

Efficient Parameter Estimation for Multivariate Jump-Diffusions

François Guay and Gustavo Schwenkler*

November 16, 2017[†]

Abstract

This paper develops estimators of the transition density, filters, and parameters of multivariate jump-diffusion models. The drift, volatility, jump intensity, and jump magnitude are allowed to be state-dependent and non-affine. It is not necessary to diagonalize the volatility matrix. Our density and filter estimators achieve the canonical rate of convergence typically associated with exact Monte Carlo estimation. Our parameter estimators have the same asymptotic distribution as maximum likelihood estimators, which are often intractable for the class of models we consider. The results of this paper enable the empirical analysis of previously intractable models of asset prices and economic time series.

*Guay is at Cornerstone Research. Schwenkler is at Boston University. Schwenkler is the corresponding author: Boston University Questrom School of Business, Department of Finance, 595 Commonwealth Avenue, Boston, MA 02215 (gas@bu.edu).

[†]Guay acknowledges support from a Hariri Graduate Fellowship. We are grateful to Torben Andersen, Kay Giesecke, Peter Glynn, Jean Jacod, Eric Jacquier, Pierre Perron, Zhongjun Qu, Marcel Rindisbacher, Olivier Scaillet, Viktor Todorov, and the seminar participants at Boston University, Northwestern University, Stanford University, the 2015 AMS Sectional Mathematics Meetings at Loyola University, the 2016 Annual Conference of the Society for Financial Econometrics, the 2016 Conference on Financial Econometrics & Empirical Asset Pricing at Lancaster University, the 2016 Greenland Applied Probability Symposium, the 2016 Computational and Financial Econometrics Conference, the 2017 Vienna-Copenhagen Conference on Financial Econometrics, the 2017 Asian, Latinamerican, and North American Summer Meetings of the Econometrics Society, the 2017 NBER-NSF Time Series Conference, and the 2017 Trends in Econometrics conference for useful comments and suggestions, and Hang Jiang, Shenghao Wang, Qing Xu, and Hannan Zheng for excellent research assistance.

1 Introduction

Multivariate Markovian jump-diffusion processes are popular stochastic models often used in economic and financial studies. The main reason for their popularity is that they allow for flexible parametric specifications of day-to-day diffusive fluctuations and of rare unexpected jumps in asset prices, volatilities, and correlations. These features have been demonstrated to be empirically relevant in financial and economic data. In spite of their popularity, it is well-known that parameter inference for multivariate jump-diffusions based on discretely observed data is challenging because the underlying transition density is intractable outside of the class of affine models. Currently available density approximations apply primarily for high frequency observations in settings without latent variables. Often times, these approximations introduce biases that result in inexact or inefficient parameter estimators. The question of how to carry out exact and efficient parameter inference for multivariate jump-diffusions models, in particular in settings in which only low frequency data is available or latent variables are present, remains open.

In this paper, we derive Monte Carlo estimators of the transition density, posterior filters, and parameters of multivariate Markovian jump-diffusion models. To the best of our knowledge, our estimators are the first to satisfy all of the following key properties: *(i)* The estimators converge at the canonical square-root rate, *(ii)* they can be computed with finite computational resources, *(iii)* they apply for non-affine specifications, *(iv)* they can handle latent variables, and *(v)* they apply for arbitrary sampling frequencies. These properties enable exact and efficient filtering and parameter inference for a broad class of previously intractable jump-diffusion models. Our results close an important gap in the econometric literature.

We consider Markovian multivariate jump-diffusions for which the drift, volatility, jump intensity, and jump magnitude are smooth parametric functions of the state. We impose no restrictions on the coefficient functions other than they yield a well-defined stochastic process with a transition density. By taking advantage of Bayes' rule and a change of measure, we rewrite the transition density in terms of a mixture of transition densities of purely diffusive processes without jumps. Our density representation is similar

to the one of [Giesecke and Schwenkler \(2017b\)](#), who characterize the transition density of a reducible jump-diffusion process as a mixture of Gaussian densities.¹ In contrast to the representation of [Giesecke and Schwenkler \(2017b\)](#), our representation also applies to irreducible jump-diffusions. Our density representation provides a significant generalization of the well-known representations of [Dacunha-Castelle and Florens-Zmirou \(1986\)](#) and [Rogers \(1985\)](#), which apply only to univariate diffusive processes.

A benefit of our density representation is that it can be estimated via simulation. This is because it is given by an unconditional expectation of a path functional of the state. We exploit a novel randomization technique introduced by [Rhee and Glynn \(2015\)](#) to construct an unbiased estimator of the transition density. Our density estimator can be seen as a randomized multilevel Monte Carlo estimator.² It is constructed from samples derived from Euler’s discretization method with varying time steps, which are mixed and weighted adequately to ensure unbiasedness of the density estimator.³ The resulting density estimator is unbiased; its only source of error is variance that arises due to the simulation we carry out. The accuracy of our density estimator depends only on the number of Monte Carlo replications used.

We exploit our density estimator to carry out parameter inference and filtering. When the data is incomplete and only some components of the state are observed, the likelihood function is given by a filter that is difficult to evaluate because it involves taking conditional expectations of the latent factors conditioned on the observed factors. We construct an efficient estimator of the filter by applying an additional change of measure. Under the new measure, the latent and the observable factors are independent and have simple parameter-free laws. The likelihood filter can therefore be evaluated with little effort under the new measure via exact simulation. Similar estimators can also be constructed for other filters, such as those involving posterior moments of the latent factors that are popular in empirical applications. When combined with our density estimators, our approach yields filter estimators that converge at the canonical square-root rate to the true filter.

¹A process is reducible if it can be transformed to a unit volatility process; see [Aït-Sahalia \(2008\)](#). Reducibility is a restrictive assumption that is violated by many popular models.

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

³See [Kloeden and Platten \(1999\)](#) for an overview of Euler’s method.

Alternative filter estimation methodologies generally only achieve subcanonical rates of convergence; see our discussion in Section 2.2.

We propose simulated maximum likelihood estimators obtained by maximizing the likelihood estimator we construct via simulation. We show that our parameter estimators 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 mild conditions, our parameter estimators converge to true maximum likelihood estimators as the number of Monte Carlo replications grows large while keeping the data sample fixed. Our parameter estimators are consistent if the number of Monte Carlo replications grows as the number of data points grows. If the number of Monte Carlo replications grows at a faster rate than the data, our parameter estimators are also asymptotically normal with the same asymptotic variance-covariance matrix as true maximum likelihood estimators. As a result, our parameter estimators are asymptotically efficient in the sense that they have the same asymptotic standard errors as true maximum likelihood estimators.

An important implication of these results is that the Monte Carlo methodology we use to estimate the likelihood does not affect the asymptotic distribution of the resulting parameter estimators. This holds because our density estimators are unbiased. Were they not unbiased, then their biases would transfer over to the parameter estimators either by making them inconsistent or asymptotically inefficient. [Detemple et al. \(2006\)](#) establish this result in the diffusion case, and numerical experiments in [Giesecke and Schwenkler \(2017b\)](#) suggest that the same holds in the jump-diffusion case. Overall, the fact that our density estimators are unbiased is the main property that enables efficient parameter estimation in this paper.

Our density estimators have important computational features. They can be evaluated at multiple arguments without re-simulation; a single set of Monte Carlo replications suffices to evaluate the density estimators at different arguments. This property entails significant computational benefits when carrying out parameter inference. It reduces the maximization of the likelihood estimated via simulation to a deterministic problem that

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

can be solved using standard methods. Furthermore, our density estimators can be fine-tuned to minimize their variances for a given number of Monte Carlo replications. This feature makes our density estimators highly accurate in practical applications. Finally, because our density estimators are constructed from independent Monte Carlo replications, their computations can be easily parallelized, yielding further computational benefits.

A numerical case study showcases the benefits of our estimators. We consider an affine consumption growth model with a persistent long-run risk factor driving the conditional growth rate of consumption. The distribution of consumption growth is non-Gaussian, 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 et al. \(2000\)](#). The numerical results show that our density estimators are highly accurate. They are able to capture the non-Gaussian distribution of consumption growth. Our density estimators have the fastest rate of convergence across a class of competing simulation-based density estimators. This class includes a standard Euler discretization based estimator, and a Gaussian kernel estimator proposed by [Kristensen and Shin \(2012\)](#). Our filter and parameter estimators are also found to be highly accurate and efficient. Our filter estimator has the fastest rate of convergence among a class of competing estimators, which includes standard particle filter and MCMC estimators. Our parameter estimators are able to closely recover the data-generating parameters.

The rest of this paper is organized as follows. [Section 2](#) formulates the model and the estimation problem, and provides a brief literature overview. In [Section 3](#), we derive our transition density representation. We derive an unbiased estimator of the transition density estimator in [Section 4](#). Its implementation is discussed in [Section 5](#). [Sections 6](#) and [7](#) introduce our filter and parameter estimators. A numerical case study is carried out in [Section 8](#). There are several technical appendices, one containing the proofs of our results.

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

$$dX_t = \mu(X_t; \theta)dt + \Sigma(X_t; \theta)dB_t + dL_t, \quad (1)$$

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

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

with event stopping times $(T_n)_{n \geq 1}$ and jump intensity $\lambda_t = \Lambda(X_t; \theta)$ for a function $\Lambda : \mathcal{S} \times \Theta \rightarrow \mathbb{R}_+$. The jump magnitudes of the process X are determined by the function $\Gamma : \mathcal{S} \times \mathcal{D} \times \Theta \rightarrow \mathbb{R}^d$. 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 . We assume that X_0 is a random variable that takes values in \mathcal{S} according to the probability density $p_0(\cdot; \theta)$. The drift, volatility, jump intensity, jump magnitude, and initial distribution are specified by a parameter $\theta \in \Theta$ to be estimated. Here, Θ is a subset of Euclidean space.

We impose the following assumptions. First, the boundary of \mathcal{S} is unattainable. 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 $\sup_{i \geq 1} |t_i - t_{i-1}| < \infty$. We also assume for simplicity that the process N and the mark variables $(D_n)_{n \geq 1}$ 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 et al. \(2013\)](#), [Komatsu and Takeuchi \(2001\)](#), and [Takeuchi \(2002\)](#) provide sufficient conditions.

We use the following notation. A subscript in \mathbb{P}_θ or \mathbb{E}_θ 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 $1 \leq i, j, k \leq r$, write ∂_i , $\partial_{i,j}^2$, and $\partial_{i,j,k}^3$ for the first, second, and third partial derivatives with respect to $\theta_i, \theta_j, \theta_k$.

2.1 Inference problem

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

The data available for inference includes observations of a subset of the components of X at the discrete observation times. More precisely, suppose that X can be decomposed as $X = (X_o, X_l)$, where X_o is valued in $\mathcal{S}_o \subset \mathbb{R}^{d_o}$ and X_l is valued in $\mathcal{S}_l \subset \mathbb{R}^{d_l}$ with $d_o + d_l = d$ and $d_o, d_l > 0$. We assume that the component X_o can be observed without error at the times $(t_i)_{1 \leq i \leq m}$, while the component X_l is latent and can never be observed by the econometrician (therefore, the subindex “o” stands for “observed” and “l” stands for “latent”). The data available for inference is $\mathbf{D}_m = \{X_{o,t_0}, \dots, X_{o,t_m}\}$, which is a random variable valued in \mathcal{S}_1^m and measurable with respect to \mathcal{B}_1^m with $\mathcal{B}_1 = \sigma(\mathcal{S}_1)$.

Because of the Markovian structure of (1), the likelihood of θ at the data \mathbf{D}_m is

$$\mathcal{L}_m(\theta) = \mathbb{E}_{\theta^*} \left[\frac{p_0(X_0; \theta)}{p_0(X_0; \theta^*)} \prod_{i=1}^m \frac{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta)}{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta^*)} \middle| \mathbf{D}_m \right], \quad (3)$$

where $p_\Delta(x, \cdot; \theta)$ is the Radon-Nikodym density of the law of X_Δ given $X_0 = x$ with respect to the Lebesgue measure on $(\mathcal{S}, \mathcal{B})$ (the transition density of X). A maximum likelihood estimator (MLE) satisfies

$$\hat{\theta}_m \in \arg \max_{\theta \in \Theta} \mathcal{L}_m(\theta) \quad (4)$$

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

$$\nabla \mathcal{L}_m(\hat{\theta}_m) = 0. \quad (5)$$

We will make several assumptions about the likelihood and the maximum likelihood estimator. First, we assume that the likelihood is almost surely bounded: $\sup_{\theta \in \Theta} \mathcal{L}_m(\theta) < \infty$ with probability one. Second, we assume that the likelihood $\mathcal{L}_m(\theta)$ is almost surely twice

continuously differentiable for all $\theta \in \Theta$ and $m \geq 1$. Third, we assume that the maximum likelihood estimator $\hat{\theta}_m$ is consistent, asymptotically normal, and asymptotically efficient. That is, $\hat{\theta}_m \rightarrow \theta^*$ in \mathbb{P}_{θ^*} -probability and $\sqrt{m}(\hat{\theta}_m - \theta^*) \rightarrow N(0, I(\theta^*)^{-1})$ in \mathbb{P}_{θ^*} -distribution as $m \rightarrow \infty$, where $I(\theta^*)$ is the Fisher information matrix. These assumptions are standard. Sufficient conditions can be found, e.g., in [Douc and Matias \(2001\)](#), [Douc et al. \(2004\)](#), and [Fuh \(2006\)](#).

The likelihood (3) represents a filter that is difficult to compute due to several layers of difficulty. First, the density p_Δ is intractable for most jump-diffusion models outside of the affine class. Second, the expectation in (3) is taken with respect to the conditional law of $(X_{2,t_i})_{0 \leq i \leq m}$ given $(X_{1,t_i})_{0 \leq i \leq m}$, which is often difficult to evaluate. Third, the true parameter θ^* governing the expectation in (3) is unknown. We will take several steps in the following sections to resolve these issues.

2.2 Available methodologies

There is an extensive literature on parameter inference for jump-diffusion processes based on discretely observed data. [Giesecke and Schwenkler \(2017b\)](#) estimate the transition density p_Δ of reducible jump-diffusions using exact simulation techniques. The density estimator of [Giesecke and Schwenkler \(2017b\)](#) is unbiased, and computationally efficient for large computational budgets. The parameter estimators of [Giesecke and Schwenkler \(2017b\)](#) also inherit the asymptotic properties of maximum likelihood estimators. The method of [Giesecke and Schwenkler \(2017b\)](#) 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. As a result, the methodology of [Giesecke and Schwenkler \(2017b\)](#) is not generally applicable for the class of models we consider in this paper.

Inspired by the 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 (i.e., models for which the function $\Gamma(x, d; \theta)$ does not depend upon x). The coefficients of the expansion satisfy a set of interdependent par-

tial 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 of [Yu \(2007\)](#) inherit the asymptotic properties of maximum likelihood estimators in high frequency regimes when the time between consecutive observations shrinks to zero as more data points become available. Because of this, the estimators of [Yu \(2007\)](#) do not apply for the discrete data that we consider in this paper.

[Kristensen and Shin \(2012\)](#) derive nonparametric estimators of the transition density of a jump-diffusion process with state-independent coefficient functions.⁵ These authors apply a kernel estimator to samples of the jump-diffusion process derived from Euler discretization. If the kernel bandwidth shrinks to zero as the number of data points grows large, then the resulting parameter estimators inherit the asymptotic properties of maximum likelihood estimators. The results of [Kristensen and Shin \(2012\)](#) apply in our setting only when the coefficient functions μ , Σ , Γ , and Λ do not depend upon x .

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 is given in terms of solutions to ordinary differential equations. The method of [Duffie et al. \(2000\)](#) is applicable in our setting only when the model (1)-(2) is affine. [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. The method of [Lo \(1988\)](#) applies in our setting only when the coefficient functions μ , Σ , Λ , and Γ are state-independent.

[Gouriéroux et al. \(1993\)](#) and [Smith \(1993\)](#) propose methods of indirect inference that may apply for multivariate jump-diffusions. Indirect inference requires that one specifies an auxiliary model, and that one estimates the function binding the auxiliary model and the true model via simulation. The estimation of the binding function may become infeasible if it demands that we simulate exact samples of a general multivariate jump-diffusion process. Currently, there are no exact simulation techniques that apply for

⁵The assumption that the distribution of ϵ_t is independent of t and θ in equation (1) of [Kristensen and Shin \(2012\)](#) effectively restricts their model to state-independent jump-diffusions.

irreducible multivariate jump-diffusions.⁶ As a result, indirect inference methods may not be feasible for the general class of multivariate jump-diffusion models we consider.

Moment-based methods can also be used for parameter inference. [Jiang and Knight \(2002\)](#), [Chacko and Viceira \(2003\)](#), [Duffie and Glynn \(2004\)](#), and [Duffie and 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

One may also use particle filter or Markov Chain Monte Carlo (MCMC) methods to carry out parameter inference; see [Johannes and Polson \(2010\)](#) for an overview. This approach takes a Bayesian view of the world, making no assumptions about the existence of a data-generating parameter. Consequently, no consistency or asymptotic normality results can be established for the parameter estimators resulting from MCMC methods.

Another popular methodology that is applicable in our setting is the expectation-maximization (EM) algorithm of [Dempster et al. \(1977\)](#). The EM algorithm is an iterative method that alternates between an expectation step, in which an estimator of the likelihood is constructed via simulation for a given parameter, and a maximization step, in which the estimated likelihood is maximized. [Nielsen \(2000\)](#) shows that the parameter estimator derived from the EM algorithm is consistent and asymptotically normal, but not asymptotically efficient. This is due to the simulation noise that is introduced in the expectation step.

The methodology we will introduce in the following sections applies for Model (1)-(2) without major restrictions on the coefficient functions μ , Σ , Λ , and Γ , or the data \mathbf{D}_m . Unlike the aforementioned methods, our approach will yield asymptotically efficient parameter estimators for a general class of models.

⁶We refer to [Giesecke and Smelov \(2013\)](#) for the exact simulation of reducible jump-diffusions. [Henry-Labordère et al. \(2015\)](#) develop exact simulation tools for multivariate diffusions.

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

3 Density representation

We derive a tractable representation of the transition density p_Δ that can be estimated without bias via Monte Carlo simulation. We exploit a change of measure to accomplish this goal. Consider the random variable

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

for $\theta \in \Theta$ and $\ell > 0$. If $\mathbb{E}_\theta[Z_\Delta(\theta)] < \infty$, then $Z_\Delta(\theta)$ defines an equivalent probability measure \mathbb{Q}_θ on $(\Omega, \mathcal{F}_\Delta)$ given by $\mathbb{Q}_\theta[A] = \mathbb{E}_\theta[Z_\Delta(\theta)1_A]$ for any $A \in \mathcal{F}_\Delta$. The theorems of Lévy and Watanabe imply that, under \mathbb{Q}_θ and on $[0, \Delta]$, N is a Poisson process with rate ℓ . Consequently, jumps of the process X arrive at a constant rate under \mathbb{Q}_θ . 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) *The process $(X_t : t \in [0, \Delta])$ is a strong Markov process.*

(A2) *For any $\theta \in \Theta$, the variable $Z_\Delta(\theta)$ has finite mean: $\mathbb{E}_\theta[Z_\Delta(\theta)] < \infty$.*

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

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

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

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

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

Theorem 3.1 rewrites the transition density p_Δ as a mixture of transition densities of diffusion processes of the type (7). This results is an implication of Bayes' formula. Under Assumption (A1) and conditional on $(T_n, D_n, X_{T_n-})_{n \leq N_\Delta}$, 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 (A1) and conditional on $(T_n, D_n, X_{T_n-})_{n \leq N_\Delta}$, 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 $(T_n, D_n, X_{T_n-})_{n \leq N_\Delta}$. 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 (A2) guarantees that the change of measure is well-defined. It is a standard regularity assumption; Blanchet and Ruf (2013) and Protter and Shimbo (2008) give sufficient conditions. Assumption (A1) is also standard; see Protter (2004, Theorem 32). The diffusion density \tilde{p}_t exists given that we assumed that the jump-diffusion density p_t exists.

The density representation (8) generalizes the recently developed representation of Giesecke and Schwenkler (2017b), which applies whenever the jump-diffusion X can be reduced to a unit volatility process via the Lamperti transform. Under the reducibility assumption, Giesecke and Schwenkler (2017b) can simplify the dynamics of \tilde{X} in (7) to have constant drift and unit volatility. Such a simplification yields a density representation that is easier to estimate than ours. However, the reducibility assumption significantly restricts the set of applications as it imposes strong constraints on the variance-covariance matrix Σ . Many models of practical relevance are not reducible.⁸ Because we do not impose the reducibility constraint, our representation is applicable for a broader class of models than the one of Giesecke and Schwenkler (2017b).

Theorem 3.1 complements the well-known density representations of Dacunha-Castelle and 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) facilitates the derivation of conditions under which the tran-

⁸For example, the stochastic volatility model of Heston (1993) is not reducible, but it is extensively used in the options pricing literature; see Andersen et al. (2002), Eraker et al. (2003), and Eraker (2004), among many others.

sition density is smooth with respect to the parameter θ . Smoothness is necessary for consistency and asymptotic normality of maximum likelihood estimators; see Proposition 7.2. For smoothness of the density, we only require smoothness of the coefficient functions and differentiability of the diffusive density \tilde{p} . Our conditions are 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 and Takeuchi (2001), and Takeuchi (2002)).

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

(A3) *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.*

(A4) *For all $t \in [0, \Delta]$, the diffusive density $\tilde{p}_t(v, w; \theta)$ is n -times differentiable at almost all $v, w \in \mathcal{S}$ and $\theta \in \Theta$.*

(A5) *The partial derivatives up to n -th order of $\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{Z_\Delta(\theta)}$ with respect to θ are uniformly bounded in expectation in the following sense: For all $1 \leq k \leq n$ and $i_1, \dots, i_k \in \{1, \dots, r\}$,*

$$\sup_{\theta \in \Theta} \sup_{v, w \in \mathcal{S}} \mathbb{E}_\theta^\mathbb{Q} \left[\partial_{i_1, \dots, i_k}^k \frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{Z_\Delta(\theta)} \mid X_0 = v \right] < \infty.$$

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

4 Density estimator

We now construct an unbiased Monte Carlo estimator of the density

$$p_\Delta(v, w; \theta) = \mathbb{E}_\theta^\mathbb{Q} \left[\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{Z_\Delta(\theta)} \mid X_0 = v \right] \quad (9)$$

For this, we exploit a randomization technique introduced by Rhee and Glynn (2015).

4.1 Towards an unbiased estimator

Samples of N_Δ under \mathbb{Q}_θ can easily be simulated without bias from the Poisson distribution with rate ℓ . Conditional on N_Δ , we can also exactly simulate the jump times $(T_n)_{n \leq N_\Delta}$ given that their conditional distribution is the same as that of the order statistics of N_Δ uniform random variables on $[0, \Delta]$. Now, if the diffusive density \tilde{p} were known in closed form and if samples of $(X_{T_{N_\Delta}}, \frac{1}{Z_\Delta(\theta)})$ could be simulated without bias, then

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

given $X_0 = v$ would be an unbiased Monte Carlo estimator of (9). However, for most specifications of the jump-diffusion in (1), the diffusive density \tilde{p} is not known in closed form and one cannot sample exactly from the distribution of $(X_{T_{N_\Delta}}, \frac{1}{Z_\Delta(\theta)})$. We circumvent these issues step by step.

4.1.1 Euler discretization

We can generate a biased estimator of $(X_{T_{N_\Delta}}, \frac{1}{Z_\Delta(\theta)})$ using Euler discretization between jump times.⁹ To do this, we first generate exact samples of N_Δ and also samples of $(T_n)_{n \leq N_\Delta}$ conditional on N_Δ . We fix $T_0 = 0$ and $T_{N_\Delta+1} = \Delta$ for simplicity. Pick $J \in \mathbb{N}$, and set $h_n = \frac{T_n - T_{n-1}}{J}$. For $n \in \{0, \dots, N_\Delta + 1\}$ and $j \in \{0, \dots, J\}$, let $(X_{n,j}^J, Z_{n,j}^{-J})$ be an estimator of $(X_{T_n + jh_n}, \frac{1}{Z_{T_n + jh_n}(\theta)})$ constructed via Euler discretization with J Euler steps between jump times. The Euler discretization estimator can be constructed as follows:

- Initialize $X_{0,0}^J = X_0$ and $Z_{0,0}^{-J} = 1$.
- For $n \in \{0, \dots, N_\Delta + 1\}$ and $j \in \{0, \dots, J\}$, set

$$X_{n,j}^J = \begin{cases} X_{n,j-1}^J + \mu(X_{n,j-1}^J; \theta) h_n + \Sigma(X_{n,j-1}^J; \theta) (B_{jh_n} - B_{(j-1)h_n}), & j > 0 \\ X_{n-1,J}^J + \Gamma(X_{n-1,J}^J, D_n; \theta), & n > 0, j = 0 \end{cases} \quad (10)$$

$$Z_{n,j}^{-J} = \begin{cases} Z_{n,j-1}^{-J} \exp\left(\left(\ell - \Lambda(X_{n,j-1}^J; \theta)\right) h_n\right), & 1 \leq j \leq J \\ \frac{\Lambda(X_{n-1,J}^J; \theta)}{\ell} Z_{n-1,J}^{-J}, & n > 0, j = 0 \end{cases} \quad (11)$$

⁹We refer to [Jacod and Protter \(1998\)](#) for an overview of the Euler discretization method.

The construction (10)-(11) ensures that Euler discretizations between consecutive jumps are correctly pasted together by accounting for the jumps of X and $Z(\theta)$.

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 $(X_{T_{N_\Delta}}, \frac{1}{Z_\Delta(\theta)})$. Consequently,

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

is a biased estimator of the density $p_\Delta(v, w; \theta)$ in (9).

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 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 $\tilde{h} = \frac{\Delta - T_{N_\Delta}}{I}$:

$$\tilde{X}_0^I = \mathcal{X}^J, \quad \tilde{X}_i^I = \tilde{X}_{i-1}^I + \mu(\tilde{X}_{i-1}^I; \theta) \tilde{h} + \Sigma(\tilde{X}_{i-1}^I; \theta) (\tilde{B}_{i\tilde{h}} - \tilde{B}_{(i-1)\tilde{h}}) \quad (12)$$

for $1 \leq i \leq I$. Conditional on T_{N_Δ} and \mathcal{X}^J , the law of \tilde{X}_I^I is mixed Gaussian because each increment in the above Euler discretization is normally distributed. More precisely, the conditional density of \tilde{X}_I^I given T_{N_Δ} and $\tilde{X}_0 = v$ is

$$\tilde{\mathcal{P}}^I(v, w; \theta) = \int \phi(x_1; v, \tilde{h}) \prod_{i=2}^{I-1} \phi(x_i; x_{i-1}, \tilde{h}) \phi(w; x_{I-1}, \tilde{h}) dx_1 \dots dx_{I-1}, \quad (13)$$

where $\phi(\cdot; x, h)$ 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 (13) can be computed using standard numerical routines; see Section 5. We know from Bally and Talay (1996) that the difference between the density $\tilde{\mathcal{P}}^I$ and the true density \tilde{p} is of order $O(I^{-1})$. As a result, $\tilde{\mathcal{P}}^I$ serves as a first-order approximation of \tilde{p} .

We can now compute a biased estimator of the density (9), namely

$$\hat{p}_\Delta^{I,J}(v, w; \theta) = \tilde{\mathcal{P}}^I(\mathcal{X}^J, w; \theta) \mathcal{Z}^J. \quad (14)$$

The bias of the estimator (14) vanishes in the limit $I \rightarrow \infty$ and $J \rightarrow \infty$. That is,

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

4.1.3 Randomization

One drawback of the density estimator (14) is that it is biased for any finite I and J . Next, we exploit a randomization technique introduced by Rhee and Glynn (2015) to obtain an unbiased estimator of the density (9).

Let Ξ be a \mathcal{F}_0 -measurable random variable valued in \mathbb{N} . Assume that the distribution of Ξ is independent of the parameter θ and the initial value X_0 , and write $q_\xi = \mathbb{Q}_\theta[\Xi = \xi]$. Consider subsequences $(I_\xi)_{\xi \geq 0}$ and $(J_\xi)_{\xi \geq 0}$ so that $I_\xi \rightarrow \infty$ and $J_\xi \rightarrow \infty$ as $\xi \rightarrow \infty$. The asymptotic unbiasedness of the estimator (14) yields, under certain regularity conditions,

$$\begin{aligned}
p_\Delta(v, w; \theta) &= \lim_{\xi \rightarrow \infty} \mathbb{E}_\theta^\mathbb{Q} \left[\hat{p}_\Delta^{I_\xi, J_\xi}(v, w; \theta) \mid X_0 = v \right] \\
&= \hat{p}_\Delta^{I_0, J_0}(v, w; \theta) + \sum_{\xi \geq 1} \mathbb{E}_\theta^\mathbb{Q} \left[\hat{p}_\Delta^{I_\xi, J_\xi}(v, w; \theta) - \hat{p}_\Delta^{I_{\xi-1}, J_{\xi-1}}(v, w; \theta) \mid X_0 = v \right] \\
&= \hat{p}_\Delta^{I_0, J_0}(v, w; \theta) + \sum_{\xi \geq 1} \mathbb{E}_\theta^\mathbb{Q} \left[\frac{\hat{p}_\Delta^{I_\xi, J_\xi}(v, w; \theta) - \hat{p}_\Delta^{I_{\xi-1}, J_{\xi-1}}(v, w; \theta)}{q_\xi} \mid X_0 = v \right] q_\xi \\
&= \hat{p}_\Delta^{I_0, J_0}(v, w; \theta) + \mathbb{E}_\theta^\mathbb{Q} \left[\frac{\hat{p}_\Delta^{I_\Xi, J_\Xi}(v, w; \theta) - \hat{p}_\Delta^{I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_\Xi} \mid X_0 = v \right]. \quad (15)
\end{aligned}$$

The last equality follows because Ξ is \mathcal{F}_0 -measurable and independent of θ and X_0 . The calculations in (15) imply that

$$\hat{p}_\Delta^{I_0, J_0}(v, w; \theta) + \frac{\hat{p}_\Delta^{I_\Xi, J_\Xi}(v, w; \theta) - \hat{p}_\Delta^{I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_\Xi}$$

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 below.

Theorem 4.1. *Let $\Delta > 0$ be given. Fix sequences $(J_\xi)_{\xi \geq 0}$ and $(I_\xi)_{\xi \geq 0}$ such that $I_\xi \rightarrow \infty$ and $J_\xi \rightarrow \infty$ as $\xi \rightarrow \infty$. Let Ξ be an \mathcal{F}_0 -measurable random variable valued in \mathbb{N} with distribution $q_\xi = \mathbb{Q}_\theta[\Xi = \xi]$ that is independent of the parameter θ and the initial value X_0 . Let $(\mathcal{X}^J, \mathcal{Z}^J)$ be samples of $(X_{T_{N_\Delta}}, \frac{1}{Z_{\Delta}(\theta)})$ constructed via the Euler discretization (10)-(11) with J steps between jump times. In addition, let $\tilde{\mathcal{P}}^I$ be a mixed Gaussian density*

as in (13) derived from the Euler discretization (12) of \tilde{X} with I steps. Assume that the conditions of Theorem 3.1 are valid. Suppose that the following condition also holds.

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

$$\sum_{\xi \geq 1} \frac{\left\| \hat{p}_{\Delta}^{I_{\xi}, J_{\xi}}(v, w; \theta) - p_{\Delta}(v, w; \theta) \right\|_2^2}{q_{\xi}} < \infty.$$

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

$$\hat{p}_{\Delta}(v, w; \theta) = \hat{p}_{\Delta}^{I_0, J_0}(v, w; \theta) + \frac{\hat{p}_{\Delta}^{I_{\Xi}, J_{\Xi}}(v, w; \theta) - \hat{p}_{\Delta}^{I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_{\Xi}} \quad (16)$$

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

$$\mathbb{E}_{\theta}^{\mathbb{Q}}[\hat{p}_{\Delta}(v, w; \theta)] = p_{\Delta}(v, w; \theta).$$

The estimator \hat{p}_{Δ} in (16) is given in terms of the sum of a biased density and a correction term. If $I_0 = J_0 = 0$, then $\hat{p}_{\Delta}^{I_0, J_0}(v, w; \theta)$ is the density of the multivariate normal distribution with mean $v + \mu(v; \theta)\Delta$ and variance-covariance matrix $\Delta\Sigma(v; \theta)\Sigma^{\top}(v; \theta)$. This density is generally not equal to the true density $p_{\Delta}(v, w; \theta)$. Theorem 4.1 corrects for this bias by adding the random term $\frac{\hat{p}_{\Delta}^{I_{\Xi}, J_{\Xi}}(v, w; \theta) - \hat{p}_{\Delta}^{I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_{\Xi}}$. The sum of the biased density plus the correction term recover, on average, the true density.

The key advantage of the density estimator $\hat{p}_{\Delta}(v, w; \theta)$ is that it is unbiased for every $v, w \in \mathcal{S}$, $\theta \in \Theta$, and $\Delta > 0$. This property yields important benefits when performing maximum likelihood estimation based on \hat{p}_{Δ} . In particular, the unbiasedness property ensures that when one uses the density estimator \hat{p}_{Δ} for maximum likelihood inference, the resulting parameter estimators are asymptotically efficient and asymptotically unbiased; see Section 7. This is generally not possible if one were to use a biased density estimator.¹⁰

We conclude this section by remarking 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.

¹⁰We refer to Detemple et al. (2006) for a theoretical demonstration of this statement in the univariate diffusion case ($d = 1$ and $\Gamma \equiv 0$). Giesecke and Schwenkler (2017b) provide numerical evidence in support of this claim in the univariate jump-diffusion case ($d = 1$). We derive additional numerical evidence in the multivariate case ($d > 1$) in Section 8.

Proposition 4.2. *Suppose that the conditions of Theorem 4.1 hold. Furthermore, suppose that the following condition also holds:*

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

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

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,

$$\mathbb{E}_\theta^\mathbb{Q} [\partial_{i_1, \dots, i_n}^n \hat{p}_\Delta(v, w; \theta)] = \partial_{i_1, \dots, i_n}^n p_\Delta(v, w; \theta) \quad \text{for all } i_1, \dots, i_n \in \{1, \dots, r\}.$$

5 Computation of the density estimator

We specify 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 , and the methodology to compute the mixed normal density $\tilde{\mathcal{P}}^I$.

5.1 Approximating $\tilde{\mathcal{P}}^I$

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

$$\tilde{\mathcal{P}}^{H,I}(v, w; \theta) = \frac{1}{H} \sum_{\nu=1}^H \phi\left(w; \tilde{X}_{I-1}^{I,\nu}, \tilde{h}\right), \quad (17)$$

where $\tilde{X}_{I-1}^{I,\nu}$ is the ν -th sample of \tilde{X}_{I-1}^I , and $\tilde{X}_0^{I,\nu} = v$ 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 (14), and the density estimator \hat{p}_Δ remains unbiased. In other words, if we set

$$\begin{aligned} \hat{p}_\Delta^{H,I,J}(v, w; \theta) &= \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^J, w; \theta) \mathcal{Z}^J, \\ \hat{p}_\Delta(v, w; \theta) &= \hat{p}_\Delta^{H_0, I_0, J_0}(v, w; \theta) + \frac{\hat{p}_\Delta^{H_\Xi, I_\Xi, J_\Xi}(v, w; \theta) - \hat{p}_\Delta^{H_{\Xi-1}, I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_\Xi}, \end{aligned}$$

then $\hat{p}_\Delta(v, w; \theta)$ is still an unbiased estimator of $p_\Delta(v, w; \theta)$.

5.2 Finite variance

We now specify the sequences $(H_\xi, I_\xi, J_\xi)_{\xi \geq 0}$ to ensure that the variance of \hat{p}_Δ is finite. We know that $\|\hat{p}_\Delta^{H,I,J}(v, w; \theta) - p_\Delta(v, w; \theta)\|_2$ is of the order

$$O(\|\mathcal{Z}^J - Z_\Delta^{-1}(\theta)\|_2) + O\left(\left\|\tilde{\mathcal{P}}^{H,I}(v, w; \theta) - \tilde{p}_{\Delta-T_{N_\Delta}}(v, w; \theta)\right\|_2\right). \quad (18)$$

As a result, the rate of convergence of $\hat{p}_\Delta^{H,I,J}$ is the minimum between the rate of convergence of \mathcal{Z}^J and that of $\tilde{\mathcal{P}}^{H,I}$. It follows that an optimal choice of $(H_\xi, I_\xi, J_\xi)_{\xi \geq 0}$ is one that ensures that both terms in (18) converge to zero at the same rate. Now, Theorem 1 of [Rhee and Glynn \(2015\)](#) implies that the estimator \hat{p}_Δ has finite variance whenever the rate of convergence in (18) is $\frac{1}{2}$. We will thus choose $(H_\xi, I_\xi, J_\xi)_{\xi \geq 0}$ to ensure that both terms in (18) have rate of convergence $\frac{1}{2}$.

It is well-known that Euler discretization has strong rate of convergence of order $\frac{1}{2}$ (see, e.g., [Jacod and Protter \(1998\)](#)). In our case, because we carry out Euler discretization between consecutive jump times of X under \mathbb{Q}_θ , we have

$$\|\mathcal{Z}^J - Z_\Delta^{-1}(\theta)\|_2 = O(\ell \Delta J^{-1/2}). \quad (19)$$

On the other hand, a key result by [Gobet and Labart \(2008\)](#) implies that

$$\left\|\tilde{\mathcal{P}}^{H,I}(v, w; \theta) - \tilde{p}_{\Delta-T_{N_\Delta}}(v, w; \theta)\right\|_2^2 = O\left(I^{-2} + \text{Var}_\theta^\mathbb{Q}\left(\tilde{\mathcal{P}}^{H,I}(v, w; \theta)\right)\right). \quad (20)$$

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

$$\text{Var}_\theta^\mathbb{Q}\left(\tilde{\mathcal{P}}^{H,I}(v, w; \theta)\right) = O(H^{-1}I^{d/2}) \quad (21)$$

for all $\theta \in \Theta$, $\ell > 0$, and $v, w \in \mathcal{S}$ if $\det V_{I,\theta}$ is bounded away from zero. We can therefore set $H_\xi = O(I_\xi^{2+d/2})$ and $I_\xi = O(\sqrt{J_\xi})$ to equalize the rates of convergence of (19) and (20). These choices yield

$$\left\|\hat{p}_\Delta^{H_\xi, I_\xi, J_\xi}(v, w; \theta) - p_\Delta(v, w; \theta)\right\|_2 = O\left(J_\xi^{-\frac{1}{2}}\right). \quad (22)$$

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 $1/2$, as required by [Rhee and Glynn \(2015\)](#).

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

(C1) The determinant of $\Sigma(x; \theta)\Sigma(x; \theta)^\top$ is uniformly bounded from below. That is,

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

(C2) The variance of $Z_\Delta^{-1}(\theta)$ is uniformly bounded. That is,

$$\sup_{\theta \in \Theta} \sup_{v \in \mathcal{S}} \text{Var}_\theta^\mathbb{Q}(Z_\Delta^{-1}(\theta) \mid X_0 = v) < \infty.$$

Then,

$$\hat{p}_\Delta(v, w; \theta) = \hat{p}_{\Delta}^{H_0, I_0, J_0}(v, w; \theta) + \frac{\hat{p}_{\Delta}^{H_\Xi, I_\Xi, J_\Xi}(v, w; \theta) - \hat{p}_{\Delta}^{H_{\Xi-1}, I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)}{q_\Xi} \quad (23)$$

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}_Δ is finite. That is,

$$\mathbb{E}_\theta^\mathbb{Q}[\hat{p}_\Delta(v, w; \theta)] = p_\Delta(v, w; \theta) \quad \text{and} \quad \text{Var}_\theta^\mathbb{Q}(\hat{p}_\Delta(v, w; \theta)) < \infty$$

for all $v, w \in \mathcal{S}$ and $\theta \in \Theta$.

5.3 Computational costs

A single sample of \hat{p}_Δ is a noisy estimator of the density p_Δ . To reduce the noise, it is natural to consider the Monte Carlo estimator

$$\hat{p}_\Delta^K(v, w; \theta) = \frac{1}{K} \sum_{k=1}^K \hat{p}_\Delta^{(k)}(v, w; \theta),$$

where $(\hat{p}_\Delta^{(k)}(v, w; \theta))_{1 \leq k \leq K}$ are independent samples of the density estimator $\hat{p}_\Delta(v, w; \theta)$.

Because the variance of \hat{p}_Δ is finite, standard Monte Carlo theory tells us that

$$\|\hat{p}_\Delta^K(v, w; \theta) - p_\Delta(v, w; \theta)\|_2 = O(K^{-\frac{1}{2}}). \quad (24)$$

As a result, we can control the accuracy of the Monte Carlo estimator \hat{p}_Δ^K by choosing the number K of Monte Carlo replications. Large values of K will result in estimators with

small errors. However, choosing a large value for K increases the computational burden to evaluate the Monte Carlo estimator \hat{p}_Δ^K . This observation raises the following question: How much computational effort is necessary to evaluate a version of \hat{p}_Δ^K that does not exceed a given error bound with high probability?

The computational cost of an evaluation of a sample of \hat{p}_Δ is determined by the realization of J_Ξ :

$$\text{Effort}(\hat{p}_\Delta(v, w; \theta)) = O_P(J_\Xi + H_\Xi J_\Xi) = O_P\left(J_\Xi^{\frac{6+d}{4}}\right). \quad (25)$$

We can control the computational expenses required to evaluate \hat{p}_Δ^K by optimally choosing the sequence $(J_\xi)_{\xi \geq 0}$ of Euler steps and the distribution $(q_\xi)_{\xi \geq 0}$ of Ξ . It is necessary that $J_\xi \rightarrow \infty$ as $\xi \rightarrow \infty$ in order to achieve convergence. As a result, there needs to be sufficiently large positive probability that J_Ξ will be arbitrarily large. However, large values of J_Ξ are associated with large computational costs in (25). These observations suggests that we should keep Ξ concentrated around small values to minimize the computational expenses, and at the same time let J_ξ grow quickly enough as $\xi \rightarrow \infty$. Proposition 5.2 proposes an optimal choice.

Proposition 5.2. *Suppose that the conditions of Proposition 5.1 are satisfied. Fix*

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

for $\xi \geq 1$. Then:

- (1) *The Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ is an unbiased estimator of $p_\Delta(v, w; \theta)$ for any $v, w \in \mathcal{S}$ and $\theta \in \Theta$: $\mathbb{E}_\theta^\mathbb{Q}[\hat{p}_\Delta^K(v, w; \theta)] = p_\Delta(v, w; \theta)$.*
- (2) *The root-mean squared error of the Monte Carlo estimator \hat{p}_Δ^K decays at the canonical rate: $\|\hat{p}_\Delta^K(v, w; \theta) - p_\Delta(v, w; \theta)\|_2 = O(K^{-0.5})$.*
- (3) *The computational effort necessary to evaluate the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ is almost surely finite for all $v, w \in \mathcal{S}$ and $\theta \in \Theta$.*
- (4) *The computational effort necessary to evaluate the density estimator \hat{p}_Δ^K with a maximum error of $\epsilon > 0$ is of order $O_P(\epsilon^{-(3+d/2)} \log_2^2(1/\epsilon))$. That is,*

$$\|\hat{p}_\Delta^K(v, w; \theta) - p_\Delta(v, w; \theta)\|_2 \leq \epsilon \quad \Rightarrow \quad \frac{\text{Effort}(\hat{p}_\Delta^K(v, w; \theta))}{\epsilon^{-(3+d/2)} \log_2^2(1/\epsilon)} = O_P(1).$$

This is the slowest rate of divergence of $\text{Effort}(\hat{p}_\Delta^K(v, w; \theta))$ as $\epsilon \rightarrow 0$ among all possible choices of $(J_\xi)_{\xi \geq 1}$ and $(q_\xi)_{\xi \geq 1}$ with $J_\xi \rightarrow \infty$ as $\xi \rightarrow \infty$, $q_\xi \geq 0$, and $\sum_{\xi \geq 1} q_\xi = 1$.

Proposition 5.2 exposes the computational costs of our Monte Carlo density estimator. First, we see that the Monte Carlo estimator can always be evaluated with finite computational resources. Nonetheless, we also see that the computational effort necessary to evaluate the density estimator \hat{p}_Δ^K with an error of ϵ increases faster in ϵ than implied by the standard Monte Carlo theory. The computational costs associated with our density estimator may become excessively large with small probability. The reason why this occurs is that in order to achieve a small error in (24), we need to pick a large number K of Monte Carlo replications. However, if we use a large number of Monte Carlo replications to evaluate the density estimator, then there is high probability that at least one of the realizations of Ξ will be large, in which case the realization of $J_\Xi = O(2^\Xi)$ and the computational effort (25) are large.

Proposition 5.2 also shows that the rate at which the computational effort of \hat{p}_Δ^K grows in the error bound ϵ increases with the dimension d of the process X . This holds because $H_\xi = O(2^{\xi(1+d/4)})$ grows faster in ξ the larger d is, increasing the computational requirements to achieve the error bound ϵ .

In spite of the computational costs, the Monte Carlo estimator \hat{p}_Δ^K has other appealing computational features. We summarize these features in Appendix A. Most importantly, the estimator \hat{p}_Δ^K can be computed as

$$\hat{p}_\Delta^K(v, w; \theta) = \frac{1}{K} \sum_{k=1}^K P(v, w; \theta, \mathbf{R}^{(k)}), \quad (26)$$

where P is a deterministic function, and $(\mathbf{R}^{(k)})_{k=1, \dots, K}$ are i.i.d. \mathcal{F}_0 -measurable random variables that do not depend upon v , w , or θ . This means that the simulation step necessary to evaluate our density estimator \hat{p}_Δ^K can be decoupled from the parameter and the arguments at which the density is evaluated. No re-simulation is needed to evaluate the density estimator at different arguments. This feature yields important computational benefits when performing likelihood inference; see Section 7.

An additional neat computational feature of our estimator is that the jump rate $\ell > 0$ of the point process N under the measure \mathbb{Q}_θ can be chosen as to minimize the variance of a sample of our density estimator. This results in a density estimator that can achieve small errors with small computational resources. We carry out several numerical studies in Appendix 8, and show that our estimator outperforms several simulation-based density estimators in terms of accuracy and computational efficiency.

6 Filter estimation

We exploit our density estimator to deriving tractable estimators of filters of the type

$$\mathcal{E}_m(\theta; f, \theta') = \mathbb{E}_\theta [f(X_0, X_{t_1}, \dots, X_{t_m}; \theta') \mid \mathbf{D}_m] \quad (27)$$

for $\theta, \theta' \in \Theta$ and measurable functions $f : \mathcal{S}^{m+1} \times \Theta \rightarrow \mathbb{R}$. The likelihood (3) is given by $\mathcal{E}_m(\theta^*; f_m^*, \theta)$ with

$$f_m^*(x_0, x_1, \dots, x_m; \theta) = \frac{p_0(x_0; \theta)}{p_0(x_0; \theta^*)} \prod_{i=1}^m \frac{p_\Delta(x_{i-1}, x_i; \theta)}{p_\Delta(x_{i-1}, x_i; \theta^*)}. \quad (28)$$

Other filters of interest include those for which $f(x_0, x_1, \dots, x_m; \theta) = x_{l,j}^p$ for some $j \in \{0, \dots, m\}$ and $p \geq 0$. These kinds of filters give the posterior p -th moment of X_{l,t_j} conditional on $(X_{o,t_i})_{0 \leq i \leq m}$.

We exploit a change of measure to simplify the evaluation of the filter $\mathcal{E}_m(\theta; f, \theta')$. Let ψ_j^* and ϕ_j^* be strictly positive probability densities on \mathcal{S}_j for $j \in \{o, l\}$. Assume that these densities are known in closed form, and that they do not depend on the parameter θ . Define \mathbb{P}^* as the probability measure under which X_0 has density $\psi_o^* \times \psi_l^*$, and X has transition density $\phi_o^* \times \phi_l^*$. Such a probability measure can be constructed through the Radon-Nikodym density

$$\frac{d\mathbb{P}^*}{d\mathbb{P}_\theta} = \frac{\psi_o^*(X_{o,0})\psi_l^*(X_{l,0})}{p_0(X_0; \theta)} \prod_{i=1}^m \frac{\phi_o^*(X_{o,i\Delta})\phi_l^*(X_{l,i\Delta})}{p_\Delta(X_{(i-1)\Delta}, X_{i\Delta}; \theta)}.$$

Under the measure \mathbb{P}^* , the transition density of X does not depend upon the parameter θ , and the components X_o and X_l are independent of each other. These features simplify

the evaluation of filters under the measure \mathbb{P}^* , as summarized in the below proposition.¹¹

Proposition 6.1. *Fix a measurable function $f : \mathcal{S}^{m+1} \times \Theta \rightarrow \mathbb{R}$, as well as parameters $\theta, \theta' \in \Theta$. Suppose that the following condition holds.*

(D1) *The random variable $f(X_0, X_{t_1}, \dots, X_{t_m}; \theta')$ is \mathbb{P}_θ -integrable:*

$$\mathbb{E}_\theta [|f(X_0, X_{t_1}, \dots, X_{t_m}; \theta')|] < \infty.$$

Let ψ_j^ and ϕ_j^* be strictly positive densities on \mathcal{S}_j for $j \in \{o, l\}$. It holds almost surely that*

$$\mathcal{E}_m(\theta; f, \theta') = \frac{\mathcal{E}_m^*(\theta; f, \theta')}{\mathcal{E}_m^*(\theta; 1, f)}, \quad (29)$$

where

$$\mathcal{E}_m^*(\theta; f, \theta') = \mathbb{E}^* \left[f(X_0, X_{t_1}, \dots, X_{t_m}; \theta') \frac{p_0(X_0; \theta)}{\psi_l^*(X_{l,0})} \prod_{i=1}^m \frac{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta)}{\phi_l^*(X_{l,t_i})} \middle| \mathbf{D}_m \right] \quad (30)$$

Here, \mathbb{E}^* denotes the expectation operator under the measure \mathbb{P}^* .

Proposition 6.1 significantly reduces the effort necessary to evaluate the filter $\mathcal{E}_m(\theta; f, \theta')$ when combined with our density estimation methodology. Because X_{o,t_i} and X_{l,t_i} are independent under \mathbb{P}^* , the filter $\mathcal{E}_m^*(\theta; f, \theta')$ in (30) is evaluated with respect to the unconditional \mathbb{P}^* -distribution of $(X_{l,t_i})_{i=0,\dots,m}$. This distribution is given in closed form in terms of the densities ϕ_l^* and ψ_l^* , which do not depend on the true parameter θ^* or the parameter θ at which the likelihood is evaluated. As a result, exact samples of $(X_{l,t_i})_{i=0,\dots,m}$ under \mathbb{P}^* can easily be generated. If the density p_Δ were known, then we could evaluate the \mathbb{P}^* -filter $\mathcal{E}_m^*(\theta; f, \theta')$ via simulation. Nevertheless, the density p_Δ is often intractable. In those cases, we can replace the true density with our density estimator \hat{p}_Δ , and obtain an unbiased estimator of the \mathbb{P}^* -filter $\mathcal{E}_m^*(\theta; f, \theta')$. We summarize in the below proposition.

Proposition 6.2. *Suppose that the conditions of Propositions 5.1 and 6.1 are satisfied. For $i \in \{0, \dots, m\}$, let $\hat{X}_{2,i}$ be a sample of X_{2,t_i} under the measure \mathbb{P}^* . Define*

$$\hat{\mathcal{E}}_m^*(\theta; f, \theta') = f \left(\left(X_{o,0}, \hat{X}_{l,0} \right), \dots, \left(X_{o,t_m}, \hat{X}_{l,m} \right); \theta' \right) \frac{p_0 \left(\left(X_{o,0}, \hat{X}_{l,0} \right); \theta \right)}{\psi_2^* \left(\hat{X}_{l,0} \right)}$$

¹¹A similar approach is used by Giesecke and Schwenkler (2017a) to derive a tractable likelihood representation in the setting of marked point processes with incomplete data.

$$\times \prod_{i=1}^m \frac{\hat{p}_\Delta \left(\left(X_{o,t_{i-1}}, \hat{X}_{l,i-1} \right), \left(X_{o,t_i}, \hat{X}_{l,i} \right); \theta \right)}{\phi_2^* \left(\hat{X}_{l,i} \right)}. \quad (31)$$

Then, $\hat{\mathcal{E}}_m^*(\theta; f, \theta')$ is an unbiased estimator of $\mathcal{E}_m^*(\theta; f, \theta')$ for all $\theta, \theta' \in \Theta$:

$$\mathbb{E}^* \left[\hat{\mathcal{E}}_m^*(\theta; f, \theta') \mid \mathbf{D}_m \right] = \mathcal{E}_m^*(\theta; f, \theta').$$

Proposition 6.2 states that the estimator $\hat{\mathcal{E}}_m^*(\theta; f, \theta')$ of the \mathbb{P}^* -filter $\mathcal{E}_m^*(\theta; f, \theta')$ is unbiased for any $\theta, \theta' \in \Theta$ and measurable function f . This holds because the \mathbb{P}^* -sample $(\hat{X}_{l,i})_{i=0,\dots,m}$ is independent of the random variate \mathbf{R} used to compute our density estimator. We exploit the unbiasedness of the estimator $\hat{\mathcal{E}}_m^*(\theta; f, \theta')$ to construct an estimator of the \mathbb{P}_θ -filter $\mathcal{E}_m(\theta; f, \theta')$ that converges at the canonical rate.

Proposition 6.3. *Assume that the conditions of Proposition 6.2 are satisfied. Suppose that the following condition also holds:*

(D2) *The function f is square-integrable. That is, for all $\theta, \theta' \in \Theta$,*

$$\mathbb{E}_\theta \left[f^2(X_0, X_{t_1}, \dots, X_{t_m}; \theta') \right] < \infty.$$

Fix $K \in \mathbb{N}$. For $i \in \{0, \dots, m\}$, let $(\hat{X}_{2,i}^{(k)})_{k=1,\dots,K}$ be a sequence of i.i.d. samples of X_{2,t_i} under the measure \mathbb{P}^* . In addition, let $(\mathbf{R}^{(k)})_{k=1,\dots,K}$ be a sequence of i.i.d. samples of \mathbf{R} as in (26). Define

$$\begin{aligned} w_k &= \frac{p_0 \left(\left(X_{o,0}, \hat{X}_{l,0}^{(k)} \right); \theta \right)}{\psi_2^* \left(\hat{X}_{l,0}^{(k)} \right)} \prod_{i=1}^m \frac{P \left(\left(X_{o,t_{i-1}}, \hat{X}_{l,i-1}^{(k)} \right), \left(X_{o,t_i}, \hat{X}_{l,i}^{(k)} \right); \theta, \mathbf{R}^{(k)} \right)}{\phi_2^* \left(\hat{X}_{l,i}^{(k)} \right)}, \\ W &= \sum_{k=1}^K w_k, \\ \hat{\mathcal{E}}_m^K(\theta; f, \theta') &= \sum_{k=1}^K f \left(\left(X_{o,0}, \hat{X}_{l,0}^{(k)} \right), \dots, \left(X_{o,t_m}, \hat{X}_{l,m}^{(k)} \right); \theta' \right) \frac{w_k}{W}, \end{aligned} \quad (32)$$

where P is the function yielding the Monte Carlo estimator \hat{p}_Δ^K as in equation (26). Then, $\hat{\mathcal{E}}_m^K(\theta; f, \theta')$ converges almost surely to $\mathcal{E}_m(\theta; f, \theta')$ at square-root rate:

$$\left\| \hat{\mathcal{E}}_m^K(\theta; f, \theta') - \mathcal{E}_m(\theta; f, \theta') \right\|_2 = O(K^{-0.5})$$

Proposition 6.3 derives a feasible estimator of the filter $\mathcal{E}_m(\theta; f', \theta)$ that converges at square-root rate. This is the same rate of convergence associated with exact Monte Carlo simulation, which is nonetheless infeasible for the general class of jump-diffusion models considered in this paper. The filter estimator (32) can be viewed as a self-normalized importance sampling estimator in the spirit of Geweke (1989). While our filter estimator is not unbiased, it is computationally efficient because it achieves the canonical rate of convergence. Alternative filter estimation methods, such as particle filters and Markov Chain Monte Carlo methods, generally converge at slower rates. We document the computational benefits of our filter estimator $\hat{\mathcal{E}}_m^K(\theta; f, \theta')$ in Section 8.

7 Parameter estimation

Proposition 6.1 implies that the likelihood satisfies $\mathcal{L}_m(\theta) \propto \mathcal{E}_m^*(\theta^*; f_m^*, \theta)$ for

$$f_m^*(x_0, x_1, \dots, x_m; \theta) = \frac{p_0(x_0; \theta)}{p_0(x_0; \theta^*)} \prod_{i=1}^m \frac{p_\Delta(x_{i-1}, x_i; \theta)}{p_\Delta(x_{i-1}, x_i; \theta^*)}.$$

An unbiased estimator of the likelihood according to Proposition 6.2 is therefore given by

$$\hat{\mathcal{E}}_m^*(\theta^*; f^*, \theta) = \frac{p_0\left(\left(X_{o,0}, \hat{X}_{l,0}\right); \theta\right)}{\psi_2^*\left(\hat{X}_{l,0}\right)} \prod_{i=1}^m \frac{\hat{p}_\Delta\left(\left(X_{o,t_{i-1}}, \hat{X}_{l,i-1}\right), \left(X_{o,t_i}, \hat{X}_{l,i}\right); \theta\right)}{\phi_2^*\left(\hat{X}_{l,i}\right)}. \quad (33)$$

Unlike the true likelihood, the likelihood estimator $\hat{\mathcal{E}}_m^*(\theta^*; f^*, \theta)$ is feasible because it does not involve the true parameter θ^* , and because our density estimator \hat{p}_Δ can be computed with finite computational resources. As a result, the likelihood estimator (33) can be maximized to obtain parameter estimators. We now study the asymptotic properties of the resulting parameter estimators.

Define the *simulated likelihood* $\hat{L}_m^K(\theta)$ as the average of K independent samples of $\hat{\mathcal{E}}_m^*(\theta^*; f^*, \theta)$ in (33). A *simulated maximum likelihood estimator* (SMLE) $\hat{\theta}_m^K$ is defined as an almost sure maximizer of the simulated likelihood:

$$\hat{\theta}_m^K \in \arg \sup_{\theta \in \Theta} \hat{L}_m^K(\theta) \quad \text{almost surely.}$$

In the below proposition, we derive conditions under which a SML $\hat{\theta}_m^K$ is asymptotically unbiased, consistent, asymptotically normal, and asymptotically efficient as $m \rightarrow \infty$.

Proposition 7.1. *Suppose that the conditions of Proposition 6.2 are satisfied, and assume that the following conditions also hold.*

(E1) *The conditional \mathbb{P}^* -variance of $\hat{\mathcal{E}}_m^*(\theta^*; f^*, \theta)$ is almost surely uniformly bounded:*

$$\sup_{\theta \in \Theta} \text{Var}^* \left(\hat{\mathcal{E}}_m^*(\theta^*; f^*, \theta) \mid \mathbf{D}_m \right) < \infty \quad \text{almost surely.}$$

Then a SMLE $\hat{\theta}_M^K$ is an asymptotically unbiased estimator of an MLE $\hat{\theta}_m$ as $K \rightarrow \infty$. That is, there exists an MLE $\hat{\theta}_m$ such that $\hat{\theta}_m^K \rightarrow \hat{\theta}_m$ almost surely as $K \rightarrow \infty$. Furthermore, if an MLE $\hat{\theta}_m$ is (strongly) consistent, then a SMLE $\hat{\theta}_m^K$ is also (strongly) consistent, and $\hat{\theta}_m^K \rightarrow \theta^$ in \mathbb{P}_{θ^*} -probability (almost surely) as $m \rightarrow \infty$ and $K \rightarrow \infty$. Suppose that, in addition, the following conditions also holds:*

(E2) *For any $K \geq 1$, the simulated likelihood \hat{L}_m^K is twice continuously differentiable.*

(E3) *The Hessian matrix $\frac{1}{m} \nabla^2 \log \hat{L}_m^K(\hat{\theta}_m^K)$ is almost surely negative definite for all sufficiently large m and K .*

If $\frac{m}{K} \rightarrow 0$ as $m \rightarrow \infty$, then a SMLE $\hat{\theta}_m^K$ is asymptotically normal and asymptotically efficient. That is, $\sqrt{m}(\hat{\theta}_m^K - \theta^) \rightarrow N(0, I(\theta^*)^{-1})$ as $m \rightarrow \infty$. Here, $I(\theta^*)$ is the Fisher information matrix given in Proposition 7.2.*

Proposition 7.1 tells us that our simulated likelihood estimator inherits the asymptotic properties of a true MLE if $\frac{m}{K} \rightarrow 0$ as $m \rightarrow \infty$. This holds even though we are estimating the likelihood via Monte Carlo simulation. The results of Proposition 7.1 imply that the noise introduced by the simulation-based estimation approach does not negatively impact the asymptotic properties of the resulting parameter estimators. This finding is remarkable because a similar result is not available for alternative estimators. A standard approach for parameter inference in the presence of partial data is the stochastic EM algorithm; see Dempster et al. (1977), Wei and Tanner (1990), and Celeux and Diebolt (1992). Nielsen (2000) shows the parameter estimator derived from the EM Algorithm is not asymptotically efficient because its asymptotic variance-covariance matrix includes a correction term due to the simulation part of the algorithm. Other simulation-based estimators that in theory can achieve asymptotic efficiency include those derived from the

efficient method of moments (Bansal et al. (1994), Bansal et al. (1995)), the simulated method of moments (Duffie and Singleton (1993)), and indirect inference (Gouriéroux et al. (1993), Smith (1993)). In order to achieve asymptotic efficiency, these methods require that one can simulate exact samples from the law of the underlying stochastic process. To the best of our knowledge, there are currently no exact simulation method that applies to the general class of multivariate jump-diffusions considered in this paper. As a result, asymptotic efficiency based on these alternative methods cannot be achieved in practical applications. Finally, particle filtering and MCMC methods are also popular; see Eraker et al. (2003), Jacquier et al. (1994, 2004), and Johannes and Polson (2010). However, due to their Bayesian structure, no consistency or asymptotic normality results are available for the implied parameter estimators.

Assumption (E1) is mild and imposes restrictions on the \mathbb{P}^* -densities ψ_l^* and ϕ_l^* . It essentially requires that these densities are bounded away from zero wherever the density estimator \hat{p}_Δ is positive. Assumptions similar to (E1) are standard in the importance sampling literature. Assumptions (E2) and (E3) are also standard.

An important computational feature of our parameter inference approach is that we can maximize the simulated likelihood $\hat{L}_m^K(\theta)$ using standard numerical optimization routines without the need to resimulate. To see why this is the case, note that our density estimator can always be written as an analytical function of a vector of \mathcal{F}_0 -measurable and parameter-independent random variates \mathbf{R} : $\hat{p}_\Delta(v, w; \theta) = P(v, w; \theta, \mathbf{R})$; see (26) and Appendix A. Our likelihood estimator can thus be evaluated as

$$\hat{L}_m^K(\theta) = \frac{1}{K} \sum_{k=1}^K \frac{p_0\left(\left(X_{o,0}, \hat{X}_{l,0}^{(k)}\right); \theta\right)}{\psi_2^*\left(\hat{X}_{l,0}^{(k)}\right)} \prod_{i=1}^m \frac{P\left(\left(X_{o,t_{i-1}}, \hat{X}_{l,i-1}^{(k)}\right), \left(X_{o,t_i}, \hat{X}_{l,i}^{(k)}\right); \theta, \mathbf{R}^{(k)}\right)}{\phi_2^*\left(\hat{X}_{l,i}^{(k)}\right)},$$

where $(\hat{X}_{l,i}^{(k)})_{k=1,\dots,K}$ are i.i.d. samples of X_{l,t_i} under \mathbb{P}^* , and $(\mathbf{R}^{(k)})_{k=1,\dots,K}$ are i.i.d. samples of \mathbf{R} . The distribution of these samples does not depend upon the parameter θ at which the likelihood is evaluated. As a result, the samples $(\hat{X}_{l,i}^{(k)})_{k=1,\dots,K}$ and $(\mathbf{R}^{(k)})_{k=1,\dots,K}$ need only be generated once, and can be reused throughout the numerical optimization of $\hat{L}_m^K(\theta)$. This feature generates significant computational benefits in practical applications. A numerical case study carried in Appendix 8 highlights the computational advantages of our simulated likelihood estimators, and shows that they perform better than several

alternative parameter estimators.

We finalize this section by deriving a tractable expression for the asymptotic variance-covariance matrix of our SMLE.

Proposition 7.2. *The Fisher information matrix satisfies*

$$I(\theta^*) = \mathbb{E}_{\theta^*} \left[\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^m \nabla \log \mathcal{R}_i^*(\theta^*)^\top \nabla \log \mathcal{R}_i^*(\theta^*) \right] = -\mathbb{E}_{\theta^*} \left[\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^m \nabla^2 \log \mathcal{R}_i^*(\theta^*) \right],$$

where

$$\mathcal{R}_i^*(\theta) = \frac{\mathcal{E}_i^*(\theta^*; f_i^*, \theta)}{\mathcal{E}_{i-1}^*(\theta^*; f_{i-1}^*, \theta)} \quad (34)$$

for $1 \leq i \leq m$, and $\mathcal{R}_0^*(\theta) = \mathbb{E}^* \left[\frac{p_0(X_0; \theta)}{\psi_l^*(X_{l,0})} \mid \mathbf{D}_0 \right]$.

Guided by Proposition 7.2, we estimate the Fisher information matrix driving the asymptotic variance-covariance matrix of our SMLE by taking derivatives of ratios of the likelihood estimator \hat{L}_M^K . Two alternative estimators of the Fisher information matrix $I(\theta^*)$ are given by

$$\hat{I}_1 = \frac{1}{m} \sum_{i=0}^m \nabla \log \frac{\hat{L}_i^K(\hat{\theta}_m^K)}{\hat{L}_{i-1}^K(\hat{\theta}_m^K)}^\top \nabla \log \frac{\hat{L}_i^K(\hat{\theta}_m^K)}{\hat{L}_{i-1}^K(\hat{\theta}_m^K)} \quad \text{and} \quad \hat{I}_2 = -\frac{1}{m} \sum_{i=0}^m \nabla^2 \log \frac{\hat{L}_i^K(\hat{\theta}_m^K)}{\hat{L}_{i-1}^K(\hat{\theta}_m^K)}.$$

The estimators \hat{I}_1 and \hat{I}_2 are tractable because our likelihood estimator \hat{L}_m^K is tractable. Our results therefore make it feasible to estimate asymptotic standard errors in empirical applications.

8 Numerical results

This section illustrates the behavior of our estimators in a numerical case study. We consider an affine bivariate consumption growth model. Even though our methodological results are particularly powerful for non-affine models, we focus here on the affine case because the transition density is known in semi-analytical form thanks to [Duffie et al. \(2000\)](#). This yields a close-to-error-free benchmark relative to which we can measure the accuracy of our estimators.

We specify the jump-diffusion X through the following choice of coefficient functions for $\theta = (a, b, k, v, \ell_0, m) \in \Theta$, $x = (x_1, x_2) \in \mathcal{S} = \mathbb{R}^2$, and $d \in \mathcal{D} = \mathbb{R}_+$:

$$\begin{aligned}\mu(x; \theta) &= \begin{pmatrix} a + bx_2 \\ -kx_2 \end{pmatrix}, & \Sigma(x; \theta) &= \begin{pmatrix} v & 0 \\ 0 & 1 \end{pmatrix} \\ \Gamma(x, d; \theta) &= \begin{pmatrix} -md \\ 0 \end{pmatrix}, & \Lambda(x; \theta) &= \ell_0.\end{aligned}$$

The SDE (1) can be rewritten as

$$d \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} a + bX_{2,t} \\ -kX_{2,t} \end{pmatrix} dt + \begin{pmatrix} v & 0 \\ 0 & 1 \end{pmatrix} dW_t + dL_t, \quad (35)$$

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 $(D_n)_{n \geq 1}$ are i.i.d. samples of the standard exponential distribution. We fix the parameter space $\Theta = [-0.25, 0.25] \times [0, 1] \times [0, 10] \times [10^{-8}, 1] \times [10^{-8}, 10] \times [0, 5]$. The true data-generating parameter is $\theta^* = (0.02, 0.005, 0.36, 0.005, 0.017, 0.29)$, and we take $X_{1,0} = 3$ and $X_{2,0} \sim N(0, 1)$ unless specified otherwise. We assume that the component X_1 is observable ($X_o = X_1$), and the component X_2 is latent ($X_l = X_2$) so that $d_o = d_l = 1$. Due to the presence of the latent factor X_2 and the jump process L_t , the conditional distribution of X_1 is not Gaussian. Instead, the marginal distribution of X_1 is asymmetric with a heavy left tail. These features yield an interesting case study for our methodology. For simplicity, we assume that one unit of time corresponds to one year.

The SDE (35) describes a consumption growth model that nests a disaster model as is Barro (2006), and a long-run risk model similar to that of Bansal and Yaron (2004). The component X_1 stands for log-consumption; its increments represent consumption growth. Jumps in log-consumption are negatively exponentially distributed with an average magnitude of -29% . The jump intensity is constant and equal to 0.017, which implies that a jump arrives on average every 59 years. Log-consumption jumps can be interpreted as disasters, and their magnitudes and frequency are consistent with the disaster estimates of Barro (2006), Barro and Jin (2011), and Wachter (2013). The component X_2 is latent and drives the growth rate (drift) of log-consumption. This component of the model fluctuates

around zero with slowly mean-reverting dynamics. Consequently, X_2 can be interpreted as a long-run risk factor in the spirit of [Bansal and Yaron \(2004\)](#).¹²

Given that model (35) is affine, the characteristic function of X is given in terms of solutions to standard Ricatti equations (see [Duffie et al. \(2000\)](#)).¹³ As a result, we can compute the transition density p_Δ semi-analytically via Fourier inversion. We implement Fourier inversion via numerical quadrature with 500 discretization points per dimension in $[-2000, 2000]^2$.

We specify the implementation of our density estimator. We fix the sequences $J_\xi = 2^\xi$, $I_\xi = \lceil 0.1 J_\xi^{1/2} \rceil$, $H_\xi = 9 + \lceil (J_\xi^{1+d/4})/1000 \rceil$, and $q_\xi = 2^{-\xi} \xi \log_2^2(1 + \xi)$. These choices satisfy the conditions of Propositions 5.1 and 5.2 that ensure that our density estimator has finite variance and small computational costs.

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 Linux Centos 6.6 operating system. All codes are available upon request.

8.1 Density estimator

We study the accuracy of the transition density estimator \hat{p}_Δ^K . Figure 1 shows three-dimensional surface plots of the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta^*)$ for $K \in \{1000, 10000\}$ and $\Delta = 1$, which corresponds to a yearly time horizon in empirical applications. Figure 2 shows the corresponding contour plots. We see that when K is small and only few Monte Carlo replications are used, our estimator may assign large probability mass to areas in which the true density has little mass. These imprecisions vanish as the number of Monte Carlo replications grows large.

In Figure 3, we project our density estimators onto the X_1 and X_2 axes. These projected densities are proportional conditional densities of X_1 and X_2 , respectively. We also include 95% confidence bands computed via bootstrap, as well as the corresponding projections for the true density p_Δ . Confirming Theorem 4.1, our density estimator \hat{p}_Δ^K is

¹²Continuous-time long-run risk models similar to ours are studied by [REFS?](#).

¹³We fix the jump intensity as a constant in order to be able to solve the Ricatti equations in closed form without numerical error.

unbiased. The confidence bands are centered around the true density, and the widths of the confidence bands shrink as K grows large. We see that our density estimator captures the skewed non-Gaussian distribution of log-consumption.

Next, we evaluate the error of our density estimator. We randomly pick 600 points $(v, w; \theta)$ uniformly in $[2.5, 3.3] \times [-2, 2] \times \Theta$, and evaluate our density estimator at the randomly selected points and parameters. We compute the root mean squared error (RMSE) of the density estimators across the 600 combinations of v , w , and θ , and track the time it takes to compute each estimator for all 600 (v, w, θ) . Figure 4 summarizes the results of this analysis for $\Delta = 1$. Confirming the convergence result of Proposition 5.2, we see that the error of our density estimator vanishes quickly as the number of Monte Carlo samples grows.

Figures 1 through 4 also show an alternative simulation-based density estimator, namely a bivariate Gaussian kernel estimator obtained from K i.i.d. samples of X_Δ generated via Euler discretization; see Giesecke et al. (2015) and Kristensen and Shin (2012). We focus on this alternative density estimator as it is applicable for the general class of multivariate jump-diffusion models we consider in this paper without major restrictions on the coefficient functions. We fix the number of discretization steps according to the square-root rule of Duffie and Glynn (1995). Due to the discretization and the kernel estimation procedure, the alternative density estimator is biased. This bias leads to confidence bands for the alternative density estimator that are not centered around the true density in Figure 3. The bias also results in a slower rate of convergence for the alternative density estimator in Figure 4. As a result, our density estimator can achieve small error levels with smaller computational resources than the alternative density estimator.

8.2 Filter estimator

We now evaluate the computational properties of our filter estimator. Proposition 6.3 tells us that the estimator $\hat{\mathcal{E}}_m^K(\theta; f, \theta')$ converges to $\mathcal{E}_m(\theta; f, \theta')$ almost surely. We verify this convergence result numerically. For this, we implement our filter estimator $\hat{\mathcal{E}}_m^K(\theta; f, \theta')$ for $f(x_0, x_1, \dots, x_m) = (0, 1)'x_1$ and $f(x_0, x_1, \dots, x_m) = ((0, 1)'x_1)^2$, which yield estimators of the posterior mean and posterior second moment of the latent long-run risk component

one period ahead. We evaluate our filter estimator for 100 randomly chosen parameters $\theta, \theta' \in \Theta$, as well as randomly chosen realizations of the data $\mathbf{D}_m = \{X_{1,0}, X_{1,\Delta}\}$. The results of this analysis are summarized in Figure 5. Consistent with our theoretical results, we find that the error of our filter estimator vanishes as the number of Monte Carlo samples grows.

In Figure 5, we also compare our filter estimator to two alternative Markov Chain Monte Carlo alternatives:

- A Gibbs sampler, which is a particle filter that estimates the joint probability of X via Fourier inversion of the characteristic function as in Duffie et al. (2000) (“Gibbs sampler”), and
- A particle filter which estimates the joint probability of X via Euler discretization as in Johannes et al. (2009) (“Particle filter + Euler”).

Note that only the estimator “Particle filter + Euler” is generally available for the same class of models that we consider in this paper, while the Gibbs sampler is restricted to the affine class of models.

Compared to the two filter estimation alternatives, our filter estimator achieves the highest rate of convergence in Figure 5. It outperforms the estimator “Particle filter + Euler” even for small computational budgets. The estimator “Particle filter + Fourier” achieves small errors even with small computational budget. However, the good performance of the estimator “Particle filter + Fourier” is to be expected given that it exploits the affine features of the analyzed model.

A Computational properties

We summarize some of the computational features of our density estimator, including how to implement it via algorithms, and how to specify the open Poisson rate ℓ .

A.1 Multilevel Monte Carlo

In order to construct a sample of \hat{p}_Δ 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 half as many Euler steps as 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 $O(2^{\Xi-1})$ 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.

A.2 Implementation

The evaluation of the Monte Carlo estimator $\hat{p}_\Delta^K(v, w; \theta)$ requires that we generate K independent samples of a vector $\mathbf{R} = (\Xi, \mathbf{P}, \mathbf{T}, \mathbf{D}, \mathbf{W}, \mathbf{U}, \mathbf{V})$ of random variables that include:

- $\Xi \sim (q_\xi)_{\xi \geq 0}$, where $(q_\xi)_{\xi \geq 0}$ is fixed as in [Proposition 5.2](#),
- $\mathbf{P} \sim \text{Poisson}(\ell\Delta)$, which is a sample of the jump count N_Δ under \mathbb{Q}_θ ,
- $\mathbf{T} = (T_n)_{n=1, \dots, \mathbf{P}}$, which is a sample of the jump times $(T_n)_{n \leq N_\Delta}$ under \mathbb{Q}_θ conditional on $N_\Delta = \mathbf{P}$,
- A collection $\mathbf{D} = (\mathbf{D}_n)_{n=1, \dots, \mathbf{P}}$ of i.i.d. samples of jump marks from the density π ,
- A collection $\mathbf{W} = (\mathbf{W}_{n,j})_{n=0, \dots, \mathbf{P}, j=1, \dots, J_\Xi}$ of i.i.d. samples from the d -dimensional standard normal distribution with $J_\Xi = O(2^\Xi)$ that are used to construct Brownian increments for the Euler discretization in [\(10\)](#)-[\(11\)](#),
- A collection $\mathbf{U} = (\mathbf{U}_{n,i,\nu})_{n=0, \dots, \mathbf{P}, i=1, \dots, I, \nu=1, \dots, H}$ of i.i.d. d -dimensional standard normal samples with $I = O(J_{\Xi-1}^{1/2})$ and $H = O(J_{\Xi-1}^{1+d/4})$ that are independent of \mathbf{W} , which are used to construct Brownian increments for the Euler discretization in [\(12\)](#) when $J = J_{\Xi-1}$, and

- Another collection $\mathbf{V} = (\mathbf{V}_{n,i,\nu})_{n=0,\dots,P, i=1,\dots,I, \nu=1,\dots,H}$ of i.i.d. d -dimensional standard normal samples with $I = O(J_{\Xi}^{1/2})$ and $H = O(J_{\Xi}^{1+d/4})$ that are independent of \mathbf{W} , which are used to construct Brownian increments for the Euler discretization in (12) when $J = J_{\Xi}$.

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}_{Δ}^K .

Algorithm A.1 (Sampling of $\hat{p}_{\Delta}(v, w; \theta)$). *Fix $J_{\xi} = O(2^{\xi})$. Let $v, w \in \mathcal{S}$, $\theta \in \Theta$, the Poisson rate $\ell > 0$, the exponent $\rho > 1$, and i.i.d. samples $\mathbf{R}^k = (\Xi^k, \mathbf{P}^k, \mathbf{T}^k, \mathbf{D}^k, \mathbf{W}^k, \mathbf{U}^k, \mathbf{V}^k)$ 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 $j = 1$, $j = O(J_{\Xi^k})$, and $j = O(J_{\Xi^{k-1}})$ Euler steps between consecutive jumps. Use the Euler increments \mathbf{W}^k , the jump times \mathbf{T}^k , and the jump marks \mathbf{D}^k , and assume $X_0 = v$.*
- (2) *Set $H = 1$ and $I = 1$. Set $\tilde{X}_0^{I,1} = \mathcal{X}^j$ for $j = 1$, and construct the Euler discretization $(\tilde{X}_i^{I,1})_{i=1,\dots,I}$ of \tilde{X} with I Euler steps in $[0, \Delta - \mathbf{T}_{\mathbf{P}^k}^k]$ by using the Euler increments \mathbf{V}^k . Evaluate the density estimator $\tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)$ in (17).*
- (3) *Set $\hat{p}^{(0)} = \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta) \mathcal{Z}^j$ for $j = 1$.*
- (4) *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 $[0, \Delta - \mathbf{T}_{\mathbf{P}^k}^k]$ by using the Euler increments \mathbf{V}^k . Evaluate the density estimator $\tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)$ in (17).*
- (5) *Set $\hat{p}^{(2)} = \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta) \mathcal{Z}^j$ for $j = O(J_{\Xi^k})$.*
- (6) *Set $H = O(J_{\Xi-1}^{1+d/4})$ 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 $[0, \Delta - \mathbf{T}_{\mathbf{P}^k}^k]$ by using the Euler increments \mathbf{U}^k . Evaluate $\tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta)$ as in (17).*
- (7) *Set $\hat{p}^{(1)} = \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^j, w; \theta) \mathcal{Z}^j$ for $j = O(J_{\Xi^{k-1}})$.*

(8) Update \hat{p}^K as

$$\hat{p}^K + \frac{1}{K} \left(\hat{p}^{(0)} + \frac{\hat{p}^{(2)} - \hat{p}^{(1)}}{q_{\Xi^k}} \right).$$

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 A.1 is simple. Steps (1), (2), and (3) require straightforward Euler discretization; Algorithms A.2 and A.3 provide guidance. Steps (2), (5) and (6) involve basic algebraic operations. We have implemented Algorithms A.1 through A.3 in R, and the codes are available online at <http://www.gustavo-schwenkler.com>.

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

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

(2) Fix

$$X_{n,1}^0 = X_{n,0}^0 + Jh_n\mu(X_{n,0}^0; \theta) + \sqrt{h_n}\Sigma(X_{n,0}^0; \theta) \sum_{i=1}^J W_{n,i}$$

$$Z_{n,1}^0 = Z_{n,0}^0 - Jh_n Z_{n,0}^0 (\Lambda(X_{n,0}^0; \theta) - \ell).$$

(3) For $j = 1, \dots, \frac{J}{2}$, fix:

$$X_{n,j}^1 = X_{n,j-1}^1 + 2h_n\mu(X_{n,j-1}^1; \theta) + \sqrt{h_n}\Sigma(X_{n,j-1}^1; \theta) (W_{n,2j-1} + W_{n,2j})$$

$$X_{n,2j-1}^2 = X_{n,2(j-1)}^2 + h_n\mu(X_{n,2(j-1)}^2; \theta) + \sqrt{h_n}\Sigma(X_{n,2(j-1)}^2; \theta) W_{n,2j-1}$$

$$X_{n,2j}^2 = X_{n,2j-1}^2 + h_n\mu(X_{n,2j-1}^2; \theta) + \sqrt{h_n}\Sigma(X_{n,2j-1}^2; \theta) W_{n,2j}$$

$$Z_{n,j}^1 = Z_{n,j-1}^1 - 2h_n Z_{n,j-1}^1 (\Lambda(X_{n,j-1}^1; \theta) - \ell),$$

$$Z_{n,2j-1}^2 = Z_{n,2(j-1)}^2 - h_n Z_{n,2(j-1)}^2 (\Lambda(X_{n,2(j-1)}^2; \theta) - \ell),$$

$$Z_{n,2j}^2 = Z_{n,2j-1}^2 - h_n Z_{n,2j-1}^2 (\Lambda(X_{n,2j-1}^2; \theta) - \ell).$$

(4) If $n < P$, set

$$\begin{aligned}
X_{n+1,0}^0 &= X_{n,1}^0 + \Gamma(X_{n,1}^0, D_n; \theta), \\
X_{n+1,0}^1 &= X_{n,\frac{J}{2}}^1 + \Gamma(X_{n,\frac{J}{2}}^1, D_n; \theta), \\
X_{n+1,0}^2 &= X_{n,J}^2 + \Gamma(X_{n,2J}^2, D_n; \theta), \\
Z_{n+1,0}^0 &= \frac{\Lambda(X_{n,1}^0; \theta)}{\ell} Z_{n,1}^0, \\
Z_{n+1,0}^1 &= \frac{\Lambda(X_{n,\frac{J}{2}}^1; \theta)}{\ell} Z_{n,\frac{J}{2}}^1, \\
Z_{n+1,0}^2 &= \frac{\Lambda(X_{n,2J}^2; \theta)}{\ell} Z_{n,2J}^2.
\end{aligned}$$

Return $(X_{P,0}^k, Z_{P,j}^k)_{k=0,1,2}$, which are samples of $(\mathcal{X}^j, \mathcal{Z}^j)$ with $j = 1$, $j = O(J_{\Xi-1})$, and $j = O(J_{\Xi})$, respectively.

Algorithm A.3 (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 $h = \frac{\Delta - T_P}{I}$.

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

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

(3) Return

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

Algorithms A.1 through A.3 showcase an important feature of the Monte Carlo estimator \hat{p}_{Δ}^K : It can be computed as a separable function of samples of the random element $R = (\Xi, P, T, D, W, U, V)$ and the arguments $(v, w; \theta)$ of the density. This property implies that we can use the same samples of R in order to evaluate $\hat{p}_{\Delta}^K(v, w; \theta)$ at different values of v , w , or θ . In other words, there is no need to re-simulate to evaluate the density estimator at different arguments. The decoupling of the random element R from the arguments of the density estimator is inspired by the work of Beskos et al. (2009). This

feature generates important computational advantages when using the Monte estimator \hat{p}_Δ^K for parameter inference; see Section 7.

Given that one can re-use samples of \mathbf{R} to evaluate the density estimator at many different arguments, one may be tempted to store all samples of \mathbf{R} in memory and re-call them each time Algorithm A.1 is run. Such an approach is computationally inefficient when the number of Monte Carlo replications is large because it exhausts the available memory capacity. We pursue a more efficient approach in that we only generate a sequence of independent seeds that will be used to generate independent samples of \mathbf{R} . By re-using the same seeds for each call of Algorithm A.1, we are effectively re-generating the same samples of \mathbf{R} each time. However, the repeated re-generation of the same samples of \mathbf{R} is less computationally demanding than storing and re-calling these samples, reducing the costs to evaluate of density estimator \hat{p}_Δ^K .

A.3 Choice of free parameters

We can indirectly improve the accuracy of the Monte Carlo estimator \hat{p}_Δ^K by reducing the variance of each sample of \hat{p}_Δ . We have two degrees of freedom to control for the variance of \hat{p}_Δ : The choice of the Poisson rate $\ell > 0$ that determines the number of jumps of X under \mathbb{Q}_θ , and the choice of the linear correlation coefficient ρ between samples of the collections \mathbf{U} and \mathbf{V} used in Algorithm A.1. Ideally, we would choose ℓ and ρ as to minimize the local variance of the density estimator; that is:

$$\min_{\ell > 0, \rho \in [-1, 1]} \text{Var}_\theta^\mathbb{Q}(\hat{p}_\Delta(v, w; \theta)). \quad (36)$$

Nevertheless, solving (36) may be computationally burdensome, especially if we were to repeatedly solve (36) for all arguments $(v, w; \theta)$ at which we wish to estimate the transition density. To circumvent this issue, we propose heuristics to fix ℓ and ρ .

We begin with a heuristic for fixing $\ell > 0$. The Poisson rate ℓ determines the measure \mathbb{Q}_θ under which we estimate the transition density of X . The only difference between the measure \mathbb{Q}_θ and the true measure \mathbb{P}_θ governing the dynamics of X is the jump intensity of X : Jumps arrive at a stochastic intensity $\Lambda(X_t; \theta)$ under \mathbb{P}_θ , while they arrive at the constant intensity ℓ under \mathbb{Q}_θ . The theory of importance sampling tells us that the further

away ℓ lies from $\Lambda(X_t; \theta)$, the higher the \mathbb{Q}_θ -variance of $\frac{1}{Z_\Delta}$ is. This implies that if ℓ and $\Lambda(X_t; \theta)$ are very different, the variance of the unbiased estimator \hat{p}_Δ will be large. Inspired by these insights we propose to fix $\ell = \Lambda(v; \theta)$ when evaluating $\hat{p}_\Delta(v, w; \theta)$.¹⁴

Next, we propose a heuristic for fixing the correlation ρ between the samples \mathbf{U} and \mathbf{V} . Because we employ a multi-level Monte Carlo approach that uses the same Brownian increments to construct the Euler discretized samples $(\mathcal{X}^{J_\Xi}, \mathcal{Z}^{J_\Xi})$ and $(\mathcal{X}^{J_{\Xi-1}}, \mathcal{Z}^{J_{\Xi-1}})$, we know that \mathcal{X}^{J_Ξ} and $\mathcal{X}^{J_{\Xi-1}}$ are positively correlated, and so are \mathcal{Z}^{J_Ξ} and $\mathcal{Z}^{J_{\Xi-1}}$. It therefore follows that $\hat{p}_\Delta^{H_\Xi, I_\Xi, J_\Xi}(v, w; \theta)$ and $\hat{p}_\Delta^{H_{\Xi-1}, I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)$ are positively correlated if $\mathbf{U} = \mathbf{V}$. We conclude from these observations that the variance of \mathcal{D}_Δ^Ξ ,

$$\begin{aligned} \text{Var}_\theta^\mathbb{Q}(\mathcal{D}_\Delta^\Xi(v, w; \theta)) &= \text{Var}_\theta^\mathbb{Q}\left(\hat{p}_\Delta^{H_\Xi, I_\Xi, J_\Xi}(v, w; \theta)\right) + \text{Var}_\theta^\mathbb{Q}\left(\hat{p}_\Delta^{H_{\Xi-1}, I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)\right) \\ &\quad - 2\text{Cov}_\theta^\mathbb{Q}\left(\hat{p}_\Delta^{H_\Xi, I_\Xi, J_\Xi}(v, w; \theta), \hat{p}_\Delta^{H_{\Xi-1}, I_{\Xi-1}, J_{\Xi-1}}(v, w; \theta)\right) \end{aligned}$$

is smaller if \mathbf{U} is perfectly correlated with \mathbf{V} than if \mathbf{U} and \mathbf{V} are uncorrelated or negatively correlated. Inspired by these observations, we propose to fix $\rho = 1$. Such a choice yields an unbiased density estimator with small variance.

B Proofs

Proof of Theorem 3.1. The proof follows along the lines of the proof of Theorem 3.1 of Giesecke and Schwenkler (2017b). Fix $\theta \in \Theta$. For random variables $A, B \in \mathcal{F}_\Delta$, we write for simplicity \mathbb{P}^A (\mathbb{Q}^A) for the unconditional \mathbb{P}_θ -law (\mathbb{Q}_θ -law) of A , and $\mathbb{P}^{A|B}$ ($\mathbb{Q}^{A|B}$) for the conditional \mathbb{P}_θ -law (\mathbb{Q}_θ -law) of A given B . We also write \mathbb{L} for the Lebesgue measure on $(\mathcal{S}, \sigma(\mathcal{S}))$. By construction, \mathbb{P}^{X_Δ} is absolutely continuous with respect to \mathbb{L} , and

$$\frac{d\mathbb{P}^{X_\Delta}}{d\mathbb{L}} = p_\Delta(X_0, X_\Delta; \theta).$$

Now, Assumption (A2) implies that $Z_\Delta(\theta)$ is a valid Radon-Nikodym derivative, and that the local \mathbb{P}_θ -martingale $(Z_t(\theta))_{t \leq \Delta}$ given by $Z_t(\theta) = \mathbb{E}_\theta[Z_\Delta(\theta) | \mathcal{F}_t]$ is a uniformly

¹⁴Note that because we are only storing the seeds needed to generate the random variates \mathbf{R} , we can allow the Poisson rate ℓ to depend on v and θ when evaluating $\hat{p}_\Delta(v, w; \theta)$. By re-using the same seed, we can ensure that repeated evaluations of $\hat{p}_\Delta(v, w; \theta)$ yield the same result.

integrable martingale. Thus, $\mathbb{E}_\theta[Z_\Delta(\theta)] = Z_0(\theta) = 1$, and \mathbb{Q}_θ is a well-defined probability measure on $(\Omega, \mathcal{F}_\Delta)$ that is equivalent to \mathbb{P}_θ . The law of iterated expectations implies that

$$\frac{d\mathbb{P}^{X_\Delta}}{d\mathbb{Q}^{X_\Delta}} = \mathbb{E}_\theta^\mathbb{Q} \left[\frac{1}{Z_\Delta(\theta)} \middle| X_\Delta \right], \quad (37)$$

where $\mathbb{E}_\theta^\mathbb{Q}$ represents the expectation operator under \mathbb{Q}_θ . A consequence of (37) is that the law \mathbb{Q}^{X_Δ} is also absolutely continuous with respect to \mathbb{L} since

$$p_\Delta(X_0, X_\Delta; \theta) = \frac{d\mathbb{P}^{X_\Delta}}{d\mathbb{L}} = \frac{d\mathbb{P}^{X_\Delta}}{d\mathbb{Q}^{X_\Delta}} \frac{d\mathbb{Q}^{X_\Delta}}{d\mathbb{L}} = \mathbb{E}_\theta^\mathbb{Q} \left[\frac{1}{Z_\Delta(\theta)} \middle| X_\Delta \right] \frac{d\mathbb{Q}^{X_\Delta}}{d\mathbb{L}}.$$

Hence, the transition density of X under \mathbb{Q}_θ exists:

$$\frac{d\mathbb{Q}_\theta^X}{d\mathbb{L}} = q_\Delta(X_0, X_\Delta; \theta) = \frac{p_\Delta(X_0, X_\Delta; \theta)}{\mathbb{E}_\theta[1/Z_\Delta(\theta) | X_\Delta]}. \quad (38)$$

It follows that $p_\Delta(v, w; \theta) = q_\Delta(v, w; \theta)g(v, w)$, where

$$g(v, w) = \mathbb{E}_\theta^\mathbb{Q} \left[\exp \left(\int_0^\Delta (\ell - \Lambda(X_s; \theta)) ds \right) \prod_{n=1}^{N_\Delta} \frac{\Lambda(X_{T_n-}; \theta)}{\ell} \middle| X_0 = v, X_\Delta = w \right]. \quad (39)$$

We simplify (39) using an iterative argument. Write

$$\Phi_t(\theta) = \exp \left(\int_0^t (\ell - \Lambda(X_s; \theta)) ds \right) \prod_{n=1}^{N_t} \frac{\Lambda(X_{T_n-}; \theta)}{\ell}$$

By the law of iterated expectation, the Assumption (A1), and since no jump occurs between times T_{N_Δ} and Δ , it follows that (neglecting the trivial conditioning on X_0)

$$\begin{aligned} g(X_0, X_\Delta) &= \mathbb{E}_\theta^\mathbb{Q} \left[\Phi_{T_{N_\Delta}}(\theta) \mathbb{E}_\theta^\mathbb{Q} \left[\exp \left(\int_{T_{N_\Delta}}^\Delta (\ell - \Lambda(X_s; \theta)) ds \right) \middle| \mathcal{F}_{T_{N_\Delta}-}, X_\Delta \right] \middle| X_\Delta \right] \\ &= \mathbb{E}_\theta^\mathbb{Q} \left[\Phi_{T_{N_\Delta}}(\theta) \mathbb{E}_\theta^\mathbb{Q} \left[\exp \left(\int_{T_{N_\Delta}}^\Delta (\ell - \Lambda(X_s; \theta)) ds \right) \middle| X_{T_{N_\Delta}}, X_\Delta \right] \middle| X_\Delta \right] \\ &= \mathbb{E}_\theta^\mathbb{Q} \left[\Phi_{T_{N_\Delta}}(\theta) \phi_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta) \middle| X_\Delta \right], \end{aligned}$$

where

$$\phi_t(v, w; \theta) = \mathbb{E}_\theta^\mathbb{Q} \left[e^{\int_0^t (\ell - \Lambda(X_s; \theta)) ds} \middle| X_0 = v, X_t = w \right]. \quad (40)$$

We iterate the above argument by conditioning on the σ -algebras \mathcal{F}_{T_n-} and $\sigma\{X_s : s \geq T_{n+1}\}$ for $n = N_\Delta - 1, \dots, 1$, and obtain

$$g(X_0, X_\Delta) = \mathbb{E}_\theta^\mathbb{Q} \left[\phi_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta) \prod_{n=1}^{N_\Delta} \phi_{T_n-T_{n-1}}(X_{T_{n-1}}, X_{T_n-}; \theta) \frac{\Lambda(X_{T_n-}; \theta)}{\ell} \middle| X_\Delta \right].$$

The above expectation is taken with respect to the law of $\mathcal{J} = (T_n, D_n, Y_{T_n-})_{n \leq N_\Delta}$ conditioned on X_Δ . We exploit Bayes' formula to simplify this expectation. The law of total probability states that $\mathbb{Q}^{\mathcal{J}|X_\Delta} \times \mathbb{Q}^{X_\Delta} = \mathbb{Q}^{X_\Delta|\mathcal{J}} \times \mathbb{Q}^{\mathcal{J}}$, yielding the following version of Bayes formula

$$\frac{d\mathbb{Q}^{\mathcal{J}|X_\Delta}}{d\mathbb{Q}^{\mathcal{J}}} = \frac{d\mathbb{Q}^{X_\Delta|\mathcal{J}}}{d\mathbb{Q}^{X_\Delta}}.$$

Now, we know that \mathbb{Q}^{X_Δ} is absolutely continuous with respect to the Lebesgue measure: $\frac{d\mathbb{Q}^{X_\Delta}}{d\mathbb{L}} = q_\Delta(X_0, X_\Delta; \theta)$. Furthermore, because the change of measure does not alter the drift of X , Assumption (A1) and the fact that no jump occurs during the time interval $(T_{N_\Delta}, \Delta]$ imply that the conditional law $\mathbb{Q}^{X_\Delta|\mathcal{J}}$ is also absolutely continuous with respect to the Lebesgue measure with $\frac{d\mathbb{Q}^{X_\Delta|\mathcal{J}}}{d\mathbb{L}} = \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta)$. Here, \tilde{p} is the transition density of the diffusive process \tilde{X} in (7), which exists because the transition density p of X exists. Consequently,

$$\frac{d\mathbb{Q}^{\mathcal{J}|X_\Delta}}{d\mathbb{Q}^{\mathcal{J}}} = \frac{d\mathbb{Q}^{X_\Delta|\mathcal{J}}/d\mathbb{L}}{d\mathbb{Q}^{X_\Delta}/d\mathbb{L}} = \frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta)}{q_\Delta(X_0, X_\Delta; \theta)}.$$

We can therefore change the measure in (39) from the conditional law $\mathbb{Q}^{\mathcal{J}|X_\Delta}$ to the unconditional law $\mathbb{Q}^{\mathcal{J}}$ using the Radon-Nikodym derivative $\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta)}{q_\Delta(X_0, X_\Delta; \theta)}$:

$$\begin{aligned} g(X_0, X_\Delta) &= \mathbb{E}^{\mathbb{Q}^{\mathcal{J}|X_\Delta}} \left[\phi_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta) \prod_{n=1}^{N_\Delta} \phi_{T_n-T_{n-1}}(X_{T_{n-1}}, X_{T_n-}; \theta) \frac{\Lambda(X_{T_n-}; \theta)}{\ell} \right] \\ &= \mathbb{E}^{\mathbb{Q}^{\mathcal{J}}} \left[\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta)}{q_\Delta(X_0, X_\Delta; \theta)} \phi_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta) \prod_{n=1}^{N_\Delta} \phi_{T_n-T_{n-1}}(X_{T_{n-1}}, X_{T_n-}; \theta) \frac{\Lambda(X_{T_n-}; \theta)}{\ell} \right] \\ &= \mathbb{E}_\theta \left[\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta)}{q_\Delta(X_0, X_\Delta; \theta)} \phi_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, X_\Delta; \theta) \prod_{n=1}^{N_\Delta} \phi_{T_n-T_{n-1}}(X_{T_{n-1}}, X_{T_n-}; \theta) \frac{\Lambda(X_{T_n-}; \theta)}{\ell} \right]. \end{aligned}$$

We can unwind the conditional expectations ϕ the same way they were introduced, resulting in

$$g(v, w) = \mathbb{E}_\theta^\mathbb{Q} \left[\frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{q_\Delta(v, w; \theta)} \frac{1}{Z_\Delta(\theta)} \middle| X_0 = v \right].$$

The claim follows given that $p_\Delta(v, w; \theta) = q_\Delta(v, w; \theta)g(v, w)$. \square

Proof of Proposition 3.2. We show that Conditions (A1)-(A4) of Glasserman (2003, Section 7.2.2) are satisfied. Recall that

$$X_t = X_0 + \int_0^t \mu(X_s; \theta) ds + \int_0^t \sigma(X_s; \theta) dB_s + \sum_{n=1}^{N_\Delta} \Gamma(X_{T_n}, D_n; \theta) \quad (41)$$

under \mathbb{Q}_θ . Assumption (A3) implies that X_t is pathwise differentiable, and Condition (A1) of Glasserman (2003, Section 7.2.2) holds. Assumption (A4) implies that Condition (A2) of Glasserman (2003, Section 7.2.2) is satisfied. The discussion on p. 395 of Glasserman (2003) implies that it suffices if Condition (A4) holds locally over bounded subsets. Assumption (A3) implies that $\theta \mapsto Z_\Delta(\theta)$ is differentiable. Furthermore, the pathwise differentiability of X_t together with Assumption (A4) imply that $\theta \mapsto \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)$ is also differentiable. Assumption (A5) implies Condition (A4) of Glasserman (2003, Section 7.2.2). Condition (A3) of Glasserman (2003, Section 7.2.2) is naturally satisfied. \square

Proof of Theorem 4.1. The claim follows from Theorem 1 of Rhee and Glynn (2015). \square

Proof of Proposition 4.2. We will again show that Conditions (A1)-(A4) of Glasserman (2003, Section 7.2.2) are satisfied. $\hat{p}_\Delta(v, w; \theta)$ is n -times continuously differentiable in θ if and only if the estimator $\hat{p}_\Delta^{I,J}(v, w; \theta)$ is n -times continuously differentiable in θ given that q_ξ is independent of θ . The estimator \tilde{p}^I is a mixture of Gaussian densities, and is therefore n -times continuously differentiable under Assumption (A3). Moreover, \mathcal{X}^J and \mathcal{Z}^J are deterministic function of θ , and n -times continuously differentiable when Assumption (A3) holds. As a result, Conditions (A1) and (A2) of Glasserman (2003, Section 7.2.2) are satisfied. Condition (A3) is naturally satisfied. Assumption (B2) implies Condition (A4) of Glasserman (2003, Section 7.2.2). \square

Proof of Proposition 5.1. Theorem 1 of Rhee and Glynn (2015) together with Assumption (B1) implies that the estimator \hat{p}_Δ has finite variance if $\|\hat{p}_\Delta^{H,I,J}(v, w; \theta) - p_\Delta(v, w; \theta)\|_2 = O(J^{-0.5})$ as $J \rightarrow \infty$. Temporarily omitting the subindex ξ , equation (19) and (20) imply:

$$\begin{aligned} & \left\| \frac{\tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta)}{Z_\Delta(\theta)} - \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^J, w; \theta) \mathcal{Z}^J \right\|_2 \\ & \leq \left\| \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta) \right\|_2 \times \|Z_\Delta^{-1}(\theta) - \mathcal{Z}^J\|_2 \\ & \quad + \|Z_\Delta^{-1}(\theta)\|_2 \times \left\| \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta) - \tilde{\mathcal{P}}^{H,I}(\mathcal{X}^J, w; \theta) \right\|_2 \\ & = \left\| \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta) \right\|_2 \times O(J_\xi^{-1/2}) + \|Z_\Delta^{-1}(\theta)\|_2 \times O(J_\xi^{-1/2}). \end{aligned}$$

Assumption (C1) and equation (21) imply that $\left\| \tilde{p}_{\Delta-T_{N_\Delta}}(X_{T_{N_\Delta}}, w; \theta) \right\|_2 < \infty$. Together with Assumption (C2) it follows that $\|\hat{p}_\Delta^{H,I,J}(v, w; \theta) - p_\Delta(v, w; \theta)\|_2 = O(J^{-0.5})$. \square

Proof of Proposition 5.2. Claim (1) follows from Theorem 4.1 after recognizing that the choices of J_ξ and q_ξ satisfy Assumption (B1). Claim (2) follows immediately from Proposition 5.1. Claim (3) holds because $\mathbb{Q}_\theta[\Xi < \infty] = 1$ so that $J_\Xi < \infty$ and $\text{Effort}(\hat{p}_\Delta^K(v, w; \theta)) < \infty$ almost surely. For Claim (4), note that any choice of J_n and q_n needs to be fixed such that the estimator \hat{p}_Δ is unbiased. For this, we need to satisfy Assumption (B1):

$$\sum_{n \geq 0} \frac{1}{J_n q_n} < \infty.$$

We therefore set $q_n = \frac{\bar{q}_n}{J_n}$ and pick $(\bar{q}_n)_{n \geq 0}$ as to satisfy

$$\sum_{n \geq 0} \frac{1}{\bar{q}_n} < \infty. \quad (42)$$

Furthermore, in order for the random variable Ξ to be well-defined, we also need $\sum_{n \geq 0} q_n < \infty$. Thus:

$$\sum_{n \geq 0} \frac{\bar{q}_n}{J_n} < \infty.$$

Now let τ be the computational effort to evaluate a sample of \hat{p}_Δ . To determine the computational effort to evaluate \hat{p}_Δ^K , we use a key result from Section 4 of Rhee and Glynn (2015) that tells us that $\mathbb{Q}_\theta[\text{Effort}(\hat{p}_\Delta^K(v, w; \theta)) > a_K] \rightarrow 0$ as $K \rightarrow \infty$ if

$$\sum_{n \geq 0} \mathbb{Q}_\theta[\tau \geq a_n] < \infty. \quad (43)$$

We proceed to construct a sequence $(a_n)_{n \geq 0}$ with smallest possible a_n 's as well as sequences $(J_n)_{n \geq 0}$ and $(q_n)_{n \geq 0}$ so that Condition (43) is satisfied.

Suppose that $J_n = f(n)$ for an increasing and invertible function f . Equation (25) implies

$$\mathbb{Q}_\theta[\tau \geq a_n] \sim \mathbb{Q}_\theta\left[J_\Xi^{\frac{6+d}{4}} \geq a_n\right] = \mathbb{Q}_\theta\left[\Xi \geq f^{-1}\left(a_n^{\frac{4}{6+d}}\right)\right] \sim q_{f^{-1}(a_n^{4/(6+d)})},$$

where “ \sim ” denotes equal in first order. If we set $a_n = n^p \log_2^r(n)$ for some $p, r > 0$, then

$$\begin{aligned} \sum_{n \geq 0} \mathbb{Q}_\theta[\tau \geq a_n] &\sim \sum_{n \geq 0} q_{f^{-1}(a_n^{4/(6+d)})} = \sum_{n \geq 0} \frac{\bar{q}_{f^{-1}(a_n^{4/(6+d)})}}{J_{f^{-1}(a_n^{4/(6+d)})}} = \sum_{n \geq 0} \frac{\bar{q}_{f^{-1}(a_n^{4/(6+d)})}}{a_n^{4/(6+d)}} \\ &\sim \sum_{n \geq 1} \frac{\bar{q}_{f^{-1}(a_n^{4/(6+d)})}}{n^{\frac{4p}{6+d}} \log_2^r(n)}. \end{aligned} \quad (44)$$

To satisfy (42), necessarily $\bar{q}_n \rightarrow \infty$ as $n \rightarrow \infty$. As a result, convergence of $\sum_{n \geq 0} \mathbb{Q}_\theta [\tau \geq a_n]$ necessarily requires that $p \geq \frac{6+d}{4}$ and $r > 1$. Because n^p is the dominating divergence term of a_n , we fix $p = \frac{6+d}{4}$ to ensure that a_n is as small as possible for large n . We now need to fix r as small as possible to ensure convergence of $\sum_{n \geq 0} \mathbb{Q}_\theta [\tau \geq a_n]$. For this, the numerator in (45) should be at most of logarithmic order. We can accomplish this by fixing $\bar{q}_n = n \log_2^2(n)$ and $f(n) = 2^n$ so that $J_n = 2^n$. These choices imply $\bar{q}_{f^{-1}(a_n^{4/(6+d)})} \sim \log_2(n) \log_2^2(\log_2(n))$ and

$$\sum_{n \geq 0} \mathbb{Q}_\theta [\tau \geq a_n] \sim \sum_{n \geq 1} \frac{\log_2^2(\log_2(n))}{n \log_2^{r-1}(n)}. \quad (45)$$

As a result, convergence can only be achieved with $r > 2$. It follows that, if $a_K = K^{\frac{6+d}{4}} \log_2^r(n)$ for $r > 2$, then $\mathbb{Q}_\theta [\text{Effort}(\hat{p}_\Delta^K(v, w; \theta)) > a_K] \rightarrow 0$ as $K \rightarrow \infty$. This implies that

$$\frac{\text{Effort}(\hat{p}_\Delta^K(v, w; \theta))}{K^{\frac{6+d}{4}} \log_2^r(K)} = O_P(1)$$

for all $r > 2$. The smoothness of the right-hand side, and the fact that \hat{p}_Δ^K does not depend on r , then imply that

$$\frac{\text{Effort}(\hat{p}_\Delta^K(v, w; \theta))}{K^{\frac{6+d}{4}} \log_2^2(K)} = O_P(1). \quad (46)$$

Now, in order to bound the error in (24) by $\epsilon > 0$, it is necessary to pick $K = O(\epsilon^{-2})$. Plugging this value of K in equation (46) yields the claim. \square

Proof of Proposition 6.1. Given that ψ_j^* and ϕ_j^* are strictly positive densities, the Radon-Nikodym density

$$\frac{d\mathbb{P}^*}{d\mathbb{P}_{\theta^*}} = \frac{\psi_o^*(X_{o,0})\psi_l^*(X_{l,0})}{p_0(X_0; \theta^*)} \prod_{i=1}^m \frac{\phi_o^*(X_{o,i-1\Delta}, X_{o,i\Delta})\phi_l^*(X_{l,(i-1)\Delta}, X_{l,i\Delta})}{p_\Delta(X_{(i-1)\Delta}, X_{i\Delta}; \theta^*)}.$$

induces a well-defined measure change. Given that $\mathbb{E}_\theta [|f(X_0, X_{t_1}, \dots, X_{t_m})|] < \infty$, standard measure change arguments yield the claim. \square

Proof of Proposition 6.2. The random variate \mathbf{R} used by Algorithm A.1 in Appendix A is \mathcal{F}_0 -measurable and independent of X . As a result, the distribution of \mathbf{R} under \mathbb{Q}_θ is the same as under \mathbb{P}^* . The sample $(\hat{X}_{2,i})_{0 \leq i \leq m}$ is by construction independent of \mathbf{R} . The claim then follows. \square

Proof of Proposition 6.3. Geweke (1989) implies that

$$\mathbb{E}_\theta \left[\left(\hat{\mathcal{E}}_m^K(\theta; f, \theta') - \mathcal{E}_m(\theta; f, \theta') \right)^2 \middle| \mathbf{D}_m \right] = O_P(K^{-1}).$$

Taking expectations on both sides then yields the claim in light of Assumption (D2). \square

Proof of Proposition 7.1. For simplicity, write $\mathcal{E}_m^*(\theta^*; f_m^*, \theta) = L_m(\theta)$. The unbiasedness of the \mathbb{P}^* -filter $\hat{\mathcal{E}}_m^*(\theta^*; f_m^*, \theta)$ stated in Proposition 6.2 together with Assumption (E1) imply that the simulated likelihood is almost surely bounded: $\sup_{\theta \in \Theta} \hat{L}_m^K(\theta) < \infty$ with probability one. The strong law of large numbers together with Assumption (E1) yield

$$\hat{L}_m^K(\theta) \rightarrow L_m(\theta) \quad \text{almost surely as } K \rightarrow \infty,$$

and that this convergence occurs uniformly in θ at the rate $O_P(K^{-0.5})$. The uniform convergence implies that for any fixed $\epsilon > 0$, $m \geq 1$, and data \mathbf{D}_m , there exists a sufficiently large $K \geq 1$ such that

$$\sup_{\theta \in \Theta} L_m(\theta) - \epsilon \leq \sup_{\theta \in \Theta} \hat{L}_m^K(\theta) \leq \sup_{\theta \in \Theta} L_m(\theta) + \epsilon$$

Letting $\epsilon \rightarrow 0$ implies that a SMLE $\hat{\theta}_m^K$ converges to an MLE $\hat{\theta}_m$ as $K \rightarrow \infty$ almost surely. This yields the asymptotic unbiasedness of a SMLE. Asymptotic unbiasedness implies consistency given that $\lim_{m \rightarrow \infty} \lim_{K \rightarrow \infty} \hat{\theta}_m^K = \lim_{m \rightarrow \infty} \hat{\theta}_m = \theta^*$. For asymptotic normality, note that we can decompose $\sqrt{m}(\hat{\theta}_m^K - \theta^*) = \sqrt{m}(\hat{\theta}_m^K - \hat{\theta}_m) + \sqrt{m}(\hat{\theta}_m - \theta^*)$. Assumption ?? tells us that $\sqrt{m}(\hat{\theta}_m - \theta^*) \rightarrow \mathcal{N}(0, I(\theta^*)^{-1})$ in \mathbb{P}_{θ^*} -distribution. A standard second-order Taylor expansion and Assumption (E2) imply that

$$\sqrt{m}(\hat{\theta}_m^K - \hat{\theta}_m) = \left(-\frac{1}{m} \nabla^2 \log \hat{L}_m^K(\tilde{\theta}) \right)^{-1} \left(\frac{1}{\sqrt{m}} \nabla \log \hat{L}_m^K(\hat{\theta}_m) - \frac{1}{\sqrt{m}} \nabla \log L_m(\hat{\theta}_m) \right)$$

for some $\tilde{\theta}$ in a neighborhood of $\hat{\theta}_m$. The uniform convergence of \hat{L}_m^K at the rate $O_P(K^{-0.5})$ implies that

$$\hat{L}_m^K(\hat{\theta}_m) - L_m(\hat{\theta}_m) = O_P(K^{-0.5}).$$

The asymptotic unbiasedness of $\hat{\theta}_m^K$ together with Assumption (E3) yield

$$\left\| \left(-\frac{1}{m} \nabla^2 \log \hat{L}_m^K(\tilde{\theta}) \right)^{-1} \right\| < \infty$$

for sufficiently large K . As a result,

$$\sqrt{m} \left(\hat{\theta}_m^K - \hat{\theta}_m \right) = O_P \left(\sqrt{\frac{m}{K}} \right).$$

It follows that if $\frac{m}{K} \rightarrow 0$ as $m \rightarrow \infty$, then $\sqrt{m} \left(\hat{\theta}_m^K - \hat{\theta}_m \right) \rightarrow 0$ in \mathbb{P}_{θ^*} -probability. Consequently, $\sqrt{m}(\hat{\theta}_m^K - \theta^*)$ has the same limiting distribution as $\sqrt{m}(\hat{\theta}_m - \theta^*)$. The asymptotic normality and asymptotic efficiency claims follow. \square

Proof of Proposition 7.2. For f_m^* as defined in (28), we have

$$\begin{aligned} \mathbb{E}_\theta [f_m^*(X_0, X_{t_1}, \dots, X_{t_m})] &= \mathbb{E}_{\theta^*} \left[\frac{p_0(X_0; \theta)}{p_0(X_0; \theta^*)} \prod_{i=1}^m \frac{p_\Delta(X_{i-1}, X_i; \theta)}{p_\Delta(X_{i-1}, X_i; \theta^*)} \right] \\ &= \int p_0(x_0; \theta) \prod_{i=1}^m p_\Delta(x_{i-1}, x_i; \theta) dx_m \dots dx_1 dx_0 = 1 < \infty. \end{aligned}$$

Thus, the conditions of Proposition 6.1 are satisfied for the function f_m^* in (28). Now, the likelihood function can be rewritten as

$$\mathcal{L}_m(\theta) = \prod_{i=0}^m \mathcal{R}_i(\theta)$$

for $\mathcal{R}_i(\theta) = \frac{\mathcal{L}_i(\theta)}{\mathcal{L}_{i-1}(\theta)}$ for $1 \leq i \leq m$, and $\mathcal{L}_{-1}(\theta) = 1$. As a result, an MLE $\hat{\theta}_m$ satisfies the first-order condition

$$0 = \frac{1}{m} \sum_{i=0}^m \nabla \log \mathcal{R}_i(\hat{\theta}_m).$$

Furthermore, for any $\theta \in \Theta$,

$$\begin{aligned} \mathbb{E}_\theta [\mathcal{R}_i(\theta) \mid \mathbf{D}_{i-1}] &= \frac{\mathbb{E}_\theta [\mathcal{L}_i(\theta) \mid \mathbf{D}_{i-1}]}{\mathcal{L}_{i-1}(\theta)} = \frac{\mathbb{E}_{\theta^*} \left[\frac{p_0(X_0; \theta)}{p_0(X_0; \theta^*)} \prod_{j=1}^i \frac{p_\Delta(X_{t_{j-1}}, X_{t_j}; \theta)}{p_\Delta(X_{t_{j-1}}, X_{t_j}; \theta^*)} \mid \mathbf{D}_{i-1} \right]}{\mathcal{L}_{i-1}(\theta)} \\ &= \frac{\mathbb{E}_{\theta^*} \left[\frac{p_0(X_0; \theta)}{p_0(X_0; \theta^*)} \prod_{j=1}^{i-1} \frac{p_\Delta(X_{t_{j-1}}, X_{t_j}; \theta)}{p_\Delta(X_{t_{j-1}}, X_{t_j}; \theta^*)} \mathbb{E}_{\theta^*} \left[\frac{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta)}{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta^*)} \mid X_{t_{i-1}} \right] \mid \mathbf{D}_{i-1} \right]}{\mathcal{L}_{i-1}(\theta)}. \end{aligned}$$

By construction, $\mathbb{E}_{\theta^*} \left[\frac{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta)}{p_\Delta(X_{t_{i-1}}, X_{t_i}; \theta^*)} \mid X_{t_{i-1}} \right] = 1$. Therefore, $\mathbb{E}_\theta [\mathcal{R}_i(\theta) \mid \mathbf{D}_{i-1}] = 1$, and the sequence $(\mathcal{R}_i(\theta))_{i \geq 0}$ can be interpreted as a density sequence. We thus find ourselves in the standard setting of maximum likelihood inference. In this context, the Fisher information matrix satisfies

$$I(\theta^*) = \mathbb{E}_{\theta^*} \left[\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^m \nabla \log \mathcal{R}_i(\theta^*)^\top \nabla \log \mathcal{R}_i(\theta^*) \right] = -\mathbb{E}_{\theta^*} \left[\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=0}^m \nabla^2 \log \mathcal{R}_i(\theta^*) \right]$$

In addition, the change of measure implies that $\mathcal{R}_i(\theta) \propto \mathcal{R}_i^*(\theta)$ for the $\mathcal{R}_i^*(\theta)$ defined in (34). Here, the proportionality constant does not depend on θ . As a result,

$$\nabla \log \mathcal{R}_i(\theta) = \nabla \log \mathcal{R}_i^*(\theta) \quad \text{and} \quad \nabla^2 \log \mathcal{R}_i(\theta) = \nabla^2 \log \mathcal{R}_i^*(\theta)$$

for all $\theta \in \Theta$. The claim follows. \square

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 and Jesper Lund (2002), ‘An empirical investigation of continuous-time equity return models’, *The Journal of Finance* **57**(3), 1239–1284.
- Bally, Vlad and 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.
- Bansal, Ravi and Amir Yaron (2004), ‘Risks for the long run: A potential resolution of asset pricing puzzles’, *The Journal of Finance* **59**(4), 1481–1509.
- Bansal, Ravi, A.Ronald Gallant, Robert Hussey and George Tauchen (1994), Computational aspects of nonparametric simulation estimation, in D. A.Belsley, ed., ‘Computational Techniques for Econometrics and Economic Analysis’, Springer, pp. 3–22.
- Bansal, Ravi, A.Ronald Gallant, Robert Hussey and George Tauchen (1995), ‘Nonparametric estimation of structural models for high-frequency currency market data’, *Journal of Econometrics* **66**(1-2), 251 – 287.
- Barro, Robert J. (2006), ‘Rare disasters and asset markets in the twentieth century’, *The Quarterly Journal of Economics* **121**(3), 823–866.

- Barro, Robert J. and Tao Jin (2011), ‘On the size distribution of macroeconomic disasters’, *Econometrica* **79**(5), 1567–1589.
- Beskos, Alex, Omiros Papaspiliopoulos and Gareth Roberts (2009), ‘Monte carlo maximum likelihood estimation for discretely observed diffusion processes’, *Annals of Statistics* **37**, 223–245.
- Bibby, Bo Martin and Michael Sørensen (1995), ‘Martingale estimation functions for discretely observed diffusion processes’, *Bernoulli* **1**(1-2), 17–39.
- Blanchet, J. and J. Ruf (2013), A Weak Convergence Criterion Constructing Changes of Measure. Working Paper.
- Carrasco, Marine, Mikhail Chernov, Jean-Pierre Florens and 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.
- Celeux, Gilles and Jean Diebolt (1992), ‘A stochastic approximation type em algorithm for the mixture problem’, *Stochastics and Stochastic Reports* **41**(1-2), 119–134.
- Chacko, George and Luis M. Viceira (2003), ‘Spectral gmm estimation of continuous-time processes’, *Journal of Econometrics* **116**(1-2), 259–292.
- Dacunha-Castelle, D. and D. Florens-Zmirou (1986), ‘Estimation of the coefficients of a diffusion from discrete observations’, *Stochastics* **19**(4), 263–284.
- Dempster, Arthur P., Nan M. Laird and Donald B. Rubin (1977), ‘Maximum likelihood from incomplete data via the EM Algorithm’, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38.
- Detemple, Jérôme, René Garcia and Marcel Rindisbacher (2006), ‘Asymptotic properties of monte carlo estimators of diffusion processes’, *Journal of Econometrics* **134**(1), 1–68.

- Douc, Randal and Catherine Matias (2001), ‘Asymptotics of the maximum likelihood estimator for general hidden markov models’, *Bernoulli* **7**(3), 381–420.
- Douc, Randal, Eric Moulines and Tobias Rydén (2004), ‘Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime’, *The Annals of Statistics* **32**(5), 2254–2304.
- Duffie, Darrell, Jun Pan and Kenneth Singleton (2000), ‘Transform analysis and asset pricing for affine jump-diffusions’, *Econometrica* **68**, 1343–1376.
- Duffie, Darrell and Kenneth J. Singleton (1993), ‘Simulated moments estimation of markov models of asset prices’, *Econometrica* **61**(4), 929–952.
- Duffie, Darrell and Peter Glynn (1995), ‘Efficient Monte Carlo estimation of security prices’, *Annals of Applied Probability* **4**(5), 897–905.
- Duffie, Darrell and 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 and Nicholas Polson (2003), ‘The impact of jumps in volatility and returns’, *Journal of Finance* **58**(3), 1269–1300.
- Filipović, Damir, Eberhard Mayerhofer and 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.
- Fuh, Cheng-Der (2006), ‘Efficient likelihood estimation in state space models’, *The Annals of Statistics* **34**(4), 2026–2068.
- Geweke, John (1989), ‘Bayesian inference in econometric models using monte carlo integration’, *Econometrica* **57**(6), 1317–1339.

- Giesecke, Kay and Dmitry Smelov (2013), ‘Exact sampling of jump-diffusions’, *Operations Research* . In Press.
- Giesecke, Kay, Gerald Teng and Yexiang Wei (2015), Numerical solution of jump-diffusion SDEs. Working Paper.
- Giesecke, Kay and Gustavo Schwenkler (2017*a*), ‘Filtered likelihood for point processes’, *Journal of Econometrics* . Forthcoming.
- Giesecke, Kay and Gustavo Schwenkler (2017*b*), 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.
- Gobet, Emmanuel and 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 and Markus Reiß (2004), ‘Nonparametric estimation of scalar diffusions based on low frequency data’, *The Annals of Statistics* **32**(5), 2223–2253.
- Gouriéroux, Christian, Alain Monfort and Eric Renault (1993), ‘Indirect inference’, *Journal of Applied Econometrics* **8**, S85–S118.
- Henry-Labordère, Pierre, Xiaolu Tan and 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.
- Jacod, Jean and Philip Protter (1998), ‘Asymptotic error distributions for the euler method for stochastic differential equations’, *The Annals of Probability* **26**(1), 267–307.

- Jacquier, Eric, Nicholas G. Polson and Peter E. Rossi (1994), ‘Bayesian analysis of stochastic volatility models’, *Journal of Business & Economic Statistics* **12**(4), 371–389.
- Jacquier, Eric, Nicholas G. Polson and Peter E. Rossi (2004), ‘Bayesian analysis of stochastic volatility models with fat-tails and correlated errors’, *Journal of Econometrics* **122**(1), 185 – 212.
- Jiang, George J. and 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. and John L. Knight (2010), ‘Ecf estimation of markov models where the transition density is unknown’, *Econometrics Journal* **13**(2), 245–270.
- Johannes, Michael and Nicholas Polson (2010), MCMC methods for continuous-time financial econometrics, *in* Y.Aït-Sahalia and L. P.Hansen, eds, ‘Handbook of Financial Econometrics: Applications’, Vol. 2, Elsevier, chapter 13, pp. 1 – 72.
- Johannes, Michael S., Nicholas G. Polson and Jonathan R. Stroud (2009), ‘Optimal filtering of jump diffusions: Extracting latent states from asset prices’, *Review of Financial Studies* **22**(7), 2759–2799.
- Kloeden, Peter E. and Eckhard Platten (1999), *Numerical Solutions of Stochastic Differential Equations*, Stochastic Modeling and Applied Probability, Springer.
- Komatsu, Takashi and 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.
- Kristensen, Dennis and Yongseok Shin (2012), ‘Estimation of dynamic models with non-parametric 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.

- Nielsen, Søren Feodor (2000), ‘The stochastic EM Algorithm: Estimation and asymptotic results’, *Bernoulli* **6**(3), 457–489.
- 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.
- Protter, Philip and Kazuhiro Shimbo (2008), No arbitrage and general semimartingales, in S. N. Ethier, J. Feng and R. H. Stockbridge, eds, ‘Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz’, Vol. 4 of *Collections*, Institute of Mathematical Statistics, pp. 267–283.
- Rhee, Chang-Han and Peter W. Glynn (2015), ‘Unbiased estimation with square root convergence for sde models’, *Operations Research* **63**(5), 1026–1043.
- 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.
- Wachter, Jessica A. (2013), ‘Can time-varying risk of rare disasters explain aggregate stock market activity?’, *The Journal of Finance* **68**(3), 987–1035.
- Wei, Greg and Martin Tanner (1990), ‘A Monte Carlo implementation of the EM Algorithm and the poor man’s data augmentation algorithms’, *Journal of the American Statistical Association* **85**(411), 699–704.

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.

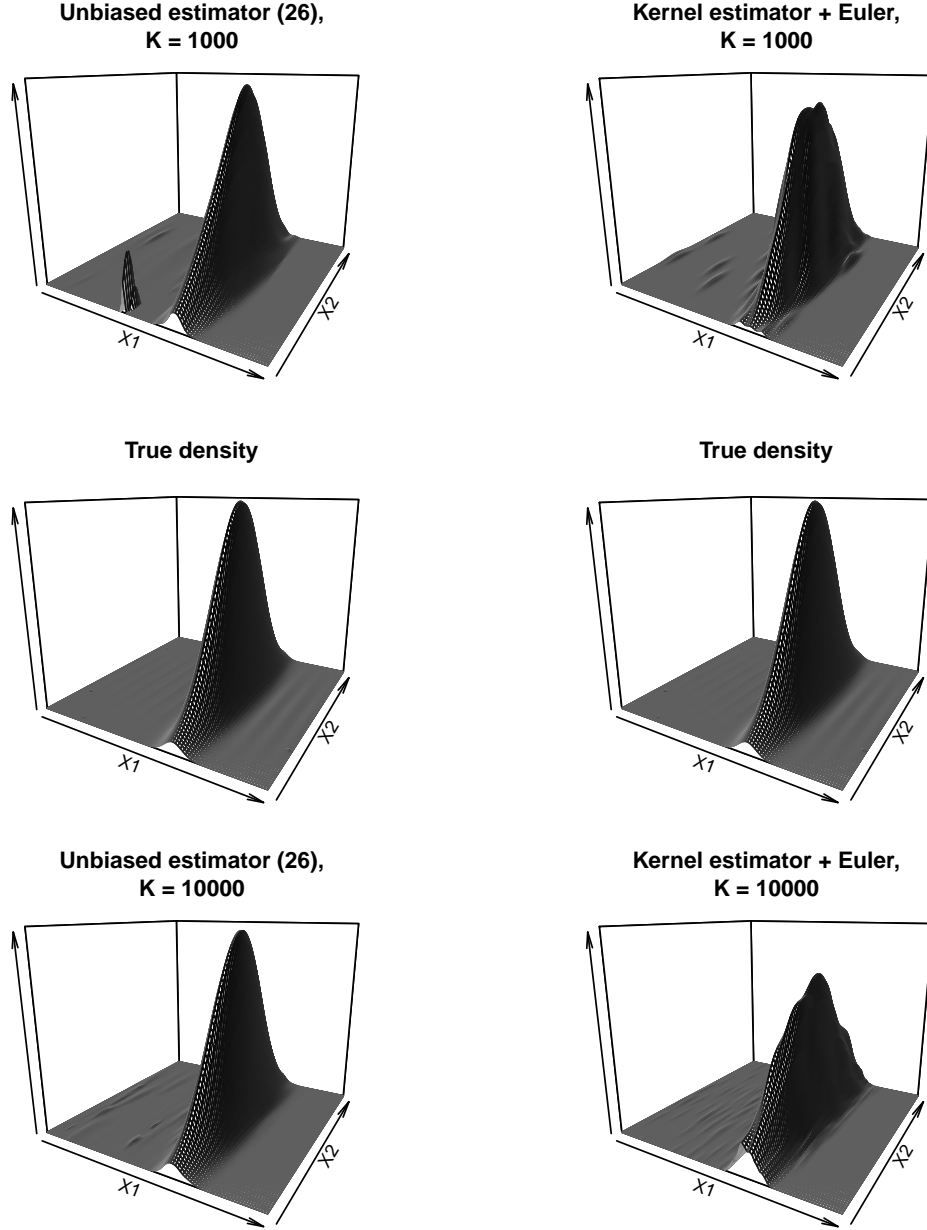


Figure 1: *Surface plots.* This figure shows the density $p_{\Delta}(v, w; \theta)$ and our unbiased estimator $\hat{p}_{\Delta}^K(v, x; \theta)$ for $v = (3, 0)$, $x = (x_1, x_2) \in [2.5, 3.3] \times [-2, 2]$, $K \in \{1000, 10000\}$, $\Delta = 1$, and $\theta = \theta^*$ (left column). It also shows surface plots for a Gaussian kernel density estimator based on K i.i.d. samples of X_{Δ} generated via Euler discretization (right column).

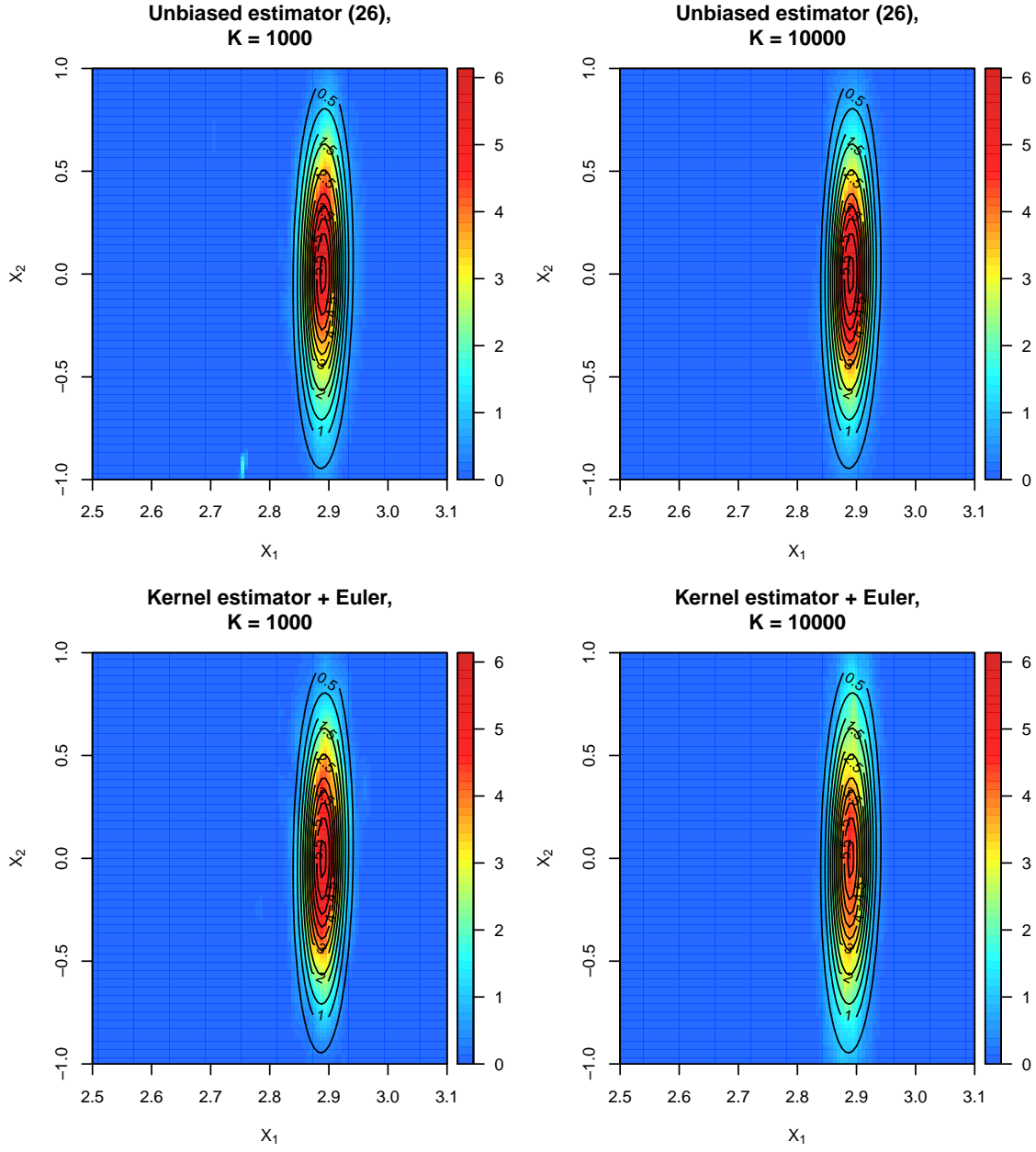


Figure 2: *Contour plots.* This figure shows a heat map of the unbiased Monte Carlo estimator $\hat{p}_\Delta^K(v, x; \theta)$ for $v = (3, 0)$, $x = (x_1, x_2) \in [2.5, 3.3] \times [-2, 2]$, $K \in \{1000, 10000\}$, $\Delta = 1$, and $\theta = \theta^*$ (top row). Overlaid is a contour plot of the true density $p_\Delta(v, w; \theta)$. The figure also shows heat maps for a Gaussian kernel density estimator based on K i.i.d. samples of X_Δ generated via Euler discretization (bottom row).

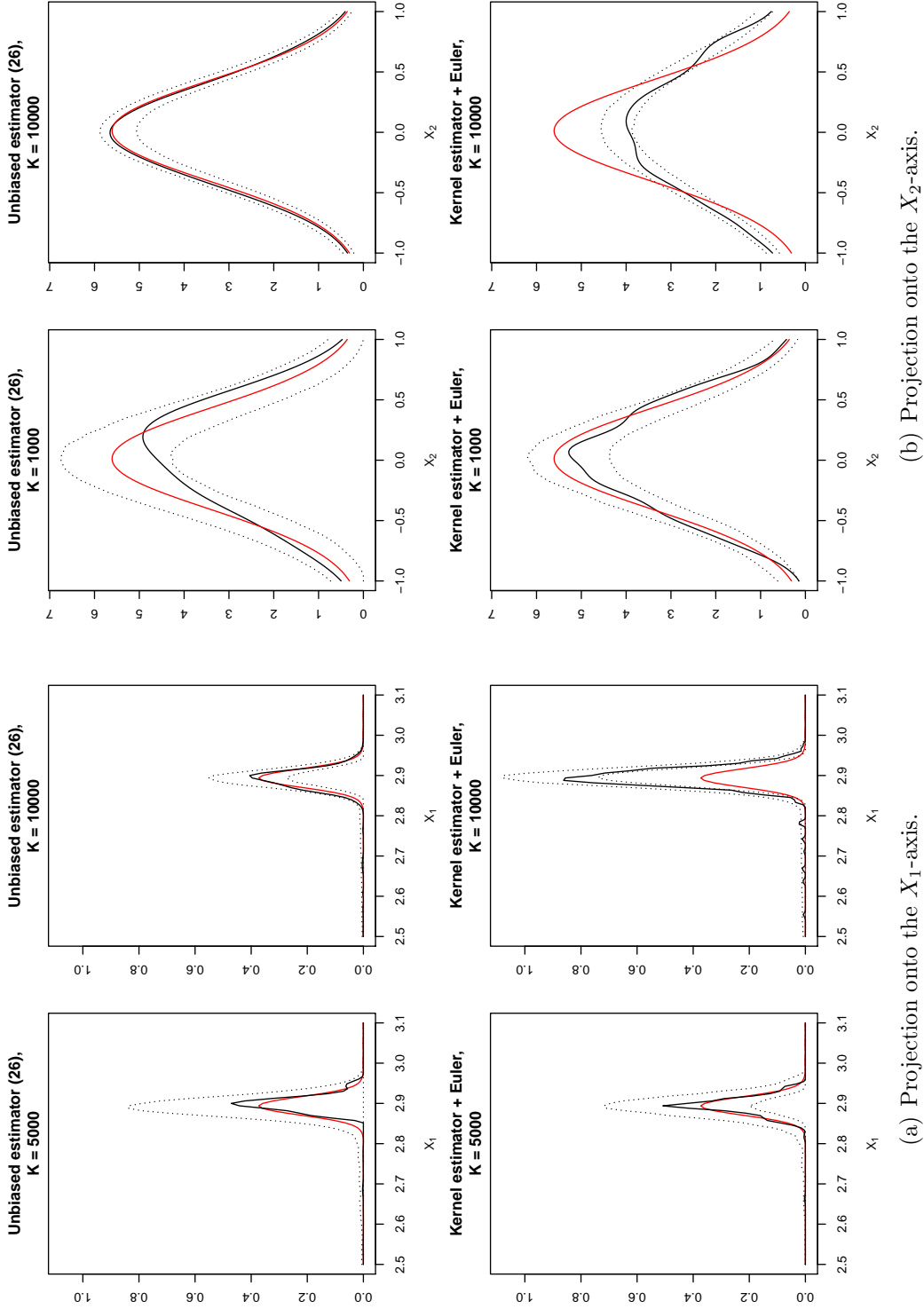


Figure 3: *Density projections.* The top rows of these figure show the projections of our unbiased density estimator $\hat{p}_\Delta^K(v, w; \theta)$ for $v = (3, 0)$, $\Delta = 1$, $\theta = \theta^*$, and $K \in \{1000, 10000\}$ (solid black lines). We take $w = (w_1, w_2)$, and fix $w_2 = 1$ for Panel (a), and $w_1 = 3$ for Panel (b). The dotted lines give 95% confidence bands computed via bootstrap with 1,000 bootstrap samples. The red lines show the true density p_Δ . The bottom row shows the analogous plots for Gaussian kernel density estimator derived from K i.i.d. samples of X_Δ generated via Euler discretization.

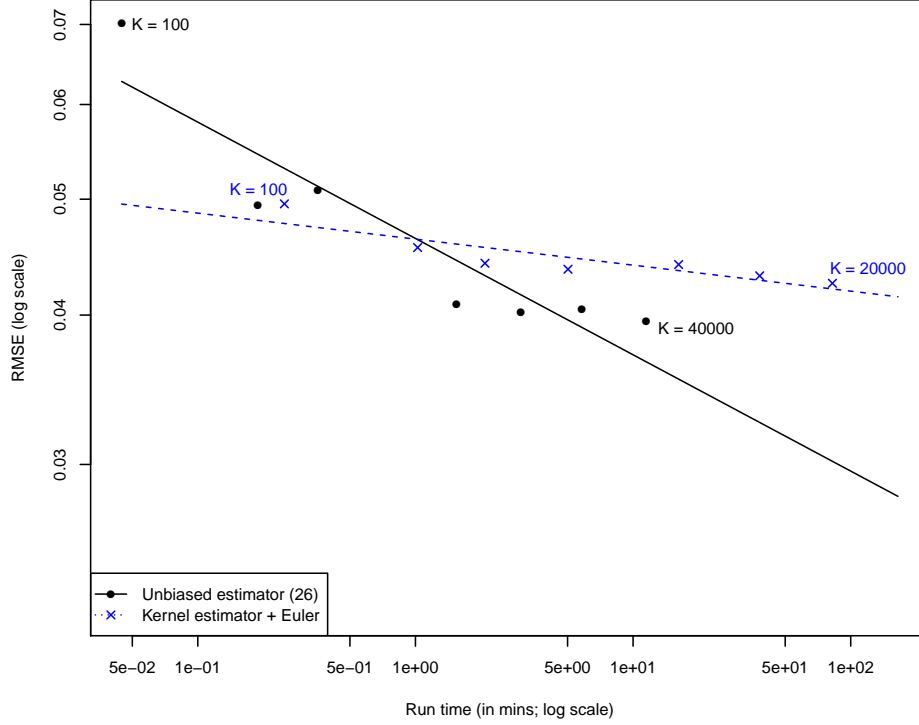
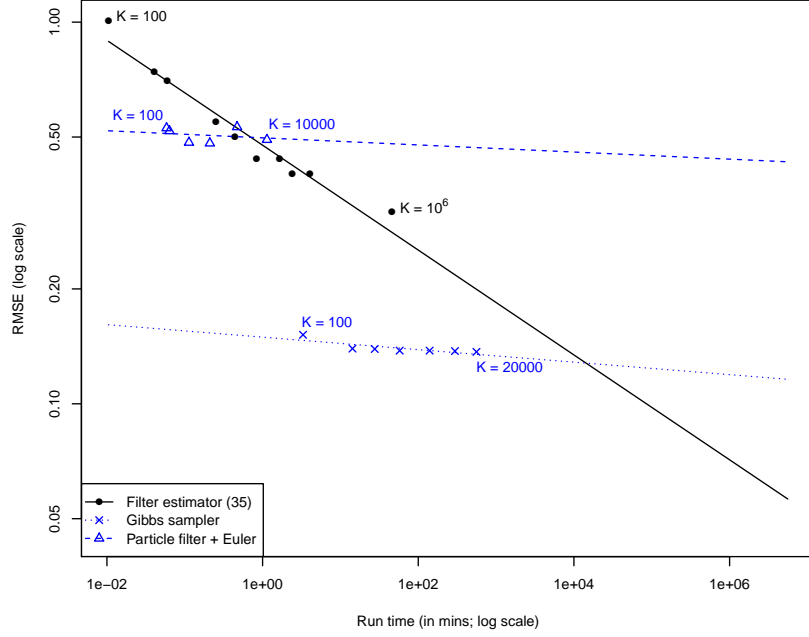
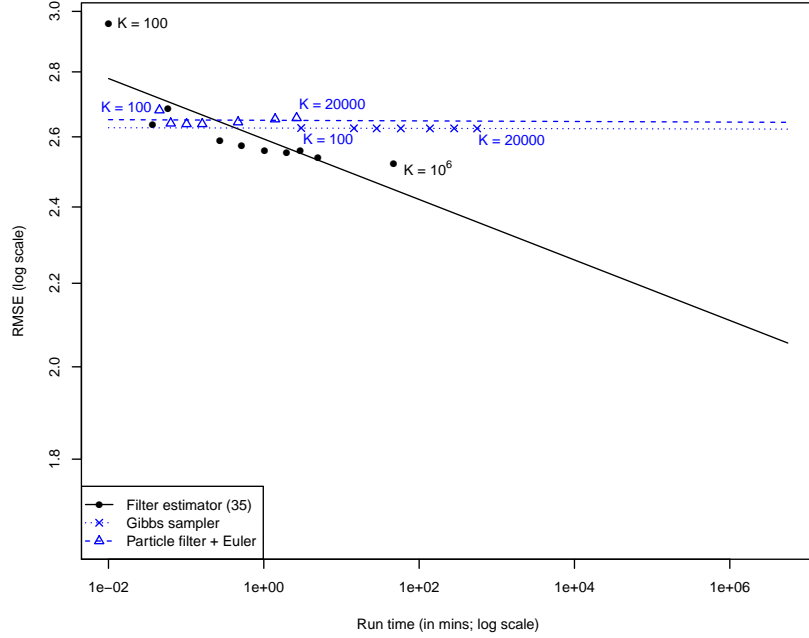


Figure 4: *Computational efficiency.* This figure plots the root mean squared error (RMSE) of the different estimators of the transition density $p_{\Delta}(v, w; \theta)$ against the time it takes to compute these estimators for 600 randomly chosen points v and w , and parameters θ . We take $\theta \in \Theta$, $v, w \in [2.5, 3.3] \times [-2, 2]$ for $\Delta = 1$. The points v, w and parameters θ at which we evaluate the density estimators are independently sampled at random from independent uniform distributions on the corresponding spaces.



(a) Posterior mean ($f(x_0, x_1, \dots, x_m) = (0, 1)'x_1$).



(b) Posterior second moment ($f(x_0, x_1, \dots, x_m) = ((0, 1)'x_1)^2$).

Figure 5: *Computational efficiency of filter estimators.* This figure plots the root mean squared error (RMSE) of different estimators of the filter $\mathcal{E}(\theta; f, \theta')$ against the time it takes to compute these estimators for 100 randomly chosen data sets D_m and parameters θ, θ' . We take $\theta, \theta' \in \Theta$ and $D_m = \{x_{1,0}, x_{1,\Delta}\}$ for $x_{1,0}, x_{1,\Delta} \in [2.5, 3.3]$ and $\Delta = 1$. The data sample D_m and parameters θ, θ' at which we evaluate the filter estimators are independently sampled at random from a uniform distribution on the corresponding spaces.