SFB
823

Nonparametric drift estimation in a Lévy driven diffusion model

Benedikt Funke

Nr. 35/2015

Discussion Paper

Nonparametric Drift Estimation in a Lévy Driven Diffusion Model

Benedikt Funke[1]

Technical University of Dortmund, Department of Mathematics, Vogelpothsweg 87, 44227 Dortmund, Germany

Abstract

In this article, a pointwise nonparametric kernel based estimator for the drift function in a Lévy driven jump diffusion model is proposed. Under ergodicity and stationarity of the underlying process X, we derive asymptotic properties as consistency and asymptotic normality of the estimator. In addition, we propose a consistent estimator of the asymptotic variance. Moreover, we show that this approach is robust under microstructure noise by using the preaveraging approach proposed in Podolskij and Vetter (2006).

Keywords: Kernel Estimator, Jump Diffusion, Microstructure Noise

1. Introduction

In recent literature, it has been shown that several financial time series possess additional jump components; see Johannes (2004) and Das (2002) in the context of jump diffusion models. Hence, it is reasonable that models for this data should invoke an additive jump component, which is responsible for indicating macro- or microeconomic shocks during the evolution of processes of interest like stock prices or volatilities. Common models, for which nonparametric estimation procedures for the state-dependent drift and volatility have been conducted, are based on jump diffusion models which incorporate a Brownian part and an additive independent Poisson-driven jump part; see Bandi and Nguyen (2003) or Johannes (2004). Hence, this class of processes allows a finite number of jumps on every finite time interval. In contrast to those jump diffusion models, models with infinite activity processes are used in current literature to model economic values of interest. A fundamental paper in this direction was proposed by Carr et al. (2002), who suggested the use of the so-called CGMY model in order to model, for instance, processes with infinite activity and finite variation. Comparable models are given by the variance gamma model (Madan et al. (1998)), the hyperbolic model (Eberlein and Keller (1995)) and, for instance, the finite moment log stable model (Carr and Wu (2003)). The first test statistic in order to determine between finite and infinite activity price processes was proposed by Aït-Sahalia and Jacod (2011), who found evidence for the presence of infinite activity jumps in their empirical findings based on transactions of two most traded stocks, namely Intel and Microsoft. In view of these papers and findings, we think that the present paper states an important generalization of the findings in Bandi and Nguyen (2003) in order to capture price processes with a jump component of infinite activity originating from a pure jump Lévy-process.
In order to handle these models adequately, we focus on the nonparametric estimation of the coefficients of those processes. In particular, we will propose a kernel based nonparametric drift estimator in a Lévy driven diffusion model for which desirable asymptotic properties such as consistency and asymptotic normality are established. The derivation of these properties takes place in a double asymptotics scheme,

[1]benedikt.funke@math.tu-dortmund.de, Tel.: +49 231 755 5917

which means that we will work with a high-frequency sample of the underlying process X observed on an increasing time interval $[0, T]$.

Nonparametric inference for jump diffusions has attained some attention in the literature until now. For example, the above mentioned paper by Bandi and Nguyen (2003) investigates nonparametric kernel estimators for the coefficients of a diffusion driven by a finite activity jump process. Moreover, Shimizu (2008) as well as Mancini and Renò (2011) proposed a threshold estimator for the unknown drift and volatility function. An alternative method was proposed in Schmisser (2014), who suggested the use of a penalized least squares estimator based on model selection. Subsequently, we will focus on the same class of stochastic processes as in Schmisser (2014) and will conduct nonparametric inference, too. In the context of discontinuous stochastic volatility models, Kanaya and Christensen (2015) proposed a two-stage kernel based estimation procedure for the drift and the dispersion coefficient of the underlying volatility. In this context, see also Comte et al. (2010) for an alternative nonparametric approach to estimate the coefficients of the underlying volatility.

In the context of nonparametric inference for diffusions possessing continuous paths, a pioneering work has been published by Bandi and Phillips (2003). They derived an asymptotic theory for pointwise estimators of state-dependent drift and volatility functions. Moreover, a recently published article by Strauch (2015) investigates the minimax optimal rates for the drift of an ergodic diffusion in the multivariate setting.

To the best of our knowledge, those rates are not yet derived in the context of nonparametric inference for jump diffusions. In the univariate and continuous framework, we refer to Hoffmann (1999), where minimax optimal rates of nonparametric estimators were derived.

The paper is organized as follows: in Section 2, the model under investigation is introduced and, moreover, the used assumptions are stated. The following section presents the main results of this paper, namely the consistency as well as the asymptotic normality of the proposed drift estimator and the consistent estimation of its asymptotic variance. Section 4 is devoted to the incorporation of microstructure noise. We will see that our estimation procedure is robust to noisy observations by using the preaveraging approach proposed by Podolskij and Vetter (2006). For this purpose, we will perform a slight modification of the estimator by working with averaged observations. Section 5 concludes the paper and gives an outlook to further research investigations. All proofs are postponed to the Appendix.

2. The model and used assumptions

Consider the following jump diffusion process $X = (X_t)_{t\geq 0}$ such that

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \xi(X_{t-})dL_t, \quad X_0 = \eta, \tag{2.1}$$

where b, σ and ξ are unknown functions and X possesses the initial distribution $\eta \sim \Gamma(dx)$ such that $E[X_0^2] < \infty$. Moreover, L is a pure jump Lévy process of the form

$$dL_t = \int_{\mathbb{R}} y(\mu(dt, dy) - \nu(dy)dt) := \int_{\mathbb{R}} y\bar{\mu}(dt, dy),$$

where μ is a Poisson random measure compensated by its intensity measure $\nu(dy)dt$. We assume that the Lévy measure $\nu(dy)$ satisfies

$$E[L^2(1)] = Var(L(1)) = \int_{\mathbb{R}} y^2\nu(dy) < \infty.$$

We also presume that W and L are independent processes and that η is independent of W and L.

Suppose that we observe X in a high frequency setting on the interval $[0, T]$ on an equidistant grid at time points $0, \Delta, 2\Delta, ..., n\Delta = T$. Our first aim will be the construction of a meaningful pointwise estimator of the drift function b at a design point x based on the available sample $X_0, X_\Delta, X_{2\Delta}, ..., X_{n\Delta} = X_T$ such that

$$\Delta \to 0 \text{ as well as } T = n\Delta \to \infty.$$

Let us now list our assumptions on the considered model, which will ensure that X is strictly stationary and ergodic.

Assumption A1

i) The functions b, σ and ξ are globally Lipschitz-continuous.

ii) The function σ is bounded away from zero (ellipticity condition) as well as uniformly bounded for all x:

$$\exists\ \sigma_1, \sigma_0 \in \mathbb{R}^+ : \forall x \in \mathbb{R} : \quad 0 < \sigma_1 \leq \sigma(x) \leq \sigma_0.$$

iii) The function ξ is non-negative and also bounded:

$$\exists\ \xi_0 \in \mathbb{R}^+ : \forall x \in \mathbb{R} : \quad 0 \leq \xi(x) \leq \xi_0.$$

iv) The function b is elastic (cf. Masuda (2007)). This means that

$$\exists\ M > 0 : \forall x \in \mathbb{R}, |x| > M : \quad xb(x) \lesssim -x^2.$$

v) The Lévy measure ν possesses the properties that

$$Var(L(1)) = \int_{\mathbb{R}} y^2 \nu(dy) < \infty, \text{ and } \quad \nu(\{0\}) = 0.$$

Under Assumption A1,i) a unique strong solution X of (2.1) exists (Masuda (2007)). Furthermore, we remark that under A1,iv) b cannot be bounded as it is required in Bandi and Nguyen (2003). Moreover, under A1,i)-v), this solution is equipped with a unique invariant probability distribution $\Gamma(dx)$. In addition, X is exponentially β-mixing, which means that

$$\exists\ \gamma > 0 : \sup_{s \in \mathbb{R}_+} \int ||P_t(x, \cdot) - \eta P_{s+t}(\cdot)|| \eta P_s(dx) =: \beta_X(t) = O(e^{-\gamma t}), \text{ as } t \to \infty,$$

where $(P_t)_{t \in \mathbb{R}_+}$ denotes the transition semigroup of the underlying process X with initial distribution η. Moreover, ηP_t denotes the distribution of X_t and $||\lambda||$ defines the total variation norm of a signed measure λ; see Masuda (2007).
Using Theorem 2.1 in Masuda (2007), we can deduce the ergodicity of X, which means that for all measurable functions $g \in L^1(\Gamma(dx))$:

$$\frac{1}{T} \int_0^T g(X_s) ds \longrightarrow \int_{\mathbb{R}} g(x) \Gamma(dx) \text{ a.s., as } T \to \infty.$$

For equivalent reformulations of A1,iv), we recommend Masuda (2007).
We remark that due to our assumptions on the Lévy measure ν and the Lipschitz-continuity of the coefficients b, σ and ξ, we have that $E[X_t^2] < \infty$, too.
Moreover, we impose that

vi) Γ is absolutely continuous with respect to the Lebesgue measure and, thus, possesses a Lebesgue density π such that $\Gamma(dx) = \pi(x)dx$.

Notice that the process X is strictly stationary, because we assumed that $X_0 = \eta \sim \Gamma(dx)$.
In order to construct a meaningful estimator of $b(x)$, we make use of the following approximations, whose analogs in the finite activity case were also used in Bandi and Nguyen (2003):

$$b(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} E[X_{t+\Delta} - X_t | X_t = x] \tag{2.2}$$

$$\tilde{\sigma}^2(x) := \sigma^2(x) + \xi^2(x) Var(L(1)) = \lim_{\Delta \to 0} \frac{1}{\Delta} E[(X_{t+\Delta} - X_t)^2 | X_t = x] \tag{2.3}$$

$$\tilde{\xi}^k(x) := \xi^k(x) c_k(L(1)) = \lim_{\Delta \to 0} \frac{1}{\Delta} E[(X_{t+\Delta} - X_t)^k | X_t = x], \tag{2.4}$$

provided that all appearing terms exist and where

$$c_k(L(1)) = \int_{\mathbb{R}} z^k \nu(dz)$$

denotes the k-th cumulant of the random variable $L(1) \equiv L_1$. In contrast to the Brownian case, higher order moments do not vanish and are all dependent on the jump component. This indicates the essential difference to the ordinary diffusion case.
Now suppose that we observe the underlying process X at n equidistant time points over the interval $[0, T]$:

$$X_0, X_\Delta, X_{2\Delta}, ..., X_{n\Delta} \equiv X_T.$$

The approximations (2.2)-(2.4) yield us to the construction of nonparametric estimators for the values of interest. Hence, it seems intuitive to use, for instance, local polynomial estimators for the estimation of $b(x)$. Therefore, we propose a Nadaraya-Watson like estimator for the unknown drift at x according to

$$\hat{b}(x) := \frac{\frac{1}{nh} \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta} - x}{h}\right) \frac{(X_{(i+1)\Delta} - X_{i\Delta})}{\Delta}}{\frac{1}{nh} \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta} - x}{h}\right)},$$

where K is a kernel function and h a bandwidth. Comparable approaches have been conducted in Bandi and Phillips (2003) for diffusions possessing continuous paths as well as in Bandi and Nguyen (2003) for diffusions with jumps which originate from a finite activity process.
We now state the following assumptions on the kernel function K.

Assumption A2

i) Let K be a bounded probability density function which is symmetric around zero, differentiable and Lipschitz-continuous.

ii) Let K fulfill

$$\int_{\mathbb{R}} z^2 K(z) dz < \infty \quad \text{and} \quad \int_{\mathbb{R}} K^2(z) dz < \infty.$$

Let us shortly remark that A2 is standard in kernel based estimation procedures. One could also allow higher order kernels, for example a kernel function K of order $l \in \mathbb{N}$, which means that

$$\int_{\mathbb{R}} K(z) dz = 1, \ \int_{\mathbb{R}} z^j K(z) dz = 0, j = 1, ..., l-1 \text{ and } \int_{\mathbb{R}} z^l K(z) dz < \infty.$$

This would cause an asymptotic bias reduction, but for the sake of simplicity we will restrict ourselves to the case of kernels of order two.
For our purposes, one can always think of the Gaussian kernel

$$K_G(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

as a toy-example for which all assumptions hold true. Of course, there are many other choices for kernel functions possible, but in the literature it is quite reputable that the choice of the kernel function is not as important as the choice of the bandwidth parameter.

3. Asymptotic results

3.1. Estimation of $b(x)$

Let us now focus on the asymptotic properties of $\hat{b}(x)$. Under appropriate assumptions on Δ, h, and n we will derive its weak consistency as well as the asymptotic normality.

Theorem 3.1 (Weak Consistency)**.** *Let Δ and h fulfill*

$$n\Delta h = Th \to \infty \text{ and } \Delta^{1/2}h^{-2} \to 0.$$

Moreover, let assumptions A1 and A2 hold true. Then, provided that $\pi(x) > 0$, we can conclude that

$$\hat{b}(x) \xrightarrow{P} b(x), \quad \text{as } n \to \infty.$$

In addition to the consistency, we will now derive the asymptotic distribution of our proposed estimator. For this purpose, we have to strengthen the assumptions in a slightly manner.

Assumption A3

i) Let the drift function b as well as the stationary density π be twice continuously differentiable.

ii) Let Δ and h satisfy
$$\Delta^{1/2}h^{-2} \to 0,\ n\Delta h^5 \to 0 \text{ and } n\Delta^2 h^{-3} \to 0.$$

iii) Let the Lévy-measure ν fulfill
$$\int_{\mathbb{R}} y^4 \nu(dy) < \infty.$$

Let us shortly remark that A3,i) is required for the determination of the rate of the bias. A3,ii) ensures that the bias term is negligible and that appearing Taylor-remainder terms are negligible, too. Finally, A3,iii) is needed for the derivation of the asymptotic distribution via a standard central limit theorem for triangular arrays of martingale difference sequences.
We are now ready to state our next important theorem.

Theorem 3.2. *Under Assumptions A1-A3, provided that $\pi(x) > 0$ and $n\Delta h \to \infty$, it holds that*

$$\sqrt{n\Delta h}(\hat{b}(x) - b(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{||K||_2^2(Var(L(1))\xi^2(x) + \sigma^2(x))}{\pi(x)}\right), \text{ as } n \to \infty.$$

Remark 3.3. By defining analogous estimators for higher conditional moments (see (2.4)), an increasing time span $T \to \infty$ is necessary for the consistent estimation, too. This fact is also pointed out for the finite activity case in Bandi and Nguyen (2003), p. 304.

Example 3.4. *By letting $\Delta \sim n^{-\alpha}$ and $h \sim n^{-\beta}$, where $\alpha, \beta > 0$, the restrictions on Δ and h can be reformulated according to*

$$\alpha + 5\beta > 1, \ 2\alpha - 3\beta < 1, \ \alpha + \beta < 1, \ 4\beta < \alpha.$$

Thus, in terms of the rate

$$G(\alpha, \beta) := \frac{1 - \alpha - \beta}{2},$$

the maximal value of G under these restrictions is given by

$$G(\alpha_{opt}, \beta_{opt}) := G(0.615, 0.077) \approx 0.154.$$

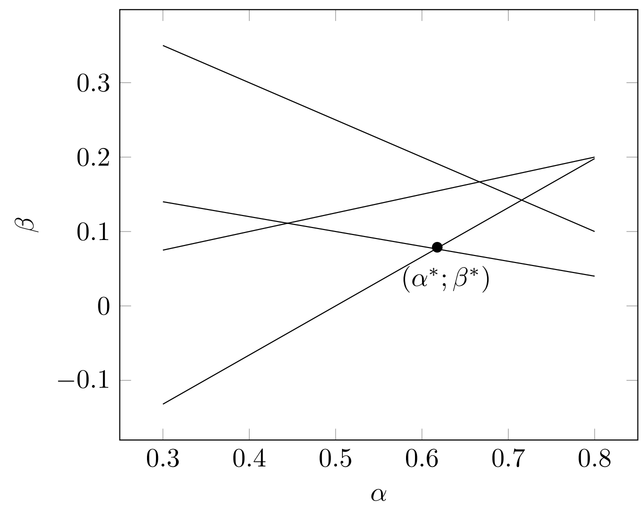

Fig. 1. Plot of the constraints on $G(\alpha, \beta)$ and the corresponding optimal coordinates $(\alpha^*; \beta^*)$.

By setting

$$n\Delta h^5 = O(1), \text{ as } n \to \infty,$$

an asymptotic bias occurs such that the AMSE of $\hat{b}(x)$ has the following form

$$\begin{aligned} \text{AMSE}(\hat{b}(x)) &= \left(\text{ABIAS}(\hat{b}(x))\right)^2 + \text{AVAR}\left(\hat{b}(x)\right) \\ &= h^4 \mu_2^2(K)\Lambda(x) + \frac{\int_{\mathbb{R}} K^2(z)dz\tilde{\sigma}^2(x)}{n\Delta h\pi(x)}, \end{aligned}$$

where

$$\Lambda(x) := \frac{b'(x)\pi'(x)}{\pi(x)} + \frac{b''(x)}{2}$$

denotes one part of the bias term,

$$\mu_2(K) = \int_{\mathbb{R}} z^2 K(z)dz$$

the second moment of K , and

$$\tilde{\sigma}^2(x) := \sigma^2(x) + Var(L(1))\xi^2(x).$$

By minimizing the AMSE with respect to h, the resulting bandwidth is called "oracle bandwidth" and in our case this optimal bandwidth $h_{opt,oracle}(x)$ has the form

$$h_{opt,oracle}(x) = (n\Delta)^{-1/5} \left(\frac{\tilde{\sigma}^2(x) \int_{\mathbb{R}} K^2(z)dz}{4\mu_2^2(K)\Lambda^2(x)\pi(x)} \right)^{-1/5}.$$

Using this bandwidth, the optimal AMSE is of order

$$AMSE\left(\hat{b}(x)\right) = O\left((n\Delta)^{-4/5}\right).$$

We remark that this order is achieved provided that

$$T^{8/5}\Delta \to 0$$

in order to ensure that h fulfills A3, ii).

3.2. Estimation of the Asymptotic Variance

In order to establish asymptotic pointwise confidence intervals, a consistent estimator of the asymptotic variance is needed. Hence, in view of Theorem 3.2, the asymptotic variance is given by

$$\overline{\sigma}^2(x) := \frac{||K||_2^2(Var(L(1))\xi^2(x) + \sigma^2(x))}{\pi(x)}.$$

As it is pointed out in the Appendix, the denominator $\pi(x)$ can be consistently estimated by

$$\hat{\pi}(x) = \frac{\Delta}{Th} \sum_{i=0}^{n-1} K\left(\frac{x - X_{i\Delta}}{h}\right) = \frac{1}{nh} \sum_{i=0}^{n-1} K\left(\frac{x - X_{i\Delta}}{h}\right).$$

In order to conduct a meaningful estimator for the numerator, we make use of approximation (2.3). Therefore, we propose the estimator of $\tilde{\sigma}^2(x)$ as follows:

$$\hat{\tilde{\sigma}}^2(x) := \frac{\frac{1}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right) \frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta}}{\frac{1}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)}.$$

For the derivation of the consistency of this estimator, we have to strengthen the assumptions of Theorem 3.1 marginally.

Theorem 3.5 (Weak Consistency)**.** *Under Assumptions A1, A2, and additionally under the assumption that the Lévy measure ν fulfills*

$$\int_{\mathbb{R}} y^4 \nu(dy) < \infty,$$

$\hat{\tilde{\sigma}}^2(x)$ is a (weak) consistent estimator for $\tilde{\sigma}^2(x)$, provided that $\pi(x) > 0$. In particular, we find out that

$$\hat{\tilde{\sigma}}^2(x) \xrightarrow{P} \tilde{\sigma}^2(x), \text{ as } n \to \infty.$$

Using this theorem, we are able to deduce approximate pointwise confidence intervals for $b(x)$ according to

$$I_x := \left[\hat{b}(x) - \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{\hat{\sigma}}(x)}{\sqrt{n\Delta h}} \leq b(x) \leq \hat{b}(x) + \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{\hat{\sigma}}(x)}{\sqrt{n\Delta h}} \right],$$

where

$$\hat{\hat{\sigma}}(x) := \frac{||K||_2^2 \hat{\hat{\sigma}}^2(x)}{\hat{\pi}(x)}$$

and provided that the bias is negligible.

4. Robustness to Microstructure Noise

In this section, we will work with noisy observations of the form

$$Y_{i\Delta} = X_{i\Delta} + \varepsilon_{i\Delta}, \ i = 1, ..., n,$$

where we assume that $\varepsilon_{i\Delta}, \ i = 1, ..., n$, are i.i.d. random variables such that

$$E[\varepsilon_{i\Delta}] = 0 \text{ and } E[\varepsilon_{i\Delta}^2] := \sigma_\varepsilon^2$$

for all i and Δ. Moreover, the processes ε and X are assumed to be independent.

It is widely known that in the case of high-frequency observation schemes, measurement errors as well as the so-called microstructure noise play significant roles and can be found in several financial data sets; see Zhang et al. (2005), Jacod et al. (2009) or Jones (2003). In the framework of Itô-diffusions, Podolskij and Vetter (2006) introduced the so-called pre-averaging approach for the nonparametric estimation of integrated volatility and quarticity based on realized bipower variation. This widely used approach has, for instance, been studied in Jacod et al. (2009) in a more general setting and in the context of stochastic volatility models in Kanaya and Christensen (2015) as well as for diffusions without jumps in Greenwood et al. (2015). We will focus on their approach for the estimation of $b(x)$ in our considered Lévy-driven diffusion model as follows.

We decompose the available sample $Y_{i\Delta}, \ i = 1, ..., n$, into $m_n := m$ subgroups of length $r_n := r$, such that $mr = n$. Without loss of generality, we assume that n can be decomposed into the product of m and r. Otherwise, one would introduce a first and a last block with a length smaller than r. Moreover, the block length as well as the number of blocks fulfill

$$r \to \infty, \quad m = \left\lfloor \frac{n}{r} \right\rfloor \to \infty, \text{ and } \Delta r \to 0 \text{ as } n \to \infty.$$

Instead of working with the contaminated raw data set $Y_{i\Delta}, \ i = 1, ..., n$, we build averages inside every block j, where $j = 1, ..., r$. Thus, we define

$$\bar{Y}_j := \bar{Y}_{j,\Delta} := \frac{1}{r} \sum_{i=1}^{r} Y_{((j-1)r+i)\Delta}$$

and analogously

$$\bar{X}_j := \bar{X}_{j,\Delta} := \frac{1}{r} \sum_{i=1}^{r} X_{((j-1)r+i)\Delta}, \text{ and } \bar{\varepsilon}_j := \bar{\varepsilon}_{j,\Delta} := \frac{1}{r} \sum_{i=1}^{r} \varepsilon_{((j-1)r+i)\Delta}.$$

Using these notations, we obtain

$$\bar{Y}_j = \bar{X}_j + \bar{\varepsilon}_j.$$

The motivation for this approach is that the averages $\bar{Y}_j$ act as approximations of the averages $\bar{X}_j$, which are, in turn, approximations of the original sample $X_{i\Delta},\ i = 1, ..., n$.
Let us now define a drift estimator $\hat{b}_Y(x)$ based on the sample $\bar{Y}_j,\ j = 1, ..., m$, according to

$$\hat{b}_Y(x) := \frac{\frac{1}{mh}\sum_{j=1}^{m-1} K\left(\frac{\bar{Y}_j - x}{h}\right)\left(\bar{Y}_{j+1} - \bar{Y}_j\right)}{\frac{r\Delta}{mh}\sum_{j=1}^{m-1} K\left(\frac{\bar{Y}_j - x}{h}\right)}.$$

In order to establish the asymptotic properties of this estimator, we will make use of the results in our previous section. For this purpose, we have to impose the following assumptions, mainly concerning the rates of convergence of the included sequences.

Assumption A4

i) The noise process $\varepsilon = (\varepsilon_t)_{t\geq 0}$ is an i.i.d. process, in particular, the random variables $\varepsilon_{i\Delta}, i = 1, ..., n$, are i.i.d. with
$$E[\varepsilon_{i\Delta}] = 0 \text{ and } E[\varepsilon_{i\Delta}^2] = \sigma_\varepsilon^2 < \infty$$
for all $0 \leq i\Delta \leq T$.

ii) The processes $X = (X_t)_{t\geq 0}$ and $\varepsilon = (\varepsilon_t)_{t\geq 0}$ are independent.

iii) The block length r fulfills
$$r \to \infty \text{ and } \Delta r \to 0 \text{ as } n \to \infty.$$
Moreover, the number of blocks m behaves like $m = \lfloor n/r \rfloor \to \infty$ as $n \to \infty$.

iv) The appearing parameters $r,\ h$ and Δ are connected via
$$n\Delta rh^5 \to 0,\ n(\Delta r)^2 h^{-3} \to 0,$$
$$(\Delta r)^{1/2} h^{-2} \to 0 \text{ and } n\Delta(rh^{3/2})^{-1} \to 0.$$

We are now ready to state our main theorem concerning the asymptotic distribution of $\hat{b}_Y(x)$.

Theorem 4.1. *Under Assumptions A1-A4, provided that $\pi(x) > 0$ and $n\Delta rh \to \infty$, it holds that*

$$\sqrt{n\Delta h}\left(\hat{b}_Y(x) - b(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{||K||_2^2(\sigma^2(x) + \xi^2(x)Var(L(1)))}{\pi(x)}\right),\ \text{as } n \to \infty.$$

Example 4.2. *In the noisy setting it would be interesting to derive an optimal rate of convergence under the assumptions on r, Δ and h, too.*
By letting

$$\Delta \sim n^{-\alpha},\ h \sim n^{-\beta},\ \text{and } r \sim n^{\rho},\quad \alpha, \beta, \rho > 0,$$

a variety of connections between α, β and ρ exists. Let again be

$$G(\alpha, \beta) := \frac{1 - \alpha - \beta}{2}$$

and maximize the function G with respect to

$$\alpha - \rho + 5\beta > 1,\ 2(\alpha - \rho) - 3\beta > 1,\ \alpha - \rho + \beta < 1,$$
$$\alpha - \rho - 4\beta > 0,\ 4\alpha - 3\rho - 3\beta > 1,\ 3\alpha - \rho - 3\beta > 1,$$
$$3\alpha - 3\beta - 2\rho > 1,\ \alpha + 2\rho - 3\beta > 1,\ -\alpha + \beta + \rho < 1,$$
$$2\alpha - 3\beta > 1,\ -2\alpha + \beta + 3\rho < 1,\ \text{and } 3\alpha - 2\rho - 3\beta > 1.$$

These different connections appear during the proofs and some of them are redundant. Nonetheless, we state them here for the sake of completeness.
The simplex algorithm yields the following result:

$$\alpha \approx 0.821, \quad \beta \approx 0.077,$$

which leads to an optimal rate under our assumptions, which is proportional to $n^{0.051}$*. This rate is slower than in the case of non-noisy data, which is not surprising. We already know that in our setting* $r\Delta$ *denotes the sampling frequency of the new random sample* $\{\bar{X}_j, j = 1, ..., m\}$ *and should behave like* Δ *in the previous section without additional noise. Therefore, we choose* $r \sim n^{0.2}$ *such that* $r\Delta \sim n^{-0.62}$*, which is approximately the optimal result for* α *in the previous section.*

5. Conclusion

In this paper we proposed a pointwise nonparametric kernel based estimator for the unknown drift function b in a Lévy-driven jump diffusion model. We derived the consistency and the asymptotic normality and, moreover, showed that this estimator is robust under microstructure noise by a slight variation of the original estimator.
In view of the works by Comte et al. (2007, 2009), it would be interesting to treat the case of an integrated diffusion process X, whose drift has to be estimated.
Furthermore, in the context of stochastic volatility models (cf. Barndorff-Nielsen and Shephard (2004) as well as Comte et al. (2010)), we think that it would be interesting to estimate the state-dependent drift of the underlying volatility in those models, too.
Finally, the question which bandwidth selection procedure is optimal is often elusive in the literature. Hence, we think that a theoretical treatment of this issue in the framework of (jump) diffusions is important and interesting, too. All mentioned extensions are under preparation and will be covered in forthcoming papers.

6. Appendix

In this section, we will present the proofs of the stated results. The following proposition is very helpful for the derivation of the upper stated theorems. It can be found in Schmisser (2014) and, therefore, we omit its proof.

Proposition 6.1. *Let* $X = (X_t)_{t\geq 0}$ *be the solution of* (2.1)*. Under assumptions A1,i)-vi), it exists a constant* $C > 0$ *such that*

$$E\left[\sup_{|s-t|\leq\Delta} (X_s - X_t)^2\right] \leq C\Delta,$$

provided that $\Delta \leq 1$*.*
Moreover, provided that

$$\int_{\mathbb{R}} y^4 \nu(dy) < \infty,$$

the fourth moments can also be bounded by

$$E\left[\sup_{|s-t|\leq\Delta} (X_s - X_t)^4\right] \leq \tilde{C}\Delta,$$

where $\Delta \leq 1$ *and* $\tilde{C}$ *denotes another positive and deterministic constant.*

We will now present the proof of Theorem 3.1.

Proof of Theorem 3.1. Focus on the following decomposition of $\hat{b}(x)$:

$$\begin{aligned}\hat{b}(x) &= \frac{\frac{1}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(X_{(i+1)\Delta}-X_{i\Delta}\right)}{\frac{\Delta}{nh}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\ &= \frac{\frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\int_{i\Delta}^{(i+1)\Delta} b(X_s)ds + \int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s + \int_{i\Delta}^{(i+1)\Delta}\xi(X_{s-})dL_s\right)}{\frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\ &= \frac{\frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right) b(X_{i\Delta})}{\frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\ &\quad + \frac{\frac{1}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\int_{i\Delta}^{(i+1)\Delta}\left(b(X_s)-b(X_{i\Delta})\right)ds + \int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s + \int_{i\Delta}^{(i+1)\Delta}\xi(X_{s-})dL_s\right)}{\frac{\Delta}{h}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)}.\end{aligned}$$

The first term of the numerator is a consistent estimate of $b(x)$. The second one is a remainder term. The last two terms are noise terms, which will play a role for the determination of the asymptotic distribution but are at this stage negligible, too.
At first, focus on the denominator

$$\begin{aligned}D(x) :&= \frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right) \\ &= \frac{1}{T}\int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)ds + \frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right) - \int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)ds \\ &= \frac{1}{T}\int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)ds + \frac{1}{Th}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left(K\left(\frac{X_{i\Delta}-x}{h}\right) - K\left(\frac{x-X_s}{h}\right)\right)ds \\ &:= \frac{1}{T}\int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)ds + F_1^n.\end{aligned}$$

We are interested in the rate of convergence of the approximation error F_1^n and consider, therefore, its L^1-distance. Under the assumption that K is Lipschitz-continuous, by Proposition 6.1, the Cauchy-Schwarz as well as the Jensen inequality, we conclude that

$$\begin{aligned}E[|F_1^n|] &\leq \frac{1}{Th}\sum_{i=0}^{n-1} E\left[\int_{i\Delta}^{(i+1)\Delta}\left|K\left(\frac{X_{i\Delta}-x}{h}\right) - K\left(\frac{x-X_s}{h}\right)\right|ds\right] \\ &\leq \frac{1}{Th}\sum_{i=0}^{n-1} E\left[\int_{i\Delta}^{(i+1)\Delta}||K'||_\infty\left|\frac{X_{i\Delta}-X_s}{h}\right|1_{[i\Delta,(i+1)\Delta]}(s)ds\right] \\ &\leq \frac{||K'||_\infty}{Th^2}\sum_{i=0}^{n-1} E\left[\left(\int_{i\Delta}^{(i+1)\Delta}(X_{i\Delta}-X_s)^2 ds\right)^{1/2}\right]\Delta^{1/2} \\ &\leq \frac{\Delta^{1/2}||K'||_\infty}{Th^2}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left(E[X_{i\Delta}-X_s)^2]\right)^{1/2}ds \\ &\lesssim \frac{\Delta^{1/2}}{Th^2}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\Delta^{1/2}ds = \frac{n\Delta^2}{Th^2} = \frac{\Delta}{h^2}.\end{aligned}$$

With the help of the ergodicity of X it holds that

$$\begin{aligned}\frac{1}{T}\cdot \mathrm{IV} &= \frac{1}{T}\int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)ds + O_P\left(\frac{\Delta}{h^2}\right)\\ &\longrightarrow \pi(x),\ \text{ as } n,T\to\infty.\end{aligned}$$

Therefore, the denominator converges to $\pi(x)$ in probability.
The first term in the numerator can be handled in an analogous way and it turns out that it converges to $b(x)\pi(x)$ in probability. By the Lipschitz continuity of b, the remainder term is of order $O_P\left(\frac{\Delta}{h^2}\right)$, too.
Let the filtration $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-algebras be given by

$$\mathcal{F}_t = \sigma(X_0,(W_s,L_s),s\leq t).$$

For the Brownian noise term, consider its L^2-distance as follows:

$$\begin{aligned}&E\left[\left(\frac{1}{Th}\sum_{i=0}^{n-1}K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s\right)^2\right]\\ &=\frac{1}{T^2h^2}\sum_{i=0}^{n-1}E\left[K^2\left(\frac{X_{i\Delta}-x}{h}\right)E\left[\int_{i\Delta}^{(i+1)\Delta}\sigma^2(X_s)ds\middle|\mathcal{F}_{i\Delta}\right]\right]\\ &\leq \sigma_1\frac{\Delta}{T^2h^2}\sum_{i=0}^{n-1}E\left[K^2\left(\frac{X_{i\Delta}-x}{h}\right)\right]\lesssim \frac{1}{Th}.\end{aligned}$$

Using the Burkholder-Davis-Gundy type inequality (see Schmisser (2014), Result 11), we have that

$$E\left[\sup_{|s-t|\leq\Delta}\left(\int_t^s \xi(X_{s_-})dL_s\right)^2\middle|\mathcal{F}_t\right]\leq C_2E\left[\int_t^{t+\Delta}\xi^2(X_u)du\middle|\mathcal{F}_t\right]\left(1+\int_{\mathbb{R}}y^2\nu(dy)\right).$$

Hence, the second noise term possesses the same rate of convergence as the first one. This finishes the proof by using the quotient limit theorem for stationary Markov processes (cf. Revuz and Yor (1999), Theorem 3.12). □

We now state the proof of the asymptotic normality of $\hat{b}(x)$.

Proof of Theorem 3.2. We start with the derivation of the numerator as follows:

$$\begin{aligned}&\sqrt{Th}\left(\frac{1}{Th}\sum_{i=0}^{n-1}K\left(\frac{X_{i\Delta}-x}{h}\right)(X_{(i+1)\Delta}-X_{i\Delta})-b(x)\right)\\ &=\sqrt{Th}\left(\frac{1}{Th}\sum_{i=0}^{n-1}K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}b(X_s)ds-b(x)\right)\\ &\quad+\sqrt{Th}\left(\frac{1}{Th}\sum_{i=0}^{n-1}K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s+\int_{i\Delta}^{(i+1)\Delta}\xi(X_{s-})dL_s\right)\right)\\ &=\sqrt{Th}\left(\frac{1}{T}\int_0^T\frac{1}{h}K\left(\frac{x-X_s}{h}\right)b(X_s)ds-b(x)+F_b^n\right)\\ &\quad+\frac{1}{\sqrt{Th}}\sum_{i=0}^{n-1}K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s+\int_{i\Delta}^{(i+1)\Delta}\xi(X_{s-})dL_s\right)\end{aligned}$$

$$= \sqrt{Th}\left(\frac{1}{T}\int_0^T \frac{1}{h}K\left(\frac{x-X_s}{h}\right)(b(X_s)-b(x))\,ds + F_b^n\right) \tag{6.5}$$
$$+ \frac{1}{\sqrt{Th}}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dW_s + \int_{i\Delta}^{(i+1)\Delta}\xi(X_{s-})dL_s\right),$$

where F_b^n denotes the approximation error, which is negligible compared to the other terms. By letting $T \to \infty$, the first term is a bias which is of order $O(h^2)$:

$$\int_{\mathbb{R}} \frac{1}{h}K\left(\frac{x-y}{h}\right)(b(y)-b(x))\,\pi(y)dy = \int_{\mathbb{R}} K(z)(b(x-zh)-b(x))\pi(x-zh)dz$$
$$= \int_{\mathbb{R}} K(z)\left(-zhb'(x) + \frac{z^2h^2}{2}b''(x) + O(h^3)\right)\left(\pi(x) - zh\pi'(x) + O(h^2)\right)dz$$
$$= \int K(z)\left(b'(x)\pi(x)(-zh) + z^2h^2b'(x)\pi'(x)\frac{z^2h^2}{2}b''(x)\pi(x)\right)dz + O(h^3)$$
$$= h^2\int z^2K(z)dz\left(b'(x)\pi'(x) + \frac{b''(x)\pi(x)}{2}\right) + O(h^3) = O(h^2).$$

By the use of $Th^5 = n\Delta h^5 \to 0$ as $n \to \infty$, this term is negligible, too.
Observe that the remaining two terms are martingale difference sequences with respect to the upper defined filtration $(\mathcal{F}_t)_{t\geq 0}$. By invoking a standard central limit theorem for arrays of martingale difference sequences as, for instance, in Shiryayev (1995), we can treat the noise terms by analogous arguments. In fact, by the boundedness of $\int_{\mathbb{R}} y^4\nu(dy)$ and the independence of L and W, we can finish the proof. □

We now state the proof of the consistency of $\hat{\hat{\sigma}}^2(x)$.

Proof of Theorem 3.5. In order to establish the consistency of $\hat{\hat{\sigma}}^2(x)$, we use Itô´s formula to decompose the squared increment as follows:

$$(X_{(i+1)\Delta} - X_{i\Delta})^2 = 2\int_{i\Delta}^{(i+1)\Delta}(X_{s_-} - X_{i\Delta})dX_s$$
$$+ \int_{i\Delta}^{(i+1)\Delta}(\sigma^2(X_s) + \xi^2(X_s)Var(L(1)))ds + \int_{i\Delta}^{(i+1)\Delta}\xi^2(X_{s_-})\int_{\mathbb{R}} y^2\bar{\mu}(dy,ds)$$
$$= 2\int_{i\Delta}^{(i+1)\Delta}(X_s - X_{i\Delta})b(X_s)ds + 2\int_{i\Delta}^{(i+1)\Delta}(X_s - X_{i\Delta})\sigma(X_s)dW_s$$
$$+ 2\int_{i\Delta}^{(i+1)\Delta}(X_{s_-} - X_{i\Delta})\xi(X_{s_-})dL_s + \int_{i\Delta}^{(i+1)\Delta}(\sigma^2(X_s) + \xi^2(X_s)Var(L(1)))ds$$
$$+ \int_{i\Delta}^{(i+1)\Delta}\xi^2(X_{s_-})\int_{\mathbb{R}} y^2\bar{\mu}(dy,ds)$$

Due to this representation, we are able to decompose the estimator $\hat{\tilde{\sigma}}^2(x)$ into the following five parts

$$\begin{aligned}
\hat{\tilde{\sigma}}^2(x) &= \frac{\frac{1}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)(X_{(i+1)\Delta}-X_{i\Delta})^2}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&= \frac{\frac{2}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}(X_s-X_{i\Delta})b(X_s)ds}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad + \frac{\frac{2}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}(X_s-X_{i\Delta})\sigma(X_s)dW_s}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad + \frac{\frac{2}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}(X_{s_-}-X_{i\Delta})\xi(X_{s_-})dL_s}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad + \frac{\frac{1}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}(\sigma^2(X_s)+\xi^2(X_s)Var(L(1)))ds}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad + \frac{\frac{1}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}\xi^2(X_{s_-})\int_{\mathbb{R}} y^2\bar{\mu}(dy,ds)}{\frac{\Delta}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)}.
\end{aligned}$$

By the use of Proposition 6.1, the first three terms of the numerator converge to zero, which is exemplified in the term, which involves the drift b:

$$\begin{aligned}
E[|I'|] &= E\left[\left|\frac{2}{Th}\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{i\Delta}^{(i+1)\Delta}(X_s-X_{i\Delta})b(X_s)ds\right|\right] \\
&\le \frac{2}{Th}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} E\left[\left|K\left(\frac{X_{i\Delta}-x}{h}\right)\right|\cdot\left|(X_s-X_{i\Delta})b(X_s)\right|\right]ds \\
&\le \frac{2||K||_\infty}{Th}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} E\left[|X_s-X_{i\Delta}|\cdot|b(X_s)|\right]ds \\
&\le \frac{2||K||_\infty}{Th}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left(E\left[(X_s-X_{i\Delta})^2\right]\right)^{1/2}\left(E[b^2(X_s)]\right)^{1/2}ds \\
&\lesssim \frac{\Delta^{1/2}}{Th}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left(E[b^2(X_s)]\right)^{1/2}ds = \frac{\Delta^{1/2}n\Delta}{Th}\left(E[b^2(X_0)]\right)^{1/2} \\
&= O\left(\frac{n\Delta^{3/2}}{Th}\right) = O\left(\frac{\Delta^{1/2}}{h}\right) = o(1), \text{ as } n\to\infty.
\end{aligned}$$

Using the ergodicity property of X, we conclude that the fourth term converges to $\tilde{\sigma}^2(x)\pi(x)$ in probability. The last term is a martingale term, which is handled in analogy to the drift case. Using the consistency of the denominator, we see that

$$\hat{\tilde{\sigma}}^2(x) \xrightarrow{P} \tilde{\sigma}^2(x), \text{ as } n, T\to\infty.$$

□

For the derivation of the proof of Theorem 4.1, we will need an analog to Proposition 6.1 in the noisy framework.

Proposition 6.2. *Under Assumptions A1,i)-vi) and $\Delta \le 1$, the following statements hold true*

$$1) \max_{1\le j\le m} E\left[(X_{jr\Delta} - X_{(j-1)r\Delta})^2\right] \lesssim r\Delta,$$

$$2) \max_{1\le j\le m} E\left[(\bar{X}_j - X_{(j-1)r\Delta})^2\right] \lesssim r\Delta,$$

$$3) \max_{1\le j\le m-1} E\left[(\bar{X}_{j+1} - \bar{X}_j)^2\right] \lesssim r\Delta.$$

Proof of Proposition 6.2. The first statement can be directly deduced by Proposition 6.1. Observe for the second statement that

$$\begin{aligned}
&E[|\bar{X}_j - X_{(j-1)r\Delta}|] \le \frac{1}{r}\sum_{i=1}^{r} E[|X_{((j-1)r+i)\Delta} - X_{(j-1)r\Delta}|] \\
&\le \frac{1}{r}\sum_{i=1}^{r} \left(E[(X_{((j-1)r+i)\Delta} - X_{(j-1)r\Delta})^2]\right)^{1/2} \\
&\lesssim \frac{1}{r}\sum_{i=1}^{r} \Bigg(\Delta \int_{(j-1)r\Delta}^{((j-1)r+i)\Delta} E[b^2(X_s)]ds + \int_{(j-1)r\Delta}^{((j-1)r+i)\Delta} E[\sigma^2(X_s)]ds \\
&\qquad + Var(L(1)) \int_{(j-1)r\Delta}^{((j-1)r+i)\Delta} E[\xi^2(X_s)]ds\Bigg)^{1/2} \\
&\le \frac{1}{r}\sum_{i=1}^{r} \left(\Delta^2 i E[b^2(X_0)] + i\Delta(||\sigma^2||_\infty + ||\xi^2||_\infty Var(L(1)))\right)^{1/2} \\
&\lesssim \frac{1}{r}\sum_{i=1}^{r} (i\Delta)^{1/2} \le \frac{1}{r}\Delta^{1/2}\left(\sum_{i=1}^{r} i\right)^{1/2}\sqrt{r} \lesssim \frac{r^{3/2}\Delta^{1/2}}{r} = (r\Delta)^{1/2}.
\end{aligned}$$

Now use Jensen´s inequality and the monotonicity of the function $t \to \sqrt{t}$ to finish the proof of statement 2).
Finally for 3), observe that

$$\begin{aligned}
E[|\bar{X}_{j+1} - \bar{X}_j|] &\le E[|\bar{X}_{j+1} - X_{jr\Delta}|] + E[|X_{jr\Delta} - X_{(j-1)r\Delta}|] \\
&\quad + E[|\bar{X}_j - X_{(j-1)r\Delta}|] \lesssim 3(r\Delta)^{1/2}
\end{aligned}$$

due to 1) and 2). □

Proof of Theorem 4.1. Recall that we observe a high-frequency sample

$$Y_{i\Delta} = X_{i\Delta} + \varepsilon_{i\Delta},\ i = 0, ..., n,$$

consisting of the original diffusion process and contaminated by additional noise. Now define

$$\hat{b}_X(x) := \frac{\frac{1}{(m-1)h}\sum_{j=1}^{m-1}\left(X_{jr\Delta} - X_{(j-1)r\Delta}\right) K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)}{\frac{r\Delta}{(m-1)h}\sum_{j=1}^{m-1} K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)} := \frac{\hat{N}_X(x)}{\hat{D}_X(x)},$$

which is the drift estimator of $b(x)$ based on a high-frequency sample $\{X_{jr\Delta} := X_{j\delta}, j = 0, ..., m\}$. The new sample is a "thinned out version" with sampling frequency $\delta = \Delta r \to 0$ of the original sample. From an asymptotic point of view, this has no impact on the derivation of the asymptotic distribution, since we work in a high-frequency setting. Therefore, and under Assumptions A1-A4, this estimator is consistent and asymptotically normally distributed due to Theorems 3.1 and 3.2:

$$\sqrt{m\delta h}\left(\hat{b}_X(x) - b(x)\right) = \sqrt{n\Delta h}\left(\hat{b}_X(x) - b(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{||K||_2^2\tilde{\sigma}^2(x)}{\pi(x)}\right), \text{ as } n \to \infty.$$

Our aim is now to prove that

$$\hat{N}_X(x) - \hat{N}_Y(x) = o_P((n\Delta h)^{-1/2}) = o_P((m\delta h)^{-1/2}) \tag{6.6}$$

as well as

$$\hat{D}_X(x) - \hat{D}_Y(x) = o_P((n\Delta h)^{-1/2}) = o_P((m\delta h)^{-1/2}), \tag{6.7}$$

where analogously

$$\hat{b}_Y(x) = \frac{\frac{1}{(m-1)h}\sum_{j=1}^{m-1} K\left(\frac{\bar{Y}_j - x}{h}\right)\left(\bar{Y}_{j+1} - \bar{Y}_j\right)}{\frac{r\Delta}{(m-1)h}\sum_{j=1}^{m-1} K\left(\frac{\bar{Y}_j - x}{h}\right)} := \frac{\hat{N}_Y(x)}{\hat{D}_Y(x)}.$$

We start with the derivation of (6.7). For this purpose, we make at first use of Proposition 6.2 as well as the Lipschitz-continuity of K:

$$\begin{aligned} E[|\hat{D}_X(x) - \hat{D}_Y(x)|] &\leq \frac{r\Delta}{(m-1)h}\sum_{j=1}^{m-1} E\left[\left|K\left(\frac{\bar{Y}_j - x}{h}\right) - K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right|\right] \\ &\leq \frac{r\Delta||K'||_\infty}{(m-1)h^2}\sum_{j=1}^{m-1} E\left[|\bar{Y}_j - X_{(j-1)r\Delta}|\right] \\ &= \frac{r\Delta||K'||_\infty}{(m-1)h^2}\sum_{j=1}^{m-1} E\left[|\bar{X}_j - X_{(j-1)r\Delta} + \bar{\varepsilon}_j|\right] \\ &\lesssim \frac{r\Delta}{(m-1)h^2}\sum_{j=1}^{m-1}\left((r\Delta)^{1/2} + \frac{\sigma_\varepsilon}{\sqrt{r}}\right) \\ &= O\left(\frac{(\Delta r)^{3/2}}{h^2} + \frac{\Delta r^{1/2}}{h^2}\right). \end{aligned}$$

To prove (6.7), both following terms have to converge to zero in probability:

$$\sqrt{n\Delta h}\left(\hat{D}_X(x) - \hat{D}_Y(x)\right) = O_P\left(\frac{(n\Delta h)^{1/2}(\Delta r)^{3/2}}{h^2} + \frac{(n\Delta h)^{1/2}\Delta r^{1/2}}{h^2}\right) \stackrel{!}{=} o_P(1), \text{ as } n \to \infty, \tag{6.8}$$

which is ensured by assumption A4.

We now turn to the difference of both numerators and decompose it as follows:

$$
\begin{aligned}
&\hat{N}_X(x) - \hat{N}_Y(x) \\
&= \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\bar{Y}_{j+1} - \bar{Y}_j\right) K\left(\frac{\bar{Y}_j - x}{h}\right) - \left(X_{jr\Delta} - X_{(j-1)r\Delta}\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\
&= \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\bar{Y}_{j+1} - \bar{Y}_j\right) \left(K\left(\frac{\bar{Y}_j - x}{h}\right) - K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right)\right) \\
&\quad + \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\left(\bar{Y}_{j+1} - \bar{Y}_j\right) - \left(X_{jr\Delta} - X_{(j-1)r\Delta}\right)\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\
&:= A_n(x) + B_n(x).
\end{aligned}
$$

Again, we determine the L^1-distance of both terms as follows:

$$
\begin{aligned}
&E[|A_n(x)|] \leq \frac{1}{(m-1)h} \sum_{j=1}^{m-1} E\left[|\bar{Y}_{j+1} - \bar{Y}_j| \cdot \left| K\left(\frac{\bar{Y}_j - x}{h}\right) - K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right| \right] \\
&\leq \frac{||K'||_\infty}{(m-1)h^2} \sum_{j=1}^{m-1} E\left[|\bar{Y}_{j+1} - \bar{Y}_j| \cdot |\bar{Y}_j - X_{(j-1)r\Delta}| \right] \\
&\lesssim \frac{||K'||_\infty}{(m-1)h^2} \sum_{j=1}^{m-1} \left(r\Delta + \frac{1}{r}\right)^{1/2} \cdot \left(r\Delta + \frac{1}{r}\right)^{1/2} = \frac{r\Delta}{h^2} + \frac{1}{rh^2},
\end{aligned}
$$

where we used the fact that

$$E\left[(\bar{\varepsilon}_{j+1} - \bar{\varepsilon}_j)^2\right] \leq 2\left(E[\bar{\varepsilon}_{j+1}^2] + E[\bar{\varepsilon}_j^2]\right) = \frac{4\sigma_\varepsilon^2}{r}.$$

Hence, we conclude that

$$E[|A_n(x)|] = O\left(\frac{r\Delta}{h^2} + \frac{1}{rh^2}\right), \text{ as } n \to \infty.$$

Furthermore, assumption A4 ensures that $A_n(x)$ possesses the proper rate $o_P((n\Delta h)^{-1/2})$:

$$\sqrt{n\Delta h} A_n(x) = O_P\left(\frac{(n\Delta h)^{1/2} r\Delta}{h^2} + \frac{(n\Delta h)^{1/2}}{rh^2}\right) = o_P(1), \text{ as } n \to \infty.$$

The second term $B_n(x)$ is, in a first stage, decomposed as follows:

$$
\begin{aligned}
B_n(x) &= \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\left(\bar{Y}_{j+1} - \bar{Y}_j\right) - \left(X_{jr\Delta} - X_{(j-1)r\Delta}\right)\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\
&= \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\left(\bar{X}_{j+1} - \bar{X}_j\right) - \left(X_{(j+1)r\Delta} - X_{jr\Delta}\right)\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\
&\quad + \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\left(X_{(j+1)r\Delta} - X_{jr\Delta}\right) - \left(X_{jr\Delta} - X_{(j-1)r\Delta}\right)\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\
&\quad + \frac{1}{(m-1)h} \sum_{j=1}^{m-1} \left(\left(\bar{\varepsilon}_{j+1} - \bar{\varepsilon}_j\right) K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) := \sum_{k=1}^{3} B_{n,k}(x).
\end{aligned}
$$

At first look at term $B_{n,3}(x)$:

$$
\begin{aligned}
B_{n,3}(x) &= \frac{1}{(m-1)h}\sum_{j=1}^{m-1}\left((\bar{\varepsilon}_{j+1}-\bar{\varepsilon}_j)\,K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)\right)\\
&= \frac{\bar{\varepsilon}_m}{(m-1)h}K\left(\frac{X_{(m-2)r\Delta}-x}{h}\right)-\frac{\bar{\varepsilon}_1}{(m-1)h}K\left(\frac{X_0-x}{h}\right)\\
&\quad+\frac{1}{(m-1)h}\sum_{j=2}^{m-1}\bar{\varepsilon}_j\left(K\left(\frac{X_{(j-2)r\Delta}-x}{h}\right)-K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)\right).
\end{aligned}
$$

Hence, the L^1-distance can be bounded by

$$
\begin{aligned}
E[|B_{n,3}(x)|] &\le \frac{1}{(m-1)h}E\left[\left|\bar{\varepsilon}_m\cdot K\left(\frac{X_{(m-2)r\Delta}-x}{h}\right)\right|\right]\\
&\quad+\frac{1}{(m-1)h}E\left[\left|\bar{\varepsilon}_1\cdot K\left(\frac{X_0-x}{h}\right)\right|\right]\\
&\quad+\frac{1}{(m-1)h}\sum_{j=2}^{m-1}E\left[|\bar{\varepsilon}_j|\cdot\left|K\left(\frac{X_{(j-2)r\Delta}-x}{h}\right)-K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)\right|\right]\\
&\le \frac{2||K||_\infty}{(m-1)h}E[|\bar{\varepsilon}_m|]+\frac{||K'||_\infty}{(m-1)h^2}\sum_{j=2}^{m-1}E\left[|\bar{\varepsilon}_j|\cdot|X_{(j-2)r\Delta}-X_{(j-1)r\Delta}|\right]\\
&\lesssim \frac{1}{mhr^{1/2}}+\frac{1}{mh^2}\sum_{j=1}^{m-2}\frac{1}{r^{1/2}}(r\Delta)^{1/2}=O\left(\frac{1}{mhr^{1/2}}+\frac{\Delta^{1/2}}{h^2}\right).
\end{aligned}
$$

Again, by assumption A4, it holds that

$$
\sqrt{n\Delta h}B_{n,3}(x)=O_P\left(\frac{(n\Delta h)^{1/2}}{(m-1)hr^{1/2}}+\frac{(n\Delta h)^{1/2}\Delta^{1/2}}{h^2}\right)=o_P(1),\ \text{as } n\to\infty.
$$

Term $B_{n,2}(x)$ can be decomposed in an analogous manner:

$$
\begin{aligned}
B_{n,2}(x) &= \frac{1}{(m-1)h}\sum_{j=1}^{m-1}\left(\left(X_{(j+1)r\Delta}-X_{jr\Delta}\right)-\left(X_{jr\Delta}-X_{(j-1)r\Delta}\right)\right)K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)\\
&= \frac{(X_{mr\Delta}-X_{(m-1)r\Delta})}{(m-1)h}K\left(\frac{X_{(m-2)r\Delta}-x}{h}\right)+\frac{(X_0-X_{r\Delta})}{(m-1)h}K\left(\frac{X_0-x}{h}\right)\\
&\quad+\frac{1}{(m-1)h}\sum_{j=1}^{m-2}(X_{(j+1)r\Delta}-X_{jr\Delta})\cdot\left(K\left(\frac{X_{(j-1)r\Delta}-x}{h}\right)-K\left(\frac{X_{jr\Delta}-x}{h}\right)\right).
\end{aligned}
$$

Now, the L^1-distance can be bounded by

$$
\begin{aligned}
E[|B_{n,2}(x)|] &\le \frac{||K||_\infty}{(m-1)h}\left(E[|X_{mr\Delta}-X_{(m-1)r\Delta}|]+E[|X_0-X_{r\Delta}|]\right)\\
&\quad+\frac{||K'||_\infty}{(m-1)h^2}\sum_{j=1}^{m-2}E[|X_{(j+1)r\Delta}-X_{jr\Delta}|\cdot|X_{(j-1)r\Delta}-X_{jr\Delta}|]\\
&\lesssim \frac{(r\Delta)^{1/2}}{(m-1)h}+\frac{(r\Delta)}{h^2}.
\end{aligned}
$$

Finally, assumption A4 guarantees that $B_{n,2}(x)$ converges fast enough to zero in probability:

$$\sqrt{n\Delta h}B_{n,2}(x) = O_P\left(\frac{(n\Delta h)^{1/2}(r\Delta)^{1/2}}{(m-1)h} + \frac{(n\Delta h)^{1/2}r\Delta}{h^2}\right) = o_P(1), \text{ as } n \to \infty.$$

The remaining term $B_{n,1}(x)$ can be decomposed according to

$$\begin{aligned} B_{n,1}(x) &= \frac{1}{(m-1)h}\sum_{j=1}^{m-1}\left(\left((\bar{X}_{j+1} - \bar{X}_j) - (X_{(j+1)r\Delta} - X_{jr\Delta})\right)K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right)\right) \\ &= \frac{(X_{r\Delta} - \bar{X}_1)}{(m-1)h}K\left(\frac{X_0 - x}{h}\right) + \frac{(\bar{X}_m - X_{(m-1)r\Delta})}{(m-1)h}K\left(\frac{X_{(m-1)r\Delta} - x}{h}\right) \\ &+ \frac{1}{(m-1)h}\sum_{j=1}^{m-2}\left(\bar{X}_{j+1} - X_{(j+1)r\Delta}\right)\left(K\left(\frac{X_{(j-1)r\Delta} - x}{h}\right) - K\left(\frac{X_{jr\Delta} - x}{h}\right)\right). \end{aligned}$$

Following this decomposition, the L^1-distance can be bounded by

$$E[|B_{n,1}(x)|] \lesssim \frac{(r\Delta)^{1/2}}{mh} + \frac{(r\Delta)}{h^2}.$$

Hence,

$$B_{n,1}(x) = o_P((n\Delta h)^{-1/2}), \text{ as } n \to \infty,$$

which finishes the proof. □

Acknowledgments

This work has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, project C5) of the German Research Foundation (DFG), which is gratefully acknowledged.

7. References

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

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

Barndorff-Nielsen, O.E., Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*. **2**, 1. 1-37.

Carr, P., Geman, H., Madan, D.B., Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business*. **75**. 305-332.

Carr, P., Wu, L. (2003). The finite moment log stable process and option pricing. *Journal of Finance*. **58**. 753-777.

Comte, F., Genon-Catalot, V., Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*. **13**, No. 2. 514-543.

Comte, F., Genon-Catalot, V., Rozenholc, Y. (2009). Nonparametric adaptive estimation for integrated diffusions. *Stochastic Processes and Their Applications.* **119**, 3. 811-834.

Comte, F., Genon-Catalot, V., Rozenholc, Y. (2010). Nonparametric estimation for a stochastic volatility model. *Finance and Stochastics* **14**, No. 1. 49-80.

Das, S. (2002). The surprise element: jumps in interest rates. *Journal of Econometrics.* **96**. 113-143.

Eberlein, E., Keller, U. (1995). Hyperbolic distributions in fincance. *Bernoulli.* **1**. 281-299.

Greenwood, P.E., Heckman, N., Lee, W., Wefelmeyer, W. (2015). Preaveraged kernel estimators for the drift function of a diffusion process in the presence of microstructure noise. Preprint.

Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications.* **79**, 1. 135-163.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications.* **119**, 2249-2276.

Johannes, M. (2004). The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *Statistics and Computing.* **3**. 135-146.

Jones, C.S. (2003). Nonlinear mean reversion in the short-term interest rate. *Rev. Financial Stud.* **16**. 793-843.

Kanaya, S., Kristensen, D. (2015). Estimation of stochastic volatility models by nonparametric filtering. The Institute for Fiscal Studies, Department of Economics, UCL. Cemmap Working Paper CWP 09/15.

Madan, D.B., Carr, P.P. Chang, E.E. (1998). The variance gamma process and option pricing. *European Finance Review.* **2**. 79-105.

Mancini, C., Renò, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics.* **160**, 1. 77-92.

Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications.* **117**, 1. 35-56.

Podolskij, M., Vetter, M. (2006). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Technical report, Ruhr-Universität Bochum.

Revuz, D., Yor, M. (1999). Continuous Martingales and Brownian Motion. 3rd Edition, Springer, Berlin.

Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and Their Applications* **124**, No. 1. 883-914.

Shimizu, Y. (2008). Some remarks on estimation of diffusion coefficients for jump-diffusions from finite samples. *Bulletin of Informatics and Cybernetics.* **40**. 51-60.

Strauch, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Processes and Their Applications* **125**, Issue 7. 2562-2602.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noise High-Frequency Data. *Journal of the American Statistical Association.* **100**, 1394-1419.