WEIGHTED NONPARAMETRIC REGRESSION

Delicado, P.
Departament d'Economia i Empresa
Universitat Pompeu Fabra
Ramon Trias Fargas, 25
08005 Barcelona (Spain)

del Río, M.
Departamento de Estadística e I.O.
Universidad Complutense
Ciudad Universitaria
28040 Madrid (Spain)

Key Words and Phrases: asymptotic behavior; fixed design; kernel regression; local polynomial estimators; smoothing.

ABSTRACT

In the fixed design regression model, additional weights are considered for the Nadaraya-Watson and Gasser-Müller kernel estimators. We study their asymptotic behavior and the relationships between new and classical estimators. For a simple family of weights, and considering the *AIMSE* as global loss criterion, we show some possible theoretical advantages. An empirical study illustrates the performance of the weighted kernel estimators in theoretical ideal situations and in simulated data sets. Also some results concerning the use of weights for local polynomial estimators are given.

1. INTRODUCTION

Some of the most popular nonparametric regression smoothers are: the Nadaraya-Watson (NW) estimator, independently proposed by Nadaraya (1964) and Watson (1964) as an estimator of the regression curve; and the Gasser-Müller (GM) estimator, based on the convolution of a kernel function with a modification of the regressogram. An extensive comparison and discussion of their merits can be found in Chu and Marron (1991).

The bulk of this paper is a proposal of variants of these estimators considering additional weights in their associated local weighted averages. In Sections 2 and 3 the new estimators are defined and their asymptotic local behaviors are studied. Section 4 examines the performance of their asymptotic integrated mean squared error for a natural family of weights and the theoretical advantages of the new estimators. Section 5 contains numerical results showing in practice the behavior of the estimators under different design and curve combinations. Finally, Section 6 deals with the use of weights for the local polynomial method.

We consider a fixed design model given by $Y_i = g(t_i) + e_i$, $i = 1, \ldots, n$. The values $t_i \in [0,1]$ are generated by a regular density f with the usual properties (see, e.g. Müller, 1988, p. 26), and there exists a $\sigma^2(\cdot)$ function such that $V(Y_i) = \sigma^2(t_i)$. Extensions of the definitions and results to random designs are possible.

We assume that the kernels are symmetric densities of order 2 supported on $[-1,1]$ and $\mathrm{Lip}^1[-1,1]$. The weights will be defined by a function $w : [0,1] \longrightarrow \mathbb{R}^+$.

We will study the asymptotic properties of the new estimators for $t \in (0,1)$. The boundary problems can be handled with the methods of Rice (1984) or Hall and Wehrly (1991) for the NW estimators and with the classical modifications for the GM case, —see e.g., Müller (1988), Sec. 5.8—.

2. WEIGHTED NW ESTIMATORS

We define the weighted NW estimator by

$$\hat{g}_1(t,w) = \hat{g}_1(t) = \frac{\sum_{i=1}^n w(t_i) K_h(t-t_i) Y_i}{\sum_{i=1}^n w(t_i) K_h(t-t_i)},$$

where $K_h(z) = K(z/h)/h$.

Its asymptotic behavior is given in the following result.

Proposition 1 *Assume that $w, f, g \in C^2[0,1]$ and $\sigma^2 \in C^1[0,1]$. If $h \longrightarrow 0$ and $nh \longrightarrow \infty$ when $n \longrightarrow \infty$, then for $t \in (0,1)$:*

$$\begin{aligned} E(\hat{g}_1(t)) &= g(t) + \frac{h^2}{2}\mu_2(K)\left(g''(t) + \frac{2g'(t)\,(w(t)f(t))'}{w(t)f(t)}\right) + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\frac{1}{nh}\right), \\ V(\hat{g}_1(t)) &= \frac{1}{nh}V(K)\frac{\sigma^2(t)}{f(t)} + \mathrm{o}\left(\frac{1}{nh}\right), \end{aligned}$$

where $\mu_2(K) = \int u^2 K(u)du$, $V(K) = \int K^2(u)du$.

REMARKS:

1) Regarding to asymptotic bias, the use of weights given by w is equivalent to consider a design function proportional to wf. In particular, similar behavior is obtained either using the NW estimator with data from a design f or the weighted estimator with weight function $w = f$ and equally spaced t_i's. This could provide advantages in certain practical situations.

2) The ideal choice $g''(t)/g'(t) = -2\,(w(t)f(t))' / (w(t)f(t))$, that is, $wf \propto (g')^{-1/2}$, would eliminate the bias term $O(h^2)$ and it would give bias of order h^4, reducing the mean squared error.

3) The inclusion of the weight function has not effect on the asymptotic variance.

3. WEIGHTED GM ESTIMATORS

The second weighted estimator is defined by

$$\hat{g}_2(t,w) = \hat{g}_2(t) = \frac{\sum_{i=1}^n Y_i \int_{s_{i-1}}^{s_i} w(u) K_h(t-u)du}{D(t)},$$

where s_i is an interpolating sequence of the t_i's, i.e. $s_0 = 0$, $t_i \leq s_i \leq t_{i+1}$, $s_n = 1$; the function $D(t) = (w * K_h)(t) = \int_0^1 w(u) K_h(t-u) du$ is introduced as a normalization factor.

Its asymptotic properties are given in the following result.

Proposition 2 *Assume that conditions of Proposition 1 hold. Then, for* $t \in (0,1)$

$$E(\hat{g}_2(t)) = g(t) + \frac{h^2}{2}\mu_2(K)\left(g''(t) + \frac{2g'(t)w'(t)}{w(t)}\right) + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\frac{1}{n}\right).$$

If furthermore $s_i = (t_i + t_{i+1})/2$*:*

$$V(\hat{g}_2(t)) = \frac{1}{nh} V(K) \frac{\sigma^2(t)}{f(t)} + \mathrm{o}\left(\frac{1}{nh}\right).$$

REMARKS:

1) The estimator $\hat{g}_2(t)$ is defined by smoothing the function $r(t,w) = w(t)Y_i$, $s_{i-1} < t \leq s_i$, by means of its convolution with K_h. Another reasonable option is to consider a direct weighting on each sum term or, equivalently, to smooth the 'weighted regressogram' defined as $\tilde{r}(t,w) = w(t_i)Y_i$, $s_{i-1} < t \leq s_i$. This scheme would lead to consider

$$\tilde{g}_2(t) = \frac{\left(\sum_{i=1}^n Y_i w(t_i) \int_{s_{i-1}}^{s_i} K_h(t-u) du\right)}{\tilde{D}(t)},$$

where $\tilde{D}(t) = \sum_{i=1}^n w(t_i) \int_{s_{i-1}}^{s_i} K_h(t-u) du$. The selected approach makes no difference since it is easy to prove that the resulting estimators are asymptotically equivalent in the sense that the leading terms of their means and variances are equal.

2) The consideration of weights for GM estimators using w gives an asymptotic bias equal to that obtained with a NW estimator under a design proportional to w. This fact, joined with the Remark 1 after Proposition 1, will permit to relate the weighted estimators with the *classical* ones.

3) The choice $g''(t)/g'(t) = -2w'(t)/w(t)$ eliminates the bias term $O(h^2)$ and will lead to bias of order h^4.

4) Again, the asymptotic variance of the classical estimator it is not affected by the weight function.

There exists the following relationship between the new estimators.

Proposition 3 *(i) Given the estimator* $\hat{g}_1(t,w)$*, there is a function* $\bar{w}$ *such that* $\hat{g}_2(t,\bar{w})$ *is asymptotically equivalent to* $\hat{g}_1(t,w)$*.*

(ii) Given $\hat{g}_2(t,w)$*, there is a function* $\tilde{w}$ *such that* $\hat{g}_1(t,\tilde{w})$ *is asymptotically equivalent to* $\hat{g}_2(t,w)$*.*

The assertion follows taking $\bar{w} = wf$ and $\tilde{w} = w/f$ in (i) and (ii), respectively.

REMARK: Consequently, given a NW estimator (respectively, a GM estimator), it is possible to find an equivalent weighted estimator of type GM (respectively, NW) by taking weights $w = f$ (respectively, $w = 1/f$).

4. GLOBAL BEHAVIOR

To study the global performance of the new estimators and to compare them with the classical ones, we restrict ourselves to weight functions of type $w(\alpha, t) = f^{\alpha}(t)$, $\alpha \in I\!R$. In another context –studying a unified approach–, Jennen-Steinmetz and Gasser (1988) considered a class of kernel estimators with variable bandwidth given by a subfamily of $w(\alpha, t)$.

For several reasons $w(\alpha, t)$ seems an appropriate way of actuation on the predictor variable. Note that the relationships indicated in the last section were made through this family. Also, according to the remark to Proposition 3, weights equal to 1 and $1/f$ or, respectively, 1 and f, provide the NW and GM estimators. So, it is natural consider 'intermediate' weights; if we use the weighted geometric means: $(1/f)^{\alpha}1^{1-\alpha}$, or, $f^{\alpha}1^{1-\alpha}$, $\alpha \in [0, 1]$, respectively, we arrive to that family.

Let us point out that for the equally spaced design this weight family reduces to the trivial function $w(\alpha, t) = 1$. We recall that if the design is given by the uniform density, the estimators NW and GW are asymptotically equivalent; therefore, the 'intermediate' estimators defined by using $w(\alpha, t)$ will be also equivalent to them. To avoid this special case, we will henceforth consider non equally spaced designs.

The evaluation of the global performance of the estimators will be made according to their integrated mean squared error, *IMSE*. For the family $w(\alpha, t)$, Proposition 3 permits considering any of the weighted estimators. Working with, for instance, the estimator $\hat{g}_2(t)$, we have as asymptotic mean squared error:

$$AMSE(\alpha, t, h) = \frac{h^4}{4}\mu_2^2(K)C(t, \alpha) + \frac{1}{nh}V(K)\frac{\sigma^2(t)}{f(t)},$$

where

$$C(t, \alpha) = \left(g''(t) + 2\alpha\frac{g'(t)f'(t)}{f(t)}\right)^2.$$

And the asymptotic *IMSE* will be

$$AIMSE(\alpha, h) = \frac{h^4}{4}\mu_2^2(K)\int C(t, \alpha)dt + \frac{1}{nh}V(K)\int\frac{\sigma^2(t)}{f(t)}dt.$$

Differentiating with respect to h and equating to zero, we have as asymptotically optimal global bandwidth

$$h(\alpha) = n^{-1/5}\left(\frac{V(K)\int\frac{\sigma^2(t)}{f(t)}dt}{\mu_2^2(K)\int C(t, \alpha)dt}\right)^{1/5}.$$

Therefore, the (asymptotically) optimal *AIMSE* is given by

$$AIMSE(\alpha, h(\alpha)) = \frac{5}{4}n^{-4/5}\left(\mu_2^2(K)\int C(t, \alpha)dt\right)^{1/5}\left(V(K)\int\frac{\sigma^2(t)}{f(t)}dt\right)^{4/5}.$$

Only the term $\psi(\alpha) = \int C(t, \alpha)dt$ depends on α; recalling the expression for $C(t, \alpha)$, the minimum of ψ is attained at

$$\alpha_0 = -\frac{1}{2}\frac{\int \frac{g''g'f'}{f}}{\int \left(\frac{g'f'}{f}\right)^2},$$

giving,

$$\psi(\alpha_0) = \int (g'')^2 - \frac{\left(\int \frac{g''g'f'}{f}\right)^2}{\int \left(\frac{g'f'}{f}\right)^2}.$$

Observe that Cauchy–Schwarz inequality provides

$$\left(\int \frac{g''g'f'}{f}\right)^2 \leq \left(\int (g'')^2\right)\left(\int \left(\frac{g'f'}{f}\right)^2\right),$$

that is, $AIMSE(\alpha_0, h(\alpha_0)) \geq 0$, with equality if and only if,

$$g'' \propto \frac{g'f'}{f} \iff \frac{g''}{g'} \propto \frac{f'}{f} \iff f = k_1(g')^{k_2}.$$

This shows that a suitable election for the α parameter of our weighting family may theoretically permit to reduce the *AIMSE*. Likewise, substantial reductions of this error can be done in 'special' situations with appropriate combinations between the functions f and g and their derivatives. Nevertheless, to exploit in practice the use of weighted estimators we will need an estimation of α_0. This will be considered in the next section.

5. EMPIRICAL STUDIES

A numerical study was conducted to analyze the behavior of our proposal. Its main objective was to evaluate the advantages of using a weighted estimator instead of usual kernel estimators. Moreover, relationships between the gains attained by weighted regression and some characteristics of the regression function (as symmetry and curvature) were also of interest. We have considered two situations: a theoretical one, where the optimal parameter α_0 is assumed to be known, and a practical one, where the optimal value of α is estimated from the data.

The first part of our empirical study is a numerical exercise. We have considered the different combinations of designs (truncated to $[0,1]$) and regression curves displayed in Table I. In an ideal theoretical frame where function g is assumed to be known, we can use the asymptotic relative efficiencies

$$RE_A(\alpha_0, h_{\text{NW}}) = \frac{AIMSE(\alpha = 1, h = h_{NW})}{AIMSE(\alpha_0, h(\alpha_0))}$$

and $RE_A(\alpha_0, h_{\text{GM}})$ (defined as above, but for $\alpha = 0$ and $h = h_{GM}$) to compare our proposal with NW and GM estimators. The computed α_0, optimal values of the parameter, and the asymptotic relative efficiencies are displayed in the upper boxes of the cells of Tables II (normal design) and III (mixed design): $RE_A(\alpha_0, h_{\text{NW}})$ above and $RE_A(\alpha_0, h_{\text{GM}})$ below.

We point out some conclusions from this study. When the curvature of function $g(t)$ becomes larger (i.e., when γ increases), α_0 increases. Functions with moderate

TABLE I: *Scheme for empirical studies.*

Regular designs	1: $N(.5, \sigma = .25)$ truncated to $[0,1]$ 2: $\frac{1}{2}U([0,1]) + \frac{1}{2}N(.5, \sigma = .25)$ truncated to $[0,1]$
$g(t)$ functions	$g(t;\beta,\gamma) = \beta t^{\gamma} + (1-\beta)(1-t)^{\gamma}$ $\beta = .5, .65, .8, .95, \quad \gamma = 2, 3, 4, 6$
Sample sizes	$n = 50, 200$
Error distribution $\epsilon_i = Y_i - g(t_i) \sim N(0, \sigma_\epsilon)$	Relative dispersion $= \frac{4\sigma_\epsilon}{Range(g(t))+4\sigma_\epsilon} = .2, .4$

curvature ($\gamma = 2$) produce α_0 values increasing when the symmetry of $g(t)$ is larger. This is explained by the expression of α_0 and by the existing symmetries in the derivative of a symmetric function.

With regard to the theoretical relative efficiency, usually $RE_A(\alpha_0, h_{\mathrm{NW}})$ is greater than $RE_A(\alpha_0, h_{\mathrm{GM}})$. In general, $RE_A(\alpha_0, h_{\mathrm{NW}}) > RE_A(\alpha_0, h_{\mathrm{GM}})$ when $\alpha_0 < .5$ or, equivalently, when the optimal parameter is closer to the GM estimator ($\alpha = 0$) than to the NW estimator ($\alpha = 1$). The closeness to 0 or 1 of α_0 (or, in real cases, its estimation) answers, in some sense, the question "which of these kernel estimators should be used, and when?", posed in Jones et al. (1994). Focusing on the utility of the inclusion of weights we note that when the curvature of g increases, the advantages obtained are more noteworthy.

The second part of the study is more interesting from a practical point of view. It compares the weighted and unweighted estimators when our basic parameter α_0, and also bandwidths h for NW and GM estimators, are estimated from data. We use plug-in estimations adapted to the theoretical expressions given in Section 4. The required pilot estimations of the derivatives of g have been obtained using Fortran routines CURVDAT (see Köhler, 1990, and Gasser et al., 1991, where all aspects concerning the CURVDAT programs are described).

The combinations of the functions $g(t)$ with other elements, defined and given in Table I, specify each simulated situation. The NW estimator is replaced by its equivalent weighted GM version. The estimator with optimal weights is calculated using its weighted GM version. Since we have used GM type estimators, we have applied the usual boundary corrections. The simulation analysis runs through 300 times.

The evaluation of the quality of the estimations is also given by relative efficiencies. Firstly, we consider the *relative efficiency* (with respect to NW estimator) based on *AIMSE for estimated parameters*:

$$RE_A(\hat{\alpha}, \hat{h}_{\mathrm{NW}}) = \frac{AIMSE(\alpha = 1, h = \hat{h}_{NW})}{AIMSE(\hat{\alpha}, h(\hat{\alpha}))}.$$

Since we have now finite samples, we can use the *IMSE* distance to measure relative efficiencies. Thus,

$$RE_I(\alpha_0, h_{\mathrm{NW}}) = \frac{IMSE(\alpha = 1, h = h_{NW})}{IMSE(\alpha_0, h(\alpha_0))}$$

permits to evaluate the finite sample efficiency of the weighted estimator in the ideal situation of known parameter α_0. Also, the quotient

$$RE_I(\hat{\alpha}, \hat{h}_{\mathrm{NW}}) = \frac{IMSE(\alpha = 1, h = \hat{h}_{NW})}{IMSE(\hat{\alpha}, h(\hat{\alpha}))}$$

takes into account the estimation of α and the finite sample situation. Similar definitions are done for GM estimator.

Main results are shown in the lower boxes of Tables II and III. The block in the left upper position shows the average value of the estimations $\hat{\alpha}$ obtained in simulation. Each of the three remaining blocks contains two numbers (the upper one for NW information and the lower one for GM). Right upper position is reserved for $RE_A(\hat{\alpha}, \hat{h})$, left lower for $RE_I(\alpha_0, h_0)$ and right lower for $RE_I(\hat{\alpha}, \hat{h})$. ($h_0$ and $\hat{h}$ represent optimal and estimated bandwidth for usual kernel estimators.)

We conclude by pointing out some results of this study. With respect to the estimation of parameter α, the averages of $\hat{\alpha}$ values are reasonably near to the optimal values α_0 and the variation is acceptable. One detected regularity is the decrease of standard error of estimated α when either β, γ or n increase and when the dispersion of residuals decreases.

Some agreement is also observed between theoretical *AIMSE*'s and the corresponding values obtained from estimations of α_0, h_{NW} and h_{GM} in simulation. The values of $RE_A(\hat{\alpha}, \hat{h})$ are always under $RE_A(\alpha_0, h_0)$ values. Thus, as it should be expected, the estimation of α implies a reduction in the asymptotic efficiency of the weighted estimator.

The values of $RE_I(\alpha_0, h_0)$ are always around the mean value of 1 and $RE_A(\alpha_0, h_0)$. Thus, when *IMSE* is used instead of *AIMSE*, the gain for using weighted estimators is not so substantial. This reduction of efficiency is more important than that implied by the estimation of α (and analyzed by $RE_A(\hat{\alpha}, \hat{h})$). If we look at values $RE_I(\hat{\alpha}, \hat{h})$ we observe that the effects (finite samples and estimation of α) are added, giving a more important reduction in the efficiency than they do separately. Nevertheless, most of these values are above one, so the weighted estimator improves NW and GM even if we estimate α and use *IMSE*'s to evaluate accuracy.

In general, the relative efficiences obtained from simulation show patterns agreeing with the theoretical ones discussed at the beginning of this section (with lower values in simulation). Therefore, our comments about the relationship between symmetry and curvature and the previous recomendations of using weights, are also valid for small sample sizes and appreciable dispersions. The previously noted agreement is clearer, as we could expect, under large sample size or small model dispersion.

TABLE II: *Normal design.*

	γ							
β	2		3		4		6	
	.208	1.80	.208	1.80	.263	1.86	.392	1.92
		1.18		1.18		1.30		1.63
.50	.332	1.32	.260	1.46	.279	1.46	.357	1.55
		1.19		1.18		1.20		1.33
	1.38	1.13	1.40	1.18	1.45	1.20	1.35	1.16
	1.08	1.04	1.08	1.03	1.15	1.05	1.25	1.11
	.181	1.82	.203	1.84	.262	1.89	.392	1.93
		1.14		1.18		1.31		1.63
.65	.265	1.42	.207	1.53	.280	1.59	.378	1.61
		1.17		1.17		1.23		1.38
	1.38	1.16	1.38	1.20	1.47	1.29	1.46	1.26
	1.07	1.03	1.09	1.03	1.13	1.07	1.30	1.17
	.130	1.90	.192	1.92	.259	1.95	.391	1.95
		1.09		1.20		1.34		1.65
.80	.156	1.56	.200	1.66	.256	1.72	.384	1.69
		1.13		1.17		1.26		1.44
	1.45	1.23	1.44	1.29	1.47	1.32	1.44	1.30
	1.05	1.01	1.09	1.03	1.16	1.09	1.27	1.18
	.089	2.02	.182	2.04	.256	2.04	.391	1.97
		1.06		1.22		1.38		1.66
.95	.092	1.76	.191	1.81	.264	1.83	.372	1.76
		1.09		1.19		1.31		1.47
	1.56	1.38	1.52	1.37	1.49	1.36	1.42	1.31
	1.03	1.00	1.11	1.05	1.16	1.11	1.33	1.25

Relative dispersion .2; $n = 200$. Table shows α_0, average value of $\hat{\alpha}$'s in simulations and relative efficiency of weighted estimator with respect to NW and GM evaluated by $RE_A(\alpha_0, h_0)$, $RE_I(\alpha_0, h_0)$, $RE_A(\hat{\alpha}, \hat{h})$ and $RE_I(\hat{\alpha}, \hat{h})$. Each cell displays:

α_0	$RE_A(\alpha_0, h_0)$
$\hat{\alpha}$	$RE_A(\hat{\alpha}, \hat{h})$
$RE_I(\alpha_0, h_0)$	$RE_I(\hat{\alpha}, \hat{h})$

TABLE III: *Mixed design.*

β	γ = 2		3		4		6	
	.761	1.04	.761	1.04	.955	1.00	1.488	1.36
		1.27		1.27		1.44		2.04
.50	.915	1.18	.729	1.17	.991	1.15	1.231	1.20
		1.13		1.13		1.21		1.34
	1.00	1.03	1.00	1.03	1.00	1.00	1.04	1.05
	1.11	1.03	1.14	1.02	1.22	1.05	1.34	1.15
	.629	1.08	.732	1.05	.946	1.00	1.487	1.37
		1.19		1.27		1.46		2.06
.65	.608	1.19	.698	1.16	.871	1.15	1.238	1.22
		1.10		1.13		1.23		1.42
	1.03	1.04	1.02	1.03	1.00	1.00	1.11	1.05
	1.11	1.02	1.18	1.05	1.21	1.08	1.47	1.21
	.414	1.17	.671	1.09	.928	1.01	1.485	1.40
		1.10		1.27		1.50		2.12
.80	.389	1.19	.596	1.17	.848	1.13	1.199	1.18
		1.06		1.14		1.24		1.48
	1.11	1.08	1.04	1.03	1.00	1.01	1.15	1.05
	1.04	.99	1.18	1.05	1.31	1.13	1.60	1.31
	.263	1.28	.615	1.15	.909	1.02	1.483	1.44
		1.06		1.29		1.57		2.19
.95	.281	1.27	.554	1.16	.794	1.12	1.258	1.20
		1.05		1.17		1.31		1.59
	1.18	1.14	1.06	1.03	1.01	1.00	1.18	1.06
	1.04	.99	1.17	1.08	1.31	1.16	1.62	1.35

Relative dispersion .2; $n = 200$. Table shows α_0, average value of $\hat{\alpha}$'s in simulations and relative efficiency of weighted estimator with respect to NW and GM evaluated by $RE_A(\alpha_0, h_0)$, $RE_I(\alpha_0, h_0)$, $RE_A(\hat{\alpha}, \hat{h})$ and $RE_I(\hat{\alpha}, \hat{h})$. Each cell displays:

α_0	$RE_A(\alpha_0, h_0)$
$\hat{\alpha}$	$RE_A(\hat{\alpha}, \hat{h})$
$RE_I(\alpha_0, h_0)$	$RE_I(\hat{\alpha}, \hat{h})$

6. WEIGHTED LOCAL POLYNOMIAL ESTIMATORS

Local polynomial (LP) regression has a long tradition and has received a renewed interest due to a set of recent findings. A complete collection of its atractive features and properties is given in Fan and Gijbels (1996). A nice introductory exposition can be found in Wand and Jones (1995), Chapter 5.

The idea of weighting can be translated to this method. Let $\hat{\beta}_w = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^{\mathrm{T}}$ the vector minimizing

$$\sum_{i=1}^{n} \Big(Y_i - \sum_{j=1}^{p} (\beta_j (t_i - t)^j) \Big)^2 K_h(t_i - t) w(t_i).$$

We define the weighted LP estimator by $\hat{g}_3(t, w; p) = \hat{g}_3(t; p) = \hat{\beta}_0$.

To study this estimator, the following notation (Ruppert and Wand, 1994) is useful. If $\mu_j = \int u^j K(u) du$, let N_p be the $(p+1) \times (p+1)$ matrix with (i,j)th entry equal to μ_{i+j-2}, and let $M_p(u)$ be the same as N_p but with the first column replaced by $(1, u, \ldots, u^p)^T$; finally, let $K_{(p)}$ be the kernel given by $K_{(p)}(u) = (|M_p(u)|/|N_p|)\, K(u)$. The asymptotic bias and variance are given by the next result.

Proposition 4 *Assume that* $w, f \in C^2[0,1], \sigma^2 \in C^1[0,1], g \in C^{p+2}[0,1]$. *If* $h \longrightarrow 0$ *and* $nh \longrightarrow \infty$ *when* $n \longrightarrow \infty$, *then for* $t \in (0,1)$ *and* p *odd*

$$E(\hat{g}_3(t;p)) = g(t) + h^{p+1} \frac{g^{(p+1)}(t)}{(p+1)!} \mu_{p+1}(K_{(p)}) + \mathrm{o}\left(h^{p+1}\right),$$

while for even p

$$E(\hat{g}_3(t;p)) = g(t) + h^{p+2} \left\{ \frac{g^{(p+1)}(t)}{(p+1)!} \frac{[w(t)f(t)]'}{w(t)f(t)} + \frac{g^{(p+2)}(t)}{(p+2)!} \right\} \mu_{p+2}(K_{(p)}) + \mathrm{o}\left(h^{p+2}\right).$$

In either case

$$V(\hat{g}_3(t;p)) = \frac{1}{nh} V(K_{(p)}) \frac{\sigma^2(t)}{f(t)} + \mathrm{o}\left(\frac{1}{nh}\right),$$

where $\mu_j(K_{(p)}) = \int u^j K_{(p)}(u) du$, $V(K_{(p)}) = \int K_{(p)}^2(u) du$.

REMARKS:

1) Let p be an even number. Taking $w(t) = 1/f(t)$, the asymptotic bias of $\hat{g}_3(t;p)$ is equal to those of $\hat{g}_3(t;p+1)$ and the LP estimator of order $p+1$. The same happens for their *AIMSE*'s. (We note that $K_{(p)} = K_{(p+1)}$ for p even.)

2) If p is an even number, it is possible to find a weight function (depending of the unknown function g) giving bias of order $\mathrm{o}\,(h^{p+2})$.

3) Consider weight functions $w(t) = f^{\alpha}(t)$ and assume p is even. It exists a theoretical α_0 such that the *AIMSE* of $\hat{g}_3(t;p)$ is lower than or equal to the *AIMSE* for the LP estimators of orders p and $p+1$.

APPENDIX: PROOFS

Lemma 1 *Assume $K \in Lip^1[-1,1]$, $s \in Lip^1[0,1]$. If $h \longrightarrow 0$ and $nh \longrightarrow \infty$ when $n \longrightarrow \infty$, then for any $t \in (0,1)$*

$$\int_0^1 f(u)s(u)K_h(t-u)du = \frac{1}{n}\sum_{i=1}^n s(t_i)K_h(t-t_i) + \mathrm{O}\left(\frac{1}{nh}\right).$$

For powers of kernel, $K^\nu, \nu \geq 1$, if $nh^\nu \longrightarrow \infty$

$$\int_0^1 f(u)s(u)K_h^\nu(t-u)du = \frac{1}{n}\sum_{i=1}^n s(t_i)K_h^\nu(t-t_i) + \mathrm{O}\left(\frac{1}{nh^\nu}\right).$$

Proof. Let $l_i = t_i - t_{i-1}$, $i = 1, \ldots, n+1$, $t_0 = 0$, $t_{n+1} = 1$; then $l_i = (nf(t_i))^{-1} + \mathrm{O}\left(1/n^2\right)$. Defining $I_t = \left\{i | \int_{t_{i-1}}^{t_i} K_h(t-u)du \neq 0\right\}$ and using mean values $\xi_i \in [t_{i-1}, t_i]$, we have

$$\begin{gathered}\int_0^1 f(u)s(u)K_h(t-u)du = \\ = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} f(u)s(u)K_h(t-u)du = \sum_{I_t} l_i f(\xi_i)s(\xi_i)K_h(t-\xi_i) = \\ = \sum_{I_t} \mathrm{O}\left(\tfrac{1}{n^2}\right) f(\xi_i)s(\xi_i)K_h(t-\xi_i) + \sum_{I_t} \tfrac{1}{nf(t_i)} f(\xi_i)s(\xi_i)K_h(t-\xi_i) = \\ = \mathrm{O}\left(\tfrac{1}{n}\right) + \sum_{I_t} \tfrac{1}{nf(t_i)} \left[f(t_i) + \mathrm{O}\left(\tfrac{1}{n}\right)\right]\left[s(t_i) + \mathrm{O}\left(\tfrac{1}{n}\right)\right]\left[K_h(t-t_i) + \mathrm{O}\left(\tfrac{1}{nh^2}\right)\right] = \\ = \tfrac{1}{n}\sum_{i=1}^n s(t_i)K_h(t-t_i) + \mathrm{O}\left(\tfrac{1}{nh}\right),\end{gathered}$$

where we have made use of $\mathrm{Card}(I_t) = \mathrm{O}\left(nh\right)$ and $K_h'(x) = \mathrm{O}\left(1/h^2\right)$. For $\nu > 1$ the result is proved with a similar argument. □

Lemma 2 *Under the assumptions of Lemma 1 and s, $f \in C^2[0,1]$:*

$$\frac{1}{n}\sum_{i=1}^n s(t_i)K_h(t-t_i) = s(t)f(t) + \frac{1}{2}h^2\mu_2(K)[s(t)f(t)]'' + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\frac{1}{nh}\right).$$

Proof. Let $p(t) = s(t)f(t)$; from Lemma 1 it follows that

$$\begin{gathered}\tfrac{1}{n}\sum_{i=1}^n s(t_i)K_h(t-t_i) + \mathrm{O}\left(\tfrac{1}{nh}\right) = \int_0^1 p(u)K_h(t-u)du = \int_{-1}^1 p(t-hx)K(x)dx = \\ = p(t) + \tfrac{1}{2}h^2 p''(t)\int_{-1}^1 x^2K(x)dx + \mathrm{o}\left(h^2\right).\end{gathered}$$

□

Proof of Proposition 1.

Define $\hat{g}_1(t) = N(t)/M(t)$, and let $p(t) = w(t)f(t)$. Lemma 2 gives

$$\begin{gathered}\tfrac{1}{n}E(N(t)) = \tfrac{1}{n}\sum_{i=1}^n w(t_i)K_h(t-t_i)g(t_i) = \\ = p(t)g(t) + \tfrac{1}{2}h^2\mu_2(K)\left[p(t)g(t)\right]'' + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\tfrac{1}{nh}\right), \\ \tfrac{1}{n}M(t) = \tfrac{1}{n}\sum_{i=1}^n w(t_i)K_h(t-t_i) = p(t) + \tfrac{1}{2}h^2\mu_2(K)p''(t) + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\tfrac{1}{nh}\right).\end{gathered}$$

Therefore, denoting $C=\mu_2(K)/2$ and, for the sake of simple presentation, omitting the dependence on t, we have

$$\begin{aligned}E\left(\hat{g}_1(t)\right) &= \frac{pg+Ch^2(pg''+2p'g'+p''g)+\mathrm{o}\left(h^2\right)+\mathrm{O}(1/nh)}{p+Ch^2p''+\mathrm{o}(h^2)+\mathrm{O}(1/nh)} = \\ &= \frac{pg+\mathrm{O}(1/nh)}{p+Ch^2p''+\mathrm{o}(h^2)+\mathrm{O}(1/nh)} + \frac{Ch^2(pg''+2p'g'+p''g)+\mathrm{o}\left(h^2\right)}{p+\mathrm{O}(h^2)} = \\ &= g - \frac{Ch^2p''g+\mathrm{o}\left(h^2\right)}{p+\mathrm{O}(h^2)} + \frac{Ch^2(pg''+2p'g'+p''g)+\mathrm{o}\left(h^2\right)}{p+\mathrm{O}(h^2)} + \mathrm{O}\left(\frac{1}{nh}\right) = \\ &= g + \frac{Ch^2R}{p} + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\frac{1}{nh}\right),\end{aligned}$$

where $R=pg''+2p'g'$, and the first part is proved.

Using Lemma 1 with ν=2, we have

$$\begin{aligned}\frac{1}{n^2}V(N(t)) &= \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^n w^2(t_i)\sigma^2(t_i)K_h^2(t-t_i)\right] = \\ &= \frac{1}{n}\left[\frac{1}{h}\int_{-1}^1 f(t-hx)\sigma^2(t-hx)w^2(t-hx)K^2(x)dx + \mathrm{O}\left(\frac{1}{nh^2}\right)\right] = \\ &= \frac{1}{nh}\left[\int_{-1}^1 (f(t)+\mathrm{o}\,(1))(\sigma^2(t)+\mathrm{o}\,(1))(w^2(t)+\mathrm{o}\,(1))K^2(x)dx\right] + \mathrm{O}\left(\frac{1}{nh^2}\right) = \\ &= \frac{1}{nh}\left(V(K)f(t)\sigma^2(t)w^2(t)+\mathrm{o}\,(1)\right).\end{aligned}$$

Using again Lemma 1

$$\begin{aligned}\frac{1}{n^2}M^2(t) &= \left[\frac{1}{n}\sum_{i=1}^n w(t_i)K_h(t-t_i)\right]^2 = \\ &= \left[\int_{-1}^1 K(x)f(t-hx)w(t-hx)dx + \mathrm{O}\left(\frac{1}{nh}\right)\right]^2 = f^2(t)w^2(t)+\mathrm{o}\,(1).\end{aligned}$$

Since $V(\hat{g}_1(t)) = V(N(t))/M^2(t)$, the second part follows. □

Lemma 3 *If* $g\in C^1[0,1]$, r *is a bounded function on* $[0,1]$, *and* s_i *is an interpolating sequence of* t_i, *i.e.* $s_0=0, t_i\le s_i\le t_{i+1}, s_n=1$, *then*

$$\int_0^1 r(u)g(u)du = \sum_{i=1}^n g(t_i)\int_{s_{i-1}}^{s_i} r(u)du + \mathrm{O}\left(\frac{1}{n}\right).$$

Proof. Observe that

$$\begin{aligned}\int_0^1 r(u)g(u)du &= \sum_{i=1}^n \int_{s_{i-1}}^{s_i} r(u)g(u)du = \\ &= \sum_{i=1}^n \int_{s_{i-1}}^{s_i} r(u)\left[g(t_i)+(u-t_i)g'(\theta_i(u))\right]du \\ &= \sum_{i=1}^n g(t_i)\int_{s_{i-1}}^{s_i} r(u) + \sum_{i=1}^n \int_{s_{i-1}}^{s_i} r(u)g'(\theta_i(u))\mathrm{O}\left(\frac{1}{n}\right)du,\end{aligned}$$

where $\theta_i(u)$ are intermediate values between t_i and u. Therefore

$$\left|\int_0^1 r(u)g(u)du - \sum_{i=1}^n g(t_i)\int_{s_{i-1}}^{s_i} r(u)du\right| \le \max(r)\max(g')\mathrm{O}\left(\frac{1}{n}\right) = \mathrm{O}\left(\frac{1}{n}\right).$$ □

Proof of Proposition 2.

By Lemma 3 we have

$$\begin{aligned}D(t)E(\hat{g}_2(t)) &= \sum_{i=n}^n g(t_i)\int_{s_{i-1}}^{s_i} w(u)K_h(t-u)du = \\ &= \int_0^1 w(u)g(u)K_h(t-u)du + \mathrm{O}\left(\frac{1}{n}\right) = \int_{-1}^1 w(t-hx)g(t-hx)K(x)dx + \mathrm{O}\left(\frac{1}{n}\right).\end{aligned}$$

Hence $D(t)E(\hat{g}_2(t)-g(t)) =$

$$= \int_{-1}^{1} w(t-hx)K(x)[g(t-hx)-g(t)]dx + \mathrm{O}\left(\frac{1}{n}\right).$$

Taylor expansions of $w(t-hx)$ and $g(t-hx)$ give straightforwardly

$$D(t)E(\hat{g}_2(t)-g(t)) = \frac{h^2}{2}\mu_2(K)\left(w(t)g''(t)+2g'(t)w'(t)\right) + \mathrm{o}\left(h^2\right) + \mathrm{O}\left(\frac{1}{n}\right).$$

Since

$$D(t) = \int_{-1}^{1} w(t-hx)K(x)dx = w(t) + \mathrm{O}\left(h^2\right),$$

the first part follows.

The proof for the second part is given, for the sake of simplicity, when $\sigma^2(t)$ is constant; the needed changes under heteroscedasticity are simples and analogous to those ones used in the second part of the proof of Proposition 1.

Recalling the definition of I_t and using

$$s_i - s_{i-1} = \frac{1}{nf(\xi_i)} + \mathrm{O}\left(\frac{1}{n^2}\right), \quad \xi_i \in [t_{i-1}, t_{i+1}]$$

(see, for instance, Müller, 1988, p.28), we have

$$D^2(t)V(\hat{g}_2(t)) = \sigma^2 \sum_{I_t}\left[\frac{1}{h}\int_{s_{i-1}}^{s_i} w(u)K\left(\frac{t-u}{h}\right)du\right]^2 =$$
$$= \frac{\sigma^2}{h^2}\sum_{I_t}\left[\frac{1}{nf(\xi_i)} + \mathrm{O}\left(\frac{1}{n^2}\right)\right] w(\eta_i)K\left(\frac{t-\eta_i}{h}\right)\int_{s_{i-1}}^{s_i} w(u)K\left(\frac{t-u}{h}\right)du,$$

where $\eta_i \in [s_{i-1}, s_i]$. Observe that

$$\int_{s_{i-1}}^{s_i} w^2(u)K^2\left(\frac{t-u}{h}\right)du =$$
$$= \int_{s_{i-1}}^{s_i} w(u)K\left(\frac{t-u}{h}\right)\left[w(\eta_i)+\mathrm{O}\left(u-\eta_i\right)\right]\left[K\left(\frac{t-\eta_i}{h}\right)+\mathrm{O}\left(\frac{u-\eta_i}{h}\right)\right]du =$$
$$= w(\eta_i)K\left(\frac{t-\eta_i}{h}\right)\int_{s_{i-1}}^{s_i} w(u)K\left(\frac{t-u}{h}\right)du + \mathrm{O}\left(\frac{h}{n}\right).$$

Since for $i \in I_t$

$$\frac{1}{nf(\xi_i)} = \frac{1}{n(f(t)+\mathrm{O}\left(h\right))},$$

we obtain $D^2(t)V(\hat{g}_2(t)) =$

$$= \frac{\sigma^2}{h^2}\sum_{I_t}\left[\frac{1}{n(f(t)+\mathrm{O}(h))} + \mathrm{O}\left(\frac{1}{n^2}\right)\right]\left[\int_{s_{i-1}}^{s_i} w^2(u)K^2\left(\frac{t-u}{h}\right)du + \mathrm{o}\left(h\right)\right] =$$
$$= \frac{\sigma^2}{nh^2(f(t)+\mathrm{O}(h))}\left[\int_0^1 w^2(u)K^2\left(\frac{t-u}{h}\right)du + \mathrm{O}\left(h^2\right)\right] +$$
$$+\mathrm{O}\left(\frac{1}{n^2h^2}\right)\int_0^1 w^2(u)K^2\left(\frac{t-u}{h}\right)du + \mathrm{O}\left(\frac{1}{n^2}\right) =$$
$$= \frac{\sigma^2}{nh(f(t)+\mathrm{O}(h))}\int_{-1}^{1} w^2(t-hx)K^2(x)dx + \mathrm{o}\left(\frac{1}{nh}\right) =$$
$$= \frac{\sigma^2}{nh(f(t)+\mathrm{O}(h))}\left[\int_{-1}^{1} w^2(t)K^2(x)dx + \mathrm{O}\left(h^2\right)\right] + \mathrm{o}\left(\frac{1}{nh}\right) =$$
$$= \frac{\sigma^2}{nhf(t)}\left[\int_{-1}^{1} w^2(t)K^2(x)dx\right] + \mathrm{o}\left(\frac{1}{nh}\right) = \frac{\sigma^2 w^2(t)}{nhf(t)}V(K) + \mathrm{o}\left(\frac{1}{nh}\right).$$

Noting that $D^2(t) = w^2(t) + \mathrm{O}\left(h^2\right)$, the result follows. □

Proof of Proposition 4.

It is based on the proof of Theorem 4.1 in Ruppert and Wand (1994).

We have

$$E(\hat{g}_3(t;p)) = e_1^{\mathrm{T}}(X_t^{\mathrm{T}} W_t X_t)^{-1} X_t^{\mathrm{T}} W_t (g(t_1), \ldots, g(t_n))^{\mathrm{T}},$$

where $X_t = (C_0(t), C_1(t), \ldots, C_p(t))$, $C_r(t) = ((t_1-t)^r, \ldots, (t_n-t)^r)^{\mathrm{T}}$, $r \geq 0$, $W_t = diag\{w(t_i)K_h(t_i - t), i = 1, \ldots, n\}$ and $e_1^{\mathrm{T}} = (1, 0, \ldots, 0)$. The expansion of $g(t_i)$ around t gives

$$bias(\hat{g}_3(t;p)) = e_1^{\mathrm{T}}(X_t^{\mathrm{T}} W_t X_t)^{-1} X_t^{\mathrm{T}} W_t (R_{p+2}(t) + R(t)),$$

where

$$R_{p+2}(t) = \textstyle\sum_{j=1}^{2} \frac{g^{(p+j)}(t)}{(p+j)!} C_{p+j}(t),$$

and $R(t)$ is a vector of remainder terms. We also can write

$$bias(\hat{g}_3(t;p)) = e_1^{\mathrm{T}} S_n^{-1} \left(\textstyle\sum_{j=1}^{2} \frac{g^{(p+j)}(t)}{(p+j)!} \frac{1}{n} X_t^{\mathrm{T}} W_t C_{p+j}(t) \right) + R^*(t), \tag{1}$$

where $S_n = n^{-1}(X_t^{\mathrm{T}} W_t X_t)$. Defining $s_w^r(t) = n^{-1}\sum_{i=1}^{n}(t_i - t)^r w(t_i) K_h(t - t_i)$, the S_n matrix has (i,j)th entry equal to $s_w^{i+j-2}(t)$, and the $(p+1)$ vector $n^{-1}X_t^{\mathrm{T}} W_t C_{p+j}(t)$ has jth element equal to $s_w^{p+j}(t)$.

Lemma 1 provides

$$\begin{aligned} s_w^r(t) &= \textstyle\int_0^1 f(u)(u-t)^r w(u) K_h(u-t)du + \mathrm{O}\left(\frac{1}{nh}\right) = \\ &= \textstyle\int_{-1}^1 (hx)^r f(t+hx) w(t+hx) K(x) dx + \mathrm{O}\left(\frac{1}{nh}\right) = \\ &= h^r p(t)\mu_r + h^{r+1} p'(t)\mu_{r+1} + \mathrm{o}\left(h^{r+1}\right) = \qquad (2) \\ &= h^r p(t)\mu_r + \mathrm{o}\left(h^r\right), \qquad (3) \end{aligned}$$

where $p(t) = f(t)w(t)$.

Let $H = diag\{1, h, \ldots, h^p\}$. It follows from (3) that

$$S_n = H\left(p(t)N_p + hp'(t)Q_p\right)H + \mathrm{o}\left(hH\mathbf{1}H\right),$$

where Q_p is the $(p+1)\times(p+1)$ matrix having (i,j)th entry equal to μ_{i+j-1} and $\mathbf{1}$ is a $(p+1)\times(p+1)$ matrix with all the elements equal to 1. Therefore

$$e_1^{\mathrm{T}} S_n = \frac{1}{p(t)}\left(e_1^{\mathrm{T}} N_p^{-1} - h\frac{p'(t)}{p(t)} e_1^{\mathrm{T}} N_p^{-1} Q_p N_p^{-1}\right) H^{-1} + \mathrm{o}\left(h\mathbf{1}H^{-1}\right). \tag{4}$$

From (2), we have similarly

$$\begin{aligned} n^{-1}X_t^{\mathrm{T}} W_t C_{p+1}(t) &= h^{p+1}p(t)H(\mu_{p+1}, \mu_{p+2}, \ldots, \mu_{2p+1})^{\mathrm{T}} + \\ &+ h^{p+2}p'(t)H(\mu_{p+2}, \mu_{p+3}, \ldots, \mu_{2p+2})^{\mathrm{T}} + \mathrm{o}\left(h^{p+2}\right), \end{aligned}$$

and

$$n^{-1}X_t^{\mathrm{T}} W_t C_{p+2}(t) = h^{p+2}p'(t)H(\mu_{p+2}, \mu_{p+3}, \ldots, \mu_{2p+2})^{\mathrm{T}} + \mathrm{o}\left(h^{p+2}\right).$$

Using these results and (4), we have that

$$bias(\hat{g}_3(t;p)) = \left\{\sum_{j=1}^{p+1}\left(N_p^{-1}\right)_{1j}\mu_{p+j+1}\right\}\frac{g^{(p+1)}(t)}{(p+1)!}h^{p+1}+$$
$$\left[\left\{\sum_{j=1}^{p+1}\left(N_p^{-1}\right)_{1j}\mu_{p+j+2}\right\}\frac{g^{(p+2)}(t)}{(p+2)!}+\right.$$
$$\left\{\sum_{j=1}^{p+1}\left(N_p^{-1}\right)_{1j}\mu_{p+j+2} - e_1^{\mathrm{T}}N_p^{-1}Q_pN_p^{-1}(\mu_{p+1},\mu_{p+2},\ldots,\mu_{2p+1})^{\mathrm{T}}\right\}$$
$$\left.\frac{g^{(p+1)}(t)}{(p+1)!}\frac{p'(t)}{p(t)}\right]h^{p+2}+\mathrm{o}\left(h^{p+2}\right).$$

The agreement of the kernel dependent constants with those given in the Proposition 4 is proved in Ruppert and Wand (1994), pp. 1365-1366.

For the variance, first note that

$$V(\hat{g}_3(t;p)) = n^{-1}e_1^{\mathrm{T}}S_n^{-1}V_nS_n^{-1}e_1, \tag{5}$$

where $V_n = n^{-1}X_t^{\mathrm{T}}W_t\Sigma_tW_tX_t$, and $\Sigma_t = diag\{\sigma_i^2(t), i=1,\ldots,n\}$.

Now, let

$$v_w^r(t) = n^{-1}\sum_{i=1}^n(t_i-t)^r w^2(t_i)K_h^2(t_i-t)\sigma^2(t_i),$$

so, the (i,j)th entry of matrix V_n is $v_w^{i+j-2}(t)$.

Using Lemma 1 with $\nu=2$, we obtain

$$v_w^r(t) = \int_0^1 f(u)(u-t)^r w^2(u)K_h^2(u-t)\sigma^2(u)du,$$

and, analogously to (3),

$$v_w^r(t) = h^{r-1}f(t)w^2(t)\sigma^2(t)\rho_r + \mathrm{o}\left(h^{r-1}\right),$$

where $\rho_r = \int u^rK^2(u)du$. Consequently, we have

$$V_n = n^{-1}f(t)w^2(t)\sigma^2(t)HR_pH + \mathrm{o}\left(\frac{1}{h}H\mathbf{1}H\right),$$

where R_p is analogous to N_p but with μ_{i+j-2} replaced by ρ_{i+j-2}.

Using (3), we obtain

$$S_n = f(t)w(t)HN_pH + \mathrm{o}\left(\frac{1}{h}H\mathbf{1}H\right).$$

Finally, we have from (5)

$$V(\hat{g}_3(t;p)) = \frac{\sigma^2(t)}{nh}e_1^{\mathrm{T}}N_p^{-1}R_pN_p^{-1}e_1 + \mathrm{o}\left(\frac{1}{nh}\right).$$

Ruppert and Wand (1994), p. 1366, prove that $e_1^{\mathrm{T}}N_p^{-1}R_pN_p^{-1}e_1 = \int K_{(p)}^2(u)du$, giving the required result. □

ACKNOWLEDGMENTS

Part of the article was written when the authors visited the Department of Statistics, University of Berkeley. In this visit, the second author was supported by a Del Amo Foundation Grant 1995. The authors are grateful to the referees and an associate editor for helpful comments and suggestions.

BIBLIOGRAPHY

Chu, C. -K. and Marron, J. S. (1991). "Choosing a kernel regression estimator", *Statist. Sci.*, **6**, 404-419.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chapman and Hall, New York.

Gasser, T., Kneip, A. and Köhler, W. (1991). "A flexible and fast method for automatic smoothing", *J. Amer. Statist. Assoc.*, **86**, 643-652.

Gasser, T. and Müller, H.-G. (1979). "Kernel estimation of regression functions". In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rossenblat, Eds.), p. 23–68. Springer-Verlag, Heidelberg.

Hall, P. and Wehrly, T. E. (1991). "A geometrical method for removing edge effects from kernel-type nonparametric regression estimators". *J. Amer. Statist. Assoc.*, **86**, 665–672.

Jennen-Steinmetz, C. and Gasser, T. (1988). "A unifying approach to nonparametric regression estimation". *J. Amer. Statist. Assoc.*, **83**, 1084–1089.

Jones, M. C., Davies, S. J. and Park, B. U. (1994). "Versions of Kernel-Type Regression Estimators". J. Amer. Statist. Assoc., 89, 825-832.

Köhler, W. (1990). "CURVDAT 90.1: Nonparametric smoothing, differentiation and density estimation. User's manual". *STATCOM*, Institut für Statistik und Computing, Neckargemünd.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.*, **46**. Springer-Verlag, New York.

Nadaraya, E. A. (1964). "On estimating regression". *Theory Prob. Appl.*, **10**, 186–190.

Rice, J. (1984). "Boundary modification for kernel regression". *Commun. Statist.-Theory and Meth.*, **13**, 893–900.

Ruppert, D. and Wand, M. P. (1994). "Multivariate locally weighted least squares regression". *Annals of Statistics*, **22**, 1346–1370.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, New York.

Watson, G. S. (1964). "Smooth regression analysis". *Sankhya,* Ser. A, **26**, 101–116.