Implied volatility surface: construction methodologies and characteristics

Cristian Homescu[*]

This version: July 9, 2011[†]

The implied volatility surface (IVS) is a fundamental building block in computational finance. We provide a survey of methodologies for constructing such surfaces. We also discuss various topics which influence the successful construction of IVS in practice: arbitrage-free conditions in both strike and time, how to perform extrapolation outside the core region, choice of calibrating functional and selection of numerical optimization algorithms, volatility surface dynamics and asymptotics.

---
[*]Email address: cristian.homescu@gmail.com
[†]Original version: July 9, 2011

Contents

1 Introduction

The geometric Brownian motion dynamics used by Black and Scholes (1973) and Merton (1973) to price options constitutes a landmark in the development of modern quantitative finance. Although it is widely acknowledged that the assumptions underlying the Black-Scholes-Merton model (denoted $\mathcal{BSM}$ for the rest of the paper) are far from realistic, the $\mathcal{BSM}$ formula remains popular with practitioners, for whom it serves as a convenient mapping device from the space of option prices to a single real number called the implied volatility (IV). This mapping of prices to implied volatilities allows for easier comparison of options prices across various strikes, expiries and underlying assets.

When the implied volatilities are plotted against the strike price at a fixed maturity, one often observes a skew or smile pattern, which has been shown to be directly related to the conditional non-normality of the underlying return risk-neutral distribution. In particular, a smile reflects fat tails in the return distribution whereas a skew indicates return distribution asymmetry. Furthermore, how the implied volatility smile varies across option maturity and calendar time reveals how the conditional return distribution non-normality varies across different conditioning horizons and over different time periods. For a fixed relative strike across several expiries one speaks of the term structure of the implied volatility.

We mention a few complications which arise in the presence of smile. Arbitrage may exist among the quoted options. Even if the original market data set does not have arbitrage, the constructed volatility surface may not be arbitrage free. The trading desks need to price European options for strikes and maturities not quoted in the market, as well as pricing and hedging more exotic options by taking the smile into account.

Therefore there are several practical reasons [62] to have a smooth and well-behaved implied volatility surface (IVS):

1. market makers quote options for strike-expiry pairs which are illiquid or not listed;

2. pricing engines, which are used to price exotic options and which are based on far more realistic assumptions than $\mathcal{BSM}$ model, are calibrated against an observed IVS;

3. the IVS given by a listed market serves as the market of primary hedging instruments against volatility and gamma risk (second-order sensitivity with respect to the spot);

4. risk managers use stress scenarios defined on the IVS to visualize and quantify the risk inherent to option portfolios.

The IVS is constructed using a discrete set of market data (implied volatilities or prices) for different strikes and maturities. Typical approaches used by financial institutions are based on:

- (local) stochastic volatility models

- Levy processes (including jump diffusion models)

- direct modeling of dynamics of the implied volatility

- parametric or semi-parametric representations

- specialized interpolation methodologies

Arbitrage conditions may be implicitly or explicitly embedded in the procedure

This paper gives an overview of such approaches, describes characteristics of volatility surfaces and provides practical details for construction of IVS.

2 Volatility surfaces based on (local) stochastic volatility models

A widely used methodology employs formulae based from stochastic volatility models to fit the set of given market data. The result is an arbitrage free procedure to interpolate the implied volatility surface. The most commonly considered stochastic volatility models are Heston and SABR and their extensions (such as time dependent parameters, etc) and we will concentrate on these models as well. Having time dependent parameters allows us to perform calibration in both strike and time directions. This is arguably better than the case of using constant parameter models in capturing inter-dependencies of different time periods. The main disadvantage when using time dependent parameters is the increase in computational time, since in many cases we do not have analytical solutions/approximations and we have to resort to numerical solutions when performing the calibration. However, for the considered time dependent models, namely Heston and SABR, (semi)analytical approximations are available, which mitigates this issue.

We will also consider the hybrid local stochastic volatility models, which are increasingly being preferred by practitioners, and describe efficient calibration procedures for such models.

2.1 Heston model and its extensions

The Heston model is a lognormal model where the square of volatility follows a Cox–Ingersoll–Ross (CIR) process. The call (and put) price has a closed formula through to a Fourier inversion of the characteristic function. Details on efficient ways to compute those formulas are given in [98], recent advances were presented in [93], while [79] contains details for improving the numerical calibration.

It is our experience, confirmed by discussions with other practitioners, that having only one set of parameters is usually not enough to match well market data corresponding to the whole range of expiries used for calibration, especially for FX markets. Consequently we need to consider time dependent parameters.

When the parameters are piecewise constant, one can still derive a recursive closed formula using a PDE/ODE methodology [114] or a Markov argument in combination with affine models [59], but formula evaluation becomes increasingly time consuming.

A better and more general approach is presented in [18], which is based on expanding the price with respect to the volatility of volatility (which is quite small in practice) and then computing the correction terms using Malliavin calculus. The resulting formula is a sum of two terms: the $\mathcal{BSM}$ price for the model without volatility of volatility, and a correction term that is a combination of Greeks of the leading term with explicit weights depending only on the model parameters.

The model is defined as

$$
\begin{aligned}
dX(t) &= \sqrt{\nu(t)}dW_1(t) - \frac{\nu(t)}{2}dt \\
d\nu(t) &= \kappa(t)\left(\theta(t) - \nu(t)\right)dt + \xi(t)\sqrt{\nu(t)}dW_2(t) \\
d\langle W_1, W_2 \rangle &= \rho(t)dt
\end{aligned}
\tag{2.1}
$$

with initial conditions

$$
\begin{aligned}
X(0) &= x_0 \\
\nu(0) &= \nu_0
\end{aligned}
\tag{2.2}
$$

Using the same notations as in [18], the price for the put option is

$$
Put(K,T) = \exp\left[-\int_0^T r(t)dt\right] \mathbb{E}\left[\left(K - \exp\left[-\int_0^T \left(r(t) - q(t)\right)dt + X(T)\right]^+\right)\right]
$$

where $r(t), q(t)$ are the risk free rate and, respectively, dividend yield, K is the strike and T is the maturity of the option.

There are two assumptions employed in the paper:

1) Parameters of the CIR process verify the following conditions

$$\xi_{inf} > 0$$
$$\left(\frac{2\kappa\theta}{\xi^2}\right)_{inf} \geq 1$$

2) Correlation is bounded away from -1 and 1

$$\|\rho_{sup}\| < 1$$

Under these assumptions, the formula for approximation is

$$P_{BS}\left(x_0, var(T)\right) + \sum_{i=1}^{2} a_i(T) \frac{\partial^{(i+1)} P_{BS}}{\partial x^i \partial y}\left(x_0, var(T)\right) + \sum_{i=1}^{2} b_{2i}(T) \frac{\partial^{(2i+2)} P_{BS}}{\partial x^{2i} \partial y^2}\left(x_0, var(T)\right)$$

where $P_{BS}(x,y)$ is the price in a $\mathcal{BSM}$ model with spot e^x, strike K, total variance y, maturity T and rates r_{eq}, q_{eq} given by

$$r_{eq} = \frac{\int_0^T r(t)dt}{T}$$
$$q_{eq} = \frac{\int_0^T q(t)dt}{T}$$

while $var(T), a_i(T), b_{2i}(T)$ have the following formulas

$$var(T) = \int_0^T \nu_0(t)dt$$

$$a_1(t) = \int_0^T e^{\kappa s}\rho(s)\xi(s)\nu_0(s)\left(\int_s^T e^{-\kappa u}du\right)ds$$

$$a_2(t) = \int_0^T e^{\kappa s}\rho(s)\xi(s)\nu_0(s)\left(\int_s^T \rho(t)\xi(t)\left(\int_t^T e^{-\kappa u}du\right)dt\right)ds$$

$$b_0(t) = \int_0^T e^{2\kappa s}\xi^2(s)\nu_0(s)\left(\int_s^T e^{-\kappa t}\left(\int_t^T e^{-\kappa u}du\right)dt\right)ds$$

$$b_2(T) = \frac{a_1^2(T)}{2}$$

$$\nu_0(t) \triangleq e^{-\kappa t}\left(\nu_0 + \int_0^T \kappa e^{\kappa s}\theta(s)ds\right)$$

The error is shown to be of order $\mathcal{O}\left(\xi_{sup}^3 T^2\right)$

While results are presented for general case, [18] also includes explicit formulas for the case of piecewise constant parameters. Numerical results show that calibration is quite effective and that

the approximation matches well the analytical solution, which requires numerical integration. They report that the use of the approximation formula enables a speed up of the valuation (and thus the calibration) by a factor 100 to 600.

We should also mention the efficient numerical approach presented in [110] for calibration of the time dependent Heston model. The constrained optimization problem is solved with an optimization procedure that combines Gauss-Newton approximation of the Hessian with a feasible point trust region SQP (sequential quadratic programming) technique developed in [146]. As discussed in a later chapter on numerical remarks for calibration, in the case of local minimizer applied to problems with multiple local/global minima, a regularization term has to be added in order to ensure robustness/stability of the solution.

2.2 SABR model and its extensions

The SABR model [85] assumes that the forward asset price $F(t)$ and its instantaneous volatility $\alpha(t)$ are driven by the following system of SDEs:

$$
\begin{aligned}
dF(t) &= \alpha(t)F^\beta(t)dW_1(t) \\
d\alpha(t) &= \nu\alpha(t)dW_2(t) \\
d\langle W_1, W_2 \rangle &= \rho dt
\end{aligned}
\tag{2.3}
$$

where is $\nu > 0$ is volatility of volatility and $\beta > 0$ is a leverage coefficient. The initial conditions are

$$
\begin{aligned}
F(0) &= F_0 \\
\alpha(0) &= \alpha_0
\end{aligned}
$$

Financial interpretation for this model is the following: $\alpha(t)$ determines the overall level of at-the-money forward volatility; β measures skew with two particular choices: $\beta = 1$ corresponding to the log-normal model with a flat skew and $\beta = 0$ corresponding to the normal model with a very pronounced skew; ρ also controls the skew shape with the choice $\rho < 0$ (respectively $\rho > 0$) yielding the negative (respectively inverse) skew and with the choice $\rho = 0$ producing a symmetric volatility smile given $\beta = 1$; ν is a measure of convexity, i.e. stochasticity of $\alpha(t)$.

Essentially, this model assumes CEV distribution (log-normal in case $\beta = 1$) for forward price $F(t)$ and log-normal distribution for instantaneous volatility $\alpha(t)$.

SABR model is widely used by practitioners, due to the fact that it has available analytical approximations. Several approaches were used in the literature for obtaining such approximations: the singular perturbation, heat kernel asymptotics, and Malliavin calculus [94, 115, 85]. Additional higher order approximations are discussed in [119](second order) and [134], up to fifth order. Details for improving the numerical calibration were given in [79]

An extension of SABR (termed lambda-SABR), and corresponding asymptotic approximations were introduced in papers by Henry-Labordere (see chapter 6 of [87]). This model is described as follows (and degenerates into SABR model when $\lambda = 0$)

$$
\begin{aligned}
dF(t) &= \alpha(t)F^\beta(t)dW_1(t) \\
d\alpha(t) &= \lambda\left(\alpha(t) - \bar{\lambda}\right) + \nu\alpha(t)dW_2(t) \\
d\langle W_1, W_2 \rangle &= \rho dt
\end{aligned}
\tag{2.4}
$$

The high order approximations in [134] were obtained for lambda-SABR model. Approximations for extended lambda-SABR model, where a drift term is added to the SDE describing $F(t)$ in (2.4), were presented in [130] and [131]. Approximations for SABR with time dependent coefficients were presented in [118], where the model was named "Dynamic SABR", and in respectively in [80], where the approach was specialized to piecewise constant parameters.

6

If one combines the results from [131, 134, 130] with the findings of [80], the result will be a model (extended lambda-SABR with piecewise constant parameters) that may be rich to capture all desired properties when constructing a volatility surface and yet tractable enough due to analytical approximations.

Alternatively, the results presented in [80] seem very promising and will be briefly described below. The procedure is based on asymptotic expansion of the bivariate transition density [147].

To simplify the notations, the set of SABR parameters is denoted by $\theta \triangleq \{\alpha, \beta, \rho, \nu\}$ and the dependence of the model's joint transition density on the model parameters by $p\left(t, F_0, \alpha_0; T, \hat{F}, \hat{A}; \theta\right)$.

The joint transition density is defined as

$$\mathbb{P}\left(\hat{F} < F(T) \leq \hat{F} + d\hat{F}, \hat{A} \leq \alpha(T) \leq \hat{A} + d\hat{A}\right) \triangleq p\left(t, \hat{\bar{F}}, \hat{\bar{A}}; T, \hat{F}, \hat{A}\right) d\hat{F} d\hat{A}$$

We follow the notations from [147], namely:

- $\hat{F}, \hat{A}$ are forward variables denoting the state values of $F(T), \alpha(T)$

- $\hat{\bar{F}}, \hat{\bar{A}}$ are backward variables denoting the state values of $F(t), \alpha(t)$

Let us denote by $\{T_1, T_2, ..., T_N\}$ the set of expiries for which we have market data we want to calibrate to; we assume that the four SABR parameters $\{\alpha, \beta, \rho. \nu\}$ are piecewise constant on each interval $[T_{i-1}, T_i]$.

The tenor-dependent SABR model then reads

$$
\begin{aligned}
dF(t, T_i) &= \alpha(t, T_i) F^{\beta_i}(t) dW_1(t) \qquad (2.5)\\
d\alpha(t, T_i) &= \nu_i \alpha(t, T_i) dW_2(t)\\
\mathbb{E}^{\mathbb{Q}^{T_i}}\left[dW_1(t) dW_2(t)\right] &= \rho_i dt
\end{aligned}
$$

where $\mathbb{E}^{\mathbb{Q}^{T_i}}$ is the expectation under the T_i forward measure $\mathbb{Q}^{T_i}$

The SDE (2.5) is considered together with

$$
\begin{aligned}
F(0, T_i) &= F_i\\
\alpha(0, T_i) &= \alpha_i
\end{aligned}
$$

The notations for SABR set of parameters and, respectively, for the dependence of the model's joint transition density on the model parameters are updated as follows

$$
\theta(T) = \begin{cases} (\alpha_0, \beta_0, \rho_0, \nu_0) & if\, T \leq T_1 \\ (\alpha_{i-1}, \beta_{i-1}, \rho_{i-1}, \nu_{i-1}) & if\, T_{i-1} < T \leq T_i \\ (\alpha_{N-1}, \beta_{N-1}, \rho_{N-1}, \nu_{N-1}) & if\, T_{N-1} < T \leq T_N \end{cases}
$$

and, respectively

$$
\begin{aligned}
p\left(0, F_0, \alpha_0; T_1, F_1; \theta_0\right)\\
p\left(T_{i-1}, F_{i-1}, A_{i-1}; T_i, F_i, A_i; \theta_{i-1}\right)\\
p\left(T_{N-1}, F_{N-1}, A_{N-1}; T_N, F_N, A_N; \theta_N\right)
\end{aligned}
$$

In the case where only the parameter dependence need to be stressed, we use the shortened notion:

$$
\begin{aligned}
p\left(0, T_1; \theta_0\right)\\
p\left(T_{i-1}, T_i; \theta_{i-1}\right)
\end{aligned}
$$

A standard SABR model describes the dynamics of a forward price process $F(t; T_i)$ maturing at a particular T_i. Forward prices associated with different maturities are martingales with respect to different forward measures defined by different zero-coupon bonds $B(t, T_i)$ as numeraires. This raises consistency issues, on both the underlying and the pricing measure, when we work with multiple option maturities simultaneously.

We address this issue by consolidating all dynamics into those of $F(t, T_N), \alpha(t, T_N)$, whose tenor is the longest among all, and express all option prices at different tenors in one terminal measure $\mathbb{Q}^{T_N}$ which is the one associated with the zero-coupon bond $B(t, T_N)$.

We may do so because we assume

• No-arbitrage between spot price $S(t)$ and all of its forward prices $F(t, T_i), i = 1...N$, at all trading time t;

• Zero-coupon bonds $B(t, T_i)$ are risk-less assets with positive values

Based on these assumptions we obtain the following formulas

$$
\begin{aligned}
F(t, T_1) \quad &= \quad \frac{S}{B(t, T_1)} \\
F(t, T_i) \quad &= F(t, T_N) \quad \frac{B(t, T_N)}{B(t, T_i)}
\end{aligned}
$$

This will enable us to convert an option on $F(\cdot, T_i)$ into an option on $F(\cdot, T_N)$. The price of a call option on $F(\cdot, T_i)$ with strike price K_j and maturity T_i then becomes

$$
V(t, T_i, K_j) = B(t, T_N) \mathbb{E}^{\mathbb{Q}^{T_N}} \left[\left(F(T_i, T_N) - \bar{K}_j \right)^+ | \mathfrak{F}_t \right]
$$

where the adjusted strike $\bar{K}_j$ is defined as

$$
\bar{K}_j \triangleq \frac{K_j}{B(T_i, T_N)}
$$

In the context of model calibration, computation of spot implied volatilities from the model relies on computation of option prices

$$
\mathbb{E}^{\mathbb{Q}^{T_N}} \left[\left(F(T_i) - \bar{K}_j \right)^+ | F_{i-1}, A_{i-1} \right] = \iint_{\mathbb{R}^2_+} \left[\left(F(T_i) - \bar{K}_j \right)^+ p(T_{i-1}, T_i; \theta_{i-1}) \right] dF_i dA_i \tag{2.6}
$$

at each tenor T_i for each equivalent strike $\bar{K}_j$.

Asymptotic expansions of a more generic joint transition density have been obtained analytically in [147] to the nth order. We should note that for simplicity we use exactly the same notations as in section 4.2 of [80], namely that the values of state variables at time t are denoted by f, α and, respectively, the values of the state variables at T are denoted by F, A.

The expansion to second order was shown to give a quite accurate approximation

$$
p_2(t, f, \alpha; T, F, A; \theta) = \frac{1}{\nu T F^\beta A^2} \left[\hat{p}_0 + \nu^2 \hat{p}_1 \sqrt{T} + \nu^2 \hat{p}_2 T \right] (\tau, u, v, \theta)
$$

where we define τ, u, v as

$$
\begin{aligned}
\tau \quad &\triangleq \quad \frac{T - t}{T} \\
u \quad &\triangleq \quad \frac{f^{1-\beta} - F^{1-\beta}}{\alpha(1-\beta)\sqrt{T}} \\
v \quad &\triangleq \quad \frac{\ln\left(\frac{\alpha}{A}\right)}{\nu\sqrt{T}}
\end{aligned}
$$

The terms $\hat{p}_0, \hat{p}_1, \hat{p}_2$ have the following formulae

$$
\begin{aligned}
\hat{p}_0\left(\tau, u, v, \theta\right) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{u^2-2\rho uv+v^2}{2\tau\left(1-\rho^2\right)}\right] \\
\hat{p}_1\left(\tau, u, v, \theta\right) &= \frac{a_{11}+\frac{a_{10}}{\tau}}{2\left(\rho^2-1\right)}\hat{p}_0\left(\tau, u, v, \theta\right) \\
\hat{p}_2\left(\tau, u, v, \theta\right) &= \frac{a_{23}\tau+a_{22}+\frac{a_{21}}{\tau}+\frac{a_{20}}{\tau^2}}{24\left(1-\rho^2\right)^2}\hat{p}_0\left(\tau, u, v, \theta\right)
\end{aligned}
$$

Explicit expressions for the polynomial functions $a_{11}, a_{10}, a_{23}, a_{22}, a_{21}, a_{20}$ are given in Eq. (42) in [147]. In terms of computational cost, it is reported in [80] that it takes about 1-10 milliseconds for an evaluation of the integral (2.6) on a 1000 by 1000 grid, using the approximation $\hat{p}_2$ as density.

2.3 Local stochastic volatility model

More and more practitioners are combining the strengths of local and stochastic volatility models, with the resulting hybrid termed local stochastic volatility (LSV) model.

Based on [110], we describe efficient procedures for calibrating one such model, namely a hybrid Heston plus local volatility model, with dynamics given by

$$
\begin{aligned}
df^{LSV}(t) &= \sigma\left(f^{LSV}(t),t\right)\sqrt{v(t)}f^{LSV}(t)dW_1(t) \\
dv(t) &= \kappa\left(\theta-v(t)\right)dt+\xi\sqrt{v(t)}dW_2(t)
\end{aligned}
$$

The calibration procedure is based on the following 2 step process:

- calibrate stochastic volatility component

- perform LSV correction

The validity of this 2-step process is due to the observation that the forward skew dynamics in stochastic volatility setting are mainly preserved under the LSV correction.

The first approach is based on the "fixed point" concept described in [126]

1. Solve forward Kolmogorov PDE (in $x = \ln\left(S/fwd\right)$ with a given estimate of $\sigma(f,t)$

$$
\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left[\frac{1}{2}v\sigma^2 p\right] - \frac{\partial}{\partial v}\left[\kappa\left(\theta-v\right)p\right] + \frac{\partial^2}{\partial x^2}\left[\frac{1}{2}v\sigma^2 p\right] + \frac{\partial^2}{\partial x\partial v}\left[\rho\sigma\xi vp\right] + \frac{\partial^2}{\partial v^2}\left[\frac{1}{2}v\xi^2 p\right]
$$

2. Use the density from 1. to compute the conditional expected value of $v(t)$ given $f^{LSV}(t)$

$$
E\left[v(t)|f^{LSV}(t)=f\right] = \frac{\int_0^\infty vp(t,f,v)dv}{\int_0^\infty p(t,f,v)dv}
$$

3. Adjust σ according to Gyongy's identity [83] for the local volatilities of the LSV model

$$
\left(\sigma_{LV}^{LSV}\right)^2(f,t) = \sigma^2(f,t)E\left[v(t)|f^{LSV}(t)=f\right] = \left(\sigma_{LV}^{Market}\right)^2(f,t)
$$

4. repeat steps 1.-3. until $\sigma(f,t)$ has converged (it was reported that in most cases 1-2 loops are sufficient)

The second approach is based on "local volatility" ratios, similar to [120, 88]. The main idea is the following: applying Gyongy's theorem [83] twice (for the starting stochastic volatility component and, respectively, for the target LSV model) avoids the need for conditional expectations.

The procedure is as follows

1. Compute the local volatilities of an LSV and an SV model via Gyongy's formula

$$\left(\sigma_{LV}^{LSV}\right)^2 (f,t) = \sigma^2(f,t) E\left[v(t)|f^{LSV}(t)=f\right] = \left(\sigma_{LV}^{Market}\right)^2 (f,t)$$

$$\left(\sigma_{LV}^{SV}\right)^2 (x,t) = E\left[v(t)|f^{SV}(t)=x\right]$$

2. Taking the ratio and solving for the unknown function $\sigma(\cdot,\cdot)$ we obtain

$$\sigma(t,f) = \frac{\sigma_{LV}^{Market}(f,t)}{\sigma_{LV}^{SV}(x,t)} \sqrt{\frac{E\left[v(t)|f^{SV}(t)=x\right]}{E\left[v(t)|f^{LSV}(t)=f\right]}} \approx using \ x = H(f,t) \approx \frac{\sigma_{LV}^{Market}(f,t)}{\sigma_{LV}^{SV}(H(f,t),t)}$$

   with an approximate, strictly monotonically increasing map $H(f,t)$

The calculation is reported to be extremely fast if the starting local volatilities are easy to compute. The resulting calibration leads to near perfect fit of the market

We should also mention a different calibration procedure for a hybrid Heston plus local volatility model, presented in [60].

3 Volatility surfaces based on Levy processes

Volatility surface representations based on Levy processes tend to better handle steep short term skews (observed especially in FX and commodity markets). In a model with continuous paths like a diffusion model, the price process behaves locally like a Brownian motion and the probability that the price of the underlying moves by a large amount over a short period of time is very small, unless one fixes an unrealistically high value of volatility. Thus in such models the prices of short term OTM options are much lower than what one observes in real markets. On the other hand, if price of underlying is allowed to jump, even when the time to maturity is very short, there is a non-negligible probability that after a sudden change in the price of the underlying the option will move in the money.

The Levy processes can be broadly divided into 2 main categories:

- jump diffusion processes: jumps are considered rare events, and in any given finite interval there are only finite many jumps

- infinite activity Levy processes: in any finite time interval there are infinitely many jumps.

The importance of a jump component when pricing options close to maturity is also pointed out in the literature, e.g., [4]. Using implied volatility surface asymptotics, the results from [111] confirm the presence of a jump component when looking at S&P option data .

Before choosing a particular parametrization, one must determine the qualitative features of the model. In the context of Levy-based models, the important questions are [42, 138]:

- Is the model a pure-jump process, a pure diffusion process or a combination of both?

- Is the jump part a compound Poisson process or, respectively, an infinite intensity Levy process?

- Is the data adequately described by a time-homogeneous Levy process or is a more general model may be required?

Well known models based on Levy processes include Variance Gamma [107], Kou [100], Normal Inverse Gaussian [12], Meixner [129, 108], CGMY [33], affine jump diffusions [55].

From a practical point of view, calibration of Levy-based models is definitely more challenging, especially since it was shown in [43, 44] that it is not sufficient to consider only time-homogeneous Levy specifications. Using a non-parametric calibration procedure, these papers have shown that Levy processes reproduce the implied volatility smile for a single maturity quite well, but when it comes to calibrating several maturities at the same time, the calibration by Levy processes becomes much less precise. Thus successful calibration procedures would have to be based on models with more complex characteristics.

To calibrate a jump-diffusion model to options of several maturities at the same time, the model must have a sufficient number of degrees of freedom to reproduce different term structures. This was demonstrated in [139] using the Bates model, for which the smile for short maturities is explained by the presence of jumps whereas the smile for longer maturities and the term structure of implied volatility is taken into account using the stochastic volatility process.

In [74] a stochastic volatility jump diffusion model is calibrated to the whole market implied volatility surface at any given time, relying on the asymptotic behavior of the moments of the underlying distribution. A forward PIDE (Partial Integro-Differential Equation) for the evolution of call option prices as functions of strike and maturity was used in [4] in an efficient calibration to market quoted option prices, in the context of adding Poisson jumps to a local volatility model.

3.1 Implied Levy volatility

An interesting concept was introduced in [45], which introduced the implied Levy space volatility and the implied Levy time volatility, and showed that both implied Levy volatilities would allow an exact fit of the market. Instead of normal distribution, as is the case for implied volatility calculation, their starting point is a distribution that was more in line with the empirical observations.

More specifically, instead of lognormal model they assume the following model

$$S(t) = S_0 \exp\left[(r - q + \omega)\, t + \sigma X(t)\right] \tag{3.1}$$

where $\sigma > 0$, r is the risk-free rate, q is the dividend yield, ω is a term that is added in order to make dynamics risk neutral, and $X = \{X(t), t \geq 0\}$ is a stochastic process that starts at zero, has stationary and independent increments distributed according to the newly selected distribution

Once one has fixed the distribution of X (which we assume as in the Brownian case to have mean zero and variance at unit time equal to 1), for a given market price one can look for the corresponding σ, which is termed the implied (space) Levy volatility, such that the model price matches exactly the market price.

To define Implied Levy Space Volatility, we start with an infinitely divisible distribution with zero mean and variance one and denote the corresponding Levy process by $X = \{X(t), t \geq 0\}$. Hence $E[X(1)] = 0$ and $Var[X(1)]$=0. We denote the characteristic function of $X(1)$ (the mother distribution) by $\phi(u) = E[\exp(iuX(1))]$. We note that, similar to a Brownian Motion, we have $E[X(t)] = 1$ and $Var[X(t)]$=t and hence $Var[\sigma X(t)] = \sigma^2 t$.

If we set ω in (3.1) to be $\omega = -\log(\phi(-\sigma i))$, we call the volatility parameter σ needed to match the model price with a given market price the implied Levy space volatility of the option.

To define the implied Levy time volatility we start from a similar Levy process X and we consider that the dynamics of the underlying are given as

$$
\begin{aligned}
S(t) &= S_0 \exp\left[\left(r - q + \omega\sigma^2\right) t + X(t)\right] \\
\omega &= \log\left(\phi(-i)\right)
\end{aligned}
\tag{3.2}
$$

Given a market price, we now use the terminology of implied Levy time volatility of the option to describe the volatility parameter σ needed to match the model price with the market price. Note that

in the $\mathcal{BSM}$ setting the distribution (and hence also the corresponding vanilla option prices) of $\sigma W(t)$ and $W(\sigma^2 t)$ is the same, namely a $\mathcal{N}(0, \sigma^2 t)$ distribution, but this is not necessary the case for the more general Levy cases.

The price of an European option is done using characteristic functions through the Carr-Madan formula [35] and the procedure is specialized to various Levy processes: normal inverse Gaussian (NIG), Meixner, etc.

4 Volatility surface based on models for the dynamics of implied volatility

In the literature there are two distinct directions for treatment and construction of volatility surfaces [36]. One approach assumes dynamics for the underlying that can accommodate the observed implied volatility smiles and skews. As we have seen in previous chapters, such approaches include stochastic volatility models as well as various Levy processes. The general procedure is to estimate the coefficients of the dynamics through calibration from observed option prices. Another approach relies on explicitly specifying dynamics of the implied volatilities, with models belonging to this class being termed "market models" of implied volatility. In general, this approach assumes that the entire implied volatility surface has known initial value and evolves continuously over time. The approach first specifies the continuous martingale component of the volatility surface, and then derives the restriction on the risk-neutral drift of the surface imposed by the requirement that all discounted asset prices be martingales. Such models are presented in [9, 61, 84, 48] An approach that was described as falling between the two categories was described in [36] and is described next

4.1 Carr and Wu approach

Similar to the market model approach, it directly models the dynamics of implied volatilities. However, instead of taking the initial implied volatility surface as given and infer the risk-neutral drift of the implied volatility dynamics, both the risk-neutral drift and the martingale component of the implied volatility dynamics are specified, from which the allowable shape for the implied volatility surface is derived. The shape of the initial implied volatility surface is guaranteed to be consistent with the specified implied volatility dynamics and, in this sense, this approach is similar to the first category.

The starting point is the assumption that a single standard Brownian motion drives the whole volatility surface, and that a second partially correlated standard Brownian motion drives the underlying security price dynamics. By enforcing the condition that the discounted prices of options and their underlying are martingales under the risk-neutral measure, one obtains a partial differential equation (PDE) that governs the shape of the implied volatility surface, termed as Vega-Gamma-Vanna-Volga (VGVV) methodology, since it links the theta of the options and their four Greeks. Plugging in the analytical solutions for the $\mathcal{BSM}$ Greeks, the PDE is reduced into an algebraic relation that links the shape of the implied volatility surface to its risk-neutral dynamics.

By parameterizing the implied variance dynamics as a mean-reverting square-root process, the algebraic equation simplifies into a quadratic equation of the implied volatility surface as a function of a standardized moneyness measure and time to maturity. The coefficients of the quadratic equation are governed by six coefficients related to the dynamics of the stock price and implied variance. This model is denoted as the square root variance (SRV) model.

Alternatively, if the implied variance dynamics is parametrized as a mean-reverting lognormal process, one obtains another quadratic equation of the implied variance as a function of log strike over spot and time to maturity. The whole implied variance surface is again determined by six coefficients related to the stock price and implied variance dynamics. This model is labeled as the lognormal variance (LNV) model.

The computational cost for calibration is quite small, since computing implied volatilities from each of the two models (SRV and LNV) is essentially based on solving a corresponding quadratic equation.

The calibration is based on setting up a state-space framework by considering the model coefficients as hidden states and regarding the finite number of implied volatility observations as noisy observations. The coefficients are inferred from the observations using an unscented Kalman filter.

Let us introduce the framework now. We note that zero rates are assumed without loss of generality. The dynamics of the stock price of the underlying are assumed to be given by

$$dS(t) = S(t)\sqrt{v(t)}dW(t)$$

with dynamics of the instantaneous return variance $v(t)$ left unspecified.

For each option struck at K and expiring at T, its implied volatility $I(t; K, T)$ follows a continuous process given by

$$dI(t; K, T) = \mu(t)dt + \omega(t)dZ(t)$$

where $Z(t)$ is a Brownian motion.

The drift $\mu(t)$ and volatility of volatility $\omega(t)$ can depend on K, T and $I(t; K, T)$

We also assume that we have the following correlation relationship

$$\rho(t)dt = E[dW(t)dZ(t)]$$

The relationship $I(t; K, T) > 0$ guarantees that there is no static arbitrage between any option at $(K; T)$ and the underlying stock and cash.

It is further required that no dynamic arbitrage (NDA) be allowed between any option at $(K; T)$ and respectively a basis option at $(K_0; T_0)$ and the stock.

For concreteness, let the basis option be a call with $C(t; T, K)$ denoting its value, and let all other options be puts, with $P(t; K, T)$ denoting the corresponding values. We can write both the basis call and other put options in terms of the $\mathcal{BSM}$ put formula:

$$
\begin{aligned}
P(t; K, T) &= BSM(S(t), I(t; K, T), t) \\
C(t; K_0, T_0) &= BSM(S(t), I(t; K_0, T_0), t) + S(t) - K_0
\end{aligned}
$$

We can form a portfolio between the two to neutralize the exposure on the volatility risk dZ

$$\frac{\partial BSM}{\partial \sigma}(S(t), I(t; K, T), t)\omega(K, T) - N^c(t)\frac{\partial BSM}{\partial \sigma}(S(t), I(t; K_0, T_0), t)\omega(K_0, T_0) = 0$$

We can further use $N^S(t)$ shares of the underlying stock to achieve delta neutrality:

$$BSM(S(t), I(t; K, T), t) - N^c(t)[1 + BSM(S(t), I(t; K_0, T_0), t)] - N^S(t) = 0$$

Since shares have no Vega, this three-asset portfolio retains zero exposure to dZ and by construction has zero exposure to dW.

By Ito's lemma, each option in this portfolio has risk-neutral drift (RND) given by

$$RND = \frac{\partial BSM}{\partial t} + \mu(t)\frac{\partial BSM}{\partial \sigma} + \frac{v(t)}{2}S^2(t)\frac{\partial^2 BSM}{\partial S^2} + \rho(t)\omega(t)\sqrt{v(t)}S(t)\frac{\partial^2 BSM}{\partial \sigma \partial S} + \frac{\omega^2(t)}{2}\frac{\partial^2 BSM}{\partial \sigma^2}$$

Note: For simplicity of notations, for the remainder of the chapter we will use B instead of BSM

No dynamic arbitrage and no rates imply that both option drifts must vanish, leading to the fundamental "PDE".

$$-\frac{\partial B}{\partial t} = \mu(t)\frac{\partial B}{\partial \sigma} + \frac{v(t)}{2}S^2(t)\frac{\partial^2 B}{\partial S^2} + \rho(t)\omega(t)\sqrt{v(t)}S(t)\frac{\partial^2 B}{\partial \sigma \partial S} + \frac{\omega^2(t)}{2}\frac{\partial^2 B}{\partial \sigma^2} \qquad (4.1)$$

The class of implied volatility surfaces defined by the fundamental "PDE" (4.1) is termed the Vega-Gamma-Vanna-Volga (VGVV) model

We should note that (4.1) is not a PDE in the traditional sense for the following reasons

- Traditionally, a PDE is specified to solve the value function. In our case, the value function $B(S(t), I(t; K, T), t)$ is well-known.

- The coefficients are deterministic in traditional PDEs, but are stochastic in (4.1)

The "PDE" is not derived to solve the value function, but rather it is used to show that the various stochastic quantities have to satisfy this particular relation to exclude dynamic arbitrage. Plugging in the $\mathcal{BSM}$ formula for B and its partial derivatives $\frac{\partial B}{\partial t}, \frac{\partial^2 B}{\partial S^2}, \frac{\partial^2 B}{\partial S \partial \sigma}, \frac{\partial^2 B}{\partial \sigma^2}$, we can reduce the "PDE" constraint into an algebraic restriction on the shape of the implied volatility surface $I(t; K, T)$

$$\frac{I^2(t; K, T)}{2} - \mu(t)I(t; K, T)\tau - \left[\frac{v(t)}{2} - \rho(t)\omega(t)\sqrt{v(t)}\sqrt{\tau}d_2 + \frac{\omega^2(t)}{2}d_1 d_2 \tau \right] = 0$$

where $\tau = T - t$

This algebraic restriction is the basis for the specific VGVV models: SRV and LNV, that we describe next.

For SRV we assume square-root implied variance dynamics

$$dI^2(t) = \kappa(t)\left[\theta(t) - I^2(t)\right]dt + 2w(t)e^{-\eta(t)(T-t)}I(t)dZ(t)$$

If we represent the implied volatility surface in terms of $\tau = T - t$ and standardized moneyness $z(t) \triangleq \frac{\ln(K/S(t)) + 0.5I^2\tau}{I\sqrt{\tau}}$, then $I(z, \tau)$ solves the following quadratic equation

$$(1 + \kappa(t))I^2(z, \tau) + \left(w^2(t)e^{-2\eta(t)\tau}\tau^{1.5}z \right)I(z, \tau)$$
$$- \left[\left(\kappa(t)\theta(t) - w^2(t)e^{-2\eta(t)\tau} \right)\tau + v(t) + 2\rho(t)\sqrt{v(t)}e^{-\eta(t)\tau}\sqrt{\tau}z + w^2(t)e^{-2\eta(t)\tau}\tau z^2 \right] = 0$$

For LNV we assume log-normal implied variance dynamics

$$dI^2(t) = \kappa(t)\left[\theta(t) - I^2(t)\right]dt + 2w(t)e^{-\eta(t)(T-t)}I(t)dZ(t)$$

If we represent the implied volatility surface in terms of $\tau = T - t$ and log relative strike $k(t) \triangleq \ln(K/S(t))$, then $I(k, \tau)$ solves the following quadratic equation

$$\frac{w^2(t)}{4}e^{-2\eta(t)\tau}\tau^2 I^4(k, \tau) + \left[1 + \kappa(t)\tau + w^2(t)e^{-2\eta(t)\tau}\tau - \rho(t)\sqrt{v(t)}w(t)e^{-\eta(t)\tau} \right]I^2(k, \tau)$$
$$- \left[v(t) + \kappa\theta(t)\tau + 2\rho(t)\sqrt{v(t)}e^{-\eta(t)\tau}k + w^2(t)e^{-2\eta(t)\tau}k^2 \right] = 0$$

For both SRV and LNV models we have six time varying stochastic coefficients:

$$\kappa(t), \theta(t), w(t), \eta(t), \rho(t), v(t)$$

Given time t values for the six coefficients, the whole implied volatility surface at time t can be found as solution to quadratic equations.

The dynamic calibration procedure treats the six coefficients as a state vector $X(t)$ and it assumes that $X(t)$ propagates like a random walk

$$X(t) = X(t-1) + \sqrt{\Sigma_X}\epsilon(t)$$

where Σ_X is a diagonal matrix. It also assumes that all implied volatilities are observed with errors IID normally distributed with error variance σ_e^2

$$y(t) = h(X(t)) + \sqrt{\Sigma_y}\epsilon(t)$$

with $h(\cdot)$ denoting the model value (quadratic solution for SRV or LNV) and $\Sigma_y = I_N \sigma_e^2$, with I_N denoting an identity matrix of dimension N

This setup introduces seven auxiliary parameters Θ that define the covariance matrix of the state and the measurement errors.

When the state propagation and the measurement equation are Gaussian linear, the Kalman filter provides efficient forecasts and updates on the mean and covariance of the state and observations. The state-propagation equations are Gaussian and linear, but the measurement functions $h(X(t))$ are not linear in the state vector. To handle the non-linearity we employ the unscented Kalman filter. For additional details the reader is referred to [36].

The procedure was applied successfully on both currency options and equity index options, and compared with Heston.

The comparison with Heston provided the following conclusions:

- generated half the root mean squared error

- explains 4% more variation

- generated errors with lower serial correlation

- can be calibrated 100 times faster

- The whole sample (573 weeks) of implied volatility surfaces can be fitted in about half a second (versus about 1 minute for Heston).

5 Volatility surface based on parametric representations

Various parametric or semi-parametric representations of the volatility surface have been considered in the literature. A recent overview was given in [62].

5.1 Polynomial parametrization

A popular representation was suggested in [56], which proposed that the implied volatility surface is modeled as a quadratic function of the moneyness $\mathcal{M} \triangleq \ln(F/K)/\sqrt{T}$

$$\sigma(\mathcal{M}, T) = b_1 + b_2 \mathcal{M} + b_3 \mathcal{M}^2 + b_4 T + b_5 \mathcal{M} T$$

This model was considered for oil markets in [28], concluding that the model gives only an "average" shape, due to its inherent property of assuming the quadratic function of volatility versus moneyness to be the same across all maturities. Note that increasing the power of the polynomial volatility function (from two to three or higher) does not really offer a solution here, since this volatility function will still be the same for all maturities.

To overcome those problems a semi parametric representation was considered in [28], where they kept quadratic parametrization of the volatility function for each maturity T, and approximate the implied volatility by a quadratic function which has time dependent coefficients.

A similar parametrization (but dependent on strike and not moneyness) was considered in [39] under the name Practitioner's BlackScholes. It was shown that outperforms some other models in terms of pricing error in sample and out of sample.

Such parametrizations may some certain drawbacks, such as:

- are not designed to ensure arbitrage-free of the resulting volatility surface

- the dynamics of the implied volatility surface may not be adequately captured

We now describe other parametrizations that may be more suitable.

5.2 Stochastic volatility inspired (SVI) parametrization

SVI is a practitioner designed parametrization [76, 77]. Very recent papers provide the theoretical framework and describe its applicability to energy markets [52, 53]. We also note that SVI procedure may be employed together with conditions for no vertical and horizontal spread arbitrage, such as in [82].

The essence of SVI is that each time slice of the implied volatility surface is fitted separately, such that in the logarithmic coordinates the implied variance curve is a hyperbola, and additional constraints are imposed that ensure no vertical/ horizontal spread arbitrage opportunities. The hyperbola is chosen because it gives the correct asymptotic representation of the variance when log-strike tends to plus or minus infinity: written as a function of $\ln\left(K/F\right)$, where K is the strike and F is the forward price, and time being fixed, the variance tends asymptotically to straight lines when $\ln\left(K/F\right) \to \pm\infty$

The parametrization form is on the implied variance:

$$\sigma^2\left[x\right] \triangleq v(\{m, s, a, b, \rho\}, x) = a + b\left(\rho\left(x - m\right) + \sqrt{\left(kx - m\right)^2 + s^2}\right)$$

where a, b, ρ, m, s are parameters which are dependent on the time slice and $x = \ln\left(K/F\right)$.

We should note that it was recently shown [127] that SVI may not be arbitrage-free in all situations. Nevertheless SVI has many advantages such as small computational time, relatively good approximation for implied volatilities for strikes deep in- and out-of-the-money. The SVI fit for equity markets is much better than for energy markets, for which [53] reported an error of maximum 4-5% for front year and respectively 1-2% for long maturities.

Quasi explicit calibration of SVI is presented in [51], based on dimension reduction for the optimization problem. The original calibration procedure is based on matching input market data $\left\{\sigma_i^{MKT}\right\}_{i=1...M}$, which becomes an optimization problem with five variables: a, b, ρ, m, s:

$$\min_{\{a,b,\rho,m,s\}} \sum_{i=1}^{N} \left(v\left[\{m, s, a, b, \rho\}, \ln\left(\frac{K_i}{F}\right)\right] - \left(\sigma_i^{MKT}\right)^2\right)^2$$

The new procedure is based on a change of variables

$$y = \frac{x - m}{s}$$

Focusing on total variance $V = vT$, the SVI parametrization becomes

$$V(y) = \alpha T + \delta y + \beta\sqrt{y^2 + 1}$$

where we have used the following notations

$$\begin{aligned} \beta &= bsT \\ \delta &= \rho bsT \\ \alpha &= aT \end{aligned}$$

We also use the notation $\bar{V}_i = \left[\sigma_i^{MKT}\right]^2 T$

Therefore, for given m and s, which is transformed into $\left\{y_i, \bar{V}_i\right\}$, we look for the solution of the 3-dimensional problem

$$\min_{\{\beta,\delta,\alpha\}} F_{\{y_i,\bar{V}_i\}}(\beta, \delta, \alpha) \tag{5.1}$$

with the objective functional for reduced dimensionality problem defined by

$$F_{\{y_i,v_i\}}(\beta, \delta, \alpha) = \sum_{i=1}^{N} w_i\left(\alpha + \delta y_i + \beta\sqrt{y_i^2 + 1} - \bar{V}_i\right)^2$$

The domain on which to solve the problem is defined as

$$\begin{cases} \beta_{MIN} \leq \beta \leq 4s \\ -\beta \leq \delta \leq \beta \\ -(4s - \beta) \leq \delta \leq (4s - \beta) \\ \alpha_{MIN} \leq \alpha \leq \bar{V}_{MAX} \end{cases}$$

For a solution $\{\beta^*, \delta^*, \alpha^*\}$ of the problem (5.1), we identify the corresponding triplet $\{a^*, b^*, \rho^*\}$ and then we solve the 2-dimensional optimization problem

$$\min_{\{m,s\}} \sum_{i=1}^{N} \left(v \left[\{m, s, a^*, b^*, \rho^*\}, \ln\left(\frac{K_i}{F}\right) \right] - v_i^{MKT} \right)^2$$

Thus the original calibration problem was cast as a combination of distinct 2-parameter optimization problem and, respectively, 3-parameter optimization problem. Because the "2+3" procedure is much less sensitive to the choice of initial guess, the resulting parameter set is more reliable and stable. For additional details the reader is referred to [51]. The SVI parametrization is performed sequentially, expiry by expiry. An enhanced procedure was presented in [82] to obtain a satisfactory term structure for SVI, which satisfies the no-calendar spread arbitrage in time while preserving the condition of no-strike arbitrage.

5.3 Entropy based parametrization

Entropic calibrations have been considered by a number of authors. It was done for risk-neutral terminal price distribution, implied volatility function and the option pricing function.

An algorithm that yields an arbitrage-free diffusion process by minimizing the relative entropy distance to a prior diffusion is described in [8]. This results in an efficient calibration method that can match an arbitrary number of option prices to any desired degree of accuracy. The algorithm can be used to interpolate, both in strike and expiration date, between implied volatilities of traded options.

Entropy maximization is employed in [31] and [30] to construct the implied risk-neutral probability density function for the terminal price of the underlying asset. The advantage of such an entropic pricing method is that it does not rely on the use of superfluous parameters, and thus avoids the issue of over fitting altogether. Furthermore, the methodology is flexible and universal in the sense that it can be applied to a wider range of calibration situations.

Most of the entropy-based calibration methodologies adopted in financial modeling, whether they are used for relative entropy minimization or for absolute entropy maximization, rely on the use of the logarithmic entropy measure of Shannon and Wiener. One drawback in the use of logarithmic entropy measures is that if the only source of information used to maximize entropy is the market prices of the vanilla options, then the resulting density function is necessarily of exponential form. On the other hand, empirical studies indicate that the tail distributions of asset prices obey power laws of the Zipf–Mandelbrot type [30] . Thus we would like to employ entropies that may recover power law distribution, such as Renyi entropy [30]

Maximization of Renyi entropy is employed to obtain arbitrage-free interpolation. The underlying theoretical idea is that, irrespective of the nature of the underlying price process, the gamma associated with a European-style vanilla option always defines a probability density function of the spot price implied by the existence of the prices for option contracts. There is a one-to-one correspondence between the pricing formula for vanilla options and the associated gamma. Therefore, given option gamma we can unambiguously recover the corresponding option pricing formula.

We present here an overview of the approach presented in [30]

Given strikes K_j, $j = 1...M$, corresponding for input market prices, maximizing the Renyi entropy yields a density function of the form:

$$p(x) = \left(\lambda + \beta_0 x + \sum_{j=1}^{M} \beta_j \left(x - K_j \right)^+ \right)^{\frac{1}{\alpha-1}} \tag{5.2}$$

The parameters $\alpha, \lambda, \beta_0, ..., \beta_M$ are calibrated by matching the input prices to the prices computed using the density function (5.2).

We exemplify for the case of call options. For each $j = 1...M$, we have to impose the matching condition to market price C_j^{MKT}

$$\bar{S}_0 - K_m - \frac{\alpha - 1}{\alpha} \sum_{m=1}^{j-1} Y_m \left[X_m \left(x \right) \right]^{\frac{\alpha}{\alpha-1}} \left(x - K_m - \frac{\alpha-1}{2\alpha-1} Y_m X_m \left(x \right) \right) \big|_{x=K_m}^{x=K_{m+1}} = C_j^{MKT}$$

where

$$X_j(x) \triangleq \lambda + \sum_{j=1}^{M} \beta_j \left(x - K_j \right)$$

$$Y_j \triangleq \left(\sum_{m=0}^{j} \beta_m \right)$$

We also impose the normalization condition

$$\int_0^\infty p(x)dx = 1 \implies \frac{\alpha - 1}{\alpha} \sum_{j=0}^{M} Y_j \left[X_j \left(x \right) \right]^{\frac{\alpha}{\alpha-1}} \big|_{x=K_j}^{x=K_{j+1}} = 1$$

Since the density function is explicitly given, is straightforward to use for calibration additional option types, such as digitals or variance swaps.

The result is described in [30] as leading to the power-law distributions often observed in the market. By construction, the input data are calibrated with a minimum number of parameters, in an efficient manner. The procedure allows for accurate recovery of tail distribution of the underlying asset implied by the prices of the derivatives. One disadvantage is that the input values are supposed to be arbitrage free, otherwise the algorithm will fail. It is possible to enhance the algorithm to handle inputs with arbitrage, but the resulting algorithm will lose some of the highly efficient characteristics, since now we need to solve systems of equations in a least square sense

5.4 Parametrization using weighted shifted lognormal distributions

A weighted sum of interpolation functions taken in a parametric family is considered in the practitioner papers [26, 27] to generate a surface without arbitrage in time and in space, while remaining as closely as possible to market data. Each function in the family is required to satisfy the no-free lunch constraints, specified later, in such way that they are preserved in the weighted sum.

In this parametric model, the price of a vanilla option price of strike K and maturity T is estimated at time $t_0 = 0$ by the weighted sum of N functionals

$$\sum_{i=1}^{N} a_i(T) F_i \left(t_0, S_0, P(T); K, T \right)$$

with $a_i(T)$ weights and $P(t) = B \left(0, t \right)$ the zero coupon bond price.

Several families F_i can satisfy the No-Free-Lunch constraints, for instance a sum of lognormal distributions, but in order to match a wide variety of volatility surfaces the model has to produce prices that lead to risk-neutral pdf of the asset prices with a pronounced skew. If all the densities are centered in the log-space around the forward value, one recovers the no-arbitrage forward pricing condition but the resulting pricing density will not display skew. However, centering the different normal densities around different locations (found appropriately) and constraining the weights to be positive, we can recover the skew. Since we can always convert a density into call prices, we can then convert a mixture of normal densities into a linear combination of $\mathcal{BSM}$ formulae.

Therefore, we can achieve that goal with a sum of shifted log-normal distributions, that is, using the $\mathcal{BSM}$ formula with shifted strike (modified by the parameters) as an interpolation function

$$F_i\left(t_0, S_0, P(T); K, T\right) = Call_{BSM}\left(t_0, S_0, P(T), \hat{K}\left(1 + \mu_i(T)\right), T, \sigma_i\right)$$

with $\hat{K}$ denoting adjusted strike.

We note that the value of strike is adjusted only if we apply the procedure for equity markets, in which case it becomes

$$\hat{K}(K, t) = K + D(0, t)$$

with $D(0, t)$ is the compounded sum of discrete dividends between 0 and t.

The no-arbitrage theory imposes time and space constraints on market prices. Hence, the time dependent parameters $a_i(t)$ and $\mu_i(t)$ are used to recover the time structure of the volatility surface. It is argued that it sufficient to use a parsimonious representation of the form

$$
\begin{aligned}
\mu_i(t) &= \mu_i^0 f\left(t, \beta_i\right) \\
a_i(t) &= \left(\sum_{i=1}^{N} \frac{a_i^0}{f\left(t, \beta_i\right)}\right) \frac{a_i^0}{f\left(t, \beta_i\right)} \\
f\left(t, \beta_i\right) &\triangleq 1 - \frac{2}{1 + \left(1 + \frac{t}{\beta}\right)^2}
\end{aligned}
$$

Making the weights and the shift parameter time-dependent to fit a large class of volatility surfaces leads to the following no-free lunch constraints, for any time t

- $a_i \geq 0$ to get convexity of the price function

- $\sum_{i=1}^{N} a_i(t) = 1$ to get a normalized risk-neutral probability

- $\sum_{i=1}^{N} a_i(t)\mu_i(t) = 1$ to keep the martingale property of the induced risk-neutral pdf

- $\mu_i(t)$ to get non-degenerate functions

The model being invariant when multiplying all the terms a_i^0 with the same factor, we impose the normalization constraint

$$\sum_{i=1}^{N} a_i^0 = 1$$

to avoid the possibility of obtaining different parameter sets which nevertheless yield the same model.

Given the N parameters and assuming a constant volatility $\sigma_i(t) = \sigma_i^0$, there are $4N - 2$ free parameters for the N-function model since we can use the constraints to express a_1^0 and, respectively, $a_1^0 \mu_1^0$ in terms of $\left\{a_i^0\right\}_{i=1...N}$ and $\left\{\mu_i^0\right\}_{i=1...N}$ (see also Appendix A of [27]).

As such, this model does not allow for the control of the long term volatility surface. Therefore, for the model to be complete we specify the time-dependent volatility such that it captures the term structure of the implied volatility surface:

$$\sigma_i(t) = \gamma_i e^{-c_i t} + d_i f(t, b_i)$$

Thus we need to solve a $7N - 2$ optimization problem. This is done in [26, 27] using a global optimizer of Differential Evolution type.

6 Volatility surface based on nonparametric representations, smoothing and interpolation

This broad set of procedures may be divided into several categories.

6.1 Arbitrage-free algorithms

Interpolation techniques to recover a globally arbitrage-free call price function have been suggested in various papers, e.g., [95, 143]. A crucial requirement for these algorithms to work that the data to be interpolated are arbitrage-free from the beginning. [95] proposes an interpolation procedure based on piecewise convex polynomials mimicking the $\mathcal{BSM}$ pricing formula. The resulting estimate of the call price function is globally arbitrage-free and so is the volatility smile computed from it. In a second step, the total (implied) variance is interpolated linearly along strikes. Cubic B-spline interpolation was employed by [143], with interpolation performed on option prices, and the shape restrictions in interpolated curves was imposed by the use of semi-smooth equations minimizing the distance between the implied risk neutral density and a prior approximation.

Instead of smoothing prices, [20] suggests to directly smooth implied volatility parametrization by means of constrained local quadratic polynomials. Let us consider that we have M expiries $\{T_j\}$ and N strikes $\{x_i\}$, while the market data is denoted by $\left\{ \sigma_i^{MKT}(T_j) \right\}$

Two approaches are considered:
• each maturity is treated separately
• all maturities are included in the cost functional to minimize
The first case implies minimization of the following (local) least squares criterion at each expiry T_j, j=1...NT

$$\min_{\left\{\alpha_0^{(j)}, \alpha_1^{(j)}, \alpha_2^{(j)}\right\}} \sum_{i=1}^{N} \left\{ \sigma_i^{MKT}(T_j) - \alpha_0^{(j)} - \alpha_1^{(j)}(x_i - x) - \alpha_2^{(j)}(x_i - x)^2 \right\} \frac{1}{h} \mathcal{K}\left[\frac{x_i - x}{h} \right]$$

where $\mathcal{K}$ is a kernel function, typically a symmetric density function with compact support. One example is the Epanechnikov kernel

$$\mathcal{K}(u) = 0.75\left(1 - u^2\right) \mathbf{1}\left[|u| \leq 1 \right]$$

with $\mathbf{1}(A)$ denoting the indicator function for a set A and h is the bandwidth which governs the trade-off between bias and variance.

The optimization problem for the second approach is

$$\min_{\left\{\left\{\alpha_0^{(j)}, \alpha_1^{(j)}, \alpha_2^{(j)}, \alpha_3^{(j)}, \alpha_4^{(j)}\right\}_{j=1..M}\right\}} \sum_{j=1}^{M} \sum_{i=1}^{N} \Psi\left(\left\{ \alpha_0^{(j)}, \alpha_1^{(j)}, \alpha_2^{(j)}, \alpha_3^{(j)}, \alpha_4^{(j)} \right\} \right) \frac{1}{h_X} \mathcal{K}\left[\frac{x_i - x}{h_X} \right] \frac{1}{h_T} \mathcal{K}\left[\frac{T_j - T}{h_T} \right]$$

with defined as

$$\Psi\left(\left\{\alpha_0^{(j)}, \alpha_1^{(j)}, \alpha_2^{(j)}, \alpha_3^{(j)}, \alpha_4^{(j)}\right\}\right) \triangleq \sigma_i^{MKT}(T_j) - \alpha_0^{(j)} - \alpha_1^{(j)}(x_i - x)$$
$$-\alpha_2^{(j)}(T_j - T) - \alpha_3^{(j)}(x_i - x)^2 - \alpha_4^{(j)}(x_i - x)(T_j - T)$$

The approach yields a volatility surface that respects the convexity conditions, but neglects the conditions on call spreads and the general price bounds. Therefore the surface may not be fully arbitrage-free. However, since convexity violations and calendar arbitrage are by far the most virulent instances of arbitrage in observed implied volatility data, the surfaces will be acceptable in most cases.

The approach in [62] is based on cubic spline smoothing of option prices rather than on interpolation. Therefore, the input data does not have to be arbitrage-free. It employs cubic splines, with constraints specifically added to the minimization problem in order to ensure that there is no arbitrage. A potential drawback for this approach is the fact that the call price function is approximated by cubic polynomials. This can turn out to be disadvantageous, since the pricing function is not in the domain of polynomials functions. It is remedied by the choice of a sufficiently dense grid in the strike dimension.

Instead of cubic splines, [102] employs constrained smoothing B-splines. This approach permits to impose monotonicity and convexity in the smoothed curve, and also through additional pointwise constraints. According to the author, the methodology has some apparent advantages on competing methodologies. It allows to impose directly the shape restrictions of no-arbitrage in the format of the curve, and is robust the aberrant observations. Robustness to outliers is tested by comparing the methodology against smoothing spline, Local Polynomial Smoothing and Nadaraya-Watson Regression. The result shows that Smoothing Spline generates an increasing and non-convex curve, while the Nadaraya-Watson and Local Polynomial approaches are affected by the more extreme points, generating slightly non convex curves.

It is mentioned in [117] that a large drawback of bi-cubic spline or B-spline models is that they require the knots to be placed on a rectangular grid. Correspondingly, it considers instead a thin-spline representation, allowing arbitrarily placed knots. This leads to a more complex representation at shorter maturities while preventing overfitting.

Thin-spline representation of implied volatility surface was also considered in [29] and section 2.4 of [96], where it was used to obtain a pre-smoothed surface that will be eventually used as starting point for building a local volatility surface.

An efficient procedure was shown in [109] for constructing the volatility surface using generic volatility parametrization for each expiry, with no-arbitrage conditions in space and time being added as constraints, while a regularization term was added to the calibrating functional based on the difference between market implied volatilities and, respectively, volatilities given by parametrization. Bid-ask spread is also included in the setup. The resulting optimization problem has a lot of sparsity/structure, characteristics that were exploited for obtaining a good fit in less than a second

6.2 Remarks on spline interpolation

The following splines are usually employed to interpolate implied volatilities
- Regular cubic splines
- Cubic B-splines
- Thin splines

Certain criteria (such as arbitrage free etc) have to be met, and relevant papers were described in the previous section . Here we just refer to several generic articles on spline interpolation.

[145] describes an approach that yields monotonic cubic spline interpolation.

Although somewhat more complicated to implement, B-splines may be preferred to cubic splines, due to its robustness to bad data, and ability to preserve monotonicity and convexity. A recent paper [99] describes a computationally efficient approach for cubic B-splines.

A possible alternative is the thin-plate spline, which gets its name from the physical process of bending a thin plate of metal. A thin plate spline is the natural two-dimensional generalization of the cubic spline, in that it is the surface of minimal curvature through a given set of two-dimensional data points.

6.3 Remarks on interpolation in time

In some situations we need to perform interpolation in time. While at a first glance it may seem straight-forward, special care has to be employed to ensure that the result still satisfies practical arbitrage conditions. For example, one should expect that there is no calendar spread arbitrage [34, 52, 76, 124]

One common approach is to perform linear interpolation in variance. A variant of it, denoted "total variance interpolation", is described in [37].

6.4 Interpolation based on fully implicit finite difference discretization of Dupire forward PDE

We present an approach described in [6, 7, 91], based on fully implicit finite difference discretization of Dupire forward PDE.

We start from the Dupire forward PDE in time-strike space

$$-\frac{\partial c}{\partial t} + \frac{1}{2} \left[\sigma \left(t, k \right) \right]^2 \frac{\partial^2 c}{\partial k^2} = 0$$

Let us consider that we have the following time grid $0 = t_0 < t_1 < ... < t_N$ and define $\triangle t_i \triangleq t_{i+1} - t_i$

A discrete (in time) version of the forward equation is

$$\frac{c \left(t_{i+1}, k \right) - c \left(t_i, k \right)}{\triangle t_i} = \frac{1}{2} \left[\sigma \left(t_i, k \right) \right]^2 \frac{\partial^2 c}{\partial k^2} \left(t_{i+1}, k \right)$$

This is similar to an implicit finite difference step. It can be rewritten as

$$\left[1 - \frac{1}{2} \triangle t_i \left[\sigma \left(t_i, k \right) \right]^2 \frac{\partial^2}{\partial k^2} \right] c \left(t_{i+1}, k \right) = c \left(t_i, k \right) \tag{6.1}$$

Let us consider that the volatility function is piecewise defined on the time interval $t_i \leq t < t_{i+1}$ and we denote by $\nu_i(k)$ the corresponding functions

$$\nu_i(k) \triangleq \sigma(t, k) \quad for \ t_i \leq t < t_{i+1}$$

Using (6.1) we can construct European (call) option prices for all discrete time points for a given a set of volatility functions $\{\nu_i(k)\}_{i=1...N}$ by recursively solving the forward system

$$\left[1 - \frac{1}{2} \triangle t_i \left[\sigma \left(t_i, k \right) \right]^2 \frac{\partial^2}{\partial k^2} \right] c \left(t_{i+1}, k \right) = c \left(t_i, k \right) \tag{6.2}$$

$$c(0, k) = \left[S(0) - k \right]^+$$

Let us discretize the strike space as $K_{MIN} = k_0 < k_1 < ... < k_M = k_{MAX}$

By replacing the differential operator $\partial^2/\partial k^2$ by the central difference operator

$$\delta_{kk} f(k) = \frac{2}{\left(k_j - k_{j-1} \right) \left(k_{j+1} - k_{j-1} \right)} f(k_{j-1}) - \frac{2}{\left(k_j - k_{j-1} \right) \left(k_{j+1} - k_j \right)} f(k_j)$$

$$+ \frac{2}{\left(k_{j+1} - k_j \right) \left(k_{j+1} - k_{j-1} \right)} f(k_{j+1})$$

we get the following finite difference scheme system

$$\left[1 - \frac{1}{2}\triangle t_i \left[\nu_i\left(k\right)\right]^2 \delta_{kk}\right] c\left(t_{i+1}, k\right) = c\left(t_i, k\right) \tag{6.3}$$
$$c(0, k) = \left[S(0) - k\right]^+$$

The matrix of the system (6.3) is tridiagonal and shown in [6] to be diagonally dominant, which allows for a well behaved matrix that can be solved efficiently using Thomas algorithm [132].Thus we can directly obtain the European option prices if we know the expressions for $\{\nu_i(k)\}_{i=1...N}$.

This suggests that we can use a bootstrapping procedure, considering that the volatility functions are defined as piecewise constant

Let us first introduce the notations for market data. We consider that we have a set of discrete option quotes $\left\{c^{MKT}\left(t_i, K_{i,p}\right)\right\}$, where $\{t_i\}$ are the expiries and $\{K_{i,p}\}_{p=1...NK(i)}$ is the set of strikes for expiry t_i.

We should note that we may have different strikes for different expiries, and that $\{K_{i,p}\}_{p=1...NK(i)}$ and, respectively, $\{k_j\}$ represent different quantities

Then the piecewise constant volatility functions are denoted as

$$\nu_i(k) \triangleq \begin{cases} ... \\ \sigma_{i,p} & for \ K_{i,p} \leq k < K_{i,p+1} \\ ... \end{cases}$$

Thus the algorithm consists of solving an optimization problem at each expiry time, namely

$$\min_{\left\{a_{i,1},...,a_{i,NK(i)}\right\}} \left[\sum_{p=1}^{NK(i)} \left(c\left(t_i, K_{i,p}\right) - c^{MKT}\left(t_i, K_{i,p}\right)\right)^2\right] \tag{6.4}$$

We remark that, when solving (6.4) by some optimization procedure, one needs to solve only one tridiagonal matrix system for each optimization iteration.

Regarding interpolation in time, two approaches are proposed in [6]. The first one is based on the formula

$$\left[1 - \frac{1}{2}\left(t - t_i\right)\left[\nu_i\left(k\right)\right]^2 \frac{\partial^2}{\partial k^2}\right] c\left(t_{i+1}, k\right) = c\left(t_i, k\right) \ \ for \ t_i < t < t_{i+1} \tag{6.5}$$

while the second one is a generalization of (6.5)

$$\left[1 - \frac{1}{2}\left(T(t) - t_i\right)\left[\nu_i\left(k\right)\right]^2 \frac{\partial^2}{\partial k^2}\right] c\left(t_{i+1}, k\right) = c\left(t_i, k\right) \ \ for \ t_i < t < t_{i+1} \tag{6.6}$$

where $T(t)$ is a function that satisfies the conditions $T(t_i) = t_i$ and $T'(t) < 0$

It is shown in [6] that option prices generated by (6.2)and (6.5) and, respectively, by (6.2) and (6.6)are consistent with the absence of arbitrage in the sense that , for any pair (t, k) we have

$$\frac{\partial c}{\partial t}\left(t, k\right) \geq 0$$
$$\frac{\partial^2 c}{\partial k^2}\left(t, k\right) \geq 0$$

7 Adjusting inputs to avoid arbitrage

Various papers have tackled the problem of finding conditions that may be necessary/and or sufficient to ensure that prices/vols are free of arbitrage [34, 46, 49, 89, 113, 125]. If one wants to adjust the set of input prices/vols to avoid arbitrage, several approaches have been described in the literature. For example, [3] presents a relatively simple method to adjust implied volatilities, such that the resulting set is both arbitrage free and also closest to the initial values (in a least-squares sense). Another algorithm is presented in section 8.3 of [32]. We present in detail the algorithm from [34], based on the observation that the absence of call spread, butterfly spread and calendar spread arbitrages is sufficient to exclude all static arbitrages from a set of option price quotes across strikes and maturities.

7.1 Carr and Madan algorithm

The main idea is as follows: given input market prices and corresponding bid ask spreads, we start from the price corresponding to first expiry ATM and adjust the prices for that expiry. We continue to the next expiry and we make sure that arbitrage constraints are satisfied both in time and strike space, while adjusting within the bid ask spread.

We present first the arbitrage constraints from [34], using notations from there. Let C_{ij} denote the given quote for a call of strike K_i and maturity T_j. We suppose that the N strikes $\{K_i\}$ form an increasing and positive sequence as do the M maturities $\{T_j\}$. Without any loss of generality, we suppose that interest rates and dividends are zero over the period ending at the longest maturity.

We augment the provided call quotes with quotes for calls of strike $K_0 = 0$. For each maturity, these additional quotes are taken to be equal to the current spot price S_0. We also take the prices at maturity $T_0 = 0$ to be $(S_0 - K_i)^+$. This gives us the augmented matrix of prices C_{ij}, with indices $i = 0..N$ and $j = 1...M$.

For each $j > 0$ we define the following quantities:

$$
\begin{aligned}
Q_{i,j} &= \frac{C_{i-1,j} - C_{i,j}}{K_i - K_{i-1}} \\
Q_{0,j} &= 0
\end{aligned}
$$

For each $i > 0$, $Q_{i,j}$ is the cost of a vertical spread which by definition is long $1/(K_i - K_{i-1})$ calls of strike K_{i-1} and short $1/(K_i - K_{i-1})$ calls of strike K_i. A graph of the payoff from this position against the terminal stock price indicates that this payoff is bounded below by zero and above by one.

We therefore require for our first test that

$$0 \leq Q_{i,j} \leq 1, \;\; i = 1...N, j = 1...M \tag{7.1}$$

Next, for each $j > 0$, we define the following quantities:

$$BSpr_{i,j} \triangleq C_{i-1,j} - \frac{K_{i+1} - K_{i-1}}{K_{i+1} - K_i} C_{i,j} + \frac{K_i - K_{i-1}}{K_{i+1} - K_i} C_{i+1,j} \;\;\; i > 0$$

For each $i > 0$ $BSpr_{i,j}$ is the cost of a butterfly spread which by definition is long the call struck at K_{i-1}, short $(K_{i+1} - K_{i-1})/(K_{i+1} - K_i)$ calls struck at K_i, and long $(K_i - K_{i-1})/(K_{i+1} - K_i)$ calls struck at K_{i+1}. A graph indicates that the butterfly spread payoff is non-negative and hence our second test requires that

$$C_{i-1,j} - \frac{K_{i+1} - K_{i-1}}{K_{i+1} - K_i} C_{i,j} + \frac{K_i - K_{i-1}}{K_{i+1} - K_i} C_{i+1,j} \geq 0$$

Equivalently, we require that

$$C_{i-1,j} - C_{i,j} \geq \frac{K_i - K_{i-1}}{K_{i+1} - K_i} \left(C_{i,j} - C_{i+1,j} \right) \tag{7.2}$$

We define
$$q_{i,j} \triangleq Q_{i,j} - Q_{i+1,j} = \frac{C_{i-1,j} - C_{i,j}}{K_i - K_{i-1}} - \frac{C_{i,j} - C_{i+1,j}}{K_{i+1} - K_i}$$

We may interpret each $q_{i,j}$ as the marginal risk-neutral probability that the stock price at maturity T_j equals K_i.

For future use, we associate with each maturity a risk-neutral probability measure defined by
$$\mathbb{Q}_j(K) = \sum_{K_j \leq K} q_{i,j}$$

A third test on the provided call quotes requires that for each discrete strike $K_i, i \geq 0$, and each discrete maturity $T_j, j \geq 0$, we have
$$C_{i,j+1} - C_{i,j} \geq 0 \tag{7.3}$$

The left-hand side of (7.3) is the cost of a calendar spread consisting of long one call of maturity T_{j+1} and short one call of maturity T_j, with both calls struck at K_i. Hence, our third test requires that calendar spreads comprised of adjacent maturity calls are not negatively priced at each maturity.

We now conclude, following [34] , the discussion on the 3 arbitrage constraints (7.1)(7.2)(7.3).

As the call pricing functions are linear interpolations of the provided quotes, we have that at each maturity T_j, calendar spreads are not negatively priced for the continuum of strikes $K > 0$. Since all convex payoffs may be represented as portfolios of calls with non-negative weights, it follows that all convex functions $\phi(S)$ are priced higher when promised at T_{j+1} than when they are promised at T_j. In turn, this ordering implies that the risk-neutral probability measures $\mathbb{Q}$ constructed above are increasing in the convex order with respect to the index j. This implies that there exists a martingale measure which is consistent with the call quotes and which is defined on some filtration that includes at least the stock price and time. Finally, it follows that the provided call quotes are free of static arbitrage by standard results in arbitrage pricing theory.

8 Characteristics of volatility surface

Many recent papers have studied various characteristics of volatility surface:

- the static and dynamic properties of the implied volatility surface must exhibit within an arbitrage-free framework

- implied volatility calculations in a (local) stochastic volatility environment, which may also include jumps or even Levy processes.

- the behavior of implied volatility in limiting cases, such as extreme strikes, short and large maturities, etc.

For completion we include a list of relevant papers: [11, 14, 21, 38, 122, 46, 48, 49, 50, 57, 63, 70, 69, 66, 65, 64, 67, 68, 71, 73, 75, 77, 78, 81, 86, 87, 93, 101, 103, 104, 106, 112, 111, 115, 118, 119, 123, 125, 127, 128, 133, 134, 135, 137, 140, 141, 16, 17, 19, 22, 23, 24, 25, 10, 58, 72, 74, 92, 97, 136, 144, 15]

The constructed volatility surface may also need to take into account the expected behavior of the volatility surface outside the core region. The core region is defined as the region of strikes for equity markets, moneyness levels for Commodity markets, deltas for FX markets for which we have observable market data. From a theoretical point of view, this behavior may be described by the asymptotics of implied volatility, while from a practical point of view this corresponds to smile extrapolation.

8.1 Asymptotics of implied volatility

Concerning the dependence with respect to strike, some major theoretical results are known in a model-independent framework. [103] related the extreme strike slopes of the implied volatility to the critical moments of the underlying through the moment formula: let $\sigma(t,x)$ denote the implied volatility of a European Call option with maturity t and strike $K = F_0 e^x$, then

$$\limsup_{x \to \infty} \frac{t\sigma(t,x)^2}{x} = \psi(u^*(t) - 1) \tag{8.1}$$

where $\psi(u) = 2 - 4\left(\sqrt{u^2 + u} - u\right)$ and $u^*(t) \triangleq \sup\{u \geq 1; E[F^u(t)]\}$ is the critical moment of the underlying price $F = (F(t))_{t \geq 0}$. An analogous formula holds for the left part of the smile, namely when $x \to -\infty$. This result was sharpened in [14, 15], relating the left hand side of (8.1) to the tail asymptotics of the distribution of $F(t)$.

In the stochastic volatility framework this formula was applied by [5] and [97], to mention but a few.

The study of short- (resp. long-) time asymptotics of the implied volatility is motivated by the research of efficient calibration strategies to the market smile at short (resp. long) maturities. Short time results have been obtained in [111, 66, 65, 64, 21], while some other works provide insights on the large-time behavior, as done by [141] in a general setting, [97] for affine stochastic volatility models or [67] for Heston model.

8.2 Smile extrapolation

It is argued in the practitioner paper [13] that a successful smile extrapolation method should deliver arbitrage-free prices for the vanilla options, i.e., the option prices must be convex functions of strike, and stay within certain bounds. In addition, the extrapolation method should ideally have the following properties:

1. It should reprice all observed vanilla options correctly.

2. The PDF, CDF and vanilla option prices should be easy to compute.

3. The method should not generate unrealistically fat tails, and if possible, it should allow us to control how fat the tails are.

4. It should be robust and flexible enough to use with a wide variety of different implied volatility surfaces.

5. It should be easy and fast to initialize for a given smile.

The paper describes two commonly used methods which do not satisfy the above wish list. The first one is to use some interpolation within the region of observed prices, and just set the implied volatility to be a constant outside of this region. This method is flawed as it introduces unstable behavior at the boundary between the smile and the flat volatility, yielding unrealistically narrow tails at extreme strikes.

The second approach is to use the same parametric form for the implied volatility derived from a model, e.g. SABR, both inside and outside the core region. There are several problems with this method. It gives us little control over the distribution; indeed this approach often leads to excessively fat tails, which can lead to risk neutral distributions that have unrealistically large probabilities of extreme movements, and have moment explosions that lead to infinite prices, even for simple products. If the methodology is dependent on usage of an asymptotic expansion, the expansion may become less accurate at strikes away from the money, leading to concave option prices, or equivalently negative

PDFs, even at modestly low strikes. Furthermore, there is no guarantee that this functional form will lead to arbitrage free prices for very large and small strikes.

That is why [13] propose to separate the interpolation and extrapolation methods.

Their method works as follows. A core region of observability, inside which we may use any standard smile interpolation method, is defined first: $K_- \leq K \leq K_+$. Outside of this region the extrapolation is done by employing a simple analytical formula for the option prices, that has the following characteristics:

- For very low strikes region, the formula-based put prices will go towards zero as the strike goes to zero, while remaining convex.

- For very high strikes region, the formula-based call prices will go towards zero as strike goes to infinity, while remaining convex.

Each of these formulas is parametrized so that we can match the option price as well as its first two derivatives at the corresponding boundary with the core region. The methodology is also able to retain a measure of control over the form of the tails at extreme strikes.

The following functional form for the extrapolation of put and, respectively, call prices was described as parsimonious yet effective:

$$
\begin{aligned}
Put(K) \quad &= K^{\mu} \quad \exp\left[a_1 + b_1 K + c_1 K^2\right] \\
Call(K) \quad &= K^{-\nu} \quad \exp\left[a_2 + \frac{b_2}{K} + \frac{c_2}{K^2}\right]
\end{aligned}
$$

We fix $\mu > 1$, which ensures that the price is zero at zero strike, and there is no probability for the underlying to be zero at maturity. Alternatively, we can choose μ to reflect our view of the fatness of the tail of the risk neutral distribution. It is easy to check that this extrapolation generates a distribution where the m-th negative moment is finite for $m < 1 - \mu$ and infinite for $m > 1 - \mu$.

We fix $\nu > 0$ to ensure that the call price approaches zero at large enough strikes. Our choice of controls the fatness of the tail; the m-th moment will be finite if $m < \nu - 1$ and infinite if $m > \nu - 1$.

The condition for matching the price and its first two derivatives at K_- and, respectively, at K_+ yields a set of linear equations for the parameters a_1, b_1, c_1 and, respectively, for a_2, b_2, c_2

9 Remarks on numerical calibration

The calibration procedure consists of finding the set of parameters (defining the volatility surface) that minimize the error between model output and market data, while satisfying some additional constraints such as "no arbitrage in strike and time", smoothness, etc. This chapter provides some details regarding the practical aspects of numerical calibration.

Let us start by making some notations: we consider that we have M expiries $\{T^{(j)}\}$ and that for each maturity $T^{(j)}$ we have $N[j]$ calibrating instruments, with strikes $K_{i,j}$, for which market data is given (as prices or implied volatilities). The bid and ask values are denoted by $Bid\left(i, T^{(j)}\right)$ and $Ask\left(i, T^{(j)}\right)$

9.1 Calibration function

The calibration function is defined in different ways

If we perform "all-at-once" calibration, then the calibration function is constructed as

$$
\Psi \triangleq \sum_{j=1}^{M} \sum_{i=1}^{N[j]} w_{i,j} \left\| ModelOutput\left(i, T^{(j)}\right) - MarketData\left(i, T^{(j)}\right) \right\|
$$

where $w_{i,j}$ are weights and $\|\cdot\|$ denotes a generic norm

If we perform sequential calibration, one expiry at the time, than the calibration functional for each expiry will be given as

$$\Psi\left[j\right] \triangleq \sum_{i=1}^{N[j]} w_{i,j} \left\| ModelOutput\left(i, T^{(j)}\right) - MarketData\left(i, T^{(j)}\right)\right\|$$

If we use a local optimizer, then we might need to add a regularization term. The regularization term the most commonly considered in the literature is of Tikhonov type. e.g., [116, 44, 110, 2]. However, since this feature is primarily employed to ensure that the minimizer does not get stuck in a local minimum, this additional term is usually not needed if we use either a global optimizer or a hybrid (combination of global and local) optimizer.

9.2 Constructing the weights in the calibration functional

The weights $w_{i,j}$ can be selected following various procedures detailed in [26, 27, 41, 47, 105] chapter 13 of [42], to mention but a few.

Practitioners usually compute the weights (see [47]) as inverse proportional to

- the square of the bid-ask spreads, to give more importance to the most liquid options.

- the square of the $\mathcal{BSM}$ Vegas (roughly equivalent to using implied volatilities, as explained below).

[105] asserts that it is statistically optimal (minimal variance of the estimates) to choose the weights as the inverse of the variance of the residuals, which is then considered to be proportional to the inverse of squared bid–ask spread.

Another practitioner paper [26] considers a combination of the 2 approaches and this is our preferred methodology.

It is known that at least for the options that are not too far from the money, the bid-ask spreads is of order of tens of basis points. This suggests that it might be better to minimize the differences of implied volatilities instead of those of the option prices, in order to have errors proportional to bid-ask spreads and to have better scaling of the cost functional. However, this method involves additional computational cost. A reasonable approximation is to minimize the square differences of option prices weighted by the $\mathcal{BSM}$ Vegas evaluated at the implied volatilities of the market option prices.

The starting point is given by setting the weights as

$$w_{i,j} = \frac{1}{\left| Bid\left(i, T^{(j)}\right) - Ask\left(i, T^{(j)}\right)\right|}$$

To simplify the notations for the remainder of the chapter we denote the model price by $ModP$ and the market price by $MktP$

We can approximate the difference in prices as follows:

$$ModP\left(i, T^{(j)}\right) - MktP\left(i, T^{(j)}\right) \approx \frac{\partial\left[MktP\left(i, T^{(j)}\right)\right]}{\partial \sigma_{IV}} \left(\sigma_{IV}^{MOD}\left(i, T^{(j)}\right) - \sigma_{IV}^{MKT}\left(i, T^{(j)}\right)\right)$$

where $\sigma_{IV}^{MOD}\left(i, T^{(j)}\right)$ and $\sigma_{IV}^{MKT}\left(i, T^{(j)}\right)$ are the implied vols corresponding to model and, respectively, market prices for strikes $K_{i,j}$ and maturities $T^{(j)}$.

We should note that this series approximation may be continued to a higher order if necessary, with a very small additional computational cost.

Using the following expression for $\mathcal{BSM}$ Vegas evaluated at the implied volatility $\sigma_{IV}^{MKT}\left(i, T^{(j)}\right)$ of the market option prices

$$\frac{\partial\left[MktP\left(i, T^{(j)}\right)\right]}{\partial \sigma_{IV}} = Vega\left(\sigma_{IV}^{MKT}\left(i, T^{(j)}\right)\right)$$

we obtain

$$\sigma_{IV}^{MOD}\left(i, T^{(j)}\right) - \sigma_{IV}^{MKT}\left(i, T^{(j)}\right) \approx \frac{1}{Vega\left(\sigma_{IV}^{MKT}\left(i, T^{(j)}\right)\right)}\left[ModP\left(i, T^{(j)}\right) - MktP\left(i, T^{(j)}\right)\right]$$

Thus we can switch from difference of implied volatilities to difference of prices. For example, for all-at-once calibration that is based on minimization of root mean squared error (RMSE) , the corresponding calibration functional can be written as

$$\Psi \triangleq \sum_{j=1}^{M}\sum_{i=1}^{N[j]} \bar{w}_{i,j}\left[ModelOutput\left(i, T^{(j)}\right) - MarketData\left(i, T^{(j)}\right)\right]^2$$

where the weights $\bar{w}_{i,j}$ are defined as

$$\bar{w}_{i,j} = \frac{1}{\left|Bid\left(i, T^{(j)}\right) - Ask\left(i, T^{(j)}\right)\right|}\left(\frac{1}{Vega\left(\sigma_{IV}^{MKT}\left(i, T^{(j)}\right)\right)}\right)^2$$

To avoid overweighting of options very far from the money we need introduce an upper limit for the weights.

9.3 Selection of numerical optimization procedure

It is quite likely that the calibration function for the volatility surface may exhibit several local (and perhaps global) minima, making standard optimization techniques somewhat unqualified according to [40]. Gradient based optimizers, for example, are likely to get stuck in a local minimum which may also be strongly dependent on the initial parameter guess. While this situation (multiple local minima) may be less common for equity markets , it is our experience that such characteristics are quite common for FX and Commodities markets. Thus global/hybrid optimization algorithms are our preferred optimization methods in conjunction with volatility surface construction.

Our favorite global optimization algorithm is based on Differential Evolution [121]. In various flavors it was shown to outperform all other global optimization algorithms when solving benchmark problems (unconstrained, bounded or constrained optimization). Various papers and presentations described successful calibrations done with Differential Evolution in finance: [26, 27, 54, 142, 40], to mention but a few.

Here is a short description of the procedure. Consider a search space Θ and a continuous function $G : \Theta \rightarrow \mathbb{R}$ to be minimized on Θ. An evolutionary algorithm with objective function G is based on the evolution of a population of candidate solutions, denoted by $X_n^N = \left\{\theta_n^i\right\}_{i=1...N}$. The basic idea is to "evolve" the population through cycles of modification (mutation) and selection in order to improve the performance of its individuals. At each iteration the population undergoes three transformations:

$$X_n^N \rightarrow V_n^N \rightarrow W_n^N \rightarrow V_{n+1}^N$$

During the mutation stage, individuals undergo independent random transformations, as if performing independent random walks in Θ, resulting in a randomly modified population V_n^N. In the crossover stage, pairs of individuals are chosen from the population to "reproduce": each pair gives birth to a new individual, which is then added to the population. This new population, denoted W_n^N, is now

evaluated using the objective function $G(\cdot)$. Elements of the population are now selected for survival according to their fitness: those with a lower value of G have a higher probability of being selected. The N individuals thus selected then form the new population X_{n+1}^N. The role of mutation is to explore the parameter space and the optimization is done through selection.

On the downside, global optimization techniques are generally more time consuming than gradient based local optimizers. Therefore, we employ a hybrid optimization procedure of 2 stages which combines the strengths of both approaches. First we run a global optimizer such as Differential Evolution for a small number of iterations, to arrive in a "neighborhood" of a global minimum. In the second stage we run a gradient-based local optimizer, using as initial guess the output from the global optimizer, which should converge much quite fast since the initial guess is assumed to be in the "correct neighborhood". An excellent resource for selecting a suitable local optimizer can be found at [1]

We should also mention that a very impressive computational speedup (as well as reducing number of necessary optimization iterations) can be achieved if the gradient of the cost functional is computed using Adjoint method in conjunction with Automatic, or Algorithmic, Differentiation (usually termed AD). Let us exemplify the computational savings. Let us assume that the calibration functional depends on P parameters, and that the computational cost for computing one instance of the calibration functional is T time units. The combination between adjoint and AD methodology is theoretically guaranteed to produce the gradient of of the calibration functional (namely all P sensitivities with respect to parameters) in a computational time that is not larger than 4-5 times the original time T for computing one instance of the calibration functional. The gradient is also very accurate, up to machine precision, and thus we eliminate any approximation error that may come from using finite difference to compute the gradient. The local optimizer can then run much more efficiently if the gradient of the calibration functional is provided explicitly. For additional details on adjoint plus AD the reader is referred to [90]

10 Conclusion

We have surveyed various methodologies for constructing the implied volatility surface. We have also discussed related topics which may contribute to the successful construction of volatility surface in practice: conditions for arbitrage/non arbitrage in both strike and time, how to perform extrapolation outside the core region, choice of calibrating functional and selection of optimization algorithms.

References

[1] Decision tree for optimization software. http://plato.asu.edu/guide.html.

[2] Y. Achdou and O. Pironneau. *Computational methods for option pricing*. SIAM, 2005.

[3] Y. Aït-Sahalia and J. Duarte. Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47, 2003.

[4] L. B. G. Andersen and J. Andreasen. Jump-diffusion processes: Volatility smile fitting and numerical methods for pricing. *Review of Derivatives Research*, 4:231–262, 2000.

[5] L.B.G. Andersen and V.V. Piterbarg. Moment explosion in stochastic volatility models. *Finance and Stochastics*, 11:29–50, 2007.

[6] J Andreasen and B. Huge. Volatility interpolation. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1694972, 2010.

[7] J Andreasen and B Huge. Volatility interpolation. *Risk magazine*, March, 2011.

[8] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Appl. Math. Finance*, 4(1):37–64, 1997.

[9] M. Avellaneda and Y. Zhu. A risk-neutral stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 1(2):289–310, 1998.

[10] E. Ayache. Can anyone solve the smile problem? a post-scriptum. http://www.ito33.com/sites/default/files/articles/0501_ayache.pdf, 2005.

[11] E. Ayache, P. Henrotte, S. Nassar, and X. Wang. Can anyone solve the smile problem? *Wilmott magazine*, January:78–96, 2004.

[12] O. Barndorff-Nielsen. Processes of normal inverse gaussian type. *Finance and Stochastics*, 2:41–68, 1998.

[13] S. Benaim, M. Dodgson, and D Kainth. An arbitrage-free method for smile extrapolation. http://www.quarchome.org/RiskTailsPaper_v5.pdf, 2009.

[14] S. Benaim and P. Friz. Regular variation and smile asymptotics. *Mathematical Finance*, 19(1):1–12, 2009.

[15] S. Benaim and P. Friz. Smile asymptotics II: Models with known moment generating functions. *Journal of Applied Probability*, 45:16–32, 2009.

[16] E. Benhamou, E. Gobet, and M. Miri. Closed forms for european options in a local volatility model. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1275872, 2008.

[17] E. Benhamou, E. Gobet, and M. Miri. Smart expansion and fast calibration for jump diffusions. *Finance and Stochastics*, 13(4)(4):563–589, 2009.

[18] E. Benhamou, E. Gobet, and M. Miri. Time dependent Heston model. *SIAM Journal of Financial Mathematics*, 1:289–325, 2010.

[19] E. Benhamou, E. Gobet, and M. Miri. Analytical formulas for local volatility model with stochastic rates. *Quantitative Finance*, 2011.

[20] M. Benko, M. R. Fengler, W. K. Hardle, and M. Kopa. On extracting information implied in options. *Computational Statistics*, 22(4):543–553, 2007.

[21] H Berestycki, J Busca, and I. Florent. Computing the implied volatility in stochastic volatility models. *Communications on Pure and Applied Mathematics*, 57:1352 – 1373, 2004.

[22] L. Bergomi. Smile Dynamics I. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1493294, 2004.

[23] L. Bergomi. Smile Dynamics II. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1493302, 2005.

[24] L. Bergomi. Smile Dynamics III. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1493308, 2008.

[25] L. Bergomi. Smile Dynamics IV. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1520443, 2009.

[26] D.A. Bloch. A practical guide to implied and local volatility. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1538808, 2010.

[27] D.A. Bloch and C.A. Coello Coello. Smiling at evolution. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1627645, 2010.

[28] S. Borovkova and F. J. Permana. Implied volatility in oil markets. *Computational Statistics and Data Analysis*, 53:2022–2039, April 2009.

[29] D Brecher. Pushing the limits of local volatility in option pricing. *Wilmott Magazine*, September:6–15, 2006.

[30] D.C. Brody, I.R.C. Buckley, and I.C. Constantinou. Option price calibration from Rényi entropy. *Physics Letters A*, 366:298–307, July 2007.

[31] P. W. Buchen and M. Kelly. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, 31(01):143–159, March 1996.

[32] H Buehler. *Volatility markets.* PhD thesis, 2008.

[33] P. Carr, H. Geman, D. Madan, and M. Yor. The fine structure of asset returns: an empirical investigation. *Journal of Business*, 75(2):305–332, 2002.

[34] P. Carr and D. B. Madan. A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3):125–130, 2005.

[35] P. Carr and D.B. Madan. Option valuation using the Fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1998.

[36] P. Carr and L. Wu. A new simple approach for constructing implied volatility surfaces. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1701685, 2010.

[37] A Castagna. Implied volatility surface. onlineathttp://www.iasonltd.com/FileUpload/files/pdf_res_3.pdf, 2007.

[38] W. Cheng, N. Costanzino, A. L. Mazzucato, J. Liechty, and V. Nistor. Closed-form asymptotics for local volatility models. http://www.personal.psu.edu/users/j/c/jcl12/Closed%20Form%20Asymptotics.pdf, 2009.

[39] P. Christoffersen and K. Jacobs. The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2)(2):291–318, May 2004.

[40] R. Cont and S. Ben Hamida. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance*, 8(4):43–76, 2005.

[41] R. Cont and P. Tankov. Calibration of jump-diffusion option pricing models: A robust non-parametric approach. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=332400, 2002.

[42] R Cont and P. Tankov. *Financial modeling with jump processes*. Chapman and Hall, 2004.

[43] R. Cont and P. Tankov. Non-parametric calibration of jump-diffusion option pricing models. *Journal of Compuattional Finance*, 7:1–49, 2004.

[44] R. Cont and P. Tankov. Retrieving Levy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, 45:1–25, 2006.

[45] J. M. Corcuera, F. Guillaume, P. Leoni, and W. Schoutens. Implied Lévy volatility. *Quantitative Finance*, 9(4):383–393, 2009.

[46] L. Cousot. Conditions on option prices for absence of arbitrage and exact calibration. *Journal of Banking and Finance*, 31(11):3377–3397, 2007.

[47] J Crosby. Practicalities of pricing exotic derivatives. http://www.john-crosby.co.uk/pdfs/JCrosby_OxfordJune2010_Exotics.pdf, 2010.

[48] T. Daglish, J. Hull, and W. Suo. Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative finance*, 7(5):507–524, 2007.

[49] M.H.A. Davis and D.G. Hobson. The range of traded option prices. *Mathematical Finance*, 17(1):1–14, 2007.

[50] M. Decamps and A. De Schepper. Atomic implied volatilities. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1279363, 2008.

[51] P. DeMarco. Quasi-explicit calibration of Gatheral SVI model. http://www.zeliade.com/whitepapers/zwp-0005.pdf, 2009.

[52] M.V. Deryabin. On implied volatility reconstruction for energy markets. *International Journal of Human and Social Sciences*, 5, 2010.

[53] M.V. Deryabin. Implied volatility surface reconstruction for energy markets: spot price modeling versus surface parametrization. *Journal of Energy Markets*, 4(2), 2011.

[54] K. Detlefsen and W. K. Hardle. Calibration design of implied volatility surfaces. *Journal of Data Science*, 6:303–312, 2008.

[55] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

[56] B. Dumas, J. Fleming, and R.E. Whaley. Implied volatility functions: Empirical tests. *The Journal of Finance*, (53), 1998.

[57] V. Durrleman. From implied to spot volatilities. *Finance and Stochastics*, 14(2):157–177, 2010.

[58] V. Durrleman and N. El Karoui. Coupling smiles. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1005332, 2007.

[59] A. Elices. Models with time-dependent parameters using transform methods: application to Heston model,. http://arxiv.org/ftp/arxiv/papers/0708/0708.2020.pdf, 2008.

[60] B. Engelmann, F. Koster, and D. Oeltz. Calibration of the Heston local stochastic volatility model: A finite volume scheme. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1823769, 2011.

[61] M. R Fengler. *Semiparametric Modeling of Implied Volatility*. Springer, 2005.

[62] M.R. Fengler. Option data and modeling BSM implied volatility. In *Handbook of computational finance*. 2010.

[63] J. Figueroa-Lopez and M. Forde. The small-maturity smile for exponential Levy models. http://arxiv.org/abs/1105.3180, 2011.

[64] M. Forde and A. Jacquier. Small-time asymptotics for implied volatility under a general local-stochastic volatility model. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1514125, 2009.

[65] M. Forde and A. Jacquier. Small-time asymptotics for implied volatility under the Heston model. *International Journal of Theoretical and Applied Finance*, 12(6):861–876, 2009.

[66] M. Forde and A. Jacquier. Small time asymptotics for the implied volatility under the Heston model. *International Journal of Theoretical and Applied Finance*, 12(6), 2009.

[67] M. Forde and A. Jacquier. The large-maturity smile for the Heston model. *Finance and Stochastics*, In Press, 2011.

[68] M. Forde and A. Jacquier. Small-time asymptotics for an uncorrelated local-stochastic volatility model. http://www.math.tu-berlin.de/~jacquier/index_files/Forde,%20Jacquier%20-%20Small%20time%20Geodesics.pdf, 2011.

[69] M. Forde, A. Jacquier, and R. Lee. The small time smile and term structure of implied volatility under the Heston model. http://math.uchicago.edu/~rl/FJL.pdf, 2011.

[70] M. Forde, A Jacquier, and A. Mijatovic. Asymptotic formulae for implied volatility under the Heston model. *Proceedings of the Royal Society A*, 466(2124):3593–3620, 2010.

[71] P. Friz, S. Gerhold, A. Gulisashvili, and S. Sturm. On refined volatility smile expansion in the Heston model. *Quantitative Finance*, In Press, 2011.

[72] P.K. Friz, J. Gatheral, and S. Benaim. Volatility derivatives and the implied volatility smile. http://www.cfr.statslab.cam.ac.uk/events/content/20056/friz.pdf, 2005.

[73] M. Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. http://www-csfi.sigmath.es.osaka-u.ac.jp/faculty/personal/fukasawa/20091202.pdf, 2010.

[74] S. Galluccio and Y. Le Cam. Implied calibration and moments asymptotics in stochastic volatility jump diffusion models. http://www.maths.univ-evry.fr/prepubli/251.pdf, 2007.

[75] K. Gao and R. Lee. Asymptotics of implied volatility to arbitrary order. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1768383, 2011.

[76] J Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. http://www.math.nyu.edu/fellows_fin_math/gatheral/madrid2004.pdf, 2004.

[77] J. Gatheral. *The Volatility Surface: A practitioner's guide*. Wiley Finance, 2006.

[78] J. Gatheral, E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang. Asymptotics of implied volatility in local volatility models. *Mathematical Finance*, 21:In Press, 2011.

[79] P. Gauthier and P.-Y. H. Rivaille. Fitting the smile: Smart parameters for SABR and Heston. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496982, 2009.

[80] P. Glasserman and Q. Wu. Forward and future implied volatility. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1636123, 2010.

[81] A. Gulisashvili. Asymptotic formulas with error estimates for call pricing functions and the implied volatility at extreme strikes. *SIAM Journal on Financial Mathematics*, 1:609–641, 2010.

[82] S. Gurrieri. A class of term structures for SVI implied volatility. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1779463, 2010.

[83] I. Gyongy. Mimicking the one-dimensional marginal distributions of processes having an Ito differential. *Probability Theory and Related Fields*, 71:501–516, 1986.

[84] R. Hafner. *Stochastic Implied Volatility: A Factor-Based Model*. Springer, 2004.

[85] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing smile risk. *Wilmott*, 1(8):84–108, 2002.

[86] P. Henry-Labordere. A general asymptotic implied volatility for stochastic volatility models. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=698601, 2005.

[87] P. Henry-Labordere. *Geometry and Modeling in Finance: Advanced Methods in Option Pricing*. Chapman and Hall, 2008.

[88] P. Henry-Labordere. Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach. *Risk Magazine*, September, 2009.

[89] S. Herzel. Arbitrage opportunities on derivatives: A linear programming approach. *Dynamics of Continuous, Discrete and Impulsive System*, 12:589–606, 2005.

[90] C. Homescu. Adjoints and automatic (algorithmic) differentiation in computational finance. papers.ssrn.com/sol3/papers.cfm?abstract_id=1828503, 2011.

[91] B Huge. Volatility interpolation. In *Quant Congress Conference*, 2010.

[92] A. Jacquier. Heston model. http://www.zeliade.com/whitepapers/zwp-0004.pdf, 2010.

[93] A. Jacquier. *Implied volatility asymptotics under affine stochastic volatility models*. PhD thesis, Imperial College, 2010.

[94] S. Johnson and B. Nonas. Arbitrage-free construction of the swaption cube. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1330869, 2009.

[95] N. Kahale. An arbitrage-free interpolation of volatilities. *Risk Magazine*, 17:102–106, 2004.

[96] R. Kamp van der. Local volatility modelling. Master's thesis, University of Twente, 2009.

[97] M. Keller-Ressel. Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance*, 21:73–98, 2011.

[98] F. Kilin. Accelerating the calibration of stochastic volatility models. *Journal of Derivatives*, 18(3):7–16, 2011.

[99] V.P Kong and B.H. Ong. Shape preserving approximation by spatial cubic splines. *Computer Aided Geometric Design*, 26:883–903, 2009.

[100] S. G. Kou. A jump diffusion model for option pricing. *Management Science*, 48:1086–1101, 2002.

[101] P. Laurence. Implied volatility, fundamental solutions, asymptotic analysis and symmetry methods. http://www.math.nyu.edu/~laurence/SabCalTechFinalApr242008.pdf, 2008.

[102] M. Laurini. Imposing no-arbitrage conditions in implied volatility surfaces using constrained smoothing splines. http://www.mlcc.com.br/artigos/cobsimp.pdf, 2007.

[103] R. Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, 14(3):469–480, 2004.

[104] R.W. Lee. Implied volatility: Statics, dynamics, and probabilistic interpretation. In *Recent Advances in Applied Probability*. Springer, 2004.

[105] E. Lindstrom, J. Strojby, M. Broden, M. Wiktorsson, and J. Holst. Sequential calibration of options. *Computational Statistics and Data Analysis*, 52(6):2877–2891, 2008.

[106] A. Lipton. The vol smile problem. *Risk Magazine*, (February):61–65, 2002.

[107] D. Madan, P. Carr, and E. Chang. The variance gamma process and option pricing model. *European Financial Review*, 2:79–195, 1998.

[108] D. Madan and M. Yor. Representing the CGMY and Meixner Levy Processes as time changed Brownian motions. *Journal of Computational Finance*, Fall:27–47, 2008.

[109] J. Maruhn. On-the-fly bid/ask-vol fitting with applications in model calibration. In *SIAM Conference on Financial Mathematics & Engineering, San Francisco, USA, November 19-20*, 2010.

[110] J. Maruhn. Calibration of stochastic and local stochastic volatility models. In *Global Derivatives Trading & Risk Management, Paris, April 11-15*, 2011.

[111] A. Medvedev and O. Scaillet. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. http://www.hec.unige.ch/www/hec/m1/RecherchePublications/CahiersRecherche/2006/simpleText/0/content_files/file8/2006.01.01.01.01.01.01.01.01.01.01.01.01.pdf, 2006.

[112] A.N. Medvedev. Implied volatility at expiration. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1090200, 2008.

[113] F. Mercurio. No-arbitrage conditions for a finite options system. http://www.fabiomercurio.it/NoArbitrage.pdf.

[114] S. Mikhailov and U. Nogel. Heston stochastic volatility model implementation, calibration and some extensions. *Wilmott Magazine*, July:74–79, 2004.

[115] J. Obloj. Fine-tune your smile: Correction to Hagan et al. http://arxiv.org/abs/0708.0998, 2008.

[116] G Orosi. Calibration of implied volatility surface using Tikhonov regularization. In *Second International Conference on Engineering Systems Management and Its Applications (ICESMA)*, 2010.

[117] G. Orosi. Improved implementation of local volatility and its application to S&P 500 index options. *Journal of Derivatives*, 17(3):53–64, 2010.

[118] Y. Osajima. The asymptotic expansion formula of implied volatility for dynamic SABR model and FX hybrid model. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=965265, 2007.

[119] L. Paulot. Asymptotic implied volatility at the second order with application to the SABR model. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1413649, 2009.

[120] V. V. Piterbarg. Markovian projection for volatility calibration. *Risk Magazine*, 20:84–89, 2007.

[121] K. Price, R. M. Storn, and J. A. Lampinen. *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.

[122] C. Rama and J. da Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2:45–60, 2002.

[123] A. Reghai. Joint dynamics using asymptotic methods. http://www.cmap.polytechnique.fr/~euroschoolmathfi10/Joint%20dynamics%20Asymptotics%20-%20Reghai%20Natixis.pdf, 2010.

[124] E. Reiner. The characteristic curve approach to arbitrage-free time interpolation of volatility. In *ICBI Global Derivatives and Risk Management*, 2004.

[125] D. Reiswich. A note on standard no-arbitrage conditions. http://ssrn.com/abstract=1619844, 2010.

[126] Y. Ren, D. Madan, and M. Qian Qian. Calibrating and pricing with embedded local volatility models. *Risk Magazine*, September:138–143, 2007.

[127] M Roper. Arbitrage free implied volatility surfaces. www.maths.usyd.edu.au/u/pubs/publist/preprints/2010/roper-9.pdf, 2010.

[128] M. Roper. Implied volatility: small time-to-expiry asymptotics in exponential Levy models. http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2010/roper-13.html., 2010.

[129] W. Schoutens. The Meixner process: theory and applications in finance. https://perswww.kuleuven.be/~u0009713/004wsreport.pdf, 2002.

[130] K. Shiraya and A. Takahashi. Pricing average options on commodities. http://park.itc.u-tokyo.ac.jp/takahashi-lab/WPs/p35.pdf, 2010.

[131] K. Shiraya, A. Takahashi, and Toda M. Pricing barrier and average options under stochastic volatility environment. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1491937, 2009.

[132] J. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, 2004 (Second Edition).

[133] A. Takahashi. An asymptotic expansion approach in finance. http://www.carf.e.u-tokyo.ac.jp/pdf/workingpaper/fseries/104.pdf, 2007.

[134] A. Takahashi, K. Takehara, and M. Toda. Computation in an asymptotic expansion method. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1413924, 2009.

[135] A. Takahashi, K. Takehara, and M. Toda. A general computation scheme for a high-order asymptotic expansion method. Technical report, CIRJE-F-787, 2011.

[136] A. Takahashi and T. Yamada. An asymptotic expansion with push-down of malliavin weights. Technical report, 2009.

[137] K. Takehara, A. Takahashi, and M. Toda. New unified computational algorithm in a high-order asymptotic expansion scheme. http://www.cirje.e.u-tokyo.ac.jp/research/dp/2010/2010cf728.pdf, 2010.

[138] P. Tankov. Levy processes in finance and risk management. http://www.math.jussieu.fr/~tankov/wccf_paper.pdf, 2007.

[139] P. Tankov and E. Voltchkova. Jump-diffusion models: a practitioner's guide. *Banques et Marches*, 99, 2009.

[140] S.M. Taylor. Explicit density approximations for local volatility models using heat kernel expansions. http://ssrn.com/paper=1800415, 2011.

[141] M. Tehranci. Asymptotics of implied volatility far from maturity. *Journal of Applied Probability*, 46(3):629–650, 2009.

[142] I. Vollrath and J. Wendland. Calibration of interest rate and option models using Differential Evolution. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1367502, 2009.

[143] Y. Wang, H. Yin, and L. Qi. No-arbitrage interpolation of the option price function and its reformulation. *Journal of Optimization Theory and Applications*, 120(3):627–649, 2004.

[144] Z. Wang. 'Effective' parameters for stochastic volatility models. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1071768, 2007.

[145] G. Wolberg and I. Alfy. An energy-minimization framework for monotonic cubic spline interpolation. *Journal of Computational and Applied Mathematics*, 143:145–188, 2002.

[146] S. J. Wright and M. J. Tenny. A feasible trust-region Sequential Quadratic Programming algorithm. *SIAM Journal on Optimization*, 14:1074–1105, 2004.

[147] Q. Wu. Series expansion of the SABR joint density. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1367966, 2010.