Probability in the Engineering and Informational Sciences, **35**, 2021, 975–1004.
doi:10.1017/S0269964820000315

METHOD OF MOMENTS ESTIMATION FOR LÉVY-DRIVEN ORNSTEIN–UHLENBECK STOCHASTIC VOLATILITY MODELS

Xiangyu Yang
School of Management, Fudan University, Shanghai, China
E-mail: xyyang19@fudan.edu.cn

Yanfeng Wu
Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China
E-mail: wuyanfeng@jxufe.edu.cn

Zeyu Zheng
Department of Industrial Engineering and Operations Research, University of California Berkeley, Berkeley, CA, USA
E-mail: zyzheng@berkeley.edu

Jian-Qiang Hu
School of Management, Fudan University, Shanghai, China
E-mail: hujq@fudan.edu.cn

This paper studies the parameter estimation for Ornstein–Uhlenbeck stochastic volatility models driven by Lévy processes. We propose computationally efficient estimators based on the method of moments that are robust to model misspecification. We develop an analytical framework that enables closed-form representation of model parameters in terms of the moments and autocorrelations of observed underlying processes. Under moderate assumptions, which are typically much weaker than those for likelihood methods, we prove large-sample behaviors for our proposed estimators, including strong consistency and asymptotic normality. Our estimators obtain the canonical square-root convergence rate and are shown through numerical experiments to outperform likelihood-based methods.

Keywords: consistency and asymptotic normality, method of moments, parameter estimation, stochastic volatility model

1. INTRODUCTION

Volatility plays a central role in the pricing of derivative securities according to Ghysels *et al.* [14]. The Black–Scholes European options pricing model, despite being the most prevalent choice that offers symbolic indicators such as implied volatility, restrictively assumes constant volatility that may not be practically sufficient. Empirical evidence from financial markets, such as volatility clustering, the dependence between increments, and volatility smiles, indicates that the assumption of constant volatility is inappropriate (see [10]); thus, adequately quantifying volatility is critical to capture these important features observed in real markets. Stochastic volatility (SV) models are developed for this purpose and capture the time-varying volatility in financial markets. The early SV models include discrete-time models (see [28]) and diffusion-based continuous-time SV models (e.g., [19]) which do not incorporate jumps that are widely observable in financial time series nor can they model realistic short-term implied volatility patterns. The second-generation continuous-time SV models add jumps, for example, the BNS model proposed by Barndorff-Nielsen and Shephard [3] where the volatility behaves according to an Ornstein–Uhlenbeck (OU) process, driven by a positive Lévy process without Gaussian component. However, despite practical relevance, the parameter estimation for continuous-time SV models has presented a formidable computational challenge for practical implementation. Two major categories of estimation methods are moment-based inference and likelihood-based inference. This paper presents a new moment-based approach to solve the seemly complicated problem by succinct methods.

Traditionally, for parameter estimation of SV models, the focus has largely been likelihood-based inference. There are two major likelihood-based methods known as Markov chain Monte Carlo (MCMC) and maximum likelihood estimation (MLE). MCMC requires assuming a prior distribution of the parameters to be solved, and then, iteratively sampling variables and parameters in a Bayesian framework until the Markov chain converges (e.g., see [12,15,16,21,25]). This approach usually suffers from slow convergence to the Markov chain equilibrium distribution and may become computationally prohibitive. In contrast to MCMC, the MLE approach takes a frequentist view and avoids the prior dependence between variables and parameters. However, traditional MLE needs the analytical form of state transition density function which is often intractable in SV models. Even worse, the MLE approach in many situations leads to non-convex optimization problems that are challenging to solve. Therefore, various approximation methods have been proposed, among which two prevail schemes are quasi-maximum likelihood estimation (QMLE) and simulated maximum likelihood estimation (SMLE). Harvey *et al.* [17] propose a quasi-maximum likelihood approach relying on transforming the model into a state-space form. Ruiz [26] introduces a linear system constructed by treating the logarithm of volatility as a hidden variable and analyzes sample properties of a QMLE estimator based on the Kalman filter. SMLE introduces additional hidden states to be simulated between two observable states to reduce the bias in estimating the likelihood Durham and Gallant [9]. Recently, Peng *et al.* [24] propose a gradient-based simulated MLE method using characteristic functions to estimate parameters in general Lévy-driven SV models.

However, these MLE methods are still computationally intensive and may require excessive simulation efforts; also too many assumptions that may be hard to satisfy. Financial markets give rise to numerous real-time decision-making instances that demand computationally fast and robust estimators that require fewer samplings. Method of moments (MM) is a case in point, but not much is known. Few important contributions include the simulated methods of moments (SMM), the generalized method of moments (GMM), and the efficient methods of moments (EMM). SMM is, in fact, a special case of indirect inference estimation (IIE) that is characterized by the use of an auxiliary model to capture aspects of the data on which the estimates are based. The basic idea is to generate simulated series from the "true" models and then estimate parameters by matching simulated data moments with actual data moments numerically. Duffie and Singleton [8] provide a consistent estimator of parameters of a dynamic system in which the state vector follows a time-homogeneous

Markov process. GMM can be applied when we have more sample moment conditions than the number of parameters by choosing the estimates that minimize a certain norm of the sample averages of the moment conditions. GMM estimation of an SV model is given by Andersen *et al.* [1] and Bregantini [5]. EMM, introduced by Bansal *et al.* [2] and developed in [13], is a variant of SMM, matching the efficiency of the MLE with the flexibility of the GMM procedure, see [1] for estimating and testing the SV model by EMM. Generally, the major issue of the moment-based inference is its statistically inefficiency, that is, the higher-order moments are used, the greater potential for estimation biases to occur. For instance, see [27], even if merely estimating the Lévy-driven OU process, that is, the latent volatility process in BNS models, the moment-based method confirms the existence of this phenomenon when using higher-order moments; nevertheless, we can ameliorate the estimation effect through some improvements in computational techniques (e.g., see [29]).

In this paper, we develop the first tractable MM parameter estimators for one type of the BNS model. To this end, based on tractable properties of the BNS model, we first deduce the analytic form of the moments needed for MM and propose an efficient method for estimating the compensation rate parameter λ without relying on any information other than the observed data. (Note that in this paper, expectation, variance and autocovariance are collectively called "moments" in a broad sense.) Next, we provide a closed-form representation of model parameters in terms of moments. We combine these analytical representations and numerical computation to further enhance our MM estimators. Specifically, we reduce the use of one dimension of the simultaneous equations that significantly improves the performance of estimation. Our approach manages to mitigate the drawback of MM that a single high-order moment presented in the estimator may significantly reduce estimation efficiency.

Under moderate conditions, we prove large-sample behaviors for our MM estimators, including strong consistency and asymptotic normality. In addition, we have compared MM with likelihood-based methods, including MCMC and MLE numerically. We can see the strengths and weaknesses of MM, and some empirical suggestions for subsequent applications of this method are provided. As a classical parameter estimation method, the greatest advantage of MM is its simplicity and computational efficiency. We believe our method has the capacity to outperform most previous iterative algorithms in the aspect of computational speed. In the previous literature, MM is hardly used to estimate the parameters of SV models, let alone on the BNS model while our work fills the gap.

The rest of this paper is organized as follows. In Section 2, we provide some preliminary results and discuss the parameter estimation procedure of the SV model in Section 3. In Section 4, we prove the strong consistency and asymptotic normality for our estimators. In Section 5, the comparison between MM and likelihood-based methods is provided, and we present extensive numerical experiment results of MM. Section 6 concludes this paper.

2. PRELIMINARIES

2.1. SV Model

We consider the following SV model first proposed by Barndorff-Nielsen and Shephard [3]:

$$dr(t) = (\mu + \beta v(t))\,dt + \sqrt{v(t)}\,dw(t), \tag{1}$$

$$dv(t) = -\lambda v(t)\,dt + dz(\lambda t), \tag{2}$$

where $r(t)$ is the log price of an asset at time t, $v(t)$ is the spot volatility at time t, $w(t)$ is a standard Brownian motion, $z(t)$ is a compound Poisson process with arrival rate a and

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

jump size distribution $\mathcal{J}(\cdot, b)$ (where b is a parameter in the distribution), μ is the drift, and β is the volatility risk premium coefficient. In financial markets, the discrete-time series of the log of asset prices $r(t)$ can be observed, but SV $v(t)$ is unobservable. We want to estimate parameters in Eqs. (1) and (2) from the discretely observable price series.

Suppose the asset price is observed at $t = n\Delta$ $(n = 0, 1, \ldots, N)$, where Δ is the time unit. Define

$$\begin{aligned}
v_n &\triangleq v(n\Delta), \\
z_n &\triangleq z(\lambda n\Delta), \\
\Delta v_n &\triangleq \int_{(n-1)\Delta}^{n\Delta} dv(s) = v_n - v_{n-1}, \\
\Delta z_n &\triangleq \int_{(n-1)\Delta}^{n\Delta} dz(\lambda s) = z_n - z_{n-1}, \\
\Delta e z_n &\triangleq \int_{(n-1)\Delta}^{n\Delta} e^{\lambda(s-n\Delta)}\, dz(\lambda s), \\
y_n &\triangleq \int_{(n-1)\Delta}^{n\Delta} dr(t), \\
q_n &\triangleq \int_{(n-1)\Delta}^{n\Delta} v(t)\, dt,
\end{aligned}$$

for $n = 1, \ldots, N$. It is clear that $\{(\Delta z_n, \Delta e z_n),\ n = 1, \ldots, N\}$ are i.i.d. (independent and identically distributed); hence, we sometimes use a generic $(\Delta z, \Delta e z)$ to represent $(\Delta z_n, \Delta e z_n)$ when there is no confusion. Note that the sequence $\{y_n\}$ represents the aggregate returns over intervals of length Δ. In the framework of MM, we take $\{y_n\}$ as observable data to estimate the parameters in Eqs. (1) and (2).

2.2. Moments Equations

Based on Eqs. (1) and (2), we have

$$y_n = \mu\Delta + \beta q_n + \sqrt{q_n}\epsilon_n, \tag{3}$$

$$q_n = \frac{1}{\lambda}[\Delta z_n - \Delta v_n], \tag{4}$$

$$v_n = e^{-\lambda\Delta} v_{n-1} + \Delta e z_n, \tag{5}$$

where $\{\epsilon_n,\ n = 1, \ldots, N\}$ are i.i.d. standard normal variables. Based on Eqs. (3)–(5), we can obtain

$$\mathsf{E}[y_n] = \mu\Delta + \beta\mathsf{E}[q_n], \tag{6}$$

$$\operatorname{var}[y_n] = \beta^2 \operatorname{var}[q_n] + \mathsf{E}[q_n], \tag{7}$$

$$\operatorname{cov}(y_n, y_{n+1}) = \beta^2 \operatorname{cov}(q_n, q_{n+1}), \tag{8}$$

$$\operatorname{cov}(y_n^2, y_{n+1}) = \beta^3 \operatorname{cov}(q_n^2, q_{n+1}) + (\beta + 2\mu\Delta\beta^2) \operatorname{cov}(q_n, q_{n+1}), \tag{9}$$

$$\operatorname{cov}(y_n, y_{n+1}^2) = \beta^3 \operatorname{cov}(q_n, q_{n+1}^2) + (\beta + 2\mu\Delta\beta^2)\operatorname{cov}(q_n, q_{n+1}). \quad \textbf{(10)}$$

Eqs. (6) and (7) follow directly from Eq. (3) (noting that q_n and ϵ_n are independent of each other). To derive Eqs. (8)–(10), we first point out that if $A_1, A_2, B,$ and C are four random variables, and (A_1, A_2), B, and C are independent of each other, then

$$\operatorname{cov}(BA_1, CA_2) = \mathsf{E}[B]\mathsf{E}[C]\operatorname{cov}(A_1, A_2).$$

Therefore, we have

$$\begin{aligned}
\operatorname{cov}(y_n, y_{n+1}) &= \operatorname{cov}(\mu\Delta + \beta q_n + \sqrt{q_n}\epsilon_n, \mu\Delta + \beta q_{n+1} + \sqrt{q_{n+1}}\epsilon_{n+1}) \\
&= \beta^2 \operatorname{cov}(q_n, q_{n+1}) + \beta \operatorname{cov}(q_n, \sqrt{q_{n+1}}\epsilon_{n+1}) \\
&\quad + \beta \operatorname{cov}(\sqrt{q_n}\epsilon_n, q_{n+1}) + \operatorname{cov}(\sqrt{q_n}\epsilon_n, \sqrt{q_{n+1}}\epsilon_{n+1}) \\
&= \beta^2 \operatorname{cov}(q_n, q_{n+1}), \\
\operatorname{cov}(y_n^2, y_{n+1}) &= \beta^3 \operatorname{cov}(q_n^2, q_{n+1}) + \beta \operatorname{cov}(q_n\epsilon_n^2, q_{n+1}) + 2\mu\Delta\beta^2 \operatorname{cov}(q_n, q_{n+1}) \\
&= \beta^3 \operatorname{cov}(q_n^2, q_{n+1}) + (\beta + 2\mu\Delta\beta^2)\operatorname{cov}(q_n, q_{n+1}), \\
\operatorname{cov}(y_n, y_{n+1}^2) &= \beta^3 \operatorname{cov}(q_n, q_{n+1}^2) + \beta \operatorname{cov}(q_n, q_{n+1}\epsilon_{n+1}^2) + 2\mu\Delta\beta^2 \operatorname{cov}(q_n, q_{n+1}) \\
&= \beta^3 \operatorname{cov}(q_n, q_{n+1}^2) + (\beta + 2\mu\Delta\beta^2)\operatorname{cov}(q_n, q_{n+1}).
\end{aligned}$$

Furthermore, $\mathsf{E}[q_n]$, $\operatorname{var}[q_n]$, $\operatorname{cov}(q_n, q_{n+1})$, $\operatorname{cov}(q_n^2, q_{n+1})$, and $\operatorname{cov}(q_n, q_{n+1}^2)$ in Eqs. (6)–(10) can be calculated as follows:

$$\mathsf{E}[q_n] = a\Delta\mathsf{E}[J], \quad \textbf{(11)}$$

$$\operatorname{var}[q_n] = \frac{a}{\lambda^2}(\lambda\Delta - (1 - e^{-\lambda\Delta}))\mathsf{E}[J^2], \quad \textbf{(12)}$$

$$\operatorname{cov}(q_n, q_{n+1}) = \frac{a}{2\lambda^2}(1 - e^{-\lambda\Delta})^2\mathsf{E}[J^2], \quad \textbf{(13)}$$

$$\operatorname{cov}(q_n^2, q_{n+1}) = \frac{a^2\Delta}{\lambda^2}(1 - e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2] + \frac{a}{3\lambda^3}(1 - e^{-\lambda\Delta})^3\mathsf{E}[J^3], \quad \textbf{(14)}$$

$$\operatorname{cov}(q_n, q_{n+1}^2) = \frac{a^2\Delta}{\lambda^2}(1 - e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2] + \frac{a}{6\lambda^3}(1 - e^{-\lambda\Delta})^3(1 + e^{-\lambda\Delta})\mathsf{E}[J^3], \quad \textbf{(15)}$$

where J is the jump size of the compound Poisson process. The detailed derivations of Eqs. (11)–(15) are provided in the Appendix. By substituting Eqs. (11)–(15) into Eqs. (6)–(10), $E[y_n]$, $\operatorname{var}[y_n]$, $\operatorname{cov}(y_n, y_{n+1})$, $\operatorname{cov}(y_n^2, y_{n+1})$, and $\operatorname{cov}(y_n, y_{n+1}^2)$ can be expressed in terms of parameters μ, β, λ, a, and b, respectively. On the other hand, $E[y_n]$, $\operatorname{var}[y_n]$, $\operatorname{cov}(y_n, y_{n+1})$, $\operatorname{cov}(y_n^2, y_{n+1})$, and $\operatorname{cov}(y_n, y_{n+1}^2)$ can be estimated based on $\{Y_1, \ldots, Y_N\}$, a sample of $\{y_1, \ldots, y_N\}$ as follows:

$$\mathsf{E}[y_n] \approx \frac{1}{N}\sum_{i=1}^{N} Y_i, \tag{16}$$

$$\operatorname{var}(y_n) \approx \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2, \tag{17}$$

$$\operatorname{cov}(y_n, y_{n+1}) \approx \frac{1}{N}\sum_{i=1}^{N-1}(Y_i - \overline{Y})(Y_{i+1} - \overline{Y}), \tag{18}$$

$$\operatorname{cov}(y_n^2, y_{n+1}) \approx \frac{1}{N}\sum_{i=1}^{N-1}(Y_i^2 - \overline{Y^2})(Y_{i+1} - \overline{Y}), \tag{19}$$

$$\operatorname{cov}(y_n, y_{n+1}^2) \approx \frac{1}{N}\sum_{i=1}^{N-1}(Y_i - \overline{Y})(Y_{i+1}^2 - \overline{Y^2}), \tag{20}$$

where $\overline{Y} = (1/N)\sum_{i=1}^{N} Y_i$ and $\overline{Y^2} = (1/N)\sum_{i=1}^{N} Y_i^2$. The above estimates in combination with Eqs. (6)–(15) can then be used to estimate μ, β, λ, a, and b, which we will discuss in detail in the next section.

In what follows, we present a more efficient way to estimate λ. Similar to Eq. (8), we have

$$\operatorname{cov}(y_n, y_{n+k}) = \beta^2 \operatorname{cov}(q_n, q_{n+k}), \quad \text{for } k \geq 1. \tag{21}$$

Furthermore, from Eqs. (4) and (5), we have for $k \geq 2$

$$\begin{aligned}
\operatorname{cov}(q_n, q_{n+k}) &= \frac{1}{\lambda^2}(\operatorname{cov}(\Delta v_n, \Delta v_{n+k}) - \operatorname{cov}(\Delta z_n, \Delta v_{n+k})) \\
&= \frac{1}{\lambda^2}(\operatorname{cov}(\Delta v_n, e^{-\lambda\Delta}\Delta v_{n+k-1} + \Delta e z_{n+k} - \Delta e z_{n+k-1}) \\
&\quad - \operatorname{cov}(\Delta z_n, e^{-\lambda\Delta}\Delta v_{n+k-1} + \Delta e z_{n+k} - \Delta e z_{n+k-1})) \\
&= \frac{e^{-\lambda\Delta}}{\lambda^2}(\operatorname{cov}(\Delta v_n, \Delta v_{n+k-1}) - \operatorname{cov}(\Delta z_n, \Delta v_{n+k-1})) \\
&= e^{-\lambda\Delta}\operatorname{cov}(q_n, q_{n+k-1}).
\end{aligned}$$

Therefore, we have

$$\operatorname{cov}(y_n, y_{n+k}) = e^{-(k-1)\lambda\Delta}\operatorname{cov}(y_n, y_{n+1}), \tag{22}$$

which leads to

$$\lambda = \frac{1}{(k-1)\Delta}\ln\left(\frac{\operatorname{cov}(y_n, y_{n+1})}{\operatorname{cov}(y_n, y_{n+k})}\right). \tag{23}$$

Eq. (23) can be used to estimate λ. For example, we can construct the following estimator for λ:

$$\hat{\lambda}_n = \frac{1}{K\Delta}\sum_{k=2}^{K}\frac{1}{k-1}\ln\left(\frac{(n-k)\sum_{i=1}^{n-1}(Y_i - \overline{Y})(Y_{i+1} - \overline{Y})}{(n-1)\sum_{i=1}^{n-k}(Y_i - \overline{Y})(Y_{i+k} - \overline{Y})}\right), \tag{24}$$

where $2 \leq K < n$ (usually K takes a small value, e.g., $K \leq 10$). Once λ is estimated, other parameters can be estimated as stated earlier by using Eqs. (6)–(15) or their subset.

3. PARAMETER ESTIMATION

In this section, we discuss how to derive our moment-based estimators for λ, μ, β, a, and b from Eqs. (6)–(15) and propose two moment-based estimation methods. We show that different estimators may be derived and they can have very different statistical properties. Before proceeding, we point out that similar to Eqs. (6)–(15) other moments equations (e.g., $\mathsf{E}[(y_n - \mathsf{E}[y_n])^3]$, $\mathrm{cov}(y_n, y_{n+k}^2)$, $\mathrm{cov}(y_n^2, y_{n+k})$, $k \geq 2$) can be derived as well, based on which different types of estimators can then be developed. However, in this paper, we only use Eqs. (6)–(15) (or their subset). In the rest of this section and thereafter, we assume that λ is estimated using Eq. (24); therefore, we only need to focus on estimating μ, β, a, and b by using Eqs. (6)–(15).

First, we present the following results which can be derived directly from Eqs. (6)–(15):

$$\frac{1}{\beta} = \frac{1}{1-e^{-\lambda\Delta}}\left[\frac{2\mathrm{cov}(y_n, y_{n+1}^2) - (1+e^{-\lambda\Delta})\,\mathrm{cov}(y_n^2, y_{n+1})}{\mathrm{cov}(y_n, y_{n+1})} - 2(1-e^{-\lambda\Delta})\mathsf{E}[y_n]\right], \tag{25}$$

$$\frac{\mathsf{E}[J]}{\mathsf{E}[J^2]} = \frac{\beta^2(1-e^{-\lambda\Delta})^2\,\mathrm{var}[y_n]}{2\lambda^2\Delta\,\mathrm{cov}(y_n, y_{n+1})} - \frac{\beta^2(\lambda\Delta - (1-e^{-\lambda\Delta}))}{\lambda^2\Delta}, \tag{26}$$

$$a = \frac{2\lambda^2\,\mathrm{cov}(y_n, y_{n+1})}{\beta^2(1-e^{-\lambda\Delta})^2\mathsf{E}[J^2]}, \tag{27}$$

$$\mu = \frac{\mathsf{E}[y_n]}{\Delta} - \beta a\mathsf{E}[J]. \tag{28}$$

It is clear that Eqs. (25)–(28) can be used to estimate β, b, a, and μ, respectively (though b is not expressed explicitly in Eq. (26)), which seems to be simple and elegant. However, as our numerical results in Section 5 show, the estimators based on Eqs. (25)–(28) (which we shall call Moment Method 1—MM1) may perform poorly in some cases.

In what follows, we propose a different approach to estimate μ, β, a, and b (which we shall call Moment Method 2—MM2). We only use Eqs. (6)–(9) and Eqs. (11)–(14), which are suffice in estimating μ, β, a, and b. Different from MM1, MM2 does not provide unified closed-form expressions for different jump size distributions. In the paper, we separately provide MM2 estimators for four different jump distributions: exponential jump, deterministic jump, inverse Gaussian jump, and Pareto jump, among which some of the expressions are implicit (e.g., for inverse Gaussian jump and Pareto jump); thus, we need to resort numerical methods to obtain the estimation for the corresponding parameters. As we should see in what follows, the exact formulas of the estimators derived based on this approach are different for different jump size distributions. To illustrate, we first consider the case in which the jump size is exponential and then expand to other cases.

3.1. Exponential Jump

If J is exponentially distributed with mean b, that is, $\mathsf{E}[J^n] = b^n n!$ $(n = 1, 2, 3, \ldots)$, then Eqs. (6)–(9) can be rewritten as follows:

$$\mathsf{E}[y_n] = \mu\Delta + \beta ab\Delta,$$

$$\mathrm{var}[y_n] = \frac{2a\beta^2b^2(\lambda\Delta - 1 + e^{-\lambda\Delta})}{\lambda^2} + ab\Delta,$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

$$\operatorname{cov}(y_n, y_{n+1}) = \frac{a\beta^2 b^2 (1 - e^{-\lambda\Delta})^2}{\lambda^2},$$
$$\operatorname{cov}(y_n^2, y_{n+1}) = \beta^3 \left(\frac{6ab^3(1 - e^{-\lambda\Delta})^3}{3\lambda^3} + \frac{2a^2 b^3 \Delta (1 - e^{-\lambda\Delta})^2}{\lambda^2} \right) + (\beta + 2\mu\beta^2\Delta) \frac{ab^2(1 - e^{-\lambda\Delta})^2}{\lambda^2}.$$

Using the above four equations, we can derive

$$\beta = \frac{\operatorname{cov}(y_n, y_{n+1})}{\operatorname{cov}(y_n^2, y_{n+1}) - 2\operatorname{cov}(y_n, y_{n+1})\mathsf{E}[y_n]} \left[1 + \frac{2\lambda\Delta \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})VC} \right], \tag{29}$$
$$b\beta^2 = \frac{\lambda^2 \Delta \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})^2 VC}, \tag{30}$$
$$ab = \frac{1}{\Delta} VC, \tag{31}$$
$$\mu = \frac{\mathsf{E}[y_n]}{\Delta} - \frac{\beta}{\Delta} VC, \tag{32}$$

where

$$VC \triangleq \operatorname{var}[y_n] - \frac{2(\lambda\Delta - 1 + e^{-\lambda\Delta}) \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})^2}.$$

It is clear that β, b, a, and μ can be estimated from Eqs. (29)–(32) recursively.

3.2. Deterministic Jump

If J is deterministic, we have $\mathsf{E}[J^n] = b^n$ $(n = 1, 2, 3, \ldots)$. Therefore, only Eqs. (29) and (30) need to be modified as follows:

$$\beta = \frac{\operatorname{cov}(y_n, y_{n+1})}{\operatorname{cov}(y_n^2, y_{n+1}) - 2\operatorname{cov}(y_n, y_{n+1})\mathsf{E}[y_n]} \left[1 + \frac{4\lambda\Delta \operatorname{cov}(y_n, y_{n+1})}{3(1 - e^{-\lambda\Delta})VC} \right], \tag{33}$$
$$b\beta^2 = \frac{2\lambda^2 \Delta \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})^2 VC}. \tag{34}$$

β, b, a, and μ can then be estimated based on Eqs. (33) and (34), together with Eqs. (31) and (32).

3.3. Inverse Gaussian Jump

If J is inverse Gaussian distributed, then its first three moments are $\mathsf{E}[J] = b, \mathsf{E}[J^2] = b + b^2$, and $\mathsf{E}[J^3] = b^3 + 3b^2 + 3b$, respectively. Hence, instead of Eqs. (29) and (30), we have

$$\frac{\operatorname{cov}(y_n^2, y_{n+1})}{\operatorname{cov}(y_n, y_{n+1})} = \frac{1}{\beta} + 2\mathsf{E}[y_n] + \frac{2(1 - e^{-\lambda\Delta})}{3\lambda} \left(\beta + \frac{BB}{\beta} + \frac{\beta^3}{BB} \right), \tag{35}$$
$$\beta^2 (1 + b) = \frac{2\lambda^2 \Delta \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})^2 VC}, \tag{36}$$

where

$$BB \triangleq \frac{2\lambda^2 \Delta \operatorname{cov}(y_n, y_{n+1})}{(1 - e^{-\lambda\Delta})^2 VC}.$$

Though it is not expressed explicitly in β, Eq. (35) can be used to solve β numerically. In fact, when $\beta > 0$ (or $\beta < 0$), Eq. (35) can be transformed into a strictly convex (or concave) function with respect to β with two possibilities: two real-valued solutions or none. In our numerical experiments presented in Section 5, we adopt the following empirical policy: if there are two solutions, then we take the smaller one as our solution for β; otherwise, take the minimum (or maximum) point as our solution. Once β is obtained, then b, a, and μ can be easily estimated based on Eqs. (36), (31), and (32) as for exponential and deterministic jumps.

3.4. Pareto Jump

Finally, we consider the case in which J has the Pareto distribution. Without loss of generality, we assume that the scale parameter (the minimum value of the Pareto distribution) is equal to 1; hence, $\mathsf{E}[J^n] = b/(b-n)$, $n = 1, 2, 3, \ldots$. We have

$$\frac{ab\beta^2}{b-2} = \frac{2\lambda^2 \operatorname{cov}(y_n, y_{n+1})}{(1-e^{-\lambda\Delta})^2}, \tag{37}$$

$$a = \frac{b-1}{b\Delta} VC, \tag{38}$$

$$\beta = \frac{\lambda}{1-e^{-\lambda\Delta}} \sqrt{\frac{2(b-2)\Delta \operatorname{cov}(y_n, y_{n+1})}{(b-1)VC}}. \tag{39}$$

We can then obtain the following equation for solving b

$$\begin{aligned}\frac{\operatorname{cov}(y_n^2, y_{n+1})}{\operatorname{cov}(y_n, y_{n+1})} &= \frac{1}{\beta} + 2\mu\Delta + \beta\left[\frac{4(e^{\lambda\Delta}-1)}{3\lambda e^{\lambda\Delta}(b-3)(b-1)} + \frac{2b(-1+e^{\lambda\Delta}+3a\lambda e^{\lambda\Delta}\Delta)}{3\lambda e^{\lambda\Delta}(b-1)}\right] \\ &= \frac{1-e^{-\lambda\Delta}}{\lambda}\sqrt{\frac{(b-1)VC}{2(b-2)\Delta \operatorname{cov}(y_n, y_{n+1})}} + 2\mathsf{E}[y_n] \\ &\quad + \frac{2(b-2)}{3(b-3)}\sqrt{\frac{2(b-2)\Delta \operatorname{cov}(y_n, y_{n+1})}{(b-1)VC}}.\end{aligned} \tag{40}$$

Notice that Eq. (40) is a six-degree equation which does not have any analytical solution. Therefore, we have to resort to numerical methods (such as the R function "optimize") to solve b. Once b is obtained, then a, β, and μ can be estimated easily.

It is clear that the estimates of all the parameters can then be obtained based on the above equations if we replace the theoretical moments with the corresponding sample moments provided by Eqs. (16)–(20). We introduce the following general notations for the moments:

$$\hat{\psi}_n = (\hat{\eta}(\zeta), \hat{\gamma}_{\zeta,\zeta}(0),\ 1 \le \zeta \le r_0 \vee s_0;\ \hat{\gamma}_{r,s}(h),\ 1 \le h \le k,\ 1 \le r \le r_0,\ 1 \le s \le s_0)', \tag{41}$$

$$\psi_* = (\eta(\zeta), \gamma_{\zeta,\zeta}(0),\ 1 \le \zeta \le r_0 \vee s_0;\ \gamma_{r,s}(h),\ 1 \le h \le k,\ 1 \le r \le r_0,\ 1 \le s \le s_0)', \tag{42}$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

where

$$\hat{\eta}(\zeta) = \frac{1}{n}\sum_{i=1}^{n} y_i^{\zeta},$$

$$\hat{\gamma}_{r,s}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(y_i^r - \hat{\eta}(r))(y_{i+h}^s - \hat{\eta}(s)),$$

$$\eta(\zeta) = \mathsf{E}[y_n^{\zeta}],$$

$$\gamma_{r,s}(h) = \mathrm{cov}(y_n^r, y_{n+h}^s),$$

$$r \vee s = \max(r,s),$$

(for $h = 0, 1, \ldots, k,\ r = 1, 2, \ldots, r_0,\ s = 1, 2, \ldots, s_0,\ \zeta = 1, 2, \ldots, r_0 \vee s_0$). It is clear that $\hat{\eta}(\zeta)$ is the ζ-th sample moment and $\hat{\gamma}_{r,s}(h)$ is the sample autocovariance with lag h. For $h = 0, r = s$, the autocovariance simply reduces to the variance. Furthermore, let us define our moment-based estimators (for both MM1 and MM2) and the corresponding true values as follows:

$$\hat{\theta}_n = (\hat{\lambda}_n, \hat{\mu}_n, \hat{\beta}_n, \hat{a}_n, \hat{b}_n)', \tag{43}$$

$$\theta_* = (\lambda_*, \mu_*, \beta_*, a_*, b_*)'. \tag{44}$$

Based on the above derivations, it is not difficult to see that using either MM1 or MM2, we can always define a continuous mapping $G : \mathbb{R}^{\dim(\psi_*)} \longrightarrow \mathbb{R}^5$, such that

$$G(\hat{\psi}_n) = \hat{\theta}_n, \tag{45}$$

$$G(\psi_*) = \theta_*, \tag{46}$$

where $\dim(\psi_*)$ signifies the dimension of vector ψ_*.

For example, if we denote the moments and estimators for MM1 as

$$\hat{\psi}_n^{\mathrm{MM1}} = (\hat{\eta}(1), \hat{\gamma}_{1,1}(0), \hat{\gamma}_{1,1}(1), \hat{\gamma}_{1,1}(k), \hat{\gamma}_{2,1}(1), \hat{\gamma}_{1,2}(1))',$$

$$\psi_*^{\mathrm{MM1}} = (\eta(1), \gamma_{1,1}(0), \gamma_{1,1}(1), \gamma_{1,1}(k), \gamma_{2,1}(1), \gamma_{1,2}(1))',$$

$$\hat{\theta}_n^{\mathrm{MM1}} = (\hat{\lambda}_n^{\mathrm{MM1}}, \hat{\mu}_n^{\mathrm{MM1}}, \hat{\beta}_n^{\mathrm{MM1}}, \hat{a}_n^{\mathrm{MM1}}, \hat{b}_n^{\mathrm{MM1}})',$$

respectively, then based on Eqs. (23) and (25)–(28), we have a continuous mapping $g : \mathbb{R}^6 \longrightarrow \mathbb{R}^5$, such that

$$g(\hat{\psi}_n^{\mathrm{MM1}}) = \hat{\theta}_n^{\mathrm{MM1}}, \tag{47}$$

$$g(\psi_*^{\mathrm{MM1}}) = \theta_*. \tag{48}$$

These continuous mappings are useful in establishing consistency and asymptotic normality for our estimators in the next section.

4. LARGE-SAMPLE BEHAVIOR: STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY

In this section, we address the issues related to large-sample behaviors for our moment-based estimators (for both MM1 and MM2) derived in the previous section. First, we use

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

MM1 to illustrate the basic idea of how strong consistency and asymptotic normality can be established for our moment-based estimators.

LEMMA 4.1: *We have the following results as $n \to \infty$:*

(i) *If the sample moments derived based on MM1 are strongly consistent, that is,*

$$\hat{\psi}_n^{MM1} \longrightarrow \psi_*^{MM1} \quad a.s.,$$

then

$$\hat{\theta}_n^{MM1} \longrightarrow \theta_* \quad a.s.$$

(ii) *If the sample moments derived based on MM1 are asymptotically normal, that is, there exists a (positive definite) covariance matrix Σ_{MM1} such that*

$$\sqrt{n}\left(\hat{\psi}_n^{MM1} - \psi_*^{MM1}\right) \xrightarrow{d} N(0, \Sigma_{MM1}),$$

where $\xrightarrow{d}$ denotes the convergence in distribution, then

$$\sqrt{n}(\hat{\theta}_n^{MM1} - \theta_*) \xrightarrow{d} N(0, \mathrm{J}_g \Sigma_{MM1} \mathrm{J}_g'),$$

where J_g is the Jacobian of g.

PROOF: Using MM1, all the estimates can be uniquely expressed by moment-based quantities as expressed by Eqs. (47)–(48). Hence, (i) follows directly from the continuous mapping theorem.

On the other hand, if the sample moments are asymptotically normal, applying the delta method we have

$$\sqrt{n}(g(\hat{\psi}_n^{\mathrm{MM1}}) - g(\psi_*^{\mathrm{MM1}})) \xrightarrow{d} N(0, \mathrm{J}_g \Sigma_{\mathrm{MM1}} \mathrm{J}_g'),$$

where J_g is the Jacobian of g, a 5×6 matrix of partial derivatives with respect to the entries of g. Taking the first row of J_g as an example, it is the gradient of λ_* with respect to ψ_*^{MM1} as follows:

$$\left(0, 0, \frac{1}{(k-1)\Delta \operatorname{cov}(y_n, y_{n+1})}, \frac{1}{(1-k)\Delta \operatorname{cov}(y_n, y_{n+k})}, 0, 0\right).$$

Combined this with Eqs. (47)–(48), we have (ii). □ ■

REMARK 4.2: *In fact, using MM1 or MM2, our moment-based estimators are continuous mappings of the sample moments, they are strongly consistent and asymptotically normal if the sample moments are strongly consistent and asymptotically normal (based on the continuous mapping theorem and the delta method). Therefore, the key to proving strong consistency and asymptotic normality is to establish strong consistency and asymptotic normality for the sample moments.*

We begin with the following assumptions:

(A1) The volatility process $v(t)$ is strictly stationary.

(A2) F_v is the self-decomposable marginal distribution of $v(t)$ satisfying

$$\int_{\mathbb{R}} |v|^p F_v \, dv < \infty,$$

for some $p > 0$.

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

(A3) There exists a constant $\kappa > 0$ such that

$$E[y_1^{4+\kappa}] < \infty \quad \text{and} \quad E[|Z_1|^{2+\kappa}] < \infty,$$

where

$$Z_1 = (y_1^\zeta, (y_1^\zeta - \eta(\zeta))^2,\ 1 \le \zeta \le r_0 \vee s_0;$$
$$(y_1^r - \eta(r))(y_{1+h}^s - \eta(s)),\ 1 \le h \le k,\ 1 \le r \le r_0,\ 1 \le s \le s_0)'.$$

We point out that under some moderate conditions (e.g., see [27]), (A.1) and (A.2) are satisfied.

THEOREM 4.3: *Under (A.1)–(A.3), we have the following results as $n \to \infty$:*

(i) *Moment-based estimators are strongly consistent, that is,*

$$\hat{\theta}_n \longrightarrow \theta_* \quad a.s. \tag{49}$$

(ii) *Moment-based estimators are asymptotically normal, that is, there exists a covariance matrix Σ_θ such that*

$$\sqrt{n}(\hat{\theta}_n - \theta_*) \xrightarrow{d} N(0, \Sigma_\theta). \tag{50}$$

REMARK 4.4: *Notice that we do not give an analytical expression for Σ_θ in (50). However, as we can see in the proof of Lemma 4.1, as long as there is an explicit functional relationship between moment-based quantities and parameters (see Eqs. (25)–(40)), we can always give the closed-form expression for Σ_θ through basic calculus (e.g., derivatives of implicit function and chain rules).*

Before proving Theorem 4.1, we first introduce some definitions on mixing, which plays a central role in the proof of Theorem 4.1.

DEFINITION 4.5: *Suppose that $V = V(t)_{t \ge 0}$ is a stationary process with σ-algebras $\mathscr{F}_1 = \mathscr{F}_{(0,u)} = \sigma(\{V_t\},\ 0 \le t < u)$ and $\mathscr{F}_2 = \mathscr{F}_{[u+x,\infty)} = \sigma(\{V_t\},\ t \ge u + x)$, then*

(i) *V is called β-mixing if:*

$$\beta(x) = \sup_{A_i \in \mathscr{F}_1, B_j \in \mathscr{F}_2} \frac{1}{2} \sum_i \sum_j \mid P(A_i \cap B_j) - P(A_i)P(B_j) \mid \to 0 \quad as\ x \to \infty, \tag{51}$$

where the supremum is taken over all pairs of partitions A_i and B_j such that $A_i \in \mathscr{F}_1$ for each i and $B_j \in \mathscr{F}_2$ for each j.

(ii) *V is called β-mixing with exponential rate if it is β-mixing with*

$$\beta(x) \le C_1 e^{-q_1 x}, \quad \forall x \ge 0, \tag{52}$$

for some $C_1 > 0$ and $q_1 > 0$.

(iii) *V is called strong mixing (or α-mixing) if:*

$$\alpha(x) = \sup_{A \in \mathscr{F}_1, B \in \mathscr{F}_2} \mid P(A \cap B) - P(A)P(B) \mid \to 0 \quad as\ x \to \infty. \tag{53}$$

(iv) V *is called strong mixing with the exponential rate if it is strong mixing with*

$$\alpha(x) \le C_2 e^{-q_2 x}, \quad \forall x \ge 0, \tag{54}$$

for some $C_2 > 0$ *and* $q_2 > 0$.

The above definitions on β-mixing and strong mixing (with the exponential rate) can be easily extended to a discrete-time process. We point out that β-mixing implies strong mixing, and strong mixing implies ergodicity.

According to Theorem 4.3 in [22], we have

PROPOSITION 4.6: *Under (A.1)–(A.2),* $v(t)$ *is* β*-mixing with* $\beta(x) = O(e^{-qx})$ *for some* $q > 0$ *as* $x \to \infty$*, which implies that both* $v(t)$ *and its discrete-time correspondence* $\{v_n\}$ *are ergodic.*

On the other hand, it is not difficult to see that $\{y_n\}$ can be viewed as a generalized hidden Markov model (see Definition 3 in [7]), with the following observation kernel:

$$y_n | v_n \sim N(\mu\Delta + \beta\Delta v_n, \Delta v_n). \tag{55}$$

Since $\{v_n\}$ is β-mixing, we have the following result based on Carrasco and Chen [7] (Proposition 4 (ii)):

PROPOSITION 4.7: $\{y_n\}$ *is a generalized hidden Markov model with a hidden chain* $\{v_n\}$ *as defined by (55), and* $\{y_n\}$ *is also* β*-mixing with a decaying rate at least as fast as that of* $\{v_n\}$.

We are now ready to prove Theorem 4.1.

PROOF: Based on Propositions 4.1 and 4.2, $\{y_n\}$ is strictly stationary and ergodic. Therefore, we have the following strongly consistent result (see [4]):

$$\hat{\psi}_n \longrightarrow \psi_* \quad \text{a.s.}$$

According to the continuous mapping theorem, we have

$$G(\hat{\psi}_n) \longrightarrow G(\psi_*) \quad \text{a.s.}$$

that is, (49) holds.

Next, we prove (50). We define

$$\begin{aligned} Z_i = (&y_i^\zeta, (y_i^\zeta - \eta(\zeta))^2,\ 1 \le \zeta \le r_0 \vee s_0;\ (y_i^r - \eta(r))(y_{i+h}^s - \eta(s)), \\ &1 \le h \le k,\ 1 \le r \le r_0,\ 1 \le s \le s_0)', \end{aligned}$$

$$\sigma_{m,l} = \operatorname{cov}(Z_1^{(m)}, Z_1^{(l)}) + 2\sum_{i=1}^{\infty} \operatorname{cov}(Z_1^{(m)}, Z_{i+1}^{(l)}),$$

where $Z_i^{(m)}$ is the mth component of Z_i ($m = 1, 2, \ldots, \dim(\psi_*)$),

$$\Sigma = [\sigma_{m,l}]_{m,l=1}^{\dim(\psi_*)},$$

$$\hat{\gamma}_{r,s}^*(h) = \frac{1}{n}\sum_{i=1}^{n} (y_i^r - \eta(r))(y_{i+h}^s - \eta(s)),\ h \in \{0, \ldots, k\},\ r \in \{1, \ldots, r_0\},\ s \in \{1, \ldots, s_0\},$$

$$\hat{\psi}_n^* = (\hat{\eta}(1), \ldots, \hat{\eta}(r_0 \vee s_0),\ \hat{\gamma}_{1,1}(0), \hat{\gamma}_{1,2}^*(1), \hat{\gamma}_{2,1}^*(1), \ldots, \hat{\gamma}_{r_0,s_0}^*(k))'.$$

TABLE 1. Comparison between MM1 and MM2: exponential jump case

True values	MM1	MM2	MM1 with β given
$\lambda = 0.4$	0.400 ± 0.004	0.400 ± 0.005	0.400 ± 0.005
$\mu = 1$	-9.481 ± 339.637	0.975 ± 0.432	0.980 ± 0.737
$\beta = 2$	2.847 ± 27.361	2.006 ± 0.068	N/A
$a = 3$	610.732 ± 1034.819	3.007 ± 0.118	3.009 ± 0.194
$b = 4$	6.781 ± 7.784	3.989 ± 0.185	3.999 ± 0.137
$\lambda = 0.8$	0.798 ± 0.017	0.801 ± 0.018	0.799 ± 0.018
$\mu = 0.5$	0.594 ± 1.602	0.503 ± 0.165	0.504 ± 0.152
$\beta = -0.5$	-0.504 ± 0.088	-0.500 ± 0.013	N/A
$a = 2.5$	2.646 ± 1.076	2.503 ± 0.109	2.509 ± 0.128
$b = 7$	7.4560 ± 2.401	7.004 ± 0.296	6.996 ± 0.242
$\lambda = 0.32$	0.320 ± 0.003	0.320 ± 0.003	0.320 ± 0.003
$\mu = 9.41$	23.476 ± 2696.76	9.390 ± 3.538	9.276 ± 18.870
$\beta = 4.66$	4.138 ± 81.854	4.755 ± 0.619	N/A
$a = 6.27$	1960.21 ± 3723.69	6.277 ± 0.308	6.380 ± 1.573
$b = 5.23$	164.722 ± 239.61	5.210 ± 0.680	5.309 ± 0.701

TABLE 2. Comparison between MM1 and MM2: deterministic jump case

True values	MM1	MM2	MM1 with β given
$\lambda = 0.4$	0.400 ± 0.005	0.399 ± 0.005	0.400 ± 0.005
$\mu = 1$	-1.446 ± 80.228	1.005 ± 0.718	1.026 ± 0.450
$\beta = 2$	2.196 ± 6.653	1.999 ± 0.068	N/A
$a = 3$	37.459 ± 434.69	3.001 ± 0.187	2.994 ± 0.126
$b = 4$	4.529 ± 3.286	4.014 ± 0.252	4.006 ± 0.096
$\lambda = 0.8$	0.797 ± 0.026	0.800 ± 0.027	0.797 ± 0.028
$\mu = 0.5$	0.747 ± 1.809	0.514 ± 0.289	0.505 ± 0.136
$\beta = -0.5$	-0.513 ± 0.102	-0.500 ± 0.017	N/A
$a = 2.5$	2.758 ± 1.264	2.512 ± 0.193	2.513 ± 0.149
$b = 7$	7.324 ± 2.571	6.999 ± 0.497	6.985 ± 0.308
$\lambda = 0.32$	0.319 ± 0.003	0.320 ± 0.003	0.319 ± 0.003
$\mu = 9.41$	333.29 ± 6551.13	9.476 ± 6.709	8.827 ± 0.874
$\beta = 4.66$	-6.273 ± 212.155	4.707 ± 0.353	N/A
$a = 6.27$	11508.5 ± 18946.7	6.273 ± 0.559	6.333 ± 0.745
$b = 5.23$	65.519 ± 95.797	5.227 ± 0.542	5.232 ± 0.318

Under (A.3), our proof of asymptotic normality essentially follows that of Proposition 2 in [18]. For every $w \in \mathbb{R}^{d+2}$ satisfying $w'\Sigma w > 0$, we consider $\{w'Z_i\}$. Recall that $\{y_n\}$ is strictly stationary and β-mixing with the exponential rate; therefore, it is strictly stationary and strong mixing with the exponential rate. According to Francq and Zakoian [11] (Appendix A.3), $\{Z_i\}$ is also strictly stationary and strong mixing with the exponential rate. Furthermore, it is known that strong mixing with the exponential rate is preserved under linear transformation; hence, $\{w'Z_i\}$ is strictly stationary and strong mixing with the exponential rate. Therefore,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w'Z_i - w'\psi_0\right) \xrightarrow{d} N(0, \tilde{\sigma}^2),$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

Table 3. Comparison between MM1 and MM2: inverse Gaussian jump case

True values	MM1	MM2	MM1 with β given
$\lambda = 0.4$	0.399 ± 0.004	0.400 ± 0.005	0.400 ± 0.004
$\mu = 1$	25.449 ± 412.589	0.877 ± 0.990	1.050 ± 0.483
$\beta = 2$	-0.006 ± 33.760	2.015 ± 0.095	N/A
$a = 3$	-4.638 ± 117.017	3.051 ± 0.328	2.986 ± 0.148
$b = 4$	5.545 ± 5.961	3.967 ± 0.432	4.014 ± 0.123
$\lambda = 0.8$	0.799 ± 0.023	0.802 ± 0.025	0.801 ± 0.026
$\mu = 0.5$	0.781 ± 1.749	0.485 ± 0.289	0.488 ± 0.134
$\beta = -0.5$	-0.516 ± 0.099	-0.499 ± 0.018	N/A
$a = 2.5$	2.915 ± 1.547	2.491 ± 0.205	2.490 ± 0.153
$b = 7$	7.279 ± 2.864	7.060 ± 0.568	7.037 ± 0.333
$\lambda = 0.32$	0.320 ± 0.003	0.319 ± 0.003	0.320 ± 0.003
$\mu = 9.41$	100.672 ± 1441.56	9.072 ± 8.405	9.449 ± 11.634
$\beta = 4.66$	1.713 ± 47.307	4.674 ± 0.438	N/A
$a = 6.27$	2.152 ± 34.144	6.345 ± 0.861	6.319 ± 1.070
$b = 5.23$	98.932 ± 142.99	5.296 ± 0.887	5.269 ± 0.492

Table 4. Comparison between MM1 and MM2: Pareto jump case

True values	MM1	MM2	MM1 with β given
$\lambda = 0.4$	0.399 ± 0.006	0.400 ± 0.006	0.399 ± 0.006
$\mu = 1$	37.046 ± 568.911	0.805 ± 0.091	0.359 ± 0.107
$\beta = 2$	$-0.9.494 \pm 132.136$	1.896 ± 0.029	N/A
$a = 3$	0.472 ± 21.049	3.202 ± 0.050	3.385 ± 0.077
$b = 4$	3.036 ± 60.288	3.872 ± 0.127	4.624 ± 0.190
$\lambda = 0.8$	0.801 ± 0.121	0.799 ± 0.118	0.808 ± 0.119
$\mu = 0.5$	0.367 ± 10.172	0.504 ± 0.086	0.505 ± 0.019
$\beta = 0.5$	0.544 ± 3.468	0.500 ± 0.033	N/A
$a = 2.5$	1.869 ± 7.183	2.475 ± 0.194	2.516 ± 0.306
$b = 7$	-2.627 ± 95.950	11.774 ± 21.103	18.245 ± 136.235
$\lambda = 0.32$	0.320 ± 0.004	0.320 ± 0.004	0.320 ± 0.003
$\mu = 9.41$	53.566 ± 1.268	4.502 ± 0.419	7.573 ± 0.841
$\beta = 4.66$	-0.704 ± 0.112	3.602 ± 0.047	N/A
$a = 6.27$	5.778 ± 0.091	8.592 ± 0.149	13.105 ± 0.533
$b = 5.23$	2.027 ± 0.009	4.073 ± 0.110	-6.971 ± 1.483

where $\tilde{\sigma}^2 = \mathrm{var}(w'Z_1) + 2\sum_{i=1}^{n} \mathrm{cov}(w'Z_1, w'Z_{i+1}) = w'\Sigma w$ [20, Thm. 1.7]. In other words, for every $w \in \mathbb{R}^{d+2}$ satisfying $w'\Sigma w > 0$, we have

$$\sqrt{n}\left(w'\hat{\psi}_n^* - w'\psi_0\right) \xrightarrow{d} N(0, w'\Sigma w),$$

(noting that $\hat{\psi}_n^* = (1/n)\sum_{i=1}^{n} Z_i$). Then, by the Cramer–Wold device, we have

$$\sqrt{n}(\hat{\psi}_n^* - \psi_0) \xrightarrow{d} N(0, \Sigma). \tag{56}$$

On the other hand, following the proof of Proposition 7.3.4 of [6], we have

$$\sqrt{n}\left(w'\hat{\psi}_n^* - w'\hat{\psi}_n\right) \longrightarrow 0 \quad \text{in probability,}$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

TABLE 5. Comparison between MM2 and GSMLE

	λ	μ	β	a	b	T
True values	0.4	0.5	−0.5	2	3	–
GSMLE	0.974 ± 0.158	0.663 ± 0.126	-0.827 ± 0.190	2.757 ± 0.495	4.576 ± 0.794	4.86 h
MM2$^{(1)}$	0.399 ± 0.020	0.502 ± 0.053	-0.5002 ± 0.012	2.007 ± 0.082	2.992 ± 0.134	14.81 s
MM2$^{(2)}$	0.399 ± 0.000	0.499 ± 0.002	-0.4999 ± 0.000	1.999 ± 0.003	3.001 ± 0.005	3.25 h
True values	0.5	0.1	0.2	3	2	–
GSMLE	0.997 ± 0.014	0.009 ± 0.210	0.127 ± 0.162	2.345 ± 0.238	1.471 ± 0.214	5.10 h
MM2$^{(1)}$	0.538 ± 0.212	0.129 ± 0.097	0.194 ± 0.018	2.943 ± 0.557	2.026 ± 0.401	15.73 s
MM2$^{(2)}$	0.502 ± 0.006	0.100 ± 0.003	0.200 ± 0.000	2.991 ± 0.023	2.006 ± 0.015	4.30 h
True values	0.347	0.373	0.103	0.897	0.239	–
GSMLE	0.723 ± 0.451	0.034 ± 0.206	0.167 ± 0.128	2.182 ± 1.566	0.726 ± 0.511	4.42 h
MM2$^{(1)}$	0.044 ± 0.885	0.369 ± 0.078	0.121 ± 0.368	1.216 ± 11.594	0.014 ± 0.331	13.14 s
MM2$^{(2)}$	0.354 ± 0.075	0.373 ± 0.002	0.102 ± 0.007	0.890 ± 0.079	0.243 ± 0.023	2.21 h

that is, $\hat{\psi}_n$ has the same asymptotic behavior as $\hat{\psi}_n^*$. Combining this with (56), we can conclude that the following holds:

$$\sqrt{n}\left(\hat{\psi}_n - \psi_*\right) \xrightarrow{d} N(0, \Sigma), \tag{57}$$

that is, the sample moments are asymptotically normal. Applying the delta method to (57), we have

$$\sqrt{n}\left(G(\hat{\psi}_n) - G(\psi_*)\right) \xrightarrow{d} \left[\frac{\partial G(\psi_*)}{\partial \psi_*}\right] N(0, \Sigma), \tag{58}$$

that is, (50) holds. This completes our proof. ∎

5. NUMERICAL EXPERIMENTS

In the previous sections, we develop two different ways of estimating the parameters based on our moment-based method (MM1 and MM2). We also establish strong consistency and central limit theorem for these estimators. In this section, we numerically investigate the statistical efficiency of MM1 and MM2 and also compare them with other existing parameter estimation methods. The results can be summarized as follows:

1. First, we compare MM1 and MM2. Our numerical results show that MM2 performs much better than MM1. One reason is that MM2 uses less high-order (third-order) sample moments than MM1 does. This illustrates that our moment-based method may produce different estimators with different statistical efficiency. In general, those using less high-order sample moments would perform better. This explains why many moment-based methods that rely heavily on high-order moment estimations are usually statistically inefficient.
2. Secondly, we compare MM2 with a type of MLE method recently proposed by Peng *et al.* [23], which they call gradient-based simulated maximum likelihood estimation method (GSMLE). Our numerical results show that MM2 performs much better than GSMLE. We should point out that even though we do not provide a numerical comparison between MM2 and another traditional method, the MCMC method, the numerical results provided in [23] shows that GSMLE has several advantages

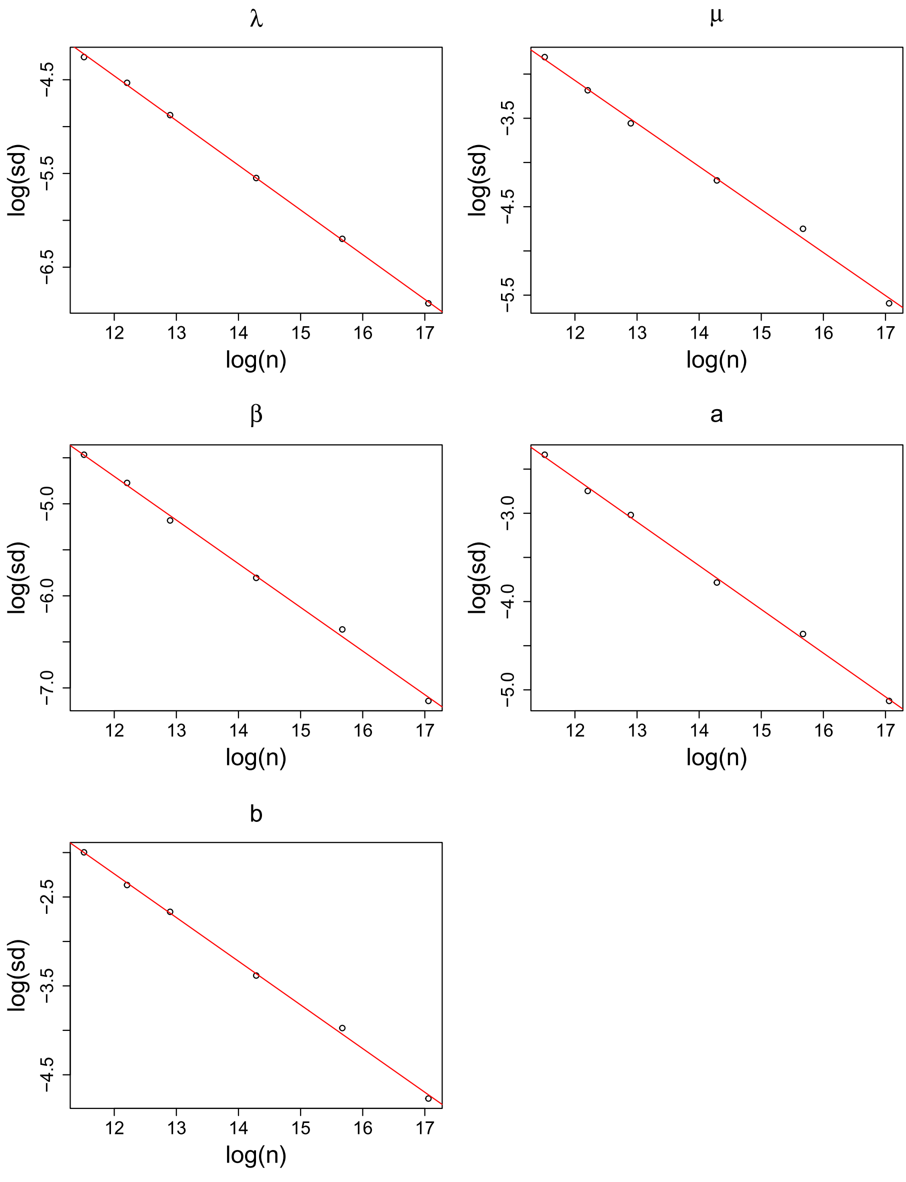

FIGURE 1. Log(sample standard deviation)–log(sample size) plot 1.

over MCMC. For example, MCMC highly depends on the initial prior distribution of the parameters, and it also converges much more lowly than GSMLE. Our own investigation also confirms these conclusions.

3. Finally, we conduct extensive numerical experiments under different parameter settings to test MM2. Our results show that MM2 performs reasonably well in most instances.

All numerical experiments reported in this section were performed on a PC with single Core(TM) m3-6Y30 processor.

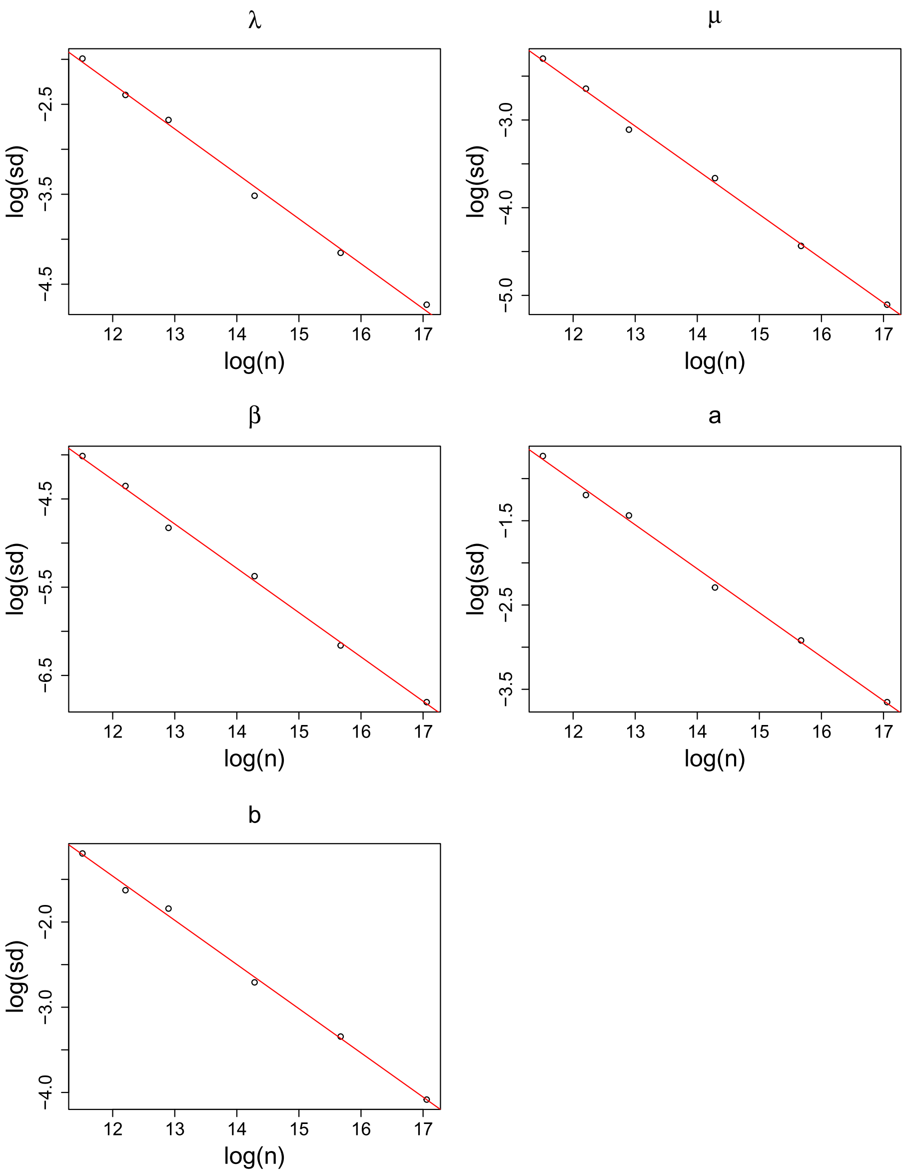

FIGURE **2.** Log(sample standard deviation)–log(sample size) plot 2.

5.1. MM1 vs. MM2

In this subsection, we compare MM1 and MM2. The estimators of MM1 Eqs. (25)–(28) are much simpler and straightforward than those of MM2; however, our extensive numerical experiments show that MM2 surprisingly performs much better than MM1. Tables 1–4 provide some numerical examples under different jump size distributions with multiple parameter settings. (Note that according to Eq. (39), β must be nonnegative in the case of Pareto jump.) For each jump size distribution, we run 400 replications with 100,000 samples for each replication. The numerical results are presented as “mean $\pm$ standard deviation” based on these 400 replications (the format remains the same for all numerical results in

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

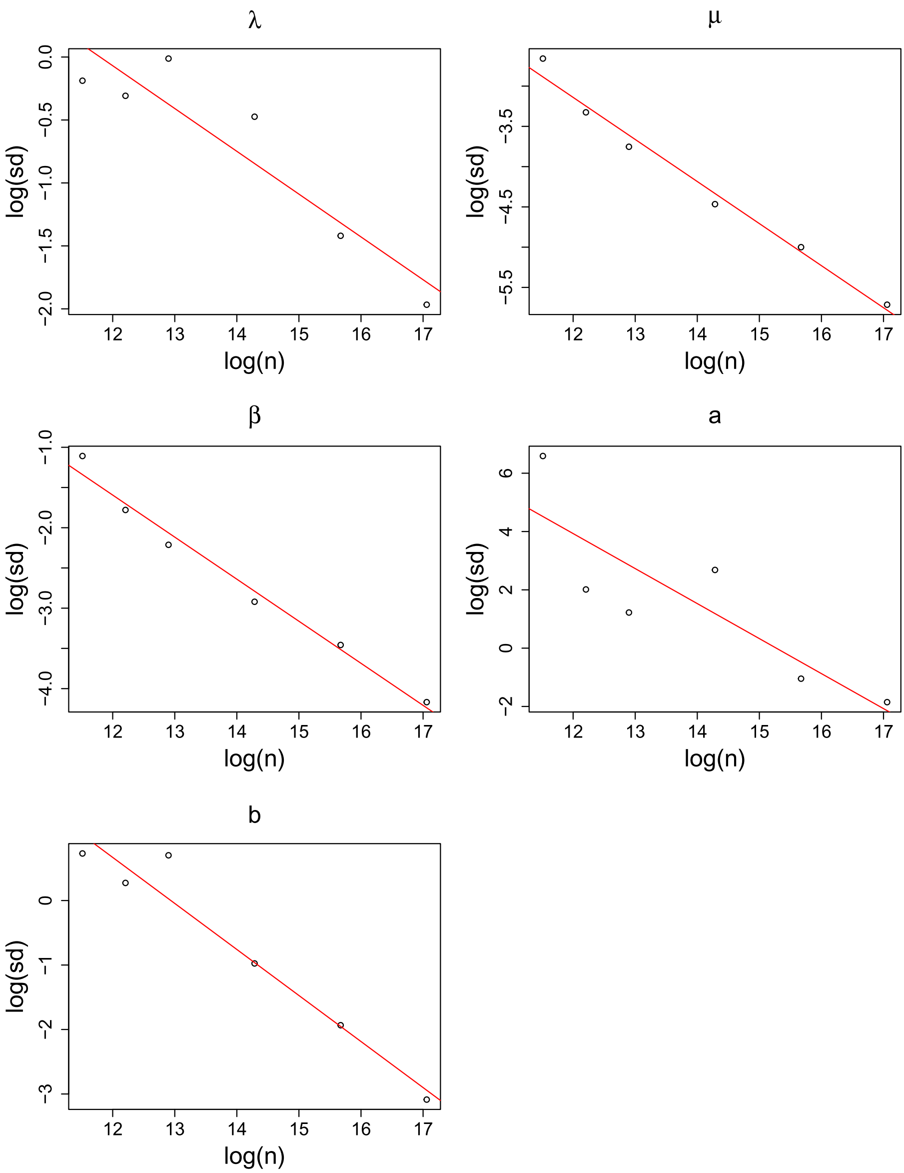

FIGURE **3.** Log(sample standard deviation)–log(sample size) plot 3.

this section). It is clear that by comparing the results in columns 2 and 3 in Tables 1–4, MM2 performs much better than MM1. In fact, MM1 performs poorly. One reason is that MM1 uses more third-order sample moments ($\text{cov}(y_n, y_{n+1}^2)$ and $\text{cov}(y_n^2, y_{n+1})$) than MM2 does ($\text{cov}(y_n^2, y_{n+1})$) in estimating β. We note that the third-order moment estimations are only explicitly used in Eq. (25) in estimating β. Hence, to further verify this hypothesis, we assume that β is given and use Eqs. (26)–(28) to estimate b, a, and μ, the numerical results are given by column 4 in Tables 1–4. It is clear that most estimation results are significantly improved. Therefore, if possible, we should avoid using high-order moments as much as possible in our moment-based estimation. In addition, we remark that MM1 is more sensitive to simulation errors than MM2 in the sense that a small disturbance in

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

simulation samples can lead to a large estimation error. For example, for Pareto jump, we have $b = 1/(1 - EJ/EJ^2) + 1$ based on MM1; however, in some simulation runs, the "EJ/EJ^2" term are slightly greater than 1 (even when the parameter β is fixed), which may produce negative estimation for b. Therefore, how to overcome this numerical problem in MM1 is an important topic for future research. In the remainder of this section, we will temporarily focus on MM2 only.

5.2. MM2 vs. Other Methods

In this subsection, we compare MM2 with GSMLE, a type of MLE method recently proposed by Peng *et al.* [23]. However, we should mention that we have done extensive numerical studies comparing MM2, GSMLE, and MCMC as well and found that in general MM2 performs much better than both GSMLE and MCMC, while GSMLE is better than MCMC. Since Peng *et al.* [23] have presented many numerical results comparing GSMLE and MCMC and our conclusion is the same as theirs, to make our numerical study more focused here, we choose to only present our experimental results on MM2 and GSMLE (for more discussions on a comparison between GSMLE and MCMC, the reader is referred to [23]). Before proceeding, we want to point out GSMLE proposed in [23] is only applicable when the volatility process follows a Gamma distribution (i.e., the jump size is exponentially distributed). Therefore, all the numerical examples considered in this subsection have exponential jump sizes. In addition, it is worth noting that many existing MCMC methods proposed in the literature are also quite restrictive in their applications. For example, they often assume that μ and β are known in Eq. (1) and focus on estimating a, b, and λ, the parameters involved in Eq. (2).

We consider three numerical examples here (they are the same as the last three examples of Table 8 in [23]). For each example, we run 40 replications (we do not run 400 replications as we do in other numerical experiments because each replication of GSMLE method requires hours of computational time). For each replication,

1. GSMLE method iterates 1,000 steps with initial values $(0.3, 0.5, 0.5, 1.5, 1.5)$, $N = 1,000$, and $M = 100$, where N is the sample size of each simulation replication and M is the number of particles as mass points to represent the posterior distribution in sequential Monte Carlo (SMC). The feasible parameter space is $[0.01, 1] \times [-1, 1] \times [-1, 1] \times [0.1, 5] \times [0.1, 5]$ and $\Delta = 1$.
2. MM2 is run for two scenarios: one with 100,000 samples ($\mathrm{MM2}^{(1)}$) and the other one with 100,000,000 samples ($\mathrm{MM2}^{(2)}$). For both scenarios, we set $\Delta = 1$ and $K = 5$ (K is the lag number for estimating λ).

The numerical results are presented in Table 5, where the last column T represents the total running time (h—hours, s—seconds). As the results in Table 5 indicate, in both scenarios, $\mathrm{MM2}^{(1)}$ and $\mathrm{MM2}^{(2)}$ are superior both in terms of statistical efficiency and running time: $\mathrm{MM2}^{(1)}$ needs much less computation time and $\mathrm{MM2}^{(2)}$ produces much better estimates. Moreover, under the same three sets of parameter settings, we plot three figures (Figures 1–3) to demonstrate the asymptotic behavior of MM2. In each figure, the x-coordinate is the logarithmic sample size, and the y-coordinate is the logarithmic sample standard deviation (corresponding to a fixed sample size). It can be seen that almost all the regression lines (red lines) are approximately with a slope of $-1/2$, which reflects the asymptotic normality of our method. In addition, we give the following remarks:

- Since GSMLE is an iterative algorithm, there are many hyper-parameters involved which can affect the quality of such an algorithm (e.g., the initial solution, the search

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

Table 6. MM2 estimates for exponential jump

λ	μ	β	a	b	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\beta}$	$\hat{a}$	$\hat{b}$
	(True values)						(Estimates)		
0.1	1	2	3	6	0.0999 ± 0.0017	1.0415 ± 1.0412	1.9976 ± 0.0649	2.9917 ± 0.1901	6.0419 ± 0.3949
0.4	1	2	3	6	0.4000 ± 0.0048	0.9894 ± 0.6761	1.9998 ± 0.0846	3.0054 ± 0.1250	6.0088 ± 0.3287
1	1	2	3	6	1.0002 ± 0.0160	0.9790 ± 0.6705	2.0440 ± 0.2560	3.0038 ± 0.1234	5.9536 ± 0.6778
0.5	**0.5**	2	3	5	0.5001 ± 0.0058	0.5107 ± 0.5555	2.0065 ± 0.0903	2.9984 ± 0.1222	5.0001 ± 0.2740
0.5	**1**	2	3	5	0.4999 ± 0.0056	0.9648 ± 0.4999	2.0007 ± 0.0899	3.0090 ± 0.1090	5.0041 ± 0.2620
0.5	**2**	2	3	5	0.4996 ± 0.0064	2.0099 ± 0.5072	2.0011 ± 0.0915	2.9991 ± 0.1108	5.0127 ± 0.2673
0.1	1	**0.5**	2	4	0.0998 ± 0.0066	0.9962 ± 0.0907	0.5004 ± 0.0126	2.0081 ± 0.1115	3.9981 ± 0.2228
0.1	1	**1**	2	4	0.1002 ± 0.0027	0.9924 ± 0.0027	1.0013 ± 0.0275	2.0089 ± 0.1210	3.9967 ± 0.2431
0.1	1	**2**	2	4	0.0999 ± 0.0017	0.9441 ± 0.4194	2.0080 ± 0.0585	2.0142 ± 0.1246	3.9832 ± 0.2485
0.8	1	2	**0.3**	6	0.7989 ± 0.0114	0.9946 ± 0.0537	2.0268 ± 0.2215	0.3011 ± 0.0110	5.9811 ± 0.6455
0.8	1	2	**1.5**	6	0.8000 ± 0.0119	0.9691 ± 0.2765	2.0153 ± 0.1936	1.5060 ± 0.0533	5.9963 ± 0.5477
0.8	1	2	**9**	6	0.7996 ± 0.0110	0.9884 ± 2.8787	2.0063 ± 0.1928	9.0149 ± 0.4928	6.0342 ± 0.5925
0.6	1	2	3	**4**	0.6005 ± 0.0075	0.9689 ± 0.4332	2.0115 ± 0.1000	3.0097 ± 0.1187	3.9840 ± 0.2260
0.6	1	2	3	**8**	0.6008 ± 0.0075	0.9941 ± 0.7972	2.0265 ± 0.1586	3.0030 ± 0.1109	7.9419 ± 0.6135
0.6	1	2	3	**12**	0.5995 ± 0.0073	0.9846 ± 1.3113	2.0105 ± 0.2448	3.0039 ± 0.1197	12.0946 ± 1.3583

TABLE 7. MM2 estimates for deterministic jump

λ	μ	β (True values)	a	b	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\beta}$ (Estimates)	$\hat{a}$	$\hat{b}$
0.1	1	2	3	6	0.1001 ± 0.0018	0.9547 ± 1.5096	2.0013 ± 0.0890	3.0179 ± 0.2668	6.0171 ± 0.5386
0.4	1	2	3	6	0.3999 ± 0.0048	0.9530 ± 1.0610	2.0054 ± 0.0715	3.0104 ± 0.1825	5.9940 ± 0.3699
1	1	2	3	6	1.0008 ± 0.0178	1.0303 ± 1.2124	2.0173 ± 0.1519	2.9954 ± 0.2109	6.0010 ± 0.5077
0.5	**0.5**	2	3	5	0.5001 ± 0.0060	0.4798 ± 0.9344	2.0028 ± 0.0763	3.0079 ± 0.1950	5.0063 ± 0.3264
0.5	**1**	2	3	5	0.4999 ± 0.0067	0.9595 ± 0.9299	2.0064 ± 0.0785	3.0096 ± 0.1910	4.9956 ± 0.3281
0.5	**2**	2	3	5	0.4996 ± 0.0067	2.0092 ± 0.8429	1.9967 ± 0.0740	2.9999 ± 0.1777	5.0263 ± 0.3070
0.1	1	**0.5**	2	4	0.1004 ± 0.0112	0.9933 ± 0.1058	0.5006 ± 0.0143	2.0080 ± 0.1198	4.0002 ± 0.2357
0.1	1	**1**	2	4	0.1003 ± 0.0037	1.0006 ± 0.2325	1.0005 ± 0.0303	2.0025 ± 0.1304	4.0088 ± 0.2619
0.1	1	**2**	2	4	0.1001 ± 0.0019	0.9510 ± 0.5497	2.0072 ± 0.0703	2.0168 ± 0.1494	3.9862 ± 0.2933
0.8	1	2	**0.3**	6	0.8005 ± 0.0125	0.9947 ± 0.0402	2.0112 ± 0.1173	0.3011 ± 0.0087	5.9740 ± 0.3036
0.8	1	2	**1.5**	6	0.7994 ± 0.0130	0.9690 ± 0.3863	2.0023 ± 0.1060	1.5071 ± 0.0694	5.9971 ± 0.3401
0.8	1	2	**9**	6	0.8007 ± 0.0116	0.9844 ± 6.0227	2.0076 ± 0.1428	9.0297 ± 1.0254	6.0429 ± 0.7087
0.6	1	2	3	**4**	0.6003 ± 0.0078	0.9911 ± 0.6960	2.0018 ± 0.0700	3.0045 ± 0.1807	4.0070 ± 0.2395
0.6	1	2	3	**8**	0.5999 ± 0.0078	0.9780 ± 1.3842	2.0047 ± 0.1004	3.0052 ± 0.1782	8.0093 ± 0.5289
0.6	1	2	3	**12**	0.5999 ± 0.0075	1.0672 ± 2.0520	2.0040 ± 0.1300	2.9972 ± 0.1780	12.0511 ± 0.9172

TABLE 8. MM2 estimates for inverse Gaussian jump

λ	μ	β (True values)	a	b	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\beta}$ (Estimates)	$\hat{a}$	$\hat{b}$
0.1	1	2	3	6	0.1000 ± 0.0017	0.9303 ± 2.0171	2.0016 ± 0.1136	3.0343 ± 0.4243	6.0483 ± 0.8081
0.4	1	2	3	6	0.4001 ± 0.0051	0.9094 ± 1.2803	2.0111 ± 0.0908	3.0254 ± 0.2621	5.9805 ± 0.5388
1	1	2	3	6	1.0005 ± 0.0152	1.0160 ± 1.4023	2.0173 ± 0.1876	3.0083 ± 0.2835	6.0275 ± 0.7802
0.5	**0.5**	2	3	5	0.5002 ± 0.0060	0.4406 ± 1.1146	2.0052 ± 0.0978	3.0256 ± 0.2832	5.0032 ± 0.4893
0.5	**1**	2	3	5	0.5000 ± 0.0064	0.9586 ± 1.0962	2.0029 ± 0.1008	3.0196 ± 0.2755	5.0136 ± 0.4911
0.5	**2**	2	3	5	0.5002 ± 0.0062	1.9760 ± 1.1520	2.0013 ± 0.0962	3.0172 ± 0.2877	5.0218 ± 0.5075
0.1	1	**0.5**	2	4	0.0995 ± 0.0101	1.0360 ± 0.0991	0.4960 ± 0.0138	1.9465 ± 0.1319	4.1233 ± 0.2879
0.1	1	**1**	2	4	0.0999 ± 0.0031	0.9922 ± 0.3414	1.0006 ± 0.0435	2.0181 ± 0.2356	4.0159 ± 0.4543
0.1	1	**2**	2	4	0.1001 ± 0.0018	0.9174 ± 0.9264	2.0108 ± 0.1173	2.0494 ± 0.3335	3.9938 ± 0.5878
0.8	1	2	**0.3**	6	0.7997 ± 0.0113	0.9978 ± 0.0734	2.0050 ± 0.1416	0.3009 ± 0.0159	6.0189 ± 0.5654
0.8	1	2	**1.5**	6	0.8006 ± 0.0112	0.9374 ± 0.5229	2.0226 ± 0.1433	1.5150 ± 0.1075	5.9507 ± 0.5974
0.8	1	2	**9**	6	0.8005 ± 0.0120	0.2303 ± 7.0828	2.0322 ± 0.1957	9.2327 ± 1.4821	5.9708 ± 1.0270
0.6	1	2	3	**4**	0.6002 ± 0.0078	0.9386 ± 1.0064	2.0098 ± 0.1101	3.0362 ± 0.3316	3.9958 ± 0.4634
0.6	1	2	3	**8**	0.5997 ± 0.0075	0.8513 ± 1.1625	2.0093 ± 0.1040	3.0358 ± 0.2359	5.9715 ± 0.5375
0.6	1	2	3	**12**	0.6011 ± 0.0076	0.9295 ± 2.2200	2.0214 ± 0.1523	3.0137 ± 0.2083	11.9445 ± 1.1552

TABLE 9. MM2 estimates for Pareto jump

λ	μ	β (true values)	a	b	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\beta}$ (estimates)	$\hat{a}$	$\hat{b}$
0.1	1	2	3	6	0.1000 ± 0.0031	0.9269 ± 0.1146	1.9045 ± 0.0331	3.1081 ± 0.0725	5.4330 ± 0.5864
0.4	1	2	3	6	0.4005 ± 0.0072	0.9611 ± 0.7830	1.8865 ± 0.0266	3.1009 ± 0.0546	5.2319 ± 0.3168
1	1	2	3	6	1.0014 ± 0.0226	0.9953 ± 0.1167	2.0049 ± 0.0545	3.0008 ± 0.0968	6.1323 ± 0.8204
0.5	**0.5**	2	3	5	0.5008 ± 0.0086	0.4754 ± 0.0749	1.9279 ± 0.0277	3.0696 ± 0.0517	4.6909 ± 0.2116
0.5	**1**	2	3	5	0.5000 ± 0.0085	0.9797 ± 0.0750	1.9284 ± 0.0290	3.0662 ± 0.0537	4.6889 ± 0.2141
0.5	**2**	2	3	5	0.5004 ± 0.0089	1.9765 ± 0.0782	1.9297 ± 0.0286	3.0666 ± 0.0551	4.6915 ± 0.2092
0.1	1	**0.5**	2	4	0.1125 ± 0.1066	0.9991 ± 0.0195	0.5030 ± 0.0429	1.9948 ± 0.2521	3.9959 ± 0.4965
0.1	1	**1**	2	4	0.1005 ± 0.0091	0.9807 ± 0.0352	0.9893 ± 0.0167	2.0374 ± 0.0393	4.0186 ± 0.1951
0.1	1	**2**	2	4	0.1002 ± 0.0029	0.8514 ± 0.0669	1.9166 ± 0.0286	2.1313 ± 0.0390	3.9329 ± 0.1671
0.8	1	2	**0.3**	6	0.8001 ± 0.0167	1.0078 ± 0.0057	2.0785 ± 0.0441	0.2896 ± 0.0070	6.4822 ± 0.4864
0.8	1	2	**1.5**	6	0.8006 ± 0.0167	1.0411 ± 0.0429	2.0757 ± 0.0439	1.4470 ± 0.0382	6.4664 ± 0.6461
0.8	1	2	**9**	6	0.8000 ± 0.0146	1.1714 ± 0.6001	2.0857 ± 0.0671	8.7372 ± 0.3857	7.4321 ± 4.7900
0.6	1	2	3	**4**	0.6006 ± 0.0102	0.9918 ± 0.0927	1.9938 ± 0.0306	3.0105 ± 0.0587	3.9948 ± 0.1339
0.6	1	2	3	**8**	0.6003 ± 0.0110	1.0536 ± 0.1231	1.9757 ± 0.0440	2.9749 ± 0.0915	8.0030 ± 8.4954
0.6	1	2	3	**12**	0.6013 ± 0.0119	1.0725 ± 0.1255	1.9740 ± 0.0459	2.9517 ± 0.1078	11.9578 ± 8.7169

space, the step size, etc.). These hyper-parameters often require substantial efforts of fine-tuning and they are very much problem-dependent. On the other hand, MM2 completely avoids this problem.

- MM2 (or any other moment-based estimation method) requires few assumptions on models and parameters, for example, it does not assume that the volatility process has the Gamma distribution or require any prior distribution for the parameters. This makes it less model-dependent and more applicable. Therefore, it tends to be more robust with respect to modeling errors.

5.3. More Numerical Experiments for MM2

In this subsection, we present extensive numbers of results for MM2. We consider four cases: (1) exponential jump, (2) deterministic jump, (3) inverse Gaussian jump, and (4) Pareto jump, and each case under 15 different parameter settings (by considering three different values for each of the five parameters, λ, μ, β, a, and b). For each parameter setting, we run 400 replications with 100,000 samples for each replication. The numerical results are presented in Tables 6–9. Overall, the results show that MM2 works reasonably well in most cases, with two exceptions: (1) the estimates of μ and (2) the estimates of b for the case with Pareto jump. For the former case, we note that μ is estimated based on $\mu = \mathsf{E}[y_n]/\Delta - \beta a \mathsf{E}[J]$, which contains the product term $\beta a \mathsf{E}[J]$. It could introduce large variances (errors) for the estimates of μ. As for the latter case, since we have to solve a six-degree polynomial Eq. (40), whose coefficients depend on various sample moments and autocovariances. This could introduce large errors for the estimates of b. Of course, as we have demonstrated in the previous subsection, the accuracy of our estimates based on MM2 can be improved if we increase the number of samples.

6. CONCLUSION

In this paper, we study the problem of parameter estimation for the Lévy-driven OU SV model by using the MM. We derive an analytical framework that enables us to derive closed-form formulas for the moments of this model. Based on these formulas, we then develop two types of computationally efficient estimations (MM1 and MM2). Though MM1 is simpler, it uses more high moments than MM2 and, therefore, is much less efficient. We also establish the large-sample results and show that the estimators we developed are strongly consistent and asymptotic normal under moderate assumptions. Finally, we provide extensive numerical results to demonstrate that our moment-based estimators are more efficient than other traditional methods such as MLE, and they produce very good estimates in most instances we tested. Our work in this paper is the first to offer an efficient moment-based method for parameter estimation of the Lévy-driven OU SV model, and it overcomes the issue of statistical inefficiency usually associated with moment-based methods. We remark that the model studied in the paper assumes the underlying Poisson process has a constant arrival rate and the jump size distribution is independent of the arrival process, which may limit the implementation of our method in financial markets. To deal with the time-varying arrival rate, one can consider using a piecewise constant rate to approximate the true rate. If the underlying process is stationary during each "piece", then we can apply our moment-based methods to estimate the parameters separately. For tackling the dependence problem, we point out that more *ad hoc* assumptions on the model structure may be needed, which is beyond the scope of this paper.

Acknowledgments

The authors thank the editor and an anonymous referee for their helpful comments and suggestions. This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grants 71720107003 and 71571048.

References

1. Andersen, T.G., Chung, H.J., & Sørensen, B.E. (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics* 91(1): 61–87.
2. Bansal, R., Gallant, A.R., Hussey, R., & Tauchen, G. (1994). Computational aspects of nonparametric simulation estimation. In Belsley DA (ed.), *Computational Techniques for Econometrics and Economic Analysis*. Advances in Computational Economics, Vol. 3. Dordrecht: Springer, pp. 3–22.
3. Barndorff-Nielsen, O.E. & Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2): 167–241.
4. Billingsley, P. (2008). *Probability and measure*. New York: John Wiley & Sons.
5. Bregantini, D. (2013). Moment-based estimation of stochastic volatility. *Journal of Banking & Finance* 37(12): 4755–4764.
6. Brockwell, P.J., Davis, R.A., & Fienberg, S.E. (1991). *Time series: theory and methods*. New York: Springer Science & Business Media.
7. Carrasco, M. & Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory* 18(1): 17–39.
8. Duffie, D. & Singleton, K.J. (1990). Simulated moments estimation of Markov models of asset prices. *Econometrica* 61(4): 929–952.
9. Durham, G.B. & Gallant, A.R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics* 20(3): 297–338.
10. Fama, E.F. (1965). The behavior of stock-market prices. *The Journal of Business* 38(1): 34–105.
11. Francq, C. & Zakoian, J.M. (2019). *GARCH models: structure, statistical inference and financial applications*. New York: John Wiley & Sons.
12. Frühwirth-Schnatter, S. & Sögner, L. (2009). Bayesian estimation of stochastic volatility models based on OU processes with marginal gamma law. *Annals of the Institute of Statistical Mathematics* 61(1): 159–179.
13. Gallant, A.R. & Tauchen, G. (1996). Which moments to match? *Econometric Theory* 12(4): 657–681.
14. Ghysels, E., Harvey, A.C., & Renault, E. (1996). 5 stochastic volatility. *Handbook of Statistics* 14: 119–191.
15. Griffin, J.E. & Steel, M.F. (2006). Inference with non-Gaussian Ornstein–Uhlenbeck processes for stochastic volatility. *Journal of Econometrics* 134(2): 605–644.
16. Griffin, J.E. & Steel, M.F. (2010). Bayesian inference with stochastic volatility models using continuous superpositions of non-Gaussian Ornstein–Uhlenbeck processes. *Computational Statistics & Data Analysis* 54(11): 2594–2608.
17. Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies* 61(2): 247–264.
18. Haug, S., Klüppelberg, C., Lindner, A., & Zapp, M. (2007). Method of moment estimation in the Cogarch (1,1) model. *The Econometrics Journal* 10(2): 320–341.
19. Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6(2): 327–343.
20. Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications* 7(4): 349–382.
21. Jasra, A., Stephens, D.A., Doucet, A., & Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scandinavian Journal of Statistics* 38(1): 1–22.
22. Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli* 10(1): 97–120.
23. Peng, Y.J., Fu, M.C., & Hu, J.Q. (2014). Gradient-based simulated maximum likelihood estimation for Lévy-driven Ornstein–Uhlenbeck stochastic volatility models. *Quantitative Finance* 14(8): 1399–1414.
24. Peng, Y., Fu, M.C., & Hu, J.Q. (2016). Gradient-based simulated maximum likelihood estimation for stochastic volatility models using characteristic functions. *Quantitative Finance* 16(9): 1393–1411.
25. Roberts, G.O., Papaspiliopoulos, O., & Dellaportas, P. (2004). Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(2): 369–393.

26. Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63(1): 289–306.
27. Spiliopoulos, K. (2009). Method of moments estimation of Ornstein-Uhlenbeck processes driven by general Lévy process. *Annales de l'ISUP* **53**: 3–18.
28. Taylor, S.J. (2008). *Modelling financial time series*. New York: World Scientific.
29. Wu, Y., Hu, J., & Zhang, X. (2019). Moment estimators for the parameters of Ornstein-Uhlenbeck processes driven by compound Poisson processes. *Discrete Event Dynamic Systems* 29(1): 57–77.

APPENDIX: DERIVATIONS OF EQS. (11)–(15)

In this appendix, we provide detailed derivations for Eqs. (11)–(15). We first present the following results which are useful in our derivations:

$$\mathsf{E}[\Delta z] = (\lambda a\Delta)\mathsf{E}[J], \tag{A.1}$$

$$\mathsf{E}[\Delta ez] = a(1-e^{-\lambda\Delta})\mathsf{E}[J], \tag{A.2}$$

$$\mathsf{E}[(\Delta z)^2] = (\lambda a\Delta)\mathsf{E}[J^2] + (\lambda a\Delta)^2(\mathsf{E}[J])^2, \tag{A.3}$$

$$\mathsf{E}[(\Delta ez)^2] = \tfrac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^2] + a^2(1-e^{-\lambda\Delta})^2(\mathsf{E}[J])^2, \tag{A.4}$$

$$\mathsf{E}[\Delta z\Delta ez] = a(1-e^{-\lambda\Delta})\mathsf{E}[J^2] + (\lambda a^2\Delta)(1-e^{-\lambda\Delta})(\mathsf{E}[J])^2, \tag{A.5}$$

$$\begin{aligned}\mathsf{E}[(\Delta z)^2\Delta ez] &= a(1-e^{-\lambda\Delta})\mathsf{E}[J^3] + 3(\lambda a^2\Delta)(1-e^{-\lambda\Delta})\mathsf{E}[J]\mathsf{E}[J^2]\\ &\quad + (\lambda a\Delta)^2 a(1-e^{-\lambda\Delta})(\mathsf{E}[J])^3,\end{aligned} \tag{A.6}$$

$$\begin{aligned}\mathsf{E}[\Delta z(\Delta ez)^2] &= \tfrac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^3] + 2a^2(1-e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2]\\ &\quad + \tfrac{1}{2}(\lambda a^2\Delta)(1-e^{-2\lambda\Delta})\mathsf{E}[J]\mathsf{E}[J^2] + (\lambda a^3\Delta)(1-e^{-\lambda\Delta})^2(\mathsf{E}[J])^3.\end{aligned} \tag{A.7}$$

To illustrate how Eqs. (A.1)–(A.7) can be derived, we only consider Eq. (A.7) in what follows since the others can be derived in a very similar manner.

Suppose that during $[0,\Delta]$, Poisson process $z(t)$ has jumps $J_1, J_2, \ldots, J_{n(\Delta)}$ at $s_1, s_2, \ldots, s_{n(\Delta)}$, respectively ($n(\Delta)$ is a Poisson random variable with mean $a\Delta$). We should point out that $s_1, s_2, \ldots, s_{n(\Delta)}$ are not ordered; hence, given $n(\Delta)$ $s_1, s_2, \ldots, s_{n(\Delta)}$ are i.i.d. uniform random variables over $[0,\Delta]$. We use s to represent a generic uniform random variable over $[0,\Delta]$. We have

$$\mathsf{E}\left[e^{\lambda(s-\Delta)}\right] = \frac{1-e^{-\lambda\Delta}}{\lambda\Delta}, \tag{A.8}$$

$$\mathsf{E}\left[e^{2\lambda(s-\Delta)}\right] = \frac{1-e^{-2\lambda\Delta}}{2\lambda\Delta}, \tag{A.9}$$

and

$$\begin{aligned}\mathsf{E}[\Delta z(\Delta ez)^2] &= \sum_{n=1}^{\infty} e^{-\lambda a\Delta}\frac{(\lambda a\Delta)^n}{n!}\mathsf{E}\left[\left(\sum_{i=1}^{n} J_i\right)\left(\sum_{i=1}^{n} e^{(s_i-1)\lambda\Delta}J_i\right)^2\right]\\ &= \sum_{n=1}^{\infty} e^{-\lambda a\Delta}\frac{(\lambda a\Delta)^n}{n!}n\mathsf{E}[e^{2\lambda(s-\Delta)}]\mathsf{E}[J^3]\\ &\quad + 2\sum_{n=2}^{\infty} e^{-\lambda a\Delta}\frac{(\lambda a\Delta)^n}{n!}n(n-1)(\mathsf{E}[e^{\lambda(s-\Delta)}])^2\mathsf{E}[J]\mathsf{E}[J^2]\\ &\quad + \sum_{n=2}^{\infty} e^{-\lambda a\Delta}\frac{(\lambda a\Delta)^n}{n!}n(n-1)\mathsf{E}[e^{2\lambda(s-\Delta)}]\mathsf{E}[J]\mathsf{E}[J^2]\end{aligned}$$

$$+\sum_{n=3}^{\infty} e^{-\lambda a\Delta}\frac{(\lambda a\Delta)^n}{n!}n(n-1)(n-2)(\mathsf{E}[e^{\lambda(s-\Delta)}])^2(\mathsf{E}[J])^3$$

$$=\frac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^3]+2a^2(1-e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2]$$

$$+\frac{1}{2}(\lambda a^2\Delta)(1-e^{-2\lambda\Delta})\mathsf{E}[J]\mathsf{E}[J^2]+(\lambda a^3\Delta)(1-e^{-\lambda\Delta})^2(\mathsf{E}[J])^3.$$

Hence, we have Eq. (A.7). Based on Eqs. (A.1)–(A.7), we have

$$\mathrm{var}[\Delta z]=(\lambda a\Delta)\mathsf{E}[J^2], \tag{A.10}$$

$$\mathrm{var}[\Delta ez]=\tfrac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^2], \tag{A.11}$$

$$\mathrm{cov}(\Delta z,\Delta ez)=a(1-e^{-\lambda\Delta})\mathsf{E}[J^2], \tag{A.12}$$

$$\mathrm{cov}((\Delta z)^2,\Delta ez)=a(1-e^{-\lambda\Delta})\mathsf{E}[J^3]+2(\lambda a^2\Delta)(1-e^{-\lambda\Delta})\mathsf{E}[J]\mathsf{E}[J^2], \tag{A.13}$$

$$\mathrm{cov}(\Delta z\Delta ez,\Delta ez)=\tfrac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^3]+a^2(1-e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2] + \tfrac{1}{2}(\lambda a^2\Delta)(1-e^{-2\lambda\Delta})\mathsf{E}[J]\mathsf{E}[J^2], \tag{A.14}$$

$$\mathrm{cov}(\Delta z,(\Delta ez)^2)=\tfrac{1}{2}a(1-e^{-2\lambda\Delta})\mathsf{E}[J^3]+2a^2(1-e^{-\lambda\Delta})^2\mathsf{E}[J]\mathsf{E}[J^2]. \tag{A.15}$$

Since $\{v_n, n\geq 1\}$ is stationary, we use a generic v to represent v_n. Noting that v_{n-1} and Δez_n in Eq. (5) are independent of each other, we can obtain:

$$\mathsf{E}[v]=a\mathsf{E}[J], \tag{A.16}$$

$$\mathsf{E}[v^2]=\frac{a}{2}\mathsf{E}[J^2]+a^2(\mathsf{E}[J])^2, \tag{A.17}$$

$$\mathsf{E}[v^3]=\frac{a}{3}\mathsf{E}[J^3]+\frac{3a^2}{2}\mathsf{E}[J]\mathsf{E}[J^2]+a^3(\mathsf{E}[J])^3. \tag{A.18}$$

Based on Eq. (4), we have

$$\mathsf{E}[q_n]=\frac{1}{\lambda}\left[\mathsf{E}[\Delta z_n]-\mathsf{E}[\Delta v_n]\right]=\lambda a\mathsf{E}[J],$$

which gives Eq. (11). Next, we consider Eq. (12). Based on Eqs. (4) and (5), we have

$$q_n=\frac{1}{\lambda}(\Delta z_n-\Delta v_n)=\frac{1}{\lambda}(\Delta z_n-\Delta ez_n+(1-e^{-\lambda\Delta})v_{n-1}),$$

hence

$$\mathrm{var}[q_n]=\frac{1}{\lambda^2}(\mathrm{var}[\Delta z_n-\Delta ez_n]+(1-e^{-\lambda\Delta})^2\mathrm{var}[v_{n-1}])$$

$$=\frac{1}{\lambda^2}(\mathrm{var}[\Delta z_n]+\mathrm{var}[\Delta ez_n]-2\,\mathrm{cov}(\Delta z_n,\Delta ez_n)+(1-e^{-\lambda\Delta})^2\mathrm{var}[v_{n-1}])$$

$$=\frac{a}{\lambda^2}(\lambda\Delta-(1-e^{-\lambda\Delta}))\mathsf{E}[J^2],$$

which gives Eq. (12). Eq. (13) can be verified as follows:

$$\lambda^2\,\mathrm{cov}(q_n,q_{n+1})=\mathrm{cov}(q_n,(1-e^{-\lambda\Delta})v_n)$$

$$=(1-e^{-\lambda\Delta})^2\,\mathrm{cov}(v_{n-1},v_n)+(1-e^{-\lambda\Delta})\,\mathrm{cov}(\Delta z_n-\Delta ez_n,v_n)$$

$$=(1-e^{-\lambda\Delta})^2e^{-\lambda\Delta}\mathrm{var}[v_{n-1}]+(1-e^{-\lambda\Delta})\,\mathrm{cov}(\Delta z_n-\Delta ez_n,\Delta ez_n)$$

$$=\frac{a}{2}(1-e^{-\lambda\Delta})^2\mathsf{E}[J^2].$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

We now consider Eq. (14). First, we have

$$\begin{aligned}\lambda^3 \operatorname{cov}(q_n^2, q_{n+1}) &= \operatorname{cov}(q_n^2, (1-e^{-\lambda\Delta})v_n)\\ &= (1-e^{-\lambda\Delta}) \operatorname{cov}((\Delta z_n - \Delta v_n)^2, v_n)\\ &= (1-e^{-\lambda\Delta})(\operatorname{cov}((\Delta z_n)^2, v_n) + \operatorname{cov}((\Delta v_n)^2, v_n) - 2\operatorname{cov}(\Delta z_n \Delta v_n, v_n)).\end{aligned} \tag{A.19}$$

The three terms in Eq. (A.19) can be calculated as follows:

$$\begin{aligned}\operatorname{cov}((\Delta z_n)^2, v_n) &= \operatorname{cov}(\Delta z_n)^2, \Delta ez_n)\\ &= a(1-e^{-\lambda\Delta})(\mathsf{E}[J^3] + 2(\lambda a \Delta)\mathsf{E}[J]\mathsf{E}[J^2]).\end{aligned} \tag{A.20}$$

$$\begin{aligned}\operatorname{cov}((\Delta v_n)^2, v_n) &= \operatorname{cov}(v_n^2 - 2v_n v_{n-1} + v_{n-1}^2, v_n)\\ &= \operatorname{cov}(v_n^2, v_n) + \operatorname{cov}(v_{n-1}^2, v_n) - 2\operatorname{cov}(v_n v_{n-1}, v_n)\\ &= \operatorname{cov}(v_n^2, v_n) + \operatorname{cov}(v_{n-1}^2, e^{-\lambda\Delta} v_{n-1} + \Delta ez_n)\\ &\quad - 2\operatorname{cov}((e^{-\lambda\Delta} v_{n-1} + \Delta ez_n) v_{n-1}, e^{-\lambda\Delta} v_{n-1} + \Delta ez_n)\\ &= (1 + e^{-\lambda\Delta} - 2e^{-2\lambda\Delta}) \operatorname{cov}(v^2, v)\\ &\quad + e^{-\lambda\Delta}\mathsf{E}[\Delta ez_n]\operatorname{var}[v_n] + \mathsf{E}[v_{n-1}]\operatorname{var}[\Delta ez_n]\\ &= \frac{a}{3}(1-e^{-\lambda\Delta})(1+2e^{-\lambda\Delta})\mathsf{E}[J^3].\end{aligned} \tag{A.21}$$

$$\begin{aligned}\operatorname{cov}(\Delta z_n \Delta v_n, v_n) &= \operatorname{cov}(\Delta z_n((e^{-\lambda\Delta} - 1)v_{n-1} + \Delta ez_n), e^{-\lambda\Delta} v_{n-1} + \Delta ez_n)\\ &= (e^{-\lambda\Delta} - 1)(e^{-\lambda\Delta}\mathsf{E}[\Delta z_n]\operatorname{var}[v_{n-1}] + \mathsf{E}[v_{n-1}]\operatorname{cov}(\Delta z_n, \Delta ez_n))\\ &\quad + \operatorname{cov}(\Delta z_n \Delta ez_n, \Delta ez_n)\\ &= \frac{a(1-e^{-\lambda\Delta})}{2}\left(\lambda a \Delta \mathsf{E}[J]\mathsf{E}[J^2] + (1+e^{-\lambda\Delta})\mathsf{E}[J^3]\right).\end{aligned} \tag{A.22}$$

Substituting Eqs. (A.20)–(A.22) into Eq. (A.19), we can obtain Eq. (14). Finally, we consider Eq. (15). Similar to Eq. (A.19), we have

$$\begin{aligned}\lambda^3 \operatorname{cov}(q_n, q_{n+1}^2) &= \operatorname{cov}(\Delta z_n - \Delta v_n, (\Delta z_{n+1} - \Delta v_{n+1})^2)\\ &= -2\operatorname{cov}(\Delta z_n, \Delta z_{n+1}\Delta v_{n+1}) + \operatorname{cov}(\Delta z_n, (\Delta v_{n+1})^2)\\ &\quad + 2\operatorname{cov}(\Delta v_n, \Delta z_{n+1}\Delta v_{n+1}) - \operatorname{cov}(\Delta v_n, (\Delta v_{n+1})^2).\end{aligned} \tag{A.23}$$

We now calculate the four terms in Eq. (A.23):

$$\begin{aligned}\operatorname{cov}(\Delta z_n, \Delta z_{n+1}\Delta v_{n+1}) &= \operatorname{cov}(\Delta z_n, \Delta z_{n+1}((e^{-\lambda\Delta} - 1)v_n + \Delta ez_{n+1}))\\ &= (e^{-\lambda\Delta} - 1)\operatorname{cov}(\Delta z_n, \Delta z_{n+1} v_n)\\ &= (e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta z_{n+1}]\operatorname{cov}(\Delta z_n, v_n)\\ &= (e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta z_{n+1}]\operatorname{cov}(\Delta z_n, \Delta ez_n)\\ &= -a(1-e^{-\lambda\Delta})^2(\lambda a \Delta)\mathsf{E}[J]\mathsf{E}[^2].\end{aligned} \tag{A.24}$$

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press

$$\begin{aligned}
\operatorname{cov}(\Delta z_n, (\Delta v_{n+1})^2) &= \operatorname{cov}(\Delta z_n, (e^{-\lambda\Delta} - 1)v_n + \Delta e z_{n+1})^2) \\
&= (e^{-\lambda\Delta} - 1)^2 \operatorname{cov}(\Delta z_n, v_n^2) + 2(e^{-\lambda\Delta} - 1) \operatorname{cov}(\Delta z_n, v_n \Delta e z_{n+1}) \\
&= (e^{-\lambda\Delta} - 1)^2 \operatorname{cov}(\Delta z_n, (e^{-\lambda\Delta} v_{n-1} + \Delta e z_n)^2) \\
&\quad + 2(e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta e z_{n+1}] \operatorname{cov}(\Delta z_n, v_n) \\
&= (e^{-\lambda\Delta} - 1)^2 (2e^{-\lambda\Delta}\mathsf{E}[v_{n-1}] \operatorname{cov}(\Delta z_n, \Delta e z_n) + \operatorname{cov}(\Delta z_n, (\Delta e z_n)^2)) \\
&\quad + 2(e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta e z_{n+1}] \operatorname{cov}(\Delta z_n, \Delta e z_n) \\
&= \frac{a}{2}(1 - e^{-\lambda\Delta})^2(1 - e^{-2\lambda\Delta})\mathsf{E}[J^3]. \qquad \textbf{(A.25)}
\end{aligned}$$

$$\begin{aligned}
\operatorname{cov}(\Delta v_n, \Delta z_{n+1}\Delta v_{n+1}) &= \operatorname{cov}(\Delta v_n, \Delta z_{n+1}((e^{-\lambda\Delta} - 1)v_n + \Delta e z_{n+1})) \\
&= (e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta z_{n+1}] \operatorname{cov}(\Delta v_n, v_n) \\
&= -(1 - e^{-\lambda\Delta})^2 \mathsf{E}[\Delta z] var(v) \\
&= -\frac{a}{2}(1 - e^{-\lambda\Delta})^2(\lambda a\Delta)\mathsf{E}[J]\mathsf{E}[J^2]. \qquad \textbf{(A.26)}
\end{aligned}$$

$$\begin{aligned}
\operatorname{cov}(\Delta v_n, (\Delta v_{n+1})^2) &= \operatorname{cov}(\Delta v_n, ((e^{-\lambda\Delta} - 1)v_n + \Delta e z_{n+1})^2) \\
&= (e^{-\lambda\Delta} - 1)^2 \operatorname{cov}(\Delta v_n, v_n^2) + 2(e^{-\lambda\Delta} - 1) \operatorname{cov}(\Delta v_n, v_n \Delta e z_{n+1}) \\
&= (e^{-\lambda\Delta} - 1)^2 (\operatorname{cov}(v_n, v_n^2) - \operatorname{cov}(v_{n-1}, (e^{-\lambda\Delta} v_{n-1} + \Delta e z_n)^2)) \\
&\quad + 2(e^{-\lambda\Delta} - 1)\mathsf{E}[\Delta e z_{n+1}] \operatorname{cov}(\Delta v_n, v_n) \\
&= (e^{-\lambda\Delta} - 1)^2 ((1 - e^{-2\lambda\Delta}) \operatorname{cov}(v, v^2) - 2e^{-\lambda\Delta}\mathsf{E}[\Delta e z]\operatorname{var}[v]) \\
&\quad - 2(e^{-\lambda\Delta} - 1)^2 \mathsf{E}[\Delta e z]\operatorname{var}[v] \\
&= \frac{a}{3}(1 - e^{-\lambda\Delta})^2(1 - e^{-2\lambda\Delta})\mathsf{E}[J^3]. \qquad \textbf{(A.27)}
\end{aligned}$$

Substituting Eqs. (A.24)–(A.27) into Eq. (A.23), we have (15). This completes our derivations for Eqs. (11)–(15).

https://doi.org/10.1017/S0269964820000315 Published online by Cambridge University Press