Asymptotic Normality of Reweighted Nadaraya-Watson Threshold Estimation for Jump Intensity Function

Rui Feng*

Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University
e-mail: rfeng.20@saif.sjtu.edu.cn

Abstract: In this paper, based on the threshold technique, we construct the reweighted Nadaraya-Watson estimator for the unknown jump intensity function of continuous-time diffusion model with finite active jumps. The asymptotic normality of the proposed estimator is obtained under some regular conditions without assuming stationarity. Moreover, Monte Carlo simulations show that the underlying estimator performs better in finite samples.

Keywords and phrases: Jump-diffusion model, Nonparametric threshold estimation, Jacod's stable convergence theorem, Bias reduction.

1. Introduction

Jumps have been widely considered in asset pricing process. Diffusion model with jumps is one of the most commonly used models in describing variables in finance and economics, such as interest rates, asset returns, exchange rates, etc. One can refer to Das [6], Ait-Sahalia and Jacod [1] and references therein. The model we analyze is described by the following stochastic differential equation:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t, \qquad t \in [0, T] \tag{1}$$

where $\{W_t, t \geq 0\}$ is a standard Brownian motion, $\{J_t, t \geq 0\}$ is a finite activity (FA) pure jump semimartingale, which we can write as $J_t = \sum_{j=1}^{N_t} \gamma_j$, where N_t is a Poisson process. Furthmore, the intensity of N_t is a stochastic process depending on X_t with the form $\lambda(X_t)$, not just a constant, more details in Cont and Tankov [5]. The total amount of noise in this model comes not only from the diffusion part, but also the jump part. The description of jumping behaviors can help explain the sharp fluctuations of the underlying assets, especially λ, which characters the frequency of jumps. As a consequence, in this paper, we mainly focus on the nonparametric estimation of the intensity parameter.

In previous work, Bandi and Nguyen [3] used higher infinitesimal moments, namely fourth and sixth moments, to construct nonparametric estimator for the jump intensity function and proved the consistency of the estimator. But they didn't obtain the central limit theorem. Mancini [15] applied the threshold

*Corresponding author

technic to identify the instants of jumps and constructed an estimator of constant intensity parameter. Furthermore, Mancini and Renò [17] employed the kernel smoothing technique to reconstruct the Nadaraya-Watson estimator for the generalized jump intensity function (not necessarily to be a constant) and provided the asymptotic normality of the underlying estimator.

To improve the finite sample performance of the estimator considered in Mancini and Renò [17], a traditional idea is to consider the local linear estimator. However, it may produce a negative result while the intensity function is assumed to be nonnegative. A practical evidence for the negativity problem of local linear estimator can be seen in Xu [21]. The reweighted Nadaraya-Watson estimation is a useful approach to solve this problem. It is guaranteed to be nonnegative in finite samples and possesses the bias properties of the local linear estimator as well. Based on the this, Xu [21] proposed a new estimator of the diffusion function; Hanif et al.[12] estimated the second infinitesimal moment of the jump diffusion model. However, there are very few literatures about the application of reweighted Nadaraya-Watson method to the estimation of intensity parameter independently. In this paper, we study the reweighted Nadaraya-Watson threshold estimation of the jump intensity function independently and obtain asymptotic normality of this estimator.

Practically, through a Monte Carlo simulation, we compare our reweighted Nadaraya-Watson estimator with the traditional Nadaraya-Watson estimator considered in Mancini and Renò [17]. The better finite-sample performance of our estimator, especially bias correction effect at boundary points, is verified. Moreover, QQ plot is employed to reveal the normality of the underlying estimator, thus confirming the main result of this paper. We finally conclude the reweighted Nadaraya-Watson estimator constructed in this paper can effectively improve the statistical behavior of Nadaraya-Watson estimator such as bias reduction and boundary correction.

An outline of the paper is as follows: Section 2 introduces the jump-diffusion model and reweighted Nadaraya-Watson estimator. In this section, we construct our estimator for jump intensity coefficient of a jump-diffusion model. Section 3 presents ordinary assumptions and main results of our paper. Asymptotic normality of our estimators is shown there. The finite-sample properties of the underlying nonparametric threshold estimators are demonstrated in Section 4. Section 5 concludes. The proof of the lemmas and the main results will be collected in Section 6.

2. Jump-diffusion model and Reweighted Nadaraya-Watson Estimator

We assume that all processes are defined on a filtered probability space (Ω, $(\mathscr{F}_t)_{t\in[0,T]}$, $\mathscr{F}$, P) satisfying the usual conditions as those in Protter[18]. The diffusion process with jumps considered in this paper is governed by the stochastic differential equation as follows:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t, \qquad t \in [0,T]$$

where $\mu(x)$ and $\sigma(x)$ are smooth functions, $\{W_t, t \geq 0\}$ is a standard Brownian motion, $\{J_t, t \geq 0\}$ is a finite activity (FA) pure jump semimartingale independent of $\{W_t, t \geq 0\}$ and is also assumed to be

$$J_t = \int_0^t \int_{\mathscr{R}} x \cdot m(ds, dx) = \sum_{i=1}^{N_t} \gamma_i,$$

where m is the jump random measure of J_t, and $N_t := \int_0^t \int_{\mathscr{R}} 1 \cdot m(ds, dx)$ is a a.s. finite Poisson process with a jump intensity stochastic process $\lambda(\cdot)$ in $L^1(\Omega \times [0, +\infty))$ depending only on X_t. $(\tau_i)_{i=1,...,N_T}$, denotes the jumping time and each γ_i, also denoted γ_{τ_i}, is the size of the jump occurred at τ_i.

Recall the Nadaraya-Watson estimator and local linear estimator:

$$\hat{\lambda}_{NW}(x, h) = \frac{\sum_{i=1}^n K_h(X_{t_{i-1}} - x) c_{i,n} I_{\{|\Delta_i X|^2 > \vartheta(\delta)\}}}{\sum_{i=1}^n K_h(X_{t_{i-1}} - x)\delta}$$

$$\hat{\lambda}_{LL}(x, h) = \frac{\sum_{i=1}^n w_{i-1}^{LL} K_h(X_{t_{i-1}} - x) c_{i,n} I_{\{|\Delta_i X|^2 > \vartheta(\delta)\}}}{\sum_{i=1}^n w_{i-1}^{LL} K_h(X_{t_{i-1}} - x)\delta}$$

where $t_i = i\delta$ with $\delta = \frac{T}{n}$ denoting the sampling frequency, $\Delta_i X = X_{t_i} - X_{t_{i-1}}$, $c_{i,n}$ is a double array of constants with $i = 1, ..., n$ and $\vartheta(\delta)$ is the threshold function. $K_h(\cdot) = K(\cdot/h)/h$. $w_i^{LL} = S_{n,2} - (X_{t_i} - x) S_{n,1}$ with $S_{n,j} = \sum_{i=1}^n (X_{t_i} - x)^j K_h(X_{t_i} - x)$, $j = 1, 2$. The local linear estimator performs better than the Nadaraya-Watson estimator in finite-samples, however the weights $w_i^{LL}(x, h)$ are not guaranteed to be nonnegative and thus the resulting estimates may be negative. The reweighted Nadaraya-Watson estimator can be used to solve the negativity problem.

The reweighted idea was first introduced by Hall and Presnell [11] as an aid to increasing performance in a range of statistical problems by altering the effective sampling distribution to reduce a source of bias or variance, without changing other aspects of the problem. Cai [4] applied this method to study nonparametric estimation of regression function for α-mixing time series at both boundary and interior points. Hall and Huang [10] employed a similar reweighting idea for monotonising general linear, kernel-type estimators. Xu [21] proposed a new functional estimator of diffusion coefficient based on the reweighted Nadaraya-Watson estimator and developed a limit theory under mild requirements.

In this problem, the reweighted Nadaraya-Watson estimator for the diffusion function is defined by

$$\hat{\lambda}_{RNW}(x, h) = \frac{\sum_{i=1}^n w_{i-1}^{RNW} K_h(X_{t_{i-1}} - x) c_{i,n} I_{\{|\Delta_i X|^2 > \vartheta(\delta)\}}}{\sum_{i=1}^n w_{i-1}^{RNW} K_h(X_{t_{i-1}} - x)\delta} \tag{2}$$

where the weights $w_i^{RNW}(x, h), i = 1, \cdots, n$, solve the following constrained optimization problem

$$w_i^{RNW}(x, h) = \underset{\{w_i\}}{argmax} \sum_{i=1}^n \log n w_i$$

subject to

$$w_i \geq 0 \qquad \sum_{i=1}^{n} w_i = 1 \tag{3}$$

and

$$\sum_{i=1}^{n} w_i (X_{t_i} - x) K_h (X_{t_i} - x) = 0 \tag{4}$$

The weights $w_i^{RNW}(x,h)$ in (2) are easy to compute. Let the order observations be $X_{(\Delta)} \leq X_{(2\Delta)} \cdots \leq X_{(n\Delta)}$ and assume that the spatial point x satisfies $X_{(\Delta)} < x < X_{(n\Delta)}$. The Lagrangian function for the constrained optimization problem is

$$\mathcal{L}_n(w_i, \alpha, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log n w_i - \alpha (\sum_{i=1}^{n} w_i - 1) - \theta \sum_{i=1}^{n} w_i (X_{t_i} - x) K_h (X_{t_i} - x),$$

where α, θ are Lagrange multipliers. The first order condition of $\mathcal{L}_n(w_i, \alpha, \theta)$ with respect to w_i, α and θ gives

$$\frac{1}{n} - \alpha w_i - \theta w_i (X_{t_i} - x) K_h (X_{t_i} - x) = 0,$$

$$\sum_{i=1}^{n} w_i = 1,$$

$$\sum_{i=1}^{n} w_i (X_{t_i} - x) K_h (X_{t_i} - x) = 0$$

from which $\alpha = 1$ and

$$w_i^{RNW}(x,h) = \frac{1}{n(1 + \theta (X_{t_i} - X) K_h (X_{t_i} - x))}$$

where θ satisfies

$$\sum_{i=1}^{n} \frac{(X_{t_i} - x) K_h (X_{t_i} - x)}{n(1 + \theta (X_{t_i} - x) K_h (X_{t_i} - x))} = 0$$

Remark 2.1. The intuition of using threshold technique lies in that when the time interval between two observations tends to zero, jumps can be detected as the threshold goes to zero slower than the modulus of continuity of the Brownian motion paths. For more details, one can refer to Mancini [15].

Moreover, assuming the jump size γ is normally distributed with mean 0 and variance σ_J^2, as is shown in Mancini and Renò [17]

$$c_{i,n} = \frac{1}{2F_{\mathcal{N}}(-\sqrt{\vartheta(\delta)}/\sigma_J)},$$

where $F_{\mathscr{N}}(x)$ is the cumulative normal distribution function.

Moreover, the variance of γ can be estimated through the method of moment by

$$m_2(\sqrt{\vartheta(\delta)}) = \sigma_J^2 + \frac{\sqrt{\vartheta(\delta)}\sigma_J exp\left(-\frac{\vartheta(\delta)}{2\sigma_J^2}\right)}{F_{\mathscr{N}}(-\sqrt{\vartheta(\delta)}\sigma_J)\sqrt{2\pi}},$$

where $m_2(\sqrt{\vartheta(\delta)})$ is the second moment of jump size $\gamma \sim \mathscr{N}(0, \sigma_J^2)$ conditional to $\mid \gamma \mid \geq \sqrt{\vartheta(\delta)}$.

Remark 2.2. The Nadaraya-Watson estimator can be seen as uniform weight $w_i = 1/n$, while the weights of local linear estimator w_i^{LL} and reweighted Nadaraya-Watson estimator w_i^{RNW} depends on the design points and locations. With simple calculation, one can see w_i^{LL} satisfy the constraint (4) automatically, which is crucial for the asymptotic bias reduction and boundary correction, especially for the right boundary points. As a consequence, imposing the constraint (4) for the optimization problem, the reweighted Nadaraya-Watson estimator possesses sample bias reduction and automatic boundary correction as local linear estimator. Beyond that, the non-negativity restriction (3) modifies the negativity problem of local linear fitting. For the numerical solution of w_i^{RNW} we need in Section 4, we adopt the algorithm suggested in Xu [21].

3. Assumptions and Main Results

We impose the following assumptions throughout the paper. Assume that $\mathscr{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$ is the range of the process X_t.

Assumption 1. *For model (1), the coefficients $\mu_t := \mu(X_{t-})$ and $\sigma_t = \sigma(X_{t-})$ are bounded, twice continuously differentiable and progressively measurable processes with càdlàg paths satisfying the following conditions:*

(i) For each $n \in \mathbb{N}$, there exists a positive constant L_n such that for any $|x| \leq n$, $|y| \leq n$,

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq L_n|x - y|;$$

(ii) There exists a positive constant c, such that for all $x \in \mathbb{R}$,

$$|\mu(x)| + |\sigma(x)| \leq c(1 + |x|);$$

(iii) $\lambda(x)$ is bounded, nonnegative and twice continuously differentiable, $\lambda^{'}(x)$ is also bounded.

Assumption 2. *The solution of model (1) is Harris recurrent.*

Assumption 3. *The kernel function $K(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is a continuous differentiable and bounded density function, such that $\int_{-\infty}^{\infty} K(u)du = 1$, $\int_{-\infty}^{\infty} |K^{'}(u)|^2 du < \infty$ and $K_i^j = \int_{-\infty}^{\infty} u^j K^i(u)du < \infty$ for $i \leq 2,\ j \leq 2$.*

Assumption 4. *There exists a fixed constant c such that $\forall \varepsilon > 0$, $P\{|\gamma_i| < \varepsilon\} \leq c\varepsilon$ and the jump sizes $\{\gamma_i\}_i$ of J_t in model (1) are independent of N_t.*

Assumption 5. *The threshold function $\vartheta(\delta) = \delta^{\eta}, \eta \in (0,1)$, and $n\delta^{1+\eta/2} \to 0$ as $\delta \to 0$.*

Assumption 6. *The bandwidth parameter is of the form $h(= h_n) = \delta^{\phi}$ with $\phi \in (0, \eta/2)$ and as $n \to \infty$, $T = n\delta \to \infty$, $\delta \to 0$ and*

$$\frac{(\delta ln\frac{1}{\delta})^{\frac{1}{2}}}{h^2} \to 0, \quad \frac{(\delta ln\frac{1}{\delta})^{\frac{1}{2}}}{h}\hat{L}_X(T,x) \xrightarrow{a.s.} 0, \;\; h\hat{L}_X(T,x) \xrightarrow{a.s.} \infty, \quad h^5\hat{L}_X(T,x) = O_P(1)$$

for x visited by X_t, where $\hat{L}_X(T,x) = \frac{1}{h}\sum_{i=1}^{n} K\big(\frac{X_{t_{i-1}}-x}{h_n}\big)\delta$.

Remark 3.1. The Assumption (1) guarantees that the SDE (1) has a unique strong solution which is adapted and right continuous with left limits on $[0,T]$, see Jacod and Shiryaev [13] for more details.

The Assumption (2) means that starting from any initial level $x \in \mathscr{D}$, the process X_t visits a generic set A an infinite number of times as $T \to \infty$, almost surely. It is crucial to achieve consistence and asymptotic normality of the underlying estimator. Besides, the Assumption (2) is a milder assumption than stationarity. To be specific, stationary processes are Harris recurrent, but Harris recurrent processes do not have to be stationary.

In the next theorem, we will obtain the corresponding asymptotic normality of reweighted Nadaraya-Watson threshold estimators (2) by letting $n \to \infty$, $T \to \infty$ and $\delta = T/n \to 0$.

Theorem 1. *For the model (1), under assumptions (1)-(6), we also assume that if $c_{i,n}$ is a double array of constants with $i = 1, ..., n$ such that $\forall x$, $\sqrt{h\hat{L}_X(T,x)}sup_i|1 - c_{i,n}| \to 0$ as $n \to \infty$, then, for each x visited by X, as $\delta \to 0$, $T \to \infty$ and $n \to \infty$, we have*

$$\sqrt{h\hat{L}_X(T,x)}\left(\hat{\lambda}_{RNW}(x) - \lambda(x) - \frac{1}{2}\lambda''(x)K_1^2 \cdot h^2\right) \xrightarrow{d} \mathscr{N}\left(0, \lambda(x)K_2^0\right),$$

Remark 3.2. For the positive recurrent case of Assumption (2), the local time $\bar{L}_X(T,x)$ increases consistently with T as

$$\frac{\bar{L}_X(T,x)}{T} \xrightarrow{a.s.} p(x) = \frac{\phi(dx)}{\phi(\mathscr{D})}, \;\; \forall x \in \mathscr{D}. \tag{5}$$

where $\phi(x)$ is the unique invariant measure of X_t. So we can deduce corollary 1 as follows by means of Theorem 1 easily with the property (5) above.

Corollary 1 Under Assumptions (1), (2) with positive Harris recurrent condition and (3)-(6), then we can obtain

$$\sqrt{n\Delta_n h}\left(\hat{\lambda}_{LL}(x) - \lambda(x) - \frac{1}{2}\lambda''(x)K_1^2 \cdot h^2\right) \xrightarrow{d} \mathscr{N}\left(0, \frac{\lambda(x)K_2^0}{p(x)}\right).$$

Remark 3.3. Optimal bandwidth selection is technically important when applying nonparametric estimation. Based on Theorem 1, the asymptotic mean-square-error (MSE) at a generic level $x \in \mathscr{D}$ is

$$\left(\frac{1}{2}\lambda''(x)K_1^2\right)^2 + \frac{\lambda(x)K_2^0}{h\hat{L}_X(T,x)}$$

To minimize the order of the asymptotic MSE as suggested in Fan and Gijbels [8], the optimal bandwidth of reweighted Nadaraya-Watson threshold estimator (2) is given

$$h_{n,opt}^{RNW} = \left(\frac{4\lambda(x)K_2^0}{\hat{L}_X(T,x)\left[\lambda''(x)K_1^2\right]^2}\right)^{\frac{1}{5}} = O_P\left(L_X(T,x)^{-\frac{1}{5}}\right).$$

For the further discussion of optimal bandwidth selection, one can refer to Aït-Sahalia and Park [2] and Wang and Zhou [20].

4. Simulation Study

In this section, we conduct a Monte Carlo simulation experiment to compare the finite-sample performance of Nadaraya-Watson threshold estimator and reweighted Nadaraya-Watson threshold estimator. For simplicity, the reweighted Nadaraya-Watson threshold estimator proposed in this section is denoted as RNW, Nadaraya-Watson estimator studied in Mancini and Renò [17] is denoted as NW.

Through the section, we consider the jump-diffusion model defined as

$$dX_t = \left(\mu + \frac{\sigma^2}{2}\right)X_t dt + \sigma X_t dW_t + dJ_t, \tag{6}$$

where $\mu = 0.087,\ \sigma = 0.178$, as once used in Fan and Zhang [9]. $J_t = \sum_{n=1}^{N_t} \gamma_{t_n}$ with the intensity function $\lambda(x) = x$ and jump size $\gamma_{t_n} \sim \mathscr{N}(0, 0.036^2)$, where t_n is the $n-$th jump of the Poisson process N_t.

One sample path of the X_t with $T = 10,\ n = 480,\ X_0 = 0.4,\ \Delta_n = \frac{T}{n} = \frac{1}{48}$ generated by the Euler-Maruyama scheme according to (6) is shown in FIG 1.

To construct the threshold estimator, we use Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ and the practical bandwidth selected as $h = 5 \cdot \hat{S} \cdot T^{-\frac{1}{5}}$ for both RNW and NW estimator, where $\hat{S}$ denotes the standard deviation of the data. Moreover, The threshold function $\vartheta(\delta) = 3\sqrt{h_t}$ is implemented as that in Mancini and Renò [17], where h_t is the filtered variance form GARCH(1, 1).

The corresponding biases of NW and RNW estimators at various quantile points of sample X_t are shown in Table 1. It can be observed that RNW estimator performs better at most of the quantile points, especially at boundary points.

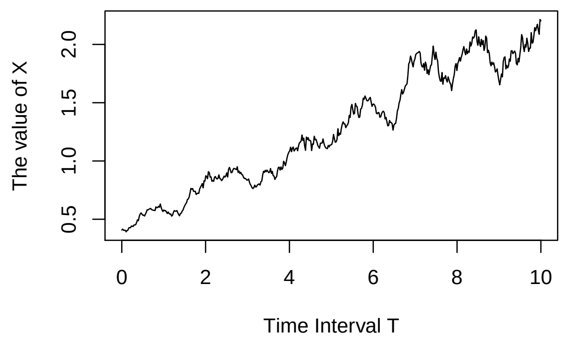

FIG 1. *One path of process X_t*

TABLE 1
The biases of NW and RNW estimators at various quantile points of sample X_t

Bias	Quantile points of sample X_t								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
NW	0.3057	0.2279	0.1503	0.0728	-0.0045	-0.0816	-0.1585	-0.2353	-0.3119
RNW	-0.1146	-0.0468	-0.0167	0.0137	0.0497	0.1070	0.1883	0.1645	-0.0325

Furthermore, we will compare the overall finite-sample performance NW estimator and RNW estimator via the following four measures used in Fan et al [7].

Measure 1: Absolute Mean Error (AME): AME $= \frac{1}{N}\left|\sum_{k=1}^{N}[\hat{\lambda}(x_k) - \lambda(x_k)]\right|$.

Measure 2: Root Mean Square Error (RMSE): RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}[\hat{\lambda}(x_k) - \lambda(x_k)]^2}$.

Measure 3: Ideal Mean Absolute Deviation Error (IMADE): IMADE $= \frac{1}{N}\sum_{k=1}^{N}\left|\hat{\lambda}(x_k) - \lambda(x_k)\right|$.

Measure 4: Relative Ideal Mean Absolute Deviation Error (RIMADE):

$$RIMADE = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|\hat{\lambda}(x_k) - \lambda(x_k)\right|}{\lambda(x_k)},$$

where $\hat{\lambda}(x)$ is the nonparametric threshold estimator of $\lambda(x)$ and $\{x_k\}_1^N$ are chosen uniformly to cover the range of sample path of X_t.

Numerical results are list in Table 2, through which we can find that the RNW estimator performs better than the NW estimator for various measures.

Finally, the QQ plots of RNW estimator for $\lambda(x) = x$ are displayed in FIG 2, which reveals the normality of the underlying nonparametric estimator and confirms the result in Theorem 1.

TABLE 2
Comparisons between NW and RNW estimators for intensity function with various measures

Measure	estimators	$\lambda(x) = x$	$\lambda(x) = 3x$	$\lambda(x) = 5x$
AME	NW	0.0038	0.3072	0.3690
	RNW	0.0002	0.1742	0.2915
RMSE	NW	0.2254	0.3165	0.3936
	RNW	0.1335	0.2408	0.3489
IMADE	NW	0.1952	0.3072	0.3690
	RNW	0.1050	0.2040	0.3186
RIMADE	NW	0.2613	0.7996	0.8979
	RNW	0.1346	0.4725	0.6793

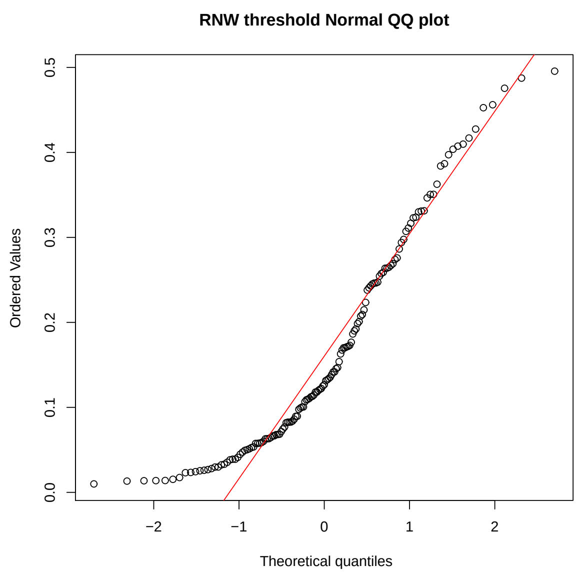

FIG 2. *QQ plot of RNW Threhold Estimators for $\lambda(x) = x$*

5. Conclusion

In this paper, combined with threshold technique, we introduce the reweighted Nadaraya-Watson estimator of the intensity function in a jump-diffusion model. Though how to identify the frequency of jumps is a meaningful topic, there are few literatures on how to estimate the jump intensity function independently. So the nonparametric estimator for the intensity function proposed in this paper is really a supplement to this field. Theoretically, compared with the popular local linear estimator, the reweighted Nadaraya-Watson estimator possesses its fine properties, especially the bias reduction and boundary correction, and further fixes the negativity problem. Practically, simulation study in Section 4 shows that, in finite samples, the underlying estimator performs better than the traditional Nadaraya-Watson estimator, and is always positive.

6. The Proof of Main Result

We recall that $\delta = \frac{T}{n}, t_i = i\delta$. Here we write $dX_t = dY_t + dJ_t$, where $dY_t = \mu(X_t)dt + \sigma(X_t)dW_t$. Denote for an integer i,

$$\Delta_i\hat{Y} = \Delta_i X I_{\{(\Delta_i X)^2 \le \vartheta(\delta)\}},\ \Delta_i\hat{J} = \Delta_i X I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}},\ \Delta_i\hat{N} = I_{\{(\Delta_i X)^2 > \vartheta(\delta)\}}.$$

For any bounded process Z, we denote by $\bar{Z} = \sup_{u\in[0,T]}|Z_u|$. Throughout this article, we use C to denote a generic constant, which may vary from line to line. By $\sigma.W$ we denote the stochastic integral of σ with respect to W. We denote by $\left(\tau_j\right)_{j\in\mathbb{N}}$ the jump instants of J_t and by $\tau^{(i)}$ the instant of the first jump in $(t_{i-1}, t_i]$, if $\Delta_i N \ge 1$.

Lemma 6.1. ***[The occupation time formula]*** *Let X_t be a semimartingale with local time $(L_X(\cdot, a))_{a\in\mathscr{D}}$. Let g be a bounded Borel measurable function. Then, a.s.*

$$\int_{-\infty}^{\infty} L_X(t,a)g(a)da = \int_0^t g(X_{s-})d[X]_s^c,$$

where $[X]^c$ is the continuous part of the quadratic variation of X and $\mathscr{D}$ denotes the the admissible range of the process of interest.

Lemma 6.2. ***[Jacod's stable convergence theorem]*** *For all n, $(\zeta_{n,i} : i \ge 1)$ defined on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F})_{t\ge 0}, P)$ are a sequence of $\mathbb{R}-$valued and $\mathscr{F}_{i\delta}-$measurable variables. Assume there exists a continuous adapted $\mathbb{R}-$valued process of finite variation B_t and a continuous adapted and increasing process C_t, for any $t > 0$, we have*

$$\sup_{0\le s\le t}\Big|\sum_{i=1}^{[s/\delta]} E\big[\zeta_{n,i} \mid \mathscr{F}_{(i-1)\delta}\big] - B_s\Big| \xrightarrow{P} 0,$$

$$\sum_{i=1}^{[t/\delta]}\Big(E\big[\zeta_{n,i}^2 \mid \mathscr{F}_{(i-1)\delta}\big] - \mathbb{E}^2\big[\zeta_{n,i} \mid \mathscr{F}_{(i-1)\delta}\big]\Big) - C_t \xrightarrow{P} 0,$$

$$\sum_{i=1}^{[t/\delta]} E\big[\zeta_{n,i}^4 \mid \mathscr{F}_{(i-1)\delta}\big] \xrightarrow{P} 0.$$

Assume also

$$\sum_{i=1}^{[t/\delta]} E\big[\zeta_{n,i}\Delta_n^i H \mid \mathscr{F}_{(i-1)\delta}\big] \xrightarrow{P} 0,$$

where either H is the Wiener process W or is any bounded martingale orthogonal (in the martingale sense) to W and $\Delta_n^i H = H_{i\delta} - H_{(i-1)\delta}$.

Then the processes

$$\sum_{i=1}^{[t/\delta]} \zeta_{n,i} \overset{\mathcal{S}-\mathcal{L}}{\longrightarrow} B_t + M_t,$$

where $\mathcal{S}-\mathcal{L}$ *denotes the stable convergence in law,* M_t *is a continuous process defined on an extension* $\big(\widetilde{\Omega}, \widetilde{P}, \widetilde{\mathscr{F}}\big)$ *of the filtered probability space* $\big(\Omega, P, \mathscr{F}\big)$ *and which, conditionally on the* $\sigma-$*filter* $\mathscr{F}$*, is a centered Gaussian* $\mathbb{R}-$*valued process with* $\widetilde{E}\big[M_t^2 \mid \mathscr{F}\big] = C_t$.

Remark 6.1. For lemma 6.2, one can refer to Jacod [14] (Lemma 4.4) for more details. The stable convergence implies the following crucial property, which is fundamental for the mixed normal distribution with random variance of the reweighted Nadaraya-Watson estimator, detailed in the proof of Theorem 1.

If $Z_n \overset{\mathcal{S}-\mathcal{L}}{\longrightarrow} Z$ and if Y_n and Y are variables defined on $(\Omega, \mathscr{F}, P)$ and with values in the same Polish space F, then

$$Y_n \overset{P}{\longrightarrow} Y \qquad \Rightarrow \qquad (Y_n,\ Z_n) \overset{\mathcal{S}-\mathcal{L}}{\longrightarrow} (Y,\ Z),$$

which implies that $Y_n \times Z_n \overset{\mathcal{S}-\mathcal{L}}{\longrightarrow} Y \times Z$ through the continuous function $g(x, y) = x \times y$.

Lemma 6.3. *For model (1), when* μ_t *and* σ_t *are bounded, then a.s for small* δ*,*

$$\sup_{i \in \{1,\ldots,n\}} \frac{\left| \int_{t_{i-1}}^{t_i} \mu_s ds + \int_{t_{i-1}}^{t_i} \sigma_s dW_s \right|}{\sqrt{2\delta \log \frac{1}{\delta}}} \leq C$$

where C *is a constant.*

This lemma characterizes the property of uniform boundedness of the increments of X_t *in* $[t_{i-1}, t_i]$ *when no jump occurs in this interval. For discussions and proof of this lemma, one can refer to Mancini [16].*

Lemma 6.4 (Mancini and Renò [17])**.** *Assume that* N_t *is a doubly stochastic Poisson process with an intensity process* λ_t*, if* $\lambda(\cdot)$ *is bounded, then uniformly for all* $i = 1, ..., n$*,*

$$P((N_{i\delta} - N_{(i-1)\delta}) \geq 1) = O(\delta),$$

$$P((N_{i\delta} - N_{(i-1)\delta}) \geq 2) = O(\delta^2).$$

Lemma 6.5. *In model (1), if assumptions* (1)−(3) *are satisfied and* $\frac{\delta ln(1/\delta)}{h^2} \to 0$ *as* $n \to \infty$*, then we have*

$$\frac{1}{h} \sum_{i=1}^{n} K^j \left(\frac{X_{t_{i-1}} - x}{h} \right) \left(\frac{X_{t_{i-1}} - x}{h} \right)^k \delta \ \xrightarrow{\text{a.s.}} \ \frac{K_j^k L_X(T, x)}{\sigma^2(x)}$$

where $K_j^k L_X(T, x) = L_X(T, x) \int_R K^j(u) u^k du$ *for all* x *visited by* X*, as* $\delta \to 0$ *and* $h \to 0$*. One can refer to Song and Wang [19] or Xu [21] for detailed proof.*

Lemma 6.6 (Xu [21]). *In model (1), if Assumptions (1) − (3) are satisfied and let K_1^0 be defined in Assumptions (3) there exists a σ-finite measure ϕ, for the lagrangian multiplier θ, we have*

$$\theta = O_{a.s}(h) \qquad \theta = \frac{hK_1^0\phi'(x)}{K_2^2\phi(x)} + O_{a.s.}(h^3)$$

Lemma 6.7. *In model (1), if Assumptions (1) − (3) are satisfied are satisfied and $\frac{\delta ln(1/\delta)}{h^2} \to 0$ as $n \to \infty$, then we have*

$$\frac{\sum_{i=1}^n \left[nw_{i-1}^{RNW} K(\frac{X_{t_{i-1}}-x}{h})\right]^j}{\sum_{i=1}^n \left[K(\frac{X_{t_{i-1}}-x}{h})\right]^j} \xrightarrow{a.s} 1$$

for $j = 1,\ 2,\ 4$.

Proof. Look at $j = 1$ first. In view of Lemma (6.6), $|\theta(X_{t_i} - x)K_h(X_{t_i} - x)| < 1$ for a sufficiently large n, and so by Taylor expansion we have

$$\begin{aligned}
&\frac{\sum_{i=1}^n nw_{i-1}^{RNW} K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} \\
= \quad & \frac{\sum_{i=1}^n \frac{1}{1+\theta(X_{t_{i-1}}-x)K_h(X_{t_{i-1}}-x)} K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} \\
= \quad & \frac{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} - \theta \frac{\sum_{i=1}^n (X_{t_{i-1}} - x)K_h(X_{t_{i-1}} - x)K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} \\
& + \theta^2 \frac{\sum_{i=1}^n ((X_{t_{i-1}} - x)K_h(X_{t_{i-1}} - x))^2 K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} + O_{a.s}(h^3)
\end{aligned}$$

Based on Lemma (6.5), we can obtain

$$\frac{\sum_{i=1}^n K^k(\frac{X_{t_{i-1}}-x}{h})(\frac{X_{t_{i-1}}-x}{h})^j}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} \xrightarrow{\text{a.s.}} \int u^j K^k(u)du$$

The right side of (6) vanishes when j is an odd integer. So by (6) and Lemma (6.6), we have

$$\frac{\sum_{i=1}^n nw_i^{RNW} K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} = \frac{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})}{\sum_{i=1}^n K(\frac{X_{t_{i-1}}-x}{h})} + O_{a.s}(h^2) \xrightarrow{\text{a.s.}} 1$$

Thus (6.7) holds for $j = 1$. Similarly we can show that (6.7) holds for $j = 2$, 4 □

Combined with Lemma 6.5, in fact we have:

$$\frac{1}{h}\sum_{i=1}^{n}\left[nw_{i-1}^{RNW}K\left(\frac{X_{t_{i-1}}-x}{h}\right)\right]^{j}\delta \xrightarrow{\text{a.s.}} \frac{K_j^0 L_X(T,x)}{\sigma^2(x)}$$

for $j = 1,\, 2,\, 4$.

6.1. The proof of Theorem 1

Proof. We show the convergence of $\hat{\lambda}_{RNW}(x)$. Denote

$$K_{i-1}^{\star} := nw_{i-1}^{RNW}K\left(\frac{X_{t_{i-1}}-x}{h}\right),$$

$$\hat{L}_X(T,x) := \frac{1}{h}\sum_{i=1}^{n}K\left(\frac{X_{t_{i-1}}-x}{h}\right)\delta,$$

and

$$\hat{L}_T^{\star}(x) := \frac{1}{h}\sum_{i=1}^{n}nw_{i-1}^{RNW}K\left(\frac{X_{t_{i-1}}-x}{h}\right)\delta.$$

Write

$$\begin{aligned}
&\sqrt{h\hat{L}_X(T,x)}(\hat{\lambda}_{RNW}(x)-\lambda(x)) \\
= \;& \sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}(c_{in}-1)I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}}{\hat{L}_T^{\star}(x)} \\
&+\sqrt{h\hat{L}_X(T,x)}\Big(\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}}{\hat{L}_T^{\star}(x)}-\lambda(x)\Big),
\end{aligned}$$

we now show that both of the two parts tend to zero in probability.

Firstly, for the second part of $\sqrt{h\hat{L}_X(T,x)}(\hat{\lambda}_{RNW}(x)-\lambda(x))$,

$$\begin{aligned}
S_{n,T} \;:=\;& \sqrt{h\hat{L}_X(T,x)}\Big(\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}}{\hat{L}_T^{\star}(x)}-\lambda(x)\Big) \\
=\;& \sqrt{h\hat{L}_X(T,x)}\Big(\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}(I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}-\Delta_i N)}{\hat{L}_T^{\star}(x)} \\
&+\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\big(\Delta_i N-\int_{t_{i-1}}^{t_i}\lambda_s ds\big)}{\hat{L}_T^{\star}(x)}+\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}\big(\lambda_s-\lambda(x)\big)ds}{\hat{L}_T^{\star}(x)}\Big) \\
:=\;& S_{1_{n,T}}+S_{2_{n,T}}+S_{3_{n,T}}.
\end{aligned}$$

As for $S_{1_{n,T}}$, it's sufficient to show that

$$\sum_{i=1}^{n}K_{i-1}^{\star}(I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}-\Delta_i N)\xrightarrow{P}0$$

Write

$$\sum_{i=1}^{n} K_{i-1}^{\star}(I_{\{(\Delta_i X)^2>\vartheta(\delta)\}} - \Delta_i N) = \sum_{i=1}^{n} K_{i-1}^{\star}(I_{\{(\Delta_i X)^2>\vartheta(\delta)\}} - \Delta_i N) \times \left[I_{\langle\Delta_i N=0\}} + I_{\{\Delta_i N=1\}} + I_{\{\Delta_i N\geq 2\}}\right]$$

We can show that each term above tends to zero in probability similarly as the detailed technical proof of (A.14)-(A.18) in Mancini and Renò [17] replacing $K\left(\frac{X_{t_{i-1}}-x}{h}\right)$ with $K_{i-1}^{\star}$.

As for $S_{2_{n,T}}$, In order to obtain

$$\sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n} K_{i-1}^{\star}\left(\Delta_i N - \int_{t_{i-1}}^{t_i}\lambda_s ds\right)}{\hat{L}_T^{\star}(x)} \xrightarrow{d} N\left(0, \lambda(x)K_2^0\right),$$

we firstly show that the numerator of

$$\frac{\frac{1}{T\sqrt{Th}}\sum_{i=1}^{n} K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)}{\frac{\hat{L}_T^{\star}(x)}{T\sqrt{T}}} := \frac{\frac{1}{T\sqrt{Th}}\sum_{i=1}^{n} K_{i-1}^{\star}\left(\Delta_i N - \int_{t_{i-1}}^{t_i}\lambda_s ds\right)}{\frac{\hat{L}_T^{\star}(x)}{T\sqrt{T}}}$$

converges stably in law to M_1, where $\bar{\nu}(dt) = N_t dt - \lambda(X_{t-})dt$, M_1 is a Gaussian martingale defined on an extension $\left(\tilde{\Omega}, \tilde{P}, \tilde{\mathscr{F}}\right)$ of our filtered probability space and having $\tilde{E}[M_1^2|\mathscr{F}] = \frac{\lambda(x)}{\sigma^2(x)} \cdot \frac{L_X(T,x)}{T^3} \cdot K_2^0$.

Denote $\sum_{i=1}^n q_i := \frac{1}{T\sqrt{Th}}\sum_{i=1}^{n} K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)$ and Jacod's stable convergence theorem tells us that the following arguments,

$$V_1 = \sum_{i=1}^{n} E_{i-1}[q_i] \xrightarrow{P} 0,$$

$$V_2 = \sum_{i=1}^{n}\left(E_{i-1}[q_i^2] - E_{i-1}^2[q_i]\right) \xrightarrow{P} \frac{\lambda(x)}{\sigma^2(x)} \cdot \frac{L_X(T,x)}{T^3} \cdot K_2^0,$$

$$V_3 = \sum_{i=1}^{n} E_{i-1}[q_i^4] \xrightarrow{P} 0,$$

$$V_4 = \sum_{i=1}^{n} E_{i-1}[q_i\Delta_i H] \xrightarrow{P} 0,$$

implies $\sum_{i=1}^n q_i \xrightarrow{\mathcal{S}-\mathcal{L}} M_1$, either $H = W$ or H is any bounded martingale orthogonal (in the martingale sense) to W, $E_{i-1}[\ \cdot\ ] = E[\ \cdot\ |X_{t_{i-1}}]$.
For V_1, considering $K_{i-1}^{\star}$ is measurable with respect to the σ-algebra generated by $\{X_u, 0 \leq u \leq t_{i-1}\}$, q_i^L is a martingale difference sequence, so $V_1 \equiv 0$.
For V_2,

$$V_2 = \sum_{i=1}^{n}\left(E_{i-1}[q_i^2] - E_{i-1}^2[q_i]\right)$$

$$\begin{aligned}
&= \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}E_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)\right)^2\right] = \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}E_{i-1}\left[\int_{t_{i-1}}^{t_i}\lambda(X_{s-})ds\right] \\
&= \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}E_{i-1}\left[\int_{t_{i-1}}^{t_i}\lambda(X_{t_{i-1}})ds + \int_{t_{i-1}}^{t_i}(\lambda(X_{s-})-\lambda(X_{t_{i-1}}))ds\right] \\
&= \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}\lambda(X_{t_{i-1}})\delta + \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}E_{i-1}\left[\int_{t_{i-1}}^{t_i}(\lambda(X_{s-})-\lambda(X_{t_{i-1}}))ds\right] \\
&:= V_{21}+V_{22}.
\end{aligned}$$

For V_{21} :

$$\begin{aligned}
V_{21} &:= \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}\lambda(X_{t_{i-1}})\delta \\
&= \frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}\int_{t_{i-1}}^{t_i}(\lambda(X_{t_{i-1}})-\lambda(X_s))ds \\
&\quad + \frac{1}{T^3h}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}(K_{i-1}^{\star 2}\lambda(X_s)-K_s^{\star 2}\lambda(X_s))ds \\
&\quad + \frac{1}{T^3h}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}K_s^{\star 2}\lambda(X_s)ds \\
&:= V_{211}+V_{212}+V_{213},
\end{aligned}$$

where $V_{211} \overset{a.s.}{\to} 0$ and $V_{212} \overset{a.s.}{\to} 0$ can be dealt with in analogy to (b3) and (b4) in Mancini and Renò [17] using Lemma 6.7 and replacing $K\left(\frac{X_{t_{i-1}}-x}{h}\right)$ with $K_{i-1}^{\star}$. Using the occupation time formula and Lemma 6.7, we can obtain that

$$V_{213} \xrightarrow{a.s.} \frac{\lambda(x)}{\sigma^2(x)} \cdot \frac{L_X(T,x)}{T^3} \cdot K_2^0.$$

For V_{22}, we define the random sets for each n,

$$I_{0,n} = \{i \in \{1,...,n\} : \Delta_i N = 0\},$$

and

$$I_{1,n} = \{i \in \{1,...,n\} : \Delta_i N \neq 0\}.$$

Applying the mean-value theorem for $\lambda(\cdot)$, neglecting the terms with $i \in I_{1,n}$ similarly as the detailed proof for Lemma 6.5 and bounding $|X_s - X_{t_{i-1}}|$ by Lemma 6.3 when $i \in I_{0,n}$, we can reach

$$\begin{aligned}
&\frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}\delta \cdot \sup_x |\lambda^{'}(x)| \cdot \sqrt{\delta ln\frac{1}{\delta}} \\
&= O_{a.s.}\big(\frac{1}{T^3h}\sum_{i\in I_{0,n}}K_{i-1}^{\star 2}\delta \cdot \sqrt{\delta ln\frac{1}{\delta}}\big)
\end{aligned}$$

$$
\begin{aligned}
&= O_{a.s.}\big(\frac{1}{T^3h}\sum_{i=1}^{n}K_{i-1}^{\star 2}\delta\cdot\sqrt{\delta ln\frac{1}{\delta}}\big) \\
&= O_{a.s.}\Big(\frac{K_2^0}{\sigma^2(x)}\cdot\frac{L_X(T,x)}{T^3}\Big)\cdot\sqrt{\delta ln\frac{1}{\delta}} \\
&\xrightarrow{a.s.} 0.
\end{aligned}
$$

For V_3, by Burkholder-Davis-Gundy inequality, Lemma 6.5 and Lemma 6.7, we get

$$
\begin{aligned}
V_3 &= \sum_{i=1}^{n}E_{i-1}[q_i^4] \\
&= \frac{1}{T^6h^2}\sum_{i=1}^{n}K_{i-1}^{\star 4}E_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)\right)^4\right] \\
&\le O_p(1)\cdot\frac{1}{T^6h^2}\sum_{i=1}^{n}K_{i-1}^{\star 4}(\delta) \\
&= O_p(1)\cdot\left(\frac{1}{T^5h}\cdot\frac{L_X(T,x)}{\sigma^2(x)T}\right) \\
&\xrightarrow{P} 0.
\end{aligned}
$$

For V_4, when $H=W$, according to the model assumption that J_t is independent of W_t, we easily get $V_4\equiv 0$. Moreover, when H is any bounded martingale orthogonal (in the martingale sense) to W,

$$
\begin{aligned}
\sum_{i=1}^{n}E_{i-1}\left[q_i\Delta_iH\right] &= \frac{1}{T\sqrt{Th}}\sum_{i=1}^{n}K_{i-1}^{\star}E_{i-1}\left[\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)\Delta_iH\right] \\
&= O_P\left(\frac{1}{T\sqrt{Th}}\sum_{i=1}^{n}K_{i-1}^{\star}E_{i-1}\left[\int_{t_{i-1}}^{t_i}\bar{\nu}(ds)\right]\right) \\
&= O_P\left(\frac{1}{T\sqrt{Th}}\sum_{i=1}^{n}K_{i-1}^{\star}\delta^{\frac{1}{2}}\right) \\
&= O_P\left(\frac{\sqrt{nh}}{T^2}\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\delta\right) \\
&= O_P\left(\frac{\sqrt{nh}}{T}\cdot\frac{L_X(T,x)}{\sigma^2(x)T}\right) \\
&\xrightarrow{P} 0,
\end{aligned}
$$

provided the bounded of H such that $\Delta_iH\le C$ for the second equality, Hölder inequality and Burkerholder-Davis-Gundy inequality for the third equality and $\frac{\sqrt{nh}}{T}=o_P(1)$.

As for $S_{3_{n,T}}$,

$$\begin{aligned}
&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}(\lambda_s-\lambda(x))ds}{\hat{L}_T^{\star}(x)}\\
=&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}(\lambda_s-\lambda_{t_{i-1}})ds}{\hat{L}_T^{\star}(x)}+\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}(\lambda_{t_{i-1}}-\lambda(x))ds}{\hat{L}_T^{\star}(x)}\\
:=&D_{1_{n,T}}+D_{2_{n,T}},
\end{aligned}$$

where $\hat{L}_T^{\star}(x)\xrightarrow{\text{a.s.}}\frac{L_X(T,x)}{\sigma^2(x)}$ by Lemma 6.5 and Lemma 6.7.

We now obtain the asymptotic bias for the expression $D_{1_{n,T}}$ and $D_{2_{n,T}}$ above. That is

$$D_{1_{n,T}}=o_P(D_{2_{n,T}})$$

By the Taylor expansion for $\lambda_{t_{i-1}}-\lambda(x)$ up to order 2

$$\lambda_{t_{i-1}}-\lambda(x)=\lambda^{'}(x)(X_{t_{i-1}}-x)+\frac{1}{2}\lambda^{''}(x)(x+\beta(X_{t_{i-1}}-x))(X_{t_{i-1}}-x)^2,$$

where β is a random variable satisfying $\beta\in[0,1]$. Then, according to Lemma 6.5, we have

$$\begin{aligned}
&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}(\lambda_{t_{i-1}}-\lambda(x))ds}{\hat{L}_T^{\star}(x)}\\
=&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\lambda^{'}(x)(X_{t_{i-1}}-x)\Delta}{\hat{L}_T^{\star}(x)}\\
&+\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\frac{1}{2}\lambda^{''}(x)(x+\beta(X_{t_{i-1}}-x))(X_{t_{i-1}}-x)^2\Delta}{\hat{L}_T^{\star}(x)}\\
\xrightarrow{\text{a.s.}}&\frac{1}{2}h^2\lambda^{''}(x)K_1^2
\end{aligned}$$

and by using the mean-value theorem to $\lambda_s-\lambda_{t_{i-1}}$ for $D_{1_{n,T}}$, we get

$$\begin{aligned}
D_{1_{n,T}}=&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}(\lambda_s-\lambda_{t_{i-1}})ds}{\hat{L}_T^{\star}(x)}\\
=&\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\int_{t_{i-1}}^{t_i}\lambda^{'}(\xi_i)(X_s-X_{t_{i-1}})ds}{\hat{L}_T^{\star}(x)}\\
\overset{\text{a.s.}}{\leq}&\frac{\sigma^2(x)}{L_X(T,x)}(\delta ln\frac{1}{\delta})^{\frac{1}{2}}*\sup_x|\lambda^{'}(x)|*\frac{1}{h}\sum_{i\in I_{0,n}}K(\frac{X_{t_{i-1}}-x}{h})\delta+2CN_1\delta\\
\to& O\Big[(\delta ln\frac{1}{\delta})^{\frac{1}{2}}\Big]\\
=&o(h^2),
\end{aligned}$$

by Lemma 6.3 when $i \in I_{0,n}$ for the $\overset{\text{a.s.}}{\leq}$ and the assumption on h and δ. Thus we prove that $D_{1_{n,T}} = o_P(D_{2_{n,T}})$, so the dominant bias arises from $D_{2_{n,T}}$, which is

$$\frac{1}{2}h^2\lambda^{''}(x)K_1^2$$

Finally, for the first part of $\sqrt{h\hat{L}_X(T,x)}(\hat{\lambda}_{LL}(x) - \lambda(x))$,

$$\begin{aligned}
&\sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}(c_{in}-1)I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}}{\hat{L}_T^{\star}(x)} \\
\leq&\sup_i|1-c_{in}|\sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}}{\hat{L}_T^{\star}(x)} \\
=&\sup_i|1-c_{in}|\sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\big(I_{\{(\Delta_i X)^2>\vartheta(\delta)\}}-\lambda(x)\delta\big)}{\hat{L}_T^{\star}(x)} \\
&+\sup_i|1-c_{in}|\sqrt{h\hat{L}_X(T,x)}\frac{\frac{1}{h}\sum_{i=1}^{n}K_{i-1}^{\star}\lambda(x)\delta}{\hat{L}_T^{\star}(x)} \\
=&\sup_i|1-c_{in}|\big(O_p(1)+\lambda(x)\sqrt{h\hat{L}_X(T,x)}\big) \\
=&O_p\big(\sup_i|1-c_{in}|\sqrt{h\hat{L}_X(T,x)}\big)\longrightarrow 0,
\end{aligned}$$

with the help of asymptotic normality for $S_{n,T}$. □

Acknowledgements

I am sincerely grateful to Hanchao Wang and Yuping Song for their useful suggestions.

References

[1] Aït-Sahalia Y, Jacod J. Testing for jumps in a discretely observed process[J]. The Annals of Statistics, 2009, 37(1): 184-222.

[2] Aït-Sahalia Y, Park J Y. Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models[J]. Journal of Econometrics, 2016, 192(1): 119-138.

[3] Bandi F M, Nguyen T H. On the functional estimation of jumpdiffusion models[J]. Journal of Econometrics, 2003, 116(1-2): 293-328.

[4] Cai Z. Weighted nadarayawatson regression estimation[J]. Statistics & probability letters, 2001, 51(3): 307-318.

[5] Tankov P. Financial modelling with jump processes[M]. CRC press, 2003.

[6] Das S R. The surprise element: jumps in interest rates[J]. Journal of Econometrics, 2002, 106(1): 27-65.

[7] Fan J, Fan Y, Jiang J. Dynamic integration of time-and state-domain methods for volatility estimation[J]. Journal of the American Statistical Association, 2007, 102(478): 618-631.
[8] Fan J, Gijbels I. Local polynomial modelling and its applications: monographs on statistics and applied probability 66[M]. CRC Press, 1996.
[9] Fan J, Zhang C. A reexamination of diffusion estimators with applications to financial model validation[J]. Journal of the American Statistical Association, 2003, 98(461): 118-134.
[10] Hall P, Huang L S. Nonparametric kernel regression subject to monotonicity constraints[J]. Annals of Statistics, 2001: 624-647.
[11] Hall P, Presnell B. Intentionally biased bootstrap methods[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1999, 61(1): 143-158.
[12] Hanif M, Wang H C, Lin Z Y. Reweighted Nadaraya-Watson estimation of jump-diffusion models[J]. Science China Mathematics, 2012, 55(5): 1005-1016.
[13] Jacod J, Shiryaev A. Limit theorems for stochastic processes[M]. Springer Science & Business Media, 2013.
[14] Jacod J. Statistics and high frequency data[J]. Statistical methods for stochastic differential equations, 2012, 124: 191-310.
[15] Mancini C. Estimation of the characteristics of the jumps of a general Poisson-diffusion model[J]. Scandinavian Actuarial Journal, 2004, 2004(1): 42-52.
[16] Mancini C. Nonparametric threshold estimation for models with stochastic diffusion coefficient and jumps[J]. Scandinavian Journal of Statistics, 2009, 36(2): 270-296.
[17] Mancini C, Ren R. Threshold estimation of Markov models with jumps and interest rate modeling[J]. Journal of Econometrics, 2011, 160(1): 77-92.
[18] Protter P E. Stochastic integration and differential equations[M]. Springer, Berlin, Heidelberg, 2005: 249-361.
[19] Song Y, Wang H. Central limit theorems of local polynomial threshold estimator for diffusion processes with jumps[J]. Scandinavian Journal of Statistics, 2018, 45(3): 644-681.
[20] Wang H, Zhou L. Bandwidth selection of nonparametric threshold estimator in jumpdiffusion models[J]. Computers & Mathematics with Applications, 2017, 73(2): 211-219.
[21] Xu K L. Reweighted functional estimation of diffusion models[J]. Econometric Theory, 2010: 541-563.