ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

Sandwiched Volterra Volatility models: theory and applications

Author:
Yuanxi He

Supervisors:
Josep Vives i Santa Eulàlia
David Márquez Carreras

Facultat de Matemàtiques i Informàtica

June 7th, 2026

Contents

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Josep Vives and David Márquez, for their guidance, support, and invaluable insights throughout the development of this thesis. Their expertise in stochastic analysis and mathematical finance, as well as their constant encouragement, have been essential in shaping both the direction and the quality of this work.

I am also deeply grateful to the faculty members of the Department of Mathematics and Informatics at the Universitat de Barcelona for providing a stimulating academic environment and for the knowledge I have gained throughout my studies. In particular, I would like to thank Carles Rovira and David Márquez, my professors in the courses Probability Theory and Stochastic Calculus, respectively.

Special thanks go to my colleagues, who made this journey both intellectually enriching and personally enjoyable. The many discussions, shared experiences, and mutual support have been invaluable during my time at university.

Finally, I would like to thank Richard, my partner, and my family for their unconditional love, patience, and encouragement. Their continuous support has been fundamental throughout my academic journey and beyond.

Abstract

This thesis begins with an introduction to mathematical finance, reviewing the historical development of asset-price models and several stylised facts observed in financial markets. We then motivate the study of fractional and rough volatility models, which are driven by fractional Brownian motion and provide a more realistic description of volatility dynamics. The necessary definitions and theoretical results concerning fractional Brownian motion are also presented.

Building on the work of Di Nunno et al. [15], we study the Sandwiched Volterra Volatility (SVV) model. In particular, we reformulate the general SVV framework in a one-dimensional setting, which allows for a more transparent exposition of the model and more detailed proofs of several key results. We further illustrate the model through a number of examples. In addition, we investigate the Malliavin differentiability of both the volatility and asset-price processes within the SVV framework and provide rigorous proofs of the corresponding results.

Finally, we analyse classical rough volatility models in order to characterise those that can be represented within the Sandwiched Volterra Volatility (SVV) framework. We show that most of the models studied fall into this class, with the exception of the Rough Fractional Stochastic Volatility model (for Hurst parameter $H \neq \frac{1}{2}$), the rough Heston model, and the quadratic rough Heston model.

1 Inroduction

The framework of mathematical finance began with Bachelier's doctoral thesis in 1900 [4], which proposed Brownian motion to model the evolution of asset prices in financial markets using a continuous-time framework. For this purpose, probability theory and stochastic analysis were introduced as fundamental tools of analysis in finance. Later, in the same year 1959, Osborne [44] and Samuelson [48] presented the geometric Brownian motion to model the logarithm of asset prices. In 1973, Merton and, independently, Black and Scholes developed the famous Black–Scholes formula for pricing vanilla call and put options on the underlying financial assets (see [7] and [39]):

$$C(S,t) = S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \quad \text{and} \quad P(S,t) = Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1),$$

with

$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = d_1 - \sigma\sqrt{T-t},$$

where S_t is the asset price at time $t \in [0,T]$, with t representing the current time and $T > 0$ a fixed maturity date. Furthermore, K is the strike price, r is the risk-free short-term interest rate, σ is the volatility parameter, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

The Black-Scholes formula ensures the price dynamics following the geometric Brownian motion:

$$S_t = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}, \qquad \mu \in \mathbb{R}, S_0, \sigma > 0,$$

or, as a stochastic differential equation (SDE),

$$dS_t = rS_t dt + \sigma S_t dW_t. \tag{1.1}$$

Which was introduced by Osborne [44] and Samuelson [48], by modelling the logarithm of asset prices so that the resulting price process remains strictly positive. This formulation represented a major innovation and was subsequently widely adopted by practitioners as a benchmark for the pricing and valuation of financial derivatives.

This study of price evolution was accompanied by the fundamental principle of no arbitrage, which states that in an efficient market there are no arbitrage opportunities, that is, no possibilities of generating riskless profits. Consequently, the modelling of risk, specifically the volatility, which quantifies the degree of variability of an asset's price and is statistically measured by the standard deviation, became a central focus of financial modelling.

However, the financial crisis of Black Monday in 1987 showed that the Osborne and Samuelson model's assumption that the logarithms of asset prices follow a Brownian motion is not correct, so the prices of vanilla options in derivatives markets do not coincide with the predictions of the Black-Scholes model. One of the important false assumptions was that the

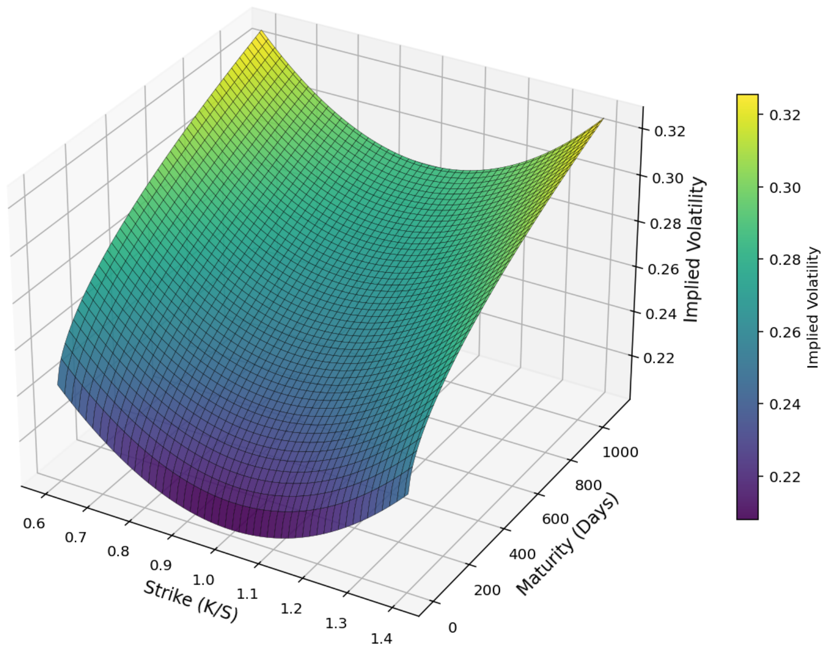

Figure 1: Implied Volatility Surface using simulated data

volatility is constant. So, to solve this divergence and to indicate the real market price of a vanilla option, one would have to use a different volatility, which is known as the implied volatility. It consists of looking for the volatility that, when placed in the Black-Scholes formula, gives us the real value of the option, which is a function of the strike price and the time to expiry date of the option, creating the volatility surface, and the shape of these functions generates the so-called volatility smile phenomenon in the financial market (see Figure 1).

Observation. An important qualitative property of the Black–Scholes formulas is that both the call and put option prices are strictly increasing functions of the volatility parameter σ. More precisely,

$$\frac{\partial C}{\partial \sigma} > 0 \qquad \text{and} \qquad \frac{\partial P}{\partial \sigma} > 0,$$

for $t < T$. Indeed, differentiating the Black–Scholes formulas with respect to σ gives

$$\frac{\partial C}{\partial \sigma} = \frac{\partial P}{\partial \sigma} = S_t\sqrt{T-t}\,\varphi(d_1),$$

where $\varphi(\cdot)$ denotes the density function of the standard normal distribution,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Since $\varphi(d_1) > 0$ for every $d_1 \in \mathbb{R}$, together with $S_t > 0$ and $T - t > 0$, it follows that

$$\frac{\partial C}{\partial \sigma} > 0 \qquad \text{and} \qquad \frac{\partial P}{\partial \sigma} > 0.$$

Therefore, both call and put option prices are strictly increasing with respect to the volatility σ. Intuitively, greater volatility increases the probability of large favorable movements in the underlying asset price, while the holder's losses remain limited by the option payoff structure.

Furthermore, other empirical evidences show some properties of the asset prices and their volatility, such as fat tails and non-Gaussian distribution of the log-returns of the financial assets, the leverage effect, that is, there is a negative correlation between the dynamics of the volatility and the return of asset prices, the clustering of high and low volatility seasons presenting a mean-reversion effect that is that the volatility tends to return to the mean level of its long-run distribution.

In addition, the log-return defined by

$$r_t := \log\left(\frac{S(t+\Delta)}{S(t)}\right)$$

exhibits long-range dependence, i.e. the autocorrelation function of the absolute log-return process

$$\rho(\tau) := \mathrm{Corr}(|r_t|, |r_{t+\tau}|)$$

remains positive while decaying slowly, typically as

$$\rho(\tau) \sim C\, \tau^{-\beta} \quad \text{as } \tau \to \infty,$$

for some constant $C > 0$ and $\beta \in (0, \frac{1}{2}]$.

From that moment, the need to develop appropriate pricing formulas for financial derivatives became evident. It was therefore essential to construct more general models capable of accounting for non-constant volatility. Among the first extensions of the Black-Scholes framework was the constant elasticity of variance (CEV) model introduced by Cox [13] in 1976. Later, local volatility models were proposed independently by Derman and Kani [14] and Dupire [18]. These models assume that volatility is a deterministic function of the current asset price and time. However, despite their greater flexibility, they still proved insufficient to fully capture the complex dynamics observed in financial markets.

An alternative approach is provided by stochastic volatility models, which are the main subject of this thesis. In these models, volatility itself is assumed to follow a stochastic process, possibly correlated with the underlying asset price process, allowing a more realistic description of market phenomena.

Firstly, we can rewrite equation (1.1) by introducing an appropriate stochastic volatility process $\{\sigma_t\}_{t\geq 0}$ instead of the deterministic coefficient σ:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t, \tag{1.2}$$

where μ_t is called the drift process and σ_t the diffusion process.

We naturally assume the diffusion process, i.e., the volatility, to be non-negative. This process may be driven by another Brownian motion, denoted by $\{B_t\}_{t\geq 0}$, correlated with the Brownian motion W_t appearing in (1.2). In order to account for the leverage effect, we assume an appropriate dependence structure between B_t and W_t, namely $\mathbb{E}[W_t B_t] = \rho t$ with $-1 < \rho < 0$.

Another important property of stochastic volatility models is mean reversion, which allows one to capture the volatility clustering effect. A common approach is to consider a drift term of the form $\theta_1(\theta_2 - \sigma_t)$ in the SDE for the volatility. This drift pulls σ_t back toward the level θ_2 whenever it deviates from it. The parameter θ_1 determines the speed of mean reversion.

Here, we present some classical stochastic volatility models from the literature, among many others.

- Hull and White (1987) [31] introduced a model in which the squared volatility $\{\sigma_t^2\}_{t\geq 0}$ itself follows a geometric Brownian motion. In this case, the price and volatility satisfy the stochastic differential equations

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t,$$
$$d\sigma_t^2 = \theta_1 \sigma_t^2 dt + \theta_2 \sigma_t^2 dB_t,$$

  respectively, where B_t and W_t are Brownian motions that may be correlated in order to capture the leverage effect. Note that the volatility process is strictly positive but not mean-reverting.

- Stein and Stein (1991) [49] considered the volatility to follow an Ornstein–Uhlenbeck process, i.e.,

$$dS_t = \mu S_t dt + v_t S_t dW_t,$$
$$dv_t = \theta_1(\theta_2 - v_t)dt + \theta_3 dB_t,$$

  where μ, θ_1, θ_2 and θ_3 are fixed constants. Note that v_t is the instantaneous variance. Since the Ornstein–Uhlenbeck process is Gaussian, it can take negative values with positive probability. This issue may be addressed by considering the absolute value of σ_t or by introducing a reflecting barrier in the volatility dynamics.

- The Heston model (1993) [30] is defined by the system of stochastic differential equations

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sqrt{v_t}\,S_t\,dW_t, \\ dv_t &= \theta_1(\theta_2 - v_t)\,dt + \theta_3\sqrt{v_t}\,dB_t. \end{aligned} \tag{1.3}$$

  In this framework, the volatility process σ_t follows a Cox–Ingersoll–Ross (CIR) square-root process, originally introduced by Cox, Ingersoll and Ross (1985) [12]. This specification ensures strict positivity of the volatility, provided that the Feller condition $2\theta_1\theta_2 \geq \theta_3^2$ holds. The volatility evolves as a mean-reverting stochastic process and is typically assumed to be correlated with the Brownian motion driving the asset price.

We first define the risk-neutral measure and show how it can be used to obtain a driftless representation of the model.

Definition 1.1 (Risk-neutral measure)**.** *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space and let $\{S_t^0\}_{t\geq 0}$ denote the money market account. A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is called a* risk-neutral measure *(or* equivalent martingale measure*) if*

$$\mathbb{Q} \sim \mathbb{P},$$

and for every traded asset price process $\{S_t\}_{t\geq 0}$, the discounted price process

$$\tilde{S}_t := \frac{S_t}{S_t^0}$$

is a $\mathbb{Q}$-martingale. Equivalently,

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{S_T}{S_T^0}\,\Big|\,\mathcal{F}_t\right] = \frac{S_t}{S_t^0}, \qquad 0 \leq t \leq T.$$

Remark 1.2. *In the above models, the drift term $\mu S_t\,dt$ can be removed by passing to discounted prices or, equivalently, by working under the risk-neutral measure. Let $\{S_t^0\}_{t\geq 0}$ denote the money market account, defined by*

$$dS_t^0 = rS_t^0\,dt, \qquad S_0^0 = 1.$$

The discounted asset price is

$$\tilde{S}_t := \frac{S_t}{S_t^0} = e^{-rt}S_t.$$

Under the risk-neutral measure $\mathbb{Q}$, the process $\tilde{S}_t$ is a martingale and satisfies

$$d\tilde{S}_t = \sigma_t\tilde{S}_t\,dW_t^{\mathbb{Q}},$$

where $W^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. Hence, the drift term vanishes.
Equivalently, under $\mathbb{Q}$,

$$F_t := \mathbb{E}^{\mathbb{Q}}[S_T \mid \mathcal{F}_t] = S_t e^{r(T-t)}.$$

Since F_t is itself a conditional expectation under $\mathbb{Q}$, it is a $\mathbb{Q}$-martingale and therefore evolves without drift.

- Hagan et al. (2002) [28] introduced the stochastic alpha-beta-rho (SABR) model of the form

$$
\begin{aligned}
dF_t &= v_t F_t^\beta dW_t, \\
dv_t &= \alpha v_t dB_t,
\end{aligned}
$$

  where $\alpha \geq 0$, $0 \leq \beta \leq 1$, and $\mathbb{E}[W_t B_t] = \rho t$ with $-1 < \rho < 1$. Note that the case $\beta \in \left(0, \frac{1}{2}\right)$ is mathematically delicate and requires special treatment, see [9].

- Gatheral (2008) [24] proposed the double CEV model of the form

$$
\begin{aligned}
dF_t &= F_t \sqrt{v_t} dW_t, \\
dv_t &= -\kappa(v_t - v_t')dt + \eta_1 v_t^\alpha dB_t, \\
dv_t' &= -c(v_t' - \theta)dt + \eta_2 (v_t')^\beta dB_t',
\end{aligned}
$$

  where $\alpha, \beta \in \left[\frac{1}{2}, 1\right]$. Notice that the case when $\alpha = \beta = 1/2$ is a double Heston model, the case when $\alpha = \beta = 1$ is a double Lognormal model, and the general case is the double CEV. All such models involve a short-term variance level v that reverts to a moving level v' at rate κ. v' reverts to the long-term level θ at the slower rate $c < \kappa$.

These stochastic volatility models are widely used in both academia and industry, as they provide a realistic description of the implied volatility surfaces observed in financial markets.

Nevertheless, these models exhibit important limitations. In particular, they fail to capture long-range dependence in the volatility dynamics, as highlighted by Comte and Renault (1998) [10]. Empirical analyses of financial time series reveal persistent autocorrelations in volatility over long time horizons. This persistence is reflected in option markets, where the amplitude of the implied volatility smile decays much more slowly with maturity than predicted by standard Brownian diffusion models, suggesting the presence of long-memory effects, see [25]. This discrepancy reflects the intrinsic limitations of Brownian motion, whose structural properties constrain the modelling capabilities of standard diffusion-based stochastic volatility frameworks. Consequently, substantial research efforts have been devoted to extending the stochastic volatility paradigm in order to overcome these shortcomings.

2 Motivation

An alternative approach to modelling volatility is based on fractional processes, an idea initiated by Comte and Renault (1998) [10], who used fractional Brownian motion (fBm) with Hurst index $H \geq \frac{1}{2}$ to describe the long-term behaviour of the volatility surface.

One way to extend the stochastic volatility models described before is to replace the standard Brownian driver B with a process capable of capturing the desired stylised facts. Perhaps the most commonly used choice in the literature is fractional Brownian motion $B^H = \{B_t^H\}_{t\geq 0}$, which is a generalisation that deviates significantly from standard Brownian motion and semimartingales. It is closely based on fractional calculus. For a complete reference on the topic, we refer to the excellent book by Biagini et al. [6].

Definition 2.1. *Let $H \in (0,1)$. A fractional Brownian motion (fBm) of Hurst parameter H is a continuous and centered Gaussian process $B^H = \left\{B_t^H\right\}_{t\geq 0}$ with the covariance function*

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

It admits the Volterra-type representation

$$B_t^H = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\int_{-\infty}^{0}\left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right)dB_s + \frac{1}{\Gamma\left(H+\frac{1}{2}\right)}\int_0^t (t-s)^{H-\frac{1}{2}}dB_s,$$

where $B = \{B_t\}_{t\geq 0}$ is a standard Brownian motion.

Fractional Brownian motion is the only stochastic process that satisfies the following properties; for further details, see [36], [38] and the Section 3 in this thesis.

1. $B_0^H = 0$ a.s. and $\mathbb{E}[B_t^H] = 0$ for all $t \geq 0$.
2. B^H is a Gaussian process and $\mathbb{E}[(B_t^H)^2] = t^{2H}$, $t \geq 0$, for all $H \in (0,1)$.
3. **Self-similarity**. For all $a > 0$, $\left(a^{-H}B_{at}^H\right)_{t\geq 0} \overset{\text{law}}{=} \left(B_t^H\right)_{t\geq 0}$.
4. **Stationarity of increments**. For all $h > 0$, $\left(B_{t+h}^H - B_h^H\right)_{t\geq 0} \overset{\text{law}}{=} \left(B_t^H\right)_{t\geq 0}$.
5. **Correlation of increments**. For $H \neq \frac{1}{2}$ and for all $h > 0$, the increments are not independent. Indeed, the covariance function between $B_{t+h}^H - B_t^H$ and $B_{s+h}^H - B_s^H$ with $s + h \leq t$ and $t - s = nh$ is

   $$\rho_H(n) = \frac{1}{2}h^{2H}\left[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\right],$$

   which depends on the Hurst parameter. We have three cases:

- If $H > \frac{1}{2}$, then the increments are positively correlated. This is usually called long-memory dependence and persistence (proved by Theorem 3.2 below).
- If $H = \frac{1}{2}$, then the increments are independent.
- If $H < \frac{1}{2}$, then the increments are negatively correlated.

Remark 2.2. *Clearly, when $H = 1/2$, the fBm coincides with a standard Brownian motion, and when $H = 1$, we have that $B_t^H = tB_1^H$ almost surely for all $t \geq 0$.*

Proof. We take the covariance of the fBm for $H = 1/2$, for $s < t$,

$$\mathbb{E}[B_t^{\frac{1}{2}} B_s^{\frac{1}{2}}] = \frac{1}{2}\left(t + s - |t - s|\right) = s,$$

which is the same as for the standard Brownian motion.

When H=1, we take the variance of both terms, since B^H is a centered Gaussian process, then $\mathbb{E}[B_t^H] = 0$.

$$\begin{aligned}
\mathbb{E}\left[\left(B_t^H - tB_1^H\right)^2\right] &= \mathbb{E}\left[\left(B_t^H\right)^2 - 2tB_t^H B_1^H + t^2\left(B_1^H\right)^2\right] \\
&= \mathbb{E}\left[\left(B_t^H\right)^2\right] - 2t\mathbb{E}\left[B_t^H B_1^H\right] + t^2\mathbb{E}\left[\left(B_1^H\right)^2\right] \\
&= t^2 - 2t\left[\frac{1}{2}\left(t^2 + 1 - (1-t)^2\right)\right] + t^2 \\
&= 2t^2 - 2t^2 = 0,
\end{aligned}$$

therefore, $B_t^H = tB_1^H$ almost surely. □

In volatility modelling, two distinct regimes are typically considered: $H > \frac{1}{2}$ and $H < \frac{1}{2}$.

Fractional models with long memory $\left(H > \frac{1}{2}\right)$. Observe that

$$\mathbb{E}[B_1^H\left(B_n^H - B_{n-1}^H\right)] = \frac{1}{2}\left(n^{2H} - 2(n-1)^{2H} + (n-2)^{2H}\right) \sim H(2H-1)n^{2H-2}, \tag{2.1}$$

as $n \to \infty$.

Hence, if $H \in \left(\frac{1}{2}, 1\right)$, the autocorrelation function behaves as $O(n^{-\beta})$ with $\beta \in (0,1)$, revealing the long-memory property. In particular, if $H \in \left(\frac{3}{4}, 1\right)$, the behaviour in (2.1) matches empirical estimates $0 < \beta \leq \frac{1}{2}$ for absolute log-returns highlighted previously.

Comte and Renault [10] proposed the first continuous-time fractional volatility model of the form

$$\sigma_t = \theta_1 \exp\left\{\theta_2 \int_0^t e^{-\theta_3(t-s)}\, dB_s^H\right\},$$

which is capable of reproducing volatility persistence and explaining the slow decay of the smile amplitude of implied volatility surfaces as $T \to \infty$.

We now present several relevant contributions concerning stochastic volatility models driven by fractional Brownian motion with $H > \frac{1}{2}$.

- Rosenbaum (2008) [46] considers the model

$$\begin{aligned} dS_t &= \sigma_t dW_t, \\ \sigma_t &= F\left(\int_0^t a(t,u)\, dB_u^H + f(t)\xi_0\right), \end{aligned}$$

  where F, a, and f are nuisance parameters and ξ_0 is a random initial condition.

- Mishura and Yurchenko-Tytarenko [40], [41] and [42] proposed

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma(Y_t) S_t dW_t, \\ dY_t &= \frac{1}{2}\left(\frac{\theta_1}{Y_t} - \theta_2 Y_t\right) dt + \frac{\theta_3}{2} dB_t^H, \end{aligned}$$

  where $\mu, \theta_1, \theta_2, \theta_3 > 0$ and σ is a given function with sublinear growth. This model can be regarded as a fractional extension of the Heston model (1.3), since the process $X_t = Y_t^2$ satisfies the pathwise SDE

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} dB_t^H.$$

As a final remark, we emphasise the intrinsic connection between long memory and self-similarity in fractional Brownian motion. Quantifying long-range dependence is statistically challenging, as it typically requires the analysis of very long time series, which may be affected by non-stationarity. However, when the data are assumed to possess self-similarity, information about their long-term behaviour can be extracted from high-frequency observations over comparatively short time horizons.

Rough volatility models $\left(H < \frac{1}{2}\right)$. In Alòs et al. [1], the short-term behaviour of the volatility surface was studied, showing that fractional noises with Hurst index less than $\frac{1}{2}$ replicate key short-term features of the implied volatility surface. This pioneering work initiated a vast literature in which such fractional models are referred to as rough volatility models.

Empirical evidence suggests that volatility observed in markets is best described by models driven by fractional processes with a small Hurst index. This leads naturally to generalisations of classical stochastic volatility equations, such as the Heston model, into the rough regime. Rough volatility models are particularly effective for short-maturity data.

For instance, Alòs et al. ([1], Section 7.2.1) analysed the short-maturity at-the-money implied volatility skew

$$\Psi(t,T) := \left| \frac{\partial}{\partial \kappa} \hat{\sigma}(t,T,\kappa) \right|_{\kappa=0},$$

where $\kappa := \log \frac{K}{e^{r(T-t)}S_t}$ denotes the log-moneyness. They proved that, for $H > \frac{1}{2}$, the skew at fixed t does not behave as $O((T-t)^{-\beta})$ with $\beta \approx \frac{1}{2}$ as $T \downarrow t$. However, in Section 7.2.2, they showed that a stochastic volatility model driven by Riemann–Liouville fractional Brownian motion with $H \in \left(0, \frac{1}{2}\right)$ produces the correct skew asymptotics $O((T-t)^{-\frac{1}{2}+H})$ as $T \downarrow t$. Gatheral et al. [25] further concluded that fractional Brownian motions with a very small Hurst index provide the most accurate description of volatility dynamics.

Some notable contributions in this area include:

- Gatheral et al. (2014) [25] proposed the rough fractional stochastic volatility (RFSV) model

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t dW_t, \\ \sigma_t &= \exp\left\{\theta_1 + \theta_2 \int_{-\infty}^{t} e^{-\theta_3(t-s)}\, dB_s^H\right\}, \end{aligned}$$

- Bayer et al. (2015) [5] and Jacquier et al. (2017) [34] introduced the rough Bergomi model

$$\begin{aligned} dS_t &= -\frac{1}{2} v_t S_t dt + \sqrt{v_t} S_t dW_t, \\ v_t &= \theta_0(t) \exp\left\{\int_0^t (t-s)^{H-\frac{1}{2}}\, dB_s - \theta_1^2 t^{2H}\right\}, \end{aligned}$$

  where $\theta_1 > 0$ and θ_0 is a deterministic function.

- Guyon (2019) [27] and Lacombe et al. (2021) [37] proposed the mixed rough Bergomi model

$$\begin{aligned} dS_t &= -\frac{1}{2} v_t S_t dt + \sqrt{v_t} S_t dB_t, \\ v_t &= \theta_0(t) \Big(\delta \exp\left\{2\theta_1 \int_0^t (t-s)^{H-\frac{1}{2}}\, dW_s - \theta_1^2 t^{2H}\right\} \\ &\quad + (1-\delta) \exp\left\{2\theta_2 \int_0^t (t-s)^{H-\frac{1}{2}}\, dW_s - \theta_2^2 t^{2H}\right\}\Big), \end{aligned}$$

  where $\delta \in [0,1], \theta_1, \theta_2 > 0$ and θ_0 is a deterministic function.

- Fukasawa and Gatheral (2022) [22] proposed the rough SABR model

$$\begin{aligned} dF_t &= \sqrt{v_t} \beta(F_t) dW_t, \\ v_t &= \theta_0(t) \exp\left\{\theta\sqrt{2H} \int_0^t (t-u)^{H-\frac{1}{2}}\, dB_u - \frac{1}{2}\theta^2 t^{2H}\right\}, \end{aligned}$$

where θ_0 is a positive process and β is a positive continuous function.

- Harms and Stefanovits (2019) [29] introduced the rough Stein–Stein model

$$
\begin{aligned}
dF_t &= \theta v_t F_t dW_t, \\
v_t &= \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \int_0^t (t-s)^{H-\frac{1}{2}}\, dB_s.
\end{aligned}
$$

- El Euch and Rosenbaum (2019) [19], who introduced the rough Heston model

$$
\begin{aligned}
dF_t &= \sqrt{v_t} F_t dW_t, \\
v_t &= v_0 + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \left(\theta_1(\theta_2 - v_s)\, ds + \theta_3 \sqrt{v_s}\, dB_s\right),
\end{aligned}
\tag{2.2}
$$

  and its quadratic extension (see [26] and [47])

$$
\begin{aligned}
dF_t &= \sqrt{v_t} F_t dW_t, \\
v_t &= a(Z_t - b)^2 + c, \\
Z_t &= \int_0^t \theta_1 \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \left(\theta_0(s) - Z_s\right)\, ds + \int_0^t \theta_2 \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \sqrt{a(Z_s-b)^2+c}\, dW_s,
\end{aligned}
$$

  where B, W are Brownian motions, $a, b, c, \theta_1, \theta_2, \theta_3 > 0$ and θ_0 is a deterministic function.

The rough Heston model (2.2) is especially attractive from a modelling standpoint. Besides reproducing the power-law behaviour of

$$
\Psi(T) := \left| \frac{\partial}{\partial \kappa} \sigma(T, \kappa) \right|_{\kappa=0} \sim O\left(T^{-\frac{1}{2}+H}\right), \text{ for } H \in \left(0, \frac{1}{2}\right),
$$

it is capable of capturing a particular manifestation of the Zumbach effect (see [52]). Moreover, it can be interpreted as the scaling limit of a tick-by-tick price model driven by two-dimensional Hawkes processes, thus providing a microstructural foundation for the emergence of volatility roughness.

This type of model seems especially interesting since, on the one hand, the theoretical result of Fukasawa [21] shows that, in a continuous semimartingale framework, the combination of a power-law behaviour of implied volatility and absence of arbitrage implies that the volatility paths must exhibit low Hölder regularity—comparable to that of fractional Brownian motion with $H < \frac{1}{2}$. On the other hand, Alòs et al. [1] show that, in Malliavin differentiable stochastic volatility models, the observed explosive behaviour of option sensitivities is driven by the divergence of the Malliavin derivative of the volatility process. Conveniently, fractional Brownian motion with $H < \frac{1}{2}$ exhibits precisely this type of irregularity.

From a statistical perspective, however, the interpretation of rough volatility is more subtle. Cont and Das [11] investigate empirical evidence for rough behaviour in volatility using a model-free estimation procedure based on normalised p-th variations. Their findings suggest that high-frequency estimates of volatility systematically exhibit a Hurst exponent $H < \frac{1}{2}$, even when the underlying latent volatility process is generated by a standard diffusion with smooth sample paths. This indicates that the observed roughness may partly arise as an artefact of the estimation of integrated volatility rather than reflecting the true regularity of the instantaneous volatility process. Taken together with the theoretical results of Fukasawa [21], these findings highlight a potential dichotomy between statistical estimation effects and structural no-arbitrage constraints, thereby reinforcing the relevance of rough fractional models.

3 Relevant properties of fractional Brownian motion

Having already defined fractional Brownian motion, we now present some of its key properties that are related to the stylised facts discussed in Sections 1 and 2. For completeness, we provide direct proofs of some results, while for others we refer the reader to the existing literature, as detailed proofs fall outside the scope of this thesis.

Definition 3.1. *A stationary sequence $\{X_n\}_{n\in\mathbb{N}}$ exhibits long-range dependence if the auto-covariance functions $\rho(n) := cov(X_k, X_{k+n})$ satisfy*

$$\lim_{n\to\infty} \frac{\rho(n)}{cn^{\alpha}} = 1,$$

for some constant c and $\alpha \in (0,1)$. In this case, the dependence between X_k and X_{k+n} decays slowly as $n \to \infty$ and

$$\sum_{n=1}^{\infty} \rho(n) = \infty.$$

Theorem 3.2 (Long-range dependence)**.** *Let $X_k := B_k^H - B_{k-1}^H$ and $X_{k+n} := B_{k+n}^H - B_{k+n-1}^H$, and for $H \in (\frac{1}{2}, 1)$ we have the long-range dependence property, and*

$$\rho_H(n) = \frac{1}{2}\left[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\right] \sim H(2H-1)n^{2H-2},$$

as $n \to \infty$ and therefore

$$\begin{cases} \sum_{j=1}^{\infty} |\rho_H(j)| < \infty & \text{for } H < \frac{1}{2} \\ \sum_{j=1}^{\infty} |\rho_H(j)| = \infty & \text{for } H > \frac{1}{2} \end{cases}.$$

Proof. We begin the proof by rewriting the autocovariance function

$$\begin{aligned} \rho_H(n) &= \frac{1}{2}\left[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\right] \\ &= \frac{1}{2}n^{2H}\left[\left(1+\frac{1}{n}\right)^{2H} + \left(1-\frac{1}{n}\right)^{2H} - 2\right] \\ &= \frac{1}{2}n^{2H} f(n^{-1}), \end{aligned}$$

where $f(x) = (1+x)^{2H} + (1-x)^{2H} - 2$.

Applying the Taylor expansion of $f(x)$ at the origin, we will get

$$\begin{aligned} f(x) &= f(0) + f'(0)x - + \frac{1}{2}f''(0)x^2 + \mathcal{O}(x^3) \\ &= 0 + (2Hx - 2Hx) + \frac{1}{2}\left[2H(2H-1) + 2H(2H-1)\right]x^2 + \mathcal{O}(x^3) \\ &= \frac{1}{2}\left[4H(2H-1)\right]x^2 + \mathcal{O}(x^3) \\ &= 2H(2H-1)x^2 + \mathcal{O}(x^3). \end{aligned}$$

Therefore, as $n \to \infty$, we have

$$\rho_H(n) = \frac{1}{2} n^{2H} 2H(2H-1)n^{-2} = H(2H-1)n^{2H-2}.$$

We conclude the proof by Definition 3.1, we have

$$\sum_{n=1}^{\infty} \rho_H(n) = \infty,$$

since $\alpha = 2 - 2H \in (0,1)$ only if $H \in \left(\frac{1}{2}, 1\right)$. $\square$

Theorem 3.3 (Hölder continuity)**.** *Let $\{B_t^H\}_{t\geq 0}$ be a fractional Brownian motion with Hurst parameter $H \in (0,1)$. Then, the fBm is almost surely Hölder continuous for any order $\alpha < H$. That is, there exists a constant $C > 0$ such that for all $s, t \geq 0$,*

$$|B_t^H - B_s^H| \leq C|t-s|^{\alpha}.$$

Proof. We refer to Section 1.6 of Biagnini et al. [6], where this is proved by applying the Kolmogorov criterion. $\square$

Theorem 3.4. *Let $\{B_t^H\}_{t\geq 0}$ be a fractional Brownian motion with Hurst parameter $H \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$. Then, $\{B_t^H\}_{t\geq 0}$ is not a semimartingale, nor a Markov process.*

Proof. For the proof, we refer to the article by Roger [45] computing the p-variation of the fBm process and analysing its asymptotic behaviour. $\square$

Nonetheless, in the paper by Cheridito [8] it is introduced and proved that for a fBm $\{B_t^H\}_{t\geq 0}$ with a Hurst parameter $H \in \left(\frac{3}{4}, 1\right)$, and with $\{W_t\}_{t\geq 0}$ an independent standard Brownian motion, the process $M_t^H := B_t^H + W_t$ is a weak semimartingale.

Proposition 3.5. *Let $H \in (0,1)$. Almost all sample paths of fractional Brownian motion $B^H(t,\omega)$ are not differentiable for all $t_0 \in [0,\infty)$. In fact, fixing $\omega \in \Omega$,*

$$\limsup_{t \to t_0} \left| \frac{B_t^H - B_{t_0}^H}{t - t_0} \right| = \infty,$$

with probability one.

Proof. We refer to the paper by Mandelbrot and Van Ness [38]. $\square$

Simulations. Here we simulate several sample paths of fractional Brownian motion for different values of the Hurst parameter H. As shown in Figure 2, one can clearly observe that the smoothness of the paths increases with H: for $H > \frac{1}{2}$ the increments are positively correlated, leading to smoother trajectories, whereas for $H < \frac{1}{2}$ they are negatively correlated, resulting in rougher paths.

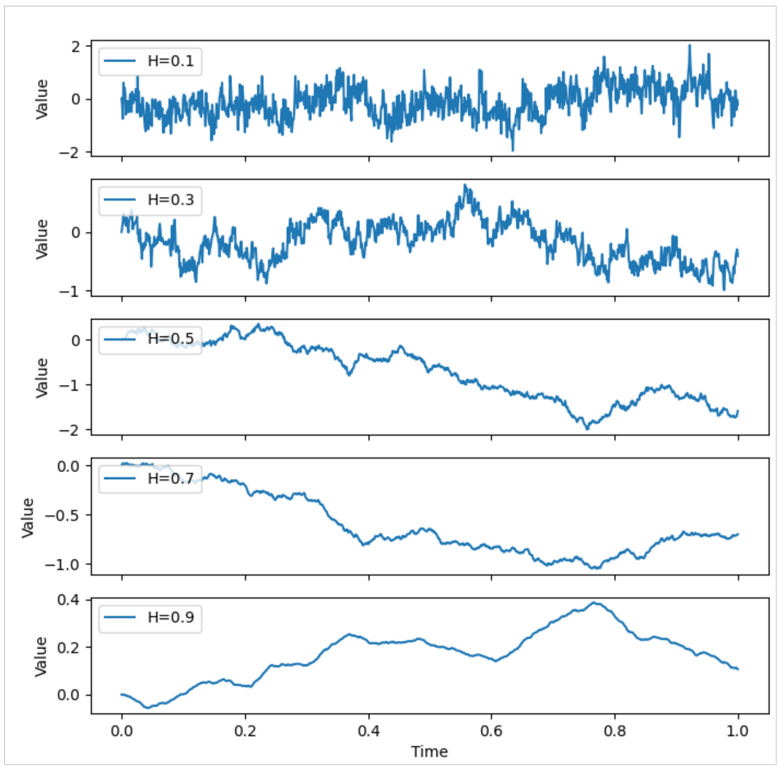

Figure 2: Simulation of fractional Brownian motion for different Hurst parameters H

4 The Sandwiched Volterra Volatility Model

In this section, we will introduce a family of sandwiched Volterra volatility (SVV) models based on the framework developed by Di Nunno, et al.[15]. This class of models extends the classical rough volatility paradigm by incorporating a general Volterra convolution structure that captures the long-range dependence and self-similarity properties mentioned in Section 1. Specifically, the volatility process is defined as a Volterra integral of a Lévy noise, allowing for both Gaussian and non-Gaussian driving mechanisms, which provides greater flexibility in modelling the stylised facts of asset returns such as leverage effects, volatility clustering, and heavy tails. The authors establish fundamental properties of these processes, including existence, uniqueness, and Malliavin differentiability under suitable regularity conditions on the kernel and drift coefficients, detailed in Section 5.

Our analysis builds upon this foundational work by examining the joint dynamics of asset price and volatility within a one-dimensional setting, where the correlation structure between the price and volatility innovations plays a crucial role in determining the implied volatility surface. Moreover, we investigate the small-time asymptotic behaviour of the at-the-money volatility skew, establishing a power law that generalises known results for fractional volatility models to the broader class of Volterra-driven processes. The flexibility of the SVV framework allows us to consider both rough volatility regimes, where the Hurst parameter satisfies $H \in (0, 1/2)$, and smoother regimes, depending on the choice of the Volterra kernel. This unified approach provides a comprehensive treatment of volatility modeling that encompasses many existing models as special cases, including the rough Bergomi model [5], [34], the rough SABR model [22] and the rough Stein-Stein model [29], explained in Section 6.

4.1 Model setting

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the canonical two-dimensional Wiener space, i.e.

$$\Omega := C_0([0,T]; \mathbb{R}^2),$$

and let $W = (W_1, W_2)$ be the corresponding two-dimensional Brownian motion defined by

$$W(t,\omega) = \omega(t), \quad t \in [0,T], \ \omega \in \Omega.$$

Denote by $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ the $\mathbb{P}$-augmented filtration generated by W.

4.1.1 Choice of Brownian representation

In the modelling framework above, different equivalent representations of a correlated two-dimensional Brownian motion can be employed depending on analytical convenience. Starting from a correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad \rho \in [-1, 1],$$

we construct a two-dimensional Brownian motion $B = (B_1, B_2)$ via a Cholesky decomposition $\Sigma = \mathcal{L}\mathcal{L}^\top$, i.e.

$$B(t) = \mathcal{L}W(t),$$

where W is a standard two-dimensional Brownian motion and

$$\mathcal{L} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}.$$

This yields the covariance structure $\mathbb{E}[B(t)B(s)^\top] = (t \wedge s)\Sigma$.

Alternatively, one may perform a further orthogonalisation by introducing a new Brownian motion $\widetilde{W} := \mathcal{U}^{-1}B(t)$, where $\mathcal{U}$ is the upper triangular matrix obtained from the Cholesky decomposition of Σ^{-1}:

$$\mathcal{U} = \begin{pmatrix} \frac{1}{\sqrt{1-\rho^2}} & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & 1 \end{pmatrix}.$$

In this formulation, the original Brownian components become

$$B_1(t) = \frac{1}{\sqrt{1-\rho^2}}\widetilde{W}_1(t) - \frac{\rho}{\sqrt{1-\rho^2}}\widetilde{W}_2(t), \qquad B_2(t) = \widetilde{W}_2(t).$$

Hence B_2 is driven solely by $\widetilde{W}_2$, while B_1 depends on both. Both constructions are equivalent in law, generate the same filtration, and the choice between them is a matter of convenience. The second representation is particularly useful for Malliavin calculus (see Section 5), as it isolates the noise driving the volatility component.

4.1.2 Sandwiched Volterra Volatility model

We now introduce the notation

$$B^S(t) := B_1(t), \qquad B^Y(t) := B_2(t).$$

Since the transformation from W to B is invertible and deterministic, the filtration generated by B coincides with the underlying Brownian filtration $\mathbb{F}$.

We now introduce the one-dimensional Sandwiched Volterra Volatility (SVV) model, following [15]. The price process S and the volatility process Y satisfy

$$S(t) = S(0) + \int_0^t \mu(s)S(s)ds + \int_0^t Y(s)S(s)dB^S(s),$$

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))ds + \int_0^t \mathcal{K}(t,s)dB^Y(s), \tag{4.1}$$

where:

- μ is a deterministic continuous function;
- B^S and B^Y are correlated Brownian motions;
- $\mathcal{K}$ is a square-integrable Volterra kernel such that the process

$$Z(t) := \int_0^t \mathcal{K}(t,s)\, dB^Y(s)$$

  is Hölder continuous of order $H \in (0,1)$;
- b is a drift function satisfying the conditions below.

Example 4.1 (Volterra kernels)**.**

- **Fractional kernel (rough volatility):**

$$\mathcal{K}(t,s) = C(t-s)^{H-\frac{1}{2}}, \qquad H \in (0,1),$$

  which yields the Riemann–Liouville fractional Brownian motion. The resulting process Z is Hölder continuous of order H.

- **Exponential kernel:**

$$\mathcal{K}(t,s) = e^{-\lambda(t-s)}, \qquad \lambda > 0.$$

  In this case Z satisfies the SDE $dZ(t) = -\lambda Z(t)dt + dB^Y(t)$, i.e. Z is an Ornstein–Uhlenbeck process.

- **Power-law kernel:**

$$\mathcal{K}(t,s) = (t-s)^{\beta} e^{-\lambda(t-s)}, \qquad \beta > -\frac{1}{2},\ \lambda \geq 0.$$

  This combines short-memory exponential decay with long-memory power-law behaviour near zero.

Assume that there exist deterministic continuous functions φ and ψ such that $0 < \varphi(t) < \psi(t)$ for all $t \in [0,T]$, and constants $c > 0$, $\gamma > \frac{1}{H} - 1$, and $y_* > 0$ such that

$$\begin{aligned} b(t,y) &\geq \frac{c}{(y-\varphi(t))^{\gamma}}, \qquad y \in (\varphi(t), \varphi(t) + y_*), \\ b(t,y) &\leq -\frac{c}{(\psi(t)-y)^{\gamma}}, \quad y \in (\psi(t) - y_*, \psi(t)). \end{aligned}$$

Under these assumptions, the process Y remains confined between φ and ψ, namely

$$\varphi(t) < Y(t) < \psi(t), \quad t \in [0,T].$$

For this reason, processes of the form (4.1) are referred to as *sandwiched processes*. This structure connects the model to uncertain volatility frameworks (see [2]).

4.1.3 Stochastic volatility process

Let $\mathcal{K} : [0,T]^2 \to \mathbb{R}$ be a measurable function satisfying the following conditions.

Assumption A (Kernel assumptions)**.**

(i) $\mathcal{K}$ is a square-integrable Volterra kernel, i.e. $\mathcal{K}(t,s) = 0$ for $t < s$, and

$$\int_0^T \int_0^T \mathcal{K}^2(t,s)\, ds\, dt < \infty,$$

together with

$$\int_0^t \mathcal{K}^2(t,s)\, ds < \infty, \quad \int_s^T \mathcal{K}^2(t,s)\, dt < \infty, \quad \forall\, s,t \in [0,T];$$

(ii) there exists $H \in (0,1)$ such that for any $\lambda \in (0,H)$,

$$\int_0^t \big(\mathcal{K}(t,u) - \mathcal{K}(s,u)\big)^2 du \le C_\lambda |t-s|^{2\lambda}, \quad 0 \le s \le t \le T, \tag{4.2}$$

where $C_\lambda > 0$ is a constant.

Under Assumption A(i), one can define the Gaussian Volterra process

$$Z(t) := \int_0^t \mathcal{K}(t,s)\, dB^Y(s), \quad t \in [0,T]. \tag{4.3}$$

Moreover, Assumption A(ii) ensures regularity of the trajectories, as stated below.

Proposition 4.2 ([3], Theorem 1 and Corollary 4)**.** *The Gaussian process $Z(t)$ admits a modification with a.s. Hölder continuous paths of any order $\lambda \in (0,H)$ if and only if (4.2) holds. Furthermore, for any $\lambda \in (0,H)$, the random variable*

$$\Lambda_\lambda := \sup_{0 \le s < t \le T} \frac{|Z(t) - Z(s)|}{|t-s|^\lambda} \tag{4.4}$$

has finite moments of all orders.

Proof. We prove the proposition in two parts.

Sufficiency ($\Leftarrow$). Assume that (4.2) holds for some $H \in (0,1)$ and all $\lambda \in (0,H)$. Fix $\lambda \in (0,H)$. For any $0 \le s < t \le T$, the increment $Z(t) - Z(s)$ is a centered Gaussian random variable. Its variance can be computed as follows:

$$\mathbb{E}\left[(Z(t) - Z(s))^2\right] = \int_0^T \big(\mathcal{K}(t,u)\mathbb{1}_{[0,t]}(u) - \mathcal{K}(s,u)\mathbb{1}_{[0,s]}(u)\big)^2 du.$$

Since $\mathcal{K}$ is a Volterra kernel ($\mathcal{K}(t,u)=0$ for $u>t$), we have for $u \le s$:

$$\mathcal{K}(t,u)\mathbb{1}_{[0,t]}(u) - \mathcal{K}(s,u)\mathbb{1}_{[0,s]}(u) = \mathcal{K}(t,u) - \mathcal{K}(s,u),$$

while for $s < u \le t$, we have $\mathcal{K}(t,u)\mathbb{1}_{[0,t]}(u) = \mathcal{K}(t,u)$ and $\mathcal{K}(s,u)\mathbb{1}_{[0,s]}(u) = 0$. Therefore,

$$\mathbb{E}\left[(Z(t)-Z(s))^2\right] = \int_0^s \left(\mathcal{K}(t,u)-\mathcal{K}(s,u)\right)^2 du + \int_s^t \mathcal{K}(t,u)^2 du.$$

Now, by (4.2), the first term is bounded by $C_\lambda |t-s|^{2\lambda}$. For the second term, note that by Assumption A(ii) with $s=0$, we have

$$\int_s^t \mathcal{K}(t,u)^2 du \le \int_0^t \mathcal{K}(t,u)^2 du < \infty,$$

but we need a bound of order $|t-s|^{2\lambda}$. Using the fact that $\int_0^t \mathcal{K}(t,u)^2 du$ is a non-decreasing function in t (since $\mathcal{K}$ is a Volterra kernel), we can write

$$\int_s^t \mathcal{K}(t,u)^2 du \le \int_0^t \mathcal{K}(t,u)^2 du - \int_0^s \mathcal{K}(s,u)^2 du + \int_0^s \left(\mathcal{K}(s,u)^2 - \mathcal{K}(t,u)^2\right) du.$$

However, a more standard approach (see [3, Theorem 1]) uses the fact that (4.2) implies the existence of a constant C such that

$$\mathbb{E}\left[(Z(t)-Z(s))^2\right] \le C|t-s|^{2\lambda}$$

for any $\lambda < H$. Indeed, both terms can be controlled using (4.2) and the square-integrability of $\mathcal{K}$.
Thus, for any $p \ge 1$, since $Z(t)-Z(s)$ is Gaussian, we have

$$\mathbb{E}\left[|Z(t)-Z(s)|^{2p}\right] \le C_p \left(\mathbb{E}\left[(Z(t)-Z(s))^2\right]\right)^p \le C_p' |t-s|^{2p\lambda}.$$

By the Kolmogorov–Čentsov continuity theorem (see [35], Chapter 2.2), there exists a modification of Z with Hölder continuous paths of any order $\lambda' < \lambda$. Since λ can be chosen arbitrarily close to H, we obtain Hölder continuity of any order $\alpha < H$. Moreover, the same theorem yields that Λ_λ defined in (4.4) has finite moments of all orders for any $\lambda < H$.

Necessity ($\Rightarrow$). Conversely, assume that Z has a Hölder continuous modification with exponent $\lambda \in (0, H)$, i.e., there exists a random variable G such that

$$|Z(t)-Z(s)| \le G|t-s|^\lambda \quad \text{a.s.}$$

for all $0 \le s < t \le T$. Since $Z(t)-Z(s)$ is Gaussian, its variance satisfies

$$\mathbb{E}\left[(Z(t)-Z(s))^2\right] \le \mathbb{E}\left[G^2\right]|t-s|^{2\lambda}.$$

Now observe that

$$\mathbb{E}\left[(Z(t)-Z(s))^2\right]=\int_0^s\big(\mathcal{K}(t,u)-\mathcal{K}(s,u)\big)^2du+\int_s^t\mathcal{K}(t,u)^2du.$$

Both terms on the right-hand side are non-negative. Therefore, for all $0\le s\le t\le T$,

$$\int_0^s\big(\mathcal{K}(t,u)-\mathcal{K}(s,u)\big)^2du\le\mathbb{E}\left[G^2\right]|t-s|^{2\lambda}.$$

This gives (4.2) for the given λ, with $C_\lambda=\mathbb{E}[G^2]$. Since the argument holds for any λ for which Z admits a Hölder continuous modification, and by assumption this holds for all $\lambda<H$, we have proved the necessary condition.

Finite moments. The statement that Λ_λ has finite moments of all orders follows from the Kolmogorov–Čentsov theorem and the Gaussianity of Z. Specifically, by the Kolmogorov–Čentsov theorem, the condition

$$\mathbb{E}\left[|Z(t)-Z(s)|^{2p}\right]\le C|t-s|^{2p\lambda+1}$$

for some $p>1/\lambda$ guarantees the existence of a modification with Hölder continuous paths and implies that $\mathbb{E}[|\Lambda_\lambda|^q]<\infty$ for appropriate q. Since Gaussian increments satisfy moment bounds for all p, we obtain finite moments of all orders. □

Remark 4.3. *In view of Proposition 4.2, we fix a modification of Z such that, for all $\omega\in\Omega$,*

$$|Z(t)-Z(s)|\le\Lambda_\lambda|t-s|^\lambda,\quad\lambda\in(0,H).$$

This Hölder continuous version will be used throughout.

Let $\varphi,\psi:[0,T]\to\mathbb{R}$ be functions such that:

1. φ and ψ are Hölder continuous of order H;
2. for all $t\in[0,T]$,
$$0<\varphi(t)<\psi(t).$$

For $a_1,a_2\in\left[0,\frac{1}{2}\|\psi-\varphi\|_\infty\right)$, define

$$\mathcal{D}_{a_1,a_2}:=\{(t,y)\in[0,T]\times\mathbb{R}_+:y\in(\varphi(t)+a_1,\psi(t)-a_2)\}.\tag{4.5}$$

Consider now a function $b:\mathcal{D}_{0,0}\to\mathbb{R}$ satisfying the following conditions.

Assumption B (Sandwiched drift)**.**

(i) $b\in C(\mathcal{D}_{0,0})$;

(ii) there exist $c > 0$, $p > 1$ such that for any $\varepsilon \in \left(0, \min\left\{1, \frac{1}{2}\|\psi - \varphi\|_\infty\right\}\right)$,

$$|b(t_1, y_1) - b(t_2, y_2)| \leq \frac{c}{\varepsilon^p}\left(|y_1 - y_2| + |t_1 - t_2|^\lambda\right),$$

for all $(t_i, y_i) \in \mathcal{D}_{\varepsilon,\varepsilon}$, for $i = 1, 2$;

(iii) there exist $\gamma > \frac{1}{H} - 1$ and $y_* \in \left(0, \frac{1}{2}\|\psi - \varphi\|_\infty\right)$ such that

$$b(t, y) \geq \frac{c}{(y - \varphi(t))^\gamma}, \quad (t, y) \in \mathcal{D}_{0,0} \setminus \mathcal{D}_{y_*,0},$$

$$b(t, y) \leq -\frac{c}{(\psi(t) - y)^\gamma}, \quad (t, y) \in \mathcal{D}_{0,0} \setminus \mathcal{D}_{0,y_*};$$

(iv) $\partial_y b$ exists, is continuous, and satisfies

$$\partial_y b(t, y) < c, \quad (t, y) \in \mathcal{D}_{0,0}.$$

We define the *sandwiched volatility process* Y as

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))\, ds + Z(t), \quad t \in [0, T], \tag{4.6}$$

where $Y(0) \in (\varphi(0), \psi(0))$ is deterministic, Z is given by (4.3) and b satisfies Assumption B.

Remark 4.4. *Equation* (4.6) *admits a unique pathwise solution for almost all* $\omega \in \Omega$ *provided that the corresponding* $Z(t, \omega)$ *of the noise is Hölder continuous up to the order* H *and this happens a.s. by Remark 4.3 (see [16]). Moreover, for any* $\lambda \in \left(\frac{1}{\gamma+1}, H\right)$, *where* γ *is from Assumption B (iii), there exist constants* $L_1, L_2, \alpha > 0$ *such that or all* $t \in [0, T]$:

$$\varphi(t) + \frac{L_1}{(L_2 + \Lambda_\lambda)^\alpha} \leq Y(t) \leq \psi(t) - \frac{L_1}{(L_2 + \Lambda_\lambda)^\alpha}.$$

In particular, Y *remains strictly between* φ *and* ψ *almost surely.*
Furthermore, for any $r > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} \frac{1}{Y^r(t)}\right] < \infty,$$

and

$$\mathbb{E}\left[\sup_{t\in[0,T]} \frac{1}{(Y(t) - \varphi(t))^r}\right] < \infty, \quad \mathbb{E}\left[\sup_{t\in[0,T]} \frac{1}{(\psi(t) - Y(t))^r}\right] < \infty.$$

Proof. The equation (4.6) is a Volterra equation of the form

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))\, ds + Z(t),$$

where $Z(t) = \int_0^t \mathcal{K}(t, s)\, dB^Y(s)$. By Proposition 4.2 and Remark 4.3, the path $Z(\cdot, \omega)$ is Hölder continuous of any order $\lambda < H$ for almost every $\omega \in \Omega$. For such ω, we treat the equation pathwise as a deterministic Volterra equation with Hölder continuous forcing. Under Assumption B, the drift b satisfies:

- local Lipschitz continuity in y on $\mathcal{D}_{\varepsilon,\varepsilon}$ (by condition (ii));
- explosion conditions (iii) at the boundaries φ and ψ.

A standard localization argument (see [16, Theorem 3.2]) guarantees the existence of a unique global solution $Y(\cdot,\omega)$ that remains in $(\varphi(t),\psi(t))$ for all $t\in[0,T]$. Hence, for almost every ω, a unique pathwise solution exists.
Now fix $\lambda \in \left(\frac{1}{\gamma+1}, H\right)$ and define the Hölder norm

$$\Lambda_\lambda := \sup_{0\le s<t\le T} \frac{|Z(t)-Z(s)|}{|t-s|^\lambda},$$

which is almost surely finite by Proposition 4.2 and has finite moments of all orders. Consider the reciprocal distance to the lower boundary:

$$R(t) := \frac{1}{Y(t)-\varphi(t)}.$$

Since $Y(0)>\varphi(0)$, $R(0)$ is finite. We aim to show that $R(t)$ remains bounded by a function of Λ_λ. Using the dynamics of Y and the deterministic nature of φ, we formally compute

$$dR(t) = -\frac{1}{(Y(t)-\varphi(t))^2}\big(b(t,Y(t))-\varphi'(t)\big)dt + \frac{1}{(Y(t)-\varphi(t))^3}d\langle Z\rangle(t) - \frac{1}{(Y(t)-\varphi(t))^2}dZ(t),$$

where $d\langle Z\rangle(t) = \mathcal{K}(t,t)^2dt$ (the quadratic variation of Z exists and is absolutely continuous under Assumption A; see e.g. [3]). Let $C_Z := \sup_{t\in[0,T]}\mathcal{K}(t,t)^2$ (which is finite by the square-integrability conditions).
By Assumption B(iii), for $Y(t)$ sufficiently close to $\varphi(t)$ we have $b(t,Y(t)) \ge c/(Y(t)-\varphi(t))^\gamma$. Hence,

$$-\frac{b(t,Y(t))-\varphi'(t)}{(Y(t)-\varphi(t))^2} \le -\frac{c}{(Y(t)-\varphi(t))^{\gamma+2}} + \frac{\|\varphi'\|_\infty}{(Y(t)-\varphi(t))^2}.$$

In terms of R, this becomes

$$-\frac{b-\varphi'}{(Y-\varphi)^2} \le -cR^{\gamma+2} + \|\varphi'\|_\infty R^2.$$

The quadratic variation term gives

$$\frac{d\langle Z\rangle(t)}{(Y-\varphi)^3} \le C_Z R^3 dt.$$

Therefore,

$$dR(t) \le \big(-cR(t)^{\gamma+2} + \|\varphi'\|_\infty R(t)^2 + C_Z R(t)^3\big)dt - R(t)^2 dZ(t).$$

Consider the deterministic initial value problem

$$\frac{dr}{dt} = -cr(t)^{\gamma+2} + \|\varphi'\|_\infty r(t)^2 + C_Z r(t)^3, \qquad r(0) = R(0).$$

Since $\gamma + 2 > 2$ (as $\gamma > 0$), the term $-cr^{\gamma+2}$ dominates for large r. The right-hand side is negative for sufficiently large r, so $r(t)$ remains bounded and does not explode in finite time. Moreover, the solution is unique and depends continuously on initial data. Using a stochastic comparison theorem (e.g., the fact that the drift of R is pointwise less than or equal to that of r, and the martingale term does not push R upward in a way that violates the comparison, see [16, Lemma 4.1]), we obtain $R(t) \leq r(t)$ almost surely for all $t \in [0, T]$. Consequently,

$$Y(t) - \varphi(t) = \frac{1}{R(t)} \geq \frac{1}{r(t)}.$$

The constants $\|\varphi'\|_\infty$ and C_Z depend only on φ and the kernel $\mathcal{K}$, not on Λ_λ. However, the martingale term $M(t) := \int_0^t R(s)^2 dZ(s)$ can be controlled using the Hölder continuity of Z. Specifically, for any t,

$$\left| \int_0^t R(s)^2 dZ(s) \right| \leq \Lambda_\lambda \int_0^t R(s)^2 \lambda s^{\lambda-1} ds,$$

which introduces Λ_λ multiplicatively. A refined analysis (see [16, Theorem 3.2]) shows that the comparison ODE must be replaced by a random differential inequality that incorporates Λ_λ. Solving this inequality yields the existence of constants $L_1, L_2 > 0$ and $\alpha > 0$, depending only on c, γ, λ, T and φ, such that

$$Y(t) - \varphi(t) \geq \frac{L_1}{(L_2 + \Lambda_\lambda)^\alpha}, \qquad \forall t \in [0, T].$$

The upper bound

$$\psi(t) - Y(t) \geq \frac{L_1}{(L_2 + \Lambda_\lambda)^\alpha}$$

follows by symmetry: apply the same argument to $\tilde{Y}(t) := \psi(t) - Y(t)$. The drift condition (iii) for the upper boundary gives $b(t, Y(t)) \leq -c/(\psi(t) - Y(t))^\gamma$, which translates into a similar inequality for $\tilde{Y}$. The same constants L_1, L_2, α can be chosen (possibly after taking the minimum of the two constants).
From the sandwiching bounds, we have

$$\frac{1}{Y(t) - \varphi(t)} \leq \frac{(L_2 + \Lambda_\lambda)^\alpha}{L_1}.$$

Since Λ_λ has finite moments of all orders by Proposition 4.2, for any $r > 0$,

$$\mathbb{E}\left[\sup_{t \in [0,T]} \frac{1}{(Y(t) - \varphi(t))^r} \right] \leq \mathbb{E}\left[\frac{(L_2 + \Lambda_\lambda)^{\alpha r}}{L_1^r} \right] < \infty.$$

The same holds for $1/(\psi(t) - Y(t))^r$.

For $1/Y(t)^r$, note that $Y(t) \geq \varphi(t) > 0$ (since φ is continuous and positive). Hence,

$$\frac{1}{Y(t)^r} \leq \frac{1}{\varphi(t)^r} \leq \frac{1}{\min_{t \in [0,T]} \varphi(t)^r} < \infty,$$

so the expectation is trivially finite. This completes the proof. $\square$

4.1.4 Price process

The price process S is defined as the solution to

$$S(t) = S(0) + \int_0^t \mu(s)S(s)\,ds + \int_0^t Y(s)S(s)\,dB^S(s), \tag{4.7}$$

where $S(0) > 0$ and $\mu : [0,T] \to \mathbb{R}$ is an H-Hölder continuous function.

Theorem 4.5. *Equation* (4.7) *admits a unique solution given by*

$$S(t) = S(0)\exp\left\{\int_0^t \left(\mu(s) - \frac{Y^2(s)}{2}\right) ds + \int_0^t Y(s)\,dB^S(s)\right\}. \tag{4.8}$$

Moreover, for any $r \in \mathbb{R}$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} S^r(t)\right] < \infty.$$

Proof. Uniqueness follows from standard results for SDEs with Lipschitz coefficients, since the sandwiched volatility process Y is bounded. Define the process

$$X(t) = \int_0^t \left(\mu(s) - \frac{Y(s)^2}{2}\right) ds + \int_0^t Y(s)\,dB^S(s).$$

Then the candidate solution can be written as $S(t) = f(X(t))$ with $f(x) = S(0)e^x$. Applying Itô's formula to $f(X(t))$ yields

$$dS(t) = f'(X(t))\,dX(t) + \frac{1}{2}f''(X(t))\,d\langle X\rangle(t).$$

Since $f'(x) = f''(x) = f(x) = S(0)e^x = S(t)$ when evaluated at $x = X(t)$, we have

$$dS(t) = S(t)\,dX(t) + \frac{1}{2}S(t)\,d\langle X\rangle(t).$$

The differential $dX(t)$ is given by

$$dX(t) = \left(\mu(t) - \frac{Y(t)^2}{2}\right) dt + Y(t)\,dB^S(t),$$

and the quadratic variation is

$$d\langle X\rangle(t) = Y(t)^2\,dt.$$

Substituting these into the expression for $dS(t)$ gives

$$dS(t) = S(t)\left[\left(\mu(t) - \frac{Y(t)^2}{2}\right) dt + Y(t)\,dB^S(t)\right] + \frac{1}{2}S(t)Y(t)^2\,dt.$$

Simplifying,

$$
\begin{aligned}
dS(t) &= S(t)\mu(t)\,dt - \frac{1}{2}S(t)Y(t)^2\,dt + \frac{1}{2}S(t)Y(t)^2\,dt + S(t)Y(t)\,dB^S(t),\\
dS(t) &= S(t)\mu(t)\,dt + S(t)Y(t)\,dB^S(t).
\end{aligned}
$$

This is precisely the SDE (4.7) in differential form. Moreover, at $t = 0$,

$$
S(0) = S(0)\exp\left\{\int_0^0 \left(\mu(s) - \frac{Y^2(s)}{2}\right) ds + \int_0^0 Y(s)\,dB^S(s)\right\} = S(0),
$$

so the initial condition is satisfied. Hence, (4.8) is the unique solution.

Next, using the explicit form of S and since $Y(t)$ is bounded above by $\psi(t)$, one can see that there exists a constant $C > 0$ such that

$$
\begin{aligned}
S^r(t) &= S^r(0)\exp\left\{r\int_0^t \mu(s)ds - \frac{r}{2}\int_0^t Y^2(s)ds + r\int_0^t Y(s)dB^S(s)\right\}\\
&\leq S^r(0)\exp\left\{|r|T\max_{s\in[0,T]}|\mu(s)| + \frac{|r|T}{2}\max_{s\in[0,T]}\psi^2(s)\right\}\exp\left\{r\int_0^t Y(s)dB^S(s)\right\}\\
&= S^r(0)\exp\left\{|r|T\max_{s\in[0,T]}|\mu(s)| + \frac{|r|T}{2}\max_{s\in[0,T]}\psi^2(s)\right\}\exp\left\{\frac{r^2}{2}\int_0^t Y^2(s)ds\right\}\times\\
&\qquad\times\exp\left\{\int_0^t rY(s)dB^S(s) - \frac{1}{2}\int_0^t (rY(s))^2 ds\right\}\\
&\leq c\exp\left\{\int_0^t rY(s)dB^S(s) - \frac{1}{2}\int_0^t (rY(s))^2 ds\right\}\\
&:= c\mathcal{E}_t\{rY\cdot B^S\},
\end{aligned}
\tag{4.9}
$$

where

$$
c := S^r(0)\exp\left\{|r|T\max_{s\in[0,T]}|\mu(s)| + \frac{|r|(|r|+1)T}{2}\max_{s\in[0,T]}\psi^2(s)\right\}
$$

and

$$
\mathcal{E}_t\{rY\cdot B^S\} := \exp\left\{\int_0^t rY(s)dB^S(s) - \frac{1}{2}\int_0^t (rY(s))^2 ds\right\}, \quad t\in[0,T].
$$

Note that the Novikov's criterion immediately yields that the process $\mathcal{E}_t\{rY\cdot B^S\}$, $t\in[0,T]$, is a uniformly integrable martingale such that

$$
\mathbb{E}\left[\mathcal{E}_t\{rY\cdot B^S\}\right] = 1, \quad t\in[0,T],
$$

and, moreover,

$$
\mathcal{E}_t\{rY\cdot B^S\} = 1 + \int_0^t \mathcal{E}_s\{rY\cdot B^S\}rY(s)dB^S(s), \quad t\in[0,T].
$$

By the Burkholder–Davis–Gundy inequality, there exists a constant $C_1 > 0$ such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]} \mathcal{E}_t\{rY \cdot B^S\}\right] &= 1 + \mathbb{E}\left[\sup_{t\in[0,T]} \int_0^t \mathcal{E}_s\{rY \cdot B^S\} rY(s) dB^S(s)\right] \\
&\le 1 + C_1 \mathbb{E}\left[\left(\int_0^T \mathcal{E}_s^2\{rY \cdot B^S\} Y^2(s) ds\right)^{\frac{1}{2}}\right] \\
&\le 1 + C_1 \sup_{s\in[0,T]} \psi(s) \left(\int_0^T \mathbb{E}\left[\mathcal{E}_s^2\{rY \cdot B^S\}\right] ds\right)^{\frac{1}{2}}.
\end{aligned}
$$

By Novikov's criterion, $\mathbb{E}\left[\mathcal{E}_s(2rY \cdot B^S)\right] = 1$, so

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{E}_s^2\{rY \cdot B^S\}\right] &= \mathbb{E}\left[\mathcal{E}_s\{2rY \cdot B^S\} \exp\left\{\int_0^s (rY(u))^2 du\right\}\right] \\
&\le \exp\left\{r^2 T \max_{s\in[0,T]} \psi^2(s)\right\} \mathbb{E}\left[\mathcal{E}_s\{2rY \cdot B^S\}\right] \\
&= \exp\left\{r^2 T \max_{s\in[0,T]} \psi^2(s)\right\} < \infty
\end{aligned}
$$

and hence, taking into account (4.9), we obtain that

$$
\mathbb{E}\left[\sup_{t\in[0,T]} S^r(t)\right] \le c\left(1 + C_1\sqrt{T} \sup_{s\in[0,T]} \psi(s) \exp\left\{\frac{r^2 T}{2} \max_{s\in[0,T]} \psi^2(s)\right\}\right),
$$

which yields the required result. □

4.1.5 Numéraire and discounted price

Let the discounted price process be

$$
\widetilde{S}(t) := e^{-rt} S(t)
$$

where $r \ge 0$ is the instantaneous interest rate. This satisfies

$$
\widetilde{S}(t) = S(0) + \int_0^t \widetilde{\mu}(s)\widetilde{S}(s)\, ds + \int_0^t Y(s)\widetilde{S}(s)\, dB^S(s),
$$

where $\widetilde{\mu} = \mu - r$.

4.1.6 Incompleteness and martingale measures

The model is driven by

$$
\begin{pmatrix} B^S(t) \\ B^Y(t) \end{pmatrix} = \mathcal{L} W(t),
$$

where $W = (W_1, W_2)$ is a two-dimensional Brownian motion and

$$\mathcal{L} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

is invertible and nondegenerate lower triangular matrix.

Lemma 4.6. *A positive local $\mathbb{P}$-martingale w.r.t. $\mathbb{F}$ M with $M(0) = 1$ a.s. satisfies that $M\widetilde{S}$ is a local martingale if and only if*

$$M(t) = \mathcal{E}_t \left\{ -\int_0^{\cdot} \frac{\widetilde{\mu}(s)}{Y(s)} dW_1(s) + \int_0^{\cdot} \xi_2(s) dW_2(s) \right\}, \tag{4.10}$$

where ξ_2 is a predictable process such that $\int_0^{\cdot} \xi_2(s) dW_2(s)$ is well-defined local martingale.

Proof. By the virtue of the local martingale representation theorem (see [35, Section 3.2.D]), M is a local $\mathbb{P}$-martingale w.r.t. $\mathbb{F}$ if and only if there exist predictable processes $\zeta_j = \{\zeta_j(t)\}_{t\in[0,T]}$, $j = 1, 2$, such that

$$\mathbb{P}\left(\int_0^T \zeta_j^2(s) ds < \infty \right) = 1, \quad j = 1, 2,$$

and

$$M(t) = 1 + \sum_{j=1}^{2} \int_0^t \zeta_j(s) dW_j(s). \tag{4.11}$$

By Itô's formula,

$$\begin{aligned} d(M(t)\widetilde{S}(t)) &= (\tilde{\mu}(t)M(t) + Y(t)\zeta_1(t))\, \widetilde{S}(t)dt \\ &\quad + (M(t)Y(t) + \zeta_1(t))\, \widetilde{S}(t)dW_1(t) + \zeta_2(t)\widetilde{S}(t)dW_1(t), \end{aligned}$$

and thus $M\widetilde{S}$ is a local martingale if and only if

$$\int_0^t (\tilde{\mu}(s)M(s) + Y(s)\zeta_1(s))\, \widetilde{S}(s)ds = 0, \quad t \in [0, T].$$

This, in turn, implies that

$$\zeta_1(t) = -M(t)\frac{\tilde{\mu}(t)}{Y(t)},$$

almost everywhere w.r.t. $dt \otimes d\mathbb{P}$. Since $\mathcal{L}$ is lower-triangular and each $Y(t)$ is positive a.s., we have that

$$\mathcal{L}_{11}\, \zeta_1(t) = -M(t)\frac{\tilde{\mu}(t)}{Y(t)}$$

almost everywhere w.r.t. $dt \otimes d\mathbb{P}$ and hence

$$\zeta_1(t) = \mathcal{L}_{11}^{-1} \left(-M(t)\frac{\tilde{\mu}(t)}{Y(t)} \right) \tag{4.12}$$

almost everywhere w.r.t. $dt \otimes d\mathbb{P}$. Using notation $\mathcal{L}_{11}^{-1} = 1$ and substituting (4.12) to (4.11), we see that the process M should satisfy the SDE of the form

$$dM(t) = -M(t)\frac{\widetilde{\mu}(t)}{Y(t)}dW_1(t) + \zeta_2(t)dW_2(t).$$

This SDE has a unique solution of the form

$$\begin{aligned} M(t) &= \mathcal{E}_t\left\{-\int_0^{\cdot} \frac{\widetilde{\mu}(s)}{Y(s)}dW_1(s)\right\} \times \\ &\quad \times \left(1 + \int_0^t \mathcal{E}_u^{-1}\left\{-\int_0^{\cdot} \frac{\widetilde{\mu}(s)}{Y(s)}dW_1(s)\right\}\zeta_2(u)dW_2(u)\right) \\ &=: \mathcal{E}_t\left\{-\int_0^{\cdot} \frac{\widetilde{\mu}(s)}{Y(s)}dW_1(s)\right\}\left(1 + \int_0^t \widetilde{\zeta}_2(u)dW_2(u)\right). \end{aligned} \tag{4.13}$$

Furthermore, since the process M is positive, the continuous process $\widetilde{M}$ defined by

$$\widetilde{M}(t) := 1 + \int_0^t \widetilde{\zeta}_2(s)dW_2(s)$$

must also be positive, thus it can be represented as

$$\widetilde{M}(t) = \mathcal{E}_t\left(\int_0^{\cdot} \xi_2(s)dW_2(s)\right), \tag{4.14}$$

with ξ_2 being a predictable process such that the corresponding stochastic integrals are well-defined. It remains to notice that $\widetilde{\zeta}_2$ can be arbitrary predictable stochastic processes such that the corresponding stochastic integrals are well-defined, thus ξ_2 can also be chosen arbitrarily. Taking into account (4.13) and (4.14), we get the representation (4.10). □

Such densities define candidate martingale measures. When M is uniformly integrable, $\mathbb{Q}$ defined by $d\mathbb{Q} = M(T)d\mathbb{P}$ is a probability measure.

Theorem 4.7. *The market is arbitrage-free and incomplete.*

Proof. Since the process Y is bounded away from zero by the corresponding $\min_{t\in[0,T]} \varphi(t) > 0$, thus, by Novikov's condition,

$$M(t) = \mathcal{E}_t\left\{-\int_0^{\cdot} \frac{\widetilde{\mu}(s)}{Y(s)}dW_1(s)\right\}$$

is a uniformly integrable martingale (this case corresponds to $\xi_2 \equiv 0$ in (4.10)). Therefore the associated $\mathbb{Q}$ is a probability measure, i.e. the market is arbitrage free.
Furthermore, again by Novikov's condition, any ξ_2 such that

$$\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_0^T \xi_2^2(t)dt\right\}\right] < \infty,$$

which means M is a uniformly integrable martingale, thus the market is incomplete. □

Equivalently, using the representation $B = \mathcal{L}W$, the model can be written in terms of the standard Brownian motion $W = (W_1, W_2)$ as

$$\begin{aligned} S(t) &= S(0) + \int_0^t \mu(s)S(s)\,ds + \int_0^t Y(s)S(s)\,dW_1(s), \\ Y(t) &= Y(0) + \int_0^t b(s, Y(s))\,ds + \int_0^t \mathcal{K}(t,s)\big(\rho dW_1(s) + \sqrt{1-\rho^2}dW_2(s)\big), \end{aligned} \tag{4.15}$$

where $\mathcal{K}$ satisfies Assumption A, b satisfies Assumption B, and μ is an H-Hölder continuous function.

Finally, applying the transformation introduced in Subsection 4.1.1, the system (4.15) can be rewritten in terms of a new Brownian motion $\widetilde{W} = (\widetilde{W}_1, \widetilde{W}_2)$ as

$$\begin{aligned} S(t) &= S(0) + \int_0^t \mu(s)S(s)\,ds + \int_0^t Y(s)S(s)\left(\frac{1}{\sqrt{1-\rho^2}}d\widetilde{W}_1(s) - \frac{\rho}{\sqrt{1-\rho^2}}d\widetilde{W}_2(s)\right), \\ Y(t) &= Y(0) + \int_0^t b(s, Y(s))\,ds + \int_0^t \mathcal{K}(t,s)\,d\widetilde{W}_2(s), \end{aligned}$$

where the transformation is given by

$$\begin{pmatrix} \frac{1}{\sqrt{1-\rho^2}} & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \widetilde{W}_1 \\ \widetilde{W}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}.$$

4.1.7 Implied volatility surface generated by the SVV model

A widely accepted criterion for assessing the adequacy of a financial model is the shape of the implied volatility surface it generates as we discussed in Section 1. In this subsection, we analyse qualitative properties of the implied volatility surfaces produced by the SVV model. We assume that the short-term interest rate r is constant and consider the SVV model given by

$$\begin{aligned} S(t) &= S(0) + r\int_0^t S(s)\,ds + \int_0^t Y(s)S(s)\left(\sqrt{1-\rho^2}\,dB_1(s) + \rho\,dB_2(s)\right), \\ Y(t) &= Y(0) + \int_0^t b(s, Y(s))\,ds + \int_0^t \mathcal{K}(t,s)\,dB_2(s), \end{aligned} \tag{4.16}$$

where B_1 and B_2 are independent Brownian motions, $\rho \in (-1, 1)$, the kernel $\mathcal{K}$ satisfies Assumption A, and the drift b satisfies Assumption B, ensuring that

$$0 < \min_{t\in[0,T]} \varphi(t) < \varphi(t) < Y(t) < \psi(t) < \max_{t\in[0,T]} \psi(t) < \infty. \tag{4.17}$$

Remark 4.8. *In the particular case of the SVV model* (4.16)*, the discounted price process* $\widetilde{S} = \{e^{-rt}S(t),\ t \in [0,T]\}$ *satisfies the stochastic differential equation*

$$\widetilde{S}(t) = S(0) + \int_0^t Y(s)\widetilde{S}(s)\left(\sqrt{1-\rho^2}\, dB_1(s) + \rho\, dB_2(s)\right),$$

and is therefore a martingale. Consequently, under model (4.16)*, the arbitrage-free price* Π *of a European call option with maturity* T *and strike* K *is given by*

$$\Pi_{SVV}(T,K) = e^{-rT}\mathbb{E}\left[(S(T)-K)_+\right].$$

We also recall the Black–Scholes price of a European call option with maturity T, strike K, and volatility $\sigma > 0$:

$$\Pi_{BS}(T,K;\sigma) := S(0)\Phi\left(\frac{\log\frac{e^{rT}S(0)}{K} + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right) - Ke^{-rT}\Phi\left(\frac{\log\frac{e^{rT}S(0)}{K} - \frac{\sigma^2}{2}T}{\sigma\sqrt{T}}\right),$$

where Φ denotes the cumulative distribution function of the standard normal distribution. We define the implied volatility surface $(T,K) \mapsto \widehat{\sigma}(T,K)$ associated with the SVV model by

$$\Pi_{BS}(T,K;\widehat{\sigma}(T,K)) = \Pi_{SVV}(T,K).$$

Remark 4.9. *The implied volatility* $\widehat{\sigma}(T,K)$ *is well-defined and unique for all* $T,K > 0$*, since the mapping* $\sigma \mapsto \Pi_{BS}(T,K;\sigma)$ *is strictly increasing. Moreover,*

$$\begin{aligned}
\lim_{\sigma\to 0+}\Pi_{BS}(T,K;\sigma) &= (S(0) - e^{-rT}K)_+ \\
&= (\mathbb{E}[e^{-rT}S(T)] - e^{-rT}K)_+ \\
&\le \mathbb{E}\left[(e^{-rT}S(T) - e^{-rT}K)_+\right] = \Pi_{SVV}(T,K),
\end{aligned}$$

which guarantees existence.

We first establish that the implied volatility surface generated by the SVV model is bounded.

Theorem 4.10. *For all* $T,K > 0$,

$$\min_{t\in[0,T]}\varphi(t) \le \widehat{\sigma}(T,K) \le \max_{t\in[0,T]}\psi(t).$$

Proof. Fix $T,K > 0$ and denote

$$\sigma_- := \min_{t\in[0,T]}\varphi(t), \qquad \sigma_+ := \max_{t\in[0,T]}\psi(t).$$

By assumption (4.17), the volatility process satisfies $\sigma_- \le Y(t) \le \sigma_+$ for all $t \in [0,T]$ almost surely.

It follows from the comparison result of Avellaneda, Levy, and Parás [2] that the corresponding option prices satisfy

$$\Pi_{BS}(T,K;\sigma_-) \le \Pi_{SVV}(T,K) \le \Pi_{BS}(T,K;\sigma_+).$$

On the other hand, for fixed $T, K > 0$, the mapping $\sigma \mapsto \Pi_{BS}(T,K;\sigma)$ is continuous and strictly increasing on $(0,\infty)$. By definition of the implied volatility $\widehat{\sigma}(T,K)$,

$$\Pi_{BS}(T,K;\widehat{\sigma}(T,K)) = \Pi_{SVV}(T,K).$$

Therefore, the above inequalities imply

$$\Pi_{BS}(T,K;\sigma_-) \le \Pi_{BS}(T,K;\widehat{\sigma}(T,K)) \le \Pi_{BS}(T,K;\sigma_+).$$

Using the strict monotonicity of $\Pi_{BS}(T,K;\cdot)$, we conclude that

$$\sigma_- \le \widehat{\sigma}(T,K) \le \sigma_+,$$

which completes the proof. □

Remark 4.11. *Theorem 4.10 suggests that the upper bound function ψ should be chosen such that $\max_{t\in[0,T]}\psi(t)$ exceeds the maximal implied volatility observed in the market. Under typical market conditions, taking $\max_{t\in[0,T]}\psi(t) = 1$ is sufficient.*

Define $\kappa := \log \frac{K}{e^{rT}S(0)}$ and introduce the log-moneyness parametrization

$$\widehat{\sigma}_{\text{log-m}}(T,\kappa) := \widehat{\sigma}\big(T, S(0)e^{\kappa+rT}\big).$$

Empirical evidence (see [23]) shows that the at-the-money slope of the implied volatility surface steepens as $T \to 0$, following the heuristic relation

$$\left|\frac{\widehat{\sigma}_{\text{emp}}(T,\kappa) - \widehat{\sigma}_{\text{emp}}(T,\kappa')}{\kappa-\kappa'}\right| \propto T^{-\frac{1}{2}+H}, \quad \kappa,\kappa' \approx 0, \quad H \in \left(0,\frac{1}{2}\right),$$

where $\widehat{\sigma}_{\text{emp}}$ is the empirical volatility from the market.

To reproduce this behaviour, one seeks models for which

$$\left|\frac{\partial\widehat{\sigma}_{\text{log-m}}}{\partial\kappa}(T,\kappa)\right|_{\kappa=0} = O(T^{-\frac{1}{2}+H}), \quad T \to 0. \tag{4.18}$$

The SVV model (4.16) satisfies (4.18) under suitable assumptions on the kernel $\mathcal{K}$.

Theorem 4.12. *Let $H \in \left(\frac{1}{6},\frac{1}{2}\right)$ and $\rho < 0$ in (4.16). Take*

$$b(t,y) = \frac{\theta_1(t)}{(y-\varphi(t))^{\gamma_1}} - \frac{\theta_2(t)}{(\psi(t)-y)^{\gamma_2}} + a(t,y),$$

where

- $\gamma_1 > \frac{1}{H} - 1$, $\gamma_2 > \frac{1}{H} - 1$;
- *the functions* θ_1, θ_2: $[0,T] \to \mathbb{R}$ *are strictly positive and continuous;*
- *the function* a: $[0,T] \times \mathbb{R} \to \mathbb{R}$ *is locally Lipschitz in* y *uniformly in* t, *i.e. for any* $N > 0$ *there exists a constant* $C_N > 0$ *that does not depend on* t *such that*

$$|a(t,y_2) - a(t,y_1)| \le C_N |y_2 - y_1|, \quad t \in [0,T], \quad y_1, y_2 \in [-N, N];$$

- a: $[0,T] \times \mathbb{R} \to \mathbb{R}$ *is two times differentiable w.r.t. the spatial variable* y *with* a, a'_y, a''_{yy} *all being continuous on* $[0,T] \times \mathbb{R}$

and assume that the kernel $\mathcal{K}$ *is such that*

- *for any* $0 \le s < t \le T$, *there is a constant* $C > 0$ *such that*

$$|\mathcal{K}(t,s)| \le C|t-s|^{-\frac{1}{2}+H},$$

- *there exists a constant* $K_Y > 0$ *such that*

$$\frac{1}{\tau^{\frac{3}{2}+H}} \int_0^\tau \int_s^\tau \mathcal{K}(t,s) dt ds \to K_Y, \quad \tau \to 0+.$$

Then the power law (4.18) *holds for the SVV model* (4.16).

Example 4.13. Following equation (4.16), we give an example of SVV model below, letting $r = 0$ and $\rho = -0.5$:

$$S(t) = S(0) + \int_0^t Y(s)S(s) \left(\sqrt{0.75}\, dB_1(s) - 0.5\, dB_2(s) \right)$$
$$Y(t) = Y(0) + \int_0^t \left(\frac{0.005}{(Y(s) - 0.05)^3} - \frac{0.005}{(1 - Y(s))^3} + 0.05\,(0.5 - Y(s)) \right) ds + \int_0^t \mathcal{K}(t,s)\, dB_2(s),$$

where $\mathcal{K}$ can be chosen as in Example 4.1.

Proof. The proof can be found in [17], Theorems 4 and 5. □

For a more detailed numerical study, we refer to [15], Section 2.3.

5 Malliavin Calculus

5.1 Preliminaries on Malliavin calculus

In this section, we briefly recall the main elements of Malliavin calculus that will be used throughout this work. For a comprehensive treatment, we refer to standard references such as [43].

Gaussian Framework. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space supporting a standard Brownian motion $B = \{B(t),\, t \in [0,T]\}$. We denote by $\{\mathcal{F}_t\}_{t\in[0,T]}$ the natural filtration of B, augmented to satisfy the usual conditions (see Appendix A).
Let $\mathcal{H} := L^2([0,T])$ be the Hilbert space, equipped with the inner product

$$\langle f, g\rangle_{\mathcal{H}} := \int_0^T f(t)g(t)\,dt.$$

Smooth Random Variables. We define the class of smooth random variables $\mathcal{S}$ as

$$\mathcal{S} := \left\{F = f\big(B(h_1), \ldots, B(h_n)\big) \,:\, n \geq 1,\ h_i \in \mathcal{H},\ f \in C_b^\infty(\mathbb{R}^n)\right\},$$

where

$$B(h) := \int_0^T h(t)\,dB(t).$$

Malliavin Derivative. For $F \in \mathcal{S}$ of the form $F = f(B(h_1), \ldots, B(h_n))$, the Malliavin derivative DF is the $\mathcal{H}$-valued random variable defined by

$$D_t F = \sum_{i=1}^n \partial_i f(B(h_1), \ldots, B(h_n))\, h_i(t), \quad t \in [0,T].$$

The operator D is closable from $L^2(\Omega)$ into $L^2(\Omega; \mathcal{H})$, and we denote by $\mathbb{D}^{1,2}$ the closure of $\mathcal{S}$ under the norm

$$\|F\|_{1,2}^2 := \mathbb{E}[|F|^2] + \mathbb{E}[\|DF\|_{\mathcal{H}}^2].$$

Higher-order Derivatives. The derivative operator can be iterated. For $k \geq 1$, we define $\mathbb{D}^{k,2}$ as the set of random variables with Malliavin derivatives up to order k in L^2, endowed with the norm

$$\|F\|_{k,2}^2 := \mathbb{E}[|F|^2] + \sum_{j=1}^k \mathbb{E}\big[\|D^j F\|_{\mathcal{H}^{\otimes j}}^2\big].$$

Divergence Operator. The divergence operator δ is defined as the adjoint of the Malliavin derivative. A process $u \in L^2(\Omega; \mathcal{H})$ belongs to the domain of δ if there exists a random variable $\delta(u) \in L^2(\Omega)$ such that

$$\mathbb{E}[F\,\delta(u)] = \mathbb{E}\big[\langle DF, u\rangle_{\mathcal{H}}\big], \quad \text{for all } F \in \mathcal{S}.$$

In this case, $\delta(u)$ is called the Skorokhod integral of u. If u is adapted, then $\delta(u)$ coincides with the Itô integral.

Ornstein–Uhlenbeck Operator. The Ornstein–Uhlenbeck operator L is defined by

$$L := -\delta D,$$

with domain $\mathbb{D}^{2,2}$. This operator plays a central role in the analysis of Gaussian functionals.

Clark–Ocone Formula. If $F \in \mathbb{D}^{1,2}$, then

$$F = \mathbb{E}[F] + \int_0^T \mathbb{E}[D_t F \mid \mathcal{F}_t]\, dB(t),$$

which provides a stochastic representation of F.

Non-degeneracy and Densities. A key application of Malliavin calculus is the study of the existence and smoothness of densities. By the Bouleau–Hirsch criterion, if

$$F \in \mathbb{D}^{1,2} \quad \text{and} \quad \|DF\|_{\mathcal{H}} > 0 \quad \text{a.s.},$$

then the law of F is absolutely continuous with respect to the Lebesgue measure. Moreover, if

$$F \in \mathbb{D}^{\infty} \quad \text{and} \quad \|DF\|_{\mathcal{H}}^{-1} \in L^p(\Omega) \quad \text{for all } p \geq 1,$$

then F admits a smooth density.

5.2 Malliavin differentiability of volatility and price w.r.t. the initial Brownian motion W

This subsection is devoted to the Malliavin differentiability of the volatility and price processes within the SVV model. The strategy is as follows. First, we employ the classical characterisation of the Malliavin derivative via stochastic Gâteaux differentiability to establish differentiability of the sandwiched volatility processes Y. Then, we derive a suitable chain rule and apply it to the representation (4.8) in order to prove differentiability of the price processes S.

5.2.1 Malliavin differentiability of volatility.

To introduce Malliavin differentiability, we first fix the underlying isonormal Gaussian process. A natural choice is the initial Brownian motion W. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is the canonical 2-dimensional Wiener space, and $W = (W_1, W_2)$ is the corresponding Brownian motion. The volatility processes $Y = \{Y(t),\, t \in [0,T]\}$ defined as

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))\, ds + Z(t), \quad t \in [0,T],$$

where $Z(t) = \int_0^t \mathcal{K}(t,s)\, dB^Y(s)$ and

$$B^Y := \sum_{j=1}^{2} \ell_{2,j} W_j = \rho W_1 + \sqrt{1-\rho^2} W_2.$$

Let $\mathcal{H}$ denote the associated Cameron–Martin space, consisting of all functions $F = (F_1, F_2) \in C_0([0,T]; \mathbb{R}^2)$ such that

$$F(t) = \left(\int_0^t f_1(s)\, ds, \int_0^t f_2(s)\, ds \right), \quad f \in L^2([0,T]; \mathbb{R}^2).$$

We write $F = \int_0^\cdot f(s)\, ds$. Denote by $\mathbb{D}^{1,2}$ the space of Malliavin differentiable random variables.

We will rely on the following classical result.

Theorem 5.1 ([50], Theorem 3.1). *A random variable $\eta \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ belongs to $\mathbb{D}^{1,2}$ if and only if:*

(1) η is ray absolutely continuous*, i.e., for any $F \in \mathcal{H}$, the mapping $\varepsilon \mapsto \eta(\omega + \varepsilon F)$ is absolutely continuous;*

(2) η is stochastically Gâteaux differentiable*, i.e., there exists*

$$D\eta \in L^2(\Omega; L^2([0,T]; \mathbb{R}^2))$$

such that for all $F = \int_0^\cdot f(s)\, ds \in \mathcal{H}$,

$$\frac{\eta(\omega + \varepsilon F) - \eta(\omega)}{\varepsilon} \xrightarrow{\mathbb{P}} \langle D\eta, f \rangle_{L^2([0,T];\mathbb{R}^2)}.$$

as $\varepsilon \to 0$. Therefore, the random vector $D\eta$ from (2) is the Malliavin derivative of η.

To apply Theorem 5.1, we interpret the shifted process $Y(\omega + \varepsilon F, t)$ as the solution $Y^{\varepsilon,F}(t)$ of

$$Y^{\varepsilon,F}(\omega, t) = Y(0) + \int_0^t b(s, Y^{\varepsilon,F}(\omega, s))\, ds + Z^{\varepsilon,F}(\omega, t),$$

where $Z^{\varepsilon,F}(\omega,t) := Z(\omega+\varepsilon F,t)$.
The key observation is that the shifted noise admits the decomposition

$$Z^{\varepsilon,F}(\omega,t) = Z(\omega,t) \ +\varepsilon \sum_{j=1}^{2} \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s)\, ds,$$

which separates the random and deterministic contributions with the following proposition.

Proposition 5.2. *Let $\varepsilon \geq 0$ and $F = \left(\int_0^t f_1(s)ds, \int_0^t f_2(s)ds\right) \in \mathcal{H}$. Then there exists $\Omega_{\varepsilon,F}$, $\mathbb{P}(\Omega_{\varepsilon,F}) = 1$, such that for any $\omega \in \Omega_{\varepsilon,F}$*

$$Z^{\varepsilon,F}(\omega,t) = Z(\omega,t) + \varepsilon \sum_{j=1}^{2} \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds, \quad t \in [0,T].$$

Proof. By definition, for any $t \in [0,T]$

$$\begin{aligned} Z(\omega,t) &= L^2\text{-}\lim_{n\to\infty} \sum_{l=0}^{k_n-1} K^{n,l}(t) \left(B^Y(\omega,t_{n,l+1}) - B^Y(\omega,t_{n,l})\right) \\ &= L^2\text{-}\lim_{n\to\infty} \sum_{j=1}^{2} \ell_{2,j} \sum_{l=0}^{k_n-1} K^{n,l}(t) \left(W_j(\omega,t_{n,l+1}) - W_j(\omega,t_{n,l})\right) \\ &=: L^2\text{-}\lim_{n\to\infty} Z^n(\omega,t), \end{aligned}$$

where $\pi^n = \{0 = t_{n,0} < t_{n,1} < ... < t_{n,k_n} = t\}$, $n \geq 1$, is a system of partitions such that $\sup_{l=0,\dots,k_n-1}(t_{n,l+1} - t_{n,l}) \to 0$ as $n \to \infty$ and $\mathcal{K}^n(t,s) := \sum_{l=0}^{k_n-1} K^{n,l}(t) \mathbb{1}_{[t_{n,l},t_{n,l+1})}(s)$ is such that $\int_0^t \left(\mathcal{K}(t,s) - \mathcal{K}^n(t,s)\right)^2 ds \to 0$, $n \to \infty$.

By the Riesz theorem, there exist $\Omega'_t \subset \Omega$, $\mathbb{P}(\Omega'_t) = 1$, and a subsequence of partitions that will again be denoted by $\{\pi^n,\ n \geq 1\}$ such that for all $\omega \in \Omega'_t$

$$Z(\omega,t) = \lim_{n\to\infty} Z^n(\omega,t). \tag{5.1}$$

Moreover, note that, by the definition of the probability space and Brownian motion W, for any $\omega = (\omega_1,\omega_2) \in \Omega = C_0([0,T];\mathbb{R}^2)$

$$\begin{aligned} Z^n(\omega,t) &= \sum_{j=1}^{2} \ell_{2,j} \sum_{l=0}^{k_n-1} K^{n,l}(t) \left(W_j(\omega,t_{n,l+1}) - W_j(\omega,t_{n,l})\right) \\ &= \sum_{j=1}^{2} \ell_{2,j} \sum_{l=0}^{k_n-1} K^{n,l}(t) \left(\omega_j(t_{n,l+1}) - \omega_j(t_{n,l})\right), \end{aligned}$$

whence, for all $\omega \in \Omega$,

$$Z^n(\omega + \varepsilon F, t) = Z^n(\omega) + \varepsilon \sum_{j=1}^{2} \ell_{2,j} \sum_{l=0}^{k_n - 1} K^{n,l}(t) \left(F_j(t_{n,l+1}) - F_j(t_{n,l})\right)$$

and thus, for all $\omega \in \Omega_t'$,

$$\begin{aligned} \lim_{n\to\infty} Z^n(\omega + \varepsilon F, t) &= \lim_{n\to\infty} Z^n(\omega) + \lim_{n\to\infty} \varepsilon \sum_{j=1}^{2} \ell_{2,j} \sum_{l=0}^{k_n - 1} K^{n,l}(t) \left(F_j(t_{n,l+1}) - F_j(t_{n,l})\right) \\ &= Z(\omega, t) + \varepsilon \sum_{j=1}^{2} \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds. \end{aligned} \tag{5.2}$$

Now, denote $\Omega_{\varepsilon,F,t} := \{\omega \in \Omega_t' : \ \omega + \varepsilon F \in \Omega_t'\}$. Then $\mathbb{P}(\Omega_{\varepsilon,F,t}) = 1$ since, by the Cameron-Martin theorem,

$$\mathbb{P}(\Omega_{\varepsilon,F,t}) = \int_\Omega \mathbb{1}_{\Omega_t'}(\omega + \varepsilon F) d\mathbb{P}(\omega) = \mathbb{E}\left[\mathbb{1}_{\Omega_t'} \exp\left\{\varepsilon \mathcal{I}_f - \frac{\varepsilon^2}{2} \|\mathcal{I}_f\|^2_{L^2(\Omega)}\right\}\right] = 1,$$

where $\mathcal{I}_f := \sum_{j=1}^2 \int_0^T f_j(s) dW_j(s)$, and therefore, taking into account (5.1) and (5.2), for any $\omega \in \Omega_{\varepsilon,F,t}$

$$Z(\omega + \varepsilon F, t) = Z(\omega, t) + \varepsilon \sum_{j=1}^{2} \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds. \tag{5.3}$$

Finally, it remains to notice that, for the fixed $\varepsilon \geq 0$ and $F \in \mathcal{H}$, both left- and right-hand sides of (5.3) are continuous w.r.t. t, whence we can put

$$\Omega_{\varepsilon,F} := \bigcap_{t \in \mathbb{Q} \cap [0,T]} \Omega_{\varepsilon,F,t}.$$

□

We now state the main result of this subsection.

Theorem 5.3. *Under Assumptions A and B, for all $t \in [0,T]$, one has $Y(t) \in \mathbb{D}^{1,2}$ and*

$$\begin{aligned} DY(t) &= (D^1 Y(t), D^2 Y(t)), \\ D_u^j Y(t) &= \ell_{2,j} \left(\mathcal{K}(t,u) + \int_u^t \mathcal{K}(s,u) \frac{\partial b}{\partial y}(s, Y(s)) e^{\int_s^t \frac{\partial b}{\partial y}(v, Y(v))\, dv} ds \right) \mathbb{1}_{[0,t]}(u). \end{aligned} \tag{5.4}$$

Proof. We will split the proof into three steps. First, we verify that $DY(t)$ given by (5.4) indeed satisfies

$$\sum_{j=1}^{2} \mathbb{E} \int_0^T \left(D_u^j Y_i(t)\right)^2 du < \infty$$

and then check the conditions of Theorem 5.1.

By Remark 4.4, for all $s \in [0,T]$: $(s, Y(s)) \in \mathcal{D}_{\frac{1}{\xi(0)}, \frac{1}{\xi(0)}}$, where

$$\xi(0) := \frac{(L_2 + \Lambda_\lambda)^\alpha}{L_1}$$

with Λ_λ being defined by (4.4) and L_1, L_2, α being some deterministic positive constants that depend only on $\lambda \in \left(\frac{1}{\gamma+1}, H\right)$ and the shape of the drift b. Whence, due to Assumption B(ii),

$$\left|\frac{\partial b}{\partial y}(s, Y(s))\right| \leq c\xi^p(0), \quad s \in [0,T].$$

Moreover, by Assumption B(iv),

$$e^{\int_s^t \frac{\partial b_i}{\partial y}(v, Y_i(v))dv} \leq e^{cT}.$$

Therefore

$$\begin{aligned} \mathbb{E}\int_0^T \left(D_u^j Y(t)\right)^2 du \leq 2\ell_{2,j}^2 \int_0^t \mathcal{K}^2(t,u)du \\ + 2Tc^2 e^{2cT} \ell_{2,j}^2 \mathbb{E}\left[\xi^{2p}(0)\right] \int_0^t \int_u^t \mathcal{K}^2(s,u)dsdu < \infty, \end{aligned}$$

because $\int_0^t \mathcal{K}^2(t,s)ds < \infty$ by Assumption A(i) and the random variable $\xi(0)$ has moments of all orders.

Let $F \in \mathcal{H}$ be fixed. Then, as explained above, the random process $\{Y^{\varepsilon,F}(t),\ \varepsilon \geq 0\}$ defined for every $\varepsilon \geq 0$ by the equation

$$Y^{\varepsilon,F}(t) = Y_0 + \int_0^t b(s, Y^{\varepsilon,F}(s))ds + Z(t) + \varepsilon \sum_{j=1}^2 \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds \tag{5.5}$$

is a version of the process $\{Y(\omega + \varepsilon F, t),\ \varepsilon \geq 0\}$. Let us prove that this version is Lipschitz w.r.t. ε and hence absolutely continuous. First, fix $\lambda \in \left(\frac{1}{\gamma+1}, H\right)$, where H is from Assumption A(ii) and γ is from Assumption B(iii), and take an arbitrary $\varepsilon^* > 0$. Then, by Remark 4.3 and Assumption A(ii), for any $\varepsilon \in [0, \varepsilon^*]$

$$\begin{aligned} |Z(t_1) - Z(t_2)| + \varepsilon \left| \sum_{j=1}^2 \ell_{2,j} \int_0^{t_1} \mathcal{K}(t_1,s) f_j(s) ds - \sum_{j=1}^2 \ell_{2,j} \int_0^{t_2} \mathcal{K}(t_2,s) f_j(s) ds \right| \\ \leq \Lambda_\lambda |t_1 - t_2|^\lambda + \varepsilon \left(\int_0^{t_1 \vee t_2} (\mathcal{K}(t_1,s) - \mathcal{K}(t_2,s))^2 ds \right)^{\frac{1}{2}} \left\| \sum_{j=1}^2 \ell_{2,j} f_j \right\|_{L^2([0,T])} \end{aligned}$$

$$\leq \left(\Lambda_\lambda + \varepsilon^* C_\lambda \left\| \sum_{j=1}^{2} \ell_{2,j} f_j \right\|_{L^2([0,T])}\right) |t_1 - t_2|^\lambda,$$

which implies by Remark 4.4, that $(t, Y^{\varepsilon,F}(t)) \in \mathcal{D}_{\frac{1}{\xi(\varepsilon^*)},\frac{1}{\xi(\varepsilon^*)}}$ for any $\varepsilon \in [0, \varepsilon^*]$ and $t \in [0,T]$, where $\mathcal{D}_{\cdot,\cdot}$ is defined by (4.5) and

$$\xi(\varepsilon^*) := \frac{\left(L_2 + \Lambda_\lambda + \varepsilon^* C_\lambda \left\| \sum_{j=1}^{2} \ell_{2,j} f_j \right\|_{L^2([0,T])}\right)^\alpha}{L_1} \tag{5.6}$$

for some deterministic positive constants L_1, L_2 and α that depend only on λ and the shape of the drift b. This implies, by Assumption B(ii), that, for all $s \in [0,T]$ and $0 \leq \varepsilon_1 \leq \varepsilon_2 \leq \varepsilon^*$,

$$\left|b(s, Y^{\varepsilon_2,F}(s)) - b(s, Y^{\varepsilon_1,F}(s))\right| \leq c\xi^p(\varepsilon^*) \left|Y^{\varepsilon_2,F}(s) - Y^{\varepsilon_1,F}(s)\right|,$$

whence

$$\begin{aligned}
|Y^{\varepsilon_2,F}(t) - Y^{\varepsilon_1,F}(t)| &\leq \int_0^t \left|b(s, Y^{\varepsilon_2,F}(s)) - b(s, Y^{\varepsilon_1,F}(s))\right| ds \\
&\quad + |\varepsilon_2 - \varepsilon_1| \left| \sum_{j=1}^{2} \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds \right| \\
&\leq c\xi^p(\varepsilon^*) \int_0^t \left|Y^{\varepsilon_2,F}(s) - Y^{\varepsilon_1,F}(s)\right| ds \\
&\quad + |\varepsilon_2 - \varepsilon_1| \left(\int_0^T \mathcal{K}^2(t,s) ds\right)^{\frac{1}{2}} \left\| \sum_{j=1}^{2} \ell_{2,j} f_j \right\|_{L^2([0,T])},
\end{aligned}$$

where $\int_0^T \mathcal{K}_i^2(t,s) ds < \infty$ by Assumption A(i). Now, it follows from Gronwall's inequality that, for all $0 \leq \varepsilon_1 \leq \varepsilon_2 \leq \varepsilon^*$,

$$|Y^{\varepsilon_2,F}(t) - Y^{\varepsilon_1,F}(t)| \leq \left(\int_0^T \mathcal{K}^2(t,s) ds\right)^{\frac{1}{2}} \left\| \sum_{j=1}^{2} \ell_{2,j} f_j \right\|_{L^2([0,T])} e^{cT\xi^p(\varepsilon^*)} |\varepsilon_2 - \varepsilon_1|,$$

which implies the desired absolute continuity w.r.t. ε.

Again, it is sufficient to compute the limit in probability of the form

$$\mathbb{P}\text{-}\lim_{\varepsilon \to 0} \frac{Y^{\varepsilon,F}(t) - Y^{0,F}(t)}{\varepsilon},$$

where $Y^{\varepsilon,F}(t)$ is defined by (5.5).

Using the mean-value theorem, we obtain that

$$
\begin{aligned}
Y^{\varepsilon,F}(t) - Y^{0,F}(t) &= \int_0^t (b(s, Y^{\varepsilon,F}(s)) - b(s, Y^{0,F}(s))ds \\
&\quad + \varepsilon \sum_{j=1}^2 \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds \\
&= \int_0^t \Theta_\varepsilon(s)(Y^{\varepsilon,F}(s) - Y^{0,F}(s))ds \\
&\quad + \varepsilon \sum_{j=1}^2 \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds,
\end{aligned}
\tag{5.7}
$$

where $\Theta_\varepsilon(s) := \frac{\partial b}{\partial y}\left(s, Y^{0,F}(s) + \theta_\varepsilon(s)(Y^{\varepsilon,F}(s) - Y^{0,F}(s))\right)$ with $\theta_\varepsilon(s)$ being some value between 0 and 1. It is easy to verify that (5.7) implies the representation of the form

$$
Y^{\varepsilon,F}(t) - Y^{0,F}(t) = \varepsilon \sum_{j=1}^2 \int_0^t \exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\} dG_j(s),
$$

where

$$
G_j(t) := \ell_{2,j} \int_0^t \mathcal{K}(t,s) f_j(s) ds.
$$

Note that, just like in Step 1, it follows from Remark 4.4 that for all $\varepsilon \in [0,1]$ and $s \in [0,T]$ $(s, Y^{\varepsilon,F}(s)) \in \mathcal{D}_{\frac{1}{\xi(1)}, \frac{1}{\xi(1)}}$, where $\xi(1)$ is defined via (5.6). This, together with continuity of $\frac{\partial b}{\partial y}$ due to Assumption B(iv) and compactness of $\overline{\mathcal{D}_{\frac{1}{\xi(1)}, \frac{1}{\xi(1)}}}$, implies that there exists a finite positive random variable η such that $|\Theta_\varepsilon(s)| \le \eta$ for all $\varepsilon \in [0,1]$ and $s \in [0,T]$. Moreover, the mapping $s \mapsto \exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\}$, $s \in [0,t]$, is Lipschitz and whence the integrals $\int_0^t \exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\} dG_j(s)$, $j = 1,2$, are well-defined as pathwise limits of Riemann-Stieltjes integral sums due to the classical results of Young ([51, Section 10], see also [20, Chapter 6] for a more recent overview on the subject).

Now, applying the integration by parts formula, we obtain

$$
\begin{aligned}
\frac{Y^{\varepsilon,F}(t) - Y^{0,F}(t)}{\varepsilon} &= \sum_{j=1}^2 \int_0^t \exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\} dG_j(s) \\
&= \sum_{j=1}^2 \left(G_j(t) - \int_0^t G_j(s) d\left(\exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\}\right)\right) \\
&= \sum_{j=1}^2 \left(G_j(t) + \int_0^t G_j(s)\Theta_\varepsilon(s) \exp\left\{\int_s^t \Theta_\varepsilon(v)dv\right\} ds\right),
\end{aligned}
$$

and, taking into account the definition of $G_{i,j}$,

$$\frac{Y^{\varepsilon,F}(t) - Y^{0,F}(t)}{\varepsilon} = \sum_{j=1}^{2} \int_0^T D_j(\varepsilon, t, u) f_j(u) du$$

with

$$D_j(\varepsilon, t, u) := \ell_{2,j} \left(\mathcal{K}(t, u) + \int_u^t \mathcal{K}(s, u) \Theta_\varepsilon(s) \exp\left\{ \int_s^t \Theta_\varepsilon(v) dv \right\} ds \right) \mathbb{1}_{[0,t]}(u).$$

Now, note that for all $s \in [0, T]$ and $\varepsilon \in [0, 1]$

$$|\Theta_\varepsilon(s)| \exp\left\{ \int_s^t \Theta_\varepsilon(v) dv \right\} < \eta e^{cT},$$

where c is from Assumption B(iv), and therefore we can apply the dominated convergence theorem to obtain that

$$\frac{Y^{\varepsilon,F}(t) - Y^{0,F}(t)}{\varepsilon} \to \langle DY(t), f \rangle_{L^2([0,T];\mathbb{R}^2)} \quad a.s., \quad \varepsilon \to 0,$$

where $DY(t)$ has exactly the form (5.4), which finishes the proof. □

Remark 5.4. *Higher-order Malliavin differentiability of Y can be established under stronger regularity assumptions on b (see [17]). Moreover, Theorem 5.3 plays a key role in reproducing power-law behaviour of the implied volatility skew via suitable choices of Volterra kernels.*

5.2.2 Malliavin differentiability of price.

Consider the log-price process

$$X(t) = \log S(t) = X(0) + \int_0^t \left(\mu(s) - \frac{Y^2(s)}{2} \right) ds + \int_0^t Y(s)\, dB^S(s).$$

To handle nonlinear functions such as $S(t) = e^{X(t)}$, we employ the following generalised chain rule.

Proposition 5.5. *Let $\eta = (\eta_1, ..., \eta_k) \in (\mathbb{D}^{1,2})^k$ and $f \in C^1(\mathbb{R}^k; \mathbb{R})$ be a real-valued function such that*

1) if $\frac{\partial f}{\partial x_j}$ is unbounded, then there exist sequences $\{r_n,\ n \geq 1\}$, $\{s_{j,n},\ n \geq 1\}$, $r_n, s_{j,n} \to \infty$ as $n \to \infty$, such that

$$\left| \frac{\partial f}{\partial x_j}(x_1, ..., x_k) \right| \geq s_{j,n}$$

implies that $|f(x_1, ..., x_k)| \geq r_n$;

2) $\mathbb{E}\left[|f(\eta)|^2 + \left\|\sum_{j=1}^k \frac{\partial f}{\partial x_j}(\eta) D\eta_j\right\|^2_{L^2([0,T];\mathbb{R}^2)}\right] < \infty.$

Then $f(\eta) \in \mathbb{D}^{1,2}$ *and* $Df(\eta) = \sum_{j=1}^k \frac{\partial f}{\partial x_j}(\eta) D\eta_j$.

Proof. Let $\phi \in C_c^\infty(\mathbb{R})$ be a smooth, compactly supported function satisfying:

- $\phi(x) = x$ for all $|x| \leq 1$,
- $|\phi(x)| \leq |x|$ for all $x \in \mathbb{R}$,
- $\|\phi'\|_\infty := \sup_{z\in\mathbb{R}} |\phi'(z)| < \infty$.

For each $m \in \mathbb{N}$, define $f_m(x) := m\phi\left(\frac{f(x)}{m}\right)$. We first show that each f_m has bounded partial derivatives.
Fix $j \in \{1, \ldots, k\}$. We consider two cases.

Case 1: $\frac{\partial f}{\partial x_j}$ is bounded. Then

$$\frac{\partial f_m}{\partial x_j}(x) = \frac{\partial f}{\partial x_j}(x)\phi'\left(\frac{f(x)}{m}\right),$$

and consequently

$$\left|\frac{\partial f_m}{\partial x_j}(x)\right| \leq \left\|\frac{\partial f}{\partial x_j}\right\|_\infty \|\phi'\|_\infty < \infty.$$

Case 2: $\frac{\partial f}{\partial x_j}$ is unbounded. By assumption (1), there exist sequences $\{r_n\}$, $\{s_{j,n}\}$ with $r_n, s_{j,n} \to \infty$ such that

$$\left|\frac{\partial f}{\partial x_j}(x)\right| \geq s_{j,n} \implies |f(x)| \geq r_n.$$

Choose n^* sufficiently large such that

$$\frac{r_{n^*}}{m} > \max\{|z| : z \in \operatorname{supp}(\phi)\}.$$

Now fix an arbitrary $x \in \mathbb{R}^k$ and consider two subcases.

Subcase 2a: $\left|\frac{\partial f}{\partial x_j}(x)\right| \geq s_{j,n^*}$. Then by the implication above, $|f(x)| \geq r_{n^*}$, so

$$\left|\frac{f(x)}{m}\right| \geq \frac{r_{n^*}}{m} > \max\{|z| : z \in \operatorname{supp}(\phi)\}.$$

Hence $\frac{f(x)}{m} \notin \operatorname{supp}(\phi)$, implying $\phi'\left(\frac{f(x)}{m}\right) = 0$, and therefore $\frac{\partial f_m}{\partial x_j}(x) = 0$.

Subcase 2b: $\left|\frac{\partial f}{\partial x_j}(x)\right| < s_{j,n^*}$. Then

$$\left|\frac{\partial f_m}{\partial x_j}(x)\right| = \left|\frac{\partial f}{\partial x_j}(x)\right| \left|\phi'\left(\frac{f(x)}{m}\right)\right| \leq s_{j,n^*}\|\phi'\|_\infty.$$

Combining both subcases, we obtain for all $x \in \mathbb{R}^k$:

$$\left|\frac{\partial f_m}{\partial x_j}(x)\right| \leq s_{j,n^*}\|\phi'\|_\infty < \infty.$$

Thus each partial derivative $\frac{\partial f_m}{\partial x_j}$ is bounded, and consequently $f_m \in C_b^1(\mathbb{R}^k)$. Since $\eta \in (\mathbb{D}^{1,2})^k$ and f_m has bounded partial derivatives, the classical chain rule (see e.g., [43, Proposition 1.2.3]) yields $f_m(\eta) \in \mathbb{D}^{1,2}$ and

$$Df_m(\eta) = \sum_{j=1}^k \frac{\partial f_m}{\partial x_j}(\eta) D\eta_j.$$

Now we analyse the convergence. By construction, $|f_m(\eta)| \leq |f(\eta)|$ for all m, and

$$\lim_{m\to\infty} f_m(\eta) = f(\eta)$$

almost surely. For the derivatives, note that

$$\frac{\partial f_m}{\partial x_j}(\eta) = \frac{\partial f}{\partial x_j}(\eta)\phi'\left(\frac{f(\eta)}{m}\right).$$

As $m \to \infty$, $\phi'\left(\frac{f(\eta)}{m}\right) \to \phi'(0) = 1$ almost surely. Moreover,

$$\left|\frac{\partial f_m}{\partial x_j}(\eta)\right| \leq \left|\frac{\partial f}{\partial x_j}(\eta)\right| \|\phi'\|_\infty,$$

and the right-hand side is square-integrable by assumption (2). Therefore, by the dominated convergence theorem,

$$\lim_{m\to\infty} \mathbb{E}\left[|f_m(\eta) - f(\eta)|^2\right] = 0,$$

and

$$\lim_{m\to\infty} \mathbb{E}\left[\left\|Df_m(\eta) - \sum_{j=1}^k \frac{\partial f}{\partial x_j}(\eta) D\eta_j\right\|^2_{L^2([0,T];\mathbb{R}^2)}\right] = 0.$$

The closedness of the Malliavin derivative operator D (see [43, Proposition 1.2.1]) implies that $f(\eta) \in \mathbb{D}^{1,2}$ and

$$Df(\eta) = \sum_{j=1}^k \frac{\partial f}{\partial x_j}(\eta) D\eta_j,$$

completing the proof. □

Applying this result yields the Malliavin derivative of the log-price.

Corollary 5.6. *For any $t \in [0,T]$, $X_i(t) \in \mathbb{D}^{1,2}$ and its Malliavin derivative $DX(t) = (D^1X(t), D^2X(t))$ has the form*

$$D_u^j X(t) = \left(-\int_0^t Y(s) D_u^j Y(s) ds + \ell_{1,j} Y(u) + \int_0^t D_u^j Y(s) dB_1^S(s) \right) \mathbb{1}_{[0,t]}(u).$$

Proof. First, note that if $u > t$, then the claim is true due to [43, Corollary 1.2.1]. Let now $u \leq t$. By the definition of the Malliavin derivative and the fact that deterministic functions have vanishing Malliavin derivative,

$$D_u^j \left[\int_0^t \mu(s) ds \right] = 0. \tag{5.8}$$

Next, we verify that $Y^2(s) \in \mathbb{D}^{1,2}$ for any $s \in [0,T]$. Since Y is bounded by Assumption A, we have

$$\mathbb{E}\left[Y^2(s) + \|Y(s)DY(s)\|^2_{L^2([0,T];\mathbb{R}^2)} \right] \leq \|Y\|^2_\infty + \|Y\|^2_\infty \mathbb{E}\left[\|DY(s)\|^2_{L^2([0,T];\mathbb{R}^2)} \right] < \infty,$$

where the finiteness follows from $Y(s) \in \mathbb{D}^{1,2}$. Proposition 5.5 then applies with $f(x) = x^2$, which satisfies condition (1) trivially since its derivative $2x$ is unbounded but $f(x) \to \infty$ as $|x| \to \infty$. Hence $Y^2(s) \in \mathbb{D}^{1,2}$ and $DY^2(s) = 2Y(s)DY(s)$.

By the closedness of the Malliavin derivative and the fact that the integral is a limit of Riemann sums, we can interchange the derivative and the integral:

$$D_u^j \left[\frac{1}{2} \int_0^t Y^2(s) ds \right] = \frac{1}{2} \int_0^t D_u^j Y^2(s) ds = \int_0^t Y(s) D_u^j Y(s) ds, \quad j = 1, 2. \tag{5.9}$$

Finally, using the adaptedness and boundedness of Y together with [43, Proposition 1.3.2], we compute the Malliavin derivative of the Itô integral. Recall that $B^S(t) = \ell_{1,1} W_1(t) + \ell_{1,2} W_2(t)$. Then

$$\begin{aligned} D_u^j \left[\int_0^t Y(s) dB_1^S(s) \right] &= D_u^j \left[\sum_{k=1}^2 \int_0^t \ell_{1,k} Y(s) dW_k(s) \right] \\ &= \ell_{1,j} Y(u) + \sum_{k=1}^2 \ell_{1,k} \int_0^t D_u^j Y(s) dW_k(s) \\ &= \ell_{1,j} Y(u) + \int_0^t D_u^j Y(s) dB_1^S(s), \end{aligned} \tag{5.10}$$

where the second equality uses the fact that $D_u^j W_k(s) = \delta_{jk} \mathbb{1}_{[0,s]}(u)$ and the chain rule for Malliavin derivatives of Itô integrals. Combining (5.8), (5.9), and (5.10) yields the desired expression. $\square$

5.3 Malliavin differentiability w.r.t. a transformed Brownian motion. Absolute continuity

To establish absolute continuity of the law of $X(t)$, it is sufficient to verify

$$\|D.X(t)\|_{L^2([0,T];\mathbb{R}^2)} > 0 \quad \text{a.s.}$$

However, the representation above is not convenient for this purpose. We therefore introduce a transformed Brownian motion.

Recall that $\widetilde{W} := \mathcal{U}^{-1}B(t)$ (see Section 4.1.1) is again a standard Brownian motion, and we have the relation

$$B^S(t) = \frac{1}{\sqrt{1-\rho^2}}\widetilde{W}_1(t) - \frac{\rho}{\sqrt{1-\rho^2}}\widetilde{W}_2(t), \qquad B^Y(t) = \widetilde{W}_2(t).$$

Denote by $\widetilde{D}$ the Malliavin derivative with respect to $\widetilde{W}$. The key advantage of this transformation is that the driving noise for Y becomes orthogonal to $\widetilde{W}_1$.

Theorem 5.7. *Let Assumptions A and B hold. Then, for any $t \in (0,T]$,*

$$\widetilde{D}^1_v X(t) = u_{1,1}Y(v)\mathbb{1}_{[0,t]}(v)$$

and

$$\widetilde{D}^1_v S(t) = u_{1,1}S(t)Y(v)\mathbb{1}_{[0,t]}(v),$$

where $u_{1,1}$ is the upper left element of $\mathcal{U}$. In particular,

$$\|\widetilde{D}.X(t)\|_{L^2([0,T];\mathbb{R}^2)} > 0 \quad a.s.$$

Proof. We compute $\widetilde{D}^1_v X(t)$ by differentiating the representation of $X(t)$ in terms of $\widetilde{W}$. Recall that

$$X(t) = X(0) + \int_0^t \left(\mu(s) - \frac{Y^2(s)}{2}\right) ds + \int_0^t Y(s)dB^S(s).$$

Express B^S in terms of $\widetilde{W}$:

$$B^S(s) = u_{1,1}\widetilde{W}_1(s) + u_{1,2}\widetilde{W}_2(s).$$

The Itô integral becomes

$$\int_0^t Y(s)dB^S(s) = u_{1,1}\int_0^t Y(s)d\widetilde{W}_1(s) + u_{1,2}\int_0^t Y(s)d\widetilde{W}_2(s).$$

Now we apply the Malliavin derivative $\widetilde{D}^1$ with respect to $\widetilde{W}_1$. The deterministic terms vanish. For the term involving $\widetilde{W}_1$, we use [43, Proposition 1.3.2]:

$$\widetilde{D}^1_v\left[u_{1,1}\int_0^t Y(s)d\widetilde{W}_1(s)\right] = u_{1,1}Y(v)\mathbb{1}_{[0,t]}(v) + u_{1,1}\int_0^t \widetilde{D}^1_v Y(s)d\widetilde{W}_1(s).$$

For the term involving $\widetilde{W}_2$:

$$\widetilde{D}_v^1\left[u_{1,2}\int_0^t Y(s)d\widetilde{W}_2(s)\right] = u_{1,2}\int_0^t \widetilde{D}_v^1 Y(s)d\widetilde{W}_2(s).$$

It remains to show that $\widetilde{D}_v^1 Y(s) = 0$ for all s, v. Since Y satisfies an SDE driven by B^Y, and $B^Y = u_{2,2}\widetilde{W}_2$ (as $\widetilde{W}_1$ does not appear), Y is adapted to the filtration generated by $\widetilde{W}_2$. The Malliavin derivative $\widetilde{D}^1$ with respect to $\widetilde{W}_1$ of any random variable that is measurable with respect to $\sigma(\widetilde{W}_2(s) : s \leq T)$ is zero. This follows from the fact that $\widetilde{W}_1$ and $\widetilde{W}_2$ are independent Brownian motions, and the Malliavin derivative with respect to $\widetilde{W}_1$ vanishes on functionals of $\widetilde{W}_2$ (see [43, Proposition 1.2.4]). Therefore, $\widetilde{D}_v^1 Y(s) = 0$ for all v, s. Consequently, the integral terms involving $\widetilde{D}^1 Y$ vanish, yielding

$$\widetilde{D}_v^1 X(t) = u_{1,1} Y(v)\mathbb{1}_{[0,t]}(v).$$

For $S(t) = e^{X(t)}$, we apply Proposition 5.5 with $f(x) = e^x$. Since $f'(x) = e^x$ is unbounded but $f(x) \to \infty$ as $x \to \infty$, condition (1) is satisfied with $r_n = e^{s_n}$ for any sequence $s_n \to \infty$. Condition (2) holds because $\mathbb{E}[e^{2X(t)}] < \infty$ by the properties of geometric Brownian motion and $\widetilde{D}^1 X(t)$ is square-integrable. Hence,

$$\widetilde{D}_v^1 S(t) = S(t)\widetilde{D}_v^1 X(t) = u_{1,1} S(t) Y(v)\mathbb{1}_{[0,t]}(v).$$

Finally, we compute the L^2-norm:

$$\|\widetilde{D}.X(t)\|^2_{L^2([0,T];\mathbb{R}^2)} = \int_0^T |u_{1,1}Y(v)\mathbb{1}_{[0,t]}(v)|^2 dv = u_{1,1}^2 \int_0^t Y(v)^2 dv.$$

Since $u_{1,1} > 0$ (the Cholesky factor has positive diagonal entries) and $Y(v)$ is strictly positive almost surely for all $v \in [0, T]$ by Assumption A (or an appropriate positivity condition), we have $\int_0^t Y(v)^2 dv > 0$ almost surely. Therefore,

$$\|\widetilde{D}.X(t)\|_{L^2([0,T];\mathbb{R}^2)} > 0 \quad \text{a.s.},$$

completing the proof. □

Corollary 5.8. *For each $t \in (0, T]$, the law of $X(t)$ (and thus $S(t) = e^{X(t)}$) admits a continuous and bounded density.*

Proof. By Theorem 5.7, we have

$$\|\widetilde{D}.X(t)\|_{L^2([0,T];\mathbb{R}^2)} > 0 \quad \text{a.s.}$$

Moreover, $\widetilde{D}X(t)$ is adapted to the filtration generated by $\widetilde{W}$ (specifically, it is a function of Y and thus adapted). The Malliavin derivative $\widetilde{D}X(t)$ belongs to $\mathbb{D}^{1,2}$ and satisfies the non-degeneracy condition. By the Malliavin regularity criterion (see e.g., [43, Proposition

2.1.1]), the law of $X(t)$ is absolutely continuous with respect to Lebesgue measure and admits a continuous and bounded density.

For $S(t) = e^{X(t)}$, since the exponential function is smooth and strictly monotone on $\mathbb{R}$, the change of variables formula implies that $S(t)$ also admits a density. Specifically, if $p_{X(t)}(x)$ denotes the density of $X(t)$, then the density of $S(t)$ is

$$p_{S(t)}(s) = \frac{1}{s} p_{X(t)}(\log s) \mathbb{1}_{(0,\infty)}(s),$$

which is continuous and bounded on $(0, \infty)$ because $p_{X(t)}$ is continuous and bounded on $\mathbb{R}$ and $s \mapsto 1/s$ is continuous on $(0, \infty)$. Note that $S(t)$ is almost surely positive, so its density is supported on $(0, \infty)$. $\square$

6 Rough Volatility Models Within the SVV Framework

The stochastic Volterra volatility (SVV) framework developed by Di Nunno et al. [15] provides a unified approach that encompasses several important rough volatility models. A model belongs to this family if its volatility process can be represented as a Volterra integral of a Gaussian process with a kernel satisfying the regularity conditions in Assumption A.

6.1 Classification of rough volatility models within the SVV Framework

We analyse whether each model can be represented in the canonical SVV form:

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))ds + \int_0^t \mathcal{K}(t,s)dB^Y(s),$$

where $\mathcal{K}$ satisfies Assumption A, and the drift b is locally Lipschitz.

Model	**Fits SVV?**	**Remarks**
RFSV	If $H = 1/2$	Requires rewriting to separate drift and Volterra diffusion.
Rough Bergomi	Yes	$\log \sigma_t$ fits the SVV form with zero drift and fractional kernel.
Mixed Rough Bergomi	Yes	Each exponential component fits; convex combination does not affect volatility representation.
Rough SABR	Yes	$\log \sigma_t$ fits SVV form with zero drift and fractional kernel.
Rough Stein-Stein	Yes	σ_t fits directly with zero drift.
Rough Heston	No	The drift is inside the Volterra convolution; cannot be separated.
Quadratic Rough Heston	No	Same issue as rough Heston; drift is inside the Volterra convolution.

Table 1: Classification of rough volatility models within the SVV framework

6.1.1 Detailed analysis

1. Rough Fractional Stochastic Volatility (RFSV) Model [25] The model is

$$\sigma(t) = \exp\left\{\theta_1 + \theta_2 \int_{-\infty}^t e^{-\theta_3(t-s)}dB^H(s)\right\}.$$

Define $Y(t) = \log \sigma(t) - \theta_1 = \theta_2 \int_{-\infty}^{t} e^{-\theta_3(t-s)} dB^H(s)$. This can be rewritten as

$$Y(t) = \theta_2 e^{-\theta_3 t} Y(0) + \theta_2 \int_0^t e^{-\theta_3(t-s)} dB^H(s),$$

where $Y(0) = \int_{-\infty}^{0} e^{\theta_3 s} dB^H(s)$. However, to fit the SVV form, we need $Y(t) = Y(0) + \int_0^t b(s, Y(s))ds + \int_0^t \mathcal{K}(t,s)dB^Y(s)$. For the RFSV model:

- The process satisfies the SDE $dY(t) = -\theta_3 Y(t)dt + \theta_2 dB_t^H$ when the driving noise is standard Brownian motion (i.e., $H = 1/2$). For fractional Brownian motion with $H \neq 1/2$, the representation is more complex.
- If B^H is a standard Brownian motion ($H = 1/2$), then $Y(t)$ is an Ornstein-Uhlenbeck process, which fits the SVV form with $\mathcal{K}(t,s) = \theta_2$ (constant kernel) and drift $b(s,y) = -\theta_3 y$.
- For general $H \in (0, 1/2)$, the process is driven by fractional Brownian motion, which does not fit the SVV framework as the noise is not a semimartingale.

Conclusion: Only for $H = 1/2$ (the classical case) does this fit. For rough volatility ($H < 1/2$), the RFSV model does not fit the SVV form because the driving noise is fractional Brownian motion, not a standard Brownian motion integrated against a Volterra kernel.

2. Rough Bergomi Model [5, 34] The model is

$$\begin{aligned} dS(t) &= \sqrt{v(t)} S(t) dW(t), \\ v(t) &= \theta_0(t) \exp\left\{ \int_0^t (t-s)^{H-\frac{1}{2}} dB(s) - \theta_1^2 t^{2H} \right\}. \end{aligned}$$

Define $Y(t) = \log v(t) - \log \theta_0(t) + \theta_1^2 t^{2H} = \int_0^t (t-s)^{H-1/2} dB(s)$. Then:

- $Y(t)$ is a Volterra process of the form $Y(t) = \int_0^t \mathcal{K}(t,s)dB(s)$ with $\mathcal{K}(t,s) = (t-s)^{H-1/2}$.
- This fits the SVV form with $Y(0) = 0$, $b \equiv 0$, and $\mathcal{K}(t,s) = (t-s)^{H-1/2}$.
- The kernel satisfies Assumption A for $H \in (0, 1/2)$.
- The volatility is recovered as $v(t) = \theta_0(t) \exp\{Y(t) - \theta_1^2 t^{2H}\}$, where the deterministic term $-\theta_1^2 t^{2H}$ ensures $\mathbb{E}[v(t)] = \theta_0(t)$.

Conclusion: Yes, with $Y(t)$ defined as the centered log-volatility.

3. Mixed Rough Bergomi Model [27, 37] The model is

$$
\begin{aligned}
dS(t) &= \sqrt{v(t)}S(t)dB(t), \\
v(t) &= \theta_0(t)\left(\delta \exp\left\{2\theta_1 \int_0^t (t-s)^{H-\frac{1}{2}}dW(s) - \theta_1^2 t^{2H}\right\}\right. \\
&\qquad \left. +(1-\delta)\exp\left\{2\theta_2 \int_0^t (t-s)^{H-\frac{1}{2}}dW(s) - \theta_2^2 t^{2H}\right\}\right).
\end{aligned}
$$

Define

$$
Y_1(t) = 2\theta_1 \int_0^t (t-s)^{H-1/2}dW(s), \quad Y_2(t) = 2\theta_2 \int_0^t (t-s)^{H-1/2}dW(s).
$$

Then:

- Both $Y_1(t)$ and $Y_2(t)$ are Volterra processes of the form $\int_0^t \mathcal{K}(t,s)dW(s)$ with kernel $\mathcal{K}(t,s) = (t-s)^{H-1/2}$ (up to constant multiples).
- Each fits the SVV form with zero drift.
- The volatility is a convex combination of exponentials of these processes: $v(t) = \theta_0(t)(\delta e^{Y_1(t)-\theta_1^2 t^{2H}} + (1-\delta)e^{Y_2(t)-\theta_2^2 t^{2H}})$.
- While the volatility itself is not a single Volterra process, each component is. The SVV framework describes the dynamics of $Y(t)$ (the log-volatility or a related process). Here, we would take $Y(t)$ to be, say, $Y_1(t)$, and the mixed structure appears at the level of v_t, not $Y(t)$.

Conclusion: Yes, if we take $Y(t)$ as the driving Volterra process (e.g., $Y(t) = \int_0^t (t-s)^{H-1/2}dW(s)$), then v_t is a function of $Y(t)$ (through two exponentials with different scalings), which is covered by the chain rule.

4. Rough SABR Model [22] The model is

$$
\begin{aligned}
dS(t) &= \sqrt{v(t)}\beta(S(t))dW(t), \\
v(t) &= \theta_0(t)\exp\left\{\theta\sqrt{2H}\int_0^t (t-u)^{H-\frac{1}{2}}dB(u) - \frac{1}{2}\theta^2 t^{2H}\right\}.
\end{aligned}
$$

Define $Y(t) = \log v_t - \log \theta_0(t) + \frac{1}{2}\theta^2 t^{2H} = \theta\sqrt{2H}\int_0^t (t-u)^{H-1/2}dB_u$. Then:

- $Y(t)$ is a Volterra process with kernel $\mathcal{K}(t,u) = \theta\sqrt{2H}(t-u)^{H-1/2}$.
- This fits the SVV form with $Y(0) = 0$, $b \equiv 0$, and $\mathcal{K}$ as above.
- The volatility is recovered as $v(t) = \theta_0(t)\exp\{Y(t) - \frac{1}{2}\theta^2 t^{2H}\}$.

Conclusion: Yes, exactly like the rough Bergomi model.

5. Rough Stein-Stein Model [29] The model is

$$dS(t) = \theta v(t)S(t)dW(t),$$
$$v(t) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \int_0^t (t-s)^{H-\frac{1}{2}} dB(s).$$

Define $Y(t) = v(t)$. Then:

- $Y(t)$ is directly a Volterra process with kernel $\mathcal{K}(t,s) = \frac{1}{\Gamma(H+1/2)}(t-s)^{H-1/2}$.
- This fits the SVV form with $Y(0) = 0$, $b \equiv 0$, and $\mathcal{K}$ as above.
- No exponential transformation is required.

Conclusion: Yes, directly.

6. Rough Heston Model [19] The model is

$$dS(t) = \sqrt{v(t)}S(t)dW(t),$$
$$v(t) = v(0) + \int_0^t \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \Big(\theta_1(\theta_2 - v(s))\, ds + \theta_3\sqrt{v(s)}\, dB(s)\Big).$$

This is a Volterra equation where both the drift and diffusion appear inside the Volterra convolution:

$$v_t = v_0 + \int_0^t \mathcal{K}(t,s)\theta_1(\theta_2 - v(s))ds + \int_0^t \mathcal{K}(t,s)\theta_3\sqrt{v(s)}dB(s).$$

The canonical SVV form requires:

$$Y(t) = Y(0) + \int_0^t b(s, Y(s))ds + \int_0^t \mathcal{K}(t,s)dB^Y(s).$$

In the rough Heston model:

- The drift term $\theta_1(\theta_2 - v(s))$ is inside the Volterra convolution with $\mathcal{K}(t,s)$, not outside as a standard Itô integral.
- The kernel multiplies both the drift and the diffusion, which is a different structure.
- This is a Volterra-type SDE, not a standard SDE with Volterra noise.

Conclusion: No, the rough Heston model does not fit the canonical SVV form because the drift is inside the Volterra convolution.

7. Quadratic Rough Heston Model [26, 47] The model is

$$
\begin{aligned}
dS(t) &= \sqrt{v(t)}S(t)dW(t), \\
v(t) &= a(Z(t)-b)^2 + c, \\
Z(t) &= \int_0^t \theta_1 \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \left(\theta_0(s) - Z(s)\right) ds \\
&\quad + \int_0^t \theta_2 \frac{(t-s)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \sqrt{a(Z(s)-b)^2+c}\, dW(s).
\end{aligned}
$$

The same issue arises: $Z(t)$ satisfies a Volterra equation where both drift and diffusion are inside the Volterra convolution. This is the same structural form as the rough Heston model. **Conclusion:** No, for the same reason as rough Heston.

6.1.2 Summary

1. **Models that fit the SVV form directly:**
   - Rough Bergomi: $\log v(t)$ is a Volterra process with zero drift.
   - Mixed Rough Bergomi: Each component is a Volterra process; the convex combination is handled via chain rule.
   - Rough SABR: Same structure as rough Bergomi.
   - Rough Stein-Stein: $v(t)$ itself is a Volterra process.
2. **Models with special case:**
   - RFSV: For $H = 1/2$ it fits as an Ornstein-Uhlenbeck process; for $H < 1/2$ it is driven by fractional Brownian motion, which does not fit the SVV framework.
3. **Models that do NOT fit the SVV form:**
   - Rough Heston: Drift is inside the Volterra convolution.
   - Quadratic Rough Heston: Same structure as rough Heston.

The key distinction is that in the canonical SVV form, the Volterra kernel appears only in the diffusion term, multiplying the Brownian increment. The drift is a standard Itô integral with respect to time. Models where the drift also appears inside the Volterra convolution (like rough Heston) have a different structure and are not covered by this SVV framework.

7 Conclusion

In this work, we have investigated fractional and rough volatility models from both theoretical and structural perspectives. After introducing the mathematical foundations of asset-price modelling and fractional Brownian motion, we motivated the use of rough volatility models as a framework capable of reproducing important empirical features of financial markets.

Our main focus was the Sandwiched Volterra Volatility (SVV) model introduced by Di Nunno et al. [15] We reformulated the general framework in a one-dimensional setting, providing a more accessible presentation of the model and establishing several fundamental results with detailed proofs. Furthermore, we examined the Malliavin differentiability of both the volatility and asset-price processes, thereby extending the theoretical understanding of the SVV framework and confirming its analytical tractability.

The final part of the thesis was devoted to identifying which classical rough volatility models can be represented within the SVV framework. We showed that a large class of rough volatility models admits such a representation, highlighting the flexibility and generality of the SVV approach. At the same time, we identified notable exceptions, namely the Rough Fractional Stochastic Volatility model for $H \neq \frac{1}{2}$, the rough Heston model, and the quadratic rough Heston model, which fall outside the SVV class.

Overall, our results demonstrate that the SVV framework provides a unifying perspective for a broad family of rough volatility models while retaining strong analytical properties. These findings contribute to a deeper understanding of the relationships between different rough volatility specifications and may serve as a basis for future research on the theoretical properties, numerical methods, and practical applications of Volterra-type stochastic volatility models.

A Appendix

This appendix provides essential definitions and theorems necessary for understanding the main text. For full details, see Karatzas and Shreve [35].

A.1 Stochastic calculus

We provide formal definitions of concepts in stochastic analysis.

Definition A.1 (Probability Space)**.** *Let Ω be a set and $\mathcal{F}$ a σ-algebra on Ω, so that $(\Omega, \mathcal{F})$ is a measurable space.*
A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a probability measure if it is a measure and satisfies

$$\mathbb{P}(\Omega) = 1.$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Definition A.2 (Stochastic Processes)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process $X_t := \{X(t) : t \in [0, T]\}$ is a family of random variables such that for each $t \in [0, T]$ and $T \in (0, \infty)$ (interpreted as the notion of time), each $X(t)$ is a measurable function $X_t : \Omega \to \mathbb{R}$. That is, for each $t \in [0, T]$ and each Borel set $B \subseteq \mathbb{R}$, the set $\{\omega \in \Omega : X(t, \omega) \in B\}$ belongs to $\mathcal{F}$.*

Definition A.3 (Adaptedness)**.** *A filtration $\{\mathcal{F}_t\}_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family of σ-algebras contained in $\mathcal{F}$, i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$, for all $0 \leq s < t$. A continuous time stochastic process X_t is said to be adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if X_t is $\mathcal{F}_t$-measurable, for all $t \in [0, T]$.*

Definition A.4 (Martingale)**.** *A continuous time stochastic process X_t is a martingale if it is $\mathcal{F}_t$-adapted for all $t \in [0, T]$, it is integrable, i.e. $\mathbb{E}[|X_t|] < \infty$ for all $t \in [0, T]$, and $\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$ for all $0 \leq s \leq t$.*

Definition A.5 (Brownian motion / Wiener process)**.** *A stochastic process $W(t)$ for $t \in [0, T]$ is called a Brownian motion (or Wiener process) if:*

- *$W(0) = 0$ almost surely;*
- *it has independent increments;*
- *for any $0 \leq s < t$, the increment $W(t) - W(s)$ follows a Gaussian distribution*

$$W(t) - W(s) \sim \mathcal{N}(0, t - s);$$

It is well-known that a stochastic equation of the form

$$X(t) = x_0 + \int_0^t f(s, X_s)\, ds + \int_0^t g(s, X_s)\, dW(s)$$

can be written in differential form as the stochastic differential equation (SDE)

$$\begin{aligned} dX(t) &= f(t, X_t)\,dt + g(t, X_t)\,dW(t), \quad t \in [0, T], \\ X(0) &= x_0 \in \mathbb{R}. \end{aligned} \tag{A.1}$$

where $f(t, X_t)$ is a deterministic drift function, perturbed by a diffusive function $g(t, X_t)$ and $dW(t)$ which is the increment of the Wiener process.

A.2 Stochastic integrals

This subsection aims to motivate the reader to understand the sensitivity and complexity of dealing with stochastic integrals, although it can be skipped for a general reading.

It is noteworthy that if $g(t, X_t) = 0$, (A.1) reduces to an ordinary differential equation. Therefore, our focus lies in understanding the stochastic integral $\int_0^t g(s, X_s)\,dW(s)$.

An important observation regarding this type of integral is that it cannot be solved using classical Riemann or Lebesgue integration methods, nor the Riemann-Stieltjes integral. This is because the continuous sample paths of a Wiener process lack bounded variation on any bounded time interval, due to the non-differentiability of Wiener processes. Thus, our current objective is to define the stochastic integral $\int_0^T X(t)\,dW(t)$, where we integrate with respect to the Wiener process.

Given a deterministic function $f(t) : \mathbb{R} \to \mathbb{R}$, the integral $\int_0^T f(t)dt$ can be approximated by the Riemann sum:

$$\sum_{i=0}^{N-1} f(t_i)(t_{i+1} - t_i), \tag{A.2}$$

where $t_i = idt$ with the step size $dt = T/N$ for some positive integer N. Note that if we take the limit as $dt \to 0$, we can obtain the integral. Then, we discretize the stochastic integral using the forward Euler method as follows:

$$\sum_{i=0}^{N-1} f(t_i)(W(t_{i+1}) - W(t_i)) = \sum_{i=0}^{N-1} f(t_i)\Delta W_i.$$

Let $f(t_i) \equiv W(t_i)$ and take the expectation. By the properties of the independent increments and the Gaussian distribution of the Wiener process, we obtain:

$$\mathbb{E}\left[\sum_{i=0}^{N-1} W(t_i)\Delta W_i\right] = \sum_{i=0}^{N-1} \mathbb{E}[W(t_i)\Delta W_i] = \sum_{i=0}^{N-1} \mathbb{E}[W(t_i)]\mathbb{E}[\Delta W_i] = 0.$$

Now, if we discretize it using the backward Euler method:

$$\sum_{i=0}^{N-1} W(t_{i+1})\Delta W_i,$$

taking the expectation will yield a different result:

$$\sum_{i=0}^{N-1} \mathbb{E}[W(t_{i+1})\Delta W_i] = \sum_{i=0}^{N-1} \mathbb{E}[W(t_i)\Delta W_i]\mathbb{E}[(\Delta W_i)^2] = \sum_{i=0}^{N-1} \Delta t = T.$$

Furthermore, if we use the trapezoidal or midpoint method which leads to the Stratonovich integral:

$$\sum_{i=0}^{N-1} \frac{W(t_{i+1}) + W(t_i)}{2} \Delta W_i,$$

we will find:

$$\sum_{i=0}^{N-1} \mathbb{E}\left[\frac{W(t_{i+1}) + W(t_i)}{2} \Delta W_i\right] = \frac{T}{2}.$$

As we see, all three methods do not converge to the same limit. This shows the main problem when dealing with stochastic processes due to their sensitivity to the points used to define the heights of the rectangles. In this thesis, we will focus on the forward Euler method as in (A.2), known as the Itô integral.

A.3 Itô formula

Now we define rigorously the Itô integral mentioned above (see Itô's own works [32] and [33]). To start we shall consider the probability space $(\Omega, \mathbb{F}, \mathbb{P})$, where $\mathbb{F}$ is the σ-algebra containing an increasing family $\{\mathcal{F}_t\}_{t\geq 0}$, and we say the Wiener process $W(t)$ is a martingale i.e. $W(t)$ is $\mathcal{F}_t$-adapted with $\mathbb{E}[W(t)|\mathcal{F}_0] = W(0) = 0$ and $\mathbb{E}[W(t) - W(s)|\mathcal{F}_s] = W(s) - W(s) = 0$ almost surely, for all $0 \leq s \leq t$.

Definition A.6 (Itô integral)**.** *Let $W(t)$ be a Brownian motion and let $X(t)$ be an adapted process such that $\mathbb{E}\int_0^T X(t)^2 dt < \infty$.*
For simple processes of the form

$$X(t) = \sum_{i=0}^{n-1} X_{t_i} \mathbf{1}_{[t_i, t_{i+1})}(t),$$

the Itô integral is defined as

$$\int_0^T X(t)\, dW(t) := \sum_{i=0}^{n-1} X_{t_i}\big(W_{t_{i+1}} - W_{t_i}\big).$$

For general square-integrable adapted processes, the integral is defined as the L^2-limit of such simple processes.

One can observe that the Itô integral has mean 0 as in the previous section, that is,

$$\mathbb{E}[I(X(t))] = \mathbb{E}\left[\int_0^T X(t)\, dW(t)\right] = 0,$$

by the property of the Wiener process, and this leads to the following theorem.

Theorem A.7 (Itô isometry)**.** *Let $W : [0,T] \times \Omega \to \mathbb{R}$ denote the Wiener process and let $X : [0,T] \times \Omega \to \mathbb{R}$ be an $\mathcal{F}_t$-adapted stochastic process and defined on the Hilbert space $L^2(\Omega, \mathbb{F}, \mathbb{P})$, where L^2 denotes the space of square-integrable functions, then*

$$\mathbb{E}[I(X_t)^2] = \mathbb{E}\left[\left(\int_0^T X_t \, dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T X_t^2 \, dt\right].$$

As a consequence, the Itô integral respects the inner products, i.e.

$$\mathbb{E}[I(X(t))I(Y(t))] = \mathbb{E}\left[\left(\int_0^T X(t) \, dW(t)\right)\left(\int_0^T Y(t) \, dW(t)\right)\right] = \mathbb{E}\left[\int_0^T X(t)Y(t) \, dt\right].$$

Nonetheless, the classical chain rule fails, here we state the multidimensional Itô lemma.

Definition A.8. *Let $m \in \mathbb{N}$. A vector $\{W_t^1, W_t^2, ..., W_t^m\}_{t \in [0,T]}$ of m independent Brownian motions defined on a common probability space $(\Omega, \mathbb{F}, \mathbb{P})$ is called a standard m-dimensional Brownian motion.*

Definition A.9. *Let $d \in \mathbb{N}$. A d-dimensional Itô process is a vector $\{X_t^1, X_t^2, ..., X_t^d\}_{t \in [t_0,T]}$ of Itô processes of form*

$$X_t^i = X_0^i + \int_{t_0}^t \mu_s^i ds + \sum_{j=1}^m \int_{t_0}^t \sigma_s^{ij} dW_s^j, \; t \in [t_0, T], \; i \in \{1, ..., d\}, \tag{A.3}$$

where for all $i \in \{1, ..., d\}$ and $j \in \{1, .., m\}$ we have $X_0^i \in \mathbb{R}$, $\{\mu_t^i\}_{t \in [t_0,T]}$ is an adapted stochastic process such that

$$\mathbb{P}\left(\int_{t_0}^T |\mu_t^i| dt < \infty\right) = 1$$

and $\{\sigma_t^{ij}\}_{t \in [t_0,T]}$ is $\{\mathcal{F}_t\}_{t \in [t_0,T]}$-adapted such that $\mathbb{P}\left(\int_{t_0}^T |\sigma_t^i|^2 dt < \infty\right) = 1$.

Theorem A.10 (Multidimensional Itô lemma)**.** *Let $\{X_t\}_{t \in [t_0,T]}$ be a d-dimensional Itô process of the form* (A.3) *and $Y_t = U(t, X_t) : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, for $0 \leq t \leq T$, where U is continuously differentiable in the first variable t and twice continuously differentiable in the other variables $x_1, ..., x_d$. Then*

$$dY_t = \frac{\partial U}{\partial t}(t, X_t)dt + \sum_{i=1}^d \frac{\partial U}{\partial x_i}(t, X_t)dX_t^i + \frac{1}{2}\sum_{i,k=1}^d \frac{\partial U}{\partial x_i x_k}(t, X_t)dX_t^i dX_t^k.$$

A.4 Existence and uniqueness of SDE solutions

Here we present some sufficient, but not necessary conditions to guarantee the existence and uniqueness of the solution of a stochastic differential of a form as (A.1), regarding its drift and diffusive functions.

Definition A.11 (Lipschitz continuity). *A real-valued function $f(x) : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous if there exists a constant $K > 0$ such that*

$$|f(x) - f(y)| \leq K|x - y|, \text{ for all } x, y \in \mathbb{R}.$$

Furthermore, it is locally Lipschitz continuous if $\forall n \in \mathbb{N}$ there exists a constant $K_n > 0$ such that

$$|f(x) - f(y)| \leq K_n|x - y|, \text{ for all } |x| \leq n \text{ and } |y| \leq n.$$

Proposition A.12. *A real-valued function $f(x) : \mathbb{R} \to \mathbb{R}$ satisfies a global Lipschitz condition if it has a uniformly bounded first derivative.*

Proof. By the mean value theorem, for any $x, y \in \mathbb{R}$, we have $|f(x) - f(y)| = |x - y||f'(\epsilon)|$, where $\epsilon \in [x, y]$. If $f'(x)$ is bounded, then there exists a constant $K > 0$, such that $|f'(\epsilon)| \leq K$, therefore it satisfies the Lipschitz continuity. □

Definition A.13 (Linear growth). *A real-valued function $f(x) : \mathbb{R} \to \mathbb{R}$ has a linear growth bound if there exists a constant $K > 0$ such that*

$$|f(x)|^2 \leq K^2(1 + |x|^2), \text{ for all } x, y \in \mathbb{R}.$$

Theorem A.14 (See [35], Chapter 5). *Assume that functions $f(t, x)$, $g(t, x) : [0, T] \times \mathbb{R} \to \mathbb{R}$ are Lipschitz continuous and have a linear growth bound in the second variable x. Then, the stochastic differential equation*

$$\begin{cases} dX_t = f(X_t)dt + g(X_t)dW_t, \ \forall t \in [0, T] \\ X_0 = x_0 \in \mathbb{R} \end{cases}$$

admits a unique continuous exact solution $\{X_t\}_{t\in[0,T]}$. More precisely,

$$\mathbb{P}\left(\int_0^T |f(X_t)| + |g(X_t)|^2 \, dt < \infty\right) = 1$$

and

$$\mathbb{P}\left(X_t = x_0 + \int_0^t f(X_t)\, dt + \int_0^t g(X_t)\, dW_t\right) = 1.$$

A.5 Novikov condition and Girsanov theorem

Definition A.15 (Local Martingale). *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions. An adapted càdlàg process $\{M(t)\}_{t\in[0,T]}$ is called a* local martingale *if there exists an increasing sequence of stopping times $\{\tau_n\}_{n\geq1}$ such that*

$$\tau_n \uparrow \infty \quad a.s.,$$

and for every n, the stopped process

$$M^{\tau_n}(t) := M(t \wedge \tau_n)$$

is a martingale with respect to $\{\mathcal{F}_t\}$.

Definition A.16 (Doléans-Dade Exponential). *Let $\{M(t)\}_{t\in[0,T]}$ be a continuous local martingale with $M(0) = 0$. The Doléans-Dade exponential (or stochastic exponential) of M is the unique process $\{\mathcal{E}(M)(t)\}_{t\in[0,T]}$ satisfying the stochastic differential equation*

$$d\mathcal{E}(M)(t) = \mathcal{E}(M)(t)\, dM(t), \qquad \mathcal{E}(M)(0) = 1.$$

The explicit solution is given by

$$\mathcal{E}(M)(t) = \exp\left(M(t) - \frac{1}{2}\langle M\rangle(t)\right),$$

where $\langle M\rangle(t)$ denotes the quadratic variation of M.

The Doléans-Dade exponential has the following properties:

1. If M is a local martingale, then $\mathcal{E}(M)$ is a local martingale.
2. If M is a martingale, $\mathcal{E}(M)$ is not necessarily a martingale; it is only a local martingale.
3. For two continuous local martingales M and N, the following multiplicative property holds:
$$\mathcal{E}(M)\mathcal{E}(N) = \mathcal{E}\left(M + N + \langle M, N\rangle\right).$$

Theorem A.17 (Novikov Condition). *Let $\{W(t)\}_{t\in[0,T]}$ be a one-dimensional standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions. Let $\{\theta(t)\}_{t\in[0,T]}$ be a progressively measurable real-valued process such that*

$$\int_0^T \theta(t)^2\, dt < \infty \quad a.s.$$

Define the Doléans-Dade exponential

$$\mathcal{E}_T(\theta \cdot W) := \exp\left(\int_0^T \theta(t)\, dW(t) - \frac{1}{2}\int_0^T \theta(t)^2\, dt\right).$$

If the Novikov condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \theta(t)^2\, dt\right)\right] < \infty,$$

holds, then $\{\mathcal{E}_t(\theta \cdot W)\}_{t\in[0,T]}$ is a martingale, and consequently

$$\mathbb{E}\big[\mathcal{E}_T(\theta \cdot W)\big] = 1.$$

Theorem A.18 (Girsanov Theorem). *Let $\{W(t)\}_{t\in[0,T]}$ be a one-dimensional standard Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. Let $\{\theta(t)\}_{t\in[0,T]}$ be a progressively measurable real-valued process satisfying the Novikov condition*

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{1}{2}\int_0^T \theta(t)^2\, dt\right)\right] < \infty.$$

Define the probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}_T(\theta \cdot W) = \exp\left(\int_0^T \theta(t)\, dW(t) - \frac{1}{2}\int_0^T \theta(t)^2\, dt\right).$$

Then $\mathbb{Q}$ is a probability measure equivalent to $\mathbb{P}$, and the process

$$\widetilde{W}(t) := W(t) - \int_0^t \theta(s)\, ds, \qquad t \in [0, T],$$

is a one-dimensional standard Brownian motion under $\mathbb{Q}$.

Corollary A.19 (Girsanov for Semimartingales)**.** *Under the same assumptions, if a process $\{X(t)\}_{t\in[0,T]}$ has the $\mathbb{P}$-decomposition*

$$X(t) = X(0) + \int_0^t \mu(s)\, ds + \int_0^t \sigma(s) \cdot dW(s),$$

then under $\mathbb{Q}$, X admits the decomposition

$$X(t) = X(0) + \int_0^t \big(\mu(s) + \sigma(s) \cdot \theta(s)\big)\, ds + \int_0^t \sigma(s) \cdot d\widetilde{W}(s),$$

where $\widetilde{W}$ is a $\mathbb{Q}$-Brownian motion.

References

[1] Alòs, E., León, J. A., and Vives, J. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and stochastics 11*, 4 (2007), 571–589.

[2] Avellaneda, M., Levy, A., and Parás, A. Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied mathematical finance 2*, 2 (1995), 73–88.

[3] Azmoodeh, E., Sottinen, T., Viitasaari, L., and Yazigi, A. Necessary and sufficient conditions for Hölder continuity of Gaussian processes. *Statistics & Probability Letters 94* (2014), 230 – 235.

[4] Bachelier, L. Théorie de la spéculation. *Annales Scientifiques de l'Ecole Normale Superieure. Quatrieme Serie 17* (1900), 21–86.

[5] Bayer, C., Friz, P., and Gatheral, J. Pricing under rough volatility. *Quantitative finance 16*, 6 (2015), 887–904.

[6] Biagini, F., Hu, Y., Øksendal, B., and Zhang, T. *Stochastic Calculus for Fractional Brownian Motion and Applications.* Probability and Its Applications. Springer, 2008.

[7] Black, F., and Scholes, M. The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 3 (1973), 637–654.

[8] Cheridito, P. Mixed fractional brownian motion. *Bernoulli 7*, 6 (2001), 913–934.

[9] Cherny, A. S., and Engelbert, H.-J. *Singular Stochastic Differential Equations.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[10] Comte, F., and Renault, E. Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 4 (1998), 291–323.

[11] Cont, R., and Das, P. Rough volatility: fact or artefact?, 2023.

[12] Cox, J. C., Ingersoll, J. E., and Ross, S. A. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society 53*, 2 (1985), 385–407.

[13] Cox, J. C., and Ross, S. A. The valuation of options for alternative stochastic processes. *Journal of financial economics 3*, 1–2 (1976), 145–166.

[14] Derman, E., and Kani, I. Riding on the smile. *Risk 7*, 2 (1994), 32–39.

[15] Di Nunno, G., Mishura, Y., and Yurchenko-Tytarenko, A. Option pricing in Volterra sandwiched volatility model. *ArXiv:2209.10688* (2022).

[16] Di Nunno, G., Mishura, Y., and Yurchenko-Tytarenko, A. Sandwiched SDEs with unbounded drift driven by Hölder noises. *Advances in applied probability 55*, 3 (2023), 927–964.

[17] Di Nunno, G., and Yurchenko-Tytarenko, A. Power law in Sandwiched Volterra Volatility model. *Modern Stochastics Theory and Applications* (2024), 169–194.

[18] Dupire, B. Pricing with a smile. *Risk 7* (1994), 18–20.

[19] El Euch, O., and Rosenbaum, M. The characteristic function of rough Heston models. *Mathematical Finance 29*, 1 (2019), 3–38.

[20] Friz, P. K., and Victoir, N. B. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[21] Fukasawa, M. Volatility has to be rough. *Quantitative finance 21*, 1 (2021), 1–8.

[22] Fukasawa, M., and Gatheral, J. A rough SABR formula. *Frontiers of Mathematical Finance 1*, 1 (2022), 81.

[23] Gatheral, J. *The volatility surface: A practitioner's guide*. John Wiley & Sons, 2006.

[24] Gatheral, J. Consistent modeling of SPX and VIX options. Presented at The 5th World Congress of the Bachelier Finance Society, 2008.

[25] Gatheral, J., Jaisson, T., and Rosenbaum, M. Volatility is rough. *Quantitative Finance 18*, 6 (2014), 933–949.

[26] Gatheral, J., Jusselin, P., and Rosenbaum, M. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *arXiv [q-fin.MF]* (2020).

[27] Guyon, J. The joint S&P 500/VIX smile calibration puzzle solved. *SSRN Electronic Journal* (2019).

[28] Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. Managing smile risk. *Wilmott Magazine 1* (01 2002), 84–108.

[29] Harms, P., and Stefanovits, D. Affine representations of fractional processes with applications in mathematical finance. *Stochastic processes and their applications 129*, 4 (2019), 1185–1228.

[30] Heston, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies 6*, 2 (1993), 327–343.

[31] Hull, J., and White, A. The pricing of options on assets with stochastic volatilities. *The journal of finance 42*, 2 (1987), 281–300.

[32] Itô, K. Stochastic integral. *Proceedings of the Imperial Academy 20*, 8 (1944), 519 – 524.

[33] Itô, K. On a stochastic integral equation. *Proceedings of the Japan Academy 22*, 2 (1946), 32 – 35.

[34] Jacquier, A., Martini, C., and Muguruza, A. On VIX futures in the rough Bergomi model. *Quantitative finance 18*, 1 (2017), 45–61.

[35] Karatzas, I., and Shreve, S. E. *Brownian Motion and Stochastic Calculus*, 2 ed., vol. 113 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1991.

[36] Kolmogorov, A. Wiener spirals and some other interesting curves in Hilbert space. *Doklady AN SSSR 26*, 2 (1940).

[37] Lacombe, C., Muguruza, A., and Stone, H. Asymptotics for volatility derivatives in multi-factor rough volatility models. *Mathematics and financial economics 15*, 3 (2021), 545–577.

[38] Mandelbrot, B. B., and Van Ness, J. W. Fractional brownian motions, fractional noises and applications. *SIAM review 10*, 4 (1968), 422–437.

[39] Merton, R. C. Theory of rational option pricing. *The Bell journal of economics and management science 4*, 1 (1973), 141.

[40] Mishura, Y., and Yurchenko-Tytarenko, A. Fractional Cox–Ingersoll–Ross process with non-zero "mean". *Modern Stochastics: Theory and Applications 5*, 1 (2018), 99–111.

[41] Mishura, Y., and Yurchenko-Tytarenko, A. Fractional Cox–Ingersoll–Ross process with small Hurst indices. *Modern Stochastics: Theory and Applications 6*, 1 (2018), 13–39.

[42] Mishura, Y., and Yurchenko-Tytarenko, A. Approximating expected value of an option with non-Lipschitz payoff in fractional Heston-type model. *International Journal of Theoretical and Applied Finance 23*, 05 (July 2020), 2050031.

[43] Nualart, D. *The Malliavin calculus and related topics*, 2 ed. Springer-Verlag, 2006.

[44] Osborne, M. F. M. Brownian motion in the stock market. *Operations Research 7*, 2 (1959), 145–173.

[45] Rogers, L. Arbitrage with fractional brownian motion. *Mathematical Finance 7* (1997), 95–105.

[46] Rosenbaum, M. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic processes and their applications 118*, 8 (2008), 1434–1462.

[47] Rosenbaum, M., and Zhang, J. Deep calibration of the quadratic rough Heston model. *arXiv [q-fin.CP]* (2021).

[48] Samuelson, P. A. Rational theory of warrant pricing. *Industrial Management Review 6*, 2 (1959), 13–39.

[49] Stein, E. M., and Stein, J. C. Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies 4* (1991), 727–752.

[50] Sugita, H. On a characterization of the Sobolev spaces over an abstract Wiener space. *Journal of Mathematics of Kyoto University 25*, 4 (1985), 717–725.

[51] Young, L. C. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica 67* (1936), 251–282.

[52] Zumbach, G. Volatility conditional on price trends. *Quantitative finance 10*, 4 (2010), 431–442.