Bandwidth Selection for Continuous-Time Markov Processes*

Federico M. Bandi,* Valentina Corradi,** and Guillermo Moloche***
*Johns Hopkins University, **University of Warwick, and ***University of Chicago

October 2009

Abstract

We propose a theoretical approach to bandwidth choice for continuous-time Markov processes. We do so in the context of stationary and nonstationary processes of the recurrent kind. The procedure consists of two steps. In the first step, by invoking local Gaussianity, we suggest an automated bandwidth selection method which maximizes the probability that the standardized data are a collection of normal draws. In the case of diffusions, for instance, this procedure selects a bandwidth which only ensures consistency of the infinitesimal variance estimator, not of the drift estimator. Additionally, the procedure does not guarantee that the rate conditions for asymptotic normality of the infinitesimal variance estimator are satisfied. In the second step, we propose tests of the hypothesis that the bandwidth(s) are either "too small" or "too big" to satisfy *all* necessary rate conditions for consistency and asymptotic normality. The suggested statistics rely on a randomized procedure based on the idea of conditional inference. Importantly, if the null is rejected, then the first-stage bandwidths are kept. Otherwise, the outcomes of the tests indicate whether larger or smaller bandwidths should be selected. We study scalar and multivariate diffusion processes, jump-diffusion processes, as well as processes measured with error as is the case, for instance, for stochastic volatility modelling by virtue of preliminary high-frequency spot variance estimates. The finite sample joint behavior of our proposed automated bandwidth selection method, as well as that of the associated (second-step) randomized procedure, are studied via Monte Carlo simulation.

Keywords: Bandwidth selection, recurrence, Continuous-time Markov processes.

*We are grateful to Christian-Yann Robert for his useful comments. We thank Marine Carrasco, Jean-Marie Dufour, Andrew Patton, Anders Rahbek, Eric Renault, and Olivier Scaillet for suggestions. We also thank the seminar participants at the Universitè de Montreal, the Oxford-Man Institute, the Cass Conference "What went wrong? Financial Engineering, Financial Econometrics, and the Current Stress" (December 5-6, 2008, London), the "Toulouse Financial Econometrics conference" (May 15-16, 2009, Toulouse), the Stevanovich Center-CREATES conference "Financial Econometrics and Statistics: Current Themes and New Directions" (June 4-6, 2009, Skagen, Denmark), the First European SoFiE conference (June 10-12, 2009, Geneva, Switzerland) and the Conference on Recent Developments in Time Series (September 14-15, 2009, Granger Centre, University of Nottingham, UK) for discussions. Daniel Wilhelm provided outstanding research assistance. For some of our procedures, the codes are available on our websites.

1 Introduction

Following influential, early work on fully nonparametric infinitesimal volatility estimation and testing for scalar diffusion processes (e.g., Brugiére, 1991, Corradi and White, 1999, Florens-Zmirou, 1993, and Jacod 1997), the recent nonparametric literature in continuous time has largely focused on the full system. Emphasis might, for instance, be also placed on the estimation of the first infinitesimal moment (the drift) in the diffusion case (Stanton, 1997, among others) and, in the case of jump-diffusions, on the high-order infinitesimal moments (Johannes, 2004, *inter alia*).

Motivated by the need to completely characterize the system's dynamics, Bandi and Phillips (2003) have established consistency and asymptotic (mixed) normality for Nadaraya-Watson kernel estimators of both the drift and the diffusion function of recurrent (and, hence, possibly nonstationary) scalar diffusion processes (see, also, Fan and Zhang, 2003, and Moloche, 2004, for local polynomial estimates under stationarity and recurrence, respectively). Their results rely on a double asymptotic design in which the interval between discretely-sampled observations approaches zero, *in-fill* asymptotics, and the time span diverges to infinity, *long-span* asymptotics. A significant difference between a stationary (or positive recurrent) diffusion and a nonstationary (or null recurrent) one is that in the former case the local time grows linearly with the time span, while in the latter case it grows at a slower (and, generally, unknown) rate. Because the rate of divergence of local time affects the rate of convergence of the functional estimates of the process moments, this observation is theoretically, and empirically, important. Bandi and Moloche (2004) have generalized the results in Bandi and Phillips (2003) to the case of multidimensional diffusion processes. Importantly, in the multidimensional case a well-defined notion of local time no longer exists and one has to rely on the more general notion of occupation density. In both the scalar and the multidimensional case, consistency and (mixed) normality of the drift and variance estimator (and, hence, of the full system's dynamics) rely on the proper choice of the bandwidth parameters, i.e., on the rate at which the bandwidths approach zero as the interval between discretely-sampled observations goes to zero and the corresponding occupation densities (or local times, in the scalar case) diverge to infinity.

Admittedly, in the context of the functional estimation of continuous-time Markov models, the appropriate choice of window width is a largely unresolved issue. While it is recognized that infinitesimal conditional moment estimation in continuous time and conditional moment estimation in discrete time impose different requirements on the optimal window width for estimation accuracy (see, e.g., Bandi and Phillips, 2003, and Bandi and Moloche, 2004, for discussions), there is an overwhelming tendency in the continuous-time literature to employ bandwidth selection methods which can only be justified in more traditional set-ups of the regression type. Cross-validation procedures applied to the estimation of the drift and infinitesimal variance of scalar diffusion processes are typical examples. Yet, to the best of our knowledge, even in the stationary case, no theoretical discussion has been provided to automatically select the window width in continuous-time models of the types routinely used in the nonparametric finance literature. Furthermore, for both discrete and continuous-time processes, bandwidth selection is particularly delicate in the null recurrent (nonstationary) case since, as said, the bandwidth's vanishing rate ought to depend on the divergence rate of the number of visits to open sets in the range of the process but the latter is unknown, in general. In discrete time, important progress on the issue of bandwidth selection has been made by Karlsen and Tjostheim (2001) for β-null recurrent processes and by Guerre (2004) for general recurrent processes. The continuous-time case poses additional complications in that not only one has to adapt to the level of recurrence in the estimation domain but, also, to the rate at which the interval between discretely-sampled observations vanishes asymptotically.

This paper attempts to fill this important gap in the continuous-time econometrics literature by proposing a theoretical approach to automated bandwidth choice. The approach is designed for widely-employed classes of continuous-time Markov processes, such as scalar and multivariate diffusion processes and jump-diffusion processes, and is justified under mild assumptions on their statistical properties, stationarity not being required. Our solution to the problem is novel and may also be applied to discrete-time models, as outlined in Section 8.

In the diffusion case, the intuition of our approach is as follows. Consider kernel estimates of drift and diffusion function ($\widehat{\mu}_{h^{dr}}$ and $\widehat{\sigma}_{h^{dif}}$). Assume these estimates are obtained by selecting different smoothing sequences. Invoking the local Gaussianity property which diffusion models readily imply as a useful prior on the distributional feature of the standardized data, we maximize the probability that the standardized data $\left(\frac{(X_{t+\Delta}-X_t)-\widehat{\mu}_{h^{dr}}(X_t)\Delta}{\widehat{\sigma}_{h^{dif}}(X_t)\sqrt{\Delta}}\right)$ is a collection of draws from a Gaussian distribution by choosing the relevant smoothing sequences (h^{dr} and h^{dif}) accordingly. This procedure selects a bandwidth h^{dif} which ensures the consistency of the infinitesimal variance estimator but, in spite of its sound empirical performance (see Section 7), does not select a bandwidth h^{dr} which ensures the theoretical consistency of the drift function. Also, the automatically-chosen bandwidths do not necessarily satisfy the rate conditions required for (mean zero) asymptotic normality. To overcome this issue, for each infinitesimal moment, we propose a test of the null hypothesis that one or more rate conditions (for consistency and normality) are violated versus the alternative that all rate conditions are satisfied. The suggested statistics (separately specified for drift and diffusion) rely on a randomized procedure based on the idea of conditional inference, along the lines of Corradi and Swanson (2006). If the null is rejected, then the selected bandwidth is kept, otherwise the outcome of the procedure suggests whether we should select a larger or a smaller bandwidth. We proceed sequentially, until the null is rejected. Because the probability of rejecting the null when the it is false is asymptotically one at each step, our approach does not suffer from a sequential bias problem.

Our emphasis on recurrence is empirically-motivated, theoretical generality being only a by-product. Under general recurrence properties, the bandwidth's rate conditions are not a function of T (the time span or the number of observations) as in stationary time-series analysis. They are a function of the number of visits to each level at which functional estimation is conducted. Importantly, however, even for stationary processes (which are, as emphasized, a sub-case of the class of recurrent processes) choosing the bandwidth rate as a function of the empirical occupation times is bound to provide a more objective solution to the bandwidth selection problem than choosing it based on a theoretical (and, hence, purely hypothetical) divergence rate of the occupation times equal to T. This point is, of course, particularly compelling when dealing with highly dependent, but possibly stationary, time-series of the type routinely encountered in fields such as finance. These processes return to values in their range very slowly and, thus, even though they may be stationary, have occupation densities which hardly diverge at the "theoretical" T rate.

We begin by considering the case of bandwidth selection for scalar diffusion models (Section 2). We then extend our analysis to scalar jump-diffusion processes (Section 3). The case of a diffusion observed with error is presented in Section 4. Stochastic variance processes filtered from high-frequency financial data may, of course, be regarded as processes observed with error. We evaluate the case of stochastic volatility explicitly and discuss bandwidth selection for diffusion models applied to market microstructure noise-contaminated spot variance estimates in Section 5. In Section 6 we study the multivariate diffusion case. Section 7 provides a Monte Carlo study. Section 8 contains final remarks. All proofs are collected in the Appendix.

2 Scalar diffusion processes

2.1 The framework

We consider the following class of one-factor models,

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,$$

where $\{W_t : t = 1, ..., T\}$ is a standard Brownian motion. Our objective is to provide suitable nonparametric estimates of the drift term $\mu(a)$ and of the infinitesimal variance $\sigma^2(a)$. To this extent, we assume availability of a sample of N equidistant observations and denote the discrete interval between two successive observations as $\Delta_{N,T} = T/N$, where T defines the time span. Specifically, we observe the diffusion skeleton $X_{\Delta_{N,T}}, X_{2\Delta_{N,T}}, ..., X_{N\Delta_{N,T}}$. In what follows, we require $N, T \to \infty$, $\Delta_{N,T} \to 0$ (in-fill asymptotics), and $T = \Delta_{N,T}N \to \infty$ (long-span asymptotics) for consistency of the moment estimates. As in Stanton (1997), Bandi and Phillips (2003), and Johannes (2004), *inter alia*, we construct the following estimators of the drift and infinitesimal variance, respectively:

$$\widehat{\mu}_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} K\left(\frac{X_{j\Delta_{N,T}}-a}{h^{dr}_{N,T}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h^{dr}_{N,T}}\right)}, \tag{1}$$

and

$$\widehat{\sigma}^2_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} K\left(\frac{X_{j\Delta_{N,T}}-a}{h^{dif}_{n,T}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)^2}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h^{dif}_{N,T}}\right)}. \tag{2}$$

We denote by $\mathbf{h} = \left(h^{dr}_{N,T}, h^{dif}_{N,T}\right) \in H \subset R^2_+$ a bivariate vector bandwidth belonging to the set H contained in the positive plane R^2_+. This vector is our object of econometric interest. Assumption 1 guarantees existence of a unique, recurrent solution to X. Assumption 2 outlines the conditions imposed on the kernel function $K(.)$ in Eqs. (1) and (2). The same conditions on the kernel function are also employed in the following sections.

Assumption 1.

(i) *$\mu(.)$ and $\sigma(.)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$, where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions are at least twice continuously differentiable. Hence, they satisfy local Lipschitz and growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_1^J and C_2^J so that, for all x and y in J,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1^J |x - y|,$$

and

$$|\mu(x)| + |\sigma(x)| \leq C_2^J \{1 + |x|\}.$$

(ii) *$\sigma^2(.) > 0$ on $\mathfrak{D}$.*

(iii) *We define $S(\alpha)$, the natural scale function, as*

$$S(\alpha) = \int_c^{\alpha} \exp\left\{ \int_c^y \left[-\frac{2\mu(x)}{\sigma^2(x)} \right] dx \right\} dy,$$

where c is a generic fixed number belonging to $\mathfrak{D}$. We require $S(\alpha)$ to satisfy

$$\lim_{\alpha \to l} S(\alpha) = -\infty.$$

and

$$\lim_{\alpha \to u} S(\alpha) = \infty.$$

Assumption 2. *The kernel $K(.)$ is a continuously differentiable, symmetric and nonnegative function whose derivative $K'(.)$ is absolutely integrable and for which*

$$\int_{-\infty}^{\infty} K(s) ds = 1, \quad \mathbf{K}_2 = \int_{-\infty}^{\infty} K^2(s) ds < \infty, \quad \sup_s K(s) < C_3,$$

and

$$\int_{-\infty}^{\infty} s^2 K(s) ds < \infty.$$

In what follows, the symbol $\overline{L}_X(T, a)$ denotes the chronological local time of X at T and a, i.e., the number of calendar time units spent by the process around a in the time interval $[0, T]$.

Proposition 1 (Bandi and Phillips, 2003): Let Assumptions 1 and 2 hold.

(i) Let $\overline{\Delta}_{N,\overline{T}} = \overline{T}/N$ with $\overline{T}$ fixed. If $\lim_{N\to\infty} \frac{1}{h_{N,\overline{T}}} \left(\overline{\Delta}_{N,\overline{T}} \log \frac{1}{\overline{\Delta}_{N,\overline{T}}} \right)^{1/2} \to 0$, then

$$\widehat{\overline{L}}_X(\overline{T}, a) - \overline{L}_X(\overline{T}, a) = o_{a.s.}(1),$$

where $\widehat{\overline{L}}_X(\overline{T}, a) = \frac{\overline{\Delta}_{N,\overline{T}}}{h_{N,\overline{T}}} \sum_{j=1}^N K\left(\frac{X_{j\overline{\Delta}_{N,\overline{T}}} - a}{h_{N,\overline{T}}} \right)$.

- *The drift estimator*

Let (ii) $h_{N,T}^{dr}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ and (iii) $\frac{\overline{L}_X(T,a)}{h_{N,T}^{dr}}\left(\Delta_{N,T}\log\frac{1}{\Delta_{N,T}}\right)^{1/2} \overset{a.s.}{\to} 0$, then:

$$\widehat{\mu}_{N,T}(a)-\mu(a)=o_{a.s.}(1).$$

Further, if (iv) $h_{N,T}^{dr,5}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$, then:

$$\sqrt{h_{N,T}^{dr}\widehat{\overline{L}}_X(T,a)}\left(\widehat{\mu}_{N,T}(a)-\mu(a)\right) \Rightarrow N\left(0,\mathbf{K}_2\sigma^2(a)\right).$$

- *The diffusion estimator*

If (iii) holds with $h_{N,T}^{dr}$ replaced by $h_{N,T}^{dif}$, then:[1]

$$\widehat{\sigma}_{N,T}^2(a)-\sigma^2(a)=o_{a.s.}(1).$$

Further, if (iv') $\frac{h_{N,T}^{dif,5}\overline{L}_X(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0$, then:

$$\sqrt{\frac{h_{N,T}^{dif}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}}}\left(\widehat{\sigma}_{N,T}^2(a)-\sigma^2(a)\right) \Rightarrow N\left(0,2\mathbf{K}_2\sigma^4(a)\right).$$

It is evident from the proposition above (as well as classical logic based on nonparametric moment estimation in discrete time) that consistency and asymptotic normality of the drift and variance estimator crucially rely on appropriate choice of the smoothing parameter(s). To this extent, two issues ought to be addressed. First, usual data-driven methods often employed in empirical work in continuous-time finance, such as cross-validation, are not theoretically justified and may not necessarily work in the presence of in-fill asymptotics and nonstationarity. Second, while in the positive recurrent case $\overline{L}_X(T,a)/T \overset{p}{\to} f_X(a)$, where $f_X(a)$ denotes the stationary probability density at a of the process X, in the null recurrent case $\overline{L}_X(T,a)/T \overset{p}{\to} 0$. Under null recurrence, as emphasized earlier, $\overline{L}_X(T,a)$ grows at a (generally unknown) rate which is slower than T.[2] Since the bandwidth's vanishing rate depends on this unknown rate, appropriate bandwidth selection in the null recurrent case is particularly delicate.

We shall proceed in two steps. In the first step, we introduce an adaptive bandwidth selection method which ensures consistency of the diffusion estimator but only guarantees that $\widehat{\mu}_{N,T}(a)-\mu(a)=o_p\left(\Delta_{N,T}^{-1/2}\right)$. In the second step, we employ a randomized procedure to test whether the bandwidth selected in the first stage violate any of the rate conditions (ii)-(iii)-(iv) for the drift and (iii)-(iv') for the diffusion. This second step is conducted separately for drift and diffusion. Should we reject the null, then we would rely on the previously-chosen bandwidth. Alternatively, because the outcome of the procedure gives us information about whether the selected bandwidth is too small or too large, we iterate until the null is rejected.

[1]Note that (iii) ensures that $\frac{h_{N,T}^{dif}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}} \to \infty$.

[2]The Brownian motion case is an exception for which the rate is known and $\overline{L}_X(T,a)/\sqrt{T}=O_p(1)$.

2.2 First step: A residual-based procedure

Consider the estimated residual series

$$\left\{ \widehat{\varepsilon}_{i\Delta_{N,T}} = \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \ : \ i = 2, ..., \Delta_{N,T}^{-1}\overline{T} \right\},$$

assuming, for notational simplicity, that $\Delta_{N,T}^{-1}$ is an integer. In light of the normality of the driving Brownian motion, over small time intervals $\Delta_{N,T}$ the residual series is roughly standard normally distributed. Our minimization problem requires finding

$$\widehat{\mathbf{h}}_{N,T} \in H \subset R_+^2 : \rho\left(F_{\overline{N}}^{\widehat{\mathbf{h}}_{N,T}}, \Phi\right) = \theta_{\overline{N}} \tag{3}$$

with $\theta_{\overline{N}} \downarrow 0$ as $\overline{N} = \Delta_{N,T}^{-1}\overline{T} \to \infty$, where $F_{\overline{N}}^{\widehat{\mathbf{h}}_{N,T}}$ denotes the empirical cumulative distribution of the estimated residuals $\widehat{\varepsilon}_{i\Delta_{N,T}}$, Φ is the cumulative distribution of the standard normal random variable, and $\rho\,(.,.)$ is a distance metric.

It is noted that the criterion is defined over a fixed time span $\overline{T}$ whereas the estimators, mainly for consistency of the drift, are defined over an enlarging span of time T. We define the criterion over a fixed time span to avoid theoretical imbalances in the case of nonstationary diffusions. This point is discussed in Bandi and Phillips (2007). From an empirical standpoint, fixing the sample span over which the criterion is minimized and enlarging the time span over which the nonparametric estimators are computed is immaterial. It simply amounts to splitting the sample into two parts, i.e. $(0, \overline{T}]$ and $(\overline{T}, T]$. The entire sample (from 0 to T) is used to compute $\widehat{\mu}_{N,T}(.)$ and $\widehat{\sigma}_{N,T}(.)$. The first part of the sample (from 0 to $\overline{T}$) is used to define the minimization problem.[3]

We focus on the Kolmogorov-Smirnov distance, but a different distance measure may, of course, be employed. We define the target bandwidth sequence $\mathbf{h}_{N,T}^* = (h_{N,T}^{dr}, h_{N,T}^{dif})^*$ as the bandwidth sequence which guarantees that the empirical distribution function of the standardized data converges uniformly to the standard normal distribution function as $N, T \to \infty$ with $\frac{T}{N} \to 0$ (and, of course, with $\overline{N} = \overline{T}\Delta_{N,T}^{-1} \to \infty$). We will first characterize its properties (in Theorem 1). Subsequently, we will show that it exists and that $\widehat{\mathbf{h}}_{N,T}$ is asymptotically equivalent to it (in Theorem 2).

Theorem 1. *A vector bandwidth* $\mathbf{h}_{N,T}^* = (h_{N,T}^{dr}, h_{N,T}^{dif})^*$ *satisfies*

$$\mathbf{h}_{N,T}^* = \mathbf{h} \in H : \sup_x \left| F_{\overline{N}}^{\mathbf{h}}(x) - \Phi(x) \right| \underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\to}} 0 \tag{4}$$

if and only if

[3] This statement can easily be reconciled with our theoretical framework. Assume $T = \sqrt{N}$, for instance. Then, the observations are equispaced at $\left\{\frac{1}{\sqrt{N}}, \frac{2}{\sqrt{N}}, ..., 1, 1 + \frac{1}{\sqrt{N}}, ..., \sqrt{N}\right\}$ since $\frac{T}{N} = \frac{1}{\sqrt{N}}$. We can now split the sample in two parts, namely observations in $(0, \overline{T}]$ and observations in $(\overline{T}, T]$. Assume, without loss of generality, that $\overline{T} = 1$. Also, assume that there are $\overline{N}$ equispaced observations in the first part of the sample. Then, $\frac{1}{\overline{N}} = \frac{1}{\sqrt{N}}$. This implies that the number of observations in the first part of the sample, which is defined over a fixed time span $\overline{T}$, grows with $\sqrt{N}$, whereas the number of observations in the second part of the sample grows with N. In practice one can choose $\overline{T}$ relatively large.

$$\sup_{a\in\mathfrak{D}}\left|\widehat{\mu}_{N,T}\left(a,h_{N,T}^{dr}\right)-\mu(a)\right|=o_p\left(\frac{1}{\sqrt{\Delta_{N,T}}}\right), \tag{5}$$

and

$$\sup_{a\in\mathfrak{D}}\left|\widehat{\sigma}_{N,T}\left(a,h_{N,T}^{dif}\right)-\sigma(a)\right|=o_p\left(1\right). \tag{6}$$

■

Theorem 2. *Let Assumptions 1 and 2 hold. (i) There exists a vector bandwidth* $\mathbf{h}_{N,T}^{*}=(h_{N,T}^{dr},h_{N,T}^{dif})^{*}$ *so that*

$$\mathbf{h}_{N,T}^{*}=\mathbf{h}\in H:\sup_{x}\left|F_{\overline{N}}^{\mathbf{h}}(x)-\Phi(x)\right|\underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\to}}0 \tag{7}$$

and

$$\mathbf{h}_{N,T}^{*}=\left(h_{N,T}^{dr},h_{N,T}^{dif}\right)^{*}\underset{N,T\to\infty,\Delta_{N,T}\to 0}{\to}0.$$

(ii) If

$$\widehat{\mathbf{h}}_{N,T}=\mathbf{h}\in H:\sup_{x}\left|F_{\overline{N}}^{\mathbf{h}}(x)-\Phi(x)\right|=\theta_{\overline{N}} \tag{8}$$

with $\theta_{\overline{N}}\downarrow 0$ *as* $\overline{N}\to\infty$, *then*

$$\widehat{\mathbf{h}}_{N,T}/\mathbf{h}_{N,T}^{*}\underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\to}}1.$$

■

Theorem 2 guarantees the existence of a bandwidth vector $\widehat{\mathbf{h}}_{N,T}$ ensuring that our proposed criterion has a solution. This solution guarantees uniform consistency (in probability) of the variance estimator but, despite being empirically very sensible as we show below through simulations (see Section 7), fails to guarantee theoretical consistency of the drift estimator. In addition, the selected diffusion bandwidth does not ensure asymptotic normality of the diffusion estimator. A second procedure is therefore needed in order to verify whether the resulting bandwidths satisfy *all* rate conditions needed for consistency and asymptotic normality of *both* estimators.

Given Proposition 1, we now need to check whether $h_{N,T}^{dr}$ is small enough as to satisfy $h_{N,T}^{dr,5}\overline{L}_X(T,a)\overset{a.s.}{\to}0\ \forall a\in\mathfrak{D}$ and large enough as to satisfy $\min\left\{h_{N,T}^{dr}\overline{L}_X(T,a),\frac{h_{N,T}^{dr}}{(\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\overline{L}_X(T,a)}\right\}\overset{a.s.}{\to}\infty\ \forall a\in\mathfrak{D}$. Similarly, we need to check whether $h_{N,T}^{dif}$ is small enough as to satisfy $\frac{h_{N,T}^{dif,5}\overline{L}_X(T,a)}{\Delta_{N,T}}\overset{a.s.}{\to}0\ \forall a\in\mathfrak{D}$ and large enough as to satisfy $\frac{h_{N,T}^{dif}}{(\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\overline{L}_X(T,a)}\overset{a.s.}{\to}\infty\ \forall a\in\mathfrak{D}$.

2.3 Second step: A randomized procedure

Let $\widehat{\mathbf{h}}_{N,T} = \left(\widehat{h}^{dr}_{N,T}, \widehat{h}^{dif}_{N,T}\right)$ be defined as $\widehat{\mathbf{h}}_{N,T} = \arg\min_{\mathbf{h}} \left| F^{\mathbf{h}}_{N}(x) - \Phi(x) \right|$. We begin by verifying whether $\widehat{h}^{dr}_{N,T}$ satisfies conditions (ii), (iii), and (iv) in Proposition 1. Next, we will turn to $\widehat{h}^{dif}_{N,T}$, whose requirements are slightly different.

It is immediate to see that (ii) and (iii) require the bandwidth not to approach zero too fast, thus only one of the two is binding. Condition (iv) instead requires the bandwidth to approach zero fast enough. It is important to rule out the possibility of a bandwidth which is too large to satisfy (iv) and too small to satisfy the most stringent between (ii) and (iii). To this extent, we only ought to provide primitive conditions on N and T. If (iv) is violated, then $h^{dr}_{N,T}$ goes to zero not faster than $\overline{L}_X(T,a)^{-1/5}$. This ensures that (ii) is satisfied, but does not ensure that (iii) is satisfied. For (iii) to be satisfied when (iv) is not, we need $\overline{L}_X(T,a)^{6/5}\Delta^{1/2}_{N,T}\log(1/\Delta_{N,T}) \to 0$. Because $\overline{L}_X(T,a)$ can grow at most at rate T, a sufficient condition is therefore $N/T^{17/5} \to \infty$.

Provided $N/T^{17/5} \to \infty$, there are three possibilities (see Figure 1). First, we have chosen the right bandwidth and thus $\widehat{h}^{dr}_{N,T}$ satisfies (ii), (iii), and (iv). Second, we have chosen too large a bandwidth, so that (ii) and (iii) hold, but (iv) is violated. Third, we have chosen too small a bandwidth, so that either (ii) or (iii) is violated (or both) but (iv) holds. Hence, at most one set of conditions can be violated, namely either (iv) or the most stringent between (ii) and (iii). To this extent, we consider the following hypotheses:

$$H^{dr}_0 \,:\, \widehat{h}^{dr,5}_{N,T}\widehat{\overline{L}}_X(T,a) \overset{a.s.}{\to} \infty \text{ or } \max\left\{\frac{1}{\widehat{h}^{dr}_{N,T}\widehat{\overline{L}}_X(T,a)}, \frac{\widehat{\overline{L}}_X(T,a)\Delta^{1/2}_{N,T}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}^{dr}_{N,T}}\right\} \overset{a.s.}{\to} \infty,$$

$$H^{dr}_A \,:\, \widehat{h}^{dr,5}_{N,T}\widehat{\overline{L}}_X(T,a) \overset{a.s.}{\to} 0 \text{ and } \max\left\{\frac{1}{\widehat{h}^{dr}_{N,T}\widehat{\overline{L}}_X(T,a)}, \frac{\widehat{\overline{L}}_X(T,a)\Delta^{1/2}_{N,T}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}^{dr}_{N,T}}\right\} \overset{a.s.}{\to} 0.$$

The null is that either $\widehat{h}^{dr,5}_{N,T}\widehat{\overline{L}}_X(T,a) \overset{a.s.}{\to} \infty$, (iv) is violated, or $\min\left\{\widehat{h}^{dr}_{N,T}\widehat{\overline{L}}_X(T,a), \frac{\widehat{h}^{dr}_{N,T}}{\widehat{\overline{L}}_X(T,a)\Delta^{1/2}_{N,T}\log^{1/2}(1/\Delta_{N,T})}\right\} \overset{a.s.}{\to} 0$, $(ii)\wedge(iii)$ is violated. Since it is impossible that neither $(ii)\wedge(iii)$ nor (iv) hold, the alternative is that both $(ii)\wedge(iii)$ and (iv) hold. Thus, if we reject the null, we can rely on $\widehat{h}^{dr}_{N,T}$ for drift estimation.

If, instead, we fail to reject the null, depending on which condition we fail to reject, we know whether we have chosen a bandwidth which is too small or one which is too large. Suppose that the selected bandwidth is too large, we proceed sequentially by choosing a smaller bandwidth until we reject the null. Because at all steps the probability of rejecting the null when it is wrong is asymptotically one, the procedure does not suffer from the well-known sequential bias issue.

Importantly, rejection of the null, as stated above, does not rule out the possibility that $\widehat{h}^{dr,5}_{N,T}\widehat{\overline{L}}_X(T,a) = O_p(1)$ (if $\widehat{h}^{dr}_{N,T} \propto \widehat{\overline{L}}_X(T,a)^{-1/5}$) or $\min\left\{\widehat{h}^{dr}_{N,T}\widehat{\overline{L}}_X(T,a), \frac{\widehat{h}^{dr}_{N,T}}{\widehat{\overline{L}}_X(T,a)\Delta^{1/2}_{N,T}\log^{1/2}(1/\Delta_{N,T})}\right\} = O_p(1)$ (if $\widehat{h}^{dr}_{N,T} \propto \widehat{\overline{L}}_X(T,a)^{-(1/5+\alpha/2)}$ with $\alpha > 0$ and $N \propto T^{17/5+\alpha}$ or if $\widehat{h}^{dr}_{N,T} \propto \widehat{\overline{L}}_X(T,a)^{-1}$). Also, it does not ensure that conditions (ii), (iii), and (iv) hold for all evaluation points $a \in \mathfrak{D}$. Hence, we re-formulate the

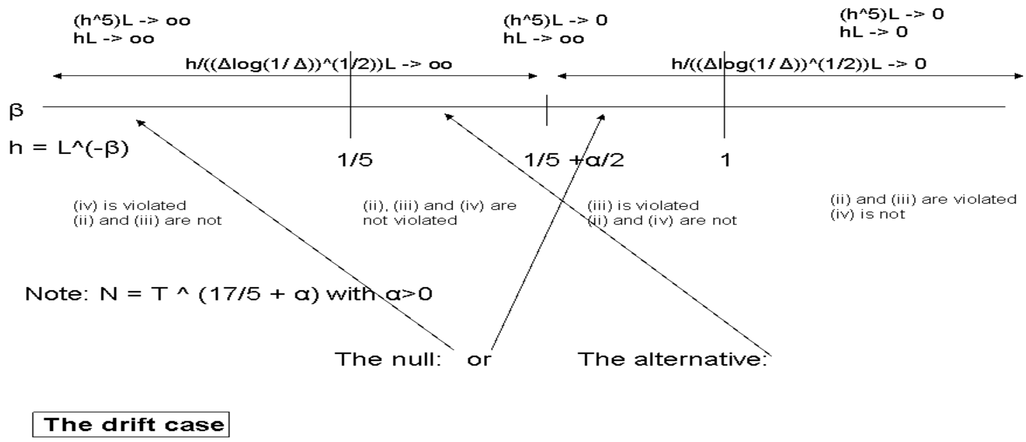

Figure 1: Graphical representation of the drift bandwidth test

hypotheses as follows:

$$H_0^{\prime,dr} \quad : \quad \int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \overset{a.s.}{\to} \infty$$

$$\text{or} \;\; \max\left\{ \frac{1}{\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(1+\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a}, \int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dr,(1+\varepsilon)}}\mathrm{d}a \right\} \overset{a.s.}{\to} \infty$$

for $\mathcal{A} \subset \mathfrak{D}$, and $\varepsilon > 0$ arbitrarily small, versus

$$H_A^{\prime,dr} : \text{ negation of } H_0^{\prime,dr}.$$

The role of the integral over $\mathcal{A}$, and of $\varepsilon > 0$, is to ensure that rejection of the null implies $\min\left\{ \int_{\mathcal{A}} \widehat{h}_{N,T}^{dr} \widehat{\overline{L}}_X(T,a)\mathrm{d}a, \frac{\widehat{h}_{N,T}^{dr}}{\int_{\mathcal{A}} \widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log(1/\Delta_{N,T})\mathrm{d}a} \right\} \overset{a.s.}{\to} \infty$ and $\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,5} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \overset{a.s.}{\to} 0$. However, of course, if we choose an ε which is not small enough, we run the risk of not having a bandwidth sequence for which $H_0^{\prime,dr}$ is rejected. Hereafter, we consider the following statistic:

$$V_{R,N,T} = \min\left\{ \widetilde{V}_{1,R,N,T}, \;\min\left\{ \widetilde{V}_{2,R,N,T}, \widetilde{V}_{3,R,N,T} \right\} \right\},$$

where for $i = 1, 2, 3$

$$\widetilde{V}_{i,R,N,T} = \int_U V_{i,R,N,T}^2(u)\pi(u)du,$$

with $U = [\underline{u}, \overline{u}]$ being a compact set, $\int_U \pi(u)du = 1$, $\pi(u) \geq 0$ for all $u \in U$, and

$$V_{i,R,N,T}(u) = \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\left\{ v_{i,j,N,T} \leq u \right\} - \frac{1}{2} \right)$$

and

$$v_{1,j,N,T} = \left(\exp \int_{\mathcal{A}} \left(\widehat{h}_{N,T}^{dr,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)\right)^{1/2} \eta_{1,j},$$

$$v_{2,j,N,T} = \left(\exp\left(\left(\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(1+\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)^{-1}\right)\right)^{1/2} \eta_{2,j},$$

$$v_{3,j,N,T} = \left(\exp\left(\int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dr,(1+\varepsilon)}}\mathrm{d}a\right)\right)^{1/2} \eta_{3,j}, \tag{9}$$

with $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3)^\intercal \sim \text{iid}N(0, I_{3R})$.

In what follows, let the symbols P^* and d^* denote convergence in probability and in distribution under P^*, which is the probability law governing the simulated random variables $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3$, i.e., a standard normal, conditional on the sample. Also, let E^* and Var^* denote the mean and variance operators under P^*. Furthermore, with the notation $a.s. - P$ we mean: for all samples but a set of measure 0.

Suppose that $\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \overset{a.s.}{\to} \infty$. Then, conditionally on the sample and $a.s. - P$, $v_{1,j,N,T}$ diverges to ∞ with probability $1/2$ and to $-\infty$ with probability $1/2$. Thus, as $N, T \to \infty$, for any $u \in U$, $1\{v_{1,j,N,T} \leq u\}$ will be distributed as a Bernoulli random variable with parameter $1/2$. Further note that as $N, T \to \infty$, for any $u \in U$, $1\{v_{1,j,N,T} \leq u\}$ is equal to either 1 or 0, regardless of the evaluation point u, and so as $N, T, R \to \infty$, for all $u, u' \in U$, $\frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\{v_{1,j,N,T} \leq u\} - \frac{1}{2}\right)$ and $\frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\{v_{1,j,N,T} \leq u'\} - \frac{1}{2}\right)$ will converge in d^*-distribution to the same standard normal random variable. Hence, $\widetilde{V}_{1,R,N,T} \overset{d^*}{\to} \chi_1^2$ $a.s. - P$. It is now immediate to notice that for all $u \in U$, $V_{1,R,N,T}^2(u)$ and $\widetilde{V}_{1,R,N,T}$ have the same limiting distribution. The reason why we are averaging over U is simply because the finite sample type I and type II errors may indeed depend on the particular evaluation point. As for the alternative, if $\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \overset{a.s.}{\to} 0$, (or, if $\int_{\mathcal{A}} \widehat{h}_{N,T}^{dr,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)\mathrm{d}a = O_{a.s.}(1)$), then $v_{1,j,N,T}$, as $N, T \to \infty$, conditionally on the sample and $a.s. - P$, will converge to a (mixed) zero mean normal random variable. Thus, $\frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\{v_{1,j,N,T} \leq u\} - \frac{1}{2}\right)$ will diverge to infinity at speed $\sqrt{R}$ if $u \neq 0$ $a.s. - P$.

Importantly, the two conditions stated in the null hypothesis are the negation of (ii), (iii), and (iv) in Proposition 1, respectively.[4] As mentioned, *only* one of the conditions stated under the null is false, simply because the criterion cannot select a bandwidth which is too small (for the most stringent between (ii) and (iii) to be satisfied) and, at the same time, too large (for (iv) to be satisfied). Hence, either $\widetilde{V}_{1,R,N,T}$ or $\min\left\{\widetilde{V}_{2,R,N,T}, \widetilde{V}_{3,R,N,T}\right\}$ has to diverge under the null. Thus, $\min\left\{\widetilde{V}_{1,R,N,T},\ \min\left\{\widetilde{V}_{2,R,N,T}, \widetilde{V}_{3,R,N,T}\right\}\right\}$, conditional on the sample, and for all samples but a set of measure zero, is asymptotically χ_1^2 under the null and diverges under the alternative. If we reject the null, then conditions (ii), (iii), and (iv) in Proposition 1 are satisfied. Otherwise, if, for instance, $\widetilde{V}_{1,R,N,T} = \min\left\{\widetilde{V}_{1,R,N,T},\ \min\left\{\widetilde{V}_{2,R,N,T}, \widetilde{V}_{3,R,N,T}\right\}\right\} \leq 3.84$ and we fail to reject the null, then $\widehat{h}_{N,T}^{dr}$ is

[4] It should be noted that the rate conditions in Proposition 1 are stated in terms of $\overline{L}_X(T,a)$ instead of $\widehat{\overline{L}}_X(T,a)$. However, $\frac{\widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dr}} \overset{a.s.}{\to} 0$ if, and only if, $\frac{\overline{L}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dr}} \overset{a.s.}{\to} 0$, but this ensures that $\widehat{\overline{L}}_X(T,a) - \overline{L}_X(\sup\{t : X_t = a\}, a) = o_{a.s.}(1)$ (Bandi and Phillips, 2003, Corollary 1).

too large (and condition (iv) is violated). The same testing procedure should therefore be repeated until

$$\widetilde{h}_{N,T}^{dr} = \max\left\{h < \widehat{h}_{N,T}^{dr} : \text{ s.t. } H_0' \text{ is rejected}\right\}.$$

In other words, the proposed procedure gives us a way to learn whether the conditions for consistency and (mean zero) mixed normality of the drift are satisfied. If they are not, it gives us a way to understand which condition is not satisfied and modify the bandwidth accordingly.

Theorem 3. Let Assumption 1 and 2 hold. Assume $T, N, R \to \infty$, $N/T^{17/5} \to \infty$, and $R/T \to 0$.[5]
(i) Under $H_0^{\prime,dr}$,

$$V_{R,N,T} \overset{d^*}{\to} \chi_1^2 \ a.s. - P.$$

(ii) Under $H_A^{\prime,dr}$, there are $\delta, \eta > 0$ so that

$$P^*\left(R^{-1+\eta} V_{R,N,T} > \delta\right) \to 1 \ a.s. - P.$$

■

The test has appealing features. Specification tests generally assume correct specification under the null. In our case, the bandwidth is correctly specified under the alternative. This is helpful in that, in theory, rejection of the null at the 5% level gives us 95% confidence that the alternative is true and the assumed bandwidth is correctly specified. Since we stop as soon as we reject the null, we do not have a sequential bias problem. Further, the critical values (those of a chi-squared random variable with 1 degree of freedom) are readily tabulated. Reliance on a classical distribution makes testing, as well as adaptation of the bandwidth in either direction should the null not be rejected, rather straightforward. It should be stressed that the limiting distribution in Theorem 3 is driven by the added randomness $\boldsymbol{\eta}$, conditional on the sample and for all samples but a set of measure zero. Nonetheless, whenever we reject the null, for all samples and for 95% of random draws $\boldsymbol{\eta}$, the alternative is true, and so keeping the selected bandwidth is the right choice.

We now turn to $h_{N,T}^{dif}$. We will ensure that $h_{N,T}^{dif}$ is small enough as to satisfy $\frac{h_{N,T}^{dif,5}\overline{L}_X(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0 \ \forall a \in \mathfrak{D}$, and large enough as to satisfy $\frac{h_{N,T}^{dif}}{(\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\overline{L}_X(T,a)} \to \infty$. In order to rule out the possibility that any bandwidth rate is either too slow to satisfy the former condition or too fast to satisfy the latter, it suffices to require that $N/T^5 \to \infty$.

We can now state the hypothesis of interest as:

$$H_0^{dif} : \int_{\mathcal{A}} \frac{\widehat{h}_{N,T}^{dif,(5-\varepsilon)}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}} da \overset{a.s.}{\to} \infty \text{ or } \int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dif,(1+\varepsilon)}} da \overset{a.s.}{\to} \infty$$

for $\mathcal{A} \subset \mathfrak{D}$, and $\varepsilon > 0$ arbitrarily small, versus

$$H_A' : \text{ negation of } H_0'.$$

[5] The condition $R/T \to 0$ is necessary only for the case in which the local time diverges at a logarithmic rate. If the local time diverges at rate T^a $a > 0$, then R can grow as fast as, or faster than, T. Thus, we drop the condition in the statement of Theorem 4.

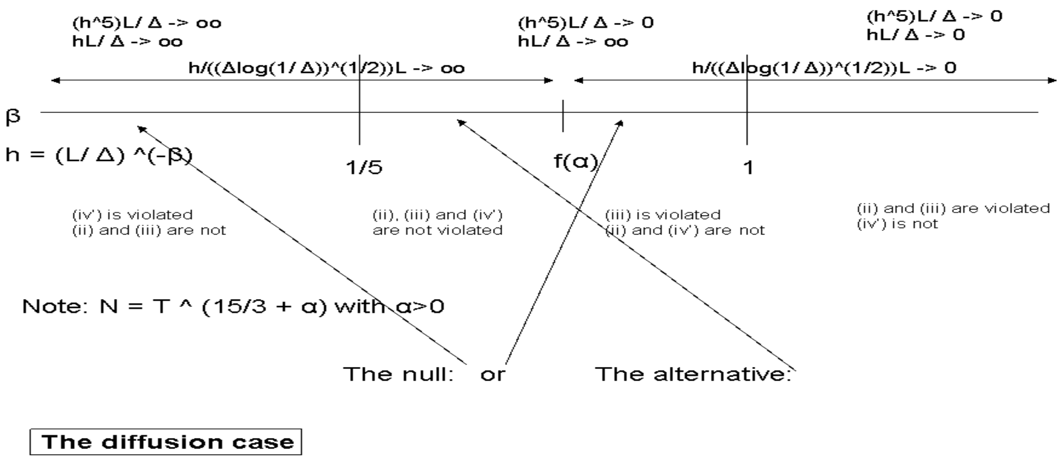

Figure 2: Graphical representation of the diffusion bandwidth test

Remark 1. We note that, contrary to the drift case, we are not writing the second condition in the null hypothesis as

$$\max\left\{\frac{\Delta_{N,T}}{\int_{\mathcal{A}}\widehat{h}_{N,T}^{dif,(1+\varepsilon)}\widehat{\overline{L}}_X(T,a)\mathrm{d}a},\int_{\mathcal{A}}\frac{\widehat{\overline{L}}_X(T,a)\Delta_{N,T}^{1/2}\log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dif,(1+\varepsilon)}}\mathrm{d}a\right\}\stackrel{a.s.}{\to}\infty. \tag{10}$$

In fact, in spite of the fact that $\frac{\widehat{h}_{N,T}^{dif}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}}$ is the rate of convergence of the diffusion estimator, we do not need to explicitly require its divergence (in Proposition 1, for example). If (iii) is satisfied for the diffusion estimator, then $\frac{\widehat{h}_{N,T}^{dif}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}}$ is guaranteed to diverge. In other words, the maximum in Eq. (10) is always the second term and the first term can be dropped. The graphical manifestation of this result is the fact that, in Figure 2, $f(a)<\frac{1}{2}$. In the case of the drift, the maximum may vary depending on α (see Figure 1). For instance, if α is larger than $\frac{8}{5}$, then the maximum condition is always $\frac{1}{\int_{\mathcal{A}}\widehat{h}_{N,T}^{dr}\widehat{\overline{L}}_X(T,a)\mathrm{d}a}$ since $\frac{1}{5}+\frac{\alpha}{2}>1$.

Consider the following statistic:

$$VD_{R,N,T}=\min\left\{\widetilde{VD}_{1,R,N,T},\ \widetilde{VD}_{2,R,N,T}\right\},$$

where for $i=1,2$

$$\widetilde{VD}_{i,R,N,T}=\int_U VD_{i,R,N,T}^2(u)\pi(u)du,$$

U and π defined as above, and

$$VD_{i,R,N,T}(u)=\frac{2}{\sqrt{R}}\sum_{j=1}^{R}\left(1\left\{vd_{i,j,N,T}\leq u\right\}-\frac{1}{2}\right)$$

with

$$vd_{1,j,N,T} = \left(\exp \int_{\mathcal{A}} \left(\frac{\widehat{h}_{N,T}^{dif,(5-\varepsilon)} \widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}} \mathrm{d}a \right) \right)^{1/2} \eta_{1,j},$$

$$vd_{2,j,N,T} = \left(\exp \left(\int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a) \Delta_{N,T}^{1/2} \log^{1/2}(1/\Delta_{N,T})}{\widehat{h}_{N,T}^{dif,(1+\varepsilon)}} \mathrm{d}a \right) \right)^{1/2} \eta_{2,j},$$

with $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)^\intercal \sim \mathrm{iid} N(0, I_{2R})$.

Theorem 4. Let Assumption 1 and 2 hold. Assume $T, N, R \to \infty$ and $N/T^5 \to \infty$.
(i) Under H_0^{dif},

$$VD_{R,N,T} \xrightarrow{d^*} \chi_1^2 \ a.s. - P.$$

(ii) Under H_A^{dif}, there are $\delta, \eta > 0$ such that

$$P^* \left(R^{-1+\eta} VD_{R,N,T} > \delta \right) \to 1 \ a.s. - P.$$

■

Remark 2 (The local polynomial and local linear case). Our discussion has focused on classical Nadaraya-Watson kernel estimates. We will continue to do so throughout this paper. This said, the methods readily apply to alternative kernel estimators when appropriately modified, if needed. For example, they apply (unchanged) to the local linear estimates studied by Fan and Zhang (2003) and Moloche (2004).

3 Jump-diffusion processes

We now study the problem of bandwidth selection in the context of processes with discontinuous sample paths. Consider the class of jump-diffusion models

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t + \mathrm{d}J_t,$$

where $\{J_t : t = 1, ..., T\}$ is a Poisson jump process with infinitesimal intensity $\lambda(X_t)\mathrm{d}t$ and jump size c. Let $c = c(X_t, y)$, where y is a random variable with stationary distribution $f_y(.)$.

We begin by assuming existence of consistent estimates of $\mu(.)$ and $\sigma(.)$ in the presence of jumps ($\widehat{\mu}_{N,T}(.)$ and $\widehat{\sigma}_{N,T}^2(.)$). Later we show how these estimates can be defined. Write, as earlier,

$$\widehat{\varepsilon}_{i\Delta_{N,T}} = \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}}$$

for $i = 2, ..., \Delta_{N,T}^{-1}\overline{T}$. We note that

$$
\begin{aligned}
\widehat{\varepsilon}_{i\Delta_{N,T}} &= \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \\
&= \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \mu(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\left(\sigma(X_{(i-1)\Delta_{N,T}}) + o_p(1)\right)\sqrt{\Delta_{N,T}}} + o_p(1) \\
&\approx \frac{\sigma(X_{(i-1)\Delta_{N,T}})\left(W_{i\Delta_{N,T}} - W_{(i-1)\Delta_{N,T}}\right)}{\left(\sigma(X_{(i-1)\Delta_{N,T}}) + o_p(1)\right)\sqrt{\Delta_{N,T}}} + \frac{J_{i\Delta_{N,T}} - J_{(i-1)\Delta_{N,T}}}{\left(\sigma(X_{(i-1)\Delta_{N,T}}) + o_p(1)\right)\sqrt{\Delta_{N,T}}} + o_p(1) \\
&\approx N(0,1) + \frac{J_{i\Delta_{N,T}} - J_{(i-1)\Delta_{N,T}}}{\sigma(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} + o_p(1). \qquad (11)
\end{aligned}
$$

If there is a jump at $i\Delta_{N,T}$, $\left(J_{i\Delta_{N,T}} - J_{(i-1)\Delta_{N,T}}\right) = O_p(1)$. However, over a finite time span $\overline{T}$, there will only be a finite number of times in which $1\{\widehat{\varepsilon}_{i\Delta_{N,T}} \leq x\}$ is 1 instead of 0 or viceversa, because of jumps. Thus,

$$
\frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}} 1\{\widehat{\varepsilon}_{i\Delta_{N,T}} \leq x\} = \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}} 1\{\widehat{\varepsilon}^{c}_{i\Delta_{N,T}} \leq x\} + \frac{O_p(1)}{\overline{N}},
$$

where $\widehat{\varepsilon}^{c}_{i\Delta_{N,T}}$ is the residual that would prevail in the continuous case. Hence, the same criterion as in Subsection 2.2 can be applied to the case with jumps.

It still remains to establish conditions under which we have consistent estimates of the infinitesimal moments in the presence of jumps. Hereafter, we rely on the following assumption:

Assumption 3.

(i) *$\mu(.)$, $\sigma(.)$, $c(.,y)$, and $\lambda(.)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l,u)$ with $-\infty \leq l < u \leq \infty$, where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. All functions are at least twice continuously differentiable. They satisfy local Lipschitz and growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_4^J, C_5^J, and C_6^J so that, for all x and z in J,*

$$
|\mu(x) - \mu(z)| + |\sigma(x) - \sigma(z)| + \lambda(x)\int_Y |c(x,y) - c(z,y)|\,\Pi(\mathrm{d}y) \leq C_4^J|x - z|,
$$

and

$$
|\mu(x)| + |\sigma(x)| + \lambda(x)\int_Y |c(x,y))|\,\Pi(\mathrm{d}y) \leq C_5^J\{1 + |x|\},
$$

and for $\alpha > 2$,

$$
\lambda(x)\int_Y |c(x,y))|^{\alpha}\,\Pi(\mathrm{d}y) \leq C_6^J\{1 + |x|^{\alpha}\},
$$

(ii) $\lambda(.) > 0$ and $\sigma^2(.) > 0$ *on* $\mathfrak{D}$.

(iii) *$\mu(.)$, $\sigma(.)$, $c(.,y)$, and $\lambda(.)$ are such that the solution is recurrent.*

In what follows, we consider two alternative scenarios. First, we establish the validity of our bandwidth selection procedure for all infinitesimal moments under parametric assumptions on the jump component. Second, without making parametric assumptions on the jump component, we discuss bandwidth selection for the purpose of consistent (and asymptotically normal) estimation of the system's drift and infinitesimal variance. In the former case, we incur the risk of incorrectly specifying the jump distribution but completely identify the system's dynamics. The procedure is, in spirit, semiparametric. In the latter case, we are agnostic about the jump distribution, but can only identify the process' drift (possibly inclusive of the first conditional jump moment) and the process' infinitesimal volatility, while remaining fully nonparametric. If interest is on the full system's dynamics, one should employ the procedure in Subsection 3.1. If interest is solely on the volatility of the continuous component of the process, then the methods in Subsection 3.2 are arguably preferable. As we will show, in fact, the diffusion's kernel estimator converges at a faster rate in this second case.

3.1 Consistent estimation of all infinitesimal moments

In order to separate the moments of the continuous component from those of the jump component, we ought to properly correct the kernel estimators considered in the previous section. Following Bandi and Nguyen (2003) and Johannes (2004), define

$$\widehat{\mu}_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T,1}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T,1}}\right)} - \widehat{\lambda}_{\mathbf{h}_{n,T}}(X_t)\widehat{\mathrm{E}}_{y,\mathbf{h}_{n,T}}\left(c(X_t,y)\right) \tag{12}$$

and

$$\widehat{\sigma}^2_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T,2}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)^2}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T,2}}\right)} - \widehat{\lambda}_{\mathbf{h}_{n,T}}(X_t)\widehat{\mathrm{E}}_{y,\mathbf{h}_{n,T}}\left(c(X_t,y)^2\right). \tag{13}$$

Since the intensity estimator $\widehat{\lambda}(.)$, as well as the jump size moment estimator, $\widehat{\mathrm{E}}_y\left(c(.,y)^j\right)$ with $j=1,2$ depend, in general, on higher-order infinitesimal moment estimates, we make explicit their dependence on a (vector-)bandwidth $\mathbf{h}_{n,T}$ and write $\widehat{\lambda}_{\mathbf{h}_{n,T}}(.)$ and $\widehat{\mathrm{E}}_{y,\mathbf{h}_{n,T}}\left(c(.,y)^2\right)$, as above.

We are now more specific. Identification of $\lambda(.)$ and the moments of the jumps may hinge on parametric assumptions on $f_y(.)$, i.e., the probability distribution of the jump size. Assume, for instance, $c(X_t,y)=y$ and $f_y(.)=N(0,\sigma_y^2)$, but alternative specifications may, of course, be invoked along the lines of, e.g., Bandi and Renò (2008). Then, from Bandi and Nguyen (2003) and Johannes (2004), one can write

$$
\begin{aligned}
\widehat{\mathrm{E}}_{y,\mathbf{h}_{n,T}}\left(c(X_t,y)^2\right) &= \left(\widehat{\sigma}_y^2\right)_{N,T} = \frac{1}{\overline{N}}\sum_{j=1}^{\overline{N}} \frac{\widehat{M}^6_{N,T,h_6}(X_{j\Delta_{n,T}})}{5\widehat{M}^4_{N,T,h_4}(X_{j\Delta_{n,T}})}, \\
\widehat{\lambda}_{\mathbf{h}_{n,T}}(a) &= \frac{\widehat{M}^4_{N,T,h_4}(a)}{3\left(\widehat{\sigma}_y^4\right)_{N,T}},
\end{aligned}
$$

with

$$
\widehat{M}^j_{N,T,h_k}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T,k}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)^j}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T,k}}\right)} \qquad j = 1, \ldots
$$

Since the mean of the jump size is zero, Eq. (12) and Eq. (13) become, in this case:

$$
\begin{aligned}
\widehat{\mu}_{N,T}(a) &= \widehat{M}^1_{N,T,h_1}(a), && (14)\\
\widehat{\sigma}^2_{N,T}(a) &= \widehat{M}^2_{N,T,h_2}(a) - \frac{\widehat{M}^4_{N,T,h_4}(a)}{3\left(\frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}} \frac{\widehat{M}^6_{N,T,h_6}(X_{i\Delta_{n,T}})}{5\widehat{M}^4_{N,T,h_4}(X_{i\Delta_{n,T}})}\right)^2} \left(\frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}} \frac{\widehat{M}^6_{N,T,h_6}(X_{i\Delta_{n,T}})}{5\widehat{M}^4_{N,T,h_4}(X_{i\Delta_{n,T}})}\right), && (15)
\end{aligned}
$$

with $\mathbf{h}_{n,T} = (h_1, h_2, h_4, h_6)$. In other words, optimization of the criterion in Subsection 2.2 will now depend on four bandwidths whose properties are laid out below.

Proposition 2 (Bandi and Nguyen, 2003): Let Assumption 3 hold.

(i) Let $\overline{\Delta}_{N,\overline{T}} = \overline{T}/N$ with $\overline{T}$ fixed. If $\lim_{N\to\infty} \frac{1}{h_{N,\overline{T}}}\left(\overline{\Delta}_{N,\overline{T}} \log \frac{1}{\overline{\Delta}_{N,\overline{T}}}\right)^{1/2} \to 0$, then

$$
\widehat{\overline{L}}_X(\overline{T}, a) - \overline{L}_X(\overline{T}, a) = o_{a.s.}(1),
$$

where $\widehat{\overline{L}}_X(\overline{T}, a) = \frac{\overline{\Delta}_{N,\overline{T}}}{h_{N,\overline{T}}} \sum_{j=1}^{N} K\left(\frac{X_{j\overline{\Delta}_{N,\overline{T}}}-a}{h_{N,\overline{T}}}\right)$.

- *The infinitesimal moments*

If (ii) $h_{N,T,k}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ and (iii) $\frac{\overline{L}_X(T,a)}{h_{N,T,k}}\left(\Delta_{N,T}\log\frac{1}{\Delta_{N,T}}\right)^{1/2} \overset{a.s.}{\to} 0$, then:

$$
\widehat{M}^k_{N,T,h_k}(a) - M^k(a) = o_{a.s.}(1).
$$

If, in addition, (iv) $h^{dr,5}_{N,T,k}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$, then:

$$
\sqrt{h_{N,T,k}\widehat{\overline{L}}_X(T,a)}\left(\widehat{M}^k_{N,T,h_k}(a) - M^k(a)\right) \Rightarrow N\left(0, \mathbf{K}_2 M^{2k}(a)\right).
$$

From the proposition above, we note that all moments estimators converge to their limit at the same rate, $\sqrt{h_{N,T,k}\widehat{\overline{L}}_X(T,a)}$. Importantly, it is theoretically sound to employ the same rate condition for all moments. This is in sharp contrast with the continuous semimartingale case in which the drift estimator converges at a slower rate than the infinitesimal variance estimator. In the continuous case, in fact, one ought to use different bandwidth rates, since, from the conditions in Proposition 1, we require $h_{N,T}^{dr}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ (for the drift bandwidth) and $h_{N,T}^{dif}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$ (for the diffusion bandwidth).[6]

Let $\widehat{\mu}_{N,T,h}(a)$ and $\widehat{\sigma}^2_{N,T,h}(a)$ be defined as in Eq. (14) and Eq. (15) with $h_1 = h_2 = h_4 = h_6 = h$. We can now select h in such a way as to minimize $\sup_x |F_{N,T,h}(x) - \Phi(x)|$, where $F_{N,T,h}(x)$ is the empirical distribution of $\widehat{\varepsilon}$ (as defined in (11)) evaluated at x. Given the nature of the bandwidth requirements from Proposition 2, the second-step procedure can be carried out as in the continuous drift case. Similarly, the asymptotic behavior of the second-step procedure is as established in Theorem 3.[7]

Needless to say, misspecification of the parametric distribution of the jump component will, in general, result in failure of the statement in Theorem 2 since, in this case, there might not exist a bandwidth for which $\sup_x |F_{N,T,h}(x) - \Phi(x)| = o_p(1)$. We now turn to a procedure which does not impose parametric assumptions on the process' discontinuities at the cost of solely identifying the moments of the process' continuous component.

3.2 Consistent estimation of the drift and infinitesimal variance

Should we be unwilling to make parametric assumptions on the distribution of the jump component, we may still consistently estimate the infinitesimal variance term. The only maintained assumption about the jump component in this subsection is that J_t is a process of finite activity. Define $\widehat{\sigma}^2_{\mathbf{J},N,T}(a)$ as:

$$\widehat{\sigma}^2_{\mathbf{J},N,T}(a) = \frac{\mu_{2/p}^{-p}}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-p} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right) \Pi_{i=1}^{p} \left| X_{(j+i)\Delta_{N,T}} - X_{(j+i-1)\Delta_{N,T}} \right|^{\frac{2}{p}}}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}, \tag{16}$$

where $\mu_k = \mathrm{E}\left(|Z|^k\right)$, with Z denoting a standard normal random variable, and $2 < p < \overline{p} < \infty$.

Corradi and Distaso (2008) have studied the properties of this class of estimators for the case $\Delta_{N,\overline{T}} = \frac{\overline{T}}{N}$ with $\overline{T}$ finite. They have shown that, under mild conditions, $\widehat{\sigma}^2_{\mathbf{J},N,T}(a)$ identifies $\sigma^2(a)$ consistently in the presence of finite activity jumps. Since we are also dealing with finite activity jumps, with probability one we can have at most a finite number of jumps over a finite time span. As the time span increases indefinitely, the number of jumps increases roughly at the same rate. Provided $p > 2$, asymptotic mixed normality follows under the same rate conditions as in the continuous semimartingale case. Theorem 5 states the relevant result.

[6] Consider conditions (*iii*) and (*iv'*) in Proposition 1. From (*iv'*), we notice that $\Delta_{N,T}$ has to vanish at a slower rate than $h_{N,T}^{dif,5}\overline{L}_X(T,a)$. Set $\Delta_{N,T} = O\left(h_{N,T}^{dif,5-\delta}\overline{L}_X(T,a)\right)$ with $\delta > 0$ arbitrarily small. Now, plugging this condition into (*iii*) and ignoring the logarithm, we obtain

$$\frac{\overline{L}_X(T,a)}{h_{N,T}^{dif}}\sqrt{h_{N,T}^{dif,5-\varepsilon}\overline{L}_X(T,a)} = \overline{L}_X^{3/2}(T,a)h_{N,T}^{dif,3/2-\varepsilon/2} \overset{a.s.}{\to} 0,$$

which implies $h_{N,T}^{dif}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$ but, of course, this is in contraddiction with (*ii*) in the drift case (see Proposition 1).

[7] Simulations suggest that it is sometimes very beneficial to select a smaller bandwidth for the infinitesimal second moment than for the first and higher-order moments (see, e.g., Bandi and Renò, 2008). One may therefore set $h_1 = h_4 = h_6$ with h_2 left unrestricted. In this case our criterion results in the choice of two bandwidths like in the continuous case.

Theorem 5. Let Assumption 3 hold and let $p > 2$. If (i) $\frac{\overline{L}_X(T,a)}{h^{dif}_{N,T}}\left(\Delta_{N,T}\log\frac{1}{\Delta_{N,T}}\right)^{1/2} \overset{a.s.}{\to} 0$ and (ii) $\frac{h^{dif,5}_{N,T}\overline{L}_X(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0$, then

$$\sqrt{\frac{h^{dif}_{N,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}}}\left(\widehat{\sigma}^2_{\mathbf{J},N,T}(a)-\sigma^2(a)\right) \Rightarrow N\left(0,\gamma_p\mathbf{K}_2\sigma^4(a)\right),$$

where

$$\gamma_p = \frac{\mu^p_{4/p}-(2p-1)\mu^{2p}_{2/p}+2\left(\mu^{p-1}_{4/p}\mu^2_{2/p}+\mu^{p-2}_{4/p}\mu^4_{2/p}+...+\mu^{p-(p-1)}_{4/p}\mu^{2(p-1)}_{2/p}\right)}{\mu^{2p}_{2/p}}.$$

■

Let

$$\widehat{\varepsilon}_{i\Delta_{N,T}} = \frac{X_{i\Delta_{N,T}}-X_{(i-1)\Delta_{N,T}}-\widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\widehat{\sigma}_{\mathbf{J},N,T}(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}},$$

where $\widehat{\mu}_{N,T}$ is defined as in (14) and $\widehat{\sigma}^2_{\mathbf{J},N,T}(a)$ is as in (16) above with $p > 2$. We may now select $h_{N,T} = (h^{dr}_{N,T}, h^{dif}_{N,T})$ so as to minimize $\sup_x|F_{N,T,h}(x)-\Phi(x)|$, where $F_{N,T,h}(x)$ is the empirical distribution of $\widehat{\varepsilon}$. Subsequently, we can verify the rate conditions as in the continuous semimartingale drift and diffusion case. In other words, Theorems 3 and 4 apply.

Of course, if the jump size does not have mean zero, the procedure only identifies the sum of the drift component and the compensator (see, e.g., Eq. 12) while remaining consistent for the diffusive volatility. Should this be the case, then one has to resort to parametric assumptions, as in the previous subsection, to identify the continuous drift component, if needed.

4 Diffusions observed with error (or microstructure noise)

We now assume that the process X_t is contaminated by measurement error and write observations from the observable process Y_t as

$$Y_{i\Delta_{N,T}} = X_{i\Delta_{N,T}} + \epsilon_{i\Delta_{N,T}}, \tag{17}$$

where $a^{-1/2}_{N,T}\epsilon_{i\Delta_{N,T}}$ is an i.i.d. sequence with mean zero, variance 1, and such that $\mathrm{E}\left(\epsilon^k_{i\Delta_{N,T}}\right) = O\left(a^{k/2}_{N,T}\right)$ $(k \geq 2)$ for $a_{N,T} \to 0$ as $N, T \to \infty$.

We provide estimates of the first two infinitesimal moments which are robust to this type of measurement error. In this context, we establish conditions for consistency and asymptotic normality. We then turn to the issue of automatic bandwidth choice. Write

$$\widehat{\mu}_{N,l,T}(a) = \frac{\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-a}{h^{dr}_{N,l,T}}\right)\Delta^{-1}_{l,T}\left(Y_{(jB+b)\Delta_{N,T}}-Y_{((j-1)B+b)\Delta_{N,T}}\right)}{\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-a}{h^{dr}_{N,l,T}}\right)}, \tag{18}$$

where $Bl = N$ and $\Delta_{l,T} = T/l$. As for the diffusion:g

$$\widehat{\sigma}^2_{N,l,T}(a) = \frac{\sum_{b=1}^{B}\sum_{j=1}^{l-1} K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-a}{h^{dif}_{N,l,T}}\right)\Delta_{l,T}^{-1}\left(Y_{(jB+b)\Delta_{N,T}} - Y_{((j-1)B+b)\Delta_{N,T}}\right)^2}{\sum_{b=1}^{B}\sum_{j=1}^{l-1} K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-a}{h^{dif}_{N,l,T}}\right)} - \Delta_{l,T}^{-1} RV_{T,N}\Delta_{N,T}, \tag{19}$$

where $RV_{T,N}\Delta_{N,T} = \Delta_{N,T}\sum_{j=1}^{N}\left(Y_{j\Delta_{N,T}} - Y_{(j-1)\Delta_{N,T}}\right)^2$. In the case of a fixed time span, the estimator in Eq. (19) has been studied by Corradi and Distaso (2008). Here, we also consider estimation of the first infinitesimal moment as in Eq. (18). In both cases, letting the time span increase without bound (which is, as always, necessary in the drift case for consistency) raises additional technical issues which ought to be dealt with.

Remark 3 (market microstructure). When $Y_{i\Delta_{N,T}}$ is an observable logarithmic *price* process (i.e., a transaction price or a mid-quote, for example), $X_{i\Delta_{N,T}}$ generally denotes the underlying, unobservable equilibrium price and $\varepsilon_{i\Delta_{N,T}}$ defines market microstructure noise. If econometric interest is placed on the drift and diffusion function of the equilibrium price process, as is generally the case, then $\widehat{\mu}_{N,l,T}(a)$ and $\widehat{\sigma}^2_{N,l,T}(a)$ will provide consistent and asymptotically normal estimates of its true infinitesimal moments (as we show below) even when contaminated price observations $Y_{i\Delta_{N,T}}$ are employed.

Remark 4. We note that the form of $\widehat{\mu}_{N,l,T}(a)$ and $\widehat{\sigma}^2_{N,l,T}(a)$ requires the use of an appropriately-chosen lower frequency l. In agreement with the two-scale estimator of Zhang, Mykland, and Aït-Sahalia (2005), ZMA henceforth, the diffusion case also requires a bias-correction term based on the higher frequency N (see, also, Aït-Sahalia, Mykland, and Zhang, 2009).

Theorem 6

- *The infinitesimal first moment*

Let Assumption 1 hold and let ϵ be defined as in Eq. (17). Also assume that $l = O(BT)$. If (i) $h^{dr}_{N,l,T}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$, (ii) $\frac{\overline{L}_X(T,a)}{h^{dr}_{N,l,T}}\left(\Delta_{l,T}\log\frac{1}{\Delta_{l,T}}\right)^{1/2} \overset{a.s.}{\to} 0$, (iii) $h^{dr,5}_{N,l,T}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$, and (iv) $\frac{N^{1/k}a^{1/2}_{N,T}\sqrt{\overline{L}_X(T,a)}}{\sqrt{h^{dr}_{N,l,T}}} \overset{a.s.}{\to} 0$, then

$$\sqrt{h^{dr}_{N,l,T}\widehat{\overline{L}}_X(T,a)}\left(\widehat{\mu}_{N,l,T}(a) - \mu(a)\right) \Rightarrow N\left(0, \frac{2}{3}\mathbf{K}_2\sigma^2(a)\right).$$

- *The infinitesimal second moment*

Let Assumption 1 hold and let ϵ be defined as in Eq. (17). Also assume that $l = O(BT)$. If (i) $\frac{\overline{L}_X(T,a)}{h^{dif}_{N,l,T}}\left(\Delta_{l,T}\log\frac{1}{\Delta_{l,T}}\right)^{1/2} \overset{a.s.}{\to} 0$, (ii) $\frac{h^{dif,5}_{N,l,T}\overline{L}_X(T,a)}{\Delta_{l,T}} \overset{a.s.}{\to} 0$, (iii) $\frac{a^2_{N,T}l}{\Delta^2_{l,T}} \to 0$, and (iv) $\frac{N^{1/k}a^{1/2}_{N,T}l^{1/2}\sqrt{\frac{h^{dif}_{N,l,T}\overline{L}_X(T,a)}{T}}}{h^{dif}_{N,l,T}} \overset{a.s.}{\to} 0$, then

$$\sqrt{\frac{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{l,T}}}\left(\widehat{\sigma}^2_{N,l,T}(a) - \sigma^2(a)\right) \Rightarrow N\left(0, \mathbf{K}_2\sigma^4(a)\right).$$

■

Both in the drift and in the diffusion case, the averaging over sub-grids reduces the constant of proportionality in the estimators' asymptotic variance (from 1 to $\frac{2}{3}$ in the drift case, from 2 to 1 in the diffusion case). The rates of convergence are also affected. In the diffusion case, since $\frac{\Delta_{N,T}}{\Delta_{l,T}} \to 0$, the rate is slower. In the drift case, the new bandwidth condition (*ii*) requires larger bandwidth choices and thus, compatibly with condition (*iii*), the actual rate may be faster. Since $N = Bl$, by choosing a smaller l, and hence a larger B, we may allow for a larger variance of the error term. This choice will in general not come at a price (in terms of convergence rate) as far as the drift is concerned, but could come at a price in the case of diffusion estimation if $\frac{h^{dif}_{N,l_1,T}}{\Delta_{l_1,T}} = o\left(\frac{h^{dif}_{N,l_2,T}}{\Delta_{l_2,T}}\right)$ with $l_1 < l_2$.

Turning to bandwidth selection, we note that our local Gaussian criterion ought to be re-adjusted in this new framework. We first propose a heuristic argument to provide intuition. Let

$$
\begin{aligned}
\widehat{u}_{i\Delta_{N,T}} &= \frac{Y_{i\Delta_{N,T}} - Y_{(i-1)\Delta_{N,T}} - \widehat{\mu}_{N,l,T}(Y_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\widehat{\sigma}_{N,l,T}(Y_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \\
&= \frac{Y_{i\Delta_{N,T}} - Y_{(i-1)\Delta_{N,T}} - \mu(Y_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\left(\sigma(Y_{(i-1)\Delta_{N,T}}) + o_p(1)\right)\sqrt{\Delta_{N,T}}} + o_p(1) \\
&= \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \mu(X_{(i-1)\Delta_{N,T}})\Delta_{N,T} + \epsilon_{i\Delta_{N,T}} - \epsilon_{(i-1)\Delta_{N,T}} - \mu'(\overline{X}_{(i-1)\Delta_{N,T}})\epsilon_{(i-1)\Delta_{N,T}}\Delta_{N,T}}{\left(\sigma(X_{(i-1)\Delta_{N,T}}) + \sigma'(\overline{X}_{(i-1)\Delta_{N,T}})\epsilon_{(i-1)\Delta_{N,T}} + o_p(1)\right)\sqrt{\Delta_{N,T}}} \\
&\quad + o_p(1) \\
&= u_{i\Delta_{N,T}} + o_p(1),
\end{aligned}
$$

where $\overline{X}_{(i-1)\Delta_{N,T}} \in \left(X_{(i-1)\Delta_{N,T}}, Y_{(i-1)\Delta_{N,T}}\right)$. In spite of the consistency of the drift and infinitesimal variance estimator, $u_{i\Delta_{N,T}}$ is in general non-Gaussian since the presence of measurement error affects $Y_{i\Delta_{N,T}} - Y_{(i-1)\Delta_{N,T}}$ and, of course, the evaluation point.

A natural solution to this issue is to use different frequencies for infinitesimal moment estimation and for bandwidth selection. For the latter, one may use a (lower) frequency at which the contamination error is expected to have little or no effect, say $\Delta_{H,T}$, with $H/N \to 0$. Provided $a_{N,T} = o(\Delta_{H,T})$, we define

$$
\begin{aligned}
\widehat{u}_{i\Delta_{H,T}} &= \frac{Y_{i\Delta_{H,T}} - Y_{(i-1)\Delta_{H,T}} - \widehat{\mu}_{N,l,T}(Y_{(i-1)\Delta_{H,T}})\Delta_{H,T}}{\widehat{\sigma}_{N,l,T}(Y_{(i-1)\Delta_{H,T}})\sqrt{\Delta_{H,T}}} \\
&= \frac{X_{i\Delta_{H,T}} - X_{(i-1)\Delta_{H,T}} - \mu(X_{(i-1)\Delta_{H,T}})\Delta_{H,T}}{\left(\sigma(X_{(i-1)\Delta_{H,T}}) + o_p(1)\right)\sqrt{\Delta_{H,T}}} + o_p(1) \\
&= u_{i\Delta_{H,T}} + o_p(1).
\end{aligned}
$$

It is now clear that $u_{i\Delta_{H,T}}$ is approximately Gaussian. The criterion defined in Section 2.2 is therefore still valid and the statements in Theorem 1 and 2 continue to apply. In finite samples, of course, the approximation is best the smallest the interval $\Delta_{H,T}$. From a practical standpoint, therefore, one has to balance the size of the implied measurement error with the accuracy of the Gaussian approximation. The highest frequency at which the measurement error appears negligible is therefore the preferable frequency.

Remark 5. In the case of high-frequency logarithmic asset prices and market microstructure noise, an appropriate frequency may be chosen in a data-driven manner, either by looking at signature plots (Andersen, Bollerslev, Diebold, and Labys 2000) or via the statistics suggested by Awartani, Corradi, and Distaso (2009).

In the second stage one needs to verify whether the bandwidths selected by the procedure in Theorem 1, say $\widehat{h}^{dr}_{N,l,T}$ and $\widehat{h}^{dif}_{N,l,T}$, satisfy all of the rate conditions in Theorem 6. We begin with the drift. Notice that N, T and the size of the measurement error $a_{N,T}$ are given. While $a_{N,T}$ is unknown in general, it may be estimated by using $\left(RV_{T,N}\Delta_{N,T}\right)/2$ as defined in Eq. (19). Given N, we fix l and B, using the fact that $N = lB$. If we choose $l = O\left(a_{N,T}^{-1}T\right)$, it is immediate to see that (ii) implies (iv). Summarizing, if $T^{17/5}/l \to 0$ and $l = O\left(a_{N,T}^{-1}T\right)$, there is a bandwidth satisfying (i)-(iv) and we can proceed along the lines of Theorem 2 by testing the hypothesis:

$$
\begin{aligned}
H_0^{dr} \quad : \quad & \int_{\mathcal{A}} \widehat{h}^{dr,5-\varepsilon}_{N,l,T}\widehat{\overline{L}}_X(T,a)\mathrm{d}a \overset{a.s.}{\to} \infty \\
& \text{or } \max\left\{ \frac{1}{\int_{\mathcal{A}} \widehat{h}^{dr,(1+\varepsilon)}_{N,l,T}\widehat{\overline{L}}_X(T,a)\mathrm{d}a}, \int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a)\Delta_{l,T}^{1/2}\log^{1/2}(1/\Delta_{l,T})}{\widehat{h}^{dr,(1+\varepsilon)}_{N,l,T}}\mathrm{d}a \right\} \overset{a.s.}{\to} \infty
\end{aligned}
$$

for $\mathcal{A} \subset \mathfrak{D}$, and $\varepsilon > 0$ arbitrarily small, versus its alternative.

We now turn to the variance estimator. If we set $l = O\left(a_{N,T}^{-2/3+\varepsilon}T^{2/3}\right)$, (iii) is always satisfied. Further, if $T^5/l \to 0$, there is a bandwidth satisfying (i) and (ii). We now test the following hypothesis:

$$
\begin{aligned}
H_0^{dif} \quad : \quad & \int_{\mathcal{A}} \frac{\widehat{h}^{dif,5-\varepsilon}_{N,l,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{l,T}}\mathrm{d}a \overset{a.s.}{\to} \infty \text{ or} \\
& \max\left\{ \int_{\mathcal{A}} \frac{N^{1/k}a_{N,T}^{1/2}l^{1/2}\sqrt{\frac{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,a)}{T}}}{h^{dif,(1+\varepsilon)}_{N,l,T}}\mathrm{d}a, \int_{\mathcal{A}} \frac{\widehat{\overline{L}}_X(T,a)\Delta_{l,T}^{1/2}\log^{1/2}(1/\Delta_{l,T})}{\widehat{h}^{dif,(1+\varepsilon)}_{N,l,T}}\mathrm{d}a, \right\} \overset{a.s.}{\to} \infty
\end{aligned}
$$

for $\mathcal{A} \subset \mathfrak{D}$, and $\varepsilon > 0$ arbitrarily small, versus its alternative.

5 Stochastic volatility

Consider now the model

$$
\begin{aligned}
dX_t &= \mu_t^X dt + v_t dW_t^X \\
df(v_t^2) &= \mu(v_t^2)dt + \sigma(v_t^2)dW_t^\sigma,
\end{aligned}
$$

where $\left\{W_t^X : t = 1, ..., T\right\}$ and $\left\{W_t^\sigma : t = 1, ..., T\right\}$ are potentially correlated Brownian motions. The function $f(x)$ may be equal to $\log(x)$ as in Jaquier, Polson, and Rossi (1994) or x as in Eraker, Johannes, and Polson (2003), for instance. Our interest is in $\mu(v_t^2)$ and $\sigma^2(v_t^2)$, the drift and the diffusion function of the spot variance process.

Volatility is latent. However, it may be filtered from prices X_t sampled at high frequency as suggested by Kristensen (2008) and Bandi and Renò (2008). To this extent, assume, as earlier, availability of N equidistant price observations with $\Delta_{N,T} = T/N$ denoting the time distance between successive data points and T denoting the time span. We again observe the price skeleton $X_{\Delta_{N,T}}, X_{2\Delta_{N,T}}, ..., X_{T\Delta_{N,T}}$. These price observations may be employed to (1) filter spot volatility (or spot variance) nonparametrically for the purpose of (2) identifying $\mu(.)$ and $\sigma^2(.)$. Using preliminary spot variance estimates $\widehat{v}_t^2$, the latter may be done by virtue of the functional estimators in Eq. (1) and (2) (Bandi and Renò, 2008, and Kanaya and Kristensen, 2008). Importantly, however, selection of the smoothing sequences h^{dr} and h^{dif} now also depends on the need to eliminate the impact of the estimation error induced by the first-step spot variance estimates.

To present the main ideas, consider spot variance estimates obtained by virtue of the classical realized variance estimator (Andersen, Bollerslev, Diebold, and Labys, 2003, and Barndorff-Nielsen and Shephard, 2002). Specifically, write

$$\widehat{v}_\tau^2 = \sum_{i=\tau - T^{-\delta_N}\Delta_{N,T}^{-1}}^{\tau + T^{-\delta_N}\Delta_{N,T}^{-1}} T^{\delta_N} \left(X_{(i+1)\Delta_{N,T}} - X_{i\Delta_{N,T}} \right)^2 . \tag{20}$$

The estimator averages $2T^{-\delta_N}\Delta_{N,T}^{-1}$ squared price differences in a local neighborhood of τ determined by the localizing factor $T^{-\delta_N}$.

Bandi and Renò (2008) introduce four additional conditions (with respect to those in Proposition 1 above) which the drift bandwidth h^{dr} and the diffusion bandwidth h^{dif} ought to satisfy for asymptotic normality of the drift and the diffusion function estimates to hold. These conditions (two for each infinitesimal moment) are *sufficient* to eliminate, asymptotically, the influence of the estimation error induced by $\widehat{v}_\tau^2$ (when used in place of the unobservable v_τ^2). Intuitively, the conditions imply that one needs to use a larger discrete interval, say $\Delta_{M,T} = \frac{T}{M}$ with $M/N \to 0$, than is used for estimating the preliminary spot variance estimates. In other words, one needs to use high-frequency data to identify spot variance $\widehat{v}^2$ and M lower frequency observations (on $\widehat{v}^2$) to identify the dynamics (through $\mu(.)$ and $\sigma^2(.)$). To this extent, call the relevant bandwidths $h_{M,T}^{dr}$ and $h_{M,T}^{dif}$.

In what follows, for conciseness, we will not discuss the origin and form of these four conditions. We refer the reader to Bandi and Renò (2008) and Appendix B to this paper for details. However, consistently with our stated goal, we discuss the implications of the four conditions for bandwidth choice. When dealing with this choice, the main technical issue is now that the rate of growth of M depends on $h_{M,T}^{dr}$, which is what one needs to find optimally, as well as on $\overline{L}_v(T,a)$ which is unknown and whose estimates depend on $h_{M,T}^{dr}$. This is an important difference from the observable case in which all N observations are used. In the drift case, one may consider optimizing over both M^{dr} and $h_{M,T}^{dr}$. Similarly, in the diffusion case one might wish to optimize over M^{dif} and $h_{M,T}^{dif}$. We leave this issue for future work and take the following approach to the problem.

As said, it is natural for applied researchers to employ N high-frequency observations to identify spot variance before using M lower frequency data (on $\widehat{v}^2$) to evaluate the dynamics. To this extent, assume M and N are fixed (with $M < N$). It can be shown (see Appendix B) that, for the drift, the implied bandwidth condition becomes:

$$h_{M,T}^{dr}\overline{L}_v(T,a)\underbrace{\left(\frac{N^{2\left(\frac{\beta}{2\beta+1}\right)}T^{-2\left(\frac{\beta}{2\beta+1}\right)}}{M^2}\right)}_{\alpha_{N,T,M}}\overset{a.s.}{\rightarrow}\infty. \tag{21}$$

where $\beta \leq \frac{1}{2}$.[8] As for the diffusion:

$$\frac{h_{M,T}^{dif}\overline{L}_v(T,a)}{\Delta_{T,M}}\underbrace{\left(\frac{NN^{-\left(\frac{2\beta}{1+2\beta}\right)\frac{1}{2\beta}}T^{\left(\frac{2\beta}{1+2\beta}\right)\frac{1}{2\beta}}T}{M^4}\right)}_{\gamma_{N,T,M}}\overset{a.s.}{\rightarrow}\infty. \tag{22}$$

In general (i.e., for empirically reasonable values of N, M, T), it is easy to see that $\alpha_{N,T,M} < 1$ and $\gamma_{N,T,M} < 1$. Hence, Eq. (21) and Eq. (22) are more stringent conditions that $h_{M,T}^{dr}\overline{L}_v(T,a) \overset{a.s.}{\rightarrow} \infty$ and $\frac{h_{M,T}^{dif}\overline{L}_v(T,a)}{\Delta_{T,M}} \overset{a.s.}{\rightarrow} \infty$ with probability one. This observation leads to the following tests.

If $\frac{M}{T^{17/5}} \rightarrow \infty$ and N, M, T are such that $\alpha_{N,T,M} \rightarrow 0$ and $T^{[1-(\frac{1}{5}+c)]\theta}\alpha_{N,T,M} \rightarrow \infty$ (with $c > 0$, where T^{θ} is the local time's divergence rate), then

$$H_0^{dr}: \widehat{h}_{M,T}^{dr,5}\widehat{\overline{L}}_v(T,a) \overset{a.s.}{\rightarrow} \infty \textit{ or } \max\left\{\frac{1}{\alpha_{N,T,M}\widehat{h}_{M,T}^{dr}\widehat{\overline{L}}_v(T,a)}, \frac{\widehat{\overline{L}}_v(T,a)\Delta_{M,T}^{1/2}\log^{1/2}(1/\Delta_{M,T})}{\widehat{h}_{M,T}^{dr}}\right\} \overset{a.s.}{\rightarrow} \infty.$$

If $\frac{M}{T^5} \rightarrow \infty$ and N, M, T are such that $\gamma_{N,T,M} \rightarrow 0$ and $M^{[1-(\frac{1}{5}+c)]}T^{[1-(\frac{1}{5}+c)](\theta-1)}\gamma_{N,T,M} \rightarrow \infty$ (with $c > 0$, where T^{θ} is the local time's divergence rate), then

$$H_0^{dif}: \frac{\widehat{h}_{M,T}^{dif,5}\widehat{\overline{L}}_v(T,a)}{\Delta_{M,T}} \overset{a.s.}{\rightarrow} \infty \textit{ or } \max\left\{\frac{\Delta_{M,T}}{\gamma_{N,T,M}\widehat{h}_{M,T}^{dif}\widehat{\overline{L}}_v(T,a)}, \frac{\widehat{\overline{L}}_v(T,a)\Delta_{M,T}^{1/2}\log^{1/2}(1/\Delta_{M,T})}{\widehat{h}_{M,T}^{dif}}\right\} \overset{a.s.}{\rightarrow} \infty.$$

6 Multivariate diffusion processes

We now turn to multidimensional diffusions. Let $X_t = (X_{1,t}, ..., X_{d,t})^{\intercal}$ and consider the stochastic differential equation

$$\mathrm{d}X_t = \boldsymbol{\mu}(X_t)\mathrm{d}t + \boldsymbol{\sigma}(X_t)\mathrm{d}\mathbf{W}_t,$$

[8]These conditions allow for the use of market microstructure noise-robust spot variance estimators. Bandi and Renò (2008) propose noise-robust *spot* variance estimators with a rate of convergence equal to $k^{\beta} = T^{-\beta\delta_N}\Delta_{N,T}^{-\beta}$ for some $\beta \leq \frac{1}{2}$. As in the case of realized variance (above), these estimators may be derived from robust *integrated* variance estimators (such as the two-scale estimator of Zhang, Mykland, and Aït-Sahalia, 2005, and the class of kernel estimators suggested by Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008b) by localizing the integrated estimates in time. Their asymptotic properties (studied in Bandi and Renò, 2008) reveal that β is, for instance, equal to 1/10 (in the case of the two-scale estimator) or 1/6 in the case of flat-top kernel estimates obtained by virtue of kernels $g(.)$ satisfying $g'(1) = 0$ and $g'(0) = 0$. For realized variance in Eq. (20) $\beta = \frac{1}{2}$.

where $\boldsymbol{\mu}(.)$ and $\boldsymbol{\sigma}(.)$ are matrix functions satisfying the regularity conditions for the existence of a recurrent solution in Bandi and Moloche (2004) and $\{\mathbf{W}_t : t = 1, ..., T\}$ is a (conformable) standard Brownian vector. Let the diffusion matrix $\boldsymbol{\Sigma}(a)$ be defined as $\boldsymbol{\Sigma}(a) = \boldsymbol{\sigma}(a)\boldsymbol{\sigma}(a)^{\intercal}$ for $x = (a_1, ..., a_d)$.

Suppose we observe $X_{\Delta_{N,T}}, X_{2\Delta_{N,T}}, ..., X_{N\Delta_{N,T}}$ with $\Delta_{N,T} = \frac{T}{N}$. Specifically, assume there is a frequency at which synchronized observations may be observed for all processes. This is standard for estimation methods relying on low-frequency observations. In principle, however, we could allow for observations recorded at random, asynchronous times and, therefore, use high-frequency data for estimation. This could be done, for example, by employing the *refresh time* approach advocated by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a). The use of refresh times, however, would require important, additional technicalities due to their randomness, and is beyond the scope of the present paper. In particular, it would require an extension of existing asymptotic (mixed) normal results for drift and infinitesimal variance estimators (in Proposition 3 below) to the case of random times.

We define Nadaraya-Watson estimators of the drift vector and covariance matrix by writing

$$\widehat{\boldsymbol{\mu}}_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} \mathbf{K}\left(\frac{X_{j\Delta_{N,T}}-a}{\mathbf{h}_{N,T}^{dr}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N,T}}\right)}{\sum_{j=1}^{N} \mathbf{K}\left(\frac{X_{j\Delta_{N,T}}-a}{\mathbf{h}_{N,T}^{dr}}\right)}$$

and

$$\widehat{\boldsymbol{\Sigma}}_{N,T}(a) = \frac{1}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-1} \mathbf{K}\left(\frac{X_{j\Delta_{N,T}}-a}{\mathbf{h}_{N,T}^{dif}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N.T}}\right)\left(X_{(j+1)\Delta_{N,T}} - X_{j\Delta_{N.T}}\right)^{\intercal}}{\sum_{j=1}^{N} \mathbf{K}\left(\frac{X_{j\Delta_{N,T}}-a}{\mathbf{h}_{N,T}^{dif}}\right)},$$

where the kernel $\mathbf{K}\left(\frac{X_{j\Delta_{N,T}}-x}{\mathbf{h}_{N,T}}\right) = \Pi_{i=1}^{d} K\left(\frac{X_{i,j\Delta_{N,T}}-x_i}{h_{i,N,T}}\right)$ is a product kernel and $K(.)$ is defined in the same manner as in Assumption 2. We denote by $\mathbf{h}_{N,T}$ the matrix bandwidth $\left(h_{1,N,T}^{dr}, ..., h_{d,N,T}^{dr}, h_{1,N,T}^{dif}, ..., h_{d,N,T}^{dif}\right)$ belonging to the set $H \subset R_{+}^{2d}$.

In the multivariate case, local time is not defined. However, the averaged kernel

$$\widehat{\overline{L}}_X(T, x) = \frac{\Delta_{N,T}}{\Pi_{i=1}^{d} h_{i,N,T}} \sum_{j=1}^{N-1} \mathbf{K}\left(\frac{X_{j\Delta_{N,T}} - a}{\mathbf{h}_{N,T}}\right)$$

will still provide an estimate of the occupation density of the process (while, at the same time, inheriting its divergence rate) as discussed in Bandi and Moloche (2004). Naturally, the divergence rate of the occupation density plays a role in the characterization of the bandwidth conditions for both the drift and the diffusion matrix.

Proposition 3 (Bandi and Moloche, 2004): Let Assumption 1 and 2 in Bandi and Moloche (2004) hold.

Assume $T, N \to \infty$ and $\Delta_{N,T} \to 0$. Assume, for all i, $h_{i,N,T} \to 0$ and

$$\left(\Delta_{n,T} \log(1/\Delta_{n,T})\right)^{1/2} / \Pi_{i=1}^{d} h_{i,N,T} \to 0.$$

Then,

$$\frac{\widehat{\overline{L}}_X(T, a)}{v(1/T)} \Rightarrow C_X \widetilde{\phi}(a)\, g_{\alpha},$$

where the function $v(1/T)$ is regularly-varying at infinity with process-specific parameter α satisfying $0 \leq \alpha \leq 1$, g_α is used here to denote the Mittag-Leffler random variable with the same process-specific parameter α, and C_X is a process-specific constant.

- *The drift estimator*

If, for all i, $h^{dr}_{i,N,T} \to 0$, $\Pi^d_{i=1} h^{dr}_{i,N,T} v(1/T) \to \infty$, and

$$\frac{v(1/T)}{\Pi^d_{i=1} h^{dr}_{i,N,T}} \left(\Delta_{N,T} \log \frac{1}{\Delta_{N,T}} \right)^{1/2} \to 0,$$

then

$$\widehat{\boldsymbol{\mu}}_{N,T}(a) - \boldsymbol{\mu}(a) \overset{a.s.}{\to} 0.$$

If, in addition, for all j, $h^{dr,5}_{j,N,T} \Pi^d_{i \neq j} h^{dr}_{i,N,T} v(1/T) \to 0$,

$$\sqrt{\Pi^d_{i=1} h^{dr}_{i,N,T} \widehat{\overline{L}}^{dr}_X(T,a)} \left(\widehat{\boldsymbol{\mu}}_{N,T}(a) - \boldsymbol{\mu}(a) \right) \Rightarrow \boldsymbol{\Sigma}^{1/2}(a) N \left(\mathbf{0}, \mathbf{K}^d_2 \mathbf{I}_d \right),$$

where $\mathbf{I}_d$ is a $d \times d$ identity matrix.

- *The diffusion estimator*

If, for all i, $h^{dif}_{i,N,T} \to 0$, $\frac{\Pi^d_{i=1} h^{dif}_{i,N,T} v(1/T)}{\Delta_{N,T}} \to \infty$, and

$$\frac{v(1/T)}{\Pi^d_{i=1} h^{dif}_{i,N,T}} \left(\Delta_{N,T} \log \frac{1}{\Delta_{N,T}} \right)^{1/2} \to 0,$$

then

$$\widehat{\boldsymbol{\Sigma}}_{N,T}(a) - \boldsymbol{\Sigma}(a) \overset{a.s.}{\to} 0.$$

If, in addition, for all j, $\frac{\left(h^{5,dif}_{j,N,T} \Pi^d_{i \neq j} h^{dif}_{i,N,T} \right) v(1/T)}{\Delta_{N,T}} \to 0$,

$$\sqrt{\frac{\Pi^d_{i=1} h^{dif}_{i,N,T} \widehat{\overline{L}}^{dif}_X(T,a)}{\Delta_{N,T}}} \text{vech} \left(\widehat{\boldsymbol{\Sigma}}_{N,T}(a) - \boldsymbol{\Sigma}(a) \right) \Rightarrow \boldsymbol{V}(a)^{1/2} N \left(\mathbf{0}, \mathbf{K^d_2 I}_d \right),$$

with $\boldsymbol{V}(a) = P_D \left(2\boldsymbol{\Sigma}(a) \otimes \boldsymbol{\Sigma}(a) \right) P_D^\intercal$, where P_D is so that $\text{vech}\boldsymbol{\Sigma}(a) = P_D \text{vec}\boldsymbol{\Sigma}(a)$.

We now turn to the first step of our bandwidth selection procedure. For $i = 2, ..., \Delta^{-1}_{N,T}\overline{T}$ define the inner product of the residual process:

$$\begin{aligned} \widehat{\varepsilon}^\intercal_{i\Delta_{N,T}} \widehat{\varepsilon}_{i\Delta_{N,T}} \quad = \quad & \left\{ \Delta^{-1}_{N,T} \left(\Delta X_{(j+1)\Delta_{N,T}} - \widehat{\boldsymbol{\mu}}_{N,T}(X_{(j+1)\Delta_{N,T}}) \Delta_{N,T} \right)^\intercal \right. \\ & \left. \widehat{\boldsymbol{\Sigma}}_{N,T}(X_{(j+1)\Delta_{N,T}})^{-1} \left(\Delta X_{(j+1)\Delta_{N,T}} - \widehat{\boldsymbol{\mu}}_{N,T}(X_{(j+1)\Delta_{N,T}}) \Delta_{N,T} \right) \right\}, \end{aligned}$$

where $\Delta X_{(j+1)\Delta_{N,T}} = X_{(j+1)\Delta_{N,T}} - X_{(j+1)\Delta_{N,T}}$. Now write:

$$\widehat{\mathbf{h}}_{N,T} = \arg\min_{\mathbf{h}} \frac{1}{\overline{N} - 1} \sup_{u \in \mathcal{D}^+} \sum_{i=2}^{\overline{N}} \left(1 \left\{ \widehat{\varepsilon}^\intercal_{i\Delta_{N,T}} \widehat{\varepsilon}_{i\Delta_{N,T}} \leq u \right\} - \Psi(u) \right)$$

and

$$\mathbf{h}_{N,T}^{*}=\mathbf{h}\in H\subset R_{+}^{2d}:\sup_{x}\left|F_{N}^{\mathbf{h}}(x)-\Psi(x)\right|\underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\to}}0,$$

where $\Psi(u)=\Pr\left(\chi_{d}^{2}\leq u\right)$, i.e., the cumulative distribution function of a Chi-squared random variable with d degrees of freedoms. Note that

$$\begin{aligned}
&\widehat{\varepsilon}_{i\Delta_{N,T}}\\
=\ &\varepsilon_{i\Delta_{N,T}}+\left(\widehat{\mathbf{\Sigma}}_{N,T}(X_{(j+1)\Delta_{N,T}})^{-1/2}-\mathbf{\Sigma}(X_{(j+1)\Delta_{N,T}})^{-1/2}\right)\Delta_{N,T}^{-1/2}\Delta X_{(j+1)\Delta_{N,T}}\\
&+\mathbf{\Sigma}(X_{(j+1)\Delta_{N,T}})^{-1/2}\left(\widehat{\boldsymbol{\mu}}_{N,T}(X_{(j+1)\Delta_{N,T}})-\boldsymbol{\mu}(X_{(j+1)\Delta_{N,T}}\right)\sqrt{\Delta_{N,T}}\\
&+\left(\widehat{\mathbf{\Sigma}}_{N,T}(X_{(j+1)\Delta_{N,T}})^{-1/2}-\mathbf{\Sigma}(X_{(j+1)\Delta_{N,T}})^{-1/2}\right)\left(\widehat{\boldsymbol{\mu}}_{N,T}(X_{(j+1)\Delta_{N,T}})-\boldsymbol{\mu}(X_{(j+1)\Delta_{N,T}}\right)\sqrt{\Delta_{N,T}},
\end{aligned}$$

where $\varepsilon_{i\Delta_{N,T}}=\mathbf{\Sigma}(X_{(j+1)\Delta_{N,T}})^{-1/2}\left(\Delta_{N,T}^{-1/2}\Delta X_{(j+1)\Delta_{N,T}}-\boldsymbol{\mu}(X_{(j+1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}\right)$, and so $\varepsilon_{i\Delta_{N,T}}^{\intercal}\varepsilon_{i\Delta_{N,T}}$ is i.i.d. χ_{d}^{2}. Hence, as $N,T\to\infty$ and $\Delta_{N,T}\to 0$, by the same arguments as in Theorem 1 and 2, $\widehat{\mathbf{h}}_{N,T}-\mathbf{h}_{N,T}^{*}\overset{p}{\to}0$ if, and only if,

$$\sup_{a\in\mathfrak{D}^{d}}\left|\widehat{\boldsymbol{\mu}}_{N,T}\left(a,\widehat{\mathbf{h}}_{N,T}^{dr}\right)-\boldsymbol{\mu}(a)\right|=o_{p}\left(\frac{1}{\sqrt{\Delta_{N,T}}}\right),\tag{23}$$

and

$$\sup_{a\in\mathfrak{D}^{d}}\operatorname{vech}\left|\widehat{\mathbf{\Sigma}}_{N,T}\left(a,\widehat{\mathbf{h}}_{N,T}^{dif}\right)-\mathbf{\Sigma}(a)\right|=o_{p}\left(1\right).\tag{24}$$

In the second step, we need to check whether $\widehat{\mathbf{h}}_{N,T}^{dr}$ is small enough as to satisfy (i) $\max_{j}h_{j,N,T}^{5,dr}\Pi_{i\neq j}^{d}h_{i,N,T}^{dr}\widehat{\overline{L}}_{X}^{dr}(T,a)\overset{a.s.}{\to}0\ \forall a\in\mathfrak{D}^{d}$ and large enough as to satisfy $\max\left\{(ii)\ \frac{1}{\Pi_{i=1}^{d}h_{i,N,T}^{dr}\widehat{\overline{L}}_{X}^{dr}(T,a)},\ (iii)\ \frac{(\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\widehat{\overline{L}}_{X}^{dr}(T,a)}{\Pi_{i=1}^{d}h_{i,N,T}^{dr}}\right\}\overset{a.s.}{\to}\infty\ \forall a\in\mathfrak{D}^{d}$. Similarly, we need to check whether $\mathbf{h}_{N,T}^{dif}$ is small enough as to satisfy $\frac{\max_{j}h_{j,N,T}^{5,dif}\Pi_{i\neq j}^{d}h_{i,N,T}^{dif}\widehat{\overline{L}}_{X}^{dr}(T,a)}{\Delta_{N,T}}\overset{a.s.}{\to}0\ \forall a\in\mathfrak{D}^{d}$ and large enough as to satisfy $\frac{\Pi_{i=1}^{d}h_{i,N,T}^{dif}}{(\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\widehat{\overline{L}}_{X}^{dif}(T,a)}\overset{a.s.}{\to}\infty$. Let us begin with the drift estimator. Without any restriction on the relative (almost-sure) order of the various bandwidths, we cannot ensure that there is a vector $\mathbf{h}_{N,T}$ so that whenever (i) is violated, (ii)-(iii) cannot be violated. This may happen when $\max_{j}h_{j,N,T}^{5,dr}\Pi_{i\neq j}^{d}h_{i,N,T}^{dr}\widehat{\overline{L}}_{X}^{dr}(T,a)\overset{a.s.}{\to}\infty$ but $\min_{j}h_{j,N,T}^{5,dr}\Pi_{i\neq j}^{d}h_{i,N,T}^{dr}\widehat{\overline{L}}_{X}^{dr}(T,a)\overset{a.s.}{\to}0$. Broadly speaking, (ii)-(iii) only depend on the product on the bandwidths, while (i) depends both on the product and on the individual bandwidths. Therefore, in order to ensure the existence of bandwidths satisfying all conditions, we need to impose some restrictions on the degree of "heterogeneity" of their almost-sure order. We require that, for all j, $h_{j,N,T}^{dr}=O_{a.s.}\left(\left(\Pi_{i\neq j}^{d}h_{i,N,T}^{dr}\right)^{1/(d-1)}\right)$, so that the bandwidths can differ from each other but are of the same almost-sure order. Given that, whenever *(i)* is violated, $\Pi_{i\neq j}^{d}h_{i,N,T}^{dr}$ approaches zero almost surely at a rate equal or slower than $\widehat{\overline{L}}_{X}^{dr}(T,a)^{-\frac{d-1}{d+4}}$, and $\Pi_{i=1}^{d}h_{i,N,T}^{dr}$ cannot approach zero at a rate faster than $\widehat{\overline{L}}_{X}^{dr}(T,a)^{-\frac{d}{d+4}}$, it is immediate to see that *(ii)* is trivially satisfied, while *(iii)* writes as

$\left((\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}\widehat{\overline{L}}_X^{dr,\frac{2d+4}{d+4}}(T,a)\right)^{-1} \geq \left((\Delta_{N,T}\log(1/\Delta_{N,T}))^{1/2}T^{\frac{2d+4}{d+4}}\right)^{-1} \to \infty$ provided $N/T^{\frac{5d+12}{d+4}} \to \infty$. Imposing the restriction $h_{j,N,T}^{dif} = O_{a.s.}\left(\left(\Pi_{i\neq j}^{d} h_{i,N,T}^{dif}\right)^{1/(d-1)}\right)$, by an analogous argument, we see that, whenever $\frac{\max_j h_{j,N,T}^{5,dif}\Pi_{i\neq j}^{d} h_{i,N,T}^{dif}\widehat{\overline{L}}_X^{dr}(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0$ is violated, *(iii)* is satisfied provided $N/T^{\frac{3d+12}{4-d}} \to 0$. Thus, if we wished to allow for $d > 3$, we would need to rely on higher-order kernels.

Testing can now be conducted as in the scalar case. However, should we reject, contrary to the scalar case, we would not have a clear-cut indication of which particular bandwidth should be made larger or smaller. In spite of this, we do have information about whether we need to increase or decrease $\Pi_{i=1}^{d} h_{i,N,T}^{dif}$ and/or $\Pi_{i=1}^{d} h_{i,N,T}^{dr}$. Future work should focus on methods to adjust iteratively individual bandwidths.

7 Simulations

The goal of this simulation study is to illustrate absolute and relative performance of our methods (as compared to existing methods in the literature, such as cross-validation) as well as fundamental issues having to do with sample frequency in the fully nonparametric estimation of continuous-time processes. To this extent, we consider two data-generating processes, namely

$$\begin{aligned}
(1) \qquad \mathrm{d}X_t &= (0.1320 - 1.5918X_t)\mathrm{d}t + 2X_t^{1.49}\mathrm{d}W_t, \qquad X_0 = 0.08,\\
(2) \qquad \mathrm{d}X_t &= (0.02 - 0.025X_t)\mathrm{d}t + 0.14X_t^{1/2}\mathrm{d}W_t, \qquad X_0 = 0.6.
\end{aligned}$$

The parameters associated with the first process may derive from the estimation of a short-term interest rate diffusion model (see, e.g., Chan et al., 1992). The parameters associated with the second process may be used to model the dynamics of stochastic variance (see, e.g., Bandi and Renò, 2008). Both processes are highly persistent.

In what follows, the standard normal density $\phi(u)$ is chosen as the kernel function for all estimates. The remaining choice variables are set as follows: $\pi(u) = \phi(u)$, $U = [-1.5, 2.5]$, $\varepsilon = 0.001$, $R = 30$, $T = 22$, and $N = 5,500$. In other words, $\Delta_{N,T} = T/N = 22/5,500 = 1/250$, thereby implying that the simulated data points can be interpreted as being daily observations over 22 years. The resulting sample size is empirically sensible and relates to much applied work in which nonparametric continuous-time models are estimated by virtue of daily data (see, e.g., Stanton, 1997, and the references therein). As we will show, while daily data may deliver accurate estimates, the very nature of continuous-time models leads to (bandwidth) conditions which may not be easily satisfied with daily sample sizes. Finally, the number of replications is equal to $1,000$.

Fig. 1 and Fig. 2 report the (average) shape of the $\widetilde{V}$ statistics for both models along with the 95% critical value of the final min-min test (3.84). We recall that the feasible bandwidth set is the one for which the $\widetilde{V}$ statistics are above the critical value. For both models, $\widetilde{V}_2$ (which only plays a role in the drift case) is never binding. Thus, the bandwidth we select is always large enough as to satisfy $h_{N,T}^{dr}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$. Importantly, while for the drift estimator the set of bandwidths for which both $\widetilde{V}_1$ and $\widetilde{V}_3$ are rejected (and, hence, all rate conditions are satisfied) is generally non-empty, for the diffusion estimator such a set is empty (model 2) or "almost empty" (model 1), on average. This is because the bandwidth conditions for diffusion estimation are considerably more stringent. In effect, $\widetilde{V}_3$ is the same for both the drift and the diffusion. However, one needs $h_{N,T}^{dr,5}\overline{L}_X(T,a) \overset{a.s.}{\to} 0$ for drift estimation and $\frac{h_{N,T}^{dif,5}\overline{L}_X(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0$ for diffusion estimation. The latter requirement implies that the diffusion bandwidth ought to be smaller than the drift bandwidth. It may therefore be the case that the bandwidth condition which is required for a vanishing diffusion bias $\left(\frac{h_{N,T}^{dif,5}\overline{L}_X(T,a)}{\Delta_{N,T}} \overset{a.s.}{\to} 0\right)$ is too small for the almost-sure requirement $\left(\frac{\overline{L}_X(T,a)\sqrt{\Delta_{N,T}\log\left(1/\Delta_{N,T}\right)}}{h_{N,T}^{dif}\Delta_{N,T}} \overset{a.s.}{\to} 0\right)$ to be satisfied.[9] This outcome is entirely a function of the discretization interval $\Delta_{N,T}$. The smaller the interval, the more likely it is for diffusion estimator's feasible bandwidth set to be non-empty and, hence, for the almost-sure condition to be satisfied along with the vanishing-bias conditions.

This discussion illustrates a fundamental difficulty with estimating continuous-time models with discretely-sampled data. When the data frequency is not high enough (as may be the case with daily data), and the relevant convergence mode is almost-sure convergence, the discrete sample path of the process might not be a "sufficiently good" approximation for its continuous counterpart. Barring complications induced by the presence of market microstructure noise, the use of high-frequency data leading to smaller $\Delta_{N,T}$ values will help drastically. Alternatively, one could envision relaxing the mode of convergence. We conjecture that weaker modes would not require the rather stringent condition $\frac{\overline{L}_X(T,a)\sqrt{\Delta_{N,T}\log\left(1/\Delta_{N,T}\right)}}{h_{N,T}^{dif}\Delta_{N,T}} \overset{a.s.}{\to} 0$, thereby leading to well-posed bandwidth sets, in general (even for diffusion estimation). Work on this issue is warranted.

We now turn to drift and diffusion estimation (Fig. 3 through 10). The bandwidth selection mechanism works as follows. We begin with the first stage. If the first-stage bandwidth falls into the set in which all rate conditions are satisfied, we stop. Otherwise, we proceed until we reach a bandwidth in the interval for which $\widetilde{V}_1, \widetilde{V}_2$, and $\widetilde{V}_3$ (or $\widetilde{V}_1$ and $\widetilde{V}_3$ for the diffusion estimator) are all above the rejection line. If such a set is empty for our chosen daily frequency (see our discussion above), we use the following stopping rule. Suppose we choose a bandwidth which is too small, as it is generally the case. Thus, $\widetilde{V}_1$ lies above the 95% rejection line whereas $\widetilde{V}_3$ is below. We select a larger bandwidth and stop whenever $\widetilde{V}_1$ reaches the 95% critical value line. The reverse applies if we start with a bandwidth which is too large. Importantly, in both cases (too small or too large a bandwidth) we stop at a bandwidth value such that $\widetilde{V}_1$ reaches the 95% critical value (or is closest to it, from the left). The justification for this choice is simple. Whenever the discretization interval is so that we cannot satisfy $\min\left\{\widetilde{V}_1, \widetilde{V}_3\right\} > 3.84$, we sacrifice $\widetilde{V}_3$ and, consequently, the conditions for almost-sure convergence (which is, as pointed out above, specific to continuous-time models and might not be "necessary" for other modes of convergence to apply, in general).

There is an overwhelming tendency in empirical work conducted using continuous-time models to employ cross-validated bandwidths. While this procedure has a well-known theoretical rationale in

[9] Recall that the condition $\frac{\overline{L}_X(T,a)\sqrt{\Delta_{N,T}\log\left(1/\Delta_{N,T}\right)}}{h_{N,T}^{dif}\Delta_{N,T}} \overset{a.s.}{\to} 0$ ensures the almost sure convergence of the local time estimator as well as almost sure convergence of the drift and variance estimators.

discrete time, to our knowledge it has not been justified in continuous time. One of the objectives of this Monte Carlo experiment is, therefore, to evaluate the relative performance of bandwidths chosen via cross-validation and bandwidths selected by means of local Gaussianity, as is the case for our first stage smoothing sequences. We will also compare cross-validated bandwidths to our full procedure, inclusive of the second stage.

We observe that, for both models, cross-validation leads to the selection of excessively large bandwidths, thereby yielding substantial oversmoothing (see Figs. 3-4 and 7-8). Cross-validation works well only if the function to estimate is very flat, as is the case for the drift function in model 2.

The first-stage bandwidth chosen via local Gaussianity is substantially smaller than that chosen via cross-validation, and leads to more accurate, i.e., less-biased, estimates, in general (see, e.g., Figs. 3, 4, and 8). Put differently, exploiting the local Gaussianity that diffusion models imply is empirically useful.

Turning to the second stage, we find that the bandwidth chosen via local Gaussianity is, in general, smaller than the second-step bandwidth (see Tables 1-4). Figs. 1 and 2 provide a complete justification for this finding. The first-stage bandwidth is likely to ensure that $\widetilde{V}_1$ lies above the rejection line and the estimators' bias is negligible. It is, however, too small for $\widetilde{V}_3$ to lie above the rejection line as well. Thus, in the second stage, we select a larger bandwidth. As emphasized earlier, this outcome is *not* due to the nature of our methods but is solely a by-product of the fine grain features of continuous-time modelling and estimation and our employed sample frequency. Indeed, the condition underlying $\widetilde{V}_3$, i.e., $\frac{\overline{L}_X(T,a)\sqrt{\Delta_{N,T}\log\left(1/\Delta_{N,T}\right)}}{h_{N,T}\Delta_{N,T}} \overset{a.s.}{\to} 0$ is hard to satisfy for small and medium sample sizes N or, alternatively, for relatively large discrete-time intervals $\Delta_{N,T}$. In other words, if we were endowed with an N sufficiently large with respect to T, then all relevant conditions would be satisfied for reasonably small bandwidths, and our criterion would capture this effect. To see this, refer to Fig. 1. For a decreasing $\Delta_{N,T}$ (i.e., going from daily data, as in our case, to high-frequency data, for instance), the $\widetilde{V}_3$ curve would move to the left thereby (1) *increasing* the likelihood of a non-empty feasible set and (2) *decreasing* the size of the feasible bandwidths.

Importantly, overall, the second stage bandwidths are smaller than the bandwidths chosen via cross-validation. Hence, while our full-blown procedure may lead to oversmoothing for insufficiently small $\Delta_{N,T}$ (model 1, for instance), the degree of oversmoothing is still smaller than that delivered by cross-validated bandwidths.

It is also worthwhile to point out that, in spite of the fact that the bandwidth rate conditions for almost-sure convergence and zero asymptotic bias are more stringent in the diffusion case than in the drift case, the full procedure leads to nonparametric diffusion estimates which are more accurate than the corresponding drift estimates (see, e.g., Fig.5 vs. Fig.6 and Fig. 9 vs. Fig.10). Indeed, the set of (small) bandwidths for which $\widetilde{V}_1$ is rejected is very limited (see Fig. 1 and 2). Hence, even if in the second stage we move to a larger bandwidth, given our stopping rule we still select a rather small bandwidth which is not too far from the one chosen in the first stage.

In sum:

1. The existence of a feasible bandwidth set guaranteeing a zero asymptotic bias and almost-sure convergence crucially depends on the discretization interval $\Delta_{N,T}$. We show that daily frequencies are generally not very problematic as far as drift estimation is concerned (provided T is large

enough, of course) but may lead to empty feasible sets in the case of diffusion estimation. The reason for this is that, in the presence of a daily $\Delta_{N,T}$, the bandwidths for which the condition in $\widetilde{V}_1$ is rejected may be too small to be compatible with the larger bandwidths required for almost-sure convergence. Consistent with our theory, increasing the sampling frequency improves matters in that it leads to smaller required bandwidths for the conditions underlying $\widetilde{V}_3$ to be satisfied.

2. In spite of the use of low (daily) frequencies, our first-stage method performs extremely well for *both* functions and drastically better than cross-validation. Cross-validation leads to substantial oversmoothing (unless, of course, the relevant functions are rather flat).

3. In the presence of daily frequencies, our two-step method may lead to some oversmoothing (model 1, for example) but continues to perform better than cross-validation, in general.

4. Importantly, there is a clear theoretical justification for the oversmoothing which might be induced by our two-step procedure (i.e., an excessively large discretization interval for almost-sure consistency to be satisfied, thereby leading to the need for a larger bandwidth - see point 1 above). Hence, in our case, sub-optimal performance (as implied by some oversmoothing in certain cases) is a by-product of the very nature of our employed (daily) discrete data, as shown theoretically and by simulation. We cannot exclude that, in the case of cross-validation, sub-optimal performance may be due to fundamental limitations of the procedure itself.

8 Conclusions and further discussions

This paper provides an automated procedure to *jointly* select all bandwidths needed to identify the dynamics of popular classes of continuous-time Markov processes. It also proposes a randomized method designed to test whether the rate conditions for almost-sure consistency and (zero mean) asymptotic normality of the moment estimates are satisfied in sample. We study applicability of our theory in scalar and multivariate models allowing for jumps, microstructure noise, and stochastic volatility.

We also illustrate (theoretically and by virtue of simulations) issues of identifications in finite sample. Our discussion highlights potential problems which might arise when estimating nonparametrically continuous-time models by virtue of discretely-sampled observations. In particular, we emphasize that the classical use of daily data may prevent the bandwidth conditions for almost-sure consistency and zero-mean asymptotic normality from being satisfied. In light of the widespread use of daily data in applied work, we view this observation as being empirically very important.

The methods proposed in this paper are of general interest. Analogous ideas may be applied to bandwidth selection for recurrent discrete-time Markov processes. Our randomized second-step procedure may also prove useful in a variety of alternative nonparametric estimation settings. Below we provide brief discussions of both issues and refer the reader to future work for complete treatments.

8.1 The discrete-time case

Consider the recurrent discrete-time kernel case. The residuals $\epsilon_t = (y_t - \mu(X_t)) / \sqrt{\sigma^2(X_t) - \mu^2(X_t)}$, where $\mu(.)$ is the conditional first moment of the y data and $\sigma^2(.)$ is the conditional second moment, are *not* locally Gaussian in general. However, it is immediate to see that, e.g., $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{E}\left(\epsilon_t^2\right) = 1$, $\mathrm{E}(\epsilon_t g(X_t)) = 0$, and $\mathrm{E}\left(\epsilon_t^2 g(X_t)\right)$ $=\mathrm{E}(g(X_t))$ for any $\mathcal{F}_X-$measurable function $g(.)$. Thus, one may select the bandwidth(s) in such a way as to minimize an appropriately-defined distance metric between sample moments of the residuals and their theoretical counterparts. Interestingly, the problem is easier than in continuous time. First, the initial criterion would yield uniform consistency of both conditional moments since, differently from our assumed continuous-time framework, the two moments would converge at the same rate (i.e., $\sqrt{h_T \widehat{L}_T(x)}$, where $\widehat{L}_T(x)$ is, as earlier, the empirical occupation density of the underlying discrete-time process). Second, the bandwidth conditions needed to be tested would closely resemble those for the drift (in Proposition 1). Importantly, however, the condition on the modulus of continuity of Brownian motion (i.e., the condition for almost-sure consistency in the continuous-time case) would not be needed. Hence, the second-step procedure would simply amount to testing whether, in-sample, the selected bandwidths are proportional to $\widehat{L}_T^{-\beta}(x)$ with $\frac{1}{5} < \beta < 1$. Also, issues of identification having to do with the coarseness of the sampling frequency (as in the continuous-time case) would not arise.

8.2 More on the second-step method

In both the stationary and the nonstationary case, irrespective of whether we operate in continuous time or in discrete time, the bandwidth conditions needed for consistency and (zero mean) asymptotic normality of kernel estimators can be expressed as functions of the process' occupation density (and its divergence rate). Even in cases for which the divergence rate of the occupation density can be quantified in closed-form (the stationary case, for example, for which it is T), relying on an in-sample assessment of the process' occupation density, rather than on purely-hypothetical divergence rates, is bound to provide a more objective evaluation of the accuracy of bandwidth choices (particularly for persistent processes). Our second-step procedure is designed to explicitly achieve this goal.

Importantly, however, our testing method may be disconnected from the first-stage method and applied to smoothing sequences selected by virtue of alternative, possibly more classical, methods of the kind routinely used in applied work. More generally, our test (and its logic) may, in principle, be extended to evaluate any choices in functional econometrics requiring the balancing of an asymptotic (and finite sample) trade-off between bias and variance. The number of sieves or the number of autocovariances in HAC estimation are possible examples. In these contexts, a test (like the one proposed in this paper) which, under the null, implies that the assumed choice is either too small or too large and provides, if the null is not rejected, an easy automated rule to adjust the initial selection in either direction appears to be very appealing.

9 Appendix A

Proof of Theorem 1. Assume $\mathbf{h}^*_{N,T} \in H$ exists and satisfies

$$\sup_x \left|F_{\overline{N}}^{\mathbf{h}}(x) - \Phi(x)\right| \underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\to}} 0. \tag{25}$$

Using the triangular inequality, write

$$\sup_x \left|F_{\overline{N}}^{\mathbf{h}}(x) - \Phi(x)\right| \geq \sup_x \left|F_{\overline{N}}^{\mathbf{h}}(x) - F_{\overline{N}}(x)\right| - \sup_x \left|F_{\overline{N}}(x) - \Phi(x)\right|,$$

where $F_{\overline{N}}(x)$ is the empirical distribution function of

$$\left\{\varepsilon_{i\Delta_{N,T}} = \frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \mu(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\sigma(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} : i = 2, ..., \overline{N}\right\}.$$

An application of the Law of Iterated Logarithm implies $\sup_x |F_{\overline{N}}(x) - \Phi(x)| = o_p(1)$. The result in Eq. (25) combined with $\sup_x |F_{\overline{N}}(x) - \Phi(x)| = o_p(1)$ yields

$$\sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - F_{\overline{N}}(x)\right| = o_p(1).$$

But,

$$\begin{aligned}
& \sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - F_{\overline{N}}(x)\right| \\
= \ & \sup_x \Big| \frac{1}{\overline{N}-1} \sum_{i=2}^{\overline{N}} \mathbf{1}\left(\frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \mu(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\sigma(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \leq x \frac{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})} \right. \\
& \left. - \frac{\left(\mu(X_{(i-1)\Delta_{N,T}}) - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\right)\Delta_{N,T}}{\sigma(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \right) \\
& - \frac{1}{\overline{N}-1} \sum_{i=2}^{\overline{N}} \mathbf{1}\left(\frac{X_{i\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}} - \mu(X_{(i-1)\Delta_{N,T}})\Delta_{N,T}}{\sigma(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}} \leq x \right) \Big| \\
= \ & o_p(1)
\end{aligned}$$

gives

$$\sup_{a \in \mathfrak{D}} \left| \frac{\widehat{\sigma}_{N,T}(a)}{\sigma(a)} - 1 \right| = o_p(1)$$

and

$$\sup_{a \in \mathfrak{D}} \left| \frac{\left(\mu(a) - \widehat{\mu}_{N,T}(a)\right)\sqrt{\Delta_{N,T}}}{\sigma(a)} \right| = o_p(1).$$

The converse follows from

$$\sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - \Phi(x)\right| \leq \sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - F_{\overline{N}}(x)\right| + \sup_x \left|F_{\overline{N}}(x) - \Phi(x)\right|.$$

■

Proof of Theorem 2. Assume $\mathbf{h}^*_{N,T} \in H$ exists. Let $\Gamma(., \varepsilon) \subset H$ be an open ball of radius ε. Then, from Eq. (8) and Eq. (7), $\forall \varepsilon > 0, \exists \delta > 0$:

$$P\left(\widehat{\mathbf{h}}_{N,T} \notin \Gamma(\mathbf{h}^*_{N,T}, \varepsilon)\right) \leq P\left(\theta_{\overline{N}} + \sup_x \left|\Phi(x) - F^{\mathbf{h}^*_{N,T}}_{\overline{N}}(x)\right| \geq \delta > 0\right) \underset{N,T\to\infty,\Delta_{N,T}\to 0}{\rightarrow} 0.$$

This proves the second part of the theorem. Now we need to show that

$$\exists \mathbf{h}^*_{N,T} = \mathbf{h} \in H : \sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - \Phi(x)\right| \underset{N,T\to\infty,\Delta_{N,T}\to 0}{\overset{p}{\rightarrow}} 0.$$

As $\sup_x |F_{\overline{N}}(x) - \Phi(x)| = o_p(1)$, and given the triangular inequality, it suffices to show that

$$\exists \mathbf{h}^*_{N,T} = \mathbf{h} \in H : \sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - F_{\overline{N}}(x)\right| \underset{N,\overline{N},T\to\infty}{\overset{p}{\rightarrow}} 0.$$

Recalling the definition of $F^{\mathbf{h}}_{\overline{N}}(x)$ in the proof of Theorem 1, note that

$$\sup_x \left|F^{\mathbf{h}}_{\overline{N}}(x) - F_{\overline{N}}(x)\right| = \sup_x |Z_{\overline{N}}(x,\mathbf{h})| + \sup_x |H_{\overline{N}}(x,\mathbf{h})|$$

where

$$\begin{aligned} Z_{\overline{N}}(x,\mathbf{h}) &= \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left\{\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}} \le \zeta_{(i-1)\Delta_{N,T}}(x)\right) - \Phi(\zeta_{(i-1)\Delta_{N,T}}(x)) - \mathbf{1}\left(\varepsilon_{i\Delta_{N,T}} \le x\right) + \Phi(x)\right\}, \\ H_{\overline{N}}(x,\mathbf{h}) &= \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left\{\Phi(\zeta_{(i-1)\Delta_{N,T}}(x)) - \Phi(x)\right\}. \end{aligned}$$

and

$$\zeta_{(i-1)\Delta_{N,T}}(x) = x\left(\frac{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right) - \frac{\left(\mu(X_{(i-1)\Delta_{N,T}}) - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})\right)\Delta_{N,T}}{\sigma_{N,T}(X_{(i-1)\Delta_{N,T}})\sqrt{\Delta_{N,T}}}.$$

We start by bounding $\sup_x |H_{\overline{N}}(x,\mathbf{h})|$. By the mean-value theorem, letting $\pi_{(i-1)\Delta_{N,T}}(x)$ be a value on the line segment connecting x and $\zeta_{(i-1)\Delta_{N,T}}(x)$,

$$\begin{aligned} &\sup_x \frac{1}{\overline{N}-1}\left|\sum_{i=2}^{\overline{N}}\left\{\Phi^{'}(\pi_{(i-1)\Delta_{N,T}}(x))\left(\zeta_{(i-1)\Delta_{N,T}}(x) - x\right)\right\}\right| \\ \le\; &\sup_x \max_i \left|x\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| \left|\frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left(\frac{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}}) - \sigma(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right)\right| \\ &+ \sup_x \max_i \left|\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| \left|\frac{\sqrt{\Delta_{N,T}}}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left(\frac{\mu(X_{(i-1)\Delta_{N,T}}) - \widehat{\mu}_{N,T}(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right)\right| \\ =\; &\sup_x \max_i \left|x\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| I_{N,T} + \sup_x \max_i \left|\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| II_{N,T} \end{aligned}$$

We begin by considering $II_{N,T}$.

$$\begin{aligned} II_{N,T} \;\le\; &\sup_x \frac{1}{\sigma(x)}\left|\frac{\sqrt{\Delta_{N,T}}}{\overline{N}}\sum_{i=1}^{\overline{N}}\frac{\sum_{j=1}^{N} K\left(\frac{X_{(j-1)\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}}}{h^{dr}_{N,T}}\right)\Delta^{-1}_{N,T}\int_{(j-1)\Delta_{N,T}}^{j\Delta_{N,T}}\left(\mu(X_s) - \mu(X_{(i-1)\Delta_{N,T}})\right)\mathrm{d}s}{\sum_{j=1}^{N} K\left(\frac{X_{(j-1)\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}}}{h^{dr}_{N,T}}\right)}\right| \\ &+ \sup_x \frac{1}{\sigma(x)}\left|\frac{\sqrt{\Delta_{N,T}}}{\overline{N}}\sum_{i=1}^{\overline{N}}\frac{\sum_{j=1}^{N} K\left(\frac{X_{(j-1)\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}}}{h^{dr}_{N,T}}\right)\Delta^{-1}_{N,T}\int_{(j-1)\Delta_{N,T}}^{j\Delta_{N,T}}\sigma(X_s)\mathrm{d}W_s}{\sum_{j=1}^{N} K\left(\frac{X_{(j-1)\Delta_{N,T}} - X_{(i-1)\Delta_{N,T}}}{h^{dr}_{N,T}}\right)}\right| \\ =\; &O_p\left(\sqrt{\Delta_{N,T}}h^{dr}_{N,T}\right) + O_p\left(\sqrt{\Delta_{N,T}\frac{1}{h^{dr}_{N,T}\inf_x \overline{L}_T(x)}}\right) = o_p(1). \end{aligned}$$

As for $\sup_x \max_i \left|x\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| I_{N,T}$, by a similar argument as that in the proof of Theorem 1 in Corradi and Distaso (2008):

$$\begin{aligned}
& \sup_x \max_i \left|x\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| \left|\frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}}\left(\frac{\widehat{\sigma}^2_{N,T}(X_{(i-1)\Delta_{N,T}})-\sigma^2(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right)\right| \\
\leq\ & \sup_x \max_i \left|x\Phi^{'}(\pi_{(i-1)\Delta_{N,T}})\right| \sup_x \frac{1}{\sigma(x)}\left|\frac{1}{\overline{N}}\sum_{i=1}^{\overline{N}}\left(\widehat{\sigma}^2_{N,T}(X_{(i-1)\Delta_{N,T}})-\sigma^2(X_{(i-1)\Delta_{N,T}})\right)\right| \\
=\ & O_p\left(\frac{1}{\sqrt{\overline{N}}}\right)=o_p(1).
\end{aligned}$$

We now turn to a bound for $\sup_x |Z_{\overline{N}}(x,\mathbf{h})|$. As in Bai (1994) and Lee and Wei (1999), we divide the real line into M_N points $-\infty = x_{N,0} < ... < x_{N,M_N} = \infty$ such that $M_N = \left[N^{1/2}\right]$ and $\Phi(x_{N,i}) = \frac{i}{M_N}$ for $i = 0, ..., M_N$. Consider now a generic value $x \in (x_{N,u}, x_{N,u+1}]$. Write

$$\begin{aligned}
Z_{\overline{N}}(x,\mathbf{h}) =\ & \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left\{\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}}\leq \zeta_{(i-1)\Delta_{N,T}}(x)\right)-\Phi(\zeta_{(i-1)\Delta_{N,T}}(x_{N,u+1}))-\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}}\leq x_{N,u+1}\right)+\Phi(x_{N,u+1})\right\} \\
& +\frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left\{\Phi(\zeta_{(i-1)\Delta_{N,T}}(x_{N,u+1}))-\Phi(\zeta_{(i-1)\Delta_{N,T}}(x))\right\} \\
& +\frac{1}{\overline{N}-1}\sum_{i=1}^{\overline{N}}\left\{\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}}\leq x_{N,u+1}\right)-\Phi(x_{N,u+1})+\Phi(x)-\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}}\leq x\right)\right\} \\
=\ & \Theta_{N1}+\Theta_{N2}+\Theta_{N3}.
\end{aligned}$$

Since the maps $x\to\sum_{i=1}^{\overline{N}}\mathbf{1}\left(\varepsilon_{i\Delta_{N,T}}\leq\zeta_{(i-1)\Delta_{N,T}}(x)\right)$ and $x\to\Phi(\zeta_{(i-1)\Delta_{N,T}}(x))$ are nondecreasing, then

$$\begin{aligned}
& \sup_x|\Theta_{N2}| \\
\leq\ & \max_u \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left|\Phi(\zeta_{(i-1)\Delta_{N,T}}(x_{N,u+1}))-\Phi(\zeta_{(i-1)\Delta_{N,T}}(x_{N,u}))\right| \\
=\ & \max_u \frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left|\Phi^{'}\left(\zeta_{(i-1)\Delta_{N,T}}(\widetilde{x}_{N,u})\right)(x_{N,u+1}-x_{N,u})\frac{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right| \\
\leq\ & \left|\sup_x\Phi^{'}(x)\right| O\left(\frac{1}{\sqrt{N}}\right)\left|\frac{1}{\overline{N}-1}\sum_{i=2}^{\overline{N}}\left(\frac{\widehat{\sigma}_{N,T}(X_{(i-1)\Delta_{N,T}})-\sigma(X_{(i-1)\Delta_{N,T}})}{\sigma(X_{(i-1)\Delta_{N,T}})}\right)\right| \\
& +\left|\sup_x\Phi^{'}(x)\right| O\left(\frac{1}{\sqrt{N}}\right) \\
=\ & \left|\sup_x\Phi^{'}(x)\right| O\left(\frac{1}{\sqrt{N}}\right) O_p\left(\frac{1}{\sqrt{N}}\right)+O\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}$$

from previous results. Then, $\sup_x|\Theta_{N2}| = o_p(1)$. Now consider Θ_{N3} and write

$$
\begin{aligned}
&\sup_x \max_u \Theta_{N3} \\
\leq\ & \frac{1}{\sqrt{N}} \sup_{p,k:|p-k|\leq\frac{1}{\sqrt{N}}} \frac{1}{\sqrt{N}} \sum_{i=1}^{\overline{N}} \mathbf{1}\left(\varepsilon_{i\Delta_{N,T}} \leq \Phi^{-1}(p)\right) - p + k - \mathbf{1}\left(\varepsilon_{i\Delta_{N,T}} \leq \Phi^{-1}(k)\right) \\
=\ & \frac{1}{\sqrt{N-1}} o_p(1)
\end{aligned}
$$

following Stute (1982) given $\varepsilon_t = O\left(\overline{\Delta}_t\right) + \overline{\varepsilon}_t$. Finally, $\sup_x |\Theta_{N1}| = o_p(1)$ can be shown similarly. This proves the stated result. ■

Proof of Theorem 3. We begin with (i). Suppose that $V_{R,N,T} = \widetilde{V}_{1,R,N,T}$. Without loss of generality, we assume that[10] $\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)$ diverges at a rate faster than $\log T$ $\forall a \in \mathfrak{D}$. First note that for all j, conditional on the sample, $v_{1,j,N,T} \simeq N\left(0, \exp\int_{\mathcal{A}}\left(\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)\right)$. Let

$$
\Omega_{N,T} = \left\{\omega : T^{-1}\left(\exp\int_{\mathcal{A}}\left(\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)\right) > \Delta \text{ for } \Delta \text{ arbitrarily large}\right\}
$$

so that, under H_0, $P\left(\lim_{N,T\to\infty}\Omega_{N,T}\right) = 1$. We shall proceed conditional on $\omega \in \Omega_{N,T}$. For any $u \in U$, assuming, without loss of generality, $u > 0$, we obtain

$$
\begin{aligned}
V_{R,N,T}(u) &= \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\left\{v_{1,j,N,T} \leq u\right\} - E^*\left(1\left\{v_{1,j,N,T} \leq u\right\}\right)\right) \\
&\quad + \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(E^*\left(1\left\{v_{1,j,N,T} \leq u\right\}\right) - \frac{1}{2}\right),
\end{aligned}
$$

where $E^*\left(1\left\{v_{1,j,N,T} \leq u\right\}\right) = 1/2 + P^*\left(0 \leq v_{1,i,N,T} \leq u\right)$. Now, uniformly in u,

$$
\begin{aligned}
&P^*\left(0 \leq v_{1,j,N,T} \leq u\right) \\
=\ & \frac{1}{\left(\pi \exp\int_{\mathcal{A}}\left(\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)\right)^{1/2}} \times \int_0^u \exp\left(-\frac{x^2}{2\exp\int_{\mathcal{A}}\left(\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\mathrm{d}a\right)}\right) dx \\
=\ & O_p\left(T^{-1/2}\right).
\end{aligned}
\tag{26}
$$

Thus,

$$
V_{R,N,T}(u) = \frac{2}{\sqrt{R}} \sum_{j=1}^{R} \left(1\left\{v_{1,j,N,T} \leq u\right\} - E^*\left(1\left\{v_{1,j,N,T} \leq u\right\}\right)\right) + o_p(1).
$$

Given (26), and recalling that $E^*\left(v_{1,s,N,T}v_{1,j,N,T}\right) = 0$ for $s \neq j$,

$$
\begin{aligned}
&Var^*\left(\frac{1}{\sqrt{R}} \sum_{j=1}^{R} \left(1\left\{v_{1,j,N,T} \leq u\right\} - E^*\left(1\left\{v_{1,j,N,T} \leq u\right\}\right)\right)\right) \\
=\ & \frac{1}{R} \sum_{j=1}^{R} \left(E^*\left(1\left\{v_{1,j,N,T} \leq u\right\} - E^*(1\left\{v_{1,j,N,T} \leq u\right\})\right)^2\right) \\
=\ & 1/4 + O_P(T^{-1/2}),
\end{aligned}
$$

[10] In fact, we could allow it to diverge at rate $\log(\log T)$ simply by using $\exp\left(\exp\left(\sup_{a\in\mathcal{D}}\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\right)\right)$ in (9) instead of $\exp\left(\sup_{a\in\mathcal{D}}\widehat{h}_{N,T}^{dr,5-\varepsilon}\widehat{\overline{L}}_X(T,a)\right)$.

where the $O_p(T^{-1/2})$ holds uniformly in u. Note that the asymptotic variance is equal to 1/4 regardless of the evaluation point u. This is an immediate consequence of that fact that, as $N, T \to \infty$, $1\{v_{1,j,N,T} \leq u\}$ takes the same value, either 0 or 1, irrespective of the evaluation point u. Hence, $\int_U V_{1,R,N,T}^2(u)\pi(u)du \xrightarrow{d^*} \int_U \chi_1^2 \pi(u)du \equiv \chi_1^2$.

We now turn to (ii). Let

$$\Omega_{N,T}^{+} = \left\{ \omega : \left(\exp \int_{\mathcal{A}} \left(\widehat{h}_{N,T}^{dr,5-\varepsilon} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \right) \right) < \Delta,\ \Delta < \infty \right\}$$

so that, under H_A, $P\left(\lim_{N,T\to\infty} \Omega_{N,T}^{+}\right) = 1$. For $\omega \in \Omega_{N,T}^{+}$, $\exp \int_{\mathcal{A}} \left(\widehat{h}_{N,T}^{dr,5-\varepsilon} \widehat{\overline{L}}_X(T,a)\mathrm{d}a \right) \xrightarrow{a.s.} M \geq 1$. Hence, $v_{1,j,N,T} \xrightarrow{d^*} N(0, M)$. Let $F(u)$ be the cumulative distribution function of a zero-mean normal random variable with variance M. Now,

$$\begin{aligned} & \frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\{v_{1,i,N,T} \leq u\} - \frac{1}{2} \right) \\ = \ & \frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\{v_{1,i,N,T} \leq u\} - F(u) \right) + 2\sqrt{R} \left(F(u) - \frac{1}{2} \right). \end{aligned} \tag{27}$$

Note that $F(u) = \frac{1}{2}$ if, and only if, $u = 0$. Thus, $R \left(\int_U \left(F(u) - \frac{1}{2} \right)^2 \pi(u)\mathrm{d}u \right)$ diverges to infinity at rate R. As for the first term on the right-hand side of Eq. (27), $\int_U \left(\frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\{v_{1,i,N,T} \leq u\} - F(u) \right) \right)^2 \pi(u)\mathrm{d}u = O_{p^*}(1)$ by the same arguments used in the proof of Theorem 1(ii) in Corradi and Swanson (2006).

Proof of Theorem 4 Similar to that of Theorem 3.

Proof of Theorem 5 Because the compensator $\lambda\left(X_{t_-}\right) \mathrm{E}_y \left(c\left(X_{t_-}, y\right) \right)$ can be treated as a component of the drift function, for notational simplicity we denote the jump component by J_t and the jumpless component by Y_t. Thus, $X_t = Y_t + J_t$. We first show that

$$\begin{aligned} & \widehat{\sigma}_{\mathbf{J},N,T}^2(a) - \frac{\mu_{2/p}^{-p}}{\Delta_{N,T}} \frac{\sum_{j=1}^{N-p} K\left(\frac{X_{j\Delta_{N,T}} - a}{h_{n,T}^{dif}} \right) \Pi_{i=1}^{p} \left| Y_{(j+i)\Delta_{N,T}} - Y_{(j+i-1)\Delta_{N,T}} \right|^{\frac{2}{p}}}{\sum_{j=1}^{N} K\left(\frac{X_{j\Delta_{N,T}} - a}{h_{N,T}^{dif}} \right)} \\ = \ & o_p\left(\sqrt{\frac{\Delta_{N,T}}{h_{N,T}^{dif} \overline{L}_X(T,a)}} \right) \end{aligned} \tag{28}$$

with $\widehat{\sigma}_{\mathbf{J},N,T}^2(a)$ defined as in Eq. (16). Hereafter, let $\Delta Y_{(j+i)\Delta_{N,T}} = \left(Y_{(j+i)\Delta_{N,T}} - Y_{(j+i-1)\Delta_{N,T}} \right)$, and let $\Delta X_{(j+i)\Delta_{N,T}}$ and $\Delta J_{(j+i)\Delta_{N,T}}$ be defined in an analogous manner. Since

$$|\Delta Y + \Delta J|^r \leq (|\Delta Y| + |\Delta J|)^r \leq |\Delta Y|^r + |\Delta J|^r$$

by the triangle inequality, monotonicity, and concavity given $r \leq 1$, it follows that

$$\sqrt{\frac{h_{N,T}^{dif}\overline{L}_X(T,a)}{\Delta_{N,T}}}\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\frac{\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)\left(\Pi_{i=1}^{p}\left|\Delta X_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}-\Pi_{i=1}^{p}\left|\Delta Y_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}\right)}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}$$

$$\begin{aligned}
&\leq \sqrt{\frac{h_{N,T}^{dif}\overline{L}_X(T,a)}{\Delta_{N,T}}}\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\frac{\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}\left(\Pi_{i=1}^{p}\left|\Delta J_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}\right.\\
&\quad\left.+\binom{p}{1}\left|\Delta Y_{(j+1)\Delta_{N,T}}\right|^{\frac{2}{p}}\Pi_{i=2}^{p}\left|\Delta J_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}+...+\binom{p}{p-1}\left|\Delta J_{(j+p)\Delta_{N,T}}\right|^{\frac{2}{p}}\Pi_{i=1}^{p-1}\left|\Delta Y_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}\right)\\
&\simeq_p \sqrt{\frac{h_{N,T}^{dif}\overline{L}_X(T,a)}{\Delta_{N,T}}}\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\frac{\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}\left(\Pi_{i=1}^{p}\delta_{(j+i)\Delta_{N,T}}\right.\\
&\quad\left.+\binom{p}{1}\left|\Delta Y_{(j+1)\Delta_{N,T}}\right|^{\frac{2}{p}}\Pi_{i=2}^{p}\delta_{(j+i)\Delta_{N,T}}+...+\binom{p}{p-1}\delta_{(j+p)\Delta_{N,T}}\Pi_{i=1}^{p-1}\left|\Delta Y_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}\right), \qquad (29)
\end{aligned}$$

where $\delta_{(j+i)\Delta_{N,T}}=1$ if $\Delta J_{(j+i)\Delta_{N,T}}\neq 0$ and $\delta_{(j+i)\Delta_{N,T}}=0$ if $\Delta J_{(j+i)\Delta_{N,T}}=0$, and $\simeq_p$ signifies "of the same probability order." Because on every fixed time span there is at most a finite number of jumps, Pr $\left(\Pi_{i=1}^{k}\delta_{(j+i)\Delta_{N,T}}=1\right)=O\left(\Delta_{N,T}^{k}\right)$ for all $k\geq 1$. Both E $\left(\Pi_{i=1}^{p}\delta_{(j+i)\Delta_{N,T}}\right)$ and $Var\left(\Pi_{i=1}^{p}\delta_{(j+i)\Delta_{N,T}}\right)$ are $O\left(\Delta_{N,T}^{p}\right)$. Since $\Pi_{i=1}^{p-k}\left|\Delta Y_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}=O_p\left(\Delta_{N,T}^{\frac{p-k}{p}}\right)$ for $k=1,...,p-1$, then the term of higher order is the last one. All other terms are of a smaller probability order. Write

$$\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\frac{\frac{\Delta_{N,T}}{h_{N,T}^{dif}\overline{L}_X(T,a)}\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)\delta_{(j+p)\Delta_{N,T}}\Pi_{i=1}^{p-1}\left|\Delta Y_{(j+i)\Delta_{N,T}}\right|^{\frac{2}{p}}}{\frac{\Delta_{N,T}}{h_{N,T}^{dif}\overline{L}_X(T,x)}\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}=O_p\left(\Delta_{N,T}^{\frac{p-1}{p}}\right).$$

Hence, the statement in Eq. (28) follows as $\sqrt{h_{N,T}^{dif}\overline{L}_X(T,a)}\Delta_{N,T}^{\frac{p-2}{2p}}\overset{a.s.}{\to}0$ for $p>2$. Finally, note that

$$\begin{aligned}
&\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\frac{\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)\Pi_{i=1}^{p}\left|Y_{(j+i)\Delta_{N,T}}-Y_{(j+i-1)\Delta_{N,T}}\right|^{\frac{2}{p}}}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}-\sigma^2(a)\\
&\simeq_p \frac{\frac{\mu_{2/p}^{-p}}{\Delta_{N,T}}\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)\left(\sigma^2\left(X_{j\Delta_{N,T}}\right)\frac{\Pi_{i=1}^{p}\left|W_{(j+i)\Delta_{N,T}}-W_{(j+i-1)\Delta_{N,T}}\right|^{\frac{2}{p}}}{\Delta_{N,T}}-\mu_{2/p}^{p}\sigma^2\left(X_{j\Delta_{N,T}}\right)\Delta_{N,T}\right)}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}\\
&\quad+\frac{\sum_{j=1}^{N-p}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{n,T}^{dif}}\right)\sigma^2\left(X_{j\Delta_{N,T}}\right)}{\sum_{j=1}^{N}K\left(\frac{X_{j\Delta_{N,T}}-a}{h_{N,T}^{dif}}\right)}-\sigma^2(a)\\
&= I_{N,T}+II_{N,T}.
\end{aligned}$$

Since, given the filtration $\mathfrak{F}_{j\Delta_{N,T}}=\sigma(X_s;s\leq j\Delta_{N,T})$, $\mathrm{E}_{j\Delta_{N,T}}\left(\sigma^2\left(X_{j\Delta_{N,T}}\right)\Pi_{i=1}^{p}\left|W_{(j+i)\Delta_{N,T}}-W_{(j+i-1)\Delta_{N,T}}\right|^{\frac{2}{p}}\right)=\mu_{2/p}^{p}\sigma^2\left(X_{j\Delta_{N,T}}\right)\Delta_{N,T}$, the term $I_{N,T}$ averages martingale differences. Using Bandi and Phillips (2003, Theorem

3), its distribution is mixed normal and its rate of convergence is $\sqrt{\frac{h_{N,T}^{dif}\widehat{\overline{L}}_X(T,a)}{\Delta_{N,T}}}$. The form of its asymptotic variance can be found along the lines of Corradi and Distaso (2008). The term $II_{N,T}$ is a bias term with order $O_p\left(h_{N,T}^{dif,2}\right)$. This proves the stated result.■

Proof of Theorem 6. We begin with the drift. Write the estimation error decomposition as

$$\begin{aligned}
&\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\left(\widehat{\mu}_{l,T}(x)-\mu(x)\right)\\
= \; & \frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)-K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\right)\\
&\times\frac{\Delta_{l,T}^{-1}\left(Y_{(jB+b)\Delta_{N,T}}-Y_{((j-1)B+b)\Delta_{N,T}}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b/T)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}+\\
&+\left(\frac{\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\left(Y_{(jB+b)\Delta_{N,T}}-Y_{((j-1)B+b)\Delta_{N,T}}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}\right.\\
&\left.-\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\mu(x)\right)\\
= \; & I_{N,T,l}+II_{N,T,l}. \qquad (30)
\end{aligned}$$

Note that

$$\begin{aligned}
&K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)-K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\\
\simeq_p\; & 1\left\{X_{((j-1)B+b)\Delta_{N,T}}\in\left(x-h_{N,l,T}^{dr},x-h_{N,l,T}^{dr}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)\cup\left(x+h_{N,l,T}^{dr},x+h_{N,l,T}^{dr}-\epsilon_{((j-1)B+b/T)\Delta_{N,T}}\right)\right\}.
\end{aligned}$$

Recalling that $\mathrm{E}\left(\epsilon_{i\Delta_{N,T}}^k\right)=O\left(a_{N,T}^{k/2}\right)$, with $a_{N,T}\to 0$ as $N,T\to\infty$, a straightforward application of Markov inequality ensures that $\sup_{i=1,\ldots,N}\left|\epsilon_{i\Delta_{N,T}}\right|=O_{a.s.}\left(N^{1/k}a_{N,T}^{1/2}\right)$. Thus,

$$\begin{aligned}
I_{N,T,l}\;\simeq_p\; & \frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}1\left\{x+h_{N,l,T}^{dr}\le X_{((j-1)B+b)\Delta_{N,T}}\le x+h_{N,l,T}^{dr}+N^{1/k}a_{N,T}^{1/2}\right\}\\
&\times\frac{\Delta_{l,T}^{-1}\left(X_{(jB+b)\Delta_{N,T}}-X_{((j-1)B+b)\Delta_{N,T}}\right)+\Delta_{l,T}^{-1}\left(\epsilon_{(jB+b)\Delta_{N,T}}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}.
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{\Delta_{N,T}}{N^{1/k}a_{N,T}^{1/2}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}1\left\{x+h_{N,l,T}^{dr}\le X_{((j-1)B+b)\Delta_{N,T}}\le x+h_{N,l,T}^{dr}+N^{1/k}a_{N,T}^{1/2}\right\}\\
= \; & O_p\left(\overline{L}_X(T,x+h_{N,l,T}^{dr})\right)=O_p(\overline{L}_X(T,x))+O_p\left(\sqrt{h_{N,l,T}^{dr}}\right)
\end{aligned}$$

then, recalling that $X_{((j-1)B+b/T)\Delta_{N,T}}$ is independent of $\epsilon_{((j-1)B+b/T)\Delta_{N,T}}$ for all j, it is easy to see that

$\mathrm{E}\left(I_{N,T,l}\right)=O_p\left(\frac{N^{1/k}a_{N,T}^{1/2}\sqrt{\overline{L}_X(T,x)}}{\sqrt{h_{N,l,T}^{dr}}}\right)$. As for the second moment,

$$\mathrm{E}\left(I_{N,T,l}^2\right)=O\left(\frac{N^{1/k}a_{N,T}^{1/2}}{h_{N,l,T}^{dr}}+\frac{a_{N,T}^{3/2}N^{1/k}\Delta_{N,T}}{h_{N,l,T}^{dr}\Delta_{l,T}^2}\right),$$

where the order of the first term can be derived as in the case of Eq. (31) and the order of the second term can be obtained as in the case of Eq. (34) below (in both cases with the indicator kernel in place of a smooth kernel). We now note that

$$\frac{N^{1/k}a_{N,T}^{1/2}\sqrt{\overline{L}_X(T,x)}}{\sqrt{h_{N,l,T}^{dr}}}\overset{a.s.}{\to}0\Rightarrow\frac{N^{1/k}a_{N,T}^{1/2}}{h_{N,l,T}^{dr}}\to 0$$

since $h_{N,l,T}^{dr}\overline{L}_X(T,x)\overset{a.s.}{\to}\infty$ (see below). Now write:

$$\begin{aligned}
&\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\left(\widehat{\mu}_{l,T}(x)-\mu(x)\right)=II_{N,T}+o_p(1)\\
=\ &\frac{\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\left(\Delta_{l,T}^{-1}\left(X_{(jB+b/T)\Delta_{N,T}}-X_{((j-1)B+b/T)\Delta_{N,T}}\right)-\mu(x)\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b/T)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}\\
&+\frac{\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b/T)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\left(\epsilon_{(jB+b/T)\Delta_{N,T}}-\epsilon_{((j-1)B+b/T)\Delta_{N,T}}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b/T)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}+o_p(1)\\
=\ &A_{N,l,T}+B_{N,l,T}+o_p(1).
\end{aligned}$$

We first show that $B_{N,l,T}$ is $o_p(1)$. Because the denominator is bounded away from zero, it suffices to show that the numerator is $o_p(1)$. Write

$$\begin{aligned}
&\operatorname{var}\left(\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\left(\epsilon_{(jB+b)\Delta_{N,T}}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)\right)\\
\simeq\ &O\left(\frac{\Delta_{N,T}^2}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)\Delta_{N,T}^{-1}\Delta_{l,T}^{-2}a_{N,T}\right)=O\left(\frac{\Delta_{N,T}a_{N,T}}{\Delta_{l,T}^2}\right)=o(1)
\end{aligned}\tag{31}$$

since $a_{N,T}\to 0$ and $l=O(BT)$. Now note that

$$\frac{\Delta_{N,T}a_{N,T}}{\Delta_{l,T}^2}\to 0\text{ and }\frac{N^{1/k}a_{N,T}^{1/2}}{h_{N,l,T}^{dr}}\to 0\Rightarrow\frac{a_{N,T}^{3/2}N^{1/k}\Delta_{N,T}}{h_{N,l,T}^{dr}\Delta_{l,T}^2}\to 0.$$

As for $A_{N,l,T}$,

$$\begin{aligned}
A_{N,l,T}\ =\ &\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\\
&\times\frac{\Delta_{l,T}^{-1}\left(\left(X_{(jB+b)\Delta_{N,T}}-X_{((j-1)B+b)\Delta_{N,T}}\right)-\mu\left(X_{((j-1)B+b)\Delta_{N,T}}\right)\Delta_{l,T}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}\\
&+\frac{\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\left(\mu\left(X_{((j-1)B+b)\Delta_{N,T}}\right)-\mu(x)\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}.
\end{aligned}\tag{32}$$

By the same argument used in the proof of Theorem 3 in Bandi and Phillips (2003) the second term on the right-hand side of Eq. (32) is $O_p\left(\sqrt{h_{N,l,T}^{dr}\overline{L}_X(T,x)}h_{N,l,T}^{dr,2}\right)$ and, of course, $o_p(1)$ if $h_{N,l,T}^{dr,5}\overline{L}_X(T,x) \overset{a.s.}{\to} 0$. As for the first term on the right-hand side of Eq. (32), write

$$\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)$$
$$\times\frac{\Delta_{l,T}^{-1}\int_{((j-1)B+b)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\left(\mu\left(X_s\right)-\mu\left(X_{((j-1)B+b)\Delta_{N,T}}\right)\right)\mathrm{d}s}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}$$
$$+\frac{\frac{\Delta_{N,T}}{\sqrt{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\int_{((j-1)B+b)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\sigma\left(X_s\right)\mathrm{d}W_s}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}. \tag{33}$$

The first term in Eq. (33) is $O_p\left(\Delta_{l,T}^{1/2}\log^{1/2}\left(1/\Delta_{l,T}\right)\sqrt{h_{N,l,T}^{dr}\overline{L}_X(T,x)}\right)$. Define the second term on the right-hand side of Eq. (33) as $A_{N,T,l}(x)$ and express its quadratic variation as

$$\begin{aligned}
&\langle A_{N,T,l}(x)\rangle\\
=\;&\frac{\Delta_{N,T}^2}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{i=1}^{B}\sum_{j=1}^{l-1}\frac{K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)K\left(\frac{X_{((j-1)B+i)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}{\left(\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\right)^2}\\
&\times\Delta_{l,T}^{-2}\left\langle\int_{((j-1)B+b/T)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\sigma\left(X_s\right)\mathrm{d}W_s,\int_{((j-1)B+i/T)\Delta_{N,T}}^{(jB+i)\Delta_{N,T}}\sigma\left(X_s\right)\mathrm{d}W_s\right\rangle\\
=\;&\frac{\Delta_{N,T}^2}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}2\sum_{b=1}^{B}\sum_{i>b}^{B}\sum_{j=1}^{l-1}\frac{K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)K\left(\frac{X_{((j-1)B+i)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}{\left(\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\right)^2}\\
&\times\Delta_{l,T}^{-2}\left\langle\int_{((j-1)B+i)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\sigma\left(X_s\right)\mathrm{d}W_s\right\rangle+o_p(1)\\
=\;&\frac{\Delta_{N,T}^2}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}2\sum_{b=1}^{B}\sum_{i>b}^{B}\sum_{j=1}^{l-1}\frac{K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)K\left(\frac{X_{((j-1)B+i)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}{\left(\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\right)^2}\\
&\times\Delta_{l,T}^{-1}\left(1-\frac{i-b}{B}\right)\sigma^2(x)\\
=\;&\frac{2}{3}\sigma^2(x)\int K^2(s)ds+o_p(1).
\end{aligned}\tag{34}$$

Finally, the limiting distribution in the statement derives from a similar argument as that in the proof of Theorem 3 in Corradi and Distaso (2008).

We now turn to the diffusion function estimator in (ii). Write the estimation error decomposition as:

$$
\begin{aligned}
& \sqrt{\frac{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}{\Delta_{l,T}}}\left(\widehat{\sigma}^2_{N,l,T}(x)-\sigma^2(x)\right) \\
= & \sqrt{\frac{\Delta_{N,T}B^{-1}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}\left(K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)-K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)\right) \\
& \times\frac{\Delta^{-1}_{l,T}\left(\left(Y_{(jB+b)\Delta_{N,T}}-Y_{((j-1)B+b)\Delta_{N,T}}\right)^2-RV_{T,N}\Delta_{N,T}\right)}{\frac{\Delta_{N,T}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)} \\
& +\frac{\sqrt{\frac{\Delta_{N,T}B^{-1}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)\Delta^{-1}_{l,T}\left(\left(Y_{(jB+b)\Delta_{N,T}}-Y_{((j-1)B+b)\Delta_{N,T}}\right)^2-RV_{T,N}\Delta_{N,T}\right)}{\frac{\Delta_{N,T}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)} \\
& -\sqrt{\frac{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}{\Delta_{l,T}}}\sigma^2(x) \\
= & C_{N,T,l}+D_{N,T,l}.
\end{aligned}
\tag{35}
$$

Expressing the kernel function as in part (i):

$$
\begin{aligned}
C_{N,T,l} \simeq_p & \sqrt{\frac{\Delta_{N,T}B^{-1}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,a)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}1\left\{x+h^{dr}_{N,l,T}\le X_{((j-1)B+b)\Delta_{N,T}}\le x+h^{dr}_{N,l,T}+N^{1/k}a^{1/2}_{N,T}\right\} \\
& \times\frac{\Delta^{-1}_{l,T}\left(\left(\left(X_{(jB+b)\Delta_{N,T}}-X_{((j-1)B+b)\Delta_{N,T}}\right)+\left(\epsilon_{(jB+b)\Delta_{N,T}}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)\right)^2-RV_{T,N}\Delta_{N,T}\right)}{\frac{\Delta_{N,T}}{h^{dif}_{N,l,T}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{Y_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dif}_{N,l,T}}\right)}.
\end{aligned}
$$

Now $\mathrm{E}(C_{N,T,l})=O\left(\frac{N^{1/k}a^{1/2}_{N,T}l^{1/2}\sqrt{\frac{h^{dif}_{N,l,T}\overline{L}_X(T,x)}{T}}}{h^{dif}_{N,l,T}}\right)$ and

$$
\mathrm{E}\left(C^2_{N,T,l}\right)=O\left(\frac{a^{1/2}_{N,T}N^{1/k}}{h^{dif}_{N,l,T}}+\frac{a^{3/2}_{N,T}N^{1/k}}{Bh^{dif}_{N,l,T}\Delta_{l,T}}+\frac{a^{5/2}_{N,T}N^{1/k}l}{h^{dif}_{N,l,T}\Delta^2_{l,T}}\right),
$$

where, again, the order of the first term is analogous to that of Eq. (38), the order of the third term is analogous to that of Eq. (36), and the order of the cross-product term is analogous to that of Eq. (37) below (in all cases with the indicator kernel in place of a smooth kernel). Note that

$$
\begin{aligned}
& \frac{N^{1/k}a^{1/2}_{N,T}l^{1/2}\sqrt{\frac{h^{dif}_{N,l,T}\overline{L}_X(T,x)}{T}}}{h^{dif}_{N,l,T}}\overset{a.s.}{\to}0 \text{ and } \sqrt{\frac{h^{dif}_{N,l,T}\overline{L}_X(T,x)}{\Delta_{l,T}}}\overset{a.s.}{\to}\infty \\
\Rightarrow\quad & \frac{a^{1/2}_{N,T}N^{1/k}}{h^{dif}_{N,l,T}}\to 0.
\end{aligned}
$$

Similarly $\frac{a_{N,T}^{3/2}N^{1/k}}{Bh_{N,l,T}^{dif}\Delta_{l,T}} \to 0$ given that $l = O(BT)$ and $\frac{a_{N,T}^{1/2}N^{1/k}}{h_{N,l,T}^{dif}} \to 0$. Now write

$$
\begin{aligned}
&\sqrt{\frac{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}{\Delta_{l,T}}}\left(\widehat{\sigma}_{N,l,T}^2(x)-\sigma^2(x)\right) = D_{N,T,l}+o_p(1)\\
=\;& \left(\frac{\sqrt{\frac{\Delta_{N,T}B^{-1}}{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\left(X_{(jB+b)\Delta_{N,T}}-X_{((j-1)B+b)\Delta_{N,T}}\right)^2}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}\right.\\
&\left.-\sqrt{\frac{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}{\Delta_{l,T}}}\sigma^2(x)\right)\\
&+\frac{\sqrt{\frac{\Delta_{N,T}B^{-1}}{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\Delta_{l,T}^{-1}\left(\left(\epsilon_{(jB+b)\Delta_{N,T}}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)^2-RV_{T,N}\Delta_{N,T}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)} \qquad (36)\\
&-2\sqrt{\frac{\Delta_{N,T}B^{-1}}{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\\
&\times\frac{\Delta_{l,T}^{-1}\left(X_{(jB+b)\Delta_{N,T}}-X_{((j-1)B+b)\Delta_{N,T}}\right)\left(\epsilon_{(jB+b)\Delta_{N,T}}-\epsilon_{((j-1)B+b)\Delta_{N,T}}\right)}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}+o_p(1) \qquad (37)\\
=\;& I_{N,l,T}+II_{N,l,T}+III_{N,l,T}+o_p(1).
\end{aligned}
$$

Now notice that $III_{N,l,T}$ is $O_p\left(\frac{a_{N,T}}{B\Delta_{l,T}}\right)$ and $II_{N,l,T}$ is $O\left(\frac{a_{N,T}^2 l}{\Delta_{l,T}^2}\right)$. If $\frac{a_{N,T}^2 l}{\Delta_{l,T}^2} \to 0$, then $\frac{a_{N,T}^{5/2}N^{1/k}l}{h_{N,l,T}^{dif}\Delta_{l,T}^2} \to 0$ since $\frac{a_{N,T}^{1/2}N^{1/k}}{h_{N,l,T}^{dif}} \to 0$. Finally,

$$
\begin{aligned}
I_{N,l,T} \;=\;& \sqrt{\frac{\Delta_{N,T}B^{-1}}{h_{N,l,T}^{dif}\widehat{\overline{L}}_X(T,x)}}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)\\
&\times\frac{\Delta_{l,T}^{-1}2\int_{((j-1)B+b)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\left(X_s-X_{((j-1)B+b)\Delta_{N,T}}\right)\sigma(X_s)\mathrm{d}W_s}{\frac{\Delta_{N,T}}{h_{N,l,T}^{dr}\widehat{\overline{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h_{N,l,T}^{dr}}\right)}+o_p(1).
\end{aligned}
$$

Noting that $\lim_{B\to\infty} B^{-2}\sum_{b=1}^{B}\sum_{i>b}^{B}\left(1-\frac{i-b}{B}\right)^2 = 1/4$, we obtain

$$
\begin{aligned}
& \langle I_{N,l,T}\rangle \\
= \; & \frac{\Delta_{N,T}}{Bh^{dif}_{N,l,T}\widehat{\widehat{L}}_X(T,x)} 4\sum_{b=1}^{B}\sum_{i>b}^{B}\sum_{j=1}^{l-1} \frac{K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)K\left(\frac{X_{((j-1)B+i)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)}{\left(\frac{\Delta_{N,T}}{h^{dr}_{N,l,T}\widehat{\widehat{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)\right)^2} \\
& \times\Delta^{-2}_{l,T}\left\langle\int_{((j-1)B+i)\Delta_{N,T}}^{(jB+b)\Delta_{N,T}}\left(\left(X_s - X_{((j-1)B+b)\Delta_{N,T}}\right)\right)\sigma\left(X_s\right)\mathrm{d}W_s\right\rangle + o_p(1) \\
= \; & \frac{\Delta_{N,T}}{Bh^{dif}_{N,l,T}\widehat{\widehat{L}}_X(T,x)} 4\sum_{b=1}^{B}\sum_{i>b}^{B}\sum_{j=1}^{l-1} \frac{K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)K\left(\frac{X_{((j-1)B+i)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)}{\left(\frac{\Delta_{N,T}}{h^{dr}_{N,l,T}\widehat{\widehat{L}}_X(T,x)}\sum_{b=1}^{B}\sum_{j=1}^{l-1}K\left(\frac{X_{((j-1)B+b)\Delta_{N,T}}-x}{h^{dr}_{N,l,T}}\right)\right)^2} \\
& \times\sigma^4\left(x\right)\left(1-\frac{i-b}{B}\right)^2 \\
= \; & \sigma^4\left(x\right)\int K^2(s)ds + o_p(1). \qquad (38)
\end{aligned}
$$

The stated result now follows. ■

10 Appendix B

Let $\Delta_{N,T} = T/N$ and $\Delta_{M,T} = T/M$, with $M < N$, be the discrete intervals used in the estimation of spot volatility (by virtue of high-frequency data) and in the estimation of the volatility drift and diffusion (by virtue of low-frequency data), respectively. Bandi and Renò (2008) have established rate conditions under which the estimation error introduced by the preliminary spot variance estimates is asymptotically negligible when estimating the variance drift and diffusion. More precisely, they present four additional conditions, two for the drift and two for the variance. The first drift condition reads:

$$\frac{TL_v^{-1/2}(T,a)}{\Delta_{M,T}h^{dr,1/2}_{M,T}T^{-\beta\delta_N}\Delta^{-\beta}_{N,T}} \overset{a.s.}{\to} 0,$$

where $\beta < \frac{1}{2}$ in the case of spot variance estimators robust to market microstructure noise and $\beta = \frac{1}{2}$ in the case of realized variance. This requires

$$M < N^{\beta}h^{dr,1/2}_{M,T}T^{-\beta(1+\delta_N)}L_v^{1/2}(T,a). \qquad (39)$$

The second drift condition reads:

$$\frac{TL_v^{-1/2}}{\Delta_{M,T}h^{dr,1/2}_{M,T}}T^{-\delta_N/2}\log\left(T^{\delta_N/2}\right)\to 0,$$

which requires

$$M < L_v^{1/2}(T,a)T^{\delta_N/2}h^{dr,1/2}_{M,T}\log(T^{-\delta_N/2}). \qquad (40)$$

By equating the right-hand sides of the inequalities in Eq. (39) and Eq. (40), as earlier, we set δ_N in such a way as to guarantee that

$$T^{\left(1+\delta_N+\frac{\delta_N}{2\beta}\right)}(\log(T^{-\delta_N/2}))^{1/\beta} = N. \qquad (41)$$

Ignoring now $\log(T^{-\delta_N/2})$, and plugging (41) into (39), one may write

$$
\begin{aligned}
M \; & < \; N^{\beta}h^{dr,1/2}_{M,T}N^{-\beta}T^{\delta_N/2}L_T^{1/2} \\
& \approx \; N^{\left(\frac{\beta}{2\beta+1}\right)}h^{dr,1/2}_{M,T}T^{-\left(\frac{\beta}{2\beta+1}\right)}L_T^{1/2}
\end{aligned}
$$

which is indeed condition (21) in Section 5. We now turn to asymptotic normality of the spot volatility's diffusion. The first condition, reads

$$\frac{TL_v^{-1/2}(T,a)}{\Delta_{M,T}^{3/2} h_{M,T}^{dif,1/2} T^{-\beta\delta_N} \Delta_{N,T}^{-\beta}} \stackrel{a.s.}{\to} 0$$

which requires

$$M < N^{\frac{2}{3}\beta} h_{M,T}^{dif1/3} T^{\frac{1}{3}-(\frac{2}{3}\beta+\frac{2}{3}\beta\delta_N)} L_v^{1/3}(T,a). \tag{42}$$

The second condition reads

$$\frac{TL_v^{-1/2}(T,a)}{\Delta_{M,T}^{3/2} h_{M,T}^{dif,1/2}} T^{-\delta_N/2} \log\left(T^{\delta_N/2}\right) \stackrel{a.s.}{\to} 0$$

which requires

$$M < L_v^{1/3}(T,a) T^{\frac{1}{3}(\delta_N+1)} h_{M,T}^{dif,1/3} \log(T^{-\delta_N/3}). \tag{43}$$

By equating the right-hand sides of Eq. (42) and Eq. (43), we can set δ_N in such a way that

$$T^{\frac{\delta_N+2\beta+2\beta\delta_N}{2\beta}} \log(T^{-\delta_N/3})^{\frac{3}{2\beta}} = N. \tag{44}$$

Thus, plugging Eq. (44) into Eq. (42), and neglecting the logarithm, we obtain:

$$\begin{aligned} M &< L_v^{1/3}(T,a) N^{1/3} T^{-\frac{\delta_N}{6\beta}} h_{M,T}^{dif,1/3} \log(T^{-\delta_N/3}) \\ &\approx L_v^{1/3}(T,a) N^{1/3} N^{-\left(\frac{2\beta}{1+2\beta}\right)\frac{1}{6\beta}} T^{\left(\frac{2\beta}{1+2\beta}\right)\frac{1}{6\beta}} h_{M,T}^{dif,1/3}. \end{aligned}$$

which amounts to condition (22) in Section 5.

References

[1] AÏT-SAHALIA, Y., P.A. MYKLAND and L. ZHANG (2009). Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. *Journal of Econometrics,* forthcoming.

[2] ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and P. LABYS (2000). Great Realizations. *Risk*, March, 105-108.

[3] ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and P. LABYS (2003). Modeling and Forecasting Realized Volatility. *Econometrica* 71, 579-625.

[4] AWARTANI, B., V. CORRADI, and W. DISTASO (2009). Assessing Market Microstructure Effects with an Application to the Dow Jones Industrial Average Stocks. *Journal of Business and Economic Statistics* 27, 251-265.

[5] BAI, J. (1994). Weak Convergence of Sequential Empirical Processes of Residuals in ARMA Models. *Annals of Statistics* 22, 2051-2061.

[6] BANDI, F.M., and G. MOLOCHE (2004). On the Functional Estimation of Multivariate Diffusion Processes. *Working paper.*

[7] BANDI, F.M., and T. NGUYEN (2003). On the Functional Estimation of Jump-Diffusion Models. *Journal of Econometrics* 116, 293-328.

[8] BANDI, F.M., and P.C.B. PHILLIPS (2003). Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* 71, 241-283.

[9] BANDI, F.M., and P.C.B. PHILLIPS (2007). A Simple Approach to the Parametric Estimation of Potentially Nonstationary Diffusions. *Journal of Econometrics* 137, 354-395.

[10] BANDI, F. M., and R. RENÒ (2008). Nonparametric Stochastic Volatility. SSRN: http://ssrn.com/abstract=1158438.

[11] BARNDORFF-NIELSEN, O.E., A. LUNDE, P.R. HANSEN, and N. SHEPHARD (2008a). Multivariate Realised Kernels: Consistent Positive Seni-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. *Working Paper.*

[12] BARNDORFF-NIELSEN, O.E., P.R. HANSEN, A. LUNDE and N. SHEPHARD (2008b). Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica* 76, 1481-1536.

[13] BRUGIÈRE, P. (1991). Estimation de la Variance d'un Processus de Diffusion dans le Cas Multidimensionnel. *C. R. Acad. Sci.,* T. 312, 1, 999-1005.

[14] CHAN, K.C., KAROLYI G.A., LONGSTAFF F.A., SANDERS A.B., (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209-1227.

[15] CORRADI, V., and W. DISTASO (2008). Diagnostic Tests for Volatility Models. *Working paper.*

[16] CORRADI, V., and N. SWANSON (2006). The Effects of Data Transformation on Common Cycle, Cointegration, and Unit Root Tests: Monte Carlo and a Simple Test. *Journal of Econometrics* 132, 195-229.

[17] CORRADI, V., and H. WHITE (1999). Specification Tests for the Variance of a Diffusion. *Journal of Time Series Analysis* 20, 253-270.

[18] FAN, J., and C. ZHANG (2003). A Re-Examination of Diffusion Estimators with Applications to Financial Model Validation. *Journal of the American Statistical Association* 98, 118-134.

[19] FLORENS-ZMIROU, D. (1993). On Estimating the Diffusion Coefficient from Discrete Observations. *Journal of Applied Probability* 30, 790-804.

[20] ERAKER, B., M. JOHANNES, and N. POLSON (2003). The Impact of Jumps in Volatility and Returns, *Journal of Finance* 58, 1269-1300.

[21] GUERRE, E. (2004). Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series. *Working Paper.*

[22] JAQUIER, E., N. POLSON, and P.E. Rossi (1994). Bayesian Analysis of Stochastic Volatility Models, *Journal of Business Economics and Statistics* 12, 371-389.

[23] JACOD, J. (1997). Nonparametric Kernel Estimation of the Diffusion Coefficient of a Diffusion. *Prépublication No. 405. du Laboratoire de Probabilités de l'Université Paris VI.*

[24] JOHANNES, M. (2004). The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *Journal of Finance* 59, 227-260.

[25] KANAYA, S., and D. KRISTENSEN (2008). Estimation of Stochastic Volatility Models by Nonparametric Filtering. *Working Paper.*

[26] KARLSEN, H.A. and V. TJOSTHEIM (2001). Nonparametric Estimation in Null Recurrent Time Series. *Annals of Statistics* 29, 372-416.

[27] KRISTENSEN, D. (2008). Nonparametric Filtering of Realized Spot Volatility: a Kernel-based Approach. *Econometric Theory,* forthcoming.

[28] LEE, S. and C.Z. WEI. (1999). On Residual Empirical Processes of Stochastic Regression Models with Application to Time Series. *Annals of Statistics* 27, 237-261.

[29] MOLOCHE, G. (2004). Local Nonparametric Estimation of Scalar Diffusions. *Working paper.*

[30] STANTON, R. (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance* 52, 1973-2002.

[31] ZHANG, L., P.A. MYKLAND and Y. AIT-SAHALIA (2005). A Tale of Two Time Scales: Determining integrated volatility with Noisy High Frequency Data. *Journal of the American Statistical Association* 100, 1394-1411.

	drift		diffusion	
bw (1st stage)	0.0746	(0.0452)	0.0053	(0.0013)
bw (2nd stage)	0.7130	(0.7956)	0.1858	(0.3592)

Table 1: Model 1: The table shows the residual-based average bandwidths (bw) and their standard deviations (in parentheses).

	drift		diffusion	
bw (1st stage)	0.6950	(0.3420)	0.0935	(0.0733)
bw (2nd stage)	0.7896	(0.9028)	0.1746	(0.1062)

Table 2: Model 1: The table shows the cross-validated average bandwidths (bw) and their standard deviations (in parentheses).

	drift		diffusion	
bw (1st stage)	0.6320	(0.5373)	0.1200	(0.1061)
bw (2nd stage)	0.5900	(0.4775)	0.1703	(0.5036)

Table 3: Model 2: The table shows the residual-based average bandwidths (bw) and their standard deviations (in parentheses).

	drift		diffusion	
bw (1st stage)	1.2867	(0.7191)	0.7339	(0.5375)
bw (2nd stage)	0.6148	(0.5815)	0.1597	(0.4946)

Table 4: Model 2: The table shows the cross-validated average bandwidths (bw) and their standard deviations (in parentheses).

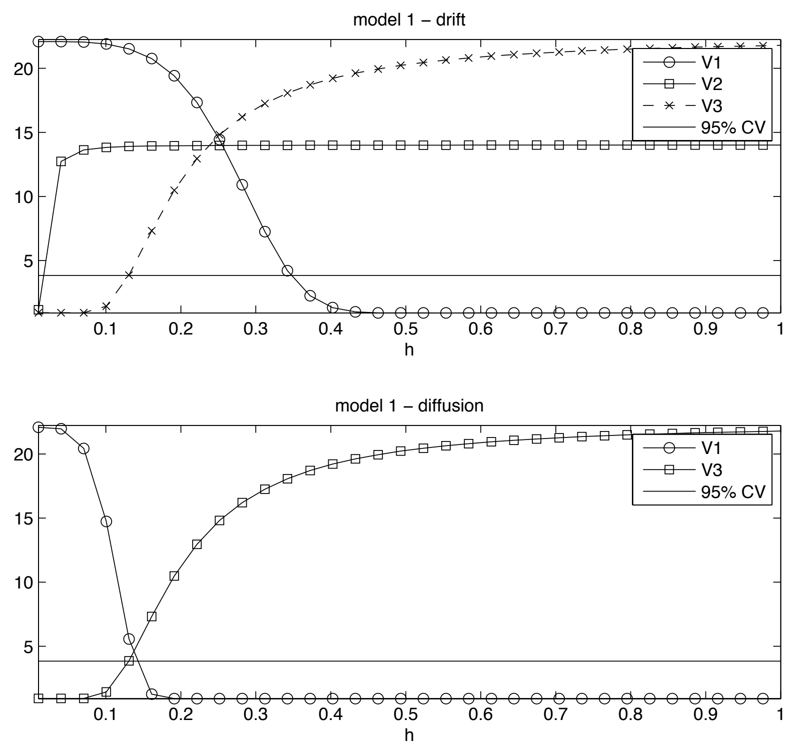

Figure 1: Model 1: The V statistics as a function of the bandwidth.

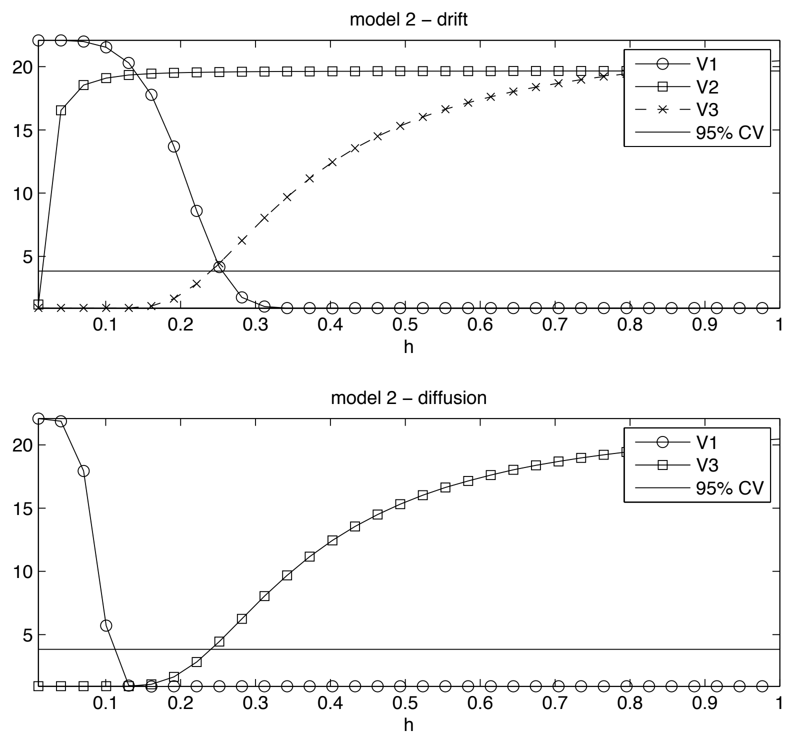

Figure 2: Model 2: The V statistics as a function of the bandwidth.

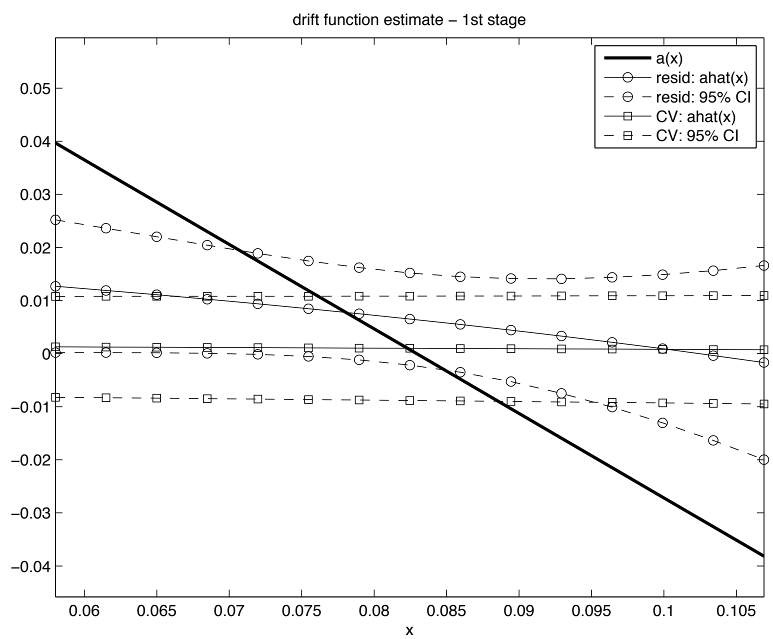

Figure 3: Model 1: drift, 1st stage

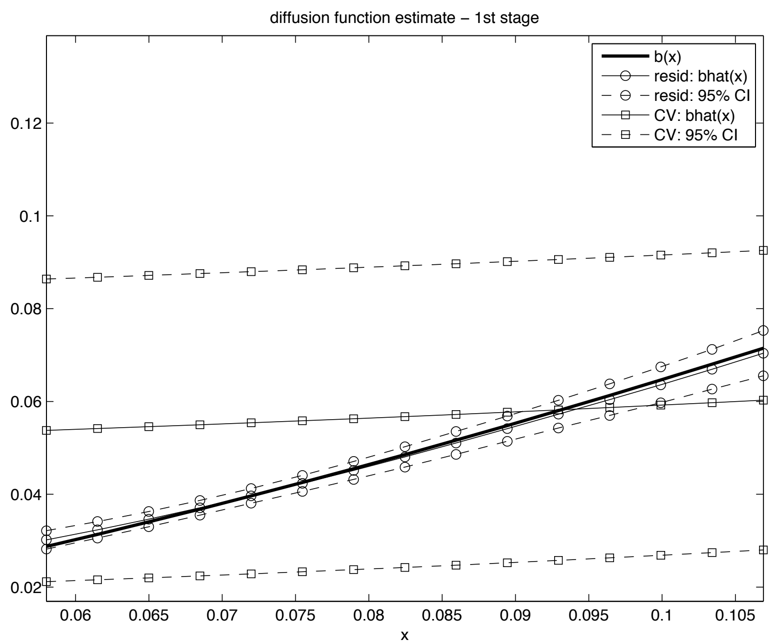

Figure 4: Model 1: diffusion, 1st stage

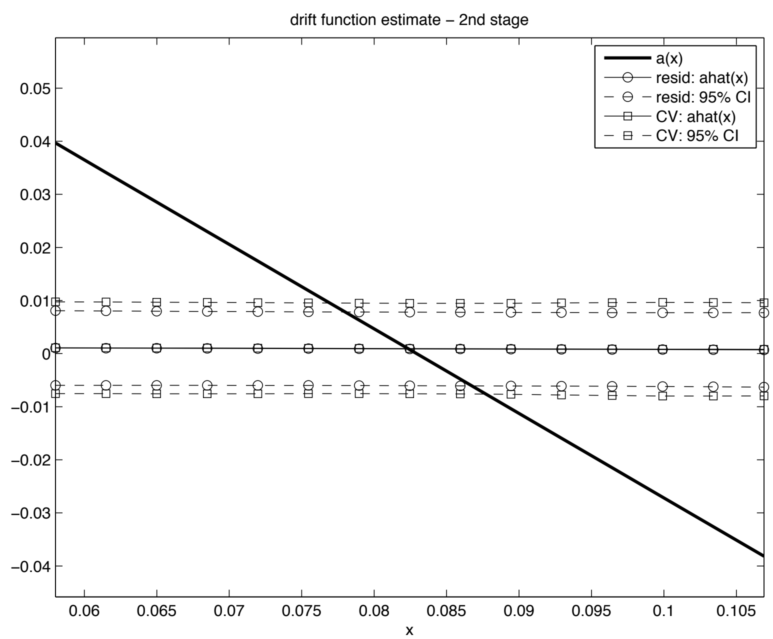

Figure 5: Model 1: drift, 2nd stage

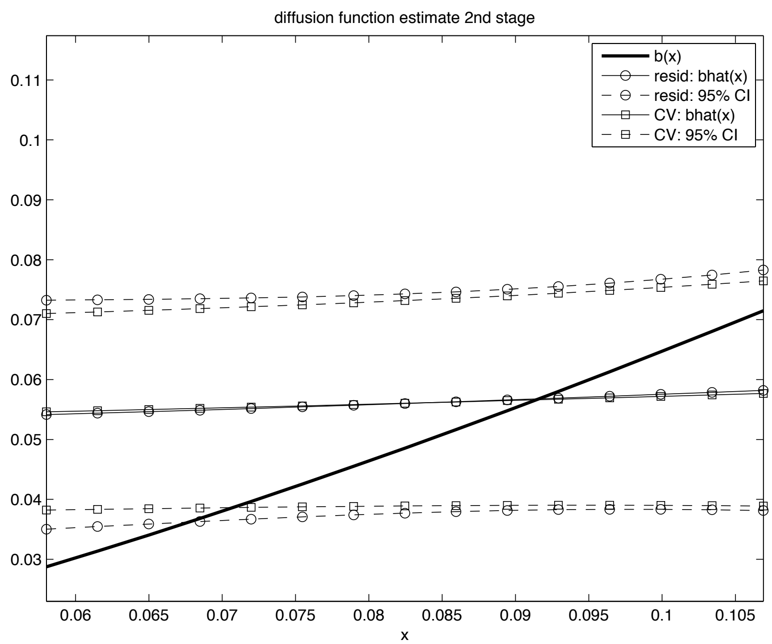

Figure 6: Model 1: diffusion, 2nd stage

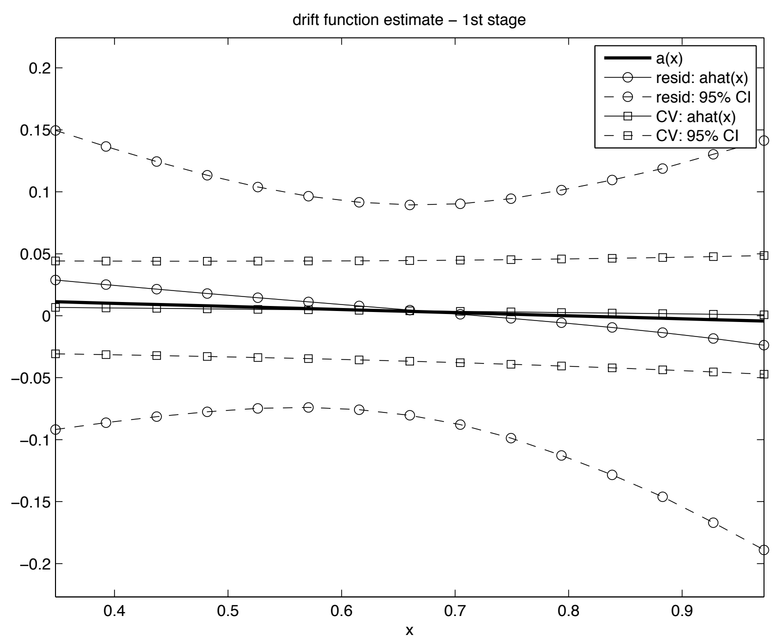

Figure 7: Model 2: drift, 1st stage

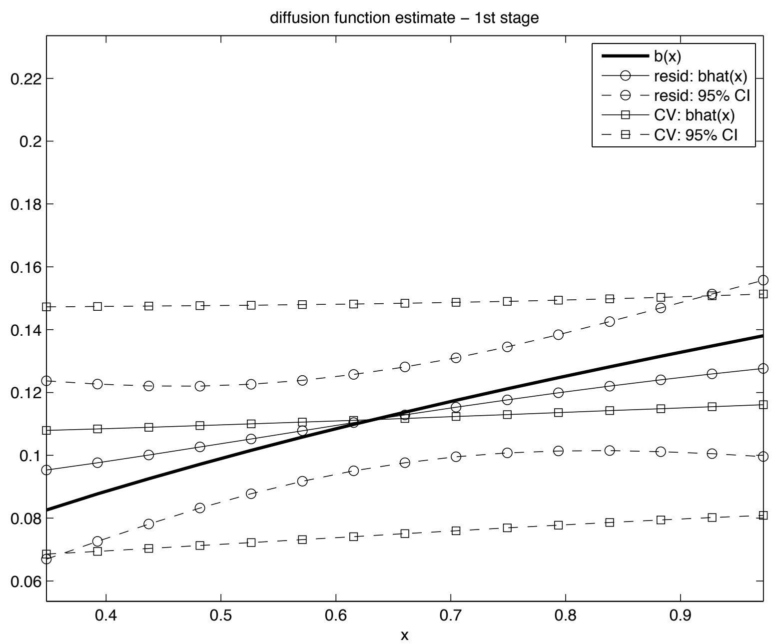

Figure 8: Model 2: diffusion, 1st stage

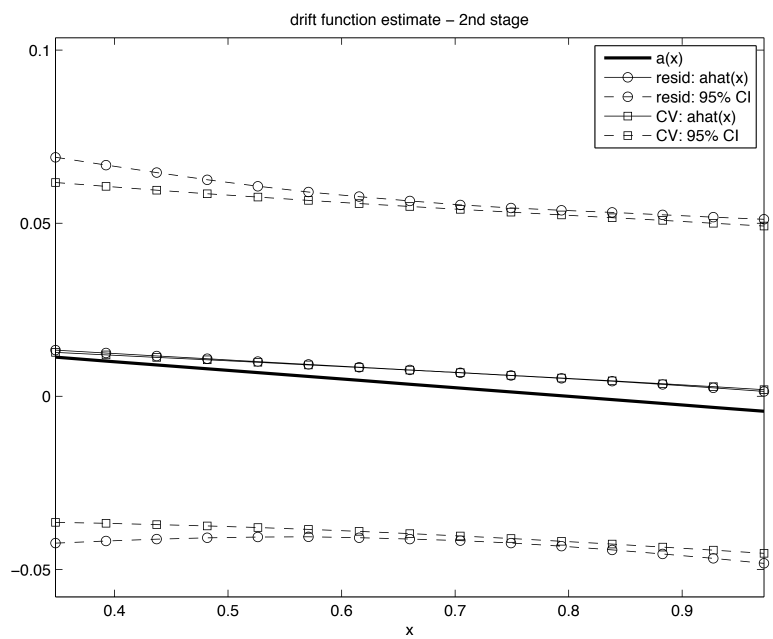

Figure 9: Model 2: drift, 2nd stage

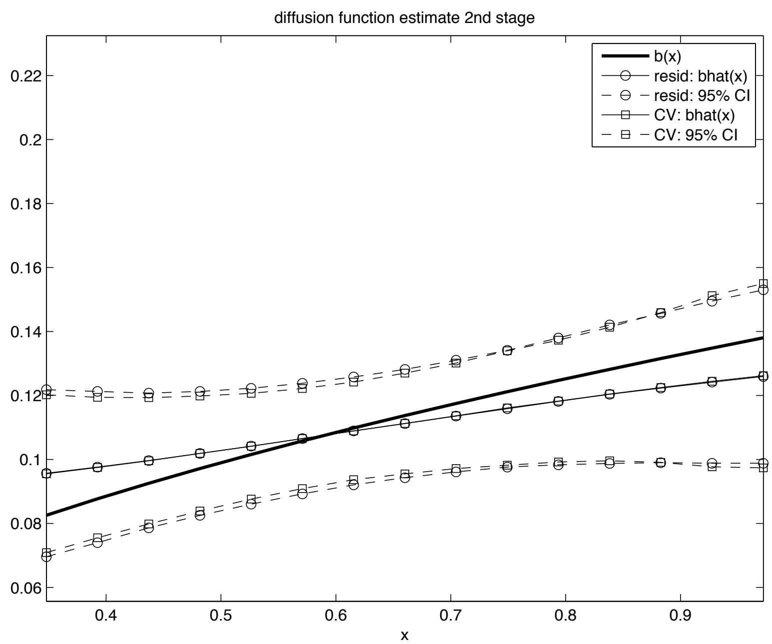

Figure 10: Model 2: diffusion, 2nd stage