Econometric Theory, **37**, 2021, 926–958.
doi:10.1017/S0266466620000389

ESTIMATION OF VOLATILITY FUNCTIONS IN JUMP DIFFUSIONS USING TRUNCATED BIPOWER INCREMENTS*

JIHYUN KIM
Toulouse School of Economics
University of Toulouse Capitole

JOON Y. PARK
Indiana University
Sungkyunkwan University

BIN WANG
Harbin Institute of Technology, Shenzhen

In this article, we introduce and analyze a new methodology to estimate the volatility functions of jump diffusion models. Our methodology relies on the standard kernel estimation technique using truncated bipower increments. The relevant asymptotics are fully developed, allowing for the time span to increase as well as the sampling interval to decrease, and accommodate both stationary and nonstationary recurrent processes. We evaluate the performance of our estimators by simulation and provide some illustrative empirical analyses.

1. INTRODUCTION

Diffusions and jump diffusions have been used widely in financial economics to model time series of various asset prices. In parallel, the statistical inference in these models has been one of the most active research areas in econometrics and statistics, to which an extensive literature has been devoted. For diffusion models without jumps, Bandi and Phillips (2003) and Aït-Sahalia and Park (2016) study nonparametric kernel estimation based on discrete samples, and establish their asymptotics as $\delta \to 0$ and $T \to \infty$, where δ and T denote, respectively, the sampling interval and the time span. (See also Fan and Zhang, 2003;

*We thank the Editor Peter C.B. Phillips, the Co-Editor Dennis Kristensen, and three anonymous referees for many helpful comments. For useful discussions, we are also grateful to Yacine Aït-Sahalia, Yoosoon Chang, Nour Meddahi, Mathieu Rosenbaum, Roberto Renò, Jun Yu, and the participants at 2015 Toulouse Financial Econometrics Conference, 2016 Princeton-QUT-SJTU-SMU Econometrics Conference and 2017 Asian Meeting of Econometric Society. Jihyun Kim is grateful to the French Government and the ANR for support under the Investissements d'Avenir program, grant ANR-17-EURE-0010. Bin Wang gratefully acknowledges financial support provided to him by Shenzhen Key Research Base of Humanities and Social Sciences. Address correspondence to Jihyun Kim, Toulouse School of Economics, Toulouse 31000, France; e-mail: jihyun.kim@tse-fr.eu.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

Aït-Sahalia and Mykland, 2003; Aït-Sahalia and Mykland, 2004; Renò, 2006, 2008; Kristensen, 2010; Yu et al., 2014; Kanaya, 2016; and Kanaya and Kristensen, 2016 for some related works.) For jump diffusion models, Bandi and Nguyen (2003) consider the nonparametric kernel estimation of total infinitesimal moments, and Mancini and Renò (2011) and Park and Wang (2018) use the threshold approach to separately identify and estimate the diffusive and jump volatilities, both relying on discrete samples. They also obtain the relevant asymptotics as $\delta \to 0$ with $T \to \infty$ or T being fixed. Bandi and Renò (2018) use a threshold method to nonparametrically estimate spot diffusive volatilities in stochastic volatility models.

In this paper, we introduce and analyze a nonparametric kernel method to estimate the volatility functions of jump diffusion models using bipower increments of discrete samples with threshold truncation.[1] The use of truncated bipower increments makes it possible to effectively disentangle the diffusive and jump volatilities of jump diffusions, at least asymptotically as $\delta \to 0$. We develop the asymptotics for local constant and local linear estimators of diffusive and jump volatility functions under $T \to \infty$ jointly with $\delta \to 0$. They are fine enough to clearly and explicitly identify the trade-offs between the bias and variance terms of the estimators which makes it possible to find the theoretical optimal bandwidths.[2] Our asymptotics are very general in that they are applicable for both stationary and nonstationary processes and for diffusions with jump processes having possibly infinite activities.

Our framework and approach are related to those of Mancini and Renò (2011) and Park and Wang (2018), which establish the asymptotics relevant for the nonparametric kernel estimation of volatility functions in jump diffusion models using squared increments of discrete samples with threshold truncation, hereafter referred to as threshold estimators.[3] Of the two, we follow Park and Wang (2018) more closely. Compared with Mancini and Renò (2011), our asymptotics are more general in several aspects. First, we let $T \to \infty$ jointly with $\delta \to 0$, while they assume that T is fixed. This is important, since the jump component in a jump diffusion model is fully identified only when $T \to \infty$. Second, they only analyze the variance term by restricting h to be of order between $\delta^{3/8}$ and $\delta^{1/3}$, which makes the bias term negligible asymptotically. In contrast, we impose no restriction on h and analyze the bias term as well as the variance term. As a result, our asymptotics

[1] We consider estimators using bipower increments with threshold, which will be referred to as threshold-bipower estimators, in Sections 3 and 4. However, for the simulation and empirical studies in Section 5, we also use the estimator using bipower increments without threshold truncation, that is, bipower estimators. Within our framework, threshold truncation has no asymptotic effects, and therefore, our asymptotics for threshold-bipower estimators are also applicable for bipower estimators using bipower increments without threshold truncation. This was shown explicitly in earlier versions of this paper.

[2] Though the optimal bandwidths are well defined theoretically, they are infeasible and their feasible versions are generally not optimal.

[3] In terms of construction of estimators for volatility functions, our kernel estimators use truncated bipower increments, while the kernel estimators in Mancini and Renò (2011) and Park and Wang (2018) use truncated squared increments.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

provide the optimal bandwidths for our estimators, which control for the bias and variance terms. Finally, we consider the local linear estimator, in addition to the local constant estimator they studied. The two estimators have bias terms that are of different orders of magnitude. Strictly speaking, however, our model is not directly comparable to theirs. We specify jump volatility as well as diffusive volatility to be state dependent, while they model intensity of jumps as being partially state dependent.

In comparison with Park and Wang (2018), this paper considers the same jump diffusion model setup, as well as estimands, that is, diffusive and jump volatility functions. However, there are some differences between this paper and Park and Wang (2018). First, the two papers consider different estimation approaches. This paper considers kernel estimators using truncated bipower increments, while Park and Wang (2018) use kernel estimators with truncated squared increments. Distinct technical arguments are required to derive the asymptotic results of our threshold bipower estimators, compared to those in Park and Wang (2018).[4] Second, the asymptotic properties of the estimators of volatility functions motivated by the two different approaches are directly comparable. More explicitly, in terms of asymptotic bias, the two approaches yield equivalent estimators for both diffusive and jump volatility functions. However, only for jump volatility functions, the estimators based on the two approaches have the same asymptotic variance. For diffusive volatility functions, the estimator using truncated bipower increments in our paper has variance that is 1.142^2 times larger than that using the threshold method in Park and Wang (2018). Third, our simulation study shows that none of the two estimators dominates the other in finite samples. Their relative finite sample performances vary across different models and sampling frequencies. For estimating diffusive volatility functions of our simulation models, the estimator using truncated bipower increments outperforms the threshold estimator in terms of both finite sample bias and variance. On the other hand, for estimating jump volatility functions, the estimator using truncated bipower increments has smaller finite sample bias, while it is largely comparable to the threshold estimator in terms of finite sample variance.[5] Therefore, it seems that the two kernel estimation methods, one using truncated bipower increments in this paper and the other one

[4]For instance, the leading variance term of our threshold bipower estimators for diffusive volatility functions contains a noncontinuous martingale difference sequence $\sigma^2(X_{(i-1)\delta})(|\Delta_i W \Delta_{i+1} W| - 2\delta/\pi)$, while the leading variance term of the threshold estimators involves a continuous martingale difference sequence $\int_{(i-1)\delta}^{i\delta}(X_t^c - X_{(i-1)\delta}^c)\sigma(X_t)dW_t$. Correspondingly, their asymptotic results, as well as required arguments, are rather different. Moreover, we also need different arguments (from those in Park and Wang, 2018) to derive asymptotic orders of error terms in the threshold bipower estimators. See, for example, Lemmas A.2–A.4 in the Mathematical Appendix and their proofs in the online supplement for more details.

[5]Our simulation shows that, compared to the threshold estimator, our threshold bipower estimator reduces 54.19% and 58.33% of finite sample biases on average (over all of our simulation models and realistic parameter values), for estimating diffusive and jump volatility functions, respectively. Moreover, our threshold bipower estimator of diffusive volatility functions has finite sample standard errors that are on average 44.82% smaller than those of the threshold estimator in Park and Wang (2018).

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

using truncated squared increments in Park and Wang (2018), are similar and each has its own particular advantage.

The approaches relying on bipower increments, thresholds, and truncated bipower increments are used widely in estimating the integrated volatilities of general semimartingales with jumps. Our kernel estimators for the diffusive and jump volatility functions of jump diffusions are indeed motivated by the use of realized variance and bipower variations to estimate the integrated volatilities of diffusive and jump components of semimartingales.[6] The relative advantages and disadvantages of the estimators for volatility functions of jump diffusions based on different approaches are largely comparable to those of the corresponding estimators for integrated volatilities of semimartingales. For the approach relying on bipower and multipower variations, see, for example, Barndorff-Nielsen and Shephard (2003, 2004, 2006) and Barndorff-Nielsen et al. (2005); Barndorff-Nielsen, Shephard, and Winkel (2006). For the threshold approach, the reader is referred to Jacod (2008, 2012), Mancini (2009), and Jacod and Rosenbaum (2013), among others. Corsi, Pirino, and Renò (2010) introduce the threshold-bipower variation approach combining the threshold and bipower variation methods to reduce the finite sample bias of bipower variation in estimating integrated diffusive volatility. For some related discussions, see also, Christensen, Oomen, and Podolskij (2010), Boudt, Croux, and Laurent (2011), and Andersen, Dobrev, and Schaumburg (2012).

Our work is also related to Bandi and Nguyen (2003), who consider nonparametric kernel estimation of infinitesimal conditional moments of increments for jump diffusions. The second infinitesimal conditional moments are given by the sum of diffusive and jump volatilities. Their approach requires neither the threshold method nor the bipower method. However, their model is generally more restrictive, and their estimator requires a parametric specification of the probability measure on jump size in order to identify diffusive and jump volatilities separately. In contrast, our approach is fully nonparametric, and we individually identify and estimate diffusive and jump volatility functions. Moreover, our asymptotics generally require less stringent technical conditions on the underlying jump diffusion models. For instance, we allow for jumps with infinite activity, whereas they restrict their jumps to be bounded with finite activity.

As expected from the earlier works cited above, our asymptotics involve the local time of underlying jump diffusion, denoted as $\ell(T,x)$, where T is the time span and x is the spatial point at which we estimate the volatility functions. Our nonparametric kernel estimates for both the diffusive and jump volatilities are consistent and asymptotically normal or mixed normal under mild regularity conditions, depending upon whether the underlying jump diffusion is stationary or nonstationary, if we choose the bandwidth parameter h appropriately. However, their asymptotic behaviors are quite different. For the estimation of diffusive

[6] In particular, our estimator of the jump volatility function relies on the same idea used in the jump test that compares realized variance and bipower variation.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

volatility, we need $h\ell(T,x) \to_p 0$ for asymptotic normality or mixed normality, though consistency generally follows. On the other hand, for the estimation of jump volatility, consistency requires $h\ell(T,x) \to_p \infty$, while asymptotic normality or mixed normality holds in general. In particular, we should have $T \to \infty$ for the consistent estimation of jump volatility, whereas diffusive volatility can be consistently estimated, even with T being fixed. Moreover, the local constant and local linear estimators have distinct leading bias terms, though they have a common leading variance term. Especially, for the estimators of diffusive volatility, the leading bias terms of the local constant and local linear estimators are of different orders of magnitude, not merely distinct from each other. If $h\ell(T,x) \to_p 0$, the bias term of the local linear estimator is of order smaller than that of the local constant estimator, this is rather unusual.

The rest of the paper is organized as follows. Section 2 introduces the model and other preliminaries such as our asymptotic setup, main assumptions and motivations. In Sections 3 and 4, the local constant and local linear estimators of diffusive and jump volatility functions are, respectively, defined and their asymptotics are developed. The simulation results and some simple empirical illustrations are provided in Section 5. Section 6 concludes the paper, and all mathematical proofs are in the Mathematical Appendix.

A word on notation. Standard notations such as $\to_p$ and $\to_d$ are used frequently, and $=_d$ signifies equality in distribution. Moreover, we use various functional notations to simplify the exposition. For function f defined on $\mathbb{R}$, we let $f_0 = f$, $f_k(x) = 1\{x \geq 0\}\int_x^\infty f_{k-1}(u)du - 1\{x < 0\}\int_{-\infty}^x f_{k-1}(u)du$ for $k = 1,2,3$, whenever they are well defined. Moreover, ι^k denotes the function $\iota^k(x) = x^k$, so that we have in particular $\iota^k f(x) = x^k f(x)$ for $k = 1,2$. By convention, we write ιf_k and f_k^2 to signify, respectively, $(\iota f)_k$ and $(f_k)^2$ for $k = 1,2,3$. We use ι to denote the integral with respect to the Lebesgue measure, and we let $\iota(f) = \int f(x)dx$ and define $\iota_k(f) = \int \iota^k f(x)dx$ for $k = 1,2$. Finally, f' and f'' denote the first and second derivatives of f, and we write $(\iota f)'$ and $(\iota f)''$ as $\iota f'$ and $\iota f''$ for simplicity. The notations introduced here will be used repeatedly throughout the paper without further reference.

2. THE MODEL AND PRELIMINARIES

We consider the jump diffusion model defined as

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + \tau(X_{t-})dJ_t, \tag{2.1}$$

where μ, σ and τ are functions defined on the domain $\mathcal{D}$ of X, W is the standard Brownian motion, and J is a jump process. We suppose that μ, σ and τ are completely unknown, in contrast to the parametric models studied in Aït-Sahalia (2002), Chen, Gao, and Tang (2008), Jeong and Park (2016), and Chang and Chen (2011), among many others, which assume that they are known up to some unknown parameters. In financial modeling, X typically represents an interest rate,

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

or the logarithmic price of a financial asset or exchange rate, with $\mathcal{D} = (0, \infty)$ or $(-\infty, \infty)$.

The jump process J is specified as

$$J_t = \int_0^t \int_{\mathbb{R}} z\Lambda(ds, dz),$$

where Λ is a Poisson random measure on $[0, \infty) \times \mathbb{R}$. For $t > 0$ and $A \subset \mathbb{R}$ given, $\Lambda\big([0, t], A\big)$ defines a Poisson process representing the number of jumps with size in a set A that occurs before time t. Throughout the paper, we assume that Poisson random measure Λ is independent of Brownian motion W and the initial value of process X. We let

$$\mathbb{E}\Lambda(dt, dz) = \lambda(dz)dt,$$

where λ is the Lévy measure associated with Λ, and denote the compensated Poisson random measure as $\Gamma(dt, dz) = \Lambda(dt, dz) - \lambda(dz)dt$. We assume that $\lambda(dz)$ is absolutely continuous with respect to the Lebesgue measure, and we write $\lambda(dz) = \lambda(z)dz$.[7] Note that $\lambda(A)$ for $A \subset \mathbb{R}$ counts the expected number of jumps with sizes in A during any time interval of unit length. Our formulation of the Poisson random measure Λ being homogeneous, λ will not change with respect to t. This is actually not restrictive since we can always have an equivalent homogeneous representation of a nonhomogeneous Poisson random measure via thinning, as illustrated in Jacod (2008) and Bollerslev and Todorov (2011).

In the development of our theory, we let the jump diffusion X be observed at intervals of length δ over time $[0, T]$ with the sample size n given by $n = T/\delta$. In what follows, we suppose that

$$X_\delta, X_{2\delta}, \ldots, X_{n\delta}$$

are observed, and we let

$$\Delta_i X = X_{i\delta} - X_{(i-1)\delta}$$

and define $\Delta_i W$ and $\Delta_i J$ from W and J similarly for $i = 1, \ldots, n$. For the development of our asymptotics, we assume that $\delta \to 0$ sufficiently fast relative to T, which we allow to be fixed, $T = \overline{T}$, or to increase up to infinity, $T \to \infty$. Furthermore, for the bandwidth parameter h used in our kernel estimators, we set $h \to 0$. In sum, our asymptotics are developed under $h \to 0$ and $\delta \to 0$ jointly with either $T = \overline{T}$ fixed or $T \to \infty$.

In all our asymptotics requiring $T \to \infty$, we assume the jump diffusion process in (2.1) to be recurrent. Therefore, we have a unique (up to constant multiples) invariant measure m on the domain $\mathcal{D}$ of X, and X is positive recurrent if $m(\mathcal{D}) < \infty$ and null recurrent if $m(\mathcal{D}) = \infty$. See, for example, Höpfner and Löcherbach

[7] We follow the usual convention and use λ to denote both the Lévy measure itself and its density with respect to the Lebesgue measure.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

(2003). If X is positive recurrent with finite m, X becomes stationary with invariant distribution given by $m/m(\mathcal{D})$.

Assumption 2.1. (a) $\sigma > 0$ and $\tau > 0$ on $\mathcal{D}$, (b) μ, σ and τ are twice continuously differentiable on $\mathcal{D}$, (c) $\alpha = \inf\left\{r : \int_{|z|\le 1} |z|^r \lambda(z)dz < \infty\right\} \in [0,1)$, (d) $\int_{|z|>1} |z|^{4+\varepsilon}\lambda(z)dz < \infty$ for some $\varepsilon > 0$, (e) $\int \lambda_1\big(z/\tau(z)\big)m(dz) < \infty$, (f) both λ and m admit continuous densities, respectively, and (g) the Darling–Kac condition holds: for some $\rho \in (0,1]$, there exists a function $\kappa : \mathbb{R}_+ \mapsto \mathbb{R}_+$ which is regularly varying of index ρ at infinity, such that for every m-integrable function g, $\frac{1}{\kappa(\eta)}\int_0^\infty \exp(-t/\eta)\mathbb{E}_x[g(X_t)]\,dt \to m(g)$ for m-almost all $x \in \mathcal{D}$ as $\eta \to \infty$.

Assumption 2.1(c) and (d) concern the vibrancy of small jumps, as well as the tail behavior of the Lévy measure λ related to large jumps in J. Note that the latter, along with τ, determines the tail behavior of the jumps in X. Under the local boundedness of τ, the value of α as defined in (c) coincides with the Blumenthal–Getoor index. The condition $\alpha < 1$ implies that the jump process J has finite variation and allows for a wide range of jump processes including compound Poisson process. However, our assumption excludes jump processes of unbounded variation, which is implied by either $\alpha \in (1,2]$, or $\alpha = 1$ with $z\lambda(dz)$ having a subpolynomial divergence rate near zero. A jump process with $\alpha = 0$ may have either finite activity or infinite activity with a subpolynomial divergence rate of $\lambda(dz)$ around zero. However, if combined with $\int_{|z|\le 1}\lambda(z)dz < \infty$, $\alpha = 0$ ensures the process to be finitely active. In (d), we require that large jumps have finite moments up to the fourth order. We introduce the technical condition in (e) to regulate the behavior of X around the boundaries. It does not seem to be stringent, though it greatly simplifies the derivation of our asymptotics in the paper. Assumption 2.1(f) and (g) are essentially the same as Assumptions 2.4 and 3.2 in Ueltzhöfer (2013), and they are introduced to ensure a central limit theorem holds for general recurrent jump diffusions. For positive recurrent jump diffusions with $m(\mathcal{D}) < \infty$, (g) holds with $\rho = 1$ and $\kappa(t) = t/m(\mathcal{D})$. See also, Touati (1987) and Höpfner and Löcherbach (2003), among others.

For the identification of our model, we need to assume

Assumption 2.2. $\int_{\mathbb{R}} z\lambda(dz) = 0 \quad \textit{and} \quad \int_{\mathbb{R}} z^2\lambda(dz) = 1.$

Clearly, we may redefine, if necessary, the drift and jump volatility to make any jump diffusion satisfy Assumption 2.2 after compensating and normalizing its jump and jump size.

The asymptotics developed in our paper heavily rely on the local time $\ell(t,x)$ of X defined in (2.1). See, for example, Sect. 6.1 of Bosq (1998) for the definition of local time. We may interpret it as an occupation density of X at point $x \in \mathcal{D}$, which yields the so-called occupation time formula

$$\int_0^T f(X_t)dt = \int_{\mathcal{D}} f(x)\ell(T,x)dx$$

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

for any positive Borelian function f. Given J has finite variation, X admits a version of local time $\ell(t,x)$ jointly continuous in (t,x). Throughout this paper, we will always refer to this bicontinuous version of local time. See, for example, Thm. 76 and its corollaries in Chap. IV of Protter (2005).

Assumption 2.3. $\bar{\ell}_h(T,x) = \sup_{|u|\le 1} \ell(T,x+hu) = O_p(\ell(T,x)^2)$.

Assumption 2.3 regulates the divergence of local time in the neighborhood of a spatial point. It is not essential and is introduced mainly to simplify our exposition by representing the asymptotic orders of error terms only as functions of $\ell(T,x)$.

We assume that the kernel function K satisfies

Assumption 2.4. $(a)K$ is nonnegative, bounded, twice continuously differentiable and has support $[-1,1]$, $(b)\int K(x)dx = 1$ and $\int xK(x)dx = 0$.

The above conditions for the kernel function K are standard, except for the boundedness of support. Though it seems nonessential, the boundedness condition greatly simplifies the proofs of our theorems in the paper.

Assumption 2.5. (a) For $f = \mu, \mu', \mu'', \iota\mu, \iota\mu', \iota\mu'', \sigma^2, \sigma^{2\prime}, \sigma^{2\prime\prime}, \tau^2, \tau^{2\prime}, \tau^{2\prime\prime}$, we have $|f(x)| \le g(x)$, where g is locally bounded and $g(x) \le c|x|^p$ at boundaries $\pm\infty$ and $g(x) \le c|x|^{-p}$ at boundary 0 for some $c > 0$ and $p \in \mathbb{R}$, (b) $\sup_{0\le t\le T} |X_t| = O_p(T^q)$ if boundaries are $\pm\infty$ and $\left(\inf_{0\le t\le T} |X_t|\right)^{-1} = O_p(T^q)$ if 0 is one of the boundaries for some $q \ge 0$.

Assumption 2.5(a) is not stringent and holds for all jump diffusions commonly used in practice, typically with $p > 0$ not so large. Likewise, Assumption 2.5(b) is also mild and satisfied widely. It holds for Brownian motion with $q = 1/2$, and also for a large class of null recurrent jump diffusions. See, for example, Jeong and Park (2016). In Assumption 2.6 below, we require that δ be sufficiently small relative to the bandwidth h we use in our kernel estimation, as well as to T.

Assumption 2.6. $\delta = o(h^3 \wedge T^{-6pq})$, where p, q are as defined in Assumption 2.5.

Assumption 2.7. For every $x \in \mathcal{D}$, there exist $\varepsilon > 0$ and $\eta > 0$ such that

$$\sup_{|y-x|\le\varepsilon} \sup_{0<t<\eta} \mathbb{E}\left[|f|^r(X_t)|X_0 = y\right] < \infty$$

with any $r \ge 1$ for $f = \mu, \sigma$, and with any $\alpha < r < 1$ for $f = \tau$.

Our asymptotics include

$$\xi(T,x) = \int_0^T \lambda_1((x-X_t)/\tau(X_t))dt \quad \text{and} \quad \phi(x) = \int \lambda_1((x-u)/\tau(u))m(u)du.$$

Under Assumption 2.1(e), $\xi(T,x)$ is an additive functional of X, and so is $\ell(T,x)$ with the Dirac measure at x (see Revuz and Yor, 1998, Prop. 2.4 in p. 410). Hence, by the ratio limit theorem (Höpfner and Löcherbach, 2003, Cor. 1.8 in p. 12), it follows that

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

$$\xi(T,x)/\ell(T,x) \to_{a.s.} \phi(x)/m(x)$$

as $T \to \infty$. The functional $\xi(T,x)$ represents the bias term originated from the jump component of our model.

Subsequently, we introduce our estimators and develop their asymptotics. In the paper, we focus on the estimation of diffusive volatility function σ and jump volatility function τ based on truncated bipower increments. In particular, we do not consider the estimation of the Lévy density λ. A consistent estimator of λ can be obtained from our estimate of jump volatility τ analogous to Park and Wang (2018), where the reader is referred for more details.[8] For the nonparametric estimation and relevant asymptotics of local time $\ell(T,x)$, drift $\mu(x)$, as well as total volatility $\omega(x)$, where $\omega^2(x) = \sigma^2(x) + \tau^2(x)$, in jump diffusion models, the reader is also referred to Park and Wang (2018). Although our asymptotics in Theorems 3.1, 3.2, 4.1, and 4.2 are obtained for threshold-bipower estimators, they are applicable also for bipower estimators using bipower variations without threshold truncation.

3. ESTIMATION OF DIFFUSIVE VOLATILITY

For the estimation of diffusive volatility $\sigma(x)$, we write

$$\begin{aligned}
&|\Delta_i X||\Delta_{i+1} X| 1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1} X| \le \delta^\beta\} \\
&= \frac{2\delta}{\pi}\sigma^2(X_{(i-1)\delta}) + \left(\sigma^2(X_{(i-1)\delta})|\Delta_i W||\Delta_{i+1} W| - \frac{2\delta}{\pi}\sigma^2(X_{(i-1)\delta})\right) \\
&\quad + \left(|\Delta_i X||\Delta_{i+1} X| 1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1} X| \le \delta^\beta\} - \sigma^2(X_{(i-1)\delta})|\Delta_i W||\Delta_{i+1} W|\right)
\end{aligned} \tag{3.1}$$

for some $\beta > 0$. Note that the second term in (3.1) has a martingale difference sequence, since in particular

$$\mathbb{E}|\Delta_i W||\Delta_{i+1} W| = \frac{2\delta}{\pi}.$$

Moreover, we will show that the third term in (3.1) becomes asymptotically negligible under appropriate conditions. Therefore, the diffusive variance $\sigma^2(x)$ may be obtained from the standard nonparametric kernel regression of scaled truncated bipower increments

$$(\pi/2\delta)|\Delta_i X||\Delta_{i+1} X| 1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1} X| \le \delta^\beta\}$$

on $(X_{(i-1)\delta})$ for $i = 1, \ldots, n$.[9]

[8] We may readily establish consistency of the resulting estimator under suitable regularity conditions, as shown in earlier versions of this paper.

[9] In comparison, Park and Wang (2018) consider the standard nonparametric kernel regression of scaled truncated squared increments $((1/\delta)(\Delta_i X)^2 1\{|\Delta_i X| \le \delta^\beta\})$ on $(X_{(i-1)\delta})$ for $i = 1, ..., n$.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

Assumption 3.1. $0 < \beta < 1/2$.

In (3.1) we consider bipower increments with threshold truncation instead of bipower increments themselves, and use only the increments $\Delta_i X$ that do not contain a jump larger than the threshold level δ^β for $\beta > 0$. As we will later show by simulation, threshold truncation generally and unambiguously has a positive effect on the finite sample performance of our estimator. Intuitively, it is clear that the truncation level δ^β has to decrease more slowly than the modulus of continuity of Brownian motion as $\delta \to 0$ to avoid truncating out continuous increments of X, and this requires $\beta < 1/2$.[10]

3.1. Local Constant Estimator

Our local constant estimator for $\sigma^2(x)$ is defined as

$$\hat{\sigma}^2(x) = \frac{P_T(K,\sigma^2)}{Q_T(K)},$$

where

$$P_T(K,\sigma^2) = \frac{\pi}{2}\frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta} - x}{h}\Big)|\Delta_i X||\Delta_{i+1} X| 1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1} X| \le \delta^\beta\},$$

$$Q_T(K) = \frac{\delta}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta} - x}{h}\Big),$$

for some $\beta > 0$, which yields the estimation error that can be decomposed as

$$\hat{\sigma}^2(x) - \sigma^2(x) = \hat{\sigma}_p^2(x) + \hat{\sigma}_q^2(x) + \hat{\sigma}_r^2(x),$$

where

$$\hat{\sigma}_p^2(x) = \frac{N_T(K,\sigma^2)}{Q_T(K)}, \qquad \hat{\sigma}_q^2(x) = \frac{M_T(K,\sigma^2)}{Q_T(K)}$$

with

$$N_T(K,\sigma^2) = \frac{\delta}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta} - x}{h}\Big)\big[\sigma^2(X_{(i-1)\delta}) - \sigma^2(x)\big],$$

$$M_T(K,\sigma^2) = \frac{\pi}{2}\frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta} - x}{h}\Big)\Big[\sigma^2(X_{(i-1)\delta})\big(|\Delta_i W||\Delta_{i+1} W| - (2/\pi)\delta\big)\Big],$$

[10] As discussed, for our simulation and empirical illustration, we also consider the estimators of diffusive and jump volatility functions based on bipower increments without threshold truncation. They are defined exactly in the same way as the estimators with threshold truncation that we introduce explicitly in the paper. They only use bipower increments themselves without threshold truncation.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

and

$$\hat{\sigma}_r^2(x) = \frac{R_T(K,\sigma^2)}{Q_T(K)}$$

with

$$R_T(K,\sigma^2) = \frac{\pi}{2}\frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\Big[|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1}X| \le \delta^\beta\} - \sigma^2(X_{(i-1)\delta})|\Delta_i W||\Delta_{i+1}W|\Big].$$

Following the usual convention, we call $\hat{\sigma}_p^2(x)$, $\hat{\sigma}_q^2(x)$, and $\hat{\sigma}_r^2(x)$, respectively, the bias, variance, and error terms of $\hat{\sigma}^2(x)$.

THEOREM 3.1. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$\hat{\sigma}_p^2(x) =_d \frac{h^2\iota_2(K)}{2}\left[\left(4\sigma^{2\prime}\left(\frac{\mu}{\sigma^2}-\frac{\sigma'}{\sigma}\right)+\sigma^{2\prime\prime}\right)(x)+\frac{8\sigma'(x)\xi(T,x)}{\sigma(x)\ell(T,x)}\right] + 4\iota(\iota K_1^2)^{1/2}\sigma'(x)h^{3/2}\ell(T,x)^{-1/2}Z_p + o_p(h^2) + o_p(h^{3/2}\ell(T,x)^{-1/2}),$$

and

$$(h\ell(T,x)/\delta)^{1/2}\hat{\sigma}_q^2(x) \to_d c(\pi)\iota(K^2)^{1/2}\sigma^2(x)Z_q,$$

where $c(\pi) = (\pi^2/4+\pi-3)^{1/2}$, Z_p *and* Z_q *are two independent standard normal random variates both of which are independent of* $\ell(T,x)$. *Moreover,*

$$\hat{\sigma}_r^2(x) = O_p(\delta^{1/2}),$$

which becomes asymptotically negligible if h is chosen appropriately.

Remark 3.1. (a) The consistency of $\hat{\sigma}^2(x)$ requires no additional assumption. In particular, it is not required to have $T \to \infty$.

(b) The asymptotic mixed normality of $\hat{\sigma}^2(x)$ holds if $h\ell(T,x) \to_p 0$. If this condition holds, the bias term $\hat{\sigma}_p^2(x)$ becomes

$$\hat{\sigma}_p^2(x) =_d 4\iota(\iota K_1^2)^{1/2}\sigma'(x)h^{3/2}\ell(T,x)^{-1/2}Z_p + o_p(h^{3/2}\ell(T,x)^{-1/2}),$$

and the variance term $\hat{\sigma}_q^2(x)$ dominates the error term $\hat{\sigma}_r^2(x)$ asymptotically. In this case, therefore, both the bias and variance terms are asymptotically mixed normal, and become normal if X is stationary.

(c) The leading term of $\hat{\sigma}_p^2(x)$ does not represent the asymptotic bias of $\hat{\sigma}^2(x)$ *if* $h\ell(T,x) \to_p 0$. Instead, it provides a variance term additional to $\hat{\sigma}_q^2(x)$. If $h\ell(T,x) \to_p \infty$, $\hat{\sigma}_p^2(x)$ has the leading term representing the asymptotic bias of $\hat{\sigma}^2(x)$. However, in this case, $\hat{\sigma}_q^2(x)$ is dominated by $\hat{\sigma}_r^2(x)$, and therefore, the variance term becomes asymptotically negligible.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

(d) In case $h\ell(T,x) \to_p 0$, the optimal bandwidth $h^*(\hat{\sigma}^2)$ for $\hat{\sigma}^2(x)$ that balances off the bias and variance terms is well defined and given by

$$h^*(\hat{\sigma}^2) = \left[c(\pi)^2 \iota(K^2)/(48\iota(\iota K_1^2))\right]^{1/4} (\sigma^2/|\sigma'|)^{1/2}(x)\delta^{1/4}.$$

The asymptotics of $\hat{\sigma}^2(x)$ are given by the term including Z_p or Z_q, respectively, in $\hat{\sigma}_p^2(x)$ or $\hat{\sigma}_q^2(x)$, depending upon whether $h \gg \delta^{1/4}$ or $h \ll \delta^{1/4}$.[11]

(e) It is very informative to compare the asymptotics of our estimator $\hat{\sigma}^2(x)$ in Theorem 3.1 with those of the local constant threshold estimator proposed by Mancini and Renò (2011). Our asymptotics here are more comparable to those in Thm. 5.1 of Park and Wang (2018). The asymptotics in Mancini and Renò (2011) do not include the bias term, since they set $\delta^{3/8} \ll h \ll \delta^{1/3}$ with T fixed, in which case the bias term is dominated by the variance term, which is shown in Park and Wang (2018). Our estimator has the same leading bias terms as the local constant threshold estimator. Moreover, the variance term of our estimator is just the constant $c(\pi)$ multiple of that of the local constant threshold estimator. The variance of our estimator is therefore $c(\pi)^2$-times bigger than the local constant threshold estimator.

3.2. Local Linear Estimator

Our local linear estimator for $\sigma^2(x)$ is given by

$$\tilde{\sigma}^2(x) = \frac{P_T(K,\sigma^2)Q_T(\iota^2 K) - P_T(\iota K,\sigma^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2 K) - Q_T(\iota K)^2},$$

where $P_T(\cdot,\sigma^2)$ and $Q_T(\cdot)$ are defined similarly as $P_T(K,\sigma^2)$ and $Q_T(K)$ introduced in the definition of the local constant estimator $\hat{\sigma}^2(x)$ in the previous section with K substituted by various functions. Its estimation error can be decomposed as

$$\tilde{\sigma}^2(x) - \sigma^2(x) = \tilde{\sigma}_p^2(x) + \tilde{\sigma}_q^2(x) + \tilde{\sigma}_r^2(x),$$

where $\tilde{\sigma}_p^2(x)$, $\tilde{\sigma}_q^2(x)$ and $\tilde{\sigma}_r^2(x)$ are, respectively, the bias, variance, and error terms of $\tilde{\sigma}^2(x)$ that are given by

$$\tilde{\sigma}_p^2(x) = \frac{N_T(K,\sigma^2)Q_T(\iota^2 K) - N_T(\iota K,\sigma^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2 K) - Q_T(\iota K)^2},$$

$$\tilde{\sigma}_q^2(x) = \frac{M_T(K,\sigma^2)Q_T(\iota^2 K) - M_T(\iota K,\sigma^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2 K) - Q_T(\iota K)^2},$$

where $N_T(\iota K,\sigma^2)$ and $M_T(\iota K,\sigma^2)$ are defined similarly as $N_T(K,\sigma^2)$ and $M_T(K,\sigma^2)$, respectively, with K replaced by ιK, and

$$\tilde{\sigma}_r^2(x) = \frac{R_T(K,\sigma^2)Q_T(\iota^2 K) - R_T(\iota K,\sigma^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2 K) - Q_T(\iota K)^2},$$

where $R_T(\iota K,\sigma^2)$ is defined as $R_T(K,\sigma^2)$ using ιK instead of K.

[11] Here and elsewhere in the paper, $p \ll q$ signifies $p = o(q)$, that is, p is negligible asymptotically relative to q.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

THEOREM 3.2. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$\tilde{\sigma}_p^2(x) = \iota_2(K)\sigma^{2\prime\prime}(x)h^2/2 + o_p(h^2),$$

and

$$(h\ell(T,x)/\delta)^{1/2}\tilde{\sigma}_q^2(x) \to_d c(\pi)\iota(K^2)^{1/2}\sigma^2(x)Z,$$

where $c(\pi) = (\pi^2/4 + \pi - 3)^{1/2}$, *Z is a standard normal random variate independent of* $\ell(T,x)$. *Moreover,*

$$\tilde{\sigma}_r^2(x) = O_p(\delta^{1/2}),$$

which becomes asymptotically negligible if h is chosen appropriately.

Remark 3.2. (a) The consistency of $\tilde{\sigma}^2(x)$ requires no additional assumption as for the consistency of $\hat{\sigma}^2(x)$.

(b) For the asymptotic (mixed) normality of $\tilde{\sigma}^2(x)$, it is necessary to choose h so that $h\ell(T,x) \to_p 0$. Otherwise, the variance term $\tilde{\sigma}_q^2(x)$ is dominated by the error term $\tilde{\sigma}_r^2(x)$, which is the same as in the case of $\hat{\sigma}^2(x)$.

(c) If $h\ell(T,x) \to_p 0$, the optimal bandwidth that balances off the bias and variance terms is well defined and given by

$$h^*(\tilde{\sigma}^2) = \left[c(\pi)^2\iota(K^2)/\iota_2(K)^2\right]^{1/5}(\sigma^2/\sigma^{2\prime\prime})^{2/5}(x)(\delta/\ell(T,x))^{1/5}.$$

Though $h^*(\tilde{\sigma}^2)$ involves $\ell(T,x)$, we may still interpret it as the bandwidth minimizing the asymptotic mean squared error as in the standard case, which is because the limit random variable Z is independent of $\ell(T,x)$. If X is stationary, we have $\ell(T,x)/T \to_{a.s.} w(x)$, where w is the invariant density of X, the optimal bandwidth $h^*(\tilde{\sigma}^2)$ may be defined as

$$h^*(\tilde{\sigma}^2) = \left[c(\pi)^2\iota(K^2)/\iota_2(K)^2\right]^{1/5}(\sigma^2/\sigma^{2\prime\prime})^{2/5}(x)w(x)^{-1/5}n^{-1/5},$$

which can be written as $h^*(\tilde{\sigma}^2) = cn^{-1/5}$ with some constant c, similarly as in the standard case.

(d) The bias term of $\tilde{\sigma}^2(x)$ is drastically different from that of $\hat{\sigma}^2(x)$. In particular, if $h\ell(T,x) \to_p 0$, they become of different orders of magnitude: It is of order h^2 for $\tilde{\sigma}^2(x)$, while of order $h^{3/2}\ell(T,x)^{-1/2}$ for $\hat{\sigma}^2(x)$. Therefore, the bias term of $\tilde{\sigma}^2(x)$ is of a smaller order compared with that of $\hat{\sigma}^2(x)$. This is unusual.

(e) The asymptotics of our estimator $\tilde{\sigma}^2(x)$ are largely comparable to those for the local linear threshold estimator in Park and Wang (2018). The two estimators have the same leading bias term. Also, as for $\hat{\sigma}^2(x)$, the variance of our estimator is the constant $c(\pi)^2$ multiple of that of the local linear threshold estimator.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

4. ESTIMATION OF JUMP VOLATILITY

For the estimation of jump volatility $\tau(x)$, we deduce from Ito's formula that

$$\begin{aligned}(\Delta_i X)^2 &= 2\int_{(i-1)\delta}^{i\delta}(X_{t-}-X_{(i-1)\delta})dX_t+\int_{(i-1)\delta}^{i\delta}d[X]_t\\ &=\int_{(i-1)\delta}^{i\delta}\sigma^2(X_t)dt+\int_{(i-1)\delta}^{i\delta}\int_{\mathbb{R}}\tau^2(X_{t-})z^2\Lambda(dt,dz)\\ &\quad+2\int_{(i-1)\delta}^{i\delta}(X_{t-}-X_{(i-1)\delta})dX_t\\ &=\delta\sigma^2(X_{(i-1)\delta})+\delta\tau^2(X_{(i-1)\delta})+\int_{(i-1)\delta}^{i\delta}\int_{\mathbb{R}}\tau^2(X_{t-})z^2\Gamma(dt,dz)\\ &\quad+2\int_{(i-1)\delta}^{i\delta}(X_{t-}-X_{(i-1)\delta})dX_t\\ &\quad+\int_{(i-1)\delta}^{i\delta}\left[\sigma^2(X_t)-\sigma^2(X_{(i-1)\delta})\right]dt+\int_{(i-1)\delta}^{i\delta}\left[\tau^2(X_t)-\tau^2(X_{(i-1)\delta})\right]dt,\end{aligned}$$

from which it follows that

$$\begin{aligned}&(\Delta_i X)^2-\frac{\pi}{2}|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X|\le\delta^\beta,|\Delta_{i+1}X|\le\delta^\beta\}\\ &=\delta\tau^2(X_{(i-1)\delta})+\int_{(i-1)\delta}^{i\delta}\int_{\mathbb{R}}\tau^2(X_{t-})z^2\Gamma(dt,dz)+2\int_{(i-1)\delta}^{i\delta}(X_{t-}-X_{(i-1)\delta})dX_t\\ &\quad-\frac{\pi}{2}\left[|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X|\le\delta^\beta,|\Delta_{i+1}X|\le\delta^\beta\}-\frac{2\delta}{\pi}\sigma^2(X_{(i-1)\delta})\right]\\ &\quad+\int_{(i-1)\delta}^{i\delta}\left[\sigma^2(X_t)-\sigma^2(X_{(i-1)\delta})\right]dt+\int_{(i-1)\delta}^{i\delta}\left[\tau^2(X_t)-\tau^2(X_{(i-1)\delta})\right]dt. \qquad \textbf{(4.1)}\end{aligned}$$

We may only consider the first two terms on the righthand side of (4.1). As will be shown, all other terms are asymptotically negligible under appropriate conditions. Furthermore, since the second term is a martingale difference sequence, we may estimate the jump variance $\tau^2(x)$ by the standard nonparametric kernel regression of $\left((1/\delta)[(\Delta_i X)^2-(\pi/2)|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X|\le\delta^\beta,|\Delta_{i+1}X|\le\delta^\beta\}]\right)$ on $(X_{(i-1)\delta})$ for $i=1,\ldots,n$.[12]

4.1. Local Constant Estimator

Our local constant estimator for $\tau^2(x)$ is given by

$$\hat{\tau}^2(x)=\frac{P_T(K,\tau^2)}{Q_T(K)},$$

[12]For estimating jump volatility function $\tau^2(\cdot)$, Park and Wang (2018) consider the standard nonparametric kernel regression of $((1/\delta)(\Delta_i X)^2 1\{|\Delta_i X|>\delta^\beta\})$ on $(X_{(i-1)\delta})$ for $i=1,\ldots,n$.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

where

$$P_T(K,\tau^2) = \frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\big[(\Delta_i X)^2 - (\pi/2)|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1}X| \le \delta^\beta\}\big],$$

and $Q_T(K)$ is as defined previously in Section 3.1. It yields the estimation error we decompose as

$$\hat{\tau}^2(x) - \tau^2(x) = \hat{\tau}_p^2(x) + \hat{\tau}_q^2(x) + \hat{\tau}_r^2(x),$$

where $\hat{\tau}_p^2(x)$, $\hat{\tau}_q^2(x)$ and $\hat{\tau}_r^2(x)$ are, respectively, the bias, variance, and error terms of $\hat{\tau}^2(x)$ defined as

$$\hat{\tau}_p^2(x) = \frac{N_T(K,\tau^2)}{Q_T(K)}, \qquad \hat{\tau}_q^2(x) = \frac{M_T(K,\tau^2)}{Q_T(K)}$$

with

$$N_T(K,\tau^2) = \frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\int_{(i-1)\delta}^{i\delta}\big[\tau^2(X_t)-\tau^2(x)\big]dt,$$

$$M_T(K,\tau^2) = \frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\int_{(i-1)\delta}^{i\delta}\int_{\mathbb{R}} \tau^2(X_{t-})z^2\Gamma(dt,dz),$$

and

$$\hat{\tau}_r^2(x) = \frac{R_T(K,\tau^2)}{Q_T(K)} = \frac{A_T(K)+B_T(K)+C_T(K)}{Q_T(K)}$$

with

$$A_T(K) = 2\frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\int_{(i-1)\delta}^{i\delta}(X_{t-}-X_{(i-1)\delta})dX_t,$$

$$B_T(K) = \frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\int_{(i-1)\delta}^{i\delta}\big[\sigma^2(X_t)-\sigma^2(X_{(i-1)\delta})\big]dt,$$

$$C_T(K) = -\frac{\pi}{2}\frac{1}{h}\sum_{i=1}^{n-1} K\Big(\frac{X_{(i-1)\delta}-x}{h}\Big)\left[|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X| \le \delta^\beta, |\Delta_{i+1}X| \le \delta^\beta\} - \frac{2\delta}{\pi}\sigma^2(X_{(i-1)\delta})\right].$$

The decomposition can be readily obtained from (4.1).

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

THEOREM 4.1. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$\hat{\tau}_p^2(x) = \frac{h^2 \iota_2(K)}{2}\left[\left(4\tau^{2\prime}\left(\frac{\mu}{\sigma^2} - \frac{\sigma'}{\sigma}\right) + \tau^{2\prime\prime}\right)(x) + \frac{4\tau^{2\prime}(x)\xi(T,x)}{\sigma^2(x)\ell(T,x)}\right] + o_p(h^2) + O_p(h^{3/2}\ell(T,x)^{-1/2}),$$

and

$$[h\ell(T,x)]^{1/2}\hat{\tau}_q^2(x) \to_d \iota(K^2)^{1/2}\kappa^2\tau^2(x)Z,$$

where $\kappa^4 = \int_{\mathbb{R}} z^4\lambda(dz)$ *and* Z *is a standard normal random variate independent of* $\ell(T,x)$. *Moreover,*

$$\hat{\tau}_r^2(x) = O_p(\delta^{1/2}) + O_p(\delta^{1/2}(h\ell(T,x))^{-1/2}) + o_p(h^2) + o_p(h^{1/2}\ell(T,x)^{-1/2}),$$

which becomes asymptotically negligible if h *is chosen appropriately.*

Remark 4.1. (a) The consistency of $\hat{\tau}^2(x)$ requires $h\ell(T,x) \to_p \infty$, for which it is necessary to have $T \to \infty$. This is in sharp contrast to the estimators of diffusive volatility, which become consistent even if T is fixed. To estimate jump volatility consistently, the number of jumps should increase to infinity and the information on jumps needs to be fully revealed in the limit.

(b) The bias term $\hat{\tau}_p^2(x)$ includes a random term $\xi(T,x)/\ell(T,x)$. However, the term converges a.s. to $\phi(x)/m(x)$ as $T \to \infty$. We may therefore replace $\xi(T,x)/\ell(T,x)$ with $\phi(x)/m(x)$, if we set $T \to \infty$. The variance term is asymptotically mixed normal, which becomes normal if X is stationary.

(c) If $h\ell(T,x) \to_p \infty$, the optimal bandwidth that balances off the bias and variance terms is well defined and given by

$$h^*(\hat{\tau}^2) = c(K)(\kappa\tau(x))^{4/5}\left[\tau^{2\prime\prime} + 4\tau'(\mu - \sigma\sigma' + \phi/m)/\sigma^2\right]^{-2/5}(x)\ell(T,x)^{-1/5}$$

with $c(K) = \iota(K^2)^{1/5}/\iota_2(K)^{2/5}$. As discussed, we may interpret $h^*(\hat{\tau}^2)$ as the bandwidth minimizing the asymptotic mean squared error even for nonstationary X, since $\ell(T,x)$ is independent of the limit normal random variate Z. For stationary X with invariant density w, we have

$$h^*(\hat{\tau}^2) = c(K)(\kappa\tau(x))^{4/5}\left[\tau^{2\prime\prime} + 4\tau'(\mu - \sigma\sigma' + \phi/m)/\sigma^2\right]^{-2/5}(x)w(x)^{-1/5}T^{-1/5},$$

which can be written as $h^*(\hat{\tau}^2) = cT^{-1/5}$ with some constant c, similar to discrete samples of size T.

(d) The asymptotics of both the bias and variance terms of $\hat{\tau}^2(x)$ are exactly the same as those of the local constant jump volatility threshold estimator in Thm. 5.3 of Park and Wang (2018). We only have some additional terms in the error term, which are asymptotically negligible compared to the asymptotic leading bias and variance terms.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

4.2. Local Linear Estimator

Our local linear estimator of $\tau^2(x)$ is given by

$$\tilde{\tau}^2(x) = \frac{P_T(K,\tau^2)Q_T(\iota^2K) - P_T(\iota K,\tau^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2K) - Q_T(\iota K)^2},$$

where $P_T(\iota K,\tau^2)$ is defined as $P_T(K,\tau^2)$ with K substituted by ιK. Similarly as before, we may decompose its estimation error as

$$\tilde{\tau}^2(x) - \tau^2(x) = \tilde{\tau}_p^2(x) + \tilde{\tau}_q^2(x) + \tilde{\tau}_r^2(x),$$

where

$$\tilde{\tau}_p^2(x) = \frac{N_T(K,\tau^2)Q_T(\iota^2K) - N_T(\iota K,\tau^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2K) - Q_T(\iota K)^2},$$

$$\tilde{\tau}_q^2(x) = \frac{M_T(K,\tau^2)Q_T(\iota^2K) - M_T(\iota K,\tau^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2K) - Q_T(\iota K)^2},$$

$$\tilde{\tau}_r^2(x) = \frac{R_T(K,\tau^2)Q_T(\iota^2K) - R_T(\iota K,\tau^2)Q_T(\iota K)}{Q_T(K)Q_T(\iota^2K) - Q_T(\iota K)^2}$$

with $N_T(\iota K,\tau^2)$, $M_T(\iota K,\tau^2)$, and $R_T(\iota K,\tau^2)$ defined, respectively, as $N_T(K,\tau^2)$, $M_T(K,\tau^2)$ and $R_T(K,\tau^2)$ using ιK instead of K.

THEOREM 4.2. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$\tilde{\tau}_p^2(x) = \iota_2(K)\tau^{2\prime\prime}(x)h^2/2 + o_p(h^2),$$

and

$$[h\ell(T,x)]^{1/2}\tilde{\tau}_q^2(x) \to_d \iota(K^2)^{1/2}\kappa^2\tau^2(x)Z,$$

where $\kappa^4 = \int_{\mathbb{R}} z^4\lambda(dz)$ *and* Z *is a standard normal random variate independent of* $\ell(T,x)$. *Moreover,*

$$\tilde{\tau}_r^2(x) = O_p(\delta^{1/2}) + O_p(\delta^{1/2}(h\ell(T,x))^{-1/2}) + o_p(h^2) + o_p(h^{1/2}\ell(T,x)^{-1/2}),$$

which becomes asymptotically negligible if h *is chosen appropriately.*

Remark 4.2. (a) As for $\hat{\tau}^2(x)$, the consistency of $\tilde{\tau}^2(x)$ also requires $h\ell(T,x) \to_p \infty$ and it is necessary to let $T \to \infty$.

(b) In contrast with $\hat{\tau}^2(x)$, the bias term $\tilde{\tau}_p^2(x)$ of $\tilde{\tau}^2(x)$ does not include any random term. Unlike in the estimation of diffusive volatility, however, the bias terms of $\hat{\tau}^2(x)$ and $\tilde{\tau}^2(x)$ have leading terms of the same order.

(c) If $h\ell(T,x) \to_p \infty$, the optimal bandwidth that balances off the bias and variance terms is well defined and given by

$$h^*(\tilde{\tau}^2) = c(K)(\kappa^2\tau^2/\tau^{2\prime\prime})^{2/5}(x)\ell(T,x)^{-1/5}$$

with $c(K) = \iota(K^2)^{1/5}/\iota_2(K)^{2/5}$. As discussed, it minimizes the asymptotic mean squared error for nonstationary as well as stationary X. As for $\hat{\tau}^2(x)$, the optimal

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

bandwidth of $\tilde{\tau}^2(x)$ can be written as $h^*(\tilde{\tau}^2) = cT^{-1/5}$ with some c, if X is stationary.

(d) The asymptotic leading bias and variance terms of $\tilde{\tau}^2(x)$ are the same as those of the local linear jump volatility threshold estimator in Thm. 5.4 of Park and Wang (2018). We have a few additional error terms that are of smaller magnitude, compared to the asymptotic leading bias and variance terms.

5. SIMULATION AND EMPIRICAL ILLUSTRATION

For our simulations and empirical studies, we consider our estimators for the diffusive and jump volatility functions in jump diffusions. In particular, the performances of our estimators are compared with those of the threshold estimators studied in Mancini and Renò (2011) and Park and Wang (2018), and also with those of the estimators using bipower increments, which simply use $|\Delta_i X||\Delta_{i+1}X|$ and $|\Delta_i X|^2-(\pi/2)|\Delta_i X||\Delta_{i+1}X|$, in place of $|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X| \le \varpi(\delta), |\Delta_{i+1}X| \le \varpi(\delta)\}$ and $|\Delta_i X|^2 - (\pi/2)|\Delta_i X||\Delta_{i+1}X|1\{|\Delta_i X| \le \varpi(\delta), |\Delta_{i+1}X| \le \varpi(\delta)\}$, respectively, for the estimation of diffusive and jump volatilities, with an appropriate choice of threshold level $\varpi(\delta)$.

To choose the threshold level $\varpi(\delta)$ for the threshold-bipower estimators and the threshold estimators, we use the approaches in Aït-Sahalia and Jacod (2009) and Mancini and Renò (2011). For the first approach, we let $\varpi_{a,b}(\delta) = a\sqrt{BV_T/T}\delta^b$ with tuning parameters $a > 0$ and $0 < b < 1/2$, where $BV_T = (\pi/2)\sum_{i=1}^{n-1}|\Delta_i X||\Delta_{i+1}X|$ is the bipower variation of X on the time span $[0,T]$, as suggested by Aït-Sahalia and Jacod (2009). The parameters, a and b, have no impact on the limiting behavior of the threshold estimators as long as $a > 0$ and $1/(4-2\alpha) < b < 1/2$, see, for example, Park and Wang (2018). In finite samples, however, the performance of the truncation estimators is expected to be dependent upon the choice of a and b. For the second approach, we use $\varpi_c(i) = c\hat{v}_i$ with tuning parameter c for each $\Delta_i X$, $i = 1, \ldots, n$, where $(\hat{v}_i)$ is the square root of the estimated filtered volatility based on an auxiliary GARCH(1,1) model.[13]

5.1. Simulation

Our simulations are based on three different jump diffusion models. The first model, Model 1, is given by

$$dX_t = \left[\alpha_1(\alpha_2 - X_t) - (\mu_\theta \nu)X_t\right]dt + \beta\sqrt{X_t}dW_t + X_{t-}(\exp(\theta Z_t) - 1)\,dN_t(\nu),$$

[13]The auxiliary GARCH(1,1) model used in Mancini and Renò (2011) is given by $\Delta_i X = \alpha + \sqrt{v_i^2}\varepsilon_i$ with $v_i^2 = \beta_0 + \beta_1(\Delta_{i-1}X)^2 + \beta_2 v_{i-1}^2$, where (ε_i) are standardized i.i.d. innovations, and α, β_0, β_1, β_2 are constants. The estimated filtered volatility $\hat{v}_i^2$ is set to be $\hat{\beta}_0 + \hat{\beta}_1(\Delta_{i-1}X)^2 + \hat{\beta}_2\hat{v}_{i-1}^2$ with initial value $\hat{v}_0^2$ as the unconditional sample variance of $(X_{i\delta})$, and $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ as the averages of estimates for β_0, β_1, and β_2, respectively, over a number of long simulations.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

where Z is i.i.d. standard normal process, $N(\nu)$ is Poisson process with intensity ν, and $\mu_\theta = \mathbb{E}(\exp(\theta Z_t) - 1) = \exp(\theta^2/2) - 1$, with parameter values given by

$$(\alpha_1, \alpha_2, \beta, \nu, \theta) = (0.2338, 0.0508, \exp(-7.1707/2), 5.3967, 0.0263).$$

As usual, we set Z and $N(\nu)$ to be independent of W. Model 1 is a modified version of the SV_1J model estimated in Andersen, Benzoni, and Lund (2004) using weekly 3-month Treasury Bills rate data for the sample period 01/06/1954 to 06/28/2000. We set β to be the estimated mean of the volatility process of the SV_1J model.[14]

The second and third models, referred to as Models 2 and 3, are specified as

$$dX_t = \alpha X_t(\beta + X_t^2)^{\kappa/2-1}dt + \sqrt{\gamma}\left(\beta + X_{t-}^2\right)^{\kappa/4}(dW_t + \theta Z_t dN_t(\nu)),$$

where Z and $N(\nu)$ are specified as in Model 1, and the parameter values for $(\alpha, \beta, \gamma, \kappa, \nu, \theta)$ are given by

$$\left(-1.6289 \times 10^{-5}, 3.1809 \times 10^{-4}, 4.518 \times 10^{-3}, 0.1188, 47134.8, 6.986 \times 10^{-3}\right),$$
$$\left(-5.5511 \times 10^{-3}, 0.1678, 8.235 \times 10^{-5}, -5.3722, 12676.6, 0.0101\right)$$

for Models 2 and 3, respectively. Models 2 and 3 are the generalized Höpfner–Kutoyants model considered in Kim and Park (2017b), with the parameter values estimated by approximated MLE using the demedianed logarithm of exchange rates of GBP/USD and JPY/USD, respectively, at 1-min frequency from 01/01/2004 to 06/30/2015.

The simulation models, Model 1 and Models 2 and 3, may be rewritten as our jump diffusion model in (2.1) with

$$\tau(x) = x/w, \quad dJ_t = w(\exp(\theta Z_t) - 1)dN_t(\nu)$$

and

$$\tau(x) = \theta\sqrt{\gamma}(\beta + x^2)^{\kappa/4}/w, \quad dJ_t = wZ_t dN_t(\nu),$$

where $w > 0$ is a constant we set $1/w^2 = \nu\mathbb{E}(\exp(\theta Z_t) - 1)^2 = \nu(e^{2\theta^2} - 2e^{\theta^2/2} + 1)$ for given ν and θ and $1/w^2 = \nu$ for given ν, respectively, to normalize the corresponding Lévy measure as in Assumption 2.2. Under the required normalization, the Lévy measure has densities

$$\lambda(z) = \nu/(\theta z + \theta w)\phi\left(\left[\log(z/w + 1)\right]/\theta\right)$$

and

$$\lambda(z) = (\nu/w)\phi(z/w),$$

where ϕ is the standard normal density, respectively, for Model 1 and Models 2 and 3.

[14] For the simulation study in the Web Appendix, Mancini and Renò (2011) use a modified version of the SV_1J-SD model in Andersen et al. (2004). In the modified SV_1J-SD model, the parameter α_2 is set to follow a diffusion process instead of a constant as in our Model 1.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

In the simulation, we set the time span $T = 50$ with the sampling interval $\delta = 1/250$ for Model 1, and $T = 5$ with $\delta = 1/(250 \times 24 \times 60)$ for Models 2 and 3. These sampling intervals for Model 1 and Models 2 and 3 correspond to the daily and 1-minute observations. We use the standard normal kernel function and the optimal bandwidths as given in our paper and Park and Wang (2018), assuming the knowledge of the full specification of our simulation models. This is to focus on the relative evaluation of our approach and the threshold approach, net of the effect from the bandwidth choice on their performance. For the threshold-bipower and threshold estimators, we set the threshold level given by Aït-Sahalia and Jacod (2009) with $a = 5$ and $b = 0.49$, that is, $\varpi_{5,0.49}(\delta)$. Moreover, for estimation of diffusive volatility function, we also consider the threshold estimator with the threshold level given by an auxiliary GARCH(1,1) model with $c = 3$, that is, $\varpi_3(i)$, and the bandwidth $h = 3\varsigma n^{-1/5}$ with the sample standard deviation ς of $(X_{i\delta})$, as in Mancini and Renò (2011).[15]

For the comparison between our estimators using truncated bipower increments and the existing threshold estimators, as well as the bipower estimators, we consider the relative biases and standard errors of diffusive and jump volatilities averaged over their ordinates x_j's for $1 \le j \le M$ equal-spaced within $\underline{x} \le x_j \le \bar{x}$ for appropriately chosen lower and upper boundaries $\underline{x}$ and $\bar{x}$ with the number of iterations $N = 2,000$.[16] More precisely, if $\hat{v}_i^2(x)$ is an estimate for the diffusive or jump volatility v^2 at x in iteration i for $1 \le i \le N$, then the averaged relative bias and standard error are obtained as

$$\frac{1}{M}\sum_{j=1}^{M}\left(\bar{\hat{v}}^2(x_j)\Big/v^2(x_j) - 1\right) \quad \text{and} \quad \frac{1}{M}\sum_{j=1}^{M}\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{v}_i^2(x_j) - \bar{\hat{v}}^2(x_j)\right)^2}\Big/v^2(x_j)\right),$$

where $\bar{\hat{v}}^2(x_j)$ is the simulated mean of $\hat{v}_i^2(x_j)$ at point x_j, that is, $\bar{\hat{v}}^2(x_j) = N^{-1}\sum_{i=1}^{N}\hat{v}_i^2(x_j)$.

Figures 5.1 and 5.2 present the biases and standard errors of diffusive and jump volatility function estimators for various values of θ. The parameter θ determines the size of jumps and also the magnitude of jump volatility relative to that of diffusive volatility in our simulation models: As θ increases, both the jump size and the relative magnitude of jump volatility increase. In our simulations, we consider the ranges of values for θ, which appear to be realistic and relevant in practical applications. Overall, the relative performances of the three estimators - including the estimator using truncated bipower increments, threshold estimator and bipower estimator - vary across different models and different ranges of values for θ. None of the three estimators uniformly dominates the other two. The relative performance of the threshold estimator depends crucially on the threshold level.

[15] Mancini and Renò (2011) consider the threshold estimator for the state dependent jump intensity function instead of jump volatility function. Therefore, for the purpose of comparison, we only consider their estimate of diffusive volatility function.

[16] The bounds are set to be $[0.03, 0.06]$ for Model 1, and $[0.005, 0.04]$ for Models 2 and 3.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

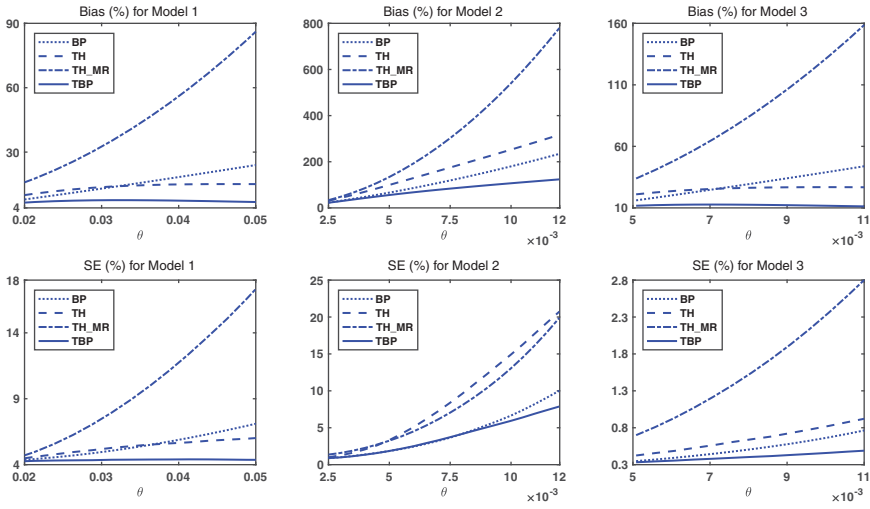


FIGURE 5.1. Biases and standard errors of diffusive volatility function estimators.
Notes: The relative biases (upper panel) and standard errors (lower panel) of diffusive volatility function estimators are presented as functions of θ. The left, middle, right panels are for Models 1, 2, and 3, respectively. The dotted, dashed, solid, and dash-dot lines are, respectively, for the bipower estimator (BP), the threshold estimator with threshold given by Aït-Sahalia and Jacod (2009) (TH), our threshold-bipower estimator (TBP), and the threshold estimator with threshold given by Mancini and Renò (2011) (TH_MR).

With the threshold level given by Aït-Sahalia and Jacod (2009), the threshold estimator generally performs reasonably well. However, if the threshold level given by Mancini and Renò (2011) is used, the threshold estimator is less satisfactory, as shown in Figure 5.1.

For large values of θ, the relative biases of the threshold estimator are in general smaller than those of the estimator using bipower increments in most cases. This is well expected since, as θ increases, the size of jumps increases and the threshold approach becomes more effective in detecting jump increments than the approach based on the bipower increments. Conversely, for small values of θ, the threshold approach does not work well in discriminating diffusive and jump increments, which makes the biases of the estimators using bipower increments smaller than those of the threshold estimators in most cases. On the other hand, in terms of standard errors, the estimator using bipower increments generally outperforms the threshold estimator. This can be more clearly seen in Figure 5.2, where we only consider the threshold estimator with threshold level given by Aït-Sahalia and Jacod (2009).

Our threshold-bipower estimator performs very well. It tends to behave more like the estimator using bipower increments, in case θ is small and the threshold approach is not supposed to work properly, whereas it performs similar to the threshold estimator in case θ is large in which case the threshold approach becomes

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

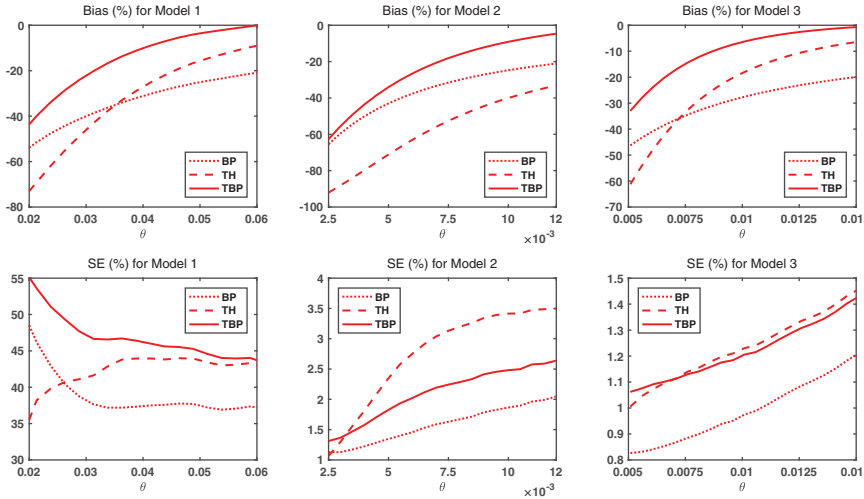


FIGURE 5.2. Biases and standard errors of jump volatility function estimators.
Notes: The relative biases (upper panel) and standard errors (lower panel) of jump volatility function estimators are presented as functions of θ. The left, middle, right panels are for Models 1, 2, and 3, respectively. The dotted, dashed, and solid lines are, respectively, for the bipower estimator (BP), the threshold estimator with threshold given by Aït-Sahalia and Jacod (2009) (TH), and our threshold-bipower estimator (TBP).

more effective and performs well. This is expected from our construction of the threshold-bipower estimator. For our simulation models with practically relevant values of θ, the threshold-bipower estimator for diffusive volatility performs best both in terms of relative biases and standard errors. The threshold-bipower estimator yields the smallest biases also for jump volatility, though its relative standard errors are generally larger than those of the estimator using bipower increments.

5.2. Empirical Illustration

For our empirical illustrations, we use the same data set as that used in Mancini and Renò (2011), which includes two interest rate time series, the 7-day Eurodollar deposit rate and 3-month Treasury Bill rate, at daily frequency from June 1, 1973 to February 24, 1995, with a total of 5,505 observations. Both of the interest rates have been used widely as proxies for the unobserved short rate or instantaneous rate. The reader is referred to Mancini and Renò (2011) and the references cited there for more discussions. Though they move very close to each other, the two interest rates show distinct time series characteristics. The 7-day Eurodollar deposit rate frequently contains large spikes, which are believed to be induced by calendar and liquidity effects typical for interest rate instruments with very short maturities.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

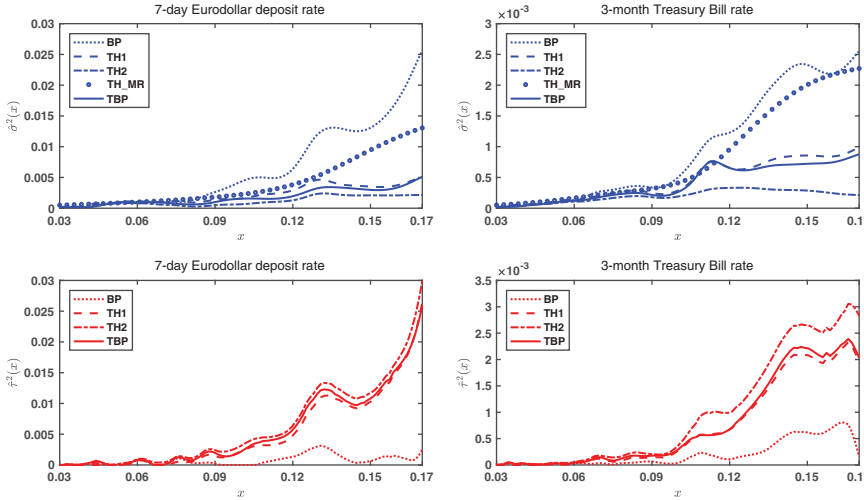


FIGURE 5.3. Estimated volatility functions.

Notes: Presented are the estimates of diffusive (upper panel) and jump (lower panel) volatility functions. The left and right panels are for the 7-day Eurodollar deposit rate and the 3-month Treasury Bill rate, respectively. The solid lines are the estimators using truncated bipower increments with threshold level $\varpi_{5,0.49}(\delta)$ (TBP), whereas the dashed, dash-dot and circle-marker lines are the threshold estimators with threshold levels $\varpi_{5,0.49}(\delta)$ (TH1), $\varpi_{3,0.49}(\delta)$ (TH2), and $\varpi_3(i)$ (TH_MR), respectively, and the dotted lines are the bipower estimators (BP).

On the other hand, the 3-month Treasury Bill rate has not shown any frequent occurrence of sizeable jumps.

The estimated diffusive and jump volatility functions of the two interest rate time series are presented in Figure 5.3. The estimates of volatility functions are obtained at each of 100 equispaced points over the range $[0.03, 0.17]$, which contains 96.2% and 96.86% of the observations of the 7-day Eurodollar deposit rate and 3-month Treasury Bill rate time series, respectively.[17] For comparisons, we consider five different estimators: our estimator using truncated bipower increments with $\varpi_{5,0.49}(\delta)$, the two threshold estimators relying on the threshold levels $\varpi_{3,0.49}(\delta)$ and $\varpi_{5,0.49}(\delta)$, the bipower estimator, as well as the threshold estimator with the threshold levels given by $\varpi_3(i)$. For the first four estimators, we use their respective optimal bandwidths at each of 100 equispaced points, obtained assuming that Model 1 in our simulation study is the true underlying model. For the last estimator, we employ the same bandwidth as the one used in Mancini and Renò (2011).

Figure 5.3 shows that the five estimators we consider here provide quite distinctive volatilities, especially as the level of interest rate increases. In general, threshold estimators are sensitive to the choice of a threshold. For diffusive

[17] For their empirical study, Mancini and Renò (2011) consider the range $[0.0175, 0.1625]$, which is similar to ours.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

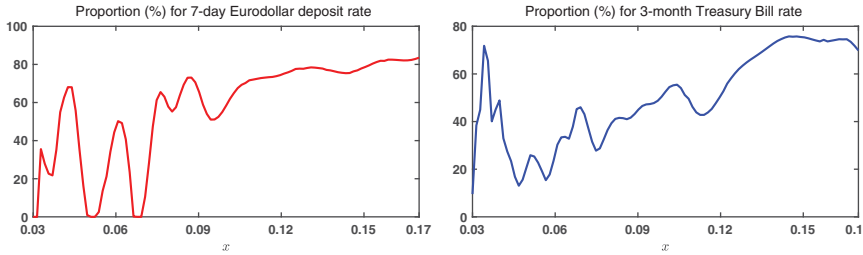


FIGURE 5.4. Proportion of jump volatility.

Notes: Presented are estimated proportions of total volatility contributed by jumps, that is, $(\hat{\tau}^2/\hat{\omega}^2)(x) \times 100\%$, using threshold-bipower estimators. The left and right panels are for the 7-day Eurodollar deposit rate and the 3-month Treasury Bill rate, respectively.

volatilities, the estimator using bipower increments yields the largest value of all five estimators, especially when the level of interest rate is high. In particular, for the 7-day Eurodollar deposit rate, all other estimators, including the threshold-bipower estimator, produce volatilities that are significantly smaller than the estimator using bipower increments. We believe that this is due to a positive bias existing in the estimator using bipower increments, which we observed from our simulations and identified as an adverse effect from the presence of large jumps in underlying jump diffusion models. For the volatility estimates of the 3-month Treasury Bill rate, in contrast, all other estimates based on the threshold approach are more or less evenly spread, though the estimates using bipower increments are largest at all levels of interest rate. For the diffusive volatility function, the estimator proposed by Mancini and Renò (2011) yields the values closest to those from the estimator using bipower increments. Note that jump and diffusive volatilities may be regarded as mirror images of each other, since the former is obtained as the residual from the total volatility net of the latter. The comparison among different estimates of jump volatility is thus essentially identical to those of diffusive volatility. Finally, as expected, the magnitudes of volatilities for the 7-day Eurodollar deposit rate are significantly larger than those for the 3-month Treasury Bill rate (Figure 5.4).

6. CONCLUSION

This paper proposes nonparametric estimations of volatility functions of jump diffusion models using truncated bipower increments. We establish asymptotics for the local constant and local linear estimators of diffusive and jump volatility functions. Our asymptotics are applicable for nonstationary, as well as stationary, jump diffusions, and are fine enough to derive the optimal bandwidths minimizing the asymptotic mean squared errors of all estimators considered in the paper.

Our threshold bipower estimators and their asymptotics are comparable to the threshold estimators using truncated squared increments proposed by Mancini

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

and Renò (2011) and their asymptotics developed in Park and Wang (2018). For estimating diffusive volatility functions, our threshold bipower estimators share the same asymptotic biases as, while have asymptotic variances that are 1.142^2 times larger than, the corresponding threshold estimators using truncated squared increments. On the other hand, for estimating jump volatility functions, our threshold bipower estimators have the same asymptotic biases and variances as those corresponding threshold estimators using truncated squared increments. The simulation study shows that our threshold bipower estimators outperform the threshold estimators using truncated squared increments in terms of finite sample biases and variances for estimating diffusive volatility functions, as well as finite sample biases for estimating jump volatility functions, across all of the three simulation models and realistic ranges of parameter values. In terms of finite sample variances for estimating jump volatility functions, the relative performance results of our threshold bipower estimator are mixed across different simulation models and parameter values, compared to the threshold estimator using truncated squared increments.

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be found at: https://doi.org/10.1017/S0266466620000389

REFERENCES

Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.

Aït-Sahalia, Y. & J. Jacod (2009) Testing for jumps in a discretely observed process. *Annals of Statistics* 37, 184–222.

Aït-Sahalia, Y. & P.A. Mykland (2003) The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71, 483–549.

Aït-Sahalia, Y. & P.A. Mykland (2004) Estimators of diffusions with randomly spaced discrete observations, A general theory. *Annals of Statistics* 32, 2186–2222.

Aït-Sahalia, Y. & J.Y. Park (2016) Bandwidth selection & asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *Journal of Econometrics* 192, 119–138.

Andersen, T.G., L. Benzoni, & J. Lund (2004) Stochastic volatility, mean drift, and jumps in the short-term interest rate. Working Paper, Northwestern University.

Andersen, T.G., D. Dobrev, & E. Schaumburg (2012) Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75–93.

Bandi, F.M. & T.H. Nguyen (2003) On the functional estimation of jump-diffusion models. *Journal of Econometrics* 116, 293–328.

Bandi, F.M. & P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.

Bandi, F.M. & R. Renò (2018) Nonparametric stochastic volatility. *Econometric Theory* 34, 1207–1255.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, M. Podolskij, & N. Shephard (2005) A central limit theorem for realised power and bipower variations of continuous semimartingales. In Y. Kabanov

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

& R. Lipster, (eds.), *From Stochastic Calculus to Mathematical Finance, Festschrift for Albert Shiryaev*, pp. 33–68. Springer.
Barndorff-Nielsen, O.E. & N. Shephard (2003) Realized power variation and stochastic volatility. *Bernoulli* 9, 243–265.
Barndorff-Nielsen, O.E. & N. Shephard (2004) Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–37.
Barndorff-Nielsen, O.E. & N. Shephard (2006) Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1–30.
Barndorff-Nielsen, O.E., N. Shephard, & M. Winkel (2006) Limit theorems for multipower variation in the presence of jumps. *Stochastic Processes and their Applications* 5, 796–806.
Bollerslev, T. & V. Todorov (2011) Estimation of jump tails. *Econometrica* 79, 1727–1783.
Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes*. Springer.
Boudt, K., C. Croux, & S. Laurent (2011) Outlyingness weighted covariation. *Journal of Financial Econometrics* 9, 657–684.
Chang, J. & S.X. Chen (2011) On the approximate maximum likelihood estimation for diffusion processes. *Annals of Statistics* 39, 2820–2851.
Chen, S.X., J. Gao, & C.Y. Tang (2008) A test for model specification of diffusion processes. *Annals of Statistics* 36, 167–198.
Christensen, K., R. Oomen, & M. Podolskij (2010) Realised quantile-based estimation of the integrated variance. *Journal of Econometrics* 159, 74–98.
Corsi, F., D. Pirino, & R. Renò (2010) Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159, 276–288.
Fan, J. & C. Zhang (2003) A reexamination of diffusion estimators with applications to financial model validation. *Journal of the American Statistical Association* 98, 118–134.
Höpfner, R. & E. Löcherbach (2003) *Limit Theorems for Null Recurrent Markov Processes*. American Mathematical Society.
Jacod, J. (2008) Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications* 118, 517–559.
Jacod, J. (2012) Statistics and high frequency data. In M. Kessler, A. Lindner, & M. Sorensen (eds.), *Statistical Methods for Stochastic Differential Equations*, pp. 191–310. Chapman & Hall.
Jacod, J. & M. Rosenbaum (2013) Quarticity and other functionals of volatility: Efficient estimation. *Annals of Statistics* 41, 1462–1484.
Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes*. Springer.
Jeong, M. & J.Y. Park (2016) An asymptotic theory of jump diffusion model. Working Paper, Indiana University.
Kanaya, S. (2016) Uniform convergence rates of kernel-based nonparametric estimators for continuous time diffusion processes: A damping function approach. *Econometric Theory* 32, 1–41.
Kanaya, S. & D. Kristensen (2016) Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* 32, 861–916.
Kim, J. & J.Y. Park (2017a) Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics* 196, 27–54.
Kim, J. & J.Y. Park (2017b) Unit root, mean reversion and nonstationarity in Financial Time Series. Working Paper, Indiana University.
Kristensen, D. (2010) Nonparametric filtering of the realised spot volatility: A kernel-based approach. *Econometric Theory* 26, 60–93.
Mancini, C. (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36, 270–296.
Mancini, C. & R. Renò (2011) Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics* 160, 77–92.
Park, J.Y. & Wang, B. (2018) Nonparametric estimation of jump diffusion models. Working Paper, Indiana University.
Protter, P.E. (2005) *Stochastic Integration and Differential Equations*. Springer.

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

Renò, R. (2006) Nonparametric estimation of stochastic volatility models. *Economics Letters* 90, 390–395.

Renò, R. (2008) Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory* 24, 1174–1206.

Revuz, D. & M. Yor (1998) *Continuous Martingales and Brownian Motion*. Springer.

Touati, A. (1987) Théorèmes limites pour des processus de Markov récurrents. *Comptes Rendus de I'Académie des Sciences-Series I-Mathematics* 305, 841–844.

Ueltzhöfer, F.A.J. (2013) On non-parametric estimation of the Lévy kernel of Markov processes. *Stochastic Processes and their Applications* 123, 3663–3709.

Yu, C., Y. Fang, Z. Li, B. Zhang, & X. Zhao (2014) Non-parametric estimation of high-frequency spot volatility for Brownian semimartingale with jumps. *Journal of Time Series Analysis* 35, 572–591.

APPENDIX

The proofs of subsequential lemmas and theorems rely heavily on Park and Wang (2018), which will be referred to simply as PW. We use $\|\cdot\|$ to denote the euclidean norm, and $\|\cdot\|_\infty$ to signify the supremum norm of functions. For the process X specified in (2.1), we let $T(f) = \sup_{0\le t\le T} |f|(X_t)$ for any function $f : \mathcal{D} \to \mathbb{R}$. Moreover, we let $X = X^c + X^d$, where X^c and X^d represent, respectively, the continuous and jump parts of X. Subsequently, we write $X^c = X^c_1 + X^c_2$ with $dX^c_{1t} = \mu(X_t)dt$ and $dX^c_{2t} = \sigma(X_t)dW_t$, and define $J(\varepsilon)$ and $X^d(\varepsilon)$ as $dJ_t(\varepsilon) = \int_{|z|<\varepsilon} z\Lambda(dt,dz)$ and $dX^d_t = \int_{|z|<\varepsilon} z\tau(X_{t-})\Lambda(dt,dz)$ for any $\varepsilon > 0$, with the convention $J_t = J_t(\infty)$ and $X^d = X^d(\infty)$. We define $\ell[T,x] = \sigma^2(x)\ell(T,x)$ as the sojourn time measured by the continuous part of quadratic variation of X around x over the time interval $[0,T]$. For any stochastic process Z, we write $\Delta_i Z = Z_{i\delta} - Z_{(i-1)\delta}$ for $i = 1,\dots,n$, and $\Delta Z_t = Z_t - Z_{t-}$ for $t > 0$. The notations introduced here will be used repeatedly without any further reference.

In the sequel, we let $f : [-1,1] \to \mathbb{R}$ be nonnegative, bounded and twice continuously differentiable, and define $f_{x,h}(y) = f((y-x)/h)$. We denote by $(\mathcal{F}_t)$ the natural filtration of X, and write $\mathbb{E}_t(\cdot)$ to be $\mathbb{E}(\cdot|\mathcal{F}_t)$. We write κ_T for $\kappa(T)$ (in Assumption 2.1(g)) for notational brevity. For $a \in \mathbb{R}$, we use $\lfloor a \rfloor$ ($\lceil a \rceil$) to denote the biggest (smallest) integer b such that $b \le a (b \ge a)$. Moreover, we use "$A_T \le_p B_T$" to denote $A_T = O_p(B_T)$. We often use the notion of L-domination and Lenglart inequality in Def. 1.3.29 and Lem. 1.3.30 in p. 35 of Jacod and Shiryaev (2003). Note in particular that if A_T is L-dominated by B_T and B_T is predictable, then we have $A_T \le_p B_T$ due to Lenglart inequality. This will simply be referred to as Lenglart domination property. Finally, some of our subsequent asymptotics rely on the modulus of continuity of diffusion given by

$$\sup_{0\le s,t\le T}\sup_{|t-s|\le\delta}\left|\int_s^t g(X_u)dW_u\right| = O_p\left(\delta^{1/2}T(g)\sqrt{\log(T/\delta)}\right)$$

as $\delta \to 0$, which is established in Lem. B2 of Kim and Park (2017a).

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

A. USEFUL LEMMAS

Lemma A.1. *Let (i)* $\iota_1(f) = 0$, *(ii)* $g : \mathcal{D} \mapsto \mathbb{R}$ *be twice continuously differentiable, and (iii) Assumptions 2.1 and 2.3 hold. Then*

$$\begin{aligned}
&\frac{1}{h^3}\int_0^T f_{x,h}(X_t)[g(X_t) - g(x)]dt \\
&\quad = \frac{\iota_2(f)}{2}\left[4g'\left(\frac{\mu}{\sigma^2} - \frac{\sigma'}{\sigma}\right) + g''\right](x)\ell(T,x) + \frac{2\iota_2(f)g'(x)\xi(T,x)}{\sigma^2(x)} \\
&\qquad + \frac{2g'(x)}{h\sigma^2(x)}\left(\int_0^T \int_{\mathbb{R}} \int_{X_{t-}}^{X_{t-}+z\tau(X_{t-})} (\iota f_1)_{x,h}(u)du\Gamma(dt,dz)\right. \\
&\qquad \left. + \int_0^T [(\iota f_1)_{x,h}\sigma](X_t)dW_t\right) + o_p(\ell(T,x)).
\end{aligned}$$

Lemma A.2. *Let (i)* $f, \chi : [-1,1] \mapsto \mathbb{R}$ *and* $g, \varphi : \mathcal{D} \mapsto \mathbb{R}$ *be twice continuously differentiable, (ii) Assumptions 2.1, 2.3, 2.5, and 2.6 hold. If we let*

$$M_T = \frac{1}{\sqrt{h}}\sum_{i=1}^{n-1} (\chi_{x,h}\varphi)(X_{(i-1)\delta})\Delta_i W$$

$$N_T = \frac{1}{\sqrt{\delta h}}\sum_{i=1}^{n-1} (f_{x,h}g)(X_{(i-1)\delta})\left(|\Delta_i W||\Delta_{i+1} W| - \omega\delta\right)$$

with $\omega = 2/\pi$, *then*

$$(M_T, N_T) =_d \ell(T,x)^{1/2} Z(1 + o_p(1)),$$

where Z is a bivariate normal random vector independent of $\ell(T,x)$, *and has covariance matrix* $\Sigma = diag\,(\iota(\chi^2)\varphi^2(x), c(\pi)\iota(f^2)g^2(x))$ *with* $c(\pi) = 1 + 2\omega - 3\omega^2$.

Lemma A.3. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$R_T(K,\sigma^2) = O_p\left(\delta^{1/2}\ell(T,x)\right),$$

with $R_T(K,\sigma^2)$ *defined in Section 3.1.*

Lemma A.4. *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, and 3.1 hold. Then*

$$R_T(K,\tau^2) = O_p\left(\delta^{1/2}\ell(T,x) + \delta^{1/2}h^{-1/2}\ell(T,x)^{1/2}\right) + o_p\left(h^2\ell(T,x) + h^{1/2}\ell(T,x)^{1/2}\right),$$

with $R_T(K,\tau^2)$ *defined in Section 4.1.*

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

B. PROOFS OF THEOREMS

Proof of Theorem 3.1. For $\hat{\sigma}_p^2(x)$, we rewrite $N_T(K,\sigma^2)$ as

$$\begin{aligned} N_T(K,\sigma^2) &= \frac{1}{h}\int_0^T K_{x,h}(X_t)\left[\sigma^2(X_t)-\sigma^2(x)\right]dt \\ &\quad - \frac{1}{h}\sum_{i=1}^{n} K_{x,h}(X_{(i-1)\delta})\int_{(i-1)\delta}^{i\delta}\left[\sigma^2(X_t)-\sigma^2(X_{(i-1)\delta})\right]dt \\ &\quad + \frac{1}{h}\sum_{i=1}^{n}\int_{(i-1)\delta}^{i\delta}\left[K_{x,h}(X_{(i-1)\delta})-K_{x,h}(X_t)\right]\left[\sigma^2(X_t)-\sigma^2(x)\right]dt. \end{aligned} \tag{B.1}$$

We may apply Lemmas A.16 and A.17, respectively, in PW to the second and third terms in (B.1) and show that they are both of order $o_p(h^2\ell(T,x))$ under $\delta = o(h^3 \wedge T^{-6pq})$, and therefore asymptotically negligible.

By Lemma A.1 (with $f = K$ and $g = \sigma^2$), we may write the first term in (B.1) as

$$\begin{aligned} &\frac{h^2\iota_2(K)}{2}\left[\left(4\sigma^{2\prime}\left(\frac{\mu}{\sigma^2}-\frac{\sigma'}{\sigma}\right)+\sigma^{2\prime\prime}\right)(x)+\frac{8\sigma'(x)\xi(T,x)}{\sigma(x)\ell(T,x)}\right]\ell(T,x) \\ &\qquad + 4h(\sigma'/\sigma)(x)\,(A_T+B_T+C_T)+o_p(h^2\ell(T,x)), \end{aligned} \tag{B.2}$$

where

$$A_T = \sum_{i=1}^{n-1}[(\iota K_1)_{x,h}\sigma](X_{(i-1)\delta})\Delta_i W, \quad B_T = \int_0^T\int_{\mathbb{R}}\int_{X_{t-}}^{X_{t-}+z\tau(X_{t-})}(\iota K_1)_{x,h}(u)du\Gamma(dt,dz)$$

$$C_T = \sum_{i=1}^{n-1}\int_{(i-1)\delta}^{i\delta}\left([(\iota K_1)_{x,h}\sigma](X_t)-[(\iota K_1)_{x,h}\sigma](X_{(i-1)\delta})\right)dW_t.$$

For A_T, it follows from Lemma A.2 that

$$\sqrt{\frac{1}{h\ell(T,x)}}A_T \to_d \iota(\iota K_1^2)^{1/2}\sigma(x)Z_p, \tag{B.3}$$

where Z_p is a standard normal random variate independent of $\ell(T,x)$.

Next, the predictable quadratic variation of B_T is given by

$$\begin{aligned} \langle B\rangle_T &= \int_0^T\int_{\mathbb{R}}\left[\int_{X_t}^{X_t+z\tau(X_t)}(\iota K_1)_{x,h}(v)dv\right]^2\lambda(dz)dt \\ &\le 2\|\iota K_1\|h\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|\int_u^{u+z\tau(u)}(\iota K_1)_{x,h}(v)dv\right|\lambda(dz)\ell(T,u)du \\ &\le 2\|\iota K_1\|h^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\lambda_1|\left(\frac{x-u+hv}{\tau(u)}\right)|\iota K_1|(v)dv\ell(T,u)du = O_p(h^2\ell(T,x)), \end{aligned}$$

where the last equality follows from Lemma A.2 in PW. We therefore have

$$B_T = O_p(h\ell(T,x)^{1/2}), \tag{B.4}$$

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

since $B_T^2 = O_p(h^2\ell(T,x))$ by Lenglart domination property. For C_T, we have

$$\begin{aligned}
[C]_T &= \sum_{i=1}^{n-1} \int_{(i-1)\delta}^{i\delta} \left([(\iota K_1)_{x,h}\sigma](X_t) - [(\iota K_1)_{x,h}\sigma](X_{(i-1)\delta})\right)^2 dt \\
&\le \sum_{i=1}^{n-1} \int_{(i-1)\delta}^{i\delta} \left([(\iota K_1)_{x,h}\sigma]^2(X_t) - [(\iota K_1)_{x,h}\sigma]^2(X_{(i-1)\delta})\right) dt \\
&\quad - 2\sum_{i=1}^{n-1} [(\iota K_1)_{x,h}\sigma](X_{(i-1)\delta}) \int_{(i-1)\delta}^{i\delta} \left([(\iota K_1)_{x,h}\sigma](X_t) - [(\iota K_1)_{x,h}\sigma](X_{(i-1)\delta})\right) dt \\
&= o_p(h\ell(T,x)),
\end{aligned}$$

where the last equality follows from Lemma A.14 in PW with $f_{x,h}$ replaced by $[(\iota K_1)_{x,h}\sigma]^2$, and Lem. A.16 in PW with both $f_{x,h}$ and g replaced by $[(\iota K_1)_{x,h}\sigma]$, under $\delta = o(h^2 \wedge T^{-2pq})$. Therefore,

$$B_T = o_p(h^{1/2}\ell(T,x)^{1/2}). \tag{B.5}$$

Moreover, As shown in Lem. A.9 and A.14 in PW, we have

$$Q_T(K) = \ell(T,x)(1 + o_p(1)) \tag{B.6}$$

under $\delta = o_p(h^2)$. Then, the stated results for $\hat{\sigma}_p^2(x)$ follows immediately from (B.1) to (B.6). Moreover, we may apply Lemmas A.2, A.3, and (B.6) to deduce the stated results for $\hat{\sigma}_q^2(x)$ and $\hat{\sigma}_r^2(x)$, which completes the proof. ■

Proof of Theorem 3.2. Let A_T, B_T, and C_T be defined as in the proof of Theorem 3.1. For $\tilde{\sigma}_p^2(x)$, we may deduce similarly as $Q_T(K)$ that

$$Q_T(\iota^2 K) = \iota_2(K)\ell(T,x)(1 + o_p(1)), \tag{B.7}$$

which, together with (B.1) and (B.2), implies that

$$N_T(K,\sigma^2)Q_T(\iota^2 K) = F_T + G_T + o_p(h^2\ell(T,x)^2), \tag{B.8}$$

where

$$\begin{aligned}
F_T &= \frac{h^2\iota_2^2(K)}{2}\left[\left(4\sigma^{2\prime}\left(\frac{\mu}{\sigma^2} - \frac{\sigma'}{\sigma}\right) + \sigma^{2\prime\prime}\right)(x) + \frac{8\sigma'(x)\xi(T,x)}{\sigma(x)\ell(T,x)}\right]\ell(T,x)^2, \\
G_T &= 4h(\sigma'/\sigma)(x)(A_T + B_T + C_T)Q_T(\iota^2 K),
\end{aligned}$$

and $A_T + B_T + C_T = O_p(h^{1/2}\ell(T,x)^{1/2})$ as shown in the proof of Theorem 3.1.

We may decompose $N_T(\iota K, \sigma^2)$ similarly as $N_T(K,\sigma^2)$ in (B.1) and (B.2) and deduce that

$$\begin{aligned}
N_T(\iota K, \sigma^2) &= \frac{1}{h}\int_0^T (\iota K)_{x,h}(X_t)\left[\sigma^2(X_t) - \sigma^2(x)\right]dt + o_p(h^2\ell(T,x)) \\
&= \iota_2(K)\sigma^{2\prime}(x)h\ell(T,x) + o_p(h\ell(T,x))
\end{aligned} \tag{B.9}$$

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

under the condition of $\delta = o(h^3 \wedge T^{-6pq})$. Moreover, by (C.2)–(C.5) in the proof of Lemma A.1,

$$\begin{aligned} Q_T(\iota K) &= \int_{-\infty}^{\infty} (\iota K)(u)\ell(T,x+hu) + O_p\left(\delta h^{-2}\ell(T,x)\right) \\ &= 2\iota_2(K)h\left[\left(\frac{\mu}{\sigma^2} - \frac{\sigma'}{\sigma}\right)(x) + \frac{\xi(T,x)}{\sigma^2(x)\ell(T,x)}\right]\ell(T,x) \\ &\quad + 2\sigma^{-2}(x)(A_T + B_T + C_T) + o_p(h\ell(T,x)), \end{aligned} \tag{B.10}$$

under $\delta = o(h^3 \wedge T^{-6pq})$. Therefore, we have

$$N_T(\iota K, \sigma^2)Q_T(\iota K) = U_T + V_T + o_p(h^2\ell(T,x)^2), \tag{B.11}$$

where

$$\begin{aligned} U_T &= \frac{h^2\iota_2^2(K)}{2}\left[\left(4\sigma^{2\prime}\left(\frac{\mu}{\sigma^2} - \frac{\sigma'}{\sigma}\right)\right)(x) + \frac{8\sigma'(x)\xi(T,x)}{\sigma(x)\ell(T,x)}\right]\ell(T,x)^2 \\ V_T &= 2\sigma^{-2}(x)(A_T + B_T + C_T)N_T(\iota K, \sigma^2). \end{aligned}$$

As shown below, from (B.8) and (B.11), we may identify the bias terms and additional variance terms of $N_T(K,\sigma^2)Q_T(\iota^2 K)$ and $N_T(\iota K,\sigma^2)Q_T(\iota K)$.

For the bias terms, we have

$$F_T - U_T = \iota_2^2(K)(\sigma^{2\prime\prime}/2)(x)h^2\ell(T,x)^2. \tag{B.12}$$

For the additional variance terms, we write $G_T - V_T = (2/\sigma^2)(x)(A_T + B_T + C_T)D_T$, where

$$D_T = \sigma^{2\prime}(x)\delta\sum_{i=1}^{n}(\iota^2 K)_{x,h}(X_{(i-1)\delta}) - \frac{\delta}{h}\sum_{i=1}^{n}(\iota K)_{x,h}(X_{(i-1)\delta})\left(\sigma^2(X_{(i-1)\delta}) - \sigma^2(x)\right).$$

However, using analogous arguments as for (B.6), we have

$$\delta\sum_{i=1}^{n}(\iota^2 K)_{x,h}(X_{(i-1)\delta}) = \int_0^T (\iota^2 K)_{x,h}(X_t)dt + o_p(h^2\ell(T,x)).$$

Moreover, similar to $N_T(K,\sigma^2)$, we have

$$\begin{aligned} &\frac{\delta}{h}\sum_{i=1}^{n}(\iota K)_{x,h}(X_{(i-1)\delta})\left(\sigma^2(X_{(i-1)\delta}) - \sigma^2(x)\right) \\ &\quad = \frac{1}{h}\int_0^T (\iota K)_{x,h}(X_t)\left(\sigma^2(X_t) - \sigma^2(x)\right)dt + o_p(h^2\ell(T,x)) \end{aligned}$$

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

under $\delta = o(h^3 \wedge T^{-6pq})$, from which it follows that

$$\begin{aligned} D_T &\le \left| \frac{1}{h} \int_0^T (\iota K)_{x,h}(X_t) \left(\sigma^2(X_t) - \sigma^2(x) - \sigma^{2\prime}(x)(X_t - x) \right) dt + o_p(h^2 \ell(T,x)) \right| \\ &\le \left(\sup_{|u-x| \le h} \left| \frac{\sigma^{2\prime\prime}}{2} \right| (u) \right) \frac{1}{h} \int_0^T |\iota K|_{x,h}(X_t)(X_t - x)^2 dt + o_p(h^2 \ell(T,x)) \\ &\le_p h \int_0^T \left| \iota^3 K \right|_{x,h}(X_t) dt + o_p(h^2 \ell(T,x)) = O_p(h^2 \ell(T,x)), \end{aligned}$$

and we have

$$G_T - V_T = (2/\sigma^2)(x)(A_T + B_T + C_T) O_p(h^2 \ell(T,x)) = O_p\left(h^{5/2} \ell(T,x)^{3/2}\right). \tag{B.13}$$

Therefore, it follows immediately from (B.8), (B.11), (B.12), and (B.13) that

$$N_T(K, \sigma^2) Q_T(\iota^2 K) - N_T(\iota K, \sigma^2) Q_T(\iota K) = \iota_2^2(K)(\sigma^{2\prime\prime}/2)(x) h^2 \ell(T,x)^2 (1 + o_p(1)), \tag{B.14}$$

and we may deduce from (B.6), (B.7), and (B.10) that

$$Q_T(\iota^2 K) Q_T(K) - Q_T(\iota K)^2 = \iota_2(K) \ell(T,x)^2 (1 + o_p(1)). \tag{B.15}$$

The stated result for $\tilde{\sigma}_p^2(x)$ easily follows from (B.14) and (B.15).

For $\tilde{\sigma}_q^2(x)$, we note that $M_T(\iota K, \sigma^2) = O_p\left(\sqrt{\delta \ell(T,x)/h}\right)$, which follows exactly as for $M_T(K, \sigma^2)$ in the proof of Theorem 3.1, and deduce that

$$\sqrt{\frac{h\ell(T,x)}{\delta}} \frac{M_T(K,\sigma^2) Q_T(\iota^2 K) - M_T(\iota K, \sigma^2) Q_T(\iota K)}{Q_T(K) Q_T(\iota^2 K) - Q_T(\iota K)^2} = \sqrt{\frac{h}{\delta \ell(T,x)}} M_T(K, \sigma^2)(1 + o_p(1)),$$

from (B.7), (B.10), and (B.15), from which and Lemma A.2, the stated result follows immediately.

For $\tilde{\sigma}_r^2(x)$, in an analogous way as for $R_T(K, \sigma^2)$ in Lemma A.3, we may readily show that $R_T(\iota K, \sigma^2) = O_p(\delta^{1/2} \ell(T,x))$, from which, together with (B.7), (B.10), (B.15), and Lemma A.3, the stated result for $\tilde{\sigma}_r^2(x)$ follows immediately. ■

Proof of Theorem 4.1. The arguments for $\hat{\tau}_p^2(x)$ are analogous to those of $\hat{\sigma}_p^2(x)$ and therefore omitted. For $\hat{\tau}_q^2(x)$, we write $M_T(K, \tau^2) = U_T + V_T$, where

$$\begin{aligned} U_T &= \frac{1}{h} \sum_{i=1}^n \left(K_{x,h} \tau^2 \right)(X_{(i-1)\delta}) \int_{(i-1)\delta}^{i\delta} \int_{\mathbb{R}} z^2 \Gamma(dt, dz) \\ V_T &= \frac{1}{h} \sum_{i=1}^n K_{x,h}(X_{(i-1)\delta}) \int_{(i-1)\delta}^{i\delta} \int_{\mathbb{R}} z^2 \left(\tau^2(X_{t-}) - \tau^2(X_{(i-1)\delta}) \right) \Gamma(dt, dz). \end{aligned}$$

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press

For V_T, we may readily deduce that

$$
\begin{aligned}
V_T^2 &\le_p \frac{\iota_4(\lambda)}{h^2}\sum_{i=1}^{n} K_{x,h}^2(X_{(i-1)\delta})\int_{(i-1)\delta}^{i\delta}\left(\tau^2(X_t)-\tau^2(X_{(i-1)\delta})\right)^2 dt \\
&\le_p T([\tau^{2\prime}]^2)\frac{\iota_4(\lambda)}{h^2}\sum_{i=1}^{n} K_{x,h}^2(X_{(i-1)\delta})\int_{(i-1)\delta}^{i\delta}\int_{(i-1)\delta}^{t}(\sigma^2+\tau^2)(X_s)dsdt \\
&\le \delta T([\tau^{2\prime}]^2)\frac{\iota_4(\lambda)}{h}\left(\frac{1}{h}\sum_{i=1}^{n} K_{x,h}^2(X_{(i-1)\delta})\int_{(i-1)\delta}^{i\delta}\left(\sigma^2+\tau^2\right)(X_t)dt\right) \\
&= O_p\left(\delta T^{2pq}h^{-1}\ell(T,x)\right),
\end{aligned}
$$

where the first inequality in probability follows from Lenglart domination property, the second inequality from Mean value theorem and Lenglart domination property, and the last equality from Lem. A.16, A.14, and A.9 in PW under $\delta = o(h^2 \wedge T^{-6pq})$. Therefore, we have

$$
V_T = O_p\left(\delta^{1/2}T^{pq}h^{-1/2}\ell(T,x)^{1/2}\right). \tag{B.16}
$$

For U_T, we note that

$$
\langle U\rangle_T = \frac{\iota_4(\lambda)\delta}{h^2}\sum_{i=1}^{n}\left(K_{x,h}\tau^2\right)^2(X_{(i-1)\delta}) = \iota_4(\lambda)\iota(K^2)\tau^4(x)h^{-1}\ell(T,x)(1+o_p(1)).
$$

Then we may apply analogous arguments as for Lemma A.2 to show that

$$
\sqrt{\frac{h}{\ell(T,x)}}U_T \to_d \left[\iota(K^2)\iota_4(\lambda)\right]^{1/2}\tau^2(x)Z, \tag{B.17}
$$

where Z is a standard normal random variate independent of $\ell(T,x)$. The stated result for $\hat{\tau}_r^2(x)$ then follows from (B.6), (B.16), and (B.17). Finally, the stated result for $\hat{\tau}_r^2(x)$ readily follows from Lemma A.4 and (B.6). ■

Proof of Theorem 4.2. The derivations of the stated results for $\tilde{\tau}_p^2(x)$, $\tilde{\tau}_q^2(x)$ and $\tilde{\tau}_r^2(x)$ are completely analogously to those for $\tilde{\sigma}_p^2(x)$, $\tilde{\sigma}_q^2(x)$, and $\tilde{\sigma}_r^2(x)$, respectively, and therefore, omitted. ■

https://doi.org/10.1017/S0266466620000389 Published online by Cambridge University Press