Yuping Song¹ / Weijie Hou¹ / Shengyi Zhou¹

Variance reduction estimation for return models with jumps using gamma asymmetric kernels

¹ Shanghai Normal University, Shanghai, China, E-mail: songyuping@shnu.edu.cn. <https://orcid.org/0000-0002-7506-1719>.

Abstract:

This paper discusses Nadaraya-Watson estimators for the unknown coefficients in second-order diffusion model with jumps constructed with Gamma asymmetric kernels. Compared with existing nonparametric estimators constructed with Gaussian symmetric kernels, local constant smoothing using Gamma asymmetric kernels possesses some extra advantages such as boundary bias correction, variance reduction and resistance to sparse design points, which is validated through theoretical details and finite sample simulation study. Under the regular conditions, the weak consistency and the asymptotic normality of these estimators are presented. Finally, the statistical advantages of the nonparametric estimators are depicted through 5-minute high-frequency data from Shenzhen Stock Exchange in China.

Keywords: continuous-time return model, high frequency financial data, Nadaraya-Watson estimator, resistance to sparse design, variance and bias reduction

DOI: 10.1515/snde-2018-0001

1 Introduction

In financial modeling and econometrics, diffusion models are widely used to describe uncertainty of changes in native economic variables and asset prices, one of the most famous models is the Black & Scholes asset pricing model. Moreover, as an extension of continuous trajectory process, diffusion model with jumps has a wide range of applications in the field of finance and economics. Jump-diffusion process X_t is represented by the following stochastic differential equation:

$$dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathcal{E}} c(X_{t-}, z)r(\omega, dt, dz), \quad (1)$$

where W_t is a standard Brownian motion, $\mu(\cdot)$ and $\sigma(\cdot)$ are the infinitesimal conditional drift and variance respectively, $\mathcal{E} = \mathbb{R} \setminus \{0\}$, $r(\omega, dt, dz) = (p - q)(dt, dz)$, $p(dt, dz)$ is a time-homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ independent of W_t , and $q(dt, dz)$ is its intensity measure, that is, $E[p(dt, dz)] = q(dt, dz) = f(z)dzdt$, $f(z)$ is its Lévy density. Moreover, jump factor in model (1) is introduced statistically to capture the fatter tail behavior of underlying asset and economically to accommodate impact of sudden and large shocks to financial markets, such as macroeconomic announcements or a dramatic interest rate cut by the Federal Reserve, more details discussed in Johannes (2004). However, on the one hand, because the Brownian motion is not differentiated at any point in probability 1, the mentioned model (1) above can not represent the integrated and differentiated process for integrated economic phenomena, that is, the observation of dynamic process at the discrete time is sometimes not easy to obtain, while the cumulation of the process can be directly observed; on the other hand, model (1) and its subsequent derivative models were mainly involved with the price of asset, while less models characterized returns of asset which is a complete and scale-free summary of the investment opportunity for average investors. Fortunately, the second-order diffusion process with jumps (2) provides a valuable model for the solution of these problems.

Here we extend the continuous autoregression of order two in Nicolau (2007) to a discontinuous and realistic case with jumps for return series based on model (1), which is called as the second-order jump-diffusion model (2),

$$\begin{cases} dY_t = X_{t-}dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathcal{E}} c(X_{t-}, z)r(\omega, dt, dz), \end{cases} \quad (2)$$

Yuping Song is the corresponding author.

©2019 Walter de Gruyter GmbH, Berlin/Boston.

where X_t represents the continuously compounded return of underlying assets and Y_t denotes the asset price by means of the cumulation of the returns plus initial asset value. It also can model the integrated econometric phenomena, for example, continuously compounded return series X_t of underlying assets can not be directly observed, while its realization Y_t such as the prices of assets can be directly observed.

Despite of the above two advantages, model (2) can be used commonly in empirical financial for another three advantages. Firstly, model (2) can accommodate nonstationary original process and be made stationary by differencing, which is a more widely used technique in empirical financial, which is verified through the Augmented Dickey-Fuller test statistic in the empirical analysis part. Secondly, compared with the model in Nicolau (2007), model (2) accommodate the impact of sudden and large shocks in combination with jump component, which is also testified through the test statistic proposed in Barndorff-Nielsen and Shephard (2006) for empirical financial data. Thirdly, model (2) can also be employed in empirical finance for integrated volatility. In empirical financial, Nicolau (2008) employed the second-order diffusion process to study the return series of three companies and three stock indices in USA and etc. Based on the generalized likelihood ratio statistics, Yan and Mei (2016) validated the parameter forms of the volatility function for returns of stock indices in Europe, America, Asia (Hong Kong, Japan, Korea) and four companies in USA stock market.

For the return model characterized by second-order diffusion process, many scholars considered parametric and nonparametric estimation for the unknown coefficients in model (2). For the case with $c(x, z) \equiv 0$, Gloter (2001) obtained the parameter estimators for the second-order Ornstein-Uhlenbeck process, and Gloter (2006) gave the minimum deviation parameter estimation for the drift and diffusion coefficients. Not assuming the specific form of the coefficients, Nicolau (2007) systematically studied Nadaraya-Watson estimators, Wang and Lin (2011) presented the local linear estimations using symmetric kernel for bias correction, Hanif (2015) discussed the Nadaraya-Watson estimators using Gamma asymmetric kernel which didn't coincide with the asymmetric kernel-based method introduced in Chen (2000). For model (2) with $c(x, z) \neq 0$, Song (2017) provided the nonparametric estimators for the infinitesimal coefficients $\mu(x)$ and $\sigma^2(x) + \int_{\mathbb{R}} c^2(x, z)f(z)dz$ in high frequency data based on Gaussian symmetric kernel, Chen and Zhang (2015) discussed the local linear estimators for them based on symmetric kernels, Song, Lin, and Wang (2013) proposed a re-weighted Nadaraya-Watson estimator for the infinitesimal conditional expectation, Funke and Schmisser (2018) constructed adaptive nonparametric estimator for the drift coefficient based on penalized least squares method.

However, the current research on the bias corrections for the unknown coefficients in the second-order jump-diffusion process is mainly focused on the simple bias representation (mainly using local linear or re-weighted nonparametric method). According to the asymptotic distribution theorem of the estimators, we can easily find that the mean square error of the estimators is mainly from two parts, one is the mean error of the estimator and the other is the perturbation caused by the asymptotic normal variance. The previous research work does not take variance reduction into full account for these unknown quantities. Based on Gamma asymmetric kernel fully discussed in Chen (2000), Kristensen (2010) considered a kernel-based approach for the realized integrated volatility estimation, Gospodinov and Hirukawa (2012) constructed nonparametric estimators for scalar diffusion models of spot interest rates. Nadaraya-Watson estimators using Gamma asymmetric kernel have some benefits such as variable bandwidth, variance reduction and resistance to sparse design. Compared with Nadaraya-Watson estimators constructed with Gaussian symmetric kernels, firstly, the curve shapes of Gamma asymmetric kernel vary with the smoothing parameter and the location of the design point similar as the variable bandwidth methods. Secondly, the finite variance of the curve estimation decreases, especially for sparse design points, due to the fact that the smaller asymptotic convergence rate and their variances vary along with the location of the design point x , which is shown in Theorem 2 and Theorem 3. Thirdly, Gamma asymmetric kernels are free from boundary bias (allowing a larger bandwidth to pool more data) and achieve the optimal rate of convergence in mean square error within the class of nonnegative kernel estimators with order two. Fourthly, the estimators constructed with Gamma kernels possesses another advantage: nonnegativity for volatility function compared with local linear estimators. In conclusion, asymmetric kernels is a combination of a boundary correction, variance reduction device and a "variable bandwidth" method.

In this paper, we will propose Nadaraya-Watson estimators for $\mu(x)$ and $\sigma^2(x) + \int_{\mathbb{R}} c^2(x, z)f(z)dz$ using Gamma asymmetric kernels in model (2) for variance reduction. It is organized as follows. In Section 2, we propose Nadaraya-Watson estimators with asymmetric kernels and some ordinary assumptions for model (2). We present the asymptotic results in Section 3. Section 4 presents the finite sample performance for two models through Monte Carlo simulation study. Estimators are depicted through some empirical financial data in Section 5. Section 6 concludes. Some technical lemmas and the main proofs are explicitly shown in Appendix A.2.

2 Nadaraya-Watson estimators with asymmetric kernels and assumptions

According to the asymmetric kernels used in Chen (2000), the Gamma kernel function is defined as

$$K_{G(x/h+1,h)}(u) = \frac{u^{x/h} \exp(-u/h)}{h^{x/h+1} \Gamma(x/h+1)} \quad 0 \leq u \leq \infty, \quad (3)$$

where $\Gamma(m) = \int_0^\infty y^{m-1} \exp(-y) dy$, $m > 0$ is the Gamma function and h is the smoothing parameter. Since the density of Gamma distribution has support $[0, \infty)$, the Gamma kernel function does not generate boundary bias for nonnegative variables or nonnegative part of underlying variables. Here we use modified Gamma kernel function $K_{G(x/h+1,h)}(u)$ instead of $K_{G(x/h,h)}(u)$ due to the fact $K_{G(x/h,h)}(u)$ is unbounded near at $x = 0$. The Gamma function has shapes varying with the design point x , which is nonnegative and asymmetric, especially at the boundary points. Furthermore, the asymptotic variance of the nonparametric Gamma kernel estimation depends on the design point x , which yields the optimal rate of convergence in mean integrated squared error of nonnegative kernels.

In this paper we consider the Nadaraya-Watson estimators with Gamma asymmetric kernel for the unknown coefficients of the second-order jump-diffusion model. Different from model (1), nonparametric estimations constructed for the coefficients in second-order jump-diffusion model (2) give rise to new challenges for two main reasons.

On the one hand, we usually get observations $\{Y_{i\Delta_n}; i = 1, 2, \dots\}$ rather than $\{X_{i\Delta_n}; i = 1, 2, \dots\}$. The value of X_{t_i} cannot be obtained from $Y_{t_i} = Y_0 + \int_0^{t_i} X_s ds$ in a fixed sample intervals. Additionally, nonparametric estimations of the unknown qualities in model (2) could not be constructed on the observations $\{Y_{i\Delta_n}; i = 1, 2, \dots\}$ due to the unknown conditional distribution. As Nicolau (2007) showed, with observations $\{Y_{i\Delta_n}; i = 1, 2, \dots\}$ and given that

$$Y_{i\Delta_n} - Y_{(i-1)\Delta_n} = \int_{(i-1)\Delta_n}^{i\Delta_n} X_u du,$$

we can obtain an approximation value of $X_{i\Delta_n}$ by

$$\tilde{X}_{i\Delta_n} = \frac{Y_{i\Delta_n} - Y_{(i-1)\Delta_n}}{\Delta_n}. \quad (4)$$

On the other hand, the Markov properties for statistical inference of unknown qualities in model (2) based on the samples $\{\tilde{X}_{i\Delta_n}; i = 1, 2, \dots\}$ should be built, which are infinitesimal conditional expectations characterized by infinitesimal operators. Fortunately, under Lemma 1 in subsection 7.2 we can build the following infinitesimal conditional expectations for model (2)

$$E \left[\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} | \mathcal{F}_{(i-1)\Delta_n} \right] = \mu(X_{(i-1)\Delta_n}) + O_p(\Delta_n), \quad (5)$$

$$E \left[\frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} | \mathcal{F}_{(i-1)\Delta_n} \right] = \frac{2}{3} \sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3} \int_{\mathcal{G}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n). \quad (6)$$

where $\mathcal{F}_t = \sigma\{X_s, s \leq t\}$. One can refer to Appendix A in Song, Lin, and Wang (2013) for detailed calculations.

We briefly discuss the Gamma Nadaraya-Watson estimators for the coefficients in model (2) based on $\{\tilde{X}_{i\Delta_n}; i = 0, 1, 2, \dots\}$. We firstly construct Gamma Nadaraya-Watson estimators for them based on equations (5) and (6). The Nadaraya-Watson estimators for $\mu(x)$ and $M(x) := \sigma^2(x) + \int_{\mathcal{G}} c(x, z) f(z) dz$ based on the Gamma asymmetric kernel are solutions to the following optimal problem:

$$\arg \min_{\mu(x)} \sum_{i=1}^n \left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \mu(x) \right)^2 K_{G(x/h_n+1,h_n)}(\tilde{X}_{(i-1)\Delta_n}), \quad (7)$$

$$\arg \min_{M(x)} \sum_{i=1}^n \left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - M(x) \right)^2 K_{G(x/h_n+1,h_n)}(\tilde{X}_{(i-1)\Delta_n}), \quad (8)$$

where $K_{G(x/h_n+1, h_n)}(\cdot)$ is a asymmetric Gamma kernel function and the smoothing parameter $h_n \rightarrow 0$ as $n \rightarrow \infty$. The solutions to (7) and (8) are

$$\hat{a}_n(x) = \frac{A_n(x)}{\hat{p}_n(x)}, \quad \hat{b}_n(x) = \frac{B_n(x)}{\hat{p}_n(x)}, \quad (9)$$

where $\hat{p}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n})$, $A_n(x) = \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n}$, $B_n(x) = \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) \frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n}$.

According to the property of Gamma kernel function, we can know that the asymptotic variance of Gamma Nadaraya-Watson estimators for the unknown estimators varies with the design point x in the theorems. For convenience of describing the different asymptotic properties, denote by “interior x ” and “boundary x ” for a design point x that satisfies $x/h_n \rightarrow \infty$ and $x/h_n \rightarrow \kappa$ for some $\kappa > 0$ as $n, T \rightarrow \infty$, respectively.

We now present some assumptions used in the paper. In what follows, let $\mathcal{D} = (l, u)$ with $l \geq -\infty$ and $u \leq \infty$ denote the admissible range of the process X_t , K denotes $K_{G(x/h_n+1, h_n)}$.

Assumption 1

- i. (Local Lipschitz continuity) For each $n \in \mathbb{N}$, there exist a constant L_n and a function $\zeta_n : \mathcal{E} \rightarrow \mathbb{R}_+$ with $\int_{\mathcal{E}} \zeta_n^2(z) f(z) dz < \infty$ such that, for any $|x| \leq n, |y| \leq n, z \in \mathcal{E}$,

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq L_n |x - y|, \quad |c(x, z) - c(y, z)| \leq \zeta_n(z) |x - y|.$$

- ii. (Linear growthness) For each $n \in \mathbb{N}$, there exist ζ_n as above and C , such that for all $x \in \mathbb{R}, z \in \mathcal{E}$,

$$|\mu(x)| + |\sigma(x)| \leq C(1 + |x|), \quad |c(x, z)| \leq \zeta_n(z)(1 + |x|).$$

Remark 1

Assumption 1 guarantees the existence and uniqueness of a solution to X_t in model (1) on the probability space $(\Omega, \mathcal{F}, P)$, see Jacod and Shiryaev (2003).

Assumption 2

The process X_t is ergodic and stationary with a finite invariant measure $\phi(x)$. Furthermore, The process X_t is ρ -mixing with $\sum_{i \geq 1} \rho(i\Delta_n) = O(\frac{1}{\Delta_n^\alpha})$, $n \rightarrow \infty$, where $\alpha < \frac{1}{2}$.

Remark 2

The finite invariant measure implies that the process X_t is positive Harris recurrent with the stationary probability measure $p(x) = \frac{\phi(x)}{\phi(\mathcal{D})}$, $\forall x \in \mathcal{D}$. The hypothesis that X_t is a stationary process is obviously a plausible assumption because for major integrated time series data, a simple differentiation generally assures stationarity. The same condition yielding information on the rate of decay of ρ -mixing coefficients for X_t was mentioned in Assumption 3 of Gugushvili and Spereij (2012).

Assumption 3

For any $2 \leq i \leq n$, g is a differentiable function on $\mathbb{R}$ and $\xi_{n,i} = \theta X_{(i-1)\Delta_n} + (1-\theta)\tilde{X}_{(i-1)\Delta_n}$, $0 \leq \theta \leq 1$, the conditions hold:

- $\lim_{h_n \rightarrow 0} E[|h_n K'(\xi_{n,i}) g(X_{(i-1)\Delta_n})|] < \infty$,
- $\lim_{h_n \rightarrow 0} h_n^{1/2} E[|h_n^2 K'^2(\xi_{n,i}) g(X_{(i-1)\Delta_n})|] < \infty$ for “interior x ”,
- $\lim_{h_n \rightarrow 0} h_n E[|h_n^2 K'^2(\xi_{n,i}) g(X_{(i-1)\Delta_n})|] < \infty$ for “boundary x ”.

Remark 3

According to the procedure for Assumption 3 in Appendix (7.1), we can easily deduce the following results:

- $\lim_{h_n \rightarrow 0} E[|K(X_{(i-1)\Delta_n}) g(X_{(i-1)\Delta_n})|] < \infty$;
- $\lim_{h_n \rightarrow 0} h_n^{1/2} E[|K^2(X_{(i-1)\Delta_n}) g(X_{(i-1)\Delta_n})|] < \infty$ for “interior x ”;

iii. $\lim_{h_n \rightarrow 0} h_n E[|K^2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] < \infty$ for “boundary x ”.

Assumption 4

For all $p \geq 1$, $\sup_{t \geq 0} E[|X_t|^p] < \infty$, and $\int_{\mathbb{R}} |z|^p f(z) dz < \infty$.

Remark 4

This assumption guarantees that Lemma 1 can be used properly throughout the article. If X_t is a Lévy process with bounded jumps (i.e. $\sup_t |\Delta X_t| \leq C < \infty$ almost surely, where C is a nonrandom constant), then $E\{|X_t|^n\} < \infty \forall n$, that is, X_t has bounded moments of all orders, see Protter (2004). This condition is widely used in the estimation of an ergodic diffusion or jump-diffusion from discrete observations, see Florens-Zmirou (1989), Kessler (1997), and Shimizu and Yoshida (2006).

Assumption 5

$\Delta_n \rightarrow 0$, $h_n \rightarrow 0$, $\frac{n\Delta_n}{h_n} \sqrt{\Delta_n \log\left(\frac{1}{\Delta_n}\right)} \rightarrow 0$, $h_n n \Delta_n^{1+\alpha} \rightarrow \infty$, as $n \rightarrow \infty$.

Remark 5

The relationship between h_n and Δ_n is similar as the stationary case in Hanif (2016), (b1), (b2) of A8 in Nicolau (2007) and Assumption 7 in Song (2017). Wang and Zhou (2017) presented the optimal bandwidth of symmetric kernel nonparametric threshold estimator of diffusion function in jump-diffusion models. We will select the optimal smoothing parameter h_n for Gamma asymmetric kernel estimation of second-order jump-diffusion models by means of minimizing the mean square error (MSE) in Remark 6.

3 Large sample properties.

Based on the above assumptions and the lemmas in the following proof procedure part, we have the following asymptotic properties. To simplify notations, we define $x \in \mathcal{D}$ to be a

“interior x ” if “ $x/h_n \rightarrow \infty$ ” or “boundary x ” if “ $x/h_n \rightarrow \kappa$ ”

Theorem 1

If Assumption 1–Assumption 5 hold, then

$$\hat{p}_n(x) := \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) \xrightarrow{p} p(x).$$

Theorem 2

i. Under the Assumption 1–Assumption 5, we have

$$\hat{a}_n(x) := \frac{A_n(x)}{\hat{p}_n(x)} \xrightarrow{p} \mu(x).$$

ii. Under the Assumption 1–Assumption 5, for “interior x ”, if $h_n = O((n\Delta_n)^{-2/5})$, then

$$\sqrt{n\Delta_n h_n^{1/2}} (\hat{a}_n(x) - \mu(x) - h_n B_{\hat{a}_n(x)}) \xrightarrow{d} N\left(0, \frac{M(x)}{2\sqrt{\pi} x^{1/2} p(x)}\right),$$

and for “boundary x ”, if $h_n = O((n\Delta_n)^{-1/3})$, then

$$\sqrt{n\Delta_n h_n} (\hat{a}_n(x) - \mu(x) - h_n B_{\hat{a}_n(x)}) \xrightarrow{d} N\left(0, \frac{M(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)p(x)}\right),$$

where $B_{\hat{a}_n(x)}$ denotes the bias of the estimator of $\hat{a}_n(x)$, that is

$$B_{\hat{a}_n(x)} = \mu'(x) \left\{1 + \frac{x p'(x)}{p(x)}\right\} + \frac{x}{2} \mu''(x).$$

Theorem 3

i. Under the Assumptions Assumption 1–Assumption 5, we have

$$\hat{b}_n(x) := \frac{B_n(x)}{\hat{p}_n(x)} \xrightarrow{p} \sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz.$$

ii. Under the Assumption 1–Assumption 5, for “interior x ”, if $h_n = O((n\Delta_n)^{-2/5})$, then

$$\sqrt{n\Delta_n h_n^{1/2}} (\hat{b}_n(x) - M(x) - h_n B_{\hat{b}_n(x)}) \xrightarrow{d} N\left(0, \frac{\int_{\mathcal{E}} c^4(x, z) f(z) dz}{2\sqrt{\pi} x^{1/2} p(x)}\right),$$

for “boundary x ”, if $h_n = O((n\Delta_n)^{-1/3})$, then

$$\sqrt{n\Delta_n h_n} (\hat{b}_n(x) - M(x) - h_n B_{\hat{b}_n(x)}) \xrightarrow{d} N\left(0, \frac{\int_{\mathcal{E}} c^4(x, z) f(z) dz \Gamma(2\kappa + 1)}{2^{2\kappa+1} \Gamma^2(\kappa + 1) p(x)}\right),$$

where $B_{\hat{b}_n(x)}$ denotes the bias of the estimator of $\hat{b}_n(x)$, that is

$$B_{\hat{b}_n(x)} = M'(x) \left\{ 1 + \frac{x p'(x)}{p(x)} \right\} + \frac{x}{2} M''(x).$$

Remark 6

Theorem 2 and Theorem 3 give the weak consistency and the asymptotic normality of Nadaraya-Watson estimators for the unknown coefficients in model (2) using Gamma asymmetric kernels. In the practical applications, it is very important to consider the choice of the smoothing parameter h_n for the nonparametric estimation using asymmetric kernels. Here we will select the optimal smoothing parameter h_n based on the mean square error (MSE) according to Theorem 2 and Theorem 3. Take $\mu(x)$ for example, for “interior x ”, the optimal smoothing parameter h_n is

$$h_{n,opt} = \left(\frac{1}{n\Delta_n} \cdot \frac{M(x)}{2\sqrt{\pi} x^{1/2} p(x)} \cdot \frac{1}{B_{\hat{a}_n(x)}^2} \right)^{\frac{2}{5}} = O_p \left(\frac{1}{n\Delta_n} \right)^{\frac{2}{5}},$$

for “boundary x ”, the optimal smoothing parameter h_n is

$$h_{n,opt} = \left(\frac{1}{n\Delta_n} \cdot \frac{M(x) \Gamma(2\kappa + 1)}{2^{2\kappa+1} \Gamma^2(\kappa + 1) p(x)} \cdot \frac{1}{B_{\hat{a}_n(x)}^2} \right)^{\frac{1}{3}} = O_p \left(\frac{1}{n\Delta_n} \right)^{\frac{1}{3}}.$$

Note that the optimal bandwidth doesn't coincide with that of the stationary case in Bandi and Nguyen (2003) nor that in Song (2017) in which the optimal smoothing parameters h_n are the same for all design points and are of order $O_p \left(\frac{1}{n\Delta_n} \right)^{\frac{1}{5}}$.

Remark 7

The pointwise estimators considered here and in Chen and Zhang (2015), Song (2017), and Song, Lin, and Wang (2013) are based on kernel approach, whereas the adaptive estimator studied in Funke and Schmisser (2018) was done based on a model selection approach and focused on L^2 -risk. Moreover, Funke and Schmisser (2018) only focused on the regression-type estimator for the drift coefficient and could not provide the exact bias term or the central limit theorem. However, we and Chen and Zhang (2015), Song (2017), and Song, Lin, and Wang (2013) investigated nonparametric estimator of the drift function as well as the volatility function and proved the consistency and asymptotically normal distribution for the underlying estimators.

In our case of variance reduction for the nonparametric estimator, we only focus on the kernel-based estimators for the unknown functions such as drift and volatility, which can provide the central limit theorems. Therefore, in the following parts such as Remark 9 or the simulation section, we will focus on the theoretical and finite-sample numerical comparisons between the proposed approach and the existing methods mentioned in Song (2017), Chen and Zhang (2015), and Song, Lin, and Wang (2013), not Funke and Schmisser (2018).

Remark 8

Here we briefly describe the procedure for constructing the confidence intervals based on asymptotic normality for the further theoretical and numerical comparisons. Take $\mu(x)$ for example. The asymptotic normality of Nadaraya-Watson estimator using Gamma asymmetric kernel for $\mu(x)$ in this paper is different from that in Song (2017), the local linear estimator in Chen and Zhang (2015) and Song, Lin, and Wang (2013) constructed with Gaussian symmetric kernel. For brevity, the estimator proposed in this paper is denoted as AS, the estimator in Song (2017), Chen and Zhang (2015), and Song, Lin, and Wang (2013) as GS. The asymptotic normality for AS is shown in Theorem 2 and the asymptotic normality for GS was

$$\sqrt{h_n n \Delta_n} (\hat{\mu}_n^{GS}(x) - \mu(x)) \xrightarrow{d} N\left(0, \frac{M(x)}{2\sqrt{\pi}p(x)}\right),$$

where $\hat{\mu}_n^{GS}(x)$ denotes the GS estimator for $\mu(x)$.

For “interior x ”, if the smoothing parameter $h_n = O((n\Delta_n)^{-2/5})$, the normal confidence interval for $\mu(x)$ using Gamma asymmetric kernel and Gaussian symmetric kernel at the significance level $100(1 - \alpha)\%$ are constructed as follows,

$$\begin{aligned} I_{\mu, \alpha}^{AS} &= \left[\hat{a}_n(x) - h_n \cdot \hat{B}_{\hat{a}_n(x)} - z_{1-\alpha/2} \cdot \frac{1}{\sqrt{n\Delta_n h_n^{1/2}}} \cdot \sqrt{\frac{\hat{b}_n(x)}{2\sqrt{\pi}x^{1/2}\hat{p}_n^{AS}(x)}}, \right. \\ &\quad \left. \hat{a}_n(x) - h_n \cdot \hat{B}_{\hat{a}_n(x)} + z_{1-\alpha/2} \cdot \frac{1}{\sqrt{n\Delta_n h_n^{1/2}}} \cdot \sqrt{\frac{\hat{b}_n(x)}{2\sqrt{\pi}x^{1/2}\hat{p}_n^{AS}(x)}} \right], \\ I_{\mu, \alpha}^{GS} &= \left[\hat{\mu}_n^{GS}(x) - z_{1-\alpha/2} \cdot \frac{1}{\sqrt{n\Delta_n h_n}} \cdot \sqrt{\frac{\hat{M}_n^{GS}(x)}{2\sqrt{\pi}\hat{p}_n^{GS}(x)}}, \right. \\ &\quad \left. \hat{\mu}_n^{GS}(x) + z_{1-\alpha/2} \cdot \frac{1}{\sqrt{n\Delta_n h_n}} \cdot \sqrt{\frac{\hat{M}_n^{GS}(x)}{2\sqrt{\pi}\hat{p}_n^{GS}(x)}} \right], \end{aligned}$$

where $\hat{a}_n(x)$, $\hat{b}_n(x)$, $\hat{\mu}_n^{GS}(x)$, $\hat{M}_n^{GS}(x)$ denote the estimators of $\mu(x)$, $M(x)$ in (9) using Gamma asymmetric kernel or Gaussian symmetric kernel mentioned in Song (2017), Chen and Zhang (2015), and Song, Lin, and Wang (2013), respectively. $z_{1-\alpha/2}$ is the inverse CDF for the standard normal distribution evaluated at $1 - \alpha/2$. Moreover, the consistent estimators for $p(x)$ are $\hat{p}_n^{AS}(x) = \frac{1}{n} \sum_{i=1}^n K_{\text{Gamma}(x/h_n+1, h)}(\tilde{X}_{(i-1)\Delta_n})$, $\hat{p}_n^{GS}(x) = \frac{1}{nh_n} \sum_{i=1}^n K_{\text{Gaussian}}\left(\frac{x - \tilde{X}_{(i-1)\Delta_n}}{h_n}\right)$. Furthermore, the estimators for the derivative $\mu'(x)$, $\mu''(x)$ and $p'(x)$ in $\hat{B}_{\hat{a}_n(x)}$ of $I_{\mu, \alpha}^{AS}$ can be estimated by taking the derivative of the estimators of $\mu(x)$ in (9) and $\hat{p}_n^{AS}(x)$ above using Gamma asymmetric kernel, one can refer to Fan and Gijbels (1996) for more similar details. For “boundary x ”, one can construct the confidence intervals based on asymptotic normality similarly as the above procedure.

Similarly, the normal confidence interval for $M(x)$ at a spatial point x using Gamma asymmetric kernel and Gaussian symmetric kernel at the significance level $100(1 - \alpha)\%$ can be constructed. The final interval for $M(x)$ should be taken as the intersection of $I_{M, \alpha}^{AS}$ or $I_{M, \alpha}^{GS}$ with $[0, +\infty)$ to coincide with the nonnegativity of the conditional variance $M(x)$.

Remark 9

There are two main differences: on one hand, the convergence rate of Nadaraya-Watson estimator using Gamma asymmetric kernel is different for the location of the design point x such as “interior x ” and “boundary x ”; on the other hand, the variance of Nadaraya-Watson estimator using Gamma asymmetric kernel is inversely proportional to the design x , which shows that the variance decreases as the design point x increases.

Here we briefly commented the theoretical comparison for the convergence rate and the variance of Nadaraya-Watson estimator for “interior x ” and “boundary x ” using Gamma asymmetric kernel or Gaussian symmetric kernel, which dominate the length of the confidence interval. For “interior x ”, the convergence rate of Nadaraya-Watson estimator based on Gamma asymmetric kernel is $\frac{1}{\sqrt{n\Delta_n h_n^{1/2}}}$, which is much smaller than

$\frac{1}{\sqrt{n\Delta_n h_n}}$ of that based on Gaussian symmetric kernel with a given h_n . Additionally, compared with the ones using Gaussian symmetric kernel, the variance of local constant estimator using Gamma asymmetric kernel is inversely proportional to the design x , which shows that the length of the confidence interval decreases as the design point x increases.

For “boundary x ”, although the convergence rate of Nadaraya-Watson estimator based on Gamma asymmetric kernel is the same as that based on Gaussian symmetric kernel, the coefficient in their variance differs a little such as $\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}$ for Gamma asymmetric kernel and $\frac{1}{2\sqrt{\pi}}$ for Gaussian symmetric kernel. Under numeral calculations, from Table 1 we can conclude that when $\kappa \leq 0.7$, the variance based on Gamma asymmetric kernel is larger, whereas when $\kappa \geq 0.75$, the variance based on Gamma asymmetric kernel is smaller than that based on Gaussian symmetric kernel. This reveals that the closer to the origin, the shorter length of confidence interval based on Gaussian symmetric kernel, whereas away from the origin such as the sparse design boundary points, the shorter length of confidence interval based on Gamma asymmetric kernel.

Table 1: The difference between $\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}$ and $\frac{1}{2\sqrt{\pi}}$ in the variance for various κ .

Value of κ	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6
Difference	0.0993	0.0838	0.0701	0.0577	0.0464	0.0362	0.0269	0.0183
Value of κ	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1
Difference	0.0103	0.003	-0.004	-0.01	-0.016	-0.022	-0.027	-0.032
Value of κ	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Difference	-0.053	-0.07	-0.083	-0.095	-0.104	-0.112	-0.12	-0.126
Value of κ	3.25	3.5	3.75	4	4.25	4.5	4.75	5
Difference	-0.132	-0.137	-0.141	-0.145	-0.149	-0.153	-0.156	-0.159

Note that the difference for a κ means the value of $\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}$ minus $\frac{1}{2\sqrt{\pi}}$.

Remark 10

In contrary to the second-order diffusion model without jumps in Hanif (2015), the second infinitesimal moment estimator for second-order jump-diffusion model has a rate of convergence that is the same as the rate of convergence of the first infinitesimal moment estimator. Apparently, this is due to the presence of discontinuous breaks that have an equal impact on all the functional estimates. As Johannes (2004) pointed out, for the conditional variance of interest rate changes, not only diffusion play a certain role, but also jumps account for more than half at lower interest level rates, almost two-thirds at higher interest level rates, which dominate the conditional volatility of interest rate changes. Thus, it is extremely important to estimate the conditional variance as $\sigma^2(x) + \int_{\mathcal{G}} c^2(x, z)f(z)dz$ which reflects the fluctuation of the underlying asset.

4 Monte carlo simulation study

In this section, compared with the existing kernel methods mentioned in Song (2017), Chen and Zhang (2015), and Song, Lin, and Wang (2013) using Gaussian kernels, the benefits of Nadaraya-Watson estimators for the drift function $\mu(x)$ and the volatility function $M(x)$ using the asymmetric kernels are demonstrated through the finite-sample performance for various second-order jump-diffusion models. For brevity, the estimator proposed in this paper is denoted as AS, the estimator in Song (2017) as NW, the estimator in Chen and Zhang (2015) as LL and the estimator in Song, Lin, and Wang (2013) as RNW.

Throughout this section, we employ Gamma kernel $K_{G(x/h_n+1, h_n)}(u) = \frac{u^{x/h_n} \exp(-u/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)}$ and Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. The common bandwidths for the estimators investigated in Song (2017), Chen and Zhang (2015), and Song, Lin, and Wang (2013) using Gaussian kernels are selected as $h = c\hat{S}(n\Delta_n)^{-\frac{1}{5}} = c\hat{S}T^{-\frac{1}{5}}$, where $\hat{S}$ denotes the standard deviation of the data and c represents different constants for different estimators with $c = 2.8$ for $\hat{\mu}_n(x)$ and $c = 1.3$ for $\hat{M}_n(x)$, one can refer to part (4.2) in Xu and Phillips (2011) for more details. In addition, the normal confidence level is assumed to be 95%.

Example 1: Gaussian jumps.

In this subsection, a simple simulation experiment is considered for the model (10) aimed at evaluating the better finite-sampling performance of nonparametric estimators (9) constructed with Gamma asymmetric kernels

$$\begin{cases} dY_t = X_t dt, \\ dX_t = -X_t dt + \sqrt{0.01 + 0.01X_t^2} dW_t + dJ_t, \end{cases} \quad (10)$$

where the coefficients of continuous part are similar as the ones used in Nicolau (2007) and J_t is a compound Poisson jump process, that is, $J_t = \sum_{n=1}^{N_t} Z_{t_n}$ with arrival intensity $\lambda \cdot T = 20$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ corresponding to Bandi and Nguyen (2003), where t_n is the n th jump of the Poisson process N_t . Taking the integral from 0 to t in the second expression of (10), we obtain

$$X_t = -\int_0^t X_{s-} ds + \int_0^t \sqrt{0.01 + 0.01X_{s-}^2} dW_t + \sum_{n=1}^{N_t} Z_{t_n}. \quad (11)$$

By (11) we have

$$Y_t = \int_0^t X_{s-} ds = -(X_t - \int_0^t \sqrt{0.01 + 0.01X_{s-}^2} dW_t - \sum_{n=1}^{N_t} Z_{t_n}). \quad (12)$$

X_t can be sampled by the Euler-Maruyama scheme according to (11), which will be detailed in the following algorithm (one can refer to Cont and Tankov (2004)). One sample trajectory of the differentiated process X_t and integrated process Y_t with $T = 10$, $n = 1000$, $X_0 = 0$ and $Y_0 = 100$ using Algorithm 1 is shown in Figure 1. Through the observation of Figure 1(B), we can find the features of the integrated process Y_t : absent mean-reversion, persistent shocks, time-dependent mean and variance, nonnormality, etc.

Figure 1: Sample Paths of X_t and Y_t for Model (10).

(A) The differentiated process X_t . (B) The integrated process Y_t .

Algorithm 1 (Simulation for trajectories of second-order jump-diffusion model).

Procedures:

Step 1: generate a standard normal random variate V and transform it into $D_i = \sqrt{0.01 + 0.01X_{t_{i-1}}^2} \times \sqrt{\Delta t_i} \times V$, where $\Delta t_i = t_i - t_{i-1} = \frac{T}{n}$ is the observation time frequency;

Step 2: generate a Poisson random variate N with intensity $\lambda \cdot T = 20$;

Step 3: generate N random variables τ_i uniformly distributed in $[0, T]$, which correspond the jump times;

Step 4: generate N random variables $Z_{\tau_i} \sim \mathcal{N}(0, 0.036^2)$, which correspond the jump sizes;

One trajectory for X_t is

$$X_{t_i} = X_{t_{i-1}} - X_{t_{i-1}} \times \Delta t_i + D_i + 1_{\{t_{i-1} \leq \tau_i < t_i\}} \times Z_{\tau_i}.$$

Step 5: By substitution of X_{t_i} in (12), Y_{t_i} can be sampled.

Here we will select the optimal smoothing parameter h_n^{AS} for Nadaraya-Watson estimation using asymmetric kernels based on the mean square error (MSE) according to Theorem 2 and Theorem 3 as shown in Remark 6. Similarly as the common bandwidths selected for estimators constructed with Gaussian kernels, the data-driven bandwidth is represented as $h_n^{AS} = c\hat{\sigma}T^{-\frac{2}{5}}$ for “interior x ” or $h = c\hat{\sigma}T^{-\frac{1}{3}}$ for “boundary x ”, where c is an unknown parameter to be selected. Curve of $MSE(c)$ for the unknown coefficients in model (10) based on Gamma kernels with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ is depicted in Figure 2, from which we can select the optimal parameter $c_{opt} = 1.44$ for drift function and $c_{opt} = 1.11$ for volatility function.

Figure 2: Curve of $MSE(c)$ for the unknown coefficients in model (10) based on Gamma kernels with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

(A) Curve of $MSE(c)$ for drift function. (B) Curve of $MSE(c)$ for volatility function.

The Nadaraya-Watson estimators for drift function $\mu(x) = -x$ and volatility function $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$ in model (10) constructed with Gaussian and Gamma kernels from a sample with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ are shown in Figure 3. We can observe that AS performs better than the others such as NW, LL and RNW, especially at the boundary points, which is also verified through the biases calculated at various points of the sample X_t , in Table 2. Additionally, Figure 4 represents 95% Monte Carlo confidence intervals for these estimators of the unknown coefficients based on Remark 8, and Table 3 shows the length of confidence bands of various estimators for drift and volatility functions. They show that the lengths of 95% Monte Carlo confidence intervals for $\mu(x)$ constructed with Gamma asymmetric kernels are almost shorter

than those constructed with Gaussian symmetric kernels and those for $M(x)$ using Gamma asymmetric kernels are shorter than the others with Gaussian symmetric kernels especially at the sparse design boundary point, which coincides with the conclusion in Remark 9 and reveals variance reduction. In addition, at the sparse design boundary point, the true value of $\mu(x)$ or $M(x)$ cannot fall in the 95% Monte Carlo confidence bands of some nonparametric estimators using Gaussian symmetric kernels, but the 95% Monte Carlo confidence bands constructed with Nadaraya-Watson estimator with Gamma asymmetric kernel.

Figure 3: Various Estimators for unknown coefficients in model (10) with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

(A) Estimators for drift function $\mu(x) = -x$. (B) Estimators for volatility function $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$.

Figure 4: 95% Monte Carlo confidence intervals for unknown coefficients based on various estimators with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

(A) Confidence bands for drift function $\mu(x) = -x$. (B) Confidence bands for volatility function $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$.

Table 2: The Biases of various Estimators for $\mu(x) = -x$ and $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$ at various points of sample X_t with $T = 10$, $n = 1000$, $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

		Bias($\times 10^{-2}$)									
		Various points of Sample X_t									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$\mu(x)$	NW	1.1837	1.9252	2.6625	3.3955	4.1243	4.8489	5.5693	6.2857	6.9979	7.7062
	LL	-0.2409	-0.5381	-0.9528	-1.4934	-2.1687	-2.9872	-3.9576	-5.0882	-6.3875	-7.8635
	RNW	1.1450	0.9574	0.7512	0.0650	-0.6313	-1.5302	-2.4320	-3.3810	-3.8591	-4.2691
	AS	0.0635	0.2263	0.2611	0.2118	0.1113	-0.0170	-0.1563	-0.2954	-0.4265	-0.5447
$M(x)$	NW	-0.0688	-0.0617	-0.0544	-0.0473	-0.0403	-0.0338	-0.0278	-0.0224	-0.0177	-0.0138
	LL	-0.0655	-0.0521	-0.0386	-0.0256	-0.0137	-0.0033	0.0055	0.0125	0.0175	0.0204
	RNW	-0.0511	-0.0295	-0.0049	0.0110	0.0213	0.0220	0.0237	0.0221	0.0192	0.0179
	AS	0.0199	0.0066	-0.0034	-0.0097	-0.0126	-0.0128	-0.0112	-0.0087	-0.0058	-0.0033

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

We will assess the small-sampling performance of Nadaraya-Watson estimators constructed with Gamma asymmetric kernels and those constructed with Gaussian symmetric kernels, for drift function $\mu(x)$ and $M(x)$ via the Mean Square Errors (MSE)

Table 3: The Length of Confidence Bands of various Estimators for $\mu(x) = -x$ and $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$ at various points of sample X_t with $T = 10$, $n = 1000$, $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

Length of Confidence Bands		Various points of Sample Xt									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$\mu(x)$	NW	0.1162	0.1171	0.1185	0.1202	0.1224	0.1250	0.1280	0.1315	0.1356	0.1402
	LL	0.1161	0.1175	0.1197	0.1227	0.1267	0.1319	0.1382	0.1458	0.1547	0.1651
	RNW	0.1195	0.1190	0.1202	0.1191	0.1198	0.1232	0.1271	0.1282	0.1339	0.1323
	AS	0.1094	0.1089	0.1097	0.1117	0.1147	0.1186	0.1236	0.1294	0.1362	0.1439
$M(x)(\times 10^{-3})$	NW	2.3163	2.3716	2.4557	2.5698	2.7159	2.8965	3.1148	3.3748	3.6817	4.0425
	LL	2.3289	2.4108	2.5249	2.6699	2.8449	3.0494	3.2838	3.5490	3.8477	4.1838
	RNW	2.3375	2.3494	2.4965	2.6226	2.7295	3.0321	3.4769	3.8691	3.9116	4.1262
	AS	5.5245	5.0540	4.7451	4.5218	4.3498	4.2108	4.0943	3.9933	3.9034	3.8216

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

The confidence intervals of various estimators are calculated as those mentioned in Remark 8.

$$MSE = \frac{1}{m} \sum_{k=0}^m \{\hat{\mu}(x_k) - \mu(x_k)\}^2 \quad (13)$$

and

$$MSE = \frac{1}{m} \sum_{k=0}^m \{\hat{M}(x_k) - M(x_k)\}^2, \quad (14)$$

where $\hat{\mu}(x)$ or $\hat{M}(x)$ are the various estimators of $\mu(x)$ or $M(x)$, respectively and $\{x_k\}_1^m$ are chosen uniformly to cover the range of sample path of X_t .

The mean square error (MSE) of various estimators for the drift function $\mu(x)$ or volatility function $M(x)$ are calculated under various lengths of observation time intervals T ($= 50, 100, 500$) and sample sizes n ($= 500, 1000, 5000$) with $\Delta_n = \frac{T}{n}$ are listed in Table 4 and Table 7. Table 5 gives the results on these MSE under different jump sizes Z_n and sample sizes n ($= 500, 1000, 5000$). Table 6 and Table 8 show these values of MSE under different jump arrival intensity λ and sample sizes T ($= 10, 50, 100$) over 500 replicates.

Table 4: MSE ($\times 10^{-3}$) of various estimators for three lengths of time interval (T) and three sample sizes for $\mu(x) = -x$ with arrival intensity $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ over 500 replicates.

Time Span	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$T = 10$	NW	2.5630	1.6007	1.4672
	LL	1.3561	0.7464	0.1899
	RNW	3.9662	1.3093	0.6185
	AS	0.6328	0.1136	0.1121
$T = 50$	NW	2.2359	2.1395	1.5274
	LL	1.3682	0.8460	0.5652
	RNW	2.0140	2.1506	1.1965
	AS	0.4683	0.3137	0.1577
$T = 100$	NW	2.6235	2.1786	1.9483
	LL	1.8212	1.0449	0.8200
	RNW	2.5772	2.2098	1.6114
	AS	0.8070	0.5841	0.5215

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 5: MSE ($\times 10^{-3}$) of various estimators for three types of jump size Z_n and three sample sizes for $\mu(x) = -x$ with arrival intensity $\lambda = 2$ and $T = 10$ over 500 replicates.

Jump Size	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$Z_n \sim \mathcal{N}(0, 0.036^2)$	NW	2.5630	1.6007	1.4672
	LL	1.3561	0.7464	0.1899
	RNW	3.9662	1.3093	0.6185

$Z_n \sim \mathcal{N}(0, 1)$	AS	0.6328	0.1136	0.1121
	NW	11.1193	3.5369	2.3887
	LL	7.2906	3.4295	1.8619
	RNW	12.3186	4.1276	2.0337
$Z_n \sim \text{Cauchy}(0, 1)$	AS	6.8088	1.6764	0.6840
	NW	36.2567	31.1746	7.2090
	LL	13.9267	10.8175	5.4620
	RNW	10.3666	5.2346	8.3132
	AS	5.3919	4.8091	2.8424

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 6: MSE ($\times 10^{-3}$) of various estimators for three types of jump arrival intensity λ and three sample sizes for $\mu(x) = -x$ with $T = 10$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ over 500 replicates.

Jump Intensity	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$\lambda = 1$	NW	2.2185	1.4198	0.9437
	LL	1.0616	0.7367	0.4891
	RNW	3.6126	1.4294	0.5364
	AS	0.4407	0.0437	0.2213
$\lambda = 2$	NW	2.5630	1.6007	1.4672
	LL	1.3561	0.7464	0.1899
	RNW	3.9662	1.3093	0.6185
	AS	0.6328	0.1136	0.1121
$\lambda = 5$	NW	4.4235	4.0884	1.9067
	LL	3.9030	1.8035	0.6593
	RNW	5.0194	1.6432	0.7351
	AS	2.5704	1.6625	0.3750

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 7: MSE ($\times 10^{-5}$) of various estimators for three lengths of time interval (T) and three sample sizes for $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$ with arrival intensity $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ over 500 replicates.

Time Span	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$T = 10$	NW	0.4190	0.2985	0.1958
	LL	0.3682	0.2610	0.1798
	RNW	1.3295	0.2312	0.1792
	AS	0.2608	0.2254	0.1705
$T = 50$	NW	0.4276	0.3754	0.3099
	LL	0.3797	0.3234	0.2579
	RNW	1.1575	0.3654	0.2874
	AS	0.3554	0.2922	0.1997
$T = 100$	NW	0.4544	0.4231	0.3303
	LL	0.4433	0.3803	0.3174
	RNW	1.5675	0.3601	0.2911
	AS	0.4047	0.3644	0.2756

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 8: MSE ($\times 10^{-5}$) of various estimators for three types of jump arrival intensity λ and three sample sizes for $M(x) = 0.01 + 0.01 \times x^2 + \lambda \times 0.036^2$ with $T = 10$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$ over 500 replicates.

Jump Intensity	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$\lambda = 1$	NW	0.1753	0.1348	0.0461
	LL	0.1627	0.1223	0.0323
	RNW	0.2179	0.1037	0.0385
	AS	0.1618	0.0889	0.0134

$\lambda = 2$	NW	0.4190	0.2985	0.1958
	LL	0.3682	0.2610	0.1798
	RNW	1.3295	0.2312	0.1792
	AS	0.2608	0.2254	0.1705
$\lambda = 5$	NW	2.2810	1.7232	0.8455
	LL	1.7250	1.2891	0.5958
	RNW	2.0331	1.0947	0.6711
	AS	1.5373	1.2534	0.5889

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

From Table 4–Table 6 or Table 7–Table 8, we can get the following findings.

- The Nadaraya-Watson estimator for $\mu(x)$ or $M(x)$ constructed with Gamma asymmetric kernels performs a little better than that constructed with Gaussian symmetric kernels in terms of MSE under different time spans, different sample sizes, different jump sizes and different arrival intensity. This is mainly due to the fact that the smaller coverage rate for “interior x ” and the smaller asymptotic variance for “boundary x ”;
- From Table 4 and Table 7, for the same time interval T , as the sample sizes n tends larger, the performances of the estimators constructed with Gamma asymmetric kernels or Gaussian symmetric kernels for $\mu(x)$ or $M(x)$ are improved due to the fact that more information for estimation procedure is sampled as $\Delta_n \rightarrow 0$. However, for the same sample sizes n , as the time interval T expands larger, the performance of the estimators for $\mu(x)$ or $M(x)$ gets worse due to the fact that more jumps happens in larger time interval T in steps 3 of Algorithm 1;
- From Table 5, Table 6 or Table 8, for the same jump size or the same jump arrival intensity, as the sample sizes n tends larger, the performances of the estimators for $\mu(x)$ or $M(x)$ are also improved due to the fact that more jump information for estimation procedure is collected as $\Delta_n \rightarrow 0$. However, for the same sample sizes n , as the amplitude or frequency of jump becomes larger, the MSE of the estimators gradually becomes larger because of larger and more jumps which is more likely to produce outliers;
- To some extent, the previous remark confirms that the drift and volatility functions cannot be identified in a fixed time span, which corresponds to the results of Theorem 2 and Theorem 3.

Figure 5 and Figure 6 give the QQ plots of Nadaraya-Watson estimators for the drift function $\mu(x)$ and conditional variance function $M(x)$ constructed with Gamma asymmetric kernels at boundary and interior points, that is 10%, 50% and 90% quantile points of X_t , with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$. This reveals the normality of the estimators of the drift function $\mu(x)$ and conditional variance function $M(x)$ constructed with Gamma asymmetric kernels, which confirms the results in Theorem 2 and Theorem 3.

Figure 5: QQ plots of Nadaraya-Watson estimator for $\mu(x) = -x$ using Gamma kernels at boundary and interior points, that is 10%, 50% and 90% quantile points of X_t , with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$. (A) QQ plot for left point. (B) QQ plot for median point. (C) QQ plot for right point.

Figure 6: QQ plots of Nadaraya-Watson estimator for $M(x) = 0.01 + 0.01 \times x^2 + 2 \times 0.036^2$ using Gamma kernels at boundary and interior points, that is 10%, 50% and 90% quantile points of X_t , with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, 0.036^2)$.

(A) QQ plot for left point. (B) QQ plot for median point. (C) QQ plot for right point.

Example 2: Mixed Gaussian jumps (The Variance Gamma model).

In this subsection, a repeated experiment is conducted for another model (15) aimed at evaluating the better finite-sampling performance of nonparametric estimators (9) constructed with Gamma asymmetric kernels

$$\begin{cases} dY_t = X_{t-}dt, \\ dX_t = -X_{t-}dt + \sqrt{0.01 + 0.01X_{t-}^2}dW_t + dJ_t, \end{cases} \quad (15)$$

where the jump size in J_t is assumed $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$ and $V \sim G(\nu) = 1/\Gamma(1/b)(1/b)((1/b)\nu)^{1/b-1}e^{-(1/b)\nu}I_{\{\nu \geq 0\}}$. $\Gamma(\cdot)$ is the gamma function, $b = 0.023$ is the variance of V and $\sigma_f = 0.02$ is the scale parameter. One can refer to Madan and Seneta (1990) and Bandi and Nguyen (2003) for more details about this variance Gamma model.

For this case, based on the sample X_t generated by the model (15), compared with nonparametric kernel estimators constructed with Gaussian symmetric kernels, we can also observe the better finite-sample properties such as smaller bias and shorter length of confidence bands, especially at the sparse design boundary point, for Nadaraya-Watson estimators of drift and volatility coefficients using Gamma asymmetric kernel from Figure 7–Figure 11 and Table 9–Table 14, which confirms the bias corrections and variance reduction shown in Theorem 2 and Theorem 3.

Figure 7: Sample Paths of X_t and Y_t for Model (15).
 (A) The differentiated process X_t . (B) The integrated process Y_t .

Figure 8: Various Estimators for unknown coefficients in model (15) with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_j^2 V)$.

(A) Estimators for drift function $\mu(x) = -x$. (B) Estimators for volatility function $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$.

Figure 9: 95% Monte Carlo confidence intervals for unknown functions based on various estimators with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$.

(A) Confidence bands for drift function $\mu(x) = -x$. (B) Confidence bands for volatility function $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$.

Figure 10: QQ plots of Nadaraya-Watson estimator for $\mu(x) = -x$ using Gamma kernels at boundary and interior points, that is 10%, 50% and 90% quantile points of X_t , with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$. (A) QQ plot for left point. (B) QQ plot for median point. (C) QQ plot for right point.

Figure 11: QQ plots of Nadaraya-Watson estimator for $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$ using Gamma kernels at boundary and interior points, that is 10%, 50% and 90% quantile points of X_t , with $\lambda = 2$, $T = 10$, $n = 1000$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$.

(A) QQ plot for median point. (B) QQ plot for left point. (C) QQ plot for right point.

Table 9: The Biases of various Estimators for $\mu(x) = -x$ and $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$ at various points of sample X_t with $T = 10$, $n = 1000$, $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$.

		Bias($\times 10^{-2}$)										Various points of Sample X_t	
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1		
$\mu(x)$	NW	0.2128	0.9130	1.6131	2.3128	3.0119	3.7102	4.4074	5.1033	5.7978	6.4908		
	LL	-0.7333	-0.8928	-1.0771	-1.2929	-1.5465	-1.8444	-2.1929	-2.5979	-3.0654	-3.6009		
	RNW	0.6624	1.4098	1.7277	2.1851	2.4255	2.7918	2.9901	3.9263	5.0019	5.2460		
	AS	-0.0418	-0.1094	-0.0797	0.0203	0.1746	0.3725	0.6072	0.8736	1.1683	1.4886		
$M(x)$	NW	-0.0978	-0.0916	-0.0852	-0.0786	-0.0718	-0.0649	-0.0578	-0.0508	-0.0438	-0.0370		
	LL	-0.0730	-0.0587	-0.0442	-0.0297	-0.0154	-0.0015	0.0118	0.0244	0.0365	0.0481		
	RNW	-0.0832	-0.0739	-0.0631	-0.0576	-0.0534	-0.0503	-0.0449	-0.0362	-0.0254	-0.0154		
	AS	-0.0701	-0.0600	-0.0443	-0.0270	-0.0107	0.0031	0.0138	0.0215	0.0266	0.0294		

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 10: The Length of Confidence Bands of various Estimators for $\mu(x) = -x$ and $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$ at various points of sample X_t with $T = 10$, $n = 1000$, $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$.

Length of Confidence Bands		Various points of Sample X_t									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1

		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$\mu(x)$	NW	0.1142	0.1149	0.1159	0.1174	0.1192	0.1213	0.1238	0.1267	0.1300	0.1336
	LL	0.1142	0.1149	0.1161	0.1177	0.1195	0.1217	0.1241	0.1266	0.1293	0.1321
	RNW	0.1140	0.1140	0.1149	0.1150	0.1203	0.1205	0.1238	0.1253	0.1284	0.1324
	AS	0.1128	0.1105	0.1098	0.1102	0.1116	0.1137	0.1165	0.1199	0.1239	0.1285
$M(x)(\times 10^{-3})$	NW	2.0770	2.1493	2.2411	2.3542	2.4908	2.6534	2.8452	3.0694	3.3304	3.6329
	LL	2.1337	2.2316	2.3534	2.5010	2.6765	2.8824	3.1218	3.3983	3.7166	4.0824
	RNW	2.2759	2.3440	2.3464	2.3263	2.4848	2.5520	2.6678	2.7784	2.8982	2.9283
	AS	3.2112	2.9896	2.8535	2.7643	2.7011	2.6525	2.6119	2.5758	2.5425	2.5109

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 11: MSE ($\times 10^{-3}$) of various estimators for three lengths of time interval (T) and three sample sizes for $\mu(x) = -x$ with arrival intensity $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$ over 500 replicates.

Time Span	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$T = 10$	NW	1.9532	1.7212	1.4911
	LL	0.7893	0.7146	0.2343
	RNW	0.9114	0.1365	0.0952
	AS	0.4311	0.2988	0.0793
$T = 50$	NW	2.0867	2.1055	0.7967
	LL	0.8819	0.9252	0.6314
	RNW	1.3678	1.7768	0.3483
	AS	0.6815	0.5261	0.3247
$T = 100$	NW	2.1932	2.1095	1.8885
	LL	1.0132	0.9874	0.7237
	RNW	1.9735	1.2976	1.4325
	AS	0.8697	0.6206	0.4619

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 12: MSE ($\times 10^{-3}$) of various estimators for three types of jump arrival intensity λ and three sample sizes for $\mu(x) = -x$ with $T = 10$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$ over 500 replicates.

Jump Intensity	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$\lambda = 1$	NW	2.7592	1.4668	1.6902
	LL	1.7704	0.6505	0.4611
	RNW	1.7553	1.0336	1.4220
	AS	0.9762	0.0823	0.0451
$\lambda = 2$	NW	1.9532	1.7212	1.4911
	LL	0.7893	0.7146	0.2343
	RNW	0.9114	0.1365	0.0952
	AS	0.4311	0.2988	0.0793
$\lambda = 5$	NW	4.5659	3.5692	2.7208
	LL	2.8368	1.0792	0.6258
	RNW	4.3618	2.7649	1.5333
	AS	2.0147	0.5393	0.3959

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 13: MSE ($\times 10^{-6}$) of various estimators for three lengths of time interval (T) and three sample sizes for $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$ with arrival intensity $\lambda = 2$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$ over 500 replicates.

Time Span	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$T = 10$	NW	1.4521	0.4921	0.4243
	LL	1.2956	0.4034	0.3161
	RNW	1.4099	0.3227	0.1765

$T = 50$	AS	0.6517	0.2513	0.2352
	NW	2.1177	0.7930	0.4396
	LL	1.5550	0.7269	0.3591
	RNW	1.4316	0.5831	0.5036
$T = 100$	AS	1.1266	0.5464	0.2685
	NW	3.1073	2.2689	0.5252
	LL	3.0142	1.7343	0.4222
	RNW	3.5333	1.6595	0.4016
	AS	2.3565	1.4350	0.3709

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

Table 14: MSE ($\times 10^{-6}$) of various estimators for three types of jump arrival intensity λ and three sample sizes for $M(x) = 0.01 + 0.01 \times x^2 + 0.02^2$ with $T = 10$ and jump size $Z_n \sim \mathcal{N}(0, \sigma_f^2 V)$ over 500 replicates.

Jump Intensity	Estimators	$n = 500$	$n = 1000$	$n = 5000$
$\lambda = 1$	NW	0.4717	0.2311	0.1088
	LL	0.3043	0.1507	0.0887
	RNW	0.3186	0.2322	0.0536
	AS	0.2488	0.1007	0.0126
$\lambda = 2$	NW	1.4521	0.4921	0.4243
	LL	1.2956	0.4034	0.3161
	RNW	1.4099	0.3227	0.1765
	AS	0.6517	0.2513	0.2352
$\lambda = 5$	NW	2.6312	1.7956	0.5335
	LL	2.3927	1.5227	0.4344
	RNW	2.1225	1.1585	0.4085
	AS	2.0745	0.8151	0.4297

Note that NW denotes Nadaraya-Watson estimator, LL denotes local linear estimator, RNW denotes reweighted Nadaraya-Watson estimator and AS denotes the estimator (9) proposed in this paper.

5 Empirical analysis

In this section, we apply the second-order jump-diffusion to model the return of stock index between Jul 2014 and Dec 2014 from Shenzhen Stock Exchange in China under 5-minute high-frequency data, and then apply the Nadaraya-Watson estimators to estimate the unknown coefficients in model (2) based on Gamma asymmetric kernels and Gaussian symmetric kernels. One can refer to Supplementary Material for the empirical data and codes.

We assume that

$$\begin{cases} d \log Y_t = X_t dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathcal{G}} c(X_{t-}, z)r(\omega, dt, dz), \end{cases} \quad (16)$$

where $\log Y_t$ is the log integrated process for stock index or commodity price and X_t is the latent process for the log-returns. According to (4), we can get the proxy of the latent process

$$\tilde{X}_{i\Delta_n} = \frac{\log Y_{i\Delta_n} - \log Y_{(i-1)\Delta_n}}{\Delta_n}. \quad (17)$$

The plots of the stock index and its proxy (17) of Shenzhen composite index in 5-minute high frequency data are shown in Figure 12.

Figure 12: Time Series and Proxy of Shenzhen Composite Index (2014) from July 03, 2014 to Dec 31, 2014. (A) Shenzhen Composite Index (2014). (B) Proxy of Shenzhen Composite Index (2014).

First, based on the Augmented Dickey-Fuller test statistic, we can easily get that the null hypothesis of non-stationarity is accepted at the 5% significance level for the stock index Y_t , but is rejected for the proxy of X_t , which confirms the assumption of stationarity by differencing. Then, we test the existence of jumps for the proxy X_t through the test statistic proposed in Barndorff-Nielsen and Shephard (2006) (denoted by BS Statistic). For the 5-minute high frequency data between Jul 2014 and Dec 2014 from Shenzhen Stock Exchange, the value of BS Statistic is -4.5961 , which exceeds $[-1.96, 1.96]$, so there exists jumps in high frequency data at the 5% significance level, which confirms the validity of model (2) for the return of stock index.

Finally, we will construct 95% normal confidence intervals for the unknown coefficients based on Gamma asymmetric kernels and Gaussian symmetric kernels under (17) and $\Delta_n = \frac{1}{48}$ for 5-minute data ($t = 1$ meaning 1 day). The 95% normal confidence bands for the drift and volatility functions are demonstrated in Figure 13. All the quantities are computed at 120 equally spaced nonnegative ordered $\tilde{X}_{i\Delta_n}$ from 0.01 to 0.601. For a more intuitive comparison of the lengths of normal confidence intervals of the drift and volatility coefficients based on Gaussian kernels and Gamma kernels for Shenzhen Composite Index (2014), here the ratios of the length of confidence bands constructed with Gaussian symmetric kernel (CB-GSK) to that constructed with Gamma asymmetric kernel (CB-GAK) are shown in Figure 14. Note that the blue dotted lines in Figure 8 represent the ratio value of one.

Figure 13: 95% normal confidence intervals of the drift and volatility coefficients for Shenzhen Composite Index (2014) based on Gamma kernels and Gaussian kernels.

Note that Nadaraya-Watson is denoted as NW in this figure. (A) Drift Confidence Band. (B) Volatility Confidence Band.

Figure 14: The ratios between the lengths of normal confidence intervals of the drift and volatility coefficients based on Gaussian kernels and Gamma kernels for Shenzhen Composite Index (2014).

(A) Length Ratios for Drift Estimators. (B) Length Ratios for Volatility Estimators.

From Figure 13 and Figure 14, we can observe the following findings.

- As for the quantities close to zero, the ratios are less than one, which coincides with the discussion in Remark 9 that the closer to the boundary point or the larger bandwidth for fixed “boundary x ”, the shorter the length of confidence interval based on Gaussian symmetric kernel.
- As the points increase, especially at the sparse points, CB-GSK tends to be longer than CB-GAK and the ratios gradually become larger and greater than 1 (excluding a small amount of points around 0.5 for volatility estimator), which effectively verifies the properties of efficiency gains and resistance to sparse points of Nadaraya-Watson smoothing using Gamma asymmetric kernel through the real high frequency financial data.

6 Conclusion

In this paper, Nadaraya-Watson estimators constructed with Gamma asymmetric kernels for the unknown coefficients in second-order jump-diffusion model are investigated. Compared with estimation constructed with Gaussian symmetric kernels, the local constant smoothing using Gamma asymmetric kernels possesses some attractive advantages such as boundary free bias, variance reduction and resistance to sparse design points, which is validated through Monte Carlo simulation study. Theoretically, under mild regularity conditions, the estimators constructed with Gamma asymmetric kernels possess the consistency and asymptotic normality for large sample and the advantages such as bias reduction and shorter length of confidence bands through simulation experiments for finite sample.

Empirically, the estimators are illustrated through stock composite index from Shenzhen Stock Exchange in China under 5-minute high-frequency data and possess some advantages mentioned above. This means the second-order jump-diffusion model may be an alternative model to describe the dynamic of some financial data, especially to explain integrated economic phenomena, that is the current observation in empirical finance usually behaves as the cumulation of all past perturbations.

Acknowledgments

The authors would like to thank the editor and two anonymous referees for their valuable suggestions, which greatly improved our paper. This research work is supported by Ministry of Education, Humanities and Social Sciences Project (No. 18YJCZH153), National Statistical Science Research Project (No. 2018LZ05), the General Research Fund of Shanghai Normal University (No. SK201720), Funding Programs for Youth Teachers of Shanghai Colleges and Universities (No. A-9103-18-104001), Key Subject of Quantitative Economics and Academic Innovation Team Construction Project in Shanghai Normal University and Shanghai Training Program of Innovation and Entrepreneurship for Undergraduates (NO. 201710270149).

A Proofs

A.1 Procedure for Assumption 3

Notice that the expectation with respect to the distribution of $\xi_{n,i}$ depends on the stationary densities of $X_{n,i}$ and $\tilde{X}_{n,i}$ because $\xi_{n,i}$ is a convex linear combination of $X_{n,i}$ and $\tilde{X}_{n,i}$.

For the case (i): $E[|h_n K'(X_{(i-1)\Delta_n})|] < \infty$. For $K_{G(x/h_n+1, h_n)}(u)$, its first-order derivative has the form of $K'_{G(x/h_n+1, h_n)}(u) = \left(\frac{x}{h_n}\right) \frac{u^{x/h_n-1} \exp(-u/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)} - \left(\frac{1}{h_n}\right) \frac{u^{x/h_n} \exp(-u/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)} := \frac{1}{h_n} K_1(u) + \frac{1}{h_n} K_2(u)$. Then using the well-known properties of the Γ function, the mean of Gamma distribution and the derivative of the function $gp(x) := g(x) \cdot p(x)$ for stationary process X_t , we have

$$\begin{aligned}
& E[|h_n K'(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] \\
&= E[|K_1(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] + E[|K_2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] \\
&= x \frac{h_n^{x/h_n} \Gamma(x/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)} \int_0^\infty \frac{y^{x/h_n-1} \exp(-y/h_n)}{h_n^{x/h_n} \Gamma(x/h_n)} gp(y) dy \\
&\quad + \int_0^\infty \frac{y^{x/h_n} \exp(-y/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)} gp(y) dy \\
&= \int_0^\infty \frac{y^{x/h_n-1} \exp(-y/h_n)}{h_n^{x/h_n} \Gamma(x/h_n)} gp(y) dy + \int_0^\infty \frac{y^{x/h_n} \exp(-y/h_n)}{h_n^{x/h_n+1} \Gamma(x/h_n+1)} gp(y) dy \\
&= E[gp(\xi_1)] + E[gp(\xi_2)] \\
&= E[gp(E(\xi_1) + \xi_1 - E(\xi_1))] + E[gp(E(\xi_2) + \xi_2 - E(\xi_2))] \\
&= 2gp(x) + O(h_n) < \infty,
\end{aligned}$$

where $\xi_1 \stackrel{\mathcal{D}}{=} G(x/h_n, h_n)$, $\xi_2 \stackrel{\mathcal{D}}{=} G(x/h_n + 1, h_n)$ and G denotes the Gamma distribution.

For the case (ii):

$$E[|h_n^2 K'^2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] \leq 2E[|K_1^2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] + 2E[|K_2^2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|].$$

Now we only deal with the first part (the second part can be dealt with in the similar way). Note that

$K_1(u) = \frac{u^{x/h_n-1} \exp(-u/h_n)}{h_n^{x/h_n} \Gamma(x/h_n)}$ can be considered as a density function for a random variable $\xi_1 \stackrel{\mathcal{D}}{=} G(x/h_n, h_n)$. By the property of the Γ function, we have with $\eta_x \stackrel{\mathcal{D}}{=} G(2x/h_n - 1, h_n)$,

$$\begin{aligned}
& E[|K_1^2(X_{(i-1)\Delta_n})g(X_{(i-1)\Delta_n})|] \\
&= B_{h_n}(x) E[gp(\eta_x)] \\
&= B_{h_n}(x) E[gp(E(\xi_1) + \xi_1 - E(\xi_1))] = B_{h_n}(x) gp(x) \\
&\approx gp(x) \begin{cases} \frac{1}{2\sqrt{\pi}} h_n^{-1/2} x^{-1/2} & \text{if } x/h_n \rightarrow \infty \text{ ("interior } x"); \\ h_n^{-1} \frac{\Gamma(2\kappa-1)}{2^{2\kappa-1} \Gamma^2(\kappa)} & \text{if } x/h_n \rightarrow \kappa \text{ ("boundary } x"), \end{cases}
\end{aligned}$$

where $B_{h_n}(x) = \frac{h_n^{-1} \Gamma(2x/h_n - 1)}{2^{2x/h_n-1} \Gamma^2(x/h_n + 1)}$ and the last equation follows from (Chen (2000), P474). hence, the results of $\lim_{h_n \rightarrow 0} h_n^{1/2} E[|h_n^2 K'^2(\xi_{n,i})g(\xi_{n,i})|] < \infty$ for "interior x " and $\lim_{h_n \rightarrow 0} h_n E[|h_n^2 K'^2(\xi_{n,i})g(\xi_{n,i})|] < \infty$ for "boundary x " hold.

A.2 Some technical lemmas with proofs

We lay out some notations. For $x = (x_1, \dots, x_d)$, $\partial_{x_j} := \frac{\partial}{\partial x_j}$, $\partial_{x_j}^2 := \frac{\partial^2}{\partial x_j^2}$, $\partial_{x_i x_j}^2 := \frac{\partial^2}{\partial x_i \partial x_j}$, $\partial x := (\partial_{x_1}, \dots, \partial_{x_d})^*$, and $\partial_x^2 = (\partial_{x_i x_j}^2)_{1 \leq i, j \leq d}$, where $*$ stands for the transpose.

Lemma 1

(Shimizu and Yoshida, 2006) Let Z be a d -dimensional solution-process to the stochastic differential equation

$$Z_t = Z_0 + \int_0^t \mu(Z_{s-}) ds + \int_0^t \sigma(Z_{s-}) dW_s + \int_0^t \int_{\mathcal{E}} c(Z_{s-}, z) r(\omega, dt, dz),$$

where Z_0 is a random variable, $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, $\mu(x)$, $c(x, z)$ are d -dimensional vectors defined on $\mathbb{R}^d, \mathbb{R}^d \times \mathcal{E}$ respectively, $\sigma(x)$ is a $d \times d$ diagonal matrix defined on $\mathbb{R}^d$, and W_t is a d -dimensional vector of independent Brownian motions.

Let g be a $C^{2(l+1)}$ -class function whose derivatives up to $2(l+1)$ th are of polynomial growth. Assume that the coefficients $\mu(x)$, $\sigma(x)$, and $c(x, z)$ are C^{2l} -class function whose derivatives with respect to x up to $2l$ th are of polynomial growth. Under Assumption 5, the following expansion holds

$$E[g(Z_t)|\mathcal{F}_s] = \sum_{j=0}^l L^j g(Z_s) \frac{\Delta_n^j}{j!} + R, \quad (18)$$

for $t > s$ and $\Delta_n = t - s$, where $R = \int_0^{\Delta_n} \int_0^{u_1} \dots \int_0^{u_l} E[L^{l+1}g(Z_{s+u_{l+1}})|\mathcal{F}_s] du_1 \dots du_{l+1}$ is a stochastic function of order Δ_n^{l+1} , $Lg(x) = \partial_x^* g(x)\mu(x) + \frac{1}{2} \text{tr}[\partial_x^2 g(x)\sigma(x)\sigma^*(x)] + \int_{\mathcal{G}} \{g(x+c(x,z)) - g(x) - \partial_x^* g(x)c(x,z)\} f(z) dz$.

Remark 11

Consider a particularly important model:

$$\begin{cases} dY_t = X_{t-} dt, \\ dX_t = \mu(X_{t-})dt + \sigma(X_{t-})dW_t + \int_{\mathcal{G}} c(X_{t-}, z)r(w, dt, dz). \end{cases}$$

As $d = 2$, we have

$$\begin{aligned} Lg(x, y) &= x(\partial g / \partial y) + \mu(x)(\partial g / \partial x) + \frac{1}{2}\sigma^2(x)(\partial^2 g / \partial x^2) \\ &\quad + \int_{\mathcal{G}} \{g(x+c(x, z), y) - g(x, y) - \frac{\partial g}{\partial x} \cdot c(x, z)\} f(z) dz. \end{aligned} \quad (19)$$

Based on the second-order infinitesimal operator (19), we can calculate many mathematical expectations involving $\tilde{X}_{i\Delta_n}$, for instance (5) and (6) which provide the basis for estimators (7) and (8).

Lemma 2

Under Assumption 1, Assumption 2 and Assumption 5, let

$$\hat{p}_n^0(x) = \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}), \quad \hat{a}_n^0(x) = \frac{A_n^0(x)}{\hat{p}_n^0(x)}, \quad \hat{b}_n^0(x) = \frac{B_n^0(x)}{\hat{p}_n^0(x)},$$

where

$$\begin{aligned} A_n^0(x) &= \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}) \frac{X_{i\Delta_n} - X_{(i-1)\Delta_n}}{\Delta_n}, \\ B_n^0(x) &= \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}) \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}. \end{aligned}$$

We have

$$\hat{p}_n^0(x) \xrightarrow{P} p(x), \quad \hat{a}_n^0(x) \xrightarrow{P} \mu(x), \quad \hat{b}_n^0(x) \xrightarrow{P} \sigma^2(x) + \int_{\mathcal{G}} c^2(x, z)f(z)dz,$$

and for “interior x ”, if $h_n = O((n\Delta_n)^{-2/5})$, then

$$\begin{aligned} \sqrt{n\Delta_n h_n^{1/2}}(\hat{a}_n^0(x) - \mu(x) - h_n B_{\hat{a}_n^0(x)}) &\xrightarrow{d} N\left(0, \frac{M(x)}{2\sqrt{\pi}x^{1/2}p(x)}\right), \\ \sqrt{n\Delta_n h_n^{1/2}}(\hat{b}_n^0(x) - M(x) - h_n B_{\hat{b}_n^0(x)}) &\xrightarrow{d} N\left(0, \frac{\int_{\mathcal{G}} c^4(x, z)f(z)dz}{2\sqrt{\pi}x^{1/2}p(x)}\right), \end{aligned}$$

for “boundary x ”, if $h_n = O((n\Delta_n)^{-1/3})$, then

$$\begin{aligned} \sqrt{n\Delta_n h_n}(\hat{a}_n^0(x) - \mu(x) - h_n B_{\hat{a}_n^0(x)}) &\xrightarrow{d} N\left(0, \frac{M(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)p(x)}\right), \\ \sqrt{n\Delta_n h_n}(\hat{b}_n^0(x) - M(x) - h_n B_{\hat{b}_n^0(x)}) &\xrightarrow{d} N\left(0, \frac{\int_{\mathcal{G}} c^4(x, z)f(z)dz\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)p(x)}\right), \end{aligned}$$

where $B_{\hat{a}_n^0(x)}, B_{\hat{b}_n^0(x)}$ denotes the bias of the estimators of $\hat{a}_n^0(x), \hat{b}_n^0(x)$, respectively, that is

$$\begin{aligned} B_{\hat{a}_n^0(x)} &= \mu'(x)\left\{1 + \frac{xp'(x)}{p(x)}\right\} + \frac{x}{2}\mu''(x), \\ B_{\hat{b}_n^0(x)} &= M'(x)\left\{1 + \frac{xp'(x)}{p(x)}\right\} + \frac{x}{2}M''(x), \end{aligned}$$

Remark 12

The main method to obtain the asymptotic properties for estimators of model (2) is to approximate the estimators for (2) by the similar estimator for model (1) in probability such as (21) and (22). Based on the basic idea, we should first know the asymptotic properties for the Nadaraya-Watson estimator for (1) before the Theorem 2 and Theorem 3 presented. Lemma 2 give us the desired properties for Nadaraya-Watson estimator of (1).

Proof.

Corollary 2 in Hanif (2016) is the stationary case for Theorem 2 in Hanif (2016). Here we only mention the procedure for modified proof to Theorem 2 on the asymptotic normality in Hanif (2016). For convenience, we still use the same notations as that in the proof of Theorem 2 in Hanif (2016).

The bias expression $h_n \left\{ M'(x) \left\{ 1 + \frac{xp'(x)}{p(x)} \right\} + \frac{x}{2} M''(x) \right\} + o(h_n)$ can be similarly deduced as equation (B10) in Gospodinov and Hirukawa (2012) or equation (4.18) in Hanif (2016). Similarly as equations (159–174) in Bandi and Nguyen (2003), the dominant part for the variance of the asymptotic normality is D_{44} due to the relationship $C_{44} = O_P(\sqrt{\Delta_{n,T}} D_{44}) = o_p(D_{44})$. We write as Hanif (2016)

$$\begin{aligned} D_{44} &= \frac{\sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \int_Y ((X_{s-} + c)^2 - X_{s-}^2 - 2X_{i\Delta_{n,T}} c(X_{s-}, y)) \bar{v}(ds, dy)}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \\ &= \frac{\sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \int_Y ((X_{s-} + c)^2 - X_{s-}^2) \bar{v}(ds, dy)}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \\ &\quad - \frac{\sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}}) \int_{i\Delta_{n,T}}^{(i+1)\Delta_{n,T}} \int_Y 2X_{i\Delta_{n,T}} c(X_{s-}, y) \bar{v}(ds, dy)}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \\ &:= a_{n,T} + b_{n,T}. \end{aligned}$$

By the Lemma 3 in Hanif (2016), we have

$$\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}}) \xrightarrow{a.s.} \bar{L}_X(T, x).$$

According to the similar procedures as equations (130), (169) and (172) in Bandi and Nguyen (2003), we can obtain

$$\sqrt{\bar{L}_X(T, x)} \left(\frac{a_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \right) \xrightarrow{\mathcal{D}} N\left(0, A_{h_n}(x) \int_{\mathcal{E}} ((x + c(x, z))^2 - x^2) f(z) dz\right)$$

and

$$\sqrt{\bar{L}_X(T, x)} \left(\frac{b_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \right) \xrightarrow{\mathcal{D}} N\left(0, A_{h_n}(x) \int_{\mathcal{E}} 4c^2(x, z) x^2 f(z) dz\right),$$

where from Chen (2000)

$$A_{h_n}(x) = \begin{cases} \frac{h_n^{-1/2}}{2\sqrt{\pi}x^{1/2}} + o(h_n^{-1/2}) & \text{if } x/h_n \rightarrow \infty : \text{"interior } x"; \\ \frac{h_n^{-1}\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o(h_n^{-1}) & \text{if } x/h_n \rightarrow \kappa : \text{"boundary } x". \end{cases}$$

Finally, using the similar procedure as equation (173) in Bandi and Nguyen (2003), it can be shown that the limiting covariance between

$$\sqrt{\bar{L}_X(T, x)} \left(\frac{a_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \right) \text{ and } \sqrt{\bar{L}_X(T, x)} \left(\frac{b_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_{n,T}})} \right) \text{ is characterized as}$$

$$\begin{aligned} & \text{Asicov} \left(\frac{\sqrt{\bar{L}_X(T, x)} a_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_n, T})}, \frac{\sqrt{\bar{L}_X(T, x)} b_{n,T}^{Num}}{\Delta_{n,T} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{i\Delta_n, T})} \right) \\ & \xrightarrow{a.s.} -2A_{h_n}(x) \int_{\mathcal{E}} [xc(x, z)((x + c(x, z))^2 - x^2)] f(z) dz \end{aligned}$$

Hence,

$$\sqrt{\bar{L}_X(T, x)} D_{44} \xrightarrow{\mathcal{D}} N\left(0, A_{h_n}(x) \int_{\mathcal{E}} c^4(x, z) f(z) dz\right),$$

that is,

$$\begin{cases} \sqrt{\bar{L}_X(T, x) h_n^{1/2}} D_{44} \xrightarrow{\mathcal{D}} N\left(0, \frac{\int_{\mathcal{E}} c^4(x, z) f(z) dz}{2\sqrt{\pi} x^{1/2}}\right) & \text{if } x/h_n \rightarrow \infty : \text{"interior } x\text{"}; \\ \sqrt{\bar{L}_X(T, x) h_n} D_{44} \xrightarrow{\mathcal{D}} N\left(0, \frac{\Gamma(2\kappa + 1) \int_{\mathcal{E}} c^4(x, z) f(z) dz}{2^{2\kappa+1} \Gamma^2(\kappa + 1)}\right) & \text{if } x/h_n \rightarrow \kappa : \text{"boundary } x\text{"} \end{cases}$$

due to the form of $A_{h_n}(x)$.

The equation (4.19) in Hanif (2016) shows that

$$\begin{cases} \sqrt{\frac{\bar{L}_X(T, x) h_n^{1/2}}{\Delta_{n,T}}} C_{44} \xrightarrow{\mathcal{D}} N\left(0, \frac{4\sigma^4(x)}{2\sqrt{\pi} x^{1/2}}\right) & \text{if } x/h_n \rightarrow \infty : \text{"interior } x\text{"}; \\ \sqrt{\frac{\bar{L}_X(T, x) h_n}{\Delta_{n,T}}} C_{44} \xrightarrow{\mathcal{D}} N\left(0, \frac{\Gamma(2\kappa + 1) 4\sigma^4(x)}{2^{2\kappa+1} \Gamma^2(\kappa + 1)}\right) & \text{if } x/h_n \rightarrow \kappa : \text{"boundary } x\text{"} \end{cases}$$

from which we can deduce $C_{44} = O_P(\sqrt{\Delta_{n,T}} D_{44})$. $\square$

Lemma 3

Assumption 1–Assumption 5 lead to the following results,

- i. $\frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) - \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}) \xrightarrow{p} 0,$
- ii.

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} \\ & - \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}) \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} \xrightarrow{p} 0, \end{aligned}$$

- iii.

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(\tilde{X}_{(i-1)\Delta_n}) \frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} \\ & - \frac{1}{n} \sum_{i=1}^n K_{G(x/h_n+1, h_n)}(X_{(i-1)\Delta_n}) \frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} \xrightarrow{p} 0, \end{aligned}$$

Remark 13

To some extend, this lemma can simplify the proofs of Theorem 2 and Theorem 3 such as $\delta_{1,n}$ and $\delta_{2,n}$. Moreover, we can obtain the limit in probability of minuend in (ii) or (iii) of Lemma 3 by calculating the limit in probability of subtractor in (ii) or (iii) (which is easier to obtain). One can refer to Nicolau (2007) for the same idea.

Proof.

$$\text{Let } \varepsilon_{1,n} = \hat{p}_n(x) - \hat{p}_n^0(x) = \frac{1}{n} \sum_{i=1}^n K_{G(x/h_{n+1}, h_n)}(\tilde{X}_{(i-1)\Delta_n}) - \frac{1}{n} \sum_{i=1}^n K_{G(x/h_{n+1}, h_n)}(X_{(i-1)\Delta_n}).$$

$$\begin{aligned} & \max_{1 \leq i \leq n} |\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}| \\ & \leq \max_{1 \leq i \leq n} \frac{1}{\Delta_n} \left| \int_{(i-2)\Delta_n}^{(i-1)\Delta_n} (X_{s-} - X_{(i-1)\Delta_n}) ds \right| \\ & \leq \max_{1 \leq i \leq n} \sup_{(i-2)\Delta_n \leq s \leq (i-1)\Delta_n} |X_{s-} - X_{(i-1)\Delta_n}| \\ & = O_{a.s.}(\sqrt{\Delta_n \log(1/\Delta_n)}), \end{aligned} \quad (20)$$

the last asymptotic equation for the order of magnitude, one can refer to Bandi and Nguyen (2003) (Equations (94) and (95) on Pages 315–316).

By the mean-value theorem, stationarity, Assumption 3, Assumption 5 and (20), we obtain

$$\begin{aligned} E[|\varepsilon_{1,n}|] & \leq E \left[\frac{1}{n} \sum_{i=1}^n |K'(\xi_{n,i})(\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})| \right] \\ & = E[|K'(\xi_{n,2})(\tilde{X}_{\Delta_n} - X_{\Delta_n})|] \\ & \leq \sqrt{\Delta_n \log(1/\Delta_n)} E[|K'(\xi_{n,2})|] \\ & = \frac{\sqrt{\Delta_n \log(1/\Delta_n)}}{h_n} E[|h_n K'(\xi_{n,2})|] \rightarrow 0, \end{aligned}$$

where $\xi_{n,2} = \theta X_{\Delta_n} + (1-\theta)\tilde{X}_{\Delta_n}$, $0 \leq \theta \leq 1$.

Lemma 3 (i) follows from Chebyshev's inequality.

Now we prove (ii), write

$$\begin{aligned} \varepsilon_{2,n} & = A_n(x) - A_n^0(x) \\ & = \frac{1}{n} \sum_{i=1}^n K(\tilde{X}_{(i-1)\Delta_n}) \frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})}{\Delta_n} - \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})}{\Delta_n}. \end{aligned}$$

We have

$$\begin{aligned} E[\varepsilon_{2,n}] & = E \left[\frac{1}{n} \sum_{i=1}^n (K(\tilde{X}_{(i-1)\Delta_n}) - K(X_{(i-1)\Delta_n})) E \left[\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} \middle| \mathcal{F}_{(i-1)\Delta_n} \right] \right] \\ & = E \left[\frac{1}{n} \sum_{i=1}^n (K(\tilde{X}_{(i-1)\Delta_n}) - K(X_{(i-1)\Delta_n})) (\mu(X_{(i-1)\Delta_n}) + O_p(\Delta_n)) \right] \\ & = E \left[K'(\xi_{n,2})(\tilde{X}_{\Delta_n} - X_{\Delta_n}) (\mu(X_{\Delta_n}) + O_p(\Delta_n)) \right] \end{aligned}$$

by (5), the mean-value theorem and stationarity. Hence

$$\begin{aligned} |E[\varepsilon_{2,n}]| & \leq E \left[h_n |K'(\xi_{n,2})| \frac{|\tilde{X}_{\Delta_n} - X_{\Delta_n}|}{h_n} |\mu(X_{\Delta_n}) + O_p(\Delta_n)| \right] \\ & \leq \frac{\sqrt{\Delta_n \log(1/\Delta_n)}}{h_n} E[h_n |K'(\xi_{n,2})| |\mu(X_{\Delta_n}) + O_p(\Delta_n)|] \rightarrow 0, \end{aligned}$$

by Assumption 3, Assumption 5 and (20). So $E[\varepsilon_{2,n}] \rightarrow 0$.

Moreover

$$\begin{aligned} \text{Var}[\varepsilon_{2,n}] & = \frac{1}{n\Delta_n h_n} \text{Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{h_n} K'(\xi_{n,i})(\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}) \sqrt{\Delta_n} \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} \right] \\ & =: \frac{1}{n\Delta_n h_n} \text{Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n f_i \right], \end{aligned}$$

where $f_i := \sqrt{h_n} K'(\xi_{n,i}) (\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n}) \sqrt{\Delta_n} \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n}$.

We find $\text{Var}[\frac{1}{\sqrt{n}} \sum_{i=1}^n f_i] = \frac{1}{n} \sum_{i=1}^n \text{Var}[f_i] + \varepsilon_n$ where ε_n represents the sum of $\frac{2}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n$ terms involving the autocorrelations.

Firstly, we prove that $E[f_i^2] < \infty$. Now we calculate $E[f_i^2]$:

$$\begin{aligned} E[f_i^2] &= E\left[h_n K'^2(\xi_{n,i}) (\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})^2 E\left[\frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} \middle| \mathcal{F}_{(i-1)\Delta_n}\right]\right] \\ &= E\left[h_n K'^2(\xi_{n,i}) (\tilde{X}_{(i-1)\Delta_n} - X_{(i-1)\Delta_n})^2 \left(\frac{2}{3} \sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3} \int_{\mathcal{G}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n)\right)\right] \\ &\leq \frac{\Delta_n \log(1/\Delta_n)}{h_n} E\left[h_n^2 K'^2(\xi_{n,i}) \left(\frac{2}{3} \sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3} \int_{\mathcal{G}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n)\right)\right] \\ &= \frac{\Delta_n \log(1/\Delta_n)}{h_n} \left\{ \begin{array}{l} \frac{1}{h_n^{1/2}} \cdot h_n^{1/2} E\left[h_n^2 K'^2(\xi_{n,i}) \left(\frac{2}{3} \sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3} \int_{\mathcal{G}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n)\right)\right] \\ \text{(if } x/h_n \rightarrow \infty : \text{"interior } x\text{"}) \\ \frac{1}{h_n} \cdot h_n E\left[h_n^2 K'^2(\xi_{n,i}) \left(\frac{2}{3} \sigma^2(X_{(i-1)\Delta_n}) + \frac{2}{3} \int_{\mathcal{G}} c^2(X_{(i-1)\Delta_n}, z) f(z) dz + O_p(\Delta_n)\right)\right] \\ \text{(if } x/h_n \rightarrow \kappa : \text{"boundary } x\text{"}) \end{array} \right\} \\ &\approx C \begin{cases} \frac{\Delta_n \log(1/\Delta_n)}{h_n^{3/2}} & \text{if } x/h_n \rightarrow \infty (\text{"interior } x\text{"}); \\ \frac{\Delta_n \log(1/\Delta_n)}{h_n^2} & \text{if } x/h_n \rightarrow \kappa (\text{"boundary } x\text{"}) \end{cases} \\ &< \infty \end{aligned}$$

by (6) and Assumption 3 and Assumption 5.

Then, we prove that $\varepsilon_n = O(\frac{1}{\Delta_n^\alpha})$.

Similarly as that pointed out in Ditlevsen and Sørensen (2004), we know that $\{X_{i\Delta_n}, i = 1, 2, \dots\}$ and $\{\tilde{X}_{i\Delta_n}, i = 1, 2, \dots\}$ are stationary and ρ -mixing with the same size. Moreover, measurable functions of ρ -mixing processes are ρ -mixing with the same size in the space of square integrable functions according to the definition of ρ -mixing.

We have already proved $E[f_i^2] < \infty$ above, so f_i is stationary under Assumption 2 and ρ -mixing with the same size as $\{X_{i\Delta_n}, i = 1, 2, \dots\}$ and $\{\tilde{X}_{i\Delta_n}, i = 1, 2, \dots\}$. Similar argument one can refer to Nicolau (2007). From Lemma 10.1.c with $p = q = 2$ on page 132 in Lin and Bai (2010), we have

$$\begin{aligned} |\varepsilon_n| &= \frac{2}{n} \left| \sum_{j=1}^{n-1} \sum_{i=j+1}^n (Ef_i f_j - Ef_i Ef_j) \right| \\ &\leq \frac{2}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n |Ef_i f_j - Ef_i Ef_j| \\ &\leq \frac{8}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n \rho((i-j)\Delta_n) (Ef_i^2)^{\frac{1}{2}} (Ef_j^2)^{\frac{1}{2}} \\ &= \frac{8}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n \rho((i-j)\Delta_n) Ef_2^2 \end{aligned}$$

Last equation is due to the stationarity of f_i .

We have proved $E[f_2^2] < \infty$ above. Moreover, we have $\sum_{i=j+1}^n \rho((i-j)\Delta_n) = O(\frac{1}{\Delta_n^\alpha})$ under Assumption 2. Hence $|\varepsilon_n| \leq \frac{8}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n \rho((i-j)\Delta_n) Ef_2^2 = O(\frac{1}{\Delta_n^\alpha})$. So the series $\varepsilon_n = O(\frac{1}{\Delta_n^\alpha})$ under Assumption 2.

In conclusion,

$$\text{Var}[\varepsilon_{2,n}] = \frac{1}{nh_n \Delta_n} \text{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n f_i\right] = O\left(\frac{1}{nh_n \Delta_n^{1+\alpha}}\right) \rightarrow 0 \text{ as } n \rightarrow \infty$$

by Assumption 5. So Lemma 3 (ii) can be deduced from the above considerations.

Finally, write

$$\begin{aligned}\varepsilon_{3,n} &= B_n(x) - B_n^0(x) \\ &= \frac{1}{n} \sum_{i=1}^n K(\tilde{X}_{(i-1)\Delta_n}) \frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2}{\Delta_n}.\end{aligned}$$

Lemma 3 (iii) now follows:

$$\begin{aligned}E[|\varepsilon_{3,n}|] &\leq E\left[\frac{1}{n} \sum_{i=1}^n |K(\tilde{X}_{(i-1)\Delta_n}) - K(X_{(i-1)\Delta_n})| E\left[\frac{(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} \middle| \mathcal{F}_{(i-1)\Delta_n}\right]\right] \\ &\leq CE\left[K'(\xi_{n,2})|\tilde{X}_{\Delta_n} - X_{\Delta_n}|(\sigma^2(X_{\Delta_n}) + \int_{\mathcal{E}} c^2(X_{\Delta_n}, z)f(z)dz + O_p(\Delta_n))\right] \\ &\leq \frac{\sqrt{\Delta_n \log(1/\Delta_n)}}{h_n} E\left[h_n K'(\xi_{n,2})(\sigma^2(X_{\Delta_n}) + \int_{\mathcal{E}} c^2(X_{\Delta_n}, z)f(z)dz + O_p(\Delta_n))\right] \\ &\rightarrow 0\end{aligned}$$

by (6) and Assumption 3 and Assumption 5, which implies Lemma 3 (iii). $\square$

A.3 The proof of Theorem 1

Proof.

One can easily prove Theorem 1 by Lemma 2 and Lemma 3 (i). $\square$

A.4 The proof of Theorem 2

Proof.

(i) From Theorem 1 we get $\hat{p}_n(x) \xrightarrow{p} p(x)$, hence to prove

$$\hat{a}_n(x) = \frac{A_n(x)}{\hat{p}_n(x)} \xrightarrow{p} \mu(x)$$

it is sufficient to verify that

$$A_n(x) \xrightarrow{p} \mu(x)p(x).$$

From Lemma 2, we obtain

$$A_n^0(x) \xrightarrow{p} \mu(x)p(x).$$

Now we prove

$$A_n(x) - A_n^0(x) \xrightarrow{p} 0. \quad (21)$$

By Lemma 3 (ii) $A_n(x) - A_n^0(x)$ has the same limit in probability as

$$\begin{aligned}\delta_{1,n}(x) &:= \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{X_{i\Delta_n} - X_{(i-1)\Delta_n}}{\Delta_n} \\ &= \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{X_{i\Delta_n} - X_{(i-1)\Delta_n}}{\Delta_n} \right).\end{aligned}$$

We now prove that $\delta_{1,n}(x) \xrightarrow{p} 0$. Using Assumption 1, Assumption 2, Assumption 4 and Lemma 1, we find

$$\begin{aligned} E[\delta_{1,n}(x)] &= E\left[K(X_{(i-1)\Delta_n})E\left[E\left[\left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)\middle|\mathcal{F}_{i\Delta_n}\right]\middle|\mathcal{F}_{(i-1)\Delta_n}\right]\right] \\ &= \frac{\Delta_n}{2}E\left[K(X_{(i-1)\Delta_n})\left(\mu(X_{(i-1)\Delta_n})\mu'(X_{(i-1)\Delta_n}) + \frac{1}{2}\sigma^2(X_{(i-1)\Delta_n})\mu''(X_{(i-1)\Delta_n})\right.\right. \\ &\quad \left.\left.+ \int_{\mathcal{E}}\{\mu(X_{(i-1)\Delta_n} + c(X_{(i-1)\Delta_n}, z)) - \mu(X_{(i-1)\Delta_n}) - \mu'(X_{(i-1)\Delta_n}) \cdot c(X_{(i-1)\Delta_n}, z)\}f(z)dz\right)\right] \\ &= O(\Delta_n) \end{aligned}$$

and

$$\begin{aligned} Var[\delta_{1,n}(x)] &= \frac{1}{n\Delta_n h_n}Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \sqrt{h_n}K(X_{(i-1)\Delta_n})\sqrt{\Delta_n}\left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)\right] \\ &=: \frac{1}{n\Delta_n h_n}Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\right]. \end{aligned}$$

We have $Var[\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i] = \frac{1}{n}\sum_{i=1}^n Var[g_i] + \varepsilon_n$ where ε_n represents the sum of $\frac{2}{n}\sum_{j=1}^{n-1}\sum_{i=j+1}^n$ terms involving the autocorrelations. Under Assumption 2, the series $\varepsilon_n = O(\frac{1}{\Delta_n^\alpha})$, and one easily obtains $Var[\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i] = O(\frac{1}{\Delta_n^\alpha})$ if $E[g_i^2] < \infty$ with the same procedures as $\varepsilon_{2,n}$ in proof of Lemma 3 (ii). In fact

$$E[g_i^2] = E\left[h_n K^2(X_{(i-1)\Delta_n})\Delta_n\left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)^2\right]$$

is finite because

$$E\left[\Delta_n\left(\frac{\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n}}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})}{\Delta_n}\right)^2\right] = \frac{2}{3}\left(E[\sigma^2(X_0)] + E\left[\int_{\mathcal{E}} c^2(X_0, z)f(z)dz\right]\right) + O(\Delta_n)$$

by Lemma 1 and Assumption 3. In conclusion,

$$Var[\delta_{1,n}] = \frac{1}{n\Delta_n h_n}Var\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\right] = O\left(\frac{1}{nh_n\Delta_n^{1+\alpha}}\right) \rightarrow 0 \text{ as } n \rightarrow \infty$$

by Assumption 7.

(ii) By Lemma 2, for “interior x ”, if $h_n = O((n\Delta_n)^{-2/5})$,

$$U_n^0(x) := \sqrt{n\Delta_n h_n^{1/2}}(\hat{a}_n^0(x) - \mu(x) - h_n B_{\hat{a}_n^0(x)}) \xrightarrow{d} N\left(0, \frac{M(x)}{2\sqrt{\pi}x^{1/2}p(x)}\right)$$

and for “boundary x ”, if $h_n = O((n\Delta_n)^{-1/3})$,

$$U_n^0(x) := \sqrt{n\Delta_n h_n}(\hat{a}_n^0(x) - \mu(x) - h_n B_{\hat{a}_n^0(x)}) \xrightarrow{d} N\left(0, \frac{M(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)p(x)}\right).$$

By the asymptotic equivalence theorem, it suffices to prove that

$$U_n(x) - U_n^0(x) \xrightarrow{p} 0$$

where $U_n(x) = \sqrt{h_n n\Delta_n}(\hat{a}_n(x) - \mu(x))$.

From part (i) we know that

$$U_n(x) - U_n^0(x) = \sqrt{h_n n \Delta_n} \left(\frac{\delta_{1,n}(x)}{\hat{p}_n(x)} \right)$$

and

$$\frac{\delta_{1,n}(x)}{\hat{p}_n(x)} = O_p(\Delta_n) = o_p(h_n),$$

thus the assumption $h_n n \Delta_n^3 \rightarrow 0$ leads to the result. $\square$

A.5 The proof of Theorem 3

Proof.

(i) From Theorem 1 we know $\hat{p}_n(x) \xrightarrow{p} p(x)$, hence to prove

$$\hat{b}_n(x) = \frac{B_n(x)}{\hat{p}_n(x)} \xrightarrow{p} \sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz$$

it is sufficient to verify that

$$B_n(x) \xrightarrow{p} \left(\sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz \right) p(x).$$

Following Lemma 2 under the conditions of the theorem we obtain

$$B_n^0(x) \xrightarrow{p} \left(\sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz \right) p(x).$$

So, we only need to prove that

$$B_n(x) - B_n^0(x) \xrightarrow{p} 0. \quad (22)$$

By Lemma 3 (iii) $B_n(x) - B_n^0(x)$ has the same limit in probability as

$$\begin{aligned} \delta_{2,n}(x) &:= \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n} \\ &= \frac{1}{n} \sum_{i=1}^n K(X_{(i-1)\Delta_n}) \left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n} \right). \end{aligned}$$

We now prove that $\delta_{2,n}(x) \xrightarrow{p} 0$. Let $q_i = \frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}$.

By Lemma 1

$$\begin{aligned} E\left[E\left[\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} \middle| \mathcal{F}_{i\Delta_n}\right] \middle| \mathcal{F}_{(i-1)\Delta_n}\right] &= \sigma^2(X_{(i-1)\Delta_n}) + \int_{\mathcal{E}} c^2(X_{(i-1)\Delta_n}) f(z) dz + O(\Delta_n), \\ E\left[\frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n} \middle| \mathcal{F}_{(i-1)\Delta_n}\right] &= \sigma^2(X_{(i-1)\Delta_n}) + \int_{\mathcal{E}} c^2(X_{(i-1)\Delta_n}) f(z) dz + O(\Delta_n). \end{aligned}$$

One has

$$E[q_i] = E\left[\left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)\right] = O(\Delta_n).$$

Similarly as $\delta_{1,n}(x)$, we can easily verify that $E[\delta_{2,n}(x)] \rightarrow 0$ by stationarity and the above equations. On the other hand,

$$\begin{aligned} \text{Var}[\delta_{2,n}(x)] &= \frac{1}{n\Delta_n h_n} \text{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{h_n} K(X_{(i-1)\Delta_n}) \sqrt{\Delta_n} \left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)\right] \\ &=: \frac{1}{n\Delta_n h_n} \text{Var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n s_i\right]. \end{aligned}$$

Using the same arguments as in the proof of Theorem 2 (i), it is easy to conclude

$$E\left[\Delta_n \left(\frac{\frac{3}{2}(\tilde{X}_{(i+1)\Delta_n} - \tilde{X}_{i\Delta_n})^2}{\Delta_n} - \frac{(X_{i\Delta_n} - X_{(i-1)\Delta_n})^2}{\Delta_n}\right)^2\right] = O(1)$$

assures $\text{Var}[\delta_{2,n}(x)] \rightarrow 0$.

(ii) By Lemma 2, for “interior x ”, if $h_n = O((n\Delta_n)^{-2/5})$,

$$U_n^0(x) := \sqrt{n\Delta_n h_n^{1/2}} (\hat{b}_n^0(x) - M(x) - h_n B_{\hat{b}_n^0(x)}) \xrightarrow{d} N\left(0, \frac{\int_{\mathcal{E}} c^4(x, z) f(z) dz}{2\sqrt{\pi} x^{1/2} p(x)}\right)$$

and for “boundary x ”, if $h_n = O((n\Delta_n)^{-1/3})$,

$$U_n^0(x) := \sqrt{n\Delta_n h_n} (\hat{b}_n^0(x) - M(x) - h_n B_{\hat{b}_n^0(x)}) \xrightarrow{d} N\left(0, \frac{\int_{\mathcal{E}} c^4(x, z) f(z) dz \Gamma(2\kappa + 1)}{2^{2\kappa+1} \Gamma^2(\kappa + 1) p(x)}\right).$$

By the asymptotic equivalence theorem, it suffices to prove that

$$U_n(x) - U_n^0(x) \xrightarrow{p} 0$$

where $U_n(x) = \sqrt{h_n n \Delta_n} (\hat{b}_n(x) - (\sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz))$.

From part (i) we know that

$$U_n(x) - U_n^0(x) = \sqrt{h_n n \Delta_n} \left(\frac{\delta_{2,n}(x)}{\hat{p}_n(x)}\right)$$

and

$$\frac{\delta_{2,n}(x)}{\hat{p}_n(x)} = O_p(\Delta_n) = o_p(h_n),$$

Thus the assumption $h_n n \Delta_n^3 \rightarrow 0$ leads to the result. $\square$

References

- Bandi, F., and T. Nguyen. 2003. “On the Functional Estimation of Jump Diffusion Models.” *Journal of Econometrics* 116: 293–328.
- Barndorff-Nielsen, O. E., and N. Shephard. 2006. “Econometrics of Testing for Jumps in Financial Economics using Bipower Variation.” *Journal of Financial Econometrics* 4: 1–30.
- Chen, S. X. 2000. “Probability Density Function Estimation using Gamma Kernels.” *Annals of the Institute of Statistical Mathematics* 52: 471–480.
- Cont, R., and P. Tankov. 2004. *Financial Modeling with Jump Processes*. New York: Chapman and Hall/CRC.
- Chen, Y., and L. Zhang. 2015. “Local Linear Estimation of Second-Order Jump-Diffusion Model.” *Communications in Statistics-Theory and Methods* 44: 3903–3920.
- Ditlevsen, S., and M. Sørensen. 2004. “Inference for Observations of Integrated Diffusion Processes.” *Scandinavian Journal of Statistics* 31: 417–429.

- Fan, J., and I. Gijbels. 1996. *Local Polynomial Modeling and its Applications*. London: Chapman and Hall.
- Florens-Zmirou, D. 1989. "Approximate Discrete Time Schemes for Statistics of Diffusion Processes." *Statistics* 20: 547–557.
- Funke, B., and E. Schmisser. 2018. "Adaptive Nonparametric Drift Estimation of an Integrated Jump Diffusion Process." Forthcoming in *ESAIM: Probability and Statistics*. DOI: 10.1051/ps/2018005.
- Gloter, A. 2001. "Parameter Estimation for a Discret Sampling of an Integrated Ornstein-Uhlenbeck Process." *Statistics* 35: 225–243.
- Gloter, A. 2006. "Parameter Estimation for a Discretely Observed Integrated Diffusion Process." *Scandinavian Journal of Statistics* 33: 83–104.
- Gospodinov, N., and M. Hirukawa. 2012. "Nonparametric Estimation of Scalar Diffusion Processes of Interest Rates using Asymmetric Kernels." *Journal of Empirical Finance* 19: 595–609.
- Gugushvili, S., and P. Spereij. 2012. "Parametric Inference for Stochastic Differential Equations: A Smooth and Match Approach." *Latin American Journal of Probability & Mathematical Statistics* 9: 609–635.
- Hanif, M. 2015. "Nonparametric Estimation of Second-Order Diffusion Equation by using Asymmetric Kernels." *Communications in Statistics-Theory and Methods* 44: 1896–1910.
- Hanif, M. 2016. "A Nonparametric Approach to the Estimation of Jump-Diffusion Models with Asymmetric Kernels." *Cogent Mathematics* 3: 1–15.
- Jacod, J., and A. Shiryaev. 2003. *Limit Theorems for Stochastic Processes*. 2nd ed. Berlin: Springer.
- Johannes, M. S. 2004. "The Economic and Statistical Role of Jumps to Interest Rates." *Journal of Finance* 59: 227–260.
- Kessler, M. 1997. "Estimation of an Ergodic Diffusion from Discrete Observations." *Scandinavian Journal of Statistics* 24: 211–229.
- Kristensen, D. 2010. "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach." *Econometric Theory* 26: 60–93.
- Lin, Z., and Z. Bai. 2010. *Probability Inequalities*. Beijing: Science Press.
- Madan, D., and E. Seneta. 1990. "The Variance Gamma (V.G.) Model for Share Market Returns." *Journal of Business* 63: 511–524.
- Nicolau, J. 2007. "Nonparametric Estimation of Second-Order Stochastic Differential Equations." *Econometric theory* 23: 880–898.
- Nicolau, J. 2008. "Modeling Financial Time Series Thruhg Second-Order Stochastic Differential Equations." *Statistics and Probability Letters* 78: 2700–2704.
- Protter, P. 2004. *Stochastic Integration and Differential Equations*. 2nd ed. Berlin: Springer.
- Shimizu, Y., and N. Yoshida. 2006. "Estimation of Parameters for Diffusion Processes with Jumps from Discrete Observations." *Statistical Inference for Stochastic Processes* 9: 227–277.
- Song, Y. 2017. "Nonparametric Estimation for Second-Order Jump-Diffusion Model in High Frequency Data." Forthcoming in *Singapore Economic Reviews*. DOI: 10.1142/S0217590817500102.
- Song, Y., Z. Lin, and H. Wang. 2013. "Re-Weighted Nadaraya-Watson Estimation of Second-Order Jump-Diffusion Model." *Journal of Statistical Planning and Inference* 143: 730–744.
- Wang, H., and Z. Lin. 2011. "Local Linear Estimation of Second-Order Diffusion Models." *Communications in Statistics-Theory and Methods* 40: 394–407.
- Wang, H., and L. Zhou. 2017. "Bandwidth Selection of Nonparametric Threshold Estimator in Jump-Diffusion Models." *Computers & Mathematics with Applications* 73: 211–219.
- Xu, K., and P. Phillips. 2011. "Tilted Nonparametric Estimation of Volatility Functions with Empirical Applications." *Journal of Business & Economic Statistics* 29: 518–528.
- Yan, T., and C. Mei. 2016. A Test for a Parametric form of the Volatility in Second-Order Diffusion Models. *Computational Statistics*. DOI: 10.1007/s00180-016-0685-z.

Supplementary Material: The online version of this article offers supplementary material (DOI: <https://doi.org/10.1515/snde-2018-0001>).