TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Small-Time Asymptotics, Moment Explosion and The Moderate Deviations Regime

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Associate Prof. Dipl.-Ing. Dr.techn. Stefan Gerhold

Institut für Stochastik und Wirtschaftsmathematik (E105)
Forschungsgruppe Finanz- und Versicherungsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Arpad Pinter, BSc

Matrikelnummer: 0726282
Wiedner Hauptstraße 8-10
A-1040 Wien

Wien, am 18. Mai 2017

Kurzfassung

Das Ziel dieser Dissertation ist das Erlangen eines besseren Verständnisses für finanzmathematische Modelle zur Optionsbepreisung, wobei der Fokus auf stochastischen Volatilitätsmodellen und exponentiellen Lévy-Modellen liegt. In diesen Modellklassen bzw. auch in einzelnen praxisrelevanten Modellen untersuchen wir das asymptotische Verhalten von Callpreisen, impliziten Volatilitäten und dazugehörigen Größen für kurze Restlaufzeiten („*small-time*"), aber auch für extreme Ausübungspreise („*large-strike*") der Optionen. Diese Ergebnisse sind von praktischer Bedeutung zur Modellkalibrierung, zur qualitativen Modellbeurteilung und zur Wahl und Gestaltung der Modellparameter.

Diese Dissertation ist in vier Teile aufgeteilt, die sich mit unterschiedlichen Modellklassen bzw. einzelnen stochastischen Volatilitätsmodellen auseinandersetzen.

Teil I, *Small-Maturity Asymptotics for the At-the-Money Implied Volatility Slope in Lévy Models*, basiert auf der gleichnamigen Publikation [53], einer gemeinsamen Arbeit mit Stefan Gerhold und I. Cetin Gülüm, die im Journal „Applied Mathematical Finance" erschienen ist. Der zusätzliche Appendix B beweist die Resultate im CGMY-Modell ausführlich.

In diesem Teil betrachten wir die at-the-money Steigung der impliziten Volatilität, d.h. die Ableitung der impliziten Volatilität nach dem Ausübungspreis, wenn die Restlaufzeit der dazugehörigen Option gegen 0 konvergiert. Das Hauptresultat quantifiziert das Verhalten der Steigung in exponentiellen Lévy-Modellen mit unendlicher Aktivität, die eine Brownsche Komponente aufweisen. Als technisches Nebenresultat, erhalten wir mit Hilfe der Mellin-Transformation eine asymptotische Entwicklung von at-the-money digitalen Callpreisen für kurze Restlaufzeiten. Abschließend diskutieren wir, in welchen Modellen die at-the-money Steigung Auskunft über das Verhalten der Steilheit der Enden des Volatilitäts-Smiles gibt.

Teil II, *Option Pricing in the Moderate Deviations Regime*, basiert auf der gleichnamigen Publikation [45], einer gemeinsamen Arbeit mit Peter Friz und Stefan Gerhold, die im Journal „Mathematical Finance" erscheinen wird. Im zusätzlichen Appendix B sind weiterführende Resultate angeführt.

Wir betrachten Callpreise für kleine Restlaufzeiten in Diffusionsmodellen, in einem asymptotischen Regime („moderately out-of-the-money"), dass zwischen den beiden gut untersuchten Fällen at-the-money und out-of-the-money interpoliert. Im Sinne der Moderate-Deviation-Theorie erhalten wir Abschätzungen erster und höherer Ordnung

für Callpreise und implizite Volatilitäten. Die entsprechenden asymptotischen Entwicklungen beinhalten nur einfache Ausdrücke der Modellparameter, und wir zeigen, wie diese Ausdrücke für allgemeine lokale und stochastische Volatilitätsmodelle berechnet werden können. Einige numerische Berechnungen für das Heston-Modell illustrieren die Genauigkeit unserer Resultate.

Teil III, *Moment Explosion in the Rough Heston Model*, befasst sich mit einer Erweiterung des bekannten Heston-Modells, bei der die Pfade des Volalitätsprozesses weniger glatt als beim klassischen Heston-Modell sind. Wir sind am Explosionsverhalten der momentenerzeugenden Funktion interessiert, die mit Hilfe der Lösung einer fraktionalen Riccati-Differentialgleichung gegeben ist. Nach der Transformation dieser Differentialgleichung in eine nicht-lineare Volterra-Integralgleichung untersuchen wir die Explosionszeiten der Lösung bzw. der momentenerzeugenden Funktion für alle möglichen Parameterwahlen. Im Falle einer endlischen Explosionszeit geben wir obere und untere Abschätzungen dafür. Unter Verwendung dieser Abschätzungen zeigen wir die Endlichkeit der kritischen Momente im Rough-Heston-Modell.

Teil IV, *Large-Strike Asymptotics in the 3/2-Model*, widmet sich dem 3/2-Modell. Zunächst untersuchen wir die Dichtefunktion, die mittels Fourier-Transformation als ein Wegintegral in der komplexen Ebene dargestellt werden kann. Durch die Wahl eines transformierten Hankel-Weges und der expliziten Form der momentenerzeugenden Funktion ist es möglich, das asymptotische Verhalten der Dichtefunktion am positiven Ende zu bestimmen. Abschließend leiten wir daraus die asymptotische large-strike Entwicklung für die implizite Volatilität ab.

Die beiden Teile III und IV sind in Arbeit und bislang noch nicht publiziert worden.

Abstract

The aim of this doctoral thesis is to get a better understanding of option pricing models, with a focus on stochastic volatility models and exponential Lévy models. In these model classes and in individual praxis-oriented models, we investigate asymptotic behaviour of call option prices, implied volatility and related quantities, close to expiry (“small-time”) and with extreme strike values (“large-strike”). Such results are of practical relevance for model calibration, qualitative model assessment and parametrisation design.

The thesis consists of four parts addressing different model classes and individual stochastic volatility models, respectively.

Part I, *Small-Maturity Asymptotics for the At-the-Money Implied Volatility Slope in Lévy Models*, is based on the eponymous paper [53], a joint work with Stefan Gerhold and I. Cetin Gülüm, published in the journal “Applied Mathematical Finance”. The additional Appendix B proves the results in the CGMY in detail.
In this part, we consider the at-the-money strike derivative of implied volatility as the maturity tends to zero. Our main results quantify the behaviour of the slope for infinite activity exponential Lévy models including a Brownian component. As auxiliary results, we obtain asymptotic expansions of short maturity at-the-money digital call options, using Mellin transform asymptotics. Finally, we discuss when the at-the-money slope is consistent with the steepness of the smile wings, as given by Lee’s moment formula.

Part II, *Option Pricing in the Moderate Deviations Regime*, is based on the eponymous paper [45], a joint work with Peter Friz and Stefan Gerhold, which will be published in the journal “Mathematical Finance”. In the additional Appendix B, further results are stated.
We consider call option prices close to expiry in diffusion models, in an asymptotic regime (“moderately out of the money”) that interpolates between the well-studied cases of at-the-money and out-of-the-money regimes. First and higher order small-time moderate deviation estimates of call prices and implied volatilities are obtained. The expansions involve only simple expressions of the model parameters, and we show how to calculate them for generic local and stochastic volatility models. Some numerical computations for the Heston model illustrate the accuracy of our results.

Part III, *Moment Explosion in the Rough Heston Model*, focuses on an extension of the well-known Heston model, where the paths of the volatility process are rougher than in

the classic Heston model. We are interested in the blow-up behaviour of the moment generating function which is given by means of the solution of a fractional Riccati differential equation. After transforming this differential equation into a non-linear Volterra integral equation, we analyse the explosion time of the solution resp. the moment generating function for any parameter choice. In case of a finite explosion time, we give upper and lower bounds. Eventually, using these estimates, we show the finiteness of the critical moments in the rough Heston model.

Part IV, *Large-Strike Asymptotics in the 3/2-Model*, deals with the 3/2-model. At first, we consider the density function which can be expressed via Fourier transform as a contour integral in the complex plane. Choosing a transformed Hankel-type contour and using the explicitness of the moment generating function, we determine the asymptotic behaviour of the positive tail of the density function. Finally, we derive the large-strike asymptotics for the implied volatility.

Both Parts III and IV are work in progress and unpublished so far.

Acknowledgements

First, I would like to thank my supervisor Stefan Gerhold. First of all, thank you for giving me this unique opportunity to do the doctorate at TU Wien. Thank you for your support, your advice and your guidance throughout the last few years. I am very thankful that you always had time whenever I needed help, and you always gave me interesting, new ideas whenever I was stuck with a problem. Thank you!

I would like to thank Peter Friz for collaboration and helpful discussions during my visit at TU Berlin.

A special thanks goes to my colleagues and friends at FAM with whom I spent memorable, pleasant years.

Last but not least, I gratefully acknowledge financial support from the Austrian Science Fund (FWF) under grant P 24880.

Contents

Part I

Small-Maturity Asymptotics for the At-the-Money Implied Volatility Slope in Lévy Models

Chapter 1

Introduction

Recent years have seen an explosion of the literature on asymptotics of option prices and implied volatilities (see e.g. Andersen and Lipton [4] and Friz, Gerhold, Gulisashvili and Sturm [44] for many references). Such results are of practical relevance for fast model calibration, qualitative model assessment, and parametrization design. The small-time behaviour of the *level* of implied volatility in Lévy models (and generalizations) has been investigated in great detail in Boyarchenko and Levendorskiĭ [12], Figueroa-López and Forde [32], Figueroa-López, Gong and Houdré [33], Figueroa-López, Gong and Houdré [34], Roper [81] and Tankov [87]. We, on the other hand, focus on the at-the-money *slope* of implied volatility, i.e. the strike derivative, and investigate its behaviour as maturity becomes small. For diffusion models, there typically exists a limiting smile as the maturity tends to zero, and the limit slope is just the slope of this limit smile (e.g. for the Heston model, this follows from Section 5 in Durrleman [28]). Our focus is, however, on exponential Lévy models. There is no limit smile here that one could differentiate, as the implied volatility blows up off-the-money, see Tankov [87]. In fact, this is a desirable feature, since in this way Lévy models are better suited to capture the steep short maturity smiles observed in the market. But it also implies that the limiting slope cannot be deduced directly from the behaviour of implied volatility itself, and requires a separate analysis. (Note that a limiting smile does exist if maturity and log-moneyness tend to zero jointly in an appropriate way, see Mijatović and Tankov [70].)

It turns out that the presence of a Brownian component has a decisive influence: Without it, the ATM (at-the-money) slope explodes (under mild conditions). The blowup is of order $T^{-1/2}$ for many models, but may also be slower (e.g. CGMY model with $Y \in (1, 2)$; see Example 2.10). Our main results are on Lévy models *with* a Brownian component, though. We provide a result (Corollary 2.6 in Section 2.4) that translates the asymptotic behaviour of the moment generating function to that of the ATM slope. When applied to concrete models, we see that the slope may converge to a finite limit (Normal Inverse Gaussian, Meixner, CGMY models), or explode at a rate slower than $T^{-1/2}$ (generalized tempered stable model; this kind of behaviour seems to be the most realistic one, see Bayer, Friz and Gatheral [6]). Note that several studies, e.g. Aït-Sahalia [1], Aït-Sahalia and Jacod [2] and Carr and Wu [18], highlight the importance of a Brownian component when fitting to historical data or option prices. In particular, in many pure jump Lévy models ATM implied volatility converges to zero as $T \downarrow 0$ (see Proposition 5 in Tankov [87] for a precise statement), which seems undesirable.

From a practical point of view, the asymptotic slope is a useful ingredientfor model calibration: E.g. if the market slope is negative, then a simple constraint on the model parameters forces the (asymptotic) model slope to be negative, too. Our numerical tests show that the sign of the slope is reliably identified by a first order asymptotic approximation, even if the maturity is not short at all. With our formulas, the asymptotic slope (and, of course, its sign) can be easily determined from the model parameters. For instance, the slope of the NIG (Normal Inverse Gaussian) model is positive if and only if the skewness parameter satisfies $\beta > -\frac{1}{2}$.

To obtain these results, we investigate the asymptotics of at-the-money digital calls; their relation to the implied volatility slope is well known. While, for Lévy processes X, the small-time behaviour of the transition probabilities $\mathbb{P}[X_T \geq x]$ (in finance terms, digital call prices) has been well studied for $x \neq X_0$ (see e.g. Figueroa-López and Houdré [35] and the references therein), not so much is known for $x = X_0$. Still, first order asymptotics of $\mathbb{P}[X_T \geq X_0]$ are available, and this suffices if there is no Brownian component. If the Lévy process has a Brownian component, then it is well known that $\lim_{T\to 0} \mathbb{P}[X_T \geq X_0] = \frac{1}{2}$. In this case, it turns out that the second order term of $\mathbb{P}[X_T \geq X_0]$ is required to obtain slope asymptotics. For this, we use a novel approach involving the Mellin transform (w.r.t. time) of the transition probability (Sections 2.3 and 2.4). We believe that this method is of wide applicability to other problems involving time asymptotics of Lévy processes, and hope to elaborate on it in future work.

Finally, we consider the question whether a positive at-the-money slope requires the right smile wing to be the steeper one, and vice versa. Wing steepness refers to large-strike asymptotics here. It turns out that this is indeed the case for several of the infinite activity models we consider. This results in a qualitative limitation on the smile shape that these models can produce.

One of the few other works dealing with small-time Lévy slope asymptotics is the comprehensive recent paper by Andersen and Lipton [4]. Besides many other problems on various models and asymptotic regimes, they study the small-maturity ATM digital price and volatility slope for the tempered stable model (Propositions 8.4 and 8.5 in Andersen and Lipton [4]). This includes the CGMY model as a special case (see Example 2.10 for details). Their proof method is entirely different from ours, exploiting the explicit form of the characteristic function of the tempered stable model. Using mainly the dominated convergence theorem, they also analyse the convexity. We, on the other hand, assume a certain asymptotic behaviour of the characteristic function, and use its explicit expression only when calculating concrete examples. Our approach covers, e.g., the ATM slope of the generalized tempered stable, NIG, and Meixner models without additional effort.

The recent preprint Figueroa-López and Ólafsson [36] is also closely related to our work. There, the Brownian component is generalized to stochastic volatility. On the other hand, the assumptions on the Lévy measure exclude, e.g., the NIG and Meixner models. Section 2.5 has additional comments on how our results compare to those of Andersen and Lipton [4] and Figueroa-López and Ólafsson [36]. Alòs, León and Vives [3] also study the small time implied volatility slope under stochastic volatility and jumps, but the latter are assumed to have finite activity, which is not our focus. Results on the *large* time slope can be found in Forde, Jacquier and Figueroa-López [40]; see also Gatheral [50], p. 63f.

Chapter 2

Implied Volatility Slope Asymptotics

2.1 Digital Call Prices

We denote the underlying by $S = e^X$, normalized to $S_0 = 1$, and the pricing measure by $\mathbb{P}$. W.l.o.g. the interest rate is set to zero, and so S is a $\mathbb{P}$-martingale. Suppose that the log-underlying $X = (X_t)_{t\geq 0}$ is a Lévy process with characteristic triplet (b, σ^2, ν) and $X_0 = 0$. The moment generating function (mgf) of X_T is

$$M(z,T) = \mathbb{E}[e^{zX_T}] = \exp\left(T\psi(z)\right),$$

where

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + bz + \int_{\mathbb{R}} (e^{zx} - 1 - zx)\,\nu(dx). \tag{2.1}$$

This representation is valid if the Lévy process has a finite first moment, which we of course assume, as even $S_t = e^{X_t}$ should be integrable. If, in addition, X has paths of finite variation, then $\int_{\mathbb{R}} |x|\nu(dx) < \infty$, and

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + b_0 z + \int_{\mathbb{R}} (e^{zx} - 1)\,\nu(dx),$$

where the drift b_0 is defined by

$$b_0 = b - \int_{\mathbb{R}} x\,\nu(dx).$$

The following theorem collects some results about the small-time behaviour of $\mathbb{P}[X_T \geq 0]$. All of them are known, or easily obtained from known results. We are mainly interested in the case where $S = e^X$ is a martingale, and so $\mathbb{P}[X_T \geq 0]$ has the interpretation of an at-the-money digital call price. Still, we mention that this assumption is not necessary for parts (i)-(iv). In part (iv), the following condition from Rosenbaum and Tankov [83] is used:

(**H**-α) The Lévy measure ν has a density $g(x)/|x|^{1+\alpha}$, where g is a non-negative measurable function admitting left and right limits at zero:
$c_+ := \lim_{x\downarrow 0} g(x), \quad c_- := \lim_{x\uparrow 0} g(x), \quad \text{with} \quad c_+ + c_- > 0.$

Theorem 2.1. *Let X be a Lévy process with characteristic triplet (b, σ^2, ν) and $X_0 = 0$.*

(i) If X has finite variation, and $b_0 \neq 0$, then

$$\lim_{T \downarrow 0} \mathbb{P}[X_T \geq 0] = \begin{cases} 1, & b_0 > 0 \\ 0, & b_0 < 0. \end{cases}$$

(ii) If $\sigma > 0$, then $\lim_{T\downarrow 0} \mathbb{P}[X_T \geq 0] = \frac{1}{2}$.

(iii) If X is a Lévy jump diffusion, i.e. it has finite activity jumps and $\sigma > 0$, then

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{b_0}{\sigma\sqrt{2\pi}}\sqrt{T} + \mathcal{O}(T), \quad T \downarrow 0.$$

(iv) Suppose that $\sigma = 0$ and that ($\mathbf{H}$-α) holds for some $\alpha \in [1, 2)$. If $\alpha = 1$, we additionally assume $c_- = c_+ =: c$ and $\int_0^1 x^{-1}|g(x) - g(-x)|dx < \infty$. Then

$$\lim_{T\downarrow 0} \mathbb{P}[X_T \geq 0] = \begin{cases} \frac{1}{2} + \frac{1}{\pi}\arctan\frac{b^*}{\pi c} & \text{if } \alpha = 1, \\ \frac{1}{2} + \frac{\alpha}{\pi}\arctan\left(\beta\tan\left(\frac{\alpha\pi}{2}\right)\right) & \text{if } \alpha \neq 1, \end{cases}$$

where $b^* = b - \int_0^\infty (g(x) - g(-x))/x\, dx$ and $\beta = (c_+ - c_-)/(c_+ + c_-)$.

(v) If e^X is a martingale and the Lévy measure satisfies $\nu(dx) = e^{-x/2}\nu_0(dx)$, where ν_0 is a symmetric measure, then

$$\mathbb{P}[X_T \geq 0] = \Phi(-\sigma_{\text{imp}}(1, T)\sqrt{T}/2),$$

where Φ denotes the standard Gaussian cdf.

Proof. (i) We have $\mathbb{P}[X_T \geq 0] = \mathbb{P}[T^{-1}X_T \geq 0]$, but $T^{-1}X_T$ converges a.s. to b_0, by Theorem 43.20 in Sato [85].
(ii) If $\sigma > 0$, then $T^{-1/2}X_T$ converges in distribution to a centered Gaussian random variable with variance σ^2 (see Sato [85]). For further CLT-type results in this vein, see Doney and Maller [26] and Gerhold, Kleinert, Porkert and Shkolnikov [54].
(iii) Conditioning on the first jump time τ, which has an exponential distribution, we find

$$\begin{aligned} \mathbb{P}[X_T \geq 0] &= \mathbb{P}[X_T \geq 0 | \tau \leq T] \cdot \mathbb{P}[\tau \leq T] + \mathbb{P}[X_T \geq 0 | \tau > T] \cdot \mathbb{P}[\tau > T] \\ &= \mathcal{O}(T) + \mathbb{P}[\sigma W_T + b_0 T \geq 0](1 + \mathcal{O}(T)) \\ &= \mathbb{P}[\sigma W_T + b_0 T \geq 0] + \mathcal{O}(T) \\ &= \Phi(b_0\sqrt{T}/\sigma) + \mathcal{O}(T). \end{aligned} \tag{2.2}$$

Now apply the expansion

$$\Phi(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} + \mathcal{O}(x^3), \quad x \to 0. \tag{2.3}$$

(iv) By Proposition 1 in Rosenbaum and Tankov [83], the rescaled process $X_t^{\varepsilon,\alpha} := \varepsilon^{-1}X_{\varepsilon^\alpha t}$ converges in law to a strictly α-stable process $X_t^{*,\alpha}$ as $\varepsilon \downarrow 0$. Therefore

$$\lim_{T\downarrow 0}\mathbb{P}[X_T \geq 0] = \lim_{\varepsilon\downarrow 0}\mathbb{P}[\varepsilon^{-1}X_{\varepsilon^\alpha} \geq 0] = \mathbb{P}[X_1^{*,\alpha} \geq 0],$$

and it suffices to evaluate the latter probability. For $\alpha = 1$, $X_1^{*,1}$ has a Cauchy distribution with characteristic exponent

$$\log \mathbb{E}[\exp(iuX_1^{*,1})] = ib^*u - \pi c|u|,$$

hence $\mathbb{P}[X_1^{*,1} \geq 0] = \frac{1}{\pi}\arctan\frac{b^*}{\pi c}$. (Our b^* is denoted γ^* in Rosenbaum and Tankov [83].) If $1 < \alpha < 2$, then $X_1^{*,\alpha}$ has a strictly stable distribution with characteristic exponent

$$\log \mathbb{E}[\exp(iuX_1^{*,\alpha})] = -|du|^\alpha\Big(1 - i\beta\,\mathrm{sgn}(u)\tan\Big(\frac{\alpha\pi}{2}\Big)\Big),$$

where

$$d_\pm^\alpha = -\Gamma(-\alpha)\cos\Big(\frac{\alpha\pi}{2}\Big)c_\pm \geq 0, \quad d^\alpha = d_+^\alpha + d_-^\alpha, \quad \beta = \frac{d_+^\alpha - d_-^\alpha}{d^\alpha} \in (-1,1).$$

The desired expression for $\mathbb{P}[X_1^{*,\alpha} \geq 0]$ then follows from Davydov and Ibragimov [20]. See Figueroa-López and Forde [32] for further related references.
(v) Under this assumption, the market model is symmetric in the sense of Fajardo [30] and Fajardo and Mordecki [31]. The statement is Theorem 3.1 in Fajardo [30]. $\square$

The variance gamma model and the CGMY model with $0 < Y < 1$ are examples of finite variation models (of course, only when $\sigma = 0$), and so part (i) of Theorem 2.1 is applicable. Part (iii) is applicable, clearly, to the well-known jump diffusion models by Merton and Kou. In Section 2.5, we will discuss two examples for part (iv) (NIG and Meixner).

2.2 Implied Volatility Slope and Digital Options with Small Maturity

The (Black-Scholes) implied volatility is the volatility that makes the Black-Scholes call price equal the call price with underlying S:

$$C_{\mathrm{BS}}(K,T,\sigma_{\mathrm{imp}}(K,T)) = C(K,T) := \mathbb{E}[(S_T - K)^+].$$

Since no explicit expression is known for $\sigma_{\mathrm{imp}}(K,T)$ (see Gerhold [52]), many authors have investigated approximations (see e.g. the references in the introduction). The following relation between implied volatility slope and digital calls is well known, see Gatheral [50]; we give a proof for completeness. (Note that absolute continuity of S_T holds in all Lévy models of interest, see Theorem 27.4 in Sato [85], and will be assumed throughout.)

Lemma 2.2. *Suppose that the law of S_T is absolutely continuous for each $T > 0$, and that*

$$\lim_{T\downarrow 0} C(K,T) = (S_0 - K)^+, \quad K > 0. \tag{2.4}$$

Then, for $T \downarrow 0$,

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim \sqrt{\frac{2\pi}{T}} \left(\frac{1}{2} - \mathbb{P}[S_T \geq 1] - \frac{\sigma_{\mathrm{imp}}(1,T)\sqrt{T}}{2\sqrt{2\pi}} + \mathcal{O}\big((\sigma_{\mathrm{imp}}(1,T)\sqrt{T})^2 \big) \right). \tag{2.5}$$

Proof. By the implicit function theorem, the implied volatility slope has the representation

$$\partial_K \sigma_{\mathrm{imp}}(K,T) = \frac{\partial_K C(K,T) - \partial_K C_{\mathrm{BS}}(K,T,\sigma_{\mathrm{imp}}(K,T))}{\partial_\sigma C_{\mathrm{BS}}(K,T,\sigma_{\mathrm{imp}}(K,T))}.$$

Since the law of S_T is absolutely continuous, the call price $C(K,T)$ is continuously differentiable w.r.t. K, and $\partial_K C(K,T) = -\mathbb{P}[S_T \geq K]$. Inserting the explicit formulas for the Black-Scholes Vega and digital price, and specializing to the ATM case $K = S_0 = 1$, we get

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} = \frac{\Phi(-\sigma_{\mathrm{imp}}(1,T)\sqrt{T}/2) - \mathbb{P}[S_T \geq 1]}{\sqrt{T}\varphi(\sigma_{\mathrm{imp}}(1,T)\sqrt{T}/2)},$$

where Φ and φ denote the standard Gaussian cdf and density, respectively. By Proposition 4.1 in Roper and Rutkowski [82], our assumption (2.4) implies that the annualized implied volatility $\sigma_{\mathrm{imp}}(1,T)\sqrt{T}$ tends to zero as $T \downarrow 0$. (The second assumption used in Roper and Rutkowski [82] are the no-arbitrage bounds $(S_0 - K)^+ \leq C(K,T) \leq S_0$, for $K, T > 0$, but these are satisfied here because our call prices are generated by the martingale S.) Using the expansion (2.3) and $\varphi(x) = \frac{1}{\sqrt{2\pi}} + \mathcal{O}(x^2)$, we thus obtain (2.5). $\square$

The asymptotic relation (2.5) is, of course, consistent with the small-moneyness expansion presented in De Leo, Vargas, Ciliberti and Bouchaud [21], where the second order term (i.e. first derivative) of implied volatility is $\sqrt{2\pi/T}\,\big(\frac{1}{2} - \mathbb{P}[S_T \geq K]\big)$.
Lemma 2.2 shows that, in order to obtain first order asymptotics for the at-the-money (ATM) slope, we need first order asymptotics for the ATM digital call price $\mathbb{P}[S_T \geq 1]$. (Recall that $S_0 = 1$.) For models where $\lim_{T\downarrow 0} \mathbb{P}[S_T \geq 1] = \frac{1}{2}$, we need the second order term of the digital call as well, and the first order term of $\sigma_{\mathrm{imp}}(1,T)\sqrt{T}$. The limiting value $1/2$ for the ATM digital call is typical for diffusion models (see Gerhold, Kleinert, Porkert and Shkolnikov [54]), and Lévy processes that contain a Brownian motion. For infinite activity models without diffusion component, $\mathbb{P}[S_T \geq 1]$ may converge to $1/2$ as well (e.g. in the CGMY model with $Y \in (1,2)$), but other limiting values are also possible. See the examples in Section 2.5.
From part (i) of Theorem 2.1 and Lemma 2.2 we can immediately conclude the following result. Note that we assume throughout that X is such that $S = e^X$ is a martingale with $S_0 = 1$.

Proposition 2.3. *Suppose that the Lévy process X has finite variation (and thus, necessarily, that $\sigma = 0$), and that $b_0 \neq 0$. Then the ATM implied volatility slope satisfies*

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim -\sqrt{\pi/2}\,\mathrm{sgn}(b_0) \cdot T^{-1/2}, \quad T \downarrow 0.$$

Note that $T^{-1/2}$ is the fastest possible growth order for the slope, in any model (see Lee [67]).
If X is a Lévy jump diffusion with $\sigma > 0$, then by part (iii) of Theorem 2.1, (2.5), and the fact that $\sigma_{\mathrm{imp}} \to \sigma$ (implied volatility converges to spot volatility), we obtain the finite limit

$$\lim_{T\downarrow 0} \partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} = -\frac{b_0}{\sigma} - \frac{\sigma}{2}. \tag{2.6}$$

(It is understood that the substitution $K = 1$ is to be performed before the limit $T \downarrow 0$.) Notice that the expression on the right hand side of (2.6) does depend on the jump parameters, because the drift b_0, fixed by the condition $\mathbb{E}[\exp(X_1)] = 1$, depends on them. Moreover, (2.6) is consistent with the formal calculation of the variance slope

$$\lim_{T\downarrow 0} \partial_K \sigma^2_{\mathrm{imp}}(K,T)|_{K=1} = -2b_0 - \sigma^2$$

on p. 61f in Gatheral [50]. In fact (2.6) is well known for jump diffusions, see Alòs, León and Vives [3] and Yan [90].

2.3 General Remarks on Mellin Transform Asymptotics

As mentioned after Lemma 2.2, we need the second order term for the ATM digital call if we want to find the limiting slope in Lévy models with a Brownian component. While this is easy for finite activity models (see the end of the preceding section), it is more difficult in the case of infinite activity jumps. We will find this second order term using Mellin transform asymptotics. For further details and references on this technique, see e.g. Flajolet, Gourdon and Dumas [37]. The Mellin transform of a function H, locally integrable on $(0,\infty)$, is defined by

$$(\mathcal{M}H)(s) = \int_0^\infty T^{s-1} H(T)\, dT.$$

Under appropriate growth conditions on H at zero and infinity, this integral defines an analytic function in an open vertical strip of the complex plane. The function H can be recovered from its transform by Mellin inversion (see formula (7) in Flajolet, Gourdon and Dumas [37]):

$$H(T) = \frac{1}{2\pi i} \int_{\kappa - i\infty}^{\kappa + i\infty} (\mathcal{M}H)(s) T^{-s} ds, \tag{2.7}$$

where κ is a real number in the strip of analyticity of $\mathcal{M}H$. For the validity of (2.7), it suffices that H is continuous and that $y \mapsto (\mathcal{M}H)(\kappa + iy)$ is integrable. Denote by $s_0 \in \mathbb{R}$ the real part of the left boundary of the strip of analyticity. A typical situation in applications is that $\mathcal{M}H$ has a pole at s_0, and admits a meromorphic extension to a left half-plane, with further poles at $s_0 > s_1 > s_2 > \dots$ Suppose also that the meromorphic continuation satisfies growth estimates at $\pm i\infty$ which allow to shift the integration path in (2.7) to the left. We then collect the contribution of each pole by the residue theorem, and arrive at an expansion (see formula (8) in Flajolet, Gourdon and Dumas [37])

$$H(T) = \mathrm{Res}_{s=s_0}(\mathcal{M}H)(s)T^{-s} + \mathrm{Res}_{s=s_1}(\mathcal{M}H)(s)T^{-s} + \dots$$

Thus, the basic principle is that singularities s_i of the transform are mapped to terms T^{-s_i} in the asymptotic expansion of H at zero. Simple poles of $\mathcal{M}H$ yield powers of T, whereas double poles produce an additional logarithmic factor $\log T$, as seen from the expansion $T^{-s} = T^{-s_i}(1 - (\log T)(s - s_i) + \mathcal{O}((s - s_i)^2))$.

2.4 Main Results: Digital Call Prices and Slope Asymptotics

The mgf $M(z,T)$ of X_T is analytic in a strip $z_- < \mathrm{Re}(z) < z_+$, given by the critical moments

$$z_+ = \sup\{z \in \mathbb{R} \colon \mathbb{E}[e^{zX_T}] < \infty\} \tag{2.8}$$

and

$$z_- = \inf\{z \in \mathbb{R} \colon \mathbb{E}[e^{zX_T}] < \infty\}. \tag{2.9}$$

Since X is a Lévy process, the critical moments do not depend on T. We will obtain asymptotic information on the transition probabilities (i.e. digital call prices) from the Fourier representation in Lee [66]

$$\begin{aligned} \mathbb{P}[S_T \geq 1] &= \mathbb{P}[X_T \geq 0] \\ &= \frac{1}{2i\pi}\int_{a-i\infty}^{a+i\infty} \frac{M(z,T)}{z}\,dz \\ &= \frac{1}{\pi}\,\mathrm{Re}\int_0^{\infty} \frac{M(a+iy,T)}{a+iy}\,dy, \end{aligned} \tag{2.10}$$

where the real part of the vertical integration contour satisfies $a \in (0,1) \subseteq (z_-, z_+)$, and convergence of the integral is assumed throughout. We are going to analyse the asymptotic behaviour of this integral, for $T \downarrow 0$, by computing its Mellin transform. Asymptotics of the probability (digital price) $\mathbb{P}[X_T \geq 0]$ are then evident from (2.10). The linearity of $\log M$ as a function of T enables us to evaluate the Mellin transform in semi-explicit form.

Lemma 2.4. *Suppose that $S = e^X$ is a martingale, and that $\sigma > 0$. Then, for any $a \in (0,1)$, the Mellin transform of the function*

$$H(T) := \int_0^{\infty} \frac{e^{T\psi(a+iy)}}{a+iy}\,dy, \quad T > 0, \tag{2.11}$$

is given by

$$(\mathcal{M}H)(s) = \Gamma(s)F(s), \quad 0 < \mathrm{Re}(s) < \tfrac{1}{2}, \tag{2.12}$$

where

$$F(s) = \int_0^\infty \frac{(-\psi(a+iy))^{-s}}{a+iy}\,dy, \quad 0 < \mathrm{Re}(s) < \tfrac{1}{2}. \tag{2.13}$$

Moreover, $|(\mathcal{M}H)(s)|$ *decays exponentially, if* $\mathrm{Re}(s) \in (0, \frac{1}{2})$ *is fixed and* $|\mathrm{Im}(s)| \to \infty$*.*

See the Appendix A for the proof of Lemma 2.4. With the Mellin transform in hand, we now proceed to convert an expansion of the mgf at $i\infty$ to an expansion of $\mathbb{P}[X_T \geq 0]$ for $T \downarrow 0$. The following result covers, e.g. the NIG and Meixner models, and the generalized tempered stable model, all with $\sigma > 0$. See Section 2.5 for details.

Theorem 2.5. *Suppose that* $S = e^X$ *is a martingale, and that* $\sigma > 0$*. Assume further that there are constants* $a \in (0,1)$*,* $c \in \mathbb{C}$*,* $\nu \in [1,2)$ *and* $\varepsilon > 0$ *such that the Laplace exponent satisfies*

$$\psi(z) = \frac{1}{2}\sigma^2 z^2 + cz^\nu + \mathcal{O}(z^{\nu-\varepsilon}), \quad \mathrm{Re}(z) = a, \ \mathrm{Im}(z) \to \infty. \tag{2.14}$$

Then the ATM digital call price satisfies

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + C_{\tilde{\nu}} T^{\tilde{\nu}} + o(T^{\tilde{\nu}}), \quad T \downarrow 0, \tag{2.15}$$

where $C_{\tilde{\nu}} = \frac{\tilde{\nu}}{2\pi}\left(\frac{1}{2}\sigma^2\right)^{\tilde{\nu}-1} \mathrm{Im}(e^{-i\pi\tilde{\nu}}c)\Gamma(-\tilde{\nu})$ *with* $\tilde{\nu} = (2-\nu)/2 \in (0, \frac{1}{2}]$*. For* $\nu = 1$*, this simplifies to*

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{\mathrm{Re}(c)}{\sigma\sqrt{2\pi}}\sqrt{T} + o(\sqrt{T}), \quad T \downarrow 0.$$

Together with Lemma 2.2, this theorem implies the following corollary, which is our main result on the implied volatility slope as $T \downarrow 0$.

Corollary 2.6. *Under the assumptions of Theorem 2.5, the ATM implied volatility slope behaves as follows:*

(i) If $\nu = 1$*, then*

$$\lim_{T\downarrow 0} \partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} = -\frac{\mathrm{Re}(c)}{\sigma} - \frac{\sigma}{2},$$

with c *from (2.14).*

(ii) If $1 < \nu < 2$ *and* $C_{\tilde{\nu}} \neq 0$*, then*

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim -\sqrt{2\pi}\, C_{\tilde{\nu}} T^{\tilde{\nu}-1/2}, \quad T \downarrow 0.$$

Proof of Theorem 2.5. From (2.10) and (2.11) we know that

$$\mathbb{P}[X_T \geq 0] = \frac{1}{\pi} \operatorname{Re} H(T). \tag{2.16}$$

We now express $H(T)$ by the Mellin inversion formula (2.7), with $\kappa \in (0, \frac{1}{2})$. This is justified by Lemma 2.4, which yields the exponential decay of the transform $\mathcal{M}H$ along vertical rays. (Continuity of H, which is also needed for the inverse transform, is clear.) Therefore, we have

$$H(T) = \frac{1}{2\pi i} \int_{1/4-i\infty}^{1/4+i\infty} \Gamma(s) F(s) T^{-s} ds, \quad T \geq 0. \tag{2.17}$$

As outlined in Section 2.3, we now show that $\Gamma(s)F(s)$ has a meromorphic continuation, then shift the integration path in (2.17) to the left, and collect residues. It is well known that Γ is meromorphic with poles at the non-positive integers, so it suffices to discuss the continuation of F, defined in (2.13). As in the proof of Lemma 2.4, we put $h(y) := -\psi(a + iy)$, $y \geq 0$. To prove exponential decay of the desired meromorphic continuation, it is convenient to split the integral:

$$\begin{aligned} F(s) &= \int_0^{y_0} \frac{h(y)^{-s}}{a+iy} dy + \int_{y_0}^{\infty} \frac{h(y)^{-s}}{a+iy} dy \\ &=: A_0(s) + \tilde{F}(s), \quad 0 < \operatorname{Re}(s) < \tfrac{1}{2}. \end{aligned} \tag{2.18}$$

The constant $y_0 \geq 0$ will be specified later. It is easy to see that A_0 is analytic in the half-plane $\operatorname{Re}(s) < \frac{1}{2}$, and so $\tilde{F}$ captures all poles of F in that half-plane. By (2.14), the function h has the expansion (with a possibly decreased ε, to be precise)

$$h(y) = \tfrac{1}{2}\sigma^2 y^2 + \tilde{c} y^\nu + \mathcal{O}(y^{\nu-\varepsilon}), \quad y \to \infty, \tag{2.19}$$

where

$$\tilde{c} := \begin{cases} -ci^\nu & \nu > 1, \\ -(c + \sigma^2 a)i & \nu = 1. \end{cases}$$

The reason why F (or $\tilde{F}$) is not analytic at $s = 0$ is that the second integral in (2.18) fails to converge for y large. We thus subtract the following convergence-inducing integral from $\tilde{F}$:

$$\begin{aligned} \tilde{G}_1(s) &:= \int_{y_0}^{\infty} \frac{(\frac{1}{2}\sigma^2 y^2)^{-s}}{a+iy} dy \\ &= -\pi i (\tfrac{1}{2} a^2 \sigma^2)^{-s} \frac{e^{i\pi s}}{\sin 2\pi s} - \int_0^{y_0} \frac{(\frac{1}{2}\sigma^2 y^2)^{-s}}{a+iy} dy \\ &=: G_1(s) + A_1(s). \end{aligned} \tag{2.20}$$

Note that G_1 is meromorphic, and that A_1 is analytic for $\operatorname{Re}(s) < \frac{1}{2}$. From the expansion

$$h(y)^{-s} = (\tfrac{1}{2}\sigma^2 y^2)^{-s} - \frac{2\tilde{c}s}{\sigma^2} \left(\frac{\sigma^2}{2}\right)^{-s} y^{\nu-2s-2} + \mathcal{O}(y^{\nu-2\operatorname{Re}(s)-2-\varepsilon}), \quad y \to \infty, \tag{2.21}$$

for s fixed, we see that the function

$$\tilde{F}_1(s) := \int_{y_0}^{\infty} \frac{1}{a+iy}\left(h(y)^{-s} - (\tfrac{1}{2}\sigma^2 y^2)^{-s}\right) dy \tag{2.22}$$

is analytic for $-\tilde{\nu} < \mathrm{Re}(s) < \frac{1}{2}$, and, clearly, for $0 < \mathrm{Re}(s) < \frac{1}{2}$ we have

$$\tilde{F}(s) = \tilde{F}_1(s) + \tilde{G}_1(s). \tag{2.23}$$

We have thus established the meromorphic continuation of $\tilde{F}$ to the strip $-\tilde{\nu} < \mathrm{Re}(s) < \frac{1}{2}$. To continue $\tilde{F}$ even further, we look at the second term in (2.21) and define

$$\begin{aligned}
\tilde{G}_2(s) &:= -\frac{2\tilde{c}s}{\sigma^2}\left(\frac{\sigma^2}{2}\right)^{-s}\int_{y_0}^{\infty}\frac{y^{\nu-2s-2}}{a+iy}\,dy \\
&= -\frac{2\tilde{c}\pi}{\sigma^2}\left(\frac{\sigma^2}{2}\right)^{-s} s a^{\nu-2s-2}\frac{e^{(2s-\nu+3)\pi i/2}}{\sin\pi(\nu-2s)} + \frac{2\tilde{c}s}{\sigma^2}\left(\frac{\sigma^2}{2}\right)^{-s}\int_0^{y_0}\frac{y^{\nu-2s-2}}{a+iy}\,dy \\
&=: G_2(s) + A_2(s)
\end{aligned}$$

and the compensated function

$$\tilde{F}_2(s) := \int_{y_0}^{\infty}\frac{1}{a+iy}\left(h(y)^{-s} - (\tfrac{1}{2}\sigma^2y^2)^{-s} + \frac{2\tilde{c}s}{\sigma^2}\left(\frac{\sigma^2}{2}\right)^{-s} y^{\nu-2s-2}\right) dy.$$

By (2.21), the function $\tilde{F}_2$ is analytic for $\mathrm{Re}(s) \in (-\tilde{\nu}-\varepsilon/2, (\nu-1)/2)$. Moreover, by definition we have

$$\tilde{F}_1(s) = \tilde{F}_2(s) + \tilde{G}_2(s), \quad -\tilde{\nu} < \mathrm{Re}(s) < \tfrac{\nu-1}{2},$$

and so the meromorphic continuation of $\tilde{F}$ to the region $-\tilde{\nu}-\varepsilon/2 < \mathrm{Re}(s) < \frac{1}{2}$ is established.

In order to shift the integration path in (2.17) to the left, we have to ensure that the integral converges. This is the content of Lemma 2.7 below, which also yields the existence of an appropriate $y_0 \geq 0$, to be used in the definition of $\tilde{F}$ in (2.18). By the residue theorem, we obtain

$$\begin{aligned}
H(T) = \mathrm{Res}_{s=0}(\mathcal{M}H)(s)T^{-s} &+ \mathrm{Res}_{s=-\tilde{\nu}}(\mathcal{M}H)(s)T^{-s} \\
&+ \frac{1}{2\pi i}\int_{\kappa-i\infty}^{\kappa+i\infty}(\mathcal{M}H)(s)T^{-s}ds, \quad T \geq 0,
\end{aligned} \tag{2.24}$$

where $\kappa = -\tilde{\nu}-\varepsilon/4$, and $\mathcal{M}H$ now of course denotes the meromorphic continuation of the Mellin transform. We then compute the residues. According to (2.18) and (2.23), the continuation of $\mathcal{M}H$ in a neighbourhood of $s=0$ is given by $\Gamma(s)(A_0(s)+\tilde{F}_1(s)+\tilde{G}_1(s))$. Therefore,

$$\begin{aligned}
\mathrm{Res}_{s=0}(\mathcal{M}H)(s)T^{-s} &= A_0(0) + \tilde{F}_1(0) + A_1(0) + \mathrm{Res}_{s=0}\Gamma(s)G_1(s)T^{-s} \\
&= \mathrm{Res}_{s=0}\Gamma(s)G_1(s)T^{-s} \\
&= \tfrac{\pi}{2} + i(\tfrac{1}{2}\gamma - \log(a\sigma/\sqrt{2}) + \tfrac{1}{2}\log T),
\end{aligned} \tag{2.25}$$

where γ is Euler's constant. Note that $A_0(0) = -A_1(0)$ and $\tilde{F}_1(0) = 0$ by definition. The remaining residue (2.25) is straightforward to compute from (2.20) (e.g. with a computer algebra system) and has real part $\frac{1}{2}\pi$. Notice that the logarithmic term $\log T$, resulting from the *double* pole at zero (see the end of Section 2.3), appears only in the imaginary part. Recalling (2.16), we see that the first term on the right-hand side of (2.24) thus yields the first term of (2.15).
Similarly, we compute for $\nu > 1$

$$\begin{aligned}\mathrm{Res}_{s=-\tilde{\nu}}(\mathcal{M}H)(s)T^{-s} &= \mathrm{Res}_{s=-\tilde{\nu}}\Gamma(s)G_2(s)T^{-s}\\ &= \frac{\Gamma(-\tilde{\nu})}{2\pi}\left[\frac{2\tilde{c}s}{\sigma^2}\left(\frac{\sigma^2}{2}\right)^{-s}\pi a^{\nu-2s-2}e^{(2s-\nu+3)\pi i/2}T^{-s}\right]_{s=-\tilde{\nu}}.\end{aligned}$$

In the case $\nu = 1$, the function G_1 also has a pole at $-\tilde{\nu} = -\frac{1}{2}$, and we obtain

$$\begin{aligned}\mathrm{Res}_{s=-\tilde{\nu}}(\mathcal{M}H)(s)T^{-s} &= \mathrm{Res}_{s=-1/2}\Gamma(s)(G_1(s)+G_2(s))T^{-s}\\ &= \sqrt{\frac{\pi}{2}}\left(\frac{i\tilde{c}}{\sigma} - a\sigma\right)\sqrt{T}.\end{aligned}$$

A straightforward computation shows that the stated formula for $C_{\tilde{\nu}}$ is correct in both cases. The integral on the right-hand side of (2.24) is clearly $\mathcal{O}(T^{-\kappa}) = o(T^{\tilde{\nu}})$, and so the proof is complete. □

Lemma 2.7. *There is $y_0 \geq 0$ such that the meromorphic continuation of $\mathcal{M}H$ constructed in the proof of Theorem 2.5, which depends on y_0 via the definition of $\tilde{F}$ in (2.18), decays exponentially as $|\,\mathrm{Im}(s)| \to \infty$.*

Lemma 2.7 is proved in the Appendix A.

2.5 Examples

We now apply our main results (Theorem 2.5 and Corollary 2.6) to several concrete models.

Example 2.8. The NIG (Normal Inverse Gaussian) model has Laplace exponent

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + \mu z + \delta\left(\sqrt{\hat{\alpha}^2-\beta^2} - \sqrt{\hat{\alpha}^2-(\beta+z)^2}\right),$$

where $\delta > 0$, $\hat{\alpha} > \max\{\beta+1, -\beta\}$. (The notation $\hat{\alpha}$ should avoid confusion with α from Theorem 2.1.) Since S is a martingale, we must have

$$\mu = -\tfrac{1}{2}\sigma^2 + \delta\left(\sqrt{\hat{\alpha}^2-(\beta+1)^2} - \sqrt{\hat{\alpha}^2-\beta^2}\right).$$

The relation between μ and b from (2.1) is $\mu + \beta\delta/\sqrt{\hat{\alpha}^2-\beta^2} = b$, as seen from the derivative of the Laplace exponent ψ at $z = 0$. The Lévy density is

$$\frac{\nu(dx)}{dx} = \frac{\delta\hat{\alpha}}{\pi|x|}e^{\beta x}K_1\big(\hat{\alpha}|x|\big),$$

where K_1 is the modified Bessel function of second order and index 1.
First assume $\sigma = 0$. Since $K_1(x) \sim 1/x$ for $x \downarrow 0$, condition (**H**-α) is satisfied with $\alpha = 1$, with $c_+ = c_- = \delta/\pi$. The integrability condition in part (iv) of Theorem 2.1 is easily checked, and we conclude

$$\lim_{T\downarrow 0} \mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{1}{\pi}\arctan\Big(\frac{\mu}{\delta}\Big), \quad \sigma = 0.$$

Note that $b^* = \mu = b - \frac{\delta\hat{\alpha}}{\pi}\int_0^\infty K_1(\hat{\alpha}x)(e^{\beta x} - e^{-\beta x})dx$. By Lemma 2.2, the implied volatility slope of the NIG model thus satisfies

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim -\sqrt{2/\pi}\,\arctan(\mu/\delta)\cdot T^{-1/2}, \quad T\downarrow 0, \quad \sigma = 0,\ \mu \neq 0.$$

Now assume that $\sigma > 0$. Since $\sqrt{\hat{\alpha}^2 - (\beta+z)^2} = -iz + \mathcal{O}(1)$ as $\mathrm{Im}(z) \to \infty$, the expansion (2.14) becomes

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + (\mu + i)z + \mathcal{O}(1), \quad \mathrm{Re}(z) = a,\ \mathrm{Im}(z) \to \infty.$$

We can thus apply Theorem 2.5 to conclude that the ATM digital price satisfies

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{\mu}{\sigma\sqrt{2\pi}}\sqrt{T} + o(\sqrt{T}), \quad T\downarrow 0, \quad \sigma > 0.$$

By part (i) of Corollary 2.6, the limit of the implied volatility slope is given by

$$\begin{aligned}\lim_{T\downarrow 0}\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} &= -\frac{\mu}{\sigma} - \frac{\sigma}{2} \\ &= \frac{\delta}{\sigma}\big(\sqrt{\hat{\alpha}^2 - \beta^2} - \sqrt{\hat{\alpha}^2 - (\beta+1)^2}\big), \quad \sigma > 0. \end{aligned} \tag{2.26}$$

This limit is positive if and only if $\beta > -\frac{1}{2}$.

See Figure 2.1 for a numerical example. Let us stress again that we identify the correct *sign* of the slope, while we find that explicit asymptotics do not approximate the *value* of the slope very accurately. Still, in the right panel of Figure 2.1 we have zoomed in at very short maturity to show that our approximation gives the asymptotically correct tangent in this example.

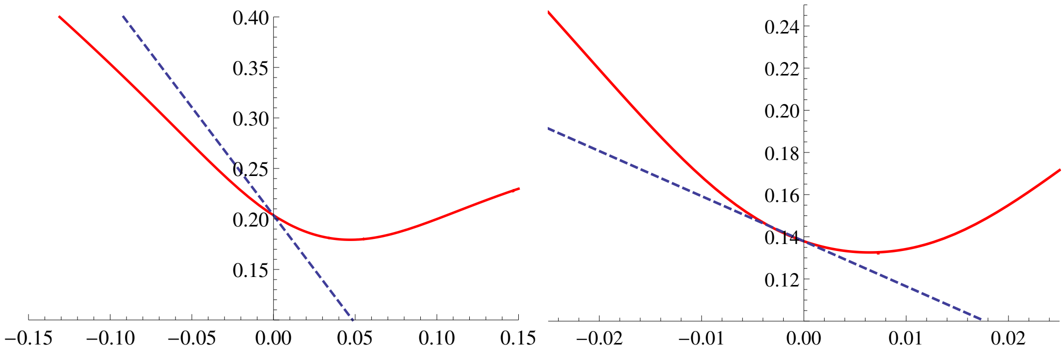

Figure 2.1: The volatility smile, as a function of log-strike, of the NIG model with parameters $\sigma = 0.085$, $\hat{\alpha} = 4.237$, $\beta = -3.55$, $\delta = 0.167$, and maturity $T = 0.1$ (left panel) respectively $T = 0.01$ (right panel). The parameters were calibrated to S&P 500 call prices from Appendix A of Bu [15]. The dashed line is the slope approximation (2.26). We did the calibration and the plots with Mathematica, using the Fourier representation of the call price.

Example 2.9. The Laplace exponent of the Meixner model is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + \mu z + 2\hat{d} \log \frac{\cos(\hat{b}/2)}{\cosh \frac{1}{2}(-\hat{a}iz - i\hat{b})},$$

where $\hat{d} > 0$, $\hat{b} \in (-\pi, \pi)$, and $0 < \hat{a} < \pi - \hat{b}$. (We follow the notation of Schoutens [86], except that we write μ instead of m, and $\hat{a}, \hat{b}, \hat{d}$ instead of a, b, d.) The Lévy density is

$$\frac{\nu(dx)}{dx} = \hat{d}\frac{\exp(\hat{b}x/\hat{a})}{x \sinh(\pi x/\hat{a})}.$$

We can proceed analogously to Example 2.8. For $\sigma = 0$ we again apply part (iv) of Theorem 2.1, with $\alpha = 1$, where now $c_+ = c_- = \hat{d}\hat{a}/\pi$. Consequently,

$$\lim_{T\downarrow 0} \mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\mu}{\hat{a}\hat{d}}\right), \quad \sigma = 0,$$

and

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim -\sqrt{2/\pi}\arctan\left(\frac{\mu}{\hat{a}\hat{d}}\right) \cdot T^{-1/2}, \quad T \downarrow 0, \quad \sigma = 0,\ \mu \neq 0.$$

Now assume $\sigma > 0$. The expansion of the Laplace exponent is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + (\mu + \hat{a}\hat{d}i)z + \mathcal{O}(1), \quad \mathrm{Re}(z) = a,\ \mathrm{Im}(z) \to \infty.$$

By Theorem 2.5, the ATM digital price in the Meixner model thus satisfies

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{\mu}{\sigma\sqrt{2\pi}}\sqrt{T} + o(\sqrt{T}), \quad T \downarrow 0.$$

The limit of the implied volatility slope is given by

$$\begin{aligned}\lim_{T\downarrow 0} \partial_K \sigma_{\text{imp}}(K,T)|_{K=1} &= -\frac{\mu}{\sigma} - \frac{\sigma}{2} \\ &= \frac{2\hat{d}}{\sigma} \log\left(\frac{\cos(\hat{b}/2)}{\cosh\frac{1}{2}(-(\hat{a}+\hat{b})i)}\right), \quad \sigma > 0.\end{aligned}$$

Example 2.10. The Laplace exponent of the CGMY model is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + \mu z + C\Gamma(-Y)((M-z)^Y - M^Y + (G+z)^Y - G^Y), \tag{2.27}$$

where we assume $C > 0$, $G > 0$, $M > 1$, $0 < Y < 2$, and $Y \neq 1$.
The case $\sigma = 0$ and $Y \in (0,1)$ need not be discussed, as it is a special case of Proposition 8.5 in Andersen and Lipton [4]. Our Proposition 2.3 could also be applied, as the CGMY process has finite variation in this case.
If $\sigma = 0$ and $Y \in (1,2)$, then the ATM digital call price converges to $\frac{1}{2}$, and the slope explodes, of order $T^{1/2-1/Y}$. This is a special case of Corollary 3.3 in Figueroa-López and Ólafsson [36]. Note that Proposition 8.5 in Andersen and Lipton [4] is not applicable here, because the constant $C_{\mathfrak{M}}$ from this proposition vanishes for the CGMY model, and so the leading term of the slope is not obtained. Theorem 2.1 (iv) from our Section 2.1 is not useful, either; it gives the correct digital call limit price $\frac{1}{2}$, but does not provide the second order term necessary to get slope asymptotics.
We now proceed to the case $\sigma > 0$, which is our main focus. The expansion of ψ at $i\infty$ is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + c_Y z^Y + \mu z + \mathcal{O}(z^{Y-1}), \quad \operatorname{Re}(z) = a,\ \operatorname{Im}(z) \to \infty,$$

with the complex constant $c_Y := C\Gamma(-Y)(1+e^{-i\pi Y})$. First assume $0 < Y < 1$. Then we proceed analogously to the preceding examples, applying Theorem 2.5 and Corollary 2.6. The ATM digital price thus satisfies

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{\mu}{\sigma\sqrt{2\pi}}\sqrt{T} + o(\sqrt{T}), \quad T \downarrow 0, \tag{2.28}$$

and the limit of the implied volatility slope is given by

$$\begin{aligned}\lim_{T\downarrow 0} \partial_K \sigma_{\text{imp}}(K,T)|_{K=1} &= -\frac{\mu}{\sigma} - \frac{\sigma}{2} \\ &= \frac{1}{\sigma} C\Gamma(-Y)((M-1)^Y - M^Y + (G+1)^Y - G^Y).\end{aligned} \tag{2.29}$$

Now assume $1 < Y < 2$. In principle, Theorem 2.5 is applicable, with $\nu = Y$; however, the constant $C_{\tilde{\nu}}$ in (2.15) is zero, and so we do not get the second term of the expansion immediately. What happens is that the Mellin transform of H (see the proof of Theorem 2.5) may have further poles in $-\frac{1}{2} < \operatorname{Re}(s) < 0$, but none of them gives a contribution, since the corresponding residues have zero real part. Therefore, (2.28) and (2.29) are true also for $1 < Y < 2$. For a rigorous proof, see Theorem B.2 in the

Appendix B. Note that (2.28) and (2.29) also follow from concurrent work by Figueroa-López and Ólafsson [36]. For $0 < Y < 1$, they also follow from Proposition 8.5 in Andersen and Lipton [4], but not for $1 < Y < 2$, because then the constant $C_{\mathfrak{M}}$ from that proposition vanishes when specializing it to the CGMY model.

In the following example, we discuss the generalized tempered stable model. The tempered stable model, which is investigated in Andersen and Lipton [4], is obtained by setting $\alpha_- = \alpha_+$.

Example 2.11. The generalized tempered stable process, see e.g. Cont and Tankov [19], is a generalization of the CGMY model, with Lévy density

$$\frac{\nu(dx)}{dx} = \frac{C_-}{|x|^{1+\alpha_-}} e^{-\lambda_-|x|} 1_{(-\infty,0)}(x) + \frac{C_+}{|x|^{1+\alpha_+}} e^{-\lambda_+|x|} 1_{(0,\infty)}(x),$$

where $\alpha_\pm < 2$ and $C_\pm, \lambda_\pm > 0$. For $\alpha_\pm \notin \{0,1\}$ the Laplace exponent of the generalized tempered stable process is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + \mu z + \Gamma(-\alpha_+)C_+\Big((\lambda_+ - z)^{\alpha_+} - \lambda_+^{\alpha_+}\Big) \\ + \Gamma(-\alpha_-)C_-\Big((\lambda_- + z)^{\alpha_-} - \lambda_-^{\alpha_-}\Big).$$

For $\sigma > 0$, $\alpha_+ \in (1,2)$, and $\alpha_- < \alpha_+$ we have the following expansion:

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + \Gamma(-\alpha_+)C_+ e^{-i\pi\alpha_+} z^{\alpha_+} + \mathcal{O}(z^{\max\{1,\alpha_-\}}), \quad \mathrm{Re}(z) = a,\ \mathrm{Im}(z) \to \infty.$$

We now apply Theorem 2.5 with $\nu = \alpha_+$, and find that the second order expansion of the ATM digital call is

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + C_{\tilde\nu} T^{\tilde\nu} + o(T^{\tilde\nu}), \quad T \downarrow 0,$$

with $\tilde\nu = 1 - \alpha_+/2 \in (0, \frac{1}{2})$ and the real constant

$$C_{\tilde\nu} = \frac{\tilde\nu}{2\pi} \left(\tfrac{1}{2}\sigma^2\right)^{\tilde\nu - 1} \Gamma(-\alpha_+)C_+ \underbrace{\mathrm{Im}(e^{-i\pi\tilde\nu} e^{-i\pi\alpha_+})}_{=\sin(-\pi(1+\alpha_+/2))} \Gamma(-\tilde\nu).$$

By Corollary 2.6 (i), the ATM implied volatility slope explodes, but slower than $T^{-1/2}$:

$$\partial_K \sigma_{\mathrm{imp}}(K,T)|_{K=1} \sim -\sqrt{2\pi} C_{\tilde\nu} T^{\tilde\nu - 1/2}, \quad T \downarrow 0.$$

Note that these results also follow from the concurrent paper Figueroa-López and Ólafsson [36], which treats tempered stable-like models.
If $\sigma > 0$ and $\alpha_+ < 1$, then part (i) of Corollary 2.6 is applicable, and formulas analogous to (2.28) and (2.29) hold.

2.6 Robustness of Lee's Moment Formula

As we have already mentioned, our first order slope approximations give limited accuracy for the size of the slope, but usually succeed at identifying its sign, i.e. whether the smile increases or decreases at the money. It is a natural question whether this sign gives information on the smile as a whole: If the slope is positive, does it follow that the right wing is steeper than the left one, and vice versa? To deal with this issue, recall Lee's moment formula in Lee [65]. Under the assumption that the critical moments z_+ and z_-, defined in (2.8) and (2.9), are finite, Lee's formula states that

$$\limsup_{k\to\infty} \frac{\sigma_{\text{imp}}(K,T)}{\sqrt{k}} = \sqrt{\frac{\Psi(z_+ - 1)}{T}} \tag{2.30}$$

and

$$\limsup_{k\to-\infty} \frac{\sigma_{\text{imp}}(K,T)}{\sqrt{-k}} = \sqrt{\frac{\Psi(-z_-)}{T}}, \tag{2.31}$$

where $T > 0$ is fixed, $k = \log K$, and $\Psi(x) := 2 - 4(\sqrt{x^2+x} - x)$. According to Lee's formula, the slopes of the wings depend on the size of the critical moments. In Lévy models, the critical moments do not depend on T. The compatibility property we seek now becomes:

$$\lim_{k\to\infty} \frac{\sigma_{\text{imp}}(K,T)}{\sqrt{k}} > \lim_{k\to-\infty} \frac{\sigma_{\text{imp}}(K,T)}{\sqrt{-k}} \quad \text{for all } T > 0 \tag{2.32}$$

if and only if

$$\partial_K \sigma_{\text{imp}}(K,T)|_{K=1} > 0 \quad \text{for all sufficiently small } T. \tag{2.33}$$

That is, the right wing of the smile is steeper than the left wing deep *out-of-the-money* if and only if the small-maturity *at-the-money* slope is positive. We now show that this is true for several infinite activity Lévy models. By our methods, this can certainly be extended to other infinite activity models. It does not hold, though, for the Merton and Kou jump diffusion models. The parameter ranges in the following theorem are the same as in the examples in Section 2.5.

Theorem 2.12. *Conditions (2.32) and (2.33) are equivalent for the following models. For the latter three, we assume that $\sigma > 0$ or $\mu \neq 0$.*

- *Variance gamma with $\sigma = 0$, $b_0 \neq 0$*
- *NIG*
- *Meixner*
- *CGMY*

Put differently, these models are *not* capable (at short maturity) of producing a smile that has, say, its minimum to the left of $\log K = k = 0$, and thus a positive ATM slope, but whose left wing is steeper than the right one.

Proof. The critical moments are clearly finite for all of these models. Moreover, it is well known that the lim sup in (2.30) and (2.31) can typically be replaced by a genuine limit, for instance using the criteria given by Benaim and Friz [7]. Their conditions on the mgf are easily verified for all our models; in fact Benaim and Friz [7] explicitly treat the variance gamma model with $b_0 = 0$ and the NIG model. We thus have to show that (2.33) is equivalent to $\Psi(z_+ - 1) > \Psi(-z_-)$. Since Ψ is strictly decreasing on $(0, \infty)$, the latter condition is equivalent to $z_+ - 1 < -z_-$. It remains to check the equivalence

$$z_+ - 1 < -z_- \quad \Longleftrightarrow \quad (2.33). \tag{2.34}$$

The mgf of the variance gamma model is (see Madan, Carr and Chang [69])

$$M(z,T) = e^{Tb_0 z}(1 - \theta\nu z - \tfrac{1}{2}\hat{\sigma}^2\nu z^2)^{-T/\nu},$$

where $\hat{\sigma}, \nu > 0$ and $\theta \in \mathbb{R}$. Its paths have finite variation, and so Proposition 2.3 shows that (2.33) is equivalent to $b_0 < 0$. The critical moments are

$$z_\pm = -\frac{\nu\theta \pm \sqrt{2\nu\hat{\sigma}^2 + \nu^2\theta^2}}{\nu\hat{\sigma}^2},$$

and we have $-z_- + 1 - z_+ = 1 + 2\theta/\hat{\sigma}^2$. This is positive if and only if

$$b_0 = \nu^{-1}\log(1 - \theta\nu - \tfrac{1}{2}\hat{\sigma}^2\nu) < 0,$$

which yields (2.34).
As for the other three models, first suppose that $\sigma > 0$. The examples in Section 2.5 show that (2.33) is equivalent to $\mu < -\frac{1}{2}\sigma^2$. The critical moments of the NIG model are $z_+ = \hat{\alpha} - \beta$ and $z_- = -\hat{\alpha} - \beta$. Therefore, $z_+ - 1 < -z_-$ if and only if $\beta > -\frac{1}{2}$, and this is indeed equivalent to

$$\mu + \tfrac{1}{2}\sigma^2 = \delta\left(\sqrt{\hat{\alpha}^2 - (\beta+1)^2} - \sqrt{\hat{\alpha}^2 - \beta^2}\right) < 0.$$

For the Meixner model, we have $z_\pm = (\pm\pi - \hat{b})/\hat{a}$, which yields $-z_- + 1 - z_+ = 1 + 2\hat{b}/\hat{a}$. On the other hand,

$$\mu + \tfrac{1}{2}\sigma^2 = -2\hat{d}\log\frac{\cos(\hat{b}/2)}{\cos((\hat{a}+\hat{b})/2)},$$

which is negative if and only if $\cos(\hat{b}/2) > \cos((\hat{a}+\hat{b})/2)$, and this is equivalent to $\hat{a} + 2\hat{b} > 0$.
Finally, in case of the CGMY model, we have

$$\mu + \tfrac{1}{2}\sigma^2 = -C\Gamma(-Y)\big((M-1)^Y - M^Y + (G+1)^Y - G^Y\big).$$

Since, for $Y \in (0,1)$, $\Gamma(-Y) < 0$ and the function $x \mapsto x^Y - (x+1)^Y$ is strictly increasing on $(0,\infty)$, we see that $\mu + \frac{1}{2}\sigma^2 < 0$ if and only if $M - 1 < G$. This is the desired condition, since the explicit expression (2.27) shows that $z_+ = M$ and $z_- = -G$. The case $Y \in (1,2)$ is analogous.

It remains to treat the case $\sigma = 0$. First, note that the critical moments do not depend on σ. Furthermore, from the examples in Section 2.5, we see that (2.33) holds if and only if $\mu < 0$. Now observe that adding a Brownian motion σW_t to a Lévy model adds $-\frac{1}{2}\sigma^2$ to the drift, if the martingale property is to be preserved. Therefore, the assertion follows from what we have already proved about $\sigma > 0$. $\square$

Chapter 3

Conclusion

Our main result (Corollary 2.6) translates asymptotics of the log-underlying's mgf to first-order asymptotics for the ATM implied volatility slope. Checking the requirements of Corollary 2.6 only requires Taylor expansion of the mgf, which has an explicit expression in all models of practical interest. Higher order expansions can be obtained by the same proof technique, if desired. They will follow in a relatively straightforward way from higher order expansions of the mgf, by collecting further residues of the Mellin transform. In future work, we hope to connect our assumptions on the mgf with properties of the Lévy triplet, which should give additional insight on how the slope depends on model characteristics.

Appendix

Appendix A Proofs of Lemmas 2.4 and 2.7

Proof of Lemma 2.4. Since $S = e^X$ is a martingale, we have $\psi'(0) = \mathbb{E}[X_1] < 0$. Then $\psi(0) = 0$ implies that $\psi(a) < 0$ for all sufficiently small $a > 0$. In fact, it easily follows from $\psi(1) = 0$ and the concavity of ψ that all $a \in (0,1)$ satisfy $\psi(a) < 0$. Let us fix such an a. From

$$\operatorname{Re}(-\psi(a+iy)) = -\psi(a) + \frac{1}{2}\sigma^2 y^2 + \int_{\mathbb{R}} e^{ax} \underbrace{(1-\cos(yx))}_{\geq 0} \nu(dx)$$

we obtain that the function $h(y) := -\psi(a+iy)$, $y \geq 0$, satisfies

$$\operatorname{Re} h(y) > \tfrac{1}{2}\sigma^2 y^2 \geq 0, \quad y \geq 0. \tag{3.1}$$

For $0 < \operatorname{Re}(s) < \frac{1}{2}$ define the function

$$g(T) = T^{\operatorname{Re}(s)-1} \int_0^\infty \frac{e^{-T\operatorname{Re}(h(y))}}{|a+iy|}\, dy, \quad T > 0.$$

Using Fubini's theorem and substituting $T\operatorname{Re}(h(y)) = u$, we then calculate for $\operatorname{Re}(s) > 0$

$$\begin{aligned}
\int_0^\infty g(T)\, dT &= \int_0^\infty \frac{1}{|a+iy|} \int_0^\infty e^{-T\operatorname{Re}(h(y))} T^{\operatorname{Re}(s)-1}\, dT\, dy \\
&= \int_0^\infty \frac{\operatorname{Re}(h(y))^{-\operatorname{Re}(s)}}{|a+iy|} \left(\int_0^\infty e^{-u} u^{\operatorname{Re}(s)-1}\, du \right) dy \\
&= \Gamma(\operatorname{Re}(s)) \int_0^\infty \frac{\operatorname{Re}(h(y))^{-\operatorname{Re}(s)}}{|a+iy|}\, dy.
\end{aligned}$$

From (3.1), we get

$$\int_0^\infty \frac{\operatorname{Re}(h(y))^{-\operatorname{Re}(s)}}{|a+iy|}\, dy \leq (\tfrac{1}{2}\sigma^2)^{-\operatorname{Re}(s)} \int_0^\infty \frac{y^{-2\operatorname{Re}(s)}}{|a+iy|}\, dy.$$

The restriction $\operatorname{Re}(s) < \frac{1}{2}$ ensures that the last integral is finite and thus the integrability of g. Using the dominated convergence theorem and Fubini's theorem, the Mellin transform of H can now be calculated as

$$\int_0^\infty H(T) T^{s-1}\, dT = \int_0^\infty \frac{1}{a+iy} \int_0^\infty e^{-Th(y)} T^{s-1}\, dT\, dy.$$

The substitution $Th(y) = u$ gives us the result. Note that $h(y)$ is in general non-real; it is easy to see, though, that Euler's integral

$$\Gamma(s) = \int_0^\infty u^{s-1} e^{-u}\, du, \quad \mathrm{Re}(s) > 0,$$

still represents the gamma function if the integration is performed along any complex ray emanating from zero, as long as the ray stays in the right half-plane. The latter holds, since $\mathrm{Re}(h(y)) > 0$.
It remains to prove the exponential decay of the Mellin transform $\mathcal{M}H(s) = \Gamma(s)F(s)$ for large $|\,\mathrm{Im}(s)|$. First, note that

$$\begin{aligned} \mathrm{Im}\,\psi(a+iy) &= by + \sigma^2 ay + \int_{\mathbb{R}} (e^{ax}\sin xy - xy)\,\nu(dx) \\ &= \mathcal{O}(y), \quad y \to \infty, \end{aligned}$$

which together with (3.1) yields the existence of an $\varepsilon > 0$ such that $|\arg h(y)| \le \frac{1}{2}\pi - \varepsilon$ for all $y \ge 0$. We then estimate, with $\mathrm{Re}(s) \in (0, \frac{1}{2})$ fixed,

$$\begin{aligned} |F(s)| &\le \int_0^\infty \frac{e^{-\mathrm{Re}(s\log h(y))}}{|a+iy|}\, dy \\ &= \int_0^\infty \frac{e^{-\mathrm{Re}(s)\log|h(y)| + \mathrm{Im}(s)\arg h(y)}}{|a+iy|}\, dy \\ &\le e^{(\pi/2-\varepsilon)|\,\mathrm{Im}(s)|} \int_0^\infty \frac{(\frac{1}{2}\sigma^2 y^2)^{-\mathrm{Re}(s)}}{|a+iy|}\, dy. \end{aligned}$$

The integral converges, and thus this estimate is good enough, since Stirling's formula yields $|\Gamma(s)| = \exp\big(-\frac{1}{2}\pi|\,\mathrm{Im}(s)|(1+o(1))\big)$. □

Proof of Lemma 2.7. Recall that, in the proof of Theorem 2.5, we defined the following meromorphic continuation of $F(s)$, to the strip $-\tilde{\nu} - \frac{1}{2}\varepsilon < \mathrm{Re}(s) < \frac{1}{2}$:

$$\begin{aligned} &A_0(s) + \tilde{G}_1(s) + \tilde{F}_1(s), \quad -\tilde{\nu} < \mathrm{Re}(s) < \tfrac{1}{2}, \\ &A_0(s) + \tilde{G}_1(s) + \tilde{G}_2(s) + \tilde{F}_2(s), \quad -\tilde{\nu} - \tfrac{1}{2}\varepsilon < \mathrm{Re}(s) < \tfrac{1}{2}(\nu - 1). \end{aligned}$$

As noted at the end of the proof of Lemma 2.4, Stirling's formula implies $|\Gamma(s)| = \exp\big(-\frac{1}{2}\pi|\,\mathrm{Im}(s)|(1+o(1))\big)$. By (2.12), it thus suffices to argue that the continuation of $F(s)$ is $\mathcal{O}(\exp((\frac{1}{2}\pi - \varepsilon)|\,\mathrm{Im}(s)|))$ for some $\varepsilon > 0$. The functions $\tilde{G}_1$ and $\tilde{G}_2$ are clearly $\mathcal{O}(1)$. As for A_0, defined in (2.18), we have

$$\begin{aligned} |A_0(s)| &\le \int_0^{y_0} \frac{e^{-\mathrm{Re}(s\log h(y))}}{|a+iy|}\, dy \\ &= \int_0^{y_0} \frac{|h(y)|^{-\mathrm{Re}(s)} e^{\mathrm{Im}(s)\arg h(y)}}{|a+iy|}\, dy. \end{aligned}$$

Now note that

$$|h(y)|^{-\operatorname{Re}(s)} \leq \begin{cases} (\frac{1}{2}\sigma^2 y^2)^{-\operatorname{Re}(s)} & 0 < \operatorname{Re}(s) < \frac{1}{2}, \\ (\max_{0\leq y\leq y_0} |h(y)|)^{-\operatorname{Re}(s)} & \operatorname{Re}(s) \leq 0, \end{cases}$$

and that

$$\exp(\operatorname{Im}(s) \arg h(y)) \leq \exp((\tfrac{\pi}{2} - \varepsilon)|\operatorname{Im}(s)|)$$

for some $\varepsilon > 0$, as argued in the proof of Lemma 2.4.
It remains to establish a bound for $\tilde{F}_1$, defined in (2.22). (The bound for $\tilde{F}_2$ is completely analogous, and we omit the details.) In what follows, we assume that $-\tilde{\nu} < \operatorname{Re}(s) < \frac{1}{2}$. By (2.19), we have (where the $\mathcal{O}$ is uniform w.r.t. s, and $y_0 \geq 0$ is still arbitrary):

$$\begin{aligned} \tilde{F}_1(s) &= \int_{y_0}^{\infty} \frac{1}{a+iy} \left((\tfrac{1}{2}\sigma^2 y^2)^{-s}(1+\mathcal{O}(y^{\nu-2}))^{-s} - (\tfrac{1}{2}\sigma^2 y^2)^{-s} \right) dy \\ &= \int_{y_0}^{\infty} \frac{1}{a+iy} (\tfrac{1}{2}\sigma^2 y^2)^{-s} \left((1+\mathcal{O}(y^{\nu-2}))^{-s} - 1 \right) dy. \end{aligned} \tag{3.2}$$

We now choose y_0 such that, for some constant $C_0 > 0$,

$$\begin{aligned} \big| \log|1+\mathcal{O}(y^{\nu-2})| \big| &\leq \tfrac{1}{4}\pi, \\ \big| \arg(1+\mathcal{O}(y^{\nu-2})) \big| &\leq \tfrac{1}{4}\pi, \\ \big| \log(1+\mathcal{O}(y^{\nu-2})) \big| &\leq C_0 y^{\nu-2}, \end{aligned}$$

hold for all $y \geq y_0$. (By a slight abuse of notation, here $\mathcal{O}(y^{\nu-2})$ of course denotes the function hiding behind the $\mathcal{O}(y^{\nu-2})$ in (3.2).) For all $w \in \mathbb{C}$ we have the estimate

$$|e^w - 1| \leq |w| e^{|\operatorname{Re}(w)|}.$$

Using this in (3.2), we find

$$\begin{aligned} \big|(1+\mathcal{O}(y^{\nu-2}))^{-s} - 1\big| &= \big|\exp(-s\log(1+\mathcal{O}(y^{\nu-2}))) - 1\big| \\ &\leq |s\log(1+\mathcal{O}(y^{\nu-2}))| \cdot \exp(|\operatorname{Re}(s\log(1+\mathcal{O}(y^{\nu-2})))|) \\ &\leq C_1 |s| y^{\nu-2} \exp(\tfrac{1}{4}\pi|\operatorname{Im}(s)|), \end{aligned}$$

where $C_1 = C_0 \exp(\frac{1}{4}\pi \sup_s |\operatorname{Re}(s)|)$, and thus

$$\begin{aligned} |\tilde{F}_1(s)| &\leq C_2 |s| e^{\frac{1}{4}\pi|\operatorname{Im}(s)|} \int_{y_0}^{\infty} y^{-2\operatorname{Re}(s)+\nu-3}\, dy \\ &= \exp\left(\tfrac{1}{4}\pi|\operatorname{Im}(s)|(1+o(1))\right). \end{aligned}$$

$\square$

Appendix B Implied Volatility Slope in the CGMY Model

This additional section is not part of the paper [53] on which the whole Part I is based. It proves the at-the-money implied volatility slope result in Corollary 2.6 for the CGMY model with parameter $Y \in (1,2)$. Note that in this model $\nu = Y$ and $C_{\tilde{\nu}} = 0$, and therefore Corollary 2.6 (ii) is not applicable.
The following lemma helps us to compute the integrals in the proof of Theorem B.2.

Lemma B.1. *Suppose there are constants $a \in (0,1)$, $\beta \geq 0$ and $y_0 \geq 0$. Then the following result holds for* $\mathrm{Re}(s) > -\beta/2$

$$\int_{y_0}^{\infty} \frac{(-z^2)^{-s}}{z^{1+\beta}}\Big|_{z=a+iy} dy = e^{i\pi s} \frac{(a+iy_0)^{-2(s+\beta/2)}}{2i(s+\beta/2)}. \tag{3.3}$$

Proof. Let $\mathbb{C}^* := \mathbb{C} \backslash (-\infty, 0]$ and consider the principal branch of the complex logarithm with $\arg(z) := \mathrm{Im}(\log(z)) \in (-\pi, \pi)$ for $z \in \mathbb{C}^*$. If $\arg(z) \in (0,\pi)$, then for a complex number w we have, using the principal branch,

$$(-z)^w = z^w e^{-i\pi w}.$$

Note that $\arg((a+iy)^2) \in (0,\pi)$ for $y > 0$. Now we can calculate the integral

$$\begin{aligned}\int_{y_0}^{\infty} \frac{(-(a+iy)^2)^{-s}}{(a+iy)^{1+\beta}} dy &= e^{i\pi s} \int_{y_0}^{\infty} (a+iy)^{-2s-\beta-1} dy \\ &= e^{i\pi s} \frac{(a+iy_0)^{-2(s+\beta/2)}}{2i(s+\beta/2)},\end{aligned}$$

and the last equality holds if and only if $2\,\mathrm{Re}(s) + \beta > 0$. □

Theorem B.2. *For $b \neq 0$ and $\sigma > 0$, the ATM digital price in the CGMY model with $Y \in (1,2)$ satisfies*

$$\mathbb{P}[X_T \geq 0] = \frac{1}{2} + \frac{b}{\sigma\sqrt{2\pi}}\sqrt{T} + o(\sqrt{T}), \quad T \to 0.$$

The implied volatility slope converges:

$$\begin{aligned}\lim_{T\to 0} \partial_K|_{K=1}\sigma_{\mathrm{imp}}(K,T) &= -\frac{b}{\sigma} - \frac{\sigma}{2} \\ &= \frac{1}{\sigma} C\Gamma(-Y)((M-1)^Y - M^Y + (G+1)^Y - G^Y).\end{aligned}$$

Proof. The Laplace exponent of the CGMY model is

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + bz + C\Gamma(-Y)((M-z)^Y - M^Y + (G+z)^Y - G^Y).$$

From the expansion, for $\mathrm{Re}(z) = a$, $\mathrm{Im}(z) \to \infty$,

$$\begin{aligned}(M-z)^Y &= e^{-i\pi Y} z^Y + \mathcal{O}(z^{Y-1}),\\ (G+z)^Y &= z^Y + \mathcal{O}(z^{Y-1}),\end{aligned}$$

we have

$$\psi(z) = \tfrac{1}{2}\sigma^2 z^2 + cz^Y + bz + \mathcal{O}(z^{Y-1}), \quad \mathrm{Re}(z) = a, \mathrm{Im}(z) \to \infty,$$

with the complex constant $c := C\Gamma(-Y)(1 + e^{-i\pi Y})$.
Observing the constant $C_{\tilde{\nu}}$ in Theorem 2.5 with $\nu = Y$ and $\tilde{\nu} = 1 - Y/2$, we see that

$$e^{-i\pi\tilde{\nu}} c = -2C\Gamma(-Y)\cos(\pi Y/2) \in \mathbb{R}, \tag{3.4}$$

and thus $C_{\tilde{\nu}} = 0$.
To find the second order term, we proceed as in the proof of Theorem 2.5. Let F, $\tilde{F}$ and A_0 be as in 2.18 and define the convergence-inducing function

$$G_1(s) := (\tfrac{1}{2}\sigma^2(a+iy_0)^2)^{-s} \frac{e^{i\pi s}}{2si}.$$

which is clearly analytic on the whole complex plane except for the single pole at $s = 0$. On the half-plane $\mathrm{Re}(s) > 0$, we compute with Lemma B.1

$$G_1(s) = \int_{y_0}^{\infty} \frac{(-\frac{1}{2}\sigma^2 z^2)^{-s}}{z}\Big|_{z=a+iy} dy \tag{3.5}$$

Note that G_1 in (3.5) is slightly different from $\tilde{G}_1$ defined in 2.20. The advantage of this convergence-inducing function G_1 is the ability of a direct computation of the integral with Lemma (B.1) and the presence of just one singularity.
With $N := \max\{n \in \mathbb{N} \colon n(Y-2) > -1\}$ we expand the function $(-\psi(z))^{-s}$, using the generalized binomial theorem,

$$\begin{aligned}(-\psi(z))^{-s} &= (-\tfrac{1}{2}\sigma^2 z^2)^{-s} \left(1 + \frac{2c}{\sigma^2} z^{Y-2} + \mathcal{O}(z^{-1})\right)^{-s}\\ &= (-\tfrac{1}{2}\sigma^2 z^2)^{-s} \sum_{n=0}^{\infty} \binom{-s}{n} \left(\frac{2c}{\sigma^2} z^{Y-2} + \mathcal{O}(z^{-1})\right)^n\\ &= (-\tfrac{1}{2}\sigma^2 z^2)^{-s} \left(1 + \sum_{n=1}^{N} c_n(s) z^{n(Y-2)} + \mathcal{O}(z^{-1})\right)\end{aligned} \tag{3.6}$$

for $\mathrm{Re}(z) = a, \mathrm{Im}(z) \to \infty$, with the polynomial $c_n(s) = \binom{-s}{n}(2c/\sigma^2)^n$.
For $n \in \{1, \ldots, N\}$ we define the convergence-inducing function

$$G_{n+1}(s) := c_n(s)(\tfrac{1}{2}\sigma^2)^{-s} e^{i\pi s} \frac{(a+iy_0)^{-2(s+n\tilde{\nu})}}{2i(s+n\tilde{\nu})},$$

which is analytic on the whole complex plane except for a single pole at $s = -n\tilde{\nu}$. On the half-plane $\mathrm{Re}(s) > n(Y-2)/2 = -n\tilde{\nu}$, we compute with Lemma B.1

$$G_{n+1}(s) = c_n(s) \int_{y_0}^{\infty} \frac{(-\frac{1}{2}\sigma^2 z^2)^{-s}}{z^{1-n(Y-2)}}\Big|_{z=a+iy} dy. \tag{3.7}$$

With $\tilde{F}$ from (2.13), the integral representation (3.5) and (3.7) and the expansion (3.6), the compensated function $F_{n+1}(s) := \tilde{F}(s) - \sum_{i=0}^{n} G_{i+1}(s)$ is analytic at least for $\mathrm{Re}(s) > \max\{-(n+1)\tilde{\nu}, -\frac{1}{2}\}$. Clearly, the identity

$$\tilde{F}(s) = F_{n+1}(s) + \sum_{i=0}^{n} G_{i+1}(s)$$

holds in the half-plane $\mathrm{Re}(s) > 0$, and thus $\tilde{F}$ is continued meromorphically to the half-plane $\mathrm{Re}(s) > \max\{-(n+1)\tilde{\nu}, -\frac{1}{2}\}$. For $n \in \{1, \ldots, N\}$ the residue of $\mathcal{M}H$ at the simple pole $s = -n\tilde{\nu}$ is

$$\mathrm{Res}_{s=-n\tilde{\nu}}(\mathcal{M}H)(s)T^{-s} = \mathrm{Res}_{s=-n\tilde{\nu}}\, G_{n+1}(s)\Gamma(s)T^{-s} \tag{3.8}$$

$$= \binom{n\tilde{\nu}}{n}(\tfrac{1}{2}\sigma^2)^{n(\tilde{\nu}-1)} \frac{(e^{-i\pi\tilde{\nu}}c)^n}{2i}\Gamma(-n\tilde{\nu})T^{n\tilde{\nu}} \in i\mathbb{R}, \tag{3.9}$$

because of the real constant (3.4). So the real part of all the poles $-n\tilde{\nu}$ with $n \in \{1, \ldots, N\}$ is zero and therefore they give no contribution to the digital price in the CGMY model.
We have to expand $(-\psi(z))^{-s}$ further. For $\mathrm{Re}(z) = a, \mathrm{Im}(z) \to \infty$,

$$(-\psi(z))^{-s} = (-\tfrac{1}{2}\sigma^2 z^2)^{-s}\left(1 + \sum_{n=1}^{N} c_n(s) z^{n(Y-2)} + d(s)z^{-1} + \mathcal{O}(z^{-1-\varepsilon})\right),$$

with some $\varepsilon > 0$ and

$$d(s) = \begin{cases} (-s)2b/\sigma^2, & \text{if } (N+1)(Y-2) \neq -1, \\ (-s)2b/\sigma^2 + c_{N+1}(s), & \text{if } (N+1)(Y-2) = -1. \end{cases}$$

Continuing our procedure, we define the meromorphic function

$$G_{N+2}(s) := d(s)(\tfrac{1}{2}\sigma^2)^{-s} e^{i\pi s} \frac{(a+iy_0)^{-2(s+1/2)}}{2i(s+1/2)}$$

which has the integral representation

$$G_{N+2}(s) = d(s) \int_{y_0}^{\infty} \frac{(-\frac{1}{2}\sigma^2 z^2)^{-s}}{z^2}\Big|_{z=a+iy} dy$$

for $\mathrm{Re}(s) > -\frac{1}{2}$. The compensated function $F_{N+2}(s) = \tilde{F}(s) - \sum_{i=0}^{N+1} G_{i+1}(s)$ is analytic for $\mathrm{Re}(s) > (-1-\varepsilon)/2$ and clearly

$$\tilde{F}(s) = F_{N+2}(s) + G_{n+2}(s)$$

is a meromorphic continuation of $\tilde{F}$ on the half-plane $\mathrm{Re}(s) > (-1-\varepsilon)/2$. With analogous calculations as in (3.8) the residue at $s = -\frac{1}{2}$ is

$$\frac{1}{\pi}\,\mathrm{Re}\big(\mathrm{Res}_{s=-1/2}(\mathcal{M}H)(s)T^{-s}\big) = \frac{b}{\sigma\sqrt{2\pi}}\sqrt{T}$$

At last, we want to mention that with similar arguments as in Lemma 2.7 a $y_0 \geq 0$ can be found such that this meromorphic continuation of $\mathcal{M}H$ decays exponentially as $|\,\mathrm{Im}(s)| \to \infty$. $\square$

Part II

Option Pricing in the Moderate Deviations Regime

Chapter 1

Introduction

Consider a European call option struck at K with remaining time to expiry $t > 0$ and no-arbitrage price[1] $C(K, t)$. Today's price of the underlying, the spot value S_0, is known and fixed. Discrete option data are available from the market, typically quoted in (Black-Scholes) implied volatilities, see Figure 1.1 below. Many option pricing models have been proposed to combine reasonable dynamics for the underlying, small number of parameters and acceptable fits to the data. However, with the notable exception of the Black-Scholes model, closed form expressions for call prices are scarce, and approximate pricing formulae have been proposed as substitute: often used to improve calibration, but also towards a better quantitative understanding of a given model. (A classic reference in this context is Gatheral [50].)

More specifically, small-maturity approximations of option prices have been studied extensively in recent years. Starting with Carr and Wu [18], it was understood that the asymptotic behaviour of $C(K, t)$ as $t \downarrow 0$ exhibits very different behaviour in the respective cases $K > S_0$ ("out-of-the-money") and $K = S_0$ ("at-the-money"). We argue that there is a significant asymptotic regime in between, namely

$$\sqrt{t} \ll K - S_0 \ll 1.$$

It has received little attention, and, to the best of our knowledge, none at all in the classical diffusion case. The aim of Part II is to fill this gap. This "moderately out-of-the-money" regime in fact reflects the reality of quoted option prices: as seen in Figure 1.1, the range of strikes tends to concentrate "around-the-money" as time to expiry becomes small. At the same time, the regime offers excellent analytic tractability.

To put our results into perspective, we recall some well-known facts on option price approximations close to expiry. We write $c(k, t)$ for the normalized call price as a function of log-moneyness $k = \log(K/S_0)$

$$C(S_0 e^k, t)/S_0 = c(k, t). \tag{1.1}$$

In general, $c(k, t)$ depends tacitly on S_0, the (fixed) spot value.[2]

[1] As we focus on stochastic volatility models, which are in general incomplete, it is understood that call prices are computed w.r.t. some fixed pricing measure.

[2] There is no spot-dependence of the normalized call price in the Black-Scholes model. This holds true, more generally, whenever dynamics for the log-price $X = \log(S/S_0)$ are specified without further spot dependence; this includes the Heston model and many other stochastic volatility models.

We start with the **at-the-money** (short: **ATM**) regime $k = 0$. In the Black-Scholes model, writing $c(k,t) = c_{\text{BS}}(0,t;\sigma)$ with volatility parameter $\sigma > 0$, we have the following ATM call price behaviour

$$c_{\text{BS}}(0,t;\sigma) \sim \frac{\sigma\sqrt{t}}{\sqrt{2\pi}}, \quad t \downarrow 0.$$

From Muhle-Karbe and Nutz [71], the same is actually true in a generic semimartingale model with diffusive component (with spot volatility $\sigma_0 = \sqrt{v_0} > 0$),

$$c(0,t) \sim \frac{\sigma_0\sqrt{t}}{\sqrt{2\pi}}, \quad t \downarrow 0, \tag{1.2}$$

and this translates to the generic ATM implied variance formula (even in presence of jumps, as long as $v_0 > 0$)

$$\sigma^2_{\text{imp}}(0,t) = v_0 + o(1), \quad t \downarrow 0.$$

(We use the notation $\sigma_{\text{imp}}(k,t)$ for the Black-Scholes implied volatility with log-moneyness k and maturity t.) Higher order terms in t will be model dependent. For instance, in the Heston case, with variance dynamics $dV_t = \lambda(\bar{v} - V_t)\,dt + \xi\sqrt{V_t}\,dW_t$, implied variance has the ATM expansion

$$\begin{aligned} \sigma^2_{\text{imp}}(0,t) &= v_0 + a(0)t + o(t), \quad t \downarrow 0, \\ a(0) &= -\frac{\xi^2}{12}\left(1 - \frac{\rho^2}{4}\right) + \frac{v_0\rho\xi}{4} + \frac{\lambda}{2}(\bar{v} - v_0). \end{aligned} \tag{1.3}$$

This is Corollary 4.4 in Forde, Jacquier and Lee [41], and we note that $a(0)$ has no easy interpretation in terms of the model parameters.

Relaxing $k = 0$ to $k_t = o(\sqrt{t})$ amounts to what we dub "**almost-ATM**" (short: **AATM**) regime.[3] (In particular, $k_t \sim t^\beta$ is in the AATM regime if and only if $\beta > 1/2$.) Again for generic semimartingale models with diffusive component and spot volatility $\sigma_0 > 0$, it is easy to see from Caravenna and Corbetta [17] and Muhle-Karbe and Nutz [71] that the ATM asymptotics (1.2) imply the almost-ATM asymptotics

$$c(k_t,t) \sim \frac{\sigma_0\sqrt{t}}{\sqrt{2\pi}}, \quad k_t = o(\sqrt{t}), \quad t \downarrow 0.$$

This fails when k_t ceases to be $o(\sqrt{t})$. Indeed, for $k_t = \theta\sqrt{t}$ with constant factor $\theta > 0$, we have, from Caravenna and Corbetta [17] and Muhle-Karbe and Nutz [71],

$$c(k_t,t) \sim \mathbb{E}[N(-\theta,\sigma_0^2)^+]\sqrt{t}, \quad t \downarrow 0,$$

[3] The term "almost-ATM" seems new, but this regime was considered by a number of authors including Caravenna and Corbetta [17] and Muhle-Karbe and Nutz [71].

where $N(-\theta, \sigma_0^2)$ stands for a Gaussian random variable with mean $-\theta$ and variance σ_0^2. This, too, holds true in the stated semimartingale generality. In any case, the proof is based on the Lévy case with non-zero diffusity v_0, and the result follows from comparison results which imply that the difference is negligible to first order. For a thorough discussion of the regime $k = \mathcal{O}(\sqrt{t})$ in the (local) diffusion case, see Pagliarani and Pascucci [78].

Beyond this regime, call price asymptotics change considerably. For instance, take an additional slowly diverging factor $\log(1/t)$,

$$k_t = \theta\sqrt{t\log(1/t)}.$$

Even in the Black-Scholes model, we now loose the $\sqrt{t}$-behaviour of call prices described above and in fact

$$c_{\mathrm{BS}}(k_t, t; \sigma) = t^{\frac{1}{2}+\frac{\theta^2}{2\sigma^2}}\ell(t),$$

for some slowly varying function $\ell(t)$, see Mijatović and Tankov [70]. On the other hand, in a genuine **out-of-the-money** (short: **OTM**) situation, with $k_t \equiv k > 0$ fixed, option values are exponentially small in diffusion models, and we are in the realm of *large deviations theory*. For instance,

$$c_{\mathrm{BS}}(k, t; \sigma) = \exp\left(-\frac{\Lambda_{\mathrm{BS}}(k)}{t}(1 + o(1))\right), \quad k > 0 \text{ fixed}, \ t \downarrow 0,$$

with $\Lambda_{\mathrm{BS}}(k) = \frac{1}{2}k^2/\sigma^2$ in the Black-Scholes model. Similar results appear in the literature, with different levels of mathematical rigor, for other and/or generic diffusion models, see Berestycki, Busca and Florent [9], Carr and Wu [18], Forde and Jacquier [38] and Paulot [79].

Throughout, we reserve the term out-of-the-money (OTM) for *fixed* OTM log-strike $k > 0$, to distinguish this regime from the *moderately* out-of-the-money regime that we now define. Our basic observation is that for

$$k_t \sim (const)t^{\beta}, \quad t \downarrow 0, \tag{1.4}$$

the cases of $\beta > \frac{1}{2}$, resp. $\beta = 0$, are covered by the afore-discussed AATM, resp. OTM, results. This leaves open a significant gap, namely $\beta \in (0, \frac{1}{2})$, which we call **moderately out-of-the-money** (short: **MOTM**). We have a threefold interest in this MOTM regime,

$$k_t \sim (const)t^{\beta}, \quad t \downarrow 0, \quad \text{for } \beta \in (0, \tfrac{1}{2}). \tag{1.5}$$

(i) First, it is related to the *reality of quoted (short-dated) option prices*, where strikes of option price data with acceptable bid-ask spreads tend to accumulate “around the

[4]This is also true for Lévy models with a Brownian component and a finite variation jump part. For Lévy models with jump part of infinite variation, the call price still decays algebraically, but slower than $\mathcal{O}(t)$. See Theorem 1 and Proposition 2 in Mijatović and Tankov [70].

Table 1.1: Asymptotic behaviour of short-maturity call options, $t \downarrow 0$

Process Type	**ATM** (at-the-money) $K = S_0$	**AATM** (almost at-the-money) $\log \frac{K}{S_0} \sim (const)t^{\beta}$, $\beta > 1/2$	**MOTM** (moderately out-of-the-money) $\log \frac{K}{S_0} \sim (const)t^{\beta}$, $0 < \beta < 1/2$	**OTM** (out-of-the-money) $\log \frac{K}{S_0} \equiv k > 0$
Black-Scholes	$\mathcal{O}(\sqrt{t})$, elementary	$\mathcal{O}(\sqrt{t})$, elementary	$\exp\left(-\frac{const}{t^{1-2\beta}}\right)$, elementary	$\exp\left(-\frac{const}{t}\right)$, elementary
Stochastic or local volatility (diffusion model)	$\mathcal{O}(\sqrt{t})$	$\mathcal{O}(\sqrt{t})$	$\exp\left(-\frac{const}{t^{1-2\beta}}\right)$	$\exp\left(-\frac{const}{t}\right)$
Jump diffusion, general semi-martingale with diff. component	$\mathcal{O}(\sqrt{t})$	$\mathcal{O}(\sqrt{t})$	$\mathcal{O}(t)$ in finite variation Lévy models[4]	$\mathcal{O}(t)$, see Bentata and Cont [8]

money", as illustrated in Figure 1.1. To account for this accumulation, we consider strikes that move closer to the money as expiry shrinks, and the simplest way to do so is to consider strikes of the order $k = \mathcal{O}(t^{\beta})$ for some $\beta > 0$. There is no reason why quoted strikes should always be almost-ATM ($\beta > 1/2$), which effectively means an extreme concentration around the money; we are thus led to study the regime (1.5).

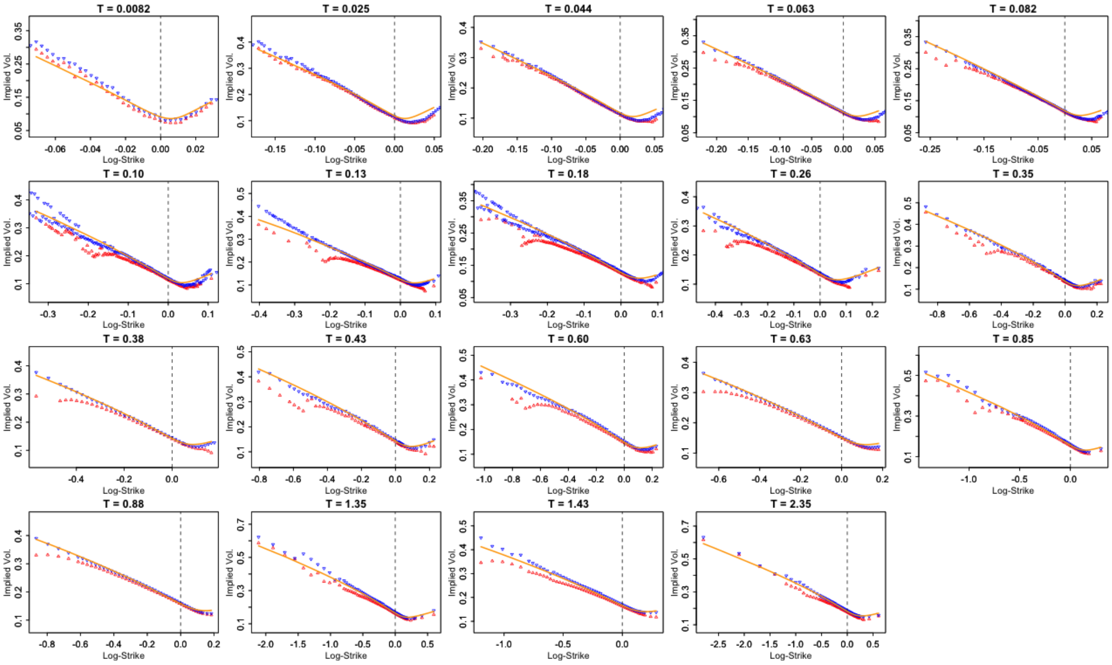

Figure 1.1: SPX volatility smiles as of 14-Aug-2013 (courtesy of J. Gatheral). Strikes of options with small remaining time to maturity ($T = 0.0082$) are about $e^{0.02} - 1 \approx 2\%$ around the money (spot); good data for a later time $T = 0.26$ already have a range of $\approx 30\%$. The highest maturity $T = 2.35$ has a range of $\approx 65\%$ around the money.

(ii) The second reason is *mathematical convenience*. In contrast to the genuine OTM regime (large deviation regime) in which the rate function $\Lambda(k)$ is notoriously difficult to analyse – often related to geodesic distance problems – MOTM naturally comes with a quadratic rate function and, most remarkably, higher order expansions are always explicitly computable in terms of the model parameters. The terminology *moderately-*OTM (MOTM) is in fact in reference to *moderate deviations theory*, which effectively interpolates between the central limit and large deviations regimes.[5] This also identifies the AATM regime as bordering the central limit regime, where asymptotics are precisely those of the Black-Scholes model, which in turn is the rescaled Gaussian (in log-coordinates) limit of a general semimartingale model with diffusive component.

(iii) Finally, our third point is that MOTM expansions naturally involve quantities very familiar to practitioners, notably spot (implied) volatility, implied volatility skew and so on.

[5]We write CLT for central limit theorem and LDP for large deviation principle. For readers unfamiliar with moderate deviations, we recall some of the basics towards the end of the introduction.

In the Black-Scholes model, it is easy to check that we have the MOTM asymptotics

$$c_{\mathrm{BS}}(k_t \equiv \theta t^{\beta}, t; \sigma) = \exp\left(- \frac{\theta^2}{2\sigma^2 t^{1-2\beta}} \big(1 + o(1)\big) \right), \quad t \downarrow 0.$$

Loosely speaking, our main results (Theorems 2.3 and 2.5 below) assert that such relations (even of higher order) are true in great generality for diffusion models, and that all quantities are computable and then related to implied volatility expansions. We note in passing that, for Lévy models, the regime (1.5) has been studied by Mijavtović and Tankov [70]; then, call prices decay algebraically rather than exponentially. For recent related results on *fractional* stochastic volatility models, see Forde and Zhang [42] and Guennoun, Jacquier and Roome [55]. Guillin [56], which considers small-noise moderate deviations of diffusions, should also be mentioned here; however, in Guillin [56] the dynamics depend on a "fast" random environment (with motivation from physics, and no obvious financial interpretation), and the non-degeneracy assumption (D) is not satisfied in our context. The recent related paper by Gao and Wang [48] contains a *first order* moderate deviation principle (MDP) for diffusions under classical regularity assumptions from SDE theory. The main difference to the bulk of our results is that we develop *higher order* expansions, until Section 2.5 where we revisit first order MOTM estimates from a moment-generating function perspective. However, in this case the underlying models (e.g. Heston) fall immediately outside the scope of Gao and Wang [48], because of the typical square-root structure of coefficients. (The matter is discussed in more detail at the beginning of Section 2.5.)

To round off the introduction, we briefly recall some background on moderate deviations. Consider the classical setting of a centered i.i.d. sequence $(X_n)_{n\geq 1}$ with finite exponential moments. Then the empirical means $\hat{X}_n := n^{-1} \sum_{k=1}^n X_k$ converge to zero (law of large numbers, LLN), and this is quantified by an LDP according to Cramér's classical theorem: $\mathbb{P}[\hat{X}_n > x] = \exp(-I(x)n + o(n))$ decreases exponentially as $n \to \infty$ for fixed $x > 0$, governed by a rate function $I(x) = \sup_{y \in \mathbb{R}} \big(yx - \log \mathbb{E}[e^{yX_1}]\big)$. On the other hand, by the CLT, $\sqrt{n}\hat{X}_n = n^{-1/2} \sum_{k=1}^n X_k$ has a Gaussian limit law. Moderate deviations cover intermediate scalings, i.e. $\sqrt{na_n}\hat{X}_n$ with $a_n \to 0$ and $na_n \to \infty$. It turns out (Theorem 3.7.1 in Dembo and Zeitouni [23]) that, for any such sequence $a_n > 0$, the family $\sqrt{na_n}\hat{X}_n$ satisfies an LDP with speed $1/a_n$ and quadratic rate function. (A natural scaling family is given by $a_n = n^{2\beta-1}$, with parameter $\beta > 0$, so that one considers $\sqrt{na_n}\hat{X}_n = n^{\beta-1} \sum_{k=1}^n X_k$. Interpolation between LDP, with LLN scaling, and CLT scaling then amounts to considering $0 < \beta < 1/2$.) This is sometimes called a *moderate deviation principle (MDP)*. Formally, an MDP is thus just a certain LDP with appropriate scaling and speed function. Still, the terminology is often useful because of the trichotomy

CLT – MDP for a range of scalings, with quadratic rate function – genuine LDP,

which occurs in many situations, not just for i.i.d. sequences of random variables. For references to some other classical moderate deviations results (e.g. on empirical measures) see Sections 6.7 and 7.4 of Dembo and Zeitouni [23]. Several authors have investigated moderate deviations in actuarial risk theory, see e.g. Fu and Shen [47] and references therein.

The rest of Part II is organized as follows. Section 2.1 contains our main results, which translate asymptotics for the transition density of the underlying into MOTM call price asymptotics. The corresponding proofs are presented in Section 2.2. Section 2.3 and Appendix A give the implied volatility expansion resulting from our call price approximations. Section 2.4 applies our main results to standard examples, namely generic local volatility models (Subsection 2.4.1), generic stochastic volatility models (Subsection 2.4.2), and the Heston model (Subsection 2.4.3). As usual, the square-root degeneracy of the Heston model makes it difficult to apply results for general stochastic volatility models, so we verify the validity of our results – if formally applied to Heston – by a direct "affine" analysis. Finally, in Section 2.5 we present a second approach at MOTM estimates, which employs the Gärtner-Ellis theorem from large deviations theory. Throughout we take zero rates, a natural simplification in view of our short-time consideration. Also, w.l.o.g. we normalize spot to $S_0 = 1$.

Chapter 2

The Moderate Deviations Regime

2.1 MOTM Option Prices via Density Asymptotics

We consider a *general stochastic volatility* model, i.e. a positive martingale $(S_t)_{t\geq 0}$ with dynamics

$$dS_t = S_t \sigma_t \, dW_t$$

and started (w.l.o.g.) at $S_0 = 1$. We assume that the stochastic volatility process $(\sigma_t)_{t\geq 0}$ itself is an Itô-diffusion, started at some deterministic value σ_0, called *spot volatility*. Recall that in any such stochastic volatility model, the local (or effective) volatility is defined by

$$\sigma^2_{\mathrm{loc}}(t,K) := \mathbb{E}[\sigma_t^2 | S_t = K].$$

As is well-known, the equivalent local volatility model

$$d\tilde{S}_t = \tilde{S}_t \sigma_{\mathrm{loc}}(t, \tilde{S}_t) \, dW_t$$

has the property that $\tilde{S}_t = S_t$ (in law) for all fixed times. See Brunick and Shreve [13] for precise technical conditions under which this holds true.[1] As a consequence, European option prices $C(K,t)$ match in both models. Recall also Dupire's formula in this context

$$\sigma^2_{\mathrm{loc}}(K,t) = \frac{\partial_t C(K,t)}{\frac{1}{2}K^2 \partial_{KK} C(K,t)}, \quad t > 0, K > 0. \tag{2.1}$$

We now state our two crucial conditions.

Assumption 2.1. For all $t > 0$, S_t has a continuous pdf $K \mapsto q(K,t)$, which behaves asymptotically as follows for small time

$$q(K,t) \sim e^{-\frac{\Lambda(k)}{t}} t^{-1/2} \gamma(k), \quad t \downarrow 0, \tag{2.2}$$

uniformly for $K = e^k$ in some neighbourhood of $S_0 = 1$. The energy function Λ is smooth in some neighbourhood of zero, with initial value $\Lambda(0) = \Lambda'(0) = 0$. Moreover, $\lim_{k\to 0} \gamma(k) = \gamma(0) > 0$.

[1] The situation is very different with jumps, see Friz, Gerhold and Yor [46].

Assumption 2.2. For $t \downarrow 0$ and $K \to S_0 = 1$, the local volatility function of $(S_t)_{t \geq 0}$ converges to spot volatility

$$\lim_{\substack{K \to S_0 \\ t \downarrow 0}} \sigma_{\text{loc}}(K, t) = \sigma_0. \tag{2.3}$$

The latter assumption is fairly harmless (in diffusion models; see the beginning of Section 2.4.2). The first assumption is potentially (very) difficult to check, but fortunately we can rely on substantial recent progress in this direction, see Deuschel, Friz, Jacquier and Violante [24], [25] and Osajima [77]. We shall see in Section 2.4.2 that both assumptions indeed hold in generic stochastic volatility models. Let us also note the fundamental relation between spot-volatility σ_0 (actually equal to implied spot volatility $\sigma_{\text{imp}}(0,0)$ here) and the Hessian of the energy function $\Lambda = \Lambda(k)$,

$$\sigma_0 = \Lambda''(0)^{-1/2}.$$

(This is well-known, see Durrleman [27], and also follows from Proposition 2.4 below.) Now we state our main result. We slightly generalize the log-strikes considered in (1.5), replacing the constant factor by an arbitrary slowly varying function ℓ.

Theorem 2.3. *Under Assumptions 2.1 and 2.2, consider moderately out-of-the-money calls, in the sense that log-strike is*

$$k_t = t^\beta \ell(t), \quad t > 0, \tag{2.4}$$

where $\ell > 0$ varies slowly at zero and $\beta \in (0, \frac{1}{2})$.

(i) The call price satisfies the moderate deviation estimate

$$\begin{aligned} c(k_t, t) &= \exp\Big(-\frac{\Lambda''(0)}{2} \frac{k_t^2}{t} (1 + o(1)) \Big) \\ &= \exp\Big(-\frac{1}{2\sigma_0^2} \frac{k_t^2}{t} (1 + o(1)) \Big), \quad t \downarrow 0. \end{aligned} \tag{2.5}$$

(ii) If we restrict β to $(0, \frac{1}{3})$, then the following moderate second order expansion holds true

$$\begin{aligned} c(k_t, t) &= \exp\Big(-\frac{1}{2}\Lambda''(0)\frac{k_t^2}{t} - \frac{1}{6}\Lambda'''(0)\frac{k_t^3}{t}(1 + o(1)) \Big) \\ &= \exp\Big(-\frac{1}{2\sigma_0^2}\frac{k_t^2}{t}\Big(1 - \frac{\mathcal{S}}{\sigma_0^2}k_t(1 + o(1))\Big) \Big), \quad t \downarrow 0, \end{aligned} \tag{2.6}$$

with spot-variance σ_0^2, equal to $\sigma_{\text{imp}}^2(0,0)$, and implied variance skew

$$\mathcal{S} = \frac{\partial}{\partial k}\Big|_{k=0} \sigma_{\text{imp}}^2(k, 0).$$

In particular, for $\ell \equiv \theta > 0$, we have the (first order) expansion

$$t^{1-2\beta} \log c(\theta t^\beta, t) \sim -\frac{\theta^2}{2\sigma_0^2}, \quad t \downarrow 0,$$

exhibiting a quadratic rate function $\theta \mapsto \theta^2/2\sigma_0^2$, typical of *moderate deviation problems.*[2]

In a nutshell, (2.5) says that inserting the time-dependent log-strike (2.4) into the fixed-strike OTM/LD approximation $c(k,t) = \exp\big(-\frac{\Lambda(k)}{t}(1+o(1))\big)$ yields a correct result, upon Taylor expanding Λ. Mind however, that this needs a proof using the specifics of our situation, in light of the fact that validity of a large deviation principle does not automatically imply a moderate deviation principle.
The quantities $\Lambda''(0), \Lambda'''(0), \ldots$ appearing above are *always computable* from the initial values and the diffusion coefficients of the stochastic volatility model. This is in stark contrast to the OTM regime, where one needs the function Λ, which is in general not available in closed form (with some famous exceptions, like the SABR model). We quote the following result on N-factor models from Osajima [77], and refer to Section 2.4.2 for detailed calculations in a two-factor stochastic volatility model.

Proposition 2.4. *Assume that $(\log S, \sigma^1, \ldots, \sigma^{N-1})$ is Markov, started at $(0, \bar\sigma_0)$ with $\bar\sigma_0 \in \mathbb{R}^{N-1}$ and $\bar\sigma_0^1 > 0$, with stochastic volatility $\sigma \equiv \sigma^1$, where the generator has (non-degenerate) principal part $\sum a^{ij}\partial_{ij}$ in the sense that a^{-1} defines a Riemannian metric. Then*

$$\Lambda(k) = \frac{1}{2b_1}k^2 - \frac{b_2}{3b_1^3}k^3 + \Big(-\frac{b_3}{4b_1^4} + \frac{b_2^2}{2b_1^5}\Big)k^4 + \mathcal{O}(k^5), \quad k \to 0,$$

where the coefficients are given by

$$\begin{aligned} b_1 &= \int_0^1 a^{11}(t, \bar\sigma_0)\,dt \\ b_2 &= \frac{3}{2}\int_0^1 (Va^{11})(t, \bar\sigma_0)\,dt \\ b_3 &= 2\int_0^1 (V^2 a^{11})(t, \bar\sigma_0)\,dt + \frac{1}{2}\int_0^1 \Gamma(a^{11}, a^{11})(t, \bar\sigma_0)\,dt, \end{aligned}$$

using the functions

$$\begin{aligned} (Vf)(t,x) &= \sum_{i=1}^N a^{1i}(t,x)\int_t^1 \frac{\partial f}{\partial x_i}(s,x)\,ds, \\ \Gamma(f,g)(t,x) &= \sum_{i,j=1}^N a^{ij}(t,x)\Big(\int_t^1 \frac{\partial f}{\partial x_i}(s,x)\,ds\Big)\Big(\int_t^1 \frac{\partial g}{\partial x_j}(s,x)\,ds\Big). \end{aligned}$$

[2] Recall that the MD rate function for a centered i.i.d. sequence $(X_n)_{n\geq 1}$ is given by $\theta \mapsto \theta^2/(2\mathrm{Var}(X_1))$. This is the "moderate" version of Cramér's theorem (Theorem 3.7.1 in Dembo and Zeitouni [23]; see also the introduction, Chapter 1).

Proof. See Osajima [77], Theorem 1(1), with $T = 1$. □

The following result presents a higher-order expansion in the MOTM regime. It yields an asymptotically equivalent expression for call prices (and not just logarithmic asymptotics).

Theorem 2.5. *Under the assumptions of Theorem 2.3, the logarithm of the call price has the refined MOTM expansion*

$$\log c(k_t, t) = -\sum_{m=2}^{\lfloor 1/\beta \rfloor} \frac{\Lambda^{(m)}(0)}{m!} \frac{k_t^m}{t} + \left(2\beta - \frac{3}{2}\right) \log \frac{1}{t} - 2 \log \ell(t) + \log\left(\gamma(0) v_0^2\right) + o(1), \quad t \downarrow 0. \tag{2.7}$$

This can be expressed equivalently as

$$c(k_t, t) \sim \gamma(0) v_0^2 \frac{t^{3/2-2\beta}}{\ell(t)^2} \exp\left(-\sum_{m=2}^{\lfloor 1/\beta \rfloor} \frac{\Lambda^{(m)}(0)}{m!} \frac{k_t^m}{t}\right), \quad t \downarrow 0.$$

If $1/\beta$ is not an integer, then k_t^m/t tends to infinity for $m = \lfloor 1/\beta \rfloor$, of order $t^{\beta\lfloor 1/\beta \rfloor - 1}$ (up to a slowly varying factor). If on the other hand $1/\beta$ is an integer, then the last summand of the sum $\sum_{m=2}^{\lfloor 1/\beta \rfloor}$ in (2.7) is of order $\ell(t)$, which means that the following term $\log(1/t)$ may be asymptotically larger. The upper summation limit $\lfloor 1/\beta \rfloor$ thus ensures that no irrelevant (i.e. $o(1)$) terms are contained in the sum. Note that $\lfloor 1/\beta \rfloor = 2$ for $\beta \in (\frac{1}{3}, \frac{1}{2})$, and $\lfloor 1/\beta \rfloor \geq 3$ for $\beta \in (0, \frac{1}{3})$, and so (2.7) is consistent with (2.5), resp. (2.6). The passage from the derivatives of the energy function to ATM derivatives of the implied volatility in the short time limit is best conducted via the BBF formula that was proved in Berestycki, Busca, and Florent [9]. (That said, theses relations are also a direct consequence of our expansions, as is pointed out in Section 2.3.) In this regard, we have

Theorem 2.6. *Suppose that Λ is a function with the properties required in Assumption 2.1, with $\Lambda''(0) = \sigma_0^{-2} = v_0^{-1}$, and that the Berestycki-Busca-Florent formula $\sigma_{\text{imp}}^2(0, k) = k^2/2\Lambda(k)$ holds. Then the small-time ATM implied variance skew and curvature, respectively, relate to Λ via*

$$\mathcal{S} := \frac{\partial}{\partial k}\Big|_{k=0} \sigma_{\text{imp}}^2(k, 0) = -\frac{1}{3} \frac{\Lambda'''(0)}{\Lambda''(0)^2} \tag{2.8}$$

and

$$\mathcal{C} := \frac{\partial^2}{\partial k^2}\Big|_{k=0} \sigma_{\text{imp}}^2(k, 0) = \frac{\frac{2}{3}\Lambda'''(0)^2 - \frac{1}{2}\Lambda''''(0)\Lambda''(0)}{3\Lambda''(0)^3}. \tag{2.9}$$

Proof. By the BBF formula and our smoothness assumptions on Λ, for $k \to 0$,

$$\begin{aligned}\sigma_{\text{imp}}^2(k,0) &= \frac{k^2}{2\Lambda(k)} \\ &= k^2\left(\Lambda''(0)k^2 + \frac{1}{3}\Lambda'''(0)k^3 + \frac{1}{12}\Lambda''''(0)k^4 + \mathcal{O}(k^5)\right)^{-1} \\ &= \frac{1}{\Lambda''(0)} - \frac{1}{3}\frac{\Lambda'''(0)}{\Lambda''(0)^2}k + \left(\frac{\frac{1}{9}\Lambda'''(0)^2 - \frac{1}{12}\Lambda''''(0)\Lambda''(0)}{\Lambda''(0)^3}\right)k^2 + \mathcal{O}(k^3).\end{aligned}$$

This implies (2.8) and (2.9). □

Proposition 2.4 combined with Theorem 2.6 allows to compute skew and curvature (and higher derivatives of the implied volatility smile, if desired) directly from the coefficients of a general stochastic volatility model. Related formulae for "general" (even non-Markovian) models also appear in the work of Durrleman (Theorem 3.1.1 in Durrleman [27]; see also Durrleman [28]). While not written in the setting of general Markovian diffusion models, and hence not in terms of the energy function Λ, they inevitably give the same results if applied to given parametric stochastic volatility models (see Section 3.1 in Durrleman [27]). However, Durrleman's work comes with some (seemingly) uncheckable assumptions, the drawbacks of which are discussed in Section 2.6 of Durrleman [27].

2.2 Proofs of the Main Results

Proof of Theorem 2.3. As the density of S_t satisfies $q = \partial_{KK}C$, we have, by Dupire's formula (2.1),

$$C(K,t) = \int_0^t \partial_s C(K,s)\,ds = \int_0^t \frac{1}{2}q(K,s)K^2\sigma_{\text{loc}}^2(K,s)\,ds.$$

Then, for $K_t = e^{k_t}$ with $k_t \downarrow 0$ as stated, we apply Assumption 2.2 as follows

$$\begin{aligned}C(K_t,t) &= \int_0^t \frac{1}{2}q(K_t,s)K_t^2\sigma_{\text{loc}}^2(K_t,s)\,ds \\ &\sim \frac{\sigma_0^2}{2}\int_0^t q(K_t,s)\,ds, \quad t\downarrow 0. \end{aligned} \tag{2.10}$$

And then, using local uniformity of our density expansion (2.2), as $t \downarrow 0$,

$$C(K_t,t) \sim \frac{\sigma_0^2\gamma(0)}{2}\int_0^t e^{-\frac{\Lambda(k_t)}{s}}s^{-1/2}\,dt \tag{2.11}$$

$$= \frac{\sigma_0^2\gamma(0)}{2}t\int_0^1 e^{-\frac{\Lambda(k_t)}{xt}}(xt)^{-1/2}\,dx$$

$$= \frac{\sigma_0^2\gamma(0)}{2}t^{1/2}\int_0^1 e^{-\frac{\Lambda(k_t)}{xt}}x^{-1/2}\,dx. \tag{2.12}$$

Because Λ is smooth at zero, and using the fact that $\Lambda(0) = \Lambda'(0) = 0$, we have

$$\frac{\Lambda(k_t)}{t} \sim \frac{1}{2}\Lambda''(0)\frac{k_t^2}{t} \to \infty \quad \text{as } t \downarrow 0.$$

For small t, the integrand in (2.12) is thus concentrated near $x = 1$, and by the Laplace method (Theorem 3.7.1 in Olver [74])

$$\int_0^1 e^{-\frac{\Lambda(k_t)}{xt}} x^{-1/2}\, dx \sim \frac{t}{\Lambda(k_t)} \exp\Big(-\frac{\Lambda(k_t)}{t}\Big), \quad t \downarrow 0. \tag{2.13}$$

Therefore,

$$\begin{aligned} C(K_t, t) &\sim \frac{\sigma_0^2\gamma(0)}{2}\frac{t^{3/2}}{\Lambda(k_t)} \exp\Big(-\frac{\Lambda(k_t)}{t}\Big) \\ &\sim v_0^2\gamma(0)\frac{t^{3/2}}{k_t^2} \exp\Big(-\frac{\Lambda(k_t)}{t}\Big), \quad t \downarrow 0, \end{aligned} \tag{2.14}$$

which implies (recall the notation c resp. C from (1.1))

$$\begin{aligned} -\log c(k_t, t) &= \frac{\Lambda(k_t)}{t} - \log\frac{t^{3/2}}{k_t^2} + \mathcal{O}(1) \qquad (2.15) \\ &= \frac{1}{t}\Big(\frac{1}{2}\Lambda''(0)k_t^2 + \frac{1}{6}\Lambda'''(0)k_t^3 + \mathcal{O}(k_t^4)\Big) + \mathcal{O}\Big(\log\frac{k_t^2}{t^{3/2}}\Big), \quad t \downarrow 0. \end{aligned}$$

To prove (i) and (ii), we thus need to argue that k_t^2/t dominates $\log(k_t^2 t^{-3/2})$ if $\beta \in (0, \frac{1}{2})$, and that k_t^3/t dominates $\log(k_t^2 t^{-3/2})$ if $\beta \in (0, \frac{1}{3})$. For $m \in \{2, 3\}$, we calculate

$$\begin{aligned} \frac{k_t^m/t}{|\log(k_t^2 t^{-3/2})|} &= \frac{t^{m\beta-1}\ell(t)^m}{|\log(t^{2\beta-3/2}\ell(t)^2)|} \\ &= \frac{t^{m\beta-1}\ell(t)^m}{|(2\beta-\frac{3}{2})\log t + 2\log\ell(t)|}. \end{aligned}$$

From Proposition 1.3.6 (i) in Bingham, Goldie, and Teugels [11], we know that $\log \ell(t) = o(\log t)$, and so

$$\frac{k_t^m/t}{|\log(k_t^2 t^{-3/2})|} \sim \frac{t^{m\beta-1}\ell(t)^m}{|(2\beta-\frac{3}{2})\log t|}, \quad t \downarrow 0.$$

This tends to infinity for $m = 2$ and $\beta \in (0, \frac{1}{2})$, and for $m = 3$ and $\beta \in (0, \frac{1}{3})$, as desired. □

Inspecting the preceding proof, it is easy to see that we can expand $\log c(k_t, t)$ further.

Proof of Theorem 2.5. Taking logs in (2.14) yields

$$\log c(k_t, t) = -\frac{\Lambda(k_t)}{t} + \log\frac{t^{3/2}}{k_t^2} + \log\big(\gamma(0)v_0^2\big) + o(1), \quad t \downarrow 0.$$

Then (2.7) follows by Taylor expanding Λ. Note that $k_t^m/t = o(1)$ for $m \geq \lfloor 1/\beta \rfloor + 1$. □

2.3 Implied Volatility

Corollary 2.1. *Under the assumptions of Theorem 2.3, let $k_t = t^\beta \ell(t)$ with $\beta \in (0, \frac{1}{3})$ and $\ell > 0$ slowly varying. Then the implied volatility has the MOTM expansion*

$$\sigma_{\mathrm{imp}}(k_t, t) = \sigma_0 - \tfrac{1}{6}\sigma_0^3 \Lambda'''(0) k_t \big(1 + o(1)\big), \quad t \downarrow 0. \tag{2.16}$$

Proof. We use our main result (Theorem 2.3) in conjunction with a transfer result of Gao and Lee [49]. As the call price tends to zero, we are in case "$(-)$" of Gao and Lee [49] (defined on p. 354 of that paper). The notation L, V therein means $L = -\log c(k_t, t)$ resp. $V = t^{1/2}\sigma_{\mathrm{imp}}(k_t, t)$, the dimensionless implied volatility. Then Corollary 7.2 of Gao and Lee [49] implies that

$$V = \frac{k_t}{\sqrt{2L}}\big(1 + \mathcal{O}(t^{1-2\beta-\varepsilon})\big) + \mathcal{O}(t^{5/2-4\beta-\varepsilon}), \quad t \downarrow 0. \tag{2.17}$$

Here, $\varepsilon > 0$ denotes an arbitrarily small constant that serves to eat up slowly varying functions in $\mathcal{O}$-estimates (See Proposition 1.3.6 (v) in Bingham, Goldie, and Teugels [11]). By part (ii) of Theorem 2.3, we have

$$2L = \frac{1}{\sigma_0^2}\frac{k_t^2}{t} + \frac{\Lambda'''(0)}{3}\frac{k_t^3}{t}\big(1 + o(1)\big), \quad t \downarrow 0.$$

Inserting this into (2.17) gives

$$\sigma_{\mathrm{imp}}(k_t, t) = t^{-1/2} k_t \left(\sigma_0 \frac{t^{1/2}}{k_t} - \frac{\sigma_0^3 \Lambda'''(0)}{6} t^{1/2}\big(1 + o(1)\big)\right) + \mathcal{O}(t^{2-4\beta-\varepsilon}), \quad t \downarrow 0,$$

which yields (2.16). □

We have no doubt that Corollary 2.1 is true for the whole MOTM regime, i.e. for all $\beta \in (0, \frac{1}{2})$, under very mild assumptions (Assumption A.1 in the Appendix A). For any *fixed* $\beta \in (0, \frac{1}{2})$, one can compute the implied volatility expansion using the results of Gao and Lee [49]. However, for β close to $\frac{1}{2}$, more and more terms are needed for the intermediate computations, and there does not seem to be a simple pattern that would allow for a general proof. The details are discussed in Appendix A, where we push the range of β for which (2.16) is proven rigorously to $0 < \beta < \frac{3}{7} \approx 0.429$. Note that the expansion in Theorem 2.5 becomes *finer* (i.e. contains more explicit terms) if β is close to zero. Suppose, on the other hand, that β is very close to $\frac{1}{2}$: Then the summands $m > \lfloor 1/\beta \rfloor = 2$ in (2.7), which are related to ATM derivatives of implied variance by Theorem 2.6 (see also paragraph (iii) in the introduction), disappear into the $o(1)$-term of (2.7).
Corollary 2.1 has some interesting consequences. Under the sheer assumption that implied volatility has a first order Taylor expansion for small maturity and small log-strike of the form

$$\sigma_{\mathrm{imp}}(k, t) = \sigma_0 + \partial_k \sigma_{\mathrm{imp}}(0, 0)\, k + o(k) + \mathcal{O}(t), \quad t \downarrow 0, k = o(1), \tag{2.18}$$

then of course in the MOTM regime, we have $t \ll k_t$, and so the k-term dominates the $\mathcal{O}(t)$-term, which in turn identifies the implied variance skew as

$$\mathcal{S} = \lim_{t\downarrow 0} \frac{2\sigma_0}{k_t}\big(\sigma_{\mathrm{imp}}(k_t,t) - \sigma_0\big). \tag{2.19}$$

On the other hand, Corollary 2.1 now implies that the right-hand side of (2.19) equals $-\frac{1}{3}\sigma_0^4\Lambda'''(0)$. We have thus arrived at an alternative proof of the skew representation (2.8) in terms of the energy function, without using the BBF formula. The curvature and higher order derivatives of the ATM smile can be dealt with similarly, if desired.

2.4 Examples

2.4.1 Generic Local Volatility Models

Clearly, Assumption 2.2 is satisfied for any local volatility model, assuming continuity of the local volatility function. We now discuss Assumption 2.1, and show how to compute our MOTM expansions. First consider the time-homogeneous local volatility model

$$dS_t = \sigma(S_t)S_t\, dW_t, \quad S_0 = 1, \tag{2.20}$$

where the deterministic function σ is C^2 on $(0,\infty)$. An expansion of the pdf $q(\cdot,t)$ of S_t has been worked out in Gatheral et al. [51]. They assume growth conditions on σ and its derivatives, which can be alleviated by the principle of not feeling the boundary (Appendix A of Gatheral et al. [51]). Proposition 2.1 therein states that

$$q(e^k,t) \sim \frac{e^{-k}u_0(0,k)}{\sqrt{2\pi t}} \exp\left(-\frac{\Lambda(k)}{t}\right), \quad t\downarrow 0, \tag{2.21}$$

uniformly in k, where the energy function is given by (cf. Varadhan [89])

$$\Lambda(k) = \frac{1}{2}\left(\int_0^k \frac{dx}{\sigma(e^x)}\right)^2,$$

and

$$u_0(0,k) = \sigma(1)^{1/2}\sigma(e^k)^{-3/2}e^{-k/2}. \tag{2.22}$$

(Recall that we normalize spot to $S_0 = 1$ throughout.) This shows that Assumption 2.1 is satisfied, with

$$\gamma(0) = \frac{1}{\sqrt{2\pi}\sigma(1)}. \tag{2.23}$$

To evaluate the expansions from Theorem 2.3, we compute the derivatives of Λ

$$\Lambda'(k) = \frac{1}{\sigma(e^k)}\int_0^k \frac{dx}{\sigma(e^x)}, \qquad \Lambda''(k) = \frac{1}{\sigma(e^k)^2} - \frac{e^k\sigma'(e^k)}{\sigma(e^k)^2}\int_0^k \frac{dx}{\sigma(e^x)},$$
$$\Lambda'''(k) = -\frac{3e^k\sigma'(e^k)}{\sigma(e^k)^3} + \left(\frac{2e^{2k}\sigma'(e^k)^2}{\sigma(e^k)^3} - \frac{\sigma(e^k)''}{\sigma(e^k)^2}\right)\int_0^k \frac{dx}{\sigma(e^x)},$$

which yield

$$\Lambda''(0) = \frac{1}{\sigma(1)^2} = \frac{1}{\sigma(S_0)^2},$$
$$\Lambda'''(0) = -\frac{3\sigma'(1)}{\sigma(1)^3} = -\frac{3\sigma'(S_0)}{\sigma(S_0)^3}. \tag{2.24}$$

Alternatively, these expressions can be obtained from Proposition 2.4. As the assumptions of Theorem 2.3 are satisfied, we obtain the following MOTM call price estimates, where $k_t = \theta t^\beta$ and $\theta > 0$

$$c(k_t, t) = \exp\left(- \frac{\theta^2}{2\sigma(1)^2 t^{1-2\beta}} \big(1 + o(1)\big) \right), \quad \beta \in (0, \tfrac{1}{2}), \ t \downarrow 0,$$
$$c(k_t, t) = \exp\left(- \frac{\theta^2}{2\sigma(1)^2 t^{1-2\beta}} - \frac{\sigma'(1)}{2\sigma(1)^3} \frac{\theta^3}{t^{1-3\beta}} \big(1 + o(1)\big) \right), \quad \beta \in (0, \tfrac{1}{3}), \ t \downarrow 0.$$

Recall from Theorem 2.6 that we denote by $\mathcal{S}$ the (limiting small-time ATM) implied *variance* skew, and so the implied *volatility* skew is given by $\mathcal{S}/(2\sigma_0)$, which equals $\mathcal{S}/(2\sigma(1)) = \mathcal{S}/(2\sigma(S_0))$ in the model (2.20). From (2.8) and (2.24), we find that the *local* skew $\sigma'(1) = \sigma'(S_0)$ equals twice the implied volatility skew,

$$\sigma'(S_0) = 2 \cdot \frac{\mathcal{S}}{2\sigma(S_0)},$$

as observed in Remark 5.2 of Henry-Labordère [59]. Generic time-inhomogeneous local volatility models

$$dS_t = \sigma(S_t, t) S_t \, dW_t$$

can be treated very similarly, using the heat kernel expansion in Section 3 of Gatheral et al. [51], itself taken from Yosida [91].

2.4.2 Generic Stochastic Volatility Models

We now discuss the results of Section 2.1 in generic stochastic volatility models. Rigorous conditions under which stochastic volatility models satisfy Assumption 2.1 can be found in Deuschel, Friz, Jacquier and Violante [24] and Osajima [77]. The function Λ is given by the Riemannian metric associated to the model: $2\Lambda(k)$ is the squared geodesic distance from $(S_0 = 1, \sigma_0)$ to $\{(K, \sigma) \colon \sigma > 0\}$ with $K = e^k$. Theorem 2.2 in Berestycki, Busca and Florent [10] gives conditions under which Assumption 2.2, concerning convergence of local volatility, is true.
Now we describe how the expressions appearing in the expansions from Theorem 2.3 can be computed explicitly in a generic two-factor stochastic volatility model

$$dS_t = S_t \sqrt{V_t} \, dW_t, \quad S_0 = 1,$$
$$dV_t = (\dots) \, dt + \xi \sqrt{V_t} \nu(V_t) \, dZ_t, \quad V_0 = v_0 > 0, \tag{2.25}$$

where $\nu\colon \mathbb{R} \to \mathbb{R}$ and $d\langle W, Z\rangle_t = \rho\, dt$. The Heston model ($\nu(v) \equiv 1$) and the 3/2-model ($\nu(v) = v$; see Lewis [68]) are special cases. The infinitesimal generator L of the stochastic process (S, V), neglecting first order terms, can be written as

$$Lf \approx \frac{1}{2} \operatorname{Tr}\left(\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} D^2 f \right), \qquad f \in C^2(\mathbb{R}^2),$$

where $D^2 f$ denotes the Hessian matrix of f, and the coefficient matrix $g(v) = (g_{ij}(v))$ is given by

$$g(v) = \begin{pmatrix} v & \rho\xi v\nu(v) \\ \rho\xi v\nu(v) & \xi^2 v\nu(v)^2 \end{pmatrix}.$$

We define the constants $b_1 = g_{11}(v_0) = v_0$ and $b_2 = \frac{3}{4}\sum_{i=1}^2 g_{1i}(v_0)\partial_i g_{11}(v_0) = \frac{3}{4}\rho\xi v_0\nu(v_0)$. If we assume that the coefficients in (2.25) are nice enough to justify application of the marginal density expansion obtained in Deuschel, Friz, Jacquier and Violante [24] or part (2) of Theorem 1 in Osajima [77], then we get the desired small-time density expansion (2.2). Moreover, thanks to Proposition 2.4,

$$\Lambda(k) = \frac{1}{2b_1}k^2 - \frac{b_2}{3b_1^3}k^3 + \mathcal{O}(k^4)$$

as $k \to 0$. Therefore, the quantities $\Lambda''(0) = v_0^{-1} = \sigma_0^{-2}$ and $\Lambda'''(0) = -\frac{3}{2}\rho\xi\nu(v_0)/v_0^2$ can easily be computed, as well as the small-time ATM implied variance skew

$$\mathcal{S} = -\frac{v_0^2}{3}\Lambda'''(0) = \frac{\rho\xi}{2}\nu(v_0).$$

Thus, all quantities appearing in our expansions (Theorem 2.3, Corollary 2.1) have very simple expressions in terms of the model parameters.

2.4.3 The Heston Model

This section contains an application of the results of Sections 2.1 and 2.3 to the familiar case of the Heston model, where many explicit "affine" computations are possible. At the beginning of Section 2.4.2, we recalled some general results implying our Assumptions 2.1 and 2.2. The Heston model is not covered by these results, but satisfies Assumptions 2.1 and 2.2 nevertheless, and thus Theorems 2.3 and 2.5 are applicable to Heston. We will explain how both assumptions can be verified rigorously by a dedicated analysis; full details would involve rather dull repetition of arguments that are found in the literature in a very similar form, and are therefore omitted. The model dynamics are

$$\begin{aligned} dS_t &= S_t\sqrt{V_t}\, dW_t, \qquad S_0 = 1, \\ dV_t &= \lambda(\bar{v} - V)\, dt + \xi\sqrt{V_t}\, dZ_t, \qquad V_0 = v_0 > 0, \end{aligned}$$

where $\bar{v}$, λ, $\xi > 0$, and $d\langle W, Z\rangle_t = \rho\, dt$ with $\rho \in (-1, 1)$. According to Forde and Jacquier [38], the first order OTM (large deviations) behaviour of the call prices is

$$t \log c_{\mathrm{He}}(k,t) \sim -\Lambda_{\mathrm{He}}(k), \quad k > 0 \text{ fixed}, \ t \downarrow 0, \tag{2.26}$$

where Λ_{He} is the (not explicitly available) Legendre transform of

$$\begin{aligned}\Gamma(p) &= \frac{v_0 p}{\xi(\bar{\rho}\cot(\xi\bar{\rho}p/2) - \rho)} \\ &= \frac{v_0 p}{\xi\big(\bar{\rho}\big(\frac{1}{\xi\bar{\rho}p/2} + \mathcal{O}(p)\big) - \rho\big)} \\ &= \frac{v_0 p}{\frac{1}{p/2} - \xi\rho + \mathcal{O}(p)} \\ &= \frac{v_0 p^2/2}{1 - p\xi\rho/2 + \mathcal{O}(p^2)} \\ &= \frac{v_0 p^2}{2}\big(1 + p\xi\rho/2 + \mathcal{O}(p^2)\big), \quad p \to 0. \end{aligned} \tag{2.27}$$

(We use the standard notation $\bar{\rho}^2 = 1 - \rho^2$.) This expansion implies

$$\Gamma''(0) = v_0 = \sigma_0^2. \tag{2.28}$$

The locally uniform density asymptotics (2.2) hold, as seen from an easy modification of the arguments in Forde, Jacquier and Lee [41]. There, the Fourier representation of the call price was analysed by the saddle point method to obtain a refinement of (2.26). Proceeding analogously for the Fourier representation of the pdf of S_t, we get the density approximation

$$\begin{aligned} q_{\mathrm{He}}(e^k, t) &= \frac{e^{-k}}{2\pi t}\int_{-\infty - ip^*(k)}^{\infty - ip^*(k)} \mathrm{Re}\big(e^{iku/t}\phi_t(-u/t)\big)\, du \\ &= \exp\Big(-\frac{\Lambda_{\mathrm{He}}(k)}{t}\Big)\frac{U(p^*(k))}{\sqrt{2\pi\Gamma''(k)}} t^{-1/2}\big(1 + o(1)\big), \quad t \downarrow 0, \end{aligned}$$

locally uniformly in k, where ϕ_t is the characteristic function of $X_t = \log S_t$, and p^* and U are defined on p. 693 of Forde, Jacquier and Lee [41]. (Note that Forde, Jacquier and Lee [41] use the notation Λ, Λ^* instead of our $\Gamma, \Lambda_{\mathrm{He}}$.) From (2.28) and the fact that $U(p^*(0)) = U(0) = 1$, we see that the factor $\gamma(k)$ from (2.2) converges to

$$\gamma(0) = \frac{1}{\sqrt{2\pi}\sigma_0} \tag{2.29}$$

as $k \to 0$.

To verify Assumption 2.2 (convergence of local volatility), the Dupire formula (2.1) can be subjected to an analysis similar to De Marco, Friz and Gerhold [22] and Friz and Gerhold [43]. More precisely, $\partial_{KK}C(K,t)$ in the numerator of (2.1) is the pdf of S_t, the

analysis of which we have just described. Virtually the same saddle point approach can be applied to the numerator $\partial_t C(K,t)$, yielding convergence of the quotient to σ_0^2.
We now calculate our MOTM asymptotic expansions for the Heston model. The Legendre transform Λ_{He} is given by $\Lambda_{\mathrm{He}}(k) = \sup_x\{kx - \Gamma(x)\}$ with maximizer $x^* = x^*(k)$. From general facts on Legendre transforms,

$$\Lambda''_{\mathrm{He}}(k) = \frac{1}{\Gamma''(x^*(k))}.$$

We have $x^*(0) = 0$, which implies

$$\Lambda''_{\mathrm{He}}(0) = \frac{1}{\Gamma''(0)} = \frac{1}{v_0}.$$

From Theorem 2.3, with $k_t = \theta t^\beta$ and $\theta > 0$, we then obtain the MOTM call price estimate

$$c_{\mathrm{He}}(k_t,t) = \exp\left(-\frac{\theta^2}{2v_0 t^{1-2\beta}}\big(1+o(1)\big)\right), \quad t \downarrow 0. \tag{2.30}$$

As for the second order expansion, from the expansion (2.27) of Γ, we clearly see that

$$\Gamma'''(0) = \frac{3}{2} v_0 \xi \rho.$$

On the other hand, a general Legendre computation gives

$$\Lambda'''_{\mathrm{He}}(k) = -\left(\frac{1}{\Gamma''(x^*(k))}\right)^2 \Gamma'''(x^*(k))\,(x^*)'(k) = -(\Lambda''_{\mathrm{He}}(k))^3\,\Gamma'''(x^*(k)).$$

Therefore,

$$\Lambda'''_{\mathrm{He}}(0) = -\frac{3}{2}\frac{\xi\rho}{v_0^2},$$

in accordance with the expression for generic two-factor models, found in Section 2.4.2. For $\beta \in (0, \frac{1}{3})$, Theorem 2.3 (ii) thus implies the second order expansion

$$c_{\mathrm{He}}(k_t,t) = \exp\left(-\frac{\theta^2}{2v_0 t^{1-2\beta}} + \frac{\xi\rho}{4v_0^2}\frac{\theta^3}{t^{1-3\beta}}\big(1+o(1)\big)\right), \quad t \downarrow 0. \tag{2.31}$$

By Theorem 2.5 and (2.29), we obtain the following refined call price expansions, as $t \downarrow 0$:

$$\log c_{\mathrm{He}}(k_t,t) = -\frac{1}{2\sigma_0^2}\frac{k_t^2}{t} + \left(\frac{3}{2} - 2\beta\right)\log t + \log\frac{\sigma_0^3}{\sqrt{2\pi}} + o(1), \quad \beta \in (\tfrac{1}{3}, \tfrac{1}{2}), \tag{2.32}$$

$$\log c_{\mathrm{He}}(k_t,t) = -\frac{1}{2\sigma_0^2}\frac{k_t^2}{t} + \frac{\xi\rho}{4v_0^2}\frac{k_t^3}{t} + \left(\frac{3}{2} - 2\beta\right)\log t + \log\frac{\sigma_0^3}{\sqrt{2\pi}} + o(1), \quad \beta \in (\tfrac{1}{4}, \tfrac{1}{3}). \tag{2.33}$$

From the relation (2.8) between implied variance skew and $\Lambda'''(0)$, we get the explicit expression $\mathcal{S}_{\mathrm{He}} = \xi\rho/2$ for the skew. This agrees with Gatheral [50], p. 35. The implied volatility expansion (2.19) becomes

$$\sigma_{\mathrm{imp}}(k_t, t) = \sigma_0 + \frac{\xi\rho}{4\sigma_0} k_t\big(1 + o(1)\big), \quad t \downarrow 0. \tag{2.34}$$

Figure 2.1 shows a good fit of this approximation, even for maturities that are not very small.

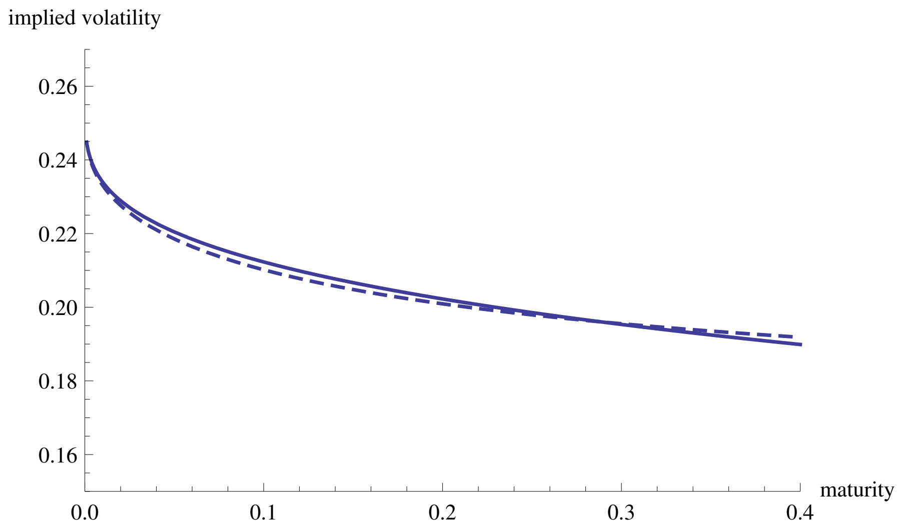

Figure 2.1: Illustration of our implied volatility expansion for the Heston model, with $\ell \equiv \theta = 0.4$ and $\beta = 0.3$. Thus, log-strike equals $k_t = 0.4\,t^{0.3}$. The model parameters are $\bar{v} = 0.0707$, $\kappa = 0.6067$, $\xi = 0.2928$, $\rho = -0.7571$, $v_0 = 0.0654$ (i.e. $\sigma_0 = 0.2557$) and $S_0 = 1$. The horizontal axis is time. The dashed curve is the exact MOTM implied volatility $\sigma_{\mathrm{imp}}(k_t, t)$. The solid curve is the approximation $\sigma_0 + \frac{\xi\rho}{4\sigma_0} k_t$ on the right hand side of (2.34).

2.5 Other Approaches at MOTM Asymptotics

In a recent paper, Gao and Wang [48] study small noise sample-path MDPs (moderate deviation principles) for SDE solutions, and specialize to the small-time regime (Corollary 4.1.2 in Gao and Wang [48]). Their asymptotic regime is in fact slightly more general than (2.4), allowing for (in our notation) any k_t satisfying $\sqrt{t} \ll k_t \ll 1$ as $t \downarrow 0$. (In the financial context, this offers no useful additional flexibility; it allows, e.g. to switch between two regimes $k_t = t^{\beta_1}$ and $k_t = t^{\beta_2}$ infinitely often as $t \downarrow 0$.) However, Gao and Wang [48] impose the assumptions of linearly bounded and locally Lipschitz coefficients. These are the typical assumptions for small-noise LDPs in the literature, but they are rarely satisfied in stochastic volatility models. In particular, their results are not directly applicable to the Heston model. The paper by Cai and Wang [16] is also worth mentioning here: It presents moderate deviations for the CIR process (the Heston variance process) and a generalization, where the exponent in the dynamics is not necessarily 1/2. The paper uses estimates tied to the (generalized) CIR stochastic differential equation.

In this section we discuss a different approach at small-time moderate deviations. While yielding only first order results, its conditions are usually easy to check for models with explicit characteristic function. Assumptions 2.1 and 2.2 are not in force here. Recall that, in the classical setting of sequences of i.i.d. random variables, a moderate deviation analogue of Cramér's theorem can be deduced by applying the Gärtner-Ellis theorem to an appropriately rescaled sequence (see Dembo and Zeitouni [23], Section 3.7). The MD short time behaviour of diffusions can be subjected to a similar analysis. Consider the log-price $X_t = \log S_t$ with $X_0 = 0$ and mgf (moment generating function)

$$M(p,t) := \mathbb{E}[e^{pX_t}]. \tag{2.35}$$

Assumption 2.1. For all $\beta \in (0, \frac{1}{2})$, the rescaled mgf satisfies

$$\lim_{t\downarrow 0} t^{1-2\beta} \log M(t^{\beta-1}p, t) = \tfrac{1}{2}\sigma_0^2 p^2, \quad p \in \mathbb{R}. \tag{2.36}$$

We expect that this assumption holds for diffusion models in considerable generality. It is easy to check that (2.36) holds for the Heston model, either by its explicit characteristic function, or, more elegantly, from the associated Riccati equations; see Appendix B. Thus, the results of the present section provide an alternative proof of the first order MOTM behaviour (2.30) of Heston call prices.

Heuristically, Assumption 2.1 can be derived from the density asymptotics in Assumption 2.1, which in turn hold in quite general diffusion settings (see Deuschel, Friz, Jacquier and Violante [24] and [25]).

$$\begin{aligned}
M(t^{\beta-1}p,t) &= \int e^{t^{\beta-1}px} q(x,t)\,dx \\
&\approx \int \exp\left(t^{\beta-1}px - \frac{\Lambda(x)}{t}\right) dx \qquad (2.37)\\
&\approx \int \exp\left(t^{\beta-1}px - \frac{\Lambda''(0)x^2}{2t}\right) dx \qquad (2.38)\\
&= \exp\left(t^{\beta-1}px - \frac{x^2}{2\sigma_0^2 t}\Big|_{x=\sigma_0^2 p t^{\beta}} (1+o(1))\right) \qquad (2.39)\\
&= \exp\left(\tfrac{1}{2}\sigma_0^2 p^2 t^{2\beta-1}(1+o(1))\right), \quad t\downarrow 0. \qquad (2.40)
\end{aligned}$$

In (2.37), we ignored that the density expansion (2.2) might not be valid globally in space; this might be made rigorous by estimating $q(x,t)$ by a Freidlin-Wentzell LD argument for x sufficiently large. As for (2.38), we can expect concentration near $x \approx 0$, because $\Lambda(x)$ increases with $|x|$. Finally, (2.39), and thus (2.40), follows from a (rigorous) application of the Laplace method. If (2.40) is correct, then (2.36) clearly follows.

The critical moment of S_t is defined by

$$p_+(t) := \sup\{p \geq 0\colon M(p,t) < \infty\}.$$

It is obvious that

$$\lim_{t\downarrow 0} t^{1-\beta} p_+(t) = \infty \tag{2.41}$$

is necessary for (2.36), i.e. $p_+(t)$ must grow faster than $t^{\beta-1}$ as $t \downarrow 0$. In the Heston model, e.g., the critical moment is of order $p_+(t) \sim (const)/t \gg t^{\beta-1}$ for small t, as follows from inverting (6.2) in Keller-Ressel [63]. On the other hand, we do not expect our results to be of much use in the presence of jumps. Indeed, suppose that (2.35) is the mgf of an exponential Lévy model. Then $p_+(t) \equiv p_+$ does not depend on t, and is finite for most models used in practice. Therefore, (2.41) cannot hold, and so Assumption 2.1 is not satisfied. The Merton jump diffusion model is one of the few Lévy models of interest that have $p_+ = \infty$, but it is easy to check that it does not satisfy (2.36), either. After this discussion of Assumption 2.1, we now give an asymptotic estimate for the distribution function of X_t (put differently, MOTM *digital call* prices) in Theorem 2.2. Then we translate this result to MOTM *call* prices in Theorem 2.3. If desired, higher order terms in (2.36) will give refined asymptotics in Theorem 2.2, using the recent refinement of the Gärtner-Ellis theorem in Gulisashvili and Teichmann [58]. (Further work will be required to translate the resulting expansions into call price asymptotics.) For other asymptotic results on option prices using the Gärtner-Ellis theorem, see e.g. Forde and Jacquier [38] and Forde and Jacquier [39].

Theorem 2.2. *Under Assumption 2.1 (and without any further assumptions on our model), for $k_t = \theta t^\beta$ with $\beta \in (0, \frac{1}{2})$ and $\theta > 0$, we have a first order MD estimate for the cdf of X_t:*

$$\mathbb{P}[X_t \geq k_t] = \exp\left(-\frac{1}{2\sigma_0^2}\frac{k_t^2}{t}\big(1+o(1)\big)\right), \quad t \downarrow 0. \tag{2.42}$$

Proof. Define

$$Z_t := t^{-\beta} X_t, \quad \text{with mgf } M_Z(s,t) = \mathbb{E}[e^{sZ_t}],$$

and

$$a_t := t^{1-2\beta} = o(1), \quad t \downarrow 0.$$

Then (2.36) is equivalent to

$$\Gamma_Z(p) := \lim_{t\downarrow 0} a_t \log M_Z(p/a_t, t) = \tfrac{1}{2}\sigma_0^2 p^2, \quad p \in \mathbb{R}.$$

As Γ_Z is finite and differentiable on $\mathbb{R}$, the Gärtner-Ellis theorem (Theorem 2.3.6 in Dembo and Zeitouni [23]) implies that $(Z_t)_{t\geq 0}$ satisfies an LDP (large deviation principle) as $t \downarrow 0$, with rate a_t and good rate function Λ_Z, the Legendre transform of Γ_Z. Trivially, Λ_Z is quadratic, too:

$$\begin{aligned}\Lambda_Z(x) &= \sup_{p\in\mathbb{R}}(px - \Gamma_Z(p)) \\ &= \sup_{p\in\mathbb{R}}(px - \tfrac{1}{2}\sigma_0^2 p^2) = \frac{x^2}{2\sigma_0^2}, \quad x \in \mathbb{R}.\end{aligned}$$

Now fix $\theta > 0$. Applying the lower estimate of the LDP to (θ, ∞) yields

$$\begin{aligned}\liminf_{t\downarrow 0} a_t \log \mathbb{P}[Z_t \geq \theta] &\geq \liminf_{t\downarrow 0} a_t \log \mathbb{P}[Z_t > \theta] \\ &\geq -\Lambda_Z(\theta) = -\frac{\theta^2}{2\sigma_0^2},\end{aligned}$$

and applying the upper estimate to $[\theta, \infty)$ yields

$$\limsup_{t\downarrow 0} a_t \log \mathbb{P}[Z_t \geq \theta] \leq -\frac{\theta^2}{2\sigma_0^2},$$

and so

$$\lim_{t\downarrow 0} a_t \log \mathbb{P}[Z_t \geq \theta] = -\frac{\theta^2}{2\sigma_0^2}.$$

This is the same as (2.42). □

As in the LD/OTM regime, first order cdf asymptotics translate readily into call price asymptotics. The proof of the following result is similar to Pham [80], p. 30f (concerning the LD regime) and Caravenna and Corbetta [17], Theorem 1.5. In the MD/MOTM regime, one can replace the condition (1.19) of Caravenna and Corbetta [17] by a mild condition on the moments of the model.

Theorem 2.3. *Let $S = e^X$ be a continuous positive martingale. Assume that, for all $p \geq 1$, its p-th moment explodes at a positive time (infinity included).*[3] *By this we mean that there is a positive $t^*(p)$ such that the mgf $\mathbb{E}[\exp(pX_t)]$ is finite for all $t \in [0, t^*(p)]$. Let $v_0 = \sigma_0^2 > 0$. Then the following are equivalent*

(i) For $k_t = \ell(t)t^\beta$, with $\beta \in (0, \frac{1}{2})$ and $\ell > 0$ slowly varying at zero, it holds that

$$\mathbb{P}[X_t \geq k_t] = \exp\left(-\frac{1}{2v_0}\frac{k_t^2}{t}(1 + o(1))\right), \quad t \downarrow 0.$$

(ii) Under the assumptions of (i), we have

$$c(k_t, t) = \exp\left(-\frac{1}{2v_0}\frac{k_t^2}{t}(1 + o(1))\right), \quad t \downarrow 0. \tag{2.43}$$

Proof. First assume (i). Let $\varepsilon > 0$ and define $\tilde{k}_t = (1 + \varepsilon)k_t$. Then

$$\begin{aligned}c(k_t, t) &\geq \mathbb{E}[(e^{X_t} - e^{k_t})^+ \mathbb{1}_{\{X_t \geq \tilde{k}_t\}}] \\ &\geq (e^{\tilde{k}_t} - e^{k_t})^+ \mathbb{P}[X_t \geq \tilde{k}_t]. \end{aligned}\tag{2.44}$$

[3] See Keller-Ressel [63] for a discussion of moment explosion in affine stochastic volatility models.

The first factor is

$$(e^{\tilde{k}_t} - e^{k_t})^+ = (\tilde{k}_t - k_t + \mathcal{O}(k_t^2))^+ = \varepsilon k_t + \mathcal{O}(k_t^2), \quad t \downarrow 0.$$

For the second factor in (2.44), we apply (i) with $\tilde{k}_t$.

$$\lim_{t\downarrow 0} \frac{t}{\tilde{k}_t^2} \log \mathbb{P}[X_t \geq \tilde{k}_t] = -\frac{1}{2v_0}.$$

Therefore,

$$\begin{aligned}\liminf_{t\downarrow 0} \frac{t}{k_t^2} \log c(k_t, t) &\geq \lim_{t\downarrow 0} \frac{t}{k_t^2} \Big(-\frac{1}{2v_0} \frac{\tilde{k}_t^2}{t} \big(1 + o(1)\big)\Big) \\ &= -\frac{(1+\varepsilon)^2}{2v_0}.\end{aligned}$$

Now let $\varepsilon \downarrow 0$ to get the desired lower bound for $c(k_t, t)$.
As for the upper bound, we let $p > 1$ and note that, by definition of $p \mapsto t^*(p)$, we have $\mathbb{E}[S_t^{p+1}] < \infty$ for all $t \in [0, t^*(p+1)]$. Define $\bar{S}_t = \sup_{0\leq u\leq t} S_u$ for $t \geq 0$. By Doob's inequality (Theorem 3.8 in Karatzas and Shreve [62]), we have

$$\mathbb{P}[\bar{S}_{t^*(p+1)} \geq s] \leq \frac{\mathbb{E}[S_{t^*(p+1)}^{p+1}]}{s^{p+1}}, \quad s > 0.$$

Hence $\bar{S}_{t^*(p+1)}$ has a finite pth moment,

$$\mathbb{E}[(\bar{S}_{t^*(p+1)})^p] = p \int_0^\infty s^{p-1} \mathbb{P}[\bar{S}_{t^*(p+1)} \geq s]\, ds < \infty.$$

By the dominated convergence theorem and the continuity of S, we thus conclude

$$\lim_{t\downarrow 0} \mathbb{E}[S_t^p] = S_0^p. \tag{2.45}$$

Now let $1/p + 1/q = 1$ and apply Hölder's inequality.

$$\begin{aligned} c(k_t, t) &= \mathbb{E}[(e^{X_t} - e^{k_t})^+ \, \mathbb{1}_{\{X_t \geq k_t\}}] \\ &\leq \mathbb{E}[((e^{X_t} - e^{k_t})^+)^p]^{1/p} \, \mathbb{P}[X_t \geq k_t]^{1/q} \\ &\leq \mathbb{E}[S_t^p]^{1/p} \, \mathbb{P}[X_t \geq k_t]^{1/q}. \end{aligned}$$

By (2.45) and (i), we obtain

$$\limsup_{t\downarrow 0} \frac{t}{k_t^2} \log c(k_t, t) \leq \frac{1}{q} \limsup_{t\downarrow 0} \frac{t}{k_t^2} \log \mathbb{P}[X_t \geq k_t] = -\frac{1}{2qv_0}.$$

Now let $p \uparrow \infty$, i.e. $q \downarrow 1$. The same argument yields the lower bound of the implication (ii) $\Longrightarrow$ (i). The remaining upper bound of (ii) $\Longrightarrow$ (i) is shown very similarly to the lower bound of the implication (i) $\Longrightarrow$ (ii). □

In the light of the general MDP result by Gao and Wang [48] quoted at the beginning of this section, it might be worth noting that Theorem 2.3 holds, with virtually the same proof, if the assumption $k_t = \ell(t)t^\beta$ is replaced by $\sqrt{t} \ll k_t \ll 1$.

Appendix

Appendix A Implied Volatility for $\beta \geq 1/3$

Assumption A.1. We refine Assumptions 2.1 and 2.2 as follows

(i) Additionally to Assumption 2.1, the relative error in (2.2) is $\mathcal{O}(t)$, locally uniformly w.r.t. k.

(ii) Convergence of local volatility (see (2.3)) can be refined to

$$\sigma_{\text{loc}}(K, t) = \sigma_0 + \mathcal{O}(t), \quad t \downarrow 0,$$

uniformly for bounded K.

(iii) In (2.2), $\gamma(k)$ satisfies

$$\gamma(0) = \frac{1}{\sqrt{2\pi}\sigma_0}.$$

While *proving* Assumption A.1 for stochastic volatility models would go well beyond the scope of the present work, there are good reasons to believe that it holds in reasonable generality. It does hold for local volatility models, which satisfy part (i) according to Proposition 2.1 of Gatheral et al. [51]. For stochastic volatility models, the approach of Deuschel, Friz, Jacquier and Violante [24] suggests that the relative error term in (2.2) has a full expansion in (integer) powers of t, which would imply (i).
Part (ii) is clear in local volatility models, just assuming smoothness of the local volatility function. In stochastic volatility models, it should be possible to refine the results of Berestycki, Busca and Florent [10] (convergence to σ_0) to a Taylor expansion.
Part (iii) is true for the Heston model (see (2.29)) and generic local volatility models (see (2.23)); the gist of the saddle point argument we applied for Heston, and the resulting expression (2.29), are not tied to that model.

Theorem A.2. *Under Assumption A.1, the statement of Corollary 2.1 holds for* $\beta \in (0, \frac{3}{7})$.

To simplify notation in the following proof, we write $f \sim_\ell g(t)$ for two functions f and g, if $f(t)/g(t) \sim \ell(t)$ as $t \downarrow 0$ for some slowly varying function ℓ. We will use this for functions of algebraic growth order, which are then "almost" asymptotically equivalent. The index ℓ in "$\sim_\ell$" is a generic symbol and does not stand for any concrete slowly varying function.

Proof of Theorem A.2. We start by improving the call price expansion from Theorem 2.5, taking into account the asymptotic errors that were made in obtaining it. By part (ii) of Assumption A.1, the relative error in (2.10) is $\mathcal{O}(t)$. Part (ii) of Assumption A.1 shows the same for (2.11). The relative error in (2.14) is $\mathcal{O}(k_t)$, as seen from

$$\Lambda(k_t)^{-1} = \frac{2v_0}{k_t^2}\big(1+\mathcal{O}(k_t)\big), \quad t \downarrow 0.$$

The only remaining source of error is the application of the Laplace method in (2.13). Here, it does not suffice to state the relative error; we have to do a little better than in the proof of Theorem 2.5. By the higher order extension of the Laplace method (Theorem 3.8.1 in Olver [74]), and because $\Lambda(k_t)/t \to \infty$ as $t \downarrow 0$, we have the integral expansion

$$\begin{aligned}\int_0^1 \exp\left(-\frac{\Lambda(k_t)}{tx}\right) x^{-1/2}\,dx &= \int_1^\infty \exp\left(-\frac{\Lambda(k_t)}{t}x\right) x^{-3/2}\,dx \\ &= \exp\left(-\frac{\Lambda(k_t)}{t}\right)\frac{t}{\Lambda(k_t)}\left(1-\frac{3}{2}\frac{t}{\Lambda(k_t)}+\mathcal{O}\left(\frac{t^2}{\Lambda(k_t)^2}\right)\right),\end{aligned}$$

with error term $t^2/\Lambda(k_t)^2 \sim_\ell t^{2(1-2\beta)}$. Therefore, our MOTM call price approximation becomes

$$c(k_t,t) = \frac{\gamma(0)\sigma_0^2}{2}\frac{t^{3/2}}{\Lambda(k_t)}\exp\left(-\frac{\Lambda(k_t)}{t}\right)\left(1-\frac{3}{2}\frac{t}{\Lambda(k_t)}+\mathcal{O}(t^{2(1-2\beta)-\varepsilon})\right),$$

where $\varepsilon > 0$ is arbitrarily small. The Taylor expansion $\Lambda(k) = \frac{1}{2}\Lambda''(0)k^2 + \frac{1}{6}\Lambda'''(0)k^3 + \mathcal{O}(k^4)$ implies

$$\begin{aligned}L_t := -\log c(k_t,t) &= \frac{\Lambda(k_t)}{t} - \log\left(\frac{\gamma(0)\sigma_0^2}{2}\frac{t^{3/2}}{\Lambda(k_t)}\right) + \frac{3}{2}\frac{t}{\Lambda(k_t)} + \mathcal{O}(t^{2(1-2\beta)-\varepsilon}) \\ &= \frac{1}{2}\Lambda''(0)\frac{k_t^2}{t} + \log\left(\frac{k_t^2}{t^{3/2}\sigma_0^4\gamma(0)}\right) + \frac{3}{\Lambda''(0)}\frac{t}{k_t^2} + \frac{1}{6}\Lambda'''(0)\frac{k_t^3}{t} + \mathcal{O}(t^{\min\{2(1-2\beta),\beta\}-\varepsilon}).\end{aligned}$$

We now translate the refined call price expansion to implied volatility asymptotics. In the proof of Corollary 2.1, we used Corollary 7.2 of Gao and Lee [49] to achieve the transfer. This would suffice for the interval $\beta \in (0, \frac{2}{5})$, too, but for the larger interval $\beta \in (0, \frac{3}{7})$ we have to take a closer look at the (arbitrary order) asymptotic machinery developed in Gao and Lee [49]. Any unexplained terminology and notation in what follows is as therein. Using Proposition 5.6, Lemma 5.8 and Example 5.13 of Gao and

Lee [49] yields the following estimates in our MOTM regime

$$|G_-(k,\phi(k,L)) - V| = \mathcal{O}\left(\frac{k}{L^{3/2}}\left(\frac{\Psi}{L^P} + \frac{1}{L^N}\right)\right) \tag{2.46}$$

$$|G_-(k,\phi(k,L)) - G_-(k,\phi(k,\hat{L}))| = \mathcal{O}\left(\frac{k}{L^{1/2}}\frac{|L-\hat{L}|}{L}\right) \tag{2.47}$$

$$\left|G_-(k,\phi(k,\hat{L})) - \frac{k}{\sqrt{2\phi(k,\hat{L})+k}}\right| = \mathcal{O}\left(\frac{k}{L^{1/2}}\frac{k^2}{L^2}\right) \tag{2.48}$$

with integers $N, P \geq 1$, $L = -\log c(k_t,t)$, an approximation $\hat{L}$ of L, dimensionless implied volatility $V := t^{1/2}\sigma_{\text{imp}}(k_t,t)$, and error estimate Ψ. We suppress the time dependence of k, L, $\hat{l}$, V, and Ψ, in order to keep the notation of Gao and Lee [49]. Note that, in the MOTM regime, $k/L \to 0$ as $t \downarrow 0$.
We have $k \sim_\ell t^\beta$ and $L \sim_\ell t^{2\beta-1}$. The factor $k/L^{1/2} \sim_\ell t^{1/2}$ in (2.46)-(2.48) corresponds to the $t^{1/2}$-term of the dimensionless implied volatility V. The error term $k^2/L^2 \sim_\ell t^{2(1-\beta)}$ in (2.48) is of negligible order. Therefore, we have to deal with the iteration scheme error $(\Psi/L^P + 1/L^N)/L$ in (2.46) and the approximation error $|L-\hat{L}|/L$ in (2.47).
We now define the iteration scheme, following Gao and Lee [49]. The choice $N = 2$ and $P = 2$ suffices for our needs. Define a 2-ply regular iteration scheme $H := \{h, \eta_1, \eta_2\}$ via the sub-log functions

$$\begin{aligned}
h(\kappa,\lambda) &:= \alpha(\kappa,\lambda) - \frac{3}{2\lambda}\\
\eta_1(\kappa,\lambda) &:= \frac{3}{2\lambda}\\
\eta_2(\kappa,\lambda) &:= \frac{3}{2}\left(\log\left(1+\frac{\alpha(\kappa,\lambda)}{\lambda}\right) - \frac{\alpha(\kappa,\lambda)}{\lambda}\right) + \frac{3}{2\lambda}\left(\left(1+\frac{\alpha(\kappa,\lambda)}{\lambda}\right)^{-1} - 1\right),
\end{aligned}$$

using the auxiliary function

$$\alpha(\kappa,\lambda) := -\frac{3}{2}\log\lambda + \log\left(\frac{\kappa}{4\sqrt{\pi}}\right).$$

The corresponding implied volatility approximation function is given by

$$\phi_H(\kappa,\lambda) := \lambda + \alpha(\kappa,\lambda) - \frac{3}{2\lambda}(\alpha(\kappa,\lambda)+1).$$

The 2-residual Ψ of H yields a sufficiently small iteration scheme error in (2.46):

$$\frac{1}{L}\left(\frac{\Psi}{L^2} + \frac{1}{L^2}\right) \sim_\ell t^{\min\{3(1-2\beta),\beta+2(1-2\beta)\}}. \tag{2.49}$$

Define $\kappa_t := \frac{1}{2\sigma_0^2}\frac{k_t^2}{t}$ for $t > 0$ and the constant $C := \frac{1}{3}\sigma_0^2\Lambda'''(0)$. By inserting $\Lambda''(0) = \sigma_0^{-2}$ and $\gamma(0) = (\sqrt{2\pi}\sigma_0)^{-1}$ (see part (iii) of Assumption A.1), we get the approximation

$$\begin{aligned}\hat{L}_t &:= \frac{1}{2\sigma_0^2}\frac{k_t^2}{t} + \ell_0(t) + 3\sigma_0^2\frac{t}{k_t^2} + \frac{1}{6}\Lambda'''(0)\frac{k_t^3}{t}\\ &= \kappa_t\left(1 + \ell_0(t)\kappa_t^{-1} + \frac{3}{2}\kappa_t^{-2} + Ck_t\right),\end{aligned}$$

where $\ell_0(t) := \log(\sqrt{2\pi}k_t^2/(\sigma_0^3 t^{3/2}))$ is slowly varying. Thus, the approximation error in (2.47) is

$$\frac{|L - \hat{L}|}{L} \sim_\ell t^{\min\{3(1-2\beta),1-\beta\}}. \tag{2.50}$$

It remains to put together all ingredients. By (2.46), (2.47) and (2.48), combined with (2.49) and (2.50), the following approximation of the dimensionless implied volatility $V_t := t^{1/2}\sigma_{\text{imp}}(k_t, t)$ holds

$$\left|\frac{k_t}{\sqrt{2\phi_H(k_t,\hat{L}_t) + k_t}} - V_t\right| \sim_\ell t^{1/2+\min\{3(1-2\beta),1-\beta\}}.$$

Further calculations show

$$\begin{aligned}&\frac{k_t}{\sqrt{2\phi_H(k_t,\hat{L}_t) + k_t}} \qquad\qquad (2.51)\\ &= \frac{k_t}{\sqrt{2\hat{L}_t}}\left(1 + \frac{k_t}{2\hat{L}_t} + \frac{1}{\hat{L}_t}\alpha(k_t,\hat{L}_t) - \frac{3}{2\hat{L}_t^2}(\alpha(k_t,\hat{L}_t) + 1)\right)^{-1/2}\\ &= \frac{k_t}{\sqrt{2\hat{L}_t}}\left(1 - \frac{1}{2\hat{L}}\alpha(k_t,\hat{L}) + \frac{3}{4\hat{L}^2}\left(\tfrac{1}{2}\alpha(k_t,\hat{L}_t)^2 + \alpha(k_t,\hat{L}) + 1\right) + \mathcal{O}(t^{\min\{3(1-2\beta),1-\beta\}-\varepsilon})\right),\end{aligned}$$

because $k_t/\hat{L}_t \sim_\ell t^{1-\beta}$, $1/\hat{L}_t^3 \sim_\ell t^{3(1-2\beta)}$, and $\alpha(k_t,\hat{L}_t) \sim_\ell 1$. Expansion of the appearing functions yields

$$\begin{aligned}&\alpha(k_t,\hat{L}_t) = -\ell_0(t) - \tfrac{3}{2}\ell_0(t)\kappa_t^{-1} + \mathcal{O}(t^{\min\{2(1-2\beta),\beta\}-\varepsilon}),\\ &\alpha(k_t,\hat{L}_t)^2 = \ell_0(t)^2 + \mathcal{O}(t^{\min\{1-2\beta,\beta\}-\varepsilon}),\\ &\frac{1}{\hat{L}_t} = \kappa_t^{-1} - \ell_0(t)\kappa_t^{-2} + \mathcal{O}(t^{\min\{3(1-2\beta),1-\beta\}-\varepsilon}),\\ &\frac{1}{\hat{L}_t^2} = \kappa_t^{-2} + \mathcal{O}(t^{\min\{3(1-2\beta),\beta+2(1-2\beta)\}-\varepsilon}),\\ &\frac{1}{\sqrt{2\hat{L}}} = (2\kappa_t)^{-1/2}\left(1 - \frac{1}{2}\ell_0(t)\kappa_t^{-1} + \frac{3}{8}\ell_0(t)^2\kappa_t^{-2} - \frac{3}{4}\kappa_t^{-2} - \frac{1}{2}Ck_t + \mathcal{O}(t^{\min\{3(1-2\beta),2\beta\}-\varepsilon})\right).\end{aligned}$$

Putting these formulas back into (2.51), we get

$$\frac{k_t}{\sqrt{2\phi_H(k_t,\hat{L}_t)+k_t}} = \sigma_0 t^{1/2}\left(1-\frac{1}{2}Ck_t+\mathcal{O}(t^{\min\{3(1-2\beta),2\beta,1-\beta\}-\varepsilon})\right) \tag{2.52}$$

$$= t^{1/2}(\sigma_0 - \tfrac{1}{6}\sigma_0^3\Lambda'''(0)k_t) + \mathcal{O}(t^{1/2+\min\{3(1-2\beta),2\beta,1-\beta\}-\varepsilon}).$$

For the second order expansion of the implied volatility to be correct, the error term should be negligible compared to k_t, which amounts to $t^{\min\{3(1-2\beta),2\beta,1-\beta\}} = o(k_t)$. This is true if and only if $\min\{3(1-2\beta),2\beta,1-\beta\} > \beta$, which is equivalent to our assumption $\beta \in (0, \frac{3}{7})$. □

For larger β, closer to $\frac{1}{2}$, the whole analysis has to be refined. A more precise iteration scheme H has to be chosen, so that the iteration scheme error in (2.49) gets smaller. Moreover, a better log-price approximation $\hat{L}_t$ has to be taken into account, using even more terms of the Laplace expansion, in order to decrease the approximation error in (2.50). It should thus be possible to reduce the error in (2.52) to

$$\mathcal{O}(t^{\min\{n(1-2\beta),2\beta,1-\beta\}-\varepsilon}), \quad t \downarrow 0,$$

where $n \in \mathbb{N}$ can be arbitrarily large. In this fashion, for any *fixed* n, it should be straightforward to provide a proof of the second order approximation (2.16) of the implied volatility for $\beta < \frac{n}{2n+1}$. That is, we have a clear procedure for any $n > 2$. For small n, say $n = 3, 4, \ldots$, this can be implemented by hand, and larger values (say, $n = 17$) are still feasible with the aid of Mathematica or similar software. In practice, as the calculations in each proof will be tied to that specific value of n, very large n remains out of reach. Here one would need a new idea to provide an argument for general n, which would then prove (2.16) for all $\beta \in (0, \frac{1}{2})$. At this moment, despite some effort, the details of such a construction elude us. Still, we believe that Assumption A.1 suffices to treat the whole interval, i.e. that Theorem A.2 holds with $\frac{3}{7}$ replaced by $\frac{1}{2}$.
Note that Tehranchi [88], which presents uniform (non-asymptotic) bounds on implied volatility, does not seem to be applicable here: For $\beta > \frac{1}{3}$, the lower bound of Proposition 4.6 therein is not tight enough, as it yields a second order term that is asymptotically larger than the second order term k_t in (2.16).

Appendix B A Moderate Deviations Result in the Heston Model

This additional section is not part of the paper [45] on which the whole Part II is based. The goal of this section is to prove Assumption 2.1 for the moment generating function (mgf) in the Heston model. The mgf of the log-price $X_t = \log(S_t)$ with $X_0 = 0$ in the Heston model, denoted as $M(u,t) := \mathbb{E}[e^{uX_t}]$, exhibits an exponential affine structure, i.e.

$$\log M(u,t) = \lambda\bar{v}\phi(u,t) + v_0\psi(u,t), \tag{2.53}$$

where $\phi(u,t) = \int_0^t \psi(u,s)\,ds$ and ψ is the solution of the Riccati differential equation

$$\frac{\partial}{\partial t}\psi(u,t) = R(u,\psi(u,t)), \tag{2.54}$$

$$\psi(u,0) = 0, \tag{2.55}$$

and the function R on the right-hand side of (2.54) is defined as

$$R(u,w) = \frac{1}{2}(u,w)\cdot A \cdot \begin{pmatrix} u \\ w \end{pmatrix} + b \cdot \begin{pmatrix} u \\ w \end{pmatrix} \tag{2.56}$$

with a symmetric, positive definite matrix A and a real vector b given by

$$A = \begin{pmatrix} 1 & \rho\xi \\ \rho\xi & \xi^2 \end{pmatrix} \quad \text{and} \quad b = (-\tfrac{1}{2}, -\lambda). \tag{2.57}$$

Although the functions ϕ and ψ can be calculated explicitly, we will only use the Riccati differential equation (2.54) with initial value (2.55), and the relationship between ϕ and ψ, respectively.

Lemma B.1. *Let $\beta \in (0, \frac{1}{2})$ and $p \in \mathbb{R}$. Then there exists a time $t_0 > 0$ such that for all $t \in (0, t_0)$ the values $\phi(pt^{\beta-1}, t)$ and $\psi(pt^{\beta-1}, t)$ are well-defined.*

Proof. The explosion time $T^*(u)$ in the Heston model, see (2.11) in Part III Chapter 2, depends on the sign of the values $e_0(u)$ and $e_1(u)$, defined in (2.9) and (2.10) in Part III Chapter 2. We have

$$e_0(u) \sim \tfrac{1}{2}\xi\rho u \quad \text{and} \quad e_1(u) \sim -\tfrac{1}{4}\xi^2\bar{\rho}^2 u^2 \quad \text{as } |u| \to \infty, \tag{2.58}$$

with the correlation $\bar{\rho} = \sqrt{1-\rho^2}$. The function $u(t) := pt^{\beta-1}$ goes to $\operatorname{sgn} p \cdot \infty$ as $t \downarrow 0$. Hence, for sufficiently small $t > 0$, the asymptotic relation for e_1 in (2.58) yields $e_1(u(t)) < 0$ and the explosion time $T^*(u(t))$ in this case is given by

$$T^*(u(t)) = \frac{1}{\sqrt{|e_1(u(t))|}}\left(\frac{\pi}{2} - \arctan\left(\frac{e_0(u(t))}{\sqrt{|e_1(u(t))|}}\right)\right).$$

It is easy to check with (2.58) that the explosion time $T^*(u(t)) \sim ct^{1-\beta}$ as $t \downarrow 0$ for some $c > 0$. Therefore, $t = o\big(T^*(u(t))\big)$ as $t \downarrow 0$, and it is possible to find $t_0 > 0$ such that $t < T^*(u(t))$ for all $t \in (0, t_0)$. As a result, the mgf $M(u(t), t)$ and hence $\phi(u(t), t)$ and $\psi(u(t), t)$ exist for all $t \in (0, t_0)$. $\square$

The solution ψ of the Riccati equation (2.54) with initial value (2.55) can be represented implicitly, see e.g. Keller-Ressel [63],

$$\int_0^{\psi(u,t)} \frac{dw}{R(u,w)} = t \tag{2.59}$$

at least for $t > 0$ in a neighbourhood of 0 and large values $|u|$. Note that for large values $|u|$ the solution ψ is obviously non-stationary. Moreover, if $|u|$ is large, both $w \mapsto R(u,w)$ and $t \mapsto \psi(u,t)$ are positive at least for small positive values.

Lemma B.2. *Let $\beta \in (0, \frac{1}{2})$ and $p \in \mathbb{R}$. Then $\psi(pt^{\beta-1}, t) = o(t^{\beta-1})$ as $t \downarrow 0$.*

Proof. Define $\Psi(t) := \psi(pt^{\beta-1}, t) \wedge g(t)$ where the function g satisfies $g(t) = o(t^{\beta-1})$ as $t \downarrow 0$ and will be specified later. From the integral representation (2.59) and the convexity of $R(u, \cdot)$ for every $u \in \mathbb{R}$, we have

$$\begin{aligned} t &= \int_0^{\psi(pt^{\beta-1},t)} \frac{dw}{R(pt^{\beta-1}, w)} \\ &\geq \int_0^{\Psi(t)} \frac{dw}{R(pt^{\beta-1}, w)} \\ &\geq \Psi(t) \frac{1}{\max\{R(pt^{\beta-1}, 0), R(pt^{\beta-1}, \Psi(t))\}}. \end{aligned} \tag{2.60}$$

By (2.56), the following asymptotic relation holds

$$R(pt^{\beta-1}, 0) \sim \tfrac{1}{2} p^2 t^{2\beta-2}, \quad t \downarrow 0. \tag{2.61}$$

Furthermore, $\Psi(t) = o(t^{\beta-1})$ as $t \downarrow 0$ and (2.61) yield

$$\begin{aligned} R(pt^{\beta-1}, \Psi(t)) &= R(pt^{\beta-1}, 0) + \tfrac{1}{2} A_{22} \Psi(t)^2 + A_{12} \Psi(t) p t^{\beta-1} + b_2 \Psi(t) \\ &= R(pt^{\beta-1}, 0) + o(t^{2\beta-2}). \end{aligned}$$

Thus,

$$R(pt^{\beta-1}, \Psi(t)) \sim R(pt^{\beta-1}, 0), \quad t \downarrow 0. \tag{2.62}$$

Combining (2.61) and (2.62) with the inequality (2.60) gives us, as $t \downarrow 0$,

$$\Psi(t) \leq t \max\{R(pt^{\beta-1}, 0), R(pt^{\beta-1}, \Psi(t))\} = \mathcal{O}(t^{2\beta-1}). \tag{2.63}$$

Now, choose $g(t) = t^{\beta-1+\varepsilon}$ and $0 < \varepsilon < \beta$. With this choice g satisfies $g(t) = o(t^{\beta-1})$, but $g(t) \neq \mathcal{O}(t^{2\beta-1})$ as $t \downarrow 0$. Consequently, for the inequality (2.63) to hold, the function $\psi(pt^{\beta-1}, t) = \mathcal{O}(t^{2\beta-1})$ as $t \downarrow 0$ and the statement follows. Note that $\Psi(t) = \psi(pt^{\beta-1}, t)$ for sufficiently small $t > 0$. □

With Lemma B.2 at hand, we are able to determine the asymptotic rate of $\psi(pt^{\beta-1}, t)$ as $t \downarrow 0$.

Theorem B.3. *Let $\beta \in (0, \frac{1}{2})$ and $p \in \mathbb{R}$. Then $\psi(pt^{\beta-1}, t) \sim \frac{1}{2} p^2 t^{2\beta-1}$ as $t \downarrow 0$.*

Proof. Define $\Psi(t) := \psi(pt^{\beta-1}, t)$. Similar to (2.60), we have the inequality

$$t \min_{w \in [0, \Psi(t)]} R(pt^{\beta-1}, w) \leq \Psi(t) \leq t \max_{w \in [0, \Psi(t)]} R(pt^{\beta-1}, w). \tag{2.64}$$

Because $R(u, \cdot)$ is a convex quadratic function for every $u \in \mathbb{R}$, the minimum can only be attained at the boundary points or the global minimum, whereas the maximum value is found at the boundary points. In (2.61) and (2.62) we determined the asymptotic rate of the maximum, hence the right-hand side of (2.64) is asymptotically equivalent

to $\frac{1}{2}p^2t^{2\beta-1}$ as $t \downarrow 0$. We only have to show now, that the left-hand side of (2.64) has the same asymptotic rate.
The global minimum of $w \mapsto R(u,w)$ is attained at $w_{\min}(u) = -(A_{12}u + b_2)/A_{22}$. If $A_{12} = 0$, then $w_{\min}(u) \equiv \bar{w}_{\min}$ is constant and by (2.56) a straightforward computation shows

$$R(pt^{\beta-1}, \bar{w}_{\min}) \sim \tfrac{1}{2}p^2t^{2\beta-1}, \quad t \downarrow 0. \tag{2.65}$$

If $A_{12} \neq 0$, then

$$w_{\min}(pt^{\beta-1}) \sim -\frac{A_{12}}{A_{22}}pt^{\beta-1}, \quad t \downarrow 0 \tag{2.66}$$

Using Lemma B.2 and (2.66), it follows that $w_{\min}(pt^{\beta-1}) \notin [0, \Psi(t)]$ for sufficiently small $t > 0$. In this case, the minimum on the left-hand side of (2.64) is attained at the boundary. Either way, combined with (2.61) and (2.62), we have established

$$t \min_{w \in [0,\Psi(t)]} R(pt^{\beta-1}, w) \sim \tfrac{1}{2}p^2t^{2\beta-1}, \quad t \downarrow 0,$$

which yields the desired asymptotic relation. □

Corollary B.4. *For all $\beta \in (0, \frac{1}{2})$, the rescaled mgf in the Heston model satisfies*

$$\lim_{t \downarrow 0} t^{1-2\beta} \log M(pt^{\beta-1}, t) = \tfrac{1}{2}\sigma_0^2 p^2, \quad p \in \mathbb{R}.$$

Proof. From Theorem B.3 we have $\psi(\lambda t^{\beta-1}, t) \sim \frac{1}{2}p^2t^{2\beta-1}$ as $t \downarrow 0$. For large values $|u|$ the function $w \mapsto R(u,w)$ is positive in a neighbourhood of 0, hence $t \mapsto \psi(u,t)$ is increasing for sufficiently small $t > 0$. Thus, we have

$$\begin{aligned} \phi(pt^{\beta-1}, t) &= \int_0^t \psi(pt^{\beta-1}, s)\, ds \\ &\leq t\psi(pt^{\beta-1}, t) \sim \tfrac{1}{2}p^2t^{2\beta}, \quad t \downarrow 0, \end{aligned}$$

so that $\phi(pt^{\beta-1}, t) = o(1)$ as $t \downarrow 0$. Using the representation (2.53), an explicit computation yields

$$\begin{aligned} \lim_{t \downarrow 0} t^{1-2\beta} \log M(pt^{\beta-1}, t) &= \lim_{t \downarrow 0} \Big(\lambda \bar{v} t^{1-2\beta}\phi(pt^{\beta-1}, t) + v_0 t^{1-2\beta}\psi(pt^{\beta-1}, t)\Big) \\ &= \tfrac{1}{2}v_0 p^2, \end{aligned}$$

where the spot variance v_0 is the squared spot volatility, $v_0 = \sigma_0^2$. □

Part III

Moment Explosion in the Rough Heston Model

Chapter 1

Introduction

In times when markets are swamped by high-frequency data and trading is done by computers with immense computational power, the need for appropriate financial models is rising fast. In recent years the importance of rough volatility models has grown considerably because empirical data has shown that historical volatility time-series exhibit a behaviour that is much rougher than the volatility processes in classic stochastic volatility models driven by Brownian motion. Modern models try to overcome these shortcomings by replacing the driving Brownian motion by a fractional Brownian motion. Microstructural foundations for rough volatility models can be found in El Euch, Fukusawa and Rosenbaum [76] and Jaisson and Rosenbaum [61].

A recent rough volatility model, put forward by Rosenbaum and El Euch [29], deals with the well-known Heston model and generalises the volatility process by adding an extra smoothness parameter $\alpha \in (\frac{1}{2}, 0)$. The square-root process given in integral form

$$V_t = v_0 + \int_0^t \lambda(\bar{v} - V_s)\, ds + \int_0^t \xi \sqrt{V_s}\, dZ_t,$$

is replaced by the modified Heston-like rough volatility process

$$V_t = v_0 + \int_0^t k(t-s)\lambda(\bar{v} - V_s)\, ds + \int_0^t k(t-s)\xi \sqrt{V_s}\, dZ_t$$

with parameter $\lambda, \bar{v}, \xi, v_0 > 0$ and kernel $k(z) = z^{\alpha-1}/\Gamma(\alpha)$. This integral kernel k in the above SDE is motivated by the Mandelbrot-van Ness representation of fractional Brownian motion W^H with Hurst parameter $H \in (0, 1)$

$$W_t^H = \frac{1}{\Gamma(H+\frac{1}{2})} \int_{-\infty}^0 \left((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}\right) dW_s + \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}}\, dW_s.$$

At this point, we want to mention that other extensions of the Heston model exist as well, see e.g. Guennoun, Jacquier and Roome [55].

Rosenbaum and El Euch [29] show in their fractional version of the Heston model, which we will call *rough Heston model* from now on, that the volatility paths are almost surely Hölder-continuous with Hölder coefficient $\alpha - \frac{1}{2} - \varepsilon$. The classic Heston model is retrieved in the case $\alpha = 1$. Furthermore, the most important feature of the model is the availability of a semi-closed form of the characteristic function which resembles the characteristic function of the Heston model, see Chapter 2 and 3 for the details. A characteristic function at hand is an exceptional feature for every practitioner. It enables to speed up the model calibration and pricing of derivatives with efficient numerical schemes.

Our interest lies in the blow-up behaviour of the moment generating function and the critical moments. In the Heston model the moment explosion time $T^*(u) = \sup\{t > 0\colon \mathbb{E}[e^{uX_t}] < \infty\}$ is explicitly given, see Andersen and Piterbarg [5]. Then the critical moments $u_+(T)$ and $u_-(T)$ can be obtained by solving the equation $T^*(u_+(T)) = T$ for fixed $T > 0$. In the rough Heston model, however, an explicit expression for T^* may be out of reach, but we prove that the moment explosion time $T^*(u)$ in the classic and in the rough Heston model is finite for the same values of $u \in \mathbb{R}$. In the cases when the moment generating function explodes, i.e. when T^* is finite, we give upper and lower bounds for T^*. Furthermore, we obtain that the critical moments $u_+(T)$ and $u_-(T)$ in the rough Heston are finite for any time $T > 0$.

Chapter 2

The Classic Heston Model

In the classic Heston model proposed by Heston [60] the asset price process $(S_t)_{t\geq 0}$ is described as

$$dS_t = S_t\sqrt{V_t}\,dW_t, \quad S_0 > 0, \tag{2.1}$$
$$dV_t = \lambda(\bar{v} - V_t)\,dt + \xi\sqrt{V_t}\,dZ_t, \quad V_0 = v_0 > 0, \tag{2.2}$$
$$d\langle W, Z\rangle_t = \rho\,dt, \tag{2.3}$$

where W and Z are correlated Brownian motions with correlation parameter $|\rho| < 1$, mean reversion rate $\lambda > 0$, long-run variance $\bar{v} > 0$, volatility of variance $\xi > 0$ and initial variance v_0.
The moment generating function (mgf) of the log-price $X_t := \log(S_t/S_0)$ for $t \geq 0$ in this model is given by

$$\mathbb{E}[e^{uX_t}] = \exp\left(\bar{v}\lambda I_t^1\psi(u,t) + v_0\psi(u,t)\right) \tag{2.4}$$

where $I_t^1 f = \int_0^\cdot f(s)\,ds$ is the classic Lebesgue integral and ψ is the unique solution of the Riccati differential equation

$$\frac{\partial}{\partial t}\psi(u,t) = R(u,\psi(u,t)), \tag{2.5}$$
$$\psi(u,0) = 0, \tag{2.6}$$

where the function R on the right-hand side of (2.5) is defined as

$$R(u,w) = c_1(u) + c_2(u)w + c_3w^2 \tag{2.7}$$

with coefficients $c_1(u) = \frac{1}{2}u(u-1)$, $c_2(u) = \rho\xi u - \lambda$ and $c_3 = \frac{1}{2}\xi^2 > 0$.
The solution of (2.5) and (2.6) can be calculated explicitly, see Heston [60]. Depending on the coefficients, the solution, and hence the moment generating function (2.4), can blow up in finite time.[1] The moment explosion time $T^*_{\mathrm{He}}(u) = \sup\{t \geq 0\colon \mathbb{E}[S_t^u] < \infty\}$ for every $u \in \mathbb{R}$ in the Heston model is given by

$$T^*_{\mathrm{He}}(u) = \begin{cases} \int_0^\infty \frac{1}{R(u,w)}\,dw, & R(u,\cdot) \text{ has no roots on } [0,\infty), \\ \infty, & \text{otherwise.} \end{cases} \tag{2.8}$$

[1] We say that a continuous function h blows up in finite time if a finite time $T^* > 0$ exists such that $\lim_{t\to T^*-}|h(t)| = \infty$. Such a time T^* is called blow-up time or explosion time.

The roots of $R(u,\cdot)$ are located at the points $\frac{1}{c_3}\left(-e_0(u) \pm \sqrt{e_1(u)}\right)$ with

$$e_0(u) = \tfrac{1}{2}c_2(u) = \tfrac{1}{2}(\rho\xi u - \lambda) \tag{2.9}$$

$$e_1(u) = e_0(u)^2 - c_3 c_1(u) = e_0(u)^2 - \tfrac{1}{4}\xi^2 u(u-1) \tag{2.10}$$

In the case $c_1(u) \leq 0$, equivalently $u \in [0,1]$, $R(u,\cdot)$ always has at least one real non-negative root, therefore $T^*_{\mathrm{He}}(u) = \infty$, i.e. there is no explosion in finite time.
In the case $c_1(u) > 0$, explicit calculations of the integral in (2.8) yield the well-known formulas, see e.g. Anderson and Piterbarg [5],

$$T^*_{\mathrm{He}}(u) = \begin{cases} \frac{1}{\sqrt{|e_1(u)|}}\left(\frac{\pi}{2} - \arctan\left(\frac{e_0(u)}{\sqrt{|e_1(u)|}}\right)\right), & e_1(u) < 0, \\ \frac{1}{2\sqrt{e_1(u)}}\log\left(\frac{e_0(u)+\sqrt{e_1(u)}}{e_0(u)-\sqrt{e_1(u)}}\right), & e_1(u) \geq 0, e_0(u) > 0, \\ \infty, & e_1(u) \geq 0, e_0(u) < 0. \end{cases} \tag{2.11}$$

Note that if $c_1(u) > 0$ and $e_1(u) \geq 0$, then $e_0(u) \neq 0$ due to the relationship (2.10).

Chapter 3

The Rough Heston Model

In the rough Heston model, put forward by Rosenbaum and El Euch [29], the asset price process $(S_t)_{t\geq 0}$ is described as

$$dS_t = S_t\sqrt{V_t}\,dW_t, \quad S_0 > 0, \tag{3.1}$$

$$V_t = v_0 + \int_0^t k(t-s)\lambda(\bar{v} - V_s)\,ds + \int_0^t k(t-s)\xi\sqrt{V_s}\,dZ_t, \tag{3.2}$$

$$d\langle W, Z\rangle_t = \rho\,dt, \tag{3.3}$$

where W and Z are correlated Brownian motions with the same parameters λ, $\bar{v}$, ξ, v_0 and ρ as in the classic Heston model (2.1)-(2.3), with an additional smoothness parameter $\alpha \in (\frac{1}{2}, 1)$.
In Rosenbaum and El Euch [29], a semi-explicit representation of the moment generating function (mgf) of the log-price $X_t = \log(S_t/S_0)$ for $t \geq 0$ in this model was established. For correlation parameter $\rho \in (-1/\sqrt{2}, 1/\sqrt{2})$, the mgf is given by

$$\mathbb{E}[e^{uX_t}] = \exp\left(\bar{v}\lambda I_t^1\psi(u,t) + v_0 I_t^{1-\alpha}\psi(u,t)\right) \tag{3.4}$$

where ψ is the unique continuous solution of the fractional-order Riccati differential equation

$$D_t^\alpha\psi(u,t) = R(u, \psi(u,t)), \tag{3.5}$$

$$I_t^{1-\alpha}\psi(u,0) = 0, \tag{3.6}$$

with the same function R as in the Heston model, defined in (2.7), and fractional integral I^α and fractional derivate D^α, defined in (3.7) and (3.8).

3.1 Fractional Integral and Fractional Derivative

Fractional calculus generalises the concept of integration and derivation by defining fractional powers of the ordinary differential and integral operator. In the literature, several kinds of fractional integrals and derivatives, defined in various ways, are considered. One of the most common types is the Riemann-Liouville fractional integral and derivate.

Definition 3.1 (Riemann-Liouville fractional integral and derivate)**.** The left-sided Riemann-Liouville fractional integral I_t^α of order $\alpha \in (0, \infty)$ started at 0 with respect to the integration variable t is given by

$$I_t^\alpha f(t) := \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s)\, ds \tag{3.7}$$

whenever the integral exists, and the left-sided Riemann-Liouville fractional derivative D_t^α of order $\alpha \in [0, 1)$ started at 0 with respect to the integration variable t is given by

$$D_t^\alpha f(t) := \frac{d}{dt} I^{1-\alpha} f(t) \tag{3.8}$$

whenever this expression exists.

The fractional derivative D_t^α can be defined for $\alpha > 1$ as well, but our focus is only on fractional derivatives with $\alpha \in (0, 1)$. Some useful results, which we will use later, are summarised in the next lemma. For further information on fractional calculus, see Samko et al. [84] and Kilbas et al. [64].

Lemma 3.2. *Let $\alpha \in (0, \infty)$ and $T > 0$.*

(i) The fractional integral and derivative of power functions can be easily calculated

$$I_t^\alpha t^\nu = t^{\nu+\alpha} \frac{\Gamma(\nu+1)}{\Gamma(\nu+\alpha+1)} \qquad \textit{for } \nu > -1, \tag{3.9}$$

$$D_t^\alpha t^\nu = t^{\nu-\alpha} \frac{\Gamma(\nu+1)}{\Gamma(\nu-\alpha+1)} \qquad \textit{for } \nu > -1+\alpha. \tag{3.10}$$

(ii) The fractional integral operators satisfy the semigroup property on $C([0, T])$, i.e.

$$I_t^{\alpha_1} I_t^{\alpha_2} = I_t^{\alpha_1+\alpha_2} \qquad \textit{for } \alpha_1, \alpha_2 \in (0, \infty). \tag{3.11}$$

(iii) For $f \in C([0, T])$ the following equation holds

$$D_t^\alpha I_t^\alpha f(t) = f(t). \tag{3.12}$$

(iv) For $f \in C([0, T])$ such that $D_t^\alpha f \in C([0, T])$ the following equation holds

$$I_t^\alpha D_t^\alpha f(t) = f(t) - \frac{I_t^{1-\alpha} f(0)}{\Gamma(\alpha)} t^{\alpha-1}. \tag{3.13}$$

Proof. (i) and (ii) are straightforward computations using the relationship between the gamma and beta function. (iii) and (iv) are special cases of Theorem 2.4 in Samko et al. [84]. □

We state the following uniqueness and existence theorem for initial value problems with fractional differential equations and the relationship to Volterra integral equations.

Theorem 3.3. *Let $0 < \alpha < 1$ and $G \subseteq \mathbb{R}$ be an open set. Let $f\colon (a,b] \times G \to \mathbb{R}$ be a function such that $f(\cdot, w) \in C_{1-\alpha}([a,b])$ for any $w \in G$. The space $C_{1-\alpha}([a,b])$ denotes the subspace of continuous functions*

$$C_{1-\alpha}([a,b]) := \{g\colon (t-a)^{1-\alpha} g(t) \in C([a,b])\}.$$

(i) If f is Lipschitz-continuous, then there exists a unique solution $\psi \in C_{1-\alpha}([a,b])$ to the initial value problem

$$D_t^\alpha \psi(t) = f(t, \psi(t)), \tag{3.14}$$
$$I^{1-\alpha}\psi(a) = \psi_0 \in \mathbb{R}. \tag{3.15}$$

(ii) If $\psi \in C_{1-\alpha}([a,b])$, then ψ satisfies the fractional integral equation (3.14) with initial value (3.15) if and only if ψ satisfies the Volterra integral equation

$$\psi(t) = \frac{\psi_0}{\Gamma(\alpha)}(t-a)^{\alpha-1} + \frac{1}{\Gamma(\alpha)} \int_a^t (t-s)^{\alpha-1} f(s, \psi(s))\, ds.$$

Proof. The proof can be found in Kilbas et al. [64], Theorem 3.10 and 3.11. □

3.2 Blow-up Behaviour of the Moment Generating Function

Our main interest lies in the blow-up behaviour of the solution ψ of the fractional-order Riccati differential equation (3.5) with initial value (3.6). Rather than dealing with the differential equation, it is more tempting to deal with the corresponding integral equation because in the literature the blow-up behaviour of non-linear Volterra integral equations is well-studied, see Brunner and Yang [14], Mydlarczyk [72] and Mydlarczyk and Okrasiński [73].

Due to the continuity of $R(u,\cdot)$ and the solution $\psi(u,\cdot)$, it follows from (3.5) that the fractional derivative $D_t^\alpha \psi(u,\cdot)$ is continuous as well. With equation (3.13) or with Theorem 3.3 (ii) the Riccati differential equation (3.5) with initial value (3.6) can be transformed into a non-linear Volterra integral equation

$$\psi(u,t) = \int_0^t k(t-s) R(u, \psi(u,s))\, ds \tag{3.16}$$

with the weakly singular kernel $k(z) := \frac{1}{\Gamma(\alpha)} z^{\alpha-1}$.

By defining the function $h(u,t) := c_3 \psi(u,t)$ and completing the square, $h(u,\cdot)$ solves

$$h(u,t) = \int_0^t k(t-s) G(u, h(u,s))\, ds \tag{3.17}$$

with non-linearity $G(u,w) = c_3 R(u, w/c_3)$. Note that G satisfies

$$G(u,w) = (w + e_0(u))^2 - e_1(u) \tag{3.18}$$

and the same coefficients $e_0(u)$ and $e_1(u)$ as in (2.9) and (2.10).

Recall the blow-up time $T^*_{\mathrm{He}}(u)$ in the Heston model (2.11). If $c_1(u) < 0$, the solution $\psi(u,\cdot)$ does not blow up in finite time. In the case $c_1(u) > 0$, the blow-up behaviour depends on the parameters $e_0(u)$ and $e_1(u)$.
In order to characterise the blow-up behaviour of the rough Heston model w.r.t. the parameter u, we distinguish between the following cases for $u \in \mathbb{R}$:

(A) $c_1(u) > 0$, $e_0(u) \geq 0$

(B) $c_1(u) > 0$, $e_0(u) < 0$ and $e_1(u) < 0$

(C) $c_1(u) > 0$, $e_0(u) < 0$ and $e_1(u) \geq 0$

(D) $c_1(u) \leq 0$

We will see later that the moment explosion time in the rough Heston model $T^*(u)$ blows up in finite time if and only if the parameter $u \in \mathbb{R}$ satisfies the conditions in the cases (A) and (B). Note that cases (A) and (B) combined are exactly the cases in which the moment explosion time $T^*_{\mathrm{He}}(u)$ in the classic Heston model is finite, cf. (2.11). In summary it can be said that the mgf of the classic and the rough Heston model blows up for the same values $u \in \mathbb{R}$.

At first, we want to give a result from Brunner and Yang [14] which characterises the blow-up behaviour of non-linear Volterra integral equations when the corresponding functions are positive and increasing.

Proposition 3.1. *Assume that $G\colon [0,\infty) \to [0,\infty)$ is continuous and the following conditions hold:*

(G1) $G(0) = 0$ *and G is strictly increasing.*

(G2) $\lim_{w\to\infty} G(w)/w = \infty$.

(P) $\phi\colon [0,\infty) \to [0,\infty)$ *is a positive, non-decreasing, continuous function.*

(K) $k\colon (0,\infty) \to [0,\infty)$ *is locally integrable and $K(t) := \int_0^t k(z)\,dz > 0$ is a non-decreasing function.*

Furthermore, assume $\lim_{t\to\infty} \phi(t) = \infty$ *and $k(z) = cz^{\alpha-1}$ for $\alpha > 0$ and $c > 0$. Then the solution h of the Volterra integral equation*

$$h(t) = \phi(t) + \int_0^t k(t-s) G(h(s))\,ds$$

blows up in finite time if and only if

$$\int_U^\infty \left(\frac{w}{G(w)}\right)^{1/\alpha} \frac{dw}{w} < \infty \tag{3.19}$$

for all $U > 0$.

Proof. This is a special case of Corollary 2.22 in Brunner and Yang [14] where G does not depend on time. □

In case (A), it is possible to define the functions ϕ and G such that all the assumptions of Proposition 3.1 are satisfied and only the integral-condition (3.19) has to be checked to determine whether the solution h of (3.17) blows up in finite time or not.

Theorem 3.2. *In case (A), the solution h of (3.17) starts at 0, is positive thereafter and blows up in finite time.*

Proof. Fix $u \in \mathbb{R}$ such that $c_1(u) > 0$ and $e_0(u) \geq 0$ and suppress the parameter u in the notation. Note that $e_0^2 - e_1 > 0$ in this case. If we write the Volterra integral equation (3.17) in the form

$$h(t) = \phi(t) + \int_0^t k(t-s)\bar{G}(h(s))\, ds$$

with non-linearity $\bar{G}(w) = w^2 + 2e_0 w$ and $\phi(t) = \frac{e_0^2 - e_1}{\Gamma(1+\alpha)} t^\alpha$, using (3.18) and (3.9), then the conditions $c_1 > 0$ and $e_0 \geq 0$ guarantee that ϕ and $\bar{G}$ are positive and strictly increasing on $(0, \infty)$ with $\phi(0) = \bar{G}(0) = 0$. Hence, the solution h is positive for positive values with $h(0) = 0$. It is easy to check that all the assumptions (G1), (G2), (P) and (K) of Proposition 3.1 are satisfied. Moreover, $\lim_{t\to\infty} \phi(t) = \infty$ and

$$\int_U^\infty \left(\frac{w}{\bar{G}(w)}\right)^{1/\alpha} \frac{dw}{w} \leq \int_U^\infty w^{-1-1/\alpha}\, dw < \infty$$

for all $U > 0$. By Proposition 3.1, the solution h blows up in finite time. □

In case (B), Proposition 3.1 can not be applied directly to the solution h of (3.17). Hence, the Volterra integral equation has to be modified in order to satisfy the assumptions of Proposition 3.1 in a way that h is still a subsolution of the modified equation, i.e. h satisfies the modified equation (3.17) with "$\geq$" instead of "$=$". At first, we need a comparison lemma for solutions and subsolutions.

Lemma 3.3. *Let $G\colon [0, \infty) \to (0, \infty)$ be an strictly increasing, continuous function and $T > 0$. Suppose that g is the unique continuous solution of the Volterra integral equation*

$$g(t) = \int_0^t k(t-s)G(g(s))\, ds, \quad t \in [0, T].$$

If h is a continuous subsolution,

$$h(t) \geq \int_0^t k(t-s)G(h(s))\,ds, \quad t \in [0,T],$$

then $h(t) \geq g(t)$ holds for all $t \in [0,T]$.

Proof. The idea of the proof is based on the proof of Lemma 2.4 in [14].
For any $c \in (0,T)$ define $h_c(t) := h(t+c)$ for $t \in [0, T-c]$. From the positivity of G, it follows that $h_c(0) = h(c) > 0$ and

$$h_c(t) \geq \int_0^t k(t-s)G(h_c(s))\,ds, \quad t \in [0, T-c].$$

Since $g(0) = 0$, it follows that $g(0) < h_c(0)$. We want to show $g < h_c$ on the whole interval $[0, T-c]$. Therefore, suppose that $t \in (0, T-c]$ exists such that $0 \leq g(s) < h_c(s)$ for all $s \in [0,t)$ and $g(t) = h_c(t)$. However, because G is strictly increasing, we have

$$0 = h_c(t) - g(t) \geq \int_0^t k(t-s)(G(h_c(s)) - G(g(s)))\,ds > 0,$$

which is a contradiction. Hence, the inequality $g(t) < h_c(t) = h(t+c)$ holds for all $t \in [0, T-c]$. Since $c \in (0,T)$ was arbitrary, the result follows easily. □

Theorem 3.4. *In case (B), the solution h of (3.17) starts at 0, is positive thereafter and blows up in finite time.*

Proof. Fix $u \in \mathbb{R}$ such that $c_1(u) > 0$, $e_0(u) < 0$ and $e_1(u) < 0$ and suppress the parameter u in the notation. Note that in this case, the non-linearity G is obviously positive by (3.18). However, G is strictly decreasing on $[0, -e_1]$. Therefore, to deal with this problem, let $0 < a < -e_1$ and define the modified non-linearity $\bar{G}_a$ as

$$\bar{G}_a(w) = \begin{cases} w^{\frac{a+e_1}{e_0}} + a, & 0 \leq w < -e_0, \\ G(w), & w \geq -e_0. \end{cases} \tag{3.20}$$

Then $\bar{G}_a$ is a positive, strictly increasing, Lipschitz-continuous function that starts at a and $\bar{G}_a \leq G$. Let $\bar{h}$ be the unique continuous solution, cf. Theorem 3.3, of the Volterra integral equation

$$\bar{h}(t) = \int_0^t k(t-s)\bar{G}_a(\bar{h}(s))\,ds = \phi(t) + \int_0^t k(t-s)\bar{G}(\bar{h}(s))\,ds$$

with $\bar{G} = \bar{G}_a - a$ and $\phi(t) = \frac{a}{\Gamma(1+\alpha)}t^\alpha$, using (3.9). Due to the positivity of ϕ and $\bar{G}$ on $(0,\infty)$, the solution $\bar{h}$ is positive on $(0,\infty)$ as well. The functions ϕ, $\bar{G}$ and k satisfy the assumptions (G1), (G2), (P) and (K) in Proposition 3.1. Furthermore, $\lim_{t\to\infty}\phi(t) = \infty$ and $\bar{G}$ satisfies (3.19). By Proposition 3.1, $\bar{h}$ blows up in finite time. Because h satisfies

(3.17) and $\bar{G}_a \leq G$, it follows that h is a subsolution of the modified Volterra integral equation

$$h(t) \geq \int_0^t k(t-s)\bar{G}_a(h(s))\,ds.$$

Now, Lemma 3.3 implies $h(t) \geq \bar{h}(t)$. Consequently, h blows up as well. □

Cases (C) and (D) are those cases where the solution h of (3.17) does not blow up in finite time. In fact, h does not blow up at all, as we will see later. The following lemma provides the key idea for both cases.

Lemma 3.5. *Let $G\colon [0,\infty) \to [0,\infty)$ be a Lipschitz-continuous function that is positive on $[0,a)$ and $G \equiv 0$ on $[a,\infty)$ for an $a > 0$. Then the unique continuous solution h, cf. Theorem 3.3, of the Volterra integral equation*

$$h(t) = \int_0^t k(t-s)G(h(s))\,ds$$

is bounded with $0 \leq h(t) \leq a$ for all $t \geq 0$.

Proof. The non-negativity of G implies $h \geq 0$. Suppose $t > 0$ exists such that $h(t) > a$. By the continuity of h, there exists $0 < t_0 < t$ that satisfies $h(t_0) = a$ and $h(s) > a$ for all $s \in (t_0, t)$. From $G \equiv 0$ on $[a,\infty)$, we have

$$\int_{t_0}^t k(t-s)G(h(s))\,ds = 0.$$

Since G is non-negative and k is decreasing,

$$\begin{aligned}
0 &< h(t) - h(t_0) \\
&= \int_{t_0}^t k(t-s)G(h(s))\,ds + \int_0^{t_0} (k(t-s) - k(t_0-s))G(h(s))\,ds \\
&= \int_0^{t_0} (k(t-s) - k(t_0-s))G(h(s))\,ds \leq 0,
\end{aligned}$$

which is a contradiction. As a result, h is bounded with $0 \leq h(t) \leq a$ for all $t \geq 0$. □

Theorem 3.6. *In case (C), the solution h of (3.17) is non-negative and bounded, and exists globally.*

Proof. Fix $u \in \mathbb{R}$ such that $c_1(u) > 0$, $e_0(u) < 0$ and $e_1(u) \geq 0$ and suppress the parameter u in the notation. Note that the inequality $0 \leq e_1 = e_0^2 - c_1c_3 < e_0^2$ implies $a := -e_0 - \sqrt{e_1} > 0$. Moreover, from (3.18), it follows that a is the smallest positive root of G. Define the non-linearity $\bar{G}$ as

$$\bar{G}(w) := \begin{cases} G(w), & 0 \leq w \leq a, \\ 0, & w > a. \end{cases}$$

Then $\bar{G}$ is a non-negative, Lipschitz-continuous function that starts at $e_0^2 - e_1 > 0$. Therefore, Lemma 3.5 yields that the unique continuous solution $\bar{h}$ of

$$\bar{h}(t) = \int_0^t k(t-s)\bar{G}(\bar{h}(s))\,ds.$$

is bounded with $0 \leq \bar{h}(t) \leq a$ for all $t \geq 0$. Since $\bar{G} = G$ on $[0, a]$, the function $\bar{h}$ solves the original Volterra integral equation

$$\bar{h}(t) = \int_0^t k(t-s)G(\bar{h}(s))\,ds$$

and from the uniqueness of the solution we obtain $h = \bar{h}$. □

Theorem 3.7. *In case (D), the solution h of (3.17) is non-positive and bounded, and exists globally.*

Proof. Fix $u \in \mathbb{R}$ such that $c_1(u) \leq 0$, which means $u \in [0,1]$, and suppress the parameter u in the notation. Note that $e_1 = e_0^2 - c_1 c_3 > e_0^2 > 0$ implies $a := \sqrt{e_1} - e_0 > 0$. Moreover, from (3.18), it follows that a is the smallest positive root of G. Define $h_- := -h$, then h_- solves

$$h_-(t) = -\int_0^t k(t-s)G(-h_-(s))\,ds. \tag{3.21}$$

If we define the non-linearity $\bar{G}$ as

$$\bar{G}(w) := \begin{cases} -G(-w), & 0 \leq w \leq a, \\ 0, & w > a, \end{cases}$$

then $\bar{G}$ is a non-negative, Lipschitz-continuous function that starts at $e_1 - e_0^2 > 0$. With Lemma 3.5 we obtain that the unique continuous solution $\bar{h}$ of

$$\bar{h}(t) = \int_0^t k(t-s)\bar{G}(\bar{h}(s))\,ds$$

is bounded with $0 \leq \bar{h}(t) \leq a$ for all $t \geq 0$. Furthermore, $\bar{h}$ solves (3.21) because $\bar{G}(w) = -G(-w)$ for all $w \in [0,a]$. The uniqueness of the solution yields $\bar{h} = h_- = -h$. Hence, the solution h is bounded with $-a \leq h(t) \leq 0$. □

We have shown in which cases the solution h of the Volterra integral equation (3.17), respectively the solution ψ of the fractional-order Riccati differential equation (3.5) with initial value (3.6), blows up in finite time. The question remains, whether the blow-up behaviour of ψ is transferred to the moment generating function (3.4) or not. The answer is given by the following lemma.

Lemma 3.8. *If f is a non-negative, continuous function that blows up in finite time with explosion time T^*, then $I_t^\alpha f$ blows up in finite time as well, with the same explosion time T^*.*
If f is a bounded continuous function, then $I_t^\alpha f$ does not blow up in finite time.

Proof. First, suppose that the non-negative, continuous function f explodes at T^* and let $M > 0$. Then we can find a small time $\varepsilon \in (0, T^*/2)$ such that $f(t) \geq M$ for all $t \in (T^* - \varepsilon, T^*)$. Hence,

$$I_t^\alpha f(t) \geq \int_{T^*-\varepsilon}^{t} k(t-s) f(s)\, ds \geq M \frac{(T^*-\varepsilon)^\alpha}{\Gamma(1+\alpha)} \geq M \frac{(T^*/2)^\alpha}{\Gamma(1+\alpha)}$$

for all $t \in (T^* - \varepsilon, T^*)$.
Now, suppose f is continuous and bounded with $M > 0$. Then we have

$$|I_t^\alpha f(t)| \leq \frac{M}{\Gamma(1+\alpha)} t^\alpha$$

for all $t \geq 0$. □

Let $T^*_{mathrmHe}$ be the moment explosion time in the classic Heston model (2.1)-(2.3) and T^* be the moment explosion time in the rough Heston model (3.1)-(3.3). Then we can summarise our findings in the following corollary.

Corollary 3.9. *For every $u \in \mathbb{R}$, the moment explosion time $T^*(u)$ blows up in finite time if and only if the moment explosion time $T^*_{\mathrm{He}}(u)$ blows up in finite time.*

3.3 Lower and Upper Bounds for the Moment Explosion Time

If we take a closer look at the cases (A) and (B) where the moment explosion time T^* is finite, we can establish lower and upper bounds for T^*.

Theorem 3.1. *In case (A) or (B), the blow-up time $T^*(u)$ of the solution $h(u, \cdot)$ of (3.17) satisfies*

$$T^*(u) \geq \Gamma(1+\alpha)^{1/\alpha} \max_{r>1} \frac{(r^\alpha - 1)^{1/\alpha}}{r(r-1)} \int_{a(u)}^{\infty} \left(\frac{w}{G(u,w)} \right)^{1/\alpha} \frac{dw}{w} \tag{3.22}$$

where $a(u) = 0$ in case (A) and $a(u) = -e_0(u) > 0$ in case (B).

Proof. The idea of the following proof is based on the proof of Lemma 2.19 in Brunner and Yang [14].
Fix u and suppress u from now on for ease of notation. It follows from Theorem 3.2 and 3.4 that in either case the solution h is non-negative, starts at 0 and $\lim_{t \to T^*-} h(t) = \infty$. For any $n \in \mathbb{N}_0$ choose

$$t_n := \min\{t > 0 : h(t) = (cr^n)^\alpha + a\}$$

with $r > 1$ and $c > 0$.
Using the inequality $k(t_n - s) < k(t_{n-1} - s)$ for $s \in (0, t_{n-1})$, the non-negativity of G and that G is strictly increasing on $[a, \infty)$, we have for $n \in \mathbb{N}$

$$\begin{aligned} h(t_n) &= \int_0^{t_{n-1}} k(t_n - s)G(h(s))\, ds + \int_{t_{n-1}}^{t_n} k(t_n - s)G(h(s))\, ds \\ &\leq h(t_{n-1}) + G(h(t_n)) \int_{t_{n-1}}^{t_n} k(t_n - s)\, ds \\ &= h(t_{n-1}) + \frac{1}{\Gamma(1+\alpha)} G(h(t_n))(t_n - t_{n-1})^{\alpha}. \end{aligned}$$

Thus, we obtain for $n \in \mathbb{N}$

$$\begin{aligned} t_n - t_{n-1} &\geq \Gamma(1+\alpha)^{1/\alpha} \left(\frac{h(t_n) - h(t_{n-1})}{G(h(t_n))} \right)^{1/\alpha} \\ &= \Gamma(1+\alpha)^{1/\alpha} (r^{\alpha} - 1)^{1/\alpha} \frac{c r^{n-1}}{G((cr^n)^{\alpha} + a)^{1/\alpha}} \\ &= \Gamma(1+\alpha)^{1/\alpha} \frac{(r^{\alpha} - 1)^{1/\alpha}}{r(r-1)} \cdot \frac{cr^{n+1} - cr^n}{G((cr^n)^{\alpha} + a)^{1/\alpha}} \\ &\geq \Gamma(1+\alpha)^{1/\alpha} \frac{(r^{\alpha} - 1)^{1/\alpha}}{r(r-1)} \int_{cr^n}^{cr^{n+1}} \left(\frac{1}{G(s^{\alpha} + a)} \right)^{1/\alpha} ds \end{aligned}$$

Finally,

$$\begin{aligned} T^* &= t_0 + \sum_{n=1}^{\infty} (t_n - t_{n-1}) \\ &\geq \Gamma(1+\alpha)^{1/\alpha} \frac{(r^{\alpha} - 1)^{1/\alpha}}{r(r-1)} \int_{cr}^{\infty} \left(\frac{1}{G(s^{\alpha} + a)} \right)^{1/\alpha} ds \end{aligned}$$

Maximization over $c > 0$, then $r > 1$, and substitution $w = s^{\alpha} + a$ yield the inequality (3.22). $\square$

For $\alpha \uparrow 1$, the right-hand side in (3.22) simplifies to

$$\int_{a(u)}^{\infty} \frac{1}{G(u,x)}\, dx = \int_{a(u)/c_3}^{\infty} \frac{1}{R(u,x)}\, dx.$$

In case (A), we have $a(u) = 0$ and therefore the right-hand side of (3.22) is exactly the moment explosion time (2.8) of the Heston model.

Theorem 3.2. *In case (A) or (B), the blow-up time $T^*(u)$ of the solution $h(u, \cdot)$ of (3.17) satisfies*

$$T^*(u) \leq 4\Gamma(1+\alpha)^{1/\alpha} \int_0^{\infty} \left(\frac{w}{\hat{G}(u,w)} \right)^{1/\alpha} \frac{dw}{w} \tag{3.23}$$

where $\hat{G} = G$ in case (A), and $\hat{G} \equiv -e_1$ on $[0, -e_0)$ and $\hat{G} = G$ on $[-e_0, \infty)$ in case (B).

Proof. The idea of the following proof is based on the proof of Lemma 2.12 in Brunner and Yang [14].
Fix u and suppress u from now on for ease of notation. From Theorem 3.2 and 3.4, in either case the solution h is positive on $(0, \infty)$, starts at 0 and $\lim_{t \to T^*-} h(t) = \infty$. For any $n \in \mathbb{N}_0$ choose

$$t_n := \max\{t < T^* : h(t) = (cr^n)^\alpha\} \tag{3.24}$$

with $r > 1$ and $c > 0$.
Define $\bar{G} := G$ in case (A) and $\bar{G} := \bar{G}_a$ from (3.20) in case (B). Since $\bar{G} \leq G$ and $\bar{G}$ is positive and strictly increasing, we have for $n \in \mathbb{N}$

$$\begin{aligned} h(t_n) &\geq \int_0^{t_n} k(t_n - s)\bar{G}(h(s))\, ds \\ &\geq \int_{t_{n-1}}^{t_n} k(t_n - s)\bar{G}(h(s))\, ds \\ &\geq \frac{1}{\Gamma(1+\alpha)} \bar{G}(h(t_{n-1}))(t_n - t_{n-1})^\alpha. \end{aligned}$$

Thus, we obtain for $n \in \mathbb{N}$

$$\begin{aligned} t_n - t_{n-1} &\leq \Gamma(1+\alpha)^{1/\alpha} \left(\frac{h(t_n)}{\bar{G}(h(t_{n-1}))} \right)^{1/\alpha} \\ &= \Gamma(1+\alpha)^{1/\alpha} \frac{cr^n}{\bar{G}((cr^{n-1})^\alpha)^{1/\alpha}} \\ &= \Gamma(1+\alpha)^{1/\alpha} \frac{r^2}{r-1} \cdot \frac{cr^{n-1} - cr^{n-2}}{\bar{G}((cr^{n-1})^\alpha)^{1/\alpha}} \\ &\leq \Gamma(1+\alpha)^{1/\alpha} \frac{r^2}{r-1} \int_{cr^{n-2}}^{cr^{n-1}} \left(\frac{1}{\bar{G}(s^\alpha)} \right)^{1/\alpha} ds \end{aligned}$$

Finally,

$$\begin{aligned} T^* &= t_0 + \sum_{n=1}^{\infty} (t_n - t_{n-1}) \\ &\leq t_0 + \Gamma(1+\alpha)^{1/\alpha} \frac{r^2}{r-1} \int_{cr^{-1}}^{\infty} \left(\frac{1}{\bar{G}(s^\alpha)} \right)^{1/\alpha} ds \end{aligned}$$

Note that from the definition of t_0, it depends on $c > 0$ and $r > 1$. The fact that h is only zero at $t = 0$ implies that $t_0 \to 0$ as $c \downarrow 0$. Taking the limit $c \downarrow 0$, then minimizing over $r > 1$ and substitution $w = s^\alpha$ yields

$$T^* \leq 4\Gamma(1+\alpha)^{1/\alpha} \int_0^\infty \left(\frac{w}{\bar{G}(w)} \right)^{1/\alpha} \frac{dw}{w}$$

In case (A), we are finished. In case (B), we have $\bar{G} = \bar{G}_a$. Then the dominated convergence theorem for $a \uparrow -e_1$ yields the inequality (3.23). $\square$

3.4 Critical Moments

With the upper bound for the moment explosion time T^* in Theorem 3.2, we can show the finiteness of the critical moments

$$u_+(T) := \sup\{u \in \mathbb{R} \colon \mathbb{E}[e^{uX_t}] < \infty\}$$
$$u_-(T) := \inf\{u \in \mathbb{R} \colon \mathbb{E}[e^{uX_t}] < \infty\}$$

for every $T > 0$.

Theorem 3.1. *In the rough Heston model (3.1)-(3.3) the critical moments $u_+(T)$ and $u_-(T)$ are finite for every $T > 0$.*

Proof. Only the finiteness of $u_+(T)$ is proven. The proof for $u_-(T)$ is similar. Denote the upper bound of $T^*(u)$ in (3.23) by $B(u)$ for all $u \in \mathbb{R}$ in the cases (A) and (B). At first, we show that for sufficiently large u, we are always in case (A) or (B), depending on the sign of the correlation parameter ρ. From (2.9) and (2.10), it is easy to see that

$$e_0(u) \sim \tfrac{1}{2}\xi\rho u \quad \text{and} \quad e_1(u) \sim -\tfrac{1}{4}\xi^2\bar{\rho}^2 u^2 \quad \text{as } u \to \infty, \tag{3.25}$$

with the correlation $\bar{\rho}^2 = 1 - \rho^2$. Thus, eventually $e_1(u) < 0$ for sufficiently large u. In the next step, we show that the upper bound $B(u)$ converges to 0 as $u \to \infty$. Indeed, in case (A) the integral in (3.23) satisfies

$$\begin{aligned}
\int_0^\infty \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} &= \int_0^u \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} + \int_u^\infty \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} \\
&\leq G(u,0)^{-1/\alpha} \int_0^u w^{-1+1/\alpha}\, dw + \int_u^\infty w^{-1-1/\alpha}\, dw \\
&\leq c u^{-1/\alpha}, \quad u \to \infty,
\end{aligned}$$

for some $c > 0$, using the monotonicity $G(u,w) \geq G(u,0)$, the inequality $G(u,w) \geq w^2$ and $G(u,0) = c_3 c_1(u) \sim \frac{1}{2}c_3 u^2$ as $u \to \infty$.
If we are eventually in case (B) as $u \to \infty$, then $-e_1(u) > 0$ and $-e_0(u) > 0$ holds for all sufficiently large u. Note, that in this case, $G(u,\cdot)$ attains its global minimum at $-e_0(u)$ and the minimum value is $-e_1(u)$. Thus, the integral in (3.23) satisfies

$$\begin{aligned}
\int_0^\infty \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} &= \int_0^{-2e_0(u)} \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} + \int_{-2e_0(u)}^\infty \left(\frac{w}{G(u,w)}\right)^{1/\alpha} \frac{dw}{w} \\
&\leq (-e_1(u))^{-1/\alpha} \int_0^{-2e_0(u)} w^{-1+1/\alpha}\, dw + 4^{1/\alpha} \int_{-2e_0(u)}^\infty w^{-1-1/\alpha}\, dw \\
&= \alpha \left((-e_1(u))^{-1/\alpha}(-2e_0(u))^{1/\alpha} + 4^{1/\alpha}(-2e_0(u))^{-1/\alpha} \right) \\
&\leq c u^{-1/\alpha}, \quad u \to \infty
\end{aligned}$$

for some $c > 0$, using the monotonicity $G(u,w) \geq -e_1(u)$, the inequality $G(u,w) \geq (w + e_0(u))^2 \geq w^2/4$ on $[-2e_0(u), \infty)$ and (3.25).
Altogether, we have $\lim_{u\to\infty} B(u) = 0$. From $0 \leq T^*(u) \leq B(u)$, the same is true for the moment explosion time T^*, i.e. $\lim_{u\to\infty} T^*(u) = 0$.
Let $T > 0$. Then there exists $u_0 \in \mathbb{R}$ such that $T^*(u) < T$ for all $u \geq u_0$. This inequality implies $\mathbb{E}[e^{uX_T}] = \infty$ for all $u \geq u_0$, and therefore $u_+(T) \leq u_0$. $\square$

Chapter 4

Conclusion

We investigated the blow-up behaviour of the moment-generating function (mgf) in the rough Heston model and we see that the mgf $\mathbb{E}[e^{uX_t}]$ in this model explodes in finite time for the same $u \in \mathbb{R}$ as in the classic Heston model.
Furthermore, we give a lower and upper bound for the explosion time T^* in the cases where the mgf blows up. Moreover, we show that the critical moments u_+ and u_- are always finite in the rough Heston model.

Part IV

Large-Strike Asymptotics in the 3/2-Model

Chapter 1

The 3/2-Model

The logarithmic stock price process $(X_t)_{t\geq 0}$ in the 3/2-model is given as the solution of the following SDE

$$\begin{aligned}
dX_t &= -\tfrac{1}{2}V_t\,dt + \sqrt{V_t}\,dW_t, && X_0 = x_0 \in \mathbb{R},\\
dV_t &= \kappa V_t(\theta - V_t)\,dt + \xi V_t^{3/2}\,dZ_t, && V_0 = v_0 > 0,\\
d\langle W, Z\rangle_t &= \rho\,dt,
\end{aligned}$$

with correlated Brownian motions W and Z and parameters $\kappa > 0$, $\theta > 0$, $\xi > 0$ and $|\rho| < 1$. Define $\bar{\rho} := \sqrt{1-\rho^2}$ and $\bar{\kappa} := 2\kappa + \xi^2$.
The moment-generating function (mgf) of X_T for $T > 0$ can be computed as

$$M(u,T) := \mathbb{E}[e^{uX_T}] = e^{ux_0}\frac{\Gamma(\mu_u - \alpha_u)}{\Gamma(\mu_u)} z_T^{\alpha_u}\, {}_1F_1(\alpha_u, \mu_u, -z_T), \tag{1.1}$$

at least for all $u \in \mathbb{C}$ in the vertical strip $a < \mathrm{Re}(u) < b$ with $a \leq 0$ and $b \geq 1$, and with the confluent hypergeometric function ${}_1F_1$ and the auxiliary functions

$$\begin{aligned}
\alpha_u &:= \frac{1}{\xi^2}(\gamma_u - \chi_u), & \gamma_u &:= \sqrt{\chi_u^2 - \xi^2 u(u-1)},\\
\mu_u &:= \frac{1}{\xi^2}(\xi^2 + 2\gamma_u), & \chi_u &:= \tfrac{1}{2}\bar{\kappa} - \rho\xi u,\\
z_T &:= \frac{2}{\xi^2\beta_T}, & \beta_T &:= \frac{v_0}{\kappa\theta}(e^{\kappa\theta T} - 1).
\end{aligned} \tag{1.2}$$

Without loss of generality, from now on, we assume $x_0 = 0$. Define the two real numbers

$$u_{\pm} := \frac{1}{2\xi\bar{\rho}^2}\left(\xi - \rho\bar{\kappa} \pm \sqrt{(\xi - \rho\bar{\kappa})^2 + \bar{\kappa}^2\bar{\rho}^2}\right), \tag{1.3}$$

which are the unique roots of the quadratic term under the square root of γ. After factorization of the polynomial, we have the following representation of γ

$$\gamma_u = \xi\bar{\rho}\sqrt{(u_+ - u)(u - u_-)}. \tag{1.4}$$

Throughout this part, we make the technical assumption

$$\mu_{u_+} - \alpha_{u_+} > 0$$

which is always satisfied if $\rho < 0$. Under this assumption, the right boundary b of the vertical strip, where equation (1.1) holds, can be extended until $b = u_+$. Note that the mgf has a branch cut along $[u_+, +\infty)$ due to the branch cut of (1.4).
For further information on the 3/2-model, see e.g. Lewis [68].

1.1 Large-Strike Asymptotics for the Density Function

We are interested in large-strike asymptotics of the density function $\varphi(k,T) := \varphi_{X_T}(k)$ of X_T for $T > 0$. With large-strike asymptotics we mean the asymptotic behaviour, if $k \to \infty$ with fixed $T > 0$. The density function φ can be expressed via Fourier-transform as

$$\varphi(k,T) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{-ku} M(u,T)\, du, \quad k \in \mathbb{R}, \tag{1.5}$$

with $a \in (u_-, u_+)$.
For the analysis, we adjust the integration path in (1.5) similar to Friz and Gerhold [43] and split it into two parts, the critical path $\mathcal{C}(k)$ and the neglectable path $\mathcal{N}(k)$, depending on the strike parameter $k \geq 1$. The critical contour $\mathcal{C}(k)$ embraces the critical moment u_+, see the left panel of Figure 1.1.

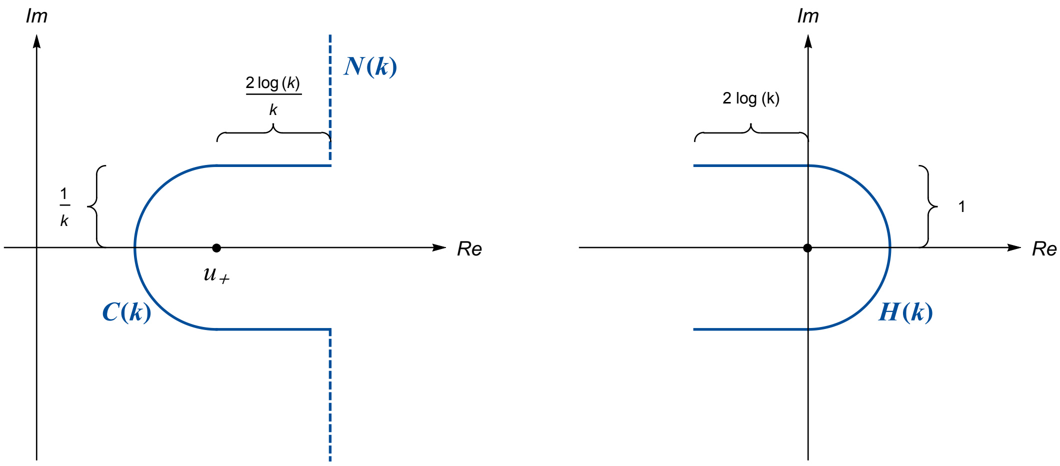

Figure 1.1: In the left panel, the critical path $\mathcal{C}(k)$ and the neglectable path $\mathcal{N}(k)$ (dashed line) are illustrated in the complex plane, whereas the right panel displays the transformed path $\mathcal{H}(k)$ after the transformation $w \mapsto u_+ - \frac{w}{k}$, where $k \geq 1$ is the strike parameter.

The critical path $\mathcal{C}(k)$ starts at $u_+ + 2\log(k)/k - i/k$, goes horizontally to $u_+ - i/k$, then clockwise along the half-circle with center u_+ and radius $1/k$ until it reaches $u_+ + i/k$, and again horizontally to the end point $u_+ + 2\log(k)/k + i/k$.
The remaining part, denoted by $\mathcal{N}(k)$, starts at the points $u_+ + 2\log(k)/k \pm i/k$ and goes straight to $u_+ + 2\log(k)/k + i\infty$ resp. $u_+ + 2\log(k)/k - i\infty$.

In our further analysis, we deal with the following representation of the density function

$$\varphi(k,T) = \frac{1}{2\pi i}\int_{\mathcal{C}(k)\cup\mathcal{N}(k)} e^{-ku} M(u,T)\,du, \quad k \in \mathbb{R}. \tag{1.6}$$

Theorem 1.1 (Large-strike asymptotics)**.** *Assume $\mu_{u_+} - \alpha_{u_+} > 0$. Then the first order term in the large-strike expansion of the density function of X_T in the 3/2-model, with $T > 0$ fixed, is given by*

$$\varphi(k,T) \sim c\frac{e^{-ku_+}}{k^{3/2}}, \quad k \to \infty, \tag{1.7}$$

where $c > 0$ can be computed explicitly.

Proof. At first, only the integral over the critical part $\mathcal{C}(k)$ in (1.6) is considered. We will show that this integral contains the whole asymptotic information of φ as $k \to \infty$ with $T > 0$ fixed. Change of variables $u = u_+ - w/k$ yields, as $k \to \infty$,

$$\frac{1}{2\pi i}\int_{\mathcal{C}(k)} e^{-ku} M(u,T)\,du = \frac{e^{-ku_+}}{k}\left(\frac{1}{2\pi i}\int_{\mathcal{H}(k)} e^{w} M\left(u_+ - \frac{w}{k}, T\right)\,dw\right) \tag{1.8}$$

where $\mathcal{H}(k)$ is the transformed path of $\mathcal{C}(k)$, see the right panel of Figure 1.1. With Lemma 1.3 the latter integral can be computed asymptotically, as $k \to \infty$,

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{\mathcal{H}(k)} e^{w} M\left(u_+ - \frac{w}{k}, T\right)\,dw \\
&\quad = M(u_+,T)\underbrace{\frac{1}{2\pi i}\int_{\mathcal{H}(k)} e^{w}\,dw}_{\mathcal{O}\left(\frac{1}{k^2}\right)} + \frac{m_1}{\sqrt{k}}\underbrace{\frac{1}{2\pi i}\int_{\mathcal{H}(k)} e^{w} w^{1/2}\,dw}_{\to 1/\Gamma\left(-\frac{1}{2}\right) = -\frac{1}{2\sqrt{\pi}}} + \underbrace{\frac{1}{2\pi i}\int_{\mathcal{H}(k)} e^{w}\mathcal{O}\left(\frac{w}{k}\right)\,dw}_{\mathcal{O}\left(\frac{1}{k}\right)}.
\end{aligned}$$

The first integral is an easy computation. In the second and third integral, we used Hankel's integral representation for the gamma function, see [75]. Therefore,

$$\frac{1}{2\pi i}\int_{\mathcal{C}(k)} e^{-ku} M(u,T)\,du \sim c\frac{e^{-ku_+}}{k^{3/2}}, \quad k \to \infty,$$

for $c = \frac{-m_1}{2\sqrt{\pi}}$. Combined with Lemma 1.2 we have the desired expansion of the density function φ. ☐

Lemma 1.2. *The integral over $\mathcal{N}(k)$ in (1.6) satisfies*

$$\frac{1}{2\pi i}\int_{\mathcal{N}(k)} e^{-ku} M(u,T)\,du = o\left(e^{-ku_+}k^{-3/2}\right), \quad k \to \infty,$$

Proof. By symmetry, it suffices to consider only the integral over the upper part of $\mathcal{N}(k)$. We choose the representing path $u_k(t) := u_+ + 2\log k/k + it$ with $t \in [1/k, \infty)$,

$$\frac{1}{2\pi i}\int_{u_k} e^{-ku} M(u,T)\,du = \frac{e^{-ku_+}}{k^2}\left(\frac{1}{2\pi}\int_{1/k}^{\infty} e^{-itk} M(u_k(t),T)\,dt\right).$$

By showing the boundedness of the latter integral, the proof is finished.
After the triangular inequality for integrals, we split the integral into two parts

$$\left| \int_{1/k}^{\infty} e^{-itk} M(u_k(t),T)\, dt \right| \le \int_{1/k}^{t_1} |M(u_k(t),T)|\, dt + \int_{t_1}^{\infty} |M(u_k(t),T)|\, dt, \qquad (1.9)$$

where $t_1 \ge 1$ will be determined later.
For the first integral in (1.9), note that $2\log k/k \in [0,1]$ for any $k \ge 1$. Recall that $M(\cdot,T)$ has a branch cut along $[u_+,\infty)$, but a continuous extension $\tilde{M}$ of M exists on the half-plane $\mathrm{Im}(s) \ge 0$. Hence $|M(\cdot,T)|$ attains a maximum value on $[u_+, u_+ + 1] + i(0,t_1]$,

$$\int_{1/k}^{t_1} |M(u_k(t),T)|\, dt \le t_1 \max_{u\in[u_+,u_++1]+i(0,t_1]} |M(u,T)| < \infty$$

In order to show the boundedness of the second integral in (1.9) and to determine $t_1 \ge 1$, we have to take a closer look at the mgf and the auxiliary functions defined in (1.2). The fact $2\log k/k \in [0,1]$ for $k \ge 1$ ensures $u_k(t) = it + \mathcal{O}(1)$ for $t \to \infty$ uniformly for all $k \ge 0$. Thus, the following asymptotic expansions of the auxiliary functions χ and γ in (1.2) hold

$$\begin{aligned}
\chi(u_k(t)) &= -i\xi\rho t + \mathcal{O}(1),\\
\gamma(u_k(t)) &= \sqrt{-\xi^2\rho^2t^2 + \xi^2t^2 + \mathcal{O}(t)} = \xi\bar{\rho}t + \mathcal{O}(1),
\end{aligned}$$

and simple computations then yield

$$\begin{aligned}
\alpha(u_k(t)) &= \tfrac{1}{\xi}(\bar{\rho} + i\rho)t + \mathcal{O}(1), && (1.10)\\
\mu(u_k(t)) &= \tfrac{2}{\xi}\bar{\rho}t + \mathcal{O}(1), &&\\
\mu(u_k(t)) - \alpha(u_k(t)) &= \tfrac{1}{\xi}(\bar{\rho} - i\rho)t + \mathcal{O}(1), && (1.11)
\end{aligned}$$

for $t \to \infty$ uniformly for all $k \ge 1$.
Due to (1.10), (1.11) and $\bar{\rho} > 0$, there exists $t_0 \ge 1$, such that $\mathrm{Re}(\mu(u_k(t)) - \alpha(u_k(t))) > 1$ and $\mathrm{Re}(\alpha(u_k(t))) > 1$ for all $k \ge 1$ and $t \ge t_0$. In particular, in this region we have $\mathrm{Re}(\mu(u_k(t))) > \mathrm{Re}(\alpha(u_k(t))) > 0$ so that we can use representation (1.25) of the confluent hypergeometric function, which reduces the mgf to

$$M(u_k(t),T) = \frac{z_T^{\alpha(u_k(t))}}{\Gamma(\alpha(u_k(t)))} \int_0^1 e^{-z_T y} y^{\alpha(u_k(t))-1}(1-y)^{\mu(u_k(t))-\alpha(u_k(t))-1}\, dy. \qquad (1.12)$$

Note that the absolute value of the integral is bounded by 1. Furthermore, we have uniformly for all $k \ge 1$

$$|z_T^{\alpha(u_k(t))}| = \exp\left(\tfrac{1}{\xi}\bar{\rho}\log(z_T)t(1+o(1))\right), \quad t \to \infty. \qquad (1.13)$$

Our choice $\mathrm{Re}(\alpha(u_k(t))) > 1$ guarantees $|\arg(\alpha(u_k(t)))| < \frac{\pi}{2}$ and Stirling's formula (1.26) is applicable to $\Gamma(\alpha(u_k(t)))$ for all $t \ge t_0$ and all $k \ge 1$. Combined with (1.10) we have, uniformly for all $k \geq 1$,

$$\begin{aligned}|\Gamma(\alpha(u_k(t)))| &\sim \sqrt{2\pi}|e^{-z}z^z z^{-1/2}|_{z=\frac{1}{\xi}(\bar\rho+i\rho)t} \\ &= \sqrt{2\pi\xi}\ x^{-1/2}\ \exp\left(\tfrac{1}{\xi}\bar\rho t\log(\tfrac{t}{\xi}) - \tfrac{1}{\xi}\rho\arg(\bar\rho+i\rho)t - \tfrac{1}{\xi}\bar\rho t\right) \\ &= \exp\left(\tfrac{1}{\xi}\bar\rho t\log t(1+o(1))\right), \quad t\to\infty. \end{aligned} \tag{1.14}$$

Putting (1.13) and (1.14) back into formula (1.12), we can find a sufficiently large $t_1 \geq t_0$ such that

$$|M(u_k(t),T)| \leq \exp\left(-(1+\varepsilon)\tfrac{1}{\xi}\bar\rho t\log t\right) \tag{1.15}$$

for all $t \geq t_1$ and all $k \geq 1$, with a constant $\varepsilon > 0$. The integrability of the right-hand side of (1.15) proves that the third integral in (1.9) is bounded. □

Lemma 1.3. *Assume $\mu_{u_+} - \alpha_{u_+} > 0$. Near the critical moment u_+, the following expansion of the mgf holds uniformly for all $w \in \mathcal{H}(k)$,*

$$M\left(u_+ - \frac{w}{k}, T\right) = M(u_+,T) + m_1\sqrt{\frac{w}{k}} + \mathcal{O}\left(\frac{w}{k}\right), \quad k\to\infty,$$

where m_1 can be computed explicitly.

Proof. First, we expand the functions χ and γ in a neighbourhood of u_+. Using representation (1.4) of γ we only have to expand $\sqrt{u-u_-} = \sqrt{(u_+-u_-)-(u_+-u)}$ near u_+. Thus, as $u \to t_+$,

$$\gamma_u = \xi\bar\rho\sqrt{u_+-u_-}(u_+-u)^{1/2} + \mathcal{O}\left((u_+-u)^{3/2}\right), \tag{1.16}$$

$$\chi_u = \chi_{u_+} + \rho\xi(u_+-u). \tag{1.17}$$

With these results, expansions for α and μ near u_+ can easily be computed,

$$\alpha_u = \alpha_{u_+} + \frac{\bar\rho}{\xi}\sqrt{u_+-u_-}(u_+-u)^{1/2} + \mathcal{O}(u_+-u), \quad u\to u_+ \tag{1.18}$$

$$\mu_u = \mu_{u_+} + 2\frac{\bar\rho}{\xi}\sqrt{u_+-u_-}(u_+-u)^{1/2} + \mathcal{O}\left((u_+-u)^{3/2}\right), \quad u\to u_+. \tag{1.19}$$

Define $u_k(w) := u_+ - \frac{w}{k}$, $w \in \mathcal{H}(k)$, for $k \geq 1$. From the uniform convergence $\sup_{w\in\mathcal{H}(k)}|u_k(w)-u_+| \to 0$ for $k\to\infty$, we have

$$\Delta\alpha := \alpha(u_k(w)) - \alpha_{u_+} = \frac{\bar\rho}{\xi}\sqrt{u_+-u_-}\left(\frac{w}{k}\right)^{1/2} + \mathcal{O}\left(\frac{w}{k}\right), \quad k\to\infty \tag{1.20}$$

$$\Delta\mu := \mu(u_k(w)) - \mu_{u_+} = 2\frac{\bar\rho}{\xi}\sqrt{u_+-u_-}\left(\frac{w}{k}\right)^{1/2} + \mathcal{O}\left(\left(\frac{w}{k}\right)^{3/2}\right), \quad k\to\infty, \tag{1.21}$$

uniformly for all $w \in \mathcal{H}(k)$. Define the function

$$\tilde M(\alpha,\mu) := \frac{\Gamma(\mu-\alpha)}{\Gamma(\mu)}(z_T)^{\alpha}{}_1F_1(\alpha,\mu,-z_T),$$

for all $(\alpha,\mu) \in \mathbb{C}^2$ where $\mu-\alpha,\mu \notin \mathbb{Z}_0^-$. In this region $\tilde{M}$ is jointly analytic in both variables. Note the relation $M(u,T) = \tilde{M}(\alpha_u,\mu_u)$. Since $\mu_{u_+} = 1$ and $\mu_{u_+} - \alpha_{u_+} > 0$, we can make a Taylor expansion of $\tilde{M}$ at the point (α_{u_+},μ_{u_+}). Combining this with (1.20) and (1.21) gives us, uniformly for all $w \in \mathcal{H}(k)$,

$$
\begin{aligned}
&M(u_k(w),T) = \tilde{M}(\alpha(u_k(w)),\mu(u_k(w)))\\
&= \tilde{M}(\alpha_{u_+},\mu_{u_+}) + \Delta\alpha\frac{\partial}{\partial\alpha}\tilde{M}(\alpha_{u_+},\mu_{u_+}) + \Delta\mu\frac{\partial}{\partial\mu}\tilde{M}(\alpha_{u_+},\mu_{u_+}) + \mathcal{O}\left((\Delta\alpha)^2\right) + \mathcal{O}\left((\Delta\mu)^2\right)\\
&= M(u_+,T) + \underbrace{\left(\frac{\partial\tilde{M}}{\partial\alpha} + 2\frac{\partial\tilde{M}}{\partial\mu}\right)(\alpha_{u_+},\mu_{u_+})\frac{\bar{\rho}}{\xi}\sqrt{u_+-u_-}}_{=:m_1}\left(\frac{w}{k}\right)^{1/2} + \mathcal{O}\left(\frac{w}{k}\right), \quad k\to\infty.
\end{aligned}
$$

□

1.2 Large-Strike Asymptotics for the Implied Volatility

From large-strike asymptotics for the density function, it is possible to obtain large strike asymptotics for the implied volatility, see Gulisashvili [57] and Friz, Gerhold, Gulisashvili and Sturm [44]. The statement is, that if the density function φ satisfies, for fixed $T>0$,

$$c_1 k^{-\xi} h(k) \le \varphi(k) \le c_2 k^{-\xi} h(k),$$

for all sufficiently large k, with $\xi > 2$, h slowly varying and constants $c_1, c_2 > 0$, then the implied volatility $\sigma_{\text{imp}}(K,T)$ satisfies

$$
\begin{aligned}
\sigma_{\text{imp}}(K,T)\frac{\sqrt{T}}{\sqrt{2}} =& \sqrt{\log K + \log\frac{1}{K^{2-\xi}h(K)} + \tfrac{1}{2}\log\log\frac{1}{K^{2-\xi}h(K)}} \\
&- \sqrt{\log\frac{1}{K^{2-\xi}h(K)} + \tfrac{1}{2}\log\log\frac{1}{K^{2-\xi}h(K)}}\\
&+ \mathcal{O}\big((\log K)^{-1}g(K)\big),
\end{aligned}
\tag{1.22}
$$

as $K\to\infty$, for every positive function g on $(0,\infty)$ and $\lim_{x\to\infty} g(x) = \infty$.

In the 3/2-model with Theorem 1.1, we have established the large-strike asymptotics for the density φ_{X_T} of the log-price $X_T = \log(S_T)$. Because the density φ_{S_T} of S_T is given by

$$\varphi_{S_T}(K) = \frac{\varphi_{X_T}(\log K)}{K}, \quad K>0,$$

we clearly have the large-strike asymptotics for φ_{S_T}

$$\varphi_{S_T}(K) \sim cK^{-(u_++1)}h(K), \quad K\to\infty, \tag{1.23}$$

with the slowly varying function $h(K) = (\log K)^{-3/2}$. Note that the critical moment u always satisfies $u_+ \ge 1$, and $u_+ = 1$ if and only if $2\xi\rho = \bar{\kappa}$. Hence, the previous statement is applicable.

Theorem 1.1. *Assume $\mu_{u_+} - \alpha_{u_+} > 0$ and $2\xi\rho \neq \bar{\kappa}$. The large-strike expansion of the implied volatility function in the 3/2-model, with $T > 0$ fixed, is given by, as $K \to \infty$,*

$$\begin{aligned}\sigma_{\text{imp}}(K,T)\frac{\sqrt{T}}{\sqrt{2}} &\sim (\sqrt{u_+} - \sqrt{u_+ - 1})\sqrt{\log K} \\ &\quad + \frac{1}{2}\left(\frac{1}{\sqrt{u_+}} - \frac{1}{\sqrt{u_+ - 1}}\right)\frac{\log\log K}{\sqrt{\log K}} + \mathcal{O}\left(\frac{\log\log\log K}{\sqrt{\log K}}\right).\end{aligned} \tag{1.24}$$

Proof. Note that if $f(K) = \mathcal{O}(g(K))$, $K \to \infty$, holds for every positive function g that tends to ∞ for $K \to \infty$, then f has to be bounded, $f(K) = \mathcal{O}(1)$ as $K \to \infty$.
A straightforward calculation, using (1.22) and (1.23), shows

$$\begin{aligned}\sigma_{\text{imp}}(K,T)\frac{\sqrt{T}}{\sqrt{2}} &= \sqrt{\log K - \log(K^{1-u_+}h(K)) - \tfrac{1}{2}\log\big(-\log(K^{1-u_+}h(K))\big)} \\ &\quad - \sqrt{-\log(K^{1-u_+}h(K)) - \tfrac{1}{2}\log\big(-\log(K^{1-u_+}h(K))\big)} \\ &\quad + \mathcal{O}\big((\log K)^{-1}\big) \\ &= \sqrt{u_+\log K + \log\log K + \mathcal{O}(\log\log\log K))} \\ &\quad - \sqrt{(u_+ - 1)\log K + \log\log K + \mathcal{O}(\log\log\log K)} \\ &\quad + \mathcal{O}\big((\log K)^{-1}\big), \\ &= \sqrt{u_+}\sqrt{\log K}\left(1 + \frac{\log\log K}{u_+\log K} + \mathcal{O}\left(\frac{\log\log\log K)}{\log K}\right)\right) \\ &\quad - \sqrt{(u_+ - 1)}\sqrt{\log K}\left(1 + \frac{\log\log K}{(u_+ - 1)\log K} + \mathcal{O}\left(\frac{\log\log\log K}{\log K}\right)\right) \\ &\quad + \mathcal{O}\big((\log K)^{-1}\big),\end{aligned}$$

as $K \to \infty$. Summing up the matching terms yields the statement. $\square$

Appendix

Appendix A Auxiliary Results

We state the following lemmas which are used in the proof of the large strike expansion of the density function. The first lemma describes a representation of the confluent hypergeometric function ${}_1F_1$, whereas the second lemma is the well-known Stirling formula for the Gamma function. For further details, see e.g. [75].

Lemma A.1. *If* $\mathrm{Re}(\mu) > \mathrm{Re}(\alpha) > 0$*, then the confluent hypergeometric function* ${}_1F_1$ *has the following integral representation*

$$ {}_1F_1(\alpha, \mu, z) = \frac{\Gamma(\mu)}{\Gamma(\alpha)\Gamma(\mu-\alpha)} \int_0^1 e^{zy} y^{\alpha-1} (1-y)^{\mu-\alpha-1}\, dy. \tag{1.25} $$

Lemma A.2 (Stirling)**.** *The Gamma function satisfies*

$$ \Gamma(z) = \sqrt{2\pi} e^{-z} z^z z^{-1/2} (1 + o(1)), \quad z \to \infty \quad \text{with } |\arg(z)| < \pi - \varepsilon. \tag{1.26} $$

where $\varepsilon > 0$

Bibliography

[1] Y. Aït-Sahalia, *Telling from discrete data whether the underlying continuous-time model is a diffusion*, Journal of Finance, 57 (2002), pp. 2075–2113.

[2] Y. Aït-Sahalia and J. Jacod, *Is Brownian motion necessary to model high-frequency data?*, Ann. Statist., 38 (2010), pp. 3093–3128.

[3] E. Alòs, J. A. León, and J. Vives, *On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility*, Finance Stoch., 11 (2007), pp. 571–589.

[4] L. Andersen and A. Lipton, *Asymptotics for exponential Lévy processes and their volatility smile: survey and new results*, Int. J. Theor. Appl. Finance, 16 (2013). Paper no. 1350001, 98 pages.

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

[6] C. Bayer, P. Friz, and J. Gatheral, *Pricing under rough volatility*, Quantitative Finance, 16 (2016), pp. 887–904.

[7] S. Benaim and P. Friz, *Smile asymptotics II: Models with known moment generating functions*, J. Appl. Probab., 45 (2008), pp. 16–32.

[8] A. Bentata and R. Cont, *Short-time asymptotics for marginal distributions of semimartingales*. Preprint, available at http://arxiv.org/abs/1202.1302, 2012.

[9] H. Berestycki, J. Busca, and I. Florent, *Asymptotics and calibration of local volatility models*, Quant. Finance, 2 (2002), pp. 61–69. Special issue on volatility modelling.

[10] H. Berestycki, J. Busca, and I. Florent, *Computing the implied volatility in stochastic volatility models*, Comm. Pure Appl. Math., 57 (2004), pp. 1352–1373.

[11] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, vol. 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1987.

[12] S. I. Boyarchenko and S. Z. Levendorskiĭ, *Non-Gaussian Merton-Black-Scholes theory*, vol. 9 of Advanced Series on Statistical Science & Applied Probability, World Scientific Publishing Co., Inc., River Edge, NJ, 2002.

[13] G. Brunick and S. Shreve, *Mimicking an Itô process by a solution of a stochastic differential equation*, Ann. Appl. Probab., 23 (2013), pp. 1584–1628.

[14] H. Brunner and Z. Yang, *Blow-up behavior of hammerstein-type volterra integral equations*, J. Integral Equations Applications, 24 (2012), pp. 487–512.

[15] Y. Bu, *Option pricing using Lévy processes*, master's thesis, Chalmers University of Technology, Göteborg, 2007.

[16] Y. Cai and S. Wang, *Central limit theorem and moderate deviation principle for CKLS model with small random perturbation*, Statist. Probab. Lett., 98 (2015), pp. 6–11.

[17] F. Caravenna and J. Corbetta, *General smile asymptotics with bounded maturity*. Preprint, available at http://arxiv.org/abs/1411.1624, 2014.

[18] P. Carr and L. Wu, *What type of process underlies options? A simple robust test*, The Journal of Finance, 58 (2003), pp. 2581–2610.

[19] R. Cont and P. Tankov, *Financial modelling with jump processes*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[20] Y. A. Davydov and I. A. Ibragimov, *On asymptotic behavior of some functionals of processes with independent increments*, Theory Probab. Appl., 16 (1971), pp. 162–167.

[21] L. De Leo, V. Vargas, S. Ciliberti, and J.-P. Bouchaud, *We've walked a million miles for one of these smiles*. Preprint, available at http://arxiv.org/abs/1203.5703, 2012.

[22] S. De Marco, P. Friz, and S. Gerhold, *Rational shapes of local volatility*, Risk, 2 (2013), pp. 82–87.

[23] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, vol. 38 of Stochastic Modelling and Applied Probability, Springer-Verlag, New York, second ed., 1998.

[24] J. D. Deuschel, P. K. Friz, A. Jacquier, and S. Violante, *Marginal density expansions for diffusions and stochastic volatility I: Theoretical foundations*, Comm. Pure Appl. Math., 67 (2014), pp. 40–82.

[25] J. D. Deuschel, P. K. Friz, A. Jacquier, and S. Violante, *Marginal density expansions for diffusions and stochastic volatility II: applications*, Comm. Pure Appl. Math., 67 (2014), pp. 321–350.

[26] R. A. Doney and R. A. Maller, *Stability and attraction to normality for Lévy processes at zero and at infinity*, J. Theoret. Probab., 15 (2002), pp. 751–792.

[27] V. Durrleman, *From implied to spot volatilities*, PhD thesis, Princeton University, 2004.

[28] V. Durrleman, *From implied to spot volatilities*, Finance Stoch., 14 (2010), pp. 157–177.

[29] O. El Euch and M. Rosenbaum, *The characteristic function of rough Heston models*, ArXiv e-prints, (2016).

[30] J. Fajardo, *Barrier style contracts under Lévy processes: an alternative approach.* Preprint, available on SSRN, 2014.

[31] J. Fajardo and E. Mordecki, *Symmetry and duality in Lévy markets*, Quant. Finance, 6 (2006), pp. 219–227.

[32] J. E. Figueroa-López and M. Forde, *The small-maturity smile for exponential Lévy models*, SIAM J. Financial Math., 3 (2012), pp. 33–65.

[33] J. E. Figueroa-López, R. Gong, and C. Houdré, *Small-time expansions of the distributions, densities, and option prices of stochastic volatility models with Lévy jumps*, Stochastic Process. Appl., 122 (2012), pp. 1808–1839.

[34] J. E. Figueroa-López, R. Gong, and C. Houdré, *High-order short-time expansions for ATM option prices of exponential Lévy models.* To appear in Mathematical Finance, 2014.

[35] J. E. Figueroa-López and C. Houdré, *Small-time expansions for the transition distributions of Lévy processes*, Stochastic Process. Appl., 119 (2009), pp. 3862–3889.

[36] J. E. Figueroa-López and S. Ólafsson, *Short-time asymptotics for the implied volatility skew under a stochastic volatility model with Lévy jumps.* Preprint, available at http://arxiv.org/abs/1502.02595, 2015.

[37] P. Flajolet, X. Gourdon, and P. Dumas, *Mellin transforms and asymptotics: harmonic sums*, Theoret. Comput. Sci., 144 (1995), pp. 3–58. Special volume on mathematical analysis of algorithms.

[38] M. Forde and A. Jacquier, *Small-time asymptotics for implied volatility under the Heston model*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 861–876.

[39] M. Forde and A. Jacquier, *Small-time asymptotics for an uncorrelated local-stochastic volatility model*, Appl. Math. Finance, 18 (2011), pp. 517–535.

[40] M. Forde, A. Jacquier, and J. E. Figueroa-López, *The large-time smile and skew for exponential Lévy models.* Preprint, 2011.

[41] M. Forde, A. Jacquier, and R. Lee, *The small-time smile and term structure of implied volatility under the Heston model*, SIAM J. Financial Math., 3 (2012), pp. 690–708.

[42] M. Forde and H. Zhang, *Asymptotics for rough stochastic volatility models.* Preprint, 2016.

[43] P. Friz and S. Gerhold, *Extrapolation analytics for Dupire's local volatility*, in Large Deviations and Asymptotic Methods in Finance, vol. 110 of Springer Proc. Math. Stat., Springer, Cham, 2015, pp. 273–286.

[44] 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.

[45] P. Friz, S. Gerhold, and A. Pinter, *Option Pricing in the Moderate Deviations Regime*, ArXiv e-prints, (2016).

[46] P. K. Friz, S. Gerhold, and M. Yor, *How to make Dupire's local volatility work with jumps*, Quant. Finance, 14 (2014), pp. 1327–1331.

[47] K.-A. Fu and X. Shen, *Moderate deviations for sums of dependent claims in a size-dependent renewal risk model*, Comm. Statist. Theory Methods, 46 (2017), pp. 3235–3243.

[48] F. Gao and S. Wang, *Asymptotic behaviors for functionals of random dynamical systems*, Stoch. Anal. Appl., 34 (2016), pp. 258–277.

[49] K. Gao and R. Lee, *Asymptotics of implied volatility to arbitrary order*, Finance Stoch., 18 (2014), pp. 349–392.

[50] J. Gatheral, *The Volatility Surface, A Practitioner's Guide*, Wiley, 2006.

[51] J. Gatheral, E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang, *Asymptotics of implied volatility in local volatility models*, Math. Finance, 22 (2012), pp. 591–620.

[52] S. Gerhold, *Can there be an explicit formula for implied volatility?*, Appl. Math. E-Notes, 13 (2013), pp. 17–24.

[53] S. Gerhold, I. C. Gülüm, and A. Pinter, *Small-maturity asymptotics for the at-the-money implied volatility slope in lévy models*, Applied Mathematical Finance, 23 (2016), pp. 135–157.

[54] S. Gerhold, M. Kleinert, P. Porkert, and M. Shkolnikov, *Small time central limit theorems for semimartingales with applications*, Stochastics, 87 (2015), pp. 723–746.

[55] H. Guennoun, A. Jacquier, and P. Roome, *Asymptotic behaviour of the fractional Heston model*. Preprint, available at http://arxiv.org/abs/1411.7653, 2014.

[56] A. Guillin, *Averaging principle of SDE with small diffusion: moderate deviations*, Ann. Probab., 31 (2003), pp. 413–443.

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

[58] A. Gulisashvili and J. Teichmann, *The Gärtner-Ellis theorem, homogenization, and affine processes*, in Large Deviations and Asymptotic Methods in Finance, vol. 110 of Springer Proc. Math. Stat., Springer, Cham, 2015, pp. 287–320.

[59] P. Henry-Labordère, *Analysis, geometry, and modeling in finance*, Chapman & Hall/CRC Financial Mathematics Series, CRC Press, Boca Raton, FL, 2009.

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

[61] T. Jaisson and M. Rosenbaum, *Limit theorems for nearly unstable hawkes processes*, Ann. Appl. Probab., 25 (2015), pp. 600–631.

[62] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.

[63] M. Keller-Ressel, *Moment explosions and long-term behavior of affine stochastic volatility models*, Math. Finance, 21 (2011), pp. 73–98.

[64] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations, Volume 204 (North-Holland Mathematics Studies)*, Elsevier Science Inc., New York, NY, USA, 2006.

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

[66] R. W. Lee, *Option pricing by transform methods: Extensions, unification, and error control*, Journal of Computational Finance, 7 (2004), pp. 51–86.

[67] R. W. Lee, *Implied volatility: statics, dynamics, and probabilistic interpretation*, in Recent advances in applied probability, Springer, New York, 2005, pp. 241–268.

[68] A. L. Lewis, *Option valuation under stochastic volatility*, Finance Press, Newport Beach, CA, 2000.

[69] D. Madan, P. Carr, and E. Chang, *The variance gamma process and option pricing*, European Finance Review, 2 (1998), pp. 79–105.

[70] A. Mijatović and P. Tankov, *A new look at short-term implied volatility in asset price models with jumps*. Preprint, available at http://arxiv.org/abs/1207.0843, 2012.

[71] J. Muhle-Karbe and M. Nutz, *Small-time asymptotics of option prices and first absolute moments*, Journal of Applied Probability, 48 (2011), pp. 1003–1020.

[72] W. Mydlarczyk, *The blow-up solutions of integral equations*, Colloquium Mathematicae, 79 (1999), pp. 147–156.

[73] W. Mydlarczyk and W. Okrasiński, *Nonlinear volterra integral equations with convolution kernels*, Bulletin of the London Mathematical Society, 35 (2003), pp. 484–490.

[74] F. W. J. Olver, *Asymptotics and special functions*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1974.

[75] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds., *NIST handbook of mathematical functions*, U.S. Department of Commerce National Institute of Standards and Technology, Washington, DC, 2010.

[76] E. E. Omar, F. Masaaki, and R. Mathieu, *The microstructural foundations of leverage effect and rough volatility*, papers, arXiv.org, 2016.

[77] Y. Osajima, *General asymptotics of Wiener functionals and application to implied volatilities*, in Large Deviations and Asymptotic Methods in Finance, vol. 110 of Springer Proc. Math. Stat., Springer, Cham, 2015, pp. 137–173.

[78] S. Pagliarani and A. Pascucci, *The parabolic Taylor formula of the implied volatility*. Preprint, available at http://arxiv.org/abs/1510.06084, 2016.

[79] L. Paulot, *Asymptotic implied volatility at the second order with application to the SABR model*, in Large Deviations and Asymptotic Methods in Finance, vol. 110 of Springer Proc. Math. Stat., Springer, Cham, 2015, pp. 37–69.

[80] H. Pham, *Large deviations in finance*. Lecture notes for the third SMAI European Summer School in Financial Mathematics, Paris, August 2010.

[81] M. Roper, *Implied volatility explosions: European calls and implied volatilities close to expiry in exponential Lévy models*. Preprint, available at http://arxiv.org/abs/0809.3305, 2008.

[82] M. Roper and M. Rutkowski, *On the relationship between the call price surface and the implied volatility surface close to expiry*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 427–441.

[83] M. Rosenbaum and P. Tankov, *Asymptotic results for time-changed Lévy processes sampled at hitting times*, Stochastic Process. Appl., 121 (2011), pp. 1607–1632.

[84] S. Samko, A. Kilbas, and O. Marichev, *Fractional Integrals and Derivatives*, Taylor & Francis, 1993.

[85] K.-i. Sato, *Lévy processes and infinitely divisible distributions*, vol. 68 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1999.

[86] W. Schoutens, *Meixner processes: Theory and applications in finance*, EURANDOM Report 2002-004, EURANDOM, Eindhoven, 2002.

[87] P. Tankov, *Pricing and hedging in exponential Lévy models: review of recent results*, in Paris-Princeton Lectures on Mathematical Finance 2010, vol. 2003 of Lecture Notes in Math., Springer, Berlin, 2011, pp. 319–359.

[88] M. R. Tehranchi, *Uniform bounds for Black-Scholes implied volatility*. Preprint, available at http://arxiv.org/abs/1512.06812, 2016.

[89] S. R. S. Varadhan, *Diffusion processes in a small time interval*, Comm. Pure Appl. Math., 20 (1967), pp. 659–685.

[90] S. Yan, *Jump risk, stock returns, and slope of implied volatility smile*, Journal of Financial Economics, 99 (2011), pp. 216–233.

[91] K. Yosida, *On the fundamental solution of the parabolic equation in a Riemannian space*, Osaka Math. J., 5 (1953), pp. 65–74.