Approximate Maximum-Likelihood for Multivariate Jump-Diffusion Models with Applications to Finance

Dennis Kristensen* Young Jun Lee† Antonio Mele‡

February 6, 2020

Abstract

This paper develops closed-form approximations of the transition densities for a general class of jump-diffusion processes that are of common usage in financial economics. These approximations rely on a series expansion of the true transition density around an auxiliary transition density which is known in closed form. Based on these expansions, we derive approximate maximum-likelihood estimators (MLEs) of the model parameters. Under regularity conditions, the approximate transition density converges towards the true density, as the number of terms in the series expansion grows. Thus, the corresponding approximate MLE converges towards the exact MLE as well. In a number of numerical examples regarding models of option evaluation and the term structure of interest rates, our method is shown to be accurate and to require only modest computation time.

JEL Classification: C13; C32; C63; G12; G13.

Keywords: Continuous-time models, jump-diffusion, transition density, stochastic volatility, closed-form approximations, maximum-likelihood estimation.

*Department of Economics, University College London; E-mail: d.kristensen@ucl.ac.uk

†Department of Economics, University of Copenhagen; E-mail: yjl@econ.ku.dk

‡Swiss Finance Institute, USI Lugano and CEPR; E-mail: antonio.mele@usi.ch

1 Introduction

Continuous-time jump-diffusion processes are used in economics and finance to describe the dynamics of state variables underlying the models of interest (see, e.g., Björk, 2009). They lead to a simple and elegant analysis of problems such as the pricing of financial assets, stochastic control and other dynamic phenomena. The main disadvantage of jump-diffusion processes is that characteristics such as moments and densities cannot be expressed in closed form except in a few special cases (most notably the class of affine models). This hampers their practical use and implementation. In particular, when taking jump-diffusion models to data, the econometrician normally has to rely on simulations in order to compute generalized method of moments (GMM) or maximum-likelihood (ML) estimators; see, for example, Beskos et al. (2006), Beskos et al. (2009), Brandt and Santa-Clara (2002), Durham and Gallant (2002), Elerian et al. (2001), Kristensen and Shin (2012) and Sermaidis et al. (2013).

We propose an approach to the approximation of transition densities of general, multivariate Markov jump-diffusion models. The approximations are in closed form, simple to implement, and computationally very fast. Under regularity conditions, the approximate densities converge to the true densities as the degree of approximation grows finer. This in turn implies that the corresponding approximate MLE is asymptotically equivalent to the exact, infeasible MLE as the degree of approximation grows. These theoretical results are illustrated through a number of numerical examples; we show that our method is quite accurate in practice and that the approximation error goes to zero quickly as we increase the degree of approximation. Moreover, the method is computationally efficient and can be implemented on a standard desktop with ease.

The approximation scheme takes as a starting point some auxiliary jump-diffusion model chosen by the researcher for which the transition density is known (as a leading case, think of this as an affine model). We then derive an explicit expression of difference between the exact, unknown transition density of interest and the auxiliary transition density. This difference takes the form of a conditional moment which, under regularity conditions, can be expressed as a series expansion in terms of the infinitesimal operator of the true model. We then propose to approximate this difference, and thereby the unknown transition density, by only including the first, few (say, five) leading terms of the series expansions. Under regularity conditions, the series expansion of this conditional moment, and thereby our approximate transition density, converge to the truth. Furthermore, we establish error bounds for any given finite number of terms which allows the researcher to gauge the degree of error incurred by using a given degree of approximation.

With the approximate transition density in hand, we can construct an approximate log-likelihood function for a given discretely sampled parametric jump-diffusion model. We call the maximizer of this approximate likelihood the approximate MLE. We derive the asymptotic properties of this estimator as sample size and number of approximate terms grows. In particular, we develop approximation rules under which the approximate MLE is first-order asymptotically equivalent to the exact, infeasible MLE.

Our approach to the approximation of the jump-diffusion density is related to the ideas originally

proposed by Aït-Sahalia (2002, 2008) and further developed in Bakshi et al. (2006), Egorov et al. (2003) and Yu (2007); see Aït-Sahalia and Kimmel (2007, 2010), Xiu (2010) and Yang (2006) for applications to estimation of interest rate and stochastic volatility models. These papers also develop approximations of transition densities based on Taylor-series type expansions. However, these existing expansions are in closed-form only in the univariate case, and in general require numerical solutions to ordinary differential equations (ODEs). This makes the implementation more difficult and computationally more burdensome. Moreover, in the general, multivariate case, the theory underlying the proposed expansions requires that the time distance between observations shrinks to zero.

By contrast, our approach leads to closed-form approximations that are readily implementable without having to solve ODEs. As such, our method is computationally more efficient in the general case. Moreover, we show consistency of our method for fixed time distance between observations. The novel point that leads to these convenient features relative to existing methods is the insight that for any two jump-diffusion models, the difference between their transition densities can be expressed as a conditional moment. Based on this observation, it is, then, straightforward to develop simple series expansions of the conditional moment in question. This insight was also employed in the closely related work of Kristensen and Mele (2011) who use the same ideas as here to develop closed-form approximations of option and bond prices in a continuous-time setting.

Finally, we would like to emphasize that our approach is radically different from other estimation methods that rely on auxiliary models such as indirect inference (II) and efficient method of moments (EMM) developed by, amongst others, Gourieroux et al. (1993) and Gallant and Tauchen (1996). We use auxiliary models as a *numerical* device to obtain approximations of the unknown likelihood; as the degree of approximation increases, we are effectively working with the correctly specified likelihood function. In contrast, II and EMM methods use auxiliary models as a statistical device, and the error from using the auxiliary one is a statistical one. One could potentially combine the two methods, thereby taking into account the approximation error in our proposal whilst constructing confidence bands etc.

The paper is organized as follows. Section 2 sets the stage and presents our proposed approximation methods for transition densities. Section 3 provides details regarding the practical implementation of our methods. Section 4 analyzes theoretical properties of the approximate densities and parameter estimators based thereon. Section 5 tests the numerical performance of our methods on a number of models in both the equity option space and the term structure of interest rates. Section 6 concludes. An Appendix gathers all proofs.

2 Closed-form approximations of transition density and likelihood

2.1 Transition densities

In this section, we derive a general approximation formula for transition densities of jump-diffusion models that cannot be solved for in closed-form. We consider a multivariate model in which a d-dimensional process, $x(t) \in \mathbb{R}^d$ solves the following stochastic differential equation (SDE):

$$dx(t) = \mu(x(t), t)\, dt + \sigma(x(t), t)\, dW(t) + J(x(t), t)\, dN(t), \tag{2.1}$$

where $W(t)$ is a d-dimensional standard Brownian motion, $N(t)$ is a jump process with intensity $\lambda(x(t), t)$, $\mu(x,t)$ and $\sigma(x,t)$ are some drift and diffusion functions, and $J(x,t)$ captures the jump-sizes and has density $v(x,t)$.

Define the infinitesimal generator L associated with Eq. (2.1),

$$\begin{aligned} Lf(x,t) &= \frac{\partial f(x,t)}{\partial t} + \sum_{i=1}^{d} \mu_i(x,t) \frac{\partial f(x,t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sigma_{ij}^2(x,t) \frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} \\ &\quad + \lambda(x(t), t) \int_{\mathbb{R}^d} [f(x+c,t) - f(x,t)]\, v(c,t)\, dc, \end{aligned} \tag{2.2}$$

where $\sigma^2(x,t) := \sigma(x,t)\,\sigma(x,t)^\top \in \mathbb{R}^{d\times d}$. Under regularity conditions, the Markov transition density, $p(y,t|x,s)$, can be expressed as the solution to the following partial differential equation (PDE):

$$Lp(y,u|x,t) = 0, \tag{2.3}$$

with boundary condition $p(y,t|x,t) = \delta(y-x)$, where $\delta(x)$ is the Dirac's delta function. In most cases, this PDE cannot be solved for in closed form and so the transition density has to be computed using numerical approximations.

Our approximation scheme takes as a starting point an auxiliary model that, in contrast to the model of interest, has a transition density that can be solved analytically. Let $x_0(t) \in \mathbb{R}^d$ solve the following SDE,

$$dx_0(t) = \mu_0(x_0(t), t)\, dt + \sigma_0(x_0(t), t)\, dW(t) + J_0(x(t), t)\, dN(t), \tag{2.4}$$

for some drift, diffusion and jump functions $\mu_0(x,t)$, $\sigma_0(x,t)$ and $J_0(x,t)$. Our objective is to obtain a suitable expansion of the transition density of interest around the one associated with the auxiliary model.

We assume that the dimension of the auxiliary model is the same as that of the actual model, i.e., $x_0(t) \in \mathbb{R}^d$. This assumption does not entail any loss of generality since we can always add constant components: Suppose that the chosen auxiliary model has a lower dimension where the

state vector $y(t) \in \mathbb{R}^m$, with $m < d$, solves, for some drift and diffusion functions μ_Y and σ_Y:

$$dy(t) = \mu_Y(y(t), t)\, dt + \sigma_Y(y(t), t)\, dW_Y(t) + q_Y(y(t), t)\, dJ_Y(t),$$

where $W_Y(t)$ and $J_Y(t)$ are a standard Brownian motion and jump process, respectively. This auxiliary model can be rewritten in the form of Eq. (2.4) by introducing an additional process $z(t) \in \mathbb{R}^{d-m}$ which solves

$$dz(t) = \mu_Z(z(t), t)\, dt + \sigma_Z(z(t), t)\, dW_Z(t) + q_Z(z(t), t)\, dJ_Z(t),$$

where $(W_Z(t), J_Z(t))$ and the $(W_Y(t), J_Z(t))$ are mutually independent. Thus, $y(t)$ and the $z(t)$ are mutually independent. We then choose $x_0(t) := \left(y(t)^\top, z(t)^\top\right)^\top \in \mathbb{R}^d$ which is a Markov jump-diffusion process with transition density $p_0((y,z), u|(y_0,z_0), t) = p_Y(y, u|y_0, t)\, p_Z(z, u|z_0, t)$, where $p_Y(y, u|y_0, t)$ and $p_Z(z, u|z_0, t)$ are the transition densities of $y(t)$ and $z(t)$, respectively.

Given the auxiliary model in Eq. (2.4), we introduce its corresponding infinitesimal operator,

$$\begin{aligned} L_0 f(x,t) &= \frac{\partial f(x,t)}{\partial t} + \sum_{i=1}^{d} \mu_{0,i}(x,t) \frac{\partial f(x,t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sigma^2_{0,ij}(x,t) \frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} \\ &\quad + \lambda_0(x(t), t) \int_{\mathbb{R}^d} [f(x+c,t) - f(x,t)]\, v_0(c,t)\, dc, \end{aligned} \tag{2.5}$$

and note that the auxiliary model's transition density, $p_0(y, u|x, t)$ say, solves

$$L_0 p_0(y, u|x, t) = 0, \tag{2.6}$$

with boundary condition $p_0(y, t|x, t) = \delta(y - x)$.

After straightforward algebra, one finds that the difference between the two transition densities, $\Delta p(y, u|x, t) \equiv p(y, u|x, t) - p_0(y, u|x, t)$, satisfies the following PDE:

$$L \Delta p(y, u|x, t) + D(y, u|x, t) = 0, \tag{2.7}$$

with boundary condition $\Delta p(y, u|x, t) = 0$. Here, the function D is defined as $D(y, u|x, t) := (L - L_0)\, p_0(y, u|x, t)$; that is,

$$\begin{aligned} D(y, u|x, t) &= \sum_{i=1}^{d} \Delta\mu_i(x,t) \frac{\partial p_0(y, u|x, t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \Delta\sigma^2_{ij}(x,t) \frac{\partial^2 p_0(y, u|x, t)}{\partial x_i \partial x_j} \\ &\quad + \Delta\lambda(x,t)\, I_1(y, u|x, t) + \lambda(x,t)\, I_2(y, u|x, t), \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} \Delta\mu_i(x,t) &= \mu_i(x,t) - \mu_{0,i}(x,t), \quad \Delta\sigma^2_{ij}(x,t) = \sigma^2_{ij}(x,t) - \sigma^2_{0,ij}(x,t), \\ \Delta\lambda(x,t) &= \lambda(x,t) - \lambda_0(x,t), \quad \Delta v(c,t) = v(c,t) - v_0(c,t), \end{aligned}$$

and

$$\begin{aligned} I_1(y,u|x,t) &= \int_{\mathbb{R}^d} \left[p_0(y,u|x+c,t) - p_0(y,u|x,t)\right] v_0(c,t)\, dc, \\ I_2(y,u|x,t) &= \int_{\mathbb{R}^d} \left[p_0(y,u|x+c,t) - p_0(y,u|x,t)\right] \Delta v(c,t)\, dc. \end{aligned}$$

Because the auxiliary model has its transition density $p_0(y,u|x,t)$ that is known in closed form, then, by construction, the function $D(y,u|x,t)$ is also known in closed form, provided we can compute the two integrals $I_1(y,u|x,t)$ and $I_2(y,u|x,t)$ analytically.

Under standard regularity conditions, we can apply the Feynman-Kac representation of the solution to the PDE in Eq. (2.7), $\Delta p(y,u|x,t)$, to obtain a representation of the unknown density, $p(y,u|x,t)$, in terms of that arising through the auxiliary model, $p_0(y,u|x,t)$. In order to be able to appeal to the Feynman-Kac representation, we assume that:

A.1 There exist solutions $p(y,u|x,t)$ and $p_0(y,u|x,t)$ to the two PDEs (2.3) and (2.6), which belong to $\mathcal{C}^{2,1}\left(\mathbb{R}^d \times [0,T]\right)$. Furthermore, for some $C, q > 0$:

$$|p(y,u|x,t) - p_0(y,u|x,t)| \leq C\left(1 + \|x\|^q\right),$$

for all $x, y \in \mathbb{R}^d$ and $0 \leq t < u \leq T$.

A.2 The function $D(y,u|x,t)$ is continuous and polynomially bounded, $D(y,u|x,t) \leq C\left(1+\|x\|^q\right)$ for all $x, y \in \mathbb{R}^d$ and $0 \leq t < u \leq T$.

These two conditions ensure that $p(y,u|x,t)$ and $p_0(y,u|x,t)$ exist as solutions to the two PDE's in Eqs. (2.3) and (2.6). Primitive sufficient conditions for the transition densities to solve these are given in Appendix A. Among these is a linear growth condition imposed on the drift and diffusion terms, which is violated in some cases. However, this growth condition is only needed to ensure that the solutions $x(t)$ and $x_0(t)$ to Eqs. (2.1) and (2.4) exist. Other conditions than the linear growth can be used to ensure these solutions do actually exist.

Theorem 2.1 *Assume that A.1-A.2 hold. Then the following identity holds for all $x, y \in \mathbb{R}^d$ and $0 \leq t < u \leq T$:*

$$p(y,u|x,t) = p_0(y,u|x,t) + \int_t^u \mathbb{E}_{x,t}\left[D(y,u|x(s),s)\right] ds, \tag{2.9}$$

where $x(t)$ satisfies Eq. (2.1), and $D(y,u|x,t)$ is given in Eq. (2.8).

The second term of the right-hand side of Eq. (2.9) delivers an exact expression for the error resulting while using the transition density of the auxiliary model in place of the true one. An appealing property of the error term is that the function $D(y,u|x,t)$ is a smooth function whenever the underlying drift, diffusion and jump component of the true and auxiliary models are smooth. Thus, for such models, we may approximate this error term $\int_t^T \mathbb{E}_{x,t}\left[D(y,u|x(s),s)\right] ds$ through a series expansion:

Definition 2.2 *Assume that the true and auxiliary models are both smooth. The N-th order approximation $p_N(y,u|x,t)$ to the unknown transition density $p(y,u|x,t)$ is given by:*

$$p_N(y,u|x,t) := p_0(y,u|x,t) + \sum_{n=0}^{N} \frac{(u-t)^{n+1}}{(n+1)!} L^n D(y,u|x,t), \tag{2.10}$$

where $L^n D(y,u|x,t) = L^{n-1} D(y,u|x,t)$, $n \geq 0$, with $L^0 D(y,u|x,t) = D(y,u|x,t)$.

Comparing Eqs. (2.9) and (2.10), we see that the integral has been replaced by a Taylor-type expansion. This type of Taylor expansion have found widespread use in econometrics and finance in the implementation of jump-diffusion models (see references cited in the introduction). The innovation of our approach relative to existing uses of such expansions is the insight that we can explicitly formulate the difference between the transition density of any two jump-diffusion models as a conditional moment which in turn can be approximated. This has two important consequences: First, and under suitable regularity conditions, the above approximation works for basically any jump-diffusion model. Second, the computation of the approximation is straightforward requiring only computation of $L^n D(y,u|x,t)$. Without any jump component, this only involves repeated differentiation of $D(y,u|x,t)$. With jump components, one needs to be able to compute integrals $I_1(y,u|x,t)$ and $I_2(y,u|x,t)$ in (2.8). These two terms can be computed either using numerical integration or Monte Carlo methods as discussed in more detail below.

Note, finally, that the expansion used in Definition 1 is only one method for approximating the right-hand side of Eq. (2.9) in Theorem 2.1. Other methods are available. For example, one could approximate the conditional expectations appearing in the right-hand side of Eq. (2.9) using Monte Carlo methods; see Appendix A for the precise algorithm.

2.2 Likelihoods

Next, we assume that the jump-diffusion model is characterized by some parameter vector $\theta \in \Theta \subseteq \mathbb{R}^k$,

$$dx(t) = \mu(x(t), t; \theta)\, dt + \sigma(x(t), t; \theta)\, dW(t) + q(x(t), t; \theta)\, dJ(t).$$

Given discrete observations of the process at time points $t_0 < t_1 < t_2 < ... < t_n$, $\{x(t_i) : i = 1, ..., n\}$, the exact maximum-likelihood estimator (MLE) is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta), \quad L_n(\theta) = \sum_{i=1}^{n} \log p(x(t_i), t_i | x(t_{i-1}), t_{i-1}; \theta),$$

where $p(y,u|x,t;\theta)$ solves (2.3) for a given value of $\theta \in \Theta$, and for some compact set Θ (see, also, condition B.2 in Section 4.2). Our approximate MLE replaces the unknown density by its approximation $p_N(y,u|x,t;\theta)$ developed in the previous section, which is now also indexed by θ. Furthermore, note that the auxiliary model shall generally include its own parameters. These parameters can be fixed at any set of values that ensure our expansions are still valid. We ex-

plain methods for choosing these values in the numerical experiments of Section 5. Formally, our estimator is defined as follows. Given the approximate transition density, the approximate MLE is

$$\hat{\theta}_{n,N} = \arg\max_{\theta\in\Theta} L_{n,N}(\theta), \quad L_{n,N}(\theta) = \sum_{i=1}^{n} \log p_N\left(x(t_i), t_i | x(t_{i-1}), t_{i-1}; \theta\right).$$

2.3 Bayesian methods

Alternatively, we may use our expansions and implement Bayesian methods. To illustrate, for a given prior $\pi(\theta)$, we may employ the MCMC algorithm proposed by **?** and obtain a Monte Carlo approximation of the (approximate) posterior distribution as

$$f_{n,N}(\theta) = \frac{\exp\left(L_{n,N}(\theta)\right)\pi(\theta)}{\int_\Theta \exp\left(L_{n,N}(\theta)\right)\pi(\theta)\, d\theta}.$$

This can in turn be used to compute point estimators such as the posterior mean, $\bar{\theta} = \int \theta f_{n,N}(\theta)\, d\theta$, and confidence intervals, such as the posterior quantiles. Our theory will focus on the properties of the approximate MLE, but most of the arguments carry over to the analysis of the above approximate posterior distribution.

3 Practical implementation

This section provides details regarding the practical implementation of our approximation methods. We first deal with the simple pure diffusion setting and, then, with the jump-diffusion case.

3.1 Diffusions

In the pure diffusion case, analytical expressions of $L^n D(y, u|x, t)$ are straightforward to obtain relying on symbolic software packages, such as Mathematica. In this case, indeed, the application of the infinitesimal operator L only involves taking partial derivatives of $D(y, u|x, t)$. To illustrate, suppose that $d = 1$ and both true and auxiliary model are homogenous, in that drifts and diffusions are independent of t; the leading term is, then, and suppressing the dependence on (y, u, x, t),

$$L^0 D = D = \Delta\mu\frac{\partial p_0}{\partial x} + \frac{1}{2}\Delta\sigma^2\frac{\partial^2 p_0}{\partial x^2}.$$

The first-order term takes the form

$$L^1 D = LD = \frac{\partial D}{\partial t} + \mu\frac{\partial D}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 D}{\partial x^2},$$

where

$$\begin{aligned}
\frac{\partial D}{\partial t} &= \Delta\mu\left(x\right)\frac{\partial^2 p_0}{\partial x \partial t} + \frac{1}{2}\Delta\sigma^2\left(x\right)\frac{\partial^3 p_0}{\partial x^2 \partial t}, \\
\frac{\partial D}{\partial x} &= \mu\left[\frac{\partial \Delta\mu}{\partial x}\frac{\partial p_0}{\partial x} + \Delta\mu\frac{\partial^2 p_0}{\partial x^2} + \frac{1}{2}\frac{\partial \Delta\sigma^2}{\partial x}\frac{\partial^2 p_0}{\partial x^2} + \frac{1}{2}\Delta\sigma^2\frac{\partial^3 p_0}{\partial x^3}\right] + \frac{1}{2}\sigma^2\frac{\partial^2 \Delta\mu}{\partial x^2}\frac{\partial p_0}{\partial x} \\
&\quad + \frac{1}{2}\sigma^2\left[2\frac{\partial \Delta\mu}{\partial x}\frac{\partial^2 p_0}{\partial x^2} + \Delta\mu\frac{\partial^3 p_0}{\partial x^3} + \frac{1}{2}\frac{\partial^2 \Delta\sigma^2}{\partial x^2}\frac{\partial^2 p_0}{\partial x^2} + \frac{\partial \Delta\sigma^2}{\partial x}\frac{\partial^3 p_0}{\partial x^3} + \frac{1}{2}\Delta\sigma^2\frac{\partial^4 p_0}{\partial x^4}\right],
\end{aligned}$$

and similarly for $\partial^2 D/\partial x^2$. One then continues sequentially to derive additional higher-order terms as desired. As can be seen from the above expression, the expansion quickly gets lengthy in the general case. Fortunately, in most specific cases, identical terms will appear in the expansion which can be collected together thereby reducing the length. Mathematica does this automatically and so in many cases, third and fourth order expansion is manageable. In addition, Mathematica enables one to translate and export symbolic formulas into software packages such as Matlab, which may then be used for fast numerical implementation of the approximate density and the associated approximate MLE.

3.2 Jump-diffusions

Next, consider the extension to the jump-diffusive case. First, write $L = L_D + L_J$ where

$$L_D f\left(x,t\right) = \frac{\partial f\left(x,t\right)}{\partial t} + \sum_{i=1}^{d}\mu_i\left(x,t\right)\frac{\partial f\left(x,t\right)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d}\sigma_{ij}^2\left(x,t\right)\frac{\partial^2 f\left(x,t\right)}{\partial x_i \partial x_j}$$

and

$$L_J f\left(x,t\right) = \lambda\left(x,t\right)\int_{\mathbb{R}^d}\left[f\left(x+c,t\right) - f\left(x,t\right)\right]v\left(c,t\right)dc$$

are the operators corresponding to the diffusive and jump component, respectively. For a given $n \geq 1$, observe that, because both operators are linear, we may interchange the order of differentiation and integration,

$$L^n D = \left(L_D + L_J\right)^n D = \sum_{i=0}^{n}\binom{n}{i}L_J^i L_D^{n-i} D.$$

For a given $N \geq 1$, we have, therefore, that

$$p_N\left(y,u|x,t\right) = p_0\left(y,u|x,t\right) + \sum_{n=0}^{N}\sum_{i=0}^{n}\frac{\left(u-t\right)^{n+1}}{\left(n+1\right)!}\binom{n}{i}L_J^i D_{n-i}\left(y,u|x,t\right), \tag{3.1}$$

where $D_i\left(y,u|x,t\right) := L_D^i D\left(y,u|x,t\right)$ is known analytically for all $i = 1,....,N$, and can be evaluated using a symbolic software package just as in the pure diffusion setting. What remains to compute is $L_J^i D_{n-i}\left(y,u|x,t\right)$, for all $i = 1,....,N$. This is a Riemann integral that is generally not available in closed form. However, note that, for a generic function $f\left(x,t\right)$ (in our case, $f\left(x,t\right) = D_{n-i}\left(y,u|x,t\right)$), $L_J^i f\left(x,t\right)$ can be approximated using sequential numerical integration

or Monte Carlo simulation. For example, letting $c_1(t), ..., c_M(t)$ be $M \geq 1$ i.i.d. draws from $v(c,t)$,

$$\hat{L}_J f(x,t) = \lambda(x,t) \frac{1}{M} \sum_{k_1=1}^{M} \left[f(x + c_{k_1}(t), t) - f(x,t) \right], \tag{3.2}$$

and

$$\begin{aligned} \hat{L}_J^2 f(x,t) &= \lambda(x,t) \frac{1}{M} \sum_{k_2=1}^{M} \left[\lambda(x + c_{k_2}(t), t) \frac{1}{M} \sum_{k_1=1}^{M} \left[f(x + c_{k_2}(t) + c_{k_1}(t), t) - f(x,t) \right] \right] \\ &\quad -\lambda^2(x,t) \frac{1}{M} \sum_{k_2=1}^{M} \left[\frac{1}{M} \sum_{k_1=1}^{M} \left[f(x + c_{k_1}(t), t) - f(x,t) \right] \right], \end{aligned}$$

are consistent estimators of $L_J f(x,t)$ and $L_J^2 f(x,t)$, respectively.

4 Theoretical properties

Under regularity conditions, we derive error bounds for the proposed approximate density. These bounds are in turn used to analyze the properties of the corresponding approximate MLE.

4.1 Properties of Approximate densities

We provide two theoretical results for $p_N(y, u|x, t)$ as defined in Eq. (2.10): The first result gives a pointwise error bound for the approximation for a given choice of N. The second one shows that indeed the approximation error vanishes as $N \to \infty$. As such, the approximation procedure is consistent. Both results follow from slight extensions of the general results found in the Appendix of Kristensen and Mele (2011).

The following results rely on the assumption that $L_J^i D_{n-i}(y, u|x, t)$ in Eqs. (3.1) can be evaluated without error. When this assumption does not hold, we will need to include one additional term in all error bounds provided below, which take into account any numerical errors made in the computation of $L_J^i D_{n-i}(y, u|x, t)$. For example, in the case where M Monte Carlo draws are employed to evaluate the integrals, as in Eq. (3.2), an additional error of order $O_P\left(1/\sqrt{M}\right)$ will appear.

The first proposition establishes an error bound for the approximation, which holds for any fixed approximation order $N \geq 1$:

Proposition 4.1 *Assume that A.1-A.2 hold; that μ, σ^2, λ and v are time-homogeneous, $\mu, \sigma^2, \lambda, v \in \mathcal{C}^{2N}(\mathbb{R}^d)$ and $D \in \mathcal{C}^{2(N+1)}(\mathbb{R}^d)$. Then p_N given in Definition 2.2 satisfies:*

$$|p_N(y, u|x, t) - p(y, u|x, t)| \leq E_N(x, y, u) \frac{(u-t)^{N+1}}{(N+1)!}, \text{ for all } (x, y) \in \mathbb{R}^{2d} \text{ and } 0 < t < u,$$

where

$$E_N(x,y,u) := \sup_{0\le t\le u} \mathbb{E}_{x,t}\left[\left\|L^{N+1}D(y,u|x(s),t)\right\|\right],$$

and L is the infinitesimal generator of $x(t)$. In particular, if μ, σ^2, λ and D are polynomially bounded, then, there exists $\bar\tau > 0$ and $\phi_{y,u} \in \mathcal{H}$ such that for $D : \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}$ defined in Eq. (2.8), it holds that

$$E_N(x,y,u) \le (1 + \|x\|^{q_N} + \|y\|^{q_N})\, e^{c_N u},$$

for some constants c_N and q_N.

This result tells us that in great generality the error decreases at a geometric rate uniformly over (x,t) in any compact interval as N increases. Florens-Zmirou (1989, Lemma 1) and Aït-Sahalia (2002) develop similar error bounds for approximations of conditional moments of diffusion processes in different contexts. In particular, letting $\Delta = u - t \to 0$, we see that the error vanishes. In practice, however, the time distance between observations is fixed.

Proposition 4.1 is not informative about the asymptotic behavior of the error terms as $N \to \infty$. In particular, we have not been able to establish bounds on q_N and c_N as N increases. To deal with the error terms for large N, we rely, instead, on results from the literature on operator theory (see Pazy, 1983). First, we introduce some additional notation and definitions. We define the spectrum and resolvent of the infinitesimal operator L as

$$\sigma(L) = \{\lambda \in \mathbb{C} : (\lambda - L) \text{ is not a bijection}\} \quad \text{and} \quad R_\lambda(L) = (\lambda - L)^{-1},\ \lambda \in \sigma(L).$$

We then introduce a function space $\mathcal{H}$, which is equipped with some function norm $\|\cdot\|_{\mathcal{H}}$ and impose the following conditions on the spectrum and resolvent in order to show that our power series expansion converges:

A.3 Given three constant $m, \omega > 0$ and $M \in \left(e^{-1}, \infty\right)$, the infinitesimal operator L given in Eq. (2.2) satisfies

$$\sigma(L) \subset \bar\sigma \equiv \{\lambda \in \mathbb{C} : |\arg(\lambda - \omega)| > \pi/2 + m\},$$

and its resolvent satisfies $\|R_\lambda(L)\| \le M/|\lambda|$ for $\lambda \in \mathbb{C}\backslash\bar\sigma(L)$.

A.4 There exists $\bar\tau > 0$ and $\phi_{y,u} \in \mathcal{H}$ such that for $D : \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}$ defined in Eq. (2.8) it holds that

$$\mathbb{E}\left[\phi_{y,t}(x(\bar\tau))\,|x(0) = x\right] = D(y,t|x,\bar\tau).$$

Moreover, the function $t \mapsto D(y,u|x,t)$ is analytic uniformly in $\|\cdot\|_{\mathcal{H}}$.

Condition (A.3) relates to the infinitesimal operator and requires that its spectrum is within $\bar\sigma$. The second condition, (A.4), imposes conditions on the error function D whose conditional moment we wish to expand. It basically requires that D can be matched through conditional moments. Both assumptions are abstract, and not easily verified for specific models. The following

proposition provides more primitive conditions for (A.3) to hold that are typically met in many diffusion models.

Proposition 4.2 *The generator L satisfies (A.3), under the following conditions:*

(i) L has a transition density $p(y,u|x,t)$ with respect to Lebesgue measure.

(ii) L has an invariant measure π satisfying

$$\pi(x)\,p(y,u|x,t) = \pi(y)\,p(x,u|y,t).$$

Conditions (i)-(ii) in Theorem 4.2 are satisfied by many standard processes used in finance. Most jump-diffusion models have a transition density, while the second condition is a generalization of time-reversibility. In particular, if the process is univariate and stationary, it is necessarily time-reversible and therefore satisfies the second condition. In conclusion, condition (A.3) holds under fairly weak conditions. We are unaware of more primitive conditions under which (A.4) holds. For a simple example where (A.2) is not satisfied, we refer to Schaumburg (2004, Example 1), who also provides additional discussion of this condition.

The following proposition ptovides theoretical foundations to the approximation in Definition 2.2:

Proposition 4.3 *Assume that A.1-A.4 hold. Then for any $|t-u| < \bar{\tau}/(Me)$, where M and $\bar{\tau}$ are given in (A.3)-(A.4), we have:*

$$\int_t^u \mathbb{E}_{x,t}\left[D(y,u|x(s),s)\right]ds = \sum_{n=0}^{\infty}\frac{(u-t)^{n+1}}{(n+1)!}D_n(y,u|x,t),$$

where D_n is given in Definition 2.2. Furthermore, the approximation p_N given in Definition 2.2 satisfies for all $|t-u| < \bar{\tau}/(Me)$,

$$\|p_N(y,u|\cdot,t) - p(y,u|\cdot,t)\|_{\mathcal{H}} \to 0, \quad N \to \infty.$$

The first part of the proposition states that under A.1-A.4, the series expansion of the integrated conditional moment indeed converges. The second part, that the approximate transition density converges, follows as a straightforward consequence of the first part. Note that in the case where M Monte Carlo draws are used to compute $L_J^i D_{n-i}(y,u|x,t)$, we need as additional requirment that $M \to \infty$.

While the above result shows consistency of the proposed approximation method as the number of terms in the series expansion grows, it is unfortunately silent about the convergence rate. It would be useful to obtain a uniform convergence rate in terms of N since this could be used to choose N to obtain a given level of error tolerance. Such extensions are left for future research.

4.2 Properties of Approximate MLE

We analyze the properties of the approximate MLE based on the series expansion of the transition density. We rely on the results in the previous section, combined with general results for approximate MLEs developed in Kristensen and Shin (2012).

We focus on the case where $x(t)$ is stationary and ergodic. We may also extend our results to the non-ergodic case, although the conditions would become a lot more complicated in this case. Therefore, we restrict ourselves to the leading case of stationarity; in particular, we rule out time-inhomomogenous processes. For notational convenience, we assume that the observations are equidistant in time and normalize time distance between observations to unity, $t_i - t_{i-1} = 1$. In the following, we write $p(y|x;\theta) := p(y, t_i|x, t_{i-1};\theta)$.

For merely technical reasons, we need to re-define the approximate MLE to include trimming-away of observations for which the approximate density is close to zero:

$$\hat{\theta}_{n,N} = \arg\max_{\theta\in\Theta} L_{n,N}(\theta), \quad L_{n,N}(\theta) = \sum_{i=1}^{n} \tau_{N,i}(a,\theta) \log p_N(x(t_i)|x(t_{i-1});\theta),$$

where

$$\tau_{N,i}(a,\theta) := \tau(p_N(x(t_i)|x(t_{i-1});\theta))$$

and $\tau_a(\cdot)$ is a continuously differentiable trimming function satisfying $\tau_a(z) = 1$ if $|z| > a$, and 0 if $|z| < a/2$. Here, $a = a(N) \to 0$ is a trimming sequence that controls the amount of trimming used. For example, we may choose $\tau_a(\cdot)$ as the indicator function in which case all observation pairs for which $p_N(x(t_i)|x(t_{i-1});\theta) \leq a/2$ are excluded in the computation of the approximate MLE.

Next, we impose the following regularity conditions on the model:

B.1 The process $x(t)$ is stationary and ergodic with $E[\|x(t)\|] < \infty$.

B.2 The parameter space Θ compact.

B.3 $|\log p(y|x;\theta)| \leq b_1(y|x)$, $\forall\theta\in\Theta$, with $\mathbb{E}\left[b_1(x(t_i)|x(t_{i-1}))^{1+\delta}\right] < \infty$ and

$$\mathbb{E}\left[\log p(x(t_i)|x(t_{i-1});\theta)\right] < \mathbb{E}\left[\log p(x(t_i)|x(t_{i-1});\theta_0)\right], \quad \forall\theta\neq\theta_0.$$

B.4 The matrix

$$i(\theta_0) = \mathbb{E}\left[\frac{\partial \log p(x(t_i)|x(t_{i-1});\theta_0)}{\partial\theta}\frac{\partial \log p(x(t_i)|x(t_{i-1});\theta_0)}{\partial\theta'}\right]$$

exists and is nonsingular.

B.5 $\|\partial^2 \log p(y|x;\theta)/(\partial\theta\partial\theta')\| \leq b_2(y|x)$ for all θ in a neighborhood of θ_0 with $\mathbb{E}[b_2(x(t_i)|x(t_{i-1}))] < \infty$.

Proposition 4.4 *Assume that A.1-A.4 and B.1-B.5 hold. Then, for any given sequence $B_N \to \infty$, the Approximate MLE satisfies for*

$$\begin{aligned} ||\hat{\theta}_{n,N} - \hat{\theta}_n|| &= O_P\left(|\log a| B_N^{-1-\delta}\right) + O_P\left(|\log a|^{-\delta}\right) \\ &\quad + O_P\left(a^{-1} \times \sup_{\theta\in\Theta} \sup_{y\in\mathbb{R}^d} \sup_{\|x\|\leq B_N} |p_N(y|x;\theta) - p(y|x;\theta)|\right). \end{aligned}$$

5 Numerical performance

We assess the performance of our transition density approximation in various jump-diffusion models. Firstly, we will consider specific univariate and multivariate diffusion models without jumps. We then consider auxiliary models and compare our approximation method with other methods. For the jump diffusion models, we select the same diffusion model, but without jumps, as an auxiliary model, and investigate how the jump diffusion process is approximated with our methods.

5.1 Cox-Ingersoll-Ross (CIR)

Consider the univariate Cox, Ingersoll and Ross (1985) (CIR) model

$$dx(t) = \beta(\alpha - x(t))\,dt + \sigma\sqrt{x(t)}dW(t),$$

where $\alpha > 0$, $\beta > 0$, and $\sigma^2 > 0$ are parameters that we collect in $\theta = \left(\alpha, \beta, \sigma^2\right)$. Provided that $q = 2\beta\alpha/\sigma^2 - 1 \geq 0$, then a well-defined solution exists on $(0, \infty)$. Conveniently, the transition density of this model is known in closed form:

$$p(y, u|x, t; \theta) = c\exp(-z - z_0)\left(\frac{z_0}{z}\right)^{q/2} I_q\left(2\sqrt{zz_0}\right),$$

where $c = 2\beta/\left(\sigma^2\left(1 - e^{-\beta(u-t)}\right)\right)$, $z = cxe^{-\beta(u-t)}$, $z_0 = cy$ and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

We will forgo this knowledge and approximate p using the Vasicek model as auxiliary model,

$$dx_0(t) = \beta_0(\alpha_0 - x_0(t))\,dt + \sigma_0 dW(t). \tag{5.1}$$

The transition density of this Gaussian model is given as

$$p_0(y, u|x, t; \theta_0) = \frac{1}{\sqrt{\pi\gamma_0^2}}\exp\left[-\frac{\left(y - \alpha_0 - (x - \alpha_0)e^{-\beta_0(u-t)}\right)^2}{\gamma_0^2}\right],$$

where $\gamma_0^2 = \sigma_0^2\left(1 - e^{-2\beta_0(u-t)}\right)/\beta_0$ and $\theta_0 = \left(\alpha_0, \beta_0, \sigma_0^2\right)$.

Since no jump components are present, and the drift functions of the true and auxiliary model can be chosen to be the same by setting $\alpha_0 = \alpha$ and $\beta_0 = \beta$, the adjustment term takes the very

simple form of:

$$D\left(y,u|x,t\right)=\frac{1}{2}\left(\sigma^{2}x-\sigma_{0}^{2}\right)\frac{\partial^{2}p_{0}\left(y,u|x,t;\theta_{0}\right)}{\sigma x^{2}}.$$

As an alternative, we could use as auxiliary model the Brownian motion,

$$\begin{aligned} dx_{0}\left(t\right) &= \mu_{0}dt+\sigma_{0}dW\left(t\right), \\ p_{0}\left(y,u|x,t;\theta_{0}\right) &= \frac{1}{\sqrt{2\pi\sigma_{0}^{2}\left(u-t\right)}}\exp\left[-\frac{\left(y-x-\mu_{0}\left(u-t\right)\right)^{2}}{2\sigma_{0}^{2}\left(u-t\right)}\right], \end{aligned}$$

in which case

$$D\left(y,u|x,t\right)=\left(\beta\left(\alpha-x\right)-\mu_{0}\right)\frac{\partial p_{0}\left(y,u|x,t;\theta_{0}\right)}{\sigma x}+\frac{1}{2}\left(\sigma^{2}x-\sigma_{0}^{2}\right)\frac{\partial^{2}p_{0}\left(y,u|x,t;\theta_{0}\right)}{\sigma x^{2}}.$$

We can improve the quality of the approximation by setting θ_0 for $p_n\left(y,u|x,t;\theta_0\left(\theta\right)\right)$ in two ways, (i) Setting both the drift and diffusion functions of the true and auxiliary models to be the same, or (ii) Setting the first two conditional moments of the true and auxiliary models to be the same. When we apply the second method to choose the auxiliary parameters, we cannot obtain conditional moments of the true model, although we may derive them through simulations of $x\left(t\right)$, which we perform relying on the Milstein's approximation scheme.

Our first exercise is to compare the approximated transition density with the true transition density with $(\alpha,\beta,\sigma)=(0.06,0.5,0.15)$. We compute the transition density $p\left(y,u|x,t\right)$ over a range of y values while fixing $x=\alpha$ and $u-t=1$. Approximated transition density is computed using up to 4th order terms of expansion. As an alternative approximation method for a benchmark, we also compute the transition density approximated by the nonparametric simulation method in Kristensen and Shin (2012). We use the same tuning parameters as in their numerical exercise in Section 4.1.1: 10 intermediate steps for Euler scheme and 500 simulated observations. Figure 1 plots the transition densities from the two approximation methods along with the true CIR transition density. Our approximation method with only 4th order expansion provides a quite accurate approximation to the true transition density, and the numerical accuracy is more satisfactory compared to that of nonparametric simulation method.

As already noted earlier, one of the main advantage of our new approximation methods is the computational efficiency. We compare the computation time of a single transition density evaluation using the true CIR, our new approximation method, and NPSL. We repeat the evaluation of transition density 1000 times and average the computation time measured in milliseconds. Up to 4th order expansion is used for the approximation of the transition density. For NPSL, we again use 10 intermediate steps for Euler scheme and 500 simulated observations. Table 1 shows that the new method can save computation time by an order of magnitude compared to existing approximation methods.

Next, we turn to the finite sample performance of the estimator. We first generate 1000 artificial data of three different sample sizes ($T=150,300,500$) from a CIR data generating process with

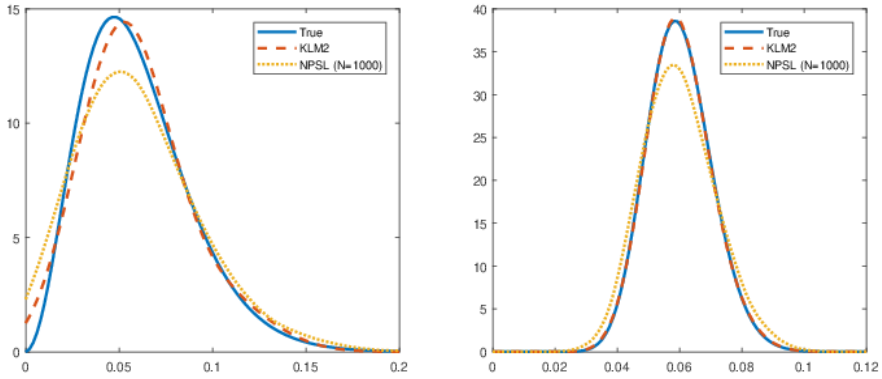

Figure 1: Approximation of transition density of CIR model using AL and NPSL at yearly (left) and monthly (right) frequency.

Table 1: Computation time for transition density using analytical expression, closed-form approximation and NPSL

	Evaluation Time	Relative to the True CIR
True CIR	0.02 ms	1.0
KLM	1.40 ms	80.3
NPSL	20.38 ms	1170.8

$(\alpha, \beta, \sigma) = (0.06, 0.5, 0.15)$. The simulation is based on Euler scheme with 10 intermediate steps, and the time distance between observations is $1/12$. Each sample path is used for estimation using true CIR likelihood and the approximated likelihood, respectively. We use the fmincon function in the Matlab Optimization Toolbox for the implementation of the estimator. As the likelihood of the CIR model is well-defined only for $q = 2\alpha\beta/\sigma^2 > 0$, we feed a nonlinear constraint function into the optimizer along with its analytic derivative information. The interior-point algorithm is used with the option 'AlwaysHonorConstraints' toggled on so as to guarantee that the nonlinear constraint is satisfied during the entire optimization procedure. As a final remark on the implementation, the true transition density of the CIR model is computed through a slightly modified form as follows:

$$p\left(y, u|x, t; \theta\right) = c \exp\left(2\sqrt{zz_0} - z - z_0\right) \left(\frac{z_0}{z}\right)^{q/2} \left[\frac{I_q\left(2\sqrt{zz_0}\right)}{\exp\left(2\sqrt{zz_0}\right)}\right],$$

where the last term in the bracket can be computed by the Matlab function `besseli`$(q, 2\sqrt{zz_0}, 1)$. The last argument, 1, indicates the Bessel function to return a value rescaled by the exponential term, which is guaranteed to be within a reasonable range.[1] Simulation results are shown in Table 2. For the α and σ estimates, we see that both bias and standard deviation and, hence, RMSE,

[1] Without rescaling, Matlab often fails to compute the CIR transition density correctly; while the value of the entire expression is in fact within a reasonable range, each term $I_q\left(2\sqrt{zz_0}\right)$ and $\exp\left(-z - z_0\right)$ separately can be beyond the precision of the floating-point to be correctly evaluated. Rescaling solves this problem.

Table 2: Performance of MLE's for CIR mode

	α			β			σ		
	KLM	NPSL	True	KLM	NPSL	True	KLM	NPSL	True
Bias	−0.0003	0.1876	−0.0002	0.0324	0.0390	0.0284	0.0000	0.0053	−0.0000
Std.dev.	0.0064	2.5106	0.0064	0.0949	0.2001	0.0914	0.0031	0.0116	0.0025
RMSE	0.0064	2.5176	0.0064	0.1003	0.2038	0.0957	0.0031	0.0128	0.0025

decrease with the sample size. The magnitude of the bias and the standard deviation is comparable to those of the true MLE, although the true MLE slightly outperforms AMLE.

Lastly, we perform a meta-analysis by comparing our results to the existing results in the Table 1 of Kristensen and Shin (2012). Their exercise provides a benchmark for the case of sample size $T = 300$ for the true MLE and the NPSMLE. The results are computed over 1000 artificial sample paths from a CIR data generating process with $(\alpha, \beta, \sigma) = (0.06, 0.5, 0.15)$ as in this paper. Their simulation is also based on Euler scheme with 10 intermediate steps, and the time distance between observations is 1/12. The final row of Table 2 shows the root mean-squared error (RMSE) of the approximated MLE (AMLE) along with the results for true MLE and NPSMLE reported in Kristensen and Shin (2012) for comparison. Of course, the true MLE outperforms both AMLE and NPSMLE, but all parameter estimators of AMLE with second order approximation significantly outperforms NPSMLE.

5.2 Aït-Sahalia (1996)

We now consider a more complicated univariate diffusion model as proposed in Aït-Sahalia (1996):

$$dx(t) = \left[a_0 + a_1 x(t) + a_2 x^2(t) + a_3/x(t)\right] dt + \sqrt{b_0 + b_1 x(t) + b_2 x^{b_3}(t)}\, dW(t). \tag{5.2}$$

with unknown parameters $\theta = (a, b)$ where $a = (a_0, a_1, a_2, a_3)$ and $b = (b_0, b_1, b_2, b_3)$. We choose again the Vasicek model in Eq. (5.1) as the auxiliary model. To ensure that the drift of Vasicek model approximate that of Aït-Sahalia's, we choose the auxiliary parameters (α_0, β_0) in the Vasicek model such that $\beta_0(\alpha_0 - x(t))$ is a linear approximation of the true drift evaluated at the point $x(t)$:

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + a_3/x \\ &\approx f(x(t)) + f'(x(t))(x - x(t)) = -f'(x(t))\left[\left(x(t) - \frac{f(x(t))}{f'(x(t))}\right) - x\right] \end{aligned}$$

Then, for each point $x(t)$, we set the auxiliary parameter values equal to

$$\begin{aligned} \alpha_0 &= x(t) - \frac{a_0 + a_1 x(t) + a_2 x^2(t) + a_3/x(t)}{a_1 + 2a_2 x(t) - a_3/x^2(t)}, \\ \beta_0 &= -\left(a_1 + 2a_2 x(t) - a_3/x^2(t)\right). \end{aligned}$$

In the same vein, we choose Vasicek instantaneous variance to match the true one at each data point, $\sigma_0^2 = b_0 + b_1 x(t) + b_2 x^{b_3}(t)$.

Our method of choosing auxiliary parameter values is based on taking into account information on data points $x(t)$, and improves the quality of the approximation of the transition density $p(x(t+1), t+1|x(t), t)$. Alternatively, we may choose the parameters $(\alpha_0, \beta_0, \sigma_0)$ by comparing first two conditional moments in two models. However, the model (5.2) does not have explicit conditional moments. In this case, we may rely on simulations based on Milstein's approximations for $x(t)$, as previously explained; in this case, conditional mean and variance are

$$\begin{aligned} \mathbb{E}[x(u)|x(t) = x] &\approx x + (u-t)\mu(x,t), \\ Var[x(u)|x(t) = x] &\approx (u-t)\sigma^2(x,t) + \frac{(u-t)^2}{2}\sigma^2(x,t)\left[\sigma'(x,t)\right]^2, \end{aligned}$$

which yields that $\sigma_0^2 = \left(b_0 + b_1 x(t) + b_2 x^{b_3}(t)\right) + (u-t)\left(b_1 + b_2 b_3 x^{b_3-1}(t)\right)^2/8$.

Note that, in this model, the transition density is unknown. We evaluate the performance of the closed-form approximation by comparing it to the Monte Carlo density estimator of Kristensen and Shin (2012). Since we are going to use this estimator as a benchmark for other models, it is useful to describe this estimator for the general Markov jump-diffusion in Eq. (2.1). To compute $p(y, u|x, t)$ from the model in Eq. (2.1), first simulate $N \geq 1$ trajectories using the Euler scheme of degree $M \geq 1$ $(m = 1, \ldots, M, i = 1, \ldots, N)$,

$$X_{1,m+1}^i = X_{1,m}^i + \mu\left(X_{1,m}^i, t + \frac{m\Delta}{M}\right)\frac{\Delta}{M} + \sigma\left(X_{1,m}^i, t + \frac{m\Delta}{M}\right)\varepsilon_{1,m}^i + J_m^i U_m^i,$$

with the recursion initialized at $X_{1,1}^i = x$ for all $i = 1, \ldots, N$. Here, $\Delta = u - t$; J_m^i are a sequence of random variable with distribution given by $J_{m+1}^i \sim \nu\left(X_{1,m}^i, t + \frac{m\Delta}{M}\right)$; U_m^i is a sequence of binomial random variable, with $p\left(U_m^i = 1\right) = \lambda\left(X_{1,m}^i, t + \frac{m\Delta}{M}\right)\frac{\Delta}{M}$; $\varepsilon_{1,m}^i$ is i.i.d. $N\left(0, I_d\frac{\Delta}{M}\right)$. We then have that $X_{1,M+1}^i \sim p(\cdot, u|x, t)$, for $i = 1, \ldots, N$, and $p(y, u|x, t)$ can be approximated by

$$\hat{p}(y, u|x, t) = \frac{1}{N}\sum_{i=1}^{N} K_h\left(X_{1,M+1}^i - y\right),$$

where $K_h(x) = K(x/h)/h^d$, $K : \mathbb{R}^d \mapsto \mathbb{R}$ is a kernel, and $h > 0$ a bandwidth.

Figure 2 plots the transition density of the model in Aït-Sahalia (1996) using the method in Kristensen and Shin (2012) and our approximation methods, respectively. Following their asymptotic property, we assume that NPSL with $N = 10^6$ in Figure 2 is the true density. The approximation methods used up to 1st and 3rd order expansion, which yields an approximated transition density quite close to the one from NPSL with large N.

In order to evaluate the finite sample performance of the approximate MLE, we perform the following Monte Carlo exercise. We generate 500 artificial sample paths of length $T = 150$ (monthly frequency) and $T = 12$ (daily frequency) from a data generating process of Aït-Sahalia (1996).

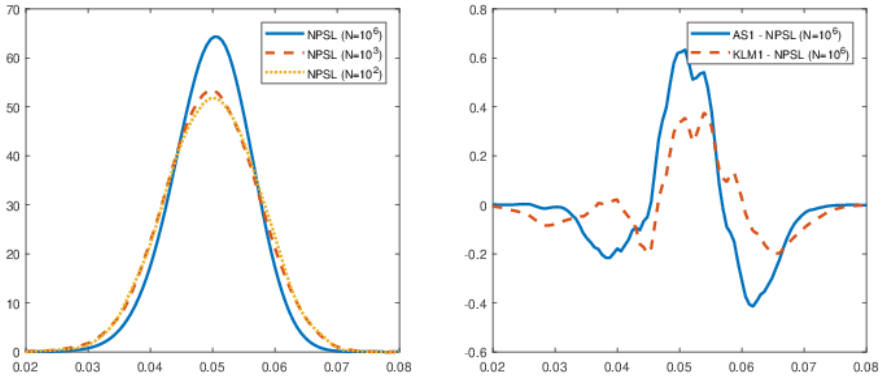

Figure 2: Approximation of transition density of AS96 model, $\Delta = 1$.

Table 3: Performance of MLE's for AS96 model computed using approximated likelihoods and NPSL. All values have been multiplied by 10^3 ($\Delta = 1/12$) and 10 ($\Delta = 1/252$).

		a_0	a_1	a_2	a_3	b_0	b_1	b_2	b_3
$\Delta = 1/12$, $T = 15$									
KLM2	Bias	−0.0399	−3.69	−78.09	0.0141	0.0004	0.0015	−0.0340	0.87
	S.D.	0.5293	12.20	179.15	0.0366	0.0063	0.0966	0.2661	7.48
	RMSE	0.5308	12.74	195.43	0.0392	0.0063	0.0966	0.2683	7.53
NPSL	Bias	−0.0174	−4.17	−64.80	0.0103	−0.0002	0.1082	−0.0325	5.65
($N = 10^2$)	S.D.	0.2524	27.95	226.98	0.0587	0.0035	0.6867	1.3557	10.28
	RMSE	0.2530	28.25	236.05	0.0596	0.0035	0.6951	1.3561	11.74
$\Delta = 1/252$, $T = 12$									
KLM1	Bias	0.1896	−0.5864	−15.98	0.0016	−0.0001	0.0001	0.0099	−0.1732
	S.D.	0.2791	5.3487	36.17	0.0114	0.0007	0.0119	0.1003	3.0918
	RMSE	0.3375	5.3807	39.54	0.0115	0.0007	0.0119	0.1007	3.0966
AS1	Bias	0.1843	−0.6975	−14.73	0.0019	−0.0001	0.0000	0.0131	−0.1047
	S.D.	0.2256	4.9134	34.19	0.0092	0.0007	0.0120	0.1121	3.3519
	RMSE	0.2913	4.9626	37.23	0.0094	0.0007	0.0120	0.1128	3.3535

True parameter values are $a = (a_0, a_1, a_2, a_3) = 10^{-3} \times (-5.652, 96.48, -534.9, 0.1041)$ and $b = (b_0, b_1, b_2, b_3) = 10^{-3} \times (0.1099, -2.007, 13.29, 2051)$. It is assumed to be the estimates in last column of Table 4 of Aït-Sahalia (1996), which corresponds to the estimates from the real sample over the period 1983-1995. Data generation is based on Milstein scheme with 10 intermediate steps. We then estimate the parameters from each sample path using approximate MLE with up to 1st and 2nd order terms of expansion. The first part of Table 3 shows that the diffusion parameters are estimated with far better accuracy compared to the drift parameters than NPSMLE. We also observe that the parameters a_2 and a_3 that are related to the higher order terms are estimated with significantly better accuracy compared to a_0 and a_1 that capture the linearity. The second part shows that our method with 1st order expansion is comparable to the method Aït-Sahalia (2002) with same order.

5.3 Affine Multifactor Term Structure Model

In a multifactor affine yield model, the instantaneous riskless rate, r_t, is a deterministic function of an $N \times 1$ vector of state variables, X_t,

$$r_t = r\left(X_t; \theta\right).$$

Under an equivalent martingale measure Q, the state vector follows the dynamics: Aït-Sahalia and Kimmel (2010) consider the following affine multi-factor model,

$$dX\left(t\right) = \left[A + BX\left(t\right)\right] dt + \Sigma S^{1/2}\left(X\left(t\right)\right) dW\left(t\right)$$

where $A \in \mathbb{R}^d$, $B \in \mathbb{R}^{d\times d}$, $\Sigma \in \mathbb{R}^{d\times d}$ and $S\left(x\left(t\right)\right) \in \mathbb{R}^{d\times d}$ satisfies $S_{ij}\left(x\left(t\right)\right) = 0$ for $i \neq j$ while $S_{ii}\left(X\left(t\right)\right) = \alpha_i + \beta_i^T X\left(t\right)$ for some $\alpha \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^d$. Dai and Singleton (2000) characterize the family of admissible affine term structure models and describe a canonical representation for each family of affine yield models. Among three families of two-factor affine models, we consider the following two models:[2]

$$\begin{aligned}
A_1\left(2\right) : d\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} &= \left[\begin{pmatrix} a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix}\right] dt + \begin{pmatrix} \sqrt{X_1\left(t\right)} & 0 \\ 0 & \sqrt{1+\sigma_{21}X_1\left(t\right)} \end{pmatrix} d\begin{pmatrix} W_1\left(t\right) \\ W_2\left(t\right) \end{pmatrix}, \\
A_2\left(2\right) : d\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} &= \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix}\right] dt + \begin{pmatrix} \sqrt{X_1\left(t\right)} & 0 \\ 0 & \sqrt{X_2\left(t\right)} \end{pmatrix} d\begin{pmatrix} W_1\left(t\right) \\ W_2\left(t\right) \end{pmatrix}.
\end{aligned}$$

For each family of affine models, the Σ matrix is equal to the identity matrix. There is no loss of generality because we can construct a new set of state variables $Y\left(t\right) = \Sigma^{-1}X\left(t\right)$. Under the constraints (i) $b_{21} = \sigma_{21} = 0$ and (ii) $b_{12} = b_{21} = 0$, the joint transition densities of $A_1\left(2\right)$ and $A_2\left(2\right)$ are the product of the marginal transition densities, respectively. For these multivariate models, we consider the following bivariate Vasicek models as auxiliary models:

$$\begin{aligned}
A_{10}\left(2\right) : d\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} &= \left[\begin{pmatrix} a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix}\right] dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} d\begin{pmatrix} W_1\left(t\right) \\ W_2\left(t\right) \end{pmatrix}, \\
A_{20}\left(2\right) : d\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} &= \left[\begin{pmatrix} a_{10} \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix}\right] dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} d\begin{pmatrix} W_1\left(t\right) \\ W_2\left(t\right) \end{pmatrix}.
\end{aligned}$$

We compare the approximated transition density to the true transition density with sets of parameters $(a_1, b_{11}, b_{21}, b_{22}, \beta_{21}) = (0.945, -0.5, 0, -0.04, 0)$ and $(a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}) = (1, 3, -0.1, 0, 0, -0.75)$ for $A_1\left(2\right)$ and $A_2\left(2\right)$, respectively. As we mentioned before, it is possible to derive the true transition densities explicitly in those settings. In figure 3-4, we compare our method with the Hermite-polynomial-based density expansion in Aït-Sahalia and Kimmel (2010). We find that

[2] In the remaining family, the dynamics of the state variables are given by

$$A_0\left(2\right) : \quad d\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} X_1\left(t\right) \\ X_2\left(t\right) \end{pmatrix} dt + d\begin{pmatrix} W_1\left(t\right) \\ W_2\left(t\right) \end{pmatrix}.$$

The transition function for this type of diffusion is known in closed-form, and is bivariate Gaussian.

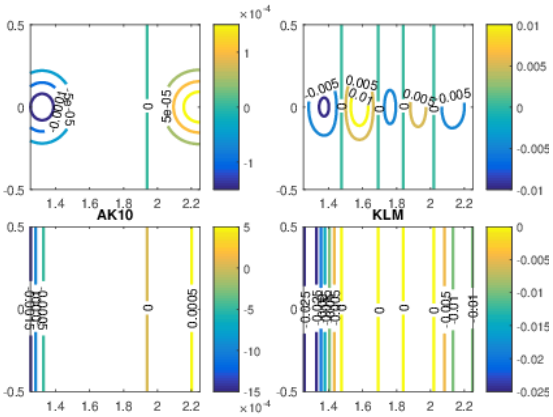

Figure 3: Approximation errors of likelihood (top) and log-likelihood (bottom) for $A_1(2)$ model, $\Delta = 1/52$

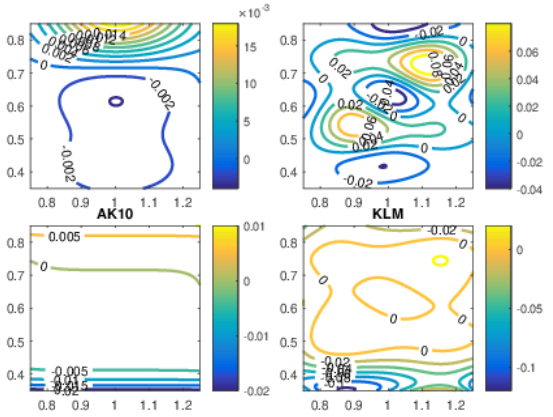

Figure 4: Approximation errors of likelihood (top) and log-likelihood (bottom) for $A_2(2)$ model, $\Delta = 1/52$

$A_{10}(2)$ and $A_{20}(2)$ have less satisfactory approximation accuracy than the one in Aït-Sahalia and Kimmel (2010), but different auxiliary models may be considered for better approximations. For example, under the constraints (i) $b_{21} = \sigma_{21} = 0$ (for $A_1(2)$) and (ii) $b_{12} = b_{21} = 0$ (for $A_2(2)$),

$$A'_{10}(2): d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \left[\begin{pmatrix} a_1 \\ a_{20} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \right] dt + \begin{pmatrix} \sqrt{X_1(t)} & 0 \\ 0 & \sigma_2 \end{pmatrix} d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix},$$

$$A'_{20}(2): d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \left[\begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \right] dt + \begin{pmatrix} \sqrt{X_1(t)} & 0 \\ 0 & \sqrt{X_2(t)} \end{pmatrix} d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}$$

are same as the original models by choosing $a_{20} = a_2$, $\sigma_2 = 1$, and $a_{10} = a_1$. Their transition densities are the products of the two marginal transition densities and they perfectly approximate $A_1(2)$ and $A_2(2)$.

5.4 Extended Heston Model

We consider the following bivariate diffusion model:

$$d\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} rX_{1t} \\ a\left(b - X_{2t}\right) \end{pmatrix} dt + \begin{pmatrix} \sqrt{\left(1-\rho^2\right) X_{2t}} X_{1t} & \rho\sqrt{X_{2t}} X_{1t} \\ 0 & \sigma\sqrt{X_{2t}} \end{pmatrix} dW_t.$$

Choi (2013) generalized the stochastic volatility model in Heston (1993) by setting X_{2t} to follow the ECIR model in Maghsoodi (1996). In terms of return of X_{1t}, $x_{1t} = \ln\left(X_{1t}\right)$ and X_{2t}, the resulting models is the time-inhomogeneous stochastic volatility model:

$$d\begin{pmatrix} x_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} r - \frac{1}{2}X_{2t} \\ a\left(b_t - X_{2t}\right) \end{pmatrix} dt + \begin{pmatrix} \sqrt{\left(1-\rho^2\right) X_{2t}} & \rho\sqrt{X_{2t}} \\ 0 & \sigma_1 e^{\sigma_2 t}\sqrt{X_{2t}} \end{pmatrix} dW_t, \tag{5.3}$$

where $b_t = \left(\sigma_1^2 d/4a\right)\exp\left(2\sigma_2 t\right)$ and d is a positive integer. We note that this model is not reducible unless $\sigma_1 = 0$ and X_{2t} is constant.

We consider two types of bivariate Vasicek models as auxiliary models,

$$d\begin{pmatrix} x_{1t} \\ X_{2t} \end{pmatrix} = \left[\begin{pmatrix} r \\ a\beta \end{pmatrix} - \begin{pmatrix} 0 & \frac{1}{2} \\ 0 & a \end{pmatrix}\begin{pmatrix} x_{1t} \\ X_{2t} \end{pmatrix}\right] dt + \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} dW\left(t\right); \tag{5.4}$$

$$d\begin{pmatrix} x_{1t} \\ X_{2t} \end{pmatrix} = \left[\begin{pmatrix} \gamma \\ a\beta \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix}\begin{pmatrix} x_{1t} \\ X_{2t} \end{pmatrix}\right] dt + \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} dW\left(t\right), \tag{5.5}$$

and choose the auxiliary parameters $\beta, \gamma, c_{11}, c_{12}, c_{22}$ by setting both the drift and diffusion functions of the true and auxiliary models to be the same

$$\beta = b_t; \quad \gamma = r - \frac{1}{2}X_{2t};$$
$$c_{11} = \sqrt{\left(1-\rho^2\right) X_{2t}}; \quad c_{12} = \rho\sqrt{X_{2t}}; \quad c_{22} = \sigma_1 e^{\sigma_2 t}\sqrt{X_{2t}}.$$

Then $D_0\left(y_1, y_2, u | x_1, x_2, t\right) = 0$.

We conduct Monte Carlo simulations to evaluate the performance of the approximate MLE. Since the true transition density of (5.3) is unknown, we generate 1000 different samples of size 500 for the weekly and daily data and size 5000 daily data by using the Euler approximation method. The sampling interval is set to $\Delta = 1/30$, where $\Delta = 1/52$ for the weekly data and $\Delta = 1/252$ for the daily observation. In addition, we set the parameter values $a = 3$, $\sigma_1 = 0.5$, $\sigma_2 = 0.001$, $\rho = -0.8$, $r = 0.03$, and $d = 9$, which are same as that in Choi (2015). The same starting values $x_{01} = \ln\left(100\right)$ and $X_{02} = 0.05625$ were used for each sample. Following the same strategies of simulating data as Aït-Sahalia and Kimmel (2007), after disposing of first 500 pairs of simulated data, every 30th pair of simulated data was put aside for estimation so that we can have a data set of frequency Δ.

The results of the simulation are presented in Table 4. Because the true transition density of (5.3) is unavailable, we cannot compute $\hat{\theta}^{(True)} - \hat{\theta}_{KLM}$ and $\hat{\theta}^{(True)} - \hat{\theta}_C$. Therefore, we show the average and standard deviation of $\hat{\theta}_{KLM} - \theta_0$ and $\hat{\theta}_C - \theta_0$. The approximated transition density with

Table 4: Monte Carlo simulation for the EHeston model: fminsearch with no option

	$\hat{\theta}_{C1}-\theta_0$		$\hat{\theta}_{C2}-\theta_0$		$\hat{\theta}_{KLM1}-\theta_0$		$\hat{\theta}_{KLM2}-\theta_0$	
θ_0	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Δ_w, $n=500$								
$a=3.00$	-0.012	0.29	0.031	0.32	0.0339	0.2877	0.0249	0.2870
$\sigma_1=0.50$	-0.0022	0.023	0.00089	0.025	0.0021	0.0234	0.0015	0.0234
$\sigma_2=0.001$	-0.00010	0.0076	-0.00028	0.0078	-0.0002	0.0069	-0.0002	0.0069
$\rho=-0.80$	0.0034	0.014	-0.00021	0.014	-0.0012	0.0132	-0.0014	0.0132
Δ_d, $n=500$								
$a=3.00$	0.082	0.57	0.075	0.56	0.0818	0.5625	0.0801	0.5622
$\sigma_1=0.50$	0.0019	0.023	0.0020	0.022	0.0004	0.0244	0.0002	0.0244
$\sigma_2=0.001$	-0.0020	0.039	-0.0016	0.038	0.0013	0.0389	0.0013	0.0389
$\rho=-0.80$	-0.00076	0.013	-0.0010	0.014	-0.0010	0.0135	-0.0011	0.0135
Δ_d, $n=5000$								
$a=3.00$	0.019	0.17	0.021	0.17	0.0186	0.1733	0.0175	0.1737
$\sigma_1=0.50$	0.000049	0.0074	0.00022	0.0073	0.000204	0.0076	0.000091	0.0076
$\sigma_2=0.001$	-0.000050	0.0013	-0.000057	0.0013	-0.000008	0.0013	-0.000007	0.0013
$\rho=-0.80$	0.00056	0.0044	0.00036	0.0044	0.000032	0.0044	-0.000038	0.0044

$\hat{\theta}_{Cn}$: nth order approximation in Choi (2015);
$\hat{\theta}_{KLM1}$, $\hat{\theta}_{KLM2}$: 1st order approximations with two different auxiliary models (5.4) and (5.5).

Table 5: Specification of jump diffusion models

Model	Specification	$J(t)$
ABNJ	$dX(t)=\mu dt+\sigma dW(t)+J(t)\,dN(t)$	$N\left(\mu_J,\sigma_J^2\right)$
VNJ	$dX(t)=\beta(\alpha-X(t))\,dt+\sigma dW(t)+J(t)\,dN(t)$	$N\left(\mu_J,\sigma_J^2\right)$
VDEJ	$dX(t)=-\beta X(t)\,dt+\sigma dW(t)+J(t)\,dN(t)$	Laplace (μ_J,σ_J)
CIREJ	$dX(t)=\beta(\alpha-X(t))\,dt+\sigma\sqrt{X(t)}dW(t)+J(t)\,dN(t)$	Exp (γ)

the second auxiliary model is lighter, so the estimation time is shorter. However, the estimation results with two different auxiliary models are comparable to that in Choi (2015).

5.5 Univariate Jump Diffusion Models

We consider jump diffusion models in Yu (2007) and Li and Chen (2016)[3]: (i) Arithmetic Brownian motions with jumps having normal distribution (ABNJ); (ii) Vasicek model with jump having normal distribution (VNJ); (iii) Vasicek model with jumps having double exponential distribution (VDEJ); (iv) CIR model with jumps having exponential distribution (CIREJ). For all jump diffusion models, $N(t)$ is a standard Poisson process with constant intensity, λ. Li and Chen (2016) consider diffusion models with jumps having normal distribution or exponential distribution, but we allow more general jump component in exponential family, e.g., the VDEJ model (Yu, 2007). Table 5 specifies jump diffusion models that we approximate, their jump distributions.

As shown before, our approximation for jump diffusions have an additional integration element. For the n-th order of expansion, we need to evaluate the following integral

$$\int_{-\infty}^{\infty} D_{n-1}(y,u|x+c,t)\,v(c,t)\,dc.$$

[3]Chenxu Li and Dachuan Chen provided their Matlab and Mathematica codes for jump diffusion models.

However, in the case of exponential family of $v(c,t)$, The above integral can be approximated by using the Gauss-Hermite or Gauss-Laguerre quadrature.[4] For example, when $J(t)$ is i.i.d. with double exponential distribution with mean zero and standard deviation σ_J, it follows from a change of variables that

$$\begin{aligned}
&\int_{-\infty}^{\infty} D_{n-1}(y,u|x+c,t)\frac{1}{2\sigma_J}e^{-\frac{|c|}{\sigma_J}}dc \\
&= \int_{0}^{\infty} \left[D_{n-1}(y,u|x+c,t)+D_{n-1}(y,u|x-c,t)\right]\frac{1}{2\sigma_J}e^{-\frac{c}{\sigma_J}}dc \\
&= \frac{1}{2}\int_{0}^{\infty} \left[D_{n-1}(y,u|x+\sigma_J z,t)+D_{n-1}(y,u|x-\sigma_J z,t)\right]e^{-z}dz.
\end{aligned}$$

Then, given the nodes and weights, z_i^{GL} and w_i^{GL}, for the Gauss-Laguerre quadrature, the approximation is in the following way:

$$\begin{aligned}
&\int_{-\infty}^{\infty} D_{n-1}(y,u|x+c,t)\frac{1}{2\sigma_J}e^{-\frac{|c|}{\sigma_J}}dc \\
&\simeq \frac{1}{2}\sum_{i=0}^{n_{GL}-1} w_i^{GL}\left[D_{n-1}\left(y,u|x+\sigma_J z_i^{GL},t\right)+D_{n-1}\left(y,u|x-\sigma_J z_i^{GL},t\right)\right].
\end{aligned}$$

We use this approximation method when we cannot obtain an exact expression of the integral with standard packages such as Mathematica, or it may not be evaluated through equally standard packages such as Matlab. An alternative way to approximate an integral is to use a Monte Carlo method, which relies on drawing random numbers. We find that Gaussian quadrature is more accurate and easier to implement than a Monte Carlo method with the extremely low computational costs.

For comparison, we truncate the infinite series transition density (ABNJ) or evaluate Fourier transform inversion of the characteristic function of $X(t)$ (other models) as benchmarks. The expressions for benchmarks are derived in the second part of Table 6. For the AMNJ model, we use the 100th sum of the above infinite series. For other models, given a characteristic function, $CF_{u-t}(v,x)=\mathbb{E}\left[\exp\left(ivX(u)\right)|X(t)=x\right]$, the transition density function is

$$p(y,u|x,t)=\frac{1}{2\pi}\int e^{-ivy}CF_{u-t}(v,x)\,dv.$$

It is efficiently approximated via numerical inversion of Fourier transform. Following Li and Chen (2016), we use an algorithm proposed in Abate and Whitt (1992).

Our approximation method begins by choosing auxiliary models corresponding to the original models. Here, we consider the original models without jump. Then, $\lambda_0(x,t)=v_0(c,t)=0$ and

[4] These quadrature procedures have been implemented in Matlab by Geert Van Damme, and available through his entries at MatlabCentral.

Table 6: Benchmarks for testing accuracy

Model	Transition density for ABNJ and $\mathbb{E}\left[\exp\left(ivX\left(u\right)\right)|X\left(t\right)=x\right]$ for other models
ABNJ	$\sum_{i=0}^{\infty}\frac{(\lambda(u-t))^{i}e^{-\lambda(u-t)}}{i!\sqrt{2\pi\left(\sigma^{2}(u-t)+i\sigma_{J}^{2}\right)}}\exp\left(-\frac{(y-x-\mu(u-t)-i\mu_{J})^{2}}{2\left(\sigma^{2}(u-t)+i\sigma_{J}^{2}\right)}\right)$
VNJ	$\exp\left(ixve^{-\beta(u-t)}+iv\alpha\left(1-e^{-\beta(u-t)}\right)+\frac{\left(e^{-2\beta(u-t)}-1\right)v^{2}\sigma^{2}}{4\beta}-\lambda\left(u-t\right)\right)$
	$\times\exp\left(\frac{\lambda}{2\beta}\int_{t}^{u}\exp\left(iv\mu_{J}e^{-\beta w}-\frac{1}{2}v^{2}\sigma_{J}^{2}e^{-2\beta w}\right)dw\right)$
VDEJ	$\exp\left(ivxe^{-\beta(u-t)}+\left(e^{-2\beta(u-t)}-1\right)v^{2}\sigma^{2}/\left(4\beta\right)\right)\left(\frac{1+e^{-2\kappa(u-t)}v^{2}\sigma_{J}^{2}}{1+v^{2}\sigma_{J}^{2}}\right)^{\lambda/2\beta}$
CIREJ	$\exp\left(\frac{ixv\beta e^{-\beta(u-t)}}{\beta-iv\sigma^{2}\left(1-e^{-\beta(u-t)}\right)/2}-\frac{2\beta\alpha}{\sigma^{2}}\log\left(\frac{\beta-iv\sigma^{2}\left(1-e^{-\beta(u-t)}\right)/2}{\beta}\right)\right)$
	$\times\exp\left(\frac{2\lambda}{2\beta-\sigma^{2}\gamma}\log\left(\frac{\beta-iv\beta e^{-\beta(u-t)}/\gamma-iv\sigma^{2}\left(1-e^{-\beta(u-t)}\right)/2}{\beta-iv\beta/\gamma}\right)\right)$

$\Delta\mu_{i}\left(x,t\right)=\Delta\sigma_{ij}^{2}\left(x,t\right)=0$ for all i,j, which lead that, in Eq. (2.8),

$$\begin{aligned}D_{0}\left(y,u|x,t\right)&=\lambda\left(x,t\right)\int\left[p_{0}\left(y,u|x+c,t\right)-p_{0}\left(y,u|x,t\right)\right]v\left(c,t\right)dc\\&=\lambda\left(x,t\right)\mathbb{E}_{c}\left[p_{0}\left(y,u|x+c,t\right)-p_{0}\left(y,u|x,t\right)\right].\end{aligned}$$

For the CIREJ model, the Vasicek model without jump is also considered as an auxiliary model.

Table 7 shows the specification of auxiliary models for approximation of the original models. We can obtain $D_{0}\left(y,u|x,t\right)$ based on these auxiliary models and jump distributions. The expressions for D_{0} are derived in the second part of Table 7 when it is available explicitly. As mentioned at the beginning of this section, in the case of exponential family of $v\left(c,t\right)$, Gauss-Hermite and Gauss-Laguerre quadratures make us approximate the above integral. For example, for the VNJ model, D_{0} and D_{1} can be derived explicitly, but we employ the Gauss-Hermite quadrature method to approximate D_{n} for $n\geq 2$.

In the numerical analysis, we employ the sets of parameters in Table 8, which are based on the existing literature. Figure 5-6 shows the approximation errors of likelihood and log-likelihood for the ABNJ model. We compute the transition density over a range of y values while fixing $x=\mu=0.2$ with weekly and daily monitoring frequencies ($\Delta:=u-t=1/52,1/252$). As an alternative approximation method for a benchmark, we also compute the transition density approximated in Li and Chen (2016). Our approximation method has a similar accuracy with Li and Chen (2016). We find that the accuracy of the approximation increases rapidly with higher order expansion and frequency.

When we use quadrature techniques, there are two types of approximation errors: our original error itself and the error resulting from the quadrature methods. We find that approximation error decreases with our approximation order or the number of nodes in the quadrature methods. Moreover, we find that, provided the original error is already tiny, our approximation error mostly results from the quadrature error. In practice, we would recommend useing larger nodes for the

Table 7: Auxiliary models for approximation of jump diffusion models

Model	Auxiliary models	Quadrature for D_n
ABNJ	$dX(t) = \mu dt + \sigma dW(t)$	No
VNJ	$dX(t) = \beta(\alpha - X(t))dt + \sigma dW(t)$	$n \geq 2$
VDEJ	$dX(t) = -\beta X(t)dt + \sigma dW(t)$	$n \geq 1$
CIREJ	(i) $dX(t) = \beta(\alpha - X(t))dt + \sigma\sqrt{X(t)}dW(t)$	$n \geq 0$
	(ii) $dX(t) = \beta(\alpha - X(t))dt + \sigma_0 dW(t)$	$n \geq 1$
	D_0/λ ($D_0(y,u\|x,t) = \lambda\mathbb{E}_c[p_0(y,u\|x+c,t) - p_0(y,u\|x,t)]$, λ: jump intensity)	
ABNJ	$\frac{\exp\left(-\frac{(y-x-\mu(u-t)-\mu_J)^2}{\sigma^2(u-t)+\sigma_J^2}\right)}{\sqrt{2\pi\left(\sigma^2(u-t)+\sigma_J^2\right)}} - \frac{\exp\left(-\frac{(y-x-\mu(u-t))^2}{\sigma^2(u-t)}\right)}{\sqrt{2\pi\sigma^2(u-t)}}$	
VNJ	$\frac{\exp\left(-\frac{\left(y-\alpha-(x+\mu_J-\alpha)e^{-\beta(u-t)}\right)^2}{\gamma^2+2e^{-2\beta(u-t)}\sigma_J^2}\right)}{\sqrt{\pi\left(\gamma^2+2e^{-2\beta(u-t)}\sigma_J^2\right)}} - \frac{\exp\left(-\frac{\left(y-\alpha-(x-\alpha)e^{-\beta(u-t)}\right)^2}{\gamma^2}\right)}{\sqrt{\pi\gamma^2}}$	
VDEJ	$\frac{e^{\frac{2ye^{\beta(u-t)}}{\sigma_J}}\left(1+\mathrm{erf}\left(\frac{\left(2\beta\sigma_J\left(x-ye^{\beta(u-t)}\right)-\sigma^2\left(e^{2\beta(u-t)}-1\right)\right)}{2\sigma_J\sqrt{\beta\sigma^2\left(e^{2\beta(u-t)}-1\right)}}\right)\right)}{4\sigma_J e^{\beta(u-t)}\exp\left(\frac{4\beta\sigma_J\left(ye^{\beta(u-t)}+x\right)-\sigma^2\left(e^{2\beta(u-t)}-1\right)}{4\beta\sigma_J^2}\right)}$	
	$+\frac{e^{\frac{2x}{\sigma_J}}\left(1-\mathrm{erf}\left(\frac{\left(2\beta\sigma_J\left(x-ye^{\beta(u-t)}\right)+\sigma^2\left(e^{2\beta(u-t)}-1\right)\right)}{2\sigma_J\sqrt{\beta\sigma^2\left(e^{2\beta(u-t)}-1\right)}}\right)\right)}{4\sigma_J e^{\beta(u-t)}\exp\left(\frac{4\beta\sigma_J\left(ye^{\beta(u-t)}+x\right)-\sigma^2\left(e^{2\beta(u-t)}-1\right)}{4\beta\sigma_J^2}\right)} - \frac{\exp\left(-\frac{\beta\left(x-ye^{\beta(u-t)}\right)^2}{\sigma^2\left(e^{2\beta(u-t)}-1\right)}\right)}{\sqrt{\pi}\sqrt{\frac{\sigma^2\left(1-e^{-2\beta(u-t)}\right)}{\beta}}}$	
CIREJ	(i) Gauss-Laguerre quadrature	
	(ii) $\frac{\gamma\mathrm{erf}\left(\sqrt{\frac{\beta}{\sigma^2x\left(e^{2\beta(u-t)}-1\right)}}\left(\alpha\left(1-e^{\beta(u-t)}\right)+\frac{\gamma\sigma^2x\left(1-e^{2\beta(u-t)}\right)}{2\beta}-\left(x-ye^{\beta(u-t)}\right)\right)\right)+\gamma}{2e^{\beta(t-u)}\exp\left(\alpha\gamma\left(1-e^{\beta(u-t)}\right)+\frac{\gamma^2\sigma^2x\left(1-e^{2\beta(u-t)}\right)}{4\beta}-\gamma\left(x-ye^{\beta(u-t)}\right)\right)}$	
	$-\frac{\sqrt{\frac{\beta}{\sigma^2x\left(e^{2\beta(u-t)}-1\right)}}\exp\left(-\frac{\beta\left((\alpha-x)e^{\beta(t-u)}-\alpha+y\right)^2}{\sigma^2x\left(1-e^{2\beta(t-u)}\right)}\right)}{\sqrt{\pi}e^{\beta(t-u)}}$	

$\mathrm{erf}(x) = 2\left(\int_0^x \exp\left(-s^2\right)ds\right)/\sqrt{\pi}$: Error function evaluated for x

Table 8: Employed set of parmeters

Model	Set of parameters	Reference
ABNJ	$(\mu,\sigma,\lambda,\mu_J,\sigma_J) = (0.2, 0.3, 0.33, 0, 0.2)$	Yu (2007)
VNJ	$(\alpha,\beta,\sigma,\lambda,\mu_J,\sigma_J) = (0, 0.5, 0.2, 0.33, 0, 0.28)$	Yu (2007)
VDEJ	$(\beta,\sigma,\lambda,\mu_J,\sigma_J) = (0.5, 0.2, 0.33, 0, 0.2)$	Yu (2007)
CIREJ	$(\alpha,\beta,\sigma,\lambda,\gamma) = (0.02, 0.6, 0.141, 0.2, 10)$	Duffie and Gârleanu (2001)

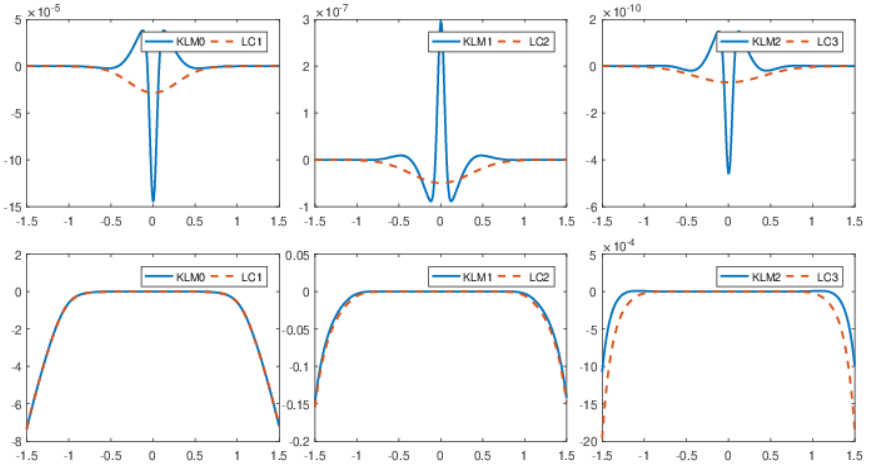

Figure 5: Approximation errors of likelihood (top) and log-likelihood (bottom) for ABNJ: $\Delta = 1/52$

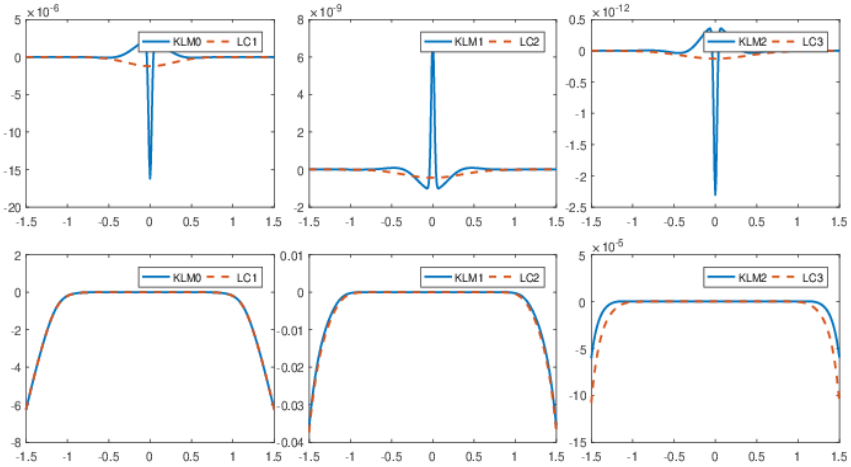

Figure 6: Approximation errors of likelihood (top) and log-likelihood (bottom) for ABNJ: $\Delta = 1/252$

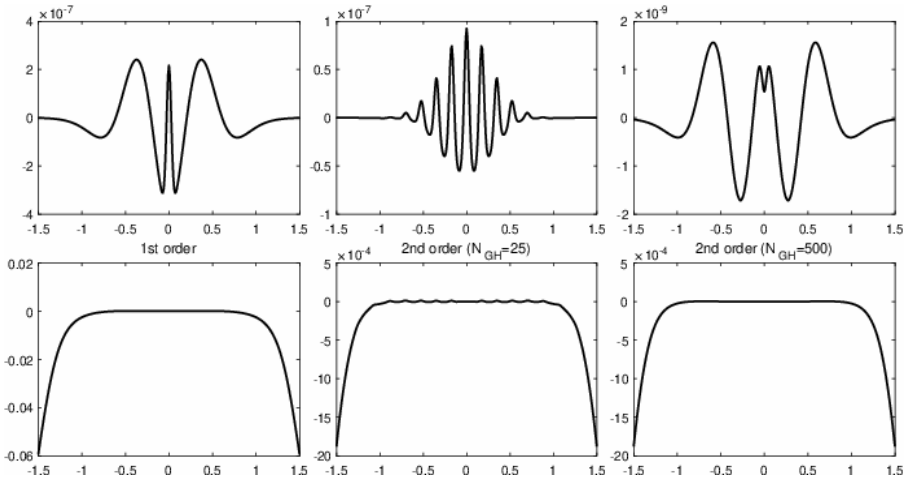

Figure 7: Approximation errors of likelihood (top) and log-likelihood (bottom) for VNJ: $\Delta = 1/52$

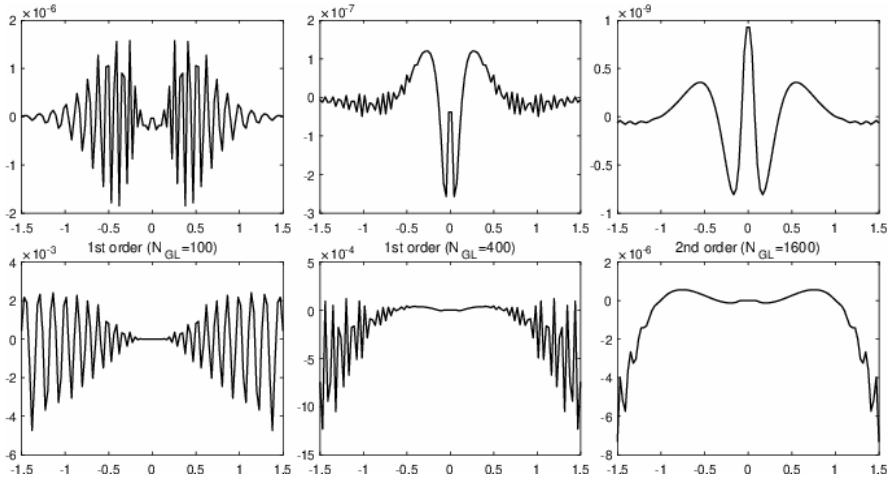

Figure 8: Approximation errors of likelihood (top) and log-likelihood (bottom) for VDEJ: $\Delta = 1/52$

approximation of the integral while using data sampled at a relatively high frequency, such as daily.

Figure 7-8 depict the approximation errors of the likelihood and log-likelihood functions for two different jump models at a weekly frequency. For the VNJ model, the approximation errors of the likelihood with large number of nodes are significantly smaller, although the errors for the log-likelihood function are quite similar. For the VDEJ model, we use the quadrature techniques from the first order approximation and a large number of nodes are required to improve approximation accuracy with higher order expansion. As a result, in the daily frequency, there is no improvement in approximation accuracy in the second order expansion compared to the first order one.

In Section 5.1, we find that the Vasicek model approximates the CIR model well. We will forgo this knowledge and approximate the transition density of the CIREJ model using the CIR model as well as the Vasicek model as auxiliary models. In figure 9, we compare our method with the

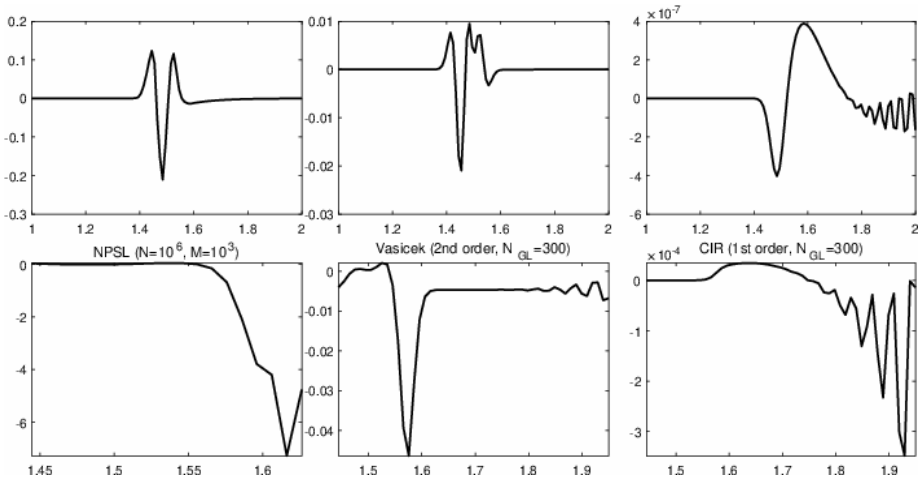

Figure 9: model with different auxiliary models (centre: Vasicek without jumps, right: CIR without jumps), $\Delta = 1/52$.

nonparametric simulation method. The choice of the auxiliary model is important. When we use our method with the Vasicek model as the auxiliary one, the approximation error mostly depend on the error from our method. We usually increase the number of nodes for the Gaussian quadrature to increase our approximation accuracy, but the accuracy of our approximation is limited when we choose the Vasicek model as an auxiliary model. However, its approximation accuracy is more satisfactory compared to that of NPSL. Our method can also save computation time maintain similar accuracy as NPSL.

5.6 Bivariate Vasicek Model with Normally Distributed Jumps

Finally, we consider the following multivariate diffusion model with jumps:

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \beta_{11} & 0 \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_1(t) \\ \alpha_2 - X_2(t) \end{pmatrix} dt + d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix} + \begin{pmatrix} J_1(t) \\ J_2(t) \end{pmatrix} dN(t), \tag{5.6}$$

where $\{(W_1(t), W_2(t))\}$ is a standard two-dimensional Brownian motion and the jump size has a bivariate normal distribution according to

$$\begin{pmatrix} J_1(t) \\ J_2(t) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{J1} \\ \mu_{J2} \end{pmatrix}, \begin{pmatrix} \sigma_{J1}^2 & 0 \\ 0 & \sigma_{J2}^2 \end{pmatrix} \right).$$

This is the bivariate Vasicek model with normally distributed jumps (BVNJ). Similar to the previous examples, the characteristic function is obtained as

$$CF_{u-t}(v_1, v_2, x_1, x_2) = \mathbb{E}\left(e^{i(v_1 X_1(t) + v_2 X_2(t))} | X(0) = (x_1, x_2) \right),$$

with $i = \sqrt{-1}$. In order to set up our benchmark, we evaluate $CF_{u-t}(v_1, v_2, x_1, x_2)$ following Li and Chen (2016). We, then, forgo this knowledge, and approximate the transition density using

the bivariate Vasicek model as an auxiliary model,

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_1(t) \\ \alpha_{02} - X_2(t) \end{pmatrix} dt + d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}, \tag{5.7}$$

where $(W_1(t), W_2(t))$ is a standard two-dimensional Brownian motion. It is equivalent to two single-dimensional diffusion models.[5] The transition density function of process (5.7) is given as

$$\begin{aligned} p_0(y,u|x,t) =& \frac{1}{\sqrt{\pi\gamma_1^2}} \exp\left[-\frac{\left(y_1 - \alpha_1 - (x_1 - \alpha_1) e^{-\beta_{11}(u-t)}\right)^2}{\gamma_1^2}\right] \\ & \times \frac{1}{\sqrt{\pi\gamma_2^2}} \exp\left[-\frac{\left(y_2 - \alpha_{02} - (x_2 - \alpha_{02}) e^{-\beta_{22}(u-t)}\right)^2}{\gamma_2^2}\right] \\ :=& p_{01}(y_1, u|x_1, t)\, p_{02}(y_2, u|x_2, t), \end{aligned}$$

where $\gamma_i^2 = \left(1 - e^{-2\beta_{ii}(u-t)}\right)/\beta_{ii}$. The adjustment term also takes the simple form:

$$\begin{aligned} & D_0(y,u|x,t) \\ =& \left(\beta_{21}(\alpha_1 - x_1) + \beta_{22}(\alpha_2 - \alpha_{02})\right) \frac{\partial p_{02}(y_2, u|x_2, t)}{\partial x_2} \\ & + \lambda\left\{\mathbb{E}_{c_1}\left[p_{01}(y_1, u|x_1 + c_1, t)\right] \mathbb{E}_{c_2}\left[p_{02}(y_2, u|x_2 + c_2, t)\right] - p_{01}(y_1, u|x_1, t)\, p_0(y_2, u|x_2, t)\right\}. \end{aligned}$$

By setting $\alpha_{02} = \alpha_2 - (\alpha_1 - x_1)\beta_{21}/\beta_{22}$, we have that

$$\begin{aligned} \frac{D_0(y,u|x,t)}{\lambda} =& \frac{\exp\left(-\frac{\beta_{22}\left(\left(e^{-\beta_{22}(u-t)}-1\right)\left(\frac{\beta_{21}(\alpha_1-x_1)}{\beta_{22}}+\alpha_2-x_2-\mu_{J2}\right)-(x_2+\mu_{J2})+y_2\right)^2}{1-\left(1-2\beta_{22}\sigma_{J2}^2\right)e^{-2\beta_{22}(u-t)}} - \frac{\beta_{11}\left(e^{-\beta_{11}(u-t)}(\alpha_1-x_1-\mu_{J1})-\alpha_1+y_1\right)^2}{1-\left(1-2\beta_{11}\sigma_{J1}^2\right)e^{-2\beta_{11}(u-t)}}\right)}{\pi\sqrt{\frac{1-e^{-2\beta_{11}(u-t)}}{\beta_{11}}}\sqrt{\frac{1-e^{-2\beta_{22}(u-t)}}{\beta_{22}}}\sqrt{1-\frac{2\sigma_{J1}^2\beta_{11}}{1-e^{2\beta_{11}(u-t)}}}\sqrt{1-\frac{2\sigma_{J2}^2\beta_{22}}{1-e^{2\beta_{22}(u-t)}}}} \\ & - \frac{\exp\left(-\frac{\beta_{22}\left(\left(e^{-\beta_{22}(u-t)}-1\right)\left(\frac{\beta_{21}(\alpha_1-x_1)}{\beta_{22}}+\alpha_2-x_2\right)-x_2+y_2\right)^2}{1-e^{-2\beta_{22}(u-t)}} - \frac{\beta_{11}\left(e^{-\beta_{11}(u-t)}(\alpha_1-x_1)-\alpha_1+y_1\right)^2}{1-e^{-2\beta_{11}(u-t)}}\right)}{\pi\sqrt{\frac{1-e^{-2\beta_{11}(u-t)}}{\beta_{11}}}\sqrt{\frac{1-e^{-2\beta_{22}(u-t)}}{\beta_{22}}}}. \end{aligned}$$

We obtain the explicit expression for $D_1(y,u|x,t)$ as well as $D_0(y,u|x,t)$ and approximate $D_2(y,u|x,t)$, $D_3(y,u|x,t)$, and other higher-order terms using Gauss-Hermite quadrature.

In the numerical experiments, we employ a set of parameters with values similar to those in Cheridito et al. (2007), i.e., $\kappa_{11} = 0.1570$, $\kappa_{21} = 0.3279$, $\kappa_{22} = 2.2883$, $\theta_1 = \theta_2 = 0$, $\lambda = 9$, $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\beta_1 = 0.3$ and $\beta_2 = 0.5$. Figure 10 illustrate the approximation errors of our density approximation for the BVNJ model. It is easy to observe that the approximation errors decrease as our approximation order or the number of nodes in quadrature methods increases. By comparing

[5] We also consider the following bivariate Vasicek model

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \beta_{11} & 0 \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 - X_1(t) \\ \alpha_2 - X_2(t) \end{pmatrix} dt + d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}.$$

Compared to the model (5.7), it is closer to the original model because it allows a nonzero β_{21} and the density expansion with this model approximate the original one well. However, we employ the Gaussian-Hermite quardrature from $D_0(y,u|x,t)$ and then it takes longer to approximate the transition density with higher order expansion. In the paper, the accuracy of our approximation with the model (5.7) is only reported.

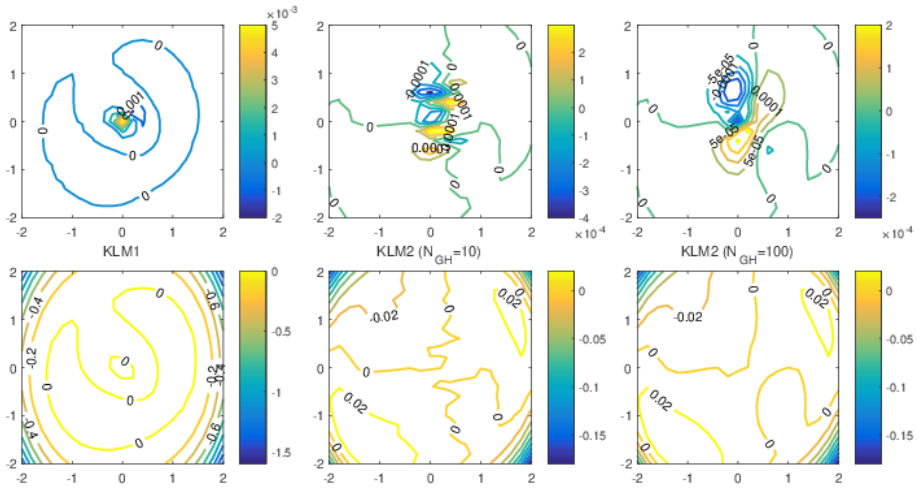

Figure 10: Approximation errors of likelihood (top) and log-likelihood (bottom) for BVNJ model, $\Delta = 1/52$.

the centre and right figures, we find that a small number of nodes in quadrature methods is sufficient to approximate the transition density with the second order approximation.

6 Conclusion

This paper develops a novel method to approximate transition densities in the context of multifactor continuous-time jump-diffusion models that are typically used in finance. These approximations rely on expanding the true transition densities around densities that are known in closed-form. The higher the number of terms in these expansions, the more accurate our approximations are. We show that our approximations may be used to implement maximum-likelihood inference as the accuracy of our density approximation increases: our approximated maximum-likelihood estimators are asymptotically equivalent to the exact ones, as the number of terms in our expansions grow. Our methods are easily applied to implement parameter estimation for many models, are numerically precise in practice, and require much less computation time than existing methods. Therefore, they may be used as adequate tools to implement practical tasks related to asset pricing, risk-management, and portfolio selection.

References

Abate, J. and W. Whitt (1992). The fourier-series method for inverting transforms of probability distributions. *Queueing Systems 10*(1), 5–87.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies 9*(2), 385–426.

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. Statist. 36*(2), 906–937.

Aït-Sahalia, Y. and R. L. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*(2), 413–452.

Aït-Sahalia, Y. and R. L. Kimmel (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics 98*(1), 113–144.

Andrews, D. W. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory 11*(3), 560–586.

Bakshi, G., N. Ju, and H. Ou-Yang (2006). Estimation of continuous-time models with an application to equity volatility dynamics. *Journal of Financial Economics 82*(1), 227–249.

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

Beskos, A., O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 333–382.

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

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

Casella, B. and G. O. Roberts (2011). Exact simulation of jump-diffusion processes with Monte Carlo applications. *Methodology and Computing in Applied Probability 13*(3), 449–473.

Cheridito, P., D. Filipović, and R. L. Kimmel (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics 83*(1), 123–170.

Choi, S. (2013). Closed-form likelihood expansions for multivariate time-inhomogeneous diffusions. *Journal of Econometrics 174*(2), 45–65.

Choi, S. (2015). Explicit form of approximate transition probability density functions of diffusion processes. *Journal of Econometrics 187*(1), 57–73.

Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *The Journal of Finance 55*(5), 1943–1978.

Duffie, D. and N. Gârleanu (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal 57*(1), 41–59.

Durham, G. B. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics 20*(3), 297–338.

Egorov, A. V., H. Li, and Y. Xu (2003). Maximum likelihood estimation of time-inhomogeneous diffusions. *Journal of Econometrics 114*(1), 107–139.

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

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics 20*(4), 547–557.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*(4), 657–681.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*(S1), S85–S118.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies 6*(2), 327–343.

Karatzas, I. and S. Shreve (1998). *Brownian Motion and Stochastic Calculus*. Springer-Verlag New York.

Kristensen, D. and A. Mele (2011). Adding and subtracting Black-Scholes: A new approach to approximating derivative prices in continuous-time models. *Journal of Financial Economics 102*(2), 390–415.

Kristensen, D. and Y. Shin (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics 167*(1), 76–94.

Li, C. and D. Chen (2016). Estimating jump-diffusions using closed-form likelihood expansions. *Journal of Econometrics 195*(1), 51–70.

Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory 4*(2), 231–247.

Maghsoodi, Y. (1996). Solution of the extended CIR term structure and bond option valuation. *Mathematical Finance 6*(1), 89–109.

Nualart, D. and W. Schoutens (2001). Backward stochastic differential equations and Feynman-Kac formula for Lévy processes, with applications in finance. *Bernoulli 7*(5), 761–776.

Pazy, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag New York.

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

Sermaidis, G., O. Papaspiliopoulos, G. O. Roberts, A. Beskos, and P. Fearnhead (2013). Markov chain monte carlo for exact inference for diffusions. *Scandinavian Journal of Statistics 40*(2), 294–321.

Xiu, D. (2010). Dissecting option prices using closed-form series expansion with an empirical study of variance risk premia. Unpublished working paper, Princeton University.

Yang, J. (2006). *Stochastic Volatility Models: Option Price Approximation, Asymptotics and Maximum Likelihood Estimation*. Ph. D. thesis, University of Illinois at Urbana-Champaign.

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

A Simulated Maximum-Likelihood

If the analytical derivation of $L^n D(y,u|x,t)$, $n \geq 0$, is prohibitively complicated, even with a symbolic software package, one can alternative use Monte Carlo methods to evaluate the right-hand side expression given in the representation given in Theorem 2.1. Instead of approximating the conditional moment on the right hand side of Eq. (2.9) through a series expansion, we here propose to simulate it. Let $x_i(t_j)$, $i = 1, ...N$ and and $j = 1, ..., M$, be N simulated trajectories evaluated at discrete time points $t_j \in [t,u]$, $i = 1, ..., M$, all initialized at $x_i(t) = x$. For example, $t_j = (t-u)\, j/M$ which generates a uniform grid on $[t,u]$. We then approximate $\int_t^u \mathbb{E}_{x,t}[D(y,u|x(s),s)]\, ds$ by $\sum_{i=1}^M \sum_{j=1}^N D(y,u|x_i(t_j),t_j)/(MN)$. This yields the following associated simulated density:

$$\hat{p}_{M,N}(y,u|x,t) = p_0(y,u|x,t) + \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} D(y,u|x_i(t_j),t_j). \tag{A.1}$$

As with the series-expansion approximation, we can define an associated (simulated) maximum likelihood estimator.

The above method assumes that we can simulate trajectories exactly. Such simulators have been developed in Beskos et al. (2009) and Casella and Roberts (2011), amongst others. For the general case, however, exact simulators are currently not available and we propose to use the Euler scheme to generate approximate simulations:

1. Compute the grid $t_j = t + (u-t) \times j/M$ for $j = 1, ..., M$.

2. For $i = 1, ..., N$: Compute

$$x_i(t_j) = x_i(t_{j-1}) + \mu(x_i(t_{j-1}), t_{j-1})\frac{1}{N} + \sigma(x_i(t_{j-1}), t_{j-1})\frac{1}{\sqrt{N}}U_i + J(x_i(t_{j-1}), t_{j-1})\, V_i$$

   where $J(x_i(t_{j-1}), t_{j-1}) \sim v(x_i(t_{j-1}), t_{j-1})$, $U_i \sim N(0, I_d)$ and $P(V_i = 1) = \lambda(x_i(t_{j-1}), t_{j-1})/N$, $j = 1, ..., M$, with initial value $x_i(t) = x$.

3. Compute $\hat{p}_{M,N}(y,u|x,t)$ given in Eq. (A.1).

The use of the Euler scheme in Step 2 imparts a discretization bias. However, this can be controlled for by choosing M large enough. In particular, as $M, N \to \infty$, $\sum_{i=1}^M \sum_{j=1}^N D(y,u|x_i(t_j),t_j)/(MN) \to^p \int_t^u \mathbb{E}_{x,t}[D(y,u|x(s),s)]\, ds$.

B Proofs

Proof of Theorem 2.1. Since p and p_0 are well-defined solutions to their partial differential equations, the difference, $\Delta p = p - p_0$, is a well-defined solution to the partial differential equation (2.7). We then need to verify that the conditions for the Feynman-Kac formula to hold are satisfied.

We use the conditions of Karatzas and Shreve (1998, Theorem 5.7.6): First, due to A.1, $\Delta p\left(y,u|x,t\right)$ belongs to $\mathcal{C}^{2,1}\left(\mathbb{R}^{d}\times\left[0,T\right]\right)$ and satisfies $\left|\Delta p\left(y,u|x,t\right)\right|\leq C\left(1+\left\|x\right\|^{q}\right)$. Second, $D\left(y,u|x,t\right)$ is continuous and satisfy the same growth condition as Δp due to A.2. All the conditions in Karatzas and Shreve (1998, Theorem 5.7.6) [see also Nualart and Schoutens, 2001] are therefore met and the result follows.

Proof of Proposition 4.1. This follows from Kristensen and Mele (2011, Proposition A.1).

Proof of Proposition 4.2. See Schaumburg (2004, Lemma 2.2).

Proof of Proposition 4.3. This follows from Kristensen and Mele (2011, Proposition A.5).

Proof of Proposition 4.4. We first apply Kristensen and Shin (2012, Proposition A.1) to obtain that

$$||\hat{\theta}_{n,N}-\hat{\theta}_{n}||=O_{P}\left(\sup_{\theta}\left|L_{n}\left(\theta\right)-L_{n,N}\left(\theta\right)\right|\right).$$

Next, we proceed as in Kristensen and Shin (2012, Proof of Theorem 1): We introduce an additional trimming function, $\tilde{\tau}_{N,i}\left(\theta,a\right)=\tau_{N,i}\left(\theta,a\right)\mathbb{I}\left\{\left\|x_{t}\right\|\leq B_{N}\right\}$, where $\mathbb{I}\left\{\cdot\right\}$ is the indicator function and $\gamma>0$, and two trimming sets,

$$\begin{aligned}A_{1,i}^{\varepsilon}(\theta,a)&=\left\{p_{N}(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)\geq\varepsilon a,\left\|x_{t}\right\|\leq B_{N}\right\},\\A_{2,i}^{\varepsilon}(\theta,a)&=\left\{p(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)\geq\varepsilon a,\left\|x_{t}\right\|\leq B_{N}\right\},\end{aligned}$$

for any $\varepsilon>0$. Defining $A_{i}^{\varepsilon}(\theta,a)=A_{1,i}^{\varepsilon}(\theta,a)\cap A_{2,i}^{\varepsilon}(\theta,a)$, it follows by the same arguments as in Andrews (1995, p. 588), $A_{2,i}^{2\varepsilon}(\theta,a)\subseteq A_{1,i}^{\varepsilon}(\theta,a)\subseteq A_{i}^{\varepsilon/2}(\theta,a)$ w.p.a.1 as $N\rightarrow\infty$ under Assumption B.1. Thus, $\mathbb{I}_{A_{2,i}^{4}(\theta,a)}\leq\mathbb{I}_{A_{1,i}^{2}(\theta,a)}\leq\tilde{\tau}_{N,i}\left(\theta,a\right)\leq\mathbb{I}_{A_{1,i}^{1/2}(\theta,a)}\leq\mathbb{I}_{A_{i}^{1/4}(\theta,a)}$. We then split up $L_{n,N}(\theta)-L_{n}(\theta)$ into three terms,

$$\begin{aligned}L_{n,N}(\theta)-L_{n}(\theta)&=\frac{1}{n}\sum_{i=1}^{n}\left[\tau_{N,i}\left(\theta,a\right)-\tilde{\tau}_{N,i}\left(\theta,a\right)\right]\log p_{N}(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\tilde{\tau}_{N,i}\left(\theta,a\right)\left[\log p_{N}(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)-\log p(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)\right]\\&\quad+\frac{1}{n}\sum_{i=1}^{T}\left[\tilde{\tau}_{N,i}\left(\theta,a\right)-1\right]\log p(x\left(t_{i}\right)|x\left(t_{i-1}\right);\theta)\\&=:S_{1}(\theta)+S_{2}(\theta)+S_{3}(\theta),\end{aligned}$$

and then wish to obtain bounds on $\sup_{\theta\in\Theta_{T}}\left|A_{i}(\theta)\right|=o_{P}(1)$, $i=1,2,3$. First,

$$\left|S_{1}(\theta)\right|\leq\frac{\left|\log a\right|}{n}\sum_{i=1}^{n}\mathbb{I}\left\{\left\|x\left(t_{i-1}\right)\right\|>B_{N}\right\}\leq\frac{\left|\log a\right|}{B_{N}^{1+\delta}}\frac{1}{n}\sum_{t=1}^{T}\left\|x\left(t_{i-1}\right)\right\|^{1+\delta}\leq\frac{\left|\log a\right|}{B_{N}^{1+\delta}}\times O_{P}(1),$$

while,

$$\begin{aligned}|S_2(\theta)| &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{A_t(1/4)}|\log p_N(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)-\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &\leq \frac{1}{a}\times\sup_{\theta\in\Theta}\sup_{y\in\mathbb{R}^d}\sup_{\|x\|\leq B_N}|p_N(y|x;\theta)-p(y|x;\theta)|.\end{aligned}$$

The final term is bounded by

$$\begin{aligned}|S_3(\theta)| &\leq \frac{1}{n}\sum_{i=1}^{n}|\tilde{\tau}_{N,i}\left(\theta,a\right)-1||\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)<4a\}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left\{\|x\left(t_{i-1}\right)\|>B_N\right\}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &= : S_{3,1}(\theta)+S_{3,2}(\theta).\end{aligned}$$

First, as $a\to 0$,

$$\begin{aligned}S_{3,1}(\theta) &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left\{p_t(y_t|x_t;\theta)<4a\right\}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)|>|\log(4a)|\}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &\leq |\log(4a)|^{-\delta}\frac{1}{n}\sum_{i=1}^{n}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)|^{1+\delta}=|\log(4a)|^{-\delta}\times O_P(1).\end{aligned}$$

where we have used Assumption A.5. Similarly, again by A.5,

$$\begin{aligned}S_{3,2}(\theta) &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\|x_t\|>B_N\}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)| \\ &\leq \left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\|x_t\|>B_N\}\right\}^{\delta/(1+\delta)}\left\{\frac{1}{n}\sum_{i=1}^{n}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)|^{1+\delta}\right\}^{1/(1+\delta)} \\ &\leq \frac{1}{B_N^{1+\delta}}\left\{\frac{1}{n}\sum_{i=1}^{n}\|x_t\|^{1+\delta}\right\}^{\delta/(1+\delta)}\left\{\frac{1}{n}\sum_{i=1}^{n}|\log p(x\left(t_i\right)|x\left(t_{i-1}\right);\theta)|^{1+\delta}\right\}^{1/(1+\delta)} \\ &= \frac{1}{B_N^{1+\delta}}\times O_P(1).\end{aligned}$$