Universität St.Gallen

Option data and modeling BSM implied volatility

Matthias R. Fengler

December 2010 Discussion Paper no. 2010-32

Department of Economics University of St. Gallen

Editor: Martina Flockerzi
University of St. Gallen
Department of Economics
Varnbüelstrasse 19
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email vwaabtass@unisg.ch
Publisher: Department of Economics
University of St. Gallen
Varnbüelstrasse 19
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Electronic Publication: http://www.vwa.unisg.ch

Option data and modeling BSM implied volatility

Matthias R. Fengler

Author's address:

Matthias R. Fengler, Prof. Dr.
Institute for Mathematics and Statistics
University of St. Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 224 2457
Fax +41 71 224 2894
Email matthias.fengler@unisg.ch
Website www. mathstat.unisg.ch

Abstract

This contribution to the Handbook of Computational Finance, Springer-Verlag, gives an overview on modeling implied volatility data. After introducing the concept of Black-Scholes-Merton implied volatility (IV), the empirical stylized facts of IV data are reviewed. We then discuss recent results on IV surface dynamics and the computational aspects of IV. The main focus is on various parametric, semi- and nonparametric modeling strategies for IV data, including ones which respect no-arbitrage bounds.

Keywords

Implied volatility

JEL Classification

G13

Forthcoming: Handbook of Computational Finance, Springer-Verlag

1 Introduction

The discovery of an explicit solution for the valuation of European style call and put options based on the assumption a Geometric Brownian motion driving the underlying asset constitutes a landmark in the development of modern financial theory. First published in Black and Scholes (1973), but relying heavily on the notion of no-arbitrage in Merton (1973), this solution is nowadays known as the Black-Scholes-Merton (BSM) option pricing formula. In recognition of this achievement, Myron Scholes and Robert C. Merton were awarded the Nobel prize in economics in 1997 (Fischer Black had already died by this time).

Although it is widely acknowledged that the assumptions underlying the BSM model are far from realistic, the BSM formula still enjoys unrivalled popularity in financial practice. This is not so much because practitioners believe in the model as a good description of market behavior, but rather because it serves as a convenient mapping device from the space of option prices to a single real number called the *(BSM-)implied volatility*. Indeed, the only unknown parameter involving the BSM formula is the volatility. Backed out of given option prices it allows for straight forward comparisons of the relative expensiveness of options across various strikes, expiries and underlying assets. In practice calls and puts are thus quoted in terms of implied volatility.

For illustration consider Figure 1 displaying implied volatility (IV) as observed on 28 Oct. 2008 and computed from options traded on the futures exchange EUREX, Frankfurt. IV is plotted against relative strikes and time to expiry. Due to institutional conventions, there is a very limited number of expiry dates, usually one to three months apart for short-dated options and six to twelve months apart for longer-dated ones, while the number of strikes for each expiry is more finely spaced. The function resulting for a fixed expiry is frequently called the 'IV smile' due to its U-shaped pattern. For a fixed (relative) strike across several expiries one speaks of the term structure of IV. Understandably, the non-flat surface, which also fluctuates from day to day, is in strong violation to the assumption of a Geometric Brownian motion underlying the BSM model.

Although IV observations are observed on this degenerate design, practitioners think of them as stemming from a smooth and well-behaved surface. This view is due to the following objectives in option portfolio management: (i) market makers quote options for strike-expiry pairs which are illiquid or not listed; (ii) pricing engines, which are used to price exotic options and which are based on far more realistic assumptions than the BSM model, are calibrated against an observed IV surface; (iii) the IV surface 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); (iv)

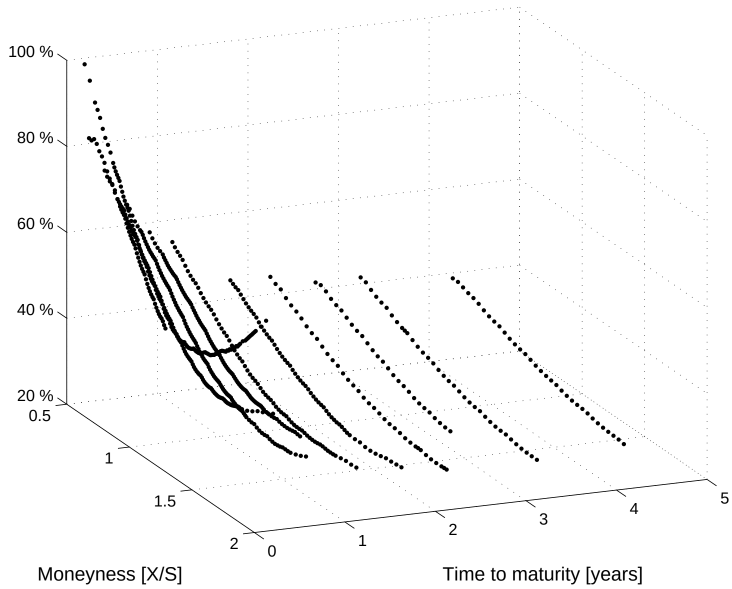

Fig. 1. IV surface of DAX index options from 28 Oct. 2008, traded at the EUREX. IV given in percent across a spot moneyness metric, time to expiry in years.

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

Each of these applications requires suitably chosen interpolation and extrapolation techniques or a fully specified model of the IV surface. This suggests the following structure of this contribution: Section 2 introduces the BSM-implied volatility and in Section 3 we outline its stylized facts. No-arbitrage constraints on the IV surface are presented in Section 4. In Section 5, recent theoretical advances on the asymptotic behavior of IV are summarized. Approximation formulae and numerical methods to recover IV from quoted option prices are reviewed in Section 6. Parametric, semi- and nonparametric modeling techniques of IV are considered in Section 7.

2 The BSM model and implied volatility

We consider an economy on the time interval $[0, T^*]$. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{0 \leq t \leq T^*}$ which is generated by a Brownian motion $(W_t)_{0 \leq t \leq T^*}$ defined on this space, see e.g. Steele (2000). A stock price $(S_t)_{0 \leq t \leq T^*}$, adapted to $(\mathcal{F}_t)_{0 \leq t \leq T^*}$ (paying no-dividends for simplicity) is modeled by the Geometric Brownian motion satisfying the stochastic differential equation

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t \ , \tag{1}$$

where μ denotes the (constant) instantaneous drift and σ^2 measures the (constant) instantaneous variance of the return process of $(\log S_t)_{t\geq 0}$. We furthermore assume the existence of a riskless money market account paying interest r. A European style call is a contingent claim paying at some expiry date T, $0 < T \leq T^*$, the amount $\psi_c(S_T) = (S_T - X)^+$, where $(\cdot)^+ \stackrel{\text{def}}{=} \max(\cdot, 0)$ and X is a fixed number, the exercise price. The payoff of a European style put is given by $\psi_p(S_T) = (X - S_T)^+$.

Under these assumptions, it can be shown that the option price $H(S_t, t)$ is a function in the space $\mathcal{C}^{2,1}\big(\mathbb{R}^+ \times (0, T)\big)$ satisfying the partial differential equation

$$0 = \frac{\partial H}{\partial t} + rS\frac{\partial H}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 H}{\partial S^2} - rH \tag{2}$$

subject to $H(S_T, T) = \psi_i(S_T)$ with $i \in \{c, p\}$.

The celebrated BSM formula for calls solving (2) with boundary condition $\psi_c(S_T)$ is found to be

$$C_t^{BSM}(X, T) = S_t \Phi(d_1) - e^{-r(T-t)} X \Phi(d_2) \ , \tag{3}$$

with

$$d_1 = \frac{\log(S_t/X) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}} \ , \tag{4}$$

$$d_2 = d_1 - \sigma\sqrt{T - t} \ , \tag{5}$$

and where $\Phi(v) = \int_{-\infty}^{v} \varphi(u)\, du$ is the cdf of the standard normal distribution with pdf $\varphi(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$ for $v \in \mathbb{R}$.

Given observed market prices $\widetilde{C}_t$, one defines – as first introduced by Latané and Rendelman (1976) – implied volatility as

$$\widehat{\sigma} : \qquad C_t^{BSM}(X, T, \widehat{\sigma}) - \widetilde{C}_t = 0 \ . \tag{6}$$

Due to monotonicity of the BSM price in σ, there exists a unique solution $\widehat{\sigma} \in \mathbb{R}^+$. Note that the definition in (6) is not confined to European options. It is also used for American options, which can be exercised at any time in $[0, T]$. In this case, as no explicit formulae for American style options exists, the option price is computed numerically, for instance by means of finite difference schemes, Randall and Tavella (2000).

In the BSM model volatility is just a constant, whereas empirically, IV displays a pronounced curvature across strikes X and different expiry days T. This gives rise to the notion of an IV surface as the mapping

$$\widehat{\sigma} : (t, X, T) \to \widehat{\sigma}_t(X, T) \ . \tag{7}$$

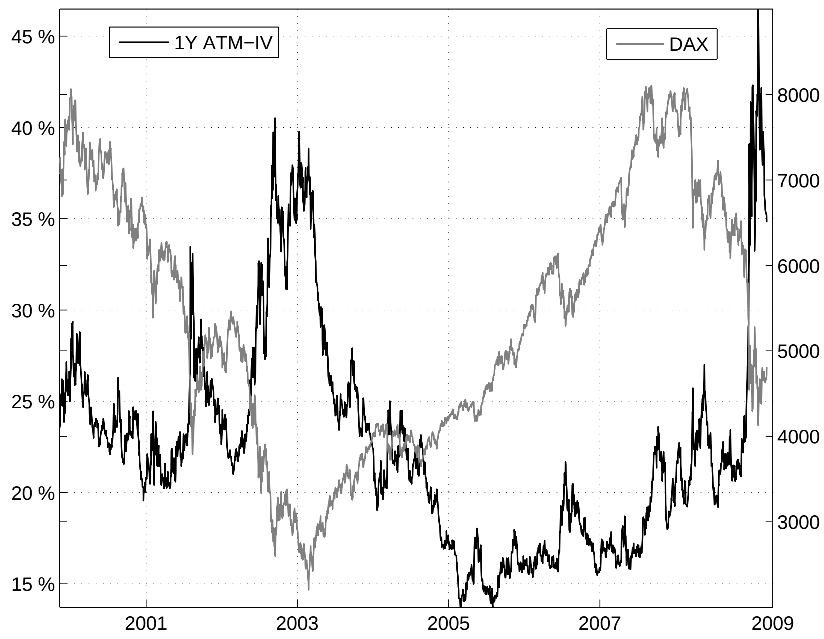

Fig. 2. Time series of 1Y ATM IV (left axis, black line) and DAX index closing prices (right axis, gray line) from 2000 to 2008.

In Figure 2, we plot the time series of 1Y at-the-money IV of DAX index options (left axis, black line) together with DAX closing prices (right axis, gray line). An option is called at-the-money (ATM) when the exercise price is equal or close to the spot (or to the forward). The index options were traded at the EUREX, Frankfurt (Germany), from 2000 to 2008. As is visible IV is subject to considerable variations. Average DAX index IV was about 22% with significantly higher levels in times of market stress, such as after the World Trade Center attacks 2001, during the pronounced bear market 2002-2003 and the financial crisis end of 2008.

3 Stylized facts of implied volatility

The IV surface displays a number of static and dynamic stylized facts which we demonstrate here using the present DAX index option data dating from 2000 to 2008. These facts can be observed for any equity index market. They similarly hold for stocks. Other asset classes may display different features, for instance, smiles may be more shallow, symmetric or even upward-sloping, but this does not fundamentally change the smile phenomenon. A more complete synthesis can be found in Rebonato (2004).

Stylized facts of IV are as follows:

1. The smile is very pronounced for short expiries and becomes flattish for longer dated options. This fact was already visible in Figure 1.
2. As noted by Rubinstein (1994) this has not always been the case. The strong asymmetry in the smile first appeared after the 1987 market turmoil.
3. For equity options, both index and stocks, the smile is negatively skewed. We define the 'skew' here by $\frac{\partial \hat{\sigma}^2}{\partial m}\Big|_{m=0}$, where m is log-forward moneyness as defined in Section 5. Figure 3 depicts the time series of the DAX index skew (left axis) for 1M and 1Y options. The skew is negative throughout and – particularly the short-term skew – increases during times of crisis. For instance, skews increase in the aftermath of the dot-com boom 2001 to 2003, or spike at 11 Sep. 2001 and during the heights of the financial crisis 2008. As theory predicts, see Section 5, the 1Y IV skew has most of the time been flatter than the 1M IV skew.
4. Fluctuations of the short-term skew are much larger. Figure 4 gives the quantiles of the skew as a function of time to expiry. Similar patterns also apply to IV levels and returns.
5. The IV surface term structure is typically upward sloping (i.e. has increasing levels of IV for longer dated options) in calm times, while in times of crisis it is downward sloping with short dated options having higher levels of IV then longer dated ones. This is seen in Figure 3 giving the difference of 1M ATM IV minus 1Y ATM in terms of percentage points on the right axis. A positive value therefore indicates a downward sloping term structure. During the financial crisis the term structure slope achieved unprecedented levels. Humped profiles can be observed as well.
6. Returns of the underlying asset and returns of IV are negatively correlated. For the present data set we find a correlation between 1M ATM IV and DAX returns of $\rho = -0.69$.
7. IV appears to be mean-reverting, see Figure 2, but it is usually difficult to confirm mean reversion statistically, since IV data is often found to be nearly integrated, see Fengler et al. (2007) for a discussion.
8. Shocks cross the IV surface are highly correlated, as can be observed from the comovements of IV levels in Figure 2 and the skew and the term structure in Figure 3. In consequence IV surface dynamics can be decomposed into a small number of driving factors, see Chapter ???Set link to Domininik's Contribution??? of this handbook.

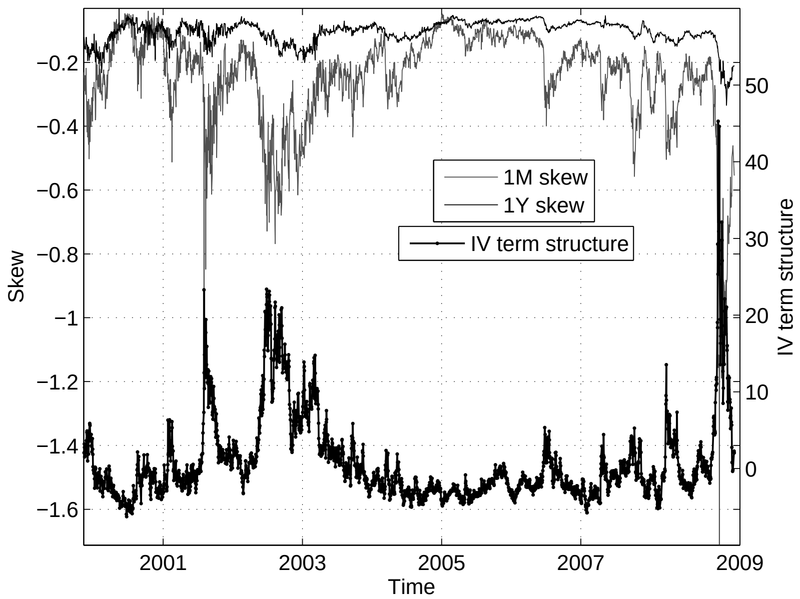

Fig. 3. Time series of 1M and 1Y IV skew (left axis, gray line and black line respectively) and time series of the IV term structure (right axis, black dotted line). Skew is defined as $\frac{\partial \hat{\sigma}^2}{\partial m}\Big|_{m=0}$, where m is log-forward moneyness. The derivative is approximated by a finite difference quotient. IV term structure is the difference between 1M ATM and 1Y ATM in terms of percentage points. Negative values indicate an upward sloping term structure.

4 Arbitrage bounds on the implied volatility surface

Despite the rich empirical behavior, IV cannot simply assume any functional form. This is due to constraints imposed by no-arbitrage principles. For IV, these constraints are very involved, but are easily stated indirectly in the option price domain. From now on, we set $t = 0$ and suppress dependence on t for sake of clarity.

We state the bounds using a (European) call option; deriving the corresponding bounds for a put is straightforward. The IV function must be such that the call price is bounded by

$$\max\left(S - e^{-rT}X, 0\right) \leq C(X,T) \leq S \; . \tag{8}$$

Moreover, the call price must be a decreasing and convex function in X, i.e.

$$-e^{-rT} \leq \frac{\partial C}{\partial X} \leq 0 \quad \text{and} \quad \frac{\partial^2 C}{\partial X^2} \geq 0 \; . \tag{9}$$

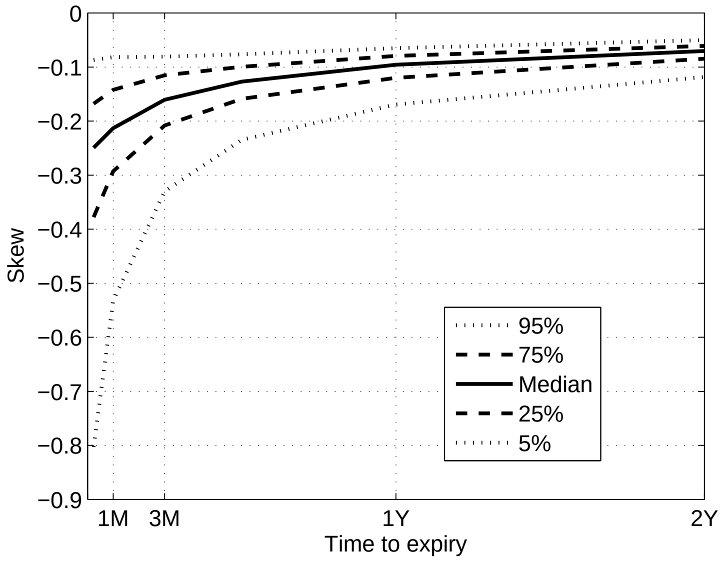

Fig. 4. Empirical quantiles of the ATM IV skew as a function of time to expiry. Skew is defined as $\left.\frac{\partial \widehat{\sigma}^2}{\partial m}\right|_{m=0}$, where m is log-forward moneyness.

To preclude calendar arbitrage, prices of American calls for the same strikes must be nondecreasing across increasing expiries. This statement does not hold for European style calls because their theta can change sign. No-arbitrage implies, however, that there exists a monotonicity relationship along forward-moneyness corrected strikes (also in the presence of dividend yield), see Reiner (2000), Gatheral (2004), Kahalé (2004), Reiner (2004), Fengler (2009). Denote by $x = X/F^T$ forward-moneyness, where F^T is a forward with expiry T, and by $T_1 < T_2$ the expiry dates of two call options whose strike prices X_1 and X_2 are related by forward-moneyness, i.e. $x_1 = x_2$. Then

$$C(X_2, T_2) \geq C(X_1, T_1) \ . \tag{10}$$

Most importantly, this results implies that total implied variance must be nondecreasing in forward-moneyness to preclude arbitrage. Defining total variance as $\nu^2(x, T) \stackrel{\text{def}}{=} \widehat{\sigma}^2(x, T)\, T$, we have

$$\nu^2(x, T_2) > \nu^2(x, T_1) \ . \tag{11}$$

Relationship (11) has the important consequence that one can visually check IV smiles for calendar arbitrage by plotting total variance across forward moneyness. If the lines intersect, (11) is violated.

5 Implied volatility surface asymptotics

Many of the following results had the nature of conjectures and were generalized and rigorously derived only very recently. Understanding the behavior of IV for far expiries and far strikes is of utter importance for extrapolation problems often arising in practice.

Throughout this section set $r = 0$ and $t = 0$. This is without loss of generality since in the presence of nonzero interest rates and dividends yields, the option and underlying asset prices may be viewed as forward prices, see Britten-Jones and Neuberger (2000). Furthermore define log-(forward) moneyness by $m \stackrel{\text{def}}{=} \log x = \log(X/S)$ and total (implied) variance by $\nu^2 \stackrel{\text{def}}{=} \widehat{\sigma}^2 T$. Let $S = (S_t)_{t \geq 0}$ be a non-negative martingale with $S_0 > 0$ under a fixed risk-neutral measure.

5.1 Far expiry asymptotics

The results of this section can be found in more detail in Tehranchi (2010) whom we follow closely.

The first theorem shows that the IV surface flattens for infinitely large expiries.

Theorem 1 (Rogers and Tehranchi (2009)). *For any $M > 0$ we have*

$$\lim_{T \to \infty} \sup_{m_1, m_2 \in [-M, M]} |\widehat{\sigma}(m_2, T) - \widehat{\sigma}(m_1, T)| = 0 \, .$$

Note that this result does not hinge upon the central limit theorem, mean-reversion of spot volatility etc., only the existence of the martingale measure. In particular, $\lim_{T \to \infty} \widehat{\sigma}(m, T)$ does not need to exist for any m.

The rate of flattening of the IV skew can be made more precise by the following result. It shows that the flattening behavior of the IV surface as described in Section 3 is not an empirical artefact, but has a well-founded theoretical underpinning (for earlier, less general arguments see Hodges (1996) and Carr and Wu (2003)).

Theorem 2 (Rogers and Tehranchi (2009)).

(i) For any $0 \leq m_1 < m_2$ we have

$$\frac{\widehat{\sigma}(m_2, T)^2 - \widehat{\sigma}(m_1, T)^2}{m_2 - m_1} \leq \frac{4}{T} \, .$$

(ii) For any $m_1 < m_2 \leq 0$

$$\frac{\widehat{\sigma}(m_2, T)^2 - \widehat{\sigma}(m_1, T)^2}{m_2 - m_1} \geq -\frac{4}{T} \, .$$

(iii) If $S_t \xrightarrow{P} 0$ *as* $t \to \infty$, *for any* $M > 0$ *we have*

$$\limsup_{T\to\infty} \sup_{m_1,\, m_2 \in [-M,M],\, m_1 \neq m_2} T \left| \frac{\widehat{\sigma}(m_2,T)^2 - \widehat{\sigma}(m_1,T)^2}{m_2 - m_1} \right| \leq 4 \,.$$

As pointed out by Rogers and Tehranchi (2009), the inequality in (*iii*) is sharp in the sense that there exists a martingale $(S_t)_{t\geq 0}$ with $S_t \xrightarrow{P} 0$ such that

$$T \frac{\partial}{\partial m} \widehat{\sigma}(m,T)^2 \to -4 \,.$$

as $T \to \infty$ uniformly for $m \in [-M, M]$. The condition $S_t \xrightarrow{P} 0$ as $t \to \infty$, is not strong. It holds for most financial models and is equivalent to the statement that

$$C(X,T) = \mathrm{E}[(S_T - X)^+] \to S_0$$

as $T \to \infty$ for some $X > 0$. The BSM formula (3) fulfills it trivially. Indeed one can show that if the stock price process does not converge to zero, then $\lim_{T\to\infty} \widehat{\sigma}(m,T) = 0$, because $\nu^2 < \infty$.

Finally Tehranchi (2009) obtains the following representation formula for IV:

Theorem 3 (Tehranchi (2009)). *For any* $M > 0$ *we have*

$$\lim_{T\to\infty} \sup_{m\in[-M,\,M]} \left| \widehat{\sigma}(m,T) - \sqrt{-\frac{8}{T} \log \mathrm{E}[S_T \wedge 1]} \right|$$

with $a \wedge b = \min(a,b)$. *Moreover there is the representation*

$$\widehat{\sigma}^2_\infty = \lim_{T\to\infty} -\frac{8}{T} \log \mathrm{E}[S_T \wedge 1] \tag{12}$$

whenever this limit is finite.

A special case of this result was derived by Lewis (2000) in the context of the Heston (1993) model. For certain model classes, such as models based on Lévy processes, the last theorem allows a direct derivation of $\widehat{\sigma}_\infty$.

The implication of these results for building an IV surface are far-reaching. The implied variance skew must be bounded by $\left|\frac{\partial \nu^2}{\partial m}\right| \leq 4$ and should decay at a rate of $1/T$ between expiries. Moreover, a constant far expiry extrapolation in $\widehat{\sigma}(m,T_n)$ beyond the last extant expiry T_n is wrong, since the IV surface does not flatten in this case. A constant far expiry extrapolation in $\nu^2(m,T_n)$ beyond T_n is fine, but may not be a very lucky choice given the comments following Theorem 2 number (*iii*).

5.2 Short expiry asymptotics

Roper and Rutkowski (2009) consider the behavior of IV towards small times to expiry. They prove

Theorem 4 (Roper and Rutkowski (2009)). *If $C(X,\epsilon) = (S-X)^+$ for some $\epsilon > 0$ then*

$$\lim_{T\to 0^+} \widehat{\sigma}(X,T) = 0\;. \tag{13}$$

Otherwise

$$\lim_{T\to 0^+} \widehat{\sigma}(X,T) = \begin{cases} \lim_{T\to 0^+} \frac{\sqrt{2\pi}C(X,T)}{S\sqrt{T}} & \text{if } S = X \\ \lim_{T\to 0^+} \frac{|\log(S/X)|}{\sqrt{-2T\log[C(X,T)-(S-X)^+]}} & \text{if } S \neq X \end{cases}\;, \tag{14}$$

in the sense that the LHS is finite (infinite) whenever the RHS is finite (infinite).

The quintessence of this theorem is twofold. First, the asymptotic behavior of $\widehat{\sigma}(X,T)$ as $T \to 0^+$ is markedly different for $S = X$ and $S \neq X$. Note that the ATM behavior of (14) is the well-established Brenner and Subrahmanyam (1988)-Feinstein (1988) formula to be presented in Section 6.1. Second, convergent IV is not a behavior coming for granted. In particular no-arbitrage does not guarantee that a limit exists, see Roper and Rutkowski (2009) for a lucid example. However, the limit of time-scaled IV exists and is zero:

$$\lim_{T\to 0^+} \nu(X,T) = \lim_{T\to 0^+} \widehat{\sigma}\sqrt{T} = 0\;. \tag{15}$$

5.3 Far strike asymptotics

Lee (2004) establishes the behavior of the IV surface as strikes tend to infinity. He finds a one-to-one correspondence between the large-strike tail and the number of moments of S_T, and the small-strike tail and the number of moments of S_T^{-1}. We retain the martingale assumption for $(S_t)_{t\geq 0}$ and $m \stackrel{\text{def}}{=} \log(X/S)$.

Theorem 5 (Lee (2004)). *Define*

$$\widetilde{p} = \sup\{p :\ \mathrm{E}\, S_T^{1+p} < \infty\} \quad \beta_R = \limsup_{m\to\infty} \frac{\nu^2}{|m|} = \limsup_{m\to\infty} \frac{\widehat{\sigma}^2}{|m|/T}\;.$$

Then $\beta_R \in [0,2]$ and

$$\widetilde{p} = \frac{1}{2\beta_R} + \frac{\beta_R}{8} - \frac{1}{2}\;,$$

with the understanding that $1/0 = \infty$. Equivalently,

$$\beta_R = 2 - 4(\sqrt{\widetilde{p}^2 + \widetilde{p}} - \widetilde{p})\;.$$

The next theorem considers the case $m \to -\infty$.

Theorem 6 (Lee (2004)). *Denote by*

$$\widetilde{q} = \sup\{q : \operatorname{E} S_T^{-q} < \infty\} \quad \beta_L = \limsup_{m \to -\infty} \frac{\nu^2}{|m|} = \limsup_{m \to -\infty} \frac{\widehat{\sigma}^2}{|m|/T} \ .$$

Then $\beta_L \in [0, 2]$ *and*

$$\widetilde{q} = \frac{1}{2\beta_L} + \frac{\beta_L}{8} - \frac{1}{2} \ ,$$

with $1/0 = \infty$*, or*

$$\beta_L = 2 - 4(\sqrt{\widetilde{q}^2 + \widetilde{q}} - \widetilde{q}) \ .$$

Roger Lee's results have again vital implications for the extrapolation of the IV surface for far strikes. They show that linear or convex skews for far strikes are wrong by the $\mathcal{O}(|m|^{1/2})$ behavior. More precisely, the IV wings should not grow faster than $|m|^{1/2}$ and not grow slower than $|m|^{1/2}$, unless the underlying asset price is supposed to have moments of all orders. The elegant solution following from these results is to extrapolate ν^2 linearly in $|m|$ with an appropriately chosen $\beta_L, \beta_R \in [0, 2]$.

6 Approximating and computing implied volatility

6.1 Approximation formulae

There is no closed-form, analytical solution to IV, even for European options. In situations when iterative procedures is not readily available, such as in the context of a spreadsheet, or when numerical approaches are not applicable, such as in real time applications, approximation formulae to IV are of high interest. Furthermore, they also serve as good inital values for the numerical schemes discussed in section 6.2.

The most simple approximation to IV, which is due to Brenner and Subrahmanyam (1988) and Feinstein (1988), is given by

$$\widehat{\sigma} \approx \sqrt{\frac{2\pi}{T}} \, \frac{C}{S} \ . \tag{16}$$

The rationale of this formula can be understood as follows. Define by $K \stackrel{\text{def}}{=} S = Xe^{-rT}$ the discounted ATM strike. The BSM formula then simplifies to

$$C = S \left(2 \, \Phi(\sigma \sqrt{T}/2) - 1 \right) \ .$$

Solving for σ yields the semi-analytical formula

$$\sigma = \frac{2}{\sqrt{T}} \ \Phi^{-1} \left(\frac{C + S}{2S} \right) \ , \tag{17}$$

where Φ^{-1} denotes the inverse function of the normal cdf. A first order Taylor expansion of (17) in the neighborhood of $\frac{1}{2}$ yields formula (16). In consequence, it is exact only, when the spot is equal to the discounted strike price.

A more accurate formula, which also holds for in-the-money (ITM) and out-of- the-money (OTM) options (calls are called OTM when $S \ll X$ and ITM when $S \gg X$), is based on a Taylor expansion of third order to Φ. It is due to Li (2005):

$$\widehat{\sigma} \approx \begin{cases} 2\,z\sqrt{\frac{2}{T}} - \frac{1}{\sqrt{T}}\sqrt{8z^2 - \frac{6\alpha}{\sqrt{2}z}} & \text{if} \quad \rho \leq 1.4 \\ \frac{1}{2\sqrt{T}}\left(\alpha + \sqrt{\alpha^2 - \frac{4(K-S)^2}{S(S+K)}}\right) & \text{if} \quad \rho > 1.4\,, \end{cases} \tag{18}$$

where $z = \cos\left[\frac{1}{3}\arccos\left(\frac{3\alpha}{\sqrt{32}}\right)\right]$, $\alpha = \frac{\sqrt{2\pi}}{S+K}(2C+K-S)$ and $\rho = |K-S|SC^{-2}$. The value of the threshold parameter ρ separating the first part, which is for nearly-ATM options, and the second part for deep ITM or OTM options, was found by Li (2005) based on numerical tests.

Other approximation formulae found in the literatur often lack a rigorous mathematical foundation. The possibly most prominent amongst these are those suggested by Corrado and Miller (1996) and Bharadia et al. (1996). The Corrado and Miller (1996) formula is given by

$$\widehat{\sigma} \approx \frac{1}{\sqrt{T}}\frac{\sqrt{2\pi}}{S+K}\left[C - \frac{S-K}{2} + \sqrt{\left(C - \frac{S-K}{2}\right)^2 - \frac{(S-K)^2}{\pi}}\right]\,. \tag{19}$$

Its relative accuracy is explained by the fact that (19) is identical to the second formula in (18) after multiplying the second term under the square root by $\frac{1}{2}(1+K/S)$, which is negligible in most cases, see Li (2005) for the details. Finally, the Bharadia et al. (1996) approximation is given by

$$\widehat{\sigma} \approx \sqrt{\frac{2\pi}{T}}\,\frac{C-(S-K)/2}{S-(S-K)/2}\,. \tag{20}$$

Isengildina-Massa et al. (2007) investigate the accuracy of six approximation formulae. According to their criteria, Corrado and Miller (1996) is the best approximation followed by Li (2005) and Bharadia et al. (1996). This finding holds uniformly also for deviations to up to 1% around ATM (somewhat unfortunate, the authors do not consider a wider range) and up to maturities of 11 months. As a matter of fact, the approximation by Brenner and Subrahmanyam (1988) and Feinstein (1988) is of competing quality for ATM options only.

6.2 Numerical computation of implied volatility

Newton-Raphson

The Newton-Raphson method, which will be the method of first choice in most cases, was suggested by Manaster and Koehler (1982). Denoting the observed market price by $\widetilde{C}$, the approach is described as

$$\sigma_{i+1} = \sigma_i - \left(C_i(\sigma_i) - \widetilde{C}\right) / \frac{\partial C}{\partial \sigma}(\sigma_i) \; , \tag{21}$$

where $C_i(\sigma_i)$ is the option price and $\frac{\partial C}{\partial \sigma}(\sigma_i)$ is the option vega computed at σ_i. The algorithm is run until a tolerance criterion, such as $|\widetilde{C} - C_{i+1}| \leq \epsilon$, is achieved; IV is given by $\widehat{\sigma} = \sigma_{i+1}$. The algorithm may fail, when the vega is close to zero, which regularly occurs for (short-dated) far ITM oder OTM options. The Newton-Raphson method has at least quadratic convergence, and combined with a good choice of the initial value, it achieves convergence within a very small number of steps. Originally, Manaster and Koehler (1982) suggested

$$\sigma_0 = \sqrt{\frac{2}{T} \, |\log(S/X) + rT|} \tag{22}$$

as initial value (setting $t = 0$). It is likely, however, that the approximation formulae discussed in section 6.1 provide initial values closer to the solution.

Regula falsi

The regula falsi is more robust than Newton-Raphson, but has linear convergence only. It is particularly useful when no closed-form expression for the vega is available, or when the price function is kinked as e.g. for American options with high probability of early exercise.

The regula falsi is initialized by two volatility estimates σ_L and σ_H with corresponding option prices $C_L(\sigma_L)$ and $C_H(\sigma_H)$ which need to include the solution. The iteration steps are:

1. Compute

$$\sigma_{i+1} = \sigma_L - \left(C_L(\sigma_L) - \widetilde{C}\right) \frac{\sigma_H - \sigma_L}{C_H(\sigma_H) - C_L(\sigma_L)} \; ; \tag{23}$$

2. if $C_{i+1}(\sigma_{i+1})$ and $C_L(\sigma_L)$ have same sign, set $\sigma_L = \sigma_{i+1}$; if $C_{i+1}(\sigma_{i+1})$ and $C_H(\sigma_H)$ have same sign, set $\sigma_H = \sigma_{i+1}$. Repeat step 1.

The algorithm is run until $|\widetilde{C} - C_i| \leq \epsilon$, where ϵ the desired tolerance. Implied volatility is $\widehat{\sigma} = \sigma_{i+1}$.

7 Models of implied volatility

7.1 Parametric models of implied volatility

Since it is often very difficult to define a single parametric function for the entire surface (see Chapter 2 in Brockhaus et al. (2000) and Dumas et al. (1998) for suggestions in this directions), a typical approach is to estimate each smile independently by some nonlinear function. The IV surface is then reconstructed by interpolating total variances along forward moneyness as is apparent from section 4. The standard method is linear interpolation. If derivatives of the IV surface with respect to time to expiry are needed, higher order polynomials for interpolation are necessary. Gatheral (2006) suggests the well-behaved cubic interpolation due to Stineman (1980). A monotonic cubic interpolation scheme can be found in Wolberg and Alfy (2002).

In practice a plethora of functional forms is used. The following selection of parametric approaches is driven by their respective popularity in three different asset classes (equity, fixed income, FX markets) and by the solid theoretical underpinnings they are derived from.

Gatheral's SVI parametrization

The stochastic volatility inspired (SVI) parametrization for the smile was introduced by Gatheral (2004) and is motivated from the asymptotic extreme strikes behavior of a IV smile, which is generated by a Heston (1993) model. It is given in terms of log-forward moneyness $m = \log(X/F)$ as

$$\widehat{\sigma}^2(m,T) = a + b\left(\rho(m-c) + \sqrt{(m-c)^2 + \theta^2}\right) , \qquad (24)$$

where $a > 0$ determines the overall level of implied variance and $b \geq 0$ (predominantly) the angle between left and right asymptotes of extreme strikes; $|\rho| \leq 1$ rotates the smile around the vertex, and θ controls the smoothness of the vertex; c translates the graph.

The beauty of Gatheral's parametrization becomes apparent observing that implied variance behaves linear in the extreme left and right wing as prescribed by the moment formula due to Lee (2004), see section 5.3. It is therefore straight forward to control the wings for no-arbitrage conditions. Indeed, comparing the slopes of the left and right wing asymptotes with Theorem 6, we find that

$$b(1 + |\rho|) \leq \frac{2}{T} ,$$

to preclude arbitrage (asymptotically) in the wings. The SVI appears to fit a wide of range smile patterns, both empirical ones and those of many stochastic volatility and pure jump models, Gatheral (2004).

The SABR parametrization

The SABR parametrization is a truncated expansion of the IV smile which is generated by the SABR model proposed by Hagan et al. (2002). SABR is an acronym for the 'stochastic $\alpha\beta\rho$ model', which is a two-factor stochastic volatility model with parameters α, the initial value of the stochastic volatility factor; $\beta \in [0,1]$, an exponent determining the dynamic relationship between the forward and the ATM volatility, where $\beta = 0$ gives rise to a 'stochastic normal' and $\beta = 1$ to a 'stochastic log-normal' behavior; $|\rho| \leq 1$, the correlation between the two Brownian motions; and $\theta > 0$, the volatility of volatility. The SABR approach is very popular in fixed income markets where each asset only has a single exercise date, such as swaption markets.

Denote by F the forward price, X is as usual the strike price. The SABR parametrization is a second order expansion given by

$$\widehat{\sigma}(X,T) = \widehat{\sigma}^0(X)\Big\{1 + \widehat{\sigma}^1(X)\, T\Big\} + \mathcal{O}(T^2) \;, \tag{25}$$

where the first term is

$$\widehat{\sigma}^0(X) = \frac{\theta}{\chi(z)} \log \frac{F}{X} \tag{26}$$

with

$$z = \frac{\theta}{\alpha}\, \frac{F^{1-\beta} - X^{1-\beta}}{1-\beta}$$

and

$$\chi(z) = \log\left(\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right) \;;$$

the second term is

$$\widehat{\sigma}^1(X) = \frac{(1-\beta)^2}{24}\frac{\alpha^2}{(FX)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\theta\alpha}{(FX)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\theta^2 \;. \tag{27}$$

Note that we display here the expansion in the corrected version as was pointed out by Obłój (2008); unlike the original fomula this version behaves consistently for $\beta \to 1$, as then $z(\beta) \to \frac{\theta}{\alpha} \log \frac{F}{X}$.

The formula (25) is involved, but explicit and can therefore be computed efficiently. For the ATM volatility, i.e. $F = X$, z and $\chi(z)$ disappear, and the first term in (25) collapses to $\widehat{\sigma}^0(F) = \alpha F^{\beta-1}$.

As a fitting strategy, it is usually recommended to obtain β from a log-log plot of historical data of the ATM IV $\widehat{\sigma}(F,F)$ against F and to exclude it from the subsequent optimizations. Parameter θ and ρ are inferred from a calibration to observed market IV; during that calibration α is found implicitly by solving for the (smallest) real root of the resulting cubic polynomial in α, given θ and ρ and the ATM IV $\widehat{\sigma}(F,F)$:

$$\alpha^3 \frac{(1-\beta)^2 T}{24\, F^{2-2\beta}} + \alpha^2 \frac{\rho\beta\theta T}{4\, F^{(1-\beta)}} + \alpha \left(1 + \frac{2-3\rho^2}{24}\theta^2 T\right) - \widehat{\sigma}(F,F) F^{1-\beta} = 0\ .$$

For further details on calibrations issues we refer to Hagan et al. (2002) and West (2005), where the latter has a specific focus on the challenges arising in illiquid markets. Alternatively, Mercurio and Pallavicini (2006) suggest a calibration procedure for all parameters (inluding β) from market data exploiting both swaption smiles and constant maturity swap spreads.

Vanna-Volga method

In terms of input information, the vanna-volga (VV) approach is probably the most parsimonious amongst all constructive methods for building an IV surface, as it relies on as few as three input observations per expiry only. It is popular in FX markets. The VV method is based on the idea of constructing a replication portfolio that is locally risk-free up to second order in spot and volatility in a fictitious setting, where the smile is flat, but varies stochastically over time. Clearly, this setting is not only fictitious, but also theoretically inconsistent, as there is no model which generates a flat smile that fluctuates stochastically. It may however be justified by the market practice of using a BSM model with a regularly updated IV as input factor. The hedging costs incurred by the replication portfolio thus constructed are then added to the flat-smile BSM price.

To fix ideas, denote the option vega by $\frac{\partial C}{\partial \sigma}$, volga by $\frac{\partial^2 C}{\partial \sigma^2}$ and vanna by $\frac{\partial^2 C}{\partial \sigma \partial S}$. We are given three market observations of IV $\widehat{\sigma}_i$ with associated strikes X_i, $i = 1,2,3$, with $X_1 < X_2 < X_3$, and same expiry dates $T_i = T$ for which the smile is to be constructed. In a first step, the VV method solves the following system of linear equations, for an arbitrary strike X and for some base volatility $\tilde{\sigma}$:

$$\begin{aligned} \frac{\partial C^{BSM}}{\partial \sigma}(X,\tilde{\sigma}) &= \sum_{i=1}^{3} w_i(X)\, \frac{\partial C^{BSM}}{\partial \sigma}(X_i,\tilde{\sigma}) \\ \frac{\partial^2 C^{BSM}}{\partial \sigma^2}(X,\tilde{\sigma}) &= \sum_{i=1}^{3} w_i(X)\, \frac{\partial^2 C^{BSM}}{\partial \sigma^2}(X_i,\tilde{\sigma}) \\ \frac{\partial^2 C^{BSM}}{\partial \sigma \partial S}(X,\tilde{\sigma}) &= \sum_{i=1}^{3} w_i(X)\, \frac{\partial^2 C^{BSM}}{\partial \sigma \partial S}(X_i,\tilde{\sigma}) \end{aligned} \tag{28}$$

The system can be solved numerically or analytically for the weights $w_i(X)$, $i = 1,2,3$. In a second step, the VV price is computed by

$$C(X) = C^{BSM}(X,\tilde{\sigma}) + \sum_{i=1}^{3} w_i(X) \left[C^{BSM}(X_i,\widehat{\sigma}_i) - C^{BSM}(X_i,\tilde{\sigma}) \right]\ , \tag{29}$$

from which one obtains IV by inverting the BSM formula. These steps need to be solved for each X to construct the VV smile. For more details on the VV method, approximation formulae for the VV smile, and numerous practical insights we refer to the lucid description presented by Castagna and Mercurio (2007). As a typical choice for the base volatility, Castagna and Mercurio (2007) suggest $\tilde{\sigma} = \sigma_2$, where σ_2 would be an ATM IV, and σ_1 and σ_3 are 25Δ put and the 25Δ call IV, respectively. As noted there, the VV method is not arbitrage-free by construction, in particular convexity can not be guaranteed, but it appears to produce arbitage-free estimates of IV surfaces for usual market conditions.

7.2 Non- and semiparametric models of implied volatility

If potential arbitrage violations in the resulting estimate are of no particular concern, virtually any non- and semiparametric method can be applied to IV data. A specific choice can often be made from practical considerations. We therefore confine this section to pointing to the relevant examples in the literature.

Piecewise quadratic or cubic polynomials to fit single smiles was applied by Shimko (1993), Malz (1997), Ané and Geman (1999) and Hafner and Wallmeier (2001). Aït-Sahalia and Lo (1998), Rosenberg (2000), Cont and da Fonseca (2002) and Fengler et al. (2003) employ a Nadaraya-Watson smoother. Higher order local polynomial smoothing of the IV surface was suggested in Fengler (2005), when the aim is to recover the local volatility function via the Dupire formula, or by Härdle et al. (2010) for estimating the empirical pricing kernel. Least-squares kernel regression was suggested in Gouriéroux et al. (1994) and Fengler and Wang (2009). Audrino and Colangelo (2009) rely on IV surface estimates based on regression trees in a forecasting study. Model selection between fully parametric, semi- and nonparametric specifications is discussed in detail in Aït-Sahalia et al. (2001).

7.3 Implied volatility modeling under no-arbitrage constraints

For certain applications, for instance for local volatility modeling, an arbitrage-free estimate of the IV surface is mandatory. Methods producing arbitrage-free estimates must respect the bounds presented in section 4. They are surveyed in this section.

Call price interpolation

Interpolation techniques to recover a globally arbitrage-free call price function have been suggested by Kahalé (2004) and Wang et al. (2004). It is crucial for these algorithms to work that the data to be interpolated are arbitrage-free from the beginning. Consider the set of pairs of strikes and call prices

$(X_i, C_i), i = 0, \ldots, n$. Then, applying to (9), the set does not admit arbitrage in strikes if the first divided differences associated with the data observe

$$-e^{-rT} < \frac{C_i - C_{i-1}}{X_i - X_{i-1}} < \frac{C_{i+1} - C_i}{X_{i+1} - X_i} < 0 \tag{30}$$

and if the price bounds (8) hold.

For interpolation Kahalé (2004) considers piecewise convex polynomials which are inspired from the BSM formula. More precisely, for a parameter vector $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top$ with $\theta_1 > 0$, $\theta_2 > 0$ consider the function

$$c(X; \Theta) = \theta_1 \Phi(d_1) - X \Phi(d_2) + \theta_3 X + \theta_4 \ , \tag{31}$$

where $d_1 = \left[\log(\theta_1/X) + 0.5\,\theta_2^2\right]/\theta_2$ and $d_2 = d_1 - \theta_2$. Clearly, $c(X;\Theta)$ is convex in strikes $X > 0$, since it differs from the BSM formula by a linear term, only. It can be shown that on a segment $[X_i, X_{i+1}]$ and for C_i, C_{i+1} and given first order deratives C'_i and C'_{i+1} there exists a unique vector Θ interpolating the observed call prices.

Kahalé (2004) proceeds in showing that for a sequence (X_i, C_i, C'_i) for $i = 0, \ldots, n+1$ with the (limit) conditions $X = 0$, $X_i < X_{i+1}$, $X_{n+1} = \infty$, $C_0 = S_0$, $C_{n+1} = 0$, $C'_0 = -e^{-rT}$ and $C'_{n+1} = 0$ and

$$C'_i < \frac{C_{i+1} - C_i}{X_{i+1} - X_i} < C'_{i+1} \tag{32}$$

for $i = 1, \ldots, n$ there exists a unique $\mathcal{C}^1$ convex function $c(X)$ described by a series of vectors Θ_i for $i = 0, \ldots, n$ interpolating observed call prices. There are $4(n+1)$ parameters in Θ_i, which are matched by $4n$ equations in the interior segments $C_i = c(X_i; \Theta_i)$ and $C'_i = c'(X_i; \Theta_i)$ for $i = 1, \ldots, n$, and four additional equations by the four limit conditions in (X_0, C_0) and (X_{n+1}, C_{n+1}).

A $\mathcal{C}^2$ convex function is obtained in the following way: For $j = 1, \ldots, n$, replace the jth condition on the first order conditions by $\gamma_j = c'(X_j; \Theta_j)$ and $\gamma_j = c'(X_j; \Theta_{j-1})$, for some $\gamma_j \in]l_j, l_{j+1}[$ and $l_j = (C_j - C_{j-1})/(X_j - X_{j-1})$. Moreover add the condition $c''(X_j; \Theta_j) = c''(X_j; \Theta_{j-1})$. This way the number of parameters is still equal to the number of constraints.

Concluding, the Kahalé (2004) algorithm for a $\mathcal{C}^2$ call price function is as follows:

1. Put $C'_0 = -e^{-rT}$, $C'_{n+1} = 0$ and $C'_i = (l_i + l_{i+1})/2$ for $i = 1, \ldots, n$, where $l_i = (C_i - C_{i-1})/(X_i - X_{i-1})$.
2. For each $j = 1, \ldots, n$ compute the $\mathcal{C}^1$ convex function with continuous second order derivative at X_j. Replace $C'_j = \gamma_j$.

Kahalé (2004) suggests to solve the algorithm using the Newton-Raphson method.

An alternative, cubic B-spline interpolation was suggested by Wang et al. (2004). For observed prices $(X_i, C_i), i = 0, \ldots, n$, $0 < a = X_0 < \ldots < X_n = b < \infty$ they consider the following minimization problem:

$$\begin{aligned}
\min \quad & ||c''(X) - e^{-rT}h(X)||_2^2 \\
\text{s.t.} \quad & c(X_i) = C_i, \qquad i = 0, \ldots, n \;, \\
& c''(X) \geq 0 \qquad X \in (0, \infty) \;,
\end{aligned} \tag{33}$$

where $||\cdot||_2$ is the (Lebesgue) L^2 norm on $[a, b]$, h some prior density (e.g., the log-normal density) and c the unknown option price function with absolutely continuous first and second order derivatives on $[a, b]$. By the Peano kernel theorem, the constraints $c(X_i) = C_i$, $i = 1, \ldots, n$ can be replaced by

$$\int_a^b B_i(X)\, c''(X)\, dX = d_i, \qquad i = 1, \ldots, n-2 \;, \tag{34}$$

where B_i is a normalized linear B-spline with the support on $[X_i, X_{i+2}]$ and d_i the second divided differences associated with the data. Wang et al. (2004) show that this infinite-dimensional optimization problem has a unique solution for $c''(X)$ and how to cast it into a finite-dimensional smooth optimization problem. The resulting function for $c(X)$ is then a cubic B-spline. Finally they devise a generalized Newton method for solving the problem with superlinear convergence.

Call price smoothing by natural cubic splines

For a sample of strikes and call prices, $\{(X_i, C_i)\}$, $X_i \in [a, b]$ for $i = 1, \ldots, n$, Fengler (2009) considers the curve estimate defined as minimizer $\widehat{g}$ of the penalized sum of squares

$$\sum_{i=1}^{n} \Big\{ C_i - g(X_i) \Big\}^2 + \lambda \int_a^b \{g''(v)\}^2\, dv \;. \tag{35}$$

The minimizer $\widehat{g}$ is a natural cubic spline, and represents a globally arbitrage-free call price function. Smoothness is controlled by the parameter $\lambda > 0$. The algorithm suggested by Fengler (2009) observes the no-arbitrage constraints (8), (9), and (10). For this purpose the natural cubic spline is converted into the value-second derivative representation suggested by Green and Silverman (1994). This allows to formulate a quadratic program solving (35). Put $g_i = g(u_i)$ and $\gamma_i = g''(u_i)$, for $i = 1, \ldots, n$, and define $g = (g_1, \ldots, g_n)^\top$ and $\gamma = (\gamma_2, \ldots, \gamma_{n-1})^\top$. By definition of a natural cubic spline, $\gamma_1 = \gamma_n = 0$. The natural spline is completely specified by the vectors g and γ, see Section 2.5 in Green and Silverman (1994) who also suggest the nonstandard notation of the entries in γ.

Sufficient and necessary conditions for g and γ to represent a valid cubic spline are formulated via the matrices Q and R. Let $h_i = u_{i+1} - u_i$ for $i = 1, \ldots, n-1$, and define the $n \times (n-2)$ matrix Q by its elements $q_{i,j}$, for $i = 1, \ldots, n$ and $j = 2, \ldots, n-1$, given by

$$q_{j-1,j} = h_{j-1}^{-1} \;, \qquad q_{j,j} = -h_{j-1}^{-1} - h_j^{-1} \;, \qquad \text{and} \;\; q_{j+1,j} = h_j^{-1} \;,$$

for $j = 2, \ldots, n-1$, and $q_{i,j} = 0$ for $|i-j| \geq 2$, where the columns of Q are numbered in the same non-standard way as the vector γ.

The $(n-2) \times (n-2)$ matrix R is symmetric and defined by its elements $r_{i,j}$ for $i, j = 2, \ldots, n-1$, given by

$$\begin{aligned} r_{i,i} &= \tfrac{1}{3}(h_{i-1} + h_i) \text{ for } i = 2, \ldots, n-1 \\ r_{i,i+1} = r_{i+1,i} &= \tfrac{1}{6} h_i \quad \text{for } i = 2, \ldots, n-2 \, , \end{aligned} \tag{36}$$

and $r_{i,j} = 0$ for $|i-j| \geq 2$. R is strictly diagonal dominant, and thus strictly positive-definite.

Arbitrage-free smoothing of the call price surface can be cast into the following iterative quadratic minimization problem. Define a $(2n-2)$-vector $y = (y_1, \ldots, y_n, 0, \ldots, 0)^\top$, a $(2n-2)$-vector $\xi = (g^\top, \gamma^\top)^\top$ and the matrices, $A = \left(Q, -R^\top\right)$ and

$$B = \begin{pmatrix} I_n & 0 \\ 0 & \lambda R \end{pmatrix} , \tag{37}$$

where I_n is the unit matrix with size n. Then

1. Estimate the IV surface by means of an initial estimate on a regular forward-moneyness grid $\mathcal{J} = [x_1, x_n] \times [T_1, T_m]$.
2. Iterate through the price surface from the last to the first expiry, and solve the following quadratic programs.
   For T_j, $j = m, \ldots, 1$, solve

$$\min_{\xi} \; -y^\top \xi + \frac{1}{2} \xi^\top B \xi \tag{38}$$

   subject to

$$\begin{aligned} A^\top \xi &= 0 \\ \gamma_i &\geq 0 \\ \tfrac{g_2 - g_1}{h_1} - \tfrac{h_1}{6}\gamma_2 &\geq -e^{-rT_j} \\ -\tfrac{g_n - g_{n-1}}{h_{n-1}} - \tfrac{h_{n-1}}{6}\gamma_{n-1} &\geq 0 \\ g_1 &\leq S_t & \text{if} \quad & j = m \\ g_i^{(j)} &< g_i^{(j+1)} & \text{if} \quad & j \in [m-1, 1] \\ & & \text{for} \quad & i = 1, \ldots, n \quad (*) \\ g_1 &\geq S_t - e^{-rT_j} u_1 \\ g_n &\geq 0 \end{aligned} \tag{39}$$

   where $\xi = (g^\top, \gamma^\top)^\top$. Note that we suppress the explicit dependence on j except in conditions $(*)$ to keep the notation more readable. Conditions $(*)$ implement (10); therefore $g_i^{(j)}$ and $g_i^{(j+1)}$ are related by forward-moneyness.

The resulting price surface is converted into IV. It can be beneficial obtain a first coarse estimate of the surface by gridding it on the estimation grid. This allows to more easily implement condition (10). The minimization problem can be solved by using the quadratic programming devices provided by standard statistical software packages. The reader is referred to Fengler (2009) for the computational details and the choice of the smoothing parameter λ. In contrast to the approach by Kahalé (2004), a potential drawback this approach suffers from 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 however by the choice of a sufficiently dense grid in the strike dimension in $\mathcal{J}$.

IV smoothing using local polynomials

As an alternative to smoothing in the call price domain Benko et al. (2007) suggest to directly smooth IV by means of constrained local quadratic polynomials. This implies minimization of the following (local) least squares criterion

$$\min_{\alpha_0,\alpha_1,\alpha_2} \sum_{i=1}^{n} \left\{ \widetilde{\sigma}_i - \alpha_0 - \alpha_1(x_i - x) - \alpha_2(x_i - x)^2 \right\}^2 \mathcal{K}_h(x - x_i) \ , \tag{40}$$

where $\widetilde{\sigma}$ is observed IV. We denote by $\mathcal{K}_h(x - x_i) = h^{-1}\mathcal{K}\left(\frac{x-x_i}{h}\right)$ and by $\mathcal{K}$ a kernel function – typically a symmetric density function with compact support, e.g. $\mathcal{K}(u) = \frac{3}{4}(1 - u^2)\mathbf{1}(|u| \le 1)$, the Epanechnikov kernel, where $\mathbf{1}(\mathcal{A})$ is the indicator function of some set $\mathcal{A}$. Finally, h is the bandwidth which governs the trade-off between bias and variance, see Härdle (1990) for the details on nonparametric regression. Since $\mathcal{K}_h$ is nonnegative within the (localization) window $[x-h, x+h]$, points outside of this interval do not have any influence on the estimator $\widehat{\sigma}(x)$.

No-arbitrage conditions in terms of IV are obtained by computing (9) for an IV adjusted BSM formula, see Brunner and Hafner (2003) among others. Expressed in forward moneyness $x = X/F$ this yields for the convexity condition

$$\begin{aligned}\frac{\partial^2 C^{BSM}}{\partial x^2} &= e^{-rT}\sqrt{T}\varphi(d_1) \\ &\times \left\{ \frac{1}{x^2\widehat{\sigma}T} + \frac{2d_1}{x\widehat{\sigma}\sqrt{T}}\frac{\partial\widehat{\sigma}}{\partial x} + \frac{d_1 d_2}{\widehat{\sigma}}\left(\frac{\partial\widehat{\sigma}}{\partial x}\right)^2 + \frac{\partial^2\widehat{\sigma}}{\partial x^2} \right\}\end{aligned} \tag{41}$$

where d_1 and d_2 are defined as in (4) and (5).

The key property of local polynomial regression is that it yields simultaneously to the regression function its derivatives. More precisely, comparing (40) with the Taylor expansion of $\widehat{\sigma}$ shows that

$$\widehat{\sigma}(x_i) = \alpha_0,\ \widehat{\sigma}'(x_i) = \alpha_1,\ \widehat{\sigma}''(x_i) = 2\alpha_2 \ . \tag{42}$$

Based on this fact Benko et al. (2007) suggest to miminize (40) subject to

$$e^{-rT}\sqrt{T}\varphi(d_1)\left\{\frac{1}{x^2\alpha_0 T}+\frac{2d_1\alpha_1}{x\alpha_0\sqrt{T}}+\frac{d_1 d_2}{\alpha_0}\left(\alpha_1\right)^2+2\alpha_2\right\}\geq 0\;, \tag{43}$$

with

$$d_1=\frac{\alpha_0^2 T/2-\log(x)}{\sigma\sqrt{T}},\;\; d_2=d_1-\alpha_0\sqrt{T}\;.$$

This leads to a nonlinear optimization problem in $\alpha_0, \alpha_1, \alpha_2$.

The case of the entire IV surface is more involved. Suppose the purpose is to estimate $\widehat{\sigma}(x,T)$ for a set of maturities $\{T_1,\ldots,T_L\}$. By (11), for a given value x, we need to ensure $\widehat{\nu}^2(x,T_l,)\leq\widehat{\nu}^2(x,T_{l'})$, for all $T_l<T_l'$. Denote by $\mathcal{K}_{h_x,h_T}(x-x_i,T_l-T_i)$ a bivariate kernel function given by the product of the two univariate kernel functions $\mathcal{K}_{h_x}(x-x_i)$ and $\mathcal{K}_{h_T}(T-T_i)$. Extending (40) linearly into the time-to-maturiy dimension then leads to the following optimization problem:

$$\begin{aligned}\min_{\alpha(l)}\quad &\sum_{l=1}^{L}\sum_{i=1}^{n}\mathcal{K}_{h_x,h_T}(x-x_i,T_l-T_i)\Big\{\widetilde{\sigma}_i-\alpha_0(l)\\ &-\alpha_1(l)(x_i-x)-\alpha_2(l)(T_i-T)-\alpha_{1,1}(l)(x_i-x)^2\\ &\qquad\qquad -\alpha_{1,2}(l)(x_i-x)(T_i-T)\Big\}^2\end{aligned} \tag{44}$$

subject to

$$\begin{aligned}&\sqrt{T_l}\varphi(d_1(l))\left\{\frac{1}{x^2\alpha_0(l)T_l}+\frac{2d_1(l)\alpha_1(l)}{x\alpha_0(l)\sqrt{T_l}}+\frac{d_1(l)d_2(l)}{a_0(l)}\alpha_1^2(l)+2\alpha_{1,1}(l)\right\}\;\geq\;0\;,\\ &d_1(l)=\frac{\alpha_0^2(l)T_l/2-\log(x)}{\alpha_0(l)\sqrt{T_l}},\qquad d_2(l)=d_1(l)-a_0(l)\sqrt{T_l},\qquad l=1,\ldots,L\\ &\qquad\qquad 2T_l\alpha_0(l)\alpha_2(l)+\alpha_0^2(l)>0\qquad l=1,\ldots,L\\ &\qquad\qquad \alpha_0^2(l)T_l<\alpha_0^2(l')T_l',\qquad T_l\;<\;T_l'\;.\end{aligned}$$

The last two conditions ensure that total implied variance is (locally) nondecreasing, since $\frac{\partial\nu^2}{\partial T}>0$ can be rewritten as $2T\alpha_0\alpha_2+\alpha_0^2>0$ for a given T, while the last conditions guarantee that total variance is increasing across the surface. From a computational view, problem (44) calculates for a given x the estimates for all given T_l in one step in order to warrant that $\widehat{\nu}$ is increasing in T.

The approach by Benko et al. (2007) yields an IV 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 IV data occurring the surfaces will be acceptable in most cases.

References

Aït-Sahalia, Y., Bickel, P. J. and Stoker, T. M. (2001). Goodness-of-fit tests for regression using kernel methods, *Journal of Econometrics* **105**: 363–412.

Aït-Sahalia, Y. and Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance* **53**: 499–548.

Ané, T. and Geman, H. (1999). Stochastic volatility and transaction time: An activity-based volatility estimator, *Journal of Risk* **2**(1): 57–69.

Audrino, F. and Colangelo, D. (2009). Semi-parametric forecasts of the implied volatility surface using regression trees, *Statistics and Computing* . Forthcoming.

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

Bharadia, M. A., Christofides, N. and Salkin, G. R. (1996). Computing the Black-Scholes implied volatility - generalization of a simple formula, *in* P. P. Boyle, F. A. Longstaff, P. Ritchken, D. M. Chance and R. R. Trippi (eds), *Advances in Futures and Options Research*, Vol. 8, JAI Press, London, pp. 15–29.

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

Brenner, M. and Subrahmanyam, M. (1988). A simple formula to compute the implied standard deviation, *Financial Analysts Journal* **44**(5): 80–83.

Britten-Jones, M. and Neuberger, A. J. (2000). Option prices, implied price processes, and stochastic volatility, *Journal of Finance* **55**(2): 839–866.

Brockhaus, O., Farkas, M., Ferraris, A., Long, D. and Overhaus, M. (2000). *Equity derivatives and market risk models*, Risk Books, London.

Brunner, B. and Hafner, R. (2003). Arbitrage-free estimation of the risk-neutral density from the implied volatility smile, *Journal of Computational Finance* **7**(1): 75–106.

Carr, P. and Wu, L. (2003). Finite moment log stable process and option pricing, *Journal of Finance* **58**(2): 753–777.

Castagna, A. and Mercurio, F. (2007). Building implied volatility surfaces from the available market quotes: a unified approach, *in* I. Nelken (ed.), *Volatility as an asset class*, Risk Books, Haymarket, London, pp. 3–59.

Cont, R. and da Fonseca, J. (2002). The dynamics of implied volatility surfaces, *Quantitative Finance* **2**(1): 45–60.

Corrado, C. J. and Miller, T. W. (1996). A note on a simple, accurate formula to compute implied standard deviations, *Journal of Banking and Finance* **20**: 595–603.

Dumas, B., Fleming, J. and Whaley, R. E. (1998). Implied volatility functions: Empirical tests, *Journal of Finance* **53**(6): 2059–2106.

Feinstein, S. (1988). A source of unbiased implied volatility, *Technical Report 88-9*, Federal Reserve Bank of Atlanta.

Fengler, M. R. (2005). *Semiparametric Modeling of Implied Volatility*, Lecture Notes in Finance, Springer-Verlag, Berlin, Heidelberg.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface, *Quantitative Finance* **9**(4): 417–428.

Fengler, M. R., Härdle, W. and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics, *Journal of Financial Econometrics* **5**(2): 189–218.

Fengler, M. R., Härdle, W. and Villa, C. (2003). The dynamics of implied volatilities: A common principle components approach, *Review of Derivatives Research* **6**: 179–202.

Fengler, M. R. and Wang, Q. (2009). Least squares kernel smoothing of the implied volatility smile, *in* W. Härdle, N. Hautsch and L. Overbeck (eds), *Applied Quantitative Finance*, 2nd edn, Springer-Verlag, Berlin, Heidelberg.

Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives, Presentation at the ICBI Global Derivatives and Risk Management, Madrid, España.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*, John Wiley & Sons, Hoboken, New Jersey.

Gouriéroux, C., Monfort, A. and Tenreiro, C. (1994). Nonparametric diagnostics for structural models, *Document de travail 9405*, CREST, Paris.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, Vol. 58 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

Hafner, R. and Wallmeier, M. (2001). The dynamics of DAX implied volatilities, *International Quarterly Journal of Finance* **1**(1): 1–27.

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

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.

Härdle, W., Okhrin, O. and Wang, W. (2010). Uniform confidence bands for pricing kernels, *SFB 649 Discussion Paper 2010-03*, Humboldt-Universität zu Berlin, Berlin.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**: 327–343.

Hodges, H. M. (1996). Arbitrage bounds on the implied volatility strike and term structures of European-style options, *Journal of Derivatives* **3**: 23–35.

Isengildina-Massa, O., Curtis, C., Bridges, W. and Nian, M. (2007). Accuracy of implied volatility approximations using "nearest-to-the-money" option premiums, *Technical report*, Southern Agricultural Economics Association.

Kahalé, N. (2004). An arbitrage-free interpolation of volatilities, *RISK* **17**(5): 102–106.

Latané, H. A. and Rendelman, J. (1976). Standard deviations of stock price ratios implied in option prices, *Journal of Finance* **31**: 369–381.

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

Li, S. (2005). A new formula for computing implied volatility, *Applied Mathematics and Computation* **170**(1): 611–625.

Malz, A. M. (1997). Estimating the probability distribution of the future exchange rate from option prices, *Journal of Derivatives* **5**(2): 18–36.

Manaster, S. and Koehler, G. (1982). The calculation of implied variances from the Black-and-Scholes model: A note, *Journal of Finance* **37**: 227–230.

Mercurio, F. and Pallavicini, A. (2006). Smiling at convexity: bridging swaption skews and CMS adjustments, *RISK* **19**(8): 64–69.

Merton, R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**(Spring): 141–183.

Obłój, J. (2008). Fine-tune your smile: Correction to hagan et al., *Wilmott Magazine* **May**.

Randall, C. and Tavella, D. (2000). *Pricing Financial Instruments: The Finite Difference Method*, John Wiley & Sons, New York.

Rebonato, R. (2004). *Volatility and Correlation*, Wiley Series in Financial Engineering, 2nd. edn, John Wiley & Son Ltd.

Reiner, E. (2000). Calendar spreads, characteristic functions, and variance interpolation. Mimeo.

Reiner, E. (2004). The characteristic curve approach to arbitrage-free time interpolation of volatility, presentation at the ICBI Global Derivatives and Risk Management, Madrid, España.

Rogers, L. C. G. and Tehranchi, M. (2009). Can the implied volatility surface move by parallel shifts?, *Finance and Stochastics* . Forthcoming.

Roper, M. and Rutkowski, M. (2009). On the relationship between the call price surface and the implied volatility surface close to expiry, *International Journal of Theoretical and Applied Finance* **12**(4): 427–441.

Rosenberg, J. (2000). Implied volatility functions: A reprise, *Journal of Derivatives* **7**: 51–64.

Rubinstein, M. (1994). Implied binomial trees, *Journal of Finance* **49**: 771–818.

Shimko, D. (1993). Bounds on probability, *RISK* **6**(4): 33–37.

Steele, J. M. (2000). *Stochastic Calculus and Financial Applications*, Springer-Verlag, Berlin, Heidelberg, New York.

Stineman, R. W. (1980). A consistently well-behaved method of interpolation, *Creative Computing* **6**(7): 54–57.

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

Tehranchi, M. (2010). Implied volatility: long maturity behavior, *in* R. Cont (ed.), *Encyclopedia of Quantitative Finance*, John Wiley & Sons. Forthcoming.

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

West, G. (2005). Calibration of the SABR model in illiquid markets, *Applied Mathematical Finance* **12**(4): 371–385.

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