Econometrics and Statistics 20 (2021) 109–130

Contents lists available at ScienceDirect

Econometrics and Statistics

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

Bias correction for local linear regression estimation using asymmetric kernels via the skewing method

Benedikt Funke[a], Masayuki Hirukawa[b,*]

[a] Department of Mathematics, Technical University of Dortmund, Vogelpothsweg 87, Dortmund 44227, Germany
[b] Faculty of Economics, Ryukoku University, 67 Tsukamoto-cho, Fukakusa, Fushimi-ku, Kyoto 612-8577, Japan

A R T I C L E I N F O

Article history:
Received 27 March 2019
Revised 27 January 2020
Accepted 28 January 2020
Available online 17 February 2020

JEL classification:
C13
C14

Keywords:
Beta kernel
Bias correction
Boundary bias
Curve estimation
Gamma kernel
Local linear regression estimation

A B S T R A C T

The skewing method, which has been originally proposed as a bias correction device for local linear regression estimation using standard symmetric kernels, is extended to the cases of asymmetric kernels. The method is defined as a convex combination of three local linear estimators. It is demonstrated that the skewed estimator using asymmetric kernels with properly chosen weights can accelerate the bias convergence from $O(b)$ to $O(b^2)$ as $b \to 0$ under sufficient smoothness of the unknown regression curve while not inflating the variance in an order of magnitude, where b is the smoothing parameter and the regressor is assumed to have at least one boundary. As a consequence, the estimator has optimal pointwise convergence of $n^{-4/9}$ when best implemented, where n is the sample size. It is noteworthy that these properties are the same as those for a local cubic regression estimator. Finite-sample properties of the skewed estimator are assessed in comparison with local linear and local cubic estimators. An application of the skewed estimation to real data is also considered.

1. Introduction

The aim of this paper is to reduce the order of magnitude in bias of local linear ("LL") regression estimators (Fan and Gijbels, 1992; Fan, 1993) smoothed by asymmetric kernels while not inflating the one in variance. The key trick for accelerating the bias convergence is the skewing method, which has been proposed originally for standard nonnegative symmetric kernels by Choi and Hall (1998). The method is defined as a convex combination of three LL estimators, namely, one standard (i.e., centered) and two shifted (i.e., left- and right-shifted) LL estimators.

Before proceeding, we should explain why we concentrate on a nonstandard smoothing technique using asymmetric kernels. Throughout it is assumed that the support of the regressor has at least one boundary (e.g., [0, 1] or $\mathbb{R}_+$). Indeed, there are many economic and financial variables that have a natural boundary at the origin; examples include wages, incomes, short-term interest rates, and insurance claims (or financial losses), to name a few. Often the variables also have right-skewed distributions. Asymmetric kernels are expected to work well for such distributions because of their freedom of boundary bias and adaptive smoothing property via changing shapes automatically across design points. Among all asymmetric kernels proposed so far, our main focuses are on the beta and gamma kernels proposed by Chen (1999, 2000b),

* Corresponding author.
E-mail addresses: benedikt.funke@mathematik.tu-dortmund.de (B. Funke), hirukawa@econ.ryukoku.ac.jp (M. Hirukawa).

https://doi.org/10.1016/j.ecosta.2020.01.004

respectively.[1] Our preference is based on empirical relevance of the kernels in economics and finance; see Table 1.1 of Hirukawa (2018) for a variety of applications of the beta and gamma kernels to estimation and testing problems.

Up until recently, the bias reduction in asymmetric kernel smoothing has been actively studied within the framework of density estimation. In fact, many authors have examined different kinds of bias correction methods that can reduce the order of magnitude in bias from $O(b)$ to $O(b^2)$ as $b \to 0$ while the one in variance is left unchanged, where b is the smoothing parameter. Interested readers may consult, for instance, Hirukawa (2010), Hirukawa and Sakudo (2014, 2015), and Funke and Kawka (2015). Hirukawa (2018, Chapter 3) also provides a survey. On the other hand, little is investigated on nonparametric regression estimation smoothed by asymmetric kernels. A few exceptions are Chen (2000a, 2002) as theoretical contributions and Gospodinov and Hirukawa (2012) as an application.[2] Specifically, Chen (2000a) examines the Gasser-Müller type regression estimator (Gasser and Müller, 1979) smoothed by the beta kernel. Subsequently, Chen (2002) explores statistical properties of LL estimators smoothed by the beta and gamma kernels. Moreover, Gospodinov and Hirukawa (2012) apply the local constant ("LC") regression estimation using the gamma kernel to nonparametric estimation of time-homogeneous drift and diffusion functions in continuous-time models that can be used to describe the underlying dynamics of spot interest rates.

However, no attempts have been made to improve the bias convergence in asymmetric kernel regression estimation, to the best of our knowledge. In this respect, our work may be viewed as a natural extension of Chen (2002) in the spirit of the aforementioned articles on bias reduction in density estimation. Following Choi and Hall (1998), we investigate two versions of the skewed estimator smoothed by asymmetric kernels. The first version is a convex combination of one standard and two shifted LL estimators. The second version, constructed as a convex combination of two shifted LL estimators only, can be viewed as a limiting form of the first version. It is demonstrated that under sufficient differentiability of the underlying regression curve, each version can accelerate the bias convergence from $O(b)$ to $O(b^2)$ while the order of magnitude in variance is preserved. As a consequence, when best implemented, the estimators have optimal pointwise convergence rate of $n^{-4/9}$ where n is the sample size. It is an improvement from the $n^{-2/5}$ rate in the usual case with no additional smoothing as in Chen (2002).

There are some advantages in incorporating asymmetric kernels into the skewing method. First, unlike Choi and Hall's (1998) original skewed estimator using symmetric kernels, it is possible to make the skewed estimator using asymmetric kernels free of edge effects. Details of the method are available in Remark 2. Second, asymmetric kernels possess the robustness property to sparse design problems. The kernels spread out as the design point moves away from a boundary, and as a result, they tend to collect more data points to smooth in the areas with fewer observations. The property continues to hold when asymmetric kernels are combined with the skewed estimation. This is documented in Theorem 2 and confirmed numerically in Section 3.

There are a few other approaches that can reduce the bias from $O(b)$ to $O(b^2)$ in nonparametric regression estimation. Examples include (i) local polynomial (e.g., local cubic) regression smoothing by Ruppert and Wand (1994) and (ii) the multiplicative bias correction method for nonparametric regression estimation by Linton and Nielsen (1994) and Jones et al. (1995, Section 4.2). Some readers may wonder why we do not instead investigate these. For (i), the reason why we do not turn to the local cubic estimation is that because higher-order polynomial estimators are subject to potential problems caused by sparse design, we must prepare for a guard against near-singularity in large matrices in practice, as argued in Choi and Hall (1998, Remark 2.3). For (ii), while the skewing method is a version of additive bias correction techniques, the multiplicative bias correction method is based on the identity $h(\cdot) \equiv g(\cdot)\{h(\cdot)/g(\cdot)\} := g(\cdot)r(\cdot)$, where $h(\cdot)$ is either the density or the conditional mean (i.e., the regression function), $g(\cdot)$ is its initial estimate, and $r(\cdot)$ serves as a correction factor. The method works well in density estimation as long as the underlying density is strictly positive at a given design point or on a certain interval, because in this case the initial density estimate $g(\cdot)$ is also likely to be strictly positive. On the other hand, the underlying regressand can take any real value. Therefore, its bias-corrected estimator is not well defined when the regression curve touches or passes through the horizontal axis.[3]

The remainder of this paper is organized as follows. Section 2 delivers convergence properties of the skewed estimators using asymmetric kernels, including the bias acceleration of the estimators. Section 3 conducts Monte Carlo simulations to examine finite-sample properties of the skewed estimators. Section 4 applies the skewed estimation to a US industry-level data and estimates a linear-homogenous production function nonparametrically. Section 5 summarizes the main results of the paper. All proofs are provided in the Appendix.

This paper adopts the following notational conventions: for $a > 0$, $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)dt$ is the gamma function; for $p, q > 0$, $B(p, q) = \int_0^1 y^{p-1}(1-y)^{q-1}dy$ denotes the beta function; and $\mathbf{1}\{\cdot\}$ signifies an indicator function. The expression '$X \overset{d}{=} Y$' 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 de-

[1] Extensions to the cases of other asymmetric kernels are straightforward and thus not pursued any further in this paper; see Section 1.2.2 of Hirukawa (2018) for a non-exhaustive list of asymmetric kernels.

[2] It is known that asymmetric kernel density estimators are not integrated to unity in general. This comes from the fact that roles of the data and design points in asymmetric kernels are not exchangeable. The lack of normalization in asymmetric kernels may cause a problem in some applications of density estimation, whereas it is not an issue when they are employed for regression estimation.

[3] Two ad hoc solutions to this problem are available in Glad (1998), actually. These are: (i) clipping the correction factor below 1/10 and above 10; and (ii) shifting all response data a distance a. Recently Scholz et al. (2015) have applied them to financial data.

sign point $x \in \mathbb{R}_+$ ($x \in [0, 1]$), we denote by "interior x" and "boundary x" a design point x that satisfies $x/b \to \infty$ (x/b and $(1-x)/b \to \infty$) and $x/b \to \kappa$ (x/b or $(1-x)/b \to \kappa$) for some $\kappa \in (0, \infty)$ as $n \to \infty$, respectively.

2. Estimators and Their Large-Sample Properties

2.1. The Estimator

Suppose that we are interested in estimating an unknown conditional mean

$$m(x) := E(Y|X = x)$$

nonparametrically using *i.i.d.* bivariate random variables $\{(Y_i, X_i)\}_{i=1}^n$. Throughout the corresponding nonparametric regression model

$$Y = m(X) + \epsilon$$

is examined, where ϵ is the regression error that satisfies $E(\epsilon|X) = 0$. Also denote the conditional variance of ϵ as $\sigma^2(\cdot) := E\left(\epsilon^2\middle|X = \cdot\right)$.

Often the economic and financial data have a boundary. Our particular focus is on the cases in which the support of the regressor X has at least one boundary. More specifically, supp(X) is assumed to be either $[0, 1]$ or $\mathbb{R}_+$. Accordingly, we employ asymmetric kernels to estimate $m(x)$ nonparametrically.

Our aim is to improve the bias property of the LL regression estimator of $m(x)$ smoothed by an asymmetric kernel while the order of magnitude in its variance remains unchanged. For this purpose, we find (α, β) that jointly minimize the local least squares problem

$$\sum_{i=1}^n \{Y_i - \alpha - \beta(X_i - x)\}^2 K_{j(x,b)}(X_i),$$

where $K_{j(x,b)}$ is some asymmetric kernel indexed by j that depends on the design point x and the smoothing parameter b. Let $\left(\hat{\alpha}_j(x), \hat{\beta}_j(x)\right)$ be the minimizer for a given x. It is well known that

$$\hat{\alpha}_j(x) = \frac{s_{j2}(x)t_{j0}(x) - s_{j1}(x)t_{j1}(x)}{s_{j0}(x)s_{j2}(x) - s_{j1}^2(x)} \text{ and}$$

$$\hat{\beta}_j(x) = \frac{s_{j0}(x)t_{j1}(x) - s_{j1}(x)t_{j0}(x)}{s_{j0}(x)s_{j2}(x) - s_{j1}^2(x)},$$

where

$$s_{jp}(x) = \frac{1}{n}\sum_{i=1}^n (X_i - x)^p K_{j(x,b)}(X_i) \text{ and} \tag{1}$$

$$t_{jp}(x) = \frac{1}{n}\sum_{i=1}^n Y_i (X_i - x)^p K_{j(x,b)}(X_i) \tag{2}$$

for $p = 0, 1, 2 \ldots$

The standard approach is to estimate $m(x)$ by the (classical) LL estimator $\hat{m}_j(x) = \hat{\alpha}_j(x)$. In what follows, a nonstandard LL estimator

$$\hat{m}_j(u|x) := \hat{\alpha}_j(x) + \hat{\beta}_j(x)(u - x) \tag{3}$$

is instead considered. The estimator $\hat{m}_j(u|x)$ can be interpreted as a shifted LL estimator in which smoothing is made at a slightly off-centered point, i.e., smoothing is made at x, which is located slightly to the left or right of the center u. Also observe that putting $u = x$ in (3) (i.e., the case in which the center coincides with x) reduces to $\hat{m}_j(x|x) = \hat{m}_j(x) = \hat{\alpha}_j(x)$, which is the classical LL estimator.

The skewed estimator is defined as a convex combination of three such LL estimators. Given weight parameters (called "weights" hereinafter) $\lambda_1, \lambda_2 > 0$ and shift parameters (called "shifts" hereinafter) $\ell_1, \ell_2 > 0$, the general form of the skewed estimator of $m(x)$ can be expressed as

$$\tilde{m}_{j,\lambda_1,\lambda_2,\ell_1,\ell_2}(x) := \frac{\lambda_1 \hat{m}_j\left(x|x - \ell_1 b^{1/2}\right) + \hat{m}_j(x|x) + \lambda_2 \hat{m}_j\left(x|x + \ell_2 b^{1/2}\right)}{\lambda_1 + 1 + \lambda_2}, \tag{4}$$

where $\hat{m}_j\left(x|x - \ell_1 b^{1/2}\right)$ and $\hat{m}_j\left(x|x + \ell_2 b^{1/2}\right)$ are called the left- and right-shifted LL estimators, respectively, hereinafter.

Unlike standard symmetric kernels, exploring statistical properties of nonparametric estimators smoothed by asymmetric kernels relies on kernel-specific and thus diversified approaches. The skewed estimator is not an exception. Then, to save space, we specialize in two asymmetric kernels, namely, the beta kernel ($j = B$)

$$K_{B(x,b)}(u) = \frac{u^{x/b}(1-u)^{(1-x)/b}}{B\{x/b+1, (1-x)/b+1\}}\mathbf{1}\{u \in [0,1]\}$$

for $x \in [0, 1]$ by Chen (1999) and the gamma kernel ($j = G$)

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

for $x \in \mathbb{R}_+$ by Chen (2000b), due to their popularity in empirical studies in economics and finance.

Furthermore, when x is located in the vicinity of a boundary, a certain choice of shifts (ℓ_1, ℓ_2) may preclude us from defining either the left- or right-shifted LL estimator. As a result, the skewed estimator is not well-defined, either. Therefore, the subsequent asymptotic results for boundary x should be interpreted as those on condition that the skewed estimator is well-defined; see Remark 2 for more details.

2.2. Convergence Properties of the Gamma Skewed Estimator

2.2.1. Regularity Conditions

We start from exploring convergence properties of the gamma skewed estimator; those of the beta skewed estimator will be delivered shortly as an extension. To develop their convergence properties, we impose the following common regularity conditions across the two estimators:

Assumption 1. The *i.i.d.* random sample $\{(Y_i, X_i)\}_{i=1}^n$ is drawn from a bivariate distribution having support on $\mathbb{R} \times \operatorname{supp}(X)$, where $\operatorname{supp}(X)$ is either $[0, 1]$ or $\mathbb{R}_+$.

Assumption 2. Let $f(\cdot)$ be the marginal pdf of X. Then, there is a neighborhood $\mathcal{N}$ around the design point $x \in \operatorname{supp}(X)$ such that $\sigma^2(\cdot)$, the second-order derivative of $f(\cdot)$ and the fourth-order derivative of $m(\cdot)$ are Hölder-continuous of order $\varsigma \in (0, 1]$ on $\mathcal{N}$. In addition, $f(x) > 0$ and $\sigma^2(x) < \infty$.

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

Assumption 1 refers to random sampling, which makes it easier to derive the dominant terms in bias and variance of the skewed estimators.[4] Fourth-order smoothness of the regression curve in Assumption 2 is crucial for bias acceleration of the skewed estimation. While only the continuity of $f(\cdot)$ is required for LC or LL estimation, the bias expansion developed below involves $f^{(2)}(\cdot)$ and thus its local Hölder-continuity in Assumption 2 is a key requirement. Local Hölder-continuity of $f^{(2)}(\cdot)$ in the neighborhood of x, for example, implies that there is a constant $C \in (0, \infty)$ such that

$$\left|f^{(2)}(u) - f^{(2)}(v)\right| \leq C|u-v|^{\varsigma},\ \forall u, v \in \mathcal{N}.$$

The condition "$\left(nb^3\right)^{-1} \to 0$" in Assumption 3 is imposed typically in the literature on bias reduction in asymmetric kernel density estimation (e.g., Hirukawa, 2010; Hirukawa and Sakudo, 2014, 2015). It is required to control the remainder term in the bias approximation of the skewed estimators, whereas a milder one "$\left(nb^2\right)^{-1} \to 0$" suffices for LC or LL estimation; see equation (2.3) of Chen (2002) for more details. It will be shown shortly that the MSE-optimal smoothing parameter for each skewed estimator becomes $b^* = O\left(n^{-2/9}\right)$ for interior x and $b^\dagger = O\left(n^{-1/5}\right)$ for boundary x (only if it is well-defined); these convergence rates are indeed within the required range.

2.2.2. Bias Approximation

Below we document the bias approximation of the gamma skewed estimator, conditional on $X = x$. In addition, as a by-product of the analysis, a second-order bias expansion (i.e., the bias approximation up to the $O(b^2)$ term) of the gamma LL estimator $\hat{m}_G(x|x) = \hat{m}_G(x)$ has been derived. The result is summarized in a separate corollary.

Theorem 1. *Suppose that for a given design point $x > 0$, weights (λ_1, λ_2) and shifts (ℓ_1, ℓ_2) in (4) are chosen as*

$$\lambda_1 = \lambda_2 = \lambda \text{ and } \ell_1 = \ell_2 = \ell = \ell(\lambda, x) = \sqrt{\left(\frac{2\lambda+1}{2\lambda}\right)x}. \tag{5}$$

Denote the gamma skewed estimator with the above weights and shifts plugged into (4) by

$$\tilde{m}_{G,\lambda}(x) = \frac{\lambda\hat{m}_G\left(x|x-\ell b^{1/2}\right) + \hat{m}_G(x|x) + \lambda\hat{m}_G\left(x|x+\ell b^{1/2}\right)}{2\lambda+1}.$$

[4] We do not consider the case in which residuals are correlated. However, it is conjectured that asymptotic results in this paper will continue to hold for weakly dependent and strictly stationary data by imposing some mixing condition (e.g., α-mixing with a suitable size), although an extension in this direction is beyond the scope of this paper.

If Assumptions 1-3 hold, then, as $n \to \infty$, the bias of $\tilde{m}_{G,\lambda}(x)$ can be approximated by $Bias\left\{\tilde{m}_{G,\lambda}(x)\right\} = \mathcal{B}_{G,\lambda}(x; m, f)b^2 + o\left(b^2\right)$, *where*

$$\mathcal{B}_{G,\lambda}(x; m, f) = -\left[\left\{1 + 2x\frac{f^{(1)}(x)}{f(x)} + \frac{x^2}{2}\frac{f^{(2)}(x)}{f(x)}\right\}m^{(2)}(x) + x\left\{\frac{5}{3} + x\frac{f^{(1)}(x)}{f(x)}\right\}m^{(3)}(x) + \frac{x^2}{4}\left(1 + \frac{1}{4\lambda}\right)m^{(4)}(x)\right].$$

Corollary 1. *Under Assumptions 1-3, as $n \to \infty$, the bias of $\hat{m}_G(x)$ can be expanded up to the second order as*

$$Bias\left\{\hat{m}_G(x)\right\} = \frac{x}{2}m^{(2)}(x)b + \left[-\left\{1 + x\frac{f^{(1)}(x)}{f(x)} - \frac{x^2}{2}\frac{f^{(2)}(x)}{f(x)} + x^2\left(\frac{f^{(1)}(x)}{f(x)}\right)^2\right\}m^{(2)}(x) + \frac{x}{3}m^{(3)}(x) + \frac{x^2}{8}m^{(4)}(x)\right]b^2 + o\left(b^2\right).$$

The theorem states that under the symmetric weights and shifts given in (5), the gamma skewed estimator can attain the bias convergence of $O(b^2)$ regardless of the position of the design point x (as long as $\tilde{m}_{G,\lambda}(x)$ is well-defined). Such weights and shifts are determined by setting coefficients on $O(b)$ and $O(b^{3/2})$ bias terms of (4) equal to zero; see the Proof of Theorem 1 in Appendix for more details. Observe that the bias acceleration in the skewed estimator can be attained by a fixed λ. In Section 2.3 below, we let $\lambda \to \infty$ and explore convergence properties of the skewed estimator in limit.

Chen (2002) shows that orders of magnitude in bias of the (standard) gamma LL estimator $\hat{m}_G(x)$ are $O(b)$ and $O(b^2)$ for interior and boundary x, respectively. Therefore, the bias property of $\tilde{m}_{G,\lambda}(x)$ indeed improves from that of the LL estimator. While the same order of magnitude in bias can be attained by local cubic smoothing, the skewing method is expected to be more robust against sparse design due to its LL-based structure.

Remark 1. There are a few noteworthy differences in the choices of shifts and weights between skewed estimators using symmetric and asymmetric kernels. The first difference is whether the shift $\ell_1 = \ell_2 = \ell$ that can accelerate the bias convergence of the skewed estimator depends on the design point x. As shown in Theorems 1 and 3, shifts of the gamma and beta skewed estimators vary with x. This is a sharp contrast to the result when a standard nonnegative symmetric kernel $K(u)$ is employed. Choi and Hall (1998) demonstrate that as in our case, symmetric weights and shifts can lead to bias acceleration from $O(h^2)$ to $O(h^4)$ in their skewed estimation using symmetric kernels, where h is the bandwidth. However, for the weight $\lambda_1 = \lambda_2 = \lambda$, they derive the shift as $\ell_1 = \ell_2 = \ell = \ell(\lambda) = \sqrt{(2\lambda + 1)\kappa_2/(2\lambda)}$, where $\kappa_2 = \int u^2 K(u)du$; to put it in another way, the shift remains unchanged regardless of the position of x in Choi and Hall's (1998) original skewed estimator.

The second difference lies in sensitivity to the choice of λ. Our Monte Carlo study in Section 3 compares finite-sample performances of asymmetric kernel-based skewed estimators with those of the skewed estimator using the symmetric, Epanechnikov kernel. The results indicate that skewed estimators smoothed by asymmetric kernels are insensitive to the choice of λ as long as it is large enough. This is not true for the Epanechnikov skewed estimator. Performances of the Epanechnikov estimator appear to be largely affected by the choice of λ. Choi and Hall (1998, Remark 2.4) provide $\lambda^* = 0.0352$ as the minimizer of the scaling factor $V(\lambda)$ in the leading variance term of the Epanechnikov estimator, and they use this value exclusively in their simulation study. Indeed, once the weight deviates from λ^*, performances of this estimator tend to deteriorate.[5] From this viewpoint, we also employ the optimal λ^* for the Epanechnikov skewed estimator in our simulation study.

Third, the points where smoothing is made in left- and right-shifted LL estimators appear to take different forms. These are $x \pm \ell h$ in Choi and Hall's (1998) original skewed estimator and $x \pm \ell b^{1/2}$ in skewed estimators using asymmetric kernels. Such 'seemingly' different expressions turn out to be equivalent if Chen's (1999, 2000b) original notation b is replaced by h^2.[6] The notational change also enables us to understand that the bias acceleration from $O(b)$ to $O(b^2)$ via the asymmetric kernel-based skewed estimation is equivalent to the one from $O(h^2)$ to $O(h^4)$ due to Choi and Hall's (1998) original skewed estimation.

Remark 2. A referee asks us how to choose (λ, ℓ) for the gamma skewed estimator so that Theorem 1 holds when x lies in the vicinity of the origin. To consider this issue, we first focus on the case in which the left-shifted LL estimator $\hat{m}_G\left(x|x - \ell b^{1/2}\right)$ can be still well-defined. Given that $\ell = \sqrt{(2\lambda + 1)x/(2\lambda)}$, we must have $x - \ell b^{1/2} \geq 0$ in this case. It follows that a possible choice of the weight λ would be any number satisfying $\lambda \geq b/\{2(x - b)\}$, provided that $x > b$; in other words, for $x \in (b, \infty)$, the bias acceleration occurs unless a very small λ is chosen. Therefore, $x \in [0, b]$ is the truly problematic case. All we can do in this case would be to define the (standard) gamma LL estimator as the gamma skewed estimator; in fact, $\tilde{m}_{G,\lambda}(0)$ (i.e., the gamma skewed estimator at the origin) collapses to $\hat{m}_G(0)$ (i.e., the gamma LL estimator at the origin). Our Monte Carlo study in Section 3 adopts this practice. Clearly, this practice can make the gamma skewed estimator free of edge effects. It is also noteworthy that the practice does not contradict the order of magnitude in bias of the skewed estimator documented in Theorem 1, because $\hat{m}_G(x)$ for boundary x itself has a $O(b^2)$ bias. The Monte Carlo study deals

[5] A referee finds this aspect in his/her own simulation study, and our preliminary simulation results confirm it. We thank him/her for running simulations and sharing results.

[6] Jones and Henderson (2007) adopt this type of notational change in developing asymmetric kernels with support on [0, 1].

with yet another skewed estimator smoothed by the beta kernel with support on [0, 1], and the practice of switching to the beta LL estimator in the vicinity of 0 or 1 is again applied to this estimator.

2.2.3. Variance Approximation

Next, the variance approximation of the gamma skewed estimator $\tilde{m}_{G,\lambda}(x)$, conditional on $X = x$, is presented.

Theorem 2. *Under Assumptions 1-3, as $n \to \infty$, the variance of $\tilde{m}_{G,\lambda}(x)$ can be approximated by*

$$Var\{\tilde{m}_{G,\lambda}(x)\} = \begin{cases} \frac{1}{nb^{1/2}} \frac{5\sigma^2(x)}{4\sqrt{\pi}\sqrt{x}f(x)} + o\left(n^{-1}b^{-1/2}\right) & for\ interior\ x \\ O\left(n^{-1}b^{-1}\right) & for\ boundary\ x \end{cases}.$$

The analytical expression of the leading variance term for boundary x is of less importance, and thus we only document its order of magnitude. Orders of magnitude in $Var\{\tilde{m}_{G,\lambda}(x)\}$ are $O\left(n^{-1}b^{-1/2}\right)$ and $O\left(n^{-1}b^{-1}\right)$ for interior and boundary x, respectively, and these are the same as those for $Var\{\hat{m}_G(x)\}$. A unique feature of the variance property is that the variance coefficient for interior x is inversely proportional to $x^{1/2}$ and thus decreases as x increases, as in other gamma density and regression estimators. This reflects that the gamma kernel tends to spread out as x increases for a given value of the smoothing parameter b; in other words, the kernel can collect more data points (or increase its effective sample size) to smooth in the areas with fewer observations. This property is particularly advantageous to estimating the regression curves whose regressors have distributions that have a long tail with sparse data, such as those of the economic and financial variables mentioned in Section 1.

Interestingly, the coefficient of the dominant term in $Var\{\tilde{m}_{G,\lambda}(x)\}$ for interior x is independent of λ (although higher-order terms may depend on it), and it follows from

$$Var\{\hat{m}_G(x)\} = \frac{1}{nb^{1/2}} \frac{\sigma^2(x)}{2\sqrt{\pi}\sqrt{x}f(x)} + o\left(n^{-1}b^{-1/2}\right)$$

that $Var\{\tilde{m}_{G,\lambda}(x)\}/Var\{\hat{m}_G(x)\} \to 5/2$ for interior x. At a first glance, this result may give us the impression that while the gamma skewed estimator does not inflate the order of magnitude in variance from the gamma LL estimator, the former inflates the variance coefficient; indeed, the increase can be attributed to the fact that variations of $\mp\ell b^{1/2}\hat{\beta}_G\left(x \pm \ell b^{1/2}\right)$ in the left- and right-shifted LL estimators (3) introduce additional variability, as the proof in Appendix indicates.

However, the above efficiency comparison is based on the implicit assumption that the same smoothing parameter b is chosen across two estimators. Rather, observe that the mean squared error ("MSE") of $\tilde{m}_{G,\lambda}(x)$ for interior x is $MSE\{\tilde{m}_{G,\lambda}(x)\} = O\left(b^4 + n^{-1}b^{-1/2}\right)$. It follows that the optimal smoothing parameter $b^*_{G,\lambda}$ that minimizes the dominant two terms in the MSE is $b^*_{G,\lambda} = O\left(n^{-2/9}\right)$, which yields the variance convergence of $O\left(n^{-8/9}\right)$. On the other hand, because $MSE\{\hat{m}_G(x)\} = O\left(b^2 + n^{-1}b^{-1/2}\right)$ for interior x, the optimal smoothing parameter for the LL estimator is $b^*_{G,LL} = O\left(n^{-2/5}\right)$, which leads to the variance convergence of $O\left(n^{-4/5}\right)$. As a result, when best implemented, $Var\{\tilde{m}_{G,\lambda}(x)\}/Var\{\hat{m}_G(x)\} = O\left(n^{-4/45}\right) \to 0$, or $\tilde{m}_{G,\lambda}(x)$ can be asymptotically more efficient than $\hat{m}_G(x)$ for interior x.

It also follows that the skewed estimator has optimal pointwise convergence of $n^{-4/9}$ for interior x. This convergence rate coincides with Stone's (1980) optimal nonparametric one under fourth-order smoothness of the unknown regression curve.

2.3. A Limit Case

So far we have seen that under a suitably chosen weighting scheme with a fixed weight λ, the skewing method can reduce the bias of the gamma LL estimator $\hat{m}_G(x)$ in an order of magnitude. Some readers may then wonder whether the value of λ can be optimized under some criterion. However, the answer is negative. The asymptotic variance for interior x does not depend on λ, nor the integrated variance based on the trimming argument as in Chen (1999, 2000b). As a result, when the mean integrated squared error ("MISE") is considered as a global performance measure, the squared bias alone varies with λ. Moreover, the leading bias term suggests the choice of a preferably large value of λ because of the reciprocal occurrence of λ in $\mathcal{B}_{G,\lambda}(x; m, f)$.

Instead, as in Choi and Hall (1998, Remarks 2.1 and 2.4), we let λ diverge to infinity and denote the limit gamma left- and right-shifted LL estimators by $\hat{m}_{G,-}(x)$ and $\hat{m}_{G,+}(x)$, respectively. These are formally expressed as

$$\hat{m}_{G,\pm}(x) := \hat{m}_G\left(x|x \pm \sqrt{xb}\right),$$

where $\lim_{\lambda\to\infty} \ell(\lambda, x) = \sqrt{x}$ is substituted in place of ℓ in $\hat{m}_G\left(x|x \pm \ell b^{1/2}\right)$. The corresponding limit gamma skewed estimator takes the form of an average of the two shifted LL estimators

$$\tilde{m}_{G,\infty}(x) := \frac{1}{2}\left\{\hat{m}_{G,-}(x) + \hat{m}_{G,+}(x)\right\}.$$

This estimator can be interpreted as a limit form of the convex combination of three LL estimators.

The next proposition provides approximations to the bias and variance of $\hat{m}_{G,\pm}(x)$ and $\tilde{m}_{G,\infty}(x)$, again conditional on $X = x$.

Proposition 1. *(i) Under Assumptions 1-3, as $n \to \infty$, the bias and variance of $\hat{m}_{G,\pm}(x)$ can be approximated by*

$$Bias\left\{\hat{m}_{G,\pm}(x)\right\} = \mp\sqrt{x}\left[\left\{\frac{3}{2} + x\frac{f^{(1)}(x)}{f(x)}\right\}m^{(2)}(x) + \frac{x}{3}m^{(3)}(x)\right]b^{3/2} + \mathcal{B}_{G,\infty}(x; m, f)b^2 + o\left(b^2\right),$$

where

$$\begin{aligned}\mathcal{B}_{G,\infty}(x; m, f) &= \lim_{\lambda\to\infty}\mathcal{B}_{G,\lambda}(x; m, f)\\ &= -\left[\left\{1 + 2x\frac{f^{(1)}(x)}{f(x)} + \frac{x^2}{2}\frac{f^{(2)}(x)}{f(x)}\right\}m^{(2)}(x) + x\left\{\frac{5}{3} + x\frac{f^{(1)}(x)}{f(x)}\right\}m^{(3)}(x) + \frac{x^2}{4}m^{(4)}(x)\right],\end{aligned}$$

and

$$Var\left\{\hat{m}_{G,\pm}(x)\right\} = \begin{cases}\frac{1}{nb^{1/2}}\frac{3\sigma^2(x)}{4\sqrt{\pi}\sqrt{x}f(x)} + o\left(n^{-1}b^{-1/2}\right) & for\ interior\ x\\ O\left(n^{-1}b^{-1}\right) & for\ boundary\ x\end{cases}.$$

(ii) Under Assumptions 1-3, as $n \to \infty$, the bias and variance of $\tilde{m}_{G,\infty}(x)$ can be approximated by

$$Bias\{\tilde{m}_{G,\infty}(x)\} = \mathcal{B}_{G,\infty}(x; m, f)b^2 + o\left(b^2\right),\ and$$

$$Var\{\tilde{m}_{G,\infty}(x)\} = \begin{cases}\frac{1}{nb^{1/2}}\frac{5\sigma^2(x)}{4\sqrt{\pi}\sqrt{x}f(x)} + o\left(n^{-1}b^{-1/2}\right) & for\ interior\ x\\ O\left(n^{-1}b^{-1}\right) & for\ boundary\ x\end{cases}.$$

Observe that $\hat{m}_{G,\pm}(x)$ alone can reduce the bias from $O(b)$ for the standard LL estimator to $O(b^{3/2})$ for interior x. Coefficients on their leading $O(b^{3/2})$ bias terms have opposite signs with the same magnitude, and thus they are offset by taking an average of the two. It follows that $\tilde{m}_{G,\infty}(x)$ can accelerate the bias convergence up until $O(b^2)$. It can be also deduced that $Var\left\{\hat{m}_{G,\pm}(x)\right\}/Var\left\{\hat{m}_G(x)\right\} \to 3/2$ for interior x and a given b.

2.4. An Extension: The Beta Skewed Estimator

Our analysis is now extended to the skewed estimators smoothed by the beta kernel. After convergence properties (i.e., the bias-variance trade-off) of the beta skewed estimator $\tilde{m}_{B,\lambda}(x)$ and the second-order expansion of the bias of the beta LL estimator $\hat{m}_B(x|x) = \hat{m}_B(x)$ are presented, those of the limit beta left- and right-shifted LL estimators $\hat{m}_{B,\pm}(x)$ and the limit beta skewed estimator $\tilde{m}_{B,\infty}(x)$ are documented. These results are again conditional on $X = x$. It can be immediately found that all the previous discussions on the asymptotic results for the gamma case continue to hold for the beta case.

Theorem 3. *Suppose that for a given design point $x \in (0, 1)$, weights (λ_1, λ_2) and shifts (ℓ_1, ℓ_2) in (4) are chosen as*

$$\lambda_1 = \lambda_2 = \lambda\ and\ \ell_1 = \ell_2 = \ell = \ell(\lambda, x) = \sqrt{\left(\frac{2\lambda+1}{2\lambda}\right)x(1-x)}.$$

Denote the beta skewed estimator with the above weights and shifts plugged into (4) by

$$\tilde{m}_{B,\lambda}(x) = \frac{\lambda\hat{m}_B\left(x|x - \ell b^{1/2}\right) + \hat{m}_B(x|x) + \lambda\hat{m}_B\left(x|x + \ell b^{1/2}\right)}{2\lambda+1}.$$

If Assumptions 1-3 hold, then, as $n \to \infty$, the bias of $\tilde{m}_{B,\lambda}(x)$ can be approximated by $Bias\left\{\tilde{m}_{B,\lambda}(x)\right\} = \mathcal{B}_{B,\lambda}(x; m, f)b^2 + o\left(b^2\right)$, where

$$\begin{aligned}\mathcal{B}_{B,\lambda}(x; m, f) = -&\left[\left\{1 - 6x(1-x) + 2x(1-x)(1-2x)\frac{f^{(1)}(x)}{f(x)} + \frac{1}{2}x^2(1-x)^2\frac{f^{(2)}(x)}{f(x)}\right\}m^{(2)}(x)\right.\\ &\left. + x(1-x)\left\{\frac{5}{3}(1-2x) + x(1-x)\frac{f^{(1)}(x)}{f(x)}\right\}m^{(3)}(x) + \frac{1}{4}\left(1 + \frac{1}{4\lambda}\right)x^2(1-x)^2m^{(4)}(x)\right].\end{aligned}$$

Moreover, the variance of $\tilde{m}_{B,\lambda}(x)$ can be approximated by

$$Var\{\tilde{m}_{B,\lambda}(x)\} = \begin{cases}\frac{1}{nb^{1/2}}\frac{5\sigma^2(x)}{4\sqrt{\pi}\sqrt{x(1-x)}f(x)} + o\left(n^{-1}b^{-1/2}\right) & for\ interior\ x\\ O\left(n^{-1}b^{-1}\right) & for\ boundary\ x\end{cases}.$$

Corollary 2. *Under Assumptions 1-3, as $n \to \infty$, the bias of $\hat{m}_B(x)$ can be expanded up to the second order as*

$$\begin{aligned} Bias\left\{\hat{m}_B(x)\right\} =& \frac{1}{2}x(1-x)m^{(2)}(x)b + \left[-\left\{1-\frac{5}{2}x(1-x)+x(1-x)(1-2x)\frac{f^{(1)}(x)}{f(x)}\right.\right. \\ &\left. -\frac{1}{2}x^2(1-x)^2\frac{f^{(2)}(x)}{f(x)} + x^2(1-x)^2\left(\frac{f^{(1)}(x)}{f(x)}\right)^2\right\}m^{(2)}(x) \\ &\left. + \frac{1}{3}x(1-x)(1-2x)m^{(3)}(x) + \frac{1}{8}x^2(1-x)^2m^{(4)}(x)\right]b^2 + o\left(b^2\right). \end{aligned}$$

Proposition 2. *(i) Under Assumptions 1-3, as $n \to \infty$, the bias and variance of $\hat{m}_{B,\pm}(x) := \hat{m}_B\left(x|x \pm \sqrt{x(1-x)b}\right)$ can be approximated by*

$$\begin{aligned} Bias\left\{\hat{m}_{B,\pm}(x)\right\} =& \mp\sqrt{x(1-x)}\left[\left\{\frac{3}{2}(1-2x)+x(1-x)\frac{f^{(1)}(x)}{f(x)}\right\}m^{(2)}(x)\right. \\ &\left. + \frac{1}{3}x(1-x)m^{(3)}(x)\right]b^{3/2} + \mathcal{B}_{B,\infty}(x;m,f)b^2 + o\left(b^2\right), \end{aligned}$$

where

$$\begin{aligned} \mathcal{B}_{B,\infty}(x;m,f) =& \lim_{\lambda\to\infty}\mathcal{B}_{B,\lambda}(x;m,f) \\ =& -\left[\left\{1-6x(1-x)+2x(1-x)(1-2x)\frac{f^{(1)}(x)}{f(x)}+\frac{1}{2}x^2(1-x)^2\frac{f^{(2)}(x)}{f(x)}\right\}m^{(2)}(x)\right. \\ &\left. + x(1-x)\left\{\frac{5}{3}(1-2x)+x(1-x)\frac{f^{(1)}(x)}{f(x)}\right\}m^{(3)}(x) + \frac{1}{4}x^2(1-x)^2m^{(4)}(x)\right] \end{aligned}$$

and

$$Var\left\{\hat{m}_{B,\pm}(x)\right\} = \begin{cases} \frac{1}{nb^{1/2}}\frac{3\sigma^2(x)}{4\sqrt{\pi}\sqrt{x(1-x)}f(x)} + o\left(n^{-1}b^{-1/2}\right) & \text{for interior } x \\ O\left(n^{-1}b^{-1}\right) & \text{for boundary } x \end{cases}.$$

(ii) Under Assumptions 1-3, as $n \to \infty$, the bias and variance of $\tilde{m}_{B,\infty}(x) := \left\{\hat{m}_{B,-}(x)+\hat{m}_{B,+}(x)\right\}/2$ can be approximated by

$$Bias\{\tilde{m}_{B,\infty}(x)\} = \mathcal{B}_{B,\infty}(x;m,f)b^2 + o\left(b^2\right), \text{ and}$$

$$Var\{\tilde{m}_{B,\infty}(x)\} = \begin{cases} \frac{1}{nb^{1/2}}\frac{5\sigma^2(x)}{4\sqrt{\pi}\sqrt{x(1-x)}f(x)} + o\left(n^{-1}b^{-1/2}\right) & \text{for interior } x \\ O\left(n^{-1}b^{-1}\right) & \text{for boundary } x \end{cases}.$$

Remark 3. Seifert and Gasser (1996, Theorem 1) argue that finite-sample variances of local polynomial regression estimators using kernels with compact support may be unbounded. This problem typically occurs when smoothing is made in sparse regions. As a remedy, Seifert and Gasser (1996) propose to increase the bandwidth locally in such regions. However, Chen (2002, Lemma 1) demonstrates that the pth-order local polynomial regression estimator using the beta kernel is immune to this problem, as long as at least $p+1$ different data points are not on the boundary of [0, 1]. It follows that in finite samples the beta LL estimator has finite variance with probability 1, and the same is true for the beta skewed estimation.

3. Finite-Sample Performance

In this section, finite-sample properties of the skewed estimators are assessed in comparison with other competing nonparametric regression estimators. Our Monte Carlo study starts from assessing the gamma skewed estimation and then proceeds to the beta skewed estimation. For each case, the data are generated from the regression model $Y = m(X) + \epsilon$, where the regressor X obeys some marginal distribution that will be specified shortly, $X \perp\!\!\!\perp \epsilon$, and $\epsilon \stackrel{d}{=} N\left(0, 0.05^2\right)$.

3.1. On the Gamma Skewed Estimator

3.1.1. Monte Carlo Design

Our first Monte Carlo study focuses on a regressor with support on $\mathbb{R}_+$. Let $X \stackrel{d}{=} G(2,1)$. Also let the regression curve $m(x)$ take one of the following functional forms:

$$\text{Model A} : m(x) = \exp(-x) + \exp\left\{-4(x-1)^2\right\}.$$

$$\text{Model B} : m(x) = -2\sqrt{\frac{x}{1+x}}\sin\left(\frac{4\pi}{\sqrt{1+x}}\right).$$

The shape of each curve can be found in Fig. 2 below. The sample size is $n \in \{100, 200, 400\}$, and 1000 replications are drawn for each combination of the model and sample size. It is worth remarking that our Monte Carlo design with Model A exactly mimics the one in Chen (2002). Hence, our results for this model are directly comparable with the results therein.

The simulation study compares finite-sample performance of the following seven regression estimators for $m(x)$: (i) the LL estimator smoothed by the symmetric, Epanechnikov kernel $\hat{m}_E(x)$; (ii) the skewed estimator with $\lambda^* = 0.0352$ smoothed by the Epanechnikov kernel $\tilde{m}_{E,\lambda^*}(x)$; (iii) the LL estimator smoothed by the gamma kernel $\hat{m}_G(x)$; (iv) the local cubic estimator smoothed by the gamma kernel $\hat{m}_{G,C}(x) = \mathbf{e}^\top \mathbf{S}_G(x)^{-1}\mathbf{T}_G(x)$, where $\mathbf{e}^\top = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$,

$$\mathbf{S}_G(x) = \begin{bmatrix} s_{G0}(x) & s_{G1}(x) & s_{G2}(x) & s_{G3}(x) \\ s_{G1}(x) & s_{G2}(x) & s_{G3}(x) & s_{G4}(x) \\ s_{G2}(x) & s_{G3}(x) & s_{G4}(x) & s_{G5}(x) \\ s_{G3}(x) & s_{G4}(x) & s_{G5}(x) & s_{G6}(x) \end{bmatrix}, \text{ and } \mathbf{T}_G(x) = \begin{bmatrix} t_{G0}(x) \\ t_{G1}(x) \\ t_{G2}(x) \\ t_{G3}(x) \end{bmatrix};$$

(v) the skewed estimator with $\lambda = 0.1$ smoothed by the gamma kernel $\tilde{m}_{G,0.1}(x)$; (vi) the skewed estimator with $\lambda = 1$ (= equally-weighted skewed estimator) smoothed by the gamma kernel $\tilde{m}_{G,1}(x)$; and (vii) the limit skewed estimator smoothed by the gamma kernel $\tilde{m}_{G,\infty}(x)$. The estimators (i)-(iv) can be viewed as competitors to the gamma skewed estimators (v)-(vii). The Epanechnikov skewed estimator (ii) is exactly the same as employed in the simulation study of Choi and Hall (1998). As mentioned in Remark 1, this estimator is sensitive to the choice of λ, and λ^* has some optimality. On the other hand, we consider three different gamma skewed estimators (v)-(vii) in order to check the sensitivity of $\tilde{m}_{G,\lambda}$ to the choice of λ. Furthermore, as recommended in Remark 2, if a gamma skewed estimator is not well-defined for a very small x, then it is replaced by the gamma LL estimator evaluated at the same design point.

Performance of an estimator $\bar{m}(x)$ is assessed via both global and local measures. The global measures include the root integrated squared error ("RISE")

$$RISE\{\bar{m}(x)\} = \sqrt{\int_0^\infty \{\bar{m}(x) - m(x)\}^2 dx},$$

and the integrated absolute deviation ("IAD")

$$IAD\{\bar{m}(x)\} = \int_0^\infty |\bar{m}(x) - m(x)| dx,$$

where each integral is approximated by the trapezoidal rule on an equally-spaced grid of 551 points over the interval [0.0, 5.5]. In addition, as the local measures, the squared bias and variance evaluated at twelve design points $x \in \{0.0, 0.5, \ldots, 5.0, 5.5\}$ over 1000 Monte Carlo samples are computed.

Furthermore, how to choose the tuning parameter is an important practical issue in kernel smoothing. We obtain the optimal global tuning parameter of an estimator by an oracle method (Case #1) and then by the cross-validation ("CV") method (Case #2). Details of each method are discussed in the corresponding section below.

3.1.2. Case #1: Oracle Tuning Parameters

By the oracle tuning parameter for an estimator $\bar{m}(x)$ we mean the optimal global tuning parameter (i.e., the bandwidth h for the Epanechnikov kernel or the smoothing parameter b for the gamma kernel) that minimizes $RISE\{\bar{m}(x)\}$. The minimizer is found on an equally-spaced grid of 40 points over the interval [0.01, 0.40].

Table 1 presents averages ("Ave") and standard deviations ("SD") of global performance measures and oracle tuning parameter values over 1000 Monte Carlo samples. Assessing an estimator with the oracle tuning parameter value plugged in can be interpreted as evaluating its potential.

It can be immediately found that global measures of each estimator shrink with the sample size, which indicates consistency. A closer look reveals that two Epanechnikov estimators perform inferior to other estimators for $n = 100$. Although their performances improve with the sample size, they do not surpass those of $\hat{m}_{G,C}(x)$, $\tilde{m}_{G,1}(x)$ and $\tilde{m}_{G,\infty}(x)$ even for $n = 400$. It is also worth remarking that average lengths of their tuning parameters are much larger than those of gamma estimators. Choosing a global bandwidth for a symmetric kernel implies that amount of smoothing is fixed everywhere. It seems that to take care of sparseness of data points on the right tail, both Epanechnikov estimators pick out fixed, long bandwidths for extrapolation to the right.

Now we turn to gamma estimators. As expected, each of four bias-corrected gamma estimators (i.e., $\hat{m}_{G,C}(x)$, $\tilde{m}_{G,0.1}(x)$, $\tilde{m}_{G,1}(x)$, and $\tilde{m}_{G,\infty}(x)$) more or less outperforms the gamma LL estimator $\hat{m}_G(x)$. Comparing global measures of three gamma skewed estimators reveals that performances of $\tilde{m}_{G,1}(x)$ and $\tilde{m}_{G,\infty}(x)$ look alike and are in general superior to those of $\tilde{m}_{G,0.1}(x)$. This can be attributed to the fact that a small weight λ inflates the leading bias coefficient of $\tilde{m}_{G,\lambda}(x)$, whereas its leading variance coefficient is insensitive to λ. Furthermore, $\tilde{m}_{G,1}(x)$ and $\tilde{m}_{G,\infty}(x)$ perform better than $\hat{m}_{G,C}(x)$ for Model A, and the results are reversed for Model B. However, margins are small for each case, and thus it may be safe to say that these estimators perform equally well. Also observe that average tuning parameter values of gamma estimators are much smaller than those of Epanechnikov estimators.

Following Chen (2002), we visualize local measures of the estimators in Fig. 1. The results from Models A and B are qualitatively similar, and thus only the former is presented. Notice that the vertical axis in each panel is the natural logarithm of either squared bias or variance. Moreover, to avoid too many lines in a panel, we include only the limit skewed

Table 1
Monte Carlo Results (Gamma Case; Oracle).

Model	n	Estimator	RISE		IAD		Tuning Parameter	
			Ave	SD	Ave	SD	Ave	SD
A	100	$\hat{m}_E(x)$	0.1042	(0.0711)	0.1483	(0.0489)	0.2797	(0.0797)
		$\tilde{m}_{E,\lambda^*}(x)$	0.1007	(0.0688)	0.1450	(0.0537)	0.2725	(0.0763)
		$\hat{m}_G(x)$	0.0812	(0.0216)	0.1383	(0.0338)	0.0225	(0.0073)
		$\hat{m}_{G,C}(x)$	0.0689	(0.0277)	0.1197	(0.0451)	0.0736	(0.0385)
		$\tilde{m}_{G,0.1}(x)$	0.0707	(0.0235)	0.1208	(0.0367)	0.0230	(0.0064)
		$\tilde{m}_{G,1}(x)$	0.0627	(0.0208)	0.1091	(0.0322)	0.0269	(0.0088)
		$\tilde{m}_{G,\infty}(x)$	0.0627	(0.0205)	0.1093	(0.0312)	0.0279	(0.0090)
	200	$\hat{m}_E(x)$	0.0675	(0.0319)	0.1103	(0.0247)	0.2465	(0.0650)
		$\tilde{m}_{E,\lambda^*}(x)$	0.0640	(0.0327)	0.1034	(0.0282)	0.2401	(0.0586)
		$\hat{m}_G(x)$	0.0654	(0.0113)	0.1092	(0.0174)	0.0206	(0.0025)
		$\hat{m}_{G,C}(x)$	0.0468	(0.0138)	0.0810	(0.0211)	0.0540	(0.0208)
		$\tilde{m}_{G,0.1}(x)$	0.0541	(0.0132)	0.0925	(0.0201)	0.0222	(0.0042)
		$\tilde{m}_{G,1}(x)$	0.0444	(0.0119)	0.0778	(0.0174)	0.0228	(0.0050)
		$\tilde{m}_{G,\infty}(x)$	0.0443	(0.0117)	0.0777	(0.0171)	0.0234	(0.0055)
	400	$\hat{m}_E(x)$	0.0473	(0.0122)	0.0830	(0.0146)	0.2182	(0.0449)
		$\tilde{m}_{E,\lambda^*}(x)$	0.0429	(0.0130)	0.0735	(0.0160)	0.2202	(0.0398)
		$\hat{m}_G(x)$	0.0583	(0.0068)	0.0936	(0.0115)	0.0202	(0.0013)
		$\hat{m}_{G,C}(x)$	0.0348	(0.0099)	0.0590	(0.0143)	0.0435	(0.0154)
		$\tilde{m}_{G,0.1}(x)$	0.0480	(0.0107)	0.0795	(0.0167)	0.0228	(0.0045)
		$\tilde{m}_{G,1}(x)$	0.0341	(0.0080)	0.0591	(0.0123)	0.0216	(0.0037)
		$\tilde{m}_{G,\infty}(x)$	0.0337	(0.0078)	0.0583	(0.0119)	0.0217	(0.0038)
B	100	$\hat{m}_E(x)$	0.7646	(0.4999)	0.7307	(0.5103)	0.3841	(0.0360)
		$\tilde{m}_{E,\lambda^*}(x)$	0.7356	(0.4812)	0.6984	(0.4966)	0.3840	(0.0352)
		$\hat{m}_G(x)$	0.1236	(0.0676)	0.1992	(0.0596)	0.0208	(0.0036)
		$\hat{m}_{G,C}(x)$	0.0706	(0.0364)	0.1120	(0.0468)	0.0787	(0.0404)
		$\tilde{m}_{G,0.1}(x)$	0.0914	(0.0751)	0.1232	(0.0567)	0.0240	(0.0062)
		$\tilde{m}_{G,1}(x)$	0.0891	(0.0742)	0.1189	(0.0516)	0.0253	(0.0073)
		$\tilde{m}_{G,\infty}(x)$	0.0905	(0.0739)	0.1213	(0.0512)	0.0250	(0.0070)
	200	$\hat{m}_E(x)$	0.2221	(0.2709)	0.2420	(0.1959)	0.3143	(0.0658)
		$\tilde{m}_{E,\lambda^*}(x)$	0.2129	(0.2588)	0.2184	(0.1915)	0.3139	(0.0652)
		$\hat{m}_G(x)$	0.0871	(0.0205)	0.1587	(0.0263)	0.0202	(0.0013)
		$\hat{m}_{G,C}(x)$	0.0463	(0.0193)	0.0731	(0.0251)	0.0584	(0.0282)
		$\tilde{m}_{G,0.1}(x)$	0.0495	(0.0253)	0.0784	(0.0229)	0.0229	(0.0046)
		$\tilde{m}_{G,1}(x)$	0.0487	(0.0249)	0.0756	(0.0206)	0.0230	(0.0048)
		$\tilde{m}_{G,\infty}(x)$	0.0497	(0.0248)	0.0773	(0.0204)	0.0228	(0.0047)
	400	$\hat{m}_E(x)$	0.0755	(0.0715)	0.1146	(0.0502)	0.2158	(0.0559)
		$\tilde{m}_{E,\lambda^*}(x)$	0.0686	(0.0711)	0.0949	(0.0494)	0.2185	(0.0534)
		$\hat{m}_G(x)$	0.0775	(0.0084)	0.1470	(0.0180)	0.0201	(0.0011)
		$\hat{m}_{G,C}(x)$	0.0336	(0.0148)	0.0513	(0.0152)	0.0486	(0.0173)
		$\tilde{m}_{G,0.1}(x)$	0.0363	(0.0106)	0.0611	(0.0156)	0.0233	(0.0047)
		$\tilde{m}_{G,1}(x)$	0.0344	(0.0102)	0.0549	(0.0122)	0.0217	(0.0039)
		$\tilde{m}_{G,\infty}(x)$	0.0350	(0.0101)	0.0560	(0.0120)	0.0216	(0.0038)

estimator among all gamma skewed estimators.[7] Notice that although the limit gamma skewed estimator for a very small x is replaced with the gamma LL estimator, their local measures differ because of difference in their smoothing parameter values. For example, $\tilde{m}_{G,\infty}(0)$ is replaced by $\hat{m}_G(0)$. Nonetheless, their squared bias and variance differ (although the gaps are very small and thus almost invisible in Fig. 1).

Fig. 1 indicates that the gamma LL estimator tends to have large squared bias. Both the gamma local cubic and limit gamma skewed estimators improve the LL estimator in terms of bias in the sense that squared bias of the former in general lies below that of the latter. We can also see that variance of each Epanechnikov estimator tends to inflate as the design point x moves away from the origin. In contrast, three gamma estimators have similar variance properties, and their variance reduction on the right tail is often substantial. These findings coincide with those in Chen (2002).

To illustrate how well each estimator can capture shapes of true regression curves, we plot pointwise averages of curve estimates over 1000 Monte Carlo samples for $n = 200$ in Fig. 2. The set of estimators used here is the same as in Fig. 1. It can be found that the gamma local cubic and limit gamma skewed estimators can capture shapes of entire regression curves (including peaks and troughs near the origin and right tails) equally well. On the other hand, two Epanechnikov estimators

[7] Based on our preliminary analysis, local performances of gamma skewed estimators with $\lambda = 1$ and $\lambda = 0.1$ are similar and inferior to those of the limit skewed estimator, respectively.

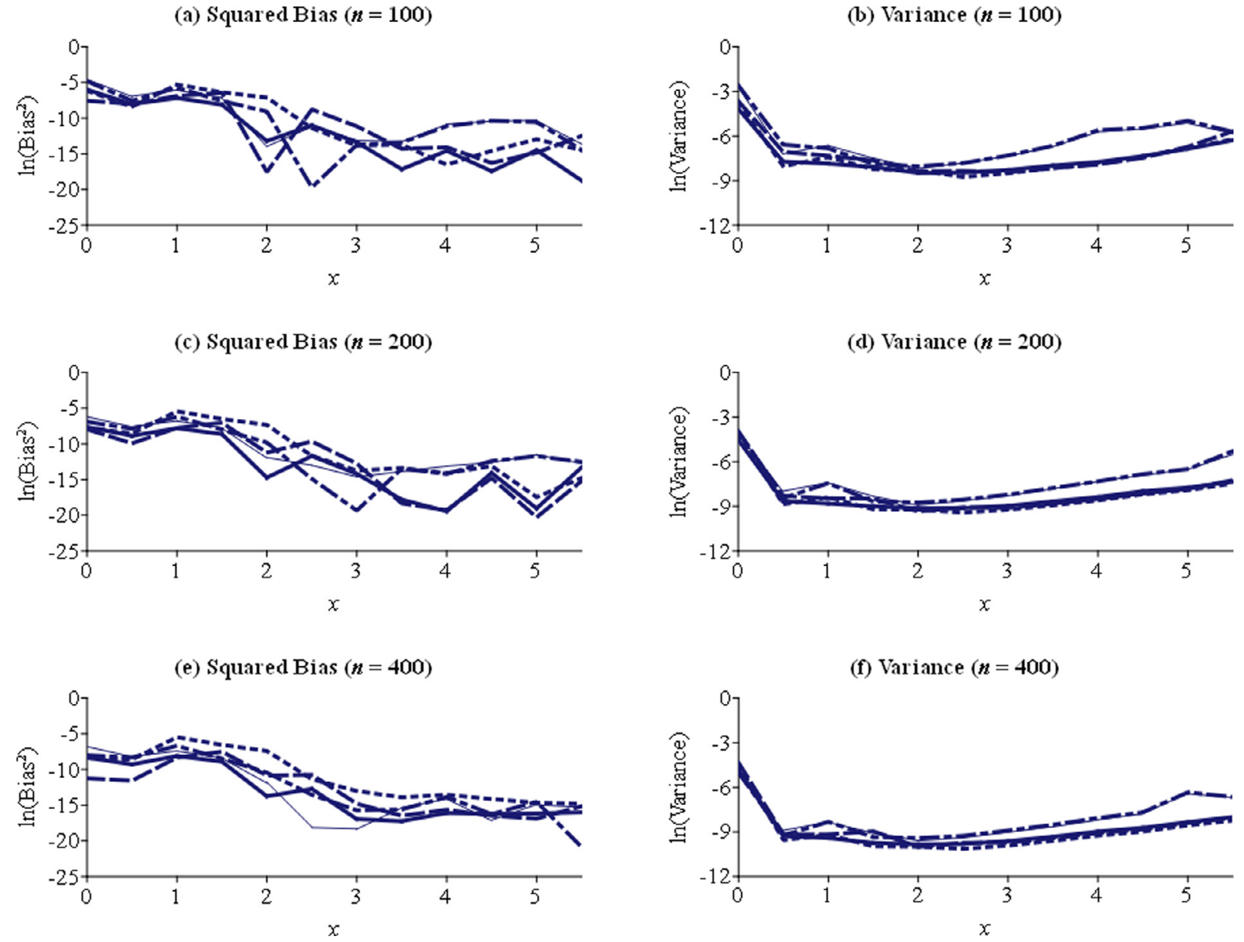

Fig. 1. Squared Bias and Variance of Curve Estimates (Gamma Case [Model A]; Oracle).
Note: Dash-dot, thin solid, dot, dash, and solid lines correspond to the Epanechnikov LL, Epanechnikov skewed, gamma LL, gamma local cubic, and limit gamma skewed estimates, respectively.

are not so impressive as the gamma estimators. Despite a long bandwidth and the sample size of 200, they yield wiggly (and often severely biased) curve estimates on the right tail. As a result of oversmoothing via the long bandwidth, they also fail to precisely capture the peak or trough in the vicinity of the origin.

In sum, unless a small weight is chosen, the gamma skewed estimator has nice finite-sample properties. Its finite-sample performance in general dominates those of the Epanechnikov LL and skewed estimators and is comparable with that of the gamma local cubic estimator. Based on these findings, we focus exclusively on the gamma skewed with $\lambda = 1$ and limit gamma skewed estimators in the analysis below and Section 4. The gamma LL estimator is also employed as a benchmark.

3.1.3. Case #2: Cross-Validated Tuning Parameters

We proceed to investigate curve estimates with data-based tuning parameter values plugged in. The leave-one-out CV is exclusively considered as the data-driven implementation method. CV is popularly employed for LL regression smoothing, and Choi and Hall (1998, Remark 2.6) themselves anticipate that it may serve as a straightforward implementation method for their skewed estimation. While a plug-in method based on the leading two terms in the MISE could be another option, it does not look attractive because in practice it requires to estimate many derivatives of m and f (as well as themselves), as indicated in Theorems 1 and 2.

Details of the CV employed in this study are as follows. Let $\bar{m}_{-i,b}(\cdot)$ be a regression estimate using the smoothing parameter b and the sample with the ith observation eliminated. Then, the CV criterion function becomes

$$CV(b) = \sum_{i=1}^{n} \{Y_i - \bar{m}_{-i,b}(X_i)\}^2.$$

The data-driven optimal global smoothing parameter is defined as the minimizer of $CV(b)$. As in the oracle case, the minimizer is found on an equally-spaced grid of 40 points over the interval [0.01, 0.40].

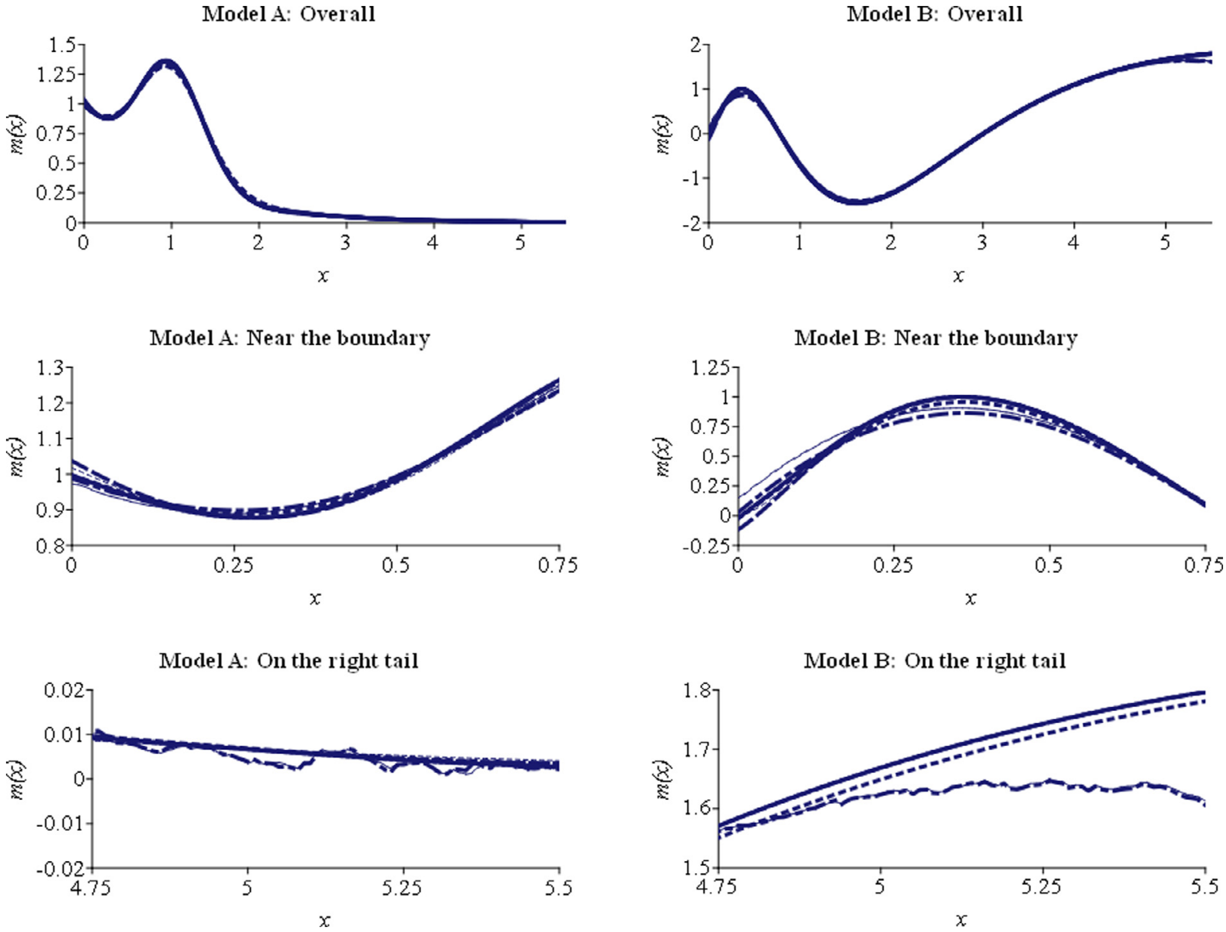

Fig. 2. Pointwise Averages of Curve Estimates (Gamma Case; Oracle; $n = 200$).
Note: Thin dash lines represent true regression curves. Dash-dot, thin solid, dot, dash, and solid lines correspond to the Epanechnikov LL, Epanechnikov skewed, gamma LL, gamma local cubic, and limit gamma skewed estimates, respectively.

Table 2 and Fig. 3 report the global and local performance measures of the gamma LL, gamma skewed with $\lambda = 1$ and limit gamma skewed estimators. The performance measures are exactly the same as in Table 1 and Fig. 1. Again the results of local measures from Models A and B are qualitatively similar, and thus only the former is presented in Fig. 3. Notice that in this case the random number generator is initialized using the same seed number as in Case #1. Therefore, Monte Carlo samples for two cases are identical and thus we may compare the results in Tables 1 and 2 or Figs. 1 and 3 directly.

It can be found in Table 2 that the CV-based smoothing parameter values are fairly close enough to the corresponding oracle ones. Because kernel estimators are sensitive to choices of tuning parameter values, both global and local performance measures become worse. However, thanks to small difference between oracle and CV-based smoothing parameter values, deterioration in the measures by switching from the former to the latter is minimal at best. As indicated in the previous section, global measures of two skewed estimators look alike again and dominate those of the gamma LL estimator. In terms of local measures, there is not much difference in variance among the three estimators. It is not surprising that variances of two skewed estimators are close to each other. This simply reflects that dominant terms of their variances are identical and independent of λ. Also observe that the skewed estimators tend to have better bias properties as the sample size increases.

All in all, Monte Carlo results indicate that gamma skewed estimators work well in practice and tend to outperform the gamma LL estimator for the sample sizes that are typically used in empirical applications. To implement the estimators, we may rely on CV.

3.2. On the Beta Skewed Estimator

The second Monte Carlo study deals with a regressor with support on [0, 1]. The regressor X is generated as absolute values of $N(0, 0.3^2)$ random variables truncated on $[-1, 1]$. The true regression curve is $m(x) = (x - 1/2)^2$. The sample size is $n \in \{100, 200, 400\}$, and the number of replications is 1000. This design is again the same as in the one in Chen (2002), and thus our results are directly comparable with the results therein.

Table 2
Monte Carlo Results (Gamma Case; CV).

Model	n	Estimator	RISE		IAD		Tuning Parameter	
			Ave	SD	Ave	SD	Ave	SD
A	100	$\hat{m}_G(x)$	0.0929	(0.0356)	0.1508	(0.0449)	0.0288	(0.0195)
		$\tilde{m}_{G,1}(x)$	0.0738	(0.0386)	0.1209	(0.0482)	0.0320	(0.0199)
		$\tilde{m}_{G,\infty}(x)$	0.0735	(0.0384)	0.1203	(0.0469)	0.0321	(0.0202)
	200	$\hat{m}_G(x)$	0.0807	(0.0214)	0.1288	(0.0308)	0.0307	(0.0143)
		$\tilde{m}_{G,1}(x)$	0.0550	(0.0209)	0.0940	(0.0317)	0.0342	(0.0154)
		$\tilde{m}_{G,\infty}(x)$	0.0534	(0.0204)	0.0912	(0.0298)	0.0338	(0.0153)
	400	$\hat{m}_G(x)$	0.0821	(0.0187)	0.1241	(0.0268)	0.0347	(0.0130)
		$\tilde{m}_{G,1}(x)$	0.0505	(0.0184)	0.0843	(0.0286)	0.0380	(0.0138)
		$\tilde{m}_{G,\infty}(x)$	0.0478	(0.0177)	0.0800	(0.0269)	0.0375	(0.0138)
B	100	$\hat{m}_G(x)$	0.1424	(0.0798)	0.2339	(0.0927)	0.0271	(0.0136)
		$\tilde{m}_{G,1}(x)$	0.1054	(0.0899)	0.1349	(0.0728)	0.0311	(0.0162)
		$\tilde{m}_{G,\infty}(x)$	0.1063	(0.0900)	0.1366	(0.0723)	0.0308	(0.0164)
	200	$\hat{m}_G(x)$	0.1140	(0.0382)	0.2129	(0.0659)	0.0298	(0.0109)
		$\tilde{m}_{G,1}(x)$	0.0638	(0.0407)	0.0941	(0.0430)	0.0339	(0.0131)
		$\tilde{m}_{G,\infty}(x)$	0.0639	(0.0401)	0.0943	(0.0416)	0.0333	(0.0129)
	400	$\hat{m}_G(x)$	0.1177	(0.0318)	0.2285	(0.0637)	0.0343	(0.0110)
		$\tilde{m}_{G,1}(x)$	0.0545	(0.0296)	0.0826	(0.0407)	0.0380	(0.0126)
		$\tilde{m}_{G,\infty}(x)$	0.0541	(0.0289)	0.0815	(0.0387)	0.0374	(0.0125)

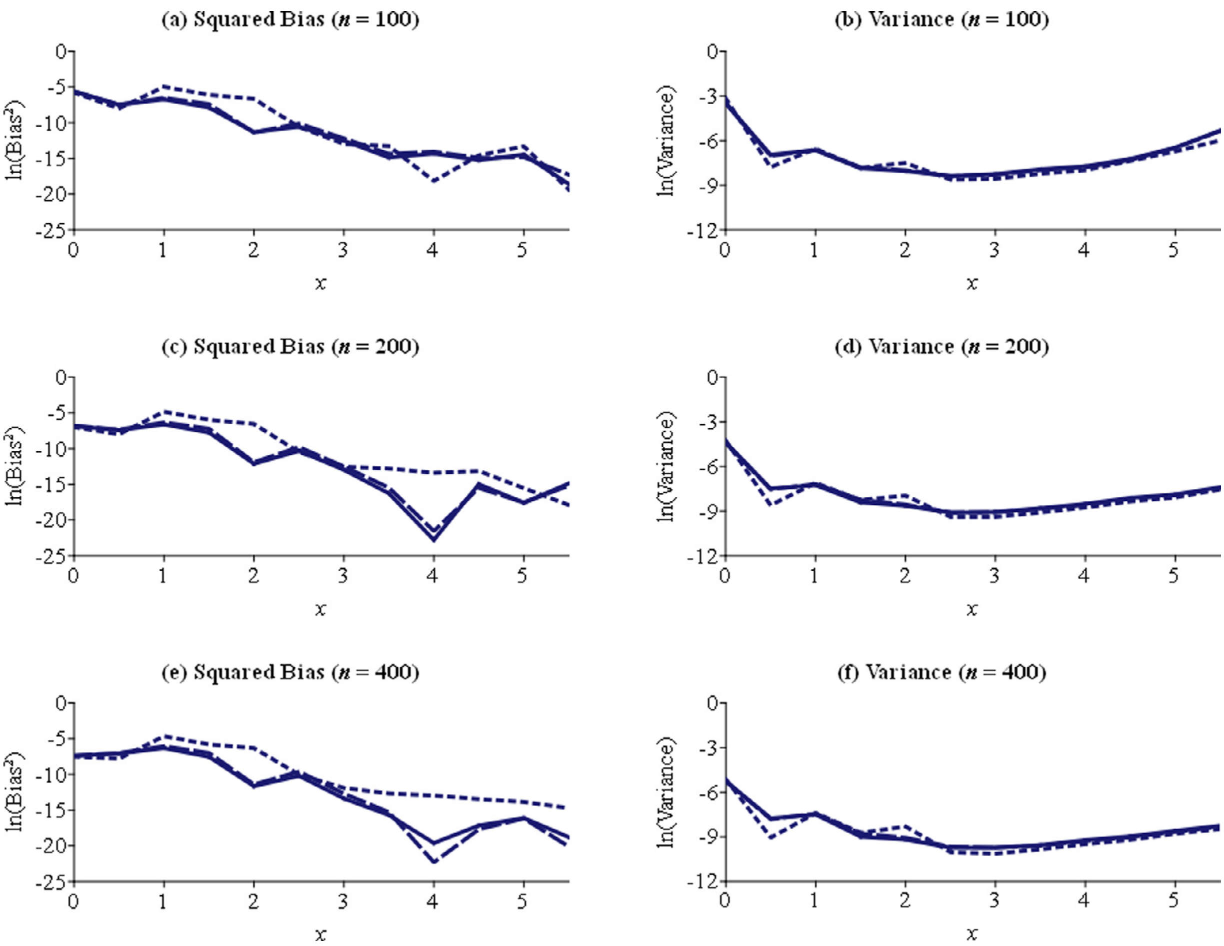

Fig. 3. Squared Bias and Variance of Curve Estimates (Gamma Case [Model A]; CV).
Note: Dot, dash, and solid lines correspond to the gamma LL, gamma skewed with $\lambda = 1$, and limit gamma skewed estimates, respectively.

Table 3
Monte Carlo Results (Beta Case; Oracle).

n	Estimator	RISE		IAD		Tuning Parameter	
		Ave	SD	Ave	SD	Ave	SD
100	$\hat{m}_E(x)$	0.0385	(0.0247)	0.0258	(0.0135)	0.3138	(0.1159)
	$\tilde{m}_{E,\lambda^*}(x)$	0.0392	(0.0246)	0.0254	(0.0139)	0.3017	(0.1122)
	$\hat{m}_B(x)$	0.0270	(0.0219)	0.0185	(0.0126)	0.0881	(0.0571)
	$\hat{m}_{B,C}(x)$	0.0354	(0.0249)	0.0227	(0.0150)	0.2541	(0.1493)
	$\tilde{m}_{B,0.1}(x)$	0.0266	(0.0221)	0.0180	(0.0128)	0.0864	(0.0564)
	$\tilde{m}_{B,1}(x)$	0.0263	(0.0226)	0.0176	(0.0132)	0.0854	(0.0545)
	$\tilde{m}_{B,\infty}(x)$	0.0264	(0.0228)	0.0176	(0.0133)	0.0844	(0.0538)
200	$\hat{m}_E(x)$	0.0237	(0.0154)	0.0172	(0.0087)	0.3194	(0.0894)
	$\tilde{m}_{E,\lambda^*}(x)$	0.0248	(0.0167)	0.0165	(0.0095)	0.2976	(0.0810)
	$\hat{m}_B(x)$	0.0198	(0.0154)	0.0138	(0.0086)	0.0762	(0.0440)
	$\hat{m}_{B,C}(x)$	0.0229	(0.0178)	0.0143	(0.0096)	0.2796	(0.1363)
	$\tilde{m}_{B,0.1}(x)$	0.0193	(0.0157)	0.0131	(0.0087)	0.0754	(0.0426)
	$\tilde{m}_{B,1}(x)$	0.0190	(0.0162)	0.0126	(0.0090)	0.0735	(0.0417)
	$\tilde{m}_{B,\infty}(x)$	0.0191	(0.0163)	0.0127	(0.0091)	0.0725	(0.0407)
400	$\hat{m}_E(x)$	0.0169	(0.0097)	0.0129	(0.0059)	0.2879	(0.0874)
	$\tilde{m}_{E,\lambda^*}(x)$	0.0178	(0.0115)	0.0118	(0.0066)	0.2572	(0.0710)
	$\hat{m}_B(x)$	0.0151	(0.0101)	0.0108	(0.0055)	0.0639	(0.0350)
	$\hat{m}_{B,C}(x)$	0.0148	(0.0112)	0.0094	(0.0057)	0.2886	(0.1315)
	$\tilde{m}_{B,0.1}(x)$	0.0145	(0.0103)	0.0099	(0.0056)	0.0642	(0.0337)
	$\tilde{m}_{B,1}(x)$	0.0143	(0.0108)	0.0095	(0.0058)	0.0619	(0.0329)
	$\tilde{m}_{B,\infty}(x)$	0.0143	(0.0109)	0.0095	(0.0059)	0.0612	(0.0319)

Table 4
Monte Carlo Results (Beta Case; CV).

n	Estimator	RISE		IAD		Tuning Parameter	
		Ave	SD	Ave	SD	Ave	SD
100	$\hat{m}_B(x)$	0.0540	(0.0976)	0.0311	(0.0410)	0.0539	(0.0331)
	$\tilde{m}_{B,1}(x)$	0.0466	(0.0601)	0.0278	(0.0272)	0.0726	(0.0381)
	$\tilde{m}_{B,\infty}(x)$	0.0472	(0.0726)	0.0280	(0.0328)	0.0728	(0.0372)
200	$\hat{m}_B(x)$	0.0382	(0.0493)	0.0214	(0.0195)	0.0394	(0.0217)
	$\tilde{m}_{B,1}(x)$	0.0304	(0.0304)	0.0179	(0.0134)	0.0613	(0.0275)
	$\tilde{m}_{B,\infty}(x)$	0.0306	(0.0306)	0.0180	(0.0135)	0.0603	(0.0271)
400	$\hat{m}_B(x)$	0.0310	(0.0511)	0.0162	(0.0142)	0.0294	(0.0151)
	$\tilde{m}_{B,1}(x)$	0.0219	(0.0178)	0.0129	(0.0079)	0.0518	(0.0211)
	$\tilde{m}_{B,\infty}(x)$	0.0221	(0.0188)	0.0129	(0.0082)	0.0508	(0.0207)

Estimators for $m(x)$ include $\hat{m}_E(x)$ (the Epanechnikov LL estimator), $\tilde{m}_{E,\lambda^*}(x)$ (the Epanechnikov skewed estimator with $\lambda^* = 0.0352$), $\hat{m}_B(x)$ (the beta LL estimator), $\hat{m}_{B,C}(x)$ (the beta local cubic estimator), $\tilde{m}_{B,0.1}(x)$ (the beta skewed estimator with $\lambda = 0.1$), $\tilde{m}_{B,1}(x)$ (the beta skewed estimator with $\lambda = 1$), and $\tilde{m}_{B,\infty}(x)$ (the limit beta skewed estimator). If a beta skewed estimator is not well-defined for x in the vicinity of 0 or 1, then it is replaced by the beta LL estimator evaluated at the same design point. To implement all these estimators, we choose their tuning parameters via the oracle and CV methods as before.

The global performance measures that assess an estimator include the RISE and IAD, where each integral is approximated by the trapezoidal rule on an equally-spaced grid of 101 points over the interval [0, 1]. As the local performance measures, the squared bias and variance evaluated at eleven design points $x \in \{0.0, 0.1, \ldots, 0.9, 1.0\}$ over 1000 Monte Carlo samples are computed.

Results of global performance measures are summarized in Tables 3 (oracle tuning parameters) and 4 (CV tuning parameters). To save space, we only report Fig. 4, which presents local measures of the estimators using oracle tuning parameters, as in Chen (2002).

The results in Tables 3 and 4 confirm our findings in Tables 1 and 2. Each estimator numerically indicates consistency in that their global performance measures shrink with the sample size. The three beta skewed estimators outperform the others. Again, performances of $\tilde{m}_{B,1}(x)$ and $\tilde{m}_{B,\infty}(x)$ look alike and are superior to those of $\tilde{m}_{B,0.1}(x)$. The only difference is that the beta cubic estimator does not perform well for $n = 100, 200$. Its performances are rather close to those of two Epanechnikov estimators. We now check the local performance measures in Fig. 4. As in Fig. 1, the vertical axis in each panel of Fig. 4 is the natural logarithm of either squared bias or variance. The figure shows that the beta cubic estimator has the smallest bias over almost the entire unit interval, whereas the limit beta skewed estimator attains the smallest variance around [0.4, 1]. The beta cubic estimator tends to have by far the largest variance around [0.1, 0.4], which appears

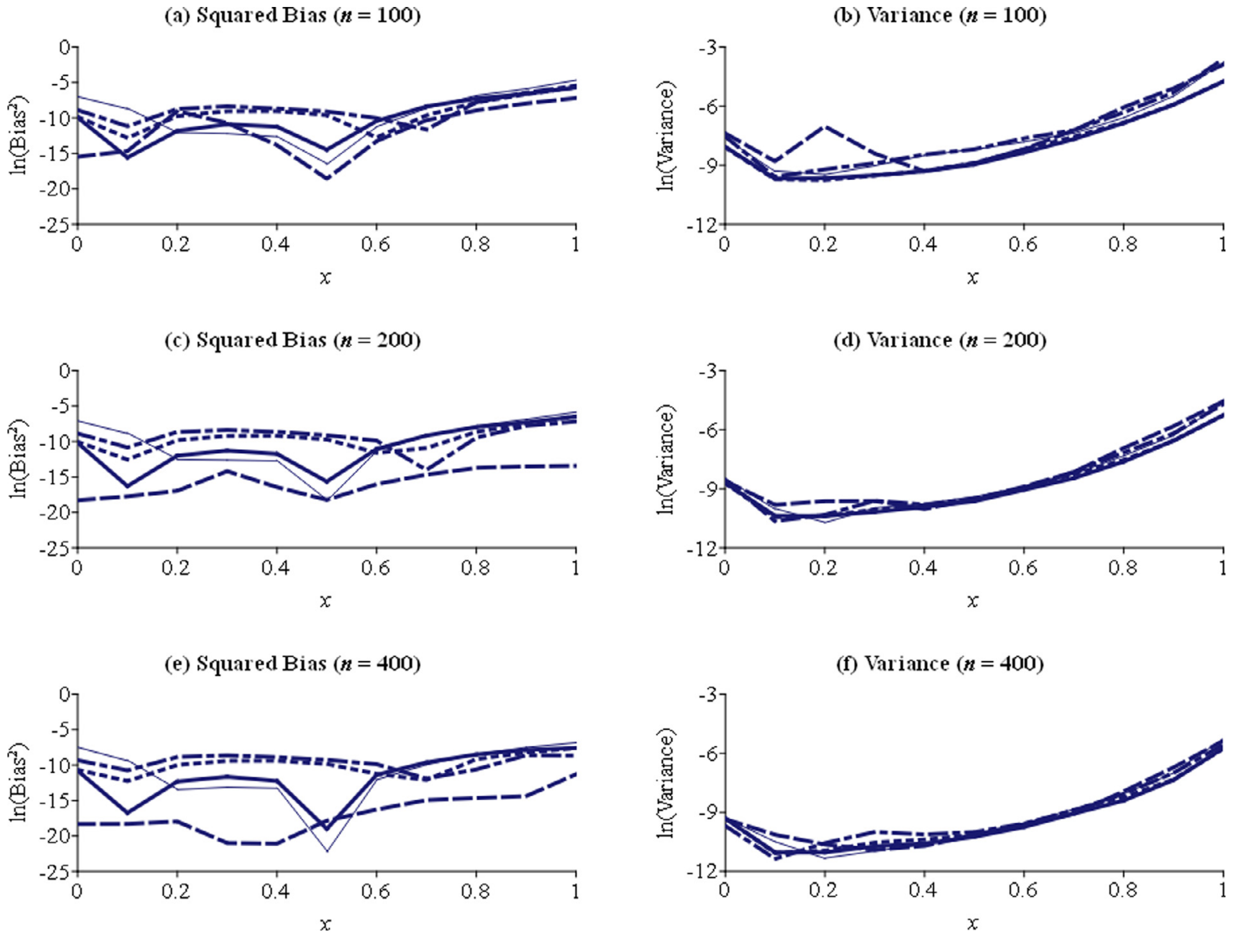

Fig. 4. Squared Bias and Variance of Curve Estimates (Beta Case; Oracle).
Note: Dash-dot, thin solid, dot, dash, and solid lines correspond to the Epanechnikov LL, Epanechnikov skewed, beta LL, beta local cubic, and limit beta skewed estimates, respectively.

to result in its mediocre global performance measures for $n = 100, 200$.[8] Furthermore, Table 4 indicates that to implement the beta skewed estimation, CV works well in practice.

4. An Empirical Illustration

4.1. The Model

In this section the skewed estimators are applied to a real data set. Inspired by Tripathi and Kim (2003), we concentrate on the problem of estimating a production function with homogeneity of degree one (or linear homogeneity) imposed. Suppose that there is the technological relationship in which two factors, namely, capital and labor inputs (K, L), can produce output Y. This relationship can be expressed as a production function and formulated in a regression format as

$$g(K, L) = E(Y|K, L) \Rightarrow Y = g(K, L) + \varepsilon$$

for an unknown function g and the regression error ε. With no further restrictions on g, we would have to run a multiple nonparametric regression of Y on $(K, L) = (k, l)$ to estimate $g(k, l)$. However, as is often the case with estimating production functions, if it is known *a priori* (or it may be reasonably assumed) that g is homogeneous of degree one in (K, L) (i.e., the technology exhibits constant returns to scale), then we can rewrite the estimation problem as that of a simple regression.

[8] It seems that the mediocre performance of the beta cubic estimator is attributed to sparseness of the data on the right boundary. The lesson learnt here is that a guard against sparseness such as the one used in Choi and Hall (1998, Section 3.2) should be considered to stabilize local polynomial estimation, regardless of the choice of a kernel.

Table 5
Summary Statistics (in millions).

Variable	Ave	SD	SK	Min	Med	Max
W	0.1319	0.1729	4.1567	0.0094	0.0793	1.5839
Z	0.1181	0.0770	2.8297	0.0281	0.0960	0.6552

Note: Ave = average; SD = standard deviation; SK = skewness; Min = minimum; Med = median; and Max = maximum.

Specifically, the linear homogeneity implies

$$g(K, L) = L \cdot g\left(\frac{K}{L}, 1\right) := L \cdot m\left(\frac{K}{L}\right)$$

so that the original regression can be reformulated as

$$\frac{Y}{L} = m\left(\frac{K}{L}\right) + \frac{\varepsilon}{L} \Rightarrow Z = m(W) + \epsilon$$

by letting $(Z, W, \epsilon) := (Y/L, K/L, \varepsilon/L)$. The entire estimation procedure then takes two steps. The first step is to run a simple nonparametric regression of Z on $W = w = k/l$ to estimate $m(w)$. Let $\bar{m}(w)$ be a nonparametric estimate of $m(w)$. Then, in the second step, it follows from the relation $w = k/l$ that $g(k, l)$ can be estimated as $\bar{g}(k, l) = l \cdot \bar{m}(w)$. Tripathi and Kim (2003) investigate LL estimation of $m(w)$ using standard symmetric kernels and develop statistical properties of the estimator.

4.2. The Data

The National Bureau of Economic Research ("NBER") makes an industry-level panel data set public on its web page (http://www.nber.org/nberces/). The data set, the NBER-CES Manufacturing Industry Database, is provided in two versions, depending on the classification of manufacturing industries; see Bartelsman and Gray (1996) and Becker et al. (2016) for more details of the database. We choose the version with industries classified by the six-digit level of the North American Industry Classification System ("NAICS") in 1997.

The 1997 NAICS version contains annual industry-level data of 473 industries from 1958 to 2011. The sample in 2000 is used throughout. Because the database includes a few candidates for each of Y, K and L, we follow Kumbhakar et al. (2012), for instance, and define *VADD* (= total value added in \$1m), *CAP* (= total real capital stock in \$1m) and *EMP* (= total employment in 1000s) as Y, K and L, respectively. Subsequently two variables $(W, Z) = (K/L, Y/L)$ can be constructed.

Before proceeding, we find a few observations with extremely large values of Z. Because they are highly likely to influence our estimation results, it is desirable to eliminate them as outliers. For this purpose, we employ the generalized boxplot by Bruffaerts et al. (2014).[9] With its aid, two observations are detected as outliers and eliminated from the sample. The eliminated industries are "Flavoring Syrup and Concentrate Manufacturing" (NAICS Code: 311930) with $(W, Z) = (0.24, 1.18)$ and "Cigarette Manufacturing" (NAICS Code: 312221) with $(W, Z) = (0.43, 2.19)$. As a result, the sample size is 471.

Table 5 presents summary statistics of (W, Z), where their units of measurement are converted in millions. The regressor W is nonnegative by construction and has a positive skewness; indeed, the scatter plot of data points in Fig. 5 indicates concentration of observations of W in the vicinity of the origin and a long right-tail of its marginal distribution. Asymmetric kernel smoothing is likely to work well with such data, and thus we are motivated to estimate the unknown curve $m(w)$ by our gamma skewed estimators.

4.3. Estimation Results

In what follows, the gamma skewed with $\lambda = 1$ and limit gamma skewed estimators, as well as the gamma LL estimator, are employed for the curve estimation. Their smoothing parameter values are chosen via CV. The values for the gamma LL, gamma skewed with $\lambda = 1$, and limit gamma skewed estimators are 0.0733, 0.0818 and 0.0843, respectively.

Furthermore, researchers often specify the production function g in a simple, parametric form. The Cobb-Douglas production function (Cobb and Douglas, 1928) is among the most popular ones. When it is homogenous of degree one, it can be expressed in a multiplicative error regression model as

$$Y = AK^{\alpha}L^{1-\alpha}u,$$

where $A > 0$ and $\alpha \in (0, 1)$ are parameters and $u > 0$ is the regression error. A simple manipulation yields

$$\ln\left(\frac{Y}{L}\right) = \ln A + \alpha \ln\left(\frac{K}{L}\right) + \ln u \Rightarrow \ln Z = \beta_0 + \beta_1 \ln W + v$$

[9] The generalized boxplot is designed for skewed and/or heavy-tailed distributions and thus fits into our context. It is available on Stata as the command `gboxplot`.

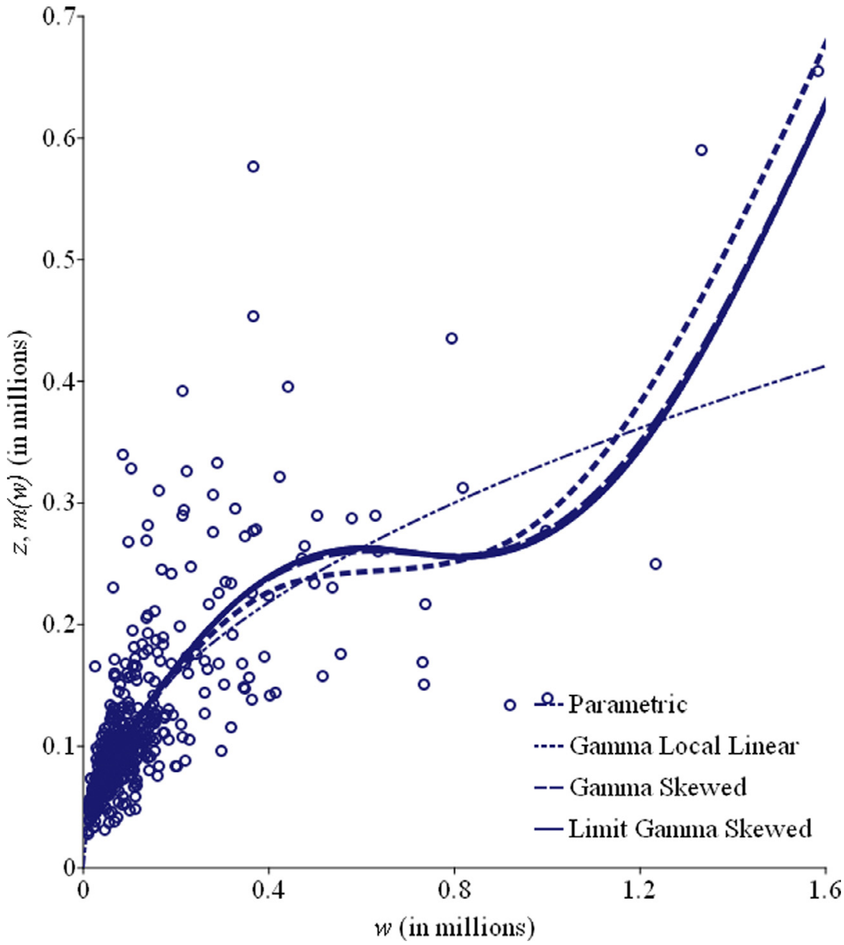

Fig. 5. Estimates of a Linear Homogeneous Production Function.

by letting $(\beta_0, \beta_1, v) := (\ln A, \alpha, \ln u)$, and parameters $(\beta_0,\ \beta_1)$ can be estimated by running the ordinary least squares ("OLS") from the regression of $\ln Z$ on $(1, \ln W)$. The OLS estimates of $(\beta_0,\ \beta_1)$ are $\left(\hat{\beta}_0, \hat{\beta}_1\right) = (-1.1537, 0.4590)$. We then employ the swearing estimate by Duan (1983) to back-transform $\hat{Z}$ from $\widehat{\ln Z}$.[10] More specifically, the final parametric estimate of $m(w)$ is $\hat{m}_{CD}(w) = \hat{\gamma}\exp\left(\hat{\beta}_0\right)\cdot w^{\hat{\beta}_1}$, where $\hat{\gamma}$ is the average of $\left\{\exp\left(\ln Z_i - \hat{\beta}_0 - \hat{\beta}_1 \ln W_i\right)\right\}_{i=1}^{n}$.

A parametric and three nonparametric curve estimates of $m(w)$, as well as a scatter plot of data points, are drawn in Fig. 5. It can be immediately found that shapes of three nonparametric estimates considerably differ from the one of the parametric one, which is globally concave by construction. In contrast, each of the nonparametric estimates appears to be roughly monotone with an inflection point from concavity to convexity around $w = 0.8$. Moreover, as deduced from the Monte Carlo results in Section 3, two skewed estimates are nearly identical. Although it is impossible to judge which estimate is closest to the true curve, the substantial difference between nonparametric and parametric estimates indicates that the parametric, Cobb-Douglas specification may be incorrect.

5. Conclusion

This paper has demonstrated that under a suitable choice of weights and shifts, the skewing method by Choi and Hall (1998) can accelerate bias convergences of LL regression estimators using asymmetric kernels from $O(b)$ to $O(b^2)$ as $b \to 0$ while not inflating their variances in an order of magnitude. A remarkable difference can be found in the choice of the shift that can lead to a faster bias convergence when symmetric and asymmetric kernels are employed. While for a given weight λ, the shift is constant regardless of the position of the design point for symmetric kernels, it varies with the design point

[10] We have also applied the back-transformation method proposed in Chapter 6 of Wooldridge (2013), which is valid under the assumption of homoscedastic normal errors. Because the result is qualitatively similar, it is unreported in Fig. 5.

for asymmetric kernels. Two versions of the skewed estimator (i.e., the case with a fixed λ and the limit case with $\lambda \to \infty$) are studied, and their large- and finite-sample properties are explored. In particular, Monte Carlo results indicate that the gamma (beta) skewed with $\lambda = 1$ and limit gamma (beta) skewed estimators have equally nice finite-sample properties. In fact, their finite-sample performances are superior to those of the Epanechnikov LL, Epanechnikov skewed and gamma (beta) LL estimators, and they are comparable with those of the gamma (beta) local cubic estimator. CV is found to be useful for implementing the skewed estimation. Finally, the skewed estimator is applied to the US manufacturing industry-level data, and a linear-homogeneous production function is estimated nonparametrically. We find that the nonparametric curve estimate substantially differs from the one implied by the popular, Cobb-Douglas specification.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgements

The authors would like to thank the co-editor Ana Colubi, an associate editor, two anonymous referees, Artem Prokhorov, and the participants of CMStatistics 2018, DAGStat 2019 and EcoSta 2019 for their constructive comments and suggestions. The first author gratefully acknowledges financial support from the German Science Foundation (DFG) through the Collaborative Research Center SFB 823. The second author gratefully acknowledges financial support from Japan Society for the Promotion of Science (grant numbers 15K03405 and 19K01595).

A Appendix

A.1. A Useful Lemma

Before proceeding, we summarize useful results on the moments of certain gamma and beta random variables as a lemma, which is frequently applied for the bias and variance approximations.

A.1.1. Lemma A.1

(i) *Let* $\theta_{jx} \overset{d}{=} G(jx/b+1, b/j)$ *for* $x \in \mathbb{R}_+$ *and* $j = 1, 2$. *Then, the moments* $\nu_{j,r} = E\left(\theta_{jx} - x\right)^r$ *for* $r = 1, 2, \ldots,$ *are:*

$$\nu_{j,1} = \frac{b}{j};$$
$$\nu_{j,2} = \frac{b}{j^2}(jx+2b);$$
$$\nu_{j,3} = \frac{b^2}{j^3}(5jx+6b);$$
$$\nu_{j,4} = \frac{b^2}{j^4}\left(3j^2x^2+26jxb+24b^2\right); \text{ and}$$
$$\nu_{j,r} = O\left(b^3\right) \text{ for } r \geq 5.$$

(ii) *Let* $\eta_{jx} \overset{d}{=} Beta\{jx/b+1, j(1-x)/b+1\}$ *for* $x \in [0, 1]$ *and* $j = 1, 2$. *Then, as* $b \to 0$, *the moments* $\upsilon_{j,r} = E\left(\eta_{jx} - x\right)^r$ *for* $r = 1, 2, \ldots,$ *can be approximated by:*

$$\upsilon_{j,1} = \left(\frac{1-2x}{j}\right)b - \left\{\frac{2(1-2x)}{j^2}\right\}b^2 + O\left(b^3\right);$$
$$\upsilon_{j,2} = \left\{\frac{x(1-x)}{j}\right\}b + \left(\frac{2-11x+11x^2}{j^2}\right)b^2 + O\left(b^3\right);$$
$$\upsilon_{j,3} = \left\{\frac{5x(1-x)(1-2x)}{j^2}\right\}b^2 + O\left(b^3\right);$$
$$\upsilon_{j,4} = \left\{\frac{3x^2(1-x)^2}{j^2}\right\}b^2 + O\left(b^3\right); \text{ and}$$
$$\upsilon_{j,r} = O\left(b^3\right) \text{ for } r \geq 5.$$

A.1.2. Proof of Lemma A.1

These results immediately follow from evaluating the moments of the gamma and beta random variables (for (i) and (ii)) and expanding them around $b = 0$ (for (ii) only). □

A.2. Proof of Theorem 1

The proof closely follows the Proof of Theorem 2.1 in Choi (1998). For brevity, the subscript G and the design point x are suppressed wherever no confusions may occur. In addition, we introduce short-handed notations such as $\tilde{m}_{\lambda,\ell} = \tilde{m}_{G,\lambda,\ell}(x)$, $\hat{m}^* = \hat{m}_G(x|x^*)$, $\mu_p = E\{s_{Gp}(x)\}$, $\mu_p^* = E\{s_{Gp}(x^*)\}$, $m^{(p)} = m^{(p)}(x)$, $m^{(p)*} = m^{(p)}(x^*)$, $f^{(p)} = f^{(p)}(x)$, $f^{(p)*} = f^{(p)}(x^*)$, and so on, where $s_{Gp}(x)$ is defined in (1). The proof takes the following steps:

1. Approximate the bias of each of three LL estimators, conditional on $X = x$, up to the $O(b^2)$ term.
2. Find the weights (λ_1, λ_2) and shifts (ℓ_1, ℓ_2) that can eliminate the $O(b)$ and $O(b^{3/2})$ terms in $Bias(\tilde{m}_{\lambda,\ell})$.
3. Substitute the results in Step 2 into the $O(b^2)$ term in $Bias(\tilde{m}_{\lambda,\ell})$.

Step 1. By a similar argument to the proof of Theorem 1 in Chen (2002), we can expand $E(\hat{m}^*)$ as

$$E(\hat{m}^*) = m^* + (x - x^*)m^{(1)*} + \frac{Q^* + R^*}{\mu_0^*\mu_2^* - \mu_1^{*2}} + O(n^{-1}b^{-1}),$$

where $R^* = R(x^*)$ is the remainder term in a smaller order of magnitude by Taylor's theorem, and $Q^* = Q(x^*)$ is in the form of

$$\begin{aligned} Q^* = {} & \frac{m^{(2)*}}{2}\left\{\left(\mu_2^{*2} - \mu_1^*\mu_3^*\right) + (x - x^*)(\mu_0^*\mu_3^* - \mu_1^*\mu_2^*)\right\} \\ & + \frac{m^{(3)*}}{6}\{(\mu_2^*\mu_3^* - \mu_1^*\mu_4^*) + (x - x^*)(\mu_0^*\mu_4^* - \mu_1^*\mu_3^*)\} \\ & + \frac{m^{(4)*}}{24}\{(\mu_2^*\mu_4^* - \mu_1^*\mu_5^*) + (x - x^*)(\mu_0^*\mu_5^* - \mu_1^*\mu_4^*)\}. \end{aligned}$$

Moreover, μ_p^* admits the expansion

$$\mu_p^* = f^*\nu_{1,p}^* + f^{(1)*}\nu_{1,p+1}^* + \frac{f^{(2)*}}{2}\nu_{1,p+2}^* + O(\nu_{1,p+3}^*),$$

where $\nu_{1,r}^* = E(\theta_{x^*} - x^*)^r$ for $\theta_{x^*} \stackrel{d}{=} G(x^*/b + 1, b)$. Then, it follows from Lemma A.1(i) that approximations to μ_p^* for $p = 0, 1, \ldots, 5$ are:

$$\begin{aligned} \mu_0^* &= f^* + \left(f^{(1)*} + \frac{x^*}{2}f^{(2)*}\right)b + O(b^2); \\ \mu_1^* &= \left(f^* + x^*f^{(1)*}\right)b + \left(2f^{(1)*} + \frac{5}{2}x^*f^{(2)*}\right)b^2 + O(b^3); \\ \mu_2^* &= x^*f^*b + \left(2f^* + 5x^*f^{(1)*} + \frac{3}{2}x^{*2}f^{(2)*}\right)b^2 + O(b^3); \\ \mu_3^* &= \left(5x^*f^* + 3x^{*2}f^{(1)*}\right)b^2 + O(b^3); \\ \mu_4^* &= 3x^{*2}f^*b^2 + O(b^3); \text{ and} \\ \mu_5^* &= O(b^3). \end{aligned}$$

Putting $x^* = x + \ell b^{1/2}$, taking Taylor expansions of $m^{(p)*}$ and μ_p^* around $x^* = x$ (i.e., $\ell b^{1/2} = 0$), doing some tedious but straightforward calculations, and recognizing that $O(n^{-1}b^{-1}) = o(b^2)$ due to Assumption 3, we finally have

$$\begin{aligned} E(\hat{m}^*) = {} & m + \frac{1}{2}(x - \ell^2)m^{(2)}b - \ell\left\{\left(\frac{3}{2} + x\frac{f^{(1)}}{f}\right)m^{(2)} + \frac{\ell^2}{3}m^{(3)}\right\}b^{3/2} \\ & + \left[-\left\{1 + x\frac{f^{(1)}}{f} - \frac{x^2}{2}\frac{f^{(2)}}{f} + x^2\left(\frac{f^{(1)}}{f}\right)^2 + \ell^2\left(\frac{f^{(1)}}{f} + x\left(\frac{f^{(2)}}{f} - \left(\frac{f^{(1)}}{f}\right)^2\right)\right)\right\}m^{(2)}\right. \\ & \left. + \left\{\frac{x}{3} - \ell^2\left(2 + x\frac{f^{(1)}}{f}\right)\right\}m^{(3)} + \left(\frac{x^2}{8} - \frac{\ell^2}{4}x - \frac{\ell^4}{8}\right)m^{(4)}\right]b^2 + o(b^2). \end{aligned}$$

The bias expansions of $\hat{m}_G(x|x - \ell_1 b^{1/2})$, $\hat{m}_G(x|x)$ and $\hat{m}_G(x|x + \ell_2 b^{1/2})$ immediately follow from letting $\ell = -\ell_1, 0, \ell_2$ on the right-hand side, respectively. In particular, a second-order bias expansion of the gamma LL estimator $\hat{m}_G(x|x) = \hat{m}_G(x)$

takes the form of

$$Bias\left\{\hat{m}_G(x)\right\} = \frac{x}{2}m^{(2)}b + \left[-\left\{1 + x\frac{f^{(1)}}{f} - \frac{x^2}{2}\frac{f^{(2)}}{f} + x^2\left(\frac{f^{(1)}}{f}\right)^2\right\}m^{(2)} + \frac{x}{3}m^{(3)} + \frac{x^2}{8}m^{(4)}\right]b^2 + o\left(b^2\right).$$

Step 2. By the results in Step 1,

$$\begin{aligned}
E\left(\tilde{m}_{\lambda,\ell}\right) &= m + \frac{1}{2(\lambda_1 + 1 + \lambda_2)}\left\{(\lambda_1 + 1 + \lambda_2)x - \left(\lambda_1\ell_1^2 + \lambda_2\ell_2^2\right)\right\}m^{(2)}b \\
&+ \frac{1}{\lambda_1 + 1 + \lambda_2}\left\{(\lambda_2\ell_2 - \lambda_1\ell_1)\left(\frac{3}{2} + x\frac{f^{(1)}}{f}\right)m^{(2)} + \frac{1}{3}\left(\lambda_2\ell_2^3 - \lambda_1\ell_1^3\right)m^{(3)}\right\}b^{3/2} \\
&+ \left[-\left\{1 + x\frac{f^{(1)}}{f} - \frac{x^2}{2}\frac{f^{(2)}}{f} + x^2\left(\frac{f^{(1)}}{f}\right)^2\right\}m^{(2)} + \frac{x}{3}m^{(3)} + \frac{x^2}{8}m^{(4)}\right. \\
&- \frac{1}{\lambda_1 + 1 + \lambda_2}\left\{\left(\lambda_1\ell_1^2 + \lambda_2\ell_2^2\right)\left(\left(\frac{f^{(1)}}{f} + x\left(\frac{f^{(2)}}{f} - \left(\frac{f^{(1)}}{f}\right)^2\right)\right)m^{(2)}\right.\right. \\
&\left.\left.+\left(2 + x\frac{f^{(1)}}{f}\right)m^{(3)}\right) + \left(\lambda_1\ell_1^2\left(\frac{x}{4} + \frac{\ell_1^2}{8}\right) + \lambda_2\ell_2^2\left(\frac{x}{4} + \frac{\ell_2^2}{8}\right)\right)m^{(4)}\right\}\right]b^2 + o\left(b^2\right).
\end{aligned}$$

Therefore, both $O(b)$ and $O(b^{3/2})$ terms vanish if and only if

$$\begin{cases} (\lambda_1 + 1 + \lambda_2)x - \left(\lambda_1\ell_1^2 + \lambda_2\ell_2^2\right) = 0 \\ \lambda_2\ell_2 - \lambda_1\ell_1 = 0 \\ \lambda_2\ell_2^3 - \lambda_1\ell_1^3 = 0 \end{cases}.$$

Solving this system for $(\lambda_1, \lambda_2, \ell_1, \ell_2)$ yields

$$\lambda_1 = \lambda_2 = \lambda \text{ and } \ell_1 = \ell_2 = \ell = \ell(\lambda, x) = \sqrt{\left(\frac{2\lambda + 1}{2\lambda}\right)x}.$$

Step 3. Substituting (λ, ℓ) into the coefficient on b^2 in $Bias\left(\tilde{m}_{\lambda,\ell}\right)$ finally gives

$$\begin{aligned}
&-\left\{1 + x\frac{f^{(1)}}{f} - \frac{x^2}{2}\frac{f^{(2)}}{f} + x^2\left(\frac{f^{(1)}}{f}\right)^2\right\}m^{(2)} + \frac{x}{3}m^{(3)} + \frac{x^2}{8}m^{(4)} - x\left[\left\{\frac{f^{(1)}}{f} + x\left(\frac{f^{(2)}}{f} - \left(\frac{f^{(1)}}{f}\right)^2\right)\right\}m^{(2)}\right. \\
&\left.+\left(2 + x\frac{f^{(1)}}{f}\right)m^{(3)} + x\left(\frac{1}{4} + \frac{2\lambda + 1}{16\lambda}\right)m^{(4)}\right] \\
&= -\left\{\left(1 + 2x\frac{f^{(1)}}{f} + \frac{x^2}{2}\frac{f^{(2)}}{f}\right)m^{(2)} + x\left(\frac{5}{3} + x\frac{f^{(1)}}{f}\right)m^{(3)} + \frac{x^2}{4}\left(1 + \frac{1}{4\lambda}\right)m^{(4)}\right\}.
\end{aligned}$$

Denoting the right-hand side by $\mathcal{B}_{G,\lambda}(x; m, f)$ completes the proof. □

A.3. Proof of Corollary 1

This has been already established in Step 1 of the proof of Theorem 1. □

A.4. Proof of Theorem 2

The proof closely follows the Proof of Theorem 2.2 in Choi (1998). In this proof, we mean by the symbols "*Var*" and "*Cov*" the conditional variance and covariance, conditional on $X = x$, respectively. In addition to the short-handed notations in the proof of Theorem 1, we use short-handed notations such as $\tilde{m}_\lambda = \tilde{m}_{G,\lambda}(x)$, $v^0 = Var\left\{\hat{m}_G(x|x)\right\}$, $c^{*\dagger} = Cov\left\{\hat{m}_G(x|x^*), \hat{m}_G\left(x|x^\dagger\right)\right\}$, $\left(\hat{\alpha}^*, \hat{\beta}^*\right) = \left(\hat{\alpha}_G(x^*), \hat{\beta}_G(x^*)\right)$, $s_p^* = s_{Gp}(x^*)$, $t_p^* = t_{Gp}(x^*)$, $\sigma^2 = \sigma^2(x)$, and so on, where $\left(x^*, x^\dagger\right) = \left(x + \ell b^{1/2}, x - \ell b^{1/2}\right)$ and $s_{Gp}(x)$ and $t_{Gp}(x)$ are defined in (1) and (2), respectively. Then,

$$Var(\tilde{m}_\lambda) = \frac{\lambda^2 v^* + v^0 + \lambda^2 v^\dagger + 2\lambda c^{0*} + 2\lambda c^{0\dagger} + 2\lambda^2 c^{*\dagger}}{(2\lambda + 1)^2}.$$

We start from working on the case for interior x and concentrate on approximating v^*. Observe that

$$v^* = Var\left(\hat{\alpha}^*\right) + \ell^2 b Var\left(\hat{\beta}^*\right) - 2\ell b^{1/2} Cov\left(\hat{\alpha}^*, \hat{\beta}^*\right) := A_1 + A_2 - 2A_3 \text{ (say)},$$

where

$$A_1 = \left(\mu_0^*\mu_2^* - \mu_1^{*2}\right)^{-2}\left\{\mu_2^{*2}Var(t_0^*) + \mu_1^{*2}Var(t_1^*) - 2\mu_1^*\mu_2^*Cov(t_0^*, t_1^*)\right\} + R_1,$$
$$A_2 = \left(\mu_0^*\mu_2^* - \mu_1^{*2}\right)^{-2}\ell^2 b\left\{\mu_0^{*2}Var(t_1^*) + \mu_1^{*2}Var(t_0^*) - 2\mu_0^*\mu_1^*Cov(t_0^*, t_1^*)\right\} + R_2,$$
$$A_3 = \left(\mu_0^*\mu_2^* - \mu_1^{*2}\right)^{-2}\ell b^{1/2}\left\{\left(\mu_0^*\mu_2^* + \mu_1^{*2}\right)Cov(t_0^*, t_1^*) - \mu_1^*\mu_2^*Var(t_0^*) - \mu_0^*\mu_1^*Var(t_1^*)\right\} + R_3,$$

and it can be shown that the remainder terms R_1, R_2 and R_3 are in smaller orders of magnitude. By the proof of Theorem 1 and Assumption 3,

$$\mu_0^* \sim f^*, \mu_1^* \sim \left(f^* + x^* f^{(1)*}\right)b \text{ and } \mu_2^* \sim x^* f^* b$$

so that

$$\mu_0^*\mu_2^* - \mu_1^{*2} \sim x^* f^{*2} b.$$

In addition,

$$Cov\left(t_p^*, t_q^*\right) = \frac{1}{n}E\left\{\sigma^2(X)(X - x^*)^{p+q}K^2_{G(x^*,b)}(X)\right\}\{1 + o(1)\},$$

where

$$E\left\{\sigma^2(X)(X - x)^{p+q}K^2_{G(x,b)}(X)\right\} = B_b(x)E\left\{(\theta_{2x} - x)^{p+q}\sigma^2(\theta_{2x})f(\theta_{2x})\right\},$$

and

$$B_b(x) := \frac{\Gamma(2x/b + 1)}{b2^{2x/b+1}\Gamma^2(x/b + 1)} \sim \frac{b^{-1/2}}{2\sqrt{\pi}\sqrt{x}}$$

by Chen (2000b). It also follows from Lemma A.1(i) that

$$E\left\{(\theta_{2x} - x)^{p+q}\sigma^2(\theta_{2x})f(\theta_{2x})\right\} \sim \begin{cases} \sigma^2 f & \text{if } p+q=0 \\ \frac{1}{2}\left\{\sigma^2 f + x\left(\sigma^2 f\right)^{(1)}\right\}b & \text{if } p+q=1 \\ \frac{x}{2}\sigma^2 f b & \text{if } p+q=2 \end{cases}.$$

After combining these results, taking a Taylor expansion around $x^* = x$ (i.e., $\ell b^{1/2} = 0$) and doing some tedious calculations, we can find that

$$A_1 = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\{1 + o(1)\},$$
$$A_2 = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(\frac{\ell^2}{2x}\right)\{1 + o(1)\},$$

and $A_3 = o\left(n^{-1}b^{-1/2}\right)$. Therefore,

$$v^* = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(1 + \frac{\ell^2}{2x}\right)\{1 + o(1)\}.$$

Expanding other terms similarly, we have the following results for interior x:

$$v^\dagger = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(1 + \frac{\ell^2}{2x}\right)\{1 + o(1)\};$$
$$v^0 = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\{1 + o(1)\};$$
$$c^{0*} = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(1 + \frac{\ell^2}{2x}\right)\{1 + o(1)\};$$
$$c^{0\dagger} = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(1 + \frac{\ell^2}{2x}\right)\{1 + o(1)\}; \text{ and}$$
$$c^{*\dagger} = \frac{1}{nb^{1/2}}\frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\left(1 + \frac{3\ell^2}{2x} + \frac{\ell^4}{x^2}\right)\{1 + o(1)\}.$$

Using these and $\ell = \sqrt{(2\lambda + 1)x/(2\lambda)}$ finally establishes that for interior x,

$$
\begin{aligned}
Var(\tilde{m}_\lambda) &= \frac{1}{nb^{1/2}} \frac{\sigma^2}{2\sqrt{\pi}\sqrt{x}f}\{1+o(1)\}\frac{1}{(2\lambda+1)^2}\left\{\lambda^2\left(1+\frac{\ell^2}{2x}\right)+1+\lambda^2\left(1+\frac{\ell^2}{2x}\right)\right.\\
&\quad \left. + 2\lambda\left(1+\frac{\ell^2}{2x}\right)+2\lambda\left(1+\frac{\ell^2}{2x}\right)+2\lambda^2\left(1+\frac{3\ell^2}{2x}+\frac{\ell^4}{x^2}\right)\right\}\\
&= \frac{1}{nb^{1/2}} \frac{5\sigma^2}{4\sqrt{\pi}\sqrt{x}f}\{1+o(1)\}.
\end{aligned}
$$

On the other hand, for boundary x, each of $B_b(x)$, $B_b(x^*)$ and $B_b(x\dagger)$ are shown to be $O\left(b^{-1}\right)$, where the coefficient on each $O\left(b^{-1}\right)$ term depends on (κ, λ). Modifying the expansions of μ_p and $E\left\{(\theta_{2x}-x)^{p+q}\sigma^2(\theta_{2x})f(\theta_{2x})\right\}$ in light of $x = \kappa b\{1+o(1)\}$ leads to $Var(\tilde{m}_\lambda) = O\left(n^{-1}b^{-1}\right)$. □

A.5. Proof of Proposition 1

The results can be obtained by taking limits of the corresponding parts as $\lambda \to \infty$ in the proofs of Theorems 1 and 2. □

A.6. Proof of Theorem 3

Replacing Part (i) of Lemma A.1 in the proofs of Theorems 1 and 2 with Part (ii) and doing some tedious but straightforward calculations can establish the theorem. □

A.7. Proof of Corollary 2

This can be established in the middle of the proof of Theorem 3. □

A.8. Proof of Proposition 2

The proof strategy is exactly the same as in the proof of Proposition 1. □

References

Bartelsman, E.J., Gray, W., 1996. The NBER manufacturing productivity database. NBER Technical Working Paper No. 205.
Becker, R., Gray, W., Marvakov, J., 2016. NBER-CES manufacturing industry database: Technical notes. Available at http://www.nber.org/nberces/nberces5811/nberces_5811_technical_notes.pdf.
Bruffaerts, C., Verardi, V., Vermandele, C., 2014. A generalized boxplot for skewed and heavy-tailed distributions. Statistics & Probability Letters 95, 110–117.
Chen, S.X., 1999. Beta kernel estimators for density functions. Computational Statistics & Data Analysis 31, 131–145.
Chen, S.X., 2000a. Beta kernel smoothers for regression curves. Statistica Sinica 10, 73–91.
Chen, S.X., 2000b. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 52, 471–480.
Chen, S.X., 2002. Local linear smoothers using asymmetric kernels. Annals of the Institute of Statistical Mathematics 54, 312–323.
Choi, E., 1998. Some Problems in Curve and Surface Estimation. Australian National University Ph.D. dissertation.
Choi, E., Hall, P., 1998. On the bias reduction in local linear smoothing. Biometrika 85, 333–345.
Cobb, C.W., Douglas, P.H., 1928. A theory of production. American Economic Review 18 (Supplement) 18, 139–165.
Duan, N., 1983. Swearing estimate: A nonparametric retransformation method. Journal of the American Statistical Association 78, 605–610.
Fan, J., 1993. Local linear regression smoothers and their minimax efficiencies. Annals of Statistics 21, 196–216.
Fan, J., Gijbels, I., 1992. Variable bandwidth and local linear regression smoothers. Annals of Statistics 20, 2008–2036.
Funke, B., Kawka, R., 2015. Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. Computational Statistics & Data Analysis 92, 148–162.
Gasser, T., Müller, H.G., 1979. Kernel estimation of regression functions. In: Gasser, T., Rosenblatt, M. (Eds.), Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2–4, 1979. Springer-Verlag, Berlin, 23–68.
Glad, I.K., 1998. Parametrically guided non-parametric regression. Scandinavian Journal of Statistics 25, 649–668.
Gospodinov, N., Hirukawa, M., 2012. Nonparametric estimation of scalar diffusion models of interest rates using asymmetric kernels. Journal of Empirical Finance 19, 595–609.
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., 2018. Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance. Springer, Singapore.
Hirukawa, M., Sakudo, M., 2014. Nonnegative bias reduction methods for density estimation using asymmetric kernels. Computational Statistics & Data Analysis 75, 112–123.
Hirukawa, M., Sakudo, M., 2015. Family of the generalised gamma kernels: A generator of asymmetric kernels for nonnegative data. Journal of Nonparametric Statistics 27, 41–63.
Jones, M.C., Henderson, D.A., 2007. Kernel-type density estimation on the unit interval. Biometrika 24, 977–984.
Jones, M.C., Linton, O., Nielsen, J.P., 1995. A simple bias reduction method for density estimation. Biometrika 82, 327–338.
Kumbhakar, S.C., Ortega-Argilés, R., Potters, L., Vivarelli, M., Voigt, P., 2012. Corporate R&D and firm efficiency: Evidence from Europe's top R&D investors. Journal of Productivity Analysis 37, 125–140.
Linton, O., Nielsen, J.P., 1994. A multiplicative bias reduction method for nonparametric regression. Statistics & Probability Letters 19, 181–187.
Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. Annals of Statistics 22, 1346–1370.
Scholz, M., Nielsen, J.P., Sperlich, S., 2015. Nonparametric prediction of stock returns based on yearly data: The long-term view. Insurance: Mathematics and Economics 65, 143–155.
Seifert, B., Gasser, T., 1996. Finite-sample variance of local polynomials: Analysis and solutions. Journal of the American Statistical Association 91, 267–275.
Stone, C.J., 1980. Optimal rates of convergence for nonparametric estimators. Annals of Statistics 8, 1348–1360.
Tripathi, G., Kim, W., 2003. Nonparametric estimation of homogeneous functions. Econometric Theory 19, 640–663.
Wooldridge, J.M., 2013. Introductory Econometrics: A Modern Approach, 5th edition, South-Western Cengage Learning, Mason, OH.