FAN_CHUNHEI_01496565

by Chun Fan

Submission date: 06-Sep-2022 12:54PM (UTC+0100)
Submission ID: 185743603
File name: FAN_CHUNHEI_01496565.pdf (2.69M)
Word count: 11019
Character count: 50275

Imperial College London

Imperial College London

Department of Mathematics

Application of Schrödinger's Bridge in Volatility Models

Author: Chun Hei Fan (CID: 01496565)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2021-2022

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

First of all, I would like to express my gratitude to my thesis supervisor Dr. Alexandre Pannier for proposing this challenging yet innovating thesis idea on volatility modelling and rough volatility. He helped me gain understanding of the topic in a limited amount of time and this thesis would not be possible without his continued guidance and support.

I would like to thank the team of the MSc Mathematics and Finance at Imperial College for their quality teaching and career support throughout the year. I would also like to thank my friends and flatmates who helped me survive my time at Imperial.

Last but not least, I owe my deepest gratitude to my parents. Their unconditional love and support is what makes me who I am today.

Abstract

This thesis focuses on a new class of volatility models that are based on Schrödinger's bridge construction [1]. Unlike the calibration of any existing stochastic volatility models, the volatility parameters are not modified. Instead, a drift term is added to the model dynamics which would alter its original probability measure. The new probability measure is defined such that the model is exactly calibrated to market option prices.

Another main objective of this thesis is to incorporate rough volatility into the model. Due to the non-Markovian nature of fractional Brownian motion, we would need to apply theories outside the domain of Itô calculus and provide original proofs to obtain a rough volatility version of the model.

Contents

List of Figures

List of Abbreviations

ATM ..... At the money

BM ..... Brownian motion

fBM ..... Fractional Brownian motion

FSVM ..... Fractional stochastic volatility model

HJB ..... Hamilton-Jacobi-Bellman

ITM ..... In the money

LSVM ..... Local stochastic volatility model

LVM ..... Local volatility model

OTM ..... Out of hthe money

PDE ..... Partial differential equation

PPDE ..... Path-dependent partial differential equation

SDE ..... Stochastic differential equation

SVM ..... Stochastic volatility model

Introduction

In derivatives pricing, the following stochastic equation is widely accepted by practitioners as the standard to model the dynamics of price prossess S:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t \tag{0.0.1}$$

where μ is the drift term and σ the volatility term.

In the Black-Scholes model, the volatility process is modelled as a constant. Along with a constant drift term, process S folllows a geometric Brownian motion which gives one of the very first mathematical equation to compute theoretical value of an option. For many years, the Black-Scholes models have become an industry standard of option pricing due to its closed-form solution and simplicity. However, the assumption of constant volatility has always been under criticism, as the non-constant implied volatility surface (see Figure 1.1) clearly violates the assumption. Therefore, it leads to the development of volatility modelling and has since become one of the most important topics in modern quantitative finance. Volatility is modelled as stochastic processes which follow particular dynamics, such as Heston, CEV, SABR etc. Model parameters are then calibrated to fit the implied volatility surfaces.

Another improvement of volatility modelling is inspired by the belief that volatility has long-memory features. Comte and Renault [2] proposed to model log-volatility using fractional Brownian motion, with Hurst parameter $H \in (\frac{1}{2}, 1]$. However, the idea of long-memory is soon replaced by rough volatility, with $H \in [0, \frac{1}{2})$, as Gatheral, Jaisson, Rosenbaum [3] and Fukasawa [4] demonstrated its capability to explain stylised facts of volatility historical time series. Despite the non-Markovian nature of fractional Brownian motion (fBM), recent developments in numerical implementations have lead to increasing popularity of rough volatility models in the industry.

In this paper, we will introduce a new class of stochastic volatility models constructed through entropy-minimalisation, similar to the construction of a Schrödinger's bridge. The idea is pioneered by Avellaneda [5] and properly formalised for its application in volatility models by Henry-Labodère [1]. The aim is to minimise the relative entropy with respect to a reference probability measure (Wiener measure), such that the calibrated measure follows a distribution implied from market prices. This method can be applied to any chosen SVM and the new model can be exactly calibrated to implied volatility surfaces. Other than that, Guyon adopted this idea in an attempt to solve the joint SPX/VIX smile calibration problem and successfully calibrate a model to both instruments [6, 7], which is not achievable by any stochastic volatility models. The next part of this paper aims to show that this method can be applied to rough volatility models.

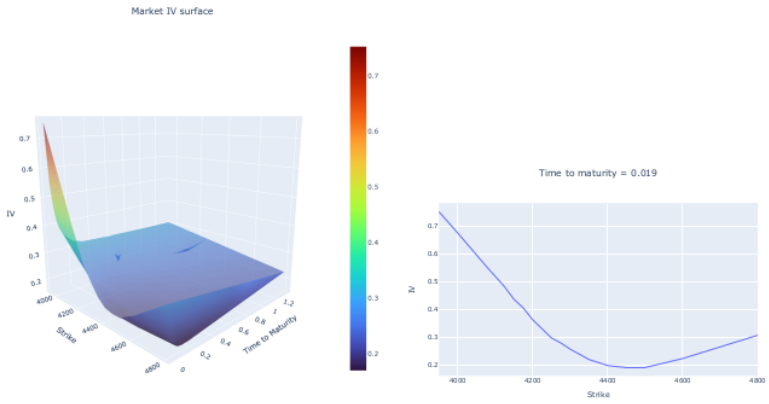
Chapter 1

Motivations

1.1 Volatility smile

Implied volatility is an important metric in option trading widely used by traders. Different to historical volatility, its value is directly inferred from the market price of an option. The observed option price is treated as an input to the Black-Scholes formula whereas the volatility is the only unknown in the equation. A root-finding algorithm, such as Newton-Raphson, is then applied to the Black-Scholes formula which outputs the implied volatility.

The volatility smile refers to the shape of the curve produced by plotting the implied volatility against the strike price of an option. The smile is commonly observed in many options and can be explained by the changing implied volatility as option moves more ITM or OTM. This contradicts the assumption of the Black-Scholes model, as volatility is supposed to be constant across strikes given the same expiration.



(a) (b)

Figure 1.1: (a) The SPX call options implied volatility surface as of August 12, 2022. The implied volatility is computed using the `py_vollib` library in Python. (b) This is an illustration of the volatility smile.

1.1.1 Local volatility

Due to the presence of volatility smile, we realise that log-returns are not normally distributed, hence we need to find a way to incorporate skewness into the model. A natural extension of the Black-Scholes model is to replace the constant volatility term with a deterministic function of time t and asset price S_t, which is deduced directly from the implied volatility surface. The local volatility function is specifically chosen such that implied volatility from the model resembles the implied volatility surface. The idea is known as local volatility model (LVM), which was first introduced by Dupire [8] in 1994. With local volatility, the dynamics is now moderated to:

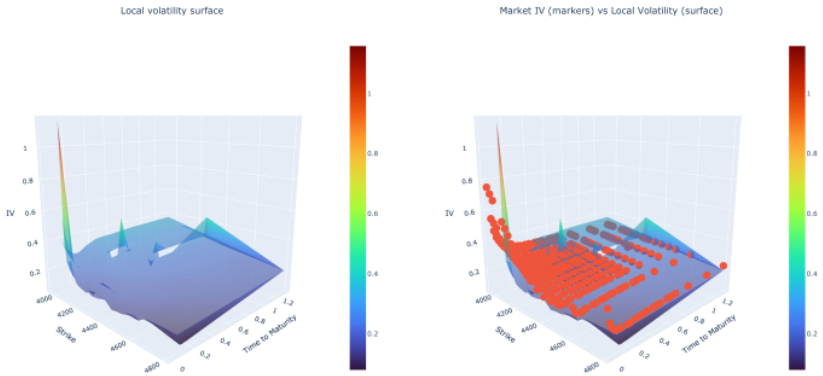
$$dS_t = \mu S_t dt + \sigma_L(t, S_t) dW_t$$

where $\sigma_L(\cdot)$ is the local volatility term.

To calibrate the model to a volatility surface, i.e. $\mathbb{E}[(S_T - K)^+] = C(T, K)$, Dupire came up with a formula for the instantaneous volatility:

$$\sigma_L^2(T, K) = \frac{1}{K^2} \frac{2\frac{\partial C(T,K)}{\partial T}}{\frac{\partial^2 C(T,K)}{\partial K^2}} \tag{1.1.1}$$

where $C(T, K)$ is the market price of a call option with maturity T and strike K.



(a) (b)

Figure 1.2: (a) This is the local volatility surface calibrated to market prices. The derivatives are interpolated using a cubic spline using `scipy.interpolate` (b) The surface is compared to available market prices which shows that the surface resembles the volatility smile in short maturities.

In theory, it does not involve any complex computations and the local volatility surface should match the implied volatility surface of vanillas, especially the skew in short maturities. As seen from Figure 1.2, there are some gaps and spikes which seem to violate the trend of the plot. This is because the volatility surface constructed from Equation (1.1.1) does not guarantee an arbitrage free surface. In this example, we adopted the approach of simply removing values in which the time

or strike derivatives are negative, which explains the gap. However, there are more sophisticated methods, specific to the data, which can eliminate the arbitrage of the surface.

Other than that, without proper volatility dynamics, local volatility underestimates the forward smile and leads to flattening of the forward volatility curve. Therefore, it is not suitable to price complex products that depends on forward volatility such as cliquet options. Moreover, the model is only consistent with present prices, which mean it has to be re-calibrated quite frequently such as hourly.

1.1.2 Stochastic volatility

An alternative to local volatility is to model the volatility as a stochastic process. This is known as stochastic volatility models (SVM), and its dynamic is generally given by:

$$\begin{aligned} dS_t = S_t a_t dW_t, \quad d\langle W, Z\rangle_t = \rho dt \\ da_t = b(a_t)dt + \sigma(a_t)dZ_t \end{aligned} \tag{1.1.2}$$

where $b(\cdot)$ and $\sigma(\cdot)$ are the drift and diffusion coefficients with $t \geq 0$, $a_t \geq 0$ and $\rho \in (-1, 1)$.

By modelling the volatility, we can introduce additional parameters to better capture market characteristics. This provides a much more realistic market dynamics, unlike local volatility which is purely constructed to fit present prices. Therefore, it is much more capable of modelling the term structure of volatility and producing a more realistic forward volatility curve. One of the SVM that we will focus on in this paper is the Heston model:

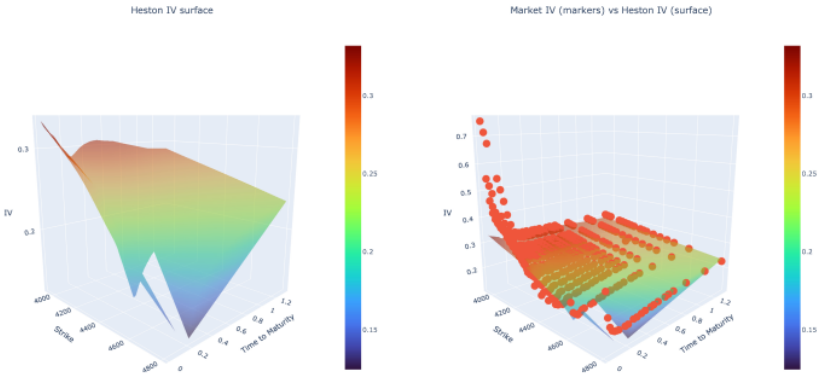
$$\begin{aligned} dS_t = S_t\sqrt{a_t} dW_t, \quad d\langle W, Z\rangle_t = \rho dt \\ da_t = \kappa(\theta - a_t)da_t + \sigma\sqrt{a_t}dZ_t \end{aligned} \tag{1.1.3}$$

where κ is the rate of mean reversion, θ the long term mean, σ the volatility of volatility and v_0 the initial volatility. These are collectively known as the Heston parameters and are restricted to positive real numbers. Unlike most SVMs, it has a semi-closed analytical formula for vanilla options (Appendix A.1).

However, SVM is harder to calibrate and usually involves numerical algorithms which are computationally expensive. In the following example (see Figure 1.3), least square minimisation is used to calibrate the Heston model to market prices to produce a volatility surface.

As seen from the figures, SVM fails to replicate the volatility surface as good as LVM, especially for short-term maturities. It could not generate enough skew to reproduce the smile curve (see Figure 1.4), however the model matches market prices with long maturities quite well. Apart from that, any changes to a single parameter would also require re-calibration of the entire model. The ATM volatility skew suggested by SVM is also inconsistent with empirical data which can be approximated by a power law.

In Figure 1.4, the graphs show the comparison of the calibrated LVM and SVM (Heston) against market implied volatilities at different maturities.



(a) (b)

Figure 1.3: (a) This is the Heston implied volatility surface with parameters $a_0 = 0.083$, $\kappa = 5.00$, $\theta = 0.090$, $\sigma = 0.984$ and $\rho = -1.00$, calibrated through `scipy.optimize`. (b) When compared to market prices, it is accurate for long maturities but it fails to capture the extra steepness in short maturities.

1.1.3 Local stochastic volatility

In an example on Double No Touch option pricing [9, Section 2.2], it is shown that LVM tends to under-price while SVM tends to over-price. The degree of inaccuracy exceeds that of the bid-ask spread which clearly shows that both models are flawed in their own ways. An attempt to capture the advantages of both SVM and LVM gives birth to local stochastic volatility models (LSVM). In simple terms, a 'mixing' parameter $\gamma \in (0, 1)$ is chosen to indicate the weights of local and stochastic volatility in the model to fit certain option prices.

A popular LSVM is introduced by Lipton [10], which follows a Heston-like dynamics:

$$\begin{aligned} dS_t &= S_t \sigma(t, S_t)\sqrt{a_t} dW_t, \quad d\langle W, Z\rangle_t = \rho dt \\ da_t &= \kappa(\theta - a_t)da_t + \gamma\sqrt{a_t} dZ_t \end{aligned}$$

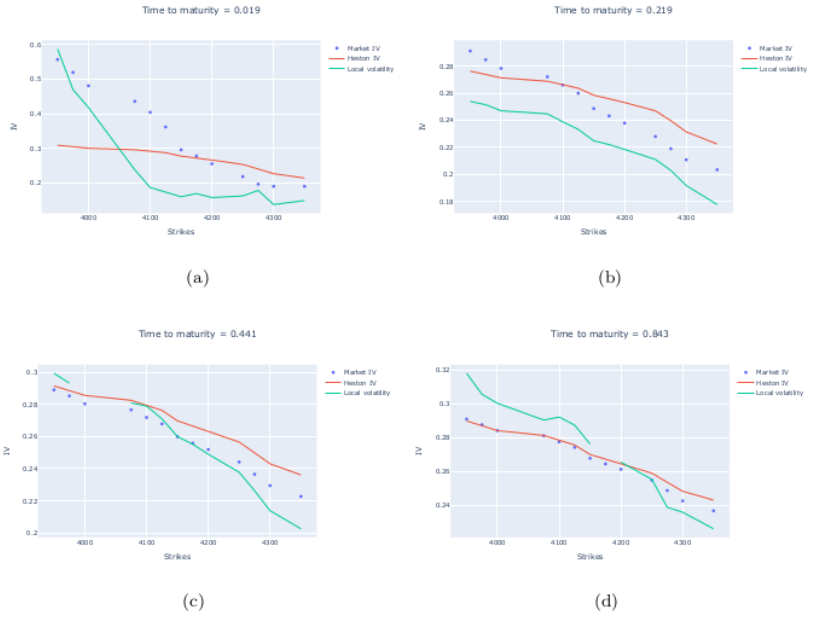
where the parameters are defined exactly as in Equation (1.1.3).

In this model, the local volatility function $\sigma(t, S_t)$ is multiplied with the stochastic volatility term $\sqrt{a_t}$, in which γ is the 'mixing' paramater mentioned earlier, and is a multiplication of the vol-of-vol parameter. The modelling of the vol-of-vol parameter is crucial in LSVM and is a major research topic in itself.

Remark 1.1.1. When $a_\cdot = 1$, LSVM simply becomes a LVM. When $\sigma(\cdot) = 1$, then it becomes a SVM.

The calibration of the local volatility function can be done by setting:

$$\sigma(t, S_t)^2 \mathbb{E}[a_t^2 | S_t] = \sigma_L(t, S_t)^2$$



(a) (b)

(c) (d)

Figure 1.4: These are the comparison of the calibrated smile from both models in different maturities. It clearly shows their respective advantage over their counterpart in short and long maturities.

where $\sigma_L(t, S_t)$ is the Dupire's local volatility defined in Equation (1.1.1).

Similar to local volatility, the surface can be calibrated to all market prices while the 'mixing' parameter allows us to retain the advantage of improved dynamics described by stochastic volatility. The main focus of this paper serves a similar purpose to LSVM, a new class of SVM which is based on martingale optimal transport [1] and allows exact calibration.

1.2 Rough volatility

Fractional stochastic volatility model (FSVM) was first introduced by Renault and Comte [2] by modelling volatility driven by a fBM. fBM is a generalisation of Brownian motion (BM) which has correlated increments, characterised by the Hurst parameter $H \in (0, 1)$ (See Appendix A.2).

A particular case of FSVM with $H < \frac{1}{2}$ is later popularised by Gatheral, Jaisson and Rosenbaum [3], known as rough volatility, which successfully replicate several empirical properties of historical volatility time series. This section presents a brief overview of the results from his paper.

1.2.1 Empirical evidence

To measure the smoothness of historical volatility, which is characterised by Hurst parameter H, we would need proxy spot volatility values[1] since they are not directly observable. Denote the volatility values on a time grid $[0, T]$ with mesh $\Delta = \frac{T}{N}$ such that $\sigma_{i\Delta} = \sigma_0 + i\Delta$ and $\sigma_0 < \sigma_\Delta < \cdots < \sigma_{N\Delta}$. First, define

[1]Daily realised variance estimates taken from the Oxford-Man Institute of Quantitative Finance, from 3 January 2000 to 28 June 2022. `https://realized.oxford-man.ox.ac.uk/data`

$$m(q,\Delta) = \frac{1}{N}\sum_{k=1}^{N}|\log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta})|^q$$

Under stationarity assumption and the law of large numbers, $m(q,\Delta)$ can be interpreted as an empirical estimate of the q^{th} moment of log-volatility increments, $\mathbb{E}[(\log\sigma_\Delta - \log\sigma_0)^q]$. We also assume

$$N^{qS_q}\log(q,\Delta) \to b_q \text{ as } \Delta \to 0$$

for some parameters $s_q > 0$ and $b_q > 0$.

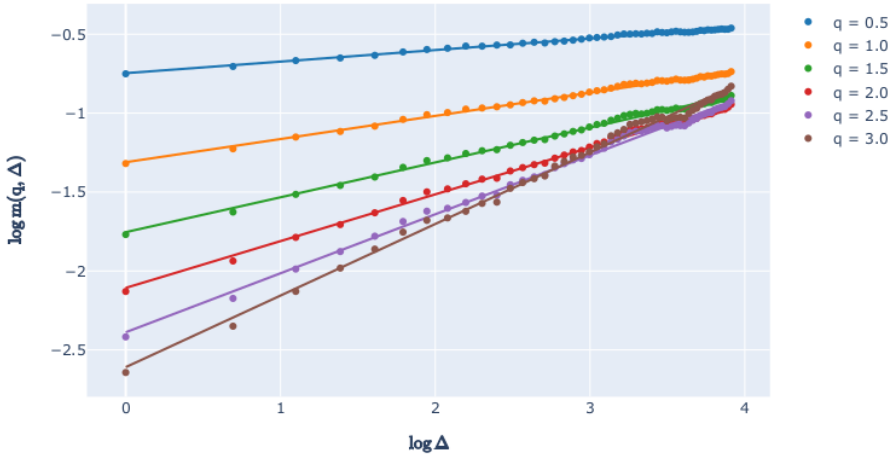


Figure 1.5: $\log m(q,\Delta)$ against $\log\Delta$

The points on Figure 1.5 can be approximated by a linear fit, which suggests that

$$\mathbb{E}[(\log\sigma_\Delta - \log\sigma_0)^q] \sim \delta^{\zeta_q}$$

where ζ_q is the slope of the curve associated with the value q.

Finally, the plot of ζ_q against q (see Figure 1.6) is again approximated by a linear fit and the slope gives us an estimate of the Hurst parameter $H \approx 0.152$.

It is shown by Fukusawa [4] that a model driven by fBM with small H reproduces the ATM volatility skew and can be approximated by the power law $\phi(\tau) \sim \tau^{H-\frac{1}{2}}$ as $\tau \to 0$ (See Figure 1.7).

Moreover, it is a well-established stylised fact that log-volatility increments follow an approximate

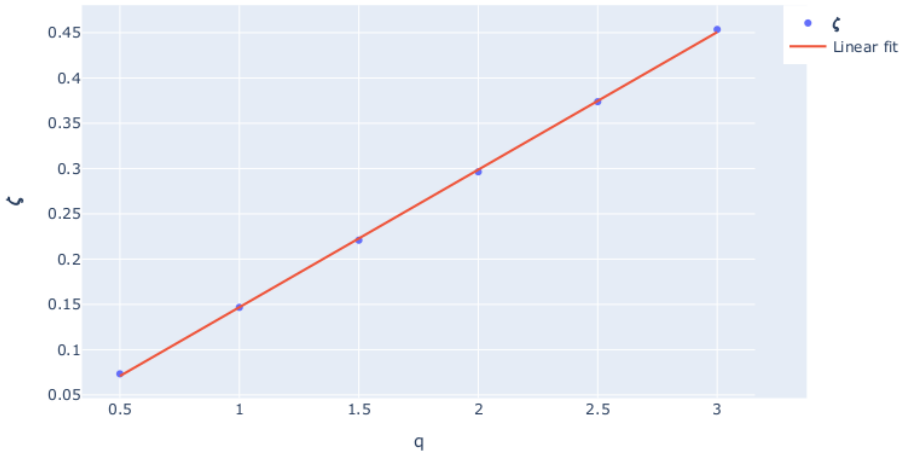


Figure 1.6: Values ζ at different q and its linear fit

Gaussian distribution. Observed data for S&P 500 seems to verify this behaviour and is fitted with a Gaussian distribution for several lags in Figure 1.8.

1.2.2 Simulation of rough volatility models

In rough volatility models, the process a follows the same dynamics as in Equation (1.1.2) but with the BM Z_t replaced by a fBM Z_t^H instead. However, it is shown that the Euler scheme only converges when $H > \frac{1}{2}$ [11], which is not applicable to rough volatility.

As a result, when simulating rough volatility models, the volatility is usually expressed by a stochastic Volterra equation of the form, which is directly derived from the fBM representation for $t \geq 0$ in Proposition A.2.4:

$$a_t = a_0 + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}} b(a_t)dt + \frac{1}{\Gamma(H+\frac{1}{2})}\int_0^t (t-s)^{H-\frac{1}{2}} \sigma(a_t)dW_t \tag{1.2.1}$$

where W_t is a standard BM.

This process can be simulated by the Euler-Maruyama algorithm, however the simulation speed is slow in terms of industrial standard. Jaber and El Euch [12] proposed a fast algorithm which reduces the algorithm's complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ or $\mathcal{O}(N \log^2 N)$ [13]. They approximate the kernel function using a Gauss-Legendre quadrature, $t^{H-\frac{1}{2}} \approx \sum_{l=1}^{N'} \omega_l e^{-x_l t}$.

Dividing $[0, t]$ into equal intervals Δt with $0 = t_0 < t_1 < \cdots < t_n = t$, the fast algorithm for the simulation of Equation (1.2.1) is given by:

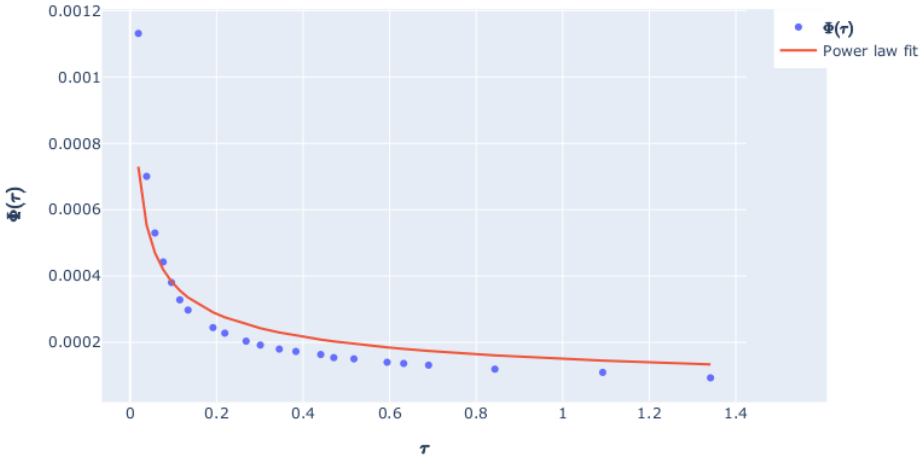


Figure 1.7: ATM volatility skew and its power law fit $\Phi(\tau) \sim \tau^{-0.348}$.

$$a_{t_k} = a_0 + \frac{\Delta t^{H+\frac{1}{2}}}{\Gamma(H+\frac{3}{2})} b(a_{t_{k-1}} + \frac{\Delta t^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \sigma(a_{t_{k-1}})(W_{t_k} - W_{t_{k-1}}) + \frac{1}{\Gamma(H+\frac{1}{2})} \sum_{l=1}^{n} \omega_l e^{-x_l \Delta t} (H_l(t_{k-1}) + J_l(t_{k-1}))$$

for $k = 1, \cdots, n$

where

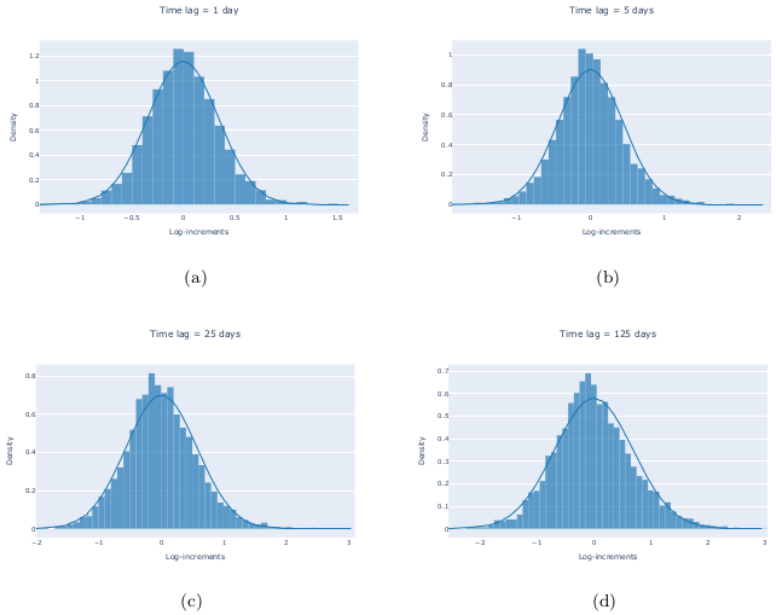
$$H_l(t_{k-1}) = \frac{b(a_{t_{k-2}})}{x_l}(1 - e^{-x_l \Delta t}) + e^{-x_l \Delta t} H_l(t_{k-2})$$
$$J_l(t_{k-1}) = e^{-x_l \Delta t} \sigma(a_{t_{k-2}}) \sqrt{\Delta t}(W_{t_{k-1}} - W_{t_{k-2}}) + e^{-x_l \Delta t} J_l(t_{k-2})$$

for $k = 2, \cdots, n$

with $H_l(t_0) = 0$ and $J_l(t_0) = 0$.

1.3 Joint calibration problem

Volatility index was first invented as a benchmark to quantify market's expectations of volatility in the near future. Due to its high negative correlation to the market, its deriviatives are widely used by traders to hedge the short term volatility of their portfolios or even as pure speculation.



(a) (b) (c) (d)

Figure 1.8: Histogram of log-increments for various lags (a) $\Delta = 1$, (b) $\Delta = 5$, (c) $\Delta = 25$, (d) $\Delta = 125$ and their normal fit (blue line).

1.3.1 VIX index

One of the most popular volatility index is the CBOE volatility index (VIX), which is designed to measure the expected 30-day volatility of SPX options. Its corresponding derivitive products, such as futures and options, also commenced trading in 2004 and 2006 respectively. The payoff of VIX futures at expiry T_1 is defined as:

$$VIX_{T_1}^2 = -\frac{2}{T_2 - T_1}\mathbb{E}_{T_1}^{\mathbb{P}}\left[\log(\frac{S_{T_2}}{S_{T_1}})\right]$$

where $T_2 = T_1 + 30$ days and S_t the price of S&P 500 index at time t.

The choice of measure $\mathbb{P}$ may seem arbitrary, but the price is actually model-independent since it can be replicated by options with expiry T_2. The corresponding call option of VIX with expiry T_1 and strike K is defined like any other ordinary options, $(VIX_{T_1} - K)^+$.

1.3.2 SPX/VIX joint calibration

Due to its popularity, the futures and options market of VIX index has reached a level of liquidity that existing models for its underlying options would need to be calibrated to VIX options as well. However, it presents a huge challenge for researchers and practitioners to build a model which jointly calibrates SPX and VIX options, especially for short maturities. This is known as the joint calibration problem.

The large negative skew of short term ATM SPX options is inconsistent with the relatively low

implied volatility of ATM VIX options. In most cases, to account for lower implied volatilities for VIX, one lowers the vol-of-vol which in turn diminishes the skew for SPX [6]. This created an issue for many existing models which fail to fit both prices at the same time. In recent years, some models have attempted to incorporate jumps to offer extra degrees of freedom which decouples the ATM skew of SPX and VIX [14]. Gatheral et al. [15] also proposed the quadratic rough Heston model, which is consistent with the strong Zumbach effect, i.e. the conditional dynamics of volatility depends on both historical price and volatility path. However, these models only resulted in an approximate fit which still have room for improvements.

Chapter 2

Schrödinger's Bridge on Stochastic Volatility

In this chapter, we present a new class of stochastic volatility models proposed by Henry-Labordère [1], in a more detailed manner. Similar to local volatility, it can be perfectly calibrated to a volatility surface and is achieved by shifting the Brownian motion in a certain direction. The drift term is determined by a concave optimisation problem which closely resembles the Schrödinger's bridge problem.

2.1 Schrödinger's bridge problem

Before moving on to volatility models, it would be appropriate to briefly outline the original Schrödinger's bridge problem. In a thought experiment introduced by Erwin Schrödinger in 1932 [16], he considered a large number of Brownian particles with initial empirical distribution μ_0 which migrates to a different distribution μ_1 at a later time t. He questioned how the particles are most likely to be transported to this new distribution. The aim is to find a new equivalent probability measure $\mathbb{P}$ that minimises the relative entropy, also known as Kullback-Leibler divergence:

$$H(\mathbb{P}||\mathbb{P}^0) = \mathbb{E}^{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{P}^0}\right] \tag{2.1.1}$$

where $\mathbb{P}^0$ is the Wiener measure, $\frac{d\mathbb{P}}{d\mathbb{P}^0}$ the Radon-Nikodym derivative and a new probability measure $\mathbb{P}$ with marginal μ_1.

The use of Schrödinger's bridge is not only restricted to its original application. In fact, many problems involving the evolution of distributions can be formulated as a Schrödinger's bridge problem, such as deep generative learning [17] and evolutionary biology [18].

2.2 Brownian motion

Consider a stochastic process X under $\mathbb{P}^0$

$$dX_t = dW_t^0$$

where W_t^0 is a Brownian motion with respect to measure $\mathbb{P}^0$.

We aim to find a drift term λ_t such that the resulting process matches a marginal distribution μ at time T under measure $\mathbb{P}$ by construction of a Schrödinger's Bridge. Let us define the entropy minimisation problem

$$\inf_{\mathbb{P}\in\mathcal{M}(\mu)} H(\mathbb{P}||\mathbb{P}^0) \tag{2.2.1}$$

where $\mathcal{M}(\mu) \coloneqq \{\mathbb{P} \sim \mathbb{P}^0 : X_T \overset{\mathbb{P}}{\sim} \mu\}$ and $H(\cdot|\mathbb{P}^0)$ is defined in Equation (2.1.1).

By Girsanov Theorem, for all $\mathcal{M}(\mu) \ni \mathbb{P} \sim \mathbb{P}^0$, the new dynamics under $\mathbb{P}$ are defined as

$$dX_t = \lambda_t dt + dW_t$$

where W_t is a Brownian motion with respect to the measure $\mathbb{P}$.

The Radon-Nikodym derivative is given by

$$\frac{d\mathbb{P}}{d\mathbb{P}^0} = \exp\left(\int_0^T \lambda_s \, dW_s + \frac{1}{2}\int_0^T \lambda_s^2 \, ds\right)$$

Since the expectation of a martingale is zero, we have

$$\inf_{\mathbb{P}\in\mathcal{M}(\mu)} H(\mathbb{P}||\mathbb{P}^0) = \inf_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{E}^{\mathbb{P}}\left[\frac{1}{2}\int_0^T \lambda_s^2 \, ds\right] \tag{2.2.2}$$

This constrained optimisation problem can be simplified to a unconstrained optimisation problem by convex duality [19]. We can consider a Lagrangian multiplier, a potential f, subjected to the constraint $\mathbb{P} \in \mathcal{M}(\mu) \iff \mathbb{E}^{\mathbb{P}}[f(X_T)] - \mathbb{E}^{\mu}[f(X_T)] = 0$. This gives us the 'dual' problem:

$$\begin{aligned}
&\inf_{\lambda\in\mathcal{A}_{0,T}} \left\{ \sup_{f\in L^1(\mu)} \mathbb{E}^{\mathbb{P}}[\frac{1}{2}\int_0^T \lambda_s^2 \, ds + f(X_T)] - \mathbb{E}^{\mu}[f(X_T)] \right\} \\
&= \sup_{f\in L^1(\mu)} \left\{ \inf_{\lambda\in\mathcal{A}_{0,T}} \mathbb{E}^{\mathbb{P}}[\frac{1}{2}\int_0^T \lambda_s^2 \, ds + f(X_T)] - \mathbb{E}^{\mu}[f(X_T)] \right\} \\
&= \sup_{f\in L^1(\mu)} \{u(0, X_0) - \mathbb{E}^{\mu}[f(X_T)]\}
\end{aligned} \tag{2.2.3}$$

where $u(t,x) \coloneqq \inf_{\lambda\in\mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{P}}[\frac{1}{2}\int_t^T \lambda_s^2 \, ds + f(X_T)|X_t = x]$ and $\mathcal{A}_{t,T}$ is the space of control λ in the interval $[t,T]$ such that the controlled diffusion admits a unique square integrable strong solution.

Consider $u(t,x)$ as the value function of a stochastic control problem. As a direct result of the Feynman-Kac theorem, we can deduce that $u(t,x)$ is the solution to the Hamilton-Jacobi-Bellman (HJB) PDE:

$$\partial_t u + \frac{1}{2}\partial_x^2 u + \inf_{\lambda \in \mathcal{A}_{t,T}} \{\frac{1}{2}\lambda_t^2 + \lambda_t \partial_x u\} = 0$$
$$u(T,x) = f(x)$$

Differentiating with respect to λ_t shows the infimum is attained at $\lambda_t^* = -\partial_x u(t,x)$. Re-substituting λ_t^* gives a Burger's like PDE:

$$\partial_t u + \frac{1}{2}\partial_x^2 u - \frac{1}{2}(\partial_x u)^2 = 0$$

By applying the transformation $u = -\log U$, the PDE now then takes the form of the heat equation:

$$\partial_t U + \frac{1}{2}\partial_x^2 U = 0$$
$$U(T,x) = e^{-f(x)}$$

which has a unique solution $U(t,x) = \mathbb{E}[e^{-f(W_T^0)}|W_t^0 = x]$ by the Feynman-Kac theorem. Therefore, the dual problem (Equation (2.2.3)) can be written as

$$\sup_{f \in L^1(\mu)} \{-\log \mathbb{E}[e^{-f(W_T^0)}|W_0^0 = X_0] - E^\mu[f]\}$$

We let f^* be the unique solution of the above equation, differentiating with respect to potential f, we obtain the following equation:

$$e^{-f^*(x)}dx = \mu(dx)e^{\frac{1}{2}\frac{(x-X_0)^2}{T}} \tag{2.2.4}$$

Proposition 2.2.1. *[1, Proposition 2.2] The dynamics of Brownian motion X under the measure $\mathbb{P}$ can be written as:*

$$dX_t = \partial_x \log \mathbb{E}[e^{-f^*(W_T^0)}|W_t^0 = x]dt + dW_t \tag{2.2.5}$$

such that $X_T \overset{\mathbb{P}}{\sim} \mu$ and f^* as defined in Equation (2.2.4).

The problem can be easily extended to fit multiple marginals, it will be explained in Section 2.3 in the context of volatility models for betting understanding.

2.3 Stochastic volatility models

To apply the idea of Schrödinger's bridge in volatility modelling, let us first consider a one-dimensional SVM of the form

$$dS_t = S_t a_t dW_t^0, \quad d\langle W^0, Z^0\rangle_t = \rho dt$$
$$da_t = b(a_t)dt + \sigma(a_t)dZ_t^0$$

To calibrate the model to market instruments, we derive the marginals $(\mu_i)_{1\leq i\leq n}$ from the market values of Call options with maturities $T_1 < T_2 < \cdots < T_n$ and strike K, where $\mu_i(K) = \partial_{KK}C(T_i, K)$. This is the risk neutral distribution of S_i such that $\mathbb{E}^{\mu_i}[S_i] = S_0$.

Similar to Section 2.2, we introduce drift terms $\lambda_t^{(1)}$, $\lambda_t^{(2)}$ to the Brownian motions respectively under new measure $\mathbb{P}$. We then decompose the process into intervals $[T_{i-1}, T_i]$, each associated with its own optimisation problem and a potential f_i. Due to the Markov property of X, the calibrated SVM is the concatenation dynamics over the intervals $[T_{i-1}, T_i]$.

Theorem 2.3.1. *[1, Proposition 3.1] For each interval, we define the optimisation problem as*

$$P_i = \sup_{f_i\in L^1(\mu_i), \Delta_{s\in(T_{i-1},T_i)}} \{-\log \mathbb{E}^{\mathbb{P}_0}[e^{-f_i(S_{T_i})-\int_{T_{i-1}}^{T_i}\Delta_s dS_s}|S_{T_{i-1}}, a_{T_{i-1}}] - \mathbb{E}^{\mu_i}[f_i]\}$$

Therefore, for all $t \in [T_{i-1}, T_i]$, the dynamics is given by

$$dS_t = S_t a_t dW_t, \quad d\langle W, Z\rangle_t = \rho dt$$
$$da_t = \big(b(a_t) + (1-\rho^2)\sigma(a_t)^2\partial_a \log \mathbb{E}^{\mathbb{P}_0}[e^{-f_i^*(S_{T_i})-\int_{T_{i-1}}^t \Delta_s^* dS_s}|S_t, a_t]\big)dt + \sigma(a_t)dZ_t$$

such that $S_{T_i} \overset{\mathbb{P}}{\sim} \mu_i$, for $1 \leq i \leq n$.

Proof. See Appendix A.3 □

Remark 2.3.2. The integral $\int_{T_{i-1}}^{T_i}\Delta_s dS_s$ is introduced to retain the martingality of process S i.e.

$$\mathbb{E}^{\mathbb{P}}[\int_{T_{i-1}}^{T_i}\Delta_s dS_s] = 0 \iff \lambda^{(1)}_{t\in(T_{i-1},T_i)} = 0 \iff S_{t\in(T_{i-1},T_i)} \text{ is a } \mathbb{P}-\text{martingale}$$

2.4 Joint calibration problem

Inspired by recent works in related research [20, 21], Guyon successfully constructed a model which can be jointly calibrated to SPX and VIX smile by transforming it into a martingale optimal transport problem [6, 7]. The model is perfectly calibrated to SPX smiles at maturities T_1, $T_2 = T_1 + 30$ days, and the VIX future and smile at maturity T_1.

Denote SPX prices with maturity T_i as S_i and VIX future with maturity T_1 as V. Let us derive the marginals (μ_1, V, μ_2) from market prices similarly by taking the second derivative of strikes.

The goal is to perform the same entropy minimisation (Equation (2.2.1)) but on $\mathcal{P}(\mu_1, V, \mu_2)$, a set of probability measures $\mathbb{P}$ such that:

$$S_{T_i} \overset{\mathbb{P}}{\sim} \mu_i, \quad V \overset{\mathbb{P}}{\sim} \mu_V, \quad \mathbb{E}^{\mathbb{P}}[S_2|S_1, V] = S_1, \quad \mathbb{E}^{\mathbb{P}}[L(\frac{S2}{S1})|S_1, V] = V^2$$

where $L(x) = -\frac{2}{T_2 - T_1}\log(x)$, i.e. the formula for VIX futures as defined in Section 1.3.1.

To dualise this optimisation problem, we would need a set of Lagrange multipliers $(f_1, f_V, f_2, \Delta_S, \Delta_L)$ to satisfy all the constraints, i.e.

$$\begin{aligned}
\mathbb{P} \in \mathcal{P}(\mu_1, V, \mu_2) \iff & \mathbb{E}^{\mathbb{P}}[f_i(S_i)] - \mathbb{E}^{\mu_i}[f_i(S_i)] = 0, \\
& \mathbb{E}^{\mathbb{P}}[f_V(V)] - \mathbb{E}^{\mu}[f_V(V)] = 0, \\
& \mathbb{E}^{\mathbb{P}}[\Delta_S(S_1, V)(S_2 - S_1)] = 0, \\
& \mathbb{E}^{\mathbb{P}}\left[\Delta_L(S_1, V)(L\frac{S_2}{S_1} - V^2)\right] = 0
\end{aligned}$$

By similar arguments in Section 2.2, there is a unique solution $\inf_{\mathbb{P}\in\mathcal{P}(\mu_1,V,\mu_2)}\{H(\mathbb{P}|\mathbb{P}^0) + \mathbb{E}^{\mathbb{P}}[\cdot]\} = -\log \mathbb{E}^{\mathbb{P}^0}[e^{-\cdot}]$ and the problem can be dualised into:

$$\begin{aligned}
\sup_{(f_1, f_V, f_2, \Delta_S, \Delta_L)} \{ & -\log \mathbb{E}^{\mathbb{P}^0}[e^{-f_1(S_1) - f_V(V) - f_2(S_2) - \Delta_S(S_1,V)(S_2 - S_1) - \Delta_L(S_1,V)(L\frac{S_2}{S_1} - V^2)}] \\
& - \mathbb{E}^{\mu_1}[f_1(S_1)] - \mathbb{E}^{V}[f_V(V)] - \mathbb{E}^{\mu_2}[f_2(S_2)]\}
\end{aligned}$$

2.5 Implementation

In this section, we will outline the steps required for the implementation of this model, suggested by Henry-Labordère [1]. We then provide details on the techniques, adopted from various articles on numerical algorithms.

2.5.1 Procedures

Single marginal

In practice, market prices are only available for a limited number of strikes, $K_1, \cdots, K_N$. Therefore, f_1 can be defined as a function of the strikes:

$$f_1(s) = \sum_{\alpha=1}^{N} \omega_\alpha (s - K_\alpha)^+$$

where ω_α are the weights for each strike K_α.

Therefore, the optimisation over function f_1 has now been reduced to the optimisation over $\omega \in \mathbb{R}^N$. The optimsation problem (Equation (2.2.3)) can be performed by gradient descent and the gradient

with respect to ω_α is given by:

$$W_\alpha(0, S_0, a_0) - \mathbb{E}^\mu[(S_{T_1} - K_\alpha)^+]$$

where W_α is the solution of the following PDE:

$$\begin{aligned} \partial_t u + \mathcal{L}^0 u - (1-\rho^2)\sigma(a)^2 \partial_a u \partial_a W_\alpha = 0 \\ W_\alpha(T_1, s, a) = (s - K_\alpha)^+ \end{aligned} \tag{2.5.1}$$

Algorithm 1. [1, Section 3.3] Gradient descent

1. Initialise the weights $\omega^{(\mathbf{0})} \in \mathbb{R}^N$, step size $\eta \in \mathbb{R}$ and set $\omega_{old} = \omega^{(\mathbf{0})}$.
2. Solve PDE (A.3.1) and (2.5.1) for each strike K_α. Then store the solutions as $u(t, s, a)$ and $W_\alpha(t, s, a)$ respectively.
3. Compute gradient ∇_ω from $W_\alpha(t, s, a)$, with
$$(\nabla_\omega)_\alpha = W_\alpha(0, S_0, a_0) - \mathbb{E}^{\mu_1}[(S_{T_1} - K_\alpha)^+]$$
4. Update $\omega_{new} = \omega_{old} - \eta \nabla_\omega$ and set weights $\omega_{old} = \omega_{new}$.
5. Repeat steps (2) - (5) until ω converges.

After performing the gradient descent algorithm, set $f_1^* = f_1^{(n)}$ and compute Δ_t^* from $u^{(n)}(t, s, a)$. By Girsanov Theorem, the price of an option with payoff Φ_t at $t < T_1$ is then given by:

$$\mathbb{E}^{\mathbb{P}}[\Phi_t] = \frac{\mathbb{E}^{\mathbb{P}^0}[\Phi_t e^{-f_1^*(S_{T_1}) - \int_0^{T_1} \Delta_s^* ds}]}{\mathbb{E}^{\mathbb{P}^0}[e^{-f_1^*(S_{T_1}) - \int_0^{T_1} \Delta_s^* ds}]}$$

where the expectation on the right hand side is computed by Monte Carlo under the original measure $\mathbb{E}^{\mathbb{E}^0}$.

Multiple marginals

To calibrate the model to multiple marginals, we would need to solve the optimisation problems P_i in Theorem 2.3.1 for each sub-interval $[T_{i-1}, T_i]$.

Algorithm 2. [1, Section 3.5] Multiple marginals calibration

1. Simulate M paths of $(S_{T_1}^{(j)}, a_{T_1}^{(j)})_{1 \le j \le M}$ using Monte Carlo under $\mathbb{P}^0$.
2. Compute the Radon-Nikodym derivative for each path:
$$G_{T_1}^{(j)} = \frac{e^{-f_1^*(S_{T_1}^{(j)}) - \int_0^{T_1} \Delta_s^{*,(j)} ds}}{e^{-u(T_0, S_{T_0}^{(j)}, a_{T_0}^{(j)})}}$$
3. Solve PDE (A.3.1) defined in $[T_2, T_1]$ and store the solution as $u(t, s, a)$.

4. Solve the following optimisation using gradient descent:

$$\sup_{f_2 \in L^1(\mu_2)} \{\frac{1}{M} \sum_{j=1}^{M} G_{T_1}^{(j)} u(T_1, S_{T_1}^{(j)}, a_{T_1}^{(j)})\}$$

5. Simulate M paths of $(S_{T_2}^{(j)}, a_{T_2}^{(j)})_{1 \leq j \leq M}$ using Monte Carlo under $\mathbb{P}^0$.

6. Update the Radon-Nikodym derivative for each path:

$$G_{T_2}^{(j)} = G_{T_1}^{(j)} \frac{e^{-f_2^*(S_{T_2}^{(j)}) - \int_{T_1}^{T_2} \Delta_s^{*,(j)} ds}}{e^{-u(T_1, S_{T_1}^{(j)}, a_{T_1}^{(j)})}}$$

7. Repeat steps (3) - (6) until it reaches the last sub-interval $[T_{n-1}, T_n]$, to obtain the value of G_{T_n}.

The price of an option with payoff $\Phi(S_{T_1}, \cdots, S_{T_n})$ is then given by:

$$\mathbb{E}^{\mathbb{P}}[\Phi] = \frac{1}{M} \sum_{j=1}^{M} G_{T_n}^{(j)} \Phi(S_{T_1}^{(j)}, \cdots, S_{T_n}^{(j)})$$

2.5.2 Numerical PDE

Solving the PDE is an important step in the implementation of this model. Presented below are the two main approaches to obtain a numerical solution of a PDE and Equation (A.3.1) will be used as an example for both approaches.

Finite differences

In this approach [22], the continuous solution of a PDE is being approximated by a discretised function. The derivatives are approximated by some finite difference operators.

Definition 2.5.1. The finite difference operators relevant to PDE (A.3.1) are defined as follows:

$\mathcal{O}(\Delta t)$ difference operator in time:

$$D_t^{-} u(t) = \pm \frac{u(t \pm \Delta t) - u(t)}{\Delta t}$$

$\mathcal{O}(\Delta x^2)$ difference operator in x:

$$D_x^{\pm} u(t) = \pm \frac{-3u(x) + 4u(x \pm \Delta x) - u(x \pm 2\Delta x)}{2\Delta x}$$

$$D_{xx} u(t) = \frac{u(x - \Delta x) - 2u(x) + u(x + \Delta x)}{\Delta x^2}$$

$\mathcal{O}(\Delta x \Delta y)$ difference operator in x, y:

$$\begin{aligned} D_{xy}^{+} u(x,y) = \frac{1}{2\Delta x^2} (&u(x - \Delta x, y - \Delta y) - u(x - \Delta x, y) - u(x + \Delta x, y) + 2u(x,y) \\ &- u(x, y - \Delta y) - u(x, y + \Delta y) + u(x + \Delta x, y + \Delta y)) \end{aligned}$$

$$D^{-}_{xy}u(x,y) = \frac{1}{2\Delta x^2}(-u(x-\Delta x, y+\Delta y) + u(x-\Delta x, y) + u(x+\Delta x, y) - 2u(x,y) \\ +u(x, y-\Delta y) + u(x, y+\Delta y) - u(x+\Delta x, y-\Delta y))$$

Algorithm 3. Finite differences

1. Discretise a grid $[0, T_1] \times [S_{min}, S_{max}] \times [a_{min}, a_{max}]$ with intervals $\Delta t, \Delta S, \Delta a$.
2. Denote $t_i = i\Delta t, S_j = S_m in + j\Delta S, a_k = a_m in + k\Delta a$.
3. Evaluate $u(T_1, s, a) = f_1(s)$ for all s in grid $[S_{min}, S_{max}]$.
4. Iterate $u(t_i, S_j, a_k)$ backwards in time using the difference operators:

$$u(t_{i-1}, S_j, a_k) = u(t_i, S_j, a_k) + \Delta t\Big(\frac{1}{2}S_i^2 a_i^2 D_{ss}u + \frac{1}{2}\sigma(a_k)^2 D_{aa}u + \rho\sigma(a_k)a_k s_j D_{sa}u \\ + b(a_k)D_a u - \frac{1}{2}(1-\rho^2)(\sigma(a_k)D_a u)^2\Big)$$

Remark 2.5.2. When function u is at the boundaries of the grid, for example $u(t_i, S_{min}, a_{max})$, the respective backward or forward difference operators on each spatial direction should be chosen carefully.

Backward SDE

Suppose the PDE (A.3.1) admits a unique solution $Y_t = u(t, s, a)$. By Itô's formula, Y_t follows the dynamics [23]:

$$\begin{aligned} dY_t &= (\partial_t u + \mathcal{L}^0 u)dt + (\sigma_t' \nabla u) \cdot d\mathbf{W}_t \\ &= (\frac{1}{2}(1-\rho^2)(\sigma(a)\partial_a u)^2)dt + (\sigma_t' \nabla u) \cdot d\mathbf{W}_t \end{aligned}$$

where $\sigma_t = \begin{pmatrix} S_t a_t & 0 \\ \rho\sigma(a_t) & \sqrt{1-\rho^2}\sigma(a_t) \end{pmatrix}$, $\nabla u = \begin{pmatrix} \partial_s u \\ \partial_a u \end{pmatrix}$ and $\mathbf{W}_t = \begin{pmatrix} W_t \\ W_t^{\perp} \end{pmatrix}$.

Define $X_t = \begin{pmatrix} S_t \\ a_t \end{pmatrix}$, $Z_t = \sigma_t' \nabla u$ and $f(t, X_t, Y_t, Z_t) = -\frac{1}{2}(1-\rho^2)(\sigma(a)\partial_a u)^2$, solving the PDE is equivalent to solving the following Backward SDE:

$$dY_t = -f(t, X_t, Y_t, Z_t)dt + Z_t \cdot d\mathbf{W}_t$$

Since Y_T is known, this SDE can then be solved backwards [24] using an Euler scheme to acquire the solution pair (Y_t, Z_t) which solves the PDE.

Algorithm 4. Backward SDE

1. Divide $[0, T_1]$ into $0 = t_0, t_1, \cdots, t_n = T_1$ with interval Δt.

2. Simulate path $X_{T_1} = (S_{T_1}, a_{T_1})$ under $\mathbb{P}^0$ and set $\Delta \mathbf{W}_{t_i} = \mathbf{W}_{t_i} - \mathbf{W}_{t_{i-1}}$.

3. Evaluate the function at terminal time T_1, i.e. $Y_{t_n} = f_1(S_{T_1})$.

4. Iterate Y_t backwards in time by the following equations:

$$Z_{t_{i-1}} = \frac{1}{\Delta t}\mathbb{E}[Y_{t_i}\Delta \mathbf{W}_{t_i}|\mathcal{F}_{t_{i-1}}]$$
$$Y_{t_{i-1}} = \mathbb{E}[Y_{t_i} + f(t_{i-1}, X_{t_{i-1}}, Y_{t_i}, Z_{t_{i-1}})|\mathcal{F}_{t_{i-1}}]$$

The estimation of conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_{t_{i-1}}]$ is first investigated in the context of American option pricing [25]. Nowadays, Malliavin's calculus remains a popular option to evaluate the expectation. However, to avoid complication, we will present both parametric and non-parametric method to estimate the expectation.

In the non parametric approach [26, Section 3], we replace $Y_{t_i} = v(\mathbf{W}_{t_i})$. Then by conditional expectation, we have:

$$\mathbb{E}[v(\mathbf{W}_{t_i})|\mathbf{W}_{t_{i-1}} = \omega] = \frac{\mathbb{E}[\delta_\omega(\mathbf{W}_{t_{i-1}})v(\mathbf{W}_{\mathbf{t_i}})]}{\mathbb{E}[\delta_\omega(\mathbf{W}_{t_{i-1}})]}$$

where δ_ω is the Dirac measure at ω.

In the one-dimensional case, this can be computed by Monte-Carlo simulations by the following formula using integration by part:

$$\mathbb{E}[v(W_i)|W_{t_{i-1}} = w] = \frac{\frac{1}{N}\sum_{j=1}^N \mathbb{1}_{W^{(j)}_{t_{i-1}} \geq w} v(W^{(j)}_{t_i})(\frac{W^{(j)}_{t_{i-1}}}{t_{i-1}} - \frac{W^{(j)}_{t_i} - W^{(j)}_{t_{i-1}}}{t_i - t_{i-1}})}{\frac{1}{N}\sum_{j=1}^N \mathbb{1}_{W^{(j)}_{t_{i-1}} \geq w} \frac{W^{(j)}_{t_{i-1}}}{t_{i-1}}}$$

However, in the multi-dimensional case, the Dirac measure is ambiguous and there doesn't seem to be a direct way to evaluate this integral. Therefore, we introduce the next method.

In the parametric regression approach [26, Section 4], we first simulate N paths of $(X_t^{(j)})_{1\leq j\leq N}$. Then estimate $\mathbb{E}[Y_{t_i}|\mathcal{F}_{t_{i-1}}]$ by a set of deterministic basis functions $\psi_{t_{i-1},m}(X_{t_{i-1}})$:

$$\mathbb{E}[Y_{t_i}|\mathcal{F}_{t_{i-1}}] = \sum_{m=1}^M \alpha_{t_{i-1},m}\psi^{(j)}_{t_{i-1},m}(X_{t_{i-1}})$$

These functions are chosen such that the first element $\psi_{t_{i-1},m}$ should mimic the real conditional expectation. Then, one should complete the basis to obtain the other basis functions such that it is orthogonal to the distribution of $X_{t_{i-1}}$. The coefficients $\alpha_{t_{i-1},m}$ are then computed using least square minimisation:

$$\sum_{j=1}^N |Y^{(j)}_{t_i} - \sum_{m=1}^M \alpha_{t_{i-1},m}\psi^{(j)}_{t_{i-1},m}(X_{t_{i-1}})|^2$$

2.5.3 Limitations

In an attempt to implement the model, there are a few difficulties that I have encountered:

- As the calibration is based on an entropy minimalisation on the reference measure, the base model plays a huge role if it can be exactly calibrated since the volatility parameters are unchanged. For example, in the Heston model, it would be about correctly modelling the vol-of-vol parameter. Therefore, only models with a certain volatility term would allow exact calibration. Guyon also mentioned in an Q&A [27] that choosing the right model might introduce too many constraints and is something that hasn't been investigated thoroughly, hence limiting its practicality.

- When implementing the finite difference method, I realise the solution tends to blow up at some point of the process. First, I suspect that the stability issue may be attributed to the violation of the Courant–Friedrichs–Lewy condition. However, even after altering the time step, the same issue persists. Therefore, the problem itself may be ill-conditioned, which means the solution is too sensitive to a small pertubation of data. Therefore, it may not be suitable for this numerical method.

- In a different attempt to numerically solve the PDE by a Backward SDE, I try the parametric regression. Typically, if Y_T is a polynomial, then one can start with the original function, followed by completing the basis. However, the function is a summation of $(x - K)^+$ for different value of strikes, and it is unclear on how the basis functions should be chosen.

Chapter 3

Schrödinger's Bridge on Rough Volatility

In this section, we aim to extend on the Schrödinger's bridge method in [1] to rough volatility models with original proofs. Since fractional Brownian motion is neither Markovian or a semi-martingale, we would require an entire new framework in order to work with these non-Markovian processes [28].

3.1 Non-Markovian framework

Let us consider a Markov diffusion process X with natural filtration $\mathcal{F}_\cdot^X$ and a continuous function g. Let us formulate the following stochastic control problem, which is a conditional expectation of the form

$$Y_t = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}[g(X_T) + \int_t^T f(s, X_s, \lambda_s) ds | \mathcal{F}_t^X]] \tag{3.1.1}$$

Due to its martingale properties, Y_t is a deterministic function of X_t, i.e. $Y = u(t, X_t)$. As shown in Section 2.2, u solves the following HJB PDE:

$$\begin{aligned} \partial_t u(t,x) + \inf_{\lambda \in \mathcal{A}_{t,T}} \{ b(t,x,\lambda_t) \partial_x u(t,x) + \frac{1}{2} \sigma(t,x,\lambda)^2 \partial_{xx} u(t,x) + f(t,x,\lambda_t) \} = 0 \\ u(T,x) = g(x) \end{aligned} \tag{3.1.2}$$

where $b(\cdot)$ and $\sigma(\cdot)$ are the drift and diffusion coefficients of process X respectively which depends on control λ.

However, when the process X is modelled by a fBM, the loss of semi-martingality implies classical Itô's calculus can no longer be used. The idea of characterising the backward problem with a PPDE is not as straightforward as the example above, and we would require new techniques to extend this idea beyond semi-martingales.

In the context of rough volatility models, let us consider the class of stochastic Volterra equations (SVE) of the form

$$X_t = X_0 + \int_0^t K(t,s)b(X_s)ds + \int_0^t K(t,s)\sigma(X_s)dW_s \tag{3.1.3}$$

where $K(t,s)$ is a deterministic kernel which blows up when $t \to s$.

Denote $\hat{b}(t,s,\omega) = K(t,s)b(X.)$ and $\hat{\sigma}(t,s,\omega) = K(t,s)\sigma(X.)$ and define the following spaces:

$$\Omega \coloneqq C^0([0,T],\mathbb{R}), \quad \bar{\Omega} \coloneqq D^0([0,T],\mathbb{R}^d), \quad \Omega_t = C^0([t,T],\mathbb{R}^d);$$

$$\Lambda \coloneqq [0,T] \times \Omega, \quad \bar{\Lambda} \coloneqq \{(t,\omega) \in [0,T] \times \bar{\Omega} : \omega\,|_{[t,T]} \in \Omega_t\}$$

We first make the following assumptions [28][Section 3.3]:

1. Equation (3.1.3) admits a weak solution and $\mathbb{E}[\sup_{[0,T]} |X_t|^p]$ is finite for $p \geq 1$.
2. For any $t \in (s,T]$, the derivatives $\partial_t \hat{b}$, $\partial_t \hat{\sigma}$ exist and $\phi \in \{\hat{b}, \hat{\sigma}\}$ satisfy the following conditions:

$$|\phi(t,s,\omega)| \leq C_0[1 + \|\omega\|_T^{\kappa_0}](t-s)^{H-\frac{1}{2}}$$

$$|\partial_t\phi(t,s,\omega)| \leq C_0[1 + \|\omega\|_T^{\kappa_0}](t-s)^{H-\frac{3}{2}}$$

In the case where $b(\cdot) = 0$, i.e. fBm, Viens and Zhang [28][Section 2.1] suggested that by introducing an auxillary process Θ, we can decompose X as follows:

$$X_s = \Theta_s^t + I_s^t, \quad t \leq s \leq T$$
$$\Theta_s^t = \int_0^t K(s,r)\sigma(X_r)dW_r, \quad I_s^t = \int_t^s K(s,r)\sigma(X_r)dW_r$$

where I_s^t is independent of $\mathcal{F}_T^{\mathcal{X}}$ and $\Theta_s^t = \mathbb{E}[X_s|\mathcal{F}_t^X]$ a martingale.

In a more general case where $b(\cdot) \neq 0$, $\Theta_s^t = \int_0^t K(s,r)b(X_r)dr + \int_0^t K(s,r)\sigma(X_r)dW_r$ is a semi-martingale instead of a martingale. We can now exploit the (semi-)martingale properties of Θ and adopt existing theories to obtain a deterministic function u in Equation (3.1.1) as a solution of some PPDE.

The function $u(t,\omega)$ now depends on the path $\omega \in \Omega$ instead of x:

$$u(t,\omega) = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}[g(\omega_T) + \int_t^T f(s,\omega_s,\lambda_s)ds|\mathcal{F}_t^X]] \tag{3.1.4}$$

such that $u \in C^{1,2}$.

Remark 3.1.1. Function u is proven to satisfy certain regularity conditions which is important

for numerical implementations when X is a fBM [28]. We assume u also has the same properties for $b(\cdot) \neq 0$ so that the following theorems also apply to general rough volatility models.

From now on, denote $\omega_t \in \mathbb{R}$ as the value of the concatenated path $(X \otimes_t \Theta^t)_\cdot = X_\cdot \mathbf{1}_{[0,t)}(\cdot) + \Theta^t_\cdot \mathbb{1}_{[t,T]}(\cdot)$ evaluated at time t, not to be confused with the path $\omega \in \Omega$.

Definition 3.1.2. Denote $\partial_t u$ as the right time derivative:

$$\partial_t(u(t,\omega)) = \lim_{\delta \downarrow 0} \frac{u(t+\delta,\omega) - u(t,\omega)}{\delta}$$

Denote $\partial_\omega u$ as the Fréchet derivative with respect to $\omega \mathbb{1}_{[t,T]}$ for some $\eta \in \Omega$:

$$u(t, \omega + \eta \mathbb{1}_{[t,T]}) - u(t,\omega) = \langle \partial_\omega u(t,\eta), \omega \rangle + o(\|\eta \mathbf{1}_{[t,T]}\|_T)$$

where

$$\langle \partial_\omega u(t,\omega), \eta \rangle = \lim_{\epsilon \to 0} \frac{u(t, \omega + \epsilon\eta \mathbb{1}_{[t,T]}) - u(t,\omega)}{\epsilon}$$

Proposition 3.1.3. *For some $\lambda \in \mathbb{R}$ and $(t,\omega) \in \bar{\Lambda}$, the Fréchet derivative has the following property:*

$$\langle \partial_\omega u(t,\omega), \lambda\eta \rangle = \lambda \langle \partial_\omega u(t,\omega), \eta \rangle$$

Proof.

$$\begin{aligned}
\langle \partial_\omega u(t,\omega), \lambda\eta \rangle &= \lim_{\epsilon \to 0} \frac{u(t, \omega + \epsilon\lambda\eta \mathbf{1}_{[t,T]}) - u(t,\omega)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{u(t, \omega + \epsilon\lambda\eta \mathbf{1}_{[t,T]}) - u(t,\omega)}{\lambda\epsilon} \lambda \\
&= \lambda \lim_{\lambda\epsilon \to 0} \frac{u(t, \omega + \epsilon\lambda\eta \mathbf{1}_{[t,T]}) - u(t,\omega)}{\lambda\epsilon} \\
&= \lambda \langle \partial_\omega u(t,\omega), \eta \rangle
\end{aligned}$$

□

As an extension of Dupire's functional Itô calculus [29], Viens and Zhang [28] derived the following the functional Itô's formula which not only depends on the path $X_{[0,t]}$ but also on $\Theta_{[t,T]}$.

Theorem 3.1.4. *[28, Theorem 3.10]*

$$du(t,\omega) = \partial_t u(t,\omega) dt + \Big(\langle \partial_\omega u(t,\omega), \hat{b}_t \rangle + \frac{1}{2} \langle \partial_{\omega\omega} u(t,\omega), (\hat{\sigma}_t, \hat{\sigma}_t) \rangle \Big) dt + \langle \partial_\omega u(t,\omega), \hat{\sigma}_t \rangle dW_t$$

Proposition 3.1.5. *The function u (3.1.4) is the solution of the following Hamilton-Jacobi-*

Bellman PPDE:

$$\begin{aligned} \partial_t u + \inf_{\lambda \in \mathcal{A}_{t,T}} \{ \langle \partial_\omega u, \hat{b}_{t,\lambda} \rangle + \frac{1}{2} \langle \partial_{\omega\omega} u, \hat{\sigma}_{t,\lambda}, \hat{\sigma}_{t,\lambda} \rangle \rangle + f(t, \omega_t, \lambda_t) \} = 0 \\ u(T, \omega) = g(\omega) \end{aligned} \tag{3.1.5}$$

Proof. See Appendix A.4. □

3.2 Fractional Brownian motion

We start off with a simple case by replacing the standard Brownian motion in Section 2.2 with a fBM. The dynamics of process X under measure $\mathbb{P}$ is then defined as

$$X_t = X_0 + \int_0^t K(t,s)(\lambda_s ds + dW_s)$$

where $K(t,s)$ is a deterministic kernel which blows up as s tends to t.

Theorem 3.2.1. *The dynamics of fractional Brownian motion under measure $\mathbb{P}$ is given by:*

$$X_t = X_0 + \int_0^t K(t,s) \Big(K(T,s) \partial_x \log \mathbb{E}[e^{-f^*(x)} | \mathcal{F}_s^X] ds + dW_s \Big)$$

Proof. Similar to Section 2.2, we obtain the relative entropy which is exactly the same as in Equation (2.1.1). We then arrive at the same dual problem in Equation (2.2.3).

$$Y_t = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{P}} [\frac{1}{2} \int_t^T \lambda_s^2 \, ds + f(X_T) | \mathcal{F}_t^X] \tag{3.2.1}$$

Note that Equation (3.2.1) only depends on the state of X at time T and λ_t is a deterministic function. Therefore, it is not dependent on previous paths of X and is considered as a simplified version of (3.1.1) [28, Section 2.2]. By the martingale decomposition of fBM in Section 3.1, we can write $Y_t = u(t, \Theta_T^t)$ such that:

$$u(t,x) = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{P}} [\frac{1}{2} \int_t^T \lambda_s^2 \, ds + f(x + I_T^t) | \mathcal{F}_t^X]$$

which gives the following equation:

$$\sup_{f \in L^1(\mu)} \{ u(0, x_0) - \mathbb{E}^{\mu}[f(X_T)] \} \tag{3.2.2}$$

By exploiting the martingale property of Θ, we can apply the standard Itô's formula to obtain a PDE representation of u instead of writing a PPDE:

$$\partial_t u + \frac{1}{2}K(T,t)^2\partial_{xx}u + \inf_{\lambda}\{\frac{1}{2}\lambda_t^2 + K(T,t)\lambda_t\partial_x u\} = 0$$
$$u(T,x) = f(x)$$

Differentiating with respect to λ_t gives $\lambda_t^* = -K(T,t)\partial_x u$. Re-substituting λ^* gives a Burger's like PDE:

$$\partial_t u + \frac{1}{2}K(T,t)^2\partial_{xx}u - \frac{1}{2}(K(T,t)\partial_x u)^2 = 0$$

Applying the same transformation $u = -\log U$, we arrive at a heat equation:

$$\partial_t U + \frac{1}{2}(K(T,t)\partial_x U)^2 = 0$$
$$U(T,x) = e^{-f(x)}$$

Re-writing the dual problem (Equation (3.2.2)) with $u(t,x) = -\log \mathbb{E}[e^{-f(x)}|\mathcal{F}_s^X]$ and denoting f^* as the unique solution of this optimisation completes the proof. □

3.3 Rough volatility models

By applying the non-Markovian framework in Section 3.1, we can now extend Theorem 2.3.1 to a general rough volatility model of the form:

$$dS_t = S_t a_t dW_t^0, \quad d\langle W^0, Z^0\rangle_t = \rho dt$$
$$a_t = a_0 + \int_0^t K(t,s)b(a_t)ds + \int_0^t K(t,s)\sigma(a_t)dZ_s^0$$

where $K(t,s) = \frac{1}{\Gamma(H+\frac{1}{2})}(t-s)^{H-\frac{1}{2}}$.

Theorem 3.3.1. *The dynamics of the calibrated model under the new measure $\mathbb{P}$ is given by:*

$$dS_t = S_t a_t dW_t, \quad d\langle W, Z\rangle_t = \rho dt$$
$$a_t = a_0 + \int_0^t K(t,s)(b(a_t) + (\lambda^*)_s^{(2)})ds + \int_0^t K(t,s)\sigma(a_t)dZ_s^0$$

where $(\lambda^*)_t^{(2)} = \sqrt{1-\rho^2}\langle \log \mathbb{E}^{\mathbb{P}^0}[e^{-f^*(S_T) - \int_t^T \Delta_s^* ds}|\mathcal{F}_t], \hat{\sigma}_t\rangle$ such that $S_T \overset{\mathbb{P}}{\sim} \mu$.

Proof. Similar to Theorem 2.3.1, we need to solve the following problem:

$$sup_{f\in L^1(\mu)}\{u(0,S_0,\omega_0) - \mathbb{E}^{\mu}[f(S_T)]\}$$

where $u(t,s,\omega) = \inf_{\lambda^{(1)},\lambda^{(2)} \in \mathcal{A}_{t,T}} \{\mathbb{E}^{\mathbb{P}}[\frac{1}{2}\int_t^{T_1} (\lambda_s^{(1)})^2 + (\lambda_s^{(2)})^2 ds + f(S_T) + \int_t^T \Delta_s ds | \mathcal{F}_t]\}$. Since Θ is now a semi-martingale, we couldn't apply the same simplification technique in Theorem 3.2.1. We would need to apply Proposition 3.1.5 such that u solves the following HJB PPDE:

$$\begin{aligned} \partial_t u + \mathcal{L}^0 u + \inf_{\lambda^{(1)},\lambda^{(2)}} \sup_{\Delta} \Big\{ \frac{1}{2}(\lambda^{(1)})^2 + \frac{1}{2}(\lambda^{(2)})^2 + \lambda^{(1)}(s\omega_t \partial_s u + s\omega_t \Delta) \\ + \rho \langle \partial_\omega u, \lambda^{(1)} \hat{\sigma}_t \rangle + \sqrt{1-\rho^2} \langle \partial_\omega u, \lambda^{(2)} \hat{\sigma}_t \rangle \Big\} = 0 \\ u(T,s,\omega) = f(s) \end{aligned}$$

where $\mathcal{L}^0 u = \frac{1}{2} s^2 \omega_t^2 \partial_{ss} u + \frac{1}{2} \langle \partial_{\omega\omega}, (\hat{\sigma}_t, \hat{\sigma}_t) \rangle + \langle \partial_\omega u, \hat{b}_t \rangle + \rho s \omega_t \partial_s \langle \partial_\omega, \hat{\sigma}_t \rangle$.

Applying Proposition 3.1.3 simplifies the equation:

$$\begin{aligned} \partial_t u + \mathcal{L}^0 u + \inf_{\lambda^{(1)},\lambda^{(2)}} \sup_{\Delta} \Big\{ \frac{1}{2}(\lambda^{(1)})^2 + \frac{1}{2}(\lambda^{(2)})^2 + \lambda^{(1)}(s\omega_t \partial_s u + \rho \langle \partial_\omega u, \hat{\sigma}_t \rangle + s\omega_t \Delta) \\ + \lambda^{(2)} \sqrt{1-\rho^2} \langle \partial_\omega u, \hat{\sigma}_t \rangle \Big\} = 0 \end{aligned} \tag{3.3.1}$$

Differentiating with respect to $\lambda_t^{(i)}$ then re-substituting the optimal controls give:

$$\begin{aligned} \lambda_t^{(1)} &= -S_t \omega_t \partial_s u(t, S_t, \omega) - \rho \langle \partial_\omega u(t, S_t, \omega), \hat{\sigma}_t \rangle - S_t \omega_t \Delta \\ \lambda_t^{(2)} &= -\sqrt{1-\rho^2} \langle \partial_\omega u(t, S_t, \omega), \hat{\sigma}_t \rangle \end{aligned}$$

$$\begin{aligned} \partial_t u + \mathcal{L}^0 u - \frac{1}{2} \langle \partial_\omega u, \hat{\sigma}_t \rangle^2 - \frac{1}{2} s^2 \omega_t^2 (\partial_s u)^2 - \rho s \omega_t \partial_s \langle \partial_\omega u, \hat{\sigma}_t \rangle \\ + \sup_{\Delta} \Big\{ \frac{1}{2} s^2 \omega_t^2 \Delta^2 - s^2 \omega_t^2 \Delta \partial_s u - \rho s \omega_t \Delta \langle \partial_\omega, \hat{\sigma}_t \rangle \Big\} = 0 \end{aligned}$$

Differentiating with respect to Δ gives $\Delta_t^* = \partial_s u(t, S_t, \omega) - \frac{\rho}{S_t \omega_t} \langle \partial_\omega u(t, S_t, \omega), \sigma_t \rangle$. Re-substituting Δ_t^* then eliminates $\lambda_t^{(1)}$ so that Equation (3.3.1) boils down to the following Burger's PPDE:

$$\partial_t u + \mathcal{L}^0 - \frac{1}{2}(1-\rho^2) \langle \partial_\omega, \hat{\sigma}_t \rangle = 0$$

It is proven by Di Girolami and Russo [30] that the path-dependent heat equation admits a unique solution. However, it is not proven for this specific Burger's PPDE. Therefore, we would need to make an assumption about the uniqueness of solution for this theorem to work. Then by analogy to Section A.3, the optimal control should be of the following form:

$$(\lambda^*)_t^{(2)} = \sqrt{1-\rho^2} \langle \log \mathbb{E}^{\mathbb{P}^0}[e^{-f^*(S_T) - \int_t^T \Delta_s^* ds} | \mathcal{F}_t], \hat{\sigma}_t \rangle$$

□

Conclusion

In this thesis, we started off by exploring the calibration of stochastic volatility and local volatility models as well as highlight their respective flaws, especially in terms of joint calibration.

We then provided a detailed walk-through of a new class of volatility model based on the construction of a Schródinger's bridge, proposed by Henry-Labordère [1]. To exactly calibrate the model to market prices, a drift is added to change the probability measure so that the resulting measure is fit to a market implied marginal. Guyon later applied the same technique and showed evidence that it is a potential solution to the long-standing SPX/VIX joint calibration problem [6, 7]. Upon confirming that volatility is rough by Gatheral et al. [3], we attempted to extend on Henry-Labordère's method to incorporate rough volatility models. The comprehensive paper by Viens and Zhang [28] provided a lot of useful results for tackling non-Markovian process.

Despite the promising potential of the application of Schrödinger's bridge in volatility modelling. Implementation has been proven to be complicated and computationally expensive, especially when rough volatility is being considered, which involves numerical solution of path-dependent PDEs. Therefore, more research should be conducted to develop more efficient numerical algorithms.

Appendix A

Appendix

A.1 Analytical Heston

One of the advantages of the Heston model over other SVM is the presence of a semi-closed analytical formula, derived by the method of characteristic functions [31].

The characteristic function of the Heston model is of the form:

$$\varphi(S_0, K, a_0, \tau; \phi) = e^{C(\tau,\phi)+D(\tau,\phi)a_0+i\phi\log(S_0)}$$

where

$$\begin{aligned} C(\tau,\phi) &= r\phi i\tau + \frac{\kappa\theta}{\sigma^2}\left[(\kappa-\rho\sigma\phi i+d)\tau - 2\log\frac{1-ge^{d\tau}}{1-g}\right] \\ D(\tau,\phi) &= \frac{\kappa-\rho\sigma\phi i+d}{\sigma^2}\left[\frac{1-e^{d\tau}}{1-ge^{d\tau}}\right], \quad g = \frac{\kappa-\rho\sigma\phi i+d}{\kappa-\rho\sigma\phi-d} \\ d &= \sqrt{(\rho\sigma\phi i-\kappa)^2+\sigma^2(\phi i+\phi^2)} \end{aligned}$$

The formula for European call option is then given by:

$$C(S_0, K, a_0, \tau) = \frac{1}{2}(S_0 - Ke^{-r\tau}) + \frac{1}{\pi}\int_0^\infty \Re[e^{r\tau}\frac{\varphi(\phi-i)}{i\phi K^{i\phi}} - K\frac{\varphi(\phi)}{i\phi K^{i\phi}}]d\phi$$

This greatly improves the efficiency on pricing European options comparing to Monte Carlo methods as well as speeds up the calibration process.

A.2 Properties of fractional Brownian motion

Definition A.2.1. A fractional Brownian motion $(W_t^H)_{t\geq 0}$ is a centered Gaussian processes with the covariance function:

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

The distribution is characterised by the Hurst parameter $H \in (0,1)$, which determines the "smoothness" of the motion.

Remark A.2.2. When $H = \frac{1}{2}$, the covariance function is $\mathbb{E}[W_t^{\frac{1}{2}} W_s^{\frac{1}{2}}] = t \wedge s$. This process is simply a standard Brownian motion.

Proposition A.2.3. *A fractional Brownian motion satisfies the following properties:*

Self-similarity: For a fixed $a > 0$, $W_{at}^H \sim a^H W_t^H$

Stationary increments: $W_t^H - W_s^H \sim W_{t-s}^H$

Dependence of increments: Assume $t_1 < s_1 < t_2 < s_2$.

$$\mathbb{E}[(W_{t_1}^H - W_{s_1}^H)(W_{t_2}^H - W_{s_2}^H)] < 0 \text{ for } H \in (0, \frac{1}{2})$$
$$\mathbb{E}[(W_{t_1}^H - W_{s_1}^H)(W_{t_2}^H - W_{s_2}^H)] > 0 \text{ for } H \in (\frac{1}{2}, 1)$$

Proposition A.2.4 (Mandelbrot-Van Ness representation)**.** *For $H \in (0, \frac{1}{2})$ and let $\gamma = \frac{1}{2} - H$, an alternative version of Mandelbrot-Van Ness representation [32] of the fractional Brownian motion is defined as:*

$$W_t^H = C_H \{ \int_{-\infty}^{t} \frac{dW_s}{(t-s)^\gamma} - \int_{-\infty}^{0} \frac{dW_s}{(-s)^\gamma} \}$$

where $C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(1/2+H)\Gamma(2-2H)}}$ and W_t is a standard Brownian motion.

A.3 Proof of Theorem 2.3.1

Proof of Theorem 2.3.1. [1, Section 3.1]
First, rewrite $dZ_t = \rho dW_t + \sqrt{1-\rho^2} dW_t^\perp$ where $\langle dW, dW^\perp \rangle_t = 0$.

Then, under measure $\mathbb{P}$, the dynamics become:

$$dS_t = S_t a_t (dW_t + \lambda_t^{(1)} dt), \quad d\langle W, W^\perp \rangle_t = \rho dt$$
$$da_t = b(a_t)dt + \sigma(a_t) \left(\rho(dW_t + \lambda_t^{(1)} dt) + \sqrt{1-\rho^2}(dW_t^\perp + \lambda_t^{(2)} dt) \right)$$

Similar to Section 2.2, the relative entropy to be minimised in P_1 given by:

$$H(\mathbb{P}|\mathbb{P}^0) = \mathbb{E}^{\mathbb{P}}[\frac{1}{2} \int_0^{T_1} (\lambda_s^{(1)})^2 + (\lambda_s^{(2)})^2 ds]$$

By the same argument, i.e. $S_{T_1} \overset{\mathbb{P}}{\sim} \mu_1 \iff \mathbb{E}^{\mathbb{P}}[f_1(S_{T_1})] - \mathbb{E}^{\mu_1}[f_1(S_{T_1})] = 0$, the potential f_1 is

introduced as a Lagrange multiplier which dualise P_1 into:

$$sup_{f_1 \in L^1(\mu_1)}\{u(0,S_0,a_0) - \mathbb{E}^{\mu_1}[f_1(S_{T_1})]\}$$

where $u(t,s,a) = \inf_{\lambda^{(1)},\lambda^{(2)} \in \mathcal{A}_{t,T}}\{\mathbb{E}^{\mathbb{P}}[\frac{1}{2}\int_t^{T_1}(\lambda^{(1)})^2 + (\lambda^{(2)})^2 ds + f_1(S_{T_1}) + \int_t^{T_1}\Delta_s ds | S_t = s, a_t = a]\}$ which solves the following HJB PDE:

$$\partial_t u + \mathcal{L}^0 u + \inf_{\lambda^{(1)},\lambda^{(2)}} \sup_{\Delta}\{\frac{1}{2}(\lambda_s^{(1)})^2 + \frac{1}{2}(\lambda_s^{(2)})^2 + \lambda^{(1)}(sa\partial_s u + \rho\sigma(a)\partial_a u + sa\Delta) + \lambda^{(2)}\sqrt{1-\rho^2}\sigma(a)\partial_a u\} = 0$$
$$u(T_1,s,a) = f_1(s)$$

where $\mathcal{L}^0$ is the is the infinitesimal generator of the process (S_t, a_t) under the original measure $\mathbb{P}^0$. Differentiating with respect to $\lambda_t^{(i)}$ gives the optimal controls:

$$\begin{aligned}\lambda_t^{(1)} &= -(S_t a_t \partial_s + \rho\sigma(a_t)\partial_a)u(t,S_t,a_t) - S_t a_t \Delta_t \\ \lambda_t^{(2)} &= -\sqrt{1-\rho^2}\sigma(a_t)\partial_a u(t,S_t,a_t)\end{aligned}$$

After re-substituting $\lambda_t^{(i)}$ and differentiating with respect to Δ, we see that $\Delta_t^* = -(\partial_s + \rho\frac{\sigma(a_t)}{S_t a_t}\partial_a)u$. The optimal value Δ_t^* forces $\lambda_t^{(1)}$ to become zero, so process S_t remains a martingale even under the new measure and is a crucial criteria in this method.

Rearranging the equation admits a Burger's like semi-linear PDE:

$$\begin{aligned}\partial_t u + \mathcal{L}^0 u - \frac{1}{2}(1-\rho^2)(\sigma(a)\partial_a u)^2 = 0 \\ u(T_1,s,a) = f_1(s)\end{aligned} \tag{A.3.1}$$

Solving the PDE and re-substituting the solution u into $\lambda_t^{(1)}$ then results in the calibrated dynamics.

These procedures are repeated on each sub-interval $[T_{i-1}, T_i]$ for $2 \le i \le n$. The concatenated dynamics is then exactly calibrated on the interval $[T_0, T_n]$. □

A.4 Path-dependent HJB

First, consider the function $u(t,x)$ in Equation (2.2.3).

Theorem A.4.1 (Dynamic programming principle)**.** *For all $s \in [t,T]$:*

$$u(t,s) = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}[\int_t^s f(r,x_r,\lambda_r)dr + u(s,x_s)|\mathcal{F}_t^X]$$

For any arbitrary control $\lambda^0 \in \mathcal{A}_{t,T}$, the theorem has an equivalent formulation:

$$u(t,x_t) \leq \mathbb{E}[\int_t^s f(r,x_r,\lambda^0)dr + u(s,x_s)|\mathcal{F}_t^X]$$

Now for the function that we are interested in, i.e. $u(t,\omega)$ in Equation (3.1.4), $x \in \mathbb{R}^n$ is replaced by a path $\omega \in \Omega$. For the dynamic programming principle to work in the space, we propose the following:

Conjecture A.4.2. For all $s \in [t,T]$ and $u(t,\omega) \in C^{1,2}(\bar{\Lambda})$:

$$u(t,\omega_t) = \inf_{\lambda \in \mathcal{A}_{t,T}} \mathbb{E}[\int_t^s f(r,\omega_r,\lambda_r)dr + u(s,\omega_s)|\mathcal{F}_t^X]$$

For any arbitrary control $\lambda^0 \in \mathcal{A}_{t,T}$, it has an equivalent formulation:

$$u(t,\omega_t) \leq \mathbb{E}[\int_t^s f(r,\omega_r,\lambda^0)dr + u(s,\omega_s)|\mathcal{F}_t^X]$$

The path-dependent HJB equation is briefly mentioned in [28, Section 4.3], however the authors did not provide any formal proof in the paper. Therefore, we try to present an original derivation of the PPDE, based on the derivation of the classical HJB equation in the textbook by Pham [33, Section 3.4].

Proof of Proposition 3.1.5. For some small positive h such that $t < t+h < T$, apply the dynamic programming principle (A.4.2):

$$u(t,\omega) \leq \mathbb{E}^{\mathbb{P}}[\int_t^{t+h} f(s,\omega_s,\lambda^0)ds + u(t+h,\omega)|\mathcal{F}_t^X] \tag{A.4.1}$$

Applying the functional Itô's formula (Theorem 3.1.4) on the interval $[t,t+h]$:

$$u(t+h,\omega) = u(t,\omega) + \int_t^{t+h} \Big(\partial_s u(s,\omega) + \mathcal{L}^{\lambda^0} u(s,\omega)\Big)ds + \int_t^{t+h} \langle \partial_\omega u(s,\omega), \hat{\sigma}_{s,\lambda}\rangle dW_s$$

where

$$\mathcal{L}^{\lambda^0} u(t,\omega) = \langle \partial_\omega u(t,\omega), \hat{b}_{t,\lambda^0}\rangle + \frac{1}{2}\langle \partial_{\omega\omega} u(t,\omega), (\hat{\sigma}_{t,\lambda^0}, \hat{\sigma}_{t,\lambda^0})\rangle$$

We then substitute $u(t+h,\omega)$ in Equation (A.4.1). By setting $h \to 0$ and the mean value theorem, we obtain the following inequality:

$$\partial_t u(t,\omega) + \mathcal{L}^{\lambda^0} u(t,\omega) + f(t,\omega_t,\lambda^0) \geq 0$$

Since λ^0 is arbitrary, we can write:

$$\partial_t u(t,\omega) + \inf_{\lambda^0 \in \mathcal{A}} \{\mathcal{L}^{\lambda^0} u(t,\omega) + f(t,\omega_t,\lambda^0)\} \geq 0 \tag{A.4.2}$$

Next, let λ_t^* be the optimal control, it gives the equality:

$$u(t,\omega) = \mathbb{E}^{\mathbb{P}}[\int_t^{t+h} f(t,\omega_t,\lambda_t^*)ds + u(t+h,\omega)|\mathcal{F}_s^X]$$

By the same arguments as above:

$$\partial_t u(t,\omega) + \mathcal{L}^{\lambda^*} u(t,\omega) + f(t,\omega_t,\lambda_t^*) = 0 \tag{A.4.3}$$

Combining Equation (A.4.2) and (A.4.3), we obtain the HJB representation of the optimal control problem:

$$\partial_t u + \inf_{\lambda\in\mathcal{A}}\{\langle\partial_\omega u, \hat{b}_{t,\lambda}\rangle + \frac{1}{2}\langle\partial_{\omega\omega} u, (\hat{\sigma}_{t,\lambda},\hat{\sigma}_{t,\lambda}))\rangle + f(t,\omega_t,\lambda_t)\} = 0$$
$$u(T,\omega) = f(\omega_T)$$

□

Bibliography

[1] Pierre Henry-Labordere. From (martingale) Schrödinger bridges to a new class of stochastic volatility model. *SSRN Electronic Journal*, Apr 2019.

[2] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

[3] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *SSRN Electronic Journal*, Oct 2014.

[4] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2010.

[5] Marco Avellaneda. Minimum-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, 01(04):447–472, 1998.

[6] Julien Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *SSRN Electronic Journal*, Jun 2019.

[7] Julien Guyon. Dispersion-constrained martingale Schrödinger problems and the exact joint S&P 500/VIX smile calibration puzzle. *SSRN Electronic Journal*, Jul 2021.

[8] Bruno Dupire. Pricing with a smile. *Risk Magazine*, 7(1), 1994.

[9] Cristian Homescu. Local stochastic volatility models: Calibration and pricing. *SSRN Electronic Journal*, Jul 2014.

[10] Alexander Lipton. The vol smile problem. *Risk Magazine*, page 61–65, Feb 2002.

[11] Yaozhong Hu, Yanghui Liu, and David Nualart. Rate of convergence and asymptotic error distribution of Euler approximation schemes for fractional diffusions. *The Annals of Applied Probability*, 26(2), 2016.

[12] Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2):309–349, 2019.

[13] Jingtang Ma and Haofei Wu. A fast algorithm for simulation of rough volatility models. *Quantitative Finance*, 22(3):447–462, 2021.

[14] Rama Cont and Thomas Kokholm. A consistent pricing model for index options and volatility derivatives. *SSRN Electronic Journal*, Sep 2009.

[15] Jim Gatheral, Paul Jusselin, and Mathieu Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *SSRN Electronic Journal*, Jan 2020.

[16] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):89–113, 2021.

[17] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via Schrödinger bridge. preprint available at webpage `proceedings.mlr.press/v139/wang21l/wang21l.pdf`, 2021.

[18] Genji Kawakita, Shunsuke Kamiya, Shuntaro Sasai, Jun Kitazono, and Masafumi Oizumi. Quantifying brain state transition cost via Schrödinger bridge. *Network Neuroscience*, 6(1):118–134, 2022.

[19] Andreas Søjmark. MATH97117 Convex Optimization: Part 2. Lecture Notes. Imperial College London, Department of Mathematics.

[20] Stefano De Marco and Pierre Henry-Labordère. Linking vanillas and VIX options: A constrained martingale optimal transport problem. *SIAM Journal on Financial Mathematics*, 6(1):1171–1194, 2015.

[21] Julien Guyon, Romain Menegaux, and Marcel Nutz. Bounds for VIX futures given S&P 500 smiles. *SSRN Electronic Journal*, Jun 2017.

[22] Jan Blechschmidt. *Numerical methods for stochastic control problems with applications in financial mathematics*. PhD thesis, Chemnitz University of Technology, 2022.

[23] Nicolas Perkowski. MATH555 Backward stochastic differential equations : An introduction. Lecture Notes, 2011. Pennsylvania State University, Department of Mathematics.

[24] Julien Guyon and Pierre Henry-Labordere. *Backward Stochastic Differential Equations*, page 173–186. Chapman and Hall/CRC, 2013.

[25] Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6(4):449–471, 2002.

[26] Bruno Bouchard. Numerical approximation of bsdes by using backward Euler schemes. European Summer School in Mathematical Finance, Aug 2016.

[27] Julien Guyon. Vix-constrained Schrödinger bridges: joint calibration of SPX and VIX smiles with continuous stochastic volatility models. `https://www.youtube.com/watch?v=zCARb_EDuHE&ab_channel=InstitutodeMatem%C3%A1ticaPuraeAplicada`.

[28] Frederi Viens and Jianfeng Zhang. A martingale approach for fractional Brownian motions and related path dependent pdes. *The Annals of Applied Probability*, 29(6), 2019.

[29] Bruno Dupire. Functional Itô calculus. *SSRN Electronic Journal*, 2009.

[30] Cristina Di Girolami and Francesco Russo. About classical solutions of the path-dependent heat equation. *Random Operators and Stochastic Equations*, 28(1):35–62, 2020.

[31] Sergei Mikhailov and Ulrich Nögel. Heston's stochastic volatility model implementation, cal-

ibration and some extensions. *Wilmott magazine*, 2003.

[32] C. Jost. On the connection between Molchan-Golosov and Mandelbrot-Van Ness representations of fractional Brownian motion. *Journal of Integral Equations and Applications*, 20(1), 2008.

[33] Huyên Pham. Continuous-time stochastic control and optimization with financial applications. *Stochastic Modelling and Applied Probability*, 2009.