Nonparametric estimation of scalar diffusion models of interest rates using asymmetric kernels ☆

Nikolay Gospodinov [a,*], Masayuki Hirukawa [b,1]

[a] Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8
[b] Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka, 572-8508, Japan

ABSTRACT

This paper proposes an asymmetric kernel-based method for nonparametric estimation of scalar diffusion models of spot interest rates. We derive the asymptotic theory for the asymmetric kernel estimators of the drift and diffusion functions for general and positive recurrent processes and illustrate the advantages of the Gamma kernel for bias correction and efficiency gains. The finite-sample properties and the practical relevance of the proposed nonparametric estimators for bond and option pricing are evaluated using actual and simulated data for U.S. interest rates.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Nonparametric methods have attracted growing attention in econometrics due to their flexibility for handling possible nonlinearities in conditional moment function estimation. This paper focuses on nonparametric estimation of time-homogeneous drift and diffusion functions in continuous-time models that are used to describe the underlying dynamics of spot interest rates. More specifically, we improve the nonparametric estimators of drift and diffusion functions of Stanton (1997) by means of a nonstandard smoothing technique. The main source for this improvement comes from asymmetric kernel functions which are employed in place of standard symmetric kernels.

Kernel smoothing has been widely applied for estimating continuous-time diffusion processes; examples of recent contributions include Florens-Zmirou (1993), Aït-Sahalia (1996a,b), Jiang and Knight (1997), Stanton (1997), Chapman and Pearson (2000), Bandi
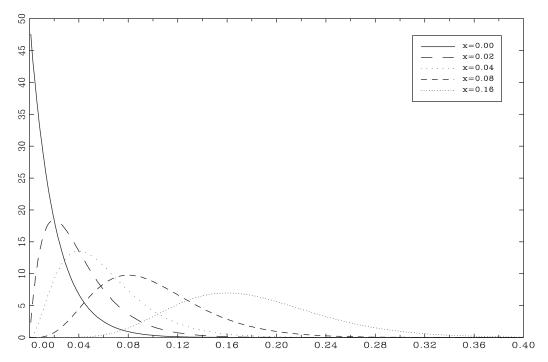
Fig. 1. Shapes of the Gamma kernel function with a fixed smoothing parameter (0.02) at various design points.

(2002), Bandi and Phillips (2003), Fan and Zhang (2003), Nicolau (2003), and Arapis and Gao (2006). An interesting situation that often arises in economics and finance is when the conditioning variables are nonnegative (i.e., they have a natural boundary at the origin) such as nominal interest rates, volatility, etc. In this case, the Nadaraya–Watson ("NW") regression estimator based on a standard symmetric kernel is not appropriate without a boundary correction for the region near the origin. This has generated a large literature on correction methods for boundary effects in nonparametric regression estimation which includes Gasser and Müller (1979), Rice (1984), Müller (1991), Fan and Gijbels (1992), Fan (1993), among others.

As a viable alternative to these boundary correction methods, Chen (2000) and Scaillet (2004) advocate the use of asymmetric kernel functions[2] that have the same support as the marginal density of the conditioning variables. In addition to eliminating boundary effects, asymmetric kernels have a number of important advantages in the analysis of economic and financial data.

First, the shape of the asymmetric kernels varies according to the position at which smoothing is made; in other words, the amount of smoothing changes in an adaptive manner. Fig. 1 plots the shapes of a Gamma kernel function for five different design points ($x = 0.00, 0.02, 0.04, 0.08$ and 0.16) at which the smoothing is performed.[3] It is worth noting that for all plotted functions, the smoothing parameter value of the Gamma kernel is fixed. In contrast, the amount of smoothing by symmetric kernels with a single bandwidth parameter is fixed everywhere. Since the distribution of interest rates is empirically characterized by a mode near the boundary and the sparse tail region, a single bandwidth does not suffice in this situation; while it is appropriate to employ a short bandwidth in the region near the boundary, a longer bandwidth is required to capture the shape of the tail. As a result, when the data indicate that the distribution contains both dense and sparse regions, it may prove useful to turn to variable bandwidth methods (e.g. Abramson, 1982; Fan and Gijbels, 1992). The adaptive smoothing property of asymmetric kernels bears some similarities to the variable bandwidth methods, but unlike these methods, the adaptive smoothing for asymmetric kernels is achieved by a single smoothing parameter which makes them much more appealing in empirical work.

Second, asymmetric kernels achieve the optimal rate of convergence (in mean integrated squared error sense) within the class of nonnegative kernel estimators. Third, unlike the case with symmetric kernels, the variances of asymmetric kernel estimators tend to decrease as the design point moves away from the boundary. This property is particularly advantageous for the analysis of interest rate data since the support of the density has a sparse region for high interest rate values. In sum, we can view asymmetric kernels as a combination of a boundary correction device and a "variable bandwidth" method.

---

[2] Strictly speaking, asymmetric kernel functions should be referred to as kernel-type weighting functions. In a slightly different context, Gouriéroux and Monfort (2006) and Jones and Henderson (2007) argue that unlike the case of symmetric kernels, the roles of the data point X and the design point x in asymmetric kernels are not exchangeable, which leads to a lack of normalization in density estimation using these kernels. Nevertheless, we follow the adopted convention in the literature for these kernel-type functions. Also, note that the lack of normalization is not an issue in regression estimation.

[3] These design points are chosen from the support of the distribution of interest rates.

Although asymmetric kernels are relatively new in the literature, several papers report favorable evidence from applying them to empirical models in economics and finance. A non-exhaustive list includes: (i) estimation of recovery rate distributions on defaulted bonds (Renault and Scaillet, 2004), (ii) income distribution estimation (Bouezmarni and Scaillet, 2005; Hagmann and Scaillet, 2007); (iii) actuarial loss distribution estimation (Gustafsson et al., 2009; Hagmann and Scaillet, 2007); (iv) hazard estimation (Bouezmarni and Rombouts, 2008); (v) regression discontinuity design (Fé, 2010); and (vi) realized integrated volatility estimation (Kristensen, 2010).

Incorporating asymmetric kernel smoothing into the nonparametric drift and diffusion estimators of Stanton (1997) is expected to shed additional light on the nonparametric estimation of spot rate diffusion models. The Monte Carlo simulations in Chapman and Pearson (2000) indicate that when the true drift is linear in the level of the spot interest rate, there are two biases in Stanton's (1997) symmetric kernel estimate of the drift function; namely, a bias near the origin and a pronounced downward bias in the region of high interest rates where the data are sparse. While the existing boundary correction methods may provide a remedy for improving the performance of the drift estimate near the boundary, they are of little help for removing the bias that occurs for high values of the spot rate. In contrast, given the appealing properties of asymmetric kernels listed above, asymmetric kernel smoothing is expected to reduce substantially both of these biases in the estimate of the drift function.

The remainder of the paper is organized as follows. Section 2 develops the asymmetric kernel estimators of the drift and diffusion functions and establishes their asymptotic properties. It also discusses the smoothing parameter selection and the implementation of a bootstrap procedure for inference. In Section 3, the proposed estimators are implemented to study the dynamics of the U.S. risk-free rate. It also reports the results from a Monte Carlo simulation experiment that examines the finite-sample performance of the symmetric and asymmetric kernel estimators of diffusion models and their practical relevance for bond and option pricing. Section 4 summarizes the main results of the paper. Assumptions and proofs are provided in the Appendix.

2. Estimation of scalar diffusion models via asymmetric kernel smoothing

2.1. NW estimators using the Gamma kernel

The estimation of scalar diffusion models plays an important role in the analysis of term structure of interest rates and derivative pricing. The general form of the underlying continuous-time process for the spot rate X_t is represented by the stochastic differential equation ("SDE")

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t,$$ (1)

where W_t is a standard Brownian motion, $\mu(X_t)$ is the drift function and $\sigma(X_t)$ is the diffusion (or instantaneous volatility) function. We assume that the initial condition X_0 is fixed, i.e. $X_0 = x_0 \in (0, \infty)$.[4]

Stanton (1997) develops a nonparametric approach to estimating the dynamics of spot interest rate driven by the scalar diffusion process (1). More specifically, Stanton (1997) uses the infinitesimal generator and a Taylor series expansion to express the first-order approximations to $\mu(X)$ and $\sigma^2(X)$ as

$$E(X_{t+\Delta} - X_t | X_t) = \mu(X_t)\Delta + o(\Delta),$$ (2)

$$E\left[(X_{t+\Delta} - X_t)^2 | X_t\right] = \sigma^2(X_t)\Delta + o(\Delta),$$ (3)

where Δ is a discrete, arbitrarily small, time step and $o(\Delta)$ denotes a remainder term which goes to 0 as $\Delta \to 0$.[5] Each of these conditional expectations at the design point x can be estimated by running a nonparametric regression. The approach by Stanton (1997) is convenient because it enables us to estimate $\mu(\cdot)$ and $\sigma^2(\cdot)$ separately, whereas the approaches by Aït-Sahalia (1996a) and Jiang and Knight (1997) require their sequential estimation; see Arapis and Gao (2006) for a comparison of these three methods.

Since the conditioning variable X_t is nonnegative, it is not appropriate to employ symmetric kernels for estimating $\mu(X)$ and $\sigma^2(X)$ without a boundary correction. The key building block for nonparametric drift and diffusion estimation that we propose in this paper is the NW kernel regression estimator based on the Gamma kernel function (Chen, 2000)[6]

$$K_{G(x/b+1,b)}(u) = \frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}1\{u \geq 0\},$$ (4)

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\exp(-y)dy, \alpha > 0$ is the Gamma function and b is the smoothing parameter. Since the density of a Gamma distribution has support $[0, \infty)$, the Gamma kernel function does not generate a boundary bias.

---

[4] This particular initialization is used in Florens-Zmirou (1993), among others. Alternative initializations are also possible.

[5] Stanton (1997) also provides higher-order approximations to $\mu(X)$ and $\sigma^2(X)$ and argues that they are more accurate. However, Fan and Zhang (2003) show that higher-order approximations tend to reduce biases at the expense of inflating variances nearly exponentially, and thus we focus only on first-order approximations.

[6] Two other asymmetric kernels, namely, the Inverse Gaussian and Reciprocal Inverse Gaussian kernels (Scaillet, 2004), can be also applied in this context, but we focus exclusively on the Gamma kernel due to its better finite-sample performance (see Gospodinov and Hirukawa, 2007).

Suppose that we observe a discrete sample $\{X_{i\Delta}\}_{i=1}^{n}$ at n equally spaced time points from the short-rate diffusion process $\{X_t : 0 \le t \le T\}$ satisfying the SDE (1). Here Δ is the step size between observations and $T = n\Delta$ is the time span of the sample. Given a design point $x > 0$, we define the Gamma-NW estimator of the drift $\mu(X)$ and diffusion function $\sigma^2(X)$ as

$$\hat{\mu}_b(x) = \frac{1}{\Delta} \frac{\sum_{i=1}^{n-1} \left(X_{(i+1)\Delta} - X_{i\Delta} \right) K_{G(x/b+1,b)}(X_{i\Delta})}{\sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})} \tag{5}$$

and

$$\hat{\sigma}_b^2(x) = \frac{1}{\Delta} \frac{\sum_{i=1}^{n-1} \left(X_{(i+1)\Delta} - X_{i\Delta} \right)^2 K_{G(x/b+1,b)}(X_{i\Delta})}{\sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})}. \tag{6}$$

2.2. Asymptotic properties of Gamma-NW estimators

In this section, we present some large sample results for the Gamma-NW estimator in the diffusion model (1). Following Bandi and Phillips (2003; abbreviated as "BP" hereafter), we explore the in-fill and long span asymptotics such that $n \to \infty$, $T \to \infty$ (long span), and $\Delta = T/n \to 0$ (in-fill). It is worth mentioning that the asymptotic results in BP are based on the assumption of symmetric, nonnegative kernel functions; see Assumption 2 in BP. None of the asymmetric kernels proposed by Chen (2000) and Scaillet (2004) can be expressed in the form of $K(X - x; b)$ for a data point X, design point x, and smoothing parameter b. Therefore, establishing the limiting theory for the Gamma estimator does not directly follow from BP and is new to the literature.

The first results are based on the assumption that the short-rate process X_t is recurrent. Let $S(x)$ denote the natural scale function of X_t which is defined, for some generic constant $c \in (0, \infty)$, as

$$S(x) = \int_c^x \exp\left[\int_c^y \left\{ -\frac{2\mu(u)}{\sigma^2(u)} \right\} du \right] dy, \tag{7}$$

and $s(x)$ be the speed function of X_t given by

$$s(x) = \frac{2}{\sigma^2(x)S'(x)}. \tag{8}$$

Because the range of X_t is $(0, \infty)$, X_t becomes recurrent if and only if $lim_{x \to 0}S(x) = -\infty$ and $lim_{x \to \infty}S(x) = \infty$ (Assumption 1(iii) in Appendix A). The full set of regularity conditions is provided in Appendix A.

In the subsequent analysis, the concept of local time plays a key role. The chronological local time of the diffusion process (1) is defined as

$$\overline{L}_X(T, x) = \frac{1}{\sigma^2(x)} \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_0^T 1_{[x, x+\epsilon)}(X_s) \sigma^2(X_s) ds \tag{9}$$

for every $x > 0$, where $1_A(\cdot)$ denotes the indicator function on the set A. $\overline{L}_X(T, x)$ is a normalized measure of the time spent by X_t in the vicinity of a generic point x. Moreover, when X_t is strictly stationary, $\overline{L}_X(T, x)/T \overset{a.s.}{\to} f(x)$, where $f(x)$ is the time-invariant marginal density of X_t; see Bosq (1998, Theorem 6.3), for instance. In contrast, for a general recurrent diffusion, $\overline{L}_X(T, x)$ diverges to infinity at a rate no faster than T. Using the Gamma kernel, $\overline{L}_X(T, x)$ for a fixed T can be (strongly) consistently estimated by

$$\hat{\overline{L}}_X(T, x, b) = \Delta \sum_{i=1}^{n} K_{G(x/b+1,b)}(X_{i\Delta}). \tag{10}$$

To describe the different asymptotic properties of the estimator depending on the position of the design point x, we denote by "interior x" and "boundary x" a design point x that satisfies $x/b \to \infty$ and $x/b \to \kappa$ for some $\kappa > 0$ as $n, T \to \infty$, respectively. Also, let "$\Rightarrow$" denote weak convergence to a random sequence. Theorem 1 states the asymptotic properties of the Gamma-NW estimator for general recurrent diffusions.

Theorem 1. *(general recurrent case)*

(i) *(Drift estimator)*
    Suppose that Assumptions 1 and 3 in Appendix A hold. In addition, if $b^{5/2}\overline{L}_X(T, x) = O_{a.s.}(1)$, then

$$\sqrt{b^{1/2}\hat{\overline{L}}_X(T, x, b)} \left\{ \hat{\mu}_b(x) - \mu(x) - B_\mu^R(x)b \right\} \Rightarrow N\left(0, \frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}} \right) \tag{11}$$

for interior x, and

$$\sqrt{b\hat{\bar{L}}_X(T,x,b)}\{\hat{\mu}_b(x)-\mu(x)\}\Rightarrow N\left(0,\frac{\Gamma(2\kappa+1)\sigma^2(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\right) \tag{12}$$

for boundary x, where $B_\mu^R(x)=\mu'(x)\{1+xs'(x)/s(x)\}+(x/2)\mu''(x)$.

(ii) *(Diffusion estimator)*
   Suppose that Assumptions 1 and 3 in Appendix A hold. In addition, if $b^{5/2}\bar{L}_X(T,x)/\Delta=O_{a.s.}(1)$, then

$$\sqrt{\frac{b^{1/2}\hat{\bar{L}}_X(T,x,b)}{\Delta}}\{\hat{\sigma}_b^2(x)-\sigma^2(x)-B_{\sigma^2}^R(x)b\}\Rightarrow N\left(0,\frac{\sigma^4(x)}{\sqrt{\pi}x^{1/2}}\right) \tag{13}$$

for interior x, and

$$\sqrt{\frac{b\hat{\bar{L}}_X(T,x,b)}{\Delta}}\{\hat{\sigma}_b^2(x)-\sigma^2(x)\}\Rightarrow N\left(0,\frac{\Gamma(2\kappa+1)\sigma^4(x)}{2^{2\kappa}\Gamma^2(\kappa+1)}\right) \tag{14}$$

for boundary x, where $B_{\sigma^2}^R(x)=\{\sigma^2(x)\}'\{1+xs'(x)/s(x)\}+(x/2)\{\sigma^2(x)\}''$.

Proof. See Appendix B.

Theorem 1 establishes the limiting behavior of the drift and diffusion function estimators when a single smoothing parameter b is used for both interior and boundary regions. For each of the results, the rate on b balances orders of magnitude in squared bias and variance for interior x. This rate undersmooths the curve for boundary x, and as a consequence, the $O(b)$ leading bias term becomes asymptotically negligible. The theorem also suggests that the diffusion estimator has a faster convergence rate than the drift estimator, regardless of the position of the design point x. This result is consistent with Remark 8 in BP. However, the chronological local time is random, and thus convergence rates of the two estimators are path-dependent (BP, Remark 4).

The existing literature often assumes stationarity or ergodicity of the process X_t. To derive the distributional theory of the Gamma-NW estimator in this case, we impose the additional condition that $\int_0^\infty s(x)dx<\infty$ (Assumption 2 in Appendix A). Using this assumption, it can be demonstrated that $\hat{\bar{L}}_X(T,x,b)/T\xrightarrow{a.s.}f(x)$ and that $s'(x)/s(x)=f'(x)/f(x)$. Accordingly, we have the following corollary.

Corollary 1. *(positive recurrent case).*

(i) *(Drift estimator)*
   Suppose that Assumptions 1, 2 and 3 in Appendix A hold. In addition, if $b=O(T^{-2/5})$, then

$$\sqrt{Tb^{1/2}}\{\hat{\mu}_b(x)-\mu(x)-B_\mu^S(x)b\}\Rightarrow N\left(0,\frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}f(x)}\right) \tag{15}$$

for interior x, and

$$\sqrt{Tb}\{\hat{\mu}_b(x)-\mu(x)\}\Rightarrow N\left(0,\frac{\Gamma(2\kappa+1)\sigma^2(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)f(x)}\right) \tag{16}$$

for boundary x, where $B_\mu^S(x)=\mu'(x)\{1+xf'(x)/f(x)\}+(x/2)\mu''(x)$.

(ii) *(Diffusion estimator)*
   Suppose that Assumptions 1, 2 and 3 in Appendix A hold. In addition, if $b=O(n^{-2/5})$, then

$$\sqrt{nb^{1/2}}\{\hat{\sigma}_b^2(x)-\sigma^2(x)-B_{\sigma^2}^S(x)b\}\Rightarrow N\left(0,\frac{\sigma^4(x)}{\sqrt{\pi}x^{1/2}f(x)}\right) \tag{17}$$

for interior x, and

$$\sqrt{nb}\{\hat{\sigma}_b^2(x)-\sigma^2(x)\}\Rightarrow N\left(0,\frac{\Gamma(2\kappa+1)\sigma^4(x)}{2^{2\kappa}\Gamma^2(\kappa+1)f(x)}\right) \tag{18}$$

for boundary x, where $B_{\sigma^2}^S(x)=\{\sigma^2(x)\}'\{1+xf'(x)/f(x)\}+(x/2)\{\sigma^2(x)\}''$.

In Corollary 1, the leading bias terms in $\hat{\mu}_b(x)$ and $\hat{\sigma}_b^2(x)$ for interior x, namely, $B_\mu^S(x)b$ and $B_{\sigma^2}^S(x)b$, are from nonparametric estimation of the conditional expectation using the Gamma kernel; on the other hand, the bias terms for boundary x become asymptotically negligible due to undersmoothing as in Theorem 1. While the bias terms are of order $O(b)$, it follows from $\Delta = T/n \rightarrow 0$ that $\hat{\mu}_b(x)$ has a slower convergence rate of order $O(T^{-1/2}b^{-1/4})$ for interior x and $O(T^{-1/2}b^{-1/2})$ for boundary x. Therefore, if the sample size is not sufficiently large, it is much harder to estimate the drift accurately, especially for design points in the region of high values where the data are sparse. However, the property of asymmetric kernel smoothing that the variance of the estimator decreases with x, is expected to provide an effective remedy.

Furthermore, Corollary 1 and the trimming argument in Chen (2000) yield the mean integrated squared error ("MISE") of each estimator given by $\text{MISE}\{\hat{\mu}_b(x)\} = O\left(b^2 + T^{-1}b^{-1/2}\right)$ and $\text{MISE}\{\hat{\sigma}_b^2(x)\} = O\left(b^2 + n^{-1}b^{-1/2}\right)$. Then, the MISE-optimal smoothing parameters are $b_\mu^* = O(T^{-2/5})$ and $b_{\sigma^2}^* = O(n^{-2/5})$ for $\hat{\mu}_b(x)$ and $\hat{\sigma}_b^2(x)$, respectively. This result suggests that a longer smoothing parameter is required to estimate the drift, as documented in Chapman and Pearson (2000, p. 367). As a consequence, the optimal MISEs are of order $\text{MISE}^*\{\hat{\mu}_b(x)\} = O\left(T^{-4/5}\right)$ and $\text{MISE}^*\{\hat{\sigma}_b^2(x)\} = O(n^{-4/5})$.

2.3. Selection of smoothing parameter

The practical implementation of the proposed nonparametric estimator requires a choice of a smoothing parameter.[7] Indeed, as argued in Chapman and Pearson (2000), the performance of the NW estimator depends crucially on the choice of a bandwidth parameter, as well as the persistence of the data. In this respect, asymmetric kernels are not an exception, even if they possess the property of adaptive smoothing with a single smoothing parameter. While the result in the previous section provides some guidance in this direction, the expression for the optimal smoothing parameter in mean squared error ("MSE") or MISE sense involves unknown quantities and a "plug-in rule" is difficult to obtain. Note also that the MSE-optimal smoothing parameter depends explicitly on the design point and, in principle, they should take different values at each x. Hagmann and Scaillet (2007), however, argue for a single, global smoothing parameter since the dependence on the design point x may deteriorate the adaptability of asymmetric kernels.

In this paper, we adopt a cross-validation ("CV") approach to choosing a global smoothing parameter for nonparametric curve estimation based on asymmetric kernels. Since the data are dependent, the leave-one-out CV is not appropriate. Instead, we work with the h-block CV version of Györfi et al. (1989) and Burman et al. (1994), where h data points on both sides of the s-th observation are removed from T observations $\{X_t\}_{t=1}^T$ and the unknown function of interest $m(X_s)$ is estimated from the remaining $T - (2h+1)$ observations. The idea behind this method is that for ergodic processes, the blocks of length h are asymptotically independent although the block size may need to shrink (at certain rate) relative to the total sample size in order to ensure the consistency of the procedure. Hall et al. (1995) establish the asymptotic properties of the h-block CV method.

In our context with equally spaced observations $\{X_{i\Delta}\}_{i=1}^n$, let $\hat{m}_{-(i-h)\Delta:(i+h)\Delta}(X_{i\Delta})$ denote the estimate of the drift $\mu(X_{i\Delta})$ or the diffusion $\sigma^2(X_{i\Delta})$ from $n - (2h+1)$ observations $\{X_\Delta, X_{2\Delta}, ..., X_{(i-h-1)\Delta}, X_{(i+h+1)\Delta}, ..., X_{n\Delta}(=X_T)\}$. Then, the smoothing parameter can be selected by minimizing the least squares cross-validation function

$$\text{CV}(b) = \arg\min_{b \in B} \sum_{i=h+1}^{n-h} \left\{ Y_{i\Delta} - \hat{m}_{-(i-h)\Delta:(i+h)\Delta}(X_{i\Delta}) \right\}^2 \psi(X_{i\Delta}), \tag{19}$$

where $Y_{i\Delta} = (X_{(i+1)\Delta} - X_{i\Delta})/\Delta$ and $(X_{(i+1)\Delta} - X_{i\Delta})^2/\Delta$ when estimating the drift and diffusion functions, respectively, $\mathcal{B}$ is a predetermined range of b, and $\psi(\cdot)$ is a weighting function that has compact support and is bounded by 1. In our numerical work, we set $\psi(\cdot) \equiv 1$. This procedure could be further extended to the modified (asymptotically optimal) h-block CV method proposed by Racine (2000).

Since formal results on the optimal selection of the block size h are not available in the literature, we propose an automatic, data-dependent choice of h that takes into account the persistence of the data. While Burman et al. (1994) and Racine (2000) adopt an ad hoc fixed fraction rule ($h = n^{1/4}$), this choice of h does not account for different degrees of persistence in the underlying process. Given some similarities between the choice of h and the selection of a bandwidth parameter in the heteroskedasticity and autocorrelation consistent ("HAC") variance estimation, we follow Andrews (1991) and modify the plug-in-rule for h to $h = (\gamma n)^{1/4}$, where $\gamma = \frac{4\rho^2}{(1-\rho)^2(1+\rho)^2}$ and $\rho < 1$ is the AR coefficient from an AR(1) regression for $\{X_{i\Delta}\}_{i=1}^n$.[8] When $\gamma = 0$ (or equivalently $\rho = 0$), this procedure naturally reduces to the leave-one-out CV for serially uncorrelated data.

2.4. Bootstrap procedure

Inference on the conditional expectations $\mu(X_t)$ and $\sigma^2(X_t)$ can be conducted using the asymptotic distributions derived in Theorem 1 or Corollary 1 above. This involves explicit estimation of unknown functions of the data that enter the asymptotic distributions. A more practical approach, that typically provides better approximations to the finite-sample distributions of the

---

[7] Bandwidth selection for kernel estimators of continuous-time diffusion models is an on-going research program. Some recent contributions tailored to continuous-time Markov processes include Bandi et al. (2010) and Kanaya and Kristensen (2011).

[8] In practice, ρ is replaced by its least squares estimate.

statistics of interest, employs simulation-based or bootstrap methods.[9] In particular, we propose a parametric bootstrap procedure based on the discretized version of model (1)

$$X_{(i+1)\Delta} - X_{i\Delta} = \mu(X_{i\Delta})\Delta + \sigma(X_{i\Delta})\sqrt{\Delta}\varepsilon_{i\Delta}, \tag{20}$$

where $\varepsilon_{i\Delta} \overset{iid}{\sim} N(0,1)$. Our bootstrap method also deals explicitly with the discretization bias induced by the discrete-time version of the underlying model. This is achieved by directly sampling from the continuous-time model with a step size set to $\triangle/M$ for some $M \geq 1$ (in our empirical application, $M = 100$) while keeping the design points the same across the bootstrap samples.[10] More specifically, using the estimated drift and diffusion functions $\hat{\mu}_b(x)$ and $\hat{\sigma}_b(x)$,[11] we set $x_0^* = X_{i\Delta}$ and compute recursively

$$x_k^* = x_{k-1}^* + \hat{\mu}_b(x_{k-1}^*)\frac{\Delta}{M} + \hat{\sigma}_b(x_{k-1}^*)\sqrt{\frac{\Delta}{M}}\varepsilon_k^*, \tag{21}$$

for $k = 1, \ldots M$, where $\varepsilon_k^* \overset{iid}{\sim} N(0,1)$. We then set $X_{(i+1)\Delta}^* = x_M^*$ and repeat this procedure for each design point to construct the bootstrap sample $\left\{ X_{(i+1)\Delta}^* \right\}_{i=1}^n$. Finally, each bootstrap series $X_{(i+1)\Delta}^* - X_{i\Delta}^*$ is used to re-estimate $\mu(X)$ and $\sigma(X)$ at the same design points.

Let $\mu_b^*(x)$ and $\sigma_b^*(x)$ denote the corresponding bootstrap estimates. Repeating this procedure a large number of times provides an approximation to the distributions of $\mu_b^*(x) - \hat{\mu}_b(x)$ and $\sigma_b^*(x) - \hat{\sigma}_b(x)$ at each design point. Then, pointwise confidence bands can be constructed as $\left[\hat{\mu}_b(x) - q_\mu^*(1 - \frac{\alpha}{2}), \hat{\mu}_b(x) - q_\mu^*(\frac{\alpha}{2}) \right]$ and $\left[\hat{\sigma}_b(x) - q_\sigma^*(1 - \frac{\alpha}{2}), \hat{\sigma}_b(x) - q_\sigma^*(\frac{\alpha}{2}) \right]$, where $q_\mu^*(\alpha)$ and $q_\sigma^*(\alpha)$ denote the α-quantiles of the bootstrap distributions of $\mu_b^*(x) - \hat{\mu}_b(x)$ and $\sigma_b^*(x) - \hat{\sigma}_b(x)$. This method can also be used for bias correction of the original estimates $\hat{\mu}_b(x)$ and $\hat{\sigma}_b(x)$ which could further improve the accuracy of the approximation in a double bootstrap procedure.

3. Empirical and simulation analysis

3.1. Dynamics of U.S. risk-free rate

One outstanding question in the analysis of the term structure of U.S. interest rates is concerned with the presence of possible nonlinearities in the underlying dynamics of the spot rate. Despite the large literature on this issue (Aït-Sahalia, 1996a,b; Stanton, 1997; Chapman and Pearson, 2000; Bandi, 2002; Durham, 2003; Fan and Zhang, 2003; Jones, 2003; Arapis and Gao, 2006; among others), there is still no consensus on the presence of statistically significant nonlinearities in short-term interest rates and their economic importance for pricing bonds and interest rate derivative products. For instance, Chapman and Pearson (2000) argue that the nonlinearity in the drift of the spot rate at high values of interest rates documented by Stanton (1997) could be spurious due to the poor finite-sample properties of Stanton's (1997) estimator. Fan and Zhang (2003) conclude that there is little evidence against linearity in the short rate drift function. Similarly, Bandi (2002), Durham (2003) and Jones (2003) find no empirical support for nonlinear mean reversion in short-term rates. In contrast, Arapis and Gao (2006) report that their specification testing strongly rejects the linearity of the short rate drift at both daily and monthly frequency.

Here, we revisit this issue by estimating the drift function using our proposed Gamma kernel estimator. The data used for estimation are 564 monthly observations for the annualized risk-free rate from the Fama files in the CRSP U.S. Government Bond Files covering the period January 1952–December 1998. As it is typically the case for interest rate data, the U.S. risk-free rate exhibits high persistence and strong conditional heteroskedasticity. Given the sensitivity of the nonparametric estimates of the drift function to the bandwidth (Chapman and Pearson, 2000), this high persistence should be properly incorporated into the selection procedure for a smoothing parameter.

Since our interest lies in identifying possible nonlinearities in the drift function of the spot rate, it proves instructive to estimate the CIR (Cox et al., 1985) model

$$dX_t = \kappa(\theta - X_t)dt + \sigma X_t^{1/2}dW_t, \tag{22}$$

which specifies the drift as a linear function in X of the form $\mu(X_t) = \kappa(\theta - X_t)$, where κ denotes the speed of mean reversion and θ is the long-run mean of X. Given the nature of the observed data for the spot rate, the unknown parameters (κ, θ, σ) are estimated from the discretized version of the model with $\Delta = 1/12$ which induces a discretization bias. To correct for the discretization bias, we follow Broze et al. (1995) and estimate the parameters of the CIR model by indirect inference using the same auxiliary model and discretization step as suggested by Broze et al. (1995). The resulting estimates for the U.S. risk-free rate are $\hat{\kappa} = 0.2804$, $\hat{\theta} = 0.0541$ and $\hat{\sigma} = 0.0876$.

Fig. 2 presents the nonparametric estimates of the drift function for the risk-free rate from the Gamma NW and Gaussian NW regressions along with the estimated parametric drift function from the CIR model. The smoothing parameter for the Gamma

---

[9] Bootstrap methods for inference in nonparametric regression models have been developed by Härdle and Marron (1991) and Franke et al. (2002), among others.

[10] We would like to thank an anonymous referee for suggesting this to us.

[11] The diffusion estimates are obtained by imposing the restriction $\sigma(0) = 0$ as in Stanton (1997) which substantially reduces the bias documented in Fan and Zhang (2003) and ensures the nonnegativity of the simulated interest rates as $\triangle \to 0$.

kernel estimator is chosen by *h*-block CV described in Section 2.3 while the smoothing parameter for the Gaussian kernel is set equal to four times the bandwidth for *iid* data as in Stanton (1997). The differences in the nonparametric estimates reveal that part of the accelerated mean reversion at higher interest rates, reported in the previous literature, may be due to larger biases of the symmetric kernel-based estimators although the Gamma kernel estimator also suggests that the mean reversion becomes stronger for interest rates larger than 11–12%. To assess the statistical significance of these findings, we also report 95% confidence bands for the Gamma kernel estimates obtained by the bootstrap procedure outlined in Section 2.4 with 999 bootstrap replications. Despite the mild nonlinearity of the nonparametric estimates, the parametric estimate of the drift function from the CIR model falls within the 95% confidence bands (with the exception of very low interest rates) and it appears that the null hypothesis of a linear drift cannot be rejected for this particular sample. Also, the nonparametric estimate based on the Gaussian kernel lies inside the 95% Gamma kernel-based confidence bands except for very large interest rates where the bias due to sparseness of the data becomes more pronounced. The Gamma kernel estimate of the diffusion function and the associated 95% bootstrap confidence bands are plotted in Fig. 3. To gain further insights into the finite-sample properties of the nonparametric estimators and the economic importance of these seemingly small statistical differences for bond and option pricing, we conduct a simulation experiment whose design closely mimics our empirical application.

3.2. Monte Carlo experiment

The data for the simulation experiment is generated from the CIR model (22) with true parameter values that are set equal to the estimates from U.S. data in the previous section, $(\kappa, \theta, \sigma) = (0.2804, 0.0541, 0.0876)$, and a time step between two consecutive observation equal to $\Delta = 1/12$ corresponding to monthly data. The CIR model is convenient because the transition and marginal densities are known and the bond and call option prices are available in closed form (Cox et al., 1985). 5000 sample paths for the spot interest rate of 600 observations ($n = 600, T = 50$) are constructed recursively by drawing random numbers from the transition non-central chi-square density and using the values for κ, θ, σ and Δ (see Chapman and Pearson, 2000, for more details).

The expressions for the price of a zero-coupon discount bond and a call option on a zero-coupon discount bond have an analytical form and are given in Cox et al. (1985). We follow Jiang (1998) and Phillips and Yu (2005) and compute the prices of a three-year zero-coupon discount bond and a one-year European call option on a three-year discount bond with a face value of $100 and an exercise price of $87 with an initial interest rate of 7% by simulating spot rate data from the estimated diffusion process. The simulated bond and derivative prices are then compared to the analytical prices based on the true values of the parameters.
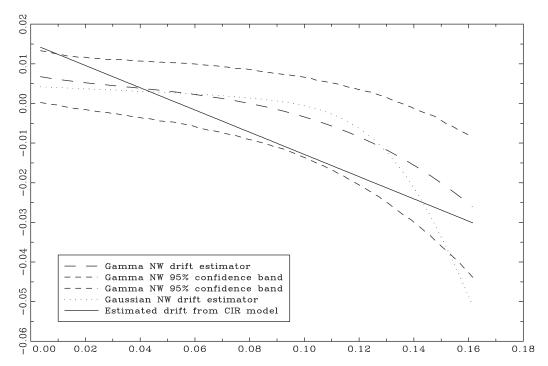

Fig. 2. Estimates of the drift function $\mu(\cdot)$ of the risk-free rate with the NW estimator based on the Gamma kernel (along with 95% bootstrap confidence bands), NW estimator based on the Gaussian kernel and indirect inference estimator of the CIR model. The smoothing parameter for the Gamma kernel estimator is chosen by *h*-block cross-validation with $h = (\gamma n)^{1/4}$, where γ is defined in Section 2.3. The smoothing parameter for the Gaussian kernel estimator is set equal to four times the bandwidth parameter for *iid* data (Stanton, 1997).
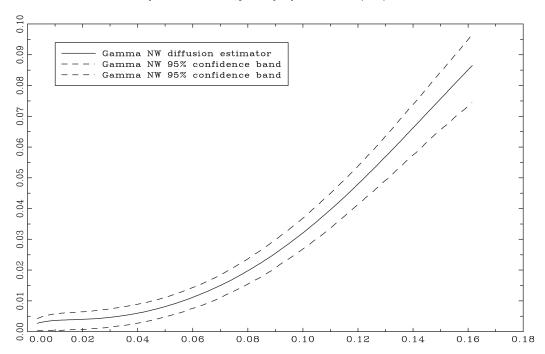
Fig. 3. Estimate of the diffusion function $\sigma(\cdot)$ of the risk-free rate with the NW estimator based on the Gamma kernel (along with 95% bootstrap confidence bands). The smoothing parameter for the Gamma kernel estimator is chosen by h-block cross-validation with $h = (\gamma n)^{1/4}$, where γ is defined in Section 2.3.

More specifically, the price of a zero-coupon bond with face value P_0 and maturity $(\tau - t)$ is computed as

$$P_t^\tau = P_0 E_t \left[\exp\left(- \int_t^\tau \tilde{X}_u du\right) \right], \tag{23}$$

where the risk-neutral process $\tilde{X}_t$ evolves as $d\tilde{X}_t = \hat{\mu}\left(\tilde{X}_t\right) dt + \hat{\sigma}\left(\tilde{X}_t\right) dW_t$, and $\hat{\mu}(\cdot)$ and $\hat{\sigma}(\cdot)$ denote the nonparametric estimates of the drift and diffusion functions, respectively.[12] The expectation is evaluated by Monte Carlo simulation using a discretized version of the dynamics of the spot rate.

The price of a call option with maturity $(m-t)$ on a zero-coupon bond with maturity $(\tau - t)$, face value P_0 and exercise price K is computed as

$$
\begin{aligned}
C_t^m &= E_t \left[\exp\left(- \int_t^m \tilde{X}_u du\right) \max(P_m^\tau - K, 0) \right] \\
&= E_t \left[\exp\left(- \int_t^m \tilde{X}_u du\right) \max\left(P_0 E_m \left[\exp\left(- \int_m^\tau \tilde{X}_u dv\right) \right] - K, 0 \right) \right],
\end{aligned} \tag{24}
$$

where $m < \tau$ and sample paths for $\tilde{X}_t$ are simulated from the nonparametrically estimated discretized model of spot rate.

We consider the NW estimators based on Gaussian and Gamma kernels with smoothing parameters that are selected by h-block CV. The average Monte Carlo estimates and 95% Monte Carlo confidence bands for the drift function are plotted in Figs. 4 and 5. While the Gamma NW kernel estimator is practically unbiased, the Gaussian NW kernel exhibits a downward bias for interest rates higher than 9–10%. Furthermore, the asymmetric kernel estimator exhibits smaller variability both near the boundary and in the tail region where the data are sparse.[13]

The economic significance of the improved estimation of diffusion models of spot rate is evaluated by comparing bond and option pricing errors based on symmetric and asymmetric kernel estimators[14] for the CIR model. The results are presented in Table 1. The bond and derivative prices based on the asymmetric kernel estimator are less biased than its Gaussian counterpart which could translate in potentially substantial financial benefits and investment profits for large portfolios. Another important feature of the results is that the Gamma-based bond and option prices enjoy much smaller variability and tighter confidence

---

[12] For simplicity, the market price of risk is assumed to be equal to zero since its computation requires another interest rate process of different maturity.

[13] The estimates of the diffusion function $\sigma(\cdot)$ are not reported to conserve space. Overall, the diffusion function estimates based on the Gamma and Gaussian kernels are similar due to the zero restriction of the diffusion function at the origin (see footnote 11). When this restriction is removed, the Gamma kernel estimator tends to perform better as a result of its intrinsic boundary correction property.

[14] Kristensen (2008) demonstrates that when the process X_t is stationary, derivative prices obtained by plug-in nonparametric estimates of the drift and diffusion functions have the same rate of convergence as parametric estimators.
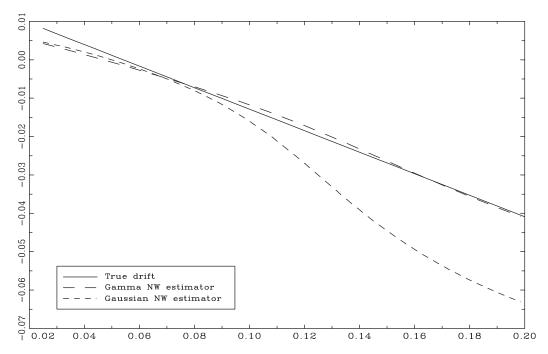
Fig. 4. Average Monte Carlo drift (Gamma and Gaussian NW) estimates from CIR model with $(\kappa, \theta, \sigma) = (0.2804, 0.0541, 0.0876)$. The smoothing parameters are selected by h-block cross validation with $h = (\gamma n)^{1/4}$, where γ is defined in Section 2.3.

intervals than the symmetric kernel-based option prices. Finally, while the confidence intervals for the Gamma kernel-based call prices appear symmetric around the median estimate, the confidence intervals for the Gaussian kernel-based call prices tend to be highly asymmetric (with long right tail) which illustrates the economic impact of the boundary problems of symmetric kernels.
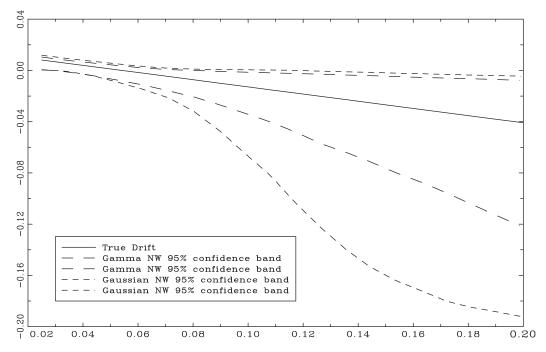

Fig. 5. 95% Monte Carlo confidence intervals of the drift (Gamma and Gaussian NW) estimates from CIR model with $(\kappa, \theta, \sigma) = (0.2804, 0.0541, 0.0876)$ and smoothing parameters selected by h-block cross validation with $h = (\gamma n)^{1/4}$, where γ is defined in Section 2.3.

Table 1
Monte Carlo statistics of bond and option prices based on nonparametric (Gaussian NW and Gamma NW) estimators in the CIR model with $\kappa = 0.2804$, $\theta = 0.0541$ and $\sigma = 0.0876$.

	Bond price	Call option price
True price	82.425	1.889
Gaussian NW estimator		
Median estimate	82.359	1.656
Standard deviation	1.322	0.515
95% confidence interval	[80.420, 85.573]	[1.014, 3.026]
Gamma NW estimator		
Median estimate	82.447	1.704
Standard deviation	1.115	0.347
95% confidence interval	[80.665, 85.058]	[1.133, 2.463]

Notes: The statistics in the table are computed from 5000 samples generated from the CIR model with $\triangle = 1/12$ and $n = 600$. The prices of a three-year zero-coupon discount bond and a one-year European call option on a three-year bond with face value of $100, strike price of $87 and initial interest rate of 7% are computed by Monte Carlo simulation.

4. Conclusion

This paper proposes an asymmetric kernel estimator for estimating the underlying scalar diffusion model of spot interest rate. The asymptotic properties of the Gamma kernel estimator of the drift and diffusion functions are established for both interior and boundary design points. We show that the Gamma kernel estimator possesses some appealing properties such as lack of boundary bias and adaptability in the amount of smoothing. The paper adopts a block cross-validation method for dependent data in choosing the smoothing parameter. The finite-sample performance of the asymmetric kernel estimator and its practical relevance for bond and option pricing are evaluated using simulated data of spot interest rate. The empirical analysis of the U.S. risk-free rate reveals the existence of some mild (and statistically insignificant) nonlinearity in the drift function although the estimated increased mean reversion at higher interest rate levels is substantially smaller than originally reported by Stanton (1997).

Appendix A. Assumptions

This appendix provides a set of assumptions for Theorem 1 and Corollary 1 that are similar to those in Bandi (2002) and BP. Our first assumption is basically the same as Assumption 1 in BP, which ensures that the SDE (1) has a unique strong solution X_t and that X_t is recurrent. Assumption 2, together with Assumption 1, implies that the process X_t is positive recurrent (or ergodic) and ensures the existence of a time-invariant distribution P^0 with density $f(x) = s(x) / \int_0^\infty s(x) dx$. Moreover, if X_0 has distribution P^0, X_t becomes strictly stationary. Assumption 3 controls the rates of convergence or divergence of the sequences used in the asymptotic results for general and positive recurrent cases.

Assumption 1.

(i) $\mu(\cdot)$ and $\sigma(\cdot)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $(0, \infty)$, where $\mathfrak{B}$ is the σ-field generated by Borel sets on $(0, \infty)$. 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 $(0, \infty)$, there exist constants C_1^J and C_2^J such that, for all $x, y \in J$, $|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1^J |x - y|$ and $|\mu(x)| + |\sigma(x)| \leq C_2^J \{1 + |x|\}$.

(ii) $\sigma^2(\cdot) > 0$ on $(0, \infty)$.

(iii) The natural scale function $S(x)$ satisfies $\lim_{x \to 0} S(x) = -\infty$ and $\lim_{x \to \infty} S(x) = \infty$.

Assumption 2. The speed function $s(x)$ satisfies $\int_0^\infty s(x) dx < \infty$.

Assumption 3. As $n, T \to \infty$, $\Delta (= T/n) \to 0$, $b (= b_{n,T}) \to 0$ such that

$$\frac{\bar{L}_X(T, x)}{b} \sqrt{\Delta \log\left(\frac{1}{\Delta}\right)} = o_{a.s.}(1). \tag{A1}$$

Appendix B. Preliminary lemma and Proof of Theorem 1

The Proof of Theorem 1 closely follows the proof of Theorem 3 in BP by adapting it to the case of the Gamma kernel. We provide only the proof for the drift estimator $\hat{\mu}_b(x)$, because the one for the diffusion estimator $\hat{\sigma}_b^2(x)$ follows similar arguments. To establish the results in Theorem 1, we need the following lemma.

Lemma 1. *Under Assumptions 1 and 3 in Appendix A,*

$$\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\mu(X_{i\Delta}) = \int_0^T K_{G(x/b+1,b)}(X_s)\mu(X_s)ds + o_{a.s.}(1). \tag{B1}$$

Proof of Lemma 1. Consider that

$$\left| \Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\mu(X_{i\Delta}) - \int_0^T K_{G(x/b+1,b)}(X_s)\mu(X_s)ds \right|$$

$$\leq \left| \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} \left\{ K_{G(x/b+1,b)}(X_{i\Delta}) - K_{G(x/b+1,b)}(X_s) \right\}\mu(X_{i\Delta})ds \right| \tag{B2}$$

$$+ \left| \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} K_{G(x/b+1,b)}(X_s)\{\mu(X_s) - \mu(X_{i\Delta})\}ds \right| + \left| \Delta K_{G(x/b+1,b)}(X_0)\mu(X_0) \right|$$

$$\equiv A_1(x) + A_2(x) + O_{a.s.}(\Delta),$$

where the bound of the third term is obtained from $X_0 = x_0$, Assumption 1 and boundedness of the kernel. Observe that

$$A_1(x) \leq \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} \left| K'_{G(x/b+1,b)}\left(\tilde{X}_{is}\right) \right| |X_s - X_{i\Delta}||\mu(X_{i\Delta})|ds \tag{B3}$$

for $\tilde{X}_{is}$ on the line segment connecting X_s and $X_{i\Delta}$. Since

$$k_{n,T} \equiv \max_{i \leq n} \sup_{i\Delta \leq s \leq (i+1)\Delta} |X_s - X_{i\Delta}| = O_{a.s.}\left(\sqrt{\Delta \log\left(\frac{1}{\Delta}\right)} \right) \tag{B4}$$

(see p. 267 in BP), $\tilde{X}_{is} = X_s + o_{a.s.}(1)$ uniformly over $i \leq n$, and thus

$$A_1 \leq k_{n,T} \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} \left| K'_{G(x/b+1,b)}(X_s + o_{a.s.}(1)) \right| |\mu(X_s + o_{a.s.}(1))|ds$$

$$= k_{n,T} \int_0^T \left| K'_{G(x/b+1,b)}(X_s + o_{a.s.}(1)) \right| |\mu(X_s + o_{a.s.}(1))|ds \tag{B5}$$

$$= \frac{k_{n,T}}{b} \int_0^\infty b \left| K'_{G(x/b+1,b)}(u + o_{a.s.}(1)) \right| |\mu(u)| \bar{L}_X(T,u)du.$$

Note that

$$K'_{G(x/b+1,b)}(u) = \left(\frac{x}{b}\right)\frac{u^{x/b-1}exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)} - \left(\frac{1}{b}\right)\frac{u^{x/b}exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}. \tag{B6}$$

Then, by the properties of the Gamma function,

$$\int_0^\infty b \left| K'_{G(x/b+1,b)}(u) \right| du = x\int_0^\infty \frac{u^{x/b-1}exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}du + \int_0^\infty \frac{u^{x/b}exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}du$$

$$= x\frac{b^{x/b}\Gamma(x/b)}{b^{x/b+1}\Gamma(x/b+1)} + 1 = 1 + 1 = 2. \tag{B7}$$

Combining this result with the continuity of $\mu(\cdot)$ and $\bar{L}_X(T,\cdot)$ and Assumption 3, we can conclude that A_1 is bounded by $O_{a.s.}\left(\frac{\bar{L}_X(T,x)}{b}\sqrt{\Delta \log\left(\frac{1}{\Delta}\right)}\right) = o_{a.s.}(1)$. Similarly, $A_2(x)$ is bounded by $O_{a.s.}\left(\frac{\bar{L}_X(T,x)}{b}\sqrt{\Delta \log\left(\frac{1}{\Delta}\right)}\right) = o_{a.s.}(1)$. ∎

Proof of Theorem 1. Observe that

$$
\mu_b(x) - \mu(x) = \frac{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\{\mu(X_{i\Delta}) - \mu(x)\}}{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})}
$$

$$
+ \frac{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\left\{\frac{X_{(i+1)\Delta} - X_{i\Delta}}{\Delta} - \mu(X_{i\Delta})\right\}}{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})}
\tag{B8}
$$

$$
\equiv B_\mu(x) + V_\mu(x).
$$

Using Lemma 1 above and Lemma 6 in BP, we have

$$
B_\mu(x) = \frac{\int_0^T K_{G(x/b+1,b)}(X_s)\{\mu(X_s) - \mu(x)\}ds + o_{a.s.}(1)}{\int_0^T K_{G(x/b+1,b)}(X_s)ds + o_{a.s.}(1)}
$$

$$
= \frac{\int_0^\infty K_{G(x/b+1,b)}(u)\{\mu(u) - \mu(x)\}s(u)du + o_{a.s.}(1)}{\int_0^\infty K_{G(x/b+1,b)}(u)s(u)du + o_{a.s.}(1)} + (s.o.),
\tag{B9}
$$

where $(s.o.)$ denotes the term that is of smaller order than the first term. Ignoring smaller order terms, taking a second-order Taylor expansion for $\mu(u)$ around $u = x$, and using $E(\theta_x - x) = b$ and $E(\theta_x - x)^2 = b(x + 2b)$ for $\theta_x \sim \text{Gamma}(x/b + 1, b)$, we obtain the following approximation of the first term:

$$
\frac{\int_0^\infty K_{G(x/b+1,b)}(u)\{\mu(u) - \mu(x)\}s(u)du}{\int_0^\infty K_{G(x/b+1,b)}(u)s(u)du} = \left[\mu'(x)\left\{1 + \frac{xs'(x)}{s(x)}\right\} + \frac{x}{2}\mu''(x)\right]b + o(b).
\tag{B10}
$$

On the other hand, $V_\mu(x)$ can be rewritten as

$$
V_\mu(x) = \frac{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\{\mu(X_s) - \mu(X_{i\Delta})\}ds}{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})}
$$

$$
+ \frac{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\sigma(X_s)dWs}{\Delta \sum_{i=1}^{n-1} K_{G(x/b+1,b)}(X_{i\Delta})}
\tag{B11}
$$

$$
\equiv V_1(x) + V_2(x),
$$

where $V_1(x)$ is bounded by $O_{a.s.}\left(\sqrt{\Delta log(\frac{1}{\Delta})}\right) = o_{a.s.}(1)$ following the same argument as in the Proof of Lemma 1 above. The leading variance term of the numerator of $V_2(x)$ can be determined by

$$
\Delta \sum_{i=0}^{n-1} K_{G(x/b+1,b)}^2(X_{i\Delta})\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\sigma^2(X_s)ds + o_{a.s.}(1)
$$

$$
= \int_0^T K_{G(x/b+1,b)}^2(X_s + o_{a.s.}(1))\left\{\sigma^2(X_s) + o_{a.s.}(1)\right\}ds + o_{a.s.}(1).
\tag{B12}
$$

Ignoring smaller order terms, and defining $A_b(x) = b^{-1}\Gamma(2x/b + 1)/\{2^{2x/b+1}\Gamma^2(x/b + 1)\}$, we have

$$
\int_0^T K_{G(x/b+1,b)}^2(X_s)\sigma^2(X_s)ds
$$

$$
= A_b(x)\int_0^T K_{G(2x/b+1,b/2)}(X_s)\sigma^2(X_s)ds
\tag{B13}
$$

$$
= A_b(x)\int_0^T K_{G(2x/b+1,b/2)}(u)\sigma^2(u)\bar{L}_X(T,u)du,
$$

where (see Chen, 2000)

$$
A_b(x) = \begin{cases} \dfrac{b^{-1/2}}{2\sqrt{\pi}x^{1/2}} + o\left(b^{-1/2}\right) & \text{for interior } x \\ \dfrac{b^{-1}\Gamma(2\kappa + 1)}{2^{2\kappa+1}\Gamma^2(\kappa + 1)} + o\left(b^{-1}\right) & \text{for boundary } x. \end{cases}
\tag{B14}
$$

Hence, for interior x,

$$b^{1/2} \int_0^T K_{G(x/b+1,b)}^2(X_s) \sigma^2(X_s) ds \overset{a.s.}{\to} \frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}} \bar{L}_X(T,x). \tag{B15}$$

It follows from the arguments in the proof of Theorem 3 in BP that

$$b^{1/4} V_2(x) \Rightarrow MN\left(0, \frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}\bar{L}_X(T,x)}\right). \tag{B16}$$

Therefore,

$$\sqrt{b^{1/2}\bar{L}_X(T,x)} V_2(x) \Rightarrow N\left(0, \frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}}\right). \tag{B17}$$

The result for interior x can be established by combining (B8), (B10) and (B17) under the assumption that $b^{5/2}\bar{L}_X(T,x) = O_{a.s.}(1)$ and then replacing $\bar{L}_X(T,x)$ by $\hat{\bar{L}}_X(T,x,b)$. The result for boundary x can be obtained similarly. ∎

References

Abramson, I.S., 1982. On bandwidth variation in kernel estimates — a square root law. Ann. Stat. 10, 1217–1223.
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. Rev. Financ. Stud. 9, 385–426.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Arapis, M., Gao, J., 2006. Empirical comparisons in short-term interest rate models using nonparametric methods. J. Financ. Econ. 4, 310–345.
Bandi, F.M., 2002. Short-term interest rate dynamics: a spatial approach. J. Financ. Econ. 65, 73–110.
Bandi, F.M., Phillips, P.C.B., 2003. Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241–283.
Bandi, F.M., V. Corradi, and Moloche, G. (2010): "Bandwidth selection for continuous-time Markov processes," Unpublished manuscript.
Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, 2nd edition. Springer-Verlag, New York.
Bouezmarni, T., Rombouts, J.V.K., 2008. Density and hazard rate estimation for censored and α-mixing data using Gamma kernels. J. Nonparametric Stat. 20, 627–643.
Bouezmarni, T., Scaillet, O., 2005. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. Econ. Theory 21, 390–412.
Broze, L., Scaillet, O., Zakoïan, J.-M., 1995. Testing for continuous-time models of the short-term interest rate. J. Empir. Finance 2, 199–223.
Burman, P., Chow, E., Nolan, D., 1994. A cross-validatory method for dependent data. Biometrika 81, 351–358.
Chapman, D.A., Pearson, N.D., 2000. Is the short rate drift actually nonlinear? J. Finance 55, 355–388.
Chen, S.X., 2000. Probability density function estimation using Gamma kernels. Ann. Inst. Stat. Math. 52, 471–480.
Cox, J.C., Ingersoll Jr., J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.
Durham, G.B., 2003. Likelihood-based specification analysis of continuous-time models of the short-term interest rate. J. Financ. Econ. 70, 463–487.
Fan, J., 1993. Local linear regression smoothers and their minimax efficiencies. Ann. Stat. 21, 196–216.
Fan, J., Gijbels, I., 1992. Variable bandwidth and local linear regression smoothers. Ann. Stat. 20, 2008–2036.
Fan, J., Zhang, C., 2003. A reexamination of diffusion estimators with applications to financial model validation. J. Am. Stat. Assoc. 98, 118–134.
Fé, D., 2010. An application of local linear regression with asymmetric kernels to regression discontinuity designs. Economics Discussion Paper Series EDP-1016. University of Manchester.
Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. J. Appl. Probab. 30, 790–804.
Franke, J., Kreiss, J.-P., Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. Bernoulli 8, 1–37.
Gasser, T., Müller, H.-G., 1979. Kernel estimation of regression functions. In: Gasser, T., Rosenblatt, M. (Eds.), Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2–4, 1979. Springer-Verlag, Berlin, pp. 23–68.
Gospodinov, N., and M. Hirukawa (2007): "Time series nonparametric regression using asymmetric kernels," CIRJE Discussion Paper No. CIRJE-F-573, CIRJE, Faculty of Economics, University of Tokyo.
Gouriéroux, C., and Monfort, A. (2006): "(Non)Consistency of the beta kernel estimator for recovery rate distribution," CREST Discussion Paper n° 2006-31.
Gustafsson, J., Hagmann, M., Nielsen, J.P., Scaillet, O., 2009. Local transformation kernel density estimation of loss distributions. J. Bus. Econ. Stat. 27, 161–175.
Györfi, L., Härdle, W., Sarda, P., Vieu, P., 1989. Nonparametric Curve Estimation from Time Series, Lecture Notes in Statistics. Springer-Verlag, Berlin.
Hagmann, M., Scaillet, O., 2007. Local multiplicative bias correction for asymmetric kernel density estimators. J. Econ. 141, 213–249.
Hall, P., Lahiri, S.N., Polzehl, J., 1995. On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. Ann. Stat. 23, 1921–1936.
Härdle, W., Marron, J.S., 1991. Bootstrap simultaneous error bars for nonparametric regression. Ann. Stat. 19, 778–796.
Jiang, G.J., 1998. Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. J. Financ. Quant. Anal. 33, 465–497.
Jiang, G.J., Knight, J.L., 1997. A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. Econ. Theory 13, 615–645.
Jones, C.S., 2003. Nonlinear mean reversion in the short-term interest rate. Rev. Financ. Stud. 16, 793–843.
Jones, M.C., Henderson, D.A., 2007. Kernel-type density estimation on the unit interval. Biometrika 24, 977–984.
Kanaya, S., Kristensen, D., 2011. "Optimal sampling and bandwidth selection for nonparametric kernel estimators of diffusion processes," Unpublished manuscript.
Kristensen, D., 2008. Estimation of partial differential equations with applications in finance. J. Econ. 144, 392–408.
Kristensen, D., 2010. Nonparametric filtering of the realized spot volatility: a kernel-based approach. Econ. Theory 26, 60–93.
Müller, H.-G., 1991. Smooth optimum kernel estimators near endpoints. Biometrika 78, 521–530.

Nicolau, J., 2003. Bias reduction in nonparametric diffusion coefficient estimation. Econ. Theory 19, 754–777.

Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. Rev. Financ. Stud. 18, 707–742.

Racine, J.S., 2000. A consistent cross-validatory method for dependent data: *hv*-block cross-validation. J. Econ. 99, 39–61.

Renault, O., Scaillet, O., 2004. On the way to recovery: a nonparametric bias free estimation of recovery rate densities. J. Bank. Financ. 28, 2915–2931.

Rice, J., 1984. Boundary modification for kernel regression. Commun. Stat. Theory Meth. 13, 893–900.

Scaillet, O., 2004. Density estimation using inverse and reciprocal inverse Gaussian kernels. J. Nonparametric Stat. 16, 217–226.

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