfloor nt\rfloor}{n} = t$, $G(0) = 0$, and the expression is well defined. □

Notice here, that the mean value theorem implies

$$\left(\mathcal{G}^\alpha Y_n\right)(t) = \sum_{i=1}^{\lfloor nt\rfloor} g\left(t_i^*\right)\left(Y_n^i - Y_n^{i-1}\right) + g\left(t^*_{\lfloor nt\rfloor+1}\right)\left(Y_n(t) - Y_n^{\lfloor nt\rfloor}\right), \tag{2.11}$$

where $t_i^* \in [t-t_i, t-t_{i-1}]$ and $t^*_{\lfloor nt\rfloor+1} \in [0, t-t_{\lfloor nt\rfloor}]$ and we use that $G(0) = 0$. This expression is closer to the usual left-point forward Euler approximation. For numerical purposes, (2.11) is much more efficient, since the integral G required in (2.10) is not necessarily available in closed form. Nevertheless, not any arbitrary choice of t_i^* gives the desired convergence from the above argument. We shall present a suitable candidate for optimal t_i^* in Section 2.4.3, which guarantees weak convergence in Hölder sense.

As could be expected, the Hurst parameter influences the speed of convergence of the scheme. We leave a formal proof to further study, but the following argument provides some intuition about the correct normalising factor: For $\alpha \in \mathfrak{R}_-^\lambda$, since $g \in \mathcal{L}^\alpha$, the approximation (2.11) reads, for any $n \geq 1$,

$$\left(\mathcal{G}^\alpha Y_n\right)(t_i) = \frac{1}{n^{1/2+\alpha}}\sum_{k=1}^{i}\left(nt_i - (k-1)\right)^\alpha L\left(t_i - t_{k-1}\right)\left(Y_n^k - Y_n^{k-1}\right)\sqrt{n}, \quad \text{for } i = 0,\dots,n.$$

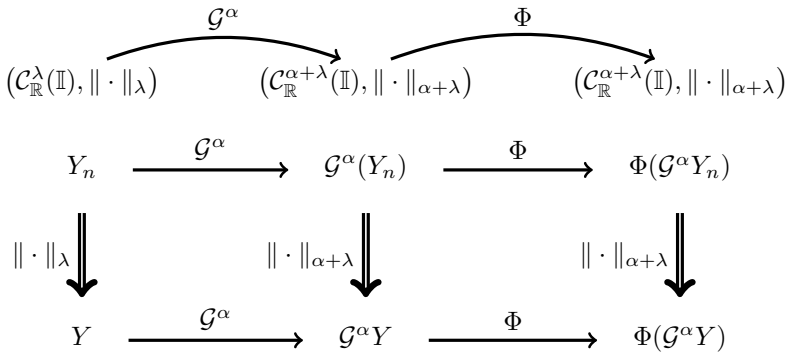
Here, $(nt_i - (k-1))^\alpha \leq t_i^\alpha$ is bounded for any $n \geq 1$, so that the normalisation factor is of order $n^{-\alpha-1/2}$. When $\alpha \in \mathfrak{R}_+^\lambda$ we rewrite (2.11) as

$$\left(\mathcal{G}^\alpha Y_n\right)(t_i) = \frac{1}{\sqrt{n}}\sum_{k=1}^{i}\left(t_i - t_{k-1}\right)^\alpha L\left(t_i - t_{k-1}\right)\left(Y_n^k - Y_n^{k-1}\right)\sqrt{n}, \quad \text{for } i = 0,\dots,n,$$

in which case, $\left(t_i - t_{k-1}\right)^\alpha \leq t_i^\alpha$ is bounded for $n \geq 1$, and the normalisation factor is of order $n^{-1/2}$. This intuition is consistent with the result by Neuenkirch and Shalaiko [NS16], who found the strong rate of convergence of the Euler scheme to be of order $\mathcal{O}(n^{-H})$ for $H < \frac{1}{2}$ for fractional Ornstein-Uhlenbeck. So far, our results hold for α-Hölder continuous functions; however, for practical purposes, it is often necessary to constrain the volatility process $(V_t)_{t\in\mathbb{I}}$ to remain strictly positive at all times. The stochastic integral $\mathcal{G}^\alpha Y$ need not be so in general. However, a simple transformation of the dynamics (e.g. exponential transformation from normal to lognormal dynamics) can easily overcome this fact. The remaining question is whether the α-Hölder continuity is preserved after this composition.

Proposition 2.3.9. *Let $(Y_n)_{n\geq 1}$ be the approximating sequence (2.6) in $\mathcal{C}_{\mathbb{R}}^\lambda(\mathbb{I})$ for $\lambda < 1/2$. Then $\left(\Phi\left(\mathcal{G}^\alpha Y_n\right)\right)$ converges weakly to $\Phi\left(\mathcal{G}^\alpha Y\right)$ in $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\right)$ for all $\alpha \in \mathfrak{R}^\lambda$.*

Proof. Drábek [Drá75] found necessary and sufficient conditions ensuring that Hölder continuity is preserved under composition (so-called Nemyckij operators), and proved that $f \circ g$ is continuous from $\left(\mathcal{C}_{\mathbb{R}}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $\left(\mathcal{C}_{\mathbb{R}}^\lambda(\mathbb{I}, \|\cdot\|_\lambda\right)$ if and only if $f \in \mathcal{C}_{\mathbb{R}}^1$. The proposition then follows by applying continuous mapping to Theorem 2.3.8 along with Drábek's continuity property. The diagram below summarises the steps with $\lambda < 1/2$. The double arrows show weak convergence, and we indicate next to them the topology in which it takes place.



□

2.3.4 Extending the weak convergence to the Skorokhod space and discussion of Conjecture 2.2.11

The Skorokhod space of càdlàg processes equipped with the Skorokhod topology has been widely used to prove weak convergence [Bil68]. The Skorokhod space of real-valued càdlàg processes equipped with the Skorokhod norm, which we denote $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$, markedly simplifies when we only consider continuous processes (as is the case of our framework with Hölder continuous processes). Billingsley [Bil68, Chapter 3 Section 12] proved that the identity $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}) \cap \mathcal{C}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}}) = (\mathcal{C}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_\infty)$ always holds. This seemingly simple statement allows us to reduce proofs of weak convergence of continuous processes in the Skorokhod topology to that in the supremum norm, usually much simpler. We start with the following straightforward observation:

Lemma 2.3.10. *For any $\lambda \in (0,1)$, the identity map is continuous from $\left(\mathcal{C}^\lambda_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_\lambda\right)$ to $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$.*

Proof. Since the identity map is linear, it suffices to check that it is bounded. For this observe that $\|f\|_\lambda = |f|_\lambda + \sup_{t\in\mathbb{I}} |f(t)| = |f|_\lambda + \|f\|_\infty > \|f\|_\infty$, where $|f|_\lambda > 0$, which concludes the proof since the Skorokhod norm in the space of continuous functions is equivalent to the supremum norm. □

Applying the Continuous Mapping Theorem twice, first with the Generalised fractional operator (Theorem 2.3.8), then with the identity map, yields the following result directly:

Theorem 2.3.11. *The sequence $(\Phi(\mathcal{G}^\alpha Y_n))$ converges weakly to $\Phi(\mathcal{G}^\alpha Y)$ in $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), ||\cdot||_{\mathcal{D}})$ for any $\alpha \in \mathfrak{R}^{1/2}$.*

The final step in the proof of our main conjecture would be to extend weak convergence to the log-stock price. For this, the following result on weak convergence of stochastic integrals $X \bullet Y := \int X^- \mathrm{d}Y$ due to Jakubowski, Memin and Pagès [JMP89], and later generalised to SDEs by Kurtz and Protter [KP91] is potentially a building block.

Theorem 2.3.12 (Kurtz and Protter [KP91])**.** *Let $(B_n)_{n\geq 1}$ be as in (2.5), N a càdlàg process on $\mathbb{I}$, and $(N_n)_{n\geq 1}$ an approximating sequence such that (N_n, B_n) converges weakly in $(\mathcal{D}_{\mathbb{R}^2}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$ to (N, B). Then, there exists a filtration $\mathcal{H}$ under which B is an $\mathcal{H}$-continuous martingale and $(N_n, B_n, N_n \bullet B_n)_{n\geq 1}$ converges weakly to $(N, B, N \bullet B)$.*

Discussion of Conjecture 2.2.11. We first note that, for $\alpha \in \mathfrak{R}^\lambda$, the first term on the right-hand side of (2.7) converges weakly to $-\frac{1}{2}\int_0^T \Phi\left(\mathcal{G}^\alpha Y\right)(s)\mathrm{d}s$ by the Continuous Mapping Theorem, as the integral is a continuous operator from $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), ||\cdot||_{\mathcal{D}})$ to itself. In order to prove the convergence of the second term in (2.7), we would like to apply [KP91, Theorem 2.2]. However, even if all conditions are satisfied, we face the issue that $\int \sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)(s)}\mathrm{d}B_n$ causes an Itô-Stratonovich issue if we apply the definition of the stochastic integral to the (linearly interpolating) approximating sequences directly. We believe that this could be overcome by introducing $\left(X_n \bullet_n Y_n\right)(t_j) := \sum_{i=0}^{\lfloor nt_j \rfloor} X_n(t_i)\left((Y_n(t_{i+1}) - Y_n(t_i)\right)$, which avoids the issue by using the left-point construction of the Itô integral. However, adapting [KP91, Theorem 2.2] to the operator $\bullet_n$ emains a topic for future work and is not formally proven here. □

2.4 Applications

2.4.1 Weak convergence of the Hybrid scheme

The Hybrid scheme (and its turbocharged version [MP18a]) introduced by Bennedsen, Lunde and Pakkanen [BLP17] is the current state-of-the-art to simulate $t\mathcal{BSS}$ processes. However, only convergence in mean-square-error was proved, but not weak convergence, which would justify the use of the scheme for path-dependent options. Unless otherwise stated, we shall denote by $\mathcal{T}_n := \{t_k = \frac{k}{n}\}_{k=0,\dots,n}$ the uniform grid on $\mathbb{I}$. We show that the Hölder convergence also holds for the case $\kappa = 1$:

Proposition 2.4.1. *The Hybrid scheme sequence $(\widetilde{\mathcal{G}}^\alpha W_n)$ for $t \in \mathbb{I}$, defined as*

$$\widetilde{\mathcal{G}}^\alpha W_n(t) := \left(\mathcal{G}^\alpha W_n\left(\frac{\lfloor nt\rfloor}{n}\right) + \mathcal{H}^\alpha W_n\left(\frac{\lfloor nt\rfloor}{n}\right)\right)(1-(nt-\lfloor nt\rfloor))+(nt-\lfloor nt\rfloor)\left(\widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt\rfloor+1}{n}\right) + \mathcal{H}^\alpha W_n\left(\frac{\lfloor nt\rfloor+1}{n}\right)\right), \tag{2.12}$$

where

$$\mathcal{H}^\alpha W_n(t_i) := -\int_{0\vee t_{i-1}}^{t_i} ng(t_i-s)\mathrm{d}s\xi_i + \int_{0\vee t_{i-1}}^{t_i} g(t_i-s)\mathrm{d}W_s, \quad i=1,\dots,n, \quad \kappa \geq 1. \tag{2.13}$$

with $\mathcal{H}^\alpha W_n(t_0) := 0$ and $\xi_i := \int_{t_{i-1}}^{t_i} \mathrm{d}W_s \sim \mathcal{N}(0,1/n)$, converges to $\mathcal{G}^\alpha W$ in $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ for $\alpha \in \mathfrak{R}^{1/2}$ and $\kappa = 1$.

Proof. We know that $\mathcal{G}^\alpha W_n$ converges to $\mathcal{G}^\alpha W$ in $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ for $\alpha \in \mathfrak{R}^{1/2}$. Thus all we need, is to prove that $\mathcal{H}_n^\alpha W$ converges to 0 in $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ for $\alpha \in \mathfrak{R}^{1/2}$. Finite dimensional convergence to 0 follows trivially. Consequently, in order to prove convergence we need to show that the approximating sequence is tight or equivalently by Theorem 2.3.4

$$\mathbb{E}\left[|\mathcal{H}^\alpha W_n(t_i) - \mathcal{H}^\alpha W_n(t_j)|^{2p}\right] \leq C(t_i-t_j)^{2p\alpha+p}$$

for $p \geq 1$ and some constant $C \geq 0$. We note that for all $i \geq 1$,

$$\begin{aligned}\mathbb{E}\left[\mathcal{H}^\alpha W_n(t_i)^2\right] &\leq \frac{1}{n}\left(L(0)\int_{0\vee t_{i-1}}^{t_i} n(t_i-s)^\alpha \mathrm{d}s\right)^2 + L(0)^2\int_{0\vee t_{i-1}}^{t_i}(t_i-s)^{2\alpha}\mathrm{d}s + 2L(0)^2\int_{0\vee t_{i-1}}^{t_i} n(t_i-s)^\alpha \mathrm{d}s\int_{0\vee t_{i-1}}^{t_i}(t_i-s)^\alpha \mathrm{d}s \\ &\leq L(0)^2\left(\left(\frac{1}{n}\right)^{2\alpha+1} + \left(\frac{1}{n}\right)^{2\alpha+1}\frac{1}{2\alpha+1} + 2\left(\frac{1}{n}\right)^{2\alpha+1}\frac{1}{(\alpha+1)^2}\right) \leq C\left(\frac{1}{n}\right)^{2\alpha+1},\end{aligned}$$

where we use the fact that $L(\cdot)$ is non increasing. Without loss of generality assume $t_j < t_i$ and take $\kappa = 1$. We have

$$\mathbb{E}\left[|\mathcal{H}^\alpha W_n(t_i) - \mathcal{H}^\alpha W_n(t_j)|^2\right] \leq \mathbb{E}\left[\mathcal{H}^\alpha W_n(t_i)^2\right] + \mathbb{E}\left[\mathcal{H}^\alpha W_n(t_j)^2\right] \leq C\left(\frac{1}{n}\right)^{2\alpha+1} \leq C(t_i-t_j)^{2\alpha+1}$$

Thus, by standard moment properties of Gaussian random variables [BLM13, Theorem 2.1] we obtain

$$\mathbb{E}\left[|\mathcal{H}^\alpha W_n(t_i) - \mathcal{H}^\alpha W_n(t_j)|^{2p}\right] \leq \tilde{C}(t_i-t_j)^{2p\alpha+p},$$

which gives the desired result.

□

We further note that Proposition 2.13 along with Conjecture 2.2.11 ensures the weak convergence of the log-stock price for the Hybrid scheme as well.

2.4.2 Application to fractional binomial trees

We consider a binomial setting for the Riemann-Liouville fractional Brownian motion $\mathcal{G}^{H-1/2}W$ with $g(u) \equiv u^{H-1/2}$, $H \in (0,1)$, for which Theorem 2.3.8 provides a weakly converging sequence. On the partition $\mathcal{T}_n$, with a Bernoulli sequence $\{\zeta_i\}_{i=1}^n$ satisfying $\mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i = -1) = \frac{1}{2}$ for all i (justified by Conjecture 2.2.11), the approximating sequence reads

$$(\mathcal{G}^{H-1/2}W_n)(t_i) := \frac{1}{\sqrt{n}} \sum_{k=1}^{i} (t_i - t_{k-1})^{H-1/2} \zeta_k, \quad \text{for } i = 0, \dots, n.$$

Figure 2.1 shows a fractional binomial tree structure for $H = 0.75$ and $H = 0.1$. Despite being symmetric, such trees cannot be recombining due to the (non-Markovian) path-dependent nature of the process. It might be possible, in principle, to modify the tree at each step to make it recombining, following the procedure developed in [ADS14] for Markovian stochastic volatility models. It is not so straightforward though, and requires a dedicated thorough analysis which we leave for future research.

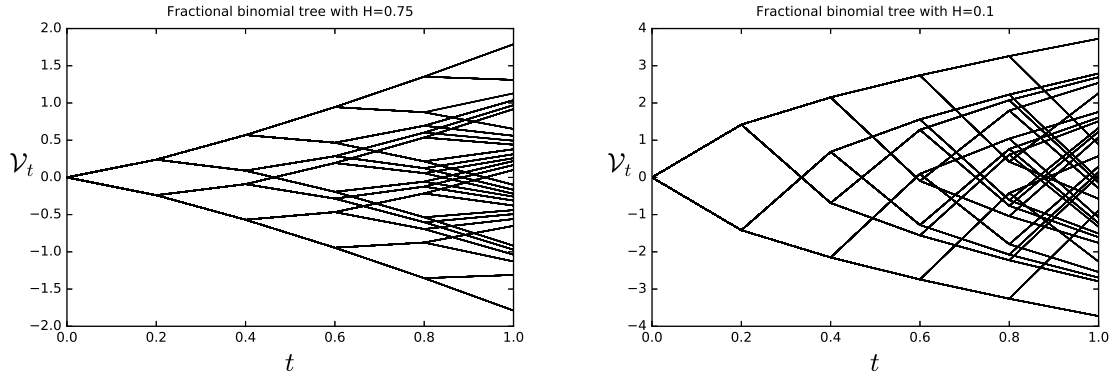


Figure 2.1: Binomial tree for the Riemann-Liouville fractional Brownian motion with $n = 5$ discretisation points for $H = 0.75$ (left) and $H = 0.1$ (right).

2.4.3 Monte-Carlo

Conjecture 2.2.11 touches upon the theoretical foundations of Monte-Carlo methods (in particular for path-dependent options) for rough volatility models. In this section we give a general and easy-to-understand recipe to implement the class of rough volatility models (2.3). For the numerical recipe to be as general as possible, we shall consider the general time partition $\mathcal{T} := \{t_i = \frac{iT}{n}\}_{i=0,\dots,n}$ on $[0,T]$ with $T > 0$.

Algorithm 2.4.2 (Simulation of rough volatility models)**.**

1. Simulate two $\mathcal{N}(0,1)$ matrices $\{\xi_{j,i}\}_{\substack{j=1,\dots,M \\ i=1,\dots,n}}$ and $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M \\ i=1,\dots,n}}$ with $\operatorname{corr}(\xi_{j,i}, \zeta_{j,i}) = \rho$;

2. simulate M paths of Y_n via[1]

$$Y_n^j(t_i) = \frac{T}{n} \sum_{k=1}^{i} b(Y_n^j(t_{k-1})) + \frac{T}{\sqrt{n}} \sum_{k=1}^{i} a\left(Y_n^j(t_{k-1})\right) \zeta_{j,k}, \quad i = 1, \dots, n \text{ and } j = 1, \dots, M,$$

and also compute

$$\Delta Y_n^j(t_i) := Y_n^j(t_i) - Y_n^j(t_{i-1}), \quad i = 1, \dots, n \text{ and } j = 1, \dots, M,$$

[1]Here, $Y_n^j(t_i)$ denotes the j-th path Y_n evaluated at the time point t_i, which is different from the notation Y_n^j in the theoretical framework above, but should not create any confusion.

3. Simulate M paths of the fractional driving process $((\mathcal{G}^\alpha Y_n)(t))_{t\in\mathcal{T}}$ using

$$(\mathcal{G}^\alpha Y_n)^j(t_i) := \sum_{k=1}^{i} g(t_{i-k+1})\Delta Y_n^j(t_k) = \sum_{k=1}^{i} g(t_k)\Delta Y_n^j(t_{i-k+1}), \quad i = 1,\dots,n \text{ and } j = 1,\dots,M.$$

The complexity of this step is in general of order $\mathcal{O}(n^2)$ (see Appendix A.1 for details). However, this step is easily implemented using discrete convolution with complexity $\mathcal{O}(n\log n)$ (see Algorithm A.1.4 in Appendix A.1 for details in the implementation). With the vectors $\mathfrak{g} := (g(t_i))_{i=1,\dots,n}$ and $\Delta Y_n^j := (\Delta Y_n^j(t_i))_{i=1,\dots,n}$ for $j = 1,\dots,M$, we can write $(\mathcal{G}^\alpha Y_n)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \Delta Y_n^j)$, for $j = 1,\dots,M$, where $*$ represents the discrete convolution operator.

4. Use the forward Euler scheme to simulate the log-stock process, for all $i = 1,\dots,n$, $j = 1,\dots,M$, as

$$X^j(t_i) = X^j(t_{i-1}) - \frac{1}{2}\frac{T}{n}\Phi\left(\mathcal{G}^\alpha Y_n\right)^j(t_{i-1}) + \sqrt{\frac{T}{n}}\sqrt{\Phi\left(\mathcal{G}^\alpha Y_n\right)^j(t_{i-1})}\xi_{j,i}.$$

Remark 2.4.3.

- When $Y = W$, we may skip step (2) and replace $\Delta Y_n^j(t_i)$ by $\sqrt{T/n}\zeta_{i,j}$ on step (3).
- Step (3) may be replaced by the Hybrid scheme algorithm [BLP17] only when $Y = W$.

Antithetic variates in Algorithm 2.4.2 are easy to implement as it suffices to consider the uncorrelated random vectors $\zeta_j := (\zeta_{j,1}, \zeta_{j,2}, \dots, \zeta_{j,n})$ and $\xi_j := (\xi_{j,1}, \xi_{j,2}, \dots, \xi_{j,n})$, for $j = 1,\dots,M$. Then $(\rho\xi_j + \overline{\rho}\zeta_j, \xi_j)$, $(\rho\xi_j - \overline{\rho}\zeta_j, \xi_j)$, $(-\rho\xi_j - \overline{\rho}\zeta_j, -\xi_j)$ and $(-\rho\xi_j + \overline{\rho}\zeta_j, -\xi_j)$, for $j = 1,\dots,M$, constitute the antithetic variates, which significantly improves the performance of the Algorithm 2.4.2 by reducing memory requirements, reducing variance and accelerating execution by exploiting symmetry of the antithetic random variables.

Enhancing performance

A standard practice in Monte-Carlo simulation is to match moments of the approximating sequence with the target process. In particular, when the process is Gaussian, matching first and second moments suffices. We only illustrate this approximation for Brownian motion: the left-point approximation (2.11) (with $Y = W$) may be modified to match moments as

$$(\mathcal{G}^\alpha W)(t_i) \approx \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{i} g\left(t_k^*\right)\zeta_k, \qquad \text{for } i = 0,\dots,n, \tag{2.14}$$

where t_k^* is chosen optimally. Since the kernel $g(\cdot)$ is deterministic, there is no confusion with the Stratonovich stochastic integral, and the resulting approximation will always converge to the Itô integral. The first two moments of $\mathcal{G}^\alpha W$ read

$$\mathbb{E}\left(\left(\mathcal{G}^\alpha W\right)(t)\right) = 0 \qquad \text{and} \qquad \mathbb{V}\left(\left(\mathcal{G}^\alpha W\right)(t)\right) = \int_0^t g(t-s)^2\mathrm{d}s.$$

The first moment of the approximating sequence (2.14) is always zero, and the second moment reads

$$\mathbb{V}\left(\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{j-1} g\left(t_k^*\right)\zeta_k\right) = \frac{1}{n}\sum_{k=1}^{j-1} g\left(t_k^*\right)^2.$$

Equating the theoretical and approximating quantities we obtain $\frac{1}{n}g(t_k^*)^2\mathrm{d}s = \int_{t_{k-1}}^{t_k} g(t-s)^2\mathrm{d}s$ for $k = 1,\dots,n$, so that the optimal evaluation point can be computed as

$$g(t_k^*) = \sqrt{n\int_{t_{k-1}}^{t_k} g(t-s)^2\mathrm{d}s}, \qquad \text{for any } k = 1,\dots,n. \tag{2.15}$$

With the optimal evaluation point the scheme is still a convolution so that Algorithm A.1.4 in Appendix A.1 can still be used for faster computations. In the Riemann-Liouville fractional Brownian motion case, $g(u) = u^{H-1/2}$,

and the optimal point can be computed in closed form as

$$t_k^* = \left(\frac{n}{2H}\left[(t-t_{k-1})^{2H} - (t-t_k)^{2H}\right]\right)^{1/(2H-1)}, \qquad \text{for each } k = 1, \dots, n.$$

Proposition 2.4.4. *The moment matching sequence* $(\hat{\mathcal{G}}^\alpha W_n)$ *defined as*

$$\hat{\mathcal{G}}^\alpha W_n(t) := \hat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(\hat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - \hat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right), t \in \mathbb{I} \tag{2.16}$$

where

$$\hat{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{i} \sqrt{n \int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s}\xi_k. \tag{2.17}$$

with ξ_k *a centered sub-Gaussian random variable with* $\mathbb{E}[(\xi_k)^2] = \frac{1}{n}$*, meaning that there exist constants* $C > 0$ *and* $v > 0$ *such that for all* $x > 0$*,* $\mathbb{P}(|\xi_k| > x) \leq Ce^{-vx^2}$*. Then, convergence to* $\mathcal{G}^\alpha W$ *holds in* $\left(\mathcal{C}_{\mathbb{R}}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2}\right)$ *for* $\alpha \in \mathfrak{R}^{1/2}$*.*

Proof. Finite dimensional convergence follows trivially from the central limit theorem as the target process is a zero mean Gaussian process, thus convergence of the covariance matrix ensures finite dimensional convergence. Consequently, in order to prove convergence we need to show that the approximating sequence is tight or equivalently by Theorem 2.3.4

$$\mathbb{E}\left[\left|\hat{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)\right|^{2p}\right] \leq C(t_i - t_s)^{2p\alpha+p}$$

for $p \geq 1$ and some constant $C \geq 0$. We have

$$\tilde{\sigma}^2 := \mathbb{E}\left[\left(\hat{\mathcal{G}}^\alpha W_n(t_i) - \hat{\mathcal{G}}^\alpha W_n(t_j)\right)^2\right] = \mathbb{E}\left[\left(\hat{\mathcal{G}}^\alpha W_n(t_i)\right)^2\right] + \mathbb{E}\left[\left(\hat{\mathcal{G}}^\alpha W_n(t_j)\right)^2\right] - 2\mathbb{E}\left[\hat{\mathcal{G}}^\alpha W_n(t_i)\hat{\mathcal{G}}^\alpha W_n(t_j)\right].$$

We note

$$\mathbb{E}\left[\left(\hat{\mathcal{G}}^\alpha W_n(t_i)\right)^2\right] = \sum_{k=1}^{i}\left(\sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s}\right)^2 = \int_0^{t_i} g(t_i - s)^2 \mathrm{d}s = \mathbb{E}\left[\left(\mathcal{G}^\alpha W(t_i)\right)^2\right]$$

and by Cauchy-Schwarz we also have

$$\mathbb{E}\left[\hat{\mathcal{G}}^\alpha W_n(t_i)\hat{\mathcal{G}}^\alpha W_n(t_j)\right] = \sum_{k=1}^{t_i \wedge t_j}\sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s \int_{t_{k-1}}^{t_k} g(t_j - s)^2 \mathrm{d}s} \geq \int_0^{t_i \wedge t_j} g(t_i - s)g(t_j - s)\mathrm{d}s = \mathbb{E}\left[\mathcal{G}^\alpha W(t_i)\mathcal{G}^\alpha W(t_j)\right].$$

We then obtain,

$$\tilde{\sigma}^2 \leq \mathbb{E}\left[\left|\mathcal{G}^\alpha W(t_i) - \mathcal{G}^\alpha W(t_j)\right|^2\right].$$

Finally, using the fact that $\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W(t_j)$ is a linear combination of sub-Gaussian random variables, which is again sub-Gaussian. We may apply a Gaussian moment inequality [BLM13, Theorem 2.1] based on the variance estimate $\tilde{\sigma}^2$ we have, to obtain

$$\mathbb{E}\left[\left|\hat{\mathcal{G}}^\alpha W_n(t_i) - \hat{\mathcal{G}}^\alpha W(t_j)\right|^{2p}\right] \leq \mathbb{E}\left[\left|\mathcal{G}^\alpha W(t_i) - \mathcal{G}^\alpha W(t_j)\right|^{2p}\right] \leq \tilde{C}(t_i - t_s)^{2p\alpha+p}.$$

□

Reducing variance

As Bayer, Friz and Gatheral [BFG16] and Bennedsen, Lunde and Pakkanen [BLP17] pointed out, a major drawback in simulating rough volatility models is the very high variance of the estimators, so that a large number of simulations are needed to produce a decent price estimate. Nevertheless, the rDonsker scheme admits a very simple conditional expectation technique which reduces both memory requirements and variance while also admitting antithetic variates. This approach is best suited for calibrating European type options. We consider

$\mathcal{F}_t^B = \sigma(B_s : s \le t)$ and $\mathcal{F}_t^W = \sigma(W_s : s \le t)$ the natural filtrations generated by the Brownian motions B and W. In particular the conditional variance process $V_t|\mathcal{F}_t^W$ is deterministic. As discussed by Romano and Touzi [RT97], and recently adapted to the rBergomi case by McCrickerd and Pakkanen [MP18a], we can decompose the stock price process as

$$\mathrm{e}^{X_t} = \mathcal{E}\left(\rho \int_0^t \sqrt{\Phi\left(\mathcal{G}^\alpha Y\right)(s)}\mathrm{d}W_s\right)\mathcal{E}\left(\sqrt{1-\rho^2}\int_0^t \sqrt{\Phi\left(\mathcal{G}^\alpha Y\right)(s)}\mathrm{d}W_s^\perp\right) := \mathrm{e}^{X_t^{||}}\mathrm{e}^{X_t^\perp},$$

and notice that

$$X_t|(\mathcal{F}_t^W \wedge \mathcal{F}_0^B) \sim \mathcal{N}\left(X_t^{||} - (1-\rho^2)\int_0^t \Phi\left(\mathcal{G}^\alpha Y\right)(s)\mathrm{d}s, (1-\rho^2)\int_0^t \Phi\left(\mathcal{G}^\alpha Y\right)(s)\mathrm{d}s\right).$$

Thus $\exp(X_t)$ becomes log-normal and the Black-Scholes closed-form formulae are valid here (European, Barrier options, maximum,...). The advantage of this approach is that the orthogonal Brownian motion $W^\perp$ is completely unnecessary for the simulation, hence the generation of random numbers is reduced to a half, yielding proportional memory saving. Not only this, but also this simple trick reduces the variance of the Monte-Carlo estimate, hence fewer simulations are needed to obtain the same precision. We present a simple algorithm to implement the rDonsker with conditional expectation and assuming that $Y = W$.

Algorithm 2.4.5 (Simulation of rough volatility models with Brownian drivers)**.** Consider the equidistant grid $\mathcal{T}$.

1. Draw a random matrix $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M/2\\ i=1,\dots,n}}$ with unit variance, and create antithetic variates $\{-\zeta_{j,i}\}_{\substack{j=1,\dots,M/2\\ i=1,\dots,n}}$;
2. Create a correlated matrix $\{\xi_{j,i}\}$ as above;
3. Simulate M paths of the fractional driving process $\mathcal{G}^\alpha W$ using discrete convolution (see Algorithm A.1.4 in Appendix A.1 for details in the implementation):

$$(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \zeta_j), \quad j = 1, \dots, M,$$

and store in memory $(1-\rho^2)\int_0^T (\mathcal{G}^\alpha W)^j(s)\mathrm{d}s \approx (1-\rho^2)\frac{T}{n}\sum_{k=0}^{n-1}(\mathcal{G}^\alpha W)^j(t_k) =: \Sigma^j$ for each $j = 1, \dots, M$;

4. use the forward Euler scheme to simulate the log-stock process, for each $i = 1, \dots, n$, $j = 1, \dots, M$, as

$$X^j(t_i) = X^j(t_{i-1}) - \frac{\rho^2}{2}\frac{T}{n}\Phi\left(\mathcal{G}^\alpha W\right)^j(t_{i-1}) + \rho\sqrt{\frac{T}{n}}\sqrt{\Phi\left(\mathcal{G}^\alpha W\right)^j(t_{i-1})}\zeta_{j,i};$$

5. Finally, we have $X^j(T) \sim \mathcal{N}(X_T^j - \Sigma^j, \Sigma^j)$ for $j = 1, \dots, M$; we may compute any option using the Black-Scholes formula. For instance a Call option with strike K would be given by $C^j(K) = \exp(X_T^j)\mathcal{N}(d_1^j) - K\mathcal{N}(d_2^j)$ for $j = 1, \dots, M$, where $d_1^j := \frac{1}{\sqrt{\Sigma^j}}(X_T^j - \log(K) + \frac{1}{2}\Sigma^j)$ and $d_2^j = d_1^j - \sqrt{\Sigma^j}$. Thus, the output of the model would be $C(K) = \frac{1}{M}\sum_{k=1}^M C^j(K)$.

The algorithm is easily adapted to the case of general diffusions Y as drivers of the volatility (see Algorithm 2.4.2 step (2)). Algorithm 2.4.2 is obviously faster than 2.4.5, especially when using control variates. Nevertheless, with the same number of paths, Algorithm 2.4.5 remarkably reduces the Monte-Carlo variance, meaning in turn that fewer simulations are needed, making it very competitive for calibration.

2.4.4 Numerical example: Rough Bergomi model

Figures 2.2-2.5 perform a numerical analysis of the Monte Carlo convergence as a function of n. It is observed that the lower H is, the larger n needs to be to achieve convergence. However, we also observe that for Cholesky, rDonsker (naive and moment match) and Hybrid schemes and $H \ge 0.1$, with $N = 252$ we already achieve a precision of $\mathcal{O}(10^{-4})$, which is equivalent to a basis point in financial terms. For $H < 0.1$ n larger than 252 might

be required, if precision is required beyond $\mathcal{O}(10^{-4})$. We also observe in 2.5 that the naive rDonsker approximation converges extremely slow for small H. Additionally, Figures 2.6-2.11 measure the price estimations compared to the Cholesky method which is taken as benchmark. We observe that the Hybrid scheme tends to be closer to the Cholesky benchmark specially for $H < 0.1$. In the range where $H \geq 0.1$ for both the Hybrid scheme and rDonsker moment-match we observe an error less than $\mathcal{O}(10^{-4})$ for $n \geq 252$. It is noteworthy mentioning that the naive rDonsker scheme has substantially worse convergence (at least an order of magnitude) than the other methods. We note that the black lines in all figures represent the 99% Monte Carlo standard deviations, hence errors below that threshold should be interpreted as noise.

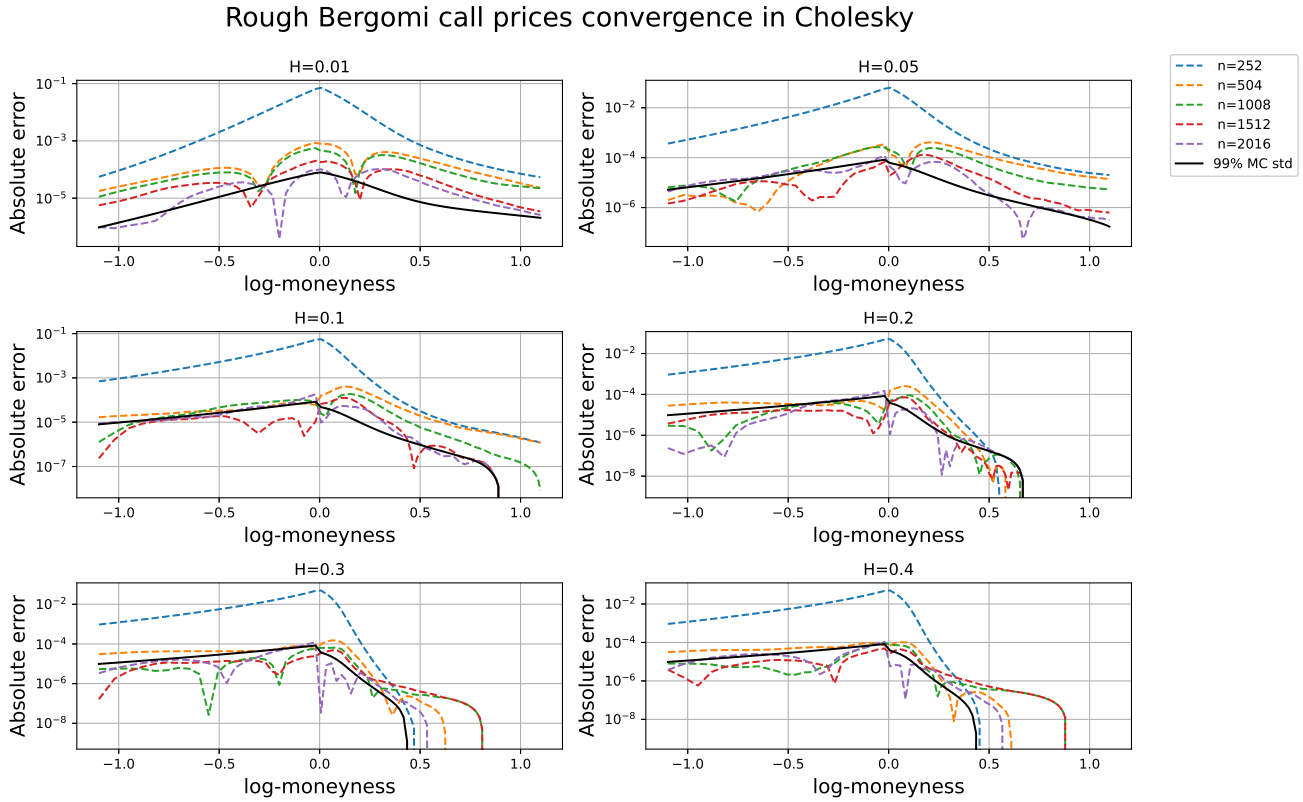


Figure 2.2: Rough Bergomi Call options price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

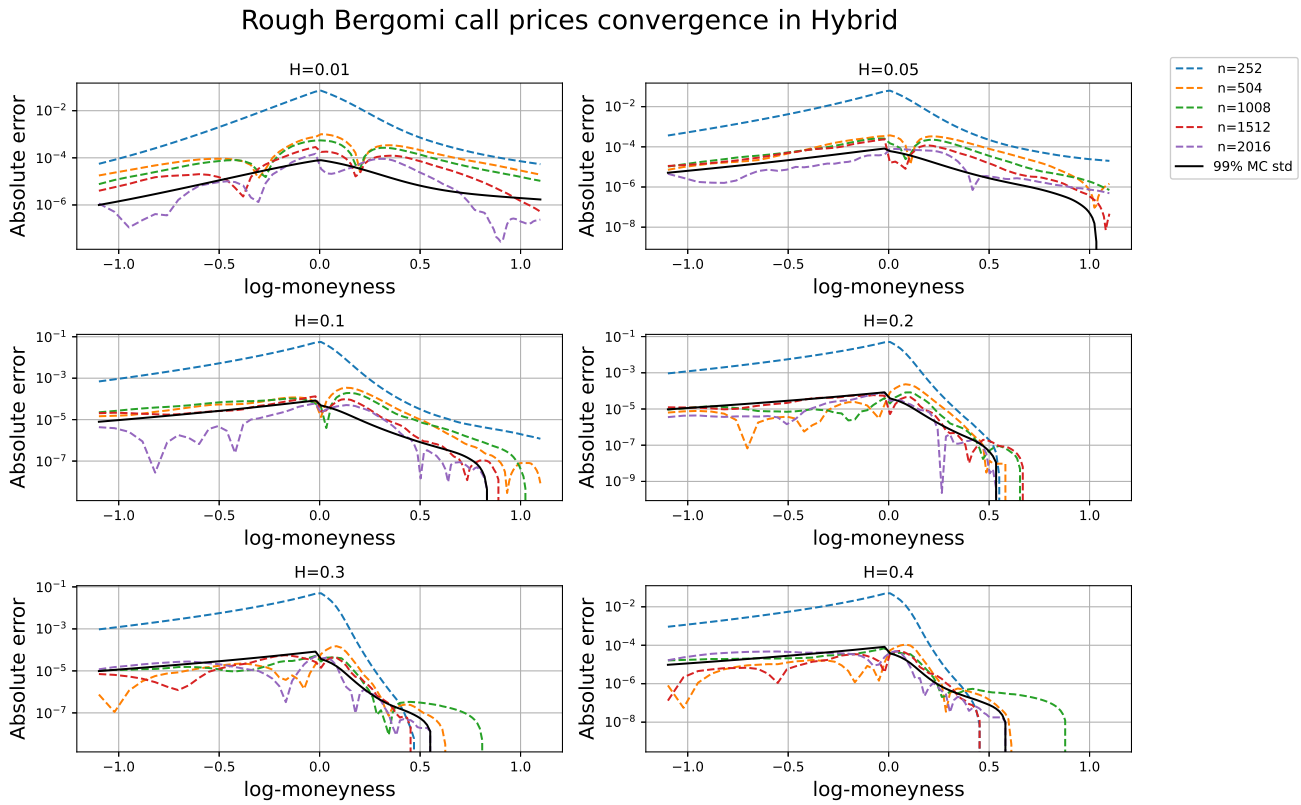


Figure 2.3: Rough Bergomi Call options price convergence using Cholesky method with $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

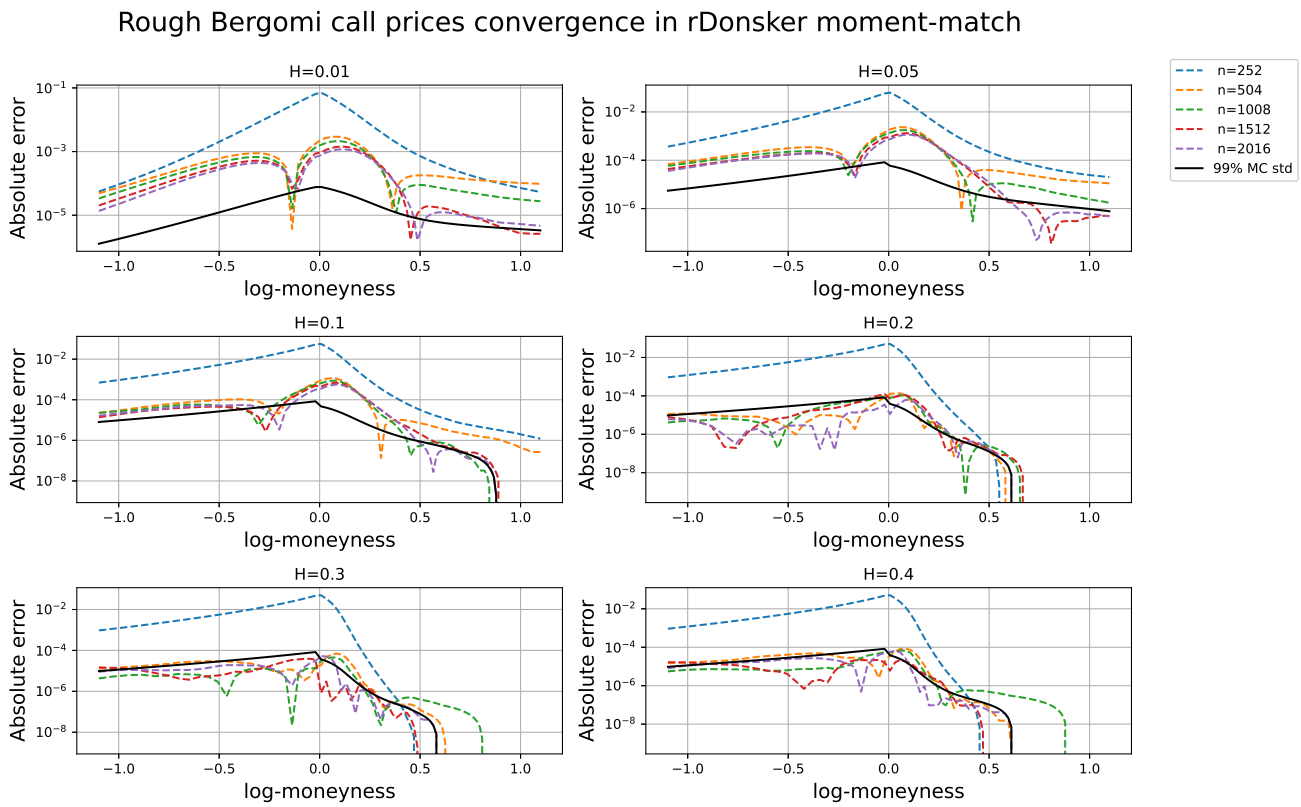


Figure 2.4: Rough Bergomi Call options price convergence using rDonsker method with moment-matching and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

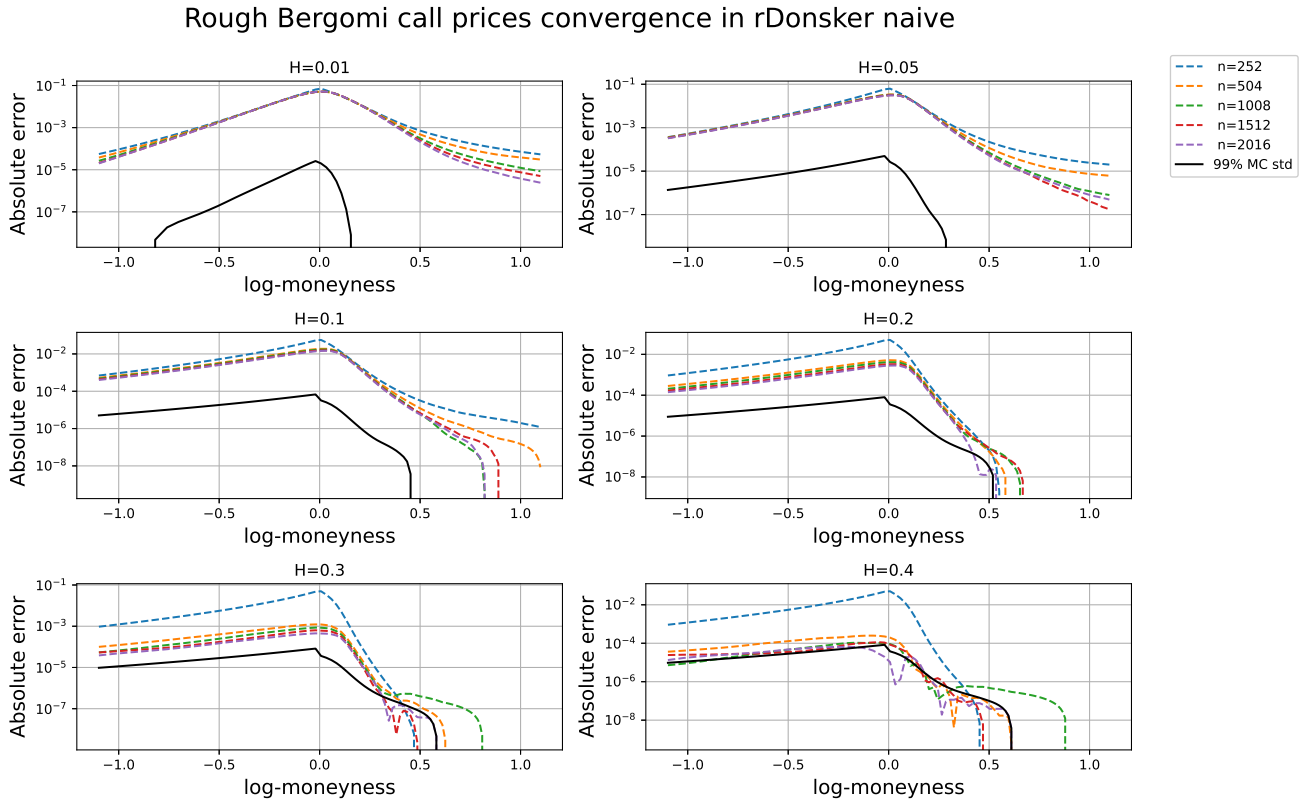


Figure 2.5: Rough Bergomi Call options price convergence using rDonsker method with left point Euler and $\xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$ the previous discretisation is $n = 126$.

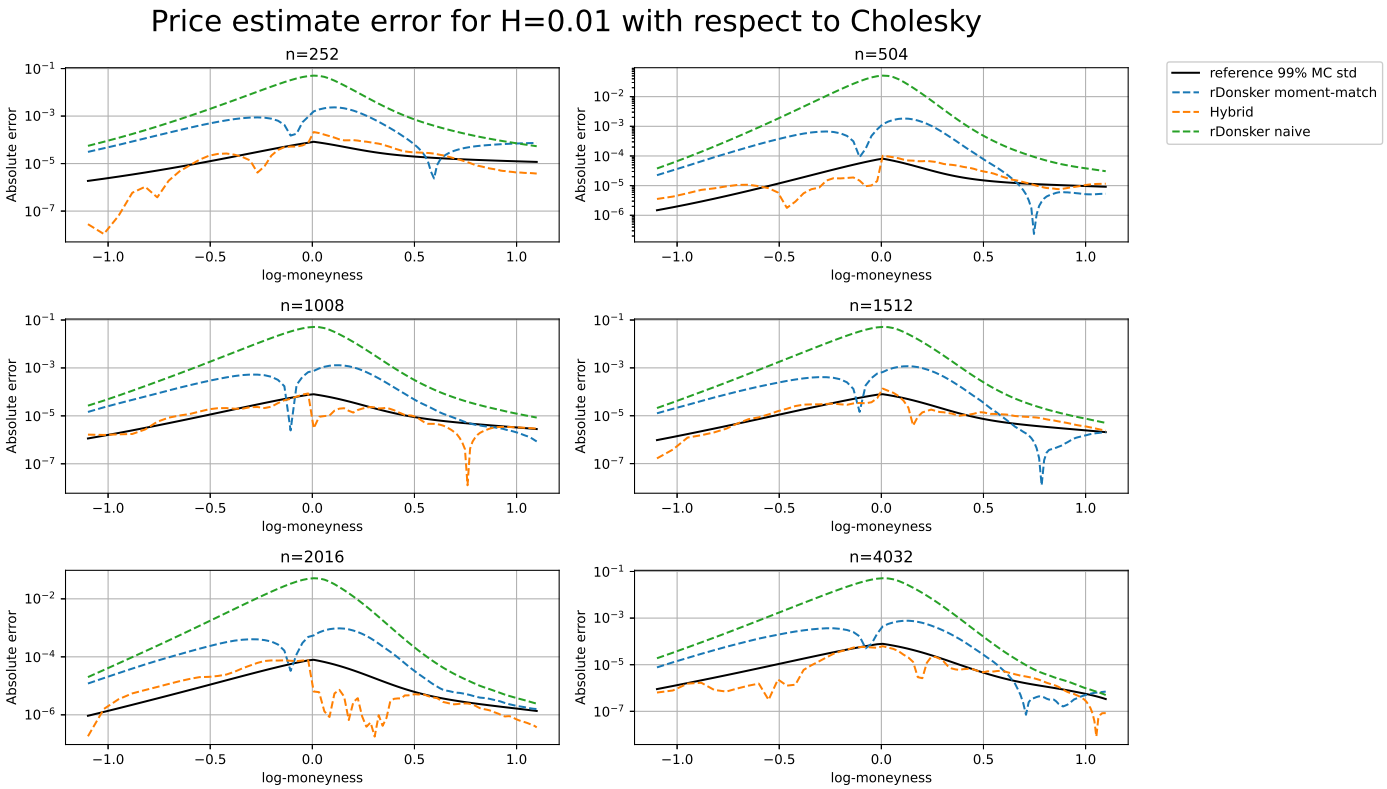


Figure 2.6: Rough Bergomi Call options price comparison with $H = 0.01, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

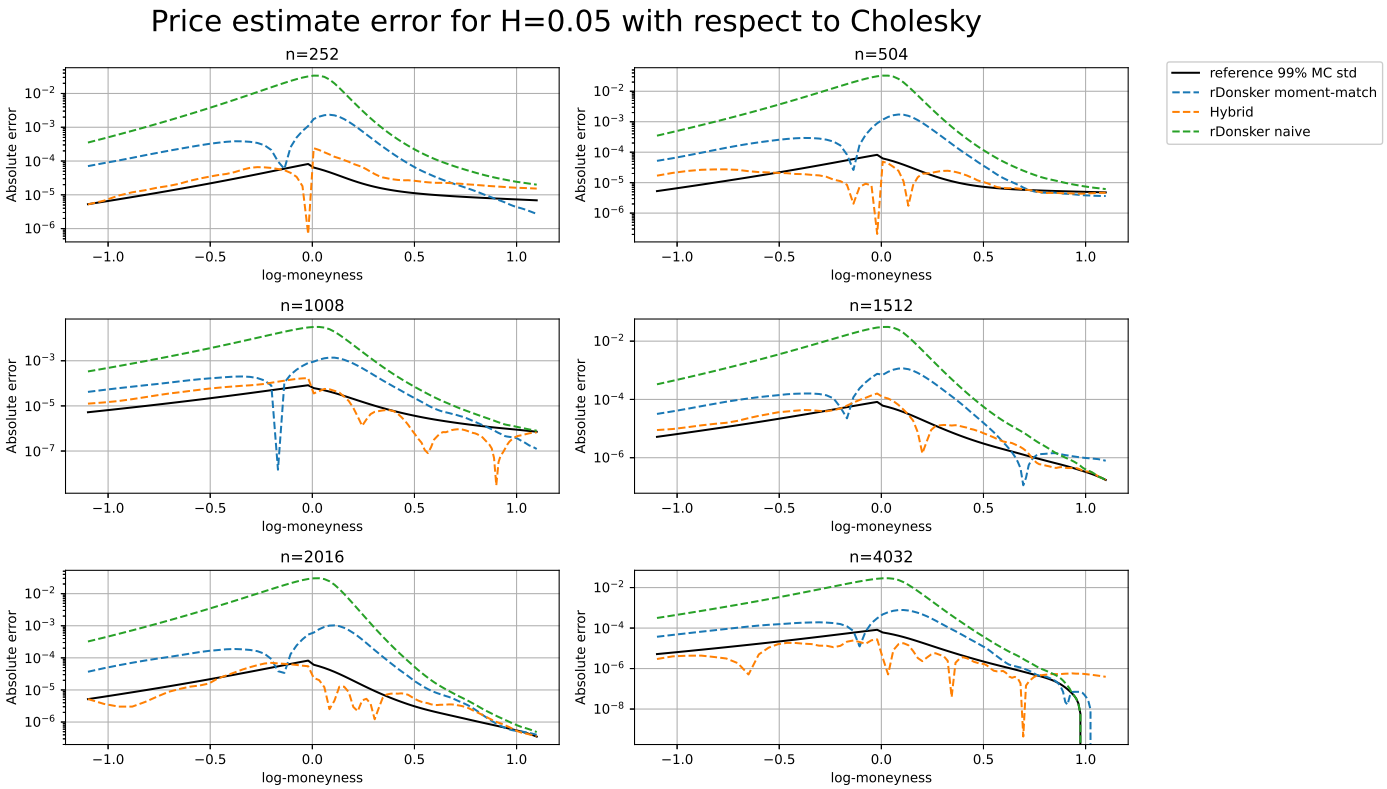


Figure 2.7: Rough Bergomi Call options price comparison with $H = 0.05, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

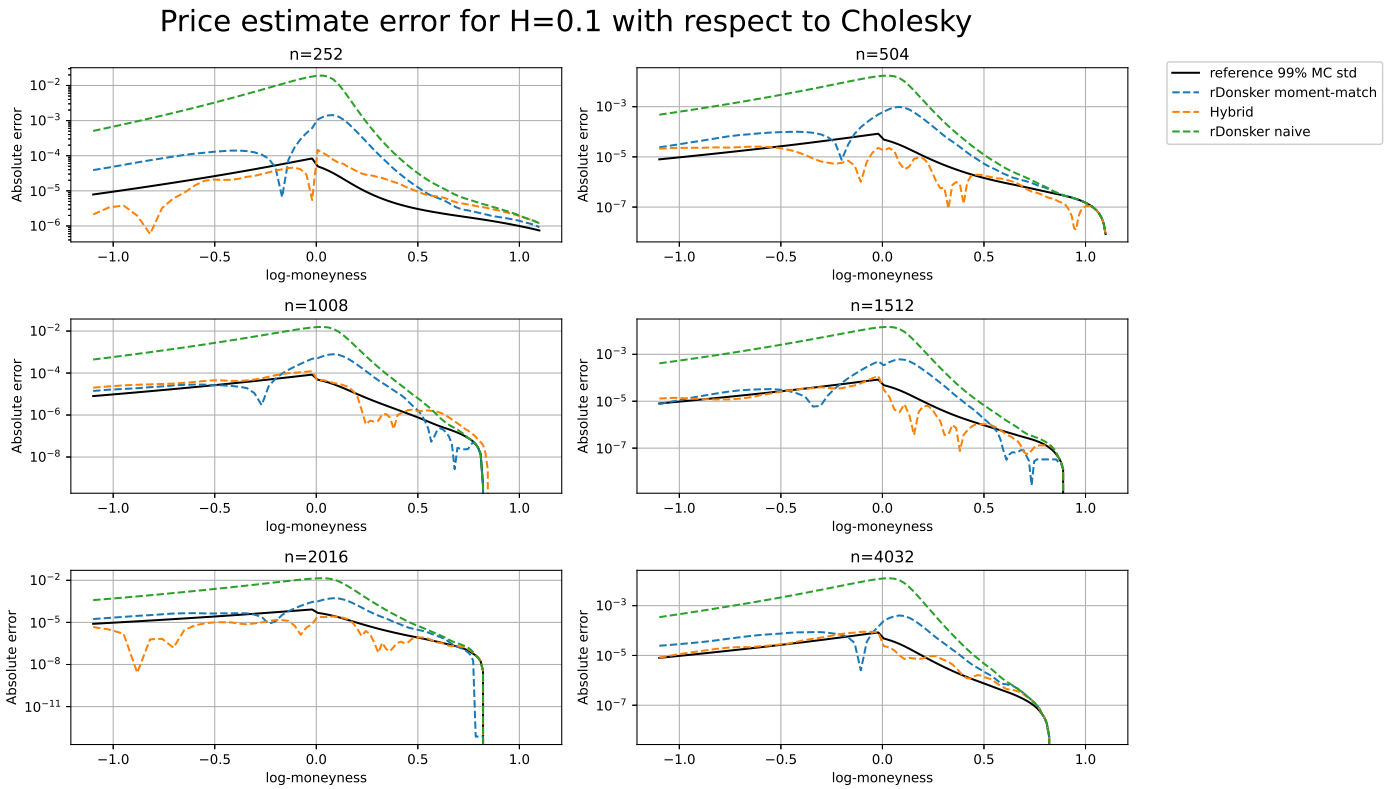


Figure 2.8: Rough Bergomi Call options price comparison with $H = 0.10, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

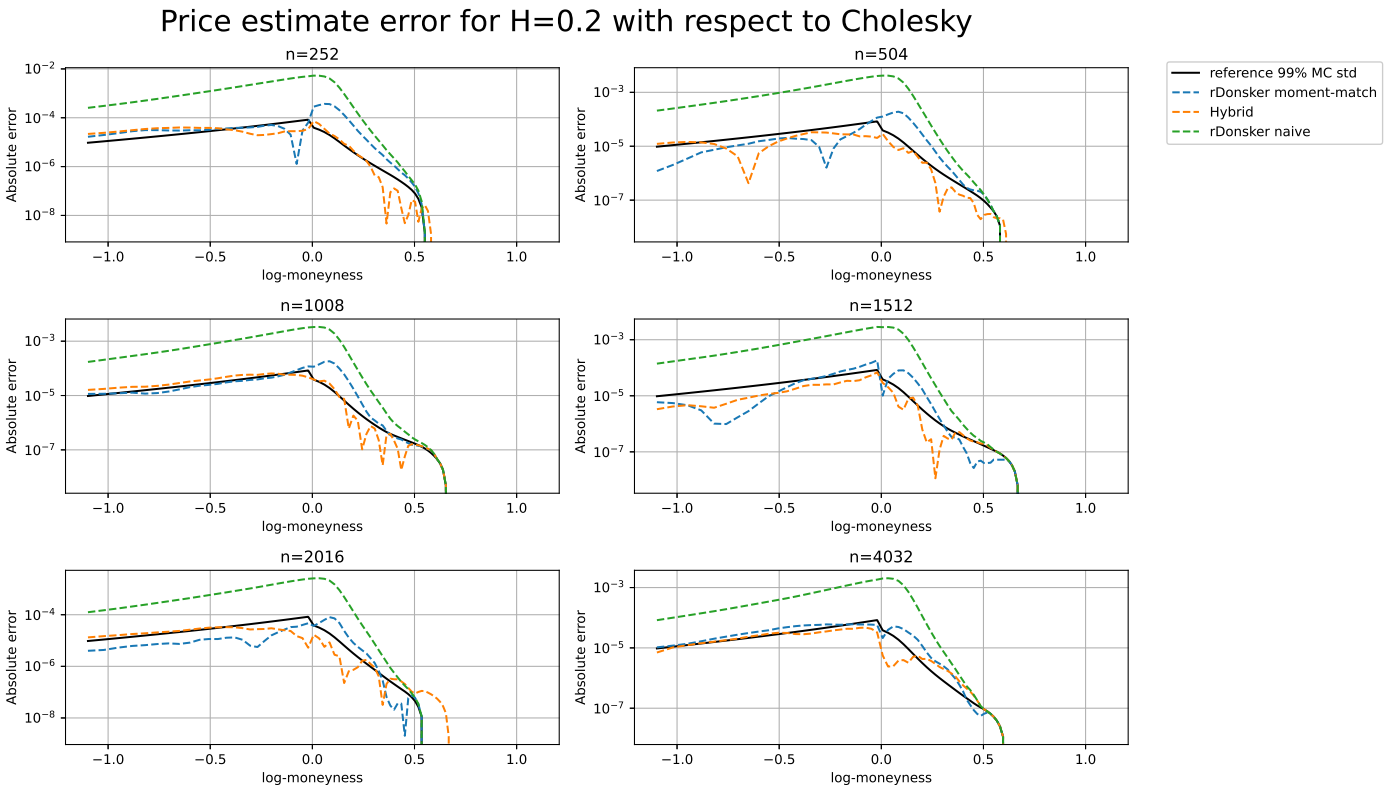


Figure 2.9: Rough Bergomi Call options price comparison with $H = 0.2, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

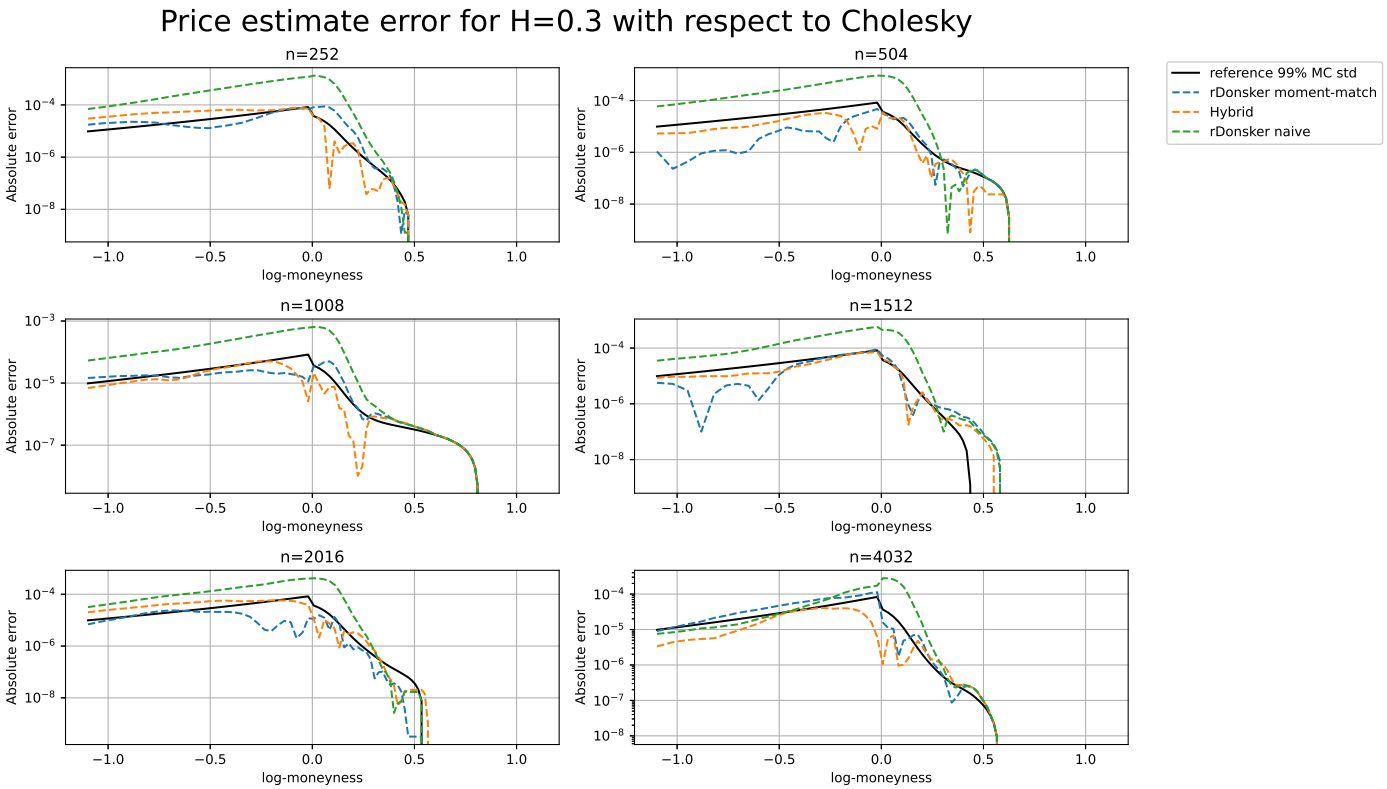


Figure 2.10: Rough Bergomi Call options price comparison with $H = 0.3, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

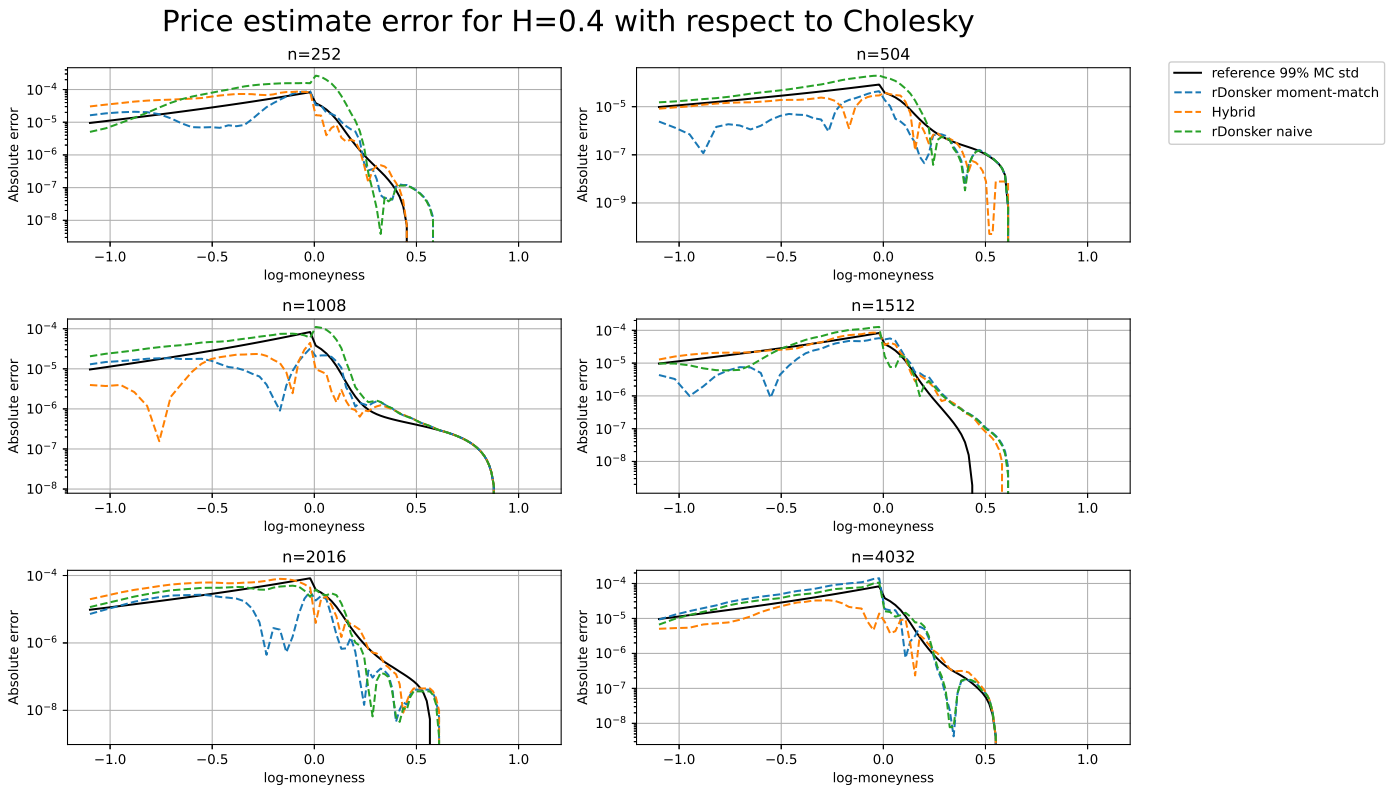


Figure 2.11: Rough Bergomi Call options price comparison with $H = 0.4, \xi_0 = 0.04, \nu = 2.3, \rho = -0.9, S_0 = 1, T = 1$ with $2 \cdot 10^6$ simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes.

2.4.5 Speed benchmark against Markovian stochastic volatility models

In this section we benchmark the speed of the rDonsker scheme against the Hybrid scheme and a classical Markovian stochastic volatility model using 10^5 simulations and averaging the speeds over 10 trials. For the former we simulate the rBergomi model [BFG16], whereas for the latter we use the classical Bergomi [Ber05] model using a forward Euler scheme in both volatility and stock price. All three schemes are implemented in `Cython` to make the comparison fair, and to obtain speeds comparable to `C++`. Figure 2.12 shows that rDonsker is about twice slower than the Markovian case whereas the Hybrid scheme is approximately 2.5 times slower, which is expected from the complexities of both schemes. However, it is remarkable that the $\mathcal{O}(n \log n)$ complexity of the FFT stays almost constant with the grid size n and the computational time grows almost linearly as in the Markovian case. We presume that this is the case since $n \ll 10000$ is relatively small. Figure 2.12 also shows that rough volatility models can be implemented very efficiently and are not particularly slower than classical stochastic volatility models.

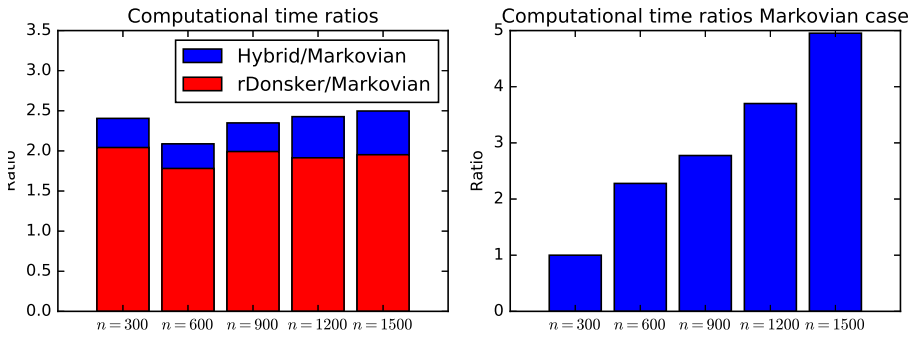


Figure 2.12: Computational time benchmark using Hybrid scheme, rDonsker and Markovian (forward Euler) for different grid sizes n.

2.4.6 Implementation guidelines and conclusion

The numerical analysis above suggests some guidelines to implement rough volatility models driven by $\mathfrak{tBSS}$ processes of the form $\mathcal{G}^{H-1/2}Y$, for some Itô diffusion Y:

$H > 0.1$	$H \in [0.05, 0.1]$	$H < 0.05$
rDonsker	choice depends on error sensitivity	Hybrid scheme

Regarding empirical estimates, Gatheral, Jaisson and Rosenbaum [GJR18] suggest that $H \approx 0.15$. Bennedsen, Lunde and Pakkanen [BLP16] give an exhaustive analysis of more than 2000 equities for which $H \in [0.05, 0.2]$. On the pricing side, Bayer, Friz and Gatheral [BFG16] and Jacquier, Martini and Muguruza [JMM18] (see Chapter 5) found that calibration routines yield $H \in [0.05, 0.10]$. Finally, Livieri, Mouti, Pallavicini and Rosenbaum [Liv+18] found evidence in options data that $H \approx 0.3$. Despite the diverse ranges found so far, there is a common agreement that $H < 1/2$.

Remark 2.4.6. The rough Heston model presented by Guennoun, Jacquier, Roome and Shi [Gue+18] is out of the scope of the Hybrid scheme. Moreover, any process of the form $\mathcal{G}^{\alpha}Y$, for some Itô diffusion Y under Assumptions 2.2.4 is, in general, out of the scope of the Hybrid scheme. This only leaves the choice of using the rDonsker scheme, for which reasonable accuracy is obtained at least for Hölder regularities greater than 0.05.

2.4.7 Bushy trees and binomial markets

Binomial trees have attracted a lot of attention from both academics and practitioners, as their apparent simplicity provides easy intuition about the dynamics of a given asset. Not only this, but they are by construction arbitrage free and allow to price path-dependent options, together with their hedging strategy. In particular, early exercise options, in particular Bermudan or American options, are usually priced using trees, as opposed to Monte-Carlo methods. The convergence stated in Conjecture 2.2.11 opens the door to the theoretical foundations to construct fractional binomial trees (note that Bernoulli random variables satisfy the conditions of the theorem). Figure 2.1 already showed binomial trees for fractional Brownian motion, but we ultimately need trees describing the dynamics of the stock price.

A binary market

We use Conjecture 2.2.11 with the independent sequences $\{\zeta_i\}_{i=1}^n$, $\{\zeta_i^\perp\}_{i=1}^n$ such that $\mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i^\perp = 1) = \mathbb{P}(\zeta_i = 1) = \mathbb{P}(\zeta_i^\perp = -1) = \frac{1}{2}$ for all i. We further define, on $\mathcal{T}$, for any $i = 1, \dots, n$,

$$B_n(t_i) = \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \left(\rho\zeta_k + \overline{\rho}\zeta_k^\perp\right),$$
$$Y_n(t_i) = \frac{T}{n} \sum_{k=1}^{i} b\left(Y_n(t_{k-1})\right) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sigma\left(Y_n(t_{k-1})\right)\zeta_k,$$

the approximating sequences to B and Y in (2.3). The approximation for X is then given by

$$X_n(t_i) = X_n(t_{i-1}) - \frac{1}{2}\frac{T}{n} \sum_{k=1}^{i} \Phi\left(\mathcal{G}^{\alpha}Y_n\right)(t_k) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sqrt{\Phi\left(\mathcal{G}^{\alpha}Y_n\right)(t_k)}\left(\rho\zeta_k + \overline{\rho}\zeta_k^\perp\right).$$

In order to construct the tree we have to consider all possible permutations of the random vectors $\{\zeta_i\}$ and $\{\zeta_i^\perp\}$. Since each random variable only takes two values, this adds up to 4^n possible combinations, hence the 'bushy tree' terminology. When $\rho \in \{-1, 1\}$, the magnitude is reduced to 2^n.

2.4.8 American options in rough volatility models

There is so far no available scheme for American options (or any early-exercise options for that matter) under rough volatility models, but the fractional trees constructed above provide a framework to do so. In the Black-Scholes model, American options can be priced using binomial trees by backward induction. American options and their pricing via simulation methods have been extensively studied in the literature (see for example Rogers [Rog03] or Longstaff and Schwartz [LS01]), however their application to non-Markovian models has not been studied nor is transparent from the implementation point of view. A key ingredient is the Snell envelope [Sne52] and the following representation by El Karoui [El 81] ($\widetilde{\mathbb{I}}$ denotes the set of stopping times with values in $\mathbb{I}$):

Definition 2.4.7. Let $(X_t)_{t\in\mathbb{I}}$ be an $(\mathcal{F}_t)_{t\in\mathbb{I}}$ adapted process, and $\tau \in \widetilde{\mathbb{I}}$. The Snell envelope $\mathcal{J}$ of X is defined as $\mathcal{J}(X)(t) := \operatorname{ess\,sup}_{\tau\in\widetilde{\mathbb{I}}} \mathbb{E}(X_\tau|\mathcal{F}_t)$ for all $t \in \mathbb{I}$.

In plain words, the Snell envelope of X is the smallest supermartingale that dominates it. Strictly speaking, it is necessary for X_τ to be uniformly integrable for any $\tau \in \widetilde{\mathbb{I}}$. Following [Kar88], an American option is nothing else than the smallest supermartingale dominating its European counterpart:

Definition 2.4.8. Let $C^e_t(k,T)$ and $P^e_t(k,T)$ denote European Call and Put prices at time t, with log-strike k and maturity T. Then the American counterparts, $C^a_t(k,T)$ and $P^a_t(k,T)$, are given by

$$C^a_t(k,T) = \mathcal{J}(C^e(k,T))(t) \qquad \text{and} \qquad P^a_t(k,T) = \mathcal{J}(P^e(k,T))(t).$$

Preservation of weak convergence under the Snell envelope map is due to Mulinacci and Pratelli [MP98], who proved that convergence takes place in the Skorokhod topology only if the Snell envelope is continuous. In our setting, the scheme for American options would be justified by the following result which builds upon Conjecture 2.2.11

Conjecture 2.4.9. For V in (2.3), if e^X is a true martingale then $\left(\mathcal{J}(e^{X_n})\right)_{n\geq 1}$ converges weakly to $\mathcal{J}\left(e^X\right)$ in the Skorokhod topology $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$.

Proof. Since the sequence $(X_n)_{n\geq 1}$ converges weakly to X in $(\mathcal{D}_{\mathbb{R}}(\mathbb{I}), \|\cdot\|_{\mathcal{D}})$ by Conjecture 2.2.11, for X in (2.3), the result follows using the Continuous Mapping Theorem if we can show that $\mathcal{J}$ is continuous. El Karoui proved in [El 81, Chapter 2.14] that the Snell envelope of an optional process, uniformly integrable for all stopping times $\tau \in \widetilde{\mathbb{I}}$, is continuous. To prove the proposition, we therefore only need to check uniform integrability of the stock price e^X. As $\mathbb{I}$ is a finite time horizon, Doob's optimal stopping theorem for martingales gives $e^{X_t} = \mathbb{E}[e^{X_1}|\mathcal{F}_t]$ for all $t \in \mathbb{I}$, thus e^X on $\mathbb{I}$ is uniformly integrable and the result follows. □

Mulinacci and Pratelli [MP98] also gave explicit conditions for weak convergence to be preserved in the Markovian case. It is trivial to see that the pricing of American options in the rough tree scheme coincides with the classical backward induction procedure. We consider continuously compounded interest rate r and dividend yield d.

Algorithm 2.4.10 (American options in rough volatility models)**.** On the equidistant grid $\mathcal{T}$,

1. construct the binomial tree using the explicit construction in Section 2.4.7 and obtain $\{S^j_t\}_{t\in\mathcal{T}, j=1,\dots,4^n}$;
2. the backward recursion for the American with exercise value $h(\cdot)$ is given by $\widetilde{h}_{t_N} := h(S_{t_N})$ and

$$\widetilde{h}_{t_i} := \mathrm{e}^{(d-r)/n}\mathbb{E}\left(\widetilde{h}_{t_{i+1}}|\mathcal{F}_{t_i}\right) \vee h(S_{t_i}), \qquad \text{for } i = N-1,\dots,0,$$

   where $\mathbb{E}(\cdot|\mathcal{F}_{t_i}) = \frac{1}{4}\left(\widetilde{h}^{++}_{t_{i+1}} + \widetilde{h}^{+-}_{t_{i+1}} + \widetilde{h}^{-+}_{t_{i+1}} + \widetilde{h}^{--}_{t_{i+1}}\right)$ and $\widetilde{h}^{\pm\pm}_{t_i}$ represents the outcome $(\zeta_i, \zeta_i^\perp) = (\pm 1, \pm 1)$ for the driving binomials, following the construction in Section 2.4.7.
3. finally, $\widetilde{h}_0$ is the price of the American option at inception of the contract.

The main computational cost of the scheme is the construction of the tree in Step 1. Once the tree is constructed, computing American prices for different options is a fast routine.

Numerical example: rough Bergomi model

The rough Bergomi model satisfies the martingale property in Theorem 2.4.9 (b) for $\rho \leq 0$ (see Gassiat [Gas19]). We construct a rough volatility tree for the rough Bergomi model [BFG16] and check the accuracy of the scheme. Figures 2.13 and 2.14 show the fractional trees for different values of H and for $\rho \in \{-1, 1\}$. Both pictures show a markedly different behaviour, but as a common property we observe that as H tends to $1/2$, the tree structure somehow becomes simpler.

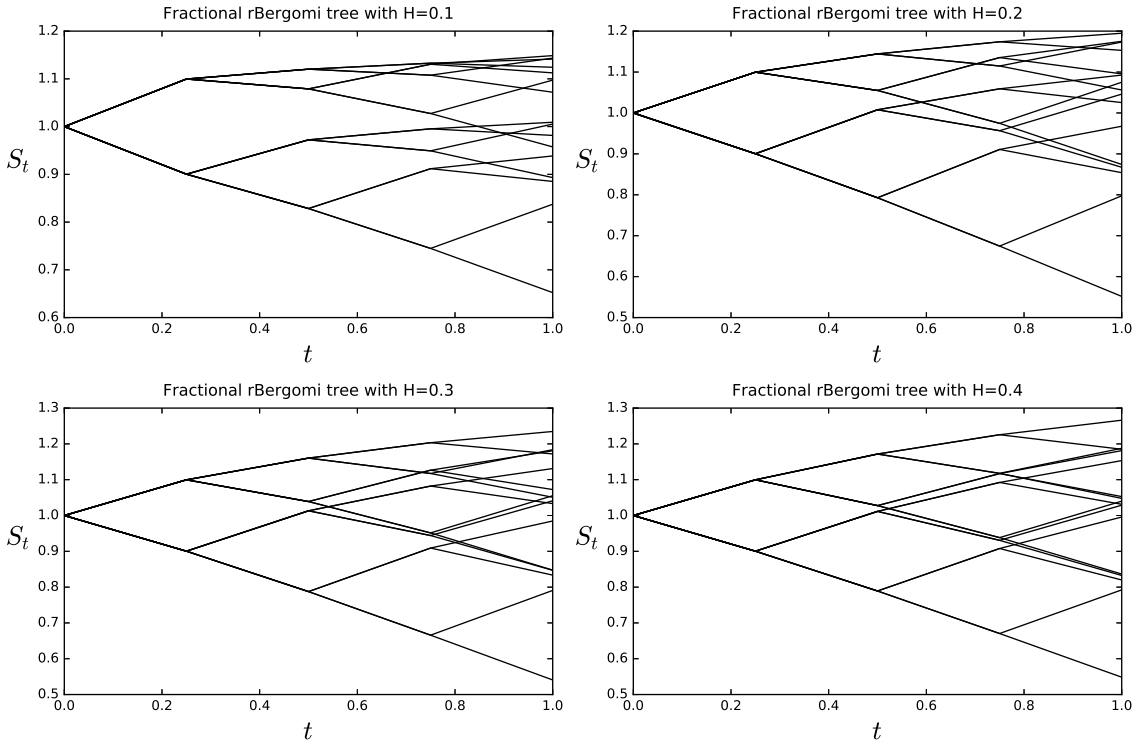


Figure 2.13: Rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 5 time steps.

European options

Figure 2.15 displays volatility smiles obtained using the tree scheme. Even though the time steps are not sufficient for small H, the fit remarkably improves when $H \geq 0.15$, and always remains inside the 95% confidence interval with respect to the Hybrid scheme. Moreover, the moment-matching approach from Section 2.4.3 shows a superior accuracy when $H \leq 0.1$, but is not sufficiently accurate. In Figure 2.16 a detailed error analysis corroborates these observations: the relative error is smaller than 3% for $H \geq 0.15$.

American options

In the context of American options, there is no benchmark to compare our result. However, the accurate results found in the previous section (at least for $H \geq 0.15$) justify the use of trees to price American options. Figure 2.17 shows the output of American and European Put prices with interest rates equal to $r = 5\%$. Interestingly, the rougher the process (the smaller the H), the larger the difference between in-the-money European and American options.

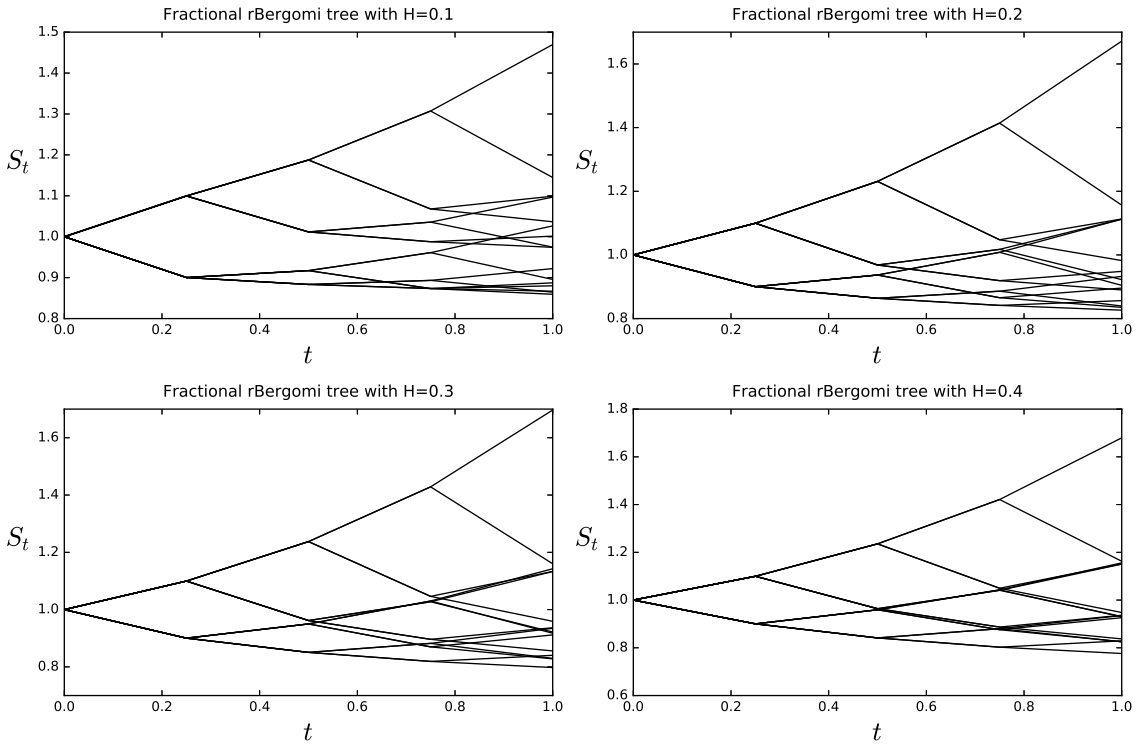


Figure 2.14: rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, 1, 0.04)$ with 5 time steps.

2.5 Summary

In this Chapter, we have formally proven weak convergence of fractional diffusions, which represent the instantaneous variance or volatility in stochastic volatility models. In spite of this, the conjecture of weak convergence results of the log-stock process by means of Euler schemes applied to rough stochastic volatility models remains a topic for future work. Furthermore, we have formally extended the definition of Riemann-Liouville fractional operators to a family of generalised fractional operators, which allow us to transform Itô diffusions into fractional difussions. We have also presented several numerical algorithms to improve the performance of Monte-Carlo simulations and explored the use of fractional binomial trees to price American options under rough volatility. Notably, we find that the computational cost of simulating non-Markovian models is roughly two times slower than their Markovian counterparts by means of efficient FFT transforms. This findings, make clear that the computational cost of rough models is much smaller than initially though in [BFG16], and advocate in favour of their use in production environments.

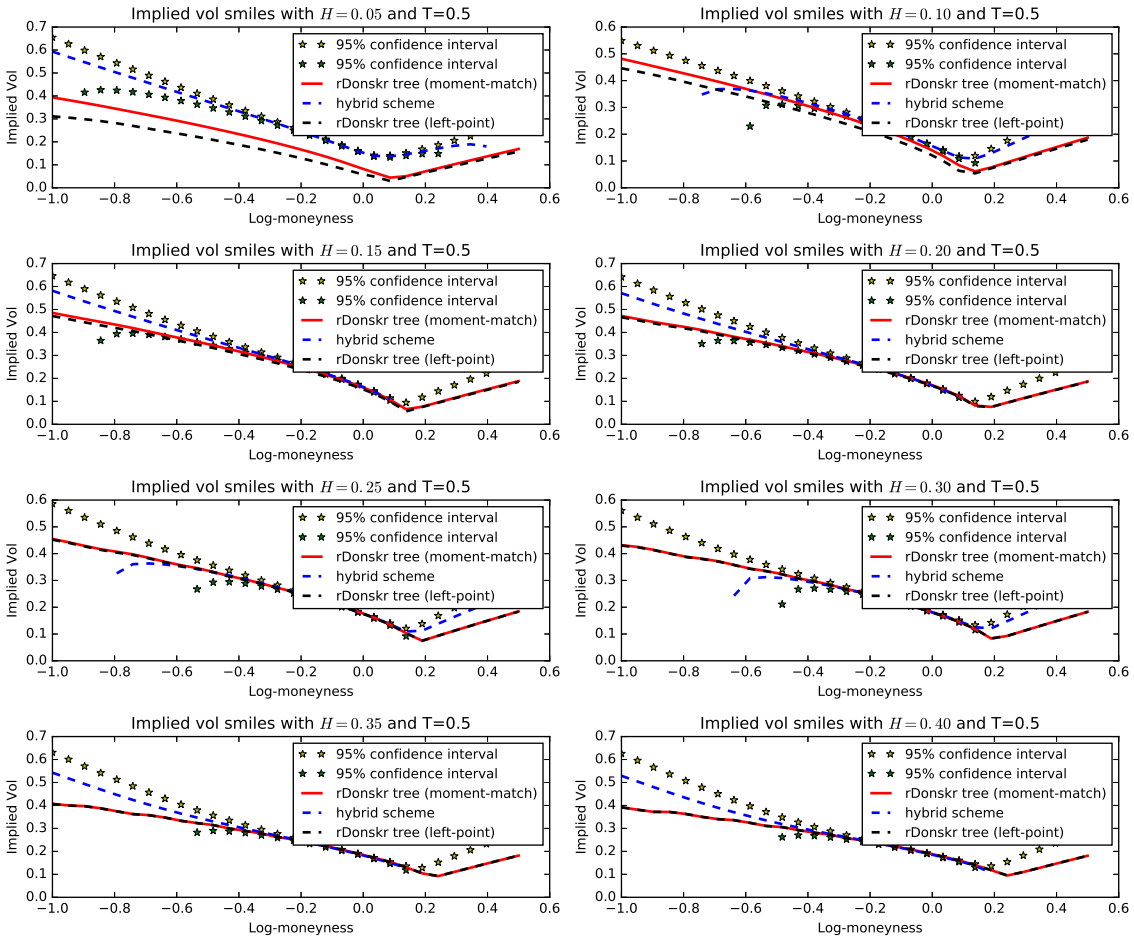


Figure 2.15: rough Bergomi trees for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps.

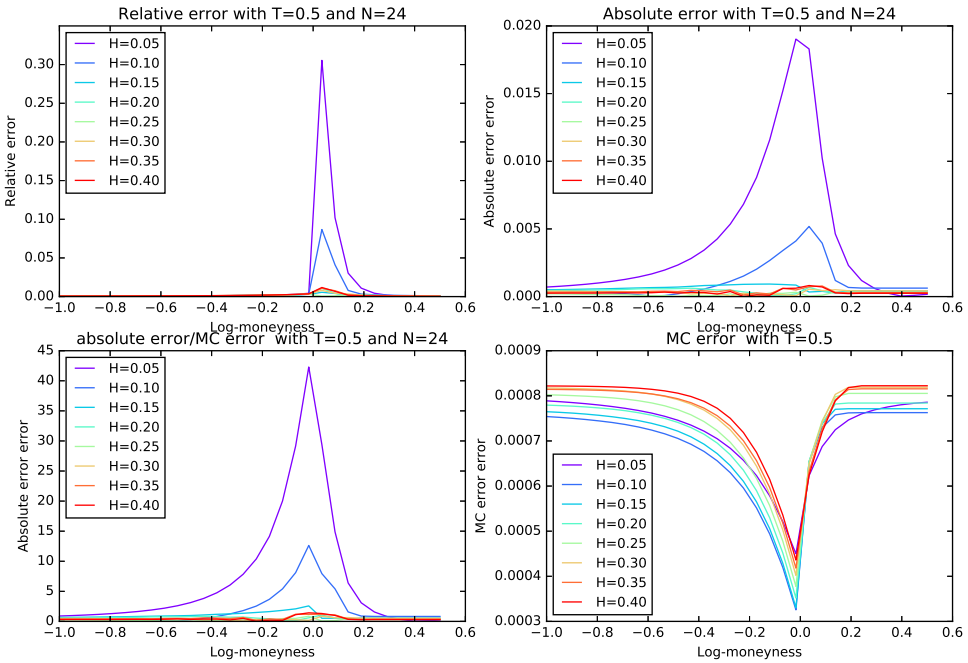


Figure 2.16: Error analysis for the rDonsker moment-match tree for different values of H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 24 time steps.

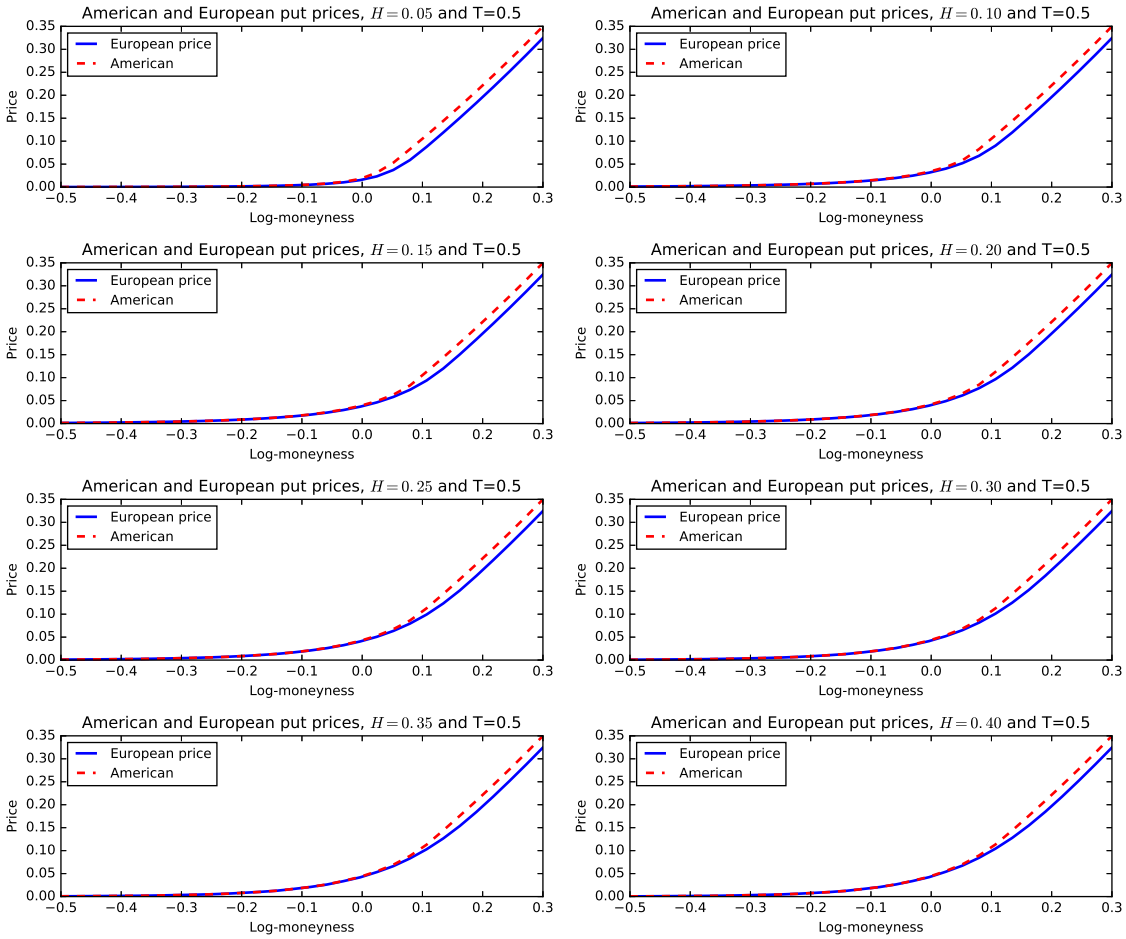


Figure 2.17: American and European Put prices in the rough Bergomi model for different values of H and $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 26 time steps.

3

Deep Learning Volatility:
A deep neural network approach to pricing and calibration in (rough) volatility models

3.1 Introduction

We note that the outcome of this chapter is a unified version of the pre-prints of [HMT21] and [BS18] that were later published as [Bay+19c] and [HMT21]. The reason to unify was that both original papers developed in parallel tacked the same issue from a different angle. A unified framework merging the two approached provided a stronger case.

Approximation methods for option prices came in all shapes and forms in the past decades and they have been extensively studied in the literature and well-understood by risk managers. Clearly, the applicability of any given option pricing method (Fourier pricing, PDE methods, asymptotic methods, Monte Carlo, ... etc.) depends on the regularity properties of the particular stochastic model at hand. Therefore, tractability of stochastic models has been one of the most decisive qualities in determining their popularity. In fact it is often a more important quality than the modelling accuracy itself: It was the (almost instantaneous) SABR asymptotic formula [Hag+02] that helped SABR become the benchmark model in fixed income desks, and similarly the convenience of Fourier pricing is largely responsible for the popularity of the Heston model, despite the well-known hiccups of these models. Needless to say that it is the very same reason (the concise Black Scholes formula) that still makes the Black-Scholes model attractive for calculations even after many generations of more realistic and more accurate stochastic market models have been developed. On the other end of the spectrum are rough volatility models, for which (despite a plethora of modelling advantages, see [BFG16; ER18; GJR18] to name a few) the necessity to rely on relatively slow Monte Carlo based pricing methods creates a major bottleneck in calibration, which has proven to be a main limiting factor with respect to industrial applications. This dichotomy can become a headache in situations when we have to weigh up the objectives of accurate pricing vs. fast calibration against one another in the choice of our pricing model. In this work we explore the possibilities provided by the availability of an algorithm that –for a choice of model parameters– directly outputs the corresponding vanilla option prices (as the Black-Scholes formula does) for a large range of maturities and strikes of a given model.

In fact, the idea of mapping model parameters to shapes of the implied volatility surface directly is not new. The stochastic volatility inspired SSVI, eSSVI surfaces (see [GJ14; HM19]) do just that: A given set of parameters is translated directly to different shapes of (arbitrage-free) implied volatility surfaces, bypassing the step of specifying

any stochastic dynamics for the underlying asset. For stochastic models that admit asymptotic expansions, such direct mappings from model parameters to (approximations of) the implied volatility surface in certain asymptotic regimes can be obtained (one example is the famous SABR formula). Such asymptotic formulae are typically limited to certain asymptotic regimes along the surface by their very nature. Complementary to asymptotic expansions we explore here a direct (approximative) mapping from different parameter combinations of stochastic models to different shapes of implied volatility surface for intermediate regimes. It's appeal is that it combines the advantages of direct parametric volatility surfaces (of the SSVI family) with the possibility to link volatility surfaces to the stochastic dynamics of the underlying asset.

In this Chapter we apply deep neural networks (merely) as powerful high-dimensional functional approximators to approximate the multidimensional pricing functionals from model parameters to option prices. The advantage of doing so via deep neural networks over standard (fixed-basis) functional approximations is that deep neural networks are agnostic to the approximation basis [GBC16]. A particular achievement in this work is to identify a network that is able to handle model families where the forward variance curve is taken as an input. The architecture is then robustly applicable to several simpler stochastic models consistently without modification. The advantage of this approach lies in moving the (often time-consuming) numerical approximation of the pricing functional into an off-line preprocessing step, and the on-line calibration part can then be performed considerably faster. This preprocessing amounts to storing the approximative direct pricing functional in form of the network weights after a supervised training procedure: Using available numerical approximations of option prices as ground truth (in a stochastic model of our choice), we train a neural network to learn an accurate approximation of the pricing functional. After training, the network outputs–for any choice of model parameters–the corresponding implied volatilities within milliseconds for a large range of maturities and strikes along the whole surface. Furthermore, we show that this procedure generalises well for unseen parameter combinations: the accuracy of price approximation of our neural network pricing functional on out-of-sample data is within the same range as the accuracy of the original numerical approximation used for training. The accuracy of this direct pricing map is demonstrated in our numerical experiments.

The work we present here is very much in the spirit of the pioneering work of Avellaneda, Carelli and Stella [ACS99]. Bearing in mind that deep neural network solutions are often challenged by concerns of generalisation and "black-box-solutions", our goal is to limit the application of neural networks to parts of the calibration process that we can control and validate. As a first step, we identify the parts of the calibration process that are mainly responsible for the prevailing calibration bottlenecks, which we will replace by a deep neural network.

In this context, we distinguish two kinds of approaches. The first, pioneered by Hernandez [Her17], seeks to learn the mapping from implied volatility surfaces to model parameters (inverse problem) directly. In [Her17], Hernandez proposes to use a neural network to learn the complete calibration routine taking market data as inputs and returning calibrated model parameters, and calibrates the popular short rate model of Hull and White [HW90] to market data in numerical experiments (In Section 3.3.1 we describe this approach in more detail). In the rest of this Chapter, we will refer to it as the one-step approach. In a second strand of research neural networks have been applied not directly to calibration problems, but simply to obtain an approximated representation of derivative valuations, i.e. of option pricing maps: For example Hutchinson, Lo and Poggio [HLP94] such as Culkin and Das [CD17] applied neural networks to learn the Black-Scholes formula and McGhee demonstrates in [McG18] a neural networks representation of the lognormal SABR model. For a detailed survey on different use cases of Neural Networks in the hedging and pricing literature we refer to [RW20]. In this Chapter we explore the advantages of shaping this second strand of research into a building block of a single two-step approach.

The two-step approach, which we highlight here, first approximates the pricing map, by a neural network (Step (i)) before calibrating the model, via traditional calibration algorithms, to market data (Step (ii)). Thereby we optimally leverage the capability of neural networks to approximate functions which are only implicitly available through input-output pairs $\{(x_i, \phi(x_i))\}_{i=1}^N$, by training a fully-connected neural network on specifically tailored, synthetically generated training data to learn an approximative representation of the pricing functional. Details of this approach and its benefits are further explained in Section 3.3.

One of the striking advantages of this two step approach is that it speeds up the (on-line part of the) calibration to the realm of just a few milliseconds. There have been several recent contributions on neural network calibrations of stochastic models [Bay+19c; Bue+19; Her17; Kon18; De +18]. Clearly, even once one settled for the two step approach, much depends on the finesse of the particular network design with respect to the performance of these networks as shown in [Bay+19c]. With the appropriate design, even non-Markovian models such as Rough Volatility can be calibrated in a reasonable time. Several of the first contributions using the two-step approach focus on a particular model or variants thereof [FG18; McG18]. Typically if a network architecture works for a more challenging class of of models (say higher dimensional or non-Markovian), it also works for simpler ones. One contribution of this work is to achieve a fast and accurate calibration of the rough Bergomi model of [BFG16] with a general forward variance curve (piecewise approximated by constant functions), which contains tWo particular difficulties: Capturing the forward variance curve as an input, which inherently poses a high-dimensionality on the problem, and the non-Markovianity coming from the rough volatility. The good performance of (the same) network architecture then also carries over to simpler models containing a forward variance curve but which do not exhibit the non-Markovianity (this is demonstrated on the the one-factor classical Bergomi model) and to classical models such as SABR and Heston. To demonstrate this, we first perform calibration experiments on simulated data and show calibration accuracy in controlled experiments. The examples of the Heston model, the one-factor Bergomi model and the rough Bergomi model (with a full forward variance curve) are included as demonstration examples in the Github repository GitHub: NN-StochVol-Calibrations. Though the speedup in calibration time can provide a considerable advantage also in the case of classical models, we consider the most pronounced breakthrough of this calibration speedup for the rough Bergomi model with full forward variance curve, where previously calibration bottlenecks have put constraints on its applicability in industry practice. Therefore in the last part of the Chapter we concentrate the presentation of our numerical experiments on this model. To showcase the speed and prowess of the approach we calibrate the rough Bergomi model to historical data and display the evolution of parameters on a dataset consisting of 10 years of SPX data.

The Chapter is organised as follows: In Sections 3.2 and 3.3 we present a neural network perspective on model calibration and recall stochastic models that are considered in later sections. In this section we also formalise our objectives about the accuracy and speed of neural network approximation of pricing functionals and the basic ideas of our training. Section 3.3.3 recalls some background on neural networks as functional approximators and some aspects of neural network training that influenced the setup of our network architecture and training design. In Section 3.3.4 we describe the network architecture and training of the price approximation network such as the calibration methods we consider. In Section 3.4 we present numerical experiments of price approximations of vanilla and some exotic options, calibration to synthetic data and to historical data.

The algorithm alongside with the presented numerical experiments and examples are provided on GitHub: NN-StochVol-Calibrations, where a code demo of our results can be downloaded. There, in addition to the models presented in the present text we also provide further examples of standard stochastic models (in particular the Heston model) where this approach is demonstrated to work well. The application can be carried over to other classical models analogously.

3.2 Model calibration

Calibration describes the procedure of tuning model parameters to fit a model surface to an empirical implied volatility surface obtained by transforming liquid European option market prices to Black-Scholes implied volatilities. A mathematically convenient approach consists of minimizing the weighted squared differences between market and model implied volatlities of $N \in \mathbb{N}$ plain vanilla European options.

Suppose that a model is parametrized by a set of parameters Θ, i.e., by $\theta \in \Theta$. Furthermore, we consider options

parametrized by a parameter $\zeta \in Z$. E.g., for put and call options we generally have $\zeta = (T, k)$, the option's maturity and log-moneyness. There might be further parameters which are needed to compute prices but can be observed on the market and, hence, do not need to be calibrated. For instance, the spot price of the underlying, the interest rate, or the forward variance curve in Bergomi-type models (see [Ber16]) falls under this type. For this quick overview, we ignore this category. We introduce the pricing map

$$(\theta, \zeta) \mapsto P(\theta, \zeta),$$

the price of an option with parameters ζ in the model with parameters θ. We are also given market prices $\mathcal{P}^{\mathrm{MKT}}(\zeta)$ for options parametrized by ζ for a (finite) subset $\zeta \in Z' \subset Z$ of all possible option parameters. Calibration now identifies a model parameter θ which minimizes a chosen distance δ between model prices $(P(\theta, \zeta))_{\zeta \in Z'}$ and market prices $\left(\mathcal{P}^{\mathrm{MKT}}(\zeta)\right)_{\zeta \in Z'}$, i.e.,

$$\widehat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \delta \left((P(\theta, \zeta))_{\zeta \in Z'}, \left(\mathcal{P}^{\mathrm{MKT}}(\zeta)\right)_{\zeta \in Z'} \right). \tag{3.1}$$

Remark 3.2.1. Financial practice often prefers to work with implied volatilities rather than option prices, and we will also do so in the numerical parts of this Chapter. For the purpose of this introduction, any mentioning of a *price* may be, mutatis mutandis, replaced by the corresponding implied volatility.

3.2.1 The numerical optimisation

The most usual choice of a distance function δ is a suitably weighted least squares function, i.e.,

$$\widehat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \sum_{\zeta \in Z'} w_\zeta \left(P(\theta, \zeta) - \mathcal{P}^{\mathrm{MKT}}(\zeta) \right)^2. \tag{3.2}$$

Here, the weights w_ζ can be chosen in order to reflect importance of an option at ζ and the reliability of the market observation $\mathcal{P}(\zeta)$. For instance, a reasonable choice might be the inverse of the bid-ask spread (see [Con10] for a motivation), which puts low weight on prices of illiquid options.

As long as the number of model parameters is smaller than the number $|Z'|$ of calibration instruments, the calibration problem is an example of an overdetermined non-linear least squares problem, where a closed-form solution hardly ever exists. Hence, numerical routines are necessary to succesfully solve (3.1) and (3.2) in particular; we outline two possible approaches:

Gradient-based optimisers

In general, for financial models the pricing map P is assumed to be smooth (at least C^1 differentiable) with respect to all its input parameters θ. In such setting, a standard necessary first order condition for optimality in (3.1) is that

$$\nabla^\theta \delta \left((P(\theta, \zeta))_{\zeta \in Z'}, \left(\mathcal{P}^{\mathrm{MKT}}(\zeta)\right)_{\zeta \in Z'} \right) = 0, \tag{3.3}$$

provided that the objective function is smooth. Then, a natural update rule is to move along the gradient via Gradient Descent i.e.

$$\theta_{i+1} = \theta_i - \lambda \nabla^\theta \delta \left((P(\theta, \zeta))_{\zeta \in Z'}, \left(\mathcal{P}^{\mathrm{MKT}}(\zeta)\right)_{\zeta \in Z'} \right), \quad \lambda > 0. \tag{3.4}$$

A common feature of gradient based optimization methods building on (3.4) is the use of the gradient $\nabla^\theta \delta \left((P(\theta, \zeta))_{\zeta \in Z'}, \left(\mathcal{P}^{\mathrm{MKT}}(\zeta)\right)_{\zeta \in Z'} \right)$, hence its correct and precise computation is crucial for subsequent success. Examples of such algorithms, are Levenberg-Marquardt [Lev44; Mar63], Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [NW06], L-BFGS-B [Zhu+97] and SLSQP [Kra88]. The main advantage of the aforementioned methods is the quick convergence towards condition (3.3). However, (3.3) only gives necessary and not sufficient conditions for optimality, hence special care must be taken with non-convex problems.

Remark 3.2.2. The appeal to use neural networks to approximate the pricing function is that they provide closed-form expressions for gradients along with their theoretical guarantee of convergence (see Theorem 3.3.7).

Gradient-free optimisers

Gradient-free optimization algorithms are gaining popularity due to the increasing number of high dimensional non-linear, non-differentiable and/or non-convex problems flourishing in many scientific fields such as biology, physics or engineering. As the name suggests, gradient-free algorithms make no C^1 assumption on the objective function. Perhaps, the most well known example is the Simplex based Nelder-Mead [NM65] algorithm. However, there are many other methods such as COBYLA [Pow94] or Differential Evolution [SP97] and we refer the reader to [RS13] for an excellent review on gradient-free methods. The main advantage of these methods is the ability to find global solutions in (3.1) regardless of the objective function. In contrast, the main drawback is a higher computational cost compared to gradient methods.

To conclude, we summarise the advantages of each approach in Table 3.1.

	Gradient-based	**Gradient-free**
Convergence Speed	Very Fast	Slow
Global Solution	Depends on problem	Always
Smooth activation function needed	Yes to apply Theorem 3.3.7	No
Accurate gradient approximation needed	Yes	No

Table 3.1: Comparison of Gradient vs. Gradient-free methods.

3.2.2 A showcase of (rough) Stochastic Volatility models

We would like to emphasize that our methodology can in principle be applied to any (classical or rough) volatility model. From the classical Black Scholes or Heston models to the rough Bergomi model of [BFG16], also to large class of rough volatility models (see Chapter 2 for a general setup). In fact the methodology is not limited to stochastic models, also parametric models of implied volatility could be used for generating training samples of abstract models, but we have not pursued this direction further. For completeness, we showcase some models for which numerical results are provided in the present text (for a few more examples we refer the interested reader to GitHub: NN-StochVol-Calibrations).

The Rough Bergomi model

In the abstract model framework, the rough Bergomi model is represented by $\mathcal{M}^{rBergomi}(\Theta^{rBergomi})$, with parameters $\theta = (\xi_0, \nu, \rho, H) \in \Theta^{rBergomi}$. On a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ the model corresponds to the following system

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t} dW_t, \quad \text{for } t > 0, \quad X_0 = 0, \\ V_t &= \xi_0(t)\mathcal{E}\left(\sqrt{2H}\nu \int_0^t (t-s)^{H-1/2} dZ_s\right), \quad \text{for } t > 0, \quad V_0 = v_0 > 0 \end{aligned} \tag{3.5}$$

where $H \in (0,1)$ denotes the Hurst parameter, $\nu > 0$, $\mathcal{E}(\cdot)$ the stochastic exponential [Dol70], and $\xi_0(\cdot) > 0$ denotes the initial forward variance curve (see [Ber16, Section 6]), and W and Z are correlated standard Brownian motions with correlation parameter $\rho \in [-1, 1]$. To fit the model parameters into our abstract model framework $\Theta^{rBergomi} \subset \mathbb{R}^n$ for some $n \in \mathbb{N}$, the initial forward variance curve $\xi_0(\cdot) > 0$ is approximated by a piecewise constant function in our numerical experiments in Sections 3.4.1, and 3.4.2.

The Bergomi model

In the general n-factor Bergomi model, the volatility is expressed as

$$V_t = \xi_0(t)\mathcal{E}\left(\eta_i \sum_{i=1}^{n} \int_0^t \exp\left(-\kappa_i(t-s)dW_s^i\right)\right) \quad \text{for } t>0, \quad V_0 = v_0 > 0, \tag{3.6}$$

where $\eta_1, \ldots, \eta_n > 0$ and $(W^1, \ldots, W^n)$ is an n-dimensional correlated Brownian motion, $\mathcal{E}(\cdot)$ the stochastic exponential [Dol70], and $\xi_0(\cdot) > 0$ denotes the initial forward variance curve, see [Ber16, Section 6] for details. In this work we consider the Bergomi model for $n = 1, 2$ in Section 3.4. Henceforth, $\mathcal{M}^{1FBergomi}(\xi_0, \beta, \eta, \rho)$ represents the 1 Factor Bergomi model, corresponding to the following dynamics:

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t \quad \text{for } t>0,\ X_0 = 0 \\ V_t &= \xi_0(t)\mathcal{E}\left(\eta \int_0^t \exp(-\beta(t-s))dZ_s\right) \quad \text{for } t>0, \quad V_0 = v_0 > 0, \end{aligned} \tag{3.7}$$

where $\nu > 0$, and W and Z are correlated standard Brownian motions with correlation parameter $\rho \in [-1, 1]$. To fit the model parameters into our abstract model framework $\Theta^{1FBergomi} \subset \mathbb{R}^n$, for some $n \in \mathbb{N}$, the initial forward variance curve $\xi_0(\cdot) > 0$ is approximated in our numerical experiments by a piecewise constant function in Sections 3.4.1, and 3.4.2.

3.3 Deep calibration

In the following sections we elaborate on the objectives and advantages of this two step calibration approach and present examples of neural network architectures, precise numerical recipes and training procedures to apply the two step calibration approach to a family of stochastic volatility models. We also present some numerical experiments and report the learning errors compared to chosen parameters of the synthetic data.

There are several advantages of separating the tasks of pricing and calibration. Above all, the most appealing reason is that it allows us to build upon the knowledge we have gained about the models in the past decades, which is of crucial importance from a risk management perspective. By its very design, (i) deep learning the price approximation combined with (ii) deterministic calibration does not cause more headache to risk managers and regulators than the corresponding stochastic models do. Designing the training as described above demonstrates how deep learning techniques can successfully extend the toolbox of financial engineering, without imposing the need for substantial changes in our risk management libraries.

3.3.1 One-step approach: Deep calibration by the inverse map

A more and more popular approach in quantitative finance (and many other fields of engineering) is to develop purely data-driven frameworks, without relying on formal models. This approach leaves the meaning of calibrated network parameters unexplained, not to mention the ambiguity about the choice of the number of network parameters and network design. This can cause major challenges towards today's regulatory requirements. In addition, issues of generalizability – how can one price exotic options in a network trained with vanilla option data, to give a simple example – are difficult to analyse, and traditional paradigms of finance – such as no arbitrage – are hard to guarantee in the absence of a model. This modeling approach is probably dominant in high frequency trading. We refer to reference [JL19] for one example of such a modeling approach for option pricing.

A second, more model based approach was proposed in the pioneering work of Hernandez [Her17], followed by several other authors such as Stone [Sto20], Dimitroff, Röder and Fries [DRF18] and many others. A main characteristic

of the neural network proposed by [Her17] is that option price approximation and parameter calibration are done in one step within the same network. Indeed, the idea is to directly learn the whole calibration problem, i.e., to learn the model parameters as a function of the market prices (typically parametrized as implied volatilities). In the formulation of Section 3.2, this means that we learn the mapping

$$\Pi^{-1} : (\mathcal{P}(\zeta))_{\zeta \in Z'} \mapsto \widehat{\theta}.$$

More precisely, [Her17] trains a deep neural network based on labelled data (x_i, y_i), $i = 1, \dots, N$, with

$$x_i = (\mathcal{P}(\zeta))_{\zeta \in Z_i'}$$

for day t_i (in the past) and the corresponding labels

$$y_i = \widehat{\theta}_i,$$

obtained from calibrating the model to the market data y_i using traditional calibration routines. The number of labelled data points N is, of course, limited to the amount of (reliable) historical market price data available.

In spite of the promising results by Hernandez [Her17] the main drawback of this approach, as Hernandez observes, is the lack of control on the function Π^{-1}. Furthermore, from a risk management perspective one has no guarantee how well the learned mapping of Π^{-1} will solve the calibration problem when exposed to unseen data. In fact, this is the behaviour observed in Hernandez [Her17], since the out of sample performance tends to differ from the in sample one, suggesting a not fully satisfactory generalisation of the learned map. We recover the same behaviour of the inverse map in our own experiments, which we included in Appendix B.

3.3.2 Two-step approach: Learning the implied volatility map of models

The two step approach is somewhere mid-way between a sole reliance on traditional pricing methods (Monte Carlo, finite elements, finite differences, Fourier methods, asymptotic methods etc.) and the direct approach described above that calibrate directly to the price data. Here, one separates the calibration procedure as described in Section 3.2 (i) We first learn (approximate) the pricing map by a neural network that maps parameters of a stochastic model to prices or implied volatilities. In other words, we set up and train (off-line) a neural network to learn the pricing map P. In a second step (ii) we calibrate (on-line) the model – as approximated by the neural network trained in step (i) – to market data using a standard calibration routine. To formalise the two step approach, for an option parametrized by ζ and a model $\mathcal{M}$ with parameters $\theta \in \Theta$ we write $\widetilde{P}(\theta, \zeta) \approx P(\theta, \zeta)$ for the approximation $\widetilde{P}$ of the true pricing map P based on a neural network. Then, in the second step, for a properly chosen distance function δ (and a properly chosen optimization algorithm) we calibrate the model by computing

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \delta \left(\left(\widetilde{P}(\theta, \zeta) \right)_{\zeta \in Z'}, (\mathcal{P}(\zeta))_{\zeta \in Z'} \right). \tag{3.8}$$

In principle, this method is not unlike traditional calibration routines, as the true option price has to be numerically approximated for all but the most simple models. This particular approximation method tends to be orders of magnitudes faster compared to other numerical approximation methods for all tested models. In particular, note that the (slow) training stage of the neural network itself only has to be done once. We will come back to comparisons of actual computational times in the numerical section of this Chapter.

At this stage, we note that the deep calibration routine is not yet specified in any details: apart from purely numerical details such as the choice of the architecture of the neural networks, the loss functions and optimization algorithms of both the training of the neural networks in stage (i) and the actual calibration in stage (ii), one particularly important choice is whether the neural network learns implied volatilities of individual options or rather a full implied volatility surface. Before discussing these details, let us already highlight some of the differences to the one-step approach of [Her17]. In principle, the one-step approach is orders of magnitude faster by construction, however we will demonstrate in Section 3.4 that the two-step approach calibrates within milliseconds making the

speed difference irrelevant in practice. Moreover, we see the main benefit of the two-step approach in the increased stability, which is influenced by two key differences:

- As the neural network is only responsible for option pricing in the model, synthetic data has to be used for training. Hence, we can easily increase the number of training data, and the training data are completely unpolluted from market imperfections.
- The two-step approach induces a natural decomposition of the overall calibration error into a pricing error (from the neural network) and a model misfit to the market data. Hence, the performance of the neural network itself is generally independent of changing market regimes – which might, of course, change the suitability of the model under consideration.

These points, in particular, imply that frequent re-training of the neural network is not needed in the two-step approach.

The two step approach: Pointwise training and implicit and grid-based training

The underlying principle of the two-step approach appears in one way or another in a number of related contributions De Spiegeleer, Madan, Reyners and Schoutens [De +18] and McGhee [McG18]. In fact, the early works of Hutchinson, Lo and Poggio [HLP94] and the more recent work of Culkin and Das [CD17]–where Deep Neural Networks are applied to learn the Black-Scholes formula–can be recognised as Step (i) of the two-step approach in a Black-Scholes context. Also Ferguson and Green [FG18] examine Step (i) of the two-step approach in [FG18] for basket options in a lognormal context and observe that the network even has a smoothing effect and increased accuracy in comparison to the underlying Monte Carlo prices. In this section, we examine its advantages and present an analysis of the objective function with the goal to enhance learning performance. Within this framework, the pointwise approach has the ability to asses the quality of $\widetilde{P}$ using Monte Carlo or PDE methods, and indeed it is superior in terms of robustness.

Pointwise learning

Step (i): Learn the map $\widetilde{P}(\theta, T, k) = \widetilde{\sigma}^{\mathcal{M}(\theta)}(T, k)$ – that is in equation (3.8) above we have $\zeta = (T, k)$. In the case of vanilla options ($\zeta = (T, k)$) one can rephrase this learning objective as an implied volatility problem: In the implied volatility problem the more informative implied volatility map $\widetilde{\sigma}^{\mathcal{M}(\theta)}(T, k)$ is learned, rather than call- or put option prices $\widetilde{P}(\theta, T, k)$. We denote the artificial neural network by $\widetilde{F}(w; \theta, \zeta)$ as a function of the weights w of the neural network, the model parameters θ and the option parameters ζ. The optimisation problem to solve is the following:

$$\widehat{\omega} := \operatorname*{argmin}_{w \in \mathbb{R}^n} \sum_{i=1}^{N_{\text{Train}}} \eta_i (\widetilde{F}(w; \theta_i, T_i, k_i) - \widetilde{\sigma}^{\mathcal{M}}(\theta_i, T_i, k_i))^2. \tag{3.9}$$

where $\eta_i \in \mathbb{R}_{>0}$ is a weight vector.

Step (ii): Solve the classical model calibration problem for the market quotes $\{\sigma_{\text{BS}}^{\text{MKT}}(k_j, T_j)\}_{j=1}^m$

$$\widehat{\theta} := \operatorname*{argmin}_{\theta \in \Theta} \sum_{j=1}^{m} \beta_j (\widetilde{F}(\widehat{w}; \theta_i, T_i, k_i) - \sigma_{\text{BS}}^{\text{MKT}}(k_j, T_j))^2.$$

for some user specified weights $\beta_j \in \mathbb{R}_{>0}$, where now the (numerical approximation of the) option price $\widetilde{P}(\theta, T, k)$ resp. implied volatility $\widetilde{\sigma}^{\mathcal{M}(\theta)}(T, k)$ is replaced by the DNN approximation $\widetilde{F}(\widehat{\omega}; \theta, T, k)$ obtained in Step (i). We note here that $\widetilde{F}(\widehat{w}; \theta_i, T_i, k_i)$ being a Neural Network, all gradients with respect to (θ, T, k) are available in closed-form and are fast to evaluate.

The critical part is, of course, the first step, as the second one merely corresponds to classical calibration against liquid options. For the first step, key issues are the choice of training data and the architecture of the neural network. Regarding the training data, the general idea is as follows:

1. Choose realistic "prior" distributions for both model parameters θ and option parameters ζ ($= (T, k)$ in the above notation). The point is that many theoretically possible parameters are very unlikely to ever occur in real markets, for both model and option parameters. Hence, it is wasteful to spend resources to learn the pricing map for, say, maturities in the range of hundreds of years. The simplest choice is to simply impose uniform distributions on truncated parameter ranges, but nothing prevents more "informed" possibilities, for instance taking into account historical distributions of estimated model parameter values or observed option parameter values.

2. Simulate model and option parameters according to the distribution chosen before and compute the corresponding option price or implied volatility, which serves as label for the respective parameter vector. The computation can be done for any available numerical method, for instance Monte Carlo simulation. As an aside, this mechanism can, of course, be used to produce training, testing and validation data in the sense of the machine learning literature.

Remark 3.3.1. Note that the above mentioned "informed" parameter distributions could also be encoded as weights into the loss function for the training of the neural network.

Remark 3.3.2. Instead of simulation of parameter values, we could also consider deterministic grids in the parameter space. In very high dimensional parameter spaces this probably becomes unfeasible due to the curse of dimensionality, but in the current context this approach may very well improve training of the neural network. We leave a comparison to future work.

Implicit & grid-based learning

We take this idea further and design an implicit form of the pricing map that is based on storing the implied volatility surface as an image given by a grid of "pixels". This image-based representation has a formative contribution in the performance of the network we present in Section 3.4. Let us denote by $\Delta := \{k_i, T_j\}_{i=1,\, j=1}^{n,\ m}$ a fixed grid of strikes and maturities, then we propose the following two step approach:

Step (i): Learn the map $\widetilde{F}(w, \theta) = \{\sigma^{\mathcal{M}(\theta)}(T_i, k_j)\}_{i=1,\, j=1}^{n,\ m}$ via neural network where the input is a parameter combination $\theta \in \Theta$ of the stochastic model $\mathcal{M}(\theta)$ and the output is a $n \times m$ grid on the implied volatility surface $\{\sigma^{\mathcal{M}(\theta)}(T_i, k_j)\}_{i=1,\, j=1}^{n,\ m}$ where $n, m \in \mathbb{N}$ are chosen appropriately (see Section 3.3.4) on a predefined fixed grid of maturities and strikes. $\widetilde{F}$ takes values in $\mathbb{R}^L$ where L = strikes $\times$ maturities $= nm$. The optimisation problem in the image-based implicit learning approach is:

$$\widehat{\omega} := \underset{w \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{N_{\text{Train}}^{\text{reduced}}} \sum_{j=1}^{L} \eta_j (\widetilde{F}(w, \theta_i)_j - \widetilde{\sigma}^{\mathcal{M}}(\theta_i, T_j, k_j))^2, \tag{3.10}$$

where $N_{\text{Train}} = N_{\text{Train}}^{\text{reduced}} \times L$ and $\eta_i \in \mathbb{R}_{>0}$ is a weight vector.

Step (ii): Solve the minimisation problem

$$\widehat{\theta} := \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{L} \beta_j (\widetilde{F}(\widehat{\omega}, \theta)_i - \sigma_{\text{BS}}^{\text{MKT}}(T_i, k_i))^2,$$

for some user specified weights $\beta_j \in \mathbb{R}_{>0}$. We note here that $\widetilde{F}(\widehat{\omega}, \theta)$ being a Neural Network, all gradients with respect to θ are available in closed-form and are fast to evaluate.

The data generation stage for the image-based approach works as in the point-wise approach, except that the option parameters $\zeta = (T, k)$ are, fixed and are no longer part of the learning algorithms – except implicitly in

the output/labels of the neural network. This is why they appear in the general objective function of pointwise learning (3.9) but no longer appear in the objective function (3.10) of the grid-based learning above. In practice, we choose a grid Δ of size 8×11.

Clearly, the neural network does depend on the grid Δ of option parameters ζ. Hence, we need to interpolate between gridpoints in order to be able to calibrate (in the calibration Step (ii)) also to such options, whose maturity and strike do not exactly lie on the grid Δ. This indirect dependence of the trained network on Δ is alluded to by the name "implicit learning".

The role of the objective function: Pointwise training versus implicit and grid-based training

Comparing the pointwise approach (characterised by the general objective function (3.9)) and the image-based approach (characterised by the objective function (3.10)), we find that both of them can be advantageous in certain situations. We highlight the connection between the two below, and elaborate on some of the respective advantages of each approach.

Equation (3.10) can be brought to the form of (3.9) equation by inserting (into (3.9)) the specification values $\theta = \theta'$, with

$$\theta'_1 = \theta_1, \ldots, \theta'_L = \theta_1, \theta'_{L+1} = \theta_2, \ldots,$$

and recalling that $L =$ strikes $\times$ maturities and $N_{\text{Train}} = N_{\text{Train}}^{\text{reduced}} \times L$. Hence, the pointwise approach is more general than the image-based one.

With this in mind we make the general note, many of the various advantages and disadvantages of both approaches can, in principle, be mitigated by careful choice of the data generation mechanism (of the training and validation datasets) and the loss function in the training.

- The biggest difference, between pointwise and image based implicit learning procedures is that image based implicit learning requires an outside (implicit) interpolation between the learned implied volatilities in order to compute the implied volatility of an option with an arbitrary strike or maturity, not aligned with the grid. At face value, this is of course an advantage of the pointwise (explicit) approach, where the interpolation is rather performed by the deep neural network. On the other hand, we note that the function $(T, k) \mapsto \sigma^{\mathcal{M}}(\theta; T, k)$ (for fixed model parameters θ) is usually a very well understood smooth function.
- Indeed, this very same structure induces a reduction of variance in the training data for the image-based approach as compared to the pointwise approach. Formally speaking, in the image based approach only the model parameters are sampled, while the strike and maturities of the underlying instruments are deterministic. As a side note, keep in mind that we should always compare the two approaches based on a fixed number N_{Train} of total training data.
- It is also easier to take into account the structure of real financial data into the data generation for the pointwise approach by adjusting the (random) sampling distribution on the surface accordingly. Clearly, not all options are equally important for the purpose of calibration, but we would like to concentrate on liquid options. It is easy to adjust the sampling distribution for strikes and maturities in the pointwise approach to take into account historical numbers of liquidity. In the grid-based approach, this can to some extent be taken into account by the choice of the weight vector $\eta_i \in \mathbb{R}_{>0}$ in (3.10), or more accurately taken into account by using non-uniform, non-tensorized, or bespoke quasirandom sampling grids, with higher density of points in regions with higher liquidity.
- The image-based approach may be seen as an efficient dimension-reduction technique as compared to the pointwise one. Of course, the price we pay is that we only learn the values of the implied volatilities on a fix grid Δ of option parameters. In this example, this price is, however, worth paying since the regularity of the volatility surface is well understood. This implies that we know very well the number and location of grid points required to get good fits globally in terms of the chosen interpolation.

In the particular calibration example presented in Section 3.4 below, the image-based approach performed somewhat better than the pointwise approach, which indicates that the variance and dimension reduction features may be more important than the other aspects in the above comparison.

Remark 3.3.3. In principle, the two-step approach is also amenable to other numerical interpolation methods. For instance, we could also use Chebyshev interpolation to approximate model implied volatilities such as [GKS19].

Remark 3.3.4. In line with Remark 3.3.3, we note that the image-based approach (in conjunction with the outside interpolation) is a hybrid between a pure DNN approximation such as the point-wise approach and a standard polynomial interpolation method, such as Chebyshev approximation, see [GKS19] for example. Of course, other, more specialized interpolation methods on the implied volatility surface are also possible, for instance using the SVI volatility parameterization, see for example [Itk15].

Reasons for the choice of grid-based implicit training as the bespoke approach

In Bayer, Horvath, Muguruza, Stemper and Tomas [Bay+19c] we provide a more detailed numerical analysis of the pointwise vs. grid-based implicit approach. However, we summarise here the reasons why the later shows superior numerical performance:

- The first advantage of implicit training is that it efficiently exploits the structure of the data. Updates in neighbouring volatility points σ_{n-1} and σ_n can be incorporated in the learning process. If the output is a full grid this effect is further enhanced. Updates of the network on each gridpoint also imply additional information for updates of the network on neighbouring gridpoints. One can say that we regard the implied volatility surface as an image with a given number of pixels.
- By evaluating the objective function on a larger set of (grid) points, injectivity of the mapping can be more easily guaranteed than in the pointwise training: Two distinct parameter combinations are less likely to yield the same value across a set of gridpoints, then if evaluated only on a single point.
- We do not limit ourselves to one specific grid on the implied volatility surface and we can easily add and evaluate additional maturities and strikes to the same set of Monte-Carlo generated paths. Note in particular that in this training design we can refine the grid on the implied volatility surface without increasing the number of training samples needed and without significantly increasing the computational time for training, as the portfolio of vanilla options on the same underlying grows with different strikes and maturities.
- In Appendix B we demonstrate that the out-of-sample generalisation is successfully achieved.

Some exotic payoffs

Our framework extends to all exotic products such as: Digital barriers, no-touch (or double no-touch) barrier, cliquets or autocallables.

We present some numerical experiments in Section 3.4.3, to demonstrate the pricing of digital barrier options. More precisely, in Section 3.4.3 we consider down-and-in such as down-and-out digital barrier options, the main building blocks of many Autocallable products. For a barrier level $B < S_0$ and maturity T the payoff is given by:

$$P^{Down-and-In}(B,T) = \mathbb{E}\left[\mathbf{1}_{\{\tau_B \leq T\}}\right] \tag{3.11}$$

$$P^{Down-and-Out}(B,T) = \mathbb{E}\left[\mathbf{1}_{\{\tau_B \geq T\}}\right] \tag{3.12}$$

where $\tau_B = \inf_t\{S_t = B\}$. In this setting, we may easily generate a grid for barrier levels and maturities $\Delta^{Barrier} := \{B_i, T_j\}_{i=1,\ j=1}^{n,\ m}$ that we can fit in the objective function specified in (3.10).

3.3.3 Some relevant properties of deep neural networks as functional approximators

Deep feed forward[1] neural networks are the most basic deep neural networks, originally designed to approximate some function F^*, which is not available in closed form but only through sample pairs of given input data x and output data $y = F^*(x)$. In a nutshell, a feed forward network defines a mapping $y = F(x, w)$ and the training determines (calibrates) the optimal values of network parameters $\widehat{w}$ that result in the best function approximation[2] $F^*(\cdot) \approx F(\cdot, \widehat{w})$ of the unknown function $F^*(\cdot)$ for the given pairs of input and output data (x, y), cf. [GBC16, Chapter 6].
To formalise this, we introduce some notation and recall some basic definitions and principles of function approximation via (feedforward) neural networks:

Definition 3.3.5 (Neural network)**.** Let $L \in \mathbb{N}$ and the tuple $(N_1, N_2 \cdots, N_L) \in \mathbb{N}^L$ denote the number of layers (depth) and the number of nodes (neurons) on each layer respectively. Furthermore, we introduce the affine functions

$$\begin{aligned} w^l : \mathbb{R}^{N_l} &\longrightarrow \mathbb{R}^{N_{l+1}} \text{ for } 1 \leq l \leq L-1 \\ x &\mapsto A^{l+1}x + b^{l+1} \end{aligned} \tag{3.13}$$

acting between layers for some $A^{l+1} \in \mathbb{R}^{N_{l+1} \times N_l}$. The vector $b^{l+1} \in \mathbb{R}^{N_{l+1}}$ denotes the *bias term* and each entry $A^{l+1}_{(i,j)}$ denotes the *weight* connecting node $i \in N_l$ of layer l with node $j \in N_{l+1}$ of layer $l+1$. For the the collection of affine functions of the form (3.13) on each layer we fix the notation $w = (w^1, \ldots, w^L)$. We call the tuple w the *network weights* for any such collection of affine functions. Then a Neural Network $F(w, \cdot) : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ is defined as the composition:

$$F := F_L \circ \cdots \circ F_1 \tag{3.14}$$

where each component is of the form $F_l := \sigma_l \circ W^l$. The function $\sigma_l : \mathbb{R} \to \mathbb{R}$ is referred to as the *activation function*. It is typically nonlinear and applied component wise on the outputs of the affine function W^l. The first and last layers, F_1 and F_L, are the *input* and *output* layers. Layers in between, $F_2 \cdots F_{L-1}$, are called *hidden layers*.

The following central result of Hornik justifies the use of neural networks as approximators for multivariate functions and their derivatives.

Theorem 3.3.6 (Universal approximation theorem (Hornik, Stinchcombe and White [HSW89]))**.** *Let $\mathcal{NN}^{\sigma}_{d_0,d_1}$ be the set of neural networks with activation function $\sigma : \mathbb{R} \mapsto \mathbb{R}$, input dimension $d_0 \in \mathbb{N}$ and output dimension $d_1 \in \mathbb{N}$. Then, if σ is continuous and non-constant, $\mathcal{NN}^{\sigma}_{d_0,1}$ is dense in $L^p(\mu)$ for all finite measures μ.*

There is a rapidly growing literature on approximation results with neural networks, see [Hor91; HSW89; Mha93; SCC18] and the references therein. Among these we would like to single out one particular result:

Theorem 3.3.7 (Universal approximation theorem for derivatives (Hornik, Stinchcombe and White [HSW90]))**.** *Let $F^* \in \mathcal{C}^n$ and $F : \mathbb{R}^{d_0} \to \mathbb{R}$ and $\mathcal{NN}^{\sigma}_{d_0,1}$ be the set of single-layer neural networks with activation function $\sigma : \mathbb{R} \mapsto \mathbb{R}$, input dimension $d_0 \in \mathbb{N}$ and output dimension 1. Then, if the (non-constant) activation function is $\sigma \in \mathcal{C}^n(\mathbb{R})$, then $\mathcal{NN}^{\sigma}_{d_0,1}$ arbitrarily approximates f and all its derivatives up to order n.*

Remark 3.3.8. Theorem 3.3.7 highlights that the smoothness properties of the activation function are of significant importance in the approximation of derivatives of the target function F^*. In particular, to guarantee the convergence of l-th order derivatives of the target function, we choose an activation function $\sigma \in C^l(\mathrm{R})$. Note that the ReLu activation function , $\sigma_{ReLu}(x) = (x)^+$ is not in $\mathcal{C}^l(\mathrm{R})$ for any $l > 0$, while $\sigma_{Elu}(x) = \alpha(e^x - 1)$ is smooth.

[1] The network is called feed forward if there are no feedback connections in which outputs of the model are fed back into itself.
[2] In our case y is a 8×11-point grid on the implied volatility surface and x are model parameters $\theta \in \Theta$, for details see Section 3.3.

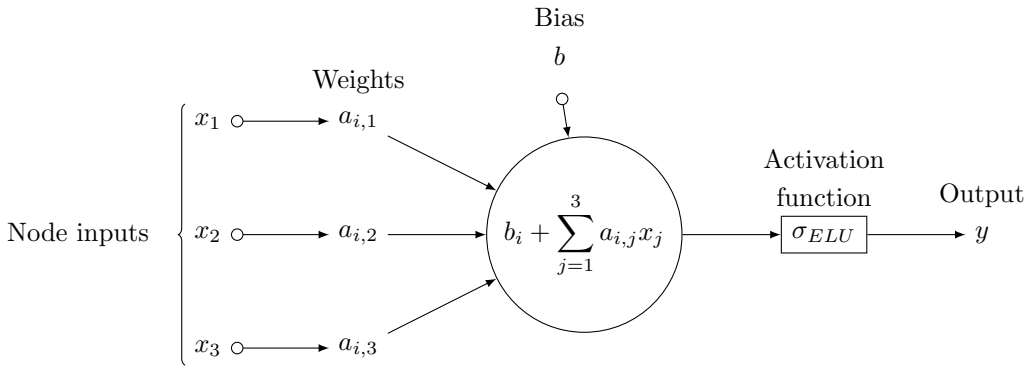


Figure 3.1: In detail neuron behaviour

The following Theorem provides theoretical bounds for the above rule of thumb and establishes a connection between the number of nodes in a network and the number of training samples needed to train it.

Theorem 3.3.9 (Estimation bounds for Neural Networks (Barron [Bar94]))**.** *Let $\mathcal{NN}^{\sigma}_{d_0,d_1}$ be the set of single-layer neural networks with Sigmoid activation function $\sigma(x) = \frac{e^x}{e^x+1}$, input dimension $d_0 \in \mathbb{N}$ and output dimension $d_1 \in \mathbb{N}$. Then:*

$$\mathbb{E}\|F^* - \hat{F}\|_2^2 \leq \mathcal{O}\left(\frac{C_f^2}{n}\right) + \mathcal{O}\left(\frac{nd_0}{N}\log N\right)$$

where n is the number of nodes, N is the training set size and C_{F^*} is the first absolut moment of the Fourier magnitude distribution of F^*.

Remark 3.3.10. Barron's [Bar94] insightful result gives a rather explicit decomposition of the error in terms of bias (model complexity) and variance:

- $\mathcal{O}\left(\frac{C_{F^*}^2}{n}\right)$ represents the model complexity, i.e. the larger n (number of nodes) the smaller the error
- $\mathcal{O}\left(\frac{nd_0}{N}\log N\right)$ represents the variance, i.e. a large n must be compensated with a large training set N in order to avoid overfitting.

Finally, we motivate the use of multi layer networks and the choice of network depth. Even though a single layer might theoretically suffice to arbitrarily approximate any continuous function,in practice the use of multiple layers dramatically improves the approximation capacities of network. We informally recall the following Theorem due to Eldan and Shamir [ES15] and refer the reader to the original paper for details.

Theorem 3.3.11 (Power of depth of Neural Networks (Eldan and Shamir [ES15]))**.** *There exists a simple (approximately radial) function on $\mathbb{R}^d$, expressible by a small 3-layer feedforward neural networks, which cannot be approximated by any 2-layer network, to more than a certain constant accuracy, unless its width is exponential in the dimension.*

Remark 3.3.12. In spite of the specific framework by Eldan and Shamir [ES15] being restrictive, it provides a theoretical justification to the power of "deep" neural networks (multiple layers) against "shallower" networks (i.e. few layers) as in [McG18] with a larger number of neurons. On the other hand, multiple findings indicate [Bay+19c; IS15] that adding hidden layers beyond 4 hidden layers does not significantly improve network performance.

3.3.4 Network architecture and training

Network architecture of the implied volatility map approximation

Here we motivate our choice of network architecture for the following numerical experiments which were inspired by the analysis in the previous sections. Our network architecture is summarised in the graph 3.2 below.

1. A fully connected feed forward neural network with 4 hidden layers (due to Theorem 3.3.11) and 30 nodes on each layers (see Figure 3.2 for a detailed representation)

2. Input dimension = n, number of model parameters

3. Output dimension = 11 strikes× 8 maturities for this experiment, but this choice of grid can be enriched or modified.

4. The four inner layers have 30 nodes each, which adding the corresponding biases results on a number

$$(n+1) \times 30 + 4 \times (1+30) \times 30 + (30+1) \times 88 = 30n + 6478$$

of network parameters to calibrate (see Section 3.3.3 for details).

5. Motivated by Theorem 3.3.7 we choose the Elu $\sigma_{Elu}(x) = \alpha(e^x - 1)$ activation function for the network and linear activation $\sigma_{Linear} = (x)$ in the output layer.

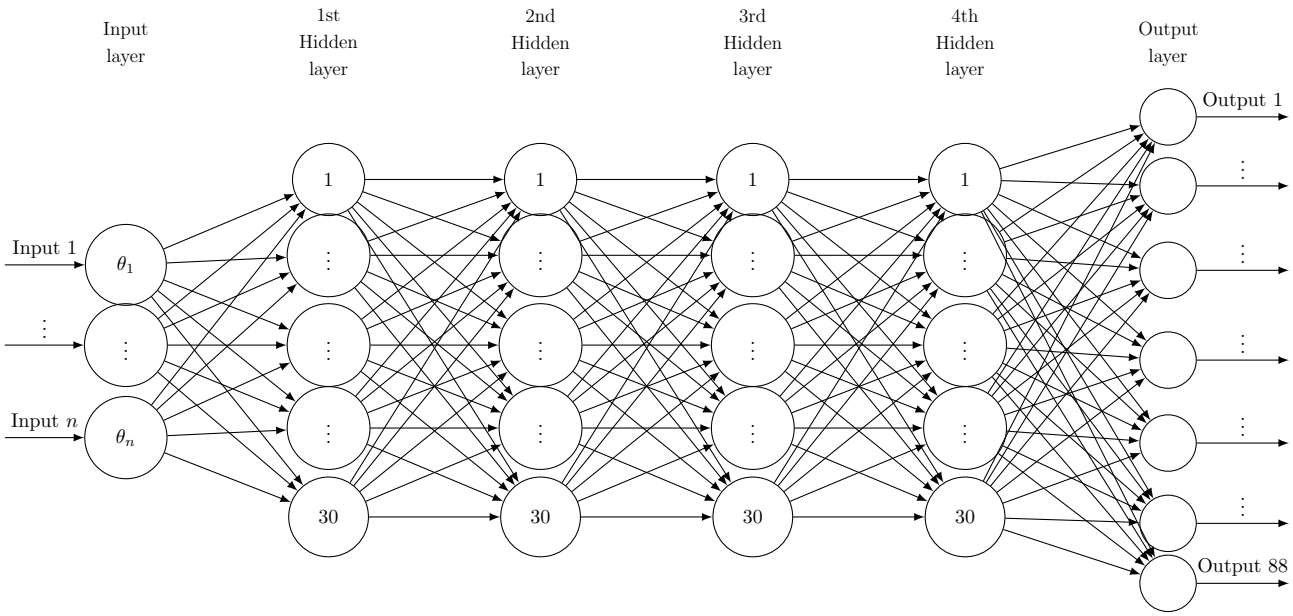


Figure 3.2: Our neural network architecture with 4 hidden layers and 30 neurons on each hidden layer, with the model parameters of the respective model on the input layer and with the 8×11 implied volatility grid on the output layer.

Training of the approximation network

We follow the common features of optimization techniques and choose mini-batches, as described in Goodfellow, Bengio and Courville [GBC16]. Typical batch size values range from around 10 to 100. In our case we started with small batch sizes and increased the batch size until training performance consistently reached a plateau. Finally, we chose batch sizes of 32, as performance is similar for batch sizes above this level, and larger batch sizes increase computation time by computing a larger number of gradients at a time.

In our training design, we use a number of regularisation techniques to speed up convergence of the training, to avoid overfitting and improve the network performance.

1) Early stopping: We choose the number of epochs as 200 and stop updating network parameters if the error has not improved in the test set for 25 steps.

2) Normalisation of model parameters: Usually, model parameters are restricted to a given domain i.e. $\theta \in [\theta_{min}, \theta_{max}]$. Then, we perform the following normalisation transform:

$$\frac{2\theta - (\theta_{max} + \theta_{min})}{\theta_{max} - \theta_{min}} \in [-1, 1].$$

3) Normalisation of implied volatilities: The normalisation of implied volatilities is a more delicate matter, since $\sigma_{BS}(T, k, \theta^{train}) \in [0, \infty)$, for each T and k. Therefore, we choose to normalise the surface subtracting the sample empirical mean and dividing by the sample standard deviation.

3.4 Numerical experiments

In our numerical experiments we demonstrate that the accuracy of the approximation network indeed remains within the accuracy of the Monte Carlo error bounds and proclaimed in the introductory sections' objectives. For this we first compute the benchmark Monte Carlo errors in Figures 3.3-3.4 and compare this with the neural network approximation errors in Figures 3.5 and 3.6. For this separation into steps (i) and (ii) to be computationally meaningful, the neural network approximation has to be a reasonably accurate approximation of the true pricing functionals and each functional evaluation (i.e. evaluation an option price for a given price and maturity) should have a considerable speed-up in comparison to the original numerical method. In this section we demonstrate that our network achieves both of these goals.

3.4.1 Numerical accuracy and speed of the price approximation for vanillas

One crucial difference that sets apart this work from direct neural network approaches, as pioneered by Hernandez [Her17], is the separation of (i) the implied volatility approximation function, mapping from parameters of the stochastic volatility model to the implied volatility surface–thereby bypassing the need for expensive Monte-Carlo simulations—and (ii) the calibration procedure, which (after this separation) becomes a simple deterministic optimisation problem. As outlined in Section 3.3.2 our aim for the Step (i) in the two-step training approach is to achieve a considerable speedup per functional evaluation of option prices while maintaining the numerical accuracy of the original pricer:

1. Approximation accuracy: here we compare the error of the approximation network to the error of Monte Carlo evaluations (defined as the 95% sample confidence interval[3]). We compute Monte Carlo prices with $60,000$ paths as reference at the nodes where we compute the implied volatility grid using Algorithm 2.4.5. In Figures 3.3 and 3.4 the approximation accuracy of the Monte Carlo method for the full implied volatility surface is computed using pointwise relative error with respect to the 95% Monte Carlo confidence interval. Figures 3.5 and 3.6 demonstrate that the same approximation accuracy for the neural network is achieved as for the Monte Carlo approximation (i.e. within a few basis points). For reference, the spread on options is around 0.2% in implied volatility terms for the most liquid and those below a year. This translates into 1% relative error for a implied volatility of 20%.

2. Approximation speed: Table 3.2 shows the CPU computation time per functional evaluation of a full surface under two different models; rBergomi 3.5 and 1 Factor Bergomi 3.7 (for a reminder see Section 3.4.1 for details).

[3]We note that confidence intervals are symmetric in price space, however this is not the case in implied volatility space as the corresponding mapping is nonlinear. Therefore, we consider the error as the maximum length of the 95% confidence interval in implied volatility space as a conservative approach.

3. For clarity we emphasize that the cost of generating a single training sample is 0.3 seconds in the case of 1F Bergomi and 0.5 seconds in the case of rBergomi. Therefore, the complete training sample generation takes 400 minutes and 670 minutes respectively.

4. In light of Chapter 2, the choice of n plays a big role in the convergence of the discrete-time process to the continuous-time one. In this case however, $n = 252$ is fixed by the user (and could take any value), thus we focus on the discrete-time process (which we aim to price via neural networks), not the continuous-time counterpart. Also, if the size of n is increased the time it takes to generate training samples would increase linearly with respect to the $n = 252$ baseline in light of Figure 2.12.

	MC Pricing 1F Bergomi Full Surface	MC Pricing rBergomi Full Surface	NN Pricing Full Surface	NN Gradient Full Surface	Speed up NN vs. MC
Piecewise constant forward variance	$300,000$ μs	$500,000$ μs	30.9 μs	113 μs	$9,000 - 16,000$

Table 3.2: Computational time of pricing map (entire implied volatility surface) and gradients via Neural Network approximation and Monte Carlo (MC). If the forward variance curve is a constant value, then the speed-up is even more pronounced

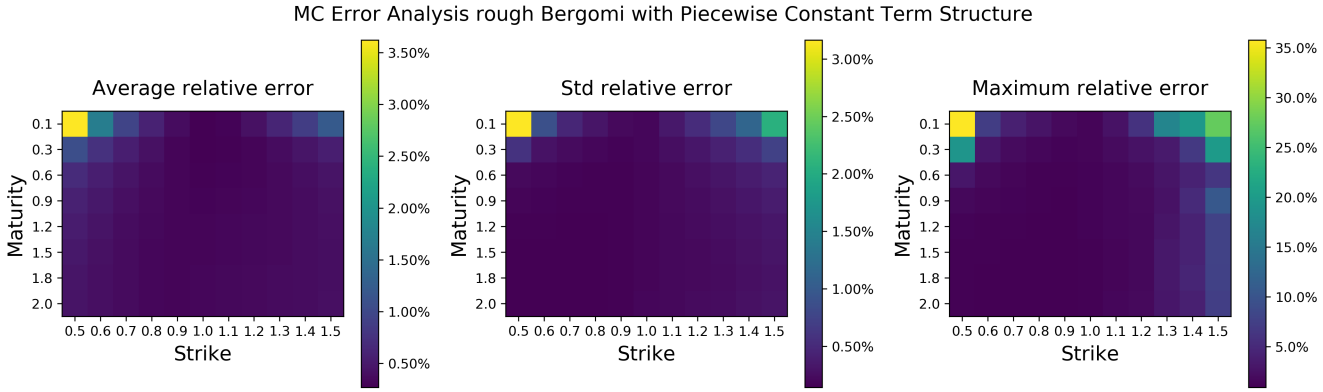


Figure 3.3: As benchmark we recall average relative errors of Monte Carlo prices computed across $80,000$ random parameter combinations of the Rough Bergomi model. Relative errors are given in terms of Average-Standard Deviation-Maximum (Left-Middle-Right) on implied volatility surfaces in the Rough Bergomi model, computed using 95% confidence intervals.

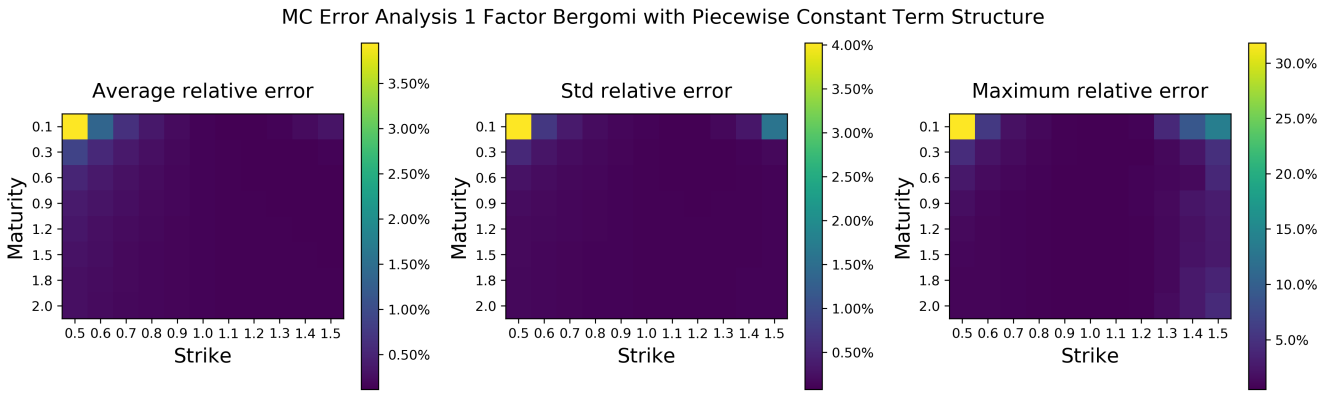


Figure 3.4: As benchmark we recall average relative errors of Monte Carlo prices computed across $80,000$ random parameter combinations of the 1 Factor Bergomi model. Relative errors are given in terms of Average-Standard Deviation-Maximum (Left-Middle-Right) on implied volatility surfaces in the 1 Factor Bergomi model, computed using 95% confidence intervals.

Neural network price approximation in (rough) Bergomi models with piecewise constant forward variance curve

We consider a piecewise constant forward variance curve $\xi_0(t) = \sum_{i-1}^{n} \xi_i \mathbf{1}_{\{t_{i-1}<t<t_i\}}$ where $t_0 = 0 < t_1 < ... < t_n$ and $\{t_i\}_i = 1, .., n$ are the option maturity dates ($n = 8$ in our case). This is the modelling approach suggested by Bergomi [Ber16]. We will consider again the rough Bergomi 3.5 and 1 Factor Bergomi models 3.7

- Normalized parameters as input and normalised implied volatilities as output
- 4 hidden layers with 30 neurons and Elu activation function
- Output layer with Linear activation function
- Total number of parameters: 6808
- Train Set: $68,000$ and Test Set: $12,000$
- Rough Bergomi sample:$(\xi_0, \nu, \rho, H) \in \mathcal{U}[0.01, 0.16]^8 \times \mathcal{U}[0.5, 4.0] \times \mathcal{U}[-0.95, -0.1] \times \mathcal{U}[0.025, 0.5]$
- 1 Factor Bergomi sample:$(\xi_0, \nu, \rho, \beta) \in \mathcal{U}[0.01, 0.16]^8 \times \mathcal{U}[0.5, 4.0] \times \mathcal{U}[-0.95, -0.1] \times \mathcal{U}[0, 10]$
- strikes=$\{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5\}$
- maturities=$\{0.1, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.0\}$
- Training data samples of Input-Output pares are computed using Algorithm 2.4.5 with $60,000$ sample paths and the spot martingale condition i.e. $\mathbb{E}[S_t] = S_0,\ t \geq 0$ as control variate.

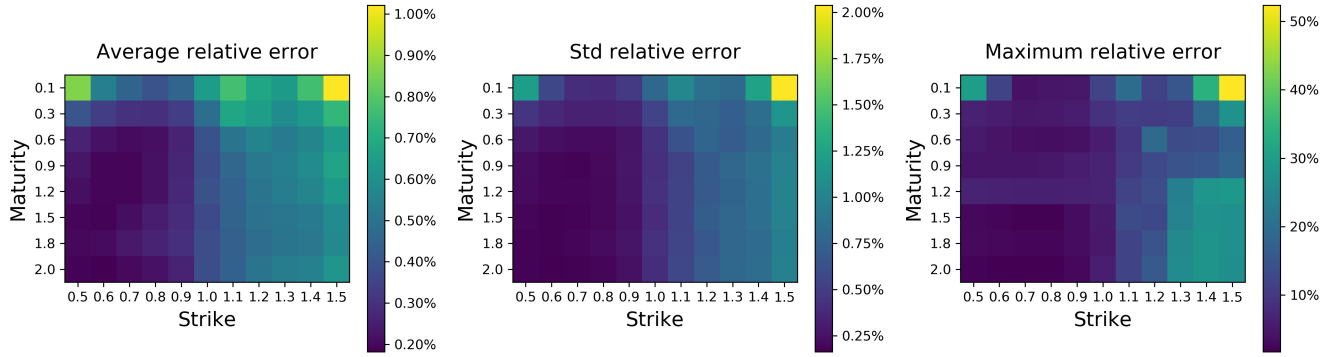


Figure 3.5: We compare surface relative errors of the neural network approximator against the Monte Carlo benchmark across all training data (68, 000 random parameter combinations)in the rough Bergomi model. Relative errors are given in terms of Average-Standard Deviation-Maximum (Left-Middle-Right).

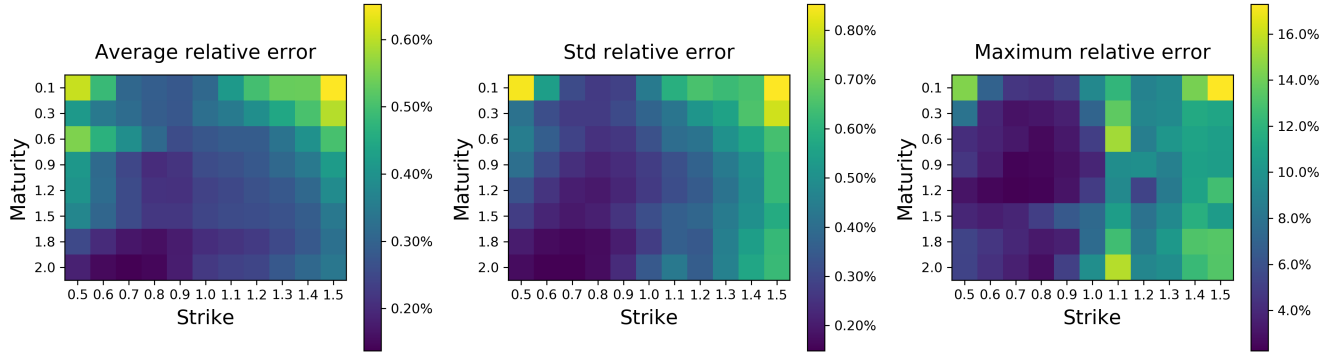


Figure 3.6: We compare surface relative errors of the neural network approximator against the Monte Carlo benchmark across all training data (68, 000 random parameter combinations)in the 1 Factor Bergomi model. Relative errors are given in terms of Average-Standard Deviation-Maximum (Left-Middle-Right).

Figures 3.5 and 3.6 show that the average (across all parameter combinations) relative error between neural network and Monte Carlo approximations is far less than 0.5% consistently (left image in Figures 3.5 and 3.6) with a standard deviation of less than 1% (middle image in Figures 3.5 and 3.6). The maximum relative error goes as far as 25%. We conclude that the methodology generalises adequately to the case of non-constant forward variances, by showing the same error behaviour.

3.4.2 Calibration speed and accuracy for implied volatility surfaces

Figure 3.7 reports average calibration times on test set for different parameter combinations on each of the models analysed. We conclude that gradient-based optimizers outperform (in terms of speed) gradient-free ones. Moreover, in Figure 3.7 one observes that computational times in gradient-free methods are heavily affected by the dimension of the parameter space, i.e. flat forward variances are much quicker to calibrate than piecewise constant ones. We find that Lavenberg-Marquardt is the most balanced optimizer in terms of speed/convergence and we choose to perform further experiments with this optimizer. The reader is encouraged to keep in mind that a wide range of optimizers is available for the calibration and the optimal selection of one is left for future research.

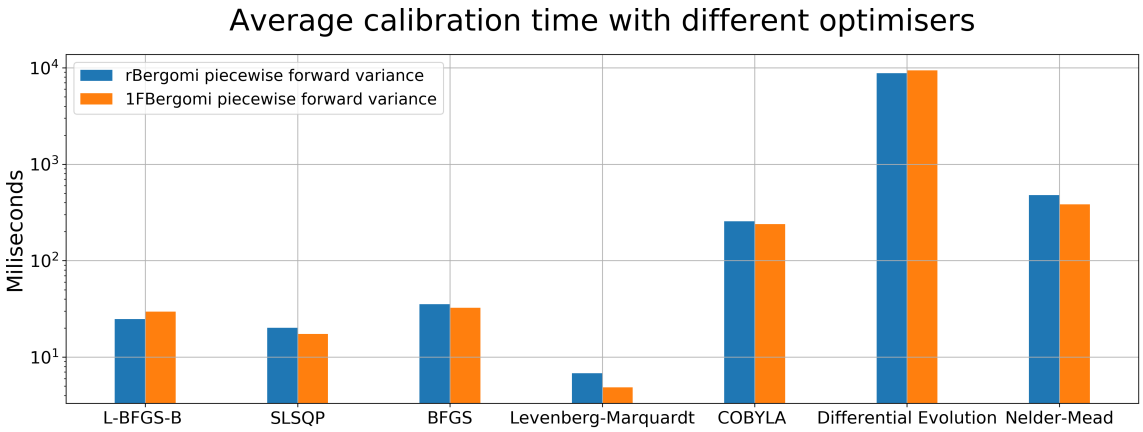


Figure 3.7: Average calibrations times for all models using a range of optimizers.

In order to assess the accuracy, we report the calibrated model parameters $\widehat{\theta}$ compared to the synthetically generated data with the set of parameters $\overline{\theta}$ that was chosen for the generation of our synthetic data. We measure the accuracy of the calibration via parameter relative error i.e.

$$E_R(\widehat{\theta}) = \frac{|\widehat{\theta} - \overline{\theta}|}{|\overline{\theta}|}$$

as well as the root mean square error (RMSE) with respect to the original surface i.e.

$$\text{RMSE}(\widehat{\theta}) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}(\widetilde{F}(\widehat{\theta})_{ij} - \sigma_{BS}^{MKT}(T_i, k_j))^2}.$$

Therefore, on one hand a measure of good calibration is a small RMSE. On the other hand, a measure of parameter sensitivity on a given model is the combined result of RMSE and parameter relative error.

A calibration experiment with simulated data in (rough) Bergomi models with piecewise constant forward variances

We consider the rough Bergomi model (3.5) and the Bergomi model (3.7) with a piecewise constant term-structure of forward variances. Figures 3.8 and 3.9 show that the 99% quantile of the RMSE is below 1% and shows that the Neural Network approach generalises properly to the piecewise constant forward variance. Again, we find that the largest relative errors per parameter are concentrated around 0, consequence of using the relative error as measure. This suggests a successful generalisation to general forward variances, which to our knowledge has not been addressed before by means of neural networks or machine learning techniques.

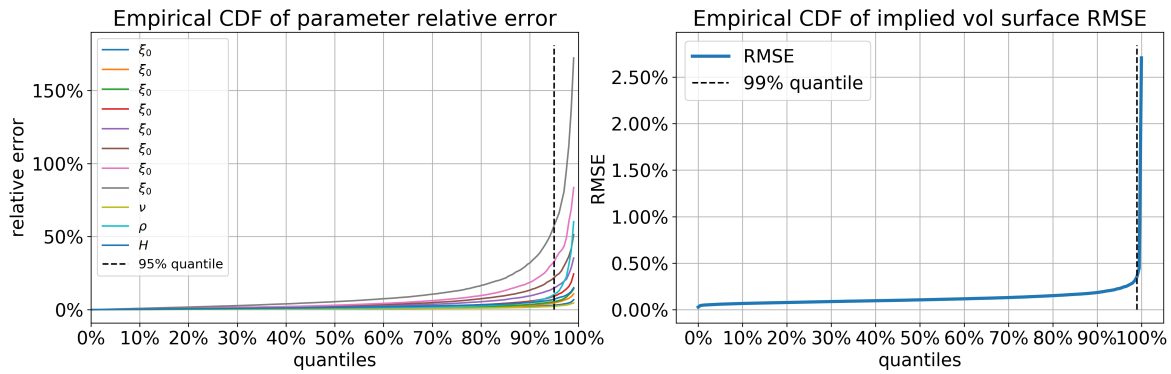


Figure 3.8: Cumulative Distribution Function (CDF) of Rough Bergomi parameter relative errors (left) and RMSE (right) after Levengerg-Marquardt calibration across test set random parameter combinations.

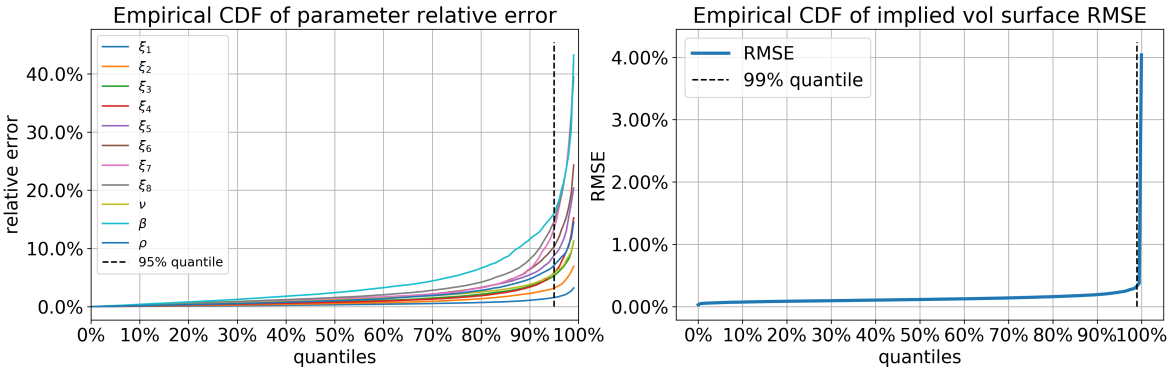


Figure 3.9: Cumulative Distribution Function (CDF) of 1 Factor Bergomi parameter relative errors (left) and RMSE (right) after Levengerg-Marquardt calibration across test set random parameter combinations.

Calibration in the rough Bergomi model with historical data

As previously mentioned, the natural use of neural network approximators is the model calibration to historical data. We discussed that as along as the approximation is accurate, the calibration task should be performed within the given tolerance. Furthermore, one should expect such tolerance to be aligned with the neural network accuracy obtained in both training and test sets.

In this section we will perform a historical calibration using the neural network approximation and compare it with that of the brute force monte carlo calibration. Precisely we seek to solve the following optimisation problem for the rough Bergomi model

$$\hat{\theta^{rBergomi}} := \underset{\theta^{rBergomi} \in \Theta^{rBergomi}}{\operatorname{argmin}} \sum_{i=1}^{5} \sum_{j=1}^{9} (\widetilde{F}(\theta)_{ij} - \sigma_{BS}^{MKT}(T_i, k_j))^2.$$

where $\theta^{rBergomi} = (\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \nu, \rho, H)$ and $\Theta^{rBergomi} = [0.01, 0.25]^5 \times [0.5, 4] \times [-1, 0] \times [0.025, 0.5]$. As for the time grid we choose

$$(T_1, T_2, T_3, T_4, T_5) := \frac{1}{12} \times (1, 3, 6, 9, 12)$$

and for the strike grid

$$k_i := 0.85 + (i-1) \times 0.05 \quad \text{for } i = 1, ..., 9.$$

Remark 3.4.1. We emphasize that $\Theta^{rBergomi} = [0.01, 0.25]^5 \times [0.5, 4] \times [-1, 0] \times [0.025, 0.5]$, i.e. the parameter

space in constrained to a bounded set that coincides with the training set. This prevents the optimisation algorithm from using the neural network in a extrapolation domain (commonly referred as out of sample environment) and makes the procedure consistent.

We consider SPX market smiles between 01/01/2010 and 18/03/2019 on the pre-specified time and strike grid. Figure 3.10 shows the historical evolution of rough Bergomi parameters calibrated to SPX using the neural network price. In particular we note that $H < \frac{1}{2}$ as previously discussed in many academic papers [ALV07; BFG16; Bay+19b; Bay+19a; BLP17; ER19; Fuk11b; GJR18; HJL19; JPS18], moreover we may confirm that under $\mathbb{Q}$, $H \in [0.1, 0.15]$ as found in Gatheral, Jaisson and Rosenbaum [GJR18] under $\mathbb{P}$. Figure 3.11, benchmarks the NN optimal fit using Levenberg-Marquardt and Differential Evolution against a brute force MC calibration via Levenberg-Marquardt. Again, we find that the discrepancy between both is below 0.2% most of the time and conclude that the Differential Evolution algorithm does outperform the Levenberg-Marquardt. This in turn, suggests that the neural network might not be precise enough on first order derivatives. This observation, is left as an open question for further research. Perhaps surprisingly, we sometimes obtain a better fit using the neural network than the MC pricing itself. This could be caused by the fact that gradients in the neural network are exact, whereas when using MC brute force calibration we resort to finite differences to approximate gradients.

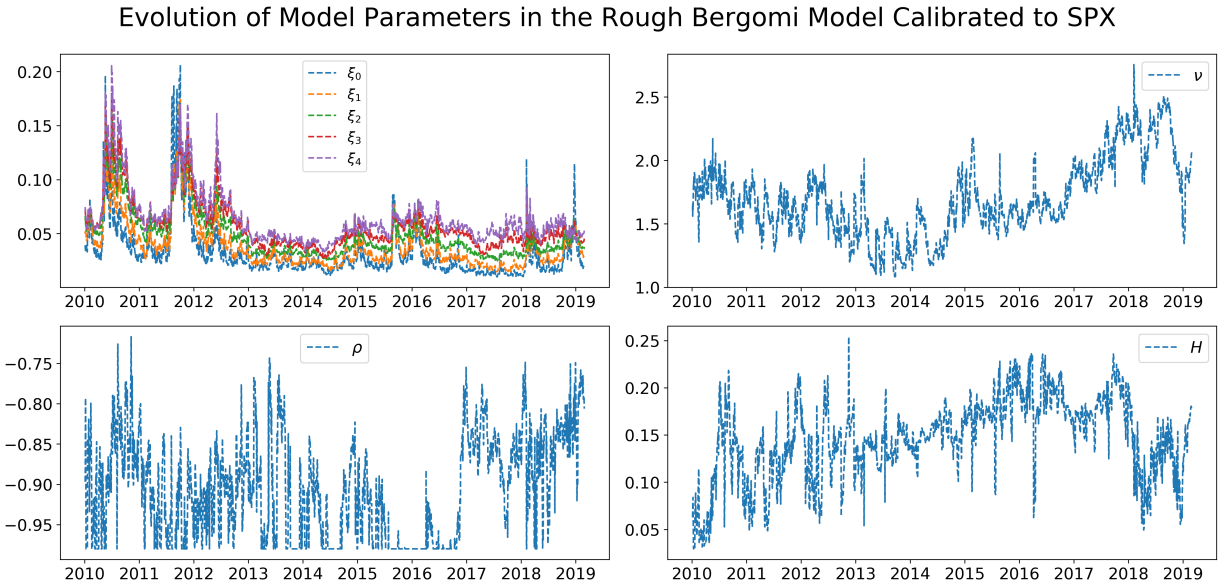


Figure 3.10: Historical Evolution of parameters in the rough Bergomi model with a piecewise constant forward variance term structure calibrated on SPX

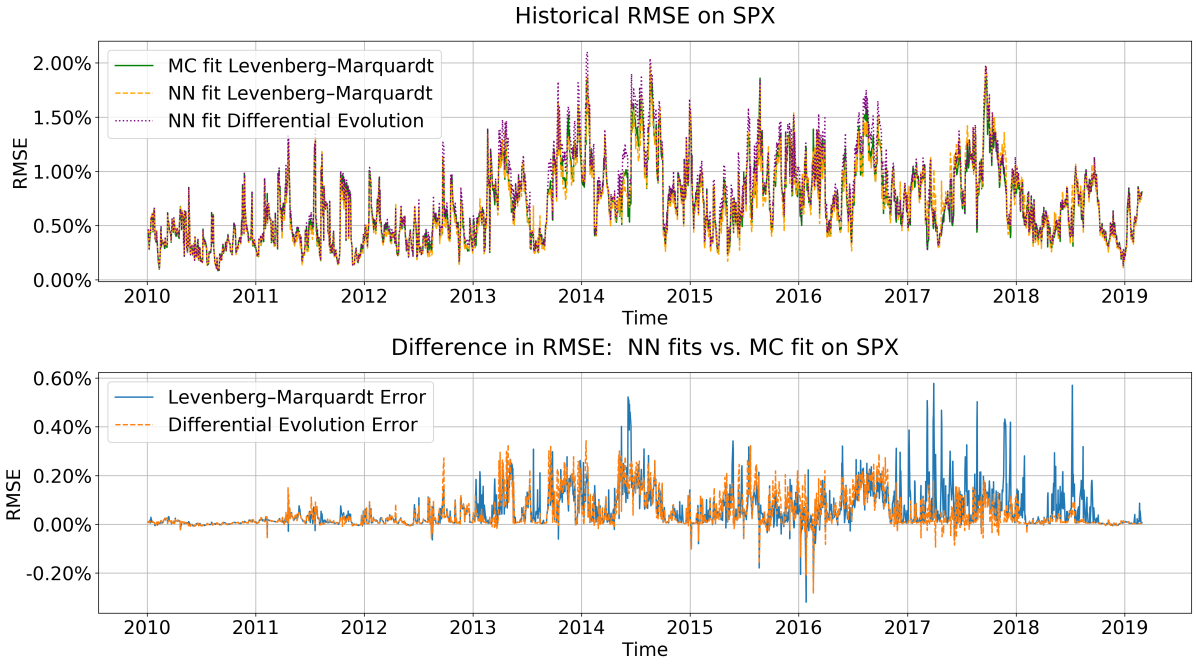


Figure 3.11: The image above compares historical RMSE obtained by the neural network best fit via Levenberg-Marquardt (dashed orange line) and Differential Evolution (dotted purple line) against the brute force MC calibration (green line) via Levenberg-Marquardt. Picture below shows the difference against MC brute force calibration.

3.4.3 Numerical experiments with barrier options in the rough Bergomi model

In this section we show that our methodology can be easily extended to exotic options. To do so we test our image-based approach on digital barrier options. We follow the same architecture and experimental design described in Section 3.4.1 for the rough Bergomi model. As described in Section 3.3.2 we adapt the objective function to the payoffs given in (3.11) and (3.12) and replace the strike grid by a barrier level grid. Figure 3.12 confirm the accuracy of the neural network approximation with average absolute errors of less than 10bps with standard deviation of 10bps.

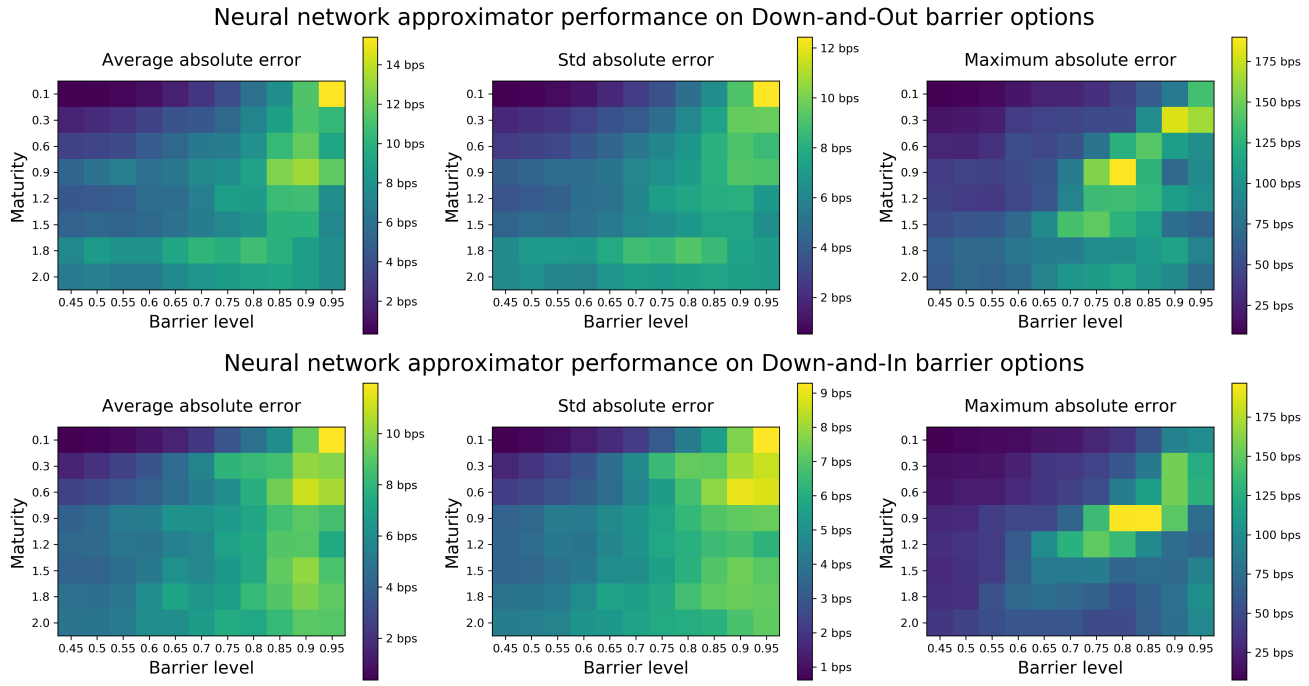


Figure 3.12: Picute above: Down-and-Out neural network absolute error analysis on test set. Picture below: Down-and-In neural network absolute error analysis on test set

3.5 Summary

Throughout this Chapter we have explored the use of deep learning methods to approximate model-based derivatives pricing functions. In particular, we have shown how to apply neural networks to learn the implied volatility generated by a (rough) stochastic volatility model and also explored some exotic payoffs. This powerfull framework, allowed us to perform calibration tasks within milliseconds whilst remaining within admissible error bounds in an out-of-sample environment applied to SPX. The novelty of our approach lies into mapping the pricing function into a grid (rather than a single point), and exploiting the structure of such grid to improve the performance of neural networks as functional approximators. This in turn, allowed us to make the use of (virtually) any stochastic volatility model as simple (and fast) as the Black-Scholes formula.

4

Not so particular about calibration:
Smile problem Resolved

4.1 Introduction

The calibration of local-stochastic volatility (LSV) models is a fundamental problem in financial modelling. In the case of low dimensional Markovian models, efficient PDE methods are already available to compute the leverage function (see Lipton [Lip02]). For more complex Markovian models, (e.g. multi factor or hybrid models) PDE methods fall into the so-called curse of dimensionality, making their use more complex and slow. The most transparent and effective solution to this LSV calibration problem in high dimensional models has been the particle method developed by Guyon and Henry-Labordère [GH12]. It is worth noting, however, that this method requires the prespecification of a kernel function and bandwidth parameter, which can be unnerving in an automatised production environment.

In this Chapter we present an algorithm that follows the principle of the particle method, without relying on non-parametric methods to obtain the leverage function. Perhaps surprisingly, we are able to obtain a closed-form expressions (under standard discrete time structure assumptions on the leverage function) that allow us to calibrate the leverage function, and apply to all Stochastic Volatility (SV) models. We further extend the results to an Euler scheme setting to obtain a calibration algorithm.

4.2 LSV framework and mathematical setting

An LSV model under the risk-neutral measure $\mathbb{Q}$ is given by

$$\frac{\mathrm{d}S_t}{S_t} = r_t \mathrm{d}t + \lambda(S_t, t)\sqrt{V_t}\mathrm{d}W_t \tag{4.1}$$

where $\lambda : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ is the so-called leverage function. W here, is a unidimensional Brownian motion and the processes V and r are assumed to be adapted to the natural filtration generated by a d-dimensional Brownian motion Z, where $d \in \mathbb{N}$. In addition, we assume the tuple (W_t, Z_t) to have the following correlation structure:

$$\Sigma = \begin{pmatrix} 1 & \rho_{WZ}^T \\ \rho_{WZ} & \Sigma_Z \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}, \quad \Sigma_Z \in \mathbb{R}^{d\times d}, \quad \rho_{WZ} \in \mathbb{R}^{d\times 1}.$$

The processes are defined on a given probability space $(\Omega, (\mathcal{G}_t)_{t\geq 0}, \mathbb{Q})$ with $\mathcal{G}_t$ being the natural filtration generated by the aforementioned Brownians. Let us further denote by $\mathcal{F}_t^W$ and $\mathcal{F}_t^Z$ the filtrations generated by W and Z respectively, such that $\mathcal{G}_t = \mathcal{F}_t^W \vee \mathcal{F}_t^Z$ holds. Throughout this chapter, we shall work with the discrete-time counterpart of (4.1) given by,

$$\log(S_{t_i}) = \log(S_{t_{i-1}}) + \left(-\frac{1}{2}V_{t_{i-1}} + r_{t_{i-1}}\right)\Delta t_i + \lambda(S_{t_{i-1}}, t_{i-1})\sqrt{V_{t_{i-1}}}\Delta W_{t_i}, \tag{4.2}$$

where $\Delta t_i = t_i - t_{i-1}$ and $\Delta W_{t_i} = W_{t_i} - W_{t_{i-1}}$, for $i = 1, ..., n$, a suitable time partition $\{0 = t_0, t_1, ..., t_n = T\}$ and a fixed time-horizon $T > 0$.

Remark 4.2.1. Dividends are neglected for sake of simplicity, but it is straightforward to consider them as in Guyon and Henry-Labordère [GH12].

4.3 The calibration of the leverage function in a nutshell

The probabilistic condition in (4.1) for the leverage function to be calibrated to market smiles is given by (see Balland [Bal05])

$$\lambda^2(K,t)\frac{\mathbb{E}^{\mathbb{Q}}[D(t)V_t|S_t = K]}{\mathbb{E}^{\mathbb{Q}}[D(t)|S_t = K]} = \sigma_{Dup}(K,t)^2 - \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t)\left(r_t - \bar{r}_t\right)\mathbb{1}_{\{S_t>K\}}\right]}{\frac{1}{2}K\frac{\partial^2 C}{dK^2}}, \tag{4.3}$$

where $D(t) = e^{-\int_0^t r_s ds}$, $ZCB(t) = \mathbb{E}^{\mathbb{Q}}[D(t)]$, $\bar{r}_t = -\frac{\partial \log(ZCB(t))}{\partial t}$. Note that $\frac{\mathbb{E}^{\mathbb{Q}}[D(t)r_t]}{\mathbb{E}^{\mathbb{Q}}[D(t)]} = \bar{r}_t$. We define $\sigma_{Dup}(K,t)$ to be the local volatility (Dupire [Dup94]) associated with the observed market i.e.

$$\sigma_{Dup}(K,T)^2 = \frac{\frac{\partial C}{\partial T} + K\frac{\partial C}{dK}\bar{r}_T}{\frac{1}{2}K^2\frac{\partial^2 C}{\partial K^2}}.$$

The SDE presented in (4.1) paired with $\lambda(\cdot,\cdot)$ given by (4.3), transforms into a McKean SDE. The existence and uniqueness of such SDE is a very intricate mathematical problem (see [JZ20]) and falls outside the scope of this work. All numerical algorithms known to solve the calibration task rely on the following standard assumption:

Assumption 4.3.1. The leverage function has the following structure:

$$\lambda(x,t) = \sum_{i=1}^{\tau(t)} f_i(x)\mathbb{1}_{\{t\in[t_{i-1},t_i)\}}, \quad \tau(t) := \underset{j\in\mathbb{N}}{\operatorname{argmin}}\{t < t_j\}$$

for functions $f_i : \mathbb{R}_+ \mapsto \mathbb{R}_+$.

Under Assumption 4.3.1 and system (4.2) the time domain is discretised and the following algorithm is constructed:

LSV Calibration Algorithm

1. Fix a time partition $\{t_0 = 0, ..., t_n = T\}$ and space partition $\{K_0 = K_{min}, ..., K_m = K_{max}\}$.
2. Initialize $\lambda(K_j, t_0) = \dfrac{\sigma_{Dup}(K_j, t_0)}{V_0}, \quad j = 0, ..., m$.

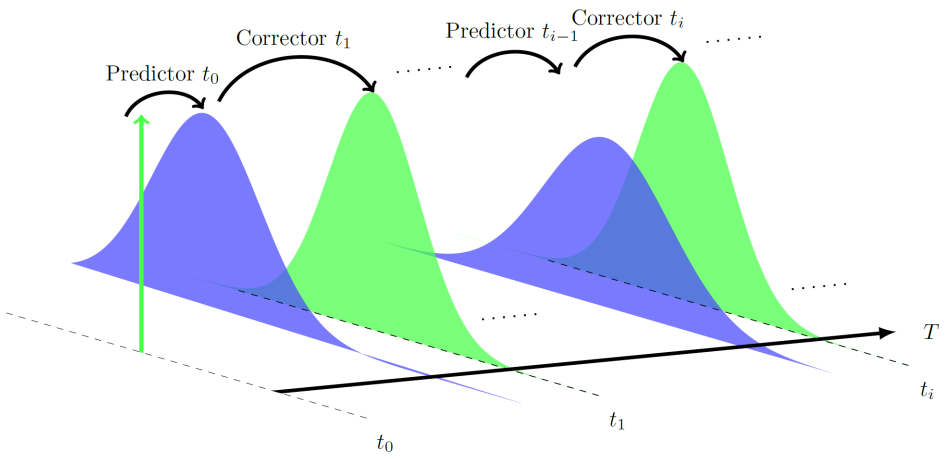


Figure 4.1: Graphical description of the forward in time LSV calibration algorithm. The t_0 distribution of the Spot is by convention a Dirac function, hence the arrow representation. The algorithm predicts the distribution of S_{t_i} implied by the model (blue density) at time t_{i-1} for time t_i with $\lambda(\cdot, t_i) = 1$ and corrects accordingly to match the σ_{Dup} market implied distribution of S_{t_i} (green density).

3. For $i = 1...n$:

   For $j = 0, ..., m$

   Step 1. Predictor:

   - Calculate $\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i} - \bar{r}_{t_i}\right)\mathbb{1}_{\{S_{t_i} > K_j\}}\right]$
   - Obtain $\dfrac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i} = K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i} = K_j]}$

   Step 2. Corrector:

   $$\bullet \lambda^2(K_j, t_i) = \left(\sigma_{Dup}(K_j, t_i)^2 - \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i} - \bar{r}_{t_i}\right)\mathbb{1}_{\{S_{t_i} > K_j\}}\right]}{\frac{1}{2}K_j \frac{\partial^2 C}{dK^2}}\right) \frac{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i} = K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i} = K_j]}.$$

As noted above, the procedure sequentially updates the leverage function λ, the critical step being the correct computation of $\frac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K_j]}$ at time t_{i-1}, meaning that we can compute $\lambda(\cdot, t_i)$ at the previous time step t_{i-1} and create a forward in time procedure. Figure 4.1 is a graphical description of the iterative procedure (forward in time) of the calibration algorithm.

4.3.1 The original particle method

In order to compute conditional expectations, Guyon and Henry-Labordère [GH12] propose a Monte Carlo based estimation via non-parametric Nadaraya-Watson kernel regression, more precisely if we denote by $\bar{S}^u_{t_i}$ and $\bar{V}^u_{t_i}$ the u-th particle obtained in a Monte Carlo simulation with M particles, then the estimator is given by:

$$\frac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i} = K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i} = K_j]} = \frac{\mathbb{E}^{\mathbb{Q}}[V_{t_i}D(t_i)\delta(S_{t_i} - K_j)]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)\delta(S_{t_i} - K_j)]} \approx \frac{\sum_{u=1}^{M} \bar{D}(t_i)^u \bar{V}^u_{t_i} K_h(\bar{S}^u_{t_i} - K_j)}{\sum_{u=1}^{M} \bar{D}(t_i)^u K_h(\bar{S}^u_{t_i} - K_j)}, \tag{4.4}$$

where $K_h(\cdot)$ is a suitable kernel function with bandwidth parameter $h > 0$. In the original article, a rule of thumb is given for the choice of h and the choice of a quartic kernel is recommended. In our numerical tests in Section 4.7, we follow these guidelines for implementing the particle method as benchmark.

4.3.2 Limitations of kernel regression: Bias, Variance and Convergence

It is well known, that non-parametric kernel regression methods suffer from a bias of order $\mathcal{O}(h^2)$ (see Härdle and Bowman [HB88]). Most importantly, the variance of the estimator is of order $\mathcal{O}\left((Mh)^{-1}\right)$. The need for a prespecified parameter h of such critical importance for the correct performance of the method, poses an inherent risk and makes it inclined to potential instabilities.

4.4 The exact formula

In order to obtain the closed-form formula for $\frac{\mathbb{E}^{\mathbb{Q}}[D(t)|S_t=K]}{\mathbb{E}^{\mathbb{Q}}[D(t)V_t|S_t=K]}$, we first obtain the following result, which is based on the classical result by Romano and Touzi [RT97].

Theorem 4.4.1. *Given model* (4.2) *and Assumption 4.3.1 we have*

$$\log S_{t_i}|\mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i} \sim \mathcal{N}\left(\mu_i, (1-\hat{\rho}^2)\sigma_i^2\right),$$

where

$$\mu_i := \log S_{t_{i-1}} + r_{t_{i-1}}\Delta t_i + \sqrt{V_{t_{i-1}}}\lambda(S_{t_{i-1}}, t_{i-1})\Delta W^{||}_{t_i} - \frac{1}{2}V_{t_{i-1}}\lambda(S_{t_{i-1}}, t_{i-1})^2\Delta t_i$$
$$\sigma_i^2 := V_{t_{i-1}}\lambda(S_{t_{i-1}}, t_{i-1})^2\Delta t_i, \quad \hat{\rho}^2 := \rho_{WZ}^T\Sigma_Z^{-1}\rho_{WZ}, \quad W^{||}_t := \rho_{WZ}^T\Sigma_Z^{-1}Z_t$$

Proof. First we see that,

$$\log S_{t_i}|\mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i} \sim \mu_i + \sqrt{1-\hat{\rho}^2}\sqrt{V_{t_{i-1}}}\lambda(S_{t_{i-1}}, t_{i-1})\Delta X_{t_i}$$

where X is a Brownian motion independent of $\mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i}$. We remark that $\lambda(S_{t_{i-1}}, t_{i-1})$ is $\mathcal{G}_{t_{i-1}}$-measurable under Assumption 4.3.1. Note also that $\mu_i \in \mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i}$ and $(1-\hat{\rho}^2)\sqrt{V_{t_{i-1}}}\lambda(S_{t_{i-1}}, t_{i-1})\Delta X_{t_i}$ has a deterministic integrand under conditioning, which by properties of the Itô integral is a centred Gaussian with variance $(1-\hat{\rho}^2)\sigma_i^2$ and the result follows. □

The next corollary gives the exact formula that we proclaimed.

Corollary 4.4.2. *Given model* (4.1) *and Assumption 4.3.1 we have*

$$\frac{\mathbb{E}^{\mathbb{Q}}[V_{t_i}D(t_i)|S_{t_i}=K]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K]} = \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)V_{t_i}\frac{e^{-\frac{1}{2}(d_i(K))^2}}{\sqrt{\sigma_i^2}}\right]}{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\frac{e^{-\frac{1}{2}(d_i(K))^2}}{\sqrt{\sigma_i^2}}\right]}, \quad d_i(K) = \frac{\mu_i - \log(K)}{\sqrt{(1-\hat{\rho}^2)\sigma_i^2}}. \tag{4.5}$$

Proof. We note that

$$\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K] = \frac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}\delta(S_{t_i}-K)]}{\mathbb{E}^{\mathbb{Q}}[\delta(S_{t_i}-K)]},$$

where $\delta(\cdot)$ represents the Dirac function at 0. Using the Tower property we further get

$$\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K] = \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)V_{t_i}\mathbb{E}^{\mathbb{Q}}[\delta(S_{t_i}-K)|\mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i}]\right]}{\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\delta(S_{t_i}-K)|\mathcal{F}^W_{t_{i-1}} \vee \mathcal{F}^Z_{t_i}]\right]},$$

where we have used that $V_{t_i}, D(t_i) \in \mathcal{F}_{t_i}^Z$. Next, we invoke Theorem 4.4.1, which gives the conditional probability density function $\phi(\cdot)$ of S:

$$\phi(x) = \frac{1}{x} \frac{e^{-\frac{1}{2}(d_i(x))^2}}{\sqrt{2\pi\sigma_i}}.$$

The result can be now directly derived by using the conditional density along with the Dirac function and the same procedure for $\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i} = K]$. □

In a Monte Carlo setting we may use Corollary 4.4.2 with the Euler discretisation of the processes involved to obtain the estimator:

$$\frac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i} = K]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i} = K]} = \frac{\mathbb{E}^{\mathbb{Q}}[V_{t_i}D(t_i)\delta(S_{t_i} - K)]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)\delta(S_{t_i} - K)]} \approx \frac{\sum_{u=1}^{M} \bar{D}(t_i)^u \bar{V}_{t_i}^u \frac{e^{-\frac{1}{2}(\bar{d}_i^u(K))^2}}{\sqrt{(\bar{\sigma}_i^u)^2}}}{\sum_{u=1}^{M} \bar{D}(t_i)^u \frac{e^{-\frac{1}{2}(\bar{d}_i^u(K))^2}}{\sqrt{(\bar{\sigma}_i^u)^2}}} \tag{4.6}$$

where u represents the u-th particle.

Remark 4.4.3. Note that expression (4.6) allows for unbiased Monte Carlo computation of the conditional expectations of the Euler scheme, without the use of external hyperparameters.

Remark 4.4.4. We note that for any pair of random variables X and Y we have,

$$\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y \mid X)) + \mathrm{Var}(\mathrm{E}(Y \mid X)).$$

Thus, one also has

$$\mathrm{Var}(\mathrm{E}(Y \mid X)) \leq \mathrm{Var}(Y).$$

Hence, the use of conditional expectations in Corollary 4.4.2, theoretically guarantees a variance reduction both for $\mathbb{E}^{\mathbb{Q}}[V_{t_i}D(t_i)\delta(S_{t_i} - K)]$ and $\mathbb{E}^{\mathbb{Q}}[D(t_i)\delta(S_{t_i} - K)]$.

4.4.1 Further exploiting the discrete nature of numerical algorithms: lognormal SV case

The results developed in Corollary 4.4.2 allow to obtain a closed-form expression for virtually any SV model. In spite of the universality of the result, there is a (mild) memory cost of storing $\sqrt{V_{t_{i-1}}}\lambda(S_{t_{i-1}}, t_{i-1})\Delta W_{t_i}^{||}$. In this section, we show how to remove this dependence. For simplicity, in this section we assume that the joint distribution of the logarithm of instantaneous variance and W in (4.1) follows a multivariate normal distribution,

$$\log(V_{t_i}) \sim \mathcal{N}(\xi_i, \nu_i). \tag{4.7}$$

Hence, the conditional distribution is also given by a lognormal random variable:

$$\begin{pmatrix} \log(V_{t_i})|\mathcal{G}_{t_{i-1}} \\ W_{t_i}|\mathcal{G}_{t_{i-1}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \widetilde{\xi}_i \\ W_{i-1} \end{pmatrix}, \begin{pmatrix} \widetilde{\nu}_i & \rho\sqrt{\widetilde{\nu}_i(t_i - t_{i-1}} \\ \sqrt{\widetilde{\nu}_i(t_i - t_{i-1}} & t_i - t_{i-1} \end{pmatrix} \right), \tag{4.8}$$

where for simplicity we consider a constant correlation parameter $\rho \in (-1, 1)$. Furthermore, we consider the processes V and r to be piecewise constant, i.e.

$$V_t := \sum_{i=1}^{\tau(t)} V_{t_{i-1}} \mathbb{1}_{\{t \in [t_{i-1}, t_i)\}}, \quad r_t := \sum_{i=1}^{\tau(t)} r_{t_{i-1}} \mathbb{1}_{\{t \in [t_{i-1}, t_i)\}}, \quad \tau(t) := \operatorname*{argmin}_{j \in \mathbb{N}} \{t < t_j\}. \tag{4.9}$$

Note that (4.9) is the usual time discretisation needed to perform numerical (forward Euler) simulation or PDE pricing; it therefore does not pose additional constraints.

Proposition 4.4.5. *Under Assumption 4.3.1 and model* (4.2) *with* V *and* r *given by* (4.7)-(4.9) *we have,*

$$\frac{\mathbb{E}^{\mathbb{Q}}[V_{t_i}D(t_i)|S_{t_i}=K]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K]} = \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\frac{\exp\left\{\widetilde{\xi}_i+\frac{1}{2\sigma_i^2(1-\rho^2)}\left((\widetilde{\mu}_i-\log(K))^2+(\rho(\widetilde{\mu}_i-\log(K))-(1-\rho^2)\widetilde{\nu}_i\sigma_i\right)^2\right\}}{\sqrt{\sigma_i^2}}\right]}{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\frac{e^{-\frac{1}{2}(\widetilde{d}_i(K))^2}}{\sqrt{\sigma_i^2}}\right]}, \tag{4.10}$$

where

$$\widetilde{\mu}_i := \log S_{t_{i-1}} + r_{t_{i-1}}\Delta t_i - \frac{1}{2}V_{t_{i-1}}\lambda(S_{t_{i-1}},t_{i-1})^2\Delta t_i$$
$$\widetilde{d}_i(K) := \frac{\widetilde{\mu}_i - \log(K)}{\sqrt{\sigma_i^2}}, \quad \rho := corr(W_{t_i}-W_{t_{i-1}}, \log(V_{t_i})|\mathcal{G}_{t_{i-1}}).$$

Proof. As in the previous result, by definition we have

$$\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K] = \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\mathbb{E}^{\mathbb{Q}}[V_{t_i}\delta(S_{t_i}-K)|\mathcal{G}_{t_{i-1}}]\right]}{\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\delta(S_{t_i}-K)|\mathcal{G}_{t_{i-1}}]\right]}.$$

Using (4.8) and (4.9), we note that in the numerator we have a bivariate Gaussian distribution with the variables:

$$\begin{pmatrix}\log(S_{t_i})|\mathcal{G}_{t_{i-1}}\\ \log(V_{t_i})|\mathcal{G}_{t_{i-1}}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\widetilde{\mu}_i\\ \widetilde{\xi}_i\end{pmatrix}, \begin{pmatrix}\sigma_i^2 & \rho\sqrt{\sigma_i^2\widetilde{\nu}_i}\\ \rho\sqrt{\sigma_i^2\widetilde{\nu}_i} & \widetilde{\nu}_i\end{pmatrix}\right).$$

Computing the conditional expectation in the numerator yields,

$$\mathbb{E}^{\mathbb{Q}}[V_{t_i}\delta(S_{t_i}-K)|\mathcal{G}_{t_{i-1}}] = \frac{\exp\left\{\widetilde{\xi}_i+\frac{1}{2\sigma_i^2(1-\rho^2)}\left((\widetilde{\mu}_i-\log(K))^2+(\rho(\widetilde{\mu}_i-\log(K))-(1-\rho^2)\widetilde{\nu}_i\sigma_i\right)^2\right\}}{\sqrt{2\pi(1-\rho^2)\sigma_i^2}},$$

and the result readily follows by simplifying terms and proceeding similarly with $\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K]$. □

Remark 4.4.6. Even though we only considered lognormal SV models in Proposition 4.4.5, it should be possible to derive analytic expressions with other dynamics as long as the conditional expectations are available in closed-form. This computations, in principle, needs to be done in a case by case basis.

4.5 The new LSV calibration algorithm

1. Fix a time partition $\{t_0=0,...,t_n=T\}$ and space partition $\{K_0=K_{min},...,K_m=K_{max}\}$.
2. Initialize $\lambda(K_j,t_0)=\dfrac{\sigma_{Dup}(K_j,t_0)}{V_0}, \quad j=0,...,m$.
3. For $i=1...n$:
   For $j=0,...,m$

   Step 1. Predictor:
   - Diffuse particles from t_{i-1} to t_i according to (4.1) and compute $\int_{t_{i-1}}^{t_i}\sqrt{V_u}\lambda(S_{t_{i-1}},u)dW_u^{||}$
   - Use formula (4.6) to obtain $\dfrac{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K_j]}$

   Step 2. Corrector:
   - $\lambda^2(K_j,t_i)=\left(\sigma_{Dup}(K_j,t_i)^2-\dfrac{\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_{t_i}\right)\mathbb{1}_{\{S_{t_i}>K_j\}}\right]}{\frac{1}{2}K_j\frac{\partial^2 C}{dK^2}}\right)\dfrac{\mathbb{E}^{\mathbb{Q}}[D(t_i)|S_{t_i}=K_j]}{\mathbb{E}^{\mathbb{Q}}[D(t_i)V_{t_i}|S_{t_i}=K_j]}$.
   - Extrapolate flat outside $[K_{min},K_{max}]$ and lay a cubic spline in between.

Remark 4.5.1. If V is lognormal one can replace (4.6) by (4.10).

Computational considerations:

In principle, the summands in expressions (4.4) and (4.6) need to be computed for all M particles. One can consider sorting particles according to spot value with cost $\mathcal{O}(M \log M)$ as in Guyon and Henry-Labordère [GH12], set lower and upper thresholds $0 < L < U$ and consider a modified kernel $\tilde{K}_h(x) = K_h(x)\mathbf{1}_{\{x\in[L,U]\}}$. Then, one is able to easily disregard particles in the sorted array.

We must note that ultimately, the savings in computational time will depend on the Monte Carlo sample size and the efficiency of the sorting procedure. If one wishes to mimic the same technique for (4.6), it is preferable to sort particles according to $\frac{\mu_i}{\sqrt{\sigma_i^2}}$ to set a threshold. Whether sorting particles or not, the computational complexity of our algorithm will be virtually the same as the original particle method as one can see from expressions (4.4) and (4.6) and Figure 4.4.

Additionally, it is preferable to use the alternative representation (see Appendix C.1)

$$\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbf{1}_{\{S_{t_i}>K_j\}}\right] = \mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\Phi(-d_i(K_j))\right] \tag{4.11}$$

$$= -\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\Phi(d_i(K_j))\right]. \tag{4.12}$$

where $\Phi(\cdot)$ is the standard Gaussian cdf. This alternative representations guarantee a variance reduction by the law of total variance. Furthermore, if one uses the sorting technique, for low strikes $\mathbf{1}_{\{S_{t_i}>K_j\}}$ will be one with high likelihood and particles cannot be disregarded. Therefore, we can use expression (4.11) for $K > S_0$ and (4.12) for $K \leq S_0$ to disregard particles and keep the computational cost controlled. We further emphasise that when extrapolating, if $K \to 0$ or $K \to \infty$ the expectations above converge to zero.

Remark 4.5.2. An alternative approach is also presented for (4.11) in [GH12] via Malliavin representation techniques.

Theoretical considerations:

Precise conditions on system (4.1)-(4.3) for a solution to exists still remains an open question. In [CMR19; GH12] it is reported that for large values of volatility of volatility the algorithm fails to converge. We tested this phenomena, but seems that the domain of convergence of our algorithm is superior to that of the particle method; in Figure 4.2 the kernel regression suggests that a fit is not possible a posteriori and one could incorrectly conclude that a solution does not exist for values of $\nu \geq 380\%$, hence the boundary of convergence the numerical scheme would be $\nu < 380\%$. On the contrary, our method clearly shows that a solution exists for $\nu = 380\%$, hence improves the convergence boundary. Whether such behaviour in the kernel regression is caused by the choice of an inadequate bandwidth parameter h remains unclear, though raises the issue of instability in the original method. Further tests with higher level of volatility of volatility showed that both our algorithm and the original particle method fail (see Figure 4.3) to match the market smiles. The precise theoretical identification of an upper bound in volatility of volatility remains unanswered.

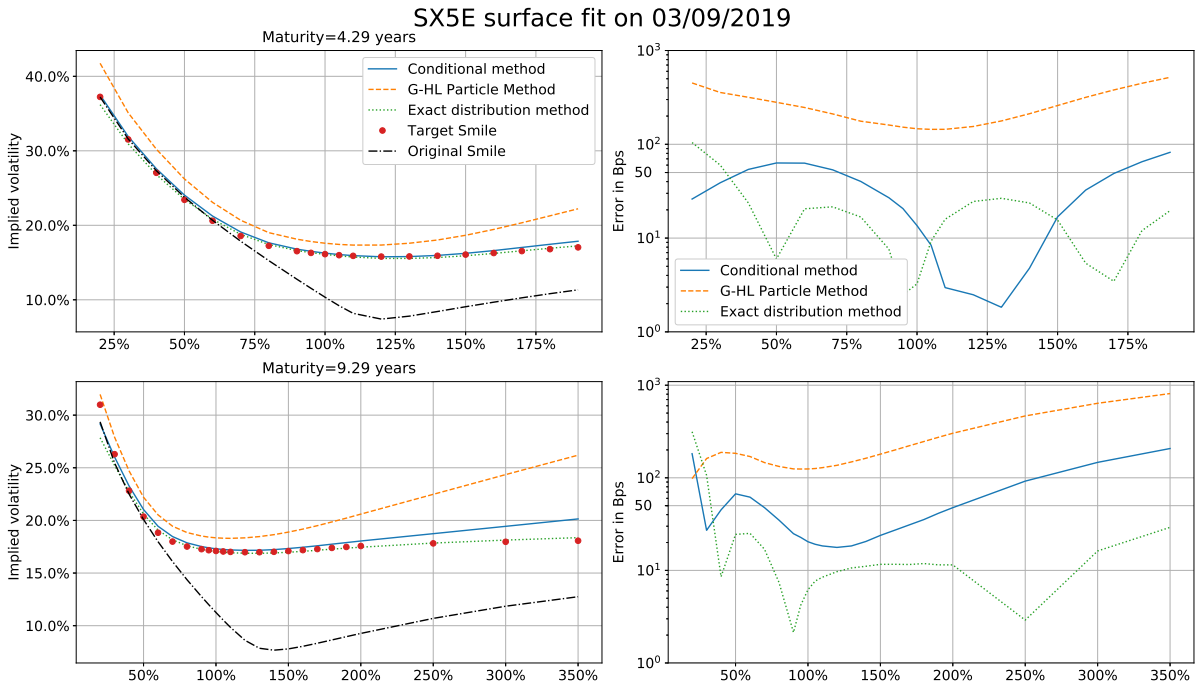


Figure 4.2: SX5E (3-Sept-2019) Implied volatility surface. Rough Bergomi parameters: $H = 0.2$, $\rho_{WZ} = -80\%$, $\beta = 0.5$, $\nu = 380\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $\rho_{WY} = 0\%$, $r_0 = 1.5\%$, $\rho_{ZY} = 0\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4).

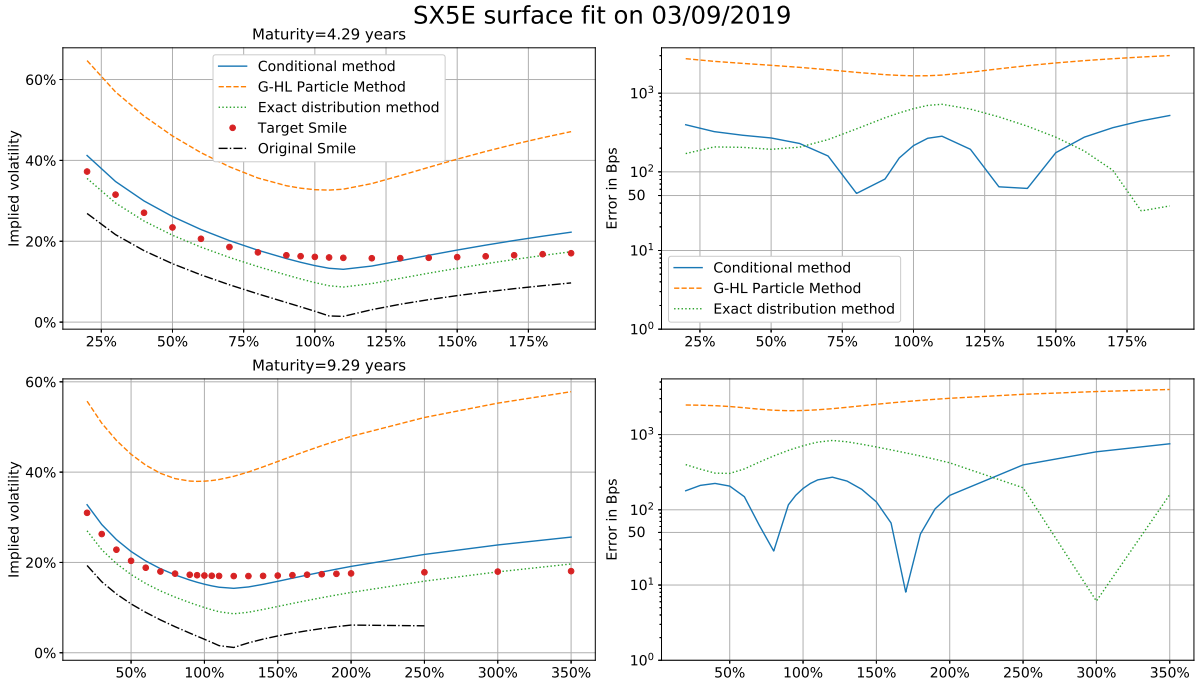


Figure 4.3: SX5E (3-Sept-2019) Implied volatility surface. Rough Bergomi parameters: $H = 0.2$, $\rho_{WZ} = -80\%$, $\beta = 0.5$, $\nu = 800\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $\rho_{WY} = 0\%$, $r_0 = 1.5\%$, $\rho_{ZY} = 0\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4).

4.6 Overture to local correlation and basket smiles

So far, we have only considered the mono-underlying case, but there is no reason why our method cannot be extended to a multi-asset environment. Let us consider an index $I_t = \sum_{k=1}^{p} \omega_k S_t^k$ with $p \in \mathbb{N}$, where

$$\begin{aligned}
\frac{\mathrm{d}I_t}{I_t} &= r_t \mathrm{d}t + \sum_{k=1}^{p} \frac{\omega_k S_t^k \lambda^i(S_t^k, t)\sqrt{V_t^k}}{I_t} \mathrm{d}W_t^k \\
\frac{\mathrm{d}S_t^k}{S_t^k} &= r_t \mathrm{d}t + \lambda^k(S_t^k, t)\sqrt{V_t^k}\mathrm{d}W_t^k, \quad k = 1, ..., p \\
\mathrm{d}[W^i, W^j]_t &:= \rho_{ij}(I_t, t)\mathrm{d}t, \quad k = 1, ..., p.
\end{aligned}$$

The condition for the index (or basket) to be calibrated to market smiles is given by (see Guyon [Guy14]):

$$\sum_{i=1}^{p}\sum_{j=1}^{p} \omega_i \omega_j \rho_{ij}(K,t) \frac{\mathbb{E}^{\mathbb{Q}}[D(t)\lambda^i(S_t^i,t)\sqrt{V_t^i}S_t^i \lambda^j(S_t^j,t)\sqrt{V_t^j}S_t^j | I_t = K]}{K^2 \mathbb{E}^{\mathbb{Q}}[D(t)|I_t = K]} = \sigma_{Dup}(K,t)^2 - \frac{\mathbb{E}^{\mathbb{Q}}\left[D(t)\,(r_t - \bar{r}_t)\,\mathbb{1}_{\{I_t > K\}}\right]}{\frac{1}{2}K\frac{\partial^2 C}{dK^2}}. \tag{4.13}$$

At this stage an assumption on the particular structure of $\rho_{ij}(\cdot,\cdot)$ is typically made, both to ensure computational feasibility and positive definiteness of the correlation matrix. Nevertheless, for the purpose of numerically solving (4.13), the problem is equivalent to being able to compute

$$\frac{\mathbb{E}^{\mathbb{Q}}[D(t)\lambda^i(S_t^i,t)\sqrt{V_t^i}S_t^i \lambda^j(S_t^j,t)\sqrt{V_t^j}S_t^j | I_t = K]}{\mathbb{E}^{\mathbb{Q}}[D(t)|I_t = K]}, \quad i,j = 1, .., p.$$

Both Corollary 4.4.2 and Proposition 4.4.5 can be easily adapted to this setting; one needs to find the appropriate filtration $\mathcal{G}_t^I$ to make the tuple $(\log S_{t_i}^i, \log S_{t_i}^j)|\mathcal{G}_{t_{i-1}}^I$ jointly normal. Therefore, similar closed-form and exact expressions can be obtained after some tedious (but not difficult) calculations.

4.7 Numerical tests with hybrid models

4.7.1 Vasicek and rough local stochastic volatility

We first consider a rough volatility model, similar to the one introduced by Bayer, Friz and Gatheral [BFG16] topped with the Vasicek stochastic interest rate model. It is well known (see Alòs, León and Vives [ALV07]) that such models reproduce a power law decay on the short time ATM skew. However, to control the long term behaviour we also add a mean-reversion parameter β. It is worth noting that non-Markovian dynamics imply that past behaviour of the volatility process influences the future behaviour.

$$\begin{aligned}
\frac{\mathrm{d}S_t}{S_t} &= r_t \mathrm{d}t + \lambda(S_t, t)\sqrt{V_t}\mathrm{d}W_t \\
V_t &= \xi_0(t)\mathcal{E}\left(\nu\sqrt{2H}\int_0^t (t-s)^{H-1/2} e^{-\beta(t-s)}\mathrm{d}Z_s\right), \quad \beta, \nu > 0,\ H \in (0, 1/2) \\
\mathrm{d}r_t &= (r_0 - \kappa r_t)\mathrm{d}t + \sigma \mathrm{d}Y_t, \quad r_0 \in \mathbb{R} \\
\mathrm{d}[W,Z]_t &= \rho_{WZ}\mathrm{d}t, \quad \mathrm{d}[W,Y]_t = \rho_{WY}\mathrm{d}t, \quad \mathrm{d}[Z,Y]_t = \rho_{ZY}\mathrm{d}t.
\end{aligned}$$

where $\mathcal{E}(x) = \exp(x - \frac{1}{2}Var(x))$. Due to its non-Markovian nature, there is no forward Kolmogorov equation known for the transition density. Hence PDE methods are out of the picture and one can only resort to simulation based methods (see Algorithm 2.4.5 for details on simulation). In Figure 4.6 we report the LSV calibration results with

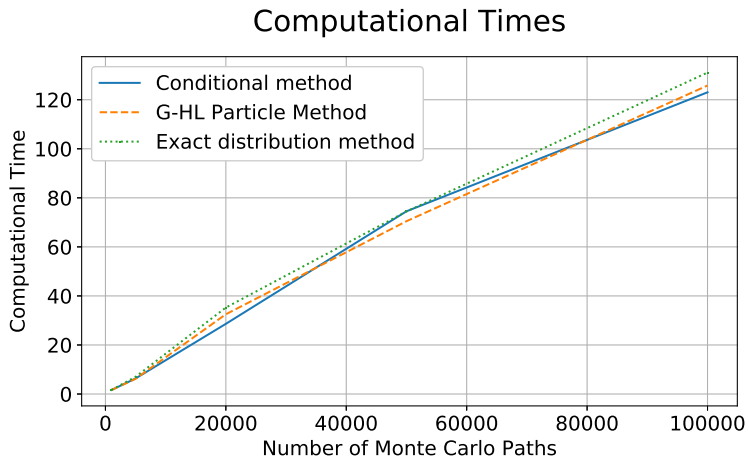


Figure 4.4: computational times relative to the G-HL particle method with 100 Monte Carlo paths, meaning that this represents the unit in the time scale.he conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4).

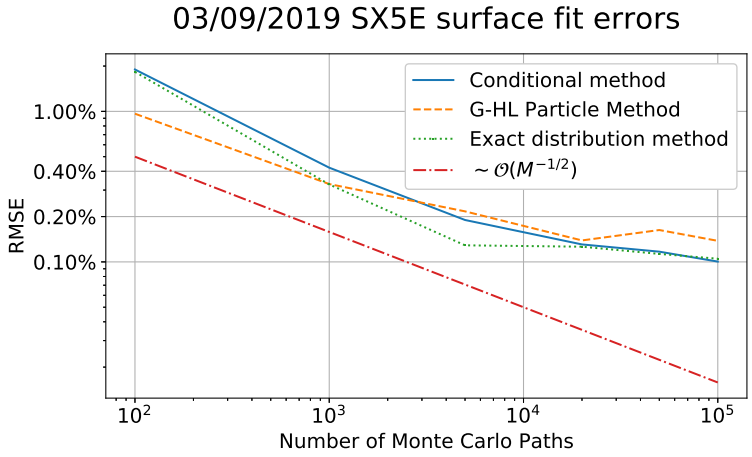


Figure 4.5: RMSE as a function of the number of Monte Carlo paths and the magnituded $\mathcal{O}(M^{-1/2})$ is also given as benchmark. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4).

50.000 Monte Carlo paths and $\frac{1}{252}$ time step. We observe that the proposed algorithm (both with the conditional approach and distribution approach) converges appropriately and performs at the level of the particle method [GH12]. For all tested maturities, we obtain an accuracy of a few basis points for reasonable strikes. In Figure 4.5 we further report computational times and RMSE as a function of Monte Carlo Paths. We do not observe a significant computational time difference between (4.6) and (4.10); The RMSE, however seems to be lower with methods (4.6) and (4.10) than (4.4), when the number of particles is greater than 10^4.

4.7.2 Vasicek and local 2F Bergomi

To test our method in a higher dimensional setting, we consider the two Factor Bergomi [Ber08] model with stochastic interest rates given by a Vasicek interest rate model. While Bergomi's model is a popular equity model due to its flexibility to fit dynamical properties of the smile, it is driven by a 4 dimensional Brownian, which poses

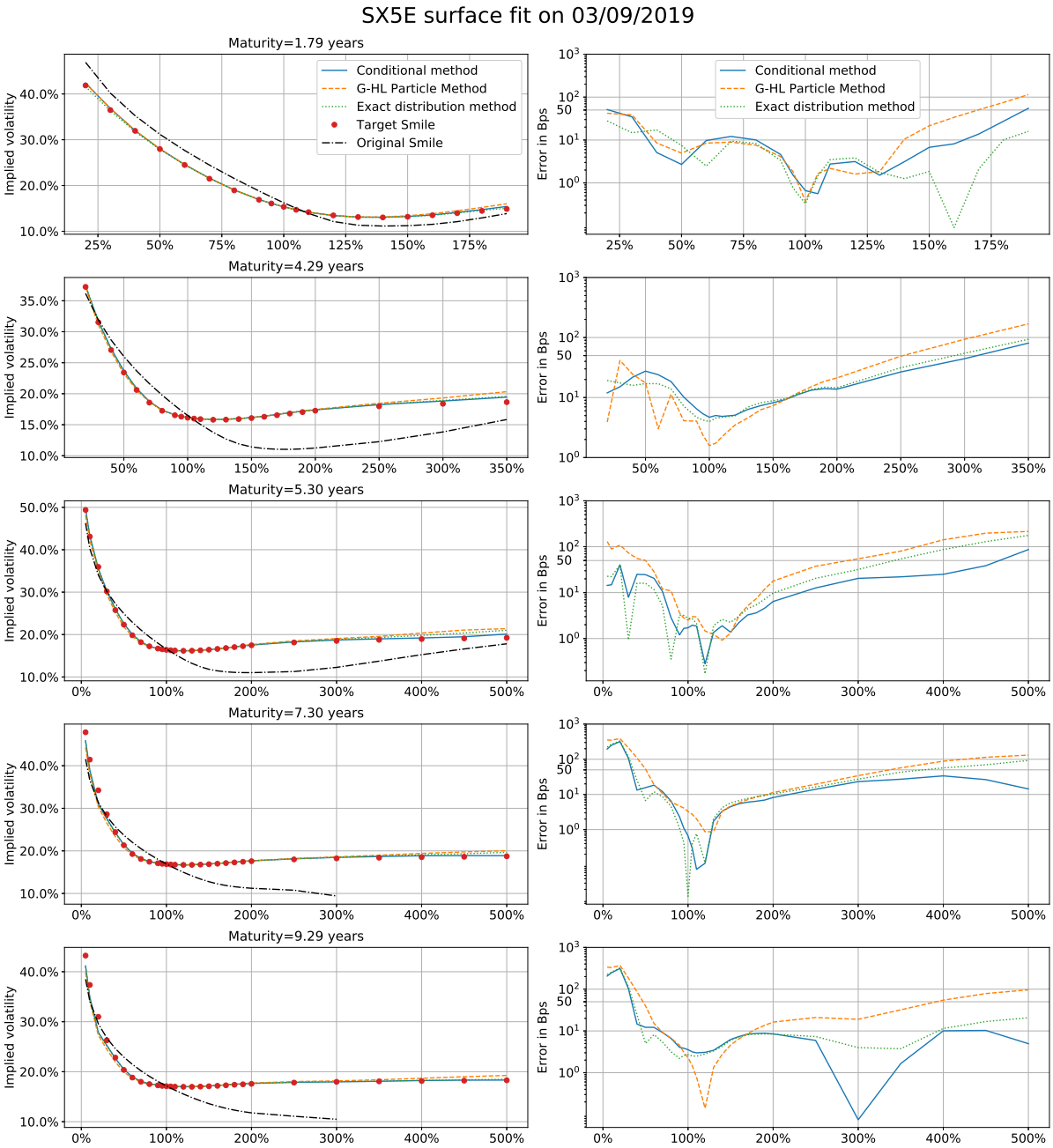


Figure 4.6: SX5E (3-Sept-2019) Implied volatility surface. Rough Bergomi parameters: $H = 0.2$, $\rho_{WZ} = -80\%$, $\beta = 0.5$, $\nu = 200\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $\rho_{WY} = 0\%$, $r_0 = 1.5\%$, $\rho_{ZY} = 0\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4) and Original Smile is the pure SV smile generated when $\lambda(\cdot, \cdot) = 1$

a remarkable challenge on a PDE framework.

$$\begin{aligned}
\frac{\mathrm{d}S_t}{S_t} &= r_t \mathrm{d}t + \lambda(S_t, t)\sqrt{V_t}\mathrm{d}W_t \\
V_t &= \xi_0(t)\mathcal{E}\left(\nu\alpha_\theta\left((1-\theta)\int_0^t e^{-\kappa_X(t-s)}\mathrm{d}X_s + \theta\int_0^t e^{-\kappa_Y(t-s)}\mathrm{d}Y_s\right)\right), \quad \kappa_X, \kappa_Y, \nu > 0,\ \theta \in [0,1] \\
\mathrm{d}r_t &= (r_0 - \kappa r_t)\mathrm{d}t + \sigma \mathrm{d}Z_t, \quad r_0 \in \mathbb{R}
\end{aligned}$$

In Figures 4.7, 4.8 and 4.9 we report the results with 50.000 Monte Carlo paths and $\frac{1}{252}$ time step. The results show again that our algorithm performs as good as the particle method and converges successfully, with accuracies of few basis points for relevant strikes as before. Furthermore, we observe a better accuracy for OTM strikes.

4.8 Summary

In this Chapter we have presented a new Monte Carlo estimation formula to calibrate the leverage function to any stochastic volatility model. Building upon the particle method developed by Guyon and Henry-Labordère [GH12], we obtained a closed-form calibration procedure that does not depend on any external parameter fine-tuning nor incurs any additional computational cost. Our experiments show that the accuracy of the method is superior to that of the original particle method and improves the convergence boundary in volatility of volatility. Overall, this new approach allows for an easier, safer and more robust implementation in an automatised production level environment.

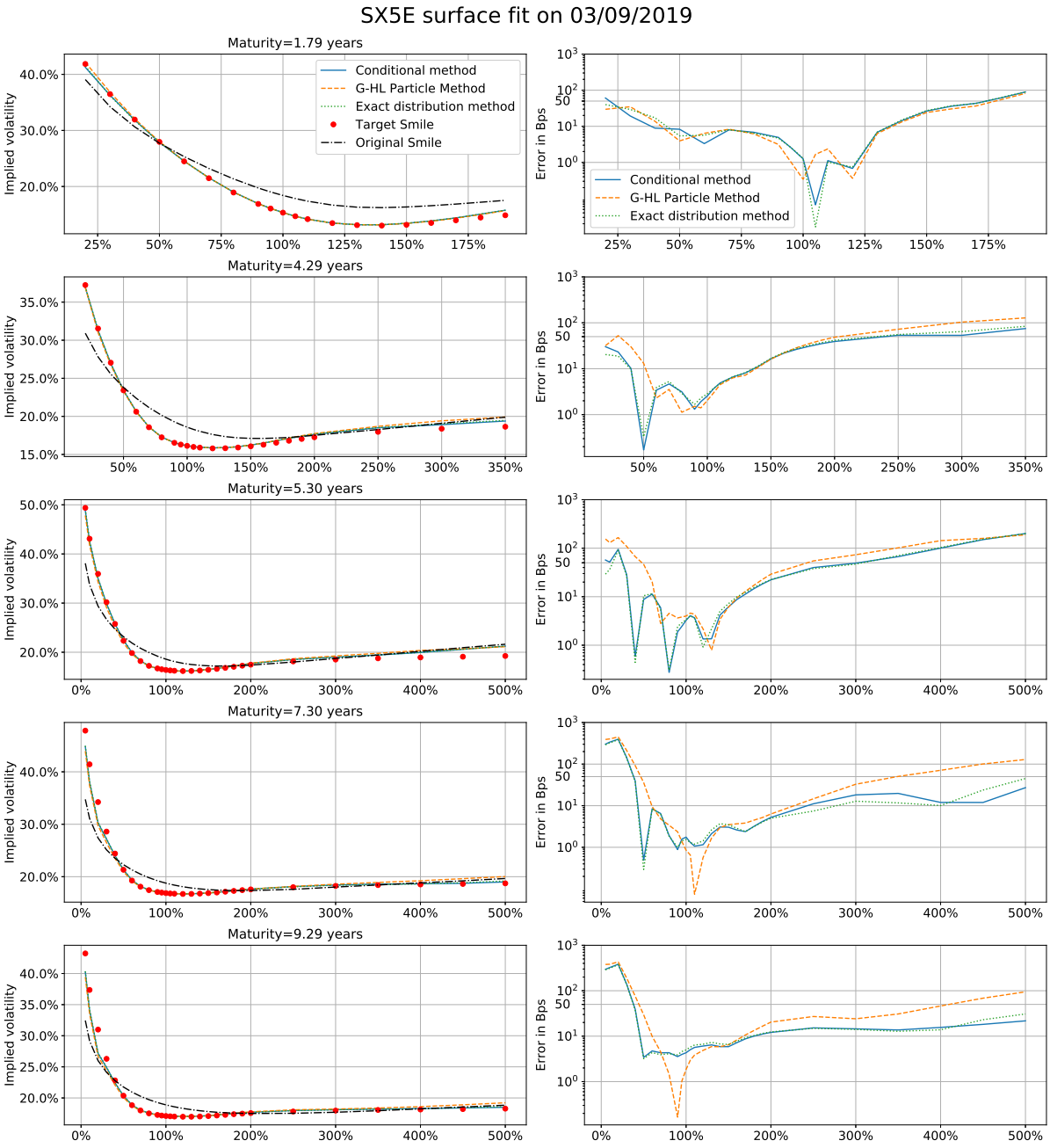


Figure 4.7: SX5E (3-Sept-2019) Implied volatility surface. 2F Bergomi parameters: $\kappa_X = 0.5$, $\kappa_Y = 8$, $\theta = 80\%$, $\rho_{WX} = -10\%$, $\rho_{WY} = -80\%$, $\rho_{XY} = 30\%$, $\nu = 400\%$, $\rho_{WZ} = \rho_{XZ} = \rho_{YZ} = 0\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $r_0 = 1.5\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4) and Original Smile is the pure SV smile generated when $\lambda(\cdot, \cdot) = 1$.

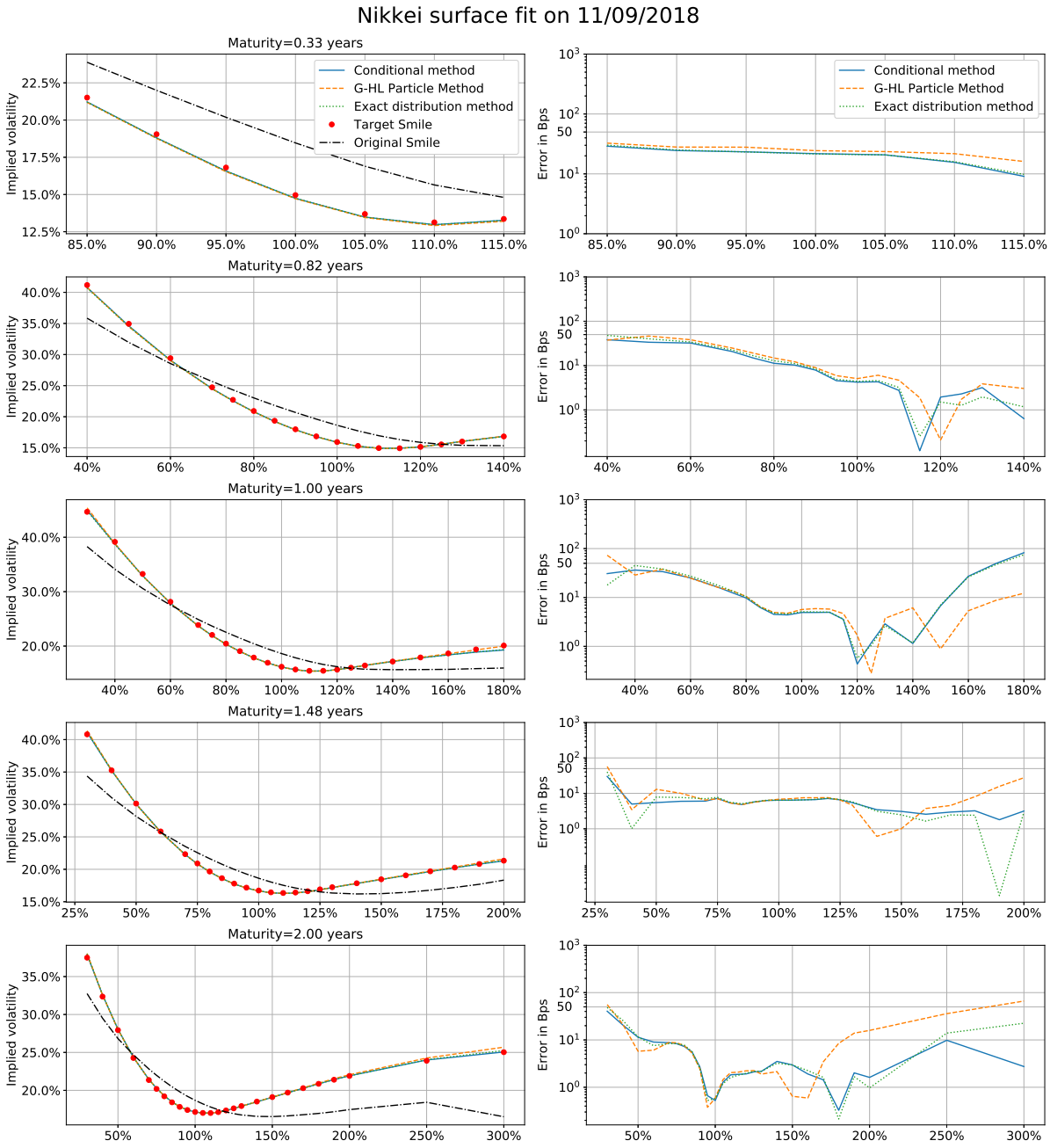


Figure 4.8: Nikkei (11-Sept-2018) Implied volatility surface. 2F Bergomi parameters: $\kappa_X = 0.9$, $\kappa_Y = 10$, $\theta = 80\%$, $\rho_{WX} = -10\%$, $\rho_{WY} = -80\%$, $\rho_{XY} = 30\%$, $\nu = 450\%$, $\rho_{WZ} = \rho_{XZ} = \rho_{YZ} = 0\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $r_0 = 1.5\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4) and Original Smile is the pure SV smile generated when $\lambda(\cdot, \cdot) = 1$.

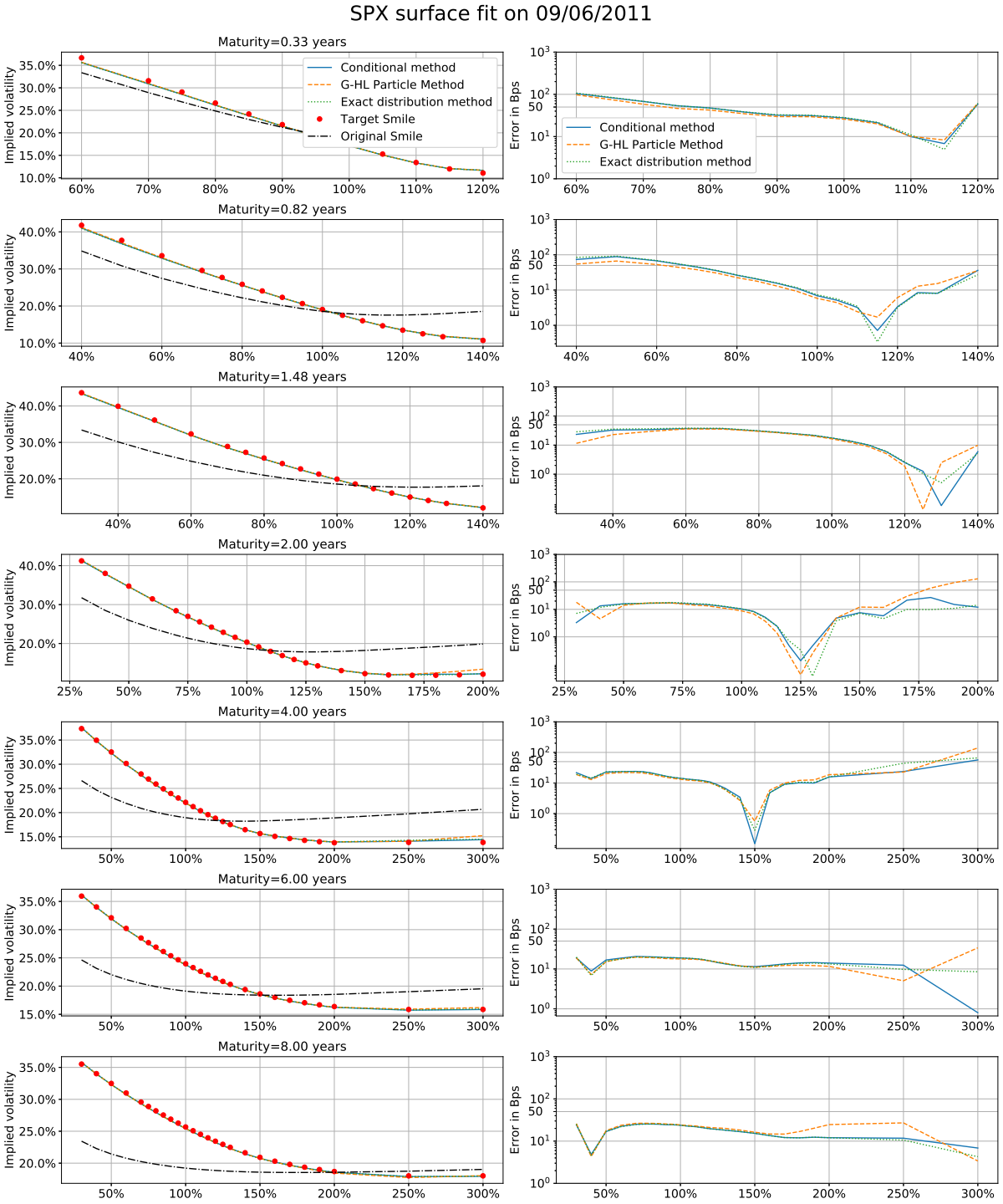


Figure 4.9: SPX (09-Jun-2011) Implied volatility surface. 2F Bergomi parameters: $\kappa_X = 0.9$, $\kappa_Y = 10$, $\theta = 80\%$, $\rho_{WX} = -10\%$, $\rho_{WY} = -80\%$, $\rho_{XY} = 50\%$, $\nu = 450\%$, $\rho_{WZ} = \rho_{XZ} = \rho_{YZ} = 0\%$. Vasicek parameters: $\kappa = 1$, $\sigma = 0.5\%$, $r_0 = 1.5\%$. The conditional and exact distribution methods are given by equations (4.6) and (4.10) respectively. G-HL Particle method is described in (4.4) and Original Smile is the pure SV smile generated when $\lambda(\cdot, \cdot) = 1$.

5

On VIX Futures in the rough Bergomi model

5.1 Introduction

Volatility, though not directly observed nor traded, is a fundamental object on financial markets, and has been the centre of attention of decades of theoretical and practical research, both to estimate it and to use it for trading purposes. The former goal has usually been carried out under the historical measure ($\mathbb{P}$) while the latter, through the introduction of volatility derivatives (VIX and related family), has been evolving under the pricing measure $\mathbb{Q}$. One of the key issues in Equity markets is, not only to fit the (SPX) implied volatility smile, but to do so jointly with a calibration of the VIX (Futures and ideally options). Gatheral's [Gat08] double mean reverting process is the leading (Markovian) continuous model in this direction, while models with jumps have been proposed abundantly by Carr and Madan [CM14] and Kokholm and Stisen [KS15]. This issue was briefly tackled by Bayer, Friz and Gatheral [BFG16] for a particular rough model (rough Bergomi), and we aim here at providing a deeper analysis of VIX dynamics under this rough model and at implementing pricing schemes for VIX Futures and options. Our main contribution is a precise link between the forward variance curve $(\xi_T(\cdot))_{T\geq 0}$ and the initial forward variance curve $\xi_0(\cdot)$ in the rough Bergomi model. This in turn, allows us not only to provide simulation methods for the VIX, but also to refine the log-normal approximation of [BFG16] for VIX Futures, matching exactly the first two moments. Finally, we develop an efficient algorithm for VIX Futures calibration, upon which we build a joint calibration method with the SPX. As opposed to the Cholesky approach in [BFG16], we adapt the hybrid-scheme by Bennedsen, Lunde and Pakkanen [BLP17] with better complexity $\mathcal{O}(n \log n)$. Assuming the universality of the Hurst parameter H across VIX and SPX allows us to compute efficiently prices recursively with complexity $\mathcal{O}(n)$. In passing, we also investigate the joint consistency of VIX and SPX in the market. The organisation of the Chapter follows accordingly: we first introduce the rough Bergomi model and its main properties (Section 5.2), before presenting its pricing power for VIX Futures (Section 5.3), and finally develop the joint calibration algorithm in Section 5.4.

5.2 Rough volatility and the rough Bergomi model

Comte and Renault [CR96] were the first to propose a stochastic volatility model in which the instantaneous volatility is driven by a fractional Brownian motion, with a Hurst index restricted to $(1/2, 1)$. Originally introduced by Alòs, León and Vives [ALV07] and Fukasawa [Fuk11b], and recently promoted further by Gatheral, Jaisson and

Rosenbaum [GJR18], stochastic volatility models with Hurst index smaller than $1/2$ are able to produce extremely good fits to observed volatility data under the physical measure $\mathbb{P}$. These models form the so-called Rough Fractional Stochastic Volatility (RFSV) family that is understood as a natural extension of classical Markovian stochastic volatility models. Our work focuses on the pricing measure $\mathbb{Q}$ and we assume through this Chapter that the model presented by Gatheral, Jaisson and Rosenbaum [GJR18] under $\mathbb{P}$ is a reasonable model. Finally, and most importantly, we follow the recent paper by Bayer, Friz and Gatheral [BFG16], in order to extend the RFSV model to pricing schemes under the measure $\mathbb{Q}$. More precisely, Bayer, Friz and Gatheral [BFG16] proposed the following model for the log stock price process $X := \log(S)$:

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2}V_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}W_t, \quad X_0 = 0 \\ V_t &= \xi_0(t)\mathcal{E}(2\nu C_H \mathcal{V}_t), \qquad V_0 > 0, \end{aligned} \tag{5.1}$$

with $\nu, \xi_0(\cdot) > 0$, $\mathcal{E}(\cdot)$ is the stochastic exponential[1] and $C_H := \sqrt{\frac{2H\Gamma(2-H_+)}{\Gamma(H_+)\Gamma(2-2H)}}$, where, for notational convenience (throughout the Chapter), we use the symbols $H_\pm := H \pm \frac{1}{2}$. All the processes are defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ supporting the two standard Brownian motions W and Z (see below). The initial forward variance curve is observed at inception, and we therefore assume without loss of generality that it is $\mathcal{F}_0$-measurable. The process $\mathcal{V}$, defined as

$$\mathcal{V}_t := \int_0^t (t-u)^{H_-} \mathrm{d}Z_u, \tag{5.2}$$

is a centred Gaussian process with covariance structure

$$\mathbb{E}(\mathcal{V}_t \mathcal{V}_s) = s^{2H} \int_0^1 \left(\frac{t}{s} - u\right)^{H_-} (1-u)^{H_-} \mathrm{d}u, \qquad \text{for any } s,t \in [0,1].$$

We shall also introduce, for any $0 \leq t_1 \leq t_2 \leq t$, the notation

$$\mathcal{V}_t^{[t_1,t_2]} := \int_{t_1}^{t_2} (t-u)^{H_-} \mathrm{d}Z_u, \tag{5.3}$$

and write $\mathcal{V}_t^{t_2} = \mathcal{V}_t^{[0,t_2]}$ whenever the lower bound of the integral is null. Note in particular that $\mathcal{V}_T^T = \mathcal{V}_T$. The two standard Brownian motions W and Z are correlated with correlation parameter $\rho \in (-1,1)$. Here, $\xi_T(t)$ denotes the forward variance observed at time T for a maturity equal to t. More precisely, if $\sigma_T^2(t)$ denotes the fair strike of a variance swap observed at time T and maturing at t, then

$$\sigma_T^2(t) = \frac{1}{t-T}\int_T^t \xi_T(u)\mathrm{d}u,$$

or equivalently $\xi_T(t) = \frac{\mathrm{d}}{\mathrm{d}t}\big((t-T)\sigma_T^2(t)\big)$. For any fixed $t > 0$, the process $(\xi_s(t))_{s\leq t}$, is a martingale, i.e. $\mathbb{E}[\xi_s(t)|\mathcal{F}_u] = \xi_u(t)$, for all $u \leq s \leq t$. Furthermore, $\mathcal{V}^T$ is a centred Gaussian process with variance

$$\mathbb{V}(\mathcal{V}_t^T) = \frac{t^{2H} - (t-T)^{2H}}{2H}, \qquad \text{for } t \geq T, \tag{5.4}$$

and covariance structure

$$\mathbb{E}\left(\mathcal{V}_t^T \mathcal{V}_s^T\right) = \int_0^T \left[(t-u)(s-u)\right]^{H_-} \mathrm{d}u = \frac{(s-t)^{H_-}}{H_+}\left\{t^{H_+}\mathrm{F}\left(\frac{-t}{s-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{s-t}\right)\right\}, \tag{5.5}$$

for any $t < s$, where we introduce the function $\mathrm{F} : \mathbb{R}_- \to \mathbb{R}$ as

$$\mathrm{F}(u) := {}_2F_1\left(-H_-, H_+, 1+H_+, u\right), \tag{5.6}$$

[1] For a continuous semimartingale Y starting from zero, the Doléans-Dade stochastic exponential [Dol70] is defined as $\mathcal{E}(Y)_t := \exp(Y_t - \frac{1}{2}[Y,Y]_{0,t})$. Here, $\mathrm{Y} = \mathcal{V}$ is a zero-mean Gaussian process, with infinite quadratic variation, and this definition does not make sense. However, the Wick exponential $\mathcal{E}(Y)_t := \exp(Y_t - \frac{1}{2}\mathbb{E}(Y_t^2))$ does, and we conveniently keep this notation without further details, following [BFG16, Remark 1.1].

and ${}_2F_1$ is the hypergeometric function [ASR88, Chapter 15].

5.3 Rough Bergomi and VIX

Bayer, Friz and Gatheral [BFG16] briefly discuss the lack of consistency of the rough Bergomi model with observed VIX options data, leading to an incorrect term structure of the VIX. In this section, we investigate in detail the dynamics of the VIX, and propose a log-normal approximation. Additionally, we investigate the viability of the model in terms of VIX Futures and options, and compare it to the approximation in [BFG16].

5.3.1 VIX Futures in the rough Bergomi model.

From now on, we fix a given maturity $T \geq 0$, and define the VIX at time T via the continuous-time monitoring formula

$$\mathrm{VIX}_T^2 := \mathbb{E}\left(\frac{1}{\Delta} \int_T^{T+\Delta} \mathrm{d}\langle X_s, X_s \rangle \mathrm{d}s \middle| \mathcal{F}_T \right),$$

where Δ is equal to 30 days. The risk-neutral formula for the VIX future $\mathfrak{V}_T$ with maturity T is then given by

$$\mathfrak{V}_T := \mathbb{E}\left(\mathrm{VIX}_T | \mathcal{F}_0\right) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left(\mathrm{d}\langle X_s, X_s \rangle | \mathcal{F}_T\right) \mathrm{d}s} \middle| \mathcal{F}_0 \right) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \xi_T(s) \mathrm{d}s} \middle| \mathcal{F}_0 \right). \tag{5.7}$$

Note that, when $T > 0$, $\xi_T(s)$ is a market input which is not $\mathcal{F}_0$-measurable, and is hence difficult to interpret only knowing $\mathcal{F}_0$. We shall make repeated use of the following random variable defined for any $t \geq T$, by

$$\eta_T(t) := \exp\left(2\nu C_H \mathcal{V}_t^T\right). \tag{5.8}$$

Proposition 5.3.1. *The VIX dynamics are given by*

$$\mathrm{VIX}_T = \left\{ \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0(t) \eta_T(t) \exp\left(\frac{\nu^2 C_H^2}{H} \left[(t-T)^{2H} - t^{2H} \right] \right) \mathrm{d}t \right\}^{1/2}.$$

Proof. Using Fubini's theorem and the instantaneous variance representation in (5.1), we can write

$$\begin{aligned} \mathrm{VIX}_T^2 &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left(V_s | \mathcal{F}_T\right) \mathrm{d}s = \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\Big[\xi_0(t) \mathcal{E}\left(2\nu C_H \mathcal{V}_t\right) | \mathcal{F}_T \Big] \mathrm{d}t \\ &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}\left[\xi_0(t) \eta_T(t) \exp\left(2\nu C_H \mathcal{V}_t^{[T,t]} - \frac{\nu^2 C_H^2 t^{2H}}{H} \right) \middle| \mathcal{F}_T \right] \mathrm{d}t, \end{aligned}$$

with $\mathcal{V}_t^{[T,t]}$ defined in (5.3). Since $\eta_T(t) \in \mathcal{F}_T$ and $\xi_0(t) \in \mathcal{F}_0$, this expression simplifies to

$$\mathrm{VIX}_T^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0(t) \eta_T(t) \mathbb{E}\left[\exp\left(2\nu C_H \mathcal{V}_t^{[T,t]} - \frac{\nu^2 C_H^2 t^{2H}}{H} \right) \middle| \mathcal{F}_T \right] \mathrm{d}t.$$

The proposition follows since $\mathcal{V}_t^{[T,t]}$ is centred Gaussian, independent of $\mathcal{F}_T$, with variance given in (5.4), and $\mathbb{E}\left(\exp\left\{ 2\nu C_H \mathcal{V}_t^{[T,t]} \right\} \middle| \mathcal{F}_T \right) = \mathbb{E}\left(\exp\left\{ 2\nu C_H \mathcal{V}_t^{[T,t]} \right\} \right) = \exp\left(\frac{\nu^2 C_H^2}{H} (t-T)^{2H} \right)$. □

The main challenge for simulation is $\eta_T(t)$. However, since the latter is independent of $\xi_0(\cdot)$, robustness of simulation schemes for the VIX will not be affected by the qualitative properties of the initial variance curve ξ_0. Since

$\mathbb{E}(V_t|\mathcal{F}_T) = \xi_T(t)$ by (5.7), the equality

$$\mathbb{E}(V_t|\mathcal{F}_T) = \xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right).$$

yields that the following representation for the forward variance curve ξ_T in the rough Bergomi model:

$$\xi_T(t) = \xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right), \qquad \text{for any } t \geq T. \tag{5.9}$$

Bayer, Friz and Gatheral [BFG16] did not derive such a representation, and their approach for pricing VIX derivatives relies on an approximation which avoids the computations developed in this section. The identity (5.9) allows for a better understanding of the process ξ_T, and for an innovative approach to price VIX derivatives.

5.3.2 Upper and lower bounds for VIX Futures

Theorem 5.3.2. *The following bounds hold for VIX Futures:*

$$\frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t \leq \mathfrak{V}_T \leq \left\{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s\right\}^{1/2}. \tag{5.10}$$

Proof. The conditional Jensen's inequality gives

$$\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) = \mathbb{E}\left(\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(s)\mathrm{d}s}\,\middle|\,\mathcal{F}_0\right) \leq \sqrt{\mathbb{E}\left(\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(s)\mathrm{d}s\,\middle|\,\mathcal{F}_0\right)}.$$

Furthermore, since ξ_0 is $\mathcal{F}_0$-adapted, Fubini's theorem along with the martingale property of ξ_T yield the upper bound $\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) \leq \sqrt{\Delta^{-1}\int_T^{T+\Delta}\xi_0(s)\mathrm{d}s}$. To obtain a lower bound we use the representation in (5.9), and Cauchy-Schwarz's inequality, and Fubini's theorem, so that

$$\begin{aligned}
\mathfrak{V}_T = \mathbb{E}(\mathrm{VIX}_T|\mathcal{F}_0) &= \mathbb{E}\left[\sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(t)\eta_T(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t}\,\middle|\,\mathcal{F}_0\right] \\
&\geq \mathbb{E}\left[\frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)\eta_T(t)}\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t\,\middle|\,\mathcal{F}_0\right] \\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\mathbb{E}\left(\sqrt{\eta_T(t)}\right)\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t \\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[t^{2H} - (t-T)^{2H}\right]\right)\exp\left(\frac{\nu^2 C_H^2}{2H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t \\
&= \frac{1}{\Delta}\int_T^{T+\Delta}\sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H} - t^{2H}\right]\right)\mathrm{d}t,
\end{aligned}$$

since $\eta_T(t)^{1/2}$ is log-normal (Proposition 5.3.1), so that $\mathbb{E}(\sqrt{\eta_T(t)}) = \exp\left(\frac{\nu^2 C_H^2}{4H}\left[t^{2H} - (t-T)^{2H}\right]\right)$. □

We perform a numerical experiment to check the tightness of the bounds obtained in Proposition 5.3.2). For this analysis we consider three qualitative scenarios for the initial forward variance curve:

$$\text{Scenario 1}: \ \xi_0(t) = 0.234^2; \quad \text{Scenario 2}: \ \xi_0(t) = 0.234^2(1+t)^2; \quad \text{Scenario 3}: \ \xi_0(t) = 0.234^2\sqrt{1+t}. \tag{5.11}$$

Figures 5.1 suggest that the lower bound given in Proposition 5.3.2) is surprisingly tight for very different shapes of ξ_0. This can be explained with the following argument: consider a simplified and deterministic version of the

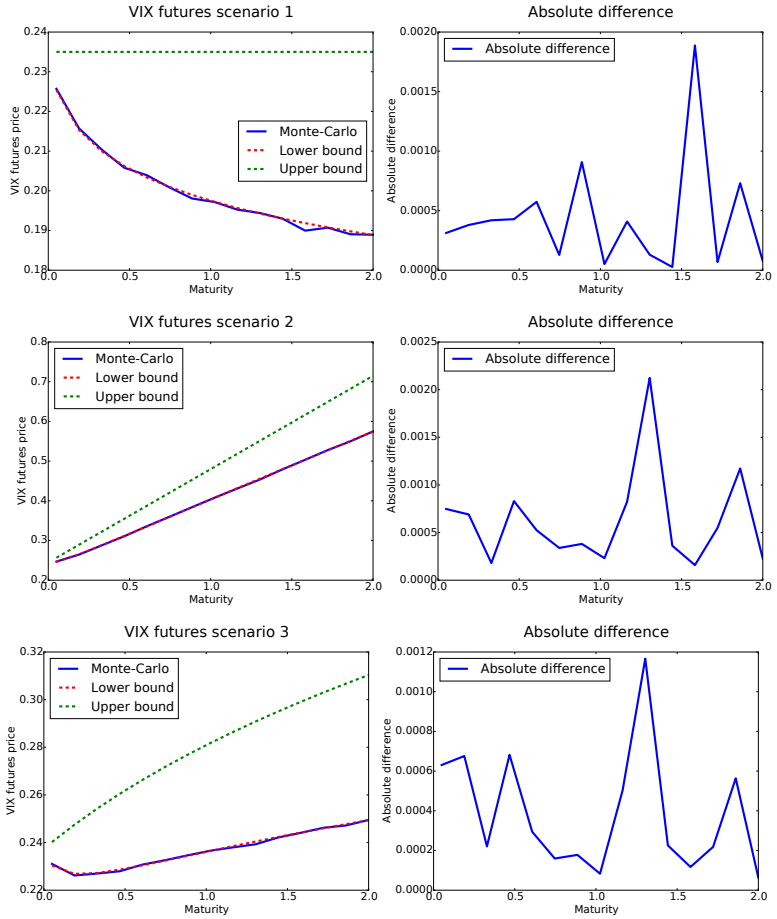


Figure 5.1: Bounds vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

VIX futures price in Proposition 5.3.1), denoted by

$$\phi(T) := \sqrt{\frac{1}{\Delta}\int_T^{T+\Delta} f(t)\mathrm{d}t} = \sqrt{f(T) + \frac{\Delta}{2}f'(T) + \frac{\Delta^2}{6}f''(T) + \mathcal{O}(\Delta^3)},$$

for some strictly positive (deterministic) function $f \in \mathcal{C}^2(\mathbb{R})$. We further introduce

$$\psi(T) := \frac{1}{\Delta}\int_T^{T+\Delta} \sqrt{f(t)}\mathrm{d}t = \sqrt{f(T)} + \frac{\Delta}{4}\frac{f'(T)}{\sqrt{f(T)}} + \mathcal{O}(\Delta^2),$$

which is the corresponding lower bound by Cauchy-Schwarz's (or Jensen's) inequality, so that

$$\phi(T)^2 - \psi(T)^2 = \Delta^2\left(\frac{f''(T)}{6} - \frac{f'(T)^2}{16f(T)}\right) + \mathcal{O}(\Delta^3). \tag{5.12}$$

Hence, we observe that for small Δ, as it is the case in VIX futures, the lower bound is very close from the original value which explains (at least for the deterministic case) the behaviour observed in Figure 5.1.

Remark 5.3.3. Recall that $f(t) = \sqrt{\xi_0(t)}\exp\left(\frac{\nu^2 C_H^2}{4H}\left[(t-T)^{2H} - t^{2H}\right]\right) > 0$ has a financial interpretation and cannot be arbitrarily small. Therefore, the term $\frac{f'(T)^2}{f(T)}$ in (5.12) does not explode. The use of the lower bound as a proxy for the price allows one, in principle, to invert the price and calibrate ξ_0 consistent with the VIX as a functional of H, ν and the current term structure of VIX Futures. This approach is left for future research.

5.3.3 Numerical implementation of VIX process

In this section, we investigate different simulation schemes for the VIX in the rough Bergomi model.

Hybrid scheme and forward Euler approach

In order to simulate the process $(\mathcal{V}_t^T)_{t\in[T,T+\Delta]}$ it is important to notice from (5.3) that the kernel has a singularity only for $\mathcal{V}_T^T$, hence we may overcome this by simulating the process using the hybrid scheme from [BLP17]. Then, we may easily extract Z and simulate $(\mathcal{V}_t^T)_{t\in(T,T+\Delta]}$ using the forward Euler scheme with complexity $\mathcal{O}(n)$. A forward Euler scheme is chosen in this case due to the fact that the kernel in (5.3) no longer has a singularity for $t \in (T, T+\Delta]$ and this method is faster than the hybrid scheme. Once $(\mathcal{V}_t^T)_{t\in[T,T+\Delta]}$ is simulated, numerical integration routines may be used to simulate the VIX process using the expression in Proposition 5.3.1. It must be pointed out that this approach is computationally expensive and memory consuming since it involves to simulate the Volterra process using the hybrid scheme with complexity of $\mathcal{O}(n \log n)$ and additionally, each $\mathcal{V}_t^T$ by forward Euler. The simulation algorithm can be summarised as follows:

Algorithm 5.3.4 (VIX simulation in the rough Bergomi model)**.** Fix a grid $\mathcal{T} = \{t_i\}_{i=0,\dots,n_T}$ and $\kappa \geq 1$.

1. Simulate the Volterra process $(\mathcal{V}_t)_{t\in[0,T]}$ using the hybrid scheme [BLP17], yielding a sample of the random variable $\mathcal{V}_T^T = \mathcal{V}_T$;

2. extract the path of the Brownian motion Z driving the Volterra process.

$$\begin{aligned} Z_{t_i} &= Z_{t_{i-1}} + n^{H-}\left(\mathcal{V}(t_i) - \mathcal{V}(t_{i-1})\right), && \text{for } i = 1,\dots,\kappa, \\ Z_{t_i} &= Z_{t_{i-1}} + \overline{Z}_{i-1}, && \text{for } i > \kappa; \end{aligned}$$

3. fix a grid $\mathfrak{T} = \{\tau_j\}_{j=0,\dots,N}$ on $[T, T+\Delta]$ and approximate the continuous-time process $\mathcal{V}^T$ by the discrete-time version $\widetilde{\mathcal{V}}^T$ defined via the following forward Euler scheme:

$$\widetilde{\mathcal{V}}_{\tau_0}^T := \mathcal{V}_T^T \qquad \text{and} \qquad \widetilde{\mathcal{V}}_{\tau_j}^T := \sum_{i=1}^{n_T} \frac{Z_{t_i} - Z_{t_{i-1}}}{(\tau_j - t_{i-1})^{-H_-}}, \qquad \text{for } j = 1,\dots,N;$$

4. compute the VIX process via numerical integration, for example using a composite trapezoidal rule:

$$\text{VIX}_T \approx \left\{ \frac{1}{\Delta} \sum_{j=0}^{N-1} \frac{Q_{T,\tau_j}^2 + Q_{T,\tau_{j+1}}^2}{2} (\tau_j - \tau_{j-1}) \right\}^{1/2},$$

where $Q_{T,\tau_j}^2 := \xi_0(\tau_j) \exp\left(2vC_H \widetilde{\mathcal{V}}_{\tau_j}^T\right) \exp\left(\frac{\nu^2 C_H^2}{H}\left((\tau_j - T)^{2H} - \tau_j^{2H}\right)\right)$.

Remark 5.3.5. Step 4 may obviously be replaced by any available numerical integration routine, but one must then carefully choose the partition in Step 3.

Remark 5.3.6. In order to analyse the convergence of the scheme, consider equidistant grids $\mathcal{T}$ and $\mathfrak{T}$ with increment sizes $h_{\mathcal{T}}$ and $h_{\mathfrak{T}}$ respectively. In the first three steps the Hybrid and forward Euler schemes yield an error of order $\mathcal{O}(h_{\mathcal{T}})$, hence we have $\mathcal{V}^T = \widetilde{\mathcal{V}}^T + \mathcal{O}(h_{\mathcal{T}})$. In addition, the trapezoidal rule [Atk89, Chapter 5] in step 4 yields an error of order $\mathcal{O}(h_{\mathfrak{T}}^2)$. Therefore the total error of convergence is of order $\mathcal{O}(h_{\mathcal{T}})$.

Truncated Cholesky approach

Alternatively, one could use the more expensive, yet exact, Cholesky decomposition to simulate $\mathcal{V}^T$ on $[T, T+\Delta]$ since its covariance structure is known from (5.5). However, computational complexity aside, with the same grid $\mathcal{T}$ as in Algorithm 5.3.4, numerical experiments suggest that the determinant of the covariance matrix is equal to zero when using more than $n_T = 8$ discretisation points. Hence, although valid in theory, the Cholesky approach is not feasible numerically. Strictly, speaking the Cholesky decomposition does not compute the determinant of the matrix, but it's stability is related to such quantity. This fact implies that there exists strong linear dependence. In fact, for any $\varepsilon > 0$, the strict inequality $\text{corr}(\mathcal{V}_{t_1}^T, \mathcal{V}_{t_1+\varepsilon}^T) < \text{corr}(\mathcal{V}_{t_2}^T, \mathcal{V}_{t_2+\varepsilon}^T)$ holds for all $T < t_1 < t_2$, as well as the following:

Proposition 5.3.7. *The limit* $\lim_{\varepsilon \downarrow 0} corr\left(\mathcal{V}_t^T, \mathcal{V}_{t+\varepsilon}^T\right) = 1$ *holds for any* $t \in [T, T+\Delta]$.

Proof. This follows readily from the continuity in L^2 of the map $t \mapsto \mathcal{V}_t^T$:

$$\begin{aligned} E\left[\left(\mathcal{V}_{t+\varepsilon}^T - \mathcal{V}_t^T\right)^2\right] &= \int_0^T [(t+\varepsilon-u)^{H_-} - (t-u)^{H_-}]^2 \mathrm{d}u \\ &= \int_0^T \left[(t+\varepsilon-u)^{2H_-} - 2[(t+\varepsilon-u)(t-u)]^{H_-} + (t-u)^{2H_-}\right] \mathrm{d}u. \end{aligned}$$

By monotone convergence, the sum of the first two terms converges to the third one, and the result follows. □

In light of Proposition 5.3.7, we model exactly the dependence structure on the first 8 grid points $t_1, \ldots, t_8$, then truncate the Cholesky decomposition and compute the correlations $\rho_i := \mathrm{corr}(\mathcal{V}_{t_{8+i}}^T, \mathcal{V}_{t_{8+i+1}}^T)$, for $i = 0, \ldots, n-9$ to approximate the process by adequately rescaling and correlating each pair of subsequent grid points. In contrast to the hybrid + forward Euler scheme, the computational complexity is much lower, since the Cholesky method is truncated with only 8 components. The VIX simulation algorithm therefore reads as follows:

Algorithm 5.3.8 (VIX simulation (truncated Cholesky))**.** Fix a grid $\mathfrak{T} = \{\tau_j\}_{j=1,\ldots,N}$ on $[T, T+\Delta]$,

(i) compute the covariance matrix of $(\mathcal{V}_{\tau_j}^T)_{j=1,\ldots,8}$ using (5.5) and simulate the process using the Cholesky method;

(ii) generate $\{\mathcal{V}_{\tau_j}^T\}_{j=1,\ldots,N}$ by correlating and rescaling using (5.5):

$$\mathcal{V}_{\tau_j}^T = \sqrt{\mathbb{V}(\mathcal{V}_{\tau_j}^T)} \left(\frac{\rho(\mathcal{V}_{\tau_{j-1}}^T, \mathcal{V}_{\tau_j}^T)\mathcal{V}_{\tau_{j-1}}^T}{\sqrt{\mathbb{V}(\mathcal{V}_{\tau_{j-1}}^T)}} + \sqrt{1 - \rho(\mathcal{V}_{\tau_{j-1}}^T, \mathcal{V}_{\tau_j}^T)^2}\mathcal{N}(0,1) \right), \quad \text{for } j = 9, \ldots, N;$$

(iii) compute the VIX via numerical integration as in Algorithm 5.3.4(4).

Singular Value Decomposition approach

As the numerical considerations in the previous section indicated, all Gaussian simulation schemes based on the computation of the determinant (including Cholesky) are not feasible to simulate the process $\mathcal{V}^T$. This is due to the fact that the driving Brownian motion in the stochastic integral representation of $\left(\mathcal{V}_t^T\right)_{0 \le t \le T}$ is always the same for all t. In addition, since $\mathcal{V}^T$ and its increments are non-stationary, many 'exact' methods (such as circulant embedding, Fourier methods or wavelets) are not applicable. This however can be bypassed by using a Singular Value Decomposition (SVD) in order to compute the square root of the covariance matrix.

Algorithm 5.3.9 (VIX simulation using SVD)**.** Fix a grid $\mathfrak{T} = \{\tau_j\}_{j=1,\ldots,N}$ on $[T, T+\Delta]$,

(i) compute the covariance matrix of $(\mathcal{V}_{\tau_j}^T)_{j=1,\ldots,N}$ using (5.5);

(ii) perform the SVD decomposition of the covariance matrix obtaining $U\Lambda U^*$, where U is a unitary matrix and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_N)$ a diagonal matrix of positive singular values;

(iii) draw a sample of a standard Gaussian $Z \sim \mathcal{N}(0, I_N)$ and generate

$$\mathcal{V}^T \sim U\Lambda^{1/2}U^*Z,$$

where $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_N})$;

(iv) compute the VIX by numerical integration as in Algorithm 5.3.4(4).

Remark 5.3.10. The complexity of the SVD decomposition [GL96, Algorithm 8.6.2] is $\mathcal{O}(N^3)$.

Remark 5.3.11. Recall that the simulation of $\mathcal{V}^T$ using SVD is exact, since U being unitary implies $U\Lambda^{1/2}U^*Z \sim \mathcal{N}(0, C)$ where $C = U\Lambda^{1/2}U^*U\Lambda^{1/2}U^* = U\Lambda U^*$. However, the numerical integration in step 4 produces an overall error of order $\mathcal{O}(h_{\mathfrak{T}}^2)$ where $h_{\mathfrak{T}} = \max_{1<j<N-1}(\tau_{j+1} - \tau_j)$.

Numerical experiment

We compute the price of VIX Futures using the simulation algorithm introduced in the previous section. We set the same parameters as in [BFG16] and [BLP17]:

$$\xi_0 = 0.235^2, \qquad H = 0.07, \qquad \nu = 1.9\frac{C_H\sqrt{2H}}{2} \approx 1.2287, \qquad \kappa = 2. \tag{5.13}$$

We perform 10^6 simulations for all three methods, with 300 equidistant points for $\mathcal{T}$ and, in the case of the HS+FE, 500 equidistant points for $\mathfrak{T}$. Figure 5.2 suggests that all methods agree qualitatively and converge to a similar output. Figure 5.2 also indicates that the Monte-Carlo standard deviation increases in T for all schemes. Figure 5.3 shows that a large number of simulations is needed in all cases to obtain reliable output. On the other hand, Figure 5.3 shows that the HS+FE method is much slower than the other two methods. Also, the truncated Cholesky method is 2.5 times faster than the SVD, which is consistent with the computational complexities discussed in the previous section. In light of this analysis, all three methods seem to approximate the prices similarly well. The computational time, however, is still too slow for reasonable calibration, and we shall only use the truncated Cholesky approach as a benchmark for the approximations in the next sections.

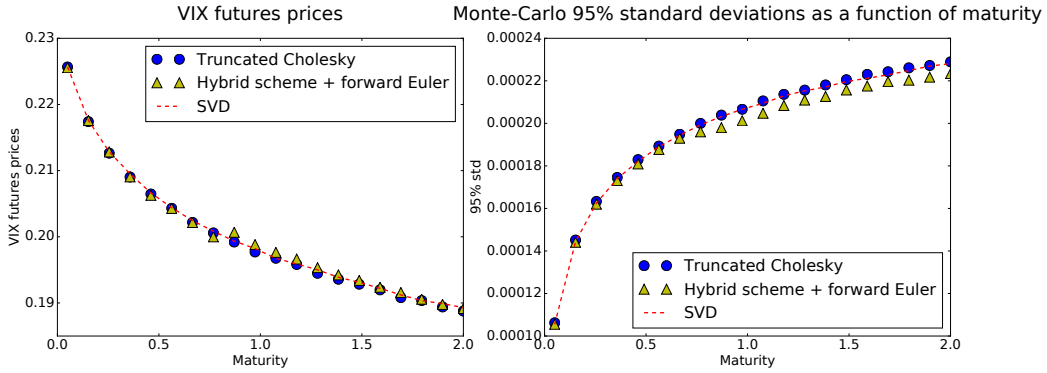


Figure 5.2: Monte-Carlo benchmarking for all three methods with 10^6 simulations and the corresponding 95% confidence level standard deviations

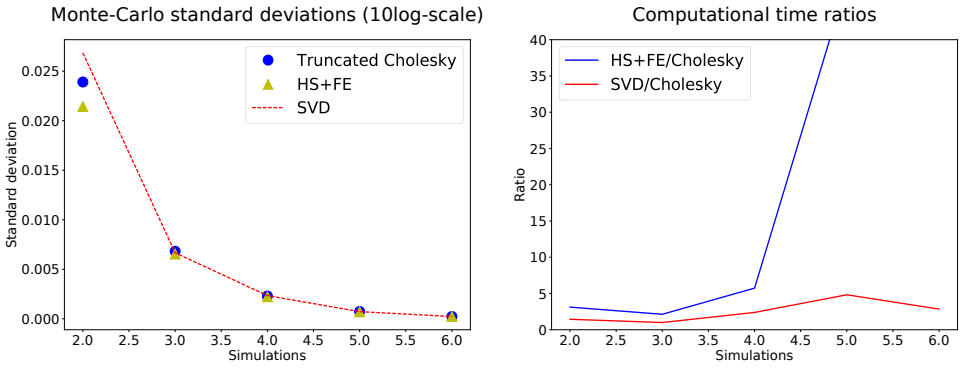


Figure 5.3: Monte-Carlo standard deviations and computational time ratios for a fixed maturity ($T = 2$ years) for all methods using efficient parallel computing.

5.3.4 The VIX process and log-normal approximations

We now investigate approximate methods to price VIX futures and options. We define the $\mathcal{F}_T$-measurable random variable $\Delta\text{VIX}_T^2 = \int_T^{T+\Delta} \xi_T(t)\mathrm{d}t$. In [BFG16], Bayer, Friz and Gatheral assumed that $\log(\Delta\text{VIX}_T^2)$ follows a Gaussian distribution, and computed directly its first and second moments, which hence fully characterise the distribution of ΔVIX_T^2. However, since the sum of log-normal random variables is not (in general) log-normal, the fact that ξ_T is log-normal does not imply that ΔVIX_T^2 is. In a different context (geometric Brownian motion

and Asian options), Dufresne [Duf04] proved that, under certain conditions, an integral of log-normal variables asymptotically converges to a log-normal. This approximation has been widely used for many applications [Duf04], and Dufresne's result motivates Bayer-Friz-Gatheral's assumption. We provide here exact formulae for the mean and variance of this distribution, and compare them numerically to those by Bayer-Friz-Gatheral.

Proposition 5.3.12. *The following holds:*

$$\sigma^2 := \mathbb{V}(\log(\Delta \mathrm{VIX}_T^2)) = -2\log \mathbb{E}(\Delta \mathrm{VIX}_T^2) + \log \mathbb{E}\left((\Delta \mathrm{VIX}_T^2)^2\right),$$
$$\mu := \mathbb{E}(\log(\Delta \mathrm{VIX}_T^2)) = \log \mathbb{E}(\Delta \mathrm{VIX}_T^2) - \frac{\sigma^2}{2}.$$

Furthermore, $\mathbb{E}(\Delta \mathrm{VIX}_T^2) = \int_T^{T+\Delta} \xi_0(t)\mathrm{d}t$ *and*

$$\mathbb{E}\left[(\Delta \mathrm{VIX}_T^2)^2\right] = \int_{[T,T+\Delta]^2} \xi_0(u)\xi_0(t) \exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H}\right]\right\} \mathrm{e}^{\overline{\Theta}_{u,t}} \mathrm{d}u\mathrm{d}t. \tag{5.14}$$

where $\overline{\Theta}_{u,t}$ *is equal to zero if* $u = t$ *and otherwise equal to* $\Theta_{u\vee t, u\wedge t}$*, where*

$$\Theta_{u,t} := 2\nu^2 C_H^2 \left\{\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + 2\frac{(u-t)^{H_-}}{H_+}\left[t^{H_+}\mathrm{F}\left(\frac{-t}{u-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{u-t}\right)\right]\right\}.$$

Remark 5.3.13. Since all the integrals in the proposition are computed over compact intervals, they are finite as long as ξ_0 is well behaved on $[T, T+\Delta]$. Assuming this is indeed the case is not restrictive in practice as ξ_0 represents the initial forward variance curve; note in particular that continuity of ξ_0 is sufficient.

Remark 5.3.14. The reason why $\overline{\Theta}_{u,t}$ is defined that way is for numerical purposes. Indeed, when $u < t$, $\Theta_{u,t}$ is not well defined since $\mathrm{F}(x)$ only makes sense when $x \leq 1$ (details on the radius of convergence of hypergeometric functions can be found in [ASR88][Page 556]), even though the integral representation (5.14) is still well defined. However, most numerical packages implement hypergeometric functions via series expansions. The trick from $\Theta_{u,t}$ to $\overline{\Theta}_{u,t}$ allows us to bypass this issue.

Proof of Proposition 5.3.12. The expectation follows directly from the tower property and Fubini's theorem. For the second moment, we use the decomposition

$$\mathbb{E}\left((\Delta \mathrm{VIX}_T^2)^2\right) = \mathbb{E}\left(\int_T^{T+\Delta}\int_T^{T+\Delta} \xi_T(u)\xi_T(t)\mathrm{d}t\mathrm{d}u\right) = \int_T^{T+\Delta}\int_T^{T+\Delta} \mathbb{E}\left(\xi_T(u)\xi_T(t)\right)\mathrm{d}t\mathrm{d}u$$

where in the last step we use that $\xi_T(s)$ is $\mathcal{F}_T$-measurable in order to apply Fubini. Using the representation obtained in (5.9), we get

$$\mathbb{E}\left(\xi_T(u)\xi_T(t)\right) = \xi_0(u)\xi_0(t)\exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H} - u^{2H} + (t-T)^{2H} - t^{2H}\right]\right\}\mathbb{E}\left(\mathrm{e}^{\vartheta_{u,t}}\right), \tag{5.15}$$

where the random variable

$$\vartheta_{u,t} := 2\nu C_H \int_0^T \left[(u-s)^{H_-} + (t-s)^{H_-}\right]\mathrm{d}Z_s$$

is Gaussian with zero expectation and variance

$$\begin{aligned}\mathbb{V}(\vartheta_{u,t}) &= 4\nu^2 C_H^2 \left(\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + 2\int_0^T (u-s)^{H_-}(t-s)^{H_-}\mathrm{d}s\right)\\ &= 4\nu^2 C_H^2 \left\{\frac{u^{2H} - (u-T)^{2H} + t^{2H} - (t-T)^{2H}}{2H} + \frac{2(u-t)^{H_-}}{H_+}\left[t^{H_+}\mathrm{F}\left(\frac{-t}{u-t}\right) - (t-T)^{H_+}\mathrm{F}\left(\frac{T-t}{u-t}\right)\right]\right\}.\end{aligned}$$

Then,

$$\mathbb{E}\left(\xi_T(u)\xi_T(t)\right) = \xi_0(u)\xi_0(t)\exp\left(\frac{\nu^2 C_H^2}{H}\left((u-T)^{2H} + (t-T)^{2H} - u^{2H} - t^{2H}\right)\right)\exp\left(\frac{\mathbb{V}(\vartheta_{u,t})}{2}\right),$$

and the second moment follows from the immediate computation

$$\begin{aligned}\mathbb{E}\left((\Delta \mathrm{VIX}_T)^2\right) &= \int_T^{T+\Delta}\int_T^{T+\Delta} \mathbb{E}\left[\xi_T(u)\xi_T(t)\right] \mathrm{d}u\mathrm{d}t \\ &= \int_T^{T+\Delta}\int_T^{T+\Delta} \xi_0(u)\xi_0(t)\exp\left\{\frac{\nu^2 C_H^2}{H}\left[(u-T)^{2H}+(t-T)^{2H}-u^{2H}-t^{2H}\right]\right\} \mathrm{e}^{\frac{1}{2}\mathbb{V}(\vartheta_{u,t})}\mathrm{d}u\mathrm{d}t,\end{aligned}$$

after defining $\Theta_{u,t} := \exp\left(\frac{1}{2}\mathbb{V}(\vartheta_{u,t})\right)$. □

In order to provide closed-form expressions for VIX Futures and options, we consider the following approximation:

Approximation 5.3.15. $\log(\Delta \mathrm{VIX}_T^2)$ is approximately Gaussian with mean μ and σ^2.

In fact, this is almost the same as in [BFG16]; however, they did not compute the variance exactly as in Proposition 5.3.12, and instead considered the lognormal approximation

Approximation 5.3.16 (Bayer-Friz-Gatheral)**.** $\log(\Delta \mathrm{VIX}_T^2)$ is approximately Gaussian with mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ given by

$$\tilde{\sigma}^2 = \frac{4\nu^2 C_H^2}{\Delta^2 H_+^2}\int_0^T \left[(T-s+\Delta)^{H_+}-(T-s)^{H_+}\right]^2 \mathrm{d}s \qquad \text{and} \qquad \tilde{\mu} = \log\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t - \frac{\tilde{\sigma}^2}{2}.$$

Lemma 5.3.17. *A VIX future is worth*

$$\mathfrak{V}_T = \begin{cases} \sqrt{\dfrac{1}{\Delta}\displaystyle\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t}\exp\left(-\dfrac{\sigma^2}{8}\right), & \textit{under Approximation 5.3.15,} \\ \sqrt{\dfrac{1}{\Delta}\displaystyle\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t}\exp\left(-\dfrac{\tilde{\sigma}^2}{8}\right), & \textit{under Approximation 5.3.16.}\end{cases}$$

Proof. Since $\Delta^{1/2}\mathrm{VIX}_T \overset{\Delta}{=} \exp\left(\frac{\mu+\sigma\mathcal{N}(0,1)}{2}\right)$, the price of a VIX future directly reads

$$\mathbb{E}\left(\mathrm{VIX}_T|\mathcal{F}_0\right) = \Delta^{-1/2}\mathbb{E}\left(\Delta^{1/2}\mathrm{VIX}_T\right) = \Delta^{-1/2}\exp\left(\frac{\mu}{2}+\frac{\sigma^2}{8}\right),$$

and the second case follows analogously. □

As opposed to the simulation schemes, the log-normal approximation does depend on the qualitative properties of the initial forward variance curve $\xi_0(\cdot)$. Hence, different curves should be analysed to check the robustness of the method. We now exploit Approximation 5.3.15 to obtain a closed-form Black-Scholes type formulae for European options on the VIX.

Lemma 5.3.18. *For $0 \le t \le T$, let $\mathfrak{V}_T(t) := \mathbb{E}\left(\mathrm{VIX}_T|\mathcal{F}_t\right)$ denote the price at time t of a VIX future maturing at T. Then, under Approximation 5.3.15, a European Call option on a VIX future maturing at T is worth*

$$\mathcal{C}_T^{\mathcal{V}} := \mathbb{E}\left[(\mathfrak{V}_T(T)-K)_+|\mathcal{F}_0\right] = \Delta^{-1/2}\sqrt{\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t}\exp\left(-\frac{\sigma^2}{8}\right)\Phi(d_1)-K\Phi(d_2),$$

where $\widetilde{K} := [\log(K^2\Delta)-\log\int_T^{T+\Delta}\xi_0(t)\mathrm{d}t+\sigma^2/2]/\sigma$, $d_1 := -\widetilde{K}+\frac{1}{2}\sigma$ and $d_2 := -\widetilde{K}$, with σ^2 as in Proposition 5.3.12

Proof. Under Approximation 5.3.15, the lemma follows directly from the following trivial computations:

$$\mathbb{E}\left(\mathfrak{V}_T(T)-K\right)_+ = \int_{\widetilde{K}}^{\infty}\left[\Delta^{-1/2}\exp\left(\frac{\mu}{2}+\frac{\sigma z}{2}\right)-K\right]_+\phi(z)\mathrm{d}z = \frac{1}{\sqrt{\Delta}}\exp\left(\frac{\mu}{2}+\frac{\sigma^2}{8}\right)\int_{\widetilde{K}}^{\infty}\phi\left(z-\frac{\sigma}{2}\right)\mathrm{d}z - K\Phi(d_2).$$

□

Numerical tests of the log-normal approximation on VIX Futures

We perform a numerical analysis of the approximation by pricing VIX Futures using the parameters in (5.13). We consider 10^6 simulations and the three qualitative scenarios introduced in (5.11) for the forward variance curve. Figures 5.4 suggest that the approximation is accurate being the difference of order 10^{-3} or less. The oscillating

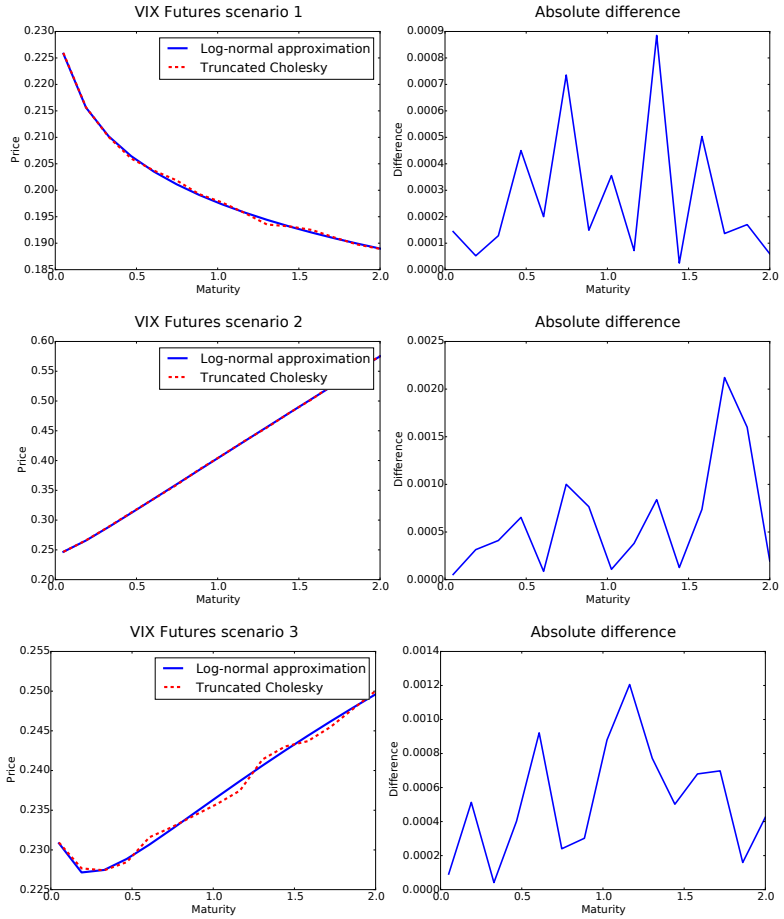


Figure 5.4: Log-normal approximations vs. Monte Carlo (Truncated Cholesky) in all three scenarios.

nature of the difference is also a good sign since it is probably caused by the Monte Carlo error and does not show any monotonicity. Furthermore, the method shows to be robust for different type of curves $\xi_0(\cdot)$. Moreover, the log-normal approximation seems to converge to the true mean, avoiding the oscillations of the Truncated Cholesky method. Therefore, we may conclude that the log-normal approximation produces a good output, with desired smoothness properties. On the other hand, we also analyse European VIX Call options repeating the previous parameters and considering the flat forward variance curve from Scenario 1. Figure 5.5 shows that the log-normal

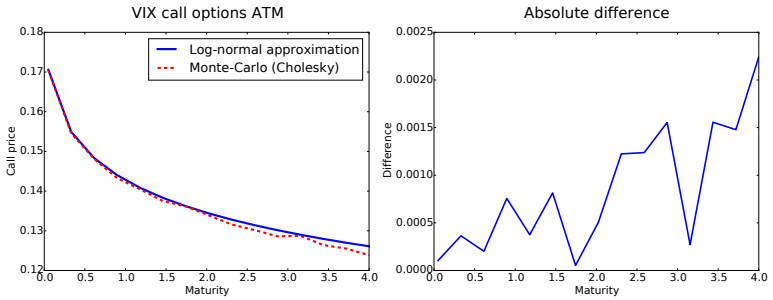


Figure 5.5: Log-normal approximation vs. Monte Carlo (Truncated Cholesky) with 10^6 simulations

approximation is accurate for at-the-money options. Nevertheless, the difference increases with maturity, and hence one must be careful when using this approximation to price options with long maturities. However, in practice, on

Equity markets, liquid maturities are only up to two years.

Numerical tests

We benchmark Bayer-Friz-Gatheral's approximation against ours with the parameters in (5.13). Figures 5.6 and 5.7 suggest that both are similar. In particular, we observe that the variance deviates as maturity increases. Nevertheless, for practical purposes, both approximations agree over a four-year horizon, long enough to cover available Futures data. Many functional forms of $\xi_0(t)$ were also tested giving similar results as the ones shown in Figures 5.6 and 5.7. We conclude that the approximation by Bayer, Friz and Gatheral is accurate enough for practical purposes and yields and reduces significantly the computational costs since the computation of $\tilde{\sigma}^2$ involves a single integral while our approximation requires a double integral (for σ^2). In particular, in order to generate Figure 5.6 our approximation was 40 times slower than Bayer, Friz and Gatheral's.

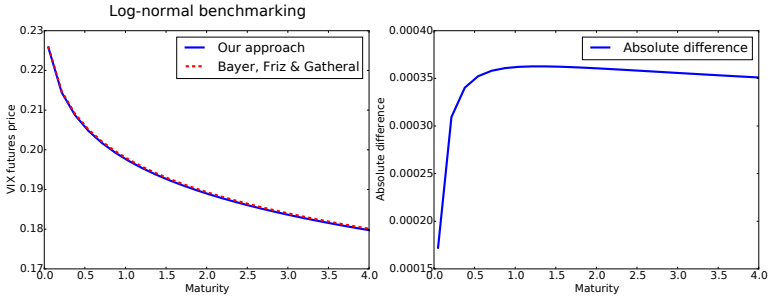


Figure 5.6: VIX Futures prices following Approximation 5.3.15 and Approximation 5.3.16.

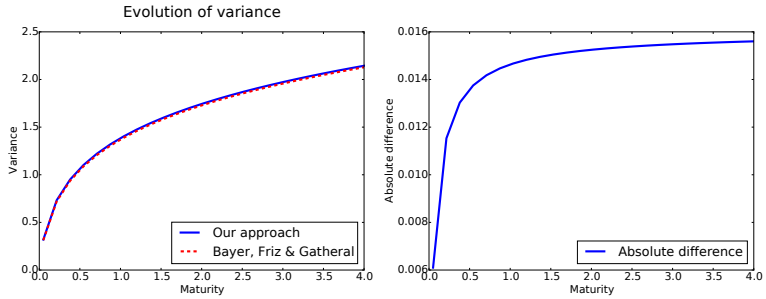


Figure 5.7: Comparison of the variance following Approximation 5.3.15 (σ^2) and Approximation 5.3.16 ($\tilde{\sigma}^2$).

5.3.5 VIX Futures calibration in rBergormi

In light of the promising results obtained in the previous sections, we create a calibration algorithm based on the log-normal approximation by Bayer, Friz and Gatheral [BFG16]. Even if our approximation seems to be more accurate the computational cost is much larger and the difference between both approaches has been shown to be very small. Moreover, the approach by Bayer, Friz and Gatheral allows to compute the gradient of the objective function in a semi-explicit form, which in terms of optimisation and calibration is extremely useful.

Objective function

To calibrate the model to VIX Futures, we first define the objective function

$$\mathcal{L}^{\mathfrak{F}}(\nu, H) := \sum_{i=1}^{N} (\mathfrak{V}_{T_i} - \mathfrak{F}_i)^2, \tag{5.16}$$

which we minimise over (ν, H). Here $(\mathfrak{F}_i)_{i=1,\dots,N}$ are the observed Futures prices on the time grid $T_1 < \ldots < T_N$, $\mathfrak{V}_{T_i} = \Delta^{-1/2}\sqrt{\int_{T_i}^{T_i+\Delta} \xi_0(t)\mathrm{d}t}\exp\left(-\frac{\tilde{\sigma}_i^2}{8}\right)$. and

$$\tilde{\sigma}_i^2 = \frac{4\nu^2 C_H^2}{H_+^2\Delta^2}\left[\frac{(T_i+\Delta)^{1+2H_+} - \Delta^{1+2H_+} + T_i^{1+2H_+}}{1+2H_+} - 2\frac{T_i^{1+H_+}\Delta^{H_+}}{1+H_+}{}_2F_1\left(-H_+, 1+H_+, 2+H_+, -\frac{T_i}{\Delta}\right)\right],$$

which is the closed-form expression of the variance in Approximation 5.3.16. The gradient of the objective function is an important source of information in many optimisation algorithms. To compute it, we differentiate the objective function in (5.16) with respect to ν and H and apply the chain rule:

$$\frac{\partial \mathcal{L}^{\mathfrak{F}}}{\partial \nu}(\nu, H) = -\frac{1}{4}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\frac{\partial \tilde{\sigma}_i^2}{\partial \nu} = -\frac{1}{2\nu}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\tilde{\sigma}_i^2,$$

where

$$\frac{\partial \tilde{\sigma}_i^2}{\partial \nu} = \frac{8\nu C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left((T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right)^2 \mathrm{d}s = \frac{2\tilde{\sigma}_i^2}{\nu}.$$

On the other hand, $\frac{\partial \mathcal{L}^{\mathfrak{F}}}{\partial H}(\nu, H) = -\frac{1}{4}\sum_{i=1}^{N}(\mathfrak{V}_{T_i} - \mathfrak{F}_i)\mathfrak{V}_{T_i}\frac{\partial \tilde{\sigma}_i^2}{\partial H}$, with

$$\begin{aligned}
\frac{\partial \tilde{\sigma}_i^2}{\partial H} &= \frac{4\nu^2\frac{\partial C_H^2}{\partial H}H_+ - 8\nu^2 C_H^2}{\Delta^2 H_+^3}\int_0^{T_i}\left((T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right)^2 \mathrm{d}s \\
&\quad + \frac{8\nu^2 C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left[(T_i - s + \Delta)^{H+\frac{1}{2}} - (T_i - s)^{H_+}\right]\left[(T_i - s + \Delta)^{H_+}\log(T_i - s + \Delta) - (T_i - s)^{H_+}\log(T_i - s)\right]\mathrm{d}s \\
&= \frac{\tilde{\sigma}_i^2(K_H - 2)}{H_+} \\
&\quad + \frac{8\nu^2 C_H^2}{\Delta^2 H_+^2}\int_0^{T_i}\left[(T_i - s + \Delta)^{H_+} - (T_i - s)^{H_+}\right]\left[(T_i - s + \Delta)^{H_+}\log(T_i - s + \Delta) - (T_i - s)^{H_+}\log(T_i - s)\right]\mathrm{d}s,
\end{aligned}$$

where

$$\frac{\partial C_H^2}{\partial H} = \frac{2\Gamma(2 - H_+)}{\Gamma(H_+)\Gamma(1 - 2H_-)}\left\{1 + \left[2\psi(2 - H_+) - \psi(H_+) - \psi(1 - 2H_-)\right]H\right\}$$

and $K_H = \frac{H_+}{H}\left\{1 + H\psi\left(2 - H_+\right) - H\left(\psi(H_+) + \psi(1 - 2H_-)\right)\right\}$, where ψ is the digamma function.

Obtaining the initial forward variance curve

The initial forward variance curve plays a crucial role in both the Bergomi [Ber05] and the rough Bergomi models [BFG16], since it is a market input. In particular it depends on the current term structure of variance swaps. Even if variance swaps are not traded in standard exchange markets, they are actively traded over-the-counter (OTC). This in turn means that there is no observable data and we must establish a valuation method for variance swaps. The celebrated static replication formula by Carr and Madan [CM01], applicable here since the underlying process is a continuous semimartingale, allows us to price any variance swap. Nevertheless, out-of-money Call and Put option prices are needed for all possible strikes. Since this information is not available in practice, one can either adopt a model-free valuation formula or directly propose a parameterisation for the implied volatility surface. The first approach involves a discretisation of the static replication formula, which is how the VIX index is computed by the Chicago Board Options Exchange. It is not easy, however, to extend this computation for large maturities (VIX is a 30-day ahead index), where liquidity of options may play a major role. On the other hand, the second approach allows to calibrate a model using available data and additionally allows to interpolate / extrapolate available option data to all strikes and maturities. We follow here the latter approach, and consider a parametric form for the total implied variance. Gatheral and Jacquier [GJ14] originally proposed such a parameterisation (called SSVI), and studied absence of arbitrage and calibration to S&P options data. We use here the

eSSVI parameterisation [HM19] for the implied volatility surface, a refinement of SSVI parametrisation, relaxing the constant correlation assumption:

$$\sigma_{\mathrm{BS}}^2(t,k)t = w(t,k) := \frac{\theta_t}{2}\left\{1 + \rho(\theta_t)\varphi(\theta_t)k + \sqrt{(\varphi(\theta_t)k + \rho(\theta_t))^2 + 1 - \rho(\theta_t)^2}\right\}, \tag{5.17}$$

for all log-(forward) moneyness $k \in \mathbb{R}$, where $(\theta_t)_{t\geq 0}$ is the observed at-the-money variance curve, and where the shape function takes the form $\varphi(\theta) = \eta\theta^{-\lambda}(1+\theta)^{\lambda-1}$. For the correlation parameter $\rho(\cdot)$ we restrict it to the following functional form:

$$\rho(\theta) = (A-C)\mathrm{e}^{-B\theta} + C, \qquad \text{for } (A,C) \in (-1,1)^2, B \geq 0, \tag{5.18}$$

ensuring that $|\rho(\cdot)| \leq 1$. We shall indistinctly refer to the total implied variance by $\sigma_{\mathrm{BS}}^2(\cdot)t$ or $w(\cdot)$, where $\sigma_{\mathrm{BS}}(\cdot)$ represents the implied volatility. Gatheral and Jacquier [GJ14] found (sufficient and almost necessary) conditions on the parameters ρ, $\varphi(\cdot)$, η and λ preventing arbitrage. In the eSSVI formulation, the correlation has a term structure (5.18), and care must be taken in order to preclude arbitrage. Concretely, following [HM19], the restriction

$$|\theta\partial_\theta(\rho(\theta)) + \rho(\theta)\gamma| \leq \gamma, \tag{5.19}$$

where $\gamma := \partial_\theta(\theta\varphi(\theta))/\varphi(\theta)$, is a necessary condition to preclude calendar spread arbitrage. To prevent butterfly arbitrage, exactly as in the SSVI parameterisation, it is sufficient [GJ14, Theorem 4.2] to check that the inequality $\theta_t\varphi^2(\theta_t)(1+|\rho(\theta_t)|) \leq 4$ holds for all maturity t. In this parameterisation, variance swaps can be computed in closed form, as proved in [HM19], based on earlier works by Gatheral [Gat06] and Fukasawa [Fuk11b; Fuk12]:

Proposition 5.3.19. *The fair strike (in total variance) of a variance swap in the eSSVI model reads*

$$\sigma_0(t)^2 t := -2\mathbb{E}\log\left(\frac{S_t}{S_0}\right) = \frac{b_t^2 + 2a_t(c_t+\theta_t)}{2a_t^2},$$

where $\chi_t := \frac{1}{4}[1-\rho(\theta_t)^2]\theta_t\varphi(\theta_t)$, *and*

$$a_t = 1 + \frac{\theta_t\varphi(\theta_t)}{2}\left(\rho(\theta_t) - \frac{\chi_t}{2}\right), \qquad b_t = \theta_t\varphi(\theta_t)\left[\chi_t - \rho(\theta_t)\right], \qquad c_t = \theta_t\varphi(\theta_t)\chi_t.$$

Recalling the relation between variance swaps and the forward variance curve, we have

$$\xi_0(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(t\sigma_0^2(t)\right) = \sigma_0^2(t) + t\frac{d}{\mathrm{d}t}\sigma_0^2(t).$$

Remark 5.3.20. In order to interpolate/extrapolate the eSSVI, it is necessary to also interpolate/extrapolate θ_t for all t. However, θ_t is only observed on a discrete set of maturity dates, and consequently, cubic splines are used to interpolate/extrapolate all other maturities.

Calibration algorithm and numerical results

We first calibrate the eSSVI parameterisation (5.17) on the SPX implied volatility surface on the 4th of December 2015. Figure 5.8 shows the fit for the shortest, medium and longest maturities available in the data set. For short maturities the eSSVI is not able to fully capture the volatility smile, however as maturity increases the fit improves remarkably. For VIX Futures, the calibration algorithm reads as follows:

Algorithm 5.3.21 (VIX Futures calibration algorithm in the rough Bergomi model)**.**

(i) Calibrate eSSVI to available SPX option data;

(ii) compute the variance swap term structure $(\sigma_0(t)^2)_{t\geq 0}$ using Proposition 5.3.19;

(iii) extract the initial forward variance curve, $\xi_0(\cdot)$ via $\xi_0(t) \approx \sigma_0^2(t) + \frac{\sigma_0^2(t+\varepsilon)-\sigma_0^2(t-\varepsilon)}{2\varepsilon}t$ (with $\varepsilon = 1E-8$);

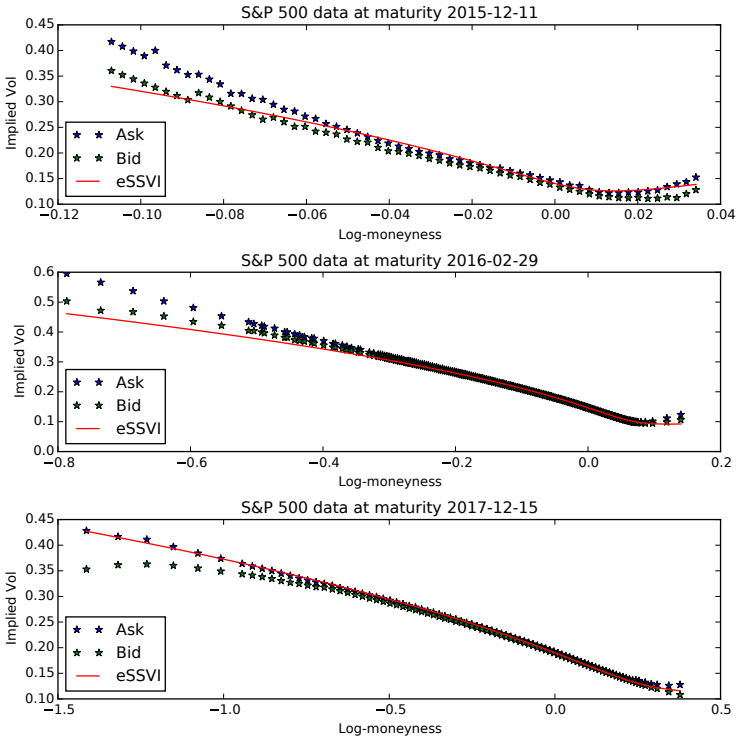


Figure 5.8: eSSVI calibration results on 4/12/2015 using traded SPX options.

(iv) minimise (over ν, H) the objective function in (5.16).

Figures 5.9-5.11 suggest that the model fits very well the observed VIX Futures term structure for different dates. Moreover, we notice that both the model and the observed data are qualitatively equal in terms of convexity/concavity. In the rough Bergomi model this information is obtained from option prices through ξ_0, which suggests a correspondence in the market between VIX futures and SPX options. However, we also observe that in all three cases the error is greater for short maturities, mimicking the calibration limits of eSSVI for short maturities, as detailed in Section 5.3.5.

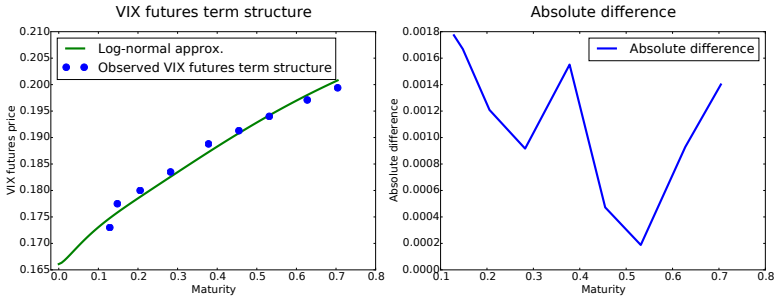


Figure 5.9: VIX Futures calibration on 4/12/2015. Optimal parameters: $(H, \nu) = (0.09237, 1.004)$.

Remark 5.3.22. The reader should recall the importance of the initial forward variance curve $(\xi_0(t))_{t\geq 0}$ in the VIX Futures process, since $(\mathfrak{V}_t)_{t\geq 0}$ depends on the whole path of ξ_0 up to time t. Therefore, even if ξ_0 is only misspecified for short maturities (as is the case of the eSSVI), this affects the whole term structure of the VIX process (for details we refer the reader to Section 5.3). Therefore, an improved ξ_0 estimation would not only increase the accuracy of the model for short maturities, but also for the whole term structure.

Remark 5.3.23. Our calibration involves two different data sets, SPX options and VIX Futures: the Vanilla quotes are extracted from the CBOE delayed option quotes page[2] and the VIX Futures from the CBOE VIX

[2]CBOE delayed option quotes: http://www.cboe.com/delayedquote/quotetable.aspx

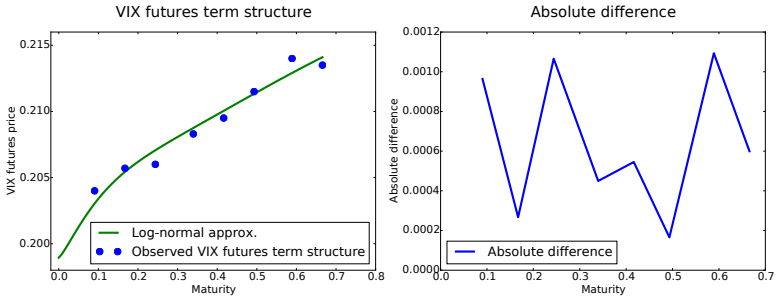


Figure 5.10: VIX Futures calibration on 22/2/2016. Optimal parameters: $(H, \nu) = (0.10093, 1.00282)$.

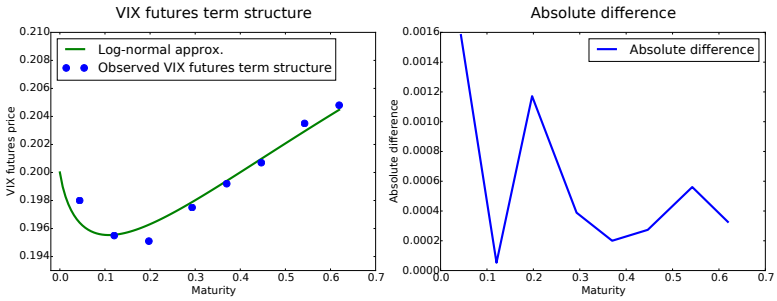


Figure 5.11: VIX Futures calibration on 4/1/2016. Optimal parameters: $(H, \nu) = (0.0509, 1.2937)$.

futures historical data page[3]. We perform an aggregation of the different Future quotes, since they are quoted on a Future per Future basis, and check the consistency between the two data sets by comparing the left end extrapolation of the Futures curve with the theoretical VIX computed from option prices.

5.4 From VIX Futures to SPX options

In this final chapter we assess whether the Hurst parameter H obtained through the VIX Futures calibration algorithm is consistent with SPX options. For this purpose, we calibrate the rough Bergomi model to SPX option data by fixing the parameter H and letting the algorithm calibrate ν and ρ. One of the main reasons to fix H is that the numerical simulation remarkably reduces its complexity from $\mathcal{O}(n \log n)$ to $\mathcal{O}(n)$, since the $\mathcal{O}(n \log n)$ complexity of the Volterra is computed only once and reused afterwards. Therefore, by fixing H the pricing scheme is much faster when several valuations are performed, as is the case of a calibration algorithm.

5.4.1 Pricing in the rough Bergomi model

We present a pricing scheme, where the Volterra process $\mathcal{V}$ is simulated using a hybrid scheme, while a standard Euler scheme generates the paths of the stock process:

Algorithm 5.4.1 (Simulation of the rough Bergomi model)**.** Consider the grid $\mathcal{T} := \{t_i\}_{i=0,\dots,n_T}$, and fix $\kappa \geq 1$.

(i) Simulate the Volterra process $\mathcal{V}$ on the grid $\mathcal{T}$ using the hybrid scheme;

(ii) simulate the variance process as $V_t = \xi_0(t)\mathcal{E}(2\nu C_H \mathcal{V}_t)$, for $t \in \mathcal{T}$;

[3]CBOE VIX futures historical data: http://cfe.cboe.com/data/historicaldata.aspx

(iii) extract the path of the Brownian motion Z driving $\mathcal{V}$:

$$\begin{aligned} Z_{t_i} &= Z_{t_{i-1}} + n^{H-}\left(\mathcal{V}(t_i) - \mathcal{V}(t_{i-1})\right), && \text{for } i = 1, \dots, \kappa, \\ Z_{t_i} &= Z_{t_{i-1}} + \overline{Z}_{i-1}, && \text{for } i > \kappa; \end{aligned}$$

compute $\{Z^{\perp}\}_{i=0}^{n_T-1}$ where $Z_i^{\perp} \overset{\Delta}{=} \mathcal{N}(0, 1/n_T)$ is an independent standard Gaussian sample;

(iv) correlate the two Brownian motions via $W_{t_i} - W_{t_{i-1}} = \rho \overline{Z}_{i-1} + \sqrt{1-\rho^2}\overline{Z}^{\perp}_{i-1}$;

(v) simulate $S_{t_i} = \exp(X_{t_i})$ using a forward Euler scheme:

$$X_{t_{i+1}} = X_{t_i} - \frac{1}{2}V_{t_i}(t_{i+1} - t_i) + \sqrt{V_{t_i}}\left(W_{t_{i+1}} - W_{t_i}\right), \qquad \text{for } i = 0, \dots, n_T - 1;$$

(vi) compute the expectation by averaging the payoff over all terminal values of each path.

5.4.2 Calibration of SPX options via VIX Futures

We first follow the calibration algorithm 5.3.21 to obtain H and ξ_0, and we then aim at minimising, over (ν, ρ), the objective function

$$\mathcal{L}^{\mathrm{C}}(\nu, \rho) := \sum_{j=1}^{L}\sum_{i=1}^{N}(\mathrm{C}_{T_{i,j}} - \mathrm{C}^{\mathrm{obs}}_{i,j})^2, \tag{5.20}$$

where $\mathrm{C}_{T_{i,j}}$ is the Call price given by the rough Bergomi model, computed using the scheme introduced in Section 5.4.1, with maturity T_i and strike $K^{(j)}$. On the other hand, $(\mathrm{C}^{\mathrm{obs}}_{i,j})_{i,j}$ is the set observed Call prices in the time grid $T_1 < \ldots < T_N$ and strike grid $K^{(1)} < \ldots < K^{(L)}$. In order to optimise the calibration algorithm, we first compute the Volterra process $\mathcal{V}$, which will then be used in a forward Euler simulation at each calibration step:

Algorithm 5.4.2 (Calibration algorithm for SPX options via VIX Futures)**.**

(i) Calibrate H and ξ_0 using the VIX Futures;

(ii) compute M paths of the Volterra process, $\{\mathcal{V}^{(u)}\}_{u=1}^{M}$ and extract the Brownian motions $\{Z^{(u)}\}_{u=1}^{M}$ driving each process. Also, compute independent Brownian motions $\{Z^{\perp(u)}\}_{u=1}^{M}$;

(iii) evaluate the Call prices in each calibration step:

$$\begin{aligned} V_t^{(u)} &= \xi_0(t)\mathcal{E}\left(2\nu C_H \mathcal{V}_t^{(u)}\right), && u = 1, \dots, M, \\ W^{(u)} &= \rho Z^{(u)} + \sqrt{1-\rho^2}Z^{\perp(u)}, && u = 1, \dots, M, \\ S^{(u)}_{t+\Delta} &= S^{(u)}_t + S^{(u)}_t\sqrt{v_t^{(u)}}\left(W^{(u)}_{t+\Delta} - W^{(u)}_t\right), && u = 1, \dots, M; \end{aligned}$$

(iv) compute the Call price for each available maturity $\{T_1, \dots, T_N\}$ and set of strikes $\{K^{(1)}, \dots, K^{(L)}\}$:

$$\mathrm{C}_{T_i,j} = \frac{1}{M}\sum_{u=1}^{M}(S^{(u)}_{T_i} - K^{(j)})_+, \quad \text{for } i = 1, \dots, N \text{ and } j = 1, \dots, L;$$

(v) minimise over (ν, ρ) the objective function $\mathcal{L}^{\mathrm{C}}(\nu, \rho)$ in (5.20).

Remark 5.4.3. Item (v) in the algorithm above may change the optimal values for ν, which was initially calibrated in (i) from the VIX Futures. Backtesting however shows that the calibration in (i) is not really affected by this.

5.4.3 Results

We calibrate the model on December 4, 2015, fixing $H = 0.09237$ obtained previously through VIX, and plot the fit in Figure 5.12. The model is not fully consistent for short maturities, which may follow from the inability of ξ_0 to

fully capture the smiles for these maturities, but the fit greatly improves with maturity. Interestingly, we observe a 20% difference between the parameter ν obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?). Nevertheless, we emphasise the importance of an accurate ξ_0 curve to improve the fit to SPX and to provide an efficient joint calibration. Note in passing that the calibration drives the correlation parameter to the lower bound of its admissible values.

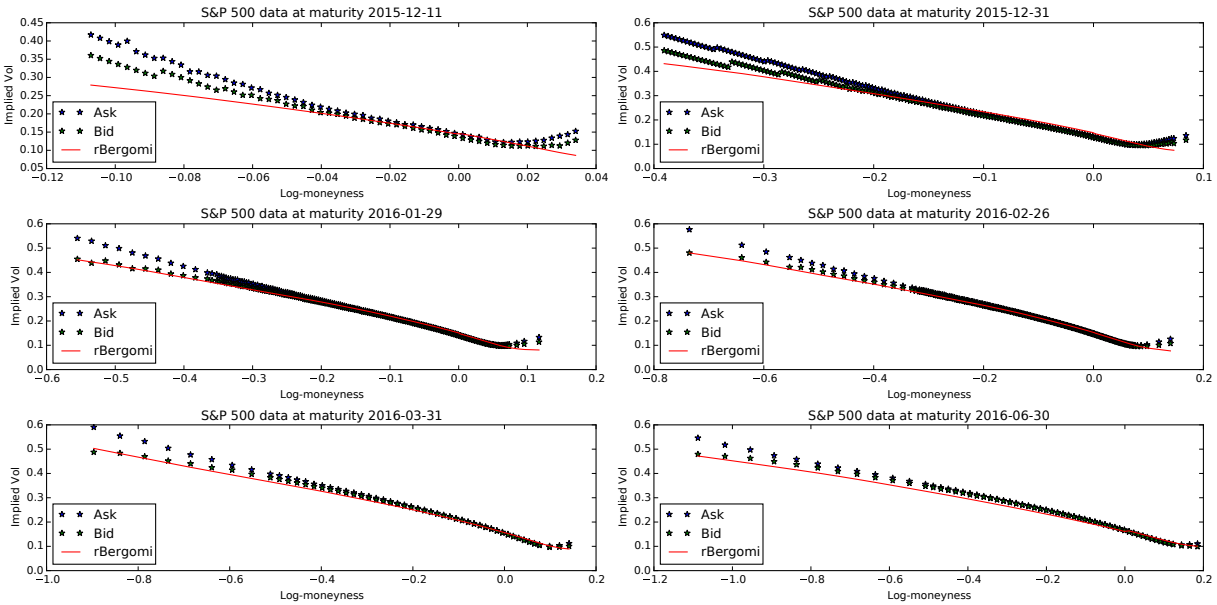


Figure 5.12: Calibration of SPX smiles on 4/12/2015. Calibrated parameters: $(\nu, \rho) = (1.19, -0.999)$.

5.5 Summary

Following the path set by Bayer, Friz and Gatheral [BFG16], we developed here a relatively fast algorithm to calibrate VIX Futures, consistently with the SPX smile, in the rough Bergomi model. To do so, we have characterised the dynamics of the VIX process under the rough Bergomi model. The clear strength of this model is that only a few parameters are needed, making the (re)calibration robust and stable. From a trader's point of view, we highlight some potential market discrepancy between the VIX and the SPX, and leave a refined analysis thereof for future research.

6

On Smile properties of volatility products

6.1 Introduction

The raison d'être of any financial model is to reproduce some behavior or dynamic that is observed in the market. For instance, in the case of vanilla options, this conundrum is translated into fitting the market implied volatility surface. To this end, several extensions of the Black-Scholes model have been proposed. A popular approach is to allow the volatility to be a stochastic process itself. It is well known that classical stochastic volatility (SV) models, where the instantaneous volatility follows a diffusion process, can explain some important properties of the implied volatility, such as its variation with respect to the strike price, described graphically as the smile or skew (see for example [RT96] or [Lee05]). A more difficult problem is to reproduce the observed term structure (dependence on the time to maturity) of the implied volatility. In particular, a popular rule of thumb (see [Lee05]) approximates the skew slope decay rate as $T^{-\frac{1}{2}}$, a property that is not satisfied for most of the classical SV models.

There have been several attempts in the literature to describe this phenomenon. Some works in the literature focus on the introduction of more volatility factors (as in the two-factor Bergomi model) that allow for two scales of mean-reversion and lead to a power-like behavior over a suciently wide range of maturities (see [Ber09]). Another approach has been introduced in [Fou+04], where the authors propose a model with time-varying coefficients, that depend on the option maturity dates. Then, the proposed calibration method is simply to fit the empirical implied volatility surface to a straight line in a composite variable (the so-called log-moneyness-to-effective-time to-maturity). In [ALV07], the authors described the short-time behavior of the implied volatility in terms of the Malliavin derivative of the volatility process and proved that volatility models driven by a fractional Brownian motion (fBm) with Hurst parameter $H < 1/2$ (dubbed in [GJR18] as 'rough volatility models') can reproduce a skew slope decay rate of $T^{H-\frac{1}{2}}$, for $H \in (0, \frac{1}{2})$.

Our main objective in this Chapter is to study the at-the-money implied (ATMI) short-time level and skew of realized variance (RV) options and VIX options. VIX options and futures are becoming increasingly popular both in the industry and academic research (see [CM14] and [HJT20] for instance). There have been many attempts to find a model that reproduces at the same time the VIX and the underlying index (the S&P 500) implied volatility surfaces (see [BB14; CM14; FS18; DeM18; Guy20b; Gat08; KS15; PS14], and the references therein). Recent research focuses on the asymptotic expansion of VIX option prices and implied volatilities, as in [BNP19].

Our analysis is based on the fact that, given a volatility derivative of the form $f(A)$, where A is a $\mathcal{F}_T$- measurable random variable, we can write $A = M_T$, where M is the forward price defined as $M_t = \mathbb{E}_t[A]$, where $\mathbb{E}_t[A]$ denotes

the conditional expectation $\mathbb{E}_t[A] = \mathbb{E}[A|\mathcal{F}_t]$. This method is very classical in math finance, for instance when working with stochastic rates. Notice that the martingale M satisfies, under some conditions, that $dM_t = \phi_t M_t \mathrm{d}W_t$, for some stochastic process ϕ depending on time to maturity T. That is, the problem reduces to study European options where the underlying is given by a stochastic volatility model, where the corresponding volatility ϕ depends on T. Our analysis can be applied to the case of rough volatility models, and our main results can be summarized as follows:

- We prove that the ATMI level and skew are of the order $O((T)^{H-\frac{1}{2}})$ for RV options and of the order $O(1)$ for VIX options, where H denotes the Hurst parameter of the volatility process.
- We construct an easy-to-apply criterion to test, given a volatility model, if the corresponding short time VIX skew is positive, as observed in real market data.

As an application of the above criterion, we prove that the short-time VIX skew can be positive or negative for the Heston model, being negative for reasonable values for the mean reversion rate. On the other hand, it is zero for the SABR, while it becomes positive with a "mixing lognormal" solution. This coincides with previous studies (see for example [BB14]) and [Ber08]), and, up to our knowledge, it gives the first analytical proof of these properties in the literature.

6.2 Malliavin calculus

We assume that the reader is familiar with the elementary notions of Malliavin calculus, as given for instance in Nualart [Nua06]. Consider a Brownian motion W defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For all $p \geq 1$, the set $\mathbb{D}_W^{1,p}$ will denote the domain of the derivative operator D with respect to the Brownian Motion W in $L^p(\Omega)$. It is well-known that, for all $p \geq 1$, $\mathbb{D}_W^{1,p}$ is a dense subset of $L^p(\Omega)$ and that D is a closed and unbounded operator from $L^p(\Omega)$ into $L^p(\Omega; L^2([0,T]))$. We will also consider the iterated derivatives D^n, for $n > 1$, whose domains will be denoted by $\mathbb{D}_W^{n,p}$. We will also make use of the notation $\mathbb{L}^{n,p} = \mathbb{D}_W^{n,p}(L^2([0,T]))$.

Consider now a process X given by the equation

$$\mathrm{d}X_t = \phi_t \mathrm{d}W_t - \frac{1}{2}\phi_t^2 \mathrm{d}t, \tag{6.1}$$

where ϕ_t is a positive and square integrable process adapted to the filtration generated by W. We will make use of the following anticipating Itô's formula (see for example [Alò06]).

Proposition 6.2.1. *Consider the model (6.1) where $\phi \in \mathbb{L}^{1,2}$ and define the process Y as $Y_t := \int_t^T \phi_s^2 \mathrm{d}s$. Let $F : [0,T] \times \mathbb{R}^2 \to \mathbb{R}$ be a function in $C^{1,2}([0,T] \times \mathbb{R}^2)$ such that there exists a positive constant C such that, for all $t \in [0,T]$, F and its partial derivatives evaluated in (t, X_t, Y_t) are bounded by C. Then it follows that*

$$\begin{aligned} F(t,X_t,Y_t) &= F(0,X_0,Y_0) + \int_0^t \partial_s F(s,X_s,Y_s)\mathrm{d}s \\ &\quad -\frac{1}{2}\int_0^t \partial_x F(s,X_s,Y_s)\phi_s^2 \mathrm{d}s \\ &\quad + \int_0^t \partial_x F(s,X_s,Y_s)\sqrt{\phi_s^2}\mathrm{d}W_s \\ &\quad - \int_0^t \partial_y F(s,X_s,Y_s)\phi_s^2 \mathrm{d}s + \int_0^t \partial_{xy}^2 F(s,X_s,Y_s)\Theta_s \mathrm{d}s \\ &\quad + \frac{1}{2}\int_0^t \partial_{xx}^2 F(s,X_s,Y_s)\phi_s^2 \mathrm{d}s, \end{aligned} \tag{6.2}$$

where $\Theta_s := (\int_s^T D_s^W \phi_r^2 \mathrm{d}r)\phi_s$.

6.3 Transforming variance options into options on the spot process of a stochastic volatility model

We study options with maturity T (expressed in years), defined by a payoff of European type

$$(A-K)_+,$$

where A is a positive and square-integrable random variable defined on a risk neutral probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Notice that, in incomplete markets, this risk neutral probability is not unique. The approach that we will follow here is the same used in [Fou+04], where it is assumed that the market selects a unique equivalent martingale measure under which derivative contracts are priced. We assume that A is $\mathcal{F}_T^W$-adapted, where $\mathcal{F}^W$ denotes the sigma-algebra generated by a Brownian motion W. Then, the price of this option at a moment $0 < t < T$ is given by

$$P_t := e^{-r(T-t)}\mathbb{E}_t\left[(A-K)_+\right],$$

where $\mathbb{E}_t$ denotes the expectation with respect to $\mathcal{F}_t$ and r is the interest rate, that we assume to be constant.

Now, consider the martingale process $M_t^T := \mathbb{E}_t[A]$, $t \in [0,T]$. By the martingale representation theorem (see, for instance, Karatzas and Shreve [KS97], theorem 3.4.15), there exists a $\mathcal{F}_t^W$-adapted process m^T such that

$$M_t^T = \mathbb{E}[M_t^T] + \int_0^t m_s^T \mathrm{d}W_s. \tag{6.3}$$

Then, M_t^T admits the following dynamics:

$$M_t^T = \mathbb{E}[M_t^T] + \int_0^t \phi_s M_s^T \mathrm{d}W_s, \tag{6.4}$$

where

$$\phi_t := \frac{m_t^T}{M_t^T}. \tag{6.5}$$

Remark 6.3.1. Note here that in (6.4) the stochastic volatility ϕ_t also depends on the time to maturity T.

Example 6.3.2 (VIX options)**.** Consider as underlying the random variable VIX_T given by

$$VIX_T = \sqrt{\frac{1}{\Delta}\mathbb{E}_T\left[\int_T^{T+\Delta} V_s \mathrm{d}s\right]}, \tag{6.6}$$

where V, the instantaneous variance process (the square of the volatility process), is positive and adapted to the filtration generated by a Brownian motion W, and where $(T, T+\Delta)$ denotes a 30-days interval. We define $M_t^{T,VIX} := \mathbb{E}_T[VIX_T]$. Then, if $VIX_T \in \mathbb{D}_W^{1,2}$, the Clark-Ocone formula and Proposition 1.2.8 in [Nua06] give us that

$$\begin{aligned} m_t^{T,VIX} &= \frac{1}{2\Delta}\mathbb{E}_t\left[\frac{1}{VIX_T}\mathbb{E}_T\left[\int_T^{T+\Delta}(D_t V_s)\mathrm{d}s\right]\right] \\ &= \frac{1}{2\Delta}\mathbb{E}_t\left[\frac{1}{VIX_T}\int_T^{T+\Delta}(D_t V_s)\mathrm{d}s\right]. \end{aligned} \tag{6.7}$$

Then, using Equality (6.5) we identify the maturity dependent stochastic volatility as

$$\phi_t^{VIX} = \frac{1}{2\Delta M_t^{T,VIX}}\mathbb{E}_t\left[\frac{1}{VIX_T}\int_T^{T+\Delta}(D_t V_s)\mathrm{d}s\right]. \tag{6.8}$$

Example 6.3.3 (Realized variance options)**.** Consider the case

$$RV_T = \frac{1}{T}\int_0^T V_s \mathrm{d}s, \tag{6.9}$$

where V is the variance process as in Example 6.3.2 and we define $M_t^{T,RV} := \mathbb{E}_t[RV_T]$. Then, the Clark-Ocone formula gives us that (6.3) holds with

$$m_t^{T,RV} = \frac{1}{T}\int_t^T \mathbb{E}_t[D_t V_s]\mathrm{d}s.$$

Therefore,

$$\phi_t^{RV} = \frac{1}{T}\frac{\int_t^T \mathbb{E}_t[D_t V_s]\mathrm{d}s}{M_t^{T,RV}}. \tag{6.10}$$

6.4 Implied volatility for variance options

Our purpose in this section is to develop the tools to study the short-time behavior of the at-the-money implied volatility (ATMI) for call options on some underlying A. In the following sections, we will apply these results to the particular case of VIX options and RV options.

Let us define the implied volatility as the quantity $I_t^T(k)$ such that

$$P_t = BS(t, \ln(\mathbb{E}_t[A]), k, I_t^T(k)),$$

where $BS(t, x, k, \sigma)$ denotes the classical Black-Sholes price for a European call on a forward price with time to maturity $T-t$, log-forward price x, log-strike price $k = \ln(K)$ and volatility σ. That is,

$$BS(t, x, k, \sigma) = e^{-r(T-t)}[e^x N(d_+(k,\sigma)) - e^k N(d_-(k,\sigma))],$$

where N denotes the cumulative probability function of the standard normal law and

$$d_\pm(k,\sigma) := \frac{x-k}{\sigma\sqrt{T-t}} \pm \frac{\sigma}{2}\sqrt{T-t}.$$

For sake of simplicity we will denote $I_T(t) := I_t^T(\ln(\mathbb{E}_t[A]))$ the corresponding ATMI. Notice that $I_T(t) = BS^{-1}(t, \ln(\mathbb{E}_t[A]), \ln(\mathbb{E}_t[A]), P_t)$. In the sequel we will write $BS^{-1}(t,u) = BS^{-1}(t, \ln(\mathbb{E}_t[A]), \ln(\mathbb{E}_t[A]), u)$.

Definition 6.4.1 (ATMI skew)**.** The ATMI skew is defined as

$$\mathcal{S}_T(t) := \frac{\partial I_t^T(k)}{\partial k}\Big|_{k=\ln(\mathbb{E}_t[A])}.$$

6.4.1 The ATMI short-time limit

Now our first objective is to study the short-time behavior of the implied volatility level of variance options. Towards this end, we will need the following technical integrability conditions.

(H1) $A \in \mathbb{L}^{2,p}$ for all $p > 1$

(H2) $\frac{1}{M_t^T} \in L^p(\Omega)$, for all $p > 1$.

(H3) (H1) and (H2) hold and the terms

$$\mathbb{E}\left[\int_0^T \frac{1}{u_s^2(T-s)}\Theta_s \mathrm{d}s\right],$$

and

$$\frac{1}{T^2}\mathbb{E}\left[\frac{1}{u_0}\int_0^T\left(\int_s^T D_s\phi_r^2 \mathrm{d}r\right)^2 \mathrm{d}s\right]$$

are well defined and tend to zero as $T \to 0$, where $u_t := \sqrt{\frac{1}{T-t}\int_t^T \phi_s^2 \mathrm{d}s}$, with ϕ defined as in (6.5) and $\Theta_s := \phi_s \int_s^T D_s\phi_r^2 \mathrm{d}r$.

(H4) There exists $\gamma \in \left(-\frac{1}{2}, 0\right]$ such that the term

$$\frac{1}{T^{\frac{1}{2}+\gamma}}\mathbb{E}\left[\sqrt{\int_0^T \phi_s^2 \mathrm{d}s}\right]$$

has a finite limit as $T \to 0$.

Remark 6.4.2. In the case of VIX options or RV options, the above hypotheses are satisfied by most of the classical stochastic volatility models. For example, in the Heston model, the bounds proved in [AE08] for the CIR model and its Malliavin derivatives imply that (H1)-(H3) hold. On the other hand, assume for example that $V_t = f(Y_t)$, where f is a positive real function, and Y is an Itô diffusion process of the form

$$dY_t = a(Y_t)\mathrm{d}t + b(Y_t)\mathrm{d}W_t,$$

for some real functions a, b. Then, under some adequate integrability conditions for f, a and b, v is a process in $\mathbb{L}^{2,2}$ and the processes Dv and D^2v have finite moments of all orders, from where it follows easily that (H1)-(H4) hold with $\gamma = 0$ both for VIX options and for RV options.
In the fractional volatility case, one of the most popular choices is to assume the variance process to be of the form

$$V_t = V_0 \exp\left(\nu W_t^H - \frac{1}{2}\nu^2 t^{2H}\right),$$

where $W_t^H := \int_0^t (t-s)^{H-\frac{1}{2}} \mathrm{d}W_s$ (see for example [BFG16]). In this case, $v \in \mathbb{L}^{2,2}$ and $D_sV_t = \nu V_t(t-s)^{H-\frac{1}{2}}$. Then, it is easy to see that (H1)-(H4) hold, with $\gamma = 0$ in the case of the VIX, and $\gamma = H - \frac{1}{2}$ in the case of RV options (see the examples in Sections 5 and 6 for further details).

Remark 6.4.3. Notice that (H1) implies that the process M_t^T is also in $\mathbb{L}^{1,p}$ for all $p > 1$, and $D_sM_t^T = \mathbb{E}_t[D_sA]$ (see Proposition 1.2.8 in Nualart (2005)). Moreover, from Clark-Ocone formula we get that $m_t^T = \mathbb{E}_t[D_tA]$, which allows us to see that $m_t^T \in \mathbb{L}^{1,2}$ and $D_sm_t^T = \mathbb{E}_t[D_sD_tA]$ for all $s < t$. This, jointly with (H2), gives us that ϕ is also in $\mathbb{L}^{1,p}$ for all $p > 1$, and

$$\begin{aligned} D_s\phi_t &= \frac{D_sm_t^T M_t^T - m_t^T D_sM_t^T}{(M_t^T)^2} \\ &= \frac{\mathbb{E}_t[D_sD_tA]\mathbb{E}_t[A] - \mathbb{E}_t[D_tA]\mathbb{E}_t[D_sA]}{(\mathbb{E}_t[A])^2}. \end{aligned}$$

The following result, the proof of which is postponed to Section A.1, is a direct consequence of the arguments in the proof of Theorems 3 and 8 in [AS19].

Theorem 6.4.4. *Assume that hypotheses (H1), (H2) and (H3) hold. Then*

$$\lim_{T\to 0}\left(I_0^T - \mathbb{E}\left[\sqrt{\frac{1}{T}\int_0^T \phi_s^2 \mathrm{d}s}\right]\right) = 0.$$

This result gives us, if (H4) holds, the following corollary.

Corollary 6.4.5. *Assume a random variable A such that hypotheses (H1), (H2), (H3) and (H4) hold. Then*

$$\lim_{T\to 0} T^{-\gamma} I_0^T = \lim_{T\to 0} \frac{1}{T^{\frac{1}{2}+\gamma}} \mathbb{E}\left[\sqrt{\int_0^T \phi_s^2 \mathrm{d}s}\right].$$

6.4.2 The short-time limit of the ATM implied volatility skew

Our main goal in this section is to study the short-time limit of the ATM implied volatility skew. In particular, we will characterize the class of stochastic volatility processes that reproduce a positive VIX skew, as observed in real market data. We will need the following technical hypotheses.

(H1') $A \in \mathbb{L}^{3,p}$ for all $p > 1$

(H5) The terms

$$\frac{1}{\sqrt{T}} \mathbb{E}\left[\int_0^T ((u_s)^2 (T-s))^{-3} \left(\int_s^T \Theta_r \mathrm{d}r\right) \Theta_s \mathrm{d}s\right]$$

and

$$\frac{1}{\sqrt{T}} \mathbb{E}\left[\int_0^T ((u_s)^2 (T-s))^{-2} \left(\int_s^T D_s \Theta_r \mathrm{d}r\right) \phi_s \mathrm{d}s\right]$$

tend to zero as $T \to 0$.

(H6) There exists $\lambda \in \left(-\frac{1}{2}, 0\right]$ such that the term

$$\mathbb{E}\left[\frac{\int_0^T \left(\int_s^T \phi_s D_s \phi_u^2 \mathrm{d}u\right) \mathrm{d}s}{u_0^3 T^{2+\lambda}}\right]$$

has a finite limit as $T \to 0$.

Theorem 6.4.6. *Consider a random variable A such that Hypotheses (H1'), (H2), (H3), (H5) and (H6) hold. Then*

$$\lim_{T\to 0} T^{-\lambda} \mathcal{S}_T(0) = \frac{1}{2} \lim_{T\to 0} \mathbb{E}\left[\frac{\int_0^T \left(\int_s^T \phi_s D_s \phi_u^2 \mathrm{d}u\right) \mathrm{d}s}{u_0^3 T^{2+\lambda}}\right] \tag{6.11}$$

Remark 6.4.7. The VIX skew is observed to be positive in real market data. It is not easy to establish *a priori* if a stochastic volatility model will reproduce this phenomenon. It is well known (see for example [FS18], [GIP17], [DeM18] or [Guy20b]) that some classical models exhibit a negative VIX skew (as the Heston) of a flat VIX skew (as the SABR). Notice that $D_s \phi_t^2 = 2\phi_t D_s \phi_t$. Then the above theorem gives us that the sign of the short-time skew is positive if

$$\lim_{s,t,T\to 0} \mathbb{E}\left[(D_s \phi_t - u(s,t))^p\right] = 0$$

for some positive process $u \in L^2([0,T]^2 \times \Omega)$ and for some $p > 1$. By (6.5), this condition holds if

$$\lim_{s,t,T\to 0} \mathbb{E}\left[\left((D_s m_t^T) M_t^T - m_t^T D_s M_t^T\right) - \hat{u}(s,t)\right)^q\right] = 0, \tag{6.12}$$

for some positive process $\hat{u} \in L^2([0,T]^2 \times \Omega)$ and some $q > 1$. Notice that (6.12) gives us a tool to identify those stochastic volatility models that can reproduce a short-time positive skew, as we will see in the next section.

6.4.3 The sign of the VIX skew in some popular models

The VIX skew is observed to be positive in real market data. It is not easy to establish *a priori* if a stochastic volatility model will reproduce this phenomenon. It is well known numerically (see for example [FS18], [GIP17], [DeM18] or [Guy20b]) that some classical models exhibit a negative VIX skew (as the Heston) of a flat VIX skew (as the SABR). We note that Remark 6.4.7 gives us a tool to identify those stochastic volatility models that can reproduce a short-time positive skew. Consequently, we make use of Remark 6.4.7 to formally show the sign of the VIX skew in some popular models.

Example 6.4.8 (The VIX skew in the SABR model)**.** Consider the SABR model given by $V = \sigma^2$, where

$$d\sigma_t = \alpha\sigma_t \mathrm{d}W_t,$$

for some positive constant α. Then $v \in \mathbb{L}^{2,2}$ and $D_s V_t = 2\alpha V_t$, $D_s D_r V_t = 4\alpha^2 V_t$, for all $r, s < t$. This implies that (H6) holds with $\lambda = 0$ and

$$\lim_{t,T\to 0} M_t^{T,VIX} = \frac{\sqrt{e^{\alpha^2\Delta}-1}\sqrt{V_0}}{\alpha\sqrt{\Delta}},$$
$$\lim_{s,t,T\to 0} D_s M_t^{T,VIX} = \frac{(e^{\alpha^2\Delta}-1)V_0}{VIX_0\alpha\Delta} = \frac{\sqrt{e^{\alpha^2\Delta}-1}\sqrt{V_0}}{\sqrt{\Delta}}$$

$$\begin{aligned} D_s m_t^{T,VIX} &= \frac{1}{2\Delta}\mathbb{E}_t\left[\frac{1}{VIX_T}\mathbb{E}_T\left[\int_T^{T+\Delta}(D_s D_t V_r)\mathrm{d}r\right]\right] \\ &- \frac{1}{4\Delta}\mathbb{E}_t\left[\frac{1}{\Delta(VIX_T)^3}\mathbb{E}_T\left[\int_T^{T+\Delta}(D_t V_r)\mathrm{d}r\right]\mathbb{E}_T\left[\int_T^{T+\Delta}(D_s V_r)\mathrm{d}r\right)\right]. \end{aligned} \tag{6.13}$$

From here, we conclude that

$$\begin{aligned} \lim_{s,t,T\to 0} D_s m_t^{T,VIX} &= \frac{4\alpha^2 V_0}{2VIX_0\Delta}\frac{e^{\alpha^2\Delta}-1}{\alpha^2} - \frac{1}{4VIX_0^3\Delta^2}\frac{4\alpha^2 V_0^2(e^{\alpha^2\Delta}-1)^2}{\alpha^4} \\ &= \frac{2\alpha\sqrt{V_0}\sqrt{e^{\alpha^2\Delta}-1}}{\sqrt{\Delta}} - \frac{\alpha\sqrt{V_0}\sqrt{e^{\alpha^2\Delta}-1}}{\sqrt{\Delta}} \\ &= \frac{\alpha\sqrt{V_0}\sqrt{e^{\alpha^2\Delta}-1}}{\sqrt{\Delta}} \end{aligned} \tag{6.14}$$

in $L^p(\Omega)$, for all $p > 1$, from where we deduce that

$$\lim_{T\to 0} D_s m_t^{T,VIX} M_t^{T,VIX} - m_t^{T,VIX} D_s M_t^{T,VIX} = \alpha\frac{\sqrt{V_0}\sqrt{e^{\alpha^2\Delta}-1}}{\sqrt{\Delta}}\frac{\sqrt{e^{\alpha^2\Delta}-1}\sqrt{V_0}}{\alpha\sqrt{\Delta}} - \left(\frac{\sqrt{e^{\alpha^2\Delta}-1}\sqrt{V_0}}{\sqrt{\Delta}}\right)^2 = 0 \tag{6.15}$$

This, jointly with (6.8) and Remark 6.4.3, gives us that the (lognormal) SABR model generates a short-time flat VIX skew.

Example 6.4.9 (The sign of the VIX skew in the Heston model)**.** For the Heston model, we have that

$$dV_t = \kappa(\theta - V_t)\mathrm{d}t + \nu\sqrt{V_t}\mathrm{d}W_t,$$

for some positive constants k, θ and ν. We know (see the proof of Proposition 4.2 in[AE08]) that if $2k\theta > 3\nu^2$, the process $\sqrt{V} \in \mathbb{L}^{2,2}$. Moreover, (H6) holds with $\lambda = 0$ and Dv is positive, which implies that ϕ is positive. Assume,

for the sake of simplicity, that $V_0 = \theta$. Then we have

$$\lim_{T\to 0} D_s m_t^{T,VIX} M_t^{T,VIX} - m_t^{T,VIX} D_s M_t^{T,VIX} = f(\Delta) := \frac{\nu^2(1-\exp(-\kappa\Delta))}{4\kappa\Delta}\left(1 - \frac{2(1-\exp(-\kappa\Delta))}{\kappa\Delta}\right),$$

in $L^2(\Omega)$. The computation of the limit is given in Appendix D.3. Under reasonable market conditions for the reversion level κ, we have that $f(\Delta)$ is negative. This, jointly with Remark 6.4.3, implies that the limit in the right-hand side of (6.11) is negative. Nevertheless, for big values of the mean reversion rate κ, (6.16) can be positive. This result fits [KS15] [6, Table VI, Figure 7] where a positive VIX skew in the Heston model is reported even for reasonable values of the mean reversion rate if the Feller condition is not satisfied.

6.5 VIX OPTIONS

This section is devoted to applying the results in Section 4 to the study of the ATMI level and skew of VIX options. First of all, let us illustrate the typical behaviour of these quantities in the market. Figure 6.1 shows the behaviour of the VIX implied volatility. We observe that the skew is clearly positive, but also decreasing as a function of time to maturity, since smiles tend to flatten.

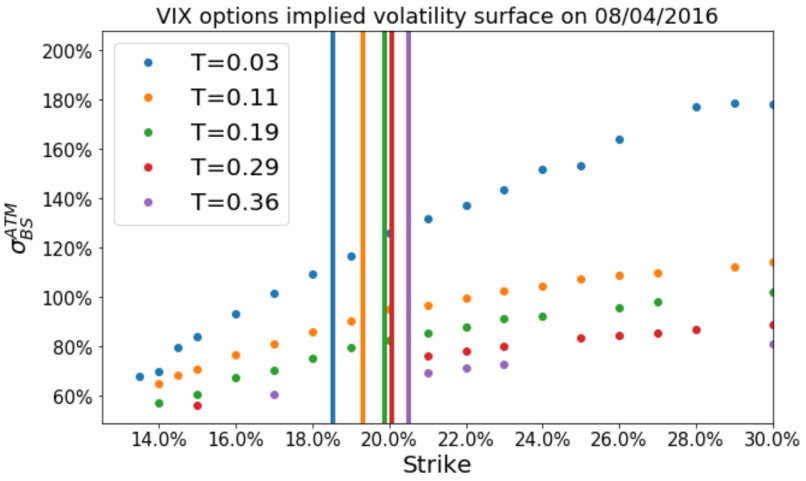


Figure 6.1: VIX options implied volatilities on 08/04/2016. Solid lines represent the VIX future prices corresponding to each maturity. Source: OptionMetrics.

The results in this section study the ATMI level and skew for models based on Gaussian processes of the form $Y_t = \int_0^t (t-s)^{H-\frac{1}{2}} g(t-s) \mathrm{d}W_s$, where H denotes the Hurst parameter, and g is a bounded deterministic function. These models are very flexible in the modeling of the volatility process (see for example [BLP17] and [BS07]).

6.5.1 ATMI of VIX options

Our aim in this subsection is to study the short-time limit of the ATMI of VIX options. Our main result is the following proposition, that proves that, for volatilities based on a Brownian stationary process this ATMI is $O(1)$ as $T \to 0$.

Proposition 6.5.1. *Consider the process VIX_T given by (6.6), with $V = f(Y)$, where f is a positive and bounded function in $\mathcal{C}^2$, and $Y_t = \int_0^t (t-s)^{H-\frac{1}{2}} g(t-s) \mathrm{d}W_s$, where $H \in (0, \frac{1}{2}]$ and $g \in \mathcal{C}_b^1$. Then it follows that*

$$\lim_{T\to 0} I_T^{VIX}(0) = \frac{f'(Y_0)}{2\Delta(VIX_0)^2}\psi(\Delta), \tag{6.16}$$

where $\psi(\Delta) := \int_0^\Delta u^{H-\frac{1}{2}} g(u)\mathrm{d}u$.

Proof. The proof of this proposition is decomposed into two steps.

Step 1 Let us see that hypotheses (H1)-(H4) hold with $\gamma = 0$. (H2) follows from the fact that f is bounded. On the other hand, as $D_s V_t = f'(Y_t)(t-s)^{H-\frac{1}{2}} g(t-s)$ and $D_r D_s V_t = f'(Y_t)(t-s)^{H-\frac{1}{2}} g(t-s)(t-r)^{H-\frac{1}{2}} g(t-r)$, a direct calculation gives us (H1). Moreover, as

$$\int_T^{T+\Delta} (D_s(V_u))\mathrm{d}u = \int_T^{T+\Delta} f'(Y_u)(u-s)^{H-\frac{1}{2}} g(s-u)\mathrm{d}s$$

and

$$\int_T^{T+\Delta} (D_r D_s(V_u))\mathrm{d}u = \int_T^{T+\Delta} f'(Y_u)(u-s)^{H-\frac{1}{2}} g(u-s)(u-r)^{H-\frac{1}{2}} g(u-r)\mathrm{d}s$$

we can easily check that

$$\phi_t^{VIX} = \frac{m_t^{T,VIX}}{M_t^{T,VIX}}$$

and

$$D_s\phi_u^2 = \frac{2\phi_u (D_s m_u^{T,VIX}) M_u^T}{(M_s^{T,VIX})^2} - \frac{2\phi_u m_u^{T,VIX} D_s M_u^T}{(M_s^T)^2}$$

are bounded in $L^p([0,T];\Omega)$ and $L^p([0,T]^2;\Omega)$. Then, (H3) and (H4) hold with $\gamma = 0$. *Step 2* As (H1)-(H4) are satisfied, Corollary 6.4.5 gives us that

$$\begin{aligned}
\lim_{T\to 0} I_T^{VIX}(0) &= \lim_{T\to 0} \frac{1}{T^{\frac{1}{2}}} \mathbb{E}\left[\sqrt{\int_0^T (\phi_s^{VIX})^2 \mathrm{d}s}\right] \\
&= \lim_{T\to 0} \frac{1}{T^{\frac{1}{2}}} \mathbb{E}\left[\sqrt{\int_0^T \left(\frac{1}{2\Delta M_s^{T,VIX}} E_s\left(\frac{1}{VIX_T}\int_T^{T+\Delta} (D_s(V_u))\mathrm{d}u\right)\right)^2 \mathrm{d}s}\right] \\
&= \frac{f'(Y_0)}{2\Delta M_0^{0,VIX} VIX_0}\left(\int_0^\Delta u^{H-\frac{1}{2}} g(u)\mathrm{d}u\right) \\
&= \frac{f'(Y_0)}{2\Delta (VIX_0)^2}\left(\int_0^\Delta u^{H-\frac{1}{2}} g(u)\mathrm{d}u\right) \\
&=: \frac{f'(Y_0)}{2\Delta (VIX_0)^2}\psi(\Delta).
\end{aligned}$$

□

The hypotheses in the previous proposition have been chosen for the sake of simplicity, but this result can be adapted to other models with adequate integrability conditions, as we can see in the following example.

Example 6.5.2 (Mixed generalized rough volatility models (MGRV))**.** We consider the following instantaneous variance dynamics:

$$V_t = V_0\left(\delta\mathcal{E}\left(\nu\sqrt{2H}\mathcal{B}_t\right) + (1-\delta)\mathcal{E}\left(\eta\sqrt{2H}\mathcal{B}_t\right)\right), \quad \nu \geq 0,\ \eta \geq 0$$

where $\mathcal{B}_t$ is a truncated Brownian semistationary process ($\mathcal{TBSS}$) of the form

$$\mathcal{B}_t = \int_0^t \exp\left(-\beta(t-s)\right)(t-s)^{H-1/2}\mathrm{d}W_s, \quad \beta \geq 0$$

(see for example [BLP17]) and $\mathcal{E}$ denotes the Wick stochastic exponential in the sense of [BFG16]. It is then easy to see that

$$\begin{aligned}
D_s V_u &= V_0\left(\delta\nu\mathcal{E}(\nu\sqrt{2H}\mathcal{B}_u) + (1-\delta)\eta\mathcal{E}(\eta\sqrt{2H}\mathcal{B}_u\right) \\
&\quad \times\sqrt{2H}(t-s)^{H-1/2}\exp(-\beta(u-s))
\end{aligned} \tag{6.17}$$

for all $s < u$. This class of models (that we will denote by MGRV) covers most of the lognormal models considered in the literature (SABR [Hag+02], (rough) Bergomi [Ber05; BFG16], etc.) Straightforward computations give us that the model MGRV satisfies hypotheses (H1')-(H6) both for VIX_T options and for RV options, with $\gamma = 0$ and $\gamma = H - \frac{1}{2}$, respectively. Applying Proposition 6.5.1 we get that, in this case,

$$\begin{aligned}
&\lim_{T\to 0} I_0^{T,VIX} \\
&= \frac{(\delta\nu + (1-\delta)\eta)\sqrt{2H}}{2\Delta}\left(\int_0^{\Delta} u^{H-\frac{1}{2}}\exp(-\beta u)\mathrm{d}u\right) \\
&= \frac{(\delta\nu + (1-\delta)\eta)\sqrt{2H}}{2\Delta}\beta^{-H-\frac{1}{2}}\Gamma_{low}\left(H+\frac{1}{2},\beta\Delta\right),
\end{aligned} \tag{6.18}$$

where $\Gamma_{low}(1+\alpha, x) = \int_0^x t^{-\alpha}e^{-t}\mathrm{d}t$ is the lower incomplete Gamma function.

In Figure 6.2 we present numerical approximations with $\delta = 1$ that support this theoretical short time limit result. In order to produce the Monte Carlo estimate, Algorithm 5.3.9 in Chapter 5 and $T = 10^{-4}$ was used.

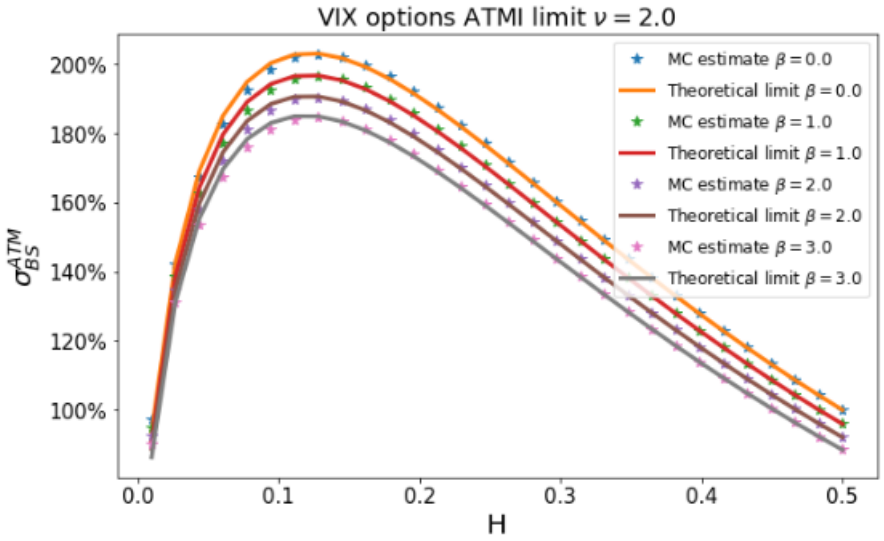


Figure 6.2: VIX ATMI Short time limit in an exponential $\mathcal{TBSS}$ with $\nu = 2$

Remark 6.5.3 (A semi closed-form formula for the ATMI level of VIX)**.** We consider the following short-time approximation inspired by (6.16):

$$I_0^{T,VIX} \approx \frac{f'(Y_0)}{V_0 2\Delta T^{\frac{1}{2}}}\sqrt{\int_0^T\left(\int_T^{T+\Delta}(u-s)^{H-\frac{1}{2}}g(u-s)\mathrm{d}u\right)^2\mathrm{d}s} \tag{6.19}$$

In Figures 6.3 and 6.4 we use formula (6.19) with a MGRV model. As previously mentioned we observe that formula (6.19) indeed performs very well, specially in the case $\delta = 1$ even for relatively large maturities. In practice, one may use approximation (6.19) to control the ATMI level of VIX in our model. Moreover, in most cases (SABR, exponential OU, exponential fBm, etc.) (6.19) admits a closed-form formula without any numerical integration.

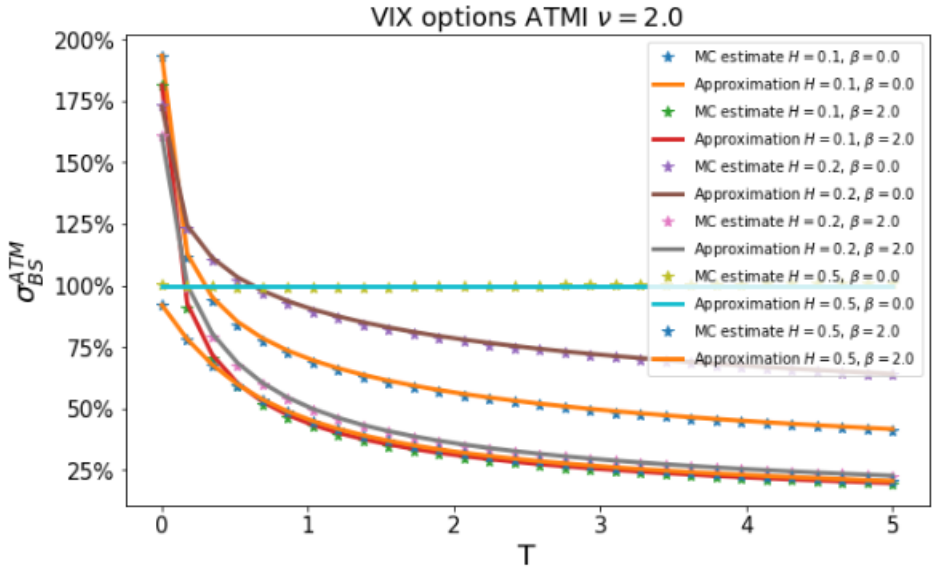


Figure 6.3: VIX ATMI in a Generalized rough volatility model with $\nu = 2$

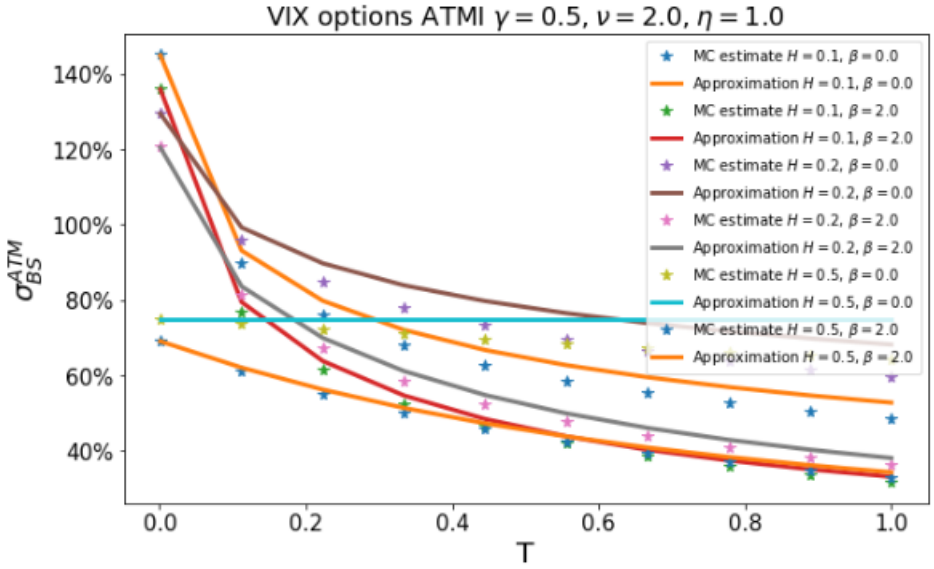


Figure 6.4: VIX ATMI in a Mixed Generalized rough volatility model with $(\delta, \nu, \eta) = (1/2, 2, 1)$

6.5.2 ATMI skew of VIX options

The following result proves that the ATMI VIX skew slope if of the order $O(1)$ as time to maturity tends to zero.

Proposition 6.5.4. *Consider an instantaneous variance model of the form*

$$V_t = V_0 f(Y_t)$$

where $Y_t = \int_0^t (t-s)^{H-1/2} \exp(-\beta(t-s)) \mathrm{d}W_s$ *for some* $H \leq \frac{1}{2}$ *and some* $\beta \geq 0$*, and where* f *is a positive and bounded function in* $\mathcal{C}^2$*. Then, the ATMI skew for VIX options is given by:*

$$\lim_{T \to 0} \mathcal{S}_T^{VIX}(0) = \frac{1}{2} \left(\frac{G(H, \Delta, \beta)}{J(H, \Delta, \beta)} \frac{f''(Y_0)}{f'(Y_0)} - \frac{J(H, \Delta, \beta)}{\Delta} \frac{f'(Y_0)}{(M_0^{0,VIX}) VIX_0} \right).$$

where

$$G(H,\Delta,\beta) = \begin{cases} (2\beta)^{-2H}\Gamma_{low}(2H, 2\beta\Delta) & \text{if } \beta > 0 \\ \dfrac{\Delta^{2H}}{2H} & \text{if } \beta = 0 \end{cases}$$

$$J(H,\Delta,\beta) = \begin{cases} \beta^{-H-1/2}\Gamma_{low}(H+1/2, \beta\Delta) & \text{if } \beta > 0 \\ \dfrac{\Delta^{H+1/2}}{H+1/2} & \text{if } \beta = 0 \end{cases}.$$

Proof. The proof is postponed to Appendix D.2 to ease the flow of the Section. □

Again, this result can be adapted to the case of MGRV models, as we can see in the following example.

Example 6.5.5 (Mixed generalized rough volatility model)**.** Consider the MGRV model, then we have that the ATMI skew for VIX options is given by

$$\lim_{T\to 0} \mathcal{S}_T^{VIX}(0) = \frac{\sqrt{2H}}{2}\left(\frac{\left(\delta\nu^2 + (1-\delta)\eta^2\right) G(H,\Delta,\beta)}{(\delta\nu + (1-\delta)\eta)\, J(H,\Delta,\beta)} - \frac{(\delta\nu + (1-\delta)\eta)\, J(H,\Delta,\beta)}{\Delta}\right).$$

Figure 6.5 shows the accuracy of the asymptotic limit with a Monte Carlo benchmark of $T = 10^{-4}$ and $\delta = 1$. The derivative was approximated using a central difference scheme. Notice that in the case $H = \frac{1}{2}$ and $\beta = 30$ the limit skew is positive, consistent with the results in [Guy20a].

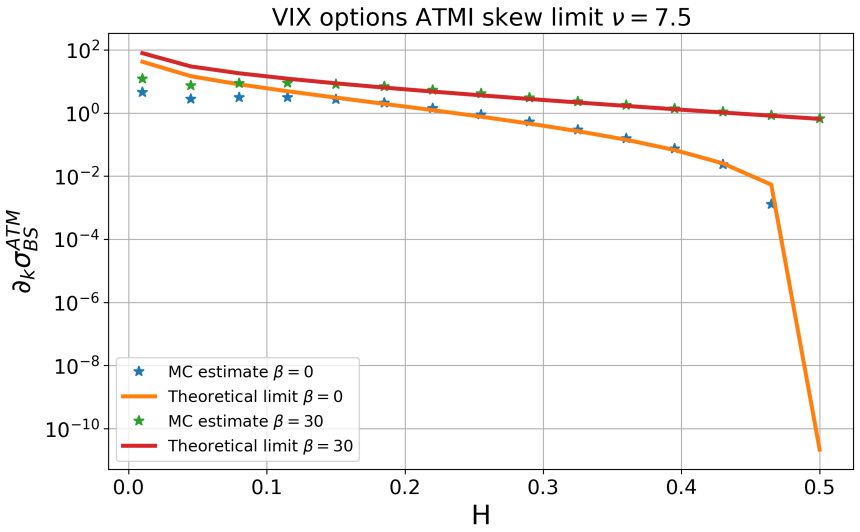


Figure 6.5: Generalized rough volatility model ATMI skew on VIX options with $\nu = 7.5$

Example 6.5.6 (The VIX skew in lognormal models)**.** Based on Figure 6.5 one may conclude that for small values of H and for big values of β one can achieve large skews, in line with the results in [Guy20a]. We show in Figure 6.6 that, when $\beta = 0$, the skew decays almost immediately. This behavior is consistent with the fact that lognormal models generate flat smiles, hence the skew should be zero. On the contrary, Figure 6.7 shows the effect of mixing lognormals (again with $\beta = 0$), which produces a much higher level of skew. This result is consistent with [DeM18] and [Guy20a], where the double exponential parametrization is studied in the framework of Gaussian processes with a power-law kernel. In particular we notice that in the mixed SABR case ($H = 1/2$ and $\beta = 0$), the skew is given by

$$\lim_{T\to 0} \mathcal{S}_T^{VIX}(0) = \frac{1}{2}\left(\frac{\left(\delta\nu^2 + (1-\delta)\eta^2\right)}{(\delta\nu + (1-\delta)\eta)} - (\delta\nu + (1-\delta)\eta)\right). \tag{6.20}$$

We clearly have that (6.20) is non-zero unless $\delta \in \{0, 1\}$ or $\nu = \eta$. Moreover, in Figure 6.7 the skew is approximately constant at this level for maturities up to 4 months.

Remark 6.5.7. In order to obtain the short-time approximating formulas we are inspired by the computations in Appendix D.2, which yield the following expression before taking the limit:

$$\begin{aligned}
&\mathcal{S}_T^{VIX}(0) \\
&\approx \frac{\sqrt{2H}}{\sqrt{T}\sqrt{\int_0^T K^2(T,\Delta,u)\mathrm{d}u}} \\
&\times \left(\frac{\left(\delta\nu^2 + (1-\delta)\eta^2\right)\int_0^T K(T,\Delta,s)\int_s^T K(T,\Delta,u) I(\Delta,T,s,u)\mathrm{d}u\mathrm{d}s}{(\delta\nu + (1-\delta)\eta)} \right. \\
&\left. - \frac{(\delta\nu + (1-\delta)\eta)\int_0^T K^2(T,\Delta,s)\int_s^T K^2(T,\Delta,u)\mathrm{d}u\mathrm{d}s}{\Delta} \right).
\end{aligned}$$

where $K(\cdot)$ and $I(\cdot)$ are defined in Appendix D.2. We observe that the approximation performs very well for maturities smaller than 6 months and could be useful to control the ATMI skew level of VIX in a given model.

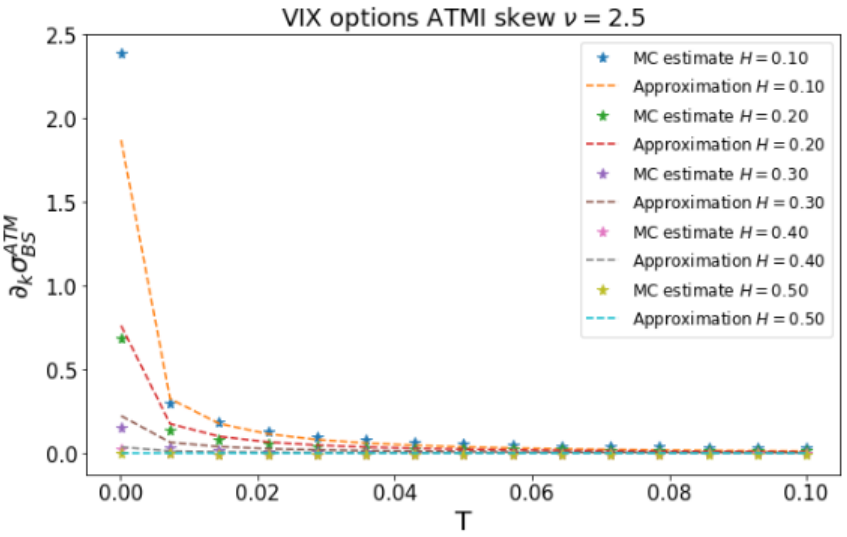


Figure 6.6: Generalized rough volatility model ATMI skew with $\nu = 2.5$

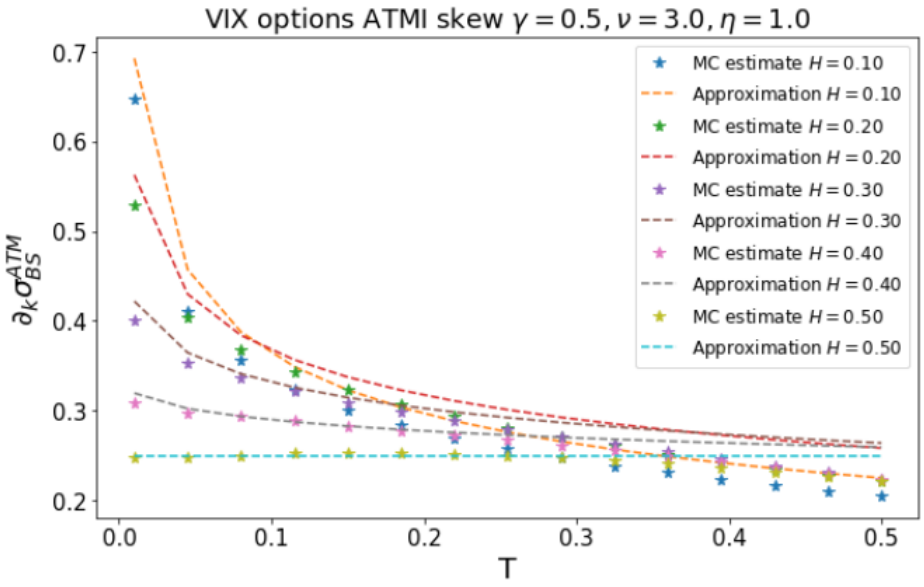


Figure 6.7: Mixed generalized rough volatility model ATMI skew with $(\delta, \nu, \eta) = (1/2, 3, 1)$

6.5.3 MGRV fit to empirical VIX options data

In this section we show how the MGRV model is able to generate the appropriate skew by calibrating the model to market smiles. To do so, we consider the model

$$V_t = \xi_0(t)\left(\delta\mathcal{E}\left(\nu\sqrt{2H}\mathcal{B}_t\right) + (1-\delta)\mathcal{E}\left(\eta\sqrt{2H}\mathcal{B}_t\right)\right), \quad \nu \geq 0,\ \eta \geq 0$$

where $\xi_0(t) = \sum_{i=1}^{n} \mathbb{1}_{\{T_{i-1}<t\leq T_i\}}\xi_i$, where $\{T_i\}_{i=1,\dots,n}$, $n \in \mathbb{N}$ are the available VIX futures dates with $T_0 = 0$ and ξ_i is chosen such that

$$VIX_{T_i}^{MKT} = \sqrt{\frac{1}{\Delta}\int_{T_i}^{T_i+\Delta} \mathbb{E}[V_s|\mathcal{F}_{T_i}]}$$

holds, where $VIX_{T_i}^{MKT}$ is the VIX future price corresponding to the maturity T_i. The calibration is performed using Algorithm 5.3.9 with 50.000 Monte Carlo paths.

Figures 6.8 and 6.9 show how the MGRV model is able to reproduce the positive VIX skew as proven in the previous section. We emphasize here that the aim of this chapter is not to show that the MGRV model is the best model to perform the VIX calibration task, as one can see in the shortest maturities in Figures 6.8 and 6.9. It is however, an example of a model for which we have obtained precise asymptotics and understood the behaviour of ATMI VIX options, without resorting to numerical schemes in the first place.

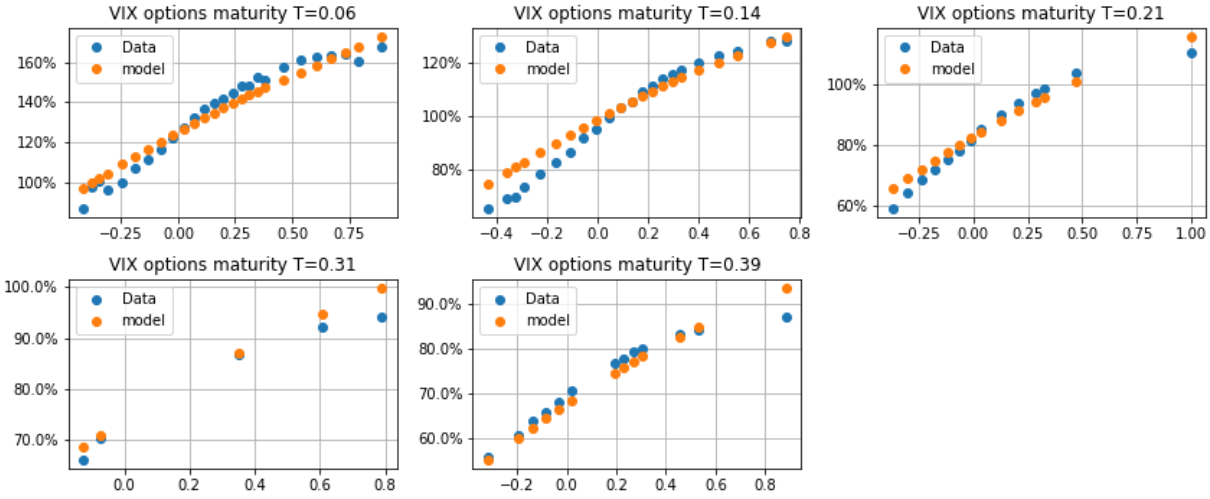


Figure 6.8: VIX options implied volatilities on 28/01/2015. Calibrated parameters $(\delta, \nu, \eta, H, \beta) = (0.45, 0.66, 3.31, 0.08, 0.02)$

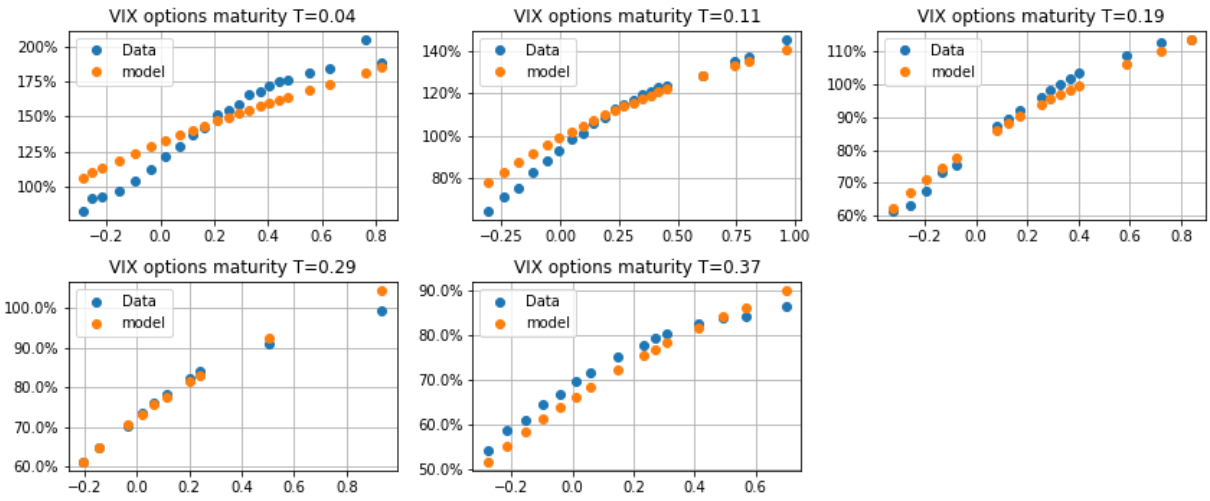


Figure 6.9: VIX options implied volatilities on 05/02/2015. Calibrated parameters $(\delta, \nu, \eta, H, \beta) = (0.42, 0.40, 2.96, 0.10, 0.01)$

6.6 Realized variance options

This section is devoted to the study of the ATMI short-time level and skew of RV options. Figure 6.10 shows a empirical term structure of the ATMI, which is usually close to a power law.

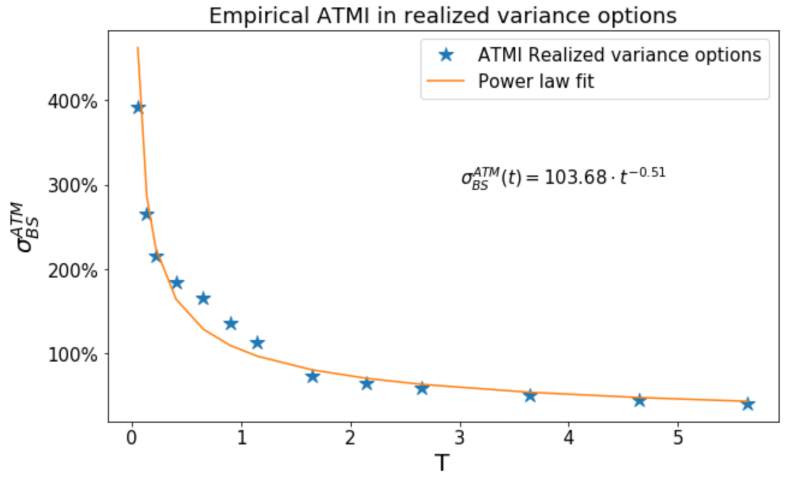


Figure 6.10: Realized variance ATMI for S&P 500 on February 27, 2018 (from OTC market quotes obtained through different brokers).

6.6.1 ATMI on realized variance options

In this section we study the short-time limit of the ATMI of RV options, whose behaviour differs from the ATMI of VIX options. Our main result is the following proposition.

Proposition 6.6.1. *Consider the process $v = f(Y)$, where f is a positive and bounded function in $\mathcal{C}^2$, and $Y = \int_0^t (t-s)^{H-\frac{1}{2}} g(t-s) \mathrm{d}W_s$, where $H \in (0, \frac{1}{2}]$ and $g \in \mathcal{C}_b^1$. Then we have that the short-time limit of the ATMI*

for RV options is given by

$$\lim_{T\to 0} T^{\frac{1}{2}-H} I_T^{RV}(0) \quad = \frac{f'(Y_0)g(0)}{(H+\frac{1}{2})\sqrt{2H+2}M_0^0}. \tag{6.21}$$

Proof. We may directly apply Corollary 6.4.5 since (H1)-(H4) hold with $\gamma = H - \frac{1}{2}$. Thus,

$$\begin{aligned}
&\lim_{T\to 0} T^{\frac{1}{2}-H} I_T^{RV}(0) \\
&= \lim_{T\to 0} \frac{1}{T^H}\mathbb{E}\left[\sqrt{\int_0^T (\phi_s^{RV})^2 \mathrm{d}s}\right] \\
&= \lim_{T\to 0} \frac{1}{T^{1+H}}\mathbb{E}\left[\sqrt{\int_0^T \left(\frac{1}{M_s^{T,RV}}\int_s^T E_s(D_s(V_u))\mathrm{d}u\right)^2 \mathrm{d}s}\right] \\
&= \frac{f'(Y_0)g(0)}{M_0^{0,RV}} \lim_{T\to 0} \frac{1}{T^{1+H}}\mathbb{E}\left[\sqrt{\int_0^T \left(\int_s^T (u-s)^{H-\frac{1}{2}}\mathrm{d}u\right)^2 \mathrm{d}s}\right] \\
&= \frac{f'(Y_0)g(0)}{(H+\frac{1}{2})M_0^{0,RV}} \lim_{T\to 0} \frac{1}{T^{1+H}}\mathbb{E}\left[\sqrt{\int_0^T (T-s)^{2H+1} \mathrm{d}s}\right] \\
&= \frac{f'(Y_0)g(0)}{(H+\frac{1}{2})\sqrt{2H+2}M_0^{0,RV}}.
\end{aligned} \tag{6.22}$$

□

Remark 6.6.2. We emphasize that (6.22) implies that, for short maturities the ATMI is of order $O(T^{H-\frac{1}{2}})$, which is consistent with real market data on equity markets as shown in Figure 6.10.

We can adapt this result to the case of MGRV models, as we can see in the following example:

Example 6.6.3 (Mixed generalized Rough volatility)**.** We consider again the MGRV model. Applying Corollary 6.4.5 we obtain the following short-time ATMI limit:

$$\begin{aligned}
&\lim_{T\to 0} T^{1/2-H} I_T^{RV}(0) \\
&= (\delta\nu + (1-\delta)\eta)\sqrt{2H} \\
&\quad \times \lim_{T\to 0} \frac{1}{T^{H+1}} \sqrt{\int_0^T \left(\int_s^T \exp\left(-\beta(u-s)\right)(u-s)^{H-1/2}\mathrm{d}u\right)^2 \mathrm{d}s} \\
&= \frac{(\delta\nu + (1-\delta)\eta)\sqrt{2H}}{(H+\frac{1}{2})\sqrt{2H+2}}.
\end{aligned}$$

Example 6.6.4 (A semi closed-form formula for the ATMI level of RV options)**.** Once again, we consider a short-time approximation of the ATMI level, inspired by the short-time limit given in Corollary 6.22. More precisely, we consider the following approximation:

$$I_T^{RV}(0) \approx \frac{f'(Y_0)}{V_0 T^{H+1}} \sqrt{\int_0^T \left(\int_s^T (u-s)^{H-\frac{1}{2}} g(u-s)\mathrm{d}u\right)^2 \mathrm{d}s} \tag{6.23}$$

In Figure 6.11 we present the numerical results for the approximation for different values of H and β in a MGRV model with $\delta = 1$. We observe a very sharp fit to the Monte Carlo estimates obtained using (2.11). Similar results are obtained in the case $\delta \neq 1$ (see Figure 6.12).

Remark 6.6.5. In light of the accuracy of the approximating scheme given in (6.23), this can be used in practice to both control the ATMI level in a model and calibrate the parameters H and β. In particular, when $\beta = 0$ this reduces to a power-law fit of order $H - \frac{1}{2}$ as previously shown in Example 6.6.3.

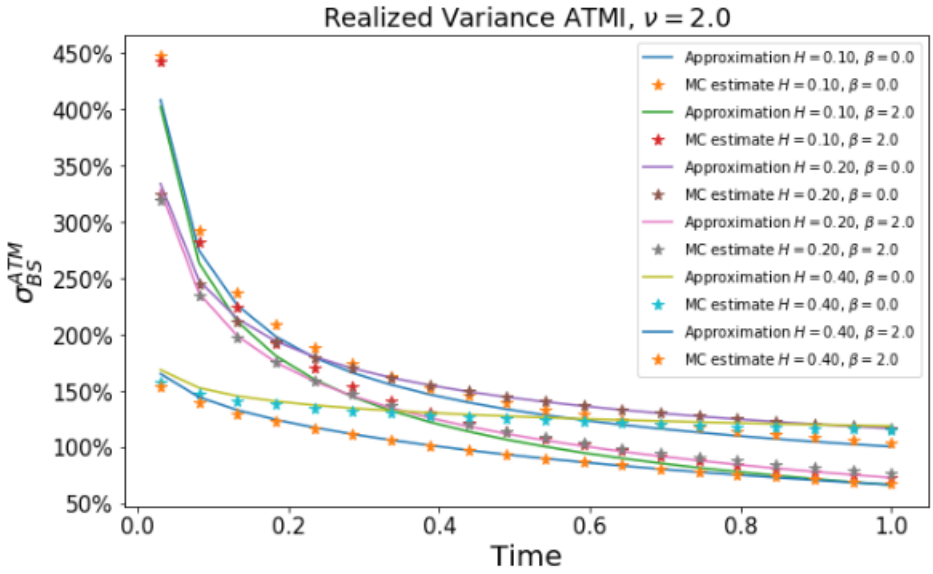


Figure 6.11: Realized variance option ATMI with $\nu = 2$

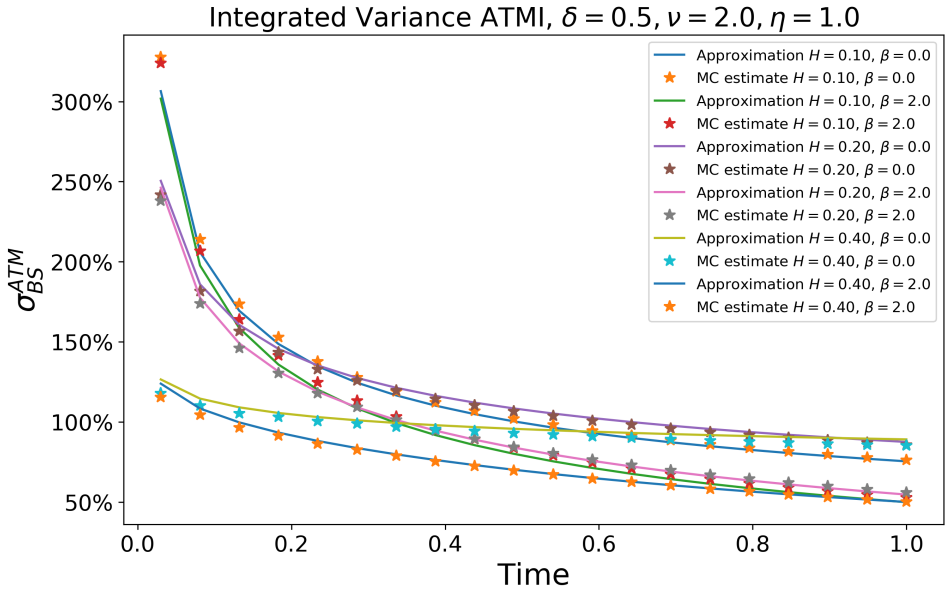


Figure 6.12: Realized variance option ATMI with $\delta = 0.5, \nu = 2$ and $\eta = 1$

6.6.2 ATMI skew for realized variance options

Proposition 6.6.6. *Consider an instantaneous variance model of the form*

$$V_t = V_0 f(Y_t)$$

where $Y_t = \int_0^t (t-s)^{H-\frac{1}{2}} g(t-s) \mathrm{d}W_s$, *with* $H \in (0, \frac{1}{2}]$ *and* $g \in \mathcal{C}_b^1$, *and* f *is a positive and bounded function in* $\mathcal{C}^2$, *and Then, the ATMI skew for RV options is given by*

$$\lim_{T \to 0} T^{\frac{1}{2}-H} \mathcal{S}_T^{RV}(0) = \left(\frac{f''(Y_0)\mathcal{I}(H)(2H+2)^{3/2}(H+1/2)}{f'(Y_0)} - \frac{f'(Y_0)}{V_0(2H+1)\sqrt{(2H+2)}} \right).$$

Proof. The proof and definition of $\mathcal{I}(H)$ are postponed to Appendix D to ease the flow of the Section. □

Example 6.6.7 (Mixed Rough Bergomi)**.** We consider the Mixed rough Bergomi model as a subclass of MGRV models with instantaneous variance given by:

$$V_t = V_0 \left(\delta \mathcal{E}\left(\nu \sqrt{2H} W_t^H \right) + (1-\delta) \mathcal{E}\left(\eta \sqrt{2H} W_t^H \right) \right),$$

where $W_t^H := \int_0^t (t-s)^{H-1/2} \mathrm{d}W_s$. Applying Proposition 6.6.6 we have that the ATMI skew for RV options is given by

$$\begin{aligned} &\lim_{T\to 0} T^{\frac{1}{2}-H} \mathcal{S}_T^{RV}(0) \\ &= \sqrt{2H} \left(\frac{\delta\nu^2 + (1-\delta)\eta^2}{\delta\nu + (1-\delta)\eta} \mathcal{I}(H)(2H+2)^{3/2}(H+1/2) - \frac{\delta\nu + (1-\delta)\eta}{(2H+1)\sqrt{(2H+2)}} \right). \end{aligned}$$

Figures 6.13 and 6.14 illustrate the accuracy of the asymptotic formula for different values of H.

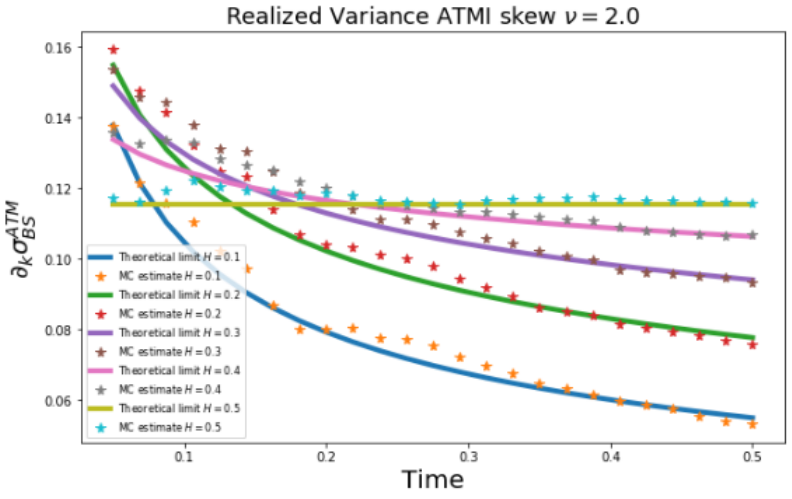


Figure 6.13: ATMI skew short time limit for RV options in a rough Bergomi model with $\nu = 2$.

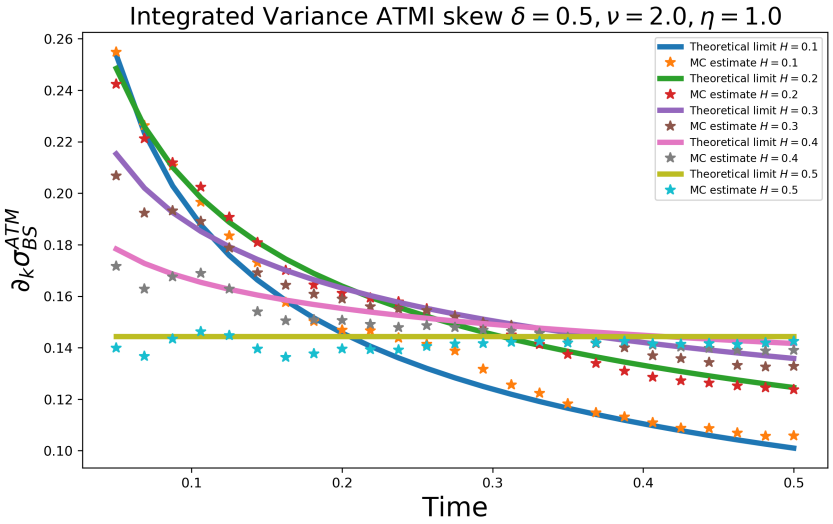


Figure 6.14: ATMI skew short time limit for RV options in a mixed rough Bergomi model with $(\delta, \nu, \eta) = (1/2, 2, 1)$.

6.7 Summary

In this Chapter we have characterised the short-time at-the-money asymptotics of options written on arbitrary payoffs, such as VIX and Realized Variance. In particular, we have found precise conditions for a model to produce a strictly positive skew on VIX options and formally proven the conjectures that 1) lognormal models have no skew 2) the Heston model produces a negative skew. Furthermore, we have discovered that for rough volatility models the short-time VIX skew tends to a constant value, whereas in realized variance options it explodes at a rate $T^{1/2-H}$. Our theoretical results are found to be constructive, in the sense that they allow to construct approximation formulas that perform, in general, very accurately for maturities $T < 1/2$.

7

Asymptotics for volatility derivatives in multi-factor rough volatility models

7.1 Introduction

Following the works by Alòs, León and Vives [ALV07], Gatheral, Jaisson, and Rosenbaum [GJR18] and Bayer, Friz and Gatheral [BFG16], rough volatility is becoming a new breed in financial modelling by generalising Bergomi's 'second generation' stochastic volatility models to a non-Markovian setting. The most basic form of (lognormal) rough volatility model is the so-called rough Bergomi model introduced in [BFG16]. Gassiat [Gas19] recently proved that such a model (under certain correlation regimes) generates true martingales for the spot process. The lack of Markovianity imposes numerous fundamental theoretical questions and practical challenges in order to make rough volatility usable in an industrial environment. On the theoretical side, Jacquier, Pakkanen, and Stone [JPS18] prove a large deviations principle for a rescaled version of the log stock price process. In this same direction, Bayer, Friz, Gulisashvili, Horvath and Stemper [Bay+19a], Forde and Zhang [FZ17], Horvath, Jacquier and Lacombe [HJL19] and most recently Friz, Gassiat and Pigato [FGP18] (to name a few) prove large deviations principles for a wide range of rough volatility models. On the practical side, competitive simulation methods are developed in Chapter 2 along with previous results by Bennedsen, Lunde and Pakkanen [BLP17] and McCrickerd and Pakkanen [MP18a]. Moreover, recent developments by Stone [Sto20] and Horvath, Muguruza and Tomas [HMT21] allow the use of neural networks for calibration; their calibration schemes are considerably faster and more accurate than existing methods for rough volatility models.

Crucially, the lack of a pricing PDE imposes a fundamental constraint on the comprehension and interpretation of rough volatility models driven, by Volterra-like Gaussian processes. The only current exception in the rough volatility literature is the rough Heston model, developed by El Euch and Rosenbaum [ER18; ER19], which allows a better understanding through the fractional PDE derived in [ER19]. Nevertheless, in this work our attention is turned to the class of models for which such a pricing PDE is unknown, and hence further theoretical results are required.

Perhaps, options on volatility itself are the most natural object to first analyse within the class of rough volatility models. In this direction, Jacquier, Martini, and Muguruza [JMM18] provide algorithms for pricing VIX options and futures. Horvath, Jacquier and Tankov [HJT20] further study VIX smiles in the presence of stochastic volatility of volatility combined with rough volatility. Nevertheless, the precise effect of model parameters (with particular interest in the Hurst parameter effect) on implied volatility smiles for VIX (or volatility derivatives in general) has been studied in Chapter 6.

The main focus of the Chapter is to derive the small-time behaviour of the realised variance process of the rough Bergomi model, as well as related but more complicated multi-factor rough volatility models, together with the small-time behaviour of options on realised variance. These results, which are interesting from a theoretical perspective, have practical applicability to the quantitative finance industry as they allow practitioners to better understand the Realised Variance smile, as well as the effect of correlation on the smile's linearity (or possibly convexity). To the best of our knowledge, this is the first work to study the small-time behaviour of options on realised variance. An additional major contribution of the present work is the numerical scheme used to compute the implied volatility smiles, using an accurate approximation of the rate function from a large deviations principle. In general rate functions are highly non-trivial to compute; our method is simple, intuitive, and accurate. The numerical methods are publicly available on GitHub: LDP-VolOptions.

Volatility options are becoming increasingly popular in the financial industry. For instance, VIX options' liquidity has consistently increased since its creation by the Chicago Board of Exchange (CBOE). One of the main popularity drivers is that volatility tends to be negatively correlated with the underlying dynamics, making it desirable for portfolio diversification. Due to the appealing nature of volatility options, their modelling has attracted the attention of many academics such as Carr, Geman, Madan and Yor [Car+05], Carr and Lee [CL09] to name a few.

For a log stock price process X defined as $X_t = -\frac{1}{2}\int_0^t V_s \mathrm{d}s + \int_0^t \sqrt{V_s}\mathrm{d}B_s, X_0 = 0$, where B is standard Brownian motion, we denote the quadratic variation of X at time t by $\langle X\rangle_t$. Then, the core object to analyse in this setting is the realised variance option with payoff

$$\left(\frac{1}{T}\int_0^T \mathrm{d}\langle X\rangle_s - K\right)^+, \tag{7.1}$$

which in turn defines the risk neutral density of the realised variance. In this work, we analyse the short time behaviour of the implied volatility given by (7.1) for a number of (rough) stochastic volatility models by means of large deviation techniques. We specifically focus on the construction of correlated factors and their effect on the distribution of the realised variance. We find the results of this Chapter consistent with that Chapter 6, which also help us characterise in close-form the implied volatility around the money. Moreover, we also obtain some asymptotic results for VIX options.

While implied volatilities for options on equities are typically convex functions of log-moneyness, giving them their "smile" moniker, implied volatility smiles for options on realised variance tend to be linear. Options on integrated variance are OTC products, and so their implied volatility smiles are not publicly available. VIX smiles are, however, and provide a good proxy for integrated variance smiles; see Figure 7.1 below for evidence of their linearity. The data also indicates both a power-law term structure ATM and its skew.

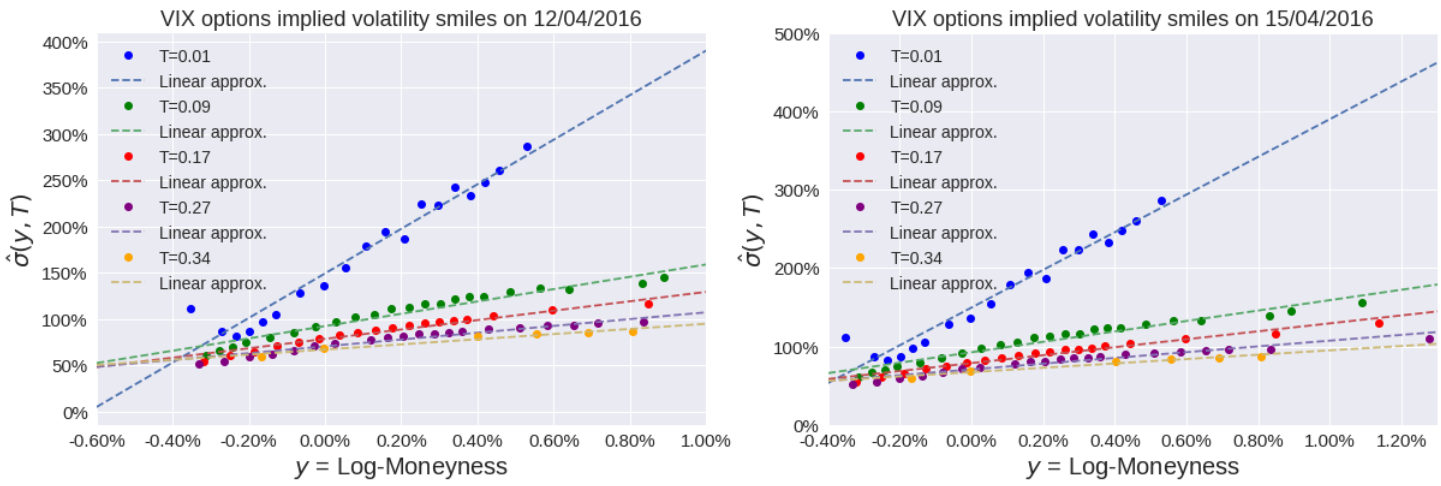


Figure 7.1: Implied volatility smiles for Call options on VIX for small maturities, close to the money. Data provided by OptionMetrics.

In spite of most of the literature agreeing on the fact that more than a single factor is needed to model volatility (see Bergomi's [Ber16] two-factor model, Avellaneda and Papanicolaou [AP19] or Horvath, Jacquier and Tankov [HJT20] for instance), there is no in-depth analysis on how to construct these (correlated) factors, nor the effect of correlation on the price of volatility derivatives and their corresponding implied volatility smiles. Our aim is to understand multi-factor models and analyse the effect of factors in implied volatility smiles. This work, to the best of our knowledge is the first to address such questions, which are of great interest to practitioners in the quantitative finance industry; it is also the first to provide a rigorous mathematical analysis of the small-time behaviour of options on integrated variance in rough volatility models.

The structure of the Chapter is as follows. Section 7.2 introduces the models, the rough Bergomi model and two closely related processes, whose small-time realised variance behaviour we study; the main results are given in Section 7.3. Section 7.4 is dedicated to numerical examples of the results attained in Section 7.3. In Section 7.5 we introduce a general variance process, which includes the rough Bergomi model for a specific choice of kernel, and briefly investigate the small-noise behaviour of VIX options in this general setting. Motivated by the numerical examples in Section 7.4, we propose a simple and very feasible approximation for the density of the realised variance for the mixed rough Bergomi model (see (7.6)) in Appendix E.1. The proofs of the main results are given in Appendix E.2; the details of the numerics are given in Appendix E.3.

Notations: Let $\mathbb{R}_+ := [0, +\infty)$ and $\mathbb{R}_+^* := (0, +\infty)$. For some index set $\mathcal{T} \subseteq \mathbb{R}_+$, the notation $L^2(\mathcal{T})$ denotes the space of real-valued square integrable functions on $\mathcal{T}$, and $\mathcal{C}(\mathcal{T}, \mathbb{R}^d)$ the space of $\mathbb{R}^d$-valued continuous functions on $\mathcal{T}$. $\mathcal{E}$ denotes the Wick stochastic exponential.

7.2 A showcase of rough volatility models

In this section we introduce the models that will be considered in the forthcoming computations. We shall always work on a given filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. For notational convenience, we introduce

$$Z_t := \int_0^t K_\alpha(s,t)\mathrm{d}W_s, \qquad \text{for any } t \in \mathcal{T}, \tag{7.2}$$

where $\alpha \in \left(-\frac{1}{2}, 0\right)$, W a standard Brownian motion, and where the kernel $K_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ reads

$$K_\alpha(s,t) := \eta\sqrt{2\alpha+1}(t-s)^\alpha, \qquad \text{for all } 0 \leq s < t, \tag{7.3}$$

for some strictly positive constant η. Note that, for any $t \geq 0$, the map $s \mapsto K_\alpha(s,t)$ belongs to $L^2(\mathcal{T})$, so that the stochastic integral (7.2) is well defined. We also define an analogous multi-dimensional version of (7.2) by

$$\mathcal{Z}_t := \left(\int_0^t K_\alpha(s,t)\mathrm{d}W_s^1, ..., \int_0^t K_\alpha(s,t)\mathrm{d}W_s^m\right) := \left(\mathcal{Z}_t^1, ..., \mathcal{Z}_t^m\right), \qquad \text{for any } t \in \mathcal{T}, \tag{7.4}$$

where $W^1, ..., W^m$ are independent Brownian motions.

Model 7.2.1 (Rough Bergomi)**.** *The rough Bergomi model, where X is the log stock price process and V is the instantaneous variance process, is then defined (see [BFG16]) as*

$$\begin{aligned} X_t &= -\frac{1}{2}\int_0^t V_s\mathrm{d}s + \int_0^t \sqrt{V_s}\mathrm{d}B_s, \qquad X_0 = 0, \\ V_t &= V_0 \exp\left(Z_t - \frac{\eta^2}{2}t^{2\alpha+1}\right), \qquad v_0 > 0, \end{aligned} \tag{7.5}$$

where the Brownian motion B is defined as $B := \rho W + \sqrt{1-\rho^2}W^\perp$ for $\rho \in [-1,1]$ and some standard Brownian motion $W^\perp$ independent of W.

Remark 7.2.2. The process $\log V$ has a modification whose sample paths are almost surely locally γ-Hölder continuous, for all $\gamma \in \left(0, \alpha + \frac{1}{2}\right)$ [JPS18, Proposition 2.2]. For rough volatility models involving a fractional Brownian motion, the sample path regularity of the log volatility process is referred to in terms of the Hurst parameter H; recall that the fractional Brownian motion has sample paths that are γ-Hölder continuous for any $\gamma \in (0, H)$ [Bia+08, Theorem 1.6.1]. By identification, therefore, we have that $\alpha = H - 1/2$.

Model 7.2.3 (Mixed rough Bergomi)**.** *The mixed rough Bergomi model is given in terms of log stock price process X and instantaneous variance process $V^{(\gamma,\nu)}$ as*

$$\begin{aligned} X_t &= -\frac{1}{2}\int_0^t V_s^{(\gamma,\nu)}\mathrm{d}s + \int_0^t \sqrt{V_s^{(\gamma,\nu)}}\mathrm{d}B_s, \qquad X_0 = 0, \\ V_t^{(\gamma,\nu)} &= V_0 \textstyle\sum_{i=1}^n \gamma_i \exp\left(\frac{\nu_i}{\eta} Z_t - \frac{\nu_i^2}{2} t^{2\alpha+1}\right), \qquad v_0 > 0 \end{aligned} \tag{7.6}$$

where $\gamma := (\gamma_1, ..., \gamma_n) \in [0,1]^n$ such that $\sum_{i=1}^n \gamma_i = 1$ and $\nu := (\nu_1, ..., \nu_n) \in \mathbb{R}^n$, such that $0 < \nu_1 < ... < \nu_n$.

The above modification of the rough Bergomi model, inspired by Bergomi [Ber08], allows a bigger slope (hence bigger skew) on the implied volatility of variance/volatility options to be created, whilst maintaining a tractable instantaneous variance form. This will be made precise in Section 7.4.2.

Model 7.2.4 (Mixed multi-factor rough Bergomi)**.** *The mixed rough Bergomi model is given in terms of log stock price process X and instantaneous variance process $V^{(\gamma,\nu,\Sigma)}$ as*

$$\begin{aligned} X_t &= -\frac{1}{2}\int_0^t V_s^{(\gamma,\nu,\Sigma)}\mathrm{d}s + \int_0^t \sqrt{V_s^{(\gamma,\nu,\Sigma)}}\mathrm{d}B_s, \qquad X_0 = 0, \\ V_t^{(\gamma,\nu,\Sigma)} &= V_0 \textstyle\sum_{i=1}^n \gamma_i \mathcal{E}\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}_t\right), \qquad V_0 > 0, \end{aligned} \tag{7.7}$$

where $\gamma := (\gamma_1, ..., \gamma_n) \in [0,1]^n$ such that $\sum_{i=1}^n \gamma_i = 1$. The vector $\nu^i = (\nu_1^i, ..., \nu_m^i) \in \mathbb{R}^m$ satisfies $0 < \nu_1^i < ... < \nu_m^i$ for all $i \in \{1, ..., n\}$. In addition, $\mathrm{L}_i \in \mathbb{R}^{m\times m}$ is a lower triangular matrix such that $\mathrm{L}_i\mathrm{L}_i^T =: \Sigma_i$ is a positive definite matrix for all $i \in \{1, ..., n\}$, denoting the covariance matrix.

For all results involving models (7.5), (7.6), and (7.7) we fix $\mathcal{T} = [0,1]$; minor adjustments to the proofs yield analogous results for more general $\mathcal{T}$. We additionally define $\beta := 2\alpha + 1 \in (0,1)$ for notational convenience.

Remark 7.2.5. In models (7.5)-(7.7) we have considered a flat or constant initial forward variance curve $V_0 > 0$. However, our framework can be easily extended to functional forms $V_0(\cdot) : \mathcal{T} \mapsto \mathbb{R}_+$ via the Contraction Principle, see Appendix E.4, as long as the mapping is continuous.

Remark 7.2.6. The reader may already have realised that the mixed multi-factor rough Bergomi defined in (7.7) is indeed general enough to cover both (7.5) and (7.6). However, we provide our theoretical results in an orderly fashion starting from (7.5) and finishing with (7.7), which we find the most natural way to increase the complexity of the model.

7.3 Small-time results for options on integrated variance

We start our theoretical analysis by considering options on realised variance, which we also refer to as integrated variance and RV interchangeably. We recall that volatility is not directly observable, nor a tradeable asset. Options on realised variance, however, exist and are traded as OTC products. Below are two examples of the payoff structure of such products:

$$(i)(RV(V)(T) - K)^+, \quad (ii)(\sqrt{RV(V)(T)} - K)^+, \quad \text{where } T, K \geq 0. \tag{7.8}$$

where we define the following $\mathcal{C}(\mathcal{T})$ operator

$$RV(f)(\cdot) : f \mapsto \frac{1}{\cdot}\int_0^\cdot f(s)\mathrm{d}s, \quad RV(f)(0) := f(0), \tag{7.9}$$

and v represents the instantaneous variance in a given stochastic volatility model. Note that $RV(V)(0) = V_0$.

Remark 7.3.1. As shown by Neuberger [Neu94], we may rewrite the variance swap in terms of the log contract as

$$\mathbb{E}[RV(V)(T)] = \mathbb{E}\left[\frac{1}{T}\int_0^T V_s \mathrm{d}s\right] = \mathbb{E}\left[-2\frac{X_T}{T}\right] \tag{7.10}$$

where $\mathbb{E}[\cdot]$ is taken under the risk-neutral measure and $S = \exp(X)$ is a risk-neutral martingale (assuming interest rates and dividends to be null). Therefore, the risk neutral pricing of $RV(V)(T)$ or options on it is fully justified by (7.10).

7.3.1 Small-time results for the rough Bergomi model

Proposition 7.3.2. *The set $\mathscr{H}^{K_\alpha} := \{\int_0^\cdot K_\alpha(s,\cdot)f(s)\mathrm{d}s : f \in L^2(\mathcal{T})\}$ defines the reproducing kernel Hilbert space for Z with inner product $\langle \int_0^\cdot K_\alpha(s,\cdot)f_1(s)\mathrm{d}s, \int_0^\cdot K_\alpha(s,\cdot)f_2(s)\mathrm{d}s\rangle_{\mathscr{H}^{K_\alpha}} = \langle f_1, f_2\rangle_{L^2(\mathcal{T})}$.*

Proof. See [JPS18, Theorem 3.1]. □

Before stating Theorem 7.3.3, we define the following function $\Lambda^Z : \mathcal{C}(\mathcal{T}) \to \mathbb{R}_+$ as $\Lambda^Z(\mathrm{x}) := \frac{1}{2}\|\mathrm{x}\|^2_{\mathscr{H}^{K_\alpha}}$, and if $\mathrm{x} \notin \mathscr{H}^{K_\alpha}$ then $\Lambda^Z(\mathrm{x}) = +\infty$.

Theorem 7.3.3. *The variance process $(V_\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function $\Lambda^V(x) := \Lambda^Z\left(\log\left(\frac{x}{V_0}\right)\right)$, where $\Lambda^V(V_0) = 0$ and $x \in \mathcal{C}(\mathcal{T})$.*

Proof. To ease the flow of the Chapter the proof is postponed to Appendix E.2.1. □

Corollary 7.3.4. *The integrated variance process $(RV(V)(t))_{t\in\mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^V$ defined as $\hat{\Lambda}^V(\mathrm{y}) := \inf\left\{\Lambda^V(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\right\}$, where $\hat{\Lambda}^V(V_0) = 0$.*

Proof. As proved in Theorem 7.3.3, the process $(V_\varepsilon)_{\varepsilon>0}$ satisfies a pathwise large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero. For small perturbations $\delta^V \in \mathcal{C}(\mathcal{T})$, we have

$$\left\|RV(V+\delta^V)(t) - RV(\delta^V)(t)\right\|_\infty \leq \sup_{t\in\mathcal{T}} \frac{1}{t}\left|\int_0^t \delta^V(s)\mathrm{d}s\right| \leq M,$$

where $M = \sup_{t\in\mathcal{T}}|\delta^V(t)|$, which is finite as $\delta^V \in \mathcal{C}(\mathcal{T})$. Clearly M tends to zero as δ^V tend to zero, and hence the operator RV is continuous with respect to the sup norm on $\mathcal{C}(\mathcal{T})$. Therefore we can apply the Contraction Principle (see Appendix E.4), and consequently the integrated variance process $RV(V_\varepsilon)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero. Clearly $RV(V_\varepsilon)(t) = RV(V)(\varepsilon t)$, for all $t \in \mathcal{T}$, and so setting $t = 1$ and mapping ε to t then yields the result. By definition, $\hat{\Lambda}^V(\mathrm{y}) := \inf\left\{\Lambda^V(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\right\}$. If we choose $\mathrm{x} \equiv V_0$ then clearly $V_0 = RV(\mathrm{x})(1)$, and $\Lambda^V(\mathrm{x}) = 0$. Since Λ^V is a norm, it is a non-negative function and therefore $\hat{\Lambda}^V(V_0) = 0$. This concludes the proof. □

Remark 7.3.5. Corollary 7.3.4 can be applied to a large number of existing results on large deviations for (rough) variance processes to get a large deviations result for the integrated (rough) variance process; for example Forde and Zhang [FZ17] and Horvath, Jacquier, and Lacombe [HJL19].

Corollary 7.3.6. *The rate function $\hat{\Lambda}^V$ is continuous.*

Proof. Indeed, as a rate function, $\hat{\Lambda}^V$ is lower semi-continuous. Moreover, as Λ^V is continuous, one can use similar arguments to [FZ17, Corollary 4.6], and deduce that $\hat{\Lambda}^V$ is upper semi-continuous, and hence is continuous. □

Before stating results on the small-time behaviour of options on integrated variance, we state that the integrated variance process $RV(V)$ satisfies a large deviations principle on $\mathbb{R}$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^V(\mathrm{e}^{\cdot})$. Then, the small-time behaviour of such options can be obtained as an application of Corollary 7.3.4.

Corollary 7.3.7. *For log moneyness $k := \log \frac{K}{RV(V)(0)} \neq 0$, the following equality holds true for Call options on integrated variance*

$$\lim_{t \downarrow 0} t^{\beta} \log \mathbb{E}\left[\left(RV(V)(t) - \mathrm{e}^{k}\right)^{+}\right] = -\mathrm{I}(k), \tag{7.11}$$

where I *is defined as as* $\mathrm{I}(x) := \inf_{y>x} \hat{\Lambda}^V(\mathrm{e}^y)$ *for* $x > 0$, $\mathrm{I}(x) := \inf_{y<x} \hat{\Lambda}^V(\mathrm{e}^y)$ *for* $x < 0$.

Similarly, for log moneyness $k := \log \frac{K}{\sqrt{RV(V)(0)}} \neq 0$,

$$\lim_{t \downarrow 0} t^{\beta} \log \mathbb{E}\left[\left(\sqrt{RV(V)(t)} - \mathrm{e}^{k}\right)^{+}\right] = -\bar{\mathrm{I}}(k), \tag{7.12}$$

where $\bar{\mathrm{I}}$ *is defined analogously as* $\bar{\mathrm{I}}(x) := \inf_{y>x} \hat{\Lambda}^V(\mathrm{e}^{2y})$ *for* $x > 0$ *and* $\bar{\mathrm{I}}(x) := \inf_{y<x} \hat{\Lambda}^V(\mathrm{e}^{2y})$ *for* $x < 0$.

Proof. The proof is postponed to Appendix E.2.2. □

As with Call options on stock price processes, we can define and study the implied volatility of such options. In the case of (7.8)(i) we define the implied volatility $\hat{\sigma}(T, k)$ to be the solution to

$$\mathbb{E}[(RV(V)(T) - e^{k})^{+}] = C_{BS}(RV(V)(0), k, T, \hat{\sigma}(T, k)), \tag{7.13}$$

where C_{BS} denotes the Call price in the Black-Scholes model. Using Corollary 7.3.7, we deduce the small-time behaviour of the implied volatility $\hat{\sigma}$, as defined in (7.13).

Corollary 7.3.8. *The small-time asymptotic behaviour of the implied volatility is given by the following limit, for a log moneyness $k \neq 0$:*

$$\lim_{t \downarrow 0} t^{1-\beta} \hat{\sigma}^2(t, k) =: \hat{\sigma}^2(k) = \frac{k^2}{2\mathrm{I}(k)}.$$

Proof. The log integrated variance process $\log RV(V)$ satisfies a large deviations principle with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^V(\mathrm{e}^{\cdot})$, which is continuous. Therefore, it follows that

$$\lim_{t \downarrow 0} t^{\beta} \log \mathbb{P}(RV(V)(t) \geq \mathrm{e}^{k}) = -\mathrm{I}(k).$$

In the Black Scholes model, i.e. a geometric Brownian motion with $S_0 = RV(V)(0)$ with constant volatility ξ, we have the following small-time implied volatility behaviour:

$$\lim_{t \downarrow 0} \xi^2 t \log \mathbb{P}(RV(v)(t) \geq \mathrm{e}^{k}) = -\frac{k^2}{2}.$$

We then apply [GL14, Corollary 7.1], identifying $\xi \equiv \hat{\sigma}(k, t)$, to conclude. □

Remark 7.3.9. Notice that the level of implied volatility in Corollary 7.3.8 has a power law behaviour as a function of time to maturity. This power law is of order $\sqrt{t^{\beta-1}}$, which is consistent with the at-the-money RV implied volatility results in Chapter 6, where the order is found to be $t^{H-1/2}$ using Malliavin Calculus techniques. Recall that $\beta = 2\alpha + 1$, and $\alpha = H - 1/2$ by Remark 7.2.2.

7.3.2 Small-time results for the mixed rough Bergomi model

Minor adjustments to Theorem 7.3.3 give the following small-time result for the mixed variance process $V^{(\gamma,\nu)}$ introduced in Model (7.6); the proof is given in Appendix E.2.3.

Theorem 7.3.10. *The mixed variance process $(V_\varepsilon^{(\gamma,\nu)})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function*

$$\Lambda^{(\gamma,\nu)}(\mathrm{x}) := \inf\{\Lambda^Z(\frac{\eta}{\nu_1}\mathrm{y}) : \mathrm{x}(\cdot) = V_0 \sum_{i=1}^n \gamma_i e^{\frac{\nu_i}{\nu_1}\mathrm{y}(\cdot)}\},$$

satisfying $\Lambda^{(\gamma,\nu)}(V_0) = 0$.

By Remark 7.3.5, we immediately get the following result for the small-time behaviour of the integrated mixed variance process $RV(V^{(\gamma,\nu)})$.

Corollary 7.3.11. *The integrated mixed variance process $(RV\left(V^{(\gamma,\nu)}\right)(t))_{t\in\mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{y}) := \inf\left\{\Lambda^{(\gamma,\nu)}(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\right\}$, where $\tilde{\Lambda}^{(\gamma,\nu)}(V_0) = 0$.*

To get the small-time implied volatility result, analogous to Corollary 7.3.8, we need the following Lemma, which is used in place of (E.3). The remainder of the proof then follows identically.

Lemma 7.3.12. *For all $t \in \mathcal{T}$ and $q > 1$ we have*

$$\mathbb{E}\left[\left(RV\left(V^{(\gamma,\nu)}\right)(t)\right)^q\right] \leq \frac{v_0^q n^{q-1}}{t^{q-1}} \exp\left(\frac{(\nu^*)^2}{2\eta^2}\left(q^2-q\right)t^{2\alpha+1}\right),$$

where $\nu^* = \max\{\nu_1, ..., \nu_n\}$.

Proof. First we note that by Hölder's inequality $(\sum_{i=1}^n x_i)^q \leq n^{q-1}\sum_{i=1}^n (x_i)^q$, for $x_i > 0$. Since, $\gamma_i \leq 1$ for $i = 1, ..., n$, we obtain

$$\mathbb{E}\left[\left(RV\left(V^{(\gamma,\nu)}\right)(t)\right)^q\right] \leq \frac{V_0^q}{t^q} n^{q-1} \sum_{i=1}^n \int_0^t \mathbb{E}\left[\exp\left(\frac{q\nu_i}{\eta}Z_s - \frac{q\nu_i^2}{2\eta^2}s^{2\alpha+1}\right)\right] \mathrm{d}s \leq \frac{V_0^q}{t^{q-1}} n^{q-1} \sum_{i=1}^n \exp\left(\frac{\nu_i^2}{2\eta^2}\left(q^2-q\right)t^{2\alpha+1}\right).$$

Choosing $\nu^* = \max\{\nu_1, ..., \nu_n\}$ the result directly follows.

□

Corollary 7.3.13. *For log moneyness $k := \log\frac{K}{RV(V^{(\gamma,\nu)})(0)} \neq 0$, the following equality holds true for Call options on integrated variance in the mixed rough Bergomi model:*

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(RV(V^{(\gamma,\nu)})(t) - \mathrm{e}^k\right)^+\right] = -\mathrm{I}(k), \tag{7.14}$$

where I *is defined as* $\mathrm{I}(x) := \inf_{y>x} \tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{e}^y)$ *for* $x > 0$, $\mathrm{I}(x) := \inf_{y<x} \tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{e}^y)$ *for* $x < 0$.

Similarly, for log moneyness $k := \log\frac{K}{\sqrt{RV(V^{(\gamma,\nu)})(0)}} \neq 0$,

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(\sqrt{RV(V^{(\gamma,\nu)})(t)} - \mathrm{e}^k\right)^+\right] = -\bar{\mathrm{I}}(k), \tag{7.15}$$

where $\bar{\mathrm{I}}$ *is defined analogously as* $\bar{\mathrm{I}}(x) := \inf_{y>x} \tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{e}^{2y})$ *for* $x > 0$ *and* $\bar{\mathrm{I}}(x) := \inf_{y<x} \tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{e}^{2y})$ *for* $x < 0$.

Proof. Follows directly from Lemma 7.3.12 and proof of Corollary 7.3.7. □

The small-time implied volatility behaviour for the mixed rough Bergomi model is then given by Corollary 7.3.8, where the function I is defined in terms of the rate function $\tilde{\Lambda}^{(\gamma,\nu)}$, as in Corollary 7.3.13, in this case.

7.3.3 Small-time results for the multi-factor rough Bergomi model

The asymptotic behaviour of the multi-factor rough Bergomi model (7.7) for the variance process is given in Theorem 7.3.14 below; note that Λ^m is the rate function associated to the reproducing kernel Hilbert space of the measure induced by $\mathcal{Z}$ on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$, denoted $\mathscr{H}_m$.The proof is given in Appendix E.2.4.

Theorem 7.3.14. *The sequence of processes* $\left(V_{\varepsilon\cdot}^{(\gamma,\nu,\Sigma)}\right)_{\varepsilon>0}$ *satisfies a large deviations principle on* $\mathcal{C}(\mathcal{T})$ *with speed* $\varepsilon^{-\beta}$ *and rate function*

$$\Lambda^{(\gamma,\nu,\Sigma)}(\mathrm{y}) = \inf\left\{\Lambda^m(\mathrm{x}) : \mathrm{x} \in \mathscr{H}_m, \mathrm{y} = V_0 \sum_{i=1}^n \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathrm{x}(1)\right)\right\},$$

satisfying $\Lambda^{(\gamma,\nu,\Sigma)}(V_0) = 0$.

As with the mixed variance process, Remark 7.3.5 gives us the following small-time result for $RV(V^{(\gamma,\nu,\Sigma)})$ straight off the bat.

Corollary 7.3.15. *The integrated variance process* $(RV\left(V^{(\gamma,\nu,\Sigma)}\right)(t))_{t\in\mathcal{T}}$ *in the multi-factor Bergomi model satisfies a large deviations principle on* $\mathbb{R}_+^*$ *as* t *tends to zero, with speed* $t^{-\beta}$ *and rate function*

$$\tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(\mathrm{y}) := \inf\left\{\Lambda^{(\gamma,\nu,\Sigma)}(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\right\},$$

where $\tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(V_0) = 0$.

We now establish the small-time behaviour for Call options on realised variance in Corollary 7.3.17, by adapting the proof of Corollary 7.3.7 as in the previous subsection. To do so we use Lemma 7.3.16 in place of (E.3). Then we attain the small-time implied volatility behaviour for the multi-factor rough Bergomi model in Corollary 7.3.8, where the function I is given by Corollary 7.3.17.

Lemma 7.3.16. *For all* $t \in \mathcal{T}$ *and* $q > 1$ *we have*

$$\mathbb{E}\left[\left(RV\left(V^{(\gamma,\nu,\Sigma)}\right)(t)\right)^q\right] \leq \frac{v_0^q n^{q-1}}{t^{q-1}} \exp\left(\frac{(\nu^*)^2}{2\eta^2}\left(q^2-q\right)t^{2\alpha+1}\right),$$

where $\nu^* = \max\{\nu_1, ..., \nu_n\}$

Proof. First we note that by Hölder's inequality $(\sum_{i=1}^n x_i)^q \leq n^{q-1}\sum_{i=1}^n (x_i)^q$, for $x_i > 0$. Since, $\gamma_i \leq 1$ for $i = 1, ..., n$, we obtain

$$\mathbb{E}\left[\left(RV\left(V^{(\gamma,\nu,\Sigma)}\right)(t)\right)^q\right] \leq \frac{V_0^q}{t^q} n^{q-1} \sum_{i=1}^n \int_0^t \mathbb{E}\left[\mathcal{E}\left(\frac{\nu^i}{\eta}\cdot \mathrm{L}_i\mathcal{Z}_s\right)^q\right] \mathrm{d}s \leq \frac{V_0^q}{t^{q-1}} n^{q-1} \sum_{i=1}^n \exp\left(\frac{\nu_i^2}{2\eta^2}\left(q^2-q\right)t^{2\alpha+1}\right).$$

Choosing $\nu^* = \max\{\nu_1, ..., \nu_n\}$ the result directly follows. □

Corollary 7.3.17. *For log moneyness* $k := \log\frac{K}{RV(V^{(\gamma,\nu,\Sigma)})(0)} \neq 0$, *the following equality holds true for Call options on integrated variance in the multi-factor rough Bergomi model:*

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(RV(V^{(\gamma,\nu,\Sigma)})(t) - \mathrm{e}^k\right)^+\right] = -\mathrm{I}(k), \tag{7.16}$$

where I *is defined as* $\mathrm{I}(x) := \inf_{y>x} \tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(\mathrm{e}^y)$ *for* $x > 0$, $\mathrm{I}(x) := \inf_{y<x} \tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(\mathrm{e}^y)$ *for* $x < 0$.

Similarly, for log moneyness $k := \log\frac{K}{\sqrt{RV(V^{(\gamma,\nu,\Sigma)})(0)}} \neq 0$,

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(\sqrt{RV(V^{(\gamma,\nu,\Sigma)})(t)} - \mathrm{e}^k\right)^+\right] = -\bar{\mathrm{I}}(k), \tag{7.17}$$

where $\bar{\mathrm{I}}$ *is defined analogously as* $\bar{\mathrm{I}}(x) := \inf_{y>x} \tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(\mathrm{e}^{2y})$ *for* $x > 0$ *and* $\bar{\mathrm{I}}(x) := \inf_{y<x} \tilde{\Lambda}^{(\gamma,\nu,\Sigma)}(\mathrm{e}^{2y})$ *for* $x < 0$.

Proof. Follows directly from Lemma 7.3.16 and the proof of Corollary 7.3.7. □

7.4 Numerical results

In this section we present numerical results for each of the three models given in Section 7.2. We also analyse the effect of each parameters in the implied volatility smile. Numerical experiments and codes are provided on GitHub: LDP-VolOptions .

7.4.1 RV smiles for rough Bergomi

We begin with numerical results for the rough Bergomi model (7.5) using Corollary 7.3.8. For the detailed numerical method we refer the reader to Appendix E.3. In Figure 7.3, we represent the rate function given in Corollary 7.3.4, which is the fundamental object to compute numerically. In particular, we notice that $\hat{\Lambda}^V$ is convex; a rigorous mathematical proof of this statement is left for future research.

More interestingly, in Figure 7.2 we provide a comparison of Corollary 7.3.8 with respect to a benchmark generated by Monte Carlo simulations, and see all smiles to follow a linear trend. In particular, we notice that Corollary 7.3.8 provides a surprisingly accurate estimate, even for relatively large maturities. As a further numerical check, in Figure 7.4 we compare recent results with the close-form at-the-money asymptotics given in Chapter 6 and once again find the correct convergence, suggesting a consistent numerical framework.

7.4.2 RV smiles for mixed rough Bergomi

We now consider the mixed rough Bergomi model (7.6) in a simplified form given by $V_t = V_0 \left(\gamma_1 \mathcal{E}(\nu_1 Z_t) + \gamma_2 \mathcal{E}(\nu_2 Z_t)\right)$. In Figure 7.5, we observe that a constraint of the type $\gamma_1\nu_1 + \gamma_2\nu_2 = 2$ in the mixed variance process (7.6) allows us to fix the at-the-money implied volatility at a given level, whilst generating different slopes for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$; as in Figure 7.2, we see that the smiles generated follow a linear trend. This is again consistent with the results found in Chapter 6.

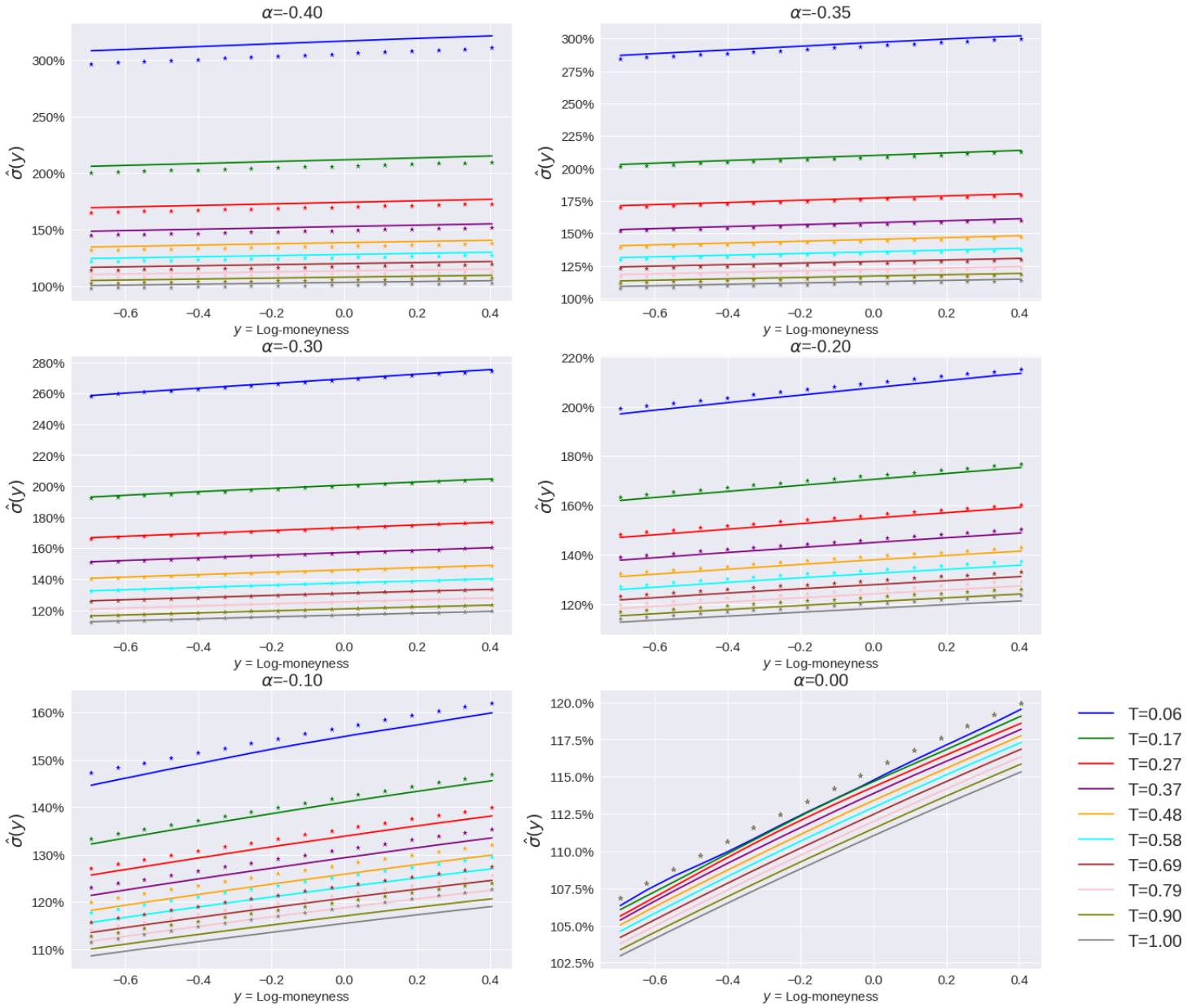


Figure 7.2: Comparison of Monte Carlo computed implied volatilities (straight lines) and LDP based implied volatilities (stars), in the rough Bergomi model, for different values of α and maturities T. We set $\eta = 2$ and $V_0 = 0.04$; for Monte Carlo we use $200,000$ simulations and $\Delta t = \frac{1}{1008}$.

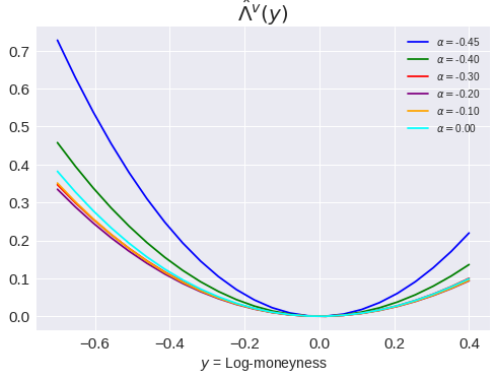


Figure 7.3: Rate function $\hat{\Lambda}^V$ for different values of α.

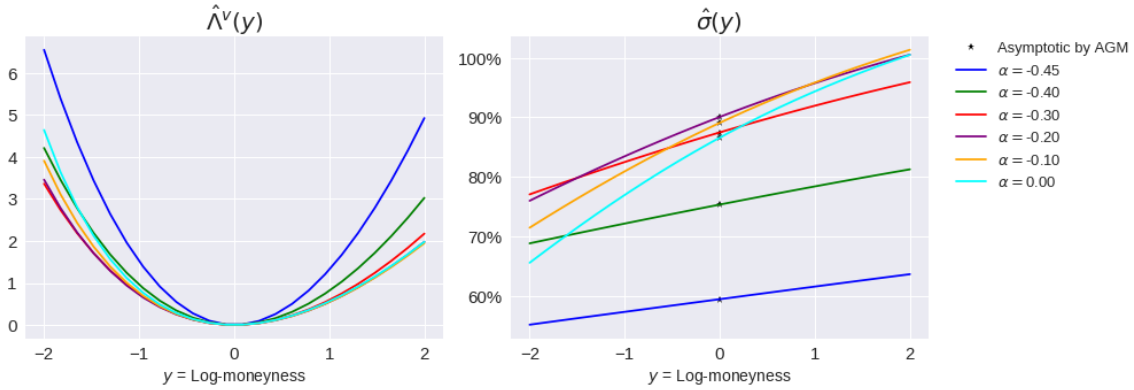


Figure 7.4: Comparison of Chapter 6 at-the-money implied volatility asymptotics and LDP based implied volatilities for different values of α, in the rough Bergomi model, with $\eta = 1.5$ and $V_0 = 0.04$.

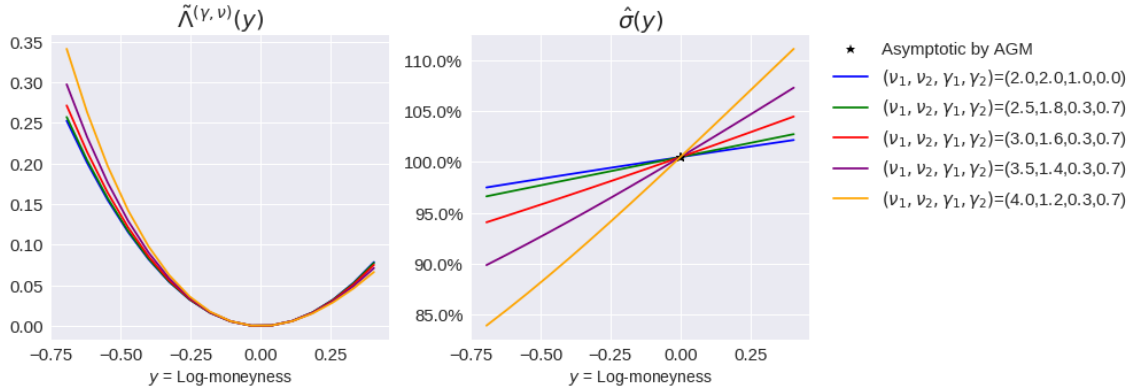


Figure 7.5: Comparison of LDP based implied volatilities for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$ in the mixed rough Bergomi process (7.6) such that $\gamma_1\nu_1 + \gamma_2\nu_2 = 2$, with $\alpha = -0.4$.

Remark 7.4.1. At this point it is important to note that the mixed rough Bergomi model 7.6 allows both the at-the-money implied volatility and its skew to be controlled through (γ, ν), whilst in the rough Bergomi model (7.5) there is not enough freedom to arbitrarily fit both quantities. Remarkably, we observe a linear pattern in Figures 7.2-7.5 for around the money strikes. In Appendix E.1 we provide an approximation scheme for the realised variance density based on the assumption of linear smiles.

7.4.3 RV smiles for mixed multi-factor rough Bergomi

We conclude our analysis by introducing the correlation effect in the implied volatility smiles, by considering the mixed multi-factor rough Bergomi model (7.7). We shall consider the following two simplified models for instantaneous variance

$$V_t = \mathcal{E}\left(\nu \int_0^t (t-s)^\alpha \mathrm{d}W_s + \eta\left(\rho \int_0^t (t-s)^\alpha \mathrm{d}W_s + \sqrt{1-\rho^2}\int_0^t (t-s)^\alpha \mathrm{d}W_s^\perp\right)\right), \tag{7.18}$$

$$V_t = \frac{1}{2}\left(\mathcal{E}\left(\nu \int_0^t (t-s)^\alpha \mathrm{d}W_s\right) + \mathcal{E}\left(\eta\left(\rho \int_0^t (t-s)^\alpha \mathrm{d}W_s + \sqrt{1-\rho^2}\int_0^t (t-s)^\alpha \mathrm{d}W_s^\perp\right)\right)\right), \tag{7.19}$$

where W and $W^\perp$ are independent standard Brownian motions and $\nu, \eta > 0$.

On one hand, Figure 7.6 shows the implied volatility smiles corresponding to (7.18). We conclude that adding up correlated factors inside the exponential does not change the behaviour of implied volatility smiles, and they still have a linear form around the money. Moreover, in this context Chapter the results in Chapter 6 still hold and we provide the asymptotic benchmark in Figure 7.6 to support our numerical scheme. On the other hand, Figure 7.7 shows the implied volatility smiles corresponding to (7.19), which are evidently non-linear around the money in the negatively correlated cases. Consequently, we can see that having a sum of exponentials, each one driven by a different (fractional) Brownian motion does indeed affect the behaviour of the convexity in the implied volatility around the money. We further superimpose a linear trend on top of the smiles in Figure 7.8 to clearly show the effect of correlation in the convexity of the smiles.

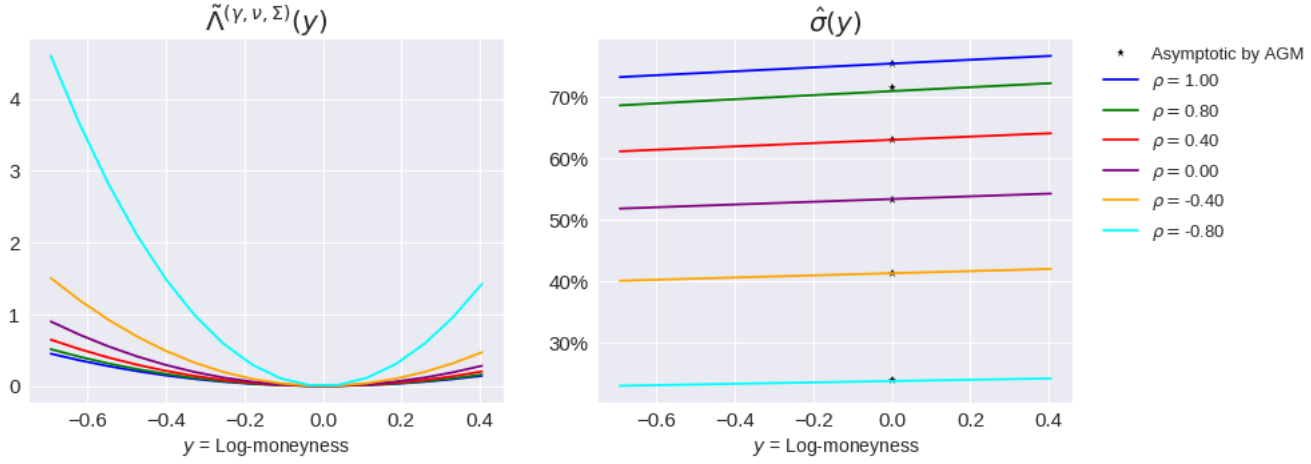


Figure 7.6: Rate function and corresponding implied volatilities for the model (7.18), with $(\alpha, \nu, \eta) = (-0.4, 0.75, 0.75)$.

Remark 7.4.2. Despite the empirical evidence given in Figure 7.1, the commitment to linear smiles from a modelling perspective is strong. The simplified mixed multi-factor rough Bergomi model (7.19), however, is sufficiently flexible to generate both linear ($\rho \geq 0$) and nonlinear ($\rho < 0$) smiles.

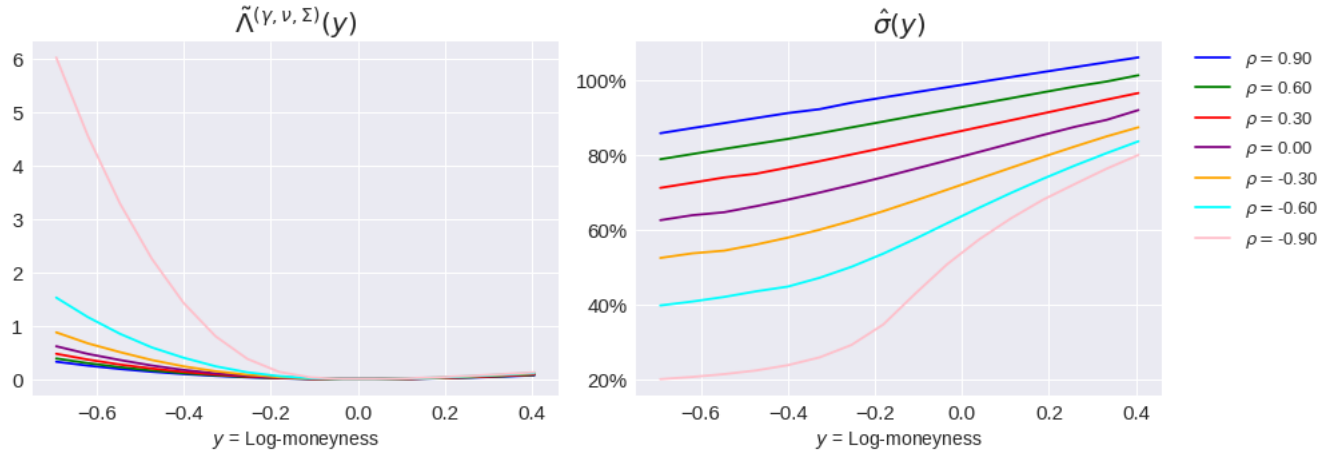


Figure 7.7: Rate function and corresponding implied volatilities for the model (7.19) with $(\alpha, \nu, \eta) = (-0.4, 1.0, 3.0)$.

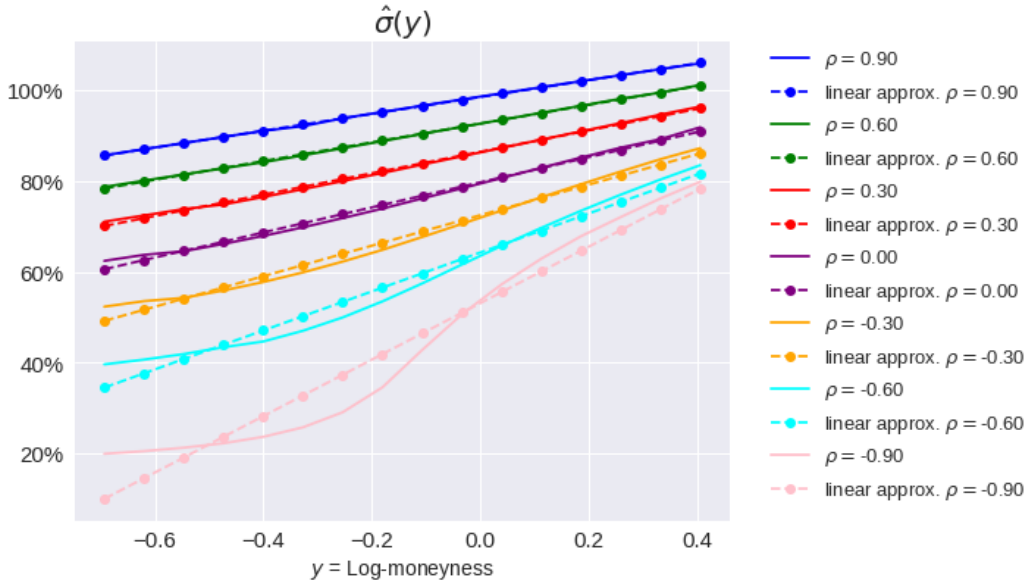


Figure 7.8: Implied volatilities and superimposed linear smiles for the model (7.19), with $(\alpha, \nu, \eta) = (-0.4, 1.0, 3.0)$.

7.5 Options on VIX

Although options on realised variance are the most natural core modelling object for stochastic volatility models, in practice the most popular variance derivative is the VIX. In this section we therefore turn our attention to the VIX and VIX options and study their asymptotic behaviour. For this section, we fix $\mathcal{T} := [0, T]$. Let us now consider the following general model $(V_t)_{t \geq 0}$ for instantaneous variance:

$$V_t = \xi_0(t) \mathcal{E}\left(\int_0^t g(t,s) \mathrm{d}W_s \right). \tag{7.20}$$

Then, the VIX process is given by

$$\mathrm{VIX}_T = \sqrt{\frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}[V_t|\mathcal{F}_T]\mathrm{d}t}.$$

We introduce the following stochastic process $(V^{g,T})_{t\in[T,T+\Delta]}$, for notational convenience, as

$$V_t^{g,T} := \int_0^T g(t,s)\mathrm{d}W_s, \tag{7.21}$$

and assume that the mapping $s \mapsto g(t,s)$ is in $L^2[0,T]$ for all $t \in [T, T+\Delta]$ such that the stochastic integral in (7.21) is well-defined.

Proposition 7.5.1. *The VIX dynamics in model* (7.20)*, where the volatility of volatility is denoted by ν, are given by*

$$VIX^2_{T,\nu} := \frac{1}{\Delta}\int_T^{T+\Delta} \xi_0(t)\exp\left(\nu V_t^{g,T} - \frac{\nu^2}{2}\mathbb{E}[(V_t^{g,T})^2]\right)\mathrm{d}t.$$

Proof. Follows directly from [JMM18, Proposition 3.1]. □

We now define the following $L^2[0,T]$ operator $\mathcal{I}^{g,T} : L^2[0,T] \to \mathcal{C}[T,T+\Delta]$, and space $\mathscr{H}^{g,T}$ as

$$\mathcal{I}^{g,T}f(\cdot) := \int_0^T g(\cdot,s)f(s)\mathrm{d}s, \quad \mathscr{H}^{g,T} := \{\mathcal{I}^{g,T}f : f \in L^2[0,T]\}, \tag{7.22}$$

where the space $\mathscr{H}^{g,T}$ is equipped with the following inner product $\langle \mathcal{I}^{g,T}f_1, \mathcal{I}^{g,T}f_2\rangle_{\mathscr{H}^{g,T}} := \langle f_1, f_2, \rangle_{L^2[0,T]}$. Note that the function g must be such that the operator $\mathcal{I}^{g,T}$ is injective so that that inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}^{g,T}}$ on $\mathscr{H}^{g,T}$ is well-defined.

Proposition 7.5.2. *Assume that there exists $h \in L^2[0,T]$ such that $\int_0^\varepsilon |h(s)|\mathrm{d}s < +\infty$ for some $\varepsilon > 0$ and $g(t,\cdot) = h(t-\cdot)$ for any $t \in [T, T+\Delta]$. Then, the space $\mathscr{H}^{g,T}$ is the reproducing kernel Hilbert space for the process $(V_t^{g,T})_{t\in[T,T+\Delta]}$.*

Proof. See Appendix E.2.5. □

Theorem 7.5.3. *For any $\gamma > 0$, the sequence of stochastic processes $(\varepsilon^{\gamma/2}V^{g,T})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}[T,T+\Delta]$ with speed $\varepsilon^{-\gamma}$ and rate function Λ^V, defined as*

$$\Lambda^V(\mathrm{x}) := \begin{cases} \frac{1}{2}\|\mathrm{x}\|^2_{\mathscr{H}^{g,T}}, & \text{if } \mathrm{x} \in \mathscr{H}^{g,T}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{7.23}$$

Proof. Direct application of the generalised Schilder's Theorem [DS89, Theorem 3.4.12]. □

Remark 7.5.4. We now introduce a Borel subset of $\mathcal{C}[T,T+\Delta]$, defined as $A := \{g \in \mathcal{C}[T,T+\Delta] : g(x) \geq 1 \text{ for all } x \in \mathbb{R}\}$. Then, by a simple application of Theorem 7.5.3 and using that the rate function Λ^V is continuous on A, we can obtain then obtain the following tail behaviour of the process $V^{g,T}$:

$$\lim_{\varepsilon\downarrow 0}\varepsilon^\gamma \log\mathbb{P}\left(V_t^{g,T} \geq \frac{1}{\varepsilon^{\gamma/2}}\right) = -\inf_{g\in A}\Lambda^V(g), \tag{7.24}$$

for any $\gamma > 0$ and $t \in [T, T+\Delta]$.

Remark 7.5.5. Let us again fix the kernel g as the rough Bergomi kernel and denote the corresponding reproducing kernel Hilbert space by $\mathscr{H}^{\eta,\alpha,T}$ and the corresponding process $V^{g,T}$ as $V^{\eta,\alpha,T}$. If $\mathrm{x} \in \mathscr{H}^{\eta,\alpha,T}$ it follows that there exists $f \in L^2[0,T]$ such that $\mathrm{x}(t) = \int_0^T \eta\sqrt{2\alpha+1}(t-s)^\alpha f(s)\mathrm{d}s$ for all $t \in [T, T+\Delta]$. Clearly, it follows that $\mathrm{x} \in \mathscr{H}^{a\eta,\alpha,T}$ for any $a > 0$, as $f \in L^2[0,T]$ implies that $\frac{1}{a}f =: f_a \in L^2[0,T]$. We can compute the norm of x in

each of these spaces to arrive at the following isometry:

$$\|\mathrm{x}\|^2_{\mathscr{H}^{a\eta,\alpha,T}} = \|f_a\|^2_{L^2[0,T]} = \frac{1}{a^2}\int_0^T f^2(s)\mathrm{d}s = \frac{1}{a^2}\|\mathrm{x}\|^2_{\mathscr{H}^{\eta,\alpha,T}}. \tag{7.25}$$

We may now amalgamate (7.23), (7.24), and (7.25) to arrive at the following statement, which tells us how the large strike behaviour scales with the vol-of-vol parameter η in the rough Bergomi model:

$$\lim_{\varepsilon\downarrow 0}\varepsilon^\gamma \log P\left(V_t^{a\eta,\alpha,T} \geq \frac{1}{\varepsilon^{\gamma/2}}\right) = \lim_{\varepsilon\downarrow 0}\varepsilon^\gamma \log\left(P\left(V_t^{\eta,\alpha,T} \geq \frac{1}{\varepsilon^{\gamma/2}}\right)^{1/a^2}\right) \tag{7.26}$$

Indeed, (7.26) tells us precisely how increasing the vol-of-vol parameter η multiplicatively by a factor a in the rough Bergomi model increases the probability that the associated process $V^{g,T}$ will exceed a certain level.

Before stating the main theorem of this section, we first define the following rescaled process:

$$V_t^{g,T,\varepsilon} := \varepsilon^{\gamma/2}V_t^{g,T}, \quad \widetilde{V}_t^{g,T,\varepsilon} := V_t^{g,T,\varepsilon} - \frac{\varepsilon^\gamma}{2}\int_0^t g^2(t,u)\mathrm{d}u + \varepsilon^{\gamma/2}, \tag{7.27}$$

for $\varepsilon \in [0,1]$, $t \in [T, T+\Delta]$. We also define the following $\mathcal{C}([T,T+\Delta]\times[0,1])$ operators $\varphi_{1,\xi_0}, \varphi_2$, which map to $\mathcal{C}([T,T+\Delta]\times[0,1])$ and $\mathcal{C}[0,1]$ respectively, as

$$(\varphi_{1,\xi_0}f)(s,\varepsilon) := \xi_0(s)\exp(f(s,\varepsilon)), \quad (\varphi_2 g)(\varepsilon) := \frac{1}{\Delta}\int_T^{T+\Delta} g(s,\varepsilon)\mathrm{d}s. \tag{7.28}$$

Note that in the definition of φ_{1,ξ_0} in (7.28) we assume ξ_0 to be a continuous, single valued, and strictly positive function on $[T,T+\Delta]$. This then implies that for every $s \in [T,T+\Delta]$, the map $\varepsilon \mapsto (\varphi_{1,\xi_0}f)(s,\varepsilon)$ is a bijection and hence has an inverse, denoted by φ^{-1}_{1,ξ_0}, which is defined as $(\varphi^{-1}_{1,\xi_0}f)(s,\varepsilon) := \log\left(\frac{f(s,\varepsilon)}{\xi_0(s)}\right)$.

Theorem 7.5.6. *For any $\gamma > 0$, the sequence of rescaled VIX processes $(e^{\varepsilon^{\gamma/2}}VIX_{T,\varepsilon^{\gamma/2}})_{\varepsilon\in[0,1]}$ satisfies a pathwise large deviations principle on $\mathcal{C}[0,1]$ with speed $\varepsilon^{-\gamma}$ and rate function*

$$\Lambda^{VIX}(\mathrm{x}) := \inf_{s\in[T,T+\Delta]}\left\{\Lambda^V\left(\log\left(\frac{\mathrm{y}(s,\cdot)}{\xi_0(s)}\right)\right) : \mathrm{x}(\cdot) = (\varphi_2\mathrm{y})(\cdot)\right\}.$$

Proof. See Appendix E.2.6. □

Remark 7.5.7. Using Theorem 7.5.6, we can deduce the small-noise, large strike behaviour of VIX options. Indeed, for the Borel subset A of $\mathcal{C}[T,T+\Delta]]$ introduced in Remark 7.5.4 we have that

$$\lim_{\varepsilon\downarrow 0}\varepsilon^\gamma \log\mathbb{P}\left(VIX_{T,\varepsilon^{\gamma/2}} \geq e^{-\varepsilon^{\gamma/2}}\right) = -\inf_{g\in A}\Lambda^{\mathrm{VIX}}(g),$$

for any $\gamma > 0$.

7.6 Summary

In this Chapter we have characterised, for the first time, the small-time behaviour of options on integrated variance in rough volatility models, using large deviations theory. Our approach has a solid theoretical basis, with very convincing corresponding numerics, which agree with observed market phenomenon and the theoretical results attained in Chapter 6. Both the theoretical and the numerical results hold for each of the three rough volatility models presented, whose complexity increases. Any of the three, with our corresponding results, would be suitable for practical use; the user would simply chose the level of complexity needed to satisfy their individual needs. Note also that the theoretical results are widely applicable, and one could very easily adapt results presented in this

work to other models where the volatility process also satisfies a large deviations principle, and whose rate function can be computed easily and accurately.

Perhaps surprisingly, we have discovered that lognormal models such as rough Bergomi [BFG16], 2 Factor Bergomi [Ber08; Ber16] and mixed versions thereof, generate linear smiles around the money for options on realised variance in log-space. This is, at the very least, a property to be taken into account when modelling volatility derivatives and, to our knowledge, has never been addressed or commented on in previous works. Whether such an assumption is realistic or not, we have in addition provided an explicit way to construct a model that generates non-linear smiles should this be required or desired.

We have also proved a pathwise large deviations principle for rescaled VIX processes, in a fairly general setting with minimal assumptions on the kernel of the stochastic integral used to define the instantaneous variance; these results are then used to establish the small-noise, large strike asymptotic behaviour of the VIX. The current set up does not allow us to deduce the small-time VIX behaviour from the pathwise large deviations principle, but this would be a very interesting area for future research. Our numerical scheme would most likely give a good approximation for the rate function and corresponding small-time VIX smiles.

8

Risk premium and rough volatility; a measure change point of view

8.1 Introduction

Recently, rough volatility models have attracted the attention of many academics as well as practitioners. In 1996, Comte and Renault [CR96] pioneered a stochastic volatility model whose increment were not independent neither Markovian. This is very much in line with the stylised fact that volatility has long-memory or long-range dependence (see Cont [Con01] for an excellent review in stylised facts). In fact, historically most of the Risk-Management computations such as Value at Risk and Expected Shortfall are based on methodologies involving GARCH [Eng82; Eng01] or related models (see [MFE15] for a complete monograph on the topic), which clearly induce non-Markovian dynamics. This in turn, implies that both the market participants and regulators are well aware of the fact that volatility has non-Markovian properties.

In spite of this empirical evidence, until Comte and Renault's [CR96] model, no continuous-time model had the non-Markovian property in the volatility process. In fact, the most famous pricing models that we covered in Chapter 1 share the common property of Markovianity, hence the current state of the processes involved suffices to charachterise future behaviour. One could, perhaps, argue that pricing and risk management are not necessarily related and that certain properties could not be preserved. In spite of this fact, the risk neutral valuation establishes a very clear bridge between the empirical or physical measure (were risk management happens) and the pricing measure through the Girsanov transform [Gir60]. This approach was recently started by Gatheral, Jaisson and Rosenbaum [GJR18] and completed by Bayer, Friz and Gatheral [BFG16], showing that the bridge between measures does not break the non-Markovianity of the volatility process.

In the present Chapter, we focus on the missing piece of the equity puzzle: the risk-premium. For an in-depth review of the topic we recommend excellent handook [R M08]. The risk-premium, in our context, is understood as the additional premium that market participants are willing to pay for securities that are not perfectly replicable (in an incomplete market setting) and is subject to participants's risk-aversion. Several attempts have been made in the past to estimate the empirical risk-premium from the market implied prices; starting from the Arrow-Debreu model [AD54; Rie88], Black-Litterman model [BL91], many variation arounds the Capital Asset Pricing Model (CAPM) such as [Cam+18], or more recently through the Ross Recovery Theorem [Ros15]. Albeit relevant, all these approaches rely on discrete-time models. In the continuous-time literature, the equity premium puzzle has

been approached by many authors, to name a few; [Aas93] takes a diffusion model with jumps, [Guo98] with the Heston model (originally for FX),[FLC14] with stochastic volatility and [XWL10] from an agent based model and equilibrium point of view. In this work, we shall focus on continuous-time stochastic volatility models in the spirit of the approach taken by Carr and Wu [CW08] and other related works, as in [BV13] and [GV16] (and references therein). We shall further ivestigate the impact of the risk-premium in $\mathbb{Q}$ volatility forecasting, which was studiend in Chernov [Che07] in a discrete-time setting.

This last chapter is organised as follows: First, we focus on the particular bridge between the Physical $\mathbb{P}$ and Pricing $\mathbb{Q}$ measures in general rough stochastic volatility models. We present sufficient conditions for such bridge to be well defined and to obtain a true martingale process under the risk-neutral measure. We further examine different approaches to model the risk premium process and finally we develop a procedure to estimate the risk-premium process under some specific assumptions, which in turn allows us to $\mathbb{Q}$ forecast volatility.

8.2 Rough volatility models and change of measure

In light of the promising empirical results presented by both Gatheral, Jaisson and Rosenbaum [GJR18] and Bennedsen, Lunde and Pakkanen [BLP16] when considering fractional processes driving volatility, we present the natural extension of the model to pricing schemes. In this section we aim at characterising the dynamics of such models under equivalent martingale measures $\mathbb{Q} \sim \mathbb{P}$. Then, due to the fundamental theorem of asset pricing [BP91; DS94] the market will be free of arbitrage opportunities by construction. As recently shown in [GJR18; BLP16] , empirical studies suggest that volatility under $\mathbb{P}$ is driven by the following dynamics:

$$\begin{cases} \dfrac{\mathrm{d}S_t}{S_t} &= \mu_t \mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^{\mathbb{P}}, \\ V_t &= \psi(t, Y_t), \\ Y_t &= \displaystyle\int_{-\infty}^{t} k(t,s)\mathrm{d}Z_s^{\mathbb{P}} \end{cases} \tag{8.1}$$

starting from $S_0 > 0$, where $Z^{\mathbb{P}}$ and $W^{\mathbb{P}}$ are respectively one and two-sided standard Brownian motions defined on a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\psi : \mathbb{R}\times\mathbb{R} \mapsto \mathbb{R}_+ := (0, \infty)$. We note that $\mathcal{F}$ should be understood as the filtration generated by the increments of the Brownian motions to avoid predictability issues in the origin. On the positive half-line, $Z^{\mathbb{P}}$ and $W^{\mathbb{P}}$ are correlated with correlation ρ, and we introduce the parameter $\overline{\rho} := \sqrt{1-\rho^2}$. We consider here a fixed time horizon $T > 0$, and denote $\mathbb{T} := [0, T]$. The choice of kernel $k(\cdot,\cdot)$ must be done such that the stochastic integral $\int_{-\infty}^{t} k(t,s)\mathrm{d}Z_s^{\mathbb{P}}$ is a well-defined Gaussian process for each $t \in \mathbb{T}$. The following examples are common choices of such kernels:

Example 8.2.1.

(i) The Mandelbrot and Van Ness' [MV68] representation of the fBM corresponds to the kernel

$$k(t,s) = C_H \left\{ (t-s)^{H-\frac{1}{2}} - (-s)^{\frac{1}{2}-H} \mathbb{1}_{\{s\leq 0\}} \right\}, \tag{8.2}$$

for $H \in (0,1)$, with $C_H := \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$.

(ii) The Gamma kernel, common in the $\mathcal{BSS}$ literature pioneered by Barndorff-Nielsen and Schmiegel [BS07], is given by

$$k(t,s) = (t-s)^{H-\frac{1}{2}} \mathrm{e}^{-\beta(t-s)}, \qquad \text{with } H \in (0,1), \beta > 0. \tag{8.3}$$

The following core result establishes sufficient conditions for a measure change to be well defined. We consider the following set of assumptions:

Assumption 8.2.2.

(i) ψ is continuous and bounded on $\mathbb{T} \times (-\infty, a]$ for each $a > 0$;

(ii) $\mathbb{E}\left[(V_t)^{-1}\right] < \infty$ for $t \in \mathbb{T}$;

(iii) μ and r are $(\mathcal{F}_t)_{t\in\mathbb{T}}$-measurable and $\mathbb{P}$-bounded;

(iv) $\inf\limits_{t\in\mathbb{T}} \left\{ \rho \int_0^t k(t,s) \dfrac{(r_s - \mu_s)}{\sqrt{\psi(s, Y_s)}} \mathrm{d}s \right\} \geq 0$, $\widehat{\mathbb{P}}_n$-almost surely for all $n \in \mathbb{N}$, where

$$\left.\frac{\mathrm{d}\widehat{\mathbb{P}}_n}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} := \mathcal{E}\left(\int_0^{\cdot} \frac{r_s - \mu_s}{\sqrt{V_s}} \mathrm{d}W_s^{\mathbb{P}} + \int_0^{\cdot} \gamma_s \mathrm{d}W_s^{\mathbb{P},\perp} \right)_{t\wedge\tau_n}, \quad \tau_n := \inf\{t \geq 0,\ Y_t = -n\}, \tag{8.4}$$

and γ is a $(\mathcal{F}_t)_{t\in\mathbb{T}}$-measurable with $M_\gamma > 0$ such that $|\gamma_t| < M_\gamma$ $\mathbb{P}$-almost surely and $\mathbb{P}_n$-almost surely, for all $n \in \mathbb{N}$ and $t \in \mathbb{T}$.

(v) $\rho \leq 0$;

In order to state the main result, define the Radon-Nikodym derivative, for each $t \in \mathbb{T}$,

$$\mathcal{D}_t^\gamma := \left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = \mathcal{E}\left(\int_0^{\cdot} \frac{r_s - \mu_s}{\sqrt{V_s}} \mathrm{d}W_s^{\mathbb{P}} + \int_0^{\cdot} \gamma_s \mathrm{d}W_s^{\mathbb{P},\perp} \right)_t, \tag{8.5}$$

where $W^{\mathbb{P},\perp}$ is a standard Brownian motion independent of $W^{\mathbb{P}}$, such that $Z_t^{\mathbb{P}} = \rho W_t^{\mathbb{P}} + \overline{\rho} W_t^{\mathbb{P}\perp}$ for $t \in \mathbb{T}$, for any $\gamma \in \Gamma$, the set of all processes satisfying Assumption 8.2.2 (iv) Here, $\mathcal{E}$ is the Dóleans-Dade stochastic exponential. This set in fact represents, in view of (8.5) the space of all possible market prices of risk. We first note the following result:

Proposition 8.2.3. *The process* $X_t := \int_0^t \dfrac{r_s - \mu_s}{\sqrt{V_s}} \mathrm{d}W_s^{\mathbb{P}} + \int_0^t \gamma_s \mathrm{d}W_s^{\mathbb{P},\perp}$ *is a local martingale on* $\mathbb{T}$.

Proof. Boundedness of γ gives that $\int_0^t \gamma_s \mathrm{d}W_s^{\mathbb{P},\perp}$ is a local martingale. For $\int_0^t \frac{r_s-\mu_s}{\sqrt{V_s}} \mathrm{d}W_s^{\mathbb{P}}$, we note that the integrand is $\mathcal{F}_s$ measurable and locally bounded with respect to the sequence of stopping times τ_n defined in (8.4), we further have $\mathbb{E}\left[\frac{(r_s-\mu_s)^2}{V_s}\right] < \infty$ for all $t \in \mathbb{T}$. This in turn, gives that X is a local martingale. □

Proposition 8.2.3 justifies the use of a Dóleans-Dade stochastic exponential in the definition of $\mathcal{D}^\gamma$. Furthermore, we obtain the following result:

Theorem 8.2.4. *Under Assumption 8.2.2, for any* $\gamma \in \Gamma$*, the following hold:*

(I) The Radon-Nikodym derivative process $\mathcal{D}^\gamma$ *in* (8.5) *is a true martingale.*

(II) Under the (arbitrage-free) equivalent risk-neutral martingale measure $\mathbb{Q}$*, we have*

$$\begin{cases} \dfrac{\mathrm{d}S_t}{S_t} &= r_t \mathrm{d}t + \sqrt{V_t}\, \mathrm{d}W_t^{\mathbb{Q}}, \qquad S_0 > 0, \\ V_t &= \psi\left(t, \displaystyle\int_{-\infty}^t k(t,s) \mathrm{d}Z_s^{\mathbb{Q}} + \int_0^t k(t,s)\lambda_s \mathrm{d}s \right), \end{cases} \tag{8.6}$$

where λ *is the market price of volatility risk defined by*

$$\lambda_s := \rho \frac{\mu_s - r_s}{\sqrt{V_s}} - \overline{\rho}\, \gamma_s,$$

and where $W^{\mathbb{Q}}$ *is a standard Brownian motion and* $Z^{\mathbb{Q}}$ *two-sided standard Brownian motion under* $\mathbb{Q}$ *defined as*

$$W_t^{\mathbb{Q}} := W_t^{\mathbb{P}} + \int_0^t \frac{\mu_s - r_s}{\sqrt{V_s}} \mathrm{d}s, \qquad \textit{and} \qquad Z_t^{\mathbb{Q}} := \begin{cases} Z_t^{\mathbb{P}} + \displaystyle\int_0^t \lambda_s \mathrm{d}s & \textit{if } t \geq 0 \\ Z_t^{\mathbb{P}} & \textit{if } t < 0 \end{cases} \tag{8.7}$$

(III) The process S is a true martingale under the risk-neutral measure $\mathbb{Q}$ on $\mathbb{T}$.

Proof. We may first rewrite $Z_t^{\mathbb{P}} = \rho W_t^{\mathbb{P}} + \overline{\rho} W_t^{\mathbb{P}\perp}$ for $t \in \mathbb{T}$, where $W^{\mathbb{P}\perp}$ is a standard Brownian motion independent of $W^{\mathbb{P}}$. In order to satisfy the no-arbitrage conditions, the change of measure for $W^{\mathbb{P}}$ is constrained by the martingale restriction on the discounted spot dynamics. On the contrary, the second Brownian motion $Z^{\mathbb{P}}$ gives freedom to the model and makes the market incomplete by the free choice of the process γ. Consequently, the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ and the corresponding Radon-Nikodym derivative directly follow from Girsanov's Theorem [Gir60] via (8.5), provided that $\mathcal{D}_t^\gamma \in L^1$ and $\mathcal{D}^\gamma$ is a true martingale. It is easy to check that $\mathcal{D}_t^\gamma \in L^1$ by Proposition 8.2.3 and, being a non-negative local martingale, it is a supermartingale, and a true martingale on $\mathbb{T}$ if and only if $\mathbb{E}[\mathcal{D}_T^\gamma] = 1$. We closely follow [Gas19] with some modifications. Define the stopping time $\tau_n := \inf\{t \geq 0,\ Y_t = -n\}$. For any $s \in \mathbb{T}$, the function $f(x) := \frac{r_s - \mu_s}{\sqrt{\psi(s,x)}}$ is bounded on $[-c, \infty)$ for any $c > 0$ since r, μ and γ are bounded, and $\psi(s,\cdot)$ bounded away from zero. Then, we have

$$1 = \mathbb{E}\left[\mathcal{D}_{T\wedge\tau_n}^\gamma\right] = \mathbb{E}\left[\mathcal{D}_T^\gamma \mathbb{1}_{\{T<\tau_n\}}\right] + \mathbb{E}\left[\mathcal{D}_{\tau_n}^\gamma \mathbb{1}_{\{\tau_n \leq T\}}\right]. \tag{8.8}$$

The first term in (8.8) converges to $\mathbb{E}[\mathcal{D}_T^\gamma]$ as n tends to infinity, yielding

$$1 - \mathbb{E}[\mathcal{D}_T^\gamma] = \lim_{n\uparrow\infty} \mathbb{E}\left[\mathcal{D}_{\tau_n}^\gamma \mathbb{1}_{\{\tau_n \leq T\}}\right].$$

Next, Girsanov's theorem implies $\mathbb{E}\left[\mathcal{D}_{\tau_n}^\gamma \mathbb{1}_{\{\tau_n \leq T\}}\right] = \widehat{\mathbb{P}}_n(\tau_n \leq T)$, where $\widehat{\mathbb{P}}_n$ is defined such that

$$\widehat{W}_t^n = W_t^{\mathbb{P}} - \int_0^{t\wedge\tau_n} \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} \mathrm{d}s$$

is a $\widehat{\mathbb{P}}_n$-Brownian motion. Then, under $\widehat{\mathbb{P}}_n$, the process Y becomes

$$Y_t = \widehat{Y}_t^n + \int_0^{t\wedge\tau_n} \rho k(t,s) \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} + \overline{\rho}\, \gamma_s \mathrm{d}s,$$

where $\widehat{Y}_t^n := \displaystyle\int_{-\infty}^t k(t,s) \mathrm{d}\widehat{Z}_s^n$ and

$$\widehat{Z}_t^n = \begin{cases} Z_t^{\mathbb{P}} - \displaystyle\int_0^{t\wedge\tau_n} \rho \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} - \overline{\rho}\, \gamma_s \mathrm{d}s & \text{if } t \geq 0 \\ Z_t^{\mathbb{P}} & \text{if } t < 0 \end{cases}$$

where $\widehat{Z}^n$ is a $\widehat{\mathbb{P}}_n$-Brownian motion. Furthermore,

$$\begin{aligned} \widehat{\mathbb{P}}_n\left(\inf_{t\geq 0} Y_t \leq -n\right) &= \widehat{\mathbb{P}}_n\left(\inf_{t\geq 0}\left\{\widehat{Y}_t^n + \int_0^t \rho k(t,s) \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} - \overline{\rho}\, \gamma_s \mathrm{d}s\right\} \leq -n\right) \\ &\leq \widehat{\mathbb{P}}_n\left(\inf_{t\geq 0} \widehat{Y}_t^n + \inf_{t\geq 0}\left\{\rho \int_0^t k(t,s) \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} \mathrm{d}s\right\} + \inf_{t\geq 0}\left\{-\overline{\rho} \int_0^t \gamma_s \mathrm{d}s\right\} \leq -n\right) \\ &\leq \widehat{\mathbb{P}}_n\left(\inf_{t\geq 0} \widehat{Y}_t^n + \inf_{t\geq 0}\left\{\rho \int_0^t k(t,s) \frac{r_s - \mu_s}{\sqrt{\psi(s, Y_s)}} \mathrm{d}s\right\} \leq -n + \overline{\rho} T M_\gamma\right), \end{aligned}$$

so that, by Assumption 8.2.2 (iv) and boundedness of $|\gamma_t| < M_\gamma$, we obtain

$$\widehat{\mathbb{P}}_n\left(\inf_{t\geq 0} Y_t \leq -n\right) \leq \widehat{\mathbb{P}}_n\left(\inf_{t\geq 0} \widehat{Y}_t^n \leq -n + \overline{\rho} T M_\gamma\right). \tag{8.9}$$

Inequality (8.9) in turn implies $\widehat{\mathbb{P}}_n(\tau_n \leq T) \leq \widehat{\mathbb{P}}_n(\widehat{\tau}_n \leq T)$, for $\widehat{\tau}_n := \inf\{t \geq 0,\ \widehat{Y}_t^n = -n + \overline{\rho} T M_\gamma\}$. Finally, we obtain

$$\lim_{n\uparrow\infty} \widehat{\mathbb{P}}_n(\tau_n \leq T) \leq \lim_{n\uparrow\infty} \widehat{\mathbb{P}}_n(\widehat{\tau}_n \leq T) = \lim_{n\uparrow\infty} \mathbb{P}\left(\inf_{t\in[0,T]} Y_t \leq -n + \overline{\rho} T M_\gamma\right) = 0.$$

and it follows that $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ is indeed a true martingale and note that $\lim_{n\uparrow\infty} \widehat{\mathbb{P}}_n = \mathbb{Q}$. More precisely, the relation (8.7) holds between the $\mathbb{P}$ and $\mathbb{Q}$ Brownian motions. The result then readily follows by direct application of the change of measure theorem.

We now prove that S is a true martingale if $\rho \leq 0$. Consider the stopping time $\iota_n := \inf\{t \geq 0,\ \int_{-\infty}^t k(t,s)\mathrm{d}Z_s^{\mathbb{Q}} = n\}$. For any $t \in \mathbb{I}$, the function $g(x) := \psi\left(t, x + \int_0^t k(t,s)\lambda_s \mathrm{d}s\right)$ is bounded $\mathbb{Q}$-almost surely on $(-\infty, a]$ by Assumption 8.2.2 (iii) and boundedness of γ, so that

$$S_0 = \mathbb{E}^{\mathbb{Q}}[S_{T\wedge\iota_n}] = \mathbb{E}^{\mathbb{Q}}[S_T \mathbb{1}_{\{T<\iota_n\}}] + \mathbb{E}^{\mathbb{Q}}[S_T \mathbb{1}_{\{T>\iota_n\}}].$$

The first term converges to $\mathbb{E}^{\mathbb{Q}}[S_T]$ as n tends to infinity, hence

$$S_0 - \mathbb{E}^{\mathbb{Q}}[S_T] = \lim_{n\uparrow\infty} \mathbb{E}^{\mathbb{Q}}[S_T \mathbb{1}_{\{T>\iota_n\}}].$$

Girsanov's Theorem further gives $\mathbb{E}^{\mathbb{Q}}[S_T \mathbb{1}_{\{T>\iota_n\}}] = S_0 \widetilde{\mathbb{P}}_n(T > \iota_n)$, where $\widetilde{\mathbb{P}}_n$ is such that

$$\widetilde{W}_t^n = W_t^{\mathbb{Q}} - \int_0^{t\wedge\iota_n} V_s \mathrm{d}s$$

is a $\widetilde{\mathbb{P}}_n$-Brownian motion. Note that for $t < \iota_n$ $\widehat{Y}_t = \widetilde{Y}_t + \rho \int_{-\infty}^t k(t,s) V_s \mathrm{d}s$, where $\widetilde{Y}_t = \int_0^t k(t,s)\mathrm{d}\widetilde{Z}_s^n$ and $\widetilde{Z}_t^n := Z_t^{\mathbb{Q}} - \rho \int_0^t k(t,s) V_s \mathrm{d}s$ is also a $\widetilde{\mathbb{P}}_n$-Brownian motion. We conclude that, if $\rho \leq 0$ then $\widehat{Y}_t \geq \widetilde{Y}_t$ and

$$\lim_{n\uparrow\infty} \widetilde{\mathbb{P}}_n(\iota_n \leq T) \leq \lim_{n\uparrow\infty} \widetilde{\mathbb{P}}_n(\widetilde{\iota}_n \leq T) = \lim_{n\uparrow\infty} \mathbb{P}(\sup_{t\in[0,T]} Y_t \leq n) = 0,$$

where $\widetilde{\iota}_n := \inf\{t \geq 0,\ \widetilde{Y}_t = n\}$, and hence S is a true martingale. □

Discussion

Consider the constant case $\gamma_s = \boldsymbol{\gamma}$ and $\boldsymbol{\mu} = \mu_s > r_s = \boldsymbol{r}$ for $\boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{r} \in \mathbb{R}$, with $\rho \leq 0$, which directly satisfies all conditions in Theorem 8.2.4. In this scenario, a sufficient condition for the change of measure to be well defined is that the physical drift must be larger than the risk-free rate. In financial terms, this means that a (risky) stock must provide a larger return than a (risk-free) bond. In the classical Black-Scholes approach, this condition is not necessary at all as any payoff can be exactly replicated. However, we see that as soon as the market becomes incomplete, some structure needs to be imposed to the physical and risk-free drifts. The precise characterisation of this structure to obtain sufficient and necessary conditions is left as further research.

8.2.1 Characterisation of rough volatility models via Generalised Fractional Operators

It is natural to represent rough volatility models in terms of fractional operators. In this section we recall the Generalised Fractional Operators (GFO) and a representation result for rough volatility models in terms of GFO. We quickly recall the GFO introduced in Definition 2.2.1:

Definition 8.2.5. For any $\beta \in (0,1)$, $\alpha \in (-\beta, 1-\beta)$ and $h \in \mathcal{C}_b^1(\mathbb{R}_+)$ such that $h'(\cdot) \leq 0$, the GFO associated to the kernel $g(x) := x^\alpha h(x)$ applied to $f \in C^\beta(\mathbb{R})$ is defined as

$$(\mathcal{G}^\alpha f)(t) := \begin{cases} \displaystyle\int_0^t (f(s) - f(0)) \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s, & \text{if } \alpha \in (0, 1-\beta), \\ \displaystyle\frac{\mathrm{d}}{\mathrm{d}t} \int_0^t (f(s) - f(0)) g(t-s)\mathrm{d}s, & \text{if } \alpha \in (-\beta, 0). \end{cases} \tag{8.10}$$

Remark 8.2.6. We may also use the shorthand notation $(\mathcal{G}^\alpha f)(u,t) := \mathbb{E}[(\mathcal{G}^\alpha f)(t)|\mathcal{F}_u]$ to ease forthcoming

computations.

Corollary 8.2.7 (GFO representation of rough volatility)**.** *Assume that*

- $k(t,s)\mathbf{1}_{\{s\geq 0\}} = (t-s)^\alpha f(t-s)$ *where* $\alpha \in (-\frac{1}{2}, 0)$ *,*$f \in \mathcal{C}_b^1$ *and* $f'(\cdot) \leq 0$*;*
- $\lambda \in \mathcal{C}^\beta$ *for some* $\beta \in (0,1]$*;*

Then, the system (8.6) *can be rewritten as*

$$\begin{cases} \dfrac{\mathrm{d}S_t}{S_t} &= r_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}W_t^{\mathbb{Q}}, \\ V_t &= \psi\left(t, \displaystyle\int_{-\infty}^0 k(t,s)\mathrm{d}Z_s^{\mathbb{Q}} + (\mathrm{k}^{H-\frac{1}{2}} Z^{\mathbb{Q}})(t) + (\mathcal{K}^{H+\frac{1}{2}}\lambda)(t)\right), \end{cases}$$

where k^α *and* $\mathcal{K}^\alpha$ *are the GFO associated with the kernels* $k(t-s) := (t-s)^\alpha f(t-s)$ *and* $K(t-s) := \int_0^t k(u-s)\mathrm{d}u$ *respectively. Moreover,* $\mathcal{K}^{H+\frac{1}{2}}\lambda \in \mathcal{C}^{\beta+H+\frac{1}{2}}$*.*

Proof. The fact that

$$\int_0^t k(t-s)\mathrm{d}Z_s^{\mathbb{Q}} = \left(\mathrm{k}^{H-\frac{1}{2}} Z^{\mathbb{Q}}\right)(t)$$

is straightforward by the properties of GFO in Corollary 2.3.7. Furthermore,

$$\int_0^t k(t-s)\lambda_s \mathrm{d}s = \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} K(t,s)\lambda_s \mathrm{d}s = \left(\mathcal{K}^{H+\frac{1}{2}}\lambda\right)(t).$$

The fact that $\mathcal{K}^{H+\frac{1}{2}}\lambda \in \mathcal{C}^{\beta+H+\frac{1}{2}}$ follows from the mapping properties of GFO (see Proposition 2.2.3).

□

Remark 8.2.8. Note that a sufficient condition for $\lambda \in \mathcal{C}^\beta$ is that the processes r, μ and γ are continuous.

Discussion

Taking into account that empirical studies suggest $H \approx 0.1$, i.e. $v \in \mathcal{C}^H(\mathbb{R}_+)$, then Corollary 8.2.7 suggests that λ itself should belong to the class of rough processes i.e. its Hölder regularity should be smaller than $\frac{1}{2}$. Despite this, we still have that for any $H > 0$, $\mathcal{K}^{H+\frac{1}{2}}\lambda \in \mathcal{C}^{2H+\frac{1}{2}}$, meaning that the risk premium has sample paths with Hölder regularity greater than $\frac{1}{2}$, regardless of the value of H.

8.3 Modelling the risk premium process: A practical approach

In practice, the process λ is directly modelled, without resorting to a change of measure starting from γ. In this section we will consider different modelling choices for the risk premium λ and analyse some properties. In spite of the formal derivation of Theorem 8.2.4, numerical and even philosophical treatment of the integral $\int_{-\infty}^t \bullet \, \mathrm{d}s$ is rather intricate. To overcome this issue, Bayer, Friz and Gatheral [BFG16] elegantly came up with the forward variance form of rough volatility following the spirit of Bergomi [Ber16] in the case where $\psi(t,x) = \mathrm{e}^{\nu x}$. We shall restrict ourselves to this functional form for the reminder of the section.

Proposition 8.3.1 (Exponential rough volatility in forward variance form)**.** *Consider* (8.6) *with* $\psi(t,x) = \mathrm{e}^{\nu x}$*, where* $\xi_0(t) = \mathbb{E}[V_t|\mathcal{F}_0]$ *is a strictly positive continuous function, and* $\nu > 0$*. Then, the risk-neutral dynamics in*

forward variance form are given by

$$\begin{cases} \dfrac{\mathrm{d}S_t}{S_t} &= r_t \mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^{\mathbb{Q}}, \\ V_t &= \xi_0(t)\exp\left(\nu\left((\mathrm{k}^{H-\frac{1}{2}} Z^{\mathbb{Q}})(t) + (\mathcal{K}^{H+\frac{1}{2}}\lambda)(t)\right)\right). \end{cases} \tag{8.11}$$

The function ξ_0 here has a precise financial meaning; it is in fact characterised as $\xi_0(t) = \mathbb{E}[V_t|\mathcal{F}_0]$, and thus represents the forward variance observed at time 0 for a future time t. In the remaining of this section, the process $X^{\mathbb{Q}}$ will denote a $\mathbb{Q}$-Brownian motion possibly correlated with $W^{\mathbb{Q}}$ and $Z^{\mathbb{Q}}$.

8.3.1 Risk premiums driven by Itô diffusion

GFO's give a natural framework to model risk premium processes driven by diffusions. The following statement shows the details of such construction.

Proposition 8.3.2. *Consider a risk premium process of the form $\lambda = \mathcal{K}^{\alpha} Y^{\mathbb{Q}} \in \mathcal{C}^{\alpha+\frac{1}{2}}$, with $\alpha \in \left(-\frac{1}{2}, 0\right)$ and*

$$\mathrm{d}Y_t^{\mathbb{Q}} = b(t, Y_t^{\mathbb{Q}})\mathrm{d}t + \sigma(t, Y_t^{\mathbb{Q}})\mathrm{d}X^{\mathbb{Q}},$$

where $b(\cdot)$, $\sigma(\cdot)$ satisfy linear-growth conditions and a weak solution exists. Then $\mathcal{K}^{H+\alpha+\frac{1}{2}} Y^{\mathbb{Q}} \in \mathcal{C}^{H+\alpha+1}$ and

$$V_t = \xi_0(t)\exp\left\{\nu\left((\mathrm{k}^{H-1/2} Z^{\mathbb{Q}})(t) + (\mathcal{K}^{H+\alpha+\frac{1}{2}} Y^{\mathbb{Q}})(t)\right)\right\}. \tag{8.12}$$

Proof. The proof directly follows from mapping [HJM17, Proposition 1.2] properties of GFO. □

Remark 8.3.3. The GFO representation of the instantaneous variance directly links the processes with their numerical implementation described in Chapter 2, which in the Gaussian case also extends to the Hybrid scheme by Bennedsen, Lunde and Pakkanen [BLP16]. In addition, using integration by parts, we obtain

$$\left(\mathcal{K}^{H+\alpha+\frac{1}{2}} Y^{\mathbb{Q}}\right)(t) = \int_0^t K(t-s)\mathrm{d}Y_s^{\mathbb{Q}}.$$

Hence, one can easily simulate the process using rDonsker scheme 2.11. Also, notice that $K(\cdot)$ is a nonsingular kernel making simulation straightforward.

8.3.2 Gaussian risk premiums

The most natural choice is to consider λ to be a Gaussian process. At least in Equity markets, this might not be a good choice since this implies (see Chapters 6 and 7) that the VIX smile and the Realized Variance option smile are almost flat. However, for other markets or options without volatility of volatility risk, a Gaussian λ is a perfectly acceptable choice. The following proposition follows from Corollary 8.2.7 and properties of GFO:

Proposition 8.3.4. *For any $\alpha \in (-\frac{1}{2}, 0)$, consider a risk premium process of the form $\lambda = \mathcal{K}^{\alpha} X^{\mathbb{Q}} \in \mathcal{C}^{\alpha+\frac{1}{2}}$. Then $\mathcal{K}^{H+\alpha+\frac{1}{2}} X^{\mathbb{Q}} \in \mathcal{C}^{H+\alpha+1}$ and*

$$V_t = \xi_0(t)\exp\left\{\nu\left[\left(\mathrm{k}^{H-1/2} Z^{\mathbb{Q}}\right)(t) + \left(\mathcal{K}^{H+\alpha+\frac{1}{2}} X^{\mathbb{Q}}\right)(t)\right]\right\}, \tag{8.13}$$

furthermore, we obtain

$$\begin{aligned} \mathbb{E}\left[V_t|\mathcal{F}_s\right] = \xi_0(t)\exp&\left\{\nu\left[\left(\mathrm{k}^{H-1/2} Z^{\mathbb{Q}}\right)(s,t) + \left(\mathcal{K}^{H+\alpha+\frac{1}{2}} X^{\mathbb{Q}}\right)(s,t)\right]\right\} \\ &\times \exp\left\{\frac{1}{2}\nu^2\left(\int_s^t k(t,s)^2\mathrm{d}s + \int_s^t K(t,s)^2\mathrm{d}s + \rho\int_s^t k(t,s)K(t,s)\mathrm{d}s\right)\right\}. \end{aligned}$$

Remark 8.3.5. Regarding applications of this model for the VIX, since the instantaneous variance is still log-Normal, the results in Chapter 5 and numerical methods therein still apply with minimal changes.

8.3.3 A risk premium driven by a CIR process

A second natural choice is to consider the Cox-Ingersoll-Ross [CIR85] (CIR) process

$$\mathrm{d}Y_s^{\mathbb{Q}} = \kappa(\theta - Y_s^{\mathbb{Q}})\mathrm{d}s + \sigma\sqrt{Y_s^{\mathbb{Q}}}\,\mathrm{d}X_s^{\mathbb{Q}},$$

with $\kappa, \theta, \sigma > 0$, such that $2\kappa\theta > \sigma^2$. As tempting as this approach might seem, it is not trivial at all to compute the basic quantity $\mathbb{E}[V_t]$ here, as the following proposition, proved in Appendix F.1, shows.

Proposition 8.3.6. *Assume that $Z^{\mathbb{Q}}$ and $X^{\mathbb{Q}}$ are independent and consider a risk premium process of the form $\lambda = \mathcal{K}^\alpha Y^{\mathbb{Q}} \in \mathcal{C}^{\alpha+\frac{1}{2}}$. Then, for any $u \leq t$,*

$$\mathbb{E}[V_t|\mathcal{F}_u] = \xi_0(t)\exp\left\{\nu\left[\left(\mathrm{k}^{H-\frac{1}{2}}Z^{\mathbb{Q}}\right)(u,t) + \left(\mathcal{K}^{H+\alpha+\frac{1}{2}}Y^{\mathbb{Q}}\right)(u,t)\right]\right\}\exp\left(\frac{\nu^2}{2}\int_u^t k(t-s)^2\mathrm{d}s\right)\mathrm{e}^{-Y_u^{\mathbb{Q}}C(u,T)-A(u,T)},$$

where C satisfies the Riccati equation

$$\nu k(T,t) - \partial_t C(t,T) + C(t,T)\theta + \frac{\sigma^2}{2}C^2(t,T) = 0,$$

with boundary condition $C(T,T) = 0$, and $A(t,T) = -\kappa\theta\int_t^T C(u,T)\mathrm{d}u$ with boundary condition $A(T,T) = 0$.

Already in the uncorrelated case the computation of $\mathbb{E}[V_t]$ becomes very costly, having to solve a PDE for each time t. In the correlated case there is no hope to obtain any semi-analytic result since one would need to compute cross terms and there is no Itô tool available in that case.

8.4 Roughly extracting the Risk Premium from the Market

In this section we consider the risk premium process λ to be deterministic, with the aim of obtaining a formula that links $\mathbb{P}$ and $\mathbb{Q}$ market observables. The following theorem shows how to infer the risk premium of the market using forecasts under the physical measure and Variance Swap prices in the pricing measure.

Theorem 8.4.1. *Consider the rough volatility model* (8.1) *under $\mathbb{P}$. If $\psi(t,x) = \mathrm{e}^{\nu x}$, $\mu_s = r_s$ for all $s \geq 0$ and $(\lambda_s)_{s\in\mathbb{R}_+} \in L^2(\mathbb{R})$ is deterministic, then*

$$\nu\overline{\rho}\int_s^t k(t,u)\lambda_u\mathrm{d}u = \log\left(\frac{\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s]}{\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_s]}\right) = \log\left(\frac{\xi_s(t)}{\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_s]}\right). \tag{8.14}$$

Proof. In a market where λ is deterministic, filtrations under $\mathbb{P}$ and $\mathbb{Q}$ agree, and we refer to the common natural filtration as $\mathcal{F}$. If $\mu = r$ almost surely, the Radon-Nikodym derivative (8.5) in Theorem 8.2.4 reads

$$\mathcal{D}^\gamma = \mathcal{E}\left(\int_0^\cdot \gamma_s\mathrm{d}W_s^{\mathbb{P}\perp}\right),$$

where $\gamma_s = -\overline{\rho}\lambda_s$, and the inverse Radon-Nikodym derivative is given by

$$\mathfrak{Q}^\gamma := \frac{1}{\mathcal{D}^\gamma} = \mathcal{E}\left(-\int_0^\cdot \gamma_s\mathrm{d}W_s^{\mathbb{Q}\perp}\right).$$

Then, the conditional change of measure formula yields

$$\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{Q}}\left[V_t\mathfrak{Q}_t^\gamma|\mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{Q}}\left[\mathfrak{Q}_t^\gamma|\mathcal{F}_s\right]}. \tag{8.15}$$

On the one hand, $\mathbb{E}^{\mathbb{Q}}\left[\mathfrak{C}_t^{\gamma}|\mathcal{F}_s\right] = \mathcal{E}\left(-\int_0^{\cdot}\gamma_u \mathrm{d}W_u^{\mathbb{Q}\perp}\right)_s$, by the properties of the stochastic exponential and Gaussian moment generating functions. On the other hand, since $Z_t^{\mathbb{Q}} = Z_t^{\mathbb{P}} - \int_0^t \lambda_s \mathrm{d}s$ for $t \in \mathbb{T}$, then

$$\mathbb{E}^{\mathbb{Q}}\left[V_t\mathfrak{C}_t^{\gamma}|\mathcal{F}_s\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\nu\left(\int_{-\infty}^t k(t,s)\mathrm{d}Z_s^{\mathbb{Q}} + \overline{\rho}\int_0^t \lambda_s k(t,s)\mathrm{d}s\right)\right)\exp\left(-\int_0^t \gamma_s \mathrm{d}W_s^{\mathbb{Q}\perp} - \frac{1}{2}\int_0^t \gamma_s^2 \mathrm{d}s\right)\middle|\mathcal{F}_s\right], \tag{8.16}$$

is just the conditional MGF of a Gaussian random variable. Applying Itô's isometry we obtain

$$\nu\int_{-\infty}^t k(t,s)\mathrm{d}Z_s^{\mathbb{Q}} - \int_0^t \gamma_s \mathrm{d}W_s^{\mathbb{Q}\perp}\middle|\mathcal{F}_s \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

with

$$\mu := \nu\int_{-\infty}^s k(t,u)\mathrm{d}Z_u^{\mathbb{Q}} - \int_0^s \gamma_u \mathrm{d}W_u^{\mathbb{Q}\perp} \qquad \text{and} \qquad \sigma^2 := \nu^2\int_s^t k^2(t,u)\mathrm{d}u + \int_s^t \gamma_u^2 \mathrm{d}u - 2\nu\overline{\rho}\int_s^t k(t,u)\gamma_u \mathrm{d}u.$$

where we have used $Z_t^{\mathbb{Q}} = \rho W_t^{\mathbb{Q}} + \overline{\rho}W_t^{\mathbb{Q}\perp}$ for $t \in \mathbb{T}$. Therefore,

$$\mathbb{E}^{\mathbb{Q}}\left[V_t\mathfrak{C}_t^{\gamma}|\mathcal{F}_s\right] = \exp\left\{\nu\overline{\rho}\int_0^t k(t,s)\lambda_s \mathrm{d}s + \mu + \frac{1}{2}\left(\sigma^2 - \int_0^t \gamma_s^2 ds\right)\right\},$$

or

$$\mathbb{E}^{\mathbb{Q}}\left[V_t\mathfrak{C}_t^{\gamma}|\mathcal{F}_s\right] = \mathbb{E}^{\mathbb{Q}}\left[\mathfrak{C}_t^{\gamma}|\mathcal{F}_s\right]\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s]\exp\left(-\nu\overline{\rho}\int_s^t k(t,u)\lambda_u \mathrm{d}u\right)$$

by using the decomposition of σ^2 as the sum of three terms. Finally, going back to (8.15) we obtain

$$\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s]\exp\left(-\nu\overline{\rho}\int_s^t k(t,u)\lambda_u \mathrm{d}u\right),$$

and the result follows. □

8.4.1 Numerical experiment with the rough Bergomi model I: estimating the risk premium

In this section we will work with the RFSV($\mathbb{P}$)-rough Bergomi model($\mathbb{Q}$):

$$\begin{array}{rl}\frac{\mathrm{d}S_t}{S_t} & = \mu_t \mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^{\mathbb{P}}, \\ V_t & = V_0\exp(\nu Z_t^H),\end{array} \quad , \quad \text{and} \quad \begin{array}{rl}\frac{\mathrm{d}S_t}{S_t} & = r_t \mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_t^{\mathbb{Q}}, \\ V_t & = \xi_0(t)\exp\left(\nu\left(\mathrm{k}^{H-\frac{1}{2}}Z^{\mathbb{Q}})(t) + (\mathcal{K}^{H+\frac{1}{2}}\lambda)(t)\right)\right),\end{array} \tag{8.17}$$

where Z^H is a the Mandelbrot and Van Ness' [MV68] defined in Example 8.2.1 (i). Under the assumption that λ is deterministic, Theorem 8.4.1 gives an explicit procedure to compute the risk-premium process. In practice however, we are only able to observe variance swap quotes in discrete times, hence it is natural to consider λ a piecewise constant function as:

Assumption 8.4.2. The deterministic process λ admits the following piecewise constant representation on the time partition $\{0 = T_0 < T_1, ..., < T_n = T\}$ for $n \in \mathbb{N}$:

$$\lambda(t) := \sum_{i=1}^n \lambda_i \mathbb{1}_{\{t\in[T_{i-1},T_i)\}}, \quad \lambda_i \in \mathbb{R} \text{ for } i = 1, .., n. \tag{8.18}$$

Similarly the forward variance ξ_0 admits the following piecewise constant representation

$$\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_0] = \xi_0(t) := \sum_{i=1}^n \xi_i \mathbb{1}_{\{t\in[T_{i-1},T_i)\}}, \quad \xi_i \in \mathbb{R} \text{ for } i = 1, .., n. \tag{8.19}$$

where

$$\xi_i := \frac{\mathcal{V}_{T_i} T_i - \mathcal{V}_{T_{i-1}} T_{i-1}}{T_i - T_{i-1}},$$

and $\mathcal{V}_T := \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{T} \int_0^T V_s \mathrm{d}s \right]$ denotes a variance swap quote from the market.

Our objective in this experiment is to estimate $\{\lambda_1, ..., \lambda_n\}$. Our dataset consists of daily Eurostoxx variance swap quotes for maturities {1M, 3M, 6M, 1Y, 2Y} (Figures 8.1 and 8.2). Figure 8.3 on the other hand, shows the daily realized volatility obtained from Oxford-Man institute data.

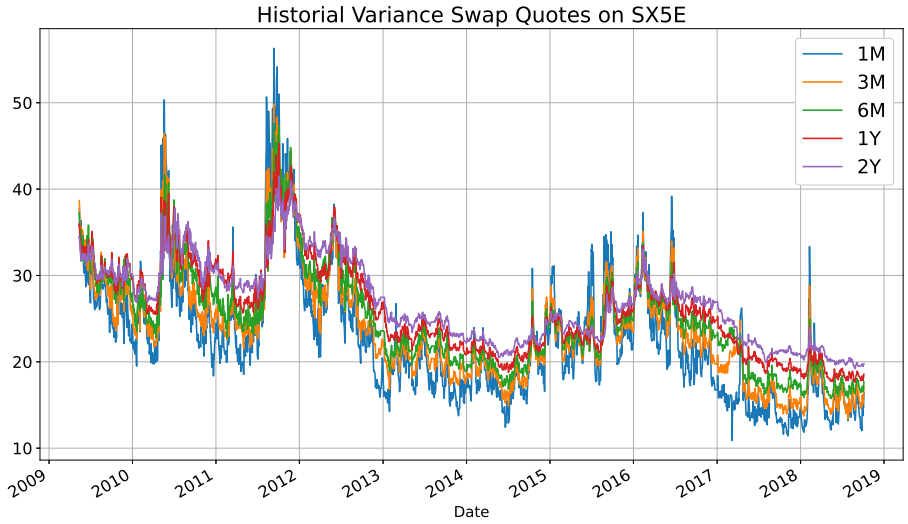


Figure 8.1: Variance Swap volatility daily quotes on SX5E

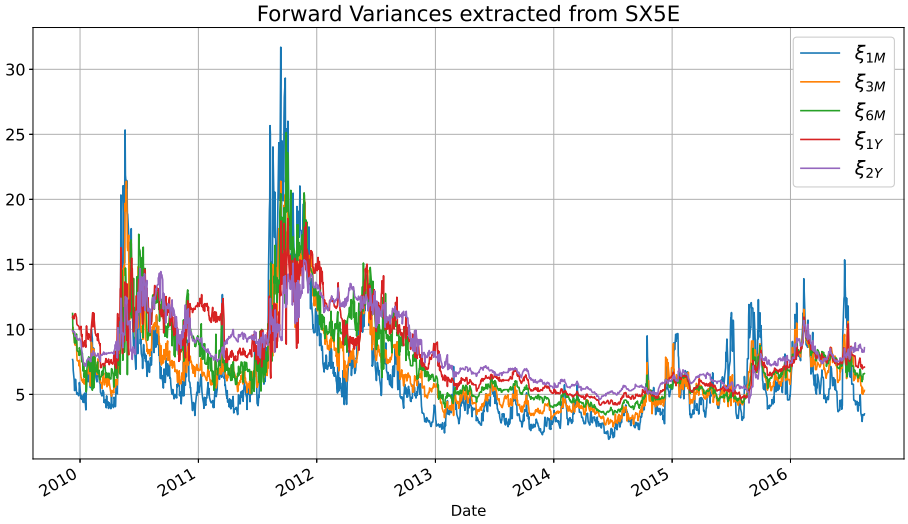


Figure 8.2: Forward variances extracted from variance swap quotes on SX5E

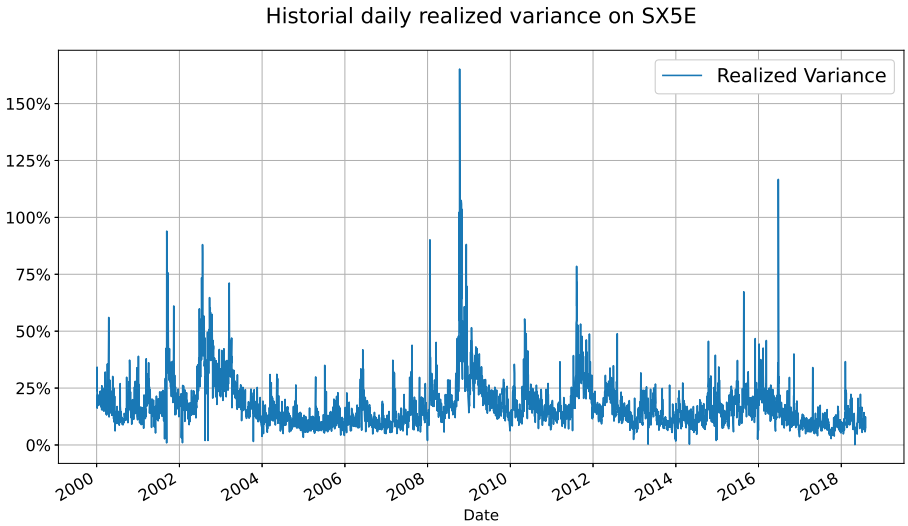


Figure 8.3: Annualised daily realized volatility on SX5E

In order to apply formula (8.14) we need the following ingredients:

(1) Parameters H, ν, ρ

(2) $\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_0]$

(3) $\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_0]$

So far we have obtained (3) from Variance Swap market quotes. Next step is to estimate (1) using historical time-series. Gatheral, Jaisson and Rosenbaum [GJR18], explain how to estimate $\widehat{H}$ and $\widehat{\nu}$ from daily volatility data as the one shown in Figure 8.3, hence we follow their approach using a 100 day rolling window (see Figure 8.4 to see the estimated values) and refer the reader to the original paper for details.

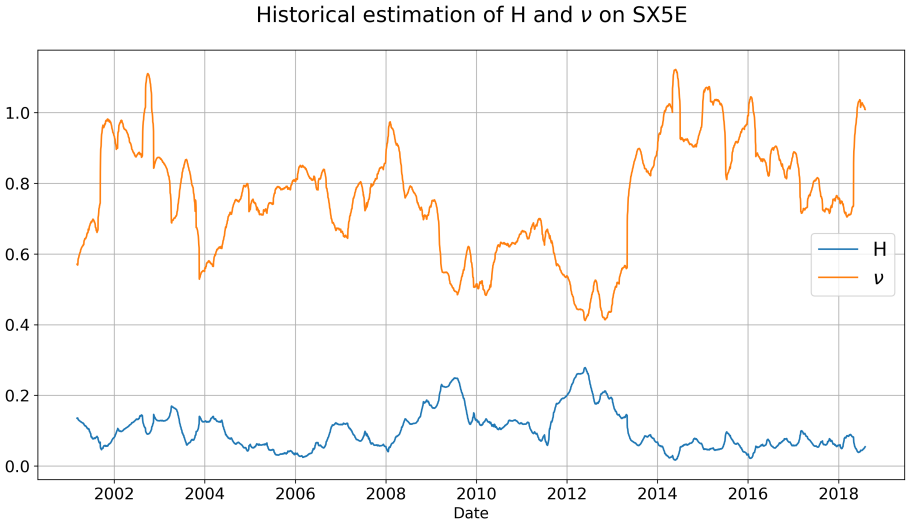


Figure 8.4: Estimated $\hat{H}$ and $\hat{\nu}$ on SX5E

In order to estimate the correlation parameter we use the formula

$$\text{Corr}(Z_t^H - Z_s^H, \int_s^t \mathrm{d}W_s) = \frac{\rho\sqrt{2H}}{H + \frac{1}{2}}, \tag{8.20}$$

which allows us to estimate the correlation with the proxy

$$\widehat{\rho} = \frac{\widehat{H} + \frac{1}{2}}{\sqrt{2\widehat{H}}} \widehat{\mathrm{Corr}} \left(\frac{\log(S_{t_i}) - \log(S_{t_{i-1}})}{\sqrt{v_{t_{i-1}}}}, \log(v_{t_i}) - \log(v_{t_{i-1}}) \right).$$

Figure 8.5 below displays the historical estimates using a estimation window of 100 days.

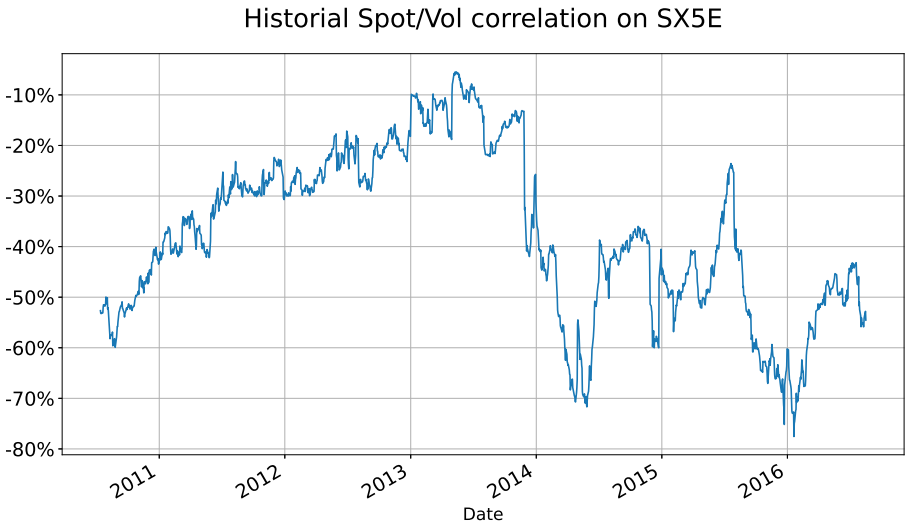


Figure 8.5: Daily correlation estimate on SX5E returns and realized volatility.

In order to forecast volatility and obtain (2) we proceed as in Gatheral, Jaisson and Rosenbaum [GJR18] and use the forecasting formula for the fractional Brownian motion due to Nuzman and Poor [NP00]:

$$Z^H_{t+\Delta} | \mathcal{F}_t \sim \mathcal{N} \left(\frac{\cos(H\pi)}{\pi} \Delta^{H+\frac{1}{2}} \int_{-\infty}^{t} \frac{Z^H_s \mathrm{d}s}{(t-s+\Delta)(t-s)^{H+\frac{1}{2}}}, \frac{C_H \Delta^{2H}}{2H} \right). \tag{8.21}$$

Finally, we orderly estimate λ_i for each $i = 1, \ldots, n$ using Theorem 8.4.1 and the piecewise constant assumption (8.18), as

$$\sum_{j=1}^{i} \lambda_j \int_{T_{j-1}}^{T_j} k(t,u) \mathrm{d}u = \frac{1}{\nu(1-\rho^2)} \log \left(\frac{\xi_0(T_i)}{\mathbb{E}^{\mathbb{P}}[v_{T_i} | \mathcal{F}_0]} \right).$$

Figure 8.6 shows the historical evolution of the risk premium process. In view of Figure 8.6 the process λ clearly seems to follow stochastic dynamics rather than deterministic ones. We leave this direction (the estimation of stochastic risk-premium processes), however, as a line of future research.

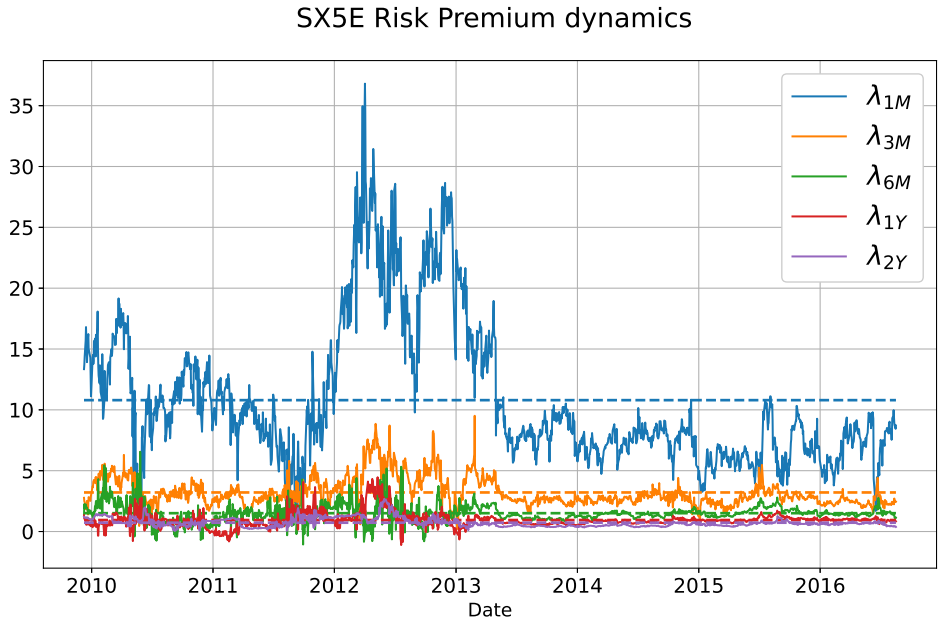


Figure 8.6: Daily risk premium dynamics on SX5E; dashed lines represent the mean value of each process with the corresponding color.

Remark 8.4.3. We would like to emphasize that it is beyond the scope of this Chapter to asses the best method to estimate ingredients (1), (2) and (3). However, we stress the importance of being able to use Theorem 8.4.1 and believe that this empirical work gives a first step towards this end. Several results in the estimation of rough volatility can be found in [CV12; FTW19; MOP20], however the case $H \in (0, \frac{1}{2})$ still remains a challenge, specially for the estimation of the correlation parameter.

8.4.2 Numerical experiment with the rough Bergomi model II: forecasting Variance Swaps

I this section we apply the previously estimated risk-premium process to forecast the price of Variance Swaps 1-day ahead. This same experiment was performend in Bayer, Friz and Gatheral [BFG16], where the authors assumed that $\mathbb{E}^{\mathbb{P}}[V_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s]$ for all $0 \leq s \leq t$ and applied Nuzman and Poor's [NP00] formula (8.21). In this section we shall drop such assumption, however under deterministic λ we still get a formula similar to (8.21):

Proposition 8.4.4 (Volatility forecasting under $\mathbb{Q}$)**.** *Consider the rough volatility model* (8.1) *under* $\mathbb{P}$*. If* $\psi(t,x) = \mathrm{e}^{\nu x}$*,* $\mu_s = r_s$ *for all* $s \geq 0$ *and* $(\lambda_s)_{s\in\mathbb{R}_+} \in L^2(\mathbb{R})$ *is deterministic, then*

$$\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}\left[V_t|\mathcal{F}_s\right] \exp\left(\overline{\rho}\nu C_H \int_s^t k(t,u)\lambda_u \mathrm{d}u\right), \tag{8.22}$$

for $0 \leq s \leq t$*.*

Proof. Using the change of measure result in Theorem 8.2.4, the conditional change of measure formula yields

$$\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}\left[V_t\mathcal{D}_t^{\gamma}|\mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{P}}\left[\mathcal{D}_t^{\gamma}|\mathcal{F}_s\right]}. \tag{8.23}$$

Then, we proceed as in the proof of Theorem 8.2.4 using properties of Gaussian MGF to obtain the result. □

Using Proposition 8.4.4, we immediately obtain a forecast formula for Variance Swaps.

Corollary 8.4.5 (Variance Swap forecasting under $\mathbb{Q}$)**.** *Under model 8.17 and assuming* $\mu_s = r_s$ *for all* $s \geq 0$ *and*

$(\lambda_s)_{s\in\mathbb{R}_+} \in L^2(\mathbb{R})$ *is deterministic, we have*

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{T}\int_{\Delta}^{T+\Delta} V_t \mathrm{d}t\right] = \frac{1}{T}\int_{\Delta}^{T+\Delta} \mathbb{E}^{\mathbb{P}}\left[V_t\right] \exp\left(\rho\nu C_H \int_0^t k(t,s)\lambda_u \mathrm{d}s\right) \mathrm{d}t.$$

Remark 8.4.6. In practice, we use the following discrete-time approximation to obtain the Variance Swap forecast:

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{T}\int_{\Delta}^{T+\Delta} V_t \mathrm{d}t\right] \approx \frac{1}{\lfloor nT\rfloor}\sum_{i=0}^{\lfloor nT\rfloor} \mathbb{E}^{\mathbb{P}}\left[V_{\Delta+\frac{i}{n}}\right] \exp\left(\rho\nu C_H \int_0^{\Delta+\frac{i}{n}} k\left(\Delta+\frac{i}{n}, s\right)\lambda_u \mathrm{d}s\right),$$

where $n = 252$, which corresponds to a time-step of $\frac{1}{252}$.

Fundamentally, Proposition 8.4.4 gives us that the risk-premium adds an extra term to the drift of the volatility process under $\mathbb{Q}$ that needs to be accounted for. We present our numerical experiments in Figures 8.7-8.9, where we have applied Corollary 8.4.5 to (8.17) using the risk-premium we obtained in Section subsec: experimentI. The main contribution here is that we do not observe the overnight effect that was mentioned in Bayer, Friz and Gatheral [BFG16] as our prediction errors oscillate around the origin. In fact, we believe that the discrepancy attributed to the overnight effect in [BFG16] is most likely coming from the risk-premium drift term obtained in (8.22). To conclude, we note that Figures 8.7-8.9 show remarkable agreement, with a superior forecasting power on short-term Variance Swap prediction. We believe that this open the door to an exciting line of research for rough volatility models in which $\mathbb{P}$ and $\mathbb{Q}$ dynamics can be exploited to obtain meaningful forecasts under both measures.

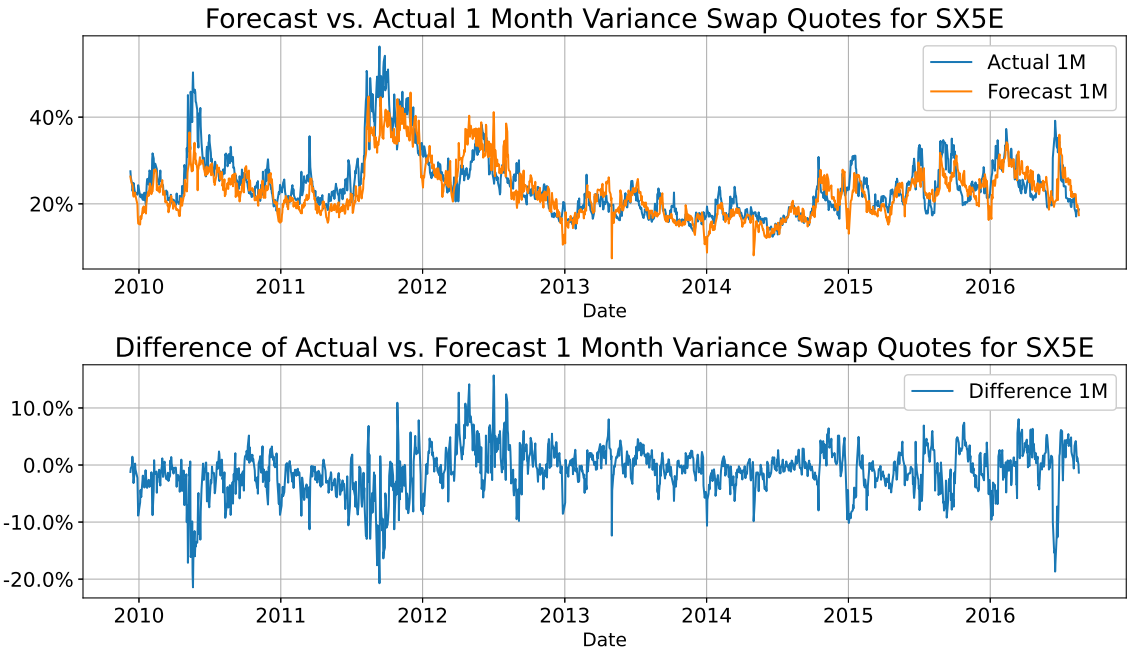


Figure 8.7: Daily 1-day ahead forecasts vs. actual 1 month Variance Swap prices on SX5E

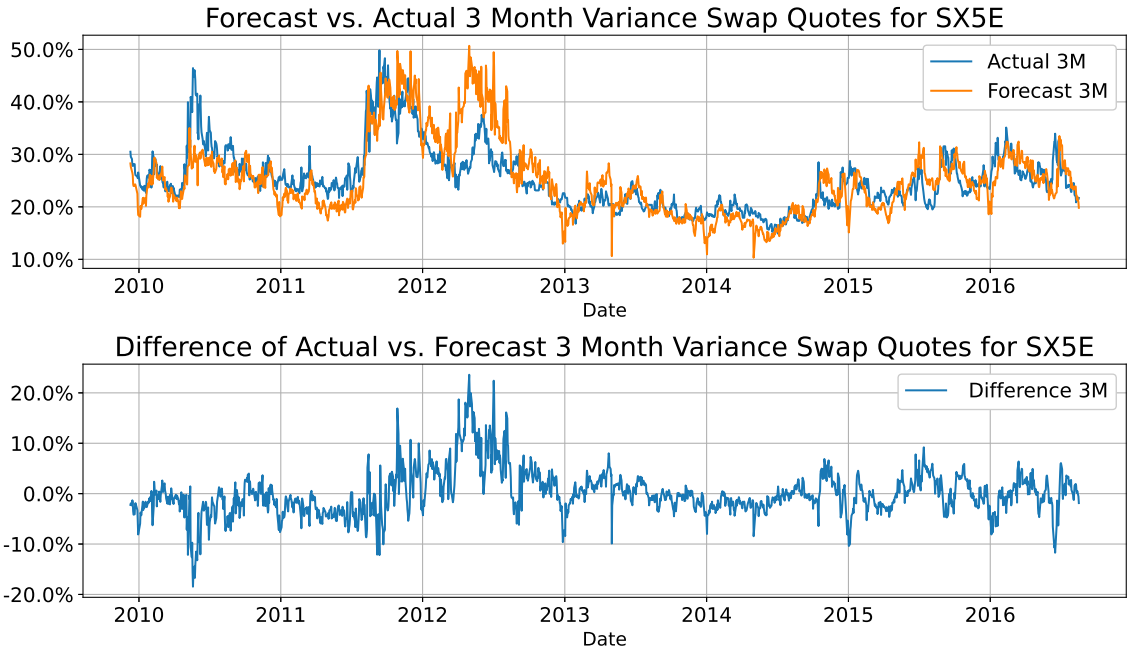


Figure 8.8: Daily 1-day ahead forecasts vs. actual 3 month Variance Swap prices on SX5E

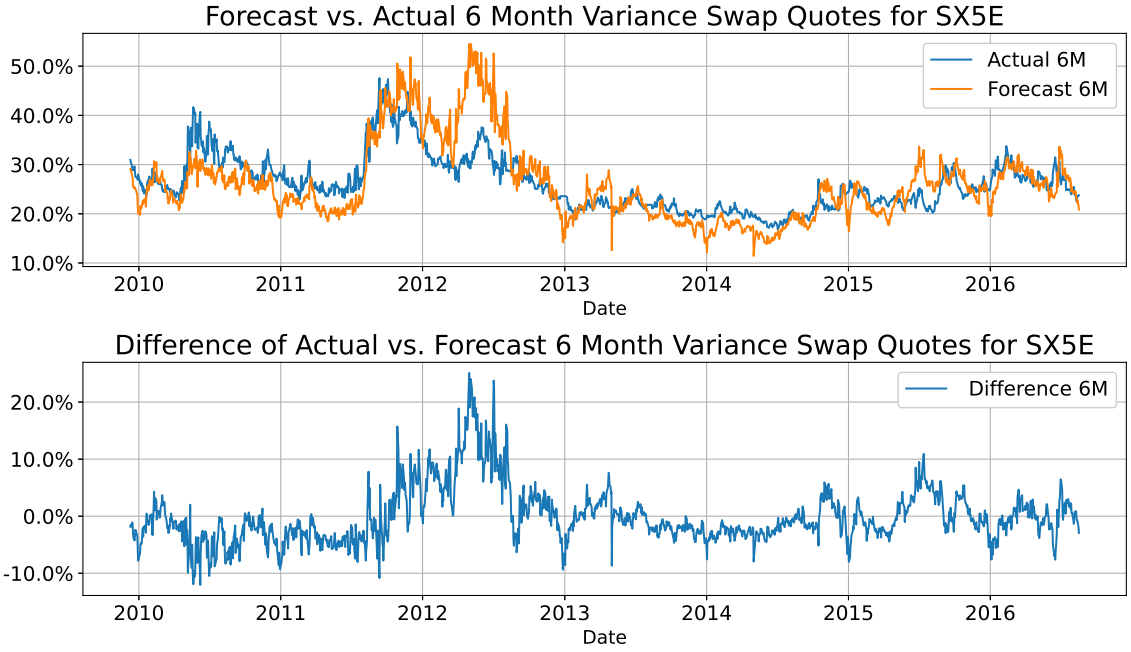


Figure 8.9: Daily 1-day ahead forecasts vs. actual 6 month Variance Swap prices on SX5E

8.5 Summary

In this Chapter we have analysed the change of measure under rough volatility models. We found sufficient conditions on the $\mathbb{P}$ for the Radon-Nikodym derivative to be well defined, as well as to yield a true martingale stock process under the risk-neutral measure $\mathbb{Q}$. We have considered several stochastic constructions of the risk premium process and finally, under the assumption of deterministic risk premium, we have developed a novel estimation method. Furthermore, we have applied the estimated risk-premium process to forecast 1-day ahead Variance Swap

quoted and found agreement between forecasted values and actual market quotes.

A

A.1 Discrete convolution

Definition A.1.1. For $\mathrm{a}, \mathrm{b} \in \mathbb{R}^n$, the discrete convolution operator $* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$(\mathrm{a} * \mathrm{b})_i := \sum_{m=0}^{i} \mathrm{a}_m \mathrm{b}_{i-m}, \quad i = 0, \dots, n-1.$$

When simulating $\mathcal{G}^\alpha W$ on the uniform partition $\mathcal{T}$, the scheme reads

$$(\mathcal{G}^\alpha W)^j(t_i) = \sum_{k=1}^{i} g(t_i - t_{k-1})\xi_k = \sum_{k=1}^{i} g(t_k)\zeta_{j,k-i+1}, \qquad \text{for } i = 1, \dots, n,$$

which has the form of the discrete convolution in Definition A.1.1. Rewritten in matrix form,

$$\begin{pmatrix} g(t_1) & 0 & \cdots & 0 \\ g(t_2) & g(t_1) & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ g(t_n) & g(t_{n-1}) & \cdots & g(t_1) \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{pmatrix},$$

it is clear that this operator yields a complexity of order $\mathcal{O}(n^2)$, which can be improved drastically.

Definition A.1.2. The Discrete Fourier Transform (DFT) of a sequence $\mathrm{c} := (c_0, c_1, ..., c_{n-1}) \in \mathbb{C}^n$ is given by

$$\widehat{f}(\mathrm{c})[j] := \sum_{k=0}^{n-1} c_k \exp\left(-\frac{2\mathtt{i}\pi jk}{n}\right), \qquad \text{for } j = 0, \dots, n-1,$$

and the Inverse DFT of c is given by

$$f(\mathrm{c})[k] := \frac{1}{n} \sum_{j=0}^{n-1} c_j \exp\left(\frac{2\mathtt{i}\pi jk}{n}\right), \qquad \text{for } k = 0, \dots, n-1.$$

In general, both transforms require a computational effort of order $\mathcal{O}(n^2)$, but the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey [CT65] exploits the symmetry and periodicity of complex exponentials of the DFT and reduces the complexity of both transforms to $\mathcal{O}(n \log n)$.

Theorem A.1.3. *For* $\mathrm{a}, \mathrm{b} \in \mathbb{R}^n$*, the identity* $(\mathrm{a} * \mathrm{b}) = f\left(\widehat{f}(\mathrm{a}) \bullet \widehat{f}(\mathrm{b})\right)$ *holds, with* $\bullet$ *the pointwise multiplication.*

This in particular implies that the complexity of the discrete convolution is reduced to $\mathcal{O}(n \log n)$ by FFT.

Algorithm A.1.4 (FFT Discrete convolution for $\mathcal{B}$)**.** On the equidistant grid $\mathcal{T}$,

1. draw a random matrix $\{\zeta_{j,i}\}_{\substack{j=1,\dots,M \\ i=1,\dots,n}}$ such that $\mathbb{V}(\zeta_{j,i}) = 1$;
2. define the vectors $\mathfrak{g} := (g(t_i))_{i=1,\dots,n}$ and $\zeta_j := (\zeta_{j,i})_{i=1,\dots,n}$, for $j = 1, \dots, M$;
3. using FFT, compute $\varphi_j := \widehat{f}(\mathfrak{g}) \cdot \widehat{f}(\zeta_j)$, for $j = 1, \dots, M$;
4. simulate M paths of $(\mathcal{G}^\alpha W)$ using FFT, as $(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}} f(\varphi_j)$ for $j = 1, \dots, M$.

In Step 2 we may replace the evaluation points $\mathfrak{g}$ by any optimal evaluation point $\{g(t_i^*)\}_{i=1}^n$ as in (2.15). Many packages offer a direct implementation of the discrete convolution such as `numpy.convolve` in `Python`. The user then only needs to pass the arguments $\mathfrak{g}$ and ξ_j, and Steps 3 and 4 are computed automatically (using efficient FFT techniques) by the function. Although the FFT step is the heaviest computation on the simulation of rough volatility models, the actual time grid $\mathcal{T}$ is not specially large ($n \ll 1000$). Hence, the fastest FFT for very large n is not essential, as the implementation is run on smaller time grids. In this aspect we find that `numpy.convolve` is a very competitive implementation.

A.2 Riemann-Liouville operators

We review here fractional operators and their mapping properties. We follow closely the excellent monograph by Samko, Kilbas and Marichev [SKM+93], as well as some classical results by Hardy and Littlewood [HL28]. However, we introduce a modification in their definition, so that the condition $f(0) = 0$ is not necessary as opposed to the original definition in Hardy and Littlewood [HL28]

Riemann-Liouville fractional operators

Definition A.2.1. For any $\lambda \in (0,1)$, $\alpha \in \mathfrak{R}^\lambda$ the left Riemann-Liouville fractional operator is defined on $\mathcal{C}^\lambda(\mathbb{I})$ as

$$(I^\alpha f)(t) := \begin{cases} \dfrac{1}{\Gamma(\alpha)} \displaystyle\int_0^t \frac{f(s) - f(0)}{(t-s)^{1-\alpha}} \mathrm{d}s, & \text{for } \alpha \in (0, 1-\lambda), \\ \left(\dfrac{\mathrm{d}}{\mathrm{d}t} I^{1+\alpha} f \right)(t) = \dfrac{1}{\Gamma(1+\alpha)} \dfrac{\mathrm{d}}{\mathrm{d}t} \displaystyle\int_0^t (t-s)^\alpha (f(s) - f(0)) \mathrm{d}s, & \text{for } \alpha \in (-\lambda, 0). \end{cases} \tag{A.1}$$

Theorem A.2.2. *For any $f \in \mathcal{C}^\lambda(\mathbb{I})$, with $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$ the identity*

$$(I^\alpha f)(t) = \psi(t),$$

holds for all $t \in \mathbb{I}$, for some $\psi \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$ satisfying $|\psi(t)| \le C t^{\lambda+\alpha}$ on $\mathbb{I}$ for some $C > 0$.

Proof. We first consider $\alpha > 0$, then may easily represent

$$(I^\alpha f)(t) = \frac{1}{\Gamma(\alpha)} \int_0^t \frac{f(u) - f(0)}{(t-u)^{1-\alpha}} \mathrm{d}u := \psi(t)$$

Since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we obtain $|\psi(t)| \le \dfrac{|f|_\lambda}{\Gamma(\alpha)} \displaystyle\int_0^t \frac{u^\lambda}{(t-u)^{1-\alpha}} \mathrm{d}u$, and hence

$$|\psi(t)| \le \frac{\Gamma(2+\lambda)|f|_\lambda}{(1+\lambda)\Gamma(\alpha+\lambda+1)} t^{\alpha+\lambda},$$

which proves the estimate for $|\psi|$. Next, we prove that $\psi \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. For this, introduce $\phi(t) := f(t) - f(0)$ and consider $t, t+h \in \mathbb{I}$ with $h > 0$,

$$\begin{aligned}\psi(t+h) - \psi(t) &= \frac{1}{\Gamma(\alpha)}\left(\int_{-h}^{t} \frac{\phi(t-u)}{(u+h)^{1-\alpha}}\mathrm{d}u - \int_0^t \frac{\phi(t-u)}{u^{1-\alpha}}\mathrm{d}u\right)\\ &= \frac{\phi(t)}{\Gamma(1+\alpha)}\left[(t+h)^\alpha - t^\alpha\right] + \frac{1}{\Gamma(\alpha)}\left(\int_{-h}^{0} \frac{\phi(t-u)-\phi(t)}{(u+h)^{1-\alpha}}\mathrm{d}u\right)\\ &+ \frac{1}{\Gamma(\alpha)}\left(\int_0^t \left[(u+h)^{\alpha-1} - u^{\alpha-1}\right]\left[\phi(t-u)-\phi(t)\right]\mathrm{d}u\right) =: J_1 + J_2 + J_3.\end{aligned}$$

We first consider J_1. If $h > t$, then

$$|J_1| \le \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^\lambda \left[(t+h)^\alpha - t^\alpha\right] \le Ch^{\lambda+\alpha}.$$

On the other hand, when $0 < h < t$, since $(1+u)^\alpha - 1 \le \alpha u$ for $u > 0$, then

$$|J_1| \le \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^{\lambda+\alpha}\left|\left(1+\frac{h}{t}\right)^\alpha - 1\right| \le Cht^{\lambda+\alpha-1} \le Ch^{\lambda+\alpha}.$$

For J_2, since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we can write

$$|J_2| \le \frac{|f|_\lambda}{\Gamma(\alpha)} \int_{-h}^{0} \frac{|u|^\lambda}{(u+h)^{1-\alpha}} \le Ch^{\lambda+\alpha}.$$

Finally,

$$|J_3| \le \frac{|f|_\lambda}{\Gamma(\alpha)} \int_0^t u^\lambda [u^{\alpha-1} - (u+h)^{\alpha-1}]\mathrm{d}u = \frac{|f|_\lambda}{\Gamma(\alpha)} h^{\lambda+\alpha} \int_0^{t/h} u^\lambda [u^{\alpha-1} - (u+1)^{\alpha-1}]\mathrm{d}u.$$

Hence, if $t \le h$, then $|J_3| \le Ch^{\lambda+\alpha}$. Likewise, if $t > h$ and $\lambda + \alpha < 1$, then $|J_3| \le Ch^{\lambda+\alpha}$ since

$$\left|u^{\alpha-1} - (u+1)^{\alpha-1}\right| = u^{\alpha-1}\left[1 - \left(1+\frac{1}{u}\right)^{\alpha-1}\right] \le Cu^{\alpha-2}.$$

Thus ψ satisfies the $(\lambda+\alpha)$-Hölder condition and belongs to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. Next, we prove the case $\alpha < 0$, for this we use Lemma 13.1 in Samko, Kilbas and Marichev [SKM+93], which gives us

$$(I^\alpha f)(t) = \frac{1}{\Gamma(1+\alpha)}\frac{\mathrm{d}}{\mathrm{d}t}\int_0^t (t-s)^\alpha (f(s) - f(0))\mathrm{d}s = \frac{1}{\Gamma(1+\alpha)}\left(\frac{f(t)-f(0)}{(t-0)^{-\alpha}} + \alpha\int_0^t \frac{f(t)-f(s)}{(t-s)^{1-\alpha}}\right) = \psi_1(t) + \psi_2(t). \tag{A.2}$$

Using (1.16) in [SKM+93] and λ-Hölder property of f directly gives that $\psi_1 \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. Likewise, a similar argument to $\alpha > 0$ gives that $\psi_2 \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$, hence we conclude the assertion.

□

Corollary A.2.3. *For any $\lambda \in (0,1)$ and $\alpha \in \mathfrak{R}^\lambda$, I^α is a continuous operator from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$.*

Proof. It is clear that I^α is a linear operator. Using the estimate in Theorem A.2.2 we have

$$\|I^\alpha f\|_{\alpha+\lambda} \le \|\psi\|_{\lambda+\alpha} \le C_1\|f\|_\lambda \|(\cdot)^{\alpha+\lambda}\|_{\lambda+\alpha} \le C\|f\|_\lambda,$$

since $|f|_\lambda \le \|f\|_\lambda$. Therefore I^α is also bounded and hence continuous. □

A.3 Additional Proofs

A.3.1 Proof of Proposition 2.2.5

Since the paths of Brownian motion are 1/2-Hölder continuous, existence (and continuity) of $\mathcal{G}^\alpha W$ is guaranteed for all $\alpha \in \mathfrak{R}^{1/2}$. When $\alpha \in \mathfrak{R}_+^{1/2}$, the kernel is smooth and square integrable, so that Itô's product rule yields (since $g(0) = 0$)

$$\begin{aligned}(\mathcal{G}^\alpha W)(t) &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)(W(s) - W(0))\mathrm{d}s = g(t)(W(0) - W(0)) - g(0)\left(W(t) - W(0)\right) + \int_0^t g(t-s)\mathrm{d}W_s, \\ &= \int_0^t g(t-s)\mathrm{d}W_s\end{aligned}$$

and the claim holds. For $\alpha \in \mathfrak{R}_-^{1/2}$, and any $\varepsilon > 0$, introduce the operator

$$\left(\mathcal{G}_\varepsilon^{1+\alpha} f\right)(t) := \int_0^{t-\varepsilon} g(t-s)(f(s) - f(0))\mathrm{d}s, \qquad \text{for all } t \in \mathbb{I},$$

which satisfies $\frac{\mathrm{d}}{\mathrm{d}t} \lim_{\varepsilon \downarrow 0} \left(\mathcal{G}_\varepsilon^{1+\alpha} f\right)(t) = \left(\mathcal{G}^{1+\alpha} f\right)(t)$ pointwise. Now, for any $t \in \mathbb{I}$, almost surely,

$$\begin{aligned}\left(\mathcal{G}_\varepsilon^{1+\alpha} W\right)(t) &= g(\varepsilon)\left(W(t-\varepsilon) - W(0)\right) - g(t)\left(W(0) - W(0)\right) + \int_0^{t-\varepsilon} \frac{\mathrm{d}}{\mathrm{d}t} g(t-s) W(s)\mathrm{d}s \\ &= g(\epsilon)W(0) + \int_0^{t-\varepsilon} g(t-s)\mathrm{d}W_s. \end{aligned} \tag{A.3}$$

Then, as ε tends to zero, the right-hand side of (A.3) tends to $\int_0^t g(t-s)\mathrm{d}W_s$, and furthermore, the convergence is uniform. On the other hand, the equalities

$$\begin{aligned}(\mathcal{G}_0^{1+\alpha} W)(t) - (\mathcal{G}_0^{1+\alpha} W)(0) &= \lim_{\varepsilon \downarrow 0} \left[(\mathcal{G}_\varepsilon^{1+\alpha} W)(t) - (\mathcal{G}_\varepsilon^{1+\alpha} W)(0)\right] = \lim_{\varepsilon \downarrow 0} \int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{G}_\varepsilon^{1+\alpha} W\right)(s)\mathrm{d}s \\ &= \int_0^t \lim_{\varepsilon \downarrow 0} \left(\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{G}_\varepsilon^{1+\alpha} W\right)(s)\mathrm{d}s = \int_0^t \left(\int_0^s g(s-u)\mathrm{d}W_u\right)\mathrm{d}s,\end{aligned}$$

hold since convergence is uniform on compacts, and the fundamental theorem of calculus concludes the proof.

B

B.1 A numerical experiment with the inverse map

To motivate the main drawbacks of the inverse map approach of Section 3.3.1, we calibrate rough Bergomi model with it, i.e., we consider the simple map

$$\Pi^{-1}(\Sigma_{\text{BS}}^{\text{rBergomi}}) \to (\hat{\xi_0}, \hat{\nu}, \hat{\rho}, \hat{H})$$

where $\Sigma_{\text{BS}}^{\text{rBergomi}} \in \mathbb{R}^{n \times m}$ is a rBergomi implied volatility surface and $(\hat{\xi_0}, \hat{\nu}, \hat{\rho}, \hat{H})$ the optimal solution to the corresponding calibration problem.

Remark B.1.1. For simplicity we consider the strikes and maturities to be fixed for all implied volatility surfaces.

Inverse Map Architecture

- 1 convolutional layer with 16 filters and 3×3 sliding window
- MaxPooling layer with 2×2 sliding window
- 50 Neuron Feedforward Layer with Elu activation function
- Output layer with linear activation function
- Total number of parameters: 10.014
- Train Set: 34.000 and Test Set: 6.000
- $(\xi_0, \nu, \rho, H) \in \mathcal{U}[0.01, 0.16] \times \mathcal{U}[0.3, 4.0] \times \mathcal{U}[-0.95, -0.1] \times \mathcal{U}[0.025, 0.5]$
- strikes=$\{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5\}$
- maturities=$\{0.1, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.0\}$
- Implied volatilities computed using Algorithm 3.5 in Horvath, Jacquier and Muguruza [HJM17] with 60.000 sample paths and the spot martingale condition i.e. $\mathbb{E}[S_t] = S_0,\ t \geq 0$ as control variate.

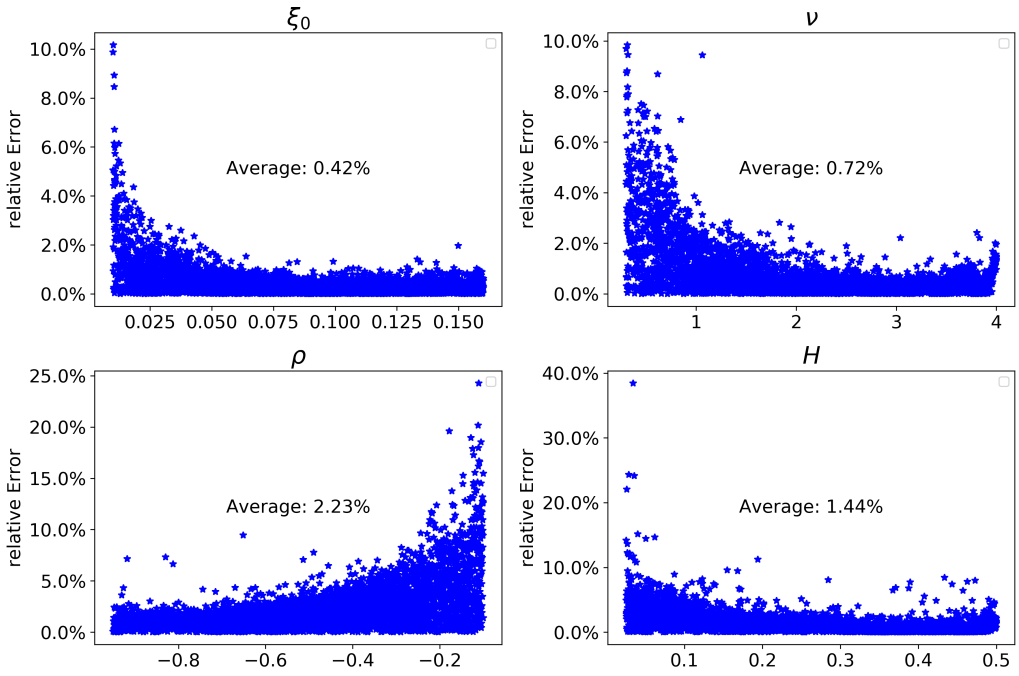


Figure B.1: Out of sample relative errors per parameter calibration

Figure B.1 shows that, indeed it is possible to approximate the inverse map and very sharply calibrate model parameters with a relatively small network. Convolutional networks make sense in this context, since a implied volatility surface has many features both in the strike and maturity direction that can be extracted, similar to image recognition problems. Notice also that the biggest error come from parameter configurations where the Monte Carlo input is more delicate i.e. very small H or very small volatility. Hence, the shape of the errors is intuitively natural and expected beforehand.

Black-Box function and "real" out of sample performance

Let us now consider "real" out of sample data, in the sense that it has not been generated by the rough Bergomi model itself. We generate implied volatility surfaces using the 2 factor Bergomi given by

$$
\begin{aligned}
dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t \\
V_t &= \xi_0(t)\mathcal{E}\left(\nu\left((1-\theta)\int_0^t \exp(-\kappa_X(t-s))dZ_s + \theta\int_0^t \exp(-\kappa_Y(t-s))dY_s\right)\right)
\end{aligned}
$$

where W, Y and Z are correlated standard Brownian motions. We feed smiles from this model into our neural network to obtain the corresponding optimal rough Bergomi parameters. We then compare these values with a direct Monte Carlo calibration via Levenberg-Marquardt [Lev44; Mar63] algorithm. Figure B.2 shows that the neural network does not generalize properly out of sample and concludes that the brute force MC method clearly beats the Inverse map approach. The results by Hernandez [Her17] also support this conclusion, since his out of sample performance (based on different historical period) is reasonably worse than the in-sample one. However, we must emphasize that when the neural network is exposed to familiar situations i.e. surfaces close to the ones generated by the rBergomi model it may work better than the MC approach (see points below the dashed red line in Figure B.2). This is likely due to delicate parameter configurations i.e. very low variance, where MC suffers to obtain accurate estimates whereas the network does not struggle that much.

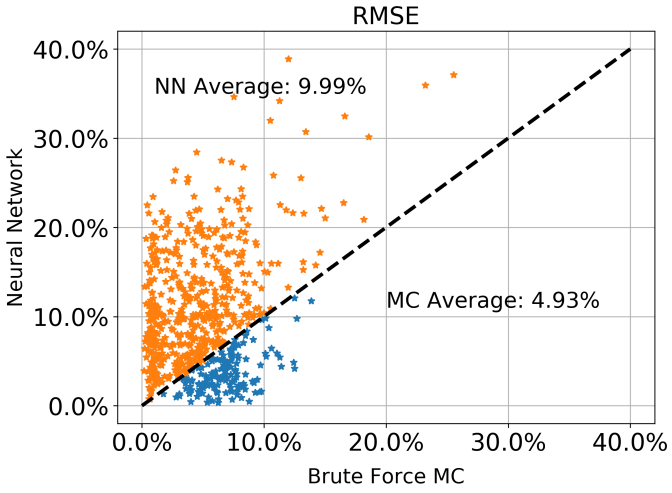


Figure B.2: Stars represent the out of sample RMSE via neural network (NN) in the one-step approach and brute force Monte Carlo (MC). Dashed black line represents the identity function.

The one-step approach does not generalise the problem to all possible settings, since by design is not possible to train Π^{-1} on all possible (arbitrage-free) market scenarios. Moreover, there is a lack of understanding in the highly non-trivial function Π^{-1}, hence from a risk-managing perspective is more difficult to justify the use of this inverse approach than of the direct approach. On the contrary, Figure B.3 reproduces the same experiment with the two-step approach. As expected, by construction the two-step approach solves the calibration problem for any market condition.

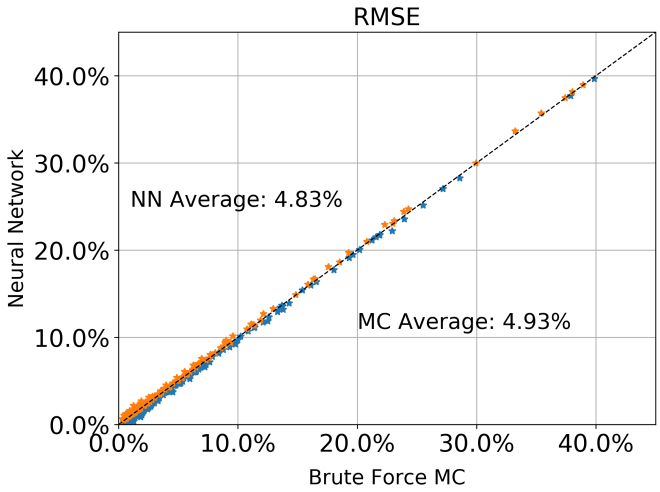


Figure B.3: Stars represent the out of sample RMSE via neural network (NN) using the two-step approach and brute force Monte Carlo (MC). Dashed black line represents the identity function.

C

C.1 Proof

We have

$$\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}>K_j\}}\right]=\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{S_{t_i}>K_j\}}|\mathcal{F}^W_{t_{i-1}}\vee\mathcal{F}^Z_{t_i}\right]\right].$$

Now, we use Theorem 4.4.1 to obtain that $S_i|\mathcal{F}^W_{t_{i-1}}\vee\mathcal{F}^Z_{t_i}$ is lognormal and obtain

$$\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}>K_j\}}\right]=\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\Phi(-d_i(K_j))\right].$$

Additionally, we note

$$0=\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\right]=\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}>K_j\}}\right]+\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}\leq K_j\}}\right].$$

Thus,

$$\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}>K_j\}}\right]=-\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\mathbb{1}_{\{S_{t_i}\leq K_j\}}\right]=-\mathbb{E}^{\mathbb{Q}}\left[D(t_i)\left(r_{t_i}-\bar{r}_t\right)\Phi(d_i(K_j))\right].$$

D

D.1 Proofs of Theorems 6.4.4 and 6.4.6

This section is devoted to the proof of Theorems 6.4.4 and 6.4.6. Towards this end we need the following auxiliary result, which is an adaptation of Theorem 4.2 in [ALV07] to the context of our problem.

Theorem D.1.1. *Assume that hypotheses (H1), (H2) and (H3) hold. Then we have that the at-the-money option price* $P_t = \mathbb{E}_t[(A_T - \mathbb{E}_t[A])_+]$ *is given by*

$$P_t = \mathbb{E}_t\left[BS\left(0, \ln(\mathbb{E}_t[A]), k, u_t\right)\right] + \frac{1}{2}\mathbb{E}_t\left[\int_0^T \frac{\partial G}{\partial x}(s, \ln(M_T(s)), \ln(\mathbb{E}_t[A]), u_s)\Theta_s \mathrm{d}s\right], \tag{D.1}$$

where $G := (\partial^2_{xx} - \partial_x)BS$.

Proof. This proof is based on the same arguments as in the proof of Theorem 4.2 in [ALV07], noting that in our case, the stochastic volatility depends on T. Denote by $\hat{BS}$ the function such that $\hat{BS}(t, x, k, \theta\sqrt{T-t}) := BS\left(t, x, k, \theta\right)$. Then we have that

$$P_t = e^{-r(T-t)}\mathbb{E}_t\left[\hat{BS}\left(T, \ln(A), \ln(\mathbb{E}_t[A]), \sqrt{\int_t^T \phi_s^2 \mathrm{d}s}\right)\right].$$

Then, applying Theorem 6.2.1,using the relationship between BS and $\hat{BS}$ and taking expectations we obtain that

$$\begin{aligned} P_t &= e^{-r(T-t)}\mathbb{E}_t\left[\hat{BS}\left(T, \ln(A), \ln(\mathbb{E}_t[A]), \sqrt{\int_t^T \phi_s^2 \mathrm{d}s}\right)\right] \\ &= \mathbb{E}_t\left[BS\left(t, \ln(\mathbb{E}_t[A]), \ln(\mathbb{E}_t[A]), u_t\right)\right] \\ &+ \mathbb{E}_t\left[\int_t^T \mathcal{L}_{BS}(u_s) BS(s, \ln(\mathbb{E}_s[A]), \ln(A), u_s)\mathrm{d}s\right] \\ &+ \mathbb{E}_t\left[\int_t^T \frac{\partial^2 BS}{\partial x \partial u}(s, \ln(\mathbb{E}_s[A]), \ln(\mathbb{E}_t[A]), u_s)\frac{1}{2u_s\sqrt{T-s}}\Theta_s \mathrm{d}s\right], \end{aligned} \tag{D.2}$$

where $\mathcal{L}_{BS}(u_s)$ denotes the Black-Scholes operator

$$\mathcal{L}_{BS}(u_s) = \left(\frac{1}{2}\left(\partial^2_{xx} - \partial_x\right) + \partial_s - r\cdot\right)u_s.$$

Now the result follows from the fact that $\mathcal{L}_{BS}(u_s)BS(s, X_s, k, u_s) = 0$ and taking into account that $\frac{\partial BS}{\partial u} = u_s\sqrt{T-s}G$. Notice that by (H3) and the fact that

$$\frac{\partial G}{\partial x}(s, x, k, \sigma) = \frac{e^x N'(d_+(k, \sigma))}{\sigma\sqrt{T-s}}\left(1 - \frac{d_+(k, \sigma)}{\sigma\sqrt{T-s}}\right),$$

the integral term in (D.2) is well defined. □

D.1.1 Proof of Theorem 6.4.4

This proof builds upon the arguments in Theorems 3 and 8 in [AS19]. By definition, the ATMI is given by

$$\begin{aligned} I_0^T &= BS^{-1}(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), P_0) \\ &= BS^{-1}(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), \Gamma_T), \end{aligned}$$

where

$$\begin{aligned} \delta_r &= E\left(BS\left(0, \ln(\mathbb{E}[A]), k, u_0\right)\right) \\ &+ \frac{1}{2}\mathbb{E}\left[\int_0^r \frac{\partial G}{\partial x}(s, \ln(M_s^T), \ln(\mathbb{E}[A]), u_s)\Theta_s \mathrm{d}s\right], \end{aligned}$$

Now, Theorem D.1.1 allows us to write

$$\begin{aligned} & I_0^T = \mathbb{E}\left[BS^{-1}(0, \delta_0)\right] \\ + \quad & \mathbb{E}\left[\int_0^T (BS^{-1})'(\ln(\mathbb{E}[A]), \delta_s)\frac{\partial G}{\partial x}(s, \ln(\mathbb{E}_s[A]), \ln(\mathbb{E}[A]), u_s)\Theta_s \mathrm{d}s\right], \end{aligned} \tag{D.3}$$

where $(BS^{-1})'$ denotes the first derivative of BS^{-1} with respect to δ. Now, as

$$\left(BS^{-1}\right)'(k, \delta_s) = \frac{1}{e^x N'(d_+(k, \delta_s))\sqrt{T}},$$

and

$$\frac{\partial G}{\partial x}(s, x, k, \sigma) = \frac{e^x N'(d_+(k, \sigma))}{\sigma\sqrt{T-s}}\left(1 - \frac{d_+(k, \sigma)}{\sigma\sqrt{T-s}}\right),$$

it is easy to see that (H3) implies that the second term in (D.3) tends to zero. For the first term, we can write

$$\begin{aligned} & BS^{-1}(0, \delta_0)) \\ = \quad & \left(BS^{-1}(0, E(BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right))\right) \\ = \quad & \mathbb{E}\left[BS^{-1}(0, BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right))\right] \\ + \quad & \mathbb{E}\left[BS^{-1}(0, \mathbb{E}\left[BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)\right])\right] \\ - \quad & BS^{-1}(0, BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)) \end{aligned}$$

Notice that the first term in the right hand side of the above equation is equal to u_0. For the last two terms we can write

$$BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)) = BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)) + \int_0^T U_s \mathrm{d}W_s, \tag{D.4}$$

where, by the Clark-Ocone formula,

$$\begin{aligned} U_s &= \mathbb{E}_s\left[D_s\left(BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)\right)\right] \\ &= \mathbb{E}_s\left[AN'(d_+\left(\ln(\mathbb{E}[A]), u_0\right))\frac{\int_s^T D_s\phi_s^2 \mathrm{d}s}{2\sqrt{T}u_0}\right]. \end{aligned} \tag{D.5}$$

Define $\Lambda_r = \mathbb{E}_r\left[BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)\right]$. Then, classical Itô's formula and (D.4) give us that

$$\begin{aligned} & \mathbb{E}\left[BS^{-1}(0, \mathbb{E}[BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)])\right] \\ - \quad & BS^{-1}(0, BS\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0\right)) \\ = \quad & \mathbb{E}\left[BS^{-1}(\ln(\mathbb{E}[A]), \Lambda_0) - BS^{-1}(\ln(\mathbb{E}[A]), \Lambda_T)\right] \\ = \quad & -\frac{1}{2}\mathbb{E}\left[\int_0^T \left(BS^{-1}\right)''\left(\ln(\mathbb{E}[A]), \Lambda_r\right) U_r^2 \mathrm{d}r\right]. \end{aligned} \tag{D.6}$$

This, jointly with (D.5), (H3) and the fact that

$$\left(BS^{-1}\right)''\left(X_0, \Lambda_r\right) = \frac{BS^{-1}(X_0, \Lambda_r)}{4\left(\exp(X_t)N'(d_+\left(X_0, BS^{-1}(X_0, \Lambda_r)\right))\right)^2},$$

allows us to prove that the last term in (D.6) tends to zero. Now the proof is complete.

D.1.2 Proof of Theorem 6.4.6

This proof follows the same principles as those given in Proposition 5.1 and Proposition 6.2 in [ALV07] and Theorem 4.5 in [AL17]. Taking partial derivatives with respect to k on the expression $P_0 = BS\left(0, \ln(\mathbb{E}[A]), k, I_0^T(k)\right)$ we obtain

$$\begin{aligned} \frac{\partial P_0}{\partial k} \quad = \quad & \frac{\partial BS}{\partial k}(0, \ln(\mathbb{E}[A]), k, I_0^T(k))) \\ & + \frac{\partial BS}{\partial \sigma}(0, \ln(\mathbb{E}[A]), k, I_0^T(k)))\frac{\partial I_0^T}{\partial k}(k). \end{aligned} \tag{D.7}$$

On the other hand, from (D.3) we deduce that

$$\frac{\partial P_0}{\partial k} = \mathbb{E}_t\left[\frac{\partial BS}{\partial k}(0, \ln(E(A)), k, u_0)\right] + \mathbb{E}\left[\int_0^T \frac{\partial F}{\partial k}(s, \ln(\mathbb{E}_s[A])), k, u_s)\Theta_s \mathrm{d}s\right], \tag{D.8}$$

where

$$F(s, x, k, \sigma) := \frac{1}{2}\frac{\partial G}{\partial x}(s, x, k, \sigma).$$

After some algebra (see [ALV07]) it is easy to see that

$$\begin{aligned} & \mathbb{E}\left[\frac{\partial BS}{\partial k}(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0) - \frac{\partial BS}{\partial k}(t, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), I_0^T))\right] \\ = \quad & \frac{1}{2}\mathbb{E}\left[\int_0^T F(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s)\Theta_s \mathrm{d}s\right], \end{aligned}$$

This, jointly with (D.7) and (D.8) implies that

$$\frac{\partial I_0^T}{\partial k}(\ln(\mathbb{E}[A]) = \frac{\mathbb{E}\left[\int_0^T L(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s)\Theta_s \mathrm{d}s\right]}{\frac{\partial BS}{\partial \sigma}\left(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), I_0^T\right)}, \tag{D.9}$$

where $L := (\frac{1}{2} + \frac{\partial}{\partial k})F$. Now

$$\begin{aligned}
& \mathbb{E}\left[\int_0^T L(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s)\Theta_s \mathrm{d}s\right] \\
= \quad & \mathbb{E}\left[L(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0)\int_0^T \Theta_s \mathrm{d}s\right] \\
+ \quad & \frac{1}{2}\mathbb{E}\left[\int_0^T \left(\frac{\partial^3}{\partial x^3} - \frac{\partial^2}{\partial x^2}\right) L(s, \ln(\mathbb{E}_s[A]), \ln(\mathbb{E}[A]), u_s)\left(\int_s^T \Theta_r \mathrm{d}r\right)\Theta_s \mathrm{d}s\right] \\
+ \quad & \mathbb{E}\left[\int_0^T \frac{\partial L}{\partial k}(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s)\left(\int_s^T D_s\Theta_r \mathrm{d}r\right)\phi_s \mathrm{d}s\right] \\
=: \quad & T_1 + T_2 + T_3. \qquad \text{(D.10)}
\end{aligned}$$

Now, a simple calculation gives us that there exists a positive constant C such that

$$\left(\frac{\partial^3}{\partial x^3} - \frac{\partial^2}{\partial x^2}\right) L(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s) \leq C\sum_{k=4}^{6}((u_s)^2(T-s))^{-\frac{k}{2}}, \qquad \text{(D.11)}$$

$$\frac{\partial L}{\partial k}(s, \ln(\mathbb{E}_s[A])), \ln(\mathbb{E}[A]), u_s) \leq C\sum_{k=3}^{4}((u_s)^2(T-s))^{-\frac{k}{2}}. \qquad \text{(D.12)}$$

Moreover,

$$L(0, \ln(\mathbb{E}[A]), \ln(\mathbb{E}[A]), u_0) = \frac{1}{2}\frac{A\exp\left(-\frac{(u_0)^2T}{8}\right)}{\sqrt{2\pi}u_0\sqrt{T}}\left[\frac{1}{(u_0)^2T} - \frac{1}{2}\right]. \qquad \text{(D.13)}$$

Then, (D.27), (D.12), (D.26) together with (D.9), (D.10) and the fact that

$$\frac{\partial BS}{\partial \sigma}(0, x, x, \sigma) = \frac{A\exp\left(-\frac{\sigma^2T}{8}\right)\sqrt{T}}{\sqrt{2\pi}}$$

allow us to complete the proof.

D.2 Proof of Proposition 6.5.4

Using Remark 6.4.3 we get

$$D_s\phi_u^2 = \frac{2\phi_u(D_sm_u^T)}{M_u^T} - \frac{2\phi_u m_u^T D_s M_u^T}{(M_u^T)^2}, \qquad \text{(D.14)}$$

for $s < u < T$. We recall the reader that in the context of VIX options we have

$$\phi_t^{VIX} = \frac{1}{M_t^{T,VIX}2\Delta}\mathbb{E}_t\left[\frac{1}{VIX_T}\int_T^{T+\Delta}(D_tV_s)\mathrm{d}s\right].$$

and we have the limiting behavior for $0 \leq t \leq T$

$$\begin{aligned}
\lim_{T\to 0}\mathbb{E}\left[\phi_t^{VIX}\right] \quad & = \lim_{T\to 0}\frac{f'(Y_0)}{M_0^{0,VIX}2\Delta VIX_0}\int_T^{T+\Delta}(r-t)^{H-1/2}\exp(-\beta(r-t))\mathrm{d}r \\
& =: \lim_{T\to 0}\frac{f'(Y_0)}{M_0^{0,VIX}2\Delta VIX_0}K(T, \Delta, t) \\
& = \frac{f'(Y_0)}{M_0^{0,VIX}2\Delta VIX_0}J(H, \Delta, \beta).
\end{aligned}$$

On one hand we have,

$$\mathbb{E}_s[D_s M_u^T] = m_s^T. \tag{D.15}$$

Hence, we may rewrite (D.14) under unconditional expectation as

$$\mathbb{E}\left[D_s\phi_u^2\right] = \mathbb{E}\left[\frac{2\phi_u(D_s m_u^T)M_u^T}{M_u^T} - \frac{2\phi_u m_u^T \phi_s}{M_u^T}\right]. \tag{D.16}$$

On the other hand,

$$\begin{aligned}
&D_s m_u^T \\
&= \frac{1}{2\Delta} D_s \mathbb{E}_u\left[\frac{1}{VIX_T}\int_T^{T+\Delta}(D_u(V_r))\mathrm{d}r\right] \\
&= \frac{1}{2\Delta}\mathbb{E}_u\left[\frac{\left(\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r\right)VIX_T - Ds(VIX_T)\int_T^{T+\Delta}(D_u(V_r))\mathrm{d}r}{(VIX_T)^2}\right] \\
&= \mathbb{E}_u\left[\frac{\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r}{2\Delta VIX_T} - \frac{\int_T^{T+\Delta}(D_s(V_r))\mathrm{d}r\int_T^{T+\Delta}(D_u(V_r))\mathrm{d}r}{4\Delta^2\Delta(VIX_T)^3}\right] \\
&= \mathbb{E}_u\left[\frac{\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r}{2\Delta VIX_T} - \frac{m_s^T m_u^T}{VIX_T}\right]
\end{aligned}$$

Therefore we further develop (D.16) into

$$\begin{aligned}
&\mathbb{E}\left[D_s\phi_u^2\right] \\
&= \mathbb{E}\left[\frac{2\phi_u\mathbb{E}_u\left[\frac{\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r}{2\Delta VIX_T} - \frac{m_s^T m_u^T}{VIX_T}\right]}{M_u^T} - \frac{2\phi_u m_u^T\phi_s}{M_s^T}\right] \\
&= \mathbb{E}\left[\frac{\phi_u\left(\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r\right)}{\Delta M_u^T VIX_T} - 2\phi_u m_u^T\phi_s\mathbb{E}_u\left[\frac{1}{M_s^T} + \frac{1}{VIX_T}\right]\right] \\
:= \quad & A(T,s,u) + B(T,s,u).
\end{aligned} \tag{D.17}$$

Taking limits for the expression $B(T,s,u)$ as in Theorem 6.4.6 up to the normalizing terms and using the definition of ϕ_s along with (D.17) we obtain

$$\begin{aligned}
& \lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s^{VIX}\left(\int_s^T B(T,s,u)\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s^{VIX}\left(\int_s^T \frac{-4\phi_u\phi_s^{VIX}m_u^{T,VIX}}{M_0^{0,VIX}}\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \frac{-4(\phi_s^{VIX})^2}{M_s^{T,VIX}}\left(\int_s^T \phi_u^{VIX}m_u^{T,VIX}\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\frac{-4(f'(Y_0))^2}{(M_0^{VIX}(0)2\Delta VIX_T)^2}\int_0^T(\phi_s^{VIX})^2\left(\int_s^T K(T,\Delta,u)^2\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\frac{-4(f'(Y_0))^4}{(M_0^{0,VIX}2\Delta VIX_T)^4}\int_0^T K(T,\Delta,s)^2\left(\int_s^T K(T,\Delta,u)^2\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \frac{-4(f'(Y_0))^4}{(M_0^{0,VIX}2\Delta VIX_0)^4}\frac{1}{2}(J(H,\Delta,\beta))^4
\end{aligned}$$

where in the last step we use $T \geq u \geq s \geq 0$ along with continuity of $K(\cdot)$. Finally, using the ATMI result that gives

$$\lim_{T\to 0} u_0 = \lim_{T\to 0} \mathbb{E}\left[\sqrt{\frac{1}{T}\int_0^T (\phi_s^{VIX})^2 \mathrm{d}s}\right] = \frac{f'(Y_0)}{2\Delta VIX_0^2} J(H,\Delta,\beta),$$

we get

$$\frac{1}{2}\lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\frac{\int_0^T \phi_s\left(\int_s^T B(T,s,u)\mathrm{d}u\right)\mathrm{d}s}{u_0^3}\right] = \lim_{T\to 0} -\frac{1}{2\Delta}\frac{f'(Y_0)}{VIX_0 M_0^{VIX,0}} J(H,\Delta,\beta) \tag{D.18}$$

On the other hand, we proceed similarly with $A(T,s,u)$. First we have

$$\begin{aligned}
&\lim_{T\to 0}\mathbb{E}\left[\frac{1}{2\Delta}\frac{\mathbb{E}_u\left[\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r\right]}{VIX_T}\right] \\
&= \frac{f''(Y_0)}{2\Delta VIX_0}\int_T^{T+\Delta}((r-u)(r-s))^{H-1/2}\exp(-\beta(2r-u-s))\mathrm{d}r \\
&= \frac{f''(Y_0)}{2\Delta VIX_0}\lim_{T\to 0} I(T,\Delta,s,u).
\end{aligned}$$

Moreover, due to the fact that $s < u < T$, we have

$$\lim_{T\to 0}\mathbb{E}\left[\frac{1}{2\Delta}\frac{\mathbb{E}_u\left[\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r\right]}{VIX_T}\right] = \frac{f''(Y_0)}{2\Delta VIX_0} G(H,\Delta,\beta).$$

Carrying similar computations, as with $B(T,s,u)$, we obtain

$$\begin{aligned}
&\lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s\left(\int_s^T A(T,s,u)\mathrm{d}u\right)\mathrm{d}s\right] \\
=\ &\lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s\left(\int_s^T \frac{\phi_u\left(\int_T^{T+\Delta} D_s(D_u(V_r))\mathrm{d}r\right)}{\Delta M_u^T VIX_T}\right)\mathrm{d}s\right] \\
=\ &\frac{f''(Y_0)\Delta^{-1}}{M_0^{0,VIX}VIX_0}\lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s^{VIX}\left(\int_s^T \phi_u^{VIX} I(T,\Delta,u,s)\mathrm{d}u\right)\mathrm{d}s\right] \\
=\ &\lim_{T\to 0}\frac{f''(Y_0)f'(Y_0)\Delta^{-2}}{(M_0^{0,VIX})^2 2VIX_0^2}\frac{1}{T^2}\mathbb{E}\left[\int_0^T \phi_s^{VIX}\left(\int_s^T K(T,\Delta,u)I(T,\Delta,u,s)\mathrm{d}u\right)\mathrm{d}s\right] \\
=\ &\frac{f''(Y_0)(f'(Y_0))^2\Delta^{-3}}{(M_0^{0,VIX})^3 4(VIX_0)^3}\lim_{T\to 0}\frac{1}{T^2}\int_0^T K(T,\Delta,s)\left(\int_s^T K(T,\Delta,u)I(T,\Delta,u,s)\mathrm{d}u\right)\mathrm{d}s \\
=\ &\frac{f''(Y_0)(f'(Y_0))^2}{(M_0^{0,VIX})^3 4(VIX_0)^3}\Delta^{-3}(J(H,\Delta,\beta))^2 G(H,\Delta,\beta).
\end{aligned}$$

were in the last step we use continuity of $I(\cdot)$ and $K(\cdot)$ and the fact $0 \leq s \leq u \leq T$. Finally, recalling the ATMI result for u_0^3 we get

$$\frac{1}{2}\lim_{T\to 0}\frac{1}{T^2}\mathbb{E}\left[\frac{\int_0^T \phi_s^{VIX}\left(\int_s^T A(T,s,u)\mathrm{d}u\right)\mathrm{d}s}{u_0^3}\right] = \frac{G(H,\Delta,\beta)}{2J(H,\Delta,\beta)}\frac{f''(Y_0)}{f'(Y_0)}. \tag{D.19}$$

Then the result follows directly using Theorem 6.4.6 along with expressions ((D.19)) and ((D.18)).

D.3 Computations of the ATMI skew sign in the Heston model

We note that $\mathbb{E}_T[V_r] = \theta + (V_T - \theta)\exp(-k(r-T))$. This implies that, for all $t < T$

$$\begin{aligned} M_t^{T,VIX} &= \mathbb{E}_t\left[\sqrt{\frac{1}{\Delta}\mathbb{E}_T\left[\int_T^{T+\Delta} V_r \mathrm{d}r\right]}\right] \\ &= \mathbb{E}_t\left[\sqrt{\theta + (V_T-\theta)\frac{1}{\Delta}\int_T^{T+\Delta}\exp(-k(r-T))\mathrm{d}r}\right] \\ &= \mathbb{E}_t\left[\sqrt{\theta + (V_T-\theta)\frac{1-\exp(-k\Delta)}{k\Delta}}\right] \\ &\to \mathbb{E}\left[\sqrt{\theta + (V_0-\theta)\frac{1-\exp(-k\Delta)}{k\Delta}}\right] \end{aligned} \tag{D.20}$$

in $L^2(\Omega)$, as $T \to 0$. We know (see [AE08]) that $v \in \mathbb{D}^{1,2}$. Then, for all $s < t < T$

$$D_s M_t^{T,VIX} = \mathbb{E}_t\left[\frac{1}{2VIX_T}\mathbb{E}_T\left[\frac{1}{\Delta}\int_T^{T+\Delta} D_s V_r \mathrm{d}r\right]\right]$$

Now, notice that

$$v_r = v_0\exp(-kr) + \theta\int_0^r \exp(-k(r-u))\mathrm{d}u + \nu\int_0^r \exp(-k(r-u))\sqrt{v_u}\mathrm{d}Wu, \tag{D.21}$$

which allows us to write

$$D_s V_r = \nu\exp(-k(r-s))\sqrt{V_s} + \nu\int_s^r \exp(-k(r-u))D_s\sqrt{v_u}\mathrm{d}Wu. \tag{D.22}$$

Then we get

$$D_s M_t^{T,VIX} \to \frac{\nu(1-\exp(-k\Delta))}{2k\Delta} \tag{D.23}$$

in $L^p(\Omega)$, for all $p > 1$, as $T \to 0$. On the other hand, (6.7) allows us to write

$$\begin{aligned} D_s m_t^{T,VIX} &= \frac{1}{2\Delta}\mathbb{E}_t\left[\frac{1}{VIX_T}\mathbb{E}_T\left[\int_T^{T+\Delta}(D_s D_t V_r)\mathrm{d}r\right]\right] \\ &- \frac{1}{4\Delta}\mathbb{E}_t\left[\frac{1}{\Delta(VIX_T)^3}\mathbb{E}_T\left[\int_T^{T+\Delta}(D_t V_r)\mathrm{d}r\right]\left(\mathbb{E}_T\left[\int_T^{T+\Delta}(D_s V_r)\mathrm{d}r\right]\right)\right] \\ &=: T_1 + T_2. \end{aligned} \tag{D.24}$$

Let us first study the term T_1. It is easy to deduce from (D.22) that, for $s < t < r$

$$\begin{aligned} D_s D_t V_r &= \nu\exp(-k(r-t))D_s\sqrt{V_t} \\ &+\nu\int_t^r \exp(-k(r-u))D_t D_s\sqrt{v_u}\mathrm{d}Wu. \end{aligned} \tag{D.25}$$

Lemma 5.3 in [AE08] gives us that, as $s, t \to 0$, $D_s\sqrt{V_t} \to \frac{\nu}{2}$ in $L^p(\Omega)$, for all $p > 1$. This implies that

$$T_1 \to \frac{\nu^2(1-\exp(-k\Delta))}{4k\Delta\sqrt{V_0}}. \tag{D.26}$$

Finally, we can write

$$T_2 \to -\frac{\nu^2(1-\exp(\Delta))^2}{4k^2\Delta^2\sqrt{V_0}}. \tag{D.27}$$

Then (D.23), (D.26) and (D.27) give us that

$$\begin{aligned}
&D_s m_t^{T,VIX} M_t^{T,VIX} - m_t^{T,VIX} D_s M_t^{T,VIX} \\
&\to \frac{\nu^2(1-\exp(-k\Delta))}{4k\Delta} - \frac{\nu^2(1-\exp(-k\Delta))^2}{4k^2\Delta^2} \\
&- \left(\frac{\nu(1-\exp(-k\Delta)}{2k\Delta}\right)^2 \\
&= \frac{\nu^2(1-\exp(-k\Delta))}{4k\Delta} - 2\frac{\nu^2(1-\exp(-k\Delta))^2}{4k^2\Delta^2} \\
&= \frac{\nu^2(1-\exp(-k\Delta))}{4k\Delta}\left(1 - \frac{2(1-\exp(-k\Delta))}{k\Delta}\right)
\end{aligned} \tag{D.28}$$

as $T \to 0$ in $L^p(\Omega)$, for all $p > 1$.

D.4 Proof of Proposition 6.6.6

Using Remark 6.4.3 we get

$$D_s\phi_u^2 = \frac{2\phi_u(D_s m_u^T)M_u^T}{(M_s^T)^2} - \frac{2\phi_u m_u^T D_s M_u^T}{(M_s^T)^2} =: A(T,s,u) + B(T,s,u)$$

First we will develop $B(T,s,u)$. On one hand we have

$$\mathbb{E}_s[D_s M_u^T] = m_s^T. \tag{D.29}$$

Taking limits for the expression in Theorem 6.4.6 up to the normalizing terms and using the definition of $\phi^R V_s$ along with (D.29) we obtain

$$\begin{aligned}
& \lim_{T\to 0} \mathbb{E}\left[\int_0^T \phi_s^{RV}\left(\int_s^T B(T,s,u)\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0} \mathbb{E}\left[\int_0^T \phi_s\left(\int_s^T \frac{-2\phi_u\phi_s^{RV} m_u^{T,RV}}{M_s^{T,RV}}\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0} \mathbb{E}\left[\int_0^T \frac{-2\phi_s^2}{M_s^{T,RV}}\left(\int_s^T \phi_u^{RV} m_u^{T,RV}\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0} \frac{-2(f'(Y_0))^2}{(M_0^{0,RV})^2(H+1/2)^2T^2}\mathbb{E}\left[\int_0^T (\phi_s^{RV})^2\left(\int_s^T (T-u)^{2H+1}\mathrm{d}u\right)\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0} \frac{-2(f'(Y_0))^2}{(M_0^{0,RV})^2(H+1/2)^2(2H+2)T^2}\mathbb{E}\left[\int_0^T (\phi_s^{RV})^2(T-s)^{2H+2}\mathrm{d}s\right] \\
= \quad & \lim_{T\to 0} \frac{-2(f'(Y_0))^4}{(M_0^{0,RV})^4(H+1/2)^4(2H+2)(4H+4)T^4}T^{4H+4} \\
= \quad & \lim_{T\to 0} \frac{-2(f'(Y_0))^4}{V_0^4(H+1/2)^4(2H+2)(4H+4)}T^{4H}
\end{aligned}$$

Finally, using the ATMI result that gives

$$\lim_{T\to 0} u_0 = \lim_{T\to 0} \mathbb{E}\left[\sqrt{\frac{1}{T}\int_0^T (\phi^{RV})_s^2 \mathrm{d}s}\right] = \frac{|f'(Y_0)|}{(H+1/2)\sqrt{2H+2V_0}} T^{H-1/2},$$

we get

$$\begin{aligned}\frac{1}{2}\lim_{T\to 0}\mathbb{E}\left[\frac{\int_0^T \phi_s^{RV}\left(\int_s^T B(T,s,u)\mathrm{d}u\right)\mathrm{d}s}{u_0^3}\right] &= \lim_{T\to 0}\frac{-f'(Y_0)(2H+2)^{3/2}}{V_0(2H+1)(2H+2)(4H+4)}T^{H+3/2} \\ &= \lim_{T\to 0}\frac{-f'(Y_0)}{V_0(2H+1)\sqrt{(2H+2)}}T^{H+3/2}.\end{aligned}$$

On the other hand, we proceed similarly with $A(T,s,u)$. First we have

$$\begin{aligned}D_s m_u^{T,RV} &= \frac{1}{T}\int_u^T \mathbb{E}_u[D_s(D_u(V_r))]\mathrm{d}r = \frac{1}{T}\int_u^T \mathbb{E}_u\left[(D_s f'(Y_r)(r-u)^{H-1/2}\right]\mathrm{d}r \\ &= \frac{1}{T}\int_u^T f''(Y_r)((r-u)(r-s))^{H-1/2}\mathrm{d}r =: I(T,u,s).\end{aligned}$$

Notice here that

$$\lim_{T\to 0}\mathbb{E}[I(T,u,s)] = \frac{1}{T}f''(Y_0)\frac{(T-u)^{H+1/2}(u-s)^{H-1/2}}{H+1/2}F\left(\frac{T-u}{s-u}\right)$$

where $F(\cdot) = {}_2F_1\left(1/2-H, H+1/2, H+3/2, \cdot\right)$ denotes the Gaussian hypergeometric function. Carrying similar computations, as with $B(T,s,u)$, we obtain

$$\begin{aligned}&\lim_{T\to 0}\mathbb{E}\left[\int_0^T \phi_s^{RV}\left(\int_s^T A(T,s,u)\mathrm{d}u\right)\mathrm{d}s\right] \\ =\ &\lim_{T\to 0}\mathbb{E}\left[\int_0^T \phi_s^{RV}\left(\int_s^T \frac{-2\phi_u^{RV}(D_s m_u^T)M_u^{T,RV}\mathrm{d}u}{(M_s^{T,RV})^2}\right)\mathrm{d}s\right] \\ =\ &\lim_{T\to 0}\frac{-2}{M_0^{0,RV}}\mathbb{E}\left[\int_0^T \phi_s^{RV}\left(\int_s^T \phi_u^{RV} I(T,u,s)\mathrm{d}u\right)\mathrm{d}s\right] \\ =\ &\lim_{T\to 0}\frac{-2f'(Y_0)}{(M_0^{0,RV})^2(H+1/2)T}\mathbb{E}\left[\int_0^T \phi_s^{RV}\int_s^T I(T,u,s)(T-u)^{H+1/2}\mathrm{d}u\mathrm{d}s\right] \\ =\ &\lim_{T\to 0}\frac{-2(f'(Y_0))^2}{(M_0^{0,RV})^3(H+1/2)^2T^2}\int_0^T (T-s)^{H+1/2}\int_s^T I(T,u,s)(T-u)^{H+1/2}\mathrm{d}u\mathrm{d}s \\ =\ &\lim_{T\to 0}\frac{-2(f'(Y_0))^2 f''(Y_0)}{(M_0^{0,RV})^3(H+1/2)^2T^3} \\ &\times\int_0^T (T-s)^{H+1/2}\int_s^T \frac{(T-u)^{2H+1}(u-s)^{H-1/2}}{H+1/2}F\left(\frac{T-u}{s-u}\right)\mathrm{d}u\mathrm{d}s \\ =\ &\lim_{T\to 0}\frac{-2f'(Y_0))^2 f''(Y_0)\mathcal{I}(H)}{V_0^3(H+1/2)^2}\,T^{4H}\end{aligned}$$

where we define $\mathcal{I}(H)$ as the finite limit

$$\mathcal{I}(H) = \lim_{T\to 0}\frac{\displaystyle\int_0^T (T-s)^{H+1/2}\int_s^T \frac{(T-u)^{2H+1}(u-s)^{H-1/2}}{H+1/2}F\left(\frac{T-u}{s-u}\right)\mathrm{d}u\mathrm{d}s}{T^{4H+3}}. \tag{D.30}$$

We may compute this limit is closed-form using the series representation (see Abramowitz and Stegun [ASR88] for details)

$$F\left(\frac{T-u}{s-u}\right) = \begin{cases} \sum_{n=0}^{\infty} \frac{(1/2-H)_n (H+1/2)}{(H+3/2+n)n!} \left(\frac{T-u}{s-u}\right)^n, & \text{if } \frac{T-u}{s-u} < 1 \\ \sum_{n=0}^{\infty} (-1)^n \frac{(1/2-H)_n}{(H+3/2)_n} \left(\frac{T-s}{u-s}\right)^{H-1/2} \left(\frac{T-u}{T-s}\right)^n & \text{otherwise} \end{cases}.$$

Consequently we have,

$$\begin{aligned}
&\int_s^T (T-u)^{2H+1}(u-s)^{H-1/2} F\left(\frac{T-u}{s-u}\right) \mathrm{d}u \\
&= \sum_{n=0}^{\infty} \frac{(1/2-H)_n}{(H+3/2)_n} \int_s^{\frac{s+T}{2}} (T-u)^{2H+1+n}(T-s)^{H-1/2-n} \mathrm{d}u \\
&+ \sum_{n=0}^{\infty} (-1)^n \frac{(1/2-H)_n (H+1/2)}{(H+3/2+n)n!} \int_{\frac{s+T}{2}}^T (T-u)^{2H+1+n}(u-s)^{H-1/2-n} \mathrm{d}u \\
&= (T-s)^{3H+3/2} \Bigg(\sum_{n=0}^{\infty} \frac{(1/2-H)_n}{(H+3/2)_n} \frac{1-2^{-2H-2-n}}{2H+2+n} \\
&+ \sum_{n=0}^{\infty} (-1)^n \frac{(1/2-H)_n (H+1/2)}{(H+3/2+n)n!} \frac{\hat{F}(n,1) - 2^{n-1/2-n}\hat{F}(n,1/2)}{H+1/2-n} \Bigg) \\
&=: (T-s)^{3H+3/2} C(H)
\end{aligned}$$

where $\hat{F}(n,x) = {}_2F_1(-n-2H-1, H+1/2-n, H+3/2-n, x)$. Finally, we get

$$\mathcal{I}(H) = \frac{C(H)}{(H+1/2)(4H+3)}.$$

We notice that in practice, $\mathcal{I}(H)$ can be also easily approximated using numerical integration packages and expression (D.30) for small T, instead of the series representation. Finally, recalling the ATMI result for u_0^3 we get

$$\frac{1}{2} \lim_{T\to 0} \mathbb{E}\left[\frac{\int_0^T \phi_s \left(\int_s^T A(T,s,u) \mathrm{d}u \right) \mathrm{d}s}{u_0^3} \right] = \lim_{T\to 0} \frac{f''(Y_0)(2H+2)^{3/2}(H+1/2)\mathcal{I}(H)}{f'(Y_0)} T^{H+3/2}.$$

To conclude the proof we make use of Theorem 6.4.6 and obtain

$$\begin{aligned}
\lim_{T\to 0} \mathcal{S}_T^{RV}(0) &= \frac{1}{2} \lim_{T\to 0} \mathbb{E}\left[\frac{\int_0^T \left(\phi_s^{RV} \int_s^T D_s (\phi_u^{RV})^2 \mathrm{d}u \right) \mathrm{d}s}{u_0^3 T^2} \right] \\
&= \frac{1}{2} \lim_{T\to 0} \mathbb{E}\left[\frac{\int_0^T \phi_s^{RV} \left(\int_s^T A(T,s,u) \mathrm{d}u \right) \mathrm{d}s}{u_0^3} + \frac{\int_0^T \phi_s^{RV} \left(\int_s^T B(T,s,u) \mathrm{d}u \right) \mathrm{d}s}{u_0^3} \right] \\
&= \lim_{T\to 0} \left(\frac{f''(Y_0)\mathcal{I}(H)(2H+2)^{3/2}(H+1/2)}{f'(Y_0)} - \frac{f'(Y_0)}{V_0(2H+1)\sqrt{(2H+2)}} \right) T^{H-1/2}.
\end{aligned}$$

E

E.1 Approximating the density of realised variance in the mixed rough Bergomi model

In light of the numerical results shown in Section 7.4 (see Figures 7.2-7.6) we identify a clear linear trend in the implied volatility smiles generated by both the rough Bergomi and mixed rough Bergomi models. Therefore, it is natural to postulate the following conjecture/approximation of log-linear smiles.

Assumption E.1.1. The implied volatility of realised variance options in the mixed rough Bergomi (7.6) model is linear in log-moneyness, and takes the following form:

$$\hat{\sigma}(K,T) = \left(T^{\beta} \left(a(\alpha,\gamma,\nu) + b(\alpha,\gamma,\nu) \log\left(\frac{K}{RV(V)(0)} \right) \right) \right)^{+}$$

where

$$a(\alpha,\gamma,\nu) = \frac{\sqrt{2\alpha+1}\sum_{i=1}^{n}\gamma_i\nu_i}{(\alpha+1)\sqrt{2\alpha+3}},$$
$$b(\alpha,\gamma,\nu) = \sqrt{2\alpha+1}\left(\frac{\sum_{i=1}^{n}\gamma_i\nu_i^2}{\sum_{i=1}^{n}\gamma_i\nu_i}\mathcal{I}(\alpha)(2\alpha+3)^{3/2}(\alpha+1) - \frac{\sum_{i=1}^{n}\gamma_i\nu_i}{(2\alpha+2)\sqrt{(2\alpha+3)}} \right),$$

with

$$\mathcal{I}(\alpha) = \frac{\left(\sum_{n=0}^{\infty} \frac{(\alpha)_n}{(\alpha+2)_n} \frac{1-2^{-2\alpha-3-n}}{2\alpha+3+n} + \sum_{n=0}^{\infty}(-1)^n \frac{(-\alpha)_n(\alpha+1)}{(\alpha+2+n)n!} \frac{\hat{F}(n,1) - 2^{n-1/2-n}\hat{F}(n,1/2)}{\alpha+1-n} \right)}{(\alpha+1)(4\alpha+5)}$$

such that $\hat{F}(n,x) = {}_2F_1(-n-2\alpha-2, \alpha+1-n, \alpha+2-n, x)$ and $(x)_n = \prod_{i=0}^{n-1}(x+i)$ represents the rising Pochhammer factorial.

Remark E.1.2. The values of the constants $a(\alpha,\gamma,\nu)$ and $b(\alpha,\gamma,\nu)$ in Assumption E.1.1, which give the level and slope of the implied volatility respectively, are given in [AGM18, Example 24 and Example 27] respectively; we generalise to n factors. These results are given in terms of the Hurst parameter H; to avoid any confusion we will continue with our use of α. Recall that, by Remark 7.2.2, $\alpha = H - 1/2$.

Proposition E.1.3. *Under Assumption E.1.1 , the density of $RV(V^{(\gamma,\nu)})(T)$ is given by*

$$\psi_{RV}(x,T) = -\phi(d_2(x))\frac{\partial d_1(x)}{\partial x}\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}d_1(x) + 1 \right), \quad x \geq 0$$

where $d_1(x) = \frac{\log(V_0)-\log(x)}{\hat{\sigma}(x,T)\sqrt{T}} + \frac{1}{2}\hat{\sigma}(x,T)\sqrt{T}$, $d_2(x) = d_1(x) - \hat{\sigma}(x,T)\sqrt{T}$ *for* $x \geq 0$ *and* $\phi(\cdot)$ *is the standard Gaussian probability density function.*

Proof. Let us denote

$$C(K,T) := \mathbb{E}[(RV(V^{(\gamma,\nu)})(T) - K)^+].$$

The well-known Breeden-Litzenberger formula [BL78] tells us that

$$\frac{\partial^2 C(x,T)}{\partial x^2}\bigg|_{x=K} = \psi_{RV}(K,T).$$

Under Assumption E.1.1, we have that

$$C(K,T) = BS(V_0, \hat{\sigma}(K,T), K, T)$$

where $BS(V_0,\sigma,K,T) = V_0\Phi(d_1) - K\Phi(d_2)$ is the Black-Scholes Call pricing formula with Φ the standard Gaussian cumulative distribution function. Then, differentiating C with respect to the strike gives

$$\frac{\partial C(x,T)}{\partial x}\bigg|_{x=K} = V_0\phi(d_1(K))\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} - x\phi(d_2(K))\frac{\partial d_2(x)}{\partial x}\bigg|_{x=K} - \Phi(d_2(K))$$

where

$$\begin{aligned}\frac{\partial d_1(x)}{\partial x} &= \frac{-\hat{\sigma}(x,T) + \log(x/V_0)a(\alpha,\gamma,\nu)T^\alpha}{x\hat{\sigma}(x,T)^2\sqrt{T}} + \frac{1}{2}\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{x} \\ &= \frac{-b(\alpha,\gamma,\nu)T^\alpha}{x\hat{\sigma}(x,T)^2\sqrt{T}} + \frac{1}{2}\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{x}\end{aligned}$$

and

$$\frac{\partial d_2(x)}{\partial x} = \frac{\partial d_1(x)}{\partial x} - \frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{x}.$$

Using the well known identity $V_0\phi(d_1(x)) = x\phi(d_2(x))$, proved in Appendix E.5, we further simplify

$$\frac{\partial C(K,T)}{\partial K} = V_0\phi(d_1(K))\left(\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{K}\right) - \Phi(d_2(K)).$$

Differentiating again we obtain,

$$\psi_{RV}(K,T) = -V_0\phi(d_1(K))\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{K}\left(d_1(K)\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} + \frac{1}{K}\right) - \phi(d_2(K))\frac{\partial d_2(x)}{\partial x}\bigg|_{x=K}.$$

Then, by using $V_0\phi(d_1(x)) = x\phi(d_2(x))$, we find that

$$\psi_{RV}(K,T) = -\phi(d_2(K))\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}\left(d_1(x)\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} + \frac{1}{K}\right) + \frac{\partial d_2(x)}{\partial x}\bigg|_{x=K}\right),$$

which we further simplify to

$$\psi_{RV}(K,T) = -\phi(d_2(K))\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K}\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}d_1(K) + 1\right),$$

and the result then follows. Note that the density $\psi_{RV}(\cdot,T)$ is indeed continuous for all $T > 0$. □

Remark E.1.4. Note that Proposition E.1.3 gives the density of $RV(V^{(\gamma,\nu)})(T)$ in closed-form. In addition, Proposition E.1.3 can be easily used to get the density of the Arithmetic Asian option under the Black-Scholes model. This would correspond to the case $\alpha = 0$ and $\nu = \sigma > 0$ as the Black-Scholes constant volatility.

Remark E.1.5. Assuming the density ψ_{RV} exists, we have the following volatility swap price:

$$\mathbb{E}[\sqrt{RV(V^{(\gamma,\nu)})(T)}] = \int_0^\infty \sqrt{x}\psi_{RV}(x,T)\mathrm{d}x.$$

In Figure E.1, we provide numerical results for the volatility swap approximation, which performs best for short maturities, due to the nature of the approximation being motivated by small-time smile behaviour. Interestingly, it captures rather accurately the short time decay of the Volatility Swap price for maturities less than 3 months; for larger maturities the absolute error does not exceed 20 basis points.

Figure E.1: Volatility Swap Monte Carlo price estimates (straight lines) and LDP based approximation (stars) for $\eta = 1.5$ and $V_0 = 0.04$; for Monte Carlo we use $200,000$ simulations and $\Delta t = \frac{1}{1008}$.

E.2 Proof of Main Results

E.2.1 Proof of Theorem 7.3.3

Proof. For $t \in \mathcal{T}$, $\varepsilon > 0$, we first define the rescaled processes

$$\begin{aligned} Z_t^\varepsilon &:= \varepsilon^{\beta/2} Z_t \overset{\mathrm{d}}{=} Z_{\varepsilon t}, \\ V_t^\varepsilon &:= V_0 \exp\left(Z_t^\varepsilon - \frac{\eta^2}{2}(\varepsilon t)^\beta \right), \end{aligned} \tag{E.1}$$

where $\beta := 2\alpha + 1 \in (0,1)$. From Schilder's Theorem [DS89, Theorem 3.4.12] and Proposition 7.3.2, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. We now prove that the two sequences of stochastic processes Z^ε and $\widetilde{Z}^\varepsilon := Z^\varepsilon - \frac{\eta^2}{2}(\varepsilon\cdot)^\beta$ are exponentially equivalent [DZ10, Definition 4.2.10]. For each $\delta > 0$ and $t \in \mathcal{T}$, there exists $\varepsilon_* := \frac{1}{t}\left(\frac{2\delta}{\eta^2}\right)^{1/\beta} > 0$ such that

$$\sup_{t\in\mathcal{T}} |Z_t^\varepsilon - \widetilde{Z}_t^\varepsilon| = \sup_{t\in\mathcal{T}} |\frac{\eta^2}{2}(\varepsilon t)^\beta| \le \delta,$$

for all $0 < \varepsilon < \varepsilon_*$. Therefore, for all $\delta > 0$, $\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta \log \mathbb{P}(\|Z^\varepsilon - \widetilde{Z}^\varepsilon\|_\infty > \delta) = -\infty$, and the two processes are indeed exponentially equivalent. Then, using [DZ10, Theorem 4.2.13], the sequence of stochastic processes $(\widetilde{Z}^\varepsilon)_{\varepsilon>0}$ also satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$, with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. Moreover, for all ε, t, we have that $V_t^\varepsilon = V_0 \exp(\widetilde{Z}_t^\varepsilon)$, where the bijective transformation $x \mapsto V_0 \exp(x)$ is clearly continuous with respect to the sup norm metric. Therefore we can apply the Contraction Principle [DZ10, Theorem 4.2.1], which is stated in Appendix E.4, concluding that the sequence of processes $(V^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function $\Lambda^Z\left(\log\left(\frac{x}{V_0}\right)\right)$. Here we have used that, for each $\varepsilon > 0$, $t \in \mathcal{T}$ and $\mathrm{x} \in \mathcal{C}(\mathcal{T})$, the inverse mapping of the bijection transformation $x \mapsto V_0 \exp(x)$ is given by $\log\left(\frac{x}{V_0}\right)$. Since, for all $t \in \mathcal{T}$ and $\varepsilon > 0$, V_t^ε and $v_{\varepsilon t}$ are equal in law, we obtain that the processes $(V_t^\varepsilon)_{t\in\mathcal{T}}$ and $(V_{\varepsilon t})_{t\in\mathcal{T}}$ are also equal in law, which concludes the theorem. Notice also that $\Lambda^V(V_0) = \Lambda^Z(0) = \|0\|^2_{\mathscr{H}^{K_\alpha}} = 0$. □

E.2.2 Proof of Corollary 7.3.7

Proof. The proof of Equation (7.11) is similar to the proof of [FZ17, Corollary 4.9], and we shall prove the lower and upper bound separately, which turn out to be equal. Firstly, as the rate function $\hat{\Lambda}^V$ is continuous on $\mathcal{C}(\mathcal{T})$, we have that, for all $k > 0$,

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{P}(\log[RV(V)(t)] > k) = -\mathrm{I}(k),$$

as an application of Corollary 7.3.4.

(1) The proof of the lower bound is exactly the same as presented in [FZ17, Appendix C] and will be omitted here; we arrive at $\liminf_{t\downarrow 0} t^\beta \log \mathbb{E}\left[(RV(V)(t) - \mathrm{e}^k)^+\right] \ge -\mathrm{I}(k)$.

(2) We establish the upper bound:
We first apply Hölder's inequality:

$$\begin{aligned} \mathbb{E}[(RV(V)(t) - \mathrm{e}^k)]^+ &= \mathbb{E}\left[(RV(V)(t) - \mathrm{e}^k)^+ \mathbb{1}_{\{RV(V)(t) \ge \mathrm{e}^k\}}\right], \\ &\le \mathbb{E}\left[\left((RV(V)(t) - \mathrm{e}^k\right)^+\right)^q\right]^{1/q} \mathbb{P}(RV(V)(t) \ge \mathrm{e}^k)^{1-1/q}, \\ &\le \mathbb{E}\left[(RV(V)(t))^q\right]^{1/q} \mathbb{P}(RV(V)(t) \ge \mathrm{e}^k)^{1-1/q} \end{aligned}$$

which holds for all $q > 1$. Thus

$$t^\beta \log \mathbb{E}[(RV(V)(t) - \mathrm{e}^k)]^+ \leq \frac{t^\beta}{q} \log \mathbb{E}\left[(RV(V)(t))^q\right] + t^\beta \left(1 - \frac{1}{q}\right) \log \mathbb{P}\left(RV(V)(t) \geq \mathrm{e}^k\right). \tag{E.2}$$

We further obtain the following inequality by applying Jensen's inequality and the fact that all moments exist for $(RV(V)(t))^q$

$$\mathbb{E}\left[(RV(V)(t))^q\right] \leq \frac{1}{t^q} \int_0^t \mathbb{E}[v_s^q] \mathrm{d}s \leq \frac{v_0^q}{t^q} \int_0^t \exp\left(\left(\frac{q^2\eta^2}{2} - \frac{q\eta^2}{2}\right) s^{2\alpha+1}\right) \mathrm{d}s \leq \frac{v_0^q}{t^{q-1}} \exp\left(\left(\frac{q^2\eta^2}{2} - \frac{q\eta^2}{2}\right) t^{2\alpha+1}\right) \tag{E.3}$$

Therefore,

$$\lim_{q\uparrow\infty} \limsup_{t\downarrow 0} \frac{t^\beta}{q} \log \mathbb{E}\left[(RV(V)(t))^q\right] \leq \lim_{q\uparrow\infty} \limsup_{t\downarrow 0} \frac{t^\beta}{q} \left(q \log V_0 - (q-1)\log t + \left(q(q-1)\frac{\eta^2}{2} t^{2\alpha+1}\right)\right) = 0.$$

Hence, taking $q \uparrow \infty$ and $t \downarrow 0$ on both sides of (E.2), we obtain by Corollary 7.3.4

$$\limsup_{t\downarrow 0} t^\beta \log \mathbb{E}[(RV(V)(t) - \mathrm{e}^k)]^+ \leq -\mathrm{I}(k),$$

which concludes the proof.

The proof of Equation (7.12) follows the same steps, after proving that the process $\sqrt{RV(V)}$ satisfies a large deviations principle on $\mathbb{R}_+$. Indeed, as the function $x \mapsto x^2$ is a continuous bijection on $\mathbb{R}_+$, we have that the square root of the integrated variance process $\sqrt{RV(V)}$ satisfies a large deviations principle on $\mathbb{R}_+$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^V((\cdot)^2)$, using [DZ10, Theorem 4.2.4]. □

E.2.3 Proof of Theorem 7.3.10

Proof. For brevity we set $n = 2$, but for larger n, identical arguments can be applied. From Schilder's Theorem [DS89, Theorem 3.4.12] and Proposition 7.3.2, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. Define the operator $f : \mathcal{C}(\mathcal{T}) \to \mathcal{C}((\mathcal{T}), \mathbb{R}^2)$ by $f(\mathrm{x}) := (\frac{\nu_1}{\eta}\mathrm{x}, \frac{\nu_2}{\eta}\mathrm{x})$, which is clearly continuous with respect to the sup-norm $\|\cdot\|_\infty$ on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$. Applying the Contraction Principle then yields that the sequence of two-dimensional processes $((\frac{\nu_1}{\eta}Z^\varepsilon, \frac{\nu_2}{\eta}Z^\varepsilon))_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ as ε tends to zero with speed $\varepsilon^{-\beta}$ and rate function

$$\tilde{\Lambda}(\mathrm{y}, \mathrm{z}) := \inf\{\Lambda^Z(\mathrm{x}) : f(\mathrm{x}) = (\mathrm{y}, \mathrm{z})\} = \inf\{\Lambda^Z(\frac{\eta}{\nu_1}\mathrm{y}) : \mathrm{z} = \frac{\nu_2}{\nu_1}\mathrm{y}\}.$$

Identical arguments to the proof of Theorem 7.3.3 give that the sequences of processes $((\frac{\nu_1}{\eta}Z^\varepsilon, \frac{\nu_2}{\eta}Z^\varepsilon))_{\varepsilon>0}$ and $((\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^\beta, \frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu^2}{2}(\varepsilon\cdot)^\beta))_{\varepsilon>0}$ are exponentially equivalent, thus satisfy the same large deviations principle, with the same rate function and the same speed.

We now define the operator $g^\gamma : \mathcal{C}(\mathcal{T}, \mathbb{R}^2) \to \mathcal{C}(\mathcal{T})$ as $g^\gamma(\mathrm{x}, \mathrm{y}) = V_0(\gamma e^{\mathrm{x}} + (1-\gamma)e^{\mathrm{y}})$. For small perturbations $\delta^{\mathrm{x}}, \delta^{\mathrm{y}} \in \mathcal{C}(\mathcal{T})$ we have that

$$\sup_{t\in\mathcal{T}} |g^\gamma(\mathrm{x} + \delta^{\mathrm{x}}, \mathrm{y} + \delta^{\mathrm{y}}) - g^\gamma(\mathrm{x}, \mathrm{y})| \leq |V_0| \left(\sup_{t\in\mathcal{T}} |\gamma e^{\mathrm{x}(t)}(e^{\delta^{\mathrm{x}}(t)} - 1)| + \sup_{t\in\mathcal{T}} |(1-\gamma)e^{\mathrm{y}(t)}(e^{\delta^{\mathrm{y}}(t)} - 1)|\right).$$

Clearly the right hand side tends to zero as $\delta^{\mathrm{x}}, \delta^{\mathrm{y}}$ tends to zero; thus the operator g^γ is continuous with respect to the sup-norm $\|\cdot\|_\infty$ on $\mathcal{C}(\mathcal{T})$. Applying the Contraction Principle then yields that the sequence of processes $(v^{(\varepsilon,\gamma,\nu)})_{\varepsilon>0} := \left(V_0\left(\gamma \exp(\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^\beta) + (1-\gamma)\exp(\frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu_2^2}{2}(\varepsilon\cdot)^\beta)\right)\right)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function

$$\mathrm{x} \mapsto \inf\{\tilde{\Lambda}(\mathrm{y}, \mathrm{z}) : \mathrm{x} = g^\gamma(\mathrm{y}, \mathrm{z})\} = \inf\{\Lambda^Z(\frac{\eta}{\nu_1}\mathrm{y}) : \mathrm{x} = g^\gamma(\mathrm{y}, \frac{\nu_2}{\nu_1}\mathrm{y})\} = \inf\{\Lambda^Z(\frac{\eta}{\nu_1}\mathrm{y}) : \mathrm{x} = V_0(\gamma e^{\mathrm{y}} + (1-\gamma)e^{\frac{\nu_2}{\nu_1}\mathrm{y}})\}.$$

Since, for all $\varepsilon > 0$ and $t \in \mathcal{T}$, $V_{\varepsilon t}^{(\gamma,\nu)}$ and $v_t^{(\varepsilon,\gamma,\nu)}$ are equal, the theorem follows immediately. Identical arguments to the proof of Theorem 7.3.3 then yield that $\Lambda^\gamma(V_0) = 0$. □

E.2.4 Proof of Theorem 7.3.14

Proof. We begin by introducing a rescaling of (7.7) for $\varepsilon > 0$, so that the system becomes

$$V_t^{(\gamma,\nu,\Sigma,\varepsilon)} := V_{\varepsilon t}^{(\gamma,\nu,\Sigma)} = V_0 \sum_{i=1}^{n} \gamma_i \mathcal{E}\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}_t^\varepsilon\right), \tag{E.4}$$

with the rescaled process $\mathcal{Z}_t^\varepsilon$ defined as

$$\mathcal{Z}_t^\varepsilon := \mathcal{Z}_{\varepsilon t} = \varepsilon^{\alpha+\frac{1}{2}} \left(\int_0^t K_\alpha(s,t)\mathrm{d}W_s^1, ..., \int_0^t K_\alpha(s,t)\mathrm{d}W_s^m\right).$$

The m-dimensional sequence of processes $(\varepsilon^{\beta/2}\mathcal{Z}^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$ as ε goes to zero with speed $\varepsilon^{-\beta}$ and rate function Λ^m defined by $\Lambda^m(\mathrm{x}) := \frac{1}{2}\|\mathrm{x}\|^2_{\mathscr{H}_m}$ for $\mathrm{x} \in \mathscr{H}_m$ and $+\infty$ otherwise. Then, $\mathscr{H}_m$ is the reproducing kernel Hilbert space of the measure induced by $\mathcal{Z}$ on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$, defined as

$$\mathscr{H}_m := \left\{(g^1,\cdots,g^m) \in \mathcal{C}(\mathcal{T},\mathbb{R}^m) : g^i(t) = \int_0^t K_\alpha(s,t) f^i(s)\mathrm{d}s, f^i \in L^2(\mathcal{T}) \text{ for all } i \in \{1\cdots m\}\right\}.$$

Then, using an extension of the proof of [HJL19, Theorem 3.6], for $i = 1,\cdots,n$, the sequence of m-dimensional processes $(\mathrm{L}_i\mathcal{Z}_\cdot^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$, as ε tends to zero with rate function $\mathrm{y} \mapsto \inf\{\Lambda^m(\mathrm{x}) : \mathrm{x} \in \mathscr{H}_m, \mathrm{y} = \mathrm{L}_i\mathrm{x}\}$ and speed $\varepsilon^{-\beta}$, where L_i is the lower triangular matrix introduced in Model (7.7). Consequently, for $i = 1,\cdots,n$ each (one-dimensional) sequence of processes $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathcal{Z}_\cdot^\varepsilon\right)_{\varepsilon>0}$ also satisfies a large deviations principle as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function $\Lambda_{\Sigma_i}(\mathrm{y}) := \inf\left\{\Lambda^m(\mathrm{x}) : \mathrm{x} \in \mathscr{H}_m, \mathrm{y} = \frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathrm{x}\right\}$.

Each sequence of processes $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathcal{Z}_\cdot^\varepsilon\right)_{\varepsilon>0}$ and $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathcal{Z}_\cdot^\varepsilon - \frac{1}{2}\nu^i\Sigma_i\nu^i(\varepsilon\cdot)^\beta\right)_{\varepsilon>0}$ are exponentially equivalent for $i = 1,\cdots,n$; therefore they satisfy the same large deviations principle with the same speed $\varepsilon^{-\beta}$ and the same rate function Λ_{Σ_i}.

We now define the operator $g^\gamma : \mathcal{C}(\mathcal{T},\mathbb{R}^n) \to \mathcal{C}(\mathcal{T})$ as

$$g^\gamma(\mathrm{x})(\cdot) := V_0 \sum_{i=1}^{n} \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot \mathrm{x}(\cdot)\right),$$

with $\mathrm{x} := (\mathrm{x}_1,\cdots,\mathrm{x}_n)$. For small perturbations $\delta^1,\cdots,\delta^n \in \mathcal{C}(\mathcal{T})$ with $\delta := (\delta^1,\cdots,\delta^n)$, we have that

$$\begin{aligned}\sup_{t\in\mathcal{T}} |g^\gamma(\mathrm{x}+\delta)(t) - g^\gamma(\mathrm{x})(t)| &= \sup_{t\in\mathcal{T}} \left| V_0 \sum_{i=1}^{n} \gamma_i \exp(\frac{\nu^i}{\eta} \cdot (\mathrm{x}(t)+\delta(t))) - \exp(\frac{\nu^i}{\eta} \cdot \mathrm{x}(t))\right| \\ &\leq \sup_{t\in\mathcal{T}} |V_0| \sum_{i=1}^{n} \left|\exp(\frac{\nu^i}{\eta} \cdot \mathrm{x}(t))(\exp(\delta(t)) - 1)\right|\end{aligned}$$

The right-hand side tends to zero as $\delta^1,\cdots,\delta^n$ tends to zero; thus the operator g^γ is continuous with respect to the sup-norm $\|\cdot\|_\infty$ on $\mathcal{C}(\mathcal{T})$. Using that $V_t^{(\gamma,\nu,\Sigma,\varepsilon)} = g^\gamma(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i\mathcal{Z}_\cdot^\varepsilon - \frac{1}{2}\nu^i\Sigma_i\nu^i(\varepsilon\cdot)^\beta)(t)$ for each $\varepsilon > 0$ and $t \in \mathcal{T}$, we can apply the Contraction Principle (see Appendix E.4), yielding that the sequence of processes $(V^{(\gamma,\nu,\Sigma,\varepsilon)})_{\varepsilon>0}$

satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function

$$\begin{aligned}\mathrm{y} &\mapsto \inf\left\{\Lambda_{\Sigma_i}(\mathrm{x}) : \mathrm{y} = V_0 \sum_{i=1}^{n} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \mathrm{x}\right)\right\} \\ &= \inf\left\{\Lambda^m(\mathrm{x}) : \mathrm{x}\in\mathcal{H}_m, \mathrm{y} = V_0 \sum_{i=1}^{n} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \mathrm{L}_i\mathrm{x}\right)\right\}.\end{aligned}$$

As with the previous two models we have that, for all $\varepsilon > 0$ and $t \in \mathcal{T}$, $V_t^{(\gamma,\nu,\Sigma,\varepsilon)}$ and $V_{\varepsilon t}^{(\gamma,\nu,\Sigma)}$ are equal in law and so the result follows directly. □

E.2.5 Proof of Proposition 7.5.2

Proof. The proof of Proposition 7.5.2, which is similar to the proofs given in [JPS18], is made up of three parts. The first part is to prove that $\left(\mathscr{H}^{g,T}, \langle\cdot,\cdot\rangle_{\mathscr{H}^{g,T}}\right)$ is a separable Hilbert space. Clearly $\mathcal{I}^{g,T}$ is surjective on $\mathscr{H}^{g,T}$. Now take $f_1, f_2 \in L^2[0,T]$ such that $\mathcal{I}^{g,T}f_1 = \mathcal{I}^{g,T}f_2$. For any $t \in [T, T+\Delta]$ it follows that $\int_0^T g(t,s)[f_1(s) - f_2(s)]\mathrm{d}s = 0$; applying the Titchmarsh convolution theorem then implies that $f_1 = f_2$ almost everywhere and so $\mathcal{I}^{g,T} : L^2[0,T] \to \mathscr{H}^{g,T}$ is a bijection. $\mathcal{I}^{g,T}$ is a linear operator, and therefore $\langle\cdot,\cdot\rangle_{\mathscr{H}^{g,T}}$ is indeed an inner product; hence $\left(\mathscr{H}^{g,T}, \langle\cdot,\cdot\rangle_{\mathscr{H}^{g,T}}\right)$ is a real inner product space. Since $L^2[0,T]$ is a complete (Hilbert) space, for every sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in L^2[0,T]$, there exists a function $\widetilde{f} \in L^2[0,T]$ such that $\{f_n\}_{n\in\mathbb{N}}$ converges to $\widetilde{f} \in L^2[0,T]$; note also that $\{\mathcal{I}^{g,T}f_n\}_{n\in\mathbb{N}}$ converges to $\mathcal{I}^{g,T}\widetilde{f}$ in $\mathscr{H}^{g,T}$. Assume for a contradiction that there exists $f \in L^2[0,T]$ such that $f \neq \widetilde{f}$ and $\{\mathcal{I}^{g,T}f_n\}_{n\in\mathbb{N}}$ converges to $\mathcal{I}^{g,T}f$ in $\mathscr{H}^{g,T}$, then, since $\mathcal{I}^{g,T}$ is a bijection, the triangle inequality yields

$$0 < \left\|\mathcal{I}^{g,T}f - \mathcal{I}^{g,T}\widetilde{f}\right\|_{\mathscr{H}^{g,T}} \leq \left\|\mathcal{I}^{g,T}f - \mathcal{I}^{g,T}f_n\right\|_{\mathscr{H}^{g,T}} + \left\|\mathcal{I}^{g,T}\widetilde{f} - \mathcal{I}^{g,T}f_n\right\|_{\mathscr{H}^{g,T}},$$

which converges to zero as n tends to infinity. Therefore $f = \widetilde{f}$, $\mathcal{I}^{g,T}f \in \mathscr{H}^{g,T}$ and $\mathscr{H}^{g,T}$ is complete, hence a real Hilbert space. Since $L^2[0,T]$ is separable with countable orthonormal basis $\{\phi_n\}_{n\in\mathbb{N}}$, then $\{\mathcal{I}^{g,T}\phi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $\mathscr{H}^{g,T}$, which is then separable.

The second part of the proof is to show that there exists a dense embedding $\iota : \mathscr{H}^{g,T} \to \mathcal{C}[T, T+\Delta]$. Since there exists $h \in L^2[0,T]$ such that $\int_0^\varepsilon |h(s)|\mathrm{d}s$ for all $\varepsilon > 0$ and $g(t,\cdot) = h(t-\cdot)$ for any $t \in [T, T+\Delta]$, we can apply by [Che08, Lemma 2.1], which tells us that $\mathscr{H}^{g,T}$ is dense in $\mathcal{C}[T, T+\Delta]$ and so we choose the embedding to be the inclusion map.

Finally we must prove that every $f^* \in \mathcal{C}[T, T+\Delta]^*$, where f^* is defined in [JPS18, Definition 3.1], is Gaussian on $\mathcal{C}[T, T+\Delta]$, with variance $\|\iota^* f^*\|_{\mathscr{H}^{g,T^*}}$, where ι^* is the dual of ι. Take $f^* \in \mathcal{C}[T, T+\Delta]^*$, then by the fact that $(\mathcal{C}[T, T+\Delta], \mathcal{H}^{g,T}, \mu)$ is a RKHS triplet by similar arguments to [FZ17, Lemma 3.1], we obtain using [JPS18, Remark 3.6] that ι^* admits an isometric embedding $\bar{\iota}^*$ such that

$$\|\bar{\iota}^* f^*\|_{\mathscr{H}^{g,T^*}} = \|f^*\|_{L^2(\mathcal{C}[T,T+\Delta],\mu)} = \int_{\mathcal{C}[T,T+\Delta]} (f^*)^2 \mathrm{d}\mu = \mathrm{VAR}(f^*),$$

where μ is the Gaussian measure induced by the process $\int_0^T g(\cdot,s)\mathrm{d}s$ on $(\mathcal{C}[T, T+\Delta], \mathscr{B}(\mathcal{C}[T, T+\Delta]))$. □

E.2.6 Proof of Theorem 7.5.6

Proof. First we recall $\widetilde{V}_t^{g,T,\varepsilon} := V_t^{g,T,\varepsilon} - \frac{\varepsilon^\gamma}{2}\int_0^t g^2(t,u)\mathrm{d}u + \varepsilon^{\gamma/2}$. We begin the proof by showing that the sequence of processes $(V^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ and $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ are exponentially equivalent [DZ10, Definition 4.2.10]. As $g(t,\cdot) \in L^2[0,T]$ for all $t \in [T, T+\Delta]$, for each $\delta > 0$ there exists $\varepsilon_* > 0$ such that $\sup_{t\in[T,T+\Delta]}\left|\varepsilon_*^{\gamma/2} - \frac{\varepsilon_*^\gamma}{2}\int_0^T g^2(t,u)\mathrm{d}u\right| \leq \delta$. Therefore, for the $\mathcal{C}[T, T+\Delta]$ norm $\|\cdot\|_\infty$ we have that for all $\varepsilon_* > \varepsilon > 0$,

$$\mathbb{P}\left(\left\|V^{g,T,\varepsilon} - \widetilde{V}^{g,T,\varepsilon}\right\|_\infty > \delta\right) = \mathbb{P}\left(\sup_{t\in[T,T+\Delta]}\left|\varepsilon^{\gamma/2} - \frac{\varepsilon^\gamma}{2}\int_0^T g^2(t,u)\mathrm{d}u\right| > \delta\right) = 0.$$

Therefore $\limsup_{\varepsilon\downarrow 0}\varepsilon^{\gamma}\log\mathbb{P}\left(\left\|V^{g,T,\varepsilon}-\widetilde{V}^{g,T,\varepsilon}\right\|_{\infty}>\delta\right)=-\infty$, and so the two sequences of processes $(V^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ and $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ are exponentially equivalent; applying [DZ10, Theorem 4.2.13] then yields that $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ satisfies a large deviations principle on $\mathcal{C}[T,T+\Delta]$ with speed $\varepsilon^{-\gamma}$ and rate function Λ^V.

We now prove that the operators φ_{1,ξ_0} and φ_2 are continuous with respect to the $\mathcal{C}([T,T+\Delta]\times[0,1])$ and $\mathcal{C}[0,1]$ $\|\cdot\|_\infty$ norms respectively. The proofs are very simple, and are included for completeness. First let us take a small perturbation $\delta^f\in\mathcal{C}([T,T+\Delta]\times[0,1])$:

$$\begin{aligned}\left\|\varphi_{1,\xi_0}(f+\delta^f)-\varphi_{1,\xi_0}(f)\right\|_\infty &= \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}\left|\xi_0(s)e^{f(s,\varepsilon)}\left(e^{\delta^f(s,\varepsilon)}-1\right)\right| \\ &\leq \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|\xi_0(s)|\sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|e^{f(s,\varepsilon)}|\sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|e^{\delta^f(s,\varepsilon)}-1|.\end{aligned}$$

Since ξ_0 is continuous on $[T,T+\Delta]$ and f is continuous on $[T,T+\Delta]\times[0,1]$, they are both bounded. Clearly $e^{\delta^f(s,\varepsilon)}-1$ tends to zero as δ^f tends to zero and hence the operator φ_{1,ξ_0} is continuous. Now take a small perturbation $\delta^f\in\mathcal{C}([T,T+\Delta]\times[0,1])$:

$$\left\|\varphi_2(f+\delta^f)-\varphi_2(f)\right\|_\infty=\sup_{\varepsilon\in[0,1]}\left|\frac{1}{\Delta}\int_T^{T+\Delta}\delta^f(s,\varepsilon)\mathrm{d}s\right|\leq M,$$

where $M:=\sup_{\varepsilon\in[0,1]}\delta^f(s,\varepsilon)$. Clearly M tends to zero as δ^f tends to zero, thus the operator φ_2 is also continuous.

For every $s\in[T,T+\Delta]$ we have the following: by an application of the Contraction Principle [DZ10, Theorem 4.2.1] and using the fact that $\varepsilon\mapsto(\varphi_{1,\xi_0}f)(s,\varepsilon)$ is a bijection for all $f\in\mathcal{C}[T,T+\Delta]$ it follows that the sequence of stochastic processes $\left(\left(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}\right)(s,\varepsilon)\right)_{\varepsilon\in[0,1]}$ satisfies a large deviations principle on $\mathcal{C}[0,1]$ as ε tends to zero with speed $\varepsilon^{-\gamma}$ and rate function

$$\hat{\Lambda}_s^V(\mathrm{y}):=\Lambda^V\left((\varphi_{1,\xi_0}\mathrm{y})^{-1}(s,\cdot)\right)=\Lambda^V\left(\log\left(\frac{\mathrm{y}(s,\cdot)}{\xi_0(s)}\right)\right).$$

A second application of the Contraction Principle then yields that the sequence of stochastic processes

$$\left(\ (\varphi_2(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\right)_{\varepsilon\in[0,1]}$$

satisfies a large deviations principle on $\mathcal{C}[0,1]$ with speed $\varepsilon^{-\gamma}$ and rate function $\Lambda^{\mathrm{VIX}}(\mathrm{x})=\inf_{s\in[T,T+\Delta]}\{\Lambda^V\left((\varphi_{1,\xi_0}\mathrm{y})^{-1}(s,\cdot)\right):\mathrm{x}(\cdot)=(\varphi_2\mathrm{y})(\cdot)\}$. By definition, the sequence of processes $\left(\ (\varphi_2(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\right)_{\varepsilon\in[0,1]}$ is almost surely equal to the rescaled VIX processes $(e^{\varepsilon^{\gamma/2}}\mathrm{VIX}_{T,\varepsilon^{\gamma/2}})_{\varepsilon\in[0,1]}$ and hence the satisfies the same large deviations principle.

□

E.3 Numerical recipes

We first consider the simple rough Bergomi (7.5) model for sake of simplicity and further develop the mixed multi-factor rough Bergomi (7.7) model in Appendix E.3.3 (which also includes (7.6)). Therefore, we tackle the numerical computation of the rate function

$$\hat{\Lambda}^V(\mathrm{y}):=\inf\left\{\Lambda^V(\mathrm{x}):\mathrm{y}=RV(\mathrm{x})(1)\right\}.$$

This problem, in turn, is equivalent to the following optimisation:

$$\hat{\Lambda}^V(\mathrm{y}):=\inf_{f\in L^2[0,1]}\left\{\frac{1}{2}||f||^2:\mathrm{y}=RV\left(\exp\left(\int_0^{\cdot}K_\alpha(u,\cdot)f(u)\mathrm{d}u\right)\right)(1)\right\}. \tag{E.5}$$

A natural approach is to consider a class of functions that is dense in $L^2[0,1]$. The Stone-Weierstrass theorem states that any continuous function on a closed interval can be uniformly approximated by a polynomial function. Consequently, we consider a polynomial basis,

$$\hat{f}^{(n)}(s) = \sum_{i=0}^{n} a_i s^i$$

such that $\{\hat{f}^{(n)}\}_{a_i \in \mathbb{R}}$ is dense in $L^2[0,1]$ as n tends to $+\infty$. Problem (E.5) may then be approximated via

$$\hat{\Lambda}_n^V(\mathrm{y}) := \inf_{a \in \mathbb{R}^{n+1}} \left\{ \frac{1}{2} ||\hat{f}^{(n)}||^2 : \mathrm{y} = RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot) \hat{f}^{(n)} \mathrm{d}u(u) \right) \right)(1) \right\},$$

where $a = (a_0, ..., a_n)$. In order to obtain the solution, first the constraint $\mathrm{y} = RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot) \hat{f}^{(n)}(u) \mathrm{d}u \right) \right)(1)$ needs to be satisfied. To accomplish this, we consider anchoring one of the coefficients in $\hat{f}^{(n)}$ such that

$$a_i^* = \underset{a_i \in \mathbb{R}}{\operatorname{argmin}} \left\{ \left(\mathrm{y} - RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot) \hat{f}^{(n)}(u) \mathrm{d}u \right) \right)(1) \right)^2 \right\} \tag{E.6}$$

and the constraint will be satisfied for all combinations of the vector $a^* = (a_0, ..., a_{i-1}, a_i^*, a_{i+1}, ..., a_n)$. Numerically, (E.6) is easily solved using a few iterations of the Newton-Raphson algorithm. Then we may easily solve

$$\inf_{a^* \in \mathbb{R}^{n+1}} \left\{ \frac{1}{2} ||\hat{f}^{(n)}||^2 \right\}$$

which will converge to the original problem (E.5) as $n \to +\infty$. The polynomial basis is particularly convenient since we have that

$$RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot) \hat{f}^{(n)}(u) \mathrm{d}u \right) \right)(1) = \int_0^1 \exp\left(\eta \sqrt{2\alpha+1} \sum_{i=0}^{n} \frac{a_i s^{\alpha+1+i} \, {}_2F_1(i+1, -\alpha, i+2, 1)}{i+1} \right) \mathrm{d}s, \tag{E.7}$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function. In particular one may store the values $\{{}_2F_1\left(i+1, -\alpha, i+2, 1\right))\}_{i=0}^n$ in the computer memory and reuse them through different iterations. In addition, the outer integral in (E.7) is efficiently computed using Gauss-Legendre quadrature i.e.

$$RV\left(\left(\int_0^{\cdot} K_\alpha(u,\cdot) \hat{f}^{(n)}(u) \mathrm{d}u \right) \right)(1) \approx \frac{1}{2} \sum_{k=1}^{m} \exp\left(\eta \sqrt{2\alpha+1} \sum_{i=1}^{n} \frac{a_i \left(\frac{1}{2}(1+p_k) \right)^{\alpha+1+i} \, {}_2F_1(i+1, -\alpha, i+2, 1)}{i+1} \right) w_k,$$

where $\{p_k, w_k\}_{k=1}^m$ are m-th order Legendre points and weights respectively.

E.3.1 Convergence Analysis of the numerical scheme

In Figure E.2 we report the absolute differences $|\hat{\Lambda}_{n+1}^V(\mathrm{y}) - \hat{\Lambda}_n^V(\mathrm{y})|$, for which we observe that the distance decreases as we increase n, the degree of the approximating polynomial. We must enphasize that both numerical routines performing the minimisation steps have a default tolerance of 10^{-8}, hence we cannot expect to obtain higher accuracies than the tolerance and that is why Figure E.2 becomes noisy once this acuracy is obtained. We note that in Figure E.2 we observe a fast convergence, and with just $n = 5$ we usually obtain accuracy of 10^{-5}.

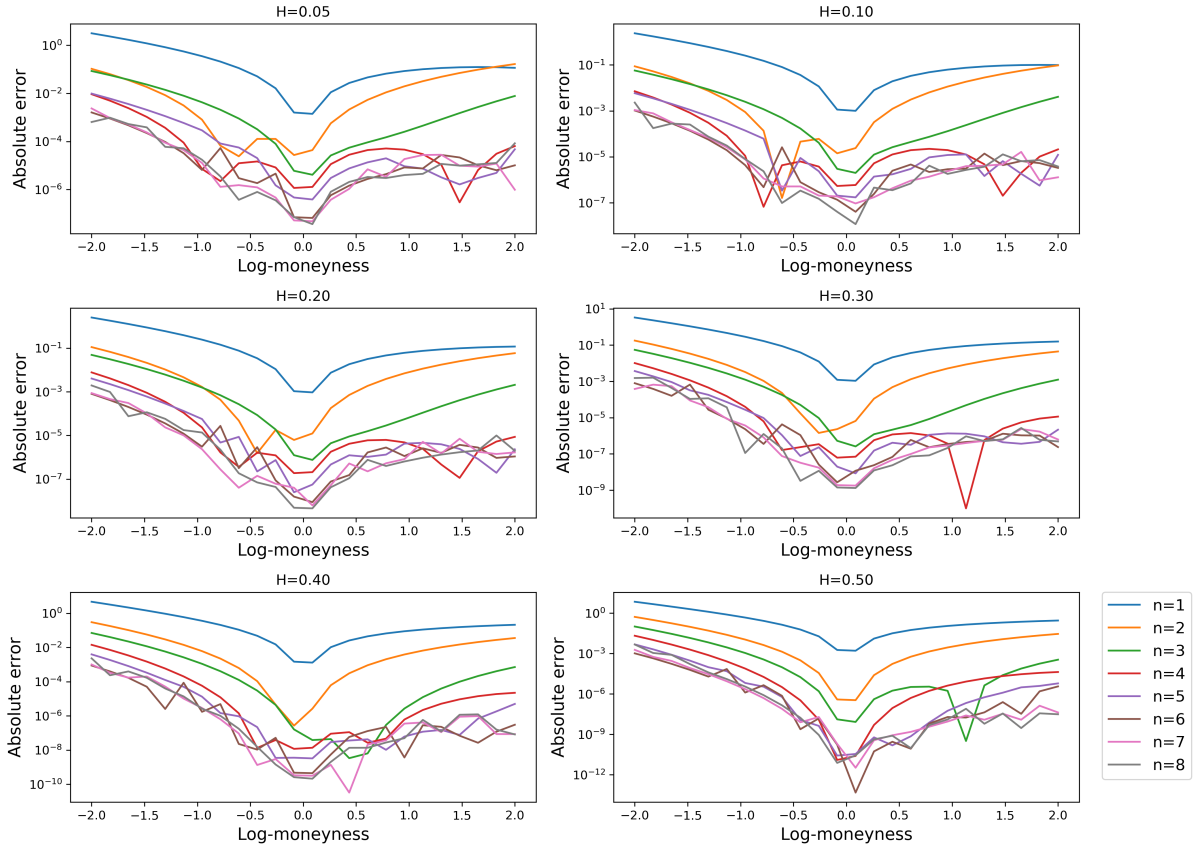


Figure E.2: Absolute difference of subsequent approximating polynomials $|\hat{\Lambda}^V_{n+1}(\mathrm{y}) - \hat{\Lambda}^V_n(\mathrm{y})|$

E.3.2 A tailor-made polynomial basis for rough volatility

We may improve the computation time of the previous approach by considering a tailor-made polynomial basis. In particular, recall the following relation

$$\int_0^s K_\alpha(u,s)u^k \mathrm{d}u = \frac{u^{\alpha+1+k}\,{}_2F_1(k+1,-\alpha,k+2)}{k+1},$$

then, for $k = -\alpha - 1$ we obtain

$$\int_0^s K_\alpha(u,s)u^{-\alpha-1} \mathrm{d}u = \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha},$$

which in turn is a constant that does not depend on the upper integral bound s.

Proposition E.3.1. *Consider the basis* $\hat{g}^{(n)}(s) = cs^{-\alpha-1} + \sum_{i=0}^n a_i s^i$, *where* $c \in \mathbb{R}$. *Then, for* $c = c^*$ *with*

$$c^* = \frac{-\alpha}{\eta\sqrt{2\alpha+1}\,{}_2F_1(-\alpha,-\alpha,1-\alpha,1)} \log\left(\frac{y}{\int_0^1 \exp\left(\eta\sqrt{2\alpha+1}\left(\sum_{i=0}^n \frac{a_i s^{\alpha+1+i}\,{}_2F_1(i+1,-\alpha,i+2,1)}{i+1}\right)\right)\mathrm{d}s}\right),$$

$\hat{g}^{(n)}(s)$ *solves* (E.6).

Proof. We have that

$$RV\left(\int_0^\cdot K_\alpha(u,\cdot)\hat{g}^{(n)}(u)\mathrm{d}u\right)(1) = \exp\left(\frac{\eta\sqrt{2\alpha+1}}{-\alpha}\,{}_2F_1\left(-\alpha,-\alpha,1-\alpha,1\right)\right) \times \int_0^1 \exp\left(\eta\sqrt{2\alpha+1}\sum_{i=0}^n \frac{a_i s^{\alpha+1+i}\,{}_2F_1\left(i+1,-\alpha,i+2,1\right)}{i+1}\right)\mathrm{d}s$$

and the proof trivially follows by solving $y = RV\left(\int_0^\cdot K_\alpha(u,\cdot)\hat{g}^{(n)}(u)\mathrm{d}u\right)(1)$. □

Remark E.3.2. Notice that Proposition E.3.1 gives a semi-closed form solution to (E.6). Then, we only need to solve

$$\inf_{(a_0,\dots,a_n)\in\mathbb{R}^{n+1}}\left\{\frac{1}{2}||\hat{g}^{(n)}||^2 : c = c^*\right\}$$

in order to recover a solution for (E.5).

Remark E.3.3. Notice that $u^{-\alpha-1} \notin L^2[0,1]$, however $u^{-\alpha-1}\mathbb{1}_{\{u>\varepsilon\}} \in L^2[0,1]$ for all $\varepsilon > 0$. Moreover,

$$\begin{aligned}\int_0^s K_\alpha(s,u)u^{-\alpha-1}\mathbb{1}_{\{u>\varepsilon\}}\mathrm{d}u &= \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha} - \frac{\varepsilon^{-\alpha}s^\alpha\ {}_2F_1(-\alpha,-\alpha,1-\alpha,\frac{\varepsilon}{t})}{-\alpha}\\ &= \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha} + \mathcal{O}(\varepsilon^{-\alpha}),\end{aligned}$$

hence for ε sufficiently small the error is bounded as long as $\alpha \neq 0$. In our applications we find that this method behaves nicely for $\alpha \in (-0.5, -0.05]$. In Figure E.3 we provide precise errors and we observe that the convergence is better for small α (which is rather surprising behaviour, as the converse is true of other approximation schemes when the volatility trajectories become more rough) as well as strikes around the money. Moreover, the truncated basis approach constitutes a 30-fold speed improvement in our numerical tests. As benchmark we consider the standard numerical algorithm introduced in (E.6), with accuracy measured by absolute error.

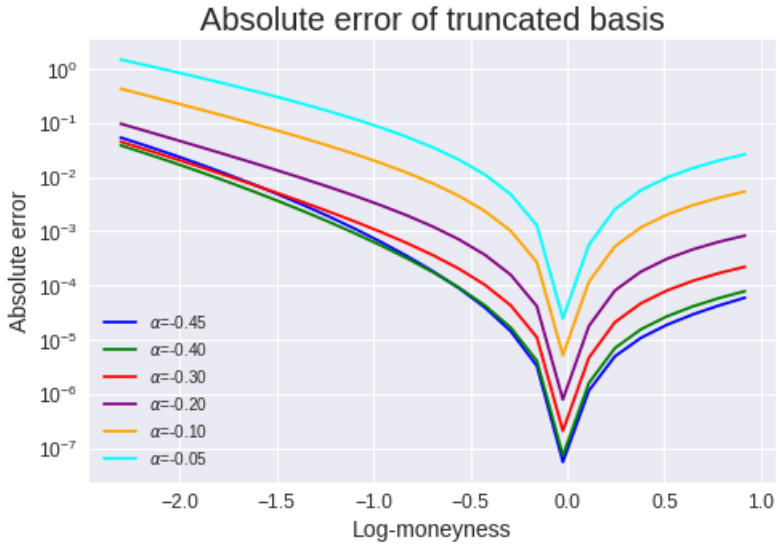


Figure E.3: Absolute error of the rate function. We consider the truncated basis approach against the standard polynomial basis with $\eta = 1.5$, $V_0 = 0.04$ and different values of α.

E.3.3 Multi-Factor case

The correlated mixed multi-factor rough Bergomi (7.7) model requires a slightly more complex setting. By Corollary 7.3.15 the rate function that we aim at is given by following multidimensional optimisation problem:

$$\hat{\Lambda}^{(v,\Sigma)}(\mathrm{y}) := \inf_{(f_1,\dots,f_n)\in L^2[0,1]}\left\{\frac{1}{2}\sum_{i=1}^n||f_i||^2 : \mathrm{y} = RV\left(\sum_{i=1}^m \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot\Sigma_i \mathfrak{f}_\cdot^{K_\alpha}\right)u\right)(1)\right\}, \tag{E.8}$$

where $\mathfrak{f}_\cdot^{K_\alpha} = \left(\int_0^\cdot K_\alpha(u,\cdot)f_1(u)\mathrm{d}u, \dots, \int_0^\cdot K_\alpha(u,\cdot)f_n(u)\mathrm{d}u\right)$. The approach to solve this problem is similar to that of (E.5). Nevertheless, in order to solve (E.8) we shall use a multi-dimensional polynomial basis

$$\left(\hat{f}_1^{(p)}(s),\dots,\hat{f}_n^{(p)}(s)\right) = \left(\sum_{i=0}^p a_i^1 s^i,\dots,\sum_{i=0}^p a_i^n s^i\right)$$

such that each $\hat{f}_i^{(p)}(s)$ for $i \in \{1, ..., n\}$ is dense as p tends to $+\infty$ in $L^2[0,1]$ by Stone-Weierstrass Theorem. Then we may equivalently solve

$$\inf_{(a_0^1,...,a_p^1,...,a_0^n,...,a_p^n)\in\mathbb{R}^{(p+1)n}} \left\{ \frac{1}{2}\sum_{i=1}^{n} ||\hat{f}_i^{(p)}||^2 : \mathrm{y} = RV\left(\sum_{i=1}^{m} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \Sigma_i \hat{\mathrm{f}}_{\cdot}^{(K_\alpha,p)}\right) u\right)(1) \right\}, \tag{E.9}$$

where $\hat{\mathrm{f}}_{\cdot}^{(K_\alpha,p)} = \left(\int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}_1^{(p)}(u)\mathrm{d}u, ..., \int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}_n^{(p)}(u)\mathrm{d}u\right)$. Then as p tends to $+\infty$, (E.9) will converge to the original problem (E.8). In order to numerically accelerate the optimisation problem in (E.9), we anchor coefficients $(a_0^1, ..., a_0^n)$ to satisfy the constraint $y = RV(\cdot)(1)$ (same way we did in the one dimensional case), that is

$$\mathfrak{a}^* := \inf_{(a_0^1,....,a_0^n)\in\mathbb{R}^n} \left\{ \left(\mathrm{y} - RV\left(\sum_{i=1}^{m} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \Sigma_i \mathrm{f}_{\cdot}^{K_\alpha}\right) u\right)(1)\right)^2 \right\}$$

where $\mathfrak{a}^* = (\mathfrak{a}_0^{1*}, ..., \mathfrak{a}_0^{n*})$ and one may use (E.7) and Gauss-Legendre quadrature to efficiently compute $RV(\cdot)(1)$. Then, the constraint will always be satisfied by construction and instead we may solve

$$\inf_{(\mathfrak{a}_0^{1*},a_1^1...,a_p^1,...,\mathfrak{a}_0^{n*},a_1^n,...,a_p^n)\in\mathbb{R}^{(p+1)n}} \left\{ \frac{1}{2}\sum_{i=1}^{n} ||\hat{f}^{(p)}||^2 \right\}. \tag{E.10}$$

E.4 Exponential Equivalence and Contraction Principle

Definition E.4.1. On a metric space $(\mathcal{Y}, d)$, two $\mathcal{Y}$-valued sequences $(X^\varepsilon)_{\varepsilon>0}$ and $(\widetilde{X}^\varepsilon)_{\varepsilon>0}$ are called exponentially equivalent (with speed h_ε) if there exist probability spaces $(\Omega, \mathcal{B}_\varepsilon, \mathbb{P}_\varepsilon)_{\varepsilon>0}$ such that for any $\varepsilon > 0$, $\mathbb{P}^\varepsilon$ is the joint law of $(\widetilde{X}^\varepsilon, X^\varepsilon)$, $\varepsilon > 0$ and, for each $\delta > 0$, the set $\left\{\omega : (\widetilde{X}^\varepsilon, X^\varepsilon) \in \Gamma_\delta\right\}$ is $\mathcal{B}_\varepsilon$-measurable, and

$$\limsup_{\varepsilon\downarrow 0} h_\varepsilon \log \mathbb{P}^\varepsilon(\Gamma_\delta) = -\infty,$$

where $\Gamma_\delta := \{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\} \subset \mathcal{Y} \times \mathcal{Y}$.

Theorem E.4.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be topological spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \to [0, \infty]$. For each $y \in \mathcal{Y}$, define $I'(y) := \inf\{I(x) : x \in \mathcal{X}, y = f(x)\}$. Then, if I controls the LDP associated with a family of probability measures $\{\mu_\varepsilon\}$ on $\mathcal{X}$, the I' controls the LDP associated with the family of probability measures $\{\mu_\varepsilon \circ f^{-1}\}$ on $\mathcal{Y}$ and I' is a good rate function on $\mathcal{Y}$.*

E.5 Proof of $V_0\phi(d_1(x)) = x\phi(d_2(x))$

In order to prove $V_0\phi(d_1(x)) = x\phi(d_2(x))$, we will prove the following equivalent result

$$(d_1(x))^2 - (d_2(x))^2 = 2\log\left(\frac{V_0}{x}\right).$$

Proof. Recall that $\phi(\cdot)$ is the standard Gaussian probability density function. Using that for $x \geq 0$, $d_1(x) =$

$\frac{\log(V_0)-\log(x)}{\hat{\sigma}(x,T)\sqrt{T}} + \frac{1}{2}\hat{\sigma}(x,T)\sqrt{T}$ and $d_2(x) = d_1(x) - \hat{\sigma}(x,T)\sqrt{T}$, we obtain

$$\begin{aligned}
(d_1(x))^2 - (d_2(x))^2 &= (d_1(x))^2 - \left(d_1(x) - \hat{\sigma}(x,T)\sqrt{T}\right)^2, \\
&= 2d_1(x)\hat{\sigma}(x,T)\sqrt{T} - T\hat{\sigma}^2(x,T), \\
&= 2\left[\log(V_0) - \log(x) + \frac{1}{2}\hat{\sigma}^2(x,T)T\right] - T\hat{\sigma}^2(x,T), \\
&= 2\log\left(\frac{V_0}{x}\right).
\end{aligned}$$

□

F

F.1 Proof of Proposition 8.3.6

Proof. To simplify the notations in the proof, we introduce the quantities $H_\pm := H \pm \frac{1}{2}$ and $H^\alpha_\pm := H + \alpha \pm \frac{1}{2}$. By independence of the driving Brownian motions we have, for any $u \leq t$,

$$\begin{aligned}\mathbb{E}[V_t|\mathcal{F}_u] = \xi_0(t) \exp\left\{\nu\left(\left(\mathrm{k}^{H_-} Z^{\mathbb{Q}}\right)(u,t) + \left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(u,t)\right)\right\} \mathbb{E}\left[\exp\left(\nu\left(\left(\mathrm{k}^{H_-} Z^{\mathbb{Q}}\right)(t) - \left(\mathrm{k}^{H_-} Z^{\mathbb{Q}}\right)(u,t)\right)\right)\right] \\ \times \mathbb{E}\left[\exp\left(\nu\left(\left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(t) - \left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(u,t)\right)\right)\right],\end{aligned}$$

where the first expected value is the MGF of a Gaussian random variable, hence

$$\mathbb{E}\left[\exp\left(\nu\left(\left(\mathrm{k}^{H_-} Z^{\mathbb{Q}}\right)(t) - \left(\mathrm{k}^{H_-} Z^{\mathbb{Q}}\right)(u,t)\right)\right)\right] = \exp\left(\frac{1}{2}\nu^2 \int_u^t k(t,s)\mathrm{d}s\right).$$

We are now interested in computing the second expectation

$$\mathbb{E}\left[\exp\left\{\nu\left(\left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(t) - \left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(u,t)\right)\right\}\right],$$

where

$$\left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(t) - \left(\mathcal{K}^{H^\alpha_+} Y^{\mathbb{Q}}\right)(u,t) = \int_u^t k(t,s) Y^{\mathbb{Q}}_s \mathrm{d}s.$$

This is, in spirit, a similar problem to computing a Bond price in the CIR model. To do so, we define

$$B(t,T) = \mathbb{E}\left[\exp\left(\nu \int_t^T k(T,s) Y^{\mathbb{Q}}_s \mathrm{d}s\right) \middle| \mathcal{F}_t\right],$$

where $t \leq T$. We note that $B(\cdot, T)$ is a semimartingale as T is fixed, therefore applying Feynman-Kac's formula we obtain

$$\left(\nu r k(T,t) + \partial_t + \kappa(\theta - r)\partial_r + \frac{\sigma^2}{2} r \partial_{rr}\right) \mathcal{B}(r,t,T) = 0.$$

where $\mathcal{B}(r,t,T)$ is such that $\mathcal{B}(Y^{\mathbb{Q}}_t, t, T)$ is a solution to the MGF. With an ansatz of the type $\mathcal{B}(r,t,T) = \exp\left[-rC(t,T) - A(t,T)\right]$, we have, at the point (r,t,T),

$$\partial_t \mathcal{B} = -(r\partial_t C(t,T) + \partial_t A(t,T))\mathcal{B}, \qquad \partial_r \mathcal{B} = -C(t,T)\mathcal{B}, \qquad \partial_{rr}\mathcal{B} = C(t,T)^2 \mathcal{B},$$

and the PDE becomes, with $r = Y_t^{\mathbb{Q}}$,

$$\left[\nu Y_t^{\mathbb{Q}} k(T,t) - \left(Y^{\mathbb{Q}}\partial_t C + \partial_t A + \kappa\left(\theta - Y_t^{\mathbb{Q}}\right)C\right) + \frac{\sigma^2}{2}C^2 Y_t^{\mathbb{Q}}\right]\mathcal{B} = 0,$$

which further simplifies to

$$\left(\nu k(T,t) - \partial_t C - \kappa C + \frac{\sigma^2}{2}C^2\right) Y_t^{\mathbb{Q}}\mathcal{B} - (\kappa\theta C + \partial_t A)\mathcal{B} = 0.$$

The second term cancels when $A(t,T) = -\kappa\theta\int_t^T C(u,T)\mathrm{d}t$, and it suffices to solve the Riccati equation

$$\nu k(T,t) - \partial_t C(t,T) - \kappa C(t,T) + \frac{\sigma^2}{2}C^2(t,T) = 0,$$

with boundary conditions $A(T,T) = C(T,T) = 0$. □

References

[Aas93] Knut K. Aase. "A Jump/Diffusion Consumption-Based Capital Asset Pricing Model and the Equity Premium Puzzle". In: *Mathematical Finance* 3.2 (1993), pp. 65–84. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9965.1993.tb00079.x.

[ACS99] Marco Avellaneda, Andrea Carelli, and Fabio Stella. "Following the Bayes path to option pricing". In: *Journal of Computational Intelligence in Finance* (1999).

[AD03] Jean Pierre Aubin and Halim Doss. "Characterization of Stochastic Viability of Any Nonsmooth Set Involving Its Generalized Contingent Curvature". In: *Stochastic Analysis and Applications* 21.5 (2003), pp. 955–981.

[AD54] Kenneth J. Arrow and Gerard Debreu. "Existence of an Equilibrium for a Competitive Economy". In: *Econometrica* 22.3 (1954), p. 265.

[ADS14] Erdinç Akyildirim, Yan Dolinsky, and H. Mete Soner. "Approximating stochastic volatility by recombinant trees". In: *Annals of Applied Probability* 24.5 (2014), pp. 2176–2205.

[AE08] Elisa Alòs and Christian-Oliver Ewald. *Malliavin Differentiability of the Heston Volatility and Applications to Option Pricing*. 2008.

[AGM18] Elisa Alòs, David García-Lorite, and Aitor Muguruza. "On smile properties of volatility derivatives and exotic products: understanding the VIX skew". In: (2018). arXiv: 1808.03610.

[AL17] Elisa Alòs and Jorge A. Léon. "On the curvature of the smile in stochastic volatility models". In: *SIAM Journal on Financial Mathematics* 8.1 (2017), pp. 373–399.

[Alò06] Elisa Alòs. "A generalization of the Hull and White formula with applications to option pricing approximation". In: *Finance and Stochastics* 10.3 (2006), pp. 353–365.

[ALV07] Elisa Alòs, Jorge A. León, and Josep Vives. "On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility". In: *Finance and Stochastics* 11.4 (2007), pp. 571–589.

[AP19] M. Avellaneda and A. Papanicolaou. "Statistics of vix futures and applications to trading volatility exchange-traded products". In: *International Journal of Theoretical and Applied Finance* 22.1 (2019).

[AS19] Elisa Alòs and Kenichiro Shiraya. "Estimating the Hurst parameter from short term volatility swaps: a Malliavin calculus approach". In: *Finance and Stochastics* 23.2 (2019), pp. 423–447.

[ASR88] Milton Abramowitz, Irene A. Stegun, and Robert H. Romer. "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables". In: *American Journal of Physics* 56.10 (1988), pp. 958–958.

[Atk89] Kendall E. Atkinson. *An introduction to numerical analysis*. 2nd Editio. Wiley & Sons, 1989, p. 693.

[Bac+13] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. "Modelling microstructure noise with mutually exciting point processes". In: *Quantitative Finance* 13.1 (2013), pp. 65–77. arXiv: 1101.3422.

[Bal05] P Balland. "Stoch-vol model with interest rate volatility". In: *ICBI conference* (2005).

[Bar+10] X. Bardina, I. Nourdin, C. Rovira, and S. Tindel. "Weak approximation of a fractional SDE". In: *Stochastic Processes and their Applications* 120.1 (2010), pp. 39–65.

[Bar94] Andrew R. Barron. "Approximation and estimation bounds for artificial neural networks". In: *Machine Learning* 14.1 (1994), pp. 115–133.

[Bay+19a] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper. "Short-time near-the-money skew in rough fractional volatility models". In: *Quantitative Finance* 19.5 (2019), pp. 779–798. arXiv: 1703.05132.

[Bay+19b] Christian Bayer, Peter K. Friz, Paul Gassiat, Jorg Martin, and Benjamin Stemper. "A regularity structure for rough volatility". In: *Mathematical Finance* (2019).

[Bay+19c] Christian Bayer, Blanka Horvath, Aitor Muguruza, Benjamin Stemper, and Mehdi Tomas. "On deep calibration of (rough) stochastic volatility models". In: (2019). arXiv: 1908.08806.

[BB14] Jan Baldeaux and Alexander Badran. "Consistent Modelling of VIX and Equity Derivatives Using a 3/2 plus Jumps Model". In: *Applied Mathematical Finance* 21.4 (2014), pp. 299–312. arXiv: 1203.5903.

[Ber05] Lorenzo Bergomi. "Smile dynamics II". In: *Risk* October (2005), pp. 67–73.

[Ber08] Lorenzo Bergomi. "Smile dynamics III". In: *Risk* October (2008), pp. 90–96.

[Ber09] Lorenzo Bergomi. "Smile dynamics IV". In: *Risk* December (2009), pp. 94–100.

[Ber16] Lorenzo Bergomi. *Stochastic volatility modeling*. CRC press, 2016.

[BFG16] Christian Bayer, Peter Friz, and Jim Gatheral. "Pricing under rough volatility". In: *Quantitative Finance* 16.6 (2016), pp. 887–904.

[Bia+08] Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic calculus for fractional Brownian motion and applications*. Springer Science & Business Media, 2008.

[Bil68] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 1968.

[BJ02] Martino Bardi and Robert Jensen. "A geometric characterization of viable sets for controlled degenerate diffusions". In: *Set-Valued Analysis* 10.2-3 (2002), pp. 129–141.

[BL78] Douglas T. Breeden and Robert H. Litzenberger. *Prices of State-Contingent Claims Implicit in Option Prices*. 1978.

[BL91] Fischer Black and Robert Litterman. "Asset Allocation: Combining Investor Views with Market Equilibrium". In: *The Journal of Fixed Income* 1.2 (1991), pp. 7–18.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. eng. Oxford: Oxford University Press, 2013, p. 496.

[BLP16] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. "Decoupling the short- and long-term behavior of stochastic volatility". In: (2016). arXiv: 1610.00332.

[BLP17] Mikkel Bennedsen, Asger Lunde, and Mikko S. Pakkanen. "Hybrid scheme for Brownian semistationary processes". In: *Finance and Stochastics* 21.4 (2017), pp. 931–965. arXiv: 1507.03004.

[BNP19] Andrea Barletta, Elisa Nicolato, and Stefano Pagliarani. "The short-time behavior of VIX-implied volatilities in a multifactor stochastic volatility framework". In: *Mathematical Finance* 29.3 (2019), pp. 928–966.

[BP12] Christian Bender and Peter Parczewski. "On the connection between discrete and continuous Wick calculus with an application to the fractional Black-Scholes model". In: *Stochastic Processes, Finance and Control*. 2012, pp. 3–40.

[BP91] Kerry Back and Stanley R. Pliska. "On the fundamental theorem of asset pricing with an infinite state space". In: *Journal of Mathematical Economics* 20.1 (1991), pp. 1–18.

[BR94] Clifford A. Ball and Antonio Roma. "Stochastic Volatility Option Pricing". In: *The Journal of Financial and Quantitative Analysis* 29.4 (1994), p. 589.

[BS07] Ole E Barndorff-Nielsen and Jürgen Schmiegel. "Ambit Processes; with Applications to Turbulence and Tumour Growth". In: *Stochastic Analysis and Applications*. Ed. by Fred Espen Benth, Giulia Di Nunno, Tom Lindstrøm, Bernt Øksendal, and Tusheng Zhang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 93–124.

[BS15] Mathias Beiglböck and Pietro Siorpaes. "Pathwise versions of the Burkholder-Davis-Gundy inequality". In: *Bernoulli* 21.1 (2015), pp. 360–373.

[BS18] Christian Bayer and Benjamin Stemper. "Deep calibration of rough stochastic volatility models". In: (2018). arXiv: 1810.03399.

[BS73] Fischer Black and Myron Scholes. *The Pricing of Options and Corporate Liabilities*. 1973.

[Bue+19] H. Buehler, L. Gonon, J. Teichmann, and B. Wood. "Deep hedging". In: *Quantitative Finance* 19.8 (2019), pp. 1271–1291. arXiv: 1802.03042.

[BV13] Ole E. Barndorff-Nielsen and Almut E. D. Veraart. "Stochastic Volatility of Volatility and Variance Risk Premia". In: *Journal of Financial Econometrics* 11.1 (2013), pp. 1–46.

[Cam+18] John Y. Campbell, Stefano Giglio, Christopher Polk, and Robert Turley. "An intertemporal CAPM with stochastic volatility". In: *Journal of Financial Economics* 128.2 (2018), pp. 207–233.

[Car+05] Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. "Pricing options on realized variance". In: *Finance and Stochastics* 9.4 (2005), pp. 453–475.

[CD17] Robert Culkin and Sanjiv R Das. "Machine learning in finance: The case of deep learning for option pricing". In: *Journal of Investment Management* 15.4 (2017), pp. 92–100.

[Čen56] N. N. Čentsov. "Weak Convergence of Stochastic Processes Whose Trajectories Have No Discontinuities of the Second Kind and the "Heuristic" Approach to the Kolmogorov-Smirnov Tests". In: *Theory of Probability & Its Applications* 1.1 (1956), pp. 140–144.

[Che07] Mikhail Chernov. *On the Role of Risk Premia in Volatility Forecasting*. 2007.

[Che08] Alexander Cherny. "Brownian moving averages have conditional full support". In: *Annals of Applied Probability* 18.5 (2008), pp. 1825–1830.

[CIR85] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. "A Theory of the Term Structure of Interest Rates". In: *Econometrica* 53.2 (1985), p. 385.

[CL09] Peter Carr and Roger Lee. "Volatility Derivatives". In: *Annual Review of Financial Economics* 1.1 (2009), pp. 319–339.

[Cla87] Dean S. Clark. "Short proof of a discrete gronwall inequality". In: *Discrete Applied Mathematics* 16.3 (1987), pp. 279–281.

[CM01] Peter Carr and Dilip Madan. "Towards a theory of volatility trading". In: *Option Pricing, Interest Rates and Risk Management, Handbooks in Mathematical Finance* (2001), pp. 458–476.

[CM14] Peter Carr and Dilip B. Madan. "Joint modeling of VIX and SPX options at a single and common maturity with risk management applications". In: *IIE Transactions (Institute of Industrial Engineers)* 46.11 (2014), pp. 1125–1131.

[CMR19] Andrei Cozma, Matthieu Mariapragassam, and Christoph Reisinger. "Calibration of a hybrid local-stochastic volatility stochastic rates model with a control variate particle method". In: *SIAM Journal on Financial Mathematics* 10.1 (2019), pp. 181–213. arXiv: 1701.06001.

[Con01] R Cont. "Empirical properties of asset returns: stylized facts and statistical issues". In: *Quantitative Finance* 1.2 (2001), pp. 223–236.

[Con10] Rama Cont. *Model Calibration*. 2010.

[CR96] F. Comte and E. Renault. "Long memory continuous time models". In: *Journal of Econometrics* 73.1 (1996), pp. 101–149.

[CT65] James W. Cooley and John W. Tukey. "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of Computation* 19.90 (1965), pp. 297–297.

[CV12] Alexandra Chronopoulou and Frederi G. Viens. "Estimation and pricing under long-memory stochastic volatility". In: *Annals of Finance* 8.2-3 (2012), pp. 379–403.

[CW08] Peter Carr and Liuren Wu. "Variance Risk Premiums". In: *The Review of Financial Studies* 22.3 (2008), pp. 1311–1341.

[De +18] Jan De Spiegeleer, Dilip B. Madan, Sofie Reyners, and Wim Schoutens. "Machine learning for quantitative finance: fast derivative pricing, hedging and fitting". In: *Quantitative Finance* 18.10 (2018), pp. 1635–1643.

[DeM18] Stefano DeMarco. "VIX derivatives in rough forward variance models". In: *Presentation, Bachelier Congress, Dublin* (2018).

[Deu+14] J. D. Deuschel, P. K. Friz, A. Jacquier, and S. Violante. "Marginal density expansions for diffusions and stochastic volatility I: Theoretical foundations". In: *Communications on Pure and Applied Mathematics* 67.1 (2014), pp. 40–82. arXiv: 1111.2462.

[DF07] Giuseppe Da Prato and Hélène Frankowska. "Stochastic viability of convex sets". In: *Journal of Mathematical Analysis and Applications* 333.1 SPEC. ISS. (2007), pp. 151–163.

[Dol70] C. Doléans-Dade. "Quelques applications de la formule de changement de variables pour les semimartingales". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 16.3 (1970), pp. 181–194.

[Don51] Monroe David Donsker. *An invariance principle for certain probability limit theorems*. Memoirs of the AMS, 1951.

[Drá75] Pavel Drábek. "Continuity of Nemyckij's operator in Hölder spaces". In: *Commentationes Mathematicae Universitatis Carolinae* 016.1 (1975), pp. 37–57.

[DRF18] Georgi Dimitroff, Dirk Röder, and Christian P Fries. "Volatility Model Calibration With Convolutional Neural Networks". In: *Econometrics: Econometric & Statistical Methods - Special Topics eJournal* (2018).

[DS89] Jean-Dominique Deuschel and Daniel W. Stroock. *Large deviations*. Academic Press, 1989, p. 307.

[DS94] Freddy Delbaen and Walter Schachermayer. "A general version of the fundamental theorem of asset pricing". In: *Mathematische Annalen* 300.1 (1994), pp. 463–520.

[Dud99] R. M. Dudley. *Uniform Central Limit Theorems*. Cambridge University Press, 1999.

[Duf04] Daniel Dufresne. *The Log-Normal Approximation in Financial and Other Computations*. 2004.

[Dup94] Bruno Dupire. "Pricing with a smile". In: *Risk* 7 (1994), pp. 18–20.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. 2nd editio. Springer-Verlag Berlin Heidelberg, 2010, p. 396.

[El 81] N. El Karoui. *Les Aspects Probabilistes Du Controle Stochastique*. Springer, Berlin, Heidelberg, 1981, pp. 73–238.

[Eng01] Robert Engle. "GARCH 101: The use of ARCH/GARCH models in applied econometrics". In: *Journal of Economic Perspectives* 15.4 (2001), pp. 157–168.

[Eng82] Robert F. Engle. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: *Econometrica* 50.4 (1982), p. 987.

[ER18] Omar El Euch and Mathieu Rosenbaum. "Perfect hedging in rough heston models". In: *Annals of Applied Probability* 28.6 (2018), pp. 3813–3856. arXiv: 1703.05049.

[ER19] Omar El Euch and Mathieu Rosenbaum. "The characteristic function of rough Heston models". In: *Mathematical Finance* 29.1 (2019), pp. 3–38.

[ES15] Ronen Eldan and Ohad Shamir. "The Power of Depth for Feedforward Neural Networks". In: *Journal of Machine Learning Research* 49.June (2015), pp. 907–940. arXiv: 1512.03965.

[FG18] Ryan Ferguson and Andrew Green. "Deeply Learning Derivatives". In: (2018). arXiv: 1809.02233.

[FGP18] Peter K. Friz, Paul Gassiat, and Paolo Pigato. "Precise asymptotics: robust stochastic volatility models". In: (2018). arXiv: 1811.00267.

[FH20] Peter K. Friz and Martin Hairer. *A Course on Rough Paths*. Springer, 2020, p. 322.

[FLC14] Ka Wai Terence Fung, Chi Keung Marco Lau, and Kwok Ho Chan. "The conditional equity premium, cross-sectional returns and stochastic volatility". In: *Economic Modelling* 38 (2014), pp. 316–327.

[Fou+04] Jean Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Solna. "Maturity cycles in implied volatility". In: *Finance and Stochastics* 8.4 (2004), pp. 451–477.

[FS18] J. P. Fouque and Y. F. Saporito. "Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options". In: *Quantitative Finance* 18.6 (2018), pp. 1003–1016. arXiv: 1706.00873.

[FTW19] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. "Is Volatility Rough ?" In: (2019). arXiv: 1905.04852.

[Fuk11a] Masaaki Fukasawa. "Asymptotic analysis for stochastic volatility: Edgeworth expansion". In: *Electronic Journal of Probability* 16 (2011), pp. 764–791.

[Fuk11b] Masaaki Fukasawa. "Asymptotic analysis for stochastic volatility: Martingale expansion". In: *Finance and Stochastics* 15.4 (2011), pp. 635–654.

[Fuk12] Masaaki Fukasawa. "THE NORMALIZING TRANSFORMATION OF THE IMPLIED VOLATILITY SMILE". In: *Mathematical Finance* 22.4 (2012), pp. 753–762. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9965.2011.00483.x.

[FV10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge University Press, 2010.

[FZ17] Martin Forde and Hongzhong Zhang. "Asymptotics for rough stochastic volatility models". In: *SIAM Journal on Financial Mathematics* 8.1 (2017), pp. 114–145. arXiv: 1610.08878.

[Gas19] Paul Gassiat. "On the martingale property in the rough bergomi model". In: *Electronic Communications in Probability* 24 (2019).

[Gat+12] Jim Gatheral, Elton P. Hsu, Peter Laurence, Cheng Ouyang, and Tai Ho Wang. "Asymptotics of implied volatility in local volatility models". In: *Mathematical Finance* 22.4 (2012), pp. 591–620.

[Gat06] Jim Gatheral. *The Volatility Surface: A Practitioner's Guide*. John Wiley & Sons, 2006, p. 210.

[Gat08] Jim Gatheral. "Consistent modelling of SPX and VIX options". In: *Presentation, Bachelier Congress, London* (2008).

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[GH12] Julien Guyon and Pierre Henry-Labordère. "Being particular about calibration". In: *Risk* (2012), pp. 92–96.

[GIP17] Stéphane Goutte, Amine Ismail, and Huyên Pham. "Regime-switching stochastic volatility model: estimation and calibration to VIX options". In: *Applied Mathematical Finance* 24.1 (2017), pp. 38–75.

[Gir60] I. V. Girsanov. "On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures". In: *Theory of Probability & Its Applications* 5.3 (1960), pp. 285–301.

[GJ14] Jim Gatheral and Antoine Jacquier. "Arbitrage-free SVI volatility surfaces". In: *Quantitative Finance* 14.1 (2014), pp. 59–71. arXiv: 1204.0646.

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. "Volatility is rough". In: *Quantitative Finance* 18.6 (2018), pp. 933–949. arXiv: 1410.3394.

[GKS19] Kathrin Glau, Daniel Kressner, and Francesco Statti. "Low-rank tensor approximation for Chebyshev interpolation in parametric option pricing". In: (2019). arXiv: 1902.04367.

[GL14] Kun Gao and Roger Lee. "Asymptotics of implied volatility to arbitrary order". In: *Finance and Stochastics* 18.2 (2014), pp. 349–392.

[GL96] Gene H Golub and Charles F Van Loan. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*. Vol. 208-209. 1996, p. 728.

[Gue+18] Hamza Guennoun, Antoine Jacquier, Patrick Roome, and Fangwei Shi. "Asymptotic behavior of the fractional heston model". In: *SIAM Journal on Financial Mathematics* 9.3 (2018), pp. 1017–1045. arXiv: 1411.7653.

[Gul18] Archil Gulisashvili. "Large deviation principle for volterra type fractional stochastic volatility models". In: *SIAM Journal on Financial Mathematics* 9.3 (2018), pp. 1102–1136.

[Guo98] Dajiang Guo. "The Risk Premium of Volatility Implicit in Currency Options". In: *Journal of Business and Economic Statistics* 16.4 (1998), pp. 498–507.

[Guy14] Julien Guyon. "Local correlation families". In: *Risk* (2014), pp. 52–58.

[Guy20a] Julien Guyon. “Inversion of convex ordering in the VIX market”. In: *Quantitative Finance* (2020), pp. 1–27.

[Guy20b] Julien Guyon. “The joint S&P 500/Vix smile calibration puzzle solved”. In: *Risk* April (2020).

[GV16] Andrea Granelli and Almut E.D. Veraart. “Modeling the variance risk premium of equity indices: The role of dependence and contagion”. In: *SIAM Journal on Financial Mathematics* 7.1 (2016), pp. 382–417.

[Hag+02] Patrick S Hagan, Deep Kumar, Andrew S Lesniewski, and Diana E Woodward. “Managing smile risk”. In: *The Best of Wilmott* 1 (2002), pp. 249–296.

[Ham00] D Hamadouche. “Invariance principles in Hölder spaces.” In: *Portugaliae Mathematica* 57.2 (2000), pp. 127–151.

[Haw71] Alan G. Hawkes. “Spectra of Some Self-Exciting and Mutually Exciting Point Processes”. In: *Biometrika* 58.1 (1971), p. 83.

[HB88] Wolfgang Hardle and Adrian W. Bowman. “Bootstrapping in Nonparametric Regression: Local Adaptive Smoothing and Confidence Bands”. In: *Journal of the American Statistical Association* 83.401 (1988), p. 102.

[Her17] Andres Hernandez. “Model calibration with neural networks”. In: *Risk* (2017).

[Hes93] Steven L Heston. “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”. In: *Review of Financial Studies* 6.2 (1993), pp. 327–43.

[HJL19] Blanka Horvath, Antoine Jacquier, and Chloé Lacombe. “Asymptotic behaviour of randomised fractional volatility models”. In: *Journal of Applied Probability* 56.2 (2019), pp. 496–523.

[HJM17] Blanka Horvath, Antoine Jacquier, and Aitor Muguruza. “Functional central limit theorems for rough volatility”. In: (2017). arXiv: 1711.03078.

[HJT20] Blanka Horvath, Antoine Jacquier, and Peter Tankov. “Volatility Options in Rough Volatility Models”. In: *SIAM Journal on Financial Mathematics* 11.2 (2020), pp. 437–469.

[HL28] G H Hardy and J E Littlewood. “Some properties of fractional integrals. I.” In: *Mathematische Zeitschrift* 27.1 (1928), pp. 565–606.

[HLP94] James M. Hutchinson, Andrew W. Lo, and Tomaso Poggio. “A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks”. In: *The Journal of Finance* 49.3 (1994), pp. 851–889.

[HM19] Sebas Hendriks and Claude Martini. “The extended SSVI volatility surface”. In: *Journal of Computational Finance* 22.5 (2019), pp. 25–39.

[HMT21] Blanka Horvath, Aitor Muguruza, and Mehdi Tomas. “Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models”. In: *Quantitative Finance* 21.1 (2021), pp. 11–27. eprint: https://doi.org/10.1080/14697688.2020.1817974.

[Hor91] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural Networks* 4.2 (1991), pp. 251–257.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5 (1989), pp. 359–366.

[HSW90] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks”. In: *Neural Networks* 3.5 (1990), pp. 551–560.

[HW87] JOHN HULL and ALAN WHITE. “The Pricing of Options on Assets with Stochastic Volatilities”. In: *The Journal of Finance* 42.2 (1987), pp. 281–300.

[HW90] John Hull and Alan White. “Pricing Interest-Rate-Derivative Securities”. In: *The Review of Financial Studies* 3.4 (1990), pp. 573–592.

[IS15] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML’15. JMLR.org, 2015, pp. 448–456.

[Itk15] Andrey Itkin. “To sigmoid-based functional description of the volatility smile”. In: *North American Journal of Economics and Finance* 31 (2015), pp. 264–291. arXiv: 1407.0256.

[JL19] H. Jang and J. Lee. “Machine learning versus econometric jump models in predictability and domain adaptability of index options”. In: *Physica A: Statistical Mechanics and its Applications* 513 (2019), pp. 74–86.

[JLP19] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. “Affine Volterra processes”. In: *Annals of Applied Probability* 29.5 (2019), pp. 3155–3200. arXiv: 1708.08796.

[JMM18] Antoine Jacquier, Claude Martini, and Aitor Muguruza. “On VIX futures in the rough Bergomi model”. In: *Quantitative Finance* 18.1 (2018), pp. 45–61.

[JMP89] A Jakubowski, J Mémin, and G Pages. “Convergence en loi des suites d’intégrales stochastiques sur l’espace *mathbbD* 1 de Skorokhod”. In: *Probability Theory and Related Fields* 81.1 (1989), pp. 111–137.

[JPS18] Antoine Jacquier, Mikko S. Pakkanen, and Henry Stone. “Pathwise large deviations for the rough Bergomi model”. In: *Journal of Applied Probability* 55.4 (2018), pp. 1078–1092. arXiv: 1706.05291.

[JR16] Thibault Jaisson and Mathieu Rosenbaum. “Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes”. In: *Annals of Applied Probability* 26.5 (2016), pp. 2860–2882. arXiv: 1504.03100.

[JZ20] Benjamin Jourdain and Alexandre Zhou. “Existence of a calibrated regime switching local volatility model”. In: *Mathematical Finance* 30.2 (2020), pp. 501–546.

[Kar88] Ioannis Karatzas. “On the pricing of American options”. In: *Applied Mathematics & Optimization* 17.1 (1988), pp. 37–60.

[Kol40] Andrei N Kolmogorov. “Wienersche spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady)”. In: *Acad. Sci. URSS (NS)* 26 (1940), pp. 115–118.

[Kon18] Alexei Kondratyev. “Curve dynamics with artificial neural networks”. In: *Risk* May (2018).

[KP91] Thomas G. Kurtz and Philip Protter. “Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations”. In: *Annals of Probability* 19.3 (1991), pp. 1035–1070.

[Kra88] Donald H Kraft. “A software package for sequential quadratic programming”. In: 1988.

[Kry96] Nikola\u\i Krylov. *Lectures on elliptic and parabolic equations in Holder spaces*. 1996.

[KS15] Thomas Kokholm and Martin Stisen. “Joint pricing of VIX and SPX options with stochastic volatility and jump models”. In: *Journal of Risk Finance* 16.1 (2015), pp. 27–48.

[KS97] Ioannis Karatzas and Steven E Shreve. “Brownian Motion and Stochastic Calculus”. In: Springer, 1997, pp. 47–127.

[Kus74] Harold J. Kushner. “On the Weak Convergence of Interpolated Markov Chains to a Diffusion”. In: *Annals of Probability* 2.1 (1974), pp. 40–50.

[Lam62a] John Lamperti. “On Convergence of Stochastic Processes”. In: *Transactions of the American Mathematical Society* 104.3 (1962), p. 430.

[Lam62b] John Lamperti. “Semi-Stable Stochastic Processes”. In: *Transactions of the American Mathematical Society* 104.1 (1962), p. 62.

[Lee05] Roger W Lee. “Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation BT - Recent Advances in Applied Probability”. In: ed. by Ricardo Baeza-Yates, Joseph Glaz, Henryk Gzyl, Jürgen Hüsler, and José Luis Palacios. Boston, MA: Springer US, 2005, pp. 241–268.

[Lev44] Kenneth Levenberg. *A method for the solution of certain non-linear problems in least squares*. 1944.

[Lip02] Alexander Lipton. “The Vol Smile Problem”. In: *Risk* February (2002).

[Liv+18] Giulia Livieri, Saad Mouti, Andrea Pallavicini, and Mathieu Rosenbaum. "Rough volatility: Evidence from option prices". In: *IISE Transactions* 50.9 (2018), pp. 767–776. arXiv: 1702.02777.

[LMS21] Chloe Lacombe, Aitor Muguruza, and Henry Stone. "Asymptotics for volatility derivatives in multi-factor rough volatility models". In: *Mathematics and Financial Economics* 15.3 (2021), pp. 545–577.

[LS01] Francis A Longstaff and Eduardo S Schwartz. "Valuing American Options by Simulation: A Simple Least-Squares Approach". In: *The Review of Financial Studies* 14.1 (2001), pp. 113–147.

[Lyo98] Terry J Lyons. "Differential equations driven by rough signals." In: *Revista Matemática Iberoamericana* 14.2 (1998), pp. 215–310.

[Mar54] Andrei Andreevich Markov. "The theory of algorithms". In: *Trudy Matematicheskogo Instituta Imeni VA Steklova* 42 (1954), pp. 3–375.

[Mar63] Donald W. Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". In: *Journal of the Society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.

[McG18] William A McGhee. "An Artificial Neural Network Representation of the SABR Stochastic Volatility Model". In: *SSRN Electronic Journal* (2018).

[MFE15] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[Mha93] H. N. Mhaskar. "Approximation properties of a multilayered feedforward artificial neural network". In: *Advances in Computational Mathematics* 1.1 (1993), pp. 61–80.

[MN11] Johannes Muhle-Karbe and Marcel Nutz. "Small-time asymptotics of option prices and first absolute moments". In: *Journal of Applied Probability* 48.4 (2011), pp. 1003–1020.

[MOP20] Stepan Mazur, Dmitry Otryakhin, and Mark Podolskij. "Estimation of the linear fractional stable motion". In: *Bernoulli* 26.1 (2020), pp. 226–252. arXiv: 1802.06373.

[MP18a] Ryan McCrickerd and Mikko S. Pakkanen. "Turbocharging Monte Carlo pricing for the rough Bergomi model". In: *Quantitative Finance* 18.11 (2018), pp. 1877–1886.

[MP18b] Maxime Morariu-Patrichi and Mikko S. Pakkanen. "Hybrid Marked Point Processes: Characterization, Existence and Uniqueness". In: *Market Microstructure and Liquidity* 04.03n04 (2018), p. 1950007. arXiv: 1707.06970.

[MP18c] Maxime Morariu-Patrichi and Mikko S. Pakkanen. "State-dependent Hawkes processes and their application to limit order book modelling". In: (2018). arXiv: 1809.08060.

[MP98] Sabrina Mulinacci and Maurizio Pratelli. "Functional convergence of Snell envelopes: Applications to American options approximations". In: *Finance and Stochastics* 2.3 (1998), pp. 311–327.

[Mug19] Aitor Muguruza. "Not so Particular about Calibration: Smile Problem Resolved". In: *arXiv preprint arXiv:1909.13366* (2019). arXiv: 1909.13366.

[MV68] Benoit B. Mandelbrot and John W. Van Ness. "Fractional Brownian Motions, Fractional Noises and Applications". In: *SIAM Review* 10.4 (1968), pp. 422–437.

[MW43] H. B. Mann and A. Wald. "On Stochastic Limit and Order Relationships". In: *Annals of Mathematical Statistics* 14.3 (1943), pp. 217–226.

[Neu94] Anthony Neuberger. "The Log Contract". In: *The Journal of Portfolio Management* 20.2 (1994), pp. 74–80.

[Nie04] Ari Nieminen. "Fractional Brownian motion and Martingale-differences". In: *Statistics and Probability Letters* 70.1 (2004), pp. 1–10.

[NM65] J A Nelder and R Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4 (1965), pp. 308–313.

[NP00] Carl J. Nuzman and H. Vincent Poor. *Linear Estimation of Self-Similar Processes via Lamperti's Transformation*. 2000.

[NR18] Eyal Neuman and Mathieu Rosenbaum. “Fractional Brownian motion with zero Hurst parameter: A rough volatility viewpoint”. In: *Electronic Communications in Probability* 23 (2018). arXiv: 1711.00427.

[NS16] Andreas Neuenkirch and Taras Shalaiko. “The Order Barrier for Strong Approximation of Rough Volatility Models”. In: (2016). arXiv: 1606.03854.

[Nua06] David Nualart. *The Malliavin Calculus and Related Topics*. Springer Science & Business Media, 2006.

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[Par14] Peter Parczewski. “A Fractional Donsker Theorem”. In: *Stochastic Analysis and Applications* 32.2 (2014), pp. 328–347.

[Par17] Peter Parczewski. “Donsker-type theorems for correlated geometric fractional Brownian motions and related processes”. In: *Electronic Communications in Probability* 22 (2017).

[Pol84] D Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

[Pow94] M. J. D. Powell. “A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation”. In: *Advances in Optimization and Numerical Analysis*. Springer Netherlands, 1994, pp. 51–67.

[PS14] Andrew Papanicolaou and Ronnie Sircar. “A regime-switching Heston model for VIX and S&P 500 implied volatilities”. In: *Quantitative Finance* 14.10 (2014), pp. 1811–1827.

[R M08] R. Mehra. *Handbook of the Equity Risk Premium: A volume in Handbooks in Finance*. Oxford: Elsevier, 2008, p. 496.

[Rie88] Thomas A. Rietz. “The equity risk premium a solution”. In: *Journal of Monetary Economics* 22.1 (1988), pp. 117–131.

[Rog03] L. C. G. Rogers. “Monte Carlo valuation of American options”. In: *Mathematical Finance* 12.3 (2003), pp. 271–286.

[Rog97] L. C.G. Rogers. “Arbitrage with fractional brownian motion”. In: *Mathematical Finance* 7.1 (1997), pp. 95–105.

[Ros15] Steve Ross. “The Recovery Theorem”. In: *Journal of Finance* 70.2 (2015), pp. 615–648.

[RS04] Alfredas Račkauskas and Charles Suquet. “Necessary and sufficient condition for the functional central limit theorem in Hölder spaces”. In: *Journal of Theoretical Probability* 17.1 (2004), pp. 221–243.

[RS13] Luis Miguel Rios and Nikolaos V. Sahinidis. “Derivative-free optimization: A review of algorithms and comparison of software implementations”. In: *Journal of Global Optimization*. Vol. 56. 3. Springer, 2013, pp. 1247–1293.

[RT96] Eric Renault and Nizar Touzi. “Option hedging and implied volatilities in a stochastic volatility model”. In: *Mathematical Finance* 6.3 (1996), pp. 279–302.

[RT97] Marc Romano and Nizar Touzi. “Contingent Claims and Market Completeness in a Stochastic Volatility Model”. In: *Mathematical Finance* 7.4 (1997), pp. 399–412.

[RW20] Johannes Ruf and Weiguan Wang. “Neural networks for option pricing and hedging: a literature review”. In: *Journal of Computational Finance, vol 24* (2020).

[SCC18] Uri Shaham, Alexander Cloninger, and Ronald R. Coifman. “Provable approximation properties for deep neural networks”. In: *Applied and Computational Harmonic Analysis* 44.3 (2018), pp. 537–557. arXiv: 1509.07385.

[SKM+93] Stefan G Samko, Anatoly A Kilbas, Oleg I Marichev, et al. *Fractional integrals and derivatives*. Vol. 1. Gordon and Breach Science Publishers, Yverdon Yverdon-les-Bains, Switzerland, 1993.

[Sne52] J. L. Snell. “Applications of Martingale System Theorems”. In: *Transactions of the American Mathematical Society* 73.2 (1952), p. 293.

[Sot01] Tommi Sottinen. “Fractional Brownian motion, random walks and binary market models”. In: *Finance and Stochastics* 5.3 (2001), pp. 343–355.

[SP97] Rainer Storn and Kenneth Price. “Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces”. In: *Journal of Global Optimization* 11.4 (1997), pp. 341–359.

[SS91] Elias M Stein and Jeremy C Stein. “Stock Price Distributions with Stochastic Volatility: An Analytic Approach”. In: *Review of Financial Studies* 4 (1991), pp. 727–752.

[Sto20] Henry Stone. “Calibrating rough volatility models: a convolutional neural network approach”. In: *Quantitative Finance* 20.3 (2020), pp. 379–392. arXiv: 1812.05315.

[Tan03] Peter Tankov. *Financial modelling with jump processes*. CRC press, 2003.

[Taq75] Murad S. Taqqu. “Weak convergence to fractional brownian motion and to the rosenblatt process”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31.4 (1975), pp. 287–302.

[TR19] Mehdi Tomas and Mathieu Rosenbaum. “From microscopic price dynamics to multidimensional rough volatility models”. In: (2019). arXiv: 1910.13338.

[Wei74] Robert Weinstock. *Calculus of variations: with applications to physics and engineering*. Courier Corporation, 1974.

[XWL10] Weidong Xu, Chongfeng Wu, and Hongyi Li. “Robust general equilibrium under stochastic volatility model”. In: *Finance Research Letters* 7.4 (2010), pp. 224–231.

[Zhu+97] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. “Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization”. In: *ACM Trans. Math. Softw.* 23.4 (1997), pp. 550–560.