Imperial College London

Imperial College London

Department of Mathematics

Fokker-Planck Calibration of one Factor Stochastic Local Volatility Models

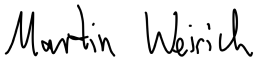
Author: Martin Maximilian WEIRICH (CID: 02274905)

A thesis submitted for the degree of

MSc in Mathematics and Finance, 2022-2023

Declaration

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



Acknowledgements

I would like to thank the amazing QPricer team for a great time during my internship. I want to especially thank Kaitai Chan for his invaluable advice for my project. Furthermore I would like to extend my gratitude to Professor Jacquier for being my academic supervisor for this thesis and also for his excellent work as the course director of the MSc .
Finally I want to thank my parents for enabling me to attend this program, and their unwavering support throughout my studies.

Abstract

Both local and stochastic volatility models are widespread and popular models for pricing derivatives. Both have distinct benefits over each other, with the local volatility model introduced by Dupire in [1] being able to perfectly fit to all currently observed option market quotes for one underlying, while featuring unrealistic volatility dynamics. Stochastic volatility models on the other hand provide realistic volatility dynamics making them suitable for pricing and hedging of exotic options depending on these dynamics, but can not consistently be perfectly calibrated to market quotes.

Stochastic local volatility models are a newer class of models that seek to combine the advantages of both of the former models and thus eliminate their drawbacks. Due to their success in accomplishing this stochastic Local volatility models have become the industry standard for Foreign Exchange and Equity Exotic derivative pricing.

However the calibration of Stochastic local volatility models is an involved and challenging problem.

In this thesis we will briefly introduce local stochastic volatility models from a theoretical standpoint and then attempt to develop, by compiling methods and ideas from existing literature, a generic framework for a stable calibration of one factor stochastic local volatility models based on numerically solving a highly non linear 2 dimensional partial differential equation using finite difference methods.

We will further exemplify this generic framework by calibrating three different local stochastic volatility models and discussing the results, highlighting in which cases satisfactory calibration can be archived, and where further improvements on our methodology are needed.

Finally we will discuss advantages and disadvantages of the three example models based on our numerical results.

Contents

List of Figures

List of Tables

Introduction

Although revolutionary for the field of option pricing, the option pricing framework introduced by Black and Scholes in [2] was quickly found to be an insufficient model of reality, as it can not be consistent with all market prices of European options for one underlying at the same time, and features very unrealistic volatility dynamics, as real world evidence shows that volatility is not constant over time. Quickly extensions of the Black-Scholes model were developed in order to correct its issues. The two major ones of these being stochastic volatility models, where not only the price but also the volatility of the underlying are modeled as stochastic processes, and Dupire's local volatility model [1], where the volatility of the underlying is modeled as a deterministic function of the price of the underlying and time. Both of these model families provide major advantages over the Black-Scholes model: Stochastic volatility models display a lot more realistic volatility dynamics than the Black-Scholes model and the local volatility model is consistent with all observed market prices of European options on a single underlying.
However stochastic volatility models in general can not fit all market prices perfectly, and the local volatility model displays unrealistic volatility dynamics. Hence it would be ideal to have a class of models which combines the strengths of both models. Stochastic local volatility models, first introduced by Jex et al in [3, Equations 4-6, page 7] in the context of trees, and further developed in a continuous setting in [4] and [5] aim to do just that.
By modelling the volatility of the underlying partially as a stochastic process and partially as a function of the underlying and time stochastic local volatility models are effectively a hybrid of the local and stochastic volatility model. It can be shown through a theorem due to Gyöngy [6], which will be introduced later, that stochastic local volatility models can inherit the property of perfectly fitting vanilla option market prices from the local volatility model. Furthermore due to the stochastic component of the volatility they can also archive realistic volatility dynamics.
The price one has to pay for this is that the calibration of stochastic local volatility models is quite involved. In this thesis we are going to study the calibration of one factor stochastic local volatility models, following a partial differential equation based approach laid out in [5]. Our aim is to provide a general easily adaptable framework combining current literature on the topic that will allow to calibrate any kind of one factor stochastic local volatility model. We will exemplify this framework by calibrating three different popular stochastic local volatility models and critiquing our numerical results.

The Thesis will be structured as follows: In Chapter 1 we will provide some mathematical background by recalling some definitions and results from the theory of stochastic differential equations that will be used throughout the thesis and moreover will allow us to formulate the two main results the calibration procedure will be based on. Then in Chapter 2 we are going to introduce the local volatility model and in Chapter 3 stochastic volatility models and discuss their respective strengths and weaknesses. Furthermore in Chapter 3 we will take a closer look at three specific stochastic volatility models, stating some of their properties that will come in handy later down then line when they will be used as the basis for the local stochastic volatility models under consideration in this thesis . Following that we will introduce stochastic volatility models in general , and the three models used throughout this thesis in particular, from a theoretical perspective in Chapter 4 and thereafter discuss their numerical calibration in Chapter 5. Finally in Chapter 6 we will share our numerical results for the calibration of all three models, use the calibrated models to price forward starting options to investigate if all three of our chosen stochastic models indeed display distinct volatility dynamics from the local volatility model and discuss the advantages and disadvantages of choosing a specific one out of these models based on the results.

Chapter 1

Stochastic Differential Equations

Since the models we consider in this thesis are all expressed as stochastic differential equations, from here on SDEs, we will quickly recall some results from the theory of SDEs that will be repeatedly used in the following chapters. Mainly the goals of this chapter are to develop the machinery to state and understand a theorem due to Gyöngy [6] which will later allow us to calibrate stochastic volatility models to market prices, and to understand the connection between the transition probability density of a stochastic process and a certain partial differential equation that will also play a big role in calibrating stochastic local volatility models.
Throughout this chapter we will assume that every probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a stochastic basis, that is, it is complete and admits a filtration $(\mathcal{F}_t)_{t\geq 0}$ that satisfies the **usual conditions** that the filtration is right continuous and $\mathcal{F}_0$ contains all $\mathbb{P}-$Null sets. Furthermore we shall assume that every stochastic basis we consider supports a m-dimensional $\mathcal{F}$-Brownian motion $(W_t)_{t\geq 0}$.
Consider the SDE

$$\begin{cases} dX_t = \mu(X_t, t)\, dt + \sigma(X_t, t)\, dW_t \\ X_0 = Z \end{cases}, \tag{1.0.1}$$

where Z is a square integrable $\mathbb{R}^n$ valued random variable independent of the filtration generated by the Brownian motion and $\mu : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^{n\times m}$ are Borel-measurable functions. Here $\mu(x, t)$ is called the drift vector, $\sigma(t, x)$ the dispersion matrix and $a(t, x) := \sigma(t, x)\sigma(t, x)^\top$ the diffusion matrix.
First and foremost we need to understand what it means to be a solution of (1.0.1). Two concepts for this that will be introduced now exist.

Definition 1.0.1 (Strong Solution)**.** Denote the filtration generated by the Brownian motion W as $\mathcal{F}^W$ and define the joint filtration of Z and W augmented by the $\mathbb{P}$- null sets $\mathcal{N}$ with $\mathcal{G}_t := \sigma((\sigma(Z) \vee \mathcal{F}_t^W) \cup \mathcal{N})$.
We say a stochastic process $X = (X_t)_{t\geq 0}$ is a strong solution of (1.0.1) if

1. X is adapted to $(\mathcal{G}_t)_{t\geq 0}$
2. $\mathbb{P}(X_0 = Z) = 1$
3. $X_t = X_0 + \int_0^t \mu(X_s, s)\, ds + \int_0^t \sigma(X_s, s)\, dW_s$ almost surely
4. $\int_0^t |\mu(X_s, s)| + \sigma_{i,j}^2(X_s, s)\, ds < \infty$ $\mathbb{P} - a.s.$ for all $t > 0$, $1 \leq i \leq n$, $1 \leq j \leq m$

Definition 1.0.2 (Weak Solution)**.** We say (1.0.1) admits a weak solution if there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$, and a pair of processes X, W adapted to this filtration, so that W is a Brownian motion and X satisfies conditions 2-4 from 1.0.1

Remark 1.0.3. The main difference between a strong and a weak solution is that for a weak solution the process X need not be a measurable with respect to the filtration generated by the Brownian motion.

With the notion of solutions of SDEs in place the next natural step is to establish existence and uniqueness results. As it turns out the familiar existence and uniqueness result from the theory of ODEs generalises to the SDE case.

Theorem 1.0.4 (Existence and uniqueness of strong Solutions)**.** *Assume that the drift vector and dispersion matrix of* (1.0.1) *are Lipschitz continuous and have linear growth, that is there exists* $K > 0$ *so that*

$$|\mu(x,t) - \mu(y,t)| + ||\sigma(x,t) - \sigma(y,t)|| \leq K|x-y|$$

and

$$|\mu(x,t)|^2 + ||\sigma(x,t)||^2 \leq K^2(1+|x|^2)$$

for every $x, y \in \mathbb{R}^n$ *and* $t \in [0,\infty)$*, where*

$$||\sigma(x,t)||^2 = \sum_i^n \sum_j^m \sigma_{i,j}(t,x)^2.$$

Then there exists a continuous, square integrable strong solution X *to the SDE* (1.0.1)*. Moreover this solution is unique in the sense that for every other strong solution* Y *to* (1.0.1) *relative to the same Brownian motion*

$$\mathbb{P}(\{X_t = Y_t, t \geq 0\}) = 1.$$

This is usually referred to as X, Y *being indistinguishable.*

For a proof of 1.0.4 we refer to [7, Theorems 2.5, 2.9, Chapter 5.2, pages 287, 289]. In one dimension this result can be extended to hold under milder assumptions on the dispersion matrix.

Theorem 1.0.5 (Yamada-Watanabe)**.** *Consider again the SDE* (1.0.1) *with* $n = m = 1$*. Assume the drift coefficient is Lipschitz continuous and the dispersion matrix satisfies*

$$||\sigma(x,t) - \sigma(y,t)|| \leq h(|x-y|),$$

where $h : [0,\infty) \to [0,\infty)$ *is strictly increasing with* $h(0) = 0$ *and for every* $\varepsilon > 0$

$$\int_0^\varepsilon \frac{1}{h^2(x)}\, dx = \infty,$$

then there exists a unique strong solution with uniqueness again understood as in 1.0.4

See [7, Proposition 2.13, Chapter 5.2, page 291,Corollary 3.23, Chapter 5.3, page 310] or [8, Theorem 1, page 164].
If it is known that a solution for a particular SDE exists, for example due to one of the two theorems presented above, the next step is then the qualitative study of this solution. As alluded to a major quantity of interest here are the marginal distributions of the solution. They can be characterised as the solution to the so called Fokker-Planck, or Kolmogorov forward, partial differential equation.

Theorem 1.0.6 (Fokker-Planck Equation)**.** *Assume the conditions on* μ, σ *in 1.0.4 hold and* X *is the strong solution of* (1.0.1)*, and its transition distribution function given that* $X_s = y$ *for* $t > s$ *admits the density* $p_{s,y}(\cdot, t)$*.*
Additionally let μ, σ *be continuous in* t*, the diffusion matrix be uniformly positive definite, meaning for all* $v, x \in \mathbb{R}^n$*,* $t \in [0,\infty)$ *there exists* $c > 0$ *so that*

$$v^\top a(t,x) v \geq c|v|,$$

and $\partial_x \mu$*,* $\partial_x a$*,* $\partial_x^2 a$ *satisfy the Lipschitz and linear growth condition in* x*.*
Then $p_{s,y}$ *is the unique solution to the PDE*

$$\begin{cases} \frac{\partial p}{\partial t} = -\sum_{i=1}^N \frac{\partial}{\partial x_i}(\mu p) + \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \frac{\partial}{\partial x_i \partial x_j}(a_{i,j} p) \\ p(s,x) = \delta_y(x) \end{cases} . \tag{1.0.2}$$

See [9, Theorem 10.9.11, page 226].
Lastly we state the main result of this chapter. A theorem by Gyöngy [6, Theorem 4.6, page 516] that shows that, given a stochastic process that is the solution to an SDE with quite general coefficients, it's possible to construct another SDE with "simpler" coefficients so that it's solution has the same one dimensional marginal distributions as the former process. As alluded to earlier this will be central to calibrate local stochastic volatility models by linking their marginal distribution to that of the local volatility model which will guarantee a perfect fit to market prices of European options.

Theorem 1.0.7 (Mimicking Theorem)**.** *Consider a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. *Let* X *be a* $\mathbb{R}^n$ *valued Ito-process given by*

$$X_t = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s, \qquad t \geq 0,$$

where μ *is a bounded,* $\mathbb{R}^n$ *valued* $(\mathcal{F}_t)_{t\geq 0}$ *adapted process and* σ *is a bounded,* $\mathbb{R}^{n\times m}$ *matrix valued process adapted to the same filtration as* μ. *Furthermore assume that the diffusion matrix is uniformly positive definite.*
Then there exist Lebesgue measurable functions $\widehat{\mu}$, $\mathbb{R}^d$ *valued, as well as* $\widehat{\sigma}$, $\mathbb{R}^{n\times m}$ *valued, defined on* $\mathbb{R}^n \times [0, \infty)$, *so that*

$$\widehat{\mu}(x,t) = E[\mu_t \,|\, X_t = x],$$
$$\widehat{\sigma}(x,t)\widehat{\sigma}^\top(x,t) = E[\sigma_t \sigma_t^\top \,|\, X_t = x],$$

$\mathbb{P}$*-almost sure, for Lebesgue almost all* t. *Moreover the SDE*

$$\begin{cases} \frac{d\widehat{X}_t}{\widehat{X}_t} & = \widehat{\mu}_t \, dt + \widehat{\sigma}_t \, d\widehat{W}_t \\ \widehat{X}_0 & = 0 \end{cases}$$

admits a weak solution $\widehat{X}$ *so that* X_t *and* $\widehat{X}_t$ *have the same one-dimensional marginal distributions for all* $t \geq 0$.

Due to the usefulness of this result for financial mathematics, that will be elucidated in the following chapters, it still receives attention and more recently has been extended in [10, Corollary 3.7, page 1592] to hold under weaker conditions on σ and μ, merely requiring
$\mathbb{E}\left[\int_0^t (|\mu_s| + ||\sigma_s \sigma_s^\top||) \, ds\right] < \infty$, with random intital value $X_0 \neq 0$, and even for suitable functions of X.

Chapter 2

The Local Volatility Model

2.1 Implied Volatility

Black and Scholes with their paper [2] , which introduced the eponymous Black-Scholes model, revolutionised the theory of derivative pricing.
Under this model, given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the risk-neutral dynamics of the stock price process $(S_t)_{t\geq 0}$ are given by the SDE

$$\frac{dS_t}{S_t} = r\,dt + \sigma\,dW_t, \qquad S_0 > 0, \tag{2.1.1}$$

where r is the instantaneous risk-free interest rate and σ the instantaneous volatility.
We will assume a slight generalisation of (2.1.1) allowing for bounded on any finite interval, time dependent but deterministic instantaneous rate $(r_t)_{t\geq 0}$ and dividend yield $(q_t)_{t\geq 0}$. The risk neutral dynamics of the stock price under this generalised model then become:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma\,dW_t, \tag{2.1.2}$$

which admits a unique strong solution up to any finite time $T > 0$ by Theorem 1.0.4.
In [2, Equation 13, page 644] a formula for the price of a European Call option at inception for an asset following (2.1.1) is derived. Assuming dynamics (2.1.2) this formula generalises as follows: Denoting the value of a European Call option on an underlying with current spot S_0 at inception with strike K and maturity $T > 0$, under the assumption that S is given by (2.1.2), by $C_{\mathrm{BS}}(S_0, K, T, \sigma)$, then

$$C_{\mathrm{BS}}(S_0, K, T, \sigma) := S_0 \mathrm{e}^{-TQ_T}\mathcal{N}(d_+) - K\mathrm{e}^{-TR_T}\mathcal{N}(d_-), \tag{2.1.3}$$

with

$$\begin{aligned} d_\pm &:= \frac{\log\left(\frac{S_0}{K}\right) + T(R_T - Q_T)}{\sigma_{\mathrm{BS}}} \pm \frac{\sigma_{\mathrm{BS}}}{2}, \\ \sigma_{\mathrm{BS}} &:= \sigma\sqrt{T}, \\ R_T &:= \frac{1}{T}\int_0^T r_s\,ds, \\ Q_T &:= \frac{1}{T}\int_0^T q_s\,ds, \end{aligned}$$

and $\mathcal{N}$ the cumulative distribution function of a standard normal random variable.
We notice that

$$\frac{\partial C_{\mathrm{BS}}}{\partial \sigma}(S_0, K, T, \sigma) = \mathrm{e}^{-TQ_T} S_0 \mathcal{N}'(d_+)\sqrt{T} > 0,$$

for all $\sigma > 0$ and thus (2.1.3) is strictly increasing in σ. Furthermore

$$\begin{aligned} \lim_{\sigma\to 0} C_{\mathrm{BS}}(S_0, K, T, \sigma) &= \mathrm{e}^{-TQ_T} S_0 - K\mathrm{e}^{-TR_T}, \\ \lim_{\sigma\to\infty} C_{\mathrm{BS}}(S_0, K, T, \sigma) &= \mathrm{e}^{-TQ_T} S_0. \end{aligned}$$

Since under our assumptions on the rate and dividend yield for no-arbitrage reasons the market price of a European Call option with strike K and maturity T $C_{\text{Mkt}}(K,T)$ has to satisfy

$$C_{\text{Mkt}}(K,T) \in (\mathrm{e}^{-TQ_T}S_0 - K\mathrm{e}^{-rT}, \mathrm{e}^{-TQ_T}S_0),$$

we can thus conclude, due to the strict monotonicity of C_{BS} in σ, that there exists an inverse mapping, the so called implied volatility

$$\sigma_{\text{imp}} : \begin{cases} (\mathrm{e}^{-TQ_T}S_0 - K\mathrm{e}^{-rT}, \mathrm{e}^{-TQ_T}S_0) \to (0,\infty) \\ C_{\text{Mkt}}(T,K) \mapsto \sigma_{\text{imp}}(K,T) \end{cases}, \tag{2.1.4}$$

so that when plugged into (2.1.3) one uniquely recovers the market price of the option, i.e.

$$C_{\text{Mkt}}(K,T) = C_{\text{BS}}(S_0, K, T, \sigma_{\text{imp}}(K,T)).$$

.

2.2 The Local Volatility Model

Since in disagreement with the Black-Scholes model European Call options exhibit implied volatilities that are not given by

$$\sigma_{\text{imp}}(K,T) = \sigma\sqrt{T},$$

for some $\sigma > 0$, but display dependence on K and T, see Figure 2.1, the Black-Scholes model can not be consistent with all market prices for the vanilla options on a given underlying. To introduce some terminology that we will use from here on out, let us point out that due to the usual shapes of $\sigma_{\text{imp}}(\cdot,T)$ for fixed T this phenomenon is usually referred to as the volatility smile or smirk, again see Figure 2.1.

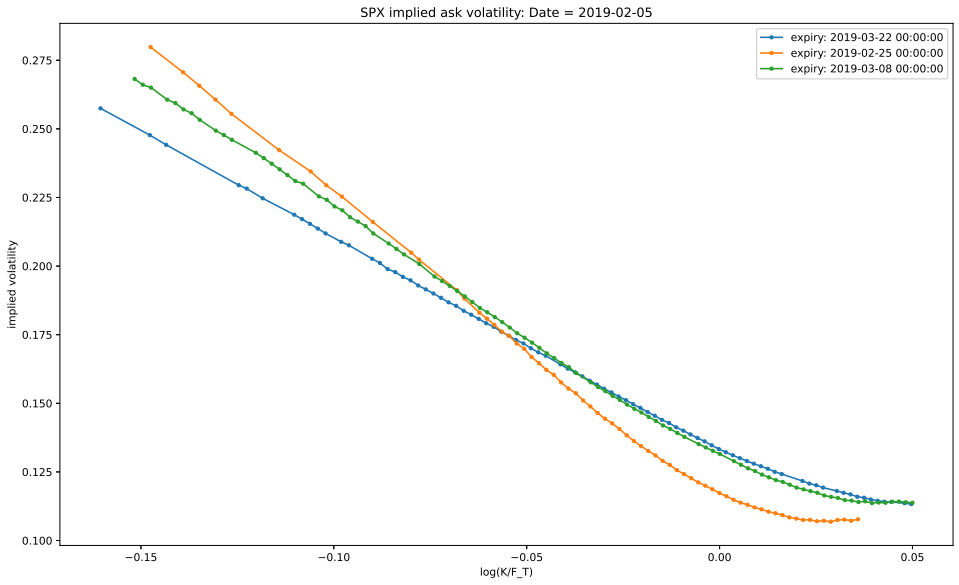


Figure 2.1: SPX ask implied volatility on 2019-02-05 at different maturities as a function of $\log(\frac{K}{F_T})$.

It is therefore desirable to find a model that is consistent with all the observed vanilla option prices at the same time.
In [1] Dupire shows this problem can be solved by a diffusion model where the instantaneous volatility is a function of t and S_t. That is he proposes the risk-neutral dynamics

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma_{loc}(S_t,t)\,dW_t. \tag{2.2.1}$$

This has the desirable effect that the local volatility model retains the completeness of the Black-Scholes model, allowing for all options to theoretically be hedged using tradable assets.
Now recalling that the price of a European option is fully determined by the marginal distribution of the underlying's process under the risk neutral measure we can interpret the process given by (2.2.1) as the mimicking process of a more complicated "true" price process obtained through an application of Theorem 1.0.7.
However as we we have no knowledge of this "true" process, we can not use 1.0.7 to infer an exact expression for the local volatility σ_{loc}. Instead in [1] Dupire makes use of Breeden and Litzenberger's observation [11, Equation 2, page 627] that the risk neutral marginal probability density may be recovered from the market prices of European Call options by differentiating twice with respect to the strike. Through equating this density with the Fokker-Planck equation for (2.2.1) Dupire obtains a formula for the local volatility term. Let us now present this derivation following [12, Pages 9-11].
Letting $p(\cdot, t)$ denote the risk neutral marginal density of the stock price at time t, the undiscounted price of an European Call with strike K is given by

$$C(K,T) = \int_K^\infty (x-K)p(x,T)\,dx. \tag{2.2.2}$$

Then by the Leibniz-rule

$$\partial_K C(K,T) = -\int_K^\infty p(x,T)\,dx, \tag{2.2.3}$$

$$\partial_K^2 C(K,T) = p(K,T). \tag{2.2.4}$$

On the other hand assuming the conditions of 1.0.6 are satisfied the Fokker-Planck equation applied to (2.2.1) yields

$$\frac{\partial p}{\partial T} = -(r_T - q_T)\frac{\partial}{\partial x}(xp) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma_{loc}^2 x^2 p).$$

Now differentiating (2.2.2) with respect to T and substituting the Fokker-Plank equation we obtain

$$\frac{\partial}{\partial T}C(K,T) = \int_K^\infty \left[-(r_T - q_T)\frac{\partial}{\partial x}(xp(x,T)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma_{loc}^2(x,T)x^2 p(x,T))\right](x-K)\,dx.$$

Integrating by parts we get, assuming that p decays to 0 sufficiently fast as $|x| \to \infty$,

$$\frac{\partial}{\partial T}C(K,T) = \int_K^\infty (r_T - q_T)xp(x,T) - \frac{1}{2}\frac{\partial}{\partial x}(\sigma_{loc}^2(x,T)x^2 p(x,T))\,dx.$$

Substituting (2.2.2) for the first term of the integral yields

$$\frac{\partial}{\partial T}C(K,T) = (r_T - q_T)\left(C(K,T) + \int_K^\infty Kp(x,T)\,dx\right) - \int_K^\infty \frac{1}{2}\frac{\partial}{\partial x}(\sigma_{loc}^2(x,T)x^2 p(x,T))\,dx.$$

Finally integrating the second integral, still assuming sufficiently fast decay of p, and substituting our expressions for the derivatives of the Call with respect to strike (2.2.3) we arrive at Dupire's equation

$$\frac{\partial}{\partial T}C(K,T) = (r_T - q_T)\left(C(K,T) - K\frac{\partial}{\partial K}C(K,T)\right) + \frac{\sigma_{loc}^2(K,T)K^2}{2}\frac{\partial^2}{\partial K^2}p(K,T),$$

which can be rearranged as

$$\sigma_{loc}^2(K,T) = \frac{\partial_T C(K,T) - (r_T - q_T)\left(C(K,T) - K\partial_K C(K,T)\right)}{\frac{1}{2}K^2\partial_K^2 C(K,T)}. \tag{2.2.5}$$

So since by construction the marginal density of the local volatility model matches the market implied risk neutral density the local volatility model will perfectly match the market prices.
However in practice we do not observe a continuum of listed options. Therefore (2.2.5) can not be directly applied to compute the local volatility for arbitrary strike and expiry, but instead first a continuous in time and strike surface of vanilla option prices needs to created. This is usually accomplished by interpolating market quotes in implied volatility either using traditional

interpolating techniques, where special care has to be taken not to create static arbitrage during interpolation, or by fitting a parametric function to the market implied volatilities.
For the latter approach a popular parametrisation is given by the SVI/SSVI family of models, see [13], that offer parametrisations of either individual volatility slices or the whole surface. A great advantage of the SSVI model over traditional interpolation is that conditions on its parameters were derived in [13] and [14] that guarantee the generated surface to be arbitrage free. On the other hand due to their parametric nature models like SSVI may struggle to fit some volatility smile shapes.
In general the proper interpolation of implied volatilities is a still actively discussed and important topic in quantitative finance which we shall not further dwell on.
Lets us instead discuss some of the limitations of the local volatility model.

2.3 Limitations of the Local Volatility Model

As discussed in the previous section the main attractive feature of the local volatility model is that it is consistent with market prices of European vanilla options since it perfectly fits the market implied marginal probability distributions. The problem with this approach is that it pins down future volatility as a deterministic function of the underlying's price which leads to unrealistic dynamics of the volatility process, since this way there is no model imposed on the conditional transition probabilities of the underlying, instead everything in the model is determined by the current observed marginal distributions. Essentially using local volatility we impose that the behavior of volatility in the future is fully determined by today's market conditions, which is an unreasonable assumption.
From a pricing perspective the lack of a model on the conditional transition probabilities is problematic since these come into effect when pricing exotic options that depend on the evolution of volatility over time.
As an example consider entering at time t_0 a forward starting Call with strike K for a future fixing date T_1 and expiry T_2, so that $t_0 < T_1 < T_2$, which has payoff $\left(\frac{S_T}{S_{T_1}} - K\right)^+$. If we knew the implied volatility surface at time T_1 we would know the value of this option at time T_1 and then could simply obtain it's current value by discounting. Clearly then the value of the option depends on the dynamics of the volatility surface over time.
Under the Black-Scholes framework (2.1.2) the price of this product can by obtained by making use of the tower property:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_{t_0}^{T_2} r_s\, ds}\left(\frac{S_T}{S_{T_1}} - K\right)^+\right] &= \mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_{t_0}^{T_1} r_s\, ds}\mathbb{E}^{\mathbb{Q}}\left[\mathrm{e}^{-\int_{T_1}^{T_2} r_s\, ds}\left(\frac{S_T}{S_{T_1}} - K\right)^+ |\mathcal{F}_{T_1}\right]\right] \\ &= \mathrm{e}^{-\int_{t_0}^{T_1} r_s\, ds} C_{\mathrm{BS}}(1, K, T_2 - T_1, \sigma).\end{aligned}$$

Given a price $C_{\mathrm{F}}(K, T_1, T_2)$ for such an option the forward implied volatility $\sigma_{\mathrm{imp}}(T_1 \,|\, K, T_2)$ is then defined as the value so that

$$\mathrm{e}^{\int_{t_0}^{T_1} r_s\, ds} C_{\mathrm{F}}(K, T_1, T_2) = C_{\mathrm{BS}}(1, K, T_2 - T_1, \sigma_{\mathrm{imp}}(T_1 \,|\, K, T_2)). \tag{2.3.1}$$

The relation (2.3.1) then allows us to study the dynamics of the volatility surface for a given model by computing the prices of forward starting options C_{F} under this model and backing out the forward implied volatility.
For the local volatility model it can then be observed that the forward volatility surface $\sigma^{\mathrm{LV}}_{\mathrm{imp}}(T_1 \,|\, \cdot, \cdot)$ flattens out as $T_1 \to \infty$.
This happens since as shown in [15, Corollary 3.7, page 240] and [15, Theorem 5.1, page 243] the implied volatility surface under any kind of martingale model necessarily flattens out in the sense that

$$\lim_{T\to\infty} \sup_{k_1,k_2\in[-M,M]} |\sigma_{imp}(k,T) - \sigma_{imp}(k,T)| = 0 \qquad \text{for all } M \in \mathbb{R},$$

and assuming the derivative with respect to k exists

$$\lim_{T\to\infty} \sup_{k\in[-M,M]} \left|\frac{\partial \sigma_{imp}(k,T)}{\partial k}\right| < 4$$

where the the spatial variable in σ_{imp} was changed to log-moneyness. It can then be inferred from Dupire's equation (2.2.5) that the local volatility skew $\frac{\partial \sigma_{loc}}{\partial k}$ flattens in the same sense, which in turn causes the forward implied volatility to flatten out too, see [12, Chapter 8, pages 102-103] for details. Additionally we will give examples in 6.2 where we compare the forward smile generated under local volatility to those of stochastic local volatility models.
This is in contrast to reality where it is observed that future implied volatility surfaces are on average more skewed than what the local volatility model predicts, meaning the local volatility model systematically miss prices options with forward starting features.

Chapter 3

Stochastic Volatility Models

The shortcomings of the local volatility model discussed in the previous section lead us to consider models in which the instantaneous volatility is modeled as a stochastic process driven by its own Brownian motion.
A general one factor stochastic volatility model under risk neutral measure takes the form

$$\begin{cases} dS_t = (r_t - q_t)S_t\,dt + \sigma_t S_t\,dW_t \\ d\sigma_t = \mu(t,\sigma_t) + \lambda(\sigma_t,t)\,dB_t \end{cases}, \tag{3.0.1}$$

where W, B are Brownian motions with covariation $[W,B]_t = \rho t$ for some $|\rho| \leq 1$. Let us remark that for the existence of a unique strong solution to the above SDE it is sufficient to establish it for the volatility process since then we can formally solve for the generalised geometric Brownian motion for S as

$$S_t = S_0 \exp\left(\int_0^t (r_s - q_s) - \frac{1}{2}\sigma_s^2\,ds + \int_0^t \sigma_s\,dW_s\right)$$

which can be verified to hold indeed by considering the stopping time $\tau_n := \inf\{t \geq 0 \,|\, \sigma_t > n\}$. Up to this stopping time σ_t is bounded and the stopped formal solution is indeed a solution to the SDE for the stopped process S^{τ_n}. Finally $\lim_{n\to\infty} S_t^{\tau_n} \to S_t$ almost surely.
However as noted in [16], to which we refer for a rigorous discussion of this matter, additional constraints on ρ and the coefficients of the volatility process are needed to ensure among others martingality and existence of moments for S.
We will now introduce some specific choices for the volatility process σ and highlight some of their features that will become important in the later part of this thesis where these processes will play the role of the stochastic component of the local stochastic volatility models.

3.1 The Scott-Chesney Model

In [17, Equation 7, page 426] Scott proposes to model the volatility as an exponential Ornstein-Uhlenbeck process, which is further developed with Chesney in [18]. For our purposes we let the mean reversion level be time dependent, essentially taking the well known one factor Hull-White interest rate model as the log-volatility process. The stochastic volatility model then takes the form

$$\begin{cases} dS_t &= (r_t - q_t)S_t\,dt + \sigma_t S_t\,dW_t \\ \sigma_t &= \exp(Y_t) \\ dY_t &= \kappa(\theta_t - Y_t)\,dt + \lambda\,dB_t \end{cases} \tag{3.1.1}$$

where $\kappa, \lambda > 0$ and θ is a bounded function with domain $[0,\infty)$. The existence of a unique strong solution for the volatility process follows from Theorem 1.0.4, moreover since the SDE for the log volatility is linear we can solve it analytically and obtain

$$Y_t = Y_0 \mathrm{e}^{-\kappa t} + \int_0^t \kappa\theta_s \mathrm{e}^{\kappa(s-t)}\,ds + \int_0^t \lambda \mathrm{e}^{\kappa(s-t)}\,dW_s. \tag{3.1.2}$$

Since the integral against Brownian motion of a deterministic function is normally distributed it follows from (3.1.2) that

$$Y_t \sim \mathcal{N}\left(Y_0 \mathrm{e}^{-\kappa t} + \int_0^t \kappa\theta_s \mathrm{e}^{\kappa(s-t)}\, ds,\ \frac{\lambda^2}{2\kappa}(1-\mathrm{e}^{-2\kappa t})\right). \tag{3.1.3}$$

This provides us with an interpretation of the parameters θ and κ : For constant $\theta > 0$ as $t \to \infty$ Y_t converges in distribution to a normal random variable with mean θ and variance $\frac{\lambda^2}{2\kappa}$. Thus θ characterises the mean reversion level and κ the speed of mean reversion of the log-volatility.
Summing up under the Scott-Chesney model, the volatility process is mean reverting and strictly positive, both desirable features agreeing with empirical observations. Furthermore the instantaneous volatility and thus also the instantaneous variance follows a log-normal distribution, which in [19, Figure 4, page 12 , Section 5, page 14] was found to be a good approximation of real realised variance behavior.
A drawback is that no closed or semi closed form formulas for European Vanilla options exists under this model, which makes calibrating the parameters to market prices by optimisation expensive forcing one to rely on estimation, which can sometimes appear to be more of an art than a science, given that instantaneous volatility is not directly observable.
Another problem is that the fat tail of the volatility distribution can cause infinite moments of the stock price, which is problematic for pricing applications.

3.2 The Mean reverting Log-Normal Model

Another popular stochastic volatility model is the so called mean reverting log-normal volatility model, also known as the inverse Gamma model, with dynamics

$$\begin{cases} dS_t &= (r_t - q_t)S_t\, dt + \sigma_t S_t\, dW_t \\ d\sigma_t &= \kappa(\theta - \sigma_t)\, dt + \lambda\sigma_t\, dB_t \end{cases} \tag{3.2.1}$$

where $\kappa, \theta, \lambda > 0$. The name log-normal here refers to the volatility being a log-normal diffusion meaning $[\sigma, \sigma]_t = \int_0^t \lambda^2\sigma_s^2\, ds$.
Existence and uniqueness of the volatility process in (3.2.1) follows from Theorem 1.0.4 and we can solve the linear SDE for the volatility as

$$\sigma_t = \sigma_0 \mathrm{e}^{-(k+\frac{1}{2}\lambda^2)t+\lambda B_t} + \kappa\theta \int_0^t \mathrm{e}^{-(\kappa+\frac{1}{2}\lambda^2)(t-s)+\lambda(B_t-B_s)}\, ds.$$

From the solution it immediately follows that for initial value $\sigma_0 > 0$ the volatility process will be strictly positive. To verify that the parameters κ, θ can be interpreted as mean reversion speed and level like in the Scott-Chesney model we calculate

$$\begin{aligned} \mathbb{E}\left[\sigma_t\right] &= \sigma_0 \mathbb{E}\left[\mathrm{e}^{-(\kappa+\frac{1}{2}\lambda^2)t+\lambda B_t}\right] + \kappa\theta\mathbb{E}\left[\int_0^t \mathrm{e}^{-(\kappa+\frac{1}{2}\lambda^2)(t-s)+\lambda(B_t-B_s)}\, ds\right] \\ &= \sigma_0 \mathrm{e}^{-\kappa t} + \kappa\theta\mathbb{E}\left[\int_0^t \mathrm{e}^{-(\kappa+\frac{1}{2}\lambda^2)(t-s)+\lambda(B_t-B_s)}\, ds\right]. \end{aligned}$$

Since the integrand in the second term is strictly positive by Fubini-Tonelli we can exchange expectation and the integral. Then

$$\mathbb{E}\left[\mathrm{e}^{-(\kappa+\frac{1}{2}\lambda^2)(t-s)+\lambda(B_t-B_s)}\right] = \mathrm{e}^{-\kappa(t-s)}.$$

Substituting this back in yields

$$\mathbb{E}\left[\sigma_t\right] = \sigma_0 \mathrm{e}^{-\kappa t} + \kappa\theta \int_0^t \mathrm{e}^{-\kappa(t-s)}\, ds = \sigma_0 \mathrm{e}^{-\kappa t} + \theta(1-\mathrm{e}^{-\kappa t}) \tag{3.2.2}$$

and confirms our interpretation. Unlike for the Scott-Chesney model though the full distribution of V_t is not known, but the stationary distribution is known to be an inverse gamma distribution [20, Appendix A.2, page 22] which has probability density function

$$p_{\Gamma^{-1}}(x\,|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1}\mathrm{e}^{-\frac{\beta}{x}} \tag{3.2.3}$$

with $\beta = \frac{2\kappa\theta}{\lambda^2}$ and $\alpha = 1 + \frac{\beta}{\theta}$. While again we lack closed form formulas for pricing of vanillas, an efficient approximation for Europan Put options was derived in [20, Theorem 3.1, page 9] and thus the parameters can be derived by calibrating the model to market prices. On the other hand as an application of [16, Theorem 2.4, page 7] S is an integrable martingale only if $\rho \leq 0$ and S_t has moments up to $m \in \mathbb{N}$ only if $\rho < -\sqrt{\frac{m-1}{m}}$ [16, Theorem 2.5,2.6, page 9].
In summary the mean reverting log-normal model retains the good properties of the Scott-Chesney model, being mean reverting, strictly positive and through being a log-normal type diffusion model displaying realistic volatility dynamics, while on top of that having a more straightforward calibration procedure. However it suffers from the moment explosion problems faced by the Scott-Chesney model as well.

3.3 The Heston Model

Finally we briefly discuss the arguably most popular stochastic volatility model, which was introduced by Heston in [21, Equations 1,4, pages 328,329] where he assumes the dynamics

$$\begin{cases} dS_t = (r_t - q_t)S_t\,dt + \sigma_t S_t\,dW_t \\ \sigma_t = \sqrt{V_t} \\ dV_t = \kappa(\theta - V_t)\,dt + \lambda\sqrt{V_t}\,dB_t \end{cases} \tag{3.3.1}$$

where $\kappa, \theta, \lambda > 0$. Here existence and uniquness of the variance and thus also the volatility process follows from the Yamada-Watanabe condition 1.0.5 with $h = \sqrt{\cdot}$. An important result about the variance process, the so called square root process, is that it remains strictly positive with probability 1 if it satisfies the Feller condition

$$\frac{2\kappa\theta}{\lambda^2} \geq 1 \tag{3.3.2}$$

see [22, Chapter 6.31, pages 98,101-103]. For the Heston model to produce prices in line with the market this condition has to often be violated though so in general under the Heston model variance and thus also volatility are not strictly positive. The mean and variance of V_t are given by [23, Equation 3.23, page 66]

$$\begin{aligned} \mathbb{E}[V_t] &= V_0 \mathrm{e}^{-\kappa t} + \theta(1 - \mathrm{e}^{-\kappa t}) \\ \mathbb{V}(V_t) &= V_0 \frac{\lambda^2}{\kappa}(\mathrm{e}^{-\kappa t} - \mathrm{e}{-2\kappa t}) + \theta\frac{\lambda^2}{2\kappa}\left(1 - \mathrm{e}^{-\kappa t}\right)^2 \end{aligned}$$

from which it can be seen that also this time κ, θ may be interpreted as the mean reversion speed and level. Additionally the distribution is known to be

$$V_t \sim \frac{\chi^2(\frac{4\kappa\theta}{\lambda^2}, 2c_t V_0 \mathrm{e}^{-\kappa t})}{2c_t}, \tag{3.3.3}$$

where $\chi^2(\alpha, \beta)$ denotes a non central χ^2 distribution with α degrees of freedom and non centrality parameter β, and $c_t := \frac{2\kappa}{(1-\mathrm{e}^{-\kappa t})\lambda^2}$ [23, page 65].
The popularity of the Heston Model stems from the fact that a semi closed form formula, in terms of the characteristic function of S_t, for the price of European Call options was derived in [21, Equations 12-18, pages 330-331]. This allows one to, since European Vanilla prices can be computed very efficiently, obtain the parameters in (3.3.1) by minimising the difference between the model and the market prices, instead of estimating them from historical data.
In summary the Heston model is the most analytically tractable among the stochastic volatility models we introduced, and has the most simple calibration procedure, but the volatility may become 0 and in [24, Figure 6, page 3175, Section 6, page 3164] the non central χ^2 distribution of the variance process was found to provide a less good fit to market data than the previous two log-normal type diffusion models.

3.4 A Solution to the short comings of the Local Volatility model?

As discussed in the previous three sections stochastic volatility models can reproduce realistically the dynamical features observed in realised volatility, like mean reversion or distribution of changes. Adding to that they don't suffer from the forward smile flattening, in the same way as the local volatility model, see [25, Section 3.3, page 9] for a discussion of this for the Heston model, and Section 6.2 where we will give examples of forward smiles under stochastic local volatility. They are thus preferable for pricing and hedging of exotic options depending on volatility dynamics.
Still stochastic volatility models like the ones we presented here have a significant weakness in comparison to the local volatility model. Due to their parametric nature they can in general not fit the whole implied volatility surface perfectly. As vanilla options are frequently used in hedging of more exotic options it is highly undesirable that a model prices the vanilla options inconsistently with the market.

Chapter 4

The Stochastic Local Volatility Model

In view of the last two chapters for the pricing of exotic options we would like to have an option pricing model that can fit the current implied volatility surface perfectly and features realistic stochastic volatility dynamics, in essence we would like to have a mixture of the local and a stochastic volatility model. Risk neutral dynamics for such a model can be written as an extension of (3.0.1) by adding a deterministic local volatility component $L(\cdot,\cdot)$, the so called Leverage function. The SDE describing this mixture model, which from now on we shall call Stochastic Local volatility model, short SLV model or SLVM, then becomes

$$\begin{cases} dS_t = (r_t - q_t)S_t\,dt + L(S_t,t)\sigma_t S_t\,dW_t \\ d\sigma_t = \mu(\sigma_t,t)\,dt + \lambda(\sigma_t,t)\,dB_t \end{cases}. \tag{4.0.1}$$

Such a kind of model, at least in the context to trees was first introduced in [3, Equations 4-6, page 7]. Another early reference to a similar model, this time in a continuous time setting, can be found in [4, Equation 8, page 64] where Lipton introduces his universal volatility model which adds jumps to (4.0.1). A pure continuous time SLVM as stated here is first introduced and discussed in [5, Equations 1,2, page 138]. In our presentation of calibrating SLV models we will follow the main approach laid out in the latter and its extensions in subsequent literature.

4.1 General calibration of SLV models and well posedness

One of our stated goals with the SLV model is to capture the whole implied volatility surface perfectly. The question then arises how L should be choosen to accomplish this. Recalling that the implied volatility surface generated by a SLV model is completely determined by the 1-dimensional marginal distributions of (4.0.1) leads us to the following result.

Theorem 4.1.1 (Calibration of SLV models)**.** *Suppose the dynamics of the underlying are described by* (4.0.1) *then the SLV model matches the market implied volatilities perfectly if and only for all* $K,T>0$

$$L(K,T) = \frac{\sigma_{loc}(K,T)}{\sqrt{\mathbb{E}[\sigma_T^2 \mid S_T = K]}} \tag{4.1.1}$$

Proof. The proof is a direct application of the mimicking theorem 1.0.7. Since the Local volatility model produces the correct marginal distributions, we just need (4.0.1) to have the same 1-dimensional marginal distributions as (2.2.1). For this to hold by the mimicking theorem, which we can apply under the mild assumptions of the drift vector and diffusion matrix being integrable, using that σ_{loc} and L are deterministic functions of K and T

$$\begin{aligned} \sigma_{loc}(K,T) &= \sqrt{\mathbb{E}[L^2(S_T,T)\sigma_T^2 \mid S_T = K]} \\ &= L(K,T)\sqrt{\mathbb{E}[\sigma_T^2 \mid S_T = K]}. \end{aligned}$$

The only if follows from the local volatility being the unique one dimensional diffusion model that matches the market perfectly. □

We note that the conditional expectation in (4.1.1) depends on the joint distribution $\mathbb{P}_{X_T}$ of S_T and σ_T. Thus the SDE describing the dynamics of a calibrated SLV model takes the general form

$$dX_t = \tilde{\mu}(X_t, t, \mathbb{P}_{X_t})\, dt + \tilde{\lambda}(X_t, t, \mathbb{P}_{X_t})\, dZ_t$$

for a 2-dimension standard Brownian motion Z. Processes of this form were first studied in [26] and are known as McKean-Vlasov processes. Altough the existence and uniqueness theorem 1.0.4 can be extended to this setting [27, Proposition 1, page 902] the Lipschitz assumption does not hold for the specific form of McKean-Vlasov process we are working with so this theorem does not apply. To make the discussion more concrete we denote the joint transition density of S, σ as p and rewrite the conditional expectation as

$$\psi(K,T) := \mathbb{E}[\sigma_T^2 \mid S_T = K] = \frac{\int_0^\infty \sigma^2 p(K,\sigma,T)\, d\sigma}{\int_0^\infty p(K,\sigma,T)\, d\sigma}. \tag{4.1.2}$$

Then (4.0.1) becomes

$$\begin{cases} dS_t = (r_t - q_t)S_t\, dt + \frac{\sigma_{loc}(S_t,t)}{\sqrt{\psi(S_t,t)}}\sigma_t S_t\, dW_t \\ d\sigma_t = \mu(\sigma_t, t)\, dt + \lambda(\sigma_t, t)\, dB_t \end{cases}. \tag{4.1.3}$$

In [28, Proposition 3, page 29] it is shown that the existence of strong solutions to this Integro-SDE can be equivalently studied by considering the associated Fokker-Planck PDE,

$$\begin{aligned} \frac{\partial p}{\partial t}(x,y,t) = &-\frac{\partial}{\partial x}[(r_t - q_t)xp(x,y,t)] - \frac{\partial}{\partial y}[\mu(y,t)p(x,y,t)] \\ &+ \frac{1}{2}\frac{\partial^2}{\partial x^2}[\frac{\sigma_{loc}^2(x,t)}{\psi(x,t)}x^2 p(x,y,t)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[\mu(y,t)p(x,y,t)] \\ &+ \rho\frac{\partial^2}{\partial x \partial y}[\lambda(y,t)\frac{\sigma_{loc}(x,t)}{\sqrt{\psi(x,t)}}yxp(x,y,t)], \end{aligned} \tag{4.1.4}$$

since the existence of triple a (S, σ, p) that is a strong solution to (4.1.3) implies the existence of a solution to (4.1.4) and vice versa.
Using this correspondence Tachet derives an existence result on bounded domains $B \subset \mathbb{R}^2$ provided that for some coordinate transformation f and $(x,y) \in B$, $\sigma = f(y)$ and f is not too far from constant in B [28, Theorem 5, page 38]. Since for numerical solutions of (4.1.4) the domain needs to be truncated and singular initial conditions need to be regularised this result in most cases guarantees existence of solutions to the associated numerical problem [28, Proposition 6, page 55], at least when the volatility of volatility is not too large. More recently existence of a solution to (4.1.3) given that $\sigma_t = c + p_{LV}(S_t,t)g(S_t,Y_t,t)$, where p_{LV} denotes the density of the local volatility model, Y is a general diffusion process of form (1.0.1) and g is a bounded function so that $g(s,y,t)\sigma_{loc}(s,t)$ is Lipschitz in (x,y) uniformly in t was shown to hold [29, Section 2.3, page 7]. The SLV models we consider in this thesis though will be based on the stochastic volatility models introduced in chapter 3 where the volatility process does not meet the assumptions required in [29]. The existence or non-existence of solutions to (4.1.3) in unrestricted domains with σ given by some popular stochastic volatility model such as the ones presented in chapter 3 remains an open problem.

Turning back to the problem of calibration, we need to calculate L given by (4.1.1). We already discussed how to calculate σ_{loc} from market data so this can be seen as given. The main problem is to find the conditional expectation ψ (4.1.2).
For this two main approaches are discussed in the literature. The first one introduced in [5] works with the Fokker-Planck equation (4.1.4), solving it for the joint density to calculate ψ by its integral representation. The other, Monte Carlo based, approach introduced in [30] and [31] works directly with the SDE (4.1.3) simulating it one step at a time and then using the empirical distribution at the next step to calculate ψ. More recently also attempts have been made to calibrate LSV models using machine learning [32] and optimal transport techniques [33].
We will adopt the Fokker-Planck PDE approach since we work with one factor stochastic volatility, meaning that the volatility process is driven by a single Brownian motion. This leads to a two dimensional Fokker-Planck equation which can numerically be handled efficiently. For multi factor stochastic volatility processes or the addition of stochastic interest rates the PDE would become

too high dimensional to be handled efficiently. In this case Monte Carlo based techniques become necessary.

4.2 The Fokker-Planck Equation calibration Algorithm

To simplify the Fokker-Planck equation for our SLV model (4.0.1) we transform the stock process into log-moneyness coordinates, and furthermore represent σ as a function of a process Y which will help us when substituting concrete stochastic volatility dynamics into the general calibration framework we shall develop as well as staying consistent with our notation for the Fokker-Planack PDE. The SDE for the SLV model is now given by

$$\begin{cases} S_t = S_0 \mathrm{e}^{X_t} \\ dX_t = [(r_t - q_t) - \frac{1}{2}\tilde{L}(X_t,t)^2 f(Y_t)^2]\, dt + \tilde{L}(X_t,t) f(Y_t)\, dW_t \\ X_0 = 0 \\ \sigma_t = f(Y_t) \\ dY_t = \mu(Y_t,t)\, dt + \lambda(Y_t,t)\, dB_t \end{cases} \tag{4.2.1}$$

with $\tilde{L}(x,t) = L(\mathrm{e}^x S_0, t)$. By abuse of notation from now on we shall just write L instead of $\tilde{L}$. The Fokker-Planck equation (4.1.4), with p now denoting the joint transition density of X and Y, is then transformed to

$$\begin{cases} \frac{\partial p}{\partial t} & = -\frac{\partial}{\partial x}[(r_t - q_t) - \frac{1}{2}L^2 f(y)^2]p - \frac{\partial}{\partial y}[\mu(y,t)]p \\ & \quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}[L^2 f(y)^2]p + \frac{1}{2}\frac{\partial^2}{\partial y^2}[\lambda^2(y,t)]p + \rho\frac{\partial^2}{\partial x \partial y}[L\lambda(y,t)f(y)]p\,. \\ p(x,y,0) & = \delta_0(x)\delta_{Y_0}(y) \end{cases} \tag{4.2.2}$$

Here δ_u denotes the Dirac measure centered in $u \in \mathbb{R}$. Let us denote by $G \subset \mathbb{R}$ the set that Y takes values in.
The following algorithm is then proposed by Ren et al in [5, page 139] to bootstrap the Leverage function :

Algorithm 1: Leverage Function Bootstrap

Discretise (4.2.2) in time between 0 and T maturity as $\{t_0, t_1, \cdots, t_N\}$;
At $t_0 = 0$: $L(0,0) = \frac{\sigma_{loc}(S_0,0)}{f(Y_0)}$ and $p(x,y,t_0) = \delta_0(x)\delta_{Y_0}(y)$;
for $i \in \{1, \cdots, N\}$ **do**
  Numerically solve (4.2.2) using $L(\cdot, t_{i-1})$ and $p(\cdot,\cdot,t_{i-1})$ to obtain $p(\cdot,\cdot,t_i)$;
  Set $\psi(x,t_i) = \frac{\int_G f(y)^2 p(x,y,t_i)\, dy}{\int_G p(x,y,t_i)\, dy}$;
  Set $L(x,t_i) = \frac{\sigma_{loc}(\mathrm{e}^x S_0, t_i)}{\sqrt{\psi(x,t_i)}}$;

Before diving into the numerical implementation of the calibration algorithm 1 let us first introduce the concrete SLV models we will work with for the remainder of this thesis.

4.3 The Madan-Qian-Ren Model

In [5, Equations 1,2, page 138] Ren et all introduce a SLV extension of the Scott-Chesney model (3.1.1), which we will refer to as the MQR model,

$$\begin{cases} dX_t = \left[(r_t - q_t) - \frac{1}{2}L^2(X_t,t)\mathrm{e}^{2Y_t}\right]\, dt + L(X_t,t)\mathrm{e}^{Y_t}\, dW_t \\ X_0 = 0 \\ dY_t = \kappa(\theta_t - Y_t)\, dt + \lambda\, dB_t \\ Y_0 = 0 \\ \theta_t := -\frac{\lambda^2}{2\kappa}(1 + \mathrm{e}^{-2\kappa t}) \end{cases} \tag{4.3.1}$$

with $\lambda, \kappa > 0$. Originally Ren et all assumed 0 correlation between the Brownian motions, but we will allow them to be correlated with instantaneous correlation $\rho \in [-1,1]$ as also done in [34, Page

2]. From (3.1.3) we can see that with the choice of θ made here $\mathbb{E}[\mathrm{e}^{2Y_t}] = 1$ for all $t \geq 0$. Then at least formally from (4.1.1) we can see that in general L should be similar to σ_{loc}. The model is thus in spirit constructed as a local volatility model with stochastic multiplier.
In Section 3.1 we commented on the Scott-Chesney model being prone to moment explosions of the stock process. For the MQR model this should be significantly mitigated due to leverage function deleveraging the volatility in the far wings.
The Fokker-Planck equation for the MQR model is given by

$$\begin{cases} \frac{\partial p}{\partial t} & = -\frac{\partial}{\partial x}[(r_t - q_t) - \frac{1}{2}L^2\mathrm{e}^{2y}]p - \frac{\partial}{\partial y}[\kappa(\theta_t - y)]p \\ & \quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}[L^2\mathrm{e}^{2y}]p + \frac{\lambda^2}{2}\frac{\partial^2}{\partial y^2}p + \rho\lambda\frac{\partial^2}{\partial x \partial y}[L\mathrm{e}^{y}]p \\ p(x,y,0) & = \delta_0(x)\delta_0(y) \end{cases} . \tag{4.3.2}$$

4.4 The Tataru-Fisher Model

In [35, CHapter 2, page 2] Tataru and Fisher introduce their SLV model based on the mean reverting log-normal model (3.2.1)

$$\begin{cases} dX_t = \left[(r_t - q_t) - \frac{1}{2}L^2(X_t,t)Y_t^2\right]\,dt + L(X_t,t)Y_t\,dW_t \\ X_0 = 0 \\ dY_t = (1 - Y_t)\,dt + \lambda Y_t\,dB_t \\ Y_0 = 1 \end{cases} . \tag{4.4.1}$$

From (3.2.2) it follows that $\mathbb{E}[Y_t] = 1$ for all $t \geq 0$, so similarly to the MQR model, the Tataru-Fisher model is constructed as a local volatility model with a stochastic multiplier, and we again expect L to be similarly shaped to σ_{loc}. Also similarly to the MQR model the moment explosion problem of the mean reverting log-normal model should be strongly mitigated by the presence of the leverage function.
The Fokker-Planck equation reads

$$\begin{cases} \frac{\partial p}{\partial t} & = -\frac{\partial}{\partial x}[(r_t - q_t) - \frac{1}{2}L^2y^2]p - \frac{\partial}{\partial y}[(1 - y)]p \\ & \quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}[L^2y^2]p + \frac{\lambda^2}{2}\frac{\partial^2}{\partial y^2}[y^2]p + \rho\lambda\frac{\partial^2}{\partial x \partial y}[Ly^2]p \\ p(x,y,0) & = \delta_0(x)\delta_1(y) \end{cases} \tag{4.4.2}$$

4.5 The Heston SLV Model

The Heston SLVM is the most studied one factor SLVM in the literature. Some early references studying its calibration using the Fokker-Planck PDE approach include [22] and [36],[37], who do not apply a transformation to the variance process in the Heston model. But as pointed out in [38, Slides 20-24] this can amplify issues with the conservation of the probability mass, which can lead to unstable or inaccurate calibrations as we will discover next chapter, that make it advisable to transform the variance process to log coordinates, essentially pushing the problematic boundary at 0 away to infinity. The dynamics of the Heston SLVM are then given as

$$\begin{cases} dX_t = \left[(r_t - q_t) - \frac{1}{2}L^2(X_t,t)\mathrm{e}^{Y_t}\right]\,dt + L(X_t,t)\mathrm{e}^{\frac{1}{2}Y_t}\,dW_t \\ X_0 = 0 \\ dY_t = \left[(\kappa\theta - \frac{1}{2}\lambda^2)\mathrm{e}^{-Y_t} - \kappa\right]\,dt + \lambda\mathrm{e}^{-\frac{1}{2}Y_t}\,dB_t \\ Y_0 = \log(V_0) \end{cases} , \tag{4.5.1}$$

for some initial variance $V_0 > 0$. Unlike the previous two models the Heston SLVM is not constructed in a way that for any parameters on average instantaneous volatility or variance are around 1. Thus we can not expect the leverage function to be similar to the local volatility. Another way to see this is that due to the on average very large value of the stochastic volatility in the Tataru-Fisher or MQR model without the leverage function these models would not be able to fit market quotes well at all and the largest part of the smile has to be explained by the leverage function. Since the stochastic volatility component for the Heston SLVM is the pure Heston model without any restricted parameters, which can capture the implied volatility surface reasonably well by itself,

we expect a larger part of the market smile to be explained by the stochastic volatility component. The Heston SLVM therefore should be interpreted as a stochastic volatility model with a local volatility correction term and not as a local volatility model with a stochastic multiplier like the other two models.
The Fokker Planck equation for (4.5.1) is given by the following:

$$\begin{cases} \frac{\partial p}{\partial t} & = -\frac{\partial}{\partial x}[(r_t - q_t) - \frac{1}{2}L^2 \mathrm{e}^y]p - \frac{\partial}{\partial y}[(\kappa\theta - \frac{1}{2}\lambda^2)\mathrm{e}^{-y} - \kappa]p \\ & \quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}[L^2\mathrm{e}^y]p + \frac{\lambda^2}{2}\frac{\partial^2}{\partial y^2}[\mathrm{e}^{-y}]p + \rho\lambda\frac{\partial^2}{\partial x \partial y}[L]p \\ p(x,y,0) & = \delta_0(x)\delta_{\log(V_0)}(y) \end{cases} . \tag{4.5.2}$$

4.6 A brief digression on mixing Weights

So far in our discussion we have assumed the correlation and volatility of volatility ρ, λ to be constant. However term structure may easily be added to the parameters by choosing some piecewise linear function $\eta : [0, \infty) \to [0, 1]$ and replacing λ and ρ by $\eta(t)\lambda$ and $\eta(t)\rho$ in the dynamics of the SLVMs respectively.
Choosing $\eta(t) = 0$ for all $t \geq 0$ lets the SLV model degenerate to the local volatility model. On the other hand choosing $\eta(t) = 1$ for all $t \geq 0$ keeps the parameters as estimated or fitted and in that sense preserves the full market observed stochastic volatility dynamics. Hence η can be seen as tuning between stochastic and local volatility in the SLV model, and is therefore referred to as mixing weight or mixing fraction. In [22, Footnote 5, page 125] Clark suggests that, at least for Foreign Exchange markets η should typically be chosen between 0.6 and 0.65. Tatataru and Fisher [35, Section 8, pages 10-11] and Tian et al [36, Chapter 5.1.1, page 57] suggest instead to use η to calibrate the SLV model to some chosen liquid exotic options or observable market behaviors by calibrating to the implied volatility surface with different choices of η and then choosing the one with the smallest pricing difference to the target exotics or the one that reproduces the desired behavior best.
Since piece wise linear parameters can easily be handled by the numerical schemes we will employ for calibration of the leverage function for notational ease we will restrict our self to constant λ, ρ and η. In this case η then simply gets absorbed into λ and ρ and we will therefore omit mixing weights in our further presentation.

Chapter 5

Numerical Implementation

For the implementation of calibration Algorithm 1 for the SLV models introduced in Sections 4.3 to 4.5 we will now establish how to numerically solve equations of the form (4.2.2) in general and give specific details for the solution of equations (4.3.2),(4.4.2) and (4.5.2) in particular. The first choice we will have to make, as this will influence the spatial discretisation for our PDE, is what kind of numerical scheme to employ. Ren et al [5] and later Tachet [28, Chapter 9, pages 73-79], Clark [22, Section 6.8] and Tian [36, Section 5.2, pages 64-72] propose to use finite difference methods for the spatial approximation of (4.2.2), in [39] and [40] the problem is solved using using finite volume methods, and Cozzi [41] considers finite element methods.
Comparing the different sources the finite difference and volume approaches seem to offer the best results when it comes to quality of calibration.
Since additionally the finite difference approach lends itself well to handling different one factor stochastic volatility models with only slight adjustments, it is the approach we choose.

5.1 Spatial Discretisation and derivative Approximation

For the execution of a finite difference scheme the spatial domain of our PDE needs to be truncated to a rectangular domain $\Omega := [x_{min}, x_{max}] \times [y_{min}, y_{max}]$ with $(0, Y_0) \in \Omega$ and further discretised into a grid

$$\mathfrak{G}_{x,y} := \{x_{min} = x_0, x_1, \cdots, x_{N-1}, x_N = x_{max}\} \times \{y_{min} = y_0, y_1, \cdots, y_{M-1}, y_M = y_{max}\}$$

where $N, M \in \mathbb{N}$ and for $1 \leq i \leq N$ and $1 \leq j \leq M$ we have that $x_{i-1} < x_i$ and $y_{j-1} < y_j$.

5.1.1 Spatial Grid Generation

As a first step in our calibration procedure the value of the grid boundary points $x_{min}, x_{max}, y_{min}, y_{max}$ needs to be determined. Since the solution to (4.2.2) is a probability density the integral of p over its domain has to be 1 at all times $t \geq 0$. Our truncated domain should then be chosen so that for T the final maturity we calibrate to

$$\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} p(x, y, t)\, dydx \approx 1 \qquad \text{for all } 0 \leq t \leq T. \tag{5.1.1}$$

This is especially crucial since, as we will further explore when talking about boundary conditions, too large losses in the total probability mass will lead to the calibration procedure becoming unstable, producing highly inaccurate results and eventually breaking. The problematic part about enforcing (5.1.1) is that, p being the solution of the problem we want to solve, we lack prior information about p. We resolve this issue by considering the x and y directions separately:
Since the marginal distribution of the calibrated stock price process should agree with the one implied from the market, we can back out the market implied cdf, see Figure 5.1 , from the implied volatility surface, and then choose x_{min}, x_{max} according to the quantiles of this distribution. This is done as follows: Denoting the discount factor between $0 \leq t \leq T$ as

$$D(t, T) := \mathrm{e}^{-\int_0^T r_s\, ds}$$

we can write the market implied cdf of S_T as

$$\begin{aligned} F^{\text{imp}}_{S_T}(K) :&= \mathbb{Q}(S_T \leq K) \\ &= 1 - \int_K^\infty p_{\text{imp}}(s,T)\, ds \\ &= 1 + \frac{1}{D(0,T)} \partial_K C_{\text{Mkt}}(K,T) \end{aligned}$$

where for the last equality we used (2.2.3), $\mathbb{Q}$ denotes the risk neutral measure and p_{imp} the market implied risk neutral density. We then find K_{min}, K_{max} so that $F^{\text{imp}}_{S_T}(K_{min}) \leq 10^{-h}$ and $F^{\text{imp}}_{S_T}(K_{max}) \geq 1 - 10^{-h}$, where we suggest to choose h between 4 and 6, and set $x_{min} := \log(\frac{K_{min}}{S_0})$ and $x_{max} := \log(\frac{K_{max}}{S_0})$.

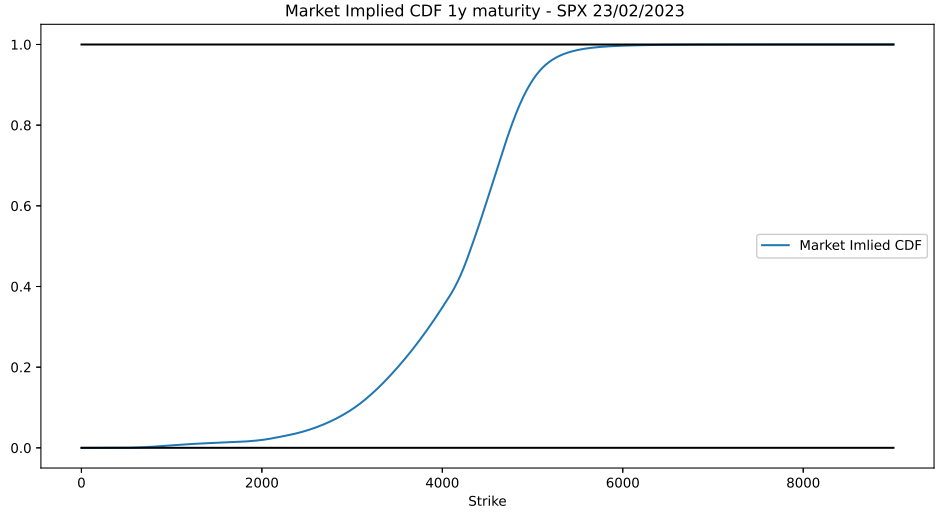
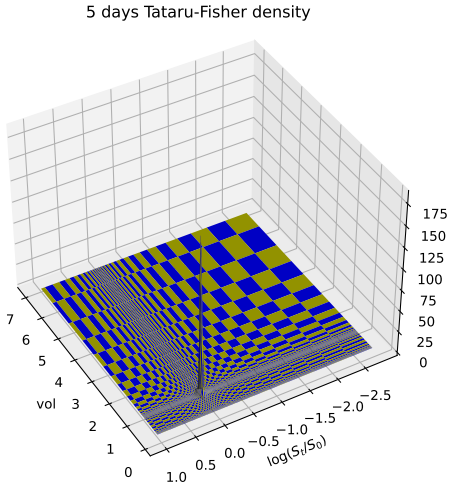


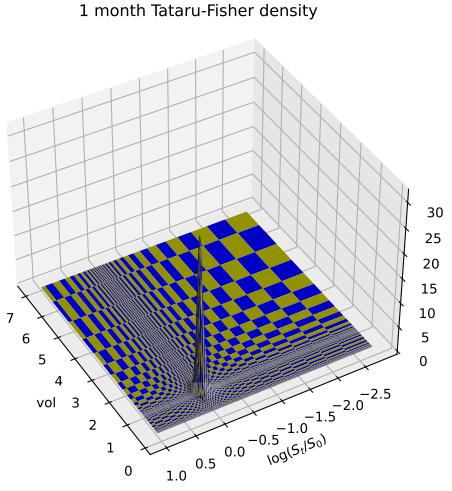
Figure 5.1: SPX market implied cdf on 2023-02-23 for 2024-02-23.

If the distribution of Y_T is known, which is the case for the MQR (3.1.3) and Heston SLVM (3.3.3), we choose y_{min} and y_{max} in accordance with x_{min} and x_{max} as quantiles of the distribution of Y_T.
If we do not know the distribution of Y_T exactly then more case specific methods have to be applied. For the Tataru-Fisher model, we choose y_{min}, y_{max} using the stationary distribution of the volatility process (3.2.3).

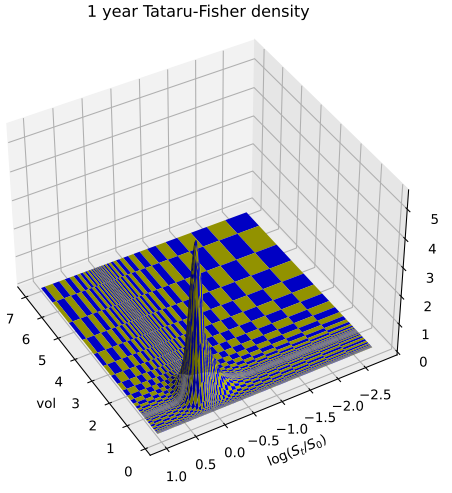
With the boundary of the grid set, the next step is to determine the location of the inner grid points. Due to the highly singular Dirac initial condition of (4.2.2), which can cause major problems for finite difference schemes due to the unbounded derivative, it is important to have a very tight grid spacing around the initial point $(0, Y_0)$ for a stable calibration. Additionally for volatility models with attractive boundaries at 0 such as the Heston SLVM when the Feller condition (3.3.2) is violated or where our truncated boundary point is close to the true boundary, as is the case with the Tataru-Fisher model, where we choose, depending on λ, y_{min} close to 0, it is important to also have a high concentration of grid points on this lower boundary. Towards the other boundaries p should be very small in general and less tight grid spacing is needed. Figure 5.2 displays this behavior of the density in case of the Tataru-Fisher model, similar plots for the MQR and Heston SLV models may be found in Appendix B.1 and B.2 .
To efficiently accomplish this tight grid spacing around some critical points, which we have found to be crucial for a stable calibration, Clark [22, page 143] recommends the use of non uniform grids. In Figure 5.3 below we show examples of the kind of grids we used for each model's calibration, with a reduced number of points for visibility.
For generating these grids we employ a procedure outlined in [36, pages 65-67]. Let us exemplify the grid construction for the y-direction of a grid like in Subfigure 5.3(b). To generate a grid with



(a) 5 day density Tataru-Fisher.

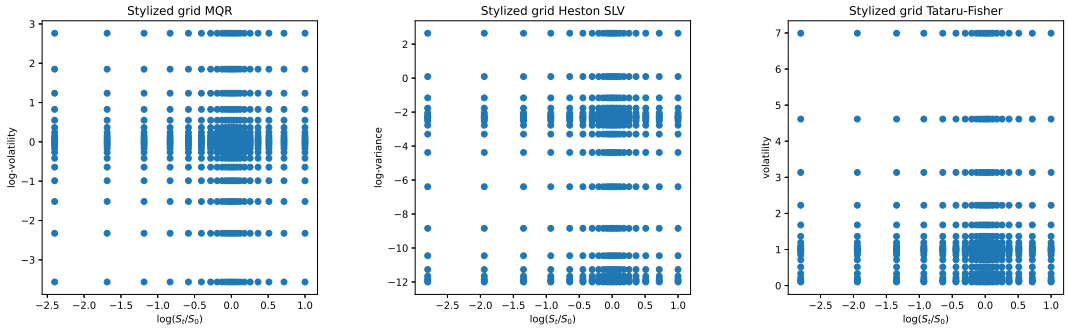


(b) 1 month density Tataru-Fisher.



(c) 1 year density Tataru-Fisher.

Figure 5.2: Tataru-Fisher density at different maturities for SPX fit, $\lambda = 0.42$, $\rho = -0.23$



(a) Stylized Grid MQR. (b) Stylized Grid Heston SLV. (c) Stylized Grid Tataru Fisher.

Figure 5.3: Stylized grids for 1 year Calibration, 20 points in each direction

$M+1$ points, including the endpoints, between y_{min}, y_{max} we first generate an uniform grid U with $M+1$ points between $[0,1]$. We then set $y_j := F(u_j)$ for $u_j \in U$, $0 \leq j \leq M$, where the coordinate transformation F is a solution to the ODE

$$\begin{cases} \frac{dF(u)}{du} = A(\left(\sum_{k=1}^{2} \frac{1}{J_k(u)}\right)^{-\frac{1}{2}} \\ J_k(u) := [\beta^2 + (F(u) - B_k)^2] \\ F(0) = y_{min} \\ F(1) = y_{max} \end{cases} \tag{5.1.2}$$

with $B_1 = Y_0$ and $B_2 = y_{min}$ and $\beta = \frac{y_{max} - y_{min}}{\tilde{U}}$ for $\tilde{U} > 0$ which controls how non uniform the grid is, larger being more non uniform. A is a constant that is adjusted so that $|F(u_M) - y_{max}| < \varepsilon$ for some tolerance level $\varepsilon > 0$. In the case of only one concentration point the ODE has an analytical solution, but for two or more concentration points it needs to be solved numerically.
We give details of the analytical solution in the case of one concentration point and the numerical procedure to solve the ODE for two concentration points in Appendix A.1.

In log-moneyness direction x we always use a non uniform grid concentrated at 0 and in volatility direction y we use a grid with one concentration point at Y_0 for the MQR model, and concentration points at the lower boundary and Y_0 for the Tataru-Fisher and Heston SLV models.

5.1.2 Spatial Derivative Approximation

We first establish how to approximate derivatives up to second order on a non uniform two dimensional grid $\mathfrak{G}_{x,y}$ using finite differences.

Theorem 5.1.1 (Approximation of Derivatives)**.** *Let g be a smooth function with domain Ω. Then for a grid $\mathfrak{G}_{x,y}$ as defined previously and $0 \leq i \leq N-1$, $0 \leq j \leq M-1$, we define $\Delta x_i := x_{i+1} - x_i$, and $\Delta y_j := y_{j+1} - y_j$, and for $0 \leq i \leq N$, $0 \leq j \leq M$ we define $g_{i,j} := g(x_i, y_j)$. The derivatives of g at the inner grid points (x_i, y_j) with $1 \leq i \leq N-1$, $1 \leq j \leq M-1$ may then be approximated by the following:*

$$\begin{aligned}
\frac{\partial}{\partial x} g_{i,j} &= -\frac{\Delta x_i}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} g_{i-1,j} + \frac{\Delta x_i - \Delta x_{i-1}}{\Delta x_i \Delta x_{i-1}} g_{i,j} + \frac{\Delta x_{i-1}}{\Delta x_i(\Delta x_i + \Delta x_{i-1})} g_{i+1,j} + \mathcal{O}((\Delta x)^2), \\
\frac{\partial^2}{\partial x^2} g_{i,j} &= \frac{2}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} g_{i-1,j} - \frac{2}{\Delta x_i \Delta x_{i-1}} g_{i,j} + \frac{2}{\Delta x_i(\Delta x_i + \Delta x_{i-1})} g_{i+1,j} + \mathcal{O}(\Delta x), \\
\frac{\partial^2}{\partial x \partial y} g_{i,j} &= \frac{g_{i+1,j+1} - g_{i+1,j-1} - g_{i-1,j+1} + g_{i-1,j-1}}{\Delta x_i \Delta y_j + \Delta x_i \Delta y_{j-1} + \Delta x_{i-1} \Delta x_j + \Delta x_{i-1} \Delta y_{j-1}} + \mathcal{O}(\Delta x + \Delta y).
\end{aligned}$$

And similarly for derivatives with respect to y. Here Δx and Δy refer to $\max\{\Delta x_0, \cdots, \Delta x_{N-1}\}$ and $\max\{\Delta y_0, \cdots, \Delta y_{M-1}\}$ respectively.

Proof. The proof is an easy application of Taylor's Theorem and can be found in Appendix A.2. □

From the theorem we may see that the spatial local truncation error of our scheme will be $\mathcal{O}(\max\{\Delta x, \Delta y\})$. When $\Delta x_i = \Delta x_{i-1}$ and $\Delta y_j = \Delta y_{j-1}$ we recover the familiar central difference derivative approximations from 5.1.1 and the local truncation error becomes $\mathcal{O}(\max\{(\Delta x)^2, (\Delta y)^2\})$. Still we found non-uniform grids to outperform uniform grids due to the higher concentration of grid points in regions where the density displays high curvature. See 5.3 where this is especially visible for short maturities when the probability density will be very concentrated around its origin.

Maturity	02/03/2023	26/03/2023	27/05/2023	25/08/2023	23/12/2023	23/02/2024
(a) Model Spot	4005.41	4014.92	4016.53	4017.64	4018.51	4018.195
h Mean Error σ_{imp}(bps)	63.4	46.2	18.4	8.1	5.4	1.8
Median Error σ_{imp}(bps)	42.5	35.2	10.7	3.1	2.9	3.1
Max Error σ_{imp}(bps)	181.9	118.7	78.7	38.9	15.6	13.8

Table 5.1: Uniform Grid

Maturity	02/03/2023	26/03/2023	27/05/2023	25/08/2023	23/12/2023	23/02/2024
(a) Model Spots	4012.23	4012.36	4012.33	4012.34	4012.36	4012.32
h Mean Error σ_{imp}(bps)	15.1	19	6.8	4.6	2.6	1.7
Median Error σ_{imp}(bps)	12.6	18.9	4.3	3	1.7	0.9
Max Error σ_{imp}(bps)	43.4	34	22.1	14.3	10.4	7.8

Table 5.2: Non uniform grid

Table 5.3: Model Spot and error in implied volatility from SLV to local volatility in bps ($10000 \cdot |\sigma_{\text{imp}}^{\text{LV}} - \sigma_{\text{imp}}^{\text{SLV}}|$) for the MQR model calibrated to SPX up to 1 year maturity using a uniform and non uniform space grid respectively. MQR prices calculated using the density. $N = 400$, $M = 325$, 365 time steps and parameters are the same. Real spot is 4012.32. For all maturities except 02/03/2023 the implied volatilities are calculated between 0.5 and 1.5 moneyness for 02/03/2023 the strike range is 0.85 to 1.1 moneyness

We now apply Theorem 5.1.1 to approximate the spatial derivatives in the Fokker-Planck equation (4.2.2). In view of the kind of time stepping schemes we will employ note that we can write the general calibration Fokker-Planck equation (4.2.2), representing all spatial derivative terms by the differential operator $\mathfrak{L}$, as

$$\frac{\partial}{\partial t} p = \mathfrak{L} p = (\mathfrak{L}_x + \mathfrak{L}_y + \mathfrak{L}_{xy}) p,$$

by splitting $\mathfrak{L}$ into 3 operators who represent the x, y and mixed derivative terms respectively. Letting now $\mathcal{L}$ be the discretised finite difference version of $\mathfrak{L}$ we approximate the Fokker-Planck PDE at $p(x_i, y_j, t) := p_{i,j}^t$ in the interior of our grid $\mathfrak{G}_{x,y}$ using 5.1.1 and obtain a semi discrete system of ODEs: For $1 \leq i \leq N-1$ and $1 \leq j \leq M-1$,

$$\frac{\partial}{\partial t} p_{i,j}^t \approx \mathcal{L}_x p_{i,j}^t + \mathcal{L}_y p_{i,j}^t + \mathcal{L}_{xy} p_{i,j}^t. \tag{5.1.3}$$

Let $L_{i,t} := L(x_i, t)$, $\mu(y_j, t) := \mu_{j,t}$ and $\lambda(y_j, t) := \lambda_{j,t}$ then the discrete finite difference operators in (5.1.3) are given explicitly by:

$$\begin{aligned}
\mathcal{L}_x p_{i,j}^t = &- \frac{\Delta x_i [-(r_t - q_t) + \frac{1}{2} L_{i-1,t}^2 f^2(y_j)] p_{i-1,j}^t}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} + \frac{(\Delta x_i - \Delta x_{i-1})[-(r_t - q_t) + \frac{1}{2} L_{i,t}^2 f^2(y_j)] p_{i,j}^t}{\Delta x_i \Delta x_{i-1}} \\
&+ \frac{\Delta x_{i-1} [-(r_t - q_t) + \frac{1}{2} L_{i+1,t}^2 f^2(y_j)] p_{i+1,j}^t}{\Delta x_i (\Delta x_i + \Delta x_{i-1})} \\
&+ \frac{f(y_j)^2}{2} \left(\frac{2 L_{i-1,t}^2 p_{i-1,j}^t}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} - \frac{2 L_{i,t}^2 p_{i,j}^t}{\Delta x_i \Delta x_{i-1}} + \frac{2 L_{i+1,t}^2 p_{i+1,j}^t}{\Delta x_i (\Delta x_i + \Delta x_{i-1})} \right),
\end{aligned} \tag{5.1.4}$$

$$\begin{aligned}
\mathcal{L}_y p_{i,j}^t = &- \frac{\Delta y_j [-\mu(y_j, t)] p_{i,j-1}^t}{\Delta y_{j-1}(\Delta y_j + \Delta y_{j-1})} + \frac{(\Delta y_j - \Delta y_{j-1})[-\mu(y_j, t)] p_{i,j}^t}{\Delta y_j \Delta y_{j-1}} + \frac{\Delta y_{j-1} [-\mu(y_{j+1}, t)] p_{i,j+1}^t}{\Delta y_j (\Delta y_j + \Delta y_{j-1})} \\
&+ \frac{1}{2} \left(\frac{2\lambda^2(y_{j-1}, t) p_{i,j-1}^t}{\Delta y_{j-1}(\Delta y_j + \Delta y_{j-1})} - \frac{2\lambda^2(y_j, t) p_{i,j}^t}{\Delta y_j \Delta y_{j-1}} + \frac{2\lambda^2(y_{j+1}, t) p_{i,j+1}^t}{\Delta y_j (\Delta y_j + \Delta y_{j-1})} \right),
\end{aligned} \tag{5.1.5}$$

$$\mathcal{L}_{xy}p^t_{i,j} = \rho\Bigg(\frac{L_{i+1,t}\lambda_{j+1,t}f(y_{j+1})p^t_{i+1,j+1} - L_{i+1,t}\lambda_{j-1,t}f(y_{j-1})p^t_{i+1,j-1}}{\Delta x_i\Delta y_j + \Delta x_i\Delta y_{j-1} + \Delta x_{i-1}\Delta x_j + \Delta x_{i-1}\Delta y_{j-1}} + \frac{-L_{i-1,t}\lambda_{j+1,t}f(y_{j+1})p^t_{i-1,j+1} + L_{i-1,t}\lambda_{j-1,t}f(y_{j-1})p^t_{i-1,j-1}}{\Delta x_i\Delta y_j + \Delta x_i\Delta y_{j-1} + \Delta x_{i-1}\Delta x_j + \Delta x_{i-1}\Delta y_{j-1}}\Bigg). \tag{5.1.6}$$

Choosing f as exp, $\mu(y,t) = \kappa(\theta_t - y)$ and $\lambda(y,t) \equiv \lambda$ constant in (5.1.4) to (5.1.6) with $\kappa, \theta_t, \lambda$ as in (4.3.1) yields the finite difference formulation of (4.3.2) the MQR model. For $f = \mathrm{Id}$ the identity, $\mu(y,t) = 1 - y$ and $\lambda(y,t) = \lambda y$ for a positive constant λ we get the Tataru-Fisher , and for $f(y) = \exp(\frac{1}{2}y_t)$, $\mu(y,t) = (\kappa\theta - \frac{1}{2}\lambda^2)\exp(-y) - \kappa$ and $\lambda(y,t) = \lambda\exp(-\frac{1}{2}y)$ the Heston SLVM. For practical implementation purposes it will be beneficial to rewrite the action of $\mathcal{L}_x$ and $\mathcal{L}_y$ on p as matrix vector operations. We define $P^t \in \mathbb{R}^{N+1\times M+1}$ with $P^t_{i,j} = p^t_{i,j}$, $0 \le i \le N$, $0 \le j \le M$. Furthermore let us denote by $P^t(j)$ the j^{th} column of P^t.
Finally we define the family of tridiagonal matrices $(\mathcal{D}^{j,t}_x)_{0\le j\le M} \subset \mathbb{R}^{N+1\times N+1}$ by

$$\mathcal{D}^{j,t}_x := \begin{pmatrix} b^{j,t}_0 & c^{j,t}_0 & 0 & \dots & \dots & 0 \\ a^{j,t}_1 & b^{j,t}_1 & c^{j,t}_1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & 0 & a^{j,t}_{N-1} & b^{j,t}_{N-1} & c^{j,t}_{N-1} \\ 0 & \dots & \dots & 0 & a^{j,t}_N & b^{j,t}_N \end{pmatrix}, \tag{5.1.7}$$

where for $1 \le i \le N-1$

$$\begin{aligned} a^{j,t}_i &:= -\frac{\Delta x_i[-(r_t - q_t) + \frac{1}{2}L^2_{i-1,t}f^2(y_j)]}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} + \frac{f(y_j)^2L^2_{i-1,t}}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})}, \\ b^{j,t}_i &:= \frac{(\Delta x_i - \Delta x_{i-1})[-(r_t - q_t) + \frac{1}{2}L^2_{i,t}f^2(y_j)]}{\Delta x_i\Delta x_{i-1}} - \frac{f(y_j)^2L^2_{i,t}}{\Delta x_i\Delta x_{i-1}}, \\ c^{j,t}_i &:= \frac{\Delta x_{i-1}[-(r_t - q_t) + \frac{1}{2}L^2_{i+1,t}f^2(y_j)]}{\Delta x_i(\Delta x_i + \Delta x_{i-1})} + \frac{f(y_j)^2L^2_{i+1,t}}{\Delta x_i(\Delta x_i + \Delta x_{i-1})}, \end{aligned}$$

and $b^j_0, c^j_0, a^j_N, b^j_N$ are determined by the boundary conditions discussed in the next section. Then

$$\mathcal{L}_xP^t(j) = \mathcal{D}^{j,t}_xP^t(j)$$

where $\mathcal{D}^{j,t}_x$ acts on $P^t(j)$ simply by matrix vector multiplication.
Similarly we define $\mathcal{D}^t_y \in \mathbb{R}^{M+1\times M+1}$ as

$$\mathcal{D}^t_y := \begin{pmatrix} \beta^t_0 & \gamma^t_0 & 0 & \dots & \dots & 0 \\ \alpha^t_1 & \beta^t_1 & \gamma^t_1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & 0 & \alpha^t_{M-1} & \beta^t_{M-1} & \gamma^t_{M-1} \\ 0 & \dots & \dots & 0 & \alpha^t_M & \beta^t_M \end{pmatrix}, \tag{5.1.8}$$

where again $\beta_0, \gamma_0, \alpha_M, \beta_M$ are determined by the boundary conditions and for $1 \le j \le M-1$

$$\begin{aligned} \alpha^t_j &:= -\frac{\Delta y_j[-\mu(y_j,t)]}{\Delta y_{j-1}(\Delta y_j + \Delta y_{j-1})} + \frac{\lambda^2(y_{j-1},t)}{\Delta y_{j-1}(\Delta y_j + \Delta y_{j-1})} \\ b^t_j &:= \frac{(\Delta y_j - \Delta y_{j-1})[-\mu(y_j,t)]}{\Delta y_j\Delta y_{j-1}} - \frac{\lambda^2(y_j,t)}{\Delta y_j\Delta y_{j-1}} \\ \gamma^t_j &:= \frac{\Delta y_{j-1}[-\mu(y_{j+1},t)]}{\Delta y_j(\Delta y_j + \Delta y_{j-1})} + \frac{\lambda^2(y_{j+1},t)}{\Delta y_j(\Delta y_j + \Delta y_{j-1})}. \end{aligned}$$

We can then represent the action of $\mathcal{L}_y$ on P^t as

$$\mathcal{L}_yP^t = (\mathcal{D}^t_y(P^t)^\top)^\top = P^t(D^t_y)^\top$$

5.1.3 Boundary Conditions

To complete the spatial discretisation we still need to determine appropriate boundary conditions for (4.2.2) on the truncated domain Ω and then deduce the still unspecified entries of (5.1.7) and (5.1.8) as well as the action of $\mathcal{L}_{xy}$ on $p^t_{i,j}$ for $(i,j) \in (\{0,N\}\times\{y_0,\cdots,y_M\}\cup\{x_0,\cdots,x_N\}\times\{0,M\})$ from them.
As pointed out previously in (5.1.1) it is crucial that the total probability mass in the truncated domain stays approximately equal to 1. Let us briefly illustrate why this is the case: Assuming for the moment 0 rates and dividends, the stock price process should be a martingale under the risk neutral measure that is we would expect that

$$S_0 = \mathbb{E}^{\mathbb{Q}}[S_T] \approx \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \mathrm{e}^x S_0 p(x,y,T)\, dydx.$$

However we observed for even somewhat small losses or gains in probability mass large losses or gains in the the spot price, calculated from the density by numerical integration, respectively. We give an example of this in Table 5.4 below.
So to keep the martingale property of the model and have a market consistent model forward it

Date	23/02/2023	02/03/2023	26/03/2023	27/05/2023	25/08/2023	23/02/2024
Probability mass	1	1	0.999929	0.998129	0.993299	0.981839
Model Spot	4012.32	4013.40	4012.04	4009.86	4003.66	3987.41

Table 5.4: Probability mass and model spot for the MQR model calibrated to SPX option data assuming 0 rates and dividends on 23/02/2023 with 0 second derivatives on boundary and 365 time steps, N, M = 400, 1 year calibration, uniform space and time grid. Probability is given to 6 decimal places.

is essential that the probability mass is properly conserved.
Problems like ours where we lack natural boundary conditions on the boundaries of our truncated domain Ω are so called open boundary problems. Classical approaches for these kind of problems are imposing so called Neumann or "natural" boundary conditions, which on a rectangular domain correspond to imposing that the first order or second order derivatives respectively vanish on their corresponding boundaries. These boundary conditions were found to be inadequate for preserving the probability mass [34, Section 5.1, pages 20,21]. For "natural" boundary conditions these findings were confirmed by our implementation, as shown in Table 5.4.
To remedy this problem of unstable mass we turn to a concept from finite volume methods, so called no flux boundary conditions that are the basis of finite volume schemes' conservatory nature. The idea is as follows:
If

$$\int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} p(x,y,t)\, dydx = 1 \qquad \text{for all } t \geq 0,$$

then

$$\partial_t \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} p(x,y,t)\, dydx = 0 \qquad \text{for all } t > 0.$$

Assuming now that we may exchange differentiation and integration we have

$$0 = \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \partial_t p(x,y,t)\, dydx = \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} (\mathfrak{L}_x + \mathfrak{L}_y + \mathfrak{L}_{xy}) p(x,y,t)\, dydx. \tag{5.1.9}$$

Splitting (5.1.9) and enforcing that each term is individually 0 yields

$$\begin{aligned}
0 &= \int_{y_{min}}^{y_{max}} \int_{x_{min}}^{x_{max}} \mathfrak{L}_x p(x,y,t)\,dxdy \\
&= \int_{y_{min}}^{y_{max}} \int_{x_{min}}^{x_{max}} -\frac{\partial}{\partial x}[(r_t - q_t) - \frac{1}{2}L(x,t)^2 f(y)^2]p(x,y,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}[L(x,t)^2 f(y)^2]p(x,y,t)\,dxdy \\
&= \int_{y_{min}}^{y_{max}} \left[\left(-(r_t - q_t) + \frac{1}{2}L^2(x,t)f(y)^2\right)p(x,y,t) + \frac{1}{2}f(y)^2\frac{\partial}{\partial x}[L^2(x,t)]p(x,y,t)\right]_{x_{min}}^{x_{max}} dy,
\end{aligned} \tag{5.1.10}$$

$$\begin{aligned}
0 &= \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \mathfrak{L}_y p(x,y,t)\,dydx \\
&= \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} -\frac{\partial}{\partial y}[\mu(y,t)]p(x,y,t) + \frac{1}{2}\frac{\partial^2}{\partial y^2}[\lambda(y,t)^2]p(x,y,t)\,dydx \\
&= \int_{x_{min}}^{x_{max}} \left[-\mu(y,t)p(x,y,t) + \frac{1}{2}\frac{\partial}{\partial y}[\lambda^2(y,t)]p(x,y,t)\right]_{y_{min}}^{y_{max}} dx,
\end{aligned} \tag{5.1.11}$$

$$\begin{aligned}
0 &= \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \mathfrak{L}_{xy} p(x,y,t)\,dydx \\
&= \int_{x_{min}}^{x_{max}} \int_{y_{min}}^{y_{max}} \rho\frac{\partial^2}{\partial x \partial y}[L(x,t)f(y)\lambda(y,t)]p(x,y,t)\,dydx \\
&= \rho\Big(L(x_{max},t)f(y_{max})p(x_{max},y_{max},t) - L(x_{max},t)f(y_{min})p(x_{max},y_{min},t) \\
&\quad - L(x_{min},t)f(y_{max})p(x_{min},y_{max},t) + L(x_{min},t)f(y_{min})p(x_{min},y_{min},t)\Big).
\end{aligned} \tag{5.1.12}$$

We can thus get sufficient boundary conditions for conservation of the total probability mass by enforcing every term appearing in (5.1.10) to (5.1.12) to be 0. At the spot boundaries this becomes: for all $t > 0$, $y \in [y_{min}, y_{max}]$

$$\begin{cases} 0 = \left((r_t - q_t) - \frac{1}{2}L^2(x_{min},t)f(y)^2\right)p(x_{min},y,t) - \frac{1}{2}f(y)^2\frac{\partial}{\partial x}[L^2(x_{min},t)]p(x_{min},y,t) \\ 0 = \left(-(r_t - q_t) + \frac{1}{2}L^2(x_{max},t)f(y)^2\right)p(x_{max},y,t) + \frac{1}{2}f(y)^2\frac{\partial}{\partial x}[L^2(x_{max},t)]p(x_{max},y,t) \end{cases}. \tag{5.1.13}$$

At the volatility boundaries: for every $t > 0$, $x \in [x_{min}, x_{max}]$

$$\begin{cases} 0 = \mu(y_{min},t)p(x,y_{min},t) - \frac{1}{2}\frac{\partial}{\partial y}[\lambda^2(y_{min},t)]p(x,y_{min},t) \\ 0 = -\mu(y_{max},t)p(x,y_{max},t) + \frac{1}{2}\frac{\partial}{\partial y}[\lambda^2(y_{max},t)]p(x,y_{max},t) \end{cases}. \tag{5.1.14}$$

And finally in the corners of our domain

$$0 = p(u,v,t) \qquad \text{for } (u,v) \in \{x_{min}.x_{max}\} \times \{y_{min}, y_{max}\}. \tag{5.1.15}$$

A theoretical justification for the no-flux boundary conditions in the case of the Heston SLVM is given in [42, Theorems 4.1,4.2, pages 8,9], where Lucic shows that, under some smoothness and boundedness assumptions that guarantee existence and uniqueness of a solution to the Heston SLV SDE (4.5.1), satisfying the no flux conditions at the volatility boundary is necessary and sufficient for a solution of the Fokker-Planck PDE (4.5.2) to be the density of the Heston SLVM (4.5.1). From a numerical point of view Clark [22, page 103] suggested to use the no-flux conditions for solving the Heston SLV Fokker-Planck equation at the volatility boundary, and Garrivier [34, Section 4.4.2, pages 12-17] employed the full no-flux boundary conditions, as presented here, for the MQR model. To our knowledge there is no discussion available about boundary conditions for the Tataru-Fisher model. We choose to use the full no-flux conditions for all three models under discussion.
With the boundary conditions (5.1.13) to (5.1.15) established, the next question is how to incorporate them into the finite difference scheme.
For this we follow a standard finite difference approach by introducing a layer of ghost points

$$\{(x_{-1}, y_j), (x_{N+1}, y_j), (x_i, y_{-1}), (x_i, y_{M+1}) \,|\, 0 \le i \le N,\, 0 \le j \le M\}$$

so that

$$\begin{cases} \Delta x_{-1} := x_0 - x_{-1} = \Delta x_0 \\ \Delta x_N := x_{N+1} - x_N = \Delta x_{N-1} \\ \Delta y_{-1} := y_0 - y_{-1} = \Delta y_0 \\ \Delta y_M := y_{M+1} - y_M = \Delta y_{M-1} \end{cases}.$$

We then discretise (5.1.13) and (5.1.14) using Theorem 5.1.1 as

$$\begin{aligned} 0 = &\left((r_t - q_t) - \frac{1}{2}L_{0,t}^2 f(y_j)^2\right) p_{0,j}^t - \frac{1}{2}f(y_j)^2\Bigg(-\frac{\Delta x_0 (L_{-1,t}^2 p_{-1,j}^t)}{\Delta x_{-1}(\Delta x_0 + \Delta x_{-1})} \\ &+ \frac{(\Delta x_0 - \Delta x_{-1})L_{0,t}^2 p_{0,j}^t}{\Delta x_0 \Delta x_{-1}} + \frac{\Delta x_{-1}(L_{1,t}^2 p_{1,j}^t)}{\Delta x_0(\Delta x_0 + \Delta x_{-1})}\Bigg), \end{aligned} \tag{5.1.16}$$

$$\begin{aligned} 0 = &\left(-(r_t - q_t) + \frac{1}{2}L_{N,t}^2 f(y_j)^2\right) p_{N,j}^t + \frac{1}{2}f(y_j)^2\Bigg(-\frac{\Delta x_N (L_{N-1,t}^2 p_{N-1,j}^t)}{\Delta x_{N-1}(\Delta x_N + \Delta x_{N-1})} \\ &+ \frac{(\Delta x_N - \Delta x_{N-1})L_{N,t}^2 p_{N,j}^t}{\Delta x_N \Delta x_{N-1}} + \frac{\Delta x_{N-1}(L_{N+1,t}^2 p_{N+1,j}^t)}{\Delta x_N(\Delta x_N + \Delta x_{N-1})}\Bigg), \end{aligned} \tag{5.1.17}$$

$$\begin{aligned} 0 = &\mu(y_0,t)p_{i,0}^t - \frac{1}{2}\Bigg(-\frac{\Delta y_0(\lambda^2(y_{-1},t)p_{i,-1}^t)}{\Delta y_{-1}(\Delta y_0 + \Delta y_{-1})} \\ &+ \frac{(\Delta y_0 - \Delta y_{-1})\lambda^2(y_0,t)p_{i,0}^t}{\Delta y_0 \Delta y_{-1}} + \frac{\Delta y_{-1}(\lambda^2(y_1,t)p_{i,1}^t)}{\Delta y_0(\Delta y_0 + \Delta y_{-1})}\Bigg), \end{aligned} \tag{5.1.18}$$

$$\begin{aligned} 0 = &-\mu(y_M,t)p_{i,M}^t + \frac{1}{2}\Bigg(-\frac{\Delta y_M(\lambda^2(y_{M-1},t)p_{i,M-1}^t)}{\Delta y_{M-1}(\Delta y_M + \Delta y_{M-1})} \\ &+ \frac{(\Delta y_M - \Delta y_{M-1})\lambda^2(y_M,t)p_{i,M}^t}{\Delta y_M \Delta y_{M-1}} + \frac{\Delta y_{M-1}(\lambda^2(y_{M+1},t)p_{i,M+1}^t)}{\Delta y_M(\Delta y_M + \Delta y_{M-1})}\Bigg), \end{aligned} \tag{5.1.19}$$

for $1 \le i \le N-1$, $1 \le j \le M-1$. By making, as suggested in [34, Page 13], the assumption $L_{-1,t} = L_{0,t} = L_{1,t}$ and $L_{N-1,t} = L_{N,t} = L_{N+1,t}$, we may solve (5.1.16) to (5.1.19) for the values of p at the ghost points in terms of values of p on the grid, which we may write as

$$\begin{aligned} p_{-1,j} &= A_{x,lower}p_{0,j} + B_{x,lower}p_{1,j}, \\ p_{N+1,j} &= A_{x,upper}p_{N-1,} + B_{x,upper}p_{N,j}, \\ p_{i,-1} &= A_{y,lower}p_{i,0} + B_{y,lower}p_{i,1}, \\ p_{i,M+1} &= A_{y,lower}p_{i,M-1} + B_{y,upper}p_{i,M+1}. \end{aligned}$$

Now recalling that in terms of the ghost points for example for $1 \le j \le N-1$

$$\begin{aligned} \mathcal{L}_x p_{0,j}^t = &-\frac{\Delta x_0[-(r_t - q_t) + \frac{1}{2}L_{-1,t}^2 f^2(y_j)]p_{-1,j}^t}{\Delta x_{-1}(\Delta x_0 + \Delta x_{-1})} + \frac{(\Delta x_0 - \Delta x_{-1})[-(r_t - q_t) + \frac{1}{2}L_{0,t}^2 f^2(y_j)]p_{0,j}^t}{\Delta x_0 \Delta x_{-1}} \\ &+ \frac{\Delta x_{-1}[-(r_t - q_t) + \frac{1}{2}L_{1,t}^2 f^2(y_j)]p_{1,j}^t}{\Delta x_0(\Delta x_0 + \Delta x_{-1})} \\ &+ \frac{f(y_j)^2}{2}\left(\frac{2L_{-1,t}^2 p_{-1,j}^t}{\Delta x_{-1}(\Delta x_0 + \Delta x_{-1})} - \frac{2L_{0,t}^2 p_{0,j}^t}{\Delta x_0 \Delta x_{-1}} + \frac{2L_{1,t}^2 p_{1,j}^t}{\Delta x_0(\Delta x_0 + \Delta x_{-1})}\right), \end{aligned}$$

we can substitute the expression for $p_{-1,j}$ obtained from (5.1.16), $\Delta x_{-1} = \Delta x_0$ and $L_{-1,t} = L_{0,t}$ to obtain an expression purely in terms of $p_{0,j}$ and $p_{1,j}$ from which we get the missing values $b_0^{j,t}, c_0^{j,t}$ in the matrix form of $\mathcal{L}_x$ (5.1.7).
Similarly we get for $1 \le j \le N-1$ $a_N^{j,t}$ and $b_N^{j,t}$ as well as $\beta_0^t, \gamma_0^t, \alpha_M^t$ and γ_M^t in (5.1.7) and (5.1.8) respectively. Additionally using the expressions for the values at the ghost points we can solve for the action of $\mathcal{L}_{xy}$ on boundary values in the same manner.
We handle the corners by setting

$$b_0^{0,t} = b_0^{M,t} = b_N^{0,t} = b_N^{M,t} = 1$$

and

$$c_0^{0,t} = c_0^{M,t} = a_N^{0,t} = a_N^{M,t} = 0,$$

as well as

$$\mathcal{L}_{xy} p_{i,j} = 0 \qquad \text{for } (i,j) \in \{0,N\} \times \{0,M\}$$

and manually setting the corners of $P_t(\mathcal{D}_y^t)^\top$ to 0.
For completeness the exact boundary terms for all three models under consideration, which we derived using SimPy [43], can be found in Appendix A.3,A.4 and A.5.
With the spatial part of the PDE handled we now turn our attention to resolving our discretised system in time.

5.2 Temporal Discretisation

Recall that originally we represented (4.2.2) as

$$\frac{\partial}{\partial t} p = \mathfrak{L} p.$$

In one space dimension partial differential equations of this form are usually numerically resolved using on of the well known explicit, implicit or Crank-Nicolson time stepping schemes, see [44, Chapter 8].
This is done by discretising $\mathfrak{L}$ and representing it as matrix. The resulting matrix will be tridiagonal , which allows to efficiently solve the linear equations resulting from the implicit or Crank-Nicoloson scheme in $\mathcal{O}(N)$ steps using the tridiagonal matrix algorithm [45, Section 2.6.3].
All these one dimensional time stepping schemes generalise to two spatial dimensions however a representation of $\mathfrak{L}$ will not be tridiagonal turning the implicit and Crank-Nicolson schemes inefficient. The explicit scheme is also not well suited for our needs as for its stability in two dimension the well known criterion

$$\Delta t \leq \frac{1}{2} \min_{0 \leq i \leq N-1, 0 \leq j \leq M-1} \{(\Delta x_i)^2, (\Delta y_j)^2\}$$

needs to hold [22, Equation 7.53, page 162]. Here Δt is the size of time steps. Especially with the non uniform grids presented in Section 5.1.1 this would require a prohibitively large amount of time steps.
A new kind of time stepping scheme is therefore needed. Recalling Section 5.1.2, if the operator $\mathfrak{L}$ is split into its directional components $\mathfrak{L}_x, \mathfrak{L}_y$ and $\mathfrak{L}_{xy}$, at least the first two can be represented as tridiagonal matrices (5.1.7) and (5.1.8). We will now introduce a scheme that makes use of this fact.
As stability properties of a time stepping scheme might restrict the choice of the time step size for now we just assume the time interval $[0,T]$ has been discretised as

$$\{t_0 := 0, t_1, \ldots, t_{L-1}, t_{\mathrm{L}} := T\} \qquad \text{for } \mathrm{L} \in \mathbb{N}$$

where $t_l < t_{l+1}$ for $0 \leq l \leq \mathrm{L}-1$, but for now leave L and $\Delta t_l := t_{l+t} - t_l$ unspecified.

5.2.1 ADI Time Stepping

As mentioned, for efficiently solving the semi discrete system

$$\frac{\partial}{\partial t} P^t = \mathcal{L} P^t$$

we want to make use of the fact that when splitting $\mathcal{L}$ into its coordinate direction and mixed derivative components $\mathcal{L}_x, \mathcal{L}_y$ and $\mathcal{L}_{xy}$ the first two can be represented as tridiagonal matrices. The main idea of so called alternating direction implicit , short ADI schemes, then is to split each time iteration into two half time steps where in each half step one coordinate direction is resolved implicitly. Since the mixed term can not be written as a tridiagonal matrix it has to be handled in an explicit fashion. The hope then is that if the contribution of the mixed term is controlled in some sense the two implicit half steps along the coordinate axis will be sufficient to guarantee stability of the scheme.

The particular ADI scheme, known as the **Douglas scheme**, we will use, recommended in [22, Equation 7.82, page 172], is an extension to handle mixed derivatives [46, Equation 6, page 342] of a scheme introduced by Douglas [47, Equations 2.6a,b , page 43].
The scheme approximates $p_{i,j}^{t_{l+1}} := p_{i,j}^{l+1}$ through an intermediate step $\mathfrak{p}_{i,j}^{l+1}$ by

$$(1 - \vartheta \Delta t_l \mathcal{L}_x)\mathfrak{p}_{i,j}^{l+1} = \Big(1 + \Delta t_l[(1-\vartheta)\mathcal{L}_x + \mathcal{L}_y + \mathcal{L}_{xy}]\Big)p_{i,j}^{l} \tag{5.2.1}$$

$$(1 - \vartheta \Delta t_l \mathcal{L}_x)p_{i,j}^{l+1} = \mathfrak{p}_{i,j}^{l+1} - \Delta t_l \vartheta \mathcal{L}_y p_{i,j}^{l} \tag{5.2.2}$$

where $0 \leq \vartheta \leq 1$ determines the implicitness of the scheme, $\vartheta = 1$ corresponding to a fully implicit handling of the unidirectional derivatives and $\vartheta = 0$ to fully explicit. In presence of the mixed derivative term the Douglas scheme is consistent of order $\mathcal{O}(\Delta t)$ [48, Page 3], and unconditionally stable for linear parabolic PDEs when $\vartheta \geq \frac{1}{2}$ [49, Theorem 3.1, page 30].
Initially there appear to be two problems with applying the scheme to the problem at hand. Firstly the application of $\mathcal{L}_x, \mathcal{L}_y$ to the intermediate solution $\mathfrak{p}$ and to p at time t_{l+1} for the implicit part of the scheme require the knowledge of p^{l+1} to calculate the leverage function $L_{i,l+1}$ which is what we want to solve for in the first place. Secondly our PDE (4.2.2) is non linear, so the stability result does not apply.
We solve both problems by simply keeping L constant in each iteration of (5.2.1)(5.2.2), that is we set $L_{i,t_{l+1}} = L_{i,t_l}$ in $(\mathcal{D}_x^{j,l+1})_{0\leq j\leq M}$ during the l^{th} step of the scheme. Effectively this also linearizes the PDE step wise making the calibration iteration unconditionally stable for $\vartheta \geq \frac{1}{2}$.
For completeness we want to mention that another popular approach in the literature is to, instead of keeping L constant, use P^l to calculate $L(\cdot, t_{l+1})$ for the l^{th} Douglas scheme step [36, Chapter 5.2.3, Page 70].

5.2.2 Temporal Grid Choice

Since as established in the last section the time stepping scheme is unconditionally stable, for $\vartheta \geq \frac{1}{2}$ which we always choose, we are completely free in the choice of our temporal grid spacing.
Since at the beginning of the calibration the density will be quite singular and changing very rapidly, we will benefit from taking very small time steps. Then over time the density will become more stable and we can get away with having larger time steps. For a visualisation of this we again refer to Figure 5.2 and Appendix B.1 B.2.
While we could use the grid generation scheme presented in 5.1.1 with one concentration point at 0 to accomplish this we rather recommend the scheme given in [22, Equation 7.8, page 175]

$$t_l := u_l^{\alpha_{\text{short}} + (\alpha_{\text{long}} - \alpha_{\text{short}})\exp(-\varphi u_l)},$$

where $u_l \in \{0, \frac{1}{\mathrm{L}}, \cdots \frac{\mathrm{L}-1}{\mathrm{L}}, 1\}$, since it gives us more control over the grid through $\alpha_{\text{short}}, \alpha_{\text{long}} > 0$ which control the concentration of grid points in the short and long term respectively, while $\varphi > 0$ controls the mixing in the medium term. Typical values we worked with were $\alpha_{\text{short}} = 1.2$, $\alpha_{\text{long}} = 2.2$ and $\varphi = 0.2$. An example of the kind of time grid this produces is given in Figure 5.4.

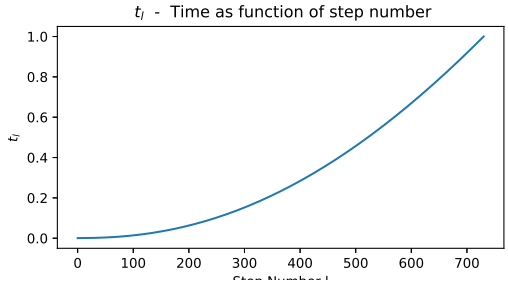
5.2.3 Approximation of the initial Condition

Even with the spatial grid being highly concentrated around $(0, Y_0)$, and using small time steps at beginning of the calibration, the Dirac initial condition remains challenging to handle for a finite difference scheme. Therefore the initial density is best approximated by a smooth function. A theoretical justification for approximating the initial condition would be that it makes Tachet's existence result for the calibration Fokker-Planck PDE discussed in Section 4.1 available.
To approximate the Dirac initial condition in a manner that is consistent with (4.2.1) we discretise the SLV SDE for one small step using the Euler scheme which leads to

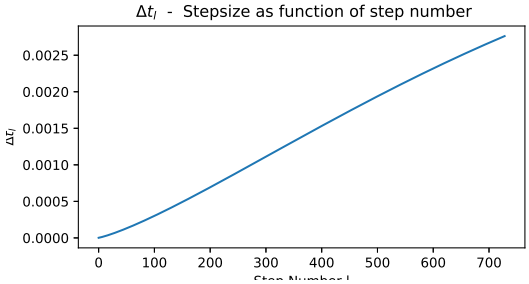
$$\begin{cases} X_1 & \approx [(r_0 - q_0) - \frac{1}{2}\sigma_{loc}^2(S_0,0)]\Delta t_0 + \sigma_{loc}(S_0,0)\sqrt{\Delta t_0}Z_1 \\ Y_1 & \approx Y_0 + \mu(Y_0,0)\Delta t_0 + \lambda(Y_0,0)\sqrt{\Delta t_0}Z_2 \end{cases}$$

where Z_1, Z_2 are standard normal random variables with correlation ρ. Hence approximately

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} [(r_0 - q_0) - \frac{1}{2}\sigma_{loc}^2(S_0,0)]\Delta t_0 \\ Y_0 + \mu(Y_0,0)\Delta t_0 \end{bmatrix}, \Delta t_0 \begin{bmatrix} \sigma_{loc}^2(S_0,0) & \rho\sigma_{loc}(S_0,0)\lambda(Y_0,0) \\ \rho\sigma_{loc}(S_0,0)\lambda(Y_0,0) & \mu(Y_0,0)^2 \end{bmatrix}\right). \tag{5.2.3}$$



(a) t_l as function of l, x-axis is l.



(b) Δt_l as function of l, x-axis is l.

Figure 5.4: Example of tight time grid for 1 year MQR calibration to SPX, 730 steps.

Denoting the mean vector of this normal distribution $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and the covariance matrix as

$$\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}$$

$p(\cdot,\cdot,t_1)$ is then approximated as the density of this normal distribution

$$p(x,y,t_1) := \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right).$$

Additionally depending on f in (4.2.1) we can, using that X_1, Y_1 follow a bi-variate normal distribution, analytically calculate the conditional expectation

$$\psi(x,t) = \mathbb{E}[f(Y_1)^2|S_{t_1} = \mathrm{e}^x S_0] = \mathbb{E}[f(Y)^2|X_1 = x],$$

which gives us L at t_1 via (4.1.1). If this is not possible the conditional expectation is calculated via numerical integration, here using the trapezoidal rule, as

$$\psi(x_i,t_1) \approx \frac{\frac{1}{2}\sum_{j=0}^{M-1}(f^2(y_j)p^1_{i,j} + f^2(y_{j+1})p^1_{i,j+1})\Delta y_i}{\frac{1}{2}\sum_{j=0}^{M-1}(p^1_{i,j}+p^1_{i,j+1})\Delta y_i}.$$

The exact normal approximations and analytical leverage functions for the MQR, Tataru-Fisher and Heston SLV model may be found in Appendix A.3,A.4 and A.5 respectively.
While this way of approximating the initial condition, or approximating it with some other normal with small variances, is quite common in literature, see for example [36, Equations 5.10,5.11, page 63], the density approximation is then usually used as the initial condition at time 0. In view of time dependent coefficients like for example θ in the MQR model, and how the density approximation is derived, we think this is inconsistent with the PDE, and suggest instead to simply start the PDE calibration from t_1 with the density approximation as initial condition, as done in [34, Section 4.6, page 20]. t_1 should neither be to small, leading to similar problems as using the Dirac condition outright, nor to large as this would give a worse approximation. We found good results with choosing t_1 corresponding to about 8 hours which depending on the daycount convention might be somewhere between $8\cdot10^{-4}$ to $1.2\cdot10^{-3}$. The time grid described in the previous section is then actually generated between t_1 and T.

5.2.4 Dealing with Drift for long Maturities

When calibrating to large maturities T of a few years, the interest rate and dividend drift term in (4.2.1) can become impactful and might move the center of the probability density, where most curvature is , out of the grid region with the highest density of grid points. To combat this for long maturities we suggest modeling the log-forward $\log(\frac{F_{t,T}}{F_{0,T}})$ to remove this drift instead. The leverage function for the log-forward can then easily be related to the one of the stock price: By Ito's lemma the SDE for the log-forward under SLV is

$$dX_{t,T} = -\frac{1}{2}L_F^2(X_{t,T},t)f^2(Y_t)\,dt + L_F^2(X_{t,T},t)f(Y_t)\,dW_t.$$

where due to our assumption of deterministic rates $L(S_t,t) = L_F(\mathrm{e}^x F_{0,T}\mathrm{e}^{-\int_t^T (r_s-q_s)\,ds}, t)$. Applying Gyöngy's Theorem 1.0.7 to this process and matching it with the process for the log-forward under local volatility, we can then see that we can calibrate L_F by simply setting $r_t - q_t = 0$ everywhere in our previous discussion.

5.3 Practical Implementation of the Calibration Algorithm

With the spatial and temporal discretisation of (4.2.2) in place we may now revisit the calibration Algorithm 1 and give its practical implementation. In the following we will denote the j^{th} column of a matrix A as $A(j)$ and the i^{th} row as $M(i,r)$. Furthermore to relate the algorithm to an implementation in a programming language all matrices and vectors are 0 indexed. It will be convenient to denote by $\mathfrak{D}_y^l$ (5.1.8) with $\beta_0^t = \beta_M^t = 1$ and $\gamma_0^t = \alpha_M^t = 0$. Lastly we will write the solution vector x of a matrix equation $Ax = b$ as $solution(A,b)$ and the identity matrix in n dimensions as $\mathbb{I} \in \mathbb{R}^{n\times n}$.

Algorithm 2: Leverage Function Bootstrap - Practical Implementation

Set up space grid $\mathfrak{G}_{x,y}$ as described in 5.1.1;
Approximate P^1 and $(L_{i,1})_{0\le i\le N}$ as in 5.2.3;
Set corners of P^1 to 0;
Set up time grid $\{t_1 \cdots, t_{\mathrm{L}} = T\}$;
for $l \in \{1, \cdots, \mathrm{L}-1\}$ **do**
  $A \leftarrow \mathcal{L}_{x,y}P^t$;
  $B \leftarrow P^t(D_y^l)^\top$;
  Set corners of B to 0;
  for $j \in \{0 \cdots M\}$ **do**
    $A(j) \leftarrow solution\Big(\mathbb{I} - \vartheta\Delta t_l D_x^{j,l+1}, \Big(\mathbb{I} + \Delta t_l(1-\vartheta)\mathcal{D}_x^{j,l}\Big)P^l(j) + \Delta t_l[A(j) + B(j)]\Big)$;
  for $i \in \{1 \cdots N-1\}$ **do**
    $A(i,r) \leftarrow solution\Big(\mathbb{I} - \vartheta\Delta t_l \mathcal{D}_y^{l+1}, A(i,r) - \vartheta\Delta t_l B(i,r)\Big)$;
  $A(0,r) \leftarrow solution\Big(\mathbb{I} - \vartheta\Delta t_l \mathfrak{D}_y^{l+1}, A(0,r) - \vartheta\Delta t_l B(0,r)\Big)$;
  $A(N,r) \leftarrow solution\Big(\mathbb{I} - \vartheta\Delta t_l \mathfrak{D}_y^{l+1}, A(0,r) - \vartheta\Delta t_l B(N,r)\Big)$;
  $P^{l+1} \leftarrow A$;
  for $i \in \{0 \cdots N\}$ **do**
    $\psi_{i,l+1} \leftarrow \frac{\frac{1}{2}\sum_{j=0}^{M-1}(f^2(y_j)p_{i,j}^{l+1} + f^2(y_{j+1})p_{i,j+1}^{l+1})\Delta y_i}{\frac{1}{2}\sum_{j=0}^{M-1}(p_{i,j}^{l+1} + p_{i,j+1}^{l+1})\Delta y_i}$;
    $L_{i,l+1} \leftarrow \frac{\sigma_{loc}(x_i,t_{l+1})}{\sqrt{\psi_{i,l+1}}}$;

Since each call of *solution* has, due to (5.1.7) and (5.1.8) being tridiagonal, complexity at most $\mathcal{O}(\max\{N,M\})$ the whole calibration algorithm has complexity $\mathcal{O}((\max\{N,M,L\})^3)$.
Depending on the grid size it can be possible that in the beginning, when calculating ψ far way from 0, the numerical integral in the denominator is 0. In this case we add small positive constants

ε, c to the denominator and numerator in the calculation of ψ respectively. In our implementation we choose $\varepsilon = 10^{-8}$ and for mean reverting models like the specific ones we consider in this thesis $c = \varepsilon\theta$ or $c = \varepsilon\theta^2$ depending on if θ is the mean reversion level of the variance like in the Heston SLV or of the volatility like in the MQR and Tataru-Fisher models.
Additionally due to numerical error in the beginning of the calibration small negative density values can appear in the low probability mass regions, these are truncated to 0.

5.4 Pricing with SLV

After having obtained obtained the family of density matrices $(P^{t_l})_{0\le l\le L}$ and the corresponding leverage function grid $(L_{i,l})_{0\le i\le N, 0\le l\le L}$, the SLV model is fully calibrated and can be used for pricing.
European Option prices for maturity t_l can be calculated directly by integration using the density matrix P^l. The price of a European Call and a Put option with strike K and maturity t_l will be given respectively by:

$$\begin{aligned} C(K,t_l)_{SLV} &= D(0,t_1)\int_{\log(\frac{K}{S_0})}^{\infty}\int_{\Omega_y} p(x,y,t_l)(\mathrm{e}^x S_0 - K)\,dydx \\ &\approx \frac{D(0,t_l)}{4}\sum_{i=k}^{N-1}\sum_{j=0}^{M-1}\Delta x_i\Delta y_j\Big(p^l_{i,j}(\mathrm{e}^{x_i}S_0-K)+p^l_{i+1,j}(\mathrm{e}^{x_{i+1}}S_0-K) \\ &\quad + p^l_{i,j+1}(\mathrm{e}^{x_i}S_0-K)+p^l_{i+1,j+1}(\mathrm{e}^{x_{i+1}}S_0-K)\Big) \qquad (5.4.1) \\ P(K,t_l)_{SLV} &= D(0,t_1)\int_{-\infty}^{\log(\frac{K}{S_0})}\int_{\Omega_y} p(x,y,t_l)(K-\mathrm{e}^x S_0)\,dydx \\ &\approx \frac{D(0,t_l)}{4}\sum_{i=0}^{k-1}\sum_{j=0}^{M-1}\Delta x_i\Delta y_j\Big(p^l_{i,j}(K-\mathrm{e}^{x_i}S_0)+p^l_{i+1,j}(K-\mathrm{e}^{x_{i+1}}S_0) \\ &\quad + p^l_{i,j+1}(K-\mathrm{e}^{x_i}S_0)+p^l_{i+1,j+1}(K-\mathrm{e}^{x_{i+1}}S_0)\Big) \qquad (5.4.2) \end{aligned}$$

where $x_k := \inf\{x_i \,|\, x_i \ge \log(\frac{K}{S_0})\}$ and Ω_y denotes the range of Y.
Since we are mainly interested in SLV models for pricing Exotic Options we wish to generate Monte Carlo sample paths of X. For this we need to interpolate the leverage function grid $(L_{i,l})_{0\le i\le N,0\le l\le L}$ to create a leverage function surface. From this we can then generate sample paths for X using the Euler scheme , while the exact path generation scheme for $\sigma = f(Y)$ is best chosen on case to case basis. To interpolate the leverage surface , we use cubic splines in X direction, and we interpolate L^2 linear in time. For x outside the calibration grid we extrapolate the surface flat that is for $x < x_{min}$ we set $L(x,t) = L(x_{min},t)$ for all $t > 0$ and similarly on the upper boundary.
To exemplify the procedure we present our simulation schemes for the MQR , Tataru-Fisher , and Heston SLV model.
For some time grid $\{0\,t_1,\cdots t_K := T\}$ which can be different from the one used during the calibration of the leverage function L we always set $X_0 = 0$ and for $0 \le k \le K-1$

$$X_{k+1} = X_k + (r_k - q_k - \frac{1}{2}L^2(X_k,t_k)\sigma_k^2)\Delta t_k + L(X_k,t_k)\sigma_k\sqrt{\Delta t_k}Z_k$$

for a family of independent standard normal random variables $(Z_k)_{0\le k\le K-1}$.
Then in the case of the MQR model we use that we know the conditional distribution of Y_t given Y_s for $t > s$ (3.1.3) so we can simulate Y exactly as $Y_0 = 0$ and

$$Y_{k+1} = Y_k\mathrm{e}^{-\kappa\Delta t_k} - \frac{\lambda^2}{2}\frac{1-\mathrm{e}^{-\kappa\Delta t_k}+\mathrm{e}^{-\kappa(2t_{k+1}+\Delta t_k)}-\mathrm{e}^{-2\kappa t_{k+1}}}{\kappa} + \sqrt{\frac{\lambda^2}{2\kappa}(1-\mathrm{e}^{-2\kappa\Delta t_k})}\tilde{Z}_k$$

for another family of independent standard normals $(\tilde{Z}_k)_{0\le k\le K-1}$ with $\mathrm{Corr}(Z_k,\tilde{Z}_k) = \rho$ and set $\sigma_k = \mathrm{e}^{Y_k}$.
For the Tataru-Fisher model we simulate the volatility process directly. Since as we recall the

volatility process here is strictly positive and the expected volatility is 1 at all times the probability of introducing negative volatility through discretisation errors is quite low so we use a simple Euler scheme for the volatility process. We set $\sigma_0 = 1$ and

$$\sigma_{k+1} = \sigma_k + (1 - \sigma_k)\Delta t_k + \lambda\sigma_k\sqrt{\Delta t_k}\tilde{Z}_k.$$

Should it happen after all that $\sigma_k < 0$ we set $\tilde{\sigma}_k := |\sigma_k|$ and use this as our process value. We will allow the volatility process to hit 0 during simulation.
For the Heston SLV model we simulate the variance process $V = \sigma^2$. Unlike for the Tataru-Fisher model, especially when the Feller condition (3.3.2) is violated, there is a significant probability for negative variance values due to the discretisation error in the simulation. To combat this we follow Gatheral's suggestion of using a Milstein scheme for the variance process as this reduces the probability for negative values [12, Equation 2.18, Chapter 2, page 22]. Then $V_0 = V_0$ and

$$V_{k+1} = V_k + \Delta t_k\kappa(\theta - V_k) + \lambda\sqrt{V_k\Delta t_k}\tilde{Z}_k - \frac{\lambda^2}{2}((\sqrt{\Delta t_k}\tilde{Z})^2 - \Delta t_k).$$

If negative values still occur they are handled by the same reflection technique we employ for the Tataru-Fisher model above.

Chapter 6

Numerical Results

We now apply the results of the last chapter by calibrating the MQR, Tataru-Fisher and Heston SLV models to the SPX implied volatility surface on the 23/02/2023 up to 23/02/2026 maturity. Unless specified otherwise we will use the following set of parameters:

- For the MQR model: $\kappa = 0.24$, $\lambda = 0.5$ and $\rho = -0.23$.
- For the Tataru-Fisher model: $\lambda = 0.5$ and $\rho = -0.23$.
- For the Heston SLV model: $\kappa = 0.6, \theta = 0.15, \lambda = 0.5, V_0 = 0.1$ and $\rho = -0.23$.

As we are mainly interested in studying the quality of the calibration archived by following Algorithm 2, which should be able to theoretically handle any reasonable stochastic volatility parameters, the stochastic parameters were chosen arbitrarily within a reasonable parameter range, only taking extra care to violate the Feller condition for the Heston SLV as to provide an extra challenge for the calibration. We attempted the calibration for a large set of parameters and as the results were quite similar we feel the choice of focusing the presentation on the above parameters is justified.
Furthermore we assumed that interest and dividend rates are constant at $r = 0.05$ and $q = 0.02$.

6.1 Quality of the Calibration

As the ultimate goal is that the calibrated SLV model will fit to the market quotes for vanilla options this is the main metric the quality of the calibration will be judged by.

6.1.1 Testing the Marginals of the Density

Our first approach to test this is by comparing implied volatilities generated by the SLV model to implied volatilities generated by the local volatility model. This is a better approach for judging the quality of the calibration than comparing the SLV implied volatilities directly to the market implied volatilities since we are calibrating the SLV model to the local volatility model after all, and thus even for a perfect calibration of the SLV model to the local volatility model there could be residual errors to the real implied volatility surface incurred from the calibration of the latter.

Our first approach to calculate the SLV implied volatilities is to compute Put and Call prices from the joint density p of X and Y, which we obtained as a by-product of the calibration of the leverage function L, using (5.4.2) and (5.4.1) and then backing out the implied volatilities $\sigma_{\text{imp}}^{SLV}$ from these prices. We also compute prices and from these implied volatilities σ_{imp}^{LV} for the same Puts and Calls using Monte Carlo and the local volatility model. Additionally we also compute $\mathbb{E}^{\mathbb{Q}}[D(0,t)S_t]$ from the density which should be equal to the current spot price and check that p integrates to approximately 1. If this is the case and $|\sigma_{\text{imp}}^{SLV} - \sigma_{\text{imp}}^{LV}|$ is small this indicates that the marginal density $p_X(x,t) := \int_{y_{min}}^{y_{max}} p(x,y,t)\,dy$ matches the density of $\log(\frac{S_t}{S_0})$ under local volatility reasonably well.
In our case we compute the risk neutral expectation and total probability mass at maturities of

1 week, 1,3,6,9,12,18,24,30 and 36 months with the start date being the 22/02/2023, and implied volatilities for the same maturities for strikes corresponding to grid points between 0.5 and 1.5 moneyness, except for the one week maturity where the strike range is between 0.9 and 1.05 moneyness, using Puts for moneyness below 1 and Calls for above. To keep the results between the models comparable in terms of computational effort we have chosen $N, M = 800$ for all three, even though for example the Heston SLV model with violated feller condition would, due the large log-variance domain and two concentration points in the grid, require more grid points in y direction than the MQR model with its normal distribution for the log-volatility. For the time steps we choose L = 1700 [1]. The results of the calibration are reported in Tables 6.1 to 6.3.

Maturity	Probability Mass	Model Spot	Mean Error (bps)	Median Error (bps)	Max Error (bps)
1w	0.999996	4012.31	3.1	2.5	10
1m	0.999998	4012.31	8.6	9.1	19.3
3m	0.999999	4012.32	2.7	3.2	4.1
6m	1.000001	4012.33	2.1	2.9	3.3
9m	0.999999	4012.32	1.3	1.5	1.2
12m	1.000001	4012.33	0.9	1	1.1
18m	1	4012.32	0.8	0.9	1.3
24m	0.999996	4012.26	0.8	0.7	1
30m	0.999963	4011.80	1.3	0.7	3.2
36m	0.999913	4011.01	2.3	0.8	5.8

Table 6.1: Results of the calibration for the MQR model. Actual spot is 4012.32. Errors are basis points in volatility: $10000|\sigma_{\text{imp}}^{\text{LSV}} - \sigma_{\text{imp}}^{\text{LV}}|$.

Maturity	Probability Mass	Model Spot	Mean Error (bps)	Median Error (bps)	Max Error (bps)
1w	0.999996	4012.31	3.1	2.5	10.5
1m	1.000004	4012.34	8.7	9.1	19.2
3m	1.000013	4012.37	2.6	3.1	4.3
6m	1.000024	4012.42	2	2.7	3.3
9m	1.000043	4012.50	1.1	1.4	1.1
12m	1.000043	4012.50	0.7	0.8	0.9
18m	1.000068	4012.60	0.6	0.7	1
24m	1.000087	4012.63	0.5	0.4	0.9
30m	1.000074	4012.26	1	0.4	2.9
36m	1.000073	4011.90	1.5	0.5	4.3

Table 6.2: Results of the calibration for the Tataru-Fisher Model. Actual spot is 4012.32. Errors are basis points in volatility: $10000|\sigma_{\text{imp}}^{\text{LSV}} - \sigma_{\text{imp}}^{\text{LV}}|$.

The results appear to be promising. For all three models the no-flux boundary conditions manage to keep the probability mass stable to a satisfactory degree, albeit we observe a somewhat larger instability for the Heston SLV model where the violated Feller conditions complicates the problem. Moreover the MQR model keeps the model spot extremely accurate up to 2 years and is still reasonably close at the three years maturity. This is similar for the Tataru-Fisher model where we observe larger but still small errors in the model spot up to 2 years and additionally smaller errors in the last year.
Together with higher loss in probability mass there is also a higher loss in the model spot for the Heston SLV model, but with retaining 99.9% of the original spot at the final maturity the model still performs acceptably. Finally the density of all three models capture the volatility surface very well for expiries between 6 months and three years, with mean and mostly even maximum errors in implied volatilities below 5 basis points. The accuracy is reduced for the very short expiries owing to the fact that even with a smooth approximation of the Dirac initial condition and tighter

[1] While this might seem excessive, this is mainly to generate sufficiently many time steps towards the beginning of the calibration which we found necessary to get a good accuracy for 1 week and 1 month maturities. With the non uniform time grid this still corresponds to taking less than 730 time steps between 1 and 3 years. We refer to Appendix B.6 for a small excerpt of convergence experiments we conducted with the MQR model for a 1 year calibration

Maturity	Probability Mass	Model Spot	Mean Error (bps)	Median Error (bps)	Max Error (bps)
1w	0.999995	4013.30	7.2	5.2	21.6
1m	1	4012.32	13.9	10.8	44
3m	1	4012.32	4.2	3.1	18.9
6m	0.999986	4012.26	2.3	1.9	8.2
9m	0.999963	4012.17	1.4	1.5	3.8
12m	0.999939	4012.01	1,1	1.2	3
18m	0.999891	4011.88	0.8	0.7	1.5
24m	0.999837	4011.61	0.7	0.8	1.3
30m	0.99975	4010.93	1.4	0.9	2.7
36m	0.999616	4009.5	2.8	1	4.8

Table 6.3: Results of the calibration for the Heston SLV model. Actual spot is 4012.32. Errors are basis points in volatility: $10000|\sigma_{\text{imp}}^{\text{LSV}} - \sigma_{\text{imp}}^{\text{LV}}|$.

grid spacing around the initial values very large derivatives are encountered around that region in the beginning of the calibration which are difficult to handle for finite difference schemes. Still the MQR and Tataru-Fisher models display mean errors of less than 10 and maximum errors of the less than 20 basis points so also here the marginal density captures the volatility surface well.
Over all we can conclude that the marginal density p_X matches the local volatility model's density quite well for all three models. Additionally we can also easily test if the marginal density of the volatility process $p_Y(y,t) := \int_{x_{min}}^{x_{max}} p(x,y,t)\,dx$ matches known moments of Y_t, or in the case of the MQR and Heston SLV models where the full distribution of the volatility process is known this distribution. We found this to be the case for all three models as well.

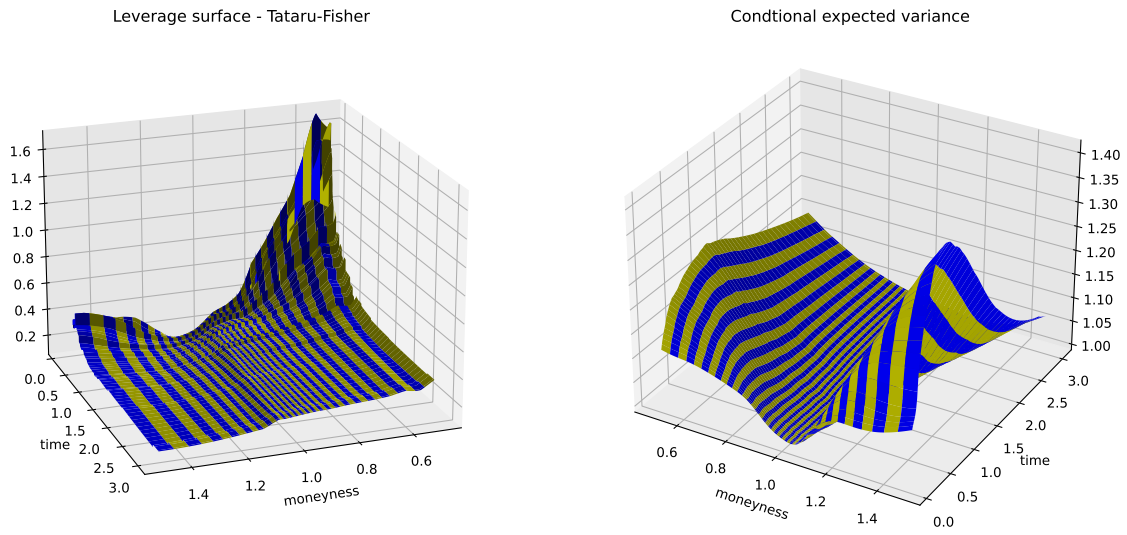


Figure 6.1: Leverage surface and square root of the conditional variance $\psi(k,t) = \sqrt{\mathbb{E}[\sigma_t^2|S_t/S_0 = k]}$ for the Tataru-Fisher model.

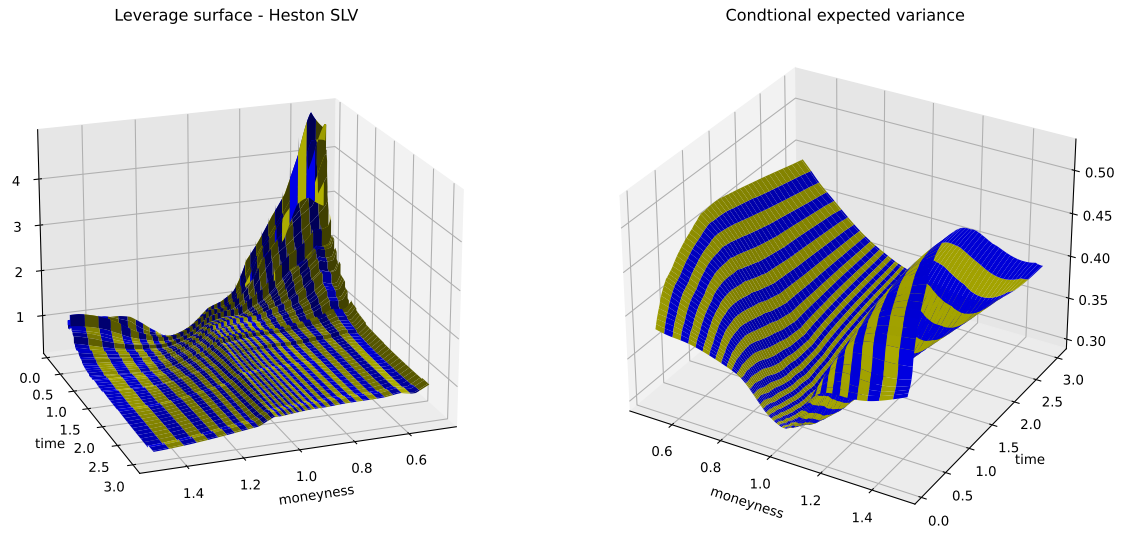


Figure 6.2: Leverage surface and square root of the conditional variance $\psi(k,t) = \sqrt{\mathbb{E}[\sigma_t^2 | S_t/S_0 = k]}$ for the Heston SLVM.

The leverage surfaces and corresponding ψ for the Tataru-Fisher and Heston SLVM can be found in Figures 6.2 and 6.1. The shape of ψ serves as a first confirmation of the interpretations made about the character of the various SLV models in Sections 4.3 to 4.5. For the Tataru-Fisher model ψ is indeed close to 1, especially around at the money, so here the leverage function will be very similar to σ_{loc} as expected. For the Heston SLVM we observe that ψ is strictly less than $\frac{1}{2}$ so here, also as expected, the leverage function will be quite far from σ_{loc}. The shape of the leverage surface also seems to be quite intuitive for both models, especially in the short term, where out of the money the leverage surfaces bumps up the volatility comparatively heavily and, then on the in the money wing, in case of the Tataru-Fisher model heavily reduces the on average quite high volatility, and for the Heston SLVM the, due to our choice of V_0, θ, quite low volatility is left around its level there. It then seems credible that this leverage surface forces the underlying stochastic volatility model to generate an implied volatility smirk with a shape similar to what we observed in Figure 2.1.
As the leverage surface and ψ for the MQR model behave similarly to the one of the Tataru-Fisher model, in line with the interpretation of this model as being more local in character, they are given in Appendix B.5 and we will not further comment on them.
Overall it seems like the SLV models have been successfully calibrated.
However just having the correct marginal densities and intuitive shapes still does not guarantee that the leverage function is correct as it depends on the joint distribution of X and Y and not only on the marginals, which leads us into the next section.

6.1.2 Testing the Leverage Function in Monte Carlo

To properly test if the leverage function is calibrated correctly we reprice vanilla options again, this time using Monte Carlo, as this will truly test the joint distribution of the stock and volatility process and thus the leverage function, with daily stepping by interpolating the leverage function into a leverage surface as described in Section 5.4, and again compute implied volatility from out of the money Puts and Calls.

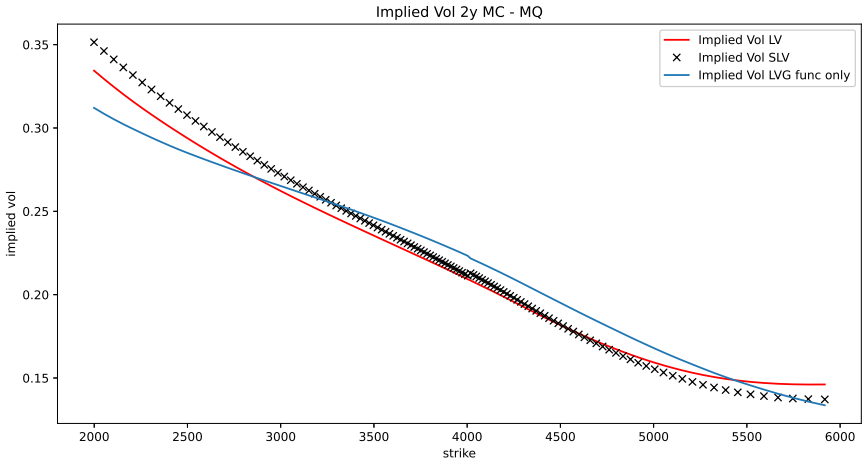
Moreover for the MQR and Tataru-Fisher model we also compute implied volatilities for the same Puts and Calls using only the leverage function, that is we price the options assuming

$$dS_t = (r_t - q_t)S_t\,dt + L(S_t, t)S_t\,dW_t$$

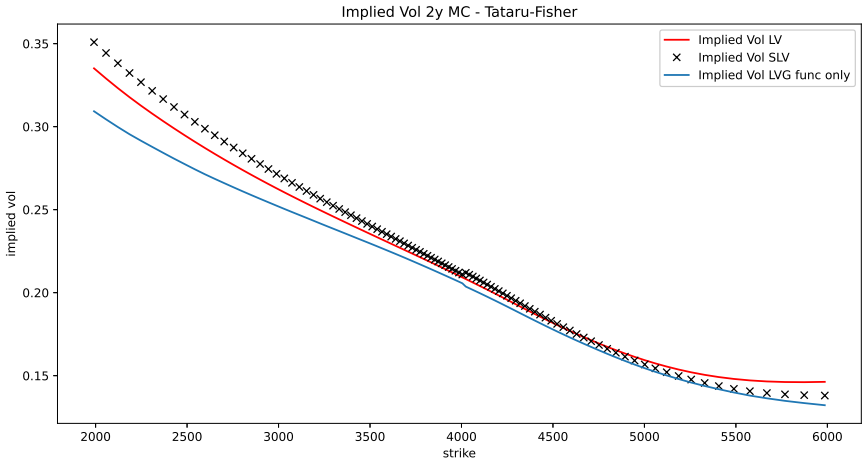
and back out corresponding implied volatilities $\sigma_{\text{imp}}^{\text{LVG}}$ to also further test our interpretation of the models being more local in nature and our expectation of the leverage function being close to σ_{loc}. If this interpretation is correct $\sigma_{\text{imp}}^{\text{LVG}}$ should be close to $\sigma_{\text{imp}}^{\text{LV}}$.
First off Figures 6.3(a) and 6.3(b) confirm the evidence from the previous section that the interpretation of the MQR and Tataru-Fisher model being mostly local volatility models with a stochastic multiplier made in Sections 4.3 and 4.4 is correct as we can see that especially for the Tataru-Fisher model around at the money most of the smile is explained by the leverage function itself. This is quite different from the Heston SLV model where the leverage function prices are not close to the actual prices at all, see Figure B.3.
On the other hand from Figure 6.3 we see that while the shape of the local volatility implied volatility surface is matched quite well by all three models there is a significant error, mean error in basis points being around 45, especially in the wings where we observe errors of up to 150 bps. Now there could be two explanations for this error: Either the chosen interpolation method for the leverage surface is insufficient or the joint density p of X, Y is wrong and thus the leverage function calculated from it, even though the marginals are correct.
To further investigate and locate the source of the error we re calibrate all three models to the same data set, setting the correlation $\rho = 0$ and keeping everything else the same.
As we see in in Figure 6.4 this eliminates the error and all three models fit the local volatility implied volatility smile close to perfectly with a mean error of around 8 bps and a maximum errors of less than 15 bps. This is consistent for all maturities, see Appendix B.4 for a further example. We can thus conclude that the error in the non zero correlation case stems from an incorrect leverage function which is caused by an incorrect joint density. Since the only difference in the calibration for the non zero correlation case is the addition of the mixed derivative term in (4.2.2) we can further pinpoint it as the root cause of the errors.

Due to the in general lengthy coefficients for the discretised mixed derivative operator (5.1.6) it is possible that there is an error in our implementation which we were not able to find. Barring this another more fundamental possible reason for the error is that the Douglas scheme (5.2.1)(5.2.2) provides insufficient handling for the mixed derivative term as it is only incorporated through an explicit step in (5.2.1). Because of this in the non zero correlation case we essentially are not fully relating $p_{i,j}^l$ to $p_{i,j}^{l+1}$ through the Fokker-Planck PDE, as a fully implicit or Crank-Nicolson scheme would, but partially merely Taylor expanding $p_{i,j}^l$ forward in time. Solutions to this problem of the Douglas Scheme present in the literature are so called predictor corrector ADI schemes, where it is used that after the second step of the Douglas scheme (5.2.2) $L_{i,l+1}$ can be calculated.

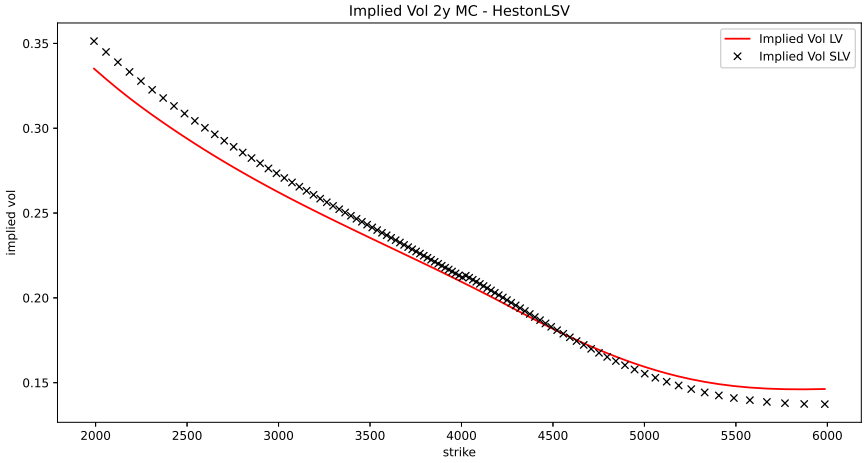
This is then used to do a second iteration similar to (5.2.1)(5.2.2) where due to the knowledge of a proxy solution, the final solution of the Douglas scheme, also the mixed derivative term can be related to $p_{i,j}^{l+1}$ through the PDE without having to solve a non tridiagonal matrix equation. A prominent example of these kind of schemes is the modified Craig-Sneyd scheme [50, Equation 1.3, page 678] which has also been used for the calibration of SLVMs in [38, Slide 9]. Using this scheme instead of the Douglas scheme might provide an improvement for the calibration in the non zero correlation case, making it an interesting option for further experimentation.



(a) 2 year maturity Monte Carlo implied volatilities $\sigma_{imp}^{\text{SLV}}$, $\sigma_{\text{imp}}^{\text{LV}}$ and $\sigma_{\text{imp}}^{\text{LVG}}$ for the MQR model. 1

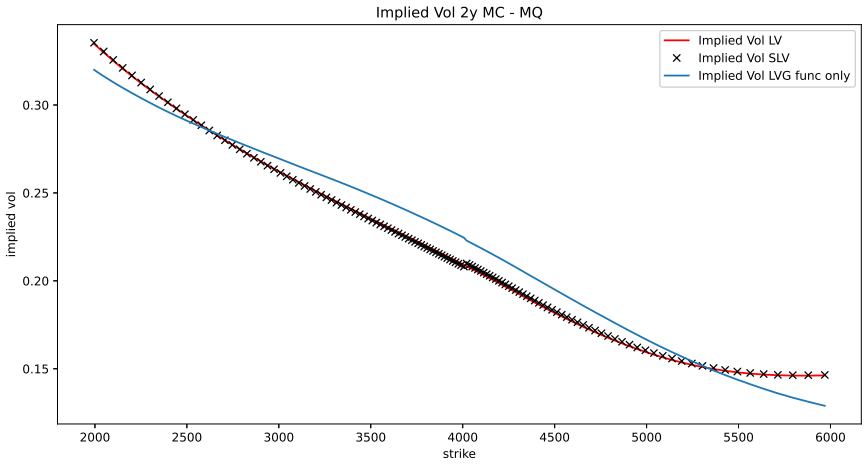


(b) 2 year maturity Monte Carlo implied volatilities $\sigma_{\text{imp}}^{\text{SLV}}$, σ_{imp}^{LV} and $\sigma_{\text{imp}}^{\text{LVG}}$ for the Tataru-Fisher model.

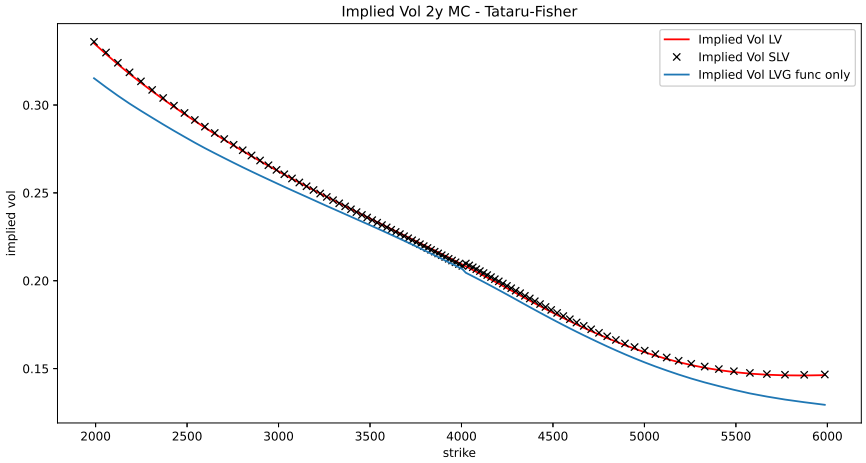


(c) 2 year maturity Monte Carlo implied volatilities $\sigma_{\text{imp}}^{\text{SLV}}$ and $\sigma_{\text{imp}}^{\text{LV}}$ for the Heston SLV model. 1

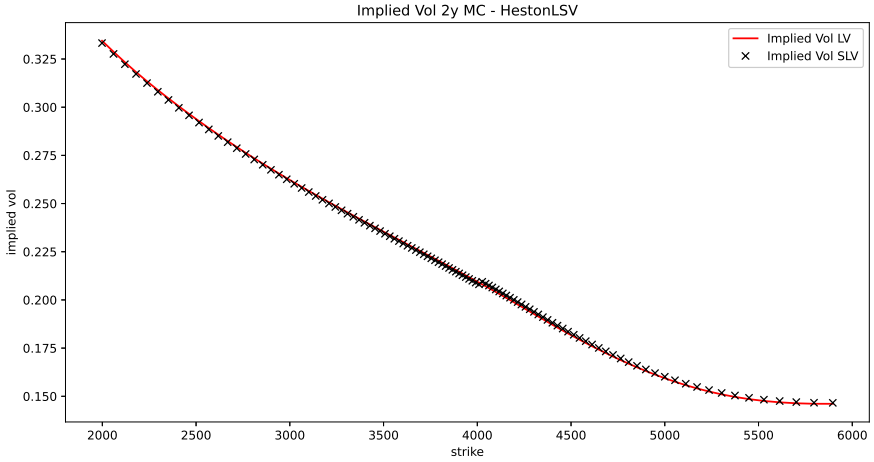
Figure 6.3: SPX implied volatilities computed via Monte Carlo with 100000 sample paths and daily steps.



(a) 2 year maturity Monte Carlo implied volatilities $\sigma_{imp}^{\text{SLV}}$, σ_{imp}^{LV} and $\sigma_{imp}^{\text{LVG}}$ for the MQR model with 0 correlation



(b) 2 year maturity Monte Carlo implied volatilities $\sigma_{imp}^{\text{SLV}}$, σ_{imp}^{LV} and $\sigma_{imp}^{\text{LVG}}$ for the Tataru-Fisher model with 0 correlation.

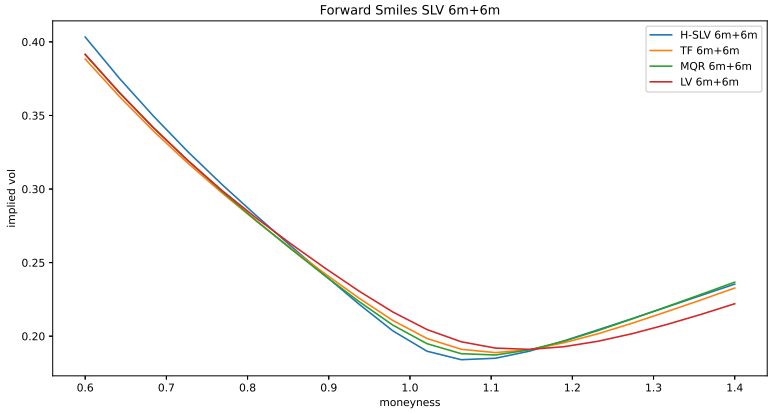


(c) 2 year maturity Monte Carlo implied volatilities $\sigma_{imp}^{\text{SLV}}$ and σ_{imp}^{LV} for the Heston SLV model with 0 correlation.

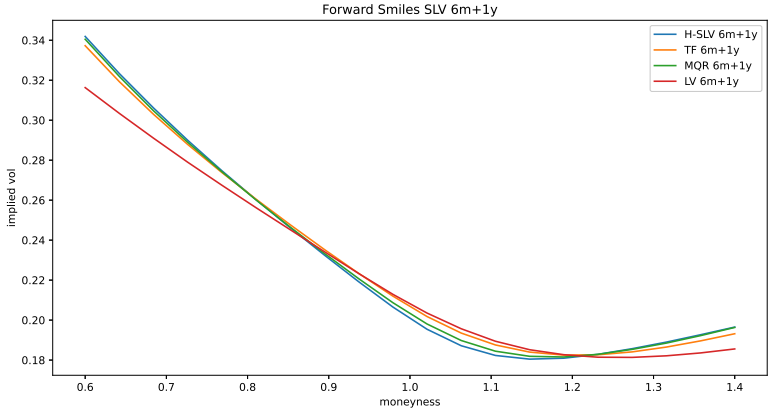
Figure 6.4: SPX implied volatilities computed via Monte Carlo with 100000 sample paths and daily steps, correlation $\rho = 0$.

6.2 Volatility Dynamics

With, at least in the zero-correlation case, fully calibrated SLVMs in place we can use them to price exotic options and compare the results to the local volatility model to see if they indeed feature volatility dynamics more akin to a stochastic volatility model. To be precise we want to investigate if the calibrated SLVMs do not suffer from the same problem of flattening forward implied volatility smiles as the local volatility model. For this we price forward starting options, introduced in Section 2.3 using Monte Carlo and then compute the forward implied volatility (2.3.1) from these prices. Figures 6.5, 6.6 shows that all three models feature more convex, less flat smiles than the local volatility model, meaning they indeed inherit the dynamic features of stochastic volatility models. Moreover we see that again the Tataru-Fisher model behaves closest to the local volatility model, in line with its intended interpretation, while the MQR model and Heston SLV with their less restricted parameter choices have even further freedom to deviate from the local volatility model.



(a) SLV SPX forward smiles $T_1 = 0.5$, $T_2 = 1$.



(b) SLV SPX forward smiles $T_1 = 0.5$, $T_2 = 1.5$.

Figure 6.5: SPX Forward implied volatility under different models as of 23/02/2023.

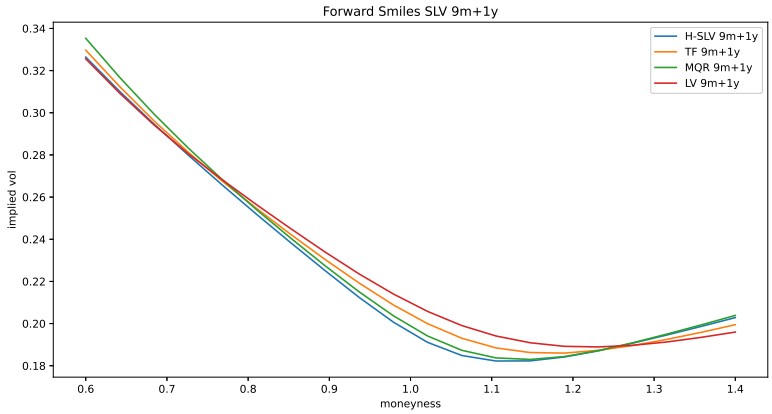


Figure 6.6: SLV SPX forward smiles $T_1 = 0.75$, $T_2 = 1.75$.

As an additional metric we also computed the forward at the money skew

$$\frac{\partial \sigma_{\text{imp}}}{\partial K}(T_1 \,|K,\, T_2)\Big|_{K=1}$$

via finite differences for all 4 models and observe, Table 6.4, that all three SLVMs exhibit more skew than the local volatility model, with the Tataru-Fisher model again being closest to the local volatility model. This again confirms that SLVMs feature richer volatility dynamics than the local volatility model.

(T_1, T_2)	Local Volatility	MQR	Tataru-Fisher	Heston SLV
$(0.5, 0.5)$	-0.1416	-0.1499	-0.1451	-0.1645
$(0.5, 1)$	-0.1113	-0.1267	-0.1214	-0.1321
$(0.75, 1)$	-0.098	-0.1137	-0.1054	-0.1139

Table 6.4: Forward at the money skew for local volatility and stochastic local volatility.

6.3 Which model to choose?

Finally let us briefly discuss which of the three implemented models we would recommend to use. In Section 6.1 we have found that when it comes to solving the Fokker-Planck equation for the density and leverage function the Heston SLV model provides the biggest challenge of the three models when trying to fit short maturity options well with the density, as well as for keeping the probability mass stable.
Still when it comes to pricing vanilla options via Monte Carlo all three models perform roughly similar, encountering the same problems in the non zero correlation case and being able to fit the smile quite well with correlation equal to zero.
On the other hand the superior stability of the probability mass and more accurate pricing of short maturity vanilla options with the fitted density inspires more confidence in the MQR and Tataru-Fisher to still be able to be calibrated under more extreme market scenarios than the one considered in this thesis. Simply put we found their calibration to be more robust.
The main advantage of the Heston SLV model is that there is a clear cut way to obtain the stochastic volatility parameters by calibrating the standard Heston model. However since the leverage function is there to correct miss calibration of the underlying pure stochastic volatility anyways very precise estimation or fitting of the stochastic parameters is not strictly necessary and one should be able to get away with any kind of reasonable stochastic volatility parameters as long as they capture the market underlying volatility dynamics reasonably well. Hence it is not a major disadvantage that for the MQR and Tataru-Fisher models the parameters have to be estimated from for example historical data , or in the case of the Tataru-Fisher model roughly calibrated with an approximation formula.
Finally when it comes to producing volatility dynamics widely different from the local volatility

model the Heston SLV model trumps, giving the most parameter freedom for the stochastic component among all three models. But as shown in Section 6.2 also the MQR and Tataru-Fisher models manage to produce dynamic behavior different to the local vol model, even when their parameter freedom is further reduced by setting $\rho = 0$, while the Tataru-Fisher model has the tendency to be closer to the local volatility model than the MQR model, owing to it being the most restricted, local in nature, model among them.
All in all we favour the MQR model for its stability, while still offering a lot of freedom in stochastic dynamics and additionally its log-normal volatility dynamics agreeing well with empirical observations as discussed in 3.1. The reliance on historical estimation for the the stochastic volatility parameters as mentioned is only a minor drawback in our opinion.

Conclusion and further Extensions

In this thesis we have presented a generic framework, based on established methods, for calibrating one factor stochastic local volatility models following the Fokker-Planck equation based approach introduced in [5]. Throughout we applied this framework to three different models, two of them, the MQR and Heston SLV model being well studied, while the third the Tataru-Fisher has received less attention. To our knowledge in doing so we have provided the first discussion about boundary conditions for the Tataru-Fisher model and through our implementation shown that also for this model no flux boundary conditions, which have already previously been applied to the Heston SLV and MQR models, are a good choice to provide a stable solution for its associated Fokker-Planck equation.
Furthermore we have shown that with our approach we can calibrate well all three models when the correlation between the stock price process and the volatility is zero. We also identified our approach to be lacking for non zero correlation, in that while the overall shape of the smile is matched decently well especially in the wings we encounter unacceptably large errors when using the calibrated models for Monte Carlo simulations. Here we offered a possible solution approach for further experimentation by suggesting to use ADI schemes with multiple predictor corrector steps to improve the handling of the problematic mixed derivative term as already done in for example [34] and [38].
Finally we verified that all three SLV models considered in this thesis can solve the forward volatility smile flattening problem of the local volatility model.
Some interesting further extensions and improvements to our approach, aside from solving the no correlation problem, would be first off to introduce jumps into the framework, similar to [4], but in this case with the main motivation to include cash dividends into the models, so that single name stocks can be more accurately modeled.
Another interesting extension focused on increasing the efficiency of our approach would be instead of only determining grid boundaries appropriate for the final maturity, and then using this grid for the whole calibration process, to let the grid boundaries $x_{min}, x_{max}, y_{min}$ and y_{max} be time dependent. This would prevent us from working with way too large grids in the beginning when the density is still very concentrated around the initial value and thus wasting computational time by solving the PDE in large areas where the density is still essentially zero.

Appendix A

Technical Details

A.1 Generating grids with ODEs

We now give details on the solution methods for the grid generation ODE (5.1.2).

$$\begin{cases} \frac{dF(u)}{du} = A\left(\sum_{k=1}^{2} \frac{1}{J_k(u)}\right)^{-\frac{1}{2}} \\ J_k(u) := [\beta^2 + (F(u) - B_k)^2] \\ F(0) = y_{min} \\ F(1) = y_{max} \end{cases}$$

We first consider the case with only one concentration point $B_1 \in [y_{min}, y_{max}]$. The ODE is then given by

$$\frac{dF}{du} = A\sqrt{\beta^2 + (F - B_1)^2}.$$

This ODE is separable so we can formally write

$$\int \frac{1}{\sqrt{\beta^2 + (F - B_1)^2}}\, dF = \int A\, du.$$

Now substituting $x := \frac{F - B_1}{\beta}$

$$\int \frac{1}{\sqrt{\beta^2 + (F - B_1)^2}}\, dF = \int \frac{1}{\sqrt{1 + x^2}}\, dx = \operatorname{arcsinh}(x) + c.$$

It follows that $\operatorname{arcsinh}(\frac{F(u) - B_1}{\beta}) = Au + c$ which is equivalent to

$$F(u) = B_1 + \beta \sinh(Au + c).$$

With the two free parameters A, c this can be solved for $F(0) = y_{min}$ and $F(1) = y_{max}$, which yields

$$F(u) = B_1 + \beta \sinh(c_1(1 - u) + c_2 u)$$

for $c_1 = \operatorname{arcsinh}(\frac{y_{min} - B_1}{\beta})$ and $c_2 = \operatorname{arcsinh}(\frac{y_{max} - B_1}{\beta})$.
In the presence of two concentration points $B_1, B_2 \in [y_{min}, y_{max}]$ we can not solve the ODE analytically anymore and numerical methods are needed. Tian [36, Algorithm 2, page 67] suggests a 4th order Rune-Kutta shooting method approach, meaning the ODE is solved numerically with initial condition $F(0) = y_{min}$ for varied A until $|F(1) - y_{max}| \leq \varepsilon$ for some $\varepsilon > 0$. Tian suggests to update A as $A_{\text{new}} = A_{\text{old}} \pm \varepsilon$ depending on if $F_{A_{\text{old}}}(1)$ is smaller or greater than y_{max} respectively. We found this updating behavior to be very slow even for for reasonably large tolerance levels ε, we thus suggest to rather update A using a bisection method, which we found to be considerably faster. Letting the initial search interval be $[A_{min}, A_{max}]$ for $0 < A_{min} < A_{max}$, $G \in \mathbb{R}^{M+1}$ a zero indexed vector, and $J(u) := \left(\sum_{k=1}^{2} \frac{1}{J_k(u)}\right)^{-\frac{1}{2}}$. Our slightly modified version of the grid generation

algorithm discussed in [36, Chatper 5.2.1, Algorithm 2, page 67] is then given by:

Algorithm 3: Grid Generation with multiple concentration Points

```
A ← (A_max + A_min)/2;
Δ ← (y_max − y_min)/M;
y_M ← y_min;
G ← {y_min, 0, ··· , 0};
while |y_m − y_max| > ε do
    for j ∈ {1, ·, M} do
        u ← G_{j−1};
        K_1 ← ΔAJ(u);
        u ← G_{j−1} + K_1/2;
        K_2 ← ΔAJ(u);
        u ← G_{j−1} + K_2/2;
        K_3 ← ΔAJ(u);
        u ← G_{j−1} + K_3;
        K4 ← ΔAJ(u);
        G_j ← G_{j−1} + (1/6)(K_1 + 2K_2 + 2K_3 + K_4);
    y_M ← G_M;
    if y_m − y_max > ε then
        A_max ← A;
    else
        A_min ← A;
    A = (A_max + A_min)/2;
return G;
```

In our implementation we used $A_{max} = 20$ and $A_{min} = 0.1$ with which we obtained quick convergence for all test grids with ε between 0.01 and 0.001.

A.2 Spatial Derivative Approximation Part 2

We provide the proof of 5.1.1.

Proof. Recall that by Taylor's theorem for $(h, l) \in \mathbb{R}^2$ and $(x, y) \in \Omega$ so that $(x + h, y + l) \in \Omega$, as g is assumed to be smooth, suppressing the argument (x, y) for ease of notation,

$$\begin{aligned} g(x+h, y+l) = g(x,y) + h\partial_x g + l\partial_y g + \frac{h^2}{2}\partial_x^2 g + hl\partial_x\partial_y g + \frac{l^2}{2}\partial_y^2 g \\ + \frac{h^3}{6}\partial_x^3 g + \frac{h^2 l}{2}\partial_x^2\partial_y g + \frac{l^2 h}{2}\partial_x\partial_y^2 g + \frac{l^3}{6}\partial_y^3 g + \mathcal{O}(h^4 + l^4). \end{aligned}$$

Thus for $\Delta x := \max\{x_i \,|\, 0 \le i \le N\}$

$$\begin{aligned} &- \frac{\Delta x_i g_{i-1,j}}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})} + \frac{(\Delta x_i - \Delta x_{i-1}) g_{i,j}}{\Delta x_i \Delta x_{i-1}} + \frac{\Delta x_{i-1} g_{i+1,j}}{\Delta x_i(\Delta x_i + \Delta x_{i-1})} \\ &= -\frac{\Delta x_i}{\Delta x_{i-1}(\Delta x_i + \Delta x_{i-1})}\left(g_{i,j} - \Delta x_{i-1}\partial_x g_{i,j} + \frac{(\Delta x_{i-1})^2}{2}\partial_x^2 g_{i,j} + \mathcal{O}((\Delta x)^2)\right) \\ &\quad + \frac{(\Delta x_i - \Delta x_{i-1}) g_{i,j}}{\Delta x_i \Delta x_{i-1}} \\ &\quad + \frac{\Delta x_{i-1}}{\Delta x_i(\Delta x_i + \Delta x_{i-1})}\left(g_{i,j} + \Delta x_i \partial_x g_{i,j} + \frac{(\Delta x_i)^2}{2}\partial_x^2 g_{i,j} + \mathcal{O}((\Delta x)^2)\right) \\ &= \partial_x g_{i,j} + \mathcal{O}((\Delta x)^2). \end{aligned}$$

Furthermore

$$
\begin{aligned}
&\frac{2}{\Delta x_{i-1}(\Delta x_i+\Delta x_{i-1})}g_{i-1,j}-\frac{2}{\Delta x_i\Delta x_{i-1}}g_{i,j}+\frac{2}{\Delta x_i(\Delta x_i+\Delta x_{i-1})}g_{i+1,j}\\
&=\frac{2}{\Delta x_i\Delta x_{i-1}(\Delta x_i+\Delta x_{i-1})}\Big(\Delta x_i[g_{i,j}-\Delta x_{i-1}\partial_x g_{i,j}+\frac{(\Delta x_{i-1})^2}{2}\partial_x^2 g_{i,j}-\frac{(\Delta x_{i-1})^3}{6}\partial^3 g_{i,j}+\mathcal{O}((\Delta x)^4)]\\
&\quad+\Delta x_{i-1}[g_{i,j}+\Delta x_i\partial_x g_{i,j}+\frac{\Delta x_i}{2}\partial_x^2 g_{i,j}+\frac{(\Delta x_i)^3}{6}\partial^3 g_{i,j}+\mathcal{O}((\Delta x)^4)]\Big)\\
&\quad-\frac{2}{\Delta x_{i-1}\Delta x_i}g_{i,j}\\
&=\partial_x^2 g_{i,j}+\frac{1}{3}\frac{(\Delta x_i)^2-(\Delta x_{i-1})^2}{\Delta x_i+\Delta x_{i-1}}\partial_x^3 g_{i,j}+\mathcal{O}((\Delta x)^2)\\
&=\partial_x^2 g_{i,j}+\mathcal{O}(\Delta x).
\end{aligned}
$$

And finally

$$
\begin{aligned}
g_{i+1,j+1}-g_{i-1,j+1}-g_{i+1,j-1}+g_{i-1,j-1}&=(\Delta x_i\Delta y_j+\Delta x_{i-1}\Delta y_j+\Delta x_i\Delta y_{j-1}+\Delta x_{i-1}\Delta y_{j-1})\partial_x\partial_y g_{i,j}\\
&\quad+((\Delta x_i)^2[\Delta y_j+\Delta y_{j-1}]-(\Delta x_{i-1})^2[\Delta y_j+\Delta y_{j-1}])\partial_x^2\partial_y\frac{g_{i,j}}{2}\\
&\quad+((\Delta y_j)^2[\Delta x_i+\Delta x_{i-1}]-(\Delta y_{j-1})^2[\Delta x_i+\Delta x_{i-1}])\partial_x\partial_y^2\frac{g_{i,j}}{2}\\
&\quad+\mathcal{O}((\Delta x)^4+(\Delta y)^4),
\end{aligned}
$$

from which it immediately follows that

$$
\partial_x\partial_y g_{i,j}=\frac{g_{i+1,j+1}-g_{i+1,j-1}-g_{i-1,j+1}+g_{i-1,j-1}}{\Delta x_i\Delta y_j+\Delta x_i\Delta y_{j-1}+\Delta x_{i-1}\Delta x_j+\Delta x_{i-1}\Delta y_{j-1}}+\mathcal{O}(\Delta x+\Delta y)
$$

□

A.3 Finite Difference Coefficients and approximated Initial Condition -MQR

In the MQR model we obtain the interior coefficients in (5.1.7) and (5.1.8) with the choice $f(y)=\exp(y)$, $\mu(y,t)=\kappa(\theta_t-y)$ and $\lambda(y,t)=\lambda$ for some constant $\lambda>0$. For this model the coefficients for no flux boundary conditions my also be found in [34, Section 4.4.2, pages 12-17]. We arrived at the same boundary conditions which are given by: For $1\leq j\leq M-1$ the boundary coefficients on the spot boundaries are given as:

$$
\begin{aligned}
b_0^{j,t}&:=(\frac{1}{\Delta x_0}-\frac{1}{(\Delta x_0)^2})L_{0,t}^2\mathrm{e}^{2y_j}-2\frac{r_t-q_t}{\Delta x_0}-\frac{1}{2}L_{0,t}^2\mathrm{e}^{2y_j}-\frac{2(r_t-q_t)^2}{L_{0,t}^2\mathrm{e}^{2y_j}},\\
c_0^{j,t}&:=\frac{L_{0,t}^2\mathrm{e}^{2y_j}}{(\Delta x_0)^2},\\
a_N^{j,t}&:=\frac{L_{N,t}^2\mathrm{e}^{2y_j}}{(\Delta x_{N-1})^2},\\
b_N^{j,t}&:=(-\frac{1}{\Delta x_{N-1}}-\frac{1}{(\Delta x_{N-1})^2})L_{N,t}^2\mathrm{e}^{2y_j}+2\frac{r_t-q_t}{\Delta x_{N-1}}-\frac{1}{2}L_{N,t}^2\mathrm{e}^{2y_j}-\frac{2(r_t-q_t)^2}{L_{N,t}^2\mathrm{e}^{2y_j}}
\end{aligned}
$$

and at the volatility boundaries as

$$\begin{aligned}
\beta_0^t &= \frac{-\lambda^2}{(\Delta y_0)^2} - \frac{2\kappa}{\lambda^2}\left(\kappa(\theta_t - y_0 + \Delta y_0) + \frac{\lambda^2}{\Delta y_0}\right)(\theta_t - y_0),\\
\gamma_0^t &= \kappa + \frac{\lambda^2}{(\Delta y_0)^2}\\
\alpha_0^t &= \kappa + \frac{\lambda^2}{(\Delta y_{M-1})^2},\\
\beta_0^t &= \frac{-\lambda^2}{(\Delta y_{M-1})^2} - \frac{2\kappa}{\lambda^2}\left(\kappa(\theta_t - y_M - \Delta y_{M-1}) - \frac{\lambda^2}{\Delta y_{M-1}}\right)(\theta_t - y_M).
\end{aligned}$$

Moreover for $1 \le j \le N-1$ we define

$$\begin{aligned}
a_0^{-,t} &:= \frac{\lambda\rho L_{0,t}\mathrm{e}^{y_{j-1}}}{4(\Delta y_{j-1} + \Delta y_j)}\left(\frac{2(r_t - q_t)}{L_{0,t}^2\mathrm{e}^{2y_{j-1}}} + 1\right),\\
a_0^{+,t} &:= -\frac{\lambda\rho L_{0,t}\mathrm{e}^{y_{j+1}}}{4(\Delta y_{j-1} + \Delta y_j)}\left(\frac{2(r_t - q_t)}{L_{0,t}^2\mathrm{e}^{2y_{j+1}}} + 1\right),\\
a_1^{-,t} &:= -\frac{\lambda\rho}{8\Delta x_0(\Delta y_{j-1} + \Delta y_j)}(L_1 - L_0)\mathrm{e}^{y_{j-1}},\\
a_1^{+,t} &:= \frac{\lambda\rho}{8\Delta x_0(\Delta y_{j-1} + \Delta y_j)}(L_1 - L_0)\mathrm{e}^{y_{j+1}},\\
a_{N-1}^{-,t} &:= -\frac{\lambda\rho}{8\Delta x_{N-1}(\Delta y_{j-1} + \Delta y_j)}(L_N - L_{N-1})\mathrm{e}^{y_{j-1}},\\
a_{N-1}^{+,t} &:= \frac{\lambda\rho}{8\Delta x_{N-1}(\Delta y_{j-1} + \Delta y_j)}(L_N - L_{N-1})\mathrm{e}^{y_{j+1}},\\
a_N^{-,t} &:= -\frac{\lambda\rho L_{N,t}\mathrm{e}^{y_{j-1}}}{4(\Delta y_{j-1} + \Delta y_j)}\left(\frac{2(r_t - q_t)}{L_{N,t}^2\mathrm{e}^{2y_{j-1}}} - 1\right),\\
a_N^{+,t} &:= \frac{\lambda\rho L_{N,t}\mathrm{e}^{y_{j+1}}}{4(\Delta y_{j-1} + \Delta y_j)}\left(\frac{2(r_t - q_t)}{L_{N,t}^2\mathrm{e}^{2y_{j+1}}} - 1\right).
\end{aligned}$$

Then

$$\mathcal{L}_{xy}p_{0,j}^t = a_0^{-,t}p_{0,j-1}^t + a_0^{+,t}p_{0,j+1}^t + a_1^{-,t}p_{1,j-1}^t + a_1^{+,t}p_{1,j+1}^t$$

and

$$\mathcal{L}_{xy}p_{N,j}^t = a_{N-1}^{-,t}p_{N-1,j-1}^t + a_{N-1}^{+,t}p_{N-1,j+1}^t + a_N^{-,t}p_{N,j-1}^t + a_N^{+,t}p_{N,j+1}^t.$$

For the volatility boundaries we define for $1 \le i \le N-1$

$$\begin{aligned}
\alpha_0^{-,t} &:= -\frac{\rho L_{i-1,t}\mathrm{e}^{y_0 - \Delta y_0}}{2(\Delta x_{i-1} + \Delta x_i)}\frac{\kappa(\theta_t - y_0)}{\lambda},\\
\alpha_0^{+,t} &:= \frac{\rho L_{i+1,t}\mathrm{e}^{y_0 - \Delta y_0}}{2(\Delta x_{i-1} + \Delta x_i)}\frac{\kappa(\theta_t - y_0)}{\lambda},\\
\alpha_1^{-,t} &:= -\frac{\lambda\rho}{8\Delta y_0(\Delta x_{i-1} + \Delta x_i)}L_{i-1,t}(\mathrm{e}^{y_1} - \mathrm{e}^{y_0 - \Delta y_0}),\\
\alpha_1^{+,t} &:= \frac{\lambda\rho}{8\Delta y_0(\Delta x_{i-1} + \Delta x_i)}L_{i+1,t}(\mathrm{e}^{y_1} - \mathrm{e}^{y_0 - \Delta y_0}),\\
\alpha_{M-1}^{-,t} &:= -\frac{\lambda\rho}{8\Delta y_{M-1}(\Delta x_{i-1} + \Delta x_i)}L_{i-1,t}(-\mathrm{e}^{y_{M-1}} + \mathrm{e}^{y_M + \Delta y_{M-1}}),\\
\alpha_{M-1}^{+,t} &:= \frac{\lambda\rho}{8\Delta y_{M-1}(\Delta x_{i-1} + \Delta x_i)}L_{i+1,t}(-\mathrm{e}^{y_{M-1}} + \mathrm{e}^{y_M + \Delta y_{M-1}}),\\
\alpha_M^{-,t} &:= -\frac{\rho L_{i-1,t}\mathrm{e}^{y_M + \Delta y_{M-1}}}{2(\Delta x_{i-1} + \Delta x_i)}\frac{\kappa(\theta_t - y_M)}{\lambda},\\
\alpha_M^{+,t} &:= \frac{\rho L_{i+1,t}\mathrm{e}^{y_M + \Delta y_{M-1}}}{2(\Delta x_{i-1} + \Delta x_i)}\frac{\kappa(\theta_t - y_M)}{\lambda}.
\end{aligned}$$

We then have

$$\mathcal{L}_{xy}p^t_{i,0} = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_0^{-,t}p^t_{i-1,0} + \alpha_0^{+,t}p^t_{i+1,0} + \alpha_1^{-,t}p^t_{i-1,1} + \alpha_1^{+,t}p^t_{i+1,1})$$

and

$$\mathcal{L}_{xy}p^t_{i,M} = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_{M-1}^{-,t}p^t_{i-1,0} + a_{M-1}^{+,t}p^t_{i+1,0} + \alpha_M^{-,t}p^t_{i-1,M} + \alpha_M^{+,t}p^t_{i+1,M}).$$

Applying the normal approximation (5.2.3) gives us, since $Y_0 = 0$,

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} [(r_0 - q_0) - \frac{1}{2}\sigma^2_{loc}(S_0,0)]\Delta t_0 \\ \kappa\theta_0\Delta t_0 \end{bmatrix}, \Delta t_0 \begin{bmatrix} \sigma^2_{loc}(S_0,0) & \rho\sigma_{loc}(S_0,0)\lambda \\ \rho\sigma_{loc}(S_0,0)\lambda & \lambda^2 \end{bmatrix}\right).$$

then conditional on X_1 Y_1 is distributed as

$$Y_1|X_1 \sim \mathcal{N}\left(\kappa\theta_0\Delta t_0 + \frac{\lambda\rho}{\sigma_{loc}(S_0,0)}(X_1 - [(r_0 - q_0) + \frac{1}{2}\sigma_{loc}(S_0,0)]\Delta t_0), \Delta t_0(1-\rho^2)\lambda^2\right)$$

so

$$\begin{aligned}\psi(x,t) &= \mathbb{E}[\mathrm{e}^{2Y_1}|X_1 = x] \\ &= \exp\left(2\left[\kappa\theta_0\Delta t_0 + \frac{\lambda\rho}{\sigma_{loc}(S_0,0)}(x - [(r_0 - q_0) + \frac{1}{2}\sigma_{loc}(S_0,0)]\Delta t_0)\right] + 2\Delta t_0(1-\rho^2)\lambda^2\right)\end{aligned}$$

A.4 Finite Difference Coefficients and approximated Initial Condition -Tataru-Fisher

In the Tataru-Fisher model we obtain the interior coefficients in (5.1.7) and (5.1.8) with the choice $f(y) = y$, $\mu(y,t) = (1-y)$ and $\lambda(y,t) = \lambda y$ for some constant $\lambda > 0$. For $1 \le j \le M-1$ the boundary coefficients on the spot boundaries are given as:

$$\begin{aligned}
b_0^{j,t} &:= (\frac{1}{\Delta x_0} - \frac{1}{(\Delta x_0)^2})L^2_{0,t}y_j^2 - 2\frac{r_t - q_t}{\Delta x_0} - \frac{1}{2}L^2_{0,t}y_j^2 - \frac{2(r_t-q_t)^2}{L^2_{0,t}y_j^2}, \\
c_0^{j,t} &:= \frac{L^2_{0,t}y_j^2}{(\Delta x_0)^2}, \\
a_N^{j,t} &:= \frac{L^2_{N,t}y_j^2}{(\Delta x_{N-1})^2}, \\
b_N^{j,t} &:= (-\frac{1}{\Delta x_{N-1}} - \frac{1}{(\Delta x_{N-1})^2})L^2_{N,t}y_j^2 + 2\frac{r_t - q_t}{\Delta x_{N-1}} - \frac{1}{2}L^2_{N,t}y_j^2 - \frac{2(r_t-q_t)^2}{L^2_{N,t}y_j^2}
\end{aligned}$$

and at the volatility boundaries as

$$\begin{aligned}
\beta_0^t &= (\Delta y_0 - y_0 + 1)\frac{2y_0 - 2}{\lambda^2(\Delta y_0 - y_0)^2} + \frac{2y_0 - 2}{\Delta y_0} - \frac{\lambda^2 y_0^2}{(\Delta y_0)^2}, \\
\gamma_0^t &= \frac{y_1 - 1}{2\Delta y_0} + (\Delta y_0 - y_0 + 1)\frac{y_1^2}{2\Delta y_0(\Delta y_0 - y_0)^2} + \frac{\lambda^2 y_1^2}{\Delta y_0^2}, \\
\alpha_0^t &= -\frac{y_{M-1} - 1}{2\Delta y_{M-1}} + (\Delta y_{M-1} + y_M - 1)\frac{y_{M-1}^2}{2\Delta y_{M-1}(\Delta y_{M-1} + y_M)^2} + \frac{\lambda^2 y_{M-1}^2}{(\Delta y_{M-1})^2}, \\
\beta_0^t &= (\Delta y_{M-1} + y_M - 1)\frac{-2y_M + 2}{\lambda^2(\Delta y_{M-1} + y_M)^2} + \frac{(-2y_M + 2)}{\Delta y_{M-1}} - \frac{\lambda^2 y_M^2}{(\Delta y_{M-1})^2}.
\end{aligned}$$

Moreover for $1 \leq j \leq N-1$ we define

$$\begin{aligned}
a_0^{-,t} &:= -4(r_t - q_t)\Delta x_0 + 2L_{0,t}^2 \Delta x_0 y_{j-1}^2, \\
a_0^{+,t} &:= 4(r_t - q_t)\Delta x_0 - 2L_{0,t}^2 \Delta x_0 y_{j+1}^2, \\
a_1^{-,t} &:= L_{1,t}^2 y_{j-1}^2 - L_{0,t} L_{1,t} y_{j-1}^2, \\
a_1^{+,t} &:= -L_{1,t}^2 y_{j+1}^2 + L_{0,t} L_{1,t} y_{j+1}^2, \\
a_{N-1}^{-,t} &:= L_{N-1,t}^2 y_{j-1}^2 - L_{N,t} L_{N-1,t} y_{j-1}^2, \\
a_{N-1}^{+,t} &:= -L_{N-1,t}^2 y_{j+1}^2 + L_{N,t} L_{N-1,t} y_{j+1}^2, \\
a_N^{-,t} &:= 4(r_t - q_t)\Delta x_{N-1} - 2L_{N,t}^2 \Delta x_{N-1} y_{j-1}^2, \\
a_N^{+,t} &:= -4(r_t - q_t)\Delta x_{N-1} + 2L_{N,t}^2 \Delta x_{N-1} y_{j+1}^2.
\end{aligned}$$

Then

$$\mathcal{L}_{xy} p_{0,j}^t = \frac{\rho\lambda}{8L_{0,t}\Delta x_0(\Delta y_j + \Delta y_{j-1})}(a_0^{-,t} p_{0,j-1}^t + a_0^{+,t} p_{0,j+1}^t + a_1^{-,t} p_{1,j-1}^t + a_1^{+,t} p_{1,j+1}^t)$$

and

$$\mathcal{L}_{xy} p_{N,j}^t = \frac{\rho\lambda}{8L_{N,t}\Delta x_{N-1}(\Delta y_j + \Delta y_{j-1})}(a_{N-1}^{-,t} p_{N-1,j-1}^t + a_{N-1}^{+,t} p_{N-1,j+1}^t + a_N^{-,t} p_{N,j-1}^t + a_N^{+,t} p_{N,j+1}^t).$$

For the volatility boundaries we define for $1 \leq i \leq N-1$

$$\begin{aligned}
\alpha_0^{-,t} &:= -Li - 1, t(1 - y_0), \\
\alpha_0^{+,t} &:= L_{i+1,t}(1 - y_0), \\
\alpha_1^{-,t} &:= 0, \\
\alpha_1^{+,t} &:= 0, \\
\alpha_{M-1}^{-,t} &:= 0, \\
\alpha_{M-1}^{+,t} &:= 0, \\
\alpha_M^{-,t} &:= -L_{i-1,t}(1 - y_M), \\
\alpha_M^{+,t} &:= L_{i+1,t}(1 - y_M).
\end{aligned}$$

We then have

$$\mathcal{L}_{xy} p_{i,0}^t = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_0^{-,t} p_{i-1,0}^t + \alpha_0^{+,t} p_{i+1,0}^t + \alpha_1^{-,t} p_{i-1,1}^t + \alpha_1^{+,t} p_{i+1,1}^t)$$

and

$$\mathcal{L}_{xy} p_{i,M}^t = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_{M-1}^{-,t} p_{i-1,0}^t + a_{M-1}^{+,t} p_{i+1,0}^t + \alpha_M^{-,t} p_{i-1,M}^t + \alpha_M^{+,t} p_{i+1,M}^t).$$

Applying the normal approximation (5.2.3) gives us, since $Y_0 = 1$,

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} [(r_0 - q_0) - \frac{1}{2}\sigma_{loc}^2(S_0, 0)]\Delta t_0 \\ 1 \end{bmatrix}, \Delta t_0 \begin{bmatrix} \sigma_{loc}^2(S_0, 0) & \rho\sigma_{loc}(S_0, 0)\lambda \\ \rho\sigma_{loc}(S_0, 0)\lambda & \lambda^2 \end{bmatrix} \right).$$

then conditional on X_1 Y_1 is distributed as

$$Y_1 | X_1 \sim \mathcal{N}\left(1 + \frac{\lambda\rho}{\sigma_{loc}(S_0, 0)}(X_1 - [(r_0 - q_0) + \frac{1}{2}\sigma_{loc}(S_0, 0)]\Delta t_0), \Delta t_0(1 - \rho^2)\lambda^2\right)$$

so

$$\begin{aligned}
\psi(x, t) &= \mathbb{E}[Y_1^2 | X_1 = x] \\
&= \Delta t_0(1 - \rho^2)\lambda^2 + \left(1 + \frac{\lambda\rho}{\sigma_{loc}(S_0, 0)}(x - [(r_0 - q_0) + \frac{1}{2}\sigma_{loc}(S_0, 0)]\Delta t_0)\right)^2
\end{aligned}$$

A.5 Finite Difference Coefficients Heston-SLV

For the Heston SLV model we obtain the interior coefficients in (5.1.7) and (5.1.8) with the choice $f(y) = \exp(\frac{1}{2}y)$, $\mu(y,t) = (\kappa\theta - \frac{1}{2}\lambda^2)\mathrm{e}^{-y} - \kappa$ and $\lambda(y,t) = \lambda\mathrm{e}^{-\frac{1}{2}y^2}$ for some constant $\lambda > 0$. For $1 \leq j \leq M-1$ the boundary coefficients on the spot boundaries are given as:

$$
\begin{aligned}
b_0^{j,t} &:= (\frac{1}{\Delta x_0} - \frac{1}{(\Delta x_0)^2})L_{0,t}^2\mathrm{e}^{y_j} - 2\frac{r_t - q_t}{\Delta x_0} - \frac{1}{2}L_{0,t}^2\mathrm{e}^{y_j} - \frac{2(r_t-q_t)^2}{L_{0,t}^2\mathrm{e}^{y_j}},\\
c_0^{j,t} &:= \frac{L_{0,t}^2\mathrm{e}^{y_j}}{(\Delta x_0)^2},\\
a_N^{j,t} &:= \frac{L_{N,t}^2\mathrm{e}^{y_j}}{(\Delta x_{N-1})^2},\\
b_N^{j,t} &:= (-\frac{1}{\Delta x_{N-1}} - \frac{1}{(\Delta x_{N-1})^2})L_{N,t}^2\mathrm{e}^{y_j} + 2\frac{r_t - q_t}{\Delta x_{N-1}} - \frac{1}{2}L_{N,t}^2\mathrm{e}^{y_j} - \frac{2(r_t-q_t)^2}{L_{N,t}^2\mathrm{e}^{y_j}}
\end{aligned}
$$

and at the volatility boundaries as

$$
\begin{aligned}
\beta_0^t &= -\frac{\lambda^2}{(\Delta y_0)^2\mathrm{e}^{y_0}} + \frac{2\kappa(-\mathrm{e}^{-y_0}\theta + 1) + \lambda^2\mathrm{e}^{-y_0})}{\Delta y_0} + \frac{2\kappa^2}{\lambda^2}(-\theta^2\mathrm{e}^{-y_0} + \theta + \theta\mathrm{e}^{-\Delta y_0} - \mathrm{e}^{-\Delta y_0 + y_0})\\
&\quad + \kappa(2\theta\mathrm{e}^{-y_0} - 1 - \mathrm{e}^{-\Delta y_0}) - \frac{1}{2}\lambda^2\mathrm{e}^{-y_0},\\
\gamma_0^t &= \frac{\lambda^2}{(\Delta y_0)^2\mathrm{e}^{y_1}} + \kappa\frac{1-\mathrm{e}^{-2\Delta y_0}}{2\Delta y_0},\\
\alpha_0^t &= \frac{\lambda^2}{(\Delta y_{M-1})^2\mathrm{e}^{y_{M-1}}} + \kappa\frac{\mathrm{e}^{2\Delta y_{M-1}} - 1}{2\Delta y_{M-1}},\\
\beta_0^t &= -\frac{\lambda^2}{(\Delta y_{M-1})^2\mathrm{e}^{y_M}} + \frac{2\kappa(\mathrm{e}^{-y_M}\theta - 1) - \lambda^2\mathrm{e}^{-y_M})}{\Delta y_{M-1}} + \frac{2\kappa^2}{\lambda^2}(-\theta^2\mathrm{e}^{-y_M} + \theta + \theta\mathrm{e}^{-\Delta y_{M-1}} - \mathrm{e}^{\Delta y_{M-1} + y_M})\\
&\quad + \kappa(2\theta\mathrm{e}^{-y_M} - 1 - \mathrm{e}^{-\Delta y_{M-1}}) - \frac{1}{2}\lambda^2\mathrm{e}^{-y_M}.
\end{aligned}
$$

Moreover for $1 \leq j \leq N-1$ we define

$$
\begin{aligned}
a_0^{-,t} &:= \frac{\lambda\rho}{4L_{0,t}(\Delta y_{j-1} + \Delta y_j)}(-2(r_t - q_t)\mathrm{e}^{-y_{j-1}} + L_{0,t}^2),\\
a_0^{+,t} &:= \frac{\lambda\rho}{4L_{0,t}(\Delta y_{j-1} + \Delta y_j)}(2(r_t - q_t)\mathrm{e}^{-y_{j+1}} - L_{0,t}^2),\\
a_1^{-,t} &:= \frac{L_{1,t}\lambda\rho(L_{1,t} - L_{0,t})}{8L_{0,t}\Delta x_0(\Delta y_{j-1} + \Delta y_j)},\\
a_1^{+,t} &:= \frac{L_{1,t}\lambda\rho(-L_{1,t} + L_{0,t})}{8L_{0,t}\Delta x_0(\Delta y_{j-1} + \Delta y_j)},\\
a_{N-1}^{-,t} &:= \frac{L_{N-1,t}\lambda\rho(L_{N,t} - L_{N-1,t})}{8L_{N,t}\Delta x_{N-1}(\Delta y_{j-1} + \Delta y_j)},\\
a_{N-1}^{+,t} &:= \frac{L_{N-1,t}\lambda\rho(-L_{N,t} + L_{N-1,t})}{8L_{N,t}\Delta x_{N-1}(\Delta y_{j-1} + \Delta y_j)},\\
a_N^{-,t} &:= \frac{\lambda\rho}{4L_{N,t}(\Delta y_{j-1} + \Delta y_j)}(-2(r_t - q_t)\mathrm{e}^{-y_{j-1}} + L_{N,t}^2),\\
a_N^{+,t} &:= \frac{\lambda\rho}{4L_{N,t}(\Delta y_{j-1} + \Delta y_j)}(2(r_t - q_t)\mathrm{e}^{-y_{j+1}} - L_{N,t}^2).
\end{aligned}
$$

Then

$$\mathcal{L}_{xy}p_{0,j}^t = a_0^{-,t}p_{0,j-1}^t + a_0^{+,t}p_{0,j+1}^t + a_1^{-,t}p_{1,j-1}^t + a_1^{+,t}p_{1,j+1}^t$$

and

$$\mathcal{L}_{xy}p_{N,j}^t = a_{N-1}^{-,t}p_{N-1,j-1}^t + a_{N-1}^{+,t}p_{N-1,j+1}^t + a_N^{-,t}p_{N,j-1}^t + a_N^{+,t}p_{N,j+1}^t.$$

For the volatility boundaries we define for $1 \leq i \leq N-1$

$$\begin{aligned}
\alpha_0^{-,t} &:= \frac{L_{i-1,t}\rho \mathrm{e}^{-\Delta y_0}}{4\lambda(\Delta x_{i-1}+\Delta x_i)}(-2\kappa\theta + 2\kappa \mathrm{e}^{y_0} + \lambda^2),\\
\alpha_0^{+,t} &:= \frac{L_{i+1,t}\rho \mathrm{e}^{-\Delta y_0}}{4\lambda(\Delta x_{i-1}+\Delta x_i)}(2\kappa\theta - 2\kappa \mathrm{e}^{y_0} - \lambda^2),\\
\alpha_1^{-,t} &:= \frac{L_{i-1,t}\lambda\rho}{8\Delta y_0(\Delta x_{i-1}+\Delta x_i)}(\mathrm{e}^{y_0-\Delta y_0-y_1} - 1),\\
\alpha_1^{+,t} &:= \frac{L_{i+1,t}\lambda\rho}{8\Delta y_0(\Delta x_{i-1}+\Delta x_i)}(-\mathrm{e}^{y_0-\Delta y_0-y_1} + 1),\\
\alpha_{M-1}^{-,t} &:= \frac{L_{i-1,t}\lambda\rho}{8\Delta y_{M-1}(\Delta x_{i-1}+\Delta x_i)}(-\mathrm{e}^{y_M+\Delta y_{M-1}-y_{M-1}} + 1),\\
\alpha_{M-1}^{+,t} &:= \frac{L_{i-1,t}\lambda\rho}{8\Delta y_{M-1}(\Delta x_{i-1}+\Delta x_i)}(\mathrm{e}^{y_M+\Delta y_{M-1}-y_{M-1}} - 1),\\
\alpha_M^{-,t} &:= \frac{L_{i-1,t}\rho \mathrm{e}^{\Delta y_{M-1}}}{4\lambda(\Delta x_{i-1}+\Delta x_i)}(-2\kappa\theta + 2\kappa \mathrm{e}^{y_M} + \lambda^2),\\
\alpha_M^{+,t} &:= \frac{L_{i-1,t}\rho \mathrm{e}^{\Delta y_{M-1}}}{4\lambda(\Delta x_{i-1}+\Delta x_i)}(2\kappa\theta - 2\kappa \mathrm{e}^{y_M} - \lambda^2).
\end{aligned}$$

We then have

$$\mathcal{L}_{xy}p_{i,0}^t = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_0^{-,t}p_{i-1,0}^t + \alpha_0^{+,t}p_{i+1,0}^t + \alpha_1^{-,t}p_{i-1,1}^t + \alpha_1^{+,t}p_{i+1,1}^t)$$

and

$$\mathcal{L}_{xy}p_{i,M}^t = \frac{\rho}{2\lambda(\Delta x_i + \Delta x_{i-1})}(\alpha_{M-1}^{-,t}p_{i-1,0}^t + a_{M-1}^{+,t}p_{i+1,0}^t + \alpha_M^{-,t}p_{i-1,M}^t + \alpha_M^{+,t}p_{i+1,M}^t).$$

Applying the normal approximation (5.2.3) gives us, since $Y_0 = \log(V_0)$,

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} [(r_0-q_0) - \frac{1}{2}\sigma_{loc}^2(S_0,0)]\Delta t_0 \\ \log(V_0) + [(\kappa\theta - \frac{1}{2}\lambda)\frac{1}{V_0} - \kappa]\Delta t_0 \end{bmatrix}, \Delta t_0 \begin{bmatrix} \sigma_{loc}^2(S_0,0) & \rho\sigma_{loc}(S_0,0)\lambda\frac{1}{\sqrt{V_0}} \\ \rho\sigma_{loc}(S_0,0)\lambda\frac{1}{\sqrt{V_0}} & \lambda^2\frac{1}{V_0} \end{bmatrix} \right).$$

then conditional on X_1 Y_1 is distributed as

$$\begin{aligned}
Y_1|X_1 \sim \mathcal{N}\Bigg(&\log(V_0) + [(\kappa\theta - \frac{1}{2}\lambda)\frac{1}{V_0} - \kappa]\Delta t_0 + \frac{\lambda\rho}{\sqrt{V_0}\sigma_{loc}(S_0,0)}(X_1 - [(r_0-q_0) + \frac{1}{2}\sigma_{loc}(S_0,0)]\Delta t_0),\\
&(1-\rho^2)\Delta t_0 \frac{\lambda^2}{V_0}\Bigg)
\end{aligned}$$

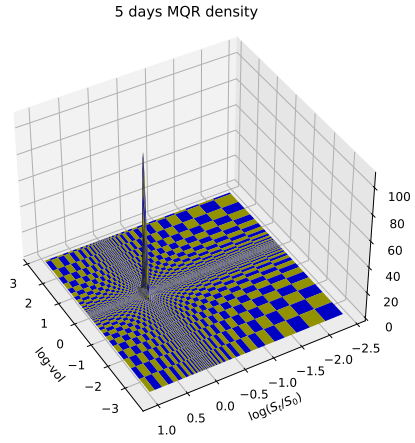
so

$$\begin{aligned}
\psi(x,t) &= \mathbb{E}[\mathrm{e}^{Y_1}|X_1 = x]\\
&= \exp\Bigg(\log(V_0) + [(\kappa\theta - \frac{1}{2}\lambda)\frac{1}{V_0} - \kappa]\Delta t_0\\
&\quad + \frac{\lambda\rho}{\sqrt{V_0}\sigma_{loc}(S_0,0)}(x - [(r_0-q_0) + \frac{1}{2}\sigma_{loc}(S_0,0)]\Delta t_0) + \quad \frac{1}{2}(1-\rho^2)\Delta t_0\frac{\lambda^2}{V_0}\Bigg)
\end{aligned}$$

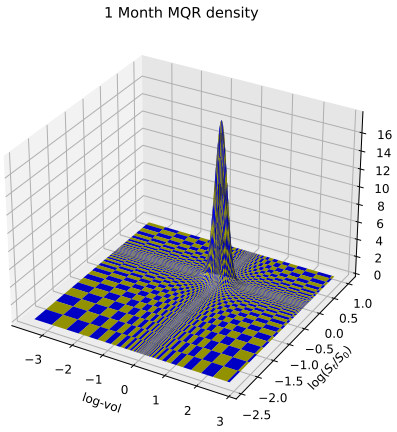
Appendix B

Additional Figures and Tables

B.1 Density Plots MQR



(a) 5 day density MQR.



(b) 1 month density MQR.

Figure B.1: MQR density at different maturities for SPX fit, $\kappa = 0.24$, $\lambda = 0.71$, $\rho = -0.23$

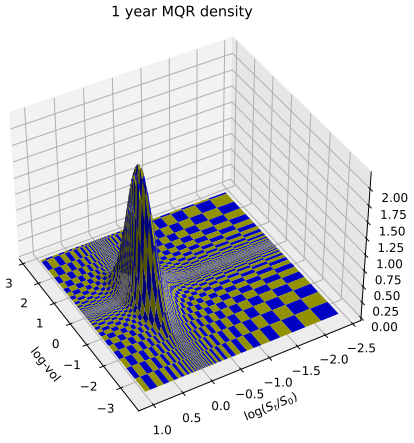
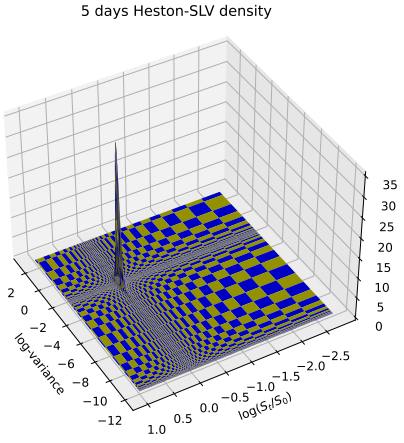
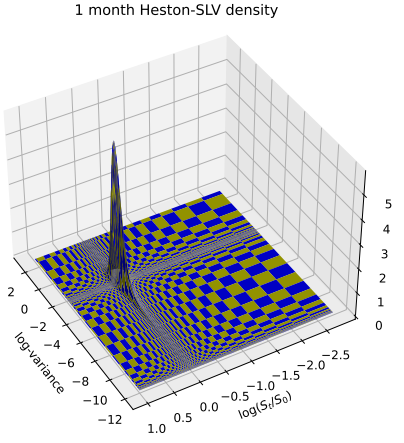


Figure B.1: MQR density 1y

B.2 Density Plots Heston SLVM



(a) 5 day density Heston-SLV.



(b) 1 month density Heston-SLV.

Figure B.2: Heston-SLV density at different maturities for SPX fit, $\kappa = 1.2$, $\theta = 0.15$, $\lambda = 0.7$, $\rho = -0.23$

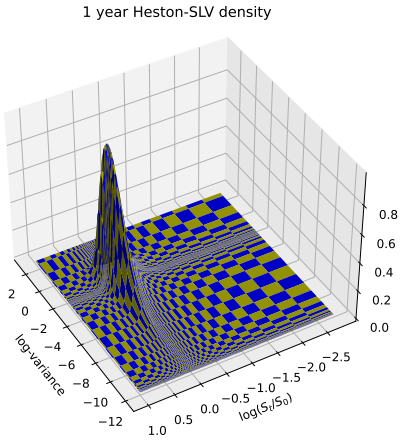


Figure B.2: 1y density Heston-SLV.

B.3 Heston SLV Leverage Function

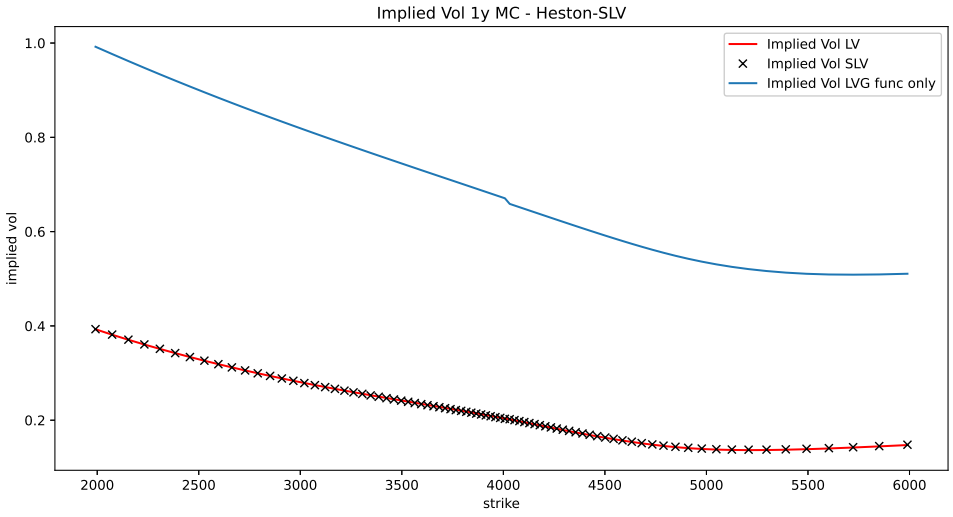
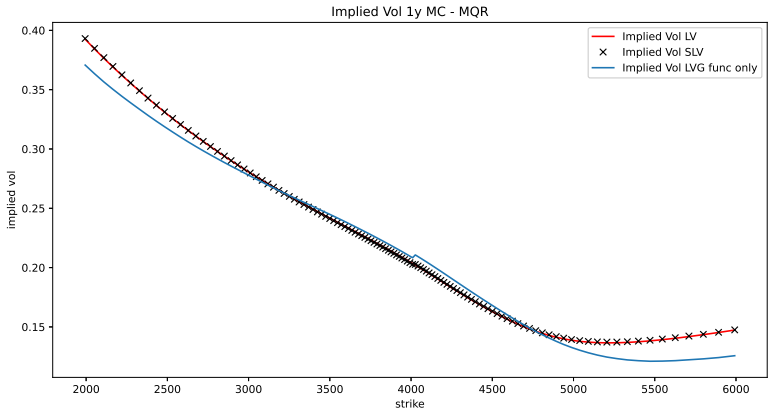
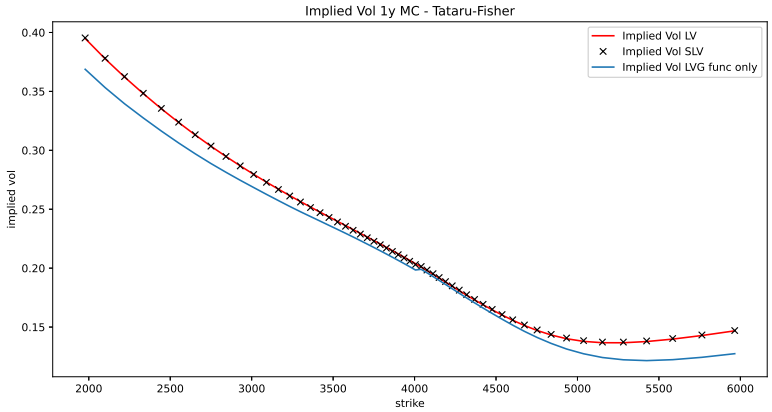


Figure B.3: 1 year maturity Monte Carlo implied volatilities $\sigma_{imp}^{\text{SLV}}$, σ_{imp}^{LV} and $\sigma_{imp}^{\text{LVG}}$ for the Heston SLV model. Correlation $\rho = 0$

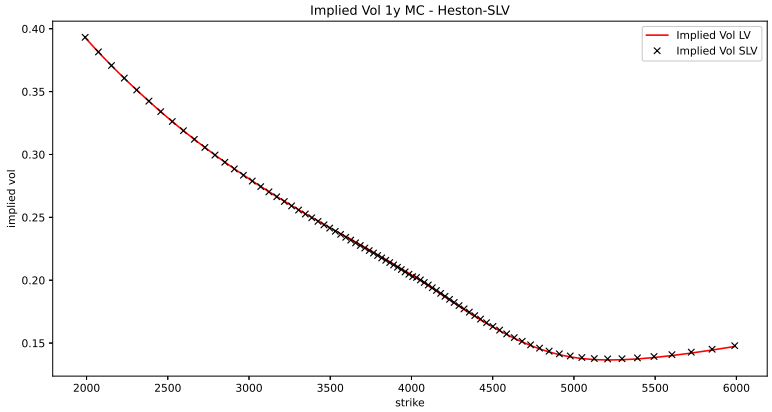
B.4 1 Year implied Volatility 1 - Monte Carlo



(a) 1 year maturity Monte Carlo implied volatilities $\sigma_{\text{imp}}^{\text{SLV}}$, $\sigma_{\text{imp}}^{\text{LV}}$ and $\sigma_{\text{imp}}^{\text{LVG}}$ for the MQR model with 0 correlation



(b) 1 year maturity Monte Carlo implied volatilities $\sigma_{\text{imp}}^{\text{SLV}}$, σ_{imp}^{LV} and $\sigma_{\text{imp}}^{\text{LVG}}$ for the Tataru-Fisher model with 0 correlation.



(c) 1 year maturity Monte Carlo implied volatilities $\sigma_{\text{imp}}^{\text{SLV}}$ and $\sigma_{\text{imp}}^{\text{LV}}$ for the Heston SLV model with 0 correlation.

Figure B.4: SPX implied volatilities computed via Monte Carlo with 100000 sample paths and daily steps, correlation $\rho = 0$. 1 year maturity.

B.5 MQR Leverage Surface

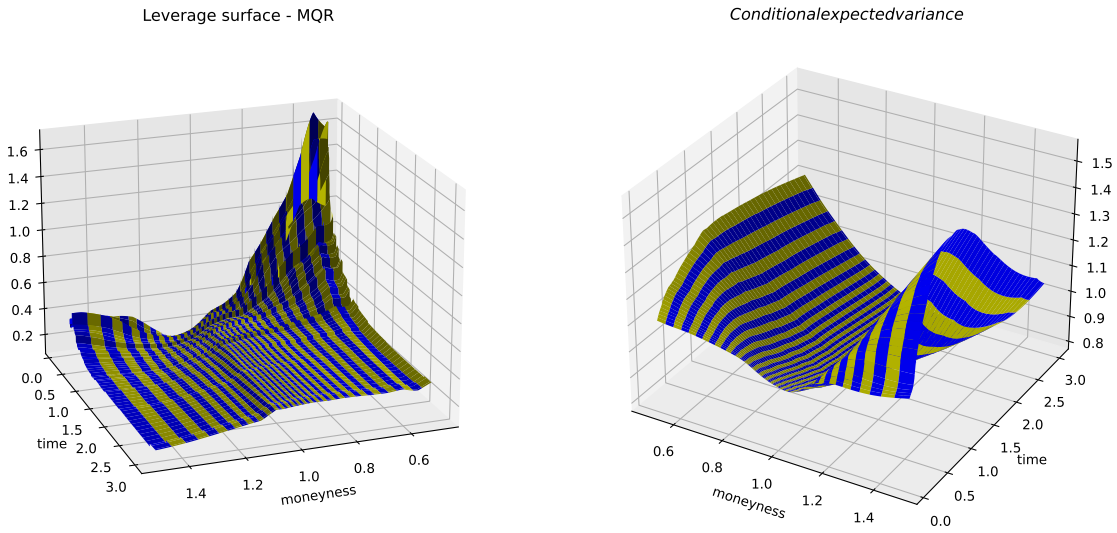


Figure B.5: Leverage surface and square root of the conditional variance $\psi(k,t) = \sqrt{\mathbb{E}[\sigma_t^2 | S_t/S_0 = x]}$ for the MQR model.

B.6 Impact of Number of Time Steps

We give below a table of errors in implied volatility in basis points for a 1 year SPX calibration of the MQR model with parameters $\kappa = 1$, $\lambda = 0.5$, $\rho = -0.31$ and constant rates and dividens $r = 0.05$, $q = 0.02$ for different combinations of N, M, L. Errors were rounded to two decimal places. The error in implied volatility for each L was calculated on the date closest to the given maturity, so technically speaking the combinations are not exactly comparable across different choices of L. Additionally for more grid points in x direction, the error is calculated for more individual strikes in the strike range between 0.5 and 1.5, 0.9 to 1.05 for the 1 week maturity, moneyness. Hence the results are also not exactly comparable across different choices of N either. The table only serves as an illustration of the importance of taking many very small time steps in the beginning of the calibration to gain an acceptable error for close maturities.

(N, M, L)	1w mean	1w max	1m mean	1m max	3m mean	3m max	1y mean	1y max
(500, 500, 80)	56.39	124.95	267.12	641.12	74.45	237.33	9.27	45.39
(500, 500, 100)	27.72	53.53	39.98	89.69	16.64	54.23	2.88	12.19
(500, 500, 150)	23.33	50.42	41.68	75.85	22.85	51.53	4.23	18.32
(500, 500, 200)	16.82	46.94	13.52	41.22	1.53	10.90	1.34	3.95
(500, 500, 250)	10.24	25.48	18.63	32.17	6.40	17.84	1.56	4.05
(500, 500, 300)	10.01	36.83	17.60	37.83	5.17	19.56	1.32	4.97
(500, 500, 365)	12.94	37.26	24.39	39.21	6.94	22.62	1.69	7.76
(500, 500, 400)	11.10	45.67	8.17	28.64	3.72	7.75	1.12	4.78
(500, 500, 550)	10.86	26.99	3.82	15.03	3.08	7.66	1.06	4.18
(500, 500, 730)	9.13	34.57	3.99	19.50	3.13	6.75	0.94	4.76
(600, 600, 80)	56.19	123.60	266.77	640.59	74.40	237.21	9.23	45.34
(600, 600, 100)	27.58	53.45	39.76	89.47	16.60	54.18	2.89	12.16
(600, 600, 150)	23.32	49.37	41.45	75.49	22.80	51.49	4.19	18.30
(600, 600, 200)	16.62	45.83	13.49	40.42	1.48	10.67	1.36	4.14
(600, 600, 250)	9.96	24.28	18.46	31.39	6.40	17.66	1.59	4.40
(600, 600, 300)	9.87	35.57	17.39	37.60	5.12	19.35	1.34	4.94
(600, 600, 365)	12.65	35.96	24.17	38.95	6.88	22.53	1.69	7.74
(600, 600, 400)	10.88	44.53	8.09	27.65	3.70	7.48	1.16	5.12
(600, 600, 550)	10.77	25.88	3.72	14.09	3.01	7.33	1.09	4.53
600, 600, 730)	8.86	33.47	3.79	18.50	3.11	6.45	0.98	5.11

Table B.1: Errors in implied volatility in bps ($10000|\sigma_{\mathrm{imp}}^{\mathrm{SLV}} - \sigma_{\mathrm{imp}}^{\mathrm{LV}}|$) for 1 year MQR calibration to SPX on 2023-02-23 for different number of grid points. For each choice of time steps the error is calculated on the closest grid date to the given maturity.

Bibliography

[1] Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[3] Michal Jex, Richard C. Henderson, and Desheng Wang. Pricing exotics under the smile. http://www.smartquant.com/references/OptionPricing/option14.pdf, 1999.

[4] Alexander Lipton. Masterclass with deutsche bank. the vol smile problem. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 15(2):61–66, 2002.

[5] Yong Ren, Dilip Madan, and M Qian Qian. Calibrating and pricing with embedded local volatility models. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 20(9):138, 2007.

[6] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an ito differential. *Probability Theory and Related Fields*, 71:501–516, 1986.

[7] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 1991.

[8] Toshio Yamada and Shinzo Watanabe. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155 – 167, 1971.

[9] H.H. Kuo. *Introduction to Stochastic Integration*. Universitext. Springer New York, 2006.

[10] Gerard P. Brunick and Steven E. Shreve. Mimicking an itô process by a solution of a stochastic differential equation. *Annals of Applied Probability*, 23:1584–1628, 2010.

[11] Douglas T. Breeden and Robert H. Litzenberger. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651, 1978.

[12] Jim Gatheral. *The volatility surface: a practitioner's guide*. John Wiley & Sons, 2011.

[13] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

[14] Gaoyue Guo, Antoine Jacquier, Claude Martini, and Leo Neufcourt. Generalized arbitrage-free svi volatility surfaces. *SIAM Journal on Financial Mathematics*, 7(1):619–641, 2016.

[15] L. C. G. Rogers and Michael Tehranchi. Can the implied volatility surface move by parallel shifts? *Finance and Stochastics*, 14:235–248, 2010.

[16] P.-L. Lions and M. Musiela. Correlations and bounds for stochastic volatility models. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 24(1):1–16, 2007.

[17] Louis O. Scott. Option pricing when the variance changes randomly: Theory, estimation, and an application. *The Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.

[18] Marc Chesney and Louis Scott. Pricing european currency options: A comparison of the modified black-scholes model and a random variance model. *Journal of Financial and Quantitative Analysis*, 24(3):267–284, 1989.

[19] Martin Tegnér and Rolf Poulsen. Volatility is log-normal—but not for the reason you think. *Risks*, 6(2), 2018.

[20] Nicolas Langrené, Geoffrey Lee, and Zili Zhu. Switching to nonaffine stochastic volatility: A closed-form expansion for the inverse gamma model. *International Journal of Theoretical and Applied Finance*, 19(05):1650031, 2016.

[21] Steven L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 04 2015.

[22] Iain J Clark. *Foreign exchange option pricing: a practitioner's guide*. John Wiley & Sons, 2011.

[23] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*, volume 2. Springer, 2006.

[24] Peter Christoffersen, Kris Jacobs, and Karim Mimouni. Volatility dynamics for the s&p500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, 23(8):3141–3189, 2010.

[25] Lorenzo Bergomi. Smile dynamics i. *Available at SSRN: https://ssrn.com/abstract=1493294*, 2004.

[26] H. P. McKean. A class of markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907–1911, 1966.

[27] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: A review of models, methods and applications. i. models and methods. *Kinetic and Related Models*, 15(6):895–1015, 2022.

[28] R Tachet. *Non-parametric model calibration in finance*. PhD thesis, PhD thesis, Ecole Centrale Paris, 2011.

[29] Mao Fabrice Djete. Non–regular mckean–vlasov equations and calibration problem in local stochastic volatility models, 2022. https://arxiv.org/abs/2208.09986.

[30] Pierre Henry-Labordère. Calibration of local stochastic volatility models to market smiles: A monte-carlo approach. *Risk Magazine, September*, 2009.

[31] Julien Guyon and Pierre Henry-Labordere. The smile calibration problem solved. *Available at SSRN 1885032 https://ssrn.com/abstract=1885032*, 2011.

[32] Christa Cuchiero, Wahid Khosrawi, and Josef Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, 8(4), 2020.

[33] Ivan Guo, Grégoire Loeper, and Shiyi Wang. Calibration of local-stochastic volatility models by optimal transport. *Mathematical Finance*, 32(1):46–77, 2022.

[34] Alexandre Garrivier. A stochastic local volatility calibration. Available at SSRN: https://ssrn.com/abstract=3617929. (May 29, 2020).

[35] Grigore Tataru and Travis Fisher. Stochastic local volatility. *Quantitative Development Group, Bloomberg Version*, 1(February 5), 2010.

[36] Yu Tian. The hybrid stochastic-local volatility model with applications in pricing fx options. *Available at SSRN 2399935 https://ssrn.com/abstract=2399935*, 2013.

[37] Yu Tian, Zili Zhu, Fima Klebaner, and Kais Hamza. A hybrid stochastic volatility model incorporating local volatility. In *2012 Fourth International Conference on Computational and Information Sciences*, pages 333–336, 2012.

[38] K. Spanderen J. Göttker-Schnetmann. Stochastic local volatility in quantlib. https://www.quantlib.org/slides/qlws14/spanderen.pdf. 2014-12-06.

[39] Bernd Engelmann, Frank Koster, and Daniel Oeltz. Calibration of the heston stochastic local volatility model: A finite volume scheme. *International Journal of Financial Engineering*, 8(01):2050048, 2021.

[40] Maarten Wyns and Jacques Du Toit. A finite volume – alternating direction implicit approach for the calibration of stochastic local volatility models. *International Journal of Computer Mathematics*, 94(11):2239–2267, 2017.

[41] DANIELE COZZI. Local stochastic volatility models. solving the smile problem with a nonlinear partial integro-differential equation, 2012. https://www.politesi.polimi.it/handle/10589/72489.

[42] Vladimir Lucic. Boundary conditions for computing densities in hybrid models via pde methods. *Stochastics*, 84(5-6):705–718, 2012.

[43] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. Sympy: symbolic computing in python. *PeerJ Computer Science*, 3:e103, January 2017.

[44] Timothy Sauer. *Numerical analysis*. Addison-Wesley Publishing Company, 2011.

[45] JW Thomas. Numerical partial differential equations: Finite difference methods. *Texts in Applied Mathematics*, 1995.

[46] I.J.D. Craig and A.D. Sneyd. An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Computers & Mathematics with Applications*, 16:341–350, 12 1988.

[47] Jr. Jim Douglas. Alternating direction methods for three space variables. *Numerische Mathematik*, 4:41–63, 1962.

[48] K.J. in 't Hout and C. Mishra. Stability of adi schemes for multidimensional diffusion equations with mixed derivative terms. *Applied Numerical Mathematics*, 74:83–94, 2013.

[49] K.J. in 't Hout and B.D. Welfert. Stability of adi schemes applied to convection–diffusion equations with mixed derivative terms. *Applied Numerical Mathematics*, 57(1):19–35, 2007.

[50] K.J. in 't Hout and B.D. Welfert. Unconditional stability of second-order adi schemes applied to multi-dimensional diffusion equations with mixed derivative terms. *Applied Numerical Mathematics*, 59(3):677–692, 2009. Selected Papers from NUMDIFF-11.