

LARGE DEVIATIONS FOR ROUGH AND COMPLETE STOCHASTIC VOLATILITY MODELS

A THESIS PRESENTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY OF IMPERIAL COLLEGE LONDON

AND THE
DIPLOMA OF IMPERIAL COLLEGE

BY
CHLOÉ ALICE LACOMBE

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE
180 QUEEN'S GATE, LONDON SW7 2BZ

JUNE 2020

I, Chloé Lacombe, 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

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Large deviations for rough and complete stochastic volatility models

ABSTRACT

Stochastic volatility models are known to capture features observed in the markets and have been widely used in the financial industry. In particular, rough volatility models, where the instantaneous volatility is driven by a fractional Brownian motion, provide a more accurate fit of the power-law decay observed on volatility smiles close to maturity.

This thesis is concerned with deriving the asymptotic behaviour of these two classes of stochastic volatility models using large deviations techniques, in order to understand the behaviour of the implied volatility. Indeed, when pricing options, a closed-form solution to the pricing problem is often not available. Practitioners are using implied volatility as a measure of option prices that is consistent for different options with different maturities and strike prices. As this implied volatility cannot be expressed in closed form either, finding asymptotic formulas is crucial. We therefore focus on obtaining the asymptotic behaviour of two rough volatility models, namely a generalised version of the Stein-Stein model where the volatility starts from a random distribution and the (multi-factor) rough Bergomi model for realised variance options. We also study a particular complete model with stochastic volatility, which can be fitted to market data perfectly via the leverage function. In each case, we obtain the asymptotic behaviour of the stock price process and the implied volatility. We produce numerical schemes to compute the implied volatility smiles.

To my family, and in memory of my grandparents.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor, Doctor Antoine Jacquier, whose experience, passion and continuous support guided me through the completion of this thesis. He made sure these past few years were both challenging and enjoyable. Our stimulating conversations never failed to motivate and inspire me and his technical expertise was essential.

I would like to thank my co-authors, Blanka Horvath, Aitor Muguruza and Henry Stone for their valuable insights and inspiring conversations.

I am grateful to Professor Johannes Muhle-Karbe and Professor Peter Tankov for agreeing to spend time reading my thesis and to be the examiners for this viva.

I am deeply indebted to Mme Rey, M. Testud and of course M. Masse who encouraged me all throughout my education and without whom I might not have pursued studying mathematics.

Finally, I am forever grateful to my parents, Adèle, Camille and all my family for their constant support and unconditional love. This would not have been possible without you, merci.

"Be happy with what you have. Be excited about what you want."

- Alan Cohen.

LIST OF FIGURES

2.1	Implied volatility smiles for Call options on VIX for small maturities, close to the money. Data provided by OptionMetrics.	56
2.2	Rate function $\hat{\Lambda}^v$ for different values of H	72
2.3	Comparison of Monte Carlo computed implied volatilities (straight lines) and LDP based implied volatilities (stars), in the rough Bergomi model, for different values of $H_- =: \alpha$ and maturities T . We set $\eta = 2$ and $v_0 = 0.04$; for Monte Carlo we use 200,000 simulations and $\Delta t = \frac{1}{1008}$	73
2.4	Comparison of Alòs, García-Lorite and Muguruza [4] at-the-money implied volatility asymptotics and LDP based implied volatilities for different values of H_- , in the rough Bergomi model, with $\eta = 1.5$ and $v_0 = 0.04$	74
2.5	Confidence intervals (dashed lines) associated with the Monte Carlo simulations, for $H = 0.3$ and different maturities T . We set $\eta = 2$ and $v_0 = 0.04$; for Monte Carlo we use 200,000 simulations and $\Delta t = \frac{1}{1008}$	74
2.6	Comparison of LDP based implied volatilities for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$ in the mixed rough Bergomi process (2.6) such that $\gamma_1 \nu_1 + \gamma_2 \nu_2 = 2$, with $H_- = -0.4$	75
2.7	Rate function and corresponding implied volatilities for the model (2.22), with $(H_-, \nu, \eta) = (-0.4, 0.75, 0.75)$	77
2.8	Rate function and corresponding implied volatilities for the model (2.23) with $(H_-, \nu, \eta) = (-0.4, 1.0, 3.0)$	77
2.9	Implied volatilities and superimposed linear smiles for the model (2.23), with $(H_-, \nu, \eta) = (-0.4, 1.0, 3.0)$	77
2.10	Absolute difference of subsequent approximating polynomials $ \hat{\Lambda}_{n+1}^v(y) - \hat{\Lambda}_n^v(y) $	84

2.11	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 H_-	86
2.12	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 $\Delta t = \frac{1}{1008}$	91
3.1	Historical VVIX vs historical VIX (13/04/2012 - 08/05/2017). <i>Source: CBOE VIX Index and VVIX Index historical data.</i>	99
3.2	Process Y simulated using the Euler scheme between 2/1/2013 and 8/5/2017.	114
3.3	Simulated (blue) and historical (red) VIX, between 2/1/2013 and 8/5/2017.	114
3.4	Simulated (blue crosses) and historical (yellow circles) VIX against historical VVIX, between 2/1/13 and 8/5/17. The red line is the linear regression fit.	114
3.5	Simulated (blue crosses) and historical (yellow circles) VIX against the recent trend of S&P500, between 2/1/13 and 8/5/17. <i>Source: CBOE VIX Index.</i>	115

LIST OF TABLES

3.1	Parameters used in the simulations	113
3.2	Initial and calibrated parameters	116

CONTENTS

0	PRELIMINARIES ON FRACTIONAL BROWNIAN MOTION, LARGE DEVIATIONS AND IMPLIED VOLATILITY	6
0.1	Fractional Brownian motion	7
0.2	Large Deviations	9
0.3	Black-Scholes and implied volatility	11
1	ASYMPTOTIC BEHAVIOUR OF RANDOMISED FRACTIONAL VOLATILITY MODELS	13
1.1	Introduction	14
1.2	Main results	17
1.2.1	Tail behaviour	18
1.2.2	Small-time behaviour	22
1.3	Applications to Implied volatility asymptotics	26
1.3.1	General fractional case	26
1.3.2	Refined asymptotic results in the special diffusive case from Sections 1.2.1 and 1.2.2	31
1.4	Proofs of the main Theorems	35
1.4.1	Proof of Theorem 13	35
1.4.2	Proof of Theorem 19	44
1.4.3	Proof of Theorem 25	48
2	ASYMPTOTICS FOR VOLATILITY DERIVATIVES IN MULTI-FACTOR ROUGH VOLATILITY MODELS	52
2.1	Introduction	53
2.2	A showcase of rough volatility models	56

2.3	Small-time results for options on integrated variance	59
2.3.1	Small-time results for the rough Bergomi model	59
2.3.2	Small-time results for the mixed rough Bergomi model	65
2.3.3	Small-time results for the multi-factor rough Bergomi model	67
2.4	Numerical results	72
2.4.1	RV smiles for rough Bergomi	72
2.4.2	RV smiles for mixed rough Bergomi	75
2.4.3	RV smiles for mixed multi-factor rough Bergomi	75
2.5	Options on VIX	78
2.6	Numerical recipes	81
2.6.1	Convergence analysis of the numerical scheme	83
2.6.2	A tailor-made polynomial basis for rough volatility	83
2.6.3	Multi-Factor case	86
2.7	Approximating the density of realised variance in the mixed rough Bergomi model	87
2.8	Conclusions	91
2.9	Proof of main results	92
2.9.1	Proof of main large deviations results	92
2.9.2	Proof of $v_0\phi(d_1(x)) = x\phi(d_2(x))$	96
3	SMALL-TIME ASYMPTOTICS FOR COMPLETE STOCHASTIC VOLATIL- ITY MODELS	97
3.1	Set up and notations	101
3.2	Theoretical analysis of the model	102
3.2.1	Boundary conditions	102
3.2.2	Stationary distribution and pricing PDE	104
3.2.3	Small-time asymptotics	107
3.3	Numerical estimation of the coefficients (α, β, γ)	111

3.3.1	Simulations	113
3.3.2	Optimisation of coefficients	115
3.4	Proofs of the main results	116
3.4.1	Proof of Proposition 79	116
3.4.2	Proof of Theorem 80(i)	117
3.4.3	Proof of Theorem 80(ii)	122
3.4.4	Proof of Proposition 87	126
3.4.5	Proof of Theorem 86	128
4	CONCLUSION	130
	REFERENCES	132

INTRODUCTION

‘All models are wrong, but some are useful’. This statement by G.E.P. Box (1987) was not used to describe financial modelling even though it seems to fit perfectly. In the financial industry, pricing options involves calculating the expectation of the payoff of the option, which can be achieved by solving a pricing partial differential equation. However, a closed-form solution is rarely available. Commonly, practitioners have adopted a general measure of option prices that can be used for different options with different underlyings: the implied volatility. The latter is the unique solution obtained when matching market option prices with Black-Scholes option prices. However, a closed-form formula for the implied volatility, function of strike K and maturity T , does not exist and finding approximating formulas as well as understanding its asymptotics is crucial. The asymptotics (small and large maturities and strikes, etc.) are used to explain the behaviour of option prices in these extreme cases, to interpret the parameters of the model and to calibrate them to market data more efficiently. Instead of performing a global optimisation problem to estimate the parameters of the model, one can use the closed-form asymptotics formulas to get a specific initial set of parameters that will accelerate the optimisation problem and overcome some of its numerical limitations. Besides, deriving the asymptotic behaviour of the implied volatility can be used to determine control variates, an efficient technique to reduce the variances of Monte Carlo estimations used to price options.

In the last three decades, the mathematical finance community has proposed various models, in particular stochastic volatility models. Such models are both easy to implement and generate implied volatility surfaces that are consistent with market data, especially on Equity and Foreign Exchange (FX) markets where options with different strikes and maturities have different

implied volatilities. There exists a large class of stochastic volatility models with different dynamics, among which the Stein-Stein model [109] in which the instantaneous volatility process follows an Ornstein-Uhlenbeck (OU) process, the Bergomi model [18] where the instantaneous volatility is assumed to be a Wick stochastic exponential, or the Heston model [73] where the instantaneous variance process follows a Cox, Ingersoll and Ross (CIR) diffusion. In all these models, the underlying stock price process $(S_t)_{t \geq 0}$ and the instantaneous volatility process $(V_t)_{t \geq 0}$ are assumed to satisfy

$$\begin{aligned} dS_t &= \sigma(V_t) (\rho dB_t + \bar{\rho} dB_t^\perp) & S_0 = s_0 > 0, \\ dV_t &= a_V(t, S_t)dt + b_V(t, S_t)dB_t, & V_0 = v_0 > 0, \end{aligned}$$

where $(B, B^\perp)$ is a two-dimensional standard Brownian motion, $\rho \in [-1, 1]$, $\bar{\rho} := \sqrt{1 - \rho^2}$ and the mappings σ , a_V and b_V will differ in each model and are such that existence and uniqueness of a strong solution (S, V) is guaranteed. In particular, stochastic volatility models capture the leverage effect observed in the Equity market via the correlation ρ , and they fit the implied volatility surface more precisely than the Black-Scholes model. They however fail to fit the ‘small-time explosion’ observed on the implied volatility smiles close to maturity, as described in [64].

To be able to reproduce the power-law behaviour of the volatility smiles close to maturity, Alòs, León and Vives [5] and Fukasawa [59] extended the work done by Comte and Renault [34, 35], who first introduced fractional volatility models with Hurst parameter $H \in (1/2, 1)$. Such models assume that the instantaneous volatility is driven by a fractional Brownian motion instead of a standard one. Alòs, León and Vives [5], Fukasawa [59] and Gatheral, Jaisson, and Rosenbaum [65] proved that rough volatility models generate the ‘small-time explosion’ of the volatility smiles close to maturity, with a Hurst parameter close to 0.1. Bayer, Friz and Gatheral [13] then extended the results and showed how to use these models to price options. An extensive literature on rough volatility exists and the reader is invited to read [14, 46, 52, 59, 69, 76, 82, 90]. One shortcoming of such models is their computational cost, and competitive simulation methods were proposed by [17, 77, 94], along with Stone [110] and Horvath, Muguruza and Tomas [79]

who used deep learning techniques to calibrate the rough volatility models (Hurst parameter and implied volatility surface), considerably improving the computational speed. An alternative to fractional volatility models are the so-called ‘jump-diffusion and exponential Lévy models’ which incorporate a jump process in the stock price dynamics (see for example [36] and [111]).

Both academics and practitioners are taking more interest in stochastic and rough volatility models due, in particular, to the precise fit to the observed implied volatility surface. Such models assume that the market is incomplete, leading to a possible non-uniqueness of option prices. Local volatility models are an example of settings that overcome this shortcoming, together with so-called path-dependent volatility models introduced for example by Guyon [70], where the instantaneous volatility is assumed to depend on the current value of the stock price as well as a set of finite path-dependent variables. The leverage function enables a perfect fit to the volatility smiles using Dupire’s equation.

As stated above, the implied volatility cannot be expressed in closed form. Using large deviations techniques to estimate the exponential decay of probability measures of rare events, in the setting of stochastic volatility or path-dependent volatility models, enables to prove small- and large-time as well as tail asymptotics of the implied volatility. Such techniques were first introduced by Schilder [107], Freidlin and Wentzell [56], Deuschel-Stroock [42] and presented, for instance, in Dembo and Zeitouni [39]. Various academics adapted these methods to mathematical finance and proposed asymptotics of the implied volatility in different settings: Forde and Zhang [52] proved the small-time asymptotics for a correlated fractional stochastic volatility model. Jacquier, Pakkanen and Stone [82] studied the small-time behaviour of the rough Bergomi model, which was then expanded to the small-time behaviour for realised variance options by [90], and Jacquier and Shi [86] studied the randomised Heston model. We refer the reader to [12, 51, 76, 80, 82, 86, 90] (to name a few) who proved large deviations results for various (rough) stochastic volatility models.

In this thesis, we provide asymptotics of the implied volatility in different option pricing settings, focusing on three stochastic volatility models. The structure of the thesis is as follows.

Chapter 0 introduces fractional Brownian motions, provides notations and reminders on the large deviations methods used in this thesis, and recalls the definition of implied volatility.

In Chapter 1, we study the asymptotic behaviour of a generalised version of the Stein-Stein stochastic volatility model. We add two extensions to the original model: (i) the instantaneous volatility process is driven by a fractional Brownian motion, and (ii) the SDE driving the instantaneous volatility process starts from a random distribution. We focus on the tail and small-time asymptotics and show how, assuming a particular tail behaviour of the random distribution, we are able to modulate the rate of explosion of the implied volatility. We also present an application to forward-start European options to get the small-time forward smile asymptotics.

In Chapter 2, we derive the small-time asymptotics for the realised variance process of the rough Bergomi model and more general multi-factor rough Bergomi models. We then apply these results to obtain the small-time behaviour of options on realised variance. A simple yet accurate numerical scheme to compute the implied volatility smiles using large deviations results is then proposed. It enables to approximate the density of the realised variance and to price volatility swaps.

Finally, in Chapter 3, we focus on complete models with stochastic volatility. Conditions on the coefficients to obtain existence and uniqueness of a strong solution are first proved before deriving small-time asymptotics of the stock price process as well as the implied volatility. We also provide a calibration of the model to the simulated VIX data, a proxy of the instantaneous volatility of the S&P 500 Index prices.

NOTATIONS

$\mathbb{R}_+$	$[0, +\infty)$
$\mathbb{R}_+^*$	$(0, +\infty)$
$\mathcal{T}$	Index set included in $\mathbb{R}_+$
$\mathcal{T}^*$	$\mathcal{T} \setminus \{0\}$
A^o	The interior of a given set A
$\overline{A}$	The closure of a given set A
$L^2(\mathcal{T})$	Space of real-valued square integrable functions on $\mathcal{T}$
$\mathcal{C}(\mathcal{T}, \mathbb{R}^d)$	For $d \geq 1$, the space of $\mathbb{R}^d$ -valued continuous functions on $\mathcal{T}$
$\mathcal{C}$	$\mathcal{C}(\mathcal{T}, \mathbb{R})$
$\mathcal{C}_b^2$	Space of twice differentiable functions on $\mathcal{T}$ with bounded partial derivatives up to the second order
$\mathcal{H}$	The space of absolutely continuous functions starting at x with square integrable derivatives, also referred as the Cameron-Martin space
BV	The space of functions of finite variations
$H_\pm$	$H \pm \frac{1}{2}$, with H the Hurst parameter
h_ε	A function satisfying $\lim_{\varepsilon \downarrow 0} h_\varepsilon = 0$
$X^\varepsilon \sim \text{LDP}(h_\varepsilon, I)$	The sequence $(X^\varepsilon)_{\varepsilon > 0}$ satisfies a large deviations principle on $\mathcal{C}$, as ε tends to zero with good rate function I and speed h_ε
$\text{supp}(X)$	The support of a random variable X
$\langle X \rangle_t$	The quadratic variation of a process X at time t
$\mathcal{E}$	The Wick stochastic exponential
$f(x) = \mathcal{O}(g(x))$	There exists $C > 0$ and $y \in \mathbb{R}$ such that $ f(x) \leq Cg(x)$, for all $x \geq y$
x^+	$\max\{x, 0\}$, for $x \in \mathbb{R}$
$\mathcal{C}_{BS}(t, e^k, \Sigma)$	Price of an European Call option, with maturity t , strike e^k , and volatility Σ
$\mathcal{C}_{obs}(t, e^k)$	Observed Call option prices in a given market, with maturity t , and strike e^k



PRELIMINARIES ON FRACTIONAL BROWNIAN MOTION, LARGE DEVIATIONS AND IMPLIED VOLATILITY

Throughout this thesis, we will always work on a given filtered probability space $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, where $(S_t)_{t \geq 0}$ denotes the stock price process and $(V_t)_{t \geq 0}$ the instantaneous volatility process. In this Chapter, we recall the definition of a fractional Brownian motion and provide a brief review of the different large deviations methods used in this thesis. We also recall the definition of implied volatility that we will use later on.

0.1 FRACTIONAL BROWNIAN MOTION

On a real, separable Banach space $(\mathcal{E}, \|\cdot\|)$, we denote by $\mathcal{B}$ the associated Borel sigma field. Letting $\mathcal{E}^*$ denote the topological dual of $\mathcal{E}$, we define a Gaussian measure as follows:

Definition 1. *A probability measure μ on $(\mathcal{E}, \mathcal{B})$ is said to be a Gaussian measure if every $\vartheta^* \in \mathcal{E}^*$ is a real Gaussian random variable $\vartheta \mapsto \langle \vartheta^*, \vartheta \rangle_{\mathcal{E}^* \mathcal{E}}$ on the probability space $(\mathcal{E}, \mathcal{B}, \mu)$.*

We associate a Gaussian process to a Gaussian measure in the usual way in [30, Section 3.2]. Particular examples of Gaussian processes, crucial for the rest of this thesis, include standard Brownian motion on the time interval $\mathcal{T}$, where $\mathcal{E} = \mathcal{C}$ equipped with the supremum norm and with the topology of uniform convergence, and $\mathcal{E}^*$ is the space of signed measures on $\mathcal{T}$. In fact, this construction applies to all (centered) continuous Gaussian processes, which are uniquely characterised by their covariance operator.

Definition 2. *A fractional Brownian motion W^H with Hurst parameter $H \in (0, 1)$ is a centered and continuous Gaussian process, starting from zero, with covariance*

$$\langle W_t^H, W_s^H \rangle = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - |t - s|^{2H} \right), \quad \text{for any } s, t \in \mathcal{T}.$$

Remark 3. *A Hurst parameter $H = \frac{1}{2}$ corresponds to a standard Brownian motion. A Hurst parameter $H \in (\frac{1}{2}, 1)$ produces a persistent process where the increments are positively correlated, and if $H \in (0, \frac{1}{2})$, the process is anti-persistent and the increments are negatively correlated.*

Of primary importance in understanding small-noise behaviours of Gaussian systems is the concept of reproducing kernel Hilbert spaces (RKHS), which we recall following [30, Definition 3.3]:

Definition 4. *Let μ be a Gaussian measure on $\mathcal{E}$ and define the map $\mathcal{R} : \mathcal{E}^* \rightarrow \mathcal{E}$ by $\mathcal{R}x^* := \int_{\mathcal{E}} \langle x^*, x \rangle x \mu(dx)$. The RKHS $\mathcal{H}_\mu$ of μ is the completion of the image $\mathcal{R}\mathcal{E}^*$ for the norm $\|\mathcal{R}x^*\|_{\mathcal{H}_\mu} := \langle x^*, \mathcal{R}x^* \rangle^{1/2}$, for all $x^* \in \mathcal{E}^*$.*

To characterise the RKHS of fractional Brownian motion, the usual tool is its Volterra representation [102]

$$W_t^H = \int_0^t K^H(t, s) dB_s, \quad (1)$$

which holds almost surely for all $t \in \mathcal{T}$, where B is a standard Brownian motion generating the same filtration as W^H , and K^H is the Volterra kernel defined, for any $s, t \in \mathcal{T}$ with $0 < s < t$, by [102, Theorem 5.2]

$$K^H(t, s) = \begin{cases} \frac{\kappa_H}{s^{H_-}} \left[(t(t-s))^{H_-} - H_- \int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}} du \right], & \text{if } H < \frac{1}{2}, \\ \frac{\kappa_H H_-}{s^{H_-}} \int_s^t \frac{u^{H_-} du}{|u-s|^{1-H_-}}, & \text{if } H > \frac{1}{2}, \\ 1, & \text{if } H = \frac{1}{2}, \end{cases} \quad (2)$$

with $H_{\pm} := H \pm \frac{1}{2}$ and $\kappa_H := \left(\frac{2H\Gamma(1-H_-)}{\Gamma(H_+)\Gamma(2-2H)} \right)^{1/2}$. For $t \in \mathcal{T}^*$, the map $K^H(t, \cdot)$ is square integrable at the origin, and the reproducing kernel Hilbert space of the fractional Brownian motion is given by

$$\mathcal{H}_{K^H} := \left\{ \int_0^t K^H(t, s) f(s) ds, t \in \mathcal{T} : f \in L^2(\mathcal{T}) \right\}$$

with inner product

$$\left\langle \int_0^\cdot K^H(\cdot, s) f_1(s) ds, \int_0^\cdot K^H(\cdot, s) f_2(s) ds \right\rangle_{\mathcal{H}_{K^H}} := \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}.$$

The notation $\mathcal{H}_{K^H}$, emphasising the link with the underlying kernel, will be useful later (in Definition 5) for more general kernels. In particular, the RKHS associated to (standard) Brownian motion ($H = 1/2$) is the Cameron-Martin space, and corresponds to the space of absolutely continuous functions starting at zero, with square integrable derivatives. In other words, for any $H \in (0, 1) \setminus \{1/2\}$, the identity $\mathcal{H}_{K^H} = \mathbb{K}_{K^H} L^2(\mathcal{T})$ holds, where $\mathbb{K}_{K^H} : L^2(\mathcal{T}) \ni f \mapsto \int_0^t K^H(t, s) f(s) ds$. This characterisation motivates the following definition:

Definition 5. For any strictly positive function $\Phi : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ such that $\Phi(t, \cdot) \in L^2(\mathcal{T})$ for any $t \in \mathcal{T}$, the corresponding RKHS is defined as

$$\mathcal{H}_\Phi := \left\{ \int_0^t \Phi(t, s) f(s) ds, t \in \mathcal{T} : f \in L^2(\mathcal{T}) \right\}, \quad (3)$$

with inner product

$$\left\langle \int_0^\cdot \Phi(\cdot, s) f_1(s) ds, \int_0^\cdot \Phi(\cdot, s) f_2(s) ds \right\rangle_{\mathcal{H}_\Phi} := \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}.$$

Reproducing Kernel Hilbert Spaces, together with their inner products, turn out to provide the right spaces to characterise large deviations rate functions.

0.2 LARGE DEVIATIONS

We provide here an overview of the different large deviations methods we will use later on. We shall use [39, Chapter 1.2] and [39, Chapter 1] as our guides through large deviations. Given a topological space $\mathcal{X}$ and the completed Borel σ -field $\mathcal{B}_\mathcal{X}$ corresponding to $\mathcal{X}$, for any $A \in \mathcal{B}_\mathcal{X}$, we consider a sequence $(X^\varepsilon)_{\varepsilon>0}$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$.

Definition 6. A (good) rate function is a lower semi-continuous mapping $I : \mathcal{X} \rightarrow [0, \infty]$ such that the level sets $\{x : I(x) \leq z\}$ are closed (compact) subsets of $\mathcal{X}$ for any $z \geq 0$.

Definition 7. The sequence $(X^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle (LDP) on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ as ε tends to zero, with speed h_ε , and rate function I , if for any Borel subset $A \subset \mathcal{X}$, the following inequalities hold:

$$-\inf_{A^o} I(\varphi) \leq \liminf_{\varepsilon \downarrow 0} h_\varepsilon \log \mathbb{P}(X^\varepsilon \in A) \leq \limsup_{\varepsilon \downarrow 0} h_\varepsilon \log \mathbb{P}(X^\varepsilon \in A) \leq -\inf_{\overline{A}} I(\varphi). \quad (4)$$

We will use the notation $X^\varepsilon \sim \text{LDP}(h_\varepsilon, I)$. Two particularly convenient tools to prove large deviations are the so-called exponential equivalence and

Contraction Principle, which we recall from [39, Definition 4.2.10] and [39, Theorem 4.2.1] as follows:

Definition 8 (Exponential equivalence). *On a metric space $(\mathcal{Y}, d)$, two $\mathcal{Y}$ -valued sequences $(X^\varepsilon)_{\varepsilon>0}$ and $(\tilde{X}^\varepsilon)_{\varepsilon>0}$ 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 and, for each $\delta > 0$, the set $\{\omega : (\tilde{X}^\varepsilon, X^\varepsilon) \in \Gamma_\delta\}$ is $\mathcal{B}_\varepsilon$ -measurable, and*

$$\limsup_{\varepsilon \downarrow 0} h_\varepsilon \log \mathbb{P}^\varepsilon (\{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\}) = -\infty.$$

Theorem 9. *Let $(X^\varepsilon)_{\varepsilon>0}$ and $(\tilde{X}^\varepsilon)_{\varepsilon>0}$ be two exponentially equivalent sequences (with speed h_ε) on some metric space. If $(X^\varepsilon) \sim \text{LDP}(h_\varepsilon, I)$ for some good rate function I , then $(\tilde{X}^\varepsilon) \sim \text{LDP}(h_\varepsilon, I)$.*

Theorem 10 (Contraction Principle). *Let $\mathcal{X}$ and $\mathcal{Y}$ be topological spaces and $f : \mathcal{X} \rightarrow \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \rightarrow [0, \infty]$. 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 $\mathcal{Y}$.*

Two standard results in large deviations are Schilder's Theorem and the large deviations principle for Gaussian Volterra processes, recalled below following [39, Theorem 5.2.3] and [42, Theorem 3.4.5].

Theorem 11 (Schilder's Theorem). *Let $(W_t)_{t \in [0,1]}$ denote a standard Brownian motion in $\mathbb{R}^d$, $d \geq 1$. For $\varepsilon > 0$, consider $(W_t^\varepsilon)_{t \in [0,1]}$ defined by $W_t^\varepsilon := \sqrt{\varepsilon} W_t$, for $t \in [0, 1]$, and ν^ε the probability measure induced by B^ε on $\mathcal{C}_0([0, 1], \mathbb{R}^d)$, the space of all continuous functions from $[0, 1]$ to $\mathbb{R}^d$ such that the value at the origin is zero.*

Then the family $(\nu^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}_0([0, 1], \mathbb{R}^d)$ when ε tends to zero, with speed ε and rate function

$$\Lambda(\varphi) = \begin{cases} \frac{1}{2} \int_0^1 |\dot{\varphi}(t)|^2 dt, & \text{if } \varphi \in \mathcal{H}, \\ +\infty, & \text{otherwise.} \end{cases} \quad (5)$$

This theorem has been generalised to Gaussian Volterra processes, using the Reproducing Kernel Hilbert spaces.

Theorem 12. *For a given Gaussian Volterra process of the form $\int_0^\cdot \Phi(\cdot, s)dB_s$, for some Volterra kernel Φ , it follows from [42, Theorem 3.4.5] that the sequence $(\varepsilon \int_0^\cdot \Phi(\cdot, s)dB_s)_{\varepsilon>0}$ satisfies a large deviations principle with speed ε^2 and rate function*

$$\Lambda_\Phi(\varphi) = \begin{cases} \frac{1}{2}\|\varphi\|_{\mathcal{H}_\Phi}, & \text{if } \varphi \in \mathcal{H}_\Phi, \\ +\infty, & \text{otherwise.} \end{cases} \quad (6)$$

An obviously special role is played by the standard Brownian motion, with $H = 1/2$, and we shall adopt the simplified notation $\mathcal{H}$ (the classical Cameron-Martin space) and Λ in place of $\mathcal{H}_{K^{1/2}}$ and $\Lambda_{K^{1/2}}$, as defined in Equation (5).

0.3 BLACK-SCHOLES AND IMPLIED VOLATILITY

In this thesis, we consider, without loss of generality, the interest rate to be zero. In the Black-Scholes-Merton model, the underlying stock price process $(S_t)_{t \geq 0}$ is assumed to be a geometric Brownian motion (where the drift under the risk-neutral measure $\mathbb{Q}$ is null) starting from $S_0 > 0$. We denote the price of a European Call option $\mathcal{C}_{BS}(t, e^k, \Sigma)$, with $t \geq 0$ the maturity of the contract, $k := \log(K)$ with $K > 0$ the strike, and $\Sigma > 0$ the instantaneous volatility of the log returns $(X_t)_{t \geq 0}$, defined by $X := \log S$.

Under the risk-neutral measure $\mathbb{Q}$, the Call option value at maturity t is $\mathbb{E}^\mathbb{Q}[(S_t - K)^+]$. Using the Black-Scholes formula, the Call option is worth

$$\mathcal{C}_{BS}(t, e^k, \Sigma) := S_0 \mathcal{N}\left(\frac{\log S_0 - k}{\Sigma\sqrt{t}} + \frac{\Sigma\sqrt{t}}{2}\right) - e^k \mathcal{N}\left(\frac{\log S_0 - k}{\Sigma\sqrt{t}} - \frac{\Sigma\sqrt{t}}{2}\right)$$

with $\mathcal{N}(\cdot)$ the cumulative distribution function of the standard Normal distribution.

Considering a market with observed Call option prices $\mathcal{C}_{obs}(t, e^k)$, with maturity t and strike e^k , we denote by $\Sigma_t(k)$ the implied volatility, defined as the unique non-negative solution to

$$\mathcal{C}_{BS}(t, e^k, \Sigma_t(k)) = \mathcal{C}_{obs}(t, e^k).$$

Note that, as the implied volatility depends on the maturity t and log-strike k , the volatility smile refers to the implied volatility as a function of k , for fixed maturity t .

1

ASYMPTOTIC BEHAVIOUR OF RANDOMISED FRACTIONAL VOLATILITY MODELS

In this Chapter, we use Large Deviations techniques to study the asymptotic behaviour of a class of small-noise diffusions driven by a fractional Brownian motion, with random starting points. In order to do so, we extend some results on sample path large deviations for such diffusions. As an application, we show how these results characterise the small-time and tail estimates of the implied volatility for rough volatility models, recently proposed in mathematical finance.*

*The main results presented in this Chapter have previously appeared in the published paper [76].

1.1 INTRODUCTION

Motivated by financial applications, we derive here asymptotic small-time and tail behaviours of the solution to a generalised version of the Stein-Stein stochastic volatility model, originally proposed in [108, 109], using large deviations techniques. Large deviations are used extensively in Physics (thermodynamics, statistical mechanics) as well as in Mathematics (information theory, stochastic analysis, mathematical finance) to estimate the exponential decay of probability measures of rare events. Varadhan [112], Schilder [107], Freidlin and Wentzell [56] proved, in different degrees of generality, large deviations principles (in $\mathbb{R}^n$ and on path space) for solutions of stochastic differential equations with small noise, and the monographs by Dembo and Zeitouni [39] and Deuschel-Stroock [42] provide a precise account of those advances (at least up to the mid-1990s). In the past decade, this set of techniques and results has been adopted by the mathematical finance community: finite-dimensional large deviations (in the sense of Gärtner-Ellis) have been used to prove small-and large-time asymptotics of implied volatility in affine models [51, 80], sample-path LDP (à la Freidlin-Wentzell [56]) have proved efficient to determine importance sampling changes of probability [66, 68, 106], and heat kernel expansions (following Ben Arous [15] and Bismut [23]), have led to a general understanding of small-time and tail behaviour of multi-dimensional diffusions [11, 12, 40, 41]. These asymptotics have overall provided a deeper understanding of the behaviour of models, and, ultimately, allow for better calibration of real data; a general overview can be found in [104].

We in particular consider two important (in light of the recent trends in the literature proposed models) extensions: (i) the SDE driving the instantaneous volatility process is started from a random distribution; this so-called ‘randomised’ type of models was recently proposed in [85, 86, 96], in particular to understand the behaviour of the so-called ‘forward volatility’; (ii) the volatility process is driven by a fractional Brownian motion. Fractional stochastic volatility models, originally proposed by Comte and Renault [35, 34] with $H \in (1/2, 1)$, have recently been extended to the case

$H \in (0, 1/2)$, and a recent flourishing activity in this area [5, 13, 59, 65, 81] has established these models as the go-to standards for estimation and calibration.

The original motivation behind randomisation of the initial starting point is rooted in financial practice, where only the initial value of the stock price process is observed directly and the instantaneous value of volatility is subject to calibration. The effect of randomisation of the initial volatility on the implied volatility surface was explored by Jacquier and Roome [84] in a simple ‘random environment’ setting, where the volatility component was assumed to follow CEV dynamics. Their results give an impetus both on the theoretical and the practical level: they solve a practical modelling problem in a simple tractable setting and at the same time raise awareness for the potential prowess of applying random evolution equations for financial modelling. In this Chapter, we follow up on this direction and blend more involved approaches (proposed by [95, 97]) from the literature around random environment and random evolution equations into our financial model, where randomness also appears in the drift and diffusion coefficients of the process. On the practical level, independently from the results of [84], Mechkov [96] goes a step further in endorsing the idea of randomising the initial volatility and makes a strong case to move away from modelling hidden variables (such as stochastic volatility) in the traditional way. He argues that starting the volatility from a fixed starting point heavily underestimates the effect of the hidden variable on the slope of the implied volatility smile, and therefore ‘hot start’ volatility models (with random starting point) significantly outperform traditional ones altogether. Indeed, both randomised models [84, 96] produce the desired explosion in the smile at short maturities. Jacquier and Shi [86] develop this further by providing a precise link between the rate of explosion of implied volatilities on the short end and the tail distribution of the initial distribution of the volatility process in a ‘randomised’ Heston model. These outputs confirm that stochastic volatility models with random starting point constitute a class of counterexamples to the long-standing belief formulated by Gatheral [64, Chapter 5], that jumps in the stock price process are needed to produce steep short-dated implied volatility skews. Another example of

broadly different design was provided by Caravenna and Corbetta [29]. In their ‘multiscaling’ model, the stock price process is continuous, while the volatility process has (carefully designed) jumps, and steepness of the smile is achieved with a heavy-tail distribution of the small-time distribution of the volatility. Rough fractional volatility models (with continuous volatility paths) have recently been proposed, and are able to capture the volatility skew [5, 13, 14, 46, 52, 59, 65, 69, 82]. In this Chapter, we analyse the combined effect of a rough fractional Brownian driver (with Hurst parameter $H \in (0, 1)$) in the volatility and a random starting point. We quantify how the tail behaviour (parametrised by a scaling coefficient $b > 0$) of the random starting point modulates the rate of explosion in the implied volatility in the presence of rough fractional volatility. In particular, we generalise [52, Corollary 4.10] to the case of a rough Ornstein-Uhlenbeck volatility process starting from a random distribution.

Finally, in a specific simplified setting we highlight how our model blends naturally into the setting of forward-start options in stochastic volatility models, whose asymptotic properties have been studied in [85]. In proving our results, we improve the large deviations literature on both SDEs with random starting points and fractional SDEs.

Denote $(X_t)_{t \geq 0}$ the log-stock price process, such that $X := \log S$, and $(V_t)_{t \geq 0}$ the instantaneous volatility process. The particular system we are interested in is

$$\begin{cases} dX_t = -\frac{1}{2}\sigma(V_t)^2 dt + \sigma(V_t) (\rho dB_t + \bar{\rho} dB_t^\perp), & X_0 = 0, \\ dV_t = (\lambda + \beta V_t) dt + \xi dW_t^H, & V_0 \sim \Theta, \end{cases} \quad (1.1)$$

where W^H is a fractional Brownian motion, with Hurst parameter $H \in (0, 1)$, $(B, B^\perp)$ is a two-dimensional standard Brownian motion, $\beta < 0$, $\lambda, \xi > 0$, $\rho \in (-1, 1)$, $\bar{\rho} := \sqrt{1 - \rho^2}$, and Θ is a square-integrable continuous random variable. In order to guarantee existence and uniqueness of a strong solution, we further assume [99, Theorem 3.1.3] that σ is Lipschitz continuous, satisfies the growth condition $|\sigma(y)| \leq C(1 + |y|)$ for $y \in \mathbb{R}$, and is differentiable with locally Hölder continuous derivative.

This Chapter is structured as follows: Section 1.2 collects the main large deviations estimates in different regimes: tail behaviour (Section 1.2.1), and small-time behaviour (Section 1.2.2). In each case, we present two different scenarios consisting of an appropriately rescaled fractional model (Theorems 13 and 19) and a simplified diffusive model (Theorems 18 and 25) with more restrictive conditions on the random starting point (allowing for simpler large deviations rate functions). Section 1.3 displays applications to implied volatility asymptotics (Corollaries 27 and 29), and presents an application to forward-start options. Proofs of the main Theorems 13, 19 and 25 can be found at the end of the Chapter, in Section 1.4.

Unless otherwise stated, we always work on a finite time horizon, say $[0, 1]$ without loss of generality, which we denote by $\mathcal{T}$, and we write $\mathcal{T}^* := \mathcal{T} \setminus \{0\}$.

1.2 MAIN RESULTS

Centrepiece of our analysis are large deviations estimates for suitably rescaled versions of (1.1). In order to prove our main results below, we make the following technical assumptions:

Assumption $\mathcal{A}_b$: There exist a measurable function $\tilde{\sigma} : \mathbb{R} \rightarrow \mathbb{R}$, with locally Hölder continuous derivative, such that $\tilde{\sigma}$ is Lipschitz continuous, satisfying the growth condition $|\tilde{\sigma}(y)| \leq C(1 + |y|)$ for $y \in \mathbb{R}$, as well as a constant $b > 0$ such that the scaling property $\lim_{\varepsilon \downarrow 0} \varepsilon^b \sigma(y/\varepsilon^b) = \tilde{\sigma}(y)$ holds true uniformly on $\mathbb{R}$.

The first rescaling (presented in Section 1.2.1) is tailored to the analysis of the tail behaviour of (1.1), while the second rescaling (Section 1.2.2) is bespoke to its short-time asymptotic properties. In addition to these asymptotic results in the general fractional case (Theorems 13 and 19), we present two special simplified diffusive cases, where particularly tractable rate functions can be obtained (Theorems 18 and 25 respectively). For this we impose stronger conditions (Assumption 16) on the random starting point. This allows us to establish, following [97] in Section 1.2.1, an exponential equivalence between (1.1) and an analogous process with fixed starting point. In Section 1.2.2, we construct a third rescaling (1.10) inspired by Mellouk [95]

in the short-time diffusive case under the assumption that the support of the random starting point is bounded. We shall work with rescaled versions V^ε of the process V in (1.1) (see (1.2) and (1.8) for specific examples), together with a function h_ε describing the speed of the large deviations estimates, for which we introduce the following assumptions:

Assumption $\mathcal{A}'_b$: There exists a family of continuous functions $(\sigma_n)_{n \geq 0}$ on $\mathbb{R}$ such that

- (i) $(\sigma_n)_{n \geq 0}$ converges uniformly to $\tilde{\sigma}$ on $\mathbb{R}$;
- (ii) for all $\delta > 0$, $\lim_{n \uparrow \infty} \limsup_{\varepsilon \downarrow 0} h_\varepsilon \log \mathbb{P}(|\sigma_n(V^\varepsilon) - \tilde{\sigma}(V^\varepsilon)| \geq \delta) = -\infty$.

Assumption $\mathcal{A}_b^\Theta$ (tail behaviour of Θ): The limit $\limsup_{\varepsilon \downarrow 0} h_\varepsilon \log \mathbb{P}(\varepsilon^b |\Theta| > 1) = -\infty$ holds.

In Assumptions $\mathcal{A}'_b$ and $\mathcal{A}_b^\Theta$ above, the large deviations speed h_ε takes the value ε^{2b} in Section 1.2.1 and ε^{4H+2b} in Section 1.2.2 respectively. The constant b (which may vary below) plays an essential role in subsequent large deviations estimates, via exponential equivalence techniques (Definition 8). Assumptions $\mathcal{A}_b$ and $\mathcal{A}'_b$ are naturally satisfied in the fractional Stein-Stein case (where $\sigma(y) \equiv y$), but imposing them allows us to state our results for more general σ , in particular when using extended Contraction Principles [93, Proposition 2.3]. Finally, from Assumption $\mathcal{A}_b$, it holds that $\tilde{\sigma}(y) := cy$, with $c \in \mathbb{R}$, but we keep the notation $\tilde{\sigma}$ for convenience.

1.2.1 TAIL BEHAVIOUR

THE GENERAL CASE

For $b, \varepsilon > 0$, introduce the rescaling $(X^\varepsilon, V^\varepsilon) := (\varepsilon^{2b} X, \varepsilon^b V)$, so that (1.1) becomes

$$\begin{cases} dX_t^\varepsilon = -\frac{\varepsilon^{2b}}{2} \sigma\left(\frac{V_t^\varepsilon}{\varepsilon^b}\right)^2 dt + \varepsilon^{2b} \sigma\left(\frac{V_t^\varepsilon}{\varepsilon^b}\right) (\rho dB_t + \bar{\rho} dB_t^\perp), & X_0^\varepsilon = 0, \\ dV_t^\varepsilon = (\varepsilon^b \lambda + \beta V_t^\varepsilon) dt + \varepsilon^b \xi dW_t^H, & V_0^\varepsilon \sim \varepsilon^b \Theta. \end{cases} \quad (1.2)$$

The particular rescaling considered here is perfectly suited for tail behaviour, as large deviations provide estimates for $\mathbb{P}(X^\varepsilon \geq 1) = \mathbb{P}(X \geq \varepsilon^{-2b})$. Our main result is as follows, and is proved in Section 1.4.1:

Theorem 13. *For any $H \in (0, 1)$, the following hold:*

- (i) *for any $b > 0$ such that $\mathcal{A}_b^\Theta$ holds, $V^\varepsilon \sim \text{LDP}(\varepsilon^{2b}, \Lambda_{FH})$, with Λ_{FH} in (6) and F^H in Lemma 14;*
- (ii) *for any $b > 0$ such that Assumptions $\mathcal{A}_b$, $\mathcal{A}'_b$, $\mathcal{A}_b^\Theta$ hold, $X^\varepsilon \sim \text{LDP}(\varepsilon^{2b}, \tilde{\Lambda})$, with $\tilde{\Lambda}$ in (1.19).*

The proof of the theorem, developed later, requires a precise analysis of the Reproducing Kernel Hilbert Spaces of the processes under consideration, and we first state the following two key ingredients (proved in Sections 1.4.1 and 1.4.1), which are also of independent interest:

Lemma 14. *For any $H \in (0, 1)$ and $\beta > 0$, there exists a standard Brownian motion Z , such that*

$$\xi \int_0^t e^{\beta(t-s)} dW_s^H = \int_0^t F^H(t, s) dZ_s, \quad (1.3)$$

holds almost surely for $t \in \mathcal{T}$, where $F^H : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ is defined for $0 < s < t$, with κ_H in (2), as

$$F^H(t, s) := \begin{cases} \frac{\xi \kappa_H}{s^{H-}} \left[[t(t-s)]^{H-} + \int_s^t \left\{ \frac{1-2H}{2u} + \beta \right\} [u(u-s)]^{H-} e^{\beta(t-u)} du \right], & \text{if } H < \frac{1}{2}, \\ \frac{\xi \kappa_H H-}{s^{H-}} \int_s^t \frac{u^{H-} e^{\beta(t-u)} du}{(u-s)^{1-H-}}, & \text{if } H > \frac{1}{2}, \\ \xi e^{\beta(t-s)}, & \text{if } H = \frac{1}{2}. \end{cases}$$

The case $\beta = 0$ (and $\xi = 1$) is excluded since, in that case, the lemma boils down to the classical Volterra representation of the fractional Brownian motion (1) and the function F^H is nothing else that K^H given in (2). For $H \in (\frac{1}{2}, 1)$, the expression for F^H is in agreement with [113, Definition 2.1], but, for $H \in (0, \frac{1}{2})$, it corrects the slightly erroneous expression therein. This

function F^H allows us to fully characterise the following RKHS, with proof postponed to Section 1.4.1:

Proposition 15. *For any $H \in (0, 1)$, $\beta > 0$, the space $\mathcal{H}_{F^H}$ is the RKHS of the Gaussian process $\xi \int_0^\cdot e^{\beta(\cdot-s)} dW_s^H$.*

THE MILLET-NUALART-SANZ APPROACH

In [97], Millet, Nualart and Sanz consider a perturbed stochastic differential equation of the form

$$dX_t^\varepsilon = b(\varepsilon, X_t^\varepsilon)dt + \varepsilon a(X_t^\varepsilon)dW_t. \quad (1.4)$$

Here, for any $\varepsilon > 0$, the functions $b(\varepsilon, \cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $a : \mathbb{R}^n \rightarrow \mathcal{M}_{(n,d)}(\mathbb{R})$ are bounded Borel measurable and uniformly Lipschitz, $b(\varepsilon, \cdot)$ converges uniformly to a function $b(\cdot)$ as ε tends to zero, W is a d -dimensional Brownian motion, and X_0^ε is an $\mathbb{R}^n$ -valued square-integrable random variable. Existence and uniqueness of a strong solution can be found in [55, Chapter 5, Theorem 2.1]. Following classical large deviations steps, consider, for any $\varphi \in \mathcal{H}$, the controlled ordinary differential equation on $\mathcal{T}$:

$$\dot{\psi}_t = a(g_t)\dot{\varphi}_t + b(\psi_t), \quad (1.5)$$

the solution flow of which, starting from $x_0 \in \mathbb{R}^n$ is denoted by $\mathcal{S}_{x_0}(\varphi)$. Millet, Nualart and Sanz [97] proved a large deviations principle [97, Theorem 4.1] for the sequence $(X^\varepsilon)_{\varepsilon>0}$ under the following assumption:

Assumption 16. *Both $a(\cdot)$ and $b(\cdot)$ belong to $\mathcal{C}_b^2$, and there exists $x_0 \in \mathbb{R}^n$ such that, for any $\delta > 0$,*

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(|X_0^\varepsilon - x_0| > \delta) = -\infty. \quad (1.6)$$

Theorem 17. *Under Assumption 16, $(X^\varepsilon)_{\varepsilon>0} \sim \text{LDP}(\varepsilon^2, I)$ with $I(\psi) = \inf\{\Lambda(\varphi) : \varphi \in \mathcal{H}, \psi = \mathcal{S}_{x_0}(\varphi)\}$.*

Condition (1.6) is an exponential equivalence property between the initial random variable X_0^ε and the constant x_0 , and ensures that large deviations are preserved under exponentially small perturbations of the starting point. Therefore, in the standard diffusion case $H = \frac{1}{2}$, it is possible to obtain a similar result to Theorem 13 with a simplified rate function (albeit with slightly more restrictions on the starting point), by using the approach considered by Millet-Nualart-Sanz in [97]. For this, we rewrite (1.4) to correspond to (1.2), albeit with stronger assumptions on the coefficients, with $W := (W_1, W_2)'$ a Brownian motion, $X^\varepsilon = (X^\varepsilon, V^\varepsilon)$, $H = 1/2$, $\varepsilon \rightarrow \varepsilon^b$, $X_0^\varepsilon = (0, \varepsilon\Theta)$, and $\bar{\rho} := \sqrt{1 - \rho^2}$,

$$b(\varepsilon, X_t^\varepsilon) = \begin{pmatrix} -\frac{1}{2}\tilde{\sigma}(V_t^\varepsilon)^2 \\ \varepsilon\lambda + \beta V_t^\varepsilon \end{pmatrix} \quad \text{and} \quad a(X_t^\varepsilon) = \begin{pmatrix} \bar{\rho}\tilde{\sigma}(V_t^\varepsilon) & \rho\tilde{\sigma}(V_t^\varepsilon) \\ 0 & \xi \end{pmatrix}.$$

The correlation between the two components of W is explicitly represented in the diffusion matrix a .

Theorem 18. *Under Assumption 16, the solution X^ε to (1.4) satisfies $(X^\varepsilon)_{\varepsilon>0} \sim \text{LDP}(\varepsilon^2, \mathbf{I})$ with $\mathbf{I}(\chi) = \inf \{\Lambda(\varphi) : \varphi \in \mathcal{H}, \chi = \mathcal{S}_{x_0}(\Psi^\rho(\varphi))\}$ and $\Psi^\rho : \mathbb{R}^2 \ni z \mapsto \begin{pmatrix} \bar{\rho} & \rho \\ 0 & 1 \end{pmatrix} z$.*

Proof. The proof of Theorem 17 relies first on proving a large deviations principle for the flow $\mathcal{S}_{x_0}$ using Schilder's theorem [39, Theorem 5.2.3], then on extending this LDP to the original system. One can easily extend it to include a correlation parameter $\rho \in (-1, 1)$, the main difference being the rate function. Indeed, since $W_2^\perp$ and W_2 are independent, Schilder's theorem yields that $\varepsilon(W_2^\perp, W_2)' \sim \text{LDP}(\varepsilon^2, \Lambda)$. Since the map Ψ^ρ is continuous on $(\mathcal{C}, \|\cdot\|_\infty)$ and $\varepsilon W' = \Psi^\rho(\varepsilon(W_2^\perp, W_2)')$, the theorem follows from the Contraction Principle giving an LDP for $\varepsilon W'$ as ε tends to zero with speed ε^2 and good rate function

$$\Lambda^\rho(\psi) := \inf \{\Lambda(\varphi) : \varphi \in \mathcal{H}, \psi = \Psi^\rho(\varphi)\}. \quad (1.7)$$

□

1.2.2 SMALL-TIME BEHAVIOUR

We now tackle the small-time behaviour of the process (1.1). Under the general set of assumptions $\mathcal{A}_b, \mathcal{A}'_b, \mathcal{A}_b^\Theta$, we need to introduce a particular rescaling, both in time and in space in order to observe some weak convergence. This is different from the classical Itô diffusion case (with fixed starting point), where solutions of such SDEs generally converge in small time. In the Itô case, though, if the distribution of the starting point has compact support, we show in Section 1.2.2 that space rescaling is not required any longer.

THE GENERAL CASE

With the rescaling $(X_t^\varepsilon, V_t^\varepsilon) := (\varepsilon^{2H+2b-1}X_{\varepsilon^2t}, \varepsilon^bV_{\varepsilon^2t})$, with $b > 0$, (1.1) becomes

$$\begin{cases} dX_t^\varepsilon = -\frac{\varepsilon^{2H+1+2b}}{2}\sigma\left(\frac{V_t^\varepsilon}{\varepsilon^b}\right)^2 dt + \varepsilon^{2H+2b}\sigma\left(\frac{V_t^\varepsilon}{\varepsilon^b}\right)(\rho dB_t + \bar{\rho}dB_t^\perp), & X_0^\varepsilon = 0, \\ dV_t^\varepsilon = (\varepsilon^{b+2}\lambda + \beta\varepsilon^2V_t^\varepsilon)dt + \varepsilon^{2H+b}\xi dW_t^H, & V_0^\varepsilon \sim \varepsilon^b\Theta. \end{cases} \quad (1.8)$$

Our main result is as follows:

Theorem 19. *For any $H \in (0, 1)$,*

- (i) *for any $b > 0$ such that $\mathcal{A}_b^\Theta$ holds $V^\varepsilon \sim \text{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$, with $\Lambda_{G_0^H}$ as in (6);*
- (ii) *for any $b > 0$ such that $\mathcal{A}_b, \mathcal{A}'_b, \mathcal{A}_b^\Theta$ hold, $X^\varepsilon \sim \text{LDP}(\varepsilon^{4H+2b}, \mathbb{I})$, with $\mathbb{I}$ defined in (1.22).*

The proof of (i) is similar to that of Theorem 13(i) and relies on proving LDP for an auxiliary process, defined in (1.17), exponentially equivalent to the original (rescaled) process V^ε . The proof of (ii) is more involved and postponed to Section 1.4.2. In order to state the following key result, define, for any $\varepsilon > 0$, G_ε^H as the function F^H in Lemma 14, replacing β by $\beta\varepsilon^2$, and

for $s, t \in \mathcal{T}$ with $0 < s < t$, its pointwise limit $G_0^H(t, s) := \lim_{\varepsilon \downarrow 0} G_\varepsilon^H(t, s)$ reads

$$G_0^H(t, s) = \begin{cases} \frac{\xi \kappa_H H_-}{s^{H_-}} \int_s^t \frac{u^{H_-}}{(u-s)^{1-H_-}} du, & \text{for } H \in \left(\frac{1}{2}, 1\right), \\ \frac{\xi \kappa_H}{s^{H_-}} \left((t(t-s))^{H_-} - H_- \int_s^t (u-s)^{H_-} u^{H_- - 1} du \right), & \text{for } H \in \left(0, \frac{1}{2}\right), \\ \xi, & \text{for } H = \frac{1}{2}. \end{cases}$$

The following proposition is similar to Proposition 15, as $G_0^H(t, \cdot) \in L^2(\mathcal{T})$ and for all $0 < s < t$, $G_0^H(t, s) > 0$, and its proof is omitted.

Proposition 20. *For any $H \in (0, 1)$, the space $\mathcal{H}_{G_0^H}$ is the RKHS of the Gaussian process $\int_0^\cdot G_0^H(\cdot, s) dZ_s$.*

SMALL-TIME ASYMPTOTICS FOR BOUNDED SUPPORT IN THE DIFFUSION CASE

In the standard case $H = \frac{1}{2}$, the rescaling in the previous subsection is not really ‘natural’, in the sense that small-time large deviations usually hold for Itô diffusions without space rescaling. In this case, using an approach introduced by Bezuidenhout [20] and further developed by Mellouk [95], we can obtain simpler large deviations estimates if the support of the initial datum Θ is bounded. A simplified version of Mellouk considers, for any $\varepsilon > 0$, the system, on $\mathcal{T}$,

$$dX_t^\varepsilon = b(X_t^\varepsilon, Z)dt + \varepsilon a(X_t^\varepsilon, Z)dW_t, \quad X_0^\varepsilon = x_0 \in \mathbb{R}^n, \quad (1.9)$$

where $b : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $a : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathcal{M}_{(n,d)}$ are bounded Borel measurable, uniformly Lipschitz continuous, Z is a random variable with compact support on $\mathbb{R}^m$ and W a d -dimensional standard Brownian motion, independent of Z . The main result of the paper is a large deviations principle on $\mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^n)$, the space of α -Hölder continuous functions, for $0 \leq \alpha < \frac{1}{2}$, for the sequence $(X^\varepsilon)_{\varepsilon > 0}$, under the following assumptions:

Assumption 21.

(H1) $b(\cdot, \cdot)$ is jointly measurable on $\mathbb{R}^n \times \mathbb{R}^m$ and there exists $C > 0$ such that, for all $x, x' \in \mathbb{R}^n$, $z, z' \in \mathbb{R}^m$,

$$|b(x, z)| \leq C(1 + |x|) \quad \text{and} \quad |b(x, z) - b(x', z')| \leq C(|x - x'| + |z - z'|).$$

(H2) $a(\cdot, \cdot)$ is jointly measurable on $\mathbb{R}^n \times \mathbb{R}^m$ and there exists $C > 0$ such that, for all $x, x' \in \mathbb{R}^n$, $z, z' \in \mathbb{R}^m$,

$$\|a(x, z)\| \leq C \quad \text{and} \quad \|a(x, z) - a(x', z')\| \leq C(|x - x'| + |z - z'|).$$

For $f \in \mathcal{H}$, $u \in \text{supp}(Z)$ and $x_0 \in \mathbb{R}^n$, let $\mathcal{S}_{x_0}(f, u)$ denote the unique solution to the controlled ODE $g_t = x_0 + \int_0^t b(g_s, u_s)ds + \int_0^t a(g_s, u_s)\dot{f}_s ds$, for $t \in \mathcal{T}$. Let us now introduce the following definition:

Definition 22. Let $\alpha \in [0, \frac{1}{2})$ and $\mathcal{B}_a$ be the ball of radius a in the α -Hölder norm. The lower semi-continuous regularisation $\check{I} : \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m) \rightarrow \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m)$ of a functional $I : \mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^m) \rightarrow \mathbb{R}$ is defined as

$$\check{I}(\psi) := \liminf_{a \downarrow 0} \{I(\varphi) : \varphi \in \mathcal{B}_a(\psi)\}.$$

Theorem 23 (Theorem 2.1 in [95]). Under Assumption 21, $(X^\varepsilon)_{\varepsilon > 0} \sim \text{LDP}(\varepsilon^2, \check{I}^\alpha)$, where (with Λ in (6))

$$I^\alpha(\psi) := \inf \{\Lambda(\varphi) : \varphi \in \mathcal{H}, \mathcal{S}_{x_0}(\varphi, u) = \psi, \text{ for some } u \in \text{supp}(Z)\}.$$

Coming back to our model, the rescaling $(X_t^\varepsilon, V_t^\varepsilon) := (X_{\varepsilon^2 t}, V_{\varepsilon^2 t})$, equivalent to that of Section 1.2.2 with $b = 0$ and $H = \frac{1}{2}$, the small-noise system (1.1), under Assumption $\mathcal{A}_b$, becomes, similarly to Section 1.2.2,

$$\begin{cases} dX_t^\varepsilon = -\frac{\varepsilon^2}{2}\tilde{\sigma}(V_t^\varepsilon)^2 dt + \varepsilon\tilde{\sigma}(V_t^\varepsilon) (\rho dB_t + \bar{\rho} dB_t^\perp), & X_0^\varepsilon = 0, \\ dV_t^\varepsilon = (\varepsilon^2\lambda + \beta\varepsilon^2 V_t^\varepsilon) dt + \varepsilon\xi dB_t, & V_0^\varepsilon \sim \Theta, \end{cases} \quad (1.10)$$

with B a standard Brownian motion. Subtracting the initial random da-

tum $X_0^\varepsilon = X_0 = (0, \Theta)'$, this system can be expressed in the form (1.9) with $X^\varepsilon = (X^\varepsilon, V^\varepsilon)$,

$$b(\varepsilon, X^\varepsilon, X_0) = \varepsilon^2 \begin{pmatrix} -\frac{1}{2}\tilde{\sigma}(V_t^\varepsilon + \Theta)^2 \\ \lambda + \beta(V_t^\varepsilon + \Theta) \end{pmatrix} \text{ and } a(X^\varepsilon, X_0) = \begin{pmatrix} \rho\tilde{\sigma}(V_t^\varepsilon + \Theta) & \bar{\rho}\tilde{\sigma}(V_t^\varepsilon + \Theta) \\ \xi & 0 \end{pmatrix}, \quad (1.11)$$

and note that $b(\varepsilon, \cdot, \cdot)$ converges to the null map as ε tends to zero. The assumptions imposed in [95] on the drift and diffusion coefficients are clearly satisfied here. While Mellouk allows the drift and diffusion to depend explicitly on external random factors, we can write our setting (dependence on a random starting point) into this framework. The large deviations estimate for the sequence $(X^\varepsilon)_{\varepsilon>0} = (X^\varepsilon, V^\varepsilon)_{\varepsilon>0}$ thus obtained is stronger than that in the previous section, as it holds on $\mathcal{C}^\alpha(\mathcal{T}, \mathbb{R}^n)$, for any $0 \leq \alpha < \frac{1}{2}$. Note that the mild conditions on the coefficients [95, (H_0) - (H_2)] are easily satisfied in our case, so the only additional assumption is the boundedness of the support on Θ . We remark here that [95] is not directly applicable to the current setting but has to be extended to include ε -dependence in the drift, and we do so following the Azencott [7]'s inspired approach developed by Peithmann [105, Subsection 2.2.1]. In order to state LDP (proved in Section 1.4.3), we impose the following assumption:

Assumption 24. *For $u \in \text{supp}(X_0)$ and $\varphi \in \mathcal{H}$, the ODE $\psi_t = \int_0^t a(\psi_s, u_s) \dot{\varphi}_s ds$ has a unique solution on $\mathcal{T}$, denoted by $\mathcal{S}_0(\varphi, u)$.*

Theorem 25. *Under Assumption 24, if Θ has compact support, then $(X^\varepsilon)_{\varepsilon>0} \sim \text{LDP}(\varepsilon^2, \check{I}^\alpha)$, with*

$$I^\alpha(\psi) := \inf \{ \Lambda^\rho(\varphi) : \varphi \in \mathcal{H}, \text{ such that } \mathcal{S}_0(\varphi, u) = \psi, \text{ for some } u \in \text{supp}(X_0) \},$$

with Λ^ρ defined in (1.7).

1.3 APPLICATIONS TO IMPLIED VOLATILITY ASYMPTOTICS

As announced in the introduction of Chapter 1, we unify here two branches of research, both aimed at reproducing the steepness of the implied volatility surface on the short end via models with continuous paths. While there are now numerous outputs [5, 13, 14, 52, 59, 65, 82] in the literature confirming that a fractional driving noise (with Hurst exponent $H < 1/2$) in the volatility leads to the observed steepness of the smile, recent results [84, 86] reproduce this effect by randomising the initial volatility in classical diffusive models. In this section we observe an explosion of the implied volatility in the small-time case, related to the tail behaviour of the random distribution Θ , independently of the Hurst parameter $H \in (0, 1)$. As recalled in Chapter 0, in the Black-Scholes-Merton model, the price of a European Call option is $\mathcal{C}_{BS}(t, e^k, \Sigma)$, with associated volatility Σ . Considering a market with observed Call option prices $\mathcal{C}_{obs}(t, e^k)$, with maturity t and strike e^k , we denote by $\Sigma_t(k)$ the implied volatility, defined as the unique non-negative solution to $\mathcal{C}_{BS}(t, e^k, \Sigma_t(k)) = \mathcal{C}_{obs}(t, e^k)$.

1.3.1 GENERAL FRACTIONAL CASE

From Theorem 19, we can deduce the asymptotic behaviour of out-of-the-money Put/Call options and the implied volatility for small maturities. We state the result below.

Corollary 26. *For any $H \in (0, 1)$ and any $b > 0$ such that Theorem 13 holds, we have the following large-strike asymptotic estimates of out-of-the-money Put/Call options:*

$$\begin{aligned} \lim_{k \uparrow \infty} \frac{1}{k} \log \mathbb{E} \left[(S_t - e^k)^+ \right] &= - \inf_{y \geq 1} \tilde{\Lambda}(\phi)|_{\phi_t=y}, \quad \text{and} \\ \lim_{k \uparrow \infty} \frac{1}{k} \log \mathbb{E} \left[(e^k - S_t)^+ \right] &= - \inf_{y \geq 1} \tilde{\Lambda}(\phi)|_{\phi_t=y}, \quad \text{with } \tilde{\Lambda} \text{ as in (1.19), and for any } t \in \mathcal{T}. \end{aligned}$$

Proof. We prove this result similarly to [67, Proposition 4.1] and apply [16, Lemma 3.2]. Indeed, fixing $t \in \mathcal{T}$ and taking ε^{-2b} to be k we have, from

Theorem 13, that

$$\lim_{k \uparrow \infty} \frac{1}{k} \log \mathbb{P}(S_t \geq e^k) = - \inf_{y \geq 1} \tilde{\Lambda}(\phi)|_{\phi_t=y}. \quad (1.12)$$

Thus, $-\log \mathbb{P}(S_t \geq e^k) \in \mathcal{R}_1$ with $\mathcal{R}_1$ the space of regularly varying functions with index 1, defined in [16, Definition 2.1]. The result follows then directly from [16, Lemma 3.2]. $\square$

Corollary 27. *(Large-strike implied volatility asymptotics) For any $H \in (0, 1)$ and any $b > 0$ such that Theorem 13 holds, we have the following large-strike asymptotic estimates of the implied volatility:*

$$\lim_{k \uparrow \infty} \frac{\Sigma_t^2(k)t}{k} = \frac{1}{2} \left(\inf_{y \geq 1} \tilde{\Lambda}(\phi)|_{\phi_t=y} \right)^{-1} \quad \text{with } \tilde{\Lambda} \text{ as in (1.19), and for any } t \in \mathcal{T}.$$

Proof. This is a straightforward application of [61, Corollary 7.1] and Corollary 26. Taking ε^{-2b} to be k we have, from Theorem 13, that $\varepsilon^{2b}X_t \sim \text{LDP}(\varepsilon^{2b}, \tilde{\Lambda})$ as ε goes to zero. Then,

$$\lim_{k \uparrow \infty} \frac{1}{k} \log \mathbb{P}(X_t \geq k) = - \inf_{y \geq 1} \tilde{\Lambda}(\phi)|_{\phi_t=y}.$$

Similarly, in Black-Scholes, $\lim_{k \uparrow \infty} \frac{\Sigma_t^2(k)}{k^2} \log \mathbb{P}(X_t \geq k) = -\frac{1}{2t}$, and the proof follows from [61, Corollary 7.1]. $\square$

Similarly, from Theorem 19 we can deduce the asymptotic behaviour of out-of-the-money Put/Call options and the implied volatility when time becomes small.

Corollary 28. *Introduce $b_H := H + b - 1/2$. For any $H \in (0, 1)$ and any $b \geq 1/2 - 2H$ such that Theorem 19 holds, we have the following small-time asymptotic estimates of out-of-the-money Put/Call options:*

$$\lim_{t \downarrow 0} t^{2H+b} \log \mathbb{E} \left(S_t - e^{kt^{-b_H}} \right)^+ = - \inf_{y \geq k} \mathbb{I}(\phi)|_{\phi_1=y}, \quad \text{with } \mathbb{I} \text{ as in (1.22)}.$$

Proof. We prove the lower and upper bounds separately, similarly to [52, Appendix C]. Firstly, taking t to be 1, ε^2 to be t and $b_H := H + b - 1/2$,

we have, from Theorem 19 that

$$\lim_{t \downarrow 0} t^{2H+b} \log \mathbb{P}(t^{bH} X \geq k) = - \inf_{y \geq k} \mathbb{I}(\phi)|_{\phi_1=y}. \quad (1.13)$$

(1) For the lower bound, for all $\delta > 0$, we have:

$$\begin{aligned} \mathbb{E} \left(S_t - e^{kt^{-bH}} \right)^+ &\geq \left(e^{k(1+\delta)t^{-bH}} - e^{kt^{-bH}} \right) \mathbb{P} \left(S_t > e^{k(1+\delta)t^{-bH}} \right), \\ &\geq e^{kt^{-bH}} \delta t^{-bH} \mathbb{P} \left(S_t > e^{k(1+\delta)t^{-bH}} \right). \end{aligned}$$

Then using (1.13) and that $\lim_{t \downarrow 0} t^{2H+b} \log t = 0$, we obtain

$$\begin{aligned} \liminf_{t \downarrow 0} t^{2H+b} \log \mathbb{E} \left(S_t - e^{kt^{-bH}} \right)^+ &= \liminf_{t \downarrow 0} t^{2H+b} \log \mathbb{P} \left(S_t > e^{k(1+\delta)t^{-bH}} \right), \\ &\geq - \inf_{y \geq k(1+\delta)} \mathbb{I}(\phi)|_{\phi_1=y}. \end{aligned}$$

Take $\delta \downarrow 0$ and use the continuity of the good rate function to obtain the desired lower bound.

(2) For the upper bound, we first note that for all $q > 1$, we have

$$\mathbb{E} \left(S_t - e^{kt^{-bH}} \right)^+ \leq \left[\mathbb{E} \left(\left(S_t - e^{kt^{-bH}} \right)^+ \right)^q \right]^{1/q} \mathbb{P} \left(S_t \geq e^{kt^{-bH}} \right)^{1-1/q},$$

leading

$$\begin{aligned} t^{2H+b} \log \mathbb{E} \left(S_t - e^{kt^{-bH}} \right)^+ &\leq \frac{t^{2H+b}}{q} \log \mathbb{E} \left[\left(\left(S_t - e^{kt^{-bH}} \right)^+ \right)^q \right] + \\ &\quad + t^{2H+b} \left(1 - \frac{1}{q} \right) \log \mathbb{P} \left(S_t \geq e^{kt^{-bH}} \right), \\ &\leq \frac{t^{2H+b}}{q} \log \mathbb{E} (S_t^q) + t^{2H+b} \left(1 - \frac{1}{q} \right) \log \mathbb{P} \left(S_t \geq e^{kt^{-bH}} \right). \end{aligned}$$

Then, similarly to [52, Appendix C], we then have $\mathbb{E}(S_t^q) = \mathbb{E}(R_1 \cdot R_2)$,

with R_1 and R_2 two non negative random variables defined by

$$\begin{aligned} R_1 &:= \exp \left\{ \frac{1}{2} (q^2 - q + q^2 \rho^2) \int_0^t \sigma^2(V_s) ds \right\}, \\ R_2 &:= \exp \left\{ q^2 \rho^2 \int_0^t \sigma^2(V_s) ds + q \rho \int_0^t \sigma(V_s) dB_s \right\}. \end{aligned}$$

Then, by using the Cauchy-Schwarz inequality and the fact that R_2^2 is a time- t value of a martingale starting from 1, we have

$$\begin{aligned} \mathbb{E}(S_t^q) &\leq \sqrt{\mathbb{E}(R_1^2 \cdot R_2^2)}, \\ &= \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 q + q^2 \rho^2) \int_0^t \sigma^2(V_s) ds \right\} \right) \times} \\ &\quad \sqrt{\mathbb{E} \left(\exp \left\{ 2q^2 \rho^2 \int_0^t \sigma^2(V_s) ds + 2q \rho \int_0^t \sigma(V_s) dB_s \right\} \right)}, \\ &= \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 q + q^2 \rho^2) \int_0^t \sigma^2(V_s) ds \right\} \right)}, \\ &\leq \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 q + q^2 \rho^2) \int_0^t 2C(1 + |V_s|^2) ds \right\} \right)}. \end{aligned}$$

We then use (1.23) taking t to be 1 and ε^2 to be t , and obtain

$$\begin{aligned} \mathbb{E}(S_t^q) &\leq \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 - q + q^2 \rho^2) t \int_0^1 2C(1 + |V_s|^2) ds \right\} \right)}, \\ &\leq \exp \left\{ Ct (q^2 - q + q^2 \rho^2) \left[1 + 4 \left(\frac{\lambda}{\beta} \right)^2 \left(\frac{e^{2\beta t} - 1}{2\beta t} + 1 \right) \right] \right\} \times \\ &\quad \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 - q + q^2 \rho^2) \frac{2C}{\beta} \Theta^2 (e^{2\beta t} - 1) \right\} \right)} \times \\ &\quad \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 - q + q^2 \rho^2) 4C\xi^2 \int_0^t |W_s^H|^2 ds \right\} \right)}. \end{aligned}$$

For $q > 1$, introduce $f(q) := 4C\xi^2 (q^2 - q + q^2 \rho^2)$. For $\varepsilon > 0$ and $u \geq 0$,

assume that $\varepsilon f(q)(u + \varepsilon)^{2H} \leq \frac{1}{4}$. We then get

$$\begin{aligned} \mathbb{E} \left(e^{f(q) \int_u^{u+\varepsilon} |W_s^H|^2 ds} \right) &\leq \frac{1}{\varepsilon} \int_u^{u+\varepsilon} \mathbb{E} \left(e^{\varepsilon f(q) |W_s^H|^2} \right) ds, \\ &= \frac{1}{\varepsilon} \int_u^{u+\varepsilon} \frac{ds}{\sqrt{1 - 2\varepsilon f(q) s^{2H}}}, \\ &\leq \sqrt{2}. \end{aligned}$$

Thus, taking $u = 0$ and $\varepsilon = t > 0$ small enough so that $f(q)t^{2H+1} \leq \frac{1}{4}$, we get

$$\begin{aligned} \mathbb{E}(S_t^q) &\leq \sqrt[4]{2} \exp \left\{ Ct (q^2 - q + q^2 \rho^2) \left[1 + 4 \left(\frac{\lambda}{\beta} \right)^2 \left(\frac{e^{2\beta t} - 1}{2\beta t} + 1 \right) \right] \right\} \times \\ &\quad \sqrt{\mathbb{E} \left(\exp \left\{ (q^2 - q + q^2 \rho^2) \frac{2C}{\beta} \Theta^2 (e^{2\beta t} - 1) \right\} \right)}. \end{aligned}$$

Finally, use Assumption $\mathcal{A}_b^\Theta$ to obtain

$$\lim_{q \uparrow \infty} \limsup_{t \downarrow 0} \frac{t^{2H+b}}{q} \log \mathbb{E}(S_t^q) \leq 0,$$

and conclude the proof. □

The following Corollary generalises [52, Corollary 4.10].

Corollary 29 (Small-time implied volatility asymptotics). *For any $H \in (0, 1)$ and any $b > 0$ such that Theorem 19 holds, the following small-time estimate is true for any $k \neq 0$:*

$$\lim_{t \downarrow 0} t^b \Sigma_t^2 (t^{1/2-H-b} k) = \frac{k^2}{2} \left(\inf_{y \geq k} \mathbb{I}(\phi) |_{\phi_1=y} \right)^{-1}, \quad \text{with } \mathbb{I} \text{ as in (1.22)}. \quad (1.14)$$

This implies that the implied volatility explodes with rate t^{-b} . For $b = 0$, it is identical to [52, Formula (26)].

Proof. This is a straightforward application of [61, Corollary 7.1] and Corollary 28. Taking ε^2 to be t and $b_H := H + b - 1/2$, from Theorem 19, $t^{b_H} X \sim \text{LDP}(t^{2H+b}, \mathbb{I})$, as t goes to zero,

$$\lim_{t \downarrow 0} t^{2H+b} \log \mathbb{P}(t^{b_H} X \geq k) = - \inf_{y \geq k} \mathbb{I}(\phi)|_{\phi_1=y}.$$

Similarly, in the Black-Scholes model, we obtain

$$\lim_{t \downarrow 0} \frac{t \Sigma_t^2(t^{-b_H} k)}{t^{-2b_H}} \log \mathbb{P}(t^{b_H} X \geq k) = \lim_{t \downarrow 0} \frac{t \Sigma_t^2(t^{-b_H} k)}{t^{-2b_H}} \log \mathbb{P}(X \geq t^{-b_H} k) = -\frac{k^2}{2},$$

and the result follows from [61, Corollary 7.1]. $\square$

1.3.2 REFINED ASYMPTOTIC RESULTS IN THE SPECIAL DIFFUSIVE CASE FROM SECTIONS 1.2.1 AND 1.2.2

LARGE-STRIKE ASYMPTOTICS

We consider here a specific case of a multidimensional diffusion, as we are only interested in studying its tail asymptotics. Let $X^\zeta := (X^{\zeta,1}, \dots, X^{\zeta,n})$ be the unique strong solution in $\mathbb{R}^n$ to

$$dX_t^\zeta = \tilde{b}(X_t^\zeta)dt + a(X_t^\zeta)dW_t, \quad X_0^\zeta = \zeta,$$

for some d -dimensional standard Brownian motion and some square integrable random variable ζ , and with $\tilde{b} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $a : \mathbb{R}^n \rightarrow \mathcal{M}_{(n,d)}(\mathbb{R})$. Consider the following scaling assumption:

Assumption 30. *There exist $b_1, \dots, b_n > 0$ with $b_1 = 2$ such that $X^{\varepsilon,\zeta} := (\varepsilon^{b_1} X^{\zeta,1}, \dots, \varepsilon^{b_n} X^{\zeta,n})$ satisfies*

$$dX_t^{\varepsilon,\zeta} = \varepsilon \tilde{b}(\varepsilon, X_t^{\varepsilon,\zeta})dt + \varepsilon a(X_t^{\varepsilon,\zeta})dW_t, \quad X_0^\zeta = (\varepsilon^{b_1} \zeta^{(1)}, \dots, \varepsilon^{b_n} \zeta^{(n)}). \quad (1.15)$$

Furthermore, $\varepsilon \tilde{b}(\varepsilon, \cdot/\varepsilon)$ converges uniformly to some function $b(\cdot)$ as ε tends to zero.

We can then state the main result about tail asymptotics:

Proposition 31. *Under Assumption 30, if there exists $x_0 \in \mathbb{R}^n$ such that for any $\delta > 0$,
 $\limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(|\zeta - x_0| > \delta) = -\infty$, and if the triple $(x_0, b(\cdot), a(\cdot))$ satisfies Assumption 16, then*

$$\lim_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X_t^{\zeta,1} \geq 1) = \lim_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X_t^{x_0,1} \geq 1), \quad \text{for any } t \in \mathcal{T}.$$

The scaling from Assumption 30 may be odd at first, but reflects the fact that components of stochastic models may each act on different scales. Consider for example the system where the instantaneous volatility is an Ornstein-Uhlenbeck process, solution to

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t^2 dt + V_t dW_t, & X_0 &= \zeta, \\ dV_t &= (\lambda + \beta V_t) dt + \xi dB_t, & V_0 &= y_0 > 0, \end{aligned}$$

where W and B are two correlated Brownian motions. The rescaling $(X^\varepsilon, V^\varepsilon) := (\varepsilon^2 X, \varepsilon V)$ (corresponding to $b_1 = 2$ and $b_2 = 1$) yields

$$\begin{aligned} dX_t^\varepsilon &= -\frac{1}{2}(V_t^\varepsilon)^2 dt + \varepsilon V_t^\varepsilon dW_t, & X_0^\varepsilon &= \varepsilon^2 \zeta, \\ dV_t^\varepsilon &= (\varepsilon \lambda + \beta V_t^\varepsilon) dt + \varepsilon \xi dB_t, & V_0^\varepsilon &= \varepsilon y_0 > 0, \end{aligned}$$

namely (1.15), and the assumptions are satisfied.

Proof. For any $\varepsilon > 0$, the pathwise rescaled process $X^{\varepsilon, \zeta}$ satisfies (1.15). The proof of the proposition relies on the (more general) theorem proved by Millet, Nualart and Sanz [97], recalled in Section 1.2.1, and whose validity is guaranteed by Assumption 16. Note first that, from standard large deviations considerations (and in particular contraction mappings), the process $X^{\varepsilon, 0}$ satisfies a large deviations principle with good rate function I given in Theorem 17. Recall that by construction $\tilde{b}(\varepsilon, \cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ satisfies $\lim_{\varepsilon \downarrow 0} \varepsilon \tilde{b}(\varepsilon, x/\varepsilon) = b(x)$ uniformly as ε tends to zero. Therefore Theorem 17 yields a large deviations principle for the sequence $(X^{\varepsilon, \zeta})_{\varepsilon \geq 0}$ as ε tends to zero, with good rate function I and speed ε^2 . In particular, for

$A := \{\psi \in \mathcal{C}(\mathcal{T}, \mathbb{R}^n), \forall x \in \mathbb{R}^n, \psi(1, x) \geq 1\}$, we have

$$\begin{cases} -\inf_{\psi \in \bar{A}} I(\psi) \leq \liminf_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(X_t^{\varepsilon, \zeta} \geq 1) \leq \limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(X_t^{\varepsilon, \zeta} \geq 1) \leq -\inf_{\psi \in \bar{A}} I(\psi), \\ -\inf_{\psi \in \bar{A}} I(\psi) \leq \liminf_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(X_t^{\varepsilon, x_0} \geq 1) \leq \limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(X_t^{\varepsilon, x_0} \geq 1) \leq -\inf_{\psi \in \bar{A}} I(\psi). \end{cases}$$

Since Λ is continuous on $\mathcal{H}$, it is upper semi-continuous on A , and so is the good rate function I by [3, Lemma 2.41]. As a good rate function, it is also lower semi-continuous, and hence continuous, on A . Translating the two sets of inequalities above in terms of εX_t^ζ and $\varepsilon X_t^{x_0}$, and using the continuity of I proves the proposition. $\square$

The assumption of the random initial condition $X_0^\zeta = \zeta$ being $\mathcal{F}_0$ -measurable distribution can be relaxed. Indeed, the filtration $\mathcal{F}$ is the filtration generated by the d -dimensional Brownian motion W . Instead, one could work on a filtration $\mathcal{F}' := (\mathcal{F}'_t)_{t \in \mathcal{T}^*}$ generated by $\mathcal{F}'_t := \sigma(\{W_u, u \leq t\} \cup \{\zeta\})$, for all $t \in \mathcal{T}^*$. Then the random initial point X_0^ζ has a $\mathcal{F}'_0$ -measurable distribution and the results above still hold, in particular Theorem 17 and Proposition 31, on the new filtered probability space $(\Omega, \mathcal{F}', (\mathcal{F}'_t)_{t \in \mathcal{T}}, \mathbb{P})$. In the context of implied volatility asymptotics, this result has the following meaning:

Corollary 32. *The tails of the smile are independent of the starting point (ζ or x_0).*

Proof. Gao and Lee [61] show that asymptotic behaviour of the implied volatility can be directly inferred from comparing tail probabilities to those of the Black-Scholes model. It is straightforward to see that the scaling of Proposition 31 is the same in Black-Scholes, and the corollary follows immediately. $\square$

SMALL-TIME ASYMPTOTICS FOR THE ‘FORWARD’ STEIN-STEIN MODEL

We are interested in a ‘forward’ process, as defined by Jacquier and Roome [84] in the context of forward-start European options:

$$\mathbb{E} \left(\frac{S_{t+\tau}}{S_t} - e^k \right)^+ = \mathbb{E} (e^{X_{t+\tau}-X_t} - e^k)^+ =: \mathbb{E} (e^{X_\tau^{(t)}} - e^k)^+,$$

with $(X_\tau^{(t)})_{\tau \geq 0}$ the so-called ‘forward’ process, defined path-wise by $X_\tau^{(t)} := X_{t+\tau} - X_t$, for some fixed $t > 0$, and for all $\tau \geq 0$. The ‘forward’ process $(X_\tau^{(t)})_{\tau \geq 0}$ then satisfies the following stochastic differential equation:

$$\begin{cases} dX_\tau^{(t)} = -\frac{1}{2}(V_\tau^{(t)})^2 d\tau + V_\tau^{(t)} dW_{1,\tau}, & X_0^{(t)} = 0, \\ dV_\tau^{(t)} = (\lambda + \beta V_\tau^{(t)}) d\tau + \xi dW_{2,\tau}, & V_0^{(t)} \sim \sigma_t. \end{cases} \quad (1.16)$$

The stochastic differential equation for $(X_\tau^{(t)}, V_\tau^{(t)})_{\tau \geq 0}$ is the same as that for $(X_t, V_t)_{t \geq 0}$, albeit with an initial random distribution (δ_0, σ_t) , where σ_t is Gaussian with mean $e^{\beta t}(\sigma_0 + \frac{\lambda}{\beta}) - \frac{\lambda}{\beta}$ and variance $\frac{\xi^2}{2\beta}(e^{2\beta t} - 1)$. We now apply the results of Section 1.2.2 to obtain small-time asymptotics for a version of the Stein-Stein ‘forward’ model with a generalised random starting point.

Proposition 33. *With the scaling $(X_\tau^\varepsilon, V_\tau^\varepsilon) := (X_{\varepsilon^2\tau}^{(t)}, V_{\varepsilon^2\tau}^{(t)})$ for $\varepsilon, t > 0$, the randomised Stein-Stein rescaled model (1.16) is the same as (1.10) with coefficients given in (1.11), with $\tilde{\sigma}(y) \equiv y$, and Theorem 25 applies.*

We can translate this result into forward implied volatility asymptotics directly using [49, Theorem 2.4]. The value of a forward-start option with forward-start date t , maturity τ and strike e^k is $\mathcal{C}_{BS}(\tau, e^k, \Sigma)$ (stationary increments). For a market with observed Call option prices $\mathcal{C}_{obs}(t, e^k)$, the forward implied volatility smile $\Sigma_{t,\tau}$ is defined as the unique non-negative solution to $\mathcal{C}_{obs}(t, e^k) = \mathcal{C}_{BS}(\tau, e^k, \Sigma_{t,\tau}(k))$.

Corollary 34. *The small-time forward smile reads $\lim_{\tau \downarrow 0} \Sigma_{t,\tau}^2(k) = \frac{k^2}{2} \left(\inf_{y \geq k} \check{I}^\alpha(\phi)|_{\phi_1=y} \right)^{-1}$, with $\check{I}^\alpha$ in Theorem 25.*

1.4 PROOFS OF THE MAIN THEOREMS

1.4.1 PROOF OF THEOREM 13

PROOF OF LEMMA 14

From [113, Definition 2.1], for any smooth function f on $\mathcal{T}$ with bounded derivatives such that $\|f\| := \int_{\mathcal{T}} (\Gamma_{H,t}^* f(t))^2 dt < \infty$, define, for $t \in \mathcal{T}$, $\int_0^t f(u) dW_u^H := \int_0^t \Gamma_{H,t}^* f(u) dZ_u$, with

$$\Gamma_{H,t}^* f(s) := -\frac{\kappa_H}{s^{H-}} \frac{d}{ds} \int_s^t u^{H-} (u-s)^{H-} f(u) du, \quad \text{for all } s \in (0, t].$$

Applying this to $f(s) := \xi e^{\beta(t-s)}$, one obtains the following:

- (1) for $H < \frac{1}{2}$, using integration by part and Leibniz's integration rule,

$$\begin{aligned} & \frac{d}{ds} \int_s^t u^{H-} (u-s)^{H-} f(u) du \\ &= \frac{d}{ds} \left\{ \frac{\xi(t-s)^{H+} t^{H-}}{H_+} \right\} + \\ &+ \frac{d}{ds} \left\{ -\xi \frac{H_-}{H_+} \int_s^t \frac{(u-s)^{H+}}{u^{1-H-}} e^{\beta(t-u)} du + \xi \frac{\beta}{H_+} \int_s^t (u-s)^{H+} u^{H-} e^{\beta(t-u)} du \right\}, \\ &= -\xi(t-s)^{H-} t^{H-} + \xi H_- \int_s^t \frac{(u-s)^{H-}}{u^{1-H-}} e^{\beta(t-u)} du - \xi \beta \int_s^t (u-s)^{H-} u^{H-} e^{\beta(t-u)} du. \end{aligned}$$

Hence,

$$\begin{aligned} \Gamma_{H,t}^* f(s) &:= -\frac{\kappa_H}{s^{H-}} \frac{d}{ds} \int_s^t u^{H-} (u-s)^{H-} f(u) du, \\ &= \frac{\xi \kappa_H}{s^{H-}} \left((t-s)^{H-} t^{H-} + \beta \int_s^t (u-s)^{H-} u^{H-} e^{\beta(t-u)} du \right) - \\ &- \frac{\xi \kappa_H}{s^{H-}} \left(H_- \int_s^t \frac{(u-s)^{H-}}{u^{1-H-}} e^{\beta(t-u)} du \right). \end{aligned}$$

(2) for $H > \frac{1}{2}$, using Leibniz's integration rule,

$$\begin{aligned} \frac{d}{ds} \int_s^t u^{H-}(u-s)^{H-} f(u) du &= \int_s^t \xi \frac{d}{ds} u^{H-}(u-s)^{H-} e^{\beta(t-u)} du, \\ &= -\xi H_- \int_s^t u_-^H(u-s)^{H-1} e^{\beta(t-u)} du, \end{aligned}$$

and hence

$$\Gamma_{H,t}^* f(s) := -\frac{\kappa_H}{s^{H-}} \frac{d}{ds} \int_s^t u^{H-}(u-s)^{H-} f(u) du = \frac{\xi \kappa_H H_-}{s^{H-}} \int_s^t u_-^H(u-s)^{H-1} e^{\beta(t-u)} du.$$

PROOF OF PROPOSITION 15

In order to prove that $\mathcal{H}_{FH}$ is the RKHS of the Gaussian process $\xi \int_0^t e^{\beta(\cdot-s)} dW_s^H$, we use [42, Theorem 3.4.12, p88] and its application in [82, proof of Theorem 3.8]. Hence, one needs to show that the operator $\mathbb{K}_{FH}$ is a bijection from $L^2(\mathcal{T})$ to $\mathcal{H}_{FH}$ and that $\mathcal{H}_{FH}$ is dense in $\mathcal{C}$.

The operator $\mathbb{K}_{FH}$ acting on $L^2(\mathcal{T})$ defined by $(\mathbb{K}_{FH} f)(t) := \int_0^t F^H(t, s) f(s) ds$ satisfies $\mathbb{K}_{FH} L^2(\mathcal{T}) =: \mathcal{H}_{FH}$ and is surjective. It is also injective on $\mathcal{H}_{FH}$: let $f \in L^2(\mathcal{T})$ a non-zero function such that $(\mathbb{K}_{FH} f) = 0$ on $\mathcal{T}^*$. Similar to [52], there is an interval $[t_1, t_2] \subset \mathcal{T}$ where f has constant sign. Using previous notations, F^H is defined, for $t \in \mathcal{T}^*$ and $s \in (0, t]$, as $F^H(t, s) := \xi \Gamma_{H,t}^* e^{\beta(t-s)}$. The function $g(s) := \xi \int_t^s u^{H-}(u-s)^{H-} e^{\beta(t-u)} du$ is increasing, hence $F^H(t, \cdot)$ is a positive function on $[0, t]$ and $\int_{t_1}^{t_2} F^H(t, s) f(s) ds > 0$, leading to a contradiction. Thus $\mathbb{K}_{FH}$ is injective on $\mathcal{H}_{FH}$, hence bijective from $L^2(\mathcal{T})$ to $\mathcal{H}_{FH}$. Since $\mathbb{K}_{FH}$ is a linear operator, one can then define $\langle \mathbb{K}_{FH} f_1, \mathbb{K}_{FH} f_2 \rangle_{\mathcal{H}_{FH}} := \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}$ as an inner product.

The second part of the proof consists in showing that $\mathcal{H}_{FH} := \mathbb{K}_{FH} L^2(\mathcal{T})$ is dense in $\mathcal{C}$. We claim that $F^H(t, \cdot)^{-1} \in L^2(\mathcal{T})$. Indeed, as shown in the proof of [113, Theorem 3.1], the following hold for all $0 < s < t$:

- if $0 < H < \frac{1}{2}$, $F^H(t, s) \geq \xi \kappa_H s^{\frac{1}{2}-H} \tau^{H-}(t-s)^{H-} > 0$, leading

$$\begin{aligned} \int_0^t \frac{ds}{F^H(t, s)^2} &\leq \frac{1}{\xi^2 \kappa_H^2 t^{2H-}} \int_0^t \left(\frac{s}{t-s} \right)^{2H-} ds, \\ &\leq \frac{1}{\xi^2 \kappa_H^2 t^{4H-}} \int_0^t s^{2H-} ds, \\ &= \frac{t^{2-2H}}{2H \xi^2 \kappa_H^2} < \infty; \end{aligned}$$

- if $\frac{1}{2} < H < 1$, $F^H(t, s) \geq \xi \kappa_H e^{-\beta(s-t)}(t-s)^{H-}$, so that

$$\begin{aligned} \int_0^t \frac{ds}{F^H(t, s)^2} &\leq \frac{1}{\xi^2 \kappa_H^2 e^{2\beta t}} \int_0^t \frac{e^{2\beta s} ds}{(t-s)^{2H-}}, \\ &\leq \frac{1}{\xi^2 \kappa_H^2 e^{2\beta t} t^{2H-}} \int_0^t e^{2\beta s} ds, \\ &= \frac{1 - e^{-2\beta t}}{2\beta \xi^2 \kappa_H^2 t^{2H-}} < \infty. \end{aligned}$$

Given $a > 0$, take for $s \in [0, t]$, $f_a(s) := \frac{s^a}{F^H(t, s)} \in L^2(\mathcal{T})$. Then, $\int_0^t F^H(t, s) f_a(s) ds = \frac{t^{a+1}}{a+1}$. Hence, $\mathcal{H}_{FH}$ contains all polynomials null at the origin, and by the Stone-Weierstrass Theorem, is dense in $\mathcal{C}$.

PROOF OF THEOREM 13(I)

Let $\bar{V}^\varepsilon$ denote the solution to the second SDE (1.2), starting from zero. The product rule for (fractional) Brownian motion [99, Theorem 3.1.4] yields, for any $t \geq 0$,

$$\bar{V}_t^\varepsilon = -\frac{\lambda \varepsilon^b}{\beta} (1 - e^{\beta t}) + \xi \varepsilon^b \int_0^t e^{\beta(t-u)} dW_u^H, \quad (1.17)$$

Since $\xi \int_0^\cdot e^{\beta(\cdot-u)} dW_u^H$ is Gaussian, Lemma 14 combined with [42, Theorem 3.4.5] yields a large deviations principle on $\mathcal{C}$ for the sequence $(\xi \varepsilon^b \int_0^\cdot e^{\beta(\cdot-s)} dW_s^H)_{\varepsilon>0}$, with speed ε^{2b} and good rate function Λ_{FH} as in (6). Since the two sequences $(\bar{V}^\varepsilon)_{\varepsilon>0}$ and $(\xi \varepsilon^b \int_0^\cdot e^{\beta(\cdot-s)} dW_s^H)_{\varepsilon>0}$ only differ by some deterministic quantity,

they are clearly exponentially equivalent: for any $\delta > 0$ and $t \in \mathcal{T}^*$,

$$\begin{aligned} & \limsup_{\varepsilon \downarrow 0} \varepsilon^{2b} \log \mathbb{P} \left(\left\| \bar{V}^\varepsilon - \xi \varepsilon^b \int_0^\cdot e^{\beta(\cdot-s)} dW_s^H \right\|_\infty > \delta \right), \\ &= \limsup_{\varepsilon \downarrow 0} \varepsilon^{2b} \log \mathbb{P} \left(\left\| \frac{\lambda \varepsilon^b}{\beta} (e^{\beta \cdot} - 1) \right\|_\infty > \delta \right), \\ &= -\infty, \end{aligned}$$

since $\beta < 0$, so that $\bar{V}^\varepsilon \sim \text{LDP}(\varepsilon^{2b}, \Lambda_{FH})$. Finally, the LDP for V^ε follows again by exponential equivalence:

$$\mathbb{P}(\|V^\varepsilon - \bar{V}^\varepsilon\|_\infty > \delta) = \mathbb{P}(\|\varepsilon^b \Theta e^{\beta \cdot}\|_\infty > \delta) \leq \mathbb{P}\left(|\Theta| > \frac{\delta}{\varepsilon^b}\right),$$

for $t, \varepsilon, \delta > 0$, and the theorem follows from Assumption $\mathcal{A}_b^\Theta$.

PROOF OF THEOREM 13(II)

We first prove large deviations for the Gaussian drivers of the process, which we then, by means of iterated Contraction Principles, translate to large deviations for the whole scaled process.

- (1) When $H \neq \frac{1}{2}$, Lemma B.1 in [52] implies that $(\rho B, W^H)$ is Gaussian, so that $\varepsilon^b(\rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2b} and good rate function

$$I_1(\varphi, \psi) := \begin{cases} \frac{1}{2} \int_{\mathcal{T}} f(s)^2 ds, & (\varphi, \psi) \in \bar{\mathcal{H}}_H, \\ +\infty, & (\varphi, \psi) \notin \bar{\mathcal{H}}_H, \end{cases} \quad (1.18)$$

where $\bar{\mathcal{H}}_H$ is the RKHS for $(\rho B, W^H)$ defined as

$$\begin{aligned} \bar{\mathcal{H}}_H := & \left\{ (\varphi, \psi) \in \mathcal{C}^2 : \varphi(\cdot) = \rho \int_0^\cdot f(s) ds \right. \\ & \left. \text{and } \psi(\cdot) = \int_0^\cdot K^H(\cdot, s) f(s) ds, \text{ for some } f \in L^2(\mathcal{T}) \right\}. \end{aligned}$$

By independence of $B^\perp$ with respect to B and W^H and using the Contraction Principle, $\varepsilon^b(\bar{\rho}B^\perp + \rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2b} and good rate function

$$I_3(\varphi, \psi) := \inf_{\varphi=u+w} \{I_1(u, \psi) + I_2(w)\} = \inf_{u \in \mathcal{H}} \{I_1(u, \psi) + I_2(\varphi - u)\},$$

where, with $\mathcal{H}_{\bar{\rho}}$ as in (3),

$$I_2(w) := \begin{cases} \frac{\bar{\rho}}{2} \int_0^1 f(s)^2 ds, & w \in \mathcal{H}_{\bar{\rho}}, \\ +\infty, & w \notin \mathcal{H}_{\bar{\rho}}. \end{cases}$$

- (2) For $H = \frac{1}{2}$, using the proof of Theorem 18, $\varepsilon^b(\bar{\rho}B^\perp + \rho B, B)'$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, when ε tends to zero with speed ε^{2b} and good rate function Λ^ρ defined in (1.7).

We now introduce the process $\tilde{X}^\varepsilon$ satisfying the following SDE $d\tilde{X}_t^\varepsilon = -\frac{1}{2}\tilde{\sigma}^2(V_t^\varepsilon)dt + \varepsilon^b\tilde{\sigma}(V_t^\varepsilon)(\rho dB_t + \bar{\rho}dB_t^\perp)$, and translate this large deviations into one for the sequence $\tilde{X}^\varepsilon$. Since the proof for the $H \neq \frac{1}{2}$ case is the same as the one for $H = \frac{1}{2}$, albeit with a different rate function, we concentrate on the former case. Since $V_t^\varepsilon = \varepsilon^b e^{\beta t} \Theta - \frac{\lambda \varepsilon^b}{\beta}(1 - e^{\beta t}) + \varepsilon^b \xi \int_0^t e^{\beta(t-u)} dW_u^H$, one can define a continuous map $G((0, 1) \times \mathcal{C} \rightarrow \mathcal{C})$, such that $(\varepsilon^b(\rho B + \bar{\rho}B^\perp), V^\varepsilon) = (\varepsilon^b(\rho B + \bar{\rho}B^\perp), G(\varepsilon, \varepsilon^b W^H)(t))$. Moreover, using $\mathcal{A}_b^\Theta$, $G(\varepsilon, \varepsilon^b W^H)$ is exponentially equivalent to $\bar{G}(\varepsilon^b W^H)$, with $\bar{G}(\varphi)(t) := \xi \int_0^t e^{\beta(t-u)} d\varphi_u$. The Contraction Principle thus yields $(\varepsilon^b(\rho B + \bar{\rho}B^\perp), V^\varepsilon) \sim \text{LDP}(\varepsilon^{2b}, I_4)$, with $I_4 : (\varphi, \psi) \mapsto \inf\{I_3(\varphi, v) \mid \psi \in \mathcal{C} : \psi = \bar{G}(v)\}$.

Under $\mathcal{A}_b'$, one can apply the extended Contraction Principle proved in [93, Proposition 2.3], so that $(\varepsilon^b(\rho B + \bar{\rho}B^\perp), \tilde{\sigma}(V^\varepsilon)) \sim \text{LDP}(\varepsilon^{2b}, I_5)$ with

$$I_5(\varphi, \psi) := \inf\{I_4(\varphi, v) \mid \psi = \tilde{\sigma}(v)\} = \inf\{I_3(\varphi, v) \mid \psi = \tilde{\sigma}(\bar{G}(v))\}.$$

Finally, as $b > 0$, the sequence of semi-martingales $(\varepsilon^b W)$ is uniformly exponentially tight (UET) in the sense of [62, Definition 1.1], and the sequence $(\tilde{\sigma}(V^\varepsilon))$ is càdlàg (Assumption $\mathcal{A}_b$), and adapted to the filtration $\mathcal{F}$. Denoting $X \cdot V := \int X dV$ the stochastic integral with respect to a

semi-martingale, Theorem 1.2 in [62] yields a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^3)$ for $(\varepsilon^b(\rho B_s + \bar{\rho} B_s^\perp), \tilde{\sigma}(V^\varepsilon), \varepsilon^b \int_0^\cdot \tilde{\sigma}(V_s^\varepsilon)(\rho dB_s + \bar{\rho} dB_s^\perp))$, with speed ε^{2b} and good rate function I_6 defined as $I_6(\phi) := I_5(\varphi, \psi)$ if $\phi = \varphi \cdot \psi$ and $\psi \in \text{BV}$ (the space of functions of finite variation), and infinity otherwise. Applying another Contraction Principle, since, for $t \in \mathcal{T}^*$,

$$\tilde{X}_t^\varepsilon = -\frac{1}{2} \int_0^t \tilde{\sigma}(V_s^\varepsilon)^2 ds + \int_0^t \varepsilon^b \tilde{\sigma}(V_s^\varepsilon) dW_s =: I\left(\tilde{\sigma}(V^\varepsilon), \tilde{\sigma}(V^\varepsilon) \cdot \varepsilon^b (\rho B_s + \bar{\rho} B_s^\perp)\right)(\varepsilon, t).$$

Hence $\tilde{X}^\varepsilon \sim \text{LDP}(\varepsilon^{2b}, \tilde{\Lambda})$ with

$$\tilde{\Lambda}(\phi) := \inf\{I_6(\chi) \mid \phi = I(\varphi, \varphi \cdot \psi), \chi = (\varphi, \psi), \psi \in \text{BV}\}. \quad (1.19)$$

The last step is proving that the processes X^ε and $\tilde{X}^\varepsilon$ are exponentially equivalent. Indeed, for $t \in \mathcal{T}^*$,

$$\begin{aligned} \|X^\varepsilon - \tilde{X}^\varepsilon\|_\infty &\leq \frac{1}{2} \left\| \int_0^\cdot \left| \varepsilon^{2b} \sigma^2\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon) \right| ds \right\|_\infty + \\ &\quad + \varepsilon^b \left\| \int_0^\cdot \left| \varepsilon^b \sigma\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon) \right| d(\rho B_s + \bar{\rho} B_s^\perp) \right\|_\infty. \end{aligned}$$

Using the linear growth assumption as well as Assumption $\mathcal{A}_b$, we have for $s \in [0, t]$,

$$\begin{aligned} \left| \varepsilon^{2b} \sigma^2\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon) \right| &\leq 2\varepsilon^{2b} C^2 \left(1 + \left| \frac{V_s^\varepsilon}{\varepsilon^b} \right|^2 \right) + 2C^2 (1 + |V_s^\varepsilon|^2), \\ &\leq 2C^2 (1 + \varepsilon^{2b} + 2|V_s^\varepsilon|^2), \end{aligned}$$

hence

$$\begin{aligned} \left\| \int_0^\cdot \left| \varepsilon^{2b} \sigma^2\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon) \right| ds \right\|_\infty &\leq \left\| \int_0^\cdot 2C^2 (1 + \varepsilon^{2b} + 2|V_s^\varepsilon|^2) ds \right\|_\infty, \\ &\leq 2C^2(1 + \varepsilon^{2b}) + 4C^2 \left\| \int_0^\cdot |V_s^\varepsilon|^2 ds \right\|_\infty. \end{aligned}$$

Since $V_s^\varepsilon = \varepsilon^b e^{\beta s} \Theta + \varepsilon^b \frac{\lambda}{\beta} (e^{\beta s} - 1) + \varepsilon^b \xi [W_s^H + \beta \int_0^s W_u^H e^{-\beta u} du]$, we obtain

for $s \in \mathcal{T}$, recalling that $\beta < 0$,

$$|V_s^\varepsilon| \leq \varepsilon^b e^{\beta s} |\Theta| + \varepsilon^b \frac{\lambda}{\beta} (e^{\beta s} - 1) + \xi \varepsilon^b |W_s^H|,$$

and

$$|V_s^\varepsilon|^2 \leq 2 \left(\varepsilon^{2b} e^{2\beta s} \Theta^2 + 2\varepsilon^{2b} \left(\frac{\lambda}{\beta} \right)^2 (e^{2\beta s} + 1) + \xi^2 \varepsilon^{2b} |W_s^H|^2 \right), \quad (1.20)$$

so that $\int_0^t |V_s^\varepsilon|^2 ds \leq 2 \left(\varepsilon^{2b} t \Theta^2 + \varepsilon^{2b} \left(\frac{\lambda}{\beta} \right)^2 \left(2t - \frac{1}{\beta} \right) + \xi^2 \varepsilon^{2b} \int_0^t |W_s^H|^2 ds \right)$.

Introducing $\mathcal{J} := \left\| \int_0^\cdot |V_s^\varepsilon|^2 ds \right\|_\infty$, we obtain for any $\delta > 0$ and $a_0 > 0$,

$$\begin{aligned} & \mathbb{P} \left(\left\| X^\varepsilon - \tilde{X}^\varepsilon \right\|_\infty > \delta \right) \\ & \leq \mathbb{P} \left(2C^2 \mathcal{J} + \varepsilon^b \left\| \int_0^\cdot \left(\varepsilon^b \sigma \left(\frac{V_s^\varepsilon}{\varepsilon^b} \right) - \tilde{\sigma}(V_s^\varepsilon) \right) d(\rho B_s + \bar{\rho} B_s^\perp) \right\|_\infty > \delta_\varepsilon \right), \\ & \leq \mathbb{P} \left(2C^2 \mathcal{J} + \varepsilon^b \left\| \int_0^\cdot \left(\varepsilon^b \sigma \left(\frac{V_s^\varepsilon}{\varepsilon^b} \right) - \tilde{\sigma}(V_s^\varepsilon) \right) d(\rho B_s + \bar{\rho} B_s^\perp) \right\|_\infty > \delta_\varepsilon \mid 2C^2 \mathcal{J} < a_0 \right) (1 - \mathfrak{P}) \\ & + \mathbb{P} \left(2C^2 \mathcal{J} + \varepsilon^b \left\| \int_0^\cdot \left(\varepsilon^b \sigma \left(\frac{V_s^\varepsilon}{\varepsilon^b} \right) - \tilde{\sigma}(V_s^\varepsilon) \right) d(\rho B_s + \bar{\rho} B_s^\perp) \right\|_\infty > \delta_\varepsilon \mid 2C^2 \mathcal{J} \geq a_0 \right) \mathfrak{P}, \\ & \leq \mathbb{P} \left(\left\| \int_0^\cdot \left(\varepsilon^b \sigma \left(\frac{V_s^\varepsilon}{\varepsilon^b} \right) - \tilde{\sigma}(V_s^\varepsilon) \right) d(\rho B_s + \bar{\rho} B_s^\perp) \right\|_\infty > \frac{\delta_\varepsilon - a_0}{\varepsilon^b} \mid 2C^2 \mathcal{J} < a_0 \right) + \mathfrak{P}, \end{aligned}$$

with $\mathfrak{P} := \mathbb{P}(2C^2 \mathcal{J} \geq a_0)$ and $\delta_\varepsilon := \delta - C^2(1 + \varepsilon^{2b})$. Using (1.20) and

introducing $\mathcal{J}_t := \int_0^t |W_s^H|^2 ds$, we obtain

$$\begin{aligned}
\mathbb{P}(\mathcal{J} > \bar{\delta}) &\leq \mathbb{P}\left(\left|\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left(2 - \frac{1}{\beta}\right) + \xi^2\varepsilon^{2b}\mathcal{J}_1\right| > \frac{\bar{\delta}}{2}\right), \\
&\leq \mathbb{P}\left(\left|\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| + \xi^2\varepsilon^{2b}\mathcal{J}_1\right| > \frac{\bar{\delta}}{2} \middle| \xi^2\varepsilon^{2b}\mathcal{J}_1 < a\right) \times \mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a) \\
&\quad + \mathbb{P}\left(\left|\varepsilon^{2b}\Theta^2 + \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| + \xi^2\varepsilon^{2b}\mathcal{J}_1\right| > \frac{\bar{\delta}}{2} \middle| \xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a\right) \times \mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a), \\
&\leq \mathbb{P}\left(\Theta^2 > \frac{1}{\varepsilon^{2b}}\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| - a\right) \middle| \xi^2\varepsilon^{2b}\mathcal{J}_1 < a\right) + \mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a), \\
&\leq \mathbb{P}\left(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| - a\right)} \middle| \xi^2\varepsilon^{2b}\mathcal{J}_1 < a\right) + \mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a),
\end{aligned}$$

for any $\bar{\delta} > 0$ and $a > 0$. One can then use Markov's inequality:

$$\mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a) \leq \frac{\mathbb{E}[\xi^2\varepsilon^{2b}\mathcal{J}_1]}{a} = \frac{\xi^2\varepsilon^{2b}}{a}\mathbb{E}\left(\int_0^1 |W_s^H|^2 ds\right).$$

Since $\mathbb{V}(W_s^H) = s^{2H}$, one obtains $\mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 \geq a) \leq \frac{\xi^2\varepsilon^{2b}}{a} \frac{1}{2H+1}$. Hence there exist $\eta \in [0, 1)$ and $\bar{\varepsilon} > 0$ such that for all $\varepsilon \leq \bar{\varepsilon}$, $\mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a) \geq 1 - \eta$.

Assumption $\mathcal{A}_b^\Theta$ then implies that there exist $M, \tilde{\varepsilon} > 0$ such that

$$\mathbb{P}\left(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| - a\right)}\right) \leq \exp\left(-\frac{M}{\varepsilon^{2b}}\right), \quad \text{for all } \varepsilon \leq \tilde{\varepsilon}.$$

Thus, for all $\varepsilon \leq \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$, Bayes' Theorem yields

$$\begin{aligned}
&\mathbb{P}\left(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| - a\right)} \middle| \xi^2\varepsilon^{2b}\mathcal{J}_1 < a\right) \\
&\leq \frac{\mathbb{P}\left(|\Theta| > \frac{1}{\varepsilon^b}\sqrt{\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b}\left(\frac{\lambda}{\beta}\right)^2\left|2 - \frac{1}{\beta}\right| - a\right)}\right)}{\mathbb{P}(\xi^2\varepsilon^{2b}\mathcal{J}_1 < a)} \leq \frac{\exp\left(-\frac{M}{\varepsilon^{2b}}\right)}{1 - \eta}.
\end{aligned}$$

Finally, for all $\varepsilon \leq \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$,

$$\begin{aligned} \mathbb{P}(\mathcal{J} > \bar{\delta}) &\leq \mathbb{P}\left(|\Theta| > \frac{1}{\varepsilon^b} \sqrt{\left(\frac{\bar{\delta}}{2} - \varepsilon^{2b} \left(\frac{\lambda}{\beta}\right)^2 \right) \left|2 - \frac{1}{\beta}\right| - a}\right) \mathbb{P}(\xi^2 \varepsilon^{2b} \mathcal{J}_1 < a) + \mathbb{P}(\xi^2 \varepsilon^{2b} \mathcal{J}_1 \geq a), \\ &\leq \frac{e^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2 \varepsilon^{2b}}{a} \frac{1}{2H_+}, \end{aligned} \quad (1.21)$$

so that for all $\varepsilon \leq \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$, $\mathbb{P}(2C^2 \mathcal{J} \geq a_0) \leq \frac{e^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2 \varepsilon^{2b}}{a} \frac{1}{2H_+}$.

Hence, there exists $\check{\eta} \in [0, 1)$ and $\check{\varepsilon} < \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$ such that for all $\varepsilon \leq \check{\varepsilon}$, $\mathbb{P}(2C^2 \mathcal{J} < a_0) \geq 1 - \check{\eta}$. Moreover, using the uniform convergence assumed in $\mathcal{A}_b$, for all $y \in \mathbb{R}$ there exists $N > 0$ and $\varepsilon_0 > 0$ such that, for all $\varepsilon < \varepsilon_0$, we have that $\left|\varepsilon^b \sigma\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon)\right| \leq N$. Hence one can then apply the Borell-TIS inequality (in particular [105][Proposition A.1]), and obtain,

$$\begin{aligned} \mathbb{P}\left(\left\|\int_0^\cdot \left(\varepsilon^b \sigma\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon)\right) d(\rho B_s + \bar{\rho} B_s^\perp)\right\|_\infty > \frac{1}{\varepsilon^b} (\delta - C^2(1 + \varepsilon^{2b}) - a_0)\right) \\ \leq 2 \exp\left\{-\frac{(\delta - C^2(1 + \varepsilon^{2b}) - a_0)^2}{2N^2 \varepsilon^{2b}}\right\}. \end{aligned}$$

Thus, for all $\varepsilon \leq \min\{\check{\varepsilon}, \varepsilon_0\}$, Bayes' Theorem yields

$$\begin{aligned} \mathbb{P}\left(\left\|\int_0^\cdot \left(\varepsilon^b \sigma\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon)\right) d(\rho B_s + \bar{\rho} B_s^\perp)\right\|_\infty > \frac{1}{\varepsilon^b} (\delta - C^2(1 + \varepsilon^{2b}) - a_0) \middle| 2C^2 \mathcal{J} < a_0\right) \\ \leq \frac{\mathbb{P}\left(\left\|\int_0^\cdot \left(\varepsilon^b \sigma\left(\frac{V_s^\varepsilon}{\varepsilon^b}\right) - \tilde{\sigma}(V_s^\varepsilon)\right) d(\rho B_s + \bar{\rho} B_s^\perp)\right\|_\infty > \frac{1}{\varepsilon^b} (\delta - C^2(1 + \varepsilon^{2b}) - a_0)\right)}{\mathbb{P}(2C^2 \mathcal{J} < a_0)}, \\ \leq \frac{2 \exp\left\{-\frac{(\delta - C^2(1 + \varepsilon^{2b}) - a_0)^2}{2N^2 \varepsilon^{2b}}\right\}}{1 - \check{\eta}}. \end{aligned}$$

Finally, for all $\varepsilon \leq \min\{\check{\varepsilon}, \varepsilon_0\}$,

$$\mathbb{P}\left(\left\|X^\varepsilon - \tilde{X}^\varepsilon\right\|_\infty > \delta\right) \leq \frac{2 \exp\left\{-\frac{(\delta - C^2(1 + \varepsilon^{2b}) - a_0)^2}{2N^2 \varepsilon^{2b}}\right\}}{1 - \check{\eta}} + \frac{e^{-M/\varepsilon^{2b}}}{1-\eta} + \frac{\xi^2 \varepsilon^{2b}}{a} \frac{1}{2H_+},$$

so that X^ε and $\tilde{X}^\varepsilon$ are exponentially equivalent and $X^\varepsilon \sim \text{LDP}(\varepsilon^{2b}, \tilde{\Lambda})$.

Remark 35. *For computational purposes, such infinite-dimensional optimisation problems are discretised over a set of basis functions, such as orthonormal polynomials. If we then consider φ continuously differentiable and I_1 finite, we can simplify $I_1(\varphi) = I_1(\varphi, \psi) = \frac{1}{2\rho^2} \int_{\mathcal{T}} [\dot{\varphi}(t)]^2 dt$. Further, if $u \in \mathcal{C}^1$, then $I_3(\varphi) \simeq \frac{1}{2} \inf_{u \in \mathcal{H} \cap \mathcal{C}^1} \left\{ \int_0^t \left(\frac{\dot{u}_s}{\rho} \right)^2 ds + \frac{1}{\rho} \int_0^t (\dot{\varphi}_s - \dot{u}_s)^2 ds \right\}$, and $I_4(\varphi) = I_3(\varphi)$.*

1.4.2 PROOF OF THEOREM 19

PROOF OF THEOREM 19(I)

Let us introduce the process $\bar{V}^\varepsilon$ defined, for $t \in \mathcal{T}^*$, as the unique solution of $d\bar{V}_t^\varepsilon = (\varepsilon^{b+2}\lambda + \beta\varepsilon^2\bar{V}_t^\varepsilon)d\tau + \varepsilon^{2H+b}\xi dW_t^H$, with $\bar{V}_0^\varepsilon = 0$, i.e.

$$\bar{V}_t^\varepsilon = -\frac{\lambda\varepsilon^b}{\beta} \left(1 - e^{\beta\varepsilon^2 t}\right) + \xi\varepsilon^{2H+b} \int_0^t e^{\beta\varepsilon^2(t-u)} dW_u^H.$$

From Lemma 14, Proposition 20 and Theorem 3.4.5 in [42], the sequence $(\varepsilon^{2H+b} \int_0^\cdot G_0^H(\cdot, s) dZ_s)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}$, with speed ε^{4H+2b} and good rate function $\Lambda_{G_0^H}$ as in (6). Moreover,

- (1) When $H \neq \frac{1}{2}$, $(\varepsilon^{2H+b} \int_0^t G_\varepsilon^H(t, s) dZ_s)_{\varepsilon>0}$ and $(\varepsilon^{2H+b} \int_0^t G_0^H(t, s) dZ_s)_{\varepsilon>0}$ are exponentially equivalent: indeed the asymptotic expansion of the Gaussian density [1, Formula (26.2.12)] yields, for any $\delta > 0$,

$$\begin{aligned} & \mathbb{P} \left(\left\| \varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s) dZ_s - \varepsilon^{2H+b} \int_0^\cdot G_0^H(\cdot, s) dZ_s \right\|_\infty > \delta \right), \\ &= \mathbb{P} \left(\left\| \int_0^\cdot (G_\varepsilon^H(\cdot, s) - G_0^H(\cdot, s)) dZ_s \right\|_\infty > \frac{\delta}{\varepsilon^{2H+b}} \right), \\ &= \mathbb{P} \left(|\mathcal{N}(0, 1)| > \frac{\delta}{\|\mathbb{V}_\varepsilon\|_\infty \varepsilon^{2H+b}} \right), \\ &= \sqrt{\frac{2}{\pi}} \frac{\|\mathbb{V}_\varepsilon\|_\infty \varepsilon^{2H+b}}{\delta} \exp \left(-\frac{\delta^2}{2\varepsilon^{4H+2b} \|\mathbb{V}_\varepsilon\|_\infty^2} \right) (1 + \mathcal{O}(\varepsilon^{4H+2b})), \end{aligned}$$

with $\mathbb{V}_\varepsilon^2 := \int_0^t (G_\varepsilon^H(t, s) - G_0^H(t, s))^2 ds$ and note that $\lim_{\varepsilon \downarrow 0} \mathbb{V}_\varepsilon^2 = 0$. Now, for $\varepsilon > 0$ and $0 < s < t$,

(i) if $H < \frac{1}{2}$,

$$\begin{aligned} [G_\varepsilon^H(t, s) - G_0^H(t, s)]^2 &\leq \frac{2\kappa_H^2 \xi^2}{s^{2H_-}} H_-^2 \left(\int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}} (e^{\beta \varepsilon^2(t-u)} - 1) du \right)^2, \\ &\quad + \frac{2\kappa_H^2 \xi^2}{s^{2H_-}} \beta^2 \varepsilon^4 \left(\int_s^t (u-s)^{H_-} u^{H_-} e^{\beta \varepsilon^2(t-u)} du \right)^2, \\ &\leq 2\kappa_H^2 \xi^2 \left[\frac{H_-^2}{s^{2H_-}} \left(\int_s^t \frac{(u-s)^{H_-}}{u^{1-H_-}} du \right)^2 \right], \\ &\quad + 2\kappa_H^2 \xi^2 \left[\frac{\beta^2 \varepsilon^4}{s^{2H_-}} \left(\int_s^t (u(u-s))^{H_-} du \right)^2 \right]. \end{aligned}$$

Using [113, Lemma A.3], we further obtain, for $\varepsilon > 0$ and $0 < s < t$, $\mathbb{V}_\varepsilon^2 \leq 2\kappa_H^2 \xi^2 [C_H t^{2H} + \beta^2 \varepsilon^4 \tilde{C}_H t^{2H+2}]$, with $C_H, \tilde{C}_H > 0$, two constants depending on H . Hence,

$$0 < \varepsilon^{4H+2b} \mathbb{V}_\varepsilon^2 \leq 2\kappa_H^2 \xi^2 \left(C_H \varepsilon^{4H+2b} t^{2H} + \beta^2 \varepsilon^{4H+2b+4} \tilde{C}_H t^{2H+2} \right);$$

(ii) if $H > \frac{1}{2}$,

$$\begin{aligned} [G_\varepsilon^H(t, s) - G_0^H(t, s)]^2 &= \frac{\kappa_H^2 \xi^2}{s^{2H_-}} H_-^2 \left(\int_s^t \frac{u^{H_-}}{(u-s)^{1-H_-}} (e^{\beta \varepsilon^2(t-u)} - 1) du \right)^2, \\ &\leq \frac{\kappa_H^2 \xi^2}{s^{2H_-}} H_-^2 \left(\int_s^t \frac{u^{H_-} du}{(u-s)^{1-H_-}} \right)^2, \\ &\leq \frac{\kappa_H^2 \xi^2}{s^{2H_-}} H_-^2 t^{2H_-} \left(\int_s^t (u-s)^{H_- - 1} du \right)^2, \\ &= \frac{\kappa_H^2 \xi^2}{s^{2H_-}} [t(t-s)]^{2H_-}. \end{aligned}$$

Hence, for $H \neq \frac{1}{2}$, as $b > 0$, $\lim_{\varepsilon \downarrow 0} \varepsilon^{4H+2b} \|\mathbb{V}_\varepsilon\|_\infty^2 = 0$ as well as $\lim_{\varepsilon \downarrow 0} \varepsilon^{2H+b} \|\mathbb{V}_\varepsilon\|_\infty = 0$.

(2) When $H = \frac{1}{2}$, $(\xi \varepsilon^{1+b} \int_0^t e^{\beta \varepsilon^2(t-s)} dB_s)_{\varepsilon > 0}$ and $(\xi \varepsilon^{1+b} B_t)_{\varepsilon > 0}$ are exponentially equivalent: the asymptotic expansion of the Gaussian density

near infinity [1, Formula (26.2.12)] yields, for any $\delta > 0$,

$$\begin{aligned}
& \mathbb{P} \left(\left\| \xi \int_0^\cdot e^{\beta \varepsilon^2(\cdot-s)} dB_s - \xi \int_0^\cdot dB_s \right\|_\infty > \frac{\delta}{\varepsilon^{1+b}} \right) \\
&= \mathbb{P} \left(\left\| \int_0^\cdot \xi (e^{\beta \varepsilon^2(\cdot-s)} - 1) dB_s \right\|_\infty > \frac{\delta}{\varepsilon^{1+b}} \right), \\
&= \mathbb{P} \left(|\mathcal{N}(0, 1)| > \frac{\delta}{\|\mathbb{V}_\varepsilon\|_\infty \varepsilon^{1+b} \xi} \right), \\
&= \sqrt{\frac{2}{\pi}} \frac{\|\mathbb{V}_\varepsilon\|_\infty \varepsilon^{1+b} \xi}{\delta} \exp \left(-\frac{\delta^2}{2\varepsilon^{2+2b} \|\mathbb{V}_\varepsilon\|_\infty^2 \xi^2} \right) (1 + \mathcal{O}(\varepsilon^{2+2b})),
\end{aligned}$$

with $\mathbb{V}_\varepsilon^2 := \int_0^t [e^{\beta \varepsilon^2(t-s)} - 1]^2 ds = t + \frac{e^{2\beta \varepsilon^2 t} - 1}{2\beta \varepsilon^2} + \frac{2(1 - e^{\beta \varepsilon^2 t})}{\beta \varepsilon^2}$. Hence $0 < \varepsilon^{2+2b} \mathbb{V}_\varepsilon^2 \leq \varepsilon^{2+2b} t + \varepsilon^{2b} \frac{e^{\beta \varepsilon^2 t} - 1}{2\beta}$, and, as $b > 0$, $\lim_{\varepsilon \downarrow 0} \varepsilon^{2+2b} \|\mathbb{V}_\varepsilon\|_\infty^2 = 0$. Besides, note that $\lim_{\varepsilon \downarrow 0} \mathbb{V}_\varepsilon^2 = 0$. The required exponential equivalence then follows, and hence for all $H \in (0, 1)$,

$$(\varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s) dZ_s)_{\varepsilon > 0} \sim \text{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H}).$$

The final step is exponential equivalence between $\overline{V}^\varepsilon$ and $(\varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s) dZ_s)_{\varepsilon > 0}$. For any $\delta, t > 0$,

$$\begin{aligned}
\mathbb{P} \left(\left\| \overline{V}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s) dZ_s \right\|_\infty > \delta \right) &= \mathbb{P} \left(\left\| \overline{V}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot e^{\beta \varepsilon^2(\cdot-s)} dW_s^H \right\|_\infty > \delta \right), \\
&= \mathbb{P} \left(\left\| -\frac{\lambda \varepsilon^b}{\beta} (1 - e^{\beta \varepsilon^2 \cdot}) \right\|_\infty > \delta \right).
\end{aligned}$$

Hence $\limsup_{\varepsilon \downarrow 0} \varepsilon^{4H+2b} \log \mathbb{P}(\|\overline{V}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot G_\varepsilon^H(\cdot, s) dZ_s\|_\infty > \delta) = -\infty$, yielding exponential equivalence and $\overline{V}^\varepsilon \sim \text{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$. Finally, since for any $\delta > 0$,

$$\mathbb{P}(\|V^\varepsilon - \overline{V}^\varepsilon\|_\infty > \delta) \leq \mathbb{P}(\varepsilon^b |\Theta| > \delta),$$

as $\beta < 0$, and $\overline{V}^\varepsilon$ and V^ε are exponentially equivalent, and $V^\varepsilon \sim \text{LDP}(\varepsilon^{4H+2b}, \Lambda_{G_0^H})$ by Assumption $\mathcal{A}_b^\Theta$.

PROOF OF THEOREM 19 (II)

The idea of the proof is the same as the proof of Theorem 13(ii).

- (1) $H \neq \frac{1}{2}$: we already established that $\varepsilon^{2H+b}(\bar{\rho}B^\perp + \rho B, W^H)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{4H+2b} and good rate function I_3 defined earlier.
- (2) $H = \frac{1}{2}$: using the proof of Theorem 18, $\varepsilon^{b+1}(\bar{\rho}B^\perp + \rho B, B)'$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, with speed ε^{2+2b} and good rate function Λ^ρ defined in (1.7).

Similar to above, we only prove the case $H \neq \frac{1}{2}$, the other one being analogous. We introduce the process $\tilde{X}^\varepsilon$ satisfying the following SDE $d\tilde{X}_t^\varepsilon = -\frac{\varepsilon^{2H+1}}{2}\tilde{\sigma}^2(V_t^\varepsilon)dt + \varepsilon^{2H+b}\tilde{\sigma}(V_t^\varepsilon)(\rho dB_t + \bar{\rho}dB_t^\perp)$, and translate this large deviations into one for the sequence $\tilde{X}^\varepsilon$.

Using that for $\varepsilon > 0$ and $t \in \mathcal{T}^*$, $V_t^\varepsilon = \varepsilon^b e^{\beta\varepsilon^2 t} \Theta - \frac{\lambda\varepsilon^b}{\beta}(1 - e^{\beta\varepsilon^2 t}) + \varepsilon^{2H+b}\xi \int_0^t e^{\beta\varepsilon^2(t-u)} dW_u^H$, one can define a continuous map $\tilde{G}((0, 1) \times \mathcal{C} \rightarrow \mathcal{C})$, such that $(\varepsilon^{2H+b}(\rho B + \bar{\rho}B^\perp), V^\varepsilon) = (\varepsilon^{2H+b}(\rho B + \bar{\rho}B^\perp), \tilde{G}(\varepsilon, \varepsilon^{2H+b}W^H)(t))$. Using $\mathcal{A}_b^\Theta$ and the asymptotic expansion of the Gaussian density near infinity [1, Formula (26.2.12)], $\tilde{G}(\varepsilon, \varepsilon^{2H+b}W^H)$ is exponentially equivalent to $\hat{G}_0(\varepsilon^{2H+b}W^H)$, defined by $\hat{G}_0(\varphi)(t) := \xi \int_0^t d\varphi_u$.

Hence, the Contraction Principle yields an LDP on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ for $(\varepsilon^{2H+b}(\rho B + \bar{\rho}B^\perp), V^\varepsilon)$, with speed ε^{4H+2b} and good rate function $\tilde{I}_4 : (\varphi, \psi) \mapsto \inf\{I_3(\varphi, v) \mid \psi \in \mathcal{C} : \psi = \hat{G}_0(v)\}$. Under $\mathcal{A}_b'$, the extended Contraction Principle [93, Proposition 2.3] yields an LDP on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ for $(\varepsilon^{2H+b}(\rho B + \bar{\rho}B^\perp), \tilde{\sigma}(V^\varepsilon))$, with speed ε^{4H+2b} and good rate function $\tilde{I}_5 : (\varphi, \psi) \mapsto \inf\{\tilde{I}_4(\varphi, v) \mid \psi = \tilde{\sigma}(v)\} = \inf\{I_3(\varphi, v) \mid \psi = \tilde{\sigma}(\hat{G}_0(v))\}$. Finally, for any $b > 0$, the sequence of semimartingales $(\varepsilon^{2H+b}W)$ is UET and the sequence $(\tilde{\sigma}(V^\varepsilon))$ is càdlàg (Assumption $\mathcal{A}_b$), and adapted to the filtration $\mathcal{F}$. Theorem 1.2 in [62] thus gives an LDP on $\mathcal{C}$ for the sequence of stochastic integrals $\varepsilon^{2H+b} \int_0^\cdot \tilde{\sigma}(V_s^\varepsilon)(\rho dB_s + \bar{\rho}dB_s^\perp)$, with speed ε^{4H+2b} and good rate function

$$\mathbb{I}(\chi) := \inf \left\{ \tilde{I}_5(\varphi, \psi) : \varphi \cdot \psi = \chi, \psi \in \text{BV} \right\}. \quad (1.22)$$

We now prove that $\tilde{X}^\varepsilon$ and $\varepsilon^{2H+b} \int_0^\cdot \tilde{\sigma}(V_s^\varepsilon)(\rho dB_s + \bar{\rho} dB_s^\perp)$ are exponentially equivalent.

$$\begin{aligned} \mathbb{P} \left(\left\| \tilde{X}^\varepsilon - \varepsilon^{2H+b} \int_0^\cdot \tilde{\sigma}(V_s^\varepsilon) dW_s \right\|_\infty > \delta \right) &= \mathbb{P} \left(\left\| -\frac{1}{2} \int_0^\cdot (\varepsilon^{H+\frac{1}{2}} \tilde{\sigma}(V_s^\varepsilon))^2 ds \right\|_\infty > \delta \right), \\ &= \mathbb{P} \left(\left\| \int_0^\cdot (\tilde{\sigma}(V_s^\varepsilon))^2 ds \right\|_\infty > \frac{2\delta}{\varepsilon^{2H+1}} \right), \\ &\leq \mathbb{P} \left(\left\| \int_0^\cdot |V_s^\varepsilon|^2 ds \right\|_\infty > \frac{\delta}{C^2 \varepsilon^{2H+1}} - 1 \right), \end{aligned}$$

using the growth condition assumption $|\tilde{\sigma}(x)| \leq C(1 + |x|)$ for $x \in \mathbb{R}$. Since $V_s^\varepsilon = \varepsilon^b e^{\beta \varepsilon^2 s} \Theta + \varepsilon^b \frac{\lambda}{\beta} (e^{\beta \varepsilon^2 s} - 1) + \varepsilon^{2H+b} \xi \left[W_s^H + \beta \varepsilon^2 \int_0^s W_u^H e^{-\beta \varepsilon^2 u} du \right]$ and $\beta < 0$, we obtain for $s \in \mathcal{T}$, $|V_s^\varepsilon| \leq \varepsilon^b e^{\beta \varepsilon^2 s} |\Theta| + \varepsilon^b \frac{\lambda}{\beta} (e^{\beta \varepsilon^2 s} - 1) + \xi \varepsilon^{2H+b} |W_s^H|$, so that

$$|V_s^\varepsilon|^2 \leq 2 \left(\varepsilon^{2b} e^{2\beta \varepsilon^2 s} \Theta^2 + 2\varepsilon^{2b} \left(\frac{\lambda}{\beta} \right)^2 (e^{2\beta \varepsilon^2 s} + 1) + \xi^2 \varepsilon^{4H+2b} |W_s^H|^2 \right), \quad (1.23)$$

and hence $\int_0^t |V_s^\varepsilon|^2 ds \leq \varepsilon^{2b} t \Theta^2 + \varepsilon^{2b} \left(\frac{\lambda}{\beta} \right)^2 \left(2t - \frac{1}{\beta \varepsilon^2} \right) + \xi^2 \varepsilon^{4H+2b} \int_0^t |W_s^H|^2 ds$. Using a similar argument to (1.21), used in the proof of Theorem 13, we have, for all $\varepsilon \leq \min\{\tilde{\varepsilon}, \bar{\varepsilon}\}$,

$$\mathbb{P} \left(\left\| \int_0^\cdot |V_s^\varepsilon|^2 ds \right\|_\infty > \frac{\delta}{C^2 \varepsilon^{2H+1}} - 1 \right) \leq \frac{e^{-M/\varepsilon^{4H+2b}}}{1 - \eta} + \frac{\xi^2 \varepsilon^{4H+2b}}{a} \frac{1}{2H_+},$$

so that $\tilde{X}^\varepsilon$ and $\varepsilon^{2H+b} \int_0^\cdot \tilde{\sigma}(V_s^\varepsilon)(\rho dB_s + \bar{\rho} dB_s^\perp)$ are exponentially equivalent.

Finally, one can use the same arguments that in Section 1.4.1, replace β by $\beta \varepsilon^2$ and the speed ε^{2b} by ε^{4H+2b} and prove that X^ε and $\tilde{X}^\varepsilon$ are exponentially equivalent. Hence $X^\varepsilon \sim \text{LDP}(\varepsilon^{4H+2b}, \mathbb{I})$.

1.4.3 PROOF OF THEOREM 25

The proof is similar to that of [105, Theorem 2.9], with some minor modifications. One key ingredient in the proof of the theorem is an exponential

equivalence between $(X^{\varepsilon, \mathbf{x}})_{\varepsilon > 0}$ and its limit system (as ε tends to zero). In this proof, we denote X^ε by $X^{\varepsilon, \mathbf{x}}$ to stress the dependence on the starting point.

Lemma 36. *Under Assumption 24, for each $R, \delta, \beta > 0$, there exists $\gamma, \rho, \varepsilon_0 > 0$ such that*

$$\varepsilon^2 \log \mathbb{P} (\|X^{\varepsilon, \mathbf{x}} - \mathcal{S}_0(\varphi, u)\| > \delta, \|\varepsilon W - \varphi\| \leq \gamma) \leq -R$$

holds when $\varepsilon \leq \varepsilon_0$ for all $u \in \text{supp}(X_0)$ and all $\varphi \in \mathcal{H}$ satisfying $\Lambda^\rho(\varphi) \leq \beta$, $\mathbf{x} \in \mathcal{B}_\rho((0, 0))$.

Proof. First, introduce the Radon-Nikodym derivative

$$D^\varepsilon(\varphi) := \exp \left\{ \frac{1}{\varepsilon} \int_{\mathcal{T}} \dot{\varphi}_s dW_s - \frac{1}{2\varepsilon^2} \int_{\mathcal{T}} \|\dot{\varphi}_s\|^2 ds \right\}.$$

Following [105], consider the family $(\bar{V}^\varepsilon)_{\varepsilon > 0}$ of (unique strong) solutions of $d\bar{V}_t^\varepsilon = c(\varepsilon, t, \bar{V}_t^\varepsilon, X_0)dt + \varepsilon a(\bar{V}_t^\varepsilon, X_0)dB_t^\varepsilon$, for $t > 0$, with initial condition $\bar{V}_0^\varepsilon = (0, 0)' \in \mathbb{R}^2$. We denote $c(\varepsilon, t, \mathbf{x}, \mathbf{v}) := b(\varepsilon, \mathbf{x}, \mathbf{v}) + a(\mathbf{x}, \mathbf{v})\dot{f}_t$, for $\mathbf{x}, \mathbf{v} \in \mathbb{R}^2$, $t \in \mathcal{T}$ and $\varepsilon > 0$, and $B_t^\varepsilon := W_t - \frac{1}{\varepsilon}\dot{f}_t$, for $t \in \mathcal{T}$. Peithmann's assumptions are here updated as:

- (i) $c : (0, \infty) \times \mathcal{T} \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ converges to $c_0 : \mathcal{T} \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ in the sense that

$$\lim_{\varepsilon \downarrow 0} \int_{\mathcal{T}} \sup_{\mathbf{x} \in \mathbb{R}^2} \|c(\varepsilon, t, \mathbf{x}, \mathbf{v}) - c_0(t, \mathbf{x}, \mathbf{v})\| dt = 0, \quad \text{for all } \mathbf{v} \in \mathbb{R}^2.$$

In particular, note that $c_0(t, \mathbf{x}, \mathbf{v}) = a(\mathbf{x}, \mathbf{v})\dot{f}_t$;

- (ii) there exists $\varrho \in L^2(\mathcal{T})$ such that for all $\mathbf{x}, \mathbf{v} \in \mathbb{R}^2$, $\|c(\varepsilon, t, \mathbf{x}, \mathbf{v})\| + \|c_0(t, \mathbf{x}, \mathbf{v})\| \leq \varrho(t)$, for $t \in \mathcal{T}$;
- (iii) there exists $\kappa \in L^1(\mathcal{T})$ such that for all $\mathbf{v}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$, $\|c_0(t, \mathbf{x}_1, \mathbf{v}) - c_0(t, \mathbf{x}_2, \mathbf{v})\| \leq \kappa(t)\|\mathbf{x}_1 - \mathbf{x}_2\|$ on $\mathcal{T}$.

Partitioning $\mathcal{T}$ into $\{t_k := \frac{kT}{n}\}_{k=0, \dots, n}$, set $\bar{V}_t^{\varepsilon, n} := \bar{V}_{t_k}^\varepsilon$ for $t_k \leq t < t_{k+1}$, and Peithmann's arguments yield

- (i) for any $\delta > 0$, $\lim_{n \uparrow \infty} \limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(\|\bar{V}^\varepsilon - \bar{V}^{\varepsilon,n}\| > \delta) = -\infty$ uniformly with respect to $v_0 \in \mathbb{R}^2$;
- (ii) given $M_t^\varepsilon := \varepsilon \int_{\mathcal{T}} a(\bar{V}_s^\varepsilon, X_0) dW_s$ and $M_t^{\varepsilon,n} := \varepsilon \int_{\mathcal{T}} a(\bar{V}_s^{\varepsilon,n}, X_0) dW_s$, for all $\delta > 0$,

$$\lim_{\beta \downarrow 0} \limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(\|M^\varepsilon - M^{\varepsilon,n}\| > \delta, \|V^\varepsilon - V^{\varepsilon,n}\| \leq \beta) = -\infty,$$

uniformly with respect to $n \in \mathbb{N}$, $y_0 \in \mathbb{R}^2$;

- (iii) for all $\delta > 0$, $\lim_{\gamma \downarrow 0} \limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \mathbb{P}(\|M^\varepsilon\| > \delta, \|\varepsilon W\| \leq \gamma) = -\infty$;
- (iv) let ζ denote the solution of the ordinary differential equation $\dot{\zeta}_t = c_0(t, \zeta_t, V_t)$ starting from $\zeta_0 = v$. For all $R > 0$, $\delta > 0$, there exists $\gamma, \rho, \varepsilon_0 > 0$ such that $\mathbb{P}(\{\|V^\varepsilon, \zeta\| > \delta\} \cap \{\|\varepsilon W\| \leq \gamma\}) \leq \exp(-R/\varepsilon^2)$, for all $v_0 \in \mathbb{R}^2$, $v \in \mathcal{B}_\rho(v_0)$ and $\varepsilon \leq \varepsilon_0$.

The theorem follows from Girsanov's theorem with Radon-Nikodym derivative $D^\varepsilon(f)$. $\square$

We start with the lower bound. For any open subset G of $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, let $\eta > 0$ and choose $\psi \in G$ such that $I^\alpha(\psi) \leq \inf_{\psi \in G} I^\alpha(\psi) + \eta$. Then, let $u \in \text{supp}(X_0)$ and $\varphi \in \mathcal{H}$ such that $\mathcal{S}_0(\varphi, u) = \psi$ and $\Lambda^\rho(\varphi) = I^\alpha(\psi)$. Let $\delta > 0$ such that $\mathcal{B}_\delta(\psi) \subset G$. Then, for each $\gamma > 0$, $x \in \mathbb{R}^2$,

$$\begin{aligned} \mathbb{P}(X^{\varepsilon,x} \in G) &\geq \mathbb{P}(\|X^{\varepsilon,x} - \psi\| \leq \delta), \\ &\geq \mathbb{P}(\|\varepsilon W - \varphi\| \leq \gamma) - \mathbb{P}(\|X^{\varepsilon,x} - \psi\| > \delta, \|\varepsilon W - \varphi\| \leq \gamma). \end{aligned}$$

Schilder's Theorem [39, Theorem 5.2.3], then yields the lower bound

$$\liminf_{\varepsilon \downarrow 0} \varepsilon^2 \mathbb{P}(\|\varepsilon W - \varphi\| \leq \gamma) \geq -\Lambda^\rho(\varphi) = -I^\alpha(\psi) \geq -\inf_{\psi \in G} I^\alpha(\psi) - \eta.$$

Then, we bound the second probability from above using Lemma 36: fix $\beta \geq \Lambda^\rho(\varphi)$ and $R > \inf_{\psi \in G} I^\alpha(\psi) + \eta$, and find $\gamma, \rho, \varepsilon_0 > 0$ such that, for $x \in \mathcal{B}_\rho((0,0)')$, $\varepsilon \leq \varepsilon_0$, the bound $\varepsilon^2 \log \mathbb{P}(\|X^{\varepsilon,x} - \psi\| > \delta, \|\varepsilon W - \varphi\| \leq \gamma) \leq -R$

holds. These two bounds then imply the required lower bound:

$$\begin{aligned} \liminf_{\varepsilon \downarrow 0} \varepsilon^2 \log \inf_{\rho \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}_\rho((0,0)')} \mathbb{P}(X^{\varepsilon, \mathbf{x}} \in G) &\geq \min \left\{ -R, -\inf_{\psi \in G} I^\alpha(\psi) - \eta \right\}, \\ &= -\inf_{\psi \in G} I^\alpha(\psi) - \eta. \end{aligned}$$

We now prove the upper bound. For any closed set F of $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$, take $\beta \in (0, \inf_{\psi \in G} I^\alpha(\psi))$ and $R > \beta$. Let $u \in \text{supp}(X_0)$ and $\psi \in \mathcal{H}$ with $I^\alpha(\psi) \leq \beta$. We find $\delta > 0$ such that $\mathcal{B}_\delta(\psi) \cap F = \emptyset$ and $\varphi \in \{\Lambda^\rho \leq \beta\}$ such that $\mathcal{S}_0(\varphi, u) = \psi$. Using Lemma 36, there exists $\gamma, \rho, \varepsilon_0 > 0$ such that for $\mathbf{x} \in \mathcal{B}_\rho((0,0)'), \varepsilon \leq \varepsilon_0$,

$$\varepsilon^2 \log \mathbb{P}(\|X^{\varepsilon, \mathbf{x}} - \psi\| > \delta, \|\varepsilon W - \varphi\| \leq \gamma) \leq -R.$$

The set $\{\mathcal{B}_\gamma(\varphi) : \psi \in \mathcal{H}, I^\alpha(\psi) \leq \beta\}$ forms a cover of the compact set $\{\Lambda^\rho(\varphi) \leq \beta\}$, so that we can extract a finite sub-cover $\{\mathcal{B}_{\gamma_i}(\varphi_i)\}_{i=1, \dots, k}$ and set $A := \cup_{i=1}^k \mathcal{B}_{\gamma_i}(\varphi_i)$ and $\psi_i := \mathcal{S}_0(\varphi_i, r)$. For any $i = 1, \dots, k$, there exist $\delta_i, \rho_i, \varepsilon_i > 0$ such that, for any $\mathbf{x} \in \mathcal{B}_{\rho_i}((0,0)'), \varepsilon \leq \varepsilon_i$,

$$\varepsilon^2 \log \mathbb{P}(\|X^{\varepsilon, \mathbf{x}} - \psi_i\| > \delta, \|\varepsilon W - \varphi_i\| \leq \gamma) \leq -R.$$

Set $\varepsilon_0 := \min\{\varepsilon_1, \dots, \varepsilon_k\}$, $\rho_0 := \min\{\rho_1, \dots, \rho_k\}$, take $\varepsilon \leq \varepsilon_0$ and $\mathbf{x} \in \mathcal{B}_{\rho_0}((0,0)').$ Since $F \cap \mathcal{B}_{\delta_i}(\psi_i) = \emptyset$ for every $i = 1, \dots, k$, we obtain

$$\begin{aligned} \mathbb{P}(X^{\varepsilon, \mathbf{x}} \in F) &\leq \mathbb{P}(X^{\varepsilon, \mathbf{x}} \in F, \varepsilon W \in A) + \mathbb{P}(\varepsilon W \in A^c) \\ &\leq \sum_{i=1}^k \mathbb{P}(\|X^{\varepsilon, \mathbf{x}} - g_i\| > \delta_i, \|\varepsilon W - \varphi_i\| \leq \gamma_i) + \mathbb{P}(\varepsilon W \in A^c) \\ &\leq k \exp\left(-\frac{R}{\varepsilon^2}\right) + \exp\left(-\frac{\beta}{\varepsilon^2}\right), \end{aligned}$$

since $I^\alpha(\psi_i) \leq \beta$ by definition of β . Finally, since $R > \beta$, the theorem follows from the upper bound

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 \log \sup_{\rho \downarrow 0} \sup_{\mathbf{x} \in \mathcal{B}_\rho(X_0^\varepsilon)} \mathbb{P}(X^{\varepsilon, \mathbf{x}} \in F) \leq -\inf_{\psi \in F} I^\alpha(\psi).$$

2

ASYMPTOTICS FOR VOLATILITY DERIVATIVES IN MULTI-FACTOR ROUGH VOLATILITY MODELS

In this Chapter, we study the small-time implied volatility smile for Realised Variance options, and investigate the effect of correlation in multi-factor models on the linearity of the smile. We also develop an approximation scheme for the Realised Variance density, allowing fast and accurate pricing of Volatility Swaps. Additionally, we establish small-noise asymptotic behaviour of a general class of VIX options in the large strike regime*.

*This Chapter is based on a joint work with Aitor Muguruza and Henry Stone. The theoretical and numerical results presented here have previously appeared in the submitted paper [90].

2.1 INTRODUCTION

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 [13]. Gassiat [63] 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 [82] 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 [14], Forde and Zhang [52], Horvath, Jacquier and Lacombe [76] and most recently Friz, Gassiat and Pigato [57] (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 Bennedsen, Lunde and Pakkanen [17], Horvath, Jacquier and Muguruza [77] and McCrickerd and Pakkanen [94]. Moreover, recent developments by Stone [110] and Horvath, Muguruza and Tomas [79] 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 partial differential equation 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 [47, 48], which allows a better understanding through the fractional PDE derived in [48]. 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 [81] provide algorithms for pricing VIX options and

futures. Horvath, Jacquier and Tankov [78] 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 not been studied until very recently in Alòs, García-Lorite and Muguruza [4].

The main focus of this 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). An additional major contribution of this Chapter 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.

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 [27], Carr and Lee [28] to name a few.

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 [19] two-factor model, Avellaneda and Papanicolaou [6] or Horvath, Jacquier and Tankov [78] 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. In this Chapter, we aim to understand multi-factor models and analyse the effect of factors in implied

volatility smiles, which are of great interest to practitioners in the quantitative finance industry.

For a log stock price process X defined as

$$X_t = -\frac{1}{2} \int_0^t v_s ds + \int_0^t \sqrt{v_s} dW_s, \quad X_0 = 0,$$

where W 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

$$\left(\frac{1}{T} \int_0^T d\langle X \rangle_s - K \right)^+, \quad (2.1)$$

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 (2.1) 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 consistent with that of Alòs, García-Lorite and Muguruza [4], 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 2.1 below for evidence of their linearity.

The structure of the Chapter is as follows. Section 2.2 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 2.3. Section 2.4 is dedicated to numerical examples of the results attained in Section 2.3. In Section 2.5 we introduce a general variance process, which includes the rough Bergomi model for a specific choice of kernel, and briefly

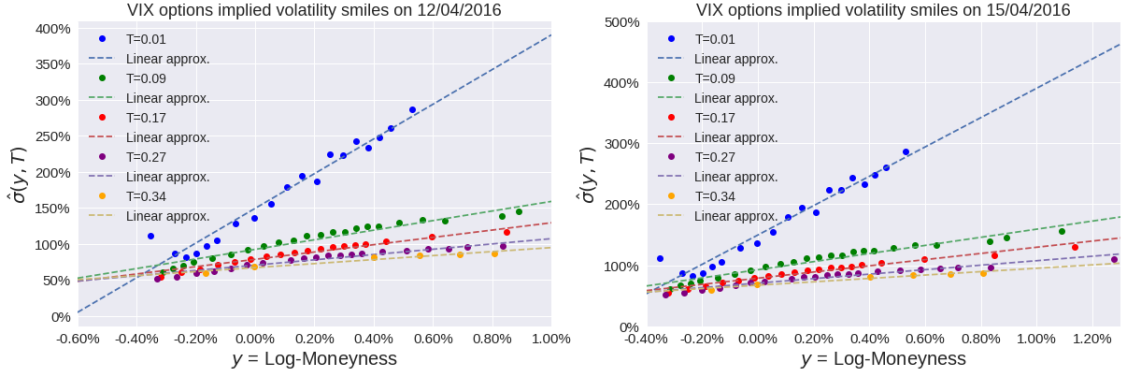


Figure 2.1: Implied volatility smiles for Call options on VIX for small maturities, close to the money. Data provided by OptionMetrics.

investigate the small-noise behaviour of VIX options in this general setting. Motivated by the numerical examples in Section 2.4, we propose a simple and very feasible approximation for the density of the realised variance for the mixed rough Bergomi model (see (2.6)) in Appendix 2.7. The proofs of the main results are given in Appendix 2.9; the details of the numerics are given in Appendix 2.6.

2.2 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

$$Z_t := \int_0^t K_{H_-}(s, t) dB_s, \quad \text{for any } t \in \mathcal{T}, \quad (2.2)$$

where $H_- := H - \frac{1}{2} \in (-\frac{1}{2}, 0)$, B a standard Brownian motion, and where the kernel $K_{H_-} : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ reads

$$K_{H_-}(s, t) := \eta \sqrt{2H}(t-s)^{H_-}, \quad \text{for all } 0 \leq s < t, \quad (2.3)$$

for some strictly positive constant η . Note that, for any $t \geq 0$, the map $s \mapsto K_{H-}(s, t)$ belongs to $L^2(\mathcal{T})$, so that the stochastic integral (2.2) is well defined. We also define an analogous multi-dimensional version of (2.2) by

$$\mathcal{Z}_t := \left(\int_0^t K_{H-}(s, t) dB_s^1, \dots, \int_0^t K_{H-}(s, t) dB_s^m \right) := (\mathcal{Z}_t^1, \dots, \mathcal{Z}_t^m), \quad \text{for } t \in \mathcal{T}, \quad (2.4)$$

where $B^1, \dots, B^m$ are independent Brownian motions.

Model 37 (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 [13]) as*

$$\begin{cases} X_t = -\frac{1}{2} \int_0^t v_s ds + \int_0^t \sqrt{v_s} dW_s, & X_0 = 0, \\ v_t = v_0 \exp \left(Z_t - \frac{\eta^2}{2} t^{2H} \right), & v_0 > 0, \end{cases} \quad (2.5)$$

where the Brownian motion W is defined as $W := \rho B + \sqrt{1 - \rho^2} B^\perp$ for $\rho \in [-1, 1]$ and some standard Brownian motion $B^\perp$ independent of B .

Remark 38. *The process $\log v$ has a modification whose sample paths are almost surely locally γ -Hölder continuous, for all $\gamma \in (0, H)$ [82, Proposition 2.2].*

Model 39 (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*

$$\begin{cases} X_t = -\frac{1}{2} \int_0^t v_s^{(\gamma, \nu)} ds + \int_0^t \sqrt{v_s^{(\gamma, \nu)}} dW_s, & X_0 = 0, \\ v_t^{(\gamma, \nu)} = v_0 \sum_{i=1}^n \gamma_i \exp \left(\frac{\nu_i}{\eta} Z_t - \frac{\nu_i^2}{2} t^{2H-1} \right), & v_0 > 0 \end{cases} \quad (2.6)$$

where $\gamma := (\gamma_1, \dots, \gamma_n) \in [0, 1]^n$ such that $\sum_{i=1}^n \gamma_i = 1$ and $\nu := (\nu_1, \dots, \nu_n) \in \mathbb{R}^n$, such that $0 < \nu_1 < \dots < \nu_n$.

The above modification of the rough Bergomi model, inspired by Bergomi [18], 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 2.4.2.

Model 40 (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*

$$\begin{cases} X_t = -\frac{1}{2} \int_0^t v_s^{(\gamma, \nu, \Sigma)} ds + \int_0^t \sqrt{v_s^{(\gamma, \nu, \Sigma)}} dW_s, & X_0 = 0, \\ v_t^{(\gamma, \nu, \Sigma)} = v_0 \sum_{i=1}^n \gamma_i \mathcal{E} \left(\frac{\nu^i}{\eta} \cdot L_i Z_t \right), & v_0 > 0, \end{cases} \quad (2.7)$$

where $\gamma := (\gamma_1, \dots, \gamma_n) \in [0, 1]^n$ such that $\sum_{i=1}^n \gamma_i = 1$. The vector $\nu^i = (\nu_1^i, \dots, \nu_m^i) \in \mathbb{R}^m$ satisfies $0 < \nu_1^i < \dots < \nu_m^i$ for all $i \in \{1, \dots, n\}$. In addition, $L_i \in \mathbb{R}^{m \times m}$ is a lower triangular matrix such that $L_i L_i^T =: \Sigma_i$ is a positive definite matrix for all $i \in \{1, \dots, n\}$, denoting the covariance matrix.

For all results involving models (2.5), (2.6), and (2.7) we fix $\mathcal{T} = [0, 1]$; minor adjustments to the proofs yield analogous results for more general $\mathcal{T}$. We additionally consider $H \in (0, \frac{1}{2})$.

Remark 41. *In models (2.5)-(2.7), 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 Theorem 10, as long as the mapping is continuous.*

Remark 42. *The reader may already have realised that the mixed multi-factor rough Bergomi defined in (2.7) is indeed general enough to cover both (2.5) and (2.6). However, we provide our theoretical results in an orderly fashion starting from (2.5) and finishing with (2.7), which we find the most natural way to increase the complexity of the model.*

Remark 43. *In place of K_{H_-} in (2.2), one may also consider more general kernels of the form $G(t, s) := K_{H_-}(t, s)L(t - s)$ where $L \in \mathcal{C}(0, \infty)$ such that the stochastic integral is well defined, $L(\cdot)$ is decreasing, hence L is a slowly-varying function at zero, that is $\lim_{x \rightarrow 0} \frac{L(tx)}{L(x)} = 1$ for any $t > 0$.*

2.3 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:

$$(i)(RV(v)(T) - K)^+, \quad (ii)(\sqrt{RV(v)(T)} - K)^+, \quad \text{where } T, K \geq 0. \quad (2.8)$$

where we define the following $\mathcal{C}(\mathcal{T})$ operator

$$RV(f)(\cdot) : f \mapsto \frac{1}{\cdot} \int_0^\cdot f(s)ds, \quad RV(f)(0) := f(0), \quad (2.9)$$

and v represents the instantaneous variance in a given stochastic volatility model. Note that $RV(v)(0) = v_0$.

Remark 44. *As shown by Neuberger [101], we may rewrite the variance swap in terms of the log contract as*

$$\mathbb{E}[RV(v)(T)] = \mathbb{E}\left[\frac{1}{T} \int_0^T v_s ds\right] = \mathbb{E}\left[-2\frac{X_T}{T}\right], \quad (2.10)$$

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 (2.10).

2.3.1 SMALL-TIME RESULTS FOR THE ROUGH BERGOMI MODEL

Proposition 45. *The set $\mathcal{H}_{K_{H_-}} := \{\int_0^\cdot K_{H_-}(s, \cdot)f(s)ds : f \in L^2(\mathcal{T})\}$ defines the reproducing kernel Hilbert space for Z with inner product*

$$\left\langle \int_0^\cdot K_{H_-}(s, \cdot)f_1(s)ds, \int_0^\cdot K_{H_-}(s, \cdot)f_2(s)ds \right\rangle_{\mathcal{H}_{K_{H_-}}} = \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}.$$

Proof. See [82, Theorem 3.1]. $\square$

Before stating Theorem 46, we define the following function $\Lambda^Z : \mathcal{C}(\mathcal{T}) \rightarrow \mathbb{R}_+$ as $\Lambda^Z(\cdot) := \Lambda_{\mathcal{H}_{K_{H-}}}(\cdot)$, with $\Lambda_{\mathcal{H}_{K_{H-}}}$ defined in Equation (6).

Theorem 46. *The variance process $(v_\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, such that $v_\varepsilon \sim \text{LDP}(\varepsilon^{2H}, \Lambda^v(x))$, with $\Lambda^v(x) := \Lambda^Z\left(\log\left(\frac{x}{v_0}\right)\right)$, where $\Lambda^v(v_0) = 0$ and $x \in \mathcal{C}(\mathcal{T})$.*

Proof. For $t \in \mathcal{T}$, $\varepsilon > 0$, we first define the rescaled processes

$$\begin{aligned} Z_t^\varepsilon &:= \varepsilon^H Z_t \stackrel{d}{=} Z_{\varepsilon t}, \\ v_t^\varepsilon &:= v_0 \exp\left(Z_t^\varepsilon - \frac{\eta^2}{2}(\varepsilon t)^{2H}\right), \end{aligned} \tag{2.11}$$

where $H \in (0, \frac{1}{2})$. From Schilder's Theorem and Proposition 45, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed ε^{2H} and rate function Λ^Z . We now prove that the two sequences of stochastic processes Z^ε and $\tilde{Z}^\varepsilon := Z^\varepsilon - \frac{\eta^2}{2}(\varepsilon \cdot)^{2H}$ are exponentially equivalent, as introduced in Definition 8. For each $\delta > 0$ and $t \in \mathcal{T}$, there exists $\varepsilon_* := \frac{1}{t} \left(\frac{2\delta}{\eta^2}\right)^{1/2H} > 0$ such that

$$\sup_{t \in \mathcal{T}} |Z_t^\varepsilon - \tilde{Z}_t^\varepsilon| = \sup_{t \in \mathcal{T}} \left| \frac{\eta^2}{2}(\varepsilon t)^{2H} \right| \leq \delta,$$

for all $0 < \varepsilon < \varepsilon_*$. Therefore, for all $\delta > 0$, $\limsup_{\varepsilon \downarrow 0} \varepsilon^{2H} \log \mathbb{P}(\|Z^\varepsilon - \tilde{Z}^\varepsilon\|_\infty > \delta) = -\infty$, and the two processes are indeed exponentially equivalent. Then, using [39, Theorem 4.2.13], the sequence of stochastic processes $(\tilde{Z}^\varepsilon)_{\varepsilon>0}$ also satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$, with speed ε^{2H} and rate function Λ^Z . Moreover, for all ε, t , we have that $v_t^\varepsilon = v_0 \exp(\tilde{Z}_t^\varepsilon)$, 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 [39, Theorem 4.2.1], which is stated in Theorem 10, concluding that the sequence of processes $(v^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed ε^{2H} 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 $x \in \mathcal{C}(\mathcal{T})$, 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 $t \in \mathcal{T}$, v_t^ε and $v_{\varepsilon t}$ are equal in law, we conclude the theorem. Notice also that $\Lambda^v(v_0) = \Lambda^Z(0) = \|0\|_{\mathcal{H}_{K_{H_-}}}^2 = 0$. $\square$

Corollary 47. *The integrated variance process $(RV(v)(t))_{t \in \mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, $RV(v)(t) \sim \text{LDP}\left(t^{2H}, \hat{\Lambda}^v\right)$ with $\hat{\Lambda}^v(y) := \inf\{\Lambda^v(x) : y = RV(x)(1)\}$, where $\hat{\Lambda}^v(v_0) = 0$.*

Proof. As proved in Theorem 46, 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

$$\|RV(v + \delta^v)(t) - RV(\delta^v)(t)\|_\infty \leq \sup_{t \in \mathcal{T}} \frac{1}{t} \left| \int_0^t \delta^v(s) ds \right| \leq M,$$

where $M = \sup_{t \in \mathcal{T}} |\delta^v(t)|$, which is finite as $\delta^v \in \mathcal{C}(\mathcal{T})$. 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 Theorem 10), and consequently the integrated variance process $RV(v_\varepsilon)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero. $RV(v_\varepsilon)(t) = RV(v)(\varepsilon t)$, for all $t \in \mathcal{T}$, and so setting $t = 1$ and mapping ε to t then yields the result. By definition, $\hat{\Lambda}^v(y) := \inf\{\Lambda^v(x) : y = RV(x)(1)\}$. If we choose $x \equiv v_0$ then $v_0 = RV(x)(1)$, and $\Lambda^v(x) = 0$. Since Λ^v is a norm, it is a non-negative function and therefore $\hat{\Lambda}^v(v_0) = 0$. This concludes the proof. $\square$

Remark 48. Corollary 47 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 [52] and Horvath, Jacquier, and Lacombe [76].

Corollary 49. 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 [52, Corollary 4.6], and deduce that $\hat{\Lambda}^v$ is upper semi-continuous, and is hence continuous. $\square$

Before stating results on the small-time behaviour of options on integrated variance, we state that the log integrated variance process $\log RV(v)$ satisfies a large deviations principle on $\mathbb{R}$ as t tends to zero, with speed t^{2H} and rate function $\hat{\Lambda}^v(e)$. Then, the small-time behaviour of such options can be obtained as an application of Corollary 47.

Corollary 50. For log moneyness $k := \log \frac{K}{RV(v)(0)} \neq 0$, the following equality holds true for Call options on integrated variance

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E} \left[\left(RV(v)(t) - e^k \right)^+ \right] = -I(k), \quad (2.12)$$

where I is defined as $I(x) := \inf_{y > x} \hat{\Lambda}^v(e^y)$ for $x > 0$, $I(x) := \inf_{y < x} \hat{\Lambda}^v(e^y)$ for $x < 0$.

Similarly, for log moneyness $k := \log \frac{K}{\sqrt{RV(v)(0)}} \neq 0$,

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E} \left[\left(\sqrt{RV(v)(t)} - e^k \right)^+ \right] = -\bar{I}(k), \quad (2.13)$$

where $\bar{I}$ is defined analogously as $\bar{I}(x) := \inf_{y > x} \hat{\Lambda}^v(e^{2y})$ for $x > 0$ and $\bar{I}(x) := \inf_{y < x} \hat{\Lambda}^v(e^{2y})$ for $x < 0$.

Proof. The proof of Equation (2.12) is similar to the proof of [52, 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 $k > 0$,

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{P}(\log[RV(v)(t)] > k) = -I(k),$$

as an application of Corollary 47.

- (1) The proof of the lower bound is exactly the same as presented in [52, Appendix C] and will be omitted here; we arrive at

$$\liminf_{t \downarrow 0} t^{2H} \log \mathbb{E} [(RV(v)(t) - e^k)^+] \geq -I(k).$$

- (2) To establish the upper bound, we apply Hölder's inequality:

$$\begin{aligned} \mathbb{E}(RV(v)(t) - e^k)^+ &= \mathbb{E} [(RV(v)(t) - e^k)^+ \mathbf{1}_{\{RV(v)(t) \geq e^k\}}], \\ &\leq \mathbb{E} \left[\left((RV(v)(t) - e^k)^+ \right)^q \right]^{1/q} \mathbb{P}(RV(v)(t) \geq e^k)^{1-1/q}, \\ &\leq \mathbb{E} [(RV(v)(t))^q]^{1/q} \mathbb{P}(RV(v)(t) \geq e^k)^{1-1/q}, \end{aligned}$$

which holds for all $q > 1$. Thus,

$$t^{2H} \log \mathbb{E}[(RV(v)(t) - e^k)^+] \leq \frac{t^{2H}}{q} \log \mathbb{E}[(RV(v)(t))^q] + t^{2H} \left(1 - \frac{1}{q}\right) \log \mathbb{P}(RV(v)(t) \geq e^k). \quad (2.14)$$

We then apply Jensen's inequality and obtain:

$$\begin{aligned} \mathbb{E} [(RV(v)(t))^q] &\leq \frac{1}{t^q} \int_0^t \mathbb{E}[v_s^q] ds, \\ &\leq \frac{v_0^q}{t^q} \int_0^t \exp \left(\left(\frac{q^2 \eta^2}{2} - \frac{q \eta^2}{2} \right) s^{2H} \right) ds, \\ &\leq \frac{v_0^q}{t^{q-1}} \exp \left(\left(\frac{q^2 \eta^2}{2} - \frac{q \eta^2}{2} \right) t^{2H} \right). \end{aligned} \quad (2.15)$$

Therefore,

$$\begin{aligned} \lim_{q \uparrow \infty} \limsup_{t \downarrow 0} \frac{t^{2H}}{q} \log \mathbb{E}[(RV(v)(t))^q] &\leq \lim_{q \uparrow \infty} \limsup_{t \downarrow 0} \frac{t^{2H}}{q} (q \log v_0 - (q-1) \log t) \\ &\quad + \lim_{q \uparrow \infty} \limsup_{t \downarrow 0} \frac{t^{2H}}{q} \left(q(q-1) \frac{\eta^2}{2} t^{2H} \right), \\ &= 0. \end{aligned}$$

Hence, we take $q \uparrow \infty$ and $t \downarrow 0$ on both sides of (2.14) and we obtain, by Corollary 47

$$\limsup_{t \downarrow 0} t^{2H} \log \mathbb{E}[(RV(v)(t) - e^k)^+] \leq -I(k).$$

The conclusion for Equation (2.12) then follows directly.

The proof of Equation (2.13) 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^{2H} and rate function $\hat{\Lambda}^v((\cdot)^2)$, using [39, Theorem 4.2.4]. $\square$

As with Call options on stock price processes, we can define and study the implied volatility of such options. In the case of (2.8)(i) we define the implied volatility $\hat{\sigma}(T, k)$ to be the solution to

$$\mathbb{E}[(RV(v)(T) - e^k)^+] = C_{BS}(RV(v)(0), k, T, \hat{\sigma}(T, k)), \quad (2.16)$$

where C_{BS} denotes the Call price in the Black-Scholes model. Using Corollary 50, we deduce the small-time behaviour of the implied volatility $\hat{\sigma}$, as defined in (2.16).

Corollary 51. *The small-time asymptotic behaviour of the implied volatility is given by the following limit, for a log moneyness $k \neq 0$:*

$$\lim_{t \downarrow 0} t^{1-2H} \hat{\sigma}^2(t, k) =: \hat{\sigma}^2(k) = \frac{k^2}{2I(k)}.$$

Proof. The log integrated variance process $\log RV(v)$ satisfies a large deviations principle with speed t^{2H} and rate function $\hat{\Lambda}^v(e^{\cdot})$, which is continuous. Therefore, it follows that

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{P}(RV(v)(t) \geq e^k) = -I(k).$$

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:

$$\lim_{t \downarrow 0} \xi^2 t \log \mathbb{P}(RV(v)(t) \geq e^k) = -\frac{k^2}{2}.$$

We then apply [61, Corollary 7.1] using Corollary 50, identifying $\xi \equiv \hat{\sigma}(k, t)$, to conclude. $\square$

Remark 52. Notice that the level of implied volatility in Corollary 51 has a power law behaviour as a function of time to maturity. This power law is of order $\sqrt{t^{2H-1}}$, which is consistent with the at-the-money RV implied volatility results by Alòs, García-Lorite and Muguruza [4], where the order is found to be t^{H-} using Malliavin Calculus techniques. Recall that $H_- := H - \frac{1}{2}$.

2.3.2 SMALL-TIME RESULTS FOR THE MIXED ROUGH BERGOMI MODEL

Minor adjustments to Theorem 46 give the following small-time result for the mixed variance process $v^{(\gamma, \nu)}$ introduced in Model (2.6); the proof is given in Section 2.9.

Theorem 53. The mixed variance process $(v_\varepsilon^{(\gamma, \nu)})_{\varepsilon > 0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$, $v_\varepsilon^{(\gamma, \nu)} \sim \text{LDP}(\varepsilon^{2H}, \Lambda^{(\gamma, \nu)})$, with

$$\Lambda^{(\gamma, \nu)}(x) := \inf \left\{ \Lambda^Z \left(\frac{\eta}{\nu_1} y \right) : x(\cdot) = v_0 \sum_{i=1}^n \gamma_i e^{\frac{\nu_i}{\nu_1} y(\cdot)} \right\},$$

satisfying $\Lambda^{(\gamma, \nu)}(v_0) = 0$.

By Remark 48, we immediately get the following result for the small-time behaviour of the integrated mixed variance process $RV(v^{(\gamma, \nu)})$.

Corollary 54. *The integrated mixed variance process $(RV(v^{(\gamma, \nu)})(t))_{t \in \mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, so that $RV(v^{(\gamma, \nu)})(t) \sim \text{LDP}\left(t^{2H}, \tilde{\Lambda}^{(\gamma, \nu)}\right)$, with $\tilde{\Lambda}^{(\gamma, \nu)}(y) := \inf\{\Lambda^{(\gamma, \nu)}(x) : y = RV(x)(1)\}$, where $\tilde{\Lambda}^{(\gamma, \nu)}(v_0) = 0$.*

To get the small-time implied volatility result, analogous to Corollary 51, we need the following Lemma, which is used in place of (2.15). The remainder of the proof then follows identically.

Lemma 55. *For all $t \in \mathcal{T}$ and $q > 1$, we have*

$$\mathbb{E}\left[\left(RV(v^{(\gamma, \nu)})(t)\right)^q\right] \leq \frac{v_0^q n^{q-1}}{t^{q-1}} \exp\left(\frac{(\nu^*)^2}{2\eta^2} (q^2 - q) t^{2H}\right),$$

where $\nu^* := \max\{\nu_1, \dots, \nu_n\}$.

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 $i = 1, \dots, n$, we obtain

$$\begin{aligned} \mathbb{E}\left[\left(RV(v^{(\gamma, \nu)})(t)\right)^q\right] &\leq \frac{v_0^q}{t^q} n^{q-1} \sum_{i=1}^n \int_0^t \mathbb{E}\left[\exp\left(\frac{q\nu_i}{\eta} Z_s - \frac{q\nu_i^2}{2\eta^2} s^{2H}\right)\right] ds, \\ &\leq \frac{v_0^q}{t^{q-1}} n^{q-1} \sum_{i=1}^n \exp\left(\frac{\nu_i^2}{2\eta^2} (q^2 - q) t^{2H}\right). \end{aligned}$$

Choosing $\nu^* = \max\{\nu_1, \dots, \nu_n\}$ the result directly follows. $\square$

Corollary 56. *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:*

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E}\left[\left(RV(v^{(\gamma, \nu)})(t) - e^k\right)^+\right] = -I(k), \quad (2.17)$$

where I is defined as $I(x) := \inf_{y > x} \tilde{\Lambda}^{(\gamma, \nu)}(e^y)$ for $x > 0$, $I(x) := \inf_{y < x} \tilde{\Lambda}^{(\gamma, \nu)}(e^y)$ for $x < 0$.

Similarly, for log moneyness $k := \log \frac{K}{\sqrt{RV(v^{(\gamma, \nu)})(0)}} \neq 0$,

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E} \left[\left(\sqrt{RV(v^{(\gamma, \nu)})(t)} - e^k \right)^+ \right] = -\bar{I}(k), \quad (2.18)$$

where $\bar{I}$ is defined analogously as $\bar{I}(x) := \inf_{y > x} \tilde{\Lambda}^{(\gamma, \nu)}(e^{2y})$ for $x > 0$ and $\bar{I}(x) := \inf_{y < x} \tilde{\Lambda}^{(\gamma, \nu)}(e^{2y})$ for $x < 0$.

Proof. Follows directly from Lemma 55 and proof of Corollary 50. $\square$

The small-time implied volatility behaviour for the mixed rough Bergomi model is then given by Corollary 51, where the function I is defined in terms of the rate function $\tilde{\Lambda}^{(\gamma, \nu)}$, as in Corollary 56, in this case.

2.3.3 SMALL-TIME RESULTS FOR THE MULTI-FACTOR ROUGH BERGOMI MODEL

The small-time behaviour of the multi-factor rough Bergomi model (2.7) can then be obtained; see Theorem 57 below; note that Λ^m is the rate function associated to the reproducing kernel Hilbert space of the measure induced by $(B^1, \dots, B^m)$ on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$.

Theorem 57. *The variance process in the multi-factor rough Bergomi model $\left(v_t^{(\gamma, \nu, \Sigma)} \right)_{t \in \mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ at t tends to zero, $\left(v_t^{(\gamma, \nu, \Sigma)} \right) \sim \text{LDP} \left(t^{2H}, \Lambda^{(\gamma, \nu, \Sigma)} \right)$, with*

$$\Lambda^{(\gamma, \nu, \Sigma)}(y) := \inf \left\{ \Lambda^m(x) : x \in \mathcal{H}_m, y = v_0 \sum_{i=1}^n \gamma_i \exp \left(\frac{\nu^i}{\eta} \cdot L_i x(1) \right) \right\},$$

satisfying $\Lambda^{(\gamma, \nu, \Sigma)}(v_0) = 0$.

Proof. We begin by introducing a small-time rescaling of (2.7) for $\varepsilon > 0$, so

that the system becomes

$$v_t^{(\gamma, \nu, \Sigma, \varepsilon)} := v_{\varepsilon t}^{(\gamma, \nu, \Sigma)} = v_0 \sum_{i=1}^n \gamma_i \mathcal{E} \left(\frac{\nu^i}{\eta} \cdot L_i \mathcal{Z}_t^\varepsilon \right), \quad (2.19)$$

with the rescaled process $\mathcal{Z}_t^\varepsilon := \mathcal{Z}_{\varepsilon t}$ defined as

$$\mathcal{Z}_t^\varepsilon = \varepsilon^H \left(\int_0^t K_{H_-}(s, t) dB_s^1, \dots, \int_0^t K_{H_-}(s, t) dB_s^m \right).$$

The m -dimensional sequence of processes $(\varepsilon^H(B^1, \dots, B^m))_{\varepsilon > 0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$ as ε goes to zero with speed ε^{2H} and rate function Λ^m defined by $\Lambda^m(x) := \frac{1}{2} \|x\|_2^2$ for $x \in \mathcal{H}_m$ and $+\infty$ otherwise, by Schilder's Theorem, recalled in Theorem 11. $\mathcal{H}_m$ is the reproducing kernel Hilbert space of the measure induced by $(W_1, \dots, W_m)$ on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$, defined as

$$\mathcal{H}_m := \left\{ (g^1, \dots, g^m) \in \mathcal{C}(\mathcal{T}, \mathbb{R}^m) : g^i(t) = \int_0^t f^i(s) ds, f^i \in L^2(\mathcal{T}) \text{ for all } i \in 1 \dots m \right\}.$$

Then, using an extension of the proof of [76, Theorem 3.3], for $i = 1, \dots, n$, the sequence of m -dimensional processes $(L_i \mathcal{Z}_t^\varepsilon)_{\varepsilon > 0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$ as ε tends to zero with speed ε^{2H} and rate function $y \mapsto \inf \{ \Lambda^m(x) : x \in \mathcal{H}_m, y = L_i x \}$ with L_i the lower triangular matrix introduced in Model (2.7). Consequently, for $i = 1, \dots, n$ each (one-dimensional) sequence of processes $\left(\frac{\nu^i}{\eta} \cdot L_i \mathcal{Z}_t^\varepsilon \right)_{\varepsilon > 0}$ also satisfies a large deviations principle as ε tends to zero, with speed ε^{2H} and rate function $\Lambda_{\Sigma_i}(y) := \inf \left\{ \Lambda^m(x) : x \in \mathcal{H}_m, y = \frac{\nu^i}{\eta} \cdot L_i x \right\}$.

Analogously to Theorem 46, each sequence of processes $\left(\frac{\nu^i}{\eta} \cdot L_i \mathcal{Z}_t^\varepsilon \right)_{\varepsilon > 0}$ and $\left(\frac{\nu^i}{\eta} \cdot L_i \mathcal{Z}_t^\varepsilon - \frac{1}{2} \nu^i \Sigma_i \nu^i (\varepsilon \cdot)^{2H} \right)_{\varepsilon > 0}$ are exponential equivalent for $i = 1 \dots n$; therefore they satisfy the same large deviations principle with the same speed ε^{2H} and the same rate function Λ_{Σ_i} .

We now define the operator $g^\gamma : \mathcal{C}(\mathcal{T}, \mathbb{R}^n) \rightarrow \mathcal{C}(\mathcal{T})$ as

$$g^\gamma(\mathbf{x})(\cdot) := v_0 \sum_{i=1}^n \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot \mathbf{x}(\cdot)\right),$$

with $\mathbf{x} := (x_1, \dots, x_n)$. For small perturbations $\delta^1, \dots, \delta^n \in \mathcal{C}(\mathcal{T})$ with $\delta := (\delta^1, \dots, \delta^n)$, we have that

$$\begin{aligned} \sup_{t \in \mathcal{T}} |g^\gamma(\mathbf{x} + \delta)(t) - g^\gamma(\mathbf{x})(t)| &= \sup_{t \in \mathcal{T}} \left| v_0 \sum_{i=1}^n \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot (\mathbf{x}(t) + \delta(t))\right) - \exp\left(\frac{\nu^i}{\eta} \cdot \mathbf{x}(t)\right) \right|, \\ &\leq \sup_{t \in \mathcal{T}} |v_0| \sum_{i=1}^n \left| \exp\left(\frac{\nu^i}{\eta} \cdot \mathbf{x}(t)\right) (\exp(\delta(t)) - 1) \right|. \end{aligned}$$

The right-hand side tends to zero as $\delta^1, \dots, \delta^n$ 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 $v_t^{(\gamma, \nu, \Sigma, \varepsilon)} = g^\gamma(\frac{\nu^i}{\eta} \cdot \mathbf{L}_i \mathcal{Z}^\varepsilon - \frac{1}{2} \nu^i \Sigma_i \nu^i (\varepsilon \cdot)^{2H})(t)$ for each $\varepsilon > 0$ and $t \in \mathcal{T}$, we can apply the Contraction Principle then yields 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 ε^{2H} and rate function

$$\begin{aligned} y &\mapsto \inf \left\{ \Lambda_{\Sigma_i}(\mathbf{x}) : y = v_0 \sum_{i=1}^n \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot \mathbf{x}\right) \right\}, \\ &= \inf \left\{ \Lambda^m(\mathbf{x}) : \mathbf{x} \in \mathcal{H}_m, y = v_0 \sum_{i=1}^n \gamma_i \exp\left(\frac{\nu^i}{\eta} \cdot \mathbf{L}_i \mathbf{x}\right) \right\}. \end{aligned}$$

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. $\square$

As with the mixed variance process, Remark 48 gives us the following small-time result for $RV(v^{(\gamma, \nu, \Sigma)})$.

Corollary 58. *The integrated variance process $(RV(v^{(\gamma, \nu, \Sigma)})(t))_{t \in \mathcal{T}}$ in the multi-factor Bergomi model satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, $RV(v^{(\gamma, \nu, \Sigma)})(t) \sim \text{LDP}(t^{2H}, \tilde{\Lambda}^{(\gamma, \nu, \Sigma)})$, with*

$$\tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(y) := \inf \left\{ \Lambda^{(\gamma, \nu, \Sigma)}(x) : y = RV(x)(1) \right\},$$

where $\tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(v_0) = 0$.

We now establish the small-time behaviour for Call options on realised variance in Corollary 60, by adapting the proof of Corollary 50 as in the previous subsection. To do so we use Lemma 59 in place of (2.15). Then we attain the small-time implied volatility behaviour for the multi-factor rough Bergomi model in Corollary 51, where the function I is given by Corollary 60.

Lemma 59. *For all $t \in \mathcal{T}$ and $q > 1$, we have*

$$\mathbb{E} \left[(RV(v^{(\gamma, \nu, \Sigma)})(t))^q \right] \leq \frac{v_0^q n^{q-1}}{t^{q-1}} \exp \left(\frac{(\nu^*)^2}{2\eta^2} (q^2 - q) t^{2H} \right),$$

where $\nu^* = \max\{\nu_1, \dots, \nu_n\}$.

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 $i = 1, \dots, n$, we obtain

$$\begin{aligned} \mathbb{E} \left[(RV(v^{(\gamma, \nu, \Sigma)})(t))^q \right] &\leq \frac{v_0^q}{t^q} n^{q-1} \sum_{i=1}^n \int_0^t \mathbb{E} \left[\mathcal{E} \left(\frac{\nu^i}{\eta} \cdot L_i \mathcal{Z}_s \right)^q \right] ds, \\ &\leq \frac{v_0^q}{t^{q-1}} n^{q-1} \sum_{i=1}^n \exp \left(\frac{\nu_i^2}{2\eta^2} (q^2 - q) t^{2H} \right). \end{aligned}$$

Choosing $\nu^* = \max\{\nu_1, \dots, \nu_n\}$ the result directly follows. $\square$

Corollary 60. *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:*

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E} \left[\left(RV(v^{(\gamma, \nu, \Sigma)})(t) - e^k \right)^+ \right] = -I(k), \quad (2.20)$$

where I is defined as $I(x) := \inf_{y > x} \tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(e^y)$ for $x > 0$, $I(x) := \inf_{y < x} \tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(e^y)$ for $x < 0$.

Similarly, for log moneyness $k := \log \frac{K}{\sqrt{RV(v^{(\gamma, \nu, \Sigma)})(0)}} \neq 0$,

$$\lim_{t \downarrow 0} t^{2H} \log \mathbb{E} \left[\left(\sqrt{RV(v^{(\gamma, \nu, \Sigma)})(t)} - e^k \right)^+ \right] = -\bar{I}(k), \quad (2.21)$$

where $\bar{I}$ is defined analogously as $\bar{I}(x) := \inf_{y > x} \tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(e^{2y})$ for $x > 0$ and $\bar{I}(x) := \inf_{y < x} \tilde{\Lambda}^{(\gamma, \nu, \Sigma)}(e^{2y})$ for $x < 0$.

Proof. Follows directly from Lemma 59 and the proof of Corollary 50. $\square$

2.4 NUMERICAL RESULTS

In this section we present numerical results for each of the three models given in Section 2.2. We also analyse the effect of each parameter in the implied volatility smile. Numerical experiments and codes are provided on [GitHub: LDP-VolOptions](#) .

2.4.1 RV SMILES FOR ROUGH BERGOMI

We begin with numerical results for the rough Bergomi model (2.5) using Corollary 51. For the detailed numerical method we refer the reader to Section 2.6. In Figure 2.2, we represent the rate function given in Corollary 47, 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.

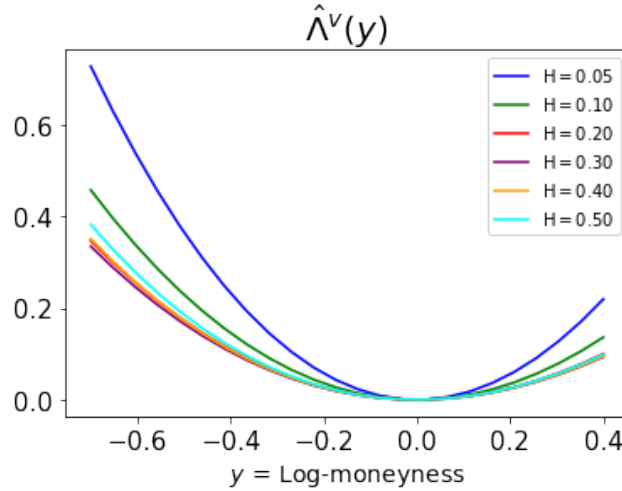


Figure 2.2: Rate function $\hat{\Lambda}^v$ for different values of H .

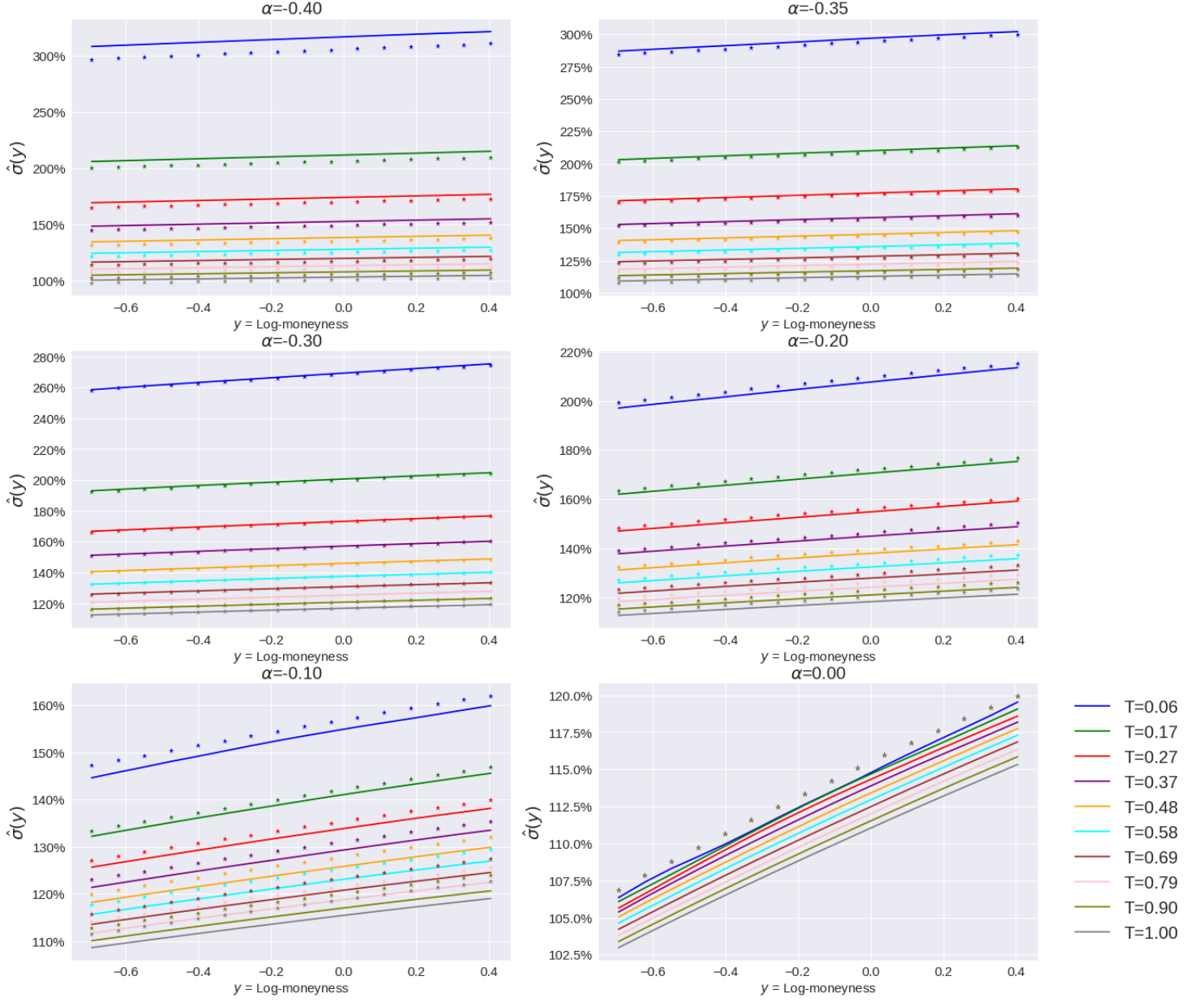


Figure 2.3: Comparison of Monte Carlo computed implied volatilities (straight lines) and LDP based implied volatilities (stars), in the rough Bergomi model, for different values of $H_- =: \alpha$ and maturities T . We set $\eta = 2$ and $v_0 = 0.04$; for Monte Carlo we use 200,000 simulations and $\Delta t = \frac{1}{1008}$.

More interestingly, in Figure 2.3 we provide a comparison of Corollary 51 with respect to a benchmark generated by Monte Carlo simulations, and see all smiles to follow a linear trend. An example of the confidence intervals associated with the simulations is displayed in Figure 2.5. In particular, we notice that Corollary 51 provides a surprisingly accurate estimate, even for relatively large maturities. As a further numerical check, in Figure 2.4 we compare our results with the close-form at-the-money asymptotics given by Alòs, García-Lorite and Muguruza [4] and once again find the correct convergence, suggesting a consistent numerical framework.

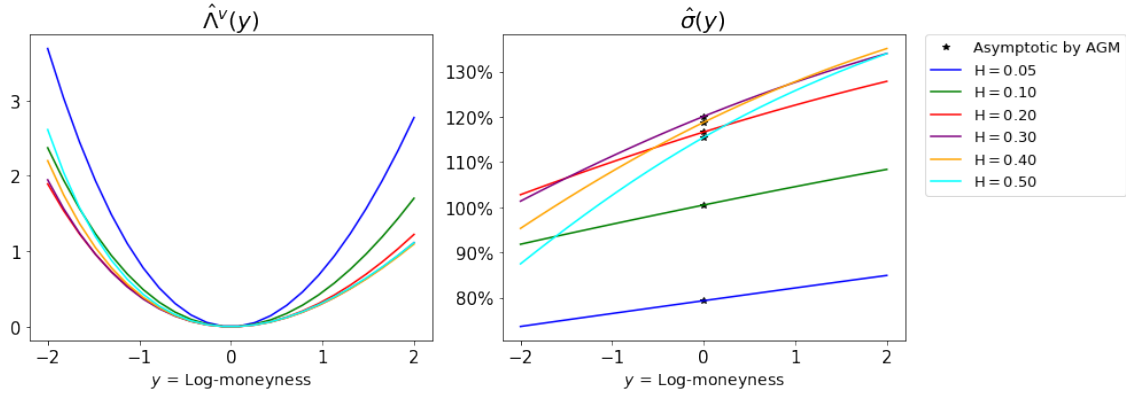


Figure 2.4: Comparison of Alòs, García-Lorite and Muguruza [4] at-the-money implied volatility asymptotics and LDP based implied volatilities for different values of H_- , in the rough Bergomi model, with $\eta = 1.5$ and $v_0 = 0.04$.

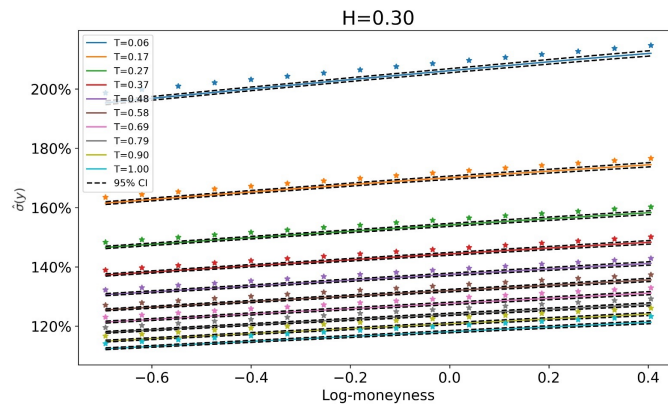


Figure 2.5: Confidence intervals (dashed lines) associated with the Monte Carlo simulations, for $H = 0.3$ and different maturities T . We set $\eta = 2$ and $v_0 = 0.04$; for Monte Carlo we use 200,000 simulations and $\Delta t = \frac{1}{1008}$.

2.4.2 RV SMILES FOR MIXED ROUGH BERGOMI

We now consider the mixed rough Bergomi model (2.6) in a simplified form given by $v_t = v_0 (\gamma_1 \mathcal{E}(\nu_1 Z_t) + \gamma_2 \mathcal{E}(\nu_2 Z_t))$. In Figure 2.6, we observe that a constraint of the type $\gamma_1 \nu_1 + \gamma_2 \nu_2 = 2$ in the mixed variance process (2.6) 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 2.3, we see that the smiles generated follow a linear trend. This is again consistent with the results found in [4].

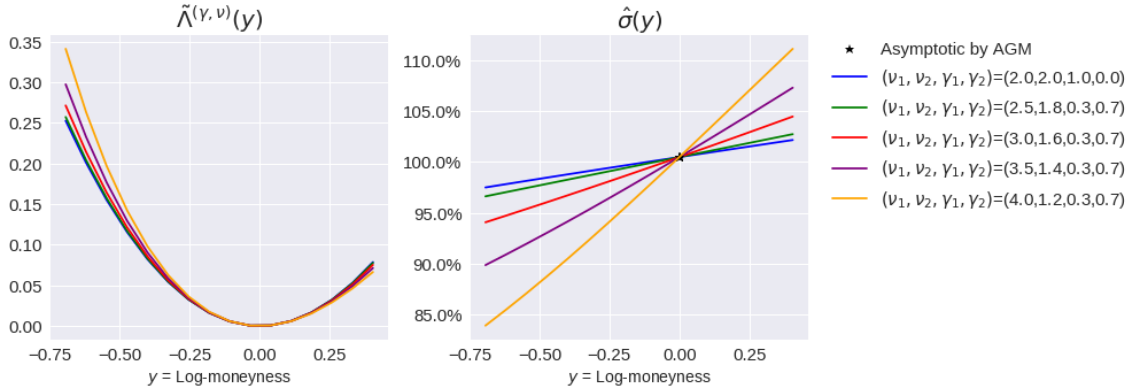


Figure 2.6: Comparison of LDP based implied volatilities for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$ in the mixed rough Bergomi process (2.6) such that $\gamma_1 \nu_1 + \gamma_2 \nu_2 = 2$, with $H_- = -0.4$.

Remark 61. *At this point it is important to note that the mixed rough Bergomi model 2.6 allows both the at-the-money implied volatility and its skew to be controlled through (γ, ν) , whilst in the rough Bergomi model (2.5) there is not enough freedom to arbitrarily fit both quantities. Remarkably, we observe a linear pattern in Figures 2.3-2.6 for around the money strikes. In Section 2.7, we provide an approximation scheme for the realised variance density based on the assumption of linear smiles.*

2.4.3 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

(2.7). We shall consider the following two simplified models for instantaneous variance

$$v_t = \mathcal{E} \left(\nu \int_0^t (t-s)^{H-} dB_s + \eta \left(\rho \int_0^t (t-s)^{H-} dB_s + \sqrt{1-\rho^2} \int_0^t (t-s)^{H-} dB_s^\perp \right) \right), \quad (2.22)$$

$$v_t = \frac{1}{2} \left(\mathcal{E} \left(\nu \int_0^t (t-s)^{H-} dB_s \right) + \mathcal{E} \left(\eta \left(\rho \int_0^t (t-s)^{H-} dB_s + \sqrt{1-\rho^2} \int_0^t (t-s)^{H-} dB_s^\perp \right) \right) \right), \quad (2.23)$$

where B and $B^\perp$ are independent standard Brownian motions and $\nu, \eta > 0$.

On one hand, Figure 2.7 shows the implied volatility smiles corresponding to (2.22). 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 [4] results still hold and we provide the asymptotic benchmark in Figure 2.7 to support our numerical scheme. On the other hand, Figure 2.8 shows the implied volatility smiles corresponding to (2.23), 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 2.9 to clearly show the effect of correlation in the convexity of the smiles.

Remark 62. *Despite the empirical evidence given in Figure 2.1, the commitment to linear smiles from a modelling perspective is strong. The simplified mixed multi-factor rough Bergomi model (2.23), however, is sufficiently flexible to generate both linear ($\rho \geq 0$) and nonlinear ($\rho < 0$) smiles.*

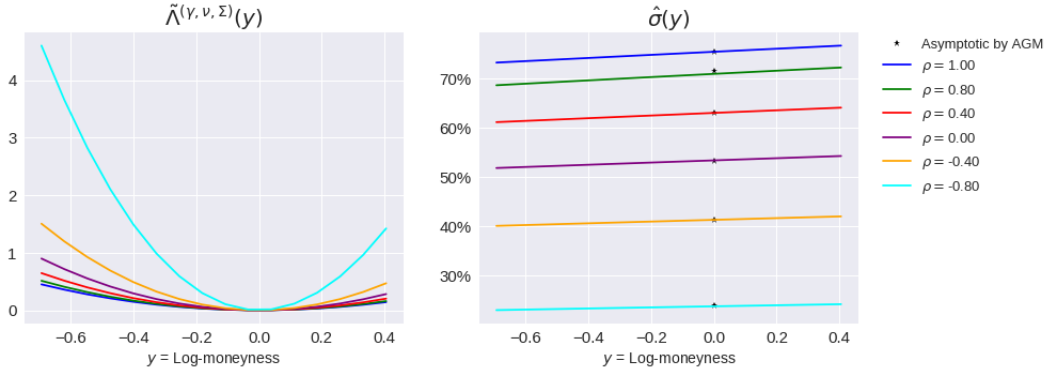


Figure 2.7: Rate function and corresponding implied volatilities for the model (2.22), with $(H_-, \nu, \eta) = (-0.4, 0.75, 0.75)$.

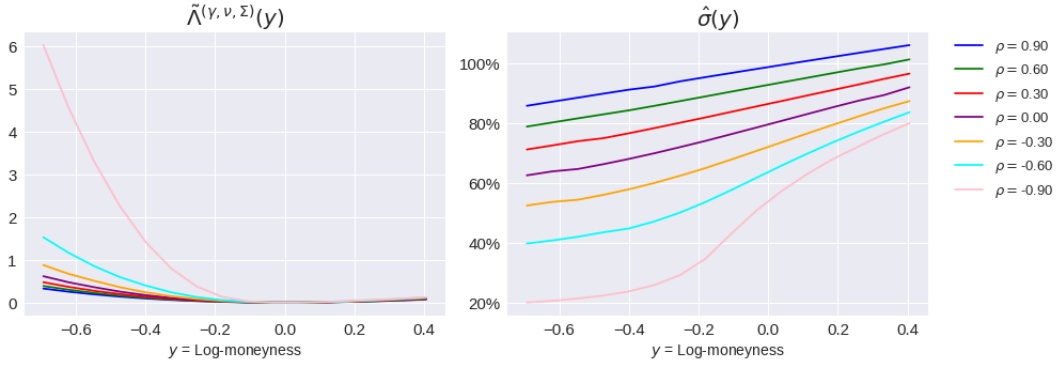


Figure 2.8: Rate function and corresponding implied volatilities for the model (2.23) with $(H_-, \nu, \eta) = (-0.4, 1.0, 3.0)$.

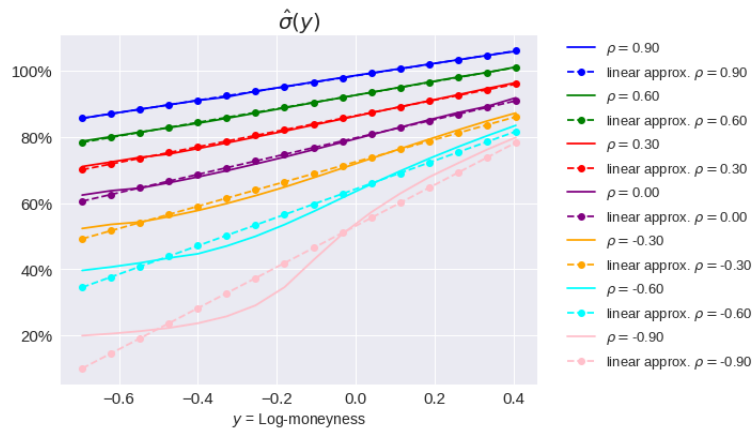


Figure 2.9: Implied volatilities and superimposed linear smiles for the model (2.23), with $(H_-, \nu, \eta) = (-0.4, 1.0, 3.0)$.

2.5 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 index 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:

$$v_t = \xi_0(t) \mathcal{E} \left(\int_0^t g(t, s) dB_s \right), \quad (2.24)$$

with $\xi_0(\cdot) > 0$ the initial forward variance curve. Then, the VIX process is given by

$$\text{VIX}_T = \sqrt{\frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}[v_t | \mathcal{F}_T] ds}.$$

We introduce the following stochastic process $(V^{g,T})_{t \in [T, T+\Delta]}$, for notational convenience, as

$$V_t^{g,T} := \int_0^T g(t, s) dB_s, \quad (2.25)$$

and assume that the mapping $s \mapsto g(t, s)$ is in $L^2(\mathcal{T})$ for all $t \in [T, T + \Delta]$ such that the stochastic integral in (2.25) is well-defined.

Proposition 63. *The VIX dynamics in model (2.24), where the volatility of volatility is denoted by ν , are given by*

$$\text{VIX}_{T,\nu}^2 := \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0(t) \exp \left(\nu V_t^{g,T} - \frac{\nu^2}{2} \mathbb{E}[(V_t^{g,T})^2] \right) dt.$$

Proof. Follows directly from [81, Proposition 3.1]. □

We now define the following $L^2(\mathcal{T})$ operator $\mathcal{I}^{g,T} : L^2(\mathcal{T}) \rightarrow \mathcal{C}[T, T + \Delta]$, and space $\mathcal{H}^{g,T}$ as

$$\mathcal{I}^{g,T} f(\cdot) := \int_0^T g(\cdot, s) f(s) ds, \quad \mathcal{H}^{g,T} := \{\mathcal{I}^{g,T} f : f \in L^2(\mathcal{T})\}, \quad (2.26)$$

where the space $\mathcal{H}^{g,T}$ is equipped with the following inner product $\langle \mathcal{I}^{g,T} f_1, \mathcal{I}^{g,T} f_2 \rangle_{\mathcal{H}^{g,T}} := \langle f_1, f_2 \rangle_{L^2(\mathcal{T})}$. Note that the function g must be such that the operator $\mathcal{I}^{g,T}$ is injective so that that inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}^{g,T}}$ on $\mathcal{H}^{g,T}$ is well-defined.

Proposition 64. *Assume that there exists $h \in L^2(\mathcal{T})$ such that $\int_0^\varepsilon |h(s)| ds < +\infty$ for some $\varepsilon > 0$ and $g(t, \cdot) = h(t - \cdot)$ for any $t \in [T, T + \Delta]$. Then, the space $\mathcal{H}^{g,T}$ is the reproducing kernel Hilbert space for the process $(V_t^{g,T})_{t \in [T, T + \Delta]}$.*

Proof. See Section 2.9. □

Theorem 65. *For any $\gamma > 0$, the sequence of stochastic processes $(\varepsilon^{\gamma/2} V^{g,T})_{\varepsilon > 0}$ satisfies a large deviations principle on $\mathcal{C}[T, T + \Delta]$ as ε goes to zero, $\varepsilon^{\gamma/2} V^{g,T} \sim \text{LDP}(\varepsilon^\gamma, \Lambda^V)$, with Λ^V , defined as*

$$\Lambda^V(\mathbf{x}) := \begin{cases} \frac{1}{2} \|\mathbf{x}\|_{\mathcal{H}^{g,T}}^2, & \text{if } \mathbf{x} \in \mathcal{H}^{g,T}, \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.27)$$

Proof. Direct application of the generalised Schilder's Theorem [42, Theorem 3.4.12]. □

Remark 66. *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 65 and using that the rate function Λ^V is continuous on A , we can obtain then obtain the following tail behaviour of the process $V^{g,T}$:*

$$\lim_{\varepsilon \downarrow 0} \varepsilon^\gamma \log \mathbb{P} \left(V_t^{g,T} \geq \frac{1}{\varepsilon^{\gamma/2}} \right) = - \inf_{g \in A} \Lambda^V(g), \quad (2.28)$$

for any $\gamma > 0$ and $t \in [T, T + \Delta]$ fixed.

Remark 67. *Let us again fix the kernel g as the rough Bergomi kernel and denote the corresponding reproducing kernel Hilbert space by $\mathcal{H}^{\eta, H-, T}$ and the corresponding process $V^{g,T}$ as $V^{\eta, H-, T}$. If $\mathbf{x} \in \mathcal{H}^{\eta, H-, T}$ it follows that there exists $f \in L^2(\mathcal{T})$ such that $\mathbf{x}(t) = \int_0^T \eta \sqrt{2H}(t-s)^{H-} f(s) ds$ for all $t \in [T, T + \Delta]$. Clearly, it follows that $\mathbf{x} \in \mathcal{H}^{a\eta, H-, T}$ for any $a > 0$, as*

$f \in L^2(\mathcal{T})$ implies that $\frac{1}{a}f =: f_a \in L^2(\mathcal{T})$. We can compute the norm of x in each of these spaces to arrive at the following isometry:

$$\|x\|_{\mathcal{H}^{a\eta, H-, T}}^2 = \|f_a\|_{L^2[0, T]}^2 = \frac{1}{a^2} \int_0^T f^2(s) ds = \frac{1}{a^2} \|x\|_{\mathcal{H}^{\eta, H-, T}}^2. \quad (2.29)$$

We may now amalgamate (2.27), (2.28), and (2.29) 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:

$$\lim_{\varepsilon \downarrow 0} \varepsilon^\gamma \log P \left(V_t^{a\eta, H-, T} \geq \frac{1}{\varepsilon^{\gamma/2}} \right) = \lim_{\varepsilon \downarrow 0} \varepsilon^\gamma \log \left(P \left(V_t^{\eta, H-, T} \geq \frac{1}{\varepsilon^{\gamma/2}} \right)^{1/a^2} \right) \quad (2.30)$$

Indeed, (2.30) 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:

$$V_t^{g, T, \varepsilon} := \varepsilon^{\gamma/2} V_t^{g, T}, \quad \tilde{V}_t^{g, T, \varepsilon} := V_t^{g, T, \varepsilon} - \frac{\varepsilon^\gamma}{2} \int_0^t g^2(t, u) du + \varepsilon^{\gamma/2}, \quad (2.31)$$

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

$$(\varphi_{1, \xi_0} f)(s, \varepsilon) := \xi_0(s) \exp(f(s, \varepsilon)), \quad (\varphi_2 g)(\varepsilon) := \frac{1}{\Delta} \int_T^{T+\Delta} g(s, \varepsilon) ds. \quad (2.32)$$

Note that in the definition of φ_{1, ξ_0} in (2.32) 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, ξ_0}^{-1} , which is defined as $(\varphi_{1, \xi_0}^{-1} f)(s, \varepsilon) := \log \left(\frac{f(s, \varepsilon)}{\xi_0(s)} \right)$.

Theorem 68. *For any $\gamma > 0$, the sequence of rescaled VIX random variables $(e^{\varepsilon^{\gamma/2}} VIX_{T, \varepsilon^{\gamma/2}})_{\varepsilon \in [0,1]}$ satisfies a pathwise large deviations principle on $\mathcal{C}[0, 1]$, $e^{\varepsilon^{\gamma/2}} VIX_{T, \varepsilon^{\gamma/2}} \sim \text{LDP}(\varepsilon^\gamma, \Lambda^{VIX})$, with*

$$\Lambda^{VIX}(\mathbf{x}) := \inf_{s \in [T, T+\Delta]} \left\{ \Lambda^V \left(\log \left(\frac{y(s, \cdot)}{\xi_0(s)} \right) \right) : \mathbf{x}(\cdot) = (\varphi_2 y)(\cdot) \right\}.$$

Proof. See Appendix 2.9. □

Remark 69. *Using Theorem 68, 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 66 we have that*

$$\lim_{\varepsilon \downarrow 0} \varepsilon^\gamma \log \mathbb{P} \left(VIX_{T, \varepsilon^{\gamma/2}} \geq e^{-\varepsilon^{\gamma/2}} \right) = - \inf_{g \in A} \Lambda^{VIX}(g),$$

for any $\gamma > 0$.

2.6 NUMERICAL RECIPES

We first consider the simple rough Bergomi (2.5) model for sake of simplicity and further develop the mixed multi-factor rough Bergomi (2.7) model in Section 2.6.3 (which also includes (2.6)). Therefore, we tackle the numerical computation of the rate function

$$\hat{\Lambda}^v(\mathbf{y}) := \inf \{ \Lambda^v(\mathbf{x}) : \mathbf{y} = RV(\mathbf{x})(1) \}.$$

This problem, in turn, is equivalent to the following optimisation:

$$\hat{\Lambda}^v(\mathbf{y}) := \inf_{f \in L^2[0,1]} \left\{ \frac{1}{2} \|f\|^2 : \mathbf{y} = RV \left(\exp \left(\int_0^\cdot K_{H_-}(u, \cdot) f(u) du \right) \right) (1) \right\}. \quad (2.33)$$

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 (2.33) may then be approximated via

$$\inf_{a \in \mathbb{R}^{n+1}} \left\{ \frac{1}{2} \|\hat{f}^{(n)}\|^2 : y = RV \left(\exp \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{f}^{(n)}(u) du \right) \right) (1) \right\},$$

where $a = (a_0, \dots, a_n)$. In order to obtain the solution, first the constraint $y = RV \left(\exp \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{f}^{(n)}(u) du \right) \right) (1)$ needs to be satisfied. To accomplish this, we consider anchoring one of the coefficients in $\hat{f}^{(n)}$ such that

$$a_i^* = \operatorname{argmin}_{a_i \in \mathbb{R}} \left\{ \left(y - RV \left(\exp \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{f}^{(n)}(u) du \right) \right) (1) \right)^2 \right\} \quad (2.34)$$

and the constraint will be satisfied for all combinations of the vector $a^* = (a_0, \dots, a_{i-1}, a_i^*, a_{i+1}, \dots, a_n)$. Numerically, (2.34) 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 (2.33) as n tends to infinity. The polynomial basis is particularly convenient since we have that

$$\begin{aligned} & RV \left(\exp \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{f}^{(n)}(u) du \right) \right) (1) \\ &= \int_0^1 \exp \left(\eta \sqrt{2H} \sum_{i=0}^n \frac{a_i s^{H_++i}}{i+1} {}_2F_1(i+1, -H_-, i+2, 1) \right) ds, \end{aligned} \quad (2.35)$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function. In particular one may store the values $\{{}_2F_1(i+1, -H_-, i+2, 1)\}_{i=0}^n$ in the computer memory

and reuse them through different iterations. In addition, the outer integral in (2.35) is efficiently computed using Gauss-Legendre quadrature i.e.

$$\begin{aligned} & RV \left(\left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{f}^{(n)}(u) du \right) \right) (1) \\ & \approx \frac{1}{2} \sum_{k=1}^m \exp \left(\eta \sqrt{2H} \sum_{i=1}^n \frac{a_i \left(\frac{1}{2}(1 + p_k) \right)^{H_+ + i}}{i + 1} {}_2F_1(i + 1, -H_-, i + 2, 1) \right) w_k, \end{aligned}$$

where $\{p_k, w_k\}_{k=1}^m$ are m -th order Legendre points and weights respectively.

2.6.1 CONVERGENCE ANALYSIS OF THE NUMERICAL SCHEME

In Figure 2.10 we report the absolute differences $|\hat{\Lambda}_{n+1}^v(y) - \hat{\Lambda}_n^v(y)|$, for which we observe that the distance decreases as we increase n , the degree of the approximating polynomial. We must emphasize 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 2.10 becomes noisy once this accuracy is obtained. We note that in Figure 2.10 we observe a fast convergence, and with just $n = 5$ we usually obtain accuracy of 10^{-5} .

2.6.2 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

$$\int_0^s K_{H_-}(u, s) u^k du = \frac{u^{H_+ + k} {}_2F_1(k + 1, -H_-, k + 2)}{k + 1},$$

then, for $k = -H_- - 1$ we obtain

$$\int_0^s K_{H_-}(u, s) u^{-H_- - 1} du = \frac{{}_2F_1(-H_-, -H_-, 1 - H_-, 1)}{-H_-},$$

which in turn is a constant that does not depend on the upper integral bound s .

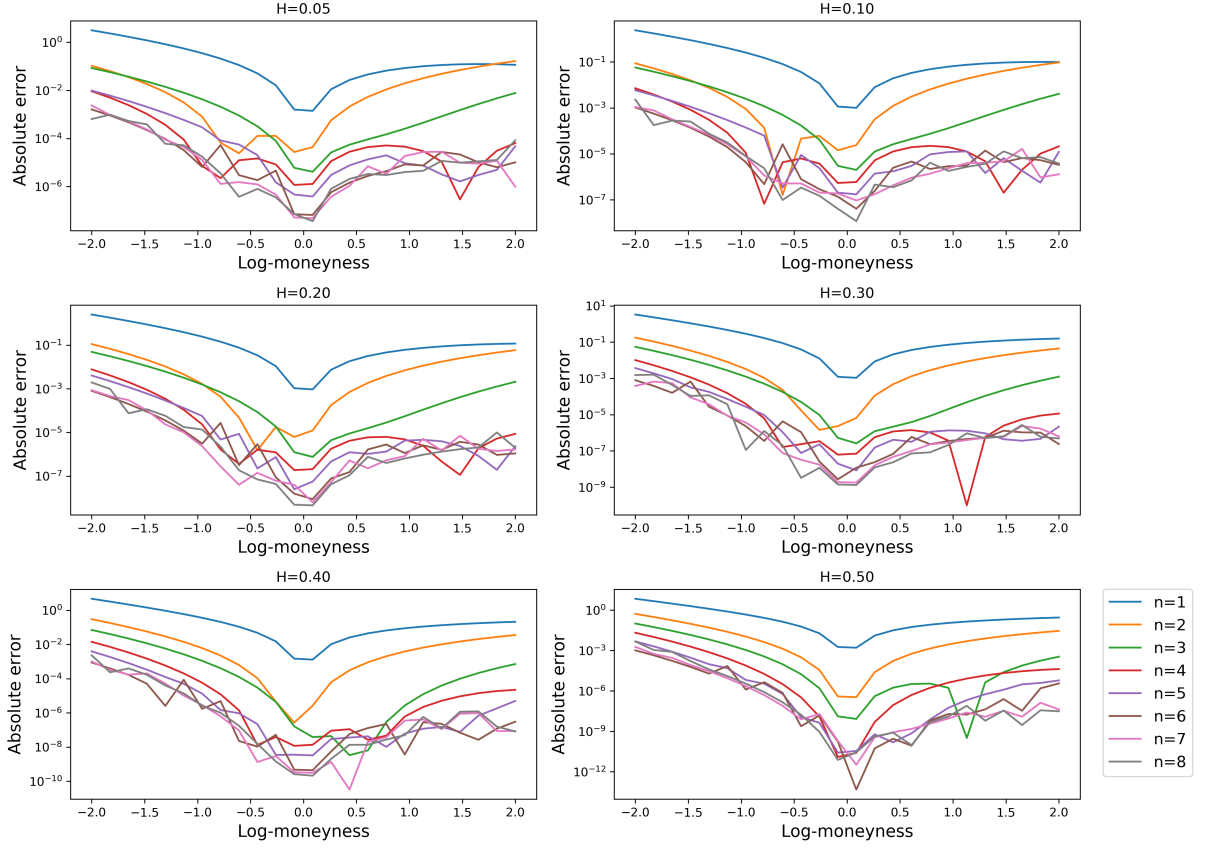


Figure 2.10: Absolute difference of subsequent approximating polynomials $|\hat{\Lambda}_{n+1}^v(y) - \hat{\Lambda}_n^v(y)|$

Proposition 70. Consider the basis $\hat{g}^{(n)}(s) = cs^{-H_- - 1} + \sum_{i=0}^n a_i s^i$ where $c \in \mathbb{R}$. Then, for $c = c^*$ with

$$c^* = \frac{-H_-}{\eta\sqrt{2H_-} {}_2F_1(-H_-, -H_-, 1 - H_-, 1)} \times \log \left(\frac{y}{\int_0^1 \exp \left(\eta\sqrt{2H_-} \left(\sum_{i=0}^n \frac{a_i s^{H_- + i}}{i+1} {}_2F_1(i+1, -H_-, i+2, 1) \right) \right) ds} \right),$$

$\hat{g}^{(n)}(s)$ solves (2.34).

Proof. We have that

$$RV \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{g}^{(n)}(u) du \right) (1) = \exp \left(\frac{\eta \sqrt{2H}}{-H_-} {}_2F_1(-H_-, -H_-, 1 - H_-, 1) \right) \\ \times \int_0^1 \exp \left(\eta \sqrt{2H} \sum_{i=0}^n \frac{a_i s^{H_+ + i} {}_2F_1(i+1, -H_-, i+2, 1)}{i+1} \right) ds,$$

and the proof trivially follows by solving $y = RV \left(\int_0^\cdot K_{H_-}(u, \cdot) \hat{g}^{(n)}(u) du \right) (1)$. $\square$

Remark 71. Notice that Proposition 70 gives a semi-closed form solution to (2.34). 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 (2.33).

Remark 72. Notice that $u^{-H_- - 1} \notin L^2[0, 1]$, however $u^{-H_- - 1} \mathbf{1}_{\{u > \varepsilon\}} \in L^2[0, 1]$ for all $\varepsilon > 0$. Moreover,

$$\int_0^s K_{H_-}(s, u) u^{-H_- - 1} \mathbf{1}_{\{u > \varepsilon\}} du = \frac{{}_2F_1(-H_-, -H_-, 1 - H_-, 1)}{-H_-} \\ - \frac{\varepsilon^{-H_-} s^{H_-} {}_2F_1(-H_-, -H_-, 1 - H_-, \frac{\varepsilon}{t})}{-H_-}, \\ = \frac{{}_2F_1(-H_-, -H_-, 1 - H_-, 1)}{-H_-} + \mathcal{O}(\varepsilon^{-H_-}),$$

hence for ε sufficiently small the error is bounded as long as $H_- \neq 0$. In our applications we find that this method behaves nicely for $H_- \in (-0.5, -0.05]$. In Figure 2.11 we provide precise errors and we observe that the convergence is better for small H_- (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 (2.34), with accuracy measured by absolute error.

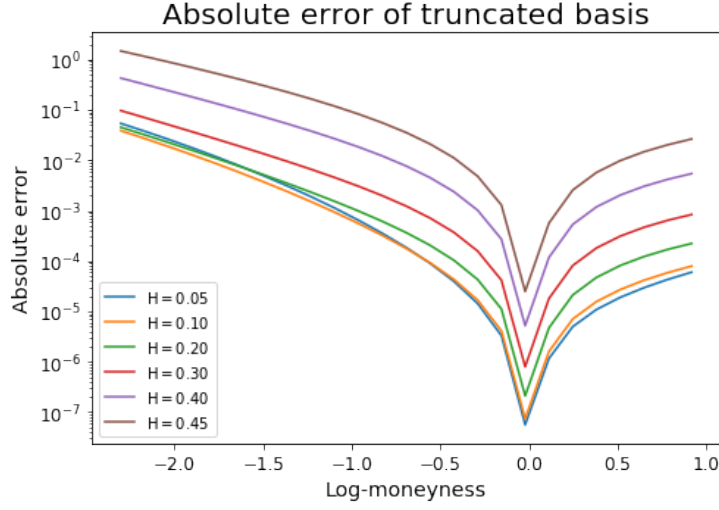


Figure 2.11: Absolute error of the rate function. We consider the truncated basis approach against the standard polynomial basis with $\eta = 1.5$, $v_0 = 0.04$ and different values of H_- .

2.6.3 MULTI-FACTOR CASE

The correlated mixed multi-factor rough Bergomi (2.7) model requires a slightly more complex setting. By Corollary 58 the rate function that we aim at is given by following multidimensional optimisation problem:

$$\hat{\Lambda}^{(v, \Sigma)}(y) := \inf_{(f_1, \dots, f_n) \in L^2[0, 1]} \left\{ \frac{1}{2} \sum_{i=1}^n \|f_i\|^2 : y = RV \left(\sum_{i=1}^m \gamma_i \exp \left(\frac{\nu^i}{\eta} \cdot \Sigma_i \mathfrak{f}^{K_{H_-}} \right) u \right) (1) \right\}, \quad (2.36)$$

where $\mathfrak{f}^{K_{H_-}} = (\int_0^\cdot K_{H_-}(u, \cdot) f_1(u) du, \dots, \int_0^\cdot K_{H_-}(u, \cdot) f_n(u) du)$. The approach to solve this problem is similar to that of (2.33). Nevertheless, in order to solve (2.36) 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 $i \in \{1, \dots, n\}$ is dense as p tends to $+\infty$ in $L^2[0, 1]$ by Stone-Weierstrass Theorem. Then we may equivalently solve

$$\inf_{(a_0^1, \dots, a_p^1, \dots, a_0^n, \dots, a_p^n) \in \mathbb{R}^{(p+1)n}} \left\{ \frac{1}{2} \sum_{i=1}^n \|\hat{f}_i^{(p)}\|^2 : y = RV \left(\sum_{i=1}^m \gamma_i \exp \left(\frac{\nu^i}{\eta} \cdot \Sigma_i \hat{f}^{(K_{H-}, p)} \right) u \right) (1) \right\}, \quad (2.37)$$

where $\hat{f}^{(K_{H-}, p)} = \left(\int_0^\cdot K_{H-}(u, \cdot) \hat{f}_1^{(p)}(u) du, \dots, \int_0^\cdot K_{H-}(u, \cdot) \hat{f}_n^{(p)}(u) du \right)$. Then as p tends to $+\infty$, (2.37) will converge to the original problem (2.36). In order to numerically accelerate the optimisation problem in (2.37), we anchor coefficients $(a_0^1, \dots, a_0^n)$ to satisfy the constraint $y = RV(\cdot)(1)$ (same way we did in the one dimensional case), that is

$$\mathbf{a}^* := \inf_{(a_0^1, \dots, a_0^n) \in \mathbb{R}^n} \left\{ \left(y - RV \left(\sum_{i=1}^m \gamma_i \exp \left(\frac{\nu^i}{\eta} \cdot \Sigma_i \hat{f}^{K_{H-}} \right) u \right) (1) \right)^2 \right\}$$

where $\mathbf{a}^* = (a_0^{1*}, \dots, a_0^{n*})$ and one may use (2.35) and Gauss-Legendre quadrature to efficiently compute $RV(\cdot)(1)$. Then, the constraint will always be satisfied by construction and instead we may solve

$$\inf_{(a_0^{1*}, a_1^1, \dots, a_p^1, \dots, a_0^{n*}, a_1^n, \dots, a_p^n) \in \mathbb{R}^{(p+1)n}} \left\{ \frac{1}{2} \sum_{i=1}^n \|\hat{f}_i^{(p)}\|^2 \right\}. \quad (2.38)$$

2.7 APPROXIMATING THE DENSITY OF REALISED VARIANCE IN THE MIXED ROUGH BERGOMI MODEL

Keller-Ressel and Muhle-karbe [88] defined a discretely sampled realised variance and proposed a method to approximate the price of options on realised variance (RV), based on calculating the price of the option on the quadratic variation using Fourier-Laplace methods. They also determined the exact price of options on realised variance in exponential Lévy models. In our non-affine stochastic volatility setting, and in light of the numerical results shown in Section 2.4 (see Figures 2.3-2.7) 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 conjec-

tured approximation of log-linear smiles and obtain an approximation of the density of $RV(v^{(\gamma, \nu)})(T)$. This conjectured approximation then enables us to price Volatility Swap options in a non-affine stochastic volatility setting and thus extend results on pricing options on realised variance to the rough Bergomi and mixed rough Bergomi models.

Approximation 73. *The implied volatility of realised variance options in the mixed rough Bergomi (2.6) model is linear in log-moneyness, and takes the following form:*

$$\hat{\sigma}(K, T) = \left(T^{2H} \left(a(H_-, \gamma, \nu) + b(H_-, \gamma, \nu) \log \left(\frac{K}{RV(v)(0)} \right) \right) \right)^+,$$

where

$$a(H_-, \gamma, \nu) = \frac{\sqrt{2H} \sum_{i=1}^n \gamma_i \nu_i}{H_+ \sqrt{2H_- + 3}},$$

$$b(H_-, \gamma, \nu) = \sqrt{2H} \left(\frac{\sum_{i=1}^n \gamma_i \nu_i^2}{\sum_{i=1}^n \gamma_i \nu_i} \mathcal{I}(H_-) (2H_- + 3)^{3/2} H_+ - \frac{\sum_{i=1}^n \gamma_i \nu_i}{2H \sqrt{(2H_- + 3)}} \right),$$

with

$$\mathcal{I}(H_-) = \frac{\sum_{n=0}^{\infty} \frac{(H_-)_n}{(H_- + 2)_n} \frac{1 - 2^{-2H_- - 3 - n}}{2H_- + 3 + n}}{H_+ (4H_- + 5)} + \frac{\sum_{n=0}^{\infty} (-1)^n \frac{(-H_-)_n H_+}{(H_- + 2 + n)!} \frac{\hat{F}(n, 1) - 2^{n-1/2-n} \hat{F}(n, 1/2)}{H_+ - n}}{H_+ (4H_- + 5)}$$

such that $\hat{F}(n, x) = {}_2F_1(-n - 2H_+, H_+ - n, H_- + 2 - n, x)$ and $(x)_n = \prod_{i=0}^{n-1} (x + i)$ represents the rising Pochhammer factorial.

Remark 74. *The values of the constants $a(H_-, \gamma, \nu)$ and $b(H_-, \gamma, \nu)$ in Assumption 73, which give the level and slope of the implied volatility respectively, are given in [4, 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 $H_- := H - \frac{1}{2}$.*

Proposition 75. *Under Approximation 73, the density of $RV(v^{(\gamma, \nu)})(T)$ is*

given by

$$\psi_{RV}(x, T) = -\phi(d_2(x)) \frac{\partial d_1(x)}{\partial x} (a(H_-, \gamma, \nu) T^H d_1(x) + 1), \quad x \geq 0,$$

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 [26] tells us that

$$\left. \frac{\partial^2 C(x, T)}{\partial x^2} \right|_{x=K} = \psi_{RV}(K, T).$$

Under Approximation 73, 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

$$\left. \frac{\partial C(x, T)}{\partial x} \right|_{x=K} = v_0 \phi(d_1(K)) \left. \frac{\partial d_1(x)}{\partial x} \right|_{x=K} - x \phi(d_2(K)) \left. \frac{\partial d_2(x)}{\partial x} \right|_{x=K} - \Phi(d_2(K))$$

where

$$\begin{aligned} \frac{\partial d_1(x)}{\partial x} &= \frac{-\hat{\sigma}(x, T) + \log(x/v_0) a(H_-, \gamma, \nu) T^{H-}}{x \hat{\sigma}(x, T)^2 \sqrt{T}} + \frac{1}{2} \frac{a(H_-, \gamma, \nu) T^{H+}}{x} \\ &= \frac{-b(H_-, \gamma, \nu) T^{H-}}{x \hat{\sigma}(x, T)^2 \sqrt{T}} + \frac{1}{2} \frac{a(H_-, \gamma, \nu) T^H}{x} \end{aligned}$$

and

$$\frac{\partial d_2(x)}{\partial x} = \frac{\partial d_1(x)}{\partial x} - \frac{a(H_-, \gamma, \nu) T^H}{x}.$$

Using the well-known identity $v_0\phi(d_1(x)) = x\phi(d_2(x))$, proved in Section 2.9.2, we further simplify

$$\frac{\partial C(K, T)}{\partial K} = v_0\phi(d_1(K)) \left(\frac{a(H_-, \gamma, \nu)T^m}{K} \right) - \Phi(d_2(K)).$$

Differentiating again we obtain,

$$\begin{aligned} \psi_{RV}(K, T) = & -v_0\phi(d_1(K)) \frac{a(H_-, \gamma, \nu)T^H}{K} \left(d_1(K) \frac{\partial d_1(x)}{\partial x} \Big|_{x=K} + \frac{1}{K} \right) \\ & - \phi(d_2(K)) \frac{\partial d_2(x)}{\partial x} \Big|_{x=K}. \end{aligned}$$

Then, by using $v_0\phi(d_1(x)) = x\phi(d_2(x))$, we find that

$$\psi_{RV}(K, T) = -\phi(d_2(K)) \left(a(H_-, \gamma, \nu)T^H \left(d_1(x) \frac{\partial d_1(x)}{\partial x} \Big|_{x=K} + \frac{1}{K} \right) + \frac{\partial d_2(x)}{\partial x} \Big|_{x=K} \right),$$

which we further simplify to

$$\psi_{RV}(K, T) = -\phi(d_2(K)) \frac{\partial d_1(x)}{\partial x} \Big|_{x=K} (a(H_-, \gamma, \nu)T^H d_1(K) + 1),$$

and the result then follows. Note that the density $\psi_{RV}(\cdot, T)$ is indeed continuous for all $T > 0$. $\square$

Remark 76. Note that Proposition 75 gives the density of $RV(v^{(\gamma, \nu)})(T)$ in closed-form. In addition, Proposition 75 can be easily used to get the density of the Arithmetic Asian option under the Black-Scholes model. This would correspond to the case $H_- = 0$ and $\nu = \sigma > 0$ as the Black-Scholes constant volatility.

Remark 77. Assuming the density ψ_{RV} exists, we have the following volatility swap price:

$$\mathbb{E}[\sqrt{RV(v^{(\gamma, \nu)})(T)}] = \int_0^\infty \sqrt{x} \psi_{RV}(x, T) dx.$$

In Figure 2.12, 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.

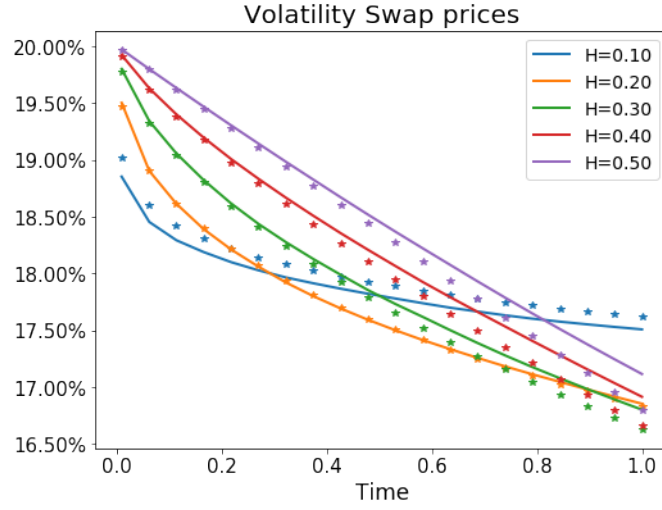


Figure 2.12: Volatility Swap Monte Carlo price estimates (straight lines) and LDP based approximation (stars) for $\eta = 1.5$ and $v_0 = 0.04$; for Monte Carlo we use 200,000 simulations and $\Delta t = \frac{1}{1008}$.

2.8 CONCLUSIONS

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 by Alòs, García-Lorite and Muguruza [4]. 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 paper 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 [13], 2 Factor Bergomi [18, 19] 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 conjectured approximation 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.

2.9 PROOF OF MAIN RESULTS

2.9.1 PROOF OF MAIN LARGE DEVIATIONS RESULTS

Proof of Theorem 53. For brevity we set $n = 2$, but for larger n , identical arguments can be applied. From Schilder's Theorem and Proposition 45, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed ε^{2H} and rate function Λ^Z . Define the operator $f : \mathcal{C}(\mathcal{T}) \rightarrow \mathcal{C}((\mathcal{T}), \mathbb{R}^2)$ by $f(x) := (\frac{\nu_1}{\eta}x, \frac{\nu_2}{\eta}x)$, 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 ε^{2H} and rate function

$$\tilde{\Lambda}(y, z) := \inf\{\Lambda^Z(x) : f(x) = (y, z)\} = \inf\{\Lambda^Z(\frac{\eta}{\nu_1}y) : z = \frac{\nu_2}{\nu_1}y\}.$$

Identical arguments to the proof of Theorem 46 give that the sequences of processes $((\frac{\nu_1}{\eta}Z^\varepsilon, \frac{\nu_2}{\eta}Z^\varepsilon))_{\varepsilon>0}$ and $((\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^{2H}, \frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu_2^2}{2}(\varepsilon\cdot)^{2H}))_{\varepsilon>0}$ are exponentially equivalent, thus satisfy the same large deviations principle, with the same rate function and the same speed.

We now define the operator $g^\gamma : \mathcal{C}(\mathcal{T}, \mathbb{R}^2) \rightarrow \mathcal{C}(\mathcal{T})$ as $g^\gamma(x, y) = v_0(\gamma e^x + (1 - \gamma)e^y)$. For small perturbations $\delta^x, \delta^y \in \mathcal{C}(\mathcal{T})$ we have that

$$\sup_{t \in \mathcal{T}} |g^\gamma(x + \delta^x, y + \delta^y) - g^\gamma(x, y)| \leq |v_0| \left(\sup_{t \in \mathcal{T}} |\gamma e^{x(t)}(e^{\delta^x(t)} - 1)| + \sup_{t \in \mathcal{T}} |(1 - \gamma)e^{y(t)}(e^{\delta^y(t)} - 1)| \right).$$

Clearly the right hand side tends to zero as δ^x, δ^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 $(v^{(\varepsilon, \gamma, \nu)})_{\varepsilon>0} := \left(v_0 \left(\gamma \exp(\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^{2H}) + (1 - \gamma) \exp(\frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu_2^2}{2}(\varepsilon\cdot)^{2H}) \right) \right)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed ε^{2H} and rate function

$$\begin{aligned} x &\mapsto \inf\{\tilde{\Lambda}(y, z) : x = g^\gamma(y, z)\} = \inf\{\Lambda^Z(\frac{\eta}{\nu_1}y) : x = g^\gamma(y, \frac{\nu_2}{\nu_1}y)\}, \\ &= \inf\{\Lambda^Z(\frac{\eta}{\nu_1}y) : x = v_0(\gamma e^y + (1 - \gamma)e^{\frac{\nu_2}{\nu_1}y})\}. \end{aligned}$$

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 46 then yield that $\Lambda^\gamma(v_0) = 0$. $\square$

Proof of Proposition 64. The proof of Proposition 64, which is similar to the proofs given in [82], is made up of three parts. The first part is to prove that $(\mathcal{H}^{g, T}, \langle \cdot, \cdot \rangle_{\mathcal{H}^{g, T}})$ is a separable Hilbert space. Clearly $\mathcal{I}^{g, T}$ is surjective on $\mathcal{H}^{g, T}$. Now take $f_1, f_2 \in L^2(\mathcal{T})$ such that $\mathcal{I}^{g, T}f_1 = \mathcal{I}^{g, T}f_2$. For any $t \in [T, T + \Delta]$ it follows that $\int_0^T g(t, s)[f_1(s) - f_2(s)]ds = 0$; applying the

Titchmarsh convolution theorem then implies that $f_1 = f_2$ almost everywhere and so $\mathcal{I}^{g,T} : L^2(\mathcal{T}) \rightarrow \mathcal{H}^{g,T}$ is a bijection. $\mathcal{I}^{g,T}$ is a linear operator, and therefore $\langle \cdot, \cdot \rangle_{\mathcal{H}^{g,T}}$ is indeed an inner product; hence $(\mathcal{H}^{g,T}, \langle \cdot, \cdot \rangle_{\mathcal{H}^{g,T}})$ is a real inner product space. Since $L^2(\mathcal{T})$ is a complete (Hilbert) space, there exists a function $\tilde{f} \in L^2(\mathcal{T})$ such that the sequence $\{f_n\}_{n \in \mathbb{N}}$ converges to $\tilde{f}$. Assume for a contradiction that $f \neq \tilde{f}$, then, since $\mathcal{I}^{g,T}$ is a bijection, the triangle inequality yields

$$0 < \left\| \mathcal{I}^{g,T} f - \mathcal{I}^{g,T} \tilde{f} \right\|_{\mathcal{H}^{g,T}} \leq \left\| \mathcal{I}^{g,T} f - \mathcal{I}^{g,T} f_n \right\|_{\mathcal{H}^{g,T}} + \left\| \mathcal{I}^{g,T} \tilde{f} - \mathcal{I}^{g,T} f_n \right\|_{\mathcal{H}^{g,T}},$$

which converges to zero as n tends to infinity. Therefore $f = \tilde{f}$, $\mathcal{I}^{g,T} f \in \mathcal{H}^{g,T}$ and $\mathcal{H}^{g,T}$ is complete, hence a real Hilbert space. Since $L^2(\mathcal{T})$ 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 $\mathcal{H}^{g,T}$, which is then separable.

The second part of the proof is to show that there exists a dense embedding $\iota : \mathcal{H}^{g,T} \rightarrow \mathcal{C}[T, T + \Delta]$. Since there exists $h \in L^2(\mathcal{T})$ such that $\int_0^\varepsilon |h(s)| ds$ for all $\varepsilon > 0$ and $g(t, \cdot) = h(t - \cdot)$ for any $t \in [T, T + \Delta]$, we can apply by [32, Lemma 2.1], which tells us that $\mathcal{H}^{g,T}$ 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 $f^* \in \mathcal{C}[T, T + \Delta]$ is Gaussian on $\mathcal{C}[T, T + \Delta]$, with variance $\|\iota^* f^*\|_{\mathcal{H}^{g,T^*}}$, where ι^* is the dual of ι . Take $f^* \in \mathcal{C}[T, T + \Delta]^*$, then [82, Remark 3.2] implies that ι^* admits an isometric embedding $\bar{\iota}^*$ such that

$$\|\bar{\iota}^* f^*\|_{\mathcal{H}^{g,T^*}} = \|f^*\|_{L^2(\mathcal{C}[T, T + \Delta], \mu)} = \int_{\mathcal{C}[T, T + \Delta]} (f^*)^2 d\mu = \text{VAR}(f^*),$$

where μ is the Gaussian measure induced by the process on $(\mathcal{C}[T, T + \Delta], \mathcal{B}(\mathcal{C}[T, T + \Delta]))$. □

Proof of Theorem 68. First we recall $\tilde{V}_t^{g,T,\varepsilon} := V_t^{g,T,\varepsilon} - \frac{\varepsilon^\gamma}{2} \int_0^t g^2(t, u) du + \varepsilon^{\gamma/2}$. We begin the proof by showing that the sequence of processes $(V_t^{g,T,\varepsilon})_{\varepsilon \in [0,1]}$ and $(\tilde{V}_t^{g,T,\varepsilon})_{\varepsilon \in [0,1]}$ are exponentially equivalent, see Definition 8. As $g(t, \cdot) \in L^2(\mathcal{T})$ for all $t \in [T, T + \Delta]$, for each $\delta > 0$ there exists $\varepsilon_* > 0$ such that $\sup_{t \in [T, T + \Delta]} \left| \varepsilon_*^{\gamma/2} - \frac{\varepsilon_*^\gamma}{2} \int_0^T g^2(t, u) du \right| \leq \delta$. Hence, for the $\mathcal{C}[T, T + \Delta]$ norm

$\|\cdot\|_\infty$ we have that for all $\varepsilon_* > \varepsilon > 0$,

$$\mathbb{P}\left(\left\|V^{g,T,\varepsilon} - \tilde{V}^{g,T,\varepsilon}\right\|_\infty > \delta\right) = \mathbb{P}\left(\sup_{t \in [T, T+\Delta]} \left|\varepsilon^{\gamma/2} - \frac{\varepsilon^\gamma}{2} \int_0^T g^2(t, u) du\right| > \delta\right) = 0.$$

Therefore $\limsup_{\varepsilon \downarrow 0} \varepsilon^\gamma \log \mathbb{P}\left(\left\|V^{g,T,\varepsilon} - \tilde{V}^{g,T,\varepsilon}\right\|_\infty > \delta\right) = -\infty$, and so the two sequences of processes $(V^{g,T,\varepsilon})_{\varepsilon \in [0,1]}$ and $(\tilde{V}^{g,T,\varepsilon})_{\varepsilon \in [0,1]}$ are exponentially equivalent; applying Theorem 9 then yields that $(\tilde{V}^{g,T,\varepsilon})_{\varepsilon \in [0,1]}$ satisfies a large deviations principle on $\mathcal{C}[T, T + \Delta]$ with speed ε^γ 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 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 $\delta^f \in \mathcal{C}([T, T + \Delta] \times [0, 1])$:

$$\left\|\varphi_2(f + \delta^f) - \varphi_2(f)\right\|_\infty = \sup_{\varepsilon \in [0,1]} \left|\frac{1}{\Delta} \int_T^{T+\Delta} \delta^f(s, \varepsilon) ds\right| \leq M,$$

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, recalled in Theorem 10 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 $\left(\left(\varphi_{1,\xi_0} \tilde{V}_s^{g,T,\varepsilon}\right)(s, \varepsilon)\right)_{\varepsilon \in [0,1]}$ satisfies a large deviations

principle on $\mathcal{C}[0, 1]$ as ε tends to zero with speed ε^γ and rate function

$$\hat{\Lambda}_s^V(y) := \Lambda^V((\varphi_{1,\xi_0}y)^{-1}(s, \cdot)) = \Lambda^V\left(\log\left(\frac{y(s, \cdot)}{\xi_0(s)}\right)\right).$$

A second application of the Contraction Principle then yields that the sequence of stochastic processes $\left((\varphi_2(\varphi_{1,\xi_0}\tilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\right)_{\varepsilon \in [0,1]}$ satisfies a large deviations principle on $\mathcal{C}[0, 1]$ with speed ε^γ and rate function $\Lambda^{\text{VIX}}(x) = \inf_{s \in [T, T+\Delta]} \{\Lambda^V((\varphi_{1,\xi_0}y)^{-1}(s, \cdot)) : x(\cdot) = (\varphi_2y)(\cdot)\}$. By definition, the sequence of processes $\left((\varphi_2(\varphi_{1,\xi_0}\tilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\right)_{\varepsilon \in [0,1]}$ is almost surely equal to the rescaled VIX processes $(e^{\varepsilon^{\gamma/2}}\text{VIX}_{T,\varepsilon^{\gamma/2}})_{\varepsilon \in [0,1]}$ and hence the satisfies the same large deviations principle. $\square$

2.9.2 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

$$\begin{aligned} (d_1(x))^2 - (d_2(x))^2 &= (d_1(x))^2 - \left(d_1(x) - \hat{\sigma}(x,T)\sqrt{T}\right)^2, \\ &= 2d_1(x)\hat{\sigma}(x,T)\sqrt{T} - T\hat{\sigma}^2(x,T), \\ &= 2\left[\log(v_0) - \log(x) + \frac{1}{2}\hat{\sigma}^2(x,T)T\right] - T\hat{\sigma}^2(x,T), \\ &= 2\log\left(\frac{v_0}{x}\right). \end{aligned}$$

$\square$

3

Small-time asymptotics for complete stochastic volatility models

In this Chapter, we study the complete model with stochastic volatility introduced by Guyon [71]. We first prove the Feller conditions on the coefficients to obtain existence and uniqueness of a strong solution for the model, characterising its behaviour at boundary points. We then use large deviations techniques to prove the small-time asymptotic behaviours. Finally, we present a numerical algorithm to estimate the coefficients of the diffusion coefficient in order to fit them to the S&P 500 and VIX market data.

In the recent years, academics showed that rough volatility models reproduce the power-law decay of the volatility smiles close to maturity accurately. These models are based on fractional Brownian motions. In particular, these Gaussian processes have infinite dependence on the past, which reflects the real-world perception of market volatility. The most well-known form of path-dependent volatility was introduced by Engle [45] and Bollerslev [25] in the early 80s. They respectively developed the (generalised) autoregressive conditional heteroscedasticity (ARCH and GARCH) models for financial time series modelling. In these models, the conditional variance of the log-prices depends on a set of exogenous and lagged endogenous variances and past residuals. Duan [43] discussed such discrete-time models and their application to option pricing. Although these models are still widely used by econometricians, in the late 90s, Hobson and Rogers [75] developed a new class of non-constant volatility models, where the instantaneous volatility is an endogenous factor defined using exponentially weighted moments of historical log-prices. Contrary to stochastic volatility models introduced earlier in this thesis, in these past-dependent continuous-time models, the market is complete as no additional source of randomness has been introduced. Hobson and Rogers [75] found the partial differential equation satisfied by a European Call option price and proved that such models generate smiles and skews of implied volatility consistent figurewith market data. The instantaneous volatility process is thus self-reinforcing, that is the volatility increases when we observe large movements in the recent past, which reflects the market perception of volatility. On the theoretical side, Mohammed and Zhang [100] proved the existence and uniqueness of strong solutions for stochastic systems with finite memory.

More recently, Guyon [70] focused on path-dependent volatility. Even if this family of models is not widely used in the industry, it does not only combine the advantages of the stochastic volatility models introduced in the previous Chapters, but adds other benefits. The volatility process σ usually depends on the current value of the stock price process S_t and a set Y_t of finite path-dependent variables, such as averages, maximums, minimums, realised

variances, etc. The general set-up studied by Guyon [70] is:

$$dS_t = S_t \sigma(t, S_t, Y_t) l(t, S_t) dW_t, \quad S_0 := s_0 > 0,$$

with W a standard Brownian motion, l the leverage function to be calibrated and Y a set of path-dependent variables to determine. Similarly to local stochastic volatility models, a path-dependent volatility is defined as $\sigma(t, S_t, Y_t) l(t, S_t)$, with l calibrated to perfectly fit the market smile of the stock price process S . The idea is therefore to choose Y and σ , and to calibrate l such that the model has the desired properties and fits to the volatility smile. In [71], Guyon selects the process Y of finite path-dependent variables to be an exponentially weighted moving average of the stock price process, similarly to Hobson and Rogers [75]. A stochastic differential equation is then obtained for Y . The existence and uniqueness of a strong solution, however, needs to be proved carefully. Moreover, one can assume a linear relation between VIX, an approximation of the instantaneous volatility, and VVIX, an approximation of the instantaneous volatility of the instantaneous volatility, as illustrated in Figure 3.1, and obtain the form for the diffusion function σ .

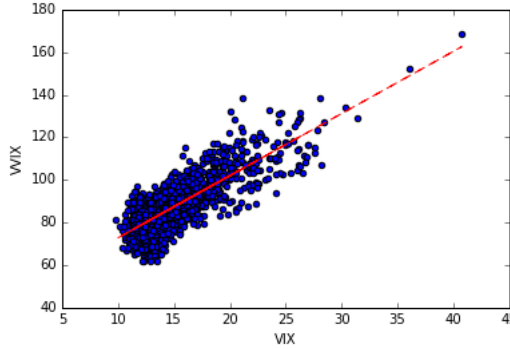


Figure 3.1: Historical VVIX vs historical VIX (13/04/2012 - 08/05/2017).

Source: CBOE VIX Index and VVIX Index historical data.

Finally, one needs to calibrate the leverage function to perfectly fit the volatility smile, using:

$$\mathbb{E}^{\mathbb{Q}}[\sigma^2(t, S_t, Y_t) | S_t] l^2(t, S_t) = \sigma_{loc}^2(t, S_t),$$

with σ_{loc} corresponding to Dupire's equation [44] for the local volatility, and $\mathbb{Q}$ the risk-neutral probability. The expectation can be approximated using the Particle Method, introduced by Guyon and Henry-Labordère [72], which enables to compute the leverage function l by simulating the paths.

Therefore, so-called path-dependent volatility models are complete, as local volatility models, and generate a unique price for options. Moreover, the model is perfectly calibrated to the volatility smile, thanks to the leverage function l . When choosing the volatility function σ , one can actually reach spot-volatility dynamics not attainable by stochastic local volatility models, thanks to the degrees of freedom present in the finite path-dependent variable Y . Besides, these models can capture the historical features of volatilities and asset prices, and reconcile the implied calibration with the historical calibration. Hence, such volatility models combine both advantages of local and stochastic volatility models and offer a range of new properties. The model introduced by Guyon [71] produces rich joint dynamics of spot, implied volatility and implied volatility of volatility, and captures the historical behaviour of S&P 500, VIX and VVIX. It extends the idea from [74] that volatility is readily observable from the history of returns in asset prices to the continuous-time setting. Heston and Nandi [74] proved that their setting is numerically close to continuous-time stochastic volatility models and easier to compute, on a daily frequency only. Guyon's model allows for intraday frequency and as the process (S, Y) obtained is Markovian, it is also easy to compute compared to other stochastic volatility models.

The aim of this Chapter is therefore to study the model introduced by Guyon [71] and to derive the small-time behaviour of the system, using large deviations techniques. In this direction, various large deviations results were proved for systems with memory, outside the scope of financial mathematics. Azencott, Geiger and Ott [9] proved a small-noise large deviations result for a system with finite discrete-time delay. More recently, Ma, Ren, Touzi and Zhang [92] obtained a large deviations principle for non-Markovian stochastic differential equations with random coefficients. In particular, in [92, Section 3.2], they provide short-maturity asymptotics for the log-stock price process. One major difference with the current Chapter is that they assume the dif-

fusion process σ to be Lipschitz continuous, strictly positive, and bounded, which is not the case here. To prove our results, one first needs to determine conditions on the coefficients of the diffusion σ , to obtain existence and uniqueness of a strong solution. Then, similar techniques to Peithmann [105] will enable to obtain small-time asymptotics for the path-dependent process Y as well as the stock price process S .

The structure of the Chapter is as follows. In Section 3.1, we set the notations and present the model of interest. Section 3.2 gathers the main theoretical results, proving existence and uniqueness of a strong solution (Section 3.2.1), deriving the stationary distribution (Section 3.2.2) and presenting pathwise large deviations estimates in Section 3.2.3, from which implied volatility asymptotics follow readily. Finally, in Section 3.3 we construct a numerical scheme to estimate the coefficients of the system for a joint calibration to S&P 500 and VIX market data.

3.1 SET UP AND NOTATIONS

The underlying process S , describing the evolution of the S&P index follows the general dynamics

$$\frac{dS_t}{S_t} = \sigma(Y_t)dW_t, \quad S_0 := s_0 > 0,$$

for some given Brownian motion W generating a filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$, where $\sigma : \mathbb{R}_+^* \rightarrow \mathbb{R}$ is non anticipative. Following Guyon [71] and Hobson and Rogers [75], we assume that the process Y is adapted to $\mathcal{F}$ and is a function of the past history of the stock S , making the latter non-Markovian, in the sense

$$Y_t := \frac{S_t}{\bar{S}_t^h}, \quad \text{for } t \in \mathcal{T}, \quad \text{where } \bar{S}_t^h := \frac{1}{h} \int_{-\infty}^t \exp \left\{ -\frac{t-u}{h} \right\} S_u du \quad (3.1)$$

is the exponential weighted moving average (EWMA) of the stock price process. The constant $h > 0$, denoting the length of the time window, is left unspecified for now. Using Itô's formula, we can summarise the dynamics

for the couple (S, Y) as

$$\begin{aligned} dS_t &= S_t \sigma(Y_t) dW_t, & S_0 &= s_0 > 0, \\ dY_t &= b(Y_t) dt + \tilde{\sigma}(Y_t) dW_t, & Y_0 &= y_0 > 0, \end{aligned} \quad (3.2)$$

with $b(y) := \frac{1}{h}y(1-y)$ and $\tilde{\sigma}(y) := y\sigma(y)$ for $y > 0$. Guyon [71] showed that, for the linear relationship between the VIX and the VVIX to hold, one needs to consider a diffusion coefficient of the form

$$\sigma(x) := -\frac{\alpha}{\beta} + \gamma x^{-\beta}, \quad (3.3)$$

with $\alpha, \beta, \gamma > 0$. In that case, $\tilde{\sigma}$ is null at $y_\sigma := \left(\frac{\beta\gamma}{\alpha}\right)^{1/\beta}$, and

$$\tilde{\sigma}(0) = \begin{cases} \text{not defined,} & \text{if } \beta > 1, \\ 0, & \text{if } \beta < 1, \\ \gamma, & \text{if } \beta = 1. \end{cases}$$

Before diving into the asymptotic behaviour of the process, we first concentrate on the existence and uniqueness of a strong solution for (3.2), depending on the values of α , β and γ , and classify the singular points 0, y_σ and infinity.

3.2 THEORETICAL ANALYSIS OF THE MODEL

3.2.1 BOUNDARY CONDITIONS

We start the analysis of the model by a detailed study of the behaviour of the system (3.2) at the boundary points 0, y_σ and ∞ . Recall [33, Definition 2.3] that a point $d \in \mathbb{R}$ is called singular if $\frac{1+|b|}{\tilde{\sigma}^2}$ is not locally integrable in a neighbourhood of d . A point that is not singular is called regular. A singular point is further called isolated if it admits a deleted neighbourhood consisting of regular points. Introduce the hitting times $\tau_x := \inf \{t \geq 0 : Y_t = x\}$ and $\tau_{x,y} := \min(\tau_x, \tau_y)$. The precise classification of boundary points is rather delicate and we only give below an informal help, and refer the interested reader to [33, Sections 2.3 and 2.4] for full and precise details:

Definition 78. Let $a > 0$.

- A point x is called a regular boundary and is said to have right-type 0, and we write $x \in \mathcal{B}_{0+}^a$, if for any $y_0 \in [x, x+a]$, there exists a unique solution defined up to $\tau_{x,a}$. This solution reaches x with strictly positive probability, $\mathbb{E}(\tau_{x,x+a})$ is finite, and $\mathbb{P}(Y_{\tau_{x,x+a}} = 0) > 0$.
- A point x is called an exit boundary and is said to have right-type 1, and we write $x \in \mathcal{B}_{1+}^a$, if for any $y_0 \in [x, x+a]$, there exists a unique solution defined up to $\tau_{x,a}$. This solution reaches x with strictly positive probability. Any solution may only leave x in the left direction.
- A point x is called an entrance boundary and is said to have right-type 3 if for any $y_0 \in (x, x+a]$, there exists a unique solution defined up to τ_{x+a} . This solution never reaches x and $\mathbb{E}(\tau_{x+a})$ is finite. We write $x \in \mathcal{B}_{3+}^a$.
- Infinity is called a recurrent boundary point if a solution cannot explode there. Moreover, if there are no singular points between y_0 and a point $z < y_0$, then the solution reaches z almost surely.

The left-types $\mathcal{B}_{1-}^a$ and $\mathcal{B}_{3-}^a$ are defined similarly.

The following proposition, proved in Section 3.4.1, characterises the behaviour of the solution to (3.2) at the origin and at infinity.

Proposition 79. *Infinity is a recurrent boundary, and the origin is regular if and only if $\beta > \frac{1}{2}$, and*

- for $\beta > \frac{1}{2}$, zero is an exit, non-entrance boundary point and belongs to $\mathcal{B}_{0+}^{y_\sigma}$;
- for $0 \leq \beta \leq \frac{1}{2}$, zero is an exit, non-entrance boundary point and belongs to $\mathcal{B}_{1+}^{y_\sigma}$.

The following theorem, proved in Section 3.4.2, is more involved and fully characterises the solution to the SDE (3.2).

Theorem 80. y_σ is an isolated singular point, and for $a > 0$, the following holds:

	Left boundary of $y_\sigma > y_0$	Right boundary of $y_\sigma < y_0$
$y_\sigma < 1$	$\mathcal{B}_{1-}^a$	$\mathcal{B}_{3+}^a$
$y_\sigma \geq 1$	$\mathcal{B}_{3-}^a$	$\mathcal{B}_{1+}^a$

3.2.2 STATIONARY DISTRIBUTION AND PRICING PDE

We now prove that the process Y introduced in (3.2) admits a stationary distribution and we derive the pricing PDE associated to (3.2).

STATIONARY DISTRIBUTION

We recall [54, Section 3.2] that a process $(Y_t)_{t>0}$ is ergodic if it admits a unique, stationary distribution Π , and for any measurable bounded function g , the almost sure limit

$$\lim_{t \uparrow \infty} \frac{1}{t} \int_0^t g(Y_s) ds = \int g(y) \Pi(dy)$$

holds. If it exists, an ergodic solution must satisfy $\mathcal{L}^* \Pi = 0$, where $\mathcal{L}^*$ is the adjoint of the infinitesimal generator $\mathcal{L}$, defined [54, Section 1.5.3] via the equality

$$\int \Psi(\xi) \mathcal{L} \phi(\xi) d\xi = \int \phi(\xi) \mathcal{L}^* \Psi(\xi) d\xi, \quad (3.4)$$

for any rapidly decaying smooth test functions ϕ and Ψ .

Proposition 81. *The infinitesimal generator $\mathcal{L}_Y$ of Y in (3.2) and its dual $\mathcal{L}_Y^*$ read*

$$(\mathcal{L}_Y f)(y) = \frac{1}{h} y(1-y) \frac{\partial}{\partial y} + \frac{1}{2} y^2 \sigma^2(y) \frac{\partial^2}{\partial y^2}, \text{ and}$$

$$(\mathcal{L}_Y^* f)(y) = -\frac{\partial}{\partial y} \left(\frac{1}{h} y(1-y) f(y) \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (y^2 \sigma^2(y) f(y)).$$

Proof. Using (3.4) and integration by part, we obtain the expression for $\mathcal{L}^*$. Given $y \in \mathbb{R}$ and $f, g : \mathbb{R} \rightarrow \mathbb{R}$ twice continuously differentiable functions

with bounded derivatives, we have

$$\begin{aligned}
\langle f, \mathcal{L}_Y^* g \rangle &= \int_{\mathbb{R}} f(y) \left[-\frac{\partial}{\partial y} \left(\frac{1}{h} y(1-y)g(y) \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (y^2 \sigma^2(y)g(y)) \right] dy, \\
&= - \int_{\mathbb{R}} \frac{\partial f}{\partial y}(y) \left\{ -\frac{1}{h} y(1-y)g(y) + \frac{1}{2} \frac{\partial}{\partial y} (y^2 \sigma^2(y)g(y)) \right\} dy, \\
&= \frac{1}{h} \int_{\mathbb{R}} \frac{\partial f}{\partial y}(y) y(1-y)g(y) dy - \frac{1}{2} \int_{\mathbb{R}} \frac{\partial f}{\partial y}(y) \frac{\partial}{\partial y} (y^2 \sigma^2(y)g(y)) dy, \\
&= \frac{1}{h} \int_{\mathbb{R}} \frac{\partial f}{\partial y}(y) y(1-y)g(y) dy + \frac{1}{2} \int_{\mathbb{R}} \frac{\partial^2 f}{\partial y^2}(y) y^2 \sigma^2(y)g(y) dy, \\
&= \int_{\mathbb{R}} g(y) \left[\frac{1}{h} y(1-y) \frac{\partial f}{\partial y}(y) + \frac{1}{2} y^2 \sigma^2(y) \frac{\partial^2 f}{\partial y^2}(y) \right] dy, \\
&= \langle \mathcal{L}_Y f, g \rangle.
\end{aligned}$$

The rapidly decaying smooth test functions f and g ensure that the boundary terms in the integration by parts are equal to zero. $\square$

In our setting, the process Y in (3.2) admits at least one stationary distribution, which can be proved easily following the arguments in [103]. However, since the map σ is not generally bounded away from zero, one cannot use classical results to prove uniqueness of the solution to the Poisson equation $\mathcal{L}^* \Phi = 0$. We leave to future endeavours this analysis.

Proposition 82. *The SDE (3.2) admits at least one stationary distribution Φ .*

Remark 83. *For $f : \mathbb{R} \rightarrow \mathbb{R}$, finding the explicit solution of the Poisson equation is tedious. Indeed,*

$$(\mathcal{L}_Y^* f)(y) = -\frac{\partial}{\partial y} \left(\frac{1}{h} y(1-y)f(y) \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (y^2 \sigma^2(y)f(y)) = 0,$$

is equivalent to

$$\begin{aligned} & \frac{1}{2}y^2 f''(y) \left[\left(\frac{\alpha}{\beta} \right)^2 - \frac{2\alpha\gamma}{\beta} y^{-\beta} + \gamma^2 y^{-2\beta} \right] \\ & + y f'(y) \left[2 \left\{ \left(\frac{\alpha}{\beta} \right)^2 - \frac{\alpha\gamma}{\beta} (2 - \beta) y^{-\beta} + (1 - \beta) \gamma^2 y^{-2\beta} \right\} - \frac{1}{h} (1 - y) \right] \\ & + f(y) \left[\left(\frac{\alpha}{\beta} \right)^2 + \frac{\alpha\gamma}{\beta} (1 - \beta) (2\alpha + \beta) y^{-\beta} + \gamma^2 (\beta - 1) (\alpha^2 (\beta - 1) + \beta) y^{-2\beta} - \frac{1}{h} (1 - 2y) \right] = 0, \end{aligned}$$

with the constraint $\int_{\mathbb{R}} f(y) dy = 1$. This is a highly non-linear problem, which does not admit any explicit solution in general. Nevertheless, the solution can be written in the integral form associated with the first order ODE satisfied by the stationary measure.

PRICING PARTIAL DIFFERENTIAL EQUATION

Consider an option with payoff $H(S_T, Y_T)$ at expiry T , and denote its price $P(t, S_t, Y_t)$ at time $t \leq T$. Since the market is complete, we can construct a self-financing and riskless replicating portfolio consisting of the derivative itself and $-(SP_S + YP_Y)/S$ units of the underlying asset S to find the corresponding pricing PDE.

Proposition 84. *Under the risk-neutral measure $\mathbb{Q}$, the pricing PDE associated to (3.2) is*

$$\left(\partial_t + rs\partial_S + \left(\frac{1-y}{h} + r \right) y\partial_y + \frac{s^2}{2}\sigma^2(y)\partial_{ss} + sx\sigma^2(y)\partial_{sy} + \frac{y^2}{2}\sigma^2(y)\partial_{yy} - r \right) P(t, s, y) = 0, \quad (3.5)$$

for all $(t, s, y) \in [0, T) \times (0, \infty)^2$, with terminal condition $P(T, s, y) = H(s, y)$.

Equation (3.5) can be rewritten as $(\mathcal{L}_Y + \mathcal{L}_1 + \mathcal{L}_2) P = 0$, with $\mathcal{L}_Y$ defined

in Proposition 81 and

$$\begin{cases} \mathcal{L}_1 := sy\sigma^2(y)\frac{\partial^2}{\partial s\partial y} + ry\frac{\partial}{\partial y}, \\ \mathcal{L}_2 := \frac{\partial}{\partial t} + \frac{s^2}{2}\sigma^2(y)\frac{\partial^2}{\partial s^2} + rs\frac{\partial}{\partial s} - r. \end{cases}$$

The operator $\mathcal{L}_2$ is the Black-Scholes infinitesimal generator with volatility $\sigma(y)$. Unfortunately, this pricing PDE does not admit an obvious explicit solution. Approximate solutions can be found by expanding the solution using perturbation methods, as developed in [54], but we leave this to future endeavours.

3.2.3 SMALL-TIME ASYMPTOTICS

We now investigate the small-time behaviour of the solution to (3.2) using large deviations techniques. Consider the log-stock price process $X := \log S$. For $\varepsilon > 0$ and $t \in \mathcal{T}$, introduce the small-time rescaling $(X_t^\varepsilon, Y_t^\varepsilon) := (X_{\varepsilon t}, Y_{\varepsilon t})$. The model becomes

$$\begin{cases} dX_t^\varepsilon = -\frac{\varepsilon}{2}\sigma^2(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t, & X_0^\varepsilon := x_0 = \log(S_0), \\ dY_t^\varepsilon = \varepsilon b(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\tilde{\sigma}(Y_t^\varepsilon)dW_t, & Y_0^\varepsilon = y_0 > 0. \end{cases} \quad (3.6)$$

Let $\overline{\mathcal{H}}$ denote the space of absolutely continuous functions starting at the origin, with square integrable derivatives, such that

$$\overline{\mathcal{H}} := \left\{ f : \mathcal{T} \rightarrow \mathbb{R} \mid f_t = \int_0^t g(s)ds, g \in L^2(\mathcal{T}), \text{ and } \inf_{t \in \mathcal{T}} f_t \geq \frac{1}{\alpha} \log \left(1 - y_0^\beta \frac{\alpha}{\beta\gamma} \right) \right\}.$$

Remark 85. When $y_0 \geq y_\sigma$, the condition

$$\inf_{t \in \mathcal{T}} f_t \geq \frac{1}{\alpha} \log \left(1 - y_0^\beta \frac{\alpha}{\beta\gamma} \right), \quad (3.7)$$

is automatically satisfied and the space $\overline{\mathcal{H}}$ boils down to the usual Cameron-Martin space. When $y_0 < y_\sigma$, Condition (3.7) is needed to ensure that the solution of the controlled ODE introduced below is positive.

The main result here is the following theorem, proved in Section 3.4.5, which states a pathwise large deviations principle for the log-stock price process. With $\mathbf{x}_0 := (x_0, y_0)$, introduce the map $I^{X,Y}$ on $\mathcal{C}(\mathcal{T}, \mathbb{R} \times \mathbb{R}_+^*)$ by

$$I^{X,Y}(g) := \inf \left\{ \Lambda(f), f \in \overline{\mathcal{H}}, \mathcal{S}^{\mathbf{x}_0}(f) = g \right\},$$

where Λ is the usual rate function of the standard Brownian motion:

$$\Lambda(f) := \begin{cases} \frac{1}{2} \int_0^T \|\dot{f}_t\|^2 dt, & \text{if } f \in \mathcal{H}, \\ +\infty, & \text{otherwise,} \end{cases}$$

and $\mathcal{S}^{\mathbf{x}_0}(f)$ on $\mathcal{T}$ is the solution to the controlled ODE $\dot{g}_t = (\sigma(g_t), \tilde{\sigma}(g_t))^\top \cdot (\dot{f}_t, \dot{f}_t)$, starting from $g_0 = \mathbf{x}_0$.

Theorem 86. *The rescaled log-stock price process X^ε satisfies a pathwise large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R})$ as ε tends to zero with speed ε and rate function*

$$I^X(g) := \inf \left\{ I^{X,Y}(h), h := (g, l), l \in \mathcal{C}(\mathcal{T}, \mathbb{R}_+^*), l_0 = y_0 \right\}, \quad (3.8)$$

The proof of the theorem relies on first obtaining a large deviations principle for the rescaled process Y^ε , which we state below (and defer its proof to Section 3.4.4). Similarly to above, denote $\mathcal{S}^y(f)(t)$ the solution to the controlled ODE $\dot{g}_t = \tilde{\sigma}(g_t)\dot{f}_t$, with $g_0 = y$.

Proposition 87. *The rescaled process Y^ε satisfies a pathwise large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}_+^*)$ as ε tends to zero with speed ε and rate function*

$$I^Y(g) := \inf \left\{ \Lambda(f), f \in \overline{\mathcal{H}}, \mathcal{S}^{y_0}(f) = g \right\}. \quad (3.9)$$

Large deviations have been used extensively in Mathematical Finance in order to derive asymptotic behaviour of the implied volatility [58]. Recall from Chapter 0 that the latter, $\Sigma_t(k)$, is the unique non-negative solution to $\mathcal{C}_{BS}(t, e^k, \Sigma_t(k)) = \mathcal{C}_{obs}(t, e^k)$, with $\mathcal{C}_{obs}(t, e^k)$ the observed Call option prices with maturity t and strike e^k , and $\mathcal{C}_{BS}$ is the Call option price in the Black-

Scholes model.

Corollary 88. *For $y_\sigma \leq \min\{y_0, 1\}$, the small-time behaviour of out-of-the-money Put/Call options is given by*

$$\begin{aligned}\lim_{t \downarrow 0} t \log \mathbb{E} (S_t - e^k)^+ &= -\inf_{y \geq k} I^X(g)|_{g(1)=y}, \quad \text{if } k > 0, \\ \lim_{t \downarrow 0} t \log \mathbb{E} (e^k - S_t)^+ &= -\inf_{y \leq k} I^X(g)|_{g(1)=y}, \quad \text{if } k < 0.\end{aligned}$$

Proof. We prove the lower and upper bounds separately, which turn out to be equal. For simplicity, we introduce the following notation, for all $k > 0$:

$$\tilde{I}^X(k) = \begin{cases} \inf_{y \geq k} I^X(g)|_{g(1)=y}, & \text{if } k > 0, \\ \inf_{y \leq k} I^X(g)|_{g(1)=y}, & \text{if } k < 0. \end{cases}$$

Firstly, as the rate function is continuous, we have for all $k \neq 0$,

$$\lim_{t \downarrow 0} t \log \mathbb{P} (S_t > e^k) = -\tilde{I}^X(k).$$

We only consider $k > 0$, the other case being symmetric.

- (1) The proof of the lower bound is similar to [52, Appendix C].

For any $\delta > 0$,

$$\mathbb{E} (S_t - e^k)^+ \geq e^k \delta \mathbb{P} (S_t > e^{k(1+\delta)}).$$

Using Theorem 86, we have that

$$\begin{aligned}\liminf_{t \downarrow 0} t \log \mathbb{E} (S_t - e^k)^+ &\geq \liminf_{t \downarrow 0} [tk + t \log \delta + t \log \mathbb{P} (S_t > e^{k(1+\delta)})], \\ &\geq -\tilde{I}^X(k(1+\delta)).\end{aligned}$$

Take $\delta \downarrow 0$, by continuity of $\tilde{I}^X(k)$, we obtain the desired lower bound.

- (2) To establish the desired upper bound, we note that for any $q > 1$, we have

$$\mathbb{E} (S_t - e^k)^+ \leq \mathbb{E} \left[\left((S_t - e^k)^+ \right)^q \right]^{1/q} \mathbb{P} (S_t \geq e^k)^{1-1/q}.$$

Thus, we have

$$t \log \mathbb{E} (S_t - e^k)^+ \leq \frac{t}{q} \mathbb{E}(S_t^q) + t \left(1 - \frac{1}{q}\right) \log \mathbb{P} (S_t \geq e^k).$$

Then, note that from Theorem 80, we have that, for $y_\sigma \leq \min\{y_0, 1\}$, the process $(Y_t)_{t \in \mathcal{T}}$ remains in $(y_\sigma, +\infty)$. The map σ is bounded on $(y_\sigma, +\infty)$ and thus using arguments similar to [50, proof of Corollary 2.1], we have $\limsup_{t \downarrow 0} \frac{t}{q} \mathbb{E}(S_t^q) \leq 0$. Therefore, for fixed $q > 1$, we have

$$\limsup_{t \downarrow 0} t \log \mathbb{E} (S_t - e^k)^+ \leq - \left(1 - \frac{1}{q}\right) \tilde{I}^X(k).$$

We finally take q to infinity and obtain the desired upper bound.

□

Corollary 89. *For $y_\sigma \leq \min\{y_0, 1\}$, the small-time asymptotic behaviour of the implied volatility is given by*

$$\lim_{t \downarrow 0} \Sigma_t(k) = \begin{cases} \frac{k^2}{2} \left(\inf_{y \geq k} I^X(g)|_{g(1)=y} \right)^{-1}, & \text{if } k > 0, \\ \frac{k^2}{2} \left(\inf_{y \leq k} I^X(g)|_{g(1)=y} \right)^{-1}, & \text{if } k < 0. \end{cases}$$

Proof. We only consider $k > 0$, the other case being symmetric. It follows from [61, Corollary 7.1] that taking $\varepsilon = t$, from Theorem 86, X satisfies a large deviations principle with speed t and rate function I^X :

$$\lim_{t \downarrow 0} t \log \mathbb{P}(X \geq k) = - \inf_{y \geq k} I^X(g)|_{g(1)=y}.$$

In the Black Scholes model, we have the following small-time implied volatility behaviour:

$$\lim_{t \downarrow 0} t \Sigma_t^2(k) \log \mathbb{P}(X \geq k) = - \frac{k^2}{2},$$

and the result follows from [61, Corollary 7.1].

□

3.3 NUMERICAL ESTIMATION OF THE COEFFICIENTS (α, β, γ)

We now devise an algorithm, using VIX and VVIX data, to find the optimal parameters (α, β, γ) appearing in the definition of the diffusion function σ in (3.3). We consider the continuously monitored formula for the VIX:

$$\text{VIX}_t^2 := \mathbb{E} \left[\frac{1}{\Delta} \int_t^{t+\Delta} d\langle X_s, X_s \rangle ds \mid \mathcal{F}_t \right] = \frac{1}{\Delta} \mathbb{E} \left[\int_t^{t+\Delta} \sigma^2(Y_u) du \mid \mathcal{F}_t \right],$$

where Δ is equal to 30 days. In order to simulate the VIX on $[0, T]$, we must first approximate the integral and the conditional expectation, and we denote $\widehat{\text{VIX}}_t$ its approximation: along a discretisation grid $\mathcal{U}_t := (u_i)_{i=0, \dots, N}$ with $u_i := t + i\frac{\Delta}{N}$, assume that $\sigma(Y_u)$ is constant on each interval, and write

$$\begin{aligned} \text{VIX}_t^2 &= \frac{1}{\Delta} \sum_{i=0}^{N-1} \int_{u_i}^{u_{i+1}} \mathbb{E}_t [\sigma^2(Y_u)] du, \\ &\simeq \frac{1}{\Delta} \sum_{i=0}^{N-1} \mathbb{E}_t [\sigma^2(Y_{u_i})] (u_{i+1} - u_i), \\ &= \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E}_t [\sigma^2(Y_{u_i})] \end{aligned}$$

The second step is to approximate the conditional expectation. Since the process Y is Markovian, then $\mathbb{E} [\sigma^2(Y_{u_i}) \mid \mathcal{F}_t] = \mathbb{E} [\sigma^2(Y_{u_i}) \mid Y_t]$, for any $u_i \in \mathcal{U}_t$. We thus approximate the conditional expectation using the empirical mean, by simulating M paths $(Y^{(1)}, \dots, Y^{(M)})$ of Y , so that

$$\frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E}_t [\sigma^2(Y_{u_i})] \simeq \frac{1}{NM} \sum_{j=1}^M \sum_{i=0}^{N-1} \sigma^2(Y_{u_i}^{(j)}) =: \widehat{\text{VIX}}_t^2. \quad (3.10)$$

Note that, in order to simulate $\widehat{\text{VIX}}_t^2$, one first needs to obtain the starting point Y_t , from which all the points $(Y_{u_i})_{i=1, \dots, N}$ can then be determined. Since

$Y_t := \frac{S_t}{\bar{S}_t^h}$ from (3.1), with

$$\bar{S}_t^h := \frac{1}{h} \int_{-\infty}^t e^{-\frac{t-v}{h}} S_v dv \simeq \frac{1}{h} \sum_{i=0}^{I-1} \int_{v_i}^{v_{i+1}} e^{-\frac{t-v}{h}} S_v dv \simeq \frac{e^{-t/h}}{h} \Delta_t \sum_{i=0}^{I-1} e^{v_i/h} S_{v_i},$$

with $\mathcal{V}_t := (v_i)_{i=0, \dots, I}$ a time-discretisation of $(-\infty, t]$, such that v_0 is the first time for which we can observe the S&P 500, and $v_I = t$; we further set the grid size to be $v_{i+1} - v_i =: \Delta_t$ for all $i = 0, \dots, I$. In our simulations, we will consider v_0 to be 25/04/2008. Empirically, in order for this approximation to be as precise as possible, one needs Δ_t to be as small as possible. However here, dealing with daily data for the S&P500 imposes $\Delta_t = 1$ day. We leave an extended analysis of this with high-frequency data to future work. Hence

$$Y_t \simeq \frac{S_t}{\bar{S}_t^h} \simeq \frac{S_t}{\frac{e^{-t/h}}{h} \Delta_t \sum_{i=0}^{I-1} e^{v_i/h} S_{v_i}} \simeq \frac{h e^{t/h} S_t}{\Delta_t \sum_{i=0}^{I-1} e^{v_i/h} S_{v_i}}. \quad (3.11)$$

Note that for any $t \in [0, T]$, the two time grids $\mathcal{V}_t$ and $\mathcal{U}_t$ do not overlap, the former taking into account time before t , the latter time after t . The algorithm to approximate $(\widehat{\text{VIX}}_t^2)_{t \in [0, T]}$ is as follows:

Algorithm:

- (i) Compute $(Y_t)_{t \in [0, T]}$ using (3.11), with S obtained from S&P 500 data;
- (ii) For $t \in [0, T]$, using Y_t from (i), simulate the auxiliary process $(\widehat{Y}_s)_{s > t}$ on the grid $\mathcal{U}_t$ as

$$\widehat{Y}_{u_{i+1}} := \psi(\widehat{Y}_{u_i}, W_{u_i}) 1_{\{\psi(\widehat{Y}_{u_i}, W_{u_i}) > D\}},$$

where

$$\psi(x, w) := x \left(1 + \frac{\Delta_t}{h} (1 - x) + \sigma(x) \Delta w \right),$$

with $\Delta_t := \frac{1}{252}$ and $\Delta W_{u_i} := W_{u_{i+1}} - W_{u_i} \simeq \frac{1}{\sqrt{252}} \mathcal{N}(0, 1)$, and obtain $\frac{1}{N} \sum_{i=0}^{N-1} \sigma^2(\widehat{Y}_{u_i})$;

- (iii) Repeat step (ii) M times to obtain $\widehat{\text{VIX}}_t^2$ from (3.10);

(iv) Repeat steps (ii)-(iii) on the grid to obtain $\widehat{\text{VIX}}^2$ on $\mathcal{V}_T \cap [0, T]$.

Remark 90. *In the Euler scheme in step (ii) of the algorithm, we introduced the threshold D to prevent the simulations from ‘exploding’ when Y approaches zero. We could alternatively apply the threshold on $\sigma(Y_t)$, but we leave the precise analysis of the (weak or strong) convergence of our discretisation for future work, hints thereabout can be picked from [89].*

3.3.1 SIMULATIONS

We use daily data to simulate VIX^2 , between January 2nd, 2013 and May 8th, 2017 (corresponding to $T = 1200$ days). The simulations are run for $h \in \{180, 30, 5\}$ days with different parameters (α, β, γ) , first taken arbitrarily, but optimised over later. Table 3.1 summarises them:

	$h = 180$	$h = 30$	$h = 5$
α	0.9	2.1	5
β	1.5	1.2	6
γ	0.8	1.9	1.7
N	500	500	500
M	500	500	500
D	0	0.01	1

Table 3.1: Parameters used in the simulations

Note first from (3.2) that the process is mean reverting to 1. The smaller the h , the larger the mean reversion $1/h$, which is clearly visible in Figure 3.2.

Using the algorithm above, the simulations for the VIX^2 , respectively for $h = 180, 30$ and 5 days, are depicted in Figure 3.3.

In all cases, the simulated VIX^2 seems to have jumps similar to those visible in the historical VIX^2 . However, when $h = 5$ days, the variations of the simulated VIX^2 are very small compared to the historical one. In order to obtain similar magnitudes, for $h = 5$ days, we had to consider a threshold $D = 1$, meaning that many simulated paths were discarded, thus creating a clear bias. We also chose $(\alpha, \beta, \gamma) = (5, 6, 1.5)$, giving unrealistic volatility

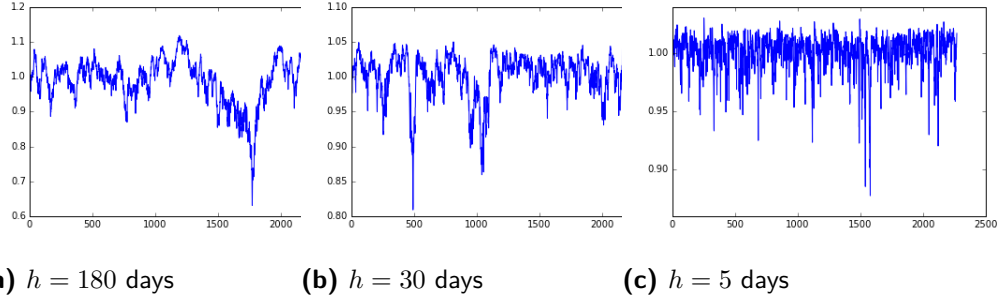


Figure 3.2: Process Y simulated using the Euler scheme between 2/1/2013 and 8/5/2017.

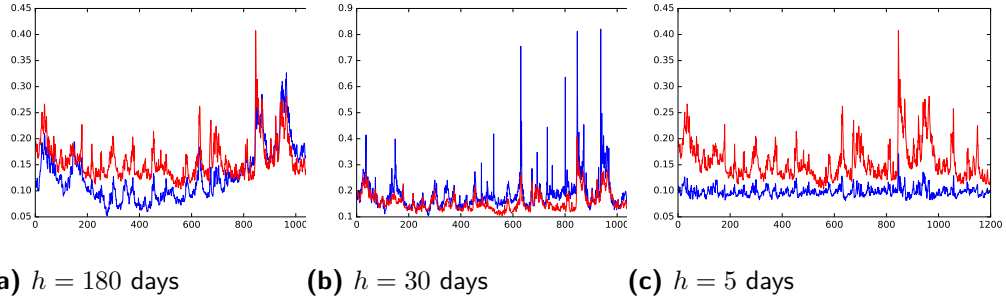


Figure 3.3: Simulated (blue) and historical (red) VIX, between 2/1/2013 and 8/5/2017.

values, for example $\sigma(1) = 0.66$ and confirming that the representation (3.3) works best for $h > \Delta = 30$. In order to analyse the results, we plot in Figure 3.4 the simulated VIX and the historical VIX against the historical VVIX as well as against the recent trend of S&P 500, that is, the process Y simulated using its definition $Y_t := \frac{S_t}{S_t^h}$ for $t \in [0, T]$ and $h \in \{5, 30, 180\}$.

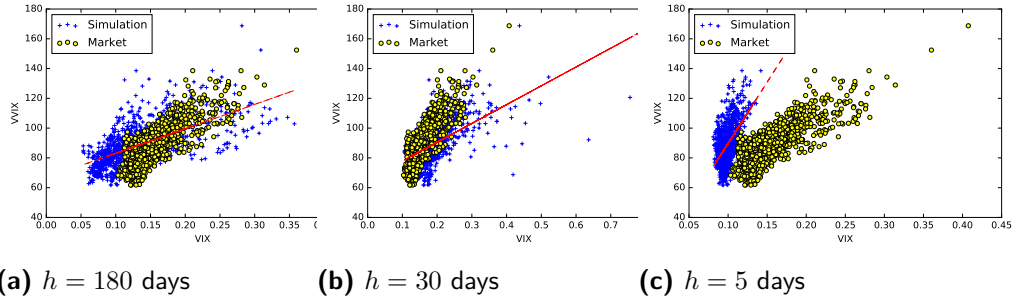


Figure 3.4: Simulated (blue crosses) and historical (yellow circles) VIX against historical VVIX, between 2/1/13 and 8/5/17. The red line is the linear regression fit.

In the first two figures (for $h = 180$ and 30 days), it seems to fit the regression of the simulated data as well. However, for $h = 5$ days, the simulated values clearly follow a different trend. This might be explained by the presence of a large threshold D in the simulations, and by the fact that the approximation for σ using the VIX data is only valid for $h > \Delta$. Finally, Figure 3.5 shows the simulated VIX and historical VIX against the recent trend of S&P 500. The latter are computed using the definition of Y changing the window h to adapt it to each case. We observe that the model does not fit market data for $h = 5$, which is consistent with the assumption that the representation of the function σ is only valid for $h > \Delta$.

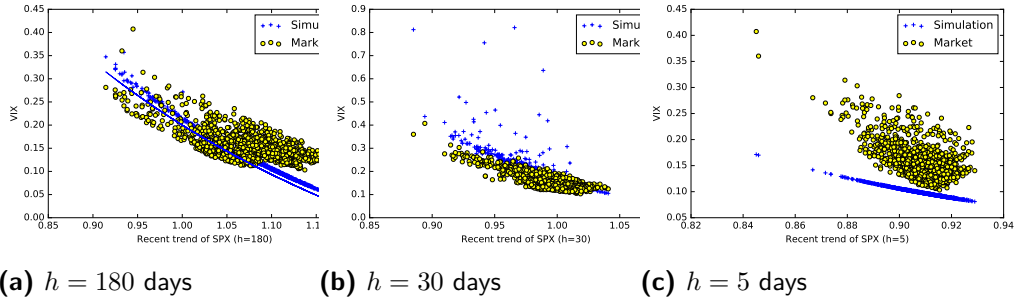


Figure 3.5: Simulated (blue crosses) and historical (yellow circles) VIX against the recent trend of S&P500, between 2/1/13 and 8/5/17. *Source: CBOE VIX Index.*

3.3.2 OPTIMISATION OF COEFFICIENTS

The calibration of the model is performed by minimising the least squared error (LSE)

$$\text{LSE}(\alpha, \beta, \gamma) := \sum_{t \in \mathcal{V}_T \cap [0, T]} \left| \text{VIX}_t - \widehat{\text{VIX}}_t(\alpha, \beta, \gamma) \right|^2$$

over the coefficients (α, β, γ) . Table 3.2 summarises the results of the optimisation, indicating the optimised parameters as well as the initial guesses and the number of runs. The optimal coefficients are close to the initial parameters $(\alpha_0, \beta_0, \gamma_0)$. This is due to starting with good guesses, in order to reduce the actual computation time, quite involved in the numerics (which can be

seen from the small number of simulations $(N, M) = (100, 200)$ considered). We leave for future work an optimisation of the algorithm in order to reduce the computational time to obtain robust optimal parameters $(\alpha^*, \beta^*, \gamma^*)$.

	$h = 180$	$h = 30$	$h = 5$
α_0	0.9	1.9	3.1
β_0	1.5	1.2	8.1
γ_0	0.8	1.8	0.5
α^*	0.88	1.96	3.15
β^*	1.52	1.21	8.01
γ^*	0.83	1.76	0.48
T	500	200	100
N	100	200	100
M	200	100	200
D	0	0.01	0

Table 3.2: Initial and calibrated parameters

3.4 PROOFS OF THE MAIN RESULTS

3.4.1 PROOF OF PROPOSITION 79

We start with the behaviour at the origin, and use the definition of a singular point to prove the statement. Regarding the first statement, for $\varepsilon > 0$,

$$\begin{aligned} \int_{-\varepsilon}^{\varepsilon} \frac{1 + |b(x)|}{\tilde{\sigma}^2(x)} dx &\leq 2 \left[1 + \max_{x \in [-\varepsilon, \varepsilon]} |b(x)| \right] \int_0^{\varepsilon} \frac{dx}{\tilde{\sigma}^2(x)}, \\ &\leq 2 \frac{[1 + \max_{x \in [-\varepsilon, \varepsilon]} |b(x)|]}{\gamma^2 + \delta} \int_0^{\varepsilon} \frac{dx}{x^{2(1-\beta)}}, \end{aligned}$$

as there exists $\delta \in (-\gamma^2, \infty)$ and $\bar{x} > 0$ such that for $x \leq \bar{x}$, $\gamma^2 - 2\frac{\alpha\gamma}{\beta}x^\beta + \frac{\alpha^2}{\beta^2}x^{2\beta} \geq \gamma^2 + \delta > 0$. This integral is finite if and only if $\beta > \frac{1}{2}$. Now, for $x > 0$, we have

$$\lim_{x \downarrow 0^+} \frac{b(x)}{\frac{1}{h}x} = \lim_{x \downarrow 0^+} (1 - x) = 1 \quad \text{and} \quad \lim_{x \downarrow 0^+} \frac{\tilde{\sigma}(x)}{\gamma x^{1-\beta}} = \lim_{x \downarrow 0^+} \left(1 - \frac{\alpha}{\beta\gamma} x^\beta \right) = 1.$$

We can then apply [33, Theorem 5.3] for $\bar{\gamma} := 2\beta - 1 \neq -1$ as $\beta > 0$, and conclude.

Regarding the behaviour at infinity, for $x > 0$, we have, as $\beta > 0$

$$\lim_{x \uparrow \infty} \frac{b(x)}{-\frac{1}{h}x^2} = 1 \quad \text{and} \quad \lim_{x \uparrow \infty} \frac{\tilde{\sigma}(x)}{-\frac{\alpha}{\beta}x} = \lim_{x \uparrow \infty} \left\{ 1 - \frac{\beta\gamma}{\alpha}x^{-\beta} \right\} = 1.$$

We can then apply [33, Theorem 5.7] for $\bar{\gamma} = 0 > -1$ and $\bar{\lambda} := -\frac{\beta^2}{h\alpha^2} < 0$ and ∞ has type A.

3.4.2 PROOF OF THEOREM 80(I)

We first show that y_σ is an isolated point. For $\varepsilon > 0$,

$$\begin{aligned} \int_{y_\sigma}^{y_\sigma + \varepsilon} \frac{1 + |b(x)|}{\tilde{\sigma}^2(x)} dx &\geq \int_{y_\sigma}^{y_\sigma + \varepsilon} \frac{dx}{\tilde{\sigma}^2(x)} = \int_{y_\sigma}^{y_\sigma + \varepsilon} \frac{dx}{x^2 \left(-\frac{\alpha}{\beta} + \gamma x^{-\beta} \right)^2} \\ &\geq K_\varepsilon \int_{y_\sigma}^{y_\sigma + \varepsilon} \frac{dx}{\left(-\frac{\alpha}{\beta} x^\beta + \gamma \right)^2} = \frac{K_\varepsilon}{\gamma^2} \int_0^\varepsilon \frac{dx}{\left(1 - \frac{\alpha}{\beta\gamma} x^\beta \left(1 + \frac{x}{y_\sigma} \right)^\beta \right)^2} \end{aligned} \quad (3.12)$$

with $K_\varepsilon := (y_\sigma + \varepsilon)^{-2(1-\beta)}$ for $\beta \in (0, 1)$, $K_\varepsilon := y_\sigma^{2(\beta-1)}$ for $\beta > 1$ and $K_\varepsilon := 1$ for $\beta = 1$.

Introduce the following quantities:

$$\begin{cases} A_\beta &:= \frac{(\beta-1)(\beta-2)}{2}, \\ B_\beta &:= A_\beta \left[A_\beta - \frac{\beta-3}{3} \right], \\ C_\beta &:= A_\beta \left[-\frac{(\beta-3)(\beta-4)}{12} + \frac{(\beta-1)(\beta-2)(\beta-3)}{3} - A_\beta^2 \right]. \end{cases} \quad (3.13)$$

A straightforward Taylor expansion around the origin then yields

$$\begin{aligned} \left[1 - \frac{\alpha}{\beta\gamma} x_\sigma^\beta \left(1 + \frac{x}{y_\sigma}\right)^\beta\right]^{-2} &= \frac{y_\sigma^2}{\beta^2 x^2} \left[1 - \frac{2A_\beta x}{y_\sigma} + (2B_\beta + A_\beta^2) \frac{x^2}{y_\sigma^2}\right] \\ &\quad + \frac{y_\sigma^2}{\beta^2 x^2} \left[2(C_\beta - A_\beta B_\beta) \frac{x^3}{y_\sigma^3} + o(x^3)\right], \end{aligned}$$

which is not integrable around zero. We thus conclude about the right behaviour of y_σ by noting that the last term in (3.12) diverges. Using similar arguments,

$$\begin{aligned} \int_{y_\sigma - \varepsilon}^{y_\sigma} \frac{1 + |b(x)|}{\tilde{\sigma}^2(x)} dx &\geq \int_{y_\sigma - \varepsilon}^{y_\sigma} \frac{dx}{\tilde{\sigma}^2(x)} = \int_{y_\sigma - \varepsilon}^{y_\sigma} \frac{dx}{x^2 \left(-\frac{\alpha}{\beta} + \gamma x^{-\beta}\right)^2}, \\ &\geq \bar{K}_\varepsilon \int_{y_\sigma - \varepsilon}^{y_\sigma} \frac{dx}{\left(-\frac{\alpha}{\beta} x^\beta + \gamma\right)^2} = \infty, \end{aligned}$$

with $\bar{K}_\varepsilon := y_\sigma^{-2(1-\beta)}$ for $\beta \in (0, 1)$, $\bar{K}_\varepsilon := (y_\sigma - \varepsilon)^{2(\beta-1)}$ for $\beta > 1$ and $\bar{K}_\varepsilon := 1$ for $\beta = 1$.

Introduce, for $a > 0$ and $x \in (0, a]$, the following functions:

$$\rho(x) := \exp\left(\int_x^a \frac{2\bar{b}(y)}{\bar{\sigma}^2(y)} dy\right), \quad \text{and} \quad \varphi(x) := \frac{1 + |\bar{b}(x)|}{\bar{\sigma}^2(x)}. \quad (3.14)$$

We then define the scale function s as:

$$s(x) := \begin{cases} \int_0^x \rho(y) dy, & \text{if } y_\sigma \geq 1, \\ -\int_x^a \rho(y) dy, & \text{if } y_\sigma < 1, \end{cases} \quad (3.15)$$

We rely here on the analysis in [33, Section 2.3]. The key ingredient is the following lemma:

Lemma 91. *The following hold:*

$$\begin{aligned} \lim_{x \downarrow 0} \rho(x) = 0, \quad \int_0^a \rho(y) dy < \infty, \quad \int_0^a \frac{\varphi(x)}{\rho(x)} dx = \infty, \quad \int_0^a \varphi(x) s(x) dx < \infty, \quad \text{if } y_\sigma \geq 1, \\ \lim_{x \downarrow 0} \rho(x) = \infty, \quad \int_0^a \rho(y) dy = \infty, \quad \int_0^a \frac{\varphi(x)}{\rho(x)} |s(x)| dx < \infty, \quad \text{if } y_\sigma < 1. \end{aligned}$$

Remark 92. *Note that the integral $\int_0^a \frac{\varphi(x)}{\rho(x)} dx$ is often referred to as the speed measure of the process Y .*

Introduce now the process $Z := (Y - y_\sigma)$, satisfying the SDE

$$dZ_t = \bar{b}(Z_t)dt + \bar{\sigma}(Z_t)dW_t, \quad Z_0 := y_0 - y_\sigma > 0.$$

with $\bar{b}(x) := b(x + y_\sigma)$ and $\bar{\sigma}(x) := \tilde{\sigma}(x + y_\sigma)$ for $x > 0$. Armed with Lemma 91, the right behaviour of Z at the origin, which corresponds to the right behaviour for the original process Y at y_σ . The case $y_\sigma \geq 1$ follows from [33, Theorem 2.13], whereas the case $y_\sigma < 1$ follows from [33, Theorem 2.16].

Proof of Lemma 91. We start with the limiting behaviour of the function ρ and its integral. A straightforward Taylor expansion around the origin yields

$$\begin{aligned} \left[1 - \frac{\alpha}{\beta\gamma} x_\sigma^\beta \left[1 + \frac{y}{y_\sigma} \right]^\beta \right]^{-2} &= \frac{y_\sigma^2}{\beta^2 y^2} \left[1 - 2A_\beta \frac{y}{y_\sigma} + (2B_\beta + A_\beta^2) \frac{y^2}{y_\sigma^2} \right] \\ &+ \frac{y_\sigma^2}{\beta^2 y^2} \left[2(C_\beta - A_\beta B_\beta) \frac{y^3}{y_\sigma^3} + o(x^3) \right], \end{aligned} \quad (3.16)$$

with $A_\beta, B_\beta, C_\beta$ introduced in (3.13). Introduce $K := \frac{2x_\sigma^{2\beta+1}(1-y_\sigma)}{h\beta^2\gamma^2} \leq 0$ as $y_\sigma \geq 1$ and $K_\beta^a := -\frac{1}{a} - \frac{2A_\beta}{y_\sigma} \log(a) + \frac{2B_\beta + A_\beta^2}{y_\sigma^2} a$.

When $y_\sigma \geq 1$, using (3.16), we obtain the asymptotic behaviour, as x approaches zero,

$$\frac{2\bar{b}(x)}{\bar{\sigma}^2(x)} = \frac{K}{x^2} \left(1 - 2A_\beta \frac{x}{y_\sigma} + (2B_\beta + A_\beta^2) \left(\frac{x}{y_\sigma} \right)^2 + 2(C_\beta - A_\beta B_\beta) \left(\frac{x}{y_\sigma} \right)^3 + o(x^3) \right). \quad (3.17)$$

As the expansion is uniform on $[x, a]$, one obtains

$$\begin{aligned}
\rho(x) &= \exp \left(2 \int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy \right), \\
&= \exp \left(\frac{K}{x} \left(1 + \frac{2A_\beta}{y_\sigma} x \log x \right) \right) \exp \left(K K_\beta^a - \frac{2B_\beta + A_\beta^2}{y_\sigma^2} Kx + o(x) \right), \\
&= \exp \left(\frac{K}{x} + K K_\beta^a \right) x^{\frac{2A_\beta K}{y_\sigma}} (1 + o(x)).
\end{aligned} \tag{3.18}$$

Since $K \leq 0$, $\rho(x)$ tends to zero as x tends to zero from above, and $\int_0^a \rho(x) dx$ is finite.

In the case $y_\sigma < 1$, the expansion (3.18) is still valid, albeit with $K \geq 0$. Therefore ρ explodes at the origin and $\int_0^a \rho(y) dy$ is infinite.

Now, it is straightforward to see that

$$\int_0^a \frac{1 + |\bar{b}(x)|}{\rho(x) \bar{\sigma}^2(x)} dx \geq \int_0^a \frac{dx}{\rho(x) \bar{\sigma}^2(x + y_\sigma)} \geq \int_0^a \frac{dx}{\bar{\sigma}^2(x + y_\sigma)},$$

which is clearly infinite., because ρ is bounded above by 1 on $(0, a]$ and from the asymptotic behaviour of the integrand around the origin, analysed in (3.12).

We now prove the last statement of the lemma, and start with the case $y_\sigma \geq 1$. With this, $\int_0^a \rho(y) dy$ is finite and $s(x) = \int_0^x \rho(y) dy$ for $x \in (0, a]$. Using (3.18), we can write the asymptotic behaviour of $s(\cdot)$ around the origin by integrating the asymptotic behaviour of $\rho(\cdot)$ around zero. Classical asymptotic expansions for integrals [98, Chapter 3.3] (note that the leading contribution arises at the right boundary of the integration domain) yields, after the change of variable $y \mapsto zx$,

$$\begin{aligned}
s(x) &= \int_0^x \rho(y) dy, \\
&= e^{K K_\beta^a x^{\frac{2A_\beta K}{y_\sigma}} + 1} \int_0^1 \exp \left(\frac{K}{zx} \right) z^{\frac{2A_\beta K}{y_\sigma}} dz, \\
&= e^{K K_\beta^a x^{\frac{2A_\beta K}{y_\sigma}} + 1} \exp \left\{ \frac{K}{x} \right\} \left(-\frac{1}{K} x + o(x) \right),
\end{aligned}$$

as x tends to zero. Combining this with (3.16) and (3.18), we obtain

$$\frac{s(x)}{\rho(x)\bar{\sigma}^2(x)} = \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2}x \left(-\frac{1}{K}x + o(x) \right) \frac{1}{x^2} (1 + o(1)) = -\frac{x_\sigma^{2\beta}}{K\beta^2\gamma^2} (1 + o(1)),$$

which is integrable on $(0, a]$ and concludes the proof, using the fact that

$$\int_0^a \frac{1 + |\bar{b}(x)|}{\rho(x)\bar{\sigma}^2(x)} s(x) dx \leq \left(1 + \max_{x \in (0, a]} |\bar{b}(x)| \right) \int_0^a \frac{s(x)}{\rho(x)\bar{\sigma}^2(x)} dx < \infty.$$

Consider now the case where $y_\sigma < 1$. $\int_0^a \frac{1 + |\bar{b}(x)|}{\rho(x)\bar{\sigma}^2(x)} |s(x)| dx \leq [1 + \max_{(0, a]} |\bar{b}|] \int_0^a \frac{|s(x)|}{\rho(x)\bar{\sigma}^2(x)} dx$. Since both $|s(x)|$ and $\rho(x)\bar{\sigma}^2(x)$ diverge to infinity at the origin, we need to study the behaviour of the integrand around zero to be able to conclude. We already proved that $\int_0^a \rho(x) dx$ is infinite for $y_\sigma < 1$. Hence, for $x \in (0, a]$, we obtain $s(x) := -\int_x^a \rho(y) dy$. For $\delta > 0$ such that $x < \delta < a$ and $x > 0$, $\int_x^a \rho(y) dy = \int_x^\delta \rho(y) dy + \int_\delta^a \rho(y) dy$. The second integral exists as the integral of a continuous function over a closed interval in $\mathbb{R}_+$. Regarding the first one, classical asymptotic expansions for integrals and (3.18), yield, after the change of variable $y \mapsto zx$,

$$\begin{aligned} \int_x^\delta \rho(y) dy &= e^{KK_\beta^a x^{1 + \frac{2A_\beta K}{y_\sigma}}} \int_1^{\delta/x} \exp\left(\frac{K}{xz}\right) z^{\frac{2A_\beta K}{y_\sigma}} dz, \\ &= e^{KK_\beta^a x^{1 + \frac{2A_\beta K}{y_\sigma}}} e^{\frac{K}{x}} \left(\frac{x}{K} + o(x) \right), \end{aligned}$$

and the asymptotic behaviour of the integrand around the origin becomes

$$\frac{|s(x)|}{\rho(x)\bar{\sigma}^2(x)} = \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2} \left(\frac{x^2}{K} + o(x^2) \right) \frac{1}{x^2} (1 + o(1)) = \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2 K} (1 + o(1)),$$

which is integrable at the origin, and the claim is proved. $\square$

3.4.3 PROOF OF THEOREM 80(II)

The strategy to prove the proposition is similar, albeit with different computations, to the previous case. Introduce, for $a > 0$ and $x \in (0, a]$,

$$\rho(x) := \exp \left(\int_x^a \frac{2\bar{b}(y)}{\bar{\sigma}^2(y)} dy \right) \quad \text{and} \quad \varphi(x) := \frac{1 + |\bar{b}(x)|}{\bar{\sigma}^2(x)}.$$

We then define the scale function s as:

$$s(x) := \begin{cases} \int_0^x \rho(y) dy, & \text{if } y_\sigma < 1, \\ -\int_x^a \rho(y) dy, & \text{if } y_\sigma \geq 1. \end{cases}$$

In order to determine the left behaviour of y_σ , we first state the following lemmas. They will enable us to decide which theorem to use from [33, Section 2.3].

Lemma 93.

$$\begin{aligned} \lim_{x \downarrow 0^+} \rho(x) = \infty, \quad \int_0^a \rho(y) dy = \infty, \quad \int_0^a \frac{\varphi(x)}{\rho(x)} |s(x)| dx < \infty, & \text{ if } y_\sigma \geq 1, \\ \lim_{x \downarrow 0^+} \rho(x) = 0, \quad \int_0^a \rho(y) dy < \infty, \quad \int_0^a \frac{\varphi(x)}{\rho(x)} dx = \infty, \quad \int_0^a \varphi(x) s(x) dx < \infty, & \text{ if } y_\sigma < 1. \end{aligned}$$

Similarly to before, introduce the process $Z := y_\sigma - Y$ satisfying the SDE

$$dZ_t = \bar{b}(Z_t) dt + \bar{\sigma}(Z_t) dW_t, \quad Z_0 := y_\sigma - y_0 > 0,$$

as well as the maps $\bar{b}(x) := -b(y_\sigma - x)$ and $\bar{\sigma}(x) := -\tilde{\sigma}(y_\sigma - x)$ for $x > 0$. With Lemma 93, we obtain the left behaviour of Z at the origin, corresponding to the left behaviour of y_σ for Y . The case $y_\sigma < 1$ follows from [33, Theorem 2.13], and the case $y_\sigma \geq 1$ from [33, Theorem 2.16], and Theorem 80(ii) follows.

Proof of Lemma 93. A straightforward Taylor expansion around the origin

yields

$$\begin{aligned} \left[1 - \frac{\alpha x_\sigma^\beta}{\beta \gamma} \left[1 - \frac{y}{y_\sigma}\right]^\beta\right]^{-2} &= \frac{y_\sigma^2}{\beta^2 y^2} \left[1 + \frac{2A_\beta y}{y_\sigma} + (2B_\beta + A_\beta^2) \frac{y^2}{y_\sigma^2}\right] \\ &+ \frac{y_\sigma^2}{\beta^2 y^2} \left[2(A_\beta B_\beta - C_\beta) \frac{y^3}{y_\sigma^3} + o(x^3)\right], \end{aligned} \quad (3.19)$$

with $A_\beta, B_\beta, C_\beta$ defined in (3.13). We start with the behaviour of the function ρ and its integrated version.

Consider first the case $y_\sigma \geq 1$. We split the range of possibilities into three possible intervals for a :

- (i) If $a < y_\sigma - 1$, then $y_\sigma - x \geq y_\sigma - a > 1$ and b is negative on $[y_\sigma - a, y_\sigma - x]$. Then, for $x \in (0, a]$,

$$\begin{aligned} \int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy &= \int_x^a \frac{-b(y_\sigma - y)}{\tilde{\sigma}^2(y_\sigma - y)} dy = - \int_x^a \frac{b(y_\sigma - y) dy}{(y_\sigma - y)^{2(1-\beta)} \gamma^2 \left(1 - \frac{\alpha}{\beta \gamma} x_\sigma^\beta \left(1 - \frac{y}{y_\sigma}\right)^\beta\right)^2}, \\ &= \frac{1}{h} \int_x^a \frac{(y_\sigma - y)^{2\beta-1} (y_\sigma - y - 1)}{\gamma^2 \left(1 - \frac{\alpha}{\beta \gamma} x_\sigma^\beta \left(1 - \frac{y}{y_\sigma}\right)^\beta\right)^2} dy, \\ &\geq \frac{(y_\sigma - a)^{2\beta-1} (y_\sigma - a - 1)}{\gamma^2 h} \int_x^a \frac{dy}{\left(1 - \frac{\alpha}{\beta \gamma} x_\sigma^\beta \left(1 - \frac{y}{y_\sigma}\right)^\beta\right)^2}, \end{aligned}$$

as $\min_{y \in [x, a]} [(y_\sigma - y)^{2\beta-1} (y_\sigma - y - 1)] = (y_\sigma - a)^{2\beta-1} (y_\sigma - a - 1) > 0$. Indeed the map $y \mapsto y^{2\beta-1} (y - 1)$ is increasing on $[y_\sigma - a, y_\sigma - x]$ because $y_\sigma - a > 1$. Noting that (3.19) is uniform on $[x, a]$, we obtain, as x approaches zero

$$\begin{aligned} &\exp \left\{ \frac{2 (y_\sigma - a)^{2\beta-1} (y_\sigma - a - 1)}{\gamma^2 h} \int_x^a \frac{dy}{\left(1 - \frac{\alpha}{\beta \gamma} x_\sigma^\beta \left(1 - \frac{y}{y_\sigma}\right)^\beta\right)^2} \right\} \\ &= \exp \left\{ \frac{\bar{K}}{x} + \bar{K} \bar{K}_\beta^a \right\} x^{-\frac{2A_\beta \bar{K}}{y_\sigma}} (1 + o(x)), \end{aligned}$$

with $\bar{K} := \frac{(y_\sigma - a)^{2\beta-1} (y_\sigma - a - 1) y_\sigma^2}{h \beta^2 \gamma^2} > 0$ and $\bar{K}_\beta^a := -\frac{1}{a} + \frac{2A_\beta}{y_\sigma} \log(a) +$

$\frac{2B_\beta + A_\beta^2}{y_\sigma^2}a$, and therefore $\lim_{x \downarrow 0} \rho(x) = \infty$ and $\int_0^a \rho(x) dx = \infty$.

(ii) If $y_\sigma - 1 \leq a < y_\sigma$, then for $x \in (0, a]$,

$$\int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy = \int_x^{y_\sigma-1} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy + \int_{y_\sigma-1}^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy.$$

Similarly to (i), one can prove that $\lim_{x \downarrow 0} \int_x^{y_\sigma-1} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy = \infty$. Then, on $(y_\sigma - 1, a]$, $\bar{\sigma}$ does not go to zero and is continuous, thus bounded; similarly, $\bar{b}$ is negative and continuous, hence bounded on $(y_\sigma - 1, a]$. Therefore $\int_{y_\sigma-1}^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy < \infty$, for $x \in (0, a]$, and

$$\lim_{x \downarrow 0} \rho(x) = \exp \left(\int_{y_\sigma-1}^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy \right) \exp \left(\int_x^{y_\sigma-1} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy \right) = \infty.$$

(iii) If $y_\sigma \leq a$, then for $x \in (0, a]$,

$$\int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy = \int_x^{y_\sigma-1} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy + \int_{y_\sigma-1}^{y_\sigma} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy + \int_{y_\sigma}^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy.$$

Similarly to (ii), it is easy to show that $\int_{y_\sigma-1}^{y_\sigma} \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy = - \int_0^1 \frac{b(y)}{\bar{\sigma}^2(y)} dy \leq 0$. Then, as $\bar{b}$ is positive on $(y_\sigma, a]$, $\int_{y_\sigma}^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy \geq 0$, and $\lim_{x \downarrow 0} \rho(x) = \infty$.

Consider now the case $y_\sigma < 1$. Using (3.19), we have the asymptotic behaviour, as x approaches zero

$$\frac{2\bar{b}(x)}{\bar{\sigma}^2(x)} = \frac{\tilde{K}}{x^2} \left(1 + 2A_\beta \frac{x}{y_\sigma} + (2B_\beta + A_\beta) \left(\frac{x}{y_\sigma} \right)^2 + 2(a_\beta B_\beta - C_\beta) \left(\frac{x}{y_\sigma} \right)^3 + o(x^3) \right), \quad (3.20)$$

with $\tilde{K} := \frac{2x_\sigma^{2\beta+1}(y_\sigma-1)}{h\beta^2\gamma^2} < 0$ as $y_\sigma < 1$. Since the expansion is again uniform

on $[x, a]$, we obtain

$$\begin{aligned}
\rho(x) &= \exp \left\{ 2 \int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy \right\}, \\
&= \exp \left\{ \frac{\tilde{K}}{x} \left(1 - \frac{2A_\beta}{y_\sigma} x \log x \right) \right\} \exp \left\{ \tilde{K} \bar{K}_\beta^a - \frac{2B_\beta + A_\beta^2}{y_\sigma^2} \tilde{K} x + o(x) \right\}, \\
&= \exp \left\{ \frac{\tilde{K}}{x} + \tilde{K} \bar{K}_\beta^a \right\} x^{-\frac{2A_\beta \tilde{K}}{y_\sigma}} (1 + o(x)). \tag{3.21}
\end{aligned}$$

Since $\tilde{K} < 0$, we easily deduce that $\lim_{x \downarrow 0} \rho(x) = 0$, and $\int_0^a \rho(x) dx$ is finite.

The middle statement in the lemma is straightforward. When $x \in (0, a]$, $\int_x^a \frac{\bar{b}(y)}{\bar{\sigma}^2(y)} dy = - \int_{y_\sigma - a}^{y_\sigma - x} \frac{b(y)}{\bar{\sigma}^2(y)} dy$. Since $0 < y_\sigma - a < y_\sigma - x < 1$, b is positive on $[y_\sigma - a, y_\sigma - x]$ and the above integral is therefore negative. Hence, ρ is bounded by 1 on $(0, a]$, and

$$\int_0^a \frac{1 + |\bar{b}(x)|}{\bar{\sigma}^2(x)} dx \geq \int_0^a \frac{dx}{\bar{\sigma}^2(y_\sigma - x)} = \infty,$$

using (3.19), which concludes the proof.

The final integrals in the lemma are rather delicate to analyse. We start with the case $y_\sigma < 1$. As $\int_0^a \rho(x) dx < \infty$, $s(x) = \int_0^x \rho(y) dy$. Using (3.21), one can obtain the asymptotic behaviour of $s(\cdot)$ around zero by integrating the asymptotic behaviour of $\rho(\cdot)$ around 0. Classical asymptotic expansions for integrals (note that the leading contribution arises at the right boundary of the integration domain) yields, after the change of variable $y \mapsto xz$,

$$\begin{aligned}
s(x) &= \int_0^x \rho(y) dy = e^{\tilde{K} \bar{K}_\beta^a} x^{1 - \frac{2A_\beta \tilde{K}}{y_\sigma}} \int_0^1 \exp \left(\frac{\tilde{K}}{xz} \right) z^{\frac{-2A_\beta \tilde{K}}{y_\sigma}} dz, \\
&= e^{\tilde{K} \bar{K}_\beta^a} x^{1 - \frac{2A_\beta \tilde{K}}{y_\sigma}} \exp \left(\frac{\tilde{K}}{x} \right) \left(-\frac{x}{\tilde{K}} + o(x) \right), \quad \text{as } x \text{ tends to zero.} \tag{3.22}
\end{aligned}$$

Therefore, combining (3.19), (3.21) and (3.22), we obtain

$$\frac{s(x)}{\rho(x)\bar{\sigma}^2(x)} = x \left(-\frac{1}{\widetilde{K}}x + o(x) \right) \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2} \frac{1}{x^2} (1 + o(1)) = -\frac{x_\sigma^{2\beta}}{\beta^2\gamma^2\widetilde{K}} (1 + o(1)),$$

which is integrable on $(0, a]$ and concludes the proof.

We now move on to the case where $y_\sigma \geq 1$. In that case,

$$\int_0^a \frac{1 + |\bar{b}(x)|}{\rho(x)\bar{\sigma}^2(x)} |s(x)| dx \leq \left[1 + \max_{(0,a]} |\bar{b}| \right] \int_0^a \frac{|s(x)|}{\rho(x)\bar{\sigma}^2(x)} dx.$$

As $\lim_{x \downarrow 0} |s(x)| = \infty$ and $\lim_{x \downarrow 0} \rho(x)\bar{\sigma}^2(x) = \infty$, one needs to study the behaviour of the integrand around zero to conclude. As $\int_0^a \rho(x) dx = \infty$, for $x \in (0, a]$, $s(x) := -\int_x^a \rho(y) dy$. For $\delta > 0$ such that $x < \delta < a$ and $x > 0$, $\int_x^a \rho(y) dy = \int_x^\delta \rho(y) dy + \int_\delta^a \rho(y) dy$. Note that the second integral is convergent as the integral of a continuous function over a closed interval of $\mathbb{R}$.

Classical asymptotic expansions for integrals [98, Chapter 3.3] and (3.21), yield, after mapping $y \mapsto zx$,

$$\begin{aligned} \int_x^\delta \rho(y) dy &= e^{\widetilde{K}\widetilde{K}_\beta^a} x^{1 - \frac{2A_\beta\widetilde{K}}{y_\sigma}} \int_1^{\delta/x} \exp\left(\frac{\widetilde{K}}{xz}\right) z^{-\frac{2A_\beta\widetilde{K}}{y_\sigma}} dz, \\ &= e^{\widetilde{K}\widetilde{K}_\beta^a} x^{1 - \frac{2A_\beta\widetilde{K}}{y_\sigma}} e^{\frac{\widetilde{K}}{x}} \left(\frac{x}{\widetilde{K}} + o(x) \right), \end{aligned}$$

and the asymptotic behaviour of the integrand around the origin is given by

$$\frac{|s(x)|}{\rho(x)\bar{\sigma}^2(x)} = \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2} \left(\frac{x^2}{\widetilde{K}} + o(x^2) \right) \frac{1 + o(1)}{x^2} = \frac{x_\sigma^{2\beta}}{\beta^2\gamma^2\widetilde{K}} (1 + o(1)),$$

which is integrable at the origin, and concludes the proof. $\square$

3.4.4 PROOF OF PROPOSITION 87

As the process Y^ε is in $\mathbb{R}_+^*$ instead of $\mathbb{R}$, we adapt the proof of [105, Theorem 2.9] to show that it satisfies a large deviations principle with speed ε and

rate function I^Y . Note first that for $y_\sigma > 0$, and in both cases $y_0 \geq y_\sigma$ and $y_0 < y_\sigma$, the function $\tilde{\sigma}$ is locally Lipschitz continuous on $\mathbb{R}_+^*$. Furthermore, for $f \in \overline{\mathcal{H}}$, the Picard-Lindelöf Theorem implies that the controlled ODE $\dot{g}_t = \tilde{\sigma}(g_t)\dot{f}_t$, with $g_0 = x$ admits the solution

$$\mathcal{S}^x(f)(t) = \left(\frac{\beta\gamma}{\alpha}\right)^{\frac{1}{\beta}} \left[e^{-\alpha \int_0^t \dot{f}_u du} \left(x^\beta \frac{\alpha}{\beta\gamma} - 1 \right) + 1 \right]^{1/\beta}, \quad \text{for } t \in \mathcal{T}, x > 0.$$

This formulation requires the term $\left[e^{-\alpha \int_0^t \dot{f}_u du} \left(x^\beta \frac{\alpha}{\beta\gamma} - 1 \right) + 1 \right]$ to be positive for all $x > 0$:

- If $y_0 \geq y_\sigma$, then $y_0^\beta \frac{\alpha}{\beta\gamma} - 1 \geq 0$ and $\mathcal{S}^{y_0}(f)$ is positive on $\mathcal{T}$;
- If $y_0 < y_\sigma$, then $y_0^\beta \frac{\alpha}{\beta\gamma} - 1 < 0$ and $\mathcal{S}^{y_0}(f)$ is positive on $\mathcal{T}$ if and only if Condition (3.7) holds.

The crucial step in [105, Theorem 2.9] is [105, Theorem 2.7], which states that if $\sqrt{\varepsilon}W$ is close to $f \in \overline{\mathcal{H}}$, then Y^ε should be close to $\mathcal{S}^{y_0}(f)$, the solution of the controlled ODE. The case of bounded and globally Lipschitz coefficients follows directly from [105, Theorem 2.7]. In order to deal with locally Lipschitz coefficients here, we need to localise. For $0 < r_\sigma < r_b$, the functions

$$\bar{b}(x) := \begin{cases} b(x), & \|x\| \leq r_b, \\ b\left(\frac{rx}{\|x\|}\right), & \|x\| > r_b, \end{cases} \quad \text{and} \quad \bar{\sigma}(x) := \begin{cases} \tilde{\sigma}(x), & \|x\| > r_\sigma, \\ \tilde{\sigma}\left(\frac{rx}{\|x\|}\right), & \|x\| \leq r_\sigma. \end{cases}$$

are bounded and globally Lipschitz continuous on $\mathbb{R}_+^*$, and clearly $\varepsilon\bar{b}(\cdot)$ converges uniformly to zero on $\mathbb{R}_+^*$ as ε goes to zero. Moreover, given $\eta \in (r_b, r_\sigma)$, there exists $0 < r \leq \min\{\eta - r_\sigma, \eta - r_b\}$ such that the δ -tube around $\mathcal{S}^{y_0}(f)$ is contained in $B_r(\eta)$. In order for this radius r to exist, one simply needs to make sure that the solution $\mathcal{S}^{y_0}(f)$ of the controlled ODE never reaches zero (explosion is impossible as infinity is recurrent), which is obvious when $y_0 \geq y_\sigma$, and guaranteed by Condition (3.7) when $y_0 < y_\sigma$.

Denote $\bar{Y}^\varepsilon$ the solution to $d\bar{Y}_t^\varepsilon = \varepsilon\bar{b}(\bar{Y}_t^\varepsilon)dt + \sqrt{\varepsilon}\bar{\sigma}(\bar{Y}_t^\varepsilon)dW_t$ with $\bar{Y}_0^\varepsilon = y_0 > 0$. Then the two sequences $(\bar{Y}^\varepsilon)_{\varepsilon>0}$ and $(Y^\varepsilon)_{\varepsilon>0}$ are identical on $B_r(\eta)$. Thus,

for each $\delta, \lambda > 0$, there exist $\xi, \zeta > 0$ such that, for all $f \in \overline{\mathcal{H}}$ with $\Lambda(f) \leq \lambda$ and $x \in B_\xi(y_0)$,

$$\mathbb{P} [\|Y^\varepsilon - \mathcal{S}^{y_0}(f)\|_\infty > \delta, \|\sqrt{\varepsilon}W - f\|_\infty \leq \zeta] = \mathbb{P} [\|\overline{Y}^\varepsilon - \mathcal{S}^{y_0}(f)\|_\infty > \delta, \|\sqrt{\varepsilon}W - f\|_\infty \leq \zeta].$$

The constants $\xi, \zeta > 0$ are such that [105, Theorem 2.7] is satisfied. Hence, for each $R, \delta, \lambda > 0$, there exist $\zeta, \xi, \varepsilon_0 > 0$ such that, for all $f \in \overline{\mathcal{H}}$ with $\Lambda(f) \leq \lambda$, $x \in B_\xi(y_0)$, $\varepsilon \leq \varepsilon_0$,

$$\mathbb{P} [\|Y^\varepsilon - \mathcal{S}^{y_0}(f)\|_\infty > \delta, \|\sqrt{\varepsilon}W - f\|_\infty \leq \zeta] \leq \exp \left\{ -\frac{R}{\varepsilon} \right\}$$

holds, and the proof follows from [105, Theorem 2.9].

3.4.5 PROOF OF THEOREM 86

To obtain a large deviations principle for X^ε , a large deviations principle for the rescaled process $X^\varepsilon := (X^\varepsilon, Y^\varepsilon)$ needs to be proved.

$$dX_t^\varepsilon = \varepsilon b(X_t^\varepsilon)dt + \sqrt{\varepsilon}a(X_t^\varepsilon)dW_t,$$

with initial condition $X_0^\varepsilon := x_0 = \begin{pmatrix} \log s_0 \\ y_0 \end{pmatrix}$ and the maps $b, a : \mathbb{R}_+^* \rightarrow \mathbb{R}^2$ defined as

$$b(X_t^\varepsilon) = \begin{pmatrix} -\frac{1}{2}\sigma^2(Y_t^\varepsilon) \\ b(Y_t^\varepsilon) \end{pmatrix} \quad \text{and} \quad a(X_t^\varepsilon) = \begin{pmatrix} \sigma(Y_t^\varepsilon) \\ \tilde{\sigma}(Y_t^\varepsilon) \end{pmatrix}.$$

These two maps are both locally Lipschitz continuous on $\mathbb{R} \times \mathbb{R}_+^*$. Solving the controlled ODE for Y^ε is sufficient to solve the controlled ODE for the process X^ε . Using the proof of Proposition 87, for $f := (f, f)$ with $f \in \overline{\mathcal{H}}$, the controlled ODE $\dot{g}_t = a(g_t)\dot{f}_t$, with $g_0 = x_0$ has a solution $g = \mathcal{S}^{x_0}(f)$ on $\mathcal{T}$. For $y_0 > y_\sigma$, the solution $\mathcal{S}^{y_0}$ is strictly positive and $\mathcal{S}^{x_0}(f)$ exists on $\mathcal{T}$ for all $f \in \overline{\mathcal{H}}$ and $x_0 \in \mathbb{R} \times \mathbb{R}_+^*$. In this case, $\overline{\mathcal{H}}$ boils down to the Cameron-Martin space. For $y_0 < y_\sigma$, Condition (3.7) ensures that $\mathcal{S}^{y_0}$ is positive. Applying [105, Theorem 2.9], the sequence X^ε then satisfies a large

deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R} \times \mathbb{R}_+^*)$ as ε tends to zero, with speed ε and rate function

$$I^{Y,X}(g) := \inf \{ \Lambda(f), f \in \overline{\mathcal{H}}, \mathcal{S}^{x_0}(f) = g \}.$$

To obtain a large deviations principle for the log-stock price X^ε , we apply the Contraction Principle [39, Theorem 4.2.1] since the projection on the first component is continuous, and the theorem follows.

4

CONCLUSION

Stochastic volatility models have always been of great interest for the mathematical finance community, to understand the dynamics and the behaviour of the implied volatility. Over the last few years, rough volatility models have been developed, as they provide important improvements to the more classical volatility models, such as the power-law decay of the volatility smiles close to maturity.

In this thesis, we studied various stochastic volatility models, with both rough and path-dependent features: in Chapter 1, a generalised version of the Stein-Stein model where the instantaneous volatility is driven by a fractional Brownian motion and starts from a random distribution; in Chapter 2, the realised variance process of the rough Bergomi model; and in Chapter 3, a complete model with stochastic volatility where the instantaneous volatility depends on an exponential weighted moving average of the stock price process. For each of these models, we used large deviations techniques to derive asymptotics for the rescaled volatility and stock price processes. We then provided asymptotics for the implied volatility, a crucial quantity for practitioners. In particular, in Chapter 1, we extended results on sample path large deviations and provided some possible techniques to overcome the

difficulties encountered when working with large deviations techniques. We also demonstrated how to modulate the effect of the Hurst parameter H as well as the randomisation of the initial volatility to reproduce the steepness of the implied volatility smile. More work needs to be carried out to develop a numerical scheme to approximate the rate function.

On the practical side, we then presented in Chapter 2 how rough volatility models, and the rough Bergomi model in particular, can be used to derive asymptotics for the Realised Variance options. We investigated the effect of correlation in the multi-factor rough Bergomi models and its influence on the smiles. As an example, we used an approximation of the Realised Variance density to price Volatility Swaps. Regarding the VIX options, we were only able to prove the small-noise, large strike asymptotics. Future research will be focused on getting the small-time behaviour of such options.

Finally in Chapter 3, we focused on a different complete model with stochastic volatility introduced by Guyon, and derived from Hobson and Rogers. After providing the conditions on the coefficients to obtain existence and uniqueness of a strong solution, we proved small-time asymptotics for the stochastic volatility model. Considering the computational time needed to optimise the coefficients in Chapter 3, the numerical scheme developed needs to be improved to obtain precise estimators and reduce the computational time.

REFERENCES

- [1] M. Abramowitz and I.A. Stegun. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical tables. Dover, New York, 1972.
- [2] R.J. Adler and J.E. Taylor. Random Fields and Geometry. Springer-Verlag, New York, 2007.
- [3] C.D. Aliprantis and K.C. Border. Infinite dimensional analysis, a hitchhiker's guide. Springer, Third Edition, 2005.
- [4] E. Alòs, D. García-Lorite and A. Muguruza. On smile properties of volatility derivatives and exotic products: understanding the VIX skew. [arXiv:1808.03610](https://arxiv.org/abs/1808.03610), 2018.
- [5] E. Alòs, J. León and J. Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4), 571-589, 2007.
- [6] M. Avellaneda and A. Papanicolaou. Statistics of VIX Futures and Their Applications to Trading Volatility Exchange-Traded Products. *The Journal of Investment Strategies*, Vol. 7, No. 2, 1-33, 2018.
- [7] R. Azencott. Grandes déviations et applications. *Lecture Notes in Mathematics*, 774:1-176, 1980.
- [8] R. Azencott. Petites perturbations aléatoires des systèmes dynamiques: développements asymptotiques. *Bulletin des sciences mathématiques*, 109: 253-308, 1985.
- [9] R. Azencott, B. Geiger and W. Ott. Large Deviations for Gaussian Diffusions with delay. *Journal of Statistical Physics*, 2017.

- [10] O. E. Barndorff-Nielsen and J. Schmiegel. Ambit processes: with applications to turbulence and tumour growth. In F. E. Benth, G. Di Nunno, T. Lindstrøm, B. Øksendal and T. Zhang (Eds.), 2007. *Stochastic analysis and applications, volume 2 of Abel Symp.*: 93-124, Springer, Berlin.
- [11] F. Baudoin and C. Ouyang. Small-time kernel expansion for solutions of stochastic differential equations driven by fractional Brownian motions. *Stochastic Processes and Applications*, 121: 759-792, 2011.
- [12] F. Baudoin and C. Ouyang. On small time asymptotics for rough differential equations driven by fractional Brownian motions. *Large Deviations and Asymptotic Methods in Finance*, Springer Proceedings in Mathematics & Statistics, 110, 2015.
- [13] C. Bayer, P. K. Friz and J. Gatheral. Pricing Under Rough Volatility. *Quantitative Finance*, 16(6): 887-904, 2016.
- [14] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, B. Stemper. Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, 19(5), 779-798, 2019.
- [15] G. Ben Arous. Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. *Stochastics*, 25: 125-153, 1988.
- [16] S. Benaïm and P. Friz. Regular Variation and Smile Asymptotics. *Mathematical Finance*, 19(1):1-12, 2009.
- [17] M. Bennedsen, A. Lunde and M.S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *Finance and Stochastics*, 19(1): 1-12, 2009.
- [18] L. Bergomi. Smile dynamics III. *Risk*, October: 90-96, 2008.
- [19] L. Bergomi. Stochastic Volatility Modeling. *Chapman & Hall/CRC financial mathematics series*, 2016.
- [20] C. Bezuidenhout. A large deviations principle for small perturbations of random evolution equations. *The Annals of Probability*, 15(2):646-658, 1987.

- [21] F. Biagini, Y. Hu, B. Øksendal and T. Zhang. Stochastic calculus for fractional Brownian motion and applications. Springer Verlag, 2008.
- [22] N. H. Bingham, C. M. Goldie and J. L. Teugels. Regular Variation, 2nd edition. Cambridge University Press, Cambridge, 1989.
- [23] J.M. Bismut. Large deviations and the Malliavin Calculus. Birkhauser, 1984.
- [24] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy* 81(3): 637-654, 1973.
- [25] T. Bollerslev. Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3): 302-327, 1986.
- [26] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in options prices. *J. Business*, 51: 621-651, 1978.
- [27] P. Carr, H. Geman, D. Madan, and M. Yor. Pricing options on realized variance. *Finance and Stochastics*, 9(4):453-475, 2005.
- [28] P. Carr and R. Lee. Volatility Derivatives. Annual Review of Financial Economics 1(1), 319-339, 2009.
- [29] F. Caravenna and J. Corbetta. The asymptotic smile of a multiscaling stochastic volatility model. *Stochastic Processes and their Applications*, 128(3): 1034-1071, 2018.
- [30] R. Carmona and M.R. Tehranchi. Interest rate models: an infinite dimensional stochastic analysis perspective. Springer, 2006.
- [31] P. Cheridito, H. Kawaguchi and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *EJP* 8(3): 1-14, 2003.
- [32] A. Cherny. Brownian moving averages have conditional full support. *Ann. App. Prob.*, 18(5): 1825-1850, 2008.
- [33] A.S. Cherny and H-J Engelbert. Singular stochastic differential equations. Springer, 2005.

- [34] F. Comte, L. Coutin and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2): 337-378, 2012
- [35] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Math. Finance*, 8(4): 291-323, 1998.
- [36] R. Cont and E. Voltchkova. A finite difference scheme for option pricing in jump-diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis* 43(4), 1596-1626, 2005.
- [37] L. Decreusefond and A. Üstünel. Stochastic analysis of the fractional Brownian motion. *Potential Analysis* 10: 177-214, 1996.
- [38] A. Dembo, J-D. Deuschel and D. Duffie. Large portfolio losses. *Finance and Stochastics*, 8: 3-16, 2004.
- [39] A. Dembo and O. Zeitouni. Large Deviations Techniques and Applications. II. Springer Verlag, 1998.
- [40] J.D. Deuschel, P.K. Friz, A. Jacquier and S. Violante. Marginal density expansions for diffusions and stochastic volatility, Part I: Theoretical foundations. *Communications on Pure and Applied Mathematics*, 67(1): 40-82, 2014.
- [41] J.D. Deuschel, P.K. Friz, A. Jacquier and S. Violante. Marginal density expansions for diffusions and stochastic volatility, Part II: Applications. *Communications on Pure and Applied Mathematics*, 67(2): 321-350, 2014.
- [42] J.D. Deuschel and D. Stroock. Large Deviations. American Mathematical Society, 342, 2001.
- [43] J.C. Duan. The GARCH Option Pricing Model. *Mathematical Finance*, 5(1):13-32, 1995.
- [44] B. Dupire. Pricing with a smile. *Risk*, 7:18-20, 1994.
- [45] R. Engle. Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom. *Econometrica*, 50: 987-1008, 1982.

- [46] O. El Euch, M. Fukasawa and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2), 241-280, 2018.
- [47] O. El Euch and M. Rosenbaum. Perfect hedging under rough Heston models. *The Annals of Applied Probability*, 28 (6): 3813-3856, 2018.
- [48] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29 (1): 3-38, 2019.
- [49] M. Forde and A. Jacquier. Small-time asymptotics for implied volatility under the Heston model. *IJTAF*, 12(6): 861-876, 2009.
- [50] M. Forde and A. Jacquier. Small-time asymptotics for an uncorrelated local-stochastic volatility model. *Applied Mathematical Finance*, 18 (6): 517-535, 2011.
- [51] M. Forde, A. Jacquier and R. Lee. The small-time smile and term structure of implied volatility under the Heston model. *SIAM Journal on Financial Mathematics*, 3(1): 690-708, 2012.
- [52] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM J. Finan. Math.*, 8(1): 114-145, 2017.
- [53] J-P. Fouque and B. Ren. Approximation for option prices under uncertain volatility. *SIAM J. Finan. Math.*, 5: 360-383, 2014.
- [54] J.P. Fouque, G. Papanicolaou, R. Sircar and K. Sølna. Multiscale stochastic volatility for equity, interest rate, and credit derivatives. *Cambridge University Press*, 2011.
- [55] A. Friedman. Stochastic Differential Equations and Applications. Volume 1. Academic Press, Inc., New York, 1975.
- [56] M. Freidlin and A. Wentzell. Random perturbations of dynamical systems. Springer Verlag, 260, 1984.
- [57] P. Friz, P. Gassiat and P. Pigato. Precise asymptotics: robust stochastic volatility models. Preprint available at [arXiv:1811.00267](https://arxiv.org/abs/1811.00267), 2018.

- [58] P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier and J. Teichmann. Large deviations and asymptotic methods in Finance. Springer Proceedings in Mathematics and Statistics, 110, 2015.
- [59] M. Fukasawa. The normalizing transformation of the implied volatility smile. *Mathematical Finance*, 22(4): 753-762, 2012.
- [60] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17: 2, 189-198, 2017.
- [61] K. Gao and R. Lee. Asymptotics of implied volatility to arbitrary order. *Finance and Stochastics* 18(2): 349-392, 2014.
- [62] J. Garcia. A large deviation principle for stochastic integrals. *Journal of Theoretical Probability*, 21(2): 476-501, 2008.
- [63] P. Gassiat. On the martingale property in the rough Bergomi model. To appear in *Electronic Communications in Probability*, 2019.
- [64] J. Gatheral. The volatility surface: A practitioner's guide. Wiley New York, 2006.
- [65] J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6): 933-949, 2018.
- [66] A. Genin and P. Tankov. Optimal importance sampling for Lévy processes. To appear in *Stochastic Processes and their Applications*, 2019.
- [67] S. Gerhold, C. Gerstenecker and A. Gulisashvili. Large deviations for fractional volatility models with non-Gaussian volatility driver. [arXiv:2003.12825](https://arxiv.org/abs/2003.12825), 2020.
- [68] P. Guasoni and S. Robertson. Optimal importance sampling with explicit formulas in continuous time. *Finance and Stochastics*, 12(1): 1-19, 2008.
- [69] A. Gulisashvili, F. Viens and X. Zhang. Small-time asymptotics for Gaussian self-similar stochastic volatility models. *Applied Mathematics & Optimization*, 2018.

- [70] J. Guyon. Path-dependent volatility. *Risk*, 2014.
- [71] J. Guyon. Path-dependent volatility: practical examples. *Global Derivatives Conference*, 2017.
- [72] J. Guyon and P. Henry-Labordère. Being particular about calibration. *Risk*, 2012.
- [73] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2): 327-343, 1993.
- [74] S.L. Heston and S. Nandi. A closed-form GARCH option valuation model. *The Review of Financial Studies*, 13(3): 585-625, 2000.
- [75] D.G. Hobson and L.C.G. Rogers. Complete models with stochastic volatility. *Mathematical Finance*, 8(1): 27-48, 1998.
- [76] B. Horvath, A. Jacquier and C. Lacombe. Asymptotic behaviour of randomised fractional volatility models. *Journal of Applied Probability*, 56(2), 496-523, 2019.
- [77] B. Horvath, A. Jacquier and A. Muguruza. Functional central limit theorems for rough volatility. [arXiv:1711.03078](https://arxiv.org/abs/1711.03078), 2017.
- [78] B. Horvath, A. Jacquier and P. Tankov. Volatility options in rough volatility models. [arXiv:1802.01641](https://arxiv.org/abs/1802.01641), 2018.
- [79] B. Horvath, A. Muguruza and M. Tomas. Deep Learning Volatility. [arXiv:1901.09647](https://arxiv.org/abs/1901.09647), 2019.
- [80] A. Jacquier, M. Keller-Ressel and A. Mijatović. Large deviations and stochastic volatility with jumps: asymptotic implied volatility for affine models. *Stochastics*, 85(2): 321-345, 2013.
- [81] A. Jacquier, C. Martini and A. Muguruza. On VIX Futures in the rough Bergomi model. *Quantitative Finance*, 18(1), 45-61, 2018.

- [82] A. Jacquier, M. Pakkanen and H. Stone. Pathwise large deviations for the Rough Bergomi model. *Journal of Applied Probability*, 55(4): 1078-1092, 2018.
- [83] A. Jacquier and P. Roome. The small-maturity Heston forward smile. *SIAM Journ. Fin. Math.*, 4(1): 831-856, 2013.
- [84] A. Jacquier and P. Roome. Asymptotics of forward implied volatility. *SIAM Journ. Fin. Math.*, 6(1): 307-351, 2013.
- [85] A. Jacquier and P. Roome. Black-Scholes in a CEV random environment: a new approach to smile modelling. *Mathematics and Financial Economics*, 12(3), 445-474, 2018.
- [86] A. Jacquier and F. Shi. The randomised Heston model. *SIAM Journal on Financial Mathematics*, 10(1), 89-129, 2019.
- [87] S. Karlin and H.M. Taylor. A second course in stochastic processes. Academic Press, 1981.
- [88] M. Keller-Ressel and J. Muhle-Karbe. Asymptotic and exact pricing of options on variance. *Finance and Stochastics*, 17(1): 107-133, 2013.
- [89] R. Koekkoek, R. Lord and D. Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2): 177-194, 2010.
- [90] C. Lacombe, A. Muguruza and H. Stone. Asymptotics for volatility derivatives in multi-factor rough volatility models [arXiv:1903.02833](https://arxiv.org/abs/1903.02833), 2019.
- [91] R. Lee. The Moment Formula for Implied Volatility at Extreme Strikes. *Mathematical Finance*, 14(3): 469-480, 2004.
- [92] J. Ma, Z. Ren, N. Touzi and J. Zhang. Large Deviations for non-Markovian diffusions and Path-dependent Eikonal equation. *Annales de l'IHP: Probabilités et Statistique*, 1196-1216, 2016.

- [93] E. Mayer-Wolf, D. Nualart and V. Pérez-Abreu. Large deviations for multiple Wiener-Itô integral processes. *Séminaire de Probabilités XXVI*, 26: 11-31, 1992.
- [94] R. McCrickerd and M. Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model *Quant. Finance*, 18(11): 1877-1886, 2018.
- [95] M. Mellouk. A large-deviation principle for random evolution equations. *Bernoulli*, 6(6): 977-999, 2000.
- [96] S. Mechkov. Hot-start initialization of the Heston model. *Risk*, November 2016.
- [97] A. Millet, D. Nualart and M. Sanz. Large Deviations for a class of anticipating stochastic differential equations. *The Annals of Probability*, 20(4): 1902-1931, 1992.
- [98] P.D. Miller. Applied asymptotic analysis. American Mathematical Society, 75, 2006.
- [99] Y.S. Mishura. Stochastic calculus for fractional Brownian motion and related processes. Springer-Verlag, 2008.
- [100] S. Mohammed and T. Zhang. Anticipating stochastic differential systems with memory. *Stochastic Processes and their Applications*, 119: 2773-2802, 2009.
- [101] A. Neuberger. The log contract. *Journal of Portfolio Management*, 20(2): 74-80, 1994.
- [102] I. Norros, E. Valkeika and J. Virtamo. An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli*, 5(4): 571-587, 1999.
- [103] E. Pardoux and A.Y. Veretennikov. On Poisson equation and diffusion approximation, II. *The Annals of Probability*, 31(3): 1166-1192, 2003.
- [104] H. Pham. Large Deviations in Finance. Third SMAI European Summer School in Financial Mathematics, 2010.

- [105] D. Peithmann. Large deviations and exit time asymptotics for diffusions and stochastic resonance. *PhD Thesis* Humboldt University, Berlin, 2007.
- [106] S. Robertson. Sample Path Large Deviations and Optimal Importance Sampling for Stochastic Volatility Models. *Stochastic Processes and Their Applications*, 120, 66-83, 2010.
- [107] M. Schilder. Asymptotic formulas for Wiener integrals. *Transactions of the AMS*, 125 : 63-85, 1966.
- [108] R. Schöbel and J. Zhu. Stochastic volatility with an Ornstein-Uhlenbeck process: an extension. *Review Fin.*, 3(1): 23-46, 1999.
- [109] E. Stein and J. Stein. Stock-price distributions with stochastic volatility - an analytic approach. *Review of Financial studies*, 4: 727-752, 1991.
- [110] H. Stone. Calibrating rough volatility models: a convolutional neural network approach. *To appear in Quantitative Finance*, available at: <https://doi.org/10.1080/14697688.2019.1654126>, 2019.
- [111] P. Tankov. Pricing and hedging in exponential Lévy models: Review of recent results. *Paris-Princeton Lectures on Mathematical Finance*, 319-359, 2010.
- [112] S.R.S. Varadhan. On the behaviour of the fundamental solution of the heat equation with variable coefficients. *Communications on Pure and Applied Mathematics*, 20: 431-455, 1967.
- [113] L. Yan, Y. Lu and Z. Xu. Some properties of the fractional Ornstein-Uhlenbeck process. *Journal of Physics A: Mathematical and Theoretical*, 41(14): 1-17, 2008.