Shapes of implied volatility with positive mass at zero*

S. De Marco[1], C. Hillairet[1], and A. Jacquier[2]

[1]CMAP, Ecole Polytechnique
[2]Imperial College London

October 4, 2013

Abstract

We study the shapes of the implied volatility when the underlying distribution has an atom at zero. We show that the behaviour at small strikes is uniquely determined by the mass of the atom at least up to the third asymptotic order, regardless of the properties of the remaining (absolutely continuous, or singular) distribution on the positive real line. We investigate the structural difference with the no-mass-at-zero case, showing how one can—a priori—distinguish between mass at the origin and a heavy-left-tailed distribution. An atom at zero is found in stochastic models with absorption at the boundary, such as the CEV process, and can be used to model default events, as in the class of jump-to-default structural models of credit risk. We numerically test our model-free result in such examples. Note that while Lee's moment formula [21] tells that implied variance is *at most* asymptotically linear in log-strike, other celebrated results for exact smile asymptotics such as [2, 14] do not apply in this setting—essentially due to the breakdown of Put-Call symmetry—and we rely here on an alternative treatment of the problem.

1 Introduction

Stochastic models are used extensively to price options and calibrate market data. In practice, such data is often quoted, not in terms of option prices, but in terms of implied volatilities. However, apart from the Black-Scholes model where the implied volatility is constant, no closed-form formula is available for most models (with or without continuous paths). Over the past decade or so, many authors have worked out approximations of this implied volatility, either in a model-free setting or for some specific models; these approximations are usually only valid in restricted regions, such as small and large maturities or extreme strikes. The latter have proved to be useful in order to extrapolate observed (and calibrated) data in an arbitrage-free way. The celebrated Lee's formula [21] was a ground-breaking model-independent result in this direction; subsequent advances were made in Benaim and Friz [2] and recently in Gulisashvili [14]. Let us quote Gulisashvili's result: let S be a non-negative martingale under a given risk-neutral measure $\mathbb{P}$ and let $P(K,T) = \mathbb{E}(K - S_T)^+$ ($C(K,T) = \mathbb{E}(S_T - K)^+$) denote the price of a Put (Call) option written on S, with strike K and maturity T. The behaviour of the implied volatility $I_T(K)$ at small strikes is related to this Put price by the asymptotic formula [14, Corollary 5.12]

$$I_T(K) = \sqrt{\frac{|\log K|}{T}}\sqrt{\psi\Big(\frac{\log P(K,T)}{\log K} - 1\Big)} + \mathcal{O}\left(\left(\log\frac{K}{P(K,T)}\right)^{-1/2}\log\log\frac{K}{P(K,T)}\right), \qquad \text{as } K \downarrow 0, \quad (1.1)$$

*We thank Valdo Durrlemann, Archil Gulisashvili, Pierre Henry-Labordère, Aleksandar Mijatović and Mike Tehranchi for stimulating discussions. SDM and CH are thankful to Nizar Touzi for his interest during the first phase of this work. SDM and AJ acknowledge funding from the Imperial College Workshop Support Grant and the London Mathematical Society for the 'Workshop on Large deviations and asymptotic methods in finance' (April 2013), during which part of this work was done. SDM and CH acknowledge funding from the research programs 'Chaire Risques Financiers' of Fondation du Risque, 'Chaire Marchés en mutation' of the Fédération Bancaire Française and 'Chaire Finance et développement durable' of EDF and Calyon. Corresponding author `demarco@cmap.polytechnique.fr`

Key words and phrases: Atomic distribution, heavy-tailed distribution, Implied Volatility, smile asymptotics, absorption at zero, CEV model.

2010 Mathematics Subject Classification: AMS 91G20, 65C50.

where the continuous function $\psi : [0,\infty] \to [0,2]$ is defined by

$$\psi(z) \equiv 2 - 4\left(\sqrt{z(z+1)} - z\right), \qquad \psi(\infty) = 0. \tag{1.2}$$

A similar formula, expressed in terms of the call price $C(K,T)$, holds as K tends to infinity. Eq. (1.1) is valid for every Put price function P such that $P(K,T) > 0$ for all $K > 0$ (in the notation of [14, Definition 1.5], the corresponding call price C belongs to the class PF_0), which is equivalent to $\mathbb{P}(S_T < K) > 0$ for all $K > 0$.[1] As shown in [14],Eq. (1.1) (together with its counterpart for $K \uparrow \infty$) allows to recover the moment formula of Lee [21] and the tail-wing formula of Benaim and Friz [2]. The derivation of (1.1) is done in two steps: first an asymptotic expansion for $I_T(K)$ as K tends to infinity is computed in terms of the call price function; then the expression for $K \downarrow 0$ is obtained via the Put-Call symmetry

$$\frac{P(KS_0,T)}{S_0} = \mathbb{E}\left(K - \frac{S_T}{S_0}\right)^+ = K\mathbb{E}_{\mathbb{Q}}\left(\frac{S_0}{S_T} - \frac{1}{K}\right)^+,$$

where $\mathbb{Q}$ is a probability measure absolutely continuous with respect to $\mathbb{P}$, defined through its Radon-Nikodym density $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} := S_T/S_0$. The Put-Call symmetry above holds if (as implicitly assumed in [14]) the law of the underlying asset price does not charge zero under $\mathbb{P}$, $\mathbb{P}(S_T = 0) = 0$. The expansion (1.1), then, is a priori not justified when $\mathbb{P}(S_T = 0) > 0$.

In certain stochastic models, an asset price is modeled with a stochastic process that accumulates mass at zero in finite time: this is the case for the Constant Elasticity of Variance (CEV) local volatility diffusions, whose fixed-time marginal distribution has a continuous part and an atom at zero under certain parameters configurations (the same phenomenon appearing for Sabr, the stochastic volatility counterpart of CEV). Also, in the past few years, the financial crisis has emphasised that potential default (namely the asset price falling to low levels) needs to be seriously taken into account by market participants. In the class of structural models, default is defined as the first time the firm's value hits a given threshold: in Collin-Dufresne et al. [8] and Coculescu et al. [7], the firm's value corresponds to the so-called solvency ratio (logarithm of assets over debt), and is modeled via an Ornstein-Uhlenbeck process. Within this modeling framework, obtaining reliable data to calibrate the model (asset volatility forecasts, capital structure leverage) is often difficult. An alternative approach, proposed by Campi et al. [4], is to refer to the underlying equity process and define the default as the first time the process hits the origin: while the equity value remains deeply related to the firm's asset and debt balance sheet, such a modeling choice avoids the application mishaps of structural models. In the model proposed in [4], the equity process hits the origin either after a jump or in a diffusive way, the continuous-path part of the equity value being modeled as a CEV diffusion with a positive probability of absorption at zero. Along the same line, we will consider in this paper asset prices that may either jump to zero, or hit zero in a diffusive way.

Note that $\mathbb{P}(S_T = 0) > 0$ implies $q^* = 0$, where q^* denotes the critical exponent of S_T, $q^* := \sup\{q \geq 0 : \mathbb{E}[S_T^{-q}] < \infty\}$. Then, Lee's moment formula for small strike yields, in full generality

$$\limsup_{K\downarrow 0} \frac{I_T(K)}{\sqrt{|\log K|}} = \sqrt{\frac{\psi(q^*)}{T}} = \sqrt{\frac{2}{T}}. \tag{1.3}$$

Tail-wing type refinements aim at finding conditions under which this lim sup can be strengthened into a genuine limit, yielding the asymptotics $I_T(K) \sim \sqrt{2|\log K|/T}$ as $K \downarrow 0$: Benaim and Friz's result [2] gives some sufficient conditions, but is limited to the case $q^* > 0$; Gulisashvili's result (1.1), however, does apply to the case $q^* = 0$ and $\mathbb{P}(S_T = 0) = 0$, and allows to formulate necessary and sufficient conditions, as done in [15] (see Theorem 2.2, where we recall the result of [15]). When the $\mathbb{P}(S_T = 0) > 0$, the asymptotic equivalence $I_T(K) \sim \sqrt{2|\log K|/T}$ as $K \downarrow 0$ always holds (this statement is 'almost' given in Lee's original paper [21]; see Proposition 2.4 and Remark 2.7 below for more details). More notably, we show that in the presence of a mass at zero and under a mild assumption on the behaviour of the cumulative distribution function $F(K) := \mathbb{P}(S_T \leq K)$ on a right neighbourhood of zero, namely there exists $\varepsilon > 0$ such that

$$F(K) - F(0) = \mathcal{O}(K^{\varepsilon}) \qquad \text{as } K \downarrow 0, \tag{1.4}$$

[1] If $\mathbb{P}(S_T < \overline{K}) = 0$ for some $0 < \overline{K} < S_0$, then $P(K,T) = 0$ for all $K \leq \overline{K}$. According to the definition of the implied volatility in [14], $I_T(K)$ is not defined for such strikes; according to our extended definition, see (1.8) below, $I_T(K)$ is identically zero for all $K \leq \overline{K}$.

the implied volatility has the form (Theorem 3.7)

$$I_T(K) = \sqrt{\frac{2|\log K|}{T}} + \frac{1}{\sqrt{T}} N^{-1}\big(\mathbb{P}(S_T=0)\big) + \frac{\left[N^{-1}\big(\mathbb{P}(S_T=0)\big)\right]^2}{2\sqrt{2T|\log K|}} + \Phi(K) \qquad \text{as } K \downarrow 0, \tag{1.5}$$

where N^{-1} is the inverse of the Gaussian cdf, and Φ satisfies $\limsup_{K\downarrow 0} \sqrt{2T|\log(K)|}\,|\Phi(K)| \leq 1$. Note that the asymptotic formula (1.5) contains an explicit third-order term, and an error term given by the function Φ, both contributing to a global $\mathcal{O}(|\log(K)|^{-1/2})$ estimate; the interplay between these two terms depends on the value of the mass at zero—we refer the reader to Remark 1 after Theorem 3.7 for more details. In order to measure the impact of Assumption (1.4), note that if the law of the stock price admits a density f in a neighbourhood of zero, which is such that $f(K) = \mathcal{O}(K^{-a})$ as $K \downarrow 0$ for some $a < 1$, then $F(K) - F(0) = \mathcal{O}(K^{1-a})$, and Condition (1.4) is fulfilled.

In Section 2, we first show that the $\mathcal{O}$ term in (1.1) is a constant (which we compute explicitly) when the stock price has a positive mass at the origin. The comparison with the first two terms of (1.5) therefore shows that the asymptotic formula (1.1) is actually still valid in this setting. We refine this result by providing the third-order term and the error estimate appearing in (1.5) in Section 3, and test the resulting formula on several examples in Section 4. Finally, in Section 5 we study the implication of the presence of a mass at zero on the Put-Call symmetry and on volatility derivatives.

Notations: We assume that for any $T > 0$, S_T is a non-negative integrable random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}[S_T] = S_0 > 0$. Risk-free interest rates are considered null, and option prices are given by expectations under the measure $\mathbb{P}$. $C(K,T) := \mathbb{E}(S_T - K)^+$ and $P(K,T) := \mathbb{E}(K - S_T)^+$ represent respectively the price of European Call and Put options with strike $K \geq 0$ and maturity $T \geq 0$, and $C_{\mathrm{BS}}(K,T;S_0,\sigma)$ and $P_{\mathrm{BS}}(K,T;S_0,\sigma)$ the corresponding Call and Put prices in the Black-Scholes model with volatility parameter σ:

$$C_{\mathrm{BS}}(K,T;S_0,\sigma) := \begin{cases} S_0 N(d_1(K,T,S_0,\sigma)) - KN(d_2(K,T,S_0,\sigma)), & \text{if } \sigma > 0, \\ (S_0 - K)^+, & \text{if } \sigma = 0, \end{cases} \tag{1.6}$$

$$P_{\mathrm{BS}}(K,T;S_0,\sigma) := \begin{cases} KN(-d_2(K,T,S_0,\sigma)) - S_0 N(-d_1(K,T,S_0,\sigma)), & \text{if } \sigma > 0, \\ (K - S_0)^+, & \text{if } \sigma = 0, \end{cases} \tag{1.7}$$

where $d_{1,2}(K,T,S_0,\sigma) := \log(S_0/K)/(\sigma\sqrt{T}) \pm \frac{1}{2}\sigma\sqrt{T}$, and N is the standard normal cumulative distribution: $N(d) = \int_{-\infty}^{d} \phi(z)\mathrm{d}z$, $\phi(z) = \mathrm{e}^{-z^2/2}/\sqrt{2\pi}$. When the spot price S_0 is fixed, it should not generate any confusion to use the same notation C_{BS} and P_{BS} for the (normalised) price with log-moneyness $x = \log(K/S_0)$:

$$S_0 C_{\mathrm{BS}}(x,T;\sigma) := C_{\mathrm{BS}}(K_x,T;S_0,\sigma) \qquad \text{and} \qquad S_0 P_{\mathrm{BS}}(x,T;\sigma) := P_{\mathrm{BS}}(K_x,T;S_0,\sigma),$$

where $K_x := S_0\mathrm{e}^x$. The definitions (1.6) of C_{BS} and (1.7) of P_{BS} are useful when both S_0 and K vary, as in Section 5. The implied volatility $I_T(x)$ is defined as the unique solution in $[0,\infty)$ to the equation

$$C(K_x,T) = S_0 C_{\mathrm{BS}}(x,T;I_T(x)). \tag{1.8}$$

Note that $I_T(x)$ is a strictly positive real number when C satisfies the strict arbitrage bounds $(S_0 - K_x)^+ < C(K_x,T) < S_0$, and it is zero whenever $C(K_x,T) = (S_0 - K_x)^+$. With a slight abuse of notation, we might also denote $I_T(K) = I_T(\log(K/S_0))$ the implied volatility as a function of strike. Finally, let us write

$$d_{1,2}(x,T,\sigma) := d_{1,2}(K_x,T,S_0,\sigma) = \frac{-x}{\sigma\sqrt{T}} \pm \frac{\sigma\sqrt{T}}{2}; \tag{1.9}$$

moreover, we denote $d_{1,2}(x,T) := d_{1,2}(x,T,I_T(x))$ whenever $I_T(x) \neq 0$.

Function asymptotics. Following usual notation, by $f(x) \sim g(x)$ (resp. $f(x) = \mathcal{O}(g(x))$, $f(x) = o(g(x))$) as $x \to x_0 \in \overline{\mathbb{R}} = [-\infty,\infty]$ when g is non null in a punctured neighbourhood of x_0, we mean that the ratio $f(x)/g(x)$ tends to one as $x \to x_0$ (resp. $\limsup_{x\to x_0} |f(x)/g(x)| < \infty$, resp. $f(x)/g(x)$ tends to zero). If g can be null around x_0, $f = o(g)$ (resp. $f = \mathcal{O}(g)$) is understood in the sense $f(x) = g(x)\phi(x)$ in a neighbourhood of x_0, for some ϕ such that $\phi(x) \to 0$ as $x \to x_0$ (resp. for some ϕ bounded around x_0). Finally, by $f(x) = g(x) + \mathcal{O}(h(x))$ as $x \to x_0$ we mean that $f(x) - g(x)$ is $\mathcal{O}(h(x))$ as $x \to x_0$.

2 Detecting the mass of the atom: the second-order behaviour

In this section we show that the existence of a positive mass at zero can be detected from the behaviour of the implied volatility at the second order. The moment formula (1.3) guarantees that $\limsup_{x\downarrow-\infty} I_T(x)^2T/|x|$ is strictly smaller than 2 when $q^* > 0$. The two situations where the lim sup reaches its maximum level 2, then, are the case $q^* = 0$ but $\mathbb{P}(S_T = 0) = 0$ (heavy left tail, but no mass at zero), and $\mathbb{P}(S_T = 0) > 0$. The former is considered by Gulisashvili [15], as we recall below; we focus on the latter, which appears to be new, in Section 2.1.

Example 2.1. In the Hull-White stochastic volatility model, the stock price process satisfies the stochastic differential equation $\mathrm{d}S_t = S_t|Z_t|\mathrm{d}W_t$, with $S_0 > 0$, and Z is a lognormal process satisfying $\mathrm{d}Z_t = \nu Z_t\mathrm{d}t + \xi Z_t\mathrm{d}B_t$, with W and B two correlated Brownian motions $\mathrm{d}\langle W, B\rangle_t = \rho\mathrm{d}t$. For all $T \geq 0$, S_T is strictly positive almost surely and (as shown in [16]) has extreme moment explosions: $q^* = 0 = p^* = \sup\{p \geq 0 : \mathbb{E}(S_T^{1+p}) < \infty\}$.

In the spirit of the tail-wing formula/statements in [2], Gulisashvili's result (1.1) allows to turn the moment formula into an asymptotic statement. This requires to study the behaviour of $\psi\big(\frac{\log P(K,T)}{\log K} - 1\big)$ as $K \downarrow 0$. The identities

$$\liminf_{K\downarrow 0} \frac{\log P(K,T)}{\log K} = 1 + q^* \tag{2.1}$$

and

$$1 + q^* = \sup\{u \geq 1 : P(K,T) = \mathcal{O}(K^u) \text{ as } K \downarrow 0\}, \tag{2.2}$$

are given in [14, Lemma 4.5] and [14, Lemma 4.6] respectively[2]. For general Put price functions, the quantity $\limsup_{K\downarrow 0} \frac{\log P(K,T)}{\log K}$ can take any value greater than $1 + q^*$. In Gulisashvili [15], conditions on the Put price function equivalent to $\limsup_{K\downarrow 0} \frac{\log P(K,T)}{\log K} = 1$ are given:

Theorem 2.2 ([15], Theorem 3.6). *If $q^* = 0$ and $\mathbb{P}(S_T = 0) = 0$, then the following statements are equivalent*

(i) $\limsup_{K\downarrow 0}[\log P(K,T)/\log K] = 1$.

(ii) $I_T(x) \sim \sqrt{2|x|/T}$*, as* $x \downarrow -\infty$.

(iii) There exist $K_1 > 0$ and a regular varying[3] function h of order -1 such that $h(1/K) \leq P(K,T)$, for $K \in (0, K_1)$.

In light of the expansion (1.1), (ii) is equivalent to $\lim_{K\downarrow 0} \psi\big(\frac{\log P(K,T)}{\log K} - 1\big) = 2$, or $\lim_{K\downarrow 0} \frac{\log P(K,T)}{\log K} = 1$ by the continuity of ψ^{-1}. Taking into account (2.1), the latter condition is equivalent to (i). In [15, Theorem 3.6], the equivalence between (ii) and (iii) is proven: the approach is first to show, when $p^* = 0$, the equivalence between the asymptotics $I_T(x) \sim \sqrt{2x/T}$, $x \uparrow \infty$, and a condition on the Call price function analogous to (iii) (see [15, Theorem 3.2]), and second to apply the Put-Call symmetry in order to transfer the result from the right to the left wing of the smile. Because of the lack of Put-Call symmetry when the law of the stock price has a mass at zero, this approach is not possible when $q^* = 0$ and $\mathbb{P}(S_T = 0) > 0$—just as it happens for the asymptotic formula (1.1). We shall get back to this point in Remark 2.6.

[2] Let us give a quick proof of (2.1) using (2.2). Assume that $\liminf_{K\downarrow 0} \frac{\log P(K,T)}{\log K} = l$: then, for any $l_2 > l$, there exists a sequence $(K_n)_n \downarrow 0$ such that $\frac{\log P(K_n,T)}{\log K_n} < l_2$, hence $P(K_n,T) > K_n^{l_2}$, for n large enough. This entails $l \geq 1 + q^*$, otherwise one would have $P(K_n,T) > K_n^{l_2}$ for some $l_2 \in (l, 1 + q^*)$, contradicting (2.2). On the other hand, $l > 1 + q^*$ implies $\frac{\log P(K,T)}{\log K} > l_1$, therefore $P(K) < K^{l_1}$, for some $l_1 \in (1 + q^*, l)$ and every K in some neighbourhood of zero, again contradicting (2.1). Then one must have $l = \liminf_{K\downarrow 0} \frac{\log P(K,T)}{\log K} = 1 + q^*$.

[3] f is regularly varying of order $\alpha \in \mathbb{R}$ if f is defined on some neighbourhood of infinity, measurable, and such that for every $\lambda > 0$, $\frac{f(\lambda x)}{f(x)} \to \lambda^\alpha$ as $x \to \infty$.

2.1 First-order behaviour

Here we confirm the asymptotic behaviour $I_T(x) \sim \sqrt{2|x|/T}$ as $x \downarrow -\infty$ when $\mathbb{P}(S_T = 0) > 0$ (under no additional assumption on the law of S_T). Lemma 2.3 contains some preliminary results on the behaviour of Put and Call prices at zero; the main result of this section is Proposition 2.4.

For every T, $C(\cdot, T)$ and $P(\cdot, T)$ are convex functions on $[0, \infty)$, hence their right derivatives $\partial_K^+ C(\cdot, T)$ and $\partial_K^+ P(\cdot, T)$ and their left derivatives $\partial_K^- C(\cdot, T)$ and $\partial_K^- P(\cdot, T)$ exist everywhere on $[0, \infty)$ and on $(0, \infty)$; finally the derivatives $\partial_K C(\cdot, T)$ and $\partial_K P(\cdot, T)$ are themselves well-defined everywhere on $(0, \infty)$ except for a countable set of points.

Lemma 2.3. *For every $K > 0$, one has*

$$\partial_K^+ P(K,T) = \mathbb{P}(S_T \leq K), \qquad \partial_K^- P(K,T) = \mathbb{P}(S_T < K),$$

$$\partial_K^+ C(K,T) = -\mathbb{P}(S_T > K), \quad \partial_K^- C(K,T) = -\mathbb{P}(S_T \geq K).$$

In particular,

$$\lim_{K \downarrow 0} \frac{P(K,T)}{K} = \mathbb{P}(S_T = 0). \tag{2.3}$$

Proof. The first part of the lemma is standard [11, Chapter 7]. The second line follows from the first by Call-Put parity. Using the equality $P(K,T) = \int_0^K \partial_K^+ P(L,T)\mathrm{d}L$ (by convexity of the Put option price) and $\lim_{K\downarrow 0} \partial_K^+ P(K,T) = \mathbb{P}(S_T = 0)$, the identity (2.3) then follows from $\lim_{K\downarrow 0} \frac{1}{K}\int_0^K \partial_K^+ P(L,T)\mathrm{d}L = \mathbb{P}(S_T = 0)$. ■

Proposition 2.4. *If $p_T := \mathbb{P}(S_T = 0) > 0$, then $I_T(x) \sim \sqrt{2|x|/T}$ as $x \downarrow -\infty$.*

Proposition 2.4 is a consequence of a more general statement that we prove in Theorem 2.9 below; we give here an independent proof of the proposition, essentially relying on Lemma 2.3, that goes through the analysis of a parameterisation of the implied volatility.[4]

Proof. Let $\overline{P}(K,T)$ be a Put price associated to an underlying distribution with no atom at zero. It follows from (2.3) that

$$\lim_{K\downarrow 0} \frac{\overline{P}(K,T)}{P(K,T)} = 0, \tag{2.4}$$

and hence $\overline{P}(K,T) < P(K,T)$ for K small enough. Set $\sigma_\gamma(x) := \sqrt{\gamma|x|/T}$ for $x < 0$ and some $\gamma > 0$, and define

$$\overline{P}_\gamma(K,T) := S_0 P_{\mathrm{BS}}(x, T; \sigma_\gamma(x)), \qquad x = \ln(K/S_0).$$

Fix now some $x^* < 0$. It is not clear a priori that σ_γ is the restriction to some interval $(-\infty, x^*)$ of an implied volatility function, so that $x \mapsto \overline{P}_\gamma(K_x, T)$ is a true Put price (restricted to $(-\infty, x^*)$). Necessary and sufficient conditions for a twice differentiable function to be an implied volatility are known from [28, Theorems 2.9 and 2.15]: in our context here, σ_γ is (the restriction to $(-\infty, x^*)$ of) a true implied volatility if and only if [28, Condition (IV.3) in Theorem 2.9]:

$$\left(1 - \frac{x\partial_x\omega}{\omega}\right)^2 - \frac{1}{4}\omega^2(\partial_x\omega)^2 + \omega\partial_{xx}^2\omega \geq 0, \tag{2.5}$$

for all $x \leq x^*$, where $\omega(x) \equiv \sqrt{T}\sigma_\gamma(x)$. Simple calculations yield the equivalent condition

$$\gamma^2|x| + 4\gamma - 4|x| < 0. \tag{2.6}$$

The positive root of (2.6), $\gamma_+(x) \equiv 2(\sqrt{1 + 1/x^2} - 1/|x|)$, is a strictly decreasing function of x and takes values between $\lim_{x\downarrow -\infty} \gamma_+(x) = 2$ and $\lim_{x\uparrow 0} \gamma_+(x) = 0$. The inequality (2.6) is satisfied when $\gamma \leq \gamma_+(x)$: it follows

[4] Also note that the asymptotic equivalence in Proposition 2.4 is not explicitly stated in Lee [21], but is actually a consequence of an argument used in the proof of [21, Theorem 4.3]. Lee's argument is recalled in Remark 2.7 for completeness.

that for any $\gamma < 2$, $\sigma_\gamma(x)$ is a true implied volatility for all $x < x^*$, where x^* is the unique solution in $(-\infty, 0)$ to the equation $\gamma_+(x^*) = \gamma$. Now

$$\begin{aligned}\partial_K \overline{P}_\gamma(K_x, T) = (S_0 \mathrm{e}^x)^{-1} \partial_x \overline{P}_\gamma(K_x, T) &= \partial_x P_{\mathrm{BS}}(x, t; \sigma_\gamma(x)) + \partial_\sigma P_{\mathrm{BS}}(x, T; \sigma_\gamma(x))\ \partial_x \sigma_\gamma(x) \\ &= N\left(-d_2(x, T, \sigma_\gamma(x))\right) - \frac{\sqrt{\gamma}}{2\sqrt{|x|}} \phi\left(d_2(x, T, \sigma_\gamma(x))\right) \\ &= N\Big(-\frac{2-\gamma}{2\sqrt{\gamma}}\sqrt{|x|}\Big) - \frac{\sqrt{\gamma}}{2\sqrt{|x|}} \phi(d_2(x, T, \sigma_\gamma(x))).\end{aligned}$$

Since ϕ is bounded, for any $\gamma < 2$, $\mu_\gamma(\{0\}) = \lim_{x\downarrow-\infty} \partial_K \overline{P}_\gamma(K_x, T) = 0$, where μ_γ is the underlying distribution associated with $\overline{P}_\gamma$. Therefore, $\overline{P}_\gamma$ is a Put price without mass at zero: it follows from (2.4) that $\overline{P}_\gamma(K, T) < P(K, T)$ for K in a neighborhood of the origin. Thus by strict monotonicity of P_{BS} with respect to the volatility parameter, $\sigma_\gamma(x) < I_T(x)$ for K_x in that same neighborhood, which implies

$$\liminf_{K\downarrow 0} \frac{T I_T(x)^2}{|x|} > \gamma \tag{2.7}$$

for any $\gamma \in (0, 2)$, and the proposition follows applying (1.3). ■

Remark 2.5. As a by-product of the previous proof, we have shown that for every $\gamma \in [0, 2)$ the map $\sigma_\gamma : (x, T) \mapsto \sqrt{\gamma|x|/T}$ is a true implied volatility function for any $T > 0$ on the interval $(-\infty, 4\gamma/(\gamma^2 - 4)]$. Indeed, the operator in (2.5) simplifies to $\Phi_\gamma(x) \equiv -\frac{1}{16x}\left((\gamma^2-4)x - 4\gamma\right)$. For any $\gamma > 0$, Φ_γ is strictly decreasing on $(-\infty, 0)$ with $\lim_{x\downarrow-\infty} \Phi_\gamma(x) = (4-\gamma^2)/16$ and $\lim_{x\uparrow 0} \Phi_\gamma(x) = -\infty$. For any $\gamma \in (0, 2)$, the equation $\Phi_\gamma(x) = 0$ on $(-\infty, 0)$ has $x^* := 4\gamma/(\gamma^2-4)$ as unique solution, hence σ_γ is a genuine implied volatility smile on $(-\infty, x^*]$, but not on $(x^*, 0]$. When $\gamma = 0$, Φ_0 is constantly equal to $1/4$, and the resulting null implied volatility on $(-\infty, 0)$ corresponds to the trivial case $S_T = S_0$ a.s. Finally, whenever $\gamma \geq 2$, $\Phi_\gamma(x) < 0$ for all $x < 0$, and arbitrage occurs (the corresponding Put price is not convex any longer). Note that the limiting case $\gamma = 2$ (compatible with Lee's moment formula (1.3)), is associated to the Put price $P(K, T) = K/2 - S_0 N(-\sqrt{2|\log K/S_0|})$, $K < S_0$, whose second derivative with respect to the strike is strictly negative. The total variance map $T \mapsto T\sigma_\gamma^2(x, T)$ is increasing (actually constant), thus ensuring that the corresponding underlying distributions are increasing in the convex order, precluding calendar spread arbitrage.

Remark 2.6. The function $h_1(K) := p_T/K$, $p_T > 0$, is regularly varying of index -1. Since $P(K, T) \geq p_T K$ for every $K \geq 0$, Condition (iii) in Theorem 2.2 is satisfied with the function h_1 when $p_T > 0$. Moreover, using the asymptotics (2.3) in Lemma 2.3, it is immediate to see that $\log P(K, T) = \log K + \mathcal{O}(1)$ as $K \downarrow 0$, hence $\log P(K, T)/\log K \to 1 = 1 + q^*$ as $K \downarrow 0$. Then, in view of Proposition 2.4, the statement of Theorem 2.2 turns out to be true also in the case $p_T = \mathbb{P}(S_T = 0) > 0$ (the Conditions (i), (ii), and (iii) of Theorem 2.2 being all satisfied in this case).

Remark 2.7. We recall here the argument in the proof of [21, Theorem 4.3] that allows to prove Proposition 2.4. Assume $\mathbb{P}(S_T = 0) > 0$. Then, for any $\gamma < 2$ there exists $x^* < 0$ such that for all $x < x^*$

$$P(K_x, T) \geq K_x \mathbb{P}(S_T = 0) > K_x\left[N(-d_2(x, T, \sigma_\gamma(x)) - \mathrm{e}^{-x} N(-d_1(x, T, \sigma_\gamma(x)))\right] = S_0 P_{\mathrm{BS}}(x, T; \sigma_\gamma(x)), \tag{2.8}$$

which proves (2.7). Indeed $N(-d_2(x, T, \sigma_\gamma(x)) = N\left(-\frac{2-\gamma}{2\sqrt{\gamma}}\sqrt{|x|}\right)$ tends to zero as $x \downarrow -\infty$, and l'Hôpital's rule implies that $\lim_{x\downarrow-\infty} \mathrm{e}^{-x} N(-d_1(x, T, \sigma_\gamma(x))) = \lim_{x\downarrow-\infty} \frac{\mathrm{e}^{-x}}{\sqrt{2\pi}} \exp\left(-\frac{(2+\gamma)^2}{8\gamma}|x|\right) \frac{2+\gamma}{4\sqrt{\gamma}} \frac{1}{\sqrt{|x|}} = 0$. Note that we do not prove here that $P_{\mathrm{BS}}(x, T; \sigma_\gamma(x))$ is a true Put price for x in some interval, as in our proof above.

2.2 Second-order behaviour

Once established that the implied volatility is asymptotic to $\sqrt{2|x|/T}$ as $x \downarrow -\infty$ if and only if $q^* = 0$ and Condition (iii) in Theorem 2.2 is fulfilled (always true when $p_T > 0$, see Remark 2.6), it remains to understand how the difference $I_T(x) - \sqrt{2|x|/T}$ behaves. The behaviour of the right wing ($x \uparrow +\infty$) has been studied by

Lee [21, Lemma 3.1], who showed that $I_T(x) - \sqrt{2|x|/T} < 0$ for x large enough, and subsequently refined by Rogers and Tehranchi [27, Theorem 5.3], who proved that $\lim_{x\uparrow+\infty}(I_T(x) - \sqrt{2x/T}) = -\infty$ for every $T > 0$. For the left wing, the situation is different, and the qualitative behaviour of the second-order term depends on the presence of a mass at zero.

Ohsaki et al. [25] study the asymptotic behaviour of $d_2(x,T)$ as $x \downarrow -\infty$. Lemma 2.8 below refines their Theorem 1: on the one hand, no assumption on the differentiability of I_T is made, on the other hand, we do not assume that the derivative of I_T has a limit at $-\infty$ (as opposed to [25, Assumption 1]).

Lemma 2.8. *The following limits hold:*

$$\begin{array}{ll} \lim_{x\downarrow-\infty} d_2(x,T) = +\infty, & \text{if } p_T = 0; \\ \lim_{x\downarrow-\infty} d_2(x,T) = -N^{-1}(p_T), & \text{if } p_T > 0. \end{array} \tag{2.9}$$

Based on Lemma 2.8, we prove the following

Theorem 2.9. *If the distribution of the stock price S_T does not charge zero ($p_T = \mathbb{P}(S_T = 0) = 0$), then*

$$\lim_{x\downarrow-\infty} \left(I_T(x) - \sqrt{2|x|/T}\right) = -\infty. \tag{2.10}$$

On the contrary, if $p_T > 0$, then

$$\lim_{x\downarrow-\infty} \left(I_T(x) - \sqrt{2|x|/T}\right) = \frac{1}{\sqrt{T}} N^{-1}(p_T). \tag{2.11}$$

Remark 2.10. Some comments are in order.

(i) Lemma 3.3 in [21] asserts that there exists x^* such that $I_T(x) - \sqrt{2|x|/T} < 0$ for all $x < x^*$ if and only if $0 \le p_T < 1/2$. In light of the new estimate (2.11), the difference $I_T(x) - \sqrt{2|x|/T}$ converges to a negative constant when $0 < p_T < 1/2$, to a positive constant when $p_T > 1/2$, and to zero when $p_T = 1/2$.

(ii) Differently put, when the left asymptotic slope of the smile is maximal ($\lim_{x\downarrow-\infty} T I_T(x)^2/|x| = 2$), the difference between an underlying distribution that charges the origin and one that does not is seen at the second-order in implied volatilities.

(iii) When $p_T > 0$, formally plugging the limit (2.3) into (1.1), the error term in Gulisashvili's asymptotic (1.1) is a constant: $\mathcal{O}\left(\left(\log \frac{K}{P(K,T)}\right)^{-1/2} \log\log \frac{K}{P(K,T)}\right) = \mathcal{O}(1)$ as $K \downarrow 0$, and the main term satisfies

$$\sqrt{\frac{|\log K|}{T}} \sqrt{\psi\left(\frac{\log P(K,T)}{\log K} - 1\right)} \sim \sqrt{\frac{|\log K|}{T}} \sqrt{\psi(0)} = \sqrt{\frac{2|\log K|}{T}}, \qquad \text{as } K \downarrow 0.$$

Therefore, comparing with (2.11), the expansion (1.1) turns out to be true also when there is a positive mass at zero. Theorem 2.9 makes the constant term precise, while Theorem 3.7 below provides an additional third-order term and an error estimate.

Proof[5] of Lemma 2.8 and Theorem 2.9. The identity

$$C_{\mathrm{BS}}(-x, T; I_T(x)) = \mathbb{E}\left(1 - \mathrm{e}^{-x}\frac{S_T}{S_0}\right)^+$$

follows from equations (A.3) and (A.4) in the Appendix. Now, since $\mathbb{E}(1 - \frac{S_T}{S_0}\mathrm{e}^{-x})^+ = \mathbb{P}(S_T = 0)$ $+\mathbb{E}\left[(1 - \frac{S_T}{S_0}\mathrm{e}^{-x})^+ \mathbb{1}_{\{S_T>0\}}\right]$, $C_{\mathrm{BS}}(-x, T, I_T(x))$ tends to $\mathbb{P}(S_T = 0)$ as $x \uparrow -\infty$ by dominated convergence. Therefore, for every $x < 0$

$$\begin{aligned} N(d_1(-x, T, I_T(x))) &= C_{\mathrm{BS}}(-x, T, I_T(x)) + \mathrm{e}^{-x} N(d_2(-x, T, I_T(x))) \\ &\le C_{\mathrm{BS}}(-x, T, I_T(x)) + \mathrm{e}^{-x} N(-\sqrt{2|x|}) \to \mathbb{P}(S_T = 0), \qquad \text{as } x \to -\infty, \end{aligned} \tag{2.12}$$

[5] We are grateful to Mike Tehranchi for sharing this neat proof with us.

where we have used the arithmetic-geometric inequality $d_2(-x,T,I_T(x)) = -\frac{|x|}{I_T(x)\sqrt{T}} - \frac{I_T(x)\sqrt{T}}{2} \leq -\sqrt{2|x|}$ and the limit $e^{z^2/2}N(-z) \to 0$ as $z \to \infty$ in the last step. Noting that $d_1(-x,T,I_T(x)) = -d_2(x,T,I_T(x))$, (2.12) proves Lemma 2.8.

Let $\mathfrak{p} := N^{-1}(\mathbb{P}(S_T = 0))$. Assume first that $\mathfrak{p} = -\infty$. Estimate (2.12) implies that for every $M > 0$ we have $d_1(-x,T,I_T(x)) = \frac{x}{I_T(x)\sqrt{T}} + \frac{I_T(x)\sqrt{T}}{2} < -M$ for x small enough, or yet $I_T(x)\sqrt{T} < -M + \sqrt{M^2 + 2|x|}$. Therefore,

$$\limsup_{x\downarrow-\infty}\left(I_T(x)\sqrt{T} - \sqrt{2|x|}\right) < -M + \limsup_{x\downarrow-\infty}(\sqrt{M^2+2|x|} - \sqrt{2|x|}) = -M$$

for every $M > 0$, which proves (2.10).

Now assume $\mathfrak{p} > -\infty$. Then for fixed $\varepsilon > 0$, we have $\mathfrak{p} - \varepsilon < d_1(-x,T,I_T(x)) < \mathfrak{p} + \varepsilon$ for x small enough. It follows that:

$$\mathfrak{p} - \varepsilon + \sqrt{2|x|} < I_T(x)\sqrt{T} = d_1(-x,T,I_T(x)) - d_2(-x,T,I_T(x)) < \mathfrak{p} + \varepsilon + \sqrt{(\mathfrak{p}+\varepsilon)^2 + 2|x|}.$$

The lower bound again follows from the arithmetic-geometric inequality for d_2, and the upper bound from the identity $d_2(-x,T,\sigma)^2 = d_1(-x,T,\sigma)^2 + 2|x|$. Therefore, $\lim_{x\downarrow-\infty}(I_T(x)\sqrt{T} - \sqrt{2|x|}) = \mathfrak{p}$, and (2.11) is proved.
■

3 An asymptotic formula for the implied volatility

Theorem 2.9 establishes that the expansion

$$I_T(x) = \sqrt{2|x|/T} + N^{-1}(p_T)/\sqrt{T} + o(1), \qquad x \downarrow -\infty \tag{3.1}$$

holds if $p_T = \mathbb{P}(S_T = 0) > 0$. In this section we refine the crude $o(1)$ term (to a $\mathcal{O}(|x|^{-1/2})$ term), providing an explicit third-order term and an error estimate.

We start with a lemma (proved in Appendix A.1), which (a) recalls the existence of the side derivatives $D^{\pm}I_T$ (nicely illustrated by Rogers and Tehranchi [27, Theorem 5.1]), and (b) provides a new estimate (3.5) on the derivatives under Condition (1.4).

Lemma 3.1. *Let $[\underline{x}_T, \overline{x}_T]$ be the smallest interval containing the support of $\log(S_T/S_0)$ (with the convention $\log 0 = -\infty$). The right derivative $D^+I_T(x)$ exists for all $x \neq \underline{x}_T$ and the left derivative $D^-_x I_T(x)$ exists for all $x \neq \overline{x}_T$. The two derivatives satisfy, for all $x \in (\underline{x}_T, \overline{x}_T)$,*

$$D^+I_T(x) = \frac{N(d_2(x,T)) - \mathbb{P}(S_T > K_x)}{\sqrt{T}\phi(d_2(x,T))} \tag{3.2}$$

$$D^-I_T(x) = \frac{N(d_2(x,T)) - \mathbb{P}(S_T \geq K_x)}{\sqrt{T}\phi(d_2(x,T))}, \tag{3.3}$$

and $D^+I_T(x)$ (resp. $D^-I_T(x)$) is null on $(-\infty, \underline{x}_T) \cup [\overline{x}_T, +\infty)$ (resp. $(-\infty, \underline{x}_T] \cup (\overline{x}_T, +\infty)$). In particular, $D^-I_T(x) \leq D^+I_T(x)$ wherever both sides exist.

If $\mathbb{P}(S_T < K_x) > 0$ for all $x < 0$, then $\underline{x}_T = -\infty$ and

$$D^-I_T(x) \geq -\frac{1}{\sqrt{2T|x|}} \tag{3.4}$$

for all $x < 0$. Moreover, if $q^* = 0$ and condition (1.4) holds, then for every $\varepsilon' < \varepsilon$ there exists $x^*_{\varepsilon'} < 0$ and a constant $c > 0$ such that, for all $x < x^*_{\varepsilon'}$,

$$D^+I_T(x) \leq c\,\mathrm{e}^{\varepsilon' x} = c\,\mathrm{e}^{-\varepsilon'|x|}. \tag{3.5}$$

Remark 3.2. If there exists $x^* < 0$ such that $D^+I_T(x) \leq 0$ for $x < x^*$, the upper bound in (3.5) can trivially be updated to zero.

Remark 3.3. Identity (3.2) can be rewritten as

$$N(d_2(x,T)) = \mathbb{P}(S_T > K_x) + \sqrt{T}\phi(d_2(x,T))D^+I_T(x).$$

Taking the limit as $x \downarrow -\infty$ and using $-N^{-1}(p_T) = N^{-1}(1-p_T) = N^{-1}(\mathbb{P}(S_T > 0))$, Lemma 2.8 entails that if $p_T > 0$, then $D^+I_T(x)$ tends to zero as $x \to -\infty$.

We shall from now on make the standing assumption that $\mathbb{P}(S_T < K_x) > 0$ for all $x < 0$. Proposition 3.5 below refines Lemma 2.8 providing an error term for $d_2(x,T)$ under Condition (1.4); it is based on the following lemma.

Lemma 3.4. *Assume $p_T > 0$. Then the following estimate holds as $x \to -\infty$:*

$$d_2(x,T) = -N^{-1}(p_T) - \phi(N^{-1}(p_T))^{-1}\big[F(K_x) - F(0)\big] + \sqrt{T}D^+I_T(x) + \mathcal{O}(F(K_x) - F(0))^2 + o(D^+I_T(x)). \quad (3.6)$$

Proof. Write $d_2(x,T) = -N^{-1}(p_T) + R(x)$, where $R(x)$ tends to zero as $x \downarrow -\infty$. Then the identity $\phi'(d) = -d\phi(d)$ yields

$$\phi(d_2(x,T)) = \phi(N^{-1}(p_T)) + N^{-1}(p_T)\phi(N^{-1}(p_T))R(x) + \mathcal{O}\left(R(x)^2\right).$$

Plugging this estimate into the identity $d_2(x,T) = N^{-1}\left(\mathbb{P}(S_T > K_x) + \sqrt{T}\phi(d_2(x,T))D^+I_T(x)\right)$ from (3.2), we obtain

$$\begin{aligned}
d_2(x,T) = N^{-1}\Big(&1 - p_T - [F(K_x) - F(0)] + \sqrt{T}\phi(N^{-1}(p_T))D^+I_T(x) \\
&+ \sqrt{T}N^{-1}(p_T)\phi(-N^{-1}(p_T))D^+I_T(x)R(x) + \mathcal{O}\big(R(x)^2D^+I_T(x)\big)\Big) \\
= &N^{-1}(1-p_T) + \phi(N^{-1}(p_T))^{-1}\left(-[F(K_x) - F(0)] + \sqrt{T}\phi(N^{-1}(p_T))D^+I_T(x)\right) \\
&+ \sqrt{T}N^{-1}(p_T)D^+I_T(x)R(x) + \mathcal{O}(F(K_x) - F(0))^2 + \mathcal{O}(D^+I_T(x))^2 \\
= &-N^{-1}(p_T) - \phi(N^{-1}(p_T))^{-1}[F(K_x) - F(0)] + \sqrt{T}D^+I_T(x) \\
&+ o(D^+I_T(x)) + \mathcal{O}(F(K_x) - F(0))^2,
\end{aligned}$$

and (3.6) is proven. ■

Proposition 3.5. *If $p_T > 0$ and Condition* (1.4) *holds, then there exists a function $\varphi : (-\infty, 0) \to \mathbb{R}$ with $\limsup_{x\downarrow-\infty} |\varphi(x)|\sqrt{2|x|} \le 1$ such that*

$$d_2(x,T) = -N^{-1}(p_T) + \varphi(x) \qquad \textit{as } x \downarrow -\infty. \quad (3.7)$$

Proof. Recall that $p_T > 0$ implies $q^* = 0$. Therefore, under condition (1.4), the bounds (3.4) and (3.5) imply

$$|D^+I_T(x)| \le \frac{1}{\sqrt{2T|x|}}, \quad (3.8)$$

for x small enough. Condition (1.4) further implies $F(K_x) - F(0) = \mathcal{O}(\mathrm{e}^{-\varepsilon|x|}) = o(1/\sqrt{|x|})$, and (3.7) follows from (3.8) and (3.6). ■

Remark 3.6. If one knows that the smile is decreasing at minus infinity, that is there exists x^* such that $D^+I_T(x) \le 0$ for $x < x^*$, then (3.8) is an immediate consequence of (3.4), without any further assumption. If moreover $p_T > 0$, according to Lemma 3.4 the estimate (3.7) holds with $\varphi(x)$ replaced by $\tilde{\varphi}(x) - \phi(N^{-1}(p_T))^{-1}\big[F(K_x) - F(0)\big] + \mathcal{O}(F(K_x) - F(0))^2$, where $\tilde{\varphi}(x) \le \frac{1}{\sqrt{2|x|}}$ for x small enough.

The following theorem contains the main result of this section.

Theorem 3.7. *Assume $p_T > 0$.*

(i) Under (1.4)*, there exists a function* $\Phi : (-\infty, 0) \to \mathbb{R}$ *satisfying* $\limsup_{x\downarrow-\infty} \sqrt{2|x|T}\,|\Phi(x)| \leq 1$ *such that*

$$I_T(x) = \sqrt{\frac{2|x|}{T}} + \frac{1}{\sqrt{T}}N^{-1}(p_T) + \frac{\left[N^{-1}(p_T)\right]^2}{2\sqrt{2T|x|}} + \Phi(x) \quad \text{as } x \downarrow -\infty. \tag{3.9}$$

(ii) If $D^+ I_T(x) \leq 0$ *for* x *small enough, there exists* $\widetilde{\Phi} : (-\infty, 0) \to \mathbb{R}$ *satisfying* $\limsup_{x\downarrow-\infty} \sqrt{2|x|T}|\widetilde{\Phi}(x)| \leq 1$ *such that, as* $x \downarrow -\infty$*,*

$$I_T(x) = \sqrt{\frac{2|x|}{T}} + \frac{1}{\sqrt{T}}N^{-1}(p_T) + \frac{[N^{-1}(p_T)]^2}{2\sqrt{2T|x|}} + \overline{\Phi}(x) + \mathcal{O}(F(K_x) - F(0))^2, \tag{3.10}$$

where

$$\overline{\Phi}(x) \equiv \widetilde{\Phi}(x) + \sqrt{\frac{2\pi}{T}} \exp\left(\frac{N^{-1}(p_T)^2}{2}\right)[F(K_x) - F(0)].$$

The following comments emphasise the practical importance of this theorem.

1. The asymptotic (3.9) contains a global $\mathcal{O}(|x|^{-1/2})$ term, given by $\frac{1}{\sqrt{2T|x|}}\left(\frac{N^{-1}(p_T)^2}{2} + \sqrt{2T|x|}\Phi(x)\right)$. Since $N^{-1}(p_T)^2$ becomes large as p_T tends to zero or one, and $\limsup_{x\downarrow-\infty} \sqrt{2|x|T}\,|\Phi(x)| \leq 1$, the contribution of the function Φ becomes negligible in front of the $N^{-1}(p_T)^2$ term when p_T is close to zero or one. In this regime, the latter is an explicit third-order term, and the function Φ is an error term. When p_T has intermediate values (close to 1/2), the sum of the two terms can be seen as a global $\mathcal{O}(|x|^{-1/2})$ estimate.

2. The inspection of (3.9) reveals a 'phase transition' in the behaviour of the implied volatility at the second-order, too: when $p_T = 1/2$, both the constant and the third-order term cancel, and (3.9) reduces to $I_T(x) = \sqrt{2|x|/T} + \Phi(x)$. In this case, the 'normalised' implied volatility $I_T(x)\sqrt{T}/\sqrt{|x|}$ converges much faster to its limit $\sqrt{2}$.

3. Assume that the underlying stock price is distributed according to the measure

$$\mu(\mathrm{d}K) = p_T\delta_0(\mathrm{d}K) + (1 - p_T)f(K)\mathrm{d}K, \tag{3.11}$$

where $p_T \in [0, 1)$ and f is a probability density function on $(0, \infty)$. If $f(K) = \mathcal{O}(K^{-a})$ as $K \downarrow 0$, for some $a < 1$, it is immediate that $F(K) - F(0) = \mathcal{O}(K^{1-a})$, and Condition (1.4) is fulfilled. Note that as soon as $f(K) \sim K^{-a}$ as $K \downarrow 0$, the restriction $a < 1$ is necessary to ensure integrability. Nearly all financial models used in practice satisfy (3.11) and the condition on f; in particular, we refer the reader to the Merton model with jump-to-default in Section 4.2.2 (whose density f tends to zero at the origin), and the CEV model in Section 4.3.1 (where the density f explodes at the origin).

4. The role played by the cumulative distribution function F in the error term $\overline{\Phi}$ in (3.10) highlights a radical difference with the no-mass-at-zero case. In the classical left tail-wing formula [2],

$$I_T(x) \sim \sqrt{|x|}\sqrt{\psi\left(\frac{-\log F(K_x)}{|x|}\right)}, \qquad \text{as } x \downarrow -\infty, \tag{3.12}$$

where ψ is defined in (1.2). Note that the logarithm of the cdf $F(K_x)$ appears in the formula, instead of the distribution function itself as in (3.10). In many stochastic volatility models, such as Heston and Stein-Stein, the cdf of the stock price satisfies, see [17, 12],

$$F(K_x) = A\,\mathrm{e}^{-\alpha_1|x| + \alpha_2\sqrt{|x|}}\,|x|^\gamma(1 + \mathcal{O}(|x|^{-1/2})) \qquad \text{as } x \downarrow -\infty, \tag{3.13}$$

for some constants $A, \alpha_1, \alpha_2 > 0$ and $\gamma \in \mathbb{R}$. Therefore, $-\log F(K_x)/|x| = \alpha_1 + \mathcal{O}(|x|^{-1/2})$, and (3.12) returns—as expected—the leading-order square root behaviour $I_T(x) \sim \sqrt{\psi(\alpha_1)|x|}$ (subsequent refinements, in analogy with (3.9), are of course possible using the precise asymptotics (3.13), as done in [12, 13]; see Remark 3.8 below for more details). For any distribution such that $F(K_x) - F(0)$ behaves as the right-hand side of (3.13), the term proportional to $F(K_x) - F(0) = \mathcal{O}(\mathrm{e}^{-\alpha_1|x|})$ in (3.10) goes to zero much faster than any of the other $\mathcal{O}(|x|^{-1/2})$ terms in (3.9).

Proof of Theorem 3.7. We first prove Theorem 3.7(i). Let us write $d_2(x)$ instead of $d_2(x,T)$. According to the definition of d_2 in (1.9), $I_T(x)$ satisfies

$$I_T(x) = \frac{1}{\sqrt{T}}\left(-d_2(x) + \sqrt{d_2^2(x) - 2x}\right) = \frac{2|x|}{\sqrt{2|x|T + Td_2(x)^2} + \sqrt{T}d_2(x)},$$

hence

$$\sqrt{T}I_T(x) - \sqrt{2|x|} = \frac{2|x| - \sqrt{4x^2 + 2|x|d_2(x)^2}}{\sqrt{2|x| + d_2(x)^2} + d_2(x)} - \frac{\sqrt{2|x|}d_2(x)}{\sqrt{2|x| + d_2(x)^2} + d_2(x)} =: A(x) - B(x). \tag{3.14}$$

The equivalence $\sqrt{f(x)} - \sqrt{f(x) + g(x)} \sim -g(x)/(2\sqrt{f(x)})$ when $f(x) \uparrow \infty$ and $g = o(f)$ allows to see that $A(x)$ satisfies

$$A(x) \sim -\frac{2|x|d_2(x,T)^2}{4|x|\sqrt{2|x|}} \sim -\frac{[N^{-1}(p_T)]^2}{2\sqrt{2|x|}} \quad \text{as } x \downarrow -\infty. \tag{3.15}$$

On the other hand, $\lim_{x\downarrow-\infty} B(x) = \lim_{x\downarrow-\infty} d_2(x,T) = -N^{-1}(p_T)$; then consider $B(x) + N^{-1}(p_T) = B_1(x) + B_2(x)$, where

$$B_1(x) := \frac{N^{-1}(p_T)d_2(x,T)}{\sqrt{2|x| + d_2(x,T)^2} + d_2(x,T)}; \qquad B_2(x) := h(x)\left(d_2(x,T) + N^{-1}(p_T)\sqrt{1 + \frac{d_2(x,T)^2}{2|x|}}\right),$$

and $h(x) := \frac{\sqrt{2|x|}}{\sqrt{2|x|+d_2(x,T)^2}+d_2(x,T)}$. $B_1(x)$ is asymptotic to $-[N^{-1}(p_T)]^2/\sqrt{2|x|}$ as $x \downarrow -\infty$. It is clear that $h(x)$ tends to one as $x \downarrow -\infty$ Using the elementary identity $\sqrt{a} - \sqrt{b} = \frac{a-b}{\sqrt{a}+\sqrt{b}}$ and taking into account Proposition 3.5, one can see that $B_2(x)$ satisfies

$$\begin{aligned}
\sqrt{2|x|}\,|B_2(x)| &= |h(x)|\left(\sqrt{2|x|}\,|\varphi(x)| + \frac{|N^{-1}(p_T)|d_2(x)^2}{\sqrt{2|x|}\left(1 + \sqrt{1 + \frac{d_2(x,T)^2}{2|x|}}\right)}\right) \\
&\leq |h(x)|\left(\sqrt{2|x|}\,|\varphi(x)| + \frac{|N^{-1}(p_T)|d_2(x)^2}{\sqrt{2|x|}}\right),
\end{aligned}$$

therefore $\limsup_{x\downarrow-\infty} \sqrt{2|x|}\,|B_2(x)| \leq 1$. Taking into account (3.15), the estimate on $B_1(x)$ and this final estimate on $B_2(x)$, it follows from (3.14) that

$$\sqrt{T}I_T(x) - \sqrt{2|x|} - N^{-1}(p_T) = A(x) - (B_1(x) + B_2(x)) = \frac{N^{-1}(p_T)^2}{2\sqrt{2|x|}} + r(x),$$

as $x \downarrow -\infty$, for some function r such that $\limsup_{x\downarrow-\infty} \sqrt{2|x|}\,|r(x)| \leq 1$. The function Φ in (3.9) is given by $\Phi(\cdot) = r(\cdot)/\sqrt{T}$. Now consider Theorem 3.7(ii). When $D^+I_T(x) \leq 0$ for x small enough, the additional term $\frac{1}{\sqrt{T}}\phi(N^{-1}(p_T))^{-1}[F(K_x) - F(0)] + \mathcal{O}(F(K_x) - F(0))^2$ comes from the second part of Lemma 3.4, arguing as in Remark 3.6. ■

Remark 3.8 (Comparison with stochastic volatility models). The asymptotic form "leading-order $\sqrt{|x|}$ term + constant + vanishing term" is typical in stochastic volatility models. Yet, the phenomenon has a different nature: when the stock price follows an exponential (hence strictly positive) diffusion process with stochastic volatility, the functional form of the implied volatility is determined by the asymptotic properties of the density of the stock close to zero. In Theorem 3.7, the same parametric form specifically relates to the presence of an atom at zero, but is rather independent of the shape of the remaining distribution on $(0,\infty)$.

1. In the (uncorrelated) Stein-Stein model, the stock price process satisfies the SDE $\mathrm{d}S_t = S_t|Z_t|\mathrm{d}W_t$ with $S_0 > 0$, $\mathrm{d}Z_t = \kappa(\theta - Z_t)\mathrm{d}t + \xi\mathrm{d}B_t$, and W, B are independent Brownian motions. Then, S_t is strictly

positive almost surely for all $t \geq 0$; Gulisashvili and Stein [17, Theorem 3.1] prove the following expansion for the implied volatility:

$$I_T(x) = \frac{1}{\sqrt{T}}\left(\gamma_1\sqrt{x} + \gamma_2 + \mathcal{O}\left(\frac{1}{\sqrt{x}}\right)\right), \quad \text{as } x \uparrow +\infty,$$

where $\gamma_1 \in (0, \sqrt{2})$ and $\gamma_2 > 0$ depend on model parameters (note that this expansion is given in [17] with an error term of the form $\mathcal{O}(\chi(x)/\sqrt{x})$, where χ is any positive increasing function on $(0, \infty)$ such that $\lim_{x\uparrow\infty}\chi(x) = \infty$. Such an asymptotic formula is actually equivalent to the same formula with $\chi(x) \equiv 1$, see [12, Remark 17]). In uncorrelated volatility models the smile is symmetric, see [26], and therefore the same expansion holds for the implied volatility when x tends to $-\infty$.

2. In the Heston model, the stock price is the unique strong solution to $\mathrm{d}S_t = S_t\sqrt{V_t}\mathrm{d}W_t$ with $S_0 > 0$, $\mathrm{d}V_t = \kappa(\theta - V_t)\mathrm{d}t + \xi\sqrt{V_t}\mathrm{d}B_t$, and $\mathrm{d}\langle W, B\rangle_t = \rho\mathrm{d}t$. Again, S_t is strictly positive almost surely for any $t \geq 0$; in [12, Eq. (4.11)], Friz et al. prove the following expansion in the Heston model as $x \downarrow -\infty$:

$$I_T(x) = \frac{1}{\sqrt{T}}\left(\rho_1\sqrt{|x|} + \rho_2 + \rho_3\frac{\log(|x|)}{\sqrt{|x|}} + \mathcal{O}\left(\frac{1}{\sqrt{|x|}}\right)\right),$$

where the coefficients $\rho_1 \in (0, \sqrt{2}), \rho_2, \rho_3$ are related to the model parameters. Note that although the first two terms in the formula for Heston implied volatility are of the same order than the corresponding terms in (3.9), the third-order term is different, because of the presence of the function $\log|x|$, and tends to zero more slowly.

3. In the uncorrelated Hull-White model (Example 2.1), Gulisashvili and Stein [16, Corollary 3.1] prove

$$I_T(x) = \frac{1}{\sqrt{T}}\left(\sqrt{2|x|} - \frac{\log|x| + \log\log|x|}{2T\xi} + \mathcal{O}(1)\right), \qquad \text{as } x \downarrow -\infty.$$

Note that the constant second-order term appearing in the expansions for the Stein-Stein and the Heston models is replaced here by a term diverging to minus infinity, in agreement with Theorem 2.9.

4 Examples with mass at zero

A distribution μ with a mass $p_T \in (0, 1)$ at zero can be written in terms of its Jordan decomposition:

$$\mu(\mathrm{d}s) = p_T\delta_0(\mathrm{d}s) + (1 - p_T)\mu_p(\mathrm{d}s), \tag{4.1}$$

where μ_p is a probability measure on $(0, \infty)$. The martingale condition $\mathbb{E}[S_T] = \int_{[0,\infty)} s\mu(\mathrm{d}s) = S_0$ imposes $p_T \neq 1$ together with the constraint $\int_{(0,\infty)} s\mu_p(\mathrm{d}s) = S_0/(1-p_T)$. In order to numerically illustrate Theorem 3.7, we compute and plot the function $J_T(x) \equiv I_T(x)\sqrt{T/|x|}$, which must tend to $\sqrt{2}$ as $x \downarrow -\infty$. We compare J_T to its second- and third-order expansions given by Theorem 2.9 and Theorem 3.7:

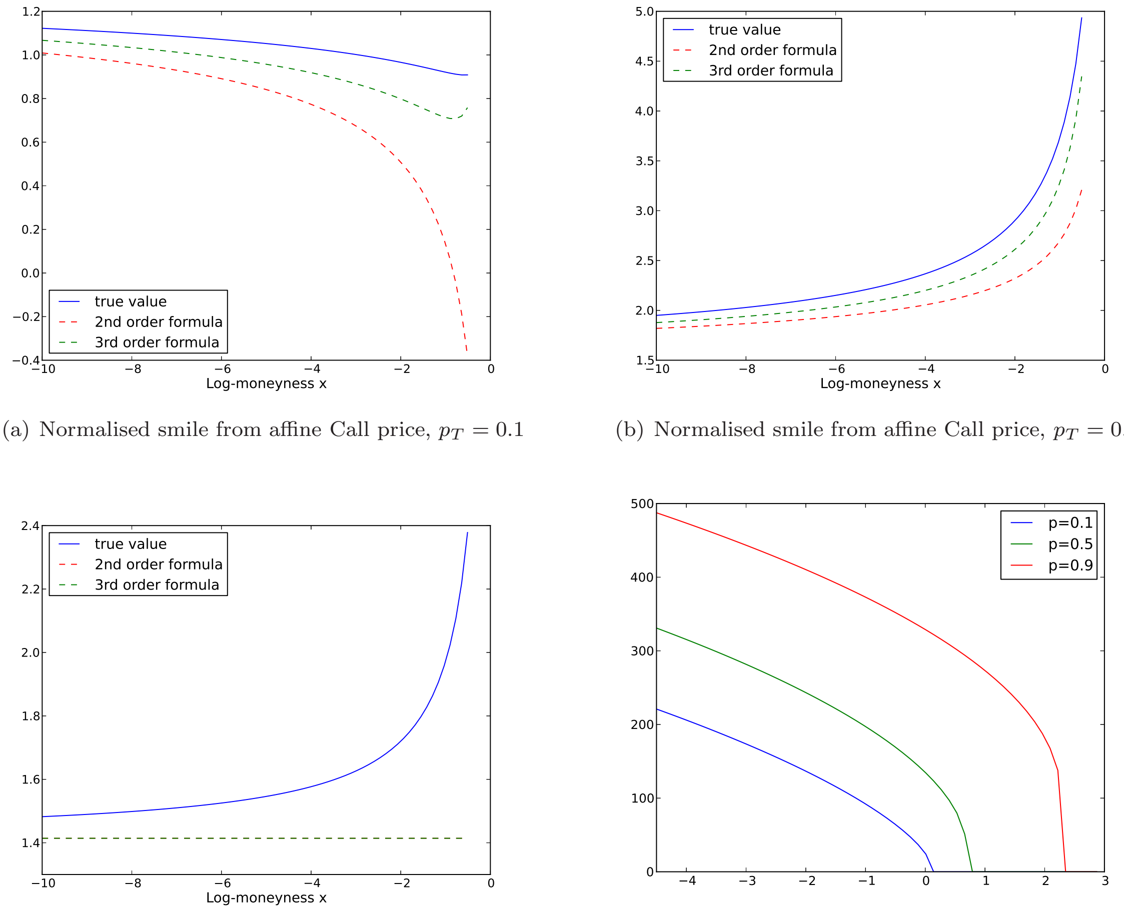
$$J_T^{(2)}(x) \equiv \sqrt{\frac{T}{|x|}}\left(\sqrt{\frac{2|x|}{T}} + \frac{1}{T}N^{-1}(p_T)\right) \qquad \text{and} \qquad J_T^{(3)}(x) \equiv J_T^{(2)}(x) + \sqrt{\frac{T}{|x|}}\frac{N^{-1}(p_T)^2}{2\sqrt{2T|x|}}. \tag{4.2}$$

4.1 A toy example

We define a piecewise affine Call price on $\mathbb{R}_+$ by setting

$$\widetilde{C}(K) = (S_0 - (1 - p_T)K)^+, \qquad K \geq 0. \tag{4.3}$$

The corresponding asset price distribution has the form (4.1), with $\mu_p(ds) = \delta_{S_0/(1-p_T)}(\mathrm{d}s)$. The cumulative distribution of μ, $F(K) = \mu([0, K]) = p_T + (1 - p_T)\mathbb{1}_{\{K \geq S_0/(1-p_T)\}}$, is constant for $K < S_0/(1 - p_T)$, hence condition (1.4) is trivially satisfied. Figure 1 shows some numerical results for $T = 1$.



(a) Normalised smile from affine Call price, $p_T = 0.1$

(b) Normalised smile from affine Call price, $p_T = 0.9$

(c) Normalised smile from affine Call price, $p_T = 0.5$

(d) Implied volatility smiles

Figure 1: Implied volatilities generated by the affine Call price in Example 4.1, $T = 1$. Figures (a), (b), (c): "normalised smile" J_T versus its approximations $J_T^{(2)}$ $J_T^{(3)}$, see (4.2). Figure (d): the corresponding implied volatility smiles. Note that $D^- I_T$ diverges at the upper bound $\overline{x}_T$ of the support of the law.

4.2 Jump-to-default models

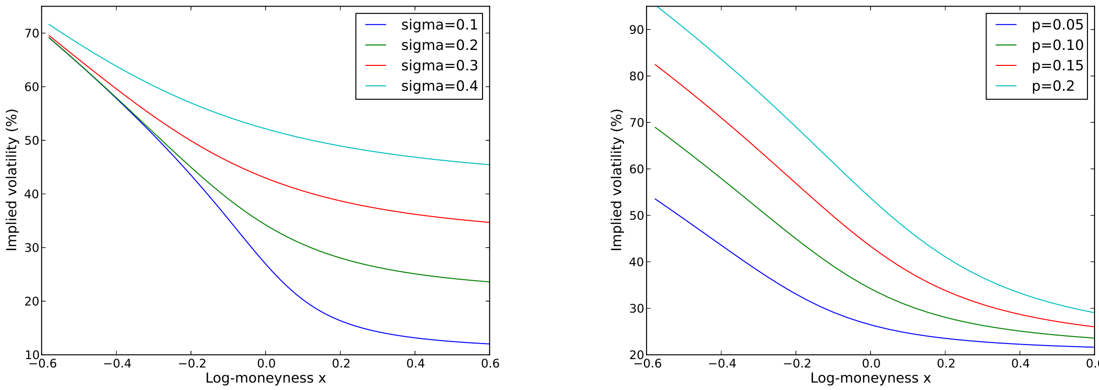
Let $(\tilde{S}_t)_{t\geq 0}$ be a strictly positive process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and τ a random time, independent of $\tilde{S}$. Set

$$S_t = \tilde{S}_t \mathbf{1}_{t<\tau}. \tag{4.4}$$

The process S jumps to zero at time τ. The fixed-time law of S_T has the form (4.1), where $p_T = \mathbb{P}(\tau \leq T)$, and μ_p is the law of $\tilde{S}_T$.

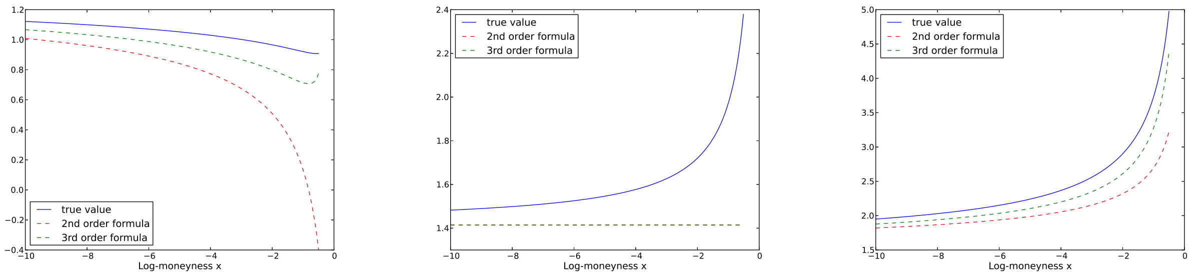
4.2.1 Merton's model with jump-to-default

In the Merton model [24], the process $\tilde{S}$ is a geometric Brownian motion with drift $\lambda > 0$, $\mathrm{d}\tilde{S}_t = \tilde{S}_t(\lambda \mathrm{d}t + \sigma \mathrm{d}W_t)$ with $\tilde{S}_0 > 0$, and τ is exponentially distributed with parameter λ, so that $p_T = \mathbb{P}(\tau \leq T) = 1 - e^{-\lambda T}$. Note that $\mathbb{E}[S_T] = \mathbb{E}[\tilde{S}_T \mathbb{1}_{\tau>T}] = \mathbb{E}[\tilde{S}_T]\mathbb{P}(\tau > T) = S_0 \mathrm{e}^{\lambda T} \mathrm{e}^{-\lambda T} = S_0$. The continuous part of the distribution of S is $\underline{\mu_p}(\mathrm{d}s) = f_{\mathrm{BS}}(s, T; S_0/(1-p_T), \sigma)\mathrm{d}s$, where $f_{\mathrm{BS}}(\cdot, T; \overline{S}, \sigma)$ is the density of a Black-Scholes stock price with mean $\overline{S}$ and volatility $\sigma > 0$. The Put price written on S reads $P(K,T) = p_T K + (1-p_T)P_{\mathrm{BS}}(K, T; S_0/(1-p_T), \sigma)$,



(a) Merton's model with mass p_T at zero, $p_T = 0.1$ (b) Merton's model with mass p_T at zero, different p_T's

Figure 2: Implied volatility smiles in the Merton's model, or Black-Scholes distribution with mass p_T at zero, with $S_0 = 1$, $T = 1$. Figure (a): $p_T = 0.1$, different values of σ. Figure (b): $\sigma = 0.2$, different values of p_T.



(a) Normalised smile, $p_T = 0.1$ (b) Normalised smile, $p_T = 0.5$ (c) Normalised smile, $p_T = 0.9$

Figure 3: Normalised implied volatility smiles in the Merton's model with $\sigma = 0.2$, $T = 1$ and mass p_T at zero. Comparison of the function J_T with $J_T^{(2)}$ and $J_T^{(3)}$, see (4.2).

and the cumulative distribution

$$F(K_x) = p_T + (1 - p_T)N\left(-d_2\left(K_x, T, \frac{S_0}{1 - p_T}, \sigma\right)\right) =: p_T + (1 - p_T)N(d_{2,p_T}(x, T, \sigma)).$$

Since $d_{2,p_T}(x, T, \sigma) \sim \frac{x}{\sigma\sqrt{T}}$ as $x \downarrow -\infty$ and $F(K_x) - F(0) = (1 - p_T)N(d_{2,p_T}(x, T, \sigma))$, the bound (A.1) yields

$$F(K_x) - F(0) \leq \frac{1 - p_T}{d_{2,p_T}(x, T, \sigma)}\phi(d_{2,p_T}(x, T, \sigma)) \leq \frac{1 - p_T}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}d_{2,p_T}(x, T, \sigma)^2\right) = \mathcal{O}\left(\exp\left(-\frac{x^2}{2\sigma^2 T}\right)\right),$$

and hence condition (1.4) is satisfied. We illustrate the validity of the expansion (3.9) in Figures 2 and 3.

Let us briefly comment on Figures 1-3 related to the toy example 4.1 and the Black-Scholes example above:

(i) Interestingly, a log-normal distribution with a constant volatility parameter σ and a mass p_T at zero produces a very pronounced skew, even for small values of p_T, see Figure 2. In analogy with displacement (see [5, Example 6.7]), this is a way of generating an implied volatility smile using only two parameters.

(ii) The different behaviours of the implied volatility foreseen by Theorem 3.7 for $p_T \approx 0$ or $p_T \approx 1$ and for $p_T = 1/2$ are confirmed in these examples (see also the CEV model in section 4.3.1). When $p_T = 1/2$, the convergence of the normalised smile $I_T(x)\sqrt{T/|x|}$ to its limit $\sqrt{2}$ has a considerably smaller bias than the case where p_T is close to 1 or zero, for which the limiting value $\sqrt{2}$ is still far in the left tail (see Figures 1(a)-1(b) and 3(a)-3(c)).

(iii) The graphics 1(a), 1(c), 1(b) and 3(a), 3(b) and 3(c) in Figures 1 and 3 are almost identical. This provides evidence of the fact that the behaviour of the implied volatility for small strike is essentially determined by the mass of the atom at zero, while the remainder of the distribution on $(0,\infty)$ has little impact.

4.2.2 Default probabilities from implied volatilities

We consider here the same model as in the previous subsection 4.2.2. Ohsaki et al. [25] study the possibility of measuring default probabilities from observed implied volatilities. Considering a firm's asset following Merton's or CreditGrades [10] model, they estimate the survival probability at time T based on the asymptotic formula $\lim_{x\downarrow-\infty} d_2(x,T) = -N^{-1}(p_T)$, and compute $d_2(x,T)$ from simulated smile data. They give evidence of the difficulty of achieving a good estimate, due to the slow speed of convergence of d_2 to its limit. For example, for a survival probability around 90%, the estimated value under the Merton model [25, Table 5] is affected by a relative error around 10%, even at very low strikes.

In Proposition 3.5, we account for the error term affecting this estimate, which is roughly $\mathcal{O}(|x|^{-1/2})$. Note however that Theorem 3.7 provides an alternative way of estimating default probabilities, which can be compared to the methodology in Ohsaki et al. [25]. Inverting the third-order formula (3.9) with respect to $-N^{-1}(p_T) =: n$ yields the equation $a(x)n^2 - n - c(x) = 0$, with $a(x) \equiv 1/(2\sqrt{2|x|})$ and $c(x) \equiv \sqrt{T}I_T(x) - \sqrt{2|x|}$. Since $1 + 4a(x)c(x) = \sqrt{\frac{2}{|x|}}(\sqrt{T}I_T(x) - \sqrt{2|x|})$ tends to zero as $x \downarrow -\infty$, Equation (3.9) has the two roots

$$n_\pm(x) = \sqrt{2|x|} \pm \sqrt{2|x| + 2\sqrt{2|x|}\left(\sqrt{T}I_T(x) - \sqrt{2|x|}\right)} \tag{4.5}$$

as soon as x is small enough. It is easy to see that $n_-(x)$ converges to $N^{-1}(1-p_T)$ while $n_+(x)$ diverges to infinity as $x \downarrow -\infty$, and hence $N(n_-(x))$ is a convergent estimator of the survival probability $1-p_T$, independent of any parametric modelling choice. Table 1 shows some numerics in the Merton model with the following parameters:

$$S_0 = 100, \quad T = 0.5, \quad \sigma = 0.3, \quad \lambda = 0.15. \tag{4.6}$$

The third row from the bottom shows the survival probability estimated from the asymptotics $\lim_{x\downarrow-\infty} d_2(x,T) = -N^{-1}(p_T)$, and provides the same values given in [25, Table 5]. The second row from the bottom estimates the survival probability from the second-order formula $I_T(x) = \sqrt{2|x|/T} + N^{-1}(p_T)/\sqrt{T}$, while the last row, based on the full third-order formula (3.9), gives the values of $N(n_-(x))$ with n_- defined in (4.5). The right column contains the exact survival probability $1 - p_T = \mathrm{e}^{-\lambda T}$. Note that, while the estimate based on second-order implied volatilities is less precise than the one based on the coefficient $d_2(x,T)$, the third-order formula is more accurate (the accuracy becoming comparable with the result of Ohsaki et al. at very low strikes). Although improved, the quality of the fit is yet unsatisfactory for the model-free formula (4.5) to be efficiently used to estimate default probabilities within the range of strikes usually available in stock markets.

When $\lambda = 0.85$, and $1 - p_T \approx 65.37\%$ (smaller default probabilities are more realistic in practice), the third-order formula (4.5) is very close to the result by Ohsaki et al. [25, Table 6], and we therefore omit it. This similarity is consistent with the fact that the derivative of the implied volatility with respect to the mass of the atom is small (hence the implied volatility is less sensitive to it) when the latter is close to 1/2.

4.3 Diffusion processes absorbed at zero

4.3.1 The CEV process

We consider here the CEV model, namely the unique strong solution to the stochastic differential equation

$$\mathrm{d}S_t = \sigma S_t^{1+\beta}\mathrm{d}W_t, \tag{4.7}$$

The process $(S_t)_{t\geq 0}$ is a true martingale if and only if $\beta \leq 0$, see [6, Chapter 6.4]. When $\beta = 0$, the SDE (4.7) reduces to the Black-Scholes SDE, and the stock price remains strictly positive almost surely for all $t \geq 0$.

Moneyness	0.5	0.4	0.3	0.2	0.1	1e-10	0
Log-moneyness x	-0.69	-0.91	-1.20	-1.61	-2.30	-23.02	$-\infty$
Survival Probability (%)							
Ohsaki et al.	75.59	77.77	79.72	81.59	83.59	90.35	92.77
Second-order IV	69.30	71.35	73.33	75.39	77.80	88.06	92.77
Third-order IV	76.77	78.81	80.59	82.27	84.07	90.37	92.77

Table 1: Survival probabilities in Merton's jump-to-default model with the parameters (4.6).

Following Chesney et al. [6] we define a new process X by $X_t := S_t^{-2\beta}/(\sigma^2\beta^2)$ up to the first time S hits zero. Itô formula yields $\mathrm{d}X_t = \delta\mathrm{d}t + 2\sqrt{X_t}\mathrm{d}W_t$, with $X_0 = S_0^{-2\beta}/(\sigma^2\beta^2) > 0$ and $\delta = 2 + 1/\beta$. The process X is a Bessel process with δ degrees of freedom (and index $\nu := \delta/2 - 1 = 1/(2\beta)$). The Feller classification (see for example Karlin et al. [20, Chapter 15, Section 6] yields the following:

- if $\delta \leq 0$, i.e. $\beta \in [-1/2, 0)$, the origin is an attainable and absorbing boundary. For every $t > 0$, the distribution μ_t of X_t on $[0,\infty)$ has a positive mass at zero and admits a density on the positive real line:

$$\mu_t(\mathrm{d}y) = \mathbb{P}(X_t = 0)\delta_0(\mathrm{d}y) + f_t(X_0, y)\mathrm{d}y,$$

  with

$$f_t(X_0, y) = \frac{1}{2t}\left(\frac{y}{X_0}\right)^{\nu/2} \exp\left(-\frac{X_0 + y}{2t}\right) I_{-\nu}\left(\frac{\sqrt{X_0 y}}{t}\right), \qquad \text{for all } y > 0,$$

  where $I_{-\nu}$ is the modified Bessel function of the first kind. Note that $\int_0^\infty f_t(X_0, y)\mathrm{d}y = \Gamma\left(-\nu, \frac{X_0}{2t}\right) < 1$, where Γ is the normalised lower incomplete Gamma function $\Gamma(v, z) := \frac{1}{\Gamma(v)}\int_0^z u^{v-1}\mathrm{e}^{-u}\mathrm{d}u$, therefore

$$\mathbb{P}(X_t = 0) = 1 - \Gamma\left(-\nu, X_0/(2t)\right) > 0.$$

- If $\delta \in (0, 2)$ ($\beta < -1/2$), the origin is attainable, but reflecting. If $\delta > 2$ ($\beta > 0$), the origin is not attainable. In both cases, $\mathbb{P}(X_t = 0) = 0$ for all t.

We can recast these results in terms of the original CEV process S, which hits zero if and only if the process X does. In the case $\beta \in [-1/2, 0)$, the density of S_T on the positive real line is given by

$$f_{S_T}(s) = -\frac{s_0^{1/2} s^{-2\beta-3/2}}{\sigma^2\beta T} \exp\left(-\frac{s_0^{-2\beta} + s^{-2\beta}}{2\sigma^2\beta^2 T}\right) I_{-\nu}\left(\frac{s_0^{-\beta} s^{-\beta}}{\sigma^2\beta^2 T}\right),$$

for any $s > 0$, and we further have $\mathbb{P}(S_T = 0) = 1 - \Gamma\left(-\nu, (2\sigma^2\beta^2 T s_0^{2\beta})^{-1}\right)$. Using the asymptotic form (see [1, Section 9.6.7]) for the modified Bessel function $I_\alpha(z) \sim \Gamma(\alpha+1)^{-1}(z/2)^\alpha$ (as $z \downarrow 0$) for positive α, together with $-\nu = 1/(2|\beta|)$, one obtains $f_{S_T}(s) \sim \mathit{const} \times s^{2|\beta|-1}$ as $s \downarrow 0$. Therefore the density of the stock price explodes at the origin when $\beta \in (-1/2, 0)$, and tends to a constant when $\beta = -1/2$, in contrast to the previous examples (where the density vanishes at the origin). As pointed out in Remark 1, the condition (1.4) on the cumulative distribution is satisfied since $2|\beta| - 1 > -1$. This CEV model can further be enhanced with an additional non-predictable independent jump-to-default, as done in [4]. This would result in augmenting the mass at zero and reducing the one on $(0, \infty)$, without affecting the shape of the density.

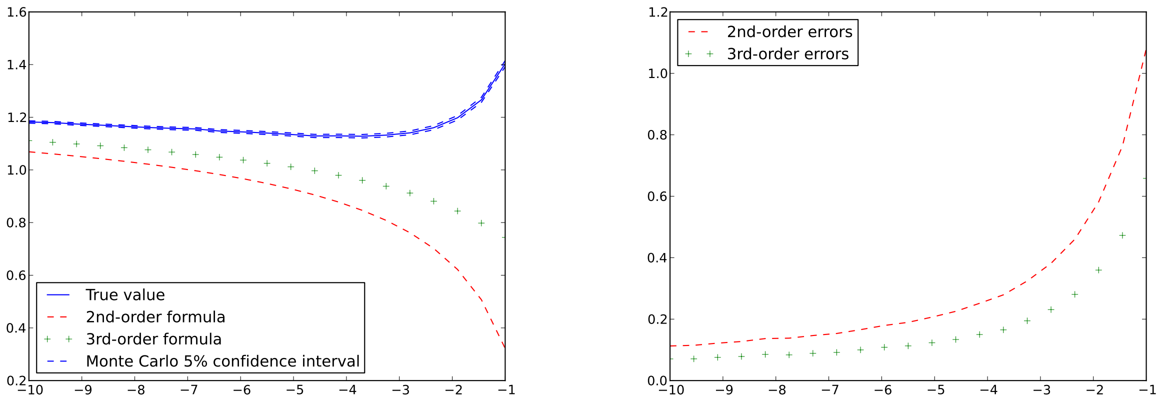
European Put option prices maturing at time $T \geq 0$ and with strike $K \geq 0$ are worth at inception

$$P(K, T) = \mathbb{E}[(K - S_T)_+] = K\mathbb{P}(S_T = 0) + \int_{(0,+\infty)} (K - s)^+ f_{S_T}(s)\mathrm{d}s. \tag{4.8}$$

We compute in Figure 4(a) such option prices and the corresponding implied volatilities in the CEV model with the following parameters:

$$s_0 = 0.1, \qquad T = 5.2, \qquad \beta = -0.4, \qquad \sigma = 0.2.$$

In this case, the mass at the origin is approximately equal to $\mathbb{P}(S_T = 0) = 0.137$. The prices are obtained with Monte Carlo simulations with 50000 paths, each drawn with 100 time steps, and the second- and third-order asymptotics are computed according to Theorem 3.7. In the Monte Carlo simulations, whenever the stock price hits the origin, it remains there since the latter is absorbing. Note that we evaluate deep out-of-the-money Puts with 5×10^4 paths, and get tight 5% confidence intervals. The accuracy of this Monte Carlo estimate is due to the fact that the simulation gives a good (small-variance) estimate of the probability of the (non-rare) event $S_T = 0$, therefore providing a good approximation of the Put option price via (4.8). The contribution of the integral term, typically subject to less accurate Monte Carlo estimates, is asymptotically smaller than the linear term for small K; its precise impact on the implied volatility is quantified by Theorem 3.7.



(a) Normalised implied volatility smile, CEV model

(b) Errors

Figure 4: Implied volatility smiles generated by the CEV model. Comparison of the normalised smile J_T with its second- and third-order approximations $J_T^{(2)}$ and $J_T^{(3)}$, see (4.2).

4.3.2 Absorbed Ornstein-Uhlenbeck process

Another example of continuous asset price dynamics that accumulates mass at zero and allows for explicit formulae for the fixed-time distribution can be built from Ornstein-Uhlenbeck (OU) processes, namely the unique strong solutions to the SDEs $\mathrm{d}\tilde{S}_t = -k\tilde{S}_t \mathrm{d}t + \sigma \mathrm{d}W_t$, with $\tilde{S}_0 = s_0 > 0$ and $k, \sigma > 0$. Then $\tilde{S}_t = s_0 \exp(-kt) + \sigma \int_0^t \exp(-k(t-u))\mathrm{d}W_u$ is a Gaussian process with mean $\mathbb{E}(\tilde{S}_t) = s_0 \exp(-kt)$ and covariance function $\mathrm{Cov}(\tilde{S}_t, \tilde{S}_s) = \frac{\sigma^2}{k} \exp(-kt) \sinh(ks)$. The origin is attainable, and we define S as the process $\tilde{S}$ stopped at the first time it hits zero: let $\tau_0 := \inf\{t \geq 0 : \tilde{S}_t = 0\}$, then $S_t := \tilde{S}_t \mathbb{1}_{\{t < \tau_0\}}$. For every $t > 0$, the law of S_t has the form $\mu_t(\mathrm{d}y) = \mathbb{P}(S_t = 0)\delta_0(\mathrm{d}y) + f_t(s_0, y)\mathrm{d}y$; from Borodin and Salminen [3], we have

$$\mathbb{P}(S_t = 0) = \mathbb{P}(\tau_0 \leq t) = \mathrm{Erfc}\left(\frac{s_0}{\sigma\sqrt{2(e^{2kt} - 1)}}\right), \quad \text{with } \mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-v^2} dv \tag{4.9}$$

and

$$f_t(s_0, y) = \mathbb{P}\Big(\tilde{S}_t \in dy, \min_{0 \leq s \leq t} \tilde{S}_s > 0\Big) = \frac{2e^{\frac{-y^2 - s_0^2 e^{-2kt}}{2\sigma^2(1 - e^{-2kt})}}}{\sqrt{2\pi\sigma^2(1 - e^{-2kt})}} \sinh(2ys_0 e^{-kt}), \qquad \text{for all } y > 0. \tag{4.10}$$

Note that $f_t(s_0, y) \sim c_t\, y$ as $y \downarrow 0$, so that $F(K) - F(0) \sim c_t K^2$ as $K \downarrow 0$, and Condition (1.4) is satisfied. Using (4.9) and (4.10), the numerical evaluation of European options is straightforward from numerical integration of μ_t or from Monte-Carlo simulation of OU paths; for small log-moneyness, the shape of the implied volatility smile (close, in the limit $x \downarrow -\infty$, to the smile of CEV and the jump-to-default models above) is again described by Theorem 3.7.

4.4 Some examples of smile parametrisations

It is interesting to note that most of the recent literature on implied volatility parameterisations seem to ignore the possibility of having a mass at the origin. Consider indeed the following parameterisation, proposed in Guo et al. [18], for the implied total variance smile $w(x,T) \equiv I_T^2(x)T$:

$$w(x,T) = \theta_T \Psi(x\xi(\theta_T)), \qquad \text{for all } (x,T) \in \mathbb{R} \times [0,\infty),$$

where

$$\xi(u) \equiv \alpha \frac{1-\mathrm{e}^{-u}}{u} \mathbb{1}_{\{u>0\}} + \mathbb{1}_{\{u=0\}} \qquad \text{and} \qquad \Psi(z) \equiv |z| + \frac{1}{2}\left(1+\sqrt{1+|z|}\right).$$

where α is a strictly positive real number, and where we let $\theta_T := \sigma^2 T$ for some $\sigma > 0$. Since the expansion $w(x,T) = \alpha|x|(1-\mathrm{e}^{-\sigma^2 T}) + \frac{1}{2}\sigma^2 T + \mathcal{O}(|x|^{-1/2})$ holds as x tends to $-\infty$, the asymptotic slope of the map $x \mapsto w(x,t)$ as $x \downarrow -\infty$ is equal to 2 if and only if $\alpha(1-\mathrm{e}^{-\sigma^2 T}) = 2$. Note further that

$$\lim_{x\downarrow-\infty} d_2\left(x, \sqrt{w(x,T)/T}\right) = -\mathrm{sgn}\left(f_T(\alpha)\right)\infty \qquad \text{and} \qquad \lim_{x\uparrow+\infty} d_1\left(x, \sqrt{w(x,T)/T}\right) = \mathrm{sgn}\left(f_T(\alpha)\right)\infty,$$

where $f_T(\alpha) \equiv \alpha(\mathrm{e}^{\sigma^2 T} - 1) - 2\mathrm{e}^{\sigma^2 T}$. For any $T \geq 0$, the function f_T is strictly increasing and is equal to zero at $\alpha^* = 2/(1-\mathrm{e}^{-\sigma^2 T})$. Therefore, whenever $\alpha < \alpha^*$, $\lim_{x\downarrow-\infty} d_2 = +\infty$ and there is no mass at zero. At the same time, in order for call prices to decrease to zero when the strike tends to infinity, we need $\lim_{x\uparrow+\infty} d_+(x, \sqrt{Tw(x,T)}) = -\infty$ (see [18]), which therefore rules out the possibility of a slope equal to 2.

5 Put-Call and Smile Symmetries

The function

$$G(K,T) = \frac{K}{S_0} P\left(\frac{S_0^2}{K}, T\right), \qquad K > 0, \tag{5.1}$$

allows to define a Black-Scholes implied volatility function $I_{G,T}$, when G is taken as a Call price with maturity T. The identity

$$I_{C,T}(K) = I_{G,T}\left(\frac{S_0^2}{K}\right), \tag{5.2}$$

is proven and used in [14, 15] to transfer the asymptotic results initially formulated for the right part of the implied volatility smile ($K \uparrow \infty$) to the left part ($K \downarrow 0$).

Proposition 5.1. *When $\mathbb{P}(S_T = 0) > 0$, the function $K \mapsto G(K,T)$ defined in (5.1) is not a call price function.*

Proof. Assume $G(\cdot,T)$ is a call price function with maturity T, then $G(K,T) = \mathbb{E}(X-K)^+$ for some integrable random variable X. Lemma 2.3 implies

$$\lim_{K\uparrow\infty} G(K,T) = \lim_{K\uparrow\infty} \frac{K}{S_0} P\left(\frac{S_0^2}{K}, T\right) = \lim_{K'\downarrow 0} \frac{S_0}{K'} P\left(K', T\right) = S_0 \mathbb{P}(S_T = 0) > 0,$$

which contradicts $\lim_{K\uparrow\infty} G(K,T) = \lim_{K\uparrow\infty} \mathbb{E}(X-K)^+ = 0$ by dominated convergence. ■

The situation where $G(\cdot,T)$ is a genuine Call price function, and moreover $G(\cdot,T) \equiv C(\cdot,T)$, is related to a symmetry of the underlying law. Denote by $\mathbb{Q}$ the probability measure defined by the Radon-Nikodym density $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = S_T/S_0$. The distribution of S_T is said to be geometrically symmetric if the distribution of S_0/S_T under the measure $\mathbb{Q}$ is the same as the distribution of S_T/S_0 under $\mathbb{P}$ (see Carr and Lee [5]). Examples include the log-normal distribution and uncorrelated stochastic volatility models (with zero risk-free rate). It is easy to see [5, Theorem 2.2 and Corollary 2.5] that geometric symmetry implies (and indeed is equivalent to) the Put-Call price symmetry

$$C(K,T) = G(K,T) \tag{5.3}$$

with G defined in (5.1). Note that (5.3) can be also written in the more "symmetric" fashion $P(K,T;S_0) = C(S_0,T;K)$, making the spot price appear explicitly. Equation (5.2) shows that Put-Call symmetry is in turn equivalent to the symmetry of the implied volatility smile with respect to the log-moneyness

$$I_T(x) = I_T(-x), \qquad \text{for all } x \in \mathbb{R}. \tag{5.4}$$

The equivalence of (5.4) and (5.3) gives the following corollary to Proposition 5.1:

Corollary 5.2. *If $\mathbb{P}(S_T = 0) > 0$, the implied volatility at expiry T cannot be symmetric in the sense of* (5.4).

Remark 5.3. Note that $\mathbb{Q}(S_T > 0) = \mathbb{E}_{\mathbb{P}}\left[(S_T/S_0)\mathbb{1}_{S_T>0}\right] = 1$, therefore S_0/S_T is $\mathbb{Q}$-almost surely well-defined also when the $\mathbb{P}$-distribution of S_T has an atom at zero. However, since $\mathbb{Q}\left(S_0/S_T > 0\right) = 1$, the $\mathbb{Q}$-distribution of S_0/S_T cannot coincide with the $\mathbb{P}$-distribution of S_T/S_0 in this case. This is another way of showing that geometric symmetry, hence the symmetry of the smile, does not hold when the $\mathbb{P}$-distribution has an atom at zero.

Remark 5.4. In [21, Theorem 4.1], under the condition of no mass at zero, Lee proves the identity $I^{\mathbb{P}}(x) = I^{\mathbb{Q}}(-x)$, where $I^{\mathbb{Q}}$ denotes the implied volatility of options written on S_0/S_T and priced under the measure $\mathbb{Q}$. Although both the functions $I^{\mathbb{P}}$ and $I^{\mathbb{Q}}$ are well-defined for any stock price distribution that is non-negative under $\mathbb{P}$, the same argument used in the proof of Proposition 5.1 shows that the identity $I^{\mathbb{P}}(x) = I^{\mathbb{Q}}(-x)$ does not hold when $\mathbb{P}(S_T = 0) > 0$.

5.1 Restricted symmetry

The property of geometric symmetry can hold for the distribution of the stock price restricted to $(0,\infty)$. This property translates into the symmetry of a modified implied volatility function that takes into account the possible mass at zero, as we now show. Recall that $p_T = \mathbb{P}(S_T = 0)$ and denote $P^p_{\text{BS}}(K,T;S_0,\sigma)$ the Put option price generated by the Black-Scholes distribution with mass at zero, with spot price S_0 and mass of the atom $p_T \in [0,1)$. The equation

$$P(K,T) = p_T K + \mathbb{E}[(K-S_T)^+ \mathbb{1}_{\{S_T>0\}}] = P^p_{\text{BS}}\left(K,T;S_0,\overline{I}^p_T(K)\right) = p_T K + (1-p_T) P_{\text{BS}}\left(K,T;\frac{S_0}{1-p_T},\overline{I}^p_T(K)\right)$$

uniquely defines a function $K \mapsto \overline{I}^p_T(K)$ for all positive K, since it is equivalent to

$$P_{\text{BS}}\left(K,T;\frac{S_0}{1-p_T},\overline{I}^p_T(K)\right) = \frac{\mathbb{E}[(K-S_T)^+\mathbb{1}_{\{S_T>0\}}]}{1-p_T} = \mathbb{E}[(K-S_T)^+|S_T>0],$$

where the right-hand side satisfies the arbitrage bounds $\left(K - \frac{S_0}{1-p_T}\right)^+ \le \mathbb{E}[(K-S_T)^+|S_T>0] \le K$. Let us set

$$I^p_T(x) := \overline{I}^p_T\left(\frac{S_0 \mathrm{e}^x}{1-p_T}\right), \qquad \text{for all } x \in \mathbb{R}. \tag{5.5}$$

I^p_T is the implied volatility generated by the distribution of the stock price on the strictly positive real line, taking into account the rescaling of the spot $\frac{S_0}{1-p_T}$. Geometric symmetry translates into the symmetry of I^p_T:

Proposition 5.5. *The following are equivalent:*

- *The $\mathbb{Q}$-distribution of $\frac{S_0}{(1-p_T)S_T}$ is the same as the $\mathbb{P}$-distribution of $\frac{(1-p_T)S_T}{S_0}$ conditional on $\{S_T>0\}$:*

$$\mathbb{P}\left(\frac{(1-p_T)S_T}{S_0} \in \mathrm{d}s \,\middle|\, S_T > 0\right) = \mathbb{Q}\left(\frac{S_0}{(1-p_T)S_T} \in \mathrm{d}s\right). \tag{5.6}$$

- *The function I^p_T defined in* (5.5) *is symmetric.*

The proof of Proposition 5.5 is provided in Appendix A.2. It would be interesting to find an explicit mapping (if any) relating the implied volatility I_T to the function I_T^p. In the spirit of the symmetry-based proof of the moment formula for small strikes in Lee [21, Theorem 4.3], such a transformation could allow to get back the small-strike asymptotics with mass at zero of Theorem 3.7 from the asymptotics of I_T for large strike, together with the symmetry of I_T^p. We shall get back to this question in future work.

Symmetries in Merton's jump-to-default model. Consider Merton's model from Section 4.2.1. The non-defaultable process $(\tilde{S}_t)_t$ is lognormal with initial value S_0 and drift coefficient λ. Itô's formula shows that the (driftless) power process $(\tilde{S}_t)^L$ with $L = 1 - 2\lambda/\sigma^2$ is lognormal, hence it is geometrically symmetric at every $t \geq 0$. If $L \neq 0$, applying the symmetry to the payoff $G(x) \equiv \varphi(x^{1/L})$, one gets $\mathbb{E}[\varphi(\tilde{S}_T/S_0)] = \mathbb{E}[(\tilde{S}_T/S_0)^L \varphi(S_0/\tilde{S}_T)]$ for any measurable positive φ, and hence

$$\mathbb{E}[\varphi(S_T)] = \varphi(0)p_T + \mathbb{E}\left[\left(\frac{S_T}{S_0}\right)^L \varphi\left(\frac{S_0^2}{S_T}\right)\mathbb{1}_{\{S_T>0\}}\right]. \tag{5.7}$$

which is the symmetry provided for Merton's model in [5, Theorem 7.2].

For fixed $T > 0$, $\tilde{S}_T$ can also be viewed as the T-value of a lognormal process with initial value $S_0 \mathrm{e}^{\lambda T} = \frac{S_0}{1-p_T}$ and no drift (equivalenty: with zero risk-free rate), and the application of the geometric symmetry of the law of $\tilde{S}_T/\left(\frac{S_0}{1-p_T}\right)$ yields the alternative formulation

$$\mathbb{E}[\varphi(S_T)] = \varphi(0)p_T + \mathbb{E}\left[\frac{S_T(1-p_T)}{S_0}\varphi\left(\frac{S_0^2}{S_T(1-p_T)^2}\right)\mathbb{1}_{\{S_T>0\}}\right]. \tag{5.8}$$

It is not clear how one can exploit (5.7) or (5.8) to derive a symmetry for the implied volatility smile I_T. The function I_T^p is symmetric though, from Proposition 5.5; more precisely, $I_T^p \equiv \sigma$ in Merton's model.

5.2 A note on volatility derivatives

For a general strictly positive semimartingale S on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a variance swap with maturity $T > 0$ is a forward contract paying the amount $\frac{1}{T}\sum_{i=1}^n \log(S_{t_i}/S_{t_{i-1}})^2$, for $0 = t_0 < \cdots < t_n = T$ (for some predefined, usually daily, time partition). In the limit as the size of the mesh tends to zero, the above sum converges in probability to the realised quadratic variation $\frac{1}{T}\langle \log S, \log S\rangle_T$. We shall not enter into the details of this here and refer the interested reader to the very thorough paper by Jarrow et al. [19]. It has been showed (see for instance Derman et al. [9]) that such an contract —equivalent to a so-called log-contract, namely a European option maturing at T and paying $\log(S_T/S_0)$— could be fully replicated by trading only in European Call and Put options, meaning that the fair strike of a variance swap is given by

$$\mathbb{E}\Big(\frac{1}{T}\langle \log S, \log S\rangle_T\Big) = \frac{2}{T}\left(\int_0^{S_0}\frac{P(K,T)}{K^2}\mathrm{d}K + \int_{S_0}^{\infty}\frac{C(K,T)}{K^2}\mathrm{d}K\right), \tag{5.9}$$

where P and C denote European Put and Call option prices with maturity T. Should the stock price have a strictly positive mass at the origin, the right-hand side of (5.9) becomes infinite. Indeed, since Lemma 2.3 implies that Put options are linear in strike, then the first integral diverges to infinity. It is not clear who is to blame here: the very (contractual) definition of the variance swap (one cannot take logarithms of S in the construction of the payoff), or the replication formula. We shall leave this discussion for further work. In any case, one should pay attention when applying equation (5.9) with Put prices that are generated by a distribution that has a mass at zero. The arithmetic variance swap—seemingly introduced in [5] and recently studied from a financial point of view in Martin [23]—has recently been mentioned in the literature. For some time partition $0 = t_0 < \cdots < t_n = T$, it pays the amount $\frac{1}{T}\sum_{i=1}^n\left(\frac{S_{t_i}-S_{t_{i-1}}}{S_0}\right)^2$ and can be fully replicated—in the limit as the size of the mesh tends to zero—by the quantity

$$\frac{2}{T}\left(\int_0^{S_0} P(K,T)\mathrm{d}K + \int_{S_0}^{\infty} C(K,T)\mathrm{d}K\right),$$

namely a linear combination (with constant weight) of European Calls and Puts with the same maturity T. The interesting thing here is that this arithmetic variance swap (called 'simple variance swap' in [23]) remains of finite value when the stock price hits the origin. In Lorig et al. [22], yet another approach has been suggested, namely to price the variance swap up to the first hitting time (excluding) of the origin.

As an alternative to variance swaps, some people advocated the use of Gamma swaps, namely contracts paying $\frac{1}{T}\sum_{i=1}^{n} S_{t_i}\log(S_{t_i}/S_{t_{i-1}})^2$ at maturity. Financially speaking, they allow the investor to have a volatility dependence (the Vega) linear in the stock price and not independent of it as in the variance swap. These also admit a model-free replication strategy, namely, as the partition becomes denser and denser in $[0,T]$:

$$\frac{2}{S_0 T}\left(\int_0^{S_0}\frac{P(K,T)}{K}\mathrm{d}K+\int_{S_0}^{\infty}\frac{C(K,T)}{K}\mathrm{d}K\right).$$

Since $\lim_{K\downarrow 0} P(K,T)/K \le 1$ (see Lemma 2.3), the first integral is always finite. The price of the Gamma swap, then, is finite as soon as the positive critical exponent p^* of S_T defined in Example 2.1 is strictly greater than 0, since this implies $C(K,T)=\mathcal{O}(K^{-u})$ for every $u<p^*$. The case $p^*=0$ may give rise to a logarithmic convergence of the Call price, and needs a specific treatment. For example, in the Hull-White model in Example 2.1, one has $C(K,T)\sim(\log K)^{\gamma}\exp(-\alpha\log\log K)$ as $K\uparrow\infty$ for some (known) $\alpha>0$, see [16], and the price of the Gamma swap is still finite.

A Appendix

A.1 Proof of Lemma 3.1

The first partial derivatives of $C_{\mathrm{BS}}(x,T;\sigma)$,

$$\partial_x C_{\mathrm{BS}}(x,T;\sigma)=-\mathrm{e}^x N\big(d_2(x,T,\sigma)\big),\qquad \partial_\sigma C_{\mathrm{BS}}(x,T;\sigma)=\mathrm{e}^x\phi\big(d_2(x,T,\sigma)\big)\sqrt{T},$$

and the well-known bound on Mills' ratio

$$\frac{N(d)}{\phi(d)}\le\frac{1}{|d|},\qquad d<0 \tag{A.1}$$

will be used in the proof. Note that using integration by parts, one can write

$$\int_{[0,K]} s\mathrm{d}F(s)=sF(s)|_0^K-\int_0^K F(s)\mathrm{d}s=KF(K)-\int_0^K F(s)\mathrm{d}s,$$

and hence $K^{-(1+\varepsilon)}\int_{[0,K]} s\mathrm{d}F(s)=K^{-(1+\varepsilon)}\int_0^K(F(K)-F(s))\mathrm{d}s$; therefore Condition (1.4) implies

$$\int_{[0,K]} s\mathrm{d}F(s)=\mathcal{O}(K^{1+\varepsilon}). \tag{A.2}$$

Proof of Lemma 3.1. First note that, from our definition of the implied volatility, $I_T(x)$ is identically zero for all $x\le\underline{x}_T$ and $x\ge\overline{x}_T$ (the Put and Call prices coincide with their payoffs in these regions), therefore the right (resp. left) derivative is identically null for $x<\underline{x}_T$ and $x\ge\overline{x}_T$ (resp. $x\le\underline{x}_T$ and $x>\overline{x}_T$).

Define $\mathcal{I}:\{(x,c)\in\mathbb{R}\times[0,\infty):(1-\mathrm{e}^x)^+\le c<1\}\to[0,\infty)$ by

$$C_{\mathrm{BS}}(x,T;\mathcal{I}(x,c))=c.$$

$\mathcal{I}$ is continuous on $\{(x,c):(1-\mathrm{e}^x)^+\le c<1\}$ and strictly positive and differentiable on $\{(x,c):(1-\mathrm{e}^x)^+<c<1\}$; the first partial derivatives are $\partial_x\mathcal{I}=-\frac{\partial_x C_{\mathrm{BS}}}{\partial_\sigma C_{\mathrm{BS}}}$ and $\partial_c\mathcal{I}=\frac{1}{\partial_\sigma C_{\mathrm{BS}}}$. Since $I_T(x)=\mathcal{I}(x,C(K_x,T)/S_0)$, applying Lemma 2.3 we have (omitting the arguments of C_{BS})

$$\begin{aligned}D^+I_T(x)&=\partial_x\mathcal{I}(x,C)+\partial_c\mathcal{I}(x,C)\frac{\partial_x^+C(K_x,T)}{S_0}=-\frac{\partial_x C_{\mathrm{BS}}}{\partial_\sigma C_{\mathrm{BS}}}-\frac{\mathrm{e}^x}{\partial_\sigma C_{\mathrm{BS}}}\mathbb{P}(S_T>K)\\&=\frac{N(d_2(x,T))-\mathbb{P}(S_T>K)}{\sqrt{T}\phi(d_2(x,T))},\end{aligned}$$

and

$$D^- I_T(x) = \frac{N(d_2(x,T)) - \mathbb{P}(S_T \geq K)}{\sqrt{T}\phi(d_2(x,T))}$$

for all $x < \overline{x}_T$, where, using the notation of (1.9), $d_2(x,T) = d_2(x,T,I_T(x))$. The inequality $D^- I_T(x) \leq D^+ I_T(x)$ is obvious from (3.2) and (3.3).

In order to prove (3.4) and (3.5), note that on the one hand,

$$1 - \mathrm{e}^{-x} + \mathrm{e}^{-x} C_{\mathrm{BS}}(x,T;\sigma) = C_{\mathrm{BS}}(-x,T;\sigma) \tag{A.3}$$

and on the other hand, put-call parity yields

$$1 - \mathrm{e}^{-x} + \mathrm{e}^{-x} \frac{C(K_x,T)}{S_0} = \mathrm{e}^{-x} \frac{P(K_x,T)}{S_0} = \mathbb{E}\Big(1 - \mathrm{e}^{-x}\frac{S_T}{S_0}\Big)^+. \tag{A.4}$$

Equations (A.3) and (A.4) yield the alternative representations $I_T(x) = \mathcal{I}(-x, \mathbb{E}(1-\mathrm{e}^{-x}S_T/S_0)^+)$ for the implied volatility, and for the derivatives:

$$\begin{aligned} D^+ I_T(x) &= \frac{\partial_x C_{\mathrm{BS}}(-x,T,I_T(x)) + \partial_x^+ \mathbb{E}(1-\mathrm{e}^{-x}S_T/S_0)^+}{\partial_\sigma C_{\mathrm{BS}}(-x,T,I_T(x))} \\ &= \frac{-N(\hat{d}_2(-x,T)) + \mathbb{E}[S_T \mathbb{1}_{\{S_T \leq S_0 \mathrm{e}^x\}}]}{\sqrt{T}\phi(\hat{d}_2(-x,T))}; \qquad \text{(A.5)} \\ D^- I_T(x) &= \frac{-N(\hat{d}_2(-x,T)) + \mathbb{E}[S_T \mathbb{1}_{\{S_T < S_0 \mathrm{e}^x\}}]}{\sqrt{T}\phi(\hat{d}_2(-x,T))}, \qquad \text{(A.6)} \end{aligned}$$

where we denote $\hat{d}_2(-x,T) := d_2(-x,T,I_T(x))$. Note that $\hat{d}_2(-x,T) < 0$ for all $x < 0$; moreover, the arithmetic mean-geometric mean inequality implies

$$-\hat{d}_2(-x,T) = \frac{|x|}{I\sqrt{T}} + \frac{I\sqrt{T}}{2} = \frac{2|x|}{2I\sqrt{T}} + \frac{I\sqrt{T}}{2} \geq \sqrt{2|x|}.$$

Now, the identity (A.6) yields

$$D^- I_T(x) \geq \frac{-N(\hat{d}_2(-x,T))}{\sqrt{T}\phi(\hat{d}_2(-x,T))} \geq -\frac{1}{\sqrt{T}|\hat{d}_2(-x,T)|} \geq -\frac{1}{\sqrt{2T|x|}}.$$

On the other hand, it follows from (A.2) that there exist $\alpha > 0$ and $x^* < 0$ such that, for all $x < x^*$,

$$\mathbb{E}\left[S_T \mathbb{1}_{\{S_T < S_0 \mathrm{e}^x\}}\right] = \int_{[0,S_0\mathrm{e}^x]} s\mathrm{d}F(s) \leq \alpha \mathrm{e}^{x(1+\varepsilon)} = \alpha \mathrm{e}^{-|x|(1+\varepsilon)}.$$

From (A.5), the inequality

$$D^+ I_T(x) \leq \frac{\mathbb{E}[S_T \mathbb{1}_{\{S_T < S_0 \mathrm{e}^x\}}]}{\sqrt{T}\phi(\hat{d}_2(-x,T))} \leq \frac{\alpha\sqrt{2\pi}}{\sqrt{T}} \mathrm{e}^{-|x|(1+\varepsilon)} \exp\left(\frac{1}{2}\hat{d}_2(-x,T)^2\right), \tag{A.7}$$

holds for every $x < x^*$. Since

$$\frac{1}{2}\hat{d}_2(-x,T)^2 = \frac{x^2}{2I_T(x)^2 T} + \frac{I_T(x)^2 T}{8} + \frac{|x|}{2},$$

it then follows from $I_T(x)^2 T \sim 2|x|$ as $x \downarrow -\infty$ (see Section 2) that $\hat{d}_2(-x,T)^2/(2|x|)$ converges to 1 as $x \downarrow -\infty$. Therefore, for any $\overline{\varepsilon}$ there exists x^{**} such that $\frac{1}{2}d_2(-x,T)^2 < (1+\overline{\varepsilon})|x|$ for all $x < x^{**}$, and applying this last estimate with $\overline{\varepsilon} = \varepsilon - \varepsilon'$, together with (A.7), the bound (3.5) follows. ■

A.2 Proof of Proposition 5.5

Referring to the decomposition (4.1) for the $\mathbb{P}$-distribution of S_T, one has

$$\mathbb{E}[(K-S_T)^+|S_T>0] = \int_0^\infty (K-s)^+\mu_p(\mathrm{d}s) = P_{\mathrm{BS}}\left(K,T;\frac{S_0}{1-p_T},\overline{I_T}^p(K)\right),$$

which is equivalent to $\overline{I_T}^p(K) = I_T^{\mu_p}(K)$, since $\int_0^\infty s\mu_p(\mathrm{d}s) = S_0/(1-p_T)$. The statement of Proposition 5.5 then follows from the classical smile symmetry (5.4) if one shows that (5.6) is equivalent to the geometric symmetry of the distribution μ_p, which we now prove.

Note that S_T has the $\mathbb{P}$-distribution of $\tilde{S}Z$, where $\tilde{S} \overset{\mathbb{P}}{\sim} \mu_p$ and Z is independent from $\tilde{S}$ with $\mathbb{P}(Z=0) = 1-\mathbb{P}(Z=1) = p_T$. For any measurable and bounded function φ,

$$\begin{aligned}
\mathbb{E}\left[\varphi\left(\frac{S_T(1-p_T)}{S_0}\right)\middle| S_T>0\right] &= \frac{1}{1-p_T}\mathbb{E}\left[\varphi\left(\frac{S_T(1-p_T)}{S_0}\right)\mathbb{1}_{\{Z>0\}}\right] \\
&= \mathbb{E}\left[\varphi\left(\frac{\tilde{S}(1-p_T)}{S_0}\right)\right] \\
&\overset{(sym)}{=} \mathbb{E}\left[\frac{\tilde{S}(1-p_T)}{S_0}\varphi\left(\frac{S_0}{\tilde{S}(1-p_T)}\right)\right] \\
&= \mathbb{E}\left[\frac{S_T}{S_0}\varphi\left(\frac{S_0}{S_T(1-p_T)}\right)\mathbb{1}_{\{Z>0\}}\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\varphi\left(\frac{S_0}{S_T(1-p_T)}\right)\mathbb{1}_{\{Z>0\}}\right] = \mathbb{E}_{\mathbb{Q}}\left[\varphi\left(\frac{S_0}{S_T(1-p_T)}\right)\right],
\end{aligned}$$

where the step $\overset{(sym)}{=}$ holds for any bounded function φ if and only if $\tilde{S}\sim\mu_p$ is geometrically symmetric, and the last step holds since $\mathbb{Q}(Z>0) = \mathbb{Q}(S_T>0) = 1$.

References

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, Tenth ed., 1972.

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

[3] A. Borodin and P. Salminen, *Handbook of Brownian Motion. Facts and Formulae*, Birkhäuser, Basel, 1996.

[4] L. Campi, S. Polbennikov, and A. Sbuelz, *Systematic equity-based credit risk: A CEV model with jump to default*, Journal of Economic Dynamics and Control, 33 (2009), pp. 93–108.

[5] P. Carr and R. Lee, *Put-call symmetry : Extensions and applications*, Mathematical Finance, 19 (2009), pp. 523–560.

[6] M. Chesney, M. Jeanblanc, and M. Yor, *Mathematical methods for financial markets.*, Springer, 2009.

[7] D. Coculescu, H. Geman, and M. Jeanblanc, *Valuation of default sensitive claims under imperfect information*, Finance and Stochastics, 12 (2008), pp. 195–218.

[8] P. Collin-Dufresne and R. Goldstein, *Do credit spreads reflect stationary leverage ratios?*, Journal of Financial, 56 (2001), pp. 1929–1957.

[9] E. Derman, K. Demeter, M. Kamal, and J. Zou, *A guide to volatility and variance swaps.*, Journal of Derivatives, 6 (4) (1999), pp. 9–32.

[10] V. Finkelstein, G. Pan, J. P. Lardy, and J. Tiemey, *Creditgrades Technical Document.* Editor Chirstopher C. Finger, RiskMetrics Group, 2002.

[11] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time, 2nd Edition*, Walter de Gruyter, Berlin, New York, 2004.

[12] P. Friz, S. Gerhold, A. Gulisashvili, and S. Sturm, *On refined volatility smile expansion in the Heston model*, Quantitative Finance, 11 (2011), pp. 1151–1164.

[13] K. Gao and R. Lee, *Asymptotics of implied volatility to arbitrary order.* Forthcoming in Finance and Stochastics. Available at http://ssrn.com/abstract=1768383, 2011.

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

[15] ———, *Asymptotic equivalence in Lee's moment formulas for the implied volatility, asset price models without moment explosions, and Piterbarg's conjecture*, International Journal of Theorical and Applied Finance, 15 (2012), pp. 1–34.

[16] A. Gulisashvili and E. M. Stein, *Implied volatility in the Hull-White model*, Mathematical Finance, 19 (2009), pp. 303–327.

[17] A. Gulisashvili and E. M. Stein, *Asymptotic behavior of the stock price distribution density and implied volatility in stochastic volatility models*, Applied Mathematics and Optimization, 61 (2010), pp. 287–315.

[18] G. Guo, A. Jacquier, C. Martini, and L. Neufcourt, *Generalised arbitrage-free SVI volatility surfaces.* Preprint arXiv:1210.7111, 2012.

[19] R. Jarrow, Y. Kchia, M. Larsson, and P. Protter, *Discretely sampled variance and volatility swaps vs their continuous approximations*, Finance and Stochastics, 17 (2) (2013), pp. 305–324.

[20] S. Karlin and H. Taylor, *A second course in stochastic processes*, Academic Press, 1981.

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

[22] M. Lorig, O. Lozano-Carbasse, and R. Mendoza-Arriaga, *Variance swaps on defaultable assets and market implied time-changes.* arxiv 1209.0697, 2013.

[23] I. Martin, *Simple variance swaps*, NBER Working Paper 16884, (2013).

[24] R. C. Merton, *Option pricing when underlying stock returns are discontinuous*, Journal of Financial Economics, 3 (1976), pp. 125–144.

[25] S. Ohsaki, t. Ozeki, Y. Umezawa, and A. Yamazaki, *A note on the Black–Scholes implied volatility with default risk*, Wilmott Magazine, 2 (June 2010), pp. 155–170.

[26] E. Renault and N. Touzi, *Option hedging and implied volatilities in a stochastic volatility model*, Mathematical Finance, 6 (1996), pp. 279–302.

[27] L. C. G. Rogers and M. R. Tehranchi, *Can the implied volatility surface move by parallel shifts?*, Finance and Stochastics, 14 (2010), pp. 235–248.

[28] M. Roper, *Arbitrage free implied volatility surfaces.* Publications and preprints of the School of Mathematics and Statistics, University of Sydney, 2010.