

Robust Reinforcement Learning under Diffusion Models for Data with Jumps

Chenyang Jiang

Department of Statistics, University of Wisconsin-Madison
Madison, WI 53706, USA

cjiang@wisc.edu

and

Donggyu Kim

Department of Economics, University of California, Riverside
Riverside, CA 92521, USA

donggyu.kim@ucr.edu

and

Alejandra Quintos*

Department of Statistics, University of Wisconsin-Madison
Madison, WI 53706, USA

alejandra.quintos@wisc.edu

and

Yazhen Wang

Department of Statistics, University of Wisconsin-Madison
Madison, WI 53706, USA

yzwang@stat.wisc.edu

September 19, 2025

Abstract

Reinforcement Learning (RL) has demonstrated considerable success in tackling complex decision-making problems across diverse domains. However, continuous-time RL remains challenging, particularly when the underlying state dynamics are modeled by stochastic differential equations (SDEs) with jump components. In this paper, we address these challenges by proposing the Mean-Square Bipower Variation Error (MSBVE) algorithm, which enhances robustness and convergence in environments with significant stochastic noise and jumps. We begin by revisiting the widely

*Supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation.

used Mean-Square Temporal Difference Error (MSTDE) algorithm and highlight its susceptibility to bias when jump components are present. In contrast, the proposed MSBVE algorithm minimizes the bipower variation error that has the form of a product of non-overlapping absolute differences, allowing it to more accurately estimate value functions in the presence of jumps. Theoretical analysis and numerical studies confirm that MSBVE outperforms MSTDE in jump-diffusion settings. These results emphasize the critical role of alternative error metrics in advancing the resilience and accuracy of RL algorithms in continuous-time stochastic systems with jumps.

Keywords: bipower variation, continuous time, jump process, stochastic differential equations, temporal difference

1 Introduction

Reinforcement Learning (RL) has emerged as a powerful framework for solving intricate decision-making challenges across a variety of domains, including robotics, finance, and healthcare. In RL, an agent interacts with a dynamic environment and learns to select actions that maximize cumulative rewards over time. A central objective in this paradigm is the accurate estimation of the value function, which quantifies the expected return associated with each state under a given policy. Precise value function estimation is crucial for policy improvement and long-term performance.

Several algorithms for estimating the value function within the RL framework have been developed for purely continuous diffusion processes. For instance, [Doya \(2000\)](#) introduced the temporal-difference (TD) algorithm tailored for the continuous-time deterministic case, while [Jia & Zhou \(2021\)](#) elucidated the convergence properties of the Mean Square TD Error (MSTDE) algorithm within continuous-time stochastic settings. These algorithms endeavor to estimate the value function by iteratively updating a parameter vector using distinct error metrics, thereby shedding light on their convergence behavior under diverse noise conditions. The optimal parameter derived from these algorithms ensures that the estimated value function closely approximates the true value function.

However, in practice, observed data often exhibit jumps, which makes the pure continuous diffusion component unobservable. These discontinuities pose significant challenges for standard RL algorithms. In many applications, especially in finance, the primary interest lies in learning models based on the latent continuous component. For instance, volatility processes are crucial for quantifying investment risk. Since jumps typically reflect unpredictable news events, they are often treated as noise, and attention is focused on estimating the volatility of the continuous part. At the very least, given the fundamentally different characteristics of the continuous and jump components, they should be learned separately.

See [Andersen et al. \(2007\)](#), [Corsi et al. \(2010\)](#), [Oh et al. \(2024\)](#). The latent continuous process captures the typical dynamics that we aim to learn. Therefore, it is crucial to develop robust RL estimation methods that can accurately recover the latent continuous component in the presence of jumps.

In this paper, we delve into the task of estimating the value function within RL frameworks operating in continuous-time settings, where state dynamics are delineated by stochastic differential equations (SDEs) potentially incorporating jump components. We begin by reviewing the Mean Square Temporal Difference Error (MSTDE) algorithm, which aims to estimate the value function by minimizing the mean-square TD error. However, we highlight its limitations in environments with jump-driven dynamics.

To overcome these challenges, we propose the Mean Square Bipower Variation Error (MSBVE) algorithm, which minimizes the bipower variation error defined as a product of non-overlapping absolute differences. This approach enhances robustness to stochastic noise and discontinuities, which leads to more accurate value function estimation in the presence of jumps. Through rigorous theoretical analysis and comprehensive simulation studies, we demonstrate the effectiveness of the MSBVE algorithm in accurately estimating the value function under challenging conditions characterized by substantial variability arising from jump components. Our results clearly show that MSBVE outperforms the MSTDE algorithm, particularly in environments where substantial jump-induced noise hinders MSTDE’s convergence. These findings highlight the critical importance of adopting alternative error metrics, such as the bipower variation error, to enhance the robustness and accuracy of RL algorithms in continuous-time settings with jump components.

The key distinction of our work lies in its emphasis on robustness. While recent studies have primarily focused on deriving optimal control policies in jump-diffusion settings, our approach aims to develop an algorithm that ensures the estimated value function remains

resilient to jump-induced noise. This leads to a more stable and reliable estimation framework for general SDEs with jumps and helps learn the latent continuous diffusion process. In contrast, [Gao et al. \(2024\)](#) and [Bo et al. \(2024\)](#) employ randomized control methods alongside q-learning to derive the optimal policy distribution for general SDEs with jumps, differing mainly in their choice of entropy terms. [Bender & Thuan \(2023\)](#) narrows its focus to the mean-variance portfolio selection problem, providing theoretical optimal solutions but not addressing general SDEs with jumps. [Guo et al. \(2023\)](#) introduces the Greedy Least-Square algorithm and derives sublinear regret bounds, though its model assumes that only the drift term depends on the state and control, excluding the diffusion and jump components. Lastly, [Denkert et al. \(2024\)](#) replaces control with a random point process, applying an actor-critic algorithm to solve a randomized problem that ultimately leads to the optimal value functions for the original system. Unlike these works, which concentrate on finding the optimal control under jump diffusions, our research addresses the challenge of value function estimation under jump dynamics and emphasizes robustness against jump noise.

The remainder of this paper is structured as follows. Section 2 provides an overview of the setup for the RL problem within continuous-time settings. In Section 3, we present the detailed descriptions of the MSTDE and MSBVE algorithms. Section 4 reports the simulation results and analyzes the performance of these algorithms. In Section 5, we apply the MSBVE algorithm to the mean-variance portfolio selection problem and compare its performance with MSTDE using the real intraday trading data. Finally, we draw conclusions in Section 6, summarizing the key findings and implications of our study. All the detailed proofs are given in appendix A.

2 Model Setup

In a general RL problem, the state space and action space are denoted by \mathcal{S} and \mathcal{A} , respectively. We consider a finite time horizon $[0, T]$, where $T < \infty$ is fixed throughout. At time t , the environment provides the current state $X_t \in \mathcal{S}$. The agent then selects an action $A_t = \pi(X_t) \in \mathcal{A}$ according to policy π . A reward $R_t = r(t, X_t, A_t)$ is calculated based on this action. Subsequently, the environment generates the next state X_{t+1} based on the previous state X_t and action A_t . In the underlying true model, the state X_t can be described by the following SDE, defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{t \geq 0})$:

$$dX_t = b(t, X_t, A_t)dt + \sigma(t, X_t, A_t)dW_t.$$

However, in practice, the observed data often contain noise introduced by jumps. This can be modeled by an SDE with jumps as follows:

$$dX_t = b(t, X_t, A_t)dt + \sigma(t, X_t, A_t)dW_t + L(t, X_t, A_t)dN_t.$$

In real-world scenarios, practical constraints often limit the ability to frequently change policies. For instance, in financial trading, risk management policies may dictate specific trading strategies that cannot be altered frequently. Therefore, this paper focuses on an RL problem with a fixed policy under the continuous-time settings. Since the policy π is fixed, $A_t = \pi(X_t)$ is a function of X_t , we use the following SDE without the action A_t to describe the state X_t :

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + L(t, X_t)dN_t, \tag{1}$$

where W_t is a m -dimensional \mathcal{F}_t -adapted Brownian motion, N_t is a n -dimensional \mathcal{F}_t -adapted counting process independent of W_t , and $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$, and $L : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times n}$ are the drift coefficient, diffusion coefficient, and jump

size, respectively. The value function $J : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^k$ is defined as:

$$J(t, x) = E \left[\int_t^T r(s, X_s) ds + h(X_T) | X_t = x \right], \quad (2)$$

where $r : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^k$ is a reward function based on the current time and state, and $h : \mathbb{R}^d \rightarrow \mathbb{R}^k$ is a reward function applied at the end of T .

We make the following standard assumptions.

Assumption 1 *The following conditions hold true.*

- (i) b, σ, L, r , and h are all continuous functions in their respective arguments;
- (ii) b and σ are uniformly Lipschitz in x , i.e., there exists a constant $C > 0$ such that

$$|b(t, x) - b(t, x')| + |\sigma(t, x) - \sigma(t, x')| \leq C|x - x'|, \quad \forall t \in [0, T], \quad x, x' \in \mathbb{R}^d;$$

- (iii) b and σ both have linear growth in x , i.e., there exists a constant $C > 0$ such that

$$|b(t, x)|^2 + |\sigma(t, x)|^2 \leq C^2(1 + |x|^2), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d;$$

- (iv) r, h , and L have polynomial growth in x , i.e., there exist constants $C > 0$ and $\nu \geq 1$ such that for all $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$|r(t, x)| \leq C(1 + |x|^\nu), \quad |h(x)| \leq C(1 + |x|^\nu), \quad |L(t, x)| \leq C(1 + |x|^\nu);$$

- (v) The total number of jumps N_T is finite with probability 1, i.e. $P(N_T < \infty) = 1$. Moreover, for all $i = 0, \dots, n - 1$, for any time interval $[t_i, t_{i+1}]$ with width $\Delta t = t_{i+1} - t_i > 0$, the probability of there existing exactly one jump within this interval tends to zero as $\Delta t \rightarrow 0$, i.e. $P(N_{t_{i+1}} - N_{t_i} = 1) \rightarrow 0$ as $\Delta t \rightarrow 0$.

Assumption 1(i)-(iii) guarantee the existence and uniqueness of a solution to the SDE in (1) (see Theorem 5.2.9 in Karatzas & Shreve (2014)). Assumption 1(iv) ensures that $J(t, x)$

is finite for any (t, x) . Assumption 1(v) ensures that the jump process has finite jumps in finite time.

Recall that J can be characterized by the following PDE based on the Feynman-Kac formula (Karatzas & Shreve (2014)):

$$\begin{cases} \mathcal{L}J(t, x) + r(t, x) = 0, & (t, x) \in [0, T] \times \mathbb{R}^d, \\ J(T, x) = h(x), \end{cases} \quad (3)$$

where

$$\mathcal{L}J(t, x) := \frac{\partial J}{\partial t}(t, x) + b(t, x) \cdot \frac{\partial J}{\partial x}(t, x) + \frac{1}{2} \sigma^2(t, x) \cdot \frac{\partial^2 J}{\partial x^2}(t, x).$$

Assumption 2 *The above PDE admits a classical solution $J \in C^{1,2}([0, T] \times \mathbb{R}^d)$ satisfying the polynomial growth condition, i.e., there exist constants $C > 0$ and $\mu \geq 1$ such that*

$$|J(t, x)| \leq C(1 + |x|^\mu), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d.$$

Assumption 2 guarantees the existence of a solution to the PDE in (3).

To analyze this RL problem, we discretize the time interval $[0, T]$ into $[t_i, t_{i+1}]$ with equal time step Δt , where $i = 0, 1, 2, \dots, n-1$ and $0 = t_0 < t_1 < \dots < t_n = T$. The goal is to estimate the value function J using a parametric family of functions J^θ with $\theta \in \Theta \subset \mathbb{R}^l, l \in \mathbb{N}^+$ (which could be reduced in complexity if we have some prior knowledge about the value function or the underlying dynamics) based on the observed states X_{t_i} and rewards $R_{t_i} = r(t_i, X_{t_i})$, without the knowledge of the explicit functional form of b, σ, L, r , and h . A key challenge is to develop an algorithm that remains robust to jumps introduced by noisy observed data, even though the true underlying model follows an SDE without jumps. We need the following assumption for J^θ . First, introduce

$$L^2_{\mathcal{F}}([0, T]) = \left\{ \kappa = \{\kappa_t, 0 \leq t \leq T\} \text{ is real-valued and } \mathcal{F}_t\text{-progressively measurable:} \right. \\ \left. E \int_0^T |\kappa_t|^2 dt < \infty \right\}.$$

Define the L^2 -norm $\|\kappa\|_{L^2} = (E \int_0^T |\kappa_t|^2 dt)^{1/2}$ on the space $L^2_{\mathcal{F}}([0, T])$.

Assumption 3 $J^\theta(t, x)$ is a smooth function of (t, x, θ) with all the derivatives existing. $E|J^\theta(t, X_t)|^2$ is a continuous function of θ for all $t \in [0, T]$ and $\theta \in \Theta$. Moreover, for all $\theta \in \Theta$, $\mathcal{L}J^\theta(\cdot, X), |\frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X)| \in L^2_{\mathcal{F}}([0, T])$, and their L^2 -norms are continuous functions of θ .

Assumption 3 ensures that $E|J^\theta(t, X_t)|^2$, $\|\mathcal{L}J^\theta(\cdot, X)\|_{L^2}$, and $\|\frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X)\|_{L^2}$ are finite in any given compact set Θ .

3 Theoretical Analysis of Algorithms

We commence our exploration by revisiting the Temporal-Difference (TD) Error proposed by Doya (2000) and the Mean-Square TD Error (MSTDE) algorithm, along with its convergence behavior described by Jia & Zhou (2021).

3.1 Deterministic Setting

In the deterministic setting, the state X_t follows an ODE:

$$dX_t = b(t, X_t)dt.$$

Since the states X_t are deterministic, the expectation equals itself, and the equation (2) becomes

$$J(t, X_t) = \int_t^T r(s, X_s)ds + h(X_T) = J(t + \Delta t, X_{t+\Delta t}) + \int_t^{t+\Delta t} r(s, X_s)ds.$$

Moving $J(t, X_t)$ to the right side and dividing both sides by Δt , we have

$$\frac{J(t + \Delta t, X_{t+\Delta t}) - J(t, X_t)}{\Delta t} + \frac{1}{\Delta t} \int_t^{t+\Delta t} r(s, X_s)ds = 0. \quad (4)$$

As $\Delta t \rightarrow 0$, the left side of (4) becomes

$$\delta_t := \frac{d}{dt}J(t, X_t) + r(t, X_t),$$

which is the the well-known TD Error. When estimating the value function J using J^θ with parameter θ , the TD Error becomes

$$\delta_t^\theta := \frac{\partial}{\partial t} J^\theta(t, X_t) + r(t, X_t).$$

Building upon the above TD Error, [Doya \(2000\)](#) introduced the mean-square TD error (MSTDE):

$$\text{MSTDE}(\theta) := \frac{1}{2} \int_0^T |\delta_t^\theta|^2 dt = \frac{1}{2} \int_0^T \left| \frac{\partial}{\partial t} J^\theta(t, X_t) + r(t, X_t) \right|^2 dt.$$

The optimal θ is determined by minimizing the above MSTDE, since if J^θ is a perfect estimator of the true value function J , then δ_t^θ would be zero. To solve the RL problem using the observed discretized data X_{t_i} and $r(t_i, X_{t_i})$, [Jia & Zhou \(2021\)](#) defined the $\text{MSTDE}_{\Delta t}(\theta)$ which is an approximation of $\text{MSTDE}(\theta)$:

$$\text{MSTDE}_{\Delta t}(\theta) := \frac{1}{2} \sum_{i=0}^{n-1} \left[\frac{J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i})}{t_{i+1} - t_i} + r(t_i, X_{t_i}) \right]^2 \Delta t,$$

and proposed an algorithm to update θ by minimizing the $\text{MSTDE}_{\Delta t}(\theta)$ using the gradient descent method.

3.2 Stochastic Setting with Continuous Dynamics

The state X_t follows a continuous SDE:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{5}$$

We first recall the findings of [El Karoui et al. \(1997\)](#), the PDE in [\(3\)](#) is connected to the following forward-backward stochastic differential equation (FBSDE):

$$\begin{cases} dX_s = b(s, X_s)ds + \sigma(s, X_s)dW_s, & s \in [t, T]; X_t = x, \\ dY_s = -r(s, X_s)ds + Z_s dW_s, & s \in [t, T]; Y_T = h(X_T). \end{cases} \tag{6}$$

The value function $J(s, X_s)$ and the solution $\{(X_s, Y_s, Z_s), t \leq s \leq T\}$ of the above FBSDE have the following relationship:

$$\begin{cases} Y_s = J(s, X_s), & s \in [t, T], \\ Z_s = \frac{\partial J}{\partial x}(s, X_s)^\dagger \sigma(s, X_s), & s \in [t, T]. \end{cases} \quad (7)$$

For any fixed $(t, x) \in [0, T] \times \mathbb{R}^d$ and $\{X_s, t \leq s \leq T\}$ solving the first equation of (6), define

$$M_s := J(s, X_s) + \int_t^s r(u, X_u) du, \quad s \in [t, T]. \quad (8)$$

By combining (6) and (7), we observe the following result:

$$\begin{aligned} dM_s &= dJ(s, X_s) + r(s, X_s) ds \\ &= dY_s + r(s, X_s) ds \\ &= Z_s dW_s. \end{aligned}$$

Assumption 1 guarantees that $E[\int_0^T Z_s^2 ds] < \infty$, which implies M_s is a martingale. The martingale property implies that for any $\Delta t > 0$,

$$E[M_{t+\Delta t} | \mathcal{F}_t] = M_t,$$

and by plugging in the formula for M , we have

$$E \left[J(t + \Delta t, X_{t+\Delta t}) + \int_t^{t+\Delta t} r(s, X_s) ds \middle| \mathcal{F}_t \right] = J(t, X_t).$$

Similar to (4), we obtain

$$E \left[\frac{J(t + \Delta t, X_{t+\Delta t}) - J(t, X_t)}{\Delta t} + \frac{1}{\Delta t} \int_t^{t+\Delta t} r(s, X_s) ds \right] = 0.$$

As Δt goes to zero, $\frac{d}{dt} J(t, X_t)$ does not exist due to the non-differentiability of Brownian motion. Although the MSTDE does not theoretically exist in this setting, we can still define $\text{MSTDE}_{\Delta t}(\theta)$ similar to the deterministic case as follows:

$$\text{MSTDE}_{\Delta t}(\theta) := \frac{1}{2} E \left[\sum_{i=0}^{n-1} \left(\frac{J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i})}{t_{i+1} - t_i} + r(t_i, X_{t_i}) \right)^2 \Delta t \right].$$

Similar to (8), we can define

$$M_s^\theta := J^\theta(s, X_s) + \int_t^s r(u, X_u) du.$$

Then, we have

$$\begin{aligned} \text{MSTDE}_{\Delta t}(\theta) &= \frac{1}{2\Delta t} E \left[\sum_{i=0}^{n-1} (J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i}) + r(t_i, X_{t_i})\Delta t)^2 \right] \\ &\approx \frac{1}{2\Delta t} E \left[\sum_{i=0}^{n-1} (J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i}) + \int_{t_i}^{t_{i+1}} r(s, X_s) ds)^2 \right] \\ &= \frac{1}{2\Delta t} E \left[\sum_{i=0}^{n-1} (M_{t_{i+1}}^\theta - M_{t_i}^\theta)^2 \right] \\ &\rightarrow \frac{1}{2\Delta t} E \langle M^\theta \rangle_T. \end{aligned}$$

The above limit holds as the size of the partition goes to 0. This provides the intuition that minimizing the $\text{MSTDE}_{\Delta t}(\theta)$ will guide the optimal parameter θ^* towards minimizing the expected quadratic variation of M^θ . By Itô's formula, we have

$$dM_t^\theta = [\mathcal{L}J^\theta(t, X_t) + r(t, X_t)] dt + \left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) dW_t.$$

Hence, we have

$$\begin{aligned} d\langle M^\theta \rangle_t &\equiv (dM_t^\theta)^2 = \left\{ [\mathcal{L}J^\theta(t, X_t) + r(t, X_t)] dt + \left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) dW_t \right\}^2 \\ &= \left[\left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) \right]^2 (dW_t)^2 + [\mathcal{L}J^\theta(t, X_t) + r(t, X_t)]^2 (dt)^2 \\ &\quad + 2 [\mathcal{L}J^\theta(t, X_t) + r(t, X_t)] dt \cdot \left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) dW_t \\ &= \left[\left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) \right]^2 dt + \text{high-order small term,} \end{aligned}$$

where the last line is due to $(dW_t)^2 = dt$ a.s. and the fact that the terms involving $(dt)^2$ and $dW_t dt$ are of higher order than dt . Therefore, minimizing the expected quadratic variation of M^θ is equivalent to minimizing $E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) \right|^2 dt \right]$. Since the

reward function $r(t_i, X_{t_i})$ actually does not contribute to the quadratic variation of M^θ as seen before, we use the following error term instead of $\text{MSTDE}_{\Delta t}(\theta)$:

$$\text{MSTDE}_{\Delta t}^*(\theta) := E \left[\sum_{i=0}^{n-1} \left(J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i}) \right)^2 \right]. \quad (9)$$

Jia & Zhou (2021) proved the following result (Theorem 2 in Jia & Zhou (2021)):

Suppose the state X_t follows the equation (5) and Assumptions 1-3 hold. Let $\theta_{\text{MSTDE}^*}^*(\Delta t) \in \arg \min_{\theta \in \Theta} \text{MSTDE}_{\Delta t}^*(\theta)$, and assume that $\theta_{\text{MSTDE}^*}^* := \lim_{\Delta t \rightarrow 0} \theta_{\text{MSTDE}^*}^*(\Delta t)$ exists. Then, we have

$$\theta_{\text{MSTDE}^*}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_t)}{\partial x} \right)^\dagger \sigma(t, X_t) \right|^2 dt \right].$$

3.3 Stochastic Setting with Jump Noise

In practice, jumps are frequently observed and are typically triggered by unexpected news or events. These jumps exhibit a fundamentally different nature from the continuous fluctuations captured by continuous diffusion processes. As a result, the presence of jumps complicates the task of learning the dynamics driven by the latent continuous diffusion process in (5). To address this, we assume that the true underlying state evolves according to a diffusion process without jumps, while the observed data are contaminated by an additive jump component as in (1). Our objective is to develop a robust algorithm that can effectively account for and mitigate the impact of jump noise in the observed process. For example, when X_t represents a stock price, we want to learn dynamics from the rate of the return, represented by $\int_0^T b(t, X_t) dt$, and from the random fluctuations due to the market volatility, represented by $\int_0^T \sigma(t, X_t) dW_t$, rather than to the jump components, represented by $\int_0^T L(t, X_t) dN_t$. However, in the presence of jumps, the MSTDE algorithm converges to the expected quadratic variation of M^θ , which differs from the target quantity of interest. In the following theorem, we formally characterize the bias introduced by the MSTDE algorithm under jump contamination.

Theorem 1 Suppose the state X_t follows the equation (1) and Assumptions 1–3 hold. Let

$$\theta_{MSTDE}^*(\Delta t) \in \arg \min_{\theta \in \Theta} MSTDE_{\Delta t}^*(\theta),$$

and assume that $\theta_{MSTDE}^* := \lim_{\Delta t \rightarrow 0} \theta_{MSTDE}^*(\Delta t)$ exists. Then, we have

$$\theta_{MSTDE}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt + \int_0^T [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right], \quad (10)$$

where $X_{t-} := \lim_{\Delta t \searrow 0} X_{t-\Delta t}$ is the left limit of X_t .

This result reflects the setting where we only observe the jump-contaminated process X_t , as given by equation (1), rather than the latent continuous path. Theorem 1 reveals that the MSTDE algorithm suffers from two sources of bias: the jump variation term, $\int_0^T [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t$, and the distortion in the gradient evaluation due to discontinuities in the observed process, $\frac{\partial J^\theta(t, X_{t-})}{\partial x}$.

To address the issue posed by the presence of jump components in the state dynamics, we introduce the Mean-Square Bipower Variation Error (MSBVE) as an alternative error term.

The MSBVE is defined as follows:

$$MSBVE_{\Delta t}(\theta) := E \left[\sum_{i=1}^{n-1} |J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i})| |J^\theta(t_i, X_{t_i}) - J^\theta(t_{i-1}, X_{t_{i-1}})| \right]. \quad (11)$$

Unlike MSTDE, which considers the squared difference between consecutive value function estimates, the MSBVE approach utilizes the product of non-overlapping absolute differences. Since jumps are rare events, there is at most one jump within a short time interval. By leveraging this property, the use of non-overlapping absolute differences helps eliminate the influence of jump components. As a result, MSBVE is capable of mitigating the impact of jump noise in the observed process. The following theorem formally demonstrates the advantage of the MSBVE approach in jump-contaminated environments.

Theorem 2 Suppose the state X_t follows the equation (1) and Assumptions 1–3 hold. Let

$$\theta_{MSBVE}^*(\Delta t) \in \arg \min_{\theta \in \Theta} MSBVE_{\Delta t}(\theta),$$

and assume that $\theta_{MSBVE}^* := \lim_{\Delta t \rightarrow 0} \theta_{MSBVE}^*(\Delta t)$ exists. Then, we have

$$\theta_{MSBVE}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right].$$

Theorem 2 shows that the proposed MSBVE estimator θ_{MSBVE}^* is free from the bias associated with the jump variation term, $\int_0^T [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t$. In other words, compared to the MSTDE approach, the MSBVE approach exhibits greater robustness to the presence of jump components. The effectiveness of MSBVE will be further illustrated through numerical studies presented in Sections 4 and 5. In the next section, we analyze the remaining bias arising from the gradient term of the value function, $\frac{\partial J^\theta(t, X_{t-})}{\partial x}$.

3.4 Bias Analysis

We can decompose the observed jump-diffusion process X_t as follows:

$$dX_t = dX_t^C + dX_t^J,$$

where $dX_t^C = b(t, X_t)dt + \sigma(t, X_t)dW_t$ and $dX_t^J = L(t, X_t)dN_t$. That is, X_t^C and X_t^J represent the continuous and jump components, respectively. Based on the latent continuous diffusion process, as discussed in Section 3.2, the oracle estimator is defined as follows:

$$\theta_{\text{Oracle}}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_t^C)}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right].$$

The proposed MSBVE estimation procedure still has a bias compared with the oracle estimator. The bias is coming from the difference between $J^\theta(t, X_{t-})$ and $J^\theta(t, X_t^C)$. More specifically, for the general vector case, by Taylor expansion, we have

$$\nabla_x J^\theta(t, X_t)^\top \nabla_x J^\theta(t, X_t) = \nabla_x J^\theta(t, X_t^C)^\top \nabla_x J^\theta(t, X_t^C) + 2\nabla_x J^\theta(t, X_t^C)^\top H_x J^\theta(t, X_t^C) X_t^J$$

$$+ (\text{tr}(H_x J^\theta(t, X_t^C) H_x J^\theta(t, X_t^C)) + \nabla_x J^\theta(t, X_t^C)^\top \nabla_x (\text{tr}(H_x J^\theta(t, X_t^C)))) \|X_t^J\|^2 + \dots,$$

where, for any differentiable function $g : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\nabla_x g(t, x_0) = \left. \frac{\partial g(t, x)}{\partial x} \right|_{x=x_0}$$

is the gradient, and

$$H_x g(t, x_0) = \left. \frac{\partial^2 g(t, x)}{\partial x \partial x^\top} \right|_{x=x_0}$$

is the Hessian matrix. Then, the bias comes from the series:

$$\text{bias}(\theta) := \sum_{i=1}^{\infty} \frac{f^{(i)}(X_t^C)}{i!} (X_t^J)^{\otimes i},$$

where $f^{(i)}(X_t^C)$ represents the i th derivative tensor of $\nabla_x J^\theta(t, X_t^C)^\top \nabla_x J^\theta(t, X_t^C)$ with respect to x , and $(X_t^J)^{\otimes i}$ denotes the i th order tensor product of X_t^J . Thus, the solution of the MSBVE procedure is

$$\theta_{\text{MSBVE}}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^T \left| \left(\frac{\partial J^\theta(t, X_t^C)}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right] + \text{bias}(\theta),$$

where the bias is a multivariate Taylor's expansion. From this result, we can find that when the value function is linear in x , the bias term $\text{bias}(\theta)$ disappears. Generally, when the value function is polynomial in x , higher-order terms remain. We will examine how this bias behaves in the following simulation and application sections.

4 Simulation Study

In our simulation setup, we consider a state process X_t , defined on the interval $t \in [0, 1]$, which evolves according to the SDE:

$$dX_t = dW_t + X_{t-} dN_t,$$

where W_t is a standard Brownian Motion and N_t is a Poisson process independent of W_t . We initialize the process with $X_0 = 0.1$. We only select those Poisson process paths where

exactly one jump occurs in $[0, 1]$, which implies the jump occurs uniformly within the interval $[0, 1]$. Under this setting, we have

$$X_t - X_{t-} = \begin{cases} X_{t-}, & \text{if } \Delta N_t = 1, \\ 0, & \text{if } \Delta N_t = 0. \end{cases}$$

Note that under this model, the jump size at time t is equal to X_{t-} , meaning the process doubles at the jump time: $X_t = 2X_{t-}$. Prior to the jump, however, X_t evolves a Brownian Motion starting from $X_0 = 0.1$. We compare the performance of the algorithms under three different forms of the value function:

(i) Linear:

$$J^\theta(t, x) = (\theta(1 - t) + 1)x;$$

(ii) Quadratic:

$$J^\theta(t, x) = \theta(1 - t)x^2 + x;$$

(iii) Exponential:

$$J^\theta(t, x) = \theta(1 - t)e^x + x.$$

For the MSBVE algorithm, we utilize stochastic gradient descent (SGD) to update the parameter θ using the error term defined in (11):

$$\text{MSBVE}_{\Delta t}(\theta) := E \left[\sum_{i=1}^{n-1} |J_{i+1}^\theta - J_i^\theta| \cdot |J_i^\theta - J_{i-1}^\theta| \right],$$

where

$$J_i^\theta = J^\theta(t_i, X_{t_i}).$$

The update rule for θ following from SGD is

$$\begin{aligned} \theta \leftarrow \theta - \alpha \times \sum_{i=1}^{n-1} & \left[\frac{\partial}{\partial \theta} (J_{i+1}^\theta - J_i^\theta) |J_i^\theta - J_{i-1}^\theta| \text{sgn}(J_{i+1}^\theta - J_i^\theta) \right. \\ & \left. + \frac{\partial}{\partial \theta} (J_i^\theta - J_{i-1}^\theta) |J_{i+1}^\theta - J_i^\theta| \text{sgn}(J_i^\theta - J_{i-1}^\theta) \right]. \end{aligned}$$

For the MSTDE algorithm, the error term in (9) becomes

$$\text{MSTDE}_{\Delta t}^*(\theta) := E \left[\sum_{i=0}^{n-1} (J_{i+1}^\theta - J_i^\theta)^2 \right],$$

and the update rule for θ follows from SGD becomes

$$\theta \leftarrow \theta - 2\alpha \times \sum_{i=0}^{n-1} (J_{i+1}^\theta - J_i^\theta) \cdot \frac{\partial}{\partial \theta} (J_{i+1}^\theta - J_i^\theta).$$

4.1 Linear Value Function

When $J^\theta(t, x) = (\theta(1-t) + 1)x$, based on the discussion of Theorem 2, the theoretical minimizer of MSBVE is

$$\begin{aligned} \theta_{\text{MSBVE}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right] \\ &= \arg \min_{\theta \in \Theta} \int_0^1 [(\theta(1-t) + 1)]^2 dt \\ &= \arg \min_{\theta \in \Theta} \left(\frac{1}{3}\theta^2 + \theta + 1 \right) \\ &= -\frac{3}{2}, \end{aligned}$$

and based on the discussion of Theorem 1, the theoretical minimizer of MSTDE is

$$\begin{aligned} \theta_{\text{MSTDE}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt + \int_0^1 [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right] \\ &= \arg \min_{\theta \in \Theta} E \left[\int_0^1 [(\theta(1-t) + 1)]^2 dt + \int_0^1 [(\theta(1-t) + 1)(X_t - X_{t-})]^2 dN_t \right] \\ &= \arg \min_{\theta \in \Theta} \left\{ \frac{1}{3}\theta^2 + \theta + 1 \right. \\ &\quad \left. + E \left[E \left(\int_0^1 [(\theta(1-t) + 1)(X_t - X_{t-})]^2 dN_t \middle| \text{the jump happens at time } u \right) \right] \right\} \\ &= \arg \min_{\theta \in \Theta} \left[\frac{1}{3}\theta^2 + \theta + 1 + E_{U \sim \text{Unif}[0,1]} ((\theta(1-U) + 1)^2 E(X_U^2 | \text{the jump happens at time } u)) \right] \\ &= \arg \min_{\theta \in \Theta} \left[\frac{1}{3}\theta^2 + \theta + 1 \right. \\ &\quad \left. + E_{U \sim \text{Unif}[0,1]} ((\theta(1-U) + 1)^2 E((W_U + 0.1)^2 | \text{the jump happens at time } u)) \right] \end{aligned}$$

$$\begin{aligned}
&= \arg \min_{\theta \in \Theta} \left[\frac{1}{3} \theta^2 + \theta + 1 + E_{U \sim \text{Unif}[0,1]} ((\theta(1-U) + 1)^2 (U + 0.01)) \right] \\
&= \arg \min_{\theta \in \Theta} \left(\frac{21}{50} \theta^2 + \frac{403}{300} \theta + \frac{151}{100} \right) \\
&= -\frac{403}{252}.
\end{aligned}$$

It is worth noting that, since the value function is linear in x , the bias term discussed in Section 3.4 vanishes. As a result, the MSBVE estimator achieves the oracle solution:

$$\theta_{\text{Oracle}}^* = \theta_{\text{MSBVE}}^* = -\frac{3}{2}.$$

4.2 Quadratic Value Function

When $J^\theta(t, x) = \theta(1-t)x^2 + x$, based on the discussion of Theorem 2, the theoretical minimizer of MSBVE is

$$\begin{aligned}
\theta_{\text{MSBVE}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} E \left[\int_0^1 [2\theta(1-t)X_{t-} + 1]^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} E \left[\int_0^u E(2\theta(1-t)(W_t + 0.1) + 1)^2 dt \right. \\
&\quad \left. + \int_u^1 E(2\theta(1-t)(W_t + W_u + 0.2) + 1)^2 dt \middle| \text{the jump happens at time } u \right] \\
&= \int_0^1 \int_0^u (4\theta^2(1-t)^2(t + 0.01) + 0.4\theta(1-t) + 1) dt du \\
&\quad + \int_0^1 \int_u^1 (4\theta^2(1-t)^2(3u + t + 0.04) + 0.8\theta(1-t) + 1) dt du \\
&= \arg \min_{\theta \in \Theta} \left(\frac{167}{300} \theta^2 + \frac{4}{15} \theta + 1 \right) \\
&= -\frac{40}{167},
\end{aligned}$$

and based on the discussion of Theorem 1, the theoretical minimizer of MSTDE is

$$\theta_{\text{MSTDE}}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt + \int_0^1 [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right]$$

$$\begin{aligned}
&= \arg \min_{\theta \in \Theta} \left[\frac{167}{300} \theta^2 + \frac{4}{15} \theta + 1 \right. \\
&\quad \left. + E \left[(3\theta(1-u)(W_u + 0.1)^2 + (W_u + 0.1))^2 \middle| \text{the jump happens at time } u \right] \right] \\
&= \arg \min_{\theta \in \Theta} \left[\frac{167}{300} \theta^2 + \frac{4}{15} \theta + 1 \right. \\
&\quad \left. + \int_0^1 (9\theta^2(1-u)^2(3u^2 + 0.06u + 0.0001) + 6\theta(1-u)(0.3u + 0.001) + (u + 0.01)) du \right] \\
&= \arg \min_{\theta \in \Theta} \left(\frac{45059}{30000} \theta^2 + \frac{1709}{3000} \theta + \frac{151}{100} \right) \\
&= -\frac{8545}{45059}.
\end{aligned}$$

For the latent continuous process, $dX_t^C = dW_t$, therefore, the oracle estimator is

$$\begin{aligned}
\theta_{\text{Oracle}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_t^C)}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} E \left[\int_0^1 [2\theta(1-t)(W_t + 0.1) + 1]^2 dt \right] \\
&= \int_0^1 (4\theta^2(1-t)^2(t + 0.01) + 0.4\theta(1-t) + 1) dt \\
&= \arg \min_{\theta \in \Theta} \left(\frac{26}{75} \theta^2 + \frac{1}{5} \theta + 1 \right) \\
&= -\frac{15}{52}.
\end{aligned}$$

Unlike the linear value function case, here the value function is nonlinear in x , so the bias term discussed in Section 3.4 does not vanish. As a result, the oracle solution and the MSBVE estimate differ: $\theta_{\text{Oracle}}^* \neq \theta_{\text{MSBVE}}^*$.

4.3 Exponential Value Function

When $J^\theta(t, x) = \theta(1-t)e^x + x$, based on the discussion of Theorem 2, the theoretical minimizer of MSBVE is

$$\theta_{\text{MSBVE}}^* \in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right]$$

$$\begin{aligned}
&= \arg \min_{\theta \in \Theta} E \left[\int_0^1 [\theta(1-t)e^{X_{t-}} + 1]^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} E \left[\int_0^u E(\theta(1-t)e^{W_t+0.1} + 1)^2 dt \right. \\
&\quad \left. + \int_u^1 E(\theta(1-t)e^{W_t+W_u+0.2} + 1)^2 dt \right] \text{ the jump happens at time } u \\
&= \arg \min_{\theta \in \Theta} \left[\int_0^1 \int_0^u (\theta^2(1-t)^2 e^{2t+0.2} + 2\theta(1-t)e^{\frac{1}{2}t+0.1} + 1) dt du \right. \\
&\quad \left. + \int_0^1 \int_u^1 (\theta^2(1-t)^2 e^{(6u+2t+0.4)} + 2\theta(1-t)e^{\frac{1}{2}(3u+t)+0.2} + 1) dt du \right] \\
&= \arg \min_{\theta \in \Theta} [3.190\theta^2 + 1.657\theta + 1] \approx -0.260,
\end{aligned}$$

and based on the discussion of Theorem 1, the theoretical minimizer of MSTDE is

$$\begin{aligned}
\theta_{\text{MSTDE}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_{t-})}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt + \int_0^1 [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right] \\
&= \arg \min_{\theta \in \Theta} [(3.190\theta^2 + 1.657\theta + 1) \\
&\quad + E \left((\theta(1-u)(e^{2W_u+0.2} - e^{W_u+0.1}) + W_u + 0.1)^2 \right) \text{ the jump happens at time } u \\
&= \arg \min_{\theta \in \Theta} \left[(3.190\theta^2 + 1.657\theta + 1) \right. \\
&\quad + \int_0^1 \theta^2(1-u)^2 (e^{8u+0.4} - 2e^{\frac{9}{2}u+0.3} + e^{2u+0.2}) du \\
&\quad \left. + \int_0^1 (2\theta(1-u)((2u+0.1)e^{2u+0.2} - (u+0.1)e^{\frac{1}{2}u+0.1}) + (u+0.01)) du \right] \\
&= \arg \min_{\theta \in \Theta} [7.607\theta^2 + 2.965\theta + 1.505] \approx -0.195.
\end{aligned}$$

For the latent continuous process, $dX_t^C = dW_t$, therefore, the oracle estimator is

$$\begin{aligned}
\theta_{\text{Oracle}}^* &\in \arg \min_{\theta \in \Theta} E \left[\int_0^1 \left| \left(\frac{\partial J^\theta(t, X_t^C)}{\partial x} \right)^\dagger \sigma(t, X_{t-}) \right|^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} E \left[\int_0^1 [\theta(1-t)e^{W_t+0.1} + 1]^2 dt \right] \\
&= \arg \min_{\theta \in \Theta} \int_0^1 (\theta^2(1-t)^2 e^{2t+0.2} + 2\theta(1-t)e^{\frac{1}{2}t+0.1} + 1) dt
\end{aligned}$$

$$\begin{aligned}
&= -\frac{\int_0^1 (1-t)e^{\frac{1}{2}t+0.1} dt}{\int_0^1 (1-t)^2 e^{2t+0.2} dt} \\
&\approx -0.901.
\end{aligned}$$

As in the quadratic case, the value function is nonlinear in x , so the bias introduced by jump contamination remains. Consequently, the MSBVE estimator does not recover the oracle solution: $\theta_{\text{Oracle}}^* \neq \theta_{\text{MSBVE}}^*$.

4.4 Results

In our simulation study, we set the time grid size as $\Delta t = 0.001$, learning rate $\alpha = 0.0005$, and initialize $\theta_0 = 0.5$. We train the model for 100000 episodes with 32 sample paths generated for each episode.

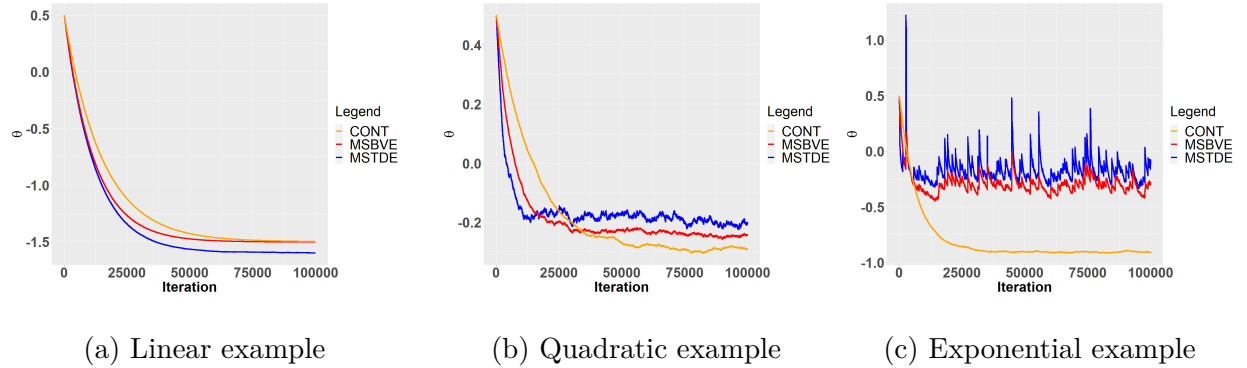


Figure 1: The convergence of θ of three different forms of value functions (linear, quadratic, and exponential)

The results shown in Figure 1 illustrate the convergence behavior of θ under three different forms of the value function. In Figure 1(a), where the value function is linear in x , the bias term $\text{bias}(\theta)$ vanishes. Consequently, the red line (MSBVE) and the orange line (Oracle estimator) converge to the same value, $-\frac{3}{2}$. However, as the value function has a high-degree polynomial, the bias increases. For instance, in Figure 1(b), where the value

function is quadratic in x , the red line (MSBVE) converges to $-\frac{40}{167}$, which is close to the blue line (MSTDE) that converges to $-\frac{8545}{45059}$. In contrast, the oracle estimator (orange line) converges to $-\frac{15}{52}$, showing a larger deviation. Similarly, in Figure 1(c), where the value function is exponential in x , the red line (MSBVE) still converges closer to the oracle estimator than the blue line (MSTDE), although the gap between MSBVE and the oracle increases compared to the linear and quadratic cases.

These findings indicate that when the value function is a low-degree polynomial in x , the MSBVE algorithm demonstrates greater robustness to stochastic processes with jumps compared to MSTDE. Specifically, MSBVE converges to a quantity associated primarily with the continuous component of the process, even in the presence of jumps. When the value function is a high-degree polynomial in x , MSBVE still outperforms MSTDE; however, the performance gap narrows, and both estimators exhibit increased deviation from the oracle estimator. Thus, it is important and interesting to develop a robust estimator that can handle the remaining bias from the gradient term of the value function, $\frac{\partial J^\theta(t, X_{t-})}{\partial x}$. We leave this for a future study.

5 Data Application

In this section, we apply the proposed MSBVE algorithm and compare it with the existing MSTDE algorithm in the context of a mean-variance portfolio selection problem. We consider a financial market with two assets over the time interval $[0, T]$: the risk-free asset follows

$$dB_t = r_f B_t dt,$$

where r_f is the risk-free rate, and the risky asset follows

$$dS_t = S_t(\mu dt + \sigma dW_t),$$

where μ is the drift, σ is the volatility, and W_t is a standard Brownian motion. Let X_t^a denote the wealth of an investor who allocates weight a_t to the risky asset at time t . Following a simplified form of the classical Merton model for continuous-time portfolio selection, the wealth process then follows an SDE:

$$dX_t^a = a_t(\sigma\rho dt + \sigma dW_t),$$

where $\rho = \frac{\mu-r_f}{\sigma}$ is the Sharpe ratio. See [Merton \(1971\)](#) for the derivation of portfolio dynamics in a diffusion market. The investor aims to minimize the variance of the wealth at time T while maintaining the expected value of the wealth at time T at a certain level z . Specifically,

$$\min_a \text{Var}(X_T^a) \quad s.t. \quad E(X_T^a) = z.$$

The theoretical optimal policy from [Jia & Zhou \(2022b\)](#) is given by

$$a_t(\rho) = -\frac{\rho}{\sigma}(X_t - w(\rho)),$$

where $w(\rho) = \frac{ze^{\rho^2 T} - X_0}{e^{\rho^2 T} - 1}$. Under this optimal policy, the wealth process becomes

$$dX_t = -\rho(X_t - w)(\rho dt + dW_t),$$

and the corresponding value function is defined as:

$$\begin{aligned} J(t, x) &= E[X_T^2 | X_t = x] - z^2 \\ &= (x - w)^2 e^{\rho^2(t-T)} - (w - z)^2. \end{aligned}$$

We parameterize the value function as

$$J^\theta(t, x) = (x - w)^2 e^{\theta^2(t-T)} - (w - z)^2,$$

and define the corresponding parametric policy as follows:

$$a_t(\theta) = -\frac{\theta}{\sigma}(x - w(\theta)),$$

where the estimated variance $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \frac{\pi}{2} \sum_{i=1}^{n-1} |\Delta X_{t_i}| \cdot |\Delta X_{t_{i-1}}|.$$

Our theoretical analysis yields two key predictions: (1) in jump-free environments, MSBVE and MSTDE should exhibit comparable performance; (2) in settings with jumps, MSBVE is expected to outperform MSTDE due to its robustness to jumps. To empirically test both settings, we perform two types of experiments: one on raw S&P 500 data (which includes jumps), and another on data where large jumps are removed.

For jump detection, we apply the threshold $\tau = 4\hat{\sigma}(\Delta t)^{0.47}$, and consider observations with $|\Delta X_t| > \tau$ to be large jumps ([Ait-Sahalia et al. 2020](#)). In practice, however, this threshold is not applied directly to the wealth process X_t , but rather to the observed S&P 500 price process S_t , since in real markets we only observe the stock price S_t and the risk free rate r_f . Using this method, we identify days with large jumps in the dataset of 5-minute S&P 500 data from 2010 to 2020. Notably, the period from 2012-10-01 to 2014-09-30 exhibits a higher frequency of jumps, with 195 jump days accounting for approximately 40% of the sample. Thus, to check the effect of jumps, we focus on this two-year dataset for model training and testing.

To simulate the jump-free scenario, we apply the thresholding procedure to remove large jumps. Specifically, we define the thresholded difference process as follows:

$$\Delta S_{t_i}^T = \begin{cases} \Delta S_{t_i}, & \text{if } |\Delta S_{t_i}| < \tau, \\ 0, & \text{if } |\Delta S_{t_i}| \geq \tau. \end{cases}$$

We can obtain the thresholded process by summarizing the thresholded difference process $S_{t_i}^T = \sum_{j=0}^{i-1} \Delta S_{t_j}^T$. We then apply MSTDE and MSBVE to the thresholded data and compare the results with those obtained from the raw data.

We use 6 months (126 trading days) of 5-minute S&P 500 data to train the allocation weights. The weights are updated after each trading day, and the algorithms are then tested on the following trading day. We consider a time horizon of $T = 1$ day with time steps $\Delta t = 1/79$. This process is repeated in a rolling-window fashion, advancing one day at a time. For each test day, we record the corresponding one-day terminal wealth, resulting in a sequence of terminal wealth values. Using this sequence, we compute the Sharpe ratio for both algorithms.

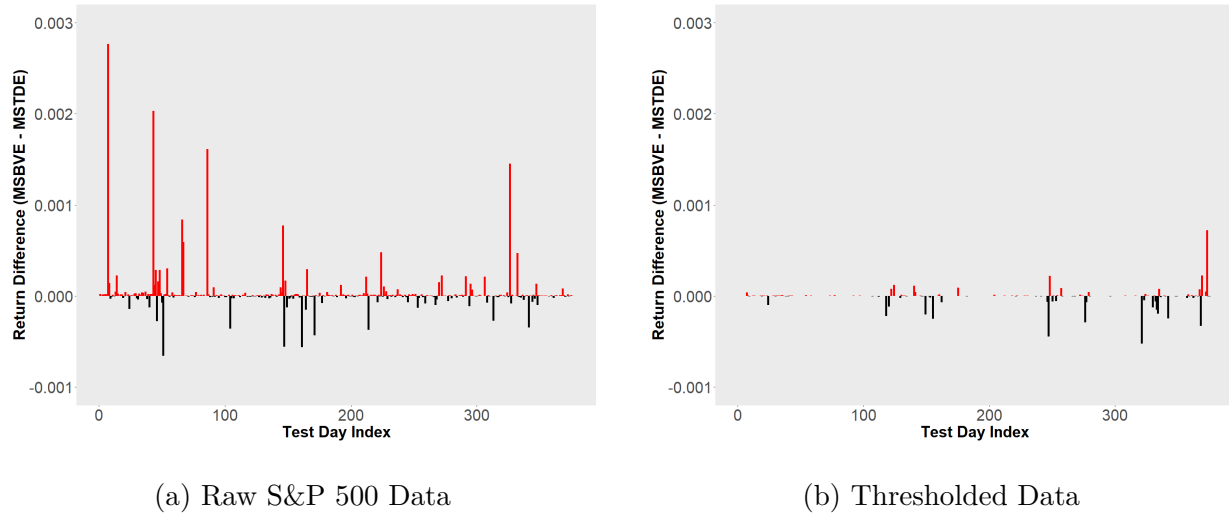


Figure 2: Return differences between MSBVE and MSTDE. Note that red indicates instances where MSBVE outperforms MSTDE, while black indicates the opposite.

Figure 2 illustrates the difference in daily returns between MSBVE and MSTDE across the testing period. Each vertical line represents the return difference on a single test day, where red indicates instances where MSBVE outperforms MSTDE, while black indicates the opposite. In Figure 2(a), which corresponds to the raw data setting, most lines are red with substantial magnitudes of outperformance. This indicates that MSBVE consistently yields higher returns than MSTDE in the presence of jumps. This observation supports our theoretical findings that MSBVE is more robust in environments with jumps. Figure 2(b) presents the performance differences after removing large jumps. In this case, red and black

lines appear more balanced, and the return differences are notably smaller, which suggests that the two algorithms perform comparably in the absence of large jumps.

To quantitatively evaluate the performance of the algorithms, we compute the Sharpe ratio under both settings and report the results in Table 1. We also include q-learning algorithm for comparison, following the implementation in Gao et al. (2024). In the raw data setting, MSBVE achieves a significantly higher Sharpe ratio than both MSTDE and q-learning, reaffirming its superior performance in environments with jumps. In contrast, under the jump-removed setting, the Sharpe ratios of the three algorithms are similar and lower than in the raw data case. These results further confirm that MSBVE performs best when applied directly to real financial data containing jumps, consistent with our theoretical findings.

Table 1: Annualized Sharpe ratio comparison of MSBVE and MSTDE

| Method | Raw S&P 500 Data | Thresholded Data |
|------------|------------------|------------------|
| MSTDE | 1.041 | 0.587 |
| MSBVE | 1.253 | 0.571 |
| q-learning | 1.055 | 0.599 |

6 Conclusion

In this paper, we explored the problem of estimating the value function in continuous-time reinforcement learning under stochastic dynamics with potential jump components. The continuous and jump parts exhibit fundamentally different characteristics, and in many applications, our primary interest lies in learning the dynamics of the continuous component. However, the latent continuous process is not directly observable due to the presence of

jumps. To address this challenge, we revisited the MSTDE algorithm and proposed a novel MSBVE algorithm, which updates a parameter vector by minimizing an alternative error metric. We have demonstrated that the MSBVE algorithm exhibits improved robustness and convergence behavior compared to the MSTDE algorithm. Importantly, when the underlying dynamics follow an SDE without jumps, the two algorithms yield equivalent results. However, when the dynamics are governed by an SDE with jump components, MSBVE places greater emphasis on the continuous part of the dynamics and is less affected by jumps, whereas MSTDE directly reflects both the continuous and jump components, often resulting in larger bias. These findings highlight that adopting alternative error metrics, such as the mean bipower variation error, can enhance the performance and reliability of reinforcement learning algorithms in continuous-time settings, particularly in the presence of jump-diffusion processes.

Moving forward, future research may explore extensions and refinements of the MSBVE algorithm, as well as investigate its applicability to real-world reinforcement learning problems characterized by complex dynamics and noisy environments. Additionally, further theoretical analysis and empirical studies could yield deeper insights into the convergence properties and computational efficiency of the proposed method, thereby contributing to the advancement of continuous-time reinforcement learning methodologies.

SUPPLEMENTARY MATERIAL

Appendix A – Proofs: Detailed proofs of all theoretical results stated in the main article.

(PDF file)

Appendix B – Code and Data: R code and datasets used to reproduce the results pre-

sented in the article. (Folder compressed as a ZIP file)

References

- Aït-Sahalia, Y., Kalnina, I. & Xiu, D. (2020), ‘High-frequency factor models and regressions’, *Journal of Econometrics* **216**(1), 86–105.
- Aït-Sahalia, Y. & Xiu, D. (2017), ‘Using principal component analysis to estimate a high dimensional factor model with high-frequency data’, *Journal of Econometrics* **201**(2), 384–399.
- Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), ‘Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility’, *The review of economics and statistics* **89**(4), 701–720.
- Barndorff-Nielsen, O. E. & Shephard, N. (2004), ‘A central limit theorem for realised power and bipower variations of continuous semimartingales’, *Journal of Financial Econometrics* **2**(1), 1–37.
- Bellman, R. (1957), *Dynamic Programming*, Princeton University Press, Princeton, NJ.
- Bender, C. & Thuan, N. T. (2023), ‘Entropy-regularized mean-variance portfolio optimization with jumps’, *arXiv preprint* .
- Bo, L., Huang, Y., Yu, X. & Zhang, T. (2024), ‘Continuous-time q-learning for jump-diffusion models under tsallis entropy’, *arXiv preprint* .
- Corsi, F., Pirino, D. & Reno, R. (2010), ‘Threshold bipower variation and the impact of jumps on volatility forecasting’, *Journal of Econometrics* **159**(2), 276–288.
- Denkert, R., Pham, H. & Warin, X. (2024), ‘Control randomisation approach for policy gradient and application to reinforcement learning in optimal switching’, *arXiv preprint* .

- Doya, K. (2000), ‘Reinforcement Learning in Continuous Time and Space’, *Neural Computation* **12**(1), 219–245.
- El Karoui, N., Mrad, M. & Mrad, M. (1997), ‘Backward stochastic differential equations in finance’, *Mathematical finance* **7**(1), 1–71.
- Gao, X., Li, L. & Zhou, X. Y. (2024), ‘Reinforcement learning for jump-diffusions, with financial applications’, *arXiv preprint* .
- Guo, X., Hu, A. & Zhang, Y. (2023), ‘Reinforcement learning for linear-convex models with jumps via stability analysis of feedback controls’, *SIAM Journal on Control and Optimization* **61**(2), 755–787.
- Jacod, J. & Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, 2nd edn, Springer.
- Jia, H. & Zhou, X. Y. (2022a), ‘Policy gradient and actor-critic learning in continuous time and space: Theory and algorithms’, *Mathematical Control and Related Fields* **12**(2), 305–338.
- Jia, H. & Zhou, X. Y. (2022b), ‘Q-learning in continuous time’, *IEEE Transactions on Automatic Control* **67**(6), 3102–3109.
- Jia, Y. & Zhou, X. Y. (2021), ‘Policy evaluation and temporal-difference learning in continuous time and space: A martingale approach’, *CoRR* **abs/2108.06655**.
- Karatzas, I. & Shreve, S. (2014), *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer New York.
- Merton, R. C. (1971), ‘Optimum consumption and portfolio rules in a continuous-time model’, *Journal of Economic Theory* **3**(4), 373–413.

- Oh, M., Kim, D. & Wang, Y. (2024), ‘Robust realized integrated beta estimator with application to dynamic analysis of integrated beta’, *Journal of Econometrics* p. 105810.
- Pham, H. (2009), *Continuous Time Stochastic Control and Optimization with Financial Applications*, Vol. 61 of *Stochastic Modelling and Applied Probability*, Springer.
- Shao, J. (2003), *Mathematical Statistics*, 2nd edn, Springer.
- Sutton, R. S. & Barto, A. G. (2018), *Reinforcement Learning: An Introduction*, 2nd edn, MIT press.

A Proofs

In this appendix, we prove Theorem 1 and Theorem 2. We first recall a lemma from Jia & Zhou (2021).

Lemma 3 (Lemma 8 in Jia & Zhou (2021)) *Let $f_h(x) = f(x) + r_h(x)$, where f is a continuous function and r_h converges to 0 uniformly on any compact set as $h \rightarrow 0$.*

(a) *Suppose $x_h^* \in \arg \min_x f_h(x) \neq \emptyset$ and $\lim_{h \rightarrow 0} x_h^* = x^*$. Then $x^* \in \arg \min_x f(x)$.*

(b) *Suppose $f_h(x_h^*) = 0$ and $\lim_{h \rightarrow 0} x_h^* = x^*$. Then $f(x^*) = 0$.*

Proof Refer to Appendix D in Jia & Zhou (2021). ■

We present the following inequalities that will be useful for the proof of Theorem 1 and Theorem 2.

Lemma 4 (Burkholder-Davis-Gundy inequality) *For any $1 \leq p < \infty$, there exists positive constants c_p, C_p such that, for all local martingales M with $M_0 = 0$ and stopping time τ , the following inequality holds:*

$$c_p E[\langle M \rangle_\tau^{p/2}] \leq E[|M_\tau^*|^p] \leq C_p E[\langle M \rangle_\tau^{p/2}],$$

where $\langle M \rangle_t$ denotes the quadratic variation of M , $M_t^* = \sup_{0 \leq s \leq t} |M_s|$ denotes the maximum process. In particular, if we consider $\tau = T$, then

$$E[|M_T|^p] \leq E[|M_T^*|^p] \leq C_p E[\langle M \rangle_T^{p/2}]. \tag{12}$$

Lemma 5 (Hölder's inequality) *Let $p, q \in [1, +\infty]$ with $1/p + 1/q = 1$.*

(a) *Let $u, v \in \mathbb{R}^n$, then we have:*

$$\sum_{i=1}^n |u_i v_i| \leq \left(\sum_{i=1}^n |u_i|^p \right)^{1/p} \left(\sum_{i=1}^n |v_i|^q \right)^{1/q}. \tag{13}$$

(b) Let Y, Z be random variables, then we have:

$$E|YZ| \leq (E|Y|^p)^{1/p} (E|Z|^q)^{1/q}. \quad (14)$$

(c) Let f, g be real-valued integrable functions, then we have:

$$\int |f(x)g(x)|dx \leq \left(\int |f(x)|^p dx \right)^{1/p} \left(\int |g(x)|^q dx \right)^{1/q}. \quad (15)$$

Lemma 6 Let $a, b, c, d \in \mathbb{R}$, if $bd = 0$, then we have:

$$\left| |a+b||c+d| - |a||c| \right| \leq |a||d| + |b||c|. \quad (16)$$

Proof If $b = 0$,

$$\begin{aligned} & |c| - |d| \leq |c+d| \leq |c| + |d| \\ \Rightarrow & |a||c| - |a||d| \leq |a||c+d| \leq |a||c| + |a||d| \\ \Rightarrow & -|a||d| \leq |a||c+d| - |a||c| \leq |a||d| \\ \Rightarrow & \left| |a||c+d| - |a||c| \right| \leq |a||d| \\ \Rightarrow & \left| |a+b||c+d| - |a||c| \right| \leq |a||d| + |b||c|. \end{aligned}$$

If $d = 0$,

$$\begin{aligned} & |a| - |b| \leq |a+b| \leq |a| + |b| \\ \Rightarrow & |a||c| - |b||c| \leq |c||a+b| \leq |a||c| + |b||c| \\ \Rightarrow & -|b||c| \leq |c||a+b| - |a||c| \leq |b||c| \\ \Rightarrow & \left| |c||a+b| - |a||c| \right| \leq |b||c| \\ \Rightarrow & \left| |a+b||c+d| - |a||c| \right| \leq |a||d| + |b||c|. \end{aligned}$$

■

We present the following standard results for asymptotic convergence and Itô process.

Lemma 7 (Theorem 1.8 in [Shao \(2003\)](#)) *Let X_n be a sequence of random variables, suppose that $X_n \xrightarrow{d} X$. Then for any $r > 0$,*

$$\lim_{n \rightarrow \infty} E\|X_n\|_r^r = E\|X\|_r^r < \infty$$

if and only if $\{\|X_n\|_r^r\}$ is uniformly integrable in the sense that

$$\lim_{t \rightarrow \infty} \sup_n E(\|X_n\|_r^r I_{\{\|X_n\|_r > t\}}) = 0.$$

A sufficient condition for uniform integrability of $\{\|X_n\|_r^r\}$ is that

$$\sup_n E\|X_n\|_r^{r+\delta} < \infty$$

for a $\delta > 0$.

Lemma 8 *Consider a semimartingale of the form $Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s$ and the partition $0 = t_0 < t_1 < \dots < t_n = T$ with $t_i = T/n$, we have*

(a) (Theorem 4.47 in [Jacod & Shiryaev \(2003\)](#))

$$\sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2 \xrightarrow{p} \int_0^T \sigma_s^2 ds.$$

(b) (Theorem 2.2 in [Barndorff-Nielsen & Shephard \(2004\)](#))

$$\sum_{i=1}^{n-1} |Y_{t_{i+1}} - Y_{t_i}| |Y_{t_i} - Y_{t_{i-1}}| \xrightarrow{p} \frac{2}{\pi} \int_0^T \sigma_s^2 ds.$$

Now, we start to prove Theorem 1.

Proof Let us start by highlighting an useful fact. First fix a sufficiently small time grid size Δt , by Assumption 1(v), when Δt is small enough, there is at most one jump within each interval $[t_i, t_{i+1}]$, hence

$$\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t$$

$$= \begin{cases} 0, & \text{if there is no jump in } [t_i, t_{i+1}] \\ J^\theta(s_i, X_{s_i}) - J^\theta(s_i^-, X_{s_i^-}), & \text{if there is a jump happens at } s_i \in [t_i, t_{i+1}]. \end{cases}$$

This implies for any $K > 0$,

$$\left\{ \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right\}^K = \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^K dN_t. \quad (17)$$

Now let's prove the theorem. By Itô's formula on $J^\theta(t_{i+1}, X_{t_{i+1}})$, we have

$$\begin{aligned} & \sum_{i=0}^{n-1} [J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i})]^2 \\ &= \sum_{i=0}^{n-1} \left\{ \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right. \\ & \quad \left. + \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right\}^2 \\ &= \sum_{i=0}^{n-1} \left\{ \left(\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right)^2 + \left[\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right]^2 \right. \\ & \quad \left. + \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right. \\ & \quad \left. + 2 \left(\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right) \left(\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right) \right. \\ & \quad \left. + 2 \left\{ \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right\} \right. \\ & \quad \left. \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\}. \end{aligned}$$

By Itô's isometry,

$$\begin{aligned} E \left\{ \sum_{i=0}^{n-1} \left[\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right]^2 \right\} &= E \left\{ \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt \right\} \\ &= E \left\{ \int_0^T \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt \right\}. \end{aligned}$$

Thus,

$$\text{MSTDE}_{\Delta t}^*(\theta)$$

$$\begin{aligned}
&= E \left\{ \int_0^T \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt + \int_0^T [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right\} \\
&+ \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right]^2 \right. \\
&+ 2 \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right] \left[\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \\
&+ 2 \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \\
&\left. \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\}.
\end{aligned}$$

We write $\text{MSTDE}_{\Delta t}^*(\theta) = QV(\theta) + R_1(\theta) + R_2(\theta)$, where

$$QV(\theta) := E \left\{ \int_0^T \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt + \int_0^T [J^\theta(t, X_t) - J^\theta(t, X_{t-})]^2 dN_t \right\},$$

$$\begin{aligned}
R_1(\theta) &:= \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right]^2 \right. \\
&\left. + 2 \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right] \left[\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\},
\end{aligned}$$

and

$$\begin{aligned}
R_2(\theta) &:= \sum_{i=0}^{n-1} E \left\{ 2 \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \right. \\
&\left. \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\}.
\end{aligned}$$

By Hölder's inequality, we obtain:

$$\begin{aligned}
|R_1(\theta)| &\leq \sum_{i=0}^{n-1} E \left\{ \int_{t_i}^{t_{i+1}} [\mathcal{L}J^\theta(t, X_{t-})]^2 dt \Delta t \right\} \\
&+ 2 \sum_{i=0}^{n-1} \left\{ E \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right]^2 E \left[\int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right]^2 \right\}^{1/2} \\
&\leq \sum_{i=0}^{n-1} E \left\{ \int_{t_i}^{t_{i+1}} [\mathcal{L}J^\theta(t, X_{t-})]^2 dt \Delta t \right\}
\end{aligned}$$

$$\begin{aligned}
& + 2 \sum_{i=0}^{n-1} \left\{ E \left[\int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^2 dt \Delta t \right] E \left[\int_{t_i}^{t_{i+1}} \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2 dt \right] \right\}^{1/2} \\
& \leq \sum_{i=0}^{n-1} E \left\{ \int_{t_i}^{t_{i+1}} [\mathcal{L}J^\theta(t, X_{t-})]^2 dt \Delta t \right\} \\
& + 2 \left\{ \sum_{i=0}^{n-1} E \left[\int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^2 dt \Delta t \right] \right. \\
& \quad \left. \cdot \sum_{i=0}^{n-1} E \left[\int_{t_i}^{t_{i+1}} \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2 dt \right] \right\}^{1/2} \\
& = \Delta t \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2}^2 + 2\sqrt{\Delta t} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2},
\end{aligned}$$

where the first line is due to (15) with $p = 2, q = 2, f = \mathcal{L}J^\theta(t, X_{t-})$ and $g = 1$, the third line is due to (15) with $p = 2, q = 2, f = \mathcal{L}J^\theta(t, X_{t-}), g = 1$ and Itô's isometry, and the fifth line is due to (13) with $p = 2, q = 2, u_i = \left\{ E \left[\int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^2 dt \Delta t \right] \right\}^{1/2}$ and $v_i = \left\{ E \left[\int_{t_i}^{t_{i+1}} \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2 dt \right] \right\}^{1/2}$.

Assumption 3 states that the L^2 -norm of $\mathcal{L}J^\theta$ and $|\frac{\partial J}{\partial x} \cdot \sigma|^2$ are continuous functions of θ , thus, for any compact set Γ , their supremes are all finite. Hence, for all $\theta \in \Gamma$,

$$\begin{aligned}
|R_1(\theta)| & \leq \Delta t \left[\sup_{\theta \in \Gamma} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \right]^2 \\
& + 2\sqrt{\Delta t} \sup_{\theta \in \Gamma} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \sup_{\theta \in \Gamma} \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2} \rightarrow 0 \text{ as } \Delta t \rightarrow 0,
\end{aligned}$$

Recall that

$$\begin{aligned}
R_2(\theta) & := \sum_{i=0}^{n-1} E \left\{ 2 \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \right. \\
& \quad \left. \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\},
\end{aligned}$$

we only need to show $R_2(\theta) \rightarrow 0$. The desired results follows from the following lemma.

Lemma 9

$$\sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \right\}$$

$$\begin{aligned}
& \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right] \Big\} \\
& \leq 2 \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \cdot T \right\}^{1/2} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \\
& + 2(C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq t \leq T} E|J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}} \left[\left\| \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2} \right\|^2 \right]^{\frac{1}{2}} \\
& \cdot \left[\frac{T}{\Delta t} \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \right]^{\frac{3}{4}}. \tag{18}
\end{aligned}$$

Proof By applying Hölder's inequality in the first summand of the LHS of (18), we obtain:

$$\begin{aligned}
& \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt \right\} \\
& \leq \sum_{i=0}^{n-1} \left\{ E \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right]^2 E \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt \right]^2 \right\}^{1/2} \\
& = \sum_{i=0}^{n-1} \left\{ P(N_{t_{i+1}} - N_{t_i} = 1) E[J^\theta(s_i, X_{s_i}) - J^\theta(s_i-, X_{s_i-})]^2 E \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt \right]^2 \right\}^{1/2} \\
& \leq \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot 4 \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \right\}^{1/2} \sum_{i=0}^{n-1} \left\{ E \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt \right]^2 \right\}^{1/2} \\
& \leq \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot 4 \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \right\}^{1/2} \\
& \quad \cdot \sum_{i=0}^{n-1} \left\{ E \left[\int_{t_i}^{t_{i+1}} [\mathcal{L}J^\theta(t, X_{t-})]^2 dt \Delta t \right] \right\}^{1/2} \\
& \leq \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot 4 \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \right\}^{1/2} \\
& \quad \cdot \left\{ \sum_{i=0}^{n-1} E \left[\int_{t_i}^{t_{i+1}} [\mathcal{L}J^\theta(t, X_{t-})]^2 dt \right] \sum_{i=0}^{n-1} \Delta t \right\}^{1/2} \\
& = 2 \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \cdot T \right\}^{1/2} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2},
\end{aligned}$$

where the second line is due to (14) with $p = 2, q = 2, Y_i = \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t$ and $Z_i = \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt$, the third line is due to (17), the fifth line is due to (15)

with $p = 2, q = 2, f = \mathcal{L}J^\theta(t, X_{t-})$ and $g = 1$, and the seventh line is due to (13) with $p = 2, q = 2, u_i = \left\{ E \left[\int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^2 dt \right] \right\}^{1/2}$ and $v_i = (\Delta t)^{1/2}$. Then for the second summand of the LHS of (18),

$$\begin{aligned}
& \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right\} \\
& \leq \sum_{i=0}^{n-1} \left[E \left| \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right|^{\frac{4}{3}} \right]^{\frac{3}{4}} \left[E \left| \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^4 \right]^{\frac{1}{4}} \\
& = \sum_{i=0}^{n-1} \left\{ \left[P(N_{t_{i+1}} - N_{t_i} = 1) E |J^\theta(s_i, X_{s_i}) - J^\theta(s_i-, X_{s_i-})|^{\frac{4}{3}} \right]^{\frac{3}{4}} \right. \\
& \quad \cdot \left. \left[E \left| \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^4 \right]^{\frac{1}{4}} \right\} \\
& \leq \sum_{i=0}^{n-1} \left\{ \left[P(N_{t_{i+1}} - N_{t_i} = 1) E |J^\theta(s_i, X_{s_i}) - J^\theta(s_i-, X_{s_i-})|^{\frac{4}{3}} \right]^{\frac{3}{4}} \right. \\
& \quad \cdot \left. \left[CE \left(\int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^2 dt \right)^2 \right]^{1/4} \right\} \\
& \leq C^{\frac{1}{4}} \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E |2J^\theta(t, X_t)|^{\frac{4}{3}} \right\}^{\frac{3}{4}} \\
& \quad \cdot \sum_{i=0}^{n-1} \left[E \left(\int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^2 dt \right)^2 \right]^{1/4} \\
& \leq C^{\frac{1}{4}} \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E |2J^\theta(t, X_t)|^{\frac{4}{3}} \right\}^{\frac{3}{4}} \\
& \quad \cdot \sum_{i=0}^{n-1} \left[E \left(\int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^4 dt \right) \Delta t \right]^{1/4} \\
& \leq (C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E |2J^\theta(t, X_t)|^{\frac{4}{3}} \right\}^{\frac{3}{4}} \\
& \quad \cdot \left[\left(\sum_{i=0}^{n-1} E \int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^4 dt \right)^{1/4} \left(\sum_{i=0}^{n-1} 1 \right)^{\frac{3}{4}} \right] \\
& \leq 2(C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq t \leq T} E |J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}} \left[\left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2}^2 \right]^{\frac{1}{2}}
\end{aligned}$$

$$\cdot \left[\frac{T}{\Delta t} \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \right]^{\frac{3}{4}},$$

where the second line is due to (14) with $p = 4/3, q = 4, Y_i = \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t$ and $Z_i = \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$, the third line is due to (17) with $K = 4/3$, the fifth line is due to (12) with $p = 4, (M_i)_t = \int_{t_i}^{t_i+t} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$ and C is a positive constant, the ninth line is due to (15) with $p = 2, q = 2, f = \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^2$ and $g = 1$, the eleventh line is due to (13) with $p = 4, q = 4/3, u_i = \left(E \int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^4 \right)^{1/4}$ and $v_i = 1$, and the thirteenth line is due to (14) with $p = 3/2, q = 3, Y = |J^\theta(t, X_t)|^{\frac{4}{3}}, Z = 1$ and the fact that $T = n\Delta t$.

Hence, we have:

$$\begin{aligned} & \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \right. \\ & \quad \cdot \left. \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right] \right\} \\ & \leq 2 \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E [J^\theta(t, X_t)]^2 \cdot T \right\}^{1/2} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \\ & \quad + 2(C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq t \leq T} E |J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}} \left[\left\| \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\| \right\|_{L^2} \right]^{\frac{1}{2}} \\ & \quad \cdot \left[\frac{T}{\Delta t} \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \right]^{\frac{3}{4}}. \end{aligned}$$

This completes the proof of Lemma 9. ■

Assumption 3 implies that for any compact set Γ , $\sup_{0 \leq t \leq T} \left\{ E |J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}}$ is bounded for all $\theta \in \Gamma$. Assumption 1(v) states that $\sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) = \mathcal{O}(\Delta t)$. Hence, for all $\theta \in \Gamma$, by Lemma 9, we have the following:

$$|R_2(\theta)| = 2 \sum_{i=0}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right] \right.$$

$$\begin{aligned}
& \cdot \left[\int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right] \Big\} \\
& \leq 4 \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E[J^\theta(t, X_t)]^2 \cdot T \right\}^{1/2} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \\
& + 4(C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq t \leq T} E|J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}} \left[\left\| \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\| \right\|_{L^2}^2 \right]^{\frac{1}{2}} \\
& \cdot \left[\frac{T}{\Delta t} \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \right]^{\frac{3}{4}} \rightarrow 0 \quad \text{as } \Delta t \rightarrow 0.
\end{aligned}$$

The desired result follows from Lemma 3. ■

Now, we present the proof of Theorem 2.

Proof Let us start by highlighting an useful fact. First fix a sufficiently small time grid size Δt , by Assumption 1(v), when Δt is small enough, there is at most one jump within each set $[t_{i-1}, t_i] \cup [t_i, t_{i+1}]$, hence at least one of $\int_{t_i}^{t_{i+1}} |J^\theta(t, X_t) - J^\theta(t, X_{t-})| dN_t$ and $\int_{t_{i-1}}^{t_i} |J^\theta(t, X_t) - J^\theta(t, X_{t-})| dN_t$ is zero. Thus,

$$\left[\int_{t_i}^{t_{i+1}} |J^\theta(t, X_t) - J^\theta(t, X_{t-})| dN_t \right] \left[\int_{t_{i-1}}^{t_i} |J^\theta(t, X_t) - J^\theta(t, X_{t-})| dN_t \right] = 0. \quad (19)$$

Now let's prove the theorem. By Itô's formula on $J^\theta(t_{i+1}, X_{t_{i+1}})$, we have

$$\begin{aligned}
& \sum_{i=1}^{n-1} \left| J^\theta(t_{i+1}, X_{t_{i+1}}) - J^\theta(t_i, X_{t_i}) \right| \left| J^\theta(t_i, X_{t_i}) - J^\theta(t_{i-1}, X_{t_{i-1}}) \right| \\
& = \sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right. \\
& + \left. \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right| \\
& \cdot \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t + \int_{t_{i-1}}^{t_i} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right|.
\end{aligned}$$

We write $\text{MSBVE}_{\Delta t}(\theta) = BV(\theta) + \tilde{R}_1(\theta) + \tilde{R}_2(\theta)$, where

$$BV(\theta) := \frac{2}{\pi} E \left\{ \int_0^T \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt \right\},$$

$$\begin{aligned} \tilde{R}_1(\theta) := & E \left[\sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right. \\ & \cdot \left. \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right] - BV(\theta), \end{aligned}$$

and

$$\begin{aligned} \tilde{R}_2(\theta) := & E \left[\sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right. \right. \\ & + \left. \left. \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right| \right. \\ & \cdot \left. \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right. \right. \\ & + \left. \left. \int_{t_{i-1}}^{t_i} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t \right| \right] \\ & - E \left[\sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right. \\ & \cdot \left. \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right] \end{aligned}$$

By Lemma 8(b), we have:

$$\begin{aligned} Y_n := & \sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \\ & \cdot \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \\ & - \frac{2}{\pi} \int_0^T \left[\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right]^2 dt \xrightarrow{p} 0. \end{aligned}$$

This implies $Y_n \xrightarrow{d} 0$. Notice that

$$\begin{aligned} & E|Y_n|^2 \\ & \leq 2E \left| \sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right. \\ & \cdot \left. \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-})dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-})dW_t \right| \right|^2 \end{aligned}$$

$$\begin{aligned}
& + 2E \left[\frac{2}{\pi} \int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2 dt \right]^2 \\
& \leq 2E \left[\sum_{i=1}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^2 \right. \\
& \quad \cdot \left. \sum_{i=1}^{n-1} \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^2 \right] \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
& \leq 2E \left[\sum_{i=0}^{n-1} \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^2 \right]^2 \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
& \leq 2n \sum_{i=0}^{n-1} E \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^4 \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
& \leq 16n \sum_{i=0}^{n-1} \left(E \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt \right|^4 + E \left| \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^4 \right) \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
& \leq 16n \sum_{i=0}^{n-1} \left[E \left(\int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^4 dt \cdot (\Delta t)^3 \right) \right. \\
& \quad \left. + CE \left(\int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^2 dt \right)^2 \right] \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
& \leq 16n \sum_{i=0}^{n-1} \left[(\Delta t)^3 E \int_{t_i}^{t_{i+1}} (\mathcal{L}J^\theta(t, X_{t-}))^4 dt + C(\Delta t) E \int_{t_i}^{t_{i+1}} \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^4 dt \right] \\
& + \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right]
\end{aligned}$$

$$\begin{aligned}
&= 16n(\Delta t) \left[(\Delta t)^2 E \int_0^T (\mathcal{L}J^\theta(t, X_{t-}))^4 dt + CE \int_0^T \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^4 dt \right] \\
&+ \frac{8}{\pi^2} E \left[\int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^4 dt \cdot T \right] \\
&= T \left[16(\Delta t)^2 \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2}^2 + (16C + \frac{8}{\pi^2}) \left\| \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2} \right\|^2 \right] < \infty,
\end{aligned}$$

where the second line is due to (13) with

$$\begin{aligned}
u_1 &= \left| \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right. \\
&\quad \left. \cdot \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right| \right|, \\
u_2 &= \left[\frac{2}{\pi} \int_0^T \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2 dt \right], v_1 = v_2 = 1, p = q = 2,
\end{aligned}$$

the fifth line is due to (13) with $u_i = \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$, $v_i = \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$, $p = q = 2$ and (15) with $f = \left(\frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right)^2$, $g = 1, p = q = 2$, the eighth line is due to adding extra positive terms, the tenth line is due to (13) with $u_i = \left| \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right|^2$, $v_i = 1, p = q = 2$, the twelfth line is due to $(u_1 + u_2)^4 \leq 8(u_1^4 + u_2^4)$ with $u_1 = \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt$ and $u_2 = \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$, the fourteenth line is due to (15) with $f = \mathcal{L}J^\theta(t, X_{t-}), g = 1, p = 4, q = 4/3$ and (12) with $p = 4, (M_i)_t = \int_{t_i}^{t_i+t} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t$ and C is a positive constant, the seventeenth line is due to (15) with $f = \left| \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) \right|^2$, $g = 1, p = q = 2$, and the twenty-first line is due to the fact that $T = n(\Delta t)$.

By Lemma 7, $\sup_n E|Y_n|^2 < \infty$ implies that $\{|Y_n|\}$ is uniformly integrable. Together with $Y_n \xrightarrow{d} 0$, this results in $E|Y_n| \rightarrow 0$, i.e., $|\tilde{R}_1(\theta)| \rightarrow 0$.

Now let's prove $|\tilde{R}_2(\theta)| \rightarrow 0$. Let

$$a_i = \int_{t_i}^{t_{i+1}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_i}^{t_{i+1}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t,$$

$$\begin{aligned}
b_i &= \int_{t_i}^{t_{i+1}} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t, \\
c_i &= \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t, \\
d_i &= \int_{t_{i-1}}^{t_i} [J^\theta(t, X_t) - J^\theta(t, X_{t-})] dN_t.
\end{aligned}$$

Recall that

$$\tilde{R}_2(\theta) = E \left[\sum_{i=1}^{n-1} |a_i + b_i| |c_i + d_i| \right] - E \left[\sum_{i=1}^{n-1} |a_i| |c_i| \right].$$

By (19) and the condition $b_i d_i = 0$, so we can use Lemma 6 with a_i, b_i, c_i and d_i , we have the following:

$$\begin{aligned}
|\tilde{R}_2(\theta)| &\leq \sum_{i=1}^{n-1} E \left\{ \left[\int_{t_i}^{t_{i+1}} |J^\theta(t, X_t) - J^\theta(t, X_{t-})| dN_t \right] \right. \\
&\quad \cdot \left[\left| \int_{t_{i+1}}^{t_{i+2}} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i+1}}^{t_{i+2}} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right| \right. \\
&\quad \left. \left. + \left| \int_{t_{i-1}}^{t_i} \mathcal{L}J^\theta(t, X_{t-}) dt + \int_{t_{i-1}}^{t_i} \frac{\partial J^\theta}{\partial x}(t, X_{t-})^\dagger \sigma(t, X_{t-}) dW_t \right| \right] \right\},
\end{aligned}$$

where $t_{-1} = t_0 = 0$, and $t_{n+1} = t_n = T$ for simplicity.

Assumption 3 implies that for any compact set Γ , $\sup_{0 \leq t \leq T} \left\{ E |J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}}$ is bounded for all $\theta \in \Gamma$. Assumption 1(v) states that $\sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) = \mathcal{O}(\Delta t)$. Hence, for all $\theta \in \Gamma$, by Lemma 9, we have the following:

$$\begin{aligned}
|\tilde{R}_2(\theta)| &\leq 8 \left\{ \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \cdot \sup_{0 \leq t \leq T} E [J^\theta(t, X_t)]^2 \cdot T \right\}^{1/2} \|\mathcal{L}J^\theta(\cdot, X)\|_{L^2} \\
&\quad + 8(C\Delta t)^{\frac{1}{4}} \left\{ \sup_{0 \leq t \leq T} E |J^\theta(t, X_t)|^2 \right\}^{\frac{1}{2}} \left[\left\| \left\| \frac{\partial J^\theta}{\partial x}(\cdot, X)^\dagger \sigma(\cdot, X) \right\|_{L^2} \right\|^2 \right]^{\frac{1}{2}} \\
&\quad \cdot \left[\frac{T}{\Delta t} \sup_{0 \leq i \leq n-1} P(N_{t_{i+1}} - N_{t_i} = 1) \right]^{\frac{3}{4}} \rightarrow 0 \quad \text{as } \Delta t \rightarrow 0.
\end{aligned}$$

The desired result follows from Lemma 3. ■