Journal of Economic Dynamics & Control 168 (2024) 104948

Contents lists available at ScienceDirect

Journal of Economic Dynamics and Control

journal homepage: www.elsevier.com/locate/jedc

# Closed-form approximations of moments and densities of continuous–time Markov models ☆

Dennis Kristensen [a,*], Young Jun Lee [b], Antonio Mele [c,d,e]

[a] Department of Economics, University College London, United Kingdom
[b] Korea Institute for International Economic Policy, Republic of Korea
[c] USI Lugano, Switzerland
[d] Swiss Finance Institute, Switzerland
[e] CEPR, United Kingdom

ARTICLE INFO

*JEL classification:*
C13
C32
C63
G12
G13

*Keywords:*
Continuous-time models
Jump-diffusion
Transition density
Stochastic volatility
Closed-form approximations
Maximum-likelihood estimation
Option pricing

ABSTRACT

This paper develops power series expansions of a general class of moment functions, including transition densities and option prices, of continuous-time Markov processes, including jump–diffusions. The proposed expansions extend the ones in Kristensen and Mele (2011) to cover general Markov processes, and nest transition density and option price expansions recently developed in the literature, thereby connecting seemingly different ideas in a unified framework. We show how the general expansion can be implemented for fully general jump–diffusion models. We provide a new theory for the validity of the expansions which shows that series expansions are not guaranteed to converge as more terms are added in general once the time span of interest gets larger than some model–specific threshold. Thus, these methods should be used with caution when applied to problems with a larger time span of interest, such as long-term options or data observed at a low frequency. At the same time, the numerical studies in this paper demonstrate good performance of the proposed implementation in practice when applied to pricing options with time to maturity below three months. Thus, our expansions are particularly well suited for pricing ultra-short-term (such as "zero–day") options.

## 1. Introduction

Continuous-time jump-diffusion processes are used in economics and finance to model the dynamics of state variables (see, e.g., Björk, 2009). They lead to a simple and elegant analysis of problems such as the pricing of financial assets, portfolio management and other dynamic phenomena. This comes at a big computational cost though: Many relevant characteristics, such as moments and densities, of such processes cannot be expressed in closed-form except in a few special cases. This hampers their practical use and implementation. This has led researchers to develop numerical methods for the computation of these. Broadly speaking, these methods fall in three categories: Finite–difference methods (Ames, 1992), simulation–based methods (see, e.g., Elerian et al., 2001; Brandt and Santa-Clara, 2002; Durham and Gallant, 2002; Beskos et al., 2009; Kristensen and Shin, 2012; Sermaidis et al., 2013) and series expansions (see, e.g., Aït-Sahalia, 2002; Bakshi et al., 2006; Yu, 2007; Aït-Sahalia, 2008; Filipović et al., 2013; Li, 2013). This paper focuses on the latter category.

☆ We would like to thank participants at the NBER–NSF Time Series Conference 2022 and Seungmoon Park for valuable comments and suggestions.

* Corresponding author.
*E-mail addresses:* d.kristensen@ucl.ac.uk (D. Kristensen), y.lee@kiep.go.kr (Y.J. Lee), antonio.mele@usi.ch (A. Mele).

https://doi.org/10.1016/j.jedc.2024.104948
Received 12 July 2023; Received in revised form 3 July 2024; Accepted 27 August 2024

Available online 19 September 2024

Most existing expansions proposed in the literature are application specific: Depending on the particular features of the chosen moment and model of the underlying stochastic process, different methods have been developed. One exception is Kristensen and Mele (2011) who developed power series expansions that covered a general class of moment functions and the transition density of multivariate diffusion processes. Their focus was on applications to option pricing but the class of expansions applies more generally. The current paper makes four contributions:

First, we demonstrate that the class of series expansions of Kristensen and Mele (2011) are easily extended to cover fully general continuous–time Markov models, including any jump–diffusion process. Thus, the proposal of Kristensen and Mele (2011) can in principle be applied to any moment of any Markov process. As part of this extension, we present a novel derivation and representation of the series expansion of Kristensen and Mele (2011). This new representation highlights important features of the original expansion that was perhaps not obvious from the analysis of Kristensen and Mele (2011).

Second, we revisit the recent work of Yang et al. (2019) and Wan and Yang (2021) and demonstrate that in fact their proposed expansions of transition densities and option prices are special cases of Kristensen and Mele (2011). Thus, at a theoretical level the expansions in Yang et al. (2019) and Wan and Yang (2021) are not new. At the same time, it should be emphasized that Yang et al. (2019) and Wan and Yang (2021) make important contributions in terms of the practical implementation of the proposal in Kristensen and Mele (2011). They develop numerical algorithms that allow for fast implementation of the general method of Kristensen and Mele (2011) when applied to transition densities and option prices of diffusion processes and a limited set of jump–diffusion processes. As such, the current paper should hopefully clarify the relationship between these three existing papers and their relative contributions to the literature.

Third, we propose a novel numerical implementation of our series expansions when applied to general jump–diffusion models. The algorithms of Kristensen and Mele (2011) and Yang et al. (2019) are restricted to pure diffusions while the extension found in Wan and Yang (2021) requires the jump component to be fully independent of the diffusive component. That is, the jump intensity and the jump sizes are not allowed to be state–dependent. Our numerical implementation allows for both to be state–dependent. We demonstrate through a series of numerical studies that our numerical method works well in practice as long as the time to maturity is not too large.

Fourth, we provide a novel theory for the validity of power series expansions of moment functions of continuous–time Markov processes used here and elsewhere in the literature, including all above references to papers employing series–based approximations. Most existing theoretical results for these expansions only show that a given moment expansion converges as the time interval over which the conditional moment is defined shrinks to zero. As such existing results provide no guarantees that the approximation error will get smaller than more terms are added to expansion; in fact, nothing rules out that the approximation error may actually explode as more terms are added. For the power series expansion to be reliable, it is desirable with conditions under which the expansions converge not only over shrinking time intervals but also over a fixed time interval. We here provide guarantees for the approximations to be numerically stable as the order of the approximation grows. Our theoretical results rely on semi–group theory as also used by, e.g., Hansen and Scheinkman (1995) to analyze the properties of continuous–time Markov processes.

Our theoretical results demonstrate that power series expansions of Markov moments may very well not converge: The chosen moment and model has to satisfy certain regularity conditions for this to hold and the time span over which the moment is computed has to stay below a certain threshold. In particular, we demonstrate that the expansions of transition densities and option prices proposed by Kristensen and Mele (2011), Yang et al. (2019) and Wan and Yang (2021) do not converge for all values of $t$, where $t$ is the time span of interest. That is, these methods will eventually fail once t gets "too large". As such, the expansions proposed in these papers and the extension to general jump–diffusions developed here should be used with care. At the same time, our numerical studies confirm good performance of our method when used to price options with time to maturity below three months. As such, our expansions are particularly well suited to approximate the price of ultra-short-term (such as "zero–day") options.

The remains of the paper are organized as follows. Section 2 presents series expansions of a broad class of moments and densities of basically any continuous–time Markov process and shows that this nests existing proposals as special cases. Section 3 analyzes the theoretical properties of the power series expansion over both shrinking and fixed time distances. In section 4, we discuss the numerical implementation of the general method when applied to general jump–diffusion models. Section 5 examines the numerical performance of our numerical algorithm. Section 6 concludes. Appendix B gathers all proofs.

## 2. Expansions of semi–groups with application to moments of Markov processes

We here first introduce the concept of semi–groups and propose a method for approximating such using a generalised version of the well-known Taylor series expansion. Next, as a leading example of the general framework, we show how our proposal can be used to obtain closed–form expansions of, potentially irregular, moments and densities of a broad class of continuous–time Markov processes.

Let $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$ be a normed function space where a given function $f \in \mathcal{F}$ and $\left\{E_t : t \geq 0\right\}$ be a semi–group with domain $\mathcal{F}$. A semi–group is a family of linear operators indexed by $0 \leq t < \infty$ from $\mathcal{F}$ onto $\mathcal{F}$[1] satisfying (i) $E_0 f = f$ and (ii) $E_{s+t} f = E_s E_t f$ for all $s, t \geq 0$. For an overview of the general theory of semi–groups with applications to Markov processes we refer the reader to Ethier and Kurtz (1986); for applications of semi–group theory in econometrics and finance, see Aït-Sahalia et al. (2010).

[1] For any given $t \geq 0$ and $f, g \in \mathcal{F}$: $E_t f \in \mathcal{F}$ and $E_t(f+g) = E_t f + E_t g$.

Often we are interested in evaluating the function $E_t f(x)$ for a particular choice of $f \in \mathcal{F}$. To economize on notation, we will in the following frequently use $u_t(x)$ to denote $E_t f(x)$ in this case,

$$u_t(x) \equiv E_t f(x). \tag{2.1}$$

Suppose that $u_t(x)$ is not available on closed form and so needs to be approximated using numerical methods; we here show how this can be approximated through a Taylor series expansion of $u_t(x)$ w.r.t. $t$ around $t=0$ when either $f$ is sufficiently regular, where the notion of "regular" will be made clear below, or it can be expressed as the limit of a regular function. The proposal is a generalisation of the one of Kristensen and Mele (2011), but the derivation will be carried out using semi–group theory which simplifies the exposition substantially compared to Kristensen and Mele (2011) and provides new insights into the expansion.

In order for the Taylor series expansion to be valid, we require $E_t f$ to be differentiable w.r.t. $t$ at $t=0$ and let $Bf = \lim_{t\to 0^+} \left(E_t f(x) - f(x)\right)/t$ denote this derivative. The operator $B$, usually referred to as the generator of $E_t$, is generally only well-defined on a subset of $\mathcal{F}$ which we denote $\mathcal{D}(B)$. Below we state the formal definitions, where we choose to work with the so–called extended generator which is defined as follows (see, e.g., Meyn and Tweedie (1993)):

**Definition 2.1.** We denote by $\mathcal{D}(B)$ the set of functions $f \in \mathcal{F}$ for which there exists $g \in \mathcal{F}$ such that, for each $t \geq 0$,

$$E_t f = f + \int_0^t E_s g\, ds, \qquad \left\| E_t |g| \right\|_{\mathcal{F}} < \infty, \tag{2.2}$$

and we write $Bf := g$ and call $B$ the (extended) generator of $E_t$.

Taking derivatives w.r.t. $t$ on both sides of eq. (2.2), we obtain the following equivalent representation of $E_t$ for any $f \in \mathcal{D}(B)$,

$$\partial_t E_t f = B E_t f, \qquad t > 0. \tag{2.3}$$

This can also informally be written as

$$E_t f = e^{Bt} f. \tag{2.4}$$

Thus, the generator $B$ fully characterizes the properties of $E_t$ when restricted to $\mathcal{D}(B)$.

If the chosen $f$ is regular in the sense that it satisfies $f \in \mathcal{D}\left(B^M\right)$ then (2.3) implies that $\partial_t^m E_t f(x) = B^m E_t f$ and so the following Taylor series expansion of $u_t(x)$ w.r.t. $t$ around $t=0$ is valid,

$$\hat{u}_t(x) \equiv \sum_{m=0}^{M} \frac{t^m}{m!} \partial_t^m u_t(x)\big|_{t=0^+} = \sum_{m=0}^{M} \frac{t^m}{m!} B^m f(x), \tag{2.5}$$

where under weak conditions $\hat{u}_t = u_t + O\left(t^M\right)$.

If the chosen $f$ is irregular in the sense that $f \notin \mathcal{D}\left(B^M\right)$ then above approximation is not well–defined/applicable; examples of such are provided below in the case of jump–diffusions. In order to still obtain a closed–form approximation to $E_t f$ in such cases, we propose to replace $f$ by a regularised version of it, denoted $u_{0,s}$, $s \geq 0$, that does belong to $\mathcal{D}\left(B^M\right)$. We require $u_{0,s}$ to satisfy the following two conditions:

**A.0** (i) $\lim_{s\to 0^+} u_{0,s} = f$ and (ii) $u_{0,s} \in \mathcal{D}\left(\left(\partial_s\right)^{M_1}\right) \cap \mathcal{D}\left(B^{M_2}\right)$ for some $s \geq 0$ and $M_1, M_2 \geq 1$.

The function $u_{0,s}$ is chosen by the researcher and needs to be available on closed form for the subsequent approximation to be operational. The choice of $u_{0,s}$ is application specific in the sense that Assumption A.0 has to be satisfied for a given choice of $f$ and $E_t$: Part (i) requires $u_{0,s}$ to converge towards the irregular function of interest $f \notin \mathcal{D}(B)$ as $s \to 0^+$. Part (ii) says that, for some $s > 0$, $u_{0,s}(x)$ is sufficiently regular in the sense that it is $M$ times continuously differentiable in $s$ and each of these derivatives belongs to $\mathcal{D}\left(B^M\right)$.

Assumption A.0 allows for a broad range of regularisers/smoothers. One choice of $u_{0,s}(x)$ which under great generality will satisfy A.0 is $u_{0,s} = E_{0,s} f$ where $E_0$ is another semi–group chosen such that $u_{0,s}$ is available on closed form. This choice clearly satisfies part (i) and if $E_0$ has similar properties as the one of interest, $E$, so that their respective generators have shared domain, then part (ii) will also hold.

Under A.0, the following identity holds:

$$f = u_{0,0} = e^{(-\partial_s)s} u_{0,s},$$

where the second equality simply states that $u_{0,0} = u_{0,s} + \int_0^s \left(-\partial_\tau\right) u_{0,\tau}\, d\tau$. Substituting this into (2.4) yields

$$u_t = e^{Bt} e^{(-\partial_s)s} u_{0,s} = e^{(-\partial_s)s} e^{Bt} u_{0,s}, \tag{2.6}$$

where the last equality uses the following fundamental result: If two infinitesimal operators, say, $B_1$ and $B_2$, commute in the sense that $B_1 B_2 f = B_2 B_1 f$ then $e^{B_1 s} e^{B_2 t} f = e^{B_1 s + B_2 t} f = e^{B_2 t} e^{B_1 s} f$. This applies to the case of $B$ and $\partial_s$, $\partial_s B = B \partial_s$, since $B$ acts on $x$ while $\partial_s$ acts on $s$. Finally, carry out a Taylor series expansion w.r.t. $(s,t)$ to obtain

$$\hat{u}_t(x) = \sum_{m_1=0}^{M_1} \sum_{m_2=0}^{M_2} \frac{(-s)^{m_1} t^{m_2}}{m_1! m_2!} B^{m_2} \partial_s^{m_1} u_{0,s}(x), \tag{2.7}$$

where the order of $\partial_s$ and $B$ can be exchanged since $\partial_s^{m_1} B^{m_2} u_{0,s} = B^{m_2} \partial_s^{m_1} u_{0,s}$. Above approximation generalises the one in (2.5): By setting $M_1 = 0$ and $s = 0$, we obtain (2.5).

The resulting approximation error is of order $O\left(s^{M_1}\right) + O\left(t^{M_2}\right)$. In particular, the above expansion will generally become more precise as $s$ gets smaller. Thus, we ideally want to choose $s$ as small as possible to reduce the approximation error. However, for the chosen value of $s \geq 0$, A.0(ii) has to be satisfied. This rules out, for example, $s = 0$ when $f$ is irregular since $u_{0,0}(x) = f(x)$.

If the approximation error is not a major concern (which is, for example, the case if the order of approximation can be chosen sufficiently large) then one can choose $s = t$ in which case $E_t f = e^{(B-\partial_t)t} u_{0,t}$ and the following special case of (2.7) can be employed,

$$\hat{u}_t(x) = \sum_{m=0}^{M} \frac{t^m}{m!} \left(B - \partial_t\right)^m u_{0,t}(x). \tag{2.8}$$

We will in the following focus on this last version.

### 2.1. Application to Markov jump–diffusions

We here apply above machinery to obtain closed–form approximations of, potentially, irregular moments and densities of a broad class of jump–diffusion processes. Consider a $d$-dimensional process, $x_t \in \mathcal{X} \subseteq \mathbb{R}^d$ that solves the following stochastic differential equation (SDE):

$$dx_t = \mu\left(x_t\right) dt + \sigma\left(x_t\right) dW_t + J_t dN_t, \tag{2.9}$$

where $\mu(x)$ and $\sigma(x)$ are the so-called drift and diffusion functions, respectively, $W_t$ is a $d$-dimensional standard Brownian motion, $N_t$ is a Poisson process with jump intensity $\lambda\left(x_t\right)$, and $J_t$ captures the jump-sizes and has conditional density $\nu\left(\cdot|x_t\right)$. The precise form of $\mu(x)$, $\sigma(x)$, $\lambda(x)$ and $\nu(\cdot|x)$ are chosen by the researcher according to the dynamic problem that is being considered and so are known to us. To keep notation simple and allow for a formal theoretical analysis, we restrict ourselves to the time–homogenous case meaning that none of the functions entering the model depend on $t$; the extension to the time–inhomogenous case can be found in Appendix A.

We are interested in computing some conditional moment $u_t(x) = E_t f(x)$, where $(t,f) \mapsto E_t f(x)$ is a mapping that takes as input any given $t \in \mathbb{R}_+$ and function $f(x)$ and returns the following conditional moment,

$$E_t f(x) \equiv \mathbb{E}\left[\exp\left(-\int_0^t r\left(x_s\right) ds\right) f\left(x_t\right) \middle| x_0 = x\right], \tag{2.10}$$

where $r(x)$ is a fixed function that discounts the future value $f\left(x_t\right)$. This family of linear operators clearly constitutes a semi–group and so we can employ our general theory to this setting.[2] We here have to choose the function space $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$ so that above conditional moment is well-defined for any $f \in \mathcal{F}$; this will be discussed in further detail in the theory section below. One could envision more complex conditional moment functions; as long as these are linear in $f$, all subsequent ideas and results will still apply with obvious modifications. Above definition of $E_t f(x)$ is sufficiently general for the applications that we have in mind as described in the following two examples:

The functions $r(x)$ and $f(x)$ entering (2.10) are chosen by the researcher according to the problem of interest. For example, with $r(x) = 0$ and $f(x) = \delta(y - x)$ for some fixed $y \in \mathcal{X}$, where $\delta(x)$ is Dirac's Delta function, $u_t(x) = p_t(y|x)$, where $p_t$ is the transition density of $x_t$,

$$\Pr\left(x_t \in \mathcal{A} | x_0 = x\right) = \int_{\mathcal{A}} p_t(y|x)\, dy, \qquad \mathcal{A} \subseteq \mathcal{X}.$$

If instead we choose $r_t(x) = r > 0$ and $f(x) = \left(\exp\left(x_1\right) - K\right)^+$ then $u_t(x)$ becomes the price of a European call option with time to maturity $t$ when the state variables $x_t$ satisfy (2.9) under the risk–neutral measure with the first component, $x_{1,t}$, being the log-price

[2] Note that we here opt for the so–called Musiela parameterization where $t$ measures the time distance between the current and some future calendar time point. One could alternatively have defined the function of interest as, for some given $T < \infty$,

$$\tilde{u}_\tau(x) = \mathbb{E}\left[\exp\left(-\int_\tau^T r\left(x_s\right) ds\right) f\left(x_T\right) \middle| x_\tau = x\right],$$

where now $\tau \leq T$ is a calendar time point. In the current time–homogenous case, it is easily seen that $\tilde{u}_\tau(x) = u_{T-\tau}(x)$, where $u_t$ was defined above.

of the underlying asset and the short-term interest rate equals the constant $r$. When $r(x) = x_2$, $u_t(x)$ is the price of the same option but now allowing for a stochastic short–term interest rate, which is the second component of $x_t$.

In most cases, an analytic expression of (2.10) is not available and $u_t(x)$ has to be computed using some form of numerical approximation. The general class of Taylor series expansions developed above provides one such with small computational costs. In order to apply our proposal, we first need to obtain the infinitesimal generator of $E_t$ in (2.10). To this end, we first introduce the infinitesimal generator $A$ of the underlying Markov process $x_t$, which corresponds to the infinitesimal generator of the simpler semi–group $\tilde{E}_t f(x) \equiv \mathbb{E}\left[ f\left(x_t\right) \middle| x_0 = x \right]$. One can show that its generator is given by, for any twice continuously differentiable and bounded function $f(x)$,

$$Af(x) = A_D f(x) + A_J f(x), \tag{2.11}$$

where $A_D f(x)$ and $A_J f(x)$ are the generators of the diffusive and jump component of $x_t$, respectively. With $\sigma^2(x) := \sigma(x)\sigma(x)^\top \in \mathbb{R}^{d\times d}$, these are given by

$$A_D f(x) = \sum_{i=1}^{d} \mu_i(x)\partial_{x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma^2_{ij,t}(x)\partial^2_{x_i,x_j} f(x), \tag{2.12}$$

and

$$A_J f(x) = \lambda(x)\int_{\mathbb{R}^d} [f(x+c) - f(x)]\nu(c|x)\,dc, \tag{2.13}$$

cf. Section 3.2. Here, $\partial_{x_i} f(x) = \partial f(x)/\left(\partial x_i\right)$, $\partial^2_{x_i,x_j} f(x) = \partial^2 f(x)/\left(\partial x_i \partial x_j\right)$ and similar for other partial derivatives. Importantly, the domain of $A$ is a subset of twice–differentiable functions and so rules out $f$ being chosen as the dirac delta function or the pay-off function of a European call option.

It can now be shown, cf. Section 3, that the generator of $E_t f(x)$ in (2.10) is $Bf(x) = Af(x) - r(x)$. In particular, $u_t(x)$ solves the following partial integro-differential equation (PIDE):

$$\partial_t u_t(x) = [A - r]u_t(x), \qquad t \geq 0, x \in \mathcal{X}, \tag{2.14}$$

with initial condition $u_0(x) = f(x)$ for all $x \in \mathcal{X}$. In the case of pure diffusions ($A_J = 0$), the reader may recognise $u_t(x) = E_t f(x)$ as the celebrated Feynman–Kac representation of the solution to (2.14) which is now a partial differential equation. As in (2.4), the solution to this PIDE can be represented in the following abstract manner: $u_t(x) = e^{(A-r)t} f(x)$, where $e^{(A-r)t}$ is the exponential of the operator $A - r$ in the sense that

$$\frac{\partial e^{(A-r)t}}{\partial t} = (A - r)e^{(A-r)t}. \tag{2.15}$$

We apply the general machinery to the problem of approximating the conditional moment function $u_t(x)$ through a Taylor series expansion. When $f \in \mathcal{D}\left((A-r)^M\right)$, we can use (2.5) yielding

$$\hat{u}_t(x) \equiv \sum_{m=0}^{M} \frac{t^m}{m!}\partial_t^m u_t(x)\big|_{t=0} = \sum_{m=0}^{M} \frac{t^m}{m!}(A-r)^m f(x), \tag{2.16}$$

for some $M \geq 1$, where the second equality uses (2.15). This type of moment approximations has found widespread use in the literature; see, e.g., (see, e.g., Aït-Sahalia, 2002; Bakshi et al., 2006; Yu, 2007; Aït-Sahalia, 2008; Filipović et al., 2013; Li, 2013). However, this expansion is not valid (well-defined) when, for example, $f(x)$ is not differentiable since the domain of the operator $A_D$ is restricted to twice differentiable functions, cf. (2.12). The transition density and option pricing examples provided above fall in this category. In such cases, we can instead employ (2.8),

$$\hat{u}_t(x) = \sum_{m=0}^{M} \frac{t^m}{m!}\left(A - r - \partial_t\right)^m u_{0,t}(x), \tag{2.17}$$

where $u_{0,t}(x)$ is chosen to satisfy Assumption A.0. A natural choice of $u_{0,t}(x)$ in the application to moments of jump-diffusion processes, as proposed by Kristensen and Mele (2011), is

$$u_{0,t}(x) = E_{0,t} f(x) = \mathbb{E}\left[ \exp\left( -\int_0^t r\left(x_{0,s}\right) ds \right) f\left(x_{0,t}\right) \middle| x_{0,0} = x \right], \tag{2.18}$$

where $x_{0,t}$ is another jump–diffusion process for which above conditional moment is available on closed form, cf. next subsection and Section 4.

### 2.2. Relationship to existing literature

As explained in the introduction, there is an existing literature on expansions of irregular moments of Markov processes. We here relate our proposal to these: We first extend the proposal of Kristensen and Mele (2011) to also cover jump–diffusions (their original proposal focused on pure diffusions) and show that it falls within above general framework. We then proceed to show that the class of series expansions of Kristensen and Mele (2011) contains as special cases the ones of Yang et al. (2019) and Wan and Yang (2021)

Recall the definition of $u_t(x) = E_t f(x)$ in the jump–diffusion example and suppose that $f$ is irregular. To approximate $u_t(x)$ in this case, Kristensen and Mele (2011) first introduces an auxiliary model on the form

$$dx_{0,t} = \mu_0(x_{0,t})\,dt + \sigma_0(x_{0,t})\,dW_t + J_{0,t}\,dN_{0,t}, \tag{2.19}$$

where $N_{0,t}$ is a Poisson process with jump intensity $\lambda_0(x)$ and $J_{0,t}$ has density $\nu_0(\cdot|x)$. We then choose $u_{0,t}$ as given in (2.18); this can be represented as the solution to the following PIDE,

$$-\partial_t u_{0,t}(x) = \left[A_0 - r\right] u_{0,t}(x), \tag{2.20}$$

with initial condition $u_{0,0}(x) = f(x)$, where $A$ has been replaced by the auxiliary model's generator, $A_0 = A_{0,D} + A_{0,J}$. Above choice of $u_{0,t}$ corresponds to one particular regulariser as introduced in Assumption A.0. Proceeding as in Kristensen and Mele (2011), we then subtract (2.20) from (2.14) and, after some straightforward manipulations, arrive at the following PIDE of $\Delta u_t(x) \equiv u_t(x) - u_{0,t}(x)$:

$$\partial_t \Delta u_t(x) = [A - r]\,\Delta u_t(x) + d_t(x), \tag{2.21}$$

where

$$d_t \equiv \left(A - A_0\right) u_{0,t}. \tag{2.22}$$

Since the initial conditions of (2.14) and (2.20) are the same, the initial condition of (2.21) becomes $\Delta u_0(x) = 0$ which is now smooth and bounded. As with $u_t(x)$, $\Delta u_t(x)$ can be represented as a moment function using Feynman-Kac formula under weak regularity conditions,

$$u_t(x) = u_{0,t}(x) + \int_0^t E_s d_s(x)\,ds. \tag{2.23}$$

The second term on the right-hand side of Eq. (2.23) delivers an exact expression of the difference between $u_t(x)$ and $u_{0,t}(x)$.

The next step utilizes the smoothness of $d_t$ to obtain a Taylor expansion w.r.t. time of this second term. We first develop a power series expansion of the integrand, $w_s(x) \equiv E_s d_s(x)$, $s \geq 0$, at $s = t$ taking the form

$$\hat{w}_s(x) \equiv \sum_{m=0}^{M-1} \frac{(s-t)^m}{m!}\,\partial_s^m w_s(x)\big|_{s=t} = \sum_{m=0}^{M-1} \frac{(s-t)^m}{m!}\left(\partial_t + A - r\right)^m d_t(x),$$

for some $M \geq 1$. Combining these last two equations, substituting the resulting expression into (2.23) and evaluating the integral $\int_0^t \hat{w}_s(x)\,ds$, we obtain the approximation originally proposed in Kristensen and Mele (2011), here extended to the general case of jump–diffusions:

$$\hat{u}_t(x) \equiv u_{0,t}(x) + \sum_{m=0}^{M-1} \frac{t^{m+1}}{(m+1)!}\left(A - r - \partial_t\right)^m d_t(x), \tag{2.24}$$

where $\left(-\partial_t + A - r\right)^0 d_t(x) = d_t(x)$. Finally, combine (2.20) and (2.22) to obtain

$$d_t = Au_{0,t} - A_0 u_{0,t} = Au_{0,t} + \left\{-\partial_t u_{0,t} - r(x)\,u_{0,t}\right\} = \left(A - r - \partial_t\right) u_{0,t}.$$

Substituting this into (2.24), we recognise the resulting expression as being identical to (2.17), where $u_{0,t}$ has been chosen as the solution to (2.20). Thus, the expansion of Kristensen and Mele (2011) is identical to the one developed in Section 2, but was presented in a more convoluted manner as can be seen from above.

Next, we demonstrate that the above class of series expansions include as special cases the approximate transition densities and option prices proposed in Yang et al. (2019) and Wan and Yang (2021). With $r = 0$ and $f(x) = \delta(y - x)$, (2.17) with $u_{0,t}$ chosen as the solution to (2.20) for some auxiliary jump–diffusion model becomes

$$\hat{p}_t(y|x) = \sum_{m=0}^{M} \frac{t^m}{m!}\left(A - \partial_t\right)^m p_{0,t}(y|x), \tag{2.25}$$

where $p_{0,t}(y|x)$ is the transition density of the auxiliary model. Now, let us first consider the transition density expansion developed in Yang et al. (2019) for pure diffusions ($A_J = 0$). Inspecting the expansion presented in eq. (10) of their paper, we recognize it to be identical to above when the auxiliary model is chosen as a Brownian motion with drift, $dx_{0,t} = \mu_0 dt + \sigma_0 dW_t$, with $\sigma_0 = \sigma(x)$. Thus, Yang et al. (2019) is a special case of Kristensen and Mele (2011). This somehow went unnoticed by the authors and we here clarify

the connection between the two papers. Second, consider the expansion of the transition density in Wan and Yang (2021) in the pure diffusion case. As explained by the authors themselves, the preferred version of the expansion used in this paper is the same as the series expansion of Yang et al. (2019) when $\mu_0 = 0$ in the auxiliary BM model. And so the pure diffusion version of Wan and Yang (2021) is also a special case of Kristensen and Mele (2011).

Next, we show that the expansion of option prices developed in Wan and Yang (2021) is again a special case of Kristensen and Mele (2011). Setting $r(x) = 0$ and $f(x) = \left(\exp\left(x_1\right) - K\right)^+$ and again using as auxiliary model a Brownian motion with drift, $\hat{u}_t(x)$ as given in (2.17) delivers an expansion of the expected pay-off of a European option where $u_{0,t}(x)$ is now the pay-off function under the Black–Scholes model. To connect this option price approximation with the corresponding proposal of Wan and Yang (2021), observe that $u_{0,t}(x) = \int f(y)\, p_{0,t}(y|x)\, dy$, where $p_{0,t}(y|x)$ is given in (4.4). Substituting this into (2.17) and changing the order of integration and differentiation yields

$$\begin{aligned}\hat{u}_t(x) &= \sum_{m=0}^{M} \frac{t^m}{m!} \left(A - \partial_t\right)^m \int f(y)\, p_{0,t}(y|x)\, dy = \int f(y) \left\{ \sum_{m=0}^{M} \frac{t^m}{m!} \left(-\partial_t + A\right)^m p_{0,t}(y|x) \right\} dy \\ &= \int f(y)\, \hat{p}_t(y|x)\, dy, \end{aligned} \tag{2.26}$$

where $\hat{p}_t(y|x)$ is the density approximation we arrived at in (2.25). Thus, for simple moment functions, such as the ones appearing in European option prices with constant interest rates, the expansion of Kristensen and Mele (2011) is equivalent to first developing the corresponding expansion for the transition density and then using this to compute the relevant moment; this was already pointed out in Appendix B of Kristensen and Mele (2011). However, in practice, it is easier to directly employ (2.17) with $u_{0,t}$ chosen as the pay-off under the Black–Scholes model since this avoids having to compute the integral $\int f(y)\, \hat{p}_t(y|x)\, dy$ after developing the expansion of the transition density.

Finally, let us consider the proposal of Wan and Yang (2021) for option pricing approximation: They take as starting point that the pay-off can be written as $u_t(x) = \int f(y)\, p_t(y|x)\, dy$, and then replace $p_t(y|x)$ by the approximation given in (2.25) with auxiliary model chosen as Brownian motion with drift. As we just demonstrated in (2.26), this is equivalent to the approximation developed in Kristensen and Mele (2011) when the auxiliary model is chosen as the Black–Scholes model since the log–price in this case follows a Brownian Motion with drift. Thus, the option price approximation of Wan and Yang (2021) is again a special case of Kristensen and Mele (2011).

## 3. Theoretical properties

We first present a general theory of series expansions on the form (2.5) when the function $f$ is regular in the sense that $f \in D(B)$. We provide two sets of results: First, we derive an error bound for any given value $M$ of the order of the expansion. Second, we provide conditions under which the error bound vanishes as $M \to \infty$ at a given value of the time horizon $t > 0$. The conditions for the second set of results come in two forms: We first provide conditions under which the proposed power series expansion converges globally, i.e., over the whole domain of $x_t$. These conditions are somewhat restrictive though and rule out certain models and functions of interest. We therefore proceed to examine how the approximation behaves on a given compact subset of the full domain, and show that the power series expansion is consistent over compact subsets under weak regularity conditions that most known models satisfy. We then apply the theory to moments of jump–diffusions on the form (2.9) and provide primitive conditions under which the expansion is valid. Some of the results presented here rely on the important insights found in the unpublished work of Schaumburg (2004) which we are indebted to.

Next, we then proceed to analyze the "smoothed" expansion (2.7). As in the regular case, we are able to derive an error bound for a given choice of $M$. But at the same time, this expansion is generally not consistent in the sense that it will not converge as $M \to \infty$ for a fixed value of $t > 0$. This is an important result since this shows that the approximation error will eventually blow up as we increase $M$. Thus, researchers should use the generalized version with caution.

### 3.1. Series expansions of "regular" moments

We take as given a semi–group $E_t : \mathcal{F} \mapsto \mathcal{F}$ as described in Section 2. As a first step in our analysis, we show that $u_t(x)$ indeed solves (2.3) if $f \in \mathcal{D}(B)$:

**Theorem 3.1.** *For any $f \in \mathcal{D}(B)$, $u_t(x) = E_t f(x)$ satisfies:*

1. *For any fixed $t \geq 0$, $x \mapsto u_t(x) \in \mathcal{D}(B)$ with $Bu_t(x) = E_t(Bf)(x)$.*
2. *If $t \mapsto E_t(Bf)$ is right-continuous at $t = 0^+$ then $u_t(x)$ solves*

$$\partial_t u_t(x) = Bu_t(x), \qquad t > 0, \qquad u_0(x) = f(x). \tag{3.1}$$

The continuity condition in the second part of the theorem is satisfied under great generality when $E_t$ is on the form (2.10). A sufficient condition is that the mapping $\{x_t : t \geq 0\}$ is Borel measurable w.r.t. the product sigma algebra, cf. p. 771 in Hansen and Scheinkman (1995). The above result, and many subsequent ones, requires the function $f$ defining $u_t(x)$ to satisfy $f \in \mathcal{D}(B)$.

Unfortunately, it rarely easy to give an explicit characterization of $\mathcal{D}(B)$. Instead, we will often work in a smaller subspace, say, $\mathcal{D}_0(B) \subseteq \mathcal{D}(B)$ which is known to us; see Section 3.2 for an example. One says that $\mathcal{D}_0(B)$ is a core of $\mathcal{D}(B)$ if it is a dense subset of the latter.

We recognize (3.1) as a generalized version of the celebrated Kolmogorov's backward equation for jump-diffusion models. In particular, it implies that $\lim_{t\to 0^+} \partial_t u_t(x) = Bf(x)$. More generally, under suitable regularity conditions, $t \mapsto u_t(x)$ will be $M \geq 1$ times differentiable with

$$\lim_{t\to 0^+} \partial_t^m u_t(x) = B^m f(x), \qquad 0 \leq m \leq M, \tag{3.2}$$

in which case the Taylor series approximation $\hat{u}_t$ given in (2.5) is valid. In order for $\hat{u}_t \to u_t$ as $M \to \infty$, we need $t \mapsto u_t$ to be analytic:

**Definition 3.2.** $t \mapsto u_t$ is said to be analytic (at $t = 0^+$) with radius $T_0 > 0$ if it is infinitely differentiable w.r.t. $t$ and satisfies

$$u_t = \lim_{M\to\infty} \hat{u}_t, \quad t \leq T_0, \tag{3.3}$$

where $\hat{u}_t$ given in (2.5).

Recall that the definition of $B$ and the convergence result (3.3) are stated w.r.t. the chosen function norm $\|\cdot\|_{\mathcal{F}}$ introduced earlier. As we shall see, different assumptions regarding the model and the chosen function $f$ defining $u$ motivate different spaces and norms. Ideally, we would like the convergence to take place uniformly over all values of $x \in \mathcal{X}$, but this will only hold for a small set of functions $f$ and models, and so in some applications it is necessary to work with the weaker $L_2$ norm.

In order for $u_t$ to be analytic, we need as a minimum that $u_t$ is infinitely differentiable so that (3.2) holds for all $m \geq 1$. This in turn requires $B^m f(x)$, $m \geq 1$, to be well–defined. That is, $f \in \mathcal{D}(B^m)$, $m \geq 1$, where the domains are defined recursively as

$$\mathcal{D}(B^m) = \left\{ f \in \mathcal{D}\left(B^{m-1}\right) : Bf \in \mathcal{D}(B) \right\} \subseteq \mathcal{D}\left(B^{m-1}\right), \qquad m = 2, 3, ...$$

The following result shows that the Taylor series $\hat{u}_t(x)$ is a valid approximation for any $f \in \mathcal{D}\left(B^{M+1}\right)$ and also provide an error bound for it:

**Theorem 3.3.** *For any $f \in \mathcal{D}\left(B^{M+1}\right)$ and $t \geq 0$, $\hat{u}_t(x)$ in (2.5) satisfies*

$$\begin{aligned} \left|u_t(x) - \hat{u}_t(x)\right| &= \left| \int_0^t \int_0^{t_1} \cdots \int_0^{t_M} B^{M+1} u_{t_{M+1}}(x)\, dt_{M+1} \cdots dt_1 \right| \\ &\leq \frac{t^{M+1}}{(M+1)!} \sup_{0\leq s\leq t} \left| B^{M+1} u_s(x) \right|. \end{aligned}$$

We recognize the error bound as a generalized version of the one that holds for a Taylor series approximation of a $M+1$ times differentiable function. The error bound can be used to show convergence of our expansion of the transition density with $M \geq 1$ fixed as the time distance between observations, corresponding to $t$, shrinks to zero. This is the standard result found in the existing literature on expansions of moments of continuous-time processes as cited in the introduction. But, based on this result alone, the corresponding approximate moment is then only guaranteed to converge towards the exact one when high-frequency data is available. That is, when $t$ shrinks to zero as the number of observations diverge. For a fixed $t$, there is guarantee for the error bound provided in the theorem not blowing up as $M \to \infty$.

We will therefore now derive conditions under which convergence will hold for a given fixed $t > 0$. From Theorem 3.3 we see that the following two conditions are necessary for convergence of $\hat{u}_t(x)$ to hold: $f \in \mathcal{D}(B^\infty)$ and $\left\| \frac{t^m}{m!} B^m f \right\|_{\mathcal{F}} \to 0$ as $m \to \infty$. The second condition will generally not hold for all $t > 0$. Formally, the radius of convergence is given by

$$T_0 = 1 / \limsup_{m\to\infty} \left\{ \|B^m f\|_{\mathcal{F}} / m! \right\}^{1/m}. \tag{3.4}$$

Often the exact value of $T_0$ cannot be derived, but it may still be possible to identify a lower bound for it. Similarly, it is in many applications difficult to provide a precise characterization of $\mathcal{D}(B^\infty) = \bigcap_{m=1}^\infty \mathcal{D}(B^m)$. One partial characterization is that it constitutes a core of $\mathcal{D}(B)$, cf. Theorem 7.4.1 of Davies (2007), so that most functions in $\mathcal{D}(B)$ also belong to $\mathcal{D}(B^\infty)$. But this provides no guarantees for that a given function in $\mathcal{D}(B)$ belongs to $\mathcal{D}(B^\infty)$.

Instead one may seek to identify a subset $\mathcal{F}_0 \subseteq \mathcal{F}$ so that (i) $\mathcal{F}_0 \subseteq \mathcal{D}(B)$ and (ii) the image $B\left(\mathcal{F}_0\right) = \left\{ Bf | f \in \mathcal{F}_0 \right\} \subseteq \mathcal{F}_0$. For a given $f \in \mathcal{F}_0$, part (i) ensures that $Bf$ is well-defined while part (ii) implies that $Bf \in \mathcal{F}_0$. In particular, (i)–(ii) guarantee that $\mathcal{F}_0 \subseteq \mathcal{D}(B^m)$ for all $m \geq 1$. As a consequence, $\mathcal{F}_0 \subseteq \mathcal{D}(B^\infty)$ thereby providing us with a partial characterization of $\mathcal{D}(B^\infty)$. In particular, for any given $f \in \mathcal{F}_0$, we have that $t \mapsto u_t$ is infinitely differentiable. The following theorem states the formal result of the above analysis:

**Theorem 3.4.** *Suppose that $f \in \mathcal{D}(B^\infty)$. Then $u_t$ is infinitely differentiable and, with the radius of convergence $T_0 \geq 0$ given in (3.4),*

$$\forall t \leq T_0 : \left\|u_t - \hat{u}_t\right\|_{\mathcal{F}} \leq \frac{\left(t/T_0\right)^{M+1}}{1 - t/T_0} \to 0 \text{ as } M \to \infty.$$

*The domain $\mathcal{D}(B^\infty)$ is a core of $\mathcal{D}(B)$. A sufficient condition for $f \in \mathcal{D}(B^\infty)$, is that $f \in \mathcal{F}_0$ for some $\mathcal{F}_0 \subseteq \mathcal{D}(B)$ satisfying $B\left(\mathcal{F}_0\right) \subseteq \mathcal{F}_0$.*

The last part of the theorem provides one sufficient condition for $u_t$ to be analytic. There are two tensions when seeking such a suitable set $\mathcal{F}_0$: First, we would like to choose $\mathcal{F}_0$ as large as possible in order to guarantee convergence of $\hat{u}_t$ over a large set of functions. But at the same time we need to restrict $\mathcal{F}_0$ so that it satisfies $B\left(\mathcal{F}_0\right) \subseteq \mathcal{F}_0$. Second, to ensure a strong convergence result, we would like to choose the norm $\|\cdot\|_{\mathcal{F}}$ as "strong" as possible, e.g., as the sup norm. But establishing $T_0 > 0$ then proves more difficult.

One way of designing the function class $\mathcal{F}_0$ is to build it from the so–called eigenfunctions of $B$. Eigenfunctions are defined in terms of the so–called spectrum of $B$,

$$\sigma(B) = \{\xi \in \mathbb{C} : (\xi I - B) \text{ is not a bijection}\} \subset \{\xi \in \mathbb{C} : \mathrm{Re}(\xi) < 0\} \cup \{0\}.$$

In particular, for any given eigenvalue $\xi \in \sigma(B)$ there exists a corresponding eigenfunction $\phi \in \mathcal{D}(B)$ so that $(\xi I - B)\phi = 0 \Leftrightarrow B\phi = \xi\phi$. This in turn implies that $\phi \in \mathcal{D}(B^\infty)$ with $B^m\phi = \xi^m\phi$. Thus,

$$E_t\phi(x) = e^{\xi t}\phi(x) = \sum_{m=0}^{\infty} \frac{t^m}{m!} B^m\phi(x),$$

which is clearly analytic and so our power series expansion will converge for any eigenfunction. The following corollary shows that in principle $\mathcal{F}_0$ can be chosen as the span of any given countable set of eigenfunctions:

**Corollary 3.5.** *For any given sequence $\left\{\left(\xi_i, \phi_i\right)\right\}_{i=1}^{\infty}$ of eigenpairs of $B$,*

$$\mathcal{F}_0 = \left\{ f = \sum_{i=1}^{\infty} \alpha_i\phi_i : \sum_{i=1}^{\infty} |\alpha_i| < \infty \right\} \subseteq \mathcal{D}(B^\infty).$$

This particular choice of $\mathcal{F}_0$ is in some cases somewhat restrictive in the sense that it may be only a small subset of $\mathcal{D}(B^\infty)$. However, in the special case of a given semi–group's spectrum being countable, we generally have that $\mathcal{F}_0 = \mathcal{D}(B^\infty)$. One example of this is so–called time reversible Markov processes whose spectra are countable with the corresponding eigenfunctions forming an orthnormal basis of $\mathcal{F}$; see, e.g., Hansen et al. (1998). But many Markov processes are irreversible and have an uncountable spectrum in which case $\mathcal{F}_0$ is a proper subset of $\mathcal{D}(B^\infty)$.

The corollary does not guarantee that for any $f \in \mathcal{F}_0$ the corresponding $u_t(x)$ is analytic – only that it is infinitely differentiable. To see the complications of ensuring analyticity, observe that, for any given $f \in \mathcal{F}_0$ with $\mathcal{F}_0$ defined above, $B^m f = \sum_{i=1}^{\infty} \alpha_i \xi_i^m \phi_i$, $m \geq 1$, so that

$$\hat{u}_t = \sum_{m=0}^{M} \frac{t^m}{m!} B^m f = \sum_{i=1}^{\infty} \left( \sum_{m=0}^{M} \frac{\left(\xi_i t\right)^m}{m!} \right) \alpha_i\phi_i.$$

Thus,

$$\left\|u_t - \hat{u}_t\right\|_{\mathcal{F}} \leq \sup_{i \geq 1} \left| \sum_{m=0}^{M} \frac{\left(\xi_i t\right)^m}{m!} - e^{-\xi_i t} \right| \sum_{i=1}^{\infty} |\alpha_i| \left\|\phi_i\right\|_{\mathcal{F}},$$

and so we need at a minimum $\sup_{i \geq 1} \left| \sum_{m=0}^{M} \frac{\left(\xi_i t\right)^m}{m!} - e^{-\xi_i t} \right| \to 0$, $M \to \infty$. But this convergence result will generally not hold; for example, if $\xi_i \in \mathbb{R}$ and $\xi_i \to \infty$ as $i \to \infty$ then convergence will fail.

In conclusion, to ensure convergence, we need to impose restrictions on the eigenvalues/the spectrum. We will now present such a set of conditions. These will involve the so–called resolvent of the generator defined as

$$R(\xi) = (\xi I - B)^{-1}, \ \xi \notin \sigma(B).$$

**Theorem 3.6.** *$t \mapsto E_t f(x)$ is analytic for all $t > 0$ and all functions $f \in \mathcal{F}$ if and only if the following two conditions are satisfied: There exists $0 < \delta < \pi/2$ and $C_B < \infty$ so that*

$$\sigma(B) \subseteq \bar{\sigma}_\delta := \{\xi \in \mathbb{C} : |\arg(\xi)| > \pi/2 + \delta\}, \tag{3.5}$$

*and, for all $\varepsilon > 0$,*

$$\|R(\xi)\|_{\mathrm{op}} \leq \frac{C_B}{|\xi|} \text{ for } \xi \in \mathbb{C} \backslash \bar{\sigma}_\delta. \tag{3.6}$$

*In particular, if $f \in E_{\tau_0}(\mathcal{F})$ for some $\tau_0 > 0$ (that is, $f(x) = E_{\tau_0} g(x)$ for some $g \in \mathcal{F}$) then*

$$\forall t < \frac{\tau_0}{C_B e} : \left\| u_t - \hat{u}_t \right\|_{\mathcal{F}} \to 0, \qquad M \to \infty.$$

The first part of the theorem states necessary and sufficient conditions for $E_t f(x)$ to be analytic at any given $t > 0$ and for *any* $f \in \mathcal{F}$. The conditions (3.5)–(3.6) ensure that the spectrum of $B$ is such that the convergence problem discussed before the theorem does not occur. This is a strong result but at the same time (3.5)–(3.6) are rather strong conditions. Moreover, they tend to be difficult to verify in practice since this requires knowledge of the spectrum $\sigma(B)$. Primitive sufficient conditions for them to hold are provided in the next section. Both the conditions and the results are relative to the chosen function space and norm $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$. By choosing $\mathcal{F}$ suitably small, we expect that (3.5)–(3.6) will hold in great generality. We give an example of this in Section 3.3.

The second part then shows that for the subclass of functions $f$ that satisfy $f(x) = E_{\tau_0} g(x)$, for some $\tau_0 > 0$ and $g$, analyticity of $E_t f(x)$ extends to $t = 0$. This part follows as a direct consequence of the first part since this implies that $E_t f(x) = E_{t+\tau_0} g(x)$ is analytic at $t = 0$. The lower bound of the radius of convergence $T_0$ depends on the degree of smoothness of $f$, as measured by $\tau_0$, and the properties of the model, specifically the bound $C_A$ on its resolvent.

The requirement $f \in E_{\tau_0}(\mathcal{F})$ is difficult to verify in a given application. In the leading case of $E_t f(x) = \mathbb{E}\left[f\left(x_t\right) | x_0 = x\right]$, the condition amounts to showing that there exists a solution $g(x)$ to the following integral equation $f(x) = \int g(y) p_{\tau_0}(y|x)\, dy$ for some $\tau_0 > 0$, assuming that $x_t$ has a transition density $p_t(y|x)$. This is a so–called Fredholm equation of the first kind; conditions for a solution to this to exist are available but not easily verified in a given application. However, it can be shown that, for any given $\tau_0$, $E_{\tau_0}(\mathcal{F})$ is dense in $\mathcal{F}$, see, e.g., Theorem 7.4.4 in Davies (2007), and so the result will hold for "almost every" $f \in \mathcal{F}$.

### 3.2. Application to jump-diffusions

We now apply the general theory to our jump–diffusion model. In the following, let $x_t$ be a weak solution to (2.9) for a given specification of $\left(\mu, \sigma^2, \lambda, \nu\right)$ with generator $A$ given in (2.11)–(2.13) and $E_t f(x) = \mathbb{E}\left[f\left(x_t\right) | x_0 = x\right]$. Two standard choices of the function space $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$ are the following: The first is the space of bounded functions equipped with a sup norm, $\|f\|_{\mathcal{F}} = \sup_{x \in \mathcal{X}} |f(x)|$. The second is the space of functions with second moments equipped with the following $L_2$ norm, $\|f\|_{\mathcal{F}}^2 = \int_{\mathcal{X}}^{\infty} f^2(x) \pi(x)\, dx$ for some weighting function $\pi(x)$. In case of $x_t$ being stationary, a natural choice for $\pi$ is the stationary marginal distribution in which case $\|f\|_{\mathcal{F}}^2 = \mathbb{E}\left[f^2\left(x_t\right)\right]$; this norm was, for example, used by Hansen and Scheinkman (1995).

We first need to get a handle on the generator of the process and its domain $\mathcal{D}(A)$. A complete characterization of $\mathcal{D}(A)$ is unfortunately not possible and we will instead only work with a subset of $\mathcal{D}(A)$ where the generator takes the form (2.11). Let $C^m = C^m(\mathcal{X})$ denote the space of functions $f(x)$ with domain $\mathcal{X}$ that are $m \geq 0$ times continuously differentiable w.r.t. $x$. If $f \in C^2$ then Ito's Lemma for jump–diffusions (see, e.g., Cont and Tankov (2003), Proposition 8.14) yields

$$\begin{aligned} f\left(x_t\right) = f\left(x_0\right) &+ \int_0^t A_D f\left(x_s\right) ds + \sum_{i : 0 \leq \tau_i \leq t} \left[ f\left(x_{\tau_i^-} + \Delta x_i\right) - f\left(x_{\tau_i^-}\right) \right] \\ &+ \sum_{i=1}^{d} \int_0^t \frac{\partial f\left(x_s\right)}{\partial x_i} \sigma_i\left(x_s\right) dW_s, \end{aligned}$$

where $A_D$ is defined in (2.12), $\sigma_i(x) = \left[\sigma_{i1}(x), ..., \sigma_{id}(x)\right]$ while $\tau_i$ and $\Delta x_i$ denote the time and the size, respectively, of the $i$th jump. Assuming $E_t |f|(x) < \infty$ and $E_t (\frac{\partial f}{\partial x_i} \sigma_i)^2(x) < \infty$, $i = 1, ..., d$, we can take conditional expectations w.r.t. the natural filtration on both sides of the above to obtain (2.2) with $A$ given in (2.11). Thus, the following is a subset of the domain of the generator,

$$\mathcal{D}_0(A) := \left\{ f \in C^2 : E_t |f| \text{ and } E_t \left\| \frac{\partial f}{\partial x} \sigma \right\|^2 \text{ exist for all } t > 0 \right\} \subseteq \mathcal{D}(A).$$

In the following we will only consider functions situated in $\mathcal{D}_0(A)$ and so not distinguish between the general generator and the one restricted to $\mathcal{D}_0(A)$. Under the assumption that $\mu$, $\sigma^2$ and $\lambda$ and $f$ all belong to $C^{2m}$, we can apply Ito's Lemma repeatedly and it follows straightforwardly that

$$\mathcal{D}_0(A^m) := \left\{ f \in C^{2m} : E_t \left| A^k f \right| \text{ and } E_t \left\| \frac{\partial A^k f}{\partial x} \sigma \right\|^2 \text{ exist for all} t > 0,\ 0 \leq k \leq m-1 \right\} \subseteq \mathcal{D}(A^m).$$

Implicit in this definition is the requirement that $\int_{\mathbb{R}^d} \left| A^k f(x + c) \right| \nu_t(c)\, dc < \infty$ for $k = 0, ..., m$. Thus, a given $f \in C^{2m}$ belongs to $\mathcal{D}_0(A^m)$ if relevant moments w.r.t. the jump measure $\nu$ and the probability measure of $x_t$ exist. For example, if $f$ and all its derivatives are bounded, $\mu$, $\sigma^2$ and $\lambda$ and all their derivatives are bounded by some function $V(x) \geq 0$ with $\mathbb{E}\left[V\left(x_t\right)\right] < \infty$, and $\nu$ has bounded support then $f \in \mathcal{D}_0(A^\infty)$. Similarly, if $f$ is a polynomial of order $q$, $\mu$, $\sigma^2$ and $\lambda$ are linear w.r.t. $x$, $\nu_t$ has all polynomial moments, and $\mathbb{E}\left[\left\|x_t\right\|^q\right] < \infty$, $0 \leq t \leq T$, then $f \in \mathcal{D}(A^\infty)$.

Since $f \in \mathcal{D}_0(A^\infty)$ is necessary for our expansion to work, we will maintain the following assumption on the model:

**A.1** (i) $\mu$, $\sigma^2$ and $\lambda$ belong to $C^\infty$ and (ii) $\sup_x \lambda(x) < \infty$.

Part (i) ensures that, under suitable moment conditions as described above, if $f \in C^\infty$ then $f \in \mathcal{D}_0(A^\infty)$. Part (ii) is imposed to simplify subsequent arguments since it entails the following result (see Pazy, 1983, Theorem 3.2.1):

**Lemma 3.1.** *Suppose that $A_J$ is a bounded operator. If $A_D$ generates an analytic semi–group then $A_D + A_J$ also generates an analytic semi–group. Under A.1(ii), $A_J$ is a bounded operator.*

Thus, for a given jump–diffusion model satisfying A.1(ii), or any other conditions ensuring $A_J$ is bounded, we only need to ensure that the diffusive component is analytic. In the following, we will implicitly assume that indeed $A_J$ is bounded and derive conditions under which $E_t$ for pure diffusion processes ($A_J = 0$) is analytic.

Ideally we would now provide primitive conditions for general jump–diffusion processes to satisfy the high-level conditions found in the theorems and corollaries stated in the previous section. This is unfortunately not possible since the spectral properties of jump–diffusions are still not fully understood. We will therefore only state results for special cases for which results do exist. At the same time we would like to emphasise that we expect the results to hold more broadly.

We first develop conditions under which polynomial moment functions are analytic. We start out with a few definitions: For a given multi-index $\alpha = (\alpha_1, ...\alpha_d) \in \mathbb{N}_0^d$ and $x = (x_1, ..., x_d)' \in \mathbb{R}^d$ let $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $(x)^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. We then let

$$\mathcal{P}_k = \left\{ p(x) = \sum_{|\alpha| \le k} c_\alpha (x)^\alpha : \alpha \in \mathbb{N}_0^d, c_\alpha \in \mathbb{R} \right\}$$

denote the family of polynomials of order $k$ and $\mathcal{P}_{k|\mathcal{X}}$ be these polynomials restricted to the domain of $x_t$. Observe here that $\mathcal{P}_{k|\mathcal{X}}$ is a finite-dimensional function space. In particular, we can choose a set of basis functions $e = (e_1, ..., e_N) \in \mathcal{P}_{k|\mathcal{X}}$, where $N = \dim \mathcal{P}_{k|\mathcal{X}}$, so for any $p \in \mathcal{P}_{k|\mathcal{X}}$ there exists $c = (c_1, ...c_N)$ so that

$$p(x) = \sum_{i=1}^N c_i e(x) = c' e(x).$$

If $\mathcal{P}_{k|\mathcal{X}}$ satisfies the two conditions of Theorem 3.4 then analyticity follows automatically from the fact that when we restrict the domain of $A$ to $\mathcal{P}_{k|\mathcal{X}}$ then it becomes a finite–dimensional operator and therefore bounded:

**Corollary 3.7.** *Suppose that $x_t$ is a polynomial process in the sense that, for all $k \ge 1$, $\mathcal{P}_{k|\mathcal{X}} \subseteq \mathcal{D}(A)$ and $A(\mathcal{P}_{k|\mathcal{X}}) \subseteq \mathcal{P}_{k|\mathcal{X}}$. Then, for any $k \ge 1$ and any $p = c'e \in \mathcal{P}_{k|\mathcal{X}}$, $u_t(x) = E_t p(x)$ is analytic with radius $+\infty$ and satisfies for all $x \in \mathcal{X}$,*

$$u_t(x) = c' \exp(t\bar{A})\, e(x) = c' \sum_{m=0}^\infty \frac{t^m}{m!} \bar{A}^m e(x),$$

*where $\bar{A} = [\bar{a}_{ij}]_{1 \le i,j \le N} \in \mathbb{R}^{N \times N}$ is defined as the solution to*

$$Ae_i = \sum_{i=1}^N \bar{a}_{ij} e_j, \qquad i = 1, ..., N.$$

*A sufficient condition for $x_t$ to be polynomial process is that $\mu \in \mathcal{P}_{1|\mathcal{X}}$, $\sigma^2 \in \mathcal{P}_{2|\mathcal{X}}$ and $\lambda \in \mathcal{P}_{2|\mathcal{X}}$.*

The second result provides primitive conditions for the high–level assumptions (3.5)–(3.6) to hold in the context of jump diffusions, where $E_t^*$ and $A^*$ denotes the so–called adjoint operators of $E_t$ and $A$, respectively:

**Corollary 3.8.** *Suppose that $E_t f(x) = \mathbb{E}[f(x_t) | x_0 = x]$ and that $\mathcal{F}$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ so that $\|f\|_{\mathcal{F}}^2 = \langle f, f \rangle$. Suppose furthermore that $x_t$ is a time–reversible Markov process in the sense that its generator is self–adjoint, $A = A^*$ (or, equivalently, $E_t = E_t^*$). Then (3.5)–(3.6) are satisfied and so $t \mapsto E_t f(x)$ is analytic for all $t > 0$.*

*Suppose that $x_t$ is a stationary diffusion process which satisfies the conditions given in either Example 1, 2 or 3 in Hansen and Scheinkman (1995). Then $x_t$ is time–reversible.*

The time–reversibility condition implies that $A$'s spectrum is discrete and contained in the negative half–line which suffices for (3.5)–(3.6) to hold. The three examples referred to in the second part of the last theorem are time–homogenous scalar diffusions, multivariate factor diffusion models, and a restricted class of multivariate diffusions; see Hansen and Scheinkman (1995) for the precise details.

Note here that the corollary imposes no smoothness conditions on $\mu$, $\sigma^2$ and $f$. This is because that $Af$ may still be well–defined even without smoothness, cf. above discussion of $\mathcal{D}(A)$. However, its particular form in these cases is generally unknown to us. Thus, in order to compute $Af$ in practice we restrict ourselves to smooth models, as in Assumption A.1(i), and smooth choices of $f$, as in $\mathcal{D}_0(A)$.

Our third result again uses Theorem 3.4 but focuses on a different class of "test functions" to obtain results for such models. We restrict the function set to

$$\mathcal{F}_0 = \left\{ f \in C^\infty : \sup_{|\alpha|\geq 0} \left\| \partial_x^\alpha f \right\|_{\mathcal{F}} < \infty \right\}, \tag{3.7}$$

where $\partial_x^\alpha f = \partial^\alpha f / (\partial x^\alpha)$, which we equip with the norm $\|f\|_{\mathcal{F}_0} = \sup_{|\alpha|\geq 0} \left\| \partial_x^\alpha f \right\|_{\mathcal{F}}$. Importantly, if $f \in \mathcal{F}_0$ then, for any $\alpha \in \mathbb{N}_0^\infty$, $\partial_x^\alpha f \in \mathcal{F}_0$ with $\left\| \partial_x^\alpha f \right\|_{\mathcal{F}_0} \leq \|f\|_{\mathcal{F}_0}$. This property of $\mathcal{F}_0$ ensures that if $\mu$ and $\sigma$ in $\mathcal{F}_0$ then $Af \in \mathcal{F}_0$ for all $f \in \mathcal{F}_0$ and so $\mathcal{F}_0 \subseteq \mathcal{D}(A^\infty)$. Moreover, the generator, when restricted to $\mathcal{F}_0$, is bounded and so the radius of convergence is infinite:

**Theorem 3.9.** *Suppose that $\mu$ and $\sigma^2$ lie in $(\mathcal{F}_0, \|\cdot\|_{\mathcal{F}_0})$ defined in (3.7). Then $\mathcal{F}_0 \subseteq \mathcal{D}(A)$ and $A : \mathcal{F}_0 \mapsto \mathcal{F}_0$ is a bounded operator. In particular, there exists $\bar{A} < \infty$ so that for any $f \in \mathcal{F}_0$ and any $t > 0$,*

$$\left\| u_t - \hat{u}_t \right\|_{\mathcal{F}_0} \leq \frac{\left(t\bar{A}\right)^{M+1}}{(M+1)!} \|f\|_{\mathcal{F}_0} \to 0 \text{ as } M \to \infty.$$

Note here that convergence holds for all $t > 0$ and that the convergence rate is super-geometric. Moreover, the result allows for a broad class of non-linear multivariate diffusion models. On the other hand, it rules out unbounded drift and diffusion terms.

### 3.3. Convergence over bounded sets

The above results are strong in the sense that they guarantee convergence w.r.t. a function norm over the full state space $\mathcal{X}$. But at the same time they are restrictive in that they do not apply to general multivariate jump–diffusion models. One way of allowing for a broader class of models and functions is to restrict attention to solutions defined on a bounded subset of $\mathcal{X}$ leading to the following class of so–called localized Cauchy problems. We here focus on the case of pure diffusions since for this class of models results exist on analytic solutions on bounded sets.

Let $\mathcal{X}_0 \subseteq \mathcal{X}$ be a bounded open set and let $u_t^*(x)$ be a function chosen by the researcher which satisfies $u_0^*(x) = f(x)$. We then consider the following "trimmed" version of the Cauchy problem for diffusion models:

$$\partial_t w_t(x) = A_D w_t(x) \text{ for } (t,x) \in (0,\infty) \times \mathcal{X}_0, \tag{3.8}$$

$$w_t(x) = u_t^*(x) \text{ for } (t,x) \in (0,\infty) \times \mathcal{X} \backslash \mathcal{X}_0, \tag{3.9}$$

with initial condition $w_0(x) = f(x)$ for $x \in \mathcal{X}_0$. We now only require the solution $w_t(x)$ to solve the Cauchy problem on a bounded open subset $\mathcal{X}_0$ of the full domain $\mathcal{X}$ and then pin down its behaviour outside of $\mathcal{X}_0$ through the pre-specified function $u^*$. The class of problems on the form (3.8)–(3.9) can be described by a semi–group $E_t$ so that $w_t = E_t f$. By choosing $\mathcal{X}_0$ as a bounded set, the requirements for this semi–group to be analytic become a lot less restrictive and essentially amount to $\mu$ and $\sigma^2$ being sufficiently smooth; see, e.g., Chapter 3 in Lunardi (1995). The following theorem states the precise conditions:

**Theorem 3.10.** *Suppose that $\mu(x)$, $\sigma^2(x)$ and $u_t^*(x)$ are analytic functions so that, for some $0 < C_0, C_1 < \infty$,*

$$\left\| \partial_x^\alpha \mu(x) \right\| \leq C_0 C_1^{-1-|\alpha|} |\alpha|!, \qquad \left\| \partial_x^\alpha \sigma^2(x) \right\| \leq C_0 C_1^{-1-|\alpha|} |\alpha|!, \qquad x \in \mathcal{X}_0; \tag{3.10}$$

*and, for some $c > 0$ and for all $x, y \in \mathcal{X}_0$, $y' \sigma^2(x) y \geq c \|y\|$. Then the solution $w_t : \mathcal{X}_0 \mapsto \mathbb{R}$ to (3.8)–(3.9) is analytic at any $t > 0$ w.r.t. the uniform norm, $\left\| u_t \right\|_{\mathcal{F},0} = \sup_{x \in \mathcal{X}_0} \left| u_t(x) \right|$.*

*Suppose furthermore that $f(x) = E_{\tau_0} g(x)$ for some continuous function $g$. Then, $w_t : \mathcal{X}_0 \mapsto \mathbb{R}$ is analytic at $t = 0$ with radius of convergence $T_0 > 1/\left(\rho \tau_0\right)$ for some $\rho = \rho(B,d) \in (0,1]$.*

This provides simple and relatively weak conditions under which a series expansion of $w_t$ will converge. But will such series expansion be a good approximation to $u_t$? By eq. (3.8) together with the initial condition, we see that

$$\partial_t^m w_t(x)\big|_{t=0^+} = A_D^m f(x), \qquad x \in \mathcal{X}_0, \quad m \geq 0.$$

Thus, under the conditions of the theorem, the power series approximation of $w_t$ shares derivatives with $u_t$ on $\mathcal{X}_0$. At the same time, the solution $w_t$ will generally differ from the global solution $u_t$. However, if we restrict $f \in D^\infty\left(A_D\right)$ then $\partial_t^m u_t(x)\big|_{t=0^+} = A_D^m f(x)$ and so $w_t(x) = u_t(x)$, $x \in \mathcal{X}_0$, and the power series will be consistent on $\mathcal{X}_0$. In particular, if we can show that $w_t(x)$ is analytic on $\mathcal{X}_0$ then the same will hold for $u_t(x)$ when considered as a function with domain $\mathcal{X}_0$. This result combined with Lemma 3.1 shows that our power series expansions converges for a very broad class of diffusion models over bounded subsets of their domains.

### 3.4. Expansion of "irregular" moments

Finally, we provide an analysis of smoothed expansions on the form (2.7). First, by following the same arguments as in Theorem 3.3, it is easily shown that if $u_{0,s}$ satisfies Assumption A.0 then $\hat{u}_t(x)$ given in (2.7) with $M_1 + M_2 = M$ satisfies

$$\left|u_t(x)-\hat{u}_t(x)\right| \leq \sum_{m_1+m_2=M} \frac{(-s)^{m_1} t^{m_2}}{m_1! m_2} \sup_{0\leq s,\tau\leq t} \left|E_\tau\left(\left(\partial_s\right)^{m_1} B^{m_2} u_{0,s}\right)(x)\right| = O\left(s^M\right)+O\left(t^M\right).$$

One could now hope for that as long as $u_{0,s}(x)$ is sufficiently regular then the expansion would converge under conditions similar to the ones in the "regular" case analyzed in the previous section as $M\to\infty$. This is unfortunately not the case. To see this, observe that the expansion (2.7) proceeds in two steps: First, we approximate $f$ by $\hat{f}(x)=\sum_{m_1=0}^{M_1}\frac{(-s)^{m_1}}{m_1!}\partial_s^{m_1}u_{0,s}(x)$; second, we plug $\hat{f}(x)$ into (2.5) to obtain $\hat{u}_t(x)=\sum_{m_2=0}^{M_2}\frac{t^{m_2}}{m_2}B^{m_2}\hat{f}(x)$. Thus, in order for the resulting expansion to converge, we need at a minimum that (i) $\hat{f}(x)\to f(x)$ and (ii) $\hat{f}(x)\in\mathcal{D}\left(B^{M_2}\right)$. Considering (i), note that

$$\begin{aligned} f(x) &= \sum_{m_1=0}^{\infty}\frac{(-s)^{m_1}}{m_1!}\partial_s^{m_1}u_{0,s}(x) = \sum_{m_1=0}^{M_1}\frac{(-s)^{m_1}}{m_1!}\partial_s^{m_1}u_{0,s}(x)+\sum_{m_1=M+1}^{\infty}\frac{(-s)^{m_1}}{m_1!}\partial_s^{m}u_{0,s}(x) \\ &=: \hat{f}(x)+\hat{e}(x), \end{aligned}$$

where $\hat{e}(x)$ is the approximation error from using $\hat{f}(x)$ in place of $f(x)$. If $f$ is irregular in the sense that $f\notin\mathcal{D}\left(B^{M_2}\right)$ for all $M_2$ large enough while at the same time $\hat{f}(x)\in\mathcal{D}\left(B^{M_2}\right)$ then obviously $\hat{e}(x)\notin\mathcal{D}\left(B^{M_2}\right)$. Thus, as $M_1$ grows, we must eventually have $\hat{f}\notin\mathcal{D}\left(B^{M_2}\right)$ in which case $\hat{u}_t(x)=\sum_{m_2=0}^{M_2}\frac{t^{m_2}}{m_2}B^{m_2}\hat{f}(x)$ is not well-defined. In practice, we expect $\hat{u}_t(x)$ in (2.7) to become numerically unstable as $M_1,M_2\to\infty$. That is, the numerical error $\hat{e}(x)$ will start blowing up.

This demonstrates that the proposed series expansions of irregular functions such as densities and option prices should be used with caution: As more terms are added to the expansions, they will most eventually become numerically unstable and produce unreliable estimates. However, as we shall see in the next section, the expansions still work well when a reasonably small number of terms are used.

## 4. Implementation of expansion for jump-diffusion models

This section provides details regarding the practical implementation of the proposed approximation in the jump–diffusion case. We here focus on the special case of $r(x)=0$ and $s=t$, in which case $u_t(x)=E_t f(x)=\mathbb{E}\left[f\left(x_t\right)|x_0=x\right]$ and

$$\hat{u}_t(x)=\sum_{m=0}^{M}\frac{t^m}{m!}\left(A-\partial_t\right)^m u_{0,t}(x). \tag{4.1}$$

This is done to avoid overly complicated notation. Most of the ideas and arguments extend to the general case.

### 4.1. Choice of smoothing function for irregular moments

Following Kristensen and Mele (2011), a simple choice of $u_{0,s}(x)$ that satisfies A.0 is $u_{0,s}(x)=E_{0,s}f(x)=\mathbb{E}\left[f\left(x_{0,s}\right)|x_{0,0}=x\right]$ where $x_{0,s}$ is chosen as the solution to an auxiliary jump–diffusion model,

$$dx_{0,t}=\mu_0\left(x_{0,t}\right)dt+\sigma_0\left(x_{0,t}\right)dW_t+J_{0,t}dN_{0,t}, \tag{4.2}$$

where $N_{0,t}$ is a Poisson process with jump intensity $\lambda_0(x)$ and $J_{0,t}$ has density $\nu_0(\cdot|x)$. The auxiliary model should be chosen so that $u_{0,t}(x)$ is available on closed form. One such model is the multivariate Brownian motion with drift model,

$$dx_{0,t}=\mu_0 dt+\sigma_0 dW_t, \tag{4.3}$$

where $\mu_0\in\mathbb{R}^d$ and $\sigma_0\in\mathbb{R}^{d\times d}$ are constants, or the multivariate Vasicek (Ornstein–Uhlenbeck) model,

$$dx_{0,t}=\left(\mu_0+Ax_{0t}\right)dt+\sigma_0 dW_t,$$

both of which have a Gaussian transition density on known form. In either case,

$$u_{0,s}(x)=\int f(y)\,p_{0,s}(y|x)\,dy,$$

where $p_{0,s}(y|x)$ is the transition density of the auxiliary model. For example, in the case of (4.3),

$$p_{0,s}(y|x)=\frac{1}{\sqrt{2\pi s\left|\sigma_0^2\right|}}\exp\left(-\frac{\left(y-x-s\mu_0\right)'\sigma_0^{-2}\left(y-x-s\mu_0\right)}{2s}\right). \tag{4.4}$$

This choice satisfies of $u_{0,s}$ satisfies condition (i) of Assumption A.0; since $p_{0,t}(y|x)$ is differentiable w.r.t. $t$ of any order then the mapping $s\mapsto u_{0,s}(x)$ has the same property. The final requirement, $u_{0,s}\in\mathcal{D}\left(A^{M_2}\right)$, has to be checked on a case by case basis.

Recall the two motivating examples of transition density and option price approximation. In the case of $f(x) = \delta(y - x)$, we get $u_{0,s}(x) = p_{0,s}(y|x)$. If $f(x) = \left(\exp\left(x_1\right) - K\right)^+$, and we set $\mu_{0,1} = r - \sigma_{0,11}^2/2$ to ensure risk–neutrality in the auxiliary model, then $u_{0,s}(x)$ takes the form of the well-known formula for the risk–neutral expected pay-off of a call option in the Black–Scholes model,

$$u_{0,s}(x) = se^{rs}\Phi\left(d_+(x,s)\right) - K\Phi\left(d_-(x,s)\right), \tag{4.5}$$

where $d_\pm(x,s) = \left(x - \log(K) + \left(r \pm \frac{1}{2}\sigma_{0,11}^2\right)s\right) / \left(\sigma_{0,11}\sqrt{s}\right)$ and $\Phi(\cdot)$ denotes the cdf of the $N(0,1)$ distribution.

### 4.2. Pure diffusion case

In the pure diffusion case, where no jump component is present so that $A_J = 0$, analytical expressions of $\left(A_D - \partial_t\right)^m u_{0,t}(x)$ are in principal straightforward to obtain relying on symbolic software packages, such as Mathematica, since $A_D$ is a differential operator. We refer to Kristensen and Mele (2011), Yang et al. (2019) and Wan and Yang (2021) for more details on this for the two leading examples of density and option price approximations and with $u_{0,t}$ chosen as the corresponding solution under (4.3).

### 4.3. Jump-diffusion case

#### 4.3.1. State–independent jump or diffusion component

Next, consider jump–diffusion models where either the diffusive component or the jump component of $x_t$ are state–independent; the latter case corresponds to the class of jump–diffusions considered in Wan and Yang (2021).

These two cases correspond to (i) $\mu(x) = \mu$ and $\sigma^2(x) = \sigma^2$ are constant or (ii) $\lambda(x) = \lambda$ and $\nu(\cdot|x) = \nu(\cdot)$ are independent of $x$, respectively. In either case, we can write $x_t = x_{D,t} + x_{J,t}$ where the diffusive component, $x_{D,t}$, and the jump component, $x_{J,t}$, are now mutually independent. As a consequence, the two generators $A_D$ and $A_J$ commute, $A_D A_J = A_J A_D$, in which case

$$u_t(x) = e^{(A_D + A_J)t} f(x) = e^{A_D t} B_{J,t}(x) = e^{A_J t} B_{D,t}(x), \tag{4.6}$$

where

$$B_{J,t}(x) = \mathbb{E}\left[f\left(x_{J,t}\right)\middle| x_{J,0} = x\right], \qquad B_{D,t}(x) = \mathbb{E}\left[f\left(x_{D,t}\right)\middle| x_{D,0} = x\right].$$

Now, consider first the case where (ii) is satisfied. In this scenario, $x_{J,t}|x_{J,0} = x$ has density

$$p_{J,t}(y|x) = \sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!}\nu_k(y - x), \tag{4.7}$$

where $\nu_k(y)$ is the density of the sum of $k$ independent jumps, $\sum_{i=1}^k J_i$, $J_i \sim \nu(\cdot)$. Since $p_{J,t}(y|x)$ is a smooth function then $x \mapsto B_{J,t}\left(\Delta x_{D,T} + x\right)$ is also a smooth function even if $f(x)$ is irregular. Thus, if $B_{J,t}(x)$ is available on closed form then the smoothing device is not needed and we can approximate $u_t$ by

$$\hat{u}_t(x) = \sum_{m=0}^{M}\frac{t^m}{m!}A_D^m B_{J,t}(x). \tag{4.8}$$

Similar, if (i) is satisfied then $x_{D,t}|x_{D,0}$ is a Brownian motion with drift and has Gaussian density as given in (4.4). Because of its simple dynamics, $B_{D,t}(x)$ is available on closed form in many cases and will again be a smooth function; if so, we propose to approximate $u_t$ by

$$\hat{u}_t(x) = \sum_{m=0}^{M}\frac{t^m}{m!}A_J^m B_{D,t}(x).$$

If closed form expressions of neither $B_{D,t}$ nor $B_{J,t}$ are available, it is still possible to simplify the computation using, for example,

$$\hat{u}_t(x) = \left[\sum_{m=0}^{M}\frac{t^m}{m!}e^{A_J t}\left(A_D - \partial_t\right)^m u_{0,t}(x)\right], \tag{4.9}$$

assuming that closed form expressions of $e^{A_J t}\left(A_D - \partial_t\right)^m u_{0,t}(x)$ can be computed. This last version is the one proposed by Wan and Yang (2021) for jump–diffusions with state–independent jumps.

#### 4.3.2. State–dependent jump and diffusive component

Finally, consider the general case where $A_J \neq 0$ and both the diffusion and jump component are state–dependent. First observe that when the jumps are state–dependent, or have a complex distribution, $A_J f(x)$ cannot be evaluated analytically for a given function $f$ in general. We propose to resolve this issue by approximating the integral part of $A_J f(x)$, $A_{J1} f(x) = \lambda(x)\int_{\mathbb{R}^d} f(x + c)\nu(c)\,dc$, by

$$\hat{A}_{J1} f(x) = \lambda(x)\sum_{s=1}^{S}\omega_s f\left(x + c_s\right), \tag{4.10}$$

where $\omega_s$ and $c_s$, $s = 1, ..., S$, are integration weights and nodes, respectively. For example, in the case of Monte Carlo integration with $S$ random draws from $\nu$, $\omega_s = 1/S$ and $c_s$ is the $s$th draw from $\nu(\cdot)$. The resulting approximate operator $\hat{A}_J f(x) = \hat{A}_{J1} f(x) - \lambda(x) f(x)$ is on closed form and so we can now continue as in the pure diffusion case. Also note that $\hat{A}_{J1} f(x) \to A_{J1} f(x)$ as $S \to \infty$ which ensures that the added numerical error can be controlled by choosing $S$ large enough.

In the case that $\upsilon(c)$ belongs to the exponential family, the generator of jump component, $A_{J1}$, is well–approximated using Gauss-Hermite or Gauss-Laguerre quadrature. For example, when $J_t$ is i.i.d. scalar with double exponential distribution with mean zero and standard deviation $\sigma_J$, it follows from a change of variables that

$$\begin{aligned}\int_{-\infty}^{\infty} f(x+c)\frac{1}{2\sigma_J}e^{-\frac{|c|}{\sigma_J}}dc &= \int_0^{\infty}\left[f(x+c)+f(x-c)\right]\frac{1}{2\sigma_J}e^{-\frac{c}{\sigma_J}}dc \\ &= \frac{1}{2}\int_0^{\infty}\left[f\left(x+\sigma_J c\right)+f\left(x-\sigma_J c\right)\right]e^{-c}dc.\end{aligned}$$

Then, given the nodes and weights, $c_s^{GL}$ and $\omega_s^{GL}$, for the Gauss-Laguerre quadrature, the approximation takes the following form:

$$\int_{-\infty}^{\infty} f(x+c)\frac{1}{2\sigma_J}e^{-\frac{|c|}{\sigma_J}}dc \simeq \frac{1}{2}\sum_{s=0}^{S_{GL}-1} w_s^{GL}\left[f\left(x+\sigma_J c_s^{GL}\right)+f\left(x-\sigma_J c_s^{GL}\right)\right].$$

We use this approximation method in our numerical studies when we cannot obtain an exact expression of the integral since we found that Gaussian quadrature is more accurate, easier to implement than Monte Carlo methods, and with low computational cost.

With $\hat{A}_{J1}$ replacing $A_{J1}$, we can now use a symbolic software package to obtain expressions of $\left(A_D + \hat{A}_J - \partial_t\right)^m u_{0,t}$, $m = 1, 2, \ldots$. For example,

$$\left(A_D + \hat{A}_J - \partial_t\right)^2 u_{0,t} = \left(A_D - \lambda - \partial_t\right)^2 u_{0,t} + \hat{A}_{J1}\left(A_D - \lambda - \partial_t\right)u_{0,t} + \left(A_{D,t} - \lambda - \partial_t\right)\hat{A}_{J1}u_{0,t} + \hat{A}_{J1}^2 u_{0,t},$$

where the evaluation of $\left(A_D - \lambda - \partial_t\right)^i u_{0,t}$, $i = 1, 2$, can be done using symbolic methods while (here in the univariate case for simplicity)

$$\begin{aligned}\left(A_D - \lambda - \partial_t\right)\hat{A}_{J1}u_{0,t} = \partial_t\left\{\lambda(x)\sum_{s=1}^{S}\omega_s u_{0,t}\left(x+c_s\right)\right\} + \mu(x)\partial_x\left\{\lambda(x)\sum_{s=1}^{S}\omega_s u_{0,t}\left(x+c_s\right)\right\} \\ + \frac{1}{2}\sigma^2(x)\partial_x^2\left\{\lambda(x)\sum_{s=1}^{S}\omega_s u_{0,t}\left(x+c_s\right)\right\} - \lambda^2(x)\sum_{s=1}^{S}\omega_s u_{0,t}\left(x+c_s\right),\end{aligned}$$

and

$$\begin{aligned}\hat{A}_{J1}^2 u_{0,t}(x) &= \lambda(x)\sum_{s_1=1}^{S}\omega_{s_1}\hat{A}_{J1}u_{0,t}\left(x+c_{s_1}\right) \\ &= \lambda(x)\sum_{s_1=1}^{S}\omega_{s_1}\left[\lambda\left(x+c_{s_1}\right)\sum_{s_2=1}^{S}\omega_{s2}u_{0,t}\left(x+c_{s_1}+c_{s_2}\right)\right].\end{aligned}$$

## 5. Numerical results

This section evaluates the approximation performance of our methods for various models. We first demonstrate that the closed–form approximations of moments will eventually diverge for $t$ large enough in Section 5.1. We do this through expansions of a regular moment and a irregular one of Ornstein-Uhlenbeck (OU) process. Subsequently, we employ the machinery to obtain approximate option prices for different jump–diffusion models of different degrees of complexity, including both one and two-factor volatility models, as well as state-independent and state-dependent jumps. In this second part, we find that as long as $t$ is small/moderate, the approximations work well.

### 5.1. Approximations of moments of the Ornstein-Uhlenbeck process

One of the key findings of Section 3 is that the proposed Taylor expansions of $u_t(x) = \mathbb{E}\left[f\left(x_t\right)|x_0 = x\right]$ will in general break down as $t$ gets large enough, cf. Theorem 3.6, and this holds true even if $f$ is a regular function. We here demonstrate this numerically through a simple example, where $x_t$ is an OU process solving

$$dx_t = -\alpha x_t dt + \sigma dW_t \tag{5.1}$$

We here focus on this simple model because its eigenfunctions and spectrum are known and available on closed form; this allows us to obtain closed-form expressions of the conditional moments that we consider below and to provide exact conditions under which

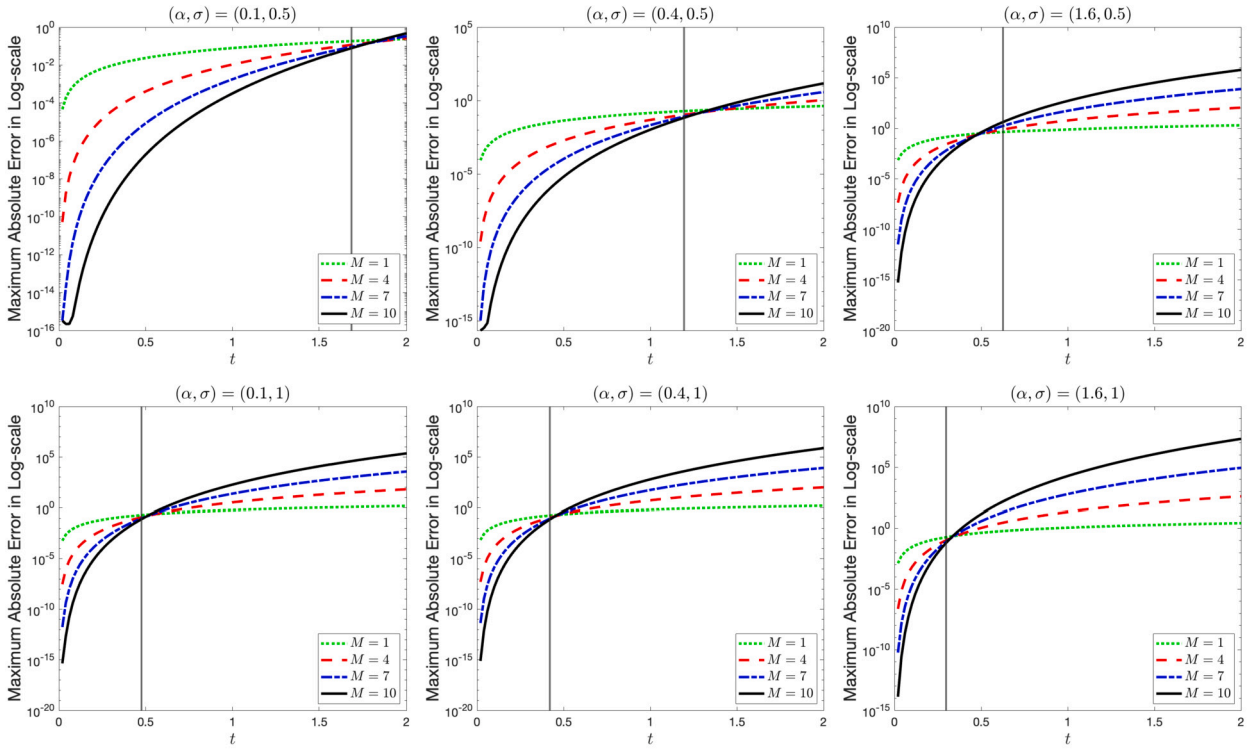


**Fig. 1.** Maximum absolute error of $M$th order approximation of $\mathbb{E}\left[\exp\left(-x_t^2\right)|x_0=x\right]$ as a function of $t$, $M=1,4,7,10$, where $x_t$ solves (5.1) with $\alpha=0.1$ (left), 0.4 (middle), 1.6 (right), and $\sigma=0.5$ (top), 1 (bottom). The vertical line indicates the threshold value $\tau_0$ below which the approximation is known to converge as $M$ diverges.

the corresponding approximation will break down. Specifically, it is well-known that the eigenvalues and eigenfunctions of the OU process's generator are $\xi_i=i\alpha$ and $\phi_i(x)=H_i\left(x\sqrt{\alpha}\right)/\left(2^{i/2}\sqrt{i!}\pi^{1/4}\right)$, where $H_i$ denotes the $i$th Hermite polynomial. Now, recall that Theorem 3.6 shows convergence for $t<\tau_0/(C_Be)$, where $\tau_0>0$ is such that $f\in E_{\tau_0}(\mathcal{F})$. For the OU example, this bound can be improved so that convergence is satisfied for all $t<\tau_0$.

Now, consider $f(x)=\exp(-x_t^2)$. This is a regular function and so we can employ the standard Taylor eseries approximation in eq. (2.16), where we set $r=0$. It can be verified that this choice of $f$ satisfies $f\in E_{\tau_0}(\mathcal{F})$ for all $\tau_0\leq\log(1+\alpha/\sigma^2)/(2\alpha)$. Thus, according to the theory, we expect the Taylor series approximation of $u_t(x)=\mathbb{E}[\exp(-x_t^2)|x_0=x]$ to break down when $t>\log(1+\alpha/\sigma^2)/(2\alpha)$. We verify this numerically in Fig. 1, where the maximum absolute approximation error, $\max_{x\in[-3.3,3.3]}|\hat{u}_t(x)-u_t(x)|$, is plotted for different choices of the parameter values and order of approximation $M$. For any given set of parameters, the approximation error diverges when $t$ gets bigger than the aforementioned threshold, indicated by a vertical line in each sub figure. As such, this numerical example highlights that Taylor series approximations of moments should be used with caution for large values of $t$.

Next, we investigate the performance of our proposed approximation of irregular moments when applied to the OU model. Specifically, we apply (2.17) to approximate the transition density, $p_t(y|x)=\mathbb{E}\left[\delta\left(y-x_t\right)|x_0=x\right]$, of the OU process. For its implementation, we need to choose a smoother. We here employ $p_{0,t}(y|x)=\exp(-(y-x)^2/(2\sigma^2t))/\sqrt{2\pi s^2t}$, which corresponds to the transition density of the process $dx_{0,t}=\sigma dW_t$. Fig. 2 plots the maximum absolute approximation error, $\max_{x\in[0,4],y\in[-3.3,3.3]}|\hat{p}_t(y|x)-p_t(y|x)|$. Compared to the approximation performance for the earlier conditional expectation with $f(x)=\exp\left(x^2\right)$, the effectiveness of the approximation diminishes more slowly with $t$ when $\alpha$ is small (e.g., 0.1 and 0.4), but more rapidly when $\alpha$ is relatively large (e.g., 1.6). However, in all cases, the approximation eventually starts deteriorating as $t$ gets large enough. This is again in line with the theory.

### 5.2. Option pricing in one-factor stochastic volatility models with jumps

In this subsection, we assess the performance of our approximations in the context of option pricing when the underlying asset's dynamics are described by a stochastic volatility model with jumps under the risk-neutral measure. Specifically, we consider a class of asset pricing models where the log-price $s_t$ of a given asset exhibits both stochastic volatility and jumps,

$$ds_t=\left(r-\delta-v_t/2-\lambda\left(v_t\right)\bar{J}\right)dt+\sqrt{v_t}dW_{1t}+\log\left(J_t+1\right)dN_t, \tag{5.2}$$

where the volatility process $v_t$ is solution to either

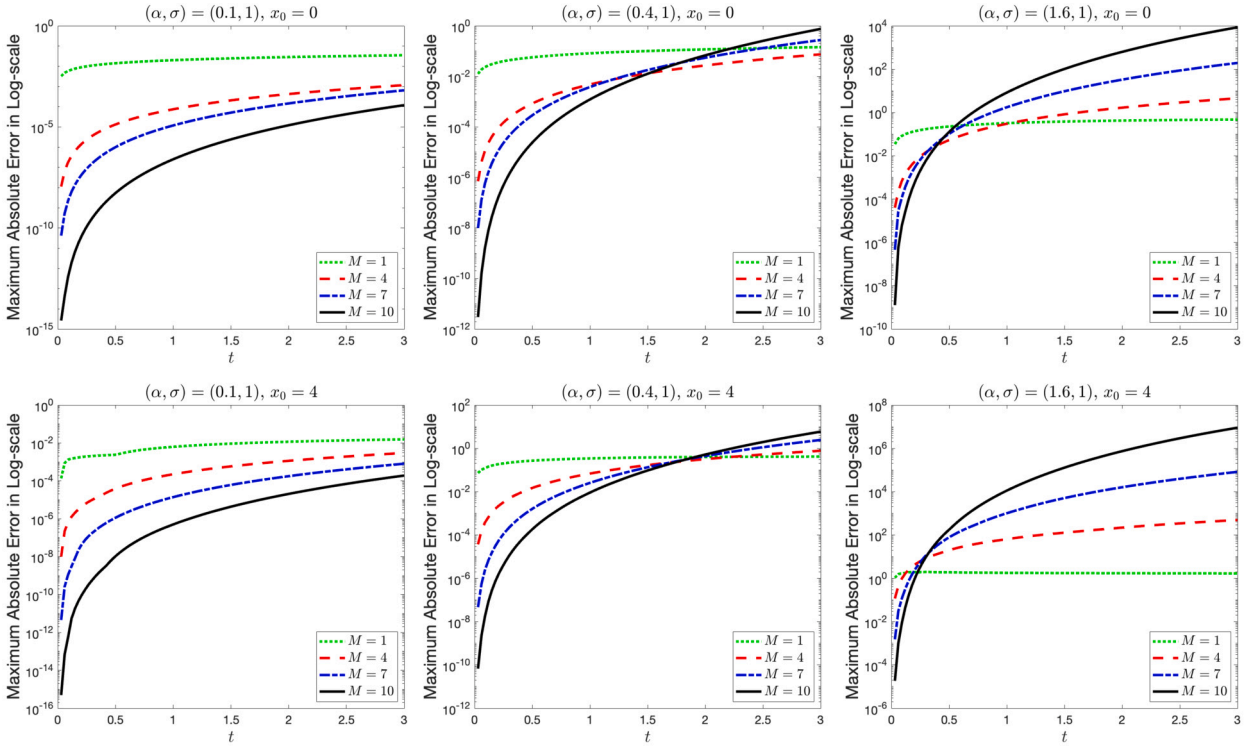


**Fig. 2.** Maximum absolute error of $M$th order approximation of $p_t(y|x)$ as a function of $t$, $M=1,4,7,10$, where $x_t$ solves (5.1) with $\alpha=0.1$ (left), 0.4 (middle), 1.6 (right), $\sigma=1$, and $x=0$ (top), 4 (bottom).

$$dv_t=\kappa_V\left(\alpha_V-v_t\right)dt+\sigma_V v_t^{\beta}\left(\rho dW_{1t}+\sqrt{1-\rho^2}dW_{2t}\right), \tag{5.3}$$

or

$$d\log v_t=\kappa_V\left(\alpha_V-\log v_t\right)dt+\sigma_V\left(\rho dW_{1t}+\sqrt{1-\rho^2}dW_{2t}\right). \tag{5.4}$$

Here, $r$ and $\delta$ are the risk-free rate and the constant dividend, respectively. To ensure that the model has a well-defined solution, $\kappa_V$, $\alpha_V$, $\sigma_V$ are restricted to be positive and $1/2\le\beta\le 1$.

The jump component consists of a Cox process $N_t$ with a jump intensity function given by $\lambda(v)=\lambda_0+\lambda_1 v$, and a random variable $J_t$ with support $[-1,\infty)$, and expectation $\bar{J}$. We include $-\lambda\left(v_t\right)\bar{J}$ in the drift as a compensator such that the jump part is a martingale. For example, if $J+1$ is chosen to be log-normally distributed with parameters $\mu_J$ and $\sigma_J$, then $\bar{J}=\exp\left(\mu_J+\sigma_J^2/2\right)-1$. Special cases of this model include Merton (1976), where both volatility and jump intensity are constant, $v_t=\sigma_0$ and $\lambda(v)=\lambda_0$. Eq. (5.2) together with either (5.3) or (5.4) is a special case of (2.9) with $x_t=\left(s_t,v_t\right)$.

This class of models encompasses those in Andersen et al. (2002) and Wan and Yang (2021), as well as several other special cases. Compared to Andersen et al. (2002), our specification allows the variance process to follow the non-affine continuous-time GARCH model ($\beta=1$) and the CEV model ($1/2<\beta<1$). Additionally, unlike Wan and Yang (2021), we accommodate state-dependent jump intensity ($\lambda_1\neq 0$), which they exclude.

We consider a European call option with payoff $f\left(x_t\right)=f\left(s_t\right)\equiv\max\left\{\exp\left(s_t\right)-K,0\right\}$ at maturity time $t>0$, where $K=100$ is the strike price. With the above model formulated under the risk-neutral measure, let $u_t(s,v)=E\left[f\left(s_t\right)|s_0=s,v_0=v\right]$ represent the expected risk-neutral payoff when the option expires in $t$ time units, given the current log stock price $s$ and volatility $v$.

Within this class of models for $s_t$, no closed-form formula for the option price exists. We implement our proposed series expansion of the unknown price, $\hat{u}_t(s,v)$, as given in (4.1), where we choose $u_{0,t}$ as the payoff under the Black-Scholes model as provided in (4.5).

In the case of state-dependent jumps, we need to compute the integration part of $A_J$ using numerical methods. Since $\log\left(J_t+1\right)$ is i.i.d. with normal distribution with mean $m_J$ and standard deviation $\sigma_J$ for all models in this section, we employ the Gauss-Hermite quadrature with varying numbers of nodes and weights, as detailed in Section 5.

To assess the numerical performance of our expansion, we will use as benchmark the option price obtained via Monte Carlo methods. Specifically, for a given choice of $t$, we simulated 10,000,000 trajectories of $x_t$ with a time-step of 10,000 per year, as described in Section 3 in Giesecke et al. (2018), and then compute the average pay-off across these trajectories. We measure the accuracy by the pointwise absolute percentage error and the maximum (over the stock price) absolute percentage errors defined

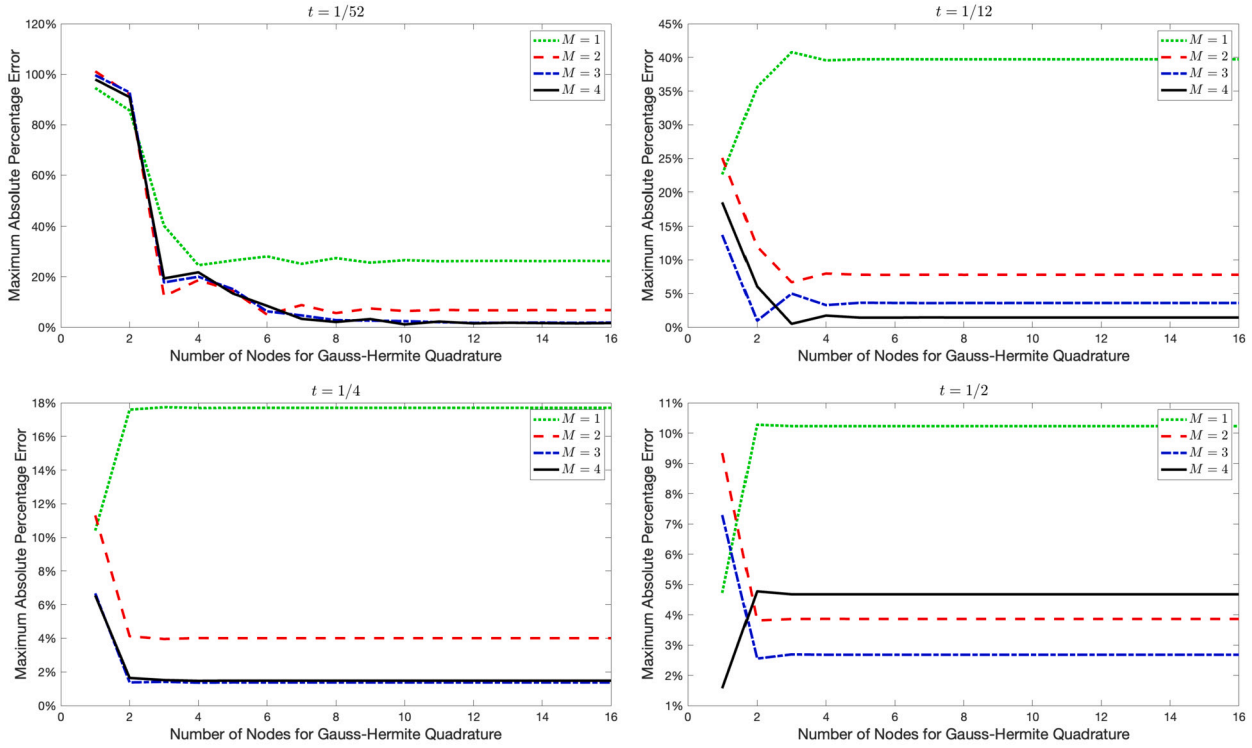


**Fig. 3.** Maximum absolute percentage error of the option price approximation for (5.2)–(5.3) as a function of number of quadrature nodes with $\lambda_1 = 0$, $\beta = 0.5$, $M = 1$, 2, 3, 4, $t = 1/52$ (top left), $1/12$ (top right), $1/4$ (bottom left), $1/2$ (bottom right), and current volatility $v = \alpha_V = 0.0416$.

as $100 \times |\hat{u}_t(s,v) - u_t^{MC}(s,v)|/u_t^{MC}(s,v)$, and $100 \times \max_{s\in[90,110]}|\hat{u}_t(s,v) - u_t^{MC}(s,v)|/u_t^{MC}(s,v)$, respectively, where $\hat{u}_t(s,v)$ and $u_t^{MC}(s,v)$ represent the series expansion and the Monte Carlo approximation of the option price, respectively.

We undertake increasingly challenging experiments to evaluate the resilience of our method in approximating option prices under progressively complex models.

Firstly, we explore the performance our method when jumps are state-independent ($\lambda_1 = 0$) and a CIR specification of the volatility process ($\beta = 0.5$) in (5.2)–(5.3). Following Wan and Yang (2021), the remaining parameter values used in this experiment are the estimates reported in Eraker (2004), $(r, \delta, \kappa_V, \alpha_V, \sigma_V, \rho, \beta) = (0.04, 0, 2.772, 0.0416, 0.203, -0.586, 0.5)$ and $(\lambda_0, \lambda_1, \mu_J, \sigma_J) = (0.504, 0, -0.018, 0.066)$.

We first examine how sensitive our proposed implementation is to the number of nodes used to approximate the jump component of the infinitesimal generator in Fig. 3. In this figure, the maximum (over $x$) absolute percentage approximation error is plotted as a function of number of nodes for different choices of the order of approximation $M$. We see that the approximation is stable with only a relatively small number of nodes and weights being used in the quadrature approximation of the jump component. For a short time to maturity ($t$ =1/52), the Gauss-Hermite quadrature with 10 nodes and weights are needed, but, for larger times to maturity ($t$ =1/12, 1/4, 1/2), only 4–5 nodes and weights are necessary. This suggests that the error in computing the integration part of the jump component is small relative to the error of our Taylor series approximation, especially when $t$ is relative large. Similar findings were made for all other subsequent models that we implemented our method on: Once 10 or more nodes were used, the method showed very little sensitivity to the number of nodes. All subsequent reported results were obtained using 16 quadrature nodes.

Fig. 4 plots the maximum percentage absolute error as a function of the current stock price $s$ when 16 nodes are used to approximate the jump component. Each of the four panels correspond to a particular time to maturity $t = 1/52, 1/12, 1/4$ and $1/2$. We make the following observations: First, for maturities $t = 1/52$ and $1/12$, the approximation error decreases as $M$ increases across all values of the stock price $s$. However, for $t = 1/4$ and $1/2$, the approximation error tends to get larger as $M$ increases for $s < K$. This seems to indicate that, similar to the previous example with the OU-process, our method starts diverging as $M$ increases as $t$ gets larger than some threshold between $1/2$ and $1/4$. Second, for a given order of approximation $M$, our method is more accurate as the time to maturity decreases. Both findings are consistent with our theoretical results.

Next, we investigate the numerical performance of our approximation for the call option under the stochastic volatility model (5.2)–(5.3) with a CEV ($\beta = 0.8$) and GARCH ($\beta = 1$) specification of the volatility process, respectively, while the jump component remains state-independent ($\lambda_1 = 0$). The parameters for the CEV and GARCH specification are from Aït-Sahalia and Kimmel (2007) and Yang and Kanniainen (2016), respectively, but we modified the parameters for the jump part to match those in Wan and Yang (2021). For the CEV specification, the parameter values used are $(r, \delta, \kappa_V, \alpha_V, \sigma_V, \rho, \beta) = (0.04, 0.015, 4, 0.05, 0.75, -0.75, 0.8)$ and $(\lambda_0, \lambda_1, \mu_J, \sigma_J) = (3, 0, -0.07, 0.07)$. For the GARCH specification, the parameter values were chosen as $(r, \delta, \kappa_V, \alpha_V, \sigma_V, \rho, \beta) =$

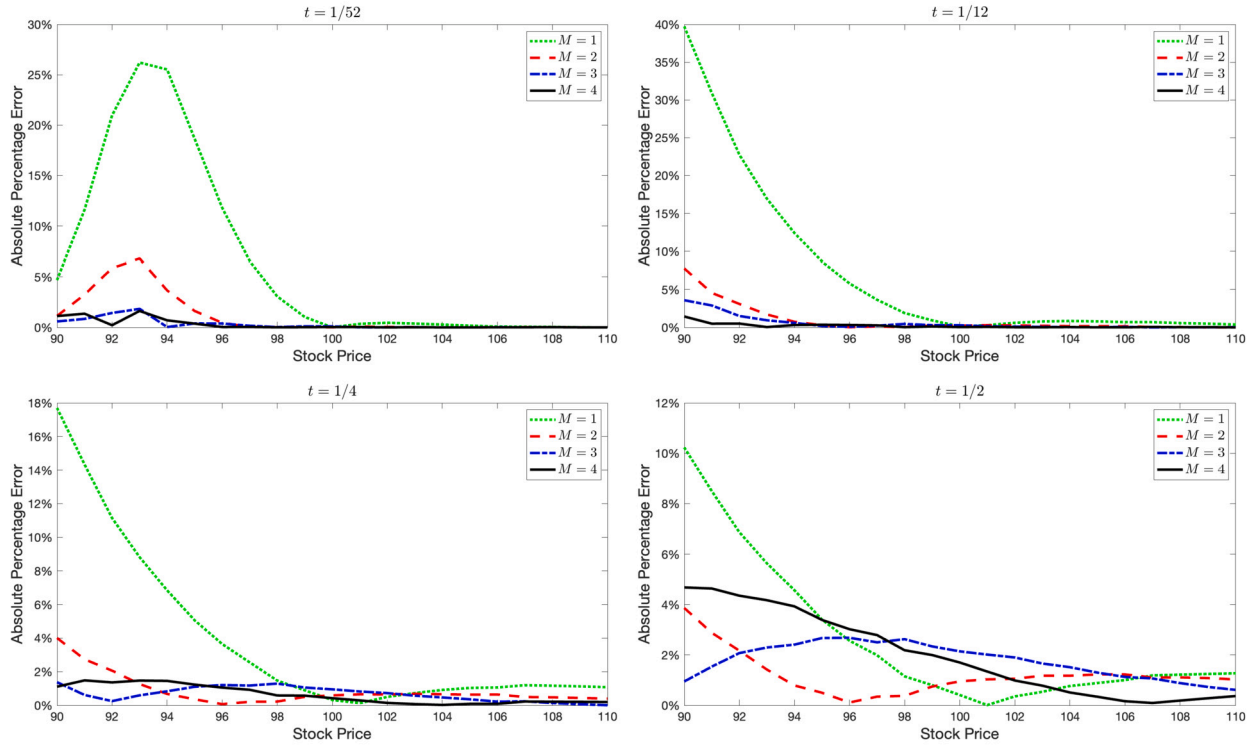


**Fig. 4.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1 = 0$, $\beta = 0.5$, $M = 1, 2, 3, 4$, $t = 1/52$ (top left), $1/12$ (top right), $1/4$ (bottom left), $1/2$ (bottom right) and current volatility $v = \alpha_V = 0.0416$.

$(0.04, 0, 3.7367, 0.023, 1.823, -0.8113, 1)$ and $\left(\lambda_0, \lambda_1, \mu_J, \sigma_J\right) = (3, 0, -0.07, 0.07)$. Sixteen quadrature nodes were used to approximate the jump component.

Fig. 5 and 6 show the performance of our method for the CEV and GARCH specification, respectively. For $t = 1/52$ and $t = 1/12$, the performance of the approximation for both specifications generally shares the same patterns as the ones reported for the CIR specification in Fig. 4 for the however, for longer time-to-maturity ($t = 1/4$ and $t = 1/2$), the performance of higher order approximations tends to get worse as $\beta$ increases. This seems to indicate that for larger values of $\beta$, closed-form approximations start to break down at a lower threshold for $t$. Over all, the performance of the approximation is good for shorter maturities and/or smaller values of $\beta$.

Next, we examine the performance of our method for the case where the option price is computed under models with state-dependent jump intensities ($\lambda_1 \neq 0$).

In Figs. 7–9, respectively, we depict the relative error of the approximation for the same three models, the SV-CIR, the SV-CEV and the SV-GARCH, but now with state-dependent jumps for an option with time-to-maturity of one month ($t = 1/12$). In each figure, from left to right, the state dependency of jump intensity ranges $\lambda_1 = 1$, 10, 30. For each of the three specifications, the remaining parameter values and the current value of volatility are kept at the same values as reported earlier.

Overall, the approximation errors for each of the three models are comparable to those of the same model with state-independent jumps ($\lambda_1 = 0$). Furthermore, the absolute percentage error is smaller for all orders of approximation for larger $\lambda_1$. This indicates that while the magnitude of $\lambda_1$ affects the level of option prices, it has little effect on the approximation error. If anything, the approximations seem to work better when there is high state dependence in the jump intensity.

Next, we consider the performance when $v_t$ solves the log-volatility model (5.4) with parameter values chosen as the estimates reported in Andersen et al. (2002), $\left(r, \delta, \kappa_V, \alpha_V, \sigma_V, \rho\right) = (0.0304, 0, 0.0145, -0.8276, 0.1153, -0.6125)$ and $\left(\lambda_0, \mu_J, \sigma_J\right) = (0.0137, -0.000125, 0.015)$. Figs. 10 and 11 display the relative error of the approximation for the call option under this model for different values of $t$ and $\lambda_1$ with $v = \alpha_V = -0.8276$. As before, 16 quadrature nodes were used throughout.

Even for the 2nd order approximation, the approximation error is quite small for all choices of time-to-maturity and $\lambda_1$. Thus, for the log-volatility specification, our method appears to first start breaking down for options relatively long time to maturity. The plotted errors are now more ragged, which we conjecture is due to bigger numerical errors in the Monte Carlo benchmark that we use for comparison.

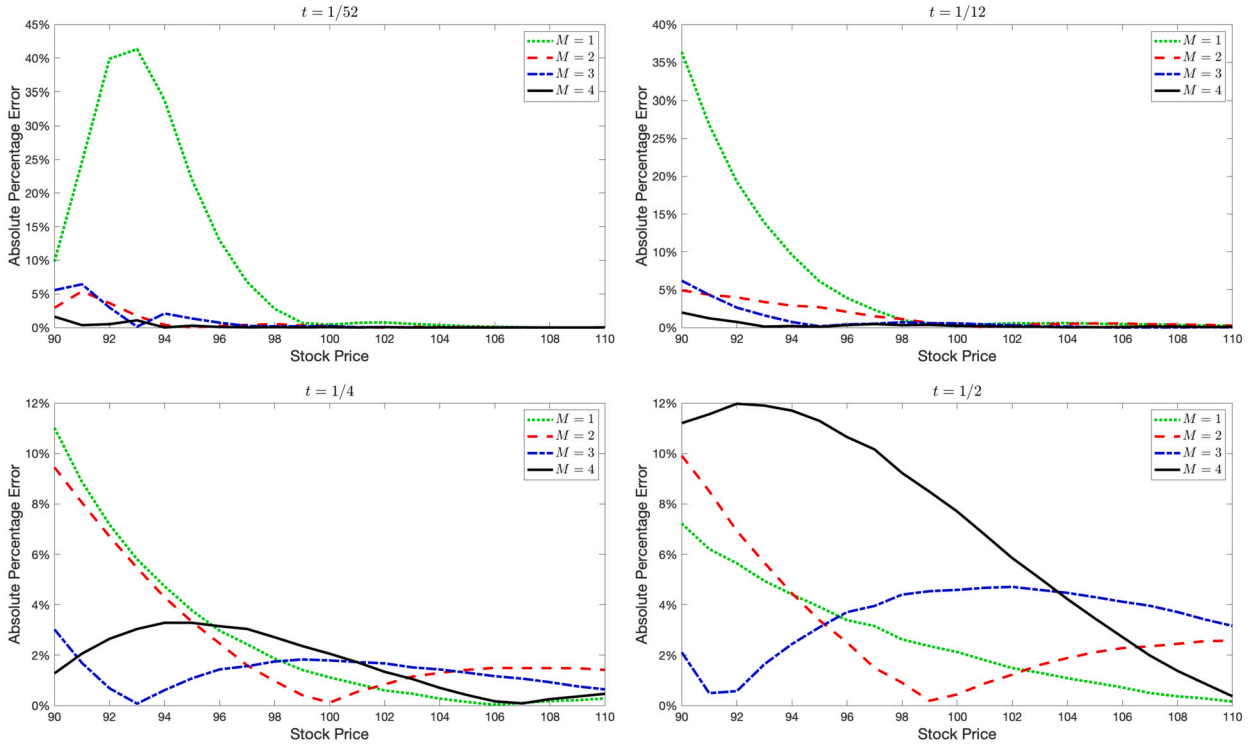


**Fig. 5.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1 = 0$, $\beta = 0.8$, $M = 1, 2, 3, 4$, $t = 1/52$ (top left), $1/12$ (top right), $1/4$ (bottom left), $1/2$ (bottom right) and current volatility $v = \alpha_V = 0.05$.

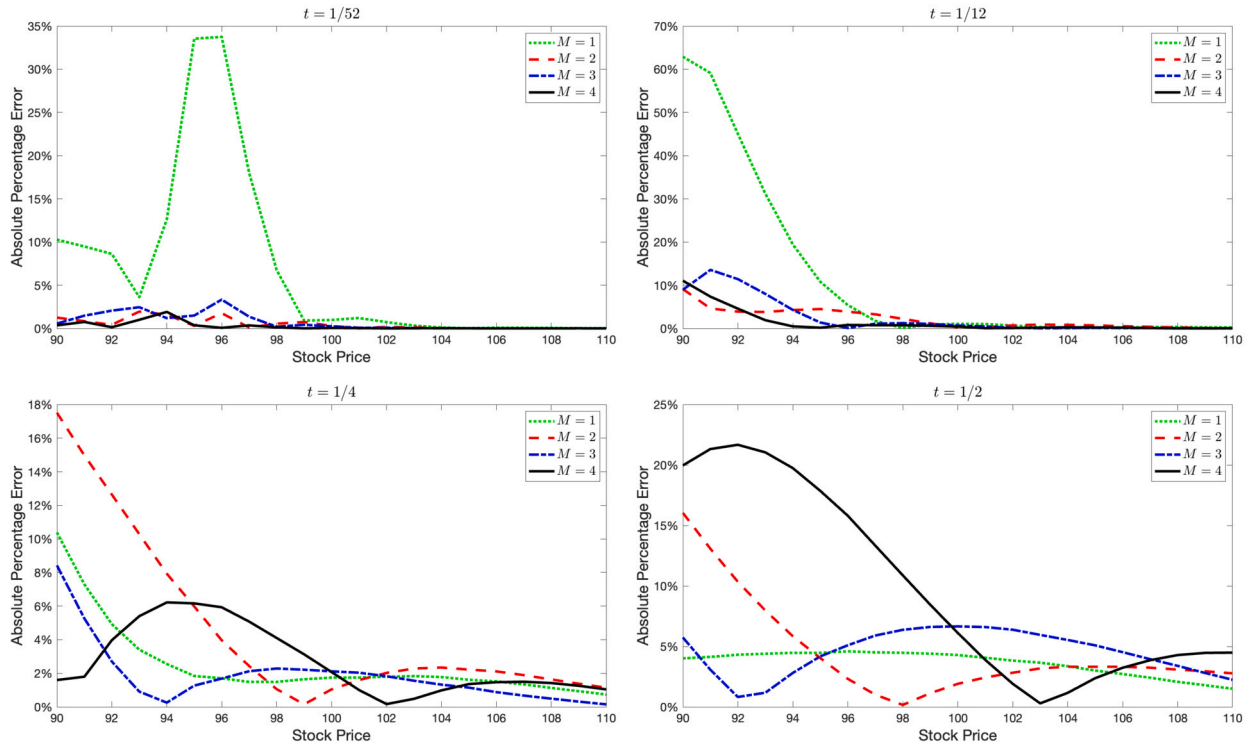


**Fig. 6.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1 = 0$, $\beta = 1.0$, $M = 1, 2, 3, 4$, $t = 1/52$ (top left), $1/12$ (top right), $1/4$ (bottom left), $1/2$ (bottom right) and current volatility $v = \alpha_V = 0.023$.

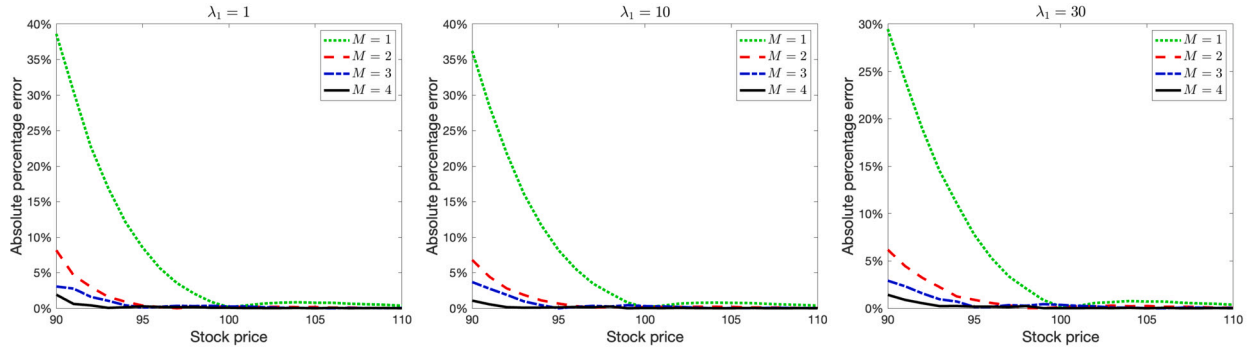


**Fig. 7.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1$ =1 (left), 10 (middle), and 30 (right), $\beta = 0.5$, $M = 1, 2, 3, 4$, $t = 1/12$ and current volatility $v = \alpha_V = 0.0416$.

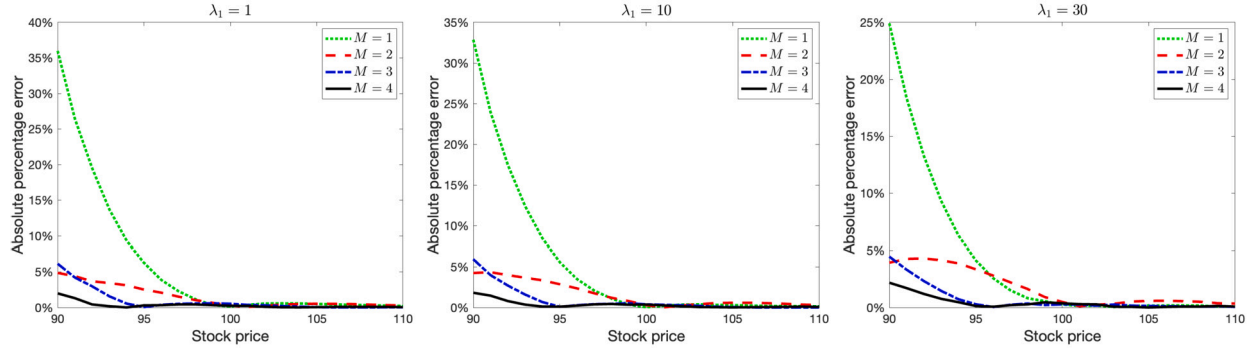


**Fig. 8.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1$ =1 (left), 10 (middle), and 30 (right), $\beta = 0.8$, $M = 1, 2, 3, 4$, $t = 1/12$ and current volatility $v = \alpha_V = 0.05$.

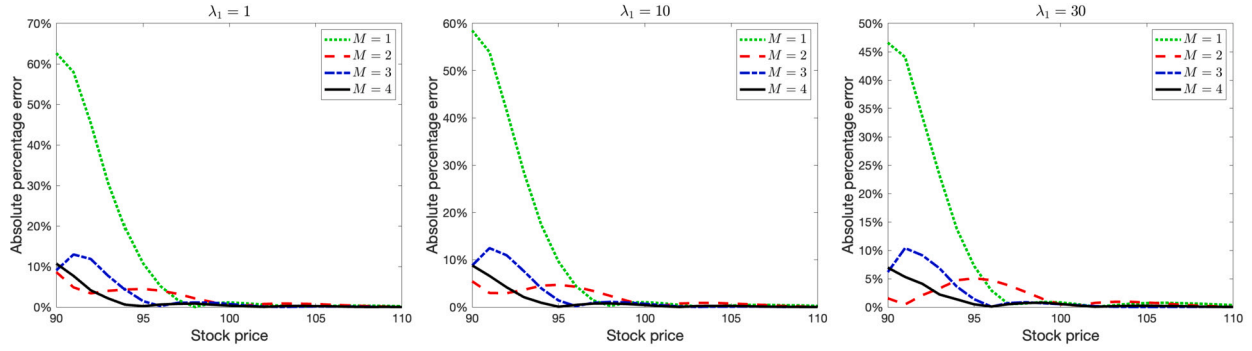


**Fig. 9.** Absolute percentage error of option price approximation for (5.2)–(5.3) as a function of current stock price $s$ with $\lambda_1$ =1 (left), 10 (middle), and 30 (right), $\beta = 1.0$, $M = 1, 2, 3, 4$, $t = 1/12$ and current volatility $v = \alpha_V = 0.023$.

### *5.3. Option pricing in a two-factor stochastic volatility model with jumps*

We now examine the performance of our method when applied to more complex models that go beyond one-factor volatility. Specifically, we consider the stochastic volatility model with two volatility factors used in Filipović et al. (2016). In this specification, the dynamics of $s_t$ under the risk-neutral measure are given by

$$\begin{aligned} ds_t &= \left(r - \delta - v_t/2 - \lambda\left(v_t, m_t\right)\bar{J}\right)dt + \sqrt{v_t}dW_{1t} + \log\left(J_t^S + 1\right)dN_t, \\ dv_t &= \kappa_V\left(m_t - v_t\right)dt + \sigma_V\sqrt{v_t}\left(\rho dW_{1t} + \sqrt{1-\rho^2}dW_{2t}\right) + J_t^V dN_t, \\ dm_t &= \kappa_m\left(\alpha_m - m_t\right)dt + \sigma_m\sqrt{m_t}W_{3t}, \end{aligned} \tag{5.5}$$

where $W_1$, $W_2$, and $W_3$ are mutually independent standard Brownian motions. Compared to the models of the previous subsection, there is a second variance factor $m_t$, which represents a stochastic level around which $v_t$ reverts. The jump component consists of: (i) $N_t$, a Cox process with a bounded intensity function given by $\lambda(v, m) = \lambda_0 + \lambda_1 v + \lambda_2 m$, and the variance jump size $J^V$ is

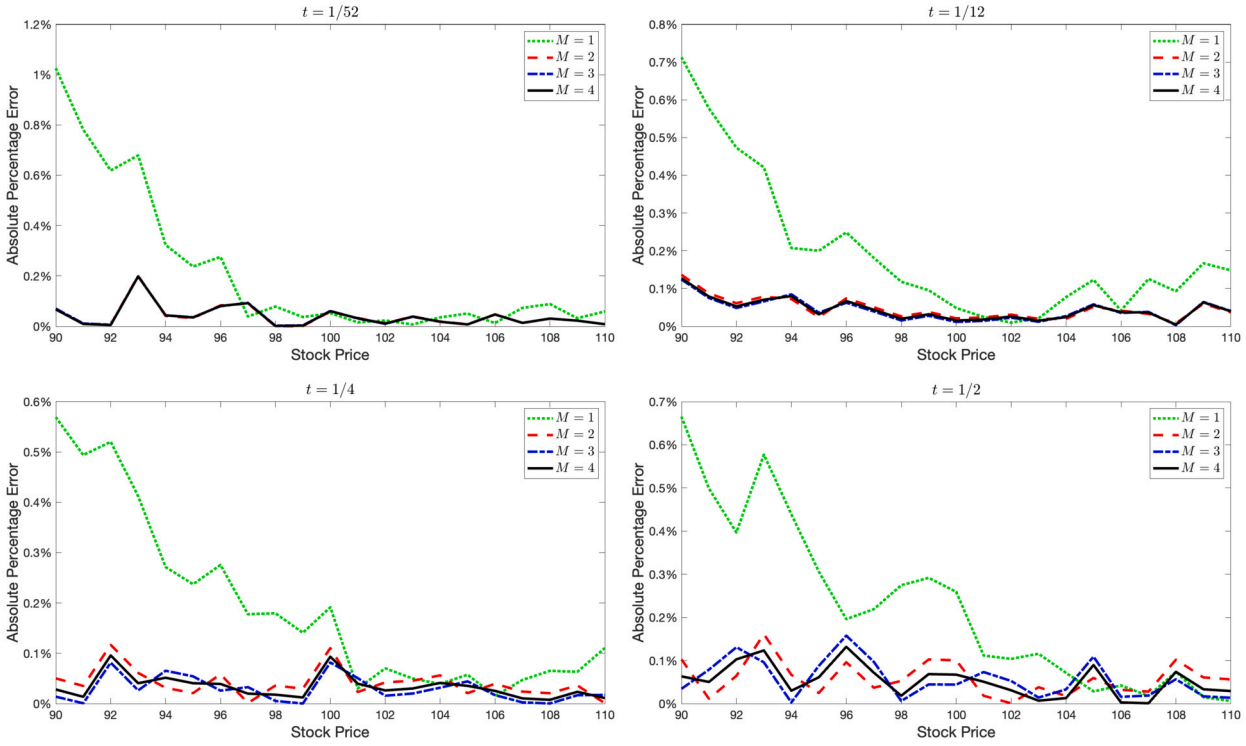


**Fig. 10.** Absolute percentage error of option price approximation for (5.2) and (5.4) with $\lambda_1 = 1.0$, and $t$ =1/52 (top left), 1/12 (top right), 1/4 (bottom left), 1/2 (bottom right) and current volatility $\log v = \alpha_V = -0.8276$.

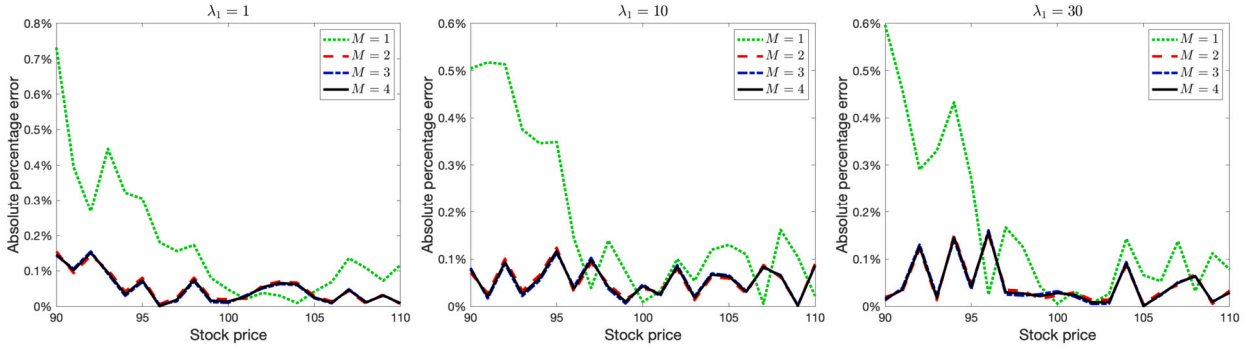


**Fig. 11.** Absolute percentage error of option price approximation for (5.2) and (5.4) with $\lambda_1$ =1 (left), 10 (centre), 30 (right), $t$ =1/12 and current volatility $\log v = \alpha_V = -0.8276$.

exponentially distributed with parameter $\mu_J^V = \mathbb{E}\left[J_t^V\right]$. The parameter values were chosen as the estimates reported in Aït-Sahalia et al. (2020): $\left(r, \delta, \kappa_V, \kappa_m, \alpha_m, \sigma_V, \sigma_m, \rho\right) = (0.04, 0.015, 3.1206, 3.3168, 0.1125, 0.394, 0.0835, -0.688)$ and $\left(\lambda_0, \lambda_1, \lambda_2\right) = (2.096, 21.225, 0)$ and $\left(\mu_J, \sigma_J, \mu_J^V\right) = (-0.012, 0.043, 0.002)$. Note that the option price is now a function of current stock price $s_0 = s$, current volatility $v_0 = v$ and current mean volatility $m_0 = m$.

Fig. 12 reports the performance of our approximation for different times to maturity. The performance of the approximation in this case shares the same patterns that we found for the one-factor volatility models: For $t = 1/52$ and $t = 1/12$, the error vanishes as $M$ increases, but for $t = 1/4$ and $t = 1/2$, the opposite pattern emerges.

## 6. Conclusion

This paper provides a general framework for developing and analyzing series expansions of moments of continuous-time Markov processes, including jump-diffusions. The expansions come in two versions depending on the features of the moment. For "regular" moments, we provide conditions under which the corresponding expansion will converge towards the actual moments as more terms are added. For the "smoothed" expansion, no such theoretical guarantees exist: The expansion will eventually become imprecise as the number of terms grows. A numerical study shows that the smoothed expansions still work well in practice when time to maturity

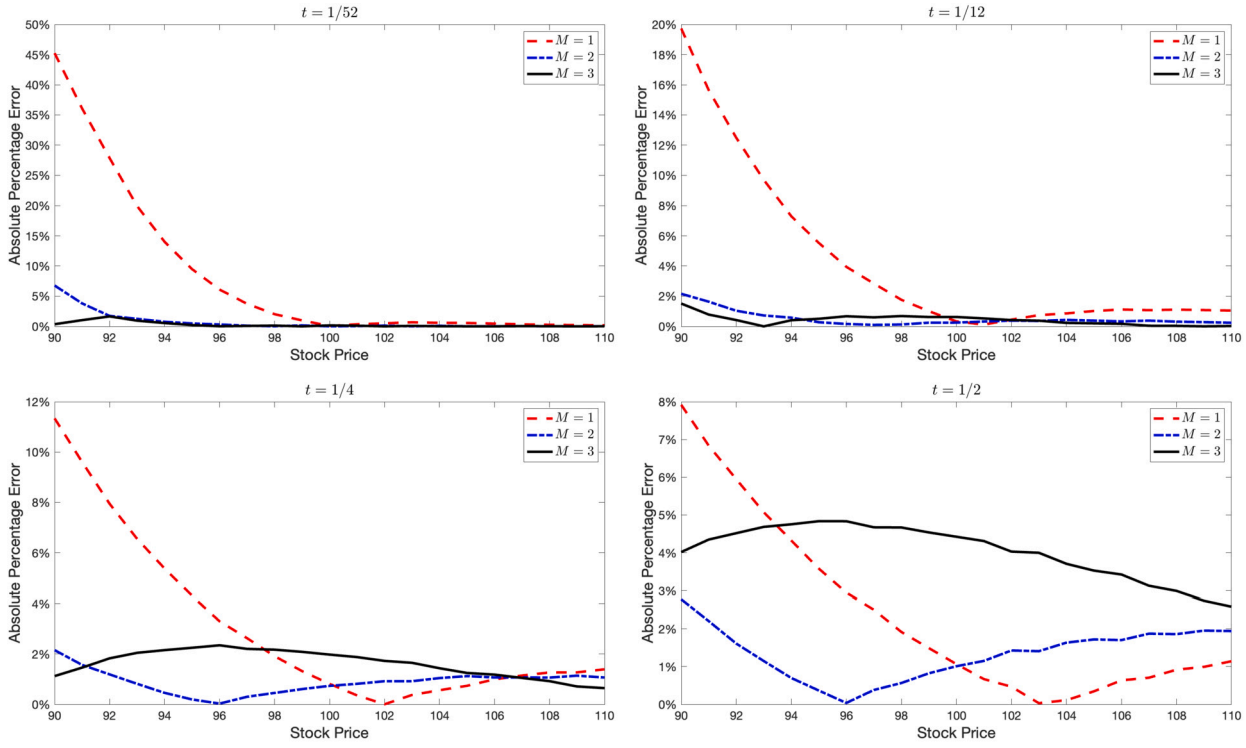


**Fig. 12.** Absolute percentage error of option price approximation for (5.5) with $M = 1$, 2, and 3, $t = 1/52$ (top left), $1/12$ (top right), $1/4$ (bottom left), $1/2$ (bottom right) and $v = m = 0.1125$.

$t$ is relatively small and even if only a relatively small number of terms are used in its implementation. However, as time to maturity increases, the method will eventually start breaking down. Thus, it should be used with care in applications where $t$ is relatively large.

## Appendix A. Extension to time-inhomogenous problems

We here present the extension of our method to handle time–inhomogenous models and problems where no closed-form solution is available to (2.20). As motivating example, consider the following extended version of the model in (2.9):

$$dx_t = \mu_t\left(x_t\right)dt + \sigma_t\left(x_t\right)dW_t + J_t\left(x_t\right)dN_t, \tag{A.1}$$

where now $\lambda_t(x)$, $\mu_t(x)$, $\sigma_t(x)$ and $\nu_t\left(x_t\right)$ are now allowed to vary with $t$. This in turn implies that the corresponding generator is also time–varying, $A_t f(x) = A_{D,t} f(x) + A_{J,t} f(x)$, where

$$A_{D,t} f(x) = \sum_{i=1}^{d} \mu_{i,t}(x)\,\partial_{x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma^2_{ij,t}(x)\,\partial^2_{x_i,x_j} f(x),$$

$$A_{J,t} f(x) = \lambda_t(x) \int_{\mathbb{R}^d} \left[f(x+c) - f(x)\right] \nu_t(c)\,dc.$$

We are interested in computing $u_{s,t}(x)$ defined as

$$u_{s,t}(x) = E_{s,t} f(x), \qquad 0 \le s \le t, \tag{A.2}$$

where

$$(s,t,f) \mapsto E_{s,t} f(x) \equiv \mathbb{E}\left[\exp\left(-\int_s^t r\left(x_u\right)du\right) f\left(x_t\right) \middle| x_s = x\right]. \tag{A.3}$$

Due to the time–inhomogeneity, the operator $E_{s,t} f(x)$ is now indexed by two time variables, $s$ and $t$. At the same time, for any fixed value of $s \ge 0$, $(t,f) \mapsto E_{s,t} f(x)$ remains a semi–group when $\mathcal{F}$ is chosen suitably. Most of the ideas and results from Sections 2–3 therefore carry over to the time–inhomogenous case with only minor differences. Below, we present the series expansion and explain how the theory applies to this.

We take as starting point a given $(s,t,f) \mapsto E_{s,t}f(x)$ where, for any given $s \geq 0$, $(t,f) \mapsto E_{s,t}f(x)$ is assumed to be semi–group on some function space $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$. In the following, we keep $s \geq 0$ fixed. We denote by $\mathcal{D}\left(B_s\right)$ the set of functions $f \in \mathcal{F}$ for which there exists $g_s \in \mathcal{F}$ such that, for each $t \geq 0$,

$$E_{s,t}f(x) = f(x) + \int_s^t E_{s,u}g_s(x)\,du, \tag{A.4}$$

and we write $B_s f(x) := g_s(x)$ and call $B_s$ the (extended) generator of $E_{s,t}$. In the motivating example above, it is easily shown by Ito's Lemma that $B_t = A_t - r$ on the space

$$\mathcal{D}_0\left(B_s\right) := \left\{ f \in C^2 \cap \mathcal{F} : E_{s,t}|f| \text{ and } E_{s,t}\left\|\frac{\partial f}{\partial x}\sigma_s\right\|^2 \text{ exist for all } t > 0 \right\} \subseteq \mathcal{D}\left(B_s\right).$$

For any regular function $f$, regular in the sense that $f \in \mathcal{D}\left(B_s^M\right)$, we have

$$\partial_t u_{s,t}(x) = B_t^m u_{s,t}(x), \quad t > 0, \tag{A.5}$$

cf. Rüschendorf et al. (2016), which corresponds to the so–called forward equation. Thus, in this case the following is a valid series expansion of $u_{s,t}(x)$:

$$\hat{u}_{s,t}(x) = \sum_{m=0}^{M} \frac{(t-s)^m}{m!} B_s^m f(x). \tag{A.6}$$

If $f$ is irregular, so that $f \notin \mathcal{D}\left(B_s\right)$, we introduce a smoothed version of it, $u_{0,s,t}(x)$ which is assumed to satisfy:

**A.0'** (i) $\lim_{t \to s^+} u_{0,s,t}(x) = f(x)$ and (ii) $u_{0,s,t} \in \mathcal{D}\left(\left(\partial_t\right)^{M_1}\right) \cap \mathcal{D}\left(B_s^{M_2}\right)$ for some $M_1, M_2 \geq 1$.

Following the same steps as in the time–homogenous case of Section 2, we obtain the following series expansion:

$$\hat{u}_{s,t}(x) = \sum_{m=0}^{M} \frac{(t-s)^m}{m!} \left(B_s - \partial_t\right)^m u_{0,s,t}(x). \tag{A.7}$$

## Appendix B. Proofs

**Proof of Theorem 3.1.** By Definition 2.1, for any $f \in \mathcal{D}(B)$,

$$u_s(x) = f(x) + \int_0^s E_w(Bf)(x)\,dw \tag{B.1}$$

Apply $E_t$ on both sides of this equation and then use that $E_s E_t = E_{s+t} = E_t E_s$ to obtain

$$E_t u_s(x) = u_s(x) + \int_0^t E_{s+w}(Bf)(x)\,dw = \int_0^t E_w\left(E_s(Bf)\right)(x)\,dw.$$

The second part of the theorem is obtained by taking derivatives w.r.t. $s$ on both sides of (B.1) and using that the right-hand side derivative equals $E_s(Bf)(x) = Bu_s(x)$ if this function is continuous w.r.t. $s$ from the right. □

**Proof of Theorem 3.3.** We expand $t \mapsto E_t f(x)$ around $E_0 f(x) = f(x)$ recursively: First, by Definition 2.1,

$$E_t f(x) = f(x) + \int_0^t E_{t_1}(Bf)(x)\,dt_1. \tag{B.2}$$

Since $Bf \in \mathcal{D}(B)$ by assumption, we can apply (B.2) again to $E_{t_1}(Bf)(x)$ yielding

$$E_{t_1}(Bf)(x) = Bf(x) + \int_0^{t_1} E_{t_2}\left(B^2 f\right)(x)\,dt_2.$$

Substitute the right-hand side of the last equation into (B.2) to obtain

$$E_t f(x) = f(x) + \int_0^t \left\{ Bf(x) + \int_0^{t_1} E_{t_2}\left(B^2 f\right)(x)\, dt_2 \right\} dt_1$$
$$= f(x) + tBf(x) + \int_0^t \int_0^{t_1} E_{t_2}\left(B^2 f\right)(x)\, dt_2 dt_1.$$

Repeating this argument $M$ more times yields the claimed result. □

**Proof of Theorem 3.4.** By definition, $\|B^m f\|_{\mathcal{F}}/m! \leq 1/T_0^m$. Thus,

$$\left\|u_t - \hat{u}_t\right\|_{\mathcal{F}} \leq \sum_{m=M+1}^{\infty} \frac{t^m}{m!} \|B^m f(x)\|_{\mathcal{F}} \leq \sum_{m=M+1}^{\infty} \left(\frac{t}{T_0}\right)^m = \frac{\left(t/T_0\right)^{M+1}}{1 - t/T_0} \to 0. \quad \square$$

**Proof of Theorem 3.6.** The first part follows from Theorem 2.5.2 of Pazy (1983). To show the second part, recall the definition of radius of convergence $T_0$ in (3.4). To bound the right hand side of (3.4), first use that $f(x) = E_{\tau_0} g(x)$ and that $B$ and $E_{\tau_0}$ commute to obtain $\|B^m f\|_{\mathcal{F}} = \left\| \left(BE_{\tau_0/m}\right)^m g \right\|_{\mathcal{F}} \leq \left\| BE_{\tau_0/m} \right\|_{\text{op}}^m \|g\|_{\mathcal{F}}$. Next, due to (3.5)–(3.6), we can apply part (d) of Theorem 2.5.2 of Pazy (1983) yielding $\left\| BE_{\tau_0/m} \right\|_{\text{op}}^m \leq \left(C_B m/\tau_0\right)^m$. In total,

$$\|B^m f\|_{\mathcal{F}}/m! \leq \left\{ \left(\frac{C_A}{\tau_0}\right)^m m^m/m! \right\}^{1/m} \|g\|_{\mathcal{F}}^{1/m} \leq \left(\frac{C_B e}{\tau_0}\right) \|g\|_{\mathcal{F}}^{1/m},$$

and we conclude that $T_0 \geq \tau_0/\left(C_B e\right)$. □

**Proof of Corollary 3.8.** With the function space being a Hilbert space, we are able to introduce the adjoint $A^*$ of the operator $A$ with corresponding semigroup $E_t^* = e^{A^* t}$. If $x_t$ indeed is reversible in the sense that $A = A^*$ then $\sigma(A) \subseteq (-\infty, 0]$ and so (3.5) is satisfied. (cf. eq. 5.8 in Hansen and Scheinkman (1995)). Moreover, by the Spectral Mapping Theorem (Rudin (1973), Theorem 10.28), the spectrum of the resolvent satisfies

$$\sigma(R(\lambda)) \setminus \{0\} = (\lambda - \sigma(A))^{-1} = \left\{ \frac{1}{\lambda - w} : w \in \sigma(A) \right\}$$

Since $A$ is self-adjoint so is $R(\lambda)$ for any $\lambda \notin \sigma(A)$. Thus,

$$\|R(\lambda)\|_{\text{op}} = \max_{w \in \sigma(R(\lambda))} |w| = \max_{w \in \sigma(A)} \frac{1}{|\lambda - w|} \leq \max_{w \leq 0} \frac{1}{|\lambda - w|} = \frac{1}{|\lambda|},$$

and so (3.6) is satisfied. □

**Proof of Theorem 3.9.** For any $f \in \mathcal{F}_0$,

$$\left\|A_D f\right\|_{\mathcal{F}_0} \leq \sum_{i=1}^d \left\|\mu_i\right\|_{\mathcal{F}} \left\| \frac{\partial f}{\partial x_i} \right\|_{\mathcal{F}} + \frac{1}{2} \sum_{i,j=1}^d \left\|\sigma_{ij}^2\right\|_{\mathcal{F}} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_{\mathcal{F}}$$
$$\leq \left( \sum_{i=1}^d \left\|\mu_i\right\|_{\mathcal{F}_0} + \frac{1}{2} \sum_{i,j=1}^d \left\|\sigma_{ij}^2\right\|_{\mathcal{F}_0} \right) \|f\|_{\mathcal{F}_0}$$
$$=: \bar{A} \|f\|_{\mathcal{F}_0},$$

where $\bar{A} < \infty$ under the assumptions of the theorem. Thus, $\|A\|_{\text{op}} = \sup_{\|f\|_{\mathcal{F}_0} \leq 1} \|Af\|_{\mathcal{F}_0} < \infty$ and so $A : \mathcal{F}_0 \mapsto \mathcal{F}_0$ is a bounded operator. This in turn implies that $\sum_{m=0}^{\infty} \frac{t^m}{m!} A^m f(x)$ is a well-defined representation of $w_t(x)$ for any $f \in \mathcal{F}_0$ and so the power series approximation is consistent. In particular,

$$\left\|\hat{u}_t - u_t\right\|_{\mathcal{F}_0} \leq \frac{t^{M+1}}{(M+1)!} \left\|A^{M+1} f\right\|_{\mathcal{F}_0} \leq \frac{\left(t\bar{A}\right)^{M+1}}{(M+1)!} \|f\|_{\mathcal{F}_0}. \quad \square$$

**Proof of Theorem 3.10.** The first part follows from Theorem 1.1 in Escauriaza et al. (2017). For the second part, first note that $w_{0,t}(x) = E_{0,t+\tau_0} g(x)$. Now, by Theorem 1.1 in Escauriaza et al. (2017), $\left| \partial_t^m w_{0,t}\big|_{t=0} \right| \leq C \left(\rho\tau_0\right)^{-m} m!$, for all $x \in \mathcal{X}_0$, for some constant $\rho = \rho(B,d) \in (0,1]$. This in turn implies that the power series expansion will converge with radius of convergence bounded by

$$T_0^{-1} = \lim_{m \to \infty} \sup \left\{ \frac{1}{m!} \left\| \partial_t^m w_{0,t}\big|_{t=0} \right\|_{\mathcal{F},0} \right\}^{1/m} \leq \frac{1}{\rho\tau_0}. \quad \square$$

## References


Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70 (1), 223–262.

Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Ann. Stat. 36 (2), 906–937.

Aït-Sahalia, Y., Hansen, L.P., Scheinkman, J.A., 2010. Operator methods for continuous-time Markov processes. In: Handbook of Financial Econometrics: Tools and Techniques. Elsevier, pp. 1–66.

Aït-Sahalia, Y., Karaman, M., Mancini, L., 2020. The term structure of equity and variance risk premia. J. Econom. 219 (2), 204–230.

Aït-Sahalia, Y., Kimmel, R.L., 2007. Maximum likelihood estimation of stochastic volatility models. J. Financ. Econ. 83 (2), 413–452.

Ames, W., 1992. Numerical Methods for Partial Differential Equations. Academic Press.

Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. J. Finance 57 (3), 1239–1284.

Bakshi, G., Ju, N., Ou-Yang, H., 2006. Estimation of continuous-time models with an application to equity volatility dynamics. J. Financ. Econ. 82 (1), 227–249.

Beskos, A., Papaspiliopoulos, O., Roberts, G., 2009. Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. Ann. Stat. 37 (1), 223–245.

Björk, T., 2009. Arbitrage Theory in Continuous Time. Oxford University Press.

Brandt, M.W., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. J. Financ. Econ. 63 (2), 161–210.

Cont, R., Tankov, P., 2003. Financial Modelling with Jump Processes. Chapman and Hall/CRC.

Davies, E.B., 2007. Linear Operators and Their Spectra. Cambridge University Press.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. J. Bus. Econ. Stat. 20 (3), 297–338.

Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69 (4), 959–993.

Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. J. Finance 59 (3), 1367–1403.

Escauriaza, L., Montaner, S., Zhang, C., 2017. Analyticity of solutions to parabolic evolutions and applications. SIAM J. Math. Anal. 49 (5), 4064–4092.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes. John Wiley & Sons, Inc.

Filipović, D., Gourier, E., Mancini, L., 2016. Quadratic variance swap models. J. Financ. Econ. 119 (1), 44–68.

Filipović, D., Mayerhofer, E., Schneider, P., 2013. Density approximations for multivariate affine jump-diffusion processes. J. Econom. 176 (2), 93–111.

Giesecke, K., Shkolnik, A., Teng, G., Wei, Y., 2018. Numerical solution of jump-diffusion SDEs. Available at SSRN 2298701.

Hansen, L.P., Scheinkman, J.A., 1995. Back to the future: generating moment implications for continuous-time Markov processes. Econometrica 63 (4), 767–804.

Hansen, L.P., Scheinkman, J.A., Touzi, N., 1998. Spectral methods for identifying scalar diffusions. J. Econom. 86 (1), 1–32.

Kristensen, D., Mele, A., 2011. Adding and subtracting Black-Scholes: a new approach to approximating derivative prices in continuous-time models. J. Financ. Econ. 102 (2), 390–415.

Kristensen, D., Shin, Y., 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood. J. Econom. 167 (1), 76–94.

Li, C., 2013. Maximum-likelihood estimation for diffusion processes via closed-form density expansions. Ann. Stat. 41 (3), 1350–1380.

Lunardi, A., 1995. Stability in fully nonlinear parabolic equations. Arch. Ration. Mech. Anal. 130 (1), 1–24.

Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. 3 (1), 125–144.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Pazy, A., 1983. Semigroups of Linear Operators and Applications to Partial Differential Equations. Springer, New York.

Rudin, W., 1973. Functional Analysis. Higher Mathematics Series. McGraw-Hill.

Rüschendorf, L., Schnurr, A., Wolf, V., 2016. Comparison of time-inhomogeneous Markov processes. Adv. Appl. Probab. 48 (4), 1015–1044.

Schaumburg, E., 2004. Estimation of Markov processes with Levy type generators. Unpublished working paper. Kellogg School of Management.

Sermaidis, G., Papaspiliopoulos, O., Roberts, G.O., Beskos, A., Fearnhead, P., 2013. Markov Chain Monte Carlo for exact inference for diffusions. Scand. J. Stat. 40 (2), 294–321.

Wan, X., Yang, N., 2021. Hermite expansion of transition densities and European option prices for multivariate diffusions with jumps. J. Econ. Dyn. Control 125, 104083.

Yang, H., Kanniainen, J., 2016. Jump and volatility dynamics for the S&P 500: evidence for infinite-activity jumps with non-affine volatility dynamics from stock and option markets. Rev. Finance 21 (2), 811–844.

Yang, N., Chen, N., Wan, X., 2019. A new delta expansion for multivariate diffusions via the Ito-Taylor expansion. J. Econom. 209 (2), 256–288.

Yu, J., 2007. Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan. J. Econom. 141 (2), 1245–1280.