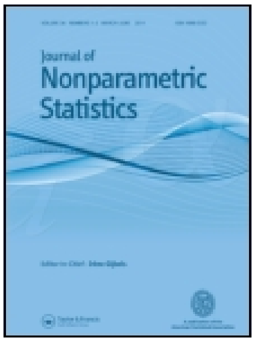

Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Family of the generalised gamma kernels: a generator of asymmetric kernels for nonnegative data

Masayuki Hirukawa & Mari Sakudo

Journal of Nonparametric Statistics, 2015
Vol. 27, No. 1, 41–63, http://dx.doi.org/10.1080/10485252.2014.998669

Taylor & Francis
Taylor & Francis Group

Family of the generalised gamma kernels: a generator of asymmetric kernels for nonnegative data

Masayuki Hirukawa[a*] and Mari Sakudo[b†]

[a]*Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka 572-8508, Japan;*
[b]*Research Institute of Capital Formation, Development Bank of Japan, 9-7, Otemachi 1-chome, Chiyoda-ku, Tokyo 100-8178, Japan*

(*Received 19 June 2013; accepted 25 November 2014*)

Unlike symmetric kernels, so far exploring asymptotics on asymmetric kernels has relied on diversified approaches. This paper proposes a family of the generalised gamma (GG) kernels that is built on the probability density function of the GG distribution [Stacy, E.W. (1962), 'A Generalization of the Gamma Distribution', *Annals of Mathematical Statistics*, 33, 1187–1192] and a few common conditions. The family can generate asymmetric kernels that share appealing properties with the modified gamma kernel [Chen, S.X. (2000), 'Probability Density Function Estimation Using Gamma Kernels', *Annals of the Institute of Statistical Mathematics*, 52, 471–480]. Asymptotics on the kernels generated from the family can be delivered by manipulating the conditions directly, as with symmetric kernels.

Keywords: asymmetric kernel; boundary correction; density estimation; kernel smoothing; generalised gamma kernels

MSC 2010: 62G07; 62G20

1. Introduction

Researchers and policy-makers are often interested in the distributions of nonnegative economic and financial variables including incomes, wages, short-term interest rates, and insurance claims. The distributions are empirically characterised by two stylised facts, namely, (i) a natural boundary at the origin and (ii) concentration of observations near the boundary and a long tail with sparse data. When their parametric specifications are found to be inappropriate, we resort to nonparametric density estimation. However, in estimating the distributions with symmetric kernels, we must make two distinct modifications for the standard smoothing technique *simultaneously*. For (i), boundary correction methods should be employed; see, for instance, Section 3 of Karunamuni and Alberts (2005) for a concise review of the methods. For (ii), adaptive smoothing such as variable bandwidth methods (e.g. Abramson 1982) would be a remedy.

Recently, asymmetric kernels with support on $[0, \infty)$ (e.g. Chen 2000; Jin and Kawczak 2003; Scaillet 2004) have emerged as a viable alternative that can accommodate the stylised facts. Although the kernels are relatively new in the literature, several papers report favourable evidence from applying them to empirical models in economics and finance; see, for instance,

*Corresponding author. Email: hirukawa@econ.setsunan.ac.jp
†Institute for Research in Contemporary Political and Economic Affairs, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan.

Section 1 of Gospodinov and Hirukawa (2012) for a non-exhaustive list of the papers. The reason why asymmetric kernels tend to work well for the distributions with two stylised facts is their property as a combination of a boundary correction device and adaptive smoothing that has effect similar to the variable bandwidth methods.

A cumbersome aspect of asymmetric kernels, however, is that so far exploring their asymptotic properties has relied on kernel-specific and thus diversified approaches. In contrast, symmetric kernels are built on a set of common conditions, and their asymptotic properties can be implied straightforwardly by the conditions. Then, this paper proposes a new class of asymmetric kernels that consists of a specific functional form and a set of common conditions. Our aim is that as long as an asymmetric kernel is generated from this class, we can deliver its asymptotic properties by manipulating the conditions directly, rather than employing the techniques peculiar to the kernel.

A key issue of establishing the new class of asymmetric kernels is to choose the shape or functional form. While an apparent shape restriction (i.e. symmetry about the origin) is imposed on symmetric kernels, there is no guidance on shapes of asymmetric kernels; actually, any shapes are admissible as long as the kernels have support on $[0, \infty)$. Recognising that most of the papers on asymmetric kernels report superior finite sample performance of the modified gamma (MG) kernel by Chen (2000), we consider the kernel a reasonable benchmark of the functional form. Accordingly, among 'close-cousins' of the gamma probability density function (pdf), the pdf of the generalised gamma (GG) distribution (Stacy 1962) is chosen as the functional form. Combining the pdf with a set of common conditions, we finally define a family of the GG kernels. As special cases, the GG kernels nest not only the MG kernel but also the newly proposed Weibull and Nakagami-*m* kernels, which are built on the Weibull and Nakagami-*m* (Nakagami 1943, 1960) pdfs, respectively.

The GG kernels are designed to preserve all appealing properties that the MG kernel possesses. First, by construction, the GG kernels are free of boundary bias and always generate nonnegative density estimates everywhere. Second, when best implemented, each GG density estimator attains Stone's (1980) optimal convergence rate in the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators. Third, the leading bias over the interior region for each GG density estimator contains only the second-order derivative of the true density over the interior region, whereas the one for density estimators using some other asymmetric kernels (e.g. the gamma kernel by Chen 2000) consists of terms involving the first- and second-order density derivatives. Fourth, the variance of the GG estimator tends to decrease as the design point moves away from the boundary. This property is particularly advantageous to estimating the distributions that have a long tail with sparse data. It is worth emphasising that these properties can be demonstrated by manipulating the common conditions directly.

This paper also demonstrates three additional properties of the GG density estimator. While the first property, applicability of two multiplicative bias correction (MBC) techniques studied in Hirukawa (2010) and Hirukawa and Sakudo (2014), is based on random sampling, the remaining two properties allow for dependent sampling. After the GG density estimator using weakly dependent data are shown to admit the same first-order bias and variance approximations as those using independent observations, it is proven that the density estimator is consistent even when the true density becomes unbounded at the boundary. The gamma and MG kernels are known to share all three properties, and thus it is demonstrated that all in all, the GG kernels are endowed with the same attractive properties.

The remainder of this paper is organised as follows. Section 2 provides the definition of a family of the GG kernels and introduces three special cases. Convergence properties of the GG density estimator are also provided. Section 3 investigates three appealing properties of the GG density estimator. Section 4 conducts Monte Carlo simulations to examine finite sample properties of the GG density estimator. Section 5 summarises the main results of the paper and suggests some research extensions. Proofs of Theorems 1 and 2 are given in the Appendix. Besides, proofs of Theorems 4–6 and comprehensive simulation results are available on the first author's webpage.

This paper adopts the following notational conventions: $\Gamma(a) = \int_0^\infty y^{a-1}\exp(-y)\,\mathrm{d}y$ $(a > 0)$ is the gamma function; and $\mathbf{1}\{\cdot\}$ signifies an indicator function. The expression '$X \stackrel{d}{=} Y$' reads 'A random variable X obeys the distribution Y.' The expression '$X_n \sim Y_n$' is used whenever $X_n/Y_n \to 1$ as $n \to \infty$. Lastly, in order to describe different asymptotic properties of an asymmetric kernel estimator across positions of the design point $x > 0$, we denote by 'interior x' and 'boundary x' a design point x that satisfies $x/b \to \infty$ and $x/b \to \kappa$ for some $0 < \kappa < \infty$ as $n \to \infty$, respectively.

2. A family of the GG kernels

2.1. *Definition*

Let Y be distributed by the GG distribution $\mathrm{GG}(\alpha, \beta, \gamma)$. Then, Y has the pdf

$$p(y; \alpha, \beta, \gamma) = \frac{\gamma y^{\alpha-1}\exp\{-(y/\beta)^\gamma\}}{\beta^\alpha \Gamma(\alpha/\gamma)}\mathbf{1}\{y \geq 0\}. \tag{1}$$

It is known that the mth uncentred moment of Y is given by

$$E(Y^m) = \beta^m \frac{\Gamma\{(\alpha+m)/\gamma\}}{\Gamma(\alpha/\gamma)}. \tag{2}$$

We now provide the definition of a family of the GG kernels that consists of several common conditions. Before proceeding, it could be beneficial to relate the definition to probability density estimation. Below we present a set of regularity conditions for asymmetric kernel density estimation.

Assumption 1 *The random sample* $\{X_i\}_{i=1}^n$ *is drawn from a univariate distribution with a pdf* f *having support on* $[0, \infty)$.

Assumption 2 f *is twice continuously differentiable.*

Assumption 3 *The smoothing parameter* $b (= b_n > 0)$ *satisfies* $b + (nb)^{-1} \to 0$ *as* $n \to \infty$.

To generate a kernel from the pdf (1), we allow (α, β, γ) to be a function of the design point $x > 0$ and the smoothing parameter b, as in Chen (2000), Jin and Kawczak (2003) and Scaillet (2004). To put it another way, whenever we refer to the triplet (α, β, γ), it should be interpreted as a short-handed notation of $(\alpha, \beta, \gamma) = (\alpha_b(x), \beta_b(x), \gamma_b(x))$ unless otherwise noted. Let the crude GG kernel be $K_{\mathrm{GG0}}(u; x, b) := p(u; \alpha, \beta, \gamma)$. Given the random sample $\{X_i\}_{i=1}^n$, we have the density estimator $\hat{f}_{\mathrm{GG0}}(x) = (1/n)\sum_{i=1}^n K_{\mathrm{GG0}}(X_i; x, b)$.

Let $\vartheta_x \stackrel{d}{=} \mathrm{GG}(\alpha, \beta, \gamma)$. Under Assumptions 2 and 3, a second-order Taylor expansion of $E\{\hat{f}_{\mathrm{GG0}}(x)\}$ around $\vartheta_x = x$ yields $E\{\hat{f}_{\mathrm{GG0}}(x)\} = f(x) + E(\vartheta_x - x)f'(x) + (1/2)E(\vartheta_x - x)^2 f''(x) + o\{E(\vartheta_x - x)^2\}$. It follows that unless $E(\vartheta_x) = x$ exactly (at least for interior x), the leading bias of $\hat{f}_{\mathrm{GG0}}$ would contain the term involving f', which is less desirable. Although there are numerous choices of (α, β, γ) that can achieve $E(\vartheta_x) = x$, we adopt the simplest resolution that we set $\beta = x$ for interior x and employ the pdf of $\mathrm{GG}(\alpha, \beta\Gamma(\alpha/\gamma)/\Gamma\{(\alpha+1)/\gamma\}, \gamma)$ (which can be obtained by the change of variable $Z := [\Gamma(\alpha/\gamma)/\Gamma\{(\alpha+1)/\gamma\}]Y$ in Equation (1)) as the kernel. In the end, we reach the following definition of a family of the GG kernels.

DEFINITION 1 *Let* $(\alpha, \beta, \gamma) = (\alpha_b(x), \beta_b(x), \gamma_b(x)) \in \mathbb{R}_+^3$ *be a continuous function of the design point* x *and the smoothing parameter* b. *For such* (α, β, γ), *consider the pdf of* $\mathrm{GG}(\alpha, \beta\Gamma(\alpha/\gamma)/\Gamma\{(\alpha+1)/\gamma\}, \gamma)$, *i.e.*

$$K_{\mathrm{GG}}(u; x, b) = \frac{\gamma u^{\alpha-1} \exp[-\{\frac{u}{\beta\Gamma(\alpha/\gamma)/\Gamma((\alpha+1)/\gamma)}\}^{\gamma}]}{\{\beta\Gamma(\alpha/\gamma)/\Gamma((\alpha+1)/\gamma)\}^{\alpha}\Gamma(\alpha/\gamma)} \mathbf{1}\{u \geq 0\}. \tag{3}$$

This pdf is said to be a family of the GG kernels if it satisfies each of the following conditions:

Condition 1.

$$\beta = \begin{cases} x & \text{for } x \geq C_1 b \\ \varphi_b(x) & \text{for } x \in [0, C_1 b) \end{cases},$$

where $0 < C_1 < \infty$ *is some constant, the function* $\varphi_b(x)$ *satisfies* $C_2 b \leq \varphi_b(x) \leq C_3 b$ *for some constants* $0 < C_2 \leq C_3 < \infty$, *and the connection between* x *and* $\varphi_b(x)$ *at* $x = C_1 b$ *is smooth.*

Condition 2. $\alpha, \gamma \geq 1$, *and for* $x \in [0, C_1 b)$, α *satisfies* $1 \leq \alpha \leq C_4$ *for some constant* $1 \leq C_4 < \infty$. *Moreover, connections of* α *and* γ *at* $x = C_1 b$, *if any, are smooth.*

Condition 3.

$$M_b(x) := \frac{\Gamma(\alpha/\gamma)\Gamma((\alpha+2)/\gamma)}{\{\Gamma((\alpha+1)/\gamma)\}^2} = \begin{cases} 1 + (C_5/x)b + o(b) & \text{for } x \geq C_1 b \\ O(1) & \text{for } x \in [0, C_1 b) \end{cases}$$

for some constant $0 < |C_5| < \infty$.

Condition 4.

$$H_b(x) := \frac{\Gamma(\alpha/\gamma)\Gamma(2\alpha/\gamma)}{2^{1/\gamma}\Gamma((\alpha+1)/\gamma)\Gamma((2\alpha-1)/\gamma)} = \begin{cases} 1 + o(1) & \text{for interior } x \\ O(1) & \text{for boundary } x \end{cases}.$$

Condition 5.

$$A_{b,\nu}(x) := \left\{\frac{\gamma\Gamma((\alpha+1)/\gamma)}{\beta}\right\}^{\nu-1} \frac{\Gamma\{(\nu(\alpha-1)+1)/\gamma\}}{\nu^{(\nu(\alpha-1)+1)/\gamma}\{\Gamma(\alpha/\gamma)\}^{2\nu-1}}$$
$$\sim \begin{cases} V_{\mathrm{I}}(\nu)(xb)^{(1-\nu)/2} & \text{for interior } x \\ V_{\mathrm{B}}(\nu)b^{1-\nu} & \text{for boundary } x \end{cases}, \quad \nu \in \mathbb{R}_+,$$

where constants $0 < V_{\mathrm{I}}(\nu), V_{\mathrm{B}}(\nu) < \infty$ *depend only on* ν.

Conditions 1 and 2 form a legitimate kernel from the GG pdf. It follows from $\varphi_b(x) = O(b)$ uniformly on $[0, C_1 b)$ in Condition 1 that the kernel is well defined in the vicinity of the origin. While $\alpha \geq 1$ in Condition 2 also ensures boundedness of the kernel near the origin, $\gamma \geq 1$ controls the tail-behaviour of the kernel and establishes an exponential rate of the tail decay. The condition also allows for each of α and γ to be a piecewise function of (x, b) like β, where the connection is made at $x = C_1 b$; the common connection point simplifies asymptotic analyses substantially. In addition, smooth connection requirements in Conditions 1 and 2 are inspired by the construction of the MG kernel; see Chen (2000, p. 473) for details. Moreover, Condition 3 and Conditions 4 and 5 are the requirements for valid approximations to the bias and variance of the GG density estimator

$$\hat{f}_{\mathrm{GG}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathrm{GG}}(X_i; x, b),$$

respectively. Although all the common conditions appear to be high-level ones, it is not hard to find a few special cases that satisfy them. Examples of the GG kernels are provided shortly.

2.2. *Convergence properties of probability density estimators using the GG kernels*

2.2.1. *Local property*

Bias-variance tradeoff. Before providing examples of the GG kernels, readers may wonder whether the kernels are a truly legitimate one, that is, they can yield a consistent density estimator. To answer this question, we begin this section with presenting the theorem on approximations to the bias and variance of $\hat{f}_{\mathrm{GG}}(x)$.

THEOREM 1 *Under Assumptions 1–3, the bias of $\hat{f}_{\mathrm{GG}}(x)$ can be approximated by* $\mathrm{Bias}\{\hat{f}_{\mathrm{GG}}(x)\} \sim B_1(x,f)b$, *where*

$$B_1(x,f) = \begin{cases} \left(\dfrac{C_5}{2}\right) x f''(x) & \text{for } x \geq C_1 b \\ \xi_b(x) f'(x) & \text{for } x \in [0, C_1 b) \end{cases},$$

and $\xi_b(x) = \{\varphi_b(x) - x\}/b = O(1)$. On the other hand, the variance of $\hat{f}_{\mathrm{GG}}(x)$ can be approximated by

$$\mathrm{Var}\{\hat{f}_{\mathrm{GG}}(x)\} \sim \begin{cases} (nb^{1/2})^{-1} V_{\mathrm{I}}(2) f(x)/\sqrt{x} & \text{for interior } x, \\ (nb)^{-1} V_{\mathrm{B}}(2) f(x) & \text{for boundary } x. \end{cases}$$

The theorem states that Conditions 1–5 do lead to familiar properties of asymmetric kernel density estimators.[1] By construction, $\hat{f}_{\mathrm{GG}}(x)$ is free of boundary bias and nonnegative everywhere. The bias of $\hat{f}_{\mathrm{GG}}(x)$ is $O(b)$, and its variance is $O\{(nb^{1/2})^{-1}\}$ for interior x and $O\{(nb)^{-1}\}$ for boundary x. Furthermore, as pointed out by Chen (2000) and Scaillet (2004), a unique feature of asymmetric kernel density estimators is that the variance coefficient decreases as x increases. This is shared with $\hat{f}_{\mathrm{GG}}(x)$ in that $V_{\mathrm{I}}(2)f(x)/\sqrt{x}$ decreases as the design point x moves away from the boundary. The shrinking variance coefficient reflects that more data points can be pooled to smooth in areas with fewer observations. This property is particularly advantageous to estimating the distributions that have a long tail with sparse data, such as those of the economic and financial variables mentioned at the beginning of Section 1.

Mean-squared error (MSE). For interior x, the MSE of $\hat{f}_{\mathrm{GG}}(x)$ can be approximated by

$$\mathrm{MSE}\{\hat{f}_{\mathrm{GG}}(x)\} \sim b^2\left(\frac{C_5^2}{4}\right)\{xf''(x)\}^2 + \frac{V_{\mathrm{I}}(2)}{nb^{1/2}}\frac{f(x)}{\sqrt{x}}. \tag{4}$$

The smoothing parameter value that minimises the right-hand side of Equation (4) is

$$b^*_{\mathrm{GG}} = \left\{\frac{V_{\mathrm{I}}(2)}{C_5^2}\right\}^{2/5}\left[\frac{f(x)/\sqrt{x}}{\{xf''(x)\}^2}\right]^{2/5} n^{-2/5}. \tag{5}$$

Observe that the MSE-optimal smoothing parameter is $O(n^{-2/5}) = O(h^{*2})$, where h^* is the MSE-optimal bandwidth for density estimators using nonnegative symmetric kernels. Therefore, when best implemented, the approximation to the MSE becomes

$$\mathrm{MSE}^*\{\hat{f}_{GG}(x)\} \sim \tfrac{5}{4}[C_5\{V_{\mathrm{I}}(2)\}^2]^{2/5}[f''(x)\{f(x)\}^2]^{2/5}n^{-4/5}. \quad (6)$$

Observe that $\mathrm{MSE}^*\{\hat{f}_{GG}(x)\}$ depends only on $f(x)$ and not on x itself. The optimal MSE of $\hat{f}_{GG}(x)$ for interior x becomes $O(n^{-4/5})$, which is also the optimal convergence rate in the MSE of nonnegative symmetric kernel density estimators. On the other hand, for boundary x, $\mathrm{MSE}\{\hat{f}_{GG}(x)\} = O(b^2 + n^{-1}b^{-1})$, which yields the MSE-optimal smoothing parameter $b_{GG}^{\dagger} = O(n^{-1/3})$ and the optimal MSE of $O(n^{-2/3})$.

2.2.2. *Global property*

The inferior rate in the optimal MSE of $\hat{f}_{GG}(x)$ for boundary x does not affect its global property. If $\int_0^\infty \{xf''(x)\}^2\,dx$ and $\int_0^\infty \{f(x)/\sqrt{x}\}\,dx$ are both finite, then applying the trimming argument in Chen (2000, p. 476) approximates the MISE of $\hat{f}_{GG}(x)$ as

$$\mathrm{MISE}\{\hat{f}_{GG}(x)\} \sim b^2\left(\frac{C_5^2}{4}\right)\int_0^\infty \{xf''(x)\}^2\,dx + \frac{V_{\mathrm{I}}(2)}{nb^{1/2}}\int_0^\infty \frac{f(x)}{\sqrt{x}}\,dx. \quad (7)$$

The smoothing parameter value that minimises the right-hand side of Equation (7) is

$$b_{GG}^{**} = \left\{\frac{V_{\mathrm{I}}(2)}{C_5^2}\right\}^{2/5}\left[\frac{\int_0^\infty \{f(x)/\sqrt{x}\}\,dx}{\int_0^\infty \{xf''(x)\}^2\,dx}\right]^{2/5} n^{-2/5}. \quad (8)$$

Therefore, when best implemented, the approximation to the MISE becomes

$$\mathrm{MISE}^{**}\{\hat{f}_{GG}(x)\} \sim \frac{5}{4}[C_5\{V_{\mathrm{I}}(2)\}^2]^{2/5}\left[\int_0^\infty \{xf''(x)\}^2\,dx\left\{\int_0^\infty \frac{f(x)}{\sqrt{x}}\,dx\right\}^4\right]^{1/5} n^{-4/5}.$$

Note that $O(n^{-4/5})$ is the optimal convergence rate of the MISE within the class of nonnegative kernel estimators in Stone's (1980) sense.

2.2.3. *A note on implementation*

Choosing the smoothing parameter b is an important practical issue. Below we consider a very simple, Silverman's (1986) rule-of-thumb type choice rule. Although in principle the rule is built on the MISE (7) as the criterion, we make a couple of modifications. First, the unknown f is replaced by a known reference density. For simplicity, we choose the pdf of $G(\mu,\omega)$ $(\mu,\omega > 0)$, that is, $g(x) = x^{\mu-1}\exp(-x/\omega)\mathbf{1}\{x \geq 0\}/\{\omega^\mu\Gamma(\mu)\}$, as the reference. Second, the criterion is modified to the asymptotic weighted mean integrated squared error (AWMISE)

$$\mathrm{AWMISE}\{\hat{f}_{GG}(x)\} := b^2\left(\frac{C_5^2}{4}\right)\int_0^\infty \{xg''(x)\}^2 w(x)\,dx + \frac{V_{\mathrm{I}}(2)}{nb^{1/2}}\int_0^\infty \frac{g(x)}{\sqrt{x}}w(x)\,dx,$$

where the weighting function $w(x) \geq 0$ must be chosen to ensure finiteness of two integrals. Given the specification of $g(x)$, it turns out that $w(x) = x^3$ fulfills this requirement. Then, the

AWMISE is simplified to

$$\text{AWMISE}\{\hat{f}_{\mathrm{GG}}(x)\} = b^2 \left\{ \frac{C_5^2 C_\mu \Gamma(2\mu)}{4^\mu \Gamma^2(\mu)} \right\} + \frac{1}{nb^{1/2}} \left\{ \frac{V_{\mathrm{I}}(2)\omega^{5/2}\Gamma(\mu+5/2)}{\Gamma(\mu)} \right\},$$

where

$$\begin{aligned} C_\mu &= \tfrac{1}{4}(\mu-2)^2(\mu-1)^2 - (\mu-2)(\mu-1)^2(\mu) \\ &\quad + \tfrac{1}{2}(3\mu-4)(\mu-1)(\mu)\left(\mu+\tfrac{1}{2}\right) - (\mu-1)(\mu)\left(\mu+\tfrac{1}{2}\right)(\mu+1) \\ &\quad + \tfrac{1}{4}(\mu)\left(\mu+\tfrac{1}{2}\right)(\mu+1)\left(\mu+\tfrac{3}{2}\right). \end{aligned}$$

As a consequence, the AWMISE-optimal smoothing parameter is given by

$$b_{\mathrm{GG}}^{\ddagger\ddagger} = \left\{ \frac{4^{\mu-1} V_{\mathrm{I}}(2)\omega^{5/2}\Gamma(\mu)\Gamma(\mu+5/2)}{C_5^2 C_\mu \Gamma(2\mu)} \right\}^{2/5} n^{-2/5}.$$

In practice, parameters (μ, ω) need to be replaced by their method-of-moments estimates $(\hat{\mu}, \hat{\omega})$.[2]

It would be possible to choose b in a more sophisticated manner. For instance, replacing two integrals in Equation (7) with their nonparametric estimators using the GG kernels, we could derive an analog to the solve-the-equation plug-in method by Sheather and Jones (1991). Alternatively, we may rely on the cross-validation method (e.g. Bouezmarni and Rombouts 2010, p. 250). Investigating these methods is left for our future research.

2.3. *Examples of the GG kernels*

This section introduces three special cases of the GG kernels. The sole reason why the three kernels are listed as examples is that for each of these kernels, approximations to the gamma functions that appear in $M_b(x)$, $H_b(x)$ and $A_{b,\nu}(x)$ in Conditions 3–5 are readily available; see the proof of Theorem 2 in the Appendix for details.

It is worth emphasising that there could be many other examples belonging to this family. As Stacy (1962, p. 1187) states, for instance, functions of a standard normal variate (e.g. its positive even powers, its modulus, and all positive powers of its modulus) will generate the GG kernels. Moreover, for a given pdf, choices of the functional forms of (α, β, γ) may not be unique, as long as they satisfy Conditions 1–5; in other words, we could even generate two or more kernels from the same pdf by making changes in (α, β, γ).

2.3.1. *Examples*

MG kernel. The MG kernel in Chen (2000) turns out to be an immediate example of the GG kernels. Put

$$(\alpha, \beta) = \begin{cases} \left(\dfrac{x}{b}, x\right) & \text{for } x \geq 2b \\ \left(\dfrac{1}{4}\left(\dfrac{x}{b}\right)^2 + 1, \dfrac{x^2}{4b} + b\right) & \text{for } x \in [0, 2b) \end{cases}$$

and $\gamma = 1$ in Equation (3). Then, it collapses to

$$K_{\mathrm{GG}}(u;x,b) = \frac{u^{\alpha-1}\exp\{-u/(\beta/\alpha)\}}{(\beta/\alpha)^{\alpha}\Gamma(\alpha)}\mathbf{1}\{u \geq 0\}, \tag{9}$$

which is the pdf of the gamma distribution $G(\alpha, \beta/\alpha)$. Observe that $\alpha = \rho_b(x)$ in Chen (2000, p. 473) and $\beta/\alpha = b$. It follows that the above kernel finally reduces to the MG kernel

$$K_{\mathrm{MG}(\rho_b(x),b)}(u) = \frac{u^{\rho_b(x)-1}\exp(-u/b)}{b^{\rho_b(x)}\Gamma\{\rho_b(x)\}}\mathbf{1}\{u \geq 0\}.$$

Weibull kernel. To obtain the Weibull (W) kernel, let

$$(\alpha, \beta) = \begin{cases} \left(\sqrt{\dfrac{2x}{b}}, x\right) & \text{for } x \geq 2b \\ \left(\dfrac{1}{2}\left(\dfrac{x}{b}\right) + 1, \dfrac{x^2}{4b} + b\right) & \text{for } x \in [0, 2b) \end{cases}$$

and $\gamma = \alpha$ in Equation (3). Then, it becomes

$$K_{\mathrm{GG}}(u;x,b) = \frac{\alpha u^{\alpha-1}\exp[-\{u/(\beta/\Gamma(1+1/\alpha))\}^{\alpha}]}{\{\beta/\Gamma(1+1/\alpha)\}^{\alpha}}\mathbf{1}\{u \geq 0\}.$$

Because the right-hand side is the pdf of the Weibull distribution $\mathrm{W}(\alpha, \beta/\Gamma(1+1/\alpha))$, the W kernel can be defined as[3]

$$K_{\mathrm{W}(\alpha,\beta/\Gamma(1+1/\alpha))}(u) = \frac{\alpha}{\beta/\Gamma(1+1/\alpha)}\left\{\frac{u}{\beta/\Gamma(1+1/\alpha)}\right\}^{\alpha-1}\exp\left[-\left\{\frac{u}{\beta/\Gamma(1+1/\alpha)}\right\}^{\alpha}\right]\mathbf{1}\{u \geq 0\}.$$

Nakagami-m kernel. Use exactly the same (α, β) as for the MG kernel but put $\gamma = 2$ in Equation (3). Then, it reduces to

$$K_{\mathrm{GG}}(u;x,b) = \frac{2u^{\alpha-1}\exp[-\{u/(\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2))\}^2]}{\{\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2)\}^{\alpha}\Gamma(\alpha/2)}\mathbf{1}\{u \geq 0\},$$

which is the pdf of the Nakagami-m distribution $\mathrm{NM}(\alpha/2, (\alpha/2)[\beta\Gamma(\alpha/2)/\Gamma\{(\alpha+1)/2\}]^2)$ due to Nakagami (1943, 1960).[4] This distribution is frequently applied in telecommunications engineering as the distribution that can describe signal intensity of short-wave fading. In the end, the Nakagami-m (NM) kernel is defined as

$$\begin{aligned} K_{\mathrm{NM}(\alpha/2,(\alpha/2)[\beta\Gamma(\alpha/2)/\Gamma\{(\alpha+1)/2\}]^2)}(u) &= \frac{2(\alpha/2)^{\alpha/2}}{[(\alpha/2)\{\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2)\}^2]^{\alpha/2}\Gamma(\alpha/2)}u^{2(\alpha/2)-1} \\ &\quad \times \exp\left[-\frac{\alpha/2}{(\alpha/2)\{\beta\Gamma(\alpha/2)/\Gamma((\alpha+1)/2)\}^2}u^2\right]\mathbf{1}\{u \geq 0\}. \end{aligned}$$

2.3.2. *Asymptotic results on density estimators using the MG, W and NM kernels*

Figure 1 plots the shapes of the MG, W and NM kernels at four different design points ($x = 0, 1, 2, 4$) at which the smoothing is performed. For reference, the gamma (G) kernel (Chen 2000) is also drawn in each panel.[5] It is worth noting that for all plotted functions, the value of the

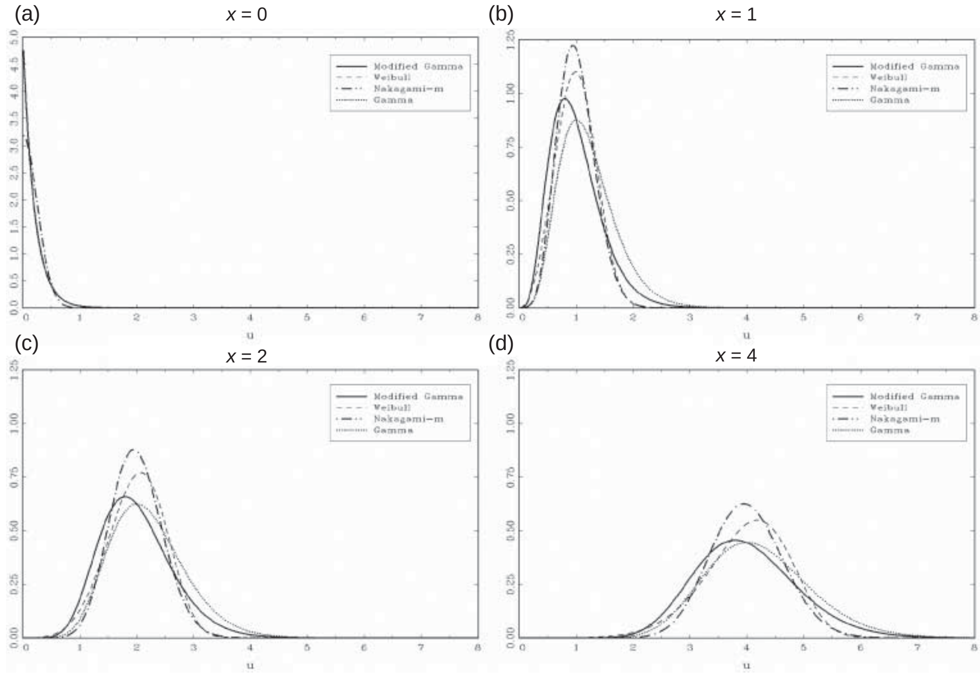

Figure 1. Shapes of the GG kernels when $b = 0.2$.

smoothing parameter is fixed at $b = 0.2$. When smoothing is made at the origin (Panel (a)), the NM kernel collapses to a half-normal pdf, whereas all others reduce to an exponential pdf. As the design point moves away from the boundary (Panels (b)–(d)), the shape of each kernel varies and becomes flatter; in other words, each kernel changes the amount of smoothing in a locally adaptive manner. It is worth emphasising that unlike variable bandwidth methods (e.g. Abramson 1982), adaptive smoothing for these kernels is achieved by a single smoothing parameter, which makes them much more appealing in empirical work. We can also observe that the more distant the design point is from the boundary, the closer the shapes become to a symmetric one; in particular, shapes of the NM kernel become almost symmetric for a large x, as its functional form suggests.

Convergence properties of density estimators using the MG, W and NM kernels are presented in Theorem 2 below. Obviously, the functional form of (α, β, γ) for each of the three kernels satisfies Conditions 1–2. Hence, to demonstrate Theorem 2, it suffices to check that Conditions 3–5 hold for each kernel.

THEOREM 2 *Let $\hat{f}_j(x)$ be the probability density estimator using the kernel $j \in \{MG, W, NM\}$. Then, under Assumptions 1–3, the bias and variance of $\hat{f}_j(x)$ can be approximated by* $\mathrm{Bias}\{\hat{f}_j(x)\} \sim B_{1,j}(x, f)b$ *and*

$$\mathrm{Var}\{\hat{f}_j(x)\} \sim \begin{cases} (nb^{1/2})^{-1} V_{\mathrm{I},j}(2) f(x)/\sqrt{x} & \text{for interior } x, \\ (nb)^{-1} V_{\mathrm{B},j}(2) f(x) & \text{for boundary } x, \end{cases}$$

where $V_{\mathrm{B},j}(2)$ can be found in the Appendix, and values of $C_{5,j}$ and explicit forms of $B_{1,j}(x, f)$ and $V_{\mathrm{I},j}(2)$ are as follows:

		$B_{1,j}(x,f)$		
j	$C_{5,j}$	$x \geq 2b$	$x \in [0,2b)$	$V_{I,j}(2)$
MG	1	$(1/2)xf''(x)$	$\xi_b(x)f'(x)$	$1/(2\sqrt{\pi})$
W	$\pi^2/12$	$(\pi^2/24)xf''(x)$	$\xi_b(x)f'(x)$	$1/(2\sqrt{2})$
NM	$1/2$	$(1/4)xf''(x)$	$\xi_b(x)f'(x)$	$1/\sqrt{2\pi}$

Moreover, $\xi_b(x) = \{(1/2)(x/b) - 1\}^2 = O(1)$ *in this case.*

It follows from Equation (5) and Theorem 2 that the MSE-optimal smoothing parameters of three kernel density estimators for interior x are

$$b^*_{\mathrm{MG}} = \left(\frac{1}{2\sqrt{\pi}}\right)^{2/5}\left[\frac{f(x)/\sqrt{x}}{\{xf''(x)\}^2}\right]^{2/5} n^{-2/5},$$

$$b^*_{\mathrm{W}} = \left\{\frac{(12/\pi^2)^2}{2^{3/2}}\right\}^{2/5}\left[\frac{f(x)/\sqrt{x}}{\{xf''(x)\}^2}\right]^{2/5} n^{-2/5},$$

$$b^*_{\mathrm{NM}} = \left(\frac{2^{3/2}}{\sqrt{\pi}}\right)^{2/5}\left[\frac{f(x)/\sqrt{x}}{\{xf''(x)\}^2}\right]^{2/5} n^{-2/5}.$$

Recognise that $b^*_{\mathrm{NM}} = 2b^*_{\mathrm{MG}}$ holds; even $b^{**}_{\mathrm{NM}} = 2b^{**}_{\mathrm{MG}}$ is the case by Equation (8), where b^{**}_{NM} and b^{**}_{MG} are MISE-optimal smoothing parameters of the NM and MG kernels, respectively. Then, by Equation (6), the optimal-MSEs of $\hat{f}_{\mathrm{MG}}(x)$ and $\hat{f}_{\mathrm{NM}}(x)$ for interior x have the relation

$$\begin{aligned}\mathrm{MSE}^*\{\hat{f}_{\mathrm{MG}}(x)\} &\sim \mathrm{MSE}^*\{\hat{f}_{\mathrm{NM}}(x)\} \\ &\sim \frac{5}{4}\left(\frac{1}{4\pi}\right)^{2/5}[f''(x)\{f(x)\}^2]^{2/5} n^{-4/5}. \end{aligned} \tag{10}$$

The right-hand side is also the optimal-MSE of the density estimator using the Gaussian kernel. In other words, when best implemented, density estimators using these kernels become first-order asymptotically equivalent, and both kernels on $[0,\infty)$ are in a sense equivalent to the Gaussian kernel on $(-\infty,\infty)$. In contrast, when best implemented, the MSE of $\hat{f}_{\mathrm{W}}(x)$ for interior x can be approximated by

$$\mathrm{MSE}^*\{\hat{f}_{\mathrm{W}}(x)\} \sim \frac{5}{4}\left(\frac{\pi^2}{96}\right)^{2/5}[f''(x)\{f(x)\}^2]^{2/5} n^{-4/5}. \tag{11}$$

Comparing the factors of Equation (10) and (11) reveal that $(5/4)\{1/(4\pi)\}^{2/5} \approx 0.454178\ldots$ and $(5/4)(\pi^2/96)^{2/5} \approx 0.503178\ldots$ Therefore, we can see that $\hat{f}_{\mathrm{W}}(x)$ is slightly inefficient than $\hat{f}_{\mathrm{MG}}(x)$ and $\hat{f}_{\mathrm{NM}}(x)$ under the best-case scenario.

3. Properties of the GG density estimator

This section explores three properties of the GG density estimator other than those described in Section 2.2. After demonstrating applicability of MBC techniques for independent observations, we show the remaining two properties for weakly dependent observations, namely, validity of the first-order bias and variance approximations in GG density estimation that are stated in Theorem 1, and weak consistency of GG density estimators when the true density is unbounded at the origin. It is known that the G and MG kernels possess all these properties. Therefore, this section establishes that these appealing properties essentially inhere in the GG kernels.

3.1. *Nonparametric MBC for GG density estimation*

We start with examining whether GG density estimators in general admit MBC techniques. While Hagmann and Scaillet (2007) and Gustafsson, Hagmann, Nielsen, and Scaillet (2009) propose semiparametric MBC procedures using asymmetric kernels, we concentrate on two classes of nonparametric MBC methods studied in Hirukawa (2010) and Hirukawa and Sakudo (2014). The first class of MBC estimators, originally proposed by Terrell and Scott (1980), is defined as

$$\tilde{f}_{\mathrm{TS,GG}}(x) = \{\hat{f}_{\mathrm{GG},b}(x)\}^{1/(1-c)}\{\hat{f}_{\mathrm{GG},b/c}(x)\}^{-c/(1-c)},$$

where $\hat{f}_{\mathrm{GG},b}(x)$ and $\hat{f}_{\mathrm{GG},b/c}(x)$ signify GG density estimators using smoothing parameters b and b/c, and $c \in (0,1)$ is some predetermined constant that does not depend on the design point x. The second class of MBC estimators due to Jones, Linton, and Nielsen (1995) is implied by the identity $f(x) = \hat{f}_{\mathrm{GG},b}(x)\{f(x)/\hat{f}_{\mathrm{GG},b}(x)\}$ and defined as

$$\tilde{f}_{\mathrm{JLN,GG}}(x) = \hat{f}_{\mathrm{GG},b}(x)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{K_{\mathrm{GG}}(X_i;x,b)}{\hat{f}_{\mathrm{GG},b}(X_i)}\right\}.$$

Both $\tilde{f}_{\mathrm{TS,GG}}(x)$ and $\tilde{f}_{\mathrm{JLN,GG}}(x)$ always generate nonnegative density estimates everywhere by construction.

To develop convergence properties of MBC estimators, we modify Assumptions 2 and 3 as follows. Discussions on these assumptions can be found in Hirukawa (2010) and Hirukawa and Sakudo (2014).

Assumption 2a *f has four continuous and bounded derivatives, and $f(x) > 0$ for a given design point $x > 0$.*

Assumption 3a *The smoothing parameter b satisfies $b + (nb^3)^{-1} \to 0$ as $n \to \infty$.*

The next theorem refers to acceleration in bias convergence via the bias correction methods. The proof is similar to the ones for Theorems 1 and 2 of Hirukawa and Sakudo (2014), and this it is omitted.

Theorem 3 *If Assumptions 1, 2a and 3a hold, and $E\{\hat{f}_{\mathrm{GG}}(x)\}$ admits the expansion $E\{\hat{f}_{\mathrm{GG}}(x)\} = f(x) + B_1(x,f)b + B_2(x,f)b^2 + o(b^2)$, where $B_1(x,f)$ (which can be found in Theorem 1) and $B_2(x,f)$ are kernel-specific functions depending on x and derivatives of f, then the biases of $\tilde{f}_{\mathrm{TS,GG}}(x)$ and $\tilde{f}_{\mathrm{JLN,GG}}(x)$ can be approximated by*

$$\mathrm{Bias}\{\tilde{f}_{\mathrm{TS,GG}}(x)\} \sim \frac{1}{c}p(x)b^2 := \frac{1}{c}\left[\frac{1}{2}\left\{\frac{B_1^2(x,f)}{f(x)}\right\} - B_2(x,f)\right]b^2,$$
$$\mathrm{Bias}\{\tilde{f}_{\mathrm{JLN,GG}}(x)\} \sim q(x)b^2 := -f(x)B_1(x,h)b^2,$$

where $B_1(x,h)$ is obtained by replacing f in $B_1(x,f)$ with $h = h(x,f) := B_1(x,f)/f(x)$.[6] *In addition, the variance of each estimator is $O\{(nb^{1/2})^{-1}\}$ for interior x and $O\{(nb)^{-1}\}$ for boundary x.*

The main point of the theorem is that under sufficient differentiability of f, the bias convergence of each MBC estimator is accelerated from $O(b)$ to $O(b^2)$ with the order of magnitude in variance left unchanged. Whether the bias convergence may speed up depends crucially on whether the second-order term in $\mathrm{Bias}\{\hat{f}_{GG}(x)\}$ is $O(b^2)$. It is worth noting that Conditions 1–5 provide no guidance on the order of magnitude in the second-order bias term. For instance, as indicated in the proof of Theorem 2, the second-order term in $\mathrm{Bias}\{\hat{f}_{W}(x)\}$ is $O(b^{3/2})$. Because both MBC techniques merely improve the bias convergence up to $O(b^{3/2})$, the theorem excludes such inferior cases.[7] In contrast, for the MG and NM kernels, each MBC technique ameliorates their bias convergence to $O(b^2)$. Explicit forms of $p(x)$ and $q(x)$ for the NM kernel are

$$p_{NM}(x) = \begin{cases} \frac{x^2}{32}\frac{\{f''(x)\}^2}{f(x)} - \frac{1}{8}\left\{\frac{1}{2}f'(x) + \frac{x}{3}f''(x) + \frac{x^2}{4}f'''(x)\right\} & \text{for } x \geq 2b, \\ \frac{1}{2}\left[\frac{\{\xi_b(x)f'(x)\}^2}{f(x)} - \left\{\left(\xi_b(x) + \frac{x}{b}\right)^2 \frac{\Gamma\left(\frac{\xi_b(x)+x/b}{2}\right)\Gamma\left(\frac{\xi_b(x)+x/b}{2}+1\right)}{\left(\Gamma\left(\frac{\xi_b(x)+x/b+1}{2}\right)\right)^2} \right.\right. \\ \qquad \left.\left. - 2\left(\frac{x}{b}\right)\xi_b(x) + \left(\frac{x}{b}\right)^2\right\}\right] & \text{for } x \in [0, 2b), \end{cases}$$

$$q_{NM}(x) = \begin{cases} -f(x)\frac{x}{4}\left\{\frac{x}{4}\frac{f''(x)}{f(x)}\right\}'' & \text{for } x \geq 2b, \\ -f(x)\xi_b(x)\left\{\frac{\xi_b(x)f'(x)}{f(x)}\right\}'' & \text{for } x \in [0, 2b), \end{cases}$$

where $\xi_b(x)$ can be found in Theorem 2; see Hirukawa and Sakudo (2014) for functional forms of $p(x)$ and $q(x)$ for the MG kernel.

To provide the variance approximation of each MBC estimator, we must additionally specify the functional form of $(\alpha, \beta, \gamma) = (\alpha_b(x), \beta_b(x), \gamma_b(x))$. For example, for each of the MG and NM kernels, $\mathrm{Var}\{\tilde{f}_{TS,GG}(x)\} \sim (nb^{1/2})^{-1}\lambda(c)V_I(2) f(x)/\sqrt{x}$ and $\mathrm{Var}\{\tilde{f}_{JLN,GG}(x)\} \sim (nb^{1/2})^{-1}V_I(2)f(x)/\sqrt{x}$ for interior x, where

$$\lambda(c) = \frac{(1+c^{5/2})(1+c)^{1/2} - 2\sqrt{2}c^{3/2}}{(1+c)^{1/2}(1-c)^2}$$

is increasing in $c \in (0,1)$ with $\lim_{c\to 0}\lambda(c) = 1$ and $\lim_{c\to 1}\lambda(c) = 27/16$. In other words, $\mathrm{Var}\{\tilde{f}_{TS,MG}(x)\}$ and $\mathrm{Var}\{\tilde{f}_{TS,NM}(x)\}$ at interior design points tend to inflate slightly after the bias correction, whereas $\mathrm{Var}\{\tilde{f}_{JLN,MG}(x)\}$ and $\mathrm{Var}\{\tilde{f}_{JLN,NM}(x)\}$ are first-order asymptotically equivalent to $\mathrm{Var}\{\hat{f}_{MG}(x)\}$ and $\mathrm{Var}\{\hat{f}_{NM}(x)\}$, respectively.

As a consequence, we can find an interesting relation between $\tilde{f}_{JLN,MG}(x)$ and $\tilde{f}_{JLN,NM}(x)$ once again. The MSEs of these estimators for interior x can be approximated by

$$\mathrm{MSE}\{\tilde{f}_{JLN,MG}(x)\} \sim \left[-f(x)\frac{x}{2}\left\{\frac{x}{2}\frac{f''(x)}{f(x)}\right\}''\right]^2 b^4 + \frac{1}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}\sqrt{x}},$$

$$\begin{aligned} \mathrm{MSE}\{\tilde{f}_{JLN,NM}(x)\} &\sim \left[-f(x)\frac{x}{4}\left\{\frac{x}{4}\frac{f''(x)}{f(x)}\right\}''\right]^2 b^4 + \frac{1}{nb^{1/2}}\frac{f(x)}{\sqrt{2\pi}\sqrt{x}} \\ &= \frac{1}{16}\left[-f(x)\frac{x}{2}\left\{\frac{x}{2}\frac{f''(x)}{f(x)}\right\}''\right]^2 b^4 + \frac{\sqrt{2}}{nb^{1/2}}\frac{f(x)}{2\sqrt{\pi}\sqrt{x}}. \end{aligned}$$

The smoothing parameter values that minimise these approximations are

$$b^*_{\mathrm{JLN,MG}} = \frac{1}{2^{2/3}}\left\{\frac{f(x)}{2\sqrt{\pi}\sqrt{x}}\right\}\left[f(x)\frac{x}{2}\left\{\frac{x}{2}\frac{f''(x)}{f(x)}\right\}''\right]^{-4/9} n^{-2/9},$$

$$b^*_{\mathrm{JLN,NM}} = 2^{1/3}\left\{\frac{f(x)}{2\sqrt{\pi}\sqrt{x}}\right\}\left[f(x)\frac{x}{2}\left\{\frac{x}{2}\frac{f''(x)}{f(x)}\right\}''\right]^{-4/9} n^{-2/9}.$$

Observe that the relation $b^*_{\mathrm{JLN,NM}} = 2b^*_{\mathrm{JLN,MG}}$ holds once again. Moreover, when best implemented, the optimal-MSEs of two estimators for interior x are first-order asymptotically equivalent, that is,

$$\mathrm{MSE}^*\left\{\tilde{f}_{\mathrm{JLN,MG}}(x)\right\} \sim \mathrm{MSE}^*\{\tilde{f}_{\mathrm{JLN,NM}}(x)\}$$

$$\sim \frac{9}{2^{8/3}}\left\{\frac{f(x)}{2\sqrt{\pi}\sqrt{x}}\right\}^{8/9}\left[f(x)\frac{x}{2}\left\{\frac{x}{2}\frac{f''(x)}{f(x)}\right\}''\right]^{2/9} n^{-8/9}.$$

3.2. *Bias and variance approximations of GG density estimators using weakly dependent observations*

The second property is concerned with estimating the marginal density from nonnegative time-series data. Examples include estimation of the distribution of important financial variables such as short-term interest rates or trading volumes, and even the baseline hazard in financial duration analysis.

To allow for weakly dependent observations in GG density estimation, we replace Assumptions 1 and 2 with the regularity conditions below. While similar conditions can be found, for instance, in Bouezmarni and Rombouts (2010), they assume strong mixing processes with an exponentially decaying mixing coefficient and exclusively study the MG kernel.[8] We relax the mixing condition and prove that their result holds for a broader class of asymmetric kernels which includes the MG kernel.

Assumption 1b *$\{X_i\}$ is a nonnegative, strictly stationary and strong mixing process with the mixing coefficient $\alpha(\ell)$ of size $-(2r-2)/(r-2)$ for some $r > 2$.*

Assumption 2b *Let $f(\cdot)$ and $f_j(\cdot,\cdot)$ be the marginal and joint densities of X_i and (X_i, X_{i+j}), respectively. Then, f is twice continuously differentiable, and f_j is uniformly bounded.*

The next theorem states that results in Theorem 1 are carried through even when positive weakly dependent observations are used.

Theorem 4 *Results in Theorem 1 still hold under Assumptions 1b, 2b and 3.*

3.3. *Convergence properties of the GG density estimator for the density unbounded at the origin*

As the third property we obtain convergence results of $\hat{f}_{\mathrm{GG}}(x)$ when the true density $f(x)$ is unbounded at $x = 0$. Often a clustering of observations near the boundary can be found in intra-day trading volumes or realised volatilities, for instance. In this case, it is highly likely that the true density has a pole at $x = 0$ (e.g. Malec and Schienle 2014). Such shapes also appear in many

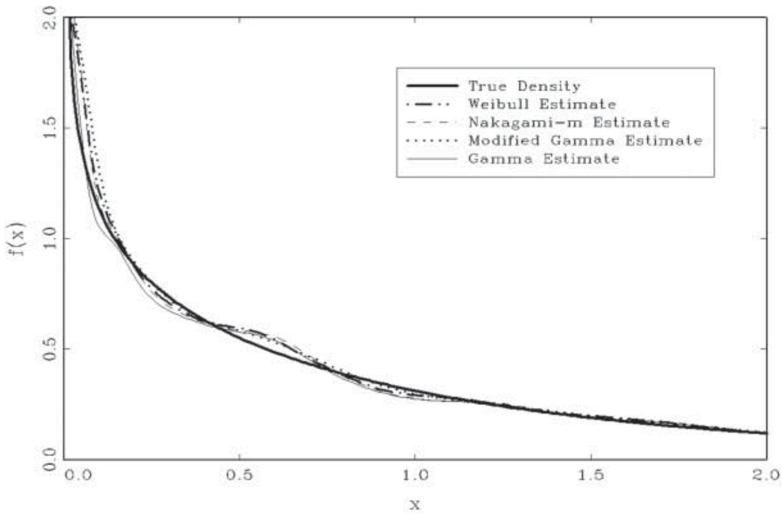

Figure 2. GG density estimates when the true distribution is $G(0.75, 1.25)$.

other applications including spectral densities of long memory processes. The following two theorems demonstrate weak consistency and the relative convergence of $\hat{f}_{GG}(x)$ as $x \to 0$, provided that the assumption on the smoothing parameter is replaced by Assumption 3c.

ASSUMPTION 3c *The smoothing parameter b satisfies $b + (nb^2)^{-1} \to 0$ as $n \to \infty$.*

THEOREM 5 *If the true density $f(x)$ is unbounded at $x = 0$ and Assumptions 1b and 3c hold, then $\hat{f}_{GG}(0) \xrightarrow{p} \infty$.*

THEOREM 6 *Suppose that the true density $f(x)$ is unbounded at $x = 0$ and continuously differentiable in the neighbourhood of the origin. If Assumptions 1b and 3c also hold, then*

$$\left| \frac{\hat{f}_{GG}(x) - f(x)}{f(x)} \right| \xrightarrow{p} 0$$

as $x \to 0$.

Bouezmarni and Scaillet (2005, p. 399) focus exclusively on random samples and find that the property in Theorem 5 is peculiar to the G and MG kernels and not shared with other asymmetric kernels such as the inverse Gaussian and reciprocal inverse Gaussian kernels by Scaillet (2004). To illustrate the result from this theorem, we prepare Figure 2, in which the W, NM, MG, and G density estimates based on 500 independent observations drawn from $G(0.75, 1.25)$ are plotted. The figure indicates that estimates using three examples of the GG kernels, as well as the G estimate, can indeed capture the shape of the density near the boundary reasonably well.

4. Finite sample performance

4.1. *Set-up*

This simulation study examines accuracy of GG density estimators. We compare finite sample performance of three GG estimators using the W, NM and MG kernels with that of three other density estimators, namely, (i) the density estimator using yet another asymmetric kernel, (ii) the one obtained via back-transforming a symmetric kernel density estimate based on the log-transformed (LT) data, and (iii) a symmetric kernel density estimator in the original scale with a proper boundary correction. For (i), because of its popularity in empirical studies, the G kernel is chosen as the asymmetric kernel that does not belong to the GG kernels. For (ii) and (iii), the LT density estimator using the Gaussian kernel and the local linear (LL) density estimator using the Epanechnikov kernel (e.g. Jones 1993) are chosen, respectively.

For each distribution in the list below, 1000 data sets of sample size $n = 100$, 200, or 500 are simulated. All density estimates are evaluated on an equally spaced grid of 500 points over the interval $[0, 5]$. Following Chen (2000) and Scaillet (2004), as the performance measure for each estimator $\bar{f}$, we compute the root integrated squared error (RISE) $\mathrm{RISE}\{\bar{f}(x)\} = \sqrt{\int_0^\infty \{\bar{f}(x) - f(x)\}^2 \, \mathrm{d}x}$. In our report, the integral is approximated over the 500 points. The smoothing parameter b (for W, NM, MG, and G) or the bandwidth h (for LT and LL) is chosen as the minimiser of the (approximated) RISE, again as in Chen (2000) and Scaillet (2004). Lastly, six distributions below are considered as truths, because of their popularity in parametric modelling of income distributions, actuarial loss distributions and baseline hazards. Distribution 5, proposed by Buch-Larsen, Nielsen, Guillén, and Bolancé (2005) for modelling distributions of insurance payments and operational risks, is known to have a Pareto-type tail. This distribution is adopted in this simulation study, from the viewpoint that it is thought to mimic the shape of income distributions well. Distribution 6, which is unbounded at the origin, is prepared to investigate the property stated in Theorem 5. Figure 3 presents the shapes of these densities.

(1) Gamma: $f(x) = x^{\alpha-1}\exp(-x/\beta)\mathbf{1}\{x \geq 0\}/\{\beta^\alpha \Gamma(\alpha)\}$, $(\alpha, \beta) = (1.5, 1)$.

(2) Weibull: $f(x) = (\alpha/\beta)(x/\beta)^{\alpha-1}\exp\{-(x/\beta)^\alpha\}\mathbf{1}\{x \geq 0\}$, $(\alpha, \beta) = (1.5, 1.5)$.

(3) Half-Normal: $f(x) = \{2/(\sqrt{2\pi}\sigma)\}\exp\{-(x-\mu)^2/(2\sigma^2)\}\mathbf{1}\{x \geq 0\}$, $(\mu, \sigma) = (0, 1.5)$.

(4) Log-Normal: $f(x) = \{1/(x\sqrt{2\pi}\sigma)\}\exp\{-(\log x - \mu)^2/(2\sigma^2)\}\mathbf{1}\{x \geq 0\}$, $(\mu, \sigma) = (0, 0.75)$.

(5) Generalised Champernowne: $f(x) = \alpha(x+c)^{\alpha-1}\{(M+c)^\alpha - c^\alpha\}\mathbf{1}\{x \geq 0\} \times \{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha\}^{-2}$, $(\alpha, M, c) = (2.5, 0.5, 0.25)$.

(6) Gamma with Pole: $f(x) = x^{\alpha-1}\exp(-x/\beta)\mathbf{1}\{x \geq 0\}/\{\beta^\alpha \Gamma(\alpha)\}$, $(\alpha, \beta) = (0.75, 1.25)$.

4.2. *Simulation results*

Table 1 presents averages and standard deviations of RISEs and averages of tuning parameter values (i.e. smoothing parameter values for W, NM, MG, and G, and bandwidth lengths for LT and LL) over 1000 Monte Carlo replications. For each distribution, results are qualitatively similar across three sample sizes. A quick inspection reveals that overall the MG estimator performs better than the G estimator; only exceptions can be found in Distribution 1 for $n = 100$ and Distribution 6 for all sample sizes. This is congruous with the results in Chen (2000) and Scaillet (2004).

Each of first three distributions is chosen so that one of the GG estimators can have clear advantage. As expected, W and MG, as well as G, perform well for Distributions 1 and 2, whereas NM dominates for Distribution 3. Poor performance of W, MG and G for Distribution 3 could be

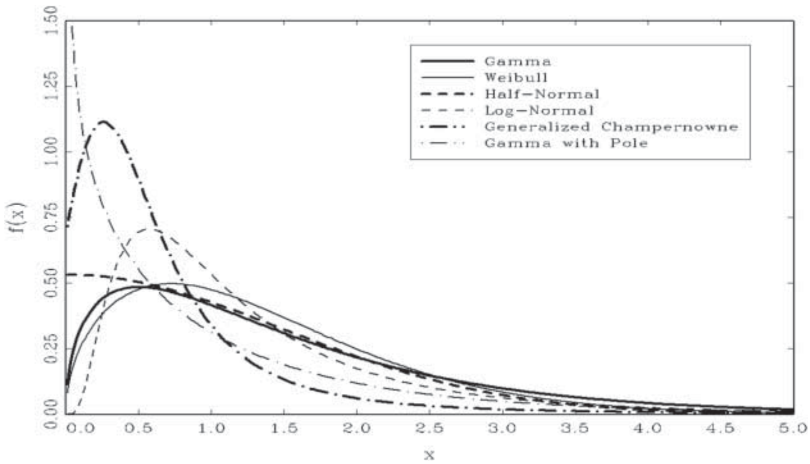

Figure 3. Shapes of true densities for Monte Carlo simulations.

attributed to their difficulty in capturing the shape of the pdf near the origin. The pdf is decreasing and satisfies the shoulder condition $f'(0) = 0$. Accordingly, the data suggest concavity of the density in the vicinity of the origin. As a consequence, these kernels tend to misinterpret the local concavity as an indication of a mode over the strictly positive region. We can also find that LL exhibits comparable performance to the GG estimators, whereas LT is consistently outperformed.

The GG estimators perform satisfactorily even for Distribution 4, which is advantageous to LT, or for Distribution 5, which could be unfavorable to all due to its Pareto-type tail. As predicted by Theorem 5, the GG estimators also exhibit reasonably good performance for Distribution 6. In contrast, LL and LT are dominated for Distributions 4 and 6 and Distribution 5, respectively.

A rationale as to why the LT estimator $\hat{f}_{LT}(x)$ does not necessarily perform well (although it is popularly applied in empirical works) may be found in its bias property. Following Wand, Marron, and Ruppert (1991), we can approximate the bias and variance of the estimator as

$$\mathrm{Bias}\{\hat{f}_{LT}(x)\} \sim \frac{1}{2}\{f(x) + 3xf'(x) + x^2 f''(x)\}h^2,$$

$$\mathrm{Var}\{\hat{f}_{LT}(x)\} \sim \begin{cases} (nh)^{-1} f(x)/(2\sqrt{\pi}x) & \text{if } x/h \to \infty, \\ (nh^2)^{-1} f(x)/(2\sqrt{\pi}\eta) & \text{if } x/h \to \eta > 0. \end{cases}$$

While $\hat{f}_{LT}(x)$ shares the shrinking variance property as the design point moves away from the boundary (which comes from the fact that the variance coefficient is proportional to $f(x)/x$) with asymmetric kernel estimators, its leading bias takes a less favourable form in that it contains three terms involving the true density and its first two derivatives.

We also make a few remarks on tuning parameters b and h. First, both h_{LT} and h_{LL} appear to be systematically long. Large values of the former may be attributed to the fact that LT observations tend to spread out toward the large (in magnitude) negative side for each distribution. Second, for each combination of the sample size and distribution (except Distribution 6), we can find a consistent ordering of b_G, b_{MG}, b_W, and b_{NM} from the smallest to the largest. Third, whether b_{NM} becomes roughly twice the size of b_{MG} as predicted in Section 2.3.2 depends crucially on distributions and sample sizes. The derivation of Equation (7) (and thus Equation (8)) is built on the trimming argument for the boundary region. Therefore, whether finite sample results support the relation relies on how fast the shrinkage of b is and/or how small the weight that the distribution puts near the boundary is. While it is hard to find the relation $b_{NM} \approx 2b_{MG}$ in Table 1,

Table 1. Averages of performance measures and tuning parameter values.

		$n = 100$		$n = 200$		$n = 500$	
		RISE	b or h	RISE	b or h	RISE	b or h
		(1) Gamma					
GG	W	0.0356 (0.0098)	0.2778	0.0294 (0.0081)	0.1701	0.0221 (0.0057)	0.0897
	NM	0.0368 (0.0091)	0.3007	0.0306 (0.0076)	0.1861	0.0232 (0.0055)	0.0975
	MG	0.0362 (0.0112)	0.1712	0.0289 (0.0088)	0.1105	0.0211 (0.0059)	0.0683
Non-GG	G	0.0358 (0.0125)	0.1404	0.0290 (0.0098)	0.0962	0.0220 (0.0066)	0.0601
	LT	0.0441 (0.0157)	0.4434	0.0348 (0.0114)	0.3820	0.0252 (0.0074)	0.3149
	LL	0.0368 (0.0116)	1.0272	0.0302 (0.0088)	0.7524	0.0234 (0.0061)	0.5152
		(2) Weibull					
GG	W	0.0374 (0.0119)	0.1870	0.0297 (0.0090)	0.1228	0.0214 (0.0058)	0.0809
	NM	0.0385 (0.0116)	0.2090	0.0307 (0.0088)	0.1382	0.0222 (0.0058)	0.0911
	MG	0.0367 (0.0127)	0.1272	0.0286 (0.0092)	0.0915	0.0204 (0.0060)	0.0634
Non-GG	G	0.0368 (0.0140)	0.1137	0.0294 (0.0103)	0.0813	0.0218 (0.0069)	0.0526
	LT	0.0470 (0.0154)	0.3730	0.0368 (0.0110)	0.3187	0.0267 (0.0073)	0.2584
	LL	0.0367 (0.0127)	0.8234	0.0294 (0.0092)	0.6600	0.0217 (0.0060)	0.5098
		(3) Half-Normal					
GG	W	0.0274 (0.0113)	0.3445	0.0225 (0.0083)	0.2745	0.0172 (0.0056)	0.1936
	NM	0.0251 (0.0120)	0.3809	0.0207 (0.0088)	0.3152	0.0158 (0.0059)	0.2354
	MG	0.0303 (0.0117)	0.2662	0.0245 (0.0087)	0.2039	0.0184 (0.0060)	0.1380
Non-GG	G	0.0327 (0.0112)	0.1750	0.0262 (0.0087)	0.1325	0.0193 (0.0057)	0.0911
	LT	0.0586 (0.0191)	0.4735	0.0457 (0.0140)	0.3997	0.0329 (0.0085)	0.3224
	LL	0.0256 (0.0119)	1.4774	0.0203 (0.0094)	1.2598	0.0147 (0.0064)	1.0193
		(4) Log-Normal					
GG	W	0.0429 (0.0153)	0.0830	0.0332 (0.0110)	0.0654	0.0242 (0.0080)	0.0471
	NM	0.0447 (0.0152)	0.0932	0.0343 (0.0108)	0.0749	0.0245 (0.0078)	0.0562
	MG	0.0416 (0.0158)	0.0624	0.0324 (0.0114)	0.0480	0.0238 (0.0081)	0.0334
Non-GG	G	0.0458 (0.0150)	0.0535	0.0360 (0.0108)	0.0390	0.0263 (0.0075)	0.0261
	LT	0.0401 (0.0166)	0.3207	0.0315 (0.0122)	0.2782	0.0232 (0.0082)	0.2310
	LL	0.0482 (0.0147)	0.4415	0.0381 (0.0106)	0.3683	0.0282 (0.0073)	0.2937
		(5) Generalised Champernowne					
GG	W	0.0477 (0.0169)	0.1324	0.0391 (0.0126)	0.0922	0.0295 (0.0101)	0.0547
	NM	0.0477 (0.0161)	0.1448	0.0390 (0.0122)	0.1033	0.0294 (0.0099)	0.0637
	MG	0.0504 (0.0190)	0.0881	0.0403 (0.0141)	0.0618	0.0298 (0.0106)	0.0388
Non-GG	G	0.0504 (0.0195)	0.0676	0.0403 (0.0151)	0.0485	0.0301 (0.0107)	0.0318
	LT	0.0700 (0.0246)	0.4451	0.0544 (0.0177)	0.3807	0.0394 (0.0117)	0.3118
	LL	0.0513 (0.0183)	0.4850	0.0413 (0.0137)	0.3618	0.0308 (0.0101)	0.2637
		(6) Gamma with Pole					
GG	W	0.0617 (0.0181)	0.0858	0.0494 (0.0136)	0.0591	0.0359 (0.0091)	0.0380
	NM	0.0652 (0.0168)	0.0855	0.0523 (0.0128)	0.0586	0.0380 (0.0088)	0.0373
	MG	0.0627 (0.0171)	0.0803	0.0500 (0.0131)	0.0554	0.0368 (0.0098)	0.0357
Non-GG	G	0.0614 (0.0202)	0.0571	0.0494 (0.0157)	0.0389	0.0363 (0.0107)	0.0242
	LT	0.0639 (0.0304)	0.7389	0.0498 (0.0213)	0.6298	0.0361 (0.0140)	0.5138
	LL	0.0650 (0.0161)	0.5280	0.0549 (0.0125)	0.3905	0.0438 (0.0099)	0.2552

Note: Numbers in parentheses are simulation standard deviations of RISEs. "b or h" denotes simulation averages of the values of smoothing parameters b for W, NM, MG, and G, or the lengths of bandwidths h for LT and LL.

in unreported simulations where the truth is GG$(5, 2, 2.5)$ (that puts a very small weight over the boundary region), we obtain $(b_{\mathrm{NM}}, b_{\mathrm{MG}}) = (0.0884, 0.0452)$, (0.0658,0.0334), (0.0460,0.0232) for $n = 100, 200, 500$ so that the relation is confirmed for each sample size.

5. Conclusion

Unlike symmetric kernels, exploring asymptotics on asymmetric kernels has relied on kernel-specific, diversified approaches. To pursue a unified approach in their asymptotics, this paper has proposed a new class of asymmetric kernels that is built on the GG pdf and a few common conditions and thus referred to as a family of the GG kernels. The family can generate asymmetric kernels that share appealing properties with the MG kernel. Asymptotics on kernels belonging to the family can be delivered by manipulating the conditions directly, as with symmetric kernels. As special cases, the family encompasses the W and NM kernels, as well as the MG kernel. This paper also demonstrates three additional properties of the GG density estimator including applicability of MBC and weak consistency for the density unbounded at the origin. Monte Carlo simulations indicate good finite sample properties of GG density estimators.

There are two possible research extensions. First, choice methods of the smoothing parameter b in GG density estimation need to be further investigated. Promising candidates include a more sophisticated plug-in method like the solve-the-equation plug-in method by Sheather and Jones (1991) and the cross-validation method. In particular, to the best of our knowledge, no equivalent method to the former has ever been proposed for asymmetric kernel density estimation. Second, goodness-of-fit tests can be built on the GG kernels. Fernandes and Grammig (2005) and Fernandes, Mendes, and Scaillet (2011) obtain favorable finite-sample results from applying the G kernel to specification testing for duration models and testing for symmetry, respectively. Combining the GG kernels with these testing procedures appear to be another good application of the kernels. These extensions are currently under authors' investigation and will be addressed in separate papers.

Acknowledgements

We would like to thank the editor Irène Gijbels, an associate editor and two anonymous referees for their comments that have substantially improved this paper. We are also grateful to Jin Seo Cho, Benedikt Funke, Bruce Hansen, Susumu Imai, Yoshihide Kakizawa, Yoshihiko Nishiyama, Katsumi Shimotsu, Nabil Zougab, and participants at Summer Workshop on Economic Theory 2012, New Zealand Econometric Study Group Meeting 2013, and seminars at the Development Bank of Japan, Kyoto University, the University of Sydney and the University of Tokyo for constructive comments and suggestions. The views expressed herein and those of the authors do not necessarily reflect the views of the Development Bank of Japan.

Conflict of interest disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first author gratefully acknowledges financial support from Japan Society of the Promotion of Science [grant number 23530259].

Notes

1. Conditions 1–5 even establish approximations to the bias and variance of nonparametric regression estimators (e.g. local constant and LL estimators) using the GG kernels.
2. Simulation results of GG density estimators with the rule-of-thumb smoothing parameters plugged in are available on the first author's webpage.
3. This definition differs from the one given in Kuruwita, Kulasekera, and Padgett (2010, Table 1), who construct their Weibull kernel from a different motivation. Apparently, their definition

$$K_{\mathrm{W}}(u;x,b)=\frac{1}{bx}\left(\frac{u}{x}\right)^{1/b-1}\exp\left\{-\left(\frac{u}{x}\right)^{1/b}\right\}\mathbf{1}\{u\geq 0\}$$

tends to be unbounded near the origin.

4. The pdf of $\mathrm{NM}(\mu, \omega)$ ($\mu \geq 1/2, \omega > 0$) is

$$p(y; \mu, \omega) = \frac{2\mu^{\mu}}{\omega^{\mu}\Gamma(\mu)} y^{2\mu-1} \exp\left(-\frac{\mu}{\omega}y^2\right) \mathbf{1}\{y \geq 0\}.$$

It is widely recognised that the Nakagami-*m* distribution was first proposed in Nakagami (1960). In reality, however, the distribution originates as early as in Nakagami (1943).
5. This kernel does not belong to the GG kernels, because it can be obtained by setting $(\alpha, \beta) = (x/b + 1, x + b)$ in Equation (9).
6. Because an error is found in the bias approximation for the Terrell and Scott (1980)-type MBC estimator in Hirukawa (2010) and Hirukawa and Sakudo (2014), we have corrected the error here. We thank Dr Nabil Zougab for pointing it out.
7. Because the third-order term in $\mathrm{Bias}\{\hat{f}_{\mathrm{W}}(x)\}$ is $O(b^2)$, it would be possible in principle to improve the bias convergence of this estimator from $O(b)$ to $O(b^2)$ by employing one of the bias correction techniques twice. However, it is doubtful whether there is much gain in practice from the bias correction via iteration, and thus we do not consider such inferior cases any further.
8. Carrasco and Chen (2002) provide the conditions that make GARCH processes stationary and β-mixing with exponential decay, whereas Chen, Hansen, and Carrasco (2010) discuss the conditions that can establish β-mixing with exponential decay in scalar diffusion processes. In relation to the latter, Feller's square-root process and its inverse, which have been employed as models of short-term interest rates by Cox, Ingersoll, and Ross (1985) and Ahn and Gao (1999), respectively, are examples of exponentially decaying β-mixing processes. Since β-mixing implies strong mixing, their assumption can cover many important applications in economics and finance.

References

Abramson, I.S. (1982), 'On Bandwidth Variation in Kernel Estimates – A Square Root Law', *Annals of Statistics*, 10, 1217–1223.

Ahn, D.H., and Gao, B. (1999), 'A Parametric Nonlinear Model of Term Structure Dynamics', *Review of Financial Studies*, 12, 721–762.

Bouezmarni, T., and Rombouts, J.V.K. (2010), 'Nonparametric Density Estimation for Positive Time Series', *Computational Statistics & Data Analysis*, 54, 245–261.

Bouezmarni, T., and Scaillet, O. (2005), 'Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data', *Econometric Theory*, 21, 390–412.

Buch-Larsen, T., Nielsen, J.P., Guillén, M., and Bolancé, C. (2005), 'Kernel Density Estimation for Heavy-Tailed Distributions Using the Champernowne Transformation', *Statistics*, 39, 503–518.

Carrasco, M., and Chen, X. (2002), 'Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models', *Econometric Theory*, 18, 17–39.

Chen, S.X. (2000), 'Probability Density Function Estimation Using Gamma Kernels', *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

Chen, X., Hansen, L.P., and Carrasco, M. (2010), 'Nonlinearity and Temporal Dependence', *Journal of Econometrics*, 155, 155–169.

Cox, J.C., Ingersoll Jr, J.E., and Ross, S.A. (1985), 'A Theory of the Term Structure of Interest Rates', *Econometrica*, 53, 385–408.

Fernandes, M., and Grammig, J. (2005), 'Nonparametric Specification Tests for Conditional Duration Models', *Journal of Econometrics*, 127, 35–68.

Fernandes, M., Mendes, E.F., and Scaillet, O. (2011), 'Testing for Symmetry and Conditional Symmetry Using Asymmetric Kernels', Swiss Finance Institute Occasional Paper Series No.11-32.

Gospodinov, N., and Hirukawa, M. (2012), 'Nonparametric Estimation of Scalar Diffusion Models of Interest Rates Using Asymmetric Kernels', *Journal of Empirical Finance*, 19, 595–609.

Gustafsson, J., Hagmann, M., Nielsen, J.P., and Scaillet, O. (2009), 'Local Transformation Kernel Density Estimation of Loss Distributions', *Journal of Business & Economic Statistics*, 27, 161–175.

Hagmann, M., and Scaillet, O. (2007), 'Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators', *Journal of Econometrics*, 141, 213–249.

Hirukawa, M. (2010), 'Nonparametric Multiplicative Bias Correction for Kernel-Type Density Estimation on the Unit Interval', *Computational Statistics & Data Analysis*, 54, 473–495.

Hirukawa, M., and Sakudo, M. (2014), 'Nonnegative Bias Reduction Methods for Density Estimation Using Asymmetric Kernels', *Computational Statistics & Data Analysis*, 75, 112–123.

Jin, X., and Kawczak, J. (2003), 'Birnbaum-Saunders and Lognormal Kernel Estimators for Modelling Durations in High Frequency Financial Data', *Annals of Economics and Finance*, 4, 103–124.

Jones, M.C. (1993), 'Simple Boundary Correction for Kernel Density Estimation', *Statistics and Computing*, 3, 135–146.

Jones, M.C., Linton, O., and Nielsen, J.P. (1995), 'A Simple Bias Reduction Method for Density Estimation', *Biometrika*, 82, 327–338.

Karunamuni, R.J., and Alberts, T. (2005), 'A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation', *Canadian Journal of Statistics*, 33, 497–509.

Kuruwita, C.N., Kulasekera, K.B., and Padgett, W.J. (2010), 'Density Estimation Using Asymmetric Kernels and Bayes Bandwidths with Censored Data', *Journal of Statistical Planning and Inference*, 140, 1765–1774.
Malec, P., and Schienle, M. (2014), 'Nonparametric Kernel Density Estimation Near the Boundary', *Computational Statistics & Data Analysis*, 72, 57–76.
Nakagami, M. (1943), 'Some Statistical Characters of Short-Wave Fading' [in Japanese], *Journal of the Institute of Electrical Communication Engineers of Japan*, 27, 145–150.
Nakagami, M. (1960), 'The *m*-Distribution – A General Formula of Intensity Distribution of Rapid Fading', in *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium Held at the University of California, Los Angeles, June 18–20, 1958*, ed. W.C. Hoffman, New York: Pergamon Press, pp. 3–36.
Scaillet, O. (2004), 'Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels', *Journal of Nonparametric Statistics*, 16, 217–226.
Sheather, S.J., and Jones, M.C. (1991), 'A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation', *Journal of the Royal Statistical Society, Series B*, 53, 683–690.
Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.
Stacy, E.W. (1962), 'A Generalization of the Gamma Distribution', *Annals of Mathematical Statistics*, 33, 1187–1192.
Stone, C.J. (1980), 'Optimal Rates of Convergence for Nonparametric Estimators', *Annals of Statistics*, 8, 1348–1360.
Terrell, G.R., and Scott, D.W. (1980), 'On Improving Convergence Rates for Nonnegative Kernel Density Estimators', *Annals of Statistics*, 8, 1160–1163.
Wand, M.P., Marron, J.S., and Ruppert, D. (1991), 'Transformations in Density Estimation', *Journal of the American Statistical Association*, 86, 343–353.

Appendix

In order to approximate the gamma function, we frequently refer to the following well-known formulae:

Stirling's formula (SF).

$$\Gamma(z+1) = \sqrt{2\pi} z^{z+1/2}\,\mathrm{e}^{-z}\left\{1 + \frac{1}{12z} + \frac{1}{288z^2} + O(z^{-3})\right\} \quad \text{as } z \to \infty.$$

Series expansion of the log gamma function (SELG).

$$\log\Gamma(1+z) = -\gamma z + \sum_{k=2}^{\infty}\frac{(-1)^k\zeta(k)}{k}z^k \quad \text{for } |z| < 1,$$

where (only in this context) $\gamma = \lim_{n\to\infty}(\sum_{k=1}^{n} k^{-1} - \log n) = 0.5772156649\ldots$ is Euler's constant, and $\zeta(k) = \sum_{n=1}^{\infty} n^{-k}$ $(k > 1)$ is the Riemann zeta function.

Legendre's duplication formula (LDF).

$$\Gamma(z)\Gamma\left(z+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2z-1}}\Gamma(2z) \quad \text{for } z > 0.$$

A.1. *Proof of Theorem 1*

Bias. Let $\theta_x \overset{d}{=} \mathrm{GG}(\alpha, \beta\Gamma(\alpha/\gamma)/\Gamma\{(\alpha+1)/\gamma\}, \gamma)$. Then, a second-order Taylor expansion of $E\{\hat{f}_{\mathrm{GG}}(x)\}$ around $\theta_x = x$ yields $E\{\hat{f}_{\mathrm{GG}}(x)\} = f(x) + E(\theta_x - x)f'(x) + (1/2)E(\theta_x - x)^2 f''(x) + o\{E(\theta_x - x)^2\}$. It follows from Equation (2) that

$$E(\theta_x^m) = \beta^m\frac{\{\Gamma(\alpha/\gamma)\}^{m-1}\Gamma\{(\alpha+m)/\gamma\}}{[\Gamma\{(\alpha+1)/\gamma\}]^m}.$$

In particular, $E(\theta_x) = \beta$ (by construction) and

$$E(\theta_x^2) = \beta^2\frac{\Gamma(\alpha/\gamma)\Gamma\{(\alpha+2)/\gamma\}}{[\Gamma\{(\alpha+1)/\gamma\}]^2} = \beta^2 M_b(x).$$

Using Conditions 1 and 3, we have, for $x \geq C_1 b$, $E(\theta_x) = x$ and $E(\theta_x^2) = x^2 + C_5 xb + o(b)$, and thus $E(\theta_x - x) = 0$ and $E(\theta_x - x)^2 \sim C_5 xb$. As a consequence, $\mathrm{Bias}\{\hat{f}_{\mathrm{GG}}(x)\} \sim (C_5/2)xf''(x)b$. On the other hand, for $x \in [0, C_1 b)$, $E(\theta_x) = \varphi_b(x) = O(b)$ and $x = O(b)$ hold, and thus $E(\theta_x - x) = [\{\varphi_b(x) - x\}/b]b := \xi_b(x)b$, where $\xi_b(x) = O(1)$. Moreover, it follows from $E(\theta_x^2) = O(b^2)$ that $E(\theta_x - x)^2 = O(b^2)$. Therefore, $\mathrm{Bias}\{\hat{f}_{\mathrm{GG}}(x)\} \sim \xi_b(x)f'(x)b$.

Variance. As usual, we consider the approximation $\mathrm{Var}\{\hat{f}_{\mathrm{GG}}(x)\} = (1/n)[E\{K_{GG}^2(X_i; x, b)\} + O(1)]$. A straightforward calculation yields, for $\nu \in \mathbb{R}_+$,

$$K_{\mathrm{GG}}^{\nu}(u; x, b) = \left[\left\{\frac{\gamma\Gamma((\alpha+1)/\gamma)}{\beta}\right\}^{\nu-1} \frac{\Gamma\{(\nu(\alpha-1)+1)/\gamma\}}{\nu^{(\nu(\alpha-1)+1)/\gamma}\{\Gamma(\alpha/\gamma)\}^{2\nu-1}}\right]$$
$$\cdot \frac{\gamma u^{\{\nu(\alpha-1)+1\}-1}\exp[-\{\frac{u}{\beta\Gamma(\alpha/\gamma)/(\nu^{1/\gamma}\Gamma((\alpha+1)/\gamma))}\}^{\gamma}]}{[\beta\Gamma(\alpha/\gamma)/\{\nu^{1/\gamma}\Gamma((\alpha+1)/\gamma)\}]^{\nu(\alpha-1)+1}\Gamma\{(\nu(\alpha-1)+1)/\gamma\}}\mathbf{1}\{u \geq 0\}$$
$$= A_{b,\nu}(x) \cdot \left\{\text{pdf of GG}\left(\nu(\alpha-1)+1, \frac{\beta\Gamma(\alpha/\gamma)}{\nu^{1/\gamma}\Gamma((\alpha+1)/\gamma)}, \gamma\right)\right\}.$$

Hence, $\mathrm{Var}\{\hat{f}_{\mathrm{GG}}(x)\} \sim (1/n)A_{b,2}(x)E\{f(\varsigma_x)\}$, where

$$\varsigma_x \stackrel{d}{=} \mathrm{GG}\left(2\alpha - 1, \frac{\beta\Gamma(\alpha/\gamma)}{2^{1/\gamma}\Gamma\{(\alpha+1)/\gamma\}}, \gamma\right).$$

By the mean-value theorem, $E\{f(\varsigma_x)\} = f(x) + E(\varsigma_x - x)f'(\bar{\varsigma}_x)$ for some $\bar{\varsigma}_x$ joining ς_x and x. It follows from Equation (2) that

$$E(\varsigma_x) = \beta\left[\frac{\Gamma(\alpha/\gamma)\Gamma(2\alpha/\gamma)}{2^{1/\gamma}\Gamma\{(\alpha+1)/\gamma\}\Gamma\{(2\alpha-1)/\gamma\}}\right] = \beta H_b(x).$$

Then, Condition 4, together with Condition 1, implies that $E(\varsigma_x - x) = o(1)$ so that $E\{f(\varsigma_x)\} \sim f(x)$ regardless of the position of x. Finally, Condition 5 establishes the approximation to the variance.

A.2. *Proof of Theorem 2*

A.2.1. *The MG kernel*

Condition 3. Observe that $M_b(x) = 1 + 1/\alpha$. For $x \geq 2b$, $M_b(x) = 1 + (1/x)b$ so that $C_{5,\mathrm{MG}} = 1$. On the other hand, for $x \in [0, 2b)$, $\alpha = (x/b)^2/4 + 1 = O(1)$, and thus $M_b(x) = O(1)$ holds.

Condition 4. Substituting $\alpha = x/b$ into $H_b(x) = 1 - (2\alpha)^{-1}$ yields $H_b(x) = 1 + O(b) = 1 + o(1)$ for interior x. On the other hand, for boundary x, $\alpha = O(1)$ regardless of whether $\alpha = x/b$ (when $x = O(b)$ and $x \geq 2b$) or $\alpha = (x/b)^2/4 + 1$ (when $x = O(b)$ and $x \in [0, 2b)$). It follows that $H_b(x) = O(1)$.

Condition 5. Because $\beta/\alpha = b$ regardless of the position of x, we have

$$A_{b,\nu}(x) = b^{1-\nu}\frac{\Gamma\{\nu(\alpha-1)+1\}}{\nu^{\nu(\alpha-1)+1}\{\Gamma(\alpha)\}^{\nu}}.$$

For interior x, SF yields, as $\alpha = x/b \to \infty$, $\Gamma(\alpha) \sim \sqrt{2\pi}\alpha^{\alpha-1/2}\mathrm{e}^{-\alpha}$ and $\Gamma\{\nu(\alpha-1)+1\} \sim \sqrt{2\pi}\{\nu(\alpha-1)\}^{\nu(\alpha-1)+1/2}\mathrm{e}^{-\nu(\alpha-1)}$. Then,

$$A_{b,\nu}(x) \sim \frac{b^{1-\nu}(\alpha-1)^{(1-\nu)/2}}{\nu^{1/2}(\sqrt{2\pi})^{\nu-1}}.$$

The result immediately follows from defining $V_{\mathrm{I,MG}}(\nu) := \{\nu^{1/2}(\sqrt{2\pi})^{\nu-1}\}^{-1}$ and recognising that $(\alpha-1)^{(1-\nu)/2} = \alpha^{(1-\nu)/2}(1-1/\alpha)^{(1-\nu)/2} \sim (x/b)^{(1-\nu)/2}$. For boundary x, the result is established by defining

$$V_{\mathrm{B,MG}}(\nu) := \begin{cases} \dfrac{\Gamma\{\nu(\kappa-1)+1\}}{\nu^{\nu(\kappa-1)+1}\{\Gamma(\kappa)\}^{\nu}} & \text{if } \dfrac{x}{b} \to \kappa \geq 2 \\ \dfrac{\Gamma((\nu/4)\kappa^2+1)}{\nu^{(\nu/4)\kappa^2+1}\{\Gamma(\kappa^2/4+1)\}^{\nu}} & \text{if } \dfrac{x}{b} \to \kappa \in (0,2) \end{cases}.$$

A.2.2. *The W kernel*

Condition 3. Observe that $M_b(x) = \Gamma(1+2/\alpha)/\{\Gamma(1+1/\alpha)\}^2$. For $x \geq 2b$, we may pick an arbitrarily small $b > 0$ so that $|2/\alpha| = |\sqrt{2b/x}| \leq 1$. Then, by SELG and $\zeta(2) = \pi^2/6$, two gamma functions admit the following approximations:

$$\Gamma\left(1+\frac{2}{\alpha}\right) = \exp\left\{\log\Gamma\left(1+\frac{2}{\alpha}\right)\right\} = 1 - \frac{2\gamma}{\alpha} + \frac{(\pi^2/3)+2\gamma^2}{\alpha^2} + O(\alpha^{-3}),$$
$$\Gamma\left(1+\frac{1}{\alpha}\right) = \exp\left\{\log\Gamma\left(1+\frac{1}{\alpha}\right)\right\} = 1 - \frac{\gamma}{\alpha} + \frac{(\pi^2/12)+(\gamma^2/2)}{\alpha^2} + O(\alpha^{-3}).$$

Applying a geometric series expansion to the approximation to $\{\Gamma(1+1/\alpha)\}^{-2}$ finally delivers $M_b(x) \sim 1 + (\pi^2/6)/\alpha^2 = 1 + \{\pi^2/(12x)\}b$ so that $C_{5,\mathrm{W}} = \pi^2/12$. On the other hand, for $x \in [0, 2b)$, $\alpha = (x/b)/2 + 1 = O(1)$, and thus $M_b(x) = O(1)$ is the case.

Condition 4. In this case, $H_b(x) = \{2^{1/\alpha}(1-1/\alpha)\Gamma(1+1/\alpha)\Gamma(1-1/\alpha)\}^{-1}$. It is easy to see that $H_b(x) = O(1)$ for boundary x. On the other hand, for interior x, $2^{1/\alpha} = \exp\{(1/\alpha)\log 2\} = 1 + O(b^{1/2})$, and $(1-1/\alpha)\Gamma(1+1/\alpha)\Gamma(1-1/\alpha) = 1 + O(b^{1/2})$ by SELG. Therefore, $H_b(x) = 1 + O(b^{1/2}) = 1 + o(1)$ holds.

Condition 5. $A_{b,\nu}(x)$ reduces to

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

For interior x, because $\alpha = \sqrt{2x/b} \to \infty$, we may approximate $\Gamma(1+1/\alpha) \sim 1$, $\Gamma\{\nu+(1-\nu)/\alpha\} \sim \Gamma(\nu)$ and $\nu^{\nu+(1-\nu)/\alpha} \sim \nu^\nu$. In addition, $(\alpha/\beta)^{\nu-1} = 2^{(\nu-1)/2}(xb)^{(1-\nu)/2}$, and thus

$$A_{b,\nu}(x) \sim 2^{(\nu-1)/2}\frac{\Gamma(\nu)}{\nu^\nu}(xb)^{(1-\nu)/2} := V_{\mathrm{I,W}}(\nu)(xb)^{(1-\nu)/2}.$$

For boundary x, the result is established by defining

$$V_{\mathrm{B,W}}(\nu) := \begin{cases} \left(\dfrac{2}{\kappa}\right)^{(\nu-1)/2} \dfrac{\{\Gamma(1+1/\sqrt{2\kappa})\}^{\nu-1}\Gamma(\nu+(1-\nu)/\sqrt{2\kappa})}{\nu^{\nu+(1-\nu)/\sqrt{2\kappa}}} & \text{if } \dfrac{x}{b} \to \kappa \geq 2, \\ \left\{\dfrac{2(\kappa+2)}{\kappa^2+4}\right\}^{\nu-1} \dfrac{\{\Gamma(1+2/(\kappa+2))\}^{\nu-1}\Gamma\{\nu+2(1-\nu)/(\kappa+2)\}}{\nu^{\nu+2(1-\nu)/(\kappa+2)}} & \text{if } \dfrac{x}{b} \to \kappa \in (0,2). \end{cases}$$

A.2.3. *The NM kernel*

Condition 3. Observe that $M_b(x) = (\alpha/2)[\Gamma(\alpha/2)/\Gamma\{(\alpha+1)/2\}]^2$. For $x \geq 2b$, it follows from LDF that $M_b(x) = (\alpha/2)[2^{\alpha-1}\{\Gamma(\alpha/2)\}^2/\{\sqrt{\pi}\Gamma(\alpha)\}]^2$. Next, SF implies that, as $\alpha = x/b \to \infty$,

$$\Gamma(\alpha) = \sqrt{2\pi}\alpha^{\alpha-1/2}\,\mathrm{e}^{-\alpha}\left\{1 + \frac{1}{12\alpha} + O(\alpha^{-2})\right\}, \tag{A1}$$

$$\Gamma\left(\frac{\alpha}{2}\right) = \sqrt{2\pi}\left(\frac{\alpha}{2}\right)^{\alpha/2-1/2}\,\mathrm{e}^{-\alpha/2}\left\{1 + \frac{1}{6\alpha} + O(\alpha^{-2})\right\}. \tag{A2}$$

Then, we find that the approximation of $M_b(x)$ takes a very simple form $M_b(x) = 1 + (2\alpha)^{-1} + O(\alpha^{-2}) = 1 + (2x)^{-1}b + o(b)$ so that $C_{5,\mathrm{NM}} = 1/2$. On the other hand, for $x \in [0, 2b)$, $\alpha = (x/b)^2/4 + 1 = O(1)$, and thus $M_b(x) = O(1)$ holds.

Condition 4. $H_b(x) = (\alpha - 1/2)\Gamma(\alpha/2)\Gamma(\alpha)/[2^{1/2}\Gamma\{(\alpha+1)/2\}\Gamma(\alpha+1/2)]$ is $O(1)$ for boundary x because of the same reason as in the proof for the MG kernel. For interior x, using LDF and then $\Gamma(2\alpha) \sim \sqrt{2\pi}(2\alpha)^{2\alpha-1/2}\,\mathrm{e}^{-2\alpha}$ as $\alpha = x/b \to \infty$, as well as Equations (A1) and (A2), yields

$$H_b(x) = \frac{\alpha-(1/2)}{2^{1/2}}\frac{2^{3\alpha-2}}{\pi}\frac{\{\Gamma(\alpha/2)\}^2\Gamma(\alpha)}{\Gamma(2\alpha)} \sim 1 - \frac{1}{2\alpha} = 1 + O(b) = 1 + o(1).$$

Condition 5. In this case,

$$A_{b,\nu}(x) = \beta^{1-\nu}\frac{2^{\nu-1}}{\nu^{\nu(\alpha-1)/2+1/2}}\frac{\{\Gamma((\alpha+1)/2)\}^{\nu-1}\Gamma\{\nu(\alpha-1)/2+1/2\}}{\{\Gamma(\alpha/2)\}^{2\nu-1}}.$$

For interior x, by LDF, $A_{b,\nu}(x)$ reduces to

$$A_{b,\nu}(x) = \beta^{1-\nu}\frac{2^{\nu-1}}{\nu^{\nu(\alpha-1)/2+1/2}}\frac{(\sqrt{\pi})^\nu\{\Gamma(\alpha)\}^{\nu-1}\Gamma\{\nu(\alpha-1)\}}{2^{(\nu-1)(\alpha-1)+\nu(\alpha-1)-1}\Gamma\{\nu(\alpha-1)/2\}\{\Gamma(\alpha/2)\}^{3\nu-2}}.$$

Next, by SF, as $\alpha = x/b \to \infty$, $\Gamma\{\nu(\alpha-1)\} \sim \sqrt{2\pi}\{\nu(\alpha-1)\}^{\nu(\alpha-1)-1/2}\,\mathrm{e}^{-\nu(\alpha-1)}$ and $\Gamma\{\nu(\alpha-1)/2\} \sim \sqrt{2\pi}\{\nu(\alpha-1)/2\}^{\nu(\alpha-1)-1/2}\,\mathrm{e}^{-\nu(\alpha-1)/2}$. Substituting these approximations, as well as Equations (A1) and (A2), we finally deduce that

$$A_{b,\nu}(x) \sim \beta^{1-\nu}\alpha^{(\nu-1)/2}\frac{\mathrm{e}^{\nu/2}}{\nu^{1/2}(\sqrt{\pi})^{\nu-1}}\left(1-\frac{1}{\alpha}\right)^{\nu(\alpha-1)/2}.$$

Moreover $\beta^{1-\nu}\alpha^{(1-\nu)/2} = (xb)^{(1-\nu)/2}$ and $(1-1/\alpha)^{\nu(\alpha-1)/2} \sim \mathrm{e}^{-\nu/2}$, and thus

$$A_{b,\nu}(x) \sim \frac{1}{\nu^{1/2}(\sqrt{\pi})^{\nu-1}}(xb)^{(1-\nu)/2} := V_{\mathrm{I,NM}}(\nu)(xb)^{(1-\nu)/2}.$$

For boundary x, the result is established by defining

$$V_{\mathrm{B,NM}}(\nu) := \begin{cases} \left(\dfrac{2}{\kappa}\right)^{\nu-1}\dfrac{\{\Gamma((\kappa+1)/2)\}^{\nu-1}\Gamma\{(\nu(\kappa-1)+1)/2\}}{\nu^{(\nu(\kappa-1)+1)/2}\{\Gamma(\kappa/2)\}^{2\nu-1}} & \text{if } \dfrac{x}{b}\to\kappa\geq 2, \\ \left(\dfrac{8}{\kappa^2+4}\right)^{\nu-1}\dfrac{\{\Gamma(\kappa^2/8+1)\}^{\nu-1}\Gamma((\nu/8)\kappa^2+1/2)}{\nu^{(\nu/8)\kappa^2+1/2}\{\Gamma(\kappa^2/8+1/2)\}^{2\nu-1}} & \text{if } \dfrac{x}{b}\to\kappa\in(0,2). \end{cases}$$