Unified moment-based modelling of integrated stochastic processes

Ioannis Kyriakou
Faculty of Actuarial Science & Insurance, Bayes Business School (formerly Cass), City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK ioannis.kyriakou@city.ac.uk

Riccardo Brignone
Department of Quantitative Finance, Faculty of Economics and Behavioural Science, University of Freiburg, Rempartstr. 16, 79098 Freiburg i. Br., Germany riccardo.brignone@finance.uni-freiburg.de

Gianluca Fusai
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone 18, 28100 Novara, Italy & Faculty of Finance, Bayes Business School (formerly Cass), City, University of London, London EC1Y 8TZ, UK gianluca.fusai@uniupo.it & gianluca.fusai.1@city.ac.uk

In this paper we present a new method for simulating integrals of stochastic processes. We focus on the nontrivial case of time integrals, conditional on the state variable levels at the endpoints of a time interval, through a moment-based probability distribution construction. We present different classes of models with important uses in finance, medicine, epidemiology, climatology, bioeconomics and physics. The method is generally applicable in well-posed moment problem settings. We study its convergence, point out its advantages through a series of numerical experiments and compare its performance against existing schemes.

Key words: Stochastic volatility; linear and nonlinear reducible models; Pearson curves; moments; simulation

1. Introduction

Time integrals of stochastic processes and their simulation feature in numerous research problems in finance, medicine, epidemiology, technology, engineering, bioeconomics and physics; the cases mentioned next are certainly non-exhaustive. In financial engineering, time integrals appear in stochastic volatility models, average options, volatility options and interest rate derivatives. In turbulent diffusion modelling and related phenomena, the position of a fluid particle at a certain time is given by the integrated velocity (e.g., see Obukhov 1959). Sums of random variables arise in wireless communications and related areas (see Nadarajah 2008), portfolio credit risk applications such as portfolio loss process modelling (see Giesecke et al. 2011 and Dassios and Zhao 2017), but also specialized areas of biomedical engineering involving signal averaging. Within the family of sigmoidal growth models, a stochastic Verhulst population model involves a stochastic integral in its explicit solution that requires accurate simulation for generating probabilistic forecasts in fields

including geoscience, oncology to describe tumor growth (e.g., see Laird 1964, and later research) or epidemic dynamics (e.g., see Shen 2020, Wu et al. 2020). A similar simulation challenge appears in a randomized Schaefer model used to describe the growth of populations living in a randomly varying environment and being harvested, but also in a Ginzburg–Landau model with uses in describing phase transition for superconductivity and other broadened applications over the years.

A long-lasting concern in the cases described earlier remains the efficient simulation of stochastic time integrals or the even more involved, as will become clearer later, *conditional* stochastic integrals. Our analysis encompasses stochastic volatility models, such as the Heston and double Heston, stochastic alpha-beta-rho (SABR), Ornstein–Uhlenbeck stochastic volatility (OU-SV), 3/2 and 4/2, but also linear models with multiplicative noise and nonlinear reducible models, such as the stochastic Verhulst, Gordon–Schaefer and Ginzburg–Landau. We extend our application to certain model variants with jumps, such as Bates, Duffie–Pan–Singleton (DPS), time-changed Lévy and self-exciting point processes. Our goal is to unify modelling and, eventually, simulation in a practicable manner that is fast and accurate.

For years, discretization of stochastic differential equations (SDEs) has featured in the literature (see, for example, Chen et al. 2012 for a review) as a possible way to go round the simulation of integrated processes, inevitably yielding a bias, which can be hard to quantify accurately, besides rendering the procedure particularly tedious. To circumvent this, attempts have been made to simulate *exactly* or, perhaps more precisely phrased, recover the $O(s^{-1/2})$ convergence rate of an unbiased Monte Carlo estimation with a total computational budget s, such as Broadie and Kaya (2006), Cai et al. (2017), Kang et al. (2017) and Li and Wu (2019), for different models. These approaches have proved to be able to produce accurate results. At their core, they rely on simulating an integrated process over a time interval conditional on its level at the interval endpoints. Although accurate, they manifest themselves into a serious demerit, as pointed out in the seminal work of Broadie and Kaya (2006), that is, the implicit need to recover the unknown distribution function of the conditional integrated process using numerical inversion of the associated Laplace transform. This can become a heavy load and almost impracticable when generating entire sample trajectories, aside from potentially introducing error, and thereby bias, and computational burden increases during numerical integration. Luckily, recent contributions, for example, by Cai et al. (2014a) have endowed us with computable error bounds of the Laplace transform inversions guaranteeing their accuracy; nevertheless, computational speed remains an issue in view also of its trade-off with accuracy, especially when integrating into a Monte Carlo simulation application. This still hinders the way between the method and the user, leaving space for further research.

Our contribution is summarized as follows. We present a unified methodological framework for modelling (conditional) integrated processes. This comprises an accurate probability curve fitting

approach based on the moments of the associated true probability distributions and a mechanism for measuring the resulting error in end quantities of interest, such as expected value functionals, in a multi-period context and different true underlying models. The true distributions are unique by their moments, i.e., they are moment-determinate, resulting in bona fide approximations whose moments are steered to the corresponding true ones. In fact, this unravels the versatility of our method, which extends to other distributions that are unique by their moments, beyond integrated processes. A deterministic recursive procedure allows us to calculate the exact error but also derive from it a computable upper bound. Although we are exactly fitting a certain number of moments, we show that the differences between the higher moments of the two distributions are insignificant, ensuring, by moment determinacy, a very accurate outcome.

We adopt a system of Pearson curves being tractable, versatile and fast in selection for varying levels of skewness and kurtosis, in parameter estimation and simulation. This choice is driven by these merits but is not restrictive as our proposed framework is built independently. The resulting simulation methodology is convergent and very accurate; we employ fast computation of higher moments based on Choudhury and Lucantoni (1996) and bypass at any stage computationally intensive Laplace transform inversion or differentiation for the computation of the moments. Another notable merit of it is that the size Δt of the time interval does not affect accuracy, as there is not any time-discretization involved. Therefore, it serves as an ideal substitute of approximations that require a large number of time steps (small Δt) to potentially secure enough accuracy, when actually a single time step of arbitrary size suffices and entire sample trajectories, with resulting undesired time increases, are unnecessary; it can still be used when access to sample trajectories is intended.

Our framework makes it possible to investigate important applications in different areas. More specifically, due to increased problem dimensionality, Monte Carlo simulation remains the method of choice for computing expected values of nonlinear functions of driving processes on several occasions, including cases of path-dependence, advanced stochastic volatility models and self-exciting point processes, where other solutions are inexistent or slow to compute. Therefore, we consider, first, the evaluation of path-dependent contracts, encompassing path-independence as a special case. Second, we explore applications in, typically, non-finance models. We cast a spotlight on the stochastic logistic model and present an illustrative simulation case study. We also revisit the Ginzburg–Landau model and its simulation and demonstrate our method's capability under stressed volatility conditions.

The remainder of the paper is structured as follows. In Section 2, we introduce moment problem determinacy that represents the cornerstone of our probability fitting approach to different models; we give also a complete classification of the Pearson distributions. In Section 3, we present relevant

financial models. In Section 4, we present our main theoretical result focused on the study of error in ultimate quantities of interest. In Section 5, we portray our random number generation mechanism using our moment-based probability distribution build-up. Section 6 presents our numerical study focused on financial applications, whereas in Section 7 we extend to other models and unfold the applicability of our simulation method through practical examples in areas such as oncology and bioeconomics. Section 8 concludes the paper. Supplementary results, materials and some additional applications are deferred to the e-companion.

2. The moment problem and moment-based approximations: the case of Pearson curves

Consider a set Ω equipped with a σ-algebra $\mathcal{F}$ and some random variable Ψ on Ω that has an unknown cumulative distribution function G but known moments $\mu_n := \mu_n(G)$. The moment problem is about deciding whether or not a given distribution is uniquely determined by the sequence of its moments.

Definition 1 (Stoyanov 2013). Let $\{\mu_0 = 1, \mu_1, \mu_2, \ldots\}$ be a sequence of real numbers and $\mathcal{I}$ a fixed interval, $\mathcal{I} \subset \mathbb{R}^1$. Suppose there is at least one distribution function $G(x)$, $x \in \mathcal{I}$, such that

$$\mu_n = \int_{\mathcal{I}} x^n dG(x), \quad n = 0, 1, 2, \ldots.$$

If G is uniquely specified by $\{\mu_n\}$, we say that the moment problem is determinate, i.e., the distribution is uniquely determined by its moments; otherwise the moment problem is indeterminate.

There are several sufficient conditions for the moment problem to be determinate or not. We consider the most relevant criterion to us.

Criterion 1 (Stoyanov 2013). Let $G(x)$, $x \in \mathbb{R}^1$, be a distribution function whose characteristic function is τ-analytic, i.e., it can be represented by a convergent power series in the interval $(-\tau, \tau)$ for some $\tau > 0$. This happens iff $\overline{\lim}_{n\to\infty} (\mu_{2n})^{1/(2n)} / (2n) < \infty$.

Then, G is uniquely determined by its moment sequence $\{\mu_n\}$ and the moment generating function exists for $|a| < c$, $c > 0$.

In this paper, we consider probability distributions that are unique by their moments, in particular, moment-determinate in the Stieltjes sense ($\mathcal{I} = [0, \infty)$), and satisfy Criterion 1. We propose a general distribution approximation whose accuracy, given moment determinacy, increases with n, but we postpone further related discussion until later in the paper.

Having explored several possibilities as part of our preliminary study, we singled out Pearson's system of distributions due to its simplicity, accuracy and fast member selection and parameter estimation for given moments, as opposed to the Johnson family of distributions (see Devroye 1986).

Other algorithms related to scale mixtures and series expansion techniques (e.g., see Hörmann et al. 2004, p. 325; Abramowitz and Stegun 1968, p. 935; Lindsay et al. 2000) were found quite inaccurate or even divergent with increasing number of moments (see Fusai and Tagliani 2002). Hence, we adhere to Pearson distributions, which we will generally denote by $\tilde{G}$, as our choice for implementing moment matching and eventually drawing random numbers from. This choice does not constrain the applicability of our proposed framework, which is built independently, as will become apparent along the way.

Let $\tilde{g}$ be the density function associated with $\tilde{G}$ and first four finite raw moments $\{\mu_1, \mu_2, \mu_3, \mu_4\}$. Let $\tilde{g}(x) = \bar{g}(x - \mu_1)$, where $\bar{g}(x)$ satisfies the differential equation

$$\frac{d\bar{g}(x)}{dx} = -\frac{c_0 + x}{c_1 + c_2 x + c_3 x^2}\bar{g}(x). \tag{1}$$

Solving equation (1) yields well-defined density functions of general form

$$\bar{g}(x) = \mathcal{C}\left(c_1 + c_2 x + c_3 x^2\right)^{-\frac{1}{2c_3}} \exp\left\{\frac{(c_2 - 2c_0c_3)\arctan\left(\frac{c_2+2c_3x}{\sqrt{4c_1c_3-c_2^2}}\right)}{c_3\sqrt{4c_1c_3 - c_2^2}}\right\},$$

where $\mathcal{C}$ is the normalizing constant and $\{c_0, c_1, c_2, c_3\}$ are the parameters that control the shape of the distribution. These are calculated during the distribution fitting and are given by

$$c_0 = c_2 := \frac{\sqrt{\beta\gamma}(\varepsilon + 3)}{10\varepsilon - 12\gamma - 18}, \quad c_1 := \frac{(4\varepsilon - 3\gamma)\beta}{10\varepsilon - 12\gamma - 18}, \quad c_3 := \frac{2\varepsilon - 3\gamma - 6}{10\varepsilon - 12\gamma - 18},$$

where

$$\beta := \mu_2 - \mu_1^2, \quad \gamma := \frac{(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3)^2}{(\mu_2 - \mu_1^2)^3}, \quad \varepsilon := \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} \tag{2}$$

are, respectively, the variance, squared skewness and kurtosis of the Pearson random variable. It may be worth noting that, although numerical methods exist making possible the Pearson system to fit more than four moments, they inevitably impact the computational effort (see Rose and Smith 2002, Chapter 5) therefore we do not consider them here. Later in Section 6, we study the closeness of the higher-order moments of a four-moment Pearson curve fit to the true ones and favourably show that, in practice, the differences remain very small also under challenging model parameterizations.

Given knowledge of the first four integer moments, we select a Pearson distribution family type based on the η-criterion by Elderton and Johnson (1969):

$$\eta := \frac{\gamma(\varepsilon + 3)^2}{4(4\varepsilon - 3\gamma)(2\varepsilon - 3\gamma - 6)}.$$

In particular, we get from it the *main* types I ($\eta<0$), IV ($0<\eta<1$) and VI ($\eta>1$); and the *transition* types, i.e., normal ($\eta=0,\varepsilon=3$), II ($\eta=0,\varepsilon<3$), III ($\eta=\pm\infty$), V ($\eta=1$) and VII ($\eta=0,\varepsilon>3$).

As said in the introduction, in this paper we focus primarily on the random variable Ψ representing (conditional) stochastic time integrals. This does not impose a limitation as, based on what said earlier, we can generally apply to moment-determinate problems; in fact, in Section EC.8.1 of the e-companion we present an additional application to the Carr–Geman–Madan–Yor (CGMY) model. In the next section, we focus the spotlight on our major application on stochastic volatility models, whereas in later sections we turn our attention to other prevalent classes of models.

3. Stochastic volatility models

Let $(\Omega,\mathcal{F},Q,\{\mathcal{F}_t\})$ be a filtered probability space where the filtration satisfies the usual conditions with $\mathcal{F}_0$ trivial. This filtered probability space supports all the processes we encounter in the sequel and Q denotes the risk neutral probability measure.

We consider a stochastic model $\left(S\left(t\right),V\left(t\right)\right)_{t\geq0}$ where S and V denote, respectively, the asset price and variance (or volatility) processes. The model is generally given by

$$\begin{cases} dS(t)=rS(t)dt+\beta\left(V(t)\right)S(t)\left(\rho dW_2(t)+\sqrt{1-\rho^2}dW_1(t)\right) \\ dV(t)=\alpha\left(V(t)\right)dt+\gamma\left(V(t)\right)dW_2(t) \end{cases}, \tag{3}$$

where W_1 and W_2 are independent standard Brownian motions, r is the continuously compounded risk-free interest rate and $\rho\in[-1,1]$ the instantaneous correlation between the two processes. The functions $\beta\left(\cdot\right)$ and $\gamma\left(\cdot\right)$ are continuously differentiable and the quotient $\beta\left(\cdot\right)/\gamma\left(\cdot\right)$ is locally integrable in the state space of V and $Q\left(\int_u^t\beta^2\left(V(s)\right)ds<\infty\right)=1$, $0\leq u<t<\infty$ (see Mijatović and Urusov 2012, Cui et al. 2017). Following Cui et al. (2021), we consider the auxiliary functions

$$w\left(x\right):=\int\frac{\beta\left(z\right)}{\gamma\left(z\right)}dz,\quad h\left(x\right):=\alpha\left(x\right)w'\left(x\right)+\frac{1}{2}\gamma^2\left(x\right)w''\left(x\right),$$

where the first resembles the Lamperti transform and the second follows from applying the infinitesimal generator of $V(t)$ to w. The following useful representation then holds:

$$\int_u^t\beta\left(V(s)\right)dW_2(s)=w\left(V(t)\right)-w\left(V(u)\right)-\int_u^t h\left(V(s)\right)ds,$$

based on which the solution of (3) can be written as

$$S(t)=\exp\left(m\left(u,t\right)+\sqrt{1-\rho^2}\int_u^t\beta\left(V(s)\right)dW_1(s)\right),$$

where

$$m\left(u,t\right):=\ln S(u)+r\left(t-u\right)-\frac{1}{2}\int_u^t\beta^2\left(V(s)\right)ds+\rho\left[w\left(V(t)\right)-w\left(V(u)\right)-\int_u^t h\left(V(s)\right)ds\right]. \tag{4}$$

Thus, the following conditional normal distribution $\mathcal{N}(m, s^2)$, for mean m and variance s^2, applies:

$$\left(\ln S(t) \middle| \ln S(u), V(u), V(t), \int_u^t \beta^2\left(V(s)\right) ds\right) \sim \mathcal{N}\left(m\left(u,t\right), s^2\left(u,t\right)\right), \tag{5}$$

where

$$s^2\left(u,t\right) := \left(1-\rho^2\right) \int_u^t \beta^2\left(V(s)\right) ds. \tag{6}$$

We summarize typical examples of the form (3) in Table 1, where $\Psi\left(u,t\right)$ corresponds to the integrated process(es) of interest following from $\int_u^t \beta^2\left(V(s)\right) ds$ for each model and represents a key quantity in our study later and $\Phi\left(t\right)$ to the relevant conditioning arguments based on V. Extended model constructions with independent jumps in the asset price process (Bates 1996), contemporaneous asset price and variance jumps of correlated magnitudes (Duffie et al. 2000) and time-changed Lévy models (Carr et al. 2003) are well known, hence, in the interest of space, we omit their details here and defer to Section 5.3 some more notes on their simulation.

We consider three typical cases of stochastic variance processes. The first is described by a Cox et al. (1985) (CIR) square-root diffusion with constant parameters θ, k, v corresponding to $dV(t) = k(\theta - V(t))dt + v\sqrt{V(t)}dW_2(t)$ in (3), implying $\alpha\left(x\right) = k(\theta - x)$ and $\gamma\left(x\right) = v\sqrt{x}$. If the Feller condition, $2k\theta \geq v^2$, is satisfied, the zero boundary is unattainable. Otherwise, it is attracting and attainable; at the zero boundary though, the process is immediately reflected into the positive domain. The variance transition is given by

$$V(t) \overset{(\text{law})}{=} \frac{v^2(1-e^{-k(t-u)})}{4k} \chi'^2_d\left(\frac{4ke^{-k(t-u)}V(u)}{v^2(1-e^{-k(t-u)})}\right),$$

where $\chi'^2_d\left(\lambda\right)$ is a noncentral chi-squared random variable with $d := 4\theta k/v^2$ degrees of freedom and noncentrality parameter λ. This process is used in models such as Heston (Heston 1993) (and its multifactor extension), 3/2 (Heston 1997, Platen 1997) and 4/2 (Grasselli 2017). In the second case, the volatility, instead, evolves according to a driftless geometric Brownian motion, i.e., $dV(t) = vV(t)dW_2(t)$, hence $\alpha\left(x\right) = 0$ and $\gamma\left(x\right) = vx$. In this case, we have that

$$\ln V(t) \sim \mathcal{N}\left(\ln V(u) - \frac{1}{2}v^2(t-u), v^2(t-u)\right). \tag{7}$$

This volatility specification appears in the SABR model. In the last case we consider, the volatility is represented by a Gaussian OU model (Scott 1987, Stein and Stein 1991, Schöbel and Zhu 1999), $dV(t) = k(\theta - V(t))dt + vdW_2(t)$, based on which $\alpha\left(x\right) = k(\theta - x)$ and $\gamma\left(x\right) = v$. Here, we have (e.g., see Li and Wu 2019) that

$$\left(V(t), \int_u^t V(s)ds\right) \sim \mathcal{N}_2\left(\underline{\mu}\left(u,t\right), \Sigma\left(u,t\right)\right),$$

Table 1 Key quantities in simulation of different models

Heston model
$\beta(x)=\sqrt{x},\ w(x)=\frac{x}{v},\ h(x)=\frac{k\theta}{v}-\frac{k}{v}x$
$\Phi(t)=V(t),\ \Psi(u,t)=\int_u^t V(s)ds$
$m(u,t)=\ln S(u)+r(t-u)-\frac{\Psi(u,t)}{2}+\frac{\rho}{v}\left(\Phi(t)-\Phi(u)-k\theta(t-u)+k\Psi(u,t)\right)$
$s^2(u,t)=(1-\rho^2)\Psi(u,t)$
Double Heston model
$\beta(x)=\sqrt{x},\ w(x)=\frac{x}{v_j},\ h(x)=\frac{k_j\theta_j}{v_j}-\frac{k_j}{v_j}x$
$\Phi(t)=(V_j(t))_j,\ \Psi(u,t)=\left(\int_u^t V_j(s)ds\right)_j$
$m(u,t)=\ln S(u)+r(t-u)+\sum_{j=1}^2\left(\frac{\rho_j}{v_j}\left(\Phi_j(t)-\Phi_j(u)-k_j\theta_j(t-u)+k_j\Psi_j(u,t)\right)-\frac{\Psi_j(u,t)}{2}\right)$
$s^2(u,t)=\sum_{j=1}^2(1-\rho_j^2)\Psi_j(u,t)$
4/2 model
$\beta(x)=a\sqrt{x}+\frac{b}{\sqrt{x}},\ w(x)=\frac{a}{v}x+\frac{b}{v}\ln x,\ h(x)=\frac{k(a\theta-b)}{v}-\frac{ka}{v}x+\left(\frac{k\theta}{v}-\frac{v}{2}\right)\frac{b}{x}$
$\Phi(t)=V(t),\ \Psi(u,t)=\left(\int_u^t V(s)ds,\int_u^t\frac{ds}{V(s)}\right)$
$m(u,t)=\ln S(u)+\left(r-ab-\frac{a\rho k\theta}{v}+\frac{b\rho k}{v}\right)(t-u)+\frac{a\rho}{v}\left(\Phi(t)-\Phi(u)\right)+\frac{b\rho}{v}\ln\frac{\Phi(t)}{\Phi(u)}$
$+\left(\frac{a\rho k}{v}-\frac{a^2}{2}\right)\Psi_1(u,t)+\left[\frac{b\rho}{v}\left(\frac{v^2}{2}-k\theta\right)-\frac{b^2}{2}\right]\Psi_2(u,t)$
$s^2(u,t)=(1-\rho^2)\left(a^2\Psi_1(u,t)+b^2\Psi_2(u,t)+2ab(t-u)\right)$
SABR model
$\beta(x)=x,\ w(x)=\frac{x}{v},\ h(x)=0$
$\Phi(t)=V(t),\ \Psi(u,t)=\frac{1}{\int_u^t V^2(s)ds}$
$m(u,t)=\ln S(u)-\frac{1}{2\Psi(u,t)}+\frac{\rho}{v}(\Phi(t)-\Phi(u))$
$s^2(u,t)=\frac{1-\rho^2}{\Psi(u,t)}$
OU-SV model
$\beta(x)=x,\ w(x)=\frac{x^2}{2v},\ h(x)=\frac{v}{2}+\frac{k\theta}{v}x-\frac{k}{v}x^2$
$\Phi(t)=\left(V(t),\int_u^t V(s)ds\right),\ \Psi(u,t)=\int_u^t V^2(s)ds$
$m(u,t)=\ln S(u)+\left(r-\frac{\rho v}{2}\right)(t-u)+\frac{\rho}{2v}(\Phi_1^2(t)-\Phi_1^2(u))-\frac{\rho k\theta}{v}\Phi_2(u,t)+\left(\frac{\rho k}{v}-\frac{1}{2}\right)\Psi(u,t)$
$s^2(u,t)=(1-\rho^2)\Psi(u,t)$
Linear & Reducible SDEs
$\Phi(t)=Y(t;a,b,\gamma)$
$\Psi(u,t)=\frac{1}{\int_u^t Y(s;a,b,\gamma)ds}$

Notes. For more details regarding the functions $\beta(x)$, $w(x)$, $h(x)$ for the Heston, Double Heston, 4/2, SABR and OU-SV models, refer to Section 3. For more information regarding the linear and reducible models, refer to Section 7. $\Psi(u,t)$ is the key random variable in each model whose conditional distribution given $\Phi(t)$ we are approximating. Where applicable, $m(u,t)$ and $s^2(u,t)$ correspond to the mean and variance of $\left(\ln S(t)\left|\ln S(u),V(u),V(t),\int_u^t\beta^2(V(s))\,ds\right.\right)$.

where $\mathcal{N}_2$ is a bivariate normal distribution with mean vector and covariance matrix

$$\underline{\mu}(u,t):=\begin{pmatrix}(V(u)-\theta)(1-\zeta(u,t))+\theta\\ \theta(t-u)+\frac{(V(u)-\theta)\zeta(u,t)}{k}\end{pmatrix}\ \text{and}\ \Sigma(u,t):=\begin{pmatrix}\frac{v^2\zeta(u,t)(2-\zeta(u,t))}{2k} & \frac{v^2\zeta^2(u,t)}{2k^2}\\ \frac{v^2\zeta^2(u,t)}{2k^2} & \frac{-v^2\zeta^2(u,t)+2kv^2\left(t-u-\frac{\zeta(u,t)}{k}\right)^2}{2k^3}\end{pmatrix},$$

respectively, for $\zeta(u,t):=1-\exp(-k(u-t))$. This requires augmenting the conditioning arguments in (5) to encompass $\int_u^t V(s)ds$.

An extension of the model specification (3) to two stochastic variance factors is due to Christoffersen et al. (2009) (a multifactor generalization is also possible), with (4) and (6) becoming

$$\begin{cases} m(u,t)=\ln S(u)+r(t-u)+\sum_{j=1}^2\left\{-\frac{1}{2}\int_u^t\beta^2(V_j(s))\,ds+\rho_j\left[w(V_j(t))-w(V_j(u))-\int_u^t h(V_j(s))\,ds\right]\right\}\\ s^2(u,t)=\sum_{j=1}^2(1-\rho_j^2)\int_u^t\beta^2(V_j(s))\,ds\end{cases}.$$

Another generalization of (3) is the SABR model (Hagan et al. 2002) with the forward asset price dynamics governed by a constant elasticity of variance (CEV) diffusion process with elasticity parameter $b \in [0,1]$ and volatility evolving according to (7):

$$\begin{cases} dS(t) = V(t)S^b(t)\left(\rho dW_2(t) + \sqrt{1-\rho^2}dW_1(t)\right) \\ dV(t) = vV(t)dW_2(t) \end{cases}. \tag{8}$$

The boundary case of $b = 1$ is straightforward and conforms to formulation (3). This opposes, though, to the more challenging case of $b \in [0,1)$ as an exact closed-form solution to (8) is not generally available. More specifically, for $\rho = 0$ and $(S(t), t \geq 0)$ with an absorbing boundary at 0 (Islah 2009, Cai et al. 2017), the exact distribution

$$Q\left(S(t) = 0 \middle| S(u), V(u), V(t), \int_u^t \beta^2\left(V(s)\right) ds\right) = 1 - Q_{\chi^2}\left(A_0; \frac{1}{1-b}\right), \tag{9}$$

$$Q\left(S(t) \leq y \middle| S(u), V(u), V(t), \int_u^t \beta^2\left(V(s)\right) ds\right) = 1 - Q_{\chi'^2}\left(A_0; \frac{1}{1-b}, C_0(y)\right) \tag{10}$$

for any $y > 0$ holds, where

$$A_0 := \frac{\left(\frac{S(u)^{1-b}}{1-b}\right)^2}{\int_u^t \beta^2\left(V(s)\right) ds}, \quad C_0(y) := \frac{\left(\frac{y^{1-b}}{1-b}\right)^2}{\int_u^t \beta^2\left(V(s)\right) ds},$$

and $Q_{\chi^2}(\cdot; d)$ and $Q_{\chi'^2}(\cdot; d, \lambda)$ denote, respectively, the chi-squared and noncentral chi-squared cumulative distribution functions. When $\rho \neq 0$ and $(S(t), t \geq 0)$ has an absorbing boundary at 0, we typically resort to the approximate distribution

$$Q\left(S(t) = 0 \middle| S(u), V(u), V(t), \int_u^t \beta^2\left(V(s)\right) ds\right) \approx 1 - Q_{\chi^2}\left(A; 1 + \frac{b}{(1-b)(1-\rho^2)}\right), \tag{11}$$

$$Q\left(S(t) \leq y \middle| S(u), V(u), V(t), \int_u^t \beta^2\left(V(s)\right) ds\right) \approx 1 - Q_{\chi'^2}\left(A; 1 + \frac{b}{(1-b)(1-\rho^2)}, C(y)\right) \tag{12}$$

for $y > 0$, where

$$A := \frac{\left(\frac{S(u)^{1-b}}{1-b} + \rho\left[w\left(V(t)\right) - w\left(V(u)\right)\right]\right)^2}{(1-\rho^2)\int_u^t \beta^2\left(V(s)\right) ds}, \quad C(y) := \frac{\left(\frac{y^{1-b}}{1-b}\right)^2}{(1-\rho^2)\int_u^t \beta^2\left(V(s)\right) ds}.$$

The previous exact and semi-exact results can be extended to the SABR model with a reflecting boundary; for brevity, we omit those here and refer to Islah (2009) (also Cai et al. 2017).

We conclude this section with an important remark, linked to the moment problem defined in the previous section. More specifically, for the models presented here the moment problem is indeed determinate, i.e., the distribution of $(\Psi(u,t)|\Phi(t))$ (see Table 1) is uniquely determined by the sequence of its moments, consistently with Criterion 1. Based on the results given in Section EC.2 of the e-companion for the different models, the corresponding moment generating functions exist for all $a \in (-c, c)$, where $c > 0$. For example, from (EC.3) for Heston, we have that $a \geq -k^2/2v^2$; from (EC.4) for SABR, $a \geq -(V(t) - V(u))^2/2v^2$; from (EC.5) for OU-SV, $a \geq -k^2/2v^2$; and for 4/2 from (EC.6), $a \geq -k^2/2v^2$ and $b \geq -(2\theta k - v^2)^2/8v^2$. Thus, for all these cases it is implied that all the moments are finite and the moment problem is determinate.

4. General framework for gauging error

The use of an approximation to the true probability distribution is likely to induce some error. In this section, we focus on how any error from the distributional approximation translates to the ultimate quantities of interest, such as expected value functionals. To this end, we, first, derive in Section 4.1 a deterministic procedure for calculating the exact error and, by exploiting this, we further derive in Section 4.2 a deterministic upper bound to it that is easily accessible and computable, assisted by useful results of proximity of two distributions presented in Section 4.3. We focus the spotlight on a general multi-period problem, which can be simplified to a single period upon path-independence, or be amended according to different path-dependent functions.

4.1. Value function backward recursion

In what follows, we consider a time horizon $T > 0$ and partition the whole time interval into N equal points: $0 = t_0 < t_1 < \cdots < t_N = T$. Where necessary, we use relevant subscripts with reference to time on the various quantities that we define on the time grid. Different functions can be defined on this grid. For example,

$$\Pi_N = (K - S_N)^+,$$

where $y^+ := \max(y, 0)$, corresponds to the payoff of a plain vanilla put option with maturity time T and fixed strike price K;

$$\Pi_{j,N} = (K - S_N)^+ \prod_{i=j}^{N} \mathbf{1}_{\{S_i \leq \varpi\}},$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator of the event $\{\cdot\}$, corresponds to a type of barrier option (for example, here, an up-and-out put) with fixed barrier level $\varpi > S_0$;

$$\Pi_{j,N} = \left(K - \bar{M}_{j,N}\right)^+,$$

where $\bar{M}_{j,N} := \min_{j \leq i \leq N} S_i$, corresponds to the payoff of a lookback put option; and,

$$\Pi_{j,N} = \left(K - \frac{1}{N-j+1} \sum_{i=j}^{N} S_i\right)^+ \tag{13}$$

equals the payoff of an Asian put option. Other functions can be accommodated based on the previous ones via standard parity relationships (we can make these available upon request; alternatively, readers may refer, for example, to Rubinstein 1991).

In the next theorem, we derive a backward iterative scheme for expected path-dependent function values. It is general in terms of driving model assumptions, therefore it can be used to compute expected values under a true or approximating distribution law; their difference gives the approximation error. We aim to be as general as possible in our main exposition, concentrating on the barrier case, which encompasses the plain vanilla and other exotic variants, and explaining subsequently how to adapt to those.

THEOREM 1. *Let $s := \ln S$ and $g_{s|\Phi}$ be the conditional density function of s_k given s_{k-1} and $\Phi_k = (\phi_k|\phi_{k-1})$. In addition, $g_\Phi(\phi_k;\phi_{k-1})$ denotes the transition density of the process Φ supported on some interval $\mathcal{I}_\Phi := (\ell_\phi, r_\phi)$, $-\infty \le \ell_\phi < r_\phi \le \infty$. Define the functions*

$$\begin{aligned} p_N(s_N,\phi_N) &= (K-e^{s_N})^+ \mathbf{1}_{\{s_N \le \ln\varpi\}}, \quad \textit{for all } \phi_N \in \mathcal{I}_\Phi, \\ p_{k-1}(s_{k-1},\phi_{k-1}) &= \int_{\mathcal{I}_\Phi}\int_{-\infty}^{\ln\varpi} p_k(s,\phi)\, g_{s|\Phi}(s|\,\phi;s_{k-1},\phi_{k-1}) g_\Phi(\phi;\phi_{k-1}) ds d\phi,\ 0<k\le N. \end{aligned} \tag{14}$$

Then, at time 0*,*

$$E_0[\Pi_{0,N}] = p_0(s_0,\phi_0).$$

Proof. We prove by induction on k that, conditional on the information at time t_k,

$$E_k[\Pi_{k,N}] = p_k(s_k,\phi_k)$$

for $k=0,\dots,N$. The result holds trivially for $k=N$ corresponding to $\Pi_{N,N}$. Suppose that

$$E_{k+1}[\Pi_{k+1,N}] = p_{k+1}(s_{k+1},\phi_{k+1})$$

also holds for arbitrary $k < N-1$. Then, by iterated expectations,

$$\begin{aligned} E_k[\Pi_{k,N}] &= E_k[E_{k+1}[\Pi_{k+1,N}]] = E_k[p_{k+1}(s_{k+1},\phi_{k+1})] \\ &= \int_{\mathcal{I}_\Phi}\int_{-\infty}^{\ln\varpi} p_{k+1}(s,\phi)\, g_{s|\Phi}(s|\,\phi;s_k,\phi_k)\, g_\Phi(\phi;\phi_k) ds d\phi = p_k(s_k,\phi_k) \end{aligned}$$

based on (14). Therefore the statement is true and, by induction,

$$E_0[\Pi_{0,N}] = p_0(s_0,\phi_0).$$

REMARK 1. First, in the case of the lookback option, we focus directly on the distribution of the extremum $\bar{M}_{k,N}$, hence we replace $\Pi_{k,N}$ by $\bar{J}_{k,N} = \prod_{i=k}^N \mathbf{1}_{\{S_i>\varpi\}}$ in Theorem 1 and we have now $p_N(s_N,\cdot) = \mathbf{1}_{\{s_N>\ln\varpi\}}$. The ultimate outcome at time 0 is $E_0\left[\bar{J}_{0,N}\right] = 1 - G_{\bar{M}_{0,N}}(\varpi)$, where $G_{\bar{M}_{0,N}}$ is the distribution function of $\bar{M}_{0,N}$, based on which $E_0[\Pi_{0,N}] = -KG_{\bar{M}_{0,N}}(0) + \int_0^K G_{\bar{M}_{0,N}}(\varpi)\, d\varpi$. Second, the plain vanilla option being path-independent with $p_N(s_N,\cdot) = (K-e^{s_N})\mathbf{1}_{\{s_N\le\ln K\}}$ results in a reduced one-period problem. Third, for the Asian option, it is necessary to employ a state-space reduction based on the new process

$$\bar{S}_k = \frac{1}{N+1}\left(\frac{K(N+1)-\sum_{i=0}^k S_i}{S_k}\right), \quad 0\le k\le N,$$

from which

$$\bar{S}_k = \frac{1}{N+1}\left(\frac{K(N+1)-\sum_{i=0}^{k-1} S_i}{S_{k-1}}\right) e^{s_{k-1}-s_k} - \frac{1}{N+1} = \bar{S}_{k-1}e^{s_{k-1}-s_k} - \frac{1}{N+1} \tag{15}$$

and, for the payoff (13),

$$\Pi_{0,N} = \left(K - \frac{1}{N+1} \sum\nolimits_{i=0}^{N} S_i \right)^+ = S_N \bar{S}_N^+.$$

Following a standard equivalent probability measure change, $E_0 [\Pi_{0,N}]$ can then be computed via backward recursion.

It is worth noting that the previous computations are simplified when $g_{s_k|\cdot}(s_k|\cdot; s_{k-1}, \cdot) = g_{s_k - s_{k-1}|\cdot}(s_k - s_{k-1}|\cdot; \cdot)$. Under a more general process s specification, where $g_{s_k|\cdot}$ does not depend on s_k and s_{k-1} only via their difference (e.g., SABR model), we must keep track of both s and the path-based variable (such as the running average, see expression 15) and (14) in Theorem 1 has to be slightly adjusted to reflect that; we omit the relevant details here, but readers may refer to Sesana et al. (2014).

The approach presented in Theorem 1 can be implemented very accurately via numerical integration/transform techniques (e.g., an adapted two-dimensional version of Lord et al. 2008). It represents a universal computational engine in terms of underlying model assumptions. For example, it can be used in conjunction with the true $g_{s|\Phi}$, but also an approximating density $\tilde{g}_{\tilde{s}|\Phi}$ (see also Section 4.3). The difference of the outcome from both implementations,

$$p_0 - \tilde{p}_0, \tag{16}$$

amounts to the compounded error from the approximation.

4.2. Value function error bound

The dimensionality of the computations involved in Theorem 1 can be considerably reduced by means of an error bound that is faster to evaluate. We develop here our main theoretical result that is of practical relevance, having the advantage of facilitating the error computation and reflecting the error propagation in time related to the backward recursion of the previous section. Following Theorem 1, we build our result based on the barrier option case, which however can be adjusted to other cases according to Remark 1.

THEOREM 2. *Let $G_{s|\Phi}$ and $\tilde{G}_{\tilde{s}|\Phi}$ be the conditional distribution functions of s_k and $\tilde{s}_k$ given s_{k-1} and $\tilde{s}_{k-1}$, respectively, and $\Phi_k = (\phi_k|\phi_{k-1})$. In addition, $E_k[\cdot]$ denotes the conditional expectation given $\Omega_k := (s_k, \tilde{s}_k, \phi_k)$. Consider*

$$\begin{aligned} \mathcal{R}_1(\Omega_{N-1}) &:= |K - \varpi| \, \Upsilon_1(\ln \varpi; \Omega_{N-1}) + \int_{-\infty}^{\ln \varpi} \Upsilon_1(s; \Omega_{N-1}) \, e^s ds, \\ \mathcal{R}_2(\phi_k, \Omega_{k-1}) &:= \int_0^1 \left| G_{s|\Phi}^{-1}(u|\, \phi_k; s_{k-1}, \phi_{k-1}) - \tilde{G}_{\tilde{s}|\Phi}^{-1}(u|\, \phi_k; \tilde{s}_{k-1}, \phi_{k-1}) \right| du, \end{aligned} \tag{17}$$

where

$$\begin{aligned}\Upsilon_1(s;\Omega_{N-1}) &:= \int_{\mathcal{I}_\Phi} \Upsilon(s;\phi,\Omega_{N-1})\, g_\Phi(\phi;\phi_{N-1})d\phi,\\ \Upsilon(s;\phi,\Omega_{N-1}) &:= \left|G_{s|\Phi}(s|\,\phi;s_{N-1},\phi_{N-1}) - \tilde{G}_{\tilde{s}|\Phi}(s|\,\phi;\tilde{s}_{N-1},\phi_{N-1})\right| \end{aligned} \tag{18}$$

for all s and ϕ. Suppose that $\overline{\mathcal{R}}_1 := \sup E_{N-2}\left[\mathcal{R}_1\left(\tilde{\Omega}_{N-1}\right)\right] < \infty$, where $\tilde{\Omega}_k := (\tilde{s}_k, \tilde{s}_k, \phi_k)$, and $\overline{\mathcal{R}}_2 := \sup E_{k-1}\left[\mathcal{R}_2(\phi_k,\Omega_{k-1})\right] < \infty$. Then,

$$|p_{k-1}(s_{k-1},\phi_{k-1}) - \tilde{p}_{k-1}(\tilde{s}_{k-1},\phi_{k-1})| \le \overline{\mathcal{R}}_1 + K(N-k)\,\overline{\mathcal{R}}_2, \quad 0 < k \le N. \tag{19}$$

Remark 2. In the one-period problem, we have that $k = N = 1$, ϖ is replaced by K and the simplifications $\overline{\mathcal{R}}_1 = \mathcal{R}_1(\Omega_0)$ and $\overline{\mathcal{R}}_2 = E_0\left[\mathcal{R}_2(\phi_1,\Omega_0)\right]$ hold naturally from the proof that follows.

Proof. From (14) for $k = N$, we have that

$$\begin{aligned} p_{N-1}(s_{N-1},\phi_{N-1}) &= \int_{-\infty}^{\ln\varpi} (K - e^s)\, g_{1,s}(s;s_{N-1},\phi_{N-1})\, ds \\ &= (K-\varpi)\, G_{1,s}(\ln\varpi;s_{N-1},\phi_{N-1}) + \int_{-\infty}^{\ln\varpi} e^s G_{1,s}(s;s_{N-1},\phi_{N-1})\, ds, \end{aligned}$$

where

$$G_{1,s}(s;s_{N-1},\phi_{N-1}) = \int_{\mathcal{I}_\Phi} G_{s|\Phi}(s|\,\phi;s_{N-1},\phi_{N-1})\, g_\Phi(\phi;\phi_{N-1})d\phi$$

and $g_{1,s}$ denotes the associated density. Similar result holds based, instead, on $\tilde{s}$. Therefore,

$$\begin{aligned} &|p_{N-1}(s_{N-1},\phi_{N-1}) - \tilde{p}_{N-1}(\tilde{s}_{N-1},\phi_{N-1})| \\ &\le |K-\varpi| \left|G_{1,s}(\ln\varpi;s_{N-1},\phi_{N-1}) - \tilde{G}_{1,\tilde{s}}(\ln\varpi;\tilde{s}_{N-1},\phi_{N-1})\right| \\ &\quad + \int_{-\infty}^{\ln\varpi} \left|G_{1,s}(s;s_{N-1},\phi_{N-1}) - \tilde{G}_{1,\tilde{s}}(s;\tilde{s}_{N-1},\phi_{N-1})\right| e^s ds \\ &= |K-\varpi|\,\Upsilon_1(\ln\varpi;s_{N-1},\tilde{s}_{N-1},\phi_{N-1}) + \int_{-\infty}^{\ln\varpi} \Upsilon_1(s;s_{N-1},\tilde{s}_{N-1},\phi_{N-1})\, e^s ds = \mathcal{R}_1(\Omega_{N-1}). \end{aligned}$$

Proceeding to the next iteration, we get based on (14) that

$$\begin{aligned} &|p_{N-2}(s_{N-2},\phi_{N-2}) - \tilde{p}_{N-2}(\tilde{s}_{N-2},\phi_{N-2})| \\ &\le \int_{\mathcal{I}_\Phi}\int_{\mathbb{R}} \left|p_{N-1}(s,\phi)\, g_{s|\Phi}(s|\,\phi;s_{N-2},\phi_{N-2}) - \tilde{p}_{N-1}(s,\phi)\, \tilde{g}_{\tilde{s}|\Phi}(s|\,\phi;s_{N-2},\phi_{N-2})\right| ds g_\Phi(\phi;\phi_{N-2})d\phi \\ &= E_{N-2}\left[|p_{N-1}(s_{N-1},\phi_{N-1}) - \tilde{p}_{N-1}(\tilde{s}_{N-1},\phi_{N-1})|\right]. \end{aligned}$$

This is further bounded by

$$E_{N-2}\left[|p_{N-1}(\tilde{s}_{N-1},\phi_{N-1}) - \tilde{p}_{N-1}(\tilde{s}_{N-1},\phi_{N-1})|\right] + E_{N-2}\left[|p_{N-1}(s_{N-1},\phi_{N-1}) - p_{N-1}(\tilde{s}_{N-1},\phi_{N-1})|\right].$$

With regard to the first term, we have that

$$E_{N-2}\left[\left|p_{N-1}\left(\tilde{s}_{N-1},\phi_{N-1}\right)-\tilde{p}_{N-1}\left(\tilde{s}_{N-1},\phi_{N-1}\right)\right|\right]\leq E_{N-2}\left[\mathcal{R}_1\left(\tilde{\Omega}_{N-1}\right)\right].$$

In relation to the second term, Taylor's theorem implies

$$\left|p_{N-1}\left(s_{N-1},\phi_{N-1}\right)-p_{N-1}\left(\tilde{s}_{N-1},\phi_{N-1}\right)\right|\simeq\left|s_{N-1}-\tilde{s}_{N-1}\right|\left|\partial_1 p_{N-1}\left(s,\phi\right)\right|,$$

where $\partial_1 p$ refers to the first-order derivative of p with respect to its first argument. We also have that

$$\partial_1 p_{k-1}\left(s_{k-1},\phi_{k-1}\right)=\int_{\mathcal{I}_\Phi}\int_{-\infty}^{\ln\varpi}\partial_1 p_k\left(s,\phi\right)g_{s|\Phi}(s|\,\phi;\phi_{k-1})g_\Phi(\phi;\phi_{k-1})dsd\phi,$$

so that $\sup|\partial_1 p_{k-1}|\leq\sup|\partial_1 p_k|$. Now $p_N\left(s\right)=\left(K-e^s\right)^+\mathbf{1}_{\{s\leq\ln\varpi\}}$ implying that $\partial_1 p_N\left(s\right)=-e^s\mathbf{1}_{\{s\leq\ln\varpi,s\leq\ln K\}}$ and $\left|\partial_1 p_N\left(s\right)\right|=e^s\mathbf{1}_{\{s\leq\ln\varpi,s\leq\ln K\}}\leq K$, hence K is a universal bound for the derivative. Therefore,

$$E_{N-2}\left[\left|p_{N-1}\left(s_{N-1},\phi_{N-1}\right)-p_{N-1}\left(\tilde{s}_{N-1},\phi_{N-1}\right)\right|\right]<KE_{N-2}\left[\left|s_{N-1}-\tilde{s}_{N-1}\right|\right],$$

where, by the inverse transform method, $s_{N-1}=G^{-1}_{s|\Phi}\left(U_{N-1}|\,\phi_{N-1};s_{N-2},\phi_{N-2}\right)$ and $\tilde{s}_{N-1}=\tilde{G}^{-1}_{\tilde{s}|\Phi}\left(U_{N-1}|\,\phi_{N-1};\tilde{s}_{N-2},\phi_{N-2}\right)$ for $U_{N-1}\sim\text{Unif}(0,1)$, so that s_{N-1} and $\tilde{s}_{N-1}$ are made as similar as possible. Then, from (17)

$$E_{N-2}\left[\left|s_{N-1}-\tilde{s}_{N-1}\right|\right]=E_{N-2}\left[\mathcal{R}_2\left(\phi_{N-1},\Omega_{N-2}\right)\right].$$

Finally, we get that

$$\left|p_{N-2}(s_{N-2},\phi_{N-2})-\tilde{p}_{N-2}(\tilde{s}_{N-2},\phi_{N-2})\right|\leq E_{N-2}\left[\mathcal{R}_1\left(\tilde{\Omega}_{N-1}\right)\right]+KE_{N-2}\left[\mathcal{R}_2\left(\phi_{N-1},\Omega_{N-2}\right)\right].$$

Therefore, the statement (19) is true for $k=N-1$. Now assume that it holds for arbitrary $k<N-1$. Then,

$$\begin{aligned}
&\left|p_{k-1}(s_{k-1},\phi_{k-1})-\tilde{p}_{k-1}(\tilde{s}_{k-1},\phi_{k-1})\right|\\
&\leq E_{k-1}\left[\left|p_k\left(\tilde{s}_k,\phi_k\right)-\tilde{p}_k\left(\tilde{s}_k,\phi_k\right)\right|\right]+E_{k-1}\left[\left|p_k\left(s_k,\phi_k\right)-p_k\left(\tilde{s}_k,\phi_k\right)\right|\right]\\
&\leq\overline{\mathcal{R}}_1+K\left(N-k-1\right)\overline{\mathcal{R}}_2+KE_{k-1}\left[\mathcal{R}_2\left(\phi_k,\Omega_{k-1}\right)\right]\leq\overline{\mathcal{R}}_1+K\left(N-k\right)\overline{\mathcal{R}}_2,
\end{aligned}$$

which concludes the proof.

4.3. Proximity of distributions with shared moments

The distance (18) is key for the practical implementation of Theorem 2. A possible way of computing this is (see Feller 1971)

$$G_{s|\Phi}\left(s|\,\phi;\cdot\right)-\tilde{G}_{\tilde{s}|\Phi}\left(s|\,\phi;\cdot\right)=\frac{2}{\pi}\int_0^\infty\frac{\sin\left(us\right)}{u}\left[\Re(\varphi_{G_{s|\Phi}}\left(u|\,\phi;\cdot\right))-\Re(\varphi_{\tilde{G}_{\tilde{s}|\Phi}}\left(u|\,\phi;\cdot\right))\right],\tag{20}$$

where $\Re(\cdot)$ denotes the real part function. (An adjusted version can be used for the quantile functions in integral 17.) $\varphi_{G_{s|\Phi}}\left(u|\,\phi;\cdot\right)=E[e^{ius}|\phi;\cdot]$ and $\varphi_{\tilde{G}_{\tilde{s}|\Phi}}\left(u|\,\phi;\cdot\right)=E[e^{iu\tilde{s}}|\phi;\cdot]$ are the relevant characteristic functions, see equation (23).

Computing (20) can be facilitated by suitable conditioning arguments for the various models presented in Section 3 (see also Table 1). Consider, for example, the Heston model. Conditional on (Ψ,Φ), its distribution law is given by

$$s_k=\alpha\left(s_{k-1},\phi_{k-1}\right)+\beta\Phi_k+\gamma\Psi_{k-1,k}+\sqrt{\delta\Psi_{k-1,k}}Y_k,\quad Y_k\sim\mathcal{N}\left(0,1\right),\tag{21}$$

where Y is independent of Φ and Ψ, $\alpha\left(s_{k-1},\phi_{k-1}\right)=s_{k-1}+(r-\beta k\theta)(t_k-t_{k-1})-\beta\phi_{k-1}$, $\gamma=\beta k-1/2$, $\beta=\rho/v$, $\delta=1-\rho^2$. Based on this,

$$G_{s|\Phi}(\,s|\,\phi;s_{k-1},\phi_{k-1})=\int_0^\infty G_{s|\Psi,\Phi}(\,s|\,x,\phi;s_{k-1},\phi_{k-1})g_{\Psi|\Phi}(x|\,\phi;\phi_{k-1})dx,\tag{22}$$

where

$$G_{s|\Psi,\Phi}(\,\cdot\,|\,x,\phi;s_{k-1},\phi_{k-1})=\frac{1}{\sqrt{2\pi\delta x}}\int_{-\infty}^{\cdot}e^{-\frac{(s-\alpha\left(s_{k-1},\phi_{k-1}\right)-\beta\phi-\gamma x)^2}{2\delta x}}ds.$$

The associated characteristic function is given by

$$\begin{aligned}\varphi_{G_{s|\Phi}}(u|\,\phi;s_{k-1},\phi_{k-1})&=\int_0^\infty e^{iu\left(\alpha\left(s_{k-1},\phi_{k-1}\right)+\beta\phi+\gamma x\right)-\frac{\delta xu^2}{2}}g_{\Psi|\Phi}(x|\,\phi;\phi_{k-1})dx&(23)\\&=e^{iu\left(\alpha\left(s_{k-1},\phi_{k-1}\right)+\beta\phi\right)}\varphi_{G_{\Psi|\Phi}}\left(u\left(\gamma+\frac{i\delta u}{2}\right)\middle|\,\phi;\phi_{k-1}\right),&(24)\end{aligned}$$

where $\varphi_{G_{\Psi|\Phi}}$ is given by (EC.3) in the e-companion Section EC.2. Introducing the approximating law for $(\tilde{\Psi}|\Phi)$, we define by analogy

$$\tilde{s}_k=\alpha\left(\tilde{s}_{k-1},\phi_{k-1}\right)+\beta\Phi_k+\gamma\tilde{\Psi}_{k-1,k}+\sqrt{\delta\tilde{\Psi}_{k-1,k}}Y_k,$$

where Y is also independent of $\tilde{\Psi}$, and replace by $\tilde{g}_{\tilde{\Psi}|\Phi}$ in (22)–(23) and $\varphi_{\tilde{G}_{\tilde{\Psi}|\Phi}}$ in (24) which we derive for the Pearson distribution in Proposition EC.1 in Section EC.1 of the e-companion. Finally, relevant representations hold for other models, see e-companion Section EC.3.

If the conditional moments of $\tilde{\Psi}$ given Φ are equal to those of Ψ given Φ, then the same holds between $\tilde{s}$ given Φ and s given Φ. This becomes obvious from the linking relationship of their moments. A straightforward application of the binomial theorem to (21) yields for general moments

$$E_{k-1}\left[s_k^n|\Phi_k\right]=\sum_{j=0}^n\binom{n}{j}\delta^{(n-j)/2}E\left[Y_k^{n-j}\right]\sum_{k=0}^j\binom{j}{k}\left(\alpha\left(s_{k-1},\phi_{k-1}\right)+\beta\Phi_k\right)^k\gamma^{j-k}E\left[\Psi_{k-1,k}^{(n+j)/2-k}\middle|\,\Phi_k\right],$$

where $E\left[Y_k^n\right]$ is equal to the double factorial $(n-1)!!$ if n is even, otherwise it is equal to 0. We note that the required integer moments of $(\Psi|\Phi)$ from 1 to n can be computed, for example, the way we explain in Section 5.2, whereas the non-integer moments are, in practice, not needed because they are multiplied by the odd moments of Y that are equal to zero. The previous moments can also be obtained from $\varphi_{G_{s|\Phi}}$ and $\varphi_{\tilde{G}_{\tilde{\Psi}|\Phi}}$.

An alternative to (20) way of computing (18) is through a bound. In the space of distribution functions of random variables, the topology given by the uniform metric (Kolmogorov 1933, Zolotarev 1983) between any pair of elements G and $\tilde{G}$ is defined, adapted to our random variable notation, as

$$\rho\left(G_{s|\Phi}, \tilde{G}_{\tilde{s}|\Phi}\right) = \sup_x \left|G_{s|\Phi}(x) - \tilde{G}_{\tilde{s}|\Phi}(x)\right|, \tag{25}$$

which represents a natural bound to (18). In addition, the Lévy distance (Lévy 1925) is defined as

$$L\left(G_{s|\Phi}, \tilde{G}_{\tilde{s}|\Phi}\right) = \inf\left\{\epsilon:\ G_{s|\Phi}(x-\epsilon) - \epsilon \leq \tilde{G}_{\tilde{s}|\Phi}(x) \leq G_{s|\Phi}(x+\epsilon) + \epsilon \text{ for all } x\right\}.$$

The two metrics are linked as follows.

LEMMA 1 **(Linnik and Ostrovskiĭ 1977)**. *For all distribution functions G and $\tilde{G}$,*

$$L\left(G, \tilde{G}\right) \leq \rho\left(G, \tilde{G}\right) \leq (1+\varrho)\, L\left(G, \tilde{G}\right), \tag{26}$$

where $\varrho = \sup_x \tilde{G}'(x)$ if $\tilde{G}$ is absolutely continuous.

Klebanov and Mkrtchyan (1986) have studied closeness in the L-metric in terms of the truncated Carleman's series whose divergence suffices for the moment problem to be determinate (Akhiezer 1965). Rachev et al. (2013, Theorem 10.3.4) allow us to restate one of the original Klebanov and Mkrtchyan's (1986) results when the higher moments of the two distributions are not coinciding but are only fairly close. Therefore, for G and $\tilde{G}$ with finite moments up to order $2m$ and

$$\left|\mu_n(G_{s|\Phi}) - \mu_n(\tilde{G}_{\tilde{s}|\Phi})\right| \leq \delta_{G,\tilde{G}}, \quad n = 1, \ldots, 2m,$$

where $0 < \delta_{G,\tilde{G}} \leq 1$, we have that

$$L\left(G_{s|\Phi}, \tilde{G}_{\tilde{s}|\Phi}\right) \leq \frac{2}{\ln\left(1+\delta_{G,\tilde{G}}^{-1/2}\right)} + \left(\frac{2\mu_{2m}}{(2m)!}\right)^{\frac{1}{2m+1}}. \tag{27}$$

This is an important result as it is not only based on moments, which is structural element of our approach, but it also implies that the error remains bounded even when the moments do not match perfectly. Whilst we are exactly fitting a certain number of moments in a well-posed moment problem setting, in practice, as we will show in Section 6.1, the higher moments of the two distributions are very close, which by moment determinacy ensures an accurate approximation. Some alternatives related to the uniform metric are presented in Sections EC.4.1–EC.4.2 of the e-companion.

5. Moment-based random number generator

We now turn our attention to the simulation problem which represents the main thrust of this research. Focusing on the models of Section 3, key quantities in them (but also of those later in Section 7) are conditional integrated processes defining moment-determinate problems, which we have generally denoted by $\Psi(u,t)$, and the relevant conditioning arguments $\Phi(t)$; see Table 1.

The relevant simulation procedure is summarized as follows:

1. Simulate $(\Phi(t)|V(u))$
2. Simulate $(\Psi(u,t)|V(u),\Phi(t))$
3. Simulate $(s(t)|s(u),V(u),\Phi(t),\Psi(u,t))$.

Steps 1 and 3 are trivial as we have seen that the conditional distributions are known; what remains a challenge is the generation of conditional samples from $\Psi(u,t)$ in step 2. Hitherto in the literature, exact simulation of the conditional $\Psi(u,t)$ has relied on numerical inversion of its Laplace transform $\mathcal{L}(a) := E\left[\exp\left(-a\Psi(u,t)\right)|\Phi(t)\right]$ (refer to e-companion Section EC.2 for the various models), which is nevertheless the hardest and most time-consuming step of the whole simulation scheme. For this reason, typical exact simulation schemes are very slow when used to simulate entire trajectories due to multiple numerical inversion of the previous Laplace transform at each time u. For ease of notation when referring to the conditional distribution of $\Psi(u,t)$, we focus hereafter on the condition on $\Phi(t)$, which is key in our problem, with the rest of the information up to time u implied.

5.1. Random number generation based on fitted Pearson curves

We aim to circumvent the previous stumbling block by proposing a new approach to simulating $(\Psi(u,t)|\Phi(t))$ relying on a Pearson curve fit to the corresponding theoretical distribution introduced in Section 2.

Once the Pearson curve type is determined, we can draw random numbers from it. First, we generate random numbers $\bar{\mathcal{Y}}$ from the selected standardized family with zero mean and variance equal to unity, based on the procedure summarized in Section EC.5 of the e-companion (see Johnson et al. 1994 for more details). Then, we rescale and shift using the true mean μ_1 and standard deviation $(\mu_2-\mu_1^2)^{1/2}$ to obtain

$$\tilde{\mathcal{Y}} := \mu_1 + \sqrt{\mu_2-\mu_1^2}\bar{\mathcal{Y}} \tag{28}$$

corresponding to a random sample from $(\Psi(u,t)|\Phi(t))$.

5.2. Computation of moments

Associated with our moment-based technique is our additional contribution of efficient computation of the integer moments. The n-th conditional moment of $\Psi(u,t)$ is traditionally given by

$$\mu_n = (-1)^n \left. \frac{\partial^n}{\partial a^n} \mathcal{L}(a) \right|_{a=0}. \tag{29}$$

However, computing the moments from (29) may not be practicable, especially for high orders, as the Laplace transform may involve special functions. For example, consider the relevant Laplace transform (EC.3) for the Heston model. Evaluating the first four moments of $\Psi(u,t)$ from (29), using, e.g., the symbolic toolbox of Mathematica, requires 307 Bessel function evaluations, which is highly computationally intensive but also endangers numerical errors.

In this paper, we bypass such kind of problem by numerical inversion of an adaptively modified moment generating function introduced by Choudhury and Lucantoni (1996). According to this,

$$\begin{aligned} \mu_n &= \frac{n!}{2nlr_n^n \alpha_n^n} \left\{ \mathcal{L}(\alpha_n r_n) + (-1)^n \mathcal{L}(-\alpha_n r_n) + 2 \sum_{j=1}^{nl-1} \Re(\mathcal{L}(\alpha_n r_n e^{\pi i j/nl}) e^{-\pi i j/l}) \right\} - \hat{e}_n, \qquad (30) \\ \hat{e}_n &:= \sum_{j=1}^{\infty} \alpha_n^{2ljn} \frac{n!}{(n+2ljn)!} \mu_{n+2ljn} 10^{-\gamma j}, \end{aligned}$$

where the latter is the error term. The choice $r_n := 10^{-\gamma/(2nl)}$ is made in order to bound the error and achieve accuracy of the order $10^{-\gamma}$. Algorithm 1 summarizes the procedure for computing any m integer moments as well as the parameter l and the adaptive α_n.

Algorithm 1 Numerical inversion of adaptively modified moment generating function

Input: m, γ, $\mathcal{L}(\cdot)$

Output: $\{\mu_n\}_{n=1}^m$

1: Set $l = \alpha_1 = 1$ and compute μ_1 from (30)

2: Compute $\alpha_2 = 1/\mu_1$ and μ_2

3: Set $l = 1 \vee 2$ and $\alpha_1 = \alpha_2 = 2\mu_1/\mu_2$ and compute new values for μ_1 and μ_2 (from 30)

4: Set $n = 3$

5: **while** $n \leq m$ **do**

6: $\quad l = 1 \vee 2$, compute $\alpha_n = (n-1)\mu_{n-2}/\mu_{n-1}$

7: $\quad$ Compute μ_n from (30)

8: $\quad n = n+1$

9: **end**

The proposed method has several merits. First, it is fast: for the first four moments, the Laplace transform is evaluated 14 times $= 2$ (for μ_1) $+ 3$ (for μ_2) $+ 4$ (for μ_3) $+ 5$ (for μ_4) (for $l = 1$ used in equation 30). Hence, for the Heston model, the Bessel function in (EC.3) is evaluated just 28 times, which is a drastic reduction from 307 times if we used (29) instead. Second, it is very accurate: using $\gamma = 11$ as per the recommendation of Choudhury and Lucantoni (1996), the error appears, consistently with them, only in the eleventh to thirteenth significant place. This is generally easy to verify by calculating just the first few terms in $\hat{e}$ based on the true moments, as the infinite series is heavily damped and the error cannot be significant. Finally, the Laplace transform can be evaluated for the whole sample of $\Phi(t)$ draws altogether (as opposed to one by one), which we favour in our application requiring multiple moment computations for different random realizations of $\Phi(t)$ due to dependence of the moments of $\Psi(u,t)$ on $\Phi(t)$.

5.3. Summary of the simulation method and extensions

Our method still hinges on the Laplace transform, but circumvents its numerical inversion in virtue of the speedy computation of the moments and simulation of the Pearson proxy. In Section 6, it is shown that the computing time is drastically reduced, benefiting the simulation of the asset price process on a set of multiple observation dates. We summarize the methodology in Algorithm 2 based on sampling from (28). If we care to simulate the terminal asset price only, we use $N = 1$.

Algorithm 2 Moment-matched conditional sampling scheme: asset price process under stochastic volatility

Input: Model parameters, terminal time T, number of monitoring dates N

Output: Asset price path $\{S(t)\}$ for $t = \{0, \Delta, 2\Delta, 3\Delta, ..., T\}$

1: Set $\Delta = \frac{T}{N}$
2: **for** $t = 0 : \Delta : T - \Delta$ **do**
3: Given $V(t)$, generate $\Phi(t+\Delta)$
4: Compute the moments of $\Psi(t, t+\Delta)$ conditional on $V(t)$ and $\Phi(t+\Delta)$ using Algorithm 1
5: Sample from the conditional $\Psi(t, t+\Delta)$ based on (28) given the moments
6: Sample $S(t+\Delta)$ given $S(t), V(t), \Phi(t+\Delta)$ and $\Psi(t, t+\Delta)$
7: **end for**
8: **return** $\{S(t)\}$ for $t = \{0, \Delta, 2\Delta, 3\Delta, ..., T\}$

The 4/2 model (see Table 1) requires some extra care as it requires sampling from $(\Psi(u,t)|\Phi(t)) = \left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)} \middle| V(t) \right)$. In Section EC.6 of the e-companion, we present an extension of our approximating framework to the multivariate case, in particular the bivariate case of relevance here. Alternatively, the sampling problem can be simplified by entirely bypassing the time integrals

using directly the conditional Laplace transform $E\left[\exp\left(-as(t)\right)|\,V(t)\right]$ (see e-companion Section EC.2).

A few more cases of models with jumps ensue. Extending to the Bates and DPS models is straightforward, following Broadie and Kaya (2006, Sections 6.1–6.2), by replacing the time integral with our Pearson fit. Similar logic applies to the simulation of the 4/2 model extension with independent jumps of Lin et al. (2017). In addition, we have a couple of cases that deserve some more attention.

5.3.1. Time-changed Lévy models A Lévy model with integrated CIR time-change is given by

$$s(t) = s(u) + r(t-u) - \ln E(e^{L(1)}) \int_u^t V(s)ds + L\left(\int_u^t V(s)ds\right),$$

where V is a square-root diffusion process and L a Lévy process. Having first simulated $\int_u^t V(s)ds$ according to the fitted Pearson distribution, we then simulate the Lévy process L on the new (stochastic) time scale. The simulation of standard Lévy processes including the variance gamma or normal inverse Gaussian is trivial (e.g., see Cont and Tankov 2004); the case of the CGMY process is much harder (e.g., see Ballotta and Kyriakou 2014) and one can use the new technique we develop in Section EC.8.1 of the e-companion.

5.3.2. Jump-diffusion model with state-dependent jump intensity Next, we draw attention to the class of self-exciting point processes, in particular, a jump-diffusion process with state-dependent drift, diffusion coefficient and jump intensity:

$$d\lambda(t) = k_\lambda(\theta_\lambda - \lambda(t))dt + \sigma_\lambda\sqrt{\lambda(t)}dW(t) + dJ(t), \tag{31}$$

where $k_\lambda, \theta_\lambda, \sigma_\lambda$ are constant, $J(t) := \beta N(t)$ is a jump process given by a compound self-exciting point process with CIR intensity, $N(t) := \sum_{i\geq 1} \mathbf{1}_{\left\{\bar{T}_i < t\right\}}$ is the total number of jumps and $\bar{T}_i$ the i-th jump time (see Dassios and Zhao 2017 for more details). Model (31) accounts for jumps, for example, in the asset price process, that are not uniformly distributed over time, but tend to appear in clusters (e.g., see Fulop and Li 2019 and Du and Luo 2019). For the sake of exemplification, we will assume that

$$s(t) = s(u) + \left(r - \frac{\sigma^2}{2}\right)(t-u) - \omega\int_u^t \lambda(s)ds + \sigma W(t) + \sum_{i=1}^{N(t)} J_i \tag{32}$$

for jump sizes $J \sim \mathcal{N}\left(\mu_J, \sigma_J^2\right)$ and $\omega := \exp\left(\mu_J + \sigma_J^2/2\right) - 1$. This model can be seen as a self-exciting extension of the one proposed in Wachter (2013) with $\beta = 0$. It is also a special case with constant variance σ^2 of the affine model in Fulop and Li (2019, Model III). (Our method is adaptable to

the full model specification; upon assuming stochastic variance with jumps, the simulation task becomes similar to that of the DPS model.)

Giesecke et al. (2011) and Dassios and Zhao (2017) show us how to simulate exactly the pairs $\left\{\left(\bar{T}_i, \lambda(\bar{T}_i)\right)\right\}_{i=1}^{N(t)}$ for $u < \bar{T}_i < t \ \forall i > 0$; nevertheless, on several occasions, such as (32), we need to be able to simulate also

$$\left(\int_{\bar{T}_i}^{\bar{T}_{i+1}} \lambda(s) ds \middle| \lambda(\bar{T}_i), \lambda(\bar{T}_{i+1}) \right). \tag{33}$$

Between two consecutive jump times $\bar{T}_i$ and $\bar{T}_{i+1}$, the process λ evolves as a CIR diffusion and our method lends itself to the efficient simulation of (33). Then, we have that

$$\int_u^t \lambda(s) ds = \sum_{i=0}^{N(t)} \int_{\bar{T}_i}^{\bar{T}_{i+1}} \lambda(s) ds + \int_{\bar{T}_{N(t)}}^t \lambda(s) ds$$

and

$$\left(s(t) \middle| s(u), \int_u^t \lambda(s) ds, \sum_{i=1}^{N(t)} J_i \right) \sim \mathcal{N}(m(u,t), s^2(u,t))$$

with

$$m(u,t) := s(u) + \left(r - \frac{1}{2}\sigma^2 \right)(t-u) - \omega \int_u^t \lambda(s) ds + \sum_{i=1}^{N(t)} J_i \text{ and } s^2(u,t) := \sigma^2 (t-u).$$

6. Numerical study

The following section is dedicated to a numerical analysis of the proposed approximation and simulation method applied to the Heston (one and two-factor), SABR, OU-SV, 3/2 and 4/2, Bates, DPS, NIGCIR (NIG with integrated CIR time-change) and SECIRJD (self-exciting point process with CIR intensity) models. We use parameter sets from the relevant literature that are practically relevant and representative of different markets and market conditions and are not benign (including high correlations, high volatility of variance, and long maturities). All parameter values and their sources are reported in Table 2. All numerical experiments are run in Matlab R2019b in Microsoft Windows 10 on a machine with an Intel(R) Core(TM) i7-9750HQ CPU @2.60GHz and 16 GB of RAM. A code is made available from `https://openaccess.city.ac.uk/id/eprint/26427/` and Section EC.9 of the e-companion.

6.1. Analysis of error and computing time

Before moving to the actual application, we study the core of our method, that is, the ability of Algorithm 1 to compute fast and accurately the moments of the conditional time integral as well as the quality of the subsequent Pearson distribution fit. In the interest of space, we present results relating only to the Heston model (parameter sets H1 and H5, being common in the literature),

Table 2 Model parameter sets

Heston model												
	$S(0)$	k	θ	v	$V(0)$	ρ	r	T				
H1	100	6.21	0.019	0.61	0.010201	-0.7	3.19%	1				
H2	100	2	0.09	1	0.09	-0.3	5%	5				
H3	100	0.5	0.04	1	0.04	-0.9	3%	1				
H4	100	0.3	0.04	0.9	0.04	-0.5	3%	1				
H5	100	1	0.09	1	0.09	-0.3	3%	1				
H6	100	6.2	0.02	0.6	0.02	-0.7	3%	1				
SABR model												
	$S(0)$	b	v	$V(0)$	ρ	r	T					
SABR1	0.05	0.3	0.6	0.4	0	0	1					
SABR2	0.05	0.3	0.6	0.4	0	0	3					
SABR3	0.05	0.3	0.6	0.4	0	0	5					
SABR4	0.5	0.5	0.4	0.5	0	0	4					
SABR5	0.04	1	0.3	0.2	-0.5	0	5					
SABR6	1	0.6	0.3	0.25	-0.5	0	20					
OU-SV model												
	$S(0)$	k	θ	v	$V^2(0)$	ρ	r	T				
OU-SV1	100	4	0.02	0.1	0.04	-0.7	9.53%	1				
OU-SV2	100	4	0.02	0.1	0.04	-0.7	9.53%	5				
OU-SV3	100	4	0.02	0.1	0.04	-0.7	9.53%	10				
Double Heston model												
	$S(0)$	$k_{1,2}$	$\theta_{1,2}$	$v_{1,2}$	$V_{1,2}(0)$	$\rho_{1,2}$	r	T				
DH1	100	0.9 1.2	0.1 0.15	0.36 0.2	0.36 0.2	-0.5 -0.5	3.00%	1				
DH2	100	1.0738 0.0326	0.1026 0.7078	0.826 1.5355	0.0028 0.0059	-0.2819 -0.687	3.00%	1				
3/2 model												
	$S(0)$	k	θ	v	$V(0)$	ρ	b	a	r	T		
3/2-1	100	1.8	0.04	0.2	0.04	-0.7	0.025	0	2%	1		
3/2-2	100	1.1705	0.6853	0.398	0.8992	-0.8637	0.0192	0	2%	1		
4/2 model												
	$S(0)$	k	θ	v	$V(0)$	ρ	b	a	r	T		
4/2-1	100	1.8	0.04	0.2	0.04	-0.7	0.025	0.3	2%	1		
4/2-2	100	1.1705	0.6853	0.398	0.8992	-0.8637	0.0192	0.0218	2%	1		
Bates model												
	$S(0)$	k	θ	v	$V(0)$	ρ	λ	ω	σ_J	r	T	
B1	100	3.99	0.014	0.2700	0.008836	-0.79	0.11	-0.12	0.15	3.19%	5	
DPS model												
	$S(0)$	k	θ	v	$V(0)$	ρ	λ	ω	σ_J	μ_v	ρ_J	r	T
DPS1	100	3.46	0.008	0.1400	0.007569	-0.82	0.47	-0.1	0.0001	0.05	-0.38	3.19%	1
NIGCIR model												
	$S(0)$	k	θ	v	$V(0)$	$\underline{\theta}$	$\underline{k}$	$\underline{\sigma}$	r	T		
NIGCIR1	100	3.99	0.014	0.27	0.008836	-11.00604	0.00294	0.84059	4%	1		
SECIRJD model												
	$S(0)$	$\lambda(0)$	k_λ	θ_λ	σ_λ	β	σ^2	μ_J	σ_J	r	T	
SECIRJD1	100	2.935	2.266	2.935	0.585	1.782	0.041	-0.018	0.002	3.19%	1	

Notes. Parameter sets H1–2, B1 (with jump model parameters $\lambda, \mu_J = \ln(1+\omega) - \sigma_J^2/2, \sigma_J$), DPS1 (with jump model parameters $\lambda, \mu_J = \ln[(1+\omega)(1-\rho_J\mu_v)] - \sigma_J^2/2, \sigma_J, \mu_v, \rho_J$) are from Broadie and Kaya (2006) and H3–6 from Glasserman and Kim (2011), DH1–2 from Gauthier and Possamaï (2010) and Zhang and Feng (2019), SABR1–3 and SABR4–6 from Cai et al. (2017) and Leitao et al. (2017), OU-SV1–3 from Li and Wu (2019), 3/2-1 and 4/2-1 from Grasselli (2017) and Callegaro et al. (2019), 3/2-2 and 4/2-2 from Gnoatto et al. (2016), NIGCIR1 (with NIG parameters $\underline{\theta}, \underline{k}, \underline{\sigma}$) from Corsaro et al. (2019, Table 2) and SECIRJD1 from Fulop and Li (2019, Table 8).

which is the most computationally complicated with associated Laplace transform (EC.3) that involves Bessel functions whose calculation generally slows down the simulation. We break down the procedure in Algorithm 2 for simulating over the one-period $[0, T]$ into the following four pieces:

Step 1 Generate a sample of $(V(T)|\,V(0))$

Step 2 Compute the first four integer moments of $\left(\int_0^T V(s)ds \middle| V(T)\right)$ using Algorithm 1

Step 3 Generate a sample of $\left(\int_0^T V(s)ds \middle| V(T)\right)$ from the Pearson curve fitted by moments

Step 4 Generate a sample of $\left(S(T) \middle| \int_0^T V(s)ds, V(T)\right)$.

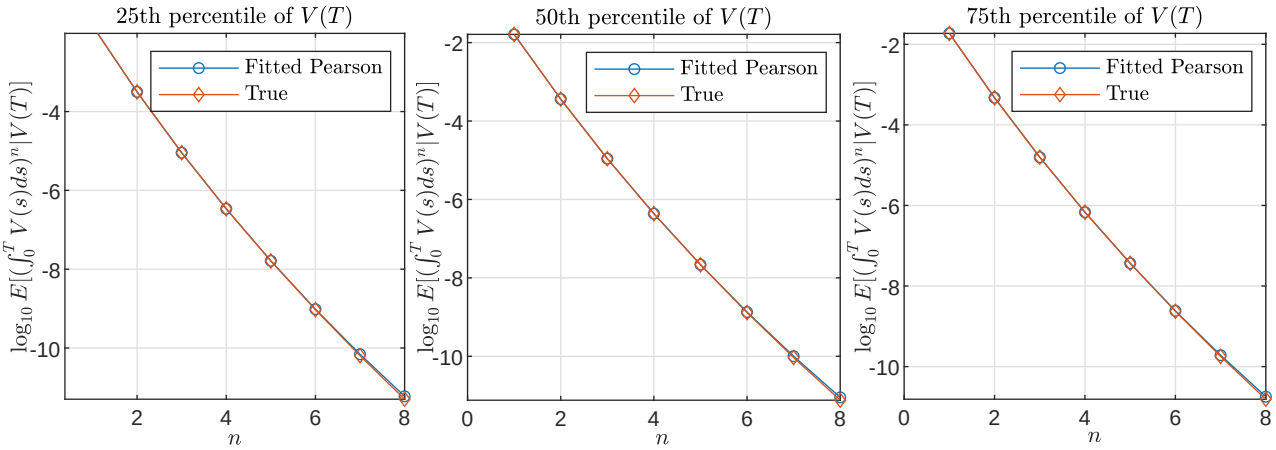
Having repeated the previous steps several times, we report the mean execution times corresponding to each step in Section EC.7.1 of the e-companion. Increasing the number of simulations results in linear increases of the computing time of each step, but with constant percentages of each with respect to the total time. Steps 1, 3 and 4 account for 4%, 11% and 8% of the total time and can be easily implemented using standard routines in numerical computing environments such as Matlab (`ncx2rnd` for step 1; `pearsrnd` for step 3; `randn` for step 4). The moments' evaluation, that is, step 2, is the dominant element occupying 77% of the total execution time. In addition to this, we compare evaluation done using Algorithm 1 and via analytical moments from equation (29). Results reported in Section EC.7.1 of the e-companion for the first four integer moments indicate a considerable speed-up using Algorithm 1, which in the case of the Heston model translates to 28 Bessel function evaluations versus 307 via equation (29).

Next, we study the potential sources of error from our moment-based approximation. First, we compare moments evaluated using Algorithm 1 and analytically from equation (29) for three different random realizations of the terminal variance $\Phi(T) = V(T)$ corresponding to the 25, 50 and 75th percentiles of that distribution, for convenience. (Similar performances were observed for different terminal variances and parameter sets.) The absolute discrepancies are extremely small, as expected based on the discussion in Section 5.2 (see also e-companion Section EC.7.1).

Second, we investigate the ability of the Pearson fit $\tilde{G}_{\tilde{\Psi}|\Phi}$ to represent the true distribution $G_{\Psi|\Phi}$ of the conditional integrated variance $(\Psi(0,T)|\Phi(T)) = \left(\int_0^T V(s)ds \middle| V(T)\right)$. To this end, we perform two tests. In the first one, we aim to assess the closeness of the Pearson moments from the true ones, i.e., having matched the first four integer moments, how far are the higher-order moments from the true ones? Adhering to the same 25, 50 and 75th percentiles of the terminal variance distribution as before for convenience, we present on a log-scale in Figure 1 the first eight integer moments. Indeed, the moments are very close and, for example, for parameters H1 and conditional on the 50th percentile of the terminal variance, we experience absolute differences $|\mu_n(G_{\Psi|\Phi}) - \mu_n(\tilde{G}_{\tilde{\Psi}|\Phi})|$ of the fifth ($n=5$) to eighth ($n=8$) moments of 0.0003, 0.0014, 0.0034 and 0.0064. Other parameter sets yield similar results insinuating that we are implicitly able to almost fit more than four moments, hence corroborating the high accuracy of our method.

Having replaced the original steps 2 and 3 of Broadie and Kaya (2006) with our approach, it also comes natural to evaluate the potential impact on accuracy. For this, we compute the true and fitted Pearson cumulative distribution functions, $G_{\Psi|\Phi}$ and $\tilde{G}_{\tilde{\Psi}|\Phi}$, and present both in Figure

Figure 1 True moments of the integrated variance in the Heston model (parameter set H1) conditional on the 25, 50, 75th percentiles of the terminal variance and corresponding moments based on fitted Pearson distribution

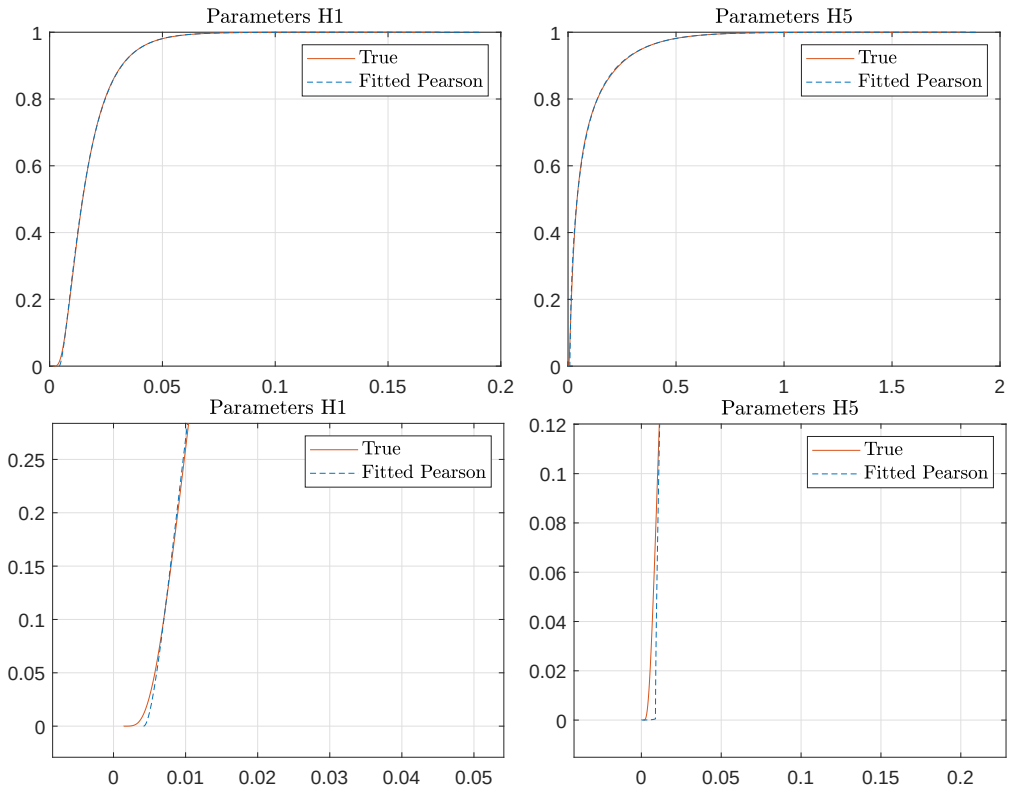


2 for terminal variance $\Phi(T) = V(T)$ corresponding to the 50th percentile only in the interest of space. The two plots are practically almost identical. Putting under the microscope the left side of the distribution (see bottom plots in Figure 2) reveals maximum absolute differences between the two of 0.001 and 0.007 for parameters H1 and H5 respectively, whereas for larger values of the integrated variance we observe even smaller differences. An implementation of a two-sample Kolmogorov–Smirnov test also leads to non-rejection of the null hypothesis that the samples are drawn from the same distribution with very significantly high p-values for the different parameter sets (smallest p-value being 0.6 for H1). We repeat this exercise in the multi-period problem applied to path-dependent quantities such as the average and the maximum of a trajectory; the relevant results are very close to the path-independent ones and are deferred to Section EC.7.1 of the e-companion.

In problems where the key question is the choice of an approximating distribution, the use of the maximum entropy principle is also popular. For this, along with our original approximating Pearson distribution function $\tilde{G}_{\tilde{\Psi}|\Phi}$ with associated density function $\tilde{g}_{\tilde{\Psi}|\Phi}$, we consider the entropy-maximizing distribution $\hat{G}_{\hat{\Psi}|\Phi}$ with density function $\hat{g}_{\hat{\Psi}|\Phi}(x) := \exp\left(-\sum_{n=0}^{m} \lambda_n x^n\right)$, where $\{\lambda_n\}$ are the Lagrange multipliers (Kapur and Kesavan 1992). For a generic continuous distribution function H and associated density function h, we define the differential entropy $\mathcal{H}[h] = -\int h(x) \ln h(x)\, dx$. The density $\hat{g}_{\hat{\Psi}|\Phi}$ is then obtained by maximizing the entropy constrained by moments. The resulting distribution $\hat{G}$ shares the same first m moments with the target true distribution G.

In Figure 3, we compare the entropy-maximizing distribution $\hat{G}$ based on m moments and our (formally four-moment) fitted Pearson distribution $\tilde{G}$ against the true distribution G. We report in the top plots $\left|H - G_{\Psi|\Phi}\right|$, in the form of a bound whose details can be found in Section EC.4.3 of the

Figure 2 True and fitted Pearson cumulative distribution functions, G and $\tilde{G}$, of the integrated variance in the Heston model (parameter sets H1 and H5) conditional on the 50th percentile of the terminal variance (top plots: full scale; bottom plots: emphasized areas of largest error)

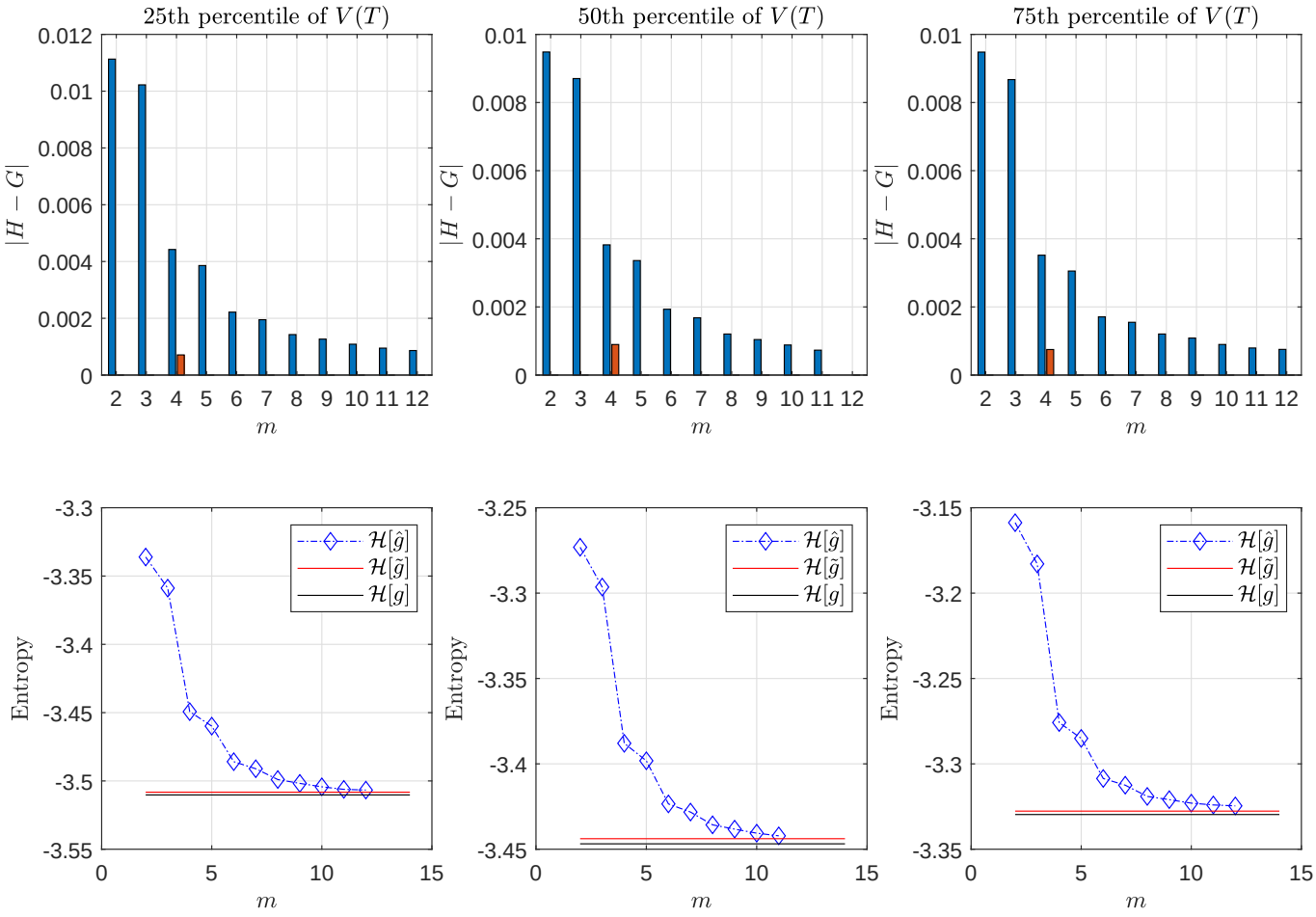


e-companion, where $H \in \{\hat{G}_{\hat{\Psi}|\Phi}, \tilde{G}_{\tilde{\Psi}|\Phi}\}$ and $G_{\Psi|\Phi}$ is the true distribution; the individual entropies can be found in the bottom plots. It is obvious that the distances are small and are reducing with increasing m. The fitted Pearson is very close in terms of entropy to the true distribution, while the entropy-maximizing distribution converges to the true with increasing m, remaining, though, even with twelve moments matched, behind the Pearson fit.

Unreported results, in the interest of space, in relation to Figures 1–3 based, instead, on the cumulative distribution functions, G and $\tilde{G}$, of $(s(T)|V(T))$ (see equation 22) remain qualitatively unchanged.

We conclude this part with a short note of some additional attempts of fitting of Pearson curves based, instead, on exponential, negative (see Cressie et al. 1981) or even fractional (see Cressie and Borkent 1986) moments, i.e., $E(e^{-m\alpha\Psi(0,T)}|\Phi(T))$ for $m = 1, \dots 4$ and given α, $E(\Psi(0,T)^{-m}|\Phi(T))$ and $E(\Psi(0,T)^{\gamma}|\Phi(T))$ for $\gamma \in \mathbb{Q}$. More specifically, the computation of the latter proved particularly slow; the application of negative moments produced some accurate simulation estimates but was not sufficiently fast. On the contrary, the use of exponential moments did speed up the computations.

Figure 3 Entropy bound



Notes. Top plots: absolute difference $\left|H(x)-G_{\Psi|\Phi}(x)\right|$ bound (EC.15) (see Section EC.4.3 of the e-companion) for true distribution G and approximation $H\in\{\hat{G}_{\hat{\Psi}|\Phi},\tilde{G}_{\tilde{\Psi}|\Phi}\}$, where $(\Psi(0,T)|\Phi(T))=\left(\int_0^T V(s)ds\middle|V(T)\right)$ in the Heston model (based on parameter set H1 and condition on the 25, 50, 75th percentiles of the terminal variance), $\hat{G}$ corresponds to the entropy-maximizing distribution with shared m first moments (blue bars) and $\tilde{G}$ to the four-moment fitted Pearson distribution (orange bar). Bottom plots: corresponding entropies $\mathcal{H}[\hat{g}_{\hat{\Psi}|\Phi}]$, $\mathcal{H}[\tilde{g}_{\tilde{\Psi}|\Phi}]$ and $\mathcal{H}[g_{\Psi|\Phi}]$.

However, the resulting errors were found to be substantially bigger, hence we did not consider this any further. More detailed results can be provided upon request.

6.2. Path-independent derivatives

In this section, we aim to assess the accuracy and speed of our proposed Algorithm 2 in the context of pricing path-independent options, in particular, European plain vanilla call options. We compare different simulation schemes starring in the literature, including those of Broadie and Kaya (2006),

Glasserman and Kim (2011), Giles (2008), Giles and Szpruch (2014), Cai et al. (2017) and Li and Wu (2019), and consider the one-factor and two-factor Heston, SABR, OU-SV, 3/2 and 4/2 models as well as models with jumps including Bates, DPS, NIGCIR and SECIRJD. Following the literature, we compute root mean square errors, $\text{RMSE} = \sqrt{\text{bias}^2 + \text{standard error}^2}$, where the $\text{bias} = p_0 - \tilde{p}_0$ is given exactly by (16) based on the difference between the true value and the Pearson proxy following from Theorem 1 and using appropriate parities for certain types of options where necessary. We report in the top panel of Table 3 the true values of at-the-money plain vanilla options, the absolute percentage biases $|p_0 - \tilde{p}_0|/p_0$, but also their upper bound (19) in Theorem 2 with (18) given from (20). We have also verified our reports having computed (16) using very fine Monte Carlo simulation estimates based on $\mathcal{M} = 10^9$ trials (we do not report these here to avoid presenting too many similar results, but we can make them available upon request). Obviously these values are consistently very small across all different parameter sets in Table 2 and translate to accuracies of 3–4 decimal places, if we rely on the exact bias, and are understated up to 1 decimal place if we rely instead on the faster upper error bound, corroborating its tightness. (Results for in-the-money and out-of-the-money options are alike and can be made available upon request.)

Table 3 True option values and biases when using the Pearson approximation

Plain vanilla option						
	H1	H2	H3	H4	H5	H6
True value	6.8061	34.9998	6.7304	7.0972	11.3743	7.0737
Abs. bias (exact)	0.019%	0.019%	0.024%	0.080%	0.020%	0.012%
Abs. bias (upper bound)	0.029%	0.025%	0.104%	0.157%	0.036%	0.015%
	SABR1	SABR2	SABR3	OU-SV1	OU-SV2	OU-SV3
True value	0.0394	0.0436	0.0447	13.2149	40.7977	62.7631
Abs. bias (exact)	0.004%	0.005%	0.000%	0.000%	0.000%	0.000%
Abs. bias (upper bound)	0.020%	0.030%	0.030%	0.019%	0.026%	0.058%
	DH1	DH2	3/2-1	3/2-2	4/2-1	4/2-2
True value	26.9504	9.3663	6.3991	2.1537	8.6592	9.8629
Abs. bias (exact)	0.001%	0.008%	0.002%	0.058%	0.077%	0.009%
Abs. bias (upper bound)	0.025%	0.080%	0.016%	0.086%	0.086%	0.080%
	B1	DPS1	NIGCIR1	SECIRJD1		
True value	20.1645	6.8619	6.6425	9.7639		
Abs. bias (exact)	0.0005%	0.0393%	0.0006%	0.0041%		
Barrier option						
	H1	H2	H3	H4	H5	H6
True value	4.9142	0.1803	6.3748	4.5714	2.6489	4.2925
Abs. bias (exact)	0.008%	0.007%	0.006%	0.009%	0.003%	0.016%
Abs. bias (upper bound)	0.031%	0.028%	0.103%	0.145%	0.039%	0.019%
Asian option						
	H1	H2	H3	H4	H5	H6
True value	3.5665	18.1576	4.1061	4.3222	6.6513	3.8590
Abs. bias (exact)	0.003%	0.015%	0.007%	0.007%	0.008%	0.015%
Abs. bias (upper bound)	0.010%	0.056%	0.114%	0.110%	0.105%	0.018%

Notes. Absolute percentage bias $|p_0 - \tilde{p}_0|/p_0$ is computed both exactly based on Theorem 1 and based on upper bound (19) in Theorem 2 with (18) given from (20). Parities for different option types have been used where necessary. Results are presented for each model (Heston, SABR, OU-SV, Double Heston, 3/2, 4/2, Bates, DPS, NIGCIR and SECIRJD) and set of parameters (see Table 2), for at-the-money European plain vanilla, barrier and Asian call options.

We compute RMSEs for varying standard error depending on the number of simulations $\mathcal{M}$ for the different models and build the speed-accuracy portrayals of the most competing methods. Numerical results are displayed in Table 4, whereas Figures 4–6 present a few relevant illustrations on a log-log scale. As $\mathcal{M}$ increases, the standard error reduces and the RMSE depends eventually on any residual error. In the case of our Algorithm 2 this is usually of the order 10^{-4}, for example, for Heston, double Heston, 3/2, Bates, DPS, NIGCIR and SECIRJD, or even 10^{-6} for the SABR model and, thus, from Figures 4 and 6 we can see how close the RMSEs are to those from methods with optimal convergence like Broadie and Kaya (2006), Giles (2008), Giles and Szpruch (2014), Cai et al. (2017) and Li and Wu (2019). Our scheme exhibits the same convergence rate but more importantly reduces the computing time, hence the shifted parallel plots to the left.

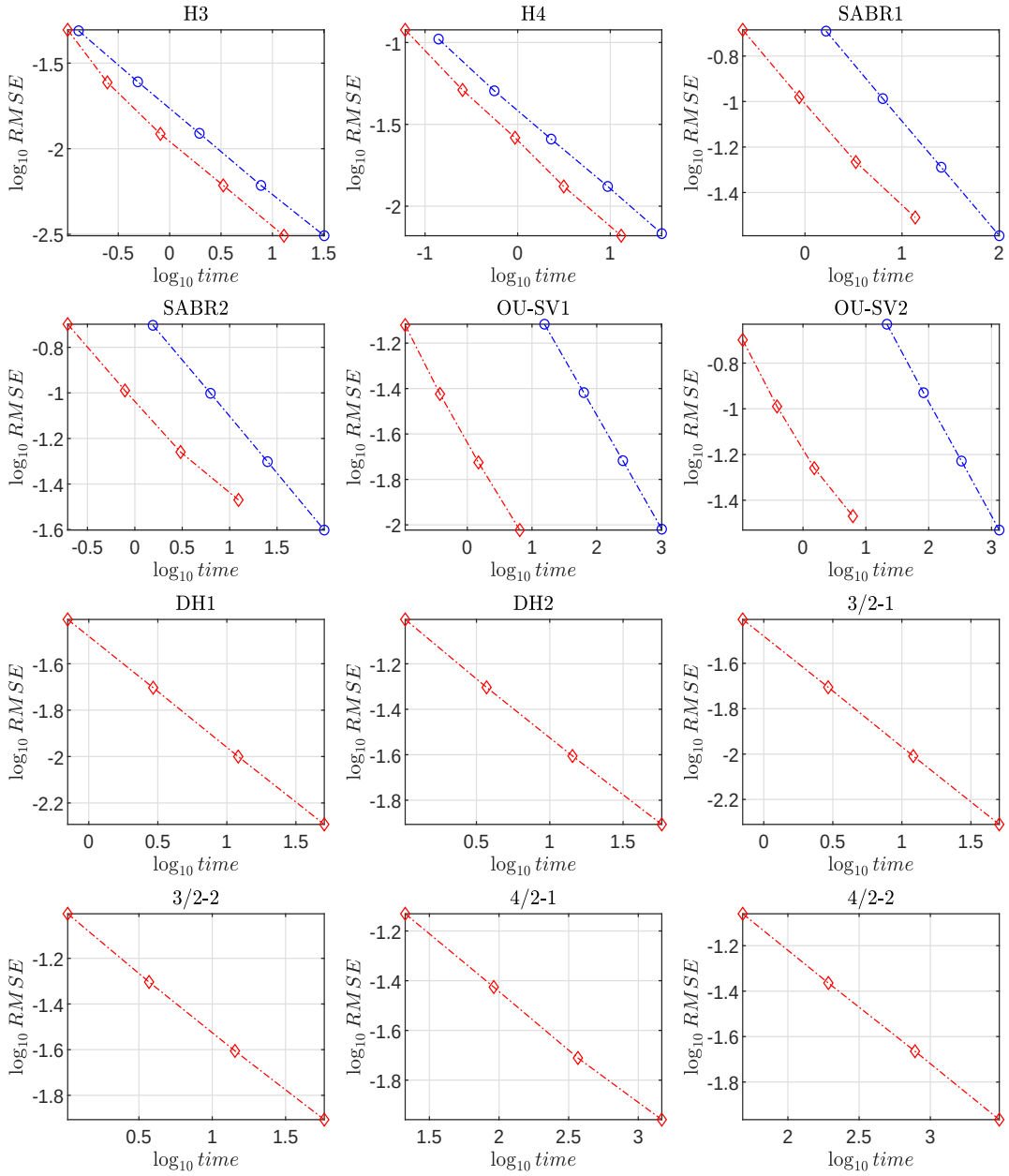
In particular, from Table 4, the computing times compared to Broadie and Kaya (2006), Li and Wu (2019) and Cai et al. (2017) decrease, respectively, by approximately a factor of 800, 190 and 8. In the case of the SABR and OU-SV models, the Laplace transforms (EC.4) and (EC.5) do not involve special functions like Heston's (EC.3), therefore the resulting speed-up gain in the Li and Wu (2019) and Cai et al. (2017) approaches is magnified in our method which gathers momentum becoming impressively faster when computing the required moments using Algorithm 1. Similarly, in relation to the expansion approach of Glasserman and Kim (2011) involving infinite summations that require truncation (here we have used 10 terms), we achieve power saving by a factor of 3–4. In the case of models with jumps, Bates and NIGCIR exhibit close computing times to Heston. The simulation of DPS and SECIRJD is slower as the conditional time integral must be simulated more times according to the (random) number of jumps (which, for the particular choice of parameter values, are expected to be more for the latter).

6.3. Path-dependent derivatives

We also study the efficiency of Algorithm 2 in generating sample paths by turning our attention to the evaluation of path-dependent contracts. As illustrative examples, we consider a barrier (up-and-out) option with a barrier level $\varpi = 120$ and an arithmetic Asian call option, both based on monthly monitoring, in the Heston model. We repeat the error computation exercise as described in the previous section and present our reports in the lower panel of Table 3. Similarly to the path-independent cases, the biases are still very small and in close agreement with their upper bounds, which is not surprising as the size of the time interval does not affect the error.

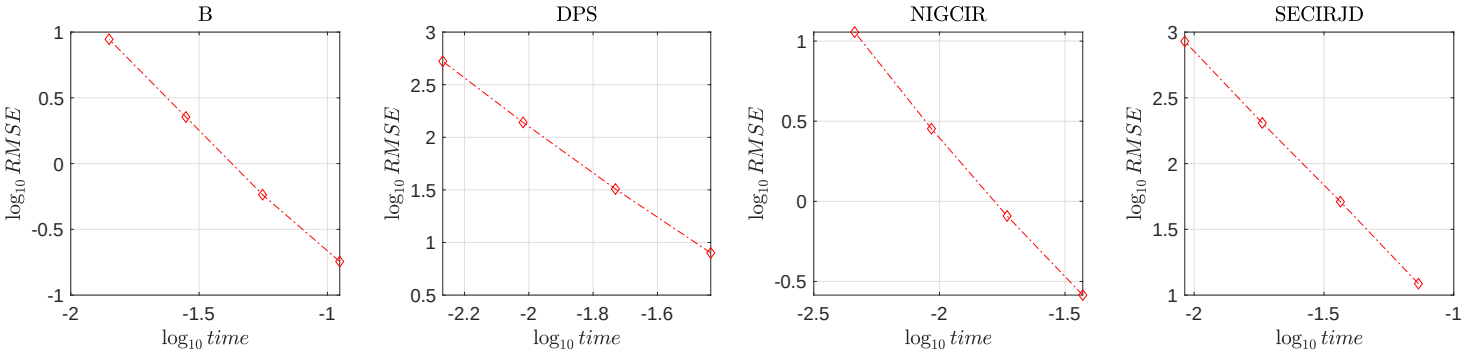
Furthermore, we compare our simulation results based on Algorithm 2 with the time-discretization scheme of Andersen (2008). For a fair comparison, we have tried different number of time steps for the latter and chosen the minimum of 250 time steps per year to have an as similar as possible RMSE (for $\mathcal{M} = 10^4$) based on pre-computed biases for the two methods. Results are

Figure 4 Speed-accuracy comparisons of Algorithm 2 and competent benchmarks for different models and parameter sets: the case of European plain vanilla option



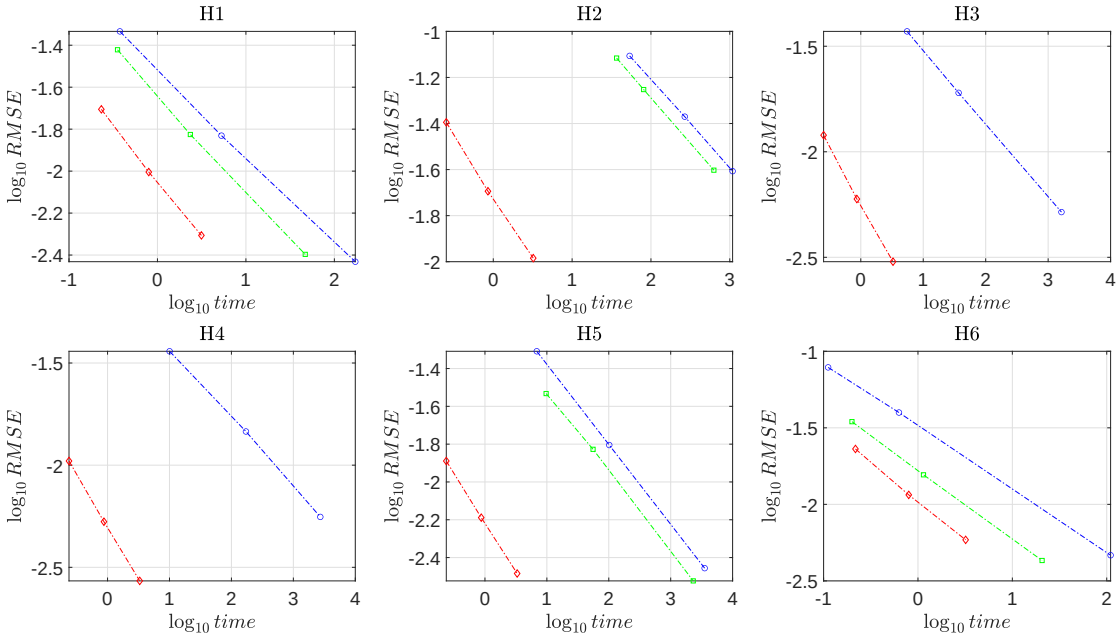
Notes. Algorithm 2: plots with diamond markers; benchmarks: plots with circle markers. Benchmarks: Glasserman and Kim (2011) (H3–4); Cai et al. (2017) (SABR1–2); Li and Wu (2019) (OU-SV1–2). For convenience, out of scale Monte Carlo benchmarks for the DH, 3/2, 4/2 models are not reported. All computing times are in seconds.

Figure 5 Speed-accuracy profiles of Algorithm 2 in the Bates, DPS, NIGCIR and SECIRJD models: the case of European plain vanilla option



Notes. All computing times are in seconds.

Figure 6 Speed-accuracy comparisons of Algorithm 2 and multi-level Monte Carlo (MLMC) methods for different parameter sets in the Heston model: the case of European plain vanilla option



Notes. Algorithm 2: plots with red diamond markers; benchmarks: plots with blue circle (green square) markers correspond to MLMC (antithetic MLMC). MLMC: geometric sequence of time step sizes $h_l = M^{-l}T$, $l = 0, 1, \ldots, L$, for $M = 4$ and $L = 6$, optimal initial number of simulations $\mathcal{M}_0 = 10^4$ (see Giles 2008), target RMSE $= \{0.05, 0.02, 0.005\}$ and simulation performed based on Euler scheme. Antithetic MLMC: based on a Milstein numerical approximation of the Heston SDE with use of antithetic variables (see Giles and Szpruch 2014). All computing times are in seconds.

Table 4 Speed-accuracy profiles of Algorithm 2 and some competent benchmarks (Broadie and Kaya 2006, Glasserman and Kim 2011, Cai et al. 2017, Li and Wu 2019) for different models and parameter sets: the case of European plain vanilla option

	Broadie–Kaya		Algorithm 2		Broadie–Kaya		Algorithm 2		Glasserman–Kim		Algorithm 2	
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H1				H2				H3			
4	0.0373	273.45	0.0372	0.21	0.2904	160.85	0.3000	0.23	0.0246	0.49	0.0244	0.25
16	0.0186	1076.72	0.0186	0.82	0.1464	643.97	0.1440	0.84	0.0123	1.95	0.0122	0.81
64	0.0093	4028.62	0.0093	3.29	0.0726	2458.06	0.0734	3.15	0.0061	7.74	0.0061	3.33
256	0.0046	16884.00	0.0046	13.21	0.0362	10467.25	0.0367	14.28	0.0031	31.96	0.0031	12.95
	Glasserman–Kim		Algorithm 2		Glasserman–Kim		Algorithm 2		Glasserman–Kim		Algorithm 2	
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H4				H5				H6			
4	0.0507	0.56	0.0513	0.25	0.0973	0.86	0.1018	0.28	0.0394	0.82	0.0411	0.31
16	0.0257	2.30	0.0262	0.94	0.0489	3.44	0.0489	0.92	0.0198	3.28	0.0205	0.80
64	0.0132	9.38	0.0132	3.14	0.0248	13.86	0.0249	3.39	0.0099	13.11	0.0103	3.14
256	0.0068	35.94	0.0066	13.11	0.0124	55.43	0.0123	13.09	0.0050	52.94	0.0053	12.28
	Cai–Song–Chen		Algorithm 2		Cai–Song–Chen		Algorithm 2		Cai–Song–Chen		Algorithm 2	
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	SABR1				SABR2				SABR3			
4	2.04e-05	1.65	2.07e-05	0.23	1.98e-05	1.55	2.00e-05	0.2	1.93e-05	1.53	1.97e-05	0.19
16	1.03e-05	6.34	1.05e-05	0.87	9.96e-06	6.28	1.02e-05	0.79	9.76e-06	6.3	9.83e-06	0.78
64	5.14e-06	25.39	5.42e-06	3.35	4.99e-06	25.12	5.50e-06	3.04	4.89e-06	25.07	4.90e-06	2.95
256	2.57e-06	100.7	3.09e-06	13.7	2.50e-06	99.64	3.39e-06	12.41	2.45e-06	99.41	2.45e-06	12.29
	Li–Wu		Algorithm 2		Li–Wu		Algorithm 2		Li–Wu		Algorithm 2	
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	OU-SV1				OU-SV2				OU-SV3			
4	0.0764	15.60	0.0757	0.11	0.2345	21.66	0.2002	0.11	0.3774	21.64	0.3382	0.10
16	0.0382	63.20	0.0377	0.38	0.1176	82.87	0.1024	0.39	0.1891	85.25	0.1687	0.40
64	0.0192	253.55	0.0188	1.49	0.0591	332.22	0.0550	1.51	0.0951	341.73	0.0839	1.49
256	0.0096	1013.31	0.0095	6.47	0.0295	1328.01	0.0339	6.22	0.0475	1366.03	0.0421	6.14
	Algorithm 2		Algorithm 2		Algorithm 2		Algorithm 2		Algorithm 2		Algorithm 2	
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	DH1		DH2		3/2-1		3/2-2		4/2-1		4/2-2	
4	0.0989	0.83	0.0107	0.74	0.0392	0.70	0.0989	0.75	0.0740	2.12	0.0872	2.18
16	0.0497	3.01	0.0053	2.96	0.0197	2.93	0.0497	3.11	0.0375	9.17	0.0432	9.75
64	0.0248	12.29	0.0027	12.11	0.0098	12.09	0.0248	12.39	0.0195	36.85	0.0216	37.29
256	0.0124	48.15	0.0014	47.89	0.0049	50.75	0.0124	51.42	0.0110	147.47	0.0108	149.12
	Algorithm 2		Algorithm 2		Algorithm 2		Algorithm 2					
$\mathcal{M} \times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time				
	B1		DPS1		NIGCIR1		SECIRJD1					
4	0.1117	0.18	0.0368	7.96	0.0372	0.26	0.0731	12.22				
16	0.0559	0.58	0.0186	32.23	0.0186	0.81	0.0366	51.30				
64	0.0281	2.26	0.0096	138.43	0.0093	2.84	0.0183	203.79				
256	0.0141	8.82	0.0054	527.43	0.0046	11.38	0.0092	849.69				

Notes. All computing times are in seconds. For convenience, out of scale Monte Carlo benchmarks for the DH, 3/2, 4/2, Bates, DPS, NIGCIR and SECIRJD models are not reported.

exhibited in Table 5 and Figure 7. We find that Andersen's scheme is faster for a small number of replications. However, as it is obvious from the graphs in Figure 7, for increasing number of simulations, the (large) bias becomes dominant and the RMSE decay slackens severely. This contrasts our method which results in moderate to high levels of precision as $\mathcal{M}$ increases. This is welcome news, as our method constitutes a valid new methodology for applications where the entire sample paths are needed.

We devote the final part of this section to the SABR model given its rising popularity in the literature. First, we apply Algorithm 2 to pricing barrier options and compare with the mSABR method of Leitao et al. (2017), which, even if non-exact, represents a satisfactory benchmark, and

Table 5 Speed-accuracy profiles of Algorithm 2 and Andersen's (2008) method in the Heston model and different parameter sets: the case of path-dependent derivatives

Asian option

	Andersen		Algorithm 2		Andersen		Algorithm 2		Andersen		Algorithm 2	
$\mathcal{M}\times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H1				H2				H3			
4	0.0261	0.98	0.0192	3.42	0.9372	5.03	0.9213	15.78	0.0239	0.86	0.0158	3.01
16	0.0201	5.06	0.0096	12.48	0.5350	27.89	0.5066	59.80	0.0195	3.91	0.0078	12.08
64	0.0183	34.74	0.0048	50.45	0.2908	331.26	0.2343	238.44	0.0183	17.69	0.0039	56.55
256	0.0179	213.71	0.0024	201.43	0.2083	3312.59	0.1172	953.92	0.0180	192.28	0.0020	272.12
	Andersen		Algorithm 2		Andersen		Algorithm 2		Andersen		Algorithm 2	
$\mathcal{M}\times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H4				H5				H6			
4	0.0334	0.86	0.0282	3.27	0.0595	1.07	0.0523	3.51	0.0284	0.95	0.0220	3.94
16	0.0231	3.93	0.0145	12.50	0.0386	4.90	0.0260	14.62	0.0211	5.12	0.0111	15.50
64	0.0194	18.22	0.0072	53.84	0.0314	21.62	0.0132	66.83	0.0188	20.08	0.0056	59.86
256	0.0184	198.61	0.0036	279.79	0.0292	186.54	0.0066	355.53	0.0182	196.87	0.0029	235.37

Barrier option

	Andersen		Algorithm 2		Andersen		Algorithm 2		Andersen		Algorithm 2	
$\mathcal{M}\times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H1				H2				H3			
4	0.0338	0.98	0.0289	3.42	0.0054	5.03	0.0023	15.78	0.0220	0.86	0.0222	3.01
16	0.0233	5.07	0.0145	12.48	0.0031	27.90	0.0012	59.80	0.0111	3.92	0.0110	12.08
64	0.0199	34.76	0.0072	50.46	0.0022	331.17	0.0006	238.46	0.0058	17.71	0.0055	56.56
256	0.0189	213.74	0.0036	201.46	0.0019	3311.68	0.0003	954.00	0.0033	192.29	0.0028	272.15
	Andersen		Algorithm 2		Andersen		Algorithm 2		Andersen		Algorithm 2	
$\mathcal{M}\times 10^4$	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time	RMSE	time
	H4				H5				H6			
4	0.0242	0.87	0.0243	3.27	0.0274	1.07	0.0243	3.51	0.0336	0.95	0.0284	3.95
16	0.0128	3.94	0.0121	12.51	0.0194	4.90	0.0122	14.62	0.0241	5.13	0.0142	15.50
64	0.0076	18.23	0.0060	53.85	0.0169	21.64	0.0061	66.83	0.0210	20.11	0.0071	59.87
256	0.0056	198.62	0.0030	279.82	0.0161	186.63	0.0030	355.57	0.0202	197.01	0.0036	235.40

Notes. All computing times are in seconds.

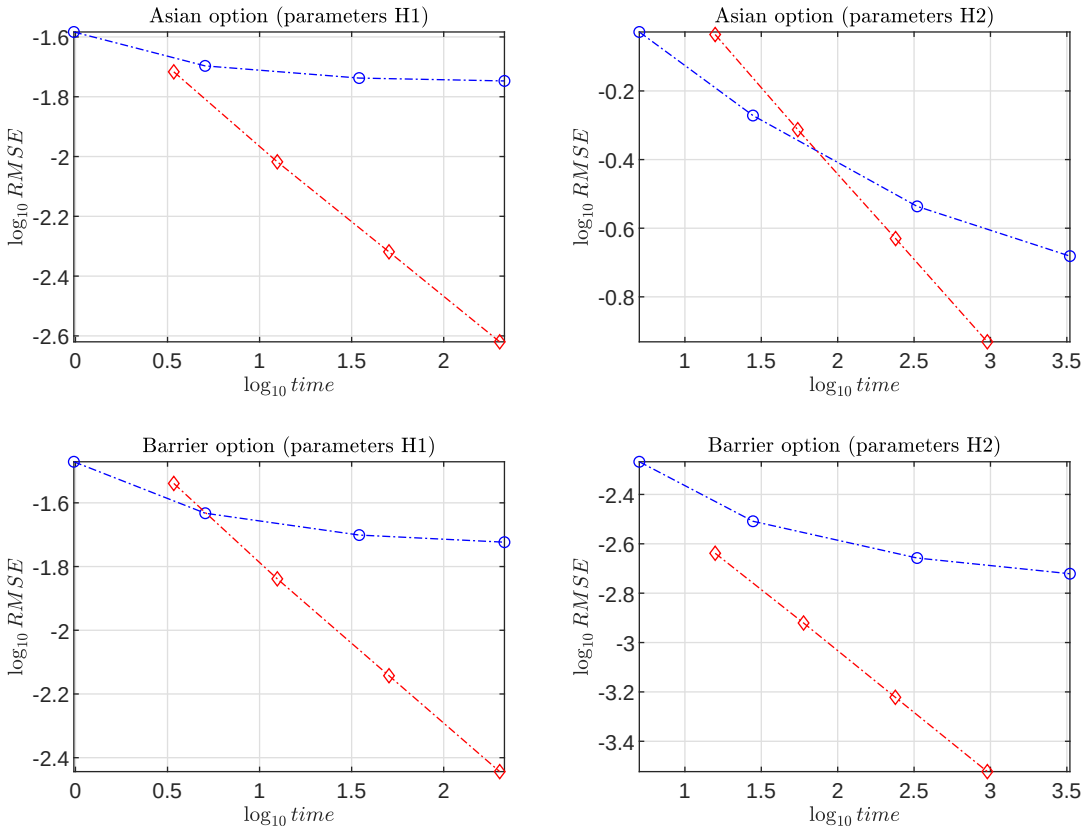
the low-bias simulation scheme of Chen et al. (2012) using the formers' choice of parameters. Based on the reports in Table 6, our Algorithm 2 and the mSABR method agree at 2–3 decimal places, whereas the discrepancies with Chen et al. (2012) are higher. This can be attributed to the larger bias induced by the small-disturbance expansion approximation of moments of the integrated variance and the potentially restrictive lognormal fit they employ. The observed increases in computing time are due to the varying maturity times (see Table 2) subject to quarterly monitoring per annum.

Table 6 Barrier option prices (with standard errors, s.e.) in the SABR model

Method	price	s.e.	time	price	s.e.	time	price	s.e.	time
	SABR4			SABR5			SABR6		
mSABR	0.0384	–	–	0.00523	–	–	0.1078	–	–
Chen et al.	0.0380	1.12e-02	3.77	0.00520	8.43e-04	4.47	0.1022	2.12e-02	21.81
Algorithm 2	0.0385	1.13e-02	21.91	0.00523	8.41e-04	25.72	0.1079	2.12e-02	102.24

Notes. All computing times are in seconds and correspond to $\mathcal{M}=10^6$ simulations and quarterly monitoring (as in Leitao et al. 2017, Section 4). Benchmarks: mSABR (Leitao et al. 2017); Chen et al. (2012). Barrier levels: $\varpi=1.2$ (SABR4); $\varpi=0.08$ (SABR5); $\varpi=2$ (SABR6).

Figure 7 Speed-accuracy profiles of Algorithm 2 and Andersen's (2008) method in the Heston model and different parameter sets: the case of path-dependent derivatives



Notes. Algorithm 2: plots with diamond markers; benchmark: plots with circle markers. Benchmark: Andersen (2008) (H1–2). All computing times are in seconds.

In what follows, we consider further applications entailing accurate simulation solutions. This is especially important in super-linear growth settings where it is well known that, for standard SDEs, the explicit Euler scheme runs into difficulties, performs poorly, or even fails to converge (see Hutzenthaler et al. 2010).

7. Extension to other models

7.1. Linear SDEs: multiplicative noise

We consider the constant-coefficients inhomogeneous model of general form

$$dX(t) = (aX(t) + c)\,dt + (bX(t) + q)\,dW(t). \tag{34}$$

From Kloeden and Platen (1992), (34) has solution

$$X(t) = Y(t;a,b,1)\left(X(0) + (c - bq)\int_0^t Y(s;a,b,-1)ds + q\int_0^t Y(s;a,b,-1)dW(s)\right),$$

where

$$Y(t;a,b,\gamma) := \exp\left(\gamma\left(a-\frac{1}{2}b^2\right)t+\gamma bW(t)\right). \tag{35}$$

The previous solution can be extended to the case of variable coefficients.

A well-known special case of (34) in finance is the Brennan and Schwartz (1980) process

$$dX(t) = \kappa\left(\theta - X(t)\right)dt + \sigma X(t)dW(t) \tag{36}$$

with $\kappa := -a > 0$, $\theta := -c/a \in \mathbb{R}$, $\sigma := b > 0$ and $q := 0$, in consistency with a more standard parameterization of the model (e.g., see Cai et al. 2014b). This, for example, has been used to model interest rate uncertainty, but has also appeared in the energy marketplace with the spot price reverting towards an equilibrium price level (which may be stochastic, that is, the so-called Pilipovic model). In Li et al. (2018), (36) is referred to as a GARCH linear SDE and is used to model stochastic default intensity, with additional references given therein for uses in modelling the volatility or variance of asset returns. Model (36) can be simulated using our Algorithm 3 (see next section) for $n = 0$, $a = -\kappa$, $b = \sigma$ and $c = \kappa\theta$.

7.2. Reducible SDEs

7.2.1. Stochastic generalized logistic (Richards) growth model This is an autonomous, nonlinear reducible SDE model with polynomial drift of degree n of general form

$$dX(t) = \left(cX(t)^n + aX(t)\right)dt + bX(t)dW(t). \tag{37}$$

The substitution x^{1-n} reduces (37) to a linear SDE with multiplicative noise, from which its explicit solution is

$$X(t) = Y(t;a,b,1)\left(X(0)^{1-n} + c(1-n)\int_0^t Y(s;a,b,n-1)ds\right)^{\frac{1}{1-n}}, \tag{38}$$

where $Y(t;a,b,\cdot)$ is given from (35).

We summarize the simulation methodology for model (38) in Algorithm 3. If we are interested in simulating X at a terminal time $T > 0$ only, we use $N = 1$. Simulation steps 3 and 6 of Algorithm 3 are trivial; for step 5, the conditional distribution of $\int_0^t Y(s;a,b,\gamma)ds$ plays key role in the implementation of our methodology. For this, let $W^{(\mu)}(t) = \mu t + W(t)$ be a Brownian motion with constant drift $\mu \in \mathbb{R}$. We also recall from Matsumoto and Yor (2005) the additive functional $A^{(\mu)}(t) = \int_0^t \exp\left(2W^{(\mu)}(s)\right)ds$. Then, it is shown (see also Cai et al. 2017) that

$$E\left[\exp\left(-\frac{u}{A^{(\mu)}(t)}\right)\middle| W^{(\mu)}(t) = w\right] = \exp\left\{-\frac{g(w,u)^2 - w^2}{2t}\right\} \tag{39}$$

for any $t > 0$ and $g(w,u) := \operatorname{arcosh}(ue^{-w} + \cosh w)$.

Algorithm 3 Moment-matched conditional sampling scheme: linear & reducible SDEs

Input: Model parameters, terminal time T, number of monitoring dates N

Output: Sample path $\{X(t)\}$ for $t=\{0,\Delta,2\Delta,3\Delta,...,T\}$

1: Set $\Delta=\frac{T}{N}$

2: **for** $t=0:\Delta:T-\Delta$ **do**

3: Given $Y(t;a,b,1)$, simulate $Y(t+\Delta;a,b,1)$

4: Compute the moments of $\int_t^{t+\Delta}Y(s;a,b,n-1)ds$ conditional on $Y(t+\Delta;a,b,1)$ using Algorithm 1

5: Simulate the conditional $\int_t^{t+\Delta}Y(s;a,b,n-1)ds$ based on (28) given the moments

6: Simulate $X(t+\Delta)$ in (38) given $X(t)$

7: **end for**

8: **return** $\{X(t)\}$ for $t=\{0,\Delta,2\Delta,3\Delta,...,T\}$

PROPOSITION 1. *The Laplace transform of* $\left(\int_0^t Y(s;a,b,\gamma)ds\right)^{-1}$ *conditional on* $Y(t;a,b,\gamma)$ *is given by*

$$E\left[\exp\left(-\frac{u}{\int_0^t Y(s;a,b,\gamma)ds}\right)\middle|\, Y(t;a,b,\gamma)\right]=\exp\left\{-\frac{g\left(\frac{\ln Y(t;a,b,\gamma)}{2},\frac{\gamma^2b^2u}{4}\right)^2-\left(\frac{\ln Y(t;a,b,\gamma)}{2}\right)^2}{\frac{1}{2}\gamma^2b^2t}\right\}. \tag{40}$$

Proof. See e-companion Section EC.1.

The range of admissible values $u\geq-\frac{2}{\gamma^2b^2}\left(\sqrt{Y(t;a,b,\gamma)}-1\right)^2$ ensures that the moment problem is determinate (see Criterion 1). Special cases of practical importance are the stochastic Verhulst and Gordon–Schaefer models for $n=2$ and the stochastic Ginzburg–Landau equation for $n=3$, which we explore further in the following sections.

7.2.2. Stochastic logistic (Verhulst) growth model The Verhulst (1838) model, also known as logistic or $\mathcal{S}$-shaped, belongs to the Richards family of sigmoidal growth models. The deterministic model is described by

$$\frac{d\tilde{X}(t)}{dt}=\lambda\left(K-\tilde{X}(t)\right)\tilde{X}(t), \tag{41}$$

where λ is the Malthusian growth coefficient and $K>0$ a finite supportable carrying capacity in an environment of finite resources. To account for seasonal variations, (41) is augmented by a random component which results in the autonomous stochastic linear-quadratic Verhulst equation

$$d\tilde{X}(t)=\tilde{\lambda}\left(1-\frac{\tilde{X}(t)}{K}\right)\tilde{X}(t)dt+\sigma\tilde{X}(t)dW(t),$$

where W is a standard Brownian motion, $X(0) \in (0,\infty)$, $\tilde{\lambda} := \lambda K$ and $\lambda, \sigma > 0$ are constant. The extinction boundary 0 is non-attractive if $\tilde{\lambda} > \sigma^2/2$ and is attractive if $\tilde{\lambda} < \sigma^2/2$. Also, there is no explosion as the boundary $+\infty$ is non-attractive. Equivalently, we can write

$$dX(t) = \left(\tilde{\lambda} X(t) - X^2(t)\right) dt + \sigma X(t) dW(t), \tag{42}$$

where $X := \lambda \tilde{X}$. In this model, an approximately exponential growth is succeeded by a slowed down linear growth, as saturation begins, and ceases at maturity. The solution exists and is unique for all $t \geq 0$, and from (37)–(38) for $n = 2$

$$X(t) = Y(t;\tilde{\lambda},\sigma,1)\left(X(0)^{-1} + \int_0^t Y(s;\tilde{\lambda},\sigma,1)ds\right)^{-1} = \frac{X(0)Y(t;\tilde{\lambda},\sigma,1)}{1 + X(0)\int_0^t Y(s;\tilde{\lambda},\sigma,1)ds}.$$

This is an environmental stochasticity model meaning that random fluctuations in the environment, such as weather, epidemics, natural disasters, crop failures, can affect the entire population. Originally, model (41) aimed at portraying the self-limiting growth of a biological population. According to this, the rate of reproduction is proportional to the existing population and the amount of available resources. A multi-dimensional version allows for affection of co-existing species by coupling together separate single species models. The logistic model appears also in branches of medicine such as oncology. In Section EC.7.2 of the e-companion, we focus the spotlight on the logistic model for tumor growth and use it to exemplify the application of our simulation method.

7.2.3. Stochastic Gordon–Schaefer model Model (37) is popular also in bioeconomics where it can be used to represent the growth of a population living in a randomly varying environment and being harvested, such as a fish population under fishing, or a wildlife population under hunting, or even a tree population in forestry.

Early important contributions on specific stochastic models for fisheries are due to Beddington and May (1977), Braumann (1985) and Hanson and Ryan (1998). The stochastic Gordon–Schaefer population growth SDE model

$$dX(t) = rX(t)\left(1 - \frac{X(t)}{K}\right) dt - H(t)dt + \sigma X(t) dW(t) \tag{43}$$

includes the term $H(t) := qEX(t)$ which represents the (e.g., fishing) harvesting rate, where $E \geq 0$ is the (fishing) effort (e.g., hours trawled or number of hooks sets per day) and $q > 0$ is the fraction of the population harvested per unit of effort, so that the difference

$$r\left(1 - \frac{X(t)}{K}\right) - qE$$

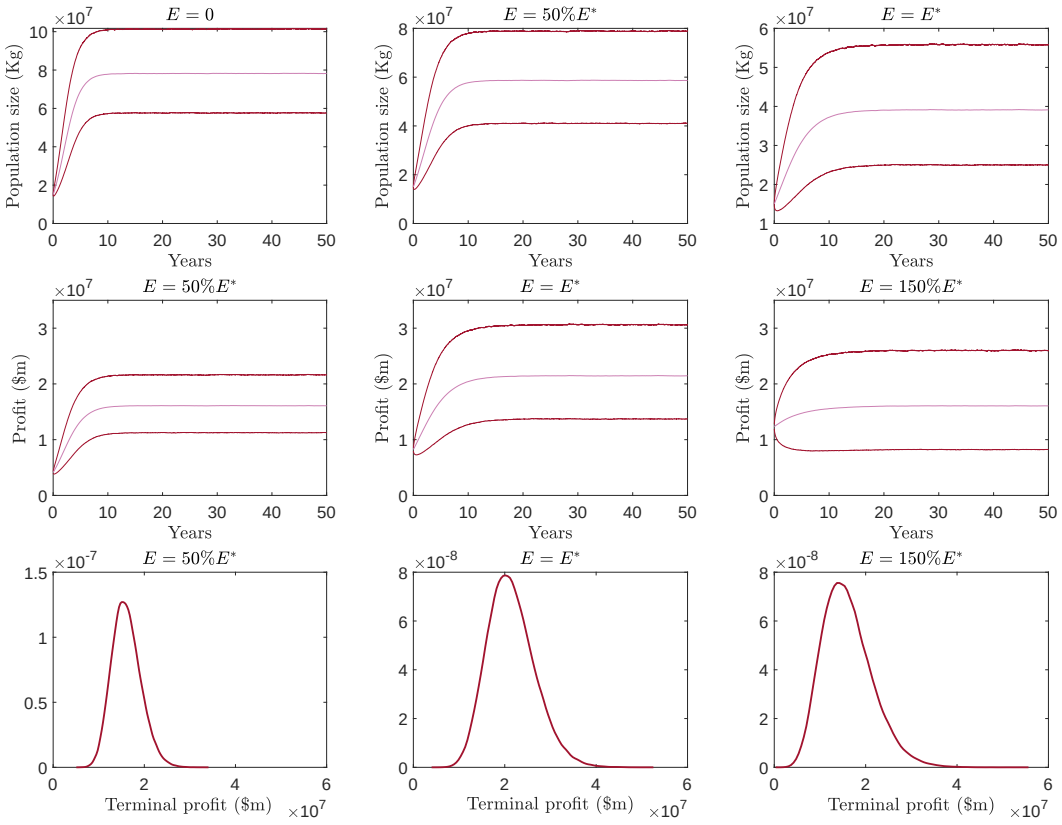
is the natural growth rate adjusted for mortality due to harvesting. Under a basic profit structure, the profit per unit time is

$$\Pi(t) = pH(t) - cE, \tag{44}$$

where the revenues (first term) depend on the harvesting rate and p is the price per unit sold, and the costs (second term) appreciate subject to c cost per unit effort per unit time. p and c can, respectively, be functions of the harvesting rate (higher rate implying lower selling price) and the effort E (increasing effort leading, for example, to overtime and therefore higher cost). Finally, from (37)–(38)

$$X(t) = Y(t; r - qE, \sigma, 1)\left(X(0)^{-1} + \frac{r}{K}\int_0^t Y(s; r - qE, \sigma, 1)ds\right)^{-1}.$$

Figure 8 Impact of varying fishing effort E on population size of Pacific halibut and fishing profit



Notes. Population size (top plots) and profit (central plots): 95% and 5% confidence bands (upper and lower lines) and mean value estimate (medium line) based on 10^5 simulated trajectories. Probability density estimates of terminal profit (bottom plots). Parameter estimates are from Hanson and Ryan (1998): $r = 0.71$, $K = 8.05 \times 10^7$ kg, $q = 3.30 \times 10^{-6}$, $p = 1.59$, $c = 96 \times 10^{-6} + 0.10 \times 10^{-6}E$. In addition, $X(0) = 1.5 \times 10^7$ kg, $\sigma = 0.2$ and $E^* = 104,540$ (optimal sustainable effort).

In Figure 8, we revisit the bioeconomic resource model framework of Hanson and Ryan (1998) for the Pacific halibut. One can determine the optimal sustainable constant effort E^* that maximizes

the expected value of the asymptotic profit. By adopting their parameter estimates, we use our method to generate population size (in kilograms) sample paths. The top panel shows the impact of increasing fishing effort starting from $E = 0$ (top-left), which reduces model (43) to (42), implying a significantly overestimated population size under a misspecified model with harvesting mortality ignored, for example, by 25%, 50% and 75% when $E = 50\%E^*$, E^* and $150\%E^*$ (not explicitly reported in the plot), respectively. In particular, the central and bottom panels focus on the impact of the fishing effort on the profit process (44). Along these panels, the middle plot corresponds to the optimal sustainable effort E^*, while as we diverge from it the profit drops. This mirrors the position of the mode of the estimated density of the terminal profit in the bottom panels, from which we observe that as E reduces below or increases above the optimal level the mode shifts to the left; the standard deviation, skewness and kurtosis also increase with E.

7.2.4. Ginzburg–Landau equation Our last application is devoted to the Ginzburg–Landau equation. In its deterministic version, this was introduced by Ginzburg and Landau (1950) to describe phase transition for superconductivity. Over the years, this model has been used in bistable systems, chemical turbulence, phase transitions in non-equilibrium systems, optics with dissipation, thermodynamics and hydrodynamics, etc. It has also played an important role as a modulation equation and served as a simple model for the transition from regular to turbulent behaviour (see Mielke 2002).

Because random noise is often unavoidable, taking into consideration stochastic disturbances is needed. A stochastic version of it is provided by Kloeden and Platen (1992) and is given by

$$dX(t) = \left(\left(\alpha + \frac{1}{2}\sigma^2\right)X(t) - \beta X^3(t)\right)dt + \sigma X(t)dW(t), \tag{45}$$

where $X(0) \in (0, \infty)$ and $\alpha \geq 0$ and $\beta, \sigma > 0$ are constant. From (38), its solution is given explicitly by

$$X(t) = \frac{X(0)Y(t; \alpha + \frac{1}{2}\sigma^2, \sigma, 1)}{\sqrt{1 + 2X^2(0)\beta \int_0^t Y(s; \alpha + \frac{1}{2}\sigma^2, \sigma, 2)ds}}.$$

Several variants of (45) with a colored noise or regime switching exist.

Here, we focus on (45) and run a simulation experiment borrowed from Hutzenthaler et al. (2010) in order to demonstrate the efficiency of our proposed scheme. Table 7 shows, for different values of the parameter σ, Monte Carlo estimates of $E(X^2(3))$ using our method as well as estimates based on different implementations of the Euler approximation. As σ increases, the bias of the Euler scheme increases and the resulting estimates become perceptibly inaccurate. For very large $\sigma = 6$ or 7, in most of the runs the Euler scheme explodes returning 'NaN' ('not-a-number'), whilst our method remains robust, even under such extremely volatile conditions, and is remarkably faster by a factor of 20.

Table 7 Simulation of $E\left(X^2(3)\right)$ (with standard errors, s.e.) in the stochastic Ginzburg–Landau model

	$\sigma=2$		$\sigma=4$		$\sigma=5$		$\sigma=6$		$\sigma=7$	
	$E\left(X^2(3)\right)$	s.e.	$E\left(X^2(3)\right)$	s.e.	$E\left(X^2(3)\right)$	s.e.	$E\left(X^2(3)\right)$	s.e.	$E\left(X^2(3)\right)$	s.e.
Algorithm 3	0.4689	4.12e-04	0.9138	1.18e-03	1.1455	1.66e-03	1.3693	2.18e-03	1.5989	2.75e-03
Euler approximation										
Batch 1	0.4556	4.10e-03	0.7553	1.10e-02	0.7106	1.33e-02	NaN	NaN	NaN	NaN
Batch 2	0.4514	4.09e-03	0.7551	1.12e-02	0.6964	1.32e-02	NaN	NaN	NaN	NaN
Batch 3	0.4576	4.16e-03	0.7582	1.10e-02	0.7065	1.33e-02	0.5191	1.43e-02	NaN	NaN
Batch 4	0.4572	4.08e-03	0.7417	1.08e-02	0.7126	1.39e-02	0.5353	1.43e-02	NaN	NaN
Batch 5	0.4608	4.09e-03	0.7274	1.05e-02	0.7043	1.36e-02	NaN	NaN	NaN	NaN
Batch 6	0.4491	4.07e-03	0.7308	1.05e-02	0.7224	1.36e-02	0.5415	1.47e-02	NaN	NaN
Batch 7	0.4595	4.08e-03	0.7393	1.08e-02	0.7310	1.42e-02	NaN	NaN	NaN	NaN
Batch 8	0.4592	4.03e-03	0.7333	1.07e-02	0.7344	1.43e-02	NaN	NaN	NaN	NaN
Batch 9	0.4589	4.10e-03	0.7539	1.11e-02	0.7108	1.34e-02	NaN	NaN	NaN	NaN
Batch 10	0.4626	4.12e-03	0.7555	1.08e-02	0.7096	1.34e-02	0.5366	1.41e-02	NaN	NaN

Notes. The number of Monte Carlo simulations used for each estimate is 10^5 for the Euler approximation (based on $1,000$ time steps) and 10^7 for Algorithm 3. Parameters are from Hutzenthaler et al. (2010): $X(0)=1$, $\alpha=0$, $\beta=1$, $\sigma=\{2,4,5,6,7\}$.

8. Conclusion

In this paper, we propose a novel method for simulating integrals of general stochastic processes. We focus our analysis on particularly hard cases of dependence on the terminal value of the process. We manage to relax the most time-consuming parts of other methods: i) the repeatedly recovered conditional distribution by numerical transform inversion; ii) the large number of time steps that mediate the actual monitoring dates with the purpose of reducing the bias of time-discretization techniques; iii) the number of terms in expansion approaches; iv) the achievement of the required moments by directly differentiating a Laplace transform.

Our method is built in a well-posed moment problem setting which encloses other distributions that are unique by their moments, beyond integrated processes. The resulting moment-based random number generation is accurate, with very small and bounded error, and fast, improving on the complexity and speed of pre-existing techniques. Bypassing the most laborious parts of others, we are able to generate entire paths of stochastic processes of interest in a reasonable time. We explore the flexibility and robustness of our method in different practical problem structures and show that it can effectively contribute to the interface of simulation with various areas in financial engineering, medicine, bioeconomics and physics and, thereby, lead to useful impact in the application domain.

Acknowledgments

We thank the Area Editor, Jeff Hong, for giving us the opportunity to respond to our reviewers' comments and revise our manuscript accordingly. We are grateful to an anonymous Associate Editor and three reviewers for their constructive feedback and motivating comments that helped us revise the paper to meet the standard expected by the journal. We are indebted to Steven Kou for fruitful comments and discussions that helped broaden the scope of this research from an early stage of it and to Russell Gerrard for abundant discussions on technical aspects. Special thanks go to Zhenyu Cui, Justin Lars Kirkby, Duy Nguyen, Robert Mnatsakanov,

Angelos Dassios, Hongbiao Zhao and Luca De Gennaro Aquino, for various discussions and contributed materials. Earlier versions of this paper were presented at the SIAM Conference on Financial Mathematics & Engineering (FM19) in Toronto, 2019 INFORMS Annual Meeting in Seattle, and XXI Workshop on Quantitative Finance in Naples; we are thankful for all the comments received. The usual disclaimer applies.

References

Abramowitz M, Stegun IA (1968) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (New York: Dover Books on Mathematics), 2 edition.

Akhiezer NI (1965) *The Classical Moment Problem and Some Related Questions in Analysis*. University Mathematical Monographs (Oliver & Boyd).

Andersen L (2008) Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance* 11(3):1–42.

Ballotta L, Kyriakou I (2014) Monte Carlo simulation of the CGMY process and option pricing. *Journal of Futures Markets* 34(12):1095–1121.

Bates DS (1996) Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Review of Financial Studies* 9(1):69–107.

Beddington JR, May RM (1977) Harvesting natural populations in a randomly fluctuating environment. *Science* 197(4302):463–465.

Braumann CA (1985) Stochastic differential equation models of fisheries in an uncertain world: Extinction probabilities, optimal fishing effort, and parameter estimation. Capasso V, Grosso E, Paveri-Fontana SL, eds., *Mathematics in Biology and Medicine*, 201–206 (New York: Springer).

Brennan MJ, Schwartz ES (1980) Analyzing convertible bonds. *The Journal of Financial and Quantitative Analysis* 15(4):907–929.

Broadie M, Glasserman P (1996) Estimating security price derivatives using simulation. *Management Science* 42(2):269–285.

Broadie M, Kaya O (2004) Exact simulation of option Greeks under stochastic volatility and jump diffusion models. Ingalls RG, Rossetti MD, Smith JS, Peters BA, eds., *Proceedings of the 2004 Winter Simulation Conference*, volume 2, 1607–1615.

Broadie M, Kaya O (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* 54(2):217–231.

Cai N, Kou SG, Liu Z (2014a) A two-sided Laplace inversion algorithm with computable error bounds and its applications in financial engineering. *Advances in Applied Probability* 46(3):766–789.

Cai N, Li C, Shi C (2014b) Closed-form expansions of discretely monitored Asian options in diffusion models. *Mathematics of Operations Research* 39:789–822.

Cai N, Song Y, Chen N (2017) Exact simulation of the SABR model. *Operations Research* 65(4):931–951.

Callegaro G, Fiorin L, Grasselli M (2019) Quantization meets Fourier: a new technology for pricing options. *Annals of Operations Research* 282:59–86.

Carr P, Geman H, Madan D, Yor M (2003) Stochastic volatility for Lévy processes. *Mathematical Finance* 13(3):345–382.

Černý A, Kyriakou I (2011) An improved convolution algorithm for discretely sampled Asian options. *Quantitative Finance* 11(3):381–389.

Chen B, Oosterlee CW, van der Weide H (2012) A low-bias simulation scheme for the SABR stochastic volatility model. *International Journal of Theoretical and Applied Finance* 15(2):125–161.

Chignola R, Schenetti A, Chiesa E, Foroni R, Sartoris S, Brendolan A, Tridente G, Andrighetto G, Liberati D (1999) Oscillating growth patterns of multicellular tumour spheroids. *Cell Proliferation* 32(1):39–48.

Choudhury GL, Lucantoni DM (1996) Numerical computation of the moments of a probability distribution from its transform. *Operations Research* 44(2):368–381.

Christoffersen P, Heston S, Jacobs K (2009) The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science* 55(12):1914–1932.

Cont R, Tankov P (2004) *Financial Modelling With Jump Processes*. Financial Mathematics Series (Boca Raton: Chapman & Hall/CRC).

Corsaro S, Kyriakou I, Marazzina D, Marino Z (2019) A general framework for pricing Asian options under stochastic volatility on parallel architectures. *European Journal of Operational Research* 272:1082–1095.

Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. *Econometrica* 53:385–407.

Cressie N, Borkent M (1986) The moment-generating function has its moments. *Journal of Statistical Planning and Inference* 13:337–344.

Cressie N, Davis AS, Folks JL, Policello GE (1981) The moment-generating function and negative integer moments. *The American Statistician* 35:148–150.

Cui Z, Kirkby JL, Lian G, Nguyen D (2017) Integral representation of probability density of stochastic volatility models and timer options. *International Journal of Theoretical and Applied Finance* 20(8):1750055–1–32.

Cui Z, Kirkby JL, Nguyen D (2021) Efficient simulation of generalized SABR and stochastic local volatility models based on Markov chain approximations. *European Journal of Operational Research* 290(3):1046–1062.

Dassios A, Zhao H (2017) Efficient simulation of clustering jumps with CIR intensity. *Operations Research* 65(6):1494–1515.

Demidenko E (2013) *Mixed Models: Theory and Applications with R*. Wiley Series in Probability and Statistics (New York: Wiley), 2 edition.

Devroye L (1986) *Non-Uniform Random Variate Generation* (New York: Springer).

Du D, Luo D (2019) The pricing of jump propagation: Evidence from spot and options markets. *Management Science* 65(5):2360–2387.

Duffie D, Pan J, Singleton K (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6):1343–1376.

Elderton WP, Johnson NL (1969) *Systems of Frequency Curves* (Cambridge: Cambridge University Press).

Esseen CG (1945) Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Mathematica* 77:1–125.

Fang F, Oosterlee CW (2008) A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing* 31(2):826–848.

Feller W (1971) *An Introduction to Probability Theory and Its Applications*, volume II (New York: John Wiley and Sons), 2 edition.

Fulop A, Li J (2019) Bayesian estimation of dynamic asset pricing models with informative observations. *Journal of Econometrics* 209(1):114–138.

Fusai G, Tagliani A (2002) An accurate valuation of Asian options using moments. *International Journal of Theoretical and Applied Finance* 5(2):147–169.

Gauthier P, Possamaï D (2010) Efficient simulation of the double Heston model, URL `https://ssrn.com/abstract=1434853`, Working Paper, Daiwa Capital Markets Europe.

Giesecke K, Kakavand H, Mousavi M (2011) Exact simulation of point processes with stochastic intensities. *Operations Research* 59:1233–1245.

Giles MB (2008) Multilevel Monte Carlo path simulation. *Operations Research* 56:607–617.

Giles MB, Szpruch L (2014) Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Annals of Applied Probability* 24:1585–1620.

Ginzburg VL, Landau LD (1950) On the theory of superconductivity. *Zh. Eksp. Teor. Fiz.* 20:1064–1082.

Glasserman P, Kim KK (2011) Gamma expansion of the Heston stochastic volatility model. *Finance and Stochastics* 15:267–296.

Gnoatto A, Grasselli M, Platen E (2016) A penny saved is a penny earned: Less expensive zero coupon bonds, URL `https://ssrn.com/abstract=2824564`, Working Paper, University of Verona.

Grasselli M (2017) The 4/2 stochastic volatility model: A unified approach for the Heston and the 3/2 model. *Mathematical Finance* 27(4):1013–1034.

Hagan PS, Kumar D, Lesniewski AS, Woodward DE (2002) Managing smile risk. *Wilmott Magazine* 84–108.

Hanson FB, Ryan D (1998) Optimal harvesting with both population and price dynamics. *Mathematical Biosciences* 148(2):129–146.

Heinrich J (2004) A guide to the Pearson Type IV distribution, available online at `https://www-cdf.fnal.gov/physics/statistics/notes/cdf6820_pearson4.pdf`.

Heston S (1997) A simple new formula for options with stochastic volatility. Technical report, Washington University of St. Louis.

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

Hörmann W, Leydold J, Derflinger G (2004) *Automatic Nonuniform Random Variate Generation*. Statistics and Computing (New York: Springer).

Hutzenthaler M, Jentzen A, Kloeden PE (2010) Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society A* 467:1563–1576.

Islah O (2009) Solving SABR in exact form and unifying it with LIBOR market model, URL `https://ssrn.com/abstract=1489428`, Working Paper, Lloyds Banking Group, Quantuply Ltd.

Johnson NL, Kotz S, Balakrishnan N (1994) *Continuous Univariate Distributions* (New York: Wiley), 2 edition.

Kang C, Kang W, Lee JM (2017) Exact simulation of the Wishart multidimensional stochastic volatility model. *Operations Research* 65(5):1190–1206.

Kapur JN, Kesavan HK (1992) Entropy optimization principles and their applications. Singh VP, Fiorentino M, eds., *Entropy and Energy Dissipation in Water Resources*, 3–20 (Dordrecht: Springer Netherlands).

Khamis SH (1954) On the reduced moment problem. *The Annals of Mathematical Statistics* 25(1):113–122.

Klebanov LB, Mkrtchyan ST (1986) Estimation of the closeness of distributions in terms of identical moments. *Journal of Soviet Mathematics* 32(1):54–60.

Kloeden PE, Platen E (1992) *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability (New York: Springer).

Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4:83–91.

Kou SG (2002) A jump-diffusion model for option pricing. *Management Science* 48(8):1086–1101.

Kullback S (1967) A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory* 13(1):126–127.

Laird AK (1964) Dynamics of tumor growth. *British Journal of Cancer* 13:490–502.

Leitao Á, Grzelak LA, Oosterlee CW (2017) On an efficient multiple time step Monte Carlo simulation of the SABR model. *Quantitative Finance* 17:1549–1565.

Lévy P (1925) *Calcul des Probabilités* (Paris: Gauthier-Villars).

Li C, Wu L (2019) Exact simulation of the Ornstein–Uhlenbeck driven stochastic volatility model. *European Journal of Operational Research* 275:768–779.

Li M, Mercurio F, Resnick S (2018) The Garch linear SDE: explicit formulas and the pricing of a quanto CDS, available online at `https://www.risk.net/media/download/1001271/download`.

Lin W, Li S, Luo A, Chern S (2017) Consistent pricing of VIX and equity derivatives with the 4/2 stochastic volatility plus jumps model. *Journal of Mathematical Analysis and Applications* 447:778–797.

Lindsay BG, Basak P (2000) Moments determine the tail of a distribution (but not much else). *The American Statistician* 54(4):248–251.

Lindsay BG, Pilla RS, Basak P (2000) Moment-based approximations of distributions using mixtures: Theory and applications. *Annals of the Institute of Statistical Mathematics* 52(2):215–230.

Linnik JV, Ostrovskiĭ IV (1977) *Decomposition of Random Variables and Vectors*, volume 48 of *Translations of Mathematical Monographs* (Providence, R. I.: American Mathematical Society).

Lord R, Fang F, Bervoets F, Oosterlee CW (2008) A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM Journal on Scientific Computing* 30:1678–1705.

Marsaglia G, Tsang WW (2000) A simple method for generating gamma variables. *ACM Transactions on Mathematical Software* 26:363–372.

Matsumoto H, Yor M (2005) Exponential functionals of Brownian motion, I: Probability laws at fixed time. *Correspondance mathématique et physique* 2:312–347.

Mielke A (2002) The Ginzburg-Landau equation in its role as a modulation equation. Fiedler B, ed., *Handbook of Dynamical Systems*, volume 2 of *Handbook of Dynamical Systems*, 759–834 (Elsevier Science).

Mijatović A, Urusov M (2012) On the martingale property of certain local martingales. *Probability Theory and Related Fields* 152(1):1–30.

Nadarajah S (2008) A review of results on sums of random variables. *Acta Applicandae Mathematicae* 103(2):131–140.

Obukhov AM (1959) Description of turbulence in terms of Lagrangian variables. volume 6 of *Advances in Geophysics*, 113–116 (Elsevier).

Parrish RS (1987) Evaluation and approximation of multivariate cumulative joint probabilities. *Journal of Statistical Computation and Simulation* 27:1–33.

Parrish RS (1990) Generating random deviates from multivariate Pearson distributions. *Computational Statistics & Data Analysis* 9:283–295.

Platen E (1997) A non-linear stochastic volatility model. *Financial Mathematics Research Report No. FMRR 005-97, Center for Financial Mathematics, Australian National University, Canberra* .

Rachev ST, Klebanov L, Stoyanov SV, Fabozzi F (2013) *The Methods of Distances in the Theory of Probability and Statistics* (New York: Springer-Verlag).

Rose C, Smith MD (2002) *Mathematical Statistics with Mathematica* (New York: Springer).

Rubinstein M (1991) Exotic options. Research Program in Finance Working Papers RPF-220, University of California at Berkeley, URL `https://EconPapers.repec.org/RePEc:ucb:calbrf:rpf-220`.

Schöbel R, Zhu J (1999) Stochastic volatility with an Ornstein–Uhlenbeck process: An extension. *European Finance Review* 3(1):23–46.

Scott L (1987) Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial Quantitative Analysis* 22(4):419–438.

Sesana D, Marazzina D, Fusai G (2014) Pricing exotic derivatives exploiting structure. *European Journal of Operational Research* 236(1):369–381.

Shen CY (2020) Logistic growth modelling of COVID-19 proliferation in China and its international implications. *International Journal of Infectious Diseases* 96(1201-9712):582–589.

Stein E, Stein J (1991) Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies* 4(4):727–752.

Stoyanov JM (2013) *Counterexamples in Probability* (New York: Dover Publications), 3 edition.

Verhulst PF (1838) Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique* 10:113–121.

Wachter JA (2013) Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance* 68(3):987–1035.

Wu K, Darcet D, Wang Q, Sornette D (2020) Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynamics* 101(3):1561–1581.

Zhang SM, Feng Y (2019) American option pricing under the double Heston model based on asymptotic expansion. *Quantitative Finance* 19(2):211–226.

Zolotarev VM (1983) Probability metrics. *Theory of Probability & Its Applications* 28(2):278–302.

E-companion to "Unified moment-based modelling of integrated stochastic processes"

EC.1. Auxiliary results

Proof of Proposition 1 We have that

$$\begin{aligned}\int_0^t Y(s;a,b,\gamma)ds &= \int_0^t \exp\left(\gamma\left(a-\frac{1}{2}b^2\right)s+\gamma bW(s)\right)ds\\ &\overset{\text{(law)}}{=} \frac{4}{\gamma^2b^2}\int_0^{\gamma^2b^2t/4}\exp\left(\frac{4}{\gamma}\left(\frac{a}{b^2}-\frac{1}{2}\right)s+2W(s)\right)ds\\ &= \frac{4}{\gamma^2b^2}\int_0^{\gamma^2b^2t/4}\exp\left(2W^{(\mu)}(s)\right)ds,\end{aligned}$$

where the second equality follows from the scaling property of the Brownian motion for $\gamma, b>0$ and $\mu := 2a/(\gamma b^2) - 1/\gamma$. In addition, by considering the condition in (39), we get

$$w = W^{(\mu)}\left(\frac{\gamma^2b^2t}{4}\right) \overset{\text{(law)}}{=} \mu\frac{\gamma^2b^2t}{4}+\frac{\gamma b}{2}W(t) = \frac{\ln Y(t;a,b,\gamma)}{2}.$$

Expression (40) then follows and the proposition is proved.

PROPOSITION EC.1. *The Pearson characteristic function is given by* $\varphi_{\bar{G}}(u) = e^{iu\mu_1}\varphi_{\bar{G}}(u)$, *where* $i=\sqrt{-1}$ *and* $\varphi_{\bar{G}}(u) = \int_{\mathbb{R}} e^{iux}\bar{g}(x)\,dx$, *with* $\bar{g}$ *solving (1), satisfies the system of ordinary differential equations*

$$\begin{cases}\varphi'_{\bar{G}}(u) = \vartheta_{\bar{G}}(u)\\ \vartheta'_{\bar{G}}(u) = \frac{uc_2+i(1-2c_3)}{iuc_3}\vartheta_{\bar{G}}(u)+\frac{c_1}{c_3}\varphi_{\bar{G}}(u)\end{cases} \tag{EC.1}$$

with set of initial conditions $(\varphi_{\bar{G}}(0), \vartheta_{\bar{G}}(0)) = (1,0)$.

Proof. From (1),

$$\begin{aligned}\int_{\mathbb{R}} e^{iux}\bar{g}(x)\,dx &= \left[e^{iux}\left(c_1+c_2x+c_3x^2\right)\bar{g}(x)\right]_{-\infty}^{\infty}\\ &\quad -\int_{\mathbb{R}} e^{iux}\left\{iu\left(c_1+c_2x+c_3x^2\right)+\left(c_2+2c_3x\right)\right\}\bar{g}(x)\,dx\\ &= -\left(iuc_1+c_2\right)\int_{\mathbb{R}} e^{iux}\bar{g}(x)\,dx-\left(iuc_2+2c_3\right)\int_{\mathbb{R}} e^{iux}x\bar{g}(x)\,dx\\ &\quad -iuc_3\int_{\mathbb{R}} e^{iux}x^2\bar{g}(x)\,dx,\end{aligned} \tag{EC.2}$$

where in the first equality $\left[e^{iux}\left(c_1+c_2x+c_3x^2\right)\bar{g}(x)\right]_{-\infty}^{\infty}=0$ by $\mu_2<\infty$. Differentiation of $\varphi_{\bar{G}}(u)$ under the integral sign is allowed because $e^{iux}\bar{g}(x)$ is twice continuously differentiable in u and once in x. Hence, we obtain $\varphi'_{\bar{G}}(u) = i\int_{\mathbb{R}} e^{iux}x\bar{g}(x)\,dx$ and $\varphi''_{\bar{G}}(u) = -\int_{\mathbb{R}} e^{iux}x^2\bar{g}(x)\,dx$, based on which (EC.2) can be written as

$$iuc_3\varphi''_{\bar{G}}(u)+\left(-uc_2+2ic_3-i\right)\varphi'_{\bar{G}}(u)+\left(c_0-iuc_1-c_2\right)\varphi_{\bar{G}}(u)=0.$$

From this, (EC.1) follows for $c_0 - c_2 = 0$ by construction. Given $\tilde{g}(x) = \bar{g}(x - \mu_1)$, we finally get $\varphi_{\tilde{G}}(u) = \int_{\mathbb{R}} e^{iu(x+\mu_1)} \bar{g}(x) dx = e^{iu\mu_1} \varphi_{\bar{G}}(u)$.

An analytic solution of (EC.1) can be obtained with the aid of a software system that allows symbolic computations, such as Mathematica, which we omit here but can make available upon request.

EC.2. Laplace transforms

The conditional Laplace transforms for the Heston, SABR, OU-SV and 4/2 models follow, respectively, from Broadie and Kaya (2006), Cai et al. (2017), Li and Wu (2019) and Grasselli (2017).

EC.2.1. Heston model

The Laplace transform of the conditional integrated variance is

$$\begin{aligned}\mathcal{L}(a) = E\left[\exp\left(-a\int_u^t V(s)ds\right)\middle| V(u), V(t)\right] = \frac{\gamma(a)e^{-(\gamma(a)-k)(t-u)/2}(1-e^{-k(t-u)})}{k(1-e^{-\gamma(a)(t-u)})} \times \\ \exp\left\{\frac{V(u)+V(t)}{v^2}\left(\frac{k(1+e^{-k(t-u)})}{1-e^{-k(t-u)}} - \frac{\gamma(a)(1+e^{-\gamma(a)(t-u)})}{1-e^{-\gamma(a)(t-u)}}\right)\right\} \times \\ \frac{I_{d/2-1}\left(\sqrt{V(u)V(t)}\frac{4\gamma(a)e^{-\gamma(a)(t-u)/2}}{v^2(1-e^{-\gamma(a)(t-u)})}\right)}{I_{d/2-1}\left(\sqrt{V(u)V(t)}\frac{4ke^{-k(t-u)/2}}{v^2(1-e^{-k(t-u)})}\right)},\end{aligned} \tag{EC.3}$$

where $\gamma(a) := \sqrt{k^2 + 2v^2 a}$ and $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind.

EC.2.2. SABR model

The Laplace transform of the conditional reciprocal of the integrated variance is

$$\begin{aligned}\mathcal{L}(a) &= E\left[\exp\left\{-a\left(\int_u^t V^2(s)ds\right)^{-1}\right\}\middle| V(u), V(t)\right] \\ &= \exp\left\{-\frac{g\left(\ln\frac{V(t)}{V(u)}, \frac{av^2}{V^2(u)}\right)^2 - \ln\left(\frac{V(t)}{V(u)}\right)^2}{2v^2(t-u)}\right\},\end{aligned} \tag{EC.4}$$

where $g(x, \lambda) := \operatorname{arcosh}(\lambda e^{-x} + \cosh x)$.

EC.2.3. OU-SV model

The Laplace transform of the conditional integrated variance is

$$\mathcal{L}(a) = E\left[\exp\left(-a\int_u^t V^2(s)ds\right)\middle| V(u), V(t), \int_u^t V(s)ds\right] = \frac{f(\gamma(a))}{f(k)}, \tag{EC.5}$$

where

$$f(x) := \frac{x^2}{2\pi\sqrt{\eta(x)}} \exp\left\{ -\frac{1}{2\eta(x)v^2} \left[2x^2 \left((V(t)+V(u)) \int_u^t V(s)ds - V(u)V(t)(t-u) \right) + x^2 \left((V^2(t)+V^2(u))(t-u) - 2(V(t)+V(u)) \int_u^t V(s)ds \right) \cosh(x(t-u)) + x \left(x^2 \left(\int_u^t V(s)ds \right)^2 - (V(t)-V(u))^2 \right) \sinh(x(t-u)) \right] \right\}$$

and

$$\eta(x) := 2 - 2\cosh(x(t-u)) + x(t-u)\sinh(x(t-u)).$$

EC.2.4. 4/2 and 3/2 models

The joint Laplace transform of $\left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)} \middle| V(u), V(t) \right)$ is

$$\begin{aligned} & E\left[\exp\left(-a_1 \int_u^t V(s)ds - a_2 \int_u^t \frac{ds}{V(s)} \right) \middle| V(u), V(t) \right] \\ &= \frac{\sqrt{\gamma(a_1)}\sinh\frac{k(t-u)}{2}}{k\sinh\frac{\sqrt{\gamma(a_1)}(t-u)}{2}} \exp\left\{ \frac{V(u)+V(t)}{v^2} \left(k\coth\frac{k(t-u)}{2} - \sqrt{\gamma(a_1)}\coth\frac{\sqrt{\gamma(a_1)}(t-u)}{2} \right) \right\} \\ &\times \frac{I_{\frac{\sqrt{(2\theta k - v^2)^2 + 8v^2 a_2}}{v^2}} \left(\frac{2\sqrt{\gamma(a_1)V(u)V(u)}}{v^2 \sinh\frac{\sqrt{\gamma(a_1)}(t-u)}{2}} \right)}{I_{\frac{2\theta k}{v^2}-1} \left(\frac{2k\sqrt{V(u)V(t)}}{v^2\sinh\frac{k(t-u)}{2}} \right)}. \end{aligned} \tag{EC.6}$$

The Laplace transform of $\left(\int_u^t \frac{ds}{V(s)} \middle| V(u), V(t) \right)$ in the 3/2 model follows as special case by setting $a_1 = 0$ in (EC.6).

In addition, we have for $s(t) = \ln S(t)$ conditional on $V(t)$ that

$$\begin{aligned} E\left[e^{us(t)} \middle| V(t) \right] &= \exp\left\{ us(0) + u\left(r - ab - \frac{a\rho k\theta}{v} + \frac{b\rho k}{v} \right) t + u^2\left(1-\rho^2\right) abt \right\} \\ &\times \exp\left\{ \frac{ua\rho}{v}\left(V(t)-V(0)\right) + \frac{ub\rho}{v}\ln\frac{V(t)}{V(0)} \right\} \\ &\times \frac{\sqrt{A_u}\sinh\frac{kt}{2}}{k\sinh\frac{\sqrt{A_u}t}{2}} \exp\left\{ \frac{V(0)+V(t)}{v^2}\left(k\coth\frac{kt}{2} - \sqrt{A_u}\coth\frac{\sqrt{A_u}t}{2} \right) \right\} \\ &\times \frac{I_{\frac{2}{v^2}\sqrt{\left(k\theta - \frac{v^2}{2}\right)^2 + 2v^2 B_u}} \left(\frac{2\sqrt{A_u V(0)V(t)}}{v^2\sinh\frac{\sqrt{A_u}t}{2}} \right)}{I_{\frac{2k\theta}{v^2}-1}\left(\frac{2k\sqrt{V(0)V(t)}}{v^2\sinh\frac{kt}{2}} \right)}, \end{aligned} \tag{EC.7}$$

with

$$\begin{aligned} A_u &= k^2 - 2v^2\left[u\left(\frac{a\rho k}{v} - \frac{1}{2}a^2 \right) + \frac{1}{2}u^2\left(1-\rho^2\right)a^2 \right], \\ B_u &= u\left[\frac{b\rho}{v}\left(\frac{v^2}{2} - k\theta \right) - \frac{b^2}{2} \right] + \frac{1}{2}u^2\left(1-\rho^2\right)b^2. \end{aligned}$$

The previous result reduces to the one for the 3/2 model by setting $a=0$.

EC.3. Model conditional representations and results

In addition to the one-factor Heston model presented in Section 4.3, we provide, next, examples of other models which admit relevant representations.

EC.3.1. Double Heston model

Conditional on $(\Psi_{k-1,k},\Phi_k)=\left(\left(\int_{t_{k-1}}^{t_k}V_1(s)ds,\int_{t_{k-1}}^{t_k}V_2(s)ds\right),(V_1(t_k),V_2(t_k))\right)$,

$$s_k=\alpha(s_{k-1},\phi_{1,k-1},\phi_{2,k-1})+\sum_{j=1}^{2}\left(\beta_j\Phi_{j,k}+\gamma_j\Psi_{j,k-1,k}+\sqrt{\delta_j\Psi_{j,k-1,k}}Y_{j,k}\right),\quad Y_{j,k}\sim\mathcal{N}(0,1),$$

where $Y_{j,k}$ is independent of $\Phi_{j,k}$ and $\Psi_{j,k-1,k}$, and $Y_{1,k}$ is independent of $Y_{2,k}$. From Table 1 in the paper, we have that $\alpha(s_{k-1},\phi_{1,k-1},\phi_{2,k-1})=s_{k-1}+r\,(t_k-t_{k-1})-\sum_{j=1}^{2}\beta_j\,(\phi_{j,k-1}+k_j\theta_j(t_k-t_{k-1}))$, $\beta_j=\rho_j/v_j$, $\gamma_j=\rho_j k_j/v_j-1/2$, $\delta_j=1-\rho_j^2$. Then,

$$\begin{aligned}&G_{s|\Phi}(\,s|\,\phi_1,\phi_2;s_{k-1},\phi_{1,k-1},\phi_{2,k-1})\\&=\int\int G_{s|\Psi,\Phi}(\,s|\,x_1,x_2,\phi_1,\phi_2;s_{k-1},\phi_{1,k-1},\phi_{2,k-1})g_{\Psi|\Phi}(\,x_1,x_2|\,\phi_1,\phi_2;\phi_{1,k-1},\phi_{2,k-1})dx_1dx_2\end{aligned}$$

is computed over the first quadrant, where

$$G_{s|\Psi,\Phi}(\cdot|\,x_1,x_2,\phi;s_{k-1},\phi_{1,k-1},\phi_{2,k-1})=\frac{\int_{-\infty}^{\cdot}\exp\left(-\frac{(s-\alpha(s_{k-1},\phi_{1,k-1},\phi_{2,k-1})-\sum_{j=1}^{2}(\beta_j\phi_j+\gamma_jx_j))^2}{2\sum_{j=1}^{2}\delta_jx_j}\right)ds}{\sqrt{2\pi\sum_{j=1}^{2}\delta_jx_j}}$$

and, by independence,

$$g_{\Psi|\Phi}(\,x_1,x_2|\,\phi_1,\phi_2;\phi_{1,k-1},\phi_{2,k-1})=g_{\Psi_1|\Phi_1}(\,x_1|\,\phi_1;\phi_{1,k-1})g_{\Psi_2|\Phi_2}(\,x_2|\,\phi_2;\phi_{2,k-1}).$$

The associated characteristic function is given by

$$\varphi_{G_{s|\Phi}}(\,u|\,\phi;s_{k-1},\phi_{k-1})=e^{iu\left(\alpha\left(s_{k-1},\phi_{1,k-1},\phi_{2,k-1}\right)+\beta_1\phi_1+\beta_2\phi_2\right)}\prod_{j=1}^{2}\varphi_{G_{\Psi_j|\Phi_j}}\left(u\left(\gamma_j+\frac{i\delta_ju}{2}\right)\middle|\,\phi_j;\phi_{j,k-1}\right),$$

where $\varphi_{G_{\Psi_j|\Phi_j}}$ is given by (EC.3).

It is possible to introduce an approximating law for Ψ, such as a Pearson distribution, and denote the approximate random variable by $\tilde{\Psi}$. The previous results imply that two univariate approximating Pearson densities are required, one for each of $(\Psi_1|\Phi_1)$ and $(\Psi_2|\Phi_2)$. Also, in (14) of Theorem 1, an integration over $\mathcal{I}_\Phi$, which is equal to the first quadrant, is now required.

EC.3.2. $4/2$ model

In this model, conditional on $(\Psi_{k-1,k}, \Phi_k) = \left(\left(\int_{t_{k-1}}^{t_k} V(s)ds, \int_{t_{k-1}}^{t_k} V^{-1}(s)ds \right), V(t_k) \right)$,

$$\begin{aligned} s_k = {} & \alpha(s_{k-1}, \phi_{k-1}) + \beta_1 \Phi_k + \beta_2 \ln \Phi_k + \gamma_1 \Psi_{1,k-1,k} + \gamma_2 \Psi_{2,k-1,k} \\ & + \sqrt{\delta_0 + \delta_1 \Psi_{1,k-1,k} + \delta_2 \Psi_{2,k-1,k}} Y_k, \quad Y_k \sim \mathcal{N}(0,1), \end{aligned}$$

where Y is independent of Φ and Ψ, $\alpha(s_{k-1}, \phi_{k-1}) = s_{k-1} + (r - ab - \beta_1 k\theta + \beta_2 k)(t_k - t_{k-1}) - \beta_1 \phi_{k-1} - \beta_2 \ln \phi_{k-1}$, $\gamma_1 = \beta_1 k - a^2/2$, $\gamma_2 = \beta_2 (v^2/2 - k\theta) - b^2/2$, $\beta_1 = a\rho/v$, $\beta_2 = b\rho/v$, $\delta_0 = 2ab(1 - \rho^2)(t_k - t_{k-1})$, $\delta_1 = a^2(1-\rho^2)$, $\delta_2 = b^2(1-\rho^2)$, and

$$G_{s|\Phi}(s|\,\phi; s_{k-1}, \phi_{k-1}) = \int\int G_{s|\Psi,\Phi}(s|\,x_1, x_2, \phi; s_{k-1}, \phi_{k-1}) g_{\Psi|\Phi}(x_1, x_2|\,\phi; \phi_{k-1}) dx_1 dx_2 \tag{EC.8}$$

is computed over the first quadrant, with

$$G_{s|\Psi,\Phi}(\cdot|\,x_1, x_2, \phi; s_{k-1}, \phi_{k-1}) = \frac{\int_{-\infty}^{\cdot} \exp\left(-\frac{(s - \alpha(s_{k-1}, \phi_{k-1}) - \beta_1 \phi - \beta_2 \ln \phi - \gamma_1 x_1 - \gamma_2 x_2)^2}{2(\delta_0 + \delta_1 x_1 + \delta_2 x_2)} \right) ds}{\sqrt{2\pi (\delta_0 + \delta_1 x_1 + \delta_2 x_2)}}.$$

The associated characteristic function is given by

$$\begin{aligned} \varphi_{G_{s|\Phi}}(u|\,\phi; s_{k-1}, \phi_{k-1}) = {} & e^{iu\left(\alpha(s_{k-1}, \phi_{k-1}) + \beta_1 \phi + \beta_2 \ln \phi + \frac{iu\delta_0}{2}\right)} \times \\ & \varphi_{G_{\Psi|\Phi}}\left(u\left(\gamma_1 + \frac{i\delta_1 u}{2}\right), u\left(\gamma_2 + \frac{i\delta_2 u}{2}\right) \middle| \, \phi; \phi_{k-1} \right), \end{aligned}$$

where $\varphi_{G_{\Psi|\Phi}}$ is given by (EC.6).

In this case, the bivariate density function $g_{\Psi|\Phi}(x_1, x_2|\,\phi; \phi_{k-1})$ in (EC.8) can be approximated by fitting a bivariate Pearson distribution (see e-companion Section EC.6). Alternatively, given (EC.7), it is possible to approximate directly $G_{s|\Phi}(s|\,\phi; s_{k-1}, \phi_{k-1})$ by fitting a univariate Pearson distribution. The $3/2$ model subcase follows by dropping the first component of Ψ.

EC.3.3. SABR model

By dint of how the distributional properties of this model are usually presented (see Section 3), we study this on the original scale, that is, in terms of S rather than $s = \ln S$, implying also minor adjustment of Theorem 1, in addition to consideration of the probability mass at $S = 0$.

Conditional on $(\Psi_{k-1,k}, \Phi_k) = \left(\left(\int_{t_{k-1}}^{t_k} V^2(s)ds \right)^{-1}, V(t_k) \right)$, we have from (9)–(12) and the inverse transform method that

$$\begin{cases} S_k = 0, & \text{if } U_k \le 1 - Q_{\chi^2} \\ S_k = G^{-1}_{S|\Psi,\Phi}(U_k|\Psi_{k-1,k}, \Phi_k; S_{k-1}, \Phi_{k-1}), & \text{if } U_k \ge 1 - Q_{\chi^2} \end{cases}, \tag{EC.9}$$

where $U_k \sim \text{Unif}(0,1)$ and $G^{-1}_{S|\Psi,\Phi}$ is the inverse of $G_{S|\Psi,\Phi}$ given exactly by (10) or approximately by (12) when $\rho = 0$ or $\rho \neq 0$, respectively. For $y > 0$,

$$G_{S|\Phi}(y|\,\phi; S_{k-1}, \phi_{k-1}) = \int Q(S_k \le y | S_{k-1}, \phi_{k-1}, \phi, x) g_{\Psi|\Phi}(x|\,\phi; \phi_{k-1}) dx$$

is computed over the positive half-line, and from (EC.9) we get that

$$\begin{aligned} Q(S_k \leq y | S_{k-1}, \phi_{k-1}, \phi, x) &= Q(U_k \leq G_{S|\Psi,\Phi}(y) | S_{k-1}, \phi_{k-1}, \phi, x) \\ &= G_{S|\Psi,\Phi}(y|\, x, \phi; S_{k-1}, \phi_{k-1}). \end{aligned}$$

The associated characteristic function is given by

$$\varphi_{G_{S|\Phi}}(u|\, \phi; S_{k-1}, \phi_{k-1}) = \int\int e^{iuy} g_{S|\Psi,\Phi}(y|\, x, \phi; S_{k-1}, \phi_{k-1}) g_{\Psi|\Phi}(x|\, \phi; \phi_{k-1}) dy dx,$$

where $g_{S|\Psi,\Phi}$ follows from differentiating (10) or (12). We can approximate $g_{\Psi|\Phi}(x|\, \phi; \phi_{k-1})$ by fitting a univariate Pearson distribution.

EC.3.4. OU-SV model

According to this model, conditional on $(\Psi_{k-1,k}, \Phi_k) = \left(\int_{t_{k-1}}^{t_k} V^2(s)ds, \left(V(t_k), \int_{t_{k-1}}^{t_k} V(s)ds\right)\right)$,

$$s_k = \alpha(s_{k-1}, \phi_{1,k-1}) + \beta_1 \Phi_{1,k}^2 + \beta_2 \Phi_{2,k-1,k} + \gamma \Psi_{k-1,k} + \sqrt{\delta \Psi_{k-1,k}} Y_k, \quad Y_k \sim \mathcal{N}(0,1),$$

where Y is independent of Φ and Ψ, $\alpha(s_{k-1}, \phi_{1,k-1}) = s_{k-1} + (r - \rho v/2)(t_k - t_{k-1}) - \beta_1 \phi_{1,k-1}^2$, $\beta_1 = \rho/(2v)$, $\beta_2 = -\rho k\theta/v$, $\gamma = \rho k/v - 1/2$, $\delta = 1 - \rho^2$, and

$$G_{s|\Phi}(s|\, \phi_1, \phi_2; s_{k-1}, \phi_{1,k-1}) = \int G_{s|\Psi,\Phi}(s|\, x, \phi_1, \phi_2; s_{k-1}, \phi_{1,k-1}) g_{\Psi|\Phi}(x|\, \phi_1, \phi_2; \phi_{1,k-1}) dx$$

is computed over $\mathbb{R}^+$, where

$$G_{s|\Psi,\Phi}(\cdot|\, x, \phi_1, \phi_2; s_{k-1}, \phi_{1,k-1}) = \frac{\int_{-\infty}^{\cdot} \exp\left(-\frac{(s - \alpha\left(s_{k-1}, \phi_{1,k-1}\right) - \beta_1 \phi_1^2 - \beta_2 \phi_2 - \gamma x)^2}{2\delta x}\right) ds}{\sqrt{2\pi\delta x}}.$$

The associated characteristic function is given by

$$\varphi_{G_{s|\Phi}}(u|\, \phi_1, \phi_2; s_{k-1}, \phi_{1,k-1}) = e^{iu\left(\alpha\left(s_{k-1}, \phi_{1,k-1}\right) + \beta_1 \phi_1^2 + \beta_2 \phi_2\right)} \varphi_{G_{\Psi|\Phi}}\left(u\left(\gamma + \frac{i\delta u}{2}\right) \middle| \phi_1, \phi_2; \phi_{1,k-1}\right),$$

where $\varphi_{G_{\Psi|\Phi}}$ is given by (EC.5). In this case, Φ has a bivariate normal distribution, therefore in (14) of Theorem 1 an integration over $\mathcal{I}_\Phi = \mathbb{R}^2$ is required. For $g_{\Psi|\Phi}(x|\, \phi_1, \phi_2; \phi_{1,k-1})$, we can use an approximating univariate Pearson density.

EC.3.5. Bates model

The main difference with Heston is the inclusion of a random jump component that is independent of the volatility process. The jump size J distribution is normal with mean $\mu_J = \ln(1+\omega) - \sigma_J^2/2$ and variance σ_J^2. Conditional on (Ψ, Φ) and the number of Poisson jumps N arriving subject to constant intensity λ,

$$s_k = \alpha\left(s_{k-1}, \phi_{k-1}\right) + \beta \Phi_k + \gamma \Psi_{k-1,k} + \sqrt{\delta \Psi_{k-1,k}} Y_k + \sum_{i=0}^{N_k} J_i, \quad Y_k \sim \mathcal{N}(0,1),$$

where Y, (Ψ,Φ), N and $\{J_i\}_i$ are mutually independent, $\alpha(s_{k-1},\phi_{k-1}) = s_{k-1} + (r - \beta k\theta - \lambda\omega)(t_k - t_{k-1}) - \beta\phi_{k-1}$, $\gamma = \beta k - 1/2$, $\beta = \rho/v$, $\delta = 1-\rho^2$, and

$$G_{s|\Phi}(s|\,\phi; s_{k-1},\phi_{k-1}) = \int G_{s|\Psi,\Phi}(s|\,x,\phi; s_{k-1},\phi_{k-1}) g_{\Psi|\Phi}(x|\,\phi;\phi_{k-1})dx,$$

where

$$G_{s|\Psi,\Phi}(\cdot|\,x,\phi; s_{k-1},\phi_{k-1}) = \sum_{i=0}^{\infty} \frac{\int_{-\infty}^{\cdot} \exp\left(-\frac{(s-\alpha\left(s_{k-1},\phi_{k-1}\right)-\beta\phi-\gamma x-i\mu_J)^2}{2(\delta x+i\sigma_J^2)}\right) ds}{\sqrt{2\pi(\delta x+i\sigma_J^2)}} \frac{\lambda^i (t_k - t_{k-1})^i e^{-\lambda(t_k-t_{k-1})}}{i!}.$$

The associated characteristic function is given by

$$\varphi_{G_{s|\Phi}}(u|\,\phi; s_{k-1},\phi_{k-1}) = e^{iu\left(\alpha\left(s_{k-1},\phi_{k-1}\right)+\beta\phi\right)+\lambda(t_k-t_{k-1})\exp\left(i\mu_J u-\sigma_J^2 u^2/2\right)} \varphi_{G_{\Psi|\Phi}}\left(u\left(\gamma+\frac{i\delta u}{2}\right)\middle|\,\phi;\phi_{k-1}\right),$$

where $\varphi_{G_{\Psi|\Phi}}$ is given by (EC.3).

For the case of jump sizes with an asymmetric double exponential distribution, interested readers may refer to Kou (2002, Theorem B.1). Finally, for $g_{\Psi|\Phi}(x|\,\phi;\phi_{k-1})$ we can use an approximating univariate Pearson density the same way as in the Heston model.

EC.4. Closeness of distributions

EC.4.1. Characteristic function-based approach

Related to the distance between distribution functions is Esseen's inequality recorded next. Consider some non-negative function $H(x)$ and $h(u) := \int_{\mathbb{R}} e^{iux} H(x)\,dx$ satisfying

$$\begin{cases} \int_{\mathbb{R}} H(x)\,dx = 1;\ b = \int_{\mathbb{R}} |x|\,H(x)\,dx < \infty \\ h(0) = 1;\ h(u) = 0 \text{ for } |u| \geq 1;\ 0 \leq |h(u)| \leq 1 \text{ for } |u| \leq 1 \end{cases}. \tag{EC.10}$$

A possible choice for H, as suggested by Esseen (1945), is $H(x) = (3/(8\pi))\left(4\sin(x/4)/x\right)^4$. Then, for the Pearson distribution function $\tilde{G}(x)$ which is differentiable almost everywhere, has bounded variation and finite ϱ (see Lemma 1 in the paper), Theorem 2a from Esseen (1945) applies. More specifically, to every arbitrary $k > 1$, there corresponds $0 < \bar{m}(k) < 1$ and $\alpha(k)$ so that

$$(2 - \bar{m}(k)) \int_{-\bar{m}(k)\alpha(k)}^{(1-\bar{m}(k))\alpha(k)} H(x)\,dx = \frac{k+1}{k}, \tag{EC.11}$$

and

$$\left|G(x) - \tilde{G}(x)\right| \leq \max\left\{\frac{k\epsilon}{2\pi} + \frac{kb\varrho}{\mathcal{T}}, \frac{\alpha(k)\,\varrho}{\mathcal{T}}\right\} \leq \frac{k\epsilon}{2\pi} + \frac{(kb+\alpha(k))\,\varrho}{\mathcal{T}}$$

with $(\mathcal{T},\epsilon)$ satisfying

$$\int_{-\mathcal{T}}^{\mathcal{T}} \left|\frac{\varphi_G(u) - \varphi_{\tilde{G}}(u)}{u}\right| du = \epsilon$$

and φ_G and $\varphi_{\tilde{G}}$ denoting the indicated characteristic functions. We can always choose $\bar{m}(k)$ sufficiently small and $\alpha(k)$ sufficiently large so that (EC.11) is fulfilled.

EC.4.2. Moment-based distance

Another possibility is the moment-based distance of Akhiezer (1965, Corollary 2.5.4), revisited later by Lindsay and Basak (2000, Theorems 1, 2).

THEOREM EC.1 **(Lindsay and Basak 2000)**. *Let any two arbitrary distributions $G(x)$ and $\tilde{G}(x)$ share their first $2m$ moments. Then, for all x,*

$$\left|G(x)-\tilde{G}(x)\right| \leq \left\{P_m'(x)\, W_m^{-1} P_m(x)\right\}^{-1}, \tag{EC.12}$$

where $P_m(x) := (1, x, x^2, \ldots, x^m)'$ and $W_m := \|\mu_{i+j}\|_{i,j=0}^{m}$ is a Hankel symmetric matrix defined by the first $2m$ moments.

The right-hand side of (EC.12) goes to 0 at the rate x^{-2m} as $x \to \infty$ giving relatively sharp tail information. An improvement to (EC.12) is due to Khamis (1954) who introduces a constant non-negative multiplier, where it exists, that is smaller than the unity.

EC.4.3. Entropy bound

For a generic continuous distribution function H with associated density function h, we define the differential entropy $\mathcal{H}[h] = -\int h(x) \ln h(x)\, dx$. Along with our original approximating Pearson distribution function $\tilde{G}$ with associated density $\tilde{g}$, we consider the entropy-maximizing distribution $\hat{G}$ with density function

$$\hat{g}(x) := \exp\left(-\sum_{n=0}^{m} \lambda_n x^n\right), \tag{EC.13}$$

where $\{\lambda_n\}$ are the Lagrange multipliers (Kapur and Kesavan 1992). This is obtained by maximizing the entropy constrained by moments. The resulting distribution $\hat{G}$ shares the same first m moments with the target true distribution G with density g.

Our aim is to derive an entropy bound to the absolute difference of each of $H \in \left\{\hat{G}, \tilde{G}\right\}$ with respect to the true G. To this end, we consider two measures of difference between density functions, namely, the divergence and variation measures

$$\mathcal{I}[g,h] := \int g(x) \ln \frac{g(x)}{h(x)} dx \quad \text{and} \quad \mathcal{V}[g,h] := \int |h(x) - g(x)|\, dx,$$

respectively. Next, we present a lower bound for the divergence measure (although more complicated ones exist in the literature).

LEMMA EC.1 **(Kullback 1967)**. *We have that*

$$\mathcal{I} \geq \frac{\mathcal{V}^2}{2} + \frac{\mathcal{V}^4}{36}. \tag{EC.14}$$

Based on the foregoing, we can state the following result.

THEOREM EC.2. *For the distributions G and $H \in \left\{\hat{G}, \tilde{G}\right\}$,*

$$|H(x) - G(x)| \leq 3\sqrt{-1 + \sqrt{1 + \frac{4\mathcal{I}[g,h]}{9}}}. \tag{EC.15}$$

In addition, $\mathcal{I}[g,\hat{g}] = \sum_{n=0}^{m} \lambda_n \mu_n(G) - \mathcal{H}[g]$.

Proof. We have that

$$|H(x) - G(x)| \leq \int |h(x) - g(x)|\, dx = \mathcal{V}[g,h].$$

Then, inequality (EC.14) solves as

$$0 \leq \mathcal{V}^2[g,h] \leq 3\sqrt{9 + 4\mathcal{I}[g,h]} - 9 = 9\left(\sqrt{1 + \frac{4}{9}\mathcal{I}[g,h]} - 1\right) \Rightarrow \mathcal{V}[g,h] \leq 3\sqrt{-1 + \sqrt{1 + \frac{4\mathcal{I}[g,h]}{9}}},$$

from which (EC.15) follows. In addition, for $H \equiv \hat{G}$, we have

$$\mathcal{I}[g,\hat{g}] = \int g(x) \ln \frac{g(x)}{\hat{g}(x)} dx = -\mathcal{H}[g] - \int g(x) \ln \hat{g}(x)\, dx = -\mathcal{H}[g] + \sum_{n=0}^{m} \lambda_n \int x^n \hat{g}(x)\, dx,$$

where the last equality is due to (EC.13) and from which the result follows.

EC.5. Sampling from the Pearson family of distributions

In what follows, we present the different Pearson distribution types and corresponding generators of random numbers $\bar{\mathcal{Y}}$ with zero mean, variance $\beta = 1$, skewness $\sqrt{\gamma}$ and kurtosis ε. To this end, we consider first a few quantities that will be used next:

$$z := -\frac{\sqrt{\gamma}(\varepsilon+3) + \operatorname{sgn}(\sqrt{\gamma}(\varepsilon+3))\sqrt{\gamma^{1/2}(\varepsilon+3)^2 - 4(4\varepsilon-3\gamma)(2\varepsilon-3\gamma-6)}}{2}, \quad a_1 := \frac{z}{2\varepsilon-3\gamma-6},$$
$$a_2 := \frac{4\varepsilon-3\gamma}{z}, \quad c_0 := \frac{4\varepsilon-3\gamma}{10\varepsilon-12\gamma-18}, \quad c_1 := \frac{\sqrt{\gamma}(\varepsilon+3)}{10\varepsilon-12\gamma-18}, \quad c_2 := \frac{2\varepsilon-3\gamma-6}{10\varepsilon-12\gamma-18},$$

where $\operatorname{sgn}(x) := x/|x|$ for $x \neq 0$ and $\operatorname{sgn}(0) := 0$. Then:

Type I (four-parameter beta)

$$\bar{\mathcal{Y}} \equiv a_1 + (a_2 - a_1)\mathcal{B}\left(\frac{c_1 + a_1}{c_2(a_2 - a_1)} + 1, -\frac{c_1 + a_2}{c_2(a_2 - a_1)} + 1\right),$$

where $\mathcal{B}(a,b)$ denotes a random number generator from a beta distribution with parameters a and b (see Devroye 1986);

Type II (symmetric four-parameter beta)

$$\bar{\mathcal{Y}} \equiv a_1 + 2|a_1|\mathcal{B}\left(\frac{c_1 + a_1}{2c_2|a_1|} + 1, \frac{c_1 + a_1}{2c_2|a_1|} + 1\right);$$

Type III (three-parameter gamma)

$$\bar{\mathcal{Y}} \equiv c_1 \Gamma\left(\frac{\frac{c_0}{c_1} - c_1}{c_1} + 1, 1\right) + a_1,$$

where $\Gamma(a, b)$ denotes a random number generator from a gamma distribution with shape parameter a and scale parameter b (see Marsaglia and Tsang 2000);

Type IV (density proportional to $\exp\left(-\nu \arctan(\frac{x-\lambda}{a})\right) / \left(1 + (\frac{x-\lambda}{a})^2\right)^m$)

$$\bar{\mathcal{Y}} \equiv \mathcal{P}_{IV}(m, \nu, a, \lambda),$$

where $\mathcal{P}_{IV}(m, \nu, a, \lambda)$ is a random number generator from a Pearson Type IV based on the exponential rejection method for log-concave densities of Devroye (1986, Section 7.2) and adapted to this case by Heinrich (2004), with

$$m := \frac{1}{2c_2}, \quad \nu := \frac{2c_1(1-m)}{\sqrt{4c_0c_2 - c_1^2}}, \quad b := 2(m-1), \quad a := \sqrt{\frac{b^2(b-1)}{b^2 + \nu^2}}, \quad \lambda := \frac{a\nu}{b};$$

Type V (inverse gamma location-scale)

$$\bar{\mathcal{Y}} \equiv -\frac{c_1 - \frac{c_1}{2c_2}}{c_2 \Gamma(\frac{1}{c_2} - 1, 1)} - \frac{c_1}{2c_2};$$

Type VI (F location-scale)

$$\begin{aligned}\bar{\mathcal{Y}} \equiv &\left(a_2 + \frac{2(m_2+1)(a_2-a_1)}{-2(m_1+m_2+1)} \mathcal{F}(2(m_2+1), -2(m_1+m_2+1))\right) \mathbf{1}_{\{a_2<0\}} \\ &+ \left(a_2 + \frac{2(m_1+1)(a_2-a_1)}{-2(m_1+m_2+1)} \mathcal{F}(2(m_1+1), -2(m_1+m_2+1))\right) \mathbf{1}_{\{a_2\geq 0\}},\end{aligned}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator of the event $\{\cdot\}$,

$$m_1 := \frac{a_1 + c_1}{c_2(a_2 - a_1)}, \quad m_2 := -\frac{a_2 + c_1}{c_2(a_2 - a_1)}$$

and $\mathcal{F}(a, b)$ is a random number generator from a Snedecor F distribution with numerator and denominator degrees of freedom a and b respectively (see Devroye 1986); and

Type VII (t location-scale)

$$\bar{\mathcal{Y}} \equiv \sqrt{\frac{c_0}{1-c_2}} \mathcal{T}\left(\frac{1}{c_2} - 1\right),$$

where $\mathcal{T}(a)$ is a random number generator from a Student's t distribution with a degrees of freedom (see Devroye 1986).

EC.6. Sampling from $\left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)}\right)$ in the $4/2$ model using a bivariate Pearson distribution approach

In this section, we focus on the special case of sampling from the pair $\left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)}\middle| V(t)\right) = (\Psi(u,t)|\Phi(t))$ in the 4/2 model (see also Table 1 of the paper).

Parrish (1987, 1990) propose a conditional nested factorization approach to simulating from a multivariate Pearson distribution; for our purposes, we consider here the bivariate case and the pair $\left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)}\right)$ in the 4/2 model. To this end, define the cross-moments

$$\mu_{r_1,r_2} = E\left[\left(\int_u^t V(s)ds\right)^{r_1}\left(\int_u^t \frac{ds}{V(s)}\right)^{r_2}\middle| V(t)\right] = \left.\frac{\partial^{r_1+r_2}\mathcal{L}(-a,-b)}{\partial a^{r_1}\partial b^{r_2}}\right|_{a=b=0},$$

where $r_1, r_2 = \{0,1,2,3,4\}$, $r_1 + r_2 \leq 4$ and

$$\mathcal{L}(a,b) = E\left[e^{-a\int_u^t V(s)ds - b\int_u^t \frac{ds}{V(s)}}\middle| V(t)\right]$$

is given by (EC.6). The simulation is then summarized in the following steps:

1. Simulate $(V(t)|V(u))$ based on $V(t) \overset{\text{(law)}}{=} \chi'^2_d(\lambda)\, v^2(1-e^{-k(t-u)})/4k$, where $\chi'^2_d(\lambda)$ is the non-central chi-squared random variable with $d = 4\theta k/v^2$ degrees of freedom and noncentrality parameter $\lambda = 4kv^{-2}e^{-k(t-u)}V(u)/(1-e^{-k(t-u)})$
2. Simulate $\left(\int_u^t \frac{ds}{V(s)}\middle| V(t)\right)$ having first fitted a Pearson curve by moments $\mu_{0,1}, \mu_{0,2}, \mu_{0,3}, \mu_{0,4}$
3. Compute $\mu_{r_1} := E\left[\left(\int_u^t V(s)ds\right)^{r_1}\middle| \int_u^t \frac{ds}{V(s)}, V(t)\right]$ via equations (EC.16)–(EC.17)
4. Simulate $\left(\int_u^t V(s)ds\middle| \int_u^t \frac{ds}{V(s)}, V(t)\right)$ which follows a Pearson distribution law, if $\left(\int_u^t V(s)ds, \int_u^t \frac{ds}{V(s)}\middle| V(t)\right)$ has a bivariate Pearson distribution (see Parrish 1987 for a proof of this).

Focusing on step 2, we have from Parrish (1990, equations 4–5) that

$$\mu_1 := \frac{b_1^\star - a_0^\star}{1-2b_{11}^\star}, \quad \mu_2 := \frac{(2b_1^\star - a_0^\star)\,\mu_1 + b_0^\star}{1-3b_{11}^\star}, \tag{EC.16}$$

$$\mu_3 := \frac{(3b_1^\star - a_0^\star)\,\mu_2 + 2b_0^\star\mu_1}{1-4b_{11}^\star}, \quad \mu_4 := \frac{(4b_1^\star - a_0^\star)\,\mu_3 + 3b_0^\star\mu_2}{1-5b_{11}^\star} \tag{EC.17}$$

for

$$a_0^\star := a_0 + a_1x_1, \quad b_0^\star := -\left(b_0 + b_1x_1 + b_{11}x_1^2\right),$$

$$b_1^\star := -\left(b_2 + b_{12}x_1\right), \quad b_{11}^\star := -b_{22}$$

with x_1 a random sample from $\left(\int_u^t \frac{ds}{V(s)} \middle| V(t)\right)$ and $a_0, a_1, b_0, b_1, b_2, b_{11}, b_{12}, b_{22}$ satisfying the system of linear equations (see Parrish 1987, Figure 3)

$$\begin{bmatrix} \mu_{0,0} & \mu_{1,0} & 0 & 0 & \mu_{0,0} & 0 & \mu_{1,0} & 2\mu_{0,1} \\ \mu_{1,0} & \mu_{2,0} & 0 & 0 & \mu_{1,0} & 0 & \mu_{2,0} & 2\mu_{1,1} \\ \mu_{0,1} & \mu_{1,1} & \mu_{0,0} & \mu_{1,0} & 2\mu_{0,1} & \mu_{2,0} & 2\mu_{1,1} & 3\mu_{0,2} \\ \mu_{2,0} & \mu_{3,0} & 0 & 0 & \mu_{2,0} & 0 & \mu_{3,0} & 2\mu_{2,1} \\ \mu_{1,1} & \mu_{2,1} & \mu_{1,0} & \mu_{2,0} & 2\mu_{1,1} & \mu_{3,0} & 2\mu_{2,1} & 3\mu_{1,2} \\ \mu_{0,2} & \mu_{1,2} & 2\mu_{0,1} & 2\mu_{1,1} & 3\mu_{0,2} & 2\mu_{2,1} & 3\mu_{1,2} & 4\mu_{0,3} \\ \mu_{3,0} & \mu_{4,0} & 0 & 0 & \mu_{3,0} & 0 & \mu_{4,0} & 2\mu_{3,1} \\ \mu_{2,1} & \mu_{3,1} & \mu_{2,0} & \mu_{3,0} & 2\mu_{2,1} & \mu_{4,0} & 2\mu_{3,1} & 3\mu_{2,2} \\ \mu_{1,2} & \mu_{2,2} & 2\mu_{1,1} & 2\mu_{2,1} & 3\mu_{1,2} & 2\mu_{3,1} & 3\mu_{2,2} & 4\mu_{1,3} \\ \mu_{0,3} & \mu_{1,3} & 3\mu_{0,2} & 3\mu_{1,2} & 4\mu_{0,3} & 3\mu_{2,2} & 4\mu_{1,3} & 5\mu_{0,4} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ b_0 \\ b_1 \\ b_2 \\ b_{11} \\ b_{12} \\ b_{22} \end{bmatrix} = - \begin{bmatrix} \mu_{0,1} \\ \mu_{1,1} \\ \mu_{0,2} \\ \mu_{2,1} \\ \mu_{1,2} \\ \mu_{0,3} \\ \mu_{3,1} \\ \mu_{2,2} \\ \mu_{1,3} \\ \mu_{0,4} \end{bmatrix}.$$

EC.7. More experimental results

EC.7.1. Supplementary results on analysis of error and computing time

Figure EC.1 Computing times (in seconds) corresponding to each implementation step of Algorithm 2 as described in Section 6.1 of the paper

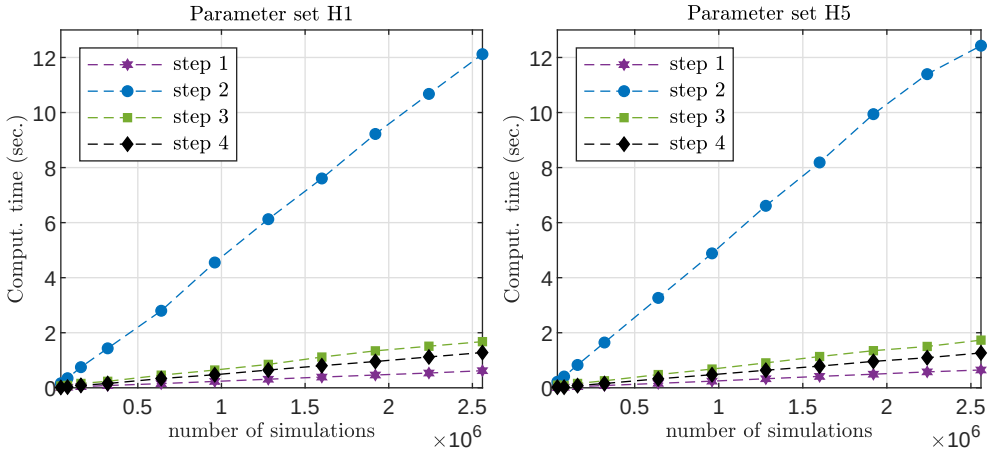


Table EC.1 Examples of differences between n-th analytical moments and moments of $\left(\int_0^T V(s)ds \middle| V(T)\right)$ computed using Algorithm 1 conditional on the 25, 50, 75th percentiles of the terminal variance

	H1			H5		
n	$V(T)=0.003$	$V(T)=0.010$	$V(T)=0.026$	$V(T)=0.001$	$V(T)=0.008$	$V(T)=0.081$
1	3.20e-11	-5.10e-11	2.27e-11	2.16e-11	3.68e-11	3.41e-12
2	2.41e-15	-2.00e-14	-1.30e-14	1.32e-13	-1.40e-13	-2.60e-13
3	6.18e-16	-5.20e-17	-7.30e-16	-2.80e-14	8.77e-14	-1.10e-14
4	-1.10e-17	-2.10e-17	1.54e-17	1.21e-14	3.15e-14	5.41e-15

Figure EC.2 Computing total times (in seconds) for the first four moments of the conditional integrated variance in the Heston model

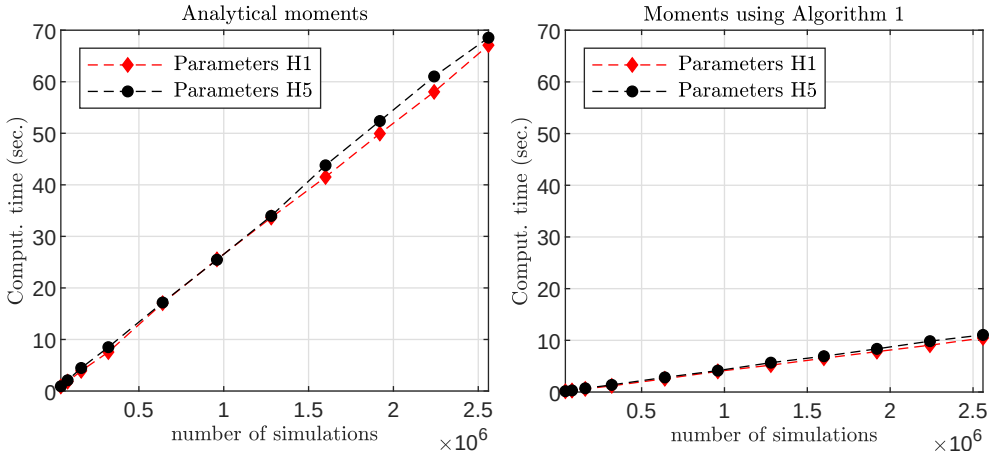
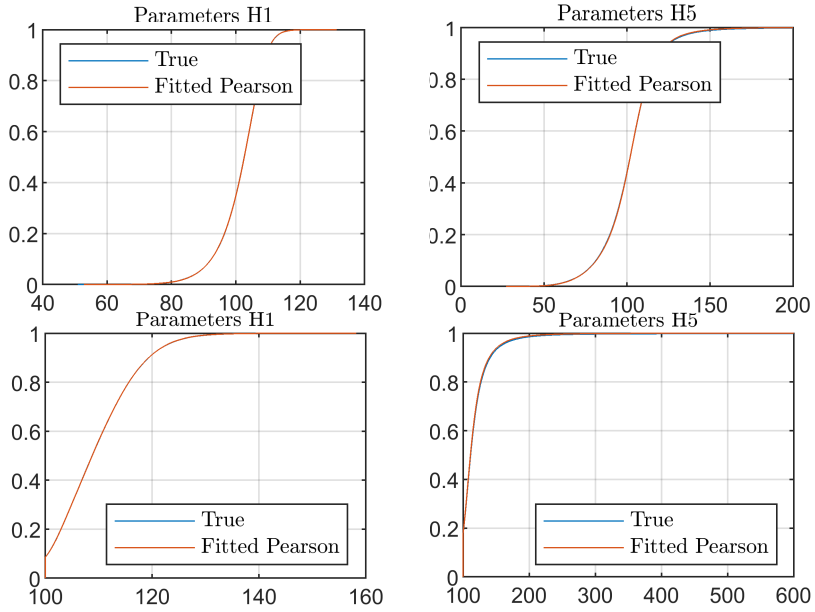


Figure EC.3 Cumulative distribution functions of $A_N = \sum_{i=0}^{N} S_i/(N+1)$ (top plots) and $M_{0,N} = \max_{0\le i\le N} S_i$ (bottom plots) in the Heston model (parameter sets H1 and H5) based on true distribution and Pearson approximation of the integrated variance conditional on the 50th percentile of the terminal variance



Notes. Top plots: largest absolute errors are 0.0015 (parameter set H1), 0.0044 (parameter set H5). Bottom plots: largest absolute errors are 0.0015 (parameter set H1), 0.0068 (parameter set H5).

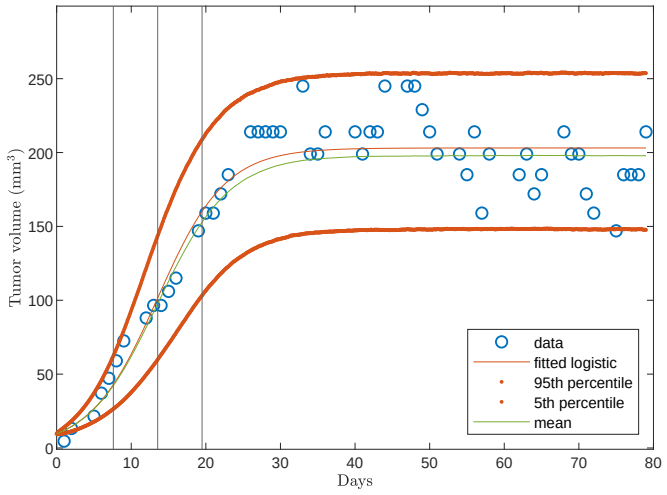
EC.7.2. The logistic model for tumor growth: a simulation case study

Cancer cells have an inflated rate of proliferation leading to a rapid tumor growth. A survey of the relevant literature reveals that, generally, small tumors grow exponentially, however the growth

slows down when they get larger. This decelerated growth typically results in a diameter (if a solid tumor) that remains constant in time. Therefore, a model that is able to accommodate these features can be a sensible choice (e.g., see Laird 1964, and later research in the field). Note that, while growth curves are relevant for modelling the untreated tumor, they cannot be applied to the treated tumor as it may decrease and regrow after a treatment such as radiation (e.g., see Demidenko 2013, Chapter 10); studying post-treatment tumor is beyond the scope of this exercise.

Here, we adopt the stochastic logistic model (42) for the growth of tumors, where we denote by $X(t)$ the tumor volume at time t. For the purposes of our illustration, we calibrate the model to growth data of multicellular tumor spheroids that include three-dimensional aggregates of cancer cells, which have been grown under controlled experimental conditions, as described in Chignola et al. (1999)and made available by Demidenko (2013). The nonlinear least-squares parameter estimates of the model are $\hat{\lambda} = 0.00109$, $\hat{K} = 203.10135$ and $\hat{X}(0) = 9.63691$; in addition, the estimated daily volatility is $\hat{\sigma} = 10.6539\%$.

Figure EC.4 Stochastic logistic model for tumor volume



Notes. Historical daily tumor growth; fitted deterministic Verhulst equation; probabilistic forecasts for tumor growth generated by model (42) in the paper: 95% and 5% confidence bands (upper and lower lines) based on 10^5 simulated trajectories; mean value estimate (medium line).

Figure EC.4 exhibits the historical daily tumor growth and the fitted deterministic Verhulst equation (red solid line). In addition, our simulation method serves as a useful tool for the analysis of the tumor growth based on model (42). First, it can be used to generate probabilistic forecasts: the medium (green) line shows the expected value forecasts and the upper and lower (orange dotted) lines are confidence bands such that there is a 90% chance that the tumor volume will not

exceed these bands. Second, it allow us to study the state of growth at certain points in time. In particular, we document three critical time points (see vertical lines in Figure EC.4) corresponding to different phases of the tumor growth *in vivo*, therefore signifying the timing of relevant guideline for treatment: T_1, when tumor slow growth (tumor vasculature) is completed since commencement of cell division; T_2, when aggressive tumor growth occurs and the maximum rate of growth is reached; T_3, when growth is slowed again (due to limited supply of oxygen, nutrients, and space). For the given model, these critical points are given by

$$\hat{T}_{1,3} = \frac{1}{\hat{\lambda}} \ln \left[\left(\pm\sqrt{3} - 2 \right) \left(1 - \frac{\hat{K}}{\hat{X}(0)} \right) \right]$$

and

$$\hat{T}_2 = \frac{1}{\hat{\lambda}} \ln \left(\frac{\hat{K}}{\hat{X}(0)} - 1 \right)$$

(inflection point). Finally, Figure EC.5 shows the tumor volume (for the 95th percentile simulated radius) at the three different phases and the corresponding simulated distributions of the tumor radii and volumes as percentages of the estimated maximum limit, $100X(\hat{T}_i)/\hat{K}\%$. At time $\hat{T}_1$, there is a large concentration of probability mass at lower volume levels but there is also a nonzero probability of volume expansion as implied by the left tail. The distribution of the relative volume becomes more symmetric and the kurtosis reduces as we approach the terminal phase at $\hat{T}_3$.

EC.8. Other applications

EC.8.1. The CGMY model

As explained, for example, in Ballotta and Kyriakou (2014), similarly to other popular Lévy models, like the variance gamma or normal inverse Gaussian, the Carr–Geman–Madan–Yor (CGMY) model can be represented as a subordinated arithmetic Brownian motion; unlike those though, the subordinator is only known via its Laplace transform. More specifically, the CGMY process is given by

$$X(t) = \theta Z(t) + W(Z(t)), \tag{EC.18}$$

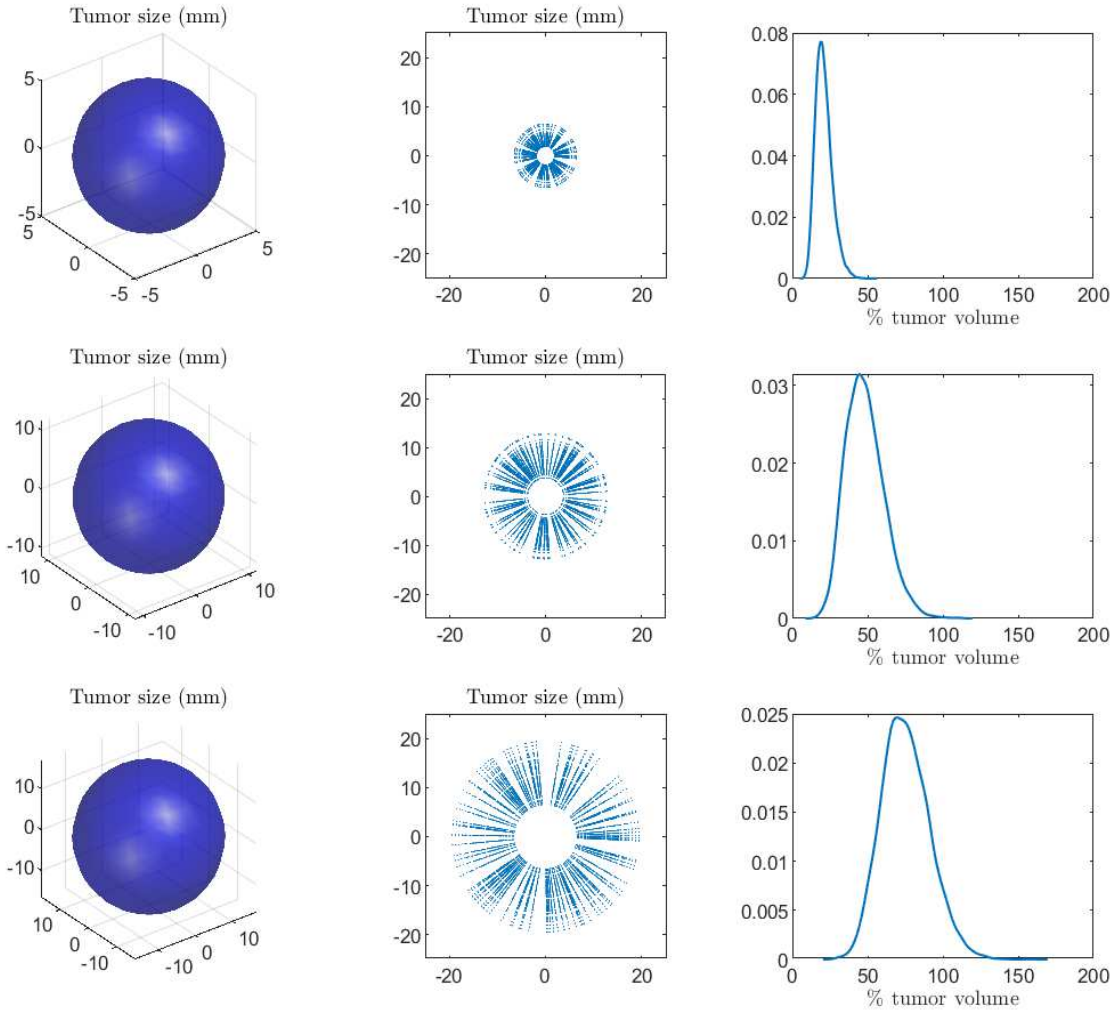
where W is a standard Brownian motion, $\theta := (G - M)/2$, $G \geq 0$, $M \geq 0$ and Z is a subordinator independent of W with Laplace transform

$$E\left[e^{-aZ(t)}\right] = \exp\left((t-u)\, C\Gamma(-Y) \left(2\left(2a + GM\right)^{Y/2} \cos\left(\xi\left(a; G, M\right) Y\right) - M^Y - G^Y \right) \right), \quad u < t, \tag{EC.19}$$

where $\xi(a; G, M) := \arctan\left(\tilde{\theta}^{-1} \sqrt{2a - \theta^2} \right)$ and $\tilde{\theta} := (G + M)/2$.

In view of the previous results, our moment-based random number generator can be adapted to the efficient simulation of the CGMY model trajectories as follows:

Figure EC.5 Tumor growth



Notes. Tumor size (3D & 2D) (left & central plots) and probability density estimates of % tumor volume with respect to estimated maximum limit (right plots) at the three different phases (top to bottom): T_1 (completion of initial growth), T_2 (aggressive tumor growth) and T_3 (terminal growth).

1. Compute the moments of $Z(t)$ (via, for example, Choudhury and Lucantoni 1996) based on (EC.19)
2. Simulate $Z(t)$ using a fitted a Pearson curve to the corresponding theoretical distribution
3. Simulate W subject to time-change $Z(t)$ and, consequently, $X(t)$ according to (EC.18).

In addition, in financial modelling, the related asset price process under the risk neutral measure is

$$S(t) = S(u)\exp\left((r+\omega)(t-u) + X(t)\right),$$

where $\omega := -C\Gamma(-Y)\left((G+1)^Y - G^Y + (M-1)^Y - M^Y\right)$. Conditional on $Z(t)$, we also have that

$$(\ln S(t)|\ln S(u), Z(t)) \sim \mathcal{N}(m, s^2),$$

where $m := \ln S(u) + (r+\omega)(t-u) + ((G-M)/2)Z(t)$ and $s^2 := Z(t)$, and the following Black–Scholes-type representation holds:

$$E\left[e^{-r(t-u)}\left(S(t)-K\right)^+ \middle| Z(t)\right] = e^{-r(t-u)}\left[e^{m+\frac{1}{2}s^2}N\left(\frac{m+s^2-\ln K}{s}\right) - KN\left(\frac{m-\ln K}{s}\right)\right],$$

where $N(\cdot) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\cdot}\exp\left(-z^2/2\right)dz$.

Some numerical results are reported in Table EC.2. The simulation of the CGMY process is remarkably fast as its increments are independent and we no longer face a conditional Laplace transform of Z, which also does not involve special functions that can severely slow down the overall execution, simplifying considerably the simulation task. Similarly to the other models in the paper, our scheme clearly converges also in this case as shown in Figure EC.6.

Table EC.2 European plain vanilla call option in the CGMY model

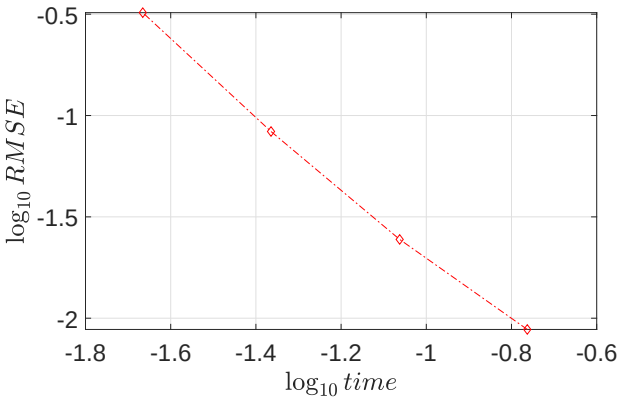
True option price	20.1965	Bias (s.e.)	0.0005	(0.001)
$\mathcal{M}\times 10^4$	4	16	64	256
RMSE	0.1726	0.0866	0.0432	0.0216
Time	0.01	0.02	0.08	0.32

Notes. True (reference) price via the method of Fang and Oosterlee (2008) based on parameters $S(0) = K = 100$, $C = 0.9795$, $G = 3.512$, $M = 10.96$, $Y = 0.8$, $r = 4\%$, $T = 1$ (Černý and Kyriakou 2011, Table 2). Bias computed using 10^9 simulation trials as the difference between the expected value of the simulation estimator (based on fitted Pearson to $Z(T)$) and the true price. RMSE $= \sqrt{\text{bias}^2 + \text{s.e.}^2}$ computed based on standard error (s.e.) for each $\mathcal{M}$ number of simulations as shown in the second row. All computing times are in seconds.

EC.8.2. Greeks

Unbiased estimation of option Greeks using Monte Carlo simulation is not as straightforward as the computation of option prices due to potential discontinuities in the option payoff function, such as, for example, in the case of barrier and digital options. In general, the standard two methods for the sensitivities are the pathwise and likelihood ratio method introduced by Broadie and Glasserman (1996). Both rely on an interchange of integration and differentiation under certain regularity conditions which are easier to justify for the likelihood ratio method, as density functions are usually

Figure EC.6 Convergence of our Monte Carlo method in the CGMY model: the case of European plain vanilla call option



smooth functions of their parameters whereas payoff functions are not. Conditioning arguments simplify the problem as they offer us direct access to standard densities. For more details, refer to Broadie and Kaya (2004, 2006).

EC.9. The code

This part contains the Matlab code for simulating the Heston model using the approach proposed in this paper, tailored to the case of a European plain vanilla option.

```
% This script contains the code to price European call options in the ...
    Heston model following the procedure proposed in the paper "Unified ...
    moment-based modelling of integrated stochastic processes" by Kyriakou, ...
    I. and Brignone, R. and Fusai, G.

S0 = 100;  % initial asset price
K = 100;   % strike price

% name of the parameter set
paramSet = 'H1';

% set parameters
switch paramSet
case 'H1'
Vstart = 0.010201; k = 6.21; theta = 0.019; v = 0.61; rho = -0.7; r = ...
    3.19/100; T = 1;  true_price = 6.8061;
case 'H2'
Vstart = 0.09;     k = 2;    theta = 0.09;  v = 1;  rho = -0.3;   r = ...
    5/100;    T = 5;  true_price = 34.9998;
end

nsimul = 10^5; % number simulations

%martCorr = 'false'; % do not apply martingale correction
martCorr = 'true';   % apply martingale correction

tic % start stopwatch timer
```

```
% implement Algorithm 2 in the paper in the Heston model
[Send, Vend] = simHeston(S0, r, T, Vstart, k, theta, v, rho, nsimul, martCorr);
% compute European call option price estimate by Monte Carlo simulation
C = normfit(exp(-r*T)*max(0, Send - K));
TT = toc; % stop timer

% compare Monte Carlo price estimate with true option price
disp(['True price: ' num2str(true_price), ', Monte Carlo price estimate: ' ...
    num2str(C) ', Comput. time (s): ' num2str(TT)])
```

EC.9.1. Functions

Heston model simulation.

```
function [Send, Vend] = simHeston(S0, r, T, Vstart, k, theta, v, rho, ...
    nsimul, martCorr)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This function simulates transitions in the Heston model using Algorithm 2 ...
    in the paper.
%
% Inputs:  - S0        (scalar or 1 x nsimul vector): initial asset price
%          - r         (scalar): risk-less rate
%          - T         (scalar): maturity in years
%          - Vstart    (scalar or 1 x nsimul vector): initial variance
%          - k         (scalar): speed of mean reversion
%          - theta     (scalar): long-run mean
%          - v         (scalar): volatility of variance
%          - rho       (scalar): correlation
%          - nsimul    (scalar): number of simulations
%          - martCorr  ('true' or 'false'): choose whether or not apply ...
    martingale correction
% Outputs: - Send      (1 x nsimul vector): asset price at maturity
%          - Vend      (1 x nsimul vector): variance at maturity

%% Step 1: Simulate terminal variance $\sigma^2(t)|\sigma^2(u)$

% degrees of freedom of noncentral chi-squared distribution
Δ = 4*theta*k/(v^2);
% noncentrality parameter
lambda = 4*k*exp(-k*T)/(v^2 * (1-exp(-k*T))) .* Vstart;
% generate \sigma^2(t)
if length(Vstart)==1
   Vend = v^2*(1-exp(-k*T))/4/k * ncx2rnd(Δ, lambda, 1, nsimul);
else
   Vend = v^2*(1-exp(-k*T))/4/k * ncx2rnd(repmat(Δ, size(lambda)), lambda);
end

%% Step 2: Compute moments of $\int_{u}^{t} \sigma^2(s) ds | ...
    \sigma^2(t),\sigma^2(u)$

% define Laplace transform of $\int_{u}^{t} \sigma^2(s) ds | ...
    \sigma^2(t),\sigma^2(u)$
lt = @(u) getCondmgfBK(-u, k, theta, v, Vend, Vstart, T, 0);
% compute first four integer moments
[mu] = Algorithm1(lt,4,nsimul);
```

```
37
38      %% Step 3: Draw random numbers from Pearson given moments (generate ...
            $\int_{u}^{t} \sigma^2(s) ds | \sigma^2(t),\sigma^2(u)$) generate draws ...
            from $\int_{u}^{t} \sigma^2(s) ds | \sigma^2(t),\sigma^2(u)$ given moments
39
40      [IVds] = pearsrndVect(mu);
41
42      %% Step 4: Simulate terminal price $S(t) | \int_{u}^{t} \sigma^2(s) ds , ...
            \sigma^2(t),\sigma^2(u)$ generate random draws from normal distribution
43
44      Z = randn(1, nsimul);
45      % compute variance
46      sigma_square = (1 - rho^2) * IVds;
47      switch martCorr
48      case 'true'
49      % apply martingale correction
50      tmp = (k*rho/v - 0.5)*IVds + rho/v*(Vend-Vstart)+ sqrt(sigma_square).*Z;
51      Mart = mean(exp(tmp));
52      Kstar = -log(Mart);
53      mu = log(S0) + r * T - 0.5 * IVds + rho * (1 / v) * (Vend - Vstart + k * ...
            IVds) + Kstar;
54      case 'false'
55      % simulate integral of \sqrt{V_t}dW(s)
56      IrtVdW = (1 / v) * (Vend - Vstart - k*theta*T + k*IVds);
57      % compute mean
58      mu = log(S0) + r * T - 0.5 * IVds + rho * IrtVdW;
59      end
60      % asset price simulation
61      Send = exp(mu + Z .* sqrt(sigma_square));
62      end
```

Conditional Laplace transform.

```
1       function  laptrans = getCondmgfBK(ltpar, k, theta, sigma, Vend, Vstart, t, u)
2       %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
3       % This function evaluates the Laplace transform of the conditional ...
            integrated variance in the Heston model.
4
5       % Inputs:  - ltpar    (vector): parameter of the Laplace transform
6       %          - k        (scalar): speed of mean reversion
7       %          - theta    (scalar): long run mean variance
8       %          - sigma    (scalar): volatility of variance process
9       %          - Vend     (vector): terminal variance
10      %          - Vstart   (vector): initial variance
11      %          - t        (scalar): final date
12      %          - u        (scalar): initial date
13      % Outputs: - laptrans (vector): numerical value of the Laplace transform
14
15      ks = k.^2;
16      gamm_a = (ks+2.*ltpar.*sigma.^2).^.5;
17
18      %1st term
19      term1num=gamm_a.*exp(-0.5*(gamm_a-k).*(t-u)).*(1-exp(-k*(t-u)));
20      term1den=k.*(1-exp(-gamm_a*(t-u)));
21      term1=term1num./term1den;
22
```

```
%2nd term
term2 = k.*(1+exp(-k*(t-u))./(1-exp(-k*(t-u)))-...
gamm_a.*(1+exp(-gamm_a*(t-u)))./(1-exp(-gamm_a*(t-u)));
term2 = exp((Vend+Vstart).*term2/sigma^2);

%%%Bessel function
nu = (-1)+2.*k*theta/sigma^2;%index Bessel function
num1 = 4.*exp((-0.5).*gamm_a.*(t-u)).*gamm_a;
den1 = (1-exp(-gamm_a.*(t-u))).*sigma.^2;
bessel1 = besseli(nu, (Vend.*Vstart).^.5.*num1./den1);

num2 = 4.*exp((-0.5).*k.*(t-u)).*k;
den2 = (1-exp(-k.*(t-u))).*sigma.^2;
bessel2 = besseli(nu, (Vend.*Vstart).^.5.*num2./den2);

laptrans = term1.*term2.*bessel1./bessel2;
```

Efficient computation of moments from Laplace transform.

```
function [mu] = Algorithm1(laptrans,N,lVend)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This function computes the moments from a Laplace transform implementing ...
    Algorithm 1 in the paper.

% Inputs:  - laptrans  (function handles): Laplace transform of conditional ...
    integrated process
%          - N         (scalar): Number of moments
%          - LVend     (Number of of different terminal variances)
% Outputs: - mu        (NrMoments x LVend matrix): matrix of moments

% prepare matrix of moments
mu = ones(N,lVend);
% set algorithm parameters
gamma = 11; L = 1;
% set r() as function of n
r = 10.^(-gamma./(2*(1:N)*L));
% n = 1
% set parameters
n = 1;  alpha(n) = 1;
% multiplicative factor
disc = factorial(n)/2./n/L./(r(n)^n)/alpha(n)^n;
% compute mu1
mu(1,:) = disc*(laptrans(alpha(n)*r(n)) + (-1)^n * laptrans(-alpha(n)*r(n)));
% n = 2
n = 2;
% compute alpha
alpha = 1./mu(1,:);
% multiplicative factor
disc = factorial(n)/2/n/L/(r(n)^n)./(alpha.^n);
% compute mu2
mu(2,:) = real(disc .* (laptrans(alpha*r(n)) + (-1)^n * ...
    laptrans(-alpha*r(n)) +2 * ...
    real(laptrans(alpha*r(n)*exp(pi*1i/n/L)).*exp(-pi*1i/L))));
% compute other moments
for n = 3:N
% compute alpha
```

```
    alpha = (n-1).*mu(n-2,:)./mu(n-1,:);
    % set indexes for the summation
    j = (1 : n-1)';
    % multiplicative factor
    disc = factorial(n)/2/n/L/(r(n)^n)./(alpha.^n);
    % calculate mun
    mu(n,:) = real(disc.*(laptrans(alpha*r(n)) + (-1)^n * laptrans(-alpha*r(n)) ...
        + ...
    2 * sum(real(laptrans(exp(pi*1i*j/n/L) * alpha*r(n)).*exp(-pi*1i*j/L) ))));
    end
    end
```

Vectorized Pearson random number generator.

```
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    % This function generates random numbers from a 4-integer-moment fitted ...
        Pearson distribution; it is a vectorized version of the built-in ...
        function "pearsrnd".

    % Inputs:  - mu         (4 x nsimul vector): matrix of moments
    % Outputs: - X          (1 x nsimul vector): Random draws from the fitted ...
        Pearson distribution

    nsimul = size(mu,2);

    % compute variance, skewness and kurtosis from moments
    Var = mu(2,:)-mu(1,:).^2;
    skew = (mu(3,:) - 3*mu(1,:).*Var- mu(1,:).^3)./(Var.^1.5);
    kurt = (mu(4,:) - 4*mu(1,:).*mu(3,:) +6*mu(1,:).^2.*mu(2,:) ...
        -3*mu(1,:).^4)./(Var.^2);

    beta1 = skew.^2;
    beta2 = kurt;
    sigma = sqrt(Var);

    % classify the distribution and find the roots of c0 + c1*x + c2*x^2
    c0 = (4*beta2 - 3*beta1); % ./ (10*beta2 - 12*beta1 - 18);
    c1 = skew .* (beta2 + 3); % ./ (10*beta2 - 12*beta1 - 18);
    c2 = (2*beta2 - 3*beta1 - 6); % ./ (10*beta2 - 12*beta1 - 18);

    cond1 = (c1==0);
    if sum(cond1)>0
    type(cond1) = (0.*double(beta2(cond1)==3) + 2.*double(beta2(cond1)<3) + ...
        7.*double(beta2(cond1) > 3));
    end
    cond2 = (c1 ≠0 & c2==0);
    if sum(cond2)>0
    type(cond2) = 3.*double(c2(cond2) == 0);
    end
    cond3 = (c1 ≠0 & c2≠0);
    if sum(cond3)>0
    kappa = c1 (cond3).^2 ./ (4*c0 (cond3).*c2(cond3));
    type(cond3) = double (kappa < 0).*1 + double (kappa≥0 & kappa≤1-eps).*4 + ...
        double (kappa≤1+eps & kappa>1-eps).*5 + ...
    double(kappa ≥ 1+eps).*6;
```

```
    tmp = -(c1(cond3) + sign (c1(cond3)).*sqrt(c1(cond3).^2 - ...
        4*c0(cond3).*c2(cond3))) ./ 2;
    a1(cond3) = tmp ./ c2(cond3);
    a2(cond3) = c0 (cond3) ./ tmp;
    %term = a1(cond3);
    for i=1:sum(cond3)
    if (real(a1(i)) > real(a2(i))), tmp = a1(i); a1(i) = a2(i); a2(i) = tmp; end
    end
    end

    denom = (10*beta2 - 12*beta1 - 18);
    % c0 = c0 ./ denom;
    c1 = c1 ./ denom;
    c2 = c2 ./ denom;
    % coefs = [c0 ; c1 ; c2];
    rtype = unique(type);
    % rtype
    %m1 = zeros(1,Nsimul); m2 = zeros(1,Nsimul);
    r = zeros(1,nsimul);
    for i=1:length(rtype)
    switch rtype(i)
    case 0
    pos0 = (type == rtype(i));
    r(pos0) = normrnd(0,1,[sum(pos0) 1]);
    case 1
    pos1 = (type == rtype(i));
    m1 = (c1(pos1) + a1(pos1)) ./ (c2 (pos1) .* (a2(pos1) - a1(pos1)));
    m2 = -(c1(pos1) + a2(pos1)) ./ (c2 (pos1) .* (a2(pos1) - a1(pos1)));
    r(pos1) = a1(pos1) + (a2(pos1) - a1(pos1)) .* betarnd(m1+1,m2+1);
    case 4
    pos4 = (type == rtype(i));
    m = 1 ./ (2*c2(pos4));
    nu = 2.*c1(pos4).*(1 - m) ./ sqrt((4.*c0(pos4).*c2(pos4) - c1(pos4).^2));
    b = 2*(m-1);
    a = sqrt(b.^2 .* (b-1) ./ (b.^2 + nu.^2)); % gives unit variance
    lambda = a.*nu ./ b; % gives zero mean
    RND = zeros(1,sum(pos4));
    for k=1:sum(pos4)
    RND(k) = pearson4rnd(m(k),nu(k),a(k),lambda(k),1);
    end
    r(pos4) = RND;
    case 6
    pos6 = (type == rtype(i));
    % F location-scale: standard support (a2,Inf) or (-Inf,a1)
    m1= (a1(pos6) + c1(pos6)) ./ (c2 (pos6).*(a2(pos6) - a1(pos6)));
    m2 = -(a2(pos6) + c1(pos6)) ./ (c2 (pos6).*(a2(pos6) - a1(pos6)));
    % a1 and a2 have the same sign, and they've been sorted so a1 < a2
    nu1 = double(a2(pos6)<0).*2.*(m2 +1) + double (a2(pos6)≥0).*2.*(m1 +1);
    nu2 =  -2*(m1 + m2 + 1);
    ranF = frnd(nu1,nu2);
    r(pos6) = double (a2(pos6)<0).* (a2(pos6) + (a2(pos6) - a1(pos6)) .* ...
        (nu1./nu2) .*ranF) + ...
    double (a2(pos6)≥0).* (a1(pos6) + (a1(pos6) - a2(pos6)) .* (nu1./nu2) .*ranF);

    end
    end

    % scale and shift
    X = abs(r.*sigma + mu(1,:));
```

```
93
94      end
95
96      function r = pearson4rnd(m,nu,a,lambda,sizeOut)
97      % PEARSON4RND Generate Pearson type 4 random variates.
98      %
99      %   Exponential rejection method for log-concave densities from
100     %   Devroye, Section VII.2. Valid only when m > 1, if called by PEARSRND.
101     %
102     %   References:
103     %      [1] Devroye, L. (1986) Non-Uniform Random Variate Generation,
104     %          Springer-Verlag.  Also available in PDF format on-line at
105     %          http://cgm.cs.mcgill.ca/¬luc/rnbookindex.html.
106     %      [2] Heinrich, J. (2004) "A Guide to the Pearson Type IV Distribution",
107     %          CDF/MEMO/STATISTICS/PUBLIC/6820, available on-line at
108     %          http://www-cdf.fnal.gov/publications/cdf6820_pearson4.pdf.
109
110     logK = -logHypGeo(m,nu/2) + (gammaln(m) - gammaln(m-.5)) - log(sqrt(pi)*a);
111
112     % generate y = arctan(x) with density g(y) = K*cos(y)^(2m-1)*exp(-nu*y)
113     b = 2*(m-1);
114     M = atan(-nu./b); % mode of y = arctan(x)
115     cosM = a ./ sqrt(b-1);
116     loggM = b.*log(cosM) - nu.*M; % log(g(mode)) + log(K)
117     invgM = exp(-loggM - logK); % 1/g(mode)
118
119     outClass = superiorfloat(m,nu,a,lambda);
120     r = zeros(sizeOut,outClass);
121     j = 1:numel(r);
122     while length(j) > 0
123     U = 4*rand(size(j)); % dist'd Unif([0,4])
124     S = (U>2); % use this to get a random +1/-1
125     U(S) = U(S) - 2; % now dist'd Unif([0,2])
126     negEstar = log(max(U,1)-(U>1)); % zero for U≤1, dist'd Exp(1) for U>1
127     X = min(U,1) - negEstar; % U or 1+Estar
128     Z = log(rand(size(j))) + negEstar; % -E or -E-Estar
129     X = M + (2*S-1).*X.*invgM;
130     k = (abs(X) < pi/2) & (Z ≤ b.*log(abs(cos(X))) - nu.*X - loggM);
131     r(j(k)) = X(k);
132     j(k) = [];
133     end
134
135     % transform, scale, and shift to standard Pearson type IV
136     r = a.*tan(r) + lambda;
137
138     end
139
140     function logF = logHypGeo(x,y)
141     % LOGHYPGEO A special case of the hypergeometric function.
142     %
143     % Returns log F(-iy,iy,x,1) = log abs(gamma(x)/gamma(x+iy))^2, where F is ...
            the complex hypergeometric function. This is based on methods described ...
            in Heinrich, J. (2004) "A Guide to the Pearson Type IV Distribution", ...
            CDF/MEMO/STATISTICS/PUBLIC/6820.
144
145     % for small x, compute (1+(y/x)^2)*...*(1+(y/(x+n))^2) which scales ...
            F(-iy,iy,x,1) to F(-iy,iy,x+n,1), which we can compute quickly if x+n is ...
            large
146     if x < 100
```

```
147       xstep = x:1:100;
148       % r = prod(1 + (y./xstep).^2);
149       logr = sum(log1p((y./xstep).^2));
150       x = xstep(end) + 1;
151       else
152       logr = 0;
153       end
154
155       % compute F(-iy,iy,x+n,1), then multiply by r to get F(-iy,iy,x,1)
156       logs = zeros(class(y)); logp = zeros(class(y)); f = zeros(class(y));
157       while logp-logs > log(eps)
158       % p = p .* (y.^2 + f.^2) ./ (x.*(f+1));
159       logp = logp + log(y.^2 + f.^2) - log(x.*(f+1));
160       x = x + 1;
161       f = f + 1;
162       % s = s + p;
163       logs = logs + log1p(exp(logp - logs));
164       end
165       % F = r.*s;
166       logF = logr + logs;
167       end
```