Efficient Parameter Estimation for Multivariate Jump-Diffusions

December 4, 2015
(Preliminary Draft)

Abstract

This paper develops an unbiased Monte Carlo estimator of the transition density of a multivariate jump-diffusion process. The drift, volatility, jump intensity, and jump magnitude are allowed to be state-dependent and non-affine. It is not necessary that the volatility matrix can be diagonalized using a change of variable or change of time. Our density estimator facilitates the parametric estimation of multivariate jump-diffusion models based on discretely observed data. Under conditions that can be verified with our density estimator, the parameter estimators we propose have the same asymptotic behavior as maximum likelihood estimators as the number of data points grows, but the observation frequency of the data is kept fixed. In a numerical case study of practical relevance, our density and parameter estimators are found to be highly accurate and computationally efficient.

1 Introduction

Multivariate jump-diffusions are popular stochastic models often used in economic and financial applications. They are used to describe the time series behavior of asset prices, volatilities, and interest rates, as well as the correlation structure of the cross section of assets. They also allow for potential discontinuities in the time series of financial and economic data. Despite their popularity, parameter inference for multivariate jump-diffusions is challenging because the underlying probability distribution is often times intractable. In this paper, we derive an unbiased Monte Carlo estimator of the transition density of a general class of multivariate jump-diffusion processes over arbitrary sample frequencies. Our density estimator can be used to perform maximum likelihood inference based on discretely observed data. Under conditions that can be verified using our density estimator, the parameter estimators we propose inherit the consistency and asymptotic normality properties of maximum likelihood estimators as the number of data points grows large.[1] Thus, the results of this paper provide a methodology to carry out statistically efficient estimation of the parameters driving the dynamics of a multivariate jump-diffusion process based on discretely observed data.

We consider a general class of Markovian multivariate jump-diffusions. The drift, volatility, jump intensity, and jump magnitude are allowed to be arbitrary parametric functions of the state. The only binding assumption is that the jump-diffusion process is well-defined in the sense that it admits a strong solution as well as a transition density. By taking advantage of Bayes' rule and a well-chosen change of measure, we rewrite the transition density of a multivariate jump-diffusion in terms of a mixture of transition densities of purely diffusive processes without jumps. Our density representation is similar to the one of Giesecke & Schwenkler (2014), who characterize the transition density of a jump-diffusion as a mixture of Gaussian densities. In contrast to Giesecke & Schwenkler (2014), though, our density representation also applies to multivariate jump-diffusion processes which are not reducible in the sense of Aït-Sahalia (2008). A process is reducible if it can be transformed to a unit volatility process, and this restrictive assumption is

[1] A maximum likelihood estimator is an almost-sure maximizer of the likelihood function.

often violated by popular multivariate jump-diffusion models. Our density representation also provides a significant generalization of the well-known representations of Dacunha-Castelle & Florens-Zmirou (1986) and Rogers (1985), which apply only to univariate diffusive processes without jumps.

A key benefit of our density representation is that it can be easily estimated via Monte Carlo simulation. This is because it is given by an unconditional expectation of a path functional of the jump-diffusion process. We exploit a novel randomization technique introduced by Glynn & Rhee (2015) to construct an unbiased estimator of the transition density. Our density estimator can be understood as a randomized multilevel Monte Carlo estimator.[2] It is constructed from samples derived from Euler's discretization method with different time steps, which are mixed and weighted adequately to ensure unbiasedness of the density estimator.[3] The accuracy of the resulting transition density estimator depends only on the number of Monte Carlo replications used.

We use the density estimator to carry out parameter inference based on discretely observed data. We construct a simulated likelihood function by replacing the uncomputable true density with our density estimator. Because the latter is unbiased, standard results ensure that the estimators that maximize the simulated likelihood inherit the asymptotic properties of true maximum likelihood estimators as the number of data points grows large while keeping the observation frequency of the data fixed.[4] Under conditions that can be verified using our density estimator, the simulated maximum likelihood estimator converges to a true maximum likelihood estimator as the number of Monte Carlo replications grows large while keeping the data sample fixed. When the number of Monte Carlo replications grows and the number of data points grows, standard conditions ensure that the simulated likelihood estimator is consistent. Furthermore, if the number of Monte Carlo replications grows at the same rate as the data grows, then a simulated maximum likelihood estimator is asymptotically normal with the same asymptotic variance-covariance matrix as a true maximum likelihood estimator. As a result, our simulated likelihood

[2]We refer to Giles (2008) for an introduction to multilevel Monte Carlo simulation.

[3]See Kloeden & Platten (1999) for an overview of Euler's method.

[4]We do not consider infill asymptotic regimes, in which the time between consecutive observations of the data shrinks as more data becomes available.

estimators are asymptotically efficient in the sense that they have the same asymptotic standard errors as true maximum likelihood estimators.

An important property of the simulated likelihood estimators we propose is that, even though they are derived from Monte Carlo simulation, their asymptotic variance-covariance matrix is the same as that of true maximum likelihood estimators. This means that the Monte Carlo methodology we use to estimate the transition density does not affect the asymptotic distribution of the resulting parameter estimators. The reason why this key property holds is that our density estimator is unbiased. Were it not unbiased, then its bias would be transferred to the parameter estimators either by making them inconsistent or asymptotically inefficient. Detemple, Garcia & Rindisbacher (2006) establish this result in the diffusion case, and we conjecture that the same holds in the jump-diffusion case. Overall, the fact that our density estimator is unbiased is the main property that enables efficient parameter estimation in this paper.

Our density estimator has important computational features. It can be evaluated at any value of the parameter and arguments of the density function without re-simulation. A single set of Monte Carlo replications suffices to evaluate the density estimator at different arguments. This property entails significant computational benefits when carrying out parameter inference, especially for large data sets. It reduces the simulated likelihood maximization problem to a deterministic problem that can be solved using standard methods. Furthermore, our density estimator can be fine-tuned to minimize its variance for a given number of Monte Carlo replications. This feature makes our density estimator highly accurate in practical applications. Because it is derived from Euler's discretization method with different time steps, several Brownian increments can be re-used when carrying out Euler discretization. This property simplifies the computational work. Finally, given that our density estimator is constructed from independent Monte Carlo replications, its computation can be easily parallelized, yielding further computation benefits.

A numerical case study showcases the benefits of our density estimator and simulated likelihood estimators. We consider a stochastic volatility model with jumps in returns and volatility. The distribution of returns is non-Gaussian, and the distribution of volatility is asymmetric and skewed. This bivariate affine model has the advantage that its transition

density is known in closed-form. It can be recovered by Fourier inversion of the characteristic function as in Duffie, Pan & Singleton (2000). Because of these properties, the stochastic volatility model provides an appropriate case study to assess the performance of our estimation methodology. The numerical results show that our density estimator is highly accurate. It is able to capture the non-Gaussian distribution of returns, as well as the asymmetric distribution of volatility, both in the centers and the tails of the distributions. The density estimator becomes more accurate as the number of Monte Carlo replications grows large. It beats a naive biased density estimator derived from Euler's method in terms of accuracy achieved using up small to medium computational budgets. Our simulated likelihood estimators are also found to be highly accurate. They are able to closely recover the data-generating parameters.

1.1 Related methods

The methodology of this paper offers several advantages for parameter inference. The density estimator we propose is easy to compute and highly accurate, yielding precise parameter estimators. Our methodology enables consistent and asymptotically efficient parameter inference. The results hold for a general class of multivariate jump-diffusions processes, and they can be applied to any times series with any frequency of observations. Alternative methodologies generally do not satisfy all of these properties.

The method closest to ours is the one of Giesecke & Schwenkler (2014), who estimate the transition density of reducible jump-diffusions using exact simulation techniques. The density estimator of Giesecke & Schwenkler (2014) is computationally efficient for large computational budgets, and unbiased. Therefore, the parameter estimators of Giesecke & Schwenkler (2014) also inherit the asymptotic properties of maximum likelihood estimators. The method of Giesecke & Schwenkler (2014) is targeted primarily towards univariate jump-diffusions, which are reducible under mild conditions. However, even some of the most basic multivariate jump-diffusions are irreducible. For example, the standard stochastic volatility model of Heston (1993) is not reducible. Unlike the estimators of Giesecke & Schwenkler (2014), ours are applicable to the class of irreducible multivariate jump-diffusions. Nevertheless, our estimators may not be computationally efficient

for large computational budgets because the costs to evaluate our density estimator may become very large with very small probability.

If the model is affine as in Duffie et al. (2000), the transition density can be recovered via Fourier inversion of the characteristic function, which satisfies a system of ordinary differential equations. However, solving these ordinary differential equations and carrying out Fourier inversion numerically is computationally challenging in the multivariate case. Lo (1988) recovers the transition density of a jump-diffusion process with constant jump intensity and state-independent jump magnitudes by solving the Fokker-Planck equations governing it. This method is computationally burdensome for large data sets because the corresponding partial differential equations need to be solved recursively across data points. In addition, the numerical solution of the Fokker-Planck equations suffers from the curse of dimensionality, making it unsuitable for multivariate applications. In contrast to the methods of Duffie et al. (2000) and Lo (1988), our density estimator also applies for non-affine models with state-dependent jumps.

Inspired by the pioneering work of Aït-Sahalia (2002, 2008), Yu (2007) derives a small-time expansion approximation of the transition density of a multivariate jump-diffusion process with state-independent jump sizes. The coefficients of his expansion satisfy a set of interdependent partial differential equations. Solving these partial differential equations is computationally burdensome when the number of expansion terms is large, and when the jump-diffusion process is not reducible. The parameter estimators derived from the density estimator of Yu (2007) inherit the asymptotic properties of maximum likelihood estimator only when the time between consecutive observations shrinks to zero as more data points become available. In contrast, our simulated likelihood estimators inherit the asymptotic properties of maximum likelihood estimators under standard conditions as the number of data points grows while keeping the observation frequency of the data fixed. This type of asymptotic regime is common in many econometric applications.[5] Furthermore, the computational effort necessary to evaluate our density estimator does not depend on the reducibility of the process.

[5]Bibby & Sørensen (1995), Florens-Zmirou (1989), Giesecke & Schwenkler (2014), and Gobet, Hoffmann & Reiß (2004) consider similar asymptotic regimes.

Kristensen & Shin (2012) derive nonparametric estimators of the transition density of a jump-diffusion process with state-independent coefficient functions.[6] These authors apply a kernel estimator to samples of the jump-diffusion process derived from Euler discretization. If the bandwidth of the kernel estimator shrinks to zero as the number of data points grows large, then the parameter estimators derived from their density estimator inherit the asymptotic properties of maximum likelihood estimators. Their density estimator and ours are similarly inexpensive from a computational point of view. However, in contrast to Kristensen & Shin (2012), our density estimator also applies to jump-diffusions with state-dependent coefficient functions.

Moment-based methods can also be used for parameter inference. Jiang & Knight (2002), Chacko & Viceira (2003), Duffie & Glynn (2004), and Duffie & Singleton (1993) propose generalized method of moments estimators for continuous-time Markov processes. Should an infinite number of moments be used to perform estimation, then the moment-based parameter estimators inherit the asymptotic properties of maximum likelihood estimators as the number of data points grows large. However, the use of an infinite number of moments is infeasible in practical applications.[7]

Gourieroux, Monfort & Renault (1993) and Smith (1993) propose methods of indirect inference that are also applicable for multivariate jump-diffusions. Indirect inference requires that one is able to simulate from the jump-diffusion model. In addition, it requires that one specifies an auxiliary model. If the latter is correctly specified, then the parameter estimators derived from indirect inference inherit the asymptotic properties of maximum likelihood estimators. Unlike indirect inference, our estimation methodology does not require the specification of auxiliary models, and our simulated likelihood estimators inherit the asymptotic properties of maximum likelihood estimators under conditions that can be verified using our density estimator. Furthermore, our methodology is applicable for a general class of multivariate jump-diffusions. This is not the case for indirect inference because the exact simulation of multivariate jump-diffusions is infeasible unless

[6]The assumption that the distribution of ϵ_t is independent of t and θ in equation (1) of Kristensen & Shin (2012) effectively restricts their model to state-independent jump-diffusions.

[7]There are few cases in which maximum likelihood efficiency can be achieved with a finite number of moments. See, e.g., Carrasco, Chernov, Florens & Ghysels (2007) and Jiang & Knight (2010).

the process is reducible.[8]

1.2 Structure of the paper

The rest of this paper is organized as follows. Section 2 formulates the model and the estimation problem. In Section 3, we derive our density representation. We introduce the density estimator in Section 4, and discuss its computational properties in Section 5. Section 6 proposes simulated likelihood estimators and summarizes their asymptotic properties. A numerical case study is carried out in Section 7.

2 Problem formulation

Fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a right-continuous, complete information filtration $(\mathcal{F}_t)_{t\geq 0}$. Let X be a jump-diffusion process valued in $\mathcal{S} \subset \mathbb{R}^d$ that is governed by the stochastic differential equation

$$\mathrm{d}X_t = \mu(X_t;\theta)\mathrm{d}t + \Sigma(X_t;\theta)\mathrm{d}B_t + \mathrm{d}L_t, \tag{1}$$

where $X_0 \in \mathcal{S}$ is fixed and known, $\mu : \mathcal{S}\times\Theta \to \mathbb{R}^d$ is the drift function, $\Sigma : \mathcal{S}\times\Theta \to \mathbb{R}^{d\times d}$ is the positive definite volatility matrix function, B is a standard d-dimensional Brownian motion, and L is a jump process of the type

$$L_t = \sum_{n=1}^{N_t} \Gamma(X_{T_n-}, D_n;\theta) \tag{2}$$

with event stopping times $(T_n)_{n\geq 1}$ and jump intensity $\lambda_t = \Lambda(X_t;\theta)$ for a function $\Lambda : \mathcal{S}\times\Theta \to \mathbb{R}_+$. Here, $X_{t-} = \lim_{s\nearrow t} X_s$. The jump magnitudes of the process X are determined by the function $\Gamma : \mathcal{S}\times\mathcal{D}\times\Theta \to \mathbb{R}^d$. The mark variables $(D_n)_{n\geq 1}$, which characterize the jumps of X, are independent and identically distributed in $\mathcal{D} \subset \mathbb{R}$ with probability density π. The drift, volatility, jump intensity, and jump size functions are specified by a parameter $\theta \in \Theta$ to be estimated, where the parameter space Θ is a subset of Euclidean space.

[8] We refer to Giesecke & Smelov (2013) for the exact simulation of reducible jump-diffusions. Henry-Labordère, Tan & Touzi (2015) develop exact simulation tools for multivariate diffusions.

Overall, X is a Markov process with infinitesimal generator for functions $f : \mathbb{R}^d \to \mathbb{R}$ with bounded and continuous first and second order derivatives given by:

$$\mathcal{A}_\theta f(x) = \sum_{i=1}^{d} \mu_i(x;\theta) \frac{\partial f(x)}{\partial x_i} + \frac{1}{2} \sum_{1 \le i,j \le d} \left(\Sigma(x;\theta)\Sigma(x;\theta)^T\right)_{i,j} \frac{\partial^2 f(x)}{\partial x_i x_j}$$
$$+ \Lambda(x;\theta) \int_{\mathcal{D}} (f(x + \Gamma(x,u;\theta)) - f(x))\pi(u)du.$$

We impose the following assumptions. First, the boundary of $\mathcal{S}$ is either unattainable or absorbing if attainable. Second, the parameter space Θ is a compact subset of $\mathbb{R}^r$ with non-empty interior. Third, there exist a unique strong solution (X, J) of the above system; sufficient conditions are given in Protter (2004). We focus on the case of constant observation frequencies, i.e., $t_i - t_{i-1} = \Delta$ for all i, although all results hold for mixed observation frequencies as long as $\sup_{i \ge 1} |t_i - t_{i-1}| < \infty$. We also assume for simplicity that the process N and the mark variables $(D_n)_{n \ge 1}$ are one-dimensional, and that the jump mark density π is parameter independent. Generalizations and extensions of these assumptions are straightforward. Finally, we assume that X admits a transition density. Cass (2009), Filipović, Mayerhofer & Schneider (2013), Komatsu & Takeuchi (2001), and Takeuchi (2002) provide sufficient conditions.

We use the following notation throughout the paper. A subscript in $\mathbb{P}_\theta$ or $\mathbb{E}_\theta$ indicates that the parameter determining the law of the stochastic process X in (1) is θ. The gradient and the Hessian matrix operators are denoted by ∇ and ∇^2, respectively. For any $1 \le \nu, \iota, \kappa \le r$, write ∂_ν, $\partial^2_{\nu,\iota}$, and $\partial^3_{\nu,\iota,\kappa}$ for the first, second, and third partial derivatives with respect to θ_ν, θ_ι, and θ_κ.

2.1 Inference problem

Suppose that there exists a parameter $\theta^* \in \operatorname{int}\Theta$ such that the paths of X satisfy the SDE (1) for $\theta = \theta^*$. We say that θ^* is the true parameter. Our goal is to estimate θ^* given a sequence of observations of X sampled at the fixed and deterministic times $0 = t_0 < \ldots < t_m < \infty$. We will use the method of maximum likelihood.

The data $\mathbf{X}_m = \{X_{t_0}, \ldots, X_{t_m}\}$ is a random variable valued in $\mathcal{S}^m$ and measurable with respect to $\mathcal{B}^m$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{S}$. The likelihood of the data

is the Radon-Nikodym density of the law of $\mathbf{X}_m$ with respect to the Lebesgue measure on $(\mathcal{S}^m, \mathcal{B}^m)$. Letting $p_t(x, .; \theta)$ be the Radon-Nikodym density of the law of X_t given $X_0 = x$ with respect to the Lebesgue measure on $(\mathcal{S}, \mathcal{B})$ (the transition density of X), the likelihood of θ at the data $\mathbf{X}_m$ takes the form

$$\mathfrak{L}_m(\theta) = \prod_{i=1}^{m} p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta) \tag{3}$$

due to the Markovian structure of (1). The maximum likelihood estimator (MLE) satisfies

$$\hat{\theta}_m \in \arg\max_{\theta \in \Theta} \mathfrak{L}_m(\theta) \tag{4}$$

almost surely. We only consider interior MLEs that satisfy the first order condition

$$\nabla \mathfrak{L}_m(\hat{\theta}_m) = 0. \tag{5}$$

Maximum likelihood inference requires that one is able to evaluate the density p_Δ. This is generally not possible for the broad class of jump-diffusion models we consider. We will therefore proceed to construct an unbiased estimator of the density p_Δ, and use this density estimator to compute maximum likelihood estimators based on (3).

3 Density representation

Consider the random variable

$$Z_\Delta(\theta) \;=\; \exp\left(\int_0^\Delta (\Lambda(X_s; \theta) - \ell)\, \mathrm{d}s\right) \prod_{n=1}^{N_\Delta} \frac{\ell}{\Lambda(X_{T_n-}; \theta)} \tag{6}$$

for $\theta \in \Theta$ and $\ell > 0$. If $\mathbb{E}_\theta[Z_\Delta(\theta)] = 1$, then $Z_\Delta(\theta)$ defines an equivalent probability measure $\mathbb{Q}_\theta$ on $(\Omega, \mathcal{F}_\Delta)$ given by $\mathbb{Q}_\theta[A] = \mathbb{E}_\theta[Z_\Delta(\theta)\mathbf{1}_A]$ for any $A \in \mathcal{F}_\Delta$. The theorems of Lévy and Watanabe imply that, under $\mathbb{Q}_\theta$ and on $[0, \Delta]$, N is a Poisson process with rate ℓ; see Brémaud (1980). Consequently, jumps of the process X arrive at a constant rate under $\mathbb{Q}_\theta$. Between jump times, X follows a diffusive process without jumps. These insights yield a novel representation of the density p_Δ, summarized in the following theorem.

Theorem 3.1. *Fix $\ell > 0$. Suppose the following assumptions hold.*

(A1) For any $\theta \in \Theta$, the variable $Z_\Delta(\theta)$ has unit expectation, $\mathbb{E}_\theta[Z_\Delta(\theta)] = 1$.

(A2) For any $\theta \in \Theta$, the process $(X_t : t \in [0, \Delta])$ is a strong Markov process under $\mathbb{Q}_\theta$.

Let $\tilde{X}$ be the solution to the SDE

$$\mathrm{d}\tilde{X}_t = \mu(\tilde{X}_t; \theta)\mathrm{d}t + \Sigma(\tilde{X}_t; \theta)\mathrm{d}\tilde{B}_t, \quad \tilde{X}_0 \in \mathcal{S}, \tag{7}$$

for a standard Brownian motion $\tilde{B}$ independent of B. Let $\tilde{p}_t(v, \cdot; \theta)$ denote the $\mathbb{P}_\theta$-transition density of $\tilde{X}_t$ given $\tilde{X}_0 = v$. Then,

$$p_\Delta(v, w; \theta) = \mathbb{E}_\theta^{\mathbb{Q}} \left[\frac{\tilde{p}_{\Delta - T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{Z_\Delta(\theta)} \,\middle|\, X_0 = v \right] \tag{8}$$

for any $0 \leq t \leq \Delta$, $v, w \in \mathcal{S}$, and $\theta \in \Theta$.

The density representation of Theorem 3.1 consists of a mixture of transition densities of diffusion processes of the type (7). It is an implication of Bayes' formula. Under Assumption (A2) and conditional on $(N_\Delta, (T_n)_{n \leq N_\Delta}, (X_{T_n})_{n \leq N_\Delta})$, that is, conditional on the number of jumps of X before time Δ, the realizations of all jump times before Δ, and the values of X at all jump times before Δ, the transition of X from time 0 to time Δ is governed only by the law of X from the last jump time T_{N_Δ} until time Δ. Given that no jumps occur in the time interval $(T_{N_\Delta}, \Delta]$, the law of X during this time interval is the same as the law of the diffusive process (7). As a result, under Assumption (A2) and conditional on $(N_\Delta, (T_n)_{n \leq N_\Delta}, (X_{T_n})_{n \leq N_\Delta})$, the density of X for a transition from v at time 0 to w at time Δ is equal to the density $\tilde{p}_{\Delta - T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)$ with $X_0 = v$. Bayes' formula tells us that we can recover the unconditional density p_Δ by integrating out according to the law of $(N_\Delta, (T_n)_{n \leq N_\Delta}, (X_{T_n})_{n \leq N_\Delta})$. This is done by taking the expectation in (8). The term $1/Z_\Delta(\theta)$ in expression (8) accounts for the change of measure, which significantly simplifies the estimation of the density in Section 4. Assumption (A1) guarantees that the change of measure is well-defined. It is a standard regularity assumption; Blanchet & Ruf (2013) give sufficient conditions. Assumption (A2) is also standard; see Protter (2004, Theorem 32). The diffusion density $\tilde{p}_t$ exists if the jump-diffusion density p_t exists.

The density representation (8) complements the recently developed density representation of Giesecke & Schwenkler (2014), who characterize the transition density of the process as a mixture of Gaussian densities. This is possible because Giesecke & Schwenkler (2014) consider a transformation of the jump-diffusion process known as the Lamperti transform, which has unit volatility. When the underlying process is univariate, the Lamperti transform exists under mild conditions. In the multivariate case, on the other hand, the Lamperti transform exists only when the process is reducible in the sense of Aït-Sahalia (2008). As a result, the density representation of Giesecke & Schwenkler (2014) is restricted to the class of reducible multivariate jump-diffusions. In contrast, we are not restricted to the class of models for which the Lamperti transform exists. Consequently, the density representation (8) also applies to irreducible processes. Many models of practical relevance are not reducible. For example, the stochastic volatility model of Heston (1993) is not reducible, but it is extensively used in the options pricing literature.[9]

Theorem 3.1 significantly extends the well-known density representations of Dacunha-Castelle & Florens-Zmirou (1986) and Rogers (1985). These representations apply only in the univariate diffusion case; i.e., when $\Gamma \equiv 0$ and $d = 1$. In contrast, our density representation also applies in the multivariate jump-diffusion case.

The representation (8) also facilitates the derivation of conditions under which the transition density is smooth with respect to the parameter θ. Smoothness is necessary for consistency and asymptotic normality of maximum likelihood estimators. For smoothness of the density, we only require smoothness of the coefficient functions and an integrability condition, which can be verified using the density estimator we introduce in Section 4. Our conditions for smoothness are easier to verify in practical settings and less restrictive than alternative conditions, which often times require that the coefficient functions have bounded derivatives of all orders (see, e.g., Cass (2009), Komatsu & Takeuchi (2001), and Takeuchi (2002)).

[9]See, e.g., Andersen, Benzoni & Lund (2002), Eraker, Johannes & Polson (2003), and Eraker (2004), among many others.

Proposition 3.2. *Suppose that the conditions of Theorem 3.1 hold. Suppose also that the following conditions hold:*

(A3) The partial derivatives up to n-th order of $\Phi_\Delta(x,y;\theta) = \tilde{p}_{\Delta-T_{N_\Delta}}(x,y;\theta)\frac{1}{Z_\Delta(\theta)}$ *are uniformly bounded in expectation in the following sense: For all* $1 \leq k \leq n$ *and* $q_1,\ldots,q_k \in \{\theta_1,\ldots,\theta_r,v,w\}$,

$$\mathbb{E}_\theta^{\mathbb{Q}}\left[\sup_{\theta\in\Theta}\sup_{v,w\in\mathcal{S}}\frac{\partial^k}{\partial q_1\ldots\partial q_k}\tilde{p}_{\Delta-T_{N_\Delta}}(v,w;\theta)\frac{1}{Z_\Delta(\theta)}\right] < \infty.$$

(A4) The drift function μ*, volatility matrix function* Σ*, jump intensity function* Λ*, jump magnitude function* Γ*, and diffusive density* $\tilde{p}$ *are n-times continuously differentiable with respect to all of their arguments.*

Then $\theta \mapsto p_\Delta(v,w;\theta)$ *is n-times continuously differentiable for any* $v,w \in \mathcal{S}$.

4 Density estimator

Evaluating the transition density of the jump-diffusion X is challenging given that the law of X is intractable in many applications. A key advantage of the density representation (8) is that it can be efficiently approximated by exploiting a randomization technique introduced by Glynn & Rhee (2015). This yields an unbiased density estimator. In this section, we introduce our density estimator, and analyze its convergence properties.

4.1 Towards an unbiased estimator

Under $\mathbb{Q}_\theta$, jumps of X arrive with constant intensity ℓ. As a result, samples of N_Δ can be simulated without bias using a standard inverse method. Conditional on N_Δ, the distribution of the jump times $(T_n)_{n\leq N_\Delta}$ is the same as that of the order statistics of N_Δ uniform random variables on $[0,\Delta]$. Samples of the jump times $(T_n)_{n\leq N_\Delta}$ conditional on N_Δ can therefore also be simulated without bias. If the diffusive density $\tilde{p}$ is known in closed form, and samples of $(X_{T_{N_\Delta}}, 1/Z_\Delta(\theta))$ can be simulated without bias. Then,

$$\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}},w;\theta)}{Z_\Delta(\theta)}$$

given $X_0 = v$ is an unbiased estimator of (8) that can be sampled exactly via Monte Carlo simulation. In most applications, however, the diffusive density $\tilde{p}$ is not known in closed form, and one cannot sample exactly from the distribution of $1/Z_\Delta(\theta)$. We circumvent these issues by taking several steps, which we summarize below.

4.1.1 Euler discretization

Note that $Z_\Delta^{-1}(\theta)$ is an exponential martingale that satisfies the following SDE under $\mathbb{Q}_\theta$:

$$\mathrm{d}Z_t^{-1}(\theta) = -Z_{t-}^{-1}(\theta)\left(\frac{\Lambda(X_{t-};\theta)}{\ell} - 1\right)(\ell \mathrm{d}t - \mathrm{d}N_t), \quad Z_0^{-1}(\theta) = 1. \tag{9}$$

We can generate an approximation of $(X_{T_{N_\Delta}}, Z_\Delta^{-1}(\theta))$ using Euler discretization. To do this, we first generate exact samples of N_Δ and also samples of $(T_n)_{n\le N_\Delta}$ conditional on N_Δ. Between sampled jump times, we approximate the dynamics of X and Z^{-1} via Euler discretization with J steps. Letting (X^J, Z^{-J}) denote the Euler discretization of (X, Z^{-1}), we initialize $X_{0,0}^J = X_0$ and $Z_{0,0}^{-J} = 1$, and set

$$X_{n,j}^J = \begin{cases} X_{n,j-1}^J + \mu\left(X_{n,j-1}^J;\theta\right)h_n + \Sigma\left(X_{n,j-1}^J;\theta\right)\left(B_{jh_n} - B_{(j-1)h_n}\right), & 1 \le j \le J, \\ X_{n-1,J}^J + \Gamma\left(X_{n-1,J}^J, D_n;\theta\right), & n > 0, j = 0, \end{cases}$$

$$Z_{n,j}^{-J} = \begin{cases} Z_{n,j-1}^{-J} - Z_{n,j-1}^{-J}\left(\Lambda\left(X_{n,j-1}^J;\theta\right) - \ell\right)h_n, & 1 \le j \le J, \\ \frac{\Lambda\left(X_{n-1,J}^J;\theta\right)}{\ell} Z_{n-1,J}^{-J}, & n > 0, j = 0, \end{cases}$$

for $0 \le n \le N_\Delta$ and $h_n = \frac{T_n - T_{n-1}}{J}$, where we have used the notation $T_0 = 0$ and $T_{N_\Delta+1} = \Delta$ for simplicity. This construction ensures that the two Euler discretizations between consecutive jump times are correctly pasted together by accounting for the jumps of X and Z^{-1}. The nature of the Euler discretization implies that $(\mathcal{X}^J, \mathcal{Z}^J) = (X_{N_\Delta,0}^J, Z_{N_\Delta,J}^{-J})$ is a biased estimator of $(X_{T_{N_\Delta}}, Z_\Delta^{-1}(\theta))$. Consequently,

$$\tilde{p}_{\Delta - T_{N_\Delta}}\left(\mathcal{X}^J, w;\theta\right)\mathcal{Z}^J$$

is a biased estimator of the density $p_\Delta(v, w;\theta)$ in (8).

4.1.2 Diffusion density

Next, we approximate the diffusion density $\tilde{p}$. This can also be done via Euler discretization. For a given sample of (N_Δ, T_{N_Δ}), we discretize the diffusive process $\tilde{X}$ between time

0 and time $\Delta - T_{N_\Delta}$ in an analogous way as for X^J, but using I Euler steps instead of J. Let $(\tilde{X}_i^I)_{0\leq i\leq I}$ denote the Euler discretization of $\tilde{X}$ with Euler step size $\tilde{h} = \frac{\Delta - T_{N_\Delta}}{I}$ obtained this way. Conditional on T_{N_Δ} and $\tilde{X}_0^I$, the law of $\tilde{X}_I^I$ is mixed Gaussian because each increment in the Euler discretization is normally distributed. More precisely, the conditional density of $\tilde{X}_I^I$ given T_{N_Δ} and $\tilde{X}_0^I = v$ is

$$\tilde{\mathcal{P}}^I(v, w; \theta) = \int \prod_{i=1}^{I} \phi\left(x_i; x_{i-1}, \tilde{h}\right) \mathrm{d}x_1 \ldots \mathrm{d}x_{I-1} \tag{10}$$

where $x_0 = v$, $x_I = w$, and $\phi(\cdot; x, h)$ is the density of the d-dimensional normal distribution with mean $x + \mu(x;\theta)h$ and variance-covariance matrix $h\Sigma(x;\theta)\Sigma^\top(x;\theta)$. The mixed normal density (10) can be computed using standard numerical routines; see Section 5. We know from Bally & Talay (1996) that the difference between the Euler density $\tilde{\mathcal{P}}^I$ and the true density $\tilde{p}$ is of order $O(I^{-1})$. Thus, $\tilde{\mathcal{P}}^I$ serves as a first-order approximation of $\tilde{p}$.

We can now compute an estimator of the density representation (8), namely

$$\hat{p}_\Delta^{I,J}(v, w; \theta) = \tilde{\mathcal{P}}^I\left(\mathcal{X}^J, w; \theta\right) \mathcal{Z}^J. \tag{11}$$

The estimator (11) can be computed for a general class of jump-diffusion models characterized by SDE's of the type (1) given that is solely based on Euler discretization. In addition, the estimator (11) is asymptotically unbiased as $I \to \infty$ and $J \to \infty$. That is,

$$\lim_{I,J\to\infty} \mathbb{E}_\theta^{\mathbb{Q}}\left[\hat{p}_\Delta^{I,J}(v, w; \theta) \,\middle|\, X_0 = v\right] = p_\Delta(v, w; \theta).$$

4.1.3 Randomization

One drawback of the density estimator (11) is that it is biased by construction for any finite I and J. If one were to carry out maximum likelihood estimation based on this biased density estimator, then the resulting parameter estimators may have a distorted asymptotic distribution even if $I \to \infty$ and $J \to \infty$ as the data sample grows. This may result in asymptotically inefficient or asymptotically biased parameter estimators.[10] To avoid these issues, we exploit a randomization technique introduced by Glynn & Rhee (2015) to construct an unbiased density estimator.

[10] See Detemple et al. (2006) and Giesecke & Schwenkler (2014).

Suppose Ξ is a random variable valued in $\mathbb{N}_0$ and measurable with respect to $\mathcal{F}_0$. Assume that the distribution of Ξ is independent of the parameter θ and the initial value X_0, and write $q_n = \mathbb{Q}_\theta[\Xi = n]$. Consider subsequences I_ξ and J_ξ so that $I_\xi, J_\xi \to \infty$ as $\xi \to \infty$. The asymptotic unbiasedness of the estimator (11) implies that, under certain regularity conditions, we can rewrite the density representation (8) as follows:

$$\begin{aligned} p_\Delta(v,w;\theta) &= \lim_{\xi\to\infty} \mathbb{E}_\theta^{\mathbb{Q}} \left[\hat{p}_\Delta^{I_\xi,J_\xi}(v,w;\theta) \,\middle|\, X_0 = v \right] \\ &= \sum_{\xi\geq 0} \mathbb{E}_\theta^{\mathbb{Q}} \left[\hat{p}_\Delta^{I_\xi,J_\xi}(v,w;\theta) - \hat{p}_\Delta^{I_{\xi-1},J_{\xi-1}}(v,w;\theta) \,\middle|\, X_0 = v \right] \\ &= \sum_{\xi\geq 0} \mathbb{E}_\theta^{\mathbb{Q}} \left[\frac{\hat{p}_\Delta^{I_\xi,J_\xi}(v,w;\theta) - \hat{p}_\Delta^{I_{\xi-1},J_{\xi-1}}(v,w;\theta)}{q_\xi} \,\middle|\, X_0 = v \right] q_\xi \\ &= \mathbb{E}_\theta^{\mathbb{Q}} \left[\frac{\hat{p}_\Delta^{I_\Xi,J_\Xi}(v,w;\theta) - \hat{p}_\Delta^{I_{\Xi-1},J_{\Xi-1}}(v,w;\theta)}{q_\Xi} \,\middle|\, X_0 = v \right] \end{aligned} \tag{12}$$

where we have set $I_{-1} = J_{-1} = 0$. The last equality follows because Ξ is $\mathcal{F}_0$-measurable and independent of θ and X_0. The calculations in (12) imply that

$$\frac{\hat{p}_\Delta^{I_\Xi,J_\Xi}(v,w;\theta) - \hat{p}_\Delta^{I_{\Xi-1},J_{\Xi-1}}(v,w;\theta)}{q_\Xi}$$

is an unbiased estimator of the transition density $p_\Delta(v,w;\theta)$.

4.2 Estimator

The steps in the previous section yield an unbiased density estimator that is applicable for a general class of jump-diffusion models. We summarize in the theorem below. For simplicity, write

$$\mathcal{D}_\Delta^\xi(v,w;\theta) = \hat{p}_\Delta^{I_\xi,J_\xi}(v,w;\theta) - \hat{p}_\Delta^{I_{\xi-1},J_{\xi-1}}(v,w;\theta).$$

Theorem 4.1. *Fix $\Delta > 0$ and sequences $(J_\xi : \xi \in \mathbb{N}_0)$ and $(I_\xi : \xi \in \mathbb{N}_0)$. Let Ξ be an $\mathcal{F}_0$-measurable random variable valued in $\mathbb{N}_0$, with distribution given by $q_\xi = \mathbb{Q}_\theta[\Xi = \xi]$ that is independent of the parameter θ and the initial value X_0. Let $(\mathcal{X}^J, \mathcal{Z}^J)$ be samples of $(X_{T_{N_\Delta}}, Z_\Delta^{-1}(\theta))$ constructed via Euler discretization with J steps between consecutive jump times. In addition, let $\tilde{\mathcal{P}}^I$ be a mixed Gaussian density as in* (10) *derived from Euler*

discretization of $\tilde{X}$ with I steps. Assume that the conditions of Theorem 3.1 are valid. In addition, suppose that the following condition also holds.

(B1) For any $\theta \in \Theta$, and $v, w \in \mathcal{S}$,

$$\sum_{\xi \geq 0} \frac{\left\| \tilde{p}_{\Delta}^{I_\xi, J_\xi}(v, w; \theta) - p_{\Delta}(v, w; \theta) \right\|_2^2}{q_\xi} < \infty.$$

Then, for any $v, w \in \mathcal{S}$ and $\theta \in \Theta$,

$$\hat{p}_{\Delta}(v, w; \theta) = \frac{\mathcal{D}_{\Delta}^{\Xi}(v, w; \theta)}{q_\Xi} \tag{13}$$

is an unbiased estimator of $p_{\Delta}(v, w; \theta)$.

The main advantage of the estimator (13) is that it is unbiased for any $v, w \in \mathcal{S}$, $\theta \in \Theta$, and $\Delta > 0$. This property generates key benefits when performing maximum likelihood estimation of the jump-diffusion model (1) based on the density estimator $\hat{p}_{\Delta}$. In particular, the unbiasedness property ensures that one can always implement a version of the density estimator $\hat{p}_{\Delta}$ which, when used for maximum likelihood inference, results in asymptotically efficient and asymptotically unbiased parameter estimators; see Giesecke & Schwenkler (2014). This is generally not possible if one were to use the biased density estimator $\hat{p}_{\Delta}^{I,J}$ in (11), as proved by Detemple et al. (2006). We will discuss in detail the implementation of the density estimator $\hat{p}_{\Delta}$ and the asymptotic properties of parameter estimators derived from this density estimator in the following sections.

We conclude this section by emphasizing that $\hat{p}_{\Delta}(v, w; \theta)$ can be differentiated under certain conditions to obtain unbiased estimators of the partial derivatives of the transition density. Partial derivatives of the density are necessary in many econometric applications.

Proposition 4.2. *Suppose that the conditions of Proposition 3.2 and Theorem 4.1 are satisfied. Furthermore, suppose:*

(B2) The partial derivatives up to n-th order of $\hat{p}_{\Delta}$ with respect to θ are uniformly bounded in expectation in the following sense: For all $1 \leq k \leq n$ and $i_1, \ldots, i_k \in \{1, \ldots, r\}$,

$$\mathbb{E}_{\theta}^{\mathbb{Q}} \left[\sup_{\theta \in \Theta} \sup_{v, w \in \mathcal{S}} \partial_{i_1, \ldots, i_k}^{k} \hat{p}_{\Delta}(v, w; \theta) \right] < \infty.$$

Then, $\theta \mapsto \hat{p}_\Delta(v, w; \theta)$ is almost-surely n-times continuously differentiable for any $v, w \in \mathcal{S}$. In addition, any n-th partial derivative of $\hat{p}_\Delta(v, w; \theta)$ with respect to θ is an unbiased estimator of the corresponding derivative of $p_\Delta(v, w; \theta)$. That is,

$$\mathbb{E}_\theta^{\mathbb{Q}} \left[\partial^n_{i_1,\ldots,i_n} \hat{p}_\Delta(v, w; \theta) \right] = \partial^n_{i_1,\ldots,i_n} p_\Delta(v, w; \theta) \quad \textit{for all} \quad i_1, \ldots, i_n \in \{1, \ldots, r\}.$$

5 Computation of the density estimator

Computing the density estimator $\hat{p}_\Delta$ requires that one specifies choices for the sequences $(I_\xi)_{\xi \geq 0}$ and $(J_\xi)_{\xi \geq 0}$, the distribution $(q_\xi)_{\xi \geq 0}$ of the random variable Ξ, the Poisson rate $\ell > 0$, and the numerical methodology to compute the mixed normal density $\tilde{\mathcal{P}}^I$. In this section, we propose an implementation of our density estimator that ensures that the density estimator has finite variance while minimizing the computational need.

5.1 Finite variance

We begin by implementing an estimator of $\tilde{\mathcal{P}}^I$. A simple unbiased estimator of $\tilde{\mathcal{P}}^I$ can be constructed via Monte Carlo simulation. For given I and T_{N_Δ}, compute H i.i.d. samples of the Euler discretization $(\tilde{X}^I_i)_{0 \leq i \leq I}$ of $\tilde{X}$ on $[0, \Delta - T_{N_\Delta}]$. Following Pedersen (1995), we estimate $\tilde{\mathcal{P}}^I$ via its Monte Carlo counterpart

$$\tilde{\mathcal{P}}^{H,I}(v, w; \theta) = \frac{1}{H} \sum_{\nu=1}^{H} \phi \left(w; \tilde{X}^{I,\nu}_{I-1}, \tilde{h} \right), \tag{14}$$

where $\tilde{h} = \frac{\Delta - T_{N_\Delta}}{I}$, $\tilde{X}^{I,\nu}_{I-1}$ is the ν-th sample of $\tilde{X}^I_{I-1}$, and $\tilde{X}^{I,\nu}_0 = v$ for all $1 \leq \nu \leq H$. This yields an unbiased estimator of $\tilde{\mathcal{P}}^I(v, w; \theta)$. We can therefore replace $\tilde{\mathcal{P}}^I$ with $\tilde{\mathcal{P}}^{H,I}$ in (11), and the density estimator $\hat{p}_\Delta$ remains unbiased. In other words, if we set

$$\begin{aligned} \hat{p}^{H,I,J}_\Delta(v, w; \theta) &= \tilde{\mathcal{P}}^{H,I} \left(\mathcal{X}^J, w; \theta \right) \mathcal{Z}^J, \\ \mathcal{D}^\xi_\Delta(v, w; \theta) &= \hat{p}^{H_\xi, I_\xi, J_\xi}_\Delta(v, w; \theta) - \hat{p}^{H_{\xi-1}, I_{\xi-1}, J_{\xi-1}}_\Delta(v, w; \theta), \end{aligned}$$

then the result of Theorem 4.1 remains unchanged.

It is well-known that Euler discretization has strong rate of convergence of order $1/2$ (see, e.g., Jacod & Protter (1998)). In our case, because we carry out Euler discretization

between consecutive jump times of X under $\mathbb{Q}_\theta$, we have

$$\left\|\mathcal{Z}^J - Z_\Delta^{-1}(\theta)\right\|_2 = O\left(\ell \Delta J^{-1/2}\right). \tag{15}$$

A key result by Gobet & Labart (2008) implies that

$$\left\|\tilde{\mathcal{P}}^{H,I}(v,w;\theta) - \tilde{p}_{\Delta - T_{N_\Delta}}(v,w;\theta)\right\|_2^2 = O\left(I^{-2} + H^{-1}\mathrm{Var}_\theta^{\mathbb{Q}}\left(\tilde{\mathcal{P}}^{1,I}(v,w;\theta)\right)\right). \tag{16}$$

Setting $V_{I,\theta} = \Sigma(\tilde{X}_{I-1}^I;\theta)\Sigma(\tilde{X}_{I-1}^I;\theta)^\top$ and $\tilde{X}_{I,\theta} = \tilde{X}_{I-1}^I + \mu(\tilde{X}_{I-1}^I;\theta)\tilde{h}$, we can show that

$$\mathrm{Var}_\theta^{\mathbb{Q}}\left(\tilde{\mathcal{P}}^{1,I}(v,w;\theta)\right) \leq \mathbb{E}_\theta^{\mathbb{Q}}\left[\frac{e^{-\frac{1}{\tilde{h}}\left(w-\tilde{X}_{I,\theta}\right)^\top V_{I,\theta}^{-1}\left(w-\tilde{X}_{I,\theta}\right)}}{\tilde{h}^d(2\pi)^d \det V_{I,\theta}}\right] = O\left(I^{d/2}\right)$$

for all $\theta \in \Theta$, $\ell > 0$, and $v, w \in \mathcal{S}$. In light of these results, we set $H_\xi = O(I_\xi^{2+d/2})$ and fix I_ξ as to equalize the rates of convergence of (15) and (16) for any given ξ. This can be achieved by selecting $I_\xi = O(\sqrt{J_\xi})$. Under sufficient regularity conditions, this choice guarantees that

$$\left\|\hat{p}_\Delta^{H_\xi, I_\xi, J_\xi}(v,w;\theta) - p_\Delta(v,w;\theta)\right\|_2 = O\left(J_\xi^{-0.5}\right).$$

In other words, the mean-squared error of the biased density estimator $\hat{p}_\Delta^{H,I,J}$ converges to zero at the canonical rate of $1/2$. We can now construct an unbiased density estimator with finite variance.

Proposition 5.1. *Fix $I_\xi = O(J_\xi^{1/2})$ and $H_\xi = O(J_\xi^{1+d/4})$ for $\xi \geq 0$. Suppose that the conditions of Theorem 4.1 are satisfied. Assume that the following conditions are also valid.*

(C1) In the limit $J \to \infty$, the following asymptotic behavior holds for any $\theta \in \Theta$:

$$\left\|\mathcal{Z}^J - \frac{1}{Z_\Delta(\theta)}\right\|_2 = O\left(J^{-0.5}\right).$$

(C2) The following asymptotic behavior holds for $v, w \in \mathcal{S}$ and $\theta \in \Theta$ in the limit $I \to \infty$:

$$\left\|\tilde{\mathcal{P}}^I(v,w;\theta) - \tilde{p}_{\Delta - T_{N_\Delta}}(v,w;\theta)\right\|_2 = O\left(I^{-1}\right).$$

(C3) The determinant of $\Sigma\Sigma^\top$ is bounded away from zero. That is,

$$\inf_{\theta\in\Theta}\inf_{x\in\mathcal{S}} \det\left(\Sigma(x;\theta)\Sigma(x;\theta)^\top\right) > 0.$$

Define

$$\hat{p}_{\Delta}^{H,I,J}(v,w;\theta) = \tilde{\mathcal{P}}^{H,I}\left(\mathcal{X}^{J},w;\theta\right)\mathcal{Z}^{J},$$

$$\mathcal{D}_{\Delta}^{\xi}(v,w;\theta) = \hat{p}_{\Delta}^{H_{\xi},I_{\xi},J_{\xi}}(v,w;\theta) - \hat{p}_{\Delta}^{H_{\xi-1},I_{\xi-1},J_{\xi-1}}(v,w;\theta).$$

Then,

$$\hat{p}_{\Delta}(v,w;\theta) = \frac{\mathcal{D}_{\Delta}^{\Xi}(v,w;\theta)}{q_{\Xi}}$$

is an unbiased estimator of $p_{\Delta}(v,w;\theta)$ *for any* $v,w \in \mathcal{S}$ *and* $\theta \in \Theta$*, and the variance of* $\hat{p}_{\Delta}$ *is finite:*

$$\left\|\hat{p}_{\Delta}(v,w;\theta) - p_{\Delta}(v,w;\theta)\right\|_2 < \infty.$$

We remark that sufficient conditions for Condition (C1) are given by Higham, Mao & Stuart (2003), Jacod & Protter (1998), and Yan (2002), among many others. Sufficient conditions for Condition (C2) are given by Bally & Talay (1996), Gobet & Labart (2008), Guyon (2006), and Konakov & Mammen (2002).

5.2 Computational properties

The density estimator $\hat{p}_{\Delta}$ has error. That is, $\hat{p}_{\Delta}(v,w;\theta) \neq p_{\Delta}(v,w;\theta)$ almost surely even though $\mathbb{E}_{\theta}^{\mathbb{Q}}[\hat{p}_{\Delta}(v,w;\theta)] = p_{\Delta}(v,w;\theta)$. A natural question to ask is: How much computational work is necessary to estimate the density so that a certain error bound is not violated with high probability? The answer to this question gives a sense of the computational complexity of a density estimator.

Given that the variance of $\hat{p}_{\Delta}$ is bounded, a starting point to evaluate the computational complexity of our density estimator is Monte Carlo simulation. Define $\hat{p}_{\Delta}^{K}(v,w;\theta)$ as the Monte Carlo estimator given by the average of K independent samples of the unbiased estimator $\hat{p}(v,w;\theta)$. It is well understood that the variance of the Monte Carlo estimator $\hat{p}_{\Delta}^{K}$ converges to zero as we let the number K of Monte Carlo samples grow infinitely large. Therefore, if we want to achieve

$$\mathbb{Q}_{\theta}\Big[\left\|\hat{p}_{\Delta}^{K}(v,w;\theta) - p_{\Delta}(v,w;\theta)\right\|_2 \leq \epsilon\Big] \geq 1-\delta$$

for some $\epsilon,\delta > 0$, we need to choose K sufficiently large.

Evaluating the Monte Carlo estimator $\hat{p}_\Delta^K$ for large K is computationally expensive. Given that $I_\Xi = O(J_\Xi^{1/2})$ and $H_\Xi = O(J_\Xi^{1+d/4})$, the computational costs are driven by the realizations of J_Ξ. The value of J_Ξ may be large whenever Ξ is large, increasing the computational effort required to evaluate $\hat{p}_\Delta^K$. These observations suggest that we can control for the computational complexity of our density estimator by optimally choosing the sequence $(J_\xi)_{\xi\geq 0}$ of Euler steps and the distribution $(q_\xi)_{\xi\geq 0}$ of Ξ. We follow Glynn & Rhee (2015) and set

$$J_\xi = O(2^\xi) \quad \text{and} \quad q_\xi = O\left(2^{-\xi}\xi \log_2^2(1+\xi)\right).$$

These choices ensures that the computational complexity of our density estimator is minimal, as indicated in the Proposition below.

Proposition 5.2. *Suppose that Assumptions (C1)-(C3) of Proposition 5.1 are satisfied. Fix $J_\xi = O(2^\xi)$, $I_\xi = O(J_\xi^{1/2})$, and $H_\xi = O(J_\xi^{1+d/4})$ for $\xi \in \mathbb{N}_0$ and some $\rho > 1$. In addition, set $q_\xi = O(2^{-\xi}\xi \log_2^2(1+\xi))$ for $\xi \in \mathbb{N}_0$. Then, $\hat{p}_\Delta^K(v,w;\theta)$ is an unbiased estimator of $p_\Delta(v,w;\theta)$ for any $v,w \in \mathcal{S}$ and $\theta \in \Theta$, and the root-mean squared error of the density estimator $\hat{p}_\Delta^K$ decays at rate $1/2$; i.e.,*

$$\left\|\hat{p}_\Delta^K(v,w;\theta) - p_\Delta(v,w;\theta)\right\|_2 = O\left(K^{-0.5}\right).$$

Furthermore, for any $\epsilon, \delta > 0$, the computational effort necessary to evaluate the density estimator $\hat{p}_\Delta^K$ so that the error bound ϵ is not violated with probability $1-\delta$ is at least of order $O\left(\epsilon^{-(3+d/2)}\log_2(1/\epsilon)\right)$. That is,

$$\mathbb{Q}_\theta\left[\left\|\hat{p}_\Delta^K(v,w;\theta) - p_\Delta(v,w;\theta)\right\|_2 \leq \epsilon\right] \geq 1-\delta \quad \Rightarrow \quad \frac{\epsilon^{-(3+d/2)}\log_2(1/\epsilon)}{\text{Effort}\left(\hat{p}_\Delta^K(v,w;\theta)\right)} = O(1).$$

This is the slowest rate of divergence of Effort$(\hat{p}_\Delta^K(v,w;\theta))$, the computational effort necessary to evaluate the Monte Carlo estimator $\hat{p}_\Delta^K$, as $K \to \infty$.

Proposition 5.2 states that the computational effort necessary to evaluate the density estimator $\hat{p}_\Delta^K$ with a maximum error of ϵ increases faster than cubicly in ϵ. In other words, the effort necessary to evaluate our density estimator grows faster than we would expect from the standard Monte Carlo theory. Furthermore, the rate at which the computational

complexity of $\hat{p}_\Delta^K$ grows increases with the dimensionality of the process X. These properties arise because $J_\Xi = O(2^\Xi)$ may become excessively large when Ξ is large, which occurs with high probability when the number of Monte Carlo samples K is large. In addition, a large number H_Ξ of Monte Carlo samples are necessary when the dimension d is large in order to control for the variance of the diffusion density estimator $\tilde{\mathcal{P}}^{H_\Xi, I_\Xi}$.

In spite of the computational costs when K is large, the Monte Carlo estimator $\hat{p}_\Delta^K$ has several features that make it appealing from a computational perspective. We describe these features below.

5.2.1 Maximum accuracy

We can control for the accuracy of the Monte Carlo estimator $\hat{p}_\Delta^K$ by controlling for the variance of $\hat{p}_\Delta$. We have one degree of freedom to control for the variance of $\hat{p}_\Delta$, namely, the choice of the Poisson rate $\ell > 0$. Small values of ℓ increase the variance of the Monte Carlo estimator $\hat{p}_\Delta(v, w; \theta)$ in its tails because the jump-diffusion density p_Δ is approximated by a Gaussian density when $\ell \approx 0$. On the other hand, large values of ℓ increase the bias in (15), therefore increasing the overall variance of our density estimator.

We fix $\ell > 0$ as to minimize the variance of the density estimator $\hat{p}_\Delta$ across the parameter and state spaces. That is, we fix

$$\ell^* = \arg\min_{\ell > 0} \max_{\theta \in \Theta} \max_{v, w \in \mathcal{S}} \mathrm{Var}_\theta^{\mathbb{Q}}\big(\hat{p}_\Delta(v, w; \theta)\big). \tag{17}$$

Such a choice for ℓ ensures that our density estimator has the smallest possible variance globally across the parameter and state spaces. This yields the most accurate Monte Carlo estimator $\hat{p}_\Delta^K$, uniformly across the parameter and state spaces.

The optimization problem (17) can be solved using a standard numerical optimization routine, such as the Nelder-Mead algorithm. It needs to be solved only once for a given jump-diffusion of the type (1) and a given $\Delta > 0$. The optimal Poisson rate ℓ^* can be reused to compute the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ for any $v, w \in \mathcal{S}$ and $\theta \in \Theta$. An unbiased estimator of the variance $\mathrm{Var}_\theta^{\mathbb{Q}}(\hat{p}_\Delta(v, w; \theta))$ can be easily constructed using independent samples of the density estimator $\hat{p}_\Delta$.

5.2.2 Multilevel Monte Carlo

In order to construct a sample of $\hat{p}_\Delta$ for a given sample of Ξ, we need to generate the Euler samples $(\mathcal{X}^j, \mathcal{Z}^j)$ based on $j = O(2^\Xi)$ and $j = O(2^{\Xi-1})$ steps. In other words, we need to run two Euler discretizations, one of which uses a fraction of the Euler steps of the other. To accomplish this task, it suffices if we sample Brownian increments for the fine Euler discretization with $O(2^\Xi)$ Euler steps, and then add up consecutive Brownian increments to obtain the increments for the coarser discretization with a fraction of Euler steps. As a result, we only need to sample once to obtain Euler discretizations with two different numbers of Euler steps. The idea of reusing Brownian increments for Euler discretizations with different numbers of Euler steps is inspired by the Multi-Level Monte Carlo method of Giles (2008). It yields important computational advantages, which we highlight in a numerical case study in Section 7.

5.3 Implementation

The evaluation of the Monte Carlo estimator $\hat{p}^K_\Delta(v, w; \theta)$ requires that we generate K independent samples of the random element $\mathsf{R} = (\Xi, \mathsf{P}, \mathsf{T}, \mathsf{D}, \mathsf{W}, \mathsf{U}, \mathsf{V})$, which contains:

- $\Xi \sim (q_\xi)_{\xi \geq 0}$, where $(q_\xi)_{\xi \geq 0}$ is fixed as in Proposition 5.2,
- $\mathsf{P} \sim \text{Poisson}(\ell\Delta)$, which is a sample of the jump count N_Δ under $\mathbb{Q}_\theta$,
- $\mathsf{T} = (T_n)_{n=1,\ldots,\mathsf{P}}$, which is a sample of the jump times $(T_n)_{n \leq N_\Delta}$ under $\mathbb{Q}_\theta$ conditional on $N_\Delta = \mathsf{P}$,
- Independent jump mark samples $\mathsf{D} = (\mathsf{D}_n)_{n=1,\ldots,\mathsf{P}}$ from the density π,
- Independent samples $\mathsf{W} = (\mathsf{W}_{n,j})_{n=0,\ldots,\mathsf{P},\, j=1,\ldots,J_\Xi}$ from the d-dimensional standard normal distribution with $J_\Xi = O(2^\Xi)$, and
- Independent samples $\mathsf{U} = (\mathsf{U}_{n,i,\nu})_{n=0,\ldots,\mathsf{P},\, i=1,\ldots,I, \nu=1,\ldots,H}$ from the d-dimensional standard normal distribution with $I = O(J_{\Xi-1}^{1/2})$ and $H = O(J_{\Xi-1}^{1+d/4})$.
- Independent samples $\mathsf{V} = (\mathsf{V}_{n,i,\nu})_{n=0,\ldots,\mathsf{P},\, i=1,\ldots,I, \nu=1,\ldots,H}$ from the d-dimensional standard normal distribution with $I = O(J_{\Xi}^{1/2})$ and $H = O(J_{\Xi}^{1+d/4})$.

The sampling of these random variables is standard; see, e.g., Glasserman (2003). The following Algorithm describes the computation of the Monte Carlo estimator $\hat{p}_\Delta^K$.

Algorithm 5.3 (Sampling of $\hat{p}_\Delta(v,w;\theta)$)**.** *Let $v, w \in \mathcal{S}$, $\theta \in \Theta$, the Poisson rate $\ell > 0$, the exponent $\rho > 1$, and i.i.d. samples $\mathsf{R}^k = (\Xi^k, \mathsf{P}^k, \mathsf{T}^k, \mathsf{D}^k, \mathsf{W}^k, \mathsf{U}^k, \mathsf{V}^k)$ for $k = 1, \dots, K$ be given. Initialize $\hat{p}^K = 0$. For $k = 1, \dots, K$, do:*

(1) Construct samples of $(\mathcal{X}^j, \mathcal{Z}^j)$ with $j = O(J_{\Xi^k})$ and $j = O(J_{\Xi^k - 1})$ Euler steps between consecutive jumps. Use the Euler increments W^k, the jump times T^k, and the jump marks D^k, and assume $X_0 = v$.

(2) Set $H = O(J_\Xi^{1+d/4})$ and $I = O(J_{\Xi^k}^{1/2})$. For $\nu = 1, \dots, H$, set $\tilde{X}_0^{I,\nu} = \mathcal{X}^j$ for $j = O(J_{\Xi^k})$, and construct the Euler discretization $(\tilde{X}_i^{I,\nu})_{i=1,\dots,I}$ of $\tilde{X}$ with I Euler steps in $[0, \Delta - \mathsf{T}^k_{\mathsf{P}^k}]$ by using the Euler increments V^k. Evaluate the density estimator $\tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)$ in (14).

(3) Set $\hat{p}^{(1)} = \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)\, \mathcal{Z}^j$ for $j = O(J_{\Xi^k})$.

(4) Set $H = O(J_{\Xi^k - 1}^{1+d/4})$ and $I = O(J_{\Xi^k - 1}^{1/2})$. For $\nu = 1, \dots, H$, set $\tilde{X}_0^{I,\nu} = \mathcal{X}^j$ for $j = O(J_{\Xi^k - 1})$, and construct the Euler discretization $(\tilde{X}_i^{I,\nu})_{i=1,\dots,I}$ of $\tilde{X}$ with I Euler steps in $[0, \Delta - \mathsf{T}^k_{\mathsf{P}^k}]$ by using the Euler increments U^k. Evaluate $\tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)$ as in (14).

(5) Set $\hat{p}^{(2)} = \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)\, \mathcal{Z}^j$ for $j = J_{\Xi^k - 1}$.

(6) Update $\hat{p}^K$ as

$$\hat{p}^K + \frac{1}{K} \frac{\hat{p}^{(1)} - \hat{p}^{(2)}}{q_{\Xi^k}}.$$

Return $\hat{p}^K$, which is an unbiased sample of $\hat{p}_\Delta^K(v, w; \theta)$.

The evaluation of our density estimator via Algorithm 5.3 is very simple. Steps (1), (2), and (3) require straightforward Euler discretization; Algorithms A.1 and A.2 in Appendix A provide guidance. Steps (2), (5) and (6) involve basic algebraic operations. We have implemented Algorithm 5.3 in R. The codes are available upon request.

Algorithm 5.3 highlights an important feature of the Monte Carlo estimator $\hat{p}_\Delta^K$: It can be computed as an analytical function of samples of the random element $\mathsf{R} = (\Xi, \mathsf{P}, \mathsf{T}, \mathsf{D}, \mathsf{W}, \mathsf{U}, \mathsf{V})$. Samples of R are independent of the parameter θ and and the pair (v, w) at which the density estimator is evaluated. Because of this, it suffices that we generate all samples of R once, and re-use these samples to evaluate $\hat{p}_\Delta^K(v, w; \theta)$ at any $\theta \in \Theta$ and $v, w \in \mathcal{S}$. This feature generates important computational advantages when using the Monte Carlo estimator $\hat{p}_\Delta^K$ for the statistical estimation of model (1).

6 Parameter inference

We derive parameter estimators based on our density estimator $\hat{p}_\Delta$, and analyze their asymptotic properties. Let $\theta^* \in \operatorname{int}\Theta$ be the true data-generating parameter. Define the simulated counterpart of the likelihood (3) as

$$\hat{\mathfrak{L}}_m^K(\theta) = \prod_{i=1}^{m} \hat{p}_\Delta^K(X_{(i-1)\Delta}, X_{i\Delta}; \theta). \tag{18}$$

A simulated maximum likelihood estimator (SMLE) $\hat{\theta}_m^K$ is an almost sure maximizer of the simulated likelihood (18). That is,

$$\hat{\theta}_m^K \in \arg\max_{\theta\in\Theta} \hat{\mathfrak{L}}_m^K(\theta). \tag{19}$$

Because the density estimator $\hat{p}_\Delta^K(\theta)$ is unbiased with finite variance, the asymptotic properties of the SMLE $\hat{\theta}_m^K$ are well understood. Giesecke & Schwenkler (2014) provide sufficient conditions that ensure that:

- A SMLE is asymptotically unbiased. That is,

  $$\hat{\theta}_m^K \to \hat{\theta}_m$$

  almost surely as $K \to \infty$.

- A SMLE is consistent, and

  $$\hat{\theta}_m^K \to \theta^*$$

  in $\mathbb{P}_{\theta^*}$-probability as $m \to \infty$ and $K \to \infty$.

- A SMLE is asymptotically normal and asymptotically efficient. More precisely,

$$\sqrt{m}(\hat{\theta}_m^K - \theta^*) \to N\left(0, \Sigma_{\theta^*}^{-1}\right)$$

if $\frac{m}{K} \to c \in [0, \infty)$ as $m \to \infty$ and $K \to \infty$, where

$$\Sigma_{\theta^*} = -\lim_{m\to\infty} \nabla^2 \log \mathfrak{L}_m(\theta^*).$$

is the Fisher information matrix.

The conditions of Giesecke & Schwenkler (2014) can be easily verified using our density estimator $\hat{p}_\Delta$.

As discussed in Section 5.3, we can separate the simulation steps from the estimation steps when evaluating the Monte Carlo density estimator $\hat{p}_\Delta^K(v, w; \theta)$. This fact generates significant computational advantages when carrying out parameter inference based on our density estimator. This is because the simulated likelihood $\hat{\mathfrak{L}}_m^K(\theta)$ becomes a deterministic function of the parameter θ and the data $\mathbf{X}_m$ once the samples of the random element R needed to evaluate $\hat{p}_\Delta^K$ have been generated. We can therefore employ standard numerical routines, such as the Nelder-Mead method, to solve the optimization problem (19).

7 Numerical results

This section illustrates the behavior of our density estimator and of simulated maximum likelihood estimators in a numerical case study. We consider a bivariate model from the affine class defined in Duffie et al. (2000). We specify the jump-diffusion X by choosing the following functions for $\theta = (a, b, k, \bar{X}, c, v, \ell_0, m, s, e) \in \mathbb{R}^2 \times \mathbb{R}_+^2 \times [-1, 1] \times \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+^2$, $X = (X_1, X_2) \in \mathcal{S} = \mathbb{R}^2$, and $D = (D_1, D_2) \in \mathcal{D} = \mathbb{R} \times \mathbb{R}_+$:

$$\mu(X; \theta) = \begin{pmatrix} a - bX_2 \\ k(\bar{X} - X_2) \end{pmatrix}, \qquad \Sigma(X; \theta) = \sqrt{X_2} \begin{pmatrix} 1 & 0 \\ cv & \sqrt{(1 - c^2)}v \end{pmatrix}$$

$$\Gamma(X, D; \theta) = \begin{pmatrix} m + sD_1 \\ -e \log(D_2) \end{pmatrix}, \qquad \Lambda(X; \theta) = \ell_0.$$

The SDE (1) in this case can be rewritten as

$$\mathrm{d}\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} a - bX_{2,t-} \\ k(\bar{X} - X_{2,t-}) \end{pmatrix} \mathrm{d}t + \sqrt{X_{2,t-}} \begin{pmatrix} 1 & 0 \\ cv & \sqrt{(1-c^2)}v \end{pmatrix} \mathrm{d}W_t + \mathrm{d}L_t, \quad (20)$$

where $L_t = \sum_{n=1}^{N_t} \Gamma(X_{T_n-}, D_n; \theta)$ and N is a counting process with intensity ℓ_0. The marks $(D_{1,n})_{n\geq 1}$ are i.i.d. samples of standard normal random variable, and $(D_{2,n})_{n\geq 1}$ are i.i.d. samples of a standard uniform random variable. We fix the parameter space $\Theta = [-0.3, 0.3] \times [-0.5, 0.5] \times [0.0001, 0.5] \times [0.0001, 0.5] \times [-1, 1] \times [0.0001, 0.5] \times [0.0001, 20] \times [-0.3, 0.3] \times [0.0001, 0.4] \times [0.0001, 0.3]$. The true data-generating parameter is $\theta^* = (0.1,$ 0, 0.1, 0.2, -0.5, 0.2, 6, -0.07, 0.1, 0.07), and $X_0 = (0, 0.1)$.

Because model (20) is affine as in Duffie et al. (2000), the characteristic function of X can be evaluated in terms of solutions of ordinary differential equations. The solutions to these ordinary differential equations are known in closed form given that the jump intensity of N is constant. As a result, the characteristic function of X is known in closed form. We can thus evaluate the true density p_Δ semi-analytically via Fourier inversion of the characteristic function. The density p_Δ derived via Fourier inversion serves as a benchmark against which we will evaluate our density estimator $\hat{p}_\Delta^K$, as well as other competing estimators. We implement Fourier inversion via numerical quadrature with 500 discretization points per dimension in $[-2000, 2000]^2$. The numerical results reported in this section are implemented in R, running on an 2×8-core 2.6 GHz Intel Xeon E5-2670, 128 GB server at Boston University with a Linus Centos 6.6 operating system. All codes used to generate the results of this section are available upon request.

The SDE (20) describes a stochastic volatility model with jumps that is commonly used in the options pricing literature; see, e.g., Andersen et al. (2002), Eraker et al. (2003), Eraker (2004). Jumps in returns are normally distributed, and jumps in volatility are exponentially distributed. Jumps in returns and volatility occur simultaneously. Brownian innovations in returns and volatility are correlated with correlation coefficient c. Because volatility is random and there are jumps, the distribution of returns is non-Gaussian. Furthermore, the distribution of volatility is asymmetric and skewed. Because of these special features, and because the true density p_Δ is known in semi-analytical form, Model (20) provides a good test case for evaluating the performance of our estimators.

7.1 Density estimator

We study the accuracy of our density estimator $\hat{p}_\Delta^K$. We fix $\Delta = 1/12$, which corresponds to a monthly time horizon. Figure 1 shows surface plots of the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ computed for $K = 1000$ and $K = 5000$. When K is small and only few Monte Carlo replications are used to evaluate our density estimator, the density estimator assigns probability mass to areas in which the true density has no mass. These spikes vanish as the number of Monte Carlo replications grows. Figure 2 shows a contour plot of the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ for $K = 5000$. Confirming the unbiasedness result of Theorem 4.1, the Monte Carlo estimator is centered around the same location as the true density.

Figures 3 and 4 plot the marginal densities of returns and volatility for $K \in \{1000, 2000, 5000\}$, together with 90% confidence bands computed from bootstrap with 1000 bootstrap samples. The marginal densities are computed via rectangular quadrature of the true density and the Monte Carlo density estimator using an equidistant grid on $[-0.5, 0.3] \times [0, 0.3]$ with 4779 grid points. It can be seen that the marginal densities derived from our estimator are close to the true marginal densities in the centers and in the tails of the distributions. Given that our density estimator has positive and finite variance, the marginal densities derived from $\hat{p}_\Delta^K$ fluctuate around the true marginal densities. However, the bandwidth of these fluctuations decreases as the number K of Monte Carlo samples grows. This confirms Proposition 5.1, which states that the mean squared error of the Monte Carlo estimator converges to zero as K grows large.

We also plot conditional densities of returns and volatility in Figures 5 and 6. Consistent with Proposition 5.1, these figures reveal that the accuracy of the Monte Carlo estimator increases as the number K of Monte Carlo samples rises. Further, these figures show that our density estimator is able to capture the asymmetric and non-Gaussian distribution of returns and volatility.

7.2 Computational complexity

We evaluate the computational complexity of our density estimator. For this, we randomly pick 500 points in $v, w \in [-0.5, 0.3] \times [0, 0.3]$, and evaluate the unbiased Monte

Carlo estimator $\hat{p}_\Delta^K$ at these 500 randomly selected points and the true parameter θ^*. We compute the root mean squared error (RMSE) of the unbiased Monte Carlo estimator across the 500 pairs (v, w), and track the time it takes to compute the density estimator for all 500 pairs (v, w). We carry out the same analysis for the Monte Carlo counterpart of the biased density estimator $\hat{p}_\Delta^{H,I,J}(v, w; \theta)$ in (11) based on K Monte Carlo samples. We adopt the square-root rule of Duffie & Glynn (1995) for the biased density estimator (11) and set $J = \sqrt{K}$. As for the unbiased density estimator, we fix $I = \sqrt{J}$ and $H = 1$ when computing the biased density estimator (11).

Figure 7 shows that our unbiased Monte Carlo estimator $\hat{p}_\Delta^K$ can achieve high accuracy even with small computational budgets. Because the Monte Carlo estimator $\hat{p}_\Delta^K$ is unbiased, it achieves smaller RMSE than the biased density estimator (11) for the computational budgets we consider. As a result, the Monte Carlo estimator $\hat{p}_\Delta^K$ is computationally efficient for the small to medium-sized computational budgets considered here. Nonetheless, Figure 7 also shows that the rate of convergence of our density estimator is not linear in the log-log scale plot. This holds because the computational effort necessary to achieve a certain error bound grows nonlinearly and faster than quadratic as the error bound shrinks to zero, as highlighted by Proposition 5.2. Consequently, we expect that our density estimator will be less computationally efficient for large computational budgets.

7.3 Simulated likelihood estimators

We carry out a simulation analysis to evaluate our simulated likelihood estimators. We simulate 100 independent samples of the data $\mathbf{X}_m = \{X_{t_0}, X_{t_1}, \ldots, X_{t_m}\}$ from its true law $\mathbb{P}_{\theta^*}$ with $t_i - t_{i-1} = \Delta = 1/12$ for all $1 \leq i \leq m$ and $m = 720$. This corresponds to 60 years of monthly data. We use Broadie & Kaya (2006) to generate exact samples of $\mathbf{X}_m$. For each data sample, we compute SMLE $\hat{\theta}_m^K$ by maximizing the simulated likelihood $\mathfrak{L}_m^K(\theta)$ for $K = 20000$. We use a Nelder-Mead method to calculate maximizers of the simulated log-likelihood.

Table 1 shows the average of the computed SMLE across all 100 data samples, as well as their empirical standard deviation. Our SMLE are able to precisely identify the true parameters. Almost all data-generating parameters are contained in a two empirical

standard deviation band around the average SMLE.

Next, we analyze the asymptotic distribution of our SMLE. For this, we compute the first four centered moments of the scaled error $\sqrt{m}(\hat{\theta}_m^K - \theta^*)$ of our SMLE empirically across the 100 samples of the data $\mathbf{X}_m$. We compare the first four empirical moments to the theoretical moments implied by the asymptotic distribution of true maximum likelihood estimators; see Theorem 6.2 of Giesecke & Schwenkler (2014). Table 2 indicates that the moments of the scaled error of our SMLE are similar to the theoretical moments implied by the asymptotic distribution of maximum likelihood estimators for most parameters. As a result, $\hat{\theta}_m^K$ has a similar distribution as true maximum likelihood estimators. In accordance with the discussion in Section 6, these numerical results suggest that the simulated likelihood estimator $\hat{\theta}_m^K$ that we propose in this paper indeed behaves similarly to true maximum likelihood estimators, and that it inherits their consistency, asymptotic normality, and asymptotic efficiency properties.

A Algorithms

Algorithm A.1 (Simultaneous sampling of $(\mathcal{X}^j, \mathcal{Z}^j)$ via Euler discretization with $j = O(J_\Xi)$ and $j = O(J_{\Xi-1})$ Euler steps between consecutive jumps for given P, T, and W)**.** *Set* $X^1_{0,0} = X^2_{0,0} = v$, $Z^1_{0,0} = Z^2_{0,0} = 1$, $\mathsf{T}_0 = 0$, $\mathsf{T}_{\mathsf{P}+1} = \Delta$, *and* $J = O(2^{\Xi-1})$. *For* $n = 0, \ldots, \mathsf{P}$, *do:*

(1) Set $h_n = \frac{\mathsf{T}_{n+1} - \mathsf{T}_n}{J}$.

(2) For $j = 1, \ldots, J$, *set:*

$$X^1_{n,j} = X^1_{n,j-1} + h_n \mu\left(X^1_{n,j-1}; \theta\right) + \sqrt{h_n}\Sigma\left(X^1_{n,j-1}; \theta\right)\left(\mathsf{W}_{n,2j-1} + \mathsf{W}_{n,2j}\right)$$

$$X^2_{n,2j-1} = X^2_{n,2(j-1)} + \frac{h_n}{2}\mu\left(X^2_{n,2(j-1)}; \theta\right) + \sqrt{\frac{h_n}{2}}\Sigma\left(X^2_{n,2(j-1)}; \theta\right)\mathsf{W}_{n,2j-1}$$

$$X^2_{n,2j} = X^2_{n,2j-1} + \frac{h_n}{2}\mu\left(X^2_{n,2j-1}; \theta\right) + \sqrt{\frac{h_n}{2}}\Sigma\left(X^2_{n,2j-1}; \theta\right)\mathsf{W}_{n,2j}$$

$$Z^1_{n,j} = Z^1_{n,j-1} - h_n Z^1_{n,j-1}\left(\Lambda\left(X^1_{n,j-1}; \theta\right) - \ell\right),$$

$$Z^2_{n,2j-1} = Z^2_{n,2(j-1)} - \frac{h_n}{2} Z^2_{n,2(j-1)}\left(\Lambda\left(X^2_{n,2(j-1)}; \theta\right) - \ell\right),$$

$$Z^2_{n,2j} = Z^2_{n,2j-1} - \frac{h_n}{2} Z^2_{n,2j-1} \left(\Lambda \left(X^2_{n,2j-1}; \theta \right) - \ell \right).$$

(3) If $n < \mathsf{P}$*, set*

$$\begin{aligned}
X^1_{n+1,0} &= X^1_{n,J} + \Gamma \left(X^1_{n,J}, \mathsf{D}_n; \theta \right), \\
X^2_{n+1,0} &= X^2_{n,2J} + \Gamma \left(X^2_{n,2J}, \mathsf{D}_n; \theta \right), \\
Z^1_{n+1,0} &= \frac{\Lambda \left(X^1_{n,J}; \theta \right)}{\ell} Z^1_{n,J}, \\
Z^2_{n+1,0} &= \frac{\Lambda \left(X^2_{n,2J}; \theta \right)}{\ell} Z^2_{n,2J}.
\end{aligned}$$

Return $(X^1_{\mathsf{P},0}, Z^1_{\mathsf{P},J})$*, which is a sample of* $(\mathcal{X}^j, \mathcal{Z}^j)$ *with* $j = O(J_{\Xi-1})$*, and* $(X^2_{\mathsf{P},0}, Z^2_{\mathsf{P},2J})$*, which is a sample of* $(\mathcal{X}^j, \mathcal{Z}^j)$ *with* $j = O(J_\Xi)$*.*

Algorithm A.2 (Sampling of the density $\tilde{\mathcal{P}}^{1,I}(v, w; \theta)$ via Euler discretization for $I = O(J_\Xi^{1/2})$ and given Δ, P, T, and V)**.** *Initialize* $\tilde{X}_0 = v$*.*

(1) Fix $h = \frac{\Delta - \mathsf{T}_\mathsf{P}}{I}$*.*

(2) For $i = 1, \ldots, I-1$*, set*

$$\tilde{X}_i = \tilde{X}_{i-1} + h\mu \left(\tilde{X}_{i-1}; \theta \right) + \sqrt{h} \Sigma \left(\tilde{X}_{i-1}; \theta \right) \mathsf{V}_i.$$

(3) Return

$$\tilde{\mathcal{P}}^{1,I}(v, w; \theta) = \phi \left(\frac{1}{\sqrt{h}} \Sigma^{-1} \left(\tilde{X}_{I-1}; \theta \right)^\top \left(w - \tilde{X}_{I-1} - \mu \left(\tilde{X}_{I-1}; \theta \right) h \right) \right).$$

References

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

Aït-Sahalia, Yacine (2008), 'Closed-form likelihood expansions for multivariate diffusions', *The Annals of Statistics* **36**(2), 906–937.

Andersen, Torben G., Luca Benzoni & Jesper Lund (2002), 'An empirical investigation of continuous-time equity return models', *The Journal of Finance* **57**(3), 1239–1284. **URL:** *http://dx.doi.org/10.1111/1540-6261.00460*

Bally, Vlad & Denis Talay (1996), 'The law of the euler scheme for stochastic differential equations: Ii. convergence rate of the density', *Monte Carlo Methods and Applications* **2**, 93–128.

Bibby, Bo Martin & Michael Sørensen (1995), 'Martingale estimation functions for discretely observed diffusion processes', *Bernoulli* **1**(1-2), 17–39.

Blanchet, J. & J. Ruf (2013), A Weak Convergence Criterion Constructing Changes of Measure. Working Paper.

Brémaud, Pierre (1980), *Point Processes and Queues – Martingale Dynamics*, Springer-Verlag, New York.

Broadie, Mark & Ozgur Kaya (2006), 'Exact simulation of stochastic volatility and other affine jump diffusion processes', *Operations Research* **54**(2), 217–231.

Carrasco, Marine, Mikhail Chernov, Jean-Pierre Florens & Eric Ghysels (2007), 'Efficient estimation of general dynamic models with a continuum of moment conditions', *Journal of Econometrics* **140**(2), 529–573.

Cass, Thomas (2009), 'Smooth densities for solutions to stochastic differential equations with jumps', *Stochastic Processes and their Applications* **119**(5), 1416 – 1435.

Chacko, George & Luis M. Viceira (2003), 'Spectral gmm estimation of continuous-time processes', *Journal of Econometrics* **116**(1-2), 259–292.

Dacunha-Castelle, D. & D. Florens-Zmirou (1986), 'Estimation of the coefficients of a diffusion from discrete observations', *Stochastics* **19**(4), 263–284.

Detemple, Jérôme, René Garcia & Marcel Rindisbacher (2006), 'Asymptotic properties of monte carlo estimators of diffusion processes', *Journal of Econometrics* **134**(1), 1–68.

Duffie, Darrell, Jun Pan & Kenneth Singleton (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**, 1343–1376.

Duffie, Darrell & Kenneth J. Singleton (1993), 'Simulated moments estimation of markov models of asset prices', *Econometrica* **61**(4), 929–952.

Duffie, Darrell & Peter Glynn (1995), 'Efficient Monte Carlo estimation of security prices', *Annals of Applied Probability* **4**(5), 897–905.

Duffie, Darrell & Peter Glynn (2004), 'Estimation of continuous-time markov processes sampled at random time intervals', *Econometrica* **72**(6), 1773–1808.

Eraker, Bjørn (2004), 'Do stock prices and volatility jump? reconciling evidence from spot and option prices', *Journal of Finance* **59**(3), 1367–1404.

Eraker, Bjørn, Michael Johannes & Nicholas Polson (2003), 'The impact of jumps in volatility and returns', *Journal of Finance* **58**(3), 1269–1300.

Filipović, Damir, Eberhard Mayerhofer & Paul Schneider (2013), 'Density approximations for multivariate affine jump-diffusion processes', *Journal of Econometrics* . In Press.

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

Giesecke, Kay & Dmitry Smelov (2013), 'Exact sampling of jump-diffusions', *Operations Research* . In Press.

Giesecke, Kay & Gustavo Schwenkler (2014), Simulated Likelihood for Discretely Observed Jump-Diffusions. Working Paper.

Giles, Michael B. (2008), 'Multilevel monte carlo path simulation', *Operations Research* **56**(3), 607–617.

Glasserman, Paul (2003), *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York.

Glynn, Peter W. & Chang-han Rhee (2015), 'Unbiased Estimation with Square Root Convergence for SDE Models', *Operations Research* . forthcoming.

Gobet, Emmanuel & Céline Labart (2008), 'Sharp estimates for the convergence of the density of the euler scheme in small time', *Electronic Communications in Probability* **13**(35), 352–363.

Gobet, Emmanuel, Marc Hoffmann & Markus Reiß (2004), 'Nonparametric estimation of scalar diffusions based on low frequency data', *The Annals of Statistics* **32**(5), 2223–2253.

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

Guyon, Julien (2006), 'Euler scheme and tempered distributions', *Stochastic Processes and their Applications* **116**(6), 877 – 904.

Henry-Labordère, Pierre, Xiaolu Tan & Nizar Touzi (2015), Exact simulation of multi-dimensional stochastic differential equations. Working Paper.

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

Higham, Desmond J., Xuerong Mao & Andrew M. Stuart (2003), 'Strong convergence of euler-type methods for nonlinear stochastic differential equations', *SIAM Journal on Numerical Analysis* **40**(3), 1041–1063.

Jacod, Jean & Philip Protter (1998), 'Asymptotic error distributions for the euler method for stochastic differential equations', *The Annals of Probability* **26**(1), 267–307.

Jiang, George J. & John L. Knight (2002), 'Estimation of continuous-time processes via the empirical characteristic function', *Journal of Business & Economic Statistics* **20**(2), 198–212.

Jiang, George J. & John L. Knight (2010), 'Ecf estimation of markov models where the transition density is unknown', *Econometrics Journal* **13**(2), 245–270.

Kloeden, Peter E. & Eckhard Platten (1999), *Numerical Solutions of Stochastic Differential Equations*, Stochastic Modeling and Applied Probability, Springer.

Komatsu, Takashi & Atsushi Takeuchi (2001), 'On the smoothness of pdf of solutions to sde of jump type', *International Journal of Differential Equations and Applications* **2**(2), 141–197.

Konakov, Valentin & Enno Mammen (2002), 'Edgeworth type expansions for euler schemes for stochastic differential equations', *Monte Carlo Methods and Applications* **8**(3), 271–285.

Kristensen, Dennis & Yongseok Shin (2012), 'Estimation of dynamic models with nonparametric simulated maximum likelihood', *Journal of Econometrics* **167**(1), 76 – 94.

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

Pedersen, Asger Roer (1995), 'Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes', *Bernoulli* **1**(3), 257–279.

Protter, Philip (2004), *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.

Rogers, L. C. G. (1985), 'Smooth transition densities for one-dimensional diffusions', *Bulletin of the London Mathematical Society* **17**(2), 157–161.

Smith, A. A., Jr (1993), 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics* **8**, S63–S84.

Takeuchi, Atsushi (2002), 'The malliavin calculus for sde with jumps and the partially hypoelliptic problem', *Osaka J. Math.* **39**(3), 523–559.

Yan, Liqing (2002), 'The euler scheme with irregular coefficients', *The Annals of Probability* **30**(3), 1172–1194.

Yu, Jialin (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.

True parameter θ^*		SMLE $\hat{\theta}_m^K$	
		M	SD
a	0.1	0.1295	0.0674
b	0	0.0008	0.0723
k	0.1	0.1058	0.0220
$\bar{X}$	0.2	0.1928	0.0860
c	-0.5	-0.5651	0.0390
v	0.2	0.1771	0.0096
ℓ_0	6	6.2828	0.4304
m	-0.07	-0.0570	0.0261
s	0.1	0.0869	0.0352
e	0.07	0.0688	0.0097

Table 1: *Simulated likelihood estimators.* This table shows the average SMLE $\hat{\theta}_m^K$ estimated from 100 independent samples of the data $\mathbf{X}_m$ with $m = 720$ (Column "M"). It also displays the standard deviation of $\hat{\theta}_m^K$ across all 100 data samples (Column "SD"). We use $K = 20000$ Monte Carlo replications to evaluate the density estimator.

	$\sqrt{m}(\hat{\theta}_m^K - \theta^*)$				Theoretical asymptotic distribution			
	M	V	S	K	M	V	S	K
a	0.7916	3.2708	0.2658	2.9229	0.0000	134.3693	0.0000	3.0000
b	0.0215	3.7636	−0.5064	3.9332	0.0000	10.0593	0.0000	3.0000
k	0.1556	0.3485	−0.0014	3.5392	0.0000	0.3548	0.0000	3.0000
$\bar{X}$	−0.1932	5.3251	−0.4290	3.3417	0.0000	272.3220	0.0000	3.0000
c	−1.7468	1.0951	0.2311	2.0267	0.0000	0.4393	0.0000	3.0000
v	−0.6145	0.0664	−0.0844	2.4919	0.0000	0.0173	0.0000	3.0000
ℓ_0	7.5883	133.3758	−1.1395	3.9449	0.0000	588.3955	0.0000	3.0000
m	0.3488	0.4905	−0.4843	2.5459	0.0000	0.6830	0.0000	3.0000
s	−0.3515	0.8921	0.0226	2.3683	0.0000	0.7190	0.0000	3.0000
e	−0.0322	0.0677	0.0424	1.5549	0.0000	0.0069	0.0000	3.0000

Table 2: *Asymptotic distribution.* This table shows the empirical mean ("M"), variance ("V"), skewness ("S"), and kurtosis ("K") of $\sqrt{m}(\hat{\theta}_m^K - \theta^*)$ estimated from 100 independent samples of the data $\mathbf{X}_m$ with $m = 720$ and $K = 20000$. It also shows the theoretical moments of the asymptotic distribution of maximum likelihood estimators. This distribution is normal with mean zero and variance-covariance matrix $\Sigma_{\theta^*}^{-1}$ for $\Sigma_{\theta^*} = -\lim_{m\to\infty} \frac{1}{m} \nabla^2 \log \mathfrak{L}(\theta^*)$ according to Theorem 6.2 of Giesecke & Schwenkler (2014).

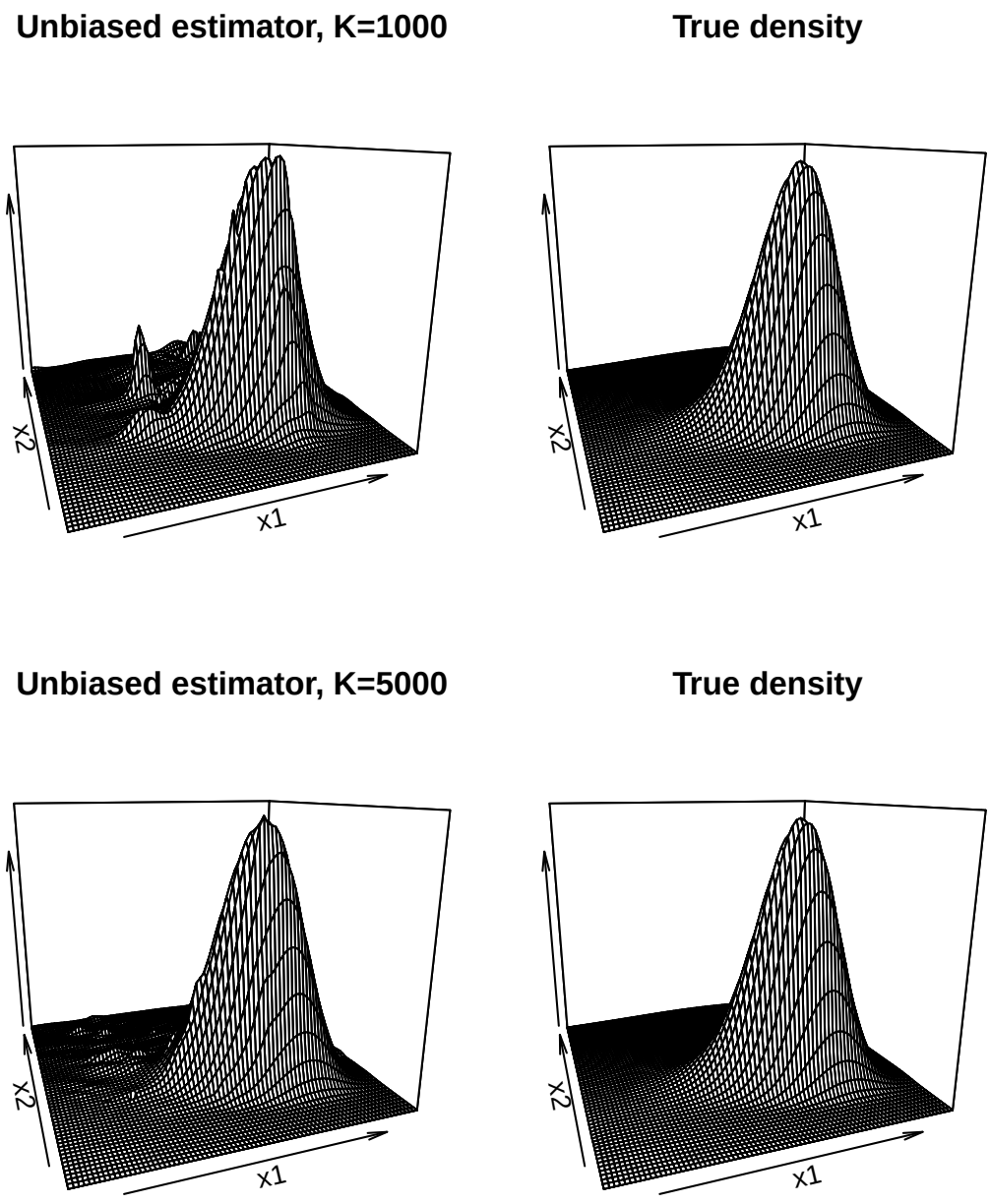

Figure 1: *Surface plots.* These figures show the surface plots of the true density $p_\Delta(v, w; \theta)$ and the unbiased Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ for $v = (0, 0.1)$, $w \in [-0.5, 0.3] \times [0, 0.3]$, and $K \in \{1000, 5000\}$.

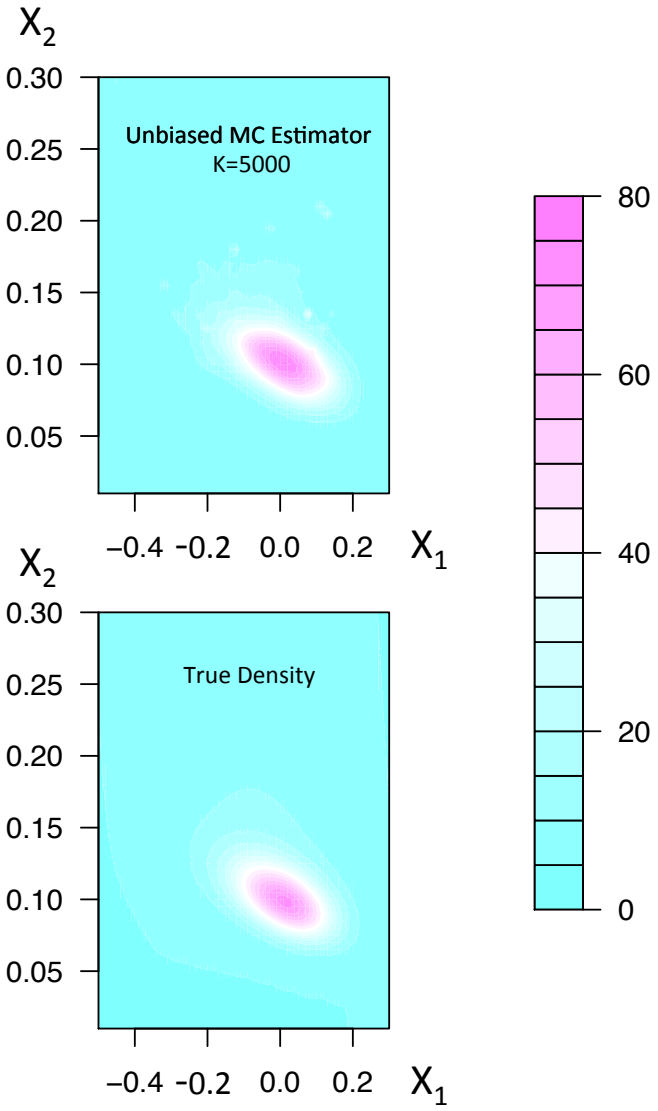

Figure 2: *Contour plots.* These figures show the contour plots of the true density $p_\Delta(v, w; \theta)$ and the unbiased Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ for $v = (0, 0.1)$, $w \in [-0.5, 0.3] \times [0, 0.3]$, and $K = 5000$.

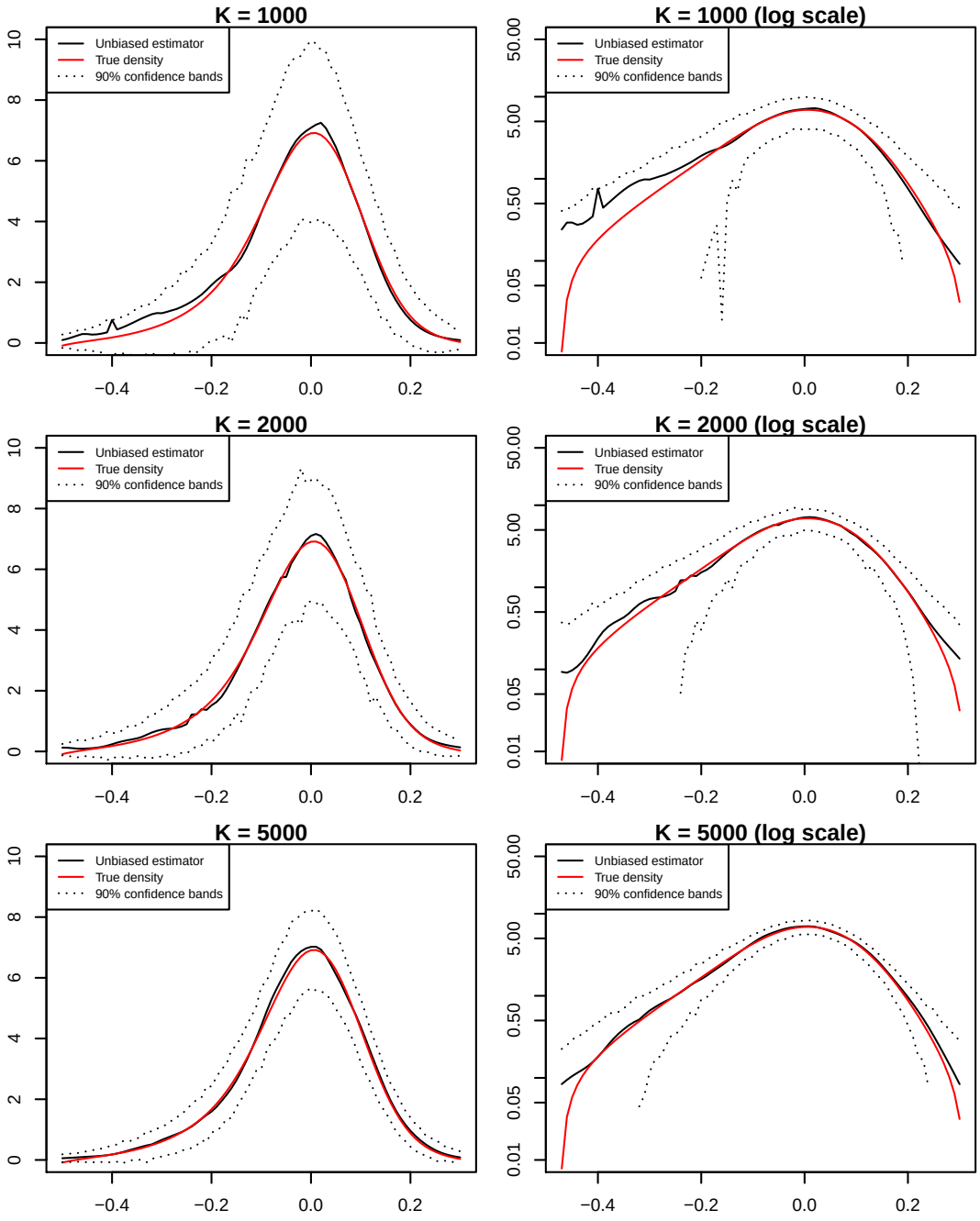

Figure 3: *Marginal density of returns*. These figures show the marginal density of returns computed via numerical quadrature along the X_2-axis given $X_0 = (0, 0.1)$. We take $K \in \{1000, 2000, 5000\}$ for the Monte Carlo estimator. Confidence bands are computed via bootstrap with 1000 bootstrap samples. The plots on the right-hand side are in log-scale.

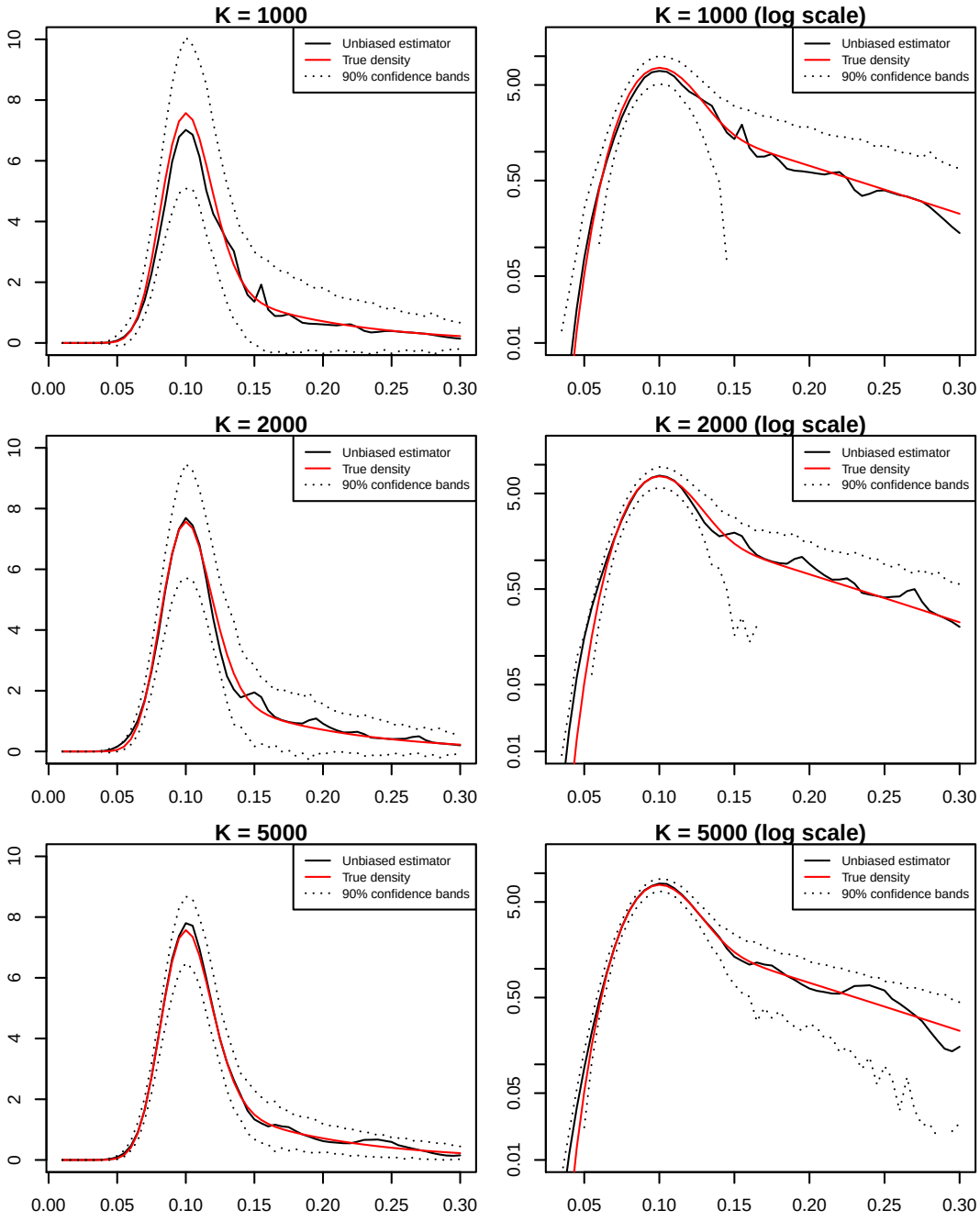

Figure 4: *Marginal density of volatility.* These figures show the marginal density of volatility computed via numerical quadrature along the X_1-axis given $X_0 = (0, 0.1)$. We take $K \in \{1000, 2000, 5000\}$ for the Monte Carlo estimator. Confidence bands are computed via bootstrap with 1000 bootstrap samples. The plots on the right side are in log-scale.

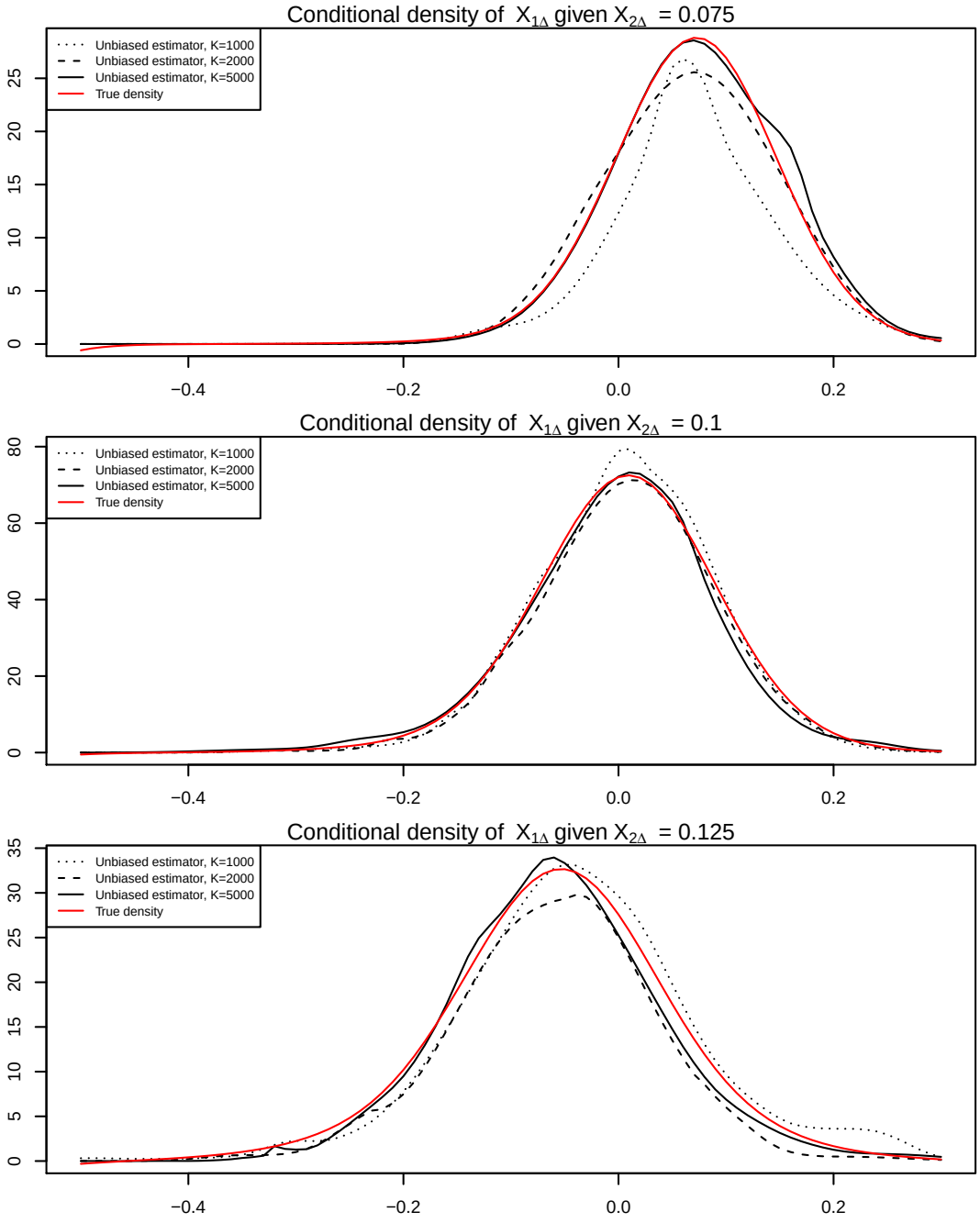

Figure 5: *Conditional density of returns.* These figures show the unnormalized conditional density of return $\hat{p}_\Delta(v, w; \theta)$ for $v = (0, 0.1)$ and $w \in [-0.5, 0.3] \times \{0.075, 0.1, 0.125\}$. We also plot samples of the corresponding Monte Carlo estimator $\hat{p}_\Delta(v, w; \theta)$ for $K \in \{1000, 2000, 5000\}$.

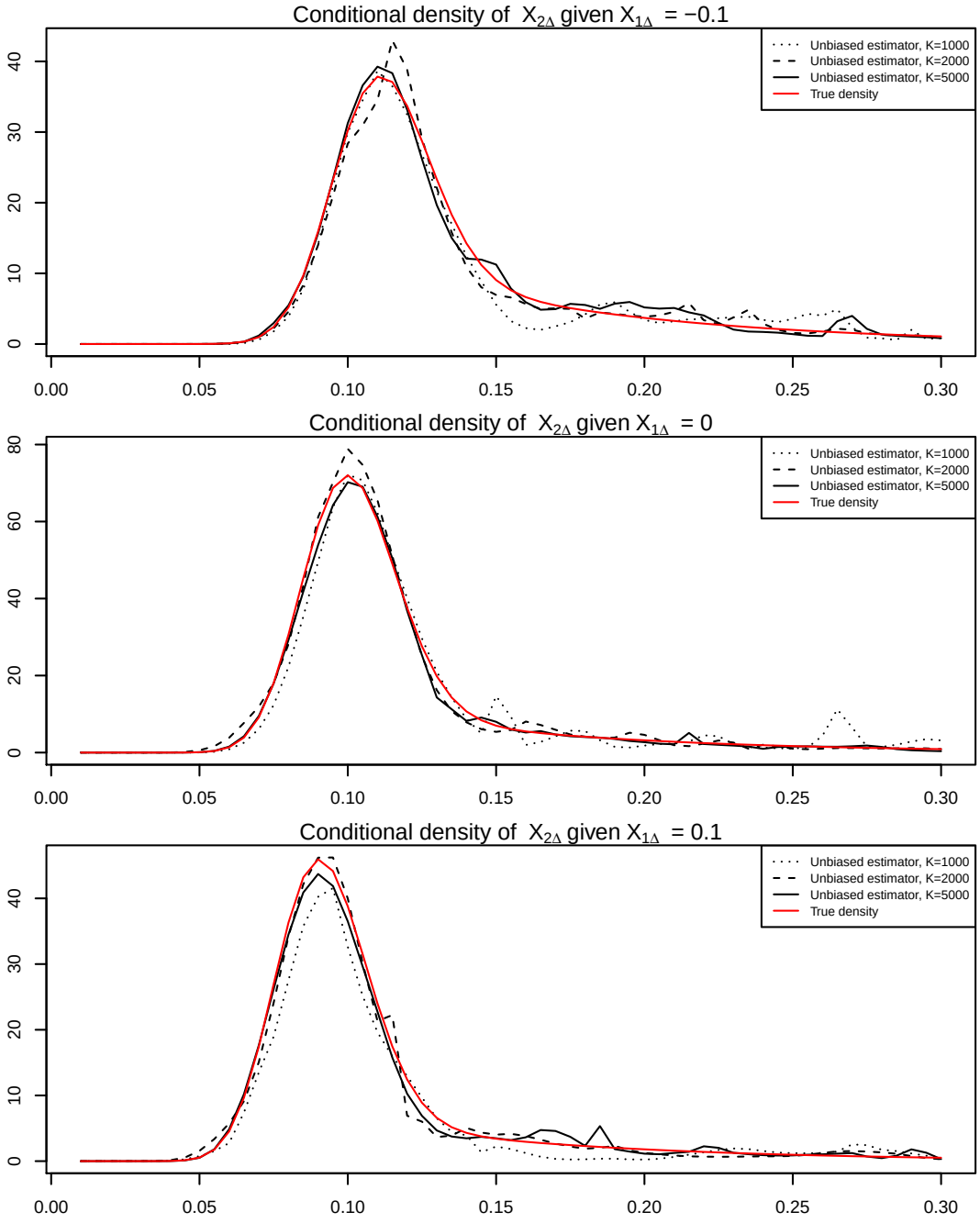

Figure 6: *Conditional density of volatility.* These figures show the unnormalized conditional density of volatility $\hat{p}_\Delta(v, w; \theta)$ for $v = (0, 0.1)$ and $w \in \{-0.1, 0, 0.1\} \times [0, 0.3]$. We also plot samples of the corresponding Monte Carlo estimator $\hat{p}_\Delta(v, w; \theta)$ for $K \in \{1000, 2000, 5000\}$.

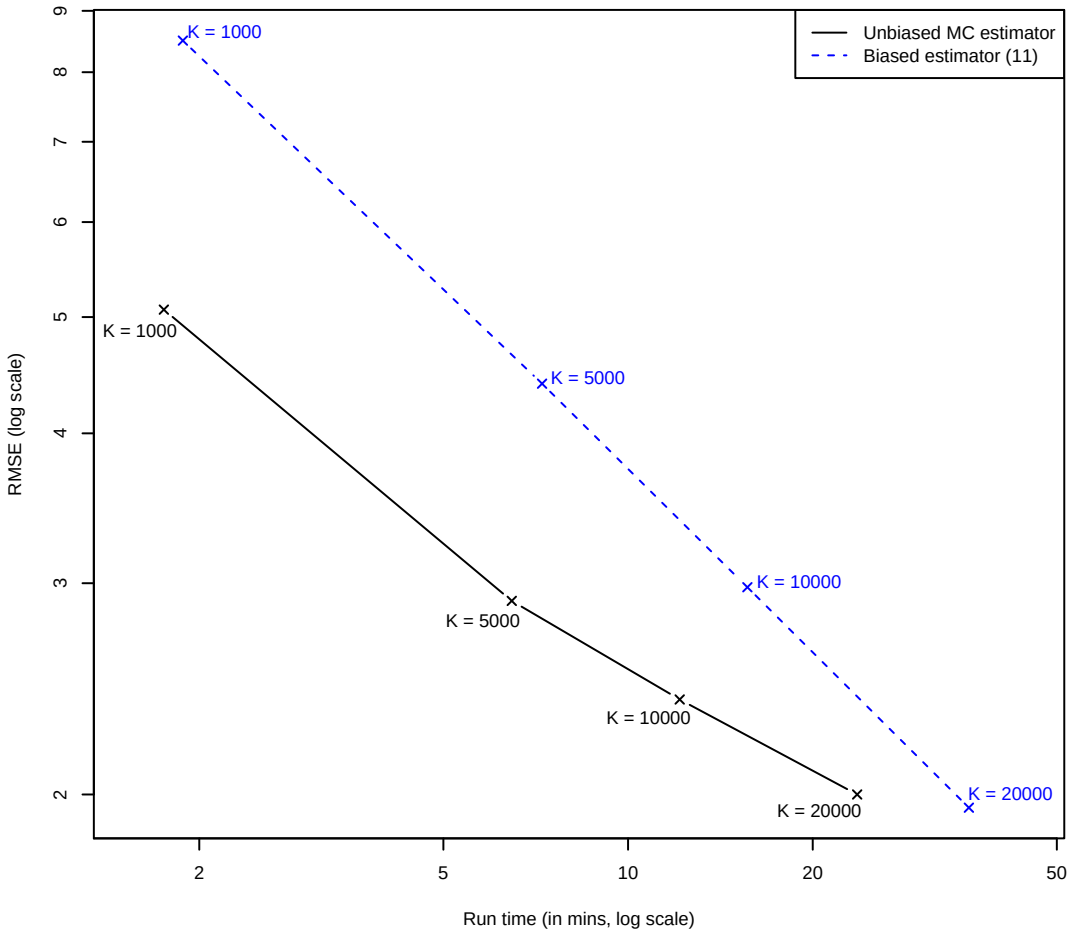

Figure 7: *Computational efficiency.* This figure plots the root mean squared error (RMSE) of the unbiased density estimator $\hat{p}_{\Delta}^{K}(v, w; \theta)$ and of the biased density estimator $\hat{p}_{\Delta}^{H,I,J}(v, w; \theta)$ in (11) against the time it takes to compute these estimators for 500 randomly chosen points $v, w \in [-0.5, 0.3] \times [0, 0.3]$. The density estimators are evaluated at the true parameter θ^*.