Computational Statistics and Data Analysis 75 (2014) 112–123

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Nonnegative bias reduction methods for density estimation using asymmetric kernels

Masayuki Hirukawa [a,*], Mari Sakudo [b,c,1]

[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

[c] Institute for Research in Contemporary Political and Economic Affairs, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan

ARTICLE INFO

Article history:
Received 2 July 2013
Received in revised form 10 December 2013
Accepted 7 January 2014
Available online 31 January 2014

Keywords:
Asymmetric kernel
Bias reduction
Boundary effect
Higher-order bias kernel

ABSTRACT

Two classes of multiplicative bias correction ("MBC") methods are applied to density estimation with support on $[0, \infty)$. It is demonstrated that under sufficient smoothness of the true density, each MBC technique reduces the order of magnitude in bias, whereas the order of magnitude in variance remains unchanged. Accordingly, the mean integrated squared error of each MBC estimator achieves a faster convergence rate of $O\left(n^{-8/9}\right)$ when best implemented, where n is the sample size. Furthermore, MBC estimators always generate nonnegative estimates by construction. Plug-in smoothing parameter choice rules for the estimators are proposed, and their finite sample performance is examined via Monte Carlo simulations.

1. Introduction

Hirukawa (2010) applied two classes of fully nonparametric multiplicative bias correction ("MBC") methods originally proposed for density estimation using symmetric kernels to estimate the density with support on $[0, 1]$ via nonstandard smoothing by the Beta kernel (Chen, 1999). This paper extends the analysis to estimating the density with support on $[0, \infty)$ by asymmetric kernels (Chen, 2000; Jin and Kawczak, 2003; Scaillet, 2004). Let $K_{j(x,b)}(\cdot)$ be the asymmetric kernel indexed by j that depends on a design point $x > 0$ and a smoothing parameter $b > 0$. Given a random sample $\{X_i\}_{i=1}^n$ drawn from a univariate distribution with density f that has support on $[0, \infty)$, the density estimator using asymmetric kernel j can be expressed as

$$\hat{f}_{j,b}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{j(x,b)}(X_i). \tag{1}$$

Throughout, the kernel j refers to one of the Gamma ("G"), Modified Gamma ("MG"), Inverse Gaussian ("IG"), Reciprocal Inverse Gaussian ("RIG"), Log-Normal ("LN"),[2] and Birnbaum–Saunders ("BS") kernels. Functional forms of these kernels are presented in Table 1. Asymmetric kernels have originally emerged as an alternative to boundary correction methods; see, for instance, Karunamuni and Alberts (2005) for a brief review of the methods. Indeed, because all kernels in Table 1 have

* Corresponding author. Tel.: +81 72 839 8095; fax: +81 72 839 8138.
E-mail addresses: hirukawa@econ.setsunan.ac.jp (M. Hirukawa), marisakudo@gmail.com (M. Sakudo).

1 Tel.: +81 3 3244 1241; fax: +81 3 3270 7084.

2 Our definition of the LN kernel slightly differs from the original one in Jin and Kawczak (2003). This definition ensures that the leading variance of the density estimator (1) becomes $n^{-1}b^{-1/2}f(x)/\left(2\sqrt{\pi}x\right)$ for a design point $x > 0$ so that $x/b \to \infty$ as $b \to 0$.

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2014.01.012

Table 1
Functional forms of asymmetric kernels.

Kernel (j)	Functional form ($u \geq 0$)
G (Chen, 2000)	$K_{G(x/b+1,b)}(u) = u^{x/b}\exp(-u/b)/\left\{b^{x/b+1}\Gamma(x/b+1)\right\}$.
MG (Chen, 2000)	$K_{MG(\rho_b(x),b)}(u) = u^{\rho_b(x)-1}\exp(-u/b)/\left[b^{\rho_b(x)}\Gamma\{\rho_b(x)\}\right]$, where $\rho_b(x) = \begin{cases} x/b & \text{for } x \geq 2b \\ (1/4)(x/b)^2+1 & \text{for } x \in [0, 2b) \end{cases}$.
IG (Scaillet, 2004)	$K_{IG(x,1/b)}(u) = \frac{1}{\sqrt{2\pi bu^3}}\exp\left\{-\frac{1}{2bx}\left(\frac{u}{x}-2+\frac{x}{u}\right)\right\}$.
RIG (Scaillet, 2004)	$K_{RIG(1/(x-b),1/b)}(u) = \frac{1}{\sqrt{2\pi bu}}\exp\left\{-\frac{x-b}{2b}\left(\frac{u}{x-b}-2+\frac{x-b}{u}\right)\right\}$.
LN (Jin and Kawczak, 2003)	$K_{LN(\log x,b)}(u) = \frac{1}{u\sqrt{2\pi b}}\exp\left\{-\frac{(\log u-\log x)^2}{2b}\right\}$.
BS (Jin and Kawczak, 2003)	$K_{BS(b^{1/2},x)}(u) = \frac{1}{2x\sqrt{2\pi b}}\left\{\left(\frac{x}{u}\right)^{1/2}+\left(\frac{x}{u}\right)^{3/2}\right\}\exp\left\{-\frac{1}{2b}\left(\frac{u}{x}-2+\frac{x}{u}\right)\right\}$.

support on $[0, \infty)$, they are free of boundary bias near the origin. Besides, the kernels have many other appealing properties, including locally adaptive smoothing via changing their shapes and 'shrinking variance' with the position of x.[3]

Below we formally define two MBC estimators built on the density estimator (1). Throughout (1) is called the bias-uncorrected ("BU") estimator to distinguish it from MBC estimators. In the spirit of Terrell and Scott (1980, abbreviated as "TS" hereafter), the first class of MBC techniques constructs a multiplicative combination of two density estimators employing the same kernel but different smoothing parameters. Let $\hat{f}_{j,b/c}(x)$ be the density estimator using asymmetric kernel j and smoothing parameter b/c, where $c \in (0, 1)$ is some predetermined constant that does not depend on the design point x. Then, the TS-MBC asymmetric kernel density estimator can be defined as

$$\tilde{f}_{TS,j}(x) = \left\{\hat{f}_{j,b}(x)\right\}^{\frac{1}{1-c}}\left\{\hat{f}_{j,b/c}(x)\right\}^{-\frac{c}{1-c}}. \tag{2}$$

On the other hand, the second class of MBC techniques due to Jones et al. (1995, abbreviated as "JLN" hereafter) utilizes a single smoothing parameter b. In light of the identity $f(x) = \hat{f}_{j,b}(x)\left\{f(x)/\hat{f}_{j,b}(x)\right\}$, the JLN-MBC asymmetric kernel density estimator can be defined as

$$\tilde{f}_{JLN,j}(x) = \hat{f}_{j,b}(x)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{K_{j(x,b)}(X_i)}{\hat{f}_{j,b}(X_i)}\right\}. \tag{3}$$

Recognize that the term inside the bracket is a natural nonparametric estimator of the bias-correction term $f(x)/\hat{f}_{j,b}(x)$. Also observe that both $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ are free of boundary bias and always generate nonnegative density estimates everywhere by construction.

Following the convention, this paper refers to the position of x as "interior x" if $x/b \to \infty$, and "boundary x" if $x/b \to \kappa$ for some constant $\kappa > 0$, as $b \to 0$. As demonstrated shortly, under sufficient differentiability of f, bias convergence of each MBC estimator is accelerated from $O(b)$ to $O(b^2)$, whereas the order of magnitude in variance remains unchanged from the one for (1), i.e. it is still $O\left\{\left(nb^{1/2}\right)^{-1}\right\}$ for interior x. Accordingly, the mean integrated squared error ("MISE") of each MBC estimator for interior x takes the form of $O\left(b^4 + n^{-1}b^{-1/2}\right)$. Therefore, when best implemented, each estimator can achieve the convergence rate of $O\left(n^{-8/9}\right)$ in MISE, which is faster than $O\left(n^{-4/5}\right)$, the MISE-optimal convergence rate within the class of nonnegative kernel estimators (Stone, 1980). Moreover, to implement MBC estimators employing the G and MG kernels, this paper proposes plug-in methods of choosing the smoothing parameter b with gamma density used as a reference.

A few articles other than Hirukawa (2010) have investigated bias reduction methods for density estimation via nonstandard smoothing when the support has a boundary. Hagmann and Scaillet (2007) and Gustafsson et al. (2009) study semi-parametric MBC methods for density estimation with support on $[0, \infty)$. Each method employs asymmetric kernels at the bias correction step after initial parametric density estimation. Unlike MBC methods in this paper, their approaches do not improve the bias convergence in order of magnitude. Moreover, Leblanc (2010) explores a bias reduction method for estimating the density with support on $[0, 1]$ using Bernstein polynomials, and establishes acceleration in bias convergence. However, he adopts an additive bias correction, and thus the bias-corrected estimator does not always generate nonnegative estimates unlike the one in Hirukawa (2010).

The remainder of this paper is organized as follows. Section 2 presents asymptotic properties of two MBC estimators. Section 3 proposes plug-in methods of choosing the smoothing parameter b for MBC estimators using the G and MG kernels, and conducts Monte Carlo simulations to check finite sample properties of the estimators. Section 4 applies two MBC techniques

[3] It is an open question whether the asymmetric kernels studied here may fit with nonparametric analysis of functional or infinite-dimensional data by Ferraty and Vieu (2006). While they consider the asymmetric kernels that take the form of $K((X-x)/b)$ for a data point X, design point x, and smoothing parameter b, none of the kernels in Table 1 can be expressed in this form.

to real data to illustrate how they work. Section 5 summarizes the main results of the paper and briefly mentions extensions of MBC to joint density estimation. Outline proofs and formulae of plug-in smoothing parameters are given in the Appendix.

This paper adopts the following notational conventions. $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\exp(-y)\,dy \ (\alpha > 0)$ denotes the gamma function. The expression '$X \overset{d}{=} Y$' reads "A random variable X obeys the distribution Y". Lastly, the expression '$X_n \sim Y_n$' is used whenever $X_n/Y_n \to 1$ as $n \to \infty$.

2. Main results

2.1. Asymptotic properties of MBC estimators

To develop convergence properties of MBC estimators, we make the following assumptions:

Assumption 1. f has four continuous and bounded derivatives, and $f(x) > 0$ for a given design point $x > 0$.

Assumption 2. The smoothing parameter $b = b_n\ (> 0)$ satisfies $b \to 0$ and $nb^{r_j+5/2} \to \infty$ as $n \to \infty$, where

$$r_j = \begin{cases} 1/2 & \text{for } j = G, MG, RIG \\ 1 & \text{for } j = LN, BS \\ 3/2 & \text{for } j = IG. \end{cases}$$

The smoothness condition on f in Assumption 1 is standard for consistency of density estimators using fourth-order kernels, whereas the positivity of $f(x)$ is required for MBC. Assumption 2 implies that the convergence rate of the smoothing parameter b is slower than $O\left\{n^{-1/(r_j+5/2)}\right\}$. We require this condition to control the order of magnitude in remainder terms when approximating the bias of each MBC estimator. It will be shown shortly that the MSE-optimal smoothing parameter for each estimator becomes $b^* = O\left(n^{-2/9}\right)$ for interior x and $b^\dagger = O\left\{n^{-1/\left(r_j+9/2\right)}\right\}$ for boundary x; these convergence rates are indeed within the required range.

Delivering bias and variance approximations for MBC estimators (2) and (3) requires (i) a second-order expansion of the bias term and (ii) an approximation to the leading variance of the BU estimator (1). For (i), Assumptions 1 and 2 imply that $E\left\{\hat{f}_{j,b}(x)\right\} = f(x) + a_{1,j}(x,f)\,b + a_{2,j}(x,f)\,b^2 + o\left(b^2\right)$, where $a_{1,j}(x,f)$ and $a_{2,j}(x,f)$ are kernel-specific functions that depend on the design point x and derivatives of f. Using properties of the random variable corresponding to each kernel, we can specify explicit forms of $a_{1,j}(x,f)$ and $a_{2,j}(x,f)$ as in Table 2. Furthermore, for (ii), the variance approximation is given by $\operatorname{Var}\left\{\hat{f}_{j,b}(x)\right\} \sim \left(nb^{1/2}\right)^{-1}\left(2\sqrt{\pi}x^{r_j}\right)^{-1} f(x) := \left(nb^{1/2}\right)^{-1} v_j(x) f(x)$ for interior x and $\operatorname{Var}\left\{\hat{f}_{j,b}(x)\right\} = O\left\{\left(nb^{r_j+1/2}\right)^{-1}\right\}$ for boundary x.

Below we present two theorems on the approximations to bias and variance terms of two MBC estimators.

Theorem 1. *If Assumptions 1 and 2 hold, then for $a_{1,j}(x,f)$ and $a_{2,j}(x,f)$, the bias of the TS-MBC estimator using kernel j can be approximated by*

$$\operatorname{Bias}\left\{\tilde{f}_{TS,j}(x)\right\} \sim \frac{1}{c(1-c)}\left[\frac{1}{2}\left\{\frac{a_{1,j}^2(x,f)}{f(x)}\right\} - a_{2,j}(x,f)\right] b^2 := \frac{1}{c(1-c)} p_j(x)\, b^2.$$

For $v_j(x)$ and r_j, the variance of the TS-MBC estimator can be approximated by

$$\operatorname{Var}\left\{\tilde{f}_{TS,j}(x)\right\} = \begin{cases} n^{-1}b^{-1/2}\lambda(c)\, v_j(x) f(x) + o\left\{\left(nb^{1/2}\right)^{-1}\right\} & \textit{for interior } x \\ O\left\{\left(nb^{r_j+1/2}\right)^{-1}\right\} & \textit{for boundary } x, \end{cases}$$

where

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

Theorem 2. *If Assumptions 1 and 2 hold, then the bias of the JLN-MBC estimator using kernel j can be approximated by*

$$\operatorname{Bias}\left\{\tilde{f}_{JLN,j}(x)\right\} \sim -f(x)\, a_{1,j}(x,h)\, b^2 := q_j(x)\, b^2,$$

where $a_{1,j}(x,h)$ is obtained by replacing $f = f(x)$ in $a_{1,j}(x,f)$ with

$$h = h_j(x,f) := \frac{a_{1,j}(x,f)}{f(x)}.$$

Table 2
Explicit forms of $a_{1,j}(x,f)$ and $a_{2,j}(x,f)$.

Kernel (j)	$a_{1,j}(x,f)$	$a_{2,j}(x,f)$
G	$f'(x)+\frac{x}{2}f''(x)$.	$f''(x)+\frac{5}{6}xf'''(x)+\frac{x^2}{8}f''''(x)$.
MG	$\begin{cases}\dfrac{x}{2}f''(x) & \text{for } x\geq 2b\\ \xi_b(x)f'(x) & \text{for } x\in[0,2b),\end{cases}$ where $\xi_b(x)=\rho_b(x)-\frac{x}{b}=O(1)$.	$\begin{cases}\dfrac{x}{3}f'''(x)+\dfrac{x^2}{8}f''''(x) & \text{for } x\geq 2b\\ \dfrac{1}{2}\left\{\xi_b^2(x)+\xi_b(x)+\dfrac{x}{b}\right\}f''(x) & \text{for } x\in[0,2b).\end{cases}$
IG	$\frac{x^3}{2}f''(x)$.	$\frac{x^5}{2}f'''(x)+\frac{x^6}{8}f''''(x)$.
RIG	$\frac{x}{2}f''(x)$.	$\frac{1}{2}f''(x)+\frac{x}{2}f'''(x)+\frac{x^2}{8}f''''(x)$.
LN	$\frac{x}{2}f'(x)+\frac{x^2}{2}f''(x)$.	$\frac{x}{8}f'(x)+\frac{7}{8}x^2f''(x)+\frac{3}{4}x^3f'''(x)+\frac{x^4}{8}f''''(x)$.
BS	$\frac{x}{2}f'(x)+\frac{x^2}{2}f''(x)$.	$\frac{3}{4}x^2f''(x)+\frac{3}{4}x^3f'''(x)+\frac{x^4}{8}f''''(x)$.

For $\upsilon_j(x)$ and r_j, the variance of the JLN-MBC estimator can be approximated by

$$\mathrm{Var}\left\{\tilde{f}_{JLN,j}(x)\right\}=\begin{cases}n^{-1}b^{-1/2}\upsilon_j(x)f(x)+o\left\{\left(nb^{1/2}\right)^{-1}\right\} & \textit{for interior } x\\ O\left\{\left(nb^{r_j+1/2}\right)^{-1}\right\} & \textit{for boundary } x.\end{cases}$$

2.2. Discussions

2.2.1. Local property

Leading bias and variance terms. Because the support of asymmetric kernels matches that of the true density f, both TS- and JLN-MBC estimators are free of boundary bias. More importantly, these estimators reduce the order of magnitude in bias from $O(b)$ to $O\left(b^2\right)$, while their variances are $O\left\{\left(nb^{1/2}\right)^{-1}\right\}$ for interior x and $O\left\{\left(nb^{r_j+1/2}\right)^{-1}\right\}$ for boundary x. Observe that orders of variances remain unchanged from those for the corresponding BU estimator (1). We can also compare $p_j(x)$ and $q_j(x)$ in leading bias terms with the corresponding ones for nonnegative symmetric kernels. As stated in Jones and Signorini (1997, Sections 3.2–3.3), when the symmetric kernels are employed, the term corresponding to $p_j(x)$ for TS-MBC is a linear combination of $f''''(x)$ and $\left\{f''(x)\right\}^2/f(x)$, and the term corresponding to $q_j(x)$ for JLN-MBC is proportional to $f(x)\left\{f''(x)/f(x)\right\}''$. The reason why $p_j(x)$ and $q_j(x)$ take more complicated forms is that while odd-order moments of symmetric kernels are exactly zero, those of asymmetric kernels around the design point x are often $O(b)$ or $O\left(b^2\right)$. As a result, extra density derivatives are included in $p_j(x)$ and $q_j(x)$.

The variance of JLN-MBC estimators is first-order asymptotically equivalent to that of the corresponding BU estimator (1) for interior x. While the semi-parametric MBC density estimator by Hagmann and Scaillet (2007) also yields the same leading variance term, this technique does not reduce the bias in order of magnitude. In contrast, when JLN-MBC is applied for the density estimation using nonnegative symmetric kernels, the leading variance term tends to be larger (although not inflated in order of magnitude) because the multiplier in the variance term involves the roughness (or squared integral) of the 'twiced' kernel (Stuetzle and Mittal, 1979). Besides, since the multiplier $\lambda(c)$ in the variance for TS-MBC estimators is increasing in $c\in(0,1)$, ranging from 1 to 27/16, the variance of these estimators tends to be larger than that of the BU estimator (1) for interior x. Lastly (but not least importantly), asymptotic variances of both TS- and JLN-MBC estimators for interior x are proportional to x^{-r_j}, even after MBC is made. The property of 'shrinking variance' with the position of x is equivalent to the strategy of using longer bandwidths over the tail region where the data are sparse. Therefore, MBC estimators share the appealing variance property with the corresponding BU estimator.

Mean squared error ("*MSE*"). For interior x, the MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ can be approximated by

$$MSE\left\{\tilde{f}_{TS,j}(x)\right\}=\frac{p_j^2(x)}{c^2(1-c)^2}b^4+n^{-1}b^{-1/2}\lambda(c)\,\upsilon_j(x)f(x)+o\left(b^4+n^{-1}b^{-1/2}\right),\quad\text{and}$$

$$MSE\left\{\tilde{f}_{JLN,j}(x)\right\}=q_j^2(x)\,b^4+n^{-1}b^{-1/2}\upsilon_j(x)f(x)+o\left(b^4+n^{-1}b^{-1/2}\right).$$

The MSE-optimal smoothing parameters are

$$b_{TS,j}^{*}=\left\{c^2(1-c)^2\lambda(c)\right\}^{2/9}\left\{\frac{\upsilon_j(x)f(x)}{8p_j^2(x)}\right\}^{2/9}n^{-2/9},\quad\text{and}$$

$$b^{*}_{JLN,j} = \left\{ \frac{\upsilon_j(x) f(x)}{8 q_j^2(x)} \right\}^{2/9} n^{-2/9},$$

which yield the optimal MSEs

$$MSE^{*}\left\{\tilde{f}_{TS,j}(x)\right\} \sim \frac{9}{8^{8/9}} \gamma(c)\, p_j^{2/9}(x) \left\{\upsilon_j(x) f(x)\right\}^{8/9} n^{-8/9}, \quad \text{and}$$

$$MSE^{*}\left\{\tilde{f}_{JLN,j}(x)\right\} \sim \frac{9}{8^{8/9}} q_j^{2/9}(x) \left\{\upsilon_j(x) f(x)\right\}^{8/9} n^{-8/9},$$

where

$$\gamma(c) = \left\{ \frac{\left(1+c^{5/2}\right)(1+c)^{1/2} - 2\sqrt{2} c^{3/2}}{c^{1/4}(1+c)^{1/2}(1-c)^{9/4}} \right\}^{8/9}.$$

Observe that the MSE-optimal smoothing parameters are $O\left(n^{-2/9}\right) = O\left(h^{*2}\right)$, where h^{*} is the MSE-optimal bandwidth for fourth-order kernel estimators, or TS- or JLN-MBC estimators using nonnegative symmetric kernels. As a result, the optimal MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ for interior x become $O\left(n^{-8/9}\right)$, as with MBC estimation using the second-order kernels. The convergence rate is faster than $O\left(n^{-4/5}\right)$, the optimal convergence rate in the MSE of the corresponding BU estimator (1) for interior x. On the other hand, for boundary x, the MSEs of $\tilde{f}_{TS,j}(x)$ and $\tilde{f}_{JLN,j}(x)$ are $O\left\{b^4 + n^{-1} b^{-(r_j+1/2)}\right\}$, which yields the MSE-optimal smoothing parameter $b^{\dagger} = O\left\{n^{-1/(r_j+9/2)}\right\}$ and the optimal MSE of $O\left\{n^{-4/(r_j+9/2)}\right\}$. The optimal convergence rate of MSEs is indeed faster than $O\left\{n^{-2/(r_j+5/2)}\right\}$, that of the BU estimator for boundary x.

2.2.2. Global property

The undesirable convergence rates over boundary regions do not affect the global properties of the MBC estimators. By the trimming argument in Chen (2000), the MISEs of the MBC estimators are

$$MISE\left\{\tilde{f}_{TS,j}(x)\right\} = \frac{b^4}{c^2(1-c)^2} \int_0^{\infty} p_j^2(x)\,dx + \frac{\lambda(c)}{n b^{1/2}} \int_0^{\infty} \upsilon_j(x) f(x)\,dx + o\left(b^4 + n^{-1} b^{-1/2}\right), \quad \text{and}$$

$$MISE\left\{\tilde{f}_{JLN,j}(x)\right\} = b^4 \int_0^{\infty} q_j^2(x)\,dx + \frac{1}{n b^{1/2}} \int_0^{\infty} \upsilon_j(x) f(x)\,dx + o\left(b^4 + n^{-1} b^{-1/2}\right),$$

provided that $p_j^2(x)$, $q_j^2(x)$, and $\upsilon_j(x)$ are integrable.[4] The MISE-optimal smoothing parameters are then given by

$$b^{**}_{TS,j} = \left\{c^2(1-c)^2 \lambda(c)\right\}^{2/9} \left\{ \frac{\int_0^{\infty} \upsilon_j(x) f(x)\,dx}{8 \int_0^{\infty} p_j^2(x)\,dx} \right\}^{2/9} n^{-2/9}, \quad \text{and}$$

$$b^{**}_{JLN,j} = \left\{ \frac{\int_0^{\infty} \upsilon_j(x) f(x)\,dx}{8 \int_0^{\infty} q_j^2(x)\,dx} \right\}^{2/9} n^{-2/9}.$$

Therefore, the optimal MISEs become

$$MISE^{**}\left\{\tilde{f}_{TS,j}(x)\right\} \sim \frac{9}{8^{8/9}} \gamma(c) \left\{\int_0^{\infty} p_j^2(x)\,dx\right\}^{2/9} \left\{\int_0^{\infty} \upsilon_j(x) f(x)\,dx\right\}^{8/9} n^{-8/9}, \quad \text{and}$$

$$MISE^{**}\left\{\tilde{f}_{JLN,j}(x)\right\} \sim \frac{9}{8^{8/9}} \left\{\int_0^{\infty} q_j^2(x)\,dx\right\}^{2/9} \left\{\int_0^{\infty} \upsilon_j(x) f(x)\,dx\right\}^{8/9} n^{-8/9}.$$

Furthermore, the multiplier $\gamma(c)$ in the optimal MISE for the TS-MBC estimator is minimized at $c^{*} \approx 0.2636$; this value is exclusively considered in subsequent analyses.

3. Finite sample performance

3.1. Monte Carlo setup

We evaluate the finite sample performance of two classes of MBC estimators via Monte Carlo simulations. From now on we concentrate on the G and MG kernels due to their popularity in the literature. This simulation study compares the following three classes of estimators: (i) BU estimators (1) [BU-G, BU-MG]; (ii) TS-MBC estimators (2) [TS-G, TS-MG]; and (iii) JLN-MBC estimators (3) [JLN-G, JLN-MG]. The value of the constant c in each TS-MBC estimator is set equal to the

[4] Whenever the integrated squared bias is considered, $p_{MG}(x)$ and $q_{MG}(x)$ refer to those for interior x (i.e. $x \geq 2b$).

Table 3
True distributions considered in Monte Carlo simulations.

Distribution	Density function $f(x)$, $x \geq 0$
1. Gamma	$x^{\alpha-1}\exp(-x/\beta)/\{\beta^{\alpha}\Gamma(\alpha)\}$, $(\alpha,\beta)=(1.5,1)$.
2. Weibull	$(\alpha/\beta)(x/\beta)^{\alpha-1}\exp\{-(x/\beta)^{\alpha}\}$, $(\alpha,\beta)=(1.5,1.5)$.
3. Half-Normal	$\frac{2}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $(\mu,\sigma)=(0,1.5)$.
4. Half-Logistic	$\left(\frac{2}{s}\right)\exp\left\{-\left(\frac{x-\mu}{s}\right)\right\}/\left[1+\exp\left\{-\left(\frac{x-\mu}{s}\right)\right\}\right]^2$, $(\mu,s)=(0,1)$.
5. Log-Normal	$\frac{1}{x\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}$, $(\mu,\sigma)=(0,0.75)$.
6. Pareto	$\rho\lambda^{\rho}/(x+\lambda)^{\rho+1}$, $(\lambda,\rho)=(1,2)$.
7. Burr	$\alpha\beta x^{\alpha-1}/(1+x^{\alpha})^{\beta+1}$, $(\alpha,\beta)=(1.5,2.5)$.
8. Generalized Gamma	$\gamma x^{\alpha-1}\exp\{-(x/\beta)^{\gamma}\}/\{\beta^{\alpha}\Gamma(\alpha/\gamma)\}$, $(\alpha,\beta,\gamma)=(5,2,2.5)$.
9. Log-Normal and Pareto Mixture 1	$p\left[\frac{1}{x\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}\right]+(1-p)\left\{\frac{\rho\lambda^{\rho}}{(x+\lambda)^{\rho+1}}\right\}$, $(p,\mu,\sigma,\lambda,\rho)=(0.7,0,0.5,1,2)$.
10. Log-Normal and Pareto Mixture 2	$p\left[\frac{1}{x\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}\right]+(1-p)\left\{\frac{\rho\lambda^{\rho}}{(x+\lambda)^{\rho+1}}\right\}$, $(p,\mu,\sigma,\lambda,\rho)=(0.3,0,0.5,1,2)$.

MISE-optimal $c^*=0.2636$. Ten true distributions are considered, as listed in Table 3. All these distributions are popularly chosen as models of, for example, the income distribution, the distribution of insurance claims and the baseline hazard.

For each distribution, 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]. As performance measures for each estimator $\bar{f}$, we compute the root integrated squared error ("RISE") $RISE\left\{\bar{f}(x)\right\}=\left[\int_0^{\infty}\left\{\bar{f}(x)-f(x)\right\}^2dx\right]^{1/2}$ and the integrated absolute bias ("IAB") $IAB\left\{\bar{f}(x)\right\}=\int_0^{\infty}\left|E\left\{\bar{f}(x)\right\}-f(x)\right|dx$. In our reports, the integrals are approximated over the 500 points, and the expected value is replaced by the simulation average.

3.2. Choices of smoothing parameters

Choosing the smoothing parameter b is an important practical issue. To expedite computations, as in Hirukawa (2010), we develop plug-in methods for TS- and JLN-MBC estimators that use gamma density as a reference.[5] The plug-in smoothing parameters for $\tilde{f}_{TS,MG}(x)$ and $\tilde{f}_{JLN,G}(x)$ (called "gamma-referenced smoothing parameters" hereafter) are defined as the minimizers of asymptotic weighted mean integrated squared errors ("AWMISEs")

$$\begin{aligned}
\hat{b}_{GR-TS} &= \arg\min_b AWMISE\left\{\tilde{f}_{TS,MG}(x)\right\}\\
&= \arg\min_b\left\{\frac{b^4}{c^2(1-c)^2}\int_0^{\infty}\tilde{p}^2_{MG}(x)\,w_{TS}(x)\,dx+\frac{\lambda(c)}{2\sqrt{\pi}nb^{1/2}}\int_0^{\infty}\frac{g(x)}{\sqrt{x}}w_{TS}(x)\,dx\right\}, \quad \text{and}\\
\hat{b}_{GR-JLN} &= \arg\min_b AWMISE\left\{\tilde{f}_{JLN,G}(x)\right\}\\
&= \arg\min_b\left\{b^4\int_0^{\infty}\tilde{q}^2_G(x)\,w_{JLN}(x)\,dx+\frac{1}{2\sqrt{\pi}nb^{1/2}}\int_0^{\infty}\frac{g(x)}{\sqrt{x}}w_{JLN}(x)\,dx\right\},
\end{aligned}$$

where $g(x)=x^{\alpha-1}\exp(-x/\beta)/\{\beta^{\alpha}\Gamma(\alpha)\}$ is the density function for the gamma distribution with parameters (α,β), and $\tilde{p}_{MG}(x)$ and $\tilde{q}_G(x)$ can be obtained by replacing $f(x)$ in $p_{MG}(x)$ and $q_G(x)$ with $g(x)$. The weighting functions are chosen as $w_{TS}(x)=x^5$ and $w_{\mathrm{JLN}}(x)=x$ to ensure finiteness of integrals. The parameters (α,β) are replaced by their estimates $\left(\hat{\alpha},\hat{\beta}\right)$ via method of moments or maximum likelihood, where the latter is exclusively used throughout. Analytical expressions of $\hat{b}_{GR-TS}$ and $\hat{b}_{GR-JLN}$, as well as $\hat{b}_{GR-BU}$ (= the gamma-referenced smoothing parameter for $\hat{f}_{MG}(x)$), are given in the Appendix.

We do not pursue the gamma-referenced smoothing parameter for $\tilde{f}_{TS,G}(x)$; since extra terms are involved in $p_G(x)$, the minimizer of its AWMISE takes a much more complicated form than $\hat{b}_{GR-TS}$. Moreover, although it is possible to derive the

[5] Alternatively, b could be chosen via a version of cross-validation methods. However, this appears to be a hard problem; indeed, even for nonnegative symmetric kernels, Jones and Signorini (1997) defer automatic bandwidth selection in this direction to future work.

Table 4
Simulation results.

Panel (a): $n = 100$

Estimator		Distribution 1		Distribution 2		Distribution 3		Distribution 4		Distribution 5	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.0389	0.0382	0.0397	0.0398	0.0350	0.0371	0.0330	0.0345	0.0710	0.0534
	SD	(0.0120)	(0.0123)	(0.0139)	(0.0138)	(0.0122)	(0.0130)	(0.0115)	(0.0122)	(0.0140)	(0.0153)
	IAB	*0.0175*	*0.0125*	*0.0183*	*0.0127*	*0.0135*	*0.0108*	*0.0137*	*0.0105*	*0.0328*	*0.0198*
BU-MG	RISE: Ave	0.0386	0.0386	0.0389	0.0396	0.0369	0.0397	0.0336	0.0364	0.0658	0.0469
	SD	(0.0113)	(0.0119)	(0.0126)	(0.0131)	(0.0144)	(0.0149)	(0.0134)	(0.0140)	(0.0158)	(0.0165)
	IAB	*0.0180*	*0.0126*	*0.0185*	*0.0125*	*0.0098*	*0.0077*	*0.0102*	*0.0076*	*0.0383*	*0.0211*
TS-G	RISE: Ave	0.0415	0.0394	0.0434	0.0414	0.0289	0.0381	0.0264	0.0345	0.0890	0.0509
	SD	(0.0110)	(0.0132)	(0.0145)	(0.0150)	(0.0131)	(0.0155)	(0.0119)	(0.0143)	(0.0118)	(0.0156)
	IAB	*0.0166*	*0.0062*	*0.0197*	*0.0056*	*0.0110*	*0.0044*	*0.0100*	*0.0037*	*0.0388*	*0.0119*
TS-MG	RISE: Ave	0.0445	0.0408	0.0436	0.0421	0.0370	0.0431	0.0362	0.0393	0.0926	0.0450
	SD	(0.0084)	(0.0127)	(0.0110)	(0.0144)	(0.0151)	(0.0177)	(0.0133)	(0.0168)	(0.0114)	(0.0158)
	IAB	*0.0235*	*0.0074*	*0.0223*	*0.0067*	*0.0170*	*0.0058*	*0.0202*	*0.0058*	*0.0563*	*0.0134*
JLN-G	RISE: Ave	0.0388	0.0365	0.0386	0.0365	0.0318	0.0327	0.0282	0.0302	0.0777	0.0566
	SD	(0.0126)	(0.0130)	(0.0154)	(0.0149)	(0.0138)	(0.0143)	(0.0127)	(0.0135)	(0.0147)	(0.0140)
	IAB	*0.0156*	*0.0104*	*0.0179*	*0.0129*	*0.0117*	*0.0099*	*0.0103*	*0.0075*	*0.0329*	*0.0210*
JLN-MG	RISE: Ave	0.0388	0.0384	0.0394	0.0385	0.0307	0.0325	0.0267	0.0303	0.0849	0.0636
	SD	(0.0090)	(0.0104)	(0.0109)	(0.0120)	(0.0141)	(0.0148)	(0.0129)	(0.0139)	(0.0129)	(0.0133)
	IAB	*0.0117*	*0.0077*	*0.0117*	*0.0085*	*0.0047*	*0.0042*	*0.0032*	*0.0023*	*0.0406*	*0.0253*
Estimator		Distribution 6		Distribution 7		Distribution 8		Distribution 9		Distribution 10	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.1287	0.0947	0.0690	0.0623	0.0525	0.0385	0.0663	0.0549	0.0756	0.0603
	SD	(0.0449)	(0.0321)	(0.0201)	(0.0206)	(0.0139)	(0.0136)	(0.0233)	(0.0168)	(0.0303)	(0.0191)
	IAB	*0.0411*	*0.0277*	*0.0230*	*0.0152*	*0.0405*	*0.0159*	*0.0338*	*0.0237*	*0.0328*	*0.0217*
BU-MG	RISE: Ave	0.1176	0.0836	0.0667	0.0616	0.0497	0.0379	0.0655	0.0560	0.0698	0.0581
	SD	(0.0448)	(0.0272)	(0.0188)	(0.0201)	(0.0139)	(0.0133)	(0.0234)	(0.0179)	(0.0275)	(0.0172)
	IAB	*0.0552*	*0.0354*	*0.0269*	*0.0174*	*0.0374*	*0.0147*	*0.0393*	*0.0267*	*0.0320*	*0.0224*
TS-G	RISE: Ave	0.1346	0.0797	0.0753	0.0626	0.0640	0.0405	0.0763	0.0529	0.0720	0.0516
	SD	(0.0498)	(0.0356)	(0.0173)	(0.0224)	(0.0124)	(0.0143)	(0.0231)	(0.0171)	(0.0337)	(0.0190)
	IAB	*0.0433*	*0.0213*	*0.0241*	*0.0087*	*0.0539*	*0.0059*	*0.0416*	*0.0178*	*0.0322*	*0.0154*
TS-MG	RISE: Ave	0.1764	0.1171	0.0841	0.0636	0.0572	0.0407	0.0795	0.0558	0.1000	0.0711
	SD	(0.0418)	(0.0307)	(0.0157)	(0.0211)	(0.0122)	(0.0142)	(0.0272)	(0.0199)	(0.0319)	(0.0221)
	IAB	*0.0967*	*0.0510*	*0.0365*	*0.0108*	*0.0461*	*0.0052*	*0.0535*	*0.0211*	*0.0603*	*0.0266*
JLN-G	RISE: Ave	0.1185	0.0762	0.0713	0.0676	0.0596	0.0366	0.0755	0.0619	0.0666	0.0506
	SD	(0.0535)	(0.0260)	(0.0215)	(0.0210)	(0.0108)	(0.0117)	(0.0279)	(0.0160)	(0.0345)	(0.0131)
	IAB	*0.0375*	*0.0219*	*0.0211*	*0.0191*	*0.0408*	*0.0186*	*0.0359*	*0.0276*	*0.0262*	*0.0173*
JLN-MG	RISE: Ave	0.1512	0.1043	0.0688	0.0663	0.0510	0.0338	0.0700	0.0595	0.0789	0.0624
	SD	(0.0523)	(0.0235)	(0.0159)	(0.0148)	(0.0096)	(0.0106)	(0.0271)	(0.0173)	(0.0350)	(0.0117)
	IAB	*0.0538*	*0.0331*	*0.0204*	*0.0179*	*0.0338*	*0.0163*	*0.0361*	*0.0269*	*0.0252*	*0.0204*

(*continued on next page*)

gamma-referenced smoothing parameter for $\tilde{f}_{JLN,MG}(x)$ in a similar way,[6] our preliminary simulation results indicate that the formula frequently generates large values and thus we do not advocate its use. From the viewpoint of practical relevance, $\hat{b}_{GR-TS}$ and $\hat{b}_{GR-JLN}$ are simply employed for $\tilde{f}_{TS,G}(x)$ and $\tilde{f}_{JLN,MG}(x)$ in our simulations, respectively. Similarly, $\hat{b}_{GR-BU}$ is chosen as the smoothing parameter for $\hat{f}_G(x)$.

Besides, a very simple formula is frequently applied in the literature (e.g. Gustafsson et al., 2009). In this respect, a "rule-of-thumb" smoothing parameter is also considered for each estimator. More precisely, we additionally employ $\hat{b}_{ROT-BU} = \hat{\sigma}_x n^{-2/5}$ for BU-G, BU-MG and $\hat{b}_{ROT-MBC} = \hat{\sigma}_x n^{-2/9}$ for four MBC estimators, where $\hat{\sigma}_x$ is the sample standard deviation of observations.

3.3. Simulation results

Table 4 presents Monte Carlo results. For each distribution, results are qualitatively similar across three sample sizes, and values of performance measures decrease with the sample size. As regards smoothing parameters, except average RISEs for Distributions 3 and 4, the gamma-referenced method in general works better than the rule-of-thumb method,

6 Choosing x^5 as the weighting function, we can derive the smoothing parameter as

$$\arg\min_b \left\{ b^4 \int_0^\infty \tilde{q}_{MG}^2(x)\, x^5 dx + \frac{1}{2\sqrt{\pi} n b^{1/2}} \int_0^\infty \frac{g(x)}{\sqrt{x}} x^5 dx \right\} = \left\{ \frac{4^\alpha \beta^{9/2} \Gamma(\alpha + 9/2)\, \Gamma(\alpha)}{4\sqrt{\pi}(\alpha - 1)^2 (\alpha - 2)^2 \Gamma(2\alpha)} \right\}^{2/9} n^{-2/9}.$$

Table 4 (*continued*)

Panel (b): $n = 200$

Estimator		Distribution 1		Distribution 2		Distribution 3		Distribution 4		Distribution 5	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.0319	0.0306	0.0319	0.0314	0.0278	0.0291	0.0261	0.0271	0.0592	0.0424
	SD	(0.0094)	(0.0097)	(0.0103)	(0.0100)	(0.0092)	(0.0096)	(0.0087)	(0.0091)	(0.0109)	(0.0112)
	IAB	*0.0140*	*0.0099*	*0.0146*	*0.0101*	*0.0105*	*0.0084*	*0.0106*	*0.0082*	*0.0272*	*0.0159*
BU-MG	RISE: Ave	0.0309	0.0303	0.0302	0.0305	0.0288	0.0309	0.0263	0.0283	0.0523	0.0362
	SD	(0.0086)	(0.0091)	(0.0093)	(0.0093)	(0.0106)	(0.0108)	(0.0100)	(0.0103)	(0.0123)	(0.0121)
	IAB	*0.0141*	*0.0098*	*0.0144*	*0.0097*	*0.0076*	*0.0060*	*0.0079*	*0.0059*	*0.0306*	*0.0164*
TS-G	RISE: Ave	0.0366	0.0309	0.0372	0.0313	0.0225	0.0285	0.0205	0.0258	0.0820	0.0403
	SD	(0.0087)	(0.0102)	(0.0109)	(0.0104)	(0.0100)	(0.0111)	(0.0092)	(0.0104)	(0.0094)	(0.0110)
	IAB	*0.0144*	*0.0052*	*0.0171*	*0.0048*	*0.0093*	*0.0035*	*0.0084*	*0.0031*	*0.0351*	*0.0102*
TS-MG	RISE: Ave	0.0375	0.0317	0.0357	0.0314	0.0284	0.0319	0.0280	0.0292	0.0837	0.0339
	SD	(0.0057)	(0.0093)	(0.0078)	(0.0098)	(0.0113)	(0.0126)	(0.0101)	(0.0121)	(0.0090)	(0.0111)
	IAB	*0.0194*	*0.0062*	*0.0189*	*0.0055*	*0.0134*	*0.0044*	*0.0161*	*0.0043*	*0.0500*	*0.0106*
JLN-G	RISE: Ave	0.0328	0.0295	0.0312	0.0286	0.0244	0.0249	0.0217	0.0229	0.0686	0.0468
	SD	(0.0097)	(0.0100)	(0.0111)	(0.0105)	(0.0104)	(0.0106)	(0.0097)	(0.0101)	(0.0114)	(0.0102)
	IAB	*0.0129*	*0.0084*	*0.0145*	*0.0102*	*0.0092*	*0.0078*	*0.0080*	*0.0057*	*0.0284*	*0.0172*
JLN-MG	RISE: Ave	0.0332	0.0315	0.0319	0.0302	0.0234	0.0246	0.0205	0.0230	0.0766	0.0534
	SD	(0.0064)	(0.0075)	(0.0073)	(0.0081)	(0.0105)	(0.0108)	(0.0097)	(0.0103)	(0.0101)	(0.0100)
	IAB	*0.0099*	*0.0065*	*0.0096*	*0.0069*	*0.0038*	*0.0033*	*0.0026*	*0.0019*	*0.0361*	*0.0214*

Estimator		Distribution 6		Distribution 7		Distribution 8		Distribution 9		Distribution 10	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.1155	0.0780	0.0572	0.0498	0.0425	0.0301	0.0571	0.0445	0.0668	0.0492
	SD	(0.0391)	(0.0236)	(0.0156)	(0.0156)	(0.0104)	(0.0098)	(0.0184)	(0.0124)	(0.0250)	(0.0139)
	IAB	*0.0366*	*0.0227*	*0.0189*	*0.0122*	*0.0325*	*0.0125*	*0.0298*	*0.0196*	*0.0291*	*0.0175*
BU-MG	RISE: Ave	0.1018	0.0649	0.0539	0.0478	0.0403	0.0297	0.0551	0.0442	0.0598	0.0458
	SD	(0.0401)	(0.0194)	(0.0143)	(0.0148)	(0.0105)	(0.0097)	(0.0181)	(0.0135)	(0.0221)	(0.0128)
	IAB	*0.0482*	*0.0279*	*0.0217*	*0.0136*	*0.0302*	*0.0117*	*0.0338*	*0.0213*	*0.0287*	*0.0181*
TS-G	RISE: Ave	0.1298	0.0677	0.0681	0.0495	0.0558	0.0301	0.0712	0.0426	0.0677	0.0422
	SD	(0.0445)	(0.0275)	(0.0133)	(0.0163)	(0.0091)	(0.0101)	(0.0191)	(0.0127)	(0.0297)	(0.0133)
	IAB	*0.0417*	*0.0184*	*0.0215*	*0.0074*	*0.0472*	*0.0048*	*0.0396*	*0.0158*	*0.0311*	*0.0139*
TS-MG	RISE: Ave	0.1706	0.1041	0.0725	0.0489	0.0502	0.0303	0.0728	0.0432	0.0951	0.0605
	SD	(0.0395)	(0.0237)	(0.0110)	(0.0147)	(0.0091)	(0.0100)	(0.0235)	(0.0144)	(0.0295)	(0.0156)
	IAB	*0.0942*	*0.0451*	*0.0310*	*0.0089*	*0.0409*	*0.0044*	*0.0501*	*0.0175*	*0.0597*	*0.0233*
JLN-G	RISE: Ave	0.1121	0.0638	0.0622	0.0580	0.0505	0.0289	0.0694	0.0533	0.0609	0.0427
	SD	(0.0484)	(0.0197)	(0.0166)	(0.0161)	(0.0079)	(0.0085)	(0.0238)	(0.0116)	(0.0308)	(0.0096)
	IAB	*0.0351*	*0.0180*	*0.0180*	*0.0158*	*0.0343*	*0.0150*	*0.0341*	*0.0243*	*0.0251*	*0.0155*
JLN-MG	RISE: Ave	0.1453	0.0905	0.0601	0.0574	0.0428	0.0265	0.0632	0.0502	0.0743	0.0552
	SD	(0.0481)	(0.0178)	(0.0110)	(0.0103)	(0.0068)	(0.0073)	(0.0220)	(0.0130)	(0.0313)	(0.0079)
	IAB	*0.0520*	*0.0280*	*0.0176*	*0.0154*	*0.0283*	*0.0131*	*0.0338*	*0.0233*	*0.0248*	*0.0194*

(*continued on next page*)

and improvement in performance measures by the former is often substantial. In particular, a smaller IAB by the gamma-referenced method can be attributed to the fact that on average it yields a smaller smoothing parameter value (although the results are unreported). Therefore, we may safely concentrate on the results from the gamma-referenced method hereafter.

Table 4 reveals that overall MBC works. For each combination of sample size and distribution, MBC estimators at least tend to deliver smaller IABs than two BU ones. The results suggest that there is no uniformly superior MBC estimator; rather, which MBC estimator performs best depends on distributions.

It is hard to judge whether for a given MBC estimator, the MG kernel improves performance measures over the G kernel, whereas it appears to be advantageous to employ the former when no MBC is made. In particular, TS-MG often performs most poorly among all six estimators. Its inferior performance can be attributed to the following two respects. First, TS-MBC estimation relies on two smoothing parameters b and b/c. Controlling both b and b/c is a cumbersome task. Because $0 < c < 1$, the density estimator using b/c tends to be oversmoothed, which is potentially a source of a large bias in every TS-MBC estimator. On the other hand, if we make b too short in order to have a reasonable length of b/c, additional variability is introduced to the other estimator using b due to undersmoothing. Second, when the MG kernel is employed, these two smoothing parameters also play a role of determining the boundary region explicitly (e.g. $[0, 2b)$ for the density estimator using b). Unless b is short enough, there is a relatively small interior region for the density estimator using b/c. This aspect is also thought to worsen the performance measures of TS-MG, in particular for small to medium sample sizes. In conclusion, when TS-MBC estimation is applied, it is desirable to use the smoothing parameter choice method that tends to provide a small value consistently. Our experiment indicates that for a given distribution and a given sample size, $\hat{b}_{GR-TS}$ on average yields a smaller value than $\hat{b}_{ROT-MBC}$ does, which explains why the former is preferable in general for TS-MBC estimation over the latter.

Table 4 (*continued*)

Panel (c): $n = 500$

Estimator		Distribution 1		Distribution 2		Distribution 3		Distribution 4		Distribution 5	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.0247	0.0231	0.0238	0.0229	0.0202	0.0212	0.0190	0.0197	0.0457	0.0313
	SD	(0.0066)	(0.0066)	(0.0072)	(0.0068)	(0.0060)	(0.0062)	(0.0058)	(0.0060)	(0.0079)	(0.0080)
	IAB	*0.0103*	*0.0073*	*0.0108*	*0.0075*	*0.0075*	*0.0060*	*0.0076*	*0.0058*	*0.0208*	*0.0119*
BU-MG	RISE: Ave	0.0230	0.0219	0.0215	0.0216	0.0209	0.0223	0.0191	0.0205	0.0380	0.0263
	SD	(0.0058)	(0.0060)	(0.0061)	(0.0061)	(0.0069)	(0.0070)	(0.0066)	(0.0067)	(0.0091)	(0.0087)
	IAB	*0.0103*	*0.0071*	*0.0100*	*0.0066*	*0.0056*	*0.0044*	*0.0058*	*0.0044*	*0.0222*	*0.0119*
TS-G	RISE: Ave	0.0316	0.0230	0.0307	0.0219	0.0160	0.0195	0.0146	0.0177	0.0731	0.0304
	SD	(0.0065)	(0.0070)	(0.0077)	(0.0069)	(0.0065)	(0.0069)	(0.0062)	(0.0065)	(0.0070)	(0.0078)
	IAB	*0.0119*	*0.0043*	*0.0142*	*0.0041*	*0.0072*	*0.0024*	*0.0066*	*0.0022*	*0.0305*	*0.0084*
TS-MG	RISE: Ave	0.0307	0.0231	0.0278	0.0218	0.0201	0.0218	0.0199	0.0200	0.0727	0.0234
	SD	(0.0036)	(0.0059)	(0.0052)	(0.0062)	(0.0071)	(0.0079)	(0.0065)	(0.0076)	(0.0070)	(0.0079)
	IAB	*0.0152*	*0.0049*	*0.0150*	*0.0039*	*0.0097*	*0.0031*	*0.0116*	*0.0031*	*0.0422*	*0.0076*
JLN-G	RISE: Ave	0.0266	0.0228	0.0239	0.0210	0.0171	0.0173	0.0152	0.0159	0.0576	0.0366
	SD	(0.0071)	(0.0069)	(0.0077)	(0.0072)	(0.0065)	(0.0066)	(0.0063)	(0.0063)	(0.0080)	(0.0070)
	IAB	*0.0098*	*0.0063*	*0.0111*	*0.0077*	*0.0066*	*0.0054*	*0.0056*	*0.0038*	*0.0232*	*0.0133*
JLN-MG	RISE: Ave	0.0276	0.0247	0.0244	0.0223	0.0164	0.0172	0.0144	0.0161	0.0658	0.0421
	SD	(0.0040)	(0.0047)	(0.0045)	(0.0050)	(0.0067)	(0.0069)	(0.0063)	(0.0066)	(0.0078)	(0.0075)
	IAB	*0.0080*	*0.0053*	*0.0074*	*0.0052*	*0.0028*	*0.0024*	*0.0020*	*0.0014*	*0.0303*	*0.0168*
Estimator		**Distribution 6**		**Distribution 7**		**Distribution 8**		**Distribution 9**		**Distribution 10**	
		ROT	GR	ROT	GR	ROT	GR	ROT	GR	ROT	GR
BU-G	RISE: Ave	0.0983	0.0587	0.0443	0.0369	0.0313	0.0217	0.0461	0.0330	0.0540	0.0370
	SD	(0.0326)	(0.0154)	(0.0111)	(0.0110)	(0.0072)	(0.0065)	(0.0141)	(0.0090)	(0.0188)	(0.0097)
	IAB	*0.0308*	*0.0170*	*0.0143*	*0.0090*	*0.0237*	*0.0088*	*0.0243*	*0.0145*	*0.0228*	*0.0126*
BU-MG	RISE: Ave	0.0824	0.0452	0.0394	0.0336	0.0297	0.0214	0.0439	0.0323	0.0472	0.0332
	SD	(0.0335)	(0.0115)	(0.0101)	(0.0099)	(0.0072)	(0.0065)	(0.0137)	(0.0091)	(0.0162)	(0.0089)
	IAB	*0.0393*	*0.0196*	*0.0157*	*0.0093*	*0.0220*	*0.0082*	*0.0272*	*0.0155*	*0.0232*	*0.0133*
TS-G	RISE: Ave	0.1228	0.0544	0.0607	0.0370	0.0459	0.0207	0.0651	0.0324	0.0598	0.0336
	SD	(0.0394)	(0.0183)	(0.0097)	(0.0116)	(0.0063)	(0.0066)	(0.0155)	(0.0088)	(0.0239)	(0.0090)
	IAB	*0.0391*	*0.0147*	*0.0183*	*0.0060*	*0.0389*	*0.0033*	*0.0365*	*0.0127*	*0.0269*	*0.0117*
TS-MG	RISE: Ave	0.1629	0.0889	0.0599	0.0352	0.0414	0.0208	0.0654	0.0313	0.0868	0.0499
	SD	(0.0368)	(0.0164)	(0.0074)	(0.0093)	(0.0061)	(0.0066)	(0.0188)	(0.0094)	(0.0249)	(0.0110)
	IAB	*0.0896*	*0.0372*	*0.0246*	*0.0067*	*0.0342*	*0.0030*	*0.0452*	*0.0136*	*0.0541*	*0.0196*
JLN-G	RISE: Ave	0.1040	0.0499	0.0526	0.0479	0.0400	0.0212	0.0619	0.0430	0.0524	0.0352
	SD	(0.0428)	(0.0133)	(0.0119)	(0.0115)	(0.0055)	(0.0057)	(0.0195)	(0.0080)	(0.0244)	(0.0068)
	IAB	*0.0325*	*0.0138*	*0.0146*	*0.0125*	*0.0270*	*0.0113*	*0.0310*	*0.0198*	*0.0219*	*0.0133*
JLN-MG	RISE: Ave	0.1370	0.0740	0.0508	0.0475	0.0332	0.0190	0.0560	0.0403	0.0662	0.0474
	SD	(0.0438)	(0.0115)	(0.0073)	(0.0068)	(0.0044)	(0.0049)	(0.0170)	(0.0085)	(0.0249)	(0.0053)
	IAB	*0.0486*	*0.0217*	*0.0142*	*0.0122*	*0.0219*	*0.0094*	*0.0304*	*0.0191*	*0.0227*	*0.0175*

Note: "ROT" and "GR" in column headings denote "rule-of-thumb" and "gamma-referenced" smoothing parameter choice methods. "Ave" and "SD" are simulation averages and standard deviations of RISEs.

Table 5
Summary statistics of real data.

Variable	Mean	SD	Min	Max	Percentiles		
					25%	50%	75%
Per Capita Income (in USD)	3,790	4,156	224	25,646	744	2,005	5,817
Belgian Female Wage (in BEF)	414	154	88	1,364	314	384	483
US Male Wage (in USD)	958	404	115	3,078	668	905	1,160

4. Empirical illustrations

This section applies the MBC density estimators to real data sets. Our focus is on income and wage data. They have a natural boundary at the origin, and their distributions are empirically characterized by concentration of observations near the boundary and a long tail with sparse data. Hence, we are motivated to estimate the densities using asymmetric kernels.

The data we use include: (a) per capita income of 114 countries (in US dollars); (b) gross hourly wage rate for 579 Belgian females (in Belgian francs); and (c) monthly earnings of 935 US males (in US dollars). The first and third data are originally used in Romer (1993) and Blackburn and Neumark (1992), and these data sets are now available under the names "`openness`" and "`wage2`" as supplemental materials for Wooldridge (2013). The second is taken from the data set "`bwage`", a supplemental material for Verbeek (2012). Table 5 reports summary statistics of the data.

The density of each data is estimated by BU-G, TS-G and JLN-G. The data are first converted to the scale-adjusted ones by dividing them by either 10^3 or 10^4, and then the resulting density estimates are back-transformed to the ones in the

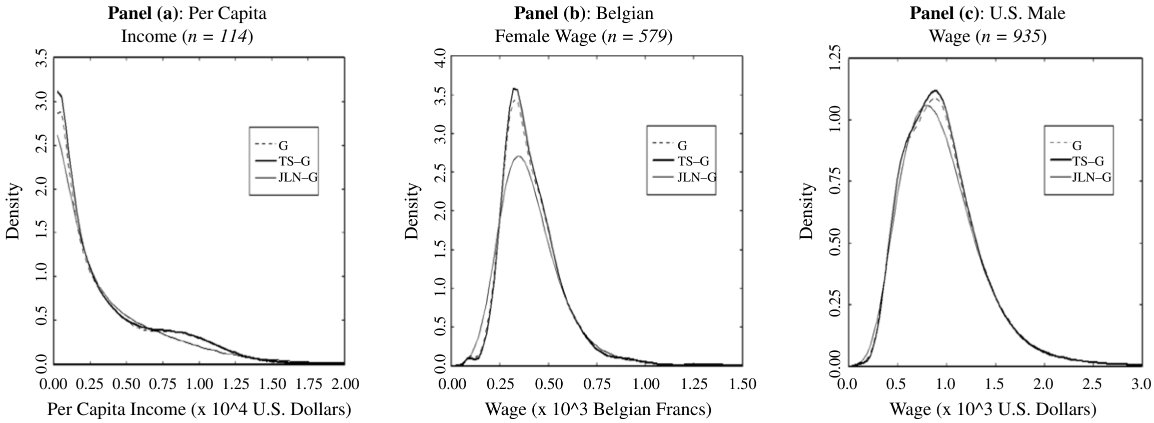

Fig. 1. Density estimates from real data.

original scale.[7] Plug-in smoothing parameters $\hat{b}_{GR-BU}$, $\hat{b}_{GR-TS}$ and $\hat{b}_{GR-JLN}$ are employed for BU-G, TS-G and JLN-G, respectively. Their values (based on the scale-adjusted data) are: $\left(\hat{b}_{GR-BU}, \hat{b}_{GR-TS}, \hat{b}_{GR-JLN}\right) = (0.0434, 0.0655, 0.1752)$ for (a); $(0.0041, 0.0054, 0.0395)$ for (b); and $(0.0105, 0.0152, 0.0677)$ for (c). The plots of density estimates are presented in Fig. 1. We can observe that while estimates from BU-G and TS-G largely look alike, the one from JLN-G is considerably different. Hirukawa (2010) reports the tendency of JLN-MBC that when the BU estimator underestimates (overestimates) the density, its corresponding JLN-MBC estimator corrects the estimate in an upward (downward) direction. Fig. 1 indicates that JLN-G indeed exhibits such smooth-out tendencies, and as a result it substantially smooths away the part that BU-G estimates wiggly.

5. Concluding remarks

This paper has demonstrated that two MBC techniques studied in Hirukawa (2010) for density estimation with support on $[0, 1]$ can be extended to density estimation with support on $[0, \infty)$ using asymmetric kernels. Under sufficient smoothness of the true density, both bias reduction methods are shown to improve the order of magnitude in bias from $O(b)$ to $O\left(b^2\right)$, while the order of magnitude in variance remains unchanged. Two classes of MBC density estimators are by construction nonnegative, and establish a faster convergence rate of $O\left(n^{-8/9}\right)$ in MSE for the interior part when best implemented. Monte Carlo simulations confirm bias reduction via two MBC methods.

While this paper deals exclusively with the univariate case, it appears to be possible to extend the analysis to joint density estimation. In the multivariate ("MV") case, the MV-BU estimator of the density f with support on $\mathbb{R}_+^d$ using a random sample $\{\mathbf{X}_i\}_{i=1}^n = \left\{(X_{1i}, \ldots, X_{di})^\top\right\}_{i=1}^n$ is $\hat{f}_{j,b}(x) = n^{-1}\sum_{i=1}^n \mathbb{K}_{j(\mathbf{x},b)}(\mathbf{X}_i)$, where $\mathbb{K}_{j(\mathbf{x},b)}(\mathbf{u}) := \prod_{\ell=1}^d K_{j(x_\ell,b)}(u_\ell)$ is the product asymmetric kernel given a design point $\mathbf{x} = (x_1, \ldots, x_d)^\top$ and a smoothing parameter $b_1 = \cdots = b_d = b$. If f has four continuous and bounded partial derivatives and $b + 1/\left\{nb^{d(r_j+1/2)+2}\right\} \to 0$, then it can be shown that each of the MV-TS-MBC estimator $\tilde{f}_{TS,j}(\mathbf{x}) = \left\{\hat{f}_{j,b}(\mathbf{x})\right\}^{1/(1-c)}\left\{\hat{f}_{j,b/c}(\mathbf{x})\right\}^{-c/(1-c)}$ for some $c \in (0, 1)$ and the MV-JLN-MBC estimator

$$\tilde{f}_{JLN,j}(\mathbf{x}) = \hat{f}_{j,b}(\mathbf{x})\left\{\frac{1}{n}\sum_{i=1}^n \frac{\mathbb{K}_{j(\mathbf{x},b)}(\mathbf{X}_i)}{\hat{f}_{j,b}(\mathbf{X}_i)}\right\}$$

admits the expansion $f(x) + O\left\{b^2 + n^{-1/2}\prod_{\ell=1}^d b^{-(1/2)\left(1/2+r_j\mathbf{1}_\ell\right)}\right\}$, where $\mathbf{1}_\ell := \mathbf{1}(x_\ell/b \to \kappa_\ell > 0)$ is the indicator function that takes 1 if x_ℓ lies in the boundary region. Analytical expressions of leading bias and variance terms and their finite sample performance will be addressed in a separate paper.

Acknowledgements

We would like to thank the Associate Editor, two anonymous referees, Benedikt Funke, Matthias Hagmann, Susumu Imai, James MacKinnon and seminar participants at Queen's University and Development Bank of Japan for insightful comments and suggestions. The first author gratefully acknowledges financial support from Japan Society for the Promotion of Science (grant number 23530259). The views expressed herein and those of the authors do not necessarily reflect the views of the Development Bank of Japan.

[7] See the horizontal axis of each panel in Fig. 1 for details.

Appendix A. Sketches of the proofs

The proof of each theorem requires kernel-specific arguments, which include Taylor expansions and properties of the random variable corresponding to the kernel. To save space, we provide only outline proofs when the Gamma kernel is employed. Full-length proofs are available upon request.

Outline Proof of Theorem 1. For the bias, it follows from the proof of Theorem 1 in Hirukawa (2010) and Assumption 2 that

$$E\left\{\tilde{f}_{TS,G}(x)\right\} = f(x) + \frac{1}{c(1-c)}\left[\frac{1}{2}\left\{\frac{a_{1,G}^2(x,f)}{f(x)}\right\} - a_{2,G}(x,f)\right]b^2 + o\left(b^2\right).$$

For the variance, recognize that

$$\operatorname{Var}\left\{\tilde{f}_{TS,G}(x)\right\} = \frac{1}{(1-c)^2}\left[\operatorname{Var}\left\{\hat{f}_{G,b}(x)\right\} - 2c\operatorname{Cov}\left\{\hat{f}_{G,b}(x), \hat{f}_{G,b/c}(x)\right\} + c^2\operatorname{Var}\left\{\hat{f}_{G,b/c}(x)\right\}\right] + O\left(n^{-1}\right).$$

Then, approximate each term as in the proof of Theorem 1 in Hirukawa (2010). ■

Outline Proof of Theorem 2. For the bias, write $h(x) = a_{1,G}(x,f)/f(x)$. It follows from the procedures in Section A.2.1 of Hirukawa (2010, pp. 490–491) that

$$E\left\{\tilde{f}_{JLN,G}(x)\right\} = f(x) - f(x)\left\{h'(x) + \frac{x}{2}h''(x)\right\}b^2 + o\left(b^2\right) = f(x) - f(x)\,a_{1,G}(x,h)\,b^2 + o\left(b^2\right).$$

For the variance, the procedures in Section A.2.2 of Hirukawa (2010, p. 492) imply that

$$\tilde{f}_{JLN,G}(x) \sim f(x)\frac{1}{n}\sum_{i=1}^{n}\frac{K_{G(x/b+1,b)}(X_i)}{f(X_i)}\left\{2 - \frac{\hat{f}_{G,b}(X_i)}{f(X_i)}\right\}.$$

By Stirling's approximation to the gamma function inside the integral and the trimming argument in Chen (2000), the leading variance on the right-hand side can be approximated by $n^{-1}A_b(x)f(x)$, where

$$A_b(x) = \frac{b^{-1}\Gamma(2x/b+1)}{2^{2x/b+1}\Gamma^2(x/b+1)} \sim \begin{cases} \dfrac{b^{-1/2}}{2\sqrt{\pi}x^{1/2}} & \text{if } x/b \to \infty \\[2ex] \dfrac{\Gamma(2\kappa+1)\,b^{-1}}{2^{2\kappa+1}\Gamma^2(\kappa+1)} & \text{if } x/b \to \kappa. \end{cases}$$

Then, the result immediately follows. ■

Appendix B. Formulae for gamma-referenced smoothing parameters

The analytical expression of $\hat{b}_{GR-TS}$ is

$$\hat{b}_{GR-TS} = \left\{c^2(1-c)^2\lambda(c)\right\}^{2/9}\left\{\frac{4^{\alpha}\beta^{9/2}\Gamma(\alpha+9/2)\,\Gamma(\alpha)}{16\sqrt{\pi}C_{TS}(\alpha)\,\Gamma(2\alpha)}\right\}^{2/9} n^{-2/9},$$

where

$$\begin{aligned} C_{TS}(\alpha) = {} & \frac{1}{36}(\alpha-2)^2\left(\alpha-\frac{3}{2}\right)^2(\alpha-1)^2 - \frac{1}{6}(\alpha-2)\left(\alpha-\frac{3}{2}\right)(\alpha-1)^2\alpha\left(\alpha+\frac{1}{2}\right) \\ & + \frac{1}{9}(\alpha-2)\left(\alpha-\frac{3}{2}\right)(\alpha-1)\alpha\left(\alpha+\frac{1}{2}\right)(\alpha+1) + \frac{1}{4}(\alpha-1)^2\alpha\left(\alpha+\frac{1}{2}\right)(\alpha+1)\left(\alpha+\frac{3}{2}\right) \\ & - \frac{1}{3}(\alpha-1)\alpha\left(\alpha+\frac{1}{2}\right)(\alpha+1)\left(\alpha+\frac{3}{2}\right)(\alpha+2) + \frac{1}{9}\alpha\left(\alpha+\frac{1}{2}\right)(\alpha+1)\left(\alpha+\frac{3}{2}\right)(\alpha+2)\left(\alpha+\frac{5}{2}\right). \end{aligned}$$

On the other hand, $\hat{b}_{GR-JLN}$ takes a much simpler form. It is given by

$$\hat{b}_{GR-JLN} = \left\{\frac{4^{\alpha}\beta^{5/2}\Gamma(\alpha+1/2)\,\Gamma(\alpha)}{4\sqrt{\pi}\Gamma(2\alpha)}\right\}^{2/9} n^{-2/9}.$$

Moreover, the gamma-referenced smoothing parameter for $\hat{f}_{MG}(x)$ is defined as

$$\begin{aligned} \hat{b}_{GR-BU} &= \arg\min_b AWMISE\left\{\hat{f}_{MG}(x)\right\} \\ &= \arg\min_b \left\{\frac{b^2}{4}\int_0^{\infty} x^2\left\{g''(x)\right\}^2 w(x)\,dx + \frac{1}{2\sqrt{\pi}nb^{1/2}}\int_0^{\infty}\frac{g(x)}{\sqrt{x}}w(x)\,dx\right\}, \end{aligned}$$

where the weighting function $w(x)$ is chosen as $w(x) = x^3$ to ensure finiteness of integrals. It follows that $\hat{b}_{GR-BU}$ can be expressed as

$$\hat{b}_{GR-BU} = \left\{ \frac{4^{\alpha} \beta^{5/2} \Gamma(\alpha + 5/2) \Gamma(\alpha)}{8\sqrt{\pi} C_{BU}(\alpha) \Gamma(2\alpha)} \right\}^{2/5} n^{-2/5},$$

where

$$C_{BU}(\alpha) = \frac{1}{4}(\alpha - 2)^2 (\alpha - 1)^2 - (\alpha - 2)(\alpha - 1)^2 \alpha + \frac{1}{2}(\alpha - 1)(3\alpha - 4)\alpha \left(\alpha + \frac{1}{2}\right) - (\alpha - 1)\alpha \left(\alpha + \frac{1}{2}\right)(\alpha + 1) + \frac{1}{4}\alpha \left(\alpha + \frac{1}{2}\right)(\alpha + 1)\left(\alpha + \frac{3}{2}\right). \quad \blacksquare$$

References

Blackburn, M., Neumark, D., 1992. Unobserved ability, efficiency wages, and interindustry wage differentials. Quart. J. Econ. 107, 1421–1436.
Chen, S.X., 1999. Beta kernel estimators for density functions. Comput. Statist. Data Anal. 31, 131–145.
Chen, S.X., 2000. Probability density function estimation using gamma kernels. Ann. Inst. Statist. Math. 52, 471–480.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer, New York, NY.
Gustafsson, J., Hagmann, M., Nielsen, J.P., Scaillet, O., 2009. Local transformation kernel density estimation of loss distributions. J. Bus. Econom. Statist. 27, 161–175.
Hagmann, M., Scaillet, O., 2007. Local multiplicative bias correction for asymmetric kernel density estimators. J. Econometrics 141, 213–249.
Hirukawa, M., 2010. Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval. Comput. Statist. Data Anal. 54, 473–495.
Jin, X., Kawczak, J., 2003. Birnbaum–Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. Ann. Econ. Finance 4, 103–124.
Jones, M.C., Linton, O., Nielsen, J.P., 1995. A simple bias reduction method for density estimation. Biometrika 82, 327–338.
Jones, M.C., Signorini, D.F., 1997. A comparison of higher-order bias kernel density estimators. J. Amer. Statist. Assoc. 92, 1063–1073.
Karunamuni, R.J., Alberts, T., 2005. On boundary correction in kernel density estimation. Stat. Methodol. 2, 191–212.
Leblanc, A., 2010. A bias-reduced approach to density estimation using Bernstein polynomials. J. Nonparametr. Stat. 22, 459–475.
Romer, D., 1993. Openness and inflation: theory and evidence. Quart. J. Econ. 108, 869–903.
Scaillet, O., 2004. Density estimation using inverse and reciprocal inverse Gaussian kernels. J. Nonparametr. Stat. 16, 217–226.
Stone, C.J., 1980. Optimal rates of convergence for nonparametric estimators. Ann. Statist. 8, 1348–1360.
Stuetzle, W., Mittal, Y., 1979. Some comments on the asymptotic behavior of robust smoothers. In: Gasser, T., Rosenblatt, M. (Eds.), Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, 2–4 April, 1979. Springer-Verlag, Berlin, pp. 191–195.
Terrell, G.R., Scott, D.W., 1980. On improving convergence rates for nonnegative kernel density estimators. Ann. Statist. 8, 1160–1163.
Verbeek, M., 2012. A Guide to Modern Econometrics, fourth ed. John Wiley & Sons, Chichester, UK.
Wooldridge, J.M., 2013. Introductory Econometrics: A Modern Approach, fifth ed. South-Western Cengage Learning, Mason, OH.