ROUGH VOLATILITY MODELS: SMALL-TIME ASYMPTOTICS AND CALIBRATION

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

BY

HENRY STONE

DEPARTMENT OF MATHEMATICS,

IMPERIAL COLLEGE LONDON,

LONDON,

SW7 2AZ.

OCTOBER 2019

Declaration

I, Henry Stone, certify that this thesis, and the research to which it refers, are the products 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

"Errors are volitional and are the portals of discovery."

James Joyce, Ulysses, 1922.

Thesis Advisors: Dr Antoine Jacquier, Dr Mikko Pakkanen Henry Stone

Abstract

Inspired by the work of Alòs, León and Vives [ALV07] and Fukasawa [Fuk17], who showed that a volatility process driven by a fractional Brownian motion generates the power law at-the-money volatility skew observed in financial market data, Gatheral, Jaisson and Rosenbaum [GJR18a] spawned a class of models now known as rough volatility models. We study the asymptotic behaviour of such models, and investigate how convolutional neural networks can be used for their calibration.

Chapter 1 serves as an introduction. We begin with implied volatility, and then introduce a number of model classes, starting with local volatility models and ending with rough volatility models, and discuss their associated asymptotic behaviour. We also introduce the theoretical tools used to prove the main results.

In Chapter 2 we study the small-time behaviour of the rough Bergomi model, introduced by Bayer, Friz, and Gatheral [BFG16]. We prove a pathwise large deviations principle for a small-noise version of the model, and use this result to establish the small-time behaviour of the rescaled log stock price process. This, in turn, allows us to characterise the small-time implied volatility behaviour of the model. Using the same theoretical framework, we are also able to establish the small-time implied volatility behaviour of the lognormal fSABR model of Akahori, Song, and Wang [ASW17].

In Chapter 3 we present small-time implied volatility asymptotics for realised variance (RV) options for a number of (rough) stochastic volatility models via a large deviations principle. We interestingly discover that these (rough) volatility models, together with others proposed in the literature, generate linear smiles around the money. We provide numerical results along with efficient and robust numerical recipes to compute the rate function; the backbone of our theoretical framework. Based on our results, we develop an approximation scheme for the density of the realised variance, which in turn allows the volatility swap density to be expressed in closed form. Lastly, we investigate different constructions of multi-factor models and how their construction affects the convexity of

the implied volatility smile. Remarkably, we identify a class of models that can generate non-linear smiles around-the-money. Additionally, we establish small-noise asymptotic behaviour of a general class of VIX options in the large strike regime.

In Chapter 4, which is self-contained, we give an introduction to machine learning and neural networks. We investigate the use of convolutional neural networks to find the Hölder exponent of simulated sample paths of the rough Bergomi model, a method which performs extremely well and is found to be robust when applied to trajectories of a fractional Brownian motion and an Ornstein-Uhlenbeck process. We then propose a novel calibration scheme for the rough Bergomi model based on our results.

Acknowledgements

I would first like to thank my supervisors, Dr Antoine Jacquier and Dr Mikko Pakkanen, for all of their help and guidance throughout the course of my PhD. Their combined and extensive knowledge proved to be invaluable on countless occasions, and their permittance and encouragement for me to pursue my own independent project on convolutional neural networks was greatly appreciated. Many thanks also to Professor Christian Bayer, Professor Damiano Brigo, and Professor Johannes Muhle-Karbe for taking the time to read my thesis and agreeing to be the examiners for my viva.

I thank everyone in the mathematical finance section at Imperial College London for a constantly stimulating research environment; in particular I thank my co-authors Chloe Lacombe and Aitor Muguruza.

I thank the EPSRC CDT in Financial Computing and Analytics for funding my PhD and providing travel expenses to a number of conferences, where I was able to present my research.

Finally I thank my sisters for always inspiring me, my parents for all the opportunities they have provided me, and my friends for their belief in me.

Contents

List of Figures

List of Tables

Notations:

$\mathscr{B}(\mathscr{X})$ denotes the Borel σ-algebra of a topological space $\mathscr{X}$.

BV denotes the space of paths of finite variation on some index set $\mathcal{T}$.

$\mathcal{C}^d := \mathcal{C}(\mathcal{T}, \mathbb{R}^d)$ denotes the space of continuous mappings from some index set $\mathcal{T}$ to $\mathbb{R}^d$. We write $\mathcal{C}^1$ as $\mathcal{C}$ for ease of notation.

δ_x denotes the Dirac delta function with mass at x.

$\mathbb{E}[\cdot]$ denotes the expectation operator.

$\mathcal{E}$ denotes the Wick stochastic exponential.

Γ denotes the Gamma function.

${}_2F_1$ denotes the Gauss hypergeometric function.

$\dot{f}$ denotes the derivative of a function f.

$\mathbb{I}_A$ denotes the indicator function of a set A.

ι denotes the inclusion map.

$\mathrm{L}^2 := L^2(\mathcal{T}, \mathbb{R})$ denotes the space of square integrable functions from some index set $\mathcal{T}$ to $\mathbb{R}$.

Φ denotes the Gaussian cumulative distribution function.

n denotes the Standard Gaussian probability density function.

$\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2.

$\mathbb{R}_+ := [0, +\infty)$ and $\mathbb{R}_+^* := (0, +\infty)$.

$x \wedge y$ denotes $\min\{x, y\}$.

$x \vee y$ denotes $\max\{x, y\}$.

$x \cdot y$ denotes the integral $\int_0^1 x(s)\mathrm{d}y(s)$.

$\langle X., Y.\rangle_t$ denotes the quadratic co-variation of two processes X and Y at time t.

For two paths x and y belonging to $\mathcal{C}$, we shall denote by z_y^x the two-dimensional path $(x, y)^\top$.

Conventions:

Any stochastic process X given throughout the paper is shorthand for $(X_t)_{t\in\mathcal{T}}$.

Unless stated, the risk-free rate of return r is assumed to be 0 throughout.

Let $(\Omega, \mathscr{A}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a given filtered probability space, where the filtration is assumed to satisfy the usual conditions. All stochastic processes given throughout the paper are assumed to live on this probability space, and further are assumed to be independent from one another unless stated otherwise.

Chapter 1

Introduction

It is widely accepted that the Black-Scholes options pricing framework, whilst offering a fast and convenient means of pricing options, suffers from a number of drawbacks that make it inadmissible for practical use. Perhaps the most significant shortcoming is the fact that the volatility is assumed to be constant, the converse of which is well-documented.

Given the parameters (S_0, K, T, r)- where $S_0 \geq 0$ is the spot price, $K \geq 0$ is the strike price, $T \geq 0$ is the maturity, and r is the risk-free rate of return- one can use the well-known Black-Scholes formula to compute option prices as a function of σ, the volatility. Similarly, if we are given some option price, it is possible to find the unique value of σ- which we will denote $\widehat{\sigma}$ and refer to as implied volatility- that corresponds to that option price within the Black-Scholes framework. Given a set of observed market Call prices at different strikes and at the same fixed maturity, the $\widehat{\sigma}$ corresponding to each Call price varies with the strike. This phenomenon is called the implied volatility smile, which directly contradicts the assumption of the Black-Scholes framework that σ is constant. Indeed, more sophisticated models than Black-Scholes are required to explain observed phenomena in market data.

An extensive number of augmented stock price models that attempt to explain the implied volatility smile have arisen; exponential Lévy models, local volatility models, and stochastic volatility models to name a few. A particular point of both theoretical

and practical interest is how the implied volatility behaves in such stock price models; especially the small and large-time behaviour. A wide variety of mathematical tools and techniques are available for the study of implied volatility asymptotics.

The addition of jumps to stock price models has been a popular extension to the Black-Scholes framework; in particular jumps may be used to explain steeper implied volatility smiles in small-time. Forde and Figueroa-López [FF12] use an exponential Lévy process to model a stock price, and give a small-time out-the-money expansion for Call option prices, which is then used to attain a small-time implied volatility expansion. Tankov and Mijatović [MT16] also consider the small-time implied volatility asymptotics of an exponential Lévy process, where the strike is a specific time dependent function. Figueroa-López, Gong, and Houdré [FGH16] give a second order at-the-money approximation for a general class of exponential Lévy models, and then deduce the small-time implied volatility; Figueroa-Lòpez and Ólaffson [FO16] then relax the conditions to their weakest possible such that the previous expansion [FGH16] remains well-defined and extend to stochastic volatility models with state independent jumps. Medvedev and Scaillet [MS07] provide the small-time implied volatility behaviour for a stochastic volatility model with jumps using PDE techniques. Using the Lewis-Lipton formula, saddlepoint methods, and known results on the decay of Fourier integrals, Andersen and Lipton [AL12] provide small-time, large-time, and extreme strike asymptotics of the implied volatility of exponential Lévy models in their survey. The focus of this thesis, however, is rough volatility models without jumps, and the above references to jump models are given for the sake of completeness.

Large deviations theory provides a powerful and precise tool to study the asymptotic behaviour of stochastic and local-stochastic volatility models, and has been used quite extensively: [FJ09], [FFF10], [FJ11], [GJR18b] for example. Jacquier and Roome [JR15] use so-called sharp large deviations techniques to arrive at small-time and large-time expansions of the forward implied volatility smile. Paulot [Pau15] uses a heat kernel expansion to study the small-time implied volatility behaviour in stochastic volatility models. Armstrong, Forde, Lorig, and Zhang [AFLZ17] also use a heat kernel expansion, as well as Laplace's method, to study the small-time implied volatility behaviour of a local-stochastic volatility model. There have been a number of recent results on the

asymptotic behaviour of rough volatility models; these are presented together with their respective rough volatility models below.

We begin this Chapter by introducing implied volatility in Section 1.1. The observed convexity of implied volatility in financial markets motivates the development of sophisticated models that may capture this convexity: in Section 1.2 we present local and stochastic volatility models capable of doing so, and then introduce rough volatility models in Section 1.3. We present the theoretical tools, the theory of Gaussian measures on (infinite dimensional) Banach spaces and large deviations theory, in Sections 1.4 and 1.5 respectively. We establish some fundamental results[1] to be used subsequently in the thesis. The Chapter finishes with an application of large deviations theory to implied volatility asymptotics, and an overview of other asymptotic methods for implied volatility analysis.

1.1 Implied Volatility

Let us begin by assuming that we model a stock price process S, starting at S_0, using any (non-negative) stochastic process, other than a geometric Brownian motion. For $(K,T) \in \mathbb{R}_+ \times \mathbb{R}_+$ we may define $\mathrm{C}(K,T) := \mathbb{E}\left[(S_T - K)^+\right]$ as the Call price function; as convention we assume $r = 0$, unless stated explicitly.

Definition 1.1.1. [Gat06, Chapter 1] For a given strike $K \geq 0$ and maturity $T \geq 0$ the implied volatility $\widehat{\sigma}(K,T)$ is defined as the unique, non-negative solution to the equation

$$\mathrm{C}_{\mathrm{BS}}(S_0, K, T, \widehat{\sigma}(K,T)) = \mathrm{C}(K,T),$$

where C_{BS} denotes the Call price within the Black-Scholes framework.

Remark 1.1.2. Note that the implied volatility is simply a means of giving and comparing the relative value of options. Furthermore, we may choose another options pricing framework rather than Black-Scholes to define implied volatility; however Black-Scholes is favoured due to its simplicity and tractability.

[1] This Chapter includes results from article [JPS18].

Let us briefly recall that empirical evidence [Gat06, Figure 3.2., Page 36] shows that the implied volatility smile steepens as the time to maturity decreases; the smile becomes more flat as the time to maturity increases. In equity markets the implied volatility smile is typically asymmetric: in particular the in-the-money (ITM) slope is steeper than the out-the-money (OTM) slope, and these features become more pronounced as the time to maturity decreases.

Remark 1.1.3. Implied volatility $\widehat{\sigma}(\cdot,\cdot)$ as function of both strike K and time to maturity T, [Gat06, Figure 3.2., Page 36], is referred to as the implied volatility surface.

1.2 Local and Stochastic Volatility Models

In this Section we give an overview of some approaches used to model volatility, starting with Dupire's local volatility model and moving on to the classical SABR and Heston stochastic volatility models. We finish with some more advanced models: Bergomi's multi-factor forward variance model and Guyon's path dependent volatility model.

1.2.1 Dupire's Local Volatility Model

One method to model volatility, originally developed by Bruno Dupire [Dup94], is to assume that volatility is a deterministic function, depending on the current stock price and the current time. Such models are referred to as local volatility models. More formally, the stock price process is defined as the unique, strong solution to

$$\mathrm{d}S_t = S_t\left(r\mathrm{d}t + \sigma(S_t,t)\mathrm{d}W_t\right), \quad S_0 > 0. \tag{1.1}$$

By defining $\mathrm{C}(K,T) := e^{-rT}\mathbb{E}\left[(S_T-K)^+\right]$ and imposing some mild conditions on the local volatility function $\sigma(\cdot,\cdot)$, it can be shown that C satisfies the following PDE:

$$\begin{cases} \dfrac{\partial \mathrm{C}}{\partial T} = \dfrac{\sigma(K,T)^2K^2}{2}\dfrac{\partial^2\mathrm{C}}{\partial K^2} - rK\dfrac{\partial \mathrm{C}}{\partial K}, \\ \mathrm{C}(K,0) = (S_0-K)^+. \end{cases} \tag{1.2}$$

From here, one can then rearrange the PDE in (1.2) to arrive at

$$\sigma(K,T) = \sqrt{2\frac{\frac{\partial \mathrm{C}}{\partial T} + rK\frac{\partial \mathrm{C}}{\partial K}}{K^2\frac{\partial^2 \mathrm{C}}{\partial K^2}}}. \tag{1.3}$$

That is, given a surface $\mathscr{C} := \{\mathrm{C}(K,T)\}_{K,T\geq 0}$ of Call prices at all strikes and maturities, one can define the local volatility function $\sigma(\cdot,\cdot)$ by (1.3) such that all Call prices in the model (1.1) are consistent with the given Call price surface $\mathscr{C}$. From here, it is possible to then express $\sigma(\cdot,\cdot)$ in terms of the Black-Scholes implied volatility $\widehat{\sigma}(\cdot,\cdot)$; see [Gat06, Pages 11-13]. Specifically, we have that

$$\sigma(K,T)^2 = \frac{\frac{\partial v}{\partial T}}{1 - \frac{x}{v}\frac{\partial v}{\partial x} + \frac{1}{4}(-\frac{1}{4} - \frac{1}{v} + \frac{x^2}{v^2})(\frac{\partial v}{\partial x})^2 + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}},$$

where $v = v(K,T) := \widehat{\sigma}(K,T)^2 T$ defines the total implied variance, and $x := \log(\frac{K}{S_0 e^{rT}})$. This in turn allows us to describe the implied volatility smile in terms of the local volatility function.

1.2.2 Classical Stochastic Volatility Models

Stochastic volatility processes, where the volatility itself is modeled as a stochastic process, are also used to explain implied volatility smiles. Additionally, stochastic volatility models allow the explicit quantification of the volatility processes' volatility.

The SABR Model

The first stochastic volatility model that we present is the stochastic alpha, beta, rho (SABR) model, introduced by Hagan, Kumar, Lesniewski, and Woodward [HKLW02]. The dynamics of the forward price S and the volatility process v are determined by the following system of SDEs:

$$\begin{aligned} \mathrm{d}S_t &= v_t S_t^{\beta}\mathrm{d}W_t, & S_0 > 0, \\ \mathrm{d}v_t &= \alpha v_t \mathrm{d}Z_t, & v_0 > 0, \\ \langle W_\cdot, Z_\cdot\rangle_t &= \rho t. \end{aligned} \tag{1.4}$$

W and Z are two standard Brownian motions; $\alpha \geq 0$, $\beta \in [0,1]$, and $\rho \in (-1,1)$. While (1.2) and (1.3) do hold statically, Hagan, Kumar, Lesniewski, and Woodward [HKLW02] show that a movement in the underlying causes $\sigma(\cdot,\cdot)$ to move, but in the opposite direction to what is observed in the market. Indeed, any movement in the underlying should cause the smile generated by the local volatility function $\sigma(\cdot,\cdot)$ to move in the same direction, whereas Dupire's local volatility function moves in the opposite direction. The SABR model, however, does generate a smile that behaves dynamically as observed market smiles do. As well as capturing observed dynamic smile behaviour, the SABR model also reproduces the shape of observed smiles from market data fairly well: see, for example, [HKLW02, Figure 3.3.].

The Heston Model

The Heston model, introduced by Steve Heston [Hes93], is a very popular stochastic volatility model that is still widely used today. The Heston model is particularly popular because the characteristic function of the log stock price process is available in closed form, at any time t. This, amongst other things, allows pricing of options by Fourier Transforms, see [CM99]. The dynamics of the stock price process S are given by:

$$\begin{aligned} \mathrm{d}S_t &= S_t\left(\mu \mathrm{d}t + \sqrt{v_t}\mathrm{d}W_t\right), & S_0 > 0, \\ \mathrm{d}v_t &= \lambda(\theta - v_t)\mathrm{d}t + \xi\sqrt{v_t}\mathrm{d}Z_t, & v_0 > 0, \\ \langle W_\cdot, Z_\cdot\rangle_t &= \rho t. \end{aligned} \tag{1.5}$$

W and Z are two standard Brownian motions; $\mu \in \mathbb{R}$; λ, θ, $\xi > 0$; $\rho \in (-1,1)$. A further assumption (the Feller condition) that $2\lambda\theta > \xi^2$ is usually imposed: this ensures that v is strictly positive, almost surely.

There is a large amount of literature on the implied volatility smile of the Heston model. Setting $\mu = 0$ and defining $X_t := \log\left(\frac{S_t}{K}\right)$, Gatheral derives the following approximation for the implied volatility of the Heston model [Gat06, Equation (3.17), page 34]:

$$\widehat{\sigma}(K,T)^2 \approx \frac{1}{T}\int_0^T \left(\theta' + (v_s - \theta')e^{-(\lambda - \rho\xi/2)s}\right)\mathrm{d}s + \rho\xi\frac{X_T}{T\Sigma_T}\int_0^T\int_0^t \mathbb{E}[v_s]e^{-(\lambda-\rho\xi/2)(t-s)}\mathrm{d}s\mathrm{d}t.$$

θ' is defined as $\frac{\theta\lambda}{\lambda - \rho\xi/2}$, and Σ_t is defined as $\int_0^t \mathbb{E}[v_s]\mathrm{d}s$. Weron and Wystrup [WW05]

show that the Heston implied volatility smile fits observed market smiles well for maturities of 0.25-1.5 years. Forde, Jacquier, and Lee [FJL12] provide a closed form small-time implied volatility expansion for the Heston model; Forde, Jacquier, and Mijatović [FJM10] give an asymptotic expansion for the implied volatility of the Heston model.

1.2.3 Advanced Stochastic Volatility Models

The SABR and Heston models described above can be seen as members of a larger class of stochastic volatility models, where the instantaneous variance is modelled as a stochastic process. Choosing to model the instantaneous variance, however, has the drawback of restricting the variance curve's shape. Bergomi suggested a refinement of these "first generation" stochastic volatility models to so-called "second generation" stochastic volatility models, where the forward variance is modelled instead. Some of the first such forward variance models were introduced by Bergomi [Ber05].

Bergomi's Multi-factor Forward Variance Model

Denoting S as the stock price process and defining $X := \log(S)$, the multi-factor Bergomi model [BG12] has the following dynamics:

$$\begin{aligned} \mathrm{d}X_t &= \sqrt{\mathcal{V}_t^t}\mathrm{d}W_t^1 - \frac{1}{2}\mathcal{V}_t^t\mathrm{d}t, & X_0 &= x_0, \\ \mathrm{d}\mathcal{V}_t^u &= \sigma(t,u,\mathcal{V}_t^u)\cdot \mathrm{d}W_t, & \mathcal{V}_0^u &> 0. \end{aligned} \tag{1.6}$$

$W = (W^1, \ldots, W^d)$ is a d-dimensional Brownian motion, and the volatility-of-volatility $\sigma = (\sigma^1, \ldots, \sigma^d)$ takes values in $\mathbb{R}^d$; $\mathcal{V}_t^u$ is the instantaneous forward variance for a future time u, observed at time t. The initial forward variance $\mathcal{V}_0^u$ can be calibrated to vanilla products in order to recover market prices. One may also choose $\mathcal{V}_0^u = \frac{\mathrm{d}}{\mathrm{d}u}(u\widehat{\sigma}_{VS}(u))$, where $\widehat{\sigma}_{VS}(u)$ is the implied volatility of a variance swap with maturity u.

Bergomi and Guyon [BG12] derive a smile approximation for the multifactor Bergomi model (1.6), by means of an expansion of the volatility-of-volatility. This provides a method to characterise the at-the-money skew and curvature of the implied volatility smile, both close to and far from maturity. An explicit smile approximation for the two

factor Bergomi model and for Heston-type models (i.e. where $v_t = \mathcal{V}_t^t$) are also provided by Bergomi and Guyon.

Path Dependent Volatility Models

Guyon [Guy14] studies a path dependent local-stochastic volatility model, where the volatility at any time t depends on the full trajectory of the stock price process S from 0 to t. Path dependent volatility models have not received the same level of attention as local volatility and stochastic volatility models over the past two decades; Hobson and Rogers [HR98] consider a path dependent volatility model, where the volatility process has dependence on exponentially weighted moments of the historic log stock price process. Guyon [Guy14] defines the path dependent volatility model as the process S that satisfies the following SDE:

$$\mathrm{d}S_t = \sigma(t, S_t, Y_t) S_t \mathrm{d}W_t, \quad S_0 > 0. \tag{1.7}$$

Here Y represents a finite set of path dependent processes, such as the weighted moving average and running maximum/minimum; W is a standard Brownian motion. Much like local-stochastic volatility models, Guyon shows that this path dependent volatility model can be calibrated to observed market smiles; they additionally can produce rich volatility dynamics that are unattainable within the local-stochastic volatility framework. The approach outlined by Guyon is to choose any set Y of path dependent processes and the function $\sigma(\cdot, \cdot, \cdot)$ such that (1.7) has the desired volatility dynamics. The model may then be calibrated to the market by multiplying by a so-called leverage function ℓ :

$$\mathrm{d}S_t = \sigma(t, S_t, Y_t) \ell(t, S_t) S_t \mathrm{d}W_t.$$

Guyon demonstrates one advantage of such a path dependent volatility model: it can be calibrated to a flat smile and still achieve a skewed forward volatility, if the function $\sigma(\cdot, \cdot, \cdot)$ is chosen correctly for the corresponding set of path dependent processes Y. Another advantage of such a model is that, since this path dependent volatility model is able to capture historical volatility patterns, market calibration may also be reconciled by historical calibration.

1.3 Rough Volatility Models

The term "rough volatility" has been adopted from [GJR18a], and refers to a stochastic volatility model where the volatility process is driven by a process with Hölder regularity strictly less than $1/2$. This is typically achieved with a fractional Brownian motion. The justification for rough volatility models is that they are able to capture the small-time power law at-the-money volatility skew observed in financial markets: see the results by Alòs, León, and Vives [ALV07], and by Fukasawa [Fuk17]. We recall the definition of a fractional Brownian motion below, [BHØZ08, Definition 1.1.1.].

Definition 1.3.1. Let $H \in (0,1)$: a fractional Brownian motion $\left(W_t^H\right)_{t\geq 0}$ is a continuous, centred Gaussian process with the following covariance function

$$\mathbb{E}\left[W_t^H W_s^H\right] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad \text{for all } s,t \geq 0.$$

Note that taking $H = 1/2$ recovers the standard Brownian motion. We now recall some of the basic properties of the fractional Brownian motion, for more details and proofs see [BHØZ08, Chapter 1].

Proposition 1.3.2. *Let the process W^H be a fractional Brownian motion. The following properties all hold.*

1. *Stationary increments: $W_{t+s}^H - W_s^H$ and W_t^H are equal in law for all $s,t \geq 0$.*
2. *Self-similarity: W_{at}^H and $a^H W_t^H$ are equal in law for all $a > 0$ and $t \geq 0$.*
3. *For $H \neq 1/2$, W^H is not a semi-martingale.*
4. *W^H has a version whose trajectories are almost surely γ-Hölder continuous, for all $\gamma \in (0,H)$.*
5. *Long range dependence: $\sum_{n=1}^{+\infty} \mathbb{E}\left[W_1^H\left(W_{n+1}^H - W_n^H\right)\right] = +\infty$ for $H \in (1/2,1)$.*
6. *Correlated increments: for $H > 1/2$, the increments of W^H are positively correlated, and in this case W^H is said to be persistent. For $H < 1/2$, the increments of W^H are negatively correlated, and in this case W^H is said to be antipersistent. See Figure 1.1.*

7. *Mandelbrot-Van Ness [MV68] integral representation: let B be a standard Brownian motion. The following representation holds almost surely for all $t \geq 0$:*

$$W_t^H = \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^{0} \left((t-s)^{H-1/2} - (-s)^{H-1/2} \right) \mathrm{d}B_s + \int_0^t (t-s)^{H-1/2} \, \mathrm{d}B_s, \tag{1.8}$$

where Γ is the standard Gamma function.

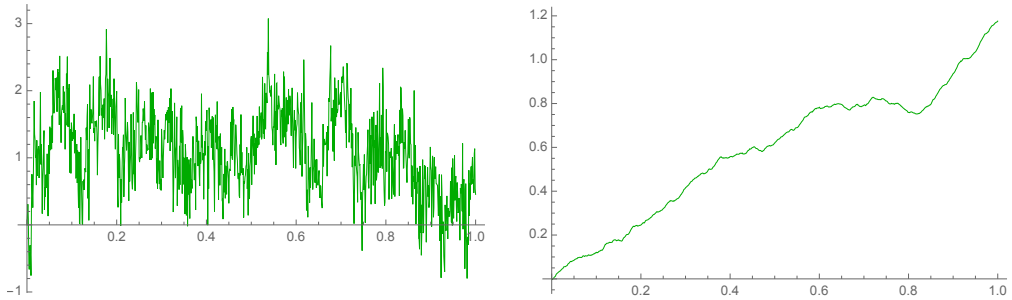

Figure 1.1: Two trajectories of a fractional Brownian motion W^H from $t = 0$ to $t = 1$ with 1001 sample points. In the left plot we set $H = 0.1$ so that W^H is antipersistent; in the right plot we set $H = 0.9$, hence W^H is persistent. Here we have used Mathematica's `'FractionalBrownianMotionProcess'` function to simulate each trajectory.

We briefly recall the definition of a semi-martingale [KS91, Definition 3.1.].

Definition 1.3.3. Let the process M be a continuous $(\mathscr{F}_t)_{t\geq 0}$ local martingale, and A be a càdlàg process of locally bounded variation, adapted to $(\mathscr{F}_t)_{t\geq 0}$. A process X is a $(\mathscr{F}_t)_{t\geq 0}$ continuous semi-martingale if X is adapted to $(\mathscr{F}_t)_{t\geq 0}$ and the following representation holds almost surely, for all $t \geq 0$: $X_t = X_0 + M_t + A_t$.

Remark 1.3.4. Fractional Brownian motion is unsuitable for the application of modelling a stock price process itself, since it is not a semi-martingale for $H \neq 1/2$: [DS94, Theorem 7.2.] tells us that if a stock price process is *not* a semi-martingale, then the "no free lunch with vanishing risk" condition is *not* satisfied. Rogers [Rog97] constructs an explicit arbitrage strategy for a fractional Brownian motion. Since volatility is not a traded asset, however, this fact does not cause issue in models where the volatility itself is driven by a fractional Brownian motion.

Remark 1.3.5. Historically, volatility was thought to exhibit persistence by some financial econometricians, and as a result fractional Brownian motion with $H > 1/2$ has been used in various models to drive volatility, such as Comte and Renault [CR98]. Comte and Renault extend Hull and White's stochastic volatility model so that the log volatility process has long memory and is mean reverting; specifically, the log volatility process is modelled as a fractional Ornstein-Uhlenbeck process.

In very recent years, a renewed interest in the fractional Brownian motion has developed, motivated by the recent work by Gatheral, Jaisson, and Rosenbaum [GJR18a]. Gatheral, Jaisson, and Rosenbaum use high frequency time series data to estimate that the log-volatility of various stock indices behaves in a similar way to a fractional Brownian motion, with $H \approx 0.1$. Very recently, Fukasawa, Takabatake, and Westphal [FTW19] use a quasi-likelihood estimator to investigate the roughness of realised volatility time series, and find H to be even smaller than 0.1.

Gatheral Jaisson, and Rosenbaum [GJR18a] propose the so-called rough fractional stochastic volatility (RFSV) model for the log volatility Y:

$$\mathrm{d}Y_t = \alpha \mathrm{d}W_t^H - \beta(Y_t - \theta)\mathrm{d}t, \quad Y_0 = y_0.$$

Here, $H \in (0, 1/2)$, $\theta \in \mathbb{R}$, and $\alpha, \beta > 0$. This is precisely the model proposed by Comte and Renault [CR98], except that Gatheral, Jaisson, and Rosenbaum [GJR18a] take $H \in (0, 1/2)$, whereas Comte and Renault take $H \in (1/2, 1)$. The RFSV model does not exhibit long memory, and this directly contradicts [CR98]. The rationale provided by Gatheral, Jaisson, and Rosenbaum [GJR18a] to explain this is that, essentially, classical methods may incorrectly identify long memory. Furthermore, while there may be some persistence present in the log volatility process, there is no long memory in terms of power law behaviour. The RFSV model is also shown to produce highly accurate forecasts for realised volatility.

1.3.1 The Rough Bergomi model

Bayer, Friz, and Gatheral [BFG16] introduce a non-Markovian generalisation of Bergomi's "second generation" stochastic volatility model, which they call the rough Bergomi

model. First define the process Z pathwise as

$$Z_t := \int_0^t K_\alpha(s,t)\mathrm{d}W_s^1, \qquad \text{for any } t \in \mathcal{T}, \tag{1.9}$$

where W^1 is a standard Brownian motion. The kernel $K_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ reads

$$K_\alpha(s,t) := \eta\sqrt{2\alpha+1}(t-s)^\alpha, \qquad \text{for all } 0 \le s < t, \tag{1.10}$$

for some strictly positive constant η and $\alpha \in (-1/2, 0)$. Note that, for any $t \ge 0$, the map $s \mapsto K_\alpha(s,t)$ belongs to L^2, so that the stochastic integral (1.9) is well defined. The rough Bergomi model is defined as:

$$\begin{aligned} S_t &= S_0 \exp\left(\int_0^t \sqrt{\mathcal{V}_s^s}\mathrm{d}B_s - \frac{1}{2}\int_0^t \mathcal{V}_s^s \mathrm{d}s\right), && S_0 > 0, \\ \mathcal{V}_t^u &= \mathbb{E}[\mathcal{V}_u^u|\mathscr{F}_t] \exp\left(\int_t^u K_\alpha(s,u)\mathrm{d}W_s^1 - \frac{\eta^2}{2}(u-t)^{2\alpha+1}\right), && \mathcal{V}_0^u > 0, \end{aligned} \tag{1.11}$$

where the process B is defined as $B := \rho W^1 + \sqrt{1-\rho^2}W^2$, for $\rho \in (-1,1)$. W^2 is a standard Brownian motion, independent of W^1. For simplicity, we set $t = 0$ and $u = t$ in the above definition of $\mathcal{V}$, and from this point onwards we refer to $\mathcal{V}$ as v; furthermore we assume that the forward variance curve is flat, i.e. we have that $\mathbb{E}[v_t|\mathscr{F}_0] =: v_0 > 0$ for all $t \ge 0$. We emphasise that *every* reference to the rough Bergomi model from this point onwards will implicitly assume this simplification of (1.11).

Proposition 1.3.6. *The process* $\log v$ *has a modification whose trajectories are almost surely locally* γ*-Hölder continuous, for all* $\gamma \in \left(0, \alpha + \frac{1}{2}\right)$.

Proof. We first prove that Z has a modification whose trajectories are γ-Hölder continuous, for all $\gamma \in (0, \alpha + \frac{1}{2})$. Firstly,

$$\begin{aligned} \mathbb{E}(|Z_t - Z_s|^2) &\le \eta^2(2\alpha+1)\left(\int_s^t |t-u|^{2\alpha}\mathrm{d}u + \int_0^s |(t-u)^\alpha - (s-u)^\alpha|^2\mathrm{d}u\right) \\ &= \eta^2|t-s|^{2\alpha+1} + \eta^2(2\alpha+1)\int_0^s |(t-u)^\alpha - (s-u)^\alpha|^2\mathrm{d}u. \end{aligned}$$

Following the change of variables $s - u = (t-s)y$ the integral becomes

$$|t-s|^{2\alpha+1}\int_0^{\frac{s}{t-s}} |(y+1)^\alpha - y^\alpha|^2\mathrm{d}y,$$

and hence $\int_0^{\frac{s}{t-s}} |(y+1)^\alpha - y^\alpha|^2 \mathrm{d}y$ is finite since $\alpha \in (-\frac{1}{2}, 0)$. Therefore there exists $K > 0$ such that $\mathbb{E}(|Z_t - Z_s|^2) \leq K|t-s|^{2\alpha+1}$ for any $s, t \geq 0$. Applying the Kolmogorov Continuity Theorem [Øks03, Theorem 2.2.3] then yields that the Gaussian process Z has a modification whose trajectories are locally γ-Hölder continuous where $\gamma \in (0, \alpha + \frac{1}{2})$, thus proving the claim. Now, for the process $\log v$, we have

$$\begin{aligned} |\log v_t - \log v_s| &= \left| Z_t - \frac{\eta^2}{2} t^{2\alpha+1} - \left(Z_s - \frac{\eta^2}{2} s^{2\alpha+1} \right) \right| \\ &\leq |Z_t - Z_s| + \frac{\eta^2}{2} \left| t^{2\alpha+1} - s^{2\alpha+1} \right| \\ &\leq C|t-s|^\gamma + \frac{\eta^2}{2} \left| t^{2\alpha+1} - s^{2\alpha+1} \right|, \end{aligned}$$

where C is a strictly positive constant, and $\gamma \in (0, \alpha + 1/2)$. Since the map $t \mapsto t^{2\alpha+1}$ is also locally γ-Hölder continuous for all $\gamma \in (0, 2\alpha + 1]$ and in particular for all $\gamma \in (0, \alpha + 1/2)$, it follows that the process $\log v$ has a modification with locally γ-Hölder continuous trajectories, for all $\gamma \in (0, \alpha + \frac{1}{2})$.

□

Remark 1.3.7. As a comparison, the fractional Brownian motion has sample paths that are γ-Hölder continuous for any $\gamma \in (0, H)$, see Proposition 1.3.2, so that the rough Bergomi model also captures this roughness by identification $\alpha = H - 1/2$; in particular these trajectories are rougher than those of the standard Brownian motion, for which $H = 1/2$.

Proposition 1.3.8. *For any $t \geq 0$, (Z_t, B_t) is a Gaussian random variable with mean zero and covariance matrix*

$$\begin{pmatrix} \eta^2 t^{2\alpha+1} & \varrho t^{\alpha+1} \\ \varrho t^{\alpha+1} & t \end{pmatrix},$$

where $\varrho = \frac{\rho\eta\sqrt{2\alpha+1}}{\alpha+1}$. Moreover (Z, B) is a two-dimensional Gaussian process. Furthermore,

$$\mathbb{E}[Z_s Z_t] = \int_0^{s\wedge t} K_\alpha(u,s) K_\alpha(u,t) \mathrm{d}u = \frac{\eta^2(2\alpha+1)}{\alpha+1} (s\wedge t)^{1+\alpha} (s\vee t)^\alpha {}_2F_1\left(-\alpha, 1; 2+\alpha; \frac{s\wedge t}{s\vee t} \right).$$

Proof. Without loss of generality, let us begin by assuming that $s < t$. This then implies that $\mathbb{E}(Z_s Z_t) = \eta^2(2\alpha+1) \int_0^s (t-u)^\alpha (s-u)^\alpha \mathrm{d}u = t^\alpha s^{1+\alpha} \int_0^1 (1-v)^\alpha (1 - sv/t)^\alpha \mathrm{d}v$,

where the second equality follows from a change of variables. Using a standard integral representation of the Gauss hypergeometric function ${}_2F_1$, it follows that $\int_0^s (t-u)^\alpha (s-u)^\alpha \mathrm{d}u = \frac{1}{\alpha+1} {}_2F_1(-\alpha, 1; \alpha+2; s/t)$, from which the result then clearly follows. □

Proposition 1.3.8 implies in particular that the process Z is not stationary, and that the following holds:

Corollary 1.3.9. *The process Z is $(\alpha + \frac{1}{2})$ self-similar: for any $a > 0$, the processes $(Z_{at})_{t\geq 0}$ and $(a^{\alpha+\frac{1}{2}} Z_t)_{t\geq 0}$ are equal in distribution.*

Note then that the parameter α can be used to describe the local and long-term behaviour of Z.

Remark 1.3.10. The process Z in (1.9) is the Holmgren-Riemann-Liouville fractional Brownian motion introduced by Lévy [Lev53], modulo some constant scaling, rather than the more commonly known fractional Brownian motion given by Mandelbrot and Van Ness (see Propoisiton 1.3.2):

$$W_t^H = \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^0 ((t-s)^{H-1/2} - (-s)^{H-1/2}) \mathrm{d}B_s + \int_0^t (t-s)^{H-1/2} \mathrm{d}B_s \right).$$

The Mandelbrot-Van Ness representation of W_t^H requires the knowledge of B from $-\infty$ to t; in comparison we only need to know W^1 from 0 to t to compute the value of Z. Also both Z and W^H are self-similar, but W^H has stationary increments whereas the increments of Z are non-stationary.

Bayer, Friz, and Gatheral [BFG16] provide an options pricing framework for the rough Bergomi model for options on the stock price S itself, as well as options on the integrated variance of S. Bayer, Friz, and Gatheral also give details on simulation, as well as showing the consistency of volatility smiles generated by the rough Bergomi model with those observed in the market. The rough Bergomi model has the advantage of being fairly simple in so far as having only three parameters (α, η, ρ). Furthermore these three parameters each have a clear relation to given features of the implied volatility surface:

- for a fixed maturity, the minimum of the implied volatility smile shifts down and right as ρ approaches -1;

- α determines the small-time decay of the term structure of the at-the-money volatility skew;
- for a fixed α, the level of the volatility skew fixes $\eta\rho$.

These properties provide some intuition as to what values the parameters should take, when calibrating the model from an observed market implied volatility surface. We consider calibration of the rough Bergomi model in more detail in Chapter 4.

1.3.2 Other Rough Volatility Models

A number of other rough volatility models have recently been proposed: [CCR12] introduces the so-called fractional Heston model, which is also studied by Guennoun, Jacquier, and Roome [GJR18b]. The dynamics of the log stock price X are given by

$$\begin{aligned} \mathrm{d}X_t &= \sqrt{v_t^d}\mathrm{d}W_t - \frac{1}{2}v_t^d\mathrm{d}t, & X_0 &= 0, \\ \mathrm{d}v_t &= \lambda(\theta - v_t)\mathrm{d}t + \xi\sqrt{v_t}\mathrm{d}Z_t, & v_0 &> 0, \\ v_t^d &= v_0^d + \mathcal{I}_{0+}^d v_t. \end{aligned} \tag{1.12}$$

W and Z are two independent, standard Brownian motions; λ, θ, ξ are strictly positive, and $d \in (-1/2, 1/2)$. The operator $\mathcal{I}_{0+}^d$ is the classical left fractional Riemann-Liouville integral operator of order d, defined on $L^1[0,1]$ where Γ denotes the standard Gamma function, as

$$\mathcal{I}_{0+}^d f(t) := \int_0^t \frac{(t-s)^{d-1}}{\Gamma(d)} f(s)\mathrm{d}s.$$

Forde and Zhang [FZ17] define the volatility pathwise as the image of a trajectory of a fractional Brownian motion under some Hölder continuous mapping. More specifically, the stock price process S and the volatility process v are determined by the following system of SDEs:

$$\begin{aligned} \mathrm{d}S_t &= S_t\sigma(v_t)\mathrm{d}W_t, & S_0 &> 0, \\ \mathrm{d}v_t &= \mathrm{d}W_t^H, & v_0 &> 0. \end{aligned} \tag{1.13}$$

W is a standard Brownian motion, and W^H is a fractional Brownian motion: both the correlated and uncorrelated cases of W and W^H are considered. The function σ is

γ-Hölder continuous, where $\gamma \in (0, 1]$; Forde and Zhang consider the cases where σ is both bounded and unbounded.

El Euch, Fukasawa, and Rosenbaum [EFR18] show that the difference of two Hawkes processes, which are used to describe buy and sell orders within a high frequency market model, converge to a rough counterpart of the Heston model, different to the fractional Heston model given in (1.12). The stock price process S in the rough Heston model has the following dynamics:

$$\begin{aligned} \mathrm{d}S_t &= S_t\sqrt{v_t}\mathrm{d}W_t, & S_0 > 0, \\ v_t &= v_0 + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}\left[\lambda(\theta - v_s)\mathrm{d}s + \lambda\xi\sqrt{v_s}\mathrm{d}B_s\right], & v_0 > 0. \end{aligned} \tag{1.14}$$

W and B are two standard Brownian motions, with correlation ρ. The parameter α is used to describe the roughness of the trajectories of the volatility process; the parameters λ, θ, ξ are strictly positive and have the same interpretation as the standard Heston model.

The characteristic function for the rough Heston model is also available [ER19, Theorem 4.1.], and is given in terms of the solution to a fractional Ricatti equation. Explicit hedging strategies for the rough Heston model are also available [ER18]. El Euch, Gatheral, and Rosenbaum also establish an approximation for the rough Heston model with the classical Heston model, by an appropriate rescaling for the vol-of-vol parameter [EGR19]. Very recently Dandapani, Jusselin, and Rosenbaum [DJR19] showed that a rescaled quadratic version of a Hawkes process converges to a refinement of the rough Heston model, which satisfies the so-called "strong Zumbach effect" (where the conditional distribution of future volatility depends on both past returns and the past volatility trajectory).

Forde, Smith and Viitasaari [FSV18] extend the expansions in [Fuk17], which are given for European options in rough volatility models driven by a two-sided fractional Brownian motion, to a rough volatility model whose volatility process is driven by a Riemann-Liouville fractional Brownian motion and whose log stock price process contains jumps. The jumps are modelled by the (infinite activity) one-sided tempered Lévy process, whose Lévy measure is $\nu(\mathrm{d}x) = \frac{Ce^{-Mx}}{x^{1+Y}}\mathrm{d}x$. The expansions in this model only require finite history, unlike the results in [Fuk17]; the expansions are then used to describe the

small-time implied volatility behaviour.

A natural question to ask is how rough volatility models behave in the limit as H tends to zero? Neuman and Rosenbaum [NR18] answer this question when the rough volatility process is driven by a fractional Brownian motion. There are many further interesting research questions on this topic that remain unanswered; for example, how does the rough Bergomi model behave as α approaches $-1/2$?

1.4 Gaussian Measures on Banach Spaces

Having provided an introduction to stochastic and rough volatility models, we now introduce the theoretical tools used to prove our main results.

Recall that a centred, i.e. mean zero, process $(Z_t)_{t\in\mathcal{T}}$, where $\mathcal{T} \subseteq \mathbb{N}$ or $\mathbb{R}$, is Gaussian if for all $n \in \mathbb{N}$ and for any $t_1, \dots, t_n \in \mathcal{T}$: the random variables $Z_{t_1}, \dots, Z_{t_n}$ are jointly Gaussian. Recall also that centred Gaussian processes are completely determined by their covariance functions. The following is a summary on Gaussian measures on Banach spaces, for the most part following Carmona and Tehranchi [CT06, Chapter 3]. Let $(\mathscr{E}, \|\cdot\|_{\mathscr{E}})$ be a real, separable Banach space and let $\mathscr{E}^*$ be the topological dual of $\mathscr{E}$, $\langle\cdot,\cdot\rangle_{\mathscr{E}^*\mathscr{E}}$ gives the duality relationship between $\mathscr{E}$ and $\mathscr{E}^*$. Recall that the topological dual $\mathscr{E}^*$ of $\mathscr{E}$ is the space of all continuous linear functionals on $\mathscr{E}$; $\langle\cdot,\cdot\rangle_{\mathscr{E}^*\mathscr{E}} : \mathscr{E}^* \times \mathscr{E} \to \mathbb{R}$ is a bilinear functional such that if $\langle x^*, x\rangle_{\mathscr{E}^*\mathscr{E}} = 0$ for all $x^* \in \mathscr{E}^*$ (resp. $x \in \mathscr{E}$) then $x = 0$ (resp. $x^* = 0$) [AB06].

The following example gives some intuition to the concept of Gaussian measures on Banach spaces and the reproducing kernel Hilbert space (RKHS), which is defined below.

Example 1.4.1. Define $\widetilde{\mathscr{E}} := \{y : [0,1] \to \mathbb{R}\}$, denote the product sigma algebra generated by cylindrical sets $\{y \in \widetilde{\mathscr{E}} : (y(t_1), \dots, y(t_n)) \in R\}$ for $t_1, \dots, t_n \in [0,1]$ and $R \in \mathscr{B}(\mathbb{R}^n)$ by $\widetilde{E}$, and let $(Z_t)_{t\in[0,1]}$ be a centred, continuous Gaussian process on some probability space $\left(\widetilde{\Omega}, \widetilde{\mathscr{A}}, \widetilde{\mathbb{P}}\right)$. Define the co-ordinate process $(Y_t)_{t\in[0,1]}$ by $Y_t(y) := y(t)$, and the co-ordinate map $\widetilde{Y} : \widetilde{\Omega} \to \widetilde{\mathscr{E}}$ by $\widetilde{Y}(y) := Z_\cdot(y)$. This map induces a measure $\overline{\lambda}$ on $\left(\widetilde{\mathscr{E}}, \widetilde{E}\right)$ given by $\overline{\lambda}(\mathcal{B}) = \widetilde{\mathbb{P}}(\widetilde{Y} \in \mathcal{B})$ that clearly is a probability measure.

Despite the fact that the space of real, continuous functions on $[0,1]$, which will be denoted $\mathcal{C} := \mathcal{C}([0,1],\mathbb{R})$ hereafter, is not a measurable subset of $\widetilde{\mathscr{E}}$, it is possible to replace $\widetilde{\mathscr{E}}$ with $\mathcal{C}$, and $\overline{\lambda}$ with its trace[2] λ, using some measure theory manipulations, since $\widetilde{Y}(\omega) \in \mathcal{C}$ for almost all $\omega \in \widetilde{\Omega}$. Hence, define λ (the trace of $\overline{\lambda}$) on $\{\mathcal{E} \cap \mathcal{C} : \mathcal{E} \in \widetilde{\mathscr{E}}\}$ such that $\lambda(\mathcal{C}) = 1$, thus giving a probability measure on $(\mathcal{C}, \mathscr{C})$ where $\mathscr{C}$ is the Borel σ-algebra of $\mathcal{C}$. For $y \in \mathcal{C}$ and any real sequence $\{\alpha_k\}_{k=1,\dots,n} \subset \mathbb{R}$, define the Gaussian random variable $\displaystyle\sum_{k=1}^{n} \alpha_k Y_{t_k}(y) = \sum_{k=1}^{n} \alpha_k Z_{t_k}(y)$ on $(\mathcal{C}, \mathscr{C}, \lambda)$. The topological dual $\mathcal{C}^*$ of $\mathcal{C}$ is the set of all signed measures on $[0,1]$, with duality given by

$$\langle \mu, y \rangle_{\mathcal{C}^*\mathcal{C}} := \int_{[0,1]} y(x)\mu(\mathrm{d}x).$$

Since $\displaystyle\int_{[0,1]} y(x)\delta_{t_k}(\mathrm{d}x) = y(t_k)$, for $\widetilde{\mu} = \sum_{k=1}^{m} \alpha_k \delta_{t_k}$, we obtain

$$\langle \widetilde{\mu}, y \rangle_{\mathcal{C}^*\mathcal{C}} = \sum_{k=1}^{m} \alpha_k \langle \delta_{t_k}, y \rangle_{\mathcal{C}^*\mathcal{C}} = \sum_{k=1}^{m} \alpha_k y(t_k),$$

which is a Gaussian random variable on $(\mathcal{C}, \mathscr{C}, \lambda)$. Any measure on $[0,1]$ may be expressed as the limit of $\sum_{k=1}^{m} \alpha_k \delta_{t_k}$ as m tends to $+\infty$, and the limit of Gaussian random variables is also Gaussian. So $\mu \in \mathcal{C}^*$, when viewed as a random variable on $\mathcal{C}$ by $\langle \mu, y, \rangle_{\mathcal{C}^*\mathcal{C}}$, is Gaussian.

Remark 1.4.2. Consider the case where the centred Gaussian process Z in Example 1.4.1 is a standard Brownian motion: the induced measure ν on $\mathcal{C}$ is called the standard Wiener measure.

Definition 1.4.3. Let $\mathscr{E}$ be a real, separable Banach space with $\mathscr{B}(\mathscr{E})$ its Borel σ-algebra. A centred Gaussian measure μ on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$ is such that every $f^* \in \mathscr{E}^*$ is a centred, real Gaussian random variable on $(\mathscr{E}, \mathscr{B}(\mathscr{E}), \mu)$ when viewed as a random variable defined by $f \mapsto \langle f^*, f \rangle_{\mathscr{E}^*\mathscr{E}}$.

The following proposition [CT06, Proposition 3.1.] characterises Gaussian measures on Banach spaces.

[2] Let μ be a finite measure on $(\Omega, \mathscr{A})$ and let C be a subset of Ω. Denote the σ-algebra generated by $\{A \cap C : A \in \mathscr{A}\}$ as $\mathscr{D}_C$. For $C \subseteq C_1 \in \mathscr{A}$, we define $\widetilde{\mu}(A \cap C) := \mu(A \cap C_1)$; $\widetilde{\mu}$ is a measure on $(C, \mathscr{D}_C)$ that in this context is called the trace of μ.

Proposition 1.4.4. *Any centred Gaussian measure μ on $\mathscr{E}$ is the law of some centred Gaussian process with continuous paths, indexed by some compact metric space.*

As Example 1.4.1 illustrates, every real-valued, continuous and centred Gaussian process induces some measure on $\mathcal{C}$.

Remark 1.4.5. By Proposition 1.4.4, one can construct a centred Gaussian probability measure μ on $\mathscr{E}$ by constructing a corresponding Gaussian process. The above argument may be extended to a d-dimensional centred Gaussian process $(Z_t)_{t\in[0,1]} = \left((Z_t^1, \ldots, Z_t^d)\right)_{t\in[0,1]}$, which induces a Gaussian measure on $\mathscr{E} = \mathcal{C}\left([0,1], \mathbb{R}^d\right) =: \mathcal{C}^d$.

We shall, from now on, let μ be a Gaussian measure on some Banach space $\mathscr{E}$. Introduce the bounded, linear operator $\Gamma : \mathscr{E}^* \to \mathscr{E}$ defined as

$$\Gamma(f^*) := \int_{\mathscr{E}} \langle f^*, f\rangle_{\mathscr{E}^*\mathscr{E}} f \mu(\mathrm{d}f), \tag{1.15}$$

and note in particular that $\langle f^*, f\rangle_{\mathscr{E}^*\mathscr{E}} f$ in the above definition is a random variable on $(\mathscr{E}, \mathscr{B}(\mathscr{E}), \mu)$.

Definition 1.4.6. Let μ be a Gaussian measure on some real, separable Banach space $\mathscr{E}$. Its reproducing kernel Hilbert space (RKHS) $\mathscr{H}_\mu$ is defined as the completion of $\Gamma(\mathscr{E}^*)$ with the following inner product $\langle \Gamma(f^*), \Gamma(g^*)\rangle_{\mathscr{H}_\mu} := \int_{\mathscr{E}} \langle f^*, f\rangle_{\mathscr{E}^*\mathscr{E}} \langle g^*, f\rangle_{\mathscr{E}^*\mathscr{E}} \mu(\mathrm{d}f)$.

Remark 1.4.7. It can be shown [CT06, Chapter 3] that $\mathscr{H}_\mu \subset \mathscr{E}$.

Example 1.4.8. As above, let Z be a centred Gaussian process and $\mathscr{E} = \mathcal{C}$. Fubini's Theorem implies that

$$\begin{aligned}
\Gamma(f^*)(t) &= \int_{\mathscr{E}} \langle f^*, f\rangle_{\mathscr{E}^*\mathscr{E}} f(t) \mu(\mathrm{d}f) \\
&= \int_{\mathscr{E}} \left(\int_{[0,1]} f(x) f^*(\mathrm{d}x) \right) f(t) \mu(\mathrm{d}f) \\
&= \int_{[0,1]} \left(\int_{\mathscr{E}} f(x) f(t) \mu(\mathrm{d}f) \right) f^*(\mathrm{d}x) \\
&= \int_{[0,1]} \rho(x,t) f^*(\mathrm{d}x),
\end{aligned}$$

where $\rho(s,t) := \mathbb{E}[Z_s Z_t]$ is the covariance function of the process Z. Note that the final equality holds because μ is a Gaussian measure on $\mathscr{E}$. The RKHS of μ is then simply the completion of Γ with the inner product given in Definition 1.4.6.

Example 1.4.9. If we take the specific case of Example 1.4.8 where the process Z is a standard Brownian motion, with ν denoting the standard Wiener measure, then

$$\Gamma(f^*)(t) = \int_{[0,1]} (x \wedge t) f^*(\mathrm{d}x) = \int_0^t x f^*(\mathrm{d}x) + \int_t^1 t f^*(\mathrm{d}x) = \int_0^t f^*([x,1])\mathrm{d}x.$$

This then implies that $\Gamma(f^*)'(t) = f^*([t,1])$. The RKHS of ν is called the Cameron-Martin space, and can be defined explicitly as

$$\mathscr{H}_\nu := \left\{ \int_0^t f(u)\mathrm{d}u : f \in \mathrm{L}^2, t \in [0,1] \right\}.$$

The inner product structure of $\mathscr{H}_\nu$ is given by

$$\left\langle \int_0^{\cdot} f_1(u)\mathrm{d}u, \int_0^{\cdot} f_2(u)\mathrm{d}u \right\rangle_{\mathscr{H}_\nu} := \langle f_1, f_2, \rangle_{\mathrm{L}^2}.$$

The following result, due to Cameron and Martin [CM44], essentially describes the law of translations on $\mathcal{C}$ by elements of $\mathscr{H}_\nu$, in terms of the Radon-Nikodym derivative.

Theorem 1.4.10. ***[Lif12, Theorem 5.1., p34]*** *For a centred Gaussian measure μ on $\mathscr{E}$ with $\mathscr{H}_\mu$ RKHS, define $\mu_h(A) := \mu(A - h)$ for all $A \in \mathscr{B}(\mathscr{E})$ and some $h \in \mathscr{E}$. Then μ_h is absolutely continuous with respect to μ if and only if $h \in \mathscr{H}_\mu$. Furthermore, if $h \in \mathscr{H}_\mu$, then the Radon-Nikodym derivative is given by*

$$\frac{\mathrm{d}\mu_h}{\mathrm{d}\mu}(\mathrm{x}) = \exp\left(f^*(\mathrm{x}) - \frac{1}{2}\|h\|^2_{\mathscr{H}_\mu} \right),$$

where $f^* \in \mathscr{E}^*$ such that $\langle f^*, f \rangle_{\mathscr{E}^*\mathscr{E}} = h$.

Example 1.4.11. For the two-dimensional centred Gaussian process (Z, B) that induces the measure μ_2 on $\mathcal{C}^2 := \mathcal{C}([0,1], \mathbb{R}^2)$, the following holds, by an application of Fubini's Theorem

$$\Gamma(f^*)(t_1,t_2) = \int_{[0,1]^2} \rho_2(s_1,s_2,t_1,t_2) f^*(\mathrm{d}s_1, \mathrm{d}s_2). \tag{1.16}$$

The function $\rho_2(s_1,s_2,t_1,t_2)$ is defined as $\rho_2(s_1,s_2,t_1,t_2) := \begin{pmatrix} \mathbb{E}[Z_{s_1}Z_{t_1}] & \mathbb{E}[Z_{s_1}B_{t_2}] \\ \mathbb{E}[Z_{s_2}B_{t_1}] & \mathbb{E}[B_{s_2}B_{t_2}] \end{pmatrix}$.

For the inclusion map $\iota : \mathscr{H}_\mu \to \mathscr{E}$, the space $\iota(\mathscr{H}_\mu)$ is dense in $\mathscr{E}$; it follows then for the adjoint map $\iota^* : \mathscr{E}^* \to \mathscr{H}_\mu^*$ that $\iota^*(\mathscr{E}^*)$ is dense in $\mathscr{H}_\mu^*$. Recall also that $\mathscr{H}_\mu$ and $\mathscr{H}_\mu^*$ are isometrically isomorphic, which we denote by $\mathscr{H}_\mu^* \simeq \mathscr{H}_\mu$, (by the Riesz representation theorem, as $\mathbb{R}$ is the underlying field). Now, consider f^* as a centred Gaussian random variable on $\mathscr{E}$ (i.e. when viewed as a random variable on $(\mathscr{E}, \mathscr{B}(\mathscr{E}), \mu)$ defined by $f \mapsto \langle f^*, f\rangle_{\mathscr{E}^*\mathscr{E}}$, see Definition 1.4.3):

$$\mathbb{E}\left[\langle f^*, f\rangle^2_{\mathscr{E}^*\mathscr{E}}\right] = \int_{\mathscr{E}} \langle f^*, g\rangle_{\mathscr{E}^*\mathscr{E}} \langle f^*, g\rangle_{\mathscr{E}^*\mathscr{E}} \mu(\mathrm{d}g),$$

where g is a dummy variable. By Definition 1.4.6 we write

$$\int_{\mathscr{E}} \langle f^*, g\rangle_{\mathscr{E}^*\mathscr{E}} \langle f^*, g\rangle_{\mathscr{E}^*\mathscr{E}} \mu(\mathrm{d}g) = \left\|f\right\|^2_{\mathscr{H}_\mu}$$

and by the above argument ($\mathscr{H}_\mu^* \simeq \mathscr{H}_\mu$) it follows that

$$\left\|f\right\|^2_{\mathscr{H}_\mu} = \left\|\iota^* f^*\right\|^2_{\mathscr{H}_\mu^*}.$$

With this in mind we now formulate an equivalent definition to Definition 1.4.6 for the RKHS of μ; see, for example [DS89, Page 88] for a similar definition.

Definition 1.4.12. Let $\mathscr{H}_\mu$ be real, separable Hilbert space such that $\mathscr{H}_\mu \subset \mathscr{E}$: it is the RKHS of μ if the following two conditions hold.

1. There exists some embedding $I : \mathscr{H}_\mu \to \mathscr{E}$, i.e. an injective continuous map whose image is dense in $\mathscr{E}$.
2. Any $f^* \in \mathscr{E}^*$ is a centred Gaussian random variable on $\mathscr{E}$ with variance $\left\|I^* f^*\right\|^2_{\mathscr{H}_\mu^*}$, where the adjoint of I is $I^* : \mathscr{E}^* \to \mathscr{H}_\mu^*$.

Remark 1.4.13. The embedding I need not necessarily be the inclusion map.

Remark 1.4.14. Given a triplet $(\mathscr{E}, \mathscr{H}_\mu, \mu)$, consider the inclusion map $\widehat{I} : \mathscr{E}^* \to L^2(\mathscr{E}, \mu)$ (we think of $\mathscr{E}^*$ as a dense subset in $\mathscr{H}_\mu^* \simeq \mathscr{H}_\mu$ by ι^*). Since $\widehat{I}$ preserves the Hilbert space structure of $L^2(\mathscr{E}, \mu)$, it is possible to extend it to an isometric embedding from $\mathscr{H}_\mu$ to $L^2(\mathscr{E}, \mu)$, i.e. $\mathcal{I} : \mathscr{H}_\mu^* \to L^2(\mathscr{E}, \mu)$ such that $\left\|\mathcal{I} f^*\right\|_{\mathscr{H}_\mu^*} = \left\|f^*\right\|_{L^2(\mathscr{E},\mu)}$.

For convenience of notation, for function $\varphi : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, introduce the L^2-operator $\mathcal{I}^\varphi$ as

$$\mathcal{I}^\varphi f := \int_0^\cdot \varphi(u, \cdot) f(u) \mathrm{d}u. \tag{1.17}$$

Whenever the function φ is constant, equal to c, we shall write $\mathcal{I}^c$ without ambiguity. Requirements on φ in the context of RKHS will be introduced below.

Proposition 1.4.15. *For the process Z defined in* (1.9)*, let μ denote the corresponding induced Gaussian measure on $\mathcal{C}$. The RKHS of μ is $\mathscr{H}^{K_\alpha} := \left\{\mathcal{I}^{K_\alpha} f : f \in \mathrm{L}^2\right\}$, with inner product $\left\langle \mathcal{I}^{K_\alpha} f_1, \mathcal{I}^{K_\alpha} f_2 \right\rangle_{\mathscr{H}^{K_\alpha}} := \langle f_1, f_2 \rangle_{\mathrm{L}^2}$.*

Proof. First we note that, by definition, the operator $\mathcal{I}^{K_\alpha}$ is surjective on $\mathscr{H}^{K_\alpha}$. Let $f_1, f_2 \in \mathrm{L}^2$ such that $\mathcal{I}^{K_\alpha} f_1(t) = \mathcal{I}^{K_\alpha} f_2(t)$ for all $t \in [0, 1]$. Then $\int_0^t (t-u)^\alpha [f_1(u) - f_2(u)] \mathrm{d}u = 0$. The Titchmarsh Convolution Theorem [Tit26, Theorem VII.] then implies that $f_1 = f_2$ almost everywhere. Hence $\mathcal{I}^{K_\alpha} : \mathrm{L}^2 \to \mathscr{H}^{K_\alpha}$ is a bijection. We can define $\left\langle \mathcal{I}^{K_\alpha} f_1, \mathcal{I}^{K_\alpha} f_2 \right\rangle_{\mathscr{H}^{K_\alpha}} := \langle f_1, f_2 \rangle_{\mathrm{L}^2}$, which does indeed define an inner product on $\mathscr{H}^{K_\alpha}$, due to the linearity of the operator $\mathcal{I}^{K_\alpha}$ and since $\langle \cdot, \cdot \rangle_{\mathrm{L}^2}$ is itself an inner product; therefore $\left(\mathscr{H}^{K_\alpha}, \langle \cdot, \cdot \rangle_{\mathscr{H}^{K_\alpha}}\right)$ is a real inner product space. In order for $\mathscr{H}^{K_\alpha}$ to satisfy Definition 1.4.12, we first need to show that it is a separable Hilbert space. Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence in L^2 such that $\left\{\mathcal{I}^{K_\alpha} f_n\right\}_{n \in \mathbb{N}}$ converges to $\mathcal{I}^{K_\alpha} f$ in L^2. Therefore $\left\|\mathcal{I}^{K_\alpha} f_n - \mathcal{I}^{K_\alpha} f_m\right\|_{\mathscr{H}^{K_\alpha}} = \left\|f_n - f_m\right\|_{\mathrm{L}^2}$ tends to zero as n and m tend to infinity. Since L^2 is a complete (Hilbert) space, there exists a function $\widetilde{f} \in \mathrm{L}^2$ such that the sequence $\{f_n\}_{n \in \mathbb{N}}$ converges to $\widetilde{f}$. Assume that $f \neq \widetilde{f}$, then, since $\mathcal{I}^{K_\alpha}$ is a bijection, the triangle inequality yields

$$0 < \left\|\mathcal{I}^{K_\alpha} f - \mathcal{I}^{K_\alpha} \widetilde{f}\right\|_{\mathscr{H}^{K_\alpha}} \leq \left\|\mathcal{I}^{K_\alpha} f - \mathcal{I}^{K_\alpha} f_n\right\|_{\mathscr{H}^{K_\alpha}} + \left\|\mathcal{I}^{K_\alpha} \widetilde{f} - \mathcal{I}^{K_\alpha} f_n\right\|_{\mathscr{H}^{K_\alpha}},$$

which converges to zero as n tends to infinity. Therefore $f = \widetilde{f}$, $\mathcal{I}^{K_\alpha} f \in \mathscr{H}^{K_\alpha}$ and $\mathscr{H}^{K_\alpha}$ is complete, so that it is a real Hilbert space. L^2 is separable with countable orthonormal basis, which we will call $\{\phi_n\}_{n \in \mathbb{N}}$; then we also get that $\left\{\mathcal{I}^{K_\alpha} \phi_n\right\}_{n \in \mathbb{N}}$ is an orthonormal basis for $\mathscr{H}^{K_\alpha}$. So, $\mathscr{H}^{K_\alpha}$ is a real, separable Hilbert space as required.
We now wish to find a dense embedding $I : \mathscr{H}^{K_\alpha} \to \mathscr{E}$, rewriting $\mathscr{H}^{K_\alpha}$ as in the Proposition. Since $\mathscr{H}^{K_\alpha} \subset \mathcal{C}$, take the embedding to be the inclusion map i.e. $I = \iota$.

Take $f_c(s) \equiv c$ for $s \in [0,1]$ where c is some arbitrary non-zero constant so that

$$\int_0^t K_\alpha(u,t) f_c(u) \mathrm{d}u = \frac{c\eta\sqrt{2\alpha+1}}{\alpha+1} t^{\alpha+1} =: g(t).$$

For any $t_1, t_2 \in [0,1]$, $g(t_1) = g(t_2)$ if and only if $t_1 = t_2$ and so we see that g is in the separating set of $\mathcal{C}$. Hence, by the Stone-Weierstrass Theorem [Will70, Theorem 44.5., page 291] it follows immediately that $\mathscr{H}^{K_\alpha}$ is dense in $\mathcal{C}$.

Finally, let us take $f^* \in \mathcal{C}^*$: since μ is a Gaussian probability measure on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$, f^* is a centred, real Gaussian random variable on $(\mathscr{E}, \mathscr{B}(\mathscr{E}), \mu)$ by Definition 1.4.3 . In turn, Remark 1.4.14 implies the following:

$$\left\| I^* f^* \right\|^2_{\mathscr{H}^{K_\alpha *}} = \left\| f^* \right\|^2_{L^2(\mathscr{E},\mu)} = \int_{\mathscr{E}} (f^*)^2 \, \mathrm{d}\mu = \mathrm{Var}(f^*).$$

So, we conclude that $\mathscr{H}^{K_\alpha} := \left\{ \mathcal{I}^{K_\alpha} f : f \in \mathrm{L}^2 \right\}$ is the RKHS of μ.

□

By applying the same arguments as Proposition 1.4.15, it is possible to generalise the result in order to give the RKHS for the measure induced by a Gaussian process Y, defined pathwise as $Y_t := \int_0^t \varphi(s,t) \mathrm{d}W_s$ for $t \in [0,1]$ and some kernel φ on $\mathbb{R}_+ \times \mathbb{R}_+$ such that the stochastic integral is indeed well-defined. We briefly present this result below, after first introducing the following assumption:

Assumption 1.4.16. For $\varphi : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ such that $\varphi(u,\cdot) \in \mathrm{L}^2$ for all $u \in [0,1]$, there exists $\phi \in L^2([-1,0], \mathbb{R})$ such that $\int_{-\varepsilon}^0 |\phi(s)| \mathrm{d}s > 0$ for all $\varepsilon > 0$ and $\varphi(s,t) = \phi(t-s)$ for all $s, t \in [-1,0]$.

Proposition 1.4.17. *Let φ satisfy Assumption 1.4.16 such that $\mathcal{I}^\varphi$ is injective on L^2. The RKHS of the measure induced the process by $\int_0^\cdot \varphi(u,\cdot) \mathrm{d}W_u$ on $\mathcal{C}$ is given by $\mathscr{H}^\varphi := \{\mathcal{I}^\varphi f : f \in \mathrm{L}^2\}$, with inner product $\langle \mathcal{I}^\varphi f_1, \mathcal{I}^\varphi f_2 \rangle_{\mathscr{H}^\varphi} := \langle f_1, f_2 \rangle_{\mathrm{L}^2}$.*

Proof. The arguments presented in the proof of Proposition 1.4.15 can be used to verify that $\mathscr{H}^\varphi$ is indeed a real, separable Hilbert space. We now wish to find a dense embedding $I : \mathscr{H}^\varphi \to \mathscr{E}$ as in Definition 1.4.12. Since $\mathscr{H}^\varphi \subset \mathcal{C}$, take the embedding to

be the inclusion map $I = \iota$. By [Che08, Lemma 2.1], the conditions on ϕ in Assumption 1.4.16 imply that $\mathscr{H}^{\varphi}$ is dense in $\mathcal{C}$. Finally, for $f^* \in \mathcal{C}^*$, the measure μ induced by the process $\int_0^{\cdot} \varphi(u,\cdot)\mathrm{d}W_s$ is a Gaussian probability measure on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$, f^* is a centred, real Gaussian random variable on $(\mathscr{E}, \mathscr{B}(\mathscr{E}), \mu)$ by Definition 1.4.3 . In turn, Remark 1.4.14 implies that I^*, the dual of I, admits an isometric embedding $\bar{I}^*$ such that $\|\bar{I}^* f^*\|^2_{(\mathscr{H}^{\varphi})^*} = \|f^*\|^2_{L^2(\mathscr{E},\mu)} = \int_{\mathscr{E}} (f^*)^2 \mathrm{d}\mu = \mathrm{Var}(f^*)$, and hence $\mathscr{H}^{\varphi}$ is the required RKHS. □

We now extend Proposition 1.4.15 to find the RKHS of μ_2, which is the Gaussian measure on the space $\mathcal{C}^2$ induced by the two-dimensional process $\left((Z_t, W^1_t)\right)_{t\in[0,1]}$, where Z and W^1 are defined in (1.9) and (1.11) respectively.

Proposition 1.4.18. *For the two-dimensional process (Z, B), let μ_2 denote the induced measure on $\mathcal{C}^2$. Then the RKHS of μ_2 is $\mathscr{H}^{K_\alpha}_\rho := \left\{\mathcal{I}^{K_\alpha}_\rho(f_1, f_2) : f_1, f_2 \in \mathrm{L}^2\right\}$ with inner product $\langle \mathcal{I}^{K_\alpha}_\rho(f_1,f_2), \mathcal{I}^{K_\alpha}_\rho(g_1,g_2)\rangle_{\mathscr{H}^{K_\alpha}_\rho} := \langle f_1, g_1\rangle_{\mathrm{L}^2} + \bar{\rho}\langle f_2, g_2\rangle_{\mathrm{L}^2}$, where the operator $\mathcal{I}^{K_\alpha}_\rho : \mathrm{L}^2 \times \mathrm{L}^2 \to \mathscr{H}^{K_\alpha}_\rho$ is defined as*

$$\mathcal{I}^{K_\alpha}_\rho(f_1, f_2) := \left(\int_0^{\cdot} K_\alpha(s,\cdot) f_1(s)\mathrm{d}s, \int_0^{\cdot} \rho f_1(s) + \bar{\rho} f_2(s)\mathrm{d}s \right).$$

Proof. First we prove that the operator $\mathcal{I}^{K_\alpha}_\rho$ is bijective. Clearly it is surjective. To prove injectivity, first note that $\int_0^{\cdot} K_\alpha(s,\cdot) f_1(s)\mathrm{d}s = \int_0^{\cdot} K_\alpha(s,\cdot) g_1(s)\mathrm{d}s$ is equivalent to $\int_0^{\cdot} K_\alpha(s,\cdot)[f_1(s) - g_1(s)]\mathrm{d}s = 0$, the Titchmarsh convolution theorem then implies that $f_1 = g_1$. Therefore $\mathcal{I}^{K_\alpha}_\rho(f_1,f_2) = \mathcal{I}^{K_\alpha}_\rho(g_1,g_2)$ implies that

$$\int_0^{\cdot} \rho f_1(s) + \bar{\rho} f_2(s)\mathrm{d}s = \int_0^{\cdot} \rho g_1(s) + \bar{\rho} g_2(s)\mathrm{d}s,$$

since $f_1 = g_1$ it then follows that $f_2 = g_2$ and so the operator $\mathcal{I}^{K_\alpha}_\rho$ is indeed bijective.

Next we prove that $\left(\mathscr{H}^{K_\alpha}_\rho, \langle\cdot,\cdot\rangle_{\mathscr{H}^{K_\alpha}_\rho}\right)$ is a real, separable Hilbert space. Note that $\langle\cdot,\cdot\rangle_{\mathscr{H}^{K_\alpha}_\rho}$ is indeed an inner product due to the linearity of $\mathcal{I}^{K_\alpha}_\rho$ and $\langle\cdot,\cdot\rangle_{\mathrm{L}^2}$ itself being an inner product. For $i = 1, 2$ let $\{f_{i,n}\}_{n\in\mathbb{N}}$ be a sequence in L^2 such that $\left\{\mathcal{I}^{K_\alpha}_\rho(f_{1,n}, f_{2,n})\right\}_{n\in\mathbb{N}}$ converges to $\mathcal{I}^{K_\alpha}_\rho(f_1,f_2)$ in L^2. Therefore

$$\left\|\mathcal{I}^{K_\alpha}_\rho(f_{1,n}, f_{2,n}) - \mathcal{I}^{K_\alpha}_\rho(f_{1,m}, f_{2,m})\right\|_{\mathscr{H}^{K_\alpha}_\rho} = \left\|f_{1,n} - f_{1,m}\right\|_{\mathrm{L}^2} + \bar{\rho}\left\|f_{2,n} - f_{2,m}\right\|_{\mathrm{L}^2}$$

tends to zero as n and m tend to infinity. Since L^2 is a complete (Hilbert) space, there exists a functions $\widetilde{f}_i \in \mathrm{L}^2$ such that the sequence $\{f_{i,n}\}_{n\in\mathbb{N}}$ converges to $\widetilde{f}_i$ for $i = 1, 2$. Assume that $f_i \neq \widetilde{f}_i$, then, since $\mathcal{I}_\rho^{K_\alpha}$ is a bijection, the triangle inequality yields

$$\begin{aligned} 0 &< \left\|\mathcal{I}_\rho^{K_\alpha}(f_1, f_2) - \mathcal{I}_\rho^{K_\alpha}(\widetilde{f}_1, \widetilde{f}_2)\right\|_{\mathscr{H}_\rho^{K_\alpha}} \\ &\leq \left\|\mathcal{I}_\rho^{K_\alpha}(f_1, f_2) - \mathcal{I}_\rho^{K_\alpha}(f_{1,n}, f_{2,n})\right\|_{\mathscr{H}_\rho^{K_\alpha}} + \left\|\mathcal{I}_\rho^{K_\alpha}(\widetilde{f}_{1,n}, \widetilde{f}_{2,n}) - \mathcal{I}_\rho^{K_\alpha}(f_{1,n}, f_{2,n})\right\|_{\mathscr{H}_\rho^{K_\alpha}}, \end{aligned}$$

which converges to zero as n tends to infinity. Therefore $f_i = \widetilde{f}_i$, $\mathcal{I}_\rho^{K_\alpha}(f_1, f_2) \in \mathscr{H}_\rho^{K_\alpha}$ and $\mathscr{H}_\rho^{K_\alpha}$ is complete, so that it is a real Hilbert space. L^2 is separable with countable orthonormal bases, which we will call $\{\phi_{i,n}\}_{n\in\mathbb{N}}$ for $i = 1, 2$; then we also get that $\left\{\mathcal{I}_\rho^{K_\alpha}(\phi_{1,n}, \phi_{2,n})\right\}_{n\in\mathbb{N}}$ is an orthonormal basis for $\mathscr{H}_\rho^{K_\alpha}$. So, $\mathscr{H}_\rho^{K_\alpha}$ is a real, separable Hilbert space as required.

Finally we prove that $\mathscr{H}_\rho^{K_\alpha}$ is dense in $\mathcal{C}^2$. Fix $\varepsilon > 0$ and $(u, v) \in \mathcal{C}^2$: because $\mathscr{H}^{K_\alpha}$ is dense in $\mathcal{C}$ we can chose $f_1^* \in \mathrm{L}^2$ such that $\| \int_0^\cdot K_\alpha(s, \cdot) f_1^*(s)\mathrm{d}s - u\|_\infty < \varepsilon/2$. Now define $v^* \in \mathcal{C}$ as $v^*(t) := v(t) - \int_0^t \rho f_1^*(s)\mathrm{d}s$. A second application of the Stone-Weirstrauss Theorem implies that there exists $f_2^* \in \mathrm{L}^2$ such that $\| \int_0^\cdot \bar{\rho} f_2^*(s)\mathrm{d}s - v^*\|_\infty < \varepsilon/2$. Together this implies that there exists $(f_1^*, f_2^*) \in \mathrm{L}^2 \times \mathrm{L}^2$ such that $\left\|\mathcal{I}_\rho^{K_\alpha}(f_1^*, f_2^*) - (u, v)\right\|_\infty < \varepsilon$ i.e. that $\mathscr{H}_\rho^{K_\alpha}$ is indeed dense in $\mathcal{C}^2$.

Lastly, for $f^* \in \mathcal{C}^{2*}$: f^* is a real, centred Gaussian random variable on $(\mathcal{C}^2, \mathscr{B}(\mathcal{C}^2), \mu_2)$ and so

$$\mathrm{Var}(f^*) = \int_{\mathcal{C}^2} (f^*)^2 \mathrm{d}\mu_2 = \|f^*\|^2_{\mathrm{L}^2(\mathcal{C}^2, \mu_2)} = \|\iota^* f^*\|^2_{(\mathscr{H}_\rho^{K_\alpha})^*}.$$

Thus, the conditions of Definition 1.4.12 have been satisfied, and we conclude that $\mathscr{H}_\rho^{K_\alpha}$ is indeed the RKHS of (Z, B).

□

1.5 Large Deviations Theory

Large deviations theory is an area of probability theory focused on characterising the exponential decay of rare events. Informally, let us consider a family of probability measures $(\mu_\varepsilon)_{\varepsilon>0}$ converging weakly to δ_x as ε tends to zero: given a set B such that

$x \in B$, large deviations theory aims to provide the exponential rate at which $\mu_\varepsilon(B^c)$ tends to zero, as ε tends to zero. We recall some facts about large deviations, as well as some results essential for proving Theorem 2.1.2, using [DS89] and [DZ10] as our guide.

Definition 1.5.1. A rate function $\Lambda : \mathscr{E} \to [0, +\infty]$ is a lower semi-continuous function: for all $\mathrm{x}_0 \in \mathscr{E}$ $\liminf_{\mathrm{x}\to\mathrm{x}_0} \Lambda(\mathrm{x}) \geq \Lambda(\mathrm{x}_0)$.

Definition 1.5.2. A family of probability measures $(\mu_\varepsilon)_{\varepsilon>0}$ on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$ is said to satisfy a large deviations principle (LDP) as ε tends to zero with speed $\varepsilon^{-\gamma}$ and rate function Λ if, for any $B \in \mathscr{B}(\mathscr{E})$,

$$-\inf_{\mathrm{x}\in B^\circ} \Lambda(\mathrm{x}) \leq \liminf_{\varepsilon\downarrow 0} \varepsilon^\gamma \log \mu_\varepsilon(B) \leq \limsup_{\varepsilon\downarrow 0} \varepsilon^\gamma \log \mu_\varepsilon(B) \leq -\inf_{\mathrm{x}\in\overline{B}} \Lambda(\mathrm{x}), \tag{1.18}$$

where $\overline{B}$ and B° denote respectively the closure and the interior of B.

Remark 1.5.3. A stochastic process X is said to satisfy an LDP as t tends to zero if the family of probability measures $(\mathbb{P}(X_t \in \cdot))_{t>0}$ satisfies an LDP as t tends to zero.

Remark 1.5.4. We say that the family of probability measures $(\mu_\varepsilon)_{\varepsilon>0}$ on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$ satisfies an LDP as ε tends to $+\infty$ with rate function Λ and speed ε^γ, if we replace ε^γ with $1/\varepsilon^\gamma$ in (1.18) and take ε tending to $+\infty$ rather than zero.

We now present two classical theorems from large deviations theory: Cramér's Theorem, [DZ10, Theorem 2.2.30.], and the Gärtner-Ellis Theorem, [DZ10, Theorem 2.3.6.], which is referred to below in Section 1.6. Let $Y_1, Y_2, \ldots$ be i.i.d. d-dimensional random vectors, and let

$$\overline{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$$

be the empirical mean of $Y_1, Y_2, \ldots$ whose law is denoted ν_n. Define the logarithmic moment generating function Λ of Y_1 as $\Lambda(y) := \log \mathbb{E}\left[e^{\langle y, Y_1\rangle}\right]$ for $y \in \mathbb{R}^d$, and the Fenchel-Legendre transform of Λ as

$$\Lambda^*(x) := \sup_{y\in\mathbb{R}^d} \left\{\langle y, x\rangle - \Lambda(y)\right\}, \quad x \in \mathbb{R}^d.$$

Theorem 1.5.5. ***(Cramér's Theorem) [Cra38]*** *If the logarithmic moment generating functions Λ exists everywhere in $\mathbb{R}^d$ then $\{\nu_n\}_{n\in\mathbb{N}}$ satisfies an LDP on $\mathbb{R}^d$ as n tends to $+\infty$ with speed n and rate function Λ^*.*

Now let $(X_t)_{t\geq 0}$ be a d-dimensional real-valued stochastic process and denote the logarithmic moment generating function of X_t by Λ_t.

Definition 1.5.6. Let $f : \mathbb{R}^d \to (-\infty, +\infty]$ with $\mathrm{dom}(f) := \left\{x \in \mathbb{R}^d : f(x) < +\infty\right\}$; f is essentially smooth if the following three conditions hold.

1. The interior of $\mathrm{dom}(f)$, $\mathrm{dom}(f)^\circ$, is non-empty.
2. f is differentiable at every point in $\mathrm{dom}(f)^\circ$.
3. For all $\{x_t\}_t$ that are sequences in $\mathrm{dom}(f)^\circ$ that converge to a boundary point of $\mathrm{dom}(f)^\circ$: $\lim_{t\to+\infty} |\nabla f(x_t)| = +\infty$.

Assumption 1.5.7. For each $y \in \mathbb{R}^d$, assume that $\Lambda(y) := \lim_{t\to 0} t\Lambda_t\left(y/t\right)$ exists as some extended real number, and denote Λ^* as the Fenchel-Legendre transform of Λ. Assume also that the origin in $\mathbb{R}^d$ belongs to the interior of $\mathrm{dom}\left(\Lambda\right) := \{y \in \mathbb{R}^d : \Lambda(y) < +\infty\}$.

Theorem 1.5.8. ***(Gärtner-Ellis Theorem)*** *Let Assumption 1.5.7 hold. If Λ is lower semi-continuous and essentially smooth in the sense of Definition 1.5.6, then the process X satisfies an LDP on $\mathbb{R}^d$ as t tends to zero with speed t^{-1} and rate function Λ^*.*

We now concentrate on large deviations for Gaussian measures. To remain consistent we denote a real, separable Banach space as $\mathscr{E}$, with its corresponding norm $\|\cdot\|_{\mathscr{E}}$. Let μ be some probability measure on $(\mathscr{E}, \mathscr{B}\left(\mathscr{E}\right))$ with

$$\int_{\mathscr{E}} \exp\left[-\mathrm{i}\langle y, x\rangle_{\mathscr{E}^*\mathscr{E}}\right] \mu(\mathrm{d}x) = \exp\left[\frac{-C_\mu(y,y)}{2}\right]$$

for $y \in \mathscr{E}^*$ and $C_\mu : \mathscr{E}^* \times \mathscr{E}^* \to [0, +\infty)$ a bilinear, symmetric map. That is, μ is a centred Gaussian measure on $(\mathscr{E}, \mathscr{B}\left(\mathscr{E}\right))$. Define $\Lambda_\mu(y) := \frac{C_\mu(y,y)}{2}$, and its Fenchel-Legendre transform $\Lambda_\mu^*(x) := \sup_{y\in\mathscr{E}^*}\left\{\langle y, x\rangle_{\mathscr{E}^*\mathscr{E}} - \Lambda_\mu(y)\right\}$ for $x \in \mathscr{E}$. The following Lemma is proved in [DS89, Lemma 3.4.2].

Lemma 1.5.9. *The following three statements hold for the centred Gaussian measure μ on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$.*

1. *There is some $\alpha \in (0, +\infty)$ such that $\int_{\mathscr{E}} \exp\left[\alpha\|x\|_{\mathscr{E}}^2\right] \mu(\mathrm{d}x)$ is finite.*

2. $C_\mu(y,y) = \int_{\mathscr{E}} \langle y,x\rangle^2_{\mathscr{E}^*\mathscr{E}} \mu(\mathrm{d}x) \leq \|y\|^2_{\mathscr{E}^*} \int_{\mathscr{E}} \|x\|^2_{\mathscr{E}} \mu(\mathrm{d}x) \in (0,+\infty)$ *for all* $y \in \mathscr{E}^*$.

3. Λ^*_μ *defines a rate function on* $\mathscr{E}$ *and satisfies* $\Lambda^*_\mu(ay) = a^2\Lambda^*_\mu(y)$ *for all* $a \in \mathbb{R}$.

For a Gaussian random variable X with distribution μ, define $X^\varepsilon := \varepsilon^{1/2}X$, whose law we will denote as μ_ε. Then the following LDP holds, [DS89, Theorem 3.4.5.].

Lemma 1.5.10. *The family of probability measures* $(\mu_\varepsilon)_{\varepsilon>0}$ *satisfies an LDP on* $\mathscr{E}$ *as* ε *tends to zero with speed* ε^{-1} *and rate function* Λ^*_μ.

Proof. Consider n i.i.d. random variables $(X^1,\dots,X^n)$ where each X^k has distribution μ. Then, let $\frac{1}{n}\sum_{k=1}^n X^k$ have distribution $\mu_{1/n}$. Lemma 1.5.9 then implies that $\int_{\mathscr{E}} \exp\left[\alpha\|x\|^2_{\mathscr{E}}\right] \mu_{1/n}(\mathrm{d}x)$ is finite for some $\alpha \in (0,+\infty)$. Together with [DS89, Theorem 3.3.11.] we conclude that $\left(\mu_{1/n}\right)_{n\geq 1}$ satisfies an LDP with rate function Λ^*_μ. Define $n(\varepsilon) := \left\lfloor \frac{1}{\varepsilon} \right\rfloor \vee 1$, $\ell(\varepsilon) := \varepsilon n(\varepsilon)$ for $\varepsilon > 0$, noting that $\ell(\varepsilon) \in [\frac{1}{2},1]$. For X with distribution $\mu_{1/n(\varepsilon)}$, it follows that $\ell(\varepsilon)^{1/2}X$ has distribution μ_ε, for $\ell(\varepsilon) \in [1-\varepsilon,1]$ and $0<\varepsilon<1$.

Let $B \subset \mathscr{E}$ closed, and define $\widetilde{B} := \left\{\ell^{-1/2}x : \text{ for all } \ell \in \left[\frac{1}{2},1\right],\, x \in B\right\}$ that is also closed:

$$\begin{aligned}
\limsup_{\varepsilon\to 0} \varepsilon \log\left(\mu_\varepsilon(B)\right) &= \limsup_{\varepsilon\to 0} \frac{\ell(\varepsilon)}{n(\varepsilon)} \log\left(\mu_{1/n(\varepsilon)}(\ell(\varepsilon)^{-1/2}B)\right) \\
&\leq \limsup_{\varepsilon\to 0} \frac{1}{n(\varepsilon)} \log\left(\mu_{1/n(\varepsilon)}(\widetilde{B})\right) = \limsup_{n\to+\infty} \frac{1}{n} \log\left(\mu_{1/n}(\widetilde{B})\right) \\
&\leq -\inf_{x\in\widetilde{B}} \Lambda^*_\mu(x).
\end{aligned}$$

Since $\inf_{x\in\widetilde{B}} \Lambda^*_\mu(x) = \inf_{\ell\in\left[\frac{1}{2},1\right]} \inf_{x\in B} \Lambda^*_\mu(\ell^{-1/2}x) = \inf_{\ell\in\left[\frac{1}{2},1\right]} \ell^{-1} \inf_{x\in B} \Lambda^*_\mu(x) = \inf_{x\in B} \Lambda^*_\mu(x)$ we attain the required upper bound for an LDP.

Now for $C \subset \mathscr{E}$ open with $x \in C$, there is some open neighbourhood O_x of x with

$\varepsilon_0 \in \left(0, \frac{1}{2}\right]$ such that $O_x \subseteq \ell(\varepsilon)^{-1/2} C$ for $0 < \varepsilon < \varepsilon_0$. It then follows that:

$$\begin{aligned}
\liminf_{\varepsilon \to 0} \varepsilon \log\left(\mu_\varepsilon(C)\right) &= \liminf_{\varepsilon \to 0} \frac{\ell(\varepsilon)}{n(\varepsilon)} \log\left(\mu_{1/n(\varepsilon)}(\ell(\varepsilon)^{-1/2} C)\right) \\
&\geq \liminf_{n \to +\infty} \frac{1}{n} \log\left(\mu_{1/n}(O_x)\right) \\
&\geq - \inf_{y \in O_x} \Lambda_\mu^*(y) \\
&\geq -\Lambda_\mu^*(x).
\end{aligned}$$

This gives the lower bound and completes the proof.

□

Remark 1.5.11. Theorem 1.5.10 implies that a standard Brownian motion $(W_t)_{t \geq 0}$ satisfies an LDP on $\mathbb{R}$ as t tends to zero with speed t^{-1}, because W_t and $\sqrt{t} W_1$ are equal in law. Furthermore, the proof of Theorem 1.5.10 stills holds for the case where $t^{\alpha+1/2} X$ has law μ_t, with speed $t^{-(2\alpha+1)}$, and the proof can be easily adapted to confirm this case.

Corollary 1.5.12. *Let ν_t be the law of Z_t, defined in* (1.9)*. Then $(\nu_t)_{t>0}$ satisfies an LDP on $\mathbb{R}$ as t tends to zero with speed $t^{-(2\alpha+1)}$ and rate function $\Lambda_\mu^*(x) := \frac{x^2}{2\eta^2}$ for $x \in \mathbb{R}$.*

Proof. Here, $\mathscr{E} = \mathbb{R}$ and $\langle u, v \rangle_{\mathscr{E}^* \mathscr{E}} = uv$. Recall that Z_t and $t^{\alpha+1/2} Z_1$ are equal in law, and the law ν of Z_1 is centred Gaussian with variance η^2:

$$\int_{\mathbb{R}} \exp\left[iyx\right] \mu(\mathrm{d}x) = \exp\left[\frac{-y^2 \eta^2}{2}\right].$$

Taking $C_\mu(x, y) = xy\eta^2$, the remainder of the proof follows from Theorem 1.5.10 and Remark 1.5.11.

□

The following two results will also be essential for establishing an LDP for the rough Bergomi model. For the proofs see [FV10, Theorem C.6.] and [DS89, Theorem 3.4.12.] respectively.

Proposition 1.5.13. ***(The Contraction Principle)*** *Let $\mathscr{E}$ and $\widetilde{\mathscr{E}}$ be two Hausdorff topological spaces and let $f : \mathscr{E} \to \widetilde{\mathscr{E}}$ be a continuous mapping. Let $(\nu_\varepsilon)_{\varepsilon>0}$, $(\widetilde{\nu}_\varepsilon)_{\varepsilon>0}$ be two families of probability measures on $(\mathscr{E}, \mathscr{B}(\mathscr{E}))$ and $(\widetilde{\mathscr{E}}, \mathscr{B}(\widetilde{\mathscr{E}}))$ respectively, such that $\widetilde{\nu}_\varepsilon = \nu_\varepsilon \circ f^{-1}$ for each $\varepsilon > 0$. If $(\nu_\varepsilon)_{\varepsilon>0}$ satisfies an LDP on $\mathscr{E}$ as ε tends to zero with rate function Λ, then $(\widetilde{\nu}_\varepsilon)_{\varepsilon>0}$ satisfies an LDP on $\widetilde{\mathscr{E}}$ as ε tends to zero with rate function*

$$\widetilde{\Lambda}(y) := \inf\{\Lambda(x) : y = f(x)\} = \inf\{\Lambda(f^{-1}(y))\}.$$

Theorem 1.5.14. *Let B be a d-dimensional Gaussian process, inducing a measure μ on $\left(\mathcal{C}^d, \mathscr{B}(\mathcal{C}^d)\right)$ with RKHS $\mathscr{H}_\mu$. Then $(\varepsilon\mu)_{\varepsilon>0}$ satisfies an LDP as ε tends to zero with speed ε^{-1} and rate function*

$$\Lambda^*_\mu(x) := \begin{cases} \frac{1}{2}\|x\|^2_{\mathscr{H}_\mu}, & \text{if } x \in \mathscr{H}_\mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

Example 1.5.15. Recall $\mathscr{H}_\nu$, the RKHS corresponding to the standard Wiener measure ν on $\mathcal{C}$ induced by a standard Brownian motion, from Example 1.4.9. Applying Theorem 1.5.14 yields that the family of measures $(\varepsilon\nu)_{\varepsilon>0}$ on $\mathcal{C}$ satisfies an LDP as ε tends to zero with speed ε^{-1} and rate function

$$\Lambda^*_\nu(f) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{f}(u)|^2 \mathrm{d}u, & \text{if } f \in \mathscr{H}_\nu, \\ +\infty, & \text{otherwise.} \end{cases}$$

This result is classically referred to as Schilder's Theorem: see, for example, [DZ10, Theorem 5.2.3.].

Before providing the final theorem of this section, [Gar08, Theorem 1.2.], it is first necessary to define a certain class of sequences of stochastic processes.

Definition 1.5.16. [Gar08, Definition 1.1.] Let $\mathscr{U}$ denote the space of simple, real valued, adapted processes Z such that $\sup_{t\geq 0}|Z_t| \leq 1$. A sequence of semi-martingales $\{Y^\varepsilon\}$ is said to be uniformly exponentially tight (UET) if, for every $t > 0$ and $c > 0$, there exists $K_{c,t} > 0$ such that

$$\limsup_{\varepsilon\to 0} \varepsilon \log\left[\sup_{Z\in\mathscr{U}} \mathbb{P}\left(\sup_{s\leq t}\left|\int_0^s Z_{u-}\mathrm{d}Y^\varepsilon_u\right| \geq K_{c,t}\right)\right] \leq \ - \ c.$$

Theorem 1.5.17. ***[Gar08, Theorem 1.2.]*** *Let $\{X^\varepsilon\}$ be a sequence of adapted, càdlàg stochastic processes, and let $\{Y^\varepsilon\}$ be a sequence of uniformly exponentially tight semi-martingales. If $\{(X^\varepsilon, Y^\varepsilon)\}$ satisfies an LDP as ε tends to zero with rate function Λ, then the sequence of processes $\left(\int_0^\cdot X_s^\varepsilon \mathrm{d}Y_s^\varepsilon\right)_{\varepsilon\geq 0}$ satisfies an LDP as ε tends to zero with rate function*

$$\widehat{\Lambda}(\varphi) := \inf\left\{\Lambda(x,y) : \varphi = \int_0^\cdot x(s)\mathrm{d}y(s), y \text{ finite variation}\right\}.$$

1.6 Application of Large Deviations Theory to Implied Volatility Asymptotics

We now focus on an application of large deviations theory to implied volatility asymptotics. We first give a short example that will demonstrate how small-time asymptotics for the implied volatility of a given model can be recovered from an LDP for that given model, and then present other relevant results.

Forde and Jacquier [FJ09] consider the Heston model (1.5), and use the closed form moment generating function of the log stock price process to establish an LDP for the log stock price process, as t tends to zero, using the Gärtner-Ellis Theorem. The LDP has rate function I^*, which is the Fenchel-Legendre transform of the continuous function I that is defined as

$$I(p) := \begin{cases} \frac{v_0 p}{\xi(\sqrt{1-\rho^2}\cot(\xi p\sqrt{1-\rho^2}/2)-p)}, & \text{if } p \in (p_-, p_+), \\ +\infty, & \text{otherwise.} \end{cases}$$

Recall the parameter ρ is the correlation between the Brownian motions driving the stock price process and the volatility process, ξ is the volatility-of-volatility, and $v_0 > 0$. The table below gives the values of p_- and p_+:

ρ	p_+	p_-
< 0	$\frac{2(\pi+\arctan(\sqrt{1-\rho^2}/\rho))}{\xi\sqrt{1-\rho^2}}$	$\frac{2\arctan(\sqrt{1-\rho^2}/\rho)}{\xi\sqrt{1-\rho^2}}$
$= 0$	$\frac{\pi}{\xi}$	$-\frac{\pi}{\xi}$
> 0	$\frac{2\arctan(\sqrt{1-\rho^2}/\rho)}{\xi\sqrt{1-\rho^2}}$	$\frac{2(\pi-\arctan(\sqrt{1-\rho^2}/\rho))}{\xi\sqrt{1-\rho^2}}$

The following Lemma is stated without proof, but can be found in Section 2 [FJ09, Corollary 2.1.] of the paper.

Lemma 1.6.1. $-\lim_{t\to 0} t \log \mathbb{E}\left[(S_t - K)^+\right] = I^*(x)$, *where* $x = \log(K/S_0) \geq 0$.

Using Lemma 1.6.1, we may now state the small-time behaviour of the implied volatility $\widehat{\sigma}(x,t)$:

$$\lim_{t\to 0} \widehat{\sigma}(x,t) = \frac{x}{\sqrt{2I^*(x)}} \quad \textit{for } x \neq 0. \tag{1.19}$$

We briefly sketch why the above limit holds; Lemma 1.6.1 implies that for $\delta > 0$, $\exp(-(I^*(x)+\delta)/t) \leq \mathbb{E}\left[(S_t - K)^+\right]$. Estimates on normally distributed random variables then imply that $\mathbb{E}\left[(S_t - K)^+\right] \leq S_0 C n(d_1)$, where C is a positive constant and d_1 is defined as

$$\frac{-x + \widehat{\sigma}(x,t)^2 t/2}{\widehat{\sigma}(x,t)\sqrt{t}}.$$

Taking logs of both sides and multiplying by t then yields

$$-(I^*(x)+\delta) \leq \frac{-x^2}{2\widehat{\sigma}(x,t)^2} + \delta.$$

Using similar techniques, Forde and Jacquier then establish that

$$-(I^*(x)-\delta) \geq -\frac{(x+\delta)^2}{2\widehat{\sigma}(x,t)^2} - \delta,$$

and from here the result is clear.

There is a wealth of literature that focuses on applying large deviations techniques to stochastic volatility models to draw inferences on the small-time, as well as large-time, implied volatility behaviour. Feng, Forde, and Fouque [FFF10] derive a small-time LDP from the moment generating function of the time changed log stock price process for the Heston model, as the volatility process mean reverts. The small-time implied volatility

behaviour is then deduced. Jacquier, Keller-Ressel, and Mijatović [JKM13] consider a class of affine stochastic volatility models with jumps: the log stock price process and the volatility process are assumed to be a two-dimensional, time-homogeneous Markov process, whose cumulant generating function has a given affine form. This class of models includes the Heston model (with and without state independent jumps), the Bates model with state dependent jumps, and the Barndorf-Nielsen-Shepard model. Jacquier, Keller-Ressel, and Mijatović [JKM13] apply the Gärtner-Ellis Theorem to attain an LDP as t tends to $+\infty$ for the rescaled log stock price process. The LDP then determines the large-time implied volatility behaviour.

More recently, attention has turned to rough volatility models and their implied volatility behaviour. The fractional Heston model (1.12), first proposed by Comte, Coutin, and Renault [CCR12], is studied in [GJR18b]. Guennoun, Jacquier, and Roome prove an LDP for the log stock price process as t tends to zero, and an LDP for the rescaled log stock price process as t tends to $+\infty$, using the Gärtner-Ellis theorem; small and large time implied volatility asymptotics are then deduced. Notably, the small-time smiles are steeper than in classical stochastic volatility models. A small-time and large-time large deviations principle for the rough Heston model [EFR18] has recently been proved by Forde, Gerhold, and Smith [FGS19].

Forde and Zhang [FZ17] consider the rough volatility model (1.13), where the volatility process is defined pathwise as the image of a Hölder continuous function acting on the trajectory of a fractional Brownian motion. An LDP as t tends to zero is attained, and used to describe small-time implied volatility asymptotics. A large-time LDP for a fractional local-stochastic volatility model is also provided.

Bayer, Friz, Gulisashvili, Horvath, and Stemper [BFGHS19] use the large devations results from [FZ17] to derive the small-time behaviour of the implied volatility in the so-called "moderate deviations regime". Horvath, Jacquier, and Lacombe [HJL19] use large deviations techniques to establish the small-time implied volatility behaviour for the generalised Stein-Stein model, where the volatility process has a random start point v_0 and is driven by a fractional Brownian motion; tail estimates are also given.

Remark 1.6.2. Large deviations techniques are, of course, not solely limited to describing asymptotic implied volatility behaviour; see for example the survey [Pha08],

where large deviations for credit risk, optimal long term investment, and rare event simulation are reviewed. Outside of mathematical finance, large deviations theory has a wide variety of applications too: for example in the random conduction of heat, in polymer chains, and in statistical mechanics, [Hol00, Chapter VII, Chapter IX] and [Tou09] respectively.

1.7 Other Asymptotic Methods for Implied Volatility

It is very much worth emphasising at this point that large deviations techniques are not the sole means of analysing the asymptotic behaviour of implied volatility within stochastic volatility models. Indeed, we finish the introduction with a brief overview of some other methods used to derive implied volatility asymptotics.

For simplicity we now assume $S_0 = 1$; we recall the definition $\mathrm{C}(x,t) := \mathbb{E}[(S_t - e^x)^+]$, $x := \log(K)$, and denote X_t to have distribution μ_t, with finite first moment. First note that

$$\mathrm{C}(x,t) = \mathbb{E}\left[e^{X_t}\mathbb{I}_{\{X_t \geq x\}}\right] - e^x \mathbb{P}(X_t \geq x) = \int_x^{+\infty} (e^y - e^x)\mu_t(\mathrm{d}y), \qquad (1.20)$$

$$\mathrm{C_{BS}}(x,t,\widehat{\sigma}(x,t)) = \frac{1}{2}\left[1 - e^x + \mathrm{erf}\left(\frac{d_-(x,t)}{\sqrt{2}}\right) - e^x \mathrm{erf}\left(\frac{d_+(x,t)}{\sqrt{2}}\right)\right],$$

where $d_\pm(x,t) := \frac{-x \pm \frac{\widehat{\sigma}^2(x,t)t}{2}}{\widehat{\sigma}(x,t)\sqrt{t}}$, and $\mathrm{erf}(z) := \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\mathrm{d}t$ denotes the error function. Using that the error function has a Taylor expansion given by $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}(z - \frac{z^3}{3} + o(z^5))$, [AS72, page 932], we see that for (nonzero) x, as t tends to zero, the following expansion for the price of a Call option holds:

$$\begin{aligned}\int_x^{+\infty} (e^y - e^x)\mu_t(\mathrm{d}y) = \frac{1-e^x}{2} + \frac{1}{\sqrt{\pi}}\Bigg[&\frac{\widehat{\sigma}(x,t)t^{1/2}\left(24(1+e^x) + 6x(1-e^x) - \widehat{\sigma}(x,t)^2 t(1+e^x)\right)}{24\sqrt{8}} \\ &- \frac{x\left(4(1-e^x) + x(1+e^x)\right)}{4\sqrt{2}\widehat{\sigma}(x,t)t^{1/2}} + \frac{x^3(1-e^x)}{6\sqrt{2}\widehat{\sigma}(x,t)^3 t^{3/2}} + o(\widehat{\sigma}(x,t)^5 t^{5/2})\Bigg] \\ &= \mathrm{C_{BS}}(x,t,\widehat{\sigma}(x,t)).\end{aligned}$$

This representation allows us to characterise the small-time asymptotic behaviour of the implied volatility $\widehat{\sigma}(\cdot,\cdot)$.

Remark 1.7.1. Clearly the at-the-money case, where $x = 0$, the above expansion may be simplified further as t tends to zero:

$$\int_0^{+\infty} (e^y - 1)\mu_t(\mathrm{d}y) = \frac{1}{\sqrt{2\pi}} \left[\widehat{\sigma}(0,t)t^{1/2} - \frac{\widehat{\sigma}(0,t)^3 t^{3/2}}{24} + o(\widehat{\sigma}(0,t)^5 t^{5/2}) \right] = \mathrm{C}_{\mathrm{BS}}(0,t,\widehat{\sigma}(0,t)).$$

Gulisashvili, Viens, and Zhang [GVZ18] consider a general class of self-similar Gaussian stochastic volatility models; an estimate on the small-time stock price density is given using the Karhunen-Loève expansion coefficients of the volatility process, together with the self-similarity parameter. From here, well-known small-time relations between Call prices and implied volatilities are used to draw conclusions on the implied volatility behaviour. Paulot [Pau15] uses a heat kernel expansion to provide a general framework for the computation of an exact Taylor expansion for the implied volatility of a stochastic volatility model, which is then applied specifically to the SABR model. Using PDE techniques and Taylor expansions, Medvedev and Scaillet [MS07] provide close to maturity implied volatility behaviour for a stochastic volatility model with jumps.

Chapter 2

Small-time Asymptotics for Selected Rough Volatility Models

In this Chapter, we turn our attention to the asymptotic behaviour of the rough Bergomi model. Our aim is to precisely characterise the small-time behaviour of the rough Bergomi implied volatility. We begin, in Section 2.1 by proving a large deviations principle for a rescaled version of the log stock price process, which then allows us to give the small-time limit of the implied volatility, under some mild assumptions. This is the main result of the Chapter[1]. We are able to use the same methodology, in fact, to establish the small-time implied volatility behaviour of the lognormal fSABR model; this is done briefly in Section 2.2.

2.1 An LDP for the Rough Bergomi Model

Let us start by defining the process X as $X_t := \log\left(\frac{S_t}{S_0}\right)$, where the stock price process S is the rough Bergomi model, defined in (1.11). Recall the process Z, defined in (1.9), and the (simplified) process v, defined in (1.11), with corresponding Brownian motion B also defined in (1.11). We fix $\mathcal{T} = [0, 1]$ for the rest of this Chapter, although our results can be easily adapted for the general interval $[0, T]$.

[1]This Chapter includes results from article [JPS18].

We now define the rescaled versions of the processes X, Z, v, B, as follows for $t, \varepsilon \geq 0$:

$$\begin{aligned} X_t^\varepsilon &:= \varepsilon^\beta X_{\varepsilon t}, \\ Z_t^\varepsilon &:= \varepsilon^{\beta/2} Z_t, \\ v_t^\varepsilon &:= \varepsilon^{1+\beta} v_0 \exp\left(Z_t^\varepsilon - \frac{\eta^2}{2} (\varepsilon t)^\beta \right), \\ B_t^\varepsilon &:= \varepsilon^{\beta/2} B_t, \end{aligned} \tag{2.1}$$

where $\beta := 2\alpha + 1$ takes values in $(0, 1)$. Note that $Z_t^\varepsilon = \varepsilon^{\alpha+1/2} Z_t$, so Z_t^ε and $Z_{\varepsilon t}$ are equal in law. This implies that the rescaled volatility process $v_\cdot^\varepsilon$ and the process $\varepsilon^{1+\beta} v_{\varepsilon \cdot}$ are also equal in law. In turn, this implies that the following representation of $X_\cdot^\varepsilon$ holds as a process:

$$\begin{aligned} X_\cdot^\varepsilon &:= \varepsilon^\beta X_{\varepsilon \cdot} \\ &\overset{\mathrm{d}}{=} \varepsilon^\beta \left(\int_0^{\varepsilon \cdot} \sqrt{v_s} \mathrm{d}B_s - \frac{1}{2} \int_0^{\varepsilon \cdot} v_s \mathrm{d}s \right) \\ &\overset{\mathrm{d}}{=} \varepsilon^\beta \left(\int_0^{\cdot} \sqrt{v_{\varepsilon s}} \mathrm{d}B_{\varepsilon s} - \frac{1}{2} \int_0^{\cdot} v_{\varepsilon s} \varepsilon \mathrm{d}s \right) \\ &\overset{\mathrm{d}}{=} \int_0^{\cdot} \sqrt{\varepsilon^{1+2\beta} v_{\varepsilon s}} \mathrm{d}B_s - \frac{1}{2} \int_0^{\cdot} \varepsilon^{1+\beta} v_{\varepsilon s} \mathrm{d}s \\ &\overset{\mathrm{d}}{=} \int_0^{\cdot} \sqrt{v_s^\varepsilon} \mathrm{d}B_s^\varepsilon - \frac{1}{2} \int_0^{\cdot} v_s^\varepsilon \mathrm{d}s. \end{aligned}$$

Remark 2.1.1. Following [GJR18a], Bayer, Friz, and Gatheral [BFG16] suggest taking $\alpha \approx -0.4$. This corresponds to $\beta \approx 0.2$. The findings of [GJR18a] are corroborated by [BLP17a], and further validated by our own findings in Chapter 4.

We now state the main result of this section: the rescaled process $(X^\varepsilon)_{\varepsilon \in \mathcal{T}}$ satisfies an LDP. We first define the operator $\mathrm{M} : \mathcal{C}^2 \to \mathcal{C}(\mathcal{T}^2, \mathbb{R}_+ \times \mathbb{R})$ as

$$(\mathrm{M}z_y^x)(t, \varepsilon) := \begin{pmatrix} (\mathfrak{m}x)(t, \varepsilon) \\ y(t) \end{pmatrix}, \quad \text{for all } t, \varepsilon \in \mathcal{T}, \tag{2.2}$$

where the operator $\mathfrak{m} : \mathcal{C} \to \mathcal{C}$ is defined by

$$(\mathfrak{m}x)(t, \varepsilon) := v_0 \varepsilon^{1+\beta} \exp\left(x(t) - \frac{\eta^2}{2} (\varepsilon t)^\beta \right), \tag{2.3}$$

and the function $\Lambda(\mathrm{z}_{y_1}^{x_1}) := \inf\left\{\Lambda^*(\mathrm{z}_{y_2}^{x_2}) : \mathrm{z}_{y_1}^{x_1} = \mathrm{M}(\mathrm{z}_{y_2}^{x_2})\right\}$ for $\mathrm{z}_{y_1}^{x_1} \in \mathcal{C}(\mathcal{T}^2, \mathbb{R}_+ \times \mathbb{R})$, where $\Lambda^*(\mathrm{z}_y^x) := \frac{1}{2}\|\mathrm{z}_y^x\|^2_{\mathscr{H}_\rho^{K_\alpha}}$ and $\mathscr{H}_\rho^{K_\alpha}$ is defined in Proposition 1.4.18. Recall z_y^x denotes the two-dimensional path $(x, y)^\top \in \mathcal{C}^2$; see **Notations** section prior to Chapter 1.

Theorem 2.1.2. *The sequence $(X^\varepsilon)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP on $\mathcal{C}$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function $\Lambda^X : \mathcal{C} \to [0, +\infty]$ defined as*

$$\Lambda^X(\varphi) := \inf\left\{\Lambda(\mathrm{z}_y^x) : \varphi = \sqrt{x} \cdot y, y \in \mathrm{BV} \cap \mathcal{C}\right\}.$$

Remark 2.1.3. Taking Proposition 1.4.17 into consideration, it is straightforward to extend the proof of Theorem 2.1.2 to the case where the process Z is defined more generally as the stochastic integral of some deterministic kernel on $\mathbb{R}_+ \times \mathbb{R}_+$. Some additional conditions, namely self-similarity, would be required on the generalised process, however, in order for the log stock price process X itself to satisfy an LDP.

We now prove Theorem 2.1.2.

Proof. Let $((Z_t, B_t))_{t\in\mathcal{T}}$ induce the measure μ on $\mathcal{C}^2$. It follows immediately from Theorem 1.5.14 and Proposition 1.4.18 that the family of measures $(\varepsilon^\beta\mu)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP as ε tends to zero with speed $\varepsilon^{-\beta}$ and rate function

$$\Lambda^*(\mathrm{z}_y^x) := \begin{cases} \frac{1}{2}\|\mathrm{z}_y^x\|^2_{\mathscr{H}_\rho^{K_\alpha}}, & \text{if } \mathrm{z}_y^x \in \mathscr{H}_\rho^{K_\alpha}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Recall that $\mathscr{H}_\rho^{K_\alpha}$ is given in Proposition 1.4.18. Note that scaling μ by ε^β is equivalent to scaling (Z, B) by $\varepsilon^{\beta/2}$. We denote this rescaled two-dimensional process as $(Z^\varepsilon, B^\varepsilon)$ as in (2.1). Pathwise, we may view $t \mapsto (Z_t^\varepsilon, B_t^\varepsilon)$ as an element of $\mathcal{C}^2$ and with that in mind, we may express $(v_t^\varepsilon, B_t^\varepsilon)$ as $\mathrm{M}(Z^\varepsilon, B^\varepsilon)(t, \varepsilon)$. Note that the choice of ε as the second argument of (t, ε) in $\mathrm{M}(Z^\varepsilon, B^\varepsilon)(t, \varepsilon)$ allows control the index and scaling of $(v_t^\varepsilon, B_t^\varepsilon)$.

First let us verify that M is indeed a continuous operator, with respect to the $\mathcal{C}(\mathcal{T}^2, \mathbb{R}_+ \times \mathbb{R})$ norm $\|\cdot\|_\infty$.

For any $(f, g)^\top \in \mathcal{C}^2$, introduce a small perturbation $(\delta^f, \delta^g) \in \mathcal{C}^2$. Then, recalling

$\mathcal{T} = [0,1]$,

$$\begin{aligned}\left\|\mathrm{M}\begin{pmatrix} f+\delta^f \\ g+\delta^g \end{pmatrix} - \mathrm{M}\begin{pmatrix} f \\ g \end{pmatrix}\right\|_\infty &= \sup_{t,\varepsilon\in\mathcal{T}} \left\{ \left| (\mathfrak{m}(f+\delta^f))(t,\varepsilon) - (\mathfrak{m}f)(t,\varepsilon) \right| + |\delta^g(t)| \right\} \\ &\leq \sup_{t,\varepsilon\in\mathcal{T}} \left\{ v_0 \varepsilon^{1+\beta} \exp\left(-\frac{\eta^2}{2}(\varepsilon t)^\beta\right) \left| \mathrm{e}^{f(t)} \right| \left| \mathrm{e}^{\delta^f(t)} - 1 \right| \right\} + \sup_{t\in\mathcal{T}} |\delta^g(t)| \\ &\leq C \sup_{t\in\mathcal{T}} \left| \mathrm{e}^{\delta^f(t)} - 1 \right| + \sup_{t\in\mathcal{T}} |\delta^g(t)|,\end{aligned}$$

for some strictly positive constant C. The right-hand side clearly tends to zero as (δ^f, δ^g) tends to zero with respect to $\|\cdot\|_\infty$, and hence M is a continuous operator.

The Contraction Principle (Proposition 1.5.13) therefore implies that the sequence $(v^\varepsilon, B^\varepsilon)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP on $\mathcal{C}(\mathcal{T}^2, \mathbb{R}_+ \times \mathbb{R})$, with speed $\varepsilon^{-\beta}$ and rate function Λ. Since M is clearly a bijection, the rate function Λ may then be expressed as $\Lambda(\mathrm{z}_{y_1}^{x_1}) = \Lambda^*\left(\mathrm{M}^{-1}(\mathrm{z}_{y_1}^{x_1})\right)$, for any $(x_1, y_1) \in \mathcal{C}^2$.

We now use Theorem 1.5.17, which tells us that the sequence of processes $\left(\int_0^\cdot \sqrt{v_s^\varepsilon}\mathrm{d}B_s^\varepsilon\right)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP as ε tends to zero. For convenience we adopt the following notation: $I(v^\varepsilon, B^\varepsilon)(t) = \int_0^t \sqrt{v_s^\varepsilon}\mathrm{d}B_s^\varepsilon$. Note that $I(v^\varepsilon, B^\varepsilon)(t) = \int_0^t \sqrt{\varepsilon^{2\alpha} v_s^\varepsilon}\sqrt{\varepsilon}\mathrm{d}B_s$ and [Gar08, Example 2.1.] demonstrates that the sequence of (semi)-martingales $\{\sqrt{\varepsilon}B_\cdot\}$ is UET in the sense of Definition 1.5.16. Furthermore, the sequence of processes $\{\sqrt{\varepsilon^{2\alpha} v_\cdot^\varepsilon}\}$ is càdlàg, and adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$. Thus, having satisfied all the requirements of Theorem 1.5.17, we conclude that the sequence of processes $I(v^\varepsilon, B^\varepsilon)$ satisfies an LDP as ε tends to zero with speed $\varepsilon^{-\beta}$ and rate function

$$\Lambda^X(\varphi) = \inf\{\Lambda(x,y) : \varphi = I(x,y),\ y \in \mathrm{BV} \cap \mathcal{C}\}.$$

Our final aim is to prove an LDP for the sequence of processes $(X^\varepsilon)_{\varepsilon\in\mathcal{T}}$, where we have that $X_t^\varepsilon = \int_0^t \sqrt{v_s^\varepsilon}\mathrm{d}B_s^\varepsilon - \frac{1}{2}\int_0^t v_s^\varepsilon \mathrm{d}s$ in law, so all that is left for us to negotiate is the drift term. The LDP for the sequence $(X^\varepsilon)_{\varepsilon\in\mathcal{T}}$ is given as ε tends to zero, so without loss of generality we may indeed set $t = 1$.

To complete the proof we show that the sequences $(X_1^\varepsilon)_{\varepsilon\in\mathcal{T}}$ and $(I(v^\varepsilon, B^\varepsilon)(1))_{\varepsilon\in\mathcal{T}}$ are exponentially equivalent. For any $\delta > 0$ it follows that

$$\mathbb{P}\left(|X_1^\varepsilon - I(v^\varepsilon, B^\varepsilon)(1)| > \delta\right) \leq \mathbb{P}\left(\int_0^1 v_s^\varepsilon \mathrm{d}s > \delta\right) \leq \mathbb{P}\left(\int_0^1 \exp(Z_s^\varepsilon)\mathrm{d}s > b_\varepsilon\right),$$

where $b_\varepsilon := \delta/v_0\varepsilon^{1+\beta}$. Note that $Z^\varepsilon_s - \frac{\eta^2}{2}(\varepsilon s)^\beta \leq Z^\varepsilon_s$ for all $s \in [0,1]$ and $\varepsilon \in \mathcal{T}$, from which the final inequality follows.

Using that $\int_0^1 \exp(Z^\varepsilon_s)\mathrm{d}s \leq \exp(\sup_{t\in[0,1]} Z^\varepsilon_t)$ almost surely, it follows that

$$\mathbb{P}\left(\int_0^1 \exp(Z^\varepsilon_s)\mathrm{d}s > b_\varepsilon\right) \leq \mathbb{P}\left(\sup_{t\in[0,1]} Z^\varepsilon_t > \log b_\varepsilon\right) = \mathbb{P}\left(\sup_{t\in[0,1]} Z_t > \frac{\log b_\varepsilon}{\varepsilon^{\beta/2}}\right).$$

The process $(Z_t)_{t\in[0,1]}$ is almost surely bounded [AT07, Theorem 1.5.4], and so we may apply the Borell-TIS inequality; a consequence of which [AT07, Theorem 2.1.1 and discussion below] implies that

$$\mathbb{P}\left(\sup_{t\in[0,1]} Z_t > \frac{\log b_\varepsilon}{\varepsilon^{\beta/2}}\right) \leq \exp\left(-\frac{1}{2}\left(\frac{\log b_\varepsilon}{\varepsilon^{\beta/2}} - \mathbb{E}\left[\sup_{t\in[0,1]} Z_t\right]\right)^2\right).$$

This then implies that

$$\begin{aligned}\varepsilon^\beta \log \mathbb{P}\left(\int_0^1 \exp(Z_s)\mathrm{d}s > b_\varepsilon\right) \\ \leq \varepsilon^\beta\left(-\frac{(\log b_\varepsilon)^2}{2\varepsilon^\beta} + \frac{\log b_\varepsilon}{\varepsilon^{\beta/2}}\mathbb{E}\left[\sup_{t\in[0,1]} Z_t\right] - \frac{1}{2}\mathbb{E}\left[\sup_{t\in[0,1]} Z_t\right]^2\right).\end{aligned}$$

Note that $\varepsilon^{\beta/2}\log b_\varepsilon$ converges to zero as ε tends to zero, which in turn implies that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^{\beta/2}\log b_\varepsilon \mathbb{E}\left[\sup_{t\in[0,1]} Z_t\right] = 0.$$

Similarly, $\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta \mathbb{E}\left[\sup_{t\in[0,1]} Z_t\right]^2 = 0$. Furthermore, it follows that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta\left(-\frac{(\log b_\varepsilon)^2}{2\varepsilon^\beta}\right) = -\infty,$$

and

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta \log \mathbb{P}\left(\sup_{t\in[0,1]} |X^\varepsilon_t - \mathbb{I}(v^\varepsilon, B^\varepsilon)(t)| > \delta\right) = -\infty.$$

Therefore $\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta \log \mathbb{P}\left(|X^\varepsilon_1 - I(v^\varepsilon, B^\varepsilon)(1)| > \delta\right) = -\infty$, which is precisely the definition of exponential equivalence [DZ10, Definition 4.2.10]. Then, by [DZ10, Theorem

4.2.13], the sequence $(X_1^\varepsilon)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP with speed $\varepsilon^{-\beta}$ and rate function Λ^X. Clearly, since 1 was chosen arbitrarily, the result for $(X^\varepsilon)_{\varepsilon\in\mathcal{T}}$ also holds.

□

Remark 2.1.4. Note that there is a degree of flexibility when choosing how to define the rescaled process $(X^\varepsilon)_{\varepsilon\in\mathcal{T}}$. For example, we may define $X_t^\varepsilon := \varepsilon^\alpha X_{\varepsilon^\gamma t}$ where $\gamma := \frac{\alpha}{\alpha/2+5/4}$. In this case we define the rescaled process $(Z^\varepsilon, B^\varepsilon)$ as $\varepsilon^{\gamma(\alpha+1/2)}(Z,B)$, and the rescaled volatility process v_t^ε as $\varepsilon^{\alpha+\gamma} v_{\varepsilon^\gamma t}$. In this case X^ε satisfies an LDP with speed $\varepsilon^{-(2\gamma(\alpha+1/2))}$ and a rate function similar to Theorem 2.1.2. This essentially falls in the category of moderate deviations, within the context of [Gui03], for the original process X; X is scaled by $1/\sqrt{t}h(t)$ where $h(t) \in [1, 1/\sqrt{t}]$ for small enough t.

Corollary 2.1.5. *The rescaled log stock price process* $\left(t^\beta X_t\right)_{t\in\mathcal{T}}$ *satisfies an LDP on* $\mathbb{R}$ *as* t *tends to zero with speed* $t^{-\beta}$ *and rate function* $\Lambda_1^X(u) := \inf\{\Lambda^X(\varphi) : \varphi(1) = u\}$, $u \in \mathbb{R}$.

Proof. Since X_1^ε and $\varepsilon^\beta X_\varepsilon$ are equal in law, $(\varepsilon^\beta X_\varepsilon)_{\varepsilon\in\mathcal{T}}$ satisfies an LDP with speed $\varepsilon^{-\beta}$ and rate function Λ^X by Theorem 2.1.2; mapping ε to t completes the proof.

□

Remark 2.1.6. Recall that Forde and Zhang [FZ17] derived pathwise large deviations for rough volatility models, with application (by scaling) to small-time asymptotics of the corresponding implied volatility. The model they consider is of the following form, for the log stock price process:

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}\sigma(Y_t)^2\mathrm{d}t + \sigma(Y_t)\mathrm{d}B_t, \\ Y_t = W_t^H, \end{cases}$$

where B is a standard Brownian motion, W^H a (possibly correlated) fractional Brownian motion. In order to prove an LDP, they consider a small-noise version of the SDE above, namely:

$$\begin{cases} \mathrm{d}X_t^\varepsilon = -\frac{1}{2}\varepsilon\sigma(Y_t)^2\mathrm{d}t + \sqrt{\varepsilon}\sigma(Y_t)\mathrm{d}B_t, \\ Y_t^\varepsilon = \varepsilon^H W_t^H. \end{cases}$$

It is of course tempting to apply their results to the rough Bergomi model. Unfortunately, the following intricacies make this impossible: firstly, they assume the function σ to have at most linear growth, whereas it is of exponential growth in rough Bergomi; secondly, their scaling assumption, allowing them to translate small-noise into small-time estimates crucially relies on the volatility process Y being driftless [FZ17, Equation (4.4)], which does not hold in rough Bergomi.

Remark 2.1.7. The RKHS structure of $\mathscr{H}_\rho^{K_\alpha}$ precisely determines the rate function Λ^X. If we were to consider the two-dimensional process (Z, B), where Z is defined in (1.9) and B is an independent Brownian motion such that $\langle Z, B\rangle_t = 0$ for all t, the corresponding RKHS would have a different structure to $\mathscr{H}_\rho^{K_\alpha}$ and thus a different inner product. The corresponding rate function, as defined in Proposition 1.5.14, would therefore be different too. Indeed, we refer the reader to Appendix A for further discussion on the matter.

Remark 2.1.8. Recall that every $\mathrm{z}_y^x \in \mathscr{H}_\rho^{K_\alpha}$ has the integral representation $(x, y) = \mathcal{I}(f_1, f_2)$, where $(f_1, f_2) \in \mathrm{L}^2 \times \mathrm{L}^2$ and the operator $\mathcal{I} = \mathcal{I}_\rho^{K_\alpha}$ as in Proposition 1.4.18. With that in mind, we may reformulate the rate function given in Theorem 2.1.2:

$$\Lambda^X(\varphi) = \inf_{\substack{\varphi = I(\mathrm{M}(\mathcal{I}(f_1,f_2)))\\ f_1,f_2\in \mathrm{L}^2}} \left\{ \frac{1}{2}\left(\int_0^1 f_1^2(u)\mathrm{d}u + \bar{\rho}\int_0^1 f_2^2(u)\mathrm{d}u \right) \right\}, \tag{2.4}$$

where the notation I and M are consistent with that in the proof of Theorem 2.1.2.

2.1.1 Implied Volatility Asymptotics

Let $\widehat{\sigma}$ denote the implied volatility, that is, for a given log-moneyness $x \in \mathbb{R}$ and maturity $t \geq 0$, the unique non-negative solution to the equation $\mathrm{C}_{\mathrm{BS}}(x, t, \widehat{\sigma}(x, t)) = \mathrm{C}(x, t)$, where C_{BS} denotes the Black-Scholes price of a vanilla Call price, and C the corresponding Call price with log-strike x and maturity t in a given (here the rough Bergomi) model.

We aim to investigate the implied volatility of the rough Bergomi model, and in order to do so we must impose the following (reasonable) assumptions, which firstly ensure that implied volatility is well defined for the rough Bergomi model and secondly allow

us to deduce the asymptotic implied volatility behaviour from the small-time log stock price behaviour.

Assumption 2.1.9. Assume the rough Bergomi model, which is a local martingale, satisfies the following:

1. the rough Bergomi model is a true martingale;
2. $xt^{-\beta}/|\log \mathbb{E}(S_t - e^{xt^{-\beta}})^+|$ tends to zero as t tends to zero.

Remark 2.1.10. Gassiat [Gas19] recently proved that the rough Bergomi model (under certain correlation regimes) generates true martingales for the spot process.

The second assumption in Assumption 2.1.9 allows us to apply [GL14, Corollary 7.1], to translate the asymptotic behaviour of the log stock price in Corollary 2.1.5 into small-time behaviour of the implied volatility, as follows:

Corollary 2.1.11. *The following holds for all $x \neq 0$:*

$$\lim_{t \downarrow 0} t^{1+\beta} \widehat{\sigma} \left(x t^{-\beta}, t \right)^2 = \begin{cases} \dfrac{x^2}{2 \inf_{y \geq x} \Lambda_1^X(y)}, & \text{if } x > 0, \\ \dfrac{x^2}{2 \inf_{y \leq x} \Lambda_1^X(y)}, & \text{if } x < 0. \end{cases} \tag{2.5}$$

2.2 The Lognormal fSABR Model

We briefly divert from the rough Bergomi model to discuss the lognormal fSABR model of Akahori, Song, and Wang [ASW17]. The lognormal fSABR is an extension of the classical SABR model (1.4), where $\beta = 1$ and the driving noise for the volatility process is a fractional Brownian motion. The fSABR model, with normalised log stock price process X, has the following dynamics:

$$\begin{aligned} X_t &= \int_0^t \sqrt{v_s} \mathrm{d}B_s - \frac{1}{2} \int_0^t v_s \mathrm{d}s, \quad X_0 = 0, \\ v_t &= v_0 \exp(\alpha B_t^H), \quad v_0 > 0. \end{aligned} \tag{2.6}$$

The process B is a standard Brownian motion, and the process B^H is a fractional Brownian motion with Hurst parameter H. The volatility of volatility parameter α is

chosen such that $\alpha \geq 0$. We chose to represent the fractional Brownian motion in terms of the Golosov-Molchan [GM69] kernel $\mathcal{K}_H$,

$$B_t^H = \int_0^t \mathcal{K}_H(s,t)\mathrm{d}W_s \quad \text{for all } t \geq 0,$$
$$\mathcal{K}_H(s,t) := \sqrt{\frac{2H\Gamma(H+1/2)\Gamma(3/2-H)}{\Gamma(2-2H)}}\frac{(t-s)^{H-1/2}}{\Gamma(H+1/2)}{}_2F_1(1/2-H, H-1/2, H+1/2, \frac{s-t}{s}),$$

where W is a standard Brownian motion such that $\langle W., B. \rangle_t = \rho t$.

Comparing the fSABR and rough Bergomi models, it is clear that we can also establish a large deviations principle for the fSABR model by applying the same methodology as the rough Bergomi model. Let us now then define the rescaled processes as follows:

$$X_t^\varepsilon := \varepsilon^{2H} X_{\varepsilon t}, \quad v_t^\varepsilon = \varepsilon^{1+2H} v_{\varepsilon t}, \quad B_t^\varepsilon := \varepsilon^H B_t.$$

By definition, it follows that the below representation of X^ε holds:

$$X_\cdot^\varepsilon \overset{d}{=} \int_0^\cdot \sqrt{v_s^\varepsilon}\mathrm{d}B_s^\varepsilon - \frac{1}{2}\int_0^\cdot v_s^\varepsilon \mathrm{d}s.$$

Finally, let us define the rate function $\Lambda : \mathcal{C} \to [0, +\infty]$ as

$$\Lambda(\varphi) := \inf_{f_1, f_2 \in \mathrm{L}^2} \Big\{ \frac{1}{2}\left(||f_1||_{\mathrm{L}^2}^2 + \bar{\rho}||f_2||_{\mathrm{L}^2}^2 \right) : \varphi = v_0 \int_0^\cdot \exp\left(\frac{\alpha \mathscr{K}_H f_1(s)}{2}\right)\left[\rho f_1(s) + \bar{\rho} f_2(s) - \frac{1}{2}\exp\left(\frac{\alpha \mathscr{K}_H f_1(s)}{2}\right)\right]\mathrm{d}s \Big\},$$

where $\mathscr{K}_H f(s) := \int_0^s \mathcal{K}_H(u,s) f(u)\mathrm{d}u$. The following theorem tells us about the small ε behaviour of the sequence of processes $(X^\varepsilon)_{\varepsilon \in \mathcal{T}}$; the proof follows that of Theorem 2.1.2, albeit with a different RKHS, and is omitted for brevity.

Theorem 2.2.1. *The sequence $(X^\varepsilon)_{\varepsilon \in \mathcal{T}}$ satisfies an LDP on $\mathcal{C}$ with speed ε^{-H} and rate function Λ.*

Corollary 2.2.2. *The rescaled log stock price process $(t^{2H}X_t)_{t \in \mathcal{T}}$ satisfies an LDP on $\mathbb{R}$ with speed t^{-H} and rate function Λ_1, which is defined analogously to Λ_1^X in Corollary 2.1.5.*

Remark 2.2.3. For small vol-of-vol α, we have the following representation of the rate function Λ:

$$\Lambda(\varphi) := \inf_{f_1, f_2 \in \mathrm{L}^2} \Big\{ \frac{1}{2} \left(||f_1||^2_{\mathrm{L}^2} + \bar{\rho} ||f_2||^2_{\mathrm{L}^2} \right) : \\ \varphi = v_0 \int_0^{\cdot} \left[\frac{\alpha}{2} \mathscr{K} f_1(s) (\rho f_1(s) + \bar{\rho} f_2(s) - 1) + \rho f_1(s) + \bar{\rho} f_2(s) - \frac{1}{2} \right] \mathrm{d}s \Big\}.$$

As with the rough Bergomi model, we may translate the small-time behaviour of the rescaled log stock price $(t^{2H} X_t)_{t \in \mathcal{T}}$ into the small-time implied volatility behaviour for the fSABR model, which we present in the Corollary below. Again, we assume that Assumption 2.1.9 holds for the fSABR model, with β replaced by H.

Corollary 2.2.4. *The following holds for all $x \neq 0$:*

$$\lim_{t \downarrow 0} t^{1+3H} \widehat{\sigma} \left(x t^{-2H}, t \right)^2 = \begin{cases} \dfrac{x^2}{2 \inf_{y \geq x} \Lambda_1(y)}, & \text{if } x > 0, \\ \dfrac{x^2}{2 \inf_{y \leq x} \Lambda_1(y)}, & \text{if } x < 0. \end{cases} \tag{2.7}$$

Chapter 3

Asymptotics for Volatility Derivatives in Multi-factor Rough Volatility Models

Having established the small-time implied volatility behaviour for the rough Bergomi and fSABR models in Chapter 2, we now shift our focus from options on spot prices to options on integrated variance. Indeed, the main aim of this Chapter is to derive the small-time behaviour of the integrated 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 integrated variance[1].

Perhaps, options on volatility itself are the most natural object to first analyse within the class of rough volatility models. In this direction, Jacquier, Martini, and Muguruza [JMM18] provide algorithms for pricing VIX options and futures. Horvath, Jacquier and Tankov [HJT18] 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 for rough volatility) has not been studied until very recently in Alòs, García-Lorite and Muguruza [AGM18].

[1]This Chapter includes results from article [LMS19].

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 [CGMY05], and Carr and Lee [CL09]. Keller-Ressel and Muhle-Karbe [KM13] study the small-time asymptotic behaviour of options on discretely sampled realised variance, where the underlying is an exponential Lévy model; we study the continuous time counterpart for rough volatility. Keller-Ressel and Muhle-Karbe find the jump distribution has a strong impact on the option's small-time behaviour; the authors also capitalise on the exponential Lévy form of the underlying to propose exact pricing methods for realised variance via the Laplace transform.

In spite of most of the literature agreeing on the fact that more than a single factor is needed to model volatility (see Bergomi's [Ber16] two-factor model, Avellaneda and Papanicolaou [AP18] or Horvath, Jacquier and Tankov [HJT18] for instance), there is no in-depth analysis on how to construct these (correlated) factors, nor the effect of correlation on the price of volatility derivatives and their corresponding implied volatility smiles. Our aim is to understand multi-factor models and analyse the effect of factors in implied volatility smiles. This Chapter, to the best of our knowledge, is the first piece of research to address such questions, which are of great interest to practitioners in the quantitative finance industry; it is also the first to provide a rigorous mathematical analysis of the small-time behaviour of options on integrated variance in rough volatility models.

For a log stock price process X defined as $X_t = \int_0^t \sqrt{v_s}\mathrm{d}B_s - \frac{1}{2}\int_0^t v_s \mathrm{d}s, X_0 = 0$, where B is standard Brownian motion, we denote the quadratic variation of X at time t by $\langle X\rangle_t$. Then, the core object to analyse in this setting is the integrated variance option with payoff

$$\left(\frac{1}{T}\int_0^T \mathrm{d}\langle X\rangle_s - K\right)^+, \tag{3.1}$$

which in turn defines the risk neutral density of the integrated variance. In this work, we analyse the small-time behaviour of the implied volatility given by (3.1) for a num-

ber 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 integrated variance. We find our results consistent with those of Alòs, García-Lorite and Muguruza [AGM18], which also help us characterise in close-form the implied volatility around the money. 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 integrated 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 3.1 below for evidence of their linearity. The data also indicates both a power-law term structure ATM and its skew.

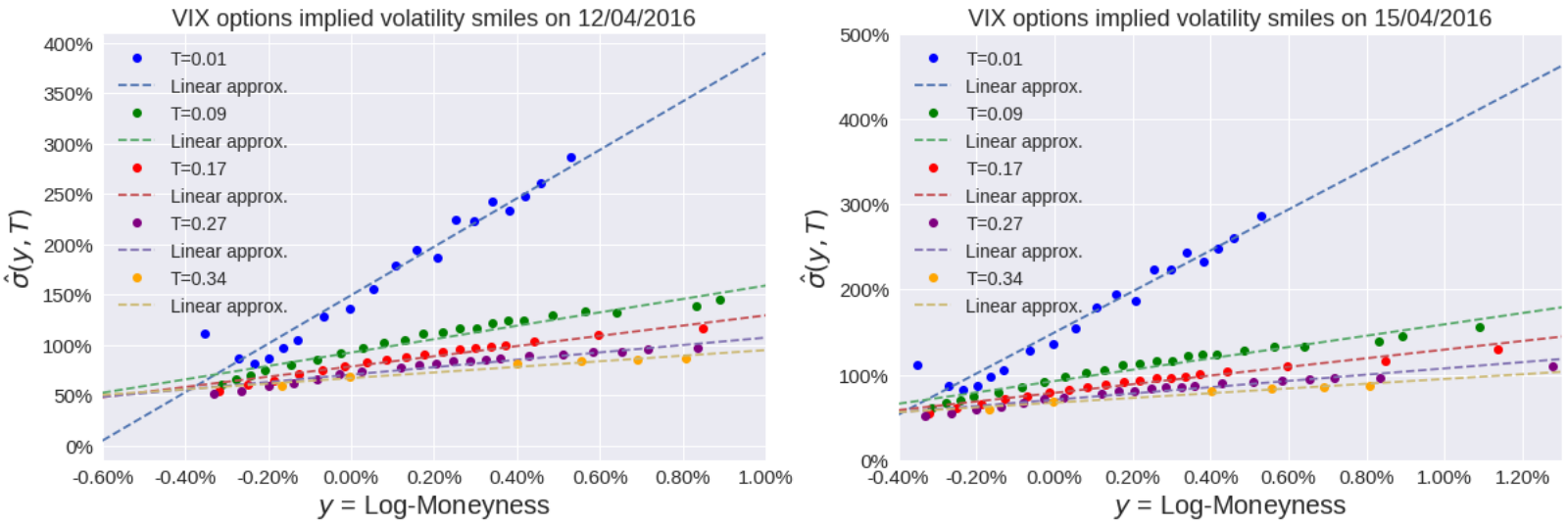

Figure 3.1: Implied volatility smiles for Call options on VIX for small maturities, close to the money. Data provided by OptionMetrics.

The structure of the Chapter is as follows. Section 3.1 introduces the models, two extensions of the rough Bergomi model, whose small-time integrated variance behaviour we study; the main results are given in Section 3.2. In Section 3.3 we provide numerical

schemes to compute the rate functions given in Section 3.2; numerical examples, which include implied volatility smiles for integrated variance, are then given in Section 3.4. Motivated by the numerical examples in Section 3.4, we propose a simple and very feasible approximation for the density of the integrated variance for the mixed rough Bergomi model (see (3.3)) in Section 3.5. We finish the Chapter with Section 3.6, where we introduce a general variance process, which includes the rough Bergomi model for a specific choice of kernel, and briefly investigate the small-noise behaviour of VIX options in this general setting.

3.1 A Showcase of Rough Volatility Models

Let us begin this Section by recalling the process Z from the rough Bergomi model, defined in (1.9). We define an analogous multi-dimensional version of (1.9) by

$$\mathcal{Z}_t := \left(\int_0^t K_\alpha(s,t) \mathrm{d}W_s^1, ..., \int_0^t K_\alpha(s,t) \mathrm{d}W_s^m \right) := \left(\mathcal{Z}_t^1, ..., \mathcal{Z}_t^m \right), \qquad \text{for any } t \in \mathcal{T}, \tag{3.2}$$

where $W^1, ..., W^m$ are independent Brownian motions. We now introduce the two extensions of the rough Bergomi model, used in the forthcoming computations.

Model 3.1.1 (Mixed rough Bergomi)**.** The mixed rough Bergomi model is given in terms of log stock price process X and instantaneous variance process $v^{(\gamma,\nu)}$ as

$$\begin{aligned} X_t &= \int_0^t \sqrt{v_s^{(\gamma,\nu)}} \mathrm{d}B_s - \frac{1}{2} \int_0^t v_s^{(\gamma,\nu)} \mathrm{d}s, \qquad && X_0 = 0, \\ v_t^{(\gamma,\nu)} &= v_0 \textstyle\sum_{i=1}^n \gamma_i \exp\left(\frac{\nu_i}{\eta} Z_t - \frac{\nu_i^2}{2} t^{2\alpha+1} \right), \qquad && v_0 > 0 \end{aligned} \tag{3.3}$$

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

The above modification of the rough Bergomi model, inspired by Bergomi [Ber08], allows a bigger slope (hence bigger skew) on the implied volatility of variance/volatility options to be created, whilst maintaining a tractable instantaneous variance form. This will be made precise in Section 3.4.2.

Model 3.1.2 (Mixed multi-factor rough Bergomi)**.** The mixed rough Bergomi model is given in terms of log stock price process X and instantaneous variance process $v^{(\gamma,\nu,\Sigma)}$ as

$$\begin{aligned} X_t &= \int_0^t \sqrt{v_s^{(\gamma,\nu,\Sigma)}} \mathrm{d}B_s - \frac{1}{2}\int_0^t v_s^{(\gamma,\nu,\Sigma)} \mathrm{d}s, \qquad X_0 = 0, \\ v_t^{(\gamma,\nu,\Sigma)} &= v_0 \textstyle\sum_{i=1}^n \gamma_i \mathcal{E}\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}_t\right), \qquad v_0 > 0, \end{aligned} \tag{3.4}$$

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

For all results involving models (1.11), (3.3), and (3.4) we fix $\mathcal{T} = [0, 1]$; minor adjustments to the proofs yield analogous results for more general $\mathcal{T}$. We recall the definition $\beta := 2\alpha + 1 \in (0, 1)$ from Chapter 2.

Remark 3.1.3. In models (1.11), (3.3), and (3.4) 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 (Proposition 1.5.13) as long as the mapping is continuous.

Remark 3.1.4. Note that the mixed multi-factor rough Bergomi defined in (3.4) is indeed general enough to cover both (1.11) and (3.3). However, we provide our theoretical results in an orderly fashion starting from (1.11) and finishing with (3.4), which we find the most natural way to increase the complexity of the model.

3.2 Small-time Results for Options on Integrated Variance

We start our theoretical analysis by considering options on integrated variance, which we also refer to as realised variance and RV interchangeably. We recall that volatility is not directly observable, nor a tradeable asset. Options on realised variance, however, exist and are traded as OTC products. Below are two examples of the payoff structure of such products:

$$(i)(RV(v)(T) - K)^+, \quad (ii)(\sqrt{RV(v)(T)} - K)^+, \quad \text{where } T, K \geq 0. \tag{3.5}$$

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

$$RV(f)(\cdot) : f \mapsto \frac{1}{\cdot}\int_0^{\cdot} f(s)\mathrm{d}s, \quad RV(f)(0) := f(0), \tag{3.6}$$

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

Remark 3.2.1. As shown by Neuberger [Neu94], we may rewrite the variance swap in terms of the log contract as

$$\mathbb{E}[RV(v)(T)] = \mathbb{E}\left[\frac{1}{T}\int_0^T v_s\mathrm{d}s\right] = \mathbb{E}\left[-2\frac{X_T}{T}\right] \tag{3.7}$$

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 (3.7).

As with Call options on stock price processes, we can define and study the implied volatility of options on integrated variance. In the case of (3.5)(i) we define the implied volatility $\hat{\sigma}(T,k)$ to be the solution to

$$\mathbb{E}[(RV(v)(T) - e^k)^+] = \mathrm{C_{BS}}(RV(v)(0), k, T, \hat{\sigma}(T,k)), \tag{3.8}$$

where $\mathrm{C_{BS}}$ denotes the Call price in the Black-Scholes model. Using Corollary 3.2.6, we deduce the small-time behaviour of the implied volatility $\hat{\sigma}$, as defined in (3.8).

3.2.1 Small-time Results for the Rough Bergomi Model

We first consider the most simple case, the rough Bergomi model. We start by proving a pathwise large deviations principle for the variance process v, Theorem 3.2.2, and then deduce the small-time behaviour of the integrated variance process $RV(v)$. This allows us to characterise the small-time integrated variance implied volatility behaviour.

Before stating Theorem 3.2.2, we define the following function $\Lambda^Z : \mathcal{C}(\mathcal{T}) \to \mathbb{R}_+$ as $\Lambda^Z(\mathrm{x}) := \frac{1}{2}\|\mathrm{x}\|^2_{\mathscr{H}^{K_\alpha}}$, and if $\mathrm{x} \notin \mathscr{H}^{K_\alpha}$ then $\Lambda^Z(\mathrm{x}) = +\infty$, where $\mathscr{H}^{K_\alpha}$ is defined in Proposition 1.4.15.

Theorem 3.2.2. *The sequence of variance processes $(v_{\varepsilon\cdot})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function $\Lambda^v(\mathrm{x}) := \Lambda^Z\left(\log\left(\frac{\mathrm{x}}{v_0}\right)\right)$, where $\Lambda^v(v_0) = 0$ and $\mathrm{x} \in \mathcal{C}(\mathcal{T})$.*

Proof of Theorem 3.2.2. For $t \in \mathcal{T}$, $\varepsilon > 0$, we first define the rescaled processes

$$\begin{aligned} Z_t^\varepsilon &:= \varepsilon^{\beta/2} Z_t \overset{\mathrm{d}}{=} Z_{\varepsilon t}, \\ v_t^\varepsilon &:= v_0 \exp\left(Z_t^\varepsilon - \frac{\eta^2}{2}(\varepsilon t)^\beta\right), \end{aligned} \tag{3.9}$$

where $\beta := 2\alpha + 1 \in (0,1)$. From Theorem 1.5.14 and Proposition 1.4.15, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. We now prove that the two sequences of stochastic processes Z^ε and $\widetilde{Z}^\varepsilon := Z^\varepsilon - \frac{\eta^2}{2}(\varepsilon\cdot)^\beta$ are exponentially equivalent [DZ10, Definition 4.2.10]. For each $\delta > 0$ and $t \in \mathcal{T}$, there exists $\varepsilon_* := \frac{1}{t}\left(\frac{2\delta}{\eta^2}\right)^{1/\beta} > 0$ such that

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

for all $0 < \varepsilon < \varepsilon_*$. Therefore, for all $\delta > 0$, $\limsup_{\varepsilon\downarrow 0} \varepsilon^\beta \log \mathbb{P}(\|Z^\varepsilon - \widetilde{Z}^\varepsilon\|_\infty > \delta) = -\infty$, and the two processes are indeed exponentially equivalent. Then, using [DZ10, Theorem 4.2.13], the sequence of stochastic processes $(\widetilde{Z}^\varepsilon)_{\varepsilon>0}$ also satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$, with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. Moreover, for all ε, t, we have that $v_t^\varepsilon = v_0 \exp(\widetilde{Z}_t^\varepsilon)$, where the bijective transformation $\mathrm{x}(t,\varepsilon) \mapsto v_0 \exp(\mathrm{x}(t,\varepsilon))$ is clearly continuous with respect to the sup norm metric. Therefore we can apply the Contraction Principle (Proposition 1.5.13), concluding that the sequence of processes $(v^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function $\Lambda^Z\left(\log\left(\frac{\mathrm{x}(1,\cdot)}{v_0}\right)\right)$. Here we have used that, for each $\varepsilon > 0$, $t \in \mathcal{T}$ and $\mathrm{x} \in \mathcal{C}(\mathcal{T} \times \mathbb{R}_+)$, the inverse mapping of the bijection transformation $\mathrm{x}(t,\varepsilon) \mapsto v_0 \exp(\mathrm{x}(t,\varepsilon))$ is given by $\log\left(\frac{\mathrm{x}(t,\varepsilon)}{v_0}\right)$. 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 $\Lambda^v(v_0) = \Lambda^Z(0) = \|0\|^2_{\mathscr{H}^{K_\alpha}} = 0$. □

Corollary 3.2.3. *The integrated variance process $(RV(v)(t))_{t\in\mathcal{T}}$ satisfies a large deviations principle on $\mathbb{R}_+^*$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^v$ defined as $\hat{\Lambda}^v(\mathrm{y}) := \inf\{\Lambda^v(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\}$, where $\hat{\Lambda}^v(v_0) = 0$.*

Proof. As proved in Theorem 3.2.2, $(v_\varepsilon)_{\varepsilon>0}$ 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

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

where $M = \sup_{t\in\mathcal{T}}|\delta^v(t)|$, which is finite as $\delta^v \in \mathcal{C}(\mathcal{T})$. Clearly M tends to zero as δ^v tends 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 (Proposition 1.5.13), and consequently the integrated variance process $RV(v_\varepsilon)$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero. Clearly $RV(v_\varepsilon)(t) = RV(v)(\varepsilon t)$, for all $t \in \mathcal{T}$, and so setting $t = 1$ and mapping ε to t then yields the result. By definition, $\hat{\Lambda}^v(\mathrm{y}) := \inf\{\Lambda^v(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\}$. If we choose $\mathrm{x} \equiv v_0$ then clearly $v_0 = RV(\mathrm{x})(1)$, and $\Lambda^v(\mathrm{x}) = 0$. Since Λ^v is a norm, it is a non-negative function and therefore $\hat{\Lambda}^v(v_0) = 0$. This concludes the proof. □

Remark 3.2.4. Corollary 3.2.3 can be applied to a large number of existing results on pathwise large deviations for rough variance processes to get a large deviations result for the integrated rough variance process; for example the fSABR model in Section 2.2 and Forde and Zhang's rough volatility model [FZ17].

Corollary 3.2.5. *The rate function $\hat{\Lambda}^v$ is continuous.*

Proof. Indeed, as a rate function, $\hat{\Lambda}^v$ is lower semi-continuous. Moreover, as Λ^v is continuous, one can use similar arguments to [FZ17, Corollary 4.6], and deduce that $\hat{\Lambda}^v$ is upper semi-continuous, and hence is continuous. □

Before stating results on the small-time behaviour of options on integrated variance, we state that the log integrated variance process $\log RV(v)$ satisfies a large deviations principle on $\mathbb{R}$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^v(\mathrm{e}^{\cdot})$. Then, the small-time behaviour of such options can be obtained as an application of Corollary 3.2.3.

Corollary 3.2.6. *For log moneyness $k := \log\frac{K}{RV(v)(0)} \neq 0$, the following equality holds true for Call options on integrated variance*

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(RV(v)(t) - \mathrm{e}^k\right)^+\right] = -\mathrm{I}(k), \tag{3.10}$$

where I *is defined as as* $\mathrm{I}(x) := \inf_{y>x} \hat{\Lambda}^v(\mathrm{e}^y)$ *for* $x > 0$, $\mathrm{I}(x) := \inf_{y<x} \hat{\Lambda}^v(\mathrm{e}^y)$ *for* $x < 0$. *Similarly, for log moneyness* $k := \log \frac{K}{\sqrt{RV(v)(0)}} \neq 0$,

$$\lim_{t\downarrow 0} t^\beta \log \mathbb{E}\left[\left(\sqrt{RV(v)(t)} - \mathrm{e}^k\right)^+\right] = -\bar{\mathrm{I}}(k), \tag{3.11}$$

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

Proof of Corollary 3.2.6. The proof of Equation (3.10) is similar to the proof of [FZ17, Corollary 4.9], and we shall prove the lower and upper bound separately, which turn out to be equal. Firstly, as the rate function $\hat{\Lambda}^v$ is continuous on $\mathcal{C}(\mathcal{T})$, we have that, for all $k > 0$,

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

as an application of Corollary 3.2.3.

(1) The proof of the lower bound is exactly the same as presented in [FZ17, Appendix C] and will be omitted here; we arrive at $\liminf_{t\downarrow 0} t^\beta \log \mathbb{E}\left[(RV(v)(t) - \mathrm{e}^k)^+\right] \geq -\mathrm{I}(k)$.

(2) We establish the upper bound:
We first apply Hölder's inequality:

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

which holds for all $q > 1$. Thus

$$t^\beta \log \mathbb{E}[(RV(v)(t) - \mathrm{e}^k)]^+ \leq \frac{t^\beta}{q} \log \mathbb{E}\left[(RV(v)(t))^q\right] + t^\beta \left(1 - \frac{1}{q}\right) \log \mathbb{P}\left(RV(v)(t) \geq \mathrm{e}^k\right). \tag{3.12}$$

We further obtain the following inequality by applying Jensen's inequality and the fact that all moments exist for $(RV(v)(t))^q$

$$\begin{aligned}\mathbb{E}\left[(RV(v)(t))^q\right] &\leq \frac{1}{t^q}\int_0^t \mathbb{E}[v_s^q]\mathrm{d}s \leq \frac{v_0^q}{t^q}\int_0^t \exp\left(\left(\frac{q^2\eta^2}{2}-\frac{q\eta^2}{2}\right)s^{2\alpha+1}\right)\mathrm{d}s \quad (3.13)\\ &\leq \frac{v_0^q}{t^{q-1}}\exp\left(\left(\frac{q^2\eta^2}{2}-\frac{q\eta^2}{2}\right)t^{2\alpha+1}\right).\end{aligned}$$

Therefore,

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

Hence, taking $q\uparrow\infty$ and $t\downarrow 0$ on both sides of (3.12), we obtain by Corollary 3.2.3

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

which concludes the proof. The proof of Equation (3.11) follows the same steps, after proving that the process $\sqrt{RV(v)}$ satisfies a large deviations principle on $\mathbb{R}_+$. Indeed, as the function $x\mapsto x^2$ is a continuous bijection on $\mathbb{R}_+$, we have that the square root of the integrated variance process $\sqrt{RV(v)}$ satisfies a large deviations principle on $\mathbb{R}_+$ as t tends to zero, with speed $t^{-\beta}$ and rate function $\hat{\Lambda}^v((\cdot)^2)$, using [DZ10, Theorem 4.2.4].

□

Corollary 3.2.7. *The small-time asymptotic behaviour of the implied volatility is given by the following limit, for a log moneyness $k\neq 0$:*

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

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

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

In the Black Scholes model, i.e. a geometric Brownian motion with $S_0 = RV(v)(0)$ with constant volatility ξ, we have the following small-time implied volatility behaviour:

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

We then apply [GL14, Corollary 7.1], identifying $\xi \equiv \hat{\sigma}(k,t)$, to conclude. □

Remark 3.2.8. Notice that the level of implied volatility in Corollary 3.2.7 has a power law behaviour as a function of time to maturity. This power law is of order $\sqrt{t^{\beta-1}}$, which is consistent with the at-the-money RV implied volatility results by Alòs, García-Lorite and Muguruza [AGM18], where the order is found to be $t^{H-1/2}$ using Malliavin Calculus techniques. Recall that $\beta = 2\alpha + 1$, and $\alpha = H - 1/2$ by Remark 1.3.7.

3.2.2 Small-time Results for the Mixed Rough Bergomi model

We now consider the mixed rough Bergomi model. As with the previous Section, our aim is to establish the small-time integrated variance implied volatility behaviour. Minor adjustments to Theorem 3.2.2 give the following result for the mixed variance process $v^{(\gamma,\nu)}$ introduced in Model (3.3).

Theorem 3.2.9. *The sequence of mixed variance processes $(v_{\varepsilon\cdot}^{(\gamma,\nu)})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function*

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

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

Proof. For brevity we set $n = 2$, but for larger n, identical arguments can be applied. From Theorem 1.5.14 and Proposition 1.4.15, we have that the sequence of processes $(Z^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ with speed $\varepsilon^{-\beta}$ and rate function Λ^Z. Define the operator $f : \mathcal{C}(\mathcal{T}) \to \mathcal{C}((\mathcal{T}), \mathbb{R}^2)$ by $f(\mathrm{x}) := (\frac{\nu_1}{\eta}\mathrm{x}, \frac{\nu_2}{\eta}\mathrm{x})$, which is clearly continuous with respect to the sup-norm $\|\cdot\|_\infty$ on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$. Applying the Contraction Principle (Proposition 1.5.13) then yields that the sequence of two-dimensional processes

$((\frac{\nu_1}{\eta}Z^\varepsilon, \frac{\nu_2}{\eta}Z^\varepsilon))_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^2)$ as ε tends to zero with speed $\varepsilon^{-\beta}$ and rate function

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

Identical arguments to the proof of Theorem 3.2.2 give that the sequences of processes $((\frac{\nu_1}{\eta}Z^\varepsilon, \frac{\nu_2}{\eta}Z^\varepsilon))_{\varepsilon>0}$ and $((\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^\beta, \frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu^2}{2}(\varepsilon\cdot)^\beta))_{\varepsilon>0}$ are exponentially equivalent, thus satisfy the same large deviations principle, with the same rate function and the same speed.

We now define the operator $g^\gamma : \mathcal{C}(\mathcal{T}, \mathbb{R}^2) \to \mathcal{C}(\mathcal{T})$ as $g^\gamma(\mathrm{x}, \mathrm{y}) = v_0(\gamma e^{\mathrm{x}} + (1-\gamma)e^{\mathrm{y}})$, where $\gamma = \gamma_1, 1-\gamma = \gamma_2$. For small perturbations $\delta^{\mathrm{x}}, \delta^{\mathrm{y}} \in \mathcal{C}(\mathcal{T})$ we have that[2]

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

Clearly the right hand side tend to zero as $\delta^{\mathrm{x}}, \delta^{\mathrm{y}}$ tend 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 (Proposition 1.5.13) then yields that the sequence of processes

$$(v^{(\varepsilon,\gamma,\nu)})_{\varepsilon>0} := \left(v_0 \left(\gamma \exp(\frac{\nu_1}{\eta}Z^\varepsilon - \frac{\nu_1^2}{2}(\varepsilon\cdot)^\beta) + (1-\gamma)\exp(\frac{\nu_2}{\eta}Z^\varepsilon - \frac{\nu_2^2}{2}(\varepsilon\cdot)^\beta) \right) \right)_{\varepsilon>0}$$

satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function

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

Since, for all $\varepsilon > 0$ and $t \in \mathcal{T}$, $v_{\varepsilon t}^{(\gamma,\nu)}$ and $v_t^{(\varepsilon,\gamma,\nu)}$ are equal in law, the theorem follows immediately. Identical arguments to the proof of Theorem 3.2.2 then yield that $\Lambda^\gamma(v_0) = 0$. □

By Remark 3.2.4, we immediately get the following result for the small-time behaviour of the integrated mixed variance process $RV(v^{(\gamma,\nu)})$.

[2] Recall we fix $\mathcal{T} = [0, 1]$ on p65.

Corollary 3.2.10. *The integrated mixed variance process* $(RV\left(v^{(\gamma,\nu)}\right)(t))_{t\in\mathcal{T}}$ *satisfies a large deviations principle on* $\mathbb{R}_+^*$ *as* t *tends to zero, with speed* $t^{-\beta}$ *and rate function* $\tilde{\Lambda}^{(\gamma,\nu)}(\mathrm{y}) := \inf\left\{\Lambda^{(\gamma,\nu)}(\mathrm{x}) : \mathrm{y} = RV(\mathrm{x})(1)\right\}$*, where* $\tilde{\Lambda}^{(\gamma,\nu)}(v_0) = 0$*.*

To get the small-time implied volatility result, analogous to Corollary 3.2.7, we need the following Lemma, which is used in place of (3.13). The remainder of the proof then follows identically.

Lemma 3.2.11. *For all* $t \in \mathcal{T}$ *and* $q > 1$ *we have*

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

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

Proof. First we note that by Hölder's inequality $(\sum_{i=1}^n x_i)^q \leq n^{q-1}\sum_{i=1}^n (x_i)^q$, for $x_i > 0$. Since, $\gamma_i \leq 1$ for $i = 1, ..., n$, we obtain

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

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

Corollary 3.2.12. *For log moneyness* $k := \log\frac{K}{RV(v^{(\gamma,\nu)})(0)} \neq 0$*, the following equality holds true for Call options on integrated variance in the mixed rough Bergomi model:*

$$\lim_{t\downarrow 0} t^\beta \log\mathbb{E}\left[\left(RV(v^{(\gamma,\nu)})(t) - \mathrm{e}^k\right)^+\right] = -\mathrm{I}(k), \tag{3.14}$$

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

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

$$\lim_{t\downarrow 0} t^\beta \log\mathbb{E}\left[\left(\sqrt{RV(v^{(\gamma,\nu)})(t)} - \mathrm{e}^k\right)^+\right] = -\bar{\mathrm{I}}(k), \tag{3.15}$$

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

Proof. Follows directly from Lemma 3.2.11 and proof of Corollary 3.2.6. □

The small-time implied volatility behaviour for the mixed rough Bergomi model is then given by Corollary 3.2.7, where the function I is defined in terms of the rate function $\tilde{\Lambda}^{(\gamma,\nu)}$, as in Corollary 3.2.12, in this case.

3.2.3 Small-time Results for the Multi-factor Rough Bergomi model

Finally we consider the multi-factor rough Bergomi model (3.4), the most general of the three cases. The asymptotic behaviour for the variance process is given in Theorem 3.2.13 below; note that Λ^m is the rate function associated to the reproducing kernel Hilbert space of the measure induced by $\mathcal{Z}$ on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$, denoted $\mathscr{H}_m$.

Theorem 3.2.13. *The sequence of processes* $\left(v_{\varepsilon\cdot}^{(\gamma,\nu,\Sigma)}\right)_{\varepsilon>0}$ *satisfies a large deviations principle on* $\mathcal{C}(\mathcal{T})$ *with speed* $\varepsilon^{-\beta}$ *and rate function*

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

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

Proof of Theorem 3.2.13. We begin by introducing a rescaling of (3.4) for $\varepsilon > 0$, so that the system becomes

$$v_t^{(\gamma,\nu,\Sigma,\varepsilon)} := v_{\varepsilon t}^{(\gamma,\nu,\Sigma)} = v_0\sum_{i=1}^n \gamma_i \mathcal{E}\left(\frac{\nu^i}{\eta}\cdot \mathrm{L}_i\mathcal{Z}_t^\varepsilon\right), \tag{3.16}$$

with the rescaled process $\mathcal{Z}_t^\varepsilon$ defined as

$$\mathcal{Z}_t^\varepsilon := \mathcal{Z}_{\varepsilon t} = \varepsilon^{\alpha+\frac{1}{2}}\left(\int_0^t K_\alpha(s,t)\mathrm{d}W_s^1, ..., \int_0^t K_\alpha(s,t)\mathrm{d}W_s^m\right).$$

The m-dimensional sequence of processes $(\varepsilon^{\beta/2}\mathcal{Z}^\varepsilon)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$ as ε goes to zero with speed $\varepsilon^{-\beta}$ and rate function Λ^m defined by $\Lambda^m(\mathrm{x}) := \frac{1}{2}\left\|\mathrm{x}\right\|^2_{\mathscr{H}_m}$ for $\mathrm{x}\in\mathscr{H}_m$ and $+\infty$ otherwise, by Theorem 1.5.14. $\mathscr{H}_m$ is the reproducing kernel Hilbert space of the measure induced by $\mathcal{Z}$ on $\mathcal{C}(\mathcal{T},\mathbb{R}^m)$, defined as

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

Then, using an extension of the proof of [HJL19, Theorem 3.6], for $i = 1, \cdots, n$, the sequence of m-dimensional processes $(\mathrm{L}_i \mathcal{Z}^\varepsilon_\cdot)_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T}, \mathbb{R}^m)$, as ε tends to zero with rate function $\mathrm{y} \mapsto \inf\{\Lambda^m(\mathrm{x}) : \mathrm{x} \in \mathscr{H}_m, \mathrm{y} = \mathrm{L}_i \mathrm{x}\}$ and speed $\varepsilon^{-\beta}$, where L_i is the lower triangular matrix introduced in Model (3.4). Consequently, for $i = 1, \cdots, n$ each (one-dimensional) sequence of processes $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}^\varepsilon_\cdot\right)_{\varepsilon>0}$ also satisfies a large deviations principle as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function $\Lambda_{\Sigma_i}(\mathrm{y}) := \inf\left\{\Lambda^m(\mathrm{x}) : \mathrm{x} \in \mathscr{H}_m, \mathrm{y} = \frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathrm{x}\right\}$.

Each sequence of processes $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}^\varepsilon_\cdot\right)_{\varepsilon>0}$ and $\left(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}^\varepsilon_\cdot - \frac{1}{2}\nu^i \Sigma_i \nu^i (\varepsilon \cdot)^\beta\right)_{\varepsilon>0}$ are exponentially equivalent for $i = 1, \cdots, n$; therefore they satisfy the same large deviations principle with the same speed $\varepsilon^{-\beta}$ and the same rate function Λ_{Σ_i}.

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

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

with $\mathrm{x} := (\mathrm{x}_1, \cdots, \mathrm{x}_n)$. For small perturbations $\delta^1, \cdots, \delta^n \in \mathcal{C}(\mathcal{T})$ with $\delta := (\delta^1, \cdots, \delta^n)$, we have that

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

The right-hand side tends to zero as $\delta^1, \cdots, \delta^n$ tends to zero; thus the operator g^γ is continuous with respect to the sup-norm $\|\cdot\|_\infty$ on $\mathcal{C}(\mathcal{T})$. Using that $v_t^{(\gamma,\nu,\Sigma,\varepsilon)} = g^\gamma(\frac{\nu^i}{\eta} \cdot \mathrm{L}_i \mathcal{Z}^\varepsilon_\cdot - \frac{1}{2}\nu^i \Sigma_i \nu^i (\varepsilon \cdot)^\beta)(t)$ for each $\varepsilon > 0$ and $t \in \mathcal{T}$, we can apply the Contraction Principle (Proposition 1.5.13), yielding that the sequence of processes $(v^{(\gamma,\nu,\Sigma,\varepsilon)})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}(\mathcal{T})$ as ε tends to zero, with speed $\varepsilon^{-\beta}$ and rate function

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

As with the previous two models we have that, for all $\varepsilon > 0$ and $t \in \mathcal{T}$, $v_t^{(\gamma,\nu,\Sigma,\varepsilon)}$ and $v_{\varepsilon t}^{(\gamma,\nu,\Sigma)}$ are equal in law and so the result follows directly. □

In the same way as with the mixed variance process, Remark 3.2.4 gives us the following small-time result for $RV(v^{(\gamma,\nu,\Sigma)})$ straight off the bat.

Corollary 3.2.14. *The integrated variance process* $\left(RV\left(v^{(\gamma,\nu,\Sigma)}\right)(t)\right)_{t\in\mathcal{T}}$ *in the multi-factor Bergomi model satisfies a large deviations principle on* $\mathbb{R}_+^*$ *as* t *tends to zero, with speed* $t^{-\beta}$ *and rate function*

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

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

We now establish the small-time behaviour for Call options on realised variance in Corollary 3.2.16, by adapting the proof of Corollary 3.2.6 as in the previous subsection. To do so we use Lemma 3.2.15 in place of (3.13). Then we attain the small-time implied volatility behaviour for the multi-factor rough Bergomi model in Corollary 3.2.7, where the function I is given by Corollary 3.2.16.

Lemma 3.2.15. *For all* $t \in \mathcal{T}$ *and* $q > 1$ *we have*

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

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

Proof. First we note that by Hölder's inequality $(\sum_{i=1}^n x_i)^q \leq n^{q-1}\sum_{i=1}^n (x_i)^q$, for $x_i > 0$. Since, $\gamma_i \leq 1$ for $i = 1, ..., n$, we obtain

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

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

Corollary 3.2.16. *For log moneyness* $k := \log \frac{K}{RV(v^{(\gamma,\nu,\Sigma)})(0)} \neq 0$*, the following equality holds true for Call options on integrated variance in the multi-factor rough Bergomi model:*

$$\lim_{t\downarrow 0} t^{\beta} \log \mathbb{E}\left[\left(RV(v^{(\gamma,\nu,\Sigma)})(t) - \mathrm{e}^{k}\right)^{+}\right] = -\mathrm{I}(k), \tag{3.17}$$

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

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

$$\lim_{t\downarrow 0} t^{\beta} \log \mathbb{E}\left[\left(\sqrt{RV(v^{(\gamma,\nu,\Sigma)})(t)} - \mathrm{e}^{k}\right)^{+}\right] = -\bar{\mathrm{I}}(k), \tag{3.18}$$

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

Proof. Follows directly from Lemma 3.2.15 and the proof of Corollary 3.2.6. □

3.3 Numerical Schemes

Having established a number of theoretical results for small-time integrated variance implied volatility in Section 3.2, we now tackle the problem of numerically computing the various rate functions with the ultimate aim of producing plots to represent the small-time integrated variance implied volatility smiles. We begin this Section by presenting the numerical schemes used to compute the rate functions in each case; we then present numerical results for each of the three models in Section 3.4 below, and analyse the effect of each parameter on the implied volatility smile.

3.3.1 Single-factor Case

We first consider the rough Bergomi (1.11) model for sake of simplicity and further develop the mixed multi-factor rough Bergomi (3.4) model in Section 3.3.3 (which also includes the mixed rough Bergomi model (3.3)). Therefore, we tackle the numerical

computation of the rate function

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

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

$$\hat{\Lambda}^v(\mathrm{y}) := \inf_{f\in \mathrm{L}^2}\left\{\frac{1}{2}\|f\|^2_{\mathrm{L}^2} : \mathrm{y} = RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot)f(u)\mathrm{d}u\right)\right)(1)\right\}. \tag{3.19}$$

A natural approach is to consider a class of functions that is dense in L^2. 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 as n tends to $+\infty$. Problem (3.19) may then be approximated via

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

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

$$a_i^* = \mathrm{argmin}_{a_i\in\mathbb{R}}\left\{\left(\mathrm{y} - RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}^{(n)}(u)\mathrm{d}u\right)\right)(1)\right)^2\right\} \tag{3.20}$$

and the constraint will be satisfied for all combinations of $a^* = (a_0, ..., a_{i-1}, a_i^*, a_{i+1}, ..., a_n)$. Numerically, (3.20) is easily solved using a few iterations of the Newton-Raphson algorithm. Then we may easily solve

$$\inf_{a^*\in\mathbb{R}^{n+1}}\left\{\frac{1}{2}||\hat{f}^{(n)}||^2_{\mathrm{L}^2}\right\}$$

which will converge to the original problem (3.19) as $n \to +\infty$. The polynomial basis is particularly convenient since we have that

$$\begin{aligned} &RV\left(\exp\left(\int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}^{(n)}(u)\mathrm{d}u\right)\right)(1) \\ &\quad= \int_0^1 \exp\left(\eta\sqrt{2\alpha+1}\sum_{i=0}^{n}\frac{a_i s^{\alpha+1+i}\ {}_2F_1(i+1,-\alpha,i+2,1)}{i+1}\right)\mathrm{d}s, \end{aligned} \tag{3.21}$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function. In particular one may store the values $\{{}_2F_1\left(i+1,-\alpha,i+2,1\right))\}_{i=0}^n$ in the computer memory and reuse them through different iterations. In addition, the outer integral in (3.21) is efficiently computed using Gauss-Legendre quadrature i.e.

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

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

3.3.2 A Tailor-made Polynomial Basis for Rough Volatility

We may improve the computation time of the previous approach by considering a tailor-made polynomial basis. In particular, recall the following relation

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

then, for $k=-\alpha-1$ we obtain

$$\int_0^s K_\alpha(u,s)u^{-\alpha-1}\mathrm{d}u = \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha},$$

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

Proposition 3.3.1. *Consider the basis* $\hat{g}^{(n)}(s)=cs^{-\alpha-1}+\sum_{i=0}^n a_i s^i$. *Then,*

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

solves (3.20).

Proof. We have that

$$\begin{aligned} RV\left(\int_0^{\cdot}K_\alpha(u,\cdot)\hat{g}^{(n)}(u)\mathrm{d}u\right)(1) &= \exp\left(\frac{\eta\sqrt{2\alpha+1}}{-\alpha}\ {}_2F_1\left(-\alpha,-\alpha,1-\alpha,1\right)\right) \\ &\quad\times\int_0^1\exp\left(\eta\sqrt{2\alpha+1}\sum_{i=0}^{n}\frac{a_i s^{\alpha+1+i}\ {}_2F_1\left(i+1,-\alpha,i+2,1\right)}{i+1}\right)\mathrm{d}s \end{aligned}$$

and the proof trivially follows by solving $y=RV\left(\int_0^{\cdot}K_\alpha(u,\cdot)\hat{g}^{(n)}(u)\mathrm{d}u\right)(1)$. □

Remark 3.3.2. Notice that Proposition 3.3.1 gives a semi-closed form solution to (3.20). 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 (3.19).

Remark 3.3.3. Notice that $u^{-\alpha-1} \notin \mathrm{L}^2$, however $u^{-\alpha-1}\mathbb{I}_{\{u>\varepsilon\}} \in \mathrm{L}^2$ for all $\varepsilon > 0$. Moreover,

$$\begin{aligned}\int_0^s K_\alpha(s,u)u^{-\alpha-1}\mathbb{I}_{\{u>\varepsilon\}}\mathrm{d}u &= \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha} - \frac{\epsilon^{-\alpha}s^\alpha\ {}_2F_1(-\alpha,-\alpha,1-\alpha,\frac{\epsilon}{t})}{-\alpha} \\ &= \frac{{}_2F_1(-\alpha,-\alpha,1-\alpha,1)}{-\alpha} + \mathcal{O}(\varepsilon^{-\alpha}),\end{aligned}$$

hence for ε sufficiently small the error is bounded as long as $\alpha \neq 0$. In our applications we find that this method behaves nicely for $\alpha \in (-0.5, -0.05]$. In Figure 3.2 we provide precise errors and we observe that the convergence is better for small α (which is rather surprising behaviour, as the converse is true of other approximation schemes when the volatility trajectories become more rough) as well as strikes around the money. Moreover, the truncated basis approach constitutes a 30-fold speed improvement in our numerical tests. As benchmark we consider the standard numerical algorithm introduced in (3.20), with accuracy measured by absolute error.

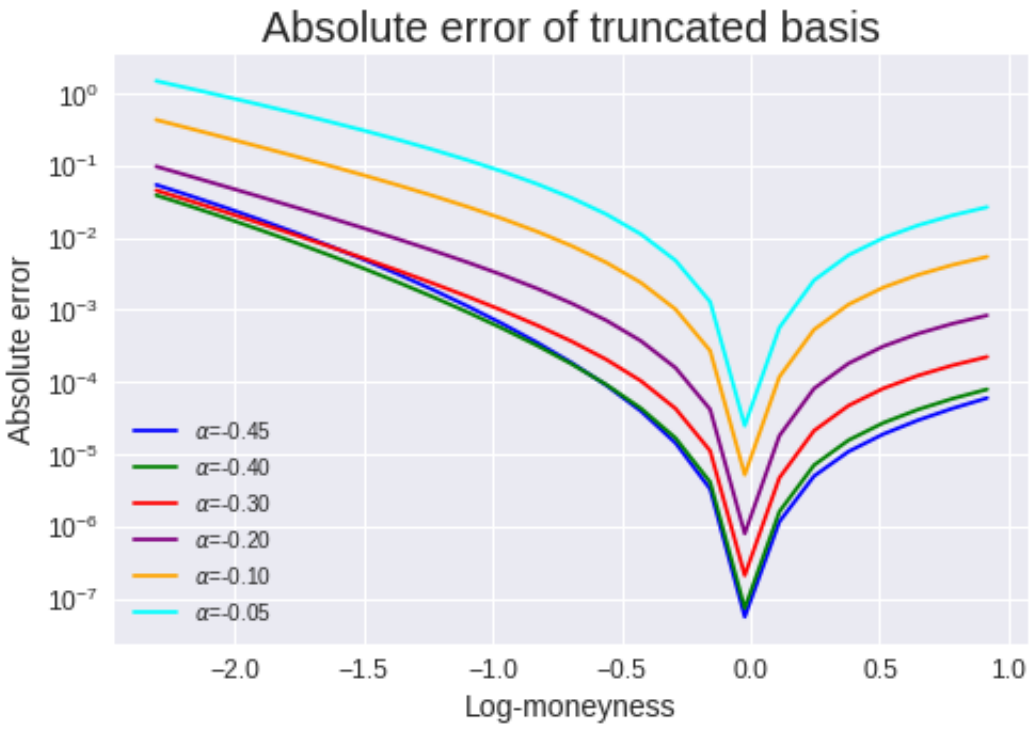

Figure 3.2: Absolute error of the rate function. We consider the truncated basis approach against the standard polynomial basis with $(\eta, v_0, n) = (1.5, 0.04, 5)$ and different values of α.

3.3.3 Multi-Factor Case

The correlated mixed multi-factor rough Bergomi (3.4) model requires a slightly more complex setting. By Corollary 3.2.14, the rate function is given by the following multi-dimensional optimisation problem:

$$\hat{\Lambda}^{(v,\Sigma)}(\mathrm{y}) := \inf_{(f_1,...,f_m)\in \mathrm{L}^2} \left\{ \frac{1}{2}\sum_{i=1}^{m} ||f_i||^2_{\mathrm{L}^2} : \mathrm{y} = RV\left(\sum_{i=1}^{m} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \Sigma_i \mathfrak{f}_{\cdot}^{K_\alpha} \right) u \right)(1) \right\}, \tag{3.22}$$

where $\mathfrak{f}_{\cdot}^{K_\alpha} = \left(\int_0^{\cdot} K_\alpha(u,\cdot) f_1(u)\mathrm{d}u, ..., \int_0^{\cdot} K_\alpha(u,\cdot) f_m(u)\mathrm{d}u \right)$. The approach to solve this problem is similar to that of (3.19). Nevertheless, in order to solve (3.22) we shall use a multi-dimensional polynomial basis

$$\left(\hat{f}_1^{(p)}(s), ..., \hat{f}_m^{(p)}(s) \right) = \left(\sum_{i=0}^{p} a_i^1 s^i, ..., \sum_{i=0}^{p} a_i^n s^i \right)$$

such that each $\hat{f}_i^{(p)}(s)$ for $i \in \{1, ..., m\}$ is dense as p tends to $+\infty$ in L^2 by Stone-Weierstrass Theorem. Then we may equivalently solve

$$\inf_{(a_0^1,...,a_p^1,...,a_0^m,...,a_p^m)\in\mathbb{R}^{(p+1)m}} \left\{ \frac{1}{2}\sum_{i=1}^{m} ||\hat{f}_i^{(p)}||^2_{\mathrm{L}^2} : \mathrm{y} = RV\left(\sum_{i=1}^{m} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \Sigma_i \hat{\mathfrak{f}}_{\cdot}^{(K_\alpha,p)} \right) u \right)(1) \right\}, \tag{3.23}$$

where $\hat{\mathfrak{f}}_{\cdot}^{(K_\alpha,p)} = \left(\int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}_1^{(p)}(u)\mathrm{d}u, ..., \int_0^{\cdot} K_\alpha(u,\cdot)\hat{f}_m^{(p)}(u)\mathrm{d}u \right)$. Then as p tends to $+\infty$, (3.23) will converge to the original problem (3.22). In order to numerically accelerate the optimisation problem in (3.23), we anchor coefficients $(a_0^1, ..., a_0^m)$ to satisfy the constraint $y = RV(\cdot)(1)$ (same way we did in the one dimensional case), that is

$$\mathfrak{a}^* := \inf_{(a_0^1,....,a_0^m)\in\mathbb{R}^m} \left\{ \left(\mathrm{y} - RV\left(\sum_{i=1}^{m} \gamma_i \exp\left(\frac{\nu^i}{\eta}\cdot \Sigma_i \mathfrak{f}_{\cdot}^{K_\alpha} \right) u \right)(1) \right)^2 \right\}$$

where $\mathfrak{a}^* = (\mathfrak{a}_0^{1*}, ..., \mathfrak{a}_0^{m*})$ and one may use (3.21) and Gauss-Legendre quadrature to efficiently compute $RV(\cdot)(1)$. Then, the constraint will always be satisfied by construction and instead we may solve

$$\inf_{(\mathfrak{a}_0^{1*},a_1^1...,a_p^1,...,\mathfrak{a}_0^{m*},a_1^m,...,a_p^m)\in\mathbb{R}^{(p+1)m}} \left\{ \frac{1}{2}\sum_{i=1}^{m} ||\hat{f}^{(p)}||^2_{\mathrm{L}^2} \right\}. \tag{3.24}$$

3.4 Integrated Variance Smiles

We now use the numerical schemes described above to plot the small-time integrated variance smiles.

3.4.1 Smiles for the Rough Bergomi Model

We begin with the rough Bergomi model (1.11), using Corollary 3.2.7. In Figure 3.3, we represent the rate function given in Corollary 3.2.3, 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.

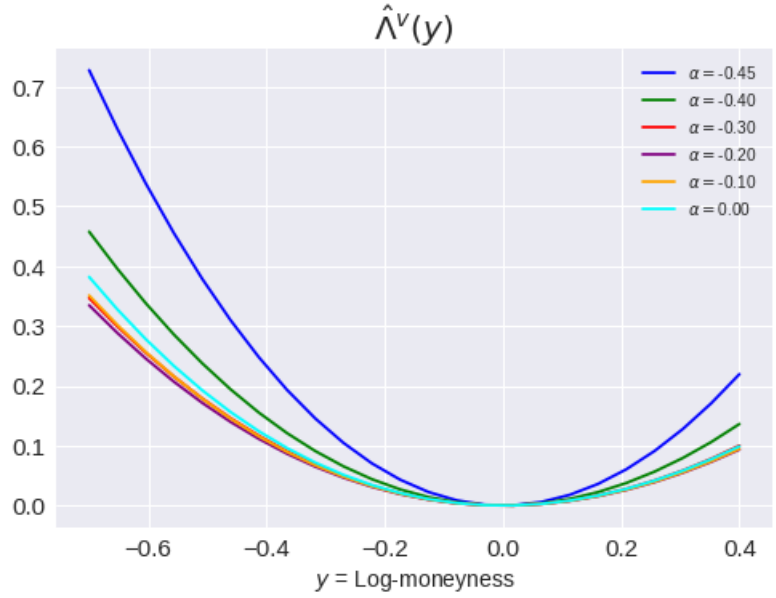

Figure 3.3: Rate function $\hat{\Lambda}^v$ for different values of α. We set $(\eta, v_0, n) = (2, 0.04, 5)$.

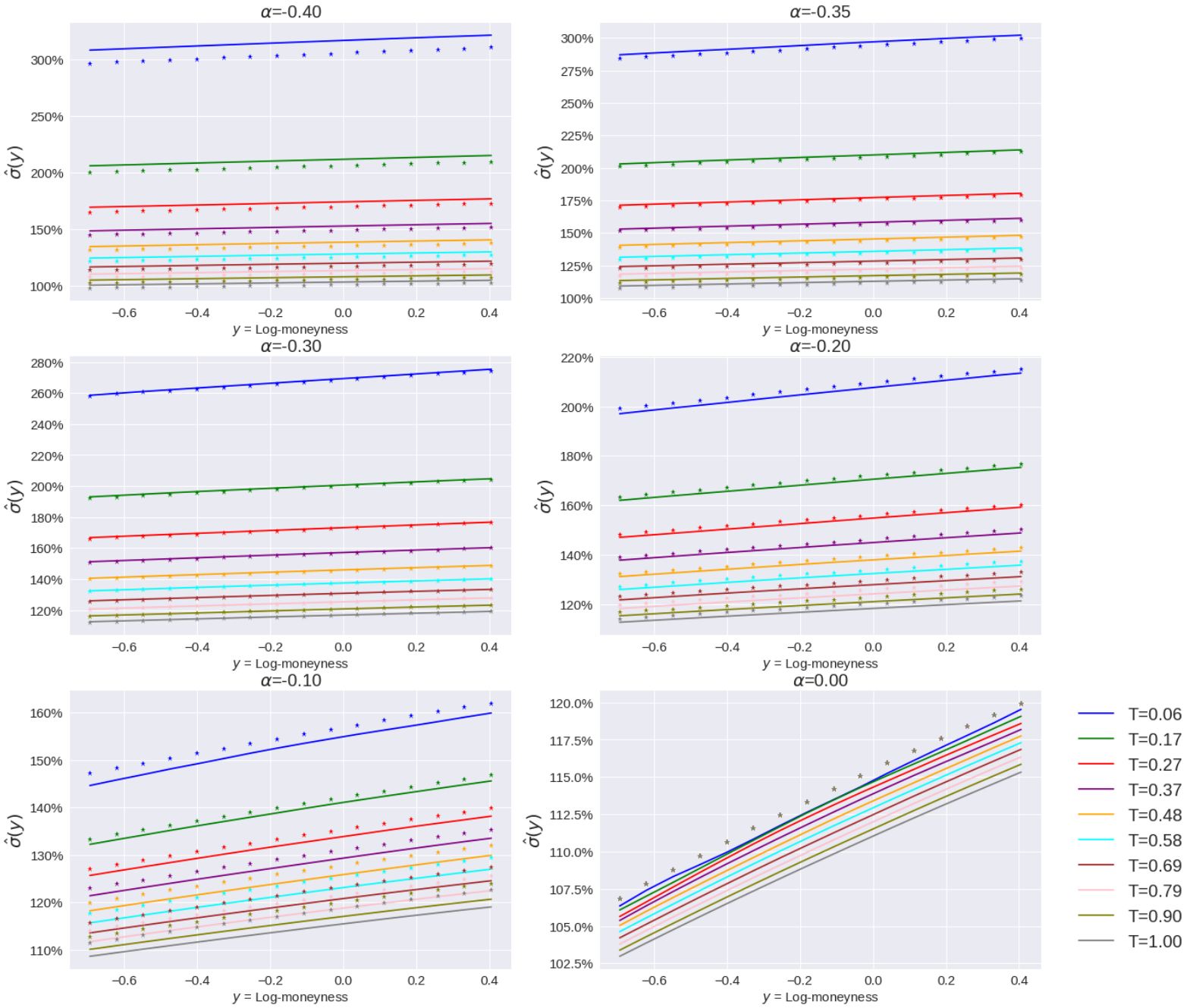

Figure 3.4: Comparison of Monte Carlo computed implied volatilities (straight lines) and LDP based implied volatilities (stars), in the rough Bergomi model, for different values of α and maturities T. We set $(\eta, v_0, n) = (2, 0.04, 5)$; for Monte Carlo we use $200,000$ simulations and $\Delta t = \frac{1}{1008}$.

More interestingly, in Figure 3.4 we provide a comparison of Corollary 3.2.7 with respect to a benchmark generated by Monte Carlo simulations, and see all smiles to follow a linear trend. In particular, we notice that Corollary 3.2.7 provides a surprisingly accurate estimate, even for relatively large maturities. As a further numerical check, in Figure 3.5 we compare our results with the close-form at-the-money asymptotics given by Alòs, García-Lorite and Muguruza [AGM18] and once again find the correct convergence, suggesting a consistent numerical framework.

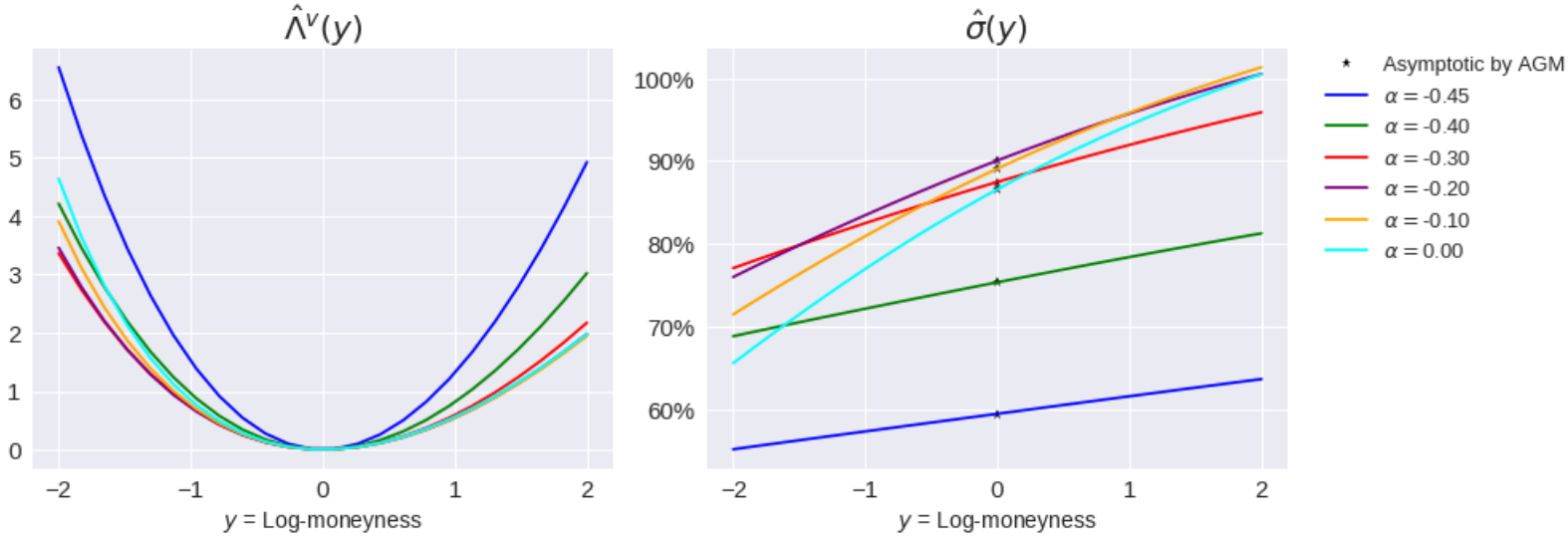

Figure 3.5: Comparison of Alòs, García-Lorite and Muguruza [AGM18] at-the-money implied volatility asymptotics and LDP based implied volatilities for different values of α, in the rough Bergomi model, with $(\eta, v_0, n) = (2, 0.04, 5)$.

3.4.2 Smiles for the Mixed Rough Bergomi Model

We now consider the mixed rough Bergomi model (3.3) in a simplified form given by $v_t = v_0 \left(\gamma_1 \mathcal{E}(\nu_1 Z_t) + \gamma_2 \mathcal{E}(\nu_2 Z_t)\right)$. In Figure 3.6, we observe that a constraint of the type $\gamma_1\nu_1 + \gamma_2\nu_2 = 2$ in the mixed variance process (3.3) 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 3.4, we see that the smiles generated follow a linear trend. This is again consistent with the results found in [AGM18].

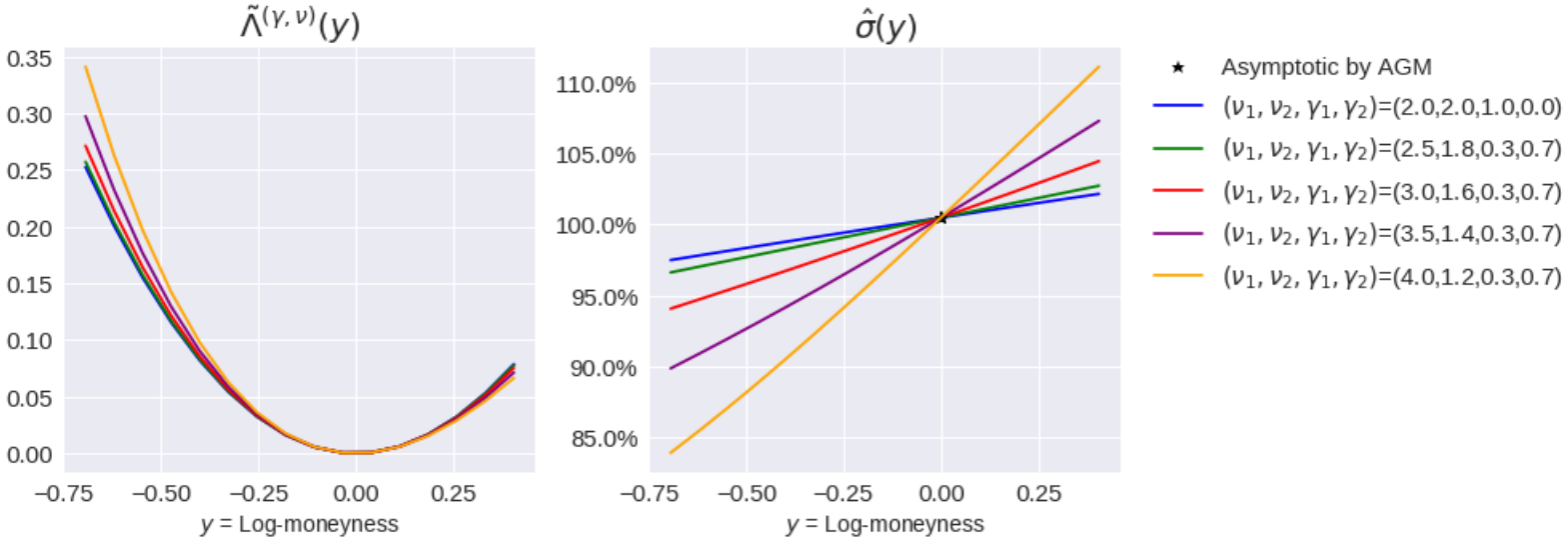

Figure 3.6: Comparison of LDP based implied volatilities for different values of $(\nu_1, \nu_2, \gamma_1, \gamma_2)$ in the mixed rough Bergomi process (3.3) such that $\gamma_1\nu_1 + \gamma_2\nu_2 = 2$, with $(\alpha, p) = (-0.4, 5)$.

Remark 3.4.1. At this point it is important to note that the mixed rough Bergomi model 3.3 allows both the at-the-money implied volatility and its skew to be controlled through (γ, ν), whilst in the rough Bergomi model (1.11) there is not enough freedom to arbitrarily fit both quantities. Remarkably, we observe a linear pattern in Figures 3.4-3.6 for around the money strikes. In Section 3.5 below we provide an approximation scheme for the realised variance density based on the assumption of linear smiles.

3.4.3 Smiles for the Mixed Multi-factor Rough Bergomi Model

We conclude our analysis by introducing the correlation effect in the implied volatility smiles, by considering the mixed multi-factor rough Bergomi model (3.4). We shall

consider the following two simplified models for instantaneous variance

$$v_t = \mathcal{E}\left(\nu \int_0^t (t-s)^\alpha \mathrm{d}W_s + \eta\left(\rho \int_0^t (t-s)^\alpha \mathrm{d}W_s + \sqrt{1-\rho^2}\int_0^t (t-s)^\alpha \mathrm{d}W_s^\perp\right)\right), \tag{3.25}$$

$$v_t = \frac{1}{2}\left(\mathcal{E}\left(\nu \int_0^t (t-s)^\alpha \mathrm{d}W_s\right) + \mathcal{E}\left(\eta\left(\rho \int_0^t (t-s)^\alpha \mathrm{d}W_s + \sqrt{1-\rho^2}\int_0^t (t-s)^\alpha \mathrm{d}W_s^\perp\right)\right)\right), \tag{3.26}$$

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

On one hand, Figure 3.7 shows the implied volatility smiles corresponding to (3.25). 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 [AGM18] results still hold and we provide the asymptotic benchmark in Figure 3.7 to support our numerical scheme. On the other hand, Figure 3.8 shows the implied volatility smiles corresponding to (3.26), 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 3.9 to clearly show the effect of correlation in the convexity of the smiles.

3.5 Realised Variance Density Approximation for the Mixed Rough Bergomi Model

In light of the numerical results shown in Section 3.3 (see Figures 3.4-3.7) 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 conjecture/approximation of log-linear smiles.

Assumption 3.5.1. The implied volatility of realised variance options in the mixed rough Bergomi (3.3) model is linear in log-moneyness, and takes the following form:

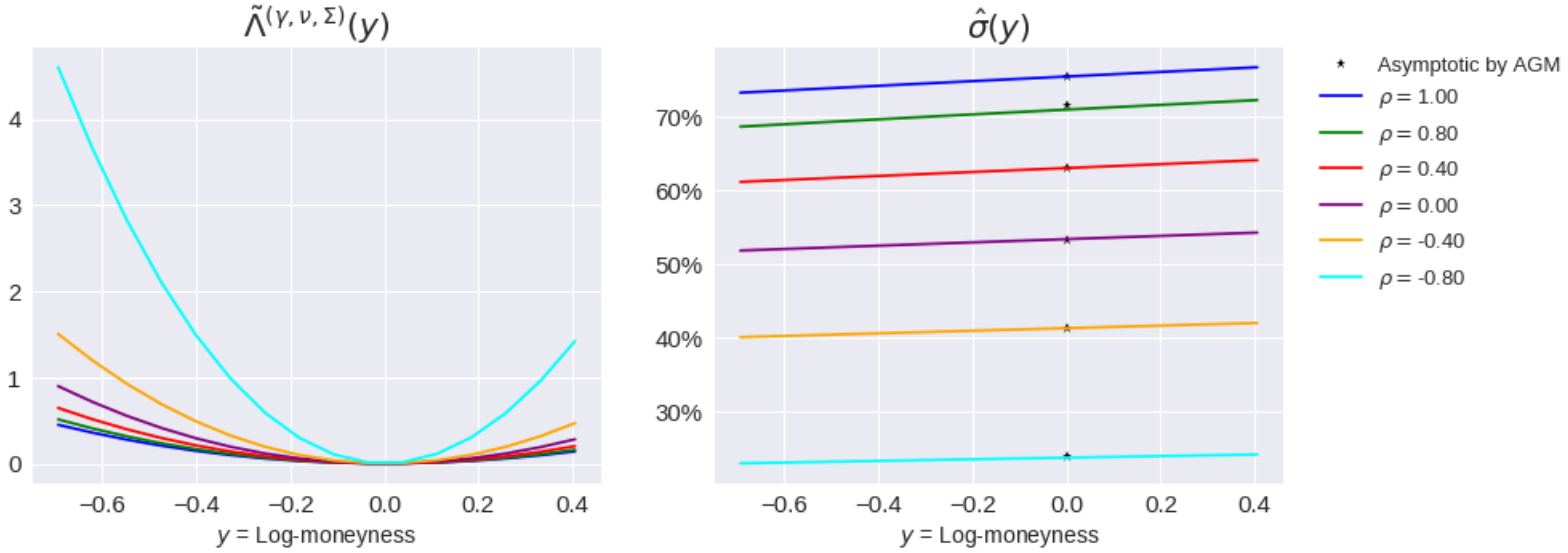

Figure 3.7: Rate function and corresponding implied volatilities for the model (3.25), with $(\alpha, \nu, \eta, p) = (-0.4, 1.0, 3.0, 5)$.

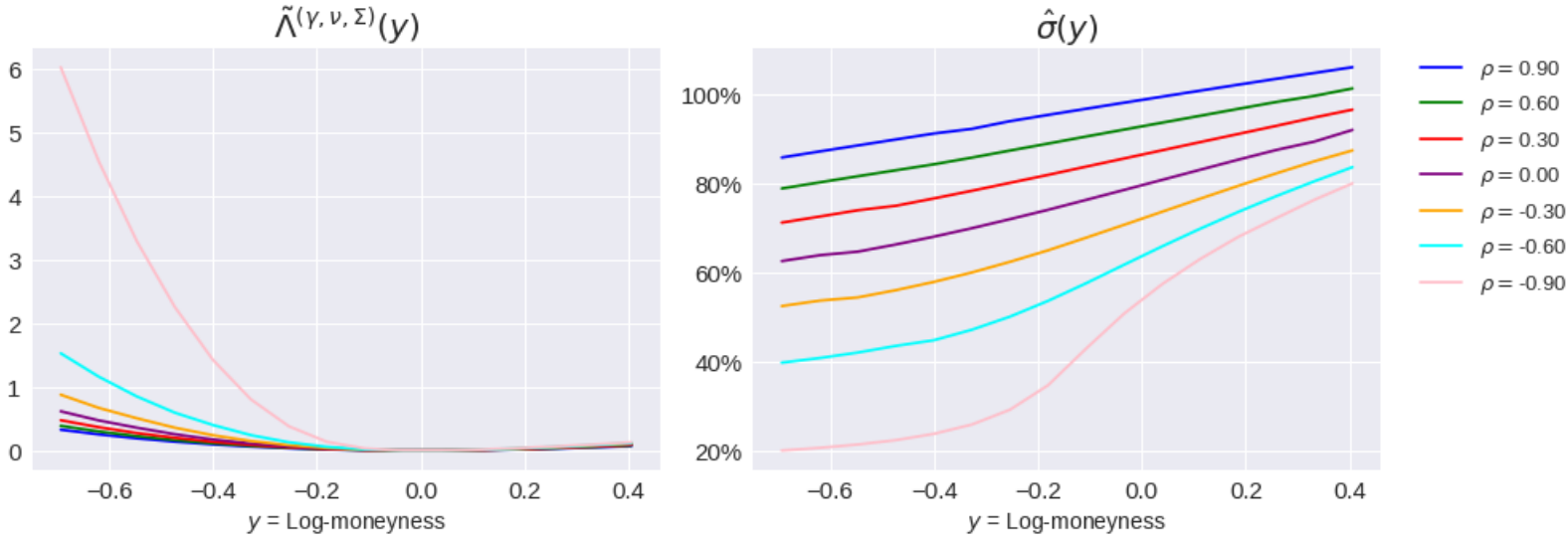

Figure 3.8: Rate function and corresponding implied volatilities for the model (3.26) with $(\alpha, \nu, \eta, p) = (-0.4, 1.0, 3.0, 5)$.

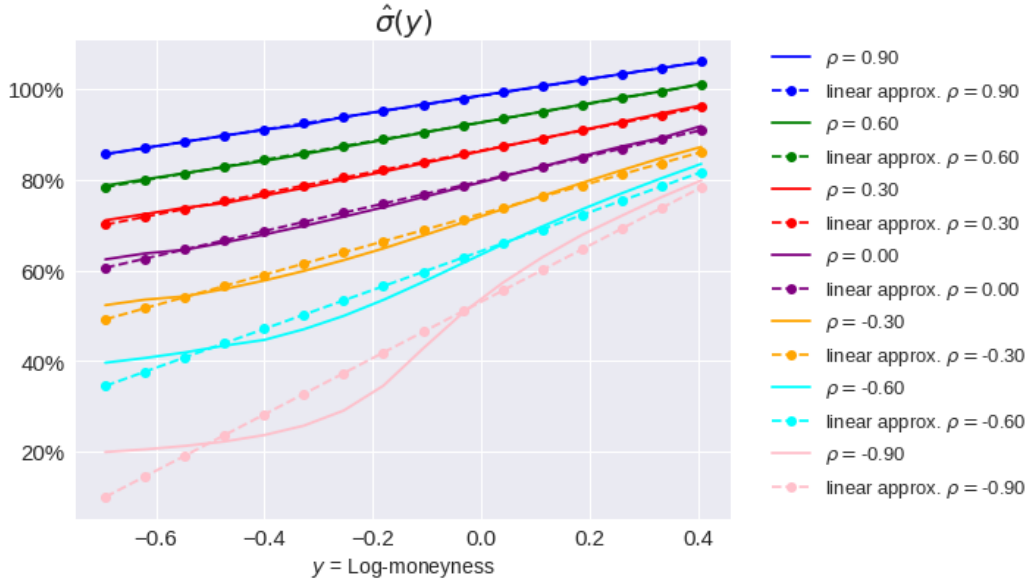

Figure 3.9: Implied volatilities and superimposed linear smiles for the model (3.26), with $(\alpha, \nu, \eta, p) = (-0.4, 1.0, 3.0, 5)$.

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

where

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

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

with

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

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

Remark 3.5.2. The values of the constants $a(\alpha, \gamma, \nu)$ and $b(\alpha, \gamma, \nu)$ in Assumption 3.5.1, which give the level and slope of the implied volatility respectively, are given in [AGM18, Example 24 and Example 27] respectively; we generalise to n factors. These results are given in terms of the Hurst parameter H; to avoid any confusion we will continue with our use of α. Recall that, by Remark 1.3.7, $\alpha = H - 1/2$.

Proposition 3.5.3. *Under Assumption 3.5.1 , the density of* $RV(v^{(\gamma,\nu)})(T)$ *is given by*

$$\psi_{RV}(x,T) = -n(d_2(x))\frac{\partial d_1(x)}{\partial x}\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}d_1(x)+1\right), \quad x \geq 0$$

where $d_1(x) = \frac{\log(v_0)-\log(x)}{\hat{\sigma}(x,T)\sqrt{T}} + \frac{1}{2}\hat{\sigma}(x,T)\sqrt{T}$, $d_2(x) = d_1(x) - \hat{\sigma}(x,T)\sqrt{T}$ *for* $x \geq 0$ *and* $n(\cdot)$ *is the standard Gaussian probability density function.*

In order to prove Proposition 3.5.3 we need the following Lemma.

Lemma 3.5.4. $v_0 n(d_1(x)) = x n(d_2(x))$

Proof. In order to prove $v_0 n(d_1(x)) = x n(d_2(x))$, we will prove the following equivalent result

$$(d_1(x))^2 - (d_2(x))^2 = 2\log\left(\frac{v_0}{x}\right).$$

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

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

□

Proof of Proposition 3.5.3. As usual let us denote

$$\mathrm{C}(K,T) := \mathbb{E}[(RV(v^{(\gamma,\nu)})(T) - K)^+].$$

The well-known Breeden-Litzenberger formula [BL78] tells us that

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

Under Assumption 3.5.1, we have that

$$\mathrm{C}(K,T) = \mathrm{C}_{BS}(v_0, \hat{\sigma}(K,T), K, T)$$

where $\mathrm{C}_{BS}(v_0,\sigma,K,T) = v_0\Phi(d_1) - K\Phi(d_2)$ is the Black-Scholes Call pricing formula with Φ the standard Gaussian cumulative distribution function. Then, differentiating C with respect to the strike gives

$$\frac{\partial \mathrm{C}(x,T)}{\partial x}\bigg|_{x=K} = v_0 n(d_1(K))\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} - xn(d_2(K))\frac{\partial d_2(x)}{\partial x}\bigg|_{x=K} - \Phi(d_2(K))$$

where

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

and

$$\frac{\partial d_2(x)}{\partial x} = \frac{\partial d_1(x)}{\partial x} - \frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{x}.$$

Using Lemma 3.5.4, we further simplify

$$\frac{\partial \mathrm{C}(K,T)}{\partial K} = v_0 n(d_1(K))\left(\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{K}\right) - \Phi(d_2(K)).$$

Differentiating again we obtain,

$$\psi_{RV}(K,T) = -v_0 n(d_1(K))\frac{a(\alpha,\gamma,\nu)T^{\alpha+1/2}}{K}\left(d_1(K)\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} + \frac{1}{K}\right) - n(d_2(K))\frac{\partial d_2(x)}{\partial x}\bigg|_{x=K}.$$

Then, by using $v_0 n(d_1(x)) = xn(d_2(x))$, we find that

$$\psi_{RV}(K,T) = -n(d_2(K))\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}\left(d_1(x)\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K} + \frac{1}{K}\right) + \frac{\partial d_2(x)}{\partial x}\bigg|_{x=K}\right),$$

which we further simplify to

$$\psi_{RV}(K,T) = -n(d_2(K))\frac{\partial d_1(x)}{\partial x}\bigg|_{x=K}\left(a(\alpha,\gamma,\nu)T^{\alpha+1/2}d_1(K)+1\right),$$

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

Remark 3.5.5. Note that Proposition 3.5.3 gives the density of $RV(v^{(\gamma,\nu)})(T)$ in closed-form. In addition, Proposition 3.5.3 can be easily used to get the density of the Arithmetic Asian option under the Black-Scholes model. This would correspond to the case $\alpha = 0$ and $\nu = \sigma > 0$ as the Black-Scholes constant volatility.

Remark 3.5.6. Assuming the density ψ_{RV} exists, we have the following volatility swap price:

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

In Figure 3.10, 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.

3.6 Asymptotic Behaviour of Options on VIX

Although options on realised variance are the most natural core modelling object for stochastic volatility models, in practice the most popular underlying for variance derivatives is the VIX. To finish the Chapter we finally turn our attention to the VIX and VIX options and study their asymptotic behaviour. For this section, we fix $\mathcal{T} := [T, T+\Delta]$. Let us now consider the following general model $(v_t)_{t\geq 0}$ for instantaneous variance:

$$v_t = \xi_0(t)\mathcal{E}\left(\int_0^t g(t,s)\mathrm{d}W_s\right). \tag{3.27}$$

Then, the VIX process is given by

$$\mathrm{VIX}_T = \sqrt{\frac{1}{\Delta}\int_T^{T+\Delta}\mathbb{E}[v_s|\mathcal{F}_T]\mathrm{d}s}.$$

Figure 3.10: Volatility Swap Monte Carlo price estimates (straight lines) and LDP based approximation (stars) for $(\eta, v_0, p) = (1.5, 0.04, 5)$; for Monte Carlo we use $200,000$ simulations and $\Delta t = \frac{1}{1008}$.

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

$$V_t^{g,T} := \int_0^T g(t,s)\mathrm{d}W_s, \tag{3.28}$$

and assume that the mapping $s \mapsto g(t,s)$ is in L^2 for all $t \in [T, T+\Delta]$ such that the stochastic integral in (3.28) is well-defined.

Proposition 3.6.1. *The VIX dynamics in model* (3.27)*, with volatility of volatility* ν*, are given by*

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

Proof. Follows directly from [JMM18, Proposition 3.1]. □

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

$$\mathcal{I}^{g,T} f(\cdot) := \int_0^T g(\cdot, s) f(s) \mathrm{d}s, \quad \mathscr{H}^{g,T} := \{\mathcal{I}^{g,T} f : f \in \mathrm{L}^2\}, \tag{3.29}$$

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

Proposition 3.6.2. *Assume that there exists $h \in \mathrm{L}^2$ such that $\int_0^\varepsilon |h(s)| \mathrm{d}s < +\infty$ for some $\varepsilon > 0$ and $g(t, \cdot) = h(t - \cdot)$ for any $t \in [T, T+\Delta]$. Then, the space $\mathscr{H}^{g,T}$ is the reproducing kernel Hilbert space for the process $(V_t^{g,T})_{t \in [T, T+\Delta]}$.*

Proof of Proposition 3.6.2. The proof of Proposition 3.6.2, which is similar to the proofs given in Chapter 1, is made up of three parts. The first part is to prove that $\left(\mathscr{H}^{g,T}, \langle \cdot, \cdot \rangle_{\mathscr{H}^{g,T}}\right)$ is a separable Hilbert space. Clearly $\mathcal{I}^{g,T}$ is surjective on $\mathscr{H}^{g,T}$. Now take $f_1, f_2 \in \mathrm{L}^2$ such that $\mathcal{I}^{g,T} f_1 = \mathcal{I}^{g,T} f_2$. For any $t \in [T, T+\Delta]$ it follows that $\int_0^T g(t, s)[f_1(s) - f_2(s)] \mathrm{d}s = 0$; applying the Titchmarsh convolution Theorem then implies that $f_1 = f_2$ almost everywhere and so $\mathcal{I}^{g,T} : \mathrm{L}^2 \to \mathscr{H}^{g,T}$ is a bijection. $\mathcal{I}^{g,T}$ is a linear operator, and therefore $\langle \cdot, \cdot \rangle_{\mathscr{H}^{g,T}}$ is indeed an inner product; hence $\left(\mathscr{H}^{g,T}, \langle \cdot, \cdot \rangle_{\mathscr{H}^{g,T}}\right)$ is a real inner product space. Since L^2 is a complete (Hilbert) space, there exists a function $\widetilde{f} \in \mathrm{L}^2$ such that the sequence $\{f_n\}_{n \in \mathbb{N}}$ converges to $\widetilde{f}$. Assume for a contradiction that $f \neq \widetilde{f}$, then, since $\mathcal{I}^{g,T}$ is a bijection, the triangle inequality yields

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

which converges to zero as n tends to infinity. Therefore $f = \widetilde{f}$, $\mathcal{I}^{g,T} f \in \mathscr{H}^{g,T}$ and $\mathscr{H}^{g,T}$ is complete, hence a real Hilbert space. Since L^2 is separable with countable orthonormal basis $\{\phi_n\}_{n \in \mathbb{N}}$, then $\{\mathcal{I}^{g,T} \phi_n\}_{n \in \mathbb{N}}$ is an orthonormal basis for $\mathscr{H}^{g,T}$, which is then separable.

The second part of the proof is to show that there exists a dense embedding $\iota : \mathscr{H}^{g,T} \to \mathcal{C}[T, T+\Delta]$. Since there exists $h \in \mathrm{L}^2$ such that $\int_0^\varepsilon |h(s)| \mathrm{d}s$ for all $\varepsilon > 0$ and $g(t, \cdot) = h(t - \cdot)$ for any $t \in [T, T+\Delta]$, we can apply [Che08, Lemma 2.1], which tells us that $\mathscr{H}^{g,T}$ is dense in $\mathcal{C}[T, T+\Delta]$ and so we choose the embedding to be the inclusion map. Finally we must prove that every $f^* \in \mathcal{C}[T, T+\Delta]$ is Gaussian on $\mathcal{C}[T, T+\Delta]$, with

variance $\|\iota^* f^*\|_{\mathscr{H}^{g,T^*}}$, where ι^* is the dual of ι. Take $f^* \in \mathcal{C}[T, T+\Delta]^*$, then Remark 1.4.14 implies that ι^* admits an isometric embedding $\bar{\iota}^*$ such that

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

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

Theorem 3.6.3. *For any $\gamma > 0$, the sequence of stochastic processes $(\varepsilon^{\gamma/2} V^{g,T})_{\varepsilon>0}$ satisfies a large deviations principle on $\mathcal{C}[T, T+\Delta]$ with speed $\varepsilon^{-\gamma}$ and rate function Λ^V, defined as*

$$\Lambda^V(\mathrm{x}) := \begin{cases} \frac{1}{2}\|\mathrm{x}\|^2_{\mathscr{H}^{g,T}}, & \text{if } \mathrm{x} \in \mathscr{H}^{g,T}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.30}$$

Proof. Direct application of Theorem 1.5.14. □

Remark 3.6.4. We now introduce a Borel subset of $\mathcal{C}[T, T+\Delta]$, defined as

$$A := \{\in \mathcal{C}[T, T+\Delta] : h(x) \ \geq \ 1 \text{ for all } x \in \mathbb{R}\}.$$

Then, by a simple application of Theorem 3.6.3 and using that the rate function Λ^V is continuous on A, we can then obtain the following tail behaviour of the process $V^{g,T}$:

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{\gamma} \log \mathbb{P}\left(V_t^{g,T} \geq \frac{1}{\varepsilon^{\gamma/2}} \right) = -\inf_{h \in A} \Lambda^V(h), \tag{3.31}$$

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

Remark 3.6.5. Let us again fix the kernel g as the rough Bergomi kernel and denote the corresponding reproducing kernel Hilbert space by $\mathscr{H}^{\eta,\alpha,T}$ (to emphasise the parameter values) and the corresponding process $V^{g,T}$ as $V^{\eta,\alpha,T}$. If $\mathrm{x} \in \mathscr{H}^{\eta,\alpha,T}$ it follows that there exists $f \in \mathrm{L}^2$ such that $\mathrm{x}(t) = \int_0^T \eta\sqrt{2\alpha+1}(t-s)^{\alpha} f(s)\mathrm{d}s$ for all $t \in [T, T+\Delta]$. Clearly, it follows that $\mathrm{x} \in \mathscr{H}^{a\eta,\alpha,T}$ for any $a > 0$, as $f \in \mathrm{L}^2$ implies that $\frac{1}{a} f =: f_a \in \mathrm{L}^2$. We can compute the norm of x in each of these spaces to arrive at the following isometry:

$$\|\mathrm{x}\|^2_{\mathscr{H}^{a\eta,\alpha,T}} = \|f_a\|^2_{\mathrm{L}^2} = \frac{1}{a^2} \int_0^T f^2(s)\mathrm{d}s = \frac{1}{a^2}\|\mathrm{x}\|^2_{\mathscr{H}^{\eta,\alpha,T}}. \tag{3.32}$$

We may now amalgamate (3.30), (3.31), and (3.32) to arrive at the following statement, which tells us how the large strike behaviour scales with the vol-of-vol parameter η in the rough Bergomi model:

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{\gamma} \log P\left(V_t^{a\eta,\alpha,T} \geq \frac{1}{\varepsilon^{\gamma/2}}\right) = \lim_{\varepsilon \downarrow 0} \varepsilon^{\gamma} \log \left(P\left(V_t^{\eta,\alpha,T} \geq \frac{1}{\varepsilon^{\gamma/2}}\right)^{1/a^2}\right). \tag{3.33}$$

Indeed, (3.33) tells us precisely how increasing the vol-of-vol parameter η multiplicatively by a factor a in the rough Bergomi model increases the probability that the associated process $V^{g,T}$ will exceed a certain level.

Before stating the main theorem of this Section, we first define the following rescaled process:

$$V_t^{g,T,\varepsilon} := \varepsilon^{\gamma/2} V_t^{g,T}, \quad \widetilde{V}_t^{g,T,\varepsilon} := V_t^{g,T,\varepsilon} - \frac{\varepsilon^{\gamma}}{2} \int_0^t g^2(t,u)\mathrm{d}u + \varepsilon^{\gamma/2}, \tag{3.34}$$

for $\varepsilon \in [0,1]$, $t \in [T, T+\Delta]$. We also define the following $\mathcal{C}\left([T,T+\Delta]\times[0,1]\right)$ operators φ_{1,ξ_0} and φ_2, which map to $\mathcal{C}\left([T,T+\Delta]\times[0,1]\right)$ and $\mathcal{C}[0,1]$ respectively, as

$$(\varphi_{1,\xi_0} f)(s,\varepsilon) := \xi_0(s)\exp(f(s,\varepsilon)), \quad (\varphi_2 g)(\varepsilon) := \frac{1}{\Delta}\int_T^{T+\Delta} g(s,\varepsilon)\mathrm{d}s. \tag{3.35}$$

Note that in the definition of φ_{1,ξ_0} in (3.35) we assume ξ_0 to be a continuous, single valued, and strictly positive function on $[T, T+\Delta]$. This then implies that for every $s \in [T, T+\Delta]$, the map $\varepsilon \mapsto (\varphi_{1,\xi_0} f)(s,\varepsilon)$ is a bijection and hence has an inverse, denoted by φ^{-1}_{1,ξ_0}, which is defined as $(\varphi^{-1}_{1,\xi_0} f)(s,\varepsilon) := \log\left(\frac{f(s,\varepsilon)}{\xi_0(s)}\right)$.

Theorem 3.6.6. *For any $\gamma > 0$, the sequence of rescaled VIX processes $\left(e^{\varepsilon^{\gamma/2}}\, VIX_{T,\varepsilon^{\gamma/2}}\right)_{\varepsilon\in[0,1]}$ satisfies a pathwise large deviations principle on $\mathcal{C}[0,1]$ with speed $\varepsilon^{-\gamma}$ and rate function*

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

Proof of Theorem 3.6.6. First we recall $\widetilde{V}_t^{g,T,\varepsilon} := V_t^{g,T,\varepsilon} - \frac{\varepsilon^{\gamma}}{2}\int_0^t g^2(t,u)\mathrm{d}u + \varepsilon^{\gamma/2}$. We begin the proof by showing that the sequence of processes $(V^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ and $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ are exponentially equivalent [DZ10, Definition 4.2.10]. As $g(t,\cdot) \in \mathrm{L}^2$ for $t \in [T, T+\Delta]$,

for each $\delta > 0$ there exists $\varepsilon_* > 0$ such that $\sup_{t\in[T,T+\Delta]} \left|\varepsilon_*^{\gamma/2} - \frac{\varepsilon_*^\gamma}{2}\int_0^T g^2(t,u)\mathrm{d}u\right| \leq \delta$. Therefore, for the $\mathcal{C}[T,T+\Delta]$ norm $\|\cdot\|_\infty$ we have that for all $\varepsilon_* > \varepsilon > 0$,

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

Therefore $\limsup_{\varepsilon\downarrow 0}\varepsilon^\gamma \log\mathbb{P}\left(\left\|V^{g,T,\varepsilon} - \widetilde{V}^{g,T,\varepsilon}\right\|_\infty > \delta\right) = -\infty$, and so the two sequences of processes $(V^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ and $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ are exponentially equivalent; applying [DZ10, Theorem 4.2.13] then yields that $(\widetilde{V}^{g,T,\varepsilon})_{\varepsilon\in[0,1]}$ satisfies a large deviations principle on $\mathcal{C}[T,T+\Delta]$ with speed $\varepsilon^{-\gamma}$ and rate function Λ^V.

We now prove that the operators φ_{1,ξ_0} and φ_2 are continuous with respect to the $\mathcal{C}([T,T+\Delta]\times[0,1])$ and $\mathcal{C}[0,1]$ $\|\cdot\|_\infty$ norms respectively. The proofs are very simple, and are included for completeness. First let us take a small perturbation $\delta^f \in \mathcal{C}([T,T+\Delta]\times[0,1])$:

$$\begin{aligned}\left\|\varphi_{1,\xi_0}(f+\delta^f) - \varphi_{1,\xi_0}(f)\right\|_\infty &= \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}\left|\xi_0(s)e^{f(s,\varepsilon)}\left(e^{\delta^f(s,\varepsilon)}-1\right)\right| \\ &\leq \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|\xi_0(s)| \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|e^{f(s,\varepsilon)}| \sup_{\substack{\varepsilon\in[0,1]\\ s\in[T,T+\Delta]}}|e^{\delta^f(s,\varepsilon)}-1|.\end{aligned}$$

Since ξ_0 is continuous on $[T,T+\Delta]$ and f is continuous on $[T,T+\Delta]\times[0,1]$, they are both bounded. Clearly $e^{\delta^f(s,\varepsilon)}-1$ tends to zero as δ^f tends to zero and hence the operator φ_{1,ξ_0} is continuous. Now take a small perturbation $\delta^f \in \mathcal{C}([T,T+\Delta]\times[0,1])$:

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

where $M := \sup_{\varepsilon\in[0,1]}\delta^f(s,\varepsilon)$. Clearly M tends to zero as δ^f tends to zero, thus the operator φ_2 is also continuous.

For every $s\in[T,T+\Delta]$ we have the following: by an application of Proposition 1.5.13 and using the fact that $\varepsilon\mapsto(\varphi_{1,\xi_0}f)(s,\varepsilon)$ is a bijection for all $f\in\mathcal{C}[T,T+\Delta]$ it follows that the sequence of stochastic processes $\left(\left(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}\right)(s,\varepsilon)\right)_{\varepsilon\in[0,1]}$ satisfies a large deviations principle on $\mathcal{C}[0,1]$ as ε tends to zero with speed $\varepsilon^{-\gamma}$ and rate function

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

A second application of Proposition 1.5.13 then yields that the sequence of stochastic processes $\Big((\varphi_2(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\Big)_{\varepsilon\in[0,1]}$ satisfies a large deviations principle on $\mathcal{C}[0,1]$ with speed $\varepsilon^{-\gamma}$ and rate function $\Lambda^{\mathrm{VIX}}(\mathrm{x}) = \inf_{s\in[T,T+\Delta]}\{\Lambda^V\left((\varphi_{1,\xi_0}\mathrm{y})^{-1}(s,\cdot)\right) : \mathrm{x}(\cdot) = (\varphi_2\mathrm{y})(\cdot)\}$. By definition, the sequence of processes $\Big((\varphi_2(\varphi_{1,\xi_0}\widetilde{V}_s^{g,T,\varepsilon}))(\varepsilon)\Big)_{\varepsilon\in[0,1]}$ is almost surely equal to the rescaled VIX processes $(e^{\varepsilon^{\gamma/2}}\mathrm{VIX}_{T,\varepsilon^{\gamma/2}})_{\varepsilon\in[0,1]}$ and hence satisfies the same large deviations principle. □

Remark 3.6.7. Using Theorem 3.6.6, we can deduce the small-noise, large strike behaviour of VIX options. Indeed, for the Borel subset A of $\mathcal{C}[T,T+\Delta]$ introduced in Remark 3.6.4 we have that, for any $\gamma > 0$,

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

Chapter 4

Calibrating Rough Volatility Models: a Convolutional Neural Networks Approach

The appeal of rough volatility models- their ability to generate power law at-the-money volatility skew and to accurately forecast future volatility- is clear. Calibration, however, is highly nontrivial and is perhaps the final stumbling block that has prevented rough volatility models being adopted by practitioners in the quantitative finance industry. The aim of this Chapter is to provide a robust, fast, and accurate means of calibrating the rough Bergomi model[1].

Calibrated parameters for a given model are typically found to be those that minimise the error between observed option prices in the market, and theoretical option prices given by the model. In the case of rough volatility models, which are non-Markovian, theoretical option prices tend to be computed via Monte Carlo methods, and therefore the computational cost of calibration via minimisation becomes too high for practical use. Instead, in this Chapter, we propose a novel calibration scheme using convolutional neural networks, which are known to give fast and accurate estimations in a wide range of applications.

[1]This Chapter includes results from article [Sto20].

We begin Chapter 4 with an introduction to machine learning and convolution neural networks (referred to as CNNs from here) using [BCG16, Chapter 5], [KNTY18], and [Yeg09, Chapter 1] as our guides. We establish the necessary machine learning terminology, and describe the structure and mechanics of a CNN.

4.1 An Introduction to Machine Learning and Convolution Neural Networks

4.1.1 Machine Learning Fundamentals

The supervised machine learning paradigm is, in its most simple form, the following:

> Given a set of input data $\{\mathbf{x}_i\}_{i\in\{1,...,N\}}$ with corresponding output data $\{\mathbf{y}_i\}_{i\in\{1,...,N\}}$ can we find a function g such that $g(\mathbf{x}_i) \approx \mathbf{y}_i$ for $i \in \{1, ..., N\}$?

Indeed, any supervised machine learning algorithm should answer the above question, albeit when the meaning of $g(\mathbf{x}_i) \approx \mathbf{y}_i$ is specified in a precise and meaningful way. The predictive performance of a machine learning algorithm is measured in terms of a loss function $\ell(\cdot, \cdot)$, which associates a loss $\ell(g(\mathbf{x}_i), \mathbf{y}_i)$ to the prediction $g(\mathbf{x}_i)$ of the true value $\mathbf{y}_i$. Clearly the loss function should satisfy $\ell(x, x) = 0$ for all x. In general, supervised learning algorithms aim to find g that minimises the expected loss $\frac{1}{N}\sum_{i=1}^{N} \ell(g(\mathbf{x}_i), \mathbf{y}_i)$. As well as minimising the expected loss, the function g should predict well on unseen data; that is, it should also satisfy $g(\mathbf{x}_j) \approx \mathbf{y}_j$ for $j \in \{N + 1, ..., M\}$ or equivalently produce a small expected loss $\frac{1}{M-N}\sum_{i=N+1}^{M} \ell(g(\mathbf{x}_i), \mathbf{y}_i)$.

Let us now introduce some relevant machine learning terminology, which will be used throughout the Chapter.

- **Classification problem:** A machine learning problem where the output data $\{\mathbf{y}_i\}_{i\in\{1,...,N\}}$ take values in some countable set.
- **Epoch:** The number of iterations executed by a machine learning algorithm on the training set.

- **Hyperparameter:** a variable that remains unchanged during the training process.
- **Overfitting:** the scenario where a machine learning algorithm learns noise in the training set, rather than the desired underlying signal. An overfitted model will perform poorly on unseen data, i.e. have a high test error compared to the training error.
- **Regression problem:** A machine learning problem where the output data $\{\mathbf{y}_i\}_{i\in\{1,\dots,N\}}$ take continuous values.
- **Test set:** the set $(\{\mathbf{x}_j\},\{\mathbf{y}_j\})_{j\in\{N+1,\dots,M\}}$, which is used to measure the predictive power of the machine learning algorithm.
- **Test loss:** the value $\frac{1}{M-N}\sum_{i=N+1}^{M}\ell(g(\mathbf{x}_i),\mathbf{y}_i)$ for a given loss function ℓ.
- **Training:** the process of finding the function g. Also referred to as the learning process.
- **Training set:** the set $(\{\mathbf{x}_i\},\{\mathbf{y}_i\})_{i\in\{1,\dots,N\}}$, which is used to find the function g.
- **Training loss:** the value $\frac{1}{N}\sum_{i=1}^{N}\ell(g(\mathbf{x}_i),\mathbf{y}_i)$ for a given loss function ℓ.
- **Validation set:** a subset of the training set, used during training to evaluate the model and test for overfitting.
- **Validation loss:** analogous definition to training and test loss, instead applied to the validation set.

Remark 4.1.1. In this introductory Section we have purely focused on supervised learning, where the input data $\{\mathbf{x}_i\}_{i\in\{1,\dots,N\}}$ has a corresponding output data $\{\mathbf{y}_i\}_{i\in\{1,\dots,N\}}$. The case where the output data is unknown is referred to as unsupervised learning.

4.1.2 Convolutions Neural Network Structure and Mechanics

Before describing the structure of a CNN we must first describe an artificial neural network. An artificial neural network is a biologically inspired system of interconnected

processing units, where each processing unit is called a layer. Inputs to each layer, apart from the first layer, are outputs from previous layers. A layer is composed of a number of nodes, and each node in a given layer is connected to the nodes in a subsequent layer, thus forming a network; each edge in this network has a weight associated to it. The first processing unit is called the input layer, and the final processing unit is the output layer. The processing unit or units between the input layer and output layer are referred to as hidden layers; typically artificial neural networks have more than one hidden layer. Figure 4.1 below illustrates the structure of a simple artificial neural network[2]. A formal, mathematical definition [BGTW19, Definition 4.1] of a neural network is given below in Definition 4.1.2.

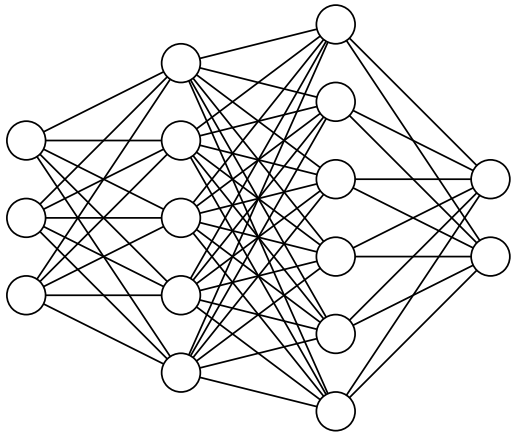

Figure 4.1: An example of a neural network, with two hidden layers. The input layer has three nodes; the hidden layers have five and six nodes respectively; the output layer has two nodes.

[2] The image was drawn using Python, and the code used is available on Github: https://gist.github.com/craffel/2d727968c3aaebd10359.

Definition 4.1.2. Let $L \in \mathbb{N}$ denote the number of layers in the neural network. The dimension of each hidden layer is denoted by $N_1, ..., N_{L-1} \in \mathbb{N}$, and the respective dimension of the input and output layer is denoted by $N_0, N_L \in \mathbb{N}$. For $A^\ell \in \mathbb{R}^{N_{\ell-1} \times N_\ell}$ and $b \in \mathbb{R}^{N_\ell}$ let the affine function $W_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$ be defined as $W_\ell(x) := A^\ell x + b^\ell$, for $\ell = 1, ..., L$. The entries of matrix A^ℓ are the weights connecting each node in layer $\ell - 1$ to layer ℓ. The neural network, with (non-linear) activation function σ, is then the function $\mathscr{N} : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ defined as the composition

$$\mathscr{N}_L(x) := W_L \circ (\sigma \circ W_{L-1}) \circ ... \circ (\sigma \circ W_1)(x). \tag{4.1}$$

The learning process of an artificial neural network essentially boils down to finding the optimal weights in each matrix A^ℓ that minimise a given loss function, which depends on the task at hand i.e. if it is a classification problem or regression problem being solved. These optimal weights are then used to formulate the function g, which can be used for predictions on the test set.

Remark 4.1.3. We take this opportunity to emphasise that the activation function σ should non-linear, otherwise the neural network is simply a composition of linear functions, and hence is itself linear.

Remark 4.1.4. Deep neural networks, those with more than one hidden layer such that $L \geq 3$, typically have superior predictive power to shallow neural networks, which only have one hidden layer such that $L = 2$. The predictive power of deep neural networks lies in the composition of many non-linear functions, which allows the neural network $\mathscr{N}_L$ to belong to a very rich space of functions. In comparison, a shallow neural network $\mathscr{N}_2$ belongs to a space of linear combinations of single non-linear functions, all of which have the same dimension. It follows, then, that adding layers is more significant for improving predictive power than increasing the number of nodes in a shallow neural network. Consequently, we can think of composition of non-linear functions as being more "powerful" than addition of non-linear functions. There is much empirical evidence to support what is written above; a good overview is given by Mhaskar, Liao, and Poggio [MLP17].

CNNs are a class of artificial neural networks, where the hidden layers can be grouped into different classes according to their purpose; one such class of hidden layer is the

eponymous convolutional layer. Below we describe the classes of hidden layers used in our CNN. Of course, this list is not exhaustive, and there exist many classes of hidden layers that we omit for means of brevity. Note also that we describe a CNN in the context of the problem we are trying to solve, where the input data are one dimensional vectors. CNNs can of course also be used on higher dimensional input data, but the fundamental structure and different roles of the hidden layers do not change.

- **Convolutional Layer:** In deep learning, the convolution operation is a method used to assign relative value to entries of input data, in our case one dimensional vectors of time series data, while simultaneously preserving spatial relationships between individual entries of input data. For a given kernel size k and an input vector of length m, the convolution operation takes entries $1, ..., k$ of the input vector and multiplies by the kernel element-wise, whose length is k. The sum of the entries of the resulting vector are then the first entry of the feature map. This operation is iterated $m + 1 - k$ times, thus incorporating every entry in the input data vector into the convolution operation. The output of the convolutional layer is the feature map.

  For example, let $(1, 2, 1, 0, 0, 3)$ be our input vector, and $(1, 0, 1)$ be our kernel; here the kernel size is 3. The first iteration of the convolution operation involves taking the element-wise multiple of $(1, 2, 1)$ and $(1, 0, 1)$: $(1, 0, 1)$ is produced and the sum, equal to 2, is computed. This is the first entry of the feature map. The resulting feature map in this example is then $(2, 2, 1, 3)$.

  Clearly, the centre of each kernel cannot overlap with the first and final entry of the input vector. Zero-padding, sometimes referred to as same-padding, preserves the dimensions of input vectors and allows more layers to be applied in the CNN: zero-padding is simply the extension of the input vector and the setting of the first and final entries as 0, while leaving the other entries unchanged. In our example, the input vector becomes $(0, 1, 2, 1, 0, 0, 3, 0)$ after zero padding.

- **Activation Layer:** The activation layer is a non-linear function σ that is applied to the output of the convolutional layer i.e. the feature map; the purpose of the activation layer is indeed to introduce non-linearity into the CNN. Examples

of activation functions include the sigmoid function and the hyperbolic tangent function. In our CNN we use the 'LeakyReLU' activation function, defined as

$$f_\alpha(x) := \begin{cases} x, & \text{if } x > 0, \\ \alpha x, & \text{otherwise.} \end{cases}$$

The LeakyReLU activation function allows a small positive gradient when the unit is inactive.

- **Max Pooling Layer:** For a given pooling size p, the max pooling layer returns a vector whose entries are the maximum among the neighbouring p entries in the feature map. For example, for feature map $(1, 3, 8, 2, 1, 0, 0, 4, 6, 1)$ and $p = 3$ the max pooling output is $(8, 8, 8, 8, 8, 2, 4, 6, 6, 6)$.

  Other pooling techniques apply the same idea, but use different functions to evaluate the neighbouring p entries in the feature map. Examples include average pooling, and L2-norm pooling, which in fact uses the Euclidean norm in mathematical nomenclature.

- **Dropout Layer:** Dropout is a well-known technique incorporated into CNNs in order to prevent overfitting. Without the addition of a dropout layer, each node in a given layer is connected to each node in the subsequent layer; dropout temporarily removes nodes from different layers in the network. The removal of nodes is random and determined by the dropout rate d, which gives the proportion of nodes to be temporarily dropped. Note that dropout is only implemented during training; during testing the weights of each node are multiplied by the dropout rate d.

  An excellent overview of the technique is given by Hinton, Krizhevsky, Salakhutdinov, Srivastava, and Sutskever [HKSSS14]. The authors provide an extensive study to show how predictive performance of CNNs, in a number of different settings, is improved using dropout.

- **Dense Layer:** Also referred to as the fully connected layer, each node in the input layer is connected to each node in the output layer as the name suggests. After being processed by the convolutional, activation, pooling, and dropout layers,

the extracted features are then mapped to the final outputs via a subset of the dense layer, an activation function is then applied subsequently. This activation function is chosen specifically for the task that the CNN is required to execute, i.e. binary/multi-class classification, or regression to output a continuous value. The final output from the dense layer has the same number of nodes as the number of classes in the output data.

Remark 4.1.5. The number of layers L, the nodes per layer N_L, the kernel size k, the pooling size p, and the dropout rate d are all examples of CNN hyperparameters.

Having described the structure, we now focus on the mechanics of training the CNN. As mentioned previously, training a CNN corresponds to finding weights in the fully connected layer, and kernels in the convolutional layers, that minimise a specific loss function. Forward propagation is the name for the process by which input data is translated to an output through layers of the CNN; it is used to give the value of the loss function, and therefore the predictive power of the CNN, for certain weights and kernels. The back-propagation algorithm is used to compute the gradient of the loss function from the error values of the loss function computed via forward propagation; weights and kernels, depending on the values of the loss function, are then updated iteratively. In the case of our CNN the Adam optimizer is used. More details on the back-propagation algorithm and the Adam optimizer can be found in [BCG16, Section 6.5, pages 200-219] and [BK17] respectively.

For the classification problem we solve below, we employ the categorical cross entropy function as the loss function; for the regression problem, the mean squared error function is used.

4.2 Classification Methodology and Results

The aim of the following two Sections is to investigate whether a CNN can learn the Hölder regularity from rough Bergomi and fractional Brownian motion (fBm) sample paths; in other words, we seek to establish if the CNN can learn 'roughness'. CNNs are known to be very powerful machine learning tools with a vast array of applications

including (but of course not limited to) image classification [HKS12], [BCGL97], [SZ15]; speech recognition [DHK13], [ADY13]; and for self-driving cars [CKSX15], [IJKW17].

In the field of mathematical finance, neural networks have recently become an area of interest for research. Bayer and Stemper [BS18] used neural networks to learn implied volatility surfaces; the network is then used as part of a wider calibration scheme for options pricing. Similarly, Horvath, Muguruza, and Thomas [HMT19] use neural networks to approximate the pricing function of vanilla options; the trained network is then used to reduce calibration time. Buehler, Gonon, Teichmann, and Wood [BGTW19] use deep reinforcement learning techniques to establish a hedging framework for derivative portfolios. Jacquier and Oumgari [JO19] use deep learning techniques to approximate the solution of the path dependent pricing PDE for a rough local-stochastic volatility model. To the best of our knowledge, however, this is the first piece of research to explore the use of CNNs to learn the Hölder regularity of a given stochastic process.

We hope to establish a robust means for calibrating rough volatility models; indeed, once the CNN has been trained we want it to perform well when making predictions on unseen data. We begin with a 'toy' classification experiment: we use CNNs to solve the classification problem for a discrete set of H values. While the problem setting may be unrealistic, and unsuitable for practical applications to calibration, it allows us to establish that CNNs are indeed a suitable tool for learning Hölder regularity. The methodology is outlined is Section 4.2.1 and the results are given in Section 4.2.2. We then move on to the regression problem, giving the methodology in Section 4.3.1 and the results in Section 4.3.2, and show that the same CNN architecture can additionally be used to solve that regression problem.

If the CNN is able to accurately learn α in rough Bergomi and H in a fBm, then we are making an important contribution to the field of mathematical finance, applied probability, and machine learning.

4.2.1 Classification Methodology

We use trajectories of fractional Brownian motion and of the normalised log volatility process $(\log(v_t/v_0))_{t\geq 0}$ of the rough Bergomi model as our input data, with the corre-

sponding Hölder regularity as the output data. We use MATLAB's `wfbm(H,L)` function to generate the fractional Brownian motion trajectories where `H` is the Hurst parameter H and `L` is the length of the fBm vector. We set `L`$= 500$, and let the Hurst parameter H take values in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. To simulate rough Bergomi we use Cholesky decomposition; this very well-known simulation technique is recommended because the resulting sample paths have the exact distribution, rather than an approximate distribution, of the normalised log volatility process of the rough Bergomi model. The code used is publicly available on Github[3].

By Proposition 1.3.6, the Hölder regularity of the normalised log volatility process is independent of the value of η; the same proposition proves that the process Z, defined in (1.9), and the normalised log volatility process have the same Hölder regularity. We therefore set $\eta = 1$ in the model (1.11) above when generating the sample paths, for simplicity, and we also ignore the deterministic drift term $t^{2\alpha+1}$.

Every member of the resulting input data set has the following form: a vector $\mathbf{x}_i$ containing 500 entries, which is a rough Bergomi or fBm sample path with a given α or H, and a label $\mathbf{y}_i \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, which corresponds to that given α or H used to generate $\mathbf{x}_i$. Recall that $H = \alpha + 1/2$ in the rough Bergomi model. We then split the input data into training and test sets; we subsequently create a validation set from part of the test set. The sizes of each training/test/validation set for the fractional Brownian motion data and rough Bergomi data are given in Tables 4.1 and 4.5 respectively.

Data set	Number of samples
Training set	20,149
Test set	10,795
Validation set	5,038

Table 4.1: Fractional Brownian motion input data size description.

We use a one dimensional CNN, since our input $\mathbf{x}_i$ are vectors, with three layers of kernels, where the kernel size for each kernel is 20 and each layer is succeeded by the Leaky ReLU activation function with alpha = 0.1; we add max pooling layers, each

[3]https://github.com/amuguruza/RoughFCLT/blob/master/rDonsker.ipynb

Data set	Number of samples
Training set	10,080
Test set	5,400
Validation set	2,520

Table 4.2: Rough Bergomi input data size description.

of size 3, and dropout layers between each layer of kernels. The values for kernel size, max pooling size, dropout rate, and rate for the Leaky ReLU activation function were chosen because they achieved the lowest error among all values tested. By no means are these hyperparameters chosen in the most optimal way but are sufficiently optimal to achieve accurate predictions, as will be shown below. We clarify the specific structure of the hidden layers of the CNN below:

- the first layer, with 32 kernels;
- max pooling layer;
- a dropout layer, with rate = 0.25;
- the second layer, with 64 kernels;
- max pooling layer;
- a dropout layer, with rate = 0.25;
- the third layer, with 128 kernels;
- max pooling layer;
- a dropout layer, with rate = 0.4;
- a dense layer with 128 units;
- a dropout layer, with rate = 0.3.

The reason for choosing this structure for our CNN, which is fairly standard for image classification in the computer science discipline, is two fold. The first is that, heuristically, images are classified by considering the values of each entry of an image matrix

together with the neighbouring entries; more emphasis is put on those neighbouring values than entries far from the entry being considered. To study the Hölder regularity of the sample path of a stochastic process, the values of neighbouring points of each entry in a sample path vector will provide the most information about the Hölder regularity of that process. For this reason we employ an image processing-type architecture. The second is to avoid the task of choosing optimal hyperparameters for the number of filters in each layer.

4.2.2 Classification Results

We train the CNN, with the architecture described above, setting `batch size`= 64 as is fairly standard, and `epochs`= 30 as these values gave the highest test accuracy and lowest categorical cross entropy loss (our chosen loss function for classification). Note that we train on the fractional Brownian motion input data and rough Bergomi input data separately. The test loss and test accuracy results for fractional Brownian motion and rough Bergomi are given in Tables 4.3 and 4.4; loss plots are given in Appendix B.1.

Test Loss	Test Accuracy
0.613	0.750

Table 4.3: Fractional Brownian motion classification results.

Test Loss	Test Accuracy
0.247	0.873

Table 4.4: Rough Bergomi classification test results.

The results presented in this Section are indeed very promising: we see that both H and α can be predicted to a high degree of accuracy using CNNs. In fact, we conjecture that a higher test accuracy could be attained if we optimised the hyperparameter values in our CNN. Additionally, invoking a k-fold cross validation could further increase the test accuracy.

4.3 Regression Methodology and Results

We now move on to solving the more realistic regression problem; the CNN will now output a continuous value for the Hölder exponent, rather than classifying the output as a member of a discrete set. We emphasise that the architecture of the CNN remains unchanged from that given above.

4.3.1 Regression Methodology

Since we are aiming to use this trained CNN to calibrate the rough Bergomi model to realised volatility market data, we focus on using simulated rough Bergomi sample paths as our input data. Once again, the simulation is done using Cholesky decomposition; we first set $\eta = 1$ and ignore the deterministic drift term as before.

For each α value we generate 5,000 rough Bergomi samples paths. We then split the input data into training and test sets; we subsequently create a validation set from part of the test set. The sizes of each training/test/validation set of the rough Bergomi data are given in Table 4.5. Since we know a-priori that $H \approx 0.1$, we choose to only generate rough Bergomi sample paths with $\alpha < 0$ i.e. $H < 1/2$. Note that we will now refer to the value of α in the rough Bergomi model in terms of $H = \alpha + 1/2$, the reason for which will be clarified below.

Data set	Number of samples
Training set	14,000
Test set	7,500
Validation set	3,500

Table 4.5: Rough Bergomi input data size description.

Selection of H values

We begin by letting H, and hence the corresponding label $\mathbf{y}_i$, take values in the discrete grid $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. We also sample 5 H values from two probability distribu-

tions: the Uniform distribution[4] on (0,0.5) and the Beta(1,9) distribution[5]. This not only allows us to avoid discreteness in the output of the network; it should also make the network more robust when it comes to calibration, as H values for historical volatility data will almost certainly not be on the discrete grid $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. Futhermore, we are also able to emphasise the "rough" values of H, i.e. $H \approx 0.1$, particularly in the case of the Beta distribution.

Recall the pdf of the Beta(α, β) distribution is given by $f_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\mathbb{I}_{(0,1)}(x)$, where the function B is defined as $\mathrm{B}(\alpha, \beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. The Beta distribution has support $[0, 1]$ and therefore we cannot sample α values from this distribution, but sampling H values does make sense. We set $\alpha = 1, \beta = 9$, so that the expected value of the Beta distribution is 0.1, in accordance with the existing empirical studies [GJR18a] and [BLP17a]. For each H value we generate 5,000 rough Bergomi samples paths.

The CNN calibration method should also be robust to the dimensions of the input data, and training the CNN on vectors of length 100 should produce similar predictive performance to the CNN trained on vectors of length 500, as in the classification case presented above. Consequently we train the regression CNN with the length of the input vector taking values in $\{100, 200, 300, 400, 500\}$.

4.3.2 Regression Results

We now move on to solving the regression problem, in order to find the Hölder exponent from sample path input data. In Section 4.3.3 we use the rough Bergomi model described above to generate our input data; in Section 4.3.4 we additionally train the CNN with the above rough Bergomi input data, except with $\eta \neq 1$, as well as with random η and H for each sample path. We also train with fBm sample paths as input data. The aim is to illustrate the robustness of this novel method using CNNs. In Section 4.3.5 we briefly investigate if the CNN approach can be extended to additionally learn the parameter η, as well the parameter H. We finish by investigating the trained CNNs' prediction accuracy of the Hölder regularity of mean-reverting Ornstein-Uhlenbeck processes in

[4]The corresponding set of possible H values is $\{0.05, 0.18, 0.29, 0.31, 0.44\}$.
[5]The corresponding set of possible H values is $\{0.02, 0.07, 0.06, 0.13, 0.22\}$.

Section 4.3.6.

4.3.3 Rough Bergomi Test Results

The sizes of each training/test/validation set of the rough Bergomi data are given in Table 4.5. We train the CNN three times: for discretely sampled H, for Uniformly sampled H, and for Beta sampled H.

We present the test results for the CNN in Tables 4.6, 4.7, and 4.8. We use the mean square error as the loss function in the CNN, and report the predictive performance of the CNN using the root mean square error (RMSE), so that the predicted value and true value of H are of the same unit of measurement. We also give the time taken, in seconds, to complete the training and testing of the CNN[6]. The loss plots are given in Appendix B.2. The Python code is available here: https://github.com/henrymstone/CNN-repository.

We train the network with the architecture described above, setting `batch size`= 64 as is fairly standard, and `epochs`= 30 as this value gave the lowest mean square error. As a comparison, we also use the least square (LS) calibration approach suggested by [GJR18a, Section 2.1] on the test set used for the CNN, and compute the loss as the root mean square error between the predicted and true values for H.

Input length	RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
100	1.041×10^{-2}	69.77	0.76	2.118×10^{-1}	591.76
200	8.196×10^{-3}	74.89	0.75	2.046×10^{-1}	622.00
300	1.096×10^{-2}	80.92	0.79	2.025×10^{-1}	635.65
400	8.263×10^{-3}	92.02	0.93	2.014×10^{-1}	634.22
500	1.232×10^{-2}	93.76	0.93	2.010×10^{-1}	627.62

Table 4.6: Test results for discretised H.

For each input length, the predictive performances of the CNNs trained on $H \sim$

[6] All computations were executed in Python, using the Keras module to build and train the network, on a Macbook pro with a 2.6 GHz Intel Core i5 processor and 8 GB 1600 MHz DDR3 memory. The code was run on Google Colaboratory, using the platform's GPU running capabilities.

Input length	RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
100	1.137×10^{-2}	66.66	0.68	1.989×10^{-1}	611.61
200	7.910×10^{-3}	72.20	0.73	1.927×10^{-1}	620.72
300	5.115×10^{-3}	79.80	0.78	1.907×10^{-1}	630.75
400	9.409×10^{-3}	86.35	0.82	1.895×10^{-1}	634.24
500	1.282×10^{-2}	96.79	0.93	1.892×10^{-1}	628.26

Table 4.7: Test results for $H \sim \text{Uniform}(0.0, 0.5)$.

Input length	RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
100	6.672×10^{-3}	70.71	0.72	1.040×10^{-1}	616.74
200	7.193×10^{-3}	73.26	0.70	9.962×10^{-2}	626.77
300	1.207×10^{-1}	80.63	0.74	9.791×10^{-2}	637.45
400	1.171×10^{-2}	87.09	0.75	9.699×10^{-2}	637.33
500	1.207×10^{-1}	94.20	0.77	9.663×10^{-2}	644.60

Table 4.8: Test results for $H \sim \text{Beta}(1, 9)$.

Uniform$(0.0, 0.5)$ and discretised H are similar, and in each case the CNN approach clearly outperforms the least square approach in terms of predictive power, by one or two orders of magnitude. For $H \sim \text{Beta}(1, 9)$, the CNN also outperforms the least square approach, again by one or two orders of magnitude, when the input vector length is 100, 200 or 400; the accuracy of the CNN is slightly poorer than the least square approach for the other input vector lengths.

As one would expect for both calibration approaches, the time taken in both cases is, in general, an increasing function of the length of the input vector for each method of sampling H. Since we are able to train the CNN using Google Colaboratory's GPUs, and Python's Keras module has been optimised for execution on GPU, the time taken for training and testing is approximately eight times less than the least square approach. The training and test times for each sampling method are all very similar.

The above analysis indicates that decreasing the length of the input vector does not significantly worsen the predictive performance of the CNN; in fact when $H \sim \text{Beta}(1, 9)$ the performance is improved, when comparing the performance of the length 100 input

vector and the length 500 input vector.

4.3.4 Robustness Test Results

Following the analysis in Section 4.3.3, we set the input vector length to be 100. We then generate our input data by letting η take values in $\{0.25, 0.8, 1.3, 2.5\}$ in the rough Bergomi model, and use discretised sampling for H to generate 5,000 sample paths for each H. We give the results in Table 4.9; as before we also include the root mean square error (RMSE) and time taken in seconds for the least square (LS) approach of [GJR18a], applied to the test set as a comparison. Plots of the training error and validation error are given in Appendix B.3.

η	RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
0.25	8.206×10^{-3}	66.81	0.63	2.122×10^{-1}	666.79
0.8	1.137×10^{-2}	66.93	0.60	2.122×10^{-1}	665.68
1.3	1.473×10^{-2}	67.22	0.66	2.122×10^{-1}	671.11
2.5	9.003×10^{-3}	67.27	0.63	2.122×10^{-1}	667.16

Table 4.9: Rough Bergomi regression results for $\eta \neq 1$ and input vector length=100.

We can see that the CNN approach vastly outperforms the LS approach in each case, both in terms of the accuracy of the predictions and the time taken. The CNN's performance for $\eta = 0.8, 1.3$ is slightly worse than for the other two values of η, but is still superior to the LS approach. Note that values for the RMSE values for the LS approach are only equal when rounded to three decimal places.

We further extend the robustness test on rough Bergomi data as follows: we begin by generating 25,000 $\eta \sim$ Uniform$(0, 3)$ and $H \sim$ Beta$(1, 9)$, and then use these values to simulate 25,000 rough Bergomi sample paths of length 100, each with its own unique and random η and H. The corresponding training, test, and validation sets thus remain the same size. The results are presented in Table 4.10; as above we include the root mean square error and time taken for the LS approach as a comparison. Plots of the training error and validation error are given in Appendix B.4.

Interestingly, when $\eta \sim$ Uniform$(0, 3)$ and $H \sim$ Beta$(1, 9)$ for each sample path, the

predictive power of the CNN is only inferior, in the worst case, by one order of magnitude compared to when the values for either η or H are fixed. Furthermore, the CNN approach maintains its significant advantage over the LS approach, both in terms of predictive power and speed.

RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
1.382×10^{-2}	66.52	0.61	1.499×10^{-1}	598.51

Table 4.10: Rough Bergomi regression results for $\eta \sim \text{Uniform}(0, 3)$, $H \sim \text{Beta}(1, 9)$, and input vector length=100.

We conclude this robustness test by using fBm sample paths, generated using Cholesky decomposition, as input data to train and test the CNN. For each H we simulate 5,000 sample paths of length 100; we employ the discretised H, $H \sim \text{Uniform}(0.0, 0.5)$, and $H \sim \text{Beta}(1, 9)$ sampling methods. The results are given in Table 4.11 and the plots of the training error and validation error are given in Appendix B.5.

Sampling	RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
Discretised	2.483×10^{-2}	72.79	0.65	2.346×10^{-1}	635.18
Uniform	2.001×10^{-2}	71.95	0.64	2.090×10^{-1}	615.74
Beta	1.945×10^{-2}	72.20	0.67	9.785×10^{-2}	622.36

Table 4.11: fBm regression results for discretised H, $H \sim \text{Uniform}(0.0, 0.5)$, and $H \sim \text{Beta}(1, 9)$.

The CNN maintains its speed advantage over the LS method for each sampling method for H, as well as maintaining superior predictive performance, by an order of magnitude. The results from this final part of the robustness test allow us to conclude that CNNs can indeed identify Hölder regularity from a set of sample paths, thus answering the question posed at the start of the Chapter.

4.3.5 Extension to Learning η

In this Subsection we briefly explore whether the CNN can learn the value of η, as well as H. We use the rough Bergomi model, with $\eta \sim \text{Uniform}(0, 3)$, $H \sim \text{Beta}(1, 9)$ for

each sample path, and input vectors of length 100 as our input data, as above; the corresponding output variable then becomes a two-dimensional vector $\mathbf{y}_i = (H_i, \eta_i)$. We continue to use the root mean square error as a measure of predictive power, and compare to the least square approach as above. The least square (LS) approach can indeed be used to estimate the value of η [GJR18a, Section 3.4], although the authors use the notation ν instead of η. The results are given in Table 4.12, with loss plots given in Appendix B.6.

RMSE (CNN)	Training Time	Test Time	RMSE (LS)	Time
0.666	71.82	0.62	1.170	613.62

Table 4.12: Regression results for learning H and η, with input vector length=100.

Encouragingly, we see that the CNN approach still outperforms the LS approach for both accuracy and time; however the CNN approach is approximately twice as accurate as the LS approach, compared to the orders of magnitude in the cases above. Note that we kept all hyperparameter values unchanged; it is likely that superior predictive power could be achieved with some tuning of the hyperparameter values in the CNN. This is not the focus of this study, however, and we leave this question to further research.

Remark 4.3.1. We conclude the above results with a remark on the speed and accuracy advantages of the CNN approach, from a theoretical perspective. The above results show the CNN method to be more accurate, by orders of magnitude, and significantly faster than the existing method suggested in [GJR18a] when estimating H on simulated rough Bergomi and fBm data. The explanation for the speed advantage of the CNN method is that once the network has been trained, estimations are made by the straightforward computation of the composition (4.1) given in Definition 4.1.2. On the other hand, the LS approach requires a number of successive regressions to be executed in order to estimate H. Clearly then the CNN method will be faster than the LS approach. We now recall our motivation that justifies the use of a CNN to estimate H: the predictive power of the CNN lies in the convolutional operator, which assigns a value to each entry $\mathbf{x}_i^j$ of an input vector $\mathbf{x}_i$; this value is determined by the relative values of the entries neighbouring $\mathbf{x}_i^j$. To estimate the Hölder regularity of a stochastic process, the values of neighbouring points of each entry in a trajectory vector will provide the most

information about the Hölder regularity of that trajectory, and thus a CNN is indeed a valid means of estimating H. The CNN is able to detect very subtle path regularity properties, via interated applications of the convolutional operator, allowing for very accurate predictions of H. Furthermore, the LS estimation is based on an approximation of the q^{th} moment of an increment of log volatility, whose accuracy is highly dependent on the choice of q, while the CNN approach is completely independent of the choice of q. Together, this explains why the CNN approach gives more accurate H estimations than the LS approach.

4.3.6 Robustness Test Results on a (Mean-reverting) Ornstein-Uhlenbeck Process

We finish this Section with a comparison of the performance of the CNN and LS approaches on a (mean-reverting) Ornstein-Uhlenbeck process. Recall that a (mean-reverting) Ornstein-Uhlenbeck process X satisfies the following SDE:

$$\mathrm{d}X_t = (a - bX_t)\mathrm{d}t + cX_t\mathrm{d}W_t, \qquad X_0 = x_0 \in \mathbb{R},$$

where W is a standard Brownian motion, and that X is γ-Hölder continuous for all $\gamma \in (0, 1/2)$; therefore we expect both approaches to estimate $H \approx 0.5$.

We set the input length to be 100 for the CNN, using the Discretised and Uniform H sampling methods detailed above[7] to train the CNN. Each trained network is then used to estimate H on trajectories of a (mean-reverting) Ornstein-Uhlenbeck process, which are simulated using the Euler-Maruyama method. We simulate 1000 trajectories of length 100, setting $(x_0, a, b) = (0.1, 1., 2.1)$, and report the mean estimated H values for the CNN and LS approaches in Table 4.13 below, for the 'high volatity' ($c = 3$) and 'low volatility' ($c = 0.3$) regimes.

The results in Table 4.13 provide compelling evidence that the CNN is indeed learning the Hölder regularity of the sample path, as the mean H estimates for the CNN are sufficiently close to 0.5. Note that the mean LS H estimates are significantly further

[7]We do not experiment with the Beta sampling method, since those H values are concentrated around 0.1, and so will produce poor estimations when we expect $H \approx 0.5$.

c	CNN (Discretised)	CNN (Uniform)	LS
3	0.46	0.42	0.22
0.3	0.49	0.44	0.67

Table 4.13: Mean H estimates for a (mean-reverting) Ornstein-Uhlenbeck process X.

from 0.5. The results provide further evidence that the CNN approach is preferable to the LS approach, which is not only slower and less accurate than the CNN approach but may also incorrectly identify roughness in data.

4.4 Calibration Using CNNs

We finish by solving the calibration problem that motivated Chapter 4. Indeed, existing calibration techniques for the rough Bergomi (rBergomi) model are still in need of development:

> "Calibration of the rBergomi model is not easy... So far, we cannot claim to have had real success with any of these approaches".
>
> Jim Gatheral, 'Rough volatility: an overview', Global Derivatives Trading and Risk Management 2017 Barcelona presentation, page 46.

The method predicts H accurately on simulated data, but in practice volatility swaps tend to be illiquid for maturities less than 8 months, and therefore it is difficult to use this method for accurate calibration in practice. Another technique, proposed by Chang [Cha14], suggests using maximum likelihood estimation to estimate H. While the method accurately predicts H from simulated fBm data, the computational cost of this approach is too high for practical application in the quantitative finance industry. Lastly we consider the least squares method of Gatheral, Jaisson, and Rosenbaum [GJR18a]. Inspired by the q^{th} moment formula[8] for increments of a fBm, the authors suggest estimation of H via a linear regression of the log of lagged q^{th} moments of the

[8] For a fBm W^H, the following holds for all $q > 0$: $E[|W^H_{t+\Delta} - W^H_t|^q] = K_q \Delta^{qH}$, where K_q is the absolute q^{th} moment of a standard normal distribution.

log volatility process against the log of the lags. The method, however, is sensitive to the choice of q; in particular the method does not perform well for higher order moments. Additonally, in Section 4.3.2, we demonstrated that the least squares method will provide erroneous H estimates for processes exhibiting mean reversion.

We begin by using the trained CNNs from Section 4.3.3 to predict the Hölder exponent of historic realised volatility data from the Oxford-Man Institute of Quantitative Finance, which is free and publicly available[9]. We choose the length of the input vectors to be 100 from the analysis given in Section 4.3.3. We took a sample of 10 different indices[10] from the 31 available; for each index we then used a time series of 200 sequential data points to create 11 vectors of length 100 (entries 0 to 100, 10 to 110, and so on) to predict the Hölder exponent for each index. We compute the root mean square error between the CNN prediction and the least square prediction, and the standard deviation of the difference between the two predictions; see Table 4.14.

Sampling Method	Root Mean Square Error	Standard Deviation
Discretised H	5.558×10^{-2}	2.900×10^{-3}
$H \sim \text{Uniform}(0.0, 0.5)$	2.444×10^{-1}	1.141×10^{-2}
$H \sim \text{Beta}(1, 9)$	4.253×10^{-2}	1.098×10^{-3}

Table 4.14: Calibration results

Indeed, we can see that this set of results is very promising. In each case, both the root mean square error and the standard deviation are small; note that both root mean squared error values and the standard deviations of discretised H and $H \sim \text{Beta}(1, 9)$ are an order of magnitude greater than for $H \sim \text{Uniform}(0.0, 0.5)$. This therefore indicates that the calibration values attained by the network are very close to those attained by the least squares approach. Note that this in turn provides further evidence that $H \approx 0.1$, further corroborating the findings of [GJR18a] and [BLP17a].

We can state, therefore, that this calibration scheme is precise enough to be used in practice, where we recommend using $H \sim \text{Beta}(1, 9)$ to train the network with input

[9] https://realized.oxford-man.ox.ac.uk/data/download

[10] AEX, All Ordinary, DAX, FTSE 100, Hang Seng, NIFTY 50, Nasdaq 100, Nikkei 225, S&P500 , Shanghai Composite.

length 100, due smaller root mean squared error and standard deviation in testing, the nondiscretised network output, and emphasis of this distribution on "rough" values of H i.e. $H \approx 0.1$.

The practical implementation of our calibration scheme is a simple two-step process. The first step is to train the CNN, with the H sampling method and input length n chosen by the practitioner; this can be done once offline, with the weights of the trained network saved to avoid unnecessary repetition for each calibration task. The second step is to input the most recent n volatility observations into the CNN, which will return the corresponding H value for those n observations. This value of $H = \alpha + 1/2$ can then be inputted into the rough Bergomi model and used for, say, pricing. We note that some testing for the optimal choice of H sampling method and input length n is required on the part of the practitioner implementing our calibration scheme.

Remark 4.4.1. We now take this opportunity to discuss the calibration methodology presented in this paper. We are treating calibration as a *supervised* learning problem when in practice it is an *unsupervised* learning problem, strictly speaking. While each vector in the input data in the regression problems above was indeed labelled with the corresponding H value, the data from the Oxford-Man Institute 'realized' library has no such labels, thus we use the least square calibration values as "true" values.

Conclusion

This thesis has focused on rough volatility models, and has made a number of important contributions both to their understanding from a theoretical perspective, and their application from a practical perspective.

On the theoretical side we have proved a number of general results for Gaussian measures on general Banach spaces, which allow us to establish the small-time log stock price, integrated variance, and pathwise VIX behaviour in a number of rough volatility models, using the theory of large deviations. We then characterise the small-time implied volatility behaviour, for Call options on stock price and integrated variance, in terms of the rate function of the large deviations principle. Crucially, we find that the parameter describing the Hölder regularity of the volatility process determines the speed of the large deviations principle, and thus the power law behaviour of the level of the implied volatility as time to maturity approaches zero, as well as implicitly defining the rate function itself. Extensions to small-time results for more general volatility processes are possible in our framework, by using lemmas and propositions provided, and would make for interesting new research topics. Investigation into the large-time behaviour of rough volatility models is also possible using our framework.

On the practical side we have provided numerical schemes to compute the rate functions for integrated variance processes; we are indeed able to plot accurate implied volatility smiles, which agree with Monte Carlo simulations. Perhaps surprisingly, we have discovered that lognormal models such as rough Bergomi (1.11), 2 Factor Bergomi [Ber08, Ber16] and mixed versions thereof, generate linear smiles around the money for options on realised variance in log-space. This is, at the very least, a property to

be taken into account when modelling volatility derivatives and, to our knowledge, has never been addressed or commented on in previous works. Whether such an assumption is realistic or not, we have in addition provided an explicit way to construct a model that generates non-linear smiles. Additionally, by approximating the realised variance density, we provide a means of pricing volatility swaps in the mixed rough Bergomi model that is fast and remarkably accurate, particularly in the close-to-maturity case. We have also provided an efficient and accurate means of calibrating the rough Bergomi model; our method is shown to be far more accurate, by orders of magnitude, and significantly faster than the existing method suggested in [GJR18a]. We have shown our method to correctly estimate H on mean-reverting trajectories, where the least squares approach fails and incorrectly identifies roughness.

Bibliography

[AB06] C.D. Aliprantis and K.C. Border. Infinite Dimensional Analysis: A Hitchhiker's Guide, Third Edition. Springer-Verlag Berlin Heidelberg, 2006.

[ADY13] O. Abdel-Hamid, L. Deng, and D. Yu. Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition. *Interspeech*, 2013.

[AFLZ17] J. Armstrong, M. Forde, M. Lorig, and H. Zhang. Small-Time Asymptotics under Local-Stochastic Volatility with a Jump-To-Default: Curvature and The Heat Kernel Expansion. *SIAM Journal on Financial Mathematics*, 8(1): 82-113, 2017.

[AGM18] E. Alòs, D. García-Lorite and A. Muguruza. On smile properties of volatility derivatives and exotic products: understanding the VIX skew. Preprint available at arXiv:1808.03610, 2018.

[AL12] L. Andersen and A. Lipton. Asymptotics For Exponential Lévy Processes and Their Volatility Smile: Survey and New Results. Survey available at SSRN:2095654, 2012.

[ALV07] E. Alòs, J.A. León and J. Vives. On the short-time behaviour of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics* 11(4): 571-589, 2007.

[AP18] M. Avellaneda and A. Papanicolaou. Statistics of VIX Futures and Their Applications to Trading Volatility Exchange-Traded Products. *The Journal of Investment Strategies* 7(2): 1-33, 2018.

[AS72] M. Abramovitz and I. Stegun. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. New York: Dover Publications, 1972.

[AS19] E. Alòs and K. Shiraya. Estimating the Hurst parameter from short term volatility swaps: a Malliavin calculus approach. *Finance and Stochastics* 23(2): 423-447, 2019.

[ASW17] J. Akahori, X. Song, and T. Wang. Probability density of lognormal fractional SABR model. Preprint available at arXiv:1702.08081, 2017.

[AT07] R.J. Adler and J.E. Taylor. Random Fields and Geometry. Springer-Verlag, New York, 2007.

[Ber05] L. Bergomi. Smile Dynamics II. *Risk*, October: 67-73, 2005.

[Ber08] L. Bergomi. Smile dynamics III. *Risk*, October: 90-96, 2008.

[Ber16] L. Bergomi. Stochastic Volatility Modeling. Chapman & Hall/CRC financial mathematics series, 2016.

[BCG16] Y. Bengio, A. Courville, and I. Goodfellow. Deep Learning. MIT Press, 2016.

[BCGL97] A.D. Back, A. Chung-Tsoi, C.L. Giles, and S. Lawrence. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1): 98-113, 1997.

[BFG16] C. Bayer, P.K. Friz and J. Gatheral. Pricing Under Rough Volatility. *Quantitative Finance*, 16(6): 887-904, 2016.

[BFGHS19] C. Bayer, P.K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper. Short-time near the money skew in rough fractional stochastic volatility models *Quantitative Finance*, 19(5): 779-798, 2019.

[BG12] L. Bergomi and J. Guyon. The Smile in Stochastic Volatility Models. Preprint available at SSRN: 1967470, 2012.

[BGTW19] H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep Hedging. *Quantitative Finance*, 19(8): 1271-1291, 2019.

[BHØZ08] F. Biagini, Y. Hu, B. Øksendal and T. Zhang. Stochastic Calculus for Fractional Brownian Motion and Applications. Springer London, 2008.

[BK17] J. Ba and D.P. Kingma. Adam: A Method for Stochastic Optimization. Preprint available at arXiv:1412.6980v9, 2017.

[BL78] D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in options prices. *Journal of Business*, 51: 621-651, 1978.

[BLP17a] M. Bennedsen, A. Lunde and M.S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. Preprint available at arXiv:1610.00332v2, 2017.

[BS07] O. E. Barndorff-Nielsen and J. Schmiegel. Ambit processes: with applications to turbulence and tumour growth. Stochastic Analysis and Applications, Volume 2 of Abel Symposium: 93-124, Springer Berlin, 2007.

[BS18] C. Bayer and B. Stemper. Deep calibration of rough stochastic volatility models. Preprint available at arXiv:1810.03399v1, 2018.

[CCR12] F. Comte, L. Coutin and E. Renault. Affine Fractional Stochastic Volatility Models. *Annals of Finance*, 8: 337-378, 2012.

[CGMY05] P. Carr, H. Geman, D. Madan, and M. Yor. Pricing options on realized variance. *Finance and Stochastics*, 9(4): 453-475, 2005.

[Cha14] Y.C. Chang. Efficiently Implementing the Maximum Likelihood Estimator for Hurst Exponent. *Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2014*, 2014.

[Che08] A. Cherny. Brownian moving averages have conditional full support. *Annals of Applied Probability*, 18(5): 1825-1850, 2008.

[CKSX15] C. Chen, A. Kornhauser, A. Seff, and J. Xiao. DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. *IEEE International Conference on Computer Vision (ICCV)*: 2722-2730, 2015.

[CL09] P. Carr and R. Lee. Volatility Derivatives. *Annual Review of Financial Economics* 1(1): 319-339, 2009.

[CM44] R.H. Cameron and W.T. Martin. Transformations of Wiener Integrals Under Translations. *Annals of Mathematics*, 45(2): 386-396, 1944.

[CM99] P. Carr and D. Madan. Option Pricing and the Fast Fourier Transform. *Journal of Computational Finance*, 2(4): 61-73, 1999.

[Cra38] H. Cramér. Sur un Nouveau Théorème-Limite de la Théorie des Probabilités. *Actualités Scientifiques et Industrielles*, 736: 5-23, 1938.

[CR98] F. Comte and E. Renault. Long Memory in Continuous-Time Stochastic Volatility Models. *Mathematical Finance*, 8(4): 291-323, 1998.

[CT06] R. Carmona and M. Tehranchi. Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective. Springer-Verlag Berlin Heidelberg, 2006.

[Dup94] B. Dupire. Pricing With a Smile. *Risk*, 7: 18-20, 1994.

[DHK13] L. Deng, G. Hinton, and B. Kingsbury. New types of deep neural network learning for speech recognition and related applications: an overview. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.

[DJR19] A. Dandapani, P. Jusselin and M. Rosenbaum. From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect. Preprint available at arXiv:1907.06151, 2019.

[DS89] J.D. Deuschel and D.W. Stroock. Large Deviations. Academic Press Inc., 1989.

[DS94] F. Delbaen and W. Schachermayer. A Generalised Version of the Fundamental Theorem of Asset Pricing. *Mathematische Annalen*, 300(1): 463-520, 1994.

[DZ10] A. Dembo and O. Zeitouni. Large Deviations Theory and Applications, Second Edition. Springer-Verlag Berlin Heidelberg, 2010.

[EGR19] O. El Euch, J. Gatheral, and M. Rosenbaum. Roughening Heston. *Risk*, May: 84-89, 2019.

[ER19] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance* 29(1) : 3-38, 2019.

[ER18] O. El Euch and M. Rosenbaum. Perfect hedging in rough Heston models. To appear *The Annals of Applied Probability*, 28(6): 3813-3856, 2018.

[EFR18] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics* 22: 241-280, 2018.

[Fuk17] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance* 17(2): 189-198, 2017.

[FFF10] J. Feng, M. Forde, and J. P. Fouque. Asymptotics For a Fast Mean-Reverting Heston Model. *SIAM Journal on Financial Mathematics*, 1(1): 126-141, 2010.

[FF12] J.E. Figueroa-López and M. Forde. The Small-Maturity Smile for Exponential Lévy Models. *Siam Journal on Financial Mathematics*, 3(1): 33-65, 2012.

[FGH16] J.E. Figueroa-López, R. Gong, and C. Houdré. High-Order Short-Time Expansions for ATM Options Prices of Exponential Lévy Models. *Mathematical Finance*, 26(3): 516-557, 2016.

[FGS19] M. Forde, S. Gerhold, and B. Smith. Small-time and large-time smile behaviour for the Rough Heston model. Preprint available at arXiv:1906.09034, 2019.

[FJ09] M. Forde and A. Jacquier. Small-Time Asymptotics for Implied Volatility Under The Heston Model. *IJTAF*, 12(6): 861-876, 2009.

[FJ11] M. Forde and A. Jacquier. Small-Time Asymptotics for an Uncorrelated Local-Stochastic Volatility Model. *Applied Mathematical Finance*, 18(6): 517- 535, 2011.

[FJL12] M. Forde, A. Jacquier, and R. Lee. The Small-Time Smile and Term Structure of Implied Volatility Under the Heston Model. *SIAM Journal on Financial Mathematics*, 3(1): 690-708, 2012.

[FJM10] M. Forde, A. Jacquier, and A. Mijatović. Asymptotic Formulae For Implied Volatility in the Heston Model. *Proceedings of the Royal Society A*, 466: 3593-3620, 2010.

[FO16] J.E. Figueroa-López and S. Ólaffson. Short-Time Expansions for Close-to-the-Money Options Under a Lévy Jump Model with Stochastic Volatility. *Finance and Stochastics*, 20(1): 219-265, 2016.

[FTW19] M. Fukasawa, T. Takabatake, and R. Westphal. Is Volatility Rough? Preprint available at arXiv:1905.04852, 2019.

[FV10] P.K. Friz and N.B. Victoir. Multidimensional Stochastic Processes as Rough Paths: Theory and Application. Cambridge University Press, 2010.

[FSV18] M. Forde, B. Smith and L. Viitasaari. Rough unbounded volatility with CGMY jumps conditioned on a finite history and the Rough Heston model - small-time Edgeworth expansions and the prediction formula for the Riemann-Liouville process. Preprint, 2018.

[FZ17] M. Forde and H. Zhang. Asymptotics for Rough Stochastic Volatility Models. *SIAM Journal on Financial Mathematics* 8: 114-145, 2017.

[Gar08] J. Garcia. A Large Deviations Principle for Stochastic Integrals. *Journal of Theoretical Probability*, 21: 476-501, 2008.

[Gas19] P. Gassiat. On the martingale property in the rough Bergomi model. *To appear in Electronic Communications in Probability*, 2019.

[Gat06] J. Gatheral. The Volatility Surface: A Practioner's Guide. Wiley Finance, 2006.

[Gui03] A. Guillin. Averaging Principle of SDE with Small Diffusion: Moderate Deviations. *The Annals of Probability*, 31(1): 413-443, 2003.

[Guy14] J. Guyon. Path Dependent Volatility. *Risk*, Oct: 2014.

[GL14] K. Gao and R. Lee. Asymptotics of implied volatility to arbitrary order. *Finance and Stochastics*, 18(2): 349-392, 2014.

[GJR18a] J. Gatheral, T. Jaisson and M. Rosenbaum. Volatility is Rough. *Quantitative Finance*, 18(6): 933-949, 2018.

[GJR18b] H. Guennoun, A. Jacquier and P. Roome. Asymptotics of the Fractional Heston Model. *SIAM Journal on Financial Mathematics*, 9(3): 1017-1045, 2018.

[GM69] J. Golosov and G. Molchan. Gaussian Stationary Processes with Asymptotic Power Spectrum. *Soviet Mathematics Doklady*, 10(1): 134-137, 1969.

[GVZ18] A. Gulisashvili, F. Viens and X. Zhang. Small-time Asymptotics for Gaussian Self-Similar Stochastic Volatility Models. *Applied Mathematics & Optimization*, 77(2): 1-41, 2018.

[Hes93] S. Heston. A Closed-Form Solution For Options with Stochastic Volatility, with Application to Bond and Currency Options. *Review of Financial Studies*, 6(2): 327-343, 1993.

[Hol00] F. den Hollander. Large Deviations. *Fields Institute Monographs*, 14: American Mathematical Society, 2000.

[HJL19] B. Horvath, A. Jacquier, and C. Lacombe. Asymptotic behaviour of randomised fractional volatility models. *Journal of Applied Probability*, 56(2): 496-523, 2019.

[HJT18] B. Horvath, A. Jacquier and P. Tankov. Volatility options in rough volatility models. Preprint available at arXiv:1802.01641, 2018.

[HMT19] B. Horvath, A. Muguruza, and M. Tomas. Deep Learning Volatility. Preprint available at arXiv:1901.09647, 2019.

[HKLW02] P.S. Hagan, D. Kumar, A. Lesniewski, and D.E. Woodward. Managing Smile Risk. *Wilmott Magazine*, 2002.

[HKS12] G. E. Hinton, A. Krizhevsky, and I. Sutskever. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25: 2012.

[HKSSS14] G. Hinton, A. Krizhevsky, R. Salakhutdinov, N. Srivastava, and I. Sutskever. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15: 1929-1958, 2014.

[HR98] D.G. Hobson and L.C.G. Rogers. Complete Models with Stochastic Volatility. *Mathematical Finance*, 8(1): 27-48, 1998.

[IJKW17] F.N. Iandola, P.H. Jin, K. Keutzer, and B. Wu. SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*: 446-454, 2017.

[JKM13] A. Jacquier, M. Keller-Ressel, and A. Mijatović. Large Deviations and Stochastic Volatility Models with Jumps: Asymptotic Implied Volatility for Affine Models. *Stochastics*, 85(2): 321-345, 2013.

[JMM18] A. Jacquier, C. Martini, A. Muguruza. On VIX Futures in the rough Bergomi model. *Quantitative Finance*, 18(1): 45-61, 2018.

[JO19] A. Jacquier and M. Oumgari. Deep PPDEs for rough local stochastic volatility. Preprint available at arXiv:1906.02551, 2019.

[JR15] A. Jacquier and P. Roome. Asymptotics of Forward Implied Volatility. *SIAM Journal on Financial Mathematics*, 6(1): 307-351, 2015.

[JPS18] A. Jacquier, M. Pakkanen, and H. Stone. Pathwise Large Deviations for the Rough Bergomi Model. *Journal of Applied Probability* 55(4): 1078-1092, 2018.

[KM13] M. Keller-Ressel and J. Muhle-Karbe. Asymptotic and Exact Pricing of Options on Variance. *Finance and Stochastics* 17(1): 107133, 2013.

[KS91] I. Karatzas and S.E. Shreve. Brownian Motion and Stochastic Calculus: Second Edition. Springer-Verlag New York, 1991.

[KNTY18] R. Kinh Gian Do, M. Nishio, K. Togashi, and R. Yamashita. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4): 611-629, 2018.

[Lev53] P. Lévy. Random functions: general theory with special references to Laplacian random functions. *University of California Publications in Statistics*, 1: 331-390, 1953.

[Lif12] M. Lifshits. Lectures on Gaussian Processes. *Springer-Verlag Berlin Heidelberg*, 2012.

[LMS19] C. Lacombe, A. Muguruza, and Henry Stone. Asymptotics for volatility derivatives in multi-factor rough volatility models. Preprint available at arXiv:1903.02833, 2019.

[MLP17] H. Mhaskar, Q. Liao, and T. Poggio. When and Why Are Deep Networks Better than Shallow Ones? *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*: 2343-2349, 2017.

[MS07] A. Medvedev and O. Scaillet. Approximation and Calibration of Short-Term Implied Volatilities Under Jump-Diffusion Stochastic Volatility. *The Review of Financial Studies*, 20(2): 427-459, 2007.

[MT16] A. Mijatović and P. Tankov. A New Look at Short-Term Implied Volatility in Asset Price Models with Jumps. *Mathematical Finance*, 26(1): 149-183, 2016.

[MV68] B.B. Mandelbrot and J.W. Van Ness. Fractional Brownian Motions, Fractional Noises, and Applications. *SIAM Review*, 10(4): 422-437, 1968.

[Neu94] A. Neuberger. The log contract. *Journal of Portfolio Management*, 20(2): 74-80, 1994.

[NR18] E. Neuman and M. Rosenbaum. Fractional Brownian motion with zero Hurst parameter: a rough volatility viewpoint. *Electronic Communications in Probability*, 23(61): 1-12, 2018.

[Øks03] B. Øksendal. Stochastic Differential Equations: An Introduction with Applications, Sixth Edition. Springer-Verlag Berlin Heidelberg, 2003.

[Pau15] L. Paulot. Asymptotic Implied Volatility at the Second Order with Application to the SABR model. Large Deviations and Asymptotic Methods in Finance, Springer International Publishing Switzerland, 2015.

[Pha08] H. Pham. Some Applications and Methods of Large Deviations in Finance and Insurance. Preprint available at arXiv:0702473v2, 2008.

[Rog97] L.C.G. Rogers. Arbitrage with Fractional Brownian Motion. *Mathematical Finance*, 7(1): 95-105, 1997.

[Sto20] H. Stone. Calibrating Rough Volatility Models: a Convolutional Neural Network approach. *Quantitative Finance*, 20(3): 379-392, 2020.

[SZ15] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. Preprint available at arXiv:1409.1556v6, 2015.

[Tit26] E.C. Titchmarsh. The Zeros of Certain Integral Functions. *Proceedings of the London Mathematical Society*, 25: 283-302, 1926.

[Tou09] H. Touchette. The Large Deviation Approach to Statistical Mechanics. Preprint available at arXiv:0804.0327v2 , 2009.

[Will70] S. Willard. General Topology. Addison-Wesley Publishing Company Inc., 1970.

[WW05] R. L. Weron and U. P. Wystrup. Heston's Model and the Smile. *Statistical Tools for Finance and Insurance*, 1: 161-181, 2005.

[Yeg09] B. Yegnanarayana. Artificial Neural Networks. PHI Learning Pvt. Ltd., 2009.

Appendix A

An LDP for the Uncorrelated Rough Bergomi Model

We treat here the special case of (1.11), where the Brownian motions W and B have zero correlation ($\rho = 0$). Following similar arguments to Proposition 1.4.18, and analogously to (1.17), we introduce the operator $\mathcal{I}^0 : \mathrm{L}^2 \times \mathrm{L}^2 \to \mathcal{C}^2$ as

$$\mathcal{I}^0(f_1, f_2) := \begin{pmatrix} \mathcal{I}^{K_\alpha} f_1 \\ \mathcal{I}^1 f_2 \end{pmatrix}, \qquad \text{for any } f_1, f_2 \in \mathrm{L}^2,$$

so that the RKHS (on $\mathcal{C}^2$) of the measure induced by (Z, B) is $\mathscr{H} := \left\{\mathcal{I}^0(f_1, f_2) : f_1, f_2 \in \mathrm{L}^2\right\}$, with inner product $\left\langle \mathcal{I}^0(f_1, f_2), \mathcal{I}^0(g_1, g_2)\right\rangle_{\mathscr{H}} := \langle f_1, g_1\rangle_{\mathrm{L}^2} + \langle f_2, g_2\rangle_{\mathrm{L}^2}$, for any $f_1, f_2, g_1, g_2 \in \mathrm{L}^2$. Similarly to Theorem 2.1.2, [DS89, Theorem 3.4.12] yields an LDP on $\mathcal{C}^2$ for $((Z^\varepsilon, B^\varepsilon))_{\varepsilon\in\mathcal{T}}$ with speed $\varepsilon^{-\beta}$ and rate function

$$\Lambda(\mathrm{z}_y^x) := \begin{cases} \frac{1}{2}\left\|\mathrm{z}_y^x\right\|_{\mathscr{H}}^2, & \text{if} (x, y)^\top \in \mathscr{H}, \\ +\infty, & \text{otherwise.} \end{cases}$$

This in turn yields an LDP for $((v^\varepsilon, B^\varepsilon))_{\varepsilon\in\mathcal{T}}$ in (3.9) on $\mathcal{C}^2$ with speed $\varepsilon^{-\beta}$ and rate function $\widetilde{\Lambda}(\mathrm{z}_y^x) := \inf\left\{\Lambda(\mathrm{z}_{y^*}^{x^*}) : \mathrm{z}_y^x = \mathrm{M}\mathrm{z}_{y^*}^{x^*}\right\}$, where the operator M is defined in (2.2). In the same vein as Theorem 2.1.2, Theorem 1.5.17 yields an LDP for $(\int_0^\cdot \sqrt{v_s^\varepsilon}\mathrm{d}B_s^\varepsilon)_{\varepsilon\in\mathcal{T}}$

on $\mathcal{C}$ with speed $\varepsilon^{-\beta}$ and rate function $\hat{\Lambda}^X$, defined as

$$\begin{aligned}\hat{\Lambda}^X(\varphi) &:= \inf\left\{\widetilde{\Lambda}(\mathrm{z}_y^x) : \varphi = x \cdot y, y \in \mathrm{BV} \cap \mathcal{C}\right\} = \inf\left\{\Lambda(\mathrm{z}_{y^*}^{x^*}) : \varphi = x \cdot y, \mathrm{z}_y^x = \mathrm{Mz}_{y^*}^{x^*}, x^*, y^* \in \mathscr{H}\right\} \\ &= \inf\left\{\Lambda(\mathrm{z}_y^x) : \varphi = x \cdot y, \mathrm{z}_y^x = \mathrm{M}(\mathcal{I}^0(f_1, f_2)), f_1, f_2 \in \mathrm{L}^2\right\} \\ &= \inf_{f_1, f_2 \in \mathrm{L}^2}\left\{\frac{1}{2}\|f_1\|_{\mathrm{L}^2}^2 + \frac{1}{2}\|f_2\|_{\mathrm{L}^2}^2 : \varphi = \int_0^{\cdot} \sqrt{\mathfrak{m}\left((\mathcal{I}^{K_\alpha} f_1)(s)\right)} f_2(s)\mathrm{d}s\right\}.\end{aligned}$$

with $\mathfrak{m}$ introduced in (2.3). Following identical an identical argument to that presented in Theorem 2.1.2, we conclude that $(X^\varepsilon)_{\varepsilon \in \mathcal{T}}$ satisfies an LDP with speed $\varepsilon^{-\beta}$ and rate function $\hat{\Lambda}^X$.

Appendix B

CNN Loss plots

In this final Appendix we present the loss ploss for the CNN in Chapter 4. We begin with the classification loss plots, and then move onto on the regression loss plots, which are presented in the same order as their respective results.

B.1 Classification Loss Plots

Figures B.1 and B.2 show that the training/validation accuracy both increase, and training/validation loss both decrease, as the number of epochs increases for the fraction Brownian motion and rough Bergomi input data respectively. This is a good indication that predictive performance improves as the number of epochs increases, without overfitting [KNTY18, Overfitting Section, Page 619].

B.2 Loss Plots for rough Bergomi with Different H Sampling

Here we plot the training and validation loss (MSE) at each epoch, for discretised H, $H \sim \text{Uniform}(0.0, 0.5)$, and $H \sim \text{Beta}(1, 9)$ in Figures B.3, B.4, and B.5 respectively. Note that the training and validation loss both tend to decrease as the number of epochs

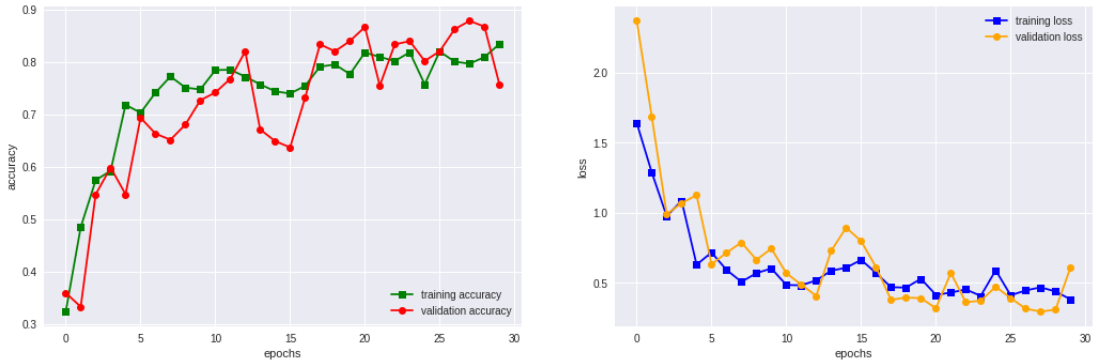

Figure B.1: Visual representation of the predictive power of the CNN on fractional Brownian motion input data.

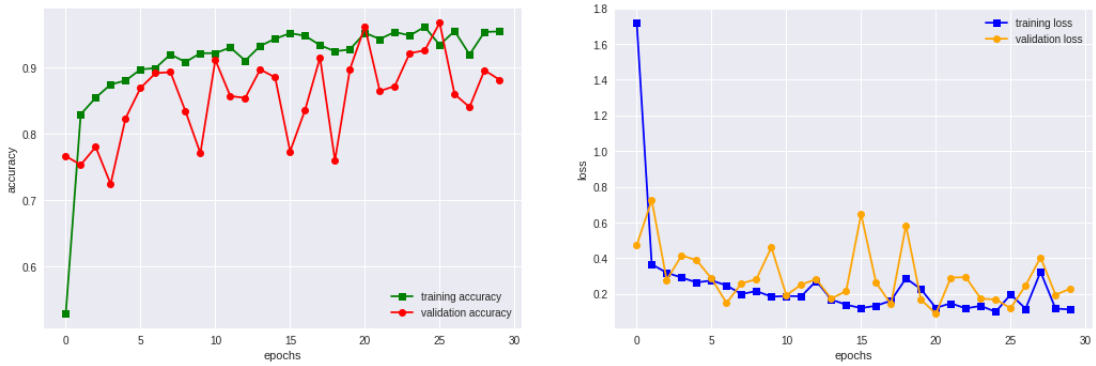

Figure B.2: Visual representation of the predictive power of the CNN on rough Bergomi input data.

increases; this again indicates predictive performance improves as the number of epochs increases, without overfitting. The only exception is Figure B.5, in the case where input length is either 300 or 500; note that these two cases also correspond to larger root mean squared error values. Recall, however, that our calibration scheme uses input vectors of length 100 so this does not pose any real problems for practical use.

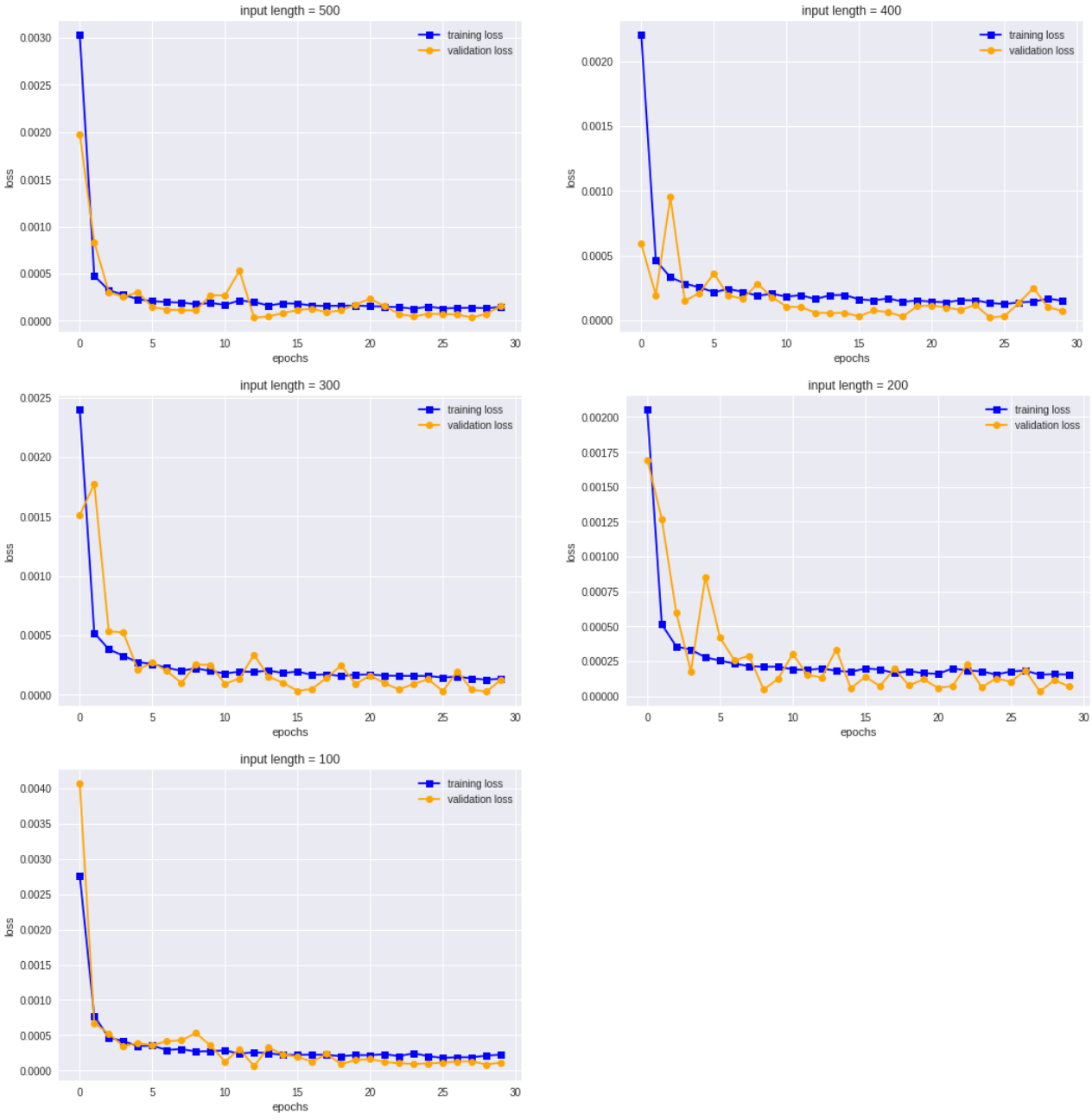

Figure B.3: Loss plots for discretised H.

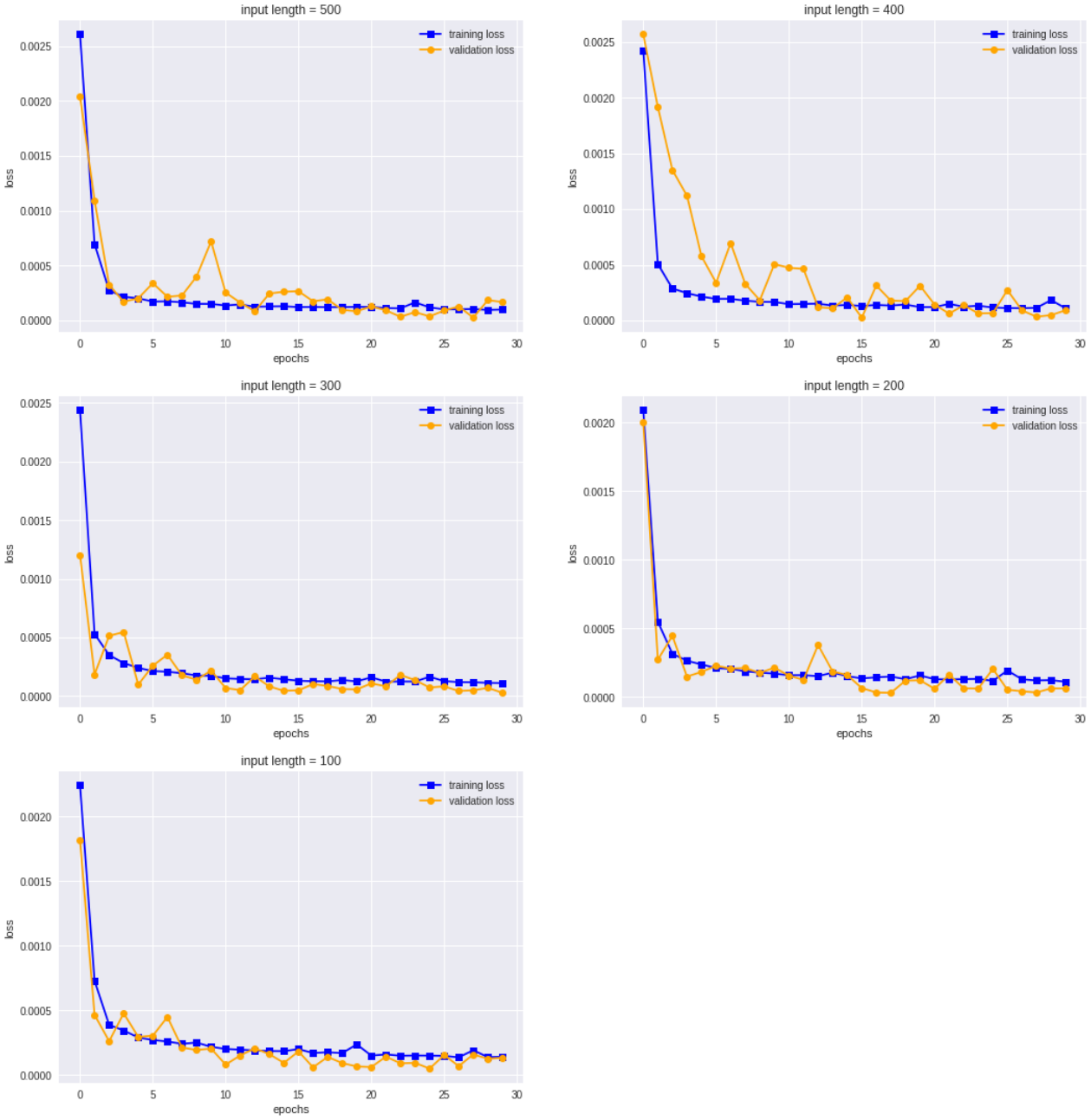

Figure B.4: Loss plots for $H \sim \text{Uniform}(0.0, 0.5)$.

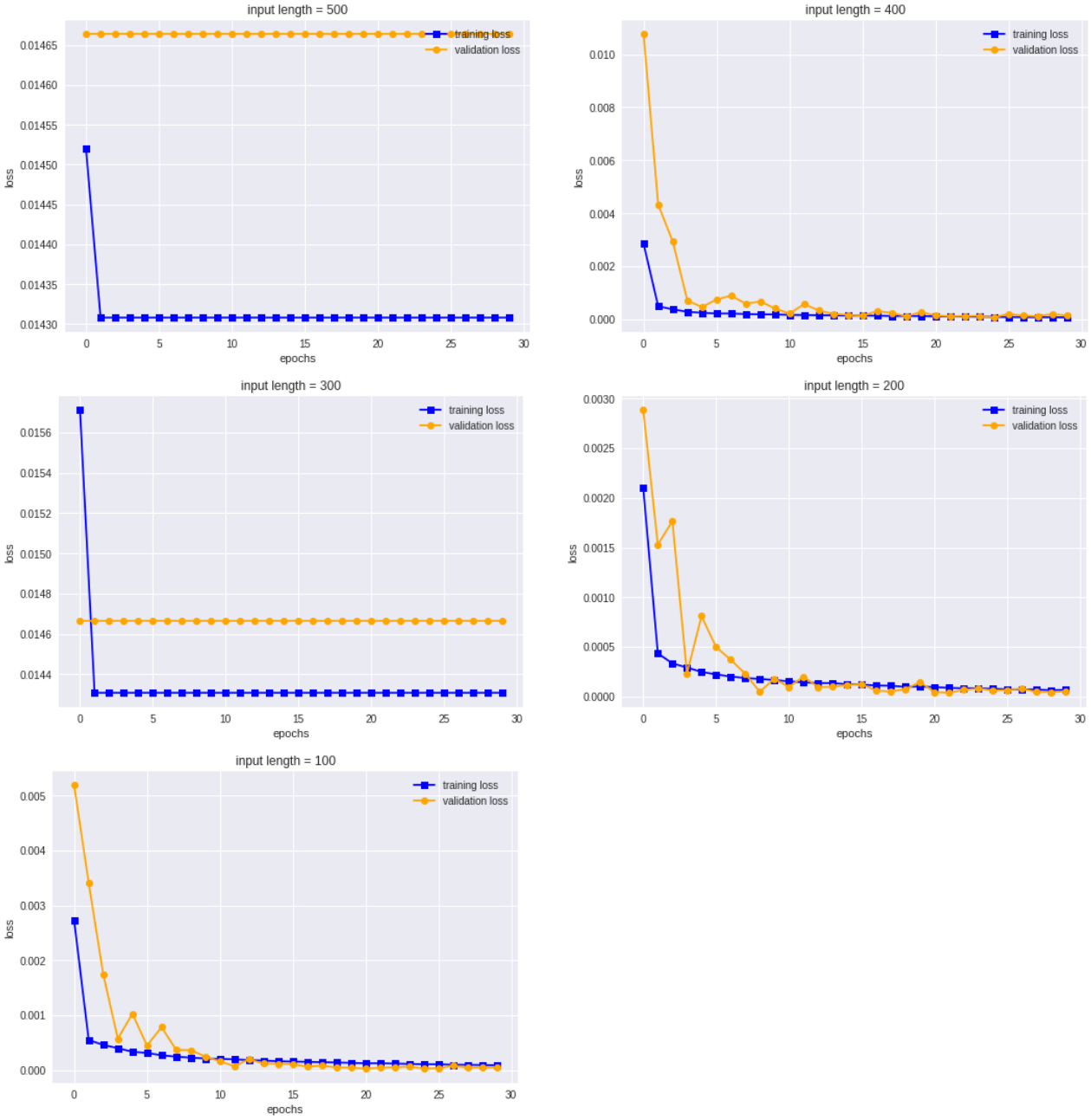

Figure B.5: Loss plots for $H \sim \text{Beta}(1, 9)$.

B.3 Loss Plots for rough Bergomi with $\eta \neq 1$

We now plot the training and validation loss (MSE) at each epoch, as above, for the rough Bergomi model with discretised H and $\eta \in \{0.25, 0.8, 1.3, 2.5\}$ in Figure B.6. We fix the input vector length to be 100. For each η value, the training loss and validation loss both decrease with each epoch.

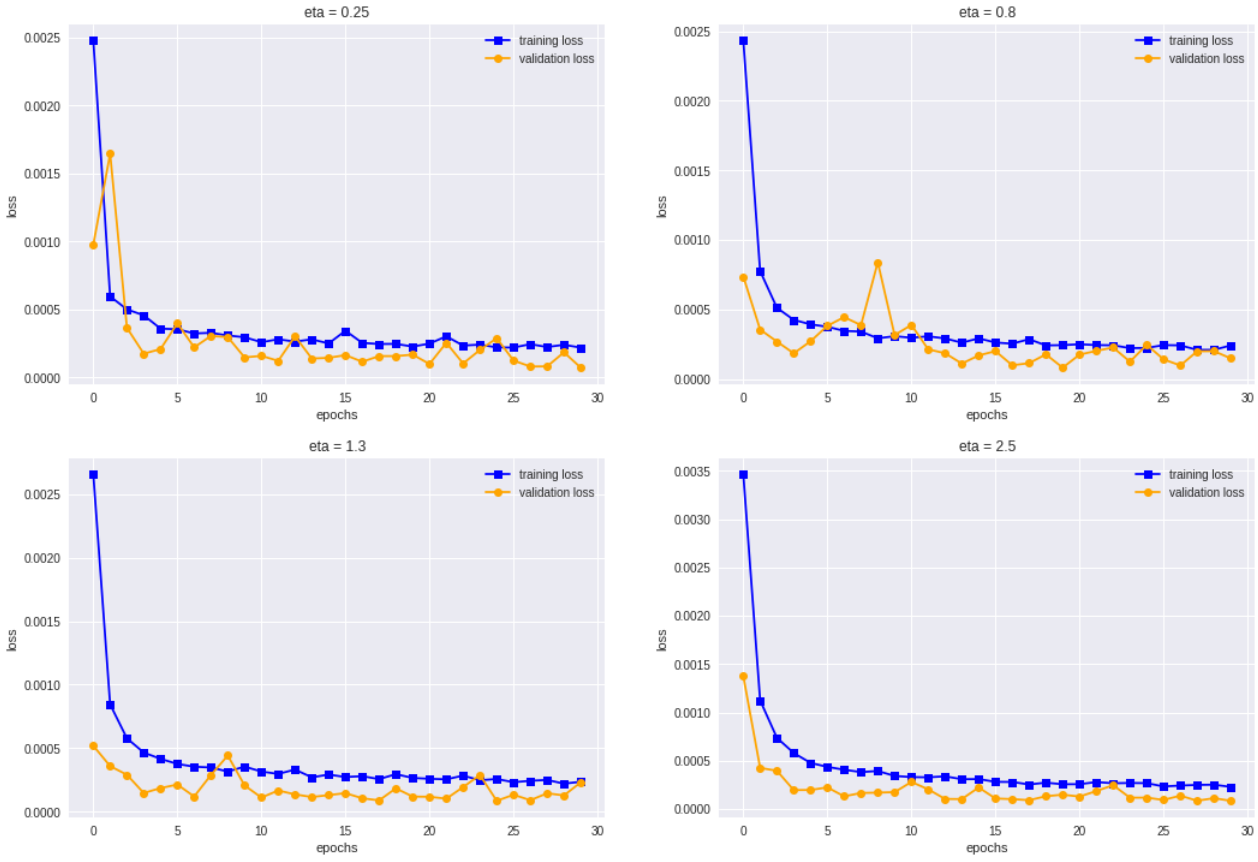

Figure B.6: Loss plots for discretised H and $\eta \in \{0.25, 0.8, 1.3, 2.5\}$.

B.4 Loss Plots for rough Bergomi with $\eta \sim$ Uniform$(0, 3)$ and $H \sim$ Beta$(1, 9)$

We now plot the training loss and validation loss (MSE) at each epoch, as above, for the rough Bergomi model with $\eta \sim \text{Uniform}(0, 3)$ and $H \sim \text{Beta}(1, 9)$ for each sample path, in Figure B.7. We fix the input vector length to be 100. Note that the training loss and validation loss both decrease with each epoch.

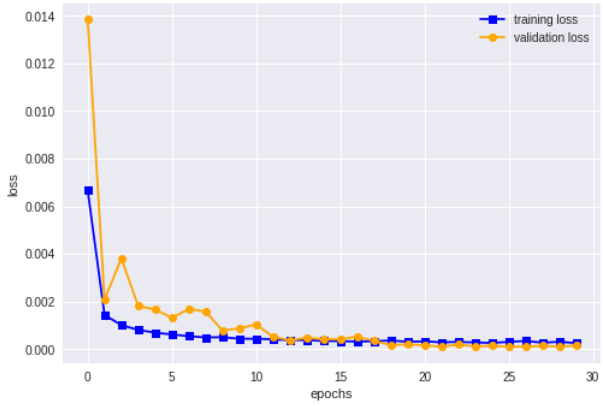

Figure B.7: Loss plots for $\eta \sim \text{Uniform}(0, 3)$ and $H \sim \text{Beta}(1, 9)$.

B.5 Loss Plots for fBm with Different H sampling

We plot the training and validation loss (MSE) at each epoch, for fBm with discretised H, $H \sim \text{Uniform}(0.0, 0.5)$, and $H \sim \text{Beta}(1, 9)$ in Figures B.8, B.9, and B.10 respectively. As in Appendix B.3 we fix the input vector length to be 100. For discretised H and $H \sim \text{Uniform}(0.0, 0.5)$ the training loss and validation loss both decrease with each epoch. For $H \sim \text{Beta}(1, 9)$, however, the training loss decreases but the validation loss remains flat. This could possibly suggest overfitting, and poor predictive performance; note that the root mean square error was indeed higher than for discretised H and $H \sim \text{Uniform}(0.0, 0.5)$.

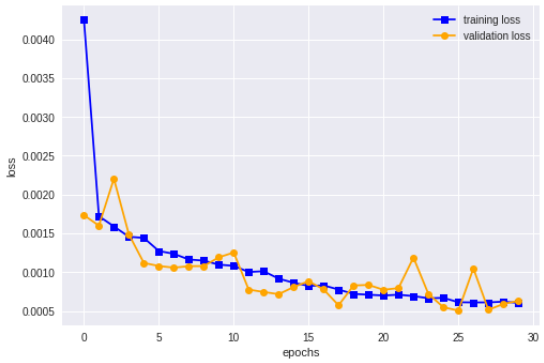

Figure B.8: Loss plot for fBm with discretised H.

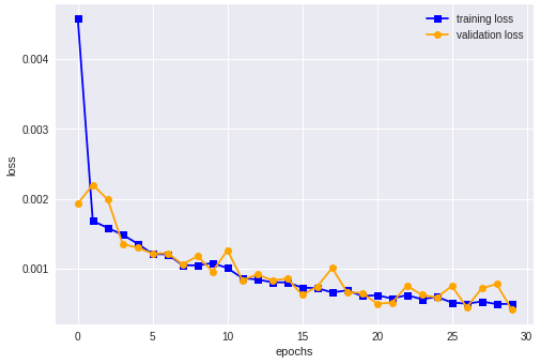

Figure B.9: Loss plot for fBm with $H \sim \text{Uniform}(0.0, 0.5)$.

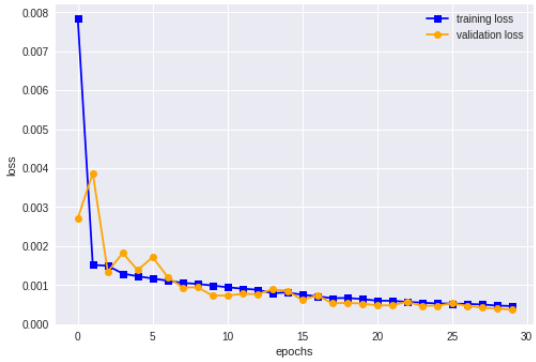

Figure B.10: Loss plot for fBm with $H \sim \text{Beta}(1, 9)$.

B.6 Loss plots for Learning η

In Figure B.11 we plot the training loss and validation loss (MSE) at each epoch, as in Appendix B.4, for the rough Bergomi model with $\eta \sim \text{Uniform}(0, 3)$ and $H \sim \text{Beta}(1, 9)$ for each sample path. Recall that in this case the CNN is learning the value of η, as well as the value of H. We fix the input vector length to be 100. Note that the training loss and validation loss both tend to decrease with each epoch.

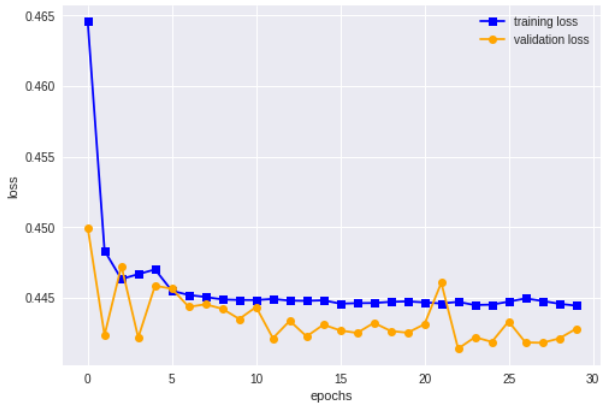

Figure B.11: Loss plots for $\eta \sim \text{Uniform}(0, 3)$ and $H \sim \text{Beta}(1, 9)$.

"But it was all right, everything was all right, the struggle was finished. He had won the victory over himself."

George Orwell, Nineteen Eighty-Four, 1949.