Econometric Theory, **26**, 2010, 541–563.
doi:10.1017/S0266466609100087

REWEIGHTED FUNCTIONAL ESTIMATION OF DIFFUSION MODELS

KE-LI XU
University of Alberta

The local linear method is popular in estimating nonparametric continuous-time diffusion models, but it may produce negative results for the diffusion (or volatility) functions and thus lead to insensible inference. We demonstrate this using U.S. interest rate data. We propose a new functional estimation method of the diffusion coefficient based on reweighting the conventional Nadaraya–Watson estimator. It preserves the appealing bias properties of the local linear estimator and is guaranteed to be nonnegative in finite samples. A limit theory is developed under mild requirements (recurrence) of the data generating mechanism without assuming stationarity or ergodicity.

1. INTRODUCTION

This paper is concerned with nonparametric estimation of the continuous-time Itô diffusion model:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t, \tag{1}$$

based on a discrete sample $\{X_\Delta, \ldots, X_{n\Delta}\}$ over the time span $[0, T]$, where $\Delta = T/n$ and $\{B_t, 0 \le t \le T\}$ is a standard Brownian motion. Here the nonrandom functions $\mu(\cdot)$ and $\sigma^2(\cdot)$, which are called infinitesimal mean function (or drift function) and infinitesimal variance function (or diffusion function, volatility function), respectively, fully characterize the transition dynamics of the time series X_t if an initial point is given and thus are of main interest to us. We assume time homogeneity of the diffusion process in the sense that two infinitesimal functions depend only on the current state, not on time. Diffusion processes have been widely used to describe the stochastic dynamics of the underlying economic and financial variables, e.g., asset returns, interest rates, exchange rates, etc.

The author has benefited from comments by the co-editor Oliver B. Linton, two anonymous referees, Donald W.K. Andrews, Zongwu Cai, Xiaohong Chen, Yuichi Kitamura, Taisuke Otsu, Peter C.B. Phillips, and other participants at the econometrics seminar at Yale University and the NBER-NSF time series conference at Montreal. Special thanks go to Peter C.B. Phillips for his guidance and encouragement. The author also gratefully acknowledges the partial financial support provided by the University of Alberta School of Business through the H.E. Pearson faculty award. Address correspondence to Ke-Li Xu, Department of Finance and Management Science, University of Alberta School of Business, Edmonton, Alberta, T6G 2R6, Canada; e-mail: keli.xu@ualberta.ca.

© Cambridge University Press, 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

There has been a large literature devoted to discovering the correct parametric forms of diffusion models that describe the evolution of various economic variables. For instance, Aït-Sahalia (1996a, Table 1) listed the existing specifications for short-term interest rate processes (see also Ahn and Gao, 1999). However, as Fan (2005) noted, "many of the [parametric] stochastic models in use are simple and convenient ones to facilitate mathematical derivations and statistical inferences"; economic or financial theories offer little guidance to choose one model instead of others. In fact, serious doubts have been cast on most of the existing short rate models. Aït-Sahalia (1996b) firmly rejected all one-factor parametric diffusion models of the spot rates previously proposed by comparing the marginal density of each parametric model to the model-free density function that is estimated nonparametrically (see also Pritsker, 1998). More recently, Hong and Li (2005) modified Aït-Sahalia's method and proposed a specification test based on the transition density, taking into account the possible serial dependence of the short rate processes. Using the same data set as in Aït-Sahalia (1996b), they still found strong evidence against most of the existing parametric models. Fan and Zhang (2003) studied a generalized likelihood ratio specification test of Fan, Zhang, and Zhang (2001), and they strongly rejected all specifications of the short rate diffusion functions considered before.

Accordingly, a nonparametric approach robust to misspecification errors provides a flexible alternative. It has been popularized recently in estimating continuous-time models in finance by the availability of large data sets.[1] Jiang and Knight (1997) proposed a two-step procedure in which the diffusion function is estimated using a kernel estimator similar to that of Florens-Zmirou (1993) and then identifying the drift function via the stationary density function (see also Arapis and Gao, 2006). Stanton (1997) estimated the drift and diffusion functions simultaneously based on the approximate conditional moment representations of the two infinitesimal moment functions. Stanton assumed stationarity, but unlike Jiang and Knight's, his method could be used even when the stationary density

TABLE 1. Estimated diffusion function based on local linear (LL) and reweighted Nadaraya–Watson (RNW) approaches, calculated using monthly data, 1 January 1934 to 1 March 2006. The bandwidth is 3.27%. Units of design points: percentage.

x	1.0	1.2	1.4	1.5	1.6	1.7
LL	−0.0032	−0.0025	−0.0017	−0.0013	−0.00084	−0.00089
RNW	0.0013	0.0015	0.0019	0.0021	0.0022	0.0024
x	1.8	1.9	2.0	2.5	3.0	4.0
LL	0.000025	0.00043	0.00086	0.0031	0.0054	0.0100
RNW	0.0027	0.0029	0.0032	0.0043	0.0057	0.0095

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

does not exist. Bandi and Phillips (2003) confirmed this and proposed a general family of estimators nesting Stanton's as special cases. They also established the asymptotic theory without assuming stationarity of the underlying process. All estimation methods mentioned earlier are based on local constant fitting and can be improved by using local linear fitting for both drift and diffusion functions, as studied by Moloche (2001) under recurrence and Fan and Zhang (2003) under strict stationarity. Local linear estimation possesses simple bias representation and corrects the boundary bias automatically, and it has become the common choice among practitioners (e.g., Durham, 2003).

It has been recognized among empirical researchers that local linear fitting (or more generally, local polynomial fitting) may assign negative weights to certain sample points, and the resulting estimator of a positive quantity (e.g., variance, density, distribution) may produce a negative result in finite samples, thus leading to insensible inferences (e.g., Hall and Tajvidi, 2000; Chen and Qin, 2002; Yu and Jones, 2004). This is also unfortunately true for the local linear estimator of the diffusion function $\sigma^2(\cdot)$ in (1) because $\sigma^2(\cdot)$ reflects the infinitesimal variation of the process around a spatial point. One might suggest truncating negative estimates as appropriate, but it does not seem credible that these volatilities would be exactly zero. In this sense, the usual local constant estimator works better because it respects the range of parameter and always gives nonnegative weights on observations, although it has inferior bias properties and boundary behavior.

In this paper we address this issue and propose a novel functional diffusion estimator, which incorporates both the advantages of the usual local constant and those of the local linear estimators. The idea is to minimally adjust the local constant estimator (or the Nadaraya–Watson estimator) subject to a discrete bias-reducing moment condition satisfied by the local linear estimator. The new estimator, which is called the *reweighted Nadaraya–Watson (RNW) estimator,* is asymptotically equivalent to the local linear estimator and inherits the nonnegativity restriction of diffusion function from the local constant estimator.

Within the discrete-time stationary time series framework, Ziegelmann (2002) and Yu and Jones (2004) proposed positive nonparametric conditional variance estimators. But their methods rely on a logarithmic transformation of the variance function, which induces an additional bias term and is therefore not asymptotically equivalent to the local linear estimator. The resulting estimators do not have closed-form representation, and the limit theory is not easy to extend to the continuous-time context.[2]

The diffusion functional estimator presented here builds on the work of Hall and Presnell (1999) and Cai (2001), who developed the reweighted estimator for the mean function in the nonparametric regression model as an alternative to the local linear estimator. This paper extends their approach to the continuous-time diffusion model to estimate the diffusion coefficient. It is straightforward to extend the methodology and theory developed here to estimate the drift coefficient. Although the estimation method proposed here is similar in concept to that of Hall and Presnell (1999) and Cai (2001), there are several important differences. First,

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

the model here is defined in continuous time, and the limit theory relies on increasing both sample sizes and sampling frequencies. Second, we do not assume the existence of the stationary density function; thus nonstationary behavior in the underlying process is allowed. We use the weaker recurrence assumption of Bandi and Phillips (2003) and characterize the convergence rate of the estimator in terms of the local time spent within the neighborhood of the spatial state under consideration. Third, we focus on the variance function of the time series rather than the mean function.

We conclude this section by considering an empirical example of the short-term interest rate dynamics to illustrate the potential problem of local polynomial estimation for the diffusion function and the effectiveness of the proposed method. We obtained monthly values of secondary market yields on three-month Treasury Bills between 1 January 1934 and 1 March 2006 ($n = 867$). The annualized three-month yields are plotted in Figure 1, and Figure 2 shows a scatter plot of monthly squared changes $(X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta$ in the short rate against the current month's rate $X_{i\Delta}$, where $\Delta = 1/12$. Nonparametric estimation of the diffusion function $\sigma^2(\cdot)$ involves fitting the monthly squared changes (vertical axis in Figure 2) locally with respect to the current short rates (horizontal axis); see Section 2. To understand that the local linear and local polynomial estimators may produce negative results for the diffusion function, we first take an extreme case in which the smoothing bandwidth is selected sufficiently large so that all observations lie in the smoothing window of each particular spatial point. With this bandwidth the local linear (local constant, local quadratic) regression with a uniform kernel

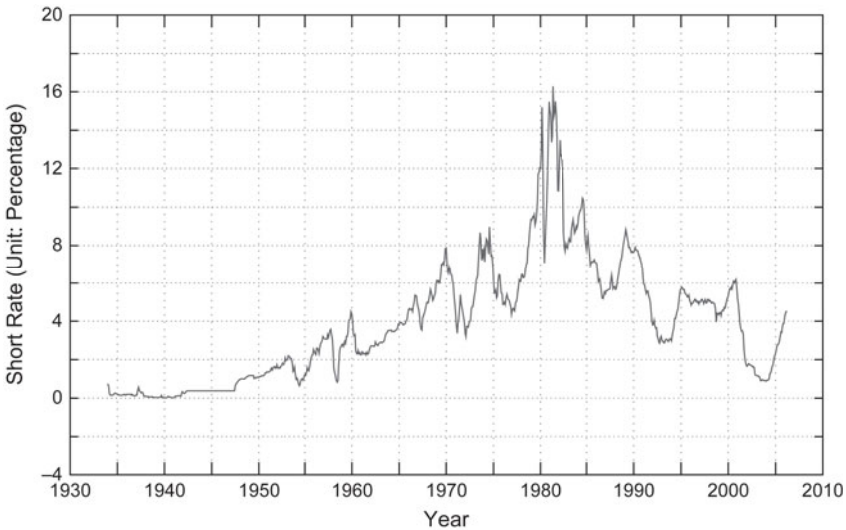


FIGURE 1. Three-month Treasury Bill: secondary market rate (percentage), 1 January 1934 to 1 March 2006. (Data from Board of Governors of the Federal Reserve System.)

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

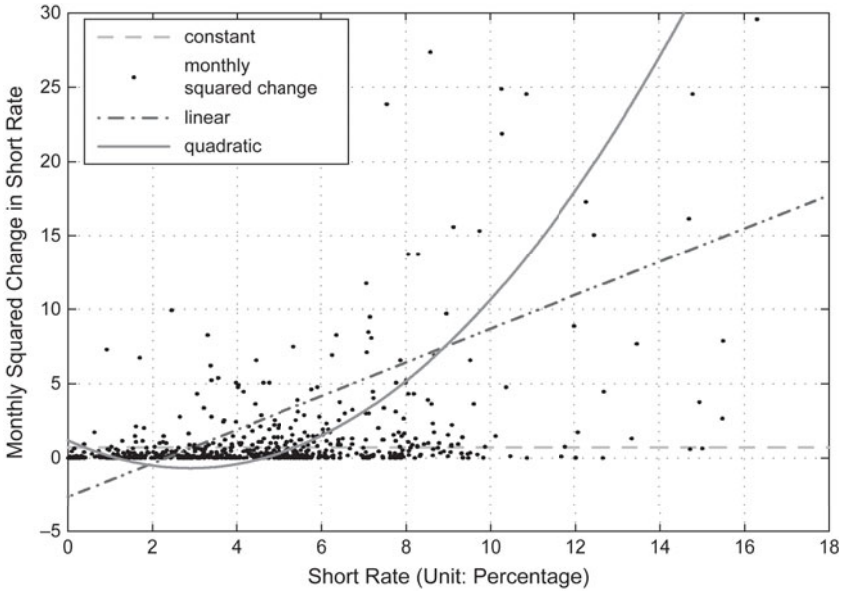


FIGURE 2. Monthly squared interest rate changes and (globally) constant, linear, and quadratic fits, 1 January 1934 to 1 March 2006.

reduces to the global linear (constant, quadratic, respectively) regression, as embodied in the curves plotted in Figure 2. Clearly the global linear and quadratic fittings give negative volatility estimates for some small values of interest rates. Now consider a realistic bandwidth selected by the rule of thumb $h = 4n^{-1/5}\widehat{s}$, where $\widehat{s}$ is the sample standard deviation of the time series, as used in Stanton (1997) and Chapman and Pearson (2000). Applying this to the data considered here, we have $h = 3.27\%$. Using this bandwidth, we present in Table 1 the local linear estimates (with Gaussian kernel) of the diffusion function at several design points and, for comparison, the estimates based on the RNW method proposed in this paper. We can see that the local linear method gives negative values for volatilities of the interest rates that are lower than 1.8%, whereas the reweighted method always provides positive and monotone volatility estimates as the interest rate levels decrease. These two methods tend to give similar volatility estimates for interest rates that are higher than 2.5%. Analogous observations can be obtained when the bandwidth is reduced to 2.18% (Table 2). The negativity problem of the local linear fitting may be mitigated by using small bandwidths, but it is known that small bandwidths lead to highly volatile results and thus unreliable inference in practical applications.

The remainder of the paper is organized as follows. Section 2 motivates the RNW functional diffusion estimator and compares it with other existing nonparametric estimators. Some implementation issues are discussed. The asymptotic

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

TABLE 2. Estimated diffusion function based on local linear (LL) and reweighted Nadaraya–Watson (RNW) approaches, calculated using monthly data, 1 January 1934 to 1 March 2006. The bandwidth is 2.18%. Units of design points: percentage.

x	1.0	1.2	1.4	1.5	1.6	1.7
LL	−0.0015	−0.00092	−0.00032	−0.0000036	0.00032	0.00065
RNW	0.0013	0.0014	0.0016	0.0017	0.0018	0.0019
x	1.8	1.9	2.0	2.5	3.0	4.0
LL	0.00098	0.0013	0.0017	0.0035	0.0054	0.0098
RNW	0.0020	0.0023	0.0026	0.0035	0.0052	0.0098

theory is given in Section 3. Section 4 discusses some possible extensions, and Section 5 concludes. The mathematical proofs of the main results are collected in an Appendix.

2. THE ESTIMATOR

2.1. Construction

In this section we introduce the reweighted Nadaraya–Watson (RNW) estimator of the diffusion function $\sigma^2(x)$ at a spatial point (or state) x. For a time-homogeneous diffusion process $\{X_t\}$ satisfying (1) and a smooth function g, the conditional expected increment can be expressed as

$$\mathrm{E}(g(X_{t+\Delta}) - g(X_t)|X_t) = \mathcal{L}g(X_t)\Delta + \frac{1}{2}\mathcal{L}^2 g(X_t)\Delta^2 + O(\Delta^3), \tag{2}$$

where the infinitesimal generator of the process $\{X_t\}$, $\mathcal{L}$, is defined by $\mathcal{L}g(y) = g'(y)\mu(y) + 1/2g''(y)\sigma^2(y)$. Setting $g(y) = y - X_t$ and $(y - X_t)^2$ in (2), we have

$$\mathrm{E}\left(\frac{X_{t+1} - X_t}{\Delta}|X_t\right) = \mu(X_t) + O(\Delta) \tag{3}$$

and

$$\mathrm{E}\left(\frac{(X_{t+1} - X_t)^2}{\Delta}|X_t\right) = \sigma^2(X_t) + O(\Delta), \tag{4}$$

respectively. Neglecting the higher order term in (4), the local polynomial estimator of $\sigma^2(x)$ is defined by $\widehat{\sigma}^2(x, h) = \beta_0$, where $\beta_0, \ldots, \beta_p$ minimize the weighted sum of squares $\sum_{i=1}^{n}\left((X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta - \sum_{j=0}^{p}\beta_j(X_{i\Delta} - x)^j\right)^2 K\left(\frac{X_{i\Delta} - x}{h}\right)$, given a kernel function $K(\cdot)$ and a smoothing bandwidth h. The local constant

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

(or Nadaraya–Watson) estimator and the local linear estimator are obtained by specifying $p = 0$ and $p = 1$, respectively, which have the following explicit-form expressions:

$$\widehat{\sigma}^2_{NW}(x,h) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)\frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta}}{\sum_{i=1}^n K_h(X_{i\Delta} - x)} \tag{5}$$

and

$$\widehat{\sigma}^2_{LL}(x,h) = \frac{\sum_{i=1}^n w_i^{LL}(x,h)K_h(X_{i\Delta} - x)\frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta}}{\sum_{i=1}^n w_i^{LL}(x,h)K_h(X_{i\Delta} - x)}, \tag{6}$$

where $w_i^{LL}(x,h) = S_{n,2} - (X_{i\Delta} - x)S_{n,1}$ with $S_{n,j} = \sum_{i=1}^n (X_{i\Delta} - x)^j K_h(X_{i\Delta} - x)$, $j = 1, 2$ and $K_h(\cdot) = K(\cdot/h)/h$. The local linear weights $w_i^{LL}(x,h)$ are not guaranteed to be nonnegative, and thus the resulting estimates may be negative.

The RNW estimator for the diffusion function $\sigma^2(x)$ is defined by

$$\widehat{\sigma}^2_{RNW}(x,h) = \frac{\sum_{i=1}^n w_i^{RNW}(x,h)K_h(X_{i\Delta} - x)\frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta}}{\sum_{i=1}^n w_i^{RNW}(x,h)K_h(X_{i\Delta} - x)}, \tag{7}$$

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

$$\left\{w_i^{RNW}(x,h)\right\} = \max_{\{w_i\}} \sum_{i=1}^n \log n w_i \quad \text{subject to} \tag{8}$$

$$w_i \geq 0, \qquad \sum_{i=1}^n w_i = 1, \tag{9}$$

and

$$\sum_{i=1}^n w_i \cdot (X_{i\Delta} - x)K_h(X_{i\Delta} - x) = 0. \tag{10}$$

Remark 1. The RNW diffusion estimator (7) has its root in the nonparametric regression (Hall and Presnell, 1999) as can be seen by noting the similarity between diffusion models and nonparametric regression models via (3) and (4). The discrete moment condition (10) is motivated by the local linear estimator whose weights $w_i^{LL}(x,h)$ satisfy (10) automatically. This condition is crucial in that it leads to many appealing properties of the local linear estimator such as bias reduction and automatic boundary correction (see Fan and Gijbels, 1996, Sect.3.2.2). When we drop the constraint (10), the constrained maximization problem (8) subject to (9) only is solved by the uniform weights $w_i = 1/n$, thus giving the local constant estimator (5).

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

Remark 2. The objective function in (8) corresponds to -1 times the Kullback–Leibler distance between the probability distribution $\{w_i\}$ and the uniform distribution $\{1/n, \ldots, 1/n\}$, although other distance measures may also be used (Cressie and Read, 1984). Thus the reweighted estimator (7) is designed to minimally adjust the usual local constant estimator such that the bias-reducing condition (10) is satisfied. The distance measure in (8) is chosen here because it leads to nice interpretation such as the log empirical likelihood ratio of Owen (2001). To illustrate this, we assume strict stationarity of the process $\{X_t\}$ here and in Remark 3, which follows, although the main results of the paper hold under much weaker conditions (cf. Section 3). Under strict stationarity, the local constant diffusion estimator (5) is the sample analog of

$$\frac{\mathrm{E}K\left(\frac{X_{i\Delta}-x}{h}\right)\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}}{\mathrm{E}K\left(\frac{X_{i\Delta}-x}{h}\right)} = \frac{\mathrm{E}K\left(\frac{X_{i\Delta}-x}{h}\right)\sigma^2(X_{i\Delta})}{\mathrm{E}K\left(\frac{X_{i\Delta}-x}{h}\right)} + O(\Delta),$$

which equals (with the higher order term neglected)

$$\frac{\sum_{i=1}^n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)}$$

if we assign the probability mass w_i to the observation $X_{i\Delta}$. Then the RNW estimator (7) follows with $\sigma^2(X_{i\Delta})$ replaced by $(X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta$. Although serial dependence is ignored in such construction, we can expect asymptotic validity of the empirical likelihood because we use essentially only spatially local data whose correlation structure is much weakened by localizing in the state domain (see Hart, 1996; Fan, 2005, Sect. 3.4).[3]

Remark 3. The asymptotic bias reduction in the estimator (7) with respect to the local constant estimator by imposing the restriction (10) is illustrated as follows. Under strict stationarity, the sampling error of the RNW estimator (7) is

$$\widehat{\sigma}^2_{RNW}(x,h) - \sigma^2(x) = \frac{\sum_{i=1}^n w_i^{RNW}(x,h)K_h(X_{i\Delta}-x)\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta} - \sigma^2(x)\right)}{\sum_{i=1}^n w_i^{RNW}(x,h)K_h(X_{i\Delta}-x)}. \tag{11}$$

The weight $w_i^{RNW}(x,h)$ is $1/n + o_p(1/n)$ as shown later (cf. (13) and (A.1)), and thus the denominator of (11) is the consistent estimator of the stationary density of the process $\{X_t\}$. By strict stationarity and (4), under certain conditions on h the expectation of the numerator of (11) is the same (except an asymptotically negligible term) as that of

$$\sum_{i=1}^n w_i^{RNW}(x,h)K_h(X_{i\Delta}-x)\Big(\sigma^2(X_{i\Delta}) - \sigma^2(x)\Big),$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

which by Taylor expansion can be written as

$$\sum_{i=1}^{n} w_i^{RNW}(x,h)K_h(X_{i\Delta}-x)\Big[(\sigma^2(x))'(X_{i\Delta}-x) + (\sigma^2(x))''\frac{(X_{i\Delta}-x)^2}{2} + \quad \text{higher order terms}\Big] \tag{12}$$

$$= \sum_{i=1}^{n} w_i^{RNW}(x,h)K_h(X_{i\Delta}-x)\Big[(\sigma^2(x))''\frac{(X_{i\Delta}-x)^2}{2} + \quad \text{higher order terms}\Big].$$

It makes clear that the restriction (10) eliminates the first term in (12) which contributes to the asymptotic bias of the local constant estimator. In Section 3 we show that the reweighted estimator (7) is asymptotically equivalent to the local linear estimator (6). A set of discrete moment conditions may be considered in addition to (10) to reduce the asymptotic bias further, but it would inflate the variance of the resulting estimator just as in the higher order local polynomial fitting.

Remark 4. The reweighting idea was first introduced by Hall and Presnell (1999) in a class of biased bootstrap techniques, which was designed to modify empirical methods "in a surgical way to alter some of their features while leaving others unchanged." Cai (2001) and Xu and Phillips (2008) extended their method to estimate the nonparametric mean function and the variance function, respectively, of discrete-time stationary mixing time series. Hall, Wolff, and Yao (1999) adapted this method in distribution function estimation so that the resulting estimates are always monotone and within the range $[0,1]$. Hall and Huang (2001) used a similar reweighting idea to monotonize general kernel-type estimators.

Remark 5. Two variants of the reweighted diffusion estimator (7) may be considered for improvement. The first is to use double smoothing as in Bandi and Phillips (2003) to reduce the asymptotic mean-squared error of the single-smoothing estimator (7). The second is to reduce the discretization bias (in terms of Δ) by building the diffusion functional estimator on the following representation instead of (4):

$$\mathrm{E}\left(\frac{(X_{t+\Delta}-X_t-\Delta\mu(X_t))^2}{\Delta}|X_t=x\right)=\sigma^2(x)+O(\Delta);$$

see Fan and Yao (1998) and Nicolau (2003). It needs a first-stage nonparametric estimator of the drift function $\mu(x)$, say, $\widetilde{\mu}(x)$, and the discretization-bias-reduced version of (7) can be constructed by replacing $(X_{(i+1)\Delta}-X_{i\Delta})^2/\Delta$ in (7) by $(X_{t+\Delta}-X_t-\Delta\widetilde{\mu}(X_t))^2/\Delta$. Stationarity would be assumed to justify such an estimator. We do not pursue these extensions further because they are found to perform quite similarly to (7) in finite samples.

2.2. Implementation

The weights $w_i^{RNW}(x,h)$ in the RNW diffusion function estimator (7) are straightforward to compute. Let the ordered observations be $X_{(\Delta)} \le X_{(2\Delta)} \le \cdots \le X_{(n\Delta)}$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

and assume that the spatial point x under consideration satisfies $X_{(\Delta)} < x < X_{(n\Delta)}$. The objective function in (8) is strictly concave on a convex set of w_i satisfying (9) and (10); thus a unique global maximum exists for an interior point of the domain. The Langrangian function for the constrained optimization problem is

$$\mathcal{L}_n^*(w_i, \gamma, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \log n w_i - \gamma\left(\sum_{i=1}^{n} w_i - 1\right) - \lambda \sum_{i=1}^{n} w_i \cdot (X_{i\Delta} - x) K_h(X_{i\Delta} - x),$$

where γ, λ are Lagrange multipliers.[4] The first-order conditions of $\mathcal{L}_n^*(w_i, \gamma, \lambda)$ with respect to w_i, γ, λ give

$$\begin{cases} 1/n - \gamma w_i - \lambda w_i \cdot (X_{i\Delta} - x) K_h(X_{i\Delta} - x) = 0, & \text{for each } i, \\ \sum_{i=1}^{n} w_i = 1, \\ \sum_{i=1}^{n} w_i \cdot (X_{i\Delta} - x) K_h(X_{i\Delta} - x) = 0, \end{cases}$$

from which we have $\gamma = 1$ and

$$w_i^{RNW}(x, h) = \frac{1}{n\left(1 + \lambda(X_{i\Delta} - x) K_h(X_{i\Delta} - x)\right)}, \tag{13}$$

where λ satisfies

$$\sum_{i=1}^{n} \frac{(X_{i\Delta} - x) K_h(X_{i\Delta} - x)}{n\left(1 + \lambda(X_{i\Delta} - x) K_h(X_{i\Delta} - x)\right)} = 0. \tag{14}$$

Thus the RNW diffusion estimator $\widehat{\sigma}_{RNW}^2(x, h)$ can be computed as follows.

Algorithm (RNW diffusion functional estimation at $\boldsymbol{x}$).

1. For a chosen value of the smoothing bandwidth h and a kernel function $K(\cdot)$, solve (14) numerically for the Lagrange multiplier λ.
2. Plug the value λ found in step 1 into (13) and get the weights $w_i^{RNW}(x, h)$.
3. Calculate $\widehat{\sigma}_{RNW}^2(x, h)$ according to (7) with the weights $w_i^{RNW}(x, h)$ obtained in step 2.

The value of λ is easy to find via a numerical search such as Newton's method by noting that the left side of (14) is a strictly decreasing function of λ. Let $q_i(x, h) = (X_{i\Delta} - x) K_h(X_{i\Delta} - x)$ and the ordered values be $q_{(1)}(x, h) \le \cdots \le q_{(n)}(x, h)$. The initial iteration interval $[\underline{\lambda}, \overline{\lambda}]$ for the numerical search can be chosen by setting w_i as one (which forces other weights to be zeros) in (13) for i such that $q_i(x, h)$ reaches the minimum and maximum, respectively, i.e., $\underline{\lambda} = q_{(n)}(x, h)^{-1}(1/n - 1)$ and $\overline{\lambda} = q_{(1)}(x, h)^{-1}(1/n - 1)$. The function on the left-hand side of (14) has positive and negative values, respectively, at $\underline{\lambda}$ and $\overline{\lambda}$ by

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

noting that $q_{(1)}(x,h) < 0$ and $q_{(n)}(x,h) > 0$, and thus the initial interval must contain λ, which solves (14). Then λ is obtained by refining the interval successively until the endpoints agree to a user-specified tolerance.

3. ASYMPTOTIC RESULTS

This section gives asymptotic results for the reweighted diffusion estimator described in Section 2. We need the following conditions.

Assumption 1. Bandi and Phillips, 2003.

(a) The initial condition $X_0 \in L^2$and is taken to be independent of $\{B_t, t \geq 0\}$.

(b) $\mu(\cdot)$ and $\sigma(\cdot)$ are measurable functions on $\mathcal{D} = (l,u)$ with $-\infty \leq l < u \leq \infty$. They are at least twice continuously differentiable and satisfy Lipschitz and growth conditions, namely, for any compact subset $S \subseteq \mathcal{D}$, there exist constants M_S and M'_S such that for all $x, y \in S$, $|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq M_S|x-y|$and $|\mu(x)| + |\sigma(x)| \leq M'_S(1+|x|)$.

(c) $\sigma^2(\cdot) > 0$ on $\mathcal{D}$.

(d) The scale function $S(\alpha) = \int_c^\alpha \exp\left(\int_c^y \left(-2\mu(x)/\sigma^2(x)\right) dx\right) dy, c \in \mathcal{D}$ satisfies $\lim_{\alpha \to l} S(\alpha) = -\infty$ and $\lim_{\alpha \to u} S(\alpha) = \infty$.

Assumption 1 assures existence and pathwise uniqueness of a strong recurrent solution to the stochastic differential equation (1). Intuitively, recurrence means that the Markov process visits every spatial point infinite times with probability one. It was shown by Bandi and Phillips (2003) to be sufficient to identify the nonparametric drift and diffusion functions from a discrete sample provided that we have a long time span and shrinking sampling intervals. See Bandi and Phillips (2009) for a nice treatment of recurrence-related concepts and their implication for statistical inference of stochastic processes. Assumption 1 does not imply existence of a stationary probability measure for X_t, which holds under stronger assumptions (see Assumption 4 later in this section); therefore nonstationarity is allowed. This is an important extension in that recent empirical studies of U.S. interest rate dynamics (see, among others, Bandi, 2002; Durham, 2003; Fan and Zhang, 2003) reveal that the estimated drift term is actually very close to zero (similar to the model of Cox, Ingersoll, and Ross, 1980). Recurrence is a reasonably weak assumption that allows for such martingale behavior.

The kernel function $K(\cdot)$ is assumed to satisfy the following conditions.

Assumption 2. The kernel function $K(\cdot)$ is a continuous differentiable, symmetric, and bounded density function with a bounded support $(-1,1)$, such that $|K(u)| < 1$ for any $u \in (-1,1)$ and $\int |K'(u)|^2 du < \infty$. Define $K_1 = \int u^2 K(u) du$, $K_2 = \int K^2(u) du$, and $v_2 = \int u^2 K^2(u)$.

The technical conditions imposed in Assumption 2 include most popular bounded supported kernel functions including the uniform, Epanechnikov, and quartic (or biweight) kernels, but exclude the triangular and triweight kernels.

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

The local time is important in analyzing a potentially nonstationary time series (Phillips, 1998, 2001). For a continuous semimartingale X_t, given a spatial point x and a time span $[0,T]$, the local time is defined as $L_X(T,x) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^T 1_{(x,x+\varepsilon)}(X_s)d[X]_s$, where $1_A(\cdot)$ is the indicator function on the set A and $[X]_s$ is the quadratic variation process of the underlying semimartingale. It gives the amount of time that the process spends in the neighborhood of x up to time T, measured in units of the quadratic variation process. For a diffusion process satisfying (1), $d[X]_s = \sigma^2(X_s)\,ds$, and thus it is convenient to consider a rescaled version of local time, $\overline{L}_X(T,x) = (1)/(\sigma^2(x))L_X(T,x)$. It can be interpreted as the sojourn time around x measured in pure time units up to time T (Park and Phillips, 1998). For a recurrent diffusion process, $\overline{L}_X(T,x)$ diverges as $T\to\infty$. We shall see that the local time plays at least two roles in our asymptotic theory. First, it can be seen as the generalized spatial density function. Second, it will be used to measure the convergence rate of the nonparametric estimator of the diffusion function. The local time $\overline{L}_X(T,x)$ can be consistently estimated by the quantity $\widehat{\overline{L}}_X(T,x) = \frac{\Delta}{h}\sum_{i=1}^n K((X_{i\Delta}-x)/h)$ for a fixed $T=\overline{T}$, under some additional regularity conditions provided in Bandi and Phillips (2003).

The limit theory developed in this section relies on the long time span ($T\to\infty$) and infill ($\Delta\to 0$) asymptotics specified as follows.

Assumption 3. Assume $n\to\infty$, $T\to\infty$, $\Delta = T/n\to 0$, and $h = h_{n,T}\to 0$ such that

$$\frac{\overline{L}_X(T,x)}{h}\sqrt{\Delta\log\frac{1}{\Delta}} = o_{a.s.}(1). \tag{15}$$

The following theorem gives strong consistency and asymptotic normality of the RNW estimator as defined in (7).

THEOREM. *Let K_1 and K_2 be as defined in Assumption 2. Under Assumptions 1–3, for each $x\in\mathcal{D}$,*

(i) $\widehat{\sigma}^2_{RNW}(x)\overset{a.s.}{\to}\sigma^2(x)$.

(ii) Furthermore, if $(h^5\overline{L}_X(T,x))/\Delta = O_{a.s.}(1)$, *then*

$$\sqrt{\frac{h\widehat{\overline{L}}_X(T,x)}{\Delta}}\left(\widehat{\sigma}^2_{RNW}(x)-\sigma^2(x)-\tfrac{h^2K_1}{2}\left(\sigma^2(x)\right)''\right)\Rightarrow\mathcal{N}\left(0,4K_2\sigma^4(x)\right). \tag{16}$$

The proofs of the theorem and some preliminary results are detailed in the Appendix. The theorem shows that the RNW estimator of the diffusion function has the same asymptotic distribution as that of the local linear estimator (Moloche, 2001). It removes the additional bias term of the local constant estimator, which depends on the speed function (see Assumption 4 later in this section) at x and its derivatives (Bandi and Phillips, 2003, p. 258). The bias term in (16) vanishes when

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

the true diffusion function is affine (e.g., in the short rate models of Vasicek, 1977; Cox, Ingersoll, and Ross, 1985; Duffie and Kan, 1996; see also Dai and Singleton, 2000).[5]

The theorem also suggests the following criterion for selecting the smoothing bandwidth:

$$h = c\left(\frac{\Delta}{\widehat{\widehat{L}}_X(T,x)}\right)^{1/5}, \tag{17}$$

where c is a constant, such that the asymptotic mean-squared error of the RNW estimator is minimized. The formula (17) implies that we use a smaller bandwidth at the spatial point x around which the local time is larger or the process visits more frequently. The constant c in (17) can be selected by the data-dependent method (e.g., cross-validation).

To facilitate comparisons with the results obtained under stationarity or ergodicity existing in the literature, we impose the following additional conditions.

Assumption 4. $\int_l^u s(y)\,dy < +\infty$, where $s(\cdot)$ is the speed function of the process X_t, namely, $s(\cdot) = 2/S'(\cdot)\sigma^2(\cdot)$.

Under Assumptions 1 and 4 the diffusion process X_t is positive recurrent (or ergodic), and the stationary distribution P^0 exists with the time-invariant density function defined as $f(x) = s(x)/\int_l^u s(y)\,dy$. Furthermore, the process X_t is stationary (in a strict sense) if the initial point X_0 has stationary distribution P^0.

By imposing positive recurrence as in Assumption 4, it can be shown (Bandi and Phillips, 2009, Thm. 3.3) that $\widehat{\widehat{L}}_X(T,x)/T \xrightarrow{a.s.} f(x)$ as $T \to \infty$. Thus the theorem holds in its weaker form:

COROLLARY. *Under Assumptions 1–4, for each $x \in \mathcal{D}$,*

(i) $\widehat{\sigma}^2_{RNW}(x) \xrightarrow{a.s.} \sigma^2(x)$.

(ii) Furthermore, if $h = O(n^{-1/5})$, *then*

$$\sqrt{nh}\left(\widehat{\sigma}^2_{RNW}(x) - \sigma^2(x) - \frac{h^2 K_1}{2}\left(\sigma^2(x)\right)''\right) \Rightarrow \mathcal{N}\left(0, 4K_2\frac{\sigma^4(x)}{f(x)}\right).$$

The smoothing bandwidth selection criterion (17) becomes $h = cn^{-1/5}$ when the process is stationary or ergodic.

Although we focus on the diffusion function in this paper, the RNW estimator for the drift function $\mu(\cdot)$ can be constructed similarly by replacing $(X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta$ in (7) by $(X_{(i+1)\Delta} - X_{i\Delta})/\Delta$, serving as an alternative to the local linear drift estimator, and the asymptotic theory can also be similarly developed. It has a slower convergence rate than that of the diffusion function estimator because of the different orders of the drift and diffusion terms at high frequencies.

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

4. EXTENSIONS

This paper has concentrated exclusively on the univariate continuous-time diffusion model. The results here naturally extend to the multivariate diffusion model, where the diffusion function is matrix-valued, describing the instantaneous variances and covariances of the underlying process. We illustrate this by considering a bivariate diffusion process $X_t = (X_{1,t}, X_{2,t})'$. The local linear estimator of the bivariate diffusion function at a spatial point $x = (x_1, x_2)'$ can be written as

$$\widehat{\sigma}^2_{LL}(x,h) = \frac{\sum_{i=1}^n w_i^{LL}(x,h)\mathbf{K}_h(X_{i\Delta}-x)\frac{(X_{(i+1)\Delta}-X_{i\Delta})(X_{(i+1)\Delta}-X_{i\Delta})'}{\Delta}}{\sum_{i=1}^n w_i^{LL}(x,h)\mathbf{K}_h(X_{i\Delta}-x)}, \tag{18}$$

where $\mathbf{K}_h(X_{i\Delta}-x) = h^{-2}K\left((X_{1,i\Delta}-x_1)/h\right)K\left((X_{2,i\Delta}-x_2)/h\right)$ is a product kernel function and

$$w_i^{LL}(x,h) = \widetilde{S}_1 - \widetilde{S}_2(X_{1,i\Delta}-x_1) + \widetilde{S}_3(X_{2,i\Delta}-x_2) \tag{19}$$

with

$$\widetilde{S}_1 = \det\begin{pmatrix}\widetilde{S}_{(2,0)} & \widetilde{S}_{(1,1)}\\ \widetilde{S}_{(1,1)} & \widetilde{S}_{(0,2)}\end{pmatrix}, \quad \widetilde{S}_2 = \det\begin{pmatrix}\widetilde{S}_{(1,0)} & \widetilde{S}_{(1,1)}\\ \widetilde{S}_{(0,1)} & \widetilde{S}_{(0,2)}\end{pmatrix}, \quad \widetilde{S}_3 = \det\begin{pmatrix}\widetilde{S}_{(1,0)} & \widetilde{S}_{(2,0)}\\ \widetilde{S}_{(0,1)} & \widetilde{S}_{(1,1)}\end{pmatrix},$$

where det means the determinant of a matrix and $\widetilde{S}_{(i,j)} = \sum_{i=1}^n (X_{1,i\Delta}-x_1)^j (X_{2,i\Delta}-x_2)^k \mathbf{K}_h(X_{i\Delta}-x)$ for $j,k = 0,1,2$. Although a limit theory of (18) is absent from the literature even under stationarity, it is expected to serve as an improvement of the local constant multivariate diffusion function estimator recently studied by Bandi and Moloche (2004) and Boudoukh, Downing, Richardson, Stanton, and Whitelan (2007), but it may produce negative estimates for the individual diffusion functions. The reweighted Nadaraya–Watson estimator can be constructed similarly as in the univariate case by replacing the $w_i^{LL}(x,h)$ in (19) by weights such that (8) is maximized subject to (9) and the following bias-reducing conditions:

$$\sum_{i=1}^n w_i \cdot (X_{1,i\Delta}-x_1)\mathbf{K}_h(X_{i\Delta}-x) = 0 \tag{20}$$

and

$$\sum_{i=1}^n w_i \cdot (X_{2,i\Delta}-x_2)\mathbf{K}_h(X_{i\Delta}-x) = 0, \tag{21}$$

which are satisfied by $w_i^{LL}(x,h)$ in (19). The resulting estimator is expected to reduce the bias of the local constant bivariate diffusion functional estimate just as local linear smoothing does, and it is always nonnegative in small samples. The asymptotic theory under recurrence is much more difficult than the univariate

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

case in light of the absence of the local time notion for multivariate diffusions, but the difficulty can be overcome by employing the density of the occupation time measure of the underlying process, as studied in Bandi and Moloche (2004).

Other extensions of the current work are possible. First, in this paper we assume time homogeneity, which can be restrictive in characterizing potential nonstationarity in the data generating process. It would be interesting to extend the approach considered here to the time-varying coefficient diffusion model, although the asymptotic results would be more demanding to establish. Another extension of practical importance is to allow for jumps in the process, which offers more flexibility of generating the higher moments such as skewness and kurtosis of financial time series. Finally, we only consider the one-factor model here, and it would be useful to extend our approach to the nonparametric estimation of volatility structure within the multifactor Heath–Jarrow–Morton framework (Jeffrey, Kristensen, Linton, Nguyen, and Phillips, 2004). In all these extensions, the reweighted idea developed here would be powerful in overcoming the negativity problem of the well-accepted local polynomial estimation method.

5. CONCLUSION

In this paper we propose a novel reweighted Nadaraya–Watson estimator for the diffusion function in a time-homogeneous scalar diffusion model based on a discrete sample of observations. The idea is to adjust the conventional Nadaraya–Watson estimator by tilting the empirical distribution minimally subject to a discrete-bias-reducing moment condition. Although the popular local linear smoothing is superior to the usual Nadaraya–Watson method because of its many desirable properties, more and more researchers have begun to be quite careful in using it when estimating the diffusion function because it may produce negative results in finite samples and thus lead to insensible inference of the model. The new diffusion estimator studied in this paper preserves the property of the local linear estimator by removing the bias in kernel-type estimates without materially affecting their variance, and it has the additional advantage of always being nonnegative. Empirical likelihood is used to shift the weights in a manner that is most suited to the data in the context of the given model. The new reweighted estimator is expected to be a very promising competitor to other estimating methods for the continuous-time models.

NOTES

1. See Cai and Hong (2003), Fan (2005), Aït-Sahalia (2007), and Gao (2007, Ch. 5) for recent surveys on this topic.

2. The multiplicative bias reduction method of Linton and Nielsen (1994) is also promising in achieving bias reduction while keeping positivity of the diffusion estimator although further verification is needed. We thank the co-editor for pointing out this.

3. See, Chen, Gao, and Tang (2008) and Xu (2006) for other recent applications of empirical likelihood in nonparametric continuous-time diffusion models.

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

4. Here we maximize the objective function in (8) times $1/n$ for computational convenience.

5. The asymptotic behavior of diffusion functional estimators can also be studied for a fixed time span T; see Jiang and Knight (1997) and Bandi and Phillips (2003, Thm. 6). We adopt here the large time span asymptotics only because it characterizes the bias term and thus facilitates comparisons with other existing estimators.

REFERENCES

Ahn, D.-H. & B. Gao (1999) A parametric nonlinear model of term structure dynamics. *Review of Financial Studies* 12, 721–762.

Aït-Sahalia, Y. (1996a) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.

Aït-Sahalia, Y. (1996b) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y. (2007) Estimating continuous-time models using discretely sampled data. In R. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress.* Cambridge University Press.

Arapis, M. & J. Gao (2006) Empirical comparisons in short-term interest rate models using nonparametric methods. *Journal of Financial Econometrics* 4, 310–345.

Bandi, F. (2002) Short-term interest rate dynamics: A spatial approach. *Journal of Financial Economics* 65, 73–110.

Bandi, F. & G. Moloche (2004) On the Functional Estimation of Multivariate Diffusion Processes. Working paper, University of Chicago.

Bandi, F. & P.C.B. Phillips (2009) Nonstationary continuous-time processes. In Y. Aït-Sahalia, & L.P. Hansen (eds.), *Handbook of Financial Econometrics*. North-Holland.

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

Boudoukh, J., C. Downing, M. Richardson, R. Stanton, & R.F. Whitelaw (2007) A Multifactor, Nonlinear, Continuous-Time Model of Interest Rates Volatility. Working paper, University of California, Berkeley.

Cai, Z. (2001) Weighted Nadaraya-Watson regression estimation. *Statistics and Probability Letters* 51, 307–318.

Cai, Z. & Y. Hong (2003) Nonparametric methods in continuous-time finance: A selective review. In M.G. Akritas & D.N. Politis (eds.), *Recent Advances and Trends in Nonparametric Statistics,* pp. 283–302. North-Holland.

Chapman, D.A. & N.D. Pearson (2000) Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355–388.

Chen, S.X., J. Gao, & C.Y. Tang (2008) A test for model specification of diffusion processes. *Annals of Statistics* 36, 167–198.

Chen, S.X. & Y. Qin (2002) Confidence interval based on a local linear smoother. *Scandinavian Journal of Statistics* 29, 89–99.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1980) An analysis of variable rate loan contracts. *Journal of Finance* 35, 389–403.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–467.

Cressie, N.A.C. & T.R.C. Read (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, series B* 46, 440–464.

Dai, Q. & K. Singleton (2000) Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.

Duffie, D. & R. Kan (1996) A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.

Durham, G.B. (2003) Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics* 70, 463–487.

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

Fan, J. (2005) A selective overview of nonparametric methods in financial econometrics (with discussion). *Statistical Science* 20, 317–357.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*. Chapman and Hall.

Fan, J. & Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.

Fan, J. & C. Zhang (2003) A re-examination of Stanton's diffusion estimations with applications to financial model validation. *Journal of the American Statistical Association* 98, 118–134.

Fan, J., C.M. Zhang, & J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability* 30, 790–804.

Gao, J. (2007) *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman and Hall/CRC.

Hall, P. & L.-S. Huang (2001) Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* 29, 624–647.

Hall, P. & B. Presnell (1999) Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society, series B* 61, 143–158.

Hall, P. & N. Tajvidi (2000) Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science* 15, 153–167.

Hall, P., R.C.L. Wolff, & Q. Yao (1999) Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* 94, 154–163.

Hart, J.D. (1996) Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* 6, 115–142.

Hong, Y. & H. Li (2005) Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* 18, 37–84.

Jeffrey, A., D. Kristensen, O. Linton, T. Nguyen, & P.C.B. Phillips (2004) Nonparametric estimation of a multifactor Heath-Jarrow-Morton model: An integrated approach. *Journal of Financial Econometrics* 2, 251–289.

Jiang, G.J. & J. Knight (1997) A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. *Econometric Theory* 13, 615–645.

Linton, O. & J.P. Nielsen (1994) A multiplicative bias reduction method for nonparametric regression. *Statistics and Probability Letters* 19, 181–187.

Moloche, G. (2001) Local Nonparametric Estimation of Scalar Diffusions. Mimeo, MIT.

Nicolau, J. (2003) Bias reduction in nonparametric diffusion coefficient estimation. *Econometric Theory* 19, 754–777.

Owen, A. (2001) *Empirical Likelihood*. Chapman and Hall/CRC.

Park, J. & P.C.B. Phillips (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion Paper 1181, Yale University.

Phillips, P.C.B. (1998) Econometric Analysis of Fisher's Equation. Cowles Foundation Discussion Paper 1180, Yale University.

Phillips, P.C.B. (2001) Descriptive econometrics for nonstationary time series with empirical illustrations. *Journal of Applied Econometrics* 16, 389–413.

Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* 11, 449–487.

Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.

Vasicek, O.A. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

Xu, K.-L. (2009) Empirical likelihood based inference for nonparametric recurrent diffusions. *Journal of Econometrics*, forthcoming.

Xu, K.-L. & P.C.B. Phillips (2008) Tilted Nonparametric Estimation of Volatility Functions with Empirical Applications. Cowles Foundation Discussion Paper 1612R, Yale University.

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

Yu, K. & M.C. Jones (2004) Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association* 99, 139–144.

Ziegelmann, F.A. (2002) Nonparametric estimation of volatility functions: The local exponential estimator. *Econometric Theory* 18, 985–991.

APPENDIX: Proofs

In this Appendix we provide the proof of the theorem and some preliminary results. The proofs are built upon results of Bandi and Phillips (2003), and readers are referred to that paper for more details. Throughout the Appendix, for simplicity we write the weights $w_i^{RNW}(x,h)$ in the RNW estimator simply as w_i. We also write $\underline{K}_{(jk)}(u) = u^j K^k(u)$, where j,k are nonnegative integers and $K(\cdot)$ is the kernel function satisfying Assumption 2. Integrals are taken over the real line unless otherwise specified.

The proof of the theorem relies on the following lemma, which characterizes the order of the Langrange multiplier λ in (13).

LEMMA. *Let $s(\cdot)$ be the speed function of the process X_t, namely, $s(\cdot) = 2/S'(\cdot)\sigma^2(\cdot)$ and let K_1 and v_2 be defined in Assumption 2. Under Assumptions 1–3,*

$$\lambda = \frac{hK_1 s'(x)}{v_2 s(x)} + O_{a.s.}(h^3). \tag{A.1}$$

Under the stronger Assumptions 1–4 (which imply existence of the time-invariant density function $f(\cdot)$ of the process X_t), the Langrange multiplier λ can be rewritten as

$$\lambda = \frac{hK_1 f'(x)}{v_2 f(x)} + O_{a.s.}(h^3),$$

which is comparable with the corresponding results obtained under stationarity in the discrete-time time series context (e.g., Cai, 2001).

Proof. Let $A_j = \frac{1}{n}\sum_{i=1}^n \underline{K}_{(jj)}((X_{i\Delta} - x)/h)$ for any integer $j \geq 1$. First we show that

$$\sum_{i=1}^n \frac{\Delta}{h}\underline{K}_{(jj)}\left(\frac{X_{i\Delta}-x}{h}\right) = \int_0^T \frac{1}{h}\underline{K}_{(jj)}\left(\frac{X_s - x}{h}\right) ds + o_{a.s.}(1). \tag{A.2}$$

Indeed, by the triangular inequality (noting $T = n\Delta$),

$$\begin{aligned}
&\left|\textstyle\sum_{i=1}^n \frac{\Delta}{h}\underline{K}_{(jj)}\left(\frac{X_{i\Delta}-x}{h}\right) - \int_0^T \frac{1}{h}\underline{K}_{(jj)}\left(\frac{X_s-x}{h}\right) ds\right| \\
&\quad = \frac{1}{h}\left|\textstyle\sum_{i=1}^n \Delta \underline{K}_{(jj)}\left(\frac{X_{i\Delta}-x}{h}\right) - \int_0^T \underline{K}_{(jj)}\left(\frac{X_s-x}{h}\right) ds\right| \\
&\quad \leq \frac{1}{h}\left|\textstyle\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} \underline{K}_{(jj)}\left(\frac{X_{i\Delta}-x}{h}\right) ds - \sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} \underline{K}_{(jj)}\left(\frac{X_s-x}{h}\right) ds\right| \\
&\qquad + \frac{\Delta}{h}\left|\underline{K}_{(jj)}\left(\frac{X_0-x}{h}\right)\right| + \frac{\Delta}{h}\left|\underline{K}_{(jj)}\left(\frac{X_T-x}{h}\right)\right|
\end{aligned}$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

$$\leq \frac{1}{h}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left|\underline{K}_{(jj)}\left(\tfrac{X_{i\Delta}-x}{h}\right)-\underline{K}_{(jj)}\left(\tfrac{X_s-x}{h}\right)\right|ds+O_{a.s.}\left(\frac{\Delta}{h}\right)$$

$$=\frac{1}{h}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left|\underline{K}'_{(jj)}\left(\tfrac{\widetilde{X}_{is}-x}{h}\right)\right|\cdot\left|\tfrac{X_{i\Delta}-X_s}{h}\right|ds+o_{a.s.}(1), \tag{A.3}$$

where $\widetilde{X}_{is}$ is some point between $X_{i\Delta}$ and X_s. The second inequality in the preceding derivation also uses the boundedness of the kernel function which is supported on $[0,1]$. The second equality uses the fact that $O_{a.s.}\left(\frac{\Delta}{h}\right)=o_{a.s.}(1)$, which is implied by (15). Define $\kappa_{n,T}=\max_{i\leq n}\sup_{i\Delta\leq s\leq(i+1)\Delta}|X_{i\Delta}-X_s|$, and it follows from Lévy's modulus of continuity of diffusions that $\kappa_{n,T}=O_{a.s.}\left(\sqrt{\Delta\log\frac{1}{\Delta}}\right)$. Then the first term on the right-hand side of (A.3) is bounded by

$$\begin{aligned}
&\frac{\kappa_{n,T}}{h^2}\int_0^T\left|\underline{K}'_{(jj)}\left(\tfrac{X_s-x}{h}+O_{a.s.}\left(\tfrac{1}{h}\sqrt{\Delta\log\tfrac{1}{\Delta}}\right)\right)\right|ds\\
&\quad=\frac{\kappa_{n,T}}{h^2}\int_{-\infty}^{\infty}\left|\underline{K}'_{(jj)}\left(\tfrac{p-x}{h}+o_{a.s.}(1)\right)\right|\overline{L}_X(T,p)\,dp\\
&\quad=\frac{\kappa_{n,T}}{h}\int_{-\infty}^{\infty}\left|\underline{K}'_{(jj)}(q+o_{a.s.}(1))\right|\overline{L}_X(T,hq+x)\,dq\\
&\quad\leq C\,\frac{\kappa_{n,T}}{h}O_{a.s.}(\overline{L}_X(T,x))\\
&\quad=o_{a.s.}(1),
\end{aligned}$$

where the first equality follows from the well-known occupation time formula (e.g., Bandi and Phillips, 2003, p. 246), the second equality from the change of variables $q=(p-x)/h$, and the third equality from (15). The inequality is by the absolute integrability of $\underline{K}'_{(jj)}(\cdot)$, which follows from (using the Cauchy–Schwarz inequality)

$$\begin{aligned}
&\int_{-1}^{1}|\underline{K}'_{(jj)}(u)|\,du\\
&\quad\leq\int_{-1}^{1}j|u^{j-1}K^j(u)|\,du+\int_{-1}^{1}j|u^jK^{j-1}(u)K'(u)|\,du\\
&\quad\leq\left(\textstyle\int_{-1}^{1}|u^{j-1}|^2\,du\cdot\int_{-1}^{1}|jK^j(u)|^2\,du\right)^{1/2}\\
&\qquad+\left(\textstyle\int_{-1}^{1}|juK^{j-1}(u)|^2\,du\cdot\int_{-1}^{1}|K'(u)|^2\,du\right)^{1/2}\\
&\quad<\infty
\end{aligned}$$

uniformly in j. Here we use the fact that for a real number a such that $|a|<1$, ja^j is uniformly (in j) bounded. Thus (A.2) holds. Let $s(x)$ be the speed function defined in the lemma. For $j\neq 2$,

$$\frac{A_j}{A_2}=\frac{\sum_{i=1}^n\frac{\Delta}{h}\underline{K}_{(jj)}\left(\frac{X_{i\Delta}-x}{h}\right)}{\sum_{i=1}^n\frac{\Delta}{h}\underline{K}_{(22)}\left(\frac{X_{i\Delta}-x}{h}\right)} \tag{A.4}$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{\int_0^T \frac{1}{h}\underline{K}_{(jj)}\left(\frac{X_s - x}{h}\right) ds + o_{a.s.}(1)}{\int_0^T \frac{1}{h}\underline{K}_{(22)}\left(\frac{X_s - x}{h}\right) ds + o_{a.s.}(1)} \\
&= \frac{\int \frac{1}{h}\underline{K}_{(jj)}\left(\frac{z-x}{h}\right) s(z)\, dz + o_{a.s.}(1)}{\int \frac{1}{h}\underline{K}_{(22)}\left(\frac{z-x}{h}\right) s(z)\, dz + o_{a.s.}(1)} + o_{a.s.}(1) \\
&= \frac{\int u^j K^j(u) s(x+uh)\, du + o_{a.s.}(1)}{\int u^2 K^2(u) s(x+uh)\, du + o_{a.s.}(1)} + o_{a.s.}(1) \\
&= \frac{\int u^j K^j(u)[s(x) + uhs'(x) + O(h^2)]\, du + o_{a.s.}(1)}{\int u^2 K^2(u)[s(x) + uhs'(x) + O(h^2)]\, du + o_{a.s.}(1)} + o_{a.s.}(1) \\
&= \begin{cases} \dfrac{h(\int u^{j+1} K^j(u)\, du) \cdot s'(x) + O(h^2) + o_{a.s.}(1)}{v_2 s(x) + O(h^2) + o_{a.s.}(1)} + o_{a.s.}(1), & j = 1, 3, 5, \ldots \\ \dfrac{\int u^j K^j(u)\, du \cdot s(x) + O(h^2) + o_{a.s.}(1)}{v_2 s(x) + O(h^2) + o_{a.s.}(1)} + o_{a.s.}(1), & j = 4, 6, 8, \ldots \end{cases} \qquad \textbf{(A.5)}
\end{aligned}
$$

where the third and fourth equalities follow from Bandi and Phillips (2003, Lem. 6, p. 248) and change of variables $(z-x)/h = u$, respectively, and the last two equalities from the symmetry of $K(\cdot)$. Let $q_i = \underline{K}_{(11)}((X_{i\Delta} - x)/h)$, then $A_j = \frac{1}{n}\sum_{i=1}^n q_i^j$. By Assumption 2

$$\max_{1 \le i \le n} |q_i| \le C < \infty. \qquad \textbf{(A.6)}$$

By (9) and (13), we have $\sum_{i=1}^n [n(1+\lambda q_i)]^{-1} = 1$ or

$$\frac{1}{n}\sum_{i=1}^n \frac{-q_i}{1+\lambda q_i} = 0, \qquad \textbf{(A.7)}$$

and therefore (noting that $1 + \lambda q_i = (nw_i)^{-1} \ge 0$)

$$
\begin{aligned}
0 &= \left| \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\frac{\lambda q_i^2}{1+\lambda q_i} - q_i \right) \right| \\
&\ge \frac{1}{n}\sum_{i=1}^n \frac{|\lambda| q_i^2}{|1+\lambda q_i|} - \left| \tfrac{1}{n}\textstyle\sum_{i=1}^n q_i \right| \\
&\ge \frac{|\lambda| A_2}{1 + C|\lambda|} - |A_1| \\
&= \frac{(A_2 - C|A_1|)|\lambda| - |A_1|}{1 + C|\lambda|},
\end{aligned}
$$

because for any i, $|1 + \lambda q_i| \le 1 + |\lambda| C$ by (A.6). Thus we have $(A_2 - C|A_1|)|\lambda| \le |A_1|$ or $(1 - C\left|A_1/A_2\right|)|\lambda| \le \left|A_1/A_2\right|$. So in view of (A.5), we have

$$|\lambda| = O_{a.s.}(h). \qquad \textbf{(A.8)}$$

By (A.7),

$$0 = \frac{1}{n}\sum_{i=1}^n \frac{q_i}{1+\lambda q_i} = A_1 - \lambda A_2 + \lambda^2 A_3 - \lambda^3 A_4 + O_{a.s.}(h^4)$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

or

$$\lambda = \frac{A_1}{A_2} + \lambda^2 \frac{A_3}{A_2} - \lambda^3 \frac{A_4}{A_2} + O_{a.s.}(h^4) = \frac{hK_1 s'(x)}{v_2 s(x)} + O_{a.s.}(h^3),$$

in view of (A.5) again. This completes the proof of the lemma. ■

Proof of the Theorem.

(i). The theorem can be proved by showing

$$\frac{\sum_{i=1}^n nw_i K_h(X_{i\Delta} - x)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n nw_i K_h(X_{i\Delta} - x)} \xrightarrow{a.s.} \sigma^2(x) \tag{A.9}$$

and

$$\frac{\sum_{i=1}^n nw_i K_h(X_{i\Delta} - x)\left(\frac{(X_{(i+1)\Delta} - X_{i\Delta})^2}{\Delta} - \sigma^2(X_{i\Delta})\right)}{\sum_{i=1}^n nw_i K_h(X_{i\Delta} - x)} \xrightarrow{a.s.} 0. \tag{A.10}$$

Note that (A.10) follows from $(X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta = \sigma^2(X_{i\Delta}) + o_{a.s.}(1)$, which has been shown in Bandi and Phillips (2003, similarly to their (7.11), p. 268). So we only need to show (A.9), which follows from

$$\frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \xrightarrow{a.s.} \sigma^2(x) \tag{A.11}$$

and

$$\frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta} - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \xrightarrow{a.s.} 1. \tag{A.12}$$

Look at (A.11) first. In view of (A.8), $|\lambda(X_{i\Delta} - x)K_h(X_{i\Delta} - x)| < 1$ for a sufficiently large n, and so by Taylor expansion we have

$$\begin{aligned}
&\frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \\
&= \frac{\sum_{i=1}^n \frac{1}{1+\lambda(X_{i\Delta} - x)K_h(X_{i\Delta} - x)} K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \\
&= \frac{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \\
&\quad - \lambda \frac{\sum_{i=1}^n (X_{i\Delta} - x)K_h(X_{i\Delta} - x)K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \\
&\quad + \lambda^2 \frac{\sum_{i=1}^n \left((X_{i\Delta} - x)K_h(X_{i\Delta} - x)\right)^2 K\left(\frac{X_{i\Delta} - x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta} - x}{h}\right)} \\
&\quad + O_{a.s.}(h^3).
\end{aligned} \tag{A.13}$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

The first term in (A.13) converges almost surely to $\sigma^2(x)$, as shown in Bandi and Phillips (2003). Following steps similar to those used in establishing (A.5), we can obtain

$$\frac{\sum_{i=1}^n \underline{K}_{(jk)}\left(\frac{X_{i\Delta}-x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)} \xrightarrow{a.s.} \int u^j K^k(u)\,du \cdot \sigma^2(x) \tag{A.14}$$

and

$$\frac{\sum_{i=1}^n \underline{K}_{(jk)}\left(\frac{X_{i\Delta}-x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)} \xrightarrow{a.s.} \int u^j K^k(u)\,du. \tag{A.15}$$

The right sides of (A.14) and (A.15) vanish when j is an odd integer. So by (A.13), (A.14), and the lemma,

$$\frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)} = \frac{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)\sigma^2(X_{i\Delta})}{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)} + O_{a.s.}(h^2) \xrightarrow{a.s.} \sigma^2(x).$$

Thus (A.11) holds. Similarly we can show (A.12). Therefore (A.9) holds, and part (i) is proved.

(ii) Write

$$\begin{aligned}\widehat{\sigma}^2_{RNW}(x) - \sigma^2(x) &= \frac{\sum_{i=1}^n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta} - \sigma^2(x)\right)}{\sum_{i=1}^n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)} \\ &= V_n + B_n,\end{aligned} \tag{A.16}$$

where

$$V_n = \frac{\frac{\Delta}{h}\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta} - \sigma^2(X_{i\Delta})\right)}{\frac{\Delta}{h}\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)},$$

$$B_n = \frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(X_{i\Delta}) - \sigma^2(x)\right)}{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)}.$$

Consider the term B_n first. We shall show that

$$B_n = \frac{h^2 K_1}{2}\left(\sigma^2(x)\right)'' + O_{a.s.}(h^3), \tag{A.17}$$

which follows from

$$\frac{\sum_{i=1}^n nw_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(X_{i\Delta}) - \sigma^2(x)\right)}{\sum_{i=1}^n K\left(\frac{X_{i\Delta}-x}{h}\right)} = \frac{h^2 K_1}{2}\left(\sigma^2(x)\right)'' + O_{a.s.}(h^3), \tag{A.18}$$

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press

in view of (A.12). In fact, by Taylor expansion the left side of (A.18) equals

$$\begin{aligned}
&\frac{\frac{\Delta}{h}\sum_{i=1}^{n} n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(X_{i\Delta})-\sigma^2(x)\right)}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad=\frac{\frac{\Delta}{h}\sum_{i=1}^{n} n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left((X_{i\Delta}-x)\left(\sigma^2(x)\right)'+\left(\sigma^2(x)\right)''\frac{(X_{i\Delta}-x)^2}{2}\right)+o_{a.s.}(h^2)}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad=\frac{\frac{\Delta}{2h}\sum_{i=1}^{n}\frac{1}{1+\lambda(X_{i\Delta}-x)K_h(X_{i\Delta}-x)}(X_{i\Delta}-x)^2 K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(x)\right)''+o_{a.s.}(h^2)}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)},
\end{aligned} \tag{A.19}$$

where the second equality follows from the bias-reducing moment condition (10). Neglecting the smaller order term, the numerator in (A.19) can be written as

$$\begin{aligned}
&\frac{\Delta h}{2}\sum_{i=1}^{n}\underline{K}_{(21)}\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(x)\right)''-\frac{\Delta h}{2}\lambda\sum_{i=1}^{n}\underline{K}_{(32)}\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(x)\right)'' \\
&\quad+\frac{\Delta h}{2}\lambda^2\sum_{i=1}^{n}\underline{K}_{(43)}\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(x)\right)''-\cdots.
\end{aligned}$$

Thus by (A.15) and the lemma, (A.19) equals

$$\frac{\frac{\Delta h}{2}\sum_{i=1}^{n}\underline{K}_{(21)}\left(\frac{X_{i\Delta}-x}{h}\right)\left(\sigma^2(x)\right)''}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)}+O_{a.s.}(h^3)=\frac{1}{2}h^2 K_1\left(\sigma^2(x)\right)''+O_{a.s.}(h^3)$$

and (A.17) holds. Next we study the term V_n in (A.16), and we shall show that

$$\sqrt{\frac{h\widehat{\overline{L}}_X(T,x)}{\Delta}}V_n\Rightarrow\mathcal{N}(0,4K_2\sigma^4(x)). \tag{A.20}$$

By the arguments used before, it can be shown that

$$\begin{aligned}
&\frac{\frac{\Delta}{h}\sum_{i=1}^{n} n w_i K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}-\sigma^2(X_{i\Delta})\right)}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)} \\
&\quad=\frac{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta}-\sigma^2(X_{i\Delta})\right)}{\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{i\Delta}-x}{h}\right)}+o_p\left(\sqrt{\frac{\Delta}{h\overline{L}_X(T,x)}}\right),
\end{aligned}$$

where $\sqrt{\frac{h\widehat{\overline{L}}_X(T,x)}{\Delta}}$ times the first part has been shown in Bandi and Phillips (2003) to converge weakly to $\mathcal{N}(0,4K_2\sigma^4(x))$. Thus (A.20) holds by (A.12). Combining (A.16), (A.17), and (A.20), part (ii) is proved, and therefore the proof of the theorem is complete. ■

https://doi.org/10.1017/S0266466609100087 Published online by Cambridge University Press