Averaged Singular Integral Estimation as a Bias Reduction Technique[1]

Miguel A. Delgado

Universidad Carlos III de Madrid, Madrid, Spain
E-mail: delgado@est-econ.uc3m.es

and

Jose M. Vidal-Sanz

London School of Economics, London, United Kingdom
E-mail: j.vidal-sanz@lse.ac.uk

This paper proposes an averaged version of singular integral estimators, whose bias achieves higher rates of convergence under smoothing assumptions. We derive exact bias bounds, without imposing smoothing assumptions, which are a basis for deriving the rates of convergence under differentiability assumptions.

AMS 2000 subject classification: 62G07.
Key words and phrases: global rates of convergence for the bias; singular integral estimators; bias reduction techniques; generalized jackknife.

1. INTRODUCTION

Let P be a probability function in $(\mathbb{R}^d, \mathbb{B}^d)$ absolutely continuous with respect to the Lebesgue measure λ, with corresponding probability density function (pdf) $f = dP/d\lambda$, which is assumed to belong to the space $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with $1 \leqslant p < \infty$.

Given a random sample $\{X_i, 1 = 1, ..., n\}$ from P, a *singular integral* (SI) estimator of f has the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}(X_i - x),$$

[1] We thank the editor, a referee, Juan A. Cuesta, and Gabor Lugosi for their helpful comments and suggestions. This research was supported by Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB98-0025.

where $m_n = m(n)$ is known as the *smoothing parameter* sequence, and $\{K_{m_n}\}_{n\in\mathbb{N}}$ as the singular integral window or kernel sequence.

The sequence $\{m_n\}_{n\in\mathbb{N}}$ is not necessarily a sequence of numbers; it may be a sequence of positive definite matrices ordered by decreasing norm, in the usual kernel estimator of a multivariate density, or the order of a polynomial, in the Fourier series estimator. The smoothing sequence belongs to some directed set $\mathbb{I}$, which is a non empty set endowed with a partial preorder $\leqslant$, such that if $m_1, m_2 \in \mathbb{I}$, then $\exists m_3 \in \mathbb{I}$ such that $m_1 \leqslant m_3$ and $m_2 \leqslant m_3$. It is assumed that $\{m_n\}_{n\in\mathbb{N}}$ diverges in $\mathbb{I}$ as $n \to \infty$, i.e., $\forall M \in \mathbb{I}$, $\exists n_M \in \mathbb{N}$ such that $m_n \geqslant M$ $\forall n \geqslant n_M$.

Some related estimators have been studied by Walter and Blum (1979), Prakasa Rao (1983, pp. 137–141) and Devroye and Györfi (1985, Chap. 12, Sect. 8), among others. The *SI* class encompasses a large number of nonparametric estimators as kernels, Fourier series estimators, Fejér sums estimators, etc. See, e.g., Butzer and Nessel (1971) and Devroye and Györfi (1985, Chap. 12, Sect. 8) for a review.

We propose the *averaged singular integral* (*ASI*) estimator of order $r \in \mathbb{N}$, defined as,

$$\hat{f}^r_n(x) = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{k=1}^{r} \frac{\phi^r_k}{k^d} \cdot K_{m_n}\left(\frac{X_i - x}{k}\right)\right] = \sum_{k=1}^{r} \phi^r_k \left[\frac{1}{n} \sum_{i=1}^{n} \frac{1}{k^d} \cdot K_{m_n}\left(\frac{X_i - x}{k}\right)\right],$$

where $\phi^r_k = (-1)^{(k-1)} \binom{r}{k}$, $k = 1, \ldots, r$, and $\{K_m\}$ is a singular integral window sequence. The *SI* estimator with window sequence $\{K_m\}$ is the *ASI* of order 1. Notice that $\sum_{k=0}^{r} \phi^r_k = 0$ and, therefore, *ASI* estimators integrate to one, though they can take negative values. The *ASI* estimator can be interpreted as an *SI* estimator with window $\sum_{k=1}^{r} \phi^r_k k^{-d} K_m(k^{-1}u)$ or as a weighted average of r *SI* estimators with weights ϕ^r_k and window $k^{-d}K_m(k^{-1}u)$.

In this paper we obtain bounds for the global bias of *ASI* estimators without smoothness assumptions on f. These bounds are useful for establishing finite sample properties, for obtaining global rates of convergence under smoothness assumptions, and for showing that, under certain conditions, the rate of convergence increases with r.

Density estimators with higher rate of convergence for the bias allow widening of the spectrum of the admissible degree of smoothing. This feature is decisive in semiparametric inference problems, where statistics are weighted averages of nonparametric estimates evaluated at data points, like in average derivatives (e.g., Powell *et al.*, 1989, and Robinson, 1989), partially linear models (e.g., Robinson, 1988), or when testing restrictions on nonparametric curves (e.g., Delgado and González-Manteiga, 2001.)

There are several bias reduction techniques. The *generalized jackknife* (Schucany and Sommers, 1977) is a weighted average of kernel estimators with different bandwidths. Weights and bandwidths depend on certain constants, which must be chosen by the practitioner. This method is related to *ASI* based on kernels (see Example 1), which is also a weighted average of kernel estimators with given weights ϕ_k^r and banwidths kH, $k=1, ..., r$. *Higher order kernels* estimation is possibly the most popular bias reduction technique; see, e.g., Singh (1979), Gasser and Müller (1984), and Gasser *et al.* (1985), among others. Another popular alternative is *local polynomial* estimation; see, e.g., Stone (1977), Cleveland (1979), and Fan and Gijbels (1996). Terrell and Scott (1980) propose a bias reduction technique based on the ratio of two kernel estimators of order two. The resulting estimator is always positive but it does not integrate to one.

The rest of the paper is organized as follows. In Section 2 we obtain universal global bounds. Section 3 establishes rates of convergence. Section 4 discusses some examples.

2. GLOBAL BIAS BOUNDS

The expected value of an *ASI* estimator of order r, $\hat{f}_n^r$, is given by

$$\alpha_{m_n}^r(f; x) = \sum_{k=1}^{r} \phi_k^r \cdot \int K_{m_n}(u)\, f(x+ku)\, du,$$

where $\{\alpha_{m_n}^r\}$ is a sequence of linear operators in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$. The *ASI* estimator of order r is globally asymptotically unbiased for all density functions $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ (i.e., $\hat{f}_n^r$ is "universally asymptotically unbiased" in L_p-norm) if

$$\lim_{n \in \mathbb{N}} \|E[\hat{f}_n^r(x)] - f(x)\|_{L_p(\lambda)} = \lim_{n \in \mathbb{N}} \|\alpha_{m_n}^r(f; x) - f(x)\|_{L_p(\lambda)} = 0,$$

for any sequence $\{m_n\}_{n \in \mathbb{N}}$ that diverges in $\mathbb{I}$. Note that we are considering a global convergence criteria in L_p-norm. This unbiasedness property is important, since we do not require smoothness assumptions to prove asymptotic unbiasedness. Smoothness conditions are only required to obtain rates of convergence for the bias.

The next theorem provides a bound for the bias of *ASI* estimators. Henceforth, we use the *smoothness modulus* of order $r \in \mathbb{N}$ in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, defined as

$$\omega_r(f, \delta)_{L_p(\lambda)} = \sup_{0 < \|h\| \leq \delta} \|\Delta_h^r(f; x)\|_{L_p(\lambda)},$$

where $\Delta_h^r(f; x) = \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} f(x+kh)$ is a higher order difference. Notice that $\omega_r(f; 0)_{L_p(\lambda)} = 0$, and for all $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, it is satisfied that $\lim_{\delta \to 0} \omega_r(f; \delta)_{L_p(\lambda)} = 0$. Define $\zeta_s(m) = \int |K_m(z)| \, \|z\|^s \, dz$ for integers $s \geqslant 0$.

THEOREM 1. *For all $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with $1 \leqslant p < \infty$,*

$$\|\alpha_m^r(f; x) - f(x)\|_{L_p(\lambda)} \leqslant \|f(x)(1 - \alpha_m^1(\mathbf{1}; x))\|_{L_p(\lambda)} + C \cdot 2^{r-1} \cdot \omega_r(f, \zeta_r(m)^{1/r})_{L_p(\lambda)},$$

with $C = 1 + \sup_{m \in \mathbb{I}} \zeta_0(m)$.

Proof. By the triangle inequality,

$$\|\alpha_m^r(f; x) - f(x)\|_{L_p(\lambda)} \leqslant \|f(x)(\alpha_m^1(\mathbf{1}; x) - 1)\|_{L_p(\lambda)} + B_m^r,$$

where

$$B_m^r = \|\alpha_m^r(f; x) - f(x)\, \alpha_m^1(\mathbf{1}; x)\|_{L_p(\lambda)}.$$

Noticing that $\alpha_m^r(\mathbf{1}; x) = \alpha_m^1(\mathbf{1}; x) = \int K_m(z - x)\, dz$ all $r > 1$, and applying the integral Minkowsky's inequality and Fubini's theorem,

$$\begin{aligned} B_m^r &= \left\| \int K_m(z) \cdot [(-1)^{1+r} \Delta_z^r(f; x)] \, dz \right\|_{L_p(\lambda)} \\ &\leqslant \int |K_m(z)| \cdot \left(\int |\Delta_z^r(f; x)|^p \, dx \right)^{1/p} dz \\ &\leqslant \int |K_m(z)| \cdot \omega_r(f; \|z\|)_{L_p(\lambda)} \, dz. \end{aligned} \tag{1}$$

Taking into account that for all $\tau > 0$, and $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, $\omega_r(f; \tau\delta)_{L_p(\lambda)} \leqslant (1 + \tau)^r \, \omega_r(f; \delta)_{L_p(\lambda)}$,

$$\omega_r\left(f; \frac{\|z\|}{\delta} \delta\right) \leqslant \left(1 + \frac{\|z\|}{\delta}\right)^r \cdot \omega_r(f; \delta)_{L_p(\lambda)} \leqslant 2^{(r-1)} \cdot \left(1 + \frac{\|z\|^r}{\delta^r}\right) \cdot \omega_r(f; \delta)_{L_p(\lambda)}, \tag{2}$$

applying C_τ-inequality. Hence, (1) and (2) imply that for all $\delta > 0$,

$$B_m^r \leqslant 2^{(r-1)} \cdot \omega_r(f; \delta)_{L_p(\lambda)} \cdot \left(\sup_{m \in \mathbb{I}} \zeta_0(m) + \frac{\zeta_r(m)}{\delta^r} \right). \tag{3}$$

First assume that $\zeta_r(m) > 0$ and taking $\delta = \zeta_r(m)^{1/r}$ in (3) we obtain that

$$B^r_m \leqslant 2^{(r-1)} \cdot (\sup_{m \in \mathbb{I}} \zeta_0(m) + 1) \cdot \omega_r(f; \zeta_r(m)^{1/r})_{L_p(\lambda)}.$$

Second, assume that $\zeta_r(m) = 0$. Then, by (3), for all $\delta > 0$,

$$B^r_m \leqslant 2^{(r-1)} \cdot (\sup_{m \in \mathbb{I}} \zeta_0(m) + 0) \cdot \omega_r(f; \delta)_{L_p(\lambda)},$$

and taking $\delta \downarrow 0$, $B^r_m = 0$, which proves the theorem. ∎

The next corollary provides sufficient conditions on $\{K_m\}$ for universally asymptotically unbiasedness.

COROLLARY 1. *Assume that* $\{K_m\}_{m \in \mathbb{I}} \subset L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ *satisfies*

(i) $\sup_{m \in \mathbb{I}} \zeta_0(m) < \infty$, (ii) $\alpha^1_m(\mathbf{1}; x) = 1$, *a.s.* $[\lambda]$, $\forall m \in \mathbb{I}$,

(iii) $\lim_{m \in \mathbb{I}} \zeta_r(m) = 0$.

Then $\hat{f}^r_n$ *is universally asymptotically unbiased in* L_p*-norm.*

The proof is immediate from Theorem 1. Conditions in Corollary 1 are satisfied for most $\{K_m\}$ sequences, as illustrated in the following example for the popular kernel estimators.

EXAMPLE 1. Let consider Kernels in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with

$$K_m(u) = \frac{1}{\det(H)} K(H^{-1}u),$$

where the smoothing parameter $m = H^{-1}$ is a definite positive matrix, ordered by decreasing $\|H\|$. The function $K(\cdot)$ satisfies, (a) $K \in L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, (b) $\int K(u)\, du = 1$, (c) $\int |K(u)|\, \|u\|^r\, du < \infty$. Then, (a) and (b) guarantee (i) and (ii) in Corollary 1, and (c) implies (iii), since,

$$\begin{aligned}\zeta_r(m) &= \frac{1}{\det(H)} \int |K(H^{-1}u)|\, \|u\|^r\, du \\ &\leqslant \|H\|^r \cdot \int |K(z)|\, \|z\|^r\, dz \xrightarrow[\|H\| \downarrow 0]{} 0.\end{aligned}$$

Theorem 1 provides a bias bound for a general class of density estimators without assuming differentiability on the underlaying density function. Furthermore, this result can be useful in order to establish rates of convergence for ASI estimators. For all $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, the rate of convergence of $\omega_r(f; \delta)_{L_p(\lambda)}$ to zero when $\delta \downarrow 0$ depends on f smoothness. In next section we show that ASI estimators can achieve the rate $O(\zeta_r(m))$ when f is smooth enough.

3. RATES OF CONVERGENCE

Let $W_p^s(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ be the Sobolev space of at least s-times weakly differentiable functions with L_p-integrable derivatives. In this section we will prove that, if $f \in W_p^r(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, then $\omega_r(f; \delta)_{L_p(\lambda)} = O(\delta^r)$. However, the ratio $O(\delta^r)$ cannot be improved, as stated in the following proposition.

PROPOSITION 1. *For all non constant* $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, $\exists C_f > 0$, *such that*, $\forall \delta \in (0, 1)$,

$$\omega_r(f; \delta)_{L_p(\lambda)} \geqslant C_f \cdot \delta^r.$$

Proof. Using that f is not constant and the smoothing modulus $\omega_r(f; 1)_{L_p(\lambda)} \neq 0$. Then, $\forall \delta \in (0, 1)$

$$\omega_r\left(f; \frac{\delta}{\delta}\right)_{L_p(\lambda)} \leqslant \left(1 + \frac{1}{\delta}\right)^r \omega(f; \delta)_{L_p(\lambda)} \leqslant 2^r \cdot \frac{1}{\delta^r} \cdot \omega_r(f; \delta)_{L_p(\lambda)},$$

and $\omega_r(f; \delta)_{L_p(\lambda)} \geqslant [2^{-r} \omega_r(f; 1)_{L_p(\lambda)}] \cdot \delta^r = C_f \cdot \delta^r$. ∎

As usual, given $\nu = (\nu_1, ..., \nu_d)$ with $\|\nu\|_1 = r$, define $D^\nu f(x) = \partial^r f(x)/\partial^{\nu_1} x_1 \cdots \partial^{\nu_d} x_d$, $x^\nu = x_1^{\nu_1} \cdots x_d^{\nu_d}$ and $\nu! = \nu_1! \cdots \nu_d!$. Next, we present the main result of this section.

THEOREM 2. *If* $f \in W_p^r(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, *then*,

$$\|\alpha_m^r(f; x) - f(x) \cdot \alpha_m^1(\mathbf{1}; x)\|_{L_p(\lambda)} \leqslant C \cdot 2^{r-1} \cdot \zeta_r(m) \cdot \left(\frac{1}{d^r} \left\| \sum_{\|\nu\|_1 = r} D^\nu f(x) \right\|_{L_p(\lambda)}\right).$$

Proof. First consider $d = 1$. It is known that

$$\|\Delta_h^r(f; x)\|_{L_p(\lambda)} \leqslant h^r \, \|D^r f(x)\|_{L_p(\lambda)},$$

see, e.g., Schumaker (1981, Eq. 2.109). Hence, applying the definition of smoothness modulus $\omega_r(f, \delta)_{L_p(\lambda)} \leqslant \delta^r \cdot \|D^\nu f(x)\|_{L_p(\lambda)}$, the result follows applying Theorem 1. Now, we extend this result to the $d > 1$ case. Define

$F(u)=f(x+u\cdot I_d)$, continuous in $u=0$ for almost everywhere $x\in\mathbb{R}$. Consider $c\in\mathbb{R}$, and

$$\Delta_c^r(F;u)=\sum_{k=0}^{r}(-1)^{k-1}\binom{r}{k}F(x+(u+k\cdot c)\cdot I_d).$$

Then, $\|\Delta_c^r(F;u)\|_{L_p(\lambda)}\leqslant c^r\ \|D^rF(u)\|_{L_p(\lambda)}$ holds at $u=0$. Notice that $D^rF(u)|_{u=0}=\sum_{\|v\|_1=r}D^vf(x)$, and

$$\Delta_c^r(F;u)|_{u=0}=\sum_{k=0}^{r}(-1)^{k-1}\binom{r}{k}F(x+k\cdot(c,\ldots,c)')=\Delta_{(c,\ldots,c)}^r(f;x).$$

Therefore, $\|\Delta_{(c,\ldots,c)}^r(f;x)\|_{L_p(\lambda)}\leqslant c^r\ \|\sum_{\|v\|_1=r}D^vf(x)\|_{L_p(\lambda)}$. Define $h=(c,\ldots,c)'$. Then

$$\|\Delta_h^r(f;x)\|_{L_p(\lambda)}\leqslant\left(\frac{\|h\|}{d}\right)^r\left\|\sum_{\|v\|_1=r}D^vf(x)\right\|_{L_p(\lambda)}.$$

Applying the definition of smoothness modulus of order r,

$$\omega_r(f,\delta)_{L_p(\lambda)}\leqslant\delta^r\cdot\frac{1}{d^r}\left\|\sum_{\|v\|_1=r}D^vf(x)\right\|_{L_p(\lambda)},\qquad\delta\geqslant 0.$$

Finally, apply Theorem 1. ▮

This result provides higher order rates of convergence for the bias whenever $\zeta_r(m)$ tends to zero faster than $\zeta_1(m)$. This requirement is trivially satisfied for kernel estimators, where the rate of convergence to zero of $(\|H\|^r)$ increases exponentially with r (see Example 1.)

It is straightforward to show that if $f\in W_p^s(\mathbb{R}^d,\mathbb{B}^d,\lambda)$, $s<r$,

$$\|\alpha_m^r(f;x)-f(x)\cdot\alpha_m^r(\mathbf{1};x)\|_{L_p(\lambda)}=O(\zeta_r(m)^{s/r}),$$

using the fact that $\omega_r(f;\delta)_{L_p(\lambda)}\leqslant\delta^s\omega_{r-s}(f;\delta)$ with $d=1$ (see, e.g., Schumaker, 1981, Theorem 2.59) and extending this result to the multivariate case reasoning as in Theorem 2.

Related results can also be obtained when f satisfies some Lipschitz conditions. Consider the high order Lipschitz space,

$$Lip(\gamma,r)_p=\{f\in L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda):\omega_r(f;\delta)_{L_p(\lambda)}\leqslant c_f\cdot\delta^\gamma\},$$

with $r\in\mathbb{N}$, $r-1<\gamma<r$. Applying Theorem 1, if $f\in Lip(\gamma,r)_p$, then

$$\|\alpha_m^r(f;x)-f(x)\cdot\alpha_m^r(\mathbf{1};x)\|_{L_p(\lambda)}=O(\zeta_r(m)^{\gamma/r}).$$

In order to illustrate the bias properties of *ASI* estimators, it is useful to consider an alternative approach, based on the usual Taylor expansion. Assume that $f \in W_p^r(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ and $\{K_m\}$ satisfies conditions in Corollary 1. Then,

$$f(x+ku)=f(x)+\sum_{j=1}^{r-1} \sum_{\|v\|_1=j} \frac{1}{v!}(ku)^v D^v f(x)$$

$$+r \sum_{\|v\|_1=r} \frac{1}{v!} \int_0^1 (1-t)^{r-1} D^v f(x+tku)(ku)^v \, dt$$

almost everywhere for all $k=1, \ldots, r$. Define $c_{vm}=\int u^v K_m(u)\, du$, possibly different than zero. Therefore,

$$E[\hat{f}_n^r(x)]-f(x)=\sum_{k=1}^{r} \phi_k^r \int_{-\infty}^{\infty} K_m(u)[f(x+ku)-f(x)]\, du$$

$$=\sum_{j=1}^{r-1} \sum_{\|v\|_1=j} \frac{c_{vm} D^v f(x)}{v!} \left(\sum_{k=1}^{r} \phi_k^r k^j \right) + \sum_{k=1}^{r} \phi_k^r k^r R_{km}^r(x),$$

where

$$R_{km}^r(x)=r \sum_{\|v\|_1=r} \frac{1}{v!} \int_{-\infty}^{\infty} \int_0^1 (1-t)^{r-1} D^v f(x+tku)\, u^v K_m(u)\, dt\, du.$$

Noticing that $\sum_{k=1}^r \phi_k^r k^j=0$, all $j=1, \ldots, r-1$, and all $r \geqslant 1$,

$$\|E[\hat{f}_n^r(x)]-f(x)\|_{L_p(\lambda)}=\left\| \sum_{k=1}^{r} \phi_k^r k^r R_{km}^r(x) \right\|_{L_p(\lambda)}=O(\zeta_r(m)). \tag{4}$$

It is difficult to compare the exact bias of *ASI* estimators with other higher order methods. Assume that K_m satisfies $c_{vm}=0$ for all m and all v such that $\|v\|_1=1, \ldots, r-1$. For instance, it happens with *higher order kernels* of order r. Then, the bias of the *ASI* estimator of order 1 is

$$\|E[\hat{f}_n^1(x)]-f(x)\|_{L_p(\lambda)}=\|R_{1m}^r(x)\|_{L_p(\lambda)}=O(\zeta_r(m)). \tag{5}$$

A comparison between (4) and (5), under general conditions, does not seem immediate.

The next section discusses some examples.

4. EXAMPLES AND DISCUSSION

In Example 1 we showed that $\zeta_r(m) = O(\|H\|^r)$ in the kernel case. This rate of convergence is also achieved by *higher order kernels* of order r. Furthermore, if K is a kernel of order r, the *ASI* of order r does not improve the rate of convergence of the *ASI* of order 1 (see (4) and (5)). In this situation, a comparison between the exact biases of both estimators is not immediate, but it is possible that an *ASI* of order r has lower bias than an *ASI* of smaller order, for some bandwidths, certain underlaying densities, and kernel choices. Table I illustrates this point. We report biases in the univariate case ($d = 1$) for different values of r and bandwidth H, when observations are standard normal and K is a Gaussian kernel of order 2.

Though the rates of convergence for $r = 1$ and $r = 2$ are identical, bias for $r = 2$ is smaller for larger values of H. Such bias improvements were not found using a kernel of order 4.

ASI kernel estimators can be computed easily from any kernel function K. The *ASI* kernel method can be of practical relevance when *higher order kernels* are unsuitable, or are difficult to compute. For instance, symmetric *higher order kernels* of order r (r even) are constructed solving $r/2$ moment equations and the resulting kernel has $r/2$ terms. However, with asymmetric kernels, which are considered for purposes of boundary modification or change-point estimation (e.g., Gasser *et al.*, 1985, and Müller 1991) r moment equations have to be solved, and the resulting *higher order kernel* has r terms, like the *ASI* estimator, which do not require to solve equations. Given any *SI* estimator designed for specific purposes, it can be transformed easily in an *ASI* estimator. Other *ASI* estimators that are not related to kernels can be considered. For instance, those based on *Fejér*, *Jackson*, *Rogosinski*, or *de la Vallée Poussin* windows. See Butzer and

TABLE I

$\|E[\hat{f}_n^r(x)] - f(x)\|^2_{L_2(\lambda)}$ for *ASI* Estimator with a Gaussian Kernel of Order 2 When $X_i \sim N(0, 1)$

	$r=1$	$r=2$	$r=3$	$r=4$
$H=1$	2.090×10^{-2}	7.409×10^{-3}	1.540×10^{-2}	2.603×10^{-2}
$H=3/4$	9.336×10^{-3}	4.018×10^{-3}	1.068×10^{-2}	1.585×10^{-2}
$H=1/2$	2.485×10^{-3}	2.485×10^{-3}	4.231×10^{-3}	3.943×10^{-3}
$H=1/4$	1.914×10^{-4}	4.973×10^{-4}	1.722×10^{-4}	2.422×10^{-5}
$H=1/8$	1.266×10^{-5}	4.517×10^{-5}	1.596×10^{-6}	2.150×10^{-7}
$H=1/10$	5.223×10^{-6}	1.940×10^{-5}	3.012×10^{-7}	6.236×10^{-8}

TABLE II

$2^{r-1} \cdot \zeta_r(m)$ Values for Different Values of m with *Jackson's* Window

	$r=1$	$r=2$	$r=3$	$r=4$
$m=50$	2.649×10^{-2}	2.420×10^{-3}	6.294×10^{-4}	1.198×10^{-3}
$m=100$	1.324×10^{-2}	6.026×10^{-4}	8.663×10^{-5}	1.498×10^{-4}
$m=500$	2.647×10^{-3}	2.407×10^{-5}	8.406×10^{-7}	1.199×10^{-6}

Nessel (1971). The next examples illustrate the convergence rate improvements of *ASI* based on *Jackson* and *Fejér* windows.

EXAMPLE 2. Let us consider the *Jackson* window in $L_p[-\pi, \pi]$,

$$K_m(u) = \frac{3}{2\pi m(2m^2+1)} \left(\frac{\sin(mu/2)}{\sin((1/2)\,u)} \right)^4.$$

with $m \in \mathbb{N}$, which satisfies conditions in Corollary 1.

Table II provides $2^{r-1} \cdot \zeta_r(m)$ values for different values of m. Computations have been carried out by numerical integration.

The rate improves as r increases when $r \leqslant 3$. Notice that $\zeta_3(m) > \zeta_4(m)$ but $2^2\zeta_3(m) < 2^3\zeta_4(m)$.

EXAMPLE 3. Let us consider the *Fejér* window in $L_p[-\pi, \pi]$,

$$K_m(u) = \frac{1}{2\pi(m+1)} \left(\frac{\sin((m+1/2)\,u)}{\sin((1/2)\,u)} \right)^2.$$

with $m \in \mathbb{N}$, which satisfies conditions in Corollary 1.

Table III provides $2^{r-1} \cdot \zeta_r(m)$ values for different values of m.

The rate does not improve with r in this case. Notice that $\zeta_1(m) > \zeta_2(m)$ but $\zeta_1(m) < 2\zeta_2(m)$.

TABLE III

$2^{r-1} \cdot \zeta_r(m)$ Values for Different Values of m with *Fejér's* Window

	$r=1$	$r=2$	$r=3$	$r=4$
$m=50$	0.0859	0.1087	0.3945	1.7974
$m=100$	0.0477	0.0549	0.1992	0.9075
$m=500$	0.0068	0.0110	0.0402	0.1834

REFERENCES

1. P. L. Butzer and R. J. Nessel, "Fourier Analysis and Approximation," Vol. 1, Birkhäuser, Basel/Stuttgart, 1971.
2. W. S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc.* **74** (1979), 829–836.
3. M. A. Delgado and W. Gonzalez-Manteiga, Significance testing in nonparametric regression based on the Bootstrap, *Ann. Statist.*, in press.
4. L. Devroye and L. Györfi, "Nonparametric Density Estimation: The L_1 View," Wiley, New York, 1985.
5. J. Fan and I. Gijbels, "Local Polynomial Modelling and Its Applications," Monographs on Statistics and Applied Probability, Vol. 66, Chapman & Hall, London, 1996.
6. T. Gasser and H. G. Müller, Optimal convergence properties of kernel estimates of derivatives of a density, *in* "Smoothing Techniques for Curve Estimation" (T. Gasser and M. Rosenblatt, Eds.), Lecture Notes in Mathematics, Vol. 757, pp. 144–154, Springer-Verlag, Berlin, 1984.
7. T. Gasser, H. G. Müller, and V. Mammitzsch, Kernels for nonparametric curve estimation, *J. Roy. Statist. Soc. Ser. B* **47** (1985), 238–252.
8. H. G. Müller, Smooth optimal kernel estimators near endpoints, *Biometrika* **78** (1991), 521–530.
9. J. L. Powell, J. H. Stock, and T. M. Stoker, Semiparametric estimation of index coefficients, *Econometrica* **57** (1989), 1403–1430.
10. B. L. S. Prakasa Rao, "Nonparametric Functional Estimation," Academic Press, London, 1983.
11. P. M. Robinson, Root-n consistent semiparametric regression, *Econometrica* **56** (1988), 931–954.
12. P. M. Robinson, Hypothesis testing in semiparametric and nonparametric models for econometric time series, *Rev. Econom. Stud.* **56** (1989), 511–534.
13. R. S. Singh, Mean squared errors of estimates of a density and its derivatives, *Biometrika* **66** (1979), 177–180.
14. W. R. Schucany and J. P. Sommers, Improvement of kernel type density estimators, *J. Amer. Statist. Assoc.* **72** (1977), 420–423.
15. L. L. Schumaker, "Spline Functions Basic Theory," Wiley, New York, 1981.
16. C. J. Stone, Consistent nonparametric regression, *Ann. Statist.* **5** (1977), 595–645.
17. G. R. Terrell and D. W. Scott, On improving convergence for nonnegative kernel density estimators, *Ann. Statist.* **8** (1980), 1160–1163.
18. G. Walter and J. R. Blum, Probability density estimation using delta sequences, *Ann. Statistics* **7** (1979), 328–340.