ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

LEVERAGED ETF IMPLIED VOLATILITIES FROM ETF DYNAMICS

This is the submitted version (pre peer-review, preprint) of the following publication:

Published Version:
Leung, T., Lorig, M., Pascucci, A. (2017). LEVERAGED ETF IMPLIED VOLATILITIES FROM ETF DYNAMICS. MATHEMATICAL FINANCE, 27(4), 1035-1068 [10.1111/mafi.12128].

Availability:
This version is available at: https://hdl.handle.net/11585/613992 since: 2017-12-19

Published:
DOI: http://doi.org/10.1111/mafi.12128

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

16 September 2026

This is the submitted version (pre-print) of:

Leung, T., Lorig, M. and Pascucci, A. (2017), LEVERAGED ETF IMPLIED VOLATILITIES FROM ETF DYNAMICS. Mathematical Finance, 27: 1035-1068.

The final published version is available online at : **https://doi.org/10.1111/mafi.12128**

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

Leveraged ETF implied volatilities from ETF dynamics

Tim Leung * Matthew Lorig † Andrea Pascucci ‡

This version: April 27, 2014

Abstract

The growth of the exhange-traded fund (ETF) industry has given rise to the trading of options written on ETFs and their leveraged counterparts (LETFs). We study the relationship between the ETF and LETF implied volatility surfaces when the underlying ETF is modeled by a general class of local-stochastic volatility models. A closed-form approximation for prices is derived for European-style options whose payoff depends on the terminal value of the ETF and/or LETF. Rigorous error bounds for this pricing approximation are established. A closed-form approximation for implied volatilities is also derived. We also discuss a scaling procedure for comparing implied volatilities across leverage ratios. The implied volatility expansions and scalings are tested in three well-known settings: CEV, Heston and SABR.

Keywords: implied volatility, local-stochastic volatility, leveraged exchange-traded fund, implied volatility scaling

1 Introduction

The market of exchange-traded funds (ETFs) has been growing at a robust pace since their introduction in 1993[1]. As of the end of 2012, the global ETF industry has over $1.8 trillion in assets under management (AUM) comprised of 4,272 products, and has seen close to $200 billion of positive capital inflows[2]. In recent years, a sub-class of ETFs, called leveraged ETFs (LETFs), has gained popularity among investors for their accessibility and liquidity for leveraged positions. These funds are designed to replicate multiples of the daily returns of some reference index or asset. For instance, the ProShares S&P 500 Ultra (SSO) and UltraPro (UPRO) are advertised to generate, respectively, 2 and 3 times of the daily returns of the S&P 500 index, minus a small expense fee. On the other hand, an LETF with a negative leverage ratio allows investors to

*Industrial Engineering & Operations Research Department, Columbia University, New York, NY 10027. **E-mail:** leung@ieor.columbia.edu.

†Department of Applied Mathematics, University of Washington, Seattle, WA 98195. **E-mail:** mattlorig@gmail.com. Work partially supported by NSF grant DMS-0739195.

‡Dipartimento di Matematica, Università di Bologna, Bologna, Italy. **E-mail:** andrea.pascucci@unibo.it.

[1]The first US-listed ETF, the SPDR S&P 500 ETF (SPY), was launched on January 29th, 1993.

[2]"2013 ETF & Investment Outlook" by David Mazza, SPDR ETF Strategy & Consulting, State Street Global Advisors. Available at http://www.spdr-etfs.com.

take a bearish position on the underlying index by longing the fund. An example is the ProShares S&P 500 UltraShort (SDS) with leverage ratio of -2. The most typical leverage ratios are $\{-3, -2, -1, 2, 3\}$. With the same reference, such as the S&P 500, these LETFs share very similar sources of randomness, but they also exhibit different path behaviors (see Cheng and Madhavan (2009) and Avellaneda and Zhang (2010)).

The use of ETFs has also led to increased trading of options written on ETFs. During 2012, the total options contract volume traded at Chicago Board Options Exchange (CBOE) is 1.06 billion contracts, of which 282 million contracts are ETF options while 473 million are equity options. This leads to an important question of consistent pricing of options on ETFs and LETFs with the same reference. Since options are commonly quoted and compared in terms of implied volatility, it is natural to consider the implied volatility relationships among LETF options, not only across strikes and maturities, but also for various leverage ratios.

In this paper, we analyze the implied volatility surfaces associated with European-style LETF options in a general class of local-stochastic volatility (LSV) models. Our approach is to (i) find an expansion for approximate LETF option prices (ii) establish rigorous error bounds for this approximation and (iii) translate the price approximation into approximate implied volatilities. Exact pricing and implied volatility formulas in a general LSV setting are obviously impossible to obtain. There are a number of approaches one could feasibly take in order to approximate European-style option prices and their associated implied volatilities. We review some recent approaches for unleveraged products here. Gatheral et al. (2012) use heat kernel methods in a local volatility setting. Benhamou et al. (2010) use a small volatility of volatility expansion for the time-dependent Heston model. More recently, Bompis and Gobet (2013) use Malliavin calculus to obtain approximations in a quite general LSV setting. And, Forde and Jacquier (2011) use the Freidlin-Wentzell theory of large deviations to analyze an uncorrelated LSV model.

In this paper, we use the polynomial operator expansion techniques developed in Lorig et al. (2013b) to obtain approximate prices and implied volatilities (see also Lorig et al. (2014) for pricing approximations for models with jumps). The reasons for basing our expansions on the methods developed in Lorig et al. (2013b) are two-fold. First, these methods allow us to consider a large class of LSV models for the ETF; many of the above mentioned methods work only for specific ETF dynamics. Second, the methods developed in Lorig et al. (2013b) generalize to larger dimensions; the ability to make this generalization will be needed when we consider jointly the dynamics of the ETF and LETF. However, without further development, the methods described in Lorig et al. (2013b) are not sufficient for the rigorous error bounds we establish in this paper. Indeed, in Lorig et al. (2013b), error bounds are established under uniform ellipticity assumption. As we shall see, the generator of the joint ETF/LETF process is not elliptic. As such, to establish rigorous error bounds for LETF option prices, we must work in this challenging non-elliptic setting.

Perhaps the most useful result of our analysis is the general expression we obtain for the implied volatility expansion. This expansion allows us to pinpoint the non-trivial role played by the leverage ratio β, and thus, relate the implied volatility surfaces between (unleveraged) ETF and LETF options. This also motivates us to apply the idea of *log-moneyness scaling*, with the objective to view the implied volatilities across leverage ratios on the same scale and orientation. In particular, for a negative leverage ratio and up to the first order in log-moneyness, the LETF implied volatility is known to be upward sloping while the ETF and

long-LETF implied volatilities are downward sloping (see Leung and Sircar (2012)). The scaling is capable of appropriately adjusting the level and shape of the implied volatility so that the ETF and LETF implied volatilities match closely under a given model. For illustration, we also test our implied volatility expansions and the log-moneyness scaling in three well-known settings: CEV, Heston and SABR, and find that they are very accurate.

In the recent paper, Leung and Sircar (2012) apply asymptotic techniques to understand the link between implied volatilities of the ETF and LETFs of different leverage ratios within a multiscale stochastic volatility framework (see Fouque et al. (2011) for a review of multiscale methods). They also introduce implied volatility scaling procedure, different from our own, in order to identify possible price discrepancies in the ETF and LETF options markets. In contrast to their work, the current paper studies the problem in a general LSV framework, which naturally includes well-known models such as CEV, Heston and SABR models, among others. Moreover, while Leung and Sircar (2012) obtain an implied volatility approximation that is linear in log-moneyness, we provide a general expression for LETF implied volatilities that is quadratic in log-moneyness. We also provide formulas for three specific models (CEV, Heston and SABR) that are cubic in log-moneyness.

Ahn et al. (2012) propose a heuristic approximation to compute LETF option prices with Heston stochastic volatility and jumps for the underlying. While they do not investigate the implied volatilities, they point out that if the underlying ETF admits the Heston (no jumps) dynamics, then the LETF also has Heston dynamics with different parameters. As a particular example of LSV models, we also obtain the same result revealed through our implied volatility expansions (see Section 6.2).

The rest of this paper proceeds as follows. In Section 2 we review how LETF dynamics are related to ETF dynamics in a general diffusion setting. We then introduce Markov dynamics for a general class of LSV models for the ETF. Next, in Section 3, we formally construct an asymptotic expansion for European-style options whose payoff depends on the terminal value of the ETF and/or LETF. Rigorous error bounds for our pricing approximation are established in Section 4. In Section 5 we translate our asymptotic expansion for prices into an asymptotic expansion for implied volatilities. We also discuss some natural scalings of the implied volatility surface of the LETF. Finally, in Section 6 we implement our implied volatility expansion in three well-known settings: CEV, Heston and SABR. Some concluding remarks are given in Section 7.

2 Leveraged ETF dynamics

We take as given an equivalent martingale measure $\mathbb{Q}$, chosen by the market on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{Q})$. The filtration $\{\mathcal{F}_t, t \geq 0\}$ represents the history of the market. All stochastic processes defined below live on this probability space and all expectations are taken with respect to $\mathbb{Q}$. For simplicity, we assume a frictionless market, no arbitrage, zero interest rates and no dividends. We will discuss how to relax these assumptions in Remark 3.2.

Let S be the price process of an Exchange-Traded Fund (ETF). We assume S can be modeled under $\mathbb{Q}$ as a strictly positive Itô diffusion. Specifically, we have

$$\text{ETF}: \qquad S_t = \mathrm{e}^{X_t}, \qquad \mathrm{d}X_t = -\frac{1}{2}\sigma_t^2 \mathrm{d}t + \sigma_t \, \mathrm{d}W_t^x, \tag{2.1}$$

where σ is a strictly positive stochastic process. Note that the drift is fixed by the volatility so that S is a martingale. Let L be the price process of a Leveraged Exchange-Traded Fund (LETF) with underlying S and with leverage ratio β. Typical values of β are $\{-3, -2, -1, 2, 3\}$. The LETF is managed as follows: for every unit of currency a trader invests in L, the LETF manager borrows $(\beta - 1)$ units of currency and invests β units of currency in S. The fund manager also typically charges the trader a small expense rate, which, for simplicity, we assume is zero. Then the dynamics of L are related to S as follows

$$\frac{\mathrm{d}L_t}{L_t} = \beta \frac{\mathrm{d}S_t}{S_t} = \beta\sigma_t \,\mathrm{d}W_t^x,$$

and thus we have

$$\text{LETF}: \qquad L_t = \mathrm{e}^{Z_t}, \qquad \mathrm{d}Z_t = -\frac{1}{2}\beta^2\sigma_t^2 \,\mathrm{d}t + \beta\sigma_t \,\mathrm{d}W_t^x. \tag{2.2}$$

Comparing (2.1) with (2.2), we observe that the volatility of L is scaled by a factor of β. Moreover, as shown by Avellaneda and Zhang (2010), one can solve explicitly the SDE for Z in order to obtain an expression for Z_t in terms of X_t and the quadratic variation (integrated variance) of X up to time t. Specifically, we have

$$Z_t - Z_0 = \beta\,(X_t - X_0) - \frac{\beta(\beta-1)}{2}\int_0^t \sigma_s^2 \,\mathrm{d}s. \tag{2.3}$$

Equation (2.3) shows that the log returns of an LETF is the sum of two terms. The first term is proportional to the log returns of the underlying ETF. The second term, which is proportional to the integrated variance of X, highlights the fact that options on LETFs are *path dependent* options. Note that, for leverage ratio $\beta \in \{-3, -2, -1, 2, 3\}$, the coefficient $-\frac{\beta(\beta-1)}{2}$ of the realized variance is strictly negative, but such a dependence is asymmetric in β.

2.1 Local-stochastic volatility framework

We now specialize to the Markov setting. We introduce an auxiliary process Y, which is intended to capture effects such as stochastic volatility. We assume that the triple (X, Y, Z) can be modeled by the following Stochastic Differential Equation (SDE):

$$\begin{aligned}
\mathrm{d}X_t &= -\frac{1}{2}\sigma^2(t, X_t, Y_t)\mathrm{d}t + \sigma(t, X_t, Y_t)\mathrm{d}W_t^x, \\
\mathrm{d}Y_t &= c(t, X_t, Y_t)\mathrm{d}t + g(t, X_t, Y_t)\mathrm{d}W_t^y, \\
\mathrm{d}Z_t &= -\frac{1}{2}\beta^2\sigma^2(t, X_t, Y_t)\mathrm{d}t + \beta\sigma(t, X_t, Y_t)\mathrm{d}W_t^x, \\
\mathrm{d}\langle W^x, W^y\rangle_t &= \rho(t, X_t, Y_t)\mathrm{d}t.
\end{aligned} \tag{2.4}$$

We assume that SDE (2.4) has a unique strong solution and that the coefficients (σ, c, ρ) are smooth. Sufficient conditions for a unique strong solution are given in Ikeda and Watanabe (1989). The class of models described by (2.4) enjoys the following features:

1. **Stochastic Volatility**: When σ and ρ are functions of (t, y) only (as they would be in a stochastic volatility model such as Heston), then the pairs (X, Y) and (Y, Z) are Markov processes. From a mathematical point of view, the lack of x-dependence in the correlation ρ and volatility σ greatly simplifies

the pricing and implied volatility analysis, since calls written on Z can be analyzed independently from calls on X.

2. **Local Volatility**: If both σ and ρ are dependent on (t, x) only (as they would be in a local volatility model such as CEV), then X alone and the pair (X, Z) are Markov processes. In this case, calls on X can be analyzed separately from Z. However, calls on Z must be analyzed in conjunction with X.

3. **Local-Stochastic Volatility**: If σ and/or ρ depend on (x, y) (as would be the case in a local-stochastic volatility setting such as SABR), then the pair (X, Y) is a Markov process, as is the triple (X, Y, Z). In this case, options on X can be analyzed independently from Z. In contrast, to analyze options on Z, one must consider the triple (X, Y, Z).

4. If $\beta = 1$, then from (2.4) we see that $\mathrm{d}X_t = \mathrm{d}Z_t$. Thus, we need only to obtain prices and implied volatilities for options written on Z. Options written on X can always be obtained by considering the special case $\beta = 1$.

3 Option pricing

Using risk-neutral pricing and the Markov property of the process (X, Y, Z), we can write the time t price of an option $u(t, x, y, z)$ with expiration date $T > t$ and payoff $\varphi(Z_T)$ as the risk-neutral expectation of the payoff

$$u(t, x, y, z) = \mathbb{E}[\varphi(Z_T)|X_t = x, Y_t = y, Z_t = z].$$

Under mild assumptions, the function u satisfies the Kolmogorov backward equation

$$(\partial_t + \mathcal{A}(t))\, u = 0, \qquad\qquad u(T, x, y, z) = \varphi(z), \tag{3.1}$$

where the operator $\mathcal{A}(t)$ is given by

$$\begin{aligned}\mathcal{A}(t) = a(t, x, y)\left(\left(\partial_x^2 - \partial_x\right) + \beta^2\left(\partial_z^2 - \partial_z\right) + 2\beta\,\partial_x\partial_z\right) \\ + b(t, x, y)\partial_y^2 + c(t, x, y)\partial_y + f(t, x, y)\left(\partial_x\partial_y + \beta\,\partial_y\partial_z\right),\end{aligned} \tag{3.2}$$

with the functions (a, b, f) defined as

$$a(t, x, y) = \tfrac{1}{2}\sigma^2(t, x, y), \qquad b(t, x, y) = \tfrac{1}{2}g^2(t, x, y), \qquad f(t, x, y) = g(t, x, y)\sigma(t, x, y)\rho(t, x, y).$$

For general (a, b, c, f), an explicit solution to (3.1) is not available. Thus, our goal is to find a closed form approximation for the option price u and derive rigorous error bounds for our approximation.

Remark 3.1. We note that the matrix of second order derivatives of $\mathcal{A}(t)$

$$\frac{1}{2}\begin{pmatrix} 2a & f & 2\beta a \\ f & 2b & \beta f \\ 2\beta a & \beta f & 2\beta^2 a \end{pmatrix}$$

is singular; the eigenvector $(\beta, 0, -1)$ corresponds to eigenvalue zero. Therefore, the operator $\mathcal{A}(t)$ is *not* elliptic. This gives rise to an additional mathematical challenge in establishing error bounds for the pricing approximation, which we will carry out in Section 4.

Remark 3.2 (Deterministic interest rates, dividends and expense ratios)**.** Suppose interest rates are a deterministic function $r(t)$ of time. Suppose also that the ETF holder receives a dividend $q(t)S_t$ per unite time, and the LETF provider charges an expense rate $c(t)L_t$ per unit time where $q(t)$ and $c(t)$ are deterministic functions. In this case options prices are computed as discounted expectations of the form

$$\begin{aligned}\widetilde{u}(t,\widetilde{x},y,\widetilde{z}) &:= \mathbb{E}[\mathrm{e}^{-\int_t^T \mathrm{d}s\, r(s)}\varphi(\widetilde{Z}_T)|\widetilde{X}_t = \widetilde{x}, Y_t = y, \widetilde{Z}_t = z],\\ \mathrm{d}\widetilde{X}_t &= \mathrm{d}X_t + (r(t) - q(t))\,\mathrm{d}t,\\ \mathrm{d}\widetilde{Z}_t &= \mathrm{d}Z_t + (r(t) - c(t) - \beta q(t))\,\mathrm{d}t,\end{aligned}$$

with (X, Y, Z) as given in (2.4). Upon making the following change of variables

$$\begin{aligned}u(t, x(t,\widetilde{x}), y, z(t,\widetilde{z})) &:= \mathrm{e}^{\int_t^T \mathrm{d}s\, r(s)}\widetilde{u}(t,\widetilde{x},y,\widetilde{z}),\\ x(t,\widetilde{x}) &:= \widetilde{x} + \int_t^T \mathrm{d}s\, r(s),\\ z(t,\widetilde{z}) &:= \widetilde{z} + \int_t^T \mathrm{d}s\, \left(r(s) - c(s) - \beta q(s)\right),\end{aligned} \tag{3.3}$$

a simple application of the chain rule reveals that u as defined in (3.3) satisfies Cauchy problem (3.1). Thus, the current framework allows us to readily accommodate these additional features.

3.1 Asymptotic prices via Taylor and Dyson series

In this section we show how Taylor and Dyson series can be used to construct asymptotic expansions of prices. Due to the heuristic nature of the argument presented in this section, we don't specify here the exact assumptions in force; those will be given precisely in Section 4. As stated above, our goal is to find a closed form approximation for the function u that solves (3.1). For notational simplicity, we will suppress all (x, y, z)-dependence, except where it is needed for clarity.

Assume temporarily that for every t the coefficients (a, b, c, f) of the operator $\mathcal{A}(t)$ are analytic in (x, y). This assumption is not necessary for our price approximation, but it will simplify the derivation that follows. We will relax this assumption in Remark 3.4. Because the coefficients (a, b, c, f) are analytic by assumption, we can expand each of these functions as a Taylor series about an arbitrary point $(\bar{x}, \bar{y}) \in \mathbb{R}^2$. More precisely, for any (t, x, y) we have:

$$\chi(t,x,y) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \chi_{n-k,k}(t)(x-\bar{x})^{n-k}(y-\bar{y})^k, \qquad \chi_{n-k,k}(t) = \frac{\partial_x^{n-k}\partial_y^k \chi(t,\bar{x},\bar{y})}{(n-k)!k!}, \qquad \chi = \{a, b, c, f\}.$$

Formally, the operator $\mathcal{A}(t)$ can now be written as

$$\mathcal{A}(t) = \mathcal{A}_0(t) + \mathcal{B}_1(t), \qquad \mathcal{B}_1(t) = \sum_{n=1}^{\infty}\mathcal{A}_n(t), \qquad \mathcal{A}_n(t) = \sum_{k=0}^{n}(x-\bar{x})^{n-k}(y-\bar{y})^k \mathcal{A}_{n-k,k}(t), \tag{3.4}$$

where

$$\mathcal{A}_{n-k,k}(t) = a_{n-k,k}(t)\left(\left(\partial_x^2 - \partial_x\right) + \beta^2\left(\partial_z^2 - \partial_z\right) + 2\beta\,\partial_x\partial_z\right) + b_{n-k,k}(t)\partial_y^2 + c_{n-k,k}(t)\partial_y + f_{n-k,k}(t)\left(\partial_x\partial_y + \beta\,\partial_y\partial_z\right),$$

Inserting expansion (3.4) for $\mathcal{A}(t)$ into Cauchy problem (3.1) we find

$$(\partial_t + \mathcal{A}_0(t))u(t) = -\mathcal{B}_1(t)u(t), \qquad u(T) = \varphi.$$

By construction, the operator $\mathcal{A}_0(t) \equiv \mathcal{A}_{0,0}(t)$ is the generator of a diffusion with coefficients that are deterministic functions of time only. By Duhamel's principle, we therefore have

$$u(t) = \mathcal{P}_0(t,T)\varphi + \int_t^T \mathrm{d}t_1\, \mathcal{P}_0(t,t_1)\mathcal{B}_1(t_1)u(t_1), \tag{3.5}$$

where $\mathcal{P}_0(t,T) = \exp\int_t^T \mathrm{d}s\,\mathcal{A}_0(s)$, is the *semigroup* of operators generated by $\mathcal{A}_0(t)$; we will provide the precise form for $\mathcal{P}_0(t,T)$ in Section 3.2. Inserting expression (3.5) for u back in to the right-hand side of (3.5) and iterating we obtain

$$\begin{aligned}
u(t) &= \mathcal{P}_0(t,T)\varphi + \int_t^T \mathrm{d}t_1\, \mathcal{P}_0(t,t_1)\mathcal{B}_1(t_1)\mathcal{P}_0(t_1,T)\varphi \\
&\quad + \int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2\, \mathcal{P}_0(t,t_1)\mathcal{B}_1(t_1)\mathcal{P}_0(t_1,t_2)\mathcal{B}_1(t_2)u(t_2) \\
&= \cdots \\
&= \mathcal{P}_0(t,T)\varphi + \sum_{k=1}^{\infty}\int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \cdots \int_{t_{k-1}}^T \mathrm{d}t_k \\
&\quad \mathcal{P}_0(t,t_1)\mathcal{B}_1(t_1)\mathcal{P}_0(t_1,t_2)\mathcal{B}_1(t_2)\cdots\mathcal{P}_0(t_{k-1},t_k)\mathcal{B}_1(t_k)\mathcal{P}_0(t_k,T)\varphi \qquad (3.6) \\
&= \mathcal{P}_0(t,T)\varphi + \sum_{n=1}^{\infty}\sum_{k=1}^{n}\int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \cdots \int_{t_{k-1}}^T \mathrm{d}t_k \\
&\quad \sum_{i\in I_{n,k}} \mathcal{P}_0(t,t_1)\mathcal{A}_{i_1}(t_1)\mathcal{P}_0(t_1,t_2)\mathcal{A}_{i_2}(t_2)\cdots\mathcal{P}_0(t_{k-1},t_k)\mathcal{A}_{i_k}(t_k)\mathcal{P}_0(t_k,T)\varphi, \qquad (3.7)
\end{aligned}$$

$$I_{n,k} = \{i = (i_1, i_2, \cdots, i_k) \in \mathbb{N}^k : i_1 + i_2 + \cdots + i_k = n\}. \tag{3.8}$$

Note that the second to last equality (3.6) is the classical Dyson series expansion of u corresponding to order zero generator $\mathcal{A}_0(t)$ and perturbation $\mathcal{B}_1(t)$. To obtain (3.7) from (3.6) we have used the fact that, by (3.4), the operator $\mathcal{B}_1(t)$ is an infinite sum. In light of expansion (3.7) we make the following definition:

Definition 3.3. We define $\bar{u}_N$ our Nth order approximation of u by

$$\bar{u}_N = \sum_{n=0}^{N} u_n, \qquad \text{where} \qquad u_0(t) := \mathcal{P}_0(t,T)\varphi, \tag{3.9}$$

and

$$u_n(t) := \sum_{k=1}^{n}\int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \cdots \int_{t_{k-1}}^T \mathrm{d}t_k$$

$$\sum_{i\in I_{n,k}} \mathcal{P}_0(t,t_1)\mathcal{A}_{i_1}(t_1)\mathcal{P}_0(t_1,t_2)\mathcal{A}_{i_2}(t_2)\cdots\mathcal{P}_0(t_{k-1},t_k)\mathcal{A}_{i_k}(t_k)\mathcal{P}_0(t_k,T)\varphi. \tag{3.10}$$

Remark 3.4. Although we have assumed analytic coefficients (a,b,c,f), from Definition 3.3 we see that, in order to construct the Nth order approximation $\bar{u}_N$, one requires only differentiability of the coefficients up to order N.

3.2 Expression for u_0

The *fundamental solution* Γ_0 corresponding to generator $\mathcal{A}_0(t)$, i.e., the solution of

$$(\partial_t + \mathcal{A}_0)\Gamma_0(t,x,y,z;T,\xi,\eta,\zeta) = 0, \qquad \Gamma_0(T,x,y,z;T,\xi,\eta,\zeta) = \delta_{\xi,\eta,\zeta}(x,y,z),$$

is given by

$$\Gamma_0(t,x,y;T,\xi,\eta,\zeta) = \frac{1}{2\pi\sqrt{|\mathbf{C}|}}\exp\left(-\frac{1}{2}\mathbf{m}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{m}\right)\delta\left(\zeta - z - \beta(\xi-x) + \beta(\beta-1)\int_0^t a_{0,0}(s)\mathrm{d}s\right),$$

where $\delta(\cdot)$ is a Dirac delta function, and the covariance matrix $\mathbf{C}$ and vector $\mathbf{m}$ are given by:

$$\mathbf{C} = \begin{pmatrix} 2\int_t^T a_{0,0}(s)\mathrm{d}s & \int_t^T f_{0,0}(s)\mathrm{d}s \\ \int_t^T f_{0,0}(s)\mathrm{d}s & 2\int_t^T b_{0,0}(s)\mathrm{d}s \end{pmatrix}, \qquad \mathbf{m} = \begin{pmatrix} \xi - x + \int_t^T a_{0,0}(s)\mathrm{d}s \\ \eta - y - \int_t^T c_{0,0}(s)\mathrm{d}s \end{pmatrix}.$$

The Dirac delta function $\delta(\cdot)$ appears in the fundamental solution because the covariance matrix of $\mathcal{A}_0$ is singular. The action of the semigroup $\mathcal{P}_0(t,T)$ generated by $\mathcal{A}_0(t)$ when acting on a function $\theta:\mathbb{R}^3\to\mathbb{R}$ is

$$\mathcal{P}_0(t,T)\theta(x,y,z) = \int_{\mathbb{R}^3}\mathrm{d}\xi\mathrm{d}\eta\mathrm{d}\zeta\,\Gamma_0(t,x,y;T,\xi,\eta,\zeta)\theta(\xi,\eta,\zeta). \tag{3.11}$$

Using (3.9) and (3.11) a direct computation gives the zeroth order approximation

$$u_0(t) \equiv u_0(t,z) := \mathcal{P}_0(t,T)\varphi(z) = \int_{\mathbb{R}}\mathrm{d}\zeta\,\frac{1}{\sqrt{2\pi s^2(t,T)}}\exp\left(\frac{-(\zeta - m(t,T))^2}{2s^2(t,T)}\right)\varphi(\zeta), \tag{3.12}$$

where the mean $m(t,T)$ and variance $s^2(t,T)$ are given by

$$m(t,T) = z - \beta^2\int_t^T \mathrm{d}t_1\,a_{0,0}(t_1), \qquad s^2(t,T) = 2\beta^2\int_t^T \mathrm{d}t_1\,a_{0,0}(t_1).$$

3.3 Expression for u_n

The following Theorem, and the ensuing proof, show that $u_n(t)$ can be written as a differential operator $\mathcal{L}_n(t,T)$ acting on $u_0(t)$.

Theorem 3.5. *The function u_n defined in* (3.10) *is given explicitly by*

$$u_n(t) = \mathcal{L}_n(t,T)u_0(t), \tag{3.13}$$

where u_0 is given by (3.12) *and*

$$\mathcal{L}_n(t,T) = \sum_{k=1}^n\int_t^T \mathrm{d}t_1\int_{t_1}^T \mathrm{d}t_2\cdots\int_{t_{k-1}}^T \mathrm{d}t_k\sum_{i\in I_{n,k}}\mathcal{G}_{i_1}(t,t_1)\mathcal{G}_{i_2}(t,t_2)\cdots\mathcal{G}_{i_k}(t,t_k), \tag{3.14}$$

with $I_{n,k}$ as defined in (3.8) *and*

$$\mathcal{G}_n(t,t_i) := \sum_{k=0}^{n} \left(\mathcal{M}_x(t,t_i) - \bar{x}\right)^{n-k} \left(\mathcal{M}_y(t,t_i) - \bar{y}\right)^k \mathcal{A}_{n-k,k}(t_i) \tag{3.15}$$

$$\mathcal{M}_x(t,t_i) := x + \int_t^{t_i} \mathrm{d}s \Big(a_{0,0}(s)\left(2\partial_x + 2\beta\partial_z - 1\right) + f_{0,0}(s)\partial_y \Big), \tag{3.16}$$

$$\mathcal{M}_y(t,t_i) := y + \int_t^{t_i} \mathrm{d}s \Big(f_{0,0}(s)\left(\partial_x + \beta\partial_z\right) + 2b_{0,0}(s)\partial_y + c_{0,0}(s) \Big). \tag{3.17}$$

Proof. The proof consists showing that the operator $\mathcal{G}_i(t,t_k)$ in (3.15) satisfies

$$\mathcal{P}_0(t,t_k)\mathcal{A}_i(t_k) = \mathcal{G}_i(t,t_k)\mathcal{P}_0(t,t_k). \tag{3.18}$$

Assuming (3.18) holds, we can use the fact that $\mathcal{P}_0(t,T)$ satisfies the semigroup property

$$\mathcal{P}_0(t,T) = \mathcal{P}_0(t,t_1)\mathcal{P}_0(t_1,t_2)\cdots\mathcal{P}_0(t_{k-1},t_k)\mathcal{P}_0(t_k,T),$$

and we can re-write (3.10) as

$$u_n(t) = \sum_{k=1}^{n} \int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \cdots \int_{t_{k-1}}^T \mathrm{d}t_k \sum_{i\in I_{n,k}} \mathcal{G}_{i_1}(t,t_1)\mathcal{G}_{i_2}(t,t_2)\cdots\mathcal{G}_{i_k}(t,t_k)\mathcal{P}_0(t,T)\varphi.$$

Noting that $\mathcal{P}_0(t,T)\varphi = u_0(t)$, equations (3.13)-(3.14) follow directly. Thus, we only need to show that $\mathcal{G}_i(t,t_k)$ satisfies (3.18). At this point, it will be helpful to introduce some compact notation:

$$\mathbf{x} = (x_1,x_2,x_3) = (x,y,z), \qquad \boldsymbol{\xi} = (\xi_1,\xi_2,\xi_3), \qquad \langle \boldsymbol{\xi}, \mathbf{x}\rangle = \sum_{i=1}^{3} \xi_i x_i.$$

It is sufficient to investigate how the operator $\mathcal{P}_0(t,t_k)\mathcal{A}_i(t_k)$ acts on the *oscillating exponential*:

$$\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) := \mathrm{e}^{\mathtt{i}\langle \boldsymbol{\xi},\mathbf{x}\rangle}.$$

First, we note that

$$\mathcal{P}_0(t,t_k)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) = \mathrm{e}^{\Phi_0(t,t_k,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}), \tag{3.19}$$

where $\Phi_0(t,t_k,\boldsymbol{\xi})$ is as given by

$$\Phi_0(t_0,T,\boldsymbol{\xi}) = \int_{t_0}^T \mathrm{d}t \Big(a_{0,0}(t)\left(\left(-\xi_1^2 - \mathtt{i}\xi_1\right) + \beta^2\left(-\xi_3^2 - \mathtt{i}\xi_3\right) - 2\beta\,\xi_1\xi_3\right) - b_{0,0}(t)\xi_2^2 + c_{0,0}(t)\mathtt{i}\xi_2 - f_{0,0}(t)\left(\xi_1\xi_2 + \beta\,\xi_2\xi_3\right) \Big).$$

Next, introducing $\nabla_{\mathbf{x}} := (\partial_{x_1},\partial_{x_z},\partial_{x_3})$, we observe that the operator $\mathcal{M}_x(t,t_k)$ and $\mathcal{M}_y(t,t_k)$ in (3.16) and (3.17) can be written as

$$\mathcal{M}_x(t,t_k) = M_x(t,t_k,-\mathtt{i}\nabla_{\mathbf{x}}), \qquad M_x(t,t_k,\boldsymbol{\xi}) = -\mathtt{i}\partial_{\xi_1}\left(\Phi_0(t,t_k,\boldsymbol{\xi}) + \mathtt{i}\langle\boldsymbol{\xi},x\rangle\right), \tag{3.20}$$

$$\mathcal{M}_y(t,t_k) = M_y(t,t_k,-\mathtt{i}\nabla_{\mathbf{x}}), \qquad M_y(t,t_k,\boldsymbol{\xi}) = -\mathtt{i}\partial_{\xi_2}\left(\Phi_0(t,t_k,\boldsymbol{\xi}) + \mathtt{i}\langle\boldsymbol{\xi},x\rangle\right). \tag{3.21}$$

Using (3.20) and (3.21) we observe that, for any natural numbers n and k, we have

$$\begin{aligned}
(-\mathtt{i}\partial_{\xi_1})^n(-\mathtt{i}\partial_{\xi_2})^k \mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) &= (-\mathtt{i}\partial_{\xi_1})^n(-\mathtt{i}\partial_{\xi_2})^{k-1} M_y(t,t_i,\boldsymbol{\xi})\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= \mathcal{M}_y(t,t_i)(-\mathtt{i}\partial_{\xi_1})^n(-\mathtt{i}\partial_{\xi_2})^{k-1}\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= (\mathcal{M}_y(t,t_i))^k(-\mathtt{i}\partial_{\xi_1})^n\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= (\mathcal{M}_y(t,t_i))^k(\mathcal{M}_x(t,t_i))^n\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= \mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}(\mathcal{M}_y(t,t_i))^k(\mathcal{M}_x(t,t_i))^n\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}).
\end{aligned}$$

Furthermore, since ∂_{ξ_1} and ∂_{ξ_2} commute, so do $\mathcal{M}_x(t,t_i)$ and $\mathcal{M}_y(t,t_i)$. Therefore,

$$(-\mathtt{i}\partial_{\xi_1} - \bar{x})^n(-\mathtt{i}\partial_{\xi_2} - \bar{y})^k\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) = (\mathcal{M}_x(t,t_i) - \bar{x})^n(\mathcal{M}_y(t,t_i) - \bar{y})^k\mathrm{e}^{\Phi_0(t,t_i,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}). \tag{3.22}$$

Next, we define $\Phi_{n,k}(t,\boldsymbol{\xi})$ the *symbol* of $\mathcal{A}_{n,k}(t)$, which is defined through the following relation

$$\mathcal{A}_{n,k}(t)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) = \Phi_{n,k}(t,\boldsymbol{\xi})\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}), \tag{3.23}$$

and given explicitly by

$$\begin{aligned}
\Phi_{n,k}(t,\boldsymbol{\xi}) := a_{n,k}(t)\Big(\left(-\xi_1^2 - \mathtt{i}\xi_1\right) + \beta^2\left(-\xi_3^2 - \mathtt{i}\xi_3\right) - 2\beta\,\xi_1\xi_3\Big) \\
- b_{n,k}(t)\xi_2^2 + c_{n,k}(t)\mathtt{i}\xi_2 - f_{n,k}(t)\left(\xi_1\xi_2 + \beta\,\xi_2\xi_3\right).
\end{aligned}$$

Finally, we compute

$$\begin{aligned}
\mathcal{P}_0(t,t_i)\mathcal{A}_n(t_i)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) &= \mathcal{P}_0(t,t_i)\sum_{k=0}^{n}(x-\bar{x})^{n-k}(y-\bar{y})^k\mathcal{A}_{n-k,k}(t_i)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= \mathcal{P}_0(t,t_i)\sum_{k=0}^{n}\Phi_{n-k,k}(t_i,\boldsymbol{\xi})(x-\bar{x})^{n-k}(y-\bar{y})^k\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) && \text{(by (3.23))} \\
&= \mathcal{P}_0(t,t_i)\sum_{k=0}^{n}\Phi_{n-k,k}(t_i,\boldsymbol{\xi})(-\mathtt{i}\partial_{\xi_1} - \bar{x})^{n-k}(-\mathtt{i}\partial_{\xi_2} - \bar{y})^k\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= \sum_{k=0}^{n}\Phi_{n-k,k}(t_i,\boldsymbol{\xi})(-\mathtt{i}\partial_{\xi_1} - \bar{x})^{n-k}(-\mathtt{i}\partial_{\xi_2} - \bar{y})^k\mathcal{P}_0(t,t_i)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) \\
&= \sum_{k=0}^{n}\Phi_{n-k,k}(t_i,\boldsymbol{\xi})(-\mathtt{i}\partial_{\xi_1} - \bar{x})^{n-k}(-\mathtt{i}\partial_{\xi_2} - \bar{y})^k\mathrm{e}^{\Phi_0(t,t_k,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) && \text{(by (3.19))} \\
&= \sum_{k=0}^{n}\Phi_{n-k,k}(t_i,\boldsymbol{\xi})(\mathcal{M}_x(t,t_i) - \bar{x})^{n-k}(\mathcal{M}_y(t,t_i) - \bar{y})^k\mathrm{e}^{\Phi_0(t,t_k,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) && \text{(by (3.22))} \\
&= \sum_{k=0}^{n}(\mathcal{M}_x(t,t_i) - \bar{x})^{n-k}(\mathcal{M}_y(t,t_i) - \bar{y})^k\mathcal{A}_{n-k,k}(t_i)\mathrm{e}^{\Phi_0(t,t_k,\boldsymbol{\xi})}\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) && \text{(by (3.23))} \\
&= \sum_{k=0}^{n}(\mathcal{M}_x(t,t_i) - \bar{x})^{n-k}(\mathcal{M}_y(t,t_i) - \bar{y})^k\mathcal{A}_{n-k,k}(t_i)\mathcal{P}_0(t,t_i)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}) && \text{(by (3.19))} \\
&= \mathcal{G}_n(t,t_i)\mathcal{P}_0(t,t_i)\mathrm{e}_{\boldsymbol{\xi}}(\mathbf{x}). && \text{(by (3.15))}
\end{aligned}$$

Thus, $\mathcal{P}_0(t,t_i)\mathcal{A}_n(t_i) = \mathcal{G}_n(t,t_i)\mathcal{P}_0(t,t_i)$, completing the proof. □

Remark 3.6. Note that, by (3.12), the order zero price u_0 is simply an integral of the option payoff φ versus a Gaussian kernel Γ_0, just as in the Black-Scholes model. From Theorem 3.5 we see that higher order terms u_n can be obtained by applying the differential operator $\mathcal{L}_n$ to u_0. The operator $\mathcal{L}_n$ acts on the backward variable z, which is present only in the Gaussian kernel Γ_0, producing (Hermite) polynomials in the forward variable ζ multiplied by Γ_0. Thus, every term in the price expansion is of the form

$$u_n(t,z) = \int_{\mathbb{R}} \mathrm{d}\zeta \, \frac{p_n(\zeta)}{\sqrt{2\pi s^2(t,T)}} \exp\left(\frac{-(\zeta - m(t,T))^2}{2s^2(t,T)}\right) \varphi(\zeta).$$

where the function p_n is a polynomial. As such, computation times for approximate prices are comparable to the Black-Scholes model.

4 Accuracy of the option-pricing approximation

The goal of this Section is to establish a rigorous error bound for the Nth order pricing approximation described in the previous Sections. We will adapt the methods from Lorig et al. (2013a), which apply only to uniformly elliptic operators, to our current case where the operator $\mathcal{A}(t)$ is singular (see Remark 3.1).

Our main error bound is given in Corollary 4.4 at the end of this Section. As a first step we introduce a process Z^ε, which is a modification of the dynamics of Z in (2.4). Specifically, we define

$$\mathrm{d}Z_t^\varepsilon := \mathrm{d}Z_t - \frac{1}{2}\varepsilon^2 \mathrm{d}t + \varepsilon \mathrm{d}W_t^z, \qquad \mathrm{d}\langle W^x, W^z\rangle = 0, \qquad \mathrm{d}\langle W^y, W^z\rangle = 0, \qquad \varepsilon \geq 0.$$

We will first establish the accuracy of our pricing approximation $\bar{u}_N^\varepsilon$ for options written on the Markov process (X, Y, Z^ε) with error estimates that are *uniform* in ε. Error bounds for the price approximation $\bar{u}_N$ for options written on the Markov process (X, Y, Z) will follow in the limit as $\varepsilon \to 0$. Here, $\bar{u}_N^\varepsilon$ is constructed just as $\bar{u}_N$ by replacing the generator $\mathcal{A}(t)$ in (3.2) with the generator of the Markov process (X, Y, Z^ε).

Introducing a process V^ε, which satisfies the following SDE:

$$\mathrm{d}V_t^\varepsilon = \frac{1}{2}\left(\beta(1-\beta)\sigma^2(t, X_t, Y_t) - \varepsilon^2\right)\mathrm{d}t + \varepsilon \mathrm{d}W_t^z,$$

we note that the dynamics of Z^ε can be written as follows

$$\mathrm{d}Z_t^\varepsilon = \beta \, \mathrm{d}X_t + \mathrm{d}V_t^\varepsilon.$$

Therefore, rather than considering the generator of (X, Y, Z^ε), we can consider the generator of (X, Y, V^ε), which we denote by $\mathcal{A}^\varepsilon(t)$. This operator separates into an operator $\mathcal{X}(t)$, which takes derivatives with respect to (x, y), and an operator $\mathcal{V}^\varepsilon(t)$, which takes derivatives with respect to v. That is,

$$\mathcal{A}^\varepsilon(t) = \mathcal{X}(t) + \mathcal{V}^\varepsilon(t),$$

where

$$\begin{aligned}
\mathcal{X}(t) &= a(t,x,y)\left(\partial_x^2 - \partial_x\right) + b(t,x,y)\partial_y^2 + f(t,x,y)\partial_x\partial_y + c(t,x,y)\partial_y, \\
\mathcal{V}^\varepsilon(t) &= \frac{\varepsilon^2}{2}\left(\partial_v^2 - \partial_v\right) + a(t,x,y)\beta(1-\beta)\partial_v.
\end{aligned}$$

Now, let us introduce A, the matrix of second order coefficients of $\mathcal{X}(t)$

$$A(t,x,y) := \frac{1}{2}\begin{pmatrix} 2a(t,x,y) & f(t,x,y) \\ f(t,x,y) & 2b(t,x,y) \end{pmatrix}.$$

Assumption 4.1. There exists a positive constant M such that:

i) *Uniform ellipticity:*

$$M^{-1}|\xi|^2 \leq \sum_{i,j=1}^{d} A_{ij}(t,x,y)\xi_i\xi_j \leq M|\xi|^2, \qquad t \in [0,T],\ (x,y),\xi \in \mathbb{R}^2.$$

ii) *Regularity and boundedness:* the coefficients $a(t,\cdot,\cdot), b(t,\cdot,\cdot), f(t,\cdot,\cdot)$ are $C^{N+1}(\mathbb{R}^2)$ with derivatives of any order bounded by M.

As a consequence of Assumption 4.1, for any fixed $\varepsilon > 0$, the operator $\mathcal{A}^\varepsilon(t)$ is uniformly elliptic. Moreover, $(\partial_t + \mathcal{A}^\varepsilon(t))$ is uniformly parabolic and has a fundamental solution, denoted by

$$\Gamma^\varepsilon = \Gamma^\varepsilon(t,x,y,v;T,x',y',v'), \qquad\qquad t < T,$$

which satisfies the following Gaussian estimates

Lemma 4.2. *Let $i,j,h,k \in \mathbb{N}_0$ with $h+k \leq N+2$, and $\bar{T} > 0$. Then we have*

$$\left|(x-x')^i(y-y')^j\partial_x^h\partial_y^k\Gamma^\varepsilon(t,x,y,v;T,x',y',v')\right| \leq \mathbf{c}_0 \cdot (T-t)^{\frac{i+j-h-k}{2}}\Gamma_{\text{heat}}^{(M,\varepsilon)}(t,x,y,v;T,x',y',v') \quad (4.1)$$

for any $x,y,v,x',y',v' \in \mathbb{R}$, $0 \leq t < T \leq \bar{T}$ and $\varepsilon \in (0,1]$. Here, $\Gamma_{\text{heat}}^{(M,\varepsilon)}$ denotes the (Gaussian) fundamental solution of the heat operator

$$\partial_t + M(\partial_{xx} + \partial_{yy}) + \frac{\varepsilon^2}{2}\partial_{vv},$$

and $\mathbf{c}_0$ is a positive constant that depends only on M, N, i, j and $\bar{T}$. In particular, the constant $\mathbf{c}_0$ is independent of ε.

Proof. Estimate (4.1) differs from the classical Gaussian estimates for parabolic equations (cf. Friedman (1964), see also Pascucci (2011); Di Francesco and Pascucci (2005) for a more recent and general presentation) because the operator $(\partial_t + \mathcal{A}^\varepsilon(t))$, while parabolic, is *not uniformly parabolic* with respect to $\varepsilon \in (0,1]$. Nevertheless, the thesis can be proved by mimicking the classical argument which is based on the parametrix method and carefully checking that the constant $\mathbf{c}_0$ is independent of ε. In particular, the main ingredient in the parametrix construction is the following uniform in ε Gaussian estimate (see for instance Proposition 3.1 in Di Francesco and Pascucci (2005)) which, in our setting, is as follows: for any fixed $(\bar{x},\bar{y}) \in \mathbb{R}^2$, define the following the operators with frozen coefficients

$$\mathcal{A}^\varepsilon_{\bar{x},\bar{y}}(t) := \mathcal{X}_{\bar{x},\bar{y}}(t) + \frac{\varepsilon^2}{2}\partial_v^2,$$
$$\mathcal{X}_{\bar{x},\bar{y}}(t) := a(t,\bar{x},\bar{y})\partial_x^2 + b(t,\bar{x},\bar{y})\partial_y^2 + f(t,\bar{x},\bar{y})\partial_x\partial_y.$$

Let $\Gamma^{\varepsilon}_{\bar{x},\bar{y}}$ and $\Gamma_{\bar{x},\bar{y}}$ be the fundamental solutions corresponding to $(\partial_t + \mathcal{A}^{\varepsilon}_{\bar{x},\bar{y}})$ and $(\partial_t + \mathcal{X}_{\bar{x},\bar{y}})$ respectively. Then for every $\bar{x},\bar{y},x,y,v,x',y',v' \in \mathbb{R}$, $0 \le t < T \le \bar{T}$ and $\varepsilon \in (0,1]$, we have

$$M^{-2}\Gamma^{(M^{-1},\varepsilon)}_{\text{heat}}(t,x,y,v;T,x',y',v') \le \Gamma^{\varepsilon}_{\bar{x},\bar{y}}(t,x,y,v;T,x',y',v') \le M^2\Gamma^{(M,\varepsilon)}_{\text{heat}}(t,x,y,v;T,x',y',v'). \quad (4.2)$$

Estimate (4.2) can be readily proved as in Proposition 3.1 in Di Francesco and Pascucci (2005), by noting that $\Gamma^{\varepsilon}_{\bar{x},\bar{y}} = \Gamma_{\bar{x},\bar{y}}\Gamma_{\varepsilon}$ where Γ_{ε} is the fundamental solution of the one-dimensional heat (parabolic) operator $(\partial_t + \frac{\varepsilon^2}{2}\partial_{vv})$. Notice that (4.2) is uniform in ε (i.e. the constants in the estimates are independent of ε). Based on this fact we can prove the estimate (4.1) with $\mathbf{c}_0$ independent of ε. □

We are now in a position to state our main error estimate. In the following Theorem, Γ^{ε} is the fundamental solution corresponding to $(\partial_t + \mathcal{A}^{\varepsilon}(t))$ and $\bar{\Gamma}^{\varepsilon}_N$ is the Nth order approximation of Γ^{ε}.

Theorem 4.3. *Let Assumption 4.1 hold and let $\bar{T} > 0$. Then we have*

$$\left|\Gamma^{\varepsilon}(t,x,y,v;T,x',y',v') - \bar{\Gamma}^{\varepsilon}_N(t,x,y,v;T,x',y',v')\right| \le \mathbf{c}_1 \cdot (T-t)^{\frac{N+1}{2}}\Gamma^{(M,\varepsilon)}_{\text{heat}}(t,x,y,v;T,x',y',v'), \quad (4.3)$$

for any $x,y,v,x',y',v' \in \mathbb{R}$, $0 \le t < T$ and $\varepsilon \in (0,1]$, where $\mathbf{c}_1$ is a positive constant that depends on $M,N,\overline{T}$ but is independent of ε.

Proof. Using the uniform in ε estimate (4.1) and the ellipticity of $\mathcal{A}^{\varepsilon}(t)$, we can repeat step by step the proof of (Lorig et al., 2013a, Theorem 3.10), carefully checking that, since $\mathbf{c}_0$ in (4.1) does not depend on ε, neither does $\mathbf{c}_1$. □

In the following Corollary, u^{ε} is the solution to PDE $(\partial_t + \mathcal{A}^{\varepsilon}(t))u^{\varepsilon} = 0$ with final data $u^{\varepsilon}(T) = \varphi$ and $\bar{u}^{\varepsilon}_N$ is the Nth order approximation of u^{ε}.

Corollary 4.4. *Let Assumption 4.1 hold and let $\bar{T} > 0$. Additionally, assume also that the final datum φ is bounded and Lipschitz continuous*[3]*. Then for any $\varepsilon \in [0,1]$ we have*

$$|u^{\varepsilon}(t,x,y,v) - \bar{u}^{\varepsilon}_N(t,x,y,v)| \le \mathbf{c}_1 \cdot (T-t)^{\frac{N+2}{2}}, \qquad 0 \le t < T,\ x,y,v \in \mathbb{R}, \quad (4.4)$$

where the constant $\mathbf{c}_1$ only depends on $M,N,\overline{T}$ and the Lipschitz constant of φ.

Proof. Estimate (4.4) follows from (4.3) by integrating against the payoff function φ against the fundamental solution Γ^{ε} and against Nth order approximation $\bar{\Gamma}^{\varepsilon}_N$. In particular, the case $\varepsilon = 0$ follows in the limit since $\mathbf{c}_1$ is independent of ε. □

We emphasize that Corollary 4.4 includes the case $\varepsilon = 0$, for which $u^{\varepsilon}|_{\varepsilon=0} \equiv u$ and $\bar{u}^{\varepsilon}_N|_{\varepsilon=0} \equiv \bar{u}_N$. As a result, we have established the accuracy of the pricing approximation $\bar{u}_N$ to the exact price u.

[3]The assumptions on the final datum can be considerably weakened to include the payoff of Call options or even discontinuous payoffs: we refer to (Lorig et al., 2013a, Theorem 3.10) for further details.

5 Implied volatility

In this Section, we translate our price expansion for a call option with payoff function $\varphi(z) = (\mathrm{e}^z - \mathrm{e}^k)^+$ into an expansion in implied volatility. To ease notation we shall suppress much of the dependence on (t, T, x, y, z, k). However, one should keep in mind that prices and implied volatilities do depend on these quantities, even if this is not explicitly indicated. We begin our analysis by recalling the definitions of the Black-Scholes call price and implied volatility

Definition 5.1. The *Black-Scholes Call price* $u^{\mathrm{BS}} : \mathbb{R}^+ \to \mathbb{R}^+$ is given by

$$u^{\mathrm{BS}}(\sigma) := \mathrm{e}^z \mathcal{N}(d_+(\sigma)) - \mathrm{e}^k \mathcal{N}(d_-(\sigma)), \qquad d_\pm(\sigma) := \frac{1}{\sigma\sqrt{\tau}} \left(z - k \pm \frac{\sigma^2 \tau}{2} \right), \qquad \tau := T - t, \quad (5.1)$$

where $\mathcal{N}$ is the CDF of a standard normal random variable.

Definition 5.2. For fixed (t, T, z, k), the *implied volatility* corresponding to a call price $u \in ((\mathrm{e}^z - \mathrm{e}^k)^+, \mathrm{e}^z)$ is defined as the unique strictly positive real solution σ of the equation

$$u^{\mathrm{BS}}(\sigma) = u, \quad (5.2)$$

where u^{BS} is given by (5.1).

Theorem 5.3. *For a European call option with payoff function $\varphi(z) = (\mathrm{e}^z - \mathrm{e}^k)^+$ we have*

$$u_0 = u^{\mathrm{BS}}(\sigma_0), \qquad \sigma_0^2 = \frac{2\beta^2}{T - t} \int_t^T \mathrm{d}s \, a_{0,0}(s).$$

Proof. The proof follows directly from (3.12) with $\varphi(z) = (\mathrm{e}^z - \mathrm{e}^k)^+$. □

From Theorem 5.3 we note that the price expansion (3.9) is of the form

$$u = u^{\mathrm{BS}}(\sigma_0) + \sum_{n=1}^{\infty} u_n. \quad (5.3)$$

As shown in Lorig et al. (2013b); Jacquier and Lorig (2013), the special form (5.3) lends itself to an expansion

$$\sigma = \sigma_0 + \eta, \qquad \eta = \sum_{n=1}^{\infty} \sigma_n,$$

of implied volatility. To see this, one expands $u^{\mathrm{BS}}(\sigma)$ as a Taylor series about the point σ_0

$$\begin{aligned} u^{\mathrm{BS}}(\sigma) &= u^{\mathrm{BS}}(\sigma_0 + \eta) \\ &= u^{\mathrm{BS}}(\sigma_0) + \eta \, \partial_\sigma u^{\mathrm{BS}}(\sigma_0) + \frac{1}{2!}\eta^2 \partial_\sigma^2 u^{\mathrm{BS}}(\sigma_0) + \frac{1}{3!}\eta^3 \partial_\sigma^3 u^{\mathrm{BS}}(\sigma_0) + \dots . \end{aligned} \quad (5.4)$$

Inserting expansions (5.3) and (5.4) into equation (5.2), one can solve iteratively for every term in the sequence $(\sigma_n)_{n \geq 1}$. The first three terms, which are enough to provide an accurate approximation of implied volatility, are

$$\sigma_1 = \frac{u_1}{\partial_\sigma u^{\mathrm{BS}}(\sigma_0)}, \qquad \sigma_2 = \frac{u_2 - \frac{1}{2}\sigma_1^2 \partial_\sigma^2 u^{\mathrm{BS}}(\sigma_0)}{\partial_\sigma u^{\mathrm{BS}}(\sigma_0)}, \qquad \sigma_3 = \frac{u_3 - \left(\sigma_2 \sigma_1 \partial_\sigma^2 + \frac{1}{3!}\sigma_1^3 \partial_\sigma^3 \right) u^{\mathrm{BS}}(\sigma_0)}{\partial_\sigma u^{\mathrm{BS}}(\sigma_0)}. \quad (5.5)$$

A general expression for the n-th order term can be found in Lorig et al. (2013b); Jacquier and Lorig (2013).

As written, the expressions in (5.5) are not particularly useful. Indeed $u^{\mathrm{BS}}(\sigma_0)$ and u_n are Gaussian integrals, which are not numerically intensive to compute, but do not give much explicit information about how implied volatility depends on $(t, T, x, y, z, k, \beta)$. However, using (5.1) a direct computation shows

$$\frac{\partial_\sigma^2 u^{\mathrm{BS}}(\sigma)}{\partial_\sigma u^{\mathrm{BS}}(\sigma)} = \frac{(k-z)^2}{\tau\sigma^3} - \frac{\tau\sigma}{4}, \qquad \frac{\partial_\sigma^3 u^{\mathrm{BS}}(\sigma)}{\partial_\sigma u^{\mathrm{BS}}(\sigma)} = \frac{(k-z)^4}{\tau^2\sigma^6} - \left(\frac{3}{\tau\sigma^4} - \frac{1}{2\sigma^2}\right)(k-z)^2 + \frac{\tau^2\sigma^2}{16} - \frac{\tau}{4}. \quad (5.6)$$

In general, every term of the form $\partial_\sigma^n u^{\mathrm{BS}}(\sigma_0)/\partial_\sigma u^{\mathrm{BS}}(\sigma_0)$ can be computed explicitly. Moreover, terms of the form $u_n/\partial_\sigma u^{\mathrm{BS}}(\sigma_0)$ can also be computed explicitly. To see this, we note from Theorems 3.5 and 5.3 that

$$u_n = \mathcal{L}_n(t,T)u_0 = \widetilde{\mathcal{L}}_n(t,T)u^{\mathrm{BS}}(\sigma_0),$$

where

$$\widetilde{\mathcal{L}}_n(t,T) = \sum_{k=1}^{n}\int_t^T \mathrm{d}t_1 \int_{t_1}^T \mathrm{d}t_2 \cdots \int_{t_{k-1}}^T \mathrm{d}t_k \sum_{i\in I_{n,k}} \mathcal{G}_{i_1}(t,t_1)\cdots\mathcal{G}_{i_{k-1}}(t,t_{k-1})\widetilde{\mathcal{G}}_{i_k}(t,t_k),$$

$$\widetilde{\mathcal{G}}_n(t,t_i) := \sum_{k=0}^{n} \left(\mathcal{M}_x(t,t_i)-\bar{x}\right)^{n-k}\left(\mathcal{M}_y(t,t_i)-\bar{y}\right)^{k} a_{n-k,k}(t_i)\beta^2(\partial_z^2-\partial_z)u^{\mathrm{BS}}(\sigma_0).$$

Thus, u_n is a finite sum of the form

$$u_n = \sum_m \mathcal{X}_{n,m}\partial_z^m(\partial_z^2-\partial_z)u^{\mathrm{BS}}(\sigma_0), \quad (5.7)$$

where the coefficients $(\mathcal{X}_{n,m})$ are (t,T,x,y)-dependent constants, which can be computed from Theorem 3.5. Now, using (5.1), a direct computation shows

$$\frac{\partial_z^m(\partial_z^2-\partial_z)u^{\mathrm{BS}}(\sigma_0)}{\partial_\sigma u^{\mathrm{BS}}(\sigma_0)} = \left(\frac{-1}{\sqrt{2\sigma_0^2\tau}}\right)^m \frac{H_n(w)}{\tau\sigma_0}, \qquad w := \frac{z-k-\frac{1}{2}\sigma_0^2\tau}{\sigma\sqrt{2\sigma_0^2\tau}}. \quad (5.8)$$

where $H_n(z) := (-1)^n \mathrm{e}^{z^2}\partial_z^n \mathrm{e}^{-z^2}$ is the n-th Hermite polynomial. Combining (5.7) with (5.8) we have

$$\frac{u_n}{\partial_\sigma u^{\mathrm{BS}}(\sigma_0)} = \sum_m \mathcal{X}_{n,m}\left(\frac{-1}{\sqrt{2\sigma_0^2\tau}}\right)^m \frac{H_n(w)}{\tau\sigma_0}. \quad (5.9)$$

Finally, from (5.6) and (5.9), we see that all terms in the implied volatility expansion (5.5) are polynomials in log-*moneyness* $\lambda := (k-z)$. Explicit expressions for $(\sigma_n)_{n\le 3}$ under different models will be given in Section 6. A general expression for $(\sigma_n)_{n\le 2}$ in the time-homogeneous LSV setting is given below. We denote by

$$\lambda = k - z, \qquad \tau = T - t, \qquad (X_t, Y_t) = (x, y),$$

and we choose the expansion point of out Taylor series approximation as $(\bar{x},\bar{y}) = (x,y)$. We have

$$\sigma_0 = |\beta|\sqrt{2a_{0,0}}, \qquad \sigma_1 = \sigma_{1,0} + \sigma_{0,1}, \qquad \sigma_2 = \sigma_{2,0} + \sigma_{1,1} + \sigma_{0,2},$$

where

$$\sigma_{1,0} = \left(\frac{\beta a_{1,0}}{2\sigma_0}\right)\lambda + \tau\left(\frac{1}{4}(-1+\beta)\sigma_0 a_{1,0}\right),$$

$$
\begin{aligned}
\sigma_{0,1} &= \left(\frac{\beta^3 a_{0,1} f_{0,0}}{2\sigma_0^3}\right)\lambda + \tau\left(\frac{\beta^2 a_{0,1}\left(2c_{0,0}+\beta f_{0,0}\right)}{4\sigma_0}\right),\\
\sigma_{2,0} &= \left(\frac{-3\beta^2 a_{1,0}^2 + 4\sigma_0^2 a_{2,0}}{12\sigma_0^3}\right)\lambda^2 + \tau\left(-\frac{\beta^2 a_{1,0}^2}{8\sigma_0} + \frac{1}{6}\sigma_0 a_{2,0}\right)\\
&\quad + \tau\left(-\frac{(-1+\beta)\left(\beta^2 a_{1,0}^2 - 8\sigma_0^2 a_{2,0}\right)}{24\beta\sigma_0}\right)\lambda\\
&\quad + \tau^2\left(\frac{\beta^2(5+2\beta(-5+2\beta))\sigma_0 a_{1,0}^2 + 8(-1+\beta)^2\sigma_0^3 a_{2,0}}{96\beta^2}\right),\\
\sigma_{1,1} &= \left(\frac{\beta^2\left(2\sigma_0^2 a_{1,1} f_{0,0} + a_{0,1}\left(-5\beta^2 a_{1,0} f_{0,0} + \sigma_0^2 f_{1,0}\right)\right)}{6\sigma_0^5}\right)\lambda^2\\
&\quad + \tau\left(\frac{\beta^2\left(2\sigma_0^2 a_{1,1} f_{0,0} + a_{0,1}\left(\beta^2 a_{1,0} f_{0,0} - 2\sigma_0^2 f_{1,0}\right)\right)}{12\sigma_0^3}\right)\\
&\quad + \tau\left(\frac{\beta\left(4\sigma_0^2 a_{1,1}\left(2c_{0,0}+(-1+2\beta)f_{0,0}\right) + a_{0,1}\left(5\beta^2 a_{1,0}\left(-2c_{0,0}+(1-2\beta)f_{0,0}\right) + 2\sigma_0^2\left(2c_{1,0}+(-1+2\beta)f_{1,0}\right)\right)\right)}{24\sigma_0^3}\right)\lambda\\
&\quad + \tau^2\left(\frac{4(-1+\beta)\sigma_0^2 a_{1,1}\left(2c_{0,0}+\beta f_{0,0}\right) + a_{0,1}\left(\beta^2 a_{1,0}\left(2(-1+\beta)c_{0,0} - \beta f_{0,0}\right) + 2(-1+\beta)\sigma_0^2\left(2c_{1,0}+\beta f_{1,0}\right)\right)}{48\sigma_0}\right),
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{0,2} &= \left(\frac{\beta^4\left(-9\beta^2 a_{0,1}^2 f_{0,0}^2 + 2\sigma_0^2\left(2a_{0,1}^2 b_{0,0} + 2a_{0,2} f_{0,0}^2 + a_{0,1} f_{0,0} f_{0,1}\right)\right)}{12\sigma_0^7}\right)\lambda^2\\
&\quad + \tau\left(\frac{24\beta^2\sigma_0^4 a_{0,2} b_{0,0} + 9\beta^6 a_{0,1}^2 f_{0,0}^2 - 4\beta^4\sigma_0^2\left(2a_{0,1}^2 b_{0,0} + 2a_{0,2} f_{0,0}^2 + a_{0,1} f_{0,0} f_{0,1}\right)}{24\sigma_0^5}\right)\\
&\quad + \tau\left(\frac{\beta^3\left(-9\beta^2 a_{0,1}^2 f_{0,0}\left(2c_{0,0}+\beta f_{0,0}\right) + 8\sigma_0^2 a_{0,2} f_{0,0}\left(2c_{0,0}+\beta f_{0,0}\right) + 4\sigma_0^2 a_{0,1}\left(c_{0,1} f_{0,0} + \left(c_{0,0}+\beta f_{0,0}\right) f_{0,1}\right)\right)}{24\sigma_0^5}\right)\lambda\\
&\quad + \tau^2\left(\frac{\beta^2\left(-3\beta^2 a_{0,1}^2 c_{0,0}\left(c_{0,0}+\beta f_{0,0}\right) + \sigma_0^2\left(-2\beta^2 a_{0,1}^2 b_{0,0} + 2a_{0,2}\left(2c_{0,0}+\beta f_{0,0}\right)^2 + a_{0,1}\left(2c_{0,0}+\beta f_{0,0}\right)\left(2c_{0,1}+\beta f_{0,1}\right)\right)\right)}{24\sigma_0^3}\right).
\end{aligned}
$$

5.1 Implied volatility and log-moneyness scaling

Let us continue to work in the time-homogenous setting. Let $\sigma_Z(\tau,\lambda)$ be the implied volatility of a call written on the LETF Z with time to maturity τ and log-moneyness $\lambda = (k-z)$ and let $\sigma_X(\tau,\lambda)$ be the implied volatility of a call written on the ETF X time to maturity τ and log-moneyness $\lambda = (k-x)$. The expressions above provide an explicit approximation for $\sigma_Z(\tau,\lambda)$ and $\sigma_X(\tau,\lambda)$ in a general time-homogeneous LSV setting (for $\sigma_X(\tau,\lambda)$, simply set $\beta = 1$). These expressions show the highly non-trivial dependence of the implied volatility on the leverage ratio β, and are useful for the purposes of calibration. The implied volatility surfaces $(\tau,\lambda) \mapsto \sigma_Z(\tau,\lambda)$ and $(\tau,\lambda) \mapsto \sigma_X(\tau,\lambda)$ can potentially behave very differently. Nevertheless, for price comparison across leverage ratios, it would practical to relate them, albeit heuristically or approximately. To this end, we now introduce some intuitive scalings. Examining the lowest order terms σ_0 and σ_1 we

observe

$$\text{LETF}: \qquad \sigma_Z \approx |\beta|\sqrt{2a_{0,0}} + |\beta| \left(\frac{a_{1,0}}{2\sqrt{2a_{0,0}}} + \frac{a_{0,1}f_{0,0}}{2(2a_{0,0})^{3/2}} \right) \frac{\lambda}{\beta} + \mathcal{O}(\tau), \tag{5.10}$$

$$\text{ETF}: \qquad \sigma_X \approx \sqrt{2a_{0,0}} + \left(\frac{a_{1,0}}{2\sqrt{2a_{0,0}}} + \frac{a_{0,1}f_{0,0}}{2(2a_{0,0})^{3/2}} \right) \lambda + \mathcal{O}(\tau). \tag{5.11}$$

Comparing σ_Z with σ_X, we see two effects from the leverage ratio β. First, the vertical axis of σ_Z is scaled by a factor of $|\beta|$. Second, the horizontal axis is scaled by a factor of $1/\beta$. In particular, this means that if $\beta < 0$ the slopes of σ_X and σ_Z will have opposite signs. For small τ the contribution of the $\mathcal{O}(\tau)$ terms in the expansion will be insignificant. In light of the above observations, it is natural to introduce $\sigma_X^{(\beta)}$ and $\sigma_Z^{(\beta)}$, the *scaled implied volatilities*, which we define as

$$\sigma_X^{(\beta)}(\tau, \lambda) := |\beta|\sigma_X(\tau, \lambda/\beta), \qquad \sigma_Z^{(\beta)}(\tau, \lambda) := \frac{1}{|\beta|}\sigma_Z(\tau, \beta\,\lambda). \tag{5.12}$$

These definitions offer two ways to link the implied volatilities surfaces σ_X and σ_Z. Viewed one way, the ETF implied volatility $\sigma_X(\tau, \lambda)$ should roughly coincide with the LETF implied volatility $\frac{1}{|\beta|}\sigma_Z(\tau, \beta\lambda)$. Conversely, the LETF implied volatility $\sigma_Z(\tau, \lambda)$ should be close to the ETF implied volatility $|\beta|\sigma_X(\tau, \lambda/\beta)$. In other words, from (5.10), (5.11) and (5.12), we see that for small τ

$$\sigma_Z(\tau, \lambda) \approx \sigma_X^{(\beta)}(\tau, \lambda), \qquad \sigma_X(\tau, \lambda) \approx \sigma_Z^{(\beta)}(\tau, \lambda). \tag{5.13}$$

In Figure 1, using empirical options data from the S&P500-based ETF and LETFs, we plot σ_Z and $\sigma_Z^{(\beta)}$, the unscaled and scaled implied volatilities, respectively. The figure demonstrates the pronounced effect of the scaling argument. Prior to scaling (left panel), the implied volatilities of the LETFs, SSO ($\beta = +2$) and SDS ($\beta = -2$), have much higher values than those of the unleveraged ETF SPY ($\beta = +1$). Moreover, the SDS implied volatility is increasing in log-moneyness. After scaling the LETF implied volatilities according to (5.12) (right panel), they are brought very close to the ETF implied volatility and they are now all downward sloping. In Section 6, we will compute explicit approximations for $\sigma_X(\tau, \lambda)$ and $\sigma_Z^{(\beta)}(\tau, \lambda)$ for three well-known models: CEV, Heston and SABR. As we shall see, although these three models induce distinct implied volatility surfaces, for small τ the role of β in relating σ_X to σ_Z will be captured by (5.13).

We emphasize, however, that the scaling alone is not sufficient to capture the complexity of the LETF implied volatility surface. Indeed, as τ increases, we expect $\sigma_Z^{(\beta)}$ to diverge from σ_X. This discrepancy is due to the integrated variance contribution to the terminal value of Z, as can be seen from (2.3). Thus, for longer maturities, an accurate approximation of the LETF implied volatility surface must include higher terms in τ. From the general implied volatility expression, we can see that the role of β in the $\mathcal{O}(\tau)$ terms is complicated and does not lend itself to a simple scaling argument. For this reason the full implied volatility expansion – not just the scaling argument – is important.

Remark 5.4. A recent paper Leung and Sircar (2012) postulates an alternative implied volatility scaling based on stochastic arguments. Given the terminal ETF value $X_\tau = k$, they compute the expected future log-moneyness $Z_\tau - z$

$$\mathbb{E}_{x,y,z}[Z_T - z | X_T = k] = \beta(k - x) - \frac{1}{2}\beta(\beta - 1) \int_0^\tau \mathbb{E}_{x,y,z}[\sigma^2(s, X_s, Y_s) | X_\tau = k]\, \mathrm{d}s, \tag{5.14}$$

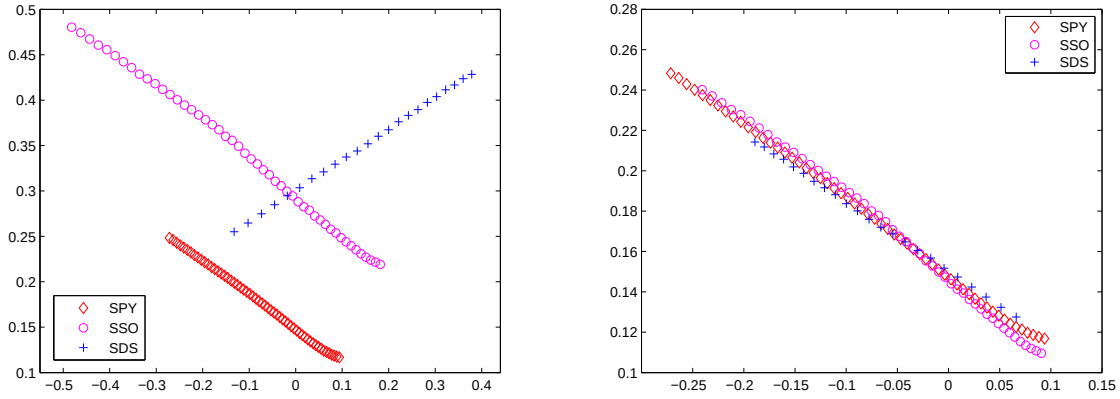

Figure 1: Left: Empirical implied volatilities $\sigma_Z(\tau,\lambda)$ plotted as a function of log-moneyness λ for SPY (red diamonds, $\beta=+1$), SSO (purple circles, $\beta=+2$), and SDS (blue crosses, $\beta=-2$) on August 15, 2013 with $\tau=155$ days to maturity. Note that the implied volatility of SDS is increasing in the LETF log-moneyness. Right: Using the same data, the scaled LETF implied volatilities $\sigma_Z^{(\beta)}(\tau,\lambda)$ nearly coincide.

where $\mathbb{E}_{x,y,z}[\cdot]=\mathbb{E}[\cdot|X_0=x,Y_0=y,Z_0=z]$. The also note, from the ETF and LETF SDEs that the volatility of Z is $|\beta|$ times the volatility of X. Using the above as heuristic, the authors propose to scale implied volatilities as follows

$$\sigma_Z(\tau,\lambda)=|\beta|\sigma_X(\tau,\beta\lambda-\frac{1}{2}\beta(\beta-1)I(\tau)),$$
$$I(\tau)=\int_0^\tau \mathbb{E}_{x,y,z}[\sigma^2(s,X_s,Y_s)|X_\tau=k]\,\mathrm{d}s.$$

In Leung and Sircar (2012), the value of $I(\tau)$ is estimated using an average from observed implied volatility. In contrast, the scaling proposed in (5.12) does not attempt to account for the integral in (5.14). Nevertheless, the effect of the integrated variance is captured by $\mathcal{O}(\tau)$ terms in the general implied volatility expansion.

6 Examples

In this Section, we provide explicit expressions for implied volatilities under three different model dynamics: CEV, Heston and SABR. Special attention will be paid to the role of β, the leverage ratio. In the examples below, we fix $(\bar{x},\bar{y})=(X_0,Y_0)$ and we evaluate implied volatilities at time $t=0$ and maturing at time $T=\tau$.

6.1 CEV

In the Constant Elasticity of Variance (CEV) local volatility model of Cox (1975), the dynamics of the underlying S are given by

$$\mathrm{d}S_t=\delta S_t^{\gamma-1}S_t\mathrm{d}W_t^x, \qquad S_0>0,$$

where, to preserve the martingale property of the process S (cf. Heston et al. (2007)), the parameter γ is assumed to be less than or equal to 1. The dynamics of $(X, Z) = (\log S, \log L)$ are

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2}\delta^2 \mathrm{e}^{2(\gamma-1)X_t}\mathrm{d}t + \delta\, \mathrm{e}^{(\gamma-1)X_t}\mathrm{d}W_t^x, & X_0 &= x := \log S_0. \\ \mathrm{d}Z_t &= -\frac{1}{2}\beta^2\delta^2 \mathrm{e}^{2(\gamma-1)X_t}\mathrm{d}t + \beta\delta\, \mathrm{e}^{(\gamma-1)X_t}\mathrm{d}W_t^x, & Z_0 &= z := \log L_0. \end{aligned}$$

The generator of (X, Z) is given by

$$\mathcal{A} = \frac{1}{2}\delta^2 \mathrm{e}^{2(\gamma-1)x}\left((\partial_x^2 - \partial_x) + \beta^2(\partial_z^2 - \partial_z) + 2\beta\partial_x\partial_z\right).$$

Thus, from (3.2), we identify

$$a(x,y) = \frac{1}{2}\delta^2 \mathrm{e}^{2(\gamma-1)x}, \qquad b(x,y) = 0, \qquad c(x,y) = 0, \qquad f(x,y) = 0. \tag{6.1}$$

Using equations (5.5), (5.6) and (5.9) we compute

$$\begin{aligned} \sigma_0 &= |\beta|\sqrt{\mathrm{e}^{2x(\gamma-1)}\delta^2}, \\ \sigma_1 &= \frac{\tau(\beta-1)(\gamma-1)\sigma_0^3}{4\beta^2} + \frac{(\gamma-1)\sigma_0}{2\beta}(k-z), \\ \sigma_2 &= \frac{\tau(\gamma-1)^2\sigma_0^3\left(4\beta^2 + t(13+2\beta(-13+6\beta))\sigma_0^2\right)}{96\beta^4} \\ &\quad + \frac{7\tau(\beta-1)(\gamma-1)^2\sigma_0^3}{24\beta^3}(k-z) + \frac{(\gamma-1)^2\sigma_0}{12\beta^2}(k-z)^2, \\ \sigma_3 &= \frac{\tau^2(\beta-1)(\gamma-1)^3\sigma_0^5\left(60\beta^2 + t\left(35-70\beta+26\beta^2\right)\sigma_0^2\right)}{384\beta^6} \\ &\quad + \frac{\tau(\gamma-1)^3\sigma_0^3\left(12\beta^2 + 5t(9+2\beta(-9+4\beta))\sigma_0^2\right)}{192\beta^5}(k-z) + \frac{7\tau(\beta-1)(\gamma-1)^3\sigma_0^3}{48\beta^4}(k-z)^2. \end{aligned}$$

We observe that the factor $(\gamma - 1)$ appears in every term of these expressions. In particular, when $\gamma = 1$, $\sigma_0 = |\beta|\delta$ and $\sigma_1 = \sigma_2 = \sigma_3 = 0$. The higher order terms also vanish since $a(x, y) = \frac{1}{2}\delta^2$ in this case (see (6.1)). Hence, just as in the Black-Scholes case, the implied volatility expansion becomes flat, as expected

In Figure 2 we plot our third order approximation of the scaled implied volatility $\sigma_Z^{(\beta)}(\tau, \lambda)$ in the CEV model with leverages $\beta = \{+2, -2\}$ and with maturities $\tau = \{0.0625, 0.125, 0.25, 0.5\}$ years. For comparison, we also plot the exact scaled implied volatility $\sigma_Z^{(\beta)}(\tau, \lambda)$ and the exact implied volatility of the ETF $\sigma_X(\tau, \lambda)$. The exact scaled implied volatility $\sigma_Z^{(\beta)}$ of the LETF is computed by obtaining call prices by Monte Carlo simulation and then my inverting the Black-Scholes formula numerically. The exact implied volatility $\sigma_X(\tau, \lambda)$ of the ETF is computed using the exact call price formula, available in Cox (1975), and then inverting the Black-Scholes formula numerically.

6.2 Heston

In the Heston model, due to Heston (1993), the dynamics of the underlying S are given by

$$\mathrm{d}S_t = \sqrt{Z_t}S_t\mathrm{d}W_t^x, \qquad S_0 > 0,$$

$$
\begin{aligned}
\mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \delta\sqrt{V_t}\mathrm{d}W_t^y, && V_0 > 0,\\
\mathrm{d}\langle W^x, W^y\rangle_t &= \rho\,\mathrm{d}t.
\end{aligned}
$$

In log notation $(X, Y, Z) := (\log S, \log V, \log L)$ we have the following dynamics

$$
\begin{aligned}
\mathrm{d}X_t &= -\frac{1}{2}\mathrm{e}^{Y_t}\mathrm{d}t + \mathrm{e}^{\frac{1}{2}Y_t}\mathrm{d}W_t^x, && X_0 = x := \log S_0,\\
\mathrm{d}Y_t &= \left((\kappa\theta - \tfrac{1}{2}\delta^2)\mathrm{e}^{-Y_t} - \kappa\right)\mathrm{d}t + \delta\,\mathrm{e}^{-\frac{1}{2}Y_t}\mathrm{d}W_t^y, && Y_0 = y := \log V_0,\\
\mathrm{d}Z_t &= -\beta^2\frac{1}{2}\mathrm{e}^{Y_t}\mathrm{d}t + \beta\mathrm{e}^{\frac{1}{2}Y_t}\mathrm{d}W_t^x, && Z_0 = z := \log L_0,\\
\mathrm{d}\langle W^x, W^y\rangle_t &= \rho\,\mathrm{d}t.
\end{aligned}
$$

The generator of (X, Y, Z) is given by

$$
\begin{aligned}
\mathcal{A} = \frac{1}{2}\mathrm{e}^y &\left((\partial_x^2 - \partial_x) + \beta^2(\partial_z^2 - \partial_z) + 2\beta\partial_x\partial_z\right)\\
&+ \left((\kappa\theta - \tfrac{1}{2}\delta^2)\mathrm{e}^{-y} - \kappa\right)\partial_y + \frac{1}{2}\delta^2\mathrm{e}^{-y}\partial_y^2 + \rho\,\delta\left(\partial_x\partial_y + \beta\partial_x\partial_z\right).
\end{aligned}
$$

Thus, from (3.2), we identify

$$
a(x,y) = \frac{1}{2}\mathrm{e}^y, \qquad b(x,y) = \frac{1}{2}\delta^2\mathrm{e}^{-y}, \qquad c(x,y) = \left((\kappa\theta - \tfrac{1}{2}\delta^2)\mathrm{e}^{-y} - \kappa\right), \qquad f(x,y) = \rho\,\delta.
$$

Using equations (5.5), (5.6) and (5.9) we obtain

$$
\sigma_0 = |\beta|\sqrt{\mathrm{e}^y}, \tag{6.2}
$$

$$
\begin{aligned}
\sigma_1 &= \frac{\tau\left(-\beta^2\left(\delta^2 - 2\theta\kappa\right) + (-2\kappa + \beta\delta\rho)\sigma_0^2\right)}{8\sigma_0} + \frac{\beta\delta\rho}{4\sigma_0}(k - z),\\
\sigma_2 &= -\frac{\tau^2\beta^4\left(\delta^2 - 2\theta\kappa\right)^2}{128\sigma_0^3} + \frac{\tau\beta^2\left(-4\tau\theta\kappa^2 - \tau\beta\delta^3\rho + 2\tau\beta\delta\theta\kappa\rho + 2\delta^2\left(8 + t\kappa + \rho^2\right)\right)}{192\sigma_0^3}\\
&\quad + \frac{\tau^2\left(5\kappa^2 + \beta\delta\left(-5\kappa\rho + \beta\delta\left(-1 + 2\rho^2\right)\right)\right)}{96\sigma_0^3}\\
&\quad + \frac{\tau\beta\delta\rho\left(5\beta^2\left(\delta^2 - 2\theta\kappa\right) + (2\kappa - \beta\delta\rho)\sigma_0^2\right)}{96\sigma_0^3}(k - z) + \frac{\beta^2\delta^2\left(2 - 5\rho^2\right)}{48\sigma_0^3}(k - z)^2,
\end{aligned} \tag{6.3}
$$

$$
\begin{aligned}
\sigma_3 &= -\frac{\tau^3\beta^6\left(\delta^2 - 2\theta\kappa\right)^3}{1024\sigma_0^5} + \frac{\tau^2\beta^4\left(\delta^2 - 2\theta\kappa\right)\left(4\tau\theta\kappa^2 + \tau\beta\delta^3\rho - 2\tau\beta\delta\theta\kappa\rho + \delta^2\left(16 - 2\tau\kappa + 20\rho^2\right)\right)}{3072\sigma_0^5}\\
&\quad + \frac{\tau^2\beta^2\left(3\delta^2\rho^2(-2\kappa + \beta\delta\rho) + \tau\kappa\left(\delta^2 - 2\theta\kappa\right)(-\kappa + \beta\delta\rho)\right)}{768\sigma_0^5}\\
&\quad + \frac{\tau^3(-2\kappa + \beta\delta\rho)\left(6\kappa^2 + \beta\delta\left(-6\kappa\rho + \beta\delta\left(-6 + 5\rho^2\right)\right)\right)}{1536\sigma_0^5}\\
&\quad + \Bigg(\frac{7\tau^2\beta^5\delta\left(\delta^2 - 2\theta\kappa\right)^2\rho}{512\sigma_0^5} + \frac{\tau^2\beta\delta\rho\left(-3\kappa^2 + \beta\delta\left(3\kappa\rho + \beta\left(\delta - 2\delta\rho^2\right)\right)\right)}{384\sigma_0^5}\\
&\qquad - \frac{\tau\beta^3\delta\rho\left(20\tau\theta\kappa^2 + 5\tau\beta\delta^3\rho - 10\tau\beta\delta\theta\kappa\rho + 2\delta^2\left(8 - 5\tau\kappa + 9\rho^2\right)\right)}{768\sigma_0^5}\Bigg)(k - z)\\
&\quad - \frac{\tau\beta^2\delta^2\left(\beta^2\left(\delta^2 - 2\theta\kappa\right)\left(-8 + 23\rho^2\right) + (2\kappa - \beta\delta\rho)\left(-2 + 7\rho^2\right)\sigma_0^2\right)}{384\sigma_0^5}(k - z)^2
\end{aligned}
$$

$$+\frac{\beta^3\delta^3\rho\left(-5+8\rho^2\right)}{96\sigma_0^5}(k-z)^3.$$

Ahn et al. (2012) noticed from the SDEs that when X has Heston dynamics with parameters (κ, θ, δ, ρ, y), then Z has Heston dynamics with parameters

$$(\kappa_Z,\theta_Z,\delta_Z,\rho_Z,y_Z)=(\kappa,\beta^2\theta,|\beta|\delta,\mathrm{sign}(\beta)\rho,y+\log\beta^2). \tag{6.4}$$

The characteristic function of X_τ is computed explicitly in Heston (1993) and Bakshi, Cao, and Chen (1997)

$$\begin{aligned}
\eta_X(\tau,x,y,\xi)&:=\log\mathbb{E}[\mathrm{e}^{\mathtt{i}\xi X_\tau}|X_0=x,Y_0=y]=\mathtt{i}\xi x+C(\tau,\xi)+D(\tau,\xi)\mathrm{e}^y,\\
C(\tau,\xi)&=\frac{\kappa\theta}{\delta^2}\left((\kappa-\rho\delta\mathtt{i}\xi+d(\xi))\tau-2\log\left[\frac{1-f(\xi)\mathrm{e}^{d(\xi)\tau}}{1-f(\xi)}\right]\right),\\
D(\tau,\xi)&=\frac{\kappa-\rho\delta\mathtt{i}\xi+d(\xi)}{\delta^2}\frac{1-\mathrm{e}^{d(\xi)\tau}}{1-f(\xi)\mathrm{e}^{d(\xi)\tau}},\\
f(\xi)&=\frac{\kappa-\rho\delta\mathtt{i}\xi+d(\xi)}{\kappa-\rho\delta\mathtt{i}\xi-d(\xi)},\\
d(\xi)&=\sqrt{\delta^2(\xi^2+\mathtt{i}\xi)+(\kappa-\rho\mathtt{i}\xi\delta)^2}.
\end{aligned}$$

Since Z also has Heston dynamics, the characteristic function of Z_t follows directly

$$\eta_Z(\tau,z,y,\xi):=\log\mathbb{E}[\mathrm{e}^{\mathtt{i}\xi Z_\tau}|Y_0=y,Z_0=z]=\eta_X(\tau,z,y,\xi)\quad\text{ with }\quad(\kappa,\theta,\delta,\rho,y)\to(\kappa_Z,\theta_Z,\delta_Z,\rho_Z,y_Z).$$

The price of a European call option with payoff $\varphi(z)=(\mathrm{e}^z-\mathrm{e}^k)^+$ can then be computed using standard Fourier methods

$$u^{\mathrm{Hes}}(\tau,z,y)=\frac{1}{2\pi}\int_{\mathbb{R}}\mathrm{d}\xi_r\,\mathrm{e}^{\eta_Z(\tau,z,y,\xi)}\widehat{\varphi}(\xi),\qquad\widehat{\varphi}(\xi)=\frac{-\mathrm{e}^{k-\mathtt{i}k\xi}}{\mathtt{i}\xi+\xi^2},\qquad\xi=\xi_r+\mathtt{i}\xi_i,\qquad\xi_i<-1. \tag{6.5}$$

Note, since the call option payoff $\varphi(z)=(\mathrm{e}^z-\mathrm{e}^k)^+$ is not in $L^1(\mathbb{R})$, its Fourier transform $\widehat{h}(\xi)$ must be computed in a generalized sense by fixing an imaginary component of the Fourier variable $\xi_i<-1$. Using (6.5) the exact implied volatility σ can be computed to solving (5.2) numerically.

Moreover, it is worth noting that relationship (6.4) can be inferred from our implied volatility expressions. Indeed, the dependence on β in expansions (6.2)-(6.3) is present *only* in the terms $\beta^2\theta,|\beta|\delta,\mathrm{sign}(\beta)\rho,y+\log\beta^2$ of the coefficients. For instance, we can write the zeroth order term $\sigma_0=\sqrt{e^{y+\log\beta^2}}=\sqrt{e^{y_Z}}$, and the coefficient of $(k-z)$ in σ_1 is $\beta\delta\rho/4\sigma_0=|\beta|\delta\mathrm{sign}(\beta)\rho/4\sigma_0=\delta_Z\rho_Z/4\sigma_0$, as per the notations in (6.4). Similar verification procedures for other terms confirm relationship (6.4).

In Figure 3 we plot our third order approximation of the scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ in the Heston model with leverages $\beta=\{+2,-2\}$ and with maturities $\tau=\{0.0625,0.125,0.25,0.5\}$ years. For comparison, we also plot the exact scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ and the exact implied volatility of the ETF $\sigma_X(\tau,\lambda)$. The exact scaled implied volatility $\sigma_Z^{(\beta)}$ of the LETF is computed by obtaining call prices from (6.5) and then my inverting the Black-Scholes formula numerically. The exact implied volatility $\sigma_X(\tau,\lambda)$ of the ETF is computed in the same manner.

6.3 SABR

The SABR model of Hagan, Kumar, Lesniewski, and Woodward (2002) is a local-stochastic volatility model in which the risk-neutral dynamics of S are given by

$$\begin{aligned} \mathrm{d}S_t &= V_t S_t^{\gamma-1} S_t \mathrm{d}W_t^x, & S_0 &> 0, \\ \mathrm{d}V_t &= \delta V_t \mathrm{d}W_t^y, & V_0 &> 0, \\ \mathrm{d}\langle W^x, W^z\rangle_t &= \rho\,\mathrm{d}t. \end{aligned}$$

In log notation $(X, Y, Z) := (\log S, \log V, \log L)$ we have, we have the following dynamics:

$$\begin{aligned} \mathrm{d}X_t &= -\frac{1}{2}\mathrm{e}^{2Y_t+2(\gamma-1)X_t}\mathrm{d}t + \mathrm{e}^{Y_t+(\gamma-1)X_t}\mathrm{d}W_t^x, & X_0 &= x := \log S_0, \\ \mathrm{d}Y_t &= -\frac{1}{2}\delta^2\mathrm{d}t + \delta\,\mathrm{d}W_t^y, & Y_0 &= y := \log V_0, \\ \mathrm{d}Z_t &= -\frac{1}{2}\beta^2\mathrm{e}^{2Y_t+2(\gamma-1)X_t}\mathrm{d}t + \beta\mathrm{e}^{Y_t+(\gamma-1)X_t}\mathrm{d}W_t^x, & Z_0 &= z := \log L_0, \\ \mathrm{d}\langle W^x, W^y\rangle_t &= \rho\,\mathrm{d}t. \end{aligned}$$

The generator of (X, Y, Z) is given by

$$\begin{aligned} \mathcal{A} = \frac{1}{2}\mathrm{e}^{2y+2(\gamma-1)x}&\left((\partial_x^2 - \partial_x) + \beta^2(\partial_z^2 - \partial_z) + 2\beta\partial_x\partial_y\right) \\ &- \frac{1}{2}\delta^2\partial_y + \frac{1}{2}\delta^2\partial_y^2 + \rho\,\delta\,\mathrm{e}^{y+(\gamma-1)x}(\partial_x\partial_y + \beta\partial_y\partial_z). \end{aligned}$$

Thus, using (3.2), we identify

$$a(x,y) = \frac{1}{2}\mathrm{e}^{2y+2(\gamma-1)x}, \qquad b(x,y) = \frac{1}{2}\delta^2, \qquad c(x,y) = -\frac{1}{2}\delta^2, \qquad f(x,y) = \rho\,\delta\,\mathrm{e}^{y+(\gamma-1)x}.$$

Using equations (5.5), (5.6) and (5.9) we compute

$$\begin{aligned} \sigma_0 &= |\beta|\sqrt{\mathrm{e}^{2y+2x(-1+\gamma)}}, & \sigma_1 &= \sigma_{1,0} + \sigma_{0,1}, \\ \sigma_2 &= \sigma_{2,0} + \sigma_{1,1} + \sigma_{0,2}, & \sigma_3 &= \sigma_{3,0} + \sigma_{2,1} + \sigma_{1,2} + \sigma_{0,3}, \end{aligned}$$

where

$$\begin{aligned} \sigma_{1,0} &= \frac{\tau(-1+\beta)(\gamma-1)\sigma_0^3}{4\beta^2} + \frac{(\gamma-1)\sigma_0}{2\beta}(k-z), \\ \sigma_{0,1} &= -\frac{1}{4}\tau\delta\sigma_0\left(\delta - \rho\,\mathrm{sign}[\beta]\sigma_0\right) + \frac{1}{2}\delta\rho\,\mathrm{sign}[\beta](k-z), \\ \sigma_{2,0} &= \frac{\tau(\gamma-1)^2\sigma_0^3\left(4\beta^2 + \tau(13 + 2\beta(-13+6\beta))\sigma_0^2\right)}{96\beta^4} \\ &\quad + \frac{7\tau(-1+\beta)(\gamma-1)^2\sigma_0^3}{24\beta^3}(k-z) + \frac{(\gamma-1)^2\sigma_0}{12\beta^2}(k-z)^2, \\ \sigma_{1,1} &= \frac{\tau(\gamma-1)\delta\sigma_0^2\left(12\rho|\beta| + \tau\sigma_0\left(-9(-1+\beta)\delta + (-11+10\beta)\rho\,\mathrm{sign}[\beta]\sigma_0\right)\right)}{48\beta^2} \\ &\quad - \frac{\tau(\gamma-1)\delta\sigma_0\left(3\delta + \frac{5(1-2\beta)\rho\sigma_0}{|\beta|}\right)}{24\beta}(k-z), \end{aligned}$$

$$
\begin{aligned}
\sigma_{0,2} &= \frac{1}{96}\tau\delta^2\sigma_0\left(32+5\tau\delta^2-12\rho^2+2\tau\sigma_0\left(-7\delta\rho\,\mathrm{sign}[\beta]+\left(-2+6\rho^2\right)\sigma_0\right)\right)\\
&\quad-\frac{1}{24}\tau\delta^2\rho\left(\delta\,\mathrm{sign}[\beta]-3\rho\sigma_0\right)+\frac{\delta^2\left(2-3\rho^2\right)}{12\sigma_0}(k-z)^2,\\
\sigma_{3,0} &= \frac{\tau^2(-1+\beta)(\gamma-1)^3\sigma_0^5\left(60\beta^2+\tau\left(35-70\beta+26\beta^2\right)\sigma_0^2\right)}{384\beta^6}\\
&\quad+\frac{\tau(\gamma-1)^3\sigma_0^3\left(12\beta^2+5\tau(9+2\beta(-9+4\beta))\sigma_0^2\right)}{192\beta^5}(k-z)+\frac{7\tau(-1+\beta)(\gamma-1)^3\sigma_0^3}{48\beta^4}(k-z)^2,\\
\sigma_{2,1} &= -\sigma_0^3\frac{\tau^2(\gamma-1)^2\delta^2}{32\beta^2}+\sigma_0^4\frac{\tau^2(-46+53\beta)(\gamma-1)^2\delta\rho|\beta|}{96\beta^4}\\
&\quad-\sigma_0^5\frac{5\tau^3(13+2\beta(-13+6\beta))(\gamma-1)^2\delta^2}{384\beta^4}+\sigma_0^6\frac{\tau^3(83+2\beta(-74+29\beta))(\gamma-1)^2\delta\rho|\beta|}{384\beta^5}\\
&\quad+\frac{\tau(\gamma-1)^2\delta\sigma_0^2\left(52\rho|\beta|+\tau\sigma_0\left(-42(-1+\beta)\delta+\frac{(31+4\beta(-31+20\beta))\rho|\beta|\sigma_0}{\beta^2}\right)\right)}{192\beta^3}(k-z)\\
&\quad-\frac{\tau(\gamma-1)^2\delta\sigma_0\left(\delta+\frac{(8-11\beta)\rho|\beta|\sigma_0}{\beta^2}\right)}{48\beta^2}(k-z)^2,\\
\sigma_{1,2} &= -\sigma_0^2\frac{7\tau^2(\gamma-1)\delta^3\rho|\beta|}{48\beta^2}+\sigma_0^3\left(\frac{\tau^2(-1+\beta)(\gamma-1)\delta^2\left(64+11\tau\delta^2\right)}{128\beta^2}+\frac{\tau^2(19+7\beta)(\gamma-1)\delta^2\rho^2}{96\beta^2}\right)\\
&\quad+\sigma_0^4\frac{\tau^3(47-42\beta)(\gamma-1)\delta^3\rho|\beta|}{192\beta^3}+\sigma_0^5\frac{\tau^3(\gamma-1)\delta^2\left(-14(-1+\beta)+(-41+34\beta)\rho^2\right)}{192\beta^2}\\
&\quad\left(\sigma_0\frac{\tau(-1+\gamma)\delta^2\left(32+5\tau\delta^2+12\rho^2\right)}{192\beta}+\sigma_0^2\frac{11\tau^2(1-2\beta)(-1+\gamma)\delta^3\rho\mathrm{Abs}[\beta]}{96\beta^3}\right)(k-z)\\
&\quad+\sigma_0^3\frac{\tau^2(-1+\gamma)\delta^2\left(-6\beta+(-19+29\beta)\rho^2\right)}{96\beta^2}(k-z)\\
&\quad-\frac{\tau(\gamma-1)\delta^2\left(3-3\beta+(-2+\beta)\rho^2\right)\sigma_0}{24\beta^2}(k-z)^2+\frac{(\gamma-1)\delta^2\left(-2+3\rho^2\right)}{24\beta\sigma_0}(k-z)^3,\\
\sigma_{0,3} &= -\sigma_0\frac{1}{128}\tau^2\delta^4\left(16+\tau\delta^2-4\rho^2\right)+\sigma_0^2\frac{\tau^2\delta^3\rho\left(104+19\tau\delta^2-36\rho^2\right)|\beta|}{384\beta}\\
&\quad\sigma_0^3\frac{1}{192}\tau^3\delta^4\left(8-21\rho^2\right)+\sigma_0^4\frac{\tau^3\delta^3\rho\left(-11+15\rho^2\right)|\beta|}{192\beta}\\
&\quad-\frac{\tau\delta^3\rho\left(8+\tau\delta^2-12\rho^2+6\tau\sigma_0\left(\frac{\delta\rho|\beta|}{\beta}+\left(1-2\rho^2\right)\sigma_0\right)\right)}{192\sqrt{\frac{1}{\beta^2}}\beta}(k-z)\\
&\quad-\frac{\tau\delta^3\rho\left(-1+\rho^2\right)|\beta|}{16\beta}(k-z)^2+\frac{\delta^3\rho\left(-5+6\rho^2\right)|\beta|}{24\beta\sigma_0^2}(k-z)^3.
\end{aligned}
$$

In Figure 4 we plot our third order approximation of the scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ in the SABR model with leverages $\beta=\{+2,-2\}$ and with maturities $\tau=\{0.0625,0.125,0.25,0.5\}$ years. For comparison, we also plot the exact scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ and the exact implied volatility of the ETF $\sigma_X(\tau,\lambda)$. The exact scaled implied volatility $\sigma_Z^{(\beta)}$ of th LETF is computed by obtaining call prices by Monte Carlo simulation and then my inverting the Black-Scholes formula numerically. The exact implied volatility $\sigma_X(\tau,\lambda)$ of the ETF is computed using the exact call price formula, available in Antonov and Spector (2012) for the special

case $\rho = 0$, and then inverting the Black-Scholes formula numerically.

7 Conclusion

In this article, starting from ETF dynamics in a general time-inhomogeneous LSV setting, we derive approximate European-style option prices written on the associated LETFs. The option price approximation requires only a normal CDF to compute. Therefore, computational times for prices are comparable to Black-Scholes. We also establish rigorous error bounds for our pricing approximation. These error bounds are established through a regularization procedure, which allows us to overcome challenges that arise when dealing with a generator $\mathcal{A}(t)$ that is *not* elliptic.

Additionally, we derive an implied volatility expansion that is fully explicit – polynomial in log-moneyness $\lambda = (k - z)$ and (for time-homogeneous models) polynomial in time to maturity. To aid in the analysis of the implied volatility surface, we discuss some natural scalings of implied volatility. Furthermore, we test our implied volatility expansion on three well-known LSV models (CEV, Heston and SABR) and find that the expansion provides an excellent approximation of the true implied volatility.

The markets for leveraged ETFs and their options continue to grow, not only in equities, but also in other sectors such as commodity, fixed-income, and currency. The question of consistent pricing, as we have investigated for equity LETF options in terms implied volatility, is also relevant to LETF options in other sectors. Naturally, the valuation of LETF options will depend on the dynamics of the LETFs and underlying price process, which may vary significantly across sectors (see e.g. Guo and Leung (2014) for commodity LETFs). Nevertheless, it is both practically and mathematically interesting to adapt the techniques in the current paper to investigate the implied volatilities across leverage ratios with different underlyings. From a market stability perspective, it is important for both investors and regulators to understand the risks and dependence structure among ETFs and the price relationships of their traded derivatives.

Acknowledgements The authors are grateful to Peter Carr, Emanuel Derman, Sebastian Jaimungal, and Ronnie Sircar for their helpful discussions, and we thank the seminar participants at Columbia University and Fields Institute for their comments.

References

Ahn, A., M. Haugh, and A. Jain (2012). Consistent pricing of options on leveraged ETFs. Working paper, Columbia University.

Antonov, A. and M. Spector (2012, march). Advanced analytics for the sabr model. *SSRN*.

Avellaneda, M. and S. Zhang (2010). Path-dependence of leveraged ETF returns. *SIAM Journal of Financial Mathematics 1*, 586–603.

Bakshi, G., C. Cao, and Z. Chen (1997, December). Empirical performance of alternative option pricing models. *Journal of Finance 52*(5), 2003–49.

Benhamou, E., E. Gobet, and M. Miri (2010). Time dependent heston model. *SIAM Journal on Financial Mathematics 1*(1), 289–325.

Bompis, R. and E. Gobet (2013). Asymptotic and non asymptotic approximations for option valuation. In *Recent developments in computational finance. Foundations, algorithms and applications.*, pp. 159–241. Hackensack, NJ: World Scientific.

Cheng, M. and A. Madhavan (2009). The dynamics of leveraged and inverse exchange traded funds. *Journal Of Investment Management 4*.

Cox, J. (1975). Notes on option pricing I: Constant elasticity of diffusions. *Unpublished draft, Stanford University*. A revised version of the paper was published by the Journal of Portfolio Management in 1996.

Di Francesco, M. and A. Pascucci (2005). On a class of degenerate parabolic equations of Kolmogorov type. *AMRX Appl. Math. Res. Express 3*, 77–116.

Forde, M. and A. Jacquier (2011). Small-time asymptotics for an uncorrelated local-stochastic volatility model. *Applied Mathematical Finance 18*(6), 517–535.

Fouque, J.-P., G. Papanicolaou, R. Sircar, and K. Solna (2011). *Multiscale stochastic volatility for equity, interest rate, and credit derivatives*. Cambridge: Cambridge University Press.

Friedman, A. (1964). *Partial differential equations of parabolic type*. Englewood Cliffs, N.J.: Prentice-Hall Inc.

Gatheral, J., E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang (2012). Asymptotics of implied volatility in local volatility models. *Math. Finance 22*(4), 591–620.

Guo, K. and T. Leung (2014). Understanding the tracking errors of commodity leveraged ETFs. Working paper, Columbia University.

Hagan, P., D. Kumar, A. Lesniewski, and D. Woodward (2002). Managing smile risk. *Wilmott Magazine 1000*, 84–108.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud. 6*(2), 327–343.

Heston, S., M. Loewenstein, and G. A. Willard (2007). Options and Bubbles. *Rev. Financ. Stud. 20*(2), 359–390.

Ikeda, N. and S. Watanabe (1989). *Stochastic differential equations and diffusion processes* (Second ed.), Volume 24 of *North-Holland Mathematical Library*. Amsterdam: North-Holland Publishing Co.

Jacquier, A. and M. Lorig (2013). The smile of certain Lévy-type models. *SIAM Journal on Financial Mathematics 4*.

Leung, T. and R. Sircar (2012). Implied volatility of leveraged ETF options. Working paper, Columbia University.

Lorig, M., S. Pagliarani, and A. Pascucci (2013a). Analytical expansions for parabolic equations. *ArXiv preprint arXiv:1312.3314*.

Lorig, M., S. Pagliarani, and A. Pascucci (2013b). Implied vol for any local-stochastic vol model. *ArXiv preprint arXiv:1306.5447*.

Lorig, M., S. Pagliarani, and A. Pascucci (2014). A family of density expansions for Lévy-type processes with default. *To appear in: Annals of Applied Probability*.

Pascucci, A. (2011). *PDE and martingale methods in option pricing*, Volume 2 of *Bocconi & Springer Series*. Milan: Springer.

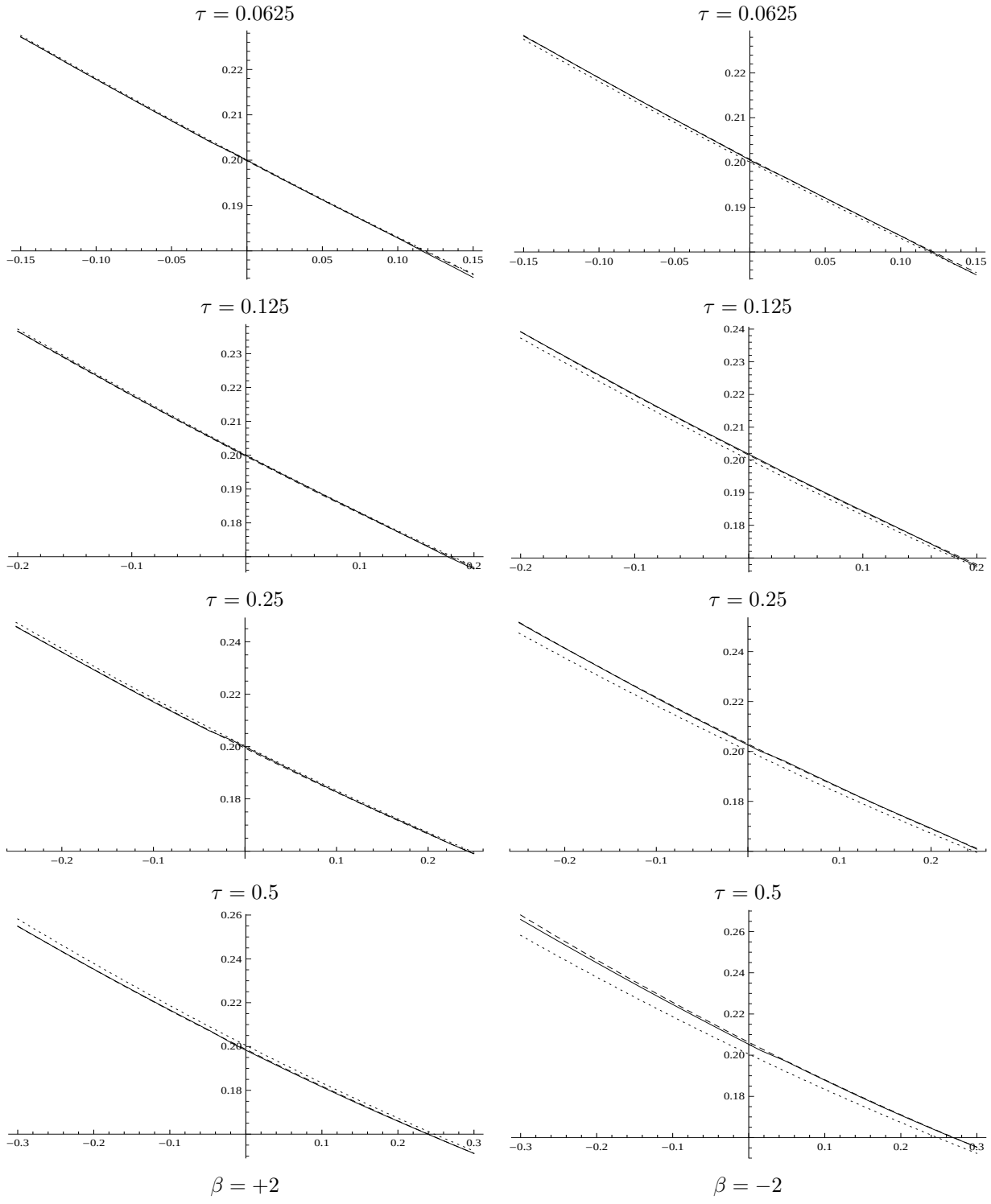

Figure 2: Exact (solid – computed by Monte Carlo) and approximate (dashed) scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ under CEV model dynamics plotted as a function of log-moneyness λ. For comparison, we also plot the exact implied volatility of the CEV model $\sigma_Z^{(1)}(\tau,\lambda) = \sigma_X(\tau,\lambda)$ (dotted). Parameters: $\delta = 0.2$, $\gamma = -0.75$, $x = 0$. For each leverage ratio ($\beta = \pm 2$), as τ increases, the solid and dotted lines diverge, while the dashed and solid lines remain close.

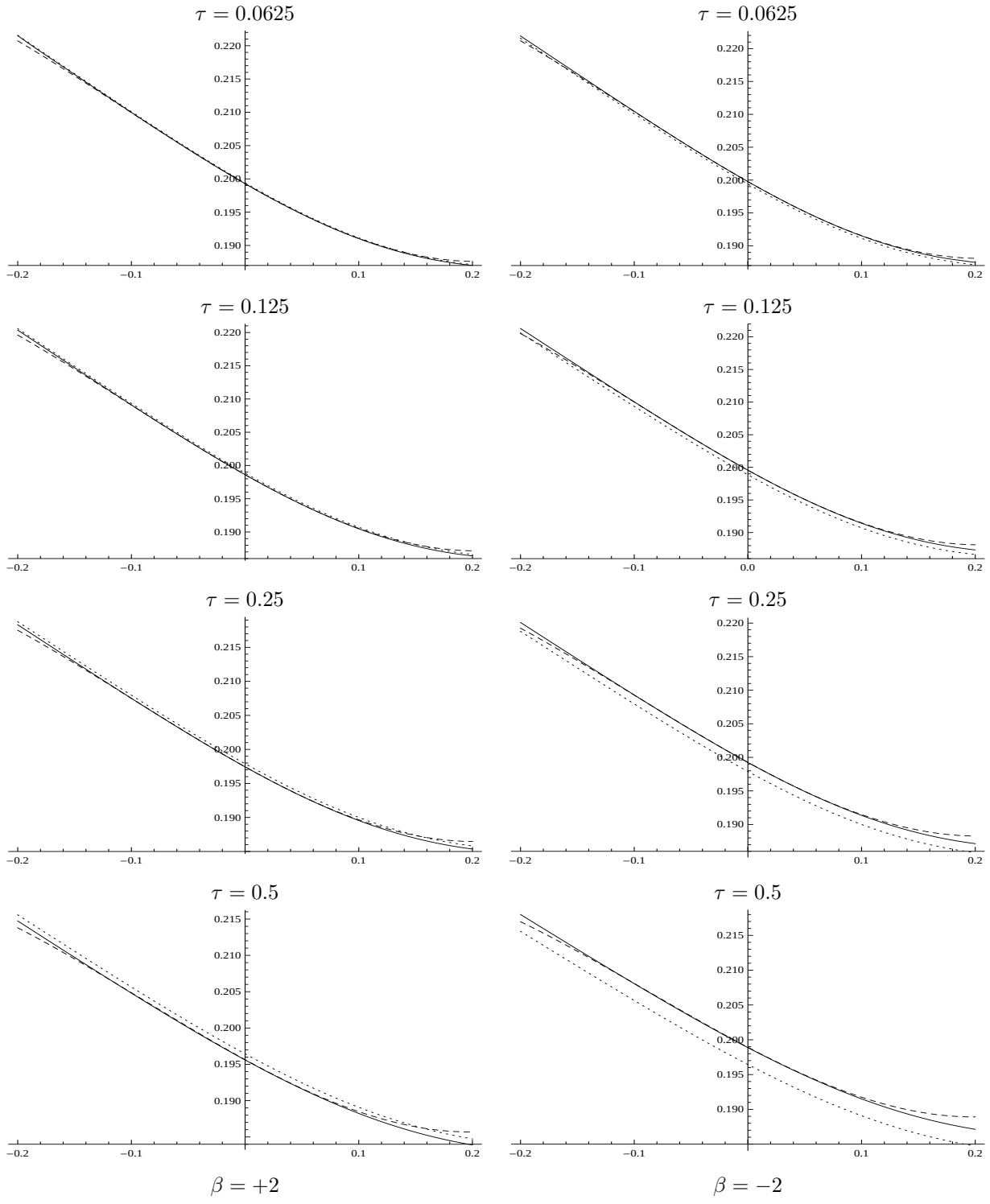

Figure 3: Exact (solid – computed by Fourier inversion) and approximate (dashed) scaled implied volatility $\sigma_Z^{(\beta)}(\tau,\lambda)$ under Heston model dynamics plotted as a function of log-moneyness λ. For comparison, we also plot the exact implied volatility of the Heston model $\sigma_Z^{(1)}(\tau,\lambda) = \sigma_X(\tau,\lambda)$ (dotted). Parameters: $\kappa = 1.15$, $\theta = 0.04$, $\delta = 0.2$, $\rho = -0.4$, $y = \log\theta$. For $\beta = \pm 2$, as τ increases, the dotted lines start to deviate from the solid lines, but the dashed and solids lines stay very close.

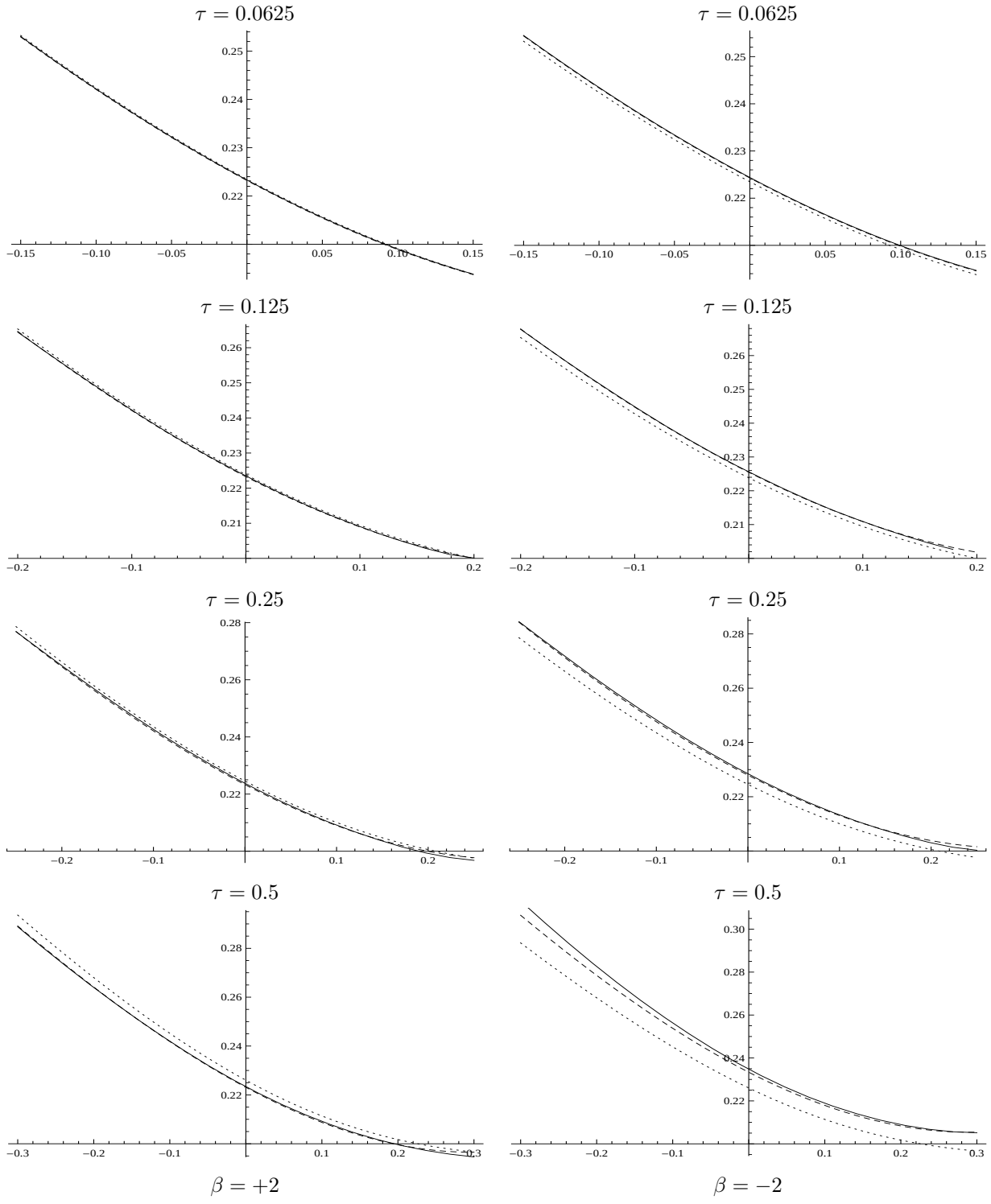

Figure 4: Exact (solid – computed by Monte Carlo) and approximate (dashed) scaled implied volatility $\sigma_Z^{(\beta)}(\tau, \lambda)$ under SABR model dynamics plotted as a function of λ. For comparison, we also plot the exact implied volatility of the SABR model $\sigma_Z^{(1)}(\tau, \lambda) = \sigma_X(\tau, \lambda)$ (dotted). Parameters: $\delta = 0.5$, $\gamma = -0.5$, $\rho = 0.0$ $x = 0$, $y = -1.5$. As expected, as τ increases, the solid and dotted lines diverge, while the dashed and solid lines remain close.