

Local polynomial expectile regression

Supplementary Material

Adam, C. · Gijbels, I.

Received: date / Revised: date

Abstract This Supplementary Material part contains the following additional material. In Section S1 a simple illustration of the relationship between quantiles and expectiles is given. The proof of Proposition 1 is given in Section S2. Further simulation results are summarized and discussed in Section S3. Additional real data applications can be found in Section S4. In Section S5 some additional explanations about the ROT-bandwidth selection in case of location-scale models are provided. Finally, Section S6 contains the proofs of Lemmas 1 and 2, that are stated in Appendix A.2.

S1 Illustration of relationship between expectiles and quantiles

An illustration of relationship (10) is provided in Figure S.1, which depicts for each value α the corresponding value $\omega(\alpha)$ such that $\tau_{\omega(\alpha)} = q_\alpha$ when the distribution of ϵ is a standard normal distribution (solid line) or Student-t distribution with 5 degrees of freedom (dashed line). For $\alpha < 0.5$ (respectively $\alpha > 0.5$) the $\omega(\alpha)$ curve lies below (respectively above) the 45° line (the grey line) where $\alpha = \omega(\alpha)$.

S2 Proof of Proposition 1

From (5) the ω th conditional expectile of Y , $\tau_\omega(x)$, can be defined as the solution of

$$\begin{aligned} \omega &= \frac{\mathbb{E}_{Y|X} [|Y - \tau_\omega(X)| \mathbb{1}\{Y \leq \tau_\omega(X)\} | X = x]}{\mathbb{E}_{Y|X} [|Y - \tau_\omega(X)| | X = x]} \\ &= \frac{\int |y - \tau_\omega(x)| \mathbb{1}\{y \leq \tau_\omega(x)\} dF_{Y|X}(y|x)}{\int |y - \tau_\omega(x)| dF_{Y|X}(y|x)} \\ &= \frac{\int_{-\infty}^{\tau_\omega(x)} (y - \tau_\omega(x)) dF_{Y|X}(y|x)}{\int_{-\infty}^{\tau_\omega(x)} (y - \tau_\omega(x)) dF_{Y|X}(y|x) - \int_{\tau_\omega(x)}^{+\infty} (y - \tau_\omega(x)) dF_{Y|X}(y|x)} \\ &= \frac{\int_{-\infty}^{\tau_\omega(x)} y dF_{Y|X}(y|x) - \int_{-\infty}^{\tau_\omega(x)} \tau_\omega(x) dF_{Y|X}(y|x)}{\int_{-\infty}^{\tau_\omega(x)} y dF_{Y|X}(y|x) - \int_{\tau_\omega(x)}^{+\infty} y dF_{Y|X}(y|x) + \int_{\tau_\omega(x)}^{+\infty} \tau_\omega(x) dF_{Y|X}(y|x)}. \end{aligned}$$

Adam, C.
Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven
E-mail: cecile.adam@kuleuven.be

Gijbels, I., Corresponding Author
Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven
E-mail: irene.gijbels@kuleuven.be

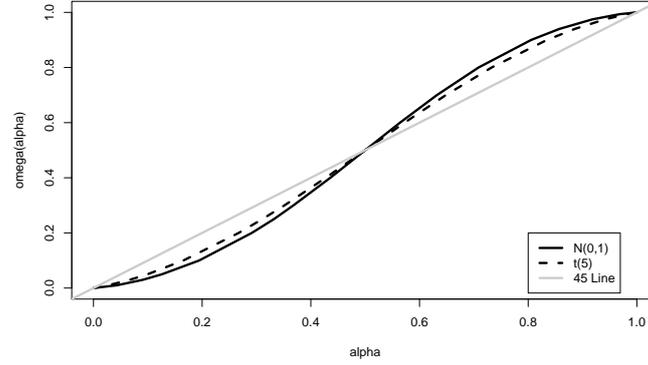


Figure S.1: One-to-one mapping between α and $\omega(\alpha)$ for a standard normal distribution and a Student-t distribution with 5 degrees of freedom as distribution for ϵ .

From the definitions of conditional expectiles and quantiles, it is clear that, for any $0 < \alpha < 1$, and any x which is a point of continuity of F_X , there exists a $\omega(\alpha, x)$ such that

$$\tau_{\omega(\alpha, x)}(X) = q_\alpha(X).$$

Henceforth,

$$\begin{aligned} \omega(\alpha, x) &= \frac{\int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x) - \int_{-\infty}^{q_\alpha(x)} q_\alpha(x) dF_{Y|X}(y|x)}{\int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x) - \int_{-\infty}^{q_\alpha(x)} q_\alpha(x) dF_{Y|X}(y|x) - \int_{q_\alpha(x)}^{+\infty} y dF_{Y|X}(y|x) + \int_{q_\alpha(x)}^{+\infty} q_\alpha(x) dF_{Y|X}(y|x)} \\ &= \frac{\int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x) - \mathbb{E}_{Y|X}[q_\alpha(X) \mathbb{1}\{Y \leq q_\alpha(X)\} | X = x]}{\mathbb{E}_{Y|X}[Y(\mathbb{1}\{Y \leq q_\alpha(X)\} - \mathbb{1}\{Y > q_\alpha(X)\}) | X = x] + \mathbb{E}_{Y|X}[q_\alpha(X)(\mathbb{1}\{Y > q_\alpha(X)\} - \mathbb{1}\{Y \leq q_\alpha(X)\}) | X = x]} \\ &= \frac{\int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x) - \alpha q_\alpha(x)}{2\mathbb{E}_{Y|X}[Y \mathbb{1}\{Y \leq q_\alpha(X)\} | X = x] - \mathbb{E}_{Y|X}[Y | X = x] + q_\alpha(x)(1 - 2\alpha)} \\ &= \frac{\alpha q_\alpha(x) - \int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x)}{2 \left[\alpha q_\alpha(x) - \int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x) \right] + \left[\mathbb{E}_{Y|X}[Y | X = x] - q_\alpha(x) \right]}. \end{aligned}$$

□

S3 Additional simulation results

S3.1 Performances of practical bandwidth selectors

In Section 6.2.1 we presented simulation results for Model 2 to illustrate the finite-sample performances of the practical bandwidth selectors. Here we show additional simulation results for Models 1 and 3.

Figure S.2 depicts kernel density estimates of $\check{h}_{\text{opt}}^{[k]} - h_{\text{opt}}$ for $k = 1, 2, 3$, for Model 1, for sample sizes $n = 100$, $n = 500$ and $n = 1000$. As before, for the purpose of visual comparison the range of the vertical and horizontal axis are kept the same for all three plots, with a vertical line indicating the position of the point zero. Firstly, all density estimates are concentrated around a positive value, meaning that mostly the bandwidth estimates are larger than the theoretical optimal bandwidth. Secondly, the density estimates of the three ROT bandwidth selectors GenROT, LSROTWith and LSROT are quite comparable. Also here, using knowledge of the error distribution (in LSROT) or not makes little difference. Thirdly, the bandwidth selector LSROTWithout is further away from the theoretical bandwidth h_{opt} . Fourthly, all bandwidth selectors improve with increasing n , but the convergence of them to h_{opt} is quite slow: for sample size $n = 100$, the difference $\hat{h} - h_{\text{opt}}$ is, for GenROT, LSROTWith

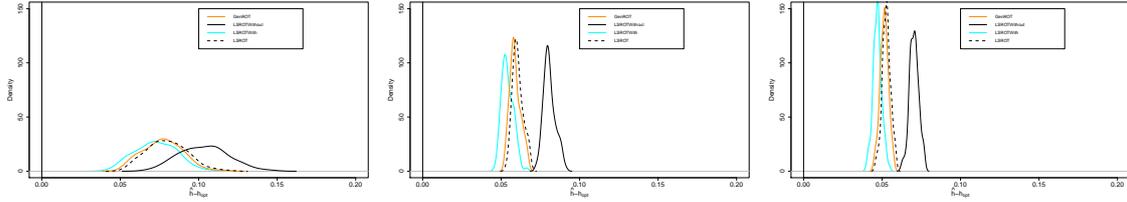


Figure S.2: Model 1. Kernel density estimates of the three ROT bandwidth selectors of Sections 5.2 and 5.3, for estimation of $\tau_{0.3}(\cdot)$. The vertical lines indicates the zero position. Sample size $n = 100$ (left), $n = 500$ (middle) and $n = 1000$ (right).

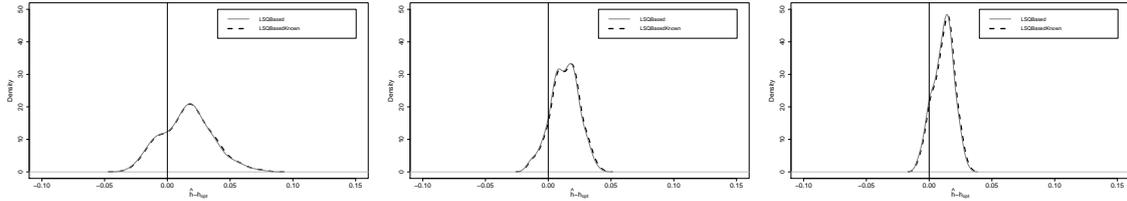


Figure S.3: Model 1. Kernel density estimates of the quantile-based bandwidth selector of Section 5.3.3, with or without assuming knowledge of the distribution of ϵ , for estimation of $\tau_{0.3}(\cdot)$. The vertical lines indicates the zero position. Sample size $n = 100$ (left), $n = 500$ (middle) and $n = 1000$ (right).

and LSROT, concentrated around 0.075, and for $n = 1000$ this mode position has shifted closer to zero, and is around 0.05.

Figure S.3 shows density estimates of the quantile-based bandwidth selector of Section 5.3.3 for Model 1, when either using knowledge about the error distribution or not. Note again the slow convergence, and also the almost indistinguishable density estimates when comparing the situations of knowing or not the error distribution.

As remarked already, from Figures S.2 and S.3 one can notice that the bandwidth selectors tend to provide larger values than their theoretical counterparts for Model 1. Recall that all bandwidth selectors are based on quite some rough estimations in the ROT-development. What is of importance of course is to investigate whether such rough bandwidth selectors can lead to good performance for the expectile estimator $\hat{\tau}_\omega(\cdot)$. We investigated this, and present results regarding this aspect for Model 1 in Section S3.2.

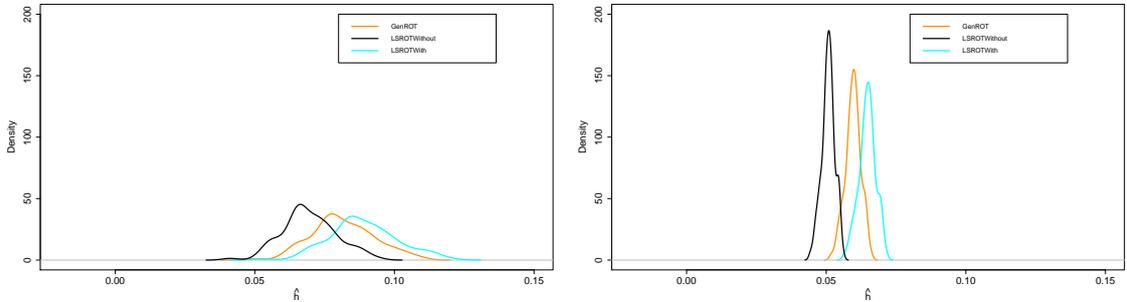


Figure S.4: Model 3. Kernel density estimates of the three ROT bandwidth selectors of Sections 5.2 and 5.3, for estimation of $\tau_{0.3}(\cdot)$. Sample size $n = 100$ (left) and $n = 1000$ (right).

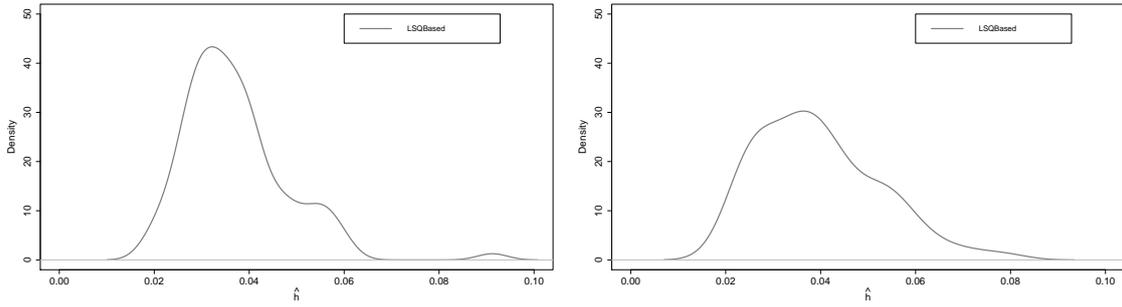


Figure S.5: Model 3. Kernel density estimates of the quantile-based bandwidth selector of Section 5.3.3 for estimation of $\tau_{0.3}(\cdot)$. Sample size $n = 100$ (left) and $n = 1000$ (right).

In Figures S.4 and S.5 we depict the kernel density estimates of all bandwidth selectors for Model 3, for sample sizes $n = 100$ and $n = 1000$. In contrast to the pictures for Models 1 and 2, the bandwidth selector values are presented on the horizontal axes. There is no position of reference now, but one can notify that the density estimates in Figure S.4 are more concentrated for $n = 1000$ than for $n = 100$. This is not the case for the density estimates in Figure S.5. Recall that the derivations in Section 5.3 might not be valid, since Model 3 is not a location-scale model. The only bandwidth selector for which the validity is guaranteed is the GenROT bandwidth selector. Note however, from Figure S.4 that the bandwidth selectors LSROTWith and LSROTWithout perform quite comparable, in contrast to the quantile-based bandwidth selector illustrated in Figure S.5.

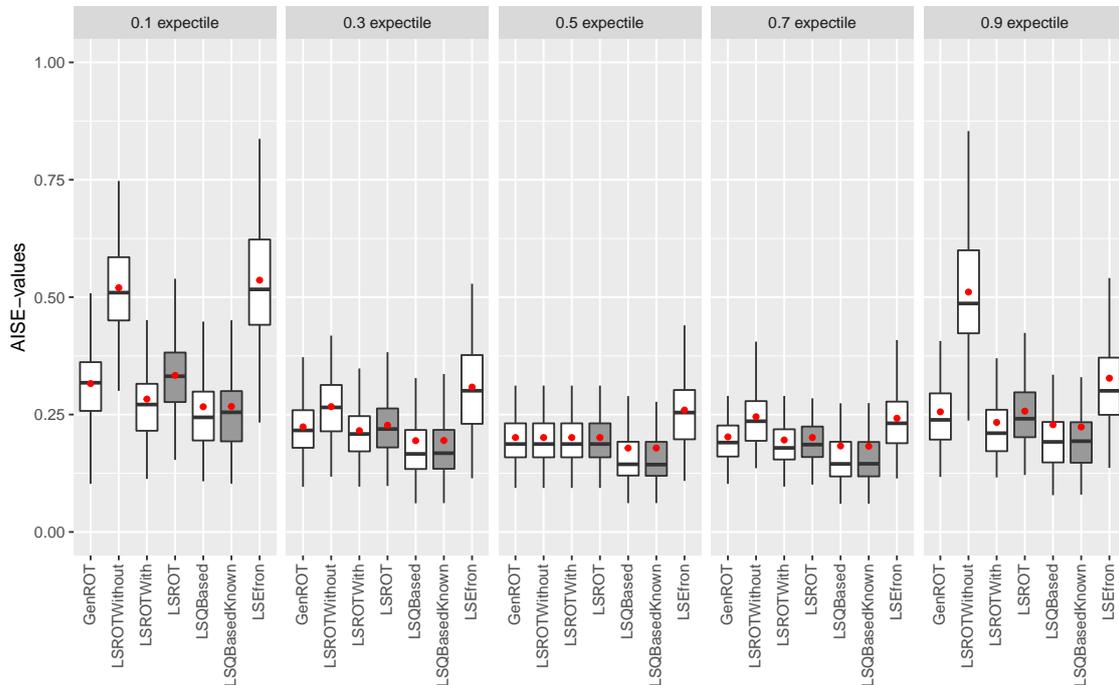


Figure S.6: Model 1. Boxplots of the AISE-values from 100 simulated samples using the different methods listed in Table 1. Grey-filled boxplots are for the cases when we assume the error distribution to be known.

S3.2 Simulation results for Model 1

Figure S.6 shows boxplots of the AISE-values of the local linear expectile regression estimates for all methods in Table 1, for the considered values of ω . As for Model 2 the performances of the local linear estimation method with any of the Rule-of-Thumb (ROT) bandwidth selectors are quite comparable (see the first four boxplots). The quantile-based bandwidth selector of Section 5.3.3 seems to be a good choice here, although the performance of some of the ROT bandwidth selectors is also good.

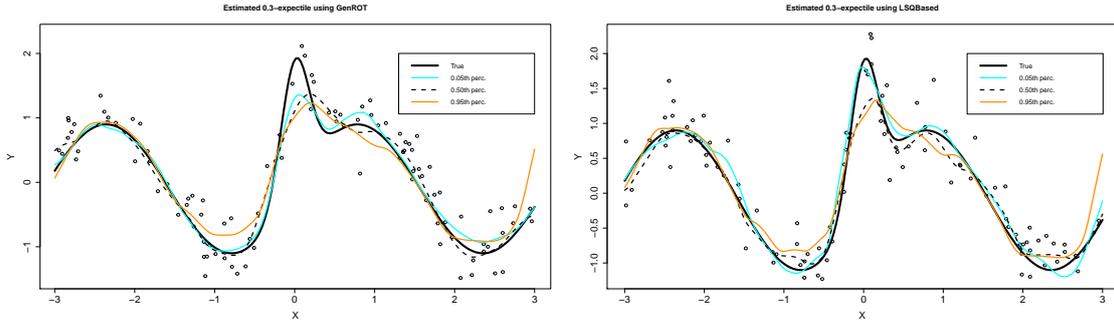


Figure S.7: *Model 1. True expectile curve $\tau_{0.3}(\cdot)$ (in black) and three representative local linear estimates: 0.05th AISE-percentile (light-grey; color blue), 0.5th AISE-percentile (dashed line), 0.95th AISE-percentile (grey; color ochre yellow), using respectively the bandwidth selection method GenROT (left panel) and the LSQBased method (right panel).*

Figure S.7 gives a graphical idea about the quality of the estimator $\hat{\tau}_\omega(\cdot)$ by presenting the true expectile curve $\tau_{0.3}(\cdot)$ together with the three representative estimates. The shown scatter plot is that of the sample with the median performance. The estimated curves are slightly smoother when using the GenROT bandwidth selector.

ω	GenROT	LSROT-Without	LSROT-With	LSROT	LSQBased	LSQBased-Known	LSEfron
MODEL 1							
0.1	23.18	24.12	50.50	27.99	49.99	21.08	4033.38
0.3	20.67	21.67	53.80	33.28	53.29	27.25	1946.12
0.5	10.52	11.55	42.28	12.41	42.36	13.98	3393.56
0.7	21.33	21.52	49.03	22.40	48.38	21.50	5251.61
0.9	23.22	25.05	54.95	23.64	51.89	24.50	7617.66
MODEL 2							
0.1	23.89	25.02	56.61	38.47	56.20	31.36	3865.96
0.3	21.11	21.68	50.03	21.94	49.38	26.38	1670.48
0.5	10.34	11.38	47.86	11.33	47.70	11.07	3700.24
0.7	20.89	22.23	66.08	20.16	65.70	21.95	5513.80
0.9	22.60	21.41	67.61	20.73	63.88	20.77	7570.03

Table S.1: *Computation times in seconds for Models 1 and 2.*

S3.3 Computational costs

In Table S.1 (first block of lines) we present average computing times for the local linear regression expectile estimator (for various values of ω) for the several bandwidth selectors (and methods) in Table 1. The Efron quantile-based approach, which was considered also in Yao and Tong (1996), has

a high computational cost. The lowest average computation cost is for the methods GenROT and LSROTWithout. The conclusions are in line with these of the simulations for Model 2 (see the second block of lines in Table S.1).

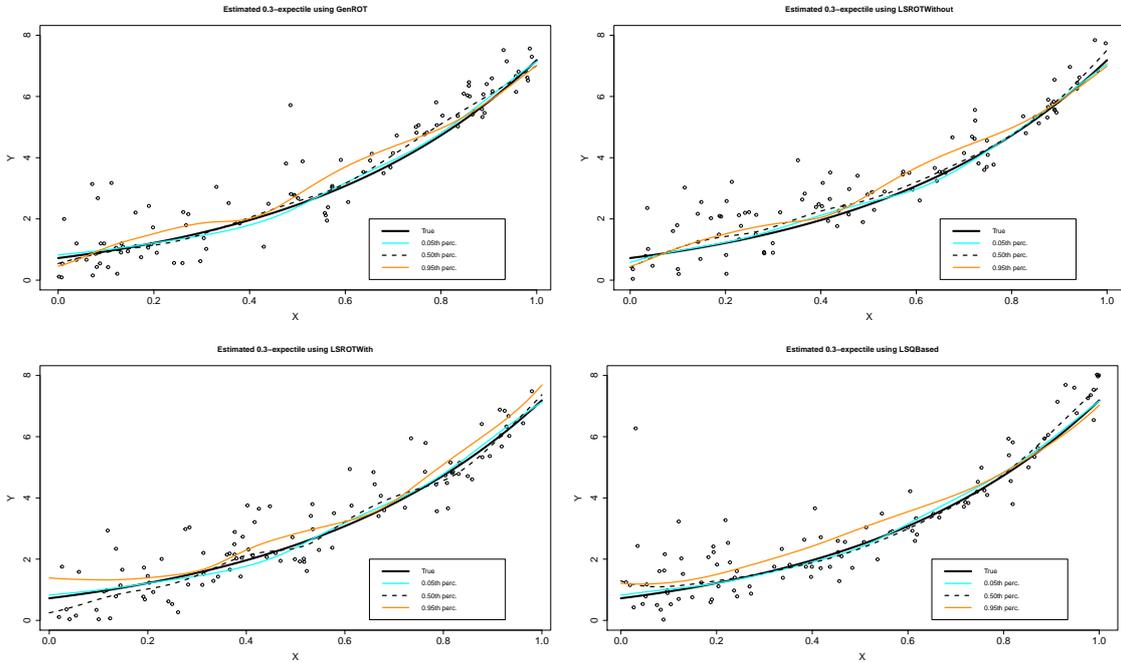


Figure S.8: *Model 3. True expectile curve $\tau_{0.3}(\cdot)$ (in black) and three representative local linear estimates: 0.05th AISE-percentile (light-grey; color blue), 0.5th AISE-percentile (dashed line), 0.95th AISE-percentile (grey; color ochre yellow), using respectively the bandwidth selection method GenROT (top left), the LSROTWithout (top right), the LSROTwith (bottom left) and the LSQBased method (bottom right).*

S3.4 Additional information on results for Model 3

Figure S.8 depicts the true $\tau_{0.3}(\cdot)$ expectile curve for this model, together with the three representative estimates ($n = 100$), for the four methods for which boxplots are shown in Figure 9. From this the results seem quite comparable for all methods.

S3.5 Finite-sample evaluation of practical implementation issues

An approximation of the minimizer of (12) is obtained through the iterative procedure of Section 3.2, using a local polynomial least squares regression estimator $\beta^{(0)}$ (see (17)) as a starting vector, and applying stopping criterion (18). In this part of the numerical study we want to get insights in two issues: (i) the number of iterations needed; and (ii) the impact of the choice for the starting vector $\beta^{(0)}$.

A first aim is to get insight in the number of iterations needed. Figure S.9 shows the convergence of the iterative algorithm when estimating $\tau_{0.7}(x)$, considering an equispaced grid of 100 values for x in the interval $[-3, 3]$. For comparison purpose we superimpose the true curve $\tau_{0.7}(x)$ (in red color). The left (respectively right) column of plots presents simulation results for sample size $n = 100$ (respectively $n = 200$). The subsequent plots in a column (from top to bottom) show the boxplots of the resulting

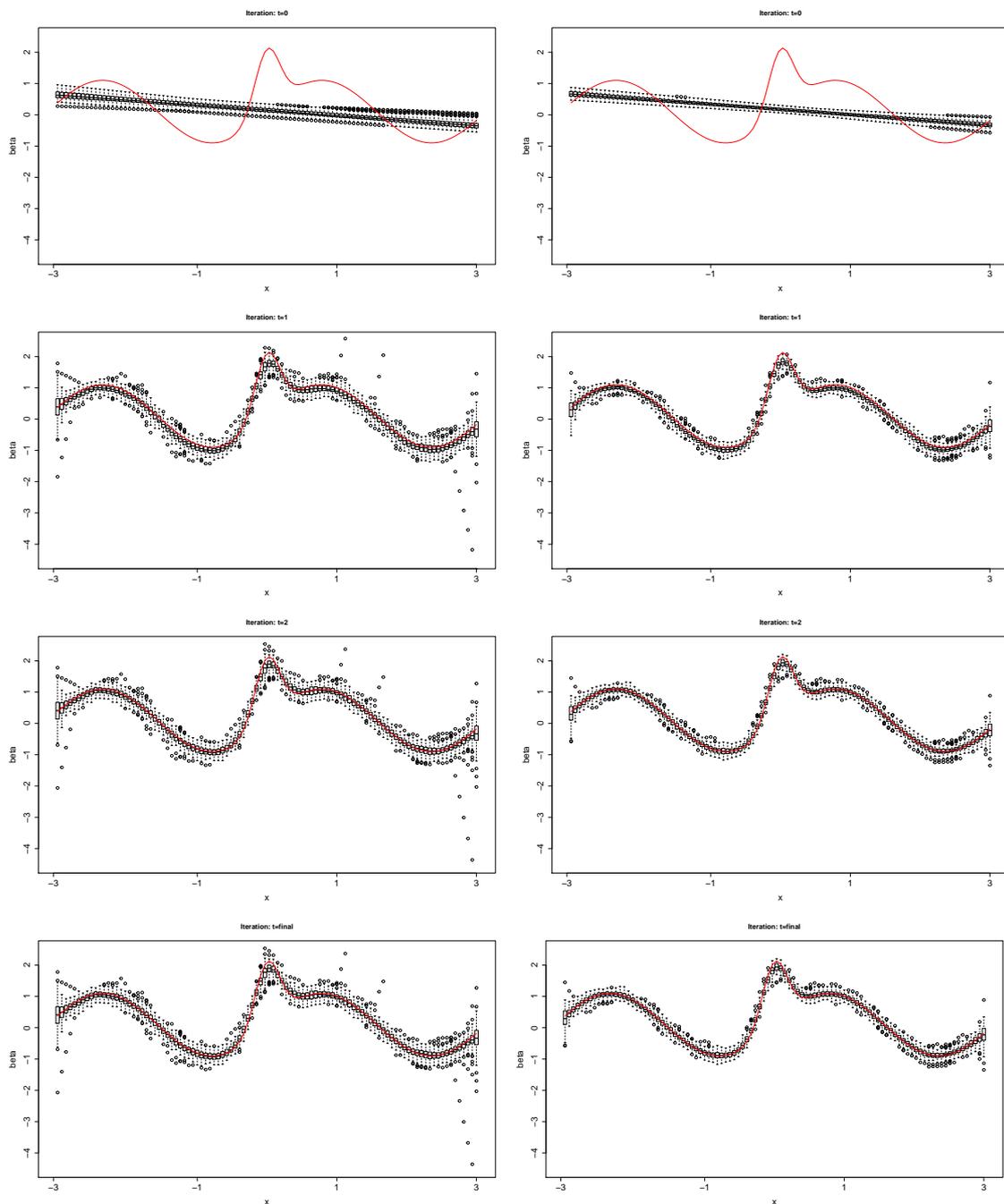


Figure S.9: Model 1. Estimation of $\tau_{0.7}(x)$. Boxplots of the $\hat{\beta}_0^{(t)}(x)$ after t iterations from 100 simulated samples of size $n = 100$ (left) and $n = 200$ (right). The true curve $\tau_{0.7}(x)$ is the red curve.

estimate of $\tau_{0.7}(x)$ after 0, 1 and 2 iterations, and after convergence of the algorithm (using the indicated stopping rule). The latter is denoted in the plots by $t=\text{final}$. Here we use for bandwidth the theoretical optimal bandwidth h_{opt} . From Figure S.9 it is seen that convergence occurs already after 2 iterations. The bias and the variance of the estimator decrease when passing from sample size $n = 100$ to $n = 200$.

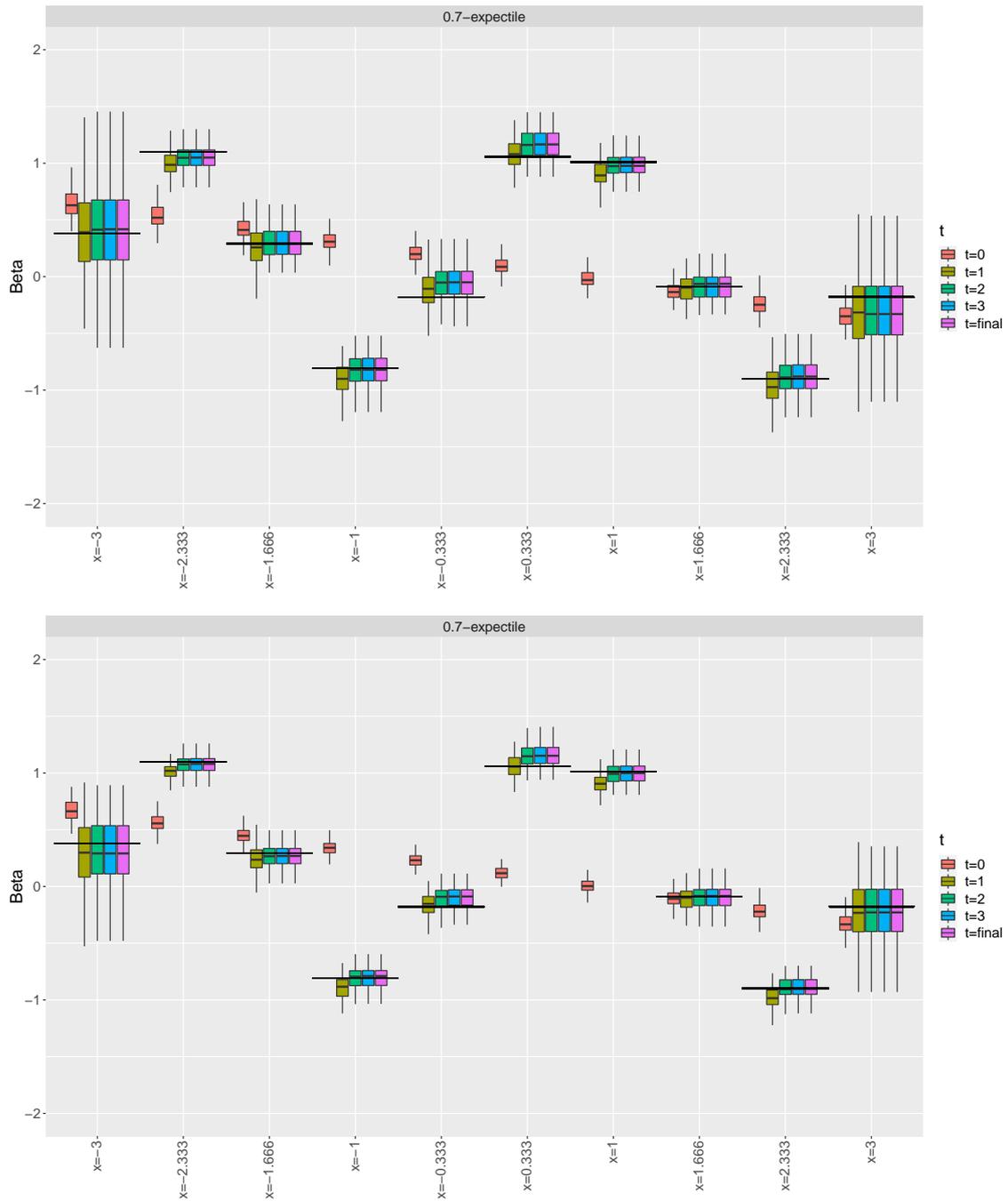


Figure S.10: *Model 1. Estimation of $\tau_{0.7}(x)$. Boxplots of the approximate estimate $\beta^{(t)}(\cdot)$ after t iteration from 100 simulated samples for each point x in a grid of 10 equispaced grid-values between -3 and 3 . True values of $\beta(\cdot)$ is added by the horizontal lines. Sample size: $n = 100$ (top panel) and $n = 200$ (bottom panel).*

Figure S.10 give a different presentation of the results for sample sizes respectively $n = 100$ and $n = 200$, for only a selection of 10 different equispaced grid-values x (see the horizontal axis). In these figures boxplots summarize the simulation results after using 0 iterations (just using the starting vector as an estimator), 1 iteration, 2 iterations, 3 iterations, and after convergence (using the stopping

criterion), indicated with $t=\text{final}$. The true value $\beta_0 = \tau_\omega(x)$ is displayed with the horizontal line crossing the boxplots. Note first of all that for some x -values the starting vector value is far away from the true (target) value $\tau_\omega(x)$. This is in particular the case for $x = \pm 1$ and $x = \pm 0.333$. If one or two iterations are used the estimates already tend to converge to the final estimator. By comparing the top and bottom panels of Figures S.10 one gets an idea about the improvements for increasing sample size.

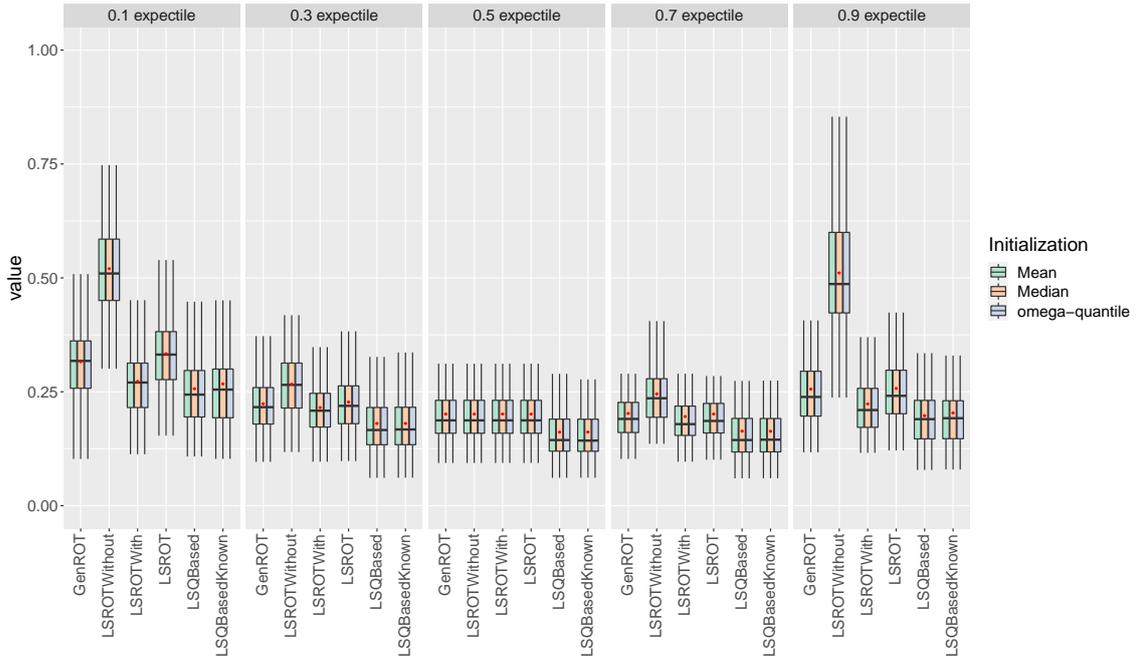


Figure S.11: Model 1. Boxplots of the AISE-values for the estimator of $\tau_\omega(x)$ from 100 simulated samples using the different methods listed in Table 1 using different starting values.

A second aim is to obtain insights in the impact of the choice of the starting vector. We considered three different methods for choosing the starting vector:

- Mean *initialization*: using a polynomial mean regression estimator, as defined in (17);
- Median *initialization*: using a polynomial median regression estimator (replacing the squared loss in (17) by the loss function $R_{0.5}(\cdot)$);
- ω -quantile *initialization*: using a polynomial ω th quantile regression estimator (replacing the squared loss in (17) by the loss function $R_\omega(\cdot)$).

We consider here estimation of $\tau_\omega(x)$, for x in the same equispaced grid on the interval $[-3, 3]$, for $\omega = 0.1, 0.3, 0.5, 0.7$ and 0.9 . For each simulated sample (of size $n = 100$) we applied the iterative procedure (until convergence with the stopping criterion) using the three different methods of initialization, and with six bandwidth selectors. Boxplots reporting on the AISE-values for estimation of $\tau_\omega(x)$, across all ten values of x in the grid, are depicted in Figure S.11. For each ω value six groups of each three boxplots are presented, referring to six bandwidth selectors and the three initialization methods. As can be seen the estimates after convergence of the iterative procedure lead to similar results for any of the three initialization methods. The findings with respect to the finite-sample performances for the different bandwidth selectors, are fully in line with these seen in Figure S.6.

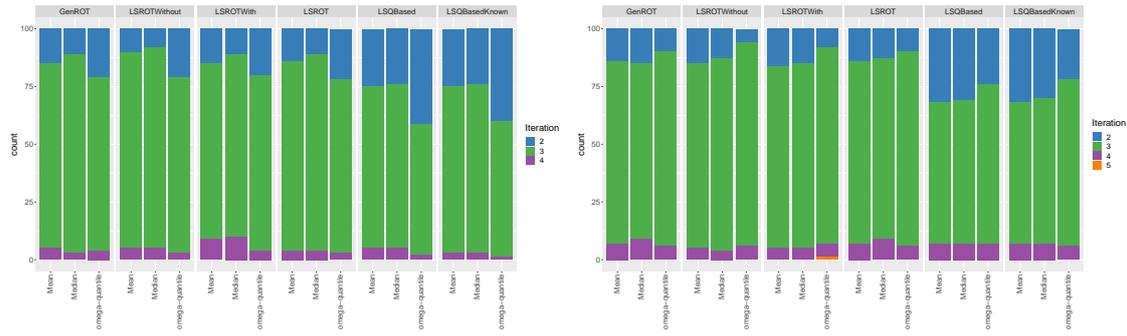


Figure S.12: Model 1. Barplots showing the percentages (over 100 simulated samples) of the number of iterations using the different methods listed in Table 1 using different starting values for one $x = -1.5$ and for the 0.3-expectile (left panel) and the 0.7-expectile (right panel).

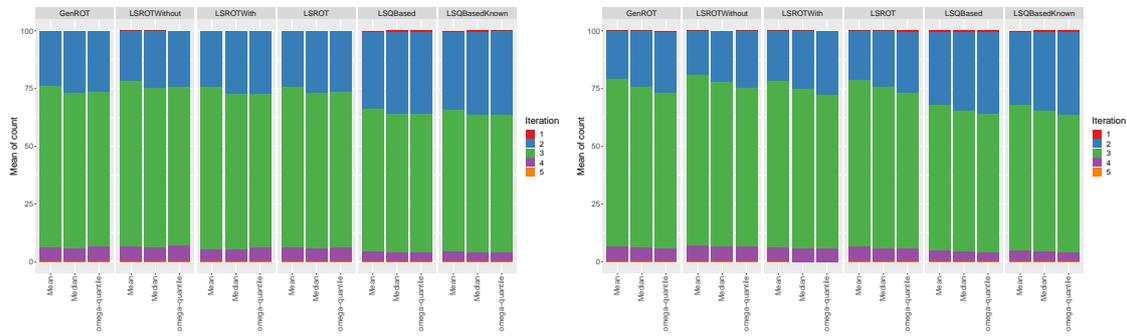


Figure S.13: Model 1. Barplots showing the percentages (over 100 simulated samples) of the number of iterations using the different methods listed in Table 1 using different starting values for 100 values of x and for the 0.3-expectile (left panel) and the 0.7-expectile (right panel).

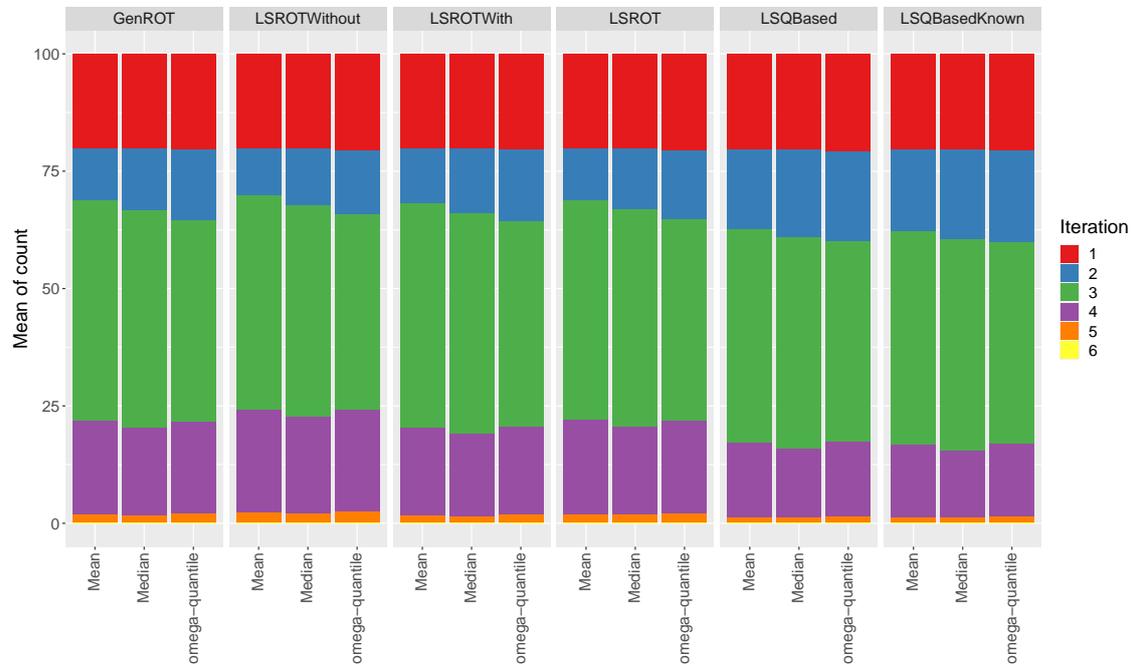


Figure S.14: Model 1. Barplots showing the percentages (over 100 simulated samples) of the number of iterations using the different methods listed in Table 1 using different starting values for 100 values of x and considering all 5 values of ω .

Although the final estimator after convergence of the algorithm appears as very close for all three initialization methods, one might wonder whether the choice of the initialization method has an impact on the number of iterations required, i.e. how the choice of initialization method possibly impacts how fast the algorithm converges. We therefore report in Figures S.12, S.13 and S.14 on the percentages of number of iterations across the 100 simulated samples (of size $n = 100$), when using the different starting values (on the horizontal axes). Figure S.12 reports on the results for estimating $\tau_\omega(x)$, with $x = -1.5$, for two different values of ω ($\omega = 0.3$ in the left panel and $\omega = 0.7$ in the right panel). In all simulations only 2,3,4 or 5 iterations (until convergence of the algorithm) were needed. Note that a similar pattern is seen when using any of the six bandwidth selection methods. Figure S.13 summarizes the results for estimating $\tau_\omega(x)$, across all values of x in a considered grid of 100 values, and for the same two values of ω . Apart from a couple of times that only one iteration was needed, the findings are the same, as for Figure S.12. Finally, Figure S.14 summarizes the obtained results for all 100 values of x and the five considered values of ω . Since for $\omega = 0.5$, the median and the expectile coincide, this leads to a number of cases where only 1 iteration was needed. In some rare cases 6 iterations were needed (a small percentage that it is even not visible on the plots). Overall, from Figures S.12, S.13 and S.14 it appears that in the large majority of simulations the algorithm converged after two or three iterations.

S3.6 Illustration of the asymptotic normality result

In Theorem 1 the asymptotic distribution for $\hat{\beta}$ the minimizer of (12) is established. Via a small simulation study we investigate the finite-sample distribution of the approximate estimator obtained via the iterative procedure discussed in Section 3.2.

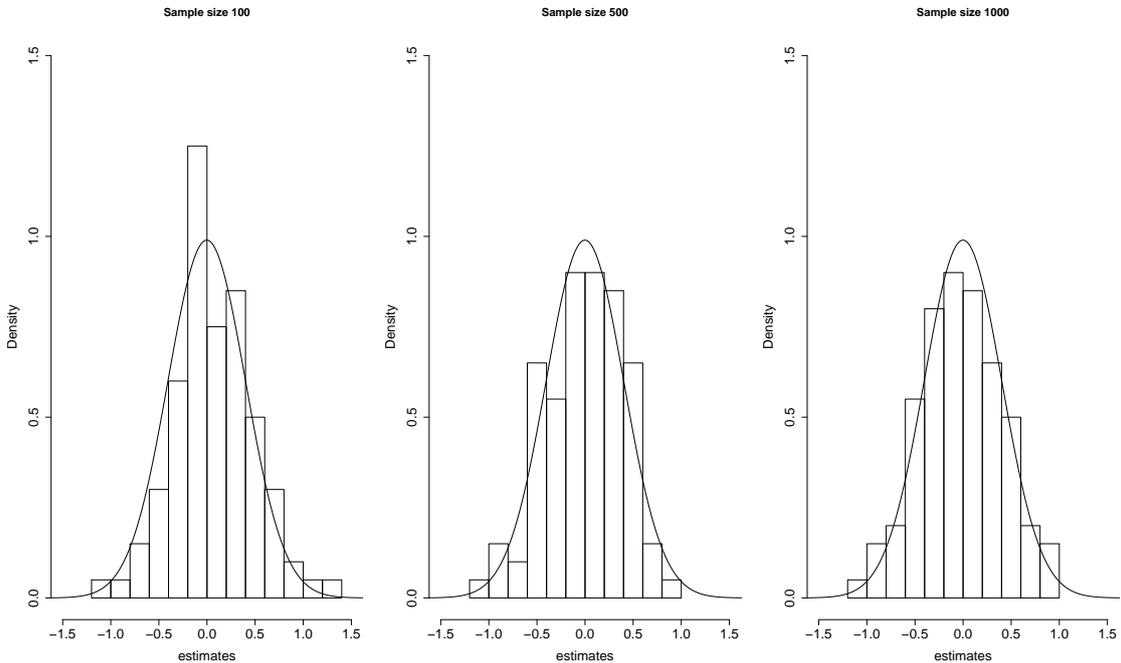


Figure S.15: *Model 1. Finite-sample distribution of $\sqrt{nh}(\hat{\tau}_\omega^{(0,A)}(x) - \tau_\omega^{(0)}(x))$, for $x = 1.5$, and its approximate (asymptotic) normal distribution imposed (solid curve).*

For given sample size ($n = 100, 500$ and 1000) we simulated 100 samples from Model 1, and use the iterative procedure of Section 3.2 to get an approximation $\hat{\tau}_\omega^{(0,A)}(x)$ of the local polynomial

expectile estimator, i.e. the first element $\hat{\tau}_\omega^0(x)$ of the minimizer of (12). Based on each sample we calculate, for $\omega = 0.3$ and the point $x = 1.5$, the estimate $\hat{\tau}_\omega^{(0,A)}(x)$. We expect that

$$\sqrt{nh} \left(\hat{\tau}_\omega^{(0,A)}(x) - \tau_\omega^{(0)}(x) - \beta_\omega^{(0)}(x)h^2 \right) \approx \mathcal{N} \left(0, (\sigma_\omega^{(0)}(x))^2 \right), \quad (\text{S.1})$$

with $\beta_\omega^{(0)}(x)$ and $(\sigma_\omega^{(0)}(x))^2$ as defined in Theorem 1, and calculated for simulation Model 1. We used the theoretical optimal bandwidth for this illustration.

We look at the histograms of the obtained 100 values for $\sqrt{nh} \left(\hat{\tau}_\omega^{(0,A)}(x) - \tau_\omega^{(0)}(x) \right)$. These histograms are depicted in Figure S.15, for the three sample sizes. Superimposed on the plots are the approximate (asymptotic) normal densities $\mathcal{N} \left(\beta_\omega^{(0)}(x)h^2; (\sigma_\omega^{(0)}(x))^2 \right)$, obtained from (S.1). As can be seen the approximation with the normal distribution improves with the sample size.

S4 Real data examples

In this section we present two additional real data examples.

- The Motorcycle impact data set of Schmidt et al. (1981). The data is from a study of the effectiveness of helmets in collisions with motorcycles and contains 133 observations. The X values are time measurements in milliseconds (ms) after a simulated impact with a post mortem human test object. The Y values are measurements of head acceleration in units of g (9.8 meters/sec²). The data are, for example, available in the R package MASS.
- The LIDAR dataset is described in Ruppert et al. (2003), among others. The data are from a light detection and ranging experiment and the data set contains 221 observations. The X values are the distances that the light travelled before it was reflected back to its source and the Y values is the logarithm of the ratio of received light from two laser sources. The data can be found in, for example, the R package SemiPar.

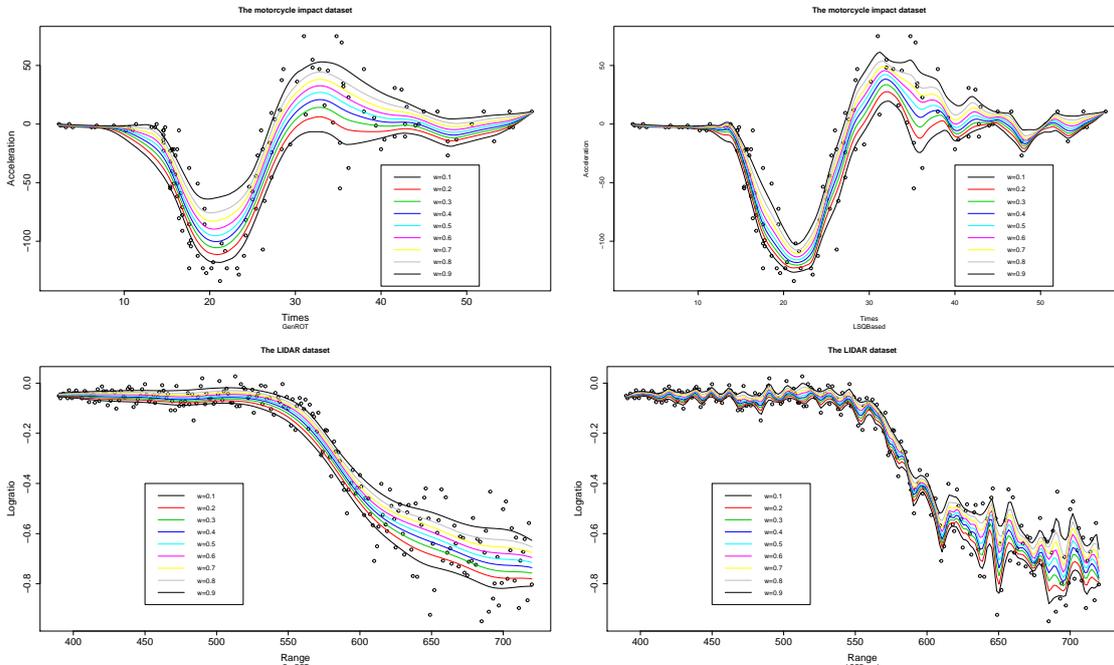


Figure S.16: Estimated expectile regression curves for $\tau_\omega(\cdot)$ for ω taking values 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 for the two data examples. Estimated expectiles using the GenROT (left panels) and the LSQBased (right panels) bandwidth selectors.

Scatterplots of the data together with the estimated expectile regression curves for ω values equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 are shown in Figure S.16. For both examples the estimated expectiles using the GenROT bandwidth selector are smoother than these using the LSQBased bandwidth selector. This is most visible for the LIDAR data. Table S.2 provides the values of the data-driven bandwidths for both data examples, for five considered ω values. Bandwidth values from the GenROT method are for all cases larger than the values from the LSQBased bandwidth selector. Note in particular the substantially larger values for the LIDAR data. The estimated expectile curves are clearly non-parallel for the motorcycle impact data and the LIDAR data. The two data sets clearly show heteroscedasticity.

ω	Motorcycle		LIDAR	
	GenROT	LSQBased	GenROT	LSQBased
0.1	2.2801	1.0691	16.3997	2.2999
0.3	2.3612	1.0544	16.2280	2.2504
0.5	2.3473	1.0531	16.0395	2.2492
0.7	2.3530	1.0711	15.9404	2.3922
0.9	2.2739	1.0727	16.3682	2.6304

Table S.2: *Motorcycle impact data and LIDAR data. Data-driven bandwidth values.*

S5 Additional explanations concerning bandwidth selectors in Sections 5.3.1 and 5.3.2

Note from the expression of the optimal theoretical bandwidth in (22) that one of the unknown quantities to be approximated is $\gamma(\omega, x)$, which in its turn requires approximations of the quantities

$$P[Y \leq \tau_\omega(X)|X = x] \quad \text{and} \quad P[Y > \tau_\omega(X)|X = x].$$

We focus our explanation here on an approximation for $P[Y \leq \tau_\omega(X)|X = x]$.

In general we start from

$$P[Y \leq \tau_\omega(X)|X = x] = P[Y - \tau_\omega(X) \leq 0|X = x] = E[\mathbb{1}\{Y - \tau_\omega(X) \leq 0\} | X = x],$$

and approximate this conditional expectation by its expectation

$$E_X \{E_{Y|X} [\mathbb{1}\{Y - \tau_\omega(X) \leq 0\} | X]\} = E\{\mathbb{1}\{Y - \tau_\omega(X) \leq 0\}\},$$

which is then estimated by an empirical version

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Y_i - \check{\tau}_\omega(X_i) \leq 0\}. \quad (\text{S.2})$$

In a location-scale model however one has knowledge of an additional structure which one should try to exploit. Recall first of all that in a location-scale model

$$\tilde{\epsilon} = Y - \tau_\omega(x) = \sigma(X) [\epsilon - \tau_{\omega,\epsilon}]. \quad (\text{S.3})$$

From this relationship between $\tilde{\epsilon}$ and ϵ one can also provide relationships between the conditional distribution function of $\tilde{\epsilon}$ given $X = x$ and the cumulative distribution function of ϵ , which we denoted by F_ϵ . One obtains, for any v ,

$$\begin{aligned} P(\tilde{\epsilon} \leq v | X = x) &= P(\sigma(X) [\epsilon - \tau_{\omega,\epsilon}] \leq v | X = x) = P\left(\epsilon \leq \frac{v}{\sigma(X)} + \tau_{\omega,\epsilon} | X = x\right) \\ &= F_\epsilon\left(\frac{v}{\sigma(x)} + \tau_{\omega,\epsilon}\right), \end{aligned}$$

where in the last line it is used that ϵ and X are independent. For the conditional p th quantile function of $\tilde{\epsilon}$ given X , with $0 < p < 1$, this leads to

$$F_{\tilde{\epsilon}|X}^{-1}(p|x) = \sigma(x) [F_{\epsilon}^{-1}(p) - \tau_{\omega,\epsilon}]. \quad (\text{S.4})$$

Recall that

$$P[Y \leq \tau_{\omega}(X)|X = x] = E[\mathbb{1}\{Y - \tau_{\omega}(X) \leq 0\} | X = x],$$

and that the ω th conditional expectile of $Y - \tau_{\omega}(X)$ given $X = x$ is zero, under the location-scale model. Hence, under such a modelling framework, one can replace the zero above by the ω th conditional expectile. These considerations then led to approximating $P[Y \leq \tau_{\omega}(X)|X = x]$ by its unconditional expectation, and hence by

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Y_i - \check{\tau}_{\omega}(X_i) \leq \check{\tau}_{\omega}(\check{\epsilon}_i)\}. \quad (\text{S.5})$$

This reasoning was used in Section 5.3.1.

In Section 5.3.2 we, in addition to the above, try to exploit that under a location-scale model one has (see (10))

$$\omega(\alpha, x) = \omega(\alpha) = \frac{\alpha F_{\epsilon}^{-1}(\alpha) - E_{\epsilon}[\epsilon \mathbb{1}\{\epsilon \leq F_{\epsilon}^{-1}(\alpha)\}]}{2E_{\epsilon}[\epsilon \mathbb{1}\{\epsilon > F_{\epsilon}^{-1}(\alpha)\}] - (1 - 2\alpha)F_{\epsilon}^{-1}(\alpha)}. \quad (\text{S.6})$$

Since we can obtain pseudo-observations for $\tilde{\epsilon} = Y - \tau_{\omega}(X)$, via the (global) parametric estimation of the ω th conditional expectile of Y given $X = x$, and hence first translate (S.6) in terms of $\tilde{\epsilon}$ instead of in terms of ϵ . This is easily done by utilizing (S.3) and (S.4).

We translate each of the terms in (S.6). We have from (S.4) that, for $0 < \alpha < 1$,

$$F_{\epsilon}^{-1}(\alpha) = \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) + \tau_{\omega,\epsilon}. \quad (\text{S.7})$$

Since $\tilde{\epsilon} = \sigma(X)[\epsilon - \tau_{\omega,\epsilon}]$,

$$\mathbb{1}\{\epsilon \leq F_{\epsilon}^{-1}(\alpha)\} = \mathbb{1}\left\{\frac{\tilde{\epsilon}}{\sigma(X)} + \tau_{\omega,\epsilon} \leq \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) + \tau_{\omega,\epsilon}\right\} = \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\}.$$

It then follows that

$$\begin{aligned} E_{\epsilon}[\epsilon \mathbb{1}\{\epsilon \leq F_{\epsilon}^{-1}(\alpha)\}] &= E\left[\left(\frac{\tilde{\epsilon}}{\sigma(X)} + \tau_{\omega,\epsilon}\right) \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\}\right] \\ &= E_X\left\{E_{\tilde{\epsilon}|X}\left[\left(\frac{\tilde{\epsilon}}{\sigma(X)} + \tau_{\omega,\epsilon}\right) \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right]\right\} \\ &= E_X\left\{\frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X}\left[\left(\tilde{\epsilon} \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\}\right) \mid X\right] + E_{\tilde{\epsilon}|X}\left[\tau_{\omega,\epsilon} \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right]\right\} \\ &= E_X\left\{\frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X}\left(\tilde{\epsilon} \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right) + \alpha \tau_{\omega,\epsilon}\right\} \\ &= E_X\left\{\frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X}\left(\tilde{\epsilon} \mathbb{1}\left\{\tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right)\right\} + \alpha \tau_{\omega,\epsilon}. \end{aligned} \quad (\text{S.8})$$

Note that the term $\alpha \tau_{\omega,\epsilon}$ drops out when combining the terms in (S.7) and (S.8) to get to the numerator in (S.6).

We proceed in a similar fashion for rewriting the terms in the denominator of (S.6). We get that

$$\begin{aligned} E_{\epsilon}[\epsilon \mathbb{1}\{\epsilon > F_{\epsilon}^{-1}(\alpha)\}] &= E_X\left\{\frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X}\left[\left(\tilde{\epsilon} \mathbb{1}\left\{\tilde{\epsilon} > F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\}\right) \mid X\right] + E_{\tilde{\epsilon}|X}\left[\tau_{\omega,\epsilon} \mathbb{1}\left\{\tilde{\epsilon} > F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right]\right\} \\ &= E_X\left\{\frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X}\left(\tilde{\epsilon} \mathbb{1}\left\{\tilde{\epsilon} > F_{\tilde{\epsilon}|X}^{-1}(\alpha|X)\right\} \mid X\right)\right\} + (1 - \alpha) \tau_{\omega,\epsilon}. \end{aligned} \quad (\text{S.9})$$

Combining (S.7), (S.8) and (S.9) and substituting these into (S.6) we obtain the following alternative expression for the right-hand side of (S.6):

$$\frac{\alpha \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) - E_X \left\{ \frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X} \left(\tilde{\epsilon} \mathbb{1} \left\{ \tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \right\} | X \right) \right\}}{2 E_X \left\{ \frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X} \left(\tilde{\epsilon} \mathbb{1} \left\{ \tilde{\epsilon} > F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \right\} | X \right) \right\} - (1 - 2\alpha) \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) + \tau_{\omega, \epsilon}}. \quad (\text{S.10})$$

The term $\tau_{\omega, \epsilon}$ in the above is unknown, and one has no access to pseudo-observations of ϵ . Put since $E[\epsilon] = 0$ and using relationship (S.3) one can replace $\tau_{\omega, \epsilon}$ in (S.10) by

$$\tau_{\omega, \epsilon} = \tau_{\omega, \epsilon} - E[\epsilon] = \tau_{\omega, \epsilon} - E_{\tilde{\epsilon}, X} \left[\frac{\tilde{\epsilon}}{\sigma(X)} + \tau_{\omega, \epsilon} \right] = -E_{\tilde{\epsilon}, X} \left[\frac{\tilde{\epsilon}}{\sigma(X)} \right]. \quad (\text{S.11})$$

Combining (S.10) and (S.11) we obtain the following expression for the right-hand side of (S.6)

$$\frac{\alpha \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) - E_X \left\{ \frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X} \left(\tilde{\epsilon} \mathbb{1} \left\{ \tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \right\} | X \right) \right\}}{2 E_X \left\{ \frac{1}{\sigma(X)} E_{\tilde{\epsilon}|X} \left(\tilde{\epsilon} \mathbb{1} \left\{ \tilde{\epsilon} > F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \right\} | X \right) \right\} - (1 - 2\alpha) \frac{1}{\sigma(X)} F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) - E_{\tilde{\epsilon}, X} \left[\frac{\tilde{\epsilon}}{\sigma(X)} \right]}.$$

Following the same reasonings as those that led to (S.5), we write

$$\mathbb{1} \left\{ \tilde{\epsilon} \leq F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \right\} = \mathbb{1} \left\{ \tilde{\epsilon} - F_{\tilde{\epsilon}|X}^{-1}(\alpha|X) \leq 0 \right\},$$

and replace the zero above by the ω th conditional expectile of $\tilde{\epsilon}$ given $X = x$. The term in the denominator involving the indicator function, is looked upon in a similar manner.

Further approximations then consist of treating $\sigma(x)$ as a constant. The remaining quantities can then be estimated, using the pseudo-observations obtained via the global parametric fit, via

$$\frac{1}{n} \sum_{i=1}^n \check{\epsilon}_i \mathbb{1} \left\{ \check{\epsilon}_i - \check{F}_{\check{\epsilon}}^{-1}(\hat{\alpha}) \leq \check{\tau}_{\omega}(\check{\epsilon}_i) \right\}, \quad \frac{1}{n} \sum_{i=1}^n \check{\epsilon}_i \mathbb{1} \left\{ \check{\epsilon}_i - \check{F}_{\check{\epsilon}}^{-1}(\hat{\alpha}) > \check{\tau}_{\omega}(\check{\epsilon}_i) \right\}, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n \check{\epsilon}_i.$$

All the above considerations resulted into (27). Although the approximations used are quite rough, this simple attempt to exploit (to a full extent) the known location-scale setting, can pay off, as is seen from the simulation results.

S6 Proofs Lemmas 1 and 2

S6.1 Proof of Lemma 1

The reason for presenting second-order expansions is that first-order terms may vanish due to the symmetry of K which causes its odd moments to be null.

For some real value ξ between X_i and x and thanks to Assumption (A4), we have

$$\begin{aligned}
E_{Y,X}[U_{n,j}] &= E_{Y,X} \left[\frac{2}{nh} \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \right] \\
&= \frac{2}{nh} \sum_{i=1}^n E_{Y,X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i] \\
&= \frac{2}{nh} \sum_{i=1}^n E_X [E_{Y|X} [L_\omega(Y_i^*) | X_i] (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i] \\
&= \frac{1}{nh} \sum_{i=1}^n E_X \left[\varphi^{(1)} \left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X_i - x)^p \middle| X_i \right) \right. \\
&\quad \left. (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \right] \\
&= \frac{1}{nh} \sum_{i=1}^n E_X \left[\varphi^{(1)} \left(\frac{\tau_\omega^{(p+1)}(x)}{(p+1)!} (X_i - x)^{p+1} + \frac{\tau_\omega^{(p+2)}(x)}{(p+2)!} (X_i - x)^{p+2} \right. \right. \\
&\quad \left. \left. + \frac{\tau_\omega^{(p+3)}(\xi)}{(p+3)!} (X_i - x)^{p+3} \middle| X_i \right) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \right] \\
&= \frac{1}{h} \int \varphi^{(1)} \left(\frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x)(v-x)^{p+1} + \frac{\tau_\omega^{(p+2)}(x)}{(p+2)!} (v-x)^{p+2} (1+o(1)) \middle| X=v \right) \\
&\quad (\text{adj}(\mathbf{S})\mathbf{z})_{j+1} K \left(\frac{v-x}{h} \right) f_X(v) dv.
\end{aligned}$$

By a change of variable and with $\mathbf{z}_u = (1, u, u^2, \dots, u^p)^\top$ we have

$$\begin{aligned}
E_{Y,X}[U_{n,j}] &= \int \varphi^{(1)} \left(\frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x)(uh)^{p+1} + \frac{\tau_\omega^{(p+2)}(x)}{(p+2)!} (uh)^{p+2} (1+o(1)) \middle| X=x+uh \right) \\
&\quad (\text{adj}(\mathbf{S})\mathbf{z}_u)_{j+1} K(u) f_X(x+uh) du.
\end{aligned}$$

By Taylor expansion we find

$$\varphi^{(1)}(t|x) = \varphi^{(1)}(0|x) + \varphi^{(2)}(0|x)t(1+O(t)) = \varphi^{(2)}(0|x)t(1+O(t)) \quad \text{as } t \rightarrow 0$$

by the definition of $\tau_\omega(x)$. Then

$$\begin{aligned}
E_{Y,X}[U_{n,j}] &= \int \varphi^{(1)} \left(\frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x)(uh)^{p+1} + \frac{\tau_\omega^{(p+2)}(x)}{(p+2)!} (uh)^{p+2} (1+o(1)) \middle| X=x+uh \right) \\
&\quad (\text{adj}(\mathbf{S})\mathbf{z}_u)_{j+1} K(u) f_X(x+uh) du \\
&= \int \varphi^{(2)}(0|X=x+uh) \left(\frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x)(uh)^{p+1} + \frac{\tau_\omega^{(p+2)}(x)}{(p+2)!} (uh)^{p+2} (1+o(1)) \right) \\
&\quad (\text{adj}(\mathbf{S})\mathbf{z}_u)_{j+1} K(u) f_X(x+uh) du \\
&= \frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x) h^{p+1} \gamma(\omega, x) (\text{adj}(\mathbf{S})\mathbf{c}_p)_{j+1} f_X(x) \\
&\quad + \frac{1}{(p+2)!} \tau_\omega^{(p+2)}(x) h^{p+2} \gamma(\omega, x) (\text{adj}(\mathbf{S})\tilde{\mathbf{c}}_p)_{j+1} f_X(x) \\
&\quad + \frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x) h^{p+2} \gamma(\omega, x) (\text{adj}(\mathbf{S})\tilde{\mathbf{c}}_p)_{j+1} f_X^{(1)}(x) + o(h^{p+2}).
\end{aligned}$$

Remark S1 If K is symmetric, we can make a remark similar to Remark 2 regarding the structure of \mathbf{S} . Using the special structure of \mathbf{S} and of \mathbf{c}_p and $\tilde{\mathbf{c}}_p$, it is seen that some elements of $\text{adj}(\mathbf{S})\tilde{\mathbf{c}}_p$ and $\text{adj}(\mathbf{S})\mathbf{c}_p$ are equal to zero. Two cases are possible:

(i) $p - j$ odd

$$\mathbb{E}_{Y,X}[U_{n,j}] = \frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x) h^{p+1} \gamma(\omega, x) (\text{adj}(\mathbf{S})\mathbf{c}_p)_{j+1} f_X(x) + o(h^{p+2})$$

(ii) $p - j$ even

$$\begin{aligned} \mathbb{E}_{Y,X}[U_{n,j}] &= \frac{1}{(p+2)!} \tau_\omega^{(p+2)}(x) h^{p+2} \gamma(\omega, x) (\text{adj}(\mathbf{S})\tilde{\mathbf{c}}_p)_{j+1} f_X(x) \\ &\quad + \frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x) h^{p+2} \gamma(\omega, x) (\text{adj}(\mathbf{S})\tilde{\mathbf{c}}_p)_{j+1} f_X^{(1)}(x) + o(h^{p+2}). \end{aligned}$$

To obtain the conditional result, we proceed as follows. Suppose that the support of K is contained in $[-M, M]$. Then,

$$|(\text{adj}(\mathbf{S})\mathbf{Z})_{j+1}| K\left(\frac{X-x}{h}\right) \leq CK\left(\frac{X-x}{h}\right).$$

It follows, by the law of iterated expectation, that

$$\begin{aligned} \mathbb{E}_X \left[(\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}] - \mathbb{E}_{Y,X}[U_{n,j}])^2 \right] &= \mathbb{E}_X \left[(\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}] - \mathbb{E}_X[\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}]])^2 \right] \\ &= \text{Var}_X[\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}]] \\ &= \text{Var}_X \left[\mathbb{E}_{Y|X} \left[\frac{2}{nh} \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| \mathcal{X} \right] \right] \\ &= \frac{4}{n^2 h^2} \text{Var}_X \left[\sum_{i=1}^n \mathbb{E}_{Y|X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| \mathcal{X}] \right] \\ &= \frac{4}{n^2 h^2} \sum_{i=1}^n \text{Var}_X [\mathbb{E}_{Y|X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| \mathcal{X}]] \\ &\quad + \frac{4}{n^2 h^2} \sum_{m \neq n} \text{Cov}_X (\mathbb{E}_{Y|X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| X = X_m], \\ &\quad \quad \quad \mathbb{E}_{Y|X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| X = X_n]) \\ &= \frac{4}{n^2 h^2} \sum_{i=1}^n \text{Var}_X [\mathbb{E}_{Y|X} [L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \middle| \mathcal{X}]] \\ &= \frac{1}{nh^2} \text{Var}_X \left[\varphi^{(1)} \left(\tau_\omega(X) - \tau_\omega(x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X-x)^p \middle| X \right) \right. \\ &\quad \left. (\text{adj}(\mathbf{S})\mathbf{Z})_{j+1} K\left(\frac{X-x}{h}\right) \right] \\ &\leq \frac{C^2}{nh^2} \mathbb{E}_X \left[\left(\varphi^{(1)} \left(\frac{1}{(p+1)!} \tau_\omega^{(p+1)}(x) (X-x)^{p+1} + o_P(h^{p+1}) \middle| X \right) \right. \right. \\ &\quad \left. \left. K\left(\frac{X-x}{h}\right) \right)^2 \right] \\ &= \frac{C^2}{nh} \int (\varphi^{(2)}(0|X = x+uh) O(h^{p+1}) K(u))^2 f_X(x+uh) du \\ &= O\left(\frac{h^{2(p+1)}}{nh}\right). \end{aligned}$$

Two cases must be taken into account

(i) $p - j$ odd: with $h \rightarrow 0$ and $nh \rightarrow \infty$,

$$\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}] = \mathbb{E}_{Y,X}[U_{n,j}] + o_P(h^{p+1}) = dh^{p+1}(1 + o_P(1))$$

since

$$\frac{h^{p+\frac{1}{2}}}{\sqrt{nh^{p+1}}} = \frac{1}{\sqrt{nh}} \rightarrow 0 \quad \text{as } nh \rightarrow \infty.$$

(ii) $p - j$ even: with $h \rightarrow 0$ and we need another assumption for this case, $nh^3 \rightarrow \infty$,

$$\mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}] = \mathbb{E}_{Y,X}[U_{n,j}] + o_P(h^{p+2}) = dh^{p+1}(1 + o_P(1))$$

since

$$\frac{h^{p+\frac{1}{2}}}{\sqrt{nh^{p+2}}} = \frac{1}{\sqrt{nh^3}} \rightarrow 0 \quad \text{as } nh^3 \rightarrow \infty.$$

Therefore, the first assertion is established.

For the second assertion we write

$$\begin{aligned} \text{Var}_{Y|X}[U_{n,j}|\mathcal{X}] &= \text{Var}_{Y|X} \left[\frac{2}{nh} \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i | \mathcal{X} \right] \\ &= \frac{1}{n^2 h^2} \text{Var}_{Y|X} \left[2 \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i | \mathcal{X} \right] \\ &= \frac{1}{n^2 h^2} \left(\mathbb{E}_{Y|X} \left[\left(2 \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i \right)^2 | \mathcal{X} \right] \right. \\ &\quad \left. - \mathbb{E}_{Y|X} \left[2 \sum_{i=1}^n L_\omega(Y_i^*) (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i | \mathcal{X} \right]^2 \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1}^2 K^2 \left(\frac{X_i - x}{h} \right) \mathbb{E}_{Y|X} [4L_\omega^2(Y_i^*) | \mathcal{X}] + o_P \left(\frac{1}{nh} \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1}^2 K^2 \left(\frac{X_i - x}{h} \right) \\ &\quad \mathbb{E}_{Y|X} \left[4L_\omega^2(Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X_i - x)^p) | \mathcal{X} \right] + o_P \left(\frac{1}{nh} \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n Z_{n,i} + o_P \left(\frac{1}{nh} \right), \end{aligned}$$

where

$$Z_{n,i} = (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1}^2 K^2 \left(\frac{X_i - x}{h} \right) \mathbb{E}_{Y|X} \left[4L_\omega^2(Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X_i - x)^p) | \mathcal{X} \right].$$

It follows from the conditional Jensen inequality and Assumption (A3) that

$$\begin{aligned} &\mathbb{E}_X [|Z_{n,i}|^{1+\delta/2}] \\ &\leq C^{2+\delta} \mathbb{E}_{X,Y} \left[\left| 2L_\omega(Y_1 - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_1 - x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X_1 - x)^p) K \left(\frac{X_1 - x}{h} \right) \right|^{2+\delta} \right] \\ &\leq C^{2+\delta} h \int \int \left| 2L_\omega(y - \tau_\omega(x) - \tau_\omega^{(1)}(x)hv - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (hv)^p) \right|^{2+\delta} \\ &\quad f_{Y|X}(y|x+hv) dy K^{2+\delta}(v) f_X(x+hv) dv \\ &= O(h). \end{aligned}$$

By the moment inequality for the sum of independent random variables we have

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n Z_{n,i} &= \frac{\mathbb{E}_X[Z_{n,i}]}{h} + o_P(1) \\ &= f_X(x) \int (2L_\omega(y - \tau_\omega(x)))^2 f_{Y|X}(y|x) dy \int (\text{adj}(\mathbf{S})\mathbf{z}_v)_{j+1}^2 K^2(v) dv + o_P(1) \end{aligned}$$

and therefore

$$\text{Var}_{Y|X}[U_{n,j}|\mathcal{X}] = \frac{1}{nh} f_X(x) \int (2L_\omega(y - \tau_\omega(x)))^2 f_{Y|X}(y|x) dy \int (\text{adj}(\mathbf{S})\mathbf{z}_v)_{j+1}^2 K^2(v) dv (1 + o_P(1)).$$

□

S6.2 Proof of Lemma 2

Denote

$$Z_{n,i}^* = (\text{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K \left(\frac{X_i - x}{h} \right) 2L_\omega \left(Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \dots - \frac{\tau_\omega^{(p)}(x)}{p!} (X_i - x)^p \right).$$

Then

$$U_{n,j} - \mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}] = \frac{1}{nh} \sum_{i=1}^n [Z_{n,i}^* - \mathbb{E}_{Y|X}[Z_{n,i}^*|X_i]].$$

Note that given \mathcal{X} , the previous equation is a sum of independent random variables. To show that

$$P \left[\frac{U_{n,j} - \mathbb{E}_{Y|X}[U_{n,j}|\mathcal{X}]}{\sqrt{\text{Var}_{Y|X}[U_{n,j}|\mathcal{X}]}} \leq t | \mathcal{X} \right] = \Phi(t) + o_p(1) \quad , \text{ as } n \rightarrow \infty,$$

it suffices to verify the conditional Lyapunov condition:

$$\frac{1}{(nh)^{2+\delta} [\text{Var}_{Y|X}[U_{n,j}|\mathcal{X}]]^{1+\delta/2}} \sum_{i=1}^n \mathbb{E}_{Y|X}[|Z_{n,i}^*|^{2+\delta} | X_i] = o_P(1)$$

for some $\delta > 0$. By Lemma 1

$$\mathbb{E}_X \left[\sum_{i=1}^n \mathbb{E}_{Y|X}[|Z_{n,i}^*|^{2+\delta} | X_i] \right] = n \mathbb{E}_X[|Z_{n,i}^*|^{2+\delta}] = O(nh).$$

Thus

$$\sum_{i=1}^n \mathbb{E}_{Y|X}[|Z_{n,i}^*|^{2+\delta} | X_i] = O_P(nh).$$

This and the second assertion of Lemma 1 imply the conditional Lyapunov condition. □

References

- Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge.
- Schmidt, G., Mattern, R., and Schüller, F. (1981). Biomechanical investigation to determine physical and traumatological differentiation criteria for the maximum load capacity of head and vertebral column with and without protective helmet under the effects of impact. Technical report, Institut für Rechtsmedizin, Universität Heidelberg.
- Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, 6(2):273–292.