Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(11): 26994–27015.
DOI: 10.3934/math.20251186
Received: 07 August 2025
Revised: 30 October 2025
Accepted: 05 November 2025
Published: 20 November 2025

Research article

Binomial jump-amplitude modeling in SDEs: A regularized stepwise estimation with computational diagnostics

Wuchen Li[1], Zhaoxiang Xu[2], Jian Xu[3], Linghui Li[1] and Liping Bai[1,*]

[1] Faculty of Innovation Engineering, Macau University of Science and Technology, Taipa, Macau, China
[2] School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, China
[3] South China University of Technology, Guangzhou, China

* **Correspondence:** Email: lipbai@must.edu.mo; Tel: +853-8897 1978.

Abstract: The presence of the jump process makes parameter estimation for stationary stochastic differential equations particularly challenging. Moreover, existing jump parameter models often suffer from significant systematic errors. This paper introduces a new stationary stochastic differential equation model with jumps, in which the jump amplitude follows a binomial distribution. This approach helps mitigate systematic errors, particularly those arising when the probability density remains nonzero for infinitely large jump amplitudes or when it becomes excessively high at zero jump size. On this basis, we use the stepwise estimation method to estimate the parameters of the model (that is, first estimate parameters of the drift and diffusion term by the tool of quadratic variation, and then estimate the parameters of the jump process), and the result has a high estimation accuracy.

Keywords: stochastic differential equations; jump amplitude; quadratic variation; binomial distribution; stepwise parameter estimation
Mathematics Subject Classification: 60H35

1. Introduction

Stochastic differential equations (SDEs) play a crucial role of applications in engineering, physics, biology, and financial data research. For general stable SDEs, if a jump process is incorporated into it, we can obtain the stochastic differential equation with jumps, which is a model can effectively capture the dynamics of signals under the influence of sudden shocks. Its general expression can be written as follows:

$$dX(t) = \mu(X(t^-), \theta)dt + \sigma(X(t^-), \theta)dW(t) + JdM(t), \quad 0 \le t \le T. \tag{1}$$

In the model above, $W(t)$ is the standard Brownian motion [1–4], $X(t^{-})$ means the state variable of the stable stochastic differential equations, $\mu(\cdot, \theta)$ represents the deterministic drift term characterizing the trend of the dynamical system, while $\sigma(\cdot, \theta)$ means the stochastic diffusion term characterizing the random fluctuations of the system induced by stochastic forces. The parameter vector θ is defined on a typically compact set. $JdM(t)$ represents the jump process term, in which J characterizes a random variable displaying the jump amplitude following a kind of probability distribution, and $M(t)$ represents the time when the jump occurs, which also follows a certain probability distribution [5].

Modeling the jump amplitude in jump processes is an important topic in academic research [6–8] . Currently, common kinds of models of the jump amplitude include the normal distribution, log-normal distribution, (double) exponential distribution [9] , generalized double exponential distribution, gamma distribution and Pareto distribution. It is recognized that the model of the log-normal distribution is surely the most commonly applied and widely studied jump size model. People used to think that this was the parametric model which can better fit the jump behavior of financial markets' asset prices [5,10]. However, this method is limited to modeling positively skewed distributions and cannot account for negatively skewed cases.

Among the kinds of jump amplitude models mentioned above, the normal distribution, the double exponential distribution, and the generalized double exponential distribution models can be directly applied to simulate jump amplitudes in both directions. These models' parameters can be adjusted for the convenience of adapting to the characteristics of different jump processes [11]. If the normal distribution has been truncated, it could also be applied to simulate unidirectional jumps, but generally speaking, normal distribution is more commonly applied to simulate bidirectional jumps as it can describe the characteristics of the jump size in both positive and negative directions [5]. Nevertheless, its systematic errors caused by symmetrical distribution compared with the generalized double exponential distribution, may be unfavorable to describing the feature of different jump amplitude distributions in different directions: Certainly, the double exponential distribution also has this problem [10].

In comparison, other models such as the gamma distribution, the Pareto distribution, the log-normal distribution, and the exponential distribution are usually applied to simulate unidirectional jumps for their probability density functions following a skewed distribution and only being meaningful when the independent variable is bigger than zero, so they are appropriate for describing the features of unidirectional jumps. For parameter models that only simulate unidirectional jumps, predecessors usually simulate the jump direction by introducing a direction variable, such as two-point distribution or uniform distribution to simulate the probability of jumps occurring in different directions. Generally speaking, how to choose the jump model depends on the specific questions and data features: the apposite model should be chosen to model the jump process in line with actual demands [11].

However, a key issue with these commonly used jump amplitude models is their significant systematic errors. For instance, when applying the double exponential distribution, the normal distribution, and the generalized double exponential distribution to simulate jump amplitude in different directions simultaneously, these models all have a kind of probability problem: The probability density is too large or even the maximum when the jump amplitude is 0 [11,12]. This problem can cause inaccurate prediction results from the chosen parameter model, especially in fields such as financial markets where negligible jump sizes are unlikely to happen. Even though the log-normal distribution, the Pareto distribution, the gamma distribution, and other parameter models are applied to simulate only unidirectional jumps, systematic errors still exist for the jump size being infinite: for instance, when you use the Pareto distribution model, because its tail decays at a polynomial rate but not an exponential rate, it simulates a unidirectional jump process when the probability density is heavy at the tail, which means that the jump size it allows may be too large [13,14]. When you use the gamma distribution to simulate a unidirectional jump, its probability density is still unbounded on the right side, so the jump amplitude to be infinite, too [15]. Additionally, when applying common models that only describe unidirectional jumps, such as the log-normal distribution and exponential distribution, these models also have infinite jump size; though the probability of their infinite jump size may not be too big, it still exists [5,10]. This may cause inaccurate prediction results of the parameter model, especially in extreme situations [16,17].

In addition to the probability distribution parameter models above, predecessors also have other methods, using stochastic processes [18] to model jump processes including their jump sizes. The Lévy distribution only has systematic errors of infinite jump size, but does not have systematic errors with jump amplitude of size 0, because the domain of the Lévy distribution covers real number axis, and the probability density is 0 at the point of 0 jump amplitude [19]. So the Lévy distribution does not have a non-zero probability density to the location of a jump amplitude of 0. The Weibull distribution could have systematic errors with a jump amplitude of size 0 and an infinite jump amplitude at the same time. Nevertheless, by adjusting the parameters, the Weibull distribution could only have the inevitable systematic error of infinite jump size [20]. Specifically, when shape parameter k of the distribution is larger, the probability density is closer to 0 at the point of 0 jump size, thus reducing the probability density of size 0. At the same time, when λ (the scale parameter) is smaller, the tail of the probability density is fatter, thus increasing the probability density of infinite jump amplitude [21]. So, by selecting suitable parameter values, the Weibull distribution could minimize the probability density of jump amplitude of size 0, and could only have systematic errors of infinite jump size. In general, the Lévy distribution only has a systematic error of infinite large jump size, while the Weibull distribution may exist both kinds of systematic errors, but the probability density of jump amplitude of size 0 could be reduced by replacing the value of parameters.

Based on the above analysis, it is reasonable to consider using alternative probability distributions to model the probability density of jump amplitudes. If binomial distribution is applied to model the distribution of jump size in unidirectional jump, it can avoid the systematic error mentioned above. It is particularly important to note that the binomial distribution could adjust the skewness by adjusting the parameters of it, which is a possible advantage [22]. Although the binomial distribution is not a continuous probability distribution model, under small sample conditions, that is, when the number of jumps is relatively small, the discontinuous distribution and probability model are not without merit.

To estimate the parameters of the complete stochastic differential equation with jumps, we recommend a step-by-step approach. First, smooth the jump shock data, estimate the parameters of the drift and diffusion terms, and then estimate the parameters of the jump process on this basis. This helps to improve the estimation accuracy of the drift and diffusion terms and avoid excessive interference of the jump process on the accuracy of parameter estimation. Before estimating the parameters of the jump process, we attempt to use the Volatility occupation time (VOT) tool to estimate the parameters of the drift and diffusion terms of SDE. VOT is a path-wise analogue of the cumulative distribution function which measures the percentage of time that the quadratic variation of a SDE is less than an artificially defined threshold over a period of time [23,24]. Its definition can be written as:

$$F_t(x) = \int_0^t 1\{V_s \le x\}\,ds, \qquad \forall x \ge 0,\ t \in [0,T]. \tag{2}$$

This tool can be used to measure the fluctuation intensity of a stochastic process over specific time intervals. Though this tool may be difficult to apply, we are inspired by it to simplify it into a quadratic variation tool, and then use it for our research. Take the partitions of the time interval $[a,b]$ as $a = t_0 < t_1 < t_2 < ... < t_N = b$, and then we can get the calculation formula of the quadratic variation for formula (1) as follows with the longest small time interval denoted as $\|\Pi\| = \max_i(t_i - t_{i-1})$:

$$[X,X]_{[a,b]} = \lim_{\|\Pi\|\to 0} \sum_{i=1}^{N} [X(t_i) - X(t_{i-1})]^2 = [X^c, X^c]_{[a,b]} + [J,J]_{[a,b]} \le \text{threshold}, \tag{3}$$

in which $[\cdot,\cdot]$ denotes the quadratic variation, X^c denotes the continuous part of Eq (1) and J denotes right continuous pure jump process of Eq (1) [25]. When the quadratic variation falls below the chosen threshold, we determine that the stochastic process has reached a stationary state.

We perform stepwise parameter estimation for stationary stochastic differential equations with jumps. The process is to first use the quadratic variation tool combined with regularized maximum likelihood estimation to estimate the parameters of the drift and diffusion terms, and then model the jump amplitude as a binomial distribution and estimate the parameters of the jump process.

2. Methodology

2.1. Experimental scenario

The Ornstein–Uhlenbeck (OU) process with jump is used to generate data. This kind of stochastic process is a stationary stochastic process that can be used to simulate the long-term stabilization of currency exchange and rates [26,27]. The Poisson process can be used to simulate the jump process with adding it to the OU process. This is equivalent to modeling a scene of currency exchange and rates with the interference of unexpected event signals [5]. The OU process with jumps can be expressed as:

$$dX(t) = \kappa(\gamma - X(t))\,dt + \sigma\,dW(t) + J\,dN(t), \quad t \in [0,T], \tag{4}$$

in which κ can adjust OU process's returning speed while it should be a real number bigger than 0 (this can ensure the process's returning to the mean value), γ means the average value which the OU process will recover to ultimately, and the σ means the degree of fluctuation of the stochastic process. In our experiments, we set a jump term having a binomial distribution jump amplitude, Bernoulli distribution jump direction, and Poisson distribution jump occurrence time to simulate the jump process, which means that $J = \xi \cdot \eta$, where $\xi \sim$ Bernoulli(p) determines jump direction, $\eta \sim$ Bernoulli(n,p) governs amplitude, and $N(t)$ represents the time when the jump occurs following the Poisson distribution.

2.2. Model formulation

Maximum likelihood estimation (MLE) is the basis in our experimental methodology [28,29]. At first, we discretize the data, and then analyze the discretized OU process on the basis of the Markov assumption [30]. The Euler-Maruyama method is applied to discretize the OU process's data to implement MLE as follows:

$$X_{t_{i+1}} = X_{t_i} + \mu(X_{t_i}, \theta)\Delta t + \sigma(X_{t_i}, \theta)\Delta W_{t_i}, \tag{5}$$

in which $\Delta t = t_{i+1} - t_i$, $\Delta W_{t_i} \sim N(0, \Delta t)$ means the Wiener process's increment over the differentiation interval $[t_i, t_{i+1}]$. The drift term $\mu(\theta, \cdot)$ and diffusion term $\sigma(\theta, \cdot)$ separately characterizes the OU process's dynamics [31]. $X_0, X_1, ..., X_n$ are presumed as the observation values of the state variables at moments $t_0, t_1, ..., t_n$, so we can acquire the conditional distribution of $X_{t_{i+1}}$ based on X_{t_i} as follows:

$$X_{t_{i+1}} \mid X_{t_i} \sim N\left(X_{t_i} + \mu(X_{t_i}, \theta)\Delta t, \sigma(X_{t_i}, \theta)^2 \Delta t\right). \tag{6}$$

2.3. Maximum likelihood estimation and regularization

Likelihood function expression $L(\theta)$'s general form can be acquired for the parameter vector θ due to Markov property assumption with the observation value of state variable X_0, X_1, ... , X_n like below:

$$L_N(\theta \mid X^{\text{known}}) = \prod_{i=1}^{N} p(X(t_i) \mid X(t_{i-1}); \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma(X_{t_i}, \theta)^2 \Delta t}} \exp\left(-\frac{(X_{t_{i+1}} - X_{t_i} - \mu(X_{t_i}, \theta)\Delta t)^2}{2\sigma(X_{t_i}, \theta)^2 \Delta t}\right), \tag{7}$$

where $p(X_{t_i}|Xt_{i-1}; \theta)$ expresses the transition probability density, while the X^{known} expresses our known observation data $X^{known} = (X(t_1), ..., X(t_N))^T$ which correspond to the moments of $\{t_i : i = 1, ..., N\}$ [32]. Afterwards, we acquire the log-likelihood function $L(\theta)$ involving parameter θ due to the observation values of state variables $X_0, X_1, ..., X_n$:

$$\log L_N(\theta \mid X^{\text{known}}) = -\frac{1}{2} \sum_{i=1}^{n-1} \left[\log\left(2\pi\sigma(X_{t_i}, \theta)^2 \Delta t\right) + \frac{(X_{t_{i+1}} - X_{t_i} - \mu(X_{t_i}, \theta)\Delta t)^2}{\sigma(X_{t_i}, \theta)^2 \Delta t}\right]. \tag{8}$$

Ultimately, we hope to solve vector θ's values of the components when $L(\theta)$ realizes its maximum value ($\theta \in \Theta$, $\quad \Theta \subset \mathbb{R}^p$ expresses a compact set), called the parameter estimator of the maximum likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} L_N(\theta \mid X^{\text{known}}). \tag{9}$$

To avoid overfitting and enhance the generalization of the model, in ML (machine learning), regularization operations are often applied [33]. The most commonly employed method of regularization is adding a penalty term to the objective function which disheartens overly complicated models. Penalty term's form depends on the vector θ and is scaled by λ, that is the regularization parameter [34].

These three kinds of regularization operations are commonly employed:

L1 (Lasso Regression) regularization: $R(\theta) = \sum_{i=1}^{p} |\theta_i|$;

L2 (Ridge Regression) regularization: $R(\theta) = \ 1/2 \sum_{i=1}^{p} \theta_i^2$;

Elastic Net Regression: $R(\theta) = \alpha \sum_{i=1}^{p} 1/2 \cdot \theta_i^2 + (1 - \alpha) \sum_{i=1}^{p} |\theta_i|$.

When elastic net regression is used, we set $\alpha = 0.5$ to execute the experiment. On this basis, we should subtract the penalty term from the expression of the original log-likelihood function, then the regularized objective function could be acquired:

$$\text{Regularized Loss} = \text{Log-Likelihood} - \lambda R(\theta),$$

in which Log-likelihood function represents the likelihood of the observation values due to the parameters vector θ, and λ means the penalty coefficient which adjusts the regularization strength. It could be selected via cross-validation or other model selection criteria. $R(\theta)$ expresses the regularization penalty term penalizing excessively large θ values: different regularization methods (e.g., lasso regression, ridge regression, elastic net regression) can penalize parameters by different means [35].

2.4. Determining the starting time of the stationary state and smoothing the data of jump impact segment

Evidently, the shock of the jump could give rise to lots of noise interference to the original OU process, and this will lead to a large error in the parameter estimation of the stationary SDE itself. In the operation of the currency and interest rate markets, the moment when the jump caused by the release of major events news is usually known by all, but the time of this impact's ending is usually unclear. The quadratic variation tool can help to judge when the interfering of the jump shock on the stochastic differential equation ends: If the quadratic variation is stably lower than the suitable judgment threshold within the selected time interval, we can judge that the interference caused by the jump has, on the whole, ended. This can help us rule out the interference and improve the parameter estimation's accuracy of stochastic differential equation. You can witness that we do not filter out the jump before entering the stationary state for the first time, and this method indeed cannot filter out the jump in this stage—the stochastic process we are concerned with oscillates violently before becoming stationary, and in fact, there is no need to screen out the jumps at this stage. What we should pay attention to is the impact interference of the jump process encountered by the overall stationary segment data.

The Eq (3) is used by us to judge when the OU process enters the stationary state and the moment of its recovering from the jump impact. At first, we try to select a suitable judgment threshold. In a time interval of appropriate length, we divide the whole interval into many small equal time groups. Calculate the quadratic variation of all equal time groups, then select their median as the threshold. After the stochastic process's starting, if the quadratic variation of three consecutive small time groups is continuously lower than the judgment threshold, we could judge that the OU process has entered the stationary state at the left endpoint of the first time group where the quadratic variation becomes lower than the threshold. When a jump shock happens, the quadratic variation of each time group is calculated by us step by step. When the quadratic variation value is lower than the judgment threshold, we think that the jump's shock has finished and the stochastic process has recovered to stationary state at the left endpoint of the first time group where the quadratic variation is lower than the threshold.

From the description before, we notice that our criteria for determining the start point of entering the stationary state is stricter than the criteria for determining the end of a jump's interference. We determine that this criterion is logical as a financial product usually needs a period of violent fluctuations before it obtains a suitable price. This initial fluctuation's amplitude is generally higher than the short-term shock of market news (that is, the object modeled by the jump process) on financial prices. Specifically in our experiments, we set the quadratic variation of three consecutive time groups to be steadily and continuously lower than the threshold before we think that the OU process has entered the stationary state. Each time group involves ten steps, which indicates that the quadratic variation of a total of thirty steps should be stably lower than the threshold. When recovering from a jump's impact, we proceed step by step and calculate the quadratic variation of the 10 steps after the jump's occurring as a time group. Based on Eq (3), and the experimental scenario we studied is that the jump time is known, then after the jump occurs, the jump process will no longer contribute to the calculation of the quadratic variation. Therefore, we can get the quadratic variation calculation formula of the time window after the jump interference ends as follows:

$$[X, X]_{[a,b]} = [X^c, X^c]_{[a,b]} = \lim_{\|\Pi\| \to 0} \sum_{i=1}^{N} [X(t_i) - X(t_{i-1})]^2 = \int_a^b \sigma^2(\theta, X(t))\, dt \leq \text{threshold}. \tag{10}$$

The next step is to smooth the jump impact segment. We propose a data substitution method by averaging the data from forward and backward SDE.

1) Forward SDE: We need to substitute all the data between the moment when the jump occurs and when the stationary state is restored. Firstly, we perform MLE on all the data of a stochastic process trajectory, as well as three regularized parameter calculations, then choose the one with the lowest mean square error (MSE). Next, using this chosen parameter estimation method, we do parameter estimation on the segment data before and after the jump impact segment, and average the parameter estimation results of the two segments to get the parameters of the OU process that substitutes the jump impact segment data. The stochastic process that substitutes the original jump impact segment data begins from the last time point before the jump's occurring, and the initial value equals to the real state variable value at this time point. The data at the end of the substituted process is generated by the new OU process itself and has no quantitative relationship with the original real state variable value of the recovery stationary moment. In other words, the end time point of the stochastic process which substitutes the original jump impact segment data is the last time point before the stationary process is restored.

2) Backward-in-time SDE: Though the method above keeps the systematic error of the substitute data endpoint to a certain extent, avoiding over-fitting when the start and end points are strictly constrained, we hope to further relieve the systematic error of the substitute data endpoint on the basis above, so as to further reduce the mean square error of the estimation result of the original OU process to be estimated. So our method is to take the average of the best parameter estimation results before and after the jump impact segment, and then use this set of average values to build a backward-in-time SDE model to substitute the data of the jump impact segment. The start point of this backward-in-time SDE is the recovery stationary moment at the end of the jump impact, and the 'final' point is the time step just next to the jump occurs, which also has no quantitative relationship with the jump's occurs. This backward-in-time SDE can be expressed in the following form:

$$-dX(t) = \frac{\hat{\kappa}(X(t+dt) - \hat{\gamma})}{1 - \hat{\kappa}dt}\, dt - \frac{\hat{\sigma}}{1 - \hat{\kappa}dt}\, dW(t). \tag{11}$$

We will replenish the proof of the above formula form in Appendix. Aso, we need to point out that backward-in-time SDE is not backward SDE, because the former generates stochastic process data in reverse time, while the latter still has a forward time flow. The only similarity between the two is that the final value of the stochastic process at the future moment is the same [36].

In the Python experiment program, Eq (11) should be expressed as below:

$$\begin{aligned} X_{\text{adj}}[i] = X_{\text{adj}}[i+1] + \text{kapa\_est} * (X_{\text{adj}}[i+1] - \text{gama\_est})/(1 - \text{kapa\_est} * dt) * dt \\ -\text{sigma\_est}/(1 - \text{kapa\_est} * dt) * (W[i+1] - W[i]). \end{aligned} \tag{12}$$

Based on this, we average the modeled data generated by the backward-in-time SDE in the jump impact segment with the data of the forward SDE to obtain the substitute segment data set.

In summary, this method averages the data of the simulated forward SDE and the simulated backward-in-time SDE, which clears up the noise of the unidirectional process at each time point to some extent and declines the systematic error before the estimated OU process returns back to the stationary state. However, this approach also introduces new systematic errors. Specifically, the data at the jump occurrence time is not quantitatively related to the final value of the simulated backward-in-time SDE. But generally speaking, the degree of reduction in parameter estimation errors by this method is obviously larger than the new systematic errors it pulls in, and we will do experiments to verify it in the following.

2.5. The parameter estimation of the jump process with jump amplitude obeying the binomial distribution

After completing the parameter estimation of the drift term and diffusion term of the OU process to be estimated, the parameter estimation of the jump process with a jump amplitude obeying the binomial distribution is performed. We use maximum likelihood estimation for the jump amplitudes in different directions, and moment estimation for the probability of different jump directions and the jump occurrence time of the Poisson process simulation, thereby obtaining the estimation results of all parameters of the jump process.

The formula for parameter estimation of the jump process is as follows.

a) Poisson process intensity : $\hat{\lambda} = \frac{N}{T}$, where N is the number of jumps that occurred, T is the total time of observation.

b) Estimation formula for jump direction probability: $\hat{p} = \frac{N_{\text{up}}}{N}$, where N_{up} is the number of upward jumps and N is the total number of jumps. Correspondingly, it is easy to see that the probability of a downward jump is $1 - p$.

c) Binomial distribution parameters for jump sizes (estimated separately for upward and downward): $\hat{n}_{\text{up}} = \max(S^{(i)}_{\text{up}})$, $\hat{p}_{\text{up}} = \frac{\bar{S}_{\text{up}}}{\hat{n}_{\text{up}}}$, $\hat{n}_{\text{down}} = \max(S^{(i)}_{\text{down}})$, $\hat{p}_{\text{down}} = \frac{\bar{S}_{\text{down}}}{\hat{n}_{\text{down}}}$, where S represents the jump size.

In this way, we have completed the parameter estimation of the entire stochastic differential equation with jumps step by step.

3. Experiments

3.1. Validity verification experiment

3.1.1. Single validity verification experiment

We set the total experiment time as $T = 10$, that is to say [0,10] is the experimental time interval. On this basis, we set 1001 time points uniformly distributed from $t = 0$ to $t = 10.0$, so the interval can be divided into 1000 steps of equal scale, meaning each step's size is $dt = 0.01$.

Then the true parameters: $\kappa = 1.0, \gamma = 3.0$, and $\sigma = 1.0$ are set to generate the OU process to be estimated, the strength of the Poisson process is set as 2.0 while the upward and downward jump sizes are set to follow the binomial distribution with parameters (2, 0.4) and (2, 0.6), respectively. The jump direction is set to follow a Bernoulli distribution, where the probability of jumping upward is set to 0.7. The regularization coefficient is set as 0.1 to run our program. The following is the algorithm pseudo code of a single experiment:

Algorithm 1: OU Process with Jumps Parameter Estimation in Single Experiment

Require: OU process parameters: κ, γ, σ, dt, T, λ_{poisson}
Require: Jump parameters: `jump_size_param_up`, `jump_size_param_down`, `jump_size_prob_up`, `jump_size_prob_down`, `jump_direction_up_prob`
Require: Simulation parameters: `n_steps_per_group`, `initial_guess`
Ensure: Estimated parameters with and without jumps, jump process parameters

Simulate OU process with jumps:
Initialize $X[0] \leftarrow 0$
for $t = 1$ to n_{steps} **do**
    if jump occurs at time t **then**
        Sample jump direction and size from binomial distribution
        Update $X[t]$ based on OU dynamics and jump magnitude
    else
        Update $X[t]$ using OU dynamics without jump
    end if
end for
Return X, `jump_times`, `jump_sizes`, `jump_directions`
Find stationary state's start point:
Compute quadratic variation over groups of size `n_steps_per_group`
Define stability threshold as median of quadratic variation
Detect first stable segment with 3 consecutive groups below threshold
Return `stationary_start`
Estimate parameters with jumps:
for each method in {`MLE`, `L1`, `L2`, `Elastic Net`} **do**
    Estimate OU parameters (κ, γ, σ) using log-likelihood
end for
Return `results_with_jumps`
Identify jump segments:
for each jump time after `stationary_start` **do**

Detect end of jump segment where variation returns to stationary
end for
Return `jump_segments`
Adjust data to remove jumps:
for each jump segment in `jump_segments` **do**
Estimate params before and after segment
Replace segment with simulated OU using average parameters
end for
Generate forward and backward adjusted data
Compute average as X_{adjusted}
Return X_{adjusted}
Estimate parameters after jump removal:
for each method in {`MLE`, `L1`, `L2`, `Elastic Net`} **do**
Estimate OU parameters from X_{adjusted}
end for
Return `results_no_jumps`
Estimate jump process parameters:
Calculate λ_{est} as jump rate
Estimate jump size/direction parameters via MLE
Return all estimated jump parameters
Output:
Print parameter estimation table, plot data with jump points, compute MSE

The execution results of Algorithm 1 are as follows (Table 1):

Table 1. Parameter estimation results of the drift and diffusion terms of the OU process before and after bridging substitution for the jump impact segment data in one single experiment.

Method	With jumps			Without jumps (after data adjustment)		
	κ	γ	σ	κ	γ	σ
MLE	1.6044	3.3399	1.6265	1.3600	3.3550	1.3686
L1	1.5841	3.3324	1.6264	1.3454	3.3476	1.3685
L2	1.5771	3.3087	1.6263	1.3463	3.3234	1.3685
Elastic Net	1.5808	3.3205	1.6264	1.3461	3.3354	1.3685
MSE with jumps						
Method	MSE(κ)	MSE(γ)	MSE(σ)	Total MSE		
L2	0.333051	0.095310	0.392287	0.820647		
Elastic Net	0.337313	0.102727	0.392331	0.832371		
L1	0.341224	0.110459	0.392379	0.844062		
MLE	0.365258	0.115565	0.392542	0.873365		
MSE after data adjustment						
Method	MSE(κ)	MSE(γ)	MSE(σ)	Total MSE		
L2	0.119937	0.104595	0.135772	0.360304		
Elastic Net	0.119760	0.112492	0.135782	0.368033		
L1	0.119288	0.120838	0.135792	0.375918		
MLE	0.129580	0.126028	0.135860	0.391468		

True parameters:

OU process: kappa = 1.0000, gamma = 3.0000, sigma = 1.0000;
Jump process: lambda = 2.0000;
Upward_jumps: jump_size_param = 2.0000, jump_size_prob = 0.4000;
Downward jumps: jump_size_param = 2.0000, jump_size_prob = 0.6000.
Jump direction: up_prob = 0.7000, down_prob = 0.3000

Estimated jump process parameters:

lambda = 1.6000;
Upward jumps: jump_size_param = 2.0000, jump_size_prob = 0.4583;
Downward jumps: jump_size_param = 1.0000, jump_size_prob = 0.5000;
Jump direction: up_prob = 0.7500, down_prob = 0.2500.

The following Figure 1 is the stochastic process trajectory generated in our single experiment, including the original trajectory before and after smoothing data replacement:

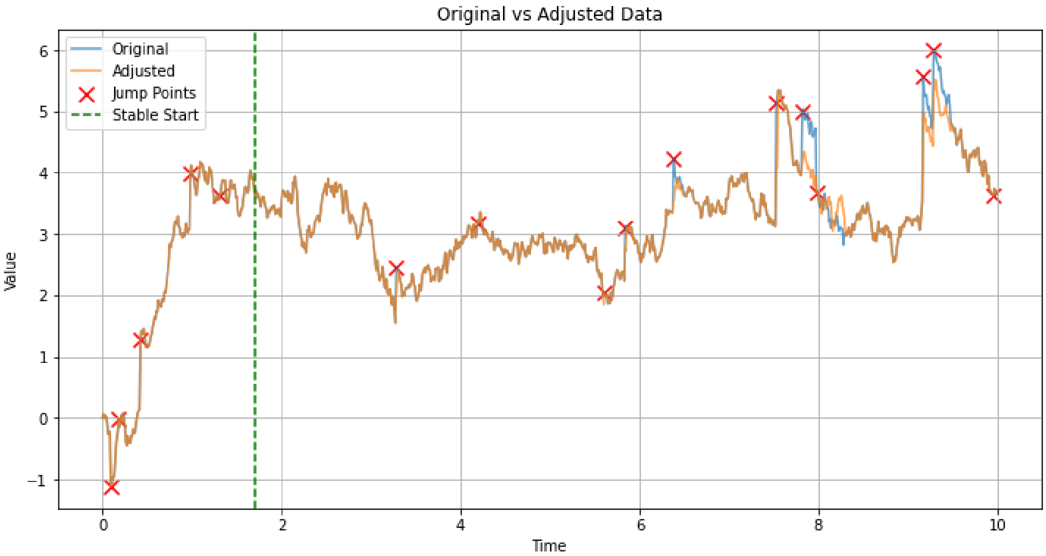

Figure 1. Images of stochastic process trajectories before and after bridging data substitution in a single experiment.

It can be seen from the results of this single experiment that when estimating the parameters of the drift and diffusion terms of the OU process, the accuracy of the result after data substitution smoothing and L2 regularization is improved by about 56% compared to the result of L2 regularization directly based on the original data (from 0.820647 to 0.360304). In addition, the parameter estimation of the jump process has a higher accuracy based on the existing methods.

However, the results of a single experiment are somewhat random, so we adjusted the parameters of the upward and downward jumps to (2, 0.4) and (1, 0.8), repeat the experiment 10 times, and take the average value to further verify the effectiveness of this distribution parameter estimation method.

3.1.2. Ten validity verification experiments and take average

The execution results of Algorithm 2 are as follows (Table 2):

Table 2. Parameter estimation results of the drift and diffusion terms of the OU process before and after bridging substitution for the jump impact segment data of 10 experiments' average.

Method	With jumps			Without jumps (after data adjustment)		
	κ	γ	σ	κ	γ	σ
MLE	1.3864	3.4693	1.7419	1.3803	3.4426	1.6453
L1	1.3695	3.4522	1.7418	1.3639	3.4260	1.6452
L2	1.3677	3.4011	1.7417	1.3614	3.3761	1.6451
Elastic Net	1.3691	3.4261	1.7417	1.3632	3.4003	1.6452
MSE with jumps						
Method	MSE(κ)	MSE(γ)	MSE(σ)	Total MSE		
L2	0.135204	0.160911	0.550053	0.846167		
Elastic Net	0.136215	0.181531	0.550125	0.867871		
L1	0.136507	0.204528	0.550200	0.891235		
MLE	0.14930	0.220258	0.550422	0.919980		
MSE after data adjustment						
Method	MSE(κ)	MSE(γ)	MSE(σ)	Total MSE		
L2	0.130645	0.141474	0.416196	0.688316		
Elastic Net	0.131894	0.160234	0.416242	0.708370		
L1	0.132434	0.181468	0.416293	0.730194		
MLE	0.144631	0.195856	0.416468	0.756955		

True parameters:

OU process: kappa = 1.0000, gamma = 3.0000, sigma = 1.0000;
Jump process: lambda = 2.0000;
Upward jumps: jump_size_param = 2.0000, jump_size_prob = 0.4000;
Downward jumps: jump_size_param = 1.0000, jump_size_prob = 0.8000;
Jump direction: up_prob = 0.7000, down_prob = 0.3000;

Average estimated jump process parameters:

lambda = 2.0500;
Upward jumps: jump_size_param = 2.0000, jump_size_prob = 0.3897;
Downward jumps: jump_size_param = 1.0000, jump_size_prob = 0.7180;
Jump direction: up_prob = 0.7361, down_prob = 0.2639.

Algorithm 2: OU Process with Jumps Parameter Estimation in 10 Experiments and Take Average

Require: Number of experiments $N_{\text{exp}} = 10$, OU parameters: (κ, γ, σ), λ_{poisson}
Require: Jump parameters: $(n_{\text{up}}, p_{\text{up}}, n_{\text{down}}, p_{\text{down}}, p_{\text{dir\_up}})$, time step dt, total time T
Ensure: Averaged estimated OU and jump parameters

Initialize empty lists: `all_results_with_jumps`, `all_results_no_jumps`, `all_best_methods`, `all_jump_params`
for $i = 1$ to N_{exp} **do**
    Generate OU process with jumps X using input parameters
    Compute quadratic variation, find stationary start point
    Estimate parameters with jumps (MLE, L1, L2, Elastic Net)
    Identify jump segments in X
    Adjust X to remove jump effects with forward/backward method
    Estimate parameters without jumps
    Estimate jump process parameters
    Store all results in respective lists
end for
Compute average results:
  `avg_results_with_jumps` ← average of `all_results_with_jumps`
  `avg_results_no_jumps` ← average of `all_results_no_jumps`
  `best_method` ← most frequent in `all_best_methods`
  `avg_jump_params` ← average of `all_jump_params`
Compute MSE for each method in both results
return `avg_results_with_jumps`, `avg_results_no_jumps`, `best_method`, `avg_jump_params`, MSE values

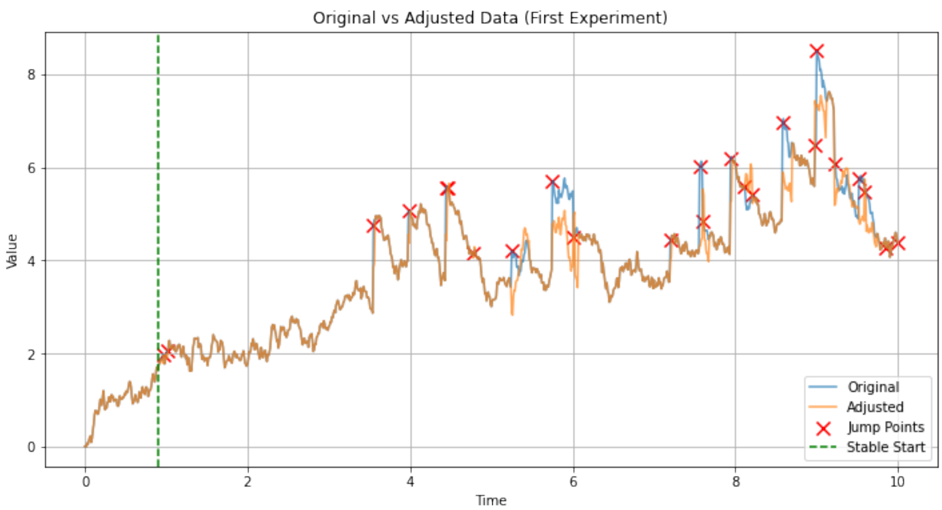

Figure 2. First experiment images of stochastic process trajectories before and after bridging data substitution in 10 experiments.

Figure 2 is the stochastic process trajectory generated by the first experiment of our ten validation experiments, including the original trajectory before and after the smoothing data replacement:

The experimental results show that the average result after 10 experiments, the accuracy of the parameter estimation results for the drift term and the diffusion term after smoothing adjustment, has increased by about 19% (from 0.846167 to 0.688316). And the parameter estimates are closer to the true values.

3.2. Comparison with previously recognized best method

A review paper published in 2022 [37] mentioned that a modified Brownian bridge (from $X(0)$ to $X(t)$) sampler method of Durham and Gallant [38] is recognized as the best stochastic process data augmentation method to date, which can effectively bridge incomplete stochastic process data for parameter estimation. In addition, after referring to the images shown in the previous experiments, it seems that under the parameter conditions of the experimental scenario we selected, selecting the median of the quadratic variation of all time groups as the threshold, determining that the stochastic process has entered a stationary state for the first time, may result in some earlier jumps not being processed, resulting in insufficient parameter estimation accuracy. So we further narrow this threshold: changing it to the 3/2 of the quadratic variation's median of all time groups, and compared it with the case where the threshold is selected as the median, comparing the parameter estimation effect of our method with the best previous method of Durham and Gallant [38]. When encountering jump interference after entering a stationary state for the first time, the threshold for judging the end of the jump shock is the same, that is to say 3/2 of the median, rather than the median. It should be emphasized that in the same set of experiments, the jump impact segments replaced by the two data replacement methods are the same. On this basis, 100 sets of experiments are conducted and the average value is taken to determine the accuracy of different methods. Similar to the previous experiments, L2 regularization always has the highest accuracy, so here we only give the results under L2 regularization in the Table 3 below.

Table 3. Comparison of the accuracy of the two experimental methods under different threshold parameters.

Threshold	Our Method			Total MSE	Durham and Gallant (2002)			Total MSE
	κ	γ	σ		κ	γ	σ	
3/2	1.3350	3.4448	1.6956	0.793847	1.4454	3.4364	1.7800	0.997166
1 (Median)	1.2656	3.5409	1.7345	0.902591	1.3402	3.4003	1.7445	0.830227

Comparing the total MSE of the above experiments, we can see that when the threshold is selected as the median*3/2, the accuracy of our method is significantly higher than the best method of the predecessors, while when the threshold is selected as the median, the parameter estimation accuracy of our method is slightly lower (0.997166 - 0.793847 > 0.902591 - 0.830227). Overall, in the scenario where the jump amplitude obeys a binomial distribution, the parameter estimation accuracy of our data substitution method combined with the backward-in-time stochastic differential equation is higher than the former best method of Durham and Gallant [38].

However, we would like to further remind you that after entering the stationary state for the first time, if the threshold for judging the end of the jump shock state is set to 2/3 of the median, the accuracy of our method will be significantly distorted—in fact, in our related experiments, the total MSE in this case is often greater than 1, even 2, and the estimation accuracy is so low that we no longer need to list it separately. We will give the above error analysis in Appendix.

4. Conclusions and outlook

The method we proposed to smooth the jump segment data of the stationary stochastic differential equation with jumps and then estimate the parameters of the drift and diffusion terms, and then estimate the parameters of the jump process itself should be more effective than estimating parameters based on the original data directly. We optimistically expect that this method will have broad application prospects in fields such as financial mathematics and signal processing.

Author contributions

W.L: Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing—Original Draft; Z.X: Visualization; J.X: Investigation; L.L: Editing; L.B: Funding Acquisition, Supervision, Review.

Use of Generative-AI tools declaration

The authors declare they have used Artificial Intelligence (AI) tools in the creation of this article. AI tools used: The data table format in the final LaTeX manuscript is AI-generated, but the data itself is obtained through true experiments. AI was not used elsewhere.

Acknowledgements

The research work was supported by Science and Technology Development Fund, Macau SAR (No. 0027/2022/AGJ) and (No. 0082/2025/AFJ) & Science and Technology Planning Project of Guangdong Province (2023A0505020007), China.

Conflict of interest

The authors declare no conflict of interest in this paper.

References

1. I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calculus*, New York: Springer-Verlag, 1991.
2. I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calculus*, New York: Springer, 1998. https://doi.org/10.1007/978-1-4612-0949-2
3. B. Øksendal, *Stochastic differential equations: An introduction with applications*, Springer, 2003.
4. D. W. Stroock, S. S. Varadhan, *Multidimensional diffusion processes*, Springer, 1979.

5. R. C. Merton, Option pricing when underlying stock returns are discontinuous, *J. Financ. Econ.*, **3** (1976), 125–144. https://doi.org/10.1016/0304-405X(76)90022-2
6. P. Carr, L. Wu, The finite moment log stable process and option pricing, *J. Finance*, **58** (2003), 753–777. https://doi.org/10.1111/1540-6261.00544
7. A. Dassios, J. Jang, Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity, *Finance Stochast.*, **7** (2003), 73–95. https://doi.org/10.1007/s007800200079
8. F. B. Hanson, J. J. Westman, Jump-diffusion stock return models in finance: Stochastic process density with uniform-jump amplitude, In: *15th International Symposium on Mathematical Theory of Networks and Systems*, 2002.
9. C. A. Ramezani, Y. Zeng, Maximum likelihood estimation of the double exponential jump-diffusion process, *Ann. Finance*, **3** (2007), 487–507. https://doi.org/10.1007/s10436-006-0062-y
10. S. G. Kou, A jump-diffusion model for option pricing, *Manage. Sci.*, **48** (2002), 1086–1101.
11. R. Cont, P. Tankov, *Financial modelling with jump processes*, New York: Chapman and Hall/CRC, 2004. https://doi.org/10.1201/9780203485217
12. Y. Aït-Sahalia, J. Jacod, Estimating the degree of activity of jumps in high frequency data, *Ann. Statist.*, **37** (2009), 2202–2244. https://doi.org/10.1214/08-AOS640
13. E. F. Fama, R. Roll, Some properties of symmetric stable distributions, *J. Amer. Statist. Assoc.*, **63** (1968), 817–836. https://doi.org/10.1080/01621459.1968.11009311
14. X. Gabaix, Power laws in economics and finance, *Annu. Rev. Econ.*, **1** (2009), 255–294. 10.1146/annurev.economics.050708.142940
15. D. B. Madan, E. Seneta, The variance gamma (VG) model for share market returns, *J. Bus.*, **63** (1990), 511–524. https://doi.org/10.1086/296519
16. T. G. Andersen, L. Benzoni, J. Lund, An empirical investigation of continuous-time equity return models, *J. Finance*, **57** (2002), 1239–1284. https://doi.org/10.1111/1540-6261.00460
17. D. S. Bates, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Rev. Financ. Stud.*, **9** (1996), 69–107. https://doi.org/10.1093/rfs/9.1.69
18. D. Applebaum, *Lévy processes and stochastic calculus*, 2Eds, New York: Cambridge University Press, 2009. https://doi.org/10.1017/CBO9780511809781
19. J. P. Nolan, *Stable distributions: Models for heavy-tailed data*, Birkhäuser, 2001.
20. H. Rinne, *The Weibull distribution: A handbook*, New York: Chapman and Hall/CRC, 2008. https://doi.org/10.1201/9781420087444
21. Z. A. Karian, E. J. Dudewicz, *Handbook of fitting statistical distributions with R*, CRC Press, 2010.
22. N. L. Johnson, A. W. Kemp, S. Kotz, *Univariate discrete distributions*, John Wiley & Sons, 2005. https://doi.org/10.1002/0471715816
23. J. Li, V. Todorov, G. Tauchen, Volatility occupation times, *Ann. Statist.*, **41** (2013), 1865–1891. https://doi.org/10.1214/13-AOS1135
24. J. Li, V. Todorov, G. Tauchen, Estimating the volatility occupation time via regularized Laplace inversion, *Economet. Theor.*, **32** (2016), 1253–1288. https://doi.org/10.1017/S0266466615000171
25. S. Shreve, *Stochastic calculus for finance II: Continuous-time models*, Springer, 2004.

26. O. E. Barndorff-Nielsen, N. Shephard, Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics, *J. R. Stat. Soc. B*, **63** (2001), 167–241. https://doi.org/10.1111/1467-9868.00282

27. O. Vasicek, An equilibrium characterization of the term structure, *J. Financ. Econ.*, **5** (1977), 177–188. https://doi.org/10.1016/0304-405X(77)90016-2

28. D. T. Gillespie, Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral, *Phys. Rev. E*, **54** (1996), 2084–2091. https://doi.org/10.1103/PhysRevE.54.2084

29. A. R. Pedersen, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scand. J. Stat.*, **22** (1995), 55–71.

30. C. Fuchs, *Inference for diffusion processes: With applications in life sciences*, Springer, 2013.

31. P. E. Kloeden, E. Platen, *Numerical solution of stochastic differential equations*, Berlin, Heidelberg: Springer, 1992. https://doi.org/10.1007/978-3-662-12616-5

32. Y. Aït-Sahalia, Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach, *Econometrica*, **70** (2002), 223–262. https://doi.org/10.1111/1468-0262.00274

33. T. Hastie, R. Tibshirani, J. H. Friedman, *The elements of statistical learning*, Springer Science & Business Media, 2001.

34. H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. B*, **67** (2005), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

35. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B*, **58** (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

36. S. V. Lototsky, B. L. Rozovskii, *Stochastic partial differential equations*, Springer, 2015.

37. P. Craigmile, R. Herbei, G. Liu, G. Schneider, Statistical inference for stochastic differential equations, *WIREs Comput. Stat.*, **15** (2022), e1585. https://doi.org/10.1002/wics.1585

38. G. B. Durham, A. R. Gallant, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *J. Bus. Econom. Statist.*, **20** (2002), 297–316.

39. W. Li, Y. Jiang, J. Xu, L. Wang, L. Bai, Enhanced parameter estimation for stable SDEs with jumps via quadratic variation, *J. Stat. Comput. Sim.*, **95** (2025), 2932–2948. https://doi.org/10.1080/00949655.2025.2511924

40. W. Li, L. Zhang, J. Xu, L. Li, L. Bai, Jump amplitude inference in SDEs with cosine kernel, *Results Appl. Math.*, **26** (2025), 100596. https://doi.org/10.1016/j.rinam.2025.100596

41. R. Hu, D. Zhang, X. Gu, Reliability analysis of a class of stochastically excited nonlinear Markovian jump systems, *Chaos Soliton. Fract.*, **155** (2022), 111737. https://doi.org/10.1016/j.chaos.2021.111737

42. A. Delaigle, I. Gijbels, Practical bandwidth selection in deconvolution kernel density estimation, *Comput. Stat. Data An.*, **45** (2004), 249–267. https://doi.org/10.1016/S0167-9473(02)00329-8

43. M. C. Jones, J. S. Marron, S. J. Sheather, A brief survey of bandwidth selection for density estimation, *J. Am. Stat. Assoc.*, **91** (1996), 401–407.

44. B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall/CRC, 1986.

45. X. D. Gu, Y. Y. Zhang, I. Mughal, Z. C. Deng, Stochastic responses of nonlinear inclined cables with an attached damper and random excitations, *Nonlinear Dynam.*, **112** (2024), 15969–15986. https://doi.org/10.1007/s11071-024-09877-1

Appendix

A.1. Proof of Eq (11)

The SDE to be estimated can be written into the following form of difference equation:

$$X(t_{i+1}) = X(t_i) + \kappa(\gamma - X(t_i))(t_{i+1} - t_i) + \sigma[W(t_{i+1}) - W(t_i)]. \tag{13}$$

For convenience and without compromising correctness, $t_{i+1} - t_i$ is rewritten as dt, that is to say, $dt = t_{i+1} - t_i$. Then we can rewrite it into reverse form like following:

$$\begin{aligned}
X(t_{i+1}) &= (1 - \kappa(t_{i+1} - t_i))X(t_i) + \kappa\gamma(t_{i+1} - t_i) + \sigma[W(t_{i+1}) - W(t_i)] \\
&= (1 - \kappa \cdot dt)X(t_i) + \kappa\gamma \cdot dt + \sigma[W(t_{i+1}) - W(t_i)], \\
X(t_i) &= \frac{1}{1-\kappa dt}X(t_{i+1}) - \frac{\kappa\gamma}{1-\kappa dt}dt - \frac{\sigma[W(t_{i+1}) - W(t_i)]}{1-\kappa dt} \\
&= \frac{1-\kappa dt}{1-\kappa dt}X(t_{i+1}) + \frac{\kappa dt}{1-\kappa dt}X(t_{i+1}) - \frac{\kappa\gamma}{1-\kappa dt}dt - \frac{\sigma[W(t_{i+1}) - W(t_i)]}{1-\kappa dt} \\
&= X(t_{i+1}) + \frac{\kappa(X(t_{i+1}) - \gamma)}{1-\kappa dt}dt - \frac{\sigma[W(t_{i+1}) - W(t_i)]}{1-\kappa dt}.
\end{aligned} \tag{14}$$

Rewriting the formula above into a differential equation form, we can get:

$$-dX(t) = \frac{\kappa(X(t+dt) - \gamma)}{1-\kappa dt}dt - \frac{\sigma}{1-\kappa dt}dW(t). \tag{15}$$

A.2. Analysis on algorithm errors

(1) It should be noted that if the threshold for determining the first entry into a stationary state is selected too small (for example, ≤ the median of quadratic variation), the jump that occurred early may not be handled yet, so it may cause large error interference to the data replacement as you have seen in the previous comparative experiments. Therefore, it is foreseeable that how to select the quadratic variation threshold that determines the initial entry into a stationary state is a direction worthy of further in-depth research.

(2) After entering the stationary stage for the first time, if the threshold for judging the end of the jump impact is set too small at this stage, the accuracy of parameter estimation will be significantly reduced. Maximum likelihood estimation is the basis of our algorithm. For the OU equation without any jump process interference, the discrete model can be obtained by discretizing it using the Euler-Maruyama method (with a fixed time step Δt):

$$X_{n+1} = (1 - \kappa\Delta t)X_n + \kappa\gamma\Delta t + \sigma\sqrt{\Delta t}\,\varepsilon_n, \tag{16}$$

in which $\varepsilon_n \sim N(0, 1)$ are independent and identically distributed. This model is equivalent to an AR(1) process:

$$X_{n+1} = aX_n + b + \eta_n, \tag{17}$$

in which $a = 1 - \kappa\Delta t, b = \kappa\gamma\Delta t, \eta_n \sim N(0, \sigma^2\Delta t)$.The parameters θ, μ, σ can be estimated from the discrete observation data $\{X_n\}_{n=0}^{N}$ by maximum likelihood estimation (MLE). As $\kappa\Delta t \in (0, 2)$, we can guarantee $|a| < 1$, the AR(1) process is stationary, and its likelihood function satisfies the regularity condition. At this time, the parameters a, b, and η_n can uniquely determine the original OU parameters:

$$\kappa = \frac{1-a}{\Delta t}, \quad \gamma = \frac{b}{\kappa\Delta t}, \quad \sigma = \sqrt{\frac{\mathrm{Var}(\eta_n)}{\Delta t}}.$$

So, the model is identifiable.

The MLE estimator $\hat{\kappa}, \hat{\gamma}, \hat{\sigma}$ is consistent under stationary conditions, that is, when the sample size

$$N \to \infty : \hat{\kappa} \xrightarrow{p} \kappa, \quad \hat{\gamma} \xrightarrow{p} \gamma, \quad \hat{\sigma} \xrightarrow{p} \sigma.$$

That is, these estimators will converge to the true parameters.

The method of Durham and Gallant [38] generated new replacement data by generating new drift and diffusion coefficients through local linearization. Their method does not depend on the amount of data before and after the jump impact segment. However, our method generates the drift and diffusion terms required for replacement data, which depend on the data estimation results before and after the jump impact segment. Therefore, different judgment thresholds will lead to different available data volumes N before and after the jump impact segment. According to the law of large numbers, the closer N is to infinity, the higher the estimation accuracy is. If the threshold is too small, N here may be small, thus reducing the final estimation accuracy. However, if the threshold is selected too large, the interference caused by the jump process in which the jump amplitude obeys the binomial distribution may not be eliminated completely, which may also lead to estimation errors. Therefore, if we want to use our method to get the most accurate results possible, we should try to find the optimal quadratic variation threshold, which may be a direction worthy of further research in the future. And the upper limit of the random error of the algorithm accuracy should be $O(dt^{1/2})$, which is error that the maximum likelihood estimation (MLE) after discretization using the Euler-Maruyama method when the jump process does not occur at all.

(3) Here we explain why we choose the median of the quadratic variation of each segment as the basis for selecting the threshold. The conditional mean and variance formulas of the OU process are:

$$\mu(t) = E[X(t) \mid X(0)] = \gamma + [X(0) - \gamma]e^{-\kappa t}, \quad V(t) = \mathrm{Var}[X(t) \mid X(0)] = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}). \tag{18}$$

When $t \to \infty$, the process reaches a stationary state, and its distribution is a normal distribution that does not depend on the initial value $X(0)$. That is:

$$X(\infty) \sim N\left(\gamma, \frac{\sigma^2}{2\kappa}\right).$$

Considering that σ^2 in Eq (10) is the integrated term, it is highly correlated with the stationary variance of the OU process in the above formula. Equation (10) fluctuates as a measured value, and the stationary variance in the above formula has a factor of 1/2, so we consider the median of the quadratic variance of each segment as the starting point for consideration. However, due to the sharp fluctuations in the initial segment and the subsequent introduction of the interference of the jump term, the actual fluctuation will be larger than the stationary variance of the OU process. Therefore, in the comparative experiment in Section 3.2, we can see that when the median is enlarged to 3/2 of the original, the accuracy of the parameter estimation will be further improved. Table 4 shows the average threshold changes and parameter estimation accuracy changes after 100 experiments for readers' reference. The parameter settings of the OU process with jumps are the same as those in Sections 3.1.2–3.2. It can be observed that when the scaling factor of the threshold is between 3/2 and 2, the decrease in total MSE is not significant, while the MSE of the γ term will increase significantly. Therefore, we estimate that the optimal scaling factor is approximately around this range.

Table 4. Changes in parameter estimation accuracy under different thresholds using forward-backward smoothing and L2 regularization over 100 experiments.

Threshold (multiple of median)	MSE(κ)	MSE(γ)	MSE(σ)	Total MSE
1/2	0.313895	0.215794	3.90979	4.43948
2/3	0.238910	0.207184	2.81288	3.25897
1	0.070543	0.292573	0.539490	0.902591
3/2	0.112225	0.197847	0.483859	0.793847
2	0.028535	0.411214	0.286359	0.726108

(4) The discussion on the asymptotic consistency and error analysis of the jump amplitude parameter is the same as that in (5) below.

(5) A common misconception is that readers might want the error caused by endpoint mismatch to be as small as possible, perhaps roughly comparable to the error caused by the randomness of Brownian motion (may be approximated as $O(\sigma \cdot dt^{1/2})$). Our original intention was to overcome this misconception and demonstrate that by freeing our minds, we can achieve even better results than the modified Brownian bridge method proposed by previous researchers.

Based on the simulation results of Li et al. [39], using the same OU process experimental scenario, but modifying the jump to be unidirectional and setting the jump amplitude to a fixed value of 1.5, the parameter estimation results using only forward SDE data replacement has been slightly better than those of Durham & Gallant [38]. The interference caused by this jump magnitude in the OU process is significantly greater than the interference caused by the binomial jump amplitude used in this paper. Building on Li et al. [39], we further introduced the backward-in-time SDE used in this paper and replaced the jump interference period by taking the forward and backward averages. The 100 times experimental results are as follows: after regularizing with L2 term, the accuracy (MSE) can be improved by 21.86% on average from 0.725033 to 0.566508. In contrast, Li et al. [39] reported an average accuracy improvement of 8.06% on average from 0.768505 to 0.706550 when using only forward data for data substitution, which has been a bit better than the accuracy of Durham & Gallant's modified Brownian bridge [38]. Based on the comparative analysis of the above data results, we believe that when the jump amplitude interferes with the original OU process more, the accuracy of the forward and backward averaging method proposed in this paper will further increase compared to the modified Brownian bridge method. This shows that endpoint mismatch has a positive impact on error analysis in this problem, while overly strict endpoint constraints in the modified Brownian bridge method can have a more negative impact on parameter estimation.

A.3. Verification of the model's practicality

(1) This paper uses simulated data from an OU process generated using the Euler-Maruyama method to verify the effectiveness of the algorithm. However, since real-world stochastic processes do not necessarily follow an OU process, it is crucial to validate the model assumptions when applying the algorithm to empirical data. A recommended practice is to first fit the OU model to the observed data and then perform diagnostic checks, such as testing the normality and independence of the model residuals. Subsequently, cross-validation should be conducted by splitting the data into training and test sets to evaluate the model's predictive accuracy and generalization performance.

(2) For a unidirectional jump distribution, the chi-square distribution can be used to test whether it follows a binomial distribution. For continuous-time stochastic processes, some researchers may think that it is not reasonable to use this discrete model to simulate jumps. Here we give our explanation: The binomial distribution is not only suitable for modeling symmetric or finite-support jumps, but can also be used to approximate continuous skewed distributions through appropriate parameter tuning. Continuous skewed distributions (such as the log-normal or gamma distribution) are often used in financial markets to describe the asymmetric nature of asset price jumps. As a discrete distribution, the binomial distribution can approximate these characteristics to a certain extent by increasing the number of trials, n, and optimizing the probability of success, p. For example, when n is large and p is biased to one side (e.g., $p < 0.5$ or $p > 0.5$), the probability mass function of the binomial distribution exhibits skewness, similar to that of a mildly skewed continuous distribution. By discretizing the jump range into a finite number of states and leveraging the central limit theorem, the standardized form of the binomial distribution, $(Y_n - np)/[np(1-p)]^{1/2}$ (where $Y_n \sim Binomial(n, p)$), can be shown to converge weakly to a normal distribution. By adjusting the combination of p and n, the cumulative distribution function of skewed distributions can be further approximated. To quantify this approximation, the error is typically related to the magnitude of $n^{-1/2}$, indicating that the accuracy of the approximation gradually improves as n increases. This property makes the binomial distribution useful for modeling jump processes with mild skewness, especially under small sample conditions (such as the OU process simulation in this study), thus providing a flexible tool for parameter estimation of jump magnitudes in financial data. If readers have other concerns, they can refer to the studies of Li et al. [40] and Hu et al. [41] to propose more reasonable improvement models[42–44].

(3) The SDE employed in this study uses the Itô form, which is consistent with the traditional financial mathematics on which this study is based and is suitable for modeling jump processes. However, the Stratonovich integral, commonly used in engineering, differs somewhat from the Itô integral. Therefore, scholars in the engineering field should be aware of the potential errors that may arise from this distinction when referring to this study [45].