On the convolution type kernel regression estimator

By

Eva Herrmann, Darmstadt

Abstract

This paper discusses modifications of the convolution type kernel regression estimator. One modification uses kernel quantile estimators and is analyzed more detailed. This regression estimator combines advantages of local polynomial and kernel regression estimators and can be applied for small to large sample size. Its properties are illustrated by simulation results and asymptotic theory. Especially the minor effect of bandwidth choice for the kernel quantile estimator on the regression estimator is demonstrated. A simple adaptation on sample size leads to an interesting regression estimator.

1 Introduction

Suppose that a nonparametric estimator of the conditional mean

$$r(t) = E(Y|T = t)$$

should be derived from a sample of n independent and identically distributed bivariate random variables. Let (T_i, Y_i) denote the sample points which are already ordered in the first variable and let $\varepsilon_i = Y_i - E(Y|T = T_i)$ denote the residuals. Then we have a nonparametric regression model in the form

$$Y_i = r(T_i) + \varepsilon_i \tag{1.1}$$

for $i = 1, \ldots, n$, where r is the unknown regression function. For the sake of simplicity we suppose that $T_1, \ldots, T_n$ and r are restricted to the unit interval. Several types of kernel regression estimators for estimating the regression function nonparametrically have been proposed and discussed in literature. Let K be a kernel function of order k and h be a bandwidth. Mainly important kernel estimators are the Nadaraya-Watson type estimator (Nadaraya, 1964, Watson, 1964)

$$\hat{r}_1(t;h) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{t-T_i}{h}\right) Y_i \Big/ \sum_{i=1}^{n} K\left(\frac{t-T_i}{h}\right),$$

the Gasser-Müller type estimator (Gasser and Müller, 1979)

$$\hat{r}_2(t;h) = \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \frac{1}{h} K\left(\frac{t-u}{h}\right) du Y_i$$

with the additional definitions $s_0 = 0$, $s_i = \frac{1}{2}(T_i + T_{i+1})$ for $i = 1, \ldots, n-1$ and $s_n = 1$, the Priestley-Chao type estimator (Priestley and Chao, 1972)

$$\hat{r}_3(t;h) = \sum_{i=2}^{n} (T_i - T_{i-1}) \frac{1}{h} K\left(\frac{t-T_i}{h}\right) Y_i$$

and the local polynomial estimator r_4 of order p with kernel weights as solution of

$$\sum_{i=1}^{n} K\left(\frac{t-T_i}{h}\right) \left(Y_i - r_4(t;h) - \sum_{j=1}^{p} a_j (t-T_i)^j\right)^2 = \min!$$

proposed e.g. by Stone (1977) and Cleveland (1979). These are closely related nonparametric regression estimators and in the situation of a regression model (1.1) with nonrandom but equidistant design-points $T_1, \ldots, T_n$ they nearly coincide. Nevertheless, there remain well-known differences between them in case of random design points. Here, the Gasser-Müller and the Priestly-Chao type kernel estimators are inefficient asymptotically since their asymptotic variance is about twice or 1.5 times the variance of the asymptotically minimax optimal linear estimator. The local polynomial estimator and the Nadaraya-Watson estimator can have infinite variance but the conditional variance equals almost everywhere the variance of the asymptotically minimax optimal linear estimator, e.g. Fan (1993). See Seifert and Gasser (1995) for a more detailed analysis of the finite sample variance of the local polynomial estimator. The Nadaraya-Watson estimator instead has an undesirable bias term for non equidistant design in addition to the asymptotic bias of the other kernel estimators. These and other aspects of the different choices of the weights for a kernel regression estimator are discussed in detail in several papers, e.g. Gasser and Engel (1990), Fan (1993), Jones, Davies and Park (1994) and the discussion papers of Chu and Marron (1991) and Hastie and Loader (1993).

Here we will mainly restrict our attention on the Gasser-Müller and the Priestley-Chao estimator. The next section will give an overview on several modifications which had been proposed recently. They are illustrated by some simulation examples. In the third section some asymptotic results are proved on optimal choice of the additional smoothing parameter. Technical parts of the proofs are deferred to the Appendix.

2 Comparison of convolution type kernel methods

We now concentrate on some modifications of the estimators $\hat{r}_2$ and $\hat{r}_3$ which are often refered to as convolution type kernel estimators. It has been shown recently by several authors that it is possible to modify these estimators in a way that bias and variance terms have the same asymptotic representations as the local polynomial estimator but also in an unconditional way. They do not share the numerical difficulties of the local polynomial estimators for sparse regions. The asymptotic representation of MSE which seemed to be most appropriate (compare e.g. Chu and Marron, 1991 and Jones et al., 1994) for kernel estimation is given by

$$MSE(\hat{r}_2^*(t;h)) = \frac{h^{2k}}{k!}\left\{r^{(k)}(t)\right\}^2 \int x^k K(x)\,dx + \frac{\sigma^2(t)\int\{K(x)\}^2\,dx}{nhf(t)} + o(h^{2k}+n^{-1}h^{-1}), \quad (2.1)$$

where $\sigma^2(t) = Var(Y|T=t)$ and f denotes the density of T. Both functions are supposed to be twice continuously differentiable on the unit interval and to be bounded away from zero. The regression function r is assumed to be at least k-times continuously differentiable.
The convolution type kernel methods discussed here all have the joint form

$$\hat{r}_2^*(t;h) = \sum_{i=1}^{n} c_i \int_{a_i}^{b_i} \tfrac{1}{h}K\left(\tfrac{t-u}{h}\right)du\,Y_i \quad (2.2)$$

of a generalized Gasser-Müller type estimator or

$$\hat{r}_3^*(t;h) = \sum_{i=1}^{n} c_i(b_i-a_i)\tfrac{1}{h}K\left(\tfrac{t-T_i}{h}\right)Y_i$$

of a generalized Priestley-Chao type estimator. Since differences between both estimators are very small we concentrate on the generalized Gasser-Müller type estimator in the following. If polynomial kernels are used the integration does not lead to further difficulties and $\hat{r}_2^*$ has the slight computational advantage of adapting to sparse regions for arbitrarily small bandwidths automatically and of summing up to 1 at least for $c_i = 1$ and $a_{i+1} = b_i$ but compare Jones et al. (1994) for a different point of view. With $a_1 = 0$, $a_i = 0.5(T_{i-1}+T_i)$ for $i = 2,\ldots,n$, $b_i = a_{i+1}$ for $i = 1,\ldots,n-1$ and $b_i = 1$, $c_i = 1$ for all $i = 1,\ldots,n$ we obtain the form of the Gasser-Müller estimator $\hat{r}_2$. This estimator was first motivated from a nonparametric regression model with fixed design variables. Its problems for the random design case stem from the variability of the differences $b_i - a_i$ which leads to an inflation of the variance of $\hat{r}_2$ and $\hat{r}_3$. Hence the proposed modifications will reduce this variability.

A different modification of the Gasser-Müller type estimator with the use of binning methods can be found in Kneip and Engel (1994). They prove that such a regression estimator does have MSE of the form (2.1).

2.1 Enlarging the integration regions

One approach is given by Hall and Turlach (1995) by defining either

$$a_i = \nu^{-1} \sum_{j=1}^{\nu} T_{i-\nu+j}, \quad b_i = \nu^{-1} \sum_{j=1}^{\nu} T_{i+j-1} \text{ and } c_i = \frac{1}{\nu - 1} \tag{2.3}$$

or $a_i = T_{i-\nu+1}$, $b_i = T_{i+\nu}$ and $c_i = 1/(2\nu - 1)$. For $\nu = 2$ and the first definitions we obtain the Gasser-Müller estimator $\hat{r}_2$. Using $\nu > 2$ may lead to values a_i, b_i which do not satisfy $a_{i+1} = b_i$ for $i = 1, \dots, n-1$. Hence the modified estimator can not be interpreted in a simple way as convolution of K and a step function with values Y_i as the original Gasser-Müller approach. Because of the additional averaging it does decrease the variance of the differences $b_i - a_i$. Hall and Turlach (1995) proved that the asymptotic representation of the mean squared error of equation (2.1) holds for all ν which tend to infinity more slowly than n^α for each $\alpha > 0$ under the usual regularity conditions. They also give a proposal to choose the additional parameter ν as largest integer smaller than or equal to $\log n$ which is motivated from simulation results.

2.2 Using quantile estimators

A second approach is sketched by Chu and Marron (1991) who propose to choose

$$a_{i+1} = b_i = \sum_{j=1}^{2\nu+2} \frac{1}{2\nu + 2} X_{i-\nu+j-1}$$

and $c_i = 1$ for $i = \nu + 1, \dots, n - \nu - 1$. Obviously the Gasser-Müller estimator $\hat{r}_2$ is obtained for $\nu = 0$. This approach is discussed by Jones et al. (1994) who use the Priestley-Chao kernel estimator and are therefore mainly interested in $b_i - a_i$. They also sketch a more general proposal with a kernel function K_s and a bandwidth g by

$$a_{i+1} = b_i = \sum_{j=1}^{n} T_j \frac{1}{g} K_s \left(\frac{(i/n) - (j/n))}{g} \right) \tag{2.4}$$

and $c_i = 1$. Using the uniform kernel as kernel function K_s and bandwidth $g = (\nu + 1)/n$ gives the discrete form above. Here a_i and b_i are kernel quantile estimators of the design distribution. This approaches still allows the interpretation as convolution of the kernel function K and a step function but the condition that $T_i \in [a_i, b_i]$ is no longer guaranteed. Nevertheless, Jones et al. (1994) state without proof that for the estimators chosen as in (2.3) additional bias terms will be of order $O(\nu^2 n^{-2})$ and mean squared error will have the form of equation (2.1) as long as ν tends to infinity with n more slowly than n and faster than nh^k. They illustrate the effect of choosing ν in a simulation example but they do not give a rule of thumb for choosing ν in practice.

In the following we will study the general form of this estimator with kernel quantile estimators for the design distribution. In order to avoid slight symmetry problems which are included in definition of (2.4) and to adapt automatically to arbitrarily small bandwidths $g > 0$ we choose

$$a_i = \sum_{j=1}^{n} T_j \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \frac{1}{g} K_s\left(\frac{\frac{i-0.5}{n+1} - v}{g}\right) dv \quad b_i = \sum_{j=1}^{n} T_j \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \frac{1}{g} K_s\left(\frac{\frac{i+0.5}{n+1} - v}{g}\right) dv \qquad (2.5)$$

for $i = 1, \ldots, n$ and $c_i = 1$ with a symmetric kernel function K_s of order $k_s \in \{2, 4\}$. For $g \leq \frac{1}{n+1}$ we obtain the Gasser-Müller estimator $\hat{r}_2$. Of course we should modify the quantile estimator in the boundary region e.g. by the use of boundary kernels instead of K_s. The unconditionally asymptotic mean squared error of this estimator is derived in the next section. Thereby we have to look at second order terms in order to find an asymptotically optimal bandwidth g which is shown to be of order $n^{-(3k+1)/((2k+1)(ks+1))}$, typically. From a practical point of view and since only second order terms of the mean squared error are minimized we may use the general proposal $g = 0.75(n+1)^{-(3k+1)/((2k+1)(ks+1))}$.

2.3 Simulation Examples

The following figures illustrate the described estimators. We have an example of $n = 50$ design points in $[0, 1]$ and have a look at the weights of the different estimators for $t = 1/2$ using the biweight kernel function $K(x) = 15/16(1 - x^2)^2 I_{[-1,1]}$ and bandwidth $h = 1/2$.

Figure 1 shows the weights associated to the observations Y_i at points Y_i for the Nadaraya-

Watson estimator $\hat{r}_1$ and the local linear estimator $\hat{r}_4$.

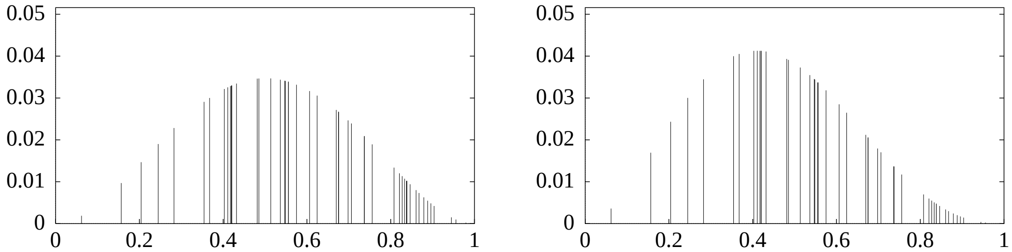

Figure 1: Weights of the Nadaraya-Watson kernel estimator (left) and of the local linear estimator (right) for a simulation example with $n = 50$ design points

The Gasser-Müller estimator $\hat{r}_2$ and the modifications proposed by Hall and Turlach are shown in Figure 2 for $\nu = 3$ to $\nu = 5$.

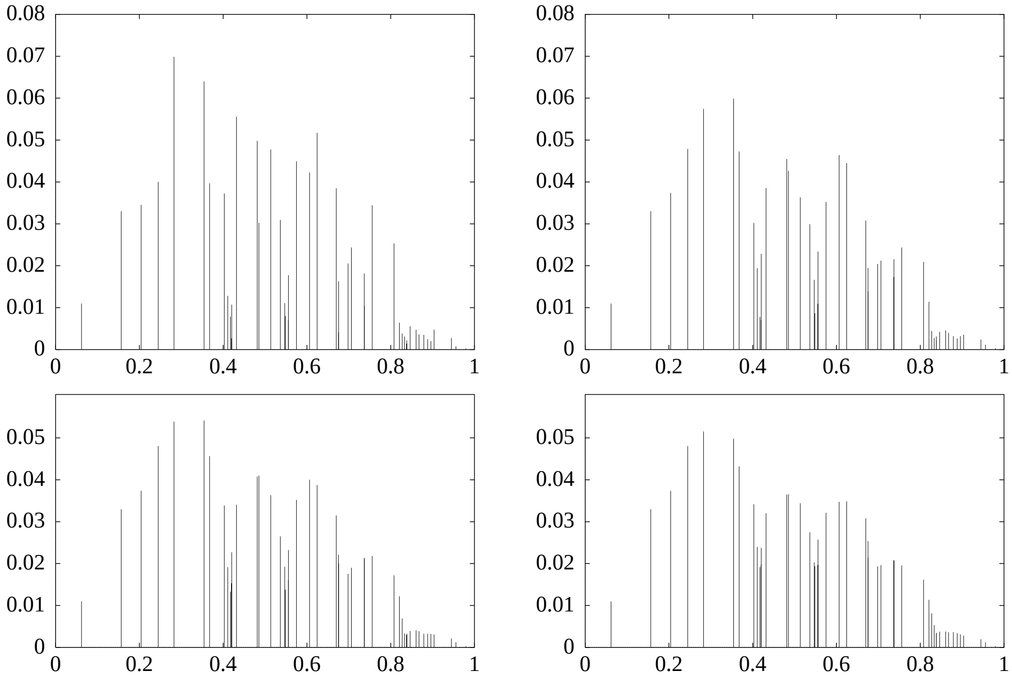

Figure 2: Weights of the Gasser-Müller kernel estimator (left, above) and of the modified estimation (2.2) with definitions (2.3) for $\nu = 3$ (right, above), $\nu = 4$ (left, below) and $\nu = 5$ (right, below). Simulation example as in Figure 1.

Figure 3 instead illustrates the modification of estimator (2.2) with definitions (2.5). Thereby the bandwidths $g = 0.1(n+1)^{-7/15}$, $g = 0.5(n+1)^{-7/15}$, $g = 0.75(n+1)^{-7/15}$ and $g = 1.0(n+1)^{-7/15}$ are used. It demonstrates that the modification mainly performs a smoothing of the weights.

For a further comparison we calculate the conditional bias and variance of these design points for regression function r_1 with $r_1(x) = 5(x - \frac{1}{2})^2$, regression function r_2 with $r_2(x) = 2 - 5x + 5\exp\{-400(x-0.5)^2\}$ (compare e.g. Seifert and Gasser, 1995 and Hall and Turlach, 1995) and $r_3(x) = 5\sin(x\pi/2)$ and constant variance σ^2. As can be expected from the figures the modifications of Figure 2 reduce variance for $\nu = 3$ to 5. For $\nu = 5$ the variance is about the same as for the local linear estimator, bias is slightly increased, typically.
The modifications of Figure 3 also reduce variance. For $g = 0.5(n+1)^{-7/15}$ we have about the same variance as for the local polynomial estimator, for $g = 0.75(n+1)^{-7/15}$ and $g = 1.0(n+1)^{-7/15}$ it is significantly smaller whereas bias is nearly unchanged here.

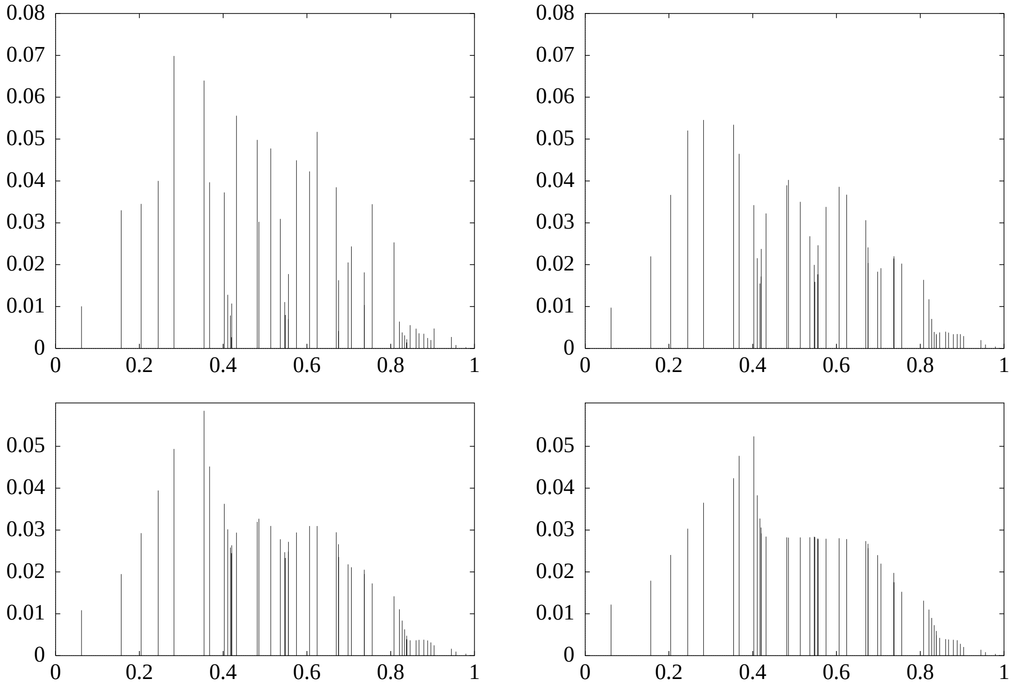

Figure 3: Weights of the Gasser-Müller kernel estimator (left, above) and of the modified estimation (2.2) with definitions (2.5) and bandwidths $g = 0.1(n+1)^{-7/15}$ (left, above), $g = 0.5(n+1)^{-7/15}$ (right, above), $g = 0.75(n+1)^{-7/15}$ (left, below) and $g = 1.0(n+1)^{-7/15}$ (right, below). Simulation example as in Figure 1.

More exact information of the modified estimators can be obtained by calculating the relative efficiency as it was defined in Seifert and Gasser (1995) and has also be done in Hall and Turlach (1995). Thereby conditional mean integrated squared error for the global bandwidths which minimizes this error are obtained from simulations. The asymptotical optimal mean integrated squared error is divided by the mean of 500 such replications. As was illustrated in

Seifert and Gasser (1995) such a measure reflects only a small part of the instability of the local polynomial estimator for small sample size. To obtain results which are comparable to the ones obtained by Seifert and Gasser (1995) and Hall and Turlach (1995) we use regression function r_2, sample sizes $n = 25$, 50, 100, 250, 500, 1000, 1500 and 2500 and constant variance function $\sigma^2 = 0.5$. The design points are obtained from uniformly distributed pseudo random numbers. Figure 4 shows the results for the local linear estimator and the Gasser-Müller estimator and the modified estimator (2.2) with definitions (2.5) and bandwidth $g = 0.75(n+1)^{-7/15}$.

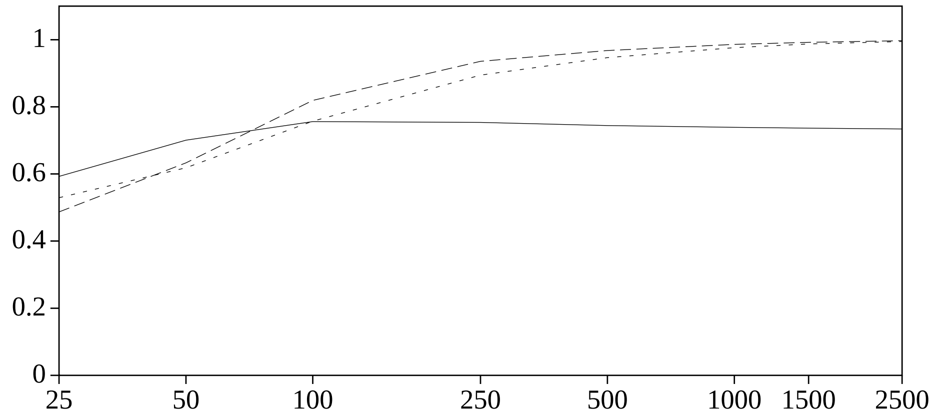

Figure 4: Relative efficiency of the local linear estimator (joined), the Gasser-Müller estimator (dashed) and the modified estimator (2.2) with (2.5) and $g = 0.75(n+1)^{-7/15}$ (dotted) for regression function r_2 and $\sigma^2 = 0.5$.

As in Hall and Turlach (1995) and Seifert and Gasser (1995) we do exclude boundary problems by computing additional observations outside the unit interval.

It is easy to see that the modifications lead to an estimator with mean integrated squared error slightly worse than the Gasser-Müller estimator for small sample sizes and a good and efficient behaviour for large sample sizes. This was also reflected by simulations with different regression design and variance functions. Figure 5 shows e.g. the analogous results for regression function r_1 with a linear design density f_1, $f_1(x) = 0.1 + 1.9 \cdot x$, and variance $\sigma^2 = 0.1$ and for regression function r_3 with a truncated $N(0.5, 0.25)$ design density and variance $\sigma^2 = 0.25$.

Summarizing these results and several others from simulations with different variances and regression functions one can state that the proposed estimator behaves better than a local linear estimator with epanechnikov weights for relatively small bandwidth and small sample size whereas it behaves much better than the Gasser-Müller kernel estimator for relatively large bandwidths and large sample size. Especially it does not share the numerical instabilities of the local polynomial estimators in sparse regions. There was only one situation observed in simulations with a variable regression function, a very small variance and small sample size where the general bandwidth $g = 0.75(n+1)^{-(3k+1)/((2k+1)(ks+1))}$ leads to a significant increase of bias and hence was too large.

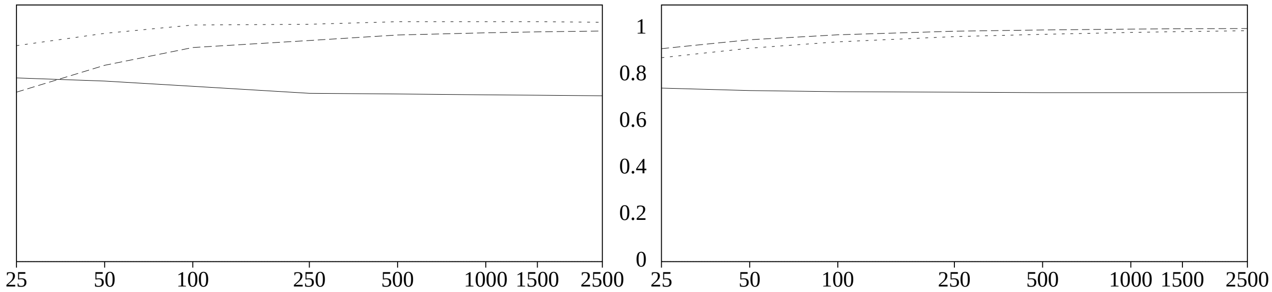

Figure 5: Relative efficiency of the local linear estimator (joined), the Gasser-Müller estimator (dashed) and the modified estimator (2.2) with (2.5) and $g = 0.75(n+1)^{-7/15}$ (dotted) for regression function r_1 and $\sigma^2 = 0.1$ and linear design density (left) and for regression function r_3 and $\sigma^2 = 0.25$ and a truncated normal design (right).

The proposed kernel estimator with the general bandwidth $g = 0.75(n+1)^{-(3k+1)/((2k+1)(ks+1))}$ can also be used for fixed design, since the differences to the unmodified Gasser-Müller estimator are very small then and vanish asymptotically. This nice property does not hold for the approach of Hall and Turlach (1995).

It can also be used for mixed design situations e.g. a design with fixed points which are slightly disturbed and similar situations. It does also perform well for multiple data points and shares the nice property of local polynomial estimators by using the mean of such points at least asymptotically. This is not true for the classical Gasser-Müller estimator, compare Chu and Marron (1991) for a discussion of this property.

3 Consistency results and asymptotically optimal bandwidth

Here we analyze mean squared error and mean integrated squared error of estimator $\hat{r}_2^*$ from equation (2.2) with definitions (2.5) unconditionally.

As before we suppose that we use a kernel K of order k and a kernel K_s of order $k_s = 2$ or $k_s = 4$ both with support $[-1, 1]$ and appropriate boundary modifications. Thereby we say as usual that a kernel L is of order ℓ if $\int x^j L(x)\,dx = \delta_{0,j}$ for $j = 0, \ldots, \ell - 1$ and $\int x^\ell L(x)\,dx \neq 0$. We also suppose that K, K_s and the respective boundary kernels satisfy the smoothness conditions which are stayed in the beginning of Appendix A.2. We will use the following abbreviations

$$\mu_j(L) = \int_{-1}^{1} x^j L(x)\,dx \text{ and } M(L) = \int_{-1}^{1} \{L(x)\}^2\,dx$$

for an arbitrary kernel function L. In the following we will use the notation $K(\cdot; t, h)$ to indicate a possible dependence of the kernel function on t and h at the boundary when using boundary kernels.

The regression function r is assumed to be at least k-times and the quantile function F^{-1} of the design variable T is assumed to be at least 4-times continuously differentiable on $[0, 1]$.

First we prove an asymptotic result on the local bias $B(t; h, g)$ of the kernel estimator.

Proposition 1: *The kernel estimator $\hat{r}_2^*(t; h)$ from (2.2) with definitions (2.5), $h = o(1)$, $g = o(1)$, $n^{-1}h^{-2} = o(1)$ and $n^{-1}g^{-2} = o(1)$ satisfies*

$$\begin{aligned} B(t; h, g) &= \frac{h^k}{k!} r^{(k)}(t)\mu_k(K) + \frac{g^{ks}}{(ks)!} r'(t)(F^{-1})^{(ks)}(F(t))\,\mu_{ks}(K_s) + R(t; h) \\ &\quad + o(g^{ks}) + O(n^{-1} + n^{-1}g^{1/2}h^{-1} + g^{2ks}h^{-1}) \end{aligned}$$

uniformly for all $t \in [h + g, 1 - h - g]$. Thereby the rest $R(t; h) = o(h^k)$ does not depend on the bandwidth g. Additionally it holds

$$\sup_{t\in[0,1]} |B(t; h, g) - \frac{h^k}{k!} r^{(k)}(t)\mu_k(K(\cdot; t, h))| = \tilde{R}(h) + O(g^{ks}) + O(n^{-1} + n^{-1}g^{1/2}h^{-1} + g^{2ks}h^{-1})$$

with a rest $\tilde{R}(h) = o(h^k)$ that does not depend on g.

Proof of Proposition 1: The typically dominating bias term is given by

$$B_1(t; h) = \int K(x; t, h) r(t - hx)\,dx - r(t)$$

and does not depend on g. With usual arguments, compare e.g. Müller (1988) we find the well known representation

$$B_1(t; h) = \frac{h^k}{k!} r^{(k)}(t)\mu_k(K) + R(t; h),$$

where R and

$$\tilde{R}(h) = \sup_{t\in[0,1]} |B_1(t;h) - \frac{h^k}{k!} r^{(k)}(t)\mu_k(K(\cdot;t,h))|$$

satisfy the proposed bounds of convergence.

We are now analyzing the additional bias term

$$\begin{aligned} B_2(t;h,g) &= B(t;h,g) - B_1(t;h) \\ &= E\left(\sum_{i=1}^{n}\int_{a_i}^{b_i} \frac{1}{h} K\left(\frac{t-u}{h};t,h\right)\{r(T_i)-r(u)\}\,du\right). \end{aligned}$$

For fixed design and $T_i \in [a_i, b_i]$ for all $i = 1,\dots,n$ which is satisfied e.g. for the Gasser-Müller estimator we obtain the known bound $|B_2| \le \frac{C}{n}$ for some global constant $C > 0$ (Müller, 1988). Since the condition $T_i \in [a_i, b_i]$ is no longer guaranteed we have to analyze B_2 more thoroughly. Denote

$$S_2 = \sum_{i=1}^{n}\int_{a_i}^{b_i} \frac{1}{h} K\left(\frac{t-u}{h};t,h\right)\{r(T_i)-r(u)\}\,du$$

and $\bar{t}_i = (F^{-1})(\frac{i}{n+1})$. We can find $\sigma_i \in (a_i, b_i)$ and $\eta_i, \tau_i, \lambda_i \in [0,1]$ with

$$\begin{aligned} S_2 &= \sum_{i=1}^{n} \frac{1}{h}(b_i-a_i)\left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right) + K\left(\frac{t-\sigma_i}{h};t,h\right) - K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}\{r(T_i)-r(\sigma_i)\} \\ &= \sum_{i=1}^{n} \frac{1}{h}(b_i-a_i)(T_i-\sigma_i) r'(\bar{t}_i) K\left(\frac{t-\bar{t}_i}{h};t,h\right) \\ &\quad + \sum_{i=1}^{n} \frac{1}{h}(b_i-a_i)(T_i-\sigma_i)\{\lambda_i(T_i-\bar{t}_i) + (1-\lambda_i)(\sigma_i-\bar{t}_i)\} r''(\eta_i) K\left(\frac{t-\tau_i}{h};t,h\right) \\ &\quad + \sum_{i=1}^{n} \frac{1}{h}(b_i-a_i)(T_i-\sigma_i)\left\{K\left(\frac{t-\sigma_i}{h};t,h\right) - K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\} r'(\mu_i). \end{aligned}$$

For the expectation of the first sum, it follows from Lemma 1 (ii) that

$$\begin{aligned} &E\left(\sum_{i=1}^{n} \frac{1}{h}(b_i-a_i)(T_i-\sigma_i) r'(\bar{t}_i) K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right) \\ &= \sum_{i=1}^{n} \frac{(F^{-1})'\left(\frac{i}{n+1}\right)}{(n+1)h} (F^{-1})^{(ks)}\left(\frac{i}{n+1}\right) r'(\bar{t}_i) K\left(\frac{t-\bar{t}_i}{h};t,h\right) \frac{g^{ks}}{ks!}\mu_{ks}(K_s(\cdot;\frac{i}{n+1},h)) \\ &\quad + O(n^{-1}) + o(g^{ks}). \end{aligned}$$

Application of Lemma 4(i) and (ii) of the Appendix on the expectation of the second sum yields

$$\begin{aligned}& E\left(\sum_{i=1}^{n}\frac{1}{h}(b_i-a_i)(T_i-\sigma_i)\{\lambda_i(T_i-\bar{t}_i)+(1-\lambda_i)(\sigma_i-\bar{t}_i)\}r''(\eta_i)K\left(\frac{t-\tau_i}{h};t,h\right)\right)\\ &= O(n^{-1}+g^{2ks}).\end{aligned}$$

Using the uniform Lipschitz continuity of kernel K and applying Lemma 4 (ii) of the Appendix proves

$$\begin{aligned}& E\left(\sum_{i=1}^{n}\frac{1}{h}(b_i-a_i)(T_i-\sigma_i)\left\{K\left(\frac{t-\sigma_i}{h};t,h\right)-K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}r'(\mu_i)\right)\\ &= O(n^{-1}g^{1/2}h^{-1}+n^{-1/2}g^{ks}h^{-1}+g^{2ks}h^{-1}).\end{aligned}$$

Hence we obtain

$$\begin{aligned}E(S_2) &= \sum_{i=1}^{n}\frac{(F^{-1})'\left(\frac{i}{n+1}\right)}{(n+1)h}(F^{-1})^{(ks)}\left(\frac{i}{n+1}\right)r'(\bar{t}_i)K\left(\frac{t-\bar{t}_i}{h};t,h\right)\frac{g^{ks}}{ks!}\mu_{ks}(K_s(\cdot;\frac{i}{n+1},h))\\ &\quad +O(n^{-1})+o(g^{ks})+O(n^{-1}g^{1/2}h^{-1}+g^{2ks}h^{-1})\\ &= O(n^{-1}+g^{ks})+O(n^{-1}g^{1/2}h^{-1}+g^{2ks}h^{-1})\end{aligned}$$

uniformly in $t\in[0,1]$. In the inner part, e.g. for $t\in[h+g,1-h-g]$ we obtain the uniform approximation

$$\begin{aligned}E(S_2) &= \frac{g^{ks}}{ks!}\mu_{ks}(K_s)\int_0^1\frac{1}{h}K\left(\frac{t-u}{h};t,h\right)r'(u)(F^{-1})^{(ks)}(F(u))\,du\\ &\quad +O(n^{-1})+o(g^{ks})+O(n^{-1}g^{1/2}h^{-1}+g^{2ks}h^{-1}).\end{aligned}$$

With the standard approximation of this integral Proposition 1 is proved. □

Proposition 1 shows that it might be possible to choose the bandwidths h and g in such a way that the leading terms both vanish. This is possible if

$$r^{(k)}(t)r'(t)(F^{-1})^{(ks)}(F(t))<0$$

is satisfied. Then both bandwidths have to be chosen locally and as consistent estimators of the optimal ones. This seems to be very difficult in practise and additionally such an approach would lead to a kernel estimator with very different bias behaviour. The idea for introducing quantile estimators instead was mainly to reduce variance and to obtain an estimator with MSE as in (2.1). Hence it might be more useful to assume $g^{ks} = o(h^k)$ as was already done by Jones et al. (1994). The next proposition deals with the asymptotic expression of the variance $V(t;g,h)$ of the kernel estimator.

Proposition 2: *The kernel estimator $\hat{r}_2^*(t;h,g)$ of equation (2.2) with definitions (2.5), $h = o(1)$, $g = o(1)$, $n^{-1}h^{-2} = o(1)$ and $n^{-1}g^{-2} = o(1)$ satisfies*

$$\begin{aligned} V(t;h,g) &= \frac{\sigma^2(t)M(K)}{f(t)nh}\left\{1 + \frac{4}{ng}M(K_s)\right\} \\ &\quad + O(n^{-2}h^{-3} + n^{-2}gh^{-4} + n^{-1}g + n^{-1}g^{ks}h^{-2} + n^{-1}g^{2ks}h^{-4} + g^{2ks} + g^{4ks}h^{-4}) \end{aligned}$$

uniformly for all $t \in [h+g, 1-h-g]$ and

$$\begin{aligned} V(t;h,g) &= \frac{\sigma^2(t)M(K(\cdot;t,h))}{f(t)nh} + O(n^{-2}g^{-1}h^{-1}) \\ &\quad + O(n^{-2}h^{-3} + n^{-2}gh^{-4} + n^{-1}g + n^{-1}g^{ks}h^{-2} + n^{-1}g^{2ks}h^{-4} + g^{2ks} + g^{4ks}h^{-4}) \end{aligned}$$

uniformly for all $t \in [0,1]$.

Proof of Proposition 2: Here we use the well known decomposition

$$Var(\hat{r}_2^*(t;h,g)) = Var(E(\hat{r}_2^*(t;h,g)|T_1,\ldots,T_n) + E(Var(\hat{r}(t;h,g)|T_1,\ldots,T_n)).$$

The first term the variance of the conditional mean is now proved to be neglectable. It follows from the Cauchy-Schwarz inequality with some suitable constant C

$$\begin{aligned} &Var(E(\hat{r}_2^*(t;h,g)|T_1,\ldots,T_n)) = Var(S_2) \\ &= Var\Big[\sum_{i=1}^n \frac{1}{h}(b_i - a_i)K\left(\frac{t-\sigma_i}{h};t,h\right)(T_i - \sigma_i)r'(\eta_i)\Big] \\ &\le 2E\Big[\sum_{i=1}^n \tfrac{1}{h}(b_i - a_i)K\left(\frac{t-\bar{t}_i}{h};t,h\right)(T_i - \sigma_i)r'(\eta_i)\Big]^2 \\ &\quad +2E\Big[\sum_{i=1}^n \tfrac{1}{h^2}(b_i - a_i)(\sigma_i - \bar{t}_i)K'\left(\frac{t-\tau_i}{h};t,h\right)(T_i - \sigma_i)r'(\eta_i)\Big]^2 \\ &\le C\max_i E\left[(n+1)^2(b_i - a_i)^2(T_i - \sigma_i)^2\left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}^2\{r'(\eta_i)\}^2\right] \\ &\quad +2\max_i E\left[\frac{(n+1)^2}{h^4}(b_i - a_i)^2(\sigma_i - \bar{t}_i)^2(T_i - \sigma_i)^2\left\{K'\left(\frac{t-\tau_i}{h};t,h\right)\right\}^2\{r'(\eta_i)\}^2\right] \\ &= O(n^{-1}g + g^{2ks}) + O(n^{-2}gh^{-4} + n^{-1}g^{2ks}h^{-4} + g^{4ks}h^{-4}). \end{aligned}$$

The last approximation follows directly from Lemma 3 (iv) and Lemma 4 (ii) of the Appendix. Typically, the conditional variance of the kernel estimator will give the leading term of the approximations.

We obtain with suitable values $\sigma_i \in (a_i, b_i)$

$$
\begin{aligned}
&(n+1)h\, Var(\hat{r}(t;h,g)|T_1,\ldots,T_n) \\
&= \sum_{i=1}^{n} \frac{1}{(n+1)h} \sigma^2(T_i)(n+1)^2(b_i-a_i)^2 \left\{K\left(\frac{t-\sigma_i}{h};t,h\right)\right\}^2 \\
&= \sum_{i=1}^{n} \frac{1}{(n+1)h} \sigma^2(\bar{t}_i)(n+1)^2(b_i-a_i)^2 \left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}^2 \\
&\quad -2\sum_{i=1}^{n} \frac{1}{(n+1)h^2} \sigma^2(\bar{t}_i)(n+1)^2(b_i-a_i)^2(\sigma_i-\bar{t}_i)K'\left(\frac{t-\bar{t}_i}{h};t,h\right) K\left(\frac{t-\bar{t}_i}{h};t,h\right) \\
&\quad +\sum_{i=1}^{n} \frac{1}{(n+1)h} (\sigma^2)'(\bar{t}_i)(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i) \left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}^2 + \bar{R}.
\end{aligned}
$$

Because of the uniform Lipschitz continuity of K' and $(\sigma^2)'$ and Lemma 3 (i) to (iii) of the Appendix we obtain the approximation $\bar{R} = O(n^{-1}h^{-2} + n^{-1/2}g^{ks}h^{-1} + g^{2ks}h^{-2})$. Additionally, Lemma 2 (ii) ensures that

$$
\sum_{i=1}^{n} \frac{1}{(n+1)h} (\sigma^2)'(\bar{t}_i)(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i) \left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}^2 = O(n^{-1})
$$

whereas it follow from Lemma 2 (iii) that

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{1}{(n+1)h^2} \sigma^2(\bar{t}_i)(n+1)^2(b_i-a_i)^2(\sigma_i-\bar{t}_i)K'\left(\frac{t-\bar{t}_i}{h};t,h\right) K\left(\frac{t-\bar{t}_i}{h};t,h\right) \\
&= O(n^{-1}h^{-1} + g^{ks}h^{-1}).
\end{aligned}
$$

Therefore Lemma 2(i) yields

$$
\begin{aligned}
&E[Var(\hat{r}(t;h,g)|T_1,\ldots,T_n)] \\
&= \sum_{i=1}^{n} \frac{1}{(n+1)^2h^2} \sigma^2(\bar{t}_i) \left\{(F^{-1})'\left(\frac{i}{n+1}\right)\right\}^2 \left\{K\left(\frac{t-\bar{t}_i}{h};t,h\right)\right\}^2 \\
&\quad +O(n^{-2}g^{-1}h^{-1} + n^{-2}h^{-3} + n^{-1}g^{ks}h^{-2} + n^{-1}g^{2ks}h^{-3}).
\end{aligned}
$$

A standard integral approximation completes the proof of the uniform bound.

The leading term in the inner part can be obtained analogously by application of Lemma 2(i) and we obtain uniformly for $t \in [h+g, 1-h-g]$

$$\begin{aligned} & E[Var(\hat{r}(t;h,g)|T_1,\ldots,T_n)] \\ &= \sum_{i=1}^{n} \frac{1}{(n+1)^2 h^2} \sigma^2(\bar{t}_i) \left\{(F^{-1})'\left(\frac{i}{n+1}\right)\right\}^2 \left\{K\left(\frac{t-\bar{t}_i}{h}; t, h\right)\right\}^2 \left\{1 + \tfrac{4}{(n+1)g} M(K_s)\right\} \\ &\quad + O(n^{-2}h^{-3} + n^{-1}g^{ks}h^{-2} + n^{-1}g^{2ks}h^{-3}). \end{aligned}$$

Hence the Proposition follows from a standard integral approximation. □

If the results of both propositions are combined it follows that MSE has the asymptotic form of equation (2.1) for all $t \in (0,1)$ as long as $g^{ks}h^{-k} \to 0$ and $gh^{-1} \to 0$ and the assumptions of these propositions are satisfied. Hence the asymptotically optimal local bandwidth h_{ASY} is given by

$$h_{\mathrm{ASY}} = \left(\frac{\sigma^2(t)(k!)^2 M(K)}{nf(t)\{r^{(k)}(t)\}^2 2k\mu_k(K)^2}\right)^{1/(2k+1)}$$

as long as $r^{(k)}(t) \neq 0$. In the following we assume that h is of optimal order, i.e. $h \asymp n^{-1/(2k+1)}$. Proposition 1 and Proposition 2 prove for $g^{ks} = o(n^{-k/(2k+1)})$ and $g = o(n^{-1/(2k+1)})$ that at least for $ks = 2$ and for all $t \in (0,1)$ the modified regression estimator satisfies

$$\begin{aligned} MSE(\hat{r}_2^*(t;h)) &= \frac{h^{2k}}{k!}\left\{r^{(k)}(t)\right\}^2 \mu_k(K)^2 + \frac{\sigma^2(t)M(K)}{nhf(t)} + R^*(t;h) \\ & h^k g^{ks} r'(t) r^{(k)}(t)(F^{-1})^{(ks)}(F(t)) \frac{\mu_k(K)\mu_{ks}(K_s)}{k!ks!} + \frac{\sigma^2(t)4M(K)M(K_s)}{n^2hgf(t)} + o(h^k g^{ks} + n^{-2}g^{-1}h^{-1}). \end{aligned}$$

with a rest term $R^*(t;h) = o(h^{2k})$ which does not depend on g. For $ks = 4$ the bounds of the propositions are not sufficient for such a statement. But if the equation above holds a bandwidth g which minimizes the influence on MSE asymptotically is given by

$$g_{\mathrm{ASY}}(h) = \left(-\frac{\sigma^2(t)4M(K)M(K_s)k!ks!}{n^2h^{k+1}f(t)r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t))\mu(K)\mu_{ks}(K_s)}\right)^{1/(ks+1)}$$

if $r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t)) < 0$ and

$$g_{\mathrm{ASY}}(h) = \left(\frac{\sigma^2(t)4M(K)M(K_s)k!(ks-1)!}{n^2h^{k+1}f(t)r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t))\mu_k(K)\mu_{ks}(K_s)}\right)^{1/(ks+1)}$$

if $r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t)) > 0$. Especially we have

$$g_{\mathrm{ASY}}(h_{\mathrm{ASY}}) = \left(-\frac{\{\sigma^2(t)\}^k r^{(k)}(t)}{n^{3k+1}\{f(t)\}^k\{r'(t)(F^{-1})^{(ks)}(F(t))\}^{2k+1}}\right)^{1/(ks+1)(2k+1)} C_1(K, K_s)$$

if $r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t)) < 0$ and

$$g_{\mathrm{ASY}}(h_{\mathrm{ASY}}) = \left(\frac{\{\sigma^2(t)\}^k r^{(k)}(t)}{n^{3k+1}\{f(t)\}^k\{r'(t)(F^{-1})^{(ks)}(F(t))\}^{2k+1}}\right)^{1/(ks+1)(2k+1)} C_2(K, K_s)$$

if $r'(t)r^{(k)}(t)(F^{-1})^{(ks)}(F(t)) > 0$ for some kernel constants C_1, C_2. The following corollary summarizes the asymptotic results on MSE.

Corollary: *The kernel estimator $\hat{r}_2^*(t; h, g)$ of equation (2.2) with definitions (2.5), $h \asymp n^{-1/(2k+1)}$ and $g \asymp n^{-(3k+1)/(2k+1)(ks+1)}$ satisfies*

$$MSE(\hat{r}_2^*(t; h)) = \frac{h^{2k}}{k!}\left\{r^{(k)}(t)\right\}^2 \mu_k(K) + \frac{\sigma^2(t)M(K)}{nhf(t)} + R^*(t; h) + O(n^{-\alpha}),$$

with $\alpha = \min\{4kks + ks + k, 2kks + ks + 5k + 2\}/((2k+1)(ks+1)) > 1$.
Additionally, for any continuous weight function w on $[0, 1]$ we obtain

$$\begin{aligned} MISE(\hat{r}_2^*(\cdot; h)) &= \int_0^1 w(t)MSE(\hat{r}_2^*(t; h))\, dt \\ &= \frac{h^{2k}}{k!}\int\left\{r^{(k)}(t)\mu_k(K(\cdot; t, h)\right\}^2 dt + \int\frac{\sigma^2(t)M(K(\cdot; t, h))}{f(t)}dt\frac{1}{nh} + R^{**}(h) + O(n^{-\alpha}), \end{aligned}$$

with α as above and a rest term $R^{**}(h) = o(h^{2k})$ which does not depend on g.

REFERENCES

Arnold, B.C., Balakrishnan, N. & Nagaraja, H.N. (1992). *A first course in order statistics.* Wiley, New York.

Chu, C.K. & Marron, J.S. (1991). Choosing a kernel regression estimator. *Statist. Sci.* **6,** 404–436.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74,** 829–836.

David, H.A. (1981). *Order Statistics (2nd ed.)* Wiley, New York.

Engel, J. & Kneip, A. (1994). A remedy for kernel regression under random design. Discussion Paper No A-431, SFB303, Universität Bonn.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *Ann. Statist.* **12,** 261–168.

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *Ann. Statist.* **13,** 428–433.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21,** 196–216.

Gasser, T. & Engel, J (1990). The choice of weights in kernel regression estimation. *Biometrika.* **77,** 377–281.

Gasser, T. & Müller, H.G. (1979). Kernel estimation of regression functions. *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757,** 23–68. Springer, New York.

Hall, P. & Turlach, B.A. (1995). Convolution and interpolation: competitors with local polynomial smoothing. Statistics Research Report SRR037-95 Australian National University, Canberra.

Hastie, T.J. & Loader, C. (1992). Local regression: automatic kernel carpentry. *Statist. Sci.* **8,** 120–143.

Jones, M.C., Davies, S.J. & Park, B.U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.* **89,** 825–832.

Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.* **9,** 141–142.

Priestley, M.B. & Chao, M.T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc., Ser. B.* **34,** 385–392.

Sheather, S.J. & Marron, J.S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.* **85,** 410–416.

Seifert, B. & Gasser, T. (1995). Finite sample analysis of local polynomials: analysis and solutions. *J. Amer. Statist. Assoc.* to appear.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5,** 595–645.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya Ser. A.* **26,** 359–372.

Xiang, X. (1994). Bahadur representation of kernel quantile estimators. *Scand. J. Statist.* **21,** 169–178.

A Appendix

Here in the Appendix some important results on order statistics and kernel quantile estimators are summarized.

A.1 On order statistics

Let $T_1, \ldots, T_{(n)}$ be the ordered sample of n independent and identically distributed continuous random variables with joint cumulative distribution function F. We assume that F is strictly increasing and that F^{-1} is at least four times continuously differentiable with bounded derivatives on $(0,1)$. Then it is well known that $T_1, \ldots, T_n$ have the same distribution as $F^{-1}(U_{(1)}), \ldots, F^{-1}(U_{(n)})$ where $U_{(1)}, \ldots, U_{(n)}$ are the order statistics from n iid. Uniform(0,1) random variables. Asymptotic properties can be derived from Taylor approximations using expressions of the central moments of the uniform order statistics, compare e.g. David (1980, section 4.6) and Arnold, Balakrishman and Nagaraja (1986, section 5.5). We obtain by Taylor expansion

$$T_i = F^{-1}\left(\tfrac{i}{n+1}\right) + \sum_{k=1}^{3} \frac{\left(U_{(i)} - \frac{i}{n+1}\right)^k}{(k!)} (F^{-1})^{(k)}\left(\tfrac{i}{n+1}\right) + \frac{\left(U_{(i)} - \frac{i}{n+1}\right)^{(4)}}{4!} (F^{-1})^{(4)}(\tau_i)$$

for some suitable $\tau_i \in (0,1)$ and all $i = 1, \ldots, n$.

A general formula of moments of the uniform order statistics $U_{(i)}$ is given e.g. in David (1980, section 3.1). It can be used to prove directly

$$E\left(U_{(i)} - \tfrac{i}{n+1}\right)^{2k} \leq \frac{C_{2k}}{(n+1)^k},\ E\left(U_{(i)} - \tfrac{i}{n+1}\right)^{2k-1} \leq \frac{C_{2k-1}}{(n+1)^k}$$

for all $i = 1, \ldots, n$ and at least for $k \leq 12$. Additionally, we have

$$E\left(U_{(i)} - \tfrac{i}{n+1}\right)\left(U_{(j)} - \tfrac{j}{n+1}\right) = \frac{\min\{i,j\}(n+1-\max\{i,j\})}{(n+1)^2(n+2)},$$

for all $i, j = 1, \ldots, n$, compare e.g. David (1980, section 3.1).

Denote $\bar{t}_i = F^{-1}\left(\frac{i}{n+1}\right)$ then it follows from the equations given above for all $i, j, k = 1, \ldots, n$ and a suitable constant C:

$$E(T_i - \bar{t}_i) \leq \frac{C}{n+1} \tag{1}$$

$$
\begin{aligned}
&E(T_i-\bar t_i)(T_j-\bar t_j)\\
&= \frac{\min\{i,j\}(n+1-\max\{i,j\})}{(n+1)^3}(F^{-1})'\left(\frac{i}{n+1}\right)(F^{-1})'\left(\frac{j}{n+1}\right)+O(n^{-2})\\
&\le \frac{C}{n+1}, \qquad (2)
\end{aligned}
$$

and additionally

$$E|T_i-\bar t_i|^\ell \le \frac{C}{(n+1)^{\ell/2}} \qquad (3)$$

at least for $\ell \le 12$. Further using the Lipschitz-continuity of $(F^{-1})'$ we can conclude

$$
\begin{aligned}
&|E\left(T_i-\bar t_i-T_j+\bar t_j\right)|\\
&= \left|\frac{i(n+1-i)}{(n+1)^3}(F^{-1})'\left(\tfrac{i}{n+1}\right)-\frac{j(n+1-j)}{(n+1)^3}(F^{-1})'\left(\tfrac{j}{n+1}\right)\right|+O(n^{-2})\\
&\le \tilde C\frac{|i-j|}{(n+1)^2} \qquad (4)
\end{aligned}
$$

and similarily

$$|E(T_k-\bar t_k)(T_i-\bar t_i-T_j+\bar t_j)| \le \tilde C\frac{|i-j|}{(n+1)^2} \qquad (5)$$

and at least for $\ell \le 4$

$$E(T_k-\bar t_k)^{2\ell}(T_i-\bar t_i-T_j+\bar t_j)^2 \le \tilde C\frac{|i-j|}{(n+1)^{\ell+2}} \qquad (6)$$

for some suitable constant $\tilde C$.

A.2 On kernel quantile estimators

Several forms of kernel quantile estimators and related quantile estimators and some aspects of bandwidth choice are described e.g. in Sheather and Marron (1990). The asymptotic deficiency and distribution was calculated by Falk (1984)and Falk (1985) respectively and a Bahadur representation was derived by Xiang (1994).

Here we need further approximations since we are mainly interested in moments on differences of these quantile estimators and also second order terms. In the following we state and proof some bounds for such moments. For a fixed bandwidth g we define a kernel estimator of the p-th quantile, $p \in [\frac{0.5}{n+1}, \frac{n+0.5}{n+1}]$ by

$$\hat q(p) = \sum_{j=1}^n w_j(p,g)\,T_j$$

with the abbreviation

$$w_j(p,g) = \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \tfrac{1}{g} K_s\left(\frac{p-u}{g}; p, g\right)\, du.$$

The kernel K_s used in the inner part of $[0,1]$ is assumed to have support$[-1,1]$, to be continuously differentiable on $\mathbb{R}$ and of order $ks = 2$ or $ks = 4$. In the boundary region we use boundary kernels of the same order. We suppose that these boundary kernels and their derivatives are Lipschitz-continuous functions of p/g in the left boundary region and of $(1-p)/g$ in the right boundary region. Besides we assume that their support lies in $[-2, p/g]$ for the left boundary and in $[-p/g, 2]$ for the right boundary region.

The kernel K used for kernel regression estimation in the inner part of $[0,1]$ is also assumed to have support$[-1,1]$, to be continuously differentiable on $\mathbb{R}$ and of order $k \geq 2$. In the boundary region we also use boundary kernels of the same order. Again we suppose that these boundary kernels and their derivatives are Lipschitz-continuous functions of p/g in the left boundary region and of $(1-p)/g$ in the right boundary region. Besides we assume that their support lies in $[-2, p/g]$ for the left boundary and in $[-p/g, 2]$ for the right boundary region. Besides we suppose that $\mu_k(K) > 0$ and $\mu_{ks}(K_s) > 0$.

For simplification of the notation we do not only assume that $g = o(1)$ and $n^{-1}g^{-1} = o(1)$ but also that $n^{-1}g^{-2} = o(1)$. We obtain by Taylor expansions of the bias terms

$$\begin{aligned} \hat{q}(p) - \sum_{j=1}^{n} w_j(p,g)(T_j - \bar{t}_j) &= \sum_{j=1}^{n} w_j(p,g)\bar{t}_j \\ &= \frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s(\cdot; p, g)\right)\left(F^{-1}\right)^{(ks)}(p) + F^{-1}(p) + O(n^{-1}) + o(g^{ks}). \end{aligned}$$

For $i = 1, \ldots, n$ this gives an uniform approximation

$$\begin{aligned} \hat{q}(p) - \bar{t}_i &= \sum_{j=1}^{n} w_j(\frac{i}{n+1}, g)(T_j - \bar{t}_j) + \frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s\left(\cdot; \frac{i}{n+1}, g\right)\right)\left(F^{-1}\right)^{(ks)}(\frac{i}{n+1}) \\ &\quad + O(n^{-1}) + o(g^{ks}) \end{aligned} \tag{7}$$

for all $p \in [\frac{i-0.5}{n+1}, \frac{i+0.5}{n+1}]$. One should note that for random p with values in $[\frac{i-0.5}{n+1}, \frac{i+0.5}{n+1}]$ the error terms bounded by $O(n^{-1}) + o(g^{ks})$ may be random terms.

Now we define $a_i = \hat{q}\left(\frac{i-0.5}{n+1}\right)$ and $b_i = \hat{q}\left(\frac{i+0.5}{n+1}\right)$ for $i = 1, \ldots, n$ and use the abbreviation

$$\bar{w}_j(i,g) = (n+1)\left\{w_j\left(\frac{i+0.5}{n+1}, g\right) - w_j\left(\frac{i-0.5}{n+1}, g\right)\right\}$$

for $j = 1, \ldots, n$. Hence we can write $(n+1)(b_i - a_i) = \sum \bar{w}_j(i,g) T_j$. Moments of this expression will be analyzed in the following lemmas.

Lemma 1: *With quantile estimators $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ defined as above we obtain*

(i) $E(n+1)(b_i - a_i) = (F^{-1})'(\frac{i}{n+1}) + O(n^{-1} + g^{ks-1})$

(ii) $E(n+1)(b_i - a_i)(T_i - \hat{q}(p_i)) =$
$-(F^{-1})'(\frac{i}{n+1})(F^{-1})^{(ks)}(\frac{i}{n+1})\frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s\left(\cdot; \frac{i}{n+1}, g\right)\right) + O(n^{-1}) + o(g^{ks})$ *for a random variable p_i with values in $[(i-0.5)/(n+1), (i+0.5)/(n+1)]$.*

Proof of Lemma 1:

(i) Firstly, it follows from Lipschitz-continuity of the kernel function

$$|\bar{w}_j(i,g)| \leq \frac{C}{ng^2} \quad \text{and} \quad \sum_{j=1}^{n} |\bar{w}_j(i,g)| \leq \frac{C}{ng} \tag{8}$$

for some suitable constant C.

Now we are analyzing $\sum_{j=1}^{n} \bar{w}_j(i,g) q(\frac{j}{n+1})$ for an arbitrary ks-times continuously differentiable function q on $[0,1]$. Since $K_s'(\cdot, p, g)$ is Lipschitz-continuous in p/g we obtain

$$\begin{aligned}
&\bar{w}_j(i,g) q(\frac{j}{n+1}) \\
&- \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \frac{n+1}{g} \left\{ K_s\left(\frac{\frac{i+0.5}{n+1} - u}{g}; \frac{i+0.5}{n+1}, g\right) - K_s\left(\frac{\frac{i-0.5}{n+1} - u}{g}; \frac{i-0.5}{n+1}, g\right) \right\} q(u)\, du \\
&= \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \frac{n+1}{g} \left\{ K_s\left(\frac{\frac{i+0.5}{n+1} - u}{g}; \frac{i+0.5}{n+1}, g\right)\right. \\
&\quad \left. - K_s\left(\frac{\frac{i-0.5}{n+1} - u}{g}; \frac{i-0.5}{n+1}, g\right) \right\} \left(\frac{j}{n+1} - u\right) q'(\frac{j}{n+1})\, du + O(n^{-3}g^{-2}) \\
&= \int_{\frac{j-0.5}{n+1}}^{\frac{j+0.5}{n+1}} \frac{n+1}{g} \left\{ K_s\left(\frac{\frac{i+0.5}{n+1} - \frac{j}{n+1}}{g}; \frac{i+0.5}{(n+1)}, g\right)\right. \\
&\quad \left. - K_s\left(\frac{\frac{i-0.5}{n+1} - \frac{j}{n+1}}{g}; \frac{i-0.5}{(n+1)}, g\right) \right\} (\frac{j}{n+1} - u) q'(\frac{j}{n+1})\, du + O(n^{-3}g^{-3}) \\
&= O(n^{-3}g^{-3}).
\end{aligned}$$

This leads to the following approximation

$$\sum_{j=1}^{n} \bar{w}_j(i,g) q(\frac{j}{n+1})$$

$$
\begin{aligned}
&= \int_0^1 \frac{n+1}{g}\left\{K_s\left(\frac{\frac{i+0.5}{n+1}-u}{g};\frac{i+0.5}{n+1},g\right)-K_s\left(\frac{\frac{i-0.5}{n+1}-u}{g};\frac{i-0.5}{n+1},g\right)\right\}q(u)\,du \\
&\quad +O(n^{-2}g^{-2}) \\
&= (n+1)\{q(\frac{i+0.5}{n+1})-q(\frac{i-0.5}{n+1})\}+O(n^{-2}g^{-2}) \\
&\quad +\int_{-1}^1\int_0^1 \frac{(gx)^{ks}}{(ks-1)!}\delta^{ks}\left\{q^{(ks)}(\frac{i+0.5}{n+1}-gx(1-\delta))K_s\left(x;\frac{i+0.5}{n+1},g\right)\right. \\
&\quad \left.-q^{(ks)}(\frac{i-0.5}{n+1}-gx(1-\delta))K_s\left(x;\frac{i-0.5}{n+1},g\right)\right\}d\delta\,dx \\
&= q'(\frac{i+0.5}{n+1})+O(n^{-1})+\begin{cases} O(g^{ks}) & \text{, for } \frac{i-1}{n+1}\geq g \text{ and } \frac{n-i}{n+1}\geq g \\ O(g^{ks-1}) & \text{, else.}\end{cases}
\end{aligned}
\tag{9}
$$

Hence we have especially the following approximation of bias terms of $(n+1)(b_i-a_i)$

$$
\begin{aligned}
(n+1)(b_i-a_i) &= \sum_{j=1}^n \bar{w}_j(i,g)F^{-1}\left(\frac{j}{n+1}\right)+\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j) \\
&= (F^{-1})'(\frac{i}{n+1})+\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)+O(n^{-1}+g^{ks-1}).
\end{aligned}
\tag{10}
$$

Now the first part of Lemma 1 follows immediately from (1), (9) and (10).

(ii) Using (7), (8) and (10) we obtain

$$
\begin{aligned}
&E(n+1)(b_i-a_i)(t_i-\hat{q}(p_i)) \\
&= E\Big(\Big\{(F^{-1})'\left(\frac{i}{n+1}\right)+\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)+O(g^{ks-1}+n^{-1}g^{-1})\Big\} \\
&\quad \Big\{-\frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s(\cdot,\frac{i}{n+1},g)\right)(F^{-1})^{(ks)}\left(\frac{i}{n+1}\right) \\
&\quad +\sum_{j=1}^n w_j(\frac{i}{n+1},g)(T_i-\bar{t}_i-T_j+\bar{t}_j)+O(n^{-1})+o(g^{ks})\Big\}\Big) \\
&= \frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s(\cdot,\frac{i}{n+1},g)\right)(F^{-1})'\left(\frac{i}{n+1}\right)(F^{-1})^{(ks)}\left(\frac{i}{n+1}\right)+O(n^{-1})+o(g^{ks}) \\
&\quad +O(1)E\sum_{j=1}^n w_j(\frac{i}{n+1},g)(T_i-\bar{t}_i-T_j+\bar{t}_j) \\
&\quad +E\sum_{j=1}^n\sum_{k=1}^n w_j(\frac{i}{n+1},g)\bar{w}_k(i,g)(T_i-\bar{t}_i-T_j+\bar{t}_j)(T_k-\bar{t}_k) \\
&\quad +O(n^{-1}+g^{ks})E\Big|\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)\Big|.
\end{aligned}
$$

Hence application of (1), (4) and (6) and of $E|\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)|=o(1)$ gives

$$
\begin{aligned}
&E(n+1)(b_i-a_i)(T_i-\hat{q}(p_i))\\
&= \frac{g^{ks}}{(ks)!}\mu_{ks}\left(K_s(\cdot,\frac{i}{n+1},g)\right)(F^{-1})'\left(\frac{i}{n+1}\right)(F^{-1})^{(ks)}\left(\frac{i}{n+1}\right)+O(n^{-1})+o(g^{ks}).
\end{aligned}
$$

This proves Lemma 1. $\square$

Lemma 2: *The quantile estimators $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ defined as above satisfy uniformly for all $i=1,\ldots,n$*

(i) $E(n+1)^2(b_i-a_i)^2=\left\{(F^{-1})'\left(\frac{i}{n+1}\right)\right\}^2+O(n^{-1}g^{-1})$
and $E(n+1)^2(b_i-a_i)^2=\left\{(F^{-1})'\left(\frac{i}{n+1}\right)\right\}^2\left\{1+\frac{4}{ng}M(K_s)\right\}+O(n^{-1})$ *for all i of the inner part, e.g. those who satisfy* $\frac{i-1}{n+1}\geq g$ *and* $\frac{n-i}{n+1}\geq g$.

(ii) $E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)=O(n^{-1})$.

(iii) $E(n+1)^2(b_i-a_i)^2(\hat{q}(p_i)-\bar{t}_i)=O(n^{-1}+g^{ks})$.
with a random variable p_i with values in $[(i-0.5)/(n+1),(i+0.5)/(n+1)]$.

Proof of Lemma 2:

(i) As described in Appendix A.1 we can expand $(n+1)^2(b_i-a_i)^2$ with use of an suitable midpoint τ_j by

$$
\begin{aligned}
&(n+1)^2(b_i-a_i)^2\\
&= \Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)\Big[\sum_{\ell=0}^{3}\tfrac{1}{\ell!}\left(U_{(j)}-\tfrac{j}{n+1}\right)^{\ell}(F^{-1})^{(\ell)}\Big(\frac{j}{n+1}\Big)+\tfrac{1}{24}\left(U_{(j)}-\tfrac{j}{n+1}\right)^{4}(F^{-1})^{(4)}(\tau_j)\Big]\Big\}^2.
\end{aligned}
$$

Using the bounds on the central moments of the uniform order statistics gives

$$
\begin{aligned}
&E(n+1)^2(b_i-a_i)^2\\
&= E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)\Big[\sum_{\ell=0}^{3}\tfrac{1}{\ell!}\left(U_{(j)}-\tfrac{j}{n+1}\right)^{\ell}(F^{-1})^{(\ell)}\Big(\frac{j}{n+1}\Big)+\tfrac{1}{24}\left(U_{(j)}-\tfrac{j}{n+1}\right)^{4}(F^{-1})^{(4)}(\tau_j)\Big]\Big\}^2\\
&= \Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(F^{-1})\Big(\frac{j}{n+1}\Big)\Big\}^2+O(n^{-2}g^{-2})
\end{aligned}
$$

$$+\frac{1}{n+1}\Big\{\sum_{j=1}^{n}\sum_{k=1}^{n}\bar{w}_j(i,g)\bar{w}_k(i,g)(F^{-1})\Big(\frac{j}{n+1}\Big)(F^{-1})''\Big(\frac{k}{n+1}\Big)\Big(\frac{k}{n+1}-\frac{k^2}{(n+1)^2}\Big)\Big\}$$

$$+\frac{2}{n+1}\Big\{\sum_{j=1}^{n}\sum_{k=1}^{n}\bar{w}_j(i,g)\bar{w}_k(i,g)(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right)\Big(\frac{\min\{j,k\}}{n+1}-\frac{jk}{(n+1)^2}\Big)\Big\}$$

$$= \Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(F^{-1})\Big(\frac{j}{n+1}\Big)\Big\}^2+O(n^{-1})$$

$$+\frac{1}{n+1}\sum_{j=1}^{n}\sum_{k=1}^{n}\bar{w}_j(i,g)\bar{w}_k(i,g)(F^{-1})\Big(\frac{j}{n+1}\Big)(F^{-1})''\Big(\frac{k}{n+1}\Big)\Big(\frac{k}{n+1}-\frac{k^2}{(n+1)^2}\Big)$$

$$+\frac{2}{n+1}\sum_{j=1}^{n}\sum_{k=1}^{n}\bar{w}_j(i,g)\bar{w}_k(i,g)(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right)\Big(\frac{\min\{j,k\}}{n+1}-\frac{j}{n+1}\frac{k}{n+1}\Big).$$

Hence application of equation (9) leads to

$$E(n+1)^2(b_i-a_i)^2 = \Big\{(F^{-1})'\left(\frac{i}{n+1}\right)\Big\}^2+O(n^{-1})$$
$$+\frac{4}{n+1}\sum_{j=1}^{n}\sum_{k\le j}\bar{w}_j(i,g)\bar{w}_k(i,g)(F^{-1})'(\frac{j}{n+1})(F^{-1})'(\frac{k}{n+1})\frac{k}{n+1}.$$

Note that analogously to equation (9) we obtain for arbitrary continuously differentiable functions q_1, q_2 the equation

$$\frac{2}{n+1}\sum_{j=1}^{n}\sum_{k\le j}\bar{w}_j(i,g)\bar{w}_k(i,g)(q_1(\frac{j}{n+1})q_2(\frac{k}{n+1})+q_1(\frac{k}{n+1})q_2(\frac{j}{n+1}))=O(n^{-1})$$

and also

$$\frac{2}{n+1}\sum_{j=1}^{n}\sum_{k\le j}\bar{w}_j(i,g)\bar{w}_k(i,g)q_1(\frac{j}{n+1})q_1(\frac{k}{n+1})=O(n^{-1}).$$

Therefore Taylor expansion of $(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right)\frac{k}{n+1}$ gives

$$E(n+1)^2(b_i-a_i)^2=\Big\{(F^{-1})'\left(\frac{i}{n+1}\right)\Big\}^2+O(n^{-1})$$
$$+\frac{4}{n+1}\sum_{j=1}^{n}\sum_{k\le j}\bar{w}_j(i,g)\bar{w}_k(i,g)\Big\{\Big[(F^{-1})'\left(\frac{i}{n+1}\right)\Big]^2\frac{i}{n+1}$$
$$+\frac{j-i}{n+1}(F^{-1})''\Big(\frac{i}{n+1}\Big)(F^{-1})'\left(\frac{i}{n+1}\right)\frac{i}{n+1}+\frac{k-i}{n+1}(F^{-1})''\Big(\frac{i}{n+1}\Big)$$
$$\times(F^{-1})'\left(\frac{i}{n+1}\right)\frac{i}{n+1}+\frac{k-i}{n+1}\Big[(F^{-1})'\left(\frac{i}{n+1}\right)\Big]^2\Big\}$$
$$= \Big\{(F^{-1})'\left(\frac{i}{n+1}\right)\Big\}^2\Big\{1+\frac{4}{n+1}\sum_{j=1}^{n}\sum_{k\le j}\bar{w}_j(i,g)\bar{w}_k(i,g)\frac{k}{n+1}\Big\}+O(n^{-1}).$$

This proves the uniform bound of part (i). In the inner part, i.e. for $\frac{i-1}{n+1} \geq g$ and $\frac{n-i}{n+1} \geq g$, integration by parts yields

$$
\begin{aligned}
&E(n+1)^2(b_i-a_i)^2 \\
&= \Big\{(F^{-1})'\Big(\frac{i}{n+1}\Big)\Big\}^2\Big\{1-\frac{4}{n+1}\int_0^1\int_0^u \frac{1}{g^4}K_s'\Big(\frac{\frac{i}{n+1}-u}{g}\Big)K_s'\Big(\frac{\frac{i}{n+1}-v}{g}\Big)\,dv\,du\Big\}+O(n^{-1}) \\
&= \Big\{(F^{-1})'\Big(\frac{i}{n+1}\Big)\Big\}^2\Big\{1+\frac{4}{n+1}\int_0^1 \frac{1}{g^2}\Big[K_s\Big(\frac{\frac{i}{n+1}-u}{g}\Big)\Big]^2\,du\Big\}+O(n^{-1}) \\
&= \Big\{(F^{-1})'\Big(\frac{i}{n+1}\Big)\Big\}^2\Big\{1+\frac{4}{ng}M(K_s)\Big\}+O(n^{-1}).
\end{aligned}
$$

This completes the proof of part (i).

(ii) This part can be proved similar to Lemma 1 (ii). Expansion of $(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)$ by equation (10) and the bounds of equation (1) and (3) yield

$$
\begin{aligned}
&|E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)| \\
&= \Big|E\Big\{\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)\Big| \\
&\leq \Big|E\Big\{\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)\bar{w}_k(i,g)(T_j-\bar{t}_j)(T_k-\bar{t}_k)(T_i-\bar{t}_i)\Big| \\
&\quad +O(1)\Big|E\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)(T_i-\bar{t}_i)\Big|+O(1)\Big|E(T_i-\bar{t}_i)\Big| \\
&= O(n^{-1})+O(1)\Big|\sum_{j=1}^n \bar{w}_j(i,g)\frac{\min\{i,j\}(n+1-\max\{i,j\})}{(n+1)^3}(F^{-1})'\Big(\frac{i}{n+1}\Big)(F^{-1})'\Big(\frac{j}{n+1}\Big)\Big|.
\end{aligned}
$$

Hence application of equations (8) and (9) gives

$$
\begin{aligned}
&|E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)| \\
&\leq O(n^{-1})+O(1)\Big|\sum_{j=1}^n \bar{w}_j(i,g)\frac{i(n+1)-ij}{(n+1)^3}(F^{-1})'\Big(\frac{i}{n+1}\Big)(F^{-1})'\Big(\frac{j}{n+1}\Big)\Big| \\
&\quad +O(1)\Big|\sum_{j=1}^{i-1} \bar{w}_j(i,g)\frac{j-i}{(n+1)^2}(F^{-1})'\Big(\frac{i}{n+1}\Big)(F^{-1})'\Big(\frac{j}{n+1}\Big)\Big| \\
&= O(n^{-1}).
\end{aligned}
$$

(iii) Applying the approximations of equation (8) and (10) we obtain

$$|E(n+1)^2(b_i-a_i)^2(\hat{q}(p_i)-\bar{t}_i)|$$

$$
\begin{aligned}
&= \Big|E\Big\{\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^n w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)+O(n^{-1}+g^{ks})\Big\}\Big| \\
&\le \Big|E\sum_{j=1}^n\sum_{k=1}^n\sum_{\ell=1}^n \bar{w}_j(i,g)\bar{w}_k(i,g)w_\ell\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)(T_k-\bar{t}_k)(T_\ell-\bar{t}_\ell)\Big|+O(n^{-1}+g^{ks}) \\
&\quad +O(1)\Big|E\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)(T_k-\bar{t}_k)\Big| \\
&\quad +O(n^{-1}+g^{ks})E\Big\{\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)\Big\}^2 \\
&\quad +O(n^{-1}+g^{ks})E\Big|\sum_{j=1}^n \bar{w}_j(i,g)(T_j-\bar{t}_j)\Big|+O(1)\Big|E\sum_{j=1}^n w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)\Big| \\
&= O(n^{-1}+g^{ks})+O(1)\Big|E\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)(T_k-\bar{t}_k)\Big|.
\end{aligned}
$$

Hence it remains to show that $E\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)(T_k-\bar{t}_k)=O(n^{-1})$. Using the asymptotic expression of equation (2) we obtain

$$
\begin{aligned}
&E\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)(T_k-\bar{t}_k) \\
&= \frac{1}{n+1}\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)\Big(\frac{\min\{j,k\}}{n+1}-\frac{j}{n+1}\frac{k}{n+1}\Big)(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right) \\
&= \frac{1}{n+1}\sum_{j=1}^n\sum_{k=1}^n \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)\Big(\frac{j}{n+1}-\frac{j}{n+1}\frac{k}{n+1}\Big)(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right) \\
&\quad +\frac{1}{n+1}\sum_{j=1}^n\sum_{k\le j} \bar{w}_j(i,g)w_k\Big(\frac{i}{n+1},g\Big)\frac{k-j}{n+1}(F^{-1})'\left(\frac{j}{n+1}\right)(F^{-1})'\left(\frac{k}{n+1}\right)=O(n^{-1})
\end{aligned}
$$

This completes the proof of Lemma 2. $\square$

Lemma 3: *For the quantile estimators $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ defined as above and a random variable p_i with values in $[(i-0.5)/(n+1),(i+0.5)/(n+1)]$ we have*

(i) $E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)^2=O(n^{-1})$.

(ii) $E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)(\hat{q}(p_i)-\bar{t}_i)=O(n^{-1}+n^{-1/2}g^{ks})$.

(iii) $E(n+1)^2(b_i-a_i)^2(\hat{q}(p_i)-\bar{t}_i)^2=O(n^{-1}+g^{2ks})$.

(iv) $E(n+1)^2(b_i-a_i)^2(T_i-\hat{q}(p_i))^2 = O(n^{-1}g+g^{2ks})$.

Proof of Lemma 3:

(i) Applying equation (10) we can write

$$\begin{aligned} & E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)^2 \\ = \quad & E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)^2. \end{aligned}$$

With the same arguments as in the proof of Lemma 2 (ii) and using the asymptotic bounds of equation (3) we obtain

$$E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)^2 = O(n^{-2}g^{-2}+n^{-3/2}g^{-1}+n^{-1}) = O(n^{-1}).$$

(ii) As in (i) using equations (3), (7) and (10) we obtain

$$\begin{aligned} & E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)(\hat{q}(p_i)-\bar{t}_i) \\ = \quad & E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)+O(g^{ks}+n^{-1})\Big\} \\ = \quad & O(n^{-1}+n^{-1/2}g^{ks}). \end{aligned}$$

(iii) Using equations (3), (7) and (10) we otain in this situation

$$\begin{aligned} & E(n+1)^2(b_i-a_i)^2(\hat{q}(p_i)-\bar{t}_i)^2 \\ = \quad & E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)+O(g^{ks}+n^{-1})\Big\}^2 \\ = \quad & E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)\Big\}^2 \\ & +O(g^{ks}+n^{-1})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big|\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)\Big| \\ & +O(g^{2ks}+n^{-2})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2 \\ = \quad & O(n^{-1}+g^{2ks}). \end{aligned}$$

(iv) Expanding the products as before and using equations (3), (7) and (10) we obtain

$$E(n+1)^2(b_i-a_i)^2(T_i-\hat{q}(p_i))^2$$

$$
\begin{aligned}
&= E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)+O(g^{ks}+n^{-1})\Big\}^2\\
&= E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)\Big\}^2\\
&\quad +O(g^{ks}+n^{-1})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big|\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)\Big|\\
&\quad +O(g^{2ks}+n^{-2})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\\
&= O(n^{-1}g+g^{2ks}).
\end{aligned}
$$

Thereby the last equation follows from the bounds of equation (4) and (6). □

Lemma 4: *For the quantile estimators $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ defined as above and a random variable p_i with values in $[(i-0.5)/(n+1),(i+0.5)/(n+1)]$ we have*

(i) $E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)^2(T_i-\hat{q}(p_i))^2=O(n^{-2}g+n^{-1}g^{2ks})$.

(ii) $E(n+1)^2(b_i-a_i)^2(T_i-\hat{q}(p_i))^2(\hat{q}(p_i)-\bar{t}_i)^2=O(n^{-2}g+n^{-1}g^{2ks}+g^{4ks})$.

Proof of Lemma 4:

(i) With similar arguments as above we obtain

$$
\begin{aligned}
&E(n+1)^2(b_i-a_i)^2(T_i-\bar{t}_i)^2(T_i-\hat{q}(p_i))^2\\
&= E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)+O(g^{ks}+n^{-1})\Big\}^2\\
&= E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)\Big\}^2\\
&\quad +O(g^{ks}+n^{-1})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)^2\Big|\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)\Big|\\
&\quad +O(g^{2ks}+n^{-2})E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2(T_i-\bar{t}_i)^2\\
&= O(n^{-2}g+n^{-1}g^{2ks}).
\end{aligned}
$$

(ii) Besides the application of equation (3), (7) and (10) we use the bounds of (6) and obtain

$$E(n+1)^2(b_i-a_i)^2(T_i-\hat{q}(p_i))^2(\hat{q}(p_i)-\bar{t}_i)^2$$

$$
\begin{aligned}
= \quad & E\Big\{\sum_{j=1}^{n}\bar{w}_j(i,g)(T_j-\bar{t}_j)+O(1)\Big\}^2\Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_i-\bar{t}_i-T_j+\bar{t}_j)+O(g^{ks}+n^{-1})\Big\}^2 \\
& \Big\{\sum_{j=1}^{n}w_j\Big(\frac{i}{n+1},g\Big)(T_j-\bar{t}_j)+O(g^{ks}+n^{-1})\Big\}^2 \\
= \quad & O(n^{-2}g+n^{-1}g^{2ks}+g^{4ks}).
\end{aligned}
$$

$\square$

Eva Herrmann
Fachbereich Mathematik
Technische Hochschule Darmstadt
Schloßgartenstraße7
D–64289 Darmstadt, F.R.G.