Weak approximations and VIX option price expansions in forward variance curve models

F. BOURGEY *†‡, S. DE MARCO† and E. GOBET †

†Centre de Mathématiques Appliquées (CMAP), CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route de Saclay, Palaiseau Cedex, 91128, France
‡Quantitative Research, Bloomberg L.P., 731 Lexington Ave, New York, NY, 10022, USA

(Received 6 May 2022; accepted 19 May 2023; published online 10 July 2023)

We provide explicit approximation formulas for VIX futures and options in forward variance models, with particular emphasis on the family of so-called Bergomi models: the one-factor Bergomi model, the rough Bergomi model, and an enhanced version of the rough model that can generate realistic positive skew for VIX smiles—introduced simultaneously by De Marco and Guyon at the Bachelier World Congress 2018, that we refer to as ‘mixed rough Bergomi model’. Following the methodology set up in previous works of Gobet and Miri on asymptotic approximations for integrated diffusion processes, we derive weak approximations for the law of the VIX, leading to option price approximations under the form of explicit combinations of Black–Scholes prices and greeks. As new contributions, we cope with the fractional integration kernel appearing in rough models, and treat the case of non-smooth payoffs, so as to encompass VIX futures, call, and put options. We stress that our approach does not rely on small maturity asymptotics, and can therefore be applied to any VIX option maturity. Our results are illustrated by several numerical experiments over a wide range of model parameter configurations, and by calibration tests to VIX market data.

Keywords: VIX options; Weak approximation; Forward variance curve; Rough volatility

MSC 2010: 60G15, 60G22, 91G20, 60H07, 34E10

1. Introduction

The VIX volatility index at a given time $T$ is, by definition, the implied volatility of a 30-day log-contract on the S&P500 index starting at $T$. Introduced in 1993, the VIX is quoted by the Chicago Board Options Exchange and is computed in practice by static replication of the log-contract from market prices of listed vanilla options on the S&P500. In 2004, futures contracts on the VIX started trading, and later on, in 2006, options on the VIX appeared. Futures and options on the VIX can be used as risk-management tools to hedge the volatility exposure of more complex options portfolios and have become extremely popular volatility derivatives.

An efficient yet parsimonious way of modeling the joint dynamics of an asset price $S_t$ (here the S&P500) and its implied volatility is precisely to target the implied variances of log-contracts (usually simply referred to as forward variances), which have the appealing feature of being driftless under the pricing measure, see Dupire (1993), Bergomi (2004), and Buehler (2006). A well-established practice, in the spirit of forward rate modeling, is to model instantaneous forward variances $(\xi^u_t)_{t\leq u}$, defined by $\xi^u_t = \frac{\xi^u_0}{2} ((u-t)\tilde{\sigma}_t(u))^2$, where $\tilde{\sigma}_t(u)^2$ denotes the implied variance of a log-contract with maturity $u$, observed at time $t$.

1.1. A class of exponential models for instantaneous forward variances

An effective class of models, encompassing the celebrated model of Bergomi (2005) and the so-called rough Bergomi model (Bayer et al. 2016), is obtained assuming that the process $(\xi^u_t)_{t\leq u}$ solves the following stochastic differential equation

$$\xi^u_t = \xi^u_0 + \int_0^t \xi^u_s K^u(s) \, dW_s, \quad t \leq u,$$

where (postponing precise assumptions to section 2) $K$ is a deterministic $L^2$ kernel and $\xi^u_0$ a given initial variance curve. In practice, the curve $u \mapsto \xi^u_0$ can be computed from observed option prices. The unique solution to (1) is of course the log-normal process $\xi^u_t = \xi^u_0 e^{\int_0^t K^u(s) \, dW_s - \frac{1}{2} \int_0^t K^u(s)^2 \, ds}$. Equivalently,
the instantaneous log-forward variance $X^u_t := \ln(\xi^u_t)$ solves
\[
dX^u_t = \frac{1}{2} K^u(t) \sigma^2 t + K^u(t) dW_t, \quad t \leq u. \tag{2}
\]

**Examples of deterministic kernels:** The one-factor Bergomi model (Bergomi 2005) corresponds to an exponential kernel of the form
\[
K^u(t) = \omega e^{-k(u-t)}, \quad \omega > 0, k \geq 0, \tag{3}
\]
where the parameter $k$ corresponds to a mean reversion speed, and $\omega$ to the volatility of forward variances. It is well known, see again (Bergomi 2005), that the choice of the exponential kernel leads to a one-dimensional Markovian representation for the forward variance curve: $\xi^u_t = \xi^u_0 f^u(t, Z_t)$ where $Z_t$ is the Ornstein–Uhlenbeck (OU) process $dZ_t = -kZ_t \, dt + dW_t$ and $\phi^u$ an explicit deterministic function (see section 2.5 for details). From a pricing and calibration point of view, such a representation is extremely convenient as it only involves a single Gaussian random variable ($n$-factors extensions of the model (1)–(3) are of course possible and have been considered in Bergomi (2005), leading to low-dimensional Markovian representations of the variance curve in terms of $n$ OU processes).

Introduced recently in Bayer et al. (2016), the rough Bergomi model corresponds to a power kernel of the form
\[
K^u(t) = \eta(u - t)^{-\frac{1}{H}}, \quad \eta > 0, H \in (0, 1/2). \tag{4}
\]
The parameter $H$ encodes the kernel decay, while $\eta$ tunes the volatility of forward variances. As opposed to (3), the kernel now explodes as $u - t \to 0$, and the curve $u \mapsto \xi^u_t$ does not admit a finite-dimensional Markovian representation anymore; on the other side, the model (1)–(3) is able to provide a parsimonious fit to the term structures of implied volatilities and implied volatility skews observed on the S&P500 market (Alòs et al. 2007, Fukasawa 2011, Bayer et al. 2016).

### 1.2. Objectives

In the instantaneous forward variance framework above, the VIX index at time $T$ is given by
\[
\text{VIX}_T^2 = \frac{\Delta}{\Delta} \int_{T}^{T+\Delta} \xi^u_t \, du = \frac{\Delta}{\Delta} \int_{T}^{T+\Delta} e^{X^u_t} \, du. \tag{5}
\]
In this work, we are concerned with the pricing of VIX options $\phi(\text{VIX}_T^2)$. For example, a call (resp. put) option on the VIX with strike $\kappa > 0$ corresponds to $\phi(x) = (\sqrt{x} - \kappa)^+ \ (\text{resp. to } \phi(x) = (\kappa - \sqrt{x})^+) \text{, while VIX futures correspond to } \phi(x) = \sqrt{x} \text{. Under (1), the price at time } t = 0 \text{ of such an option or futures contract is given by}$
\[
\mathbb{E}[\phi(\text{VIX}_T^2)] = \mathbb{E}
\left[ \phi\left( \frac{1}{\Delta} \int_{T}^{T+\Delta} \xi^u_t \, du \right) \right]. \tag{6}
\]

Despite the simplicity of the forward variance model (1), the option pricing pricing (6) is not trivial. The random variable $\text{VIX}_T^2$ is given by a continuous sum of correlated log-normal random variables, and as such, its distribution is not explicit and cannot be simulated exactly. Of course, we see an analogy with Asian option pricing, even if the problem here is structurally different, for the integration in (5) takes place with respect to the maturity dimension of the forward variance curve, as opposed to the running time variable of a Markov process as in Asian option payoffs $\phi(f^u_0 S, dt)$. The expectation (6) can eventually be approximated by coupling a discretization scheme with Monte Carlo simulation, see Jacquier et al. (2018) and Horvath et al. (2020) or Bourgey and De Marco (2022) for the implementation of a multilevel scheme, and asymptotic formulas for short maturity $T$ can also be derived, as in Alòs et al. (2022) and Lacombe et al. (2021). Here, we explore an alternative approach based on analytical approximations of the form
\[
\mathbb{E}[\phi(\text{VIX}_T^2)] = \text{Main term} + \text{Correction terms} + \text{Error}. \tag{7}
\]
Both the main and the correction terms will be easily computable using simple log-normal distributions. In addition, we aim at providing error bounds in terms of the kernel’s characteristics and the length of the time window $\Delta$, covering the case of non-smooth payoffs $\phi$.

### 1.3. Comparison with the literature on option price expansions

**Instantaneous volatility (or first-generation stochastic volatility) models:** Using Fourier-based techniques, semi-analytical formulas are derived in Zhu and Lian (2012) for VIX futures in a Heston-type stochastic volatility model with jumps, and in Goutte et al. (2017) for option prices in a regime-switching Heston model. Short-time expansions are provided in Zhao et al. (2018) for VIX futures and options in stochastic volatility models including Heston, mean-reverting CEV, and 3/2 models, and in Barletta et al. (2019) under the assumption that $\text{VIX} = \phi(t, Y_t)$ for some Markov process $Y$. Assuming multi-scale volatility modeling, some expansions for VIX and S&P500 derivatives are provided in Fouque and Saporito (2018), but their accuracy seems to deteriorate for short maturities (below four months), for which VIX derivatives are the most liquid. In all these works, the VIX model is based on a finite-dimensional Markov process, which does not encompass all models of the form (1).

**Forward variance (or second-generation stochastic volatility) models:** A recent work close in spirit to ours is Guyon (2022), where formal expansions in terms of powers of the volatility-of-volatility parameter are established for VIX futures and power payoffs in the one-factor and two-factor Bergomi models, along with their extensions to the mixed Bergomi model (Bergomi 2008). These expansions are shown to provide accurate approximations for a wide range of model parameters, covering typical values of calibrated parameters in the Equity and FX markets, and also beyond—in particular, even for large values of the volatility-of-volatility parameter. With respect to Guyon (2022), we also deal with the case of VIX call and put options, we cover more general kernel functions, and provide error estimates for our expansions. Concerning specifically the rough Bergomi model (4), several
recent papers have tackled the problem of VIX derivatives pricing in this framework. In Jacquier et al. (2018), instead of expansion formulas, upper and lower bounds for VIX futures are provided, and in Lacombe et al. (2021), large deviation theory is applied to derive small-maturity asymptotic formulas (covering one- and multi-factor mixed rough Bergomi models). In Alòs et al. (2022), the authors specifically focus on the short-maturity at-the-money implied volatility level and skew. Exploiting representations from Malliavin calculus, the authors theoretically confirm the capability of mixed log-normal models to generate a positive VIX skew (Alòs et al. 2022, Section 3.2) (as previously announced by the numerical tests in De Marco 2018, G Guyon 2018), and derive short-term asymptotic formulas for at-the-money values. For instance, in Alòs et al. (2022, Example 21) the short-term limit of the at-the-money VIX implied volatility skew is provided for the mixed rough Bergomi model, which we will consider in section 3.

1.4. More details on our contributions

Our approach follows a different path with respect to small-maturity or small-parameter asymptotics, consisting in taking advantage of a structural property of the VIX: the relatively short time-window $\Delta$ over which forward variances are integrated in (5) (recall that $\Delta \approx 1/\gamma$, when measuring time in years). The first step is to replace the arithmetic average of exponentials in (6) with their geometric average (a log-normal random variable), in the spirit of the work in Kemna and Vorst (1990) for Asian options, which then serves as a central point for deriving our asymptotic expansions. The proximity between the VIX distribution in the exponential model (1) and a log-normal distribution has already been observed and exploited in previous works (Bayer et al. 2016, Jacquier et al. 2018) in order to approximate VIX futures prices; our contribution in this article can precisely be seen as a refinement of the approximations proposed in those works, with additional error estimates.

Our mathematical analysis, although close to Gobet and Miri (2014) about averaged diffusion processes, is significantly different: first, we deal with forward curve processes that do not have a Markovian representation; second, we cover the case of payoffs $\psi$ that can fail to be smooth and are only $1/2$-Hölder; third, we also deal with model mixtures, that is, mixed Bergomi and rough Bergomi models. The terms in the resulting expansion (7) will be given by a Black–Scholes price along with explicit Black–Scholes Greeks, see theorems 2.7 and 3.1 for more details. Thanks to the integration-by-parts formula of Malliavin calculus, we can prove that the error term is still $O(\Delta^3)$ (resp. $O(\Delta^{3/2})$) for smooth payoffs, see section 3. Finally, some numerical tests on market data confirm that our approximations can be used for fast and efficient calibration of the mixed models to VIX smiles (see section 3.2).

1.5. Notations

In most of our explicit formulas and proofs, we find it convenient to factor out the dependence with respect to the initial forward variance curve $u \mapsto \xi_0^u$. In order to do so while still keeping a compact formulation, we introduce the following probability measures on the interval $A := [T, T + \Delta]$

$$v(du) := \frac{du}{A}, \quad v_0(du) := \frac{1}{A} \int_T^{T+\Delta} \xi_0^u du$$

(8)

We will denote

$$v(f) = \int_T^{T+\Delta} f(u) \frac{du}{A} = \int_A f(u) v(du)$$

and

$$v_0(f) = \int_T^{T+\Delta} f(u) v_0(du) = \int_A f(u) v_0(du)$$

(9)

the means of integrable functions $f$ with respect to the measures $v$ and $v_0$. Note that, using the notation above, we have

$$\mathbb{E}[\text{VIX}_T^2] = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u du = v(\xi_0)$$

and yet $v_0(du) = \frac{\xi_0^u}{v(\text{VIX}_T^2)} v(du)$. Of course, when the initial forward variance $u \mapsto \xi_0^u$ is constant, we have $v = v_0$. Finally, we set

$$Y_T^u := X_T^u - X_0^u = -\frac{1}{2} \int_0^T K^u(t)^2 dt + \int_0^T K^u(t) dW_t$$

(10)

We denote $\| \cdot \|_p$ the $L^p$ norm for random variables. In our error estimates and proofs, we will denote $C$ as a generic positive constant that may change from line to line and may depend on the model and option parameters, but which is in any case independent of $\Delta$, of the curve $u \mapsto \xi_0^u$, and the payoff $\psi$. For two non-negative real numbers $x$ and $y$, $x \leq y$ stands for $x \leq Cy$. We denote $\Phi$ the cumulative distribution function of the standard normal distribution.

2. Exponential forward variance models

**Assumption 2.1** The initial instantaneous forward variance curve $u \mapsto \xi_0^u$ is positive, bounded, and bounded away from zero.
Assumption 2.2 The kernel $K(\cdot)$ in (1) is such that $\int_0^T K(tj^2) \, dt < \infty$ for every $u \in [T, T + \Delta]$, for some $\Delta \geq 1$. Moreover, for any $p > 0$, there exists a positive constant $C_p$ such that

$$\frac{1}{\Delta} \int_T^{T+\Delta} e^{\int_T^t K(u^2) \, du} \, dt \leq C_p$$

for all $\Delta \leq \bar{\Delta}$.

Assumption 2.2 is a mild technical condition, which essentially means that the moments of $\xi_t^u$ are integrable over $[T, T + \Delta]$. As a consequence of assumption 2.2, all the moments of the random variable $VIX_2^T$ are also finite, by Jensen’s inequality. Using the notation we introduced in (8) for the measure $\nu_0$, under assumption 2.1 the condition (11) is equivalent to

$$\int_T^{T+\Delta} e^{\int_T^u K(u^2)} \, du \, \nu_0(du) \leq C_p'$$

for some constant $C_p'$ and all $\Delta \leq \bar{\Delta}$; we will apply assumption 2.2 under the form (12) in our estimates and proofs. It is easy to check that assumption 2.2 is satisfied by the one-factor Bergomi model (3) and the rough Bergomi model (4).

2.1. Proxy for the mean of exponentials

Recalling that $Y^T_\nu = X^\nu_T - X^\nu_0$ from (10), the random variable $VIX^2_\nu$ in (5) can be rewritten as

$$VIX^2_\nu = \frac{1}{\Delta} \int_T^{T+\Delta} \xi^u \nu_0 Y^\nu_0 \, du = \mathbb{E}[VIX^2_\nu] \int_T^{T+\Delta} \frac{\xi^u}{\mathbb{E}[VIX^2_\nu]} \nu_0 Y^\nu_0 \, du = \mathbb{E}[VIX^2_\nu] \nu_0 \left( e^{\nu_0 Y^\nu_0} \right) = v(\xi^0) \nu_0 \left( e^{\nu_0 Y^\nu_0} \right),$$

where we have used the definitions (8) and (9) for the measures $\nu, \nu_0$ and their integral means in the last identity. As addressed in the introduction, the starting point of our analysis is to approximate the arithmetic mean of exponentials $\nu_0 \left( e^{\nu_0 Y^\nu_0} \right) = \int_T^{T+\Delta} e^{\nu_0 Y^\nu_0} \nu_0(du)$ with their geometric mean $e^{\nu_0 \left( Y^\nu_0 \right)}$, which has the appealing property of being log-normal (with explicit mean and variance parameters, given in proposition 2.3 below). More precisely, we set

$$VIX^2_{\nu, P} := \mathbb{E}[VIX^2_\nu] e^{\int_T^{T+\Delta} \frac{\xi^u}{\mathbb{E}[VIX^2_\nu]} \nu_0 \left( e^{\nu_0 Y^\nu_0} \right) \, du} = v(\xi^0) e^{\nu_0 \left( Y^\nu_0 \right)},$$

where the subscript $P$ stands for proxy. Log-normal approximations of the VIX random variable in exponential forward variance models have already been exploited by several authors to derive coarse approximations of VIX futures and options prices, as in Bayer et al. (2016), Jacquier et al. (2018), and Horvath et al. (2020). In this work, we are precisely going to quantify the difference between option prices on the true VIX (13) and the corresponding prices computed on the log-normal approximation (14).

In order to work out a representation of the difference $VIX^2_\nu - VIX^2_{\nu, P}$, we introduce the interpolation

$$I(\varepsilon) := v(\xi^0) \int_A e^{\nu_0 Y^\nu_0 + \varepsilon (Y^\nu_0 - \nu_0 Y^\nu_0)} \nu_0(du), \quad \varepsilon \in [0, 1],$$

which is such that $I(0) = VIX^2_\nu$ and $I(1) = VIX^2_{\nu, P}$. Under assumptions 2.1 and 2.2, it is easy to see that the map $\varepsilon \mapsto I(\varepsilon)$ is smooth almost surely, with $n$th derivative given by

$$I^{(n)}(\varepsilon) = v(\xi^0) \times \int_A (Y^\nu_0 - \nu_0 Y^\nu_0)^n e^{\nu_0 Y^\nu_0 + \varepsilon (Y^\nu_0 - \nu_0 Y^\nu_0)} \nu_0(du).$$

Noticing that $I^{(1)}(0) = 0$, an application of Taylor’s theorem with integral remainder yields

$$VIX^2_\nu - VIX^2_{\nu, P} = I(1) - I(0)$$

$$= \int_0^1 (1 - \varepsilon) I^{(2)}(\varepsilon) \, d\varepsilon = \frac{1}{2} \int_0^1 (1 - \varepsilon)^2 I^{(3)}(\varepsilon) \, d\varepsilon.$$ 

The representations of $I(1) - I(0)$ in the second line of (17) will allow us to quantify the difference between $VIX^2_\nu$ and $VIX^2_{\nu, P}$, and to derive our expansions for the expectations functions of $VIX^2_{\nu, P}$.

As mentioned above, the keystone of our approach is the lognormal property of $VIX^2_{\nu, P}$.

**Proposition 2.3** The proxy $VIX^2_{\nu, P}$ is lognormal, that is

$\ln(VIX^2_{\nu, P}) \overset{d}{=} \mathcal{N}(\mu_\nu, \sigma^2_\nu)$,

where the mean and variance parameters are given by

$$\sigma^2_\nu := \int_0^T \nu_0(K(t))^2 \, dt,$$

$$\mu_\nu := \ln(v(\xi^0)) - \frac{1}{2} \int_0^T \nu_0(K(t))^2 \, dt.$$ 

**Proof** According to (14), $\ln(VIX^2_{\nu, P}) - \ln(v(\xi^0))$ is equal to

$$- \frac{1}{2} \int_A \left( \int_0^T K(u)^2 \, du \right) \nu_0(du) + \int_A \left( \int_0^T K(u) \, dW_t \right) \nu_0(du),$$

and we conclude using stochastic Fubini’s theorem (see e.g. Gobet and Miri 2014, Lemma 1.1).
2.2. Strong error estimates between VIX\textsuperscript{2} and its proxy VIX\textsubscript{T\_P}

To estimate the \(L^p\) norm of the difference \(\text{VIX}^2 - \text{VIX}^2\), we need some estimates for the deterministic \(L^p\) norm (over \(\mathcal{A} = \{T, T + \Delta\}\)) of the difference between the diffusion coefficient \(K^u(t)\) (resp. the drift coefficient \((K^u(t))^2\) of the log-forward variance \(\log(\xi^u_t)\) and its integral average \(v_0(K(t))\) (resp. \(v_0(K(t)^2)\)). We require that these deterministic \(L^p\) norms go to zero as \(\Delta\) goes to zero with certain rates \(d_1\) and \(d_2\), see conditions (19) and (20) in assumption 2.4.

In specific examples—the Bergomi model (3) and the rough Bergomi model (4)—we are able to check this assumption and to estimate the corresponding rates \(d_1\) and \(d_2\), see section 2.4.

We further require some conditions such that the proxy random variable is not degenerate, in the sense of (21) and (22).

**Assumption 2.4** For any \(p > 0\), there exist positive constants \(d_1, d_2, C\), and \(\Delta_0\) such that

\[
\Gamma_{\Delta, T, p} := \left( \int_0^T \left| K^u(t) \right|^2 dt \right)^{\frac{1}{2}} \leq C \Delta^{d_1},
\]

\[
\Lambda_{\Delta, T, p} := \left( \int_0^T \left[ K^u(t) \right]^2 dt \right)^{\frac{1}{2}} \leq C \Delta^{d_2}
\]

hold for all \(\Delta \leq \Delta_0\), and

\[
\sup_{\Delta \leq \Delta_0} |\mu_\Delta| \leq C, \quad (21)
\]

\[
\frac{1}{C} \leq \inf_{\Delta \leq \Delta_0} \sigma_\Delta \leq \sup_{\Delta \leq \Delta_0} \sigma_\Delta \leq C.
\]

**Proposition 2.5** Under assumptions 2.1, 2.2, and 2.4, for any \(p, q \geq 1\) we have

\[
\int_\mathcal{A} \left\| Y^u_T - v_0(Y^u_T) \right\|^q_p dt \leq \Delta^{d_1 \wedge \frac{q}{2}}.
\]

**Theorem 2.6** Under assumptions 2.1, 2.2, and 2.4, for any \(p \geq 1\) and \(n \in \mathbb{N}\) we have

\[
\sup_{c \in [0,1]} \left\| I^n(c) \right\|_{p} \leq \nu(\xi_0) \Delta^{d_1 \wedge \frac{q}{2}}.
\]

Consequently, in view of (17),

\[
\left\| \text{VIX}^2 - \text{VIX}^2\right\|_p \leq \nu(\xi_0) \Delta^{d_1 \wedge d_2}.
\]

2.3. General price expansion

In view of our discussion in section 2.1, the leading order term in the approximation of the VIX option price \(E[\varphi(\text{VIX}^2)]\) will be given by the price on the proxy \(E[\varphi(\text{VIX}^2_{T_P})]\). Assuming for a moment that the payoff function \(\varphi\) is smooth, a Taylor expansion around the point \(\text{VIX}^2\) yields

\[
E[\varphi(\text{VIX}^2)] = E[\varphi(\text{VIX}^2_{T_P})] + E[\varphi'(\text{VIX}^2_{T_P}) (\text{VIX}^2 - \text{VIX}^2_{T_P})] + E_0
\]

where the remainder \(E_0 = \int_0^1 (1 - \lambda)E[\varphi''(\lambda\text{VIX}^2 + (1 - \lambda)\text{VIX}^2_{T_P})] (\text{VIX}^2 - \text{VIX}^2_{T_P})^2] d\lambda\) will be treated as an error term. The difference \(\text{VIX}^2_{\Delta} - \text{VIX}^2_{T_P}\) can be expanded using (17): recalling from (16) the expression of the derivative \(I^{(3)}\), we get

\[
\text{VIX}^2_{\Delta} - \text{VIX}^2_{T_P} = \frac{1}{2} I^{(2)}(0) + \int_0^1 \frac{(1 - \varepsilon)^2}{2} I^{(3)}(\varepsilon) d\varepsilon
\]

so that the second expectation on the right-hand side of (26) can eventually be written as

\[
E[\varphi'(\text{VIX}^2_{T_P}) (\text{VIX}^2 - \text{VIX}^2_{T_P})]
\]

\[
= E[\varphi'(\text{VIX}^2_{T_P}) \frac{1}{2} \text{VIX}^2_{T_P}]
\]

\[
\times \int_\mathcal{A} (Y^u_T - v_0(Y^u_T))^2 v_0(d\mu) + E_1
\]

where \(E_1 = E[\varphi'(\text{VIX}^2_{T_P}) \int_0^1 (1 - \varepsilon)^2 I^{(3)}(\varepsilon) d\varepsilon]\) will be treated as a second error term. The random variable multiplying \(\varphi'(\text{VIX}^2_{T_P})\) inside (27) can be interpreted as the random weight appearing after the application of an integration-by-parts formula to higher-order derivatives of \(\varphi\)—which means that, in its turn, the expectation on the right-hand side of (27) can be rewritten in terms of a combination of derivatives of the form \(\partial^i \mathbb{E}[\varphi'(\text{VIX}^2_{T_P} e^i)]\) \(\|e\| = \infty\). The important property of such higher-order derivatives of the expectation \(\mathbb{E}[\varphi'(\text{VIX}^2_{T_P})]\) is to be explicit—they are Black–Scholes Greeks. The final expression of the expansion (26) will therefore contain a combination of a Black–Scholes price and some of its partial derivatives.

The details of the approach sketched above will be given in section A.3, here we state the final expression we obtain for the expansion (26) after the integration-by-parts procedure, see theorem 2.7 below, which is our main result in this section.
We will make use of the following coefficients \((\gamma_i)_{i=1,2,3}\):

\[
\begin{align*}
\gamma_1 &:= \frac{1}{8} \int_{0}^{T} \left( \int_{0}^{T} \left[ K^u(t) (t)^2 - v_0 (K(t))^2 \right] dt \right) v_0 (du) \\
&\quad + \frac{1}{2} \int_{0}^{T} \left( \int_{0}^{T} [K^u(t) - v_0 (K(t))^2] dt \right) v_0 (du) , \\
\gamma_2 &:= \frac{1}{2} \int_{0}^{T} \int_{0}^{T} (K^u(t) - v_0 (K(t))) dt v_0 (du) , \\
\gamma_3 &:= \frac{1}{2} \int_{0}^{T} \left( \int_{0}^{T} [K^u(t) - v_0 (K(t))] dt \right) v_0 (du) .
\end{align*}
\]

(28)

THEOREM 2.7 (Option price approximation) Let assumptions 2.1, 2.2, and 2.4 be in force, and let \(\psi : \mathbb{R} \rightarrow \mathbb{R}\) be a \(\theta\)-Hölder continuous function for some \(\theta \in (0, 1]\). The price of the VIX option with payoff \(\psi(VIX_T^2)\) is given by

\[
\mathbb{E} \left[ \psi \left( VIX_T^2 \right) \right] = \mathbb{E} \left[ \psi \left( VIX_{T,T} \right) \right] + \sum_{i=1}^{3} \gamma_i \partial^i_\theta \mathbb{E} \left[ \psi \left( VIX_{T,T}^2 e^\varepsilon \right) \right]_{\varepsilon=0} + \mathcal{O}_\varepsilon ,
\]

(29)

where \(\mathcal{O}_\varepsilon\) is an error term such that \(|\mathcal{O}_\varepsilon| \leq \Delta^{3(\theta \wedge \frac{1}{2})} \).

REMARK 1 VIX futures correspond to \(\psi(x) = \sqrt{x}\). VIX put options to \(\psi(x) = (k - \sqrt{x})^+\) and call options to \(\psi(x) = (\sqrt{x} - k)^+\). In all these cases, the function \(\psi\) is \(\frac{1}{2}\)-Hölder continuous (\(\psi\) is Lipschitz in the case of put options with strictly positive strike \(k\)).

REMARK 2 Although the payoff \(\psi\) may fail to be smooth, condition (22) ensures that the lognormal proxy \(VIX_T^2\) is not degenerate, with the effect of regularizing the map \(\varepsilon \mapsto \mathbb{E}[\psi(VIX_T^2 e^\varepsilon)]\), so that the derivatives \(\partial^i_\theta \mathbb{E}[\psi(VIX_T^2 e^\varepsilon)]\) are well-defined.

REMARK 3 The dimensional coefficients \((\gamma_i)_{i=1,2,3}\) are defined by deterministic integrals with respect to time variables. They depend on the option maturity \(T\), on the time window \(\Delta\), and on the model parameters \(\xi_0\) and \(K\), but not on the option payoff—which means that, in the case of call and put options, they can be evaluated once for all strikes. When the initial forward variance curve \(u \mapsto \xi_u^2\) is constant over the VIX time window \((T, T + \Delta)\) (which is a standard choice that can usually be made in practice), the \(\gamma_i\)'s have analytical closed-form expressions in the Bergomi model (3), see proposition 2.9. In the rough Bergomi model (4), the \(\gamma_i\)'s do not seem to admit a closed-form representation even when the initial variance curve \(\xi_u^2\) is constant, but in this case, their dependence with respect to the remaining model parameters \(\eta\) and \(H\) can be simplified, see remark 7 for more details.

For the general case of a non-constant initial curve \(\xi_u^2\), one has to appeal to deterministic quadrature to approximate the integrals with respect to the measure \(v_0(du) = \frac{\xi_u^2}{\xi_0^2} du / \Delta\).
\( \Gamma \) VIX futures with
\[
P_i^T = 2^{-i} e^{2T} + \frac{\varphi^2}{2} \quad \text{for } i \in \{0, 1, 2, 3\},
\]
\( \triangle \) VIX put options with
\[
P^\text{put}_0 := P^\text{all}_0 - P^\text{F}_0 + \kappa, \quad P^\text{put}_i = P^\text{all}_i - P^\text{F}_i \quad \text{for } i \in \{1, 2, 3\}.
\]
In particular, note that put–call parity holds for the truncated closed-form expressions for the coefficients \( \gamma_i \) in (29), that is
\[
P^\text{all}_0 - P^\text{put}_0 + \sum_{i=1}^{3} \gamma_i (P^\text{all}_i - P^\text{put}_i) = P^\text{F}_0 + \sum_{i=1}^{3} \gamma_i P^\text{F}_i - \kappa.
\]

(30)

2.4. Verification of assumption 2.4 and evaluation of the expansion in our main examples

In this section, we assume that the initial instantaneous forward variance curve \( u \mapsto \xi^0_u \) is constant over the interval \( (T, T + \Delta) \). In this case, \( v_u(du) = \nu(du) = \frac{du}{\Delta} \).

2.4.1. Bergomi model. We consider the one-factor Bergomi model (3), where the convolution kernel is \( K_u(t) = \omega e^{-t(t-u)} \) with \( \omega > 0 \) and \( k \geq 0 \). In proposition 2.9, we establish explicit formulas for the proxy’s mean and variance (18), and work out the exact asymptotics of the \( L^p \) norms (19) and (20) as \( \Delta \to 0 \). This will allow us to see that assumption 2.4 is verified for the Bergomi model. Moreover, we provide closed-form expressions for the coefficients \( \gamma_i \) appearing in the expansion (29).

Proposition 2.9 Consider the one-factor Bergomi model (3) and recall the coefficients \( \Gamma_{\Delta,T,p} \) and \( \Lambda_{\Delta,T,p} \) defined in assumption 2.4. If \( k = 0 \), then \( \Gamma_{\Delta,T,p} = \Lambda_{\Delta,T,p} = 0 \). Otherwise if \( k > 0 \), for every \( p > 0 \) we have
\[
\begin{align*}
\Gamma_{\Delta,T,p} &\sim \frac{\omega^2}{2} \frac{(1 - e^{-2kT})}{(1 - e^{-k\Delta})}, \\
\Lambda_{\Delta,T,p} &\sim \frac{\omega^2}{8k^2\Delta^2} \frac{1}{(1 - e^{-2kT})^2}.
\end{align*}
\]

(31)
The mean and variance (18) of the proxy VIX\(^2\)\(_{T,p}\) are given by
\[
\mu_p = X_0 - \frac{\omega^2}{8k^2\Delta} \frac{(1 - e^{-2kT})}{(1 - e^{-k\Delta})},
\]
\[
\sigma_p^2 = \frac{\omega^2}{2k^3\Delta^2} \frac{(1 - e^{-2kT})}{(1 - e^{-k\Delta})}.
\]
Furthermore, the coefficients \( \gamma_i \) defined in (28) have the following closed-form expressions
\[
\gamma_1 = \frac{\omega^4}{128k^4\Delta^2} \left( -1 + k\Delta \frac{1 + e^{-2k\Delta}}{1 - e^{-2k\Delta}} \right) \times \frac{(1 - e^{-2kT})^2}{(1 - e^{-2k\Delta})} \frac{1}{(2 + k\Delta)e^{-k\Delta} - 2 + k\Delta}
\]
\[
\gamma_2 = -\frac{\omega^4}{48\Delta k^5} \frac{(1 - e^{-k\Delta})^2}{(1 - e^{-2k\Delta})^2} \times \left( 2k\Delta e^{-k\Delta} + 2k\Delta + e^{-2k\Delta}(2k\Delta + 3) - 3 \right) \times \left( 1 - e^{-2kT} \right)^2,
\]
\[
\gamma_3 = \frac{\omega^4}{16k^3\Delta^3} \frac{(1 - e^{-2kT})^3}{(1 - e^{-k\Delta})^3} \times \left( k\Delta - 2 + (2 + k\Delta)e^{-k\Delta} \right).
\]

In light of (31) and of the expressions of \( \mu_p \) and \( \sigma_p \) in proposition 2.9, we have the following result.

Corollary 2.10 In the Bergomi model (3) with parameter \( k > 0 \), assumption 2.4 holds with \( d_1 = 1 \) and \( d_2 = 2 \). Consequently, theorem 2.7 holds and the error term \( \varepsilon_p \) in (29) is \( O(\Delta^3) \).

2.4.2. The rough Bergomi model. We now consider the rough Bergomi model (4). Recall that in this case, we have the fractional kernel \( K_u(t) = \eta(u - t)^H - 1 \), with \( \eta > 0 \) and \( H \in (0, 1) \).

Proposition 2.11 Consider the rough Bergomi model (4) and recall the coefficients \( \Gamma_{\Delta,T,p} \) and \( \Lambda_{\Delta,T,p} \) defined in assumption 2.4. If \( H = \frac{1}{2} \), then \( \Gamma_{\Delta,T,p} = \Lambda_{\Delta,T,p} = 0 \). Otherwise if \( H \in (0, 1) \setminus \{\frac{1}{2}\} \), for every \( p > 0 \) we have
\[
\Gamma_{\Delta,T,p} \sim \frac{(\omega^2)^2}{2H} \left( \int_0^1 \frac{1}{2H + 1} - \int_0 \Delta 2H \right) \text{ if } H \in (0, \frac{1}{2})
\]
\[
\sim \frac{(\omega^2)^2}{2(1 + p)^3} \text{ if } H \in \left(\frac{1}{2}, 1\right),
\]
and
\[
\Lambda_{\Delta,T,p} \sim \frac{(\omega^2)^2}{2H} \left( \int_0^1 \frac{1}{2H + 1} - \int_0 \Delta 2H \right) \text{ if } H \in (0, \frac{1}{2})
\]
\[
\sim \frac{(\omega^2)^2}{2(1 + p)^3} \text{ if } H \in \left(\frac{1}{2}, 1\right),
\]
where \( \text{faa}(H,p) := \frac{1}{\Gamma(H + \frac{1}{2})} (\int_0^1 (1 + s)^{-H - \frac{1}{2}} - s^{-H - \frac{1}{2}} - (H + \frac{1}{2})(1 + s)^{-H - \frac{1}{2}} d\delta[\rho](du))^{-1} \). The proxy’s mean and variance (18) are given by
\[
\mu_p = X_0 - \frac{(\omega^2)^2}{4H(2H + 1)} \text{ if } H \in (0, \frac{1}{2})
\]
\[
- \frac{\eta^2 T^{2H}}{4H} \in \mathbb{R},
\]
\[
\sigma_p^2 = \frac{(\omega^2)^2}{2H} \left( \int_0^1 \frac{1}{2H + 1} - \int_0 \Delta 2H \right) \text{ if } H \in (0, 1)
\]
\[
\sim \frac{\eta^2 T^{2H}}{2H} \text{ if } H \in (0, 1). \]

Consequently, estimates (21) and (22) hold for the rough Bergomi model.
2.5. Numerical tests of option price formulas

In this section, we test our approximation formulas for VIX futures, call, and put options. All the numerical tests (in the current section and in section 3) were performed on a MacBook Air laptop (M1, 2020) with 8GB of memory using the programming language Python 3.9.9. First of all, let us explain how we compute our reference prices.

Remark 8 Computation of the reference prices For the standard Bergomi model, recall we have the Markovian representation of forward variances

\[
\text{VIX}_T = \frac{1}{\Delta} \int_T^{T+\Delta} e^{\int_0^u g(\xi_x)} (T, X_T) \, du,
\]

where

\[
f_u(T, x) = \exp \left( \omega e^{-k(u-T)} x \right) - \frac{1}{2} \omega^2 e^{-2k(u-T)} \text{Var}(X_T), \quad \forall x \in \mathbb{R},
\]

and \((X_t)_{t \geq 0}\) is the Ornstein–Uhlenbeck process \(X_t = -k \int_0^t X_s \, ds + W_t\), with \(\text{Var}(X_T) = \frac{1}{2} e^{-2kT} \mathbb{1}_{x \geq 0} + T \mathbb{1}_{x = 0}\). Consequently, in order to price an option on \(\text{VIX}_T\) in the one-factor Bergomi model, we can (and do) rely on a two-dimensional deterministic quadrature: we couple a Gauss–Legendre scheme for the integration with respect to the time parameter \(u\), and a Gauss–Hermite scheme for the space dimension (see Bergomi 2005, Guyon 2022 for more details on the implementation).

For the rough Bergomi model, the Markovian representation is not available anymore and we rely on a discretization of the variance curve process \(u \mapsto \xi_u^2\). We approximate the integral \(\text{VIX}_T^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_u^2 (T, X_T) \, du\) using a rectangle scheme \(\text{VIX}_T^{2_n} = \frac{1}{\Delta} \sum_{n=1}^n \xi_{u_n}^2\) over a regular grid \((u_i)_{i=1,\ldots,n}\). The random vector \((\xi_u^2)\) can be simulated exactly, so that we eventually approximate \(\text{VIX}\) futures and options prices via their empirical means over i.i.d. samples of the discretized variable \(\text{VIX}_T^{2_n}\). The related weak error is known to behave as \(O(\frac{1}{\Delta})\) (independently of the value of \(H\)), see Horvath et al. (2020) and Bourgey and De Marco (2022). We can considerably reduce the variance of the estimators using an efficient control variate that is nothing but the discretized version of our lognormal proxy, see again (Bourgey and De Marco 2022) for details.

We consider three different \(\text{VIX}\) maturities \(T \in \{1, 3, 6\text{ months}\}\), and set the initial instantaneous forward variance to \(\xi_0 = 0.235^2\) (so that \(X_0 = \ln(\xi_0) \approx -2.896\)). Note we could have considered a piece-wise constant initial variance curve, constant over \((T, T + \Delta)\) for every \(T\), instead of a flat one–this will actually be the case in our calibration tests in section 3. To keep track of the error sign, we consider the signed relative error, computed as

\[
\text{Approximation Price} - \text{Reference Price} \times 100.
\]

2.5.1. Numerical tests for the rough Bergomi model. We consider options struck at \(k = 0.2\). For the reference price, we used \(M = 10^6\) Monte Carlo samples and \(n = 300\) discretization points to construct the estimators outlined in remark 8. We compare the accuracy of the expansions for various values of \(\eta\) and \(H\). In each case, we plot the reference price along with its 95% Monte Carlo confidence interval.

Quality of the approximation for different values of the vol-of-variance \(\eta\): In figure 1, we set \(H = 0.1\, \text{ and } \Delta = \frac{1}{12}\), and choose 10 evenly-spaced values of \(\eta\) ranging from 0.1 to 1.5. We observe that the smaller the \(\eta\), the more accurate our expansion formulas. Yet, even for large values of \(\eta\) (see table 4 for typical values of \(\eta\) obtained from the calibration of market \(\text{VIX}\) smiles), our approximations are extremely accurate (almost indistinguishable from the reference prices in the left plots in figure 1): absolute relative errors are less than 0.5% for the futures contract, 0.3% for the call option, and 1.4% for the put option.

Quality of the approximation for different values of \(H\): In figure 2, we set \(\eta = 1\, \text{ and } \Delta = \frac{1}{12}\), and choose 10 evenly-spaced values of \(H\) ranging from 0.05 to 0.4. Since the closer \(H\) to zero, the more singular the kernel, we expect the errors of our
Figure 1. VIX futures, call, and put option prices for different values of the volatility-of-variance parameter \( \eta \) in the rough Bergomi model (4). Top: benchmark prices obtained with the Monte Carlo procedure described in remark 8 along with our explicit expansions from theorem 2.7. Bottom: percentage relative errors between the benchmark prices and the expansions.

Figure 2. VIX futures, call, and put option prices for different values of the fractional parameter \( H \) in the rough Bergomi model (4). Top: benchmark prices obtained with the Monte Carlo procedure described in remark 8 along with our explicit expansions from theorem 2.7. Bottom: percentage relative error between the benchmark prices and the expansions.
expansions to be a decreasing function of $H$, and this is indeed what we observe in figure 2.

2.5.2. Numerical tests for the standard Bergomi model.

We now focus on the standard Bergomi model and consider VIX futures and at-the-money VIX call and put options. The reference price is computed as described in remark 8, using 80 nodes for the deterministic quadratures in the time and space dimensions.

Quality of the approximation for different values of the mean-reversion strength $k$: In figure 3, we set $\omega = 2$, $\Delta = \frac{1}{12}$, and choose 10 evenly-spaced values of $k$ ranging from 0.5 to 15. Note that the values for the call and put options are the same (for we considered at-the-money options); this is in line with the put-call parity (30) satisfied by our approximation formulas. Once again, despite the wide range of values chosen for the mean-reversion parameter $k$, we note that the approximation formulas provided by theorem 2.7 are extremely accurate: relative errors are now less than $10^{-3}$\% for the VIX futures and less than 1\% for at-the-money options.

Quality of the approximation for different values of the volatility-of-variance $\omega$: In figure 4, we set $k = 1$, $\Delta = \frac{1}{12}$, and choose 10 evenly-spaced values of $\omega$ ranging from 0.5 to 6. The results displayed in this figure are in line with the previous ones, showing a remarkable accuracy of the approximation formulas even for large values of the volatility-of-variance parameter $\omega$ (for comparison, see the largest values of $\omega$ calibrated to market VIX smiles in table 4). Relative errors are still of order $10^{-3}$\% for VIX futures and less than 1\% for at-the-money options.

2.5.3. Behavior of the error terms for different values of $\Delta$.

We wish to compare the theoretical estimates $O(\Delta^3)$ and $O(\Delta^{3H})$ given in corollaries 2.10 and 2.12 for the error terms in the standard Bergomi and rough Bergomi models with their empirical behavior. In figure 5, we plot the absolute difference between the reference price and our approximations for futures and at-the-money call and put options, for several values of the time-window $\Delta$, in logarithmic scale for the $x$-axis and $y$-axis. We consider a grid of 10 evenly-spaced values of $\Delta$ in $[0.05, 0.25]$, for both models, and set $T = 1$ month and $\xi_0 = 0.04$. The parameters $H$, $\eta$, $k$, $\omega$ are displayed above each figure.

We observe consistency with the error behavior predicted by corollaries 2.10 and 2.12 for the two models.

2.5.4. VIX implied volatility.

The implied volatility of VIX options is computed from the Black-Scholes formula, using as forward parameter the model-generated VIX futures. In figure 6, we plot the VIX smile in the rough Bergomi model obtained from the reference option prices when $\xi_0 = 0.2352$, $\eta = 1$, and $H = 0.1$, along with its approximation from theorem 2.7 and the associated signed relative error. Our approximation formula is again very accurate and yields relative errors for implied volatilities smaller (in absolute value) than 1.5\% for a one-month maturity, 0.5\% for three months, and 0.35\% for six months.

It has already been observed and reported by several authors (Bayer et al. 2016, Horvath et al. 2020) that the VIX smile generated by exponential forward variance models (1)
Weak approximations and VIX option price expansions

Figure 4. VIX futures, ATM call, and ATM put option prices for different values of the volatility-of-variance parameter $\omega$ in the standard Bergomi model (3). Top: benchmark prices obtained according to the two-dimensional quadrature described in remark 8, along with our explicit expansion from theorem 2.7. Bottom: percentage relative error between the benchmark prices and the expansions.

Figure 5. Absolute difference between the reference price and our approximation formula from theorem 2.7 for futures, ATM call, and ATM put options as a function of $\Delta$, for the rough Bergomi model (left figure) and the standard Bergomi model (right figure). Axes are both in logarithmic scale.

is almost flat, as also observed in figure 6. This is precisely a consequence of the fact that the true VIX$^2_T$ random variable is well approximated by the log-normal proxy VIX$^2_T,p$ for realistic model parameters, so that the VIX itself is not far from a log-normal random variable with a flat smile structure (as we have already pointed out, our theorems 2.6 and 2.7 precisely quantify the difference existing between the distributions of these two random variables). In practice, market VIX smiles exhibit a pronounced positive skew, which has motivated the search for more general model classes, which we now consider in section 3.

The qualitative behavior of the VIX smiles that we observe for the standard Bergomi model is very similar, as already reported in Bergomi (2005). The VIX smiles we obtained from the reference prices and the price approximations in the Bergomi model are therefore not reported, being very similar
to the implied volatility smiles in figure 6 (the approximation formulas still being very accurate when compared to the reference implied volatilities).

3. Mixed exponential models

A class of models more general than (1), and capable of appropriately capturing the behavior of market VIX smiles, was introduced in Bergomi (2008), who observed that a simple yet efficient way to twist the distribution of forward variances is to replace the exponential process $\xi_t = \xi_0 e^{\lambda V}$ stemming from the dynamics (1) with a convex combination of two exponential functions. This observation yields the extended model family

$$
\xi_t^u = \xi_0^u \left[ \lambda \mathcal{E} \left( \int_0^T K^u(t) \, dW_t \right) \right] + (1-\lambda) \mathcal{E} \left( \int_0^T K^u(t) \, dW_t \right),
$$

(33)

which we refer to as ‘mixed exponential models’ or ‘mixed Bergomi models’. In (33), $\mathcal{E}(X)$ is a shorthand notation for $e^{X / \sqrt{\text{Var}(X)}}$, and $\lambda \in [0, 1]$ is a mixing parameter allowing to tune the relative importance of each exponential term.

**Remark 9** (Mixing log-normal models) When $K^u(t) = \omega_i e^{-(u-t)^{\frac{-1}{\gamma_i}}}$, $i \in \{1,2\}$, (33) corresponds to the mixed one-factor Bergomi model introduced in Bergomi (2008), and when $K^u_j = \eta_j (u-t)^{\frac{-1}{\gamma_j}}$, (33) yields the mixed rough Bergomi model, introduced simultaneously by De Marco (2018) and Guyon (2018). It is of course possible to consider representations that encompass both the standard Bergomi and the rough Bergomi models, using kernels of the form $K^u(t) \propto e^{-(u-t)^{\frac{-1}{\gamma}}}$, as done in Alòs et al. (2022, Section 3.2). The squared VIX is of course still defined by integrated instantaneous forward variances as in (5); under (33), we see that $\text{VIX}_t^2$ is given by a convex combination of integral means of the form (13), that is

$$
\text{VIX}_t^2 = \lambda \text{VIX}_{t,1}^2 + (1-\lambda) \text{VIX}_{t,2}^2
$$

$$
:= \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u e^{Y_t^u} \, du + (1-\lambda) \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u e^{Y_t^u} \, du,
$$

(34)

where

$$
Y_{t,j}^u := \frac{1}{2} \int_0^T K^u_j(t)^2 \, dt + \int_0^T K^u_j(t) \, dW_t \quad j \in \{1,2\}.
$$

(35)

Following the approach of section 2, we approximate each integral mean with a log-normal random random variable, so that $\text{VIX}_t^2$ is eventually approximated by

$$
\text{VIX}_{t,P}^2 = \nu \left( \xi_0 \right) \text{VIX}_{t,P,1}^2 + (1-\nu) \left( \xi_0 \right) \text{VIX}_{t,P,2}^2
$$

The overall proxy $\text{VIX}_{t,P}^2$ is therefore a convex combination of correlated log-normal random variables. For $j \in \{1,2\}$, we have

$$
\ln \left( \text{VIX}_{t,P,j}^2 \right) \overset{d}{=} \mathcal{N} \left( \mu_{P,j}, \sigma_{P,j}^2 \right),
$$

(36)

where $\mu_{P,j} := -\frac{1}{2} \int_0^T v_0(K_j(t)^2) \, dt$, $\sigma_{P,j}^2 := \int_0^T v_0(K_j(t)^2) \, dt$.  

3.1. Price expansion

Let us define the analogous of the coefficients $\gamma$ in (28): for $j \in \{1,2\}$,

$$
\gamma_{1,j} := \frac{1}{8} \int_{\mathcal{A}} \left( \int_0^T \left[ K^u_j(t)^2 - v_0(K_j(t)^2) \right] \, dt \right)^2 \, v_0(du)
$$

$$
+ \frac{1}{2} \int_{\mathcal{A}} \left( \int_0^T \left[ K^u_j(t) - v_0(K_j(t)) \right]^2 \, dt \right) \, v_0(du),
$$

$$
\gamma_{2,j} := \frac{1}{8} \int_{\mathcal{A}} \left( \int_0^T \left[ K^u_j(t)^2 - v_0(K_j(t)^2) \right] \, dt \right)^2 \, v_0(du)
$$

$$
+ \frac{1}{2} \int_{\mathcal{A}} \left( \int_0^T \left[ K^u_j(t) - v_0(K_j(t)) \right]^2 \, dt \right) \, v_0(du),
$$

Figure 6. Left: VIX smiles in the rough Bergomi model for $T = 1, 3, 6$ months. Right: relative error in %. The model parameters are $\xi_0 = 0.235^2$, $\eta = 1$, $H = 0.1$. 
\[\gamma_{2,j} := -\frac{1}{2} \int_\mathcal{A} \left( \int_0^T v_0(K_j(t)) \left[ K_j^\nu(t) - v_0(K_j(t)) \right] dt \right) \times \left( \int_0^T \left[ (K_j^\nu(t))^2 - v_0(K_j(t))^2 \right] dt \right) v_0(du),\]
\[\gamma_{3,j} := \frac{1}{2} \int_\mathcal{A} \left( \int_0^T v_0(K_j(t)) \left[ K_j^\nu(t) - v_0(K_j(t)) \right] dt \right) \cdot v_0(du).\]

**Theorem 3.1** Let \( \psi \in C^2_\mathcal{A}. \) In the mixed rough Bergomi model obtained setting \( K_j = \eta(t - u)^{H - \frac{1}{2}} \) in (33), the price of an option on \( \text{VIX}_T^2 \) with payoff \( \psi \) is given by

\[\mathbb{E} \left[ \psi \left( \text{VIX}_T^2 \right) \right] = \mathbb{E} \left[ \psi \left( \text{VIX}_{T,P}^2 \right) \right] + \sum_{i=1}^2 \sum_{j=1}^2 \gamma_{i,j} P_{i,j} + \mathcal{E}_p,\]

where \( \mathcal{E}_p \) is an error term satisfying \( |\mathcal{E}_p| \leq c \Delta^{d_1,d_2} \) with \( d_1, d_2 \) given in (19)–(20), and

\[\mathbb{E} \left[ \psi \left( \text{VIX}_{T,P}^2 \right) \right] = \mathbb{E} \left[ \psi \left( \nu(\xi_0) \left[ \lambda e^{\mu_1 + \sigma_1 Z} + (1 - \lambda) e^{\mu_2 + \sigma_2 Z} \right] \right) \right],\]

\[P_{i,j} = \frac{\omega_{i,j}}{\lambda} \mathbb{E} \left[ \Psi_j \left( \mu_{i,j} + \sigma_{i,j} Z + \epsilon \right) \right] \}_{\epsilon = 0},\]

where \( \Psi_j(x) = \partial_\epsilon \psi \left( \nu(\xi_0) \left[ \lambda e^{\mu_1 + \sigma_1 Z} + (1 - \lambda) e^{\mu_2 + \sigma_2 Z} \right] \right) \) \( x = 1, 2, \) \( j = 1, 2, \)

\[\Psi_1(x) = \partial_\lambda \psi \left( \nu(\xi_0) \left[ \lambda e^{\mu_1 - \sigma_1 Z} + (1 - \lambda) e^{\mu_2 - \sigma_2 Z} \right] \right) \] \( x = 1, 2, \) \( j = 1, 2, \)

\[\Psi_2(x) = \partial_\mu \psi \left( \nu(\xi_0) \left[ \lambda e^{(\mu_1 - \sigma_1) t} + \sigma_1 \right] + (1 - \lambda) e^{(\mu_2 - \sigma_2) t} + \sigma_2 \right) \] \( x = 1, 2, \) \( j = 1, 2, \)

\[K_j^\nu(t) = (u - t)^{H - \frac{1}{2}}.\]

**Remark 10** We note that the form of (38)–(40) is specific to the mixed one-factor Bergomi model where \( K_j^\nu(t) = \omega e^{-k(t-u)} \) (same value of \( k \) for the two kernels) and to the rough Bergomi model where \( K_j^\nu(t) = \eta(t - u)^{H - \frac{1}{2}} \) (same value of \( H \)). In these cases, \( Y_{T,1}^\nu \) and \( Y_{T,2}^\nu \) are linearly dependent Gaussian variables (\( \text{Corr}(Y_{T,1}^\nu, Y_{T,2}^\nu) = 1 \)) and the VIX proxy is a function of a single Gaussian random variable,

\[\text{VIX}_{T,P}^2 \buildrel d \over = \nu(\xi_0) \left[ \lambda e^{\mu_1 + \sigma_1 Z} + (1 - \lambda) e^{\mu_2 + \sigma_2 Z} \right],\]

\[Z \overset{d}{=} \mathcal{N}(0, 1),\]

so that all the expressions in theorem 3.1 can be evaluated with efficient one-dimensional Gaussian quadratures, as we explain in detail in the following section.

Theorem 3.1 could be extended to non-smooth payoffs, using similar arguments to the proof of Theorem 2.7. Leaving this rather long analysis for future work, we focus on the current statement of theorem 3.1 for smooth payoffs in appendix, while still providing numerical tests for futures, call and put payoffs in the next section.

### 3.2. Numerical tests for option price formulas and implied volatilities

Reference prices in the mixed models are still computed as described in remark 8: in the mixed rough Bergomi model, we discretize the \( \text{VIX}_T^2 \), in (34) for \( j \in \{1, 2\} \) with a rectangle scheme and simulate exactly the discretized variable, while in the mixed standard Bergomi model, we exploit the Markovian representation in remark 8 for each term \( \text{VIX}_T^2 \), \( j \in \{1, 2\} \), and apply a two-dimensional deterministic quadrature with respect to the parameter \( u \) and to the space dimension.

#### 3.2.1. Computation of the \( P_{i,j} \) in theorem 3.1 for \( i \in \{1, 2, 3\} \) and \( j \in \{1, 2\} \). Let us drop the subscript \( P \) and denote \( \mu_j = \mu_{i,j}, \sigma_j = \sigma_{i,j}. \) Recalling that

\[\mathbb{E} \left[ \psi \left( \text{VIX}_{T,P}^2 \right) \right] = \mathbb{E} \left[ \psi \left( \nu(\xi_0) \left[ \lambda e^{\mu_1 + \sigma_1 Z} + (1 - \lambda) e^{\mu_2 + \sigma_2 Z} \right] \right) \right],\]

where \( Z \sim \mathcal{N}(0, 1) \), we can use a one-dimensional Gauss–Hermite quadrature (with 80 nodes in our tests) to evaluate the option price over the proxy \( \mathbb{E}[\psi(\text{VIX}_{T,P}^2)] \). Recall that VIX futures correspond to \( \psi(x) = x^2 \) and VIX call options to \( \psi(x) = (\sqrt{x} - \kappa)^+. \) The terms \( P_{1,1} \) and \( P_{1,2} \) for \( i \in \{1, 2, 3\} \) are given by derivatives of the expectation above with respect to a parameter, and therefore they can be recast under the form of expectations using the likelihood method (derivation of the density function). Consider the terms \( P_{1,i} \) for \( i \in \{1, 2, 3\} \). We have \( P_{1,1} = \mathbb{E}[\Psi_1(\mu_1 + \sigma_1 Z)] \), where the function \( \Psi_1 \) is explicitly given in (39). Since

\[ f_1(\epsilon) := \mathbb{E}[\Psi_1(\mu_1 + \sigma_1 Z + \epsilon)] \]

\[= \frac{1}{\sigma_1} \int_{\mathbb{R}} \Psi_1(y) \frac{\exp \left( -\frac{(y - \mu_1)^2}{2\sigma_1^2} \right)}{\sqrt{2\pi}} dy,\]

we have

\[P_{2,1} = f_1(0) = \frac{1}{\sigma_1} \int_{\mathbb{R}} \Psi_1(y) \frac{y - \mu_1}{\sigma_1} \frac{\exp \left( -\frac{(y - \mu_1)^2}{2\sigma_1^2} \right)}{\sqrt{2\pi}} dy = \frac{1}{\sigma_1} \mathbb{E}[\Psi_1(\mu_1 + \sigma_1 Z)],\]

and

\[P_{3,1} = f_1'(0) = \frac{1}{\sigma_1^2} \mathbb{E}[\left( Z^2 - 1 \right) \Psi_1(\mu_1 + \sigma_1 Z)].\]

Consequently, we can again use a one-dimensional Gauss–Hermite quadrature to evaluate the \( P_{1,i} \) for \( i \in \{1, 2, 3\} \). The terms \( P_{2,i} \) for \( i \in \{1, 2, 3\} \) are treated analogously.

In terms of complexity of the pricing procedure, note that we have replaced the Monte Carlo simulation of the VIX discretization scheme in the mixed rough Bergomi model, resp. the two-dimensional quadrature in the mixed standard Bergomi model, with one-dimensional Gaussian quadratures.

**Remark 11** Call and put option prices on the squared VIX proxy \( \text{VIX}_{T,P}^2 \), corresponding to \( \psi(x) = (x - \kappa)^+ \) and \( \psi(x) = \)
(κ − x)⁺ in (38), admit explicit expressions in terms of Black–Scholes formulas, provided one evaluates the point \( F^{-1}(κ) \), \( F^{-1} \) being the inverse of the function \( F : x \in \mathbb{R} \rightarrow v(ξ_0)\left[λ e^{(1+\eta_2)u} + (1 - λ) e^{(1+\eta_1)u}\right] \). The evaluation of \( F^{-1}(κ) \) can be performed with a simple root-finding procedure. As a consequence, in the case of call and put options on VIX², the whole expansion (37) boils down to an explicit combination of Black–Scholes prices and greeks.

3.2.2. VIX implied volatility. As done in the previous sections, we compare the reference VIX implied volatility with the approximate VIX implied volatility computed with our expansion in theorem 3.1. To test the approximation formulas on different VIX smiles, we consider two different parameter scenarios in the rough and standard Bergomi models. We have considered options maturities equal to 1, 3, and 6 months.

Implied volatility for the mixed rough Bergomi model: Recall that the mixed rough Bergomi model is obtained by injecting \( K_u = \eta_i(u - t)^{H - \frac{1}{2}} \), \( i \in \{1, 2\} \), in (33). We set \( ξ_0 = 0.2352 \) and \( H = 0.1 \); the other model parameters can be found in table 1. We evaluate the reference option prices with 10⁶ Monte Carlo samples and 300 discretization points (figure 7).

In the mixed model as well, our approximation formula proves to be very accurate: the relative error is less than 1.6% (in absolute value) for scenario 1 and less than 0.9% for scenario 2.

Implied volatility for the mixed standard Bergomi model: We perform a similar numerical analysis for the mixed one-factor standard Bergomi model, obtained setting \( K_u(t) = \omega_i e^{-k(u-t)} \), \( i \in \{1, 2\} \), in (33). We set \( ξ_0 = 0.22 \) and \( k = 1 \). The other model parameters are given in table 2.

Also, in this case, the approximation formula from theorem 3.1 turns out to be extremely accurate (relative errors are less than \( 5 \times 10^{-2} % \) for parameter scenario 3 and less than \( 2 \times 10^{-2} % \) for scenario 4), see figure 8.

3.3. Calibration to VIX market data

In this section, we perform a calibration test of the mixed rough Bergomi model and of the standard Bergomi model to market VIX smiles as of November 22, 2017, using our pricing formula (37). We decide to set \( H = 0.1 \) (resp. \( k = 1 \)) for the mixed rough (resp. for the standard) Bergomi model and calibrate the other free parameters; of course it is also possible to calibrate the parameter \( H \) (or \( k \)).

Let us describe the calibration procedure for the rough model. The model contains the initial forward variance curve \( ξ_0^2 \) and the additional parameters \( (η_1, η_2, λ) \). We can decide
Table 1. Term structure of VIX futures and model parameters for scenarios 1 and 2 in the mixed rough Bergomi model.

| Scenario | 1-month VIX futures | 3-month VIX futures | 6-month VIX futures |
|----------|---------------------|---------------------|---------------------|
| 1        | 0.218650 ± 5 × 10^{-6} | 0.206308 ± 5 × 10^{-6} | 0.196890 ± 5 × 10^{-6} |
| 2        | 0.229001 ± 3 × 10^{-6} | 0.224244 ± 3 × 10^{-6} | 0.220472 ± 3 × 10^{-6} |

Scenario | $\eta_1$ | $\eta_2$ | $\lambda$ |
|----------|----------|----------|----------|
| 1        | 1.4      | 0.7      | 0.3      |
| 2        | 0.9      | 0.0      | 0.6      |

Figure 8. VIX smiles in the mixed standard Bergomi model for $T = 1, 3, 6$ months (left figures), and relative errors between the reference implied volatilities and their approximations (right figures) for parameter scenarios 3 (top figures) and 4 (bottom figures).

Table 2. Term structure of VIX futures and model parameters for scenarios 3 and 4 in the mixed standard Bergomi model.

| Scenario | 1-month VIX futures | 3-month VIX futures | 6-month VIX futures |
|----------|---------------------|---------------------|---------------------|
| 3        | 0.172764            | 0.145976            | 0.130503            |
| 4        | 0.181527            | 0.165480            | 0.155141            |

Scenario | $\omega_1$ | $\omega_2$ | $\lambda$ |
|----------|------------|------------|----------|
| 3        | 0.5        | 6          | 0.3      |
| 4        | 10         | 2          | 0.2      |

Note that we can introduce a term structure also in the parameters $(\eta_1, \eta_2, \lambda)$ making them maturity-dependent and piece-wise constant between $T_i$ and $T_{i+1}$; the $i$-th VIX futures will be attached to its own parameter set $(\xi_i^T, \eta_i^T, \eta_2^T, \lambda^T)$. We calibrate the model sequentially from the shortest to the largest futures maturity; for each maturity, the procedure follows two steps:

to use the variance curve to match the market term structure of VIX futures exactly while using the other parameter to fit the smile of VIX options. We consider the $n = 4$ shortest VIX futures quoted on the observation date, each associated to a maturity $(T_i)_{i=1,...,n}$, with market values $(F_i)_{i=1,...,n}$. Note that we can introduce a term structure also in the parameters $(\eta_1, \eta_2, \lambda)$ making them maturity-dependent and piece-wise constant between $T_i$ and $T_{i+1}$; the $i$-th VIX futures will be attached to its own parameter set $(\xi_i^T, \eta_i^T, \eta_2^T, \lambda^T)$. We calibrate the model sequentially from the shortest to the largest futures maturity; for each maturity, the procedure follows two steps:
Figure 9. Left: term structure of VIX futures. Right: market VIX smiles calibrated using our approximate price formulas in the mixed rough Bergomi model as of November 22, 2017.

Figure 10. Left: term structure of VIX futures. Right: market VIX smiles calibrated using our approximate price formulas in the mixed standard Bergomi model as of November 22, 2017.

(1) For given \((\eta^T_1, \eta^T_2, \lambda^T)\), we set \(\xi^T_0\) as the unique solution to \(F_1 = P^F(T, \xi^T_0, \eta^T_1, \eta^T_2, \lambda^T)\) where \(P^F(\cdot)\) corresponds to the approximate price for VIX futures given by theorem 3.1.

(2) We then compute the VIX implied volatility smile (see figures 9 and 10) using the value found for \(\xi^T_0\) in the previous step and the values of the other parameters \((\eta^T_1, \eta^T_2, \lambda^T)\). We evaluate the \(L^2\) distance between the model implied volatility and the market implied volatility. Until we find a minimum, we go back to step 1.

In the \(L^2\) distance computed in step 2, implied volatilities have equal unitary weights for all strikes and maturities. In our tests, we used the function `scipy.optimize.least_squares` from the `scipy` library (Virtanen et al. 2020) in step 2. The procedure is the same for the mixed standard Bergomi model, replacing \(H\) with \(k\) and \(\eta_1, \eta_2\) with \(\omega_1, \omega_2\).

The calibration of the two models based on our approximate price formula proves to be very efficient and fast. For the standard Bergomi model, in our tests the overall procedure is 3.5 times faster than the calibration based on option pricing with a two-dimensional quadrature. More precisely, starting

| \(T \in \left[\frac{1}{12}, \frac{2}{12}\right]\) | \(T \in \left[\frac{2}{12}, \frac{3}{12}\right]\) | \(T \in \left[\frac{3}{12}, \frac{4}{12}\right]\) | \(T \in \left[\frac{4}{12}, \frac{5}{12}\right]\) |
|---|---|---|---|
| \(\xi^T_0\) | \(1.449 \times 10^{-2}\) | \(2.074 \times 10^{-2}\) | \(2.543 \times 10^{-2}\) | \(2.871 \times 10^{-2}\) |
| \(\eta^T_1\) | \(1.899\) | \(1.887\) | \(1.684\) | \(1.410\) |
| \(\eta^T_2\) | \(0.1937\) | \(0.1481\) | \(0.1482\) | \(0.1166\) |
| \(\lambda^T\) | \(0.3208\) | \(0.4849\) | \(0.5614\) | \(0.6511\) |
from the initial guess $\xi_0^T = 2 \times 10^{-2}$, $\omega_0^T = 1.5$, $\omega_1^T = 0.5$, $\lambda^T = 0.5$, and using quadratures with 120 nodes for the space and time integration, the calibration of the whole implied volatility surface takes about 10 seconds using our approximation formulas as opposed to 36 seconds when using the two-dimensional quadrature. In the mixed rough Bergomi model, the ratio of the two calibration times becomes much higher. Starting from the same initial guess (replacing $\eta_1, \eta_2$ with $\eta_1, \eta_2$) and using again Gaussian quadrature with 120 nodes in our approximation formulas, the calibration shown in figure 9 also took about 10 seconds. The whole calibration based on the Monte Carlo pricing procedure described in remark 8 is likely to take between several minutes and one hour. For comparison, an accurate Monte Carlo pricing for a single VIX futures with $10^6$ samples and 300 discretization points takes approximately 10 seconds.

The overall calibration error (defined as the sum of the mean relative error on the futures prices and the mean relative error on implied volatilities) is 0.42% for the standard Bergomi model and 0.59% for the mixed rough Bergomi model. As a sanity check, we also computed the smiles obtained with the benchmark simulation procedure in remark 8 under the calibrated parameters of tables 3 and 4. The approximation error—that is, the mean relative error of the approximation formulas with respect to the output of the simulation—is negligible with respect to the calibration error (less that one tenth), for both models.

4. Conclusion

We have studied the accuracy of weak approximations for the VIX random variable based on log-normal distributions in different classes of forward variance models, including basic and mixed Bergomi models with different convolution kernels. We have provided expansion results for VIX derivatives that are accurate, showing that the resulting approximation formulas can be exploited for a fast and efficient calibration of the VIX implied volatility surface.

Acknowledgments

The authors gratefully acknowledge financial support from the research projects Chaire Risques Financiers (École Polytechnique, Fondation du Risque and Société Générale) and Chaire Stress Test, Risk Management and Financial Steering (École Polytechnique, Fondation de l’École Polytechnique and BNP Paribas). We thank Julien Guyon, Martino Grasselli, and Mathieu Rosenbaum for feedback and stimulating discussions on the subject of this article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

F. Bourgey http://orcid.org/0000-0002-4097-6687
E. Gobet http://orcid.org/0000-0002-9940-2493

References

Alòs, E., León, J.A. and Vives, J., On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. Finance Stoch., 2007, 11(4), 571–589.
Alòs, E., García-Lorite, D. and Gonzalez, A.M., On smile properties of volatility derivatives: Understanding the VIX skews. SIAM J. Financial Math., 2022, 13(1), 32–69.
Barletta, A., Nicolato, E. and Pagliarani, S., The short-time behavior of VIX-implied volatilities in a multifactor stochastic volatility framework. Math. Finance, 2019, 29(3), 928–966.
Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. Quant. Finance, 2016, 16(6), 887–904. http://dx.doi.org/10.1080/14697688.2015.1099717
Bergomi, L., Smile dynamics. Risk, 2004, 117–123.
Bergomi, L., Smile dynamics II. Risk, 2005, 67–73.
Bergomi, L., Smile dynamics III. Risk, 2008, 90–96.
Bompis, R. and Gobet, E., Analytical approximations of local-Heston volatility model and error analysis. Math. Finance, 2018, 28(3), 920–961.
Bourgey, F., Stochastic approximations for financial risk computations. PhD thesis, Institut Polytechnique de Paris, 2020. https://tel.archives-ouvertes.fr/tel-02983018/
Bourgey, F. and De Marco, S., Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model. J. Comput. Finance., 2022, 26(2), 53–82. http://dx.doi.org/10.21314/JCF.2022.023.
Buehler, H., Consistent variance curve models. Finance Stoch., 2006, 10(2), 178–203.
Chicago Board Options Exchange. The CBOE Volatility Index-VIX, pp. 1–23, 2009. www.cboe.com/micro/vix/vixwhite.pdf.
De Marco, S., Volatility derivatives in (rough) forward variance models. Presentation at the Bachelier World Congress, Dublin, July 2018.
Dupire, B., Model art. Risk, 1993, 6(9), 118–124.
Fouque, J.-P. and Saporito, Y.F., Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. Quant. Finance, 2018, 18(6), 1003–1016.
Fukasawa, M., Asymptotic analysis for stochastic volatility: Martingale expansion. Finance Stoch., 2011, 15(4), 635–654.
A.2. Proof of theorem 2.6

Let $p \geq 1$ and $\epsilon \in [0, 1]$. Starting from (16) and repeatedly applying the generalized Minkowski and Hölder inequalities, we have

$$
\left\| f^{(n)}(\epsilon) \right\|_p \leq \int_{-\infty}^{\infty} \left\| (Y_T^\nu - v_0 (Y_T))^n e^{(Y_T^\nu - v_0 (Y_T))^+} (Y_T) \right\|_p \, du \\
\leq \int_{-\infty}^{\infty} \left\| Y_T^\nu - v_0 (Y_T) \right\|_{3p}^n \times \left\| e^{(Y_T^\nu - v_0 (Y_T))^+} \right\|_3 \, du \\
\leq \left\| e^{v_0 (Y_T)} \right\|_3 \left( \int_{-\infty}^{\infty} \left\| Y_T^\nu - v_0 (Y_T) \right\|_{3p}^n \, du \right)^{\frac{1}{n}} \\
\times \left( \int_{-\infty}^{\infty} e^{(Y_T^\nu - v_0 (Y_T))^+} \, du \right)^{\frac{n}{3}}.
$$

From (14) and (18), we have $\left\| e^{v_0 (Y_T)} \right\|_3 \leq e^{\frac{\nu^2}{\Delta}}$ which is uniformly bounded in $\Delta$ thanks to (22). Using (19) and the boundedness of $u \mapsto e^{\frac{\nu^2}{\Delta}}$, we have

$$
\left\| e^{(Y_T^\nu - v_0 (Y_T))^+} \right\|_p \leq e^{-\frac{\nu^2}{\Delta}} \left( \int_{-\infty}^{\infty} (K(t)^{\nu} - v_0 (K(t)^{\nu})) \, dt \right)^{\frac{n}{3}} \left( \int_{-\infty}^{\infty} (K(t)^{\nu} - v_0 (K(t)^{\nu}))^2 \, dt \right)^{\frac{1}{2}} \\
\leq C_1 e^{-\frac{\nu^2}{\Delta}} \left( \int_{-\infty}^{\infty} (K(t)^{\nu} - v_0 (K(t)^{\nu}))^2 \, dt \right)^{\frac{1}{2}}
$$

for some positive constants $C_1, C_2$ that depend on $p$. Integrating with respect to $u$ and applying (12) and (20), we infer that the third factor on the right-hand side of (A2) is uniformly bounded in $\Delta$. Finally, applying proposition 2.5 with $q = 2n$, we see that the second factor on the right-hand side of (A2) is bounded by $\Delta^{\frac{\nu^2}{\Delta} + \frac{n}{2}}$ up to a multiplying constant, and we obtain the estimate (24).

We now consider estimate (25). Applying the generalized Minkowski inequality to (17), we obtain

$$
\left\| \text{VIX}^2_T - \text{VIX}^2_{T,p} \right\|_p \leq \int_{0}^{1} (1 - \epsilon) \left\| f^{(2)}(\epsilon) \right\|_p \, d\epsilon \\
\leq \sup_{\epsilon \in (0, 1]} \left\| f^{(2)}(\epsilon) \right\|_p,
$$

and we can conclude applying (24) with $n = 2$. ■

A.3. Proof of theorem 2.7

Recall that $X_T^u = \ln(\xi_0^u)$, $Y_T^\nu = X_T^u - X_0^u$, $\text{VIX}^2_T = \nu(\xi_0^u) v_0 (Y_T)^\nu$, $\text{VIX}^2_{T,p} = \nu(\xi_0^u) e^{\nu (Y_T)}$, and let $[\psi]_{H^\delta}$ be the Hölder coefficient of $\psi$.

We are inspired by the techniques of Bompis and Gobet (2018), where the authors provide analytical approximations of option prices with Lipschitz payoffs in a local-Heston volatility model. In our case, to alleviate the possible lack of smoothness of $\psi$ and to overcome some degeneracy problems in the Malliavin sense (see later for more details), we introduce the Gaussian regularization $x \mapsto \psi_\varphi(x) = \mathbb{E}[\psi(x + \vartheta B_T)]$ with a positive parameter $\delta$ defined by

$$
\delta := \Delta^{\frac{1}{2}}(d_1 + \Delta),
$$

and where $B$ is a standard Brownian motion independent of $W$. Note that

$$
\psi_{\varphi,\varphi}(x) = \mathbb{E}[\psi(x + \vartheta B_T + \delta \tilde{B}_T)] = \mathbb{E}[\psi_\varphi(x + \delta B_T)],
$$

where $\tilde{B}$ is another standard Brownian motion such that $B, \tilde{B}, W$ are all independent. The functions $x \mapsto \psi_\varphi(x)$ and $x \mapsto \psi_{\varphi,\varphi}(x)$ are
smooth \((C^\infty(\mathbb{R}))\). Applying Taylor’s theorem with integral remainder to \(\varphi_{\sqrt{\Lambda}}\) at the points \(VIX_{T}^2\) and \(VIX_{\nu}^2\), one gets

\[
\mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{T}^2)] = \mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{\nu}^2)] \\
+ \frac{1}{2} \mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{T}^2 - VIX_{\nu}^2)^2] \\
+ E_0(\varphi_{\sqrt{\Lambda}}).
\]

(9)

Recall that, from (17) and (16), we have the representation formula

\[
VIX_{T}^2 - VIX_{\nu}^2 = \frac{1}{2} VIX_{\nu}^2 \int_0^T (\nu_T^2 - \nu_T (Y_T))^2 v_0 (du) \\
+ \int_0^T (1 - \lambda)^2 \int_0^T (Y) (e) \ dp. 
\]

Now, introducing the new function

\[
\Psi (x) := \varphi_{\sqrt{\Lambda}} (v (\xi_0) e^x),
\]

we have

\[
VIX_{T}^2 \varphi_{\sqrt{\Lambda}} (VIX_{\nu}^2) = v (\xi_0) e^{\nu (Y_T)} \varphi_{\sqrt{\Lambda}} (v (\xi_0) e^{\nu (Y_T)}) \\
= \Psi ^+ (v_0 (Y_T)).
\]

and the second expectation on the right-hand side of (9) can be rewritten as

\[
\mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{T}^2) (VIX_{\nu}^2 - VIX_{T}^2)] \\
= \int_A \mathbb{E}[\Psi ^+ (v_0 (Y_T)) \frac{1}{2} (\nu_T^2 - \nu_T (Y_T))^2] v_0 (du) + E_1(\varphi_{\sqrt{\Lambda}}). 
\]

At A(6)

We are going to further manipulate the expectation term appearing inside the integral \(\int_A v_0 (du)\) on the right-hand side of (9). Recalling (A1), an application of Itô’s formula to the process \((Y_T^2 - v_0 (Y_T))^2\), \(0 \leq t \leq T\), for fixed \(v\) yields

\[
\frac{1}{2} (Y_T^2 - v_0 (Y_T))^2 = \int_0^T (Y_T^2 - v_0 (Y_T)) d(Y_T^2 - v_0 (Y_T)) \\
+ \frac{1}{2} \int_0^T d(Y_T^2 - v_0 (Y_T)) \\
= \int_0^T \left[ \frac{1}{2} K''^2 (s)^2 + \frac{1}{2} \int_0^s (K (s))^2 \right] ds \\
+ (K^2 (s) - v_0 (K (s))) \ dW_s \\
+ \left[ \frac{1}{2} K''^2 (t)^2 + \frac{1}{2} v_0 (K (t))^2 \right] \\
+ (K^2 (t) - v_0 (K (t))) \ dW_t \\
+ \int_0^T \frac{1}{2} (K^2 (t) - v_0 (K (t))) \ dt.
\]

At (7)

In light of (9), the expectation \(\mathbb{E}[\Psi ^+ (v_0 (Y_T)) \frac{1}{2} (Y_T^2 - v_0 (Y_T))^2]\) we want to evaluate is equal to

\[
\mathbb{E}[\Psi ^+ (v_0 (Y_T)) \int_0^T (Y_T^2 - v_0 (Y_T))^2] \\
+ \frac{1}{2} \int_0^T \mathbb{E}[\Psi ^+ (v_0 (Y_T)) d(Y_T^2 - v_0 (Y_T))].
\]

At (A8)

since the quadratic variation \((Y_T^2 - v_0 (Y_T))^2 = \int_0^T \frac{1}{2} (K^2 (t) - v_0 (K (t)))^2 \ dt\) is deterministic. The expectation \(\mathbb{E}[\Psi ^+ (v_0 (Y_T))]\) gives rise to the first derivative of a Black-Scholes price, the variable \(v_0 (Y_T)\) being Gaussian. Our goal is to give a similar formulation for the first expectation in (9), too. To do so, we recall a formula for the integration by parts of functions of iterated Wiener integrals, under a form that is suitable for our purposes.

**Lemma A.2** \(\text{[Gobet and Miri 2014, lemma A.2.]}\) Let \(a_i, e_i, f_i, g_i, h_i : [0, T] \rightarrow \mathbb{R}\) be square-integrable and deterministic processes, and let \(I : \mathbb{R} \rightarrow \mathbb{R}\) be a bounded smooth function with bounded derivatives. Then,

\[
\mathbb{E}[\left( \int_0^T a_i dW_i (t) \right) \int_0^T f_i (t) \ dW_i (t)] \\
= \int_0^T \int_0^T \lambda_i \ d_\iota \mathbb{E}[\left( \int_0^T a_i dW_i (t) \right) (e_i dW_i (t) + f_i dW_i (t));\]

where

\[
\lambda_0 = \int_0^T \int_0^T e_i g_i ds dt, \quad \lambda_1 = \int_0^T \int_0^T (g_i a_i f_i + e_i a_i h_i) ds dt, \quad \lambda_2 = \int_0^T \int_0^T a_i f_i a_i h_i ds dt. 
\]

We set

\[
l (x) = \Psi ^+ \left( x - \frac{1}{2} \int_0^T v_0 (K (t))^2 dt \right), \quad a_i = v_0 (K (t)), \quad f_i (u) = h_i (u) = K^2 (t) - v_0 (K (t)), \quad e_i = g_i (u) = \frac{1}{2} K^2 (t)^2 + \frac{1}{2} v_0 (K (t))^2, \quad \lambda_0 (u) = \int_0^T \int_0^T e_i (u) (g_i (u) + e_i (u) a_i h_i (u)) ds dt, \quad \lambda_1 (u) = \int_0^T \int_0^T (g_i (u) a_i f_i (u) + e_i (u) a_i h_i (u)) ds dt, \quad \lambda_2 (u) = \int_0^T \int_0^T a_i f_i (u) a_i h_i (u) ds dt.
\]

and apply lemma A.2 to the first expectation in (9). Recalling that \(\Psi ^+ (x) = v (\xi_0) e^{\nu (Y_T) + x}\) for every \(x \in \mathbb{R}\), we obtain

\[
\int_A \mathbb{E}[\Psi ^+ (v_0 (Y_T)) \frac{1}{2} (Y_T^2 - v_0 (Y_T))^2] v_0 (du) \\
= \int_0^T \mathbb{E}[\Psi ^+ (v (\xi_0) e^{\nu (Y_T) + x})] v_0 (du) \\
= \sum_{i=1}^3 \gamma_i \mathbb{E}[\varphi_{\sqrt{\Lambda}} (v (\xi_0) e^{\nu (Y_T) + x})] v_0 (du) \\
\]

where

\[
\gamma_1 = \int_A \gamma (0) v_0 (du) + \frac{1}{2} \int_A \int_0^T d(Y_T^2 - v_0 (Y_T)) v_0 (du), \quad \gamma_2 = \int_A \gamma_1 (u) v_0 (du), \quad \gamma_3 = \int_A \gamma_2 (u) v_0 (du).
\]

Putting things together, recalling that \(v (\xi_0) e^{\nu (Y_T)} = VIX_{T}^2\), we have shown that

\[
\mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{T}^2)] = \mathbb{E}[\varphi_{\sqrt{\Lambda}} (VIX_{\nu}^2)]
\]
Before estimating the error terms \( E_1 \) and \( E_0 \) in (A9), we wish to get back to the true payoff function, and provide an identity analogous to (A9) for the true payoff function \( \psi \) instead of its regularized version \( \psi_{\gamma_{3a}} \). We start with some estimates of the coefficients \((\gamma)_{i=1,2,3} \).

**Lemma A.3** For every \( i \in \{1, 2, 3\}, \gamma_i = \mathcal{O}(\Delta^{2d_i/\delta_d}) \) as \( \Delta \to 0 \).

**Proof.** From (19) and (20), we have \( \gamma_1 = \mathcal{O}(\Delta^{2d_1}) + \mathcal{O}(\Delta^{2d_2}) = \mathcal{O}(\Delta^{2d_1/\delta_d}) \). Using the trivial inequality \( x^2 \leq \frac{1}{2}(x^2 + y^2) \) for \( x, y \in \mathbb{R} \) along with the Cauchy–Schwarz inequality, we see that \( |\gamma_2| \) is upper-bounded by

\[
\frac{\sigma_T^2}{2} \int_0^T \left( K(t) - \nu_0(K(t)) \right)^2 \, dt \left| \nu_0(\text{d}u) \right|
\]

Applying (19), (20), and (22), we get \( \gamma_2 = \mathcal{O}(\Delta^{d_2}) + \mathcal{O}(\Delta^{2d_2/\delta_d}) \). Finally, applying once again the Cauchy–Schwarz inequality, (20), and (22), we have

\[
\gamma_3 \leq \frac{\sigma_T^2}{2} \int_0^T \left( K(t) - \nu_0(K(t)) \right)^2 \, dt \left| \nu_0(\text{d}u) \right| = \mathcal{O}(\Delta^{d_3}).
\]

Note that, since \( \psi \) is \( \theta \)-Hölder continuous, the following useful estimate holds

\[
\sup_{x \in \mathbb{R}} |\psi(x) - \psi(0)| \leq |\psi|_{\text{H}^\theta} \left| \int \left( \delta B_T \right)^p \right| = \mathcal{O}(\delta^\theta).
\]

Consequently, in view of (A3) we have

\[
\mathbb{E} \left[ \psi \left( \text{VIX}_T^2 - \psi \left( \text{VIX}_T^2, p \right) \right) \right] = \mathbb{E} \left[ \psi_{\gamma_{3a}} \left( \text{VIX}_T^2 - \psi_{\gamma_{3a}} \left( \text{VIX}_T^2, p \right) \right) + \mathcal{O}(\Delta^{3d_i/\delta_d}) \right].
\]

For \( i \in \{1, 2, 3\} \), we also have

\[
\delta_{\gamma} \left[ \psi_{\gamma_{3a}} \left( \text{VIX}_T^2, p \right) e^\nu \right] \bigg|_{l=0} = \delta_{\gamma} \left[ \psi \left( \text{VIX}_T^2, p \right) e^\nu \right] \bigg|_{l=0} + \text{Error}_1.\]

\[
\text{Error}_1 \leq \delta^\theta \int_0^\infty \left| \delta_{\gamma} \left[ \psi \left( \text{VIX}_T^2, p \right) e^\nu \right] \bigg|_{l=0} \right| \frac{dy}{\sqrt{2\pi \sigma^2 y}}.
\]

The first inequality in (A13) is obtained applying (A10) and proposition 2.3. In order to prove the last identity in (A13), observe that the partial derivatives are related to Hermite polynomials and it is easy to prove that the integral is uniformly bounded with respect to \( \Delta \) thanks to (21) and (22). Putting Lemma A.3, (A9), (A11), and (A13) together, we get that the VIX option price \( E_0[\psi(\text{VIX}_T^2)] \) is equal to

\[
\mathbb{E} \left[ \psi \left( \text{VIX}_T^2, p \right) \right] + \sum_{i=1}^3 \gamma_i \delta_{\gamma} \left[ \psi \left( \text{VIX}_T^2, p \right) e^\nu \right] \bigg|_{l=0} + \mathcal{O}(\Delta^{3d_i/\delta_d}) + E_1 \left( \psi_{\gamma_{3a}} \right) + E_0 \left( \psi_{\gamma_{3a}} \right).
\]

**Error terms:** We now wish to establish estimates for the two error terms \( E_0(\psi_{\gamma_{3a}}) \) and \( E_1(\psi_{\gamma_{3a}}) \) defined in (A5)–(A6). Though the function \( \psi_{\gamma_{3a}} \) is smooth, establishing \( \Delta \)-uniform estimates for its first and second derivatives seem difficult. Therefore, we exploit an integration-by-parts formula to get back to \( \psi_{\gamma_{3a}} \). To do so, we use the calculus of variations with respect to the two-dimensional Brownian motion \((W_t, B_t)_{t \in [0,T]}\). We refer to the notations of Nualart (2006) for the Malliavin Sobolev spaces \( \mathcal{D}_\mathbb{P}^p \) associated to the norm \( \| \cdot \|_{\mathcal{D}_\mathbb{P}^p} \). The Malliavin derivative operator with respect to \( W \) and \( B \) is denoted \( D(\cdot):=\{D_1(\cdot), D_2(\cdot)\}_{t \in [0,T]} \). For the second-order Malliavin derivatives, we use the notation \( (D^2(\cdot))_{t \in [0,T]} \) for \( i, j \in \{1, 2\} \), and so on for higher derivatives. For any \( \lambda \in \{0, 1\} \), we set

\[
G_{\lambda} := \lambda \text{VIX}_T^2 + (1 - \lambda) \text{VIX}_T^2 p + \Delta B_T
\]

recalling (15). Even though the proxy \( \text{VIX}_T^2 p \) is a non-degenerate random variable (owing to (22) in assumption 2.4), with Malliavin covariance matrix \( \sigma_T^2 \text{VIX}_T^2 p > 0 \), the convex combination \( G_{\lambda} \) may be degenerated in the Malliavin sense, and the integration-by-parts formula may not hold. Having introduced the Gaussian perturbation \( \sqrt{2\Delta B_T} \) precisely answers this issue and corresponds to the second reason for having regularized \( \psi \). For these reasons, it is crucial to realize that the convolution identity (A4) yields the new expressions

\[
E_0(\psi_{\gamma_{3a}}) = \mathbb{E} \left[ \psi_{\gamma_{3a}} \left( \text{VIX}_T^2 \right) \int_0^1 \frac{(1 - \varepsilon)^2}{2} f^{(\varepsilon)}(\nu) \, d\nu \right],
\]

\[
E_0(\psi_{\gamma_{3a}}) = \int_0^1 \left( 1 - \lambda \right) \mathbb{E} \left[ \psi_{\gamma_{3a}} \left( \text{VIX}_T^2 \right) \int_0^1 (1 - \varepsilon)^2 f^{(\varepsilon)}(\nu) \, d\nu \right] \, d\lambda,
\]

where we have started from (A5)–(A6), and used (17).

We start with some estimates for the Sobolev norms of \( f^{(\varepsilon)}(\nu) \).

**Lemma A.4** For every \( p \geq 1 \), \( n \in \mathbb{N}, k \in \{1, 2, 3\}, \) the following estimates hold

\[
\sup_{\varepsilon \in [0,1]} \left\| f^{(\varepsilon)}(\nu) \right\|_p \leq \Delta^{n(d_i/\delta_d)}.
\]

**Proof.** It is enough to assume \( p \geq 2 \) thanks to the non-expansivity of the \( L^p \) norm. For every \( \varepsilon \in [0, 1], u \in \mathbb{A}, \) introduce

\[
E^{\varepsilon}(y) := \left( Y_T^p - \nu_0(Y_T)p \right)^n e^{u(Y_T)+u(Y_T-v_0(Y_T))},
\]

\[
f^{(\varepsilon)}(x) = v \left( \varepsilon_{0} \right) \int_\mathbb{A} E^{\varepsilon}(\nu) \nu(\text{d}u).
\]

Proceeding as in the proof of theorem 2.6 and evoking proposition 2.5 with \( p = 2n \), we have that

\[
\left\| \int_\mathbb{A} \left| E^{\varepsilon}(\nu) \right|^2 \nu(\text{d}u) \right\|^4_p \geq \int_\mathbb{A} \left| \left( Y_T^p - \nu_0(Y_T)p \right)^n e^{u(Y_T)+u(Y_T-v_0(Y_T))} \right|^2 \nu(\text{d}u) \right\|^4_p \geq \int_\mathbb{A} \left| Y_T^p - \nu_0(Y_T)p \right|^{2n} \nu(\text{d}u) \right\|^2_p \geq \Delta^{n(d_i/\delta_d)}.
\]

Observing that \( D^1_\mathbb{A}(Y_T^p - \nu_0(Y_T)) = K''(t) - \nu_0(K(t)) \), and applying the chain rule, we have

\[
D^1_\mathbb{A}(f^{(\varepsilon)}(\nu)) = \int_\mathbb{A} \left( K''(t) - \nu_0(K(t)) \right) E^{\varepsilon-1}(\nu) \nu(\text{d}u)
\]
+ \nu_0 (K(t)) \int_A C_n^\theta(e) \nu_0 (du)
+ \int_A \left[ e^{K(t) - \nu_0 (K(t))} \right] C_n^\theta(e) \nu_0 (du) =: nR_{1,n-1}(t) + R_{2,n}(t) + \epsilon R_{1,n}(t).

From Cauchy–Schwarz inequality, it holds
\[ |R_{1,n}(t)|^2 \leq \left( \int_A |K(t) - \nu_0 (K(t))|^2 \nu_0 (du) \right) \times \left( \int_A C_n^\theta(e)^2 \nu_0 (du) \right), \]
and reinserting in \( t \) and evoking (20), (A19), and taking the \( L^p \) norm, we have
\[ \left\| \left( \int_0^T |R_{1,n}(t)|^2 \, dt \right)^{\frac{1}{2}} \right\|_p \leq c \Delta^{\frac{2n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}, \quad n \geq 0. \]
Similarly, combining Cauchy–Schwarz inequality with (18), (22) and (A19),
\[ \left\| \left( \int_0^T |R_{2,n}(t)|^2 \, dt \right)^{\frac{1}{2}} \right\|_p \leq c \sigma_p \left\| \left( \int_A C_n^\theta(e)^2 \nu_0 (du) \right)^{\frac{1}{2}} \right\|_p \leq c \Delta^{\frac{n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}, \]
Noting that \( D_i^n(t^{(\epsilon)}(e)) = 0 \) and plugging estimates into (A20), we get
\[ \left\| \left( \int_0^T |D_i^n(t^{(\epsilon)}(e))|^2 \, dt \right)^{\frac{1}{2}} \right\|_p \leq c \Delta^{\frac{n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}. \]

Using (20), (A19), and Cauchy–Schwarz inequality, we get
\[
\left| S_{1,n}(t,s) \right|^2 \leq \int_A \left[ |K(t) - \nu_0 (K(t))|^2 \right] C_n^\theta(e)^2 \nu_0 (du),
\]
\[ \left\| \left( \int_0^T |S_{1,n}(t,s)|^2 \, dt \right)^{\frac{1}{2}} \right\|_p \leq c \Delta^{\frac{n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}. \]
A similar study gives that
\[ \left\| \left( \int_0^T |S_{2,n}(t,s)|^2 \, dt \right)^{\frac{1}{2}} \right\|_p \leq c \Delta^{\frac{n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}, \]
and using that \( D_i^n(t^{(\epsilon)}(e)) = D_i^n(t^{(\epsilon)}(e)) = 0 \), we obtain (A18) for \( k = 2 \). The same estimate holds for \( k = 3 \), the proof being similar we skip it.

The following lemma contains an estimate for the Sobolev norm of the Malliavin derivative of \( G_1^\theta \). The proof is a direct consequence of (A15), lemma A.4, and (24).

**Lemma A.5** For every \( p \geq 1, k \in \{ 1, 2 \} \), we have
\[ \sup_{\lambda \in [0,1]} \left\| D_{G_1^\theta}^k \big|_{L^p} \right\| = O(1). \]

**Estimate of the Malliavin covariance matrix:** The Malliavin covariance matrix of \( G_1^\theta \) is defined as
\[
\gamma_{G_1} := \int_0^T \left( |D_1(G_1^\theta)|^2 + |D_2(G_1^\theta)|^2 \right) \, dt = \int_0^T \left( |\lambda D_1(I(1)) + (1 - \lambda) \nu_0 (K(t))| I(0) \right)^2 \, dt + \sigma^2 T =: \gamma_{G_1} + \sigma^2 T, 
\]
and is invertible as \( \gamma_{G_1} \geq \sigma^2 T > 0 \). Observing that \( \sigma^2 I(0)^2 = \int_0^T |D_1(I(0)))|^2 \, dt \) and from Hölder inequality and lemma A.4, we have for any \( p \geq 2 \)
\[ \sup_{\lambda \in [0,1]} \left\| \gamma_{G_1} - \sigma^2 I(0)^2 \right\|_p = \sup_{\lambda \in [0,1]} \left\| \left( \int_0^T |\lambda D_1(I(1)) + (1 - \lambda) \nu_0 (K(t))| I(0) \right)^2 \, dt \right\|_p \leq \left\| \left( \int_0^T |\lambda D_1(I(1)) + (2 - \lambda) \nu_0 (K(t))| I(0) \right)^2 \, dt \right\|_p \leq c \Delta^{\frac{n}{d}+n} \left( d_{\epsilon,\sigma}^2 \right)^{\frac{1}{2}}. \]
Using the representation (17). The above estimate allows us to prove the next lemma, which provides uniform estimates for the Sobolev norms of the inverse of the Malliavin covariance matrix.

**Lemma A.6** For any \( p \geq 1, i \in \{ 0, 1, 2 \} \),
\[ \sup_{\lambda \in [0,1]} \left\| (\gamma_{G_1})^{-i} \right\|_{L^p} = O(1). \]

**Proof** For any \( p \geq 1, \lambda \in [0,1] \), using that \( \gamma_{G_1} = \gamma_{G_1} + \sigma^2 T \geq \sigma^2 T > 0 \), we have
\[ E \left[ (\gamma_{G_1})^{-p} \right] = E \left[ (\gamma_{G_1})^{-p} \right] + E \left[ (\gamma_{G_1})^{-p} \right] \leq (\sigma^2 T)^{-p} \left( \sigma^2 I(0)^2 - \gamma_{G_1} \right)^{-p} \leq 2^p E \left[ (\sigma^2 I(0)^2)^{-p} \right]. \]
From (22) and since \( I(0) \) is lognormal with uniformly bounded parameters \( \mu_p, \sigma_p \), the second term of the previous bound is uniformly bounded in \( \Delta \), and using Markov and Hölder inequalities, we
have for any \( q \geq 1 \),

\[
\mathbb{P} \left( \frac{\varepsilon^2 I(0)^2 - \gamma \sigma^2}{\gamma^2} \geq \frac{\sigma^2 I(0)^2}{2} \right) \leq c \left\| I(0)^{-q} \right\|_{2q}^{q} \sigma^2 I(0)^{2} - \gamma \sigma^2 \leq \mathcal{O}(\Delta^{2q} \psi^2) .
\]

Combining (A3), (A22), and choosing \( q = \frac{3}{2} \), (A23) is proved for \( i = 0 \). For the cases \( i \in [1,2] \), applying the chain rule (see Nualart 2006, lemma 2.1.6), we have for every \( i, j \in [0, T] \),

\[
D_{x}^{1}(\gamma_{-1}) = -\frac{D_{y}^{1}(G_{1})(\gamma_{1})(G_{1})}{\gamma_{1}^2} = -2\gamma_{1}^{-2} \int_{0}^{T} D_{y}^{1}(G_{1})(G_{1}) \, dx,
\]

\[
D_{s_{j}}^{0}(\gamma_{-1}) = -D_{s_{j}}^{1}(G_{1})(\gamma_{1})(G_{1}) + 2D_{s_{j}}^{0}(G_{1})(\gamma_{1})(G_{1})^{3},
\]

\[
D_{x}^{2}(\gamma_{-1}) = D_{x}^{2}(G_{1})(\gamma_{1})(G_{1}) - D_{x}^{2}(\gamma_{1})(G_{1}) = 0 ;
\]

we conclude thanks to (A18) and (A23) with \( i = 0 \).

### Integration-by-parts formula and conclusion: We can now conclude the proof of theorem 2.7. The last ingredient is the following Malliavin integration-by-parts formula.

**Proposition A.7** Let \( i \in [1, 2] \), then for any \( H \in \mathbb{D}^{\infty} \) there exist random variables \( (H_{x}, \lambda) \chi_{x}[0,1] \) such that

\[
E \left[ H \phi \left( G_{1} \right) \right] = E \left[ H_{x} \phi \left( G_{1} \right) \right],
\]

where for any \( p \geq 1 \),

\[
\sup_{x \in [0,1]} \left\| H_{x} \phi \right\|_{p} \leq c \left\| H \right\|_{L^p + \frac{3}{2}} .
\]

**Proof** Since \( G_{1} \) belongs to \( \mathbb{D}^{3} \), and is non-degenerate \( (\gamma_{1} > 0) \), the existence of \( H_{x} \lambda \), follows from Nualart (2006, Proposition 2.1.4), and its \( L^p \)-norm is controlled owing to Nualart (2006, Inequality (2.32), p. 102):

\[
\sup_{x \in [0,1]} \left\| H_{x} \phi \right\|_{p} \leq c \left\| H \right\|_{L^p + \frac{3}{2}} .
\]

The second factor on the right-hand side is finite thanks to lemmas A.5, A.6, and to Hölder inequality for the norms \( \left\| \cdot \right\|_{L^p} \) (Nualart 2006, proposition 1.5.6).

Recalling the representation (A16) and applying proposition A.7 with \( H = f^{(1)}(\epsilon) \), there exists \( H_{x} \phi(t) \) such that

\[
E_{1}(\phi \mathcal{F}_{x}) = \int_{0}^{T} \left( 1 - \frac{\varepsilon^2}{2} \right) \mathbb{E} \left[ \phi \left( G_{1} \right) f^{(3)}(\epsilon) \right] \, dx = \int_{0}^{T} \left( 1 - \frac{\varepsilon^2}{2} \right) \phi \left( G_{1} \right) H_{1}(\epsilon(\cdot)) \, dx .
\]

Therefore, applying (A18) with \( n = 3 \), we get

\[
\left\| E_{1}(\phi \mathcal{F}_{x}) \right\|_{2} \leq \sup_{x \in [0,1]} \left\| f^{(3)}(\epsilon) \right\|_{1,2} \leq c \Delta^{3} \psi \left( \frac{\lambda}{2} \right) .
\]

(A24)

In the last step of (A24), we have applied the easy estimate \( \left\| \phi \left( G_{1} \right) \right\|_{p} = \mathcal{O}(1) \), uniformly in \( \Delta \), which follows from (A10) and the sub-linearity of the (Hölder continuous) function \( \psi \).

In order to estimate \( E_{0}(\phi \mathcal{F}_{x}) \), we start from (A17) and apply proposition A.7 with \( H = f^{(2)}(\epsilon) \), \( f^{(2)}(\epsilon) \), which yields

\[
E_{0}(\phi \mathcal{F}_{x}) \leq \int_{0}^{T} \int_{0}^{T} \int_{0}^{T} (1 - \lambda) (1 - \epsilon_{1}) (1 - \epsilon_{2}) \mathbb{E} \times \left[ \phi \left( G_{1} \right) H_{2,\lambda} \left( \epsilon_{1}, \epsilon_{2} \right) \right] \, dx \, de_{1} \, de_{2} \leq c \sup_{\epsilon \in [0,1] \, \epsilon_{1} \in [0,1]} \left\| f^{(2)}(\epsilon) \right\|_{2,\lambda} \leq c \Delta^{4} \psi \left( \frac{\lambda}{2} \right) ,
\]

(A25)

where the last inequality follows from (A18) with \( n = 2 \). Now combining (A14) with (A24) and (A25), theorem 2.7 is proved.

### A.4. Proof of corollary 2.8

From proposition 2.3 and theorem 2.7, we have first

\[
P_{0}^{call} = E \left[ \psi \left( \text{VIX}^{2}_{T} \right) \right] = E \varepsilon \varepsilon N(0,1) \left[ \left( e^{\frac{\sigma}{\gamma} + \frac{2}{2} e^{-\frac{\sigma}{\gamma} + \frac{2}{2} \gamma - \gamma} \epsilon \right) \right] = C B\left( e^{\frac{\sigma}{\gamma} + \frac{2}{2} e^{-\frac{\sigma}{\gamma} + \frac{2}{2} \gamma}, \frac{\sigma}{2} \right) ,
\]

Introducing the new function \( \Psi : \epsilon \in (0, \infty) \rightarrow E \left[ \psi \left( \text{VIX}^{2}_{T} \epsilon \right) \right] \), we obtain the following standard computations yield

\[
P_{1}^{call} = \Psi \left( 0 \right) = \frac{1}{2} e^{\frac{\sigma}{\gamma} + \frac{2}{2} \Delta \text{BS} \left( e^{\frac{\sigma}{\gamma} + \frac{2}{2} \gamma}, \frac{\sigma}{2} \right)} .
\]

\[
P_{2}^{call} = \Psi \left( 0 \right) + e^{\frac{\sigma}{\gamma} + \frac{2}{2} \gamma} \Delta \text{BS} \left( e^{\frac{\sigma}{\gamma} + \frac{2}{2} \gamma}, \frac{\sigma}{2} \right) .
\]

(A26)

For \( (P_{\delta}^{\phi})_{\phi \in [0,1,2,3]} \), computations are the same as for \( (P_{\delta}^{\phi})_{\phi \in [0,1,2,3]} \) with \( k = 0 \). For put options, use the Black–Scholes put–call parity for \( P_{\delta}^{\phi} \), the identity \( \Delta^{\text{BS}}(x, y, \sigma) = -1 + \Delta \text{call}(x, y, \sigma) \) for \( \bar{P}_{\delta}^{\phi} \), and the fact that \( \Gamma \), \( \text{SpeedBS} \) are equal for call and put options with same characteristics. The put–call parity (30) directly follows.

### A.5. Proof of proposition 2.9

**Estimates (19) and (20)**: Easy computations give \( v \left( K \left( t \right) \right) = \frac{\omega^{2}}{4k^{2}} e^{-2k(t - T)} - e^{-2k(t + \Delta - t)} = \frac{\omega^{2}}{4k^{2}} e^{-2k(t - T)} \) and

\[
\int_{0}^{T} \left[ K \left( t \right) - v \left( K \left( t \right) \right) \right] \, dt = \frac{\omega^{2}}{4k^{2}} \Delta \left( 1 - e^{-2kT} \right) \times \left( 2k \Delta e^{2(t - u)} + e^{-2k \Delta} - 1 \right) .
\]

(A26)

Introducing \( v = \frac{\omega - \Delta}{4 \kappa} \), we infer that

\[
\Gamma_{\Delta, \tau, \phi} = \frac{\omega^{2}}{4k^{2}} e^{-2k(t - T)} - e^{-2k(t + \Delta - t)} = \frac{\omega^{2}}{4k^{2}} e^{-2k(t - T)} \times \left( 2k \Delta e^{2(t - u)} + e^{-2k \Delta} - 1 \right) .
\]

(A27)

We easily check that

\[
\lim_{\Delta \to 0} \left( 2k \Delta e^{2k(t - T)} + e^{-2k \Delta} - 1 \right) \Delta^{-2} = 2k \Delta \left( 1 - 2\nu \right) , \quad \text{uniformly} \nu \in [0, 1].
\]

Therefore, using \( \int_{0}^{1} \left[ 1 - 2\nu \right] \, d\nu = \frac{1}{2} \), we obtain the first asymptotics in (31).

1280 E. Bourgey et al.
Similarly, easy computations give \( v(K(t)) = \frac{a}{2\lambda} (e^{-k(T-t)} - e^{-k(T+\Delta-t)}) \) and
\[
\int_0^T \left[ K(u) - v(K(t)) \right]^2 \, du = \frac{a^2}{2\lambda^2} (1 - e^{-2kT}) \Delta \left( 1 - k\Delta e^{k(T-u)} \right)^2.
\]
(A28)

The change of variables \( v = \frac{a-T}{\Delta} \) leads to
\[
\Lambda_{\Delta,T,p} = \frac{a^2}{2k\Delta^2} \Delta \int_0^1 \frac{1}{1 + k\Delta e^{k\Delta v}} \, dv.
\]
and we still conclude using (A27).

Proxy’s mean and variance: From the definition (18) and the previous expression of \( v(K(t)) \), we easily get
\[
\mu_p = X_0 - \int_0^T \frac{a}{2k\Delta^2} e^{-(2k)(T-t)} (1 - e^{-2k\Delta}) \, dt
\]
in addition \( \mu_p \sim X_0 - \frac{a}{2k\Delta^2} (1 - e^{-2kT}) < \infty \), thus (21) holds.

Similarly, for the variance, start from (18) and the previous expression of \( v(K(t)) \), it gives
\[
\sigma_p^2 = \int_0^T \frac{a^2}{2k^2\Delta^2} e^{-(2k)(T-t)} (1 - e^{-2k\Delta})^2 \, dt
\]
clearly \( \sigma_p^2 \sim \frac{a^2}{2k^2\Delta^2} (1 - e^{-2kT}) \in (0, +\infty) \), i.e. (22) holds.

Coefficient \( \gamma_1 \): Write \( \gamma_1 = \gamma_{10} + \gamma_{11} \) where
\[
\gamma_{10} := \frac{1}{8\Delta} \int_T^{T+\Delta} \left( \int_0^T \frac{K^2(u)}{v_0} - (K(u))^2 \right) \left( 1 - e^{-2k\Delta} \right)^2 \, du,
\]
\[
\gamma_{11} := \frac{1}{2\lambda} \int_T^{T+\Delta} \left( \int_0^T \left| K^2(u) - v_0 (K(u))^2 \right| \right) \, du.
\]
From (A26) and (A28), easy computations give
\[
\gamma_{10} = \frac{a^4}{128k^4\Delta^4} \left( \frac{1}{1 + k\Delta e^{k\Delta}} \right) \left( \frac{1}{1 - e^{-2k\Delta}} \right)^2
\]
\[
\times \left( 1 - e^{-2k\Delta} \right)^2 \left( 1 - e^{-4k\Delta} \right)^2,
\]
\[
\gamma_{11} = \frac{a^4}{8k^3\Delta^2} \left( 2 + k\Delta \right) e^{-k\Delta} - 2 + k\Delta
\]
\[
\times \left( 1 - e^{-2k\Delta} \right) \left( 1 - e^{-4k\Delta} \right).
\]

Coefficient \( \gamma_2 \): Again, some easy computations give that for every \( u \in A \),
\[
\int_0^T v(K(t)) \left[ K^2(u) - v(K(t)) \right] \, dt
\]
\[
= \frac{a^2}{2k\Delta^2} \left( 1 - e^{-2kT} \right) \left( 1 - e^{-4k\Delta} \right)
\]
\[
\left( e^{-k\Delta} - 1 + k\Delta e^{k(T-u)} \right).
\]
Recalling (A26) and integrating in \( u \), we then obtain \( \gamma_2 \) after some standard computations. \( \triangleright \) Coefficient \( \gamma_3 \). Squaring the previous equality and reintegrating in \( u \), we obtain \( \gamma_3 \) after some easy computations.

A.6. Proof of proposition 2.11

Estimates (19) and (20): Easy computations first give \( v(K(t)) = \frac{a}{2\lambda} (T+\Delta-t)2H - (T-t)2H \) and
\[
\int_0^T \left[ K^2(t) - v(K(t)) \right] \, dt
\]
\[
= \frac{a^2}{2\lambda^2} \left( T+\Delta-t \right) 2H - (T-t)2H.
\]
\[
= \frac{a^2}{2\lambda^2} \left( \frac{1}{H^2} - u + H \right) 2H + \frac{\Delta + 1}{\Delta(2H+1)}.
\]

Introducing the new variable \( y = \frac{a-T}{\Delta} \), we infer that
\[
\Gamma_{\Delta,T,p} = \frac{a^2}{2H} \left( \int_0^1 \left( T+\Delta \right)^{2H} \right.
\]
\[
- \Delta \left( \frac{1}{H} \right) \frac{\Delta^{2H+1} + \Delta 2H+1 + \left( \Delta + \Delta \right)^{2H+1}}{\Delta(2H+1)} \right) \cdot
\]

Hence, as \( (T+\Delta)^{2H} - \Delta 2H+1 + \Delta 2H+1 + \left( \Delta + \Delta \right)^{2H+1} = HT^{2H-1} \Delta(2(\Delta+1) - 1) \), we have
\[
\Gamma_{\Delta,T,p} = \frac{a^2}{2H} \left( \int_0^1 \left( T+\Delta \right)^{2H} \right.
\]
\[
- \Delta \left( \frac{1}{H} \right) \frac{\Delta^{2H+1} + \Delta 2H+1 + \left( \Delta + \Delta \right)^{2H+1}}{\Delta(2H+1)} \right) \cdot
\]

Since \( f_0^1 \mid 1 - 2y \mid^p dy = \frac{1}{p+1} \), this completes the proof of the asymptotics of \( \Gamma_{\Delta,T,p} \).

Now we handle \( \Lambda_{\Delta,T,p} \). Note that
\[
\Lambda_{\Delta,T,p} = \left( \int_0^{T+\Delta} \left( \int_0^1 \frac{1}{k^2H^2} \frac{1}{(\Delta+t)^H} \right) \, dt \right) \frac{1}{H^2} \Delta\left( \int_0^1 \left( 1 - 2y \right)^p \, dy \right) \frac{1}{2}.
\]

where we have introduced the new variables \( \beta = \frac{a-T}{\Delta}, u = \frac{a-T}{\Delta}, t = \frac{T+\Delta}{2} \), and the function
\[
I : (\Delta, p) \in (0, \infty)^2 \mapsto \left( \int_0^1 \int_0^1 \left( \int_0^1 \left( (\Delta u + t)^H \right)^{2H-1} \right) \, du \right) \frac{1}{H^2} \Delta\left( \int_0^1 \left( 1 - 2y \right)^p \, dy \right) \frac{1}{2}.
\]

Integrating in \( \beta \) and defining the new variable \( s = \frac{a}{\Delta} \), we have
\[
\left( H + \frac{1}{2} \right)^2 \frac{1}{\Delta^{2H}} \left( \int_0^1 \left( 1 + \frac{1}{2} \right) \, ds \right) \frac{1}{H^2} \Delta\left( \int_0^1 \left( 1 - 2y \right)^p \, dy \right) \frac{1}{2}.
\]

where we have set \( f_0 : v \in (0, 1) \mapsto (v + s)^{H-\frac{1}{2}} \) for \( s > 0 \). When \( \Delta \to 0 \), let us show that the double integral above converges to the same integral replacing \( \frac{a}{\Delta} \) by \( +\infty \) and is finite. As the integral is increasing in \( \Delta \), it suffices to show that the square is bounded by a function \( C(s, u) \) such that \( f_0^1 \mid_0^\infty C(s, u) ds \mid^p du \ll \infty \). For \( s \in [0, 1] \), we can take
\[
C(s, u) := 3 \left( 2H+1 + \left( H + \frac{1}{2} \right)^2 \left( (\Delta u)^H + \frac{1}{2} \right) \right)
\]
while for \( s \geq 1 \), applying twice Taylor’s theorem, we choose
\[
\left| f_0^1 \mid_0^\infty \left( f_0^1 \left( 1 - f_0 \right) - f_0^1 \left( u \right) \right) \right|^2.
\]
Applying a second-order Taylor formula with integral remainder to $\varphi$ at the points VIX$^2_T$ and VIX$^2_{T,p}$, we get

$$E[\varphi(\text{VIX}^2_T)] = E[\varphi(\text{VIX}^2_{T,p})]$$

$$+ E[\varphi'(\text{VIX}^2_T)(\text{VIX}^2_T - \text{VIX}^2_{T,p})] + E_0(\varphi''),$$

(A30)

where $E_0(\varphi'')$ is an error term given by

$$E_0(\varphi'') := \int_0^T (1 - \beta) E[\varphi''(\beta \text{VIX}^2_T)]$$

$$+ (1 - \beta) \text{VIX}^2_{T,p} (\text{VIX}^2_T - \text{VIX}^2_{T,p})^2 d\beta.$$  

The expression for $E[\varphi(\text{VIX}^2_{T,p})]$ follows from (41). Noticing that $J'(0) = 0$, we can express $\text{VIX}^2_T - \text{VIX}^2_{T,p} = J(1) - J(0)$ as

$$\frac{J(2)(0)}{2} + \int_0^T (1 - \beta) \int^T \left[ \text{VIX}^2_{T,p} - \text{VIX}^2_T \right]^2 d\beta.$$ 

The variables $\text{VIX}^2_{T,p}$ have been defined in section 3. We introduce the error term

$$E_1(\varphi'') := E \left[ \varphi''(\text{VIX}^2_{T,p}) \int_0^T (1 - \beta) J'''(\beta) d\beta \right]$$

and the functions

$$\Psi_1(\alpha) := \partial_\alpha \varphi \left[ \left( \int_0^T \text{VIX}^2_T (\text{VIX}^2_T - \text{VIX}^2_{T,p}) d\alpha \right) \right]$$

$$\Psi_2(\alpha) := \partial_\alpha \varphi \left[ \left( \int_0^T \text{VIX}^2_{T,p} (\text{VIX}^2_T - \text{VIX}^2_{T,p}) d\alpha \right) \right]$$

Observing that

$$v_0(T,1) = \frac{\eta_1}{2} \int_0^T v_0(K_0(t)^c) dt + \frac{\eta_1}{2} v_0(T,Y_T),$$

and

$$v_0(T,Y_T) = \frac{\eta_2}{2} v_0(T,Y_T)^2 + \frac{\eta_2}{2} v_0(T,Y_T) v_0(T,Y_T)$$

the second term at the right-hand side of (A30) rewrites as

$$E \left[ \varphi''(\text{VIX}^2_{T,p}) (\text{VIX}^2_T - \text{VIX}^2_{T,p}) \right]$$

$$= E \varphi''(\text{VIX}^2_{T,p}) \int_0^T \frac{1}{2} (\text{VIX}^2_{T,p} - v_0(T,Y_T))^2 d\beta$$

$$+ (1 - \lambda) \text{VIX}^2_{T,p} \int_0^T \frac{1}{2} \left( v_0(T,Y_T) - v_0(T,Y_T) \right)^2 d\beta$$

$$= E \left[ \Psi_1(\text{VIX}^2_{T,p}) \right] \int_0^T \frac{1}{2} (\text{VIX}^2_{T,p} - v_0(T,Y_T))^2 d\beta$$

$$+ E \left[ \Psi_2(\text{VIX}^2_{T,p}) \right] \int_0^T \frac{1}{2} \left( v_0(T,Y_T) - v_0(T,Y_T) \right)^2 d\beta$$

where $E_0(\varphi'') = 0$.
\[ +\Psi_2 (v_0 (Y_{T,2})) \int_{\mathcal{A}} \frac{1}{2} \left( Y_{T,2}^2 - v_0 (Y_{T,2}) \right)^2 v_0 \left( du \right) \] + E_1 (\varphi').

This is similar to the setting of (A6), now with two different functions \( \Psi_1 \) and \( \Psi_2 \). Applying twice lemma A.2 to \( E \left[ \Psi_j (v_0 (Y_{T,j})) (Y_{T,j}^2 - v_0 (Y_{T,j}))^2 \right] \) for \( j = 1, 2 \) and fixed \( u \), with

\[ l(x) = \Psi_j \left( x - \frac{1}{2} \int_0^T v_0 \left( K_j (t)^2 \right) dt \right), \quad a_t = v_0 \left( K_j (t) \right), \]
\[ f_t (u) = h_t (u) = K_j^a (t) - v_0 \left( K_j (t) \right), \]
\[ e_t (u) = g_t (u) = - \frac{1}{2} K_j^a (t)^2 + \frac{1}{2} v_0 \left( K_j^a (t)^2 \right). \]

we obtain the desired expansion.

Observing that \( J^{(j)} (\varepsilon) = \lambda J_1^{(j)} (\varepsilon) + (1 - \lambda) J_2^{(j)} (\varepsilon) \), where for each \( j \in \{1, 2\} \) the function \( f_t (u) = v_0 (\xi_t) \int_{\mathcal{T}} e^{v_0 (Y_{T,j}) + \varepsilon (Y_{T,j} - v_0 (Y_{T,j}))} v_0 \left( du \right) \) is defined precisely as in (15), and conducting the same error analysis as in the proof of theorem 2.7 (without the integration by parts of Malliavin calculus, since here \( \varphi \in C^2_b \) is smooth), we retrieve the error estimate \( O\left( \Delta t^{3/2} \right) \).

\[ \square \]