CIRJE-F-573

Time Series Nonparametric Regression Using Asymmetric Kernels with an Application to Estimation of Scalar Diffusion Processes

Nikolay Gospodinov
Concordia University and CIREQ

Masayuki Hirukawa
Northern Illinois University

June 2008

Time Series Nonparametric Regression Using Asymmetric Kernels with an Application to Estimation of Scalar Diffusion Processes*

Nikolay Gospodinov† — Concordia University and CIREQ

Masayuki Hirukawa‡ — Northern Illinois University

December 2007

Abstract

This paper considers a nonstandard kernel regression for strongly mixing processes when the regressor is nonnegative. The nonparametric regression is implemented using asymmetric kernels [Gamma (Chen, 2000b), Inverse Gaussian and Reciprocal Inverse Gaussian (Scaillet, 2004) kernels] that possess some appealing properties such as lack of boundary bias and adaptability in the amount of smoothing. The paper investigates the asymptotic and finite-sample properties of the asymmetric kernel Nadaraya-Watson, local linear, and re-weighted Nadaraya-Watson estimators. Pointwise weak consistency, rates of convergence and asymptotic normality are established for each of these estimators. As an important economic application of asymmetric kernel regression estimators, we reexamine the problem of estimating scalar diffusion processes.

Keywords: Nonparametric regression; strong mixing processes; Gamma kernel; Inverse Gaussian kernel; Reciprocal Inverse Gaussian kernel; diffusion estimation.

JEL classification numbers: C13; C14; C22; E43; G13.

*We would like to thank Evan Anderson, Arthur Lewbel, Peter Phillips, Ximing Wu, and participants at the 2007 MEG Conference for helpful comments. The first author gratefully acknowledges financial support from FQRSC, IFM2 and SSHRC.

†Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec H3G 1M8, Canada; phone: (514) 848-2424 (ext. 3935); fax: (514) 848-4536; e-mail: gospodin@alcor.concordia.ca; web: http://alcor.concordia.ca/~gospodin/.

‡Department of Economics, Northern Illinois University, Zulauf Hall 515, DeKalb, IL 60115, USA; phone: (815) 753-6434; fax: (815) 752-1019; e-mail: mhirukawa@niu.edu; web: http://www.math.niu.edu/~hirukawa/.

1 Introduction

The goal of this paper is to propose a nonstandard kernel-type estimator for nonparametric regression using time series data when the support of the regressor has a boundary. Suppose that for a stationary, strongly mixing process $\{(X_t, Y_t)\}_{t=-\infty}^{\infty} \in \mathbb{R}^2$, we are interested in estimating the regression function

$$m(x) = E\{\phi(Y_t)|X_t = x\}, \quad (1)$$

where $\phi(\cdot)$ is a known measurable function. Examples of (1) include conditional distribution function and r^{th}-order conditional moment estimation of Y_t given $X_t = x$ when $\phi(Y) = \mathbf{1}\{Y \leq y\}$ and $\phi(Y) = Y^r, r > 0$, respectively, and X_t may denote a lagged value of Y_t.

An interesting situation, that often arises in economics and finance, is when the regressor X_t in (1) is nonnegative. In this case, the local constant or Nadaraya-Watson (NW) estimator (Nadaraya, 1964; Watson, 1964) based on a standard, symmetric kernel suffers from bias near the origin that does not vanish even asymptotically. This is due to the fact that the symmetric kernels assign strictly positive weights outside the support of X_t. Accordingly, several boundary correction techniques have been proposed in the context of nonparametric regression such as boundary kernels (Gasser and Müller, 1979) and Richardson extrapolation (Rice, 1984). The local linear (LL) estimator by Fan and Gijbels (1992) is also known to automatically adapt the boundary bias. On the other hand, there is a growing literature on employing asymmetric kernels as an alternative device for boundary bias correction. In density estimation for positive observations, Chen (2000b) introduces the Gamma kernel, and Scaillet (2004) proposes the Inverse Gaussian and Reciprocal Inverse Gaussian kernels.[1] These asymmetric kernels have several attractive properties. First, they are free of boundary bias because the support of the kernels match that of the density. Second, the shape of the asymmetric kernel varies according to the positions of design points, and, as a result, the amount of smoothing changes in an adaptive manner. Third, the asymmetric kernels achieve the optimal (in integrated mean squared error sense) rate of convergence within the class of nonnegative kernel estimators.

[1] Throughout this paper, we refer to asymmetric kernels as kernel functions with support on the nonnegative real line. Bouezmarni and Rolin (2003), Brown and Chen (1999), Chen (1999, 2000a), and Jones and Henderson (2007) consider estimation of density and regression functions defined over the unit interval using different versions of asymmetric kernels.

Finally, their variances decrease as the position at which smoothing is made moves away from the boundary. This property is particularly advantageous when the support of the density has sparse regions.

Subsequently, Bouezmarni and Scaillet (2005) demonstrate weak convergence of the integrated absolute error for asymmetric kernel density estimators, whereas Hagmann and Scaillet (2007) investigate the local multiplicative bias correction for asymmetric kernel density estimators that is analogous to the one by Hjort and Jones (1996) in the symmetric kernel case. Besides density estimation, Chen (2002) applies asymmetric kernels to the LL estimator, and Fernandes and Monteiro (2005) establish the central limit theorem for a class of asymmetric kernel functionals. Furthermore, while all studies cited above are based on *iid* sampling, Bouezmarni and Rombouts (2006a,b) extend asymmetric kernel density and hazard estimation to positive time series data.

In line with these recent developments, this paper proposes a nonparametric regression estimator for dependent data using asymmetric kernels. We consider the NW, LL and re-weighted Nadaraya-Watson (RNW; Hall and Presnell, 1999) estimators and study their asymptotic and finite-sample behavior. While the NW estimator includes a "design bias" term that depends on the density function of the regressor, the LL estimator is free of this bias term. On the other hand, unlike the LL estimator, the NW estimator always yields estimated values within the range of observations $\{\phi(Y_t)\}_{t=1}^{T}$ and can preserve monotonicity and nonnegativity in conditional distribution estimation or nonnegativity in conditional variance estimation, for example. The RNW estimator is known to incorporate the strengths of the NW and LL estimators and has been used for nonparametric regression estimation (Cai, 2001), quantile estimation (Hall, Wolff and Yao, 1999; Cai, 2002), and conditional density estimation (De Gooijer and Zerom, 2003). We adapt the three estimators to asymmetric kernels and strongly mixing data, and establish pointwise weak consistency and asymptotic normality. We believe that our asymptotic results constitute an important theoretical complement to the results for time series nonparametric regression with symmetric kernels such as Lu and Linton (2007) and Masry and Fan (1997). Although we focus on the single regressor case throughout, the basic idea of our methodology is expected to hold in the multiple regressor context.

As an important economic application of the asymmetric kernel regression estimators, we consider the problem of estimating time-homogeneous drift and diffusion functions in scalar diffusion processes. Using the infinitesimal generator and Taylor series expansions, Stanton (1997) derives higher-order approximation formulae of the drift and diffusion functions that are estimated nonparametrically by the NW estimator. An interesting empirical finding that emerges from this work is that the drift function for the US short-term interest rate appears to exhibit substantial nonlinearity. In contrast, Chapman and Pearson (2000) argue that the documented nonlinearity in the short rate drift could be spurious due to the poor finite-sample properties of the Stanton's (1997) estimator at high values of interest rates where the data are sparse. Fan and Zhang (2003) estimate the first-order approximations of the drift and diffusion functions by the LL estimator, and conclude that there is little evidence against linearity in the short rate drift. Bandi (2002), Durham (2003) and Jones (2003) also do not find empirical support for nonlinear mean reversion in short-term rates. We expect that the use of the asymmetric kernel estimators can shed additional light on the nonparametric estimation of spot rate diffusion models.

The remainder of the paper is organized as follows. Section 2 develops asymptotic properties of the asymmetric kernel regression estimators and discusses their practical implementation. Section 3 conducts a Monte Carlo simulation experiment that examines the finite sample performance of these estimators in the context of scalar diffusion processes for spot interest rates. Section 4 summarizes the main results of the paper. All proofs are given in the appendix.

This paper adopts the following notational conventions: $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp(-y)\, dy$, $\alpha > 0$ is the Gamma function; $G(\alpha, \beta)$, $IG(\alpha, \beta)$ and $RIG(\alpha, \beta)$ symbolize the Gamma, Inverse Gaussian, and Reciprocal Inverse Gaussian distributions with parameters (α, β), respectively; $\mathbf{1}\{\cdot\}$ is the indicator function; $\mathbb{N}$ denotes the set of positive integers $\{1, 2, ...\}$, $\lfloor x \rfloor$ signifies integer part of x; and $c\,(> 0)$ denotes a generic constant, the quantity of which varies from statement to statement. The expression '$X \stackrel{d}{=} Y$' reads "A random variable X obeys the distribution Y." For integers n and k such that $0 \le k \le n$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes the number of combinations of size k taken from n objects. Finally, the expression '$X_T \sim Y_T$' is used whenever $X_T / Y_T \to 1$ as $T \to \infty$.

2 Nonparametric Regression Using Asymmetric Kernels for Time Series Data

2.1 Nonparametric Regression Estimators

Consider the problem of estimating nonparametric regression (1) using a sample $\{(X_t, Y_t)\}_{t=1}^{T}$, where $X_t \geq 0$ is assumed throughout. For a given design point $x > 0$, the NW, LL and RNW asymmetric kernel estimators are defined as

$$\begin{aligned}
\hat{m}^{nw}(x) &= \frac{\sum_{t=1}^{T} \phi(Y_t) K_{x,b}(X_t)}{\sum_{t=1}^{T} K_{x,b}(X_t)}, \\
\hat{m}^{ll}(x) &= \sum_{t=1}^{T} w_t(x) \phi(Y_t), \\
\hat{m}^{rnw}(x) &= \frac{\sum_{t=1}^{T} \phi(Y_t) p_t(x) K_{x,b}(X_t)}{\sum_{t=1}^{T} p_t(x) K_{x,b}(X_t)},
\end{aligned}$$

where $K_{x,b}(u)$ is an asymmetric kernel function with a smoothing parameter b.

The LL estimator satisfies $\hat{m}^{ll}(x) = \hat{\beta}_0(x)$, where $\hat{\boldsymbol{\beta}}(x) = \left[\hat{\beta}_0(x), \hat{\beta}_1(x)\right]^{\intercal}$ solves the optimization problem

$$\hat{\boldsymbol{\beta}}(x) = \arg\min_{\boldsymbol{\beta}(x)} \sum_{t=1}^{T} \{\phi(Y_t) - \beta_0(x) - \beta_1(x)(X_t - x)\}^2 K_{x,b}(X_t).$$

The weight functions for the LL estimator $\{w_t(x)\}_{t=1}^{T}$ are given by

$$\begin{aligned}
w_t(x) &= \frac{1}{T} \frac{\{S_2(x) - S_1(x)(X_t - x)\} K_{x,b}(X_t)}{S_0(x) S_2(x) - S_1^2(x)}, \\
S_j(x) &= \frac{1}{T} \sum_{t=1}^{T} (X_t - x)^j K_{x,b}(X_t), \; j = 0, 1, 2.
\end{aligned}$$

On the other hand, the weight functions for the RNW estimator $\{p_t(x)\}_{t=1}^{T}$ satisfy

$$p_t(x) \geq 0, \sum_{t=1}^{T} p_t(x) = 1, \sum_{t=1}^{T} (X_t - x) p_t(x) K_{x,b}(X_t) = 0. \tag{2}$$

Since $\{p_t(x)\}_{t=1}^{T}$ that satisfy (2) are not uniquely determined, they are specified as parameters that maximize the empirical log-likelihood $\sum_{t=1}^{T} \log\{p_t(x)\}$ subject to these constraints. Then, as shown in Cai (2001, 2002), $\{p_t(x)\}_{t=1}^{T}$ are defined as

$$p_t(x) = \frac{1}{T\{1 + \lambda(X_t - x) K_{x,b}(X_t)\}}, \tag{3}$$

where λ is the Lagrange multiplier associated with $\sum_{t=1}^{T}(X_t - x)\, p_t(x)\, K_{x,b}(X_t) = 0$ that can be determined by maximizing the profile empirical log-likelihood

$$\mathcal{L}\left(\lambda; \{X_t\}_{t=1}^{T}, x\right) = \sum_{t=1}^{T} \log\{1 + \lambda (X_t - x) K_{x,b}(X_t)\}.$$

We consider several candidates for asymmetric kernels: Gamma density K_G with parameters $(x/b+1, b)$ proposed by Chen (2000b),[2] Inverse Gaussian (IG) density K_{IG} with parameters $(x, 1/b)$ and Reciprocal Inverse Gaussian (RIG) density K_{RIG} with parameters $(1/(x-b), 1/b)$ proposed by Scaillet (2004). These densities are given by

$$\begin{aligned}
K_{G(x/b+1,b)}(u) &= \frac{u^{x/b}\exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)}\mathbf{1}\{u>0\}, \\
K_{IG(x,1/b)}(u) &= \frac{1}{\sqrt{2\pi b u^3}}\exp\left\{-\frac{1}{2bx}\left(\frac{u}{x} - 2 + \frac{x}{u}\right)\right\}\mathbf{1}\{u>0\}, \\
K_{RIG(1/(x-b),1/b)}(u) &= \frac{1}{\sqrt{2\pi b u}}\exp\left\{-\frac{x-b}{2b}\left(\frac{u}{x-b} - 2 + \frac{x-b}{u}\right)\right\}\mathbf{1}\{u>0\}.
\end{aligned}$$

As is the case with symmetric kernels, the asymmetric kernel RNW estimator shares some attractive properties of both NW and LL estimators. By construction, $\min_t\{\phi(Y_t)\} \le \hat{m}^{rnw}(x) \le \max_t\{\phi(Y_t)\}$ for any x, and the RNW estimator always generates nonnegative estimates in finite samples whenever $\phi(\cdot)$ is nonnegative, as the NW estimator does. Moreover, the weight functions for the LL estimator $\{w_t(x)\}_{t=1}^{T}$ satisfy the moment conditions similar to (2)

$$\sum_{t=1}^{T} w_t(x) = 1, \sum_{t=1}^{T}(X_t - x)\, w_t(x) = 0.$$

Hence, the bias properties of the RNW estimator are expected to be as good as that of the corresponding LL estimator, and better than that of the NW estimator for interior design points.

2.2 Asymptotic Properties of Estimators

In this section we establish pointwise weak consistency with rates and asymptotic normality of the NW, LL and RNW estimators for strongly mixing processes. Before stating regularity conditions

[2] Chen (2000b) also proposes another version of the Gamma kernel function

$$K_G(u; \rho_b(x), b) = \frac{u^{\rho_b(x)-1}\exp\left(-\frac{u}{b}\right)}{b^{\rho_b(x)}\Gamma(\rho_b(x))}\mathbf{1}\{u>0\},$$

where

$$\rho_b(x) = \begin{cases} x/b & \text{if } x \ge 2b \\ (x/b)^2/4 + 1 & \text{if } x \in [0, 2b) \end{cases}.$$

However, this version is not considered here, because asymptotic properties of the LL and RNW estimators using $K_G(u; \rho_b(x), b)$ for interior x (satisfying $x/b \to \infty$) are first-order equivalent to those when $K_G(u; x/b+1, b)$ is employed.

for our main results, we provide the definition of an α-mixing process for reference. Let $\mathcal{F}_a^b$ denote the σ-algebra generated by the stationary sequence $\{(X_t, Y_t)\}_{t=a}^b$ and

$$\alpha(k) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_k^\infty} |\Pr(A\cap B) - \Pr(A)\Pr(B)|, \; k \geq 1.$$

Then, the stationary process $\{(X_t, Y_t)\}_{t=-\infty}^\infty$ is said to be strongly mixing or α-mixing if $\alpha(k) \to 0$ as $k \to \infty$ (Rosenblatt, 1956). Also, let $f(x)$ be the marginal density of the regressor X_t, and define $\sigma^2(x) = Var\{\phi(Y_t)| X_t = x\}$. To obtain our main results, the following regularity conditions are required:

(A1) For a given design point $x > 0$, $m''(x)$, $f''(x)$ and $\sigma^2(x)$ are bounded and continuous.

(A2) $\sup_{x\geq 0} f(x) \leq M_1 < \infty$, $0 < m_1 \leq \inf_{x\geq 0} f(x)$, and $\sup_{u\geq 0, v\geq 0} f_{t,s}(u, v) \leq M_2 < \infty$.

(A3) $E\left\{|\phi(Y_1)|^\delta \middle| X_1 = u\right\} \leq \alpha_0 + \alpha_1 u^l$ and $E[\max\{|\phi(Y_t)|, |\phi(Y_s)|, |\phi(Y_t)\phi(Y_s)|\}| X_t = u, X_s = v] \leq \beta_0 + \beta_1 u^m + \beta_2 v^n$, $\forall u, v \geq 0$, for some $\delta > 2$, for some $\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2 \geq 0$, and for some $l, m, n \in \mathbb{N}$.

(A4) The strong mixing coefficient $\alpha(k)$ satisfies $\sum_{k=1}^\infty k^a \{\alpha(k)\}^{1-2/\delta} < \infty$ for some $a > 1 - 2/\delta$.

(A5) The smoothing parameter $b = b_T$ satisfies

$$\begin{cases} b \to 0 \text{ and } bT \to \infty & \text{for the Gamma and RIG kernels} \\ b \to 0 \text{ and } b^2 T \to \infty & \text{for the IG kernel} \end{cases}$$

as $T \to \infty$.

(A6) There exists a sequence $s_T \in \mathbb{N}$ such that $s_T \to \infty$, $s_T = o\left\{\left(b^{1/2}T\right)^{1/2}\right\}$, and $\left(T/b^{1/2}\right)^{1/2} \alpha(s_T) \to 0$ as $T \to \infty$.

(A7) The smoothing parameter $b = b_T$ additionally satisfies $b^{5/2}T \to \gamma \in [0, \infty)$ as $T \to \infty$.

Similar conditions to (A1)-(A7) are commonly used in the literature of LL (Lu and Linton, 2007; Masry and Fan, 1997) and RNW estimation (Cai, 2001, 2002; De Gooijer and Zerom, 2003) with dependent data. The condition (A3) is inspired by Hansen (2006), who derives uniform convergence rates of nonparametric density and regression estimators using dependent data even when unbounded support kernels are employed. Both Hansen (2006) and this paper allow the two conditional moments to diverge. An important difference is that while his condition controls the divergence rates of the conditional moments in comparison with the rate of decay in tails of the marginal density of regressors, (A3) assumes the existence of polynomial dominating functions, taking into account that all three asymmetric kernels have moments of any nonnegative integer order, as indicated in the proof of Lemma B2 in the appendix.

The conditions (A5) and (A7) for the smoothing parameter b are required to establish the asymptotic normality of the estimators and ensure that the bias and the variance converge to zero, and the remainder term in the bias expression is asymptotically negligible.

(A4) implies that the strong mixing coefficient has the size $-(\delta-1)/(\delta-2)$. To establish Theorem 2 (joint asymptotic normality of regression and first-order derivative estimators), we need to replace (A4) and (A5) by the stronger conditions (A4') and (A5') stated below. Note that (A4') and (A5) are required to approximate the variance of the first-order derivative estimator, and to ensure that the variance converges to zero, respectively. In contrast, the original conditions (A4) and (A5) suffice to demonstrate the asymptotic results for the LL estimator only.

(A4') The strong mixing coefficient satisfies $\sum_{k=1}^{\infty} k^a \{\alpha(k)\}^{1-2/\delta} < \infty$ for some $a > 3(1-2/\delta)$.

(A5') The smoothing parameter $b = b_T$ satisfies

$$\begin{cases} b \to 0 \text{ and } b^3 T \to \infty & \text{for the Gamma and RIG kernels} \\ b \to 0 \text{ and } b^6 T \to \infty & \text{for the IG kernel} \end{cases}$$

as $T \to \infty$.

Now we present kernel-specific results on weak consistency and asymptotic normality of the three estimators. Since the results depend on the kernel employed, we denote the NW estimator using the Gamma kernel as $\hat{m}_G^{nw}(x)$, for example. A similar notational convention is applied to the LL and RNW estimators. We also mean by "interior x" and "boundary x" that the design point x satisfies $x/b \to \infty$ and $x/b \to \kappa > 0$ as $T \to \infty$, respectively.

Theorems 1, 2 and 3 establish the pointwise weak consistency and asymptotic normality of the asymmetric kernel NW, LL and RNW estimators for interior x.

Theorem 1. *If conditions (A1)-(A7) hold, then for interior x,*

$$\sqrt{b^{1/2}T}\left[\hat{m}_G^{nw}(x)-m(x)-\left\{m'(x)\left(1+\frac{xf'(x)}{f(x)}\right)+\frac{x}{2}m''(x)\right\}b\right]\xrightarrow{d}N(0,V_G),$$
$$\sqrt{b^{1/2}T}\left[\hat{m}_{IG}^{nw}(x)-m(x)-x^3\left\{m'(x)\frac{f'(x)}{f(x)}+\frac{1}{2}m''(x)\right\}b\right]\xrightarrow{d}N(0,V_{IG}),$$
$$\sqrt{b^{1/2}T}\left[\hat{m}_{RIG}^{nw}(x)-m(x)-x\left\{m'(x)\frac{f'(x)}{f(x)}+\frac{1}{2}m''(x)\right\}b\right]\xrightarrow{d}N(0,V_{RIG}),$$

where $V_G=\frac{1}{2\sqrt{\pi}x^{1/2}}\frac{\sigma^2(x)}{f(x)}$, $V_{IG}=\frac{1}{2\sqrt{\pi}x^{3/2}}\frac{\sigma^2(x)}{f(x)}$ *and* $V_{RIG}=V_G$.

Proof. See Appendix A. ■

Theorem 2. *If conditions (A1)-(A3), (A4'), (A5'), (A6)-(A7) hold, then for interior x,*

$$T_{b,1}\left\{\hat{\boldsymbol{\beta}}_G(x)-\boldsymbol{\beta}(x)-\begin{bmatrix}\frac{1}{2}xm''(x)b\\0\end{bmatrix}\right\}\xrightarrow{d}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&0\\0&\frac{1}{2x}\end{bmatrix}V_G\right),$$
$$T_{b,1}\left\{\hat{\boldsymbol{\beta}}_{IG}(x)-\boldsymbol{\beta}(x)-\begin{bmatrix}\frac{1}{2}x^3m''(x)b\\0\end{bmatrix}\right\}\xrightarrow{d}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&0\\0&\frac{1}{2x^3}\end{bmatrix}V_{IG}\right),$$
$$T_{b,1}\left\{\hat{\boldsymbol{\beta}}_{RIG}(x)-\boldsymbol{\beta}(x)-\begin{bmatrix}\frac{1}{2}xm''(x)b\\0\end{bmatrix}\right\}\xrightarrow{d}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&0\\0&\frac{1}{2x}\end{bmatrix}V_{RIG}\right),$$

where $\boldsymbol{\beta}(x)=[m(x),m'(x)]^{\intercal}$ *and* $T_{b,1}=\sqrt{b^{1/2}T}\begin{bmatrix}1&0\\0&b^{1/2}\end{bmatrix}$.

Proof. See Appendix A. ■

Corollary 1. *If conditions (A1)-(A7) hold, then for interior x,*

$$\sqrt{b^{1/2}T}\left\{\hat{m}_G^{ll}(x)-m(x)-\frac{1}{2}xm''(x)b\right\}\xrightarrow{d}N(0,V_G),$$
$$\sqrt{b^{1/2}T}\left\{\hat{m}_{IG}^{ll}(x)-m(x)-\frac{1}{2}x^3m''(x)b\right\}\xrightarrow{d}N(0,V_{IG}),$$
$$\sqrt{b^{1/2}T}\left\{\hat{m}_{RIG}^{ll}(x)-m(x)-\frac{1}{2}xm''(x)b\right\}\xrightarrow{d}N(0,V_{RIG}).$$

Theorem 2 and Corollary 1 can be further extended to the p^{th}-order local polynomial estimation, provided that $m(\cdot)$ has a bounded continuous p^{th}-order derivative and the mixing condition is properly strengthened.

Theorem 3. *If conditions (A1)-(A7) hold, then for interior x,*

$$\sqrt{b^{1/2}T}\left\{\hat{m}_{G}^{rnw}(x)-m(x)-\frac{1}{2}xm''(x)b\right\}\xrightarrow{d}N(0,V_{G}),$$
$$\sqrt{b^{1/2}T}\left\{\hat{m}_{IG}^{rnw}(x)-m(x)-\frac{1}{2}x^{3}m''(x)b\right\}\xrightarrow{d}N(0,V_{IG}),$$
$$\sqrt{b^{1/2}T}\left\{\hat{m}_{RIG}^{rnw}(x)-m(x)-\frac{1}{2}xm''(x)b\right\}\xrightarrow{d}N(0,V_{RIG}).$$

Proof. See Appendix A. ■

The next two theorems derive the pointwise weak consistency and asymptotic normality of NW and LL estimators for boundary x. Before proceeding, we modify the conditions (A6) and (A7). Note that two alternative replacements of (A7), namely, (A7') and (A7"), are required for asymptotic normality of NW and LL estimators, respectively.

(A6') There exists a sequence $s_T \in \mathbb{N}$ such that

$$\begin{cases} s_T\to\infty,\ s_T=o\left\{(bT)^{1/2}\right\},\ \text{and}\ (T/b)^{1/2}\alpha(s_T)\to 0 & \text{for the Gamma and RIG kernels}\\ s_T\to\infty,\ s_T=o\left\{\left(b^{2}T\right)^{1/2}\right\},\ \text{and}\ \left(T/b^{2}\right)^{1/2}\alpha(s_T)\to 0 & \text{for the IG kernel}\end{cases}$$

as $T\to\infty$.

(A7') The smoothing parameter $b=b_T$ additionally satisfies

$$\begin{cases} b^{3}T\to\gamma\in[0,\infty) & \text{for the Gamma kernel}\\ b^{10}T\to\gamma\in[0,\infty) & \text{for the IG kernel}\\ b^{5}T\to\gamma\in[0,\infty) & \text{for the RIG kernel}\end{cases}$$

as $T\to\infty$.

(A7") The smoothing parameter $b=b_T$ additionally satisfies

$$\begin{cases} b^{5}T\to\gamma\in[0,\infty) & \text{for the Gamma and RIG kernels}\\ b^{10}T\to\gamma\in[0,\infty) & \text{for the IG kernel}\end{cases}$$

as $T\to\infty$.

Theorem 4. *If conditions (A1)-(A5), (A6'), (A7') hold, then for boundary x,*

$$\sqrt{bT}\left\{\hat{m}_{G}^{nw}(x)-m(x)-m'(x)b\right\}\xrightarrow{d}N\left(0,V_{G}^{B}\right),$$
$$\sqrt{b^{2}T}\left[\hat{m}_{IG}^{nw}(x)-m(x)-\kappa^{3}\left\{m'(x)\frac{f'(x)}{f(x)}+\frac{1}{2}m''(x)\right\}b^{4}\right]\xrightarrow{d}N\left(0,V_{IG}^{B}\right),$$
$$\sqrt{bT}\left[\hat{m}_{RIG}^{nw}(x)-m(x)-(\kappa+1)\left\{m'(x)\frac{f'(x)}{f(x)}+\frac{1}{2}m''(x)\right\}b^{2}\right]\xrightarrow{d}N\left(0,V_{RIG}^{B}\right),$$

where $V_G^B = \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\frac{\sigma^2(x)}{f(x)}$, $V_{IG}^B = \frac{1}{2\sqrt{\pi}\kappa^{3/2}}\frac{\sigma^2(x)}{f(x)}$ *and* $V_{RIG}^B = \frac{\kappa^{-1/2}+(7/16)\kappa^{-3/2}+(3/32)\kappa^{-5/2}}{2\sqrt{\pi}}\frac{\sigma^2(x)}{f(x)}$.

Proof. See Appendix A. ■

Theorem 5. *If conditions (A1)-(A3), (A4'), (A5'), (A6'), (A7") hold, then for boundary x,*

$$T_{b,2}\left\{\hat{\boldsymbol{\beta}}_G(x) - \boldsymbol{\beta}(x) - \frac{m''(x)}{2}\begin{bmatrix}(\kappa-2)b^2 \\ 4b\end{bmatrix}\right\} \xrightarrow{d} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}2\kappa+5 & -2 \\ -2 & 1\end{bmatrix}\frac{V_G^B}{2(\kappa+1)}\right)$$

$$T_{b,3}\left\{\hat{\boldsymbol{\beta}}_{IG}(x) - \boldsymbol{\beta}(x) - \begin{bmatrix}\frac{1}{2}\kappa^3 m''(x) b^4 \\ 0\end{bmatrix}\right\} \xrightarrow{d} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0 \\ 0 & \frac{1}{2\kappa^3}\end{bmatrix}V_{IG}^B\right),$$

$$T_{b,2}\left\{\hat{\boldsymbol{\beta}}_{RIG}(x) - \boldsymbol{\beta}(x) - \frac{m''(x)}{2}\begin{bmatrix}(\kappa+1)b^2 \\ \left(\frac{3\kappa+5}{\kappa+1}\right)b\end{bmatrix}\right\} \xrightarrow{d} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \boldsymbol{\Sigma}_{RIG}V_{RIG}^B\right),$$

where

$$\boldsymbol{\Sigma}_{RIG} = \begin{bmatrix} 1 & -\frac{3}{4(\kappa+1)}\left\{1 - \frac{(13/48)\kappa^{-3/2}+(5/32)\kappa^{-5/2}}{\kappa^{-1/2}+(7/16)\kappa^{-3/2}+(3/32)\kappa^{-5/2}}\right\} \\ -\frac{3}{4(\kappa+1)}\left\{1 - \frac{(13/48)\kappa^{-3/2}+(5/32)\kappa^{-5/2}}{\kappa^{-1/2}+(7/16)\kappa^{-3/2}+(3/32)\kappa^{-5/2}}\right\} & \frac{\kappa}{2(\kappa+1)^2}\left\{1 + \frac{(5/8)\kappa^{-3/2}+(5/8)\kappa^{-5/2}+(33/64)\kappa^{-7/2}}{\kappa^{-1/2}+(7/16)\kappa^{-3/2}+(3/32)\kappa^{-5/2}}\right\}\end{bmatrix},$$

$T_{b,2} = \sqrt{bT}\begin{bmatrix}1 & 0 \\ 0 & b\end{bmatrix}$ *and* $T_{b,3} = \sqrt{b^2T}\begin{bmatrix}1 & 0 \\ 0 & b^2\end{bmatrix}$.

Proof. See Appendix A. ■

Corollary 2. *If conditions (A1)-(A5), (A6'), (A7") hold, then for boundary x,*

$$\sqrt{bT}\left\{\hat{m}_G^{ll}(x) - m(x) - \frac{1}{2}(\kappa-2)m''(x)b^2\right\} \xrightarrow{d} N\left(0, \frac{2\kappa+5}{2(\kappa+1)}V_G^B\right),$$

$$\sqrt{b^2T}\left\{\hat{m}_{IG}^{ll}(x) - m(x) - \frac{\kappa^3}{2}m''(x)b^4\right\} \xrightarrow{d} N\left(0, V_{IG}^B\right),$$

$$\sqrt{bT}\left\{\hat{m}_{RIG}^{ll}(x) - m(x) - \frac{1}{2}(\kappa+1)m''(x)b^2\right\} \xrightarrow{d} N\left(0, V_{RIG}^B\right).$$

2.2.1 Discussion of results

Choice of estimator and kernel function. In case of an interior design point, the results in Theorem 1, Corollary 1 and Theorem 3 reveal that the LL and RNW estimators eliminate the "design bias" term of the NW estimator without any effect on the variance. An immediate consequence of Theorem 3 and Corollary 1 is that each RNW estimator is first-order equivalent to the corresponding LL estimator, as is the case with symmetric kernels. Furthermore, we can see from Theorems 1-3 that for each of NW, LL and RNW estimators, variances decrease with x, i.e. as the position in which smoothing is made moves away from the boundary. This property is particularly advantageous when the support of the regressor has sparse regions.

Now turning our attention to the properties of the different kernel functions, we note that the estimators based on the Gamma and RIG kernels are first-order equivalent for interior x. The asymptotic bias of the IG-based estimators is larger than that of the Gamma and RIG estimators when $x > 1$; however, the larger bias is compensated by a much smaller variance. For example, in the special case of a linear function $m(x)$, the estimators $\hat{m}_{IG}^{ll}(x)$ and $\hat{m}_{IG}^{rnw}(x)$ dominate their Gamma and RIG counterparts for $x > 1$ but not for $x < 1$ which is the situation in our interest rate application.

Some interesting findings emerge from the boundary design point case. First, comparing Theorems 1 and 4 or Corollaries 1 and 2, we see that for each of the NW and LL estimators, improvement in order of magnitude in the bias term is achieved at the expense of inflating the variance. Indeed, if the smoothing parameter b is chosen to satisfy (A5) and (A7), then the bias of the NW and LL estimators becomes asymptotically negligible over the boundary region, and thus only the variance matters. Second, for the IG and RIG kernels, the LL estimator eliminates the "design bias" term of the NW estimator even over the boundary region, whereas the Gamma NW and LL estimators do not have common bias terms.

More importantly, Theorem 4 and Corollary 2 show that $\hat{m}_{G}^{nw}(x)$ and $\hat{m}_{G}^{ll}(x)$ do not share the same asymptotic variance for boundary x, which is typically the case for interior x, IG, RIG and symmetric kernels.[3] For example, for $\kappa = 0.5$, the variance of $\hat{m}_{G}^{ll}(x)$ is twice as big as the variance of $\hat{m}_{G}^{nw}(x)$ and it may well be the case that the Gamma-based NW estimator is preferred over the Gamma LL estimator even though the latter may have a smaller bias. Figure 1 plots the differences in the asymptotic variances of $\hat{m}_{G}^{nw}(x)$, $\hat{m}_{RIG}^{nw}(x)$ and $\hat{m}_{G}^{ll}(x)$ as a function of $\kappa \in [0.2, 1]$,[4] and shows the substantial efficiency advantages of the Gamma NW estimator at the extreme design points.

[3] The mean of $G(x/b+1, b)$ is not the design point x but $x+b$, whereas both $IG(x, 1/b)$ and $RIG(1/(x-b), 1/b)$ have mean x. Hence, as $x/b \to \kappa$, $S_1(x) = O_p(b)$ for the Gamma kernel, whereas $S_1(x) = O_p(b^2)$ for the IG and RIG kernels. Then, the term involving $m'(x)$ dominates the bias of $\hat{m}_{G}^{nw}(x)$ for boundary x, and as a result, $\hat{m}_{G}^{nw}(x)$ and $\hat{m}_{G}^{ll}(x)$ do not have common bias terms. Likewise, the reason why $\hat{m}_{G}^{nw}(x)$ and $\hat{m}_{G}^{ll}(x)$ do not share the same asymptotic variance for boundary x is that when the Gamma kernel is employed, the scale-adjusted outer product matrix $\mathbf{S}_T^{\dagger} = \begin{bmatrix} S_0(x) & b^{-1}S_1(x) \\ b^{-1}S_1(x) & b^{-2}S_2(x) \end{bmatrix}$ has non-negligible off-diagonal elements in the limit as $x/b \to \kappa$; for details, see Lemma A7 in Appendix A.

[4] The estimator $\hat{m}_{IG}^{nw}(x)$ is not plotted in the diagram, because it has a slower rate of convergence than $\hat{m}_{G}^{nw}(x)$, $\hat{m}_{RIG}^{nw}(x)$ and $\hat{m}_{G}^{ll}(x)$ for boundary x.

Also, unlike the case for interior x, asymptotic independence between LL regression and first-order derivative estimators does not necessarily hold for boundary x; in fact, asymptotic variance-covariance matrices of Gamma and RIG-based LL estimators in Theorem 5 have non-zero off-diagonal elements when $x/b \to \kappa$ is assumed. Moreover, we do not provide a theorem for the RNW estimator in the boundary case. The difficulty for establishing the asymptotic properties of the RNW arises from the fact that when x is located in a particularly small boundary region (of order $O(b)$), there are not enough observations less than x for the constraint (2) to hold, and, as a result, the RNW estimator is not well defined. Even though the asymptotic behavior of the RNW estimator for boundary x does not affect its global properties, these observations indicate that the numerical performance of the RNW estimator near the boundaries could be rather poor which is confirmed by our simulation results presented below.

Mean squared error. It follows directly from Theorem 3 and Corollary 1 that the mean squared errors (MSE) of the three LL (and thus RNW) estimators for interior x are approximated by

$$\begin{aligned} MSE\left\{\hat{m}_G^{ll}(x)\right\} &\sim \frac{1}{4}x^2\left\{m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}f(x)}, \\ MSE\left\{\hat{m}_{IG}^{ll}(x)\right\} &\sim \frac{1}{4}x^6\left\{m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{3/2}f(x)}, \\ MSE\left\{\hat{m}_{RIG}^{ll}(x)\right\} &\sim \frac{1}{4}x^2\left\{m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}f(x)}. \end{aligned}$$

In contrast, Theorem 1 suggests that the MSEs of their corresponding NW estimator for interior x are approximated by

$$\begin{aligned} MSE\left\{\hat{m}_G^{nw}(x)\right\} &\sim \left\{m'(x)\left(1+\frac{xf'(x)}{f(x)}\right)+\frac{x}{2}m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}f(x)}, \\ MSE\left\{\hat{m}_{IG}^{nw}(x)\right\} &\sim \left\{m'(x)\frac{x^3f'(x)}{f(x)}+\frac{x^3}{2}m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{3/2}f(x)}, \\ MSE\left\{\hat{m}_{RIG}^{nw}(x)\right\} &\sim \left\{m'(x)\frac{xf'(x)}{f(x)}+\frac{x}{2}m''(x)\right\}^2 b^2 + \frac{1}{b^{1/2}T}\frac{\sigma^2(x)}{2\sqrt{\pi}x^{1/2}f(x)}, \end{aligned}$$

and the NW estimators contain an additional "design bias" term that depends on the density of the regressor $f(x)$ while the variance terms remain unchanged. These results agree with the case of standard symmetric kernels.

Optimal smoothing parameter. From the MSE expressions, it can be easily inferred that the optimal smoothing parameters of the LL (and thus RNW) estimators for interior x are

$$\begin{aligned} b_G^* &= \left[\frac{\sigma^2(x)}{2\sqrt{\pi}\{m''(x)\}^2 f(x)}\right]^{2/5} x^{-1}T^{-2/5}, \\ b_{IG}^* &= \left[\frac{\sigma^2(x)}{2\sqrt{\pi}\{m''(x)\}^2 f(x)}\right]^{2/5} x^{-3}T^{-2/5}, \\ b_{RIG}^* &= \left[\frac{\sigma^2(x)}{2\sqrt{\pi}\{m''(x)\}^2 f(x)}\right]^{2/5} x^{-1}T^{-2/5}. \end{aligned}$$

Note that the optimal smoothing parameters are $b^* = O\left(T^{-2/5}\right) = O\left(a^{*2}\right)$, where a^* is the MSE-optimal bandwidth for the LL estimator using second-order symmetric kernels. Also, at the optimum,

$$\begin{aligned} MSE^*\left\{\hat{m}_G^{ll}(x)\right\} &\sim \frac{5}{4}\left\{\frac{\sqrt{|m''(x)|}\sigma^2(x)}{2\sqrt{\pi}f(x)}\right\}^{4/5} T^{-4/5}, \\ MSE^*\left\{\hat{m}_{IG}^{ll}(x)\right\} &\sim \frac{5}{4}\left\{\frac{\sqrt{|m''(x)|}\sigma^2(x)}{2\sqrt{\pi}f(x)}\right\}^{4/5} T^{-4/5}, \\ MSE^*\left\{\hat{m}_{RIG}^{ll}(x)\right\} &\sim \frac{5}{4}\left\{\frac{\sqrt{|m''(x)|}\sigma^2(x)}{2\sqrt{\pi}f(x)}\right\}^{4/5} T^{-4/5}, \end{aligned}$$

and each optimal MSE is identical and does not depend on x (the dependence of each optimal MSE on x comes only through $f(x)$ and $\sigma^2(x)$). In addition, the optimal MSE is the same as that of the LL estimator using the Gaussian kernel. Therefore, as argued by Chen (2000b) and Scaillet (2004), we can see that, for interior x, the three asymmetric kernels defined over $[0, \infty)$ have the same pointwise efficiency as the Gaussian kernel over $(-\infty, \infty)$.

2.3 Implementation and Selection of Smoothing Parameter

The practical implementation of the proposed nonparametric estimators requires a choice of smoothing parameter. While the previous section provides some guidance in this direction, the expressions for the optimal smoothing parameters depend on unknown functions of the data and a uniform "plug-in rule" is difficult to obtain. Note also that the optimal smoothing parameters for the asymmetric kernels depend explicitly on the design point and, in principle, they should take different values at each x. Hagmann and Scaillet (2007), however, argue for a uniform 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 uniform 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 observation t are removed from the sample and the function $m(x)$ is estimated from the remaining $T-(2h+1)$ observations. The idea behind this method is that, due to the strong mixing property of the data, 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.

Let $\hat{m}_{-(t-h):(t+h)}(X_t)$ denote the estimate from observations $1, 2, ..., t-h-1, t+h+1, ..., T$. Then, the smoothing parameter can be selected by minimizing the least squares cross-validation function

$$CV(b) = \arg\min_{b \in B_T} \sum_{t=h+1}^{T-h} \left\{ \phi(Y_t) - \hat{m}_{-(t-h):(t+h)}(X_t) \right\}^2 \psi(X_t), \tag{4}$$

where $\psi(\cdot)$ is a weighting function that has compact support and is bounded by 1. Minimizing $CV(b)$ is asymptotically equivalent to minimizing the true expected prediction error provided that h/T goes to zero at some rate as $h \to \infty$ and $T \to \infty$ (Chu, 1989; Györfi *et al.*, 1989). Alternatively, if one assumes that h is a nontrivial fraction of the sample size T so that h/T is a fixed constant as $h \to \infty$ and $T \to \infty$, $CV(b)$ has to be corrected as in Burman *et al.* (1994).[5] While the corrected $CV(b)$ of Burman *et al.* (1994) may provide a better finite-sample approximation to the true expected prediction error, this procedure is computationally more involved and in our numerical experiments the smoothing parameter is chosen by minimizing (4) with $\psi(X_t) = 1$.

3 Monte Carlo Experiment: Diffusion Models of Spot Rate

The nonparametric estimation of continuous-time diffusion processes, that are used to describe the underlying dynamics of spot interest rates, has been an active area of recent research (Bandi and Phillips, 2003; Florens-Zmirou, 1993; Jiang and Knight, 1997; Nicolau, 2003; among others). In this section, we assess the finite-sample properties of our proposed asymmetric kernel estimators in the

[5] The asymptotic optimality of the h-block cross validation bandwidths for mixing data in Chu (1989), Györfi *et al.* (1989) and Burman *et al.* (1994) is derived for symmetric kernels. While it is useful to extend these results to asymmetric kernels, it is beyond the scope of this paper.

context of a diffusion process of spot rate and evaluate the economic importance of the results in terms of computed bond and option pricing errors.

The data for the first simulation experiment is generated from the CIR model (Cox *et al.*, 1985)

$$dr_t = \kappa(\theta - r_t)\,dt + \sigma r_t^{1/2} dW_t, \tag{5}$$

where W_t is a standard Brownian motion. This 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). $5,000$ sample paths for the spot interest rate of length $T = 600$ observations are simulated using the procedure described in Chapman and Pearson (2000). After drawing an initial value from the marginal Gamma density, the interest rate process is constructed recursively by drawing random numbers from the transition non-central chi-square density and using the values for κ, θ and σ and a time step between two consecutive observation equal to $\Delta = 1/52$ corresponding to weekly data.

We consider two parameter configurations that are used in Chapman and Pearson (2000) - $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$ and $(0.85837, 0.085711, 0.1566)$, that produce persistent interest rate process with monthly autocorrelations of 0.982 and 0.931, respectively. The two specifications are calibrated to generate data with the same unconditional mean variance. The strong mixing property of the process generated by (5), is demonstrated by Carrasco *et al.* (2007).

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 5% 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.

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

$$P_t^{\tau} = P_0 E_t \left[\exp\left(-\int_t^{\tau} r_u^* du \right) \right],$$

where $r_t^* = r_t$, $dr_t^* = [\widehat{\mu}(r_t^*) - \widehat{\lambda}(r_t^*)]dt + \widehat{\sigma}(r_t^*)dW_t$, and $\widehat{\mu}(r_t^*)$, $\widehat{\sigma}(r_t^*)$ and $\widehat{\lambda}(r_t^*)$ denote the nonparametric estimates of the drift, diffusion and market price of risk functions, respectively. 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. 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 $(n - t)$ on a zero-coupon bond with maturity $(\tau - t)$, face value P_0 and exercise price K is computed as

$$\begin{aligned} C_t^n &= E_t\left[\exp\left(-\int_t^n r_u^* du\right)\max\left(P_n^\tau - K, 0\right)\right] \\ &= E_t\left[\exp\left(-\int_t^n r_u^* du\right)\max\left(P_0 E_n\left[\exp\left(-\int_N^\tau r_v^* dv\right)\right] - K, 0\right)\right], \end{aligned}$$

where $n < \tau$ and sample paths for r_t^* are simulated from the nonparametrically estimated discretized model of spot rate.

In order to evaluate if the proposed estimators capture well the shape of the true function, data are also generated from the nonlinear diffusion model of Ahn and Gao (1999)

$$dr_t = \kappa\left(\theta - r_t\right) r_t dt + \sigma r_t^{1.5} dW_t, \tag{6}$$

where the drift is a quadratic function of the interest rate. The strong mixing properties of the process generated by (6) can be inferred by verifying the conditions in Chen *et al.* (1999). As argued by Ahn and Gao (1999), $s_t = 1/r_t$ follows a square-root process with non-central chi-square transitional density which facilitates the simulation of interest rate data. The particular parameterization that we employ in simulating the data from (6) is $(\kappa, \theta, \sigma) = (3, 0.1, 1)$ which is similar to the values estimated by Ahn and Gao (1999) from actual data.

We consider the NW estimators with Gaussian and Gamma kernels and the LL and RNW estimators with Gamma kernel. The LL estimator with Gaussian kernel produces substantially larger biases than these estimators and is not reported.

First, Figures 2 to 5 present the finite-sample properties of the asymmetric NW estimators of the drift function from the CIR model. Figures 2 and 4 plot the median drift estimates of the Gamma, IG and RIG NW estimators for both parameterizations and a fixed smoothing parameter.

In agreement with the theoretical results in Section 2.2, the Gamma and RIG estimators exhibit very similar behavior and provide a very good approximation to the true drift function. In contrast, the IG drift function estimator is much more biased (the bias of the IG estimator is still substantial for larger smoothing parameters) and we do not consider this estimator further in the paper. Figures 3 and 5 plot the 90% Monte Carlo confidence bands of the Gamma and RIG estimators and reveal that the Gamma estimator is less variable than the RIG estimator especially for the more persistent specification. In the rest of the paper, we only report the results from the Gamma NW estimator noting that the RIG NW estimator delivers very similar results.

In order to compare the properties of the Gamma NW with the Gaussian NW, Gamma RNW and Gamma LL estimators, we choose a common algorithm for selecting the smoothing parameter based on h-block cross validation with $h = 30$ (our experiments with different values of h delivered very similar results.) It is interesting to note that Gamma NW and RNW select significantly smaller smoothing parameters than the Gaussian NW and Gamma LL estimators.

The median Monte Carlo estimates plotted in Figures 6 and 8 show that the Gamma NW and Gaussian NW are almost unbiased whereas the bias of the Gamma LL is rather large for both interior and boundary design points. It appears that the Gamma LL estimator is more sensitive to the high persistence in the data and its behavior improves for less persistent specifications. While the Gamma NW is only slightly less biased than the Gaussian NW, the asymmetric kernel estimator exhibits smaller variability (Figure 7) near the boundaries. The behavior of the asymmetric RNW estimator is similar to the Gamma NW estimator but it tends to be much more noisy.

Finally, Figures 9 and 10 plot the drift function estimates from the nonlinear diffusion specification of Ahn and Gao (1999). As in the case of linear drift, the Gamma kernel estimator provides a very good approximation of the true drift function. The symmetric (Gaussian) NW estimator exhibits larger bias and variability for interest rates above 9% whereas the local linear estimator again tends to perform rather poorly compared to the asymmetric kernel estimator. In summary, the Gamma NW appears to be the best performing nonparametric estimator of the drift function of highly persistent diffusion processes considered in the simulation experiments.

The economic significance of the improved estimation of diffusion models of spot rate is evaluated by comparing bond and option pricing errors based on different nonparametric estimators for the CIR model with $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$. For reference, we include also the bond and option prices computed analytically from the OLS estimates of κ, θ and σ obtained from the discretized version of the model. The results are presented in Table 1. Despite the fact that the OLS estimator uses knowledge of the true shapes of the drift and diffusion functions, the bond and especially the call option prices are substantially underestimated due mainly to the severe downward bias of the OLS estimator in autoregressive models (Phillips and Yu, 2005). In contrast, the bond and derivative prices based on both symmetric and asymmetric kernel estimators are much less biased and actually produce slightly positive pricing errors. The bias of the Gamma estimator is smaller than its Gaussian counterpart but more importantly, the Gamma-based bond and option prices enjoy much smaller variability and tighter confidence intervals than the symmetric kernel-based prices.

4 Conclusion

This paper proposes several asymmetric kernel estimators of conditional moment functions based on dependent data and nonnegative conditioning variables. The consistency, rate of convergence and asymptotic normality of these estimators are established for both interior and boundary design points. We show that the asymmetric kernel estimators possess some appealing properties such as lack of boundary bias and/or 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 estimators is evaluated in the context of a scalar diffusion process of spot interest rate. Several interesting directions for future research include construction of bootstrap confidence bands and bootstrap-based specification testing, establishing uniform rates of convergence and rate improvement via multiplicative bias correction.

A Appendix A: Proofs of Theorems

In this appendix, we present the proofs only for the Gamma kernel because the proofs for the IG and RIG kernels are similar. Note that approximations to the moments of the IG and RIG kernels can be obtained by following Scaillet (2004) and applying Lemmata B1 and B2.

A.1 Proofs of Theorems 1 and 2

The proofs of Theorems 1 and 2 require the following three lemmata. Before proceeding, define $\epsilon_t = \phi(Y_t) - m(X_t)$.

Lemma A1. *Let*

$$\mathbf{S}_T = \begin{bmatrix} S_0(x) & b^{-1/2}S_1(x) \\ b^{-1/2}S_1(x) & b^{-1}S_2(x) \end{bmatrix}.$$

If the conditions (A1)-(A5) hold, then for interior x,

$$\begin{aligned} \mathbf{S}_{G,T} \xrightarrow{p} \mathbf{S}_G &= \begin{bmatrix} 1 & 0 \\ 0 & x \end{bmatrix} f(x), \\ \mathbf{S}_{IG,T} \xrightarrow{p} \mathbf{S}_{IG} &= \begin{bmatrix} 1 & 0 \\ 0 & x^3 \end{bmatrix} f(x), \\ \mathbf{S}_{RIG,T} \xrightarrow{p} \mathbf{S}_{RIG} &= \begin{bmatrix} 1 & 0 \\ 0 & x \end{bmatrix} f(x). \end{aligned}$$

Proof of Lemma A1. Using Lemma B1,

$$E\{S_{G,j}(x)\} = E\left\{(X_1 - x)^j K_{G(x/b+1,b)}(X_1)\right\} = E\left\{(\theta_{1,x} - x)^j f(\theta_{1,x})\right\},$$

where $\theta_{1,x} \stackrel{d}{=} G(x/b+1, b)$. Taking a second-order Taylor expansion of $f(\theta_{1,x})$ around $\theta_{1,x} = x$ yields

$$(\theta_{1,x} - x)^j f(\theta_{1,x}) = (\theta_{1,x} - x)^j f(x) + (\theta_{1,x} - x)^{j+1} f'(x) + \frac{1}{2}(\theta_{1,x} - x)^{j+2} f''(x) + O_p\left\{(\theta_{1,x} - x)^{j+3}\right\}.$$

Therefore, by Lemma B2, for interior x,

$$\begin{aligned} E\{S_{G,0}(x)\} &= f(x) + O(b), \\ E\{S_{G,1}(x)\} &= \{f(x) + xf'(x)\}b + O\left(b^2\right), \\ E\{S_{G,2}(x)\} &= xf(x)b + O\left(b^2\right). \end{aligned}$$

Since strong mixing implies ergodicity, we can apply Birkhoff's ergodic theorem to establish the results. ■

Lemma A2. *Let*

$$\mathbf{t}_T^* = \begin{bmatrix} T_0^*(x) \\ b^{-1/2}T_1^*(x) \end{bmatrix} = \begin{bmatrix} T^{-1}\sum_{t=1}^T \epsilon_t K_{x,b}(X_t) \\ b^{-1/2}T^{-1}\sum_{t=1}^T (X_t - x)\epsilon_t K_{x,b}(X_t) \end{bmatrix}.$$

Also for an arbitrary vector $\mathbf{c} \in \mathbb{R}^2$, *define* $Q_T^* = \mathbf{c}^\intercal \mathbf{t}_T^*$. *If the conditions (A1)-(A3), (A4') and (A5) hold, then for interior* x,

$$\begin{aligned}
Var\left(\sqrt{b^{1/2}T}Q_{G,T}^*\right) &\to \mathbf{c}^\intercal \mathbf{V}_G \mathbf{c} = \mathbf{c}^\intercal \left\{\begin{bmatrix} \frac{1}{2\sqrt{\pi}x^{1/2}} & 0 \\ 0 & \frac{x^{1/2}}{4\sqrt{\pi}} \end{bmatrix} \sigma^2(x) f(x)\right\}\mathbf{c}, \\
Var\left(\sqrt{b^{1/2}T}Q_{IG,T}^*\right) &\to \mathbf{c}^\intercal \mathbf{V}_{IG} \mathbf{c} = \mathbf{c}^\intercal \left\{\begin{bmatrix} \frac{1}{2\sqrt{\pi}x^{3/2}} & 0 \\ 0 & \frac{x^{3/2}}{4\sqrt{\pi}} \end{bmatrix} \sigma^2(x) f(x)\right\}\mathbf{c}, \\
Var\left(\sqrt{b^{1/2}T}Q_{RIG,T}^*\right) &\to \mathbf{c}^\intercal \mathbf{V}_{RIG} \mathbf{c} = \mathbf{c}^\intercal \left\{\begin{bmatrix} \frac{1}{2\sqrt{\pi}x^{1/2}} & 0 \\ 0 & \frac{x^{1/2}}{4\sqrt{\pi}} \end{bmatrix} \sigma^2(x) f(x)\right\}\mathbf{c}.
\end{aligned}$$

Proof of Lemma A2. It suffices to demonstrate that

$$Var\left\{\sqrt{b^{1/2}T}T_{G,0}^*(x)\right\} = \frac{1}{2\sqrt{\pi}x^{1/2}}\sigma^2(x) f(x) + o(1), \tag{7}$$

$$Var\left\{\sqrt{b^{1/2}T}b^{-1/2}T_{G,1}^*(x)\right\} = \frac{x^{1/2}}{4\sqrt{\pi}}\sigma^2(x) f(x) + o(1), \tag{8}$$

$$Cov\left\{\sqrt{b^{1/2}T}T_{G,0}^*(x), \sqrt{b^{1/2}T}b^{-1/2}T_{G,1}^*(x)\right\} = o(1). \tag{9}$$

(i) Proof of (7). It follows from $E(\epsilon_t | X_t) = 0$ that

$$\begin{aligned}
Var\left\{\sqrt{b^{1/2}T}T_{G,0}^*(x)\right\} &= Var\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^T b^{1/4}\epsilon_t K_{G(x/b+1,b)}(X_t)\right\} \\
&\equiv \gamma_{G,0}(0) + 2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\gamma_{G,0}(j),
\end{aligned} \tag{10}$$

where $\gamma_{G,0}(j) = b^{1/2}E\left\{\epsilon_1\epsilon_{1+j}K_{G(x/b+1,b)}(X_1)K_{G(x/b+1,b)}(X_{1+j})\right\}$ is the j^{th}-order autocovariance of the stationary process $\left\{b^{1/4}\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$. For the first term on the right-hand side of (10), by the law of iterated expectations and Lemma B1,

$$\begin{aligned}
\gamma_{G,0}(0) &= b^{1/2}E\left\{\sigma^2(X_1)K_{G(x/b+1,b)}^2(X_1)\right\} \\
&= b^{1/2}A_{b,2}(x)E\left\{\sigma^2(\theta_{2,x})f(\theta_{2,x})\right\} \\
&= b^{1/2}\left\{\frac{b^{-1/2}x^{-1/2}}{2\sqrt{\pi}} + o\left(b^{-1/2}\right)\right\}E\left\{\sigma^2(\theta_{2,x})f(\theta_{2,x})\right\},
\end{aligned}$$

where $\theta_{2,x} \overset{d}{=} G(2x/b+1, b/2)$. Taking a Taylor expansion of $\sigma^2(\theta_{2,x}) f(\theta_{2,x})$ around $\theta_{2,x} = x$ and using Lemma B2, we have $E\left\{\sigma^2(\theta_{2,x}) f(\theta_{2,x})\right\} = \sigma^2(x) f(x) + O(b)$ so that

$$\gamma_{G,0}(0) = \frac{x^{-1/2}}{2\sqrt{\pi}} \sigma^2(x) f(x) + o(1).$$

On the other hand, for a constant a satisfying (A4'), pick a sequence $d_{0T} = \left\lfloor b^{-(1-2/\delta)/(2a)} \right\rfloor$. Then, the second term on the right-hand side of (10) is bounded by

$$\left| \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \gamma_{G,0}(j) \right| \le \sum_{j=1}^{d_{0T}-1} \left|\gamma_{G,0}(j)\right| + \sum_{j=d_{0T}}^{T-1} \left|\gamma_{G,0}(j)\right| \equiv U_1 + U_2.$$

For U_1, using $\epsilon_t = \phi(Y_t) - E\{\phi(Y_t)| X_t\}$ and the law of iterated expectations gives

$$\begin{aligned} \left|\gamma_{G,0}(j)\right| &\le b^{1/2} \left[E\left\{ E\left(|\phi(Y_1)\phi(Y_{1+j})|| X_1, X_{1+j}\right) K_{G(x/b+1,b)}(X_1) K_{G(x/b+1,b)}(X_{1+j}) \right\} \right. \\ &\quad + E\left\{ E\left(|\phi(Y_1)|| X_1, X_{1+j}\right) E\left(|\phi(Y_{1+j})|| X_{1+j}\right) K_{G(x/b+1,b)}(X_1) K_{G(x/b+1,b)}(X_{1+j}) \right\} \\ &\quad + E\left\{ E\left(|\phi(Y_1)|| X_1\right) E\left(|\phi(Y_{1+j})|| X_1, X_{1+j}\right) K_{G(x/b+1,b)}(X_1) K_{G(x/b+1,b)}(X_{1+j}) \right\} \\ &\quad \left. + E\left\{ E\left(|\phi(Y_1)|| X_1\right) E\left(|\phi(Y_{1+j})|| X_{1+j}\right) K_{G(x/b+1,b)}(X_1) K_{G(x/b+1,b)}(X_{1+j}) \right\} \right] \\ &\equiv b^{1/2} (U_{11} + U_{12} + U_{13} + U_{14}). \end{aligned}$$

As indicated in the proof of Lemma B2, $G(x/b+1, b)$ has moments of any nonnegative integer order, and all these moments are $O(1)$. Then, by (A2) and (A3),

$$\begin{aligned} U_{11} &= \int_0^\infty \int_0^\infty E\{|\phi(Y_1)\phi(Y_{1+j})|| X_1 = u, X_{1+j} = v\} \\ &\quad \cdot K_{G(x/b+1,b)}(u) K_{G(x/b+1,b)}(v) f_{1,1+j}(u,v)\, du dv \\ &\le c \int_0^\infty \int_0^\infty (\beta_0 + \beta_1 u^m + \beta_2 v^n) K_{G(x/b+1,b)}(u) K_{G(x/b+1,b)}(v)\, du dv \\ &= O(1). \end{aligned}$$

In addition, using (a conditional moment version of) Hölder's inequality,

$$E\{|\phi(Y_t)|| X_t = u\} = E\{|\phi(Y_t) \cdot 1|| X_t = u\} \le E^{1/\delta}\left\{ |\phi(Y_t)|^\delta \middle| X_t = u \right\} \cdot 1.$$

Without loss of generality, assume $\alpha_0 \ge 1$ so that $\alpha_0 + \alpha_1 u^l \ge \max\left[1, E\left\{ |\phi(Y_t)|^\delta \middle| X_t = u \right\}\right]$. Then, (A3) implies that

$$E\{|\phi(Y_t)|| X_t = u\} \le \left(\alpha_0 + \alpha_1 u^l\right)^{1/\delta} \le \alpha_0 + \alpha_1 u^l. \tag{11}$$

Using (11), (A2) and (A3), we have

$$\begin{aligned} U_{12} &= \int_0^\infty \int_0^\infty E\left\{ \left|\phi\left(Y_1\right)\right| \right| X_1 = u, X_{1+j} = v\} E\left\{ \left|\phi\left(Y_{1+j}\right)\right| \right| X_{1+j} = v\} \\ &\quad \cdot K_{G(x/b+1,b)}\left(u\right) K_{G(x/b+1,b)}\left(v\right) f_{1,1+j}\left(u,v\right) dudv \\ &\leq c \int_0^\infty \int_0^\infty \left(\beta_0 + \beta_1 u^m + \beta_2 v^n\right)\left(\alpha_0 + \alpha v^l\right) K_{G(x/b+1,b)}\left(u\right) K_{G(x/b+1,b)}\left(v\right) dudv \\ &= O\left(1\right). \end{aligned}$$

Similarly, $U_{13} \leq O\left(1\right)$ can be shown. Furthermore, by (11) and (A2),

$$\begin{aligned} U_{14} &= \int_0^\infty \int_0^\infty E\left\{ \left|\phi\left(Y_1\right)\right| \right| X_1 = u\} E\left\{ \left|\phi\left(Y_{1+j}\right)\right| \right| X_{1+j} = v\} \\ &\quad \cdot K_{G(x/b+1,b)}\left(u\right) K_{G(x/b+1,b)}\left(v\right) f_{1,1+j}\left(u,v\right) dudv \\ &\leq c\left\{ \int_0^\infty \left(\alpha_0 + \alpha_1 u^l\right) K_{G(x/b+1,b)}\left(u\right) du \right\}^2 \\ &= O\left(1\right). \end{aligned}$$

Hence, $\left|\gamma_{G,0}\left(j\right)\right| \leq O\left(b^{1/2}\right)$, which establishes that

$$U_1 \leq O\left(d_{0T} b^{1/2}\right) = O\left(b^{\{a-(1-2/\delta)\}/(2a)}\right) \to 0.$$

For U_2, we can apply Davydov's lemma (Corollary A.2 in Hall and Heyde, 1980) to obtain

$$\left|\gamma_{G,0}\left(j\right)\right| \leq 8\left\{\alpha\left(j\right)\right\}^{1-2/\delta} \left\{ E\left| b^{1/4} \epsilon_1 K_{G(x/b+1,b)}\left(X_1\right)\right|^\delta \right\}^{2/\delta}.$$

To find the bound for $E\left|b^{1/4}\epsilon_1 K_{G(x/b+1,b)}\left(X_1\right)\right|^\delta$, note that since $g\left(z\right) = z^\delta$ $\left(z \geq 0\right)$ is increasing and convex,

$$\begin{aligned} \left|x-y\right|^\delta &\leq \left(\left|x\right| + \left|y\right|\right)^\delta \\ &= \left\{ \left(\frac{1}{2}\right)\left(2\left|x\right|\right) + \left(\frac{1}{2}\right)\left(2\left|y\right|\right) \right\}^\delta \\ &\leq \left(\frac{1}{2}\right)\left(2\left|x\right|\right)^\delta + \left(\frac{1}{2}\right)\left(2\left|y\right|\right)^\delta \\ &= 2^{\delta-1}\left(\left|x\right|^\delta + \left|y\right|^\delta\right). \end{aligned}$$

Substituting $x = \phi\left(Y_1\right)$ and $y = E\left\{ \phi\left(Y_1\right) \right| X_1\}$ yields

$$\left|\epsilon_1\right|^\delta \leq 2^{\delta-1}\left[\left|\phi\left(Y_1\right)\right|^\delta + \left|E\left\{ \phi\left(Y_1\right) \right| X_1\}\right|^\delta\right] \leq 2^{\delta-1}\left[\left|\phi\left(Y_1\right)\right|^\delta + E^\delta\left\{ \left|\phi\left(Y_1\right)\right| \right| X_1\}\right]. \quad (12)$$

Then, we have

$$\begin{aligned} E\left|b^{1/4}\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} &\leq cb^{\delta/4}\left[\int_0^\infty E\left\{|\phi(Y_1)|^{\delta}\Big| X_1=u\right\} K^{\delta}_{G(x/b+1,b)}(u) f(u)\,du \right. \\ &\quad \left. +\int_0^\infty E^{\delta}\left\{|\phi(Y_1)|| X_1=u\right\} K^{\delta}_{G(x/b+1,b)}(u) f(u)\,du\right] \\ &\equiv cb^{\delta/4}(U_{21}+U_{22}). \end{aligned}$$

Again, as argued in the proof of Lemma B2, $G(\delta x/b+1, b/\delta)$ has moments of any nonnegative integer order and all these moments are $O(1)$. Then, it follows from Lemma B1, (A2), (A3), and (11) that each of U_{21} and U_{22} is bounded by

$$cA_{b,\delta}(x)\int_0^\infty \left(\alpha_0+\alpha_1 u^l\right) K_{G(\delta x/b+1,b/\delta)}(u)\,du \leq O\{A_{b,\delta}(x)\} = O\left(b^{(1-\delta)/2}\right).$$

Therefore, $E\left|b^{1/4}\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} \leq O\left(b^{1/2-\delta/4}\right)$, and thus

$$U_2 \leq O\left(b^{1/\delta-1/2}\right)\sum_{j=d_{0T}}^{T-1}\{\alpha(j)\}^{1-2/\delta} \leq O\left(b^{-(1-2/\delta)/2}\right) d_{0T}^{-a}\sum_{j=d_{0T}}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} \to 0,$$

because $O\left(b^{-(1-2/\delta)/2}\right) d_{0T}^{-a} = O(1)$, $d_{0T}\to\infty$, and $\sum_{j=1}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} < \infty$. This completes the proof of this part.

Remark. We can demonstrate (7) even after replacing (A4') by a weaker condition (A4). Observe that given (A4) and $d_{0T} = \left\lfloor b^{-(1-2/\delta)/(2a)}\right\rfloor$, each of U_1 and U_2 still becomes $o(1)$.

(ii) Proof of (8). We have

$$\begin{aligned} Var\left\{\sqrt{b^{1/2}T} b^{-1/2} T^*_{G,1}(x)\right\} &= Var\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{-1/4}(X_t-x)\epsilon_t K_{G(x/b+1,b)}(X_t)\right\} \\ &\equiv \gamma_{G,1}(0) + 2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_{G,1}(j), \end{aligned} \tag{13}$$

where $\gamma_{G,1}(j) = b^{-1/2}E\left\{(X_1-x)(X_{1+j}-x)\epsilon_1\epsilon_{1+j}K_{G(x/b+1,b)}(X_1)K_{G(x/b+1,b)}(X_{1+j})\right\}$ is the j^{th}-order autocovariance of the stationary process $\left\{b^{-1/4}(X_t-x)\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$. By the law of iterated expectations and Lemma B1, the first term on the right-hand side of (13) reduces to

$$\begin{aligned} \gamma_{G,1}(0) &= b^{-1/2}E\left\{(X_1-x)^2\sigma^2(X_1)K^2_{G(x/b+1,b)}(X_1)\right\} \\ &= b^{-1/2}A_{b,2}(x)E\left\{(\theta_{2,x}-x)^2\sigma^2(\theta_{2,x})f(\theta_{2,x})\right\} \\ &= b^{-1/2}\left\{\frac{b^{-1/2}x^{-1/2}}{2\sqrt{\pi}} + o\left(b^{-1/2}\right)\right\}E\left\{(\theta_{2,x}-x)^2\sigma^2(\theta_{2,x})f(\theta_{2,x})\right\}, \end{aligned}$$

where $\theta_{2,x} \overset{d}{=} G(2x/b+1, b/2)$. By a Taylor expansion and Lemma B2, we have

$$E\left\{(\theta_{2,x}-x)^2 \sigma^2(\theta_{2,x}) f(\theta_{2,x})\right\} = \left(\frac{xb+b^2}{2}\right)\sigma^2(x) f(x) + O\left(b^2\right)$$

so that

$$\gamma_{G,1}(0) = \frac{x^{1/2}}{4\sqrt{\pi}}\sigma^2(x) f(x) + o(1).$$

On the other hand, the second term on the right-hand side of (13) is bounded by

$$\left|\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_{G,1}(j)\right| \leq \sum_{j=1}^{d_{1T}-1}\left|\gamma_{G,1}(j)\right| + \sum_{j=d_{1T}}^{T-1}\left|\gamma_{G,1}(j)\right| \equiv V_1 + V_2,$$

where the sequence d_{1T} is defined as $d_{1T} = \left\lfloor b^{-3(1-2/\delta)/(2a)}\right\rfloor$ for a constant a satisfying (A4'). For V_1, the same logic as in part (i) yields

$$\begin{aligned}
\left|\gamma_{G,1}(j)\right| &\leq b^{-1/2}E\left[E\left\{\left|\phi(Y_1)\phi(Y_{1+j})\right|\right| X_1, X_{1+j}\right\}\left|X_1-x\right| K_{G(x/b+1,b)}(X_1) \\
&\quad \cdot\left|X_{1+j}-x\right| K_{G(x/b+1,b)}(X_{1+j})\big] \\
&\quad + b^{-1/2}E\left[E\left\{\left|\phi(Y_1)\right|\right| X_1, X_{1+j}\right\}\left|X_1-x\right| K_{G(x/b+1,b)}(X_1) \\
&\quad \cdot E\left\{\left|\phi(Y_{1+j})\right|\right| X_{1+j}\right\}\left|X_{1+j}-x\right| K_{G(x/b+1,b)}(X_{1+j})\big] \\
&\quad + b^{-1/2}E\left[E\left\{\left|\phi(Y_1)\right|\right| X_1\right\}\left|X_1-x\right| K_{G(x/b+1,b)}(X_1) \\
&\quad \cdot E\left\{\left|\phi(Y_{1+j})\right|\right| X_1, X_{1+j}\right\}\left|X_{1+j}-x\right| K_{G(x/b+1,b)}(X_{1+j})\big] \\
&\quad + b^{-1/2}E\left[E\left\{\left|\phi(Y_1)\right|\right| X_1\right\}\left|X_1-x\right| K_{G(x/b+1,b)}(X_1) \\
&\quad \cdot E\left\{\left|\phi(Y_{1+j})\right|\right| X_{1+j}\right\}\left|X_{1+j}-x\right| K_{G(x/b+1,b)}(X_{1+j})\big] \\
&\equiv V_{11} + V_{12} + V_{13} + V_{14}.
\end{aligned}$$

Observe that by (A2), we have $f^{-1}(u) \leq m_1^{-1}$ so that

$$\frac{f_{1,1+j}(u,v)}{f(u)f(v)} \leq \frac{M_2}{m_1^2} \Rightarrow f_{1,1+j}(u,v) \leq cf(u)f(v). \tag{14}$$

Using (A3),

$$\begin{aligned}
V_{11} &\leq cb^{-1/2}\left[\left\{\int_0^\infty (\beta_0+\beta_1 u^m)\left|u-x\right| K_{G(x/b+1,b)}(u) f(u)\,du\right\}\left\{\int_0^\infty \left|v-x\right| K_{G(x/b+1,b)}(v) f(v)\,dv\right\}\right. \\
&\quad \left. + \left\{\int_0^\infty \beta_2 v^n \left|v-x\right| K_{G(x/b+1,b)}(v) f(v)\,dv\right\}\left\{\int_0^\infty \left|u-x\right| K_{G(x/b+1,b)}(u) f(u)\,du\right\}\right] \\
&\equiv cb^{-1/2}\left(V_{111}V_{112} + V_{113}V_{114}\right).
\end{aligned}$$

The Cauchy-Schwarz inequality implies that

$$\begin{aligned} V_{111} &\le \left\{\beta_0 \int_0^\infty (u-x)^2 K_{G(x/b+1,b)}(u) f(u)\, du + \beta_1 \int_0^\infty u^m (u-x)^2 K_{G(x/b+1,b)}(u) f(u)\, du\right\}^{1/2} \\ &\quad \cdot \left\{\int_0^\infty (\beta_0 + \beta_1 u^m) K_{G(x/b+1,b)}(u) f(u)\, du\right\}^{1/2} \\ &\equiv (\beta_0 V_{1111} + \beta_1 V_{1112})^{1/2} V_{1113}^{1/2}. \end{aligned}$$

By a Taylor expansion and Lemma B2, we have $V_{1111} = O(b)$ and $V_{1112} = O(b)$. In addition, $V_{1113} = O(1)$, and thus $V_{111} \le O(b^{1/2})$. Similarly, each of V_{112}, V_{113} and V_{114} is at most $O(b^{1/2})$. Hence, $V_{11} \le O(b^{1/2})$. Applying the same procedure, we can also demonstrate that each of V_{12}, V_{13} and V_{14} is bounded by $O(b^{1/2})$. Hence, we can conclude that $|\gamma_{G,1}(j)| \le O(b^{1/2})$, which establishes that

$$V_1 \le O\left(d_{1T} b^{1/2}\right) = O\left(b^{\{a-3(1-2/\delta)\}/(2a)}\right) \to 0.$$

For V_2, we can apply again Davydov's lemma to obtain

$$|\gamma_{G,1}(j)| \le 8\{\alpha(j)\}^{1-2/\delta} \left\{E\left|b^{-1/4}(X_1-x)\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta}\right\}^{2/\delta}.$$

It follows from (11), (12), (14), and Lemma B1 that

$$\begin{aligned} & E\left|b^{-1/4}(X_1-x)\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} \\ \le\ & cb^{-\delta/4}\left[\int_0^\infty |u-x|^\delta E\left\{|\phi(Y_1)|^\delta \,\middle|\, X_1 = u\right\} K^\delta_{G(x/b+1,b)}(u) f(u)\, du \right. \\ & \left. + \int_0^\infty |u-x|^\delta E^\delta\{|\phi(Y_1)|| X_1 = u\} K^\delta_{G(x/b+1,b)}(u) f(u)\, du\right] \\ \le\ & cb^{-\delta/4} A_{b,\delta}(x) \int_0^\infty |u-x|^\delta \left(\alpha_0 + \alpha_1 u^l\right) K_{G(\delta x/b+1,b/\delta)}(u) f(u)\, du \\ \equiv\ & cb^{-\delta/4} A_{b,\delta}(x) V_{21}. \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} V_{21} &\le \left\{\alpha_0 \int_0^\infty |u-x|^{2\delta} K_{G(\delta x/b+1,b/\delta)}(u) f(u)\, du + \alpha_1 \int_0^\infty u^l |u-x|^{2\delta} K_{G(\delta x/b+1,b/\delta)}(u) f(u)\, du\right\}^{1/2} \\ &\quad \cdot \left\{\int_0^\infty \left(\alpha_0 + \alpha_1 u^l\right) K_{G(\delta x/b+1,b/\delta)}(u) f(u)\, du\right\}^{1/2} \\ &\equiv (\alpha_0 V_{211} + \alpha_1 V_{212})^{1/2} V_{213}^{1/2}. \end{aligned}$$

Recall that $2\delta > 4$. Hence, by Lemma B2, each of V_{211} and V_{212} is at most $O(b^2)$. Clearly,

$V_{213} = O(1)$, and thus we have $V_{21} \leq O(b)$ so that

$$E\left|b^{-1/4}(X_1 - x)\,\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} \leq b^{-\delta/4} O\left(b^{(1-\delta)/2}\right) O(b) = O\left(b^{3/2-3\delta/4}\right).$$

Therefore,

$$V_2 \leq O\left(b^{3/\delta-3/2}\right)\sum_{j=d_{1T}}^{T-1}\{\alpha(j)\}^{1-2/\delta} \leq O\left(b^{-3(1-2/\delta)/2}\right) d_{1T}^{-a}\sum_{j=d_{1T}}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} \to 0,$$

because $O\left(b^{-3(1-2/\delta)/2}\right) d_{1T}^{-a} = O(1)$, $d_{1T} \to \infty$, and $\sum_{j=1}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} < \infty$. This completes the proof of this part.

(iii) Proof of (9). We have

$$\begin{aligned}
& Cov\left\{\sqrt{b^{1/2}T}T_{G,0}^{*}(x), \sqrt{b^{1/2}T}b^{-1/2}T_{G,1}^{*}(x)\right\} \\
= \; & Cov\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{1/4}\epsilon_t K_{G(x/b+1,b)}(X_t), \frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{-1/4}(X_t - x)\,\epsilon_t K_{G(x/b+1,b)}(X_t)\right\} \\
\equiv \; & \gamma_{G,3}(0) + 2\sum_{j=1}^{T-1}\left(1 - \frac{j}{T}\right)\gamma_{G,3}(j), \qquad (15)
\end{aligned}$$

where $\gamma_{G,3}(j) = E\left\{(X_{1+j} - x)\,\epsilon_1\epsilon_{1+j} K_{G(x/b+1,b)}(X_1)\, K_{G(x/b+1,b)}(X_{1+j})\right\}$ is the j^{th}-order cross-covariance of the stationary processes $\left\{b^{1/4}\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$ and $\left\{b^{-1/4}(X_t - x)\,\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$. By the law of iterated expectations and Lemma B1, the first term on the right-hand side of (15) reduces to

$$\begin{aligned}
\gamma_{G,3}(0) &= E\left\{(X_1 - x)\,\sigma^2(X_1)\, K_{G(x/b+1,b)}^2(X_1)\right\} \\
&= A_{b,2}(x)\, E\left\{(\theta_{2,x} - x)\,\sigma^2(\theta_{2,x})\, f(\theta_{2,x})\right\} \\
&= \left\{\frac{b^{-1/2}x^{-1/2}}{2\sqrt{\pi}} + o\left(b^{-1/2}\right)\right\} E\left\{(\theta_{2,x} - x)\,\sigma^2(\theta_{2,x})\, f(\theta_{2,x})\right\},
\end{aligned}$$

where $\theta_{2,x} \stackrel{d}{=} G(2x/b+1, b/2)$. By a Taylor expansion and Lemma B2, we can see that

$$E\left\{(\theta_{2,x} - x)\,\sigma^2(\theta_{2,x})\, f(\theta_{2,x})\right\} = O(b),$$

and thus $\gamma_{G,3}(0) = o\left(b^{1/2}\right) = o(1)$. On the other hand, applying the same procedures as in parts (i) and (ii), we can also establish that the second term on the right-hand side of (15) is $o(1)$. This completes the proof. ■

Lemma A3. *If the conditions (A1)-(A3), (A4'), (A5)-(A6) hold, then for interior x,*

$$\sqrt{b^{1/2}T}Q_{G,T}^{*} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{G}\mathbf{c}\right), \sqrt{b^{1/2}T}Q_{IG,T}^{*} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{IG}\mathbf{c}\right), \sqrt{b^{1/2}T}Q_{RIG,T}^{*} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{RIG}\mathbf{c}\right).$$

Proof of Lemma A3. We employ the small-block and large-block argument. Partition the set $\{1, \ldots, T\}$ into $2q_T + 1$ subsets with large block of size r_T and small block of s_T. Put

$$q_T = \left\lfloor \frac{T}{r_T + s_T} \right\rfloor.$$

Also let $\varsigma_{G,j}^{*} = \mathbf{c}^{\intercal}\mathbf{Z}_{G,j}^{*}$, where

$$\mathbf{Z}_{G,j}^{*} = \begin{bmatrix} b^{1/4}\epsilon_{j+1}K_{G(x/b+1,b)}\left(X_{j+1}\right) \\ b^{-1/4}\left(X_{j+1} - x\right)\epsilon_{j+1}K_{G(x/b+1,b)}\left(X_{j+1}\right) \end{bmatrix}$$

for $j = 0, \ldots, T-1$. Then,

$$\sqrt{b^{1/2}T}Q_{G,T}^{*} = \frac{1}{\sqrt{T}}\sum_{j=0}^{T-1}\varsigma_{G,j}^{*}.$$

Furthermore, define the random variables, for $0 \le j \le q_T - 1$,

$$\eta_{G,j}^{*} = \sum_{i=j(r_T+s_T)}^{j(r_T+s_T)+r_T-1}\varsigma_{G,i}^{*}, \; \xi_{G,j}^{*} = \sum_{i=j(r_T+s_T)+r_T}^{(j+1)(r_T+s_T)-1}\varsigma_{G,i}^{*}, \; \xi_{G,q}^{*} = \sum_{i=q_T(r_T+s_T)}^{T-1}\varsigma_{G,i}^{*}.$$

It follows that

$$\sqrt{b^{1/2}T}Q_{G,T}^{*} = \frac{1}{\sqrt{T}}\left(\sum_{j=0}^{q_T-1}\eta_{G,j}^{*} + \sum_{j=0}^{q_T-1}\xi_{G,j}^{*} + \xi_{G,q}^{*}\right) \equiv \frac{1}{\sqrt{T}}\left(Q_{G,T,1} + Q_{G,T,2} + Q_{G,T,3}\right).$$

We will show that

$$\frac{1}{T}E\left(Q_{G,T,2}^{2}\right) \to 0, \tag{16}$$

$$\frac{1}{T}E\left(Q_{G,T,3}^{2}\right) \to 0, \tag{17}$$

$$\left|E\left\{\exp\left(itQ_{G,T,1}\right)\right\} - \prod_{j=0}^{q_T-1}E\left\{\exp\left(it\eta_{G,j}^{*}\right)\right\}\right| \to 0, \tag{18}$$

$$\frac{1}{T}\sum_{j=0}^{q_T-1}E\left(\eta_{G,j}^{*2}\right) \to \mathbf{c}^{\intercal}\mathbf{V}_{G}\mathbf{c}, \tag{19}$$

$$\frac{1}{T}\sum_{j=0}^{q_T-1}E\left[\eta_{G,j}^{*2}\mathbf{1}\left\{\left|\eta_{G,j}^{*}\right| \ge \epsilon\left(\mathbf{c}^{\intercal}\mathbf{V}_{G}\mathbf{c}\right)^{1/2}\sqrt{T}\right\}\right] \to 0 \tag{20}$$

for every $\epsilon > 0$. (16) and (17) imply that $Q_{G,T.2}$ and $Q_{G,T,3}$ are asymptotically negligible, (18) implies that the summands $\left\{\eta_{G,j}^{*}\right\}$ in $Q_{G,T,1}$ are asymptotically mutually independent, and (19) and (20) are the standard Lindeberg-Feller conditions for asymptotic normality of $Q_{G,T,1}$ under independence. Hence, the lemma follows if we can show (16)-(20).

We first choose the block sizes. (A6) implies that a sequence $\gamma_T \in \mathbb{N}$ such that $\gamma_T \to \infty$, $\gamma_T s_T / \left(b^{1/2}T\right)^{1/2} \to 0$, and $\gamma_T \left(T/b^{1/2}\right)^{1/2} \alpha\left(s_T\right) \to 0$ as $T \to \infty$. Define the large-block size r_T by

$$r_T = \left\lfloor \frac{\left(b^{1/2}T\right)^{1/2}}{\gamma_T} \right\rfloor$$

and the small-block size by s_T. It follows that

$$\frac{s_T}{r_T} \to 0,\ \frac{r_T}{T} \to 0,\ \frac{r_T}{\left(b^{1/2}T\right)^{1/2}} \to 0,\ \frac{T}{r_T}\alpha\left(s_T\right) \to 0 \tag{21}$$

as $T \to \infty$. The proofs of (16)-(20) are given subsequently.

(i) Proof of (16). Observe that

$$E\left(Q_{G,T,2}^2\right) = \sum_{j=0}^{q_T-1} Var\left(\xi_{G,j}^*\right) + \sum_{i=0}^{q_T-1} \sum_{j=0,\, j\neq i}^{q_T-1} Cov\left(\xi_{G,i}^*, \xi_{G,j}^*\right) \equiv F_1 + F_2.$$

For F_1, it follows from stationarity and Lemma A2 that

$$F_1 = q_T Var\left(\sum_{j=1}^{s_T} \varsigma_{G,j}^*\right) = q_T s_T \left\{\mathbf{c}^\intercal \mathbf{V}_G \mathbf{c} + o\left(1\right)\right\} = O\left(q_T s_T\right).$$

On the other hand, F_2 can be further rewritten as

$$F_2 = \sum_{i=0}^{q_T-1} \sum_{j=0,\, j\neq i}^{q_T-1} \sum_{l_1=0}^{s_T-1} \sum_{l_2=0}^{s_T-1} Cov\left(\xi_{G,m_i+l_1}^*, \xi_{G,m_j+l_2}^*\right),$$

where $m_j = j\left(r_T + s_T\right) + r_T$. Since $i \neq j$, we have $\left|\left(m_i + l_1\right) - \left(m_j + l_2\right)\right| \geq r_T$. Then, by stationarity,

$$\left|F_2\right| \leq 2 \sum_{l_1=0}^{T-r_T-1} \sum_{l_2=l_1+r_T}^{T-1} \left|Cov\left(\xi_{G,l_1}^*, \xi_{G,l_2}^*\right)\right| \leq 2T \sum_{j=r_T}^{T-1} \left|Cov\left(\xi_{G,0}^*, \xi_{G,j}^*\right)\right|.$$

Note that the arguments used in the proof of Lemma A2 imply that $\sum_{j=r_T}^{T-1} \left|Cov\left(\xi_{G,0}^*, \xi_{G,j}^*\right)\right| = o\left(1\right)$. Therefore, $\left|F_2\right| \leq o\left(T\right)$, and thus, by (21),

$$\frac{1}{T} E\left(Q_{G,T,2}^2\right) = O\left(\frac{q_T s_T}{T}\right) + o\left(1\right) = O\left(\frac{s_T}{r_T + s_T}\right) + o\left(1\right) \to 0.$$

(ii) Proof of (17). Using a similar argument to the one used in the proof of (16), we have, by (21),

$$\begin{aligned}\frac{1}{T}E\left(Q_{G,T,3}^{2}\right) &\leq \frac{1}{T}\left\{T-q_T\left(r_T+s_T\right)\right\}Var\left(\varsigma_{G,0}^{*}\right)+2\sum_{j=0}^{T-1}\left|Cov\left(\xi_{G,0}^{*},\xi_{G,j}^{*}\right)\right| \\ &= o\left(1\right)\left\{\mathbf{c}^{\intercal}\mathbf{V}_G\mathbf{c}+o\left(1\right)\right\}+o\left(1\right)\to 0.\end{aligned}$$

(iii) Proof of (18). Observe that $\eta_{G,a}^{*}$ is $\mathcal{F}_{i_a}^{j_a}$-measurable with $i_a = a\left(r_T+s_T\right)+1$ and $j_a = a\left(r_T+s_T\right)+r_T$. Applying Lemma B3 with $V_j=\exp\left(it\eta_{G,j}^{*}\right)$ and (21) yields

$$\left|E\left\{\exp\left(itQ_{G,T,1}\right)\right\}-\prod_{j=0}^{q_T-1}E\left\{\exp\left(it\eta_{G,j}^{*}\right)\right\}\right|\leq 16q_T\alpha\left(s_T+1\right)\sim 16\left(\frac{T}{r_T+s_T}\right)\alpha\left(s_T+1\right)\to 0.$$

(iv) Proof of (19). By stationarity and Lemma A2, we have

$$E\left(\eta_{G,j}^{*2}\right)=Var\left(\eta_{G,j}^{*}\right)=r_T\left\{\mathbf{c}^{\intercal}\mathbf{V}_G\mathbf{c}+o\left(1\right)\right\}.$$

Therefore, by (21),

$$\frac{1}{T}\sum_{j=0}^{q_T-1}E\left(\eta_{G,j}^{*2}\right)=\frac{q_Tr_T}{T}\left\{\mathbf{c}^{\intercal}\mathbf{V}_G\mathbf{c}+o\left(1\right)\right\}\sim\left(\frac{r_T}{r_T+s_T}\right)\mathbf{c}^{\intercal}\mathbf{V}_G\mathbf{c}\to\mathbf{c}^{\intercal}\mathbf{V}_G\mathbf{c}.$$

(v) Proof of (20). We employ a truncation argument because ϵ_j is not necessarily bounded. Let $\epsilon_j^L=\epsilon_j\mathbf{1}\left\{\left|\epsilon_j\right|\leq L\right\}$ for some fixed truncation point $L>0$. Also let $\varsigma_{G,j}^{*L}=\mathbf{c}^{\intercal}\mathbf{Z}_{G,j}^{*L}$, where

$$\mathbf{Z}_{G,j}^{*L}=\begin{bmatrix} b^{1/4}\epsilon_{j+1}^{L}K_{G(x/b+1,b)}\left(X_{j+1}\right) \\ b^{-1/4}\left(X_{j+1}-x\right)\epsilon_{j+1}^{L}K_{G(x/b+1,b)}\left(X_{j+1}\right)\end{bmatrix}.$$

Furthermore, define

$$Q_{G,T}^{*L}=\frac{1}{b^{1/4}T}\sum_{j=0}^{T-1}\varsigma_{G,j}^{*L},\ \eta_{G,j}^{*L}=\sum_{i=j(r_T+s_T)}^{j(r_T+s_T)+r_T-1}\varsigma_{G,i}^{*L}.$$

In addition, let $\tilde{\varsigma}_{G,j}^{*L}=\mathbf{c}^{\intercal}\tilde{\mathbf{Z}}_{G,j}^{*L}$, where

$$\tilde{\mathbf{Z}}_{G,j}^{*L}=\begin{bmatrix} b^{1/4}\tilde{\epsilon}_{j+1}^{L}K_{G(x/b+1,b)}\left(X_{j+1}\right) \\ b^{-1/4}\left(X_{j+1}-x\right)\tilde{\epsilon}_{j+1}^{L}K_{G(x/b+1,b)}\left(X_{j+1}\right)\end{bmatrix}$$

and $\tilde{\epsilon}_j^L=\epsilon_j\mathbf{1}\left\{\left|\epsilon_j\right|>L\right\}$. Finally, define

$$\tilde{Q}_{G,T}^{*L}=\frac{1}{b^{1/4}T}\sum_{j=0}^{T-1}\tilde{\varsigma}_{G,j}^{*L}$$

so that $Q_{G,T}^{*}=Q_{G,T}^{*L}+\tilde{Q}_{G,T}^{*L}$.

Since both $K_{G(x/b+1,b)}(u)$ and $uK_{G(x/b+1,b)}(u)$ are bounded above, we have

$$\left|(X_{j+1}-x)\,K_{G(x/b+1,b)}(X_{j+1})\right|<\infty \tag{22}$$

for $j=0,\ldots,T-1$ so that $\left|\varsigma^{*L}_{G,j}\right|\le cLb^{-1/4}$. Then,

$$\left|\eta^{*L}_{G,j}\right|\le cLr_T b^{-1/4}, \tag{23}$$

and thus, by (21),

$$\frac{\left|\eta^{*L}_{G,j}\right|}{\sqrt{T}}\le c\frac{r_T}{\sqrt{b^{1/2}T}}\to 0.$$

It follows that

$$\Pr\left\{\left|\eta^{*L}_{G,j}\right|\ge\epsilon\left(\mathbf{c}^\intercal\mathbf{V}_G\mathbf{c}\right)^{1/2}\sqrt{T}\right\}=0 \tag{24}$$

at all j for sufficiently large T. Then, applying (23) and (24), we have

$$\begin{aligned}&\frac{1}{T}\sum_{j=0}^{q_T-1}E\left[\left|\eta^{*L}_{G,j}\right|^2\mathbf{1}\left\{\left|\eta^{*L}_{G,j}\right|\ge\epsilon\left(\mathbf{c}^\intercal\mathbf{V}_G\mathbf{c}\right)^{1/2}\sqrt{T}\right\}\right]\\ \le\;& c\left(\frac{r_T}{\sqrt{b^{1/2}T}}\right)^2\sum_{j=0}^{q_T-1}\Pr\left\{\left|\eta^{*L}_{G,j}\right|\ge\epsilon\left(\mathbf{c}^\intercal\mathbf{V}_G\mathbf{c}\right)^{1/2}\sqrt{T}\right\}\to 0.\end{aligned}$$

In other words, (20) holds for the truncated variables. Consequently, we have the following asymptotic normality result

$$\sqrt{b^{1/2}T}Q^{*L}_{G,T}=\frac{1}{\sqrt{T}}\sum_{j=0}^{T-1}\varsigma^{*L}_{G,j}\xrightarrow{d}N\left(0,\mathbf{c}^\intercal\mathbf{V}^L_G\mathbf{c}\right), \tag{25}$$

where $\mathbf{V}^L_G=Var\left(\mathbf{Z}^{*L}_{G,j}\middle|\,X_j=x\right)$.

The remaining task for establishing (20) is to show that as first $T\to\infty$ and then $L\to\infty$,

$$b^{1/2}TVar\left(\tilde{Q}^{*L}_{G,T}\right)\to 0. \tag{26}$$

Indeed,

$$\begin{aligned}&\left|E\left\{\exp\left(it\sqrt{b^{1/2}T}Q^{*}_{G,T}\right)\right\}-\exp\left(-\frac{t^2}{2}\mathbf{c}^\intercal\mathbf{V}_G\mathbf{c}\right)\right|\\ \le\;&\left|E\left\{\exp\left(it\sqrt{b^{1/2}T}Q^{*L}_{G,T}\right)\right\}-\exp\left(-\frac{t^2}{2}\mathbf{c}^\intercal\mathbf{V}^L_G\mathbf{c}\right)\right|+\left|E\left\{\exp\left(it\sqrt{b^{1/2}T}\tilde{Q}^{*L}_{G,T}\right)\right\}-1\right|\\ &+\left|\exp\left(-\frac{t^2}{2}\mathbf{c}^\intercal\mathbf{V}^L_G\mathbf{c}\right)-\exp\left(-\frac{t^2}{2}\mathbf{c}^\intercal\mathbf{V}_G\mathbf{c}\right)\right|\\ \equiv\;&E_1+E_2+E_3.\end{aligned}$$

By (25), $E_1 \to 0$ as $T \to 0$ for every $L > 0$. $E_3 \to 0$ as first $T \to \infty$ and then $L \to \infty$, because $\mathbf{V}_G^L \to Var\left(\mathbf{Z}_{G,j}^* \middle| X_j = x\right) = \mathbf{V}_G$ by the dominated convergence theorem. We can also see that $E_2 \to 0$ as first $T \to \infty$ and then $L \to \infty$, if (26) holds. Now,

$$\lim_{T\to\infty} b^{1/2} T Var\left(\tilde{Q}_{G,T}^{*L}\right) = \mathbf{c}^\intercal Var\left(\tilde{\mathbf{Z}}_{G,j}^{*L} \middle| X_j = x\right) \mathbf{c} = \mathbf{c}^\intercal Var\left(\mathbf{Z}_{G,j}^* 1\left\{|\epsilon_j| > L\right\} \middle| X_j = x\right) \mathbf{c} \to 0$$

as $L \to \infty$ by the dominated convergence theorem. This completes the proof. ■

A.1.1 Proof of Theorem 1

Lemma A3 implies that

$$\sqrt{b^{1/2}T}\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}\left(X_t\right)\epsilon_t\right\} \xrightarrow{d} N\left(0, \frac{\sigma^2(x) f(x)}{2\sqrt{\pi}x^{1/2}}\right).$$

Also, define

$$\tilde{m}_G^{nw}(x) = \frac{\sum_{t=1}^{T} K_{G(x/b+1,b)}\left(X_t\right) m\left(X_t\right)}{\sum_{t=1}^{T} K_{G(x/b+1,b)}\left(X_t\right)} = S_{G,0}^{-1}(x)\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}\left(X_t\right) m\left(X_t\right)\right\}.$$

Then, by the definitions of $\hat{m}_G^{nw}(x)$ and ϵ_t,

$$\hat{m}_G^{nw}(x) - \tilde{m}_G^{nw}(x) = S_{G,0}^{-1}(x)\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}\left(X_t\right)\epsilon_t\right\}.$$

Therefore, by Slutsky's lemma and Lemma A1, we have

$$\sqrt{b^{1/2}T}\left\{\hat{m}_G^{nw}(x) - \tilde{m}_G^{nw}(x)\right\} \xrightarrow{d} N\left(0, \frac{1}{2\sqrt{\pi}x^{1/2}}\frac{\sigma^2(x)}{f(x)}\right). \tag{27}$$

In addition, a second-order Taylor expansion yields

$$\tilde{m}_G^{nw}(x) = m(x) + m'(x) S_{G,0}^{-1}(x) S_{G,1}(x) + \frac{m''(x)}{2} S_{G,0}^{-1}(x) S_{G,2}(x) + O_p\left\{S_{G,3}(x)\right\}, \tag{28}$$

where

$$S_{G,1}(x) = \left\{f(x) + x f'(x)\right\} b + o_p(b), \; S_{G,2}(x) = x f(x) b + o_p(b), \; S_{G,3}(x) = O_p\left(b^2\right) \tag{29}$$

by Lemma B2 and the ergodic theorem. Substituting (29) into (28) and using Lemma A1 and (A7), we can see that the left-hand side of (27) can be approximated by

$$\begin{aligned}
& \sqrt{b^{1/2}T}\left\{\hat{m}_G^{nw}(x) - \tilde{m}_G^{nw}(x)\right\} \\
= & \sqrt{b^{1/2}T}\left[\hat{m}_G^{nw}(x) - m(x) - \left\{m'(x)\left(1 + \frac{x f'(x)}{f(x)}\right) + \frac{1}{2} x m''(x)\right\} b + o_p(b)\right] \\
= & \sqrt{b^{1/2}T}\left[\hat{m}_G^{nw}(x) - m(x) - \left\{m'(x)\left(1 + \frac{x f'(x)}{f(x)}\right) + \frac{1}{2} x m''(x)\right\} b\right] + o_p(1).
\end{aligned}$$

This completes the proof. ■

A.1.2 Proof of Theorem 2

Lemma A3 and the Cramér-Wald device imply that $\sqrt{b^{1/2}T}\mathbf{t}^*_{G,T} \xrightarrow{d} N(\mathbf{0}_2, \mathbf{V}_G)$, where $\mathbf{0}_2$ is the 2×1 zero vector. Also, define

$$\tilde{\boldsymbol{\beta}}_G(x) = \left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix}\begin{bmatrix} 1 & X_t - x \end{bmatrix}\right\}^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix} m(X_t)\right\}.$$

Then, by the definitions of $\hat{\boldsymbol{\beta}}_G(x)$ and ϵ_t,

$$\begin{aligned}
& \begin{bmatrix} 1 & 0 \\ 0 & b^{1/2} \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} \\
= & \begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix}\begin{bmatrix} 1 & X_t - x \end{bmatrix}\right\}^{-1} \\
& \cdot\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix}\epsilon_t\right\} \\
= & \begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix}\begin{bmatrix} 1 & X_t - x \end{bmatrix}\right\}^{-1} \\
& \cdot\begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}\left\{\frac{1}{T}\sum_{t=1}^{T} K_{G(x/b+1,b)}(X_t)\begin{bmatrix} 1 \\ X_t - x \end{bmatrix}\epsilon_t\right\} \\
= & \mathbf{S}^{-1}_{G,T}\mathbf{t}^*_{G,T}.
\end{aligned}$$

Therefore, by Slutsky's lemma and Lemma A1, we have

$$\sqrt{b^{1/2}T}\begin{bmatrix} 1 & 0 \\ 0 & b^{1/2} \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} \xrightarrow{d} N\left(\mathbf{0}_2, \mathbf{S}^{-1}_G\mathbf{V}_G\mathbf{S}^{-1}_G\right). \tag{30}$$

In addition, a second-order Taylor expansion yields

$$\begin{aligned}
& \tilde{\boldsymbol{\beta}}_G(x) \\
= & \boldsymbol{\beta}(x) + \frac{m''(x)}{2}\begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}\mathbf{S}^{-1}_{G,T}\begin{bmatrix} 1 & 0 \\ 0 & b^{-1/2} \end{bmatrix}\begin{bmatrix} S_{G,2}(x) \\ S_{G,3}(x) \end{bmatrix} + \begin{bmatrix} O_p\{S_{G,3}(x)\} \\ O_p\left\{b^{-1}S_{G,4}(x)\right\} \end{bmatrix},
\end{aligned} \tag{31}$$

where

$$S_{G,2}(x) = xf(x)b + o_p(b),\ S_{G,3}(x) = O_p\left(b^2\right),\ S_{G,4}(x) = O_p\left(b^2\right) \tag{32}$$

by Lemma B2 and the ergodic theorem. Substituting (32) into (31) and using Lemma A1 and (A7), the left-hand side of (30) can be approximated by

$$\begin{aligned}
&\sqrt{b^{1/2}T}\begin{bmatrix} 1 & 0 \\ 0 & b^{1/2} \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} \\
= &\sqrt{b^{1/2}T}\begin{bmatrix} 1 & 0 \\ 0 & b^{1/2} \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \boldsymbol{\beta}(x) - \frac{m''(x)}{2}\begin{bmatrix} xb \\ O_p(b) \end{bmatrix} + \begin{bmatrix} o_p(b) \\ O_p(b) \end{bmatrix}\right\},
\end{aligned}$$

where $\sqrt{b^{1/2}T}o_p(b) = o_p\left(\sqrt{b^{5/2}T}\right) = o_p(1)$, and $\sqrt{b^{1/2}T}b^{1/2}O_p(b) = b^{1/2}O_p\left(\sqrt{b^{5/2}T}\right) = o_p(1)$. Therefore, (30) can be rewritten as

$$T_{b,1}\left\{\hat{\boldsymbol{\beta}}_G(x) - \boldsymbol{\beta}(x) - \begin{bmatrix} \frac{1}{2}xm''(x)b \\ 0 \end{bmatrix}\right\} + \begin{bmatrix} o_p(1) \\ o_p(1) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2x} \end{bmatrix} V_G\right)$$

by letting $T_{b,1} = \sqrt{b^{1/2}T}\begin{bmatrix} 1 & 0 \\ 0 & b^{1/2} \end{bmatrix}$, and $V_G = \frac{1}{2\sqrt{\pi}x^{1/2}}\frac{\sigma^2(x)}{f(x)}$. This completes the proof. ■

A.2 Proof of Theorem 3

The proof of Theorem 3 requires the following three lemmata. Before proceeding, we introduce some additional notation. For interior x, define

$$\begin{aligned}
b_{G,t}(x) &= \left[1 + 4\sqrt{\pi}\left\{\frac{1}{x^{1/2}} + \frac{x^{1/2}f'(x)}{f(x)}\right\}b^{1/2}(X_t - x)K_{G(x/b+1,b)}(X_t)\right]^{-1}, \\
b_{IG,t}(x) &= \left\{1 + 4\sqrt{\pi}x^{3/2}\frac{f'(x)}{f(x)}b^{1/2}(X_t - x)K_{IG(x,1/b)}(X_t)\right\}^{-1}, \\
b_{RIG,t}(x) &= \left\{1 + 4\sqrt{\pi}x^{1/2}\frac{f'(x)}{f(x)}b^{1/2}(X_t - x)K_{RIG(1/(x-b),1/b)}(X_t)\right\}^{-1}.
\end{aligned}$$

For such $b_t(x)$ depending on a particular asymmetric kernel $K_{x,b}(u)$, let

$$J_1 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{1/4}b_t(x)\epsilon_t K_{x,b}(X_t).$$

Lemma A4. *If the conditions (A1)-(A5) hold, then for interior x,*

$$Var(J_{G,1}) \to \frac{\sigma^2(x)f(x)}{2\sqrt{\pi}x^{1/2}},\ Var(J_{IG,1}) \to \frac{\sigma^2(x)f(x)}{2\sqrt{\pi}x^{3/2}},\ Var(J_{RIG,1}) \to \frac{\sigma^2(x)f(x)}{2\sqrt{\pi}x^{1/2}}.$$

Proof of Lemma A4. It follows from (22) that $b_{G,t}(x) = 1 + o_p(1)$. Then, applying the same arguments which are used to establish (7), we obtain the stated results. ■

Lemma A5. *Let*

$$W_j(x) = \frac{1}{T}\sum_{t=1}^{T}(X_t - x)^j K_{x,b}^j(X_t),\ j \in \mathbb{N}.$$

If the conditions (A1)-(A5) hold, then for interior x,

$$\begin{aligned}
W_{G,1}(x) &= \{f(x) + xf'(x)\}b + o_p(b), W_{G,2}(x) = \frac{x^{1/2}f(x)}{4\sqrt{\pi}}b^{1/2} + o_p\left(b^{1/2}\right), W_{G,3}(x) = O_p(b), \\
W_{IG,1}(x) &= x^{3/2}f'(x)b + o_p(b), W_{IG,2}(x) = \frac{x^{3/2}f(x)}{4\sqrt{\pi}}b^{1/2} + o_p\left(b^{1/2}\right), W_{IG,3}(x) = O_p(b), \\
W_{RIG,1}(x) &= xf'(x)b + o_p(b), W_{RIG,2}(x) = \frac{x^{1/2}f(x)}{4\sqrt{\pi}}b^{1/2} + o_p\left(b^{1/2}\right), W_{RIG,3}(x) = O_p(b).
\end{aligned}$$

Proof of Lemma A5. Using Lemma B1,

$$E\{W_{G,j}(x)\} = E\left\{(X_1 - x)^j K_{G(x/b+1,b)}^j(X_1)\right\} = A_{b,j}(x)E\left\{(\theta_{j,x} - x)^j f(\theta_{j,x})\right\},$$

where $\theta_{j,x} \overset{d}{=} G(jx/b + 1, b/j)$. Taking a second-order Taylor expansion and using Lemma B2, we have, for interior x,

$$\begin{aligned}
E\{W_{G,1}(x)\} &= E\{S_{G,1}(x)\} = \{f(x) + xf'(x)\}b + o(b), \\
E\{W_{G,2}(x)\} &= A_{b,2}(x)\left\{\left(\frac{xb + b^2}{2}\right)f(x) + O\left(b^2\right)\right\} = \frac{x^{1/2}f(x)}{4\sqrt{\pi}}b^{1/2} + o\left(b^{1/2}\right), \\
E\{W_{G,3}(x)\} &= A_{b,3}(x)O\left(b^2\right) = O(b).
\end{aligned}$$

Finally, the ergodic theorem establishes the results. ■

Lemma A6. *If the conditions (A1)-(A5) hold, then for interior x,*

$$\begin{aligned}
\lambda_G &= \lambda_G(x) = 4\sqrt{\pi}\left\{\frac{1}{x^{1/2}} + \frac{x^{1/2}f'(x)}{f(x)}\right\}b^{1/2}\{1 + o_p(1)\}, \\
\lambda_{IG} &= \lambda_{IG}(x) = 4\sqrt{\pi}x^{3/2}\frac{f'(x)}{f(x)}b^{1/2}\{1 + o_p(1)\}, \\
\lambda_{RIG} &= \lambda_{RIG}(x) = 4\sqrt{\pi}x^{1/2}\frac{f'(x)}{f(x)}b^{1/2}\{1 + o_p(1)\},
\end{aligned}$$

so that $p_t(x) = T^{-1}b_t(x)\{1 + o_p(1)\}$ *for* $b_t(x)$ *depending on a particular asymmetric kernel* $K_{x,b}(u)$.

Proof of Lemma A6. It follows from (22) that we can pick some constant $M_G > 0$ such that

$$\sup_{0 \le j \le T-1}\left|(X_{j+1} - x)K_{G(x/b+1,b)}(X_{j+1})\right| \le M_G < \infty.$$

Then, applying expression (6.4) in Chen and Hall (1993) and using Lemma A2, we have

$$|\lambda_G| \le \frac{|W_{G,1}(x)|}{|W_{G,2}(x)| - M_G |W_{G,1}(x)|} = O_p\left(b^{1/2}\right).$$

Furthermore, a second-order Taylor expansion of the right-hand side of

$$0 = \frac{1}{T}\sum_{t=1}^{T} \frac{(X_t - x) K_{G(x/b+1,b)}(X_t)}{1 + \lambda_G (X_t - x) K_{G(x/b+1,b)}(X_t)}$$

around $\lambda_G = 0$ gives

$$0 = W_{G,1}(x) - \lambda_G W_{G,2}(x) + \bar{\lambda}_G^2 W_{G,3}(x)$$

for some $\bar{\lambda}_G$ joining λ_G and 0. Since $\bar{\lambda}_G$ is a convex combination of λ_G and 0, we have $\bar{\lambda}_G = O_p\left(b^{1/2}\right)$ so that $\bar{\lambda}_G^2 W_{G,3}(x) = O_p\left(b^2\right)$ by Lemma A5. Therefore, substituting the results in Lemma A5 yields

$$\lambda_G = \frac{W_{G,1}(x)}{W_{G,2}(x)} + \bar{\lambda}_G^2 \frac{W_{G,3}(x)}{W_{G,2}(x)} = 4\sqrt{\pi}\left\{\frac{1}{x^{1/2}} + \frac{x^{1/2} f'(x)}{f(x)}\right\} b^{1/2}\{1 + o_p(1)\} + O_p\left(b^{3/2}\right),$$

and $p_{G,t}(x) = T^{-1} b_{G,t}(x)\{1 + o_p(1)\}$ by (3). ■

A.2.1 Proof of Theorem 3

It follows from Lemma A6 that

$$\begin{aligned}
\hat{m}_G^{rnw}(x) - m(x) &= \frac{\sum_{t=1}^{T}\{\phi(Y_t) - m(x)\} p_{G,t}(x) K_{G(x/b+1,b)}(X_t)}{\sum_{t=1}^{T} p_{G,t}(x) K_{G(x/b+1,b)}(X_t)} \\
&\equiv \left\{\left(b^{1/2} T\right)^{-1/2} J_{G,1} + J_{G,2}\right\} J_{G,3}^{-1}\{1 + o_p(1)\},
\end{aligned}$$

where

$$\begin{aligned}
J_{G,2} &= \frac{1}{T}\sum_{t=1}^{T}\{m(X_t) - m(x)\} b_{G,t}(x) K_{G(x/b+1,b)}(X_t), \\
J_{G,3} &= \frac{1}{T}\sum_{t=1}^{T} b_{G,t}(x) K_{G(x/b+1,b)}(X_t).
\end{aligned}$$

To approximate $J_{G,2}$, note that

$$\frac{1}{T}\sum_{t=1}^{T}(X_t - x)^2 K_{G(x/b+1,b)}(X_t) = S_{G,2}(x) = x f(x) b + o_p(b).$$

Then, taking a second-order Taylor expansion and using (2) and $b_{G,t}(x) = 1 + o_p(1)$, we have

$$\begin{aligned}
J_{G,2} &= \frac{1}{T}\sum_{t=1}^{T} \frac{m''(x)}{2}(X_t - x)^2 b_{G,t}(x) K_{G(x/b+1,b)}(X_t) + O_p\{S_{G,3}(x)\} \\
&= \frac{1}{2} x m''(x) f(x) b + o_p(b).
\end{aligned}$$

Similarly, $J_{G,3} = f(x) + o_p(1)$. Therefore,

$$\sqrt{b^{1/2}T}\left\{\hat{m}_G^{rnw}(x) - m(x) - \frac{1}{2}xm''(x)\,b\right\} = f^{-1}(x)\,J_{G,1} + o_p(1).$$

Finally, demonstrating the asymptotic normality of the right-hand side closely follows the arguments used in the proof of Lemma A3, and thus the details are omitted. ■

A.3 Proofs of Theorems 4 and 5

The proofs of Theorems 4 and 5 require the following three lemmata.

Lemma A7. *Let*

$$\begin{aligned}
\mathbf{S}_{G,T}^{\dagger} &= \begin{bmatrix} S_{G,0}(x) & b^{-1}S_{G,1}(x) \\ b^{-1}S_{G,1}(x) & b^{-2}S_{G,2}(x) \end{bmatrix}, \\
\mathbf{S}_{IG,T}^{\dagger} &= \begin{bmatrix} S_{IG,0}(x) & b^{-2}S_{IG,1}(x) \\ b^{-2}S_{IG,1}(x) & b^{-4}S_{IG,2}(x) \end{bmatrix}, \\
\mathbf{S}_{RIG,T}^{\dagger} &= \begin{bmatrix} S_{RIG,0}(x) & b^{-1}S_{RIG,1}(x) \\ b^{-1}S_{RIG,1}(x) & b^{-2}S_{RIG,2}(x) \end{bmatrix}.
\end{aligned}$$

If the conditions (A1)-(A5) hold, then for boundary x,

$$\mathbf{S}_{G,T}^{\dagger} \xrightarrow{p} \mathbf{S}_{G}^{\dagger}(\kappa) = \begin{bmatrix} 1 & 1 \\ 1 & \kappa+2 \end{bmatrix} f(x), \tag{33}$$

$$\mathbf{S}_{IG,T}^{\dagger} \xrightarrow{p} \mathbf{S}_{IG}^{\dagger}(\kappa) = \begin{bmatrix} 1 & 0 \\ 0 & \kappa^3 \end{bmatrix} f(x), \tag{34}$$

$$\mathbf{S}_{RIG,T}^{\dagger} \xrightarrow{p} \mathbf{S}_{RIG}^{\dagger}(\kappa) = \begin{bmatrix} 1 & 0 \\ 0 & \kappa+1 \end{bmatrix} f(x). \tag{35}$$

Proof of Lemma A7. Following the argument used in the proof of Lemma A1 and applying Lemma B2, we have

$$\begin{aligned}
E\{S_{G,0}(x)\} &= f(x) + O(b), \\
E\{S_{G,1}(x)\} &= f(x)\,b + O\left(b^2\right), \\
E\{S_{G,2}(x)\} &= (\kappa+2)\,f(x)\,b^2 + o\left(b^2\right),
\end{aligned}$$

since $x = \kappa b + o(b)$. Then, invoking the ergodic theorem establishes the results. ■

Lemma A8. *Let*

$$\mathbf{t}_T^{\dagger}=\begin{cases}\begin{bmatrix}T_0^*(x)\\ b^{-1}T_1^*(x)\end{bmatrix} & \text{for the Gamma and RIG kernels}\\ \\ \begin{bmatrix}T_0^*(x)\\ b^{-2}T_1^*(x)\end{bmatrix} & \text{for the IG kernel,}\end{cases}$$

where $T_0^*(x)$ and $T_1^*(x)$ are defined in Lemma A2. Also, for an arbitrary vector $c\in\mathbb{R}^2$, define $Q_T^{\dagger}=\mathbf{c}^{\intercal}\mathbf{t}_T^{\dagger}$. If the conditions (A1)-(A3), (A4'), (A5) hold, then for boundary x,

$$\begin{aligned}
Var\left(\sqrt{bT}Q_{G,T}^{\dagger}\right) &\to \mathbf{c}^{\intercal}\mathbf{V}_G^{\dagger}(\kappa)\,\mathbf{c}\\
&= \mathbf{c}^{\intercal}\left\{\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{\kappa+1}{2}\end{bmatrix}\frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\sigma^2(x)\,f(x)\right\}\mathbf{c},\\
Var\left(\sqrt{b^2T}Q_{IG,T}^{\dagger}\right) &\to \mathbf{c}^{\intercal}\mathbf{V}_{IG}^{\dagger}(\kappa)\,\mathbf{c}\\
&= \mathbf{c}^{\intercal}\left\{\begin{bmatrix}\frac{1}{2\sqrt{\pi}\kappa^{3/2}} & 0\\ 0 & \frac{\kappa^{3/2}}{4\sqrt{\pi}}\end{bmatrix}\sigma^2(x)\,f(x)\right\}\mathbf{c},\\
Var\left(\sqrt{bT}Q_{RIG,T}^{\dagger}\right) &\to \mathbf{c}^{\intercal}\mathbf{V}_{RIG}^{\dagger}(\kappa)\,\mathbf{c}\\
&= \mathbf{c}^{\intercal}\left\{\boldsymbol{\Sigma}^{\dagger}(\kappa)\,\sigma^2(x)\,f(x)\right\}\mathbf{c},
\end{aligned}$$

where $\boldsymbol{\Sigma}^{\dagger}(\kappa)=\frac{1}{2\sqrt{\pi}}\begin{bmatrix}\kappa^{-1/2}+\frac{7}{16}\kappa^{-3/2}+\frac{3}{32}\kappa^{-5/2} & -\frac{3}{4}\kappa^{-1/2}-\frac{1}{8}\kappa^{-3/2}+\frac{3}{64}\kappa^{-5/2}\\ -\frac{3}{4}\kappa^{-1/2}-\frac{1}{8}\kappa^{-3/2}+\frac{3}{64}\kappa^{-5/2} & \frac{1}{2}\kappa^{1/2}+\frac{17}{32}\kappa^{-1/2}+\frac{23}{64}\kappa^{-3/2}+\frac{33}{128}\kappa^{-5/2}\end{bmatrix}$.

Proof of Lemma A8. It suffices to demonstrate that

$$Var\left\{\sqrt{bT}T_{G,0}^*(x)\right\} = \frac{\Gamma(2\kappa+1)\,\sigma^2(x)\,f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}+o(1), \tag{36}$$

$$Var\left\{\sqrt{bT}b^{-1}T_{G,1}^*(x)\right\} = \left(\frac{\kappa+1}{2}\right)\frac{\Gamma(2\kappa+1)\,\sigma^2(x)\,f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}+o(1), \tag{37}$$

$$Cov\left\{\sqrt{bT}T_{G,0}^*(x),\sqrt{bT}b^{-1}T_{G,1}^*(x)\right\} = \left(\frac{1}{2}\right)\frac{\Gamma(2\kappa+1)\,\sigma^2(x)\,f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}+o(1). \tag{38}$$

(i) Proof of (36). We have

$$\begin{aligned}
Var\left\{\sqrt{bT}T_{G,0}^*(x)\right\} &= Var\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}b^{1/2}\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}\\
&\equiv \gamma_{G,0}^{\dagger}(0)+2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_{G,0}^{\dagger}(j),
\end{aligned}\tag{39}$$

where $\gamma_{G,0}^{\dagger}(j)=bE\left\{\epsilon_1\epsilon_{1+j}K_{G(x/b+1,b)}(X_1)\,K_{G(x/b+1,b)}(X_{1+j})\right\}$ is the j^{th}-order autocovariance of the stationary process $\left\{b^{1/2}\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$. Lemma B1 implies that the first term on the right-hand side of (39) reduces to

$$\gamma_{G,0}^{\dagger}(0)=bA_{b,2}(x)\,E\left\{\sigma^2(\theta_{2,x})\,f(\theta_{2,x})\right\}=b\left\{\frac{b^{-1}\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}+o\left(b^{-1}\right)\right\}E\left\{\sigma^2(\theta_{2,x})\,f(\theta_{2,x})\right\},$$

where $\theta_{2,x} \overset{d}{=} G(2x/b+1, b/2)$. Since $E\left\{\sigma^2(\theta_{2,x}) f(\theta_{2,x})\right\} = \sigma^2(x) f(x) + O(b)$ for boundary x, we have

$$\gamma_{G,0}^{\dagger}(0) = \frac{\Gamma(2\kappa+1)\sigma^2(x) f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o(1).$$

On the other hand, for a constant a satisfying (A4'), pick a sequence $d_{0T}^{\dagger} = \left\lfloor b^{-(1-2/\delta)/a} \right\rfloor$. Then, the second term on the right-hand side of (39) is bounded by

$$\left|\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_{G,0}^{\dagger}(j)\right| \leq \sum_{j=1}^{d_{0T}^{\dagger}-1}\left|\gamma_{G,0}^{\dagger}(j)\right| + \sum_{j=d_{0T}^{\dagger}}^{T-1}\left|\gamma_{G,0}^{\dagger}(j)\right| \equiv U_1^{\dagger} + U_2^{\dagger}.$$

For $U_1^{\dagger}$, $\left|\gamma_{G,0}^{\dagger}(j)\right| \leq b(U_{11}+U_{12}+U_{13}+U_{14})$, where U_{11}, U_{12}, U_{13}, and U_{14} are defined in the proof of Lemma A2. It is demonstrated in Lemma A2 that each of U_{11}, U_{12}, U_{13}, and U_{14} is bounded by $O(1)$, and thus $\left|\gamma_{G,0}^{\dagger}(j)\right| \leq O(b)$, which establishes that

$$U_1^{\dagger} \leq O\left(d_{0T}^{\dagger} b\right) = O\left(b^{\{a-(1-2/\delta)\}/a}\right) \to 0.$$

For U_2, by Davydov's lemma,

$$\left|\gamma_{G,0}^{\dagger}(j)\right| \leq 8\{\alpha(j)\}^{1-2/\delta}\left\{E\left|b^{1/2}\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta}\right\}^{2/\delta},$$

where $E\left|b^{1/2}\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} \leq cb^{\delta/2}(U_{21}+U_{22})$, and U_{21} and U_{22} are defined in the proof of Lemma A2. Applying the same arguments as in the proof of Lemma A2, it can be shown that each of U_{21} and U_{22} is bounded by $O\{A_{b,\delta}(x)\} = O\left(b^{1-\delta}\right)$. Therefore, $E\left|b^{1/4}\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^{\delta} \leq O\left(b^{1-\delta/2}\right)$, and thus

$$U_2^{\dagger} \leq O\left(b^{2/\delta-1}\right)\sum_{j=d_{0T}^{\dagger}}^{T-1}\{\alpha(j)\}^{1-2/\delta} \leq O\left(b^{-(1-2/\delta)}\right) d_{0T}^{\dagger -a}\sum_{j=d_{0T}^{\dagger}}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} \to 0,$$

because $O\left(b^{-(1-2/\delta)}\right) d_{0T}^{\dagger -a} = O(1)$, $d_{0T}^{\dagger} \to \infty$, and $\sum_{j=1}^{\infty} j^a\{\alpha(j)\}^{1-2/\delta} < \infty$. This completes the proof of this part.

Remark. As before, (36) can be demonstrated even after replacing (A4') by the weaker condition (A4). Observe that given (A4) and $d_{0T}^{\dagger} = \left\lfloor b^{-(1-2/\delta)/a} \right\rfloor$, it can be shown that each of $U_1^{\dagger}$ and $U_2^{\dagger}$ is still $o(1)$.

(ii) Proof of (37). We have

$$
\begin{aligned}
Var\left\{\sqrt{bT}b^{-1}T^{*}_{G,1}(x)\right\} &= Var\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{-1/2}(X_t - x)\,\epsilon_t K_{G(x/b+1,b)}(X_t)\right\} \\
&\equiv \gamma^{\dagger}_{G,1}(0) + 2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma^{\dagger}_{G,1}(j), \qquad (40)
\end{aligned}
$$

where $\gamma^{\dagger}_{G,1}(j) = b^{-1}E\{(X_1 - x)(X_{1+j} - x)\,\epsilon_1\epsilon_{1+j}K_{G(x/b+1,b)}(X_1)\,K_{G(x/b+1,b)}(X_{1+j})\}$ is the j^{th}-order autocovariance of the stationary process $\left\{b^{-1/2}(X_t - x)\,\epsilon_t K_{G(x/b+1,b)}(X_t)\right\}$. By Lemma B1, the first term on the right-hand side of (40) reduces to

$$
\begin{aligned}
\gamma^{\dagger}_{G,1}(0) &= b^{-1}A_{b,2}(x)\,E\left\{(\theta_{2,x} - x)^2\,\sigma^2(\theta_{2,x})\,f(\theta_{2,x})\right\} \\
&= b^{-1}\left\{\frac{b^{-1}\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o\left(b^{-1}\right)\right\} E\left\{(\theta_{2,x} - x)^2\,\sigma^2(\theta_{2,x})\,f(\theta_{2,x})\right\},
\end{aligned}
$$

where $\theta_{2,x} \overset{d}{=} G(2x/b+1, b/2)$. Using a Taylor expansion and Lemma B2 and noting that $x = \kappa b + o(b)$ for boundary x, we have

$$
E\left\{(\theta_{2,x} - x)^2\,\sigma^2(\theta_{2,x})\,f(\theta_{2,x})\right\} = \left(\frac{\kappa+1}{2}\right)\sigma^2(x)\,f(x)\,b^2 + o\left(b^2\right)
$$

so that

$$
\gamma^{\dagger}_{G,1}(0) = \left(\frac{\kappa+1}{2}\right)\frac{\Gamma(2\kappa+1)\,\sigma^2(x)\,f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)} + o(1).
$$

On the other hand, the second term on the right-hand side of (40) is bounded by

$$
\left|\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma^{\dagger}_{G,1}(j)\right| \le \sum_{j=1}^{d^{\dagger}_{1T}-1}\left|\gamma^{\dagger}_{G,1}(j)\right| + \sum_{j=d^{\dagger}_{1T}}^{T-1}\left|\gamma^{\dagger}_{G,1}(j)\right| \equiv V^{\dagger}_1 + V^{\dagger}_2,
$$

where the sequence $d^{\dagger}_{1T}$ is defined as $d^{\dagger}_{1T} = \left\lfloor b^{-3(1-2/\delta)/a}\right\rfloor$ for a constant a satisfying (A4').

To find the bound for $V_1^\dagger$, we have $\left|\gamma_{G,1}^\dagger(j)\right| \le V_{11}^\dagger + V_{12}^\dagger + V_{13}^\dagger + V_{14}^\dagger$, where

$$
\begin{aligned}
V_{11}^\dagger &= b^{-1}E\left[E\left\{|\phi(Y_1)\phi(Y_{1+j})|\,|X_1, X_{1+j}\right\}|X_1 - x|\,K_{G(x/b+1,b)}(X_1)\right. \\
&\quad \left.\cdot |X_{1+j} - x|\,K_{G(x/b+1,b)}(X_{1+j})\right], \\
V_{12}^\dagger &= b^{-1}E\left[E\left\{|\phi(Y_1)|\,|X_1, X_{1+j}\right\}|X_1 - x|\,K_{G(x/b+1,b)}(X_1)\right. \\
&\quad \left.\cdot E\left\{|\phi(Y_{1+j})|\,|X_{1+j}\right\}|X_{1+j} - x|\,K_{G(x/b+1,b)}(X_{1+j})\right], \\
V_{13}^\dagger &= b^{-1}E\left[E\left\{|\phi(Y_1)|\,|X_1\right\}|X_1 - x|\,K_{G(x/b+1,b)}(X_1)\right. \\
&\quad \left.\cdot E\left\{|\phi(Y_{1+j})|\,|X_1, X_{1+j}\right\}|X_{1+j} - x|\,K_{G(x/b+1,b)}(X_{1+j})\right], \\
V_{14}^\dagger &= b^{-1}E\left[E\left\{|\phi(Y_1)|\,|X_1\right\}|X_1 - x|\,K_{G(x/b+1,b)}(X_1)\right. \\
&\quad \left.\cdot E\left\{|\phi(Y_{1+j})|\,|X_{1+j}\right\}|X_{1+j} - x|\,K_{G(x/b+1,b)}(X_{1+j})\right].
\end{aligned}
$$

Note that $V_{11}^\dagger \le cb^{-1}(V_{111}V_{112} + V_{113}V_{114})$, where V_{111}, V_{112}, V_{113}, and V_{114} are defined in the proof of Lemma A2. Moreover, $V_{111} \le (\beta_0 V_{1111} + \beta_1 V_{1112})^{1/2} V_{1113}^{1/2}$ by the Cauchy-Schwarz inequality, where V_{1111}, V_{1112} and V_{1113} are again defined in the proof of Lemma A2. By a Taylor expansion and Lemma B2, together with $x = \kappa b + o(b)$, we have $V_{1111} = O(b^2)$ and $V_{1112} = O(b^{m+2})$. In addition, $V_{1113} = O(1)$, and thus $V_{111} \le O(b)$. Similarly, each of V_{112}, V_{113} and V_{114} is at most $O(b)$. Hence, $V_{11}^\dagger \le O(b)$. Applying the same argument, it can also be demonstrated that each of $V_{12}^\dagger$, $V_{13}^\dagger$ and $V_{14}^\dagger$ is bounded by $O(b)$. Hence, we can conclude that $\left|\gamma_{G,1}^\dagger(j)\right| \le O(b)$, which establishes that

$$V_1^\dagger \le O\left(d_{1T}^\dagger b\right) = O\left(b^{\{a-3(1-2/\delta)\}/a}\right) \to 0.$$

For V_2, Davydov's lemma implies that

$$\left|\gamma_{G,1}^\dagger(j)\right| \le 8\{\alpha(j)\}^{1-2/\delta}\left\{E\left|b^{-1/2}(X_1 - x)\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^\delta\right\}^{2/\delta},$$

where $E\left|b^{-1/2}(X_1 - x)\epsilon_1 K_{G(x/b+1,b)}(X_1)\right|^\delta \le cb^{-\delta/2}A_{b,\delta}(x)V_{21}$, and V_{21} is defined in the proof of Lemma A2. Furthermore, $V_{21} \le (\alpha_0 V_{211} + \alpha_1 V_{212})^{1/2} V_{213}^{1/2}$, where

$$
\begin{aligned}
V_{211} &= \int_0^\infty |u - x|^{2\delta} K_{G(\delta x/b+1,b/\delta)}(u) f(u)\,du, \\
V_{212} &= \int_0^\infty u^l |u - x|^{2\delta} K_{G(\delta x/b+1,b/\delta)}(u) f(u)\,du, \\
V_{213} &= \int_0^\infty \left(\alpha_0 + \alpha_1 u^l\right) K_{G(\delta x/b+1,b/\delta)}(u) f(u)\,du,
\end{aligned}
$$

where $2\delta > 4$. Then, by Lemma B2, V_{211} and V_{212} are bounded by $O\left(b^{4}\right)$ and $O\left(b^{l+4}\right)$, respectively. Clearly, $V_{213} = O\left(1\right)$, and thus we have $V_{21} \leq O\left(b^{2}\right)$. Hence, $E\left|b^{-1/2}\left(X_{1}-x\right)\epsilon_{1}K_{G(x/b+1,b)}\left(X_{1}\right)\right|^{\delta} \leq O\left(b^{3-3\delta/2}\right)$, because $A_{b,\delta}\left(x\right) = O\left(b^{1-\delta}\right)$. Therefore,

$$V_{2}^{\dagger} \leq O\left(b^{6/\delta-3}\right)\sum_{j=d_{1T}^{\dagger}}^{T-1}\left\{\alpha\left(j\right)\right\}^{1-2/\delta} \leq O\left(b^{-3(1-2/\delta)}\right)d_{1T}^{\dagger-a}\sum_{j=d_{1T}^{\dagger}}^{\infty}j^{a}\left\{\alpha\left(j\right)\right\}^{1-2/\delta} \rightarrow 0,$$

because $O\left(b^{-3(1-2/\delta)}\right)d_{1T}^{\dagger-a} = O\left(1\right)$, $d_{1T}^{\dagger} \rightarrow \infty$, and $\sum_{j=1}^{\infty}j^{a}\left\{\alpha\left(j\right)\right\}^{1-2/\delta} < \infty$. This completes the proof of this part.

(iii) Proof of (38). We have

$$\begin{aligned}
& Cov\left\{\sqrt{bT}T_{G,0}^{*}\left(x\right),\sqrt{bT}b^{-1}T_{G,1}^{*}\left(x\right)\right\} \\
= \ & Cov\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}b^{1/2}\epsilon_{t}K_{G(x/b+1,b)}\left(X_{t}\right),\frac{1}{\sqrt{T}}\sum_{t=1}^{T}b^{-1/2}\left(X_{t}-x\right)\epsilon_{t}K_{G(x/b+1,b)}\left(X_{t}\right)\right\} \\
\equiv \ & \gamma_{G,3}^{\dagger}\left(0\right)+2\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\gamma_{G,3}^{\dagger}\left(j\right), \qquad (41)
\end{aligned}$$

where $\gamma_{G,3}^{\dagger}\left(j\right) = E\left\{\left(X_{1+j}-x\right)\epsilon_{1}\epsilon_{1+j}K_{G(x/b+1,b)}\left(X_{1}\right)K_{G(x/b+1,b)}\left(X_{1+j}\right)\right\}$ is the j^{th}-order cross-covariance of the stationary processes $\left\{b^{1/2}\epsilon_{t}K_{G(x/b+1,b)}\left(X_{t}\right)\right\}$ and $\left\{b^{-1/2}\left(X_{t}-x\right)\epsilon_{t}K_{G(x/b+1,b)}\left(X_{t}\right)\right\}$. By Lemma B1, the first term on the right-hand side of (41) reduces to

$$\begin{aligned}
\gamma_{G,3}^{\dagger}\left(0\right) &= E\left\{\left(X_{1}-x\right)\sigma^{2}\left(X_{1}\right)K_{G(x/b+1,b)}^{2}\left(X_{1}\right)\right\} \\
&= A_{b,2}\left(x\right)E\left\{\left(\theta_{2,x}-x\right)\sigma^{2}\left(\theta_{2,x}\right)f\left(\theta_{2,x}\right)\right\} \\
&= \left\{\frac{b^{-1}\Gamma\left(2\kappa+1\right)}{2^{2\kappa+1}\Gamma^{2}\left(\kappa+1\right)}+o\left(b^{-1}\right)\right\}E\left\{\left(\theta_{2,x}-x\right)\sigma^{2}\left(\theta_{2,x}\right)f\left(\theta_{2,x}\right)\right\},
\end{aligned}$$

where $\theta_{2,x} \overset{d}{=} G\left(2x/b+1,b/2\right)$. Using a Taylor expansion, Lemma B2 and $x = \kappa b + o\left(b\right)$ for boundary x, we have

$$E\left\{\left(\theta_{2,x}-x\right)\sigma^{2}\left(\theta_{2,x}\right)f\left(\theta_{2,x}\right)\right\} = \frac{b}{2}\sigma^{2}\left(x\right)f\left(x\right)+O\left(b^{2}\right)$$

so that

$$\gamma_{G,3}^{\dagger}\left(0\right) = \left(\frac{1}{2}\right)\frac{\Gamma\left(2\kappa+1\right)\sigma^{2}\left(x\right)f\left(x\right)}{2^{2\kappa+1}\Gamma^{2}\left(\kappa+1\right)}+o\left(1\right).$$

On the other hand, applying the same procedures as in parts (i) and (ii), we can also establish that the second term on the right-hand side of (41) is $o\left(1\right)$. This completes the proof. ■

Lemma A9. *If the conditions (A1)-(A3), (A4'), (A5), (A6') hold, then for boundary x,*

$$\sqrt{bT}Q_{G,T}^{\dagger} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{G}^{\dagger}(\kappa)\,\mathbf{c}\right), \sqrt{b^2T}Q_{IG,T}^{\dagger} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{IG}^{\dagger}(\kappa)\,\mathbf{c}\right), \sqrt{bT}Q_{RIG,T}^{\dagger} \xrightarrow{d} N\left(0, \mathbf{c}^{\intercal}\mathbf{V}_{RIG}^{\dagger}(\kappa)\,\mathbf{c}\right).$$

Proof of Lemma A9. The proof strategies used for Lemma A3 directly apply after the following minor modifications. First, for the block sizes in the small-block and large-block argument, we can use (A6') and pick a sequence $\gamma_T \in \mathbb{N}$ such that $\gamma_T \to \infty$, $\gamma_T s_T / (bT)^{1/2} \to 0$, and $\gamma_T (T/b)^{1/2} \alpha(s_T) \to 0$ as $T \to \infty$. Define the large-block size r_T by

$$r_T = \left\lfloor \frac{(bT)^{1/2}}{\gamma_T} \right\rfloor$$

and the small-block size by s_T. It follows that

$$\frac{s_T}{r_T} \to 0, \; \frac{r_T}{T} \to 0, \; \frac{r_T}{(bT)^{1/2}} \to 0, \; \frac{T}{r_T}\alpha(s_T) \to 0$$

as $T \to \infty$. Also put

$$q_T = \left\lfloor \frac{T}{r_T + s_T} \right\rfloor.$$

Second, replace $\varsigma_{G,j}^{*}$ by $\varsigma_{G,j}^{\dagger} = \mathbf{c}^{\intercal}\mathbf{Z}_{G,j}^{\dagger}$, where

$$\mathbf{Z}_{G,j}^{\dagger} = \begin{bmatrix} b^{1/2}\epsilon_{j+1}K_{G(x/b+1,b)}(X_{j+1}) \\ b^{-1/2}(X_{j+1} - x)\,\epsilon_{j+1}K_{G(x/b+1,b)}(X_{j+1}) \end{bmatrix}$$

for $j = 0, \ldots, T-1$, so that

$$\sqrt{bT}Q_{G,T}^{\dagger} = \frac{1}{\sqrt{T}}\sum_{j=0}^{T-1}\varsigma_{G,j}^{\dagger}.$$

Then, the five statements analogous to (16)-(20) can be demonstrated in exactly the same manner.

■

A.3.1 Proof of Theorem 4

Lemma A9 implies that

$$\sqrt{bT}\left\{\frac{1}{T}\sum_{t=1}^{T}K_{G(x/b+1,b)}(X_t)\,\epsilon_t\right\} \xrightarrow{d} N\left(0, \frac{\Gamma(2\kappa+1)\,\sigma^2(x)\,f(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\right).$$

Using the same argument as in the proof of Theorem 1 and applying Slutsky's lemma and Lemma A7, we have

$$\sqrt{bT}\left\{\hat{m}_G^{nw}(x) - \tilde{m}_G^{nw}(x)\right\} \xrightarrow{d} N\left(0, \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\frac{\sigma^2(x)}{f(x)}\right), \tag{42}$$

where $\tilde{m}_G^{nw}(x)$ is defined in the proof of Theorem 1. Note that

$$S_{G,1}(x) = f(x)b + o_p(b), \; S_{G,2}(x) = O_p\left(b^2\right), \; S_{G,3}(x) = O_p\left(b^3\right), \tag{43}$$

by Lemma B2 and the ergodic theorem as $x/b \to \kappa$. Substituting (43) into (28) and using Lemma A7 and (A7'), we can see that the left-hand side of (42) can be approximated by

$$\begin{aligned}\sqrt{bT}\left\{\hat{m}_G^{nw}(x) - \tilde{m}_G^{nw}(x)\right\} &= \sqrt{bT}\left\{\hat{m}_G^{nw}(x) - m(x) - m'(x)b + o_p(b)\right\} \\ &= \sqrt{bT}\left\{\hat{m}_G^{nw}(x) - m(x) - m'(x)b\right\} + o_p(1).\end{aligned}$$

This completes the proof. ■

A.3.2 Proof of Theorem 5

Lemma A9 and the Cramér-Wald device imply that $\sqrt{bT}\mathbf{t}_{G,T}^{\dagger} \xrightarrow{d} N\left(\mathbf{0}_2, \mathbf{V}_G^{\dagger}(\kappa)\right)$. Following to the same argument as in the proof of Theorem 2, we also have

$$\begin{bmatrix} 1 & 0 \\ 0 & b \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} = \mathbf{S}_{G,T}^{\dagger-1}\mathbf{t}_{G,T}^{\dagger}.$$

Therefore, by Slutsky's lemma and Lemma A7,

$$\sqrt{bT}\begin{bmatrix} 1 & 0 \\ 0 & b \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} \xrightarrow{d} N\left(\mathbf{0}_2, \mathbf{S}_G^{\dagger-1}(\kappa)\mathbf{V}_G^{\dagger}(\kappa)\mathbf{S}_G^{\dagger-1}(\kappa)\right). \tag{44}$$

In addition, a second-order Taylor expansion yields

$$\begin{aligned}&\tilde{\boldsymbol{\beta}}_G(x) \\ = \;& \boldsymbol{\beta}(x) + \frac{m''(x)}{2}\begin{bmatrix} 1 & 0 \\ 0 & b^{-1} \end{bmatrix}\mathbf{S}_{G,T}^{\dagger-1}\begin{bmatrix} 1 & 0 \\ 0 & b^{-1} \end{bmatrix}\begin{bmatrix} S_{G,2}(x) \\ S_{G,3}(x) \end{bmatrix} + \begin{bmatrix} O_p\left\{S_{G,3}(x)\right\} \\ O_p\left\{b^{-2}S_{G,4}(x)\right\} \end{bmatrix},\end{aligned} \tag{45}$$

where

$$S_{G,2}(x) = (\kappa+2)f(x)b^2 + o_p\left(b^2\right), \; S_{G,3}(x) = (5\kappa+6)f(x)b^3 + o_p\left(b^3\right), \; S_{G,4}(x) = O_p\left(b^4\right) \tag{46}$$

by Lemma B2 and the ergodic theorem as $x/b \to \kappa$. Substituting (46) into (45) and using Lemma A7 and (A7"), we can see that the left-hand side of (44) can be approximated by

$$\begin{aligned}&\sqrt{bT}\begin{bmatrix} 1 & 0 \\ 0 & b \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \tilde{\boldsymbol{\beta}}_G(x)\right\} \\ = \;& \sqrt{bT}\begin{bmatrix} 1 & 0 \\ 0 & b \end{bmatrix}\left\{\hat{\boldsymbol{\beta}}_G(x) - \boldsymbol{\beta}(x) - \frac{m''(x)}{2}\begin{bmatrix} (\kappa-2)b^2 \\ 4b \end{bmatrix} + \begin{bmatrix} o_p\left(b^2\right) \\ o_p(b) \end{bmatrix}\right\},\end{aligned}$$

where $\sqrt{bT}o_p\left(b^2\right) = o_p\left(\sqrt{b^5 T}\right) = o_p\left(1\right)$, and $\sqrt{bT}bo_p\left(b\right) = o_p\left(\sqrt{b^5 T}\right) = o_p\left(1\right)$.

Therefore, (44) can be rewritten as

$$T_{b,2}\left\{\hat{\boldsymbol{\beta}}_G\left(x\right) - \boldsymbol{\beta}\left(x\right) - \frac{m''\left(x\right)}{2}\begin{bmatrix} \left(\kappa-2\right)b^2 \\ 4b \end{bmatrix}\right\} + \begin{bmatrix} o_p\left(1\right) \\ o_p\left(1\right) \end{bmatrix} \overset{d}{\to} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\kappa+5 & -2 \\ -2 & 1 \end{bmatrix}\frac{V_G^B}{2\left(\kappa+1\right)}\right)$$

by letting $T_{b,2} = \sqrt{bT}\begin{bmatrix} 1 & 0 \\ 0 & b \end{bmatrix}$ and $V_G^B = \frac{\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma^2(\kappa+1)}\frac{\sigma^2(x)}{f(x)}$. This completes the proof. ■

B Appendix B: Auxiliary Results

Appendix B additionally provides three lemmata that are useful to establish the theorems in this paper. Lemma B1 refers to the properties of powered asymmetric kernels with an arbitrarily chosen exponent $\nu \geq 1$. For convenience, Lemma B2 presents analytical expressions and orders of magnitude of the moments of three asymmetric kernels around the design point x. Lemma B3 restates Lemma 1.1 in Volkonskii and Razanov (1959).

Lemma B1. *For $\nu \geq 1$,*

$$\begin{aligned} K^{\nu}_{G(x/b+1,b)}\left(u\right) &= A_{b,\nu}\left(x\right)K_{G(\nu x/b+1,b/\nu)}\left(u\right), \\ K^{\nu}_{IG(x,1/b)}\left(u\right) &= B_{b,\nu}u^{3(1-\nu)/2}K_{IG(x,\nu/b)}\left(u\right), \\ K^{\nu}_{RIG(1/(x-b),1/b)}\left(u\right) &= B_{b,\nu}u^{(1-\nu)/2}K_{RIG(1/(x-b),\nu/b)}\left(u\right), \end{aligned}$$

where

$$A_{b,\nu}\left(x\right) = \frac{b^{1-\nu}\Gamma\left(\nu x/b+1\right)}{\nu^{\nu x/b+1}\Gamma^{\nu}\left(x/b+1\right)} = \begin{cases} \frac{b^{(1-\nu)/2}x^{(1-\nu)/2}}{\nu^{1/2}\left(\sqrt{2\pi}\right)^{\nu-1}} + o\left(b^{(1-\nu)/2}\right) & \text{for interior } x \\ \frac{b^{1-\nu}\Gamma(\nu\kappa+1)}{\nu^{\nu\kappa+1}\Gamma^{\nu}(\kappa+1)} + o\left(b^{1-\nu}\right) & \text{for boundary } x, \end{cases}$$

and

$$B_{b,\nu} = \frac{b^{(1-\nu)/2}}{\nu^{1/2}\left(\sqrt{2\pi}\right)^{\nu-1}}.$$

Proof of Lemma B1. A straightforward calculation yields

$$\begin{aligned} K^{\nu}_{G(x/b+1,b)}\left(u\right) &= \frac{u^{\nu x/b}\exp\left(-\nu u/b\right)}{b^{\nu x/b+\nu}\Gamma^{\nu}\left(x/b+1\right)}\mathbf{1}\left\{u>0\right\} \\ &= \left\{\frac{b^{1-\nu}\Gamma\left(\nu x/b+1\right)}{\nu^{\nu x/b+1}\Gamma^{\nu}\left(x/b+1\right)}\right\}\left[\frac{u^{\nu x/b}\exp\left\{-u/\left(b/\nu\right)\right\}}{\left(b/\nu\right)^{\nu x/b+1}\Gamma\left(\nu x/b+1\right)}\mathbf{1}\left\{u>0\right\}\right] \\ &\equiv A_{b,\nu}\left(x\right)K_{G(\nu x/b+1,b/\nu)}\left(u\right). \end{aligned}$$

Let $R(z) = \sqrt{2\pi} z^{z+1/2} \exp(-z) / \Gamma(z+1)$ for $z \geq 0$. Following the argument in Section 3 of Chen (2000b), we have

$$A_{b,\nu}(x) = \frac{b^{(1-\nu)/2} x^{(1-\nu)/2} R^{\nu}(x/b)}{\nu^{1/2} \left(\sqrt{2\pi}\right)^{\nu-1} R(\nu x/b)} = \frac{b^{(1-\nu)/2} x^{(1-\nu)/2}}{\nu^{1/2} \left(\sqrt{2\pi}\right)^{\nu-1}} + o\left(b^{(1-\nu)/2}\right)$$

for interior x. On the other hand, as $x/b \to \kappa > 0$,

$$A_{b,\nu}(x) = \frac{b^{1-\nu} \Gamma(\nu\kappa + 1)}{\nu^{\nu\kappa+1} \Gamma^{\nu}(\kappa+1)} + o\left(b^{1-\nu}\right).$$

For the IG and RIG kernels, proofs are straightforward and thus omitted. ■

Lemma B2. *Let*

$$\theta_{\nu,x} \stackrel{d}{=} G(\nu x/b + 1, b/\nu), \eta_{\nu,x} \stackrel{d}{=} IG(x, \nu/b), \xi_{\nu,x} \stackrel{d}{=} RIG(1/(x-b), \nu/b), \nu \geq 1.$$

(a)-(i) *First four moments of $\theta_{\nu,x}$ around x are*

$$\begin{aligned}
E(\theta_{\nu,x} - x) &= \frac{b}{\nu}, \\
E(\theta_{\nu,x} - x)^2 &= \frac{b(\nu x + 2b)}{\nu^2}, \\
E(\theta_{\nu,x} - x)^3 &= \frac{b^2(5\nu x + 6b)}{\nu^3}, \\
E(\theta_{\nu,x} - x)^4 &= \frac{b^2\left(3\nu^2 x^2 + 26\nu x b + 24b^2\right)}{\nu^4}.
\end{aligned}$$

(ii) *For any integer $r \geq 0$,*

$$E(\theta_{\nu,x} - x)^r = \begin{cases} O\left(b^{r/2}\right) & \text{for interior } x \text{ and even } r \\ O\left(b^{(r+1)/2}\right) & \text{for interior } x \text{ and odd } r \\ O(b^r) & \text{for boundary } x. \end{cases} \tag{47}$$

(b)-(i) *First four moments of $\eta_{\nu,x}$ around x are*

$$\begin{aligned}
E\left(\eta_{\nu,x} - x\right) &= 0, \\
E\left(\eta_{\nu,x} - x\right)^2 &= \frac{x^3 b}{\nu}, \\
E\left(\eta_{\nu,x} - x\right)^3 &= \frac{3x^5 b^2}{\nu^2}, \\
E\left(\eta_{\nu,x} - x\right)^4 &= \frac{3x^6 b^2(\nu + 5xb)}{\nu^3}.
\end{aligned}$$

(ii) *For any integer $r \geq 3$,*

$$E\left(\eta_{\nu,x}-x\right)^{r}=\begin{cases} O\left(b^{2}\right) & \text{for interior } x \\ O\left(b^{r+4}\right) & \text{for boundary } x. \end{cases} \tag{48}$$

(c)-(i) *First four moments of $\eta_{\nu,x}$ around x are*

$$\begin{aligned} E\left(\xi_{\nu,x}-x\right) &= \frac{b\left(1-\nu\right)}{\nu}, \\ E\left(\xi_{\nu,x}-x\right)^{2} &= \frac{b\left\{\nu x+\left(\nu^{2}-3\nu+3\right)b\right\}}{\nu^{2}}, \\ E\left(\xi_{\nu,x}-x\right)^{3} &= \frac{b^{2}\left\{\left(-3\nu+6\right)\nu x-\left(\nu^{3}-6\nu^{2}+15\nu-15\right)b\right\}}{\nu^{3}}, \\ E\left(\xi_{\nu,x}-x\right)^{4} &= \frac{b^{2}\left\{3\nu^{2}x^{2}+\left(6\nu^{2}-30\nu+45\right)\nu xb+\left(\nu^{4}-10\nu^{3}+45\nu^{2}-105\nu+105\right)b^{2}\right\}}{\nu^{4}}. \end{aligned}$$

(ii) *For any integer $r \geq 3$,*

$$E\left(\xi_{\nu,x}-x\right)^{r}=\begin{cases} O\left(b^{2}\right) & \text{for interior } x \\ O\left(b^{r}\right) & \text{for boundary } x. \end{cases} \tag{49}$$

Proof of Lemma B2: Part (a)-(i). We show that the recursive formula

$$E\left(\theta_{\nu,x}-x\right)^{r}=\frac{b}{\nu}E\left(\theta_{\nu,x}-x\right)^{r-1}+\left(x+\frac{b}{\nu}\right)\sum_{j=1}^{r-1}\prod_{k=0}^{j-1}\left(r-1-k\right)\left(\frac{b}{\nu}\right)^{j}E\left(\theta_{\nu,x}-x\right)^{r-1-j} \tag{50}$$

holds for $r \in \mathbb{N}$. First four moments of $\theta_{\nu,x}$ around x directly follows this formula.

By the property of the Gamma distribution, we have

$$E\left(\theta_{\nu,x}^{j}\right)=\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right).$$

Then, applying a binomial expansion and Pascal's triangle yields

$$\begin{aligned} E\left(\theta_{\nu,x}-x\right)^{r} &= \sum_{j=0}^{r}\binom{r}{j}E\left(\theta_{\nu,x}^{j}\right)\left(-x\right)^{r-j} \\ &= \sum_{j=0}^{r}\left(-1\right)^{r-j}\binom{r}{j}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &= \sum_{j=0}^{r}\left(-1\right)^{r-j}\left\{\binom{r-1}{j-1}+\binom{r-1}{j}\right\}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &= \sum_{j=0}^{r}\left(-1\right)^{r-j}\binom{r-1}{j-1}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &\quad +\sum_{j=0}^{r}\left(-1\right)^{r-j}\binom{r-1}{j}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &\equiv A\left(r\right)+B\left(r\right). \end{aligned} \tag{51}$$

By $\binom{r-1}{r}=0$,

$$\begin{aligned} B(r) &= \sum_{j=0}^{r-1}(-1)^{r-j}\binom{r-1}{j}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &= (-x)\sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{(r-1)-j} \\ &= (-x)\,E\left(\theta_{\nu,x}-x\right)^{r-1}. \end{aligned} \tag{52}$$

On the other hand, by $\binom{r-1}{-1}=0$,

$$\begin{aligned} A(r) &= \sum_{j=1}^{r}(-1)^{r-j}\binom{r-1}{j-1}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{r-j} \\ &= \sum_{j=0}^{r-1}(-1)^{r-(j+1)}\binom{r-1}{j}\left(\frac{b}{\nu}\right)^{j+1}\prod_{k=0}^{j}\left(\frac{\nu x}{b}+1+k\right)x^{r-(j+1)} \\ &= \sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j}\left\{x+\frac{(1+j)\,b}{\nu}\right\}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{(r-1)-j} \\ &= \left(x+\frac{b}{\nu}\right)\sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{(r-1)-j} \\ &\quad+\left(\frac{b}{\nu}\right)\left\{\sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j}j\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{(r-1)-j}\right\}. \end{aligned}$$

The first term on the right-hand side reduces to $(x+b/\nu)\,E\left(\theta_{\nu,x}-x\right)^{r-1}$. In addition, since $\binom{r-1}{j}j=(r-1)\binom{r-2}{j-1}$, the second term on the right-hand side reduces to

$$\left(\frac{b}{\nu}\right)(r-1)\left\{\sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{(r-1)-1}{j-1}\left(\frac{b}{\nu}\right)^{j}\prod_{k=0}^{j-1}\left(\frac{\nu x}{b}+1+k\right)x^{(r-1)-j}\right\}=\left(\frac{b}{\nu}\right)(r-1)\,A(r-1).$$

Therefore,

$$A(r)=\left(x+\frac{b}{\nu}\right)E\left(\theta_{\nu,x}-x\right)^{r-1}+\left(\frac{b}{\nu}\right)(r-1)\,A(r-1).$$

Using this recursive formula, together with $A(1)=(x+b/\nu)\,E\left(\theta_{\nu,x}-x\right)^{0}\ (=x+b/\nu)$, we have

$$A(r)=\left(x+\frac{b}{\nu}\right)\left\{E\left(\theta_{\nu,x}-x\right)^{r-1}+\sum_{j=1}^{r-1}\prod_{k=0}^{j-1}(r-1-k)\left(\frac{b}{\nu}\right)^{j}E\left(\theta_{\nu,x}-x\right)^{r-1-j}\right\}. \tag{53}$$

Finally, substituting (52) and (53) into (51) establishes (50).

Part (a)-(ii). Using (50), we prove (47) by induction for interior x and for boundary x separately.

For interior x. The result in part (i) implies that for interior x,

$$
\begin{aligned}
E\left(\theta_{\nu,x}-x\right)^{0} &= O(1)=O\left(b^{0/2}\right), \\
E\left(\theta_{\nu,x}-x\right)^{1} &= O(b)=O\left(b^{(1+1)/2}\right), \\
E\left(\theta_{\nu,x}-x\right)^{2} &= O(b)=O\left(b^{2/2}\right), \\
E\left(\theta_{\nu,x}-x\right)^{3} &= O\left(b^{2}\right)=O\left(b^{(3+1)/2}\right), \\
E\left(\theta_{\nu,x}-x\right)^{4} &= O\left(b^{2}\right)=O\left(b^{4/2}\right).
\end{aligned}
$$

Hence, (47) holds for $r=0,1,2,3,4$.

Next, suppose that (47) holds for $r=0,1,2,\ldots,s$. Then, consider the order of magnitude of $E\left(\theta_{\nu,x}-x\right)^{s+1}$. If s is odd, then $s+1$ is even, and thus we need to show that $E\left(\theta_{\nu,x}-x\right)^{s+1}=O\left(b^{(s+1)/2}\right)$. By the assumption of induction, the first term on the right-hand side of (50) is bounded by $O\left\{bE\left(\theta_{\nu,x}-x\right)^{s}\right\}=O\left(b^{(s+3)/2}\right)$. Also, by the assumption of induction,

$$
b^{j}E\left(\theta_{\nu,x}-x\right)^{s-j}=\begin{cases} b^{j}O\left(b^{(s-j)/2}\right)=O\left(b^{(s+j)/2}\right) & \text{for } j=1,3,\ldots,s \\ b^{j}O\left(b^{(s-j+1)/2}\right)=O\left(b^{(s+j+1)/2}\right) & \text{for } j=2,4,\ldots,s-1. \end{cases}
$$

Since $x+b/\nu=O(1)$ for interior x, the second term on the right-hand side of (50) is bounded by $O\left(b^{(s+1)/2}\right)$. Therefore, we have $E\left(\theta_{\nu,x}-x\right)^{s+1}=O\left(b^{(s+1)/2}\right)$.

On the other hand, if s is even, then $s+1$ is odd, and thus we need to show that $E\left(\theta_{\nu,x}-x\right)^{s+1}=O\left(b^{(s+2)/2}\right)$. By the assumption of induction, the first term on the right-hand side of (50) is bounded by $O\left\{bE\left(\theta_{\nu,x}-x\right)^{s}\right\}=O\left(b^{(s+2)/2}\right)$. Also, by the assumption of induction,

$$
b^{j}E\left(\theta_{\nu,x}-x\right)^{s-j}=\begin{cases} b^{j}O\left(b^{(s-j+1)/2}\right)=O\left(b^{(s+j+1)/2}\right) & \text{for } j=1,3,\ldots,s \\ b^{j}O\left(b^{(s-j)/2}\right)=O\left(b^{(s+j)/2}\right) & \text{for } j=2,4,\ldots,s-1. \end{cases}
$$

Then, the second term on the right-hand side of (50) is bounded by $O\left(b^{(s+2)/2}\right)$. Therefore, we have $E\left(\theta_{\nu,x}-x\right)^{s+1}=O\left(b^{(s+2)/2}\right)$, and thus (47) is proven by induction for interior x.

For boundary x. The result in part (i) implies that as $x/b\to\kappa>0$,

$$
\begin{aligned}
E\left(\theta_{\nu,x}-x\right)^{0} &= O\left(b^{0}\right), E\left(\theta_{\nu,x}-x\right)^{1}=O\left(b^{1}\right), E\left(\theta_{\nu,x}-x\right)^{2}=O\left(b^{2}\right), \\
E\left(\theta_{\nu,x}-x\right)^{3} &= O\left(b^{3}\right), E\left(\theta_{\nu,x}-x\right)^{4}=O\left(b^{4}\right).
\end{aligned}
$$

Hence, (47) holds for $r=0,1,2,3,4$.

Next, suppose that (47) holds for $r = 0, 1, 2, \ldots, s$. We need to show that $E\left(\theta_{\nu,x} - x\right)^{s+1} = O\left(b^{s+1}\right)$. By the assumption of induction, the first term on the right-hand side of (50) is bounded by $O\left\{bE\left(\theta_{\nu,x} - x\right)^{s}\right\} = O\left(b^{s+1}\right)$. Also, by the assumption of induction, $b^{j}E\left(\theta_{\nu,x} - x\right)^{s-j} = b^{j}O\left(b^{s-j}\right) = O\left(b^{s}\right)$ for $j = 1, 2, \ldots, s$. Since $x + b/\nu = O(b)$ when $x = \kappa b + o(b)$, the second term on the right-hand side of (50) is bounded by $O\left(b^{s+1}\right)$. Therefore, we have $E\left(\theta_{\nu,x} - x\right)^{s+1} = O\left(b^{s+1}\right)$, and thus (47) is proven by induction for boundary x.

Part (b)-(i). The expression (12) in Tweedie (1957)[6] gives

$$E\left(\eta_{\nu,x}^{j}\right) = x^{j} K_{j-1/2}\left(\frac{\nu}{xb}\right) K_{1/2}^{-1}\left(\frac{\nu}{xb}\right),$$

where $K_{v}(z)$ is the modified Bessel function of the second kind of order v. Applying a binomial expansion and the expressions (12) and (13) on p.80 in Watson (1944), we have

$$\begin{aligned} E\left(\eta_{\nu,x} - x\right)^{r} &= \sum_{j=0}^{r} \binom{r}{j} E\left(\eta_{\nu,x}^{j}\right)(-x)^{r-j} \\ &= \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} x^{j} K_{j-1/2}\left(\frac{\nu}{xb}\right) K_{1/2}^{-1}\left(\frac{\nu}{xb}\right) x^{r-j} \\ &= x^{r} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \sum_{k=0}^{j-1} \frac{(j-1+k)!}{2^{k} k! (j-1-k)!} \left(\frac{xb}{\nu}\right)^{k}. \end{aligned} \tag{54}$$

Then, first four moments of $\eta_{\nu,x}$ around x immediately follow.

Part (b)-(ii). By (54), we can write $E\left(\eta_{\nu,x} - x\right)^{r} = x^{r} \sum_{j=0}^{r} c_{j} (xb)^{j}$ for some constants $c_0, c_1, \ldots, c_r$. To establish (48), it suffices to show that $c_0 = c_1 = 0$ for $r \geq 3$, because if this is the case, then

$$E\left(\eta_{\nu,x} - x\right)^{r} = O\left\{x^{r} \sum_{j=2}^{r} (xb)^{j}\right\} = \begin{cases} O\left(b^{2}\right) & \text{for interior } x \\ O\left(b^{r+4}\right) & \text{for boundary } x. \end{cases}$$

Since

$$\left.\frac{(j-1+k)!}{2^{k} k! (j-1-k)!} \nu^{-k}\right|_{k=0} = 1, \ \left.\frac{(j-1+k)!}{2^{k} k! (j-1-k)!} \nu^{-k}\right|_{k=1} = \frac{1}{\nu} \frac{j!}{2(j-2)!} = \frac{1}{\nu} \binom{j}{2},$$

we need to show that

$$c_0 = \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} = 0, c_1 = \frac{1}{\nu} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{2} = 0,$$

[6] The expression (12) in Tweedie (1957) contains a typographical error. It should read "$\mu_r' = \mu^{r} K_{r-1/2}(\phi) K_{1/2}^{-1}(\phi)$."

for $r \geq 3$. The first equality immediately follows from the standard result of binomial sums. Furthermore, $\binom{0}{2} = \binom{1}{2} = 0$, and thus in order to establish the second equality, it suffices to demonstrate that

$$\sum_{j=2}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{2} = 0 \tag{55}$$

for $r \geq 3$.

We prove (55) by induction. If $r = 3$, then

$$\sum_{j=2}^{3} (-1)^{3-j} \binom{3}{j} \binom{j}{2} = (-1)^{1} \binom{3}{2} \binom{2}{2} + (-1)^{0} \binom{3}{3} \binom{3}{2} = 0.$$

Next, suppose that (55) holds for some $r \geq 3$. We want to show that (55) holds for $r + 1$. By Pascal's triangle, we have

$$\begin{aligned}
& \sum_{j=2}^{r+1} (-1)^{r+1-j} \binom{r+1}{j} \binom{j}{2} \\
= \; & \sum_{j=2}^{r+1} (-1)^{r+1-j} \left\{ \binom{r}{j-1} + \binom{r}{j} \right\} \binom{j}{2} \\
= \; & \sum_{j=2}^{r+1} (-1)^{r+1-j} \binom{r}{j-1} \binom{j}{2} + \sum_{j=2}^{r+1} (-1)^{r+1-j} \binom{r}{j} \binom{j}{2} \\
\equiv \; & C + D.
\end{aligned}$$

By $\binom{r}{r+1} = 0$ and the assumption of induction,

$$D = \sum_{j=2}^{r} (-1)^{r+1-j} \binom{r}{j} \binom{j}{2} = (-1) \sum_{j=2}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{2} = 0.$$

On the other hand, by Pascal's triangle, we have

$$\begin{aligned}
C &= \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} \binom{j+1}{2} \\
&= \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} \left\{ \binom{j}{1} + \binom{j}{2} \right\} \\
&= \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{1} + \sum_{j=1}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{2} \\
&\equiv C_1 + C_2.
\end{aligned}$$

By $\binom{1}{2} = 0$ and the assumption of induction,

$$C_2 = \sum_{j=2}^{r} (-1)^{r-j} \binom{r}{j} \binom{j}{2} = 0.$$

In addition, using $\binom{r}{j}\binom{j}{1} = r\binom{r-1}{j-1}$ and the standard result of binomial sums, we have

$$C_1 = r\sum_{j=2}^{r}(-1)^{r-j}\binom{r-1}{j-1} = r\sum_{j=1}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j} = 0.$$

Therefore,

$$\sum_{j=2}^{r+1}(-1)^{r+1-j}\binom{r+1}{j}\binom{j}{2} = 0,$$

and thus (55) is proven by induction.

Part (c)-(i). Using the expression (33) in Tweedie (1957), we have

$$E\left(\xi_{\nu,x}^{j}\right) = (x-b)^{j}\sum_{k=0}^{j}\frac{(j+k)!}{2^{k}k!\,(j-k)!}\left\{\frac{b}{\nu\,(x-b)}\right\}^{k}$$

so that

$$\begin{aligned} E\left(\xi_{\nu,x}-x\right)^{r} &= \sum_{j=0}^{r}\binom{r}{j}E\left(\xi_{\nu,x}^{j}\right)(-x)^{r-j} \\ &= \sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}(x-b)^{j}\sum_{k=0}^{j}\frac{(j+k)!}{2^{k}k!\,(j-k)!}\left\{\frac{b}{\nu\,(x-b)}\right\}^{k}x^{r-j}. \end{aligned}$$

Note that

$$\begin{aligned} (x-b)^{j}\sum_{k=0}^{j}\frac{(j+k)!}{2^{k}k!\,(j-k)!}\left\{\frac{b}{\nu\,(x-b)}\right\}^{k} &= \sum_{k=0}^{j}\frac{(j+k)!}{2^{k}k!\,(j-k)!}(x-b)^{j-k}\left(\frac{b}{\nu}\right)^{k}, \\ (x-b)^{j-k} &= \sum_{l=0}^{j-k}(-1)^{l}\binom{j-k}{l}x^{j-k-l}b^{l}. \end{aligned}$$

Hence,

$$E\left(\xi_{\nu,x}-x\right)^{r} = \sum_{j=0}^{r}\sum_{k=0}^{j}\sum_{l=0}^{j-k}(-1)^{r-j+l}\binom{r}{j}\binom{j-k}{l}\frac{(j+k)!}{2^{k}k!\,(j-k)!}\nu^{-k}x^{r-k-l}b^{k+l}. \tag{56}$$

Then, first four moments of $\xi_{\nu,x}$ around x immediately follow.

Part (c)-(ii). By (56), we can write $E\left(\xi_{\nu,x}-x\right)^{r} = \sum_{j=0}^{r}d_{j}x^{r-j}b^{j}$ for some constants $d_0, d_1, \ldots, d_r$.

To establish (56), it suffices to show that $d_0 = d_1 = 0$ for $r \geq 3$, because if this is the case, then

$$E\left(\xi_{\nu,x}-x\right)^{r} = O\left(\sum_{j=2}^{r}x^{r-j}b^{j}\right) = \begin{cases} O\left(b^{2}\right) & \text{for interior } x \\ O\left(b^{r}\right) & \text{for boundary } x. \end{cases}$$

The coefficient on $x^{r}b^{0}$ can be obtained by setting $k+l=0 \Leftrightarrow (k,l)=(0,0)$ in (56). Hence,

$$d_0 = \sum_{j=0}^{r}\sum_{k=0}^{j}\sum_{l=0}^{j-k}(-1)^{r-j+l}\binom{r}{j}\binom{j-k}{l}\frac{(j+k)!}{2^{k}k!\,(j-k)!}\nu^{-k}\Bigg|_{(k,l)=(0,0)} = \sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j},$$

which is zero by the standard result of binomial sums. Moreover, the coefficient on $x^{r-1}b^1$ can be obtained by setting $k+l=1 \Leftrightarrow (k,l)=(1,0),(0,1)$ in (56). Hence,

$$\begin{aligned}
d_1 &= \left.\sum_{j=0}^{r}\sum_{k=0}^{j}\sum_{l=0}^{j-k}(-1)^{r-j+l}\binom{r}{j}\binom{j-k}{l}\frac{(j+k)!}{2^k k!\,(j-k)!}\nu^{-k}\right|_{(k,l)=(1,0)} \\
&\quad + \left.\sum_{j=0}^{r}\sum_{k=0}^{j}\sum_{l=0}^{j-k}(-1)^{r-j+l}\binom{r}{j}\binom{j-k}{l}\frac{(j+k)!}{2^k k!\,(j-k)!}\nu^{-k}\right|_{(k,l)=(0,1)} \\
&= \frac{1}{\nu}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}\frac{(j+1)!}{2\,(j-1)!}+\sum_{j=0}^{r}(-1)^{r-j-1}\binom{r}{j}\binom{j}{1} \\
&= \frac{1}{\nu}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}\binom{j+1}{2}+r\sum_{j=0}^{r}(-1)^{r-j-1}\binom{r-1}{j-1} \\
&\equiv E+F.
\end{aligned}$$

By $\binom{1}{2}=0$,

$$E=\frac{1}{\nu}\sum_{j=1}^{r}(-1)^{r-j}\binom{r}{j}\binom{j+1}{2}=\frac{C}{\nu}=0$$

for $r\geq 3$, where C is defined in the proof of part (b)-(ii) of this lemma. Furthermore, by $\binom{r-1}{-1}=0$ and the standard result of binomial sums,

$$F=(-r)\sum_{j=1}^{r}(-1)^{(r-1)-(j-1)}\binom{r-1}{j-1}=(-r)\sum_{j=0}^{r-1}(-1)^{(r-1)-j}\binom{r-1}{j}=0.$$

Therefore, $d_1=0$ for $r\geq 3$, which completes the proof. ■

Lemma B3. (Volkonskii and Razanov, 1959) *Let $V_1,\ldots,V_L$ be strongly mixing random variables measurable with respect to the σ-algebras $\mathcal{F}_{i_1}^{j_1},\ldots,\mathcal{F}_{i_L}^{j_L}$ respectively with $1\leq i_1<j_1<i_2<\cdots<j_L\leq T$, $i_{l+1}-j_l\geq w\geq 1$ and $|V_j|\leq 1$ for $j=1,\ldots,L$. Then,*

$$\left|E\left(\prod_{j=1}^{L}V_j\right)-\prod_{j=1}^{L}E\left(V_j\right)\right|\leq 16\left(L-1\right)\alpha\left(w\right),$$

where $\alpha\left(w\right)$ is the strong mixing coefficient.

References

[1] Ahn, D. H., and B. Gao (1999): "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, 12, 721 - 762.

[2] Bandi, F. M. (2002): "Short-Term Interest Rate Dynamics: A Spatial Approach," *Journal of Financial Economics*, 65, 73 - 110.

[3] Bandi, F. M., and P. C. B. Phillips (2003): "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241 - 283.

[4] Bouezmarni, T., and J.-M. Rolin (2003): "Consistency of the Beta Kernel Density Function Estimator," *Canadian Journal of Statistics*, 31, 89 - 98.

[5] Bouezmarni, T., and J. V. K. Rombouts (2006a): "Nonparametric Density Estimation for Positive Time Series," CORE Discussion Paper 2006/85, Université Catholique de Louvain.

[6] Bouezmarni, T., and J. V. K. Rombouts (2006b): "Density and Hazard Rate Estimation for Censored and α-mixing Data Using Gamma Kernels," CORE Discussion Paper 2006/118, Université Catholique de Louvain.

[7] Bouezmarni, T., and O. Scaillet (2005): "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data," *Econometric Theory*, 21, 390 - 412.

[8] Brown, B. M., and S. X. Chen (1999): "Beta-Bernstein Smoothing for Regression Curves with Compact Support," *Scandinavian Journal of Statistics*, 26, 47 - 59.

[9] Burman, P., E. Chow, and D. Nolan (1994): "A Cross-Validatory Method for Dependent Data," *Biometrika*, 81, 351 - 358.

[10] Cai, Z. (2001): "Weighted Nadaraya-Watson Regression Estimation," *Statistics & Probability Letters*, 51, 307 - 318.

[11] Cai, Z. (2002): "Regression Quantiles for Time Series," *Econometric Theory*, 18, 169 - 192.

[12] Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2007): "Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions," *Journal of Econometrics*, 140, 529 - 573.

[13] Chapman, D. A., and N. D. Pearson (2000): "Is the Short Rate Drift Actually Nonlinear?," *Journal of Finance*, 55, 355 - 388.

[14] Chen, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131 - 145.

[15] Chen, S. X. (2000a): "Beta Kernel Smoothers for Regression Curves," *Statistica Sinica*, 10, 73 - 91.

[16] Chen, S. X. (2000b): "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471 - 480.

[17] Chen, S. X. (2002): "Local Linear Smoothers Using Asymmetric Kernels," *Annals of the Institute of Statistical Mathematics*, 54, 312 - 323.

[18] Chen, S. X., and P. Hall (1993): "Smoothed Empirical Likelihood Confidence Intervals for Quantiles," *Annals of Statistics*, 21, 1166 - 1181.

[19] Chen, X., M. Carrasco, and L. P. Hansen (1999): "Nonlinearity and Temporal Dependence," Working paper, University of Chicago.

[20] Chu, C. K. (1989): "Some Results in Nonparametric Regression," unpublished Ph.D. thesis, University of North Carolina, Chapel Hill.

[21] Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross. (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385 - 408.

[22] De Gooijer, J. G., and D. Zerom (2003): "On Conditional Density Estimation," *Statistica Neerlandica*, 57, 159 - 176.

[23] 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.

[24] Fan, J. (1993): "Local Linear Regression Smoothers and Their Minimax Efficiencies," *Annals of Statistics*, 21, 196 - 216.

[25] Fan, J., and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008 - 2036.

[26] Fan, J., and C. Zhang (2003): "A Reexamination of Diffusion Estimators with Applications to Financial Model Validation," *Journal of the American Statistical Association*, 98, 118 - 134.

[27] Fernandes, M., and P. K. Monteiro (2005): "Central Limit Theorem for Asymmetric Kernel Functionals," *Annals of the Institute of Statistical Mathematics*, 57, 425 - 442.

[28] Florens-Zmirou, D. (1993): "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30, 790 - 804.

[29] Gasser, T., and H.-G. Müller (1979): "Kernel Estimation of Regression Functions," in T. Gasser and M. Rosenblatt (eds.), *Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2 - 4, 1979.* Berlin: Springer-Verlag, 23 - 68.

[30] Györfi, L., W. Härdle, P. Sarda, and P. Vieu (1989): *Nonparametric Curve Estimation from Time Series*, Lecture Notes in Statistics 60, Berlin: Springer-Verlag.

[31] Hagmann, M., and O. Scaillet (2007): "Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators," *Journal of Econometrics*, 141, 213 - 249.

[32] Hansen, B. (2006): "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory*, forthcoming.

[33] Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

[34] Hall, P., and B. Presnell (1999): "Intentionally Biased Bootstrap Methods," *Journal of the Royal Statistical Society*, *Series B*, 61, 143 - 158.

[35] Hall, P., R. C. L. Wolff, and Q. Yao (1999): "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154 - 163.

[36] Hjort, N. L., and M. C. Jones (1996): "Locally Parametric Nonparametric Density Estimation," *Annals of Statistics*, 24, 1619 - 1647.

[37] Jiang, G. J. (1998): "Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Prices of Derivative Securities," *Journal of Financial and Quantitative Analysis*, 33, 465 - 497.

[38] Jiang, G. J., and J. L. 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.

[39] Jones, C. S. (2003): "Nonlinear Mean Reversion in the Short-Term Interest Rate." *Review of Financial Studies,* 16, 793 - 843.

[40] Jones, M. C., and D. A. Henderson (2007): "Kernel-Type Density Estimation on the Unit Interval," *Biometrika*, forthcoming.

[41] Lu, Z., and O. Linton (2007): "Local Linear Fitting under Near Epoch Dependence," *Econometric Theory*, 23, 37 - 70.

[42] Masry, E., and J. Fan (1997): "Local Polynomial Estimation of Regression Functions for Mixing Processes," *Scandinavian Journal of Statistics*, 24, 165 - 179.

[43] Nadaraya, É. A. (1964): "On Estimating Regression," *Theory of Probability and Its Applications*, 9, 141 - 142.

[44] Nicolau, J. (2003): "Bias Reduction in Nonparametric Diffusion Coefficient Estimation," *Econometric Theory*, 19, 754 - 777.

[45] Phillips, P. C. B., and J. Yu (2005): "Jackknifing Bond Option Prices," *Review of Financial Studies*, 18, 707 - 742.

[46] Rice, J. (1984): "Boundary Modification for Kernel Regression," *Communications in Statistics: Theory and Methods*, 13, 893 - 900.

[47] Rosenblatt, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Science*, *USA*, 42, 43 - 47.

[48] Scaillet, O. (2004): "Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels," *Journal of Nonparametric Statistics*, 16, 217 - 226.

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

[50] Tweedie, M. C. K. (1957): "Statistical Properties of Inverse Gaussian Distributions. I.," *Annals of Mathematical Statistics*, 28, 362 - 377.

[51] Volkonskii, V. A., and Y. A. Razanov (1959): "Some Limit Theorems for Random Functions. I," *Theory of Probability and Its Applications*, 4, 178 - 197.

[52] Watson, G. N. (1944): *A Treatise on the Theory of Bessel Functions*, 2nd edition. Cambridge: Cambridge University Press.

[53] Watson, G. S. (1964): "Smooth Regression Analysis," *Sankhyā*, *Series A*, 26, 359 - 372.

Table 1. Monte Carlo statistics of bond and option prices based on nonparametric (Gaussian NW and Gamma NW) and OLS estimators in the CIR model with $\kappa = 0.21459$, $\theta = 0.085711$ and $\sigma = 0.0783$.

	bond price	call option price
true price	0.83763	1.93118
OLS		
median estimate	0.81979	0.69922
mean estimate	0.81546	1.09872
std. deviation	0.03242	1.18950
90% confidence interval	[0.75621, 0.85952]	[0.00000, 3.46050]
Gaussian NW		
median estimate	0.84073	2.15101
mean estimate	0.83838	2.28154
std. deviation	0.03568	2.80581
90% confidence interval	[0.79089, 0.87444]	[0.02886, 4.73207]
Gamma NW		
median estimate	0.83943	2.05317
mean estimate	0.83712	2.11356
std. deviation	0.02505	1.32988
90% confidence interval	[0.79375, 0.87100]	[0.05484, 4.34292]

Notes: The statistics in the table are computed from 5,000 samples generated from the CIR model with $\triangle = 1/52$ and $T = 600$. 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 5% are computed analytically for the OLS estimator and by Monte Carlo simulation for the nonparametric estimators.

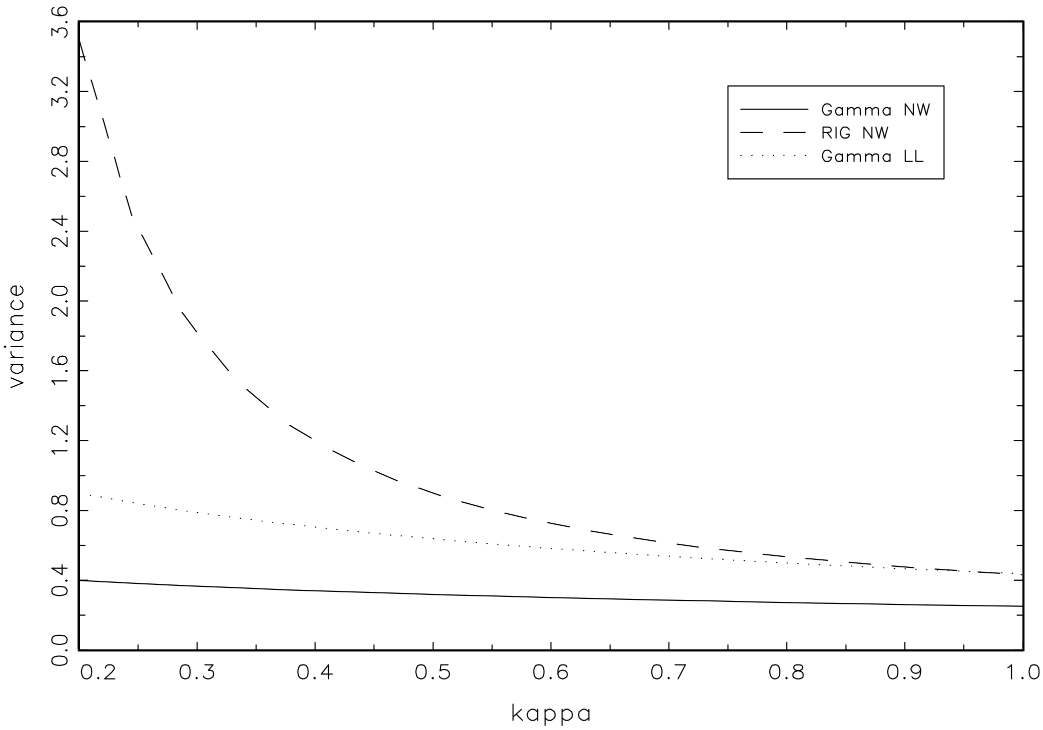

FIGURE 1. Differences in asymptotic variances of $\hat{m}_G^{nw}(x)$, $\hat{m}_{RIG}^{nw}(x)$ and $\hat{m}_G^{ll}(x)$ for boundary x as a function of $\kappa \in [0.2, 1]$.

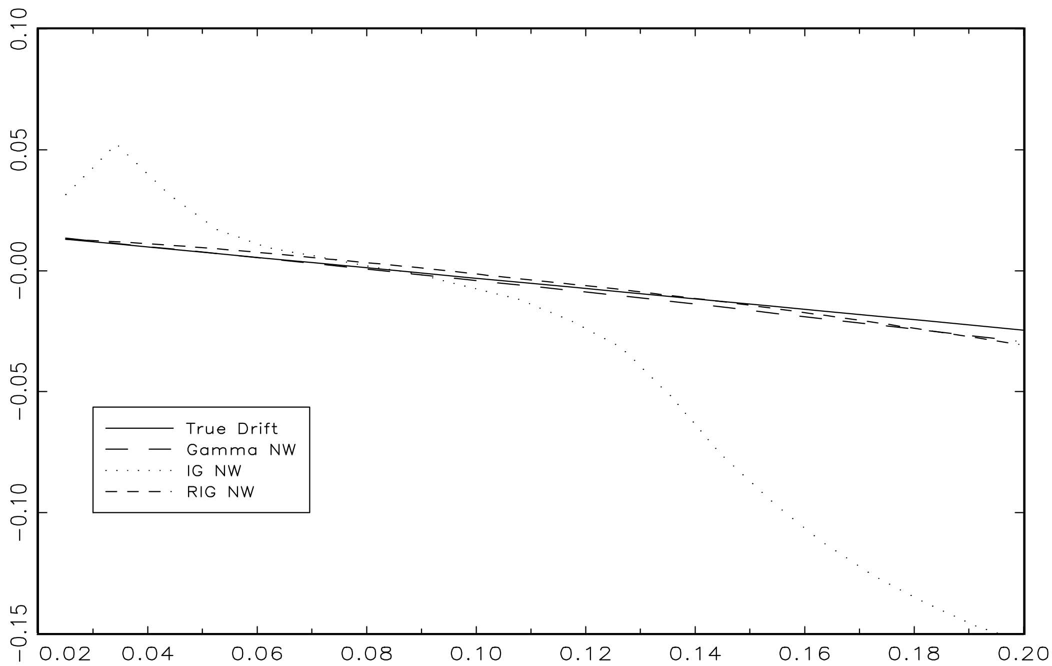

FIGURE 2. Median Monte Carlo drift estimates based on asymmetric (Gamma, IG and RIG) estimators from CIR model with $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$ and smoothing parameter equal to $2std(r_t)T^{-1/5}$, where $std(r_t)$ is the standard deviation of the data.

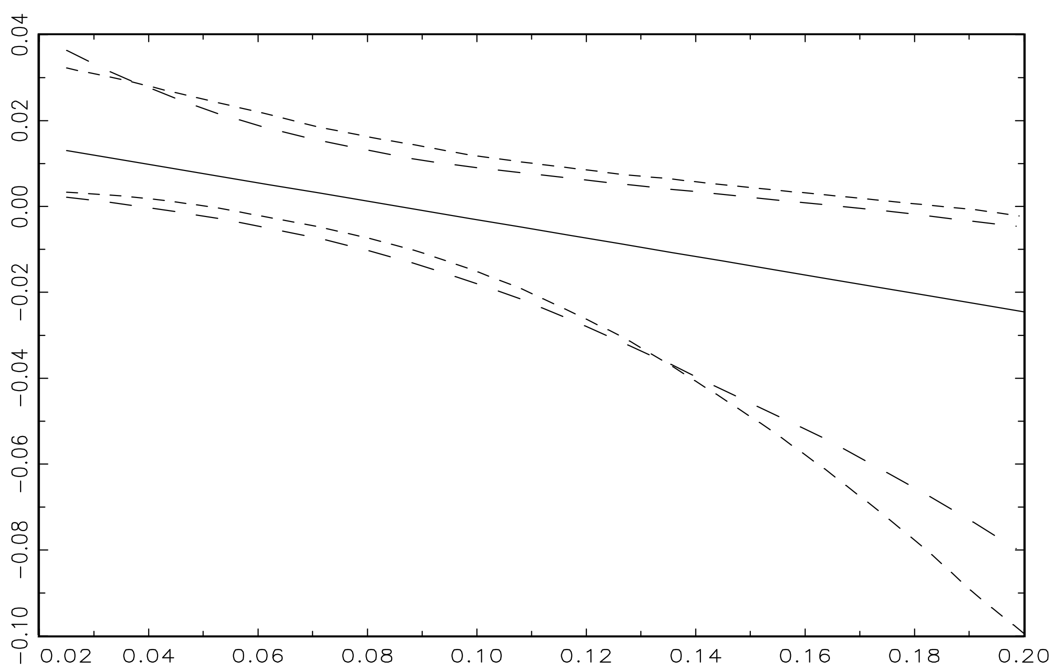

FIGURE 3. 90% Monte Carlo confidence intervals of the asymmetric kernel drift estimates from CIR model with $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$ and smoothing parameter equal to $2std(r_t)T^{-1/5}$, where $std(r_t)$ is the standard deviation of the data. Long dashes: Gamma NW estimator; short dashes: RIG NW estimator.

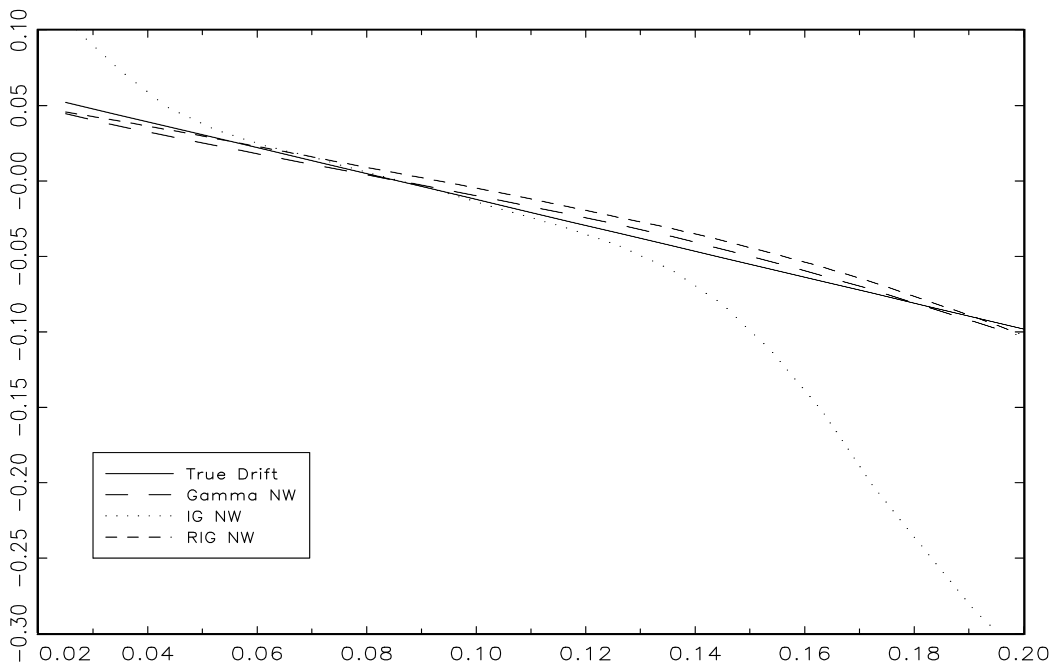

FIGURE 4. Median Monte Carlo drift estimates based on asymmetric (Gamma, IG and RIG) estimators from CIR model with $(\kappa, \theta, \sigma) = (0.85837,\ 0.085711,\ 0.1566)$ and smoothing parameter equal to $1.5std(r_t)T^{-1/5}$, where $std(r_t)$ is the standard deviation of the data.

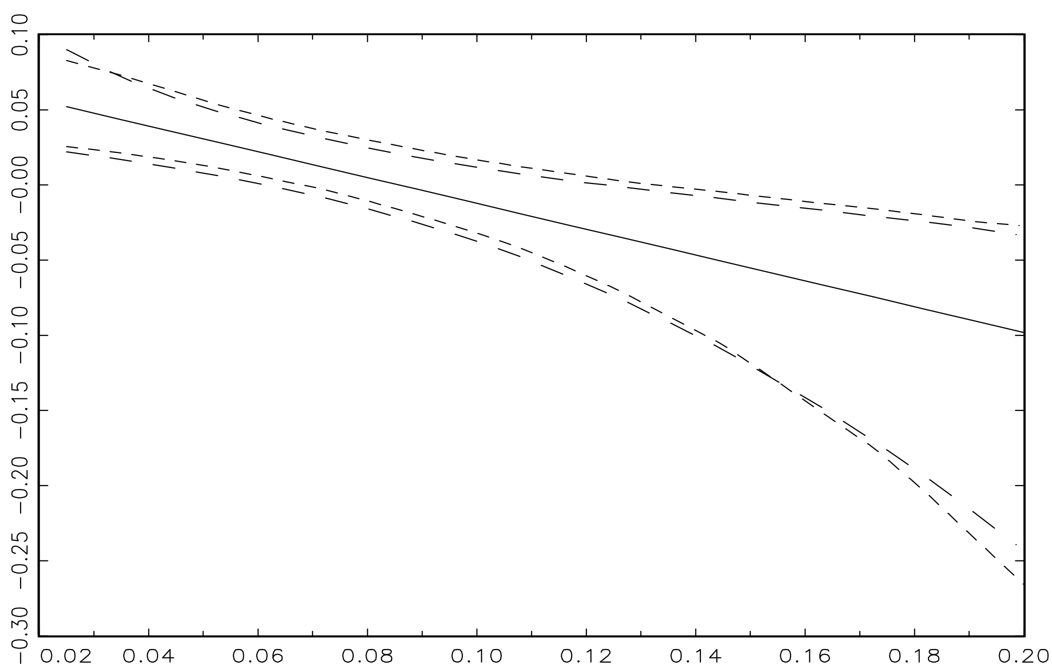

FIGURE 5. 90% Monte Carlo confidence intervals of the asymmetric kernel drift estimates from CIR model with $(\kappa, \theta, \sigma) = (0.85837, 0.085711, 0.1566)$ and smoothing parameter equal to $1.5std(r_t)T^{-1/5}$, where $std(r_t)$ is the standard deviation of the data. Long dashes: Gamma NW estimator; short dashes: RIG NW estimator.

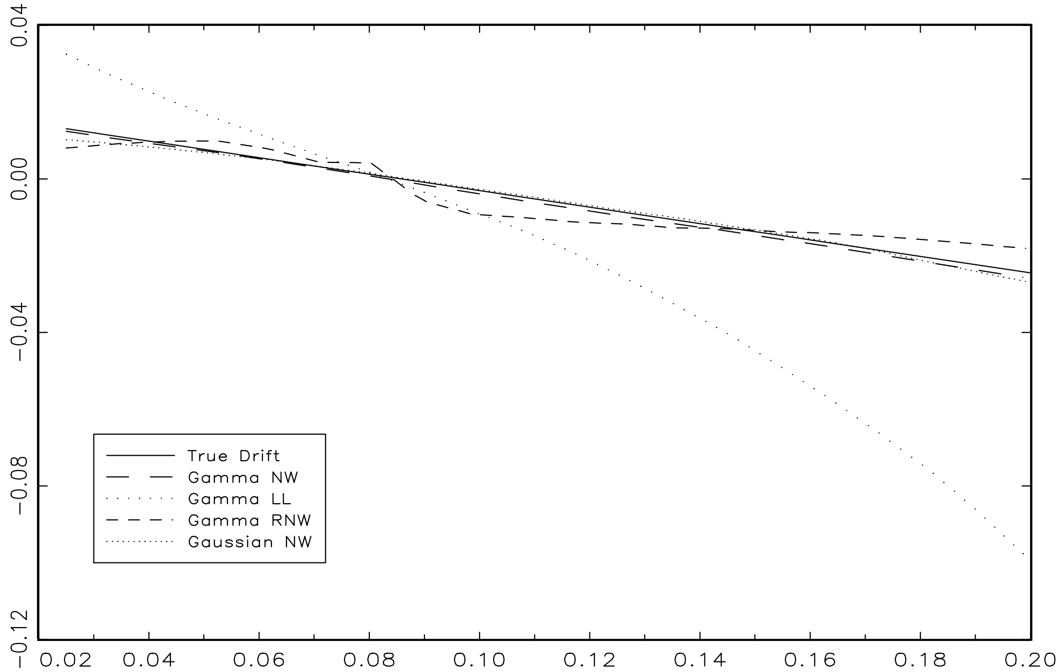

FIGURE 6. Median Monte Carlo drift estimates from CIR model with $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$. The smoothing parameter is selected by h-block cross validation with $h = 30$.

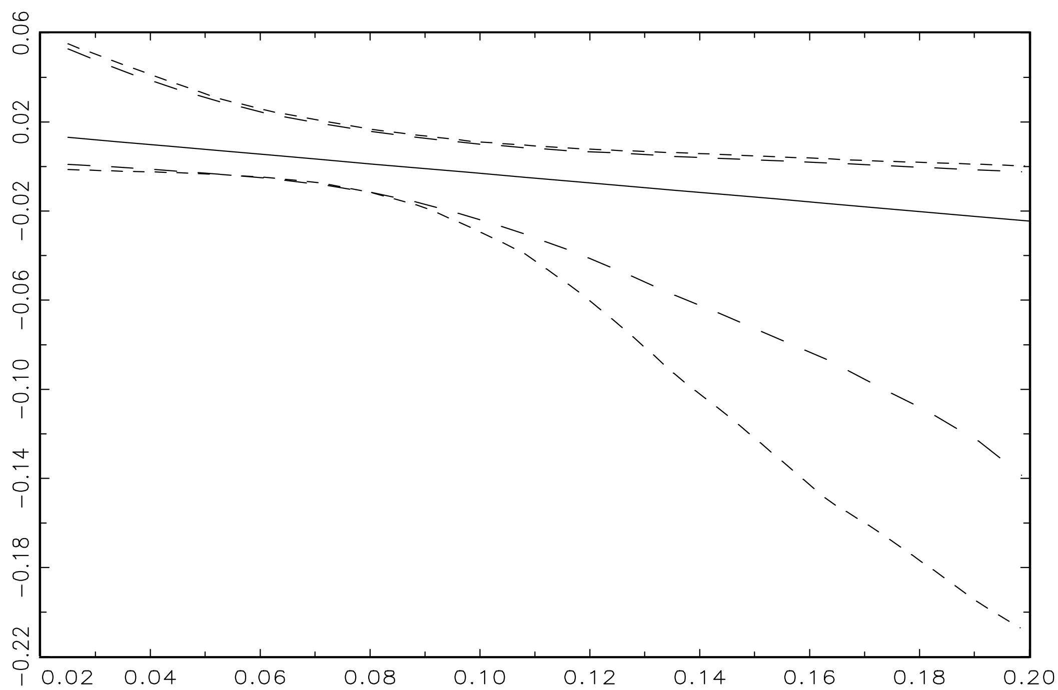

FIGURE 7. 90% Monte Carlo confidence intervals of the drift estimates from CIR model with $(\kappa, \theta, \sigma) = (0.21459, 0.085711, 0.0783)$ and smoothing parameter selected by h-block cross validation with $h = 30$. Long dashes: Gamma NW estimator; short dashes: Gaussian NW estimator.

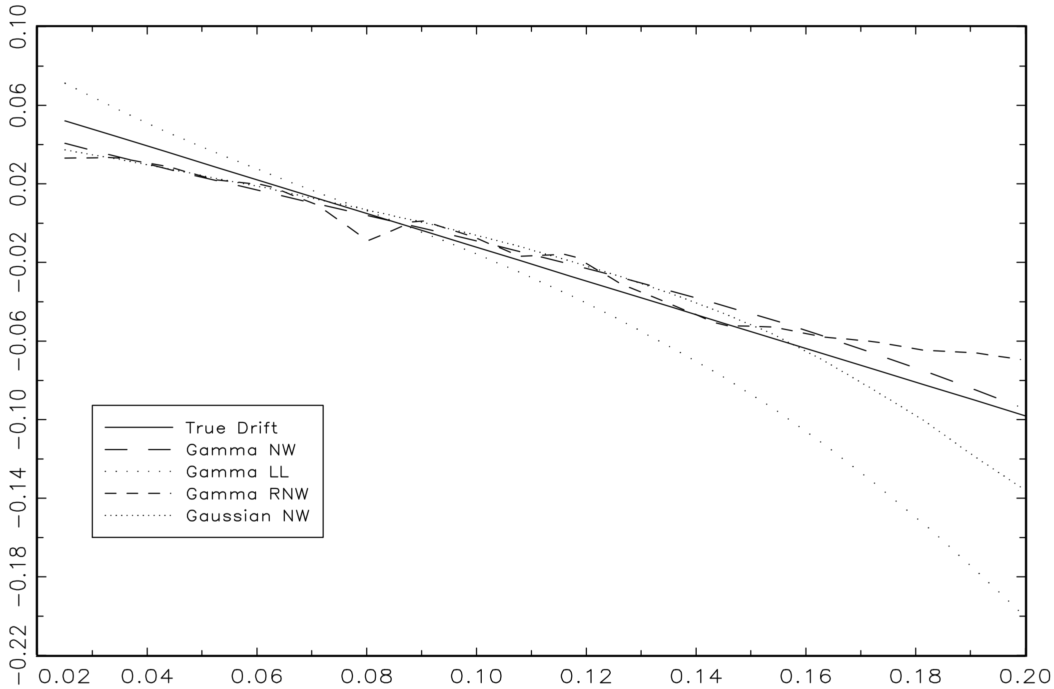

FIGURE 8. Median Monte Carlo drift estimates from CIR model with $(\kappa, \theta, \sigma) = (0.85837, 0.085711, 0.1566)$. The smoothing parameter is selected by h-block cross validation with $h = 30$.

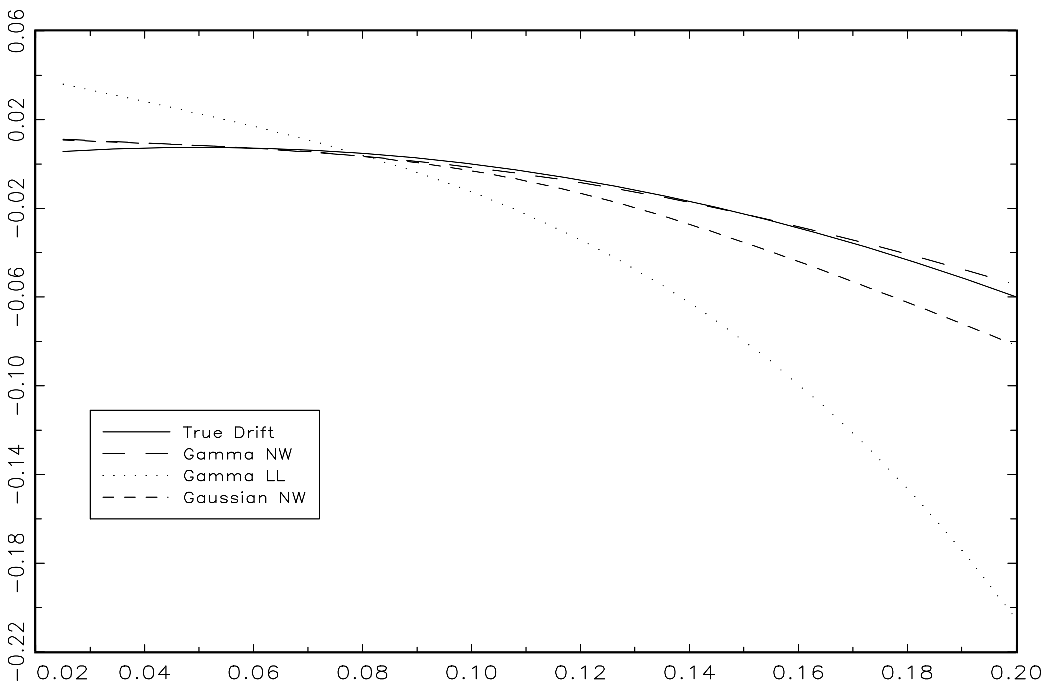

Figure 9. Median Monte Carlo drift estimates from Ahn and Gao's (1999) model with smoothing parameter selected by h-block cross validation with $h = 30$.

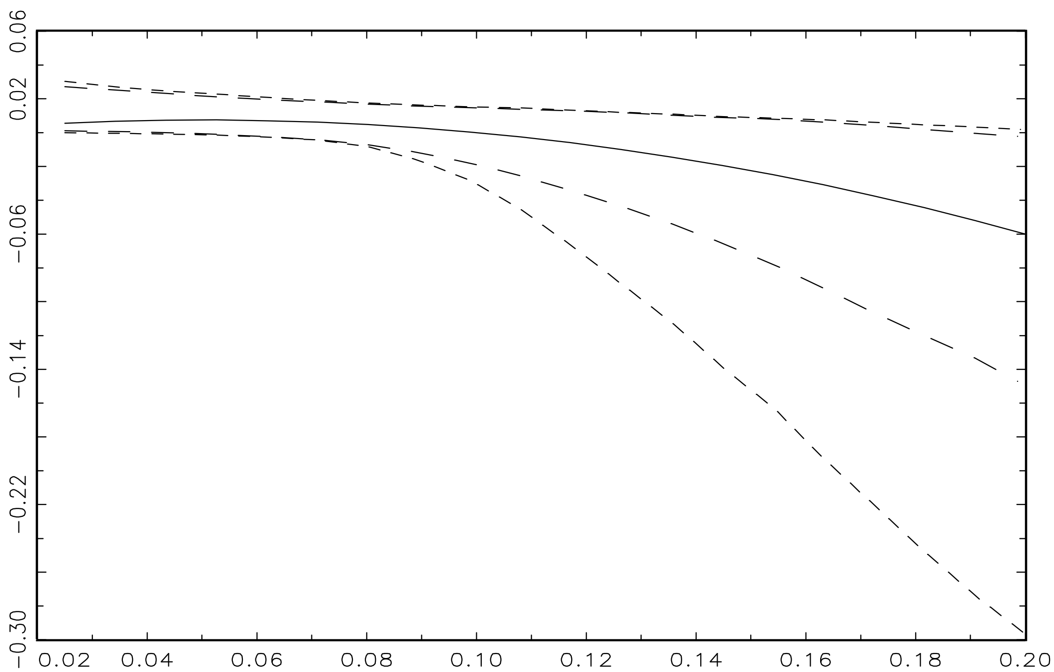

Figure 10. 90% Monte Carlo confidence intervals of the drift estimates from Ahn and Gao's (1999) model with smoothing parameter selected by h-block cross validation with $h = 30$. Long dashes: Gamma NW estimator; short dashes: Gaussian NW estimator.