Annals of the Institute of Statistical Mathematics (2022) 74:341–378
https://doi.org/10.1007/s10463-021-00799-y

Local polynomial expectile regression

C. Adam[1] · **I. Gijbels**[1]

Received: 22 May 2020 / Revised: 18 February 2021 / Accepted: 31 March 2021 /
Published online: 5 May 2021

Abstract
This paper studies local polynomial estimation of expectile regression. Expectiles and quantiles both provide a full characterization of a (conditional) distribution function, but have each their own merits and inconveniences. Local polynomial fitting as a smoothing technique has a major advantage of being simple, allowing for explicit expressions and henceforth advantages when doing inference theory. The aim of this paper is twofold: to study in detail the use of local polynomial fitting in the context of expectile regression and to contribute to the important issue of bandwidth selection, from theoretical and practical points of view. We discuss local polynomial expectile regression estimators and establish an asymptotic normality result for them. The finite-sample performance of the estimators, combined with various bandwidth selectors, is investigated in a simulation study. Some illustrations with real data examples are given.

Keywords Asymptotic normality · Bandwidth selection · Expectile regression · Local polynomial fitting · Quantile regression

1 Introduction

Among the main interests in regression analysis is to explore the influence that d covariates $\mathbf{X} = (X_1, \ldots, X_d)$ have on a variable of interest Y, the response. There is an extensive literature on flexible mean regression, in which the targeted quantity is the conditional mean of the response given the covariates, i.e. $\mathrm{E}[Y|\mathbf{X}]$.

In a nonparametric regression model, no assumptions are made on the form of the relation between the covariates and the response. See e.g. the books of Härdle

✉ I. Gijbels
irene.gijbels@kuleuven.be

C. Adam
cecile.adam@kuleuven.be

[1] Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven, Celestijnenlaan 200B, Box 2400, 3001 Leuven, Heverlee, Belgium

(1990) and Wand and Jones (1995). For nonparametric mean regression, Fan and Gijbels (1995) developed adaptive-order local polynomial fitting. Fan and Gijbels (1996) provided an extensive study of the local polynomial modelling technique and its applications to various areas. Local polynomial fitting is a popular smoothing technique due to its simplicity, the ease of computation and its nice asymptotic properties.

The conditional mean (regression), however, describes only the average effect of the response Y given the covariates $\mathbf{X}$ and if the conditional distribution of the response is skewed, it is not appropriate to describe it only via its conditional mean. Therefore, a more complete characterization of the conditional distribution of Y given the covariate vector is preferred. Quantile regression aims at estimating the conditional median or other quantiles of the response variable given the covariates and takes into account the mentioned drawback of conditional mean regression. There is a vast literature on conditional quantile regression, e.g. Koenker and Bassett (1978) for a study of the asymmetric least absolute deviation approach, and Koenker (2005) for a book dedicated to quantile regression.

An alternative to quantiles are expectiles, who also fully characterize (conditional) distributions. Newey and Powell (1987) introduced conditional expectiles in the context of linear regression models. They proposed an estimator by using an asymmetric least squares approach and showed that this estimator is asymptotically normal distributed.

Both quantiles and expectiles have some advantages and inconveniences. Quantiles and expectiles provide a complete characterization of the (conditional) distribution. A main advantage of quantiles is their appealing interpretability, whereas expectiles have a less natural interpretation. Theoretically, neither expectile nor quantile curves, for different orders of the quantiles/expectiles, can cross. However, the problem of crossing of estimated quantile curves is well known. This crossing problem is less frequently observed for estimated expectile curves. Quantiles do not need to be unique, whereas expectiles are always uniquely defined for any (conditional) distribution with a finite mean. Quantiles are robust to outliers due to the asymmetric absolute deviation loss function. In contrast, expectiles are sensitive to extreme small or large observations, given they are based on an asymmetric quadratic type of loss function. Obviously a certain sensitivity to extremal observations is preferable when it comes to risk management (e.g. in financial applications). Finally, expectiles are easier to compute than quantiles. A detailed discussion on quantile and expectile regression can be found in Schulze Waltrup et al. (2015).

Since the first papers by Newey and Powell (1987) and Efron (1991), there is an increasing interest in expectiles due to, among others, their properties as risk measures. Breckling and Chambers (1988) showed that quantiles and expectiles both belong to the class of so-called M-quantiles and later on Bellini et al. (2014) showed that the only M-quantiles that lead to coherent risk measures are expectiles. Moreover, Ziegel (2016) argued that expectiles induce the only coherent and elicitable law-invariant risk measure (i.e. with meaningful point forecasts performance). Henceforth, although expectiles are not as easy to interpret as quantiles, they have interesting properties, in particular as risk measure. Recent papers on expectiles

include Taylor (2008), Schulze Waltrup et al. (2015), Bellini and Di Bernardino (2017) and Krätschmer and Zähle (2017).

A linear model for expectile regression, as studied by Newey and Powell (1987) and Efron (1991), is often too restrictive in view of real applications. Yao and Tong (1996) studied nonparametric expectile regression in case of a univariate explanatory variable and allowing for observations that are strictly stationary and ρ-mixing. They considered a local linear estimator for regression expectiles and established its asymptotic normality. Schnabel and Eilers (2009) contributed in the issue of smoothing parameter selection by defining an asymmetric variant of cross-validation. Yang and Zou (2015) used the gradient tree boosting algorithm to derive a fully nonparametric multiple expectile regression method. Little attention so far went to good bandwidth selectors for expectile regression, based on theoretical considerations. The aim of this paper is to provide a detailed study of the use of local polynomial fitting to estimate nonparametrically a univariate expectile regression curve as well as its derivatives, including (i) asymptotic bias and variance expressions for estimators of the expectile curve and its derivatives (up to a certain order), and (ii) a discussion on theoretical optimal bandwidths as well as selection procedures for data-driven bandwidths.

The paper is organized as follows. Section 2 briefly recalls the definition of expectiles and quantiles and the existing links between the two concepts. In Sect. 3, we present the local polynomial expectile regression method, and in Sect. 4 we establish its asymptotic normality. Section 5 is devoted to the bandwidth selection issue and provides an explicit expression for the optimal bandwidth but also presents several data-driven bandwidth selectors. The performances of these are investigated in a simulations study in Sect. 6. The practical use of the methods is illustrated on a real data example in Sect. 7. We conclude with some further discussions and recommendations in Sect. 8. The proof of the main theorem is given in Appendix. Proofs of other theoretical results are provided in the Supplementary Material. Some more material, including discussion on some additional simulation results and two more real data applications can also be found in the Supplemental Material part.

2 Expectiles and quantiles

Consider first the unconditional setting with focus on a real-valued variable Y. The unconditional αth quantile, with $\alpha \in (0, 1)$, of Y is $q_\alpha := \inf_t\{t \in \mathbb{R} : P(Y \leq t) \geq \alpha\}$, which can also be obtained, in case of uniqueness, by solving the L_1-minimization problem $q_\alpha = \arg\min_{\theta\in\mathbb{R}} \mathrm{E}_Y[R_\alpha(Y - \theta)]$ with R_α the quantile check function

$$R_\alpha(y) = |\alpha - \mathbb{1}\{y \leq 0\}||y|. \tag{1}$$

With $\alpha = 0.5$, we obtain $q_{0.5}$ the (unconditional) median of Y. Figure 1 depicts the quantile check function for $\alpha = 0.4, 0.5$ and 0.6.

The unconditional ωth expectile, with $\omega \in (0, 1)$, of Y is the solution to the L_2-minimization problem

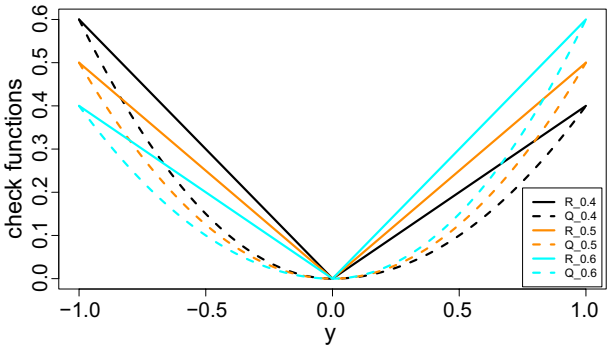


Fig. 1 Quantile check functions (solid lines) and expectile loss functions (dashed curves) for $\{\alpha, \omega\} \in \{0.4, 0.5, 0.6\}$

$$\tau_\omega = \arg\min_{\theta \in \mathbb{R}} \mathrm{E}_Y[Q_\omega(Y - \theta)]$$

with Q_ω the expectile loss function

$$Q_\omega(y) = |\omega - \mathbb{1}\{y \le 0\}| y^2. \tag{2}$$

With $\omega = 0.5$, the quantity $\tau_{0.5}$ equals $\mathrm{E}[Y]$ the (unconditional) mean of Y. For an overview of basic properties of expectiles, see e.g. Newey and Powell (1987). We, in particular, mention the following basic property. Let $\widetilde{Y} = a + bY$, with $a, b \in \mathbb{R}$, then the ωth expectile of $\widetilde{Y}$, denoted by $\tau_{\omega,\widetilde{Y}}$, is given by

$$\tau_{\omega,\widetilde{Y}} = \begin{cases} a + b\,\tau_{\omega,Y} & \text{if } \ b > 0 \\ a + b\,\tau_{1-\omega,Y} & \text{if } \ b \le 0, \end{cases} \tag{3}$$

where we denote the ωth expectile of Y by $\tau_{\omega,Y}$ to stress that it is the expectile of Y. See, for example, Newey and Powell (1987, Theorem 1, page 823) and Remillard and Abdous (1995, Theorem 1).

Figure 1 shows the expectile loss function $Q_\omega(\cdot)$ for $\omega = 0.4, 0.5$ and 0.6. Note that the check loss function $R_\alpha(\cdot)$ is not differentiable in 0, whereas the expectile loss function $Q_\omega(\cdot)$ is continuously differentiable everywhere.

Both quantiles and expectiles characterize a distribution function, although they are different in nature. Figure 2 shows the quantile curve (solid line) and the expectile curve (dotted line) of a Student-t distribution with 5 degrees of freedom. For this symmetric distribution, the mean and the median coincide as is observed on this figure at the location $\alpha = \omega = 0.5$.

In terms of interpretation, the αth quantile determines the point below which $100 \times \alpha\%$ of the mass of Y lies, i.e. $\alpha = \mathrm{E}_Y\big[\mathbb{1}\{Y \le q_\alpha\}\big]/\mathrm{E}_Y[1]$, while the ωth expectile specifies the position τ_ω such that the average distance from Y to τ_ω, when Y is below τ_ω, is $100 \times \omega\%$ of the average distance between Y and τ_ω, i.e. $\omega = \mathrm{E}_Y\big[|Y - \tau_\omega| \mathbb{1}\{Y \le \tau_\omega\}\big]/\mathrm{E}_Y[|Y - \tau_\omega|]$.

Consider (X, Y) a bivariate random vector. The concepts of quantiles and expectiles are easily extended to the conditional case, in which Y is the variable of

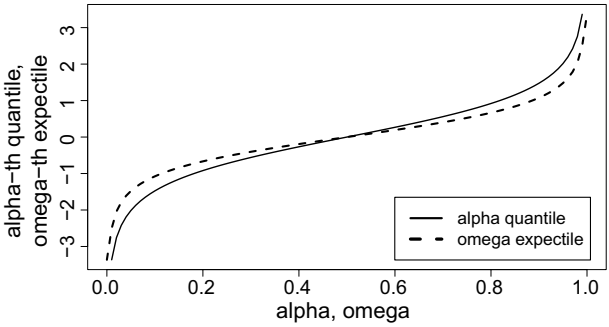


Fig. 2 Expectile function (dashed curve) and quantile function (solid curve) of a Student-t distribution with 5 degrees of freedom

interest and X a covariate. The αth conditional quantile of Y given $X = x$ is, in case of uniqueness, $q_\alpha(x) = \arg\min_{a\in\mathbb{R}} \mathrm{E}_{Y|X}[R_\alpha(Y-a)|X=x]$, with R_α the check function (see (1)). The ωth conditional expectile of Y given $X = x$, with $\omega \in (0,1)$, is

$$\tau_\omega(x) = \arg\min_{a\in\mathbb{R}} \mathrm{E}_{Y|X}\left[Q_\omega(Y-a)|X=x\right] \tag{4}$$

with Q_ω the loss function in (2). Since $Q_\omega(\cdot)$ has a first continuous derivative, $\tau_\omega(x)$ satisfies

$$\mathrm{E}_{Y|X}\left[L_\omega(Y-\tau_\omega(x))|X=x\right] = 0$$

with $L_\omega(y) = |\omega - \mathbb{1}\{y \le 0\}|y$.

Similar interpretations as in the unconditional setting hold. The αth conditional quantile determines, given $X = x$, the point below which $100\times\alpha\%$ of the mass of Y lies,

$$\alpha = \frac{\mathrm{E}_{Y|X}\left[\mathbb{1}\{Y \le q_\alpha(x)\}|X=x\right]}{\mathrm{E}_{Y|X}[1|X=x]},$$

whereas the ωth conditional expectile specifies, given $X = x$, the position $\tau_\omega(x)$ such that the average distance of Y to $\tau_\omega(x)$, when Y is below $\tau_\omega(x)$, is $100\times\omega\%$ of the average distance between Y and $\tau_\omega(x)$, i.e.

$$\omega = \frac{\mathrm{E}_{Y|X}\left[|Y-\tau_\omega(x)|\mathbb{1}\{Y \le \tau_\omega(x)\}|X=x\right]}{\mathrm{E}_{Y|X}[|Y-\tau_\omega(x)|\ |\ X=x]}. \tag{5}$$

Jones (1994) pointed out that expectiles are quantiles, not of the distribution function F_Y of Y itself but of another distribution function that is related to F_Y. Yao and Tong (1996) formulated this in an alternative way, showing that there is a one-to-one mapping (in fact a bijection) between the (conditional) quantile and the (conditional) expectile. See also De Rossi and Harvey (2009), Schulze Waltrup et al. (2015) and Yang and Zou (2015), among others. Since the existence of such a one-to-one

mapping is of crucial importance, we precise this further. Denote by F_X the cumulative distribution function of X, and with $F_{Y|X}$ the cumulative conditional distribution function of Y given X. For simplicity, we assume throughout this paper that $F_{Y|X}(y|x)$ is a continuous function in y, for all $x \in \mathbb{R}$. The link between conditional quantiles and expectiles is formally stated in Proposition 1. Since a formal statement with proof of this result does not seem to be available in the literature, we also, for completeness, provide a proof for it in Section S2 in the Supplementary Material.

Proposition 1 (One-to-one mapping) *Let x be a point of continuity of F_X, for which $|\mathrm{E}_{Y|X}(Y|X=x)| < \infty$. Then, there exists a one-to-one mapping (a bijection) $(0,1) \to (0,1) : \alpha \mapsto \omega = \omega(\alpha, x)$ such that $\tau_{\omega(\alpha,x)}(x) = q_\alpha(x)$, i.e. the $\omega(\alpha,x)$th conditional expectile equals the αth conditional quantile. Specifically, we have*

$$\omega(\alpha,x) = \frac{\alpha q_\alpha(x) - \int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x)}{2\left[\alpha q_\alpha(x) - \int_{-\infty}^{q_\alpha(x)} y dF_{Y|X}(y|x)\right] + \left[\mathrm{E}_{Y|X}(Y|X=x) - q_\alpha(x)\right]}. \tag{6}$$

The relation in (6) can be further simplified in case of a location-scale model in which

$$Y = m(X) + \sigma(X)\epsilon, \tag{7}$$

where $m(.)$ and $\sigma(.)$ are unknown functions, with $\sigma(x) > 0$ and ϵ has a continuous strictly increasing distribution function F_ϵ and quantile function F_ϵ^{-1}. Further X and ϵ are independent, with $\mathrm{E}[\epsilon] = 0$ and $\mathrm{Var}(\epsilon) = 1$. Under a location-scale model (7) the αth conditional quantile of Y given $X = x$, for $\alpha \in (0,1)$, is

$$\begin{aligned} q_\alpha(x) = \inf_y\{y : F_{Y|X}(y|x) \geq \alpha\} &= \inf_y \left\{ y : F_\epsilon\left(\frac{y - m(x)}{\sigma(x)}\right) \geq \alpha \right\} \\ &= m(x) + \sigma(x) F_\epsilon^{-1}(\alpha). \end{aligned} \tag{8}$$

Moreover, using (4) and a conditional version of (3) it is easily seen that the ωth conditional expectile of Y given $X = x$ equals (since $\sigma(x) > 0$)

$$\tau_\omega(x) = m(x) + \sigma(x)\tau_{\omega,\epsilon}, \tag{9}$$

where $\tau_{\omega,\epsilon}$ denotes the (unconditional) ωth expectile of the random variable ϵ. In case of a homoscedastic location-scale model, i.e. when $Y = m(X) + \sigma\epsilon$, equation (8) becomes $q_\alpha(x) = m(x) + \sigma F_\epsilon^{-1}(\alpha)$ and hence, for α_1 and $\alpha_2 \in (0,1)$ we have $q_{\alpha_1}(x) - q_{\alpha_2}(x) = \sigma\left(F_\epsilon^{-1}(\alpha_1) - F_\epsilon^{-1}(\alpha_2)\right)$ which is constant for all x. Hence, in a homoscedastic location-scale model, the quantile curves are parallel. From (9), it is seen that the same holds for the expectile curves in a homoscedastic location-scale model setting.

Furthermore, in case of a location-scale model (7) Yao and Tong (1996, Proposition 1) establish that the one-to-one mapping in (6) is independent of x, and reduces to

$$\omega(\alpha, x) = \omega(\alpha) = \frac{\alpha F_\epsilon^{-1}(\alpha) - \mathrm{E}_\epsilon\left[\epsilon\, \mathbb{1}\{\epsilon \le F_\epsilon^{-1}(\alpha)\}\right]}{2\mathrm{E}_\epsilon\left[\epsilon\, \mathbb{1}\{\epsilon > F_\epsilon^{-1}(\alpha)\}\right] - (1-2\alpha)F_\epsilon^{-1}(\alpha)}. \tag{10}$$

An illustration of relationship (10) is provided in Figure S.1 in the Supplementary Material.

3 Local polynomial expectile regression

3.1 Local polynomial expectile regression estimator

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample from (X, Y). The aim is to estimate the ωth conditional expectile $\tau_\omega(x)$, defined in (4), as well as derivatives of this function.

To estimate $\tau_\omega(\cdot)$ we use local polynomial fitting (see Fan and Gijbels 1996). Consider x a fixed value in the domain of the covariate X. Assume that the unknown function $\tau_\omega(\cdot)$ can be approximated by a polynomial function of degree p in a neighbourhood of x via a Taylor expansion up to order p, i.e. for z in a neighbourhood of x,

$$\tau_\omega(z) \approx \tau_\omega(x) + \tau_\omega^{(1)}(x)(z-x) + \cdots + \frac{\tau_\omega^{(p)}(x)}{p!}(z-x)^p \equiv \beta_0 + \beta_1(z-x) + \cdots + \beta_p(z-x)^p$$

where $\tau_\omega^{(j)}(x)$ denotes the jth order derivative of the function $\tau_\omega(.)$ at the point x, and we denoted $\beta_j = \frac{1}{j!}\tau_\omega^{(j)}(x)$ for $j = 0, \ldots, p$. Obviously, we need the existence of derivatives up to order p of the expectile function $\tau_\omega(\cdot)$.

We consider the minimization problem

$$\underset{\beta_0,\ldots,\beta_p}{\text{minimize}} \sum_{i=1}^{n} Q_\omega\left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j\right) K\left(\frac{X_i - x}{h}\right), \tag{11}$$

with $Q_\omega(\cdot)$ as in (2) and where $K(\cdot)$ and $h > 0$ denote, respectively, a kernel function and a bandwidth. We denote the solution of (11) by $(\widehat{\beta}_0, \ldots, \widehat{\beta}_p)$. The estimator of $\tau_\omega(x)$ is then $\widehat{\beta}_0$ and the estimator of $\tau_\omega^{(j)}(x) = \frac{d^j \tau_\omega(x)}{\mathrm{d}x^j}$ is $\widehat{\beta}_j j!$, for $j = 0, \ldots, p$. The choice of an appropriate bandwidth is important and studied in Sect. 5.

It is convenient to express the minimization problem in (11) in matrix notation, starting from

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X}_D = \begin{pmatrix} 1 & X_1 - x & (X_1 - x)^2 & \cdots & (X_1 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & (X_n - x)^2 & \cdots & (X_n - x)^p \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix},$$

$$\mathbf{H} = \mathrm{diag}\left(K\left(\frac{X_1 - x}{h}\right), \ldots, K\left(\frac{X_n - x}{h}\right)\right)$$

and $\mathbf{W} = \mathrm{diag}\,(r_1(\omega), \ldots, r_n(\omega))$ is a diagonal matrix with as ith diagonal element the weight $r_i(\omega)$ defined by $r_i(\omega) = (1-\omega)\mathbb{1}\{Y_i \le \sum_{j=0}^{p} \beta_j (X_i - x)^j\} + \omega \mathbb{1}\{Y_i > \sum_{j=0}^{p} \beta_j (X_i - x)^j\}$.

Hence, the diagonal matrix is composed of elements from the set $\{\omega, 1-\omega\}$, for a given $\omega \in (0,1)$, according to whether the observation Y_i is located above or below the polynomial function $\sum_{j=0}^{p} \beta_j (X_i - x)^j$. Note that the design matrix $\mathbf{X}_D$ and the matrices $\mathbf{H}$ and $\mathbf{W}$ in fact depend on the given value x, as also minimization problem (11). So estimation of the entire function $\tau_\omega(\cdot)$ requires solving (11) for a grid of points in the domain of X. Using the matrix notations, minimization problem (11) reads as

$$\underset{\boldsymbol{\beta}}{\text{minimize}}\ (\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta})^{\mathrm{T}}\, \mathbf{W}\mathbf{H}(\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta}). \tag{12}$$

Denote the minimizer of (12) by $\widehat{\boldsymbol{\beta}}$. The estimator $\widehat{\boldsymbol{\beta}}_0$ is the local polynomial expectile regression estimator of $\tau_\omega(x)$.

3.2 Iterative procedure

Since $r_i(\omega)$ and hence $\mathbf{W}$ depend on $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^{\mathrm{T}}$ an iterative procedure is needed to find the estimators of $\boldsymbol{\beta}$. Suppose that at step t of the iteration, we have given a value $\boldsymbol{\beta}^{(t)}$ for the vector of unknown parameters. Denote the corresponding diagonal matrix with $\mathbf{W}^{(t)}$, i.e. the diagonal matrix with as ith element the weight $r_i^{(t)}(\omega)$, given by

$$r_i^{(t)}(\omega) = \begin{cases} 1-\omega & \text{if } Y_i \le \displaystyle\sum_{j=0}^{p} \beta_j^{(t)} (X_i - x)^j \\ \omega & \text{if } Y_i > \displaystyle\sum_{j=0}^{p} \beta_j^{(t)} (X_i - x)^j. \end{cases}$$

We then need to find an improvement of the current parameter vector value $\boldsymbol{\beta}^{(t)}$ by exploiting minimization problem (12), and minimizing

$$(\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta})^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}(\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta}) \tag{13}$$

with respect to $\boldsymbol{\beta}$. Writing

$$(\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta})^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}(\mathbf{Y} - \mathbf{X}_D\boldsymbol{\beta}) = \mathbf{Y}^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{Y} - 2\boldsymbol{\beta}^T\mathbf{X}_D^T\mathbf{W}^{(t)}\mathbf{H}\mathbf{Y} + \boldsymbol{\beta}^{\mathrm{T}}\, \mathbf{X}_D^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{X}_D\boldsymbol{\beta}, \tag{14}$$

where we use the fact that the transpose of a scalar is a scalar, i.e.

$$\boldsymbol{\beta}^{\mathrm{T}}\, \mathbf{X}_D^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{Y} = (\boldsymbol{\beta}^{\mathrm{T}}\, \mathbf{X}_D^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{Y})^{\mathrm{T}} = \mathbf{Y}^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{X}_D\boldsymbol{\beta}.$$

To find the improved vector value we differentiate (14) with respect to $\boldsymbol{\beta}$, and need to solve

$$-2\mathbf{X}_D^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{Y} + 2\mathbf{X}_D^{\mathrm{T}}\, \mathbf{W}^{(t)}\mathbf{H}\mathbf{X}_D\boldsymbol{\beta} = 0, \tag{15}$$

which leads to the solution $\boldsymbol{\beta}^{(t+1)} = \left(\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D\right)^{-1} \mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{Y}$, provided that the inverse of the matrix $\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D$ exists. This solution is indeed a minimizer of (13) since the matrix of second order partial derivatives of it equals $2\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D$ obtained from (15), which is a positive definite matrix with high probability. We thus have that

$$\boldsymbol{\beta}^{(t+1)} = (\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D)^{-1} \mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{Y}. \tag{16}$$

The **iterative** procedure reads as follows.

INITIALIZATION STEP Obtain $\boldsymbol{\beta}^{(0)}$ the vector of least squares estimators

$$\boldsymbol{\beta}^{(0)} = \underset{\beta_0,\ldots,\beta_p}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2. \tag{17}$$

ITERATION STEPS For $t = 0, 1, \ldots$, obtain $\boldsymbol{\beta}^{(t+1)}$ from (16). Continue the iteration until convergence. Denote the value of $\boldsymbol{\beta}^{(t+1)}$ after convergence by $\widehat{\boldsymbol{\beta}}^{(\infty)}$.

The estimator $\widehat{\boldsymbol{\beta}}_0^{(\infty)}$ is an approximation of the local polynomial expectile regression estimator $\widehat{\boldsymbol{\beta}}_0$ of $\tau_\omega(x)$, that is obtained via the iterative procedure.

A crucial quantity in (16) is the matrix $\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D$ which we denote as $\mathbf{S}_n^{(t)}$. Denote the unit vector $\mathbf{e}_{j+1} = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathrm{T}}$, the $(p+1) \times 1$ vector with 1 on the $(j+1)$st position, and zero's everywhere else. At the iteration step t, the estimator $\widehat{\beta}_j^{(t+1)}$ of β_j is then given by

$$\boldsymbol{\beta}_j^{(t+1)} = \mathbf{e}_{j+1}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}^{(t+1)} = \mathbf{e}_{j+1}^{\mathrm{T}} \left(\mathbf{S}_n^{(t)}\right)^{-1} \mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{Y}$$

with $j = 0, \ldots, p$ and where

$$\mathbf{S}_n^{(t)} = \mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{X}_D = \begin{pmatrix} S_{n,0}^{(t)} & S_{n,1}^{(t)} & \cdots & S_{n,p}^{(t)} \\ S_{n,1}^{(t)} & S_{n,2}^{(t)} & \cdots & S_{n,p+1}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p}^{(t)} & S_{n,p+1}^{(t)} & \cdots & S_{n,2p}^{(t)} \end{pmatrix}$$

is a $(p+1) \times (p+1)$ matrix with

$$S_{n,j}^{(t)} = \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{h}\right) (X_i - x)^j \quad \text{and} \quad j = 0, \ldots, p.$$

Example 1 (*Local linear case.*)
If $p = 1$, we have to minimize

$$\sum_{i=1}^{n} Q_\omega\big(Y_i - \beta_0 - \beta_1(X_i - x)\big) K\left(\frac{X_i - x}{h}\right)$$

with respect to β_0 and β_1. In this case we have that

$$\mathbf{S}_n^{(t)} = \begin{pmatrix} S_{n,0}^{(t)} & S_{n,1}^{(t)} \\ S_{n,1}^{(t)} & S_{n,2}^{(t)} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{x}\right) & \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{x}\right)(X_i - x) \\ \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{x}\right)(X_i - x) & \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{x}\right)(X_i - x)^2 \end{pmatrix},$$

$$\mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{h}\right) Y_i \\ \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{h}\right)(X_i - x) Y_i \end{pmatrix} = \begin{pmatrix} T_{n,0}^{(t)} \\ T_{n,1}^{(t)} \end{pmatrix},$$

with $T_{n,\ell}^{(t)} = \sum_{i=1}^{n} r_i^{(t)}(\omega) K\left(\frac{X_i - x}{h}\right)(X_i - x)^\ell Y_i$ for $\ell = 0, 1$.

Therefore, the estimators $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1)^{\mathrm{T}}$ can be computed by resolving iteratively

$$\boldsymbol{\beta}^{(t+1)} = \big(\mathbf{S}_n^{(t)}\big)^{-1} \mathbf{X}_D^{\mathrm{T}} \mathbf{W}^{(t)} \mathbf{H} \mathbf{Y} = \frac{1}{S_{n,0}^{(t)} S_{n,2}^{(t)} - \left(S_{n,1}^{(t)}\right)^2} \begin{pmatrix} S_{n,2}^{(t)} T_{n,0}^{(t)} - S_{n,1}^{(t)} T_{n,1}^{(t)} \\ S_{n,0}^{(t)} T_{n,1}^{(t)} - S_{n,1}^{(t)} T_{n,0}^{(t)} \end{pmatrix},$$

which corresponds to the expression for the (approximate) estimators given by Yao and Tong (1996).

Remark 1 Since the minimizer of (12) is obtained approximatively via the iterative procedure explained above, this raises two important question: (i) does the iterative procedure converge?; (ii) when it converges, to $\widehat{\boldsymbol{\beta}}^{(\infty)}$ say, is then $\widehat{\boldsymbol{\beta}}^{(\infty)}$ equal to $\widehat{\boldsymbol{\beta}}$ the minimizer of (12)? The answer to both questions is affirmative. The answer to the second question follows from the fact that the function $Q_\omega(\cdot)$ defined in (2) is a convex function. Indeed, due to the convexity of the expectile loss function, both problems, the optimization problem (12) and the solution to equation (15), after convergence, lead to the same unique quantity, i.e. $\widehat{\boldsymbol{\beta}}^{(\infty)} = \widehat{\boldsymbol{\beta}}$. Regarding the first question, we would like to remind that the iterative procedure behind (local polynomial) expectile estimation is an iterative reweighted least squares type of algorithm. For such algorithms, the convergence has been studied. See, for example, Huber and Ronchetti (2009, Section 7.8.3). See also, among others, Wolke and Schwetlick (1988) for a convergence analysis regarding iteratively reweighted least squares algorithms involving convex criterion functions.

3.3 Practical implementation issues

When using iterative procedures in practice, some stopping rule is needed. Since we know from Remark 1 that the iterative algorithm converges, one could set a

maximum number of iterations and stop the iteration process when this maximum number of iterations is reached. Another more interesting approach is to quantify the difference between the estimator at two consecutive iteration steps, i.e. by evaluating the difference between $\boldsymbol{\beta}^{(t+1)}$ and $\boldsymbol{\beta}^{(t)}$. In our practical implementation, when focusing on estimation of $\tau_\omega(x)$, we used as a stopping criterion

$$\left|\boldsymbol{\beta}_0^{(t+1)} - \boldsymbol{\beta}_0^{(t)}\right| < 10^{-6}, \tag{18}$$

which was inspired by the computer calculation precision.

The iterative procedure described in Sect. 3.2 requires a starting vector $\boldsymbol{\beta}^{(0)}$. One possibility is to use the least squares polynomial regression estimator (17) as a starting vector. Other options to be used as starting vectors include a polynomial median regression estimator or a polynomial ωth quantile regression estimator, which are obtained by replacing in (17), the squared loss by the appropriate check loss functions $R_{0.5}(\cdot)$ or $R_\omega(\cdot)$, respectively. In a simulation study in Section S3.5 (in the Supplementary Material), we investigated the impact of these different choices of starting vector $\boldsymbol{\beta}^{(0)}$. In general, it seems that the choice of starting point has very little influence. We therefore opted for the simple choice in (17), for which an analytical expression for $\boldsymbol{\beta}^{(0)}$ is available.

In our simulation study, we found that with the stopping rule as in (18), mostly only two iterations where needed, and this no matter which method was used to choose the starting point $\boldsymbol{\beta}^{(0)}$. See Section S3.5.

4 Asymptotic results

In this section, we establish an asymptotic normality result for the local polynomial expectile regression estimators of $\tau_\omega(x)$ and its derivatives up to order p. Before stating the assumptions we introduce some notations. The moments of the kernel K and its square K^2 are denoted by, respectively,

$$\mu_j = \int u^j K(u)\mathrm{d}u \quad \text{and} \quad \nu_j = \int u^j K^2(u)\mathrm{d}u \quad \text{with} \quad j = 0, 1, \ldots, 2p.$$

Further we denote

$$\mathbf{S} = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix}, \qquad \widetilde{\mathbf{S}} = \begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \mu_2 & \mu_3 & \cdots & \mu_{p+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{p+1} & \mu_{p+2} & \cdots & \mu_{2p+1} \end{pmatrix},$$

$$\mathbf{S}^* = \begin{pmatrix} \nu_0 & \nu_1 & \cdots & \nu_p \\ \nu_1 & \nu_2 & \cdots & \nu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_p & \nu_{p+1} & \cdots & \nu_{2p} \end{pmatrix} = (\nu_{j+l})_{0\le j,l\le p}, \quad \boldsymbol{c}_p = \begin{pmatrix} \mu_{p+1} \\ \vdots \\ \mu_{2p+1} \end{pmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{c}}_p = \begin{pmatrix} \mu_{p+2} \\ \vdots \\ \mu_{2p+2} \end{pmatrix}.$$

Lastly we define

- $\varphi(t|x) = \mathrm{E}_{Y|X}\left[Q_\omega(Y - \tau_\omega(X) + t)|X = x\right]$
- $\varphi^{(1)}(t|x) = \frac{\partial\varphi(t|x)}{\partial t} = 2\mathrm{E}_{Y|X}\left[L_\omega(Y - \tau_\omega(X) + t)|X = x\right]$
- $\varphi^{(2)}(t|x) = \frac{\partial^2\varphi(t|x)}{\partial t^2} = 2(1-\omega)P[Y \le \tau_\omega(X) - t|X = x] + 2\omega P[Y > \tau_\omega(X) - t|X = x]$
- $\gamma(\omega, x) = \varphi^{(2)}(0|x) = 2(1-\omega)P[Y \le \tau_\omega(X)|X = x] + 2\omega P[Y > \tau_\omega(X)|X = x]\,.$

The following notations and assumptions ((A1)-(A6)) are needed for the theoretical results.

(A1) The quantity $\varphi^{(2)}(t|z)$, regarded upon as a function of t, is continuous in a neighbourhood of the point 0, uniformly for z in a neighbourhood of x. Furthermore, we assume that $\varphi(t|z)$, $\varphi^{(1)}(t|z)$ and $\varphi^{(2)}(t|z)$, as functions of z, are bounded and continuous in a neighbourhood of x for all small t and that $\varphi(0|x) \neq 0$.

(A2) The density function $f_X(.)$ of X has a continuous first derivative and $f_X(x) > 0$.

(A3) The function $f_{Y|X}(y|x)$ is continuous in x for each y. Moreover, there exist positive constants ξ and δ and a positive function, $H(y|x)$, such that $\sup_{|x_n - x| \le \xi} f_{Y|X}(y|x_n) \le H(y|x)$ and that

$$\int \left|2L_\omega(y - \tau_\omega(x) - \tau_\omega^{(1)}(x)(y-x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(y-x)^p)\right|^{2+\delta} H(y|x)\mathrm{d}y < \infty \quad \text{and}$$

$$\int (Q_\omega(y-t) - Q_\omega(y) + 2L_\omega(y)t)^2 H(y|x)\mathrm{d}y = o(t^2) \quad \text{as} \quad t \to 0.$$

(A4) The function $\tau_\omega(.)$ has a continuous $(p+2)$th derivative.

(A5) The kernel $K(.) \ge 0$ is a continuous density function having a bounded support.

(A6) When estimating $\tau_\omega^{(j)}(\cdot)$ and in case $p - j$ is even, we require that $nh^3 \to \infty$, as $n \to \infty$.

Theorem 1 *Under Assumptions* (A1)—(A6) *and if* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$. *Then, for* $x \in \{y : f_X(y) > 0\}$, *for each* $j = 0, \ldots, p$,

$$\sqrt{nh^{2j+1}}\left[\widehat{\tau}_\omega^{(j)}(x) - \tau_\omega^{(j)}(x) - j!\beta_\omega^{(j)}(x)h^{p+1-j}\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, j!^2\left(\sigma_\omega^{(j)}(x)\right)^2\right), \quad \text{as} \quad n \to \infty,$$

with

$$\left(\sigma_\omega^{(j)}(x)\right)^2 = \frac{\int [2L_\omega(y - \tau_\omega(x))]^2 f_{Y|X}(y|x)\mathrm{d}y \int (\mathrm{adj}(\mathbf{S})\mathbf{z}_u)_{j+1}^2 K^2(u)\mathrm{d}u}{(\gamma(\omega, x))^2 f_X(x)\det(\mathbf{S})^2},$$

$$\beta_\omega^{(j)}(x) = \frac{1}{(p+1)!}\tau_\omega^{(p+1)}(x)\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\mathbf{c}_p + \frac{1}{(p+2)!}\tau_\omega^{(p+2)}(x)h\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\widetilde{\mathbf{c}}_p + \frac{1}{(p+1)!}\tau_\omega^{(p+1)}(x)h\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\widetilde{\mathbf{c}}_p\frac{f_X^{(1)}(x)}{f_X(x)}$$

with $\mathbf{z}_u = \left(1, u, u^2, \cdots, u^p\right)^T$, $\det(\mathbf{S})$ *is the determinant of* $\mathbf{S}$ *and* $\mathrm{adj}(\mathbf{S})$ *is the adjugate matrix of* $\mathbf{S}$.

The proof of Theorem 1 is provided in Appendix.

From Theorem 1, we obtain approximations of the asymptotic variance and bias, conditionally upon $\mathcal{X} = \{X_1, \ldots, X_n\}$. We find

$$\begin{aligned}\mathrm{AVar}[\widehat{\tau}_\omega^{(j)}(x)|\mathcal{X}] =& \mathbf{e}_{j+1}^T \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_{j+1}(j!)^2 \frac{\int \left(2L_\omega(y-\tau_\omega(x))\right)^2 f_{Y|X}(y|x)\mathrm{d}y}{\gamma^2(\omega,x)f_X(x)} \frac{1}{nh^{1+2j}} \\ &+ o_P\Big(\frac{1}{nh^{1+2j}}\Big) \\ \equiv& \mathrm{ApVar}(x) + o_P\Big(\frac{1}{nh^{1+2j}}\Big),\end{aligned} \tag{19}$$

and the asymptotic expression for the conditional bias is

$$\begin{aligned}\mathrm{ABias}[\widehat{\tau}_\omega^{(j)}(x)|\mathcal{X}] =& j!\frac{1}{(p+1)!}\tau_\omega^{(p+1)}(x)\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\mathbf{c}_p h^{p+1-j} + \frac{j!}{(p+2)!}h^{p+2-j} \\ &\times \mathbf{e}_{j+1}^T\mathbf{S}^{-1}\widetilde{\mathbf{c}}_p\left(\tau_\omega^{(p+2)}(x) + (p+2)\tau_\omega^{(p+1)}(x)\frac{f_X^{(1)}(x)}{f_X(x)}\right) + o_P(h^{p+2-j}) \\ \equiv& \mathrm{ApBias}(x) + o_P(h^{p+2-j}).\end{aligned}$$

So far we only assumed that the kernel function K is a probability function. Further simplifications in the asymptotic bias are obtained when K is in addition a symmetric density. See the following remark.

Remark 2 Suppose we have that the kernel $K(.) \geq 0$ is a continuous, symmetric density function with a bounded support, hence satisfying

$$\int_{-\infty}^{+\infty} K(z)\mathrm{d}z = 1 \quad \text{and} \quad \int_{-\infty}^{+\infty} zK(z)\mathrm{d}z = 0.$$

Since K is symmetric, $\mu_{2j+1} = 0$ for $j = 0, 1, \cdots, p$, the matrices $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ have the following structure

$$\mathbf{S} = \begin{pmatrix} * & 0 & * & 0 & \cdots \\ 0 & * & 0 & * & \cdots \\ * & 0 & * & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \text{and} \quad \widetilde{\mathbf{S}} = \begin{pmatrix} 0 & * & 0 & * & \cdots \\ * & 0 & * & 0 & \cdots \\ 0 & * & 0 & * & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

with $*$ denoting any nonzero number. The matrix $\mathbf{S}^{-1}$ has a structure similar to that of $\mathbf{S}$. The vectors $\boldsymbol{c}_p$ and $\widetilde{\boldsymbol{c}}_p$ contain zero's at alternating positions.

Using the particular structures of these matrices, we have to distinguish two different cases:

- $p - j$ odd

$$\text{ABias}[\widehat{\tau}_\omega^{(j)}(x)|\mathcal{X}] = \mathbf{e}_{j+1}^T \mathbf{S}^{-1} \boldsymbol{c}_p \frac{j!}{(p+1)!} \tau_\omega^{(p+1)}(x) h^{p+1-j} + o_P(h^{p+1-j}); \tag{20}$$

- $p - j$ even

$$\text{ABias}[\widehat{\tau}_\omega^{(j)}(x)|\mathcal{X}] = \mathbf{e}_{j+1}^T \mathbf{S}^{-1} \widetilde{\boldsymbol{c}}_p \frac{j!}{(p+2)!} \left(\tau_\omega^{(p+2)}(x) + (p+2) \tau_\omega^{(p+1)}(x) \frac{f_X^{(1)}(x)}{f_X(x)} \right) h^{p+2-j} + o_P(h^{p+2-j}).$$

Remark 3 In the local linear case ($p = 1$) and under the extra assumption of Remark 2, we have

$$\mathbf{S} = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \boldsymbol{c}_p = \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{c}}_p = \begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} 0 \\ \mu_4 \end{pmatrix}.$$

The asymptotic conditional bias of $\widehat{\tau}_\omega(x)$ is

$$\text{ABias}[\widehat{\tau}_\omega(x)|\mathcal{X}] = \frac{1}{2} \tau_\omega^{(2)}(x) \mu_2 h^2 + o_P(h^2)$$

and of $\widehat{\tau}_\omega^{(1)}(x)$ is

$$\text{ABias}[\widehat{\tau}_\omega^{(1)}(x)|\mathcal{X}] = \frac{\mu_4}{\mu_2} \frac{1}{6} \left(\tau_\omega^{(3)}(x) + 3\tau_\omega^{(2)}(x) \frac{f_X^{(1)}(x)}{f_X(x)} \right) h^2 + o(h^2).$$

The asymptotic conditional variances of $\widehat{\tau}_\omega(x)$ and $\widehat{\tau}_\omega^{(1)}(x)$ are

$$\text{AVar}[\widehat{\tau}_\omega(x)|\mathcal{X}] = \nu_0 \frac{\int \left(2L_\omega(y - \tau_\omega(x))\right)^2 f_{Y|X}(y|x) \mathrm{d}y}{\gamma^2(\omega, x) f_X(x)} \frac{1}{nh} + o_P\left(\frac{1}{nh}\right)$$

$$\text{AVar}[\widehat{\tau}_\omega^{(1)}(x)|\mathcal{X}] = \frac{\nu_2}{\mu_2^2} \frac{\int \left(2L_\omega(y - \tau_\omega(x))\right)^2 f_{Y|X}(y|x) \mathrm{d}y}{\gamma^2(\omega, x) f_X(x)} \frac{1}{nh^3} + o_P\left(\frac{1}{nh^3}\right).$$

Yao and Tong (1996) obtained the asymptotic normality result for the local linear case ($p = 1$) in a setting of strictly stationary processes. Our result for $\widehat{\tau}_\omega(\cdot)$ reduces to the result in Yao and Tong (1996, see Theorem 1). For the estimation of the first derivative $\tau_\omega^{(1)}(\cdot)$ some caution is needed, since then $p - j = 0$ and even, and the approximate bias expression is as in Remark 2 in case of symmetric kernels. The result in Yao and Tong (1996, see Theorem 1) for $\widehat{\tau}_\omega^{(1)}(\cdot)$ shows a flaw here since for a symmetric kernel the term $\int u^3 K(u) \mathrm{d}u = 0$.

Remark 4 When using the iterative procedure in Sect. 3.2, this procedure is implemented with a stopping rule. See Sect. 3.3. One may then wonder whether the asymptotic normality result that is established in Theorem 1 also holds for the estimator $\boldsymbol{\beta}^{(t)}$ for a fixed number of iterations t, and thus for the approximation of the minimizer of (12) obtained via the iterative procedure described in Sect. 3.2. The

asymptotic normality result indeed continues to hold for the approximate solution. This is illustrated in Section S3.6 in the Supplementary Material, via some simulations, and is argued from theoretical side in Sect. 8.

5 Bandwidth selection

In this section, we focus on the bandwidth selection problem. For simplicity, we restrict the discussion to the case that K is a symmetric kernel and $p-j$ is odd. Firstly, we derive an expression for an optimal bandwidth. Secondly, we discuss a rule-of-thumb (ROT) bandwidth selector. Thirdly, we turn to a location-scale model (7) and discuss bandwidth selection under this specific setting.

5.1 A theoretical optimal bandwidth choice

From the approximations of the asymptotic conditional bias and variance derived in Sect. 4, we can obtain an expression for a theoretical optimal bandwidth. Based on the approximate bias and variance expressions in (19) and (20), we compute the approximate mean square error (AMSE),

$$\begin{aligned}\mathrm{AMSE}\big(\widehat{\tau}_\omega^{(j)}(x)\big) &= \big[\mathrm{ApBias}(x)\big]^2 + \mathrm{ApVariance}(x)\\ &= \left(\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\boldsymbol{c}_p\frac{j!}{(p+1)!}\tau_\omega^{(p+1)}(x)h^{p+1-j}\right)^2\\ &\quad+\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_{j+1}(j!)^2\frac{\int\big(2L_\omega(y-\tau_\omega(x))\big)^2f_{Y|X}(y|x)\mathrm{d}y}{\gamma^2(\omega,x)f_X(x)}\frac{1}{nh^{1+2j}},\end{aligned}$$

and the approximate weighted mean integrated square error (AMISE),

$$\begin{aligned}\mathrm{AMISE}\big(\widehat{\tau}_\omega^{(j)}(\cdot)\big) &= \int\mathrm{AMSE}\big(\widehat{\tau}_\omega^{(j)}(x)\big)\,k(x)\mathrm{d}x\\ &= \left(\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\boldsymbol{c}_p\frac{j!}{(p+1)!}h^{p+1-j}\right)^2\int(\tau_\omega^{(p+1)}(x))^2k(x)\mathrm{d}x\\ &\quad+\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_{j+1}(j!)^2\frac{1}{nh^{1+2j}}\int\frac{\int\big(2L_\omega(y-\tau_\omega(x))\big)^2f_{Y|X}(y|x)\mathrm{d}y}{\gamma^2(\omega,x)f_X(x)}k(x)\mathrm{d}x\end{aligned}\tag{21}$$

where $k(\cdot)\geq 0$ is some weight function. To minimize AMISE, looked upon as a function of h, we differentiate (21) with respect to h, put this derivative equal to zero and obtain

$$h^{2p+3}=\frac{\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_{j+1}(j!)^2\frac{(1+2j)}{n}\int\frac{\int\left(2L_\omega(y-\tau_\omega(x))\right)^2f_{Y|X}(y|x)\mathrm{d}y}{\gamma^2(\omega,x)f_X(x)}k(x)\mathrm{d}x}{\left(\mathbf{e}_{j+1}^T\mathbf{S}^{-1}\boldsymbol{c}_p\frac{j!}{(p+1)!}\right)^2(2(p+1-j))\int(\tau_\omega^{(p+1)}(x))^2k(x)\mathrm{d}x}.$$

This leads to the theoretical optimal constant bandwidth

$$h_{\text{opt}} = C_{p,j}(K)\left(\frac{\int \frac{\int \left(2L_\omega(y-\tau_\omega(x))\right)^2 f_{Y|X}(y|x)\mathrm{d}y}{\gamma^2(\omega,x) f_X(x)} k(x)\mathrm{d}x}{\int (\tau_\omega^{(p+1)}(x))^2 k(x)\mathrm{d}x}\right)^{1/(2p+3)} n^{-1/(2p+3)} \tag{22}$$

with

$$C_{p,j}(K) = \left(\frac{(p+1)!(1+2j)\mathbf{e}_{j+1}^T \mathbf{S}^{-1}\mathbf{S}^*\mathbf{S}^{-1}\mathbf{e}_{j+1}}{2(p+1-j)\left(\mathbf{e}_{j+1}^T \mathbf{S}^{-1}\boldsymbol{c}_p\right)^2}\right)^{1/(2p+3)}.$$

Values of the constant $C_{p,j}(K)$ for various kernels K and values p and j were tabulated in Fan and Gijbels (1996, see Table 3.2, page 67). Note that the optimal bandwidth in (22) depends on several unknown quantities $\tau_\omega^{(p+1)}(x)$, $\gamma(\omega,x)$, $f_X(x)$ and $f_{Y|X}(\cdot|x)$.

5.2 Rule-of-thumb (ROT) bandwidth selector

We firstly discuss a practical rule-of-thumb (ROT) bandwidth selection procedure that is in general applicable. For this, we follow the approach exposed in Fan and Gijbels (1996).

By taking the weight function $k(x) = k_0(x) f_X(x)$ with $k_0(\cdot) \geq 0$ a chosen weight function, we obtain from (22)

$$h_{\text{opt}} = C_{p,j}(K)\left(\frac{\int \frac{\mathrm{E}_{Y|X}[4L_\omega^2(Y-\tau_\omega(X))|X=x]}{\gamma^2(\omega,x)} k_0(x)\mathrm{d}x}{\int (\tau_\omega^{(p+1)}(x))^2 k_0(x) f_X(x)\mathrm{d}x}\right)^{1/(2p+3)} n^{-1/(2p+3)}. \tag{23}$$

A rule-of-thumb bandwidth selector is then obtained via the following procedure.

- Fit globally a parametric polynomial model of order $p+4$, and obtain the fitted model $\check{\tau}_\omega(x) = \check{\alpha}_0 + \check{\alpha}_1 x + \check{\alpha}_2 x^2 + \cdots + \check{\alpha}_{p+4} x^{p+4}$.
- Replace the unknown quantities $\mathrm{E}_{Y|X}[L_\omega^2(Y-\tau_\omega(X))|X=x]$ and

$$\begin{aligned}\gamma(\omega,x) =& 2(1-\omega)P[Y \leq \tau_\omega(X)|X=x] + 2\omega P[Y > \tau_\omega(X)|X=x] \\ =& 2(1-\omega)\mathrm{E}\left[\mathbb{1}\left\{Y-\tau_\omega(X) \leq 0\right\}|X=x\right] + 2\omega\mathrm{E}\left[\mathbb{1}\left\{Y-\tau_\omega(X) > 0\right\}|X=x\right],\end{aligned}$$

in (23) by the estimated unconditional sample versions:

$$\frac{1}{n}\sum_{i=1}^{n} L_\omega^2(Y_i - \check{\tau}_\omega(X_i))$$

$$\text{and} \quad 2(1-\omega)\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{Y_i - \check{\tau}_\omega(X_i) \le 0\right\} + 2\omega\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{Y_i - \check{\tau}_\omega(X_i) > 0\right\}.$$

- This leads to the following "rough" approximation of the optimal bandwidth in (22)

$$C_{p,j}(K)\left(\frac{\frac{4}{n}\sum_{i=1}^{n} L_\omega^2(Y_i-\check{\tau}_\omega(X_i))}{4\left((1-\omega)\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i\le\check{\tau}_\omega(X_i)\}+\omega\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i>\check{\tau}_\omega(X_i)\}\right)^2}\frac{\int k_0(x)\mathrm{d}x}{\int(\tau_\omega^{(p+1)}(x))^2 k_0(x) f_X(x)\mathrm{d}x}\right)^{1/(2p+3)} n^{-1/(2p+3)},$$

  in which the denominator can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathrm{d}^{p+1}\check{\tau}_\omega(x)}{\mathrm{d}x^{p+1}}\bigg|_{x=X_i}\right)^2 k_0(X_i).$$

- The rule-of-thumb (ROT) bandwidth selector is then defined as

$$\check{h}_{\text{opt}}^{[1]} = C_{p,j}(K)\left(\frac{\frac{1}{n}\sum_{i=1}^{n} L_\omega^2(Y_i-\check{\tau}_\omega(X_i))}{\left((1-\omega)\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i\le\check{\tau}_\omega(X_i)\}+\omega\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i>\check{\tau}_\omega(X_i)\}\right)^2}\frac{\int k_0(x)\mathrm{d}x}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d^{p+1}\check{\tau}_\omega(x)}{\mathrm{d}x^{p+1}}\Big|_{x=X_i}\right)^2 k_0(X_i)}\right)^{1/(2p+3)} n^{-1/(2p+3)}. \quad (24)$$

Although some of the above estimations are based on very rough approximations, it is seen from the simulation study that the resulting estimated expectile curves using this general ROT bandwidth selector are of very good quality.

5.3 Bandwidth selection under a location-scale model

When we are in a location-scale model (7), there are two important observations to be made: (i) the expression for the ωth expectile of Y given X in (9); and (ii) the simplified expression for the one-to-one mapping in (10). This leads to three bandwidth selectors under a location-scale model: a first one that only exploits fact (i); a second bandwidth selector in which both facts (i) and (ii) are exploited; and a third bandwidth selector in which we only exploit (ii), and link up to bandwidth selection for quantile regression, relying on the work of Yu and Jones (1998).

5.3.1 Rule-of-thumb (ROT) bandwidth selector (without the one-to-one mapping)

One of the unknown quantities appearing in (22) is $\gamma(\omega, x)$.

In a location-scale model this quantity can be simplified. Indeed, by exploiting that $Y = m(X) + \sigma(X)\epsilon$ and in particular (9) we obtain

$$\begin{aligned}\gamma(\omega, x) &= 2(1-\omega)P\{Y \leq \tau_\omega(x)|X = x\} + 2\omega P\{Y > \tau_\omega(x)|X = x\} \\ &= 2\big\{(1-\omega)P\big(\epsilon \leq \tau_{\omega,\epsilon}|X = x\big) + \omega P\big(\epsilon > \tau_{\omega,\epsilon}|X = x\big)\big\} \\ &= 2\big\{(1-\omega)P\big(\epsilon \leq \tau_{\omega,\epsilon}\big) + \omega P\big(\epsilon > \tau_{\omega,\epsilon}\big)\big\} \equiv \gamma(\omega),\end{aligned} \tag{25}$$

since ϵ and X are independent. If we know the distribution of ϵ, then we simply know the quantity $\gamma(\omega)$. However, if we do not know the distribution of ϵ we need to estimate $\gamma(\omega)$. From the location-scale model and (9), it follows that

$$\widetilde{\epsilon} = Y - \tau_\omega(X) = \sigma(X)\big[\epsilon - \tau_{\omega,\epsilon}\big],$$

and applying a conditional version of (3) it is clear that $\tau_{\omega,\widetilde{\epsilon}} = 0$. A rough way to estimate $P\big(\epsilon \leq \tau_{\omega,\epsilon}\big)$ is then to use estimates for $\widetilde{\epsilon}$ which are provided by the estimated residuals from a global parametric polynomial fit as in Sect. 5.2:

$$\check{\epsilon}_i = Y_i - \check{\tau}_\omega(X_i), \qquad \text{for } i = 1, \cdots, n.$$

A rough estimator for $P\big(\epsilon \leq \tau_{\omega,\epsilon}\big)$ is then $n^{-1}\sum_{i=1}^n \mathbb{1}\{\check{\epsilon}_i \leq \check{\tau}_\omega(\check{\epsilon}_i)\}$. See Section S5 in the Supplementary Material for some additional explanation regarding this. Subsequently, an estimator for $\gamma(\omega)$ is

$$\widehat{\gamma}(\omega) = 2\left\{(1-\omega)\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\check{\epsilon}_i \leq \check{\tau}_\omega(\check{\epsilon}_i)\} + \omega\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\check{\epsilon}_i > \check{\tau}_\omega(\check{\epsilon}_i)\}\right\}.$$

Note that the difference between this estimator for $\gamma(\omega)$ and the estimator for $\gamma(\omega, x)$ in (24) is that here we replace 0 by $\check{\tau}_\omega(\check{\epsilon}_i)$, which is possibly different for different index values i.

The above considerations lead to the bandwidth selector

$$\check{h}^{[2]}_{\text{opt}} = C_{p,j}(K)\left(\frac{\frac{1}{n}\sum_{i=1}^n 4L_\omega^2(Y_i - \check{\tau}_\omega(X_i))\int k_0(x)\mathrm{d}x}{\big(\widehat{\gamma}(\omega)\big)^2\frac{1}{n}\sum_{i=1}^n\left(\frac{d^{p+1}\check{\tau}_\omega(X_i)}{\mathrm{d}x^{p+1}}\right)^2 k_0(X_i)}\right)^{1/(2p+3)} n^{-1/(2p+3)}. \tag{26}$$

5.3.2 Rule-of-thumb (ROT) bandwidth selector (with the one-to-one mapping)

In case of a location-scale model, we can also further exploit the simple expression for the one-to-one mapping in (10) between quantiles and expectiles which is independent of the distribution of X. With this relation $\gamma(\omega(\alpha), x)$ can be written as

$$\begin{aligned}\gamma(\omega(\alpha), x) &= 2\big((1-\omega(\alpha))P\{Y \le \tau_{\omega(\alpha)}(x)|X=x\} + \omega(\alpha)P\{Y > \tau_{\omega(\alpha)}(x)|X=x\}\big)\\ &= 2\big((1-\omega(\alpha))P\{Y \le q_\alpha(x)|X=x\} + \omega(\alpha)P\{Y > q_\alpha(x)|X=x\}\big)\\ &= 2(\omega(\alpha)(1-2\alpha)+\alpha),\end{aligned}$$

with $\omega(\alpha)$ as defined in (10).

If we know the distribution of ϵ we know the precise relationship $\omega(\alpha)$. Taking as before $k(x) = k_0(x)f_X(x)$ the practical version of (22) becomes

$$\check{h}_{\text{opt}} = C_{p,j}(K)\left(\frac{\frac{1}{n}\sum_{i=1}^n L^2_{\omega(\alpha)}(Y_i - \check{\tau}_{\omega(\alpha)}(X_i))\int k_0(x)\mathrm{d}x}{(\omega(\alpha)(1-2\alpha)+\alpha)^2\frac{1}{n}\sum_{i=1}^n\left(\frac{d^{p+1}\check{\tau}_{\omega(\alpha)}(X_i)}{\mathrm{d}x^{p+1}}\right)^2 k_0(X_i)}\right)^{1/(2p+3)} n^{-1/(2p+3)}.$$

Remark 5 If we do not know the distribution of ϵ we can estimate the relation $\omega(\alpha)$ which means that for a given ω we need to find the corresponding α, which we denote by $\widehat{\alpha}$. In Section S5 in the Supplementary Material, we argue, via approximation of (10), that the approximate $\widehat{\alpha}$ is obtained by resolving

$$\omega(\widehat{\alpha}) = \frac{\check{F}^{-1}_{\check{\epsilon}}(\widehat{\alpha})\widehat{\alpha} - \frac{1}{n}\sum_{i=1}^n \check{\epsilon}_i \mathbb{1}\{\check{\epsilon}_i - \check{F}^{-1}_{\check{\epsilon}}(\widehat{\alpha}) \le \check{\tau}_\omega(\check{\epsilon}_i)\}}{\check{F}^{-1}_{\check{\epsilon}}(\widehat{\alpha})[2\widehat{\alpha}-1] + 2\frac{1}{n}\sum_{i=1}^n \check{\epsilon}_i \mathbb{1}\{\check{\epsilon}_i - \check{F}^{-1}_{\check{\epsilon}}(\widehat{\alpha}) > \check{\tau}_\omega(\check{\epsilon}_i)\} - \frac{1}{n}\sum_{i=1}^n \check{\epsilon}_i}, \tag{27}$$

with $\check{F}^{-1}_{\check{\epsilon}}(\widehat{\alpha})$ the $\widehat{\alpha}$th sample quantile of the residuals (of the global parametric polynomial fit). Using this $\omega(\widehat{\alpha})$, then leads to the data-driven bandwidth selector

$$\check{h}^{[3]}_{\text{opt}} = C_{p,j}(K)\left(\frac{\frac{1}{n}\sum_{i=1}^n L^2_{\omega(\widehat{\alpha})}(Y_i - \check{\tau}_{\omega(\widehat{\alpha})}(X_i))\int k_0(x)\mathrm{d}x}{\big(\omega(\widehat{\alpha})(1-2\widehat{\alpha})+\widehat{\alpha}\big)^2\frac{1}{n}\sum_{i=1}^n\left(\frac{\mathrm{d}^{p+1}\check{\tau}_{\omega(\widehat{\alpha})}(X_i)}{\mathrm{d}x^{p+1}}\right)^2 k_0(X_i)}\right)^{1/(2p+3)} n^{-1/(2p+3)}. \tag{28}$$

5.3.3 Quantile-based bandwidth selector

In a location-scale model when we have the simple one-to-one mapping in (10), we can also exploit the link between expectiles and quantiles by relying on data-driven bandwidth selectors that have been proposed for quantile regression. Indeed, we can use the one-to-one mapping and then working with quantiles and not expectiles. The minimizing function of (11) becomes

$$\sum_{i=1}^n Q_{\omega(\alpha)}\left(Y_i - \sum_{j=0}^p \beta_j(X_i - x)^j\right)K\left(\frac{X_i - x}{h}\right) \tag{29}$$

with $\omega(\alpha)$ as in (10).

The estimate of $\tau_\omega(x)$ is $\widehat{\beta}_0$ and the estimate of $\tau_\omega^{(j)}(x) = \frac{d^j \tau_\omega(x)}{\mathrm{d}x^j}$ be $\widehat{\beta}_j j!$. However, with the particular relationship between expectiles and quantiles we have

$$\beta_j = \frac{\tau_{\omega(\alpha)}^{(j)}(x)}{j!} = \frac{1}{j!} \frac{\mathrm{d}^j q_\alpha(x)}{\mathrm{d}x^j}.$$

Then, the minimization problem of the ωth conditional expectile of Y given $X = x$ can be seen as a minimization problem to find the αth conditional quantile of Y given $X = x$.

If the problem is seen as a minimization problem for the αth conditional quantile of Y given $X = x$, we can rely on bandwidth selectors that have been developed for quantile regression, such as the one proposed by Yu and Jones (1998) for the local linear quantile regression ($p = 1$), which was extended to the general local polynomial case in Gijbels et al. (2019). This results into:

STEP 1 Use ready-made and sophisticated methods to select h_{mean}, the optimal bandwidth choice for mean regression mean estimation (i.e. estimation of $m(\cdot)$).

STEP 2 Compute

$$h_\alpha = h_{\text{mean}} \left(\frac{\alpha(1-\alpha)}{(\phi(\boldsymbol{\Phi}^{-1}(\alpha)))^2} \right)^{1/(2p+3)}, \tag{30}$$

where $\phi(\cdot)$ and $\boldsymbol{\Phi}(\cdot)$ are, respectively, the standard normal density and cumulative distribution function.

In this case when the distribution of ϵ is known, we refer to (30) as $\check{h}_{\text{opt}}^{[5]}$.

If we do not know the distribution of ϵ, we proceed as in in Remark 5 to obtain an estimate for α in $\omega(\alpha)$. In this case, we denote the bandwidth according to (30) as $\check{h}_{\text{opt}}^{[4]}$.

5.4 Location-scale model and approach inspired by Efron (1991)

Using minimization problem (29), we can yet find another way to compute $\widehat{\omega}(\alpha)$ when the distribution of ϵ is not known, following Yao and Tong (1996) and relying on an approach of Efron (1991). The estimate of $\omega(\alpha)$, denoted by $\widehat{\omega}(\alpha) \in (0, 1)$ is determined in such a way that the proportion of data in the sample $\{(X_i, Y_i), 1 \leq i \leq n\}$ lying below the regression curve $\{y = \widehat{q}_\alpha(x) : x \in \mathbb{R}\}$ equals α. For a grid of ω-values, between 0 and 1, we vary the values of ω until that the proportion of the sample lying below the regression curves is equal to α (the value of α is fixed). In this case, the bandwidth used is h_α (see (30)). This approach is of a very different nature. It results into a different estimation procedure in case of a location-scale model with unknown error distribution.

6 Simulation study

We conducted a simulation study to investigate the finite sample performance of the local polynomial expectile estimator and the different practical bandwidth selectors in Sect. 5. For this study, we restrict to local linear fitting ($p = 1$) for estimating the expectile function $\tau_\omega(\cdot)$ ($j = 0$). A main focus will be on the quality of the bandwidth selectors.

6.1 Simulation models and settings

In the simulation study, we considered 3 simulation models:

- **Model 1**: a homoscedastic location-scale model

  $$Y = \sin(2X) + 2\exp(-16X^2) + 0.3\epsilon \qquad \text{with } X \sim U(-3, 3) \text{ and } \epsilon \sim \mathcal{N}(0, 1).$$

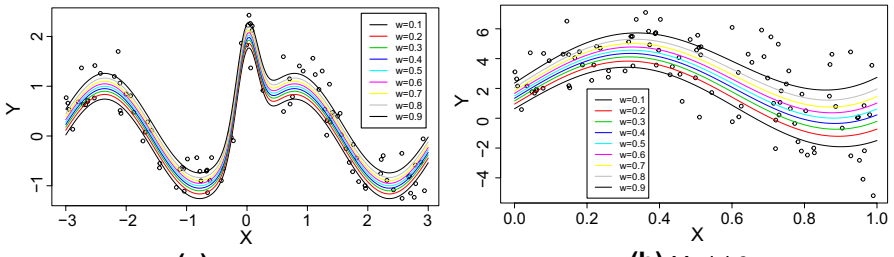
- **Model 2**: a heteroscedastic location-scale model

  $$Y = 1.5 + 2X^3 + 3\sin(5X) + \exp(0.9X)\epsilon \qquad \text{with } X \sim U(0, 1) \text{ and } \epsilon \sim \mathcal{N}(0, 1).$$

- **Model 3**: a non-location-scale model, in which the conditional density of Y given $X = x$ follows a gamma distribution $(Y \mid X = x) \sim \Gamma(\exp(5x), \exp(3x))$, with conditional density function

  $$f_{Y|X=x}(y|x) = \frac{\exp(3x)^{\exp(5x)}}{\Gamma(\exp(5x))} y^{\exp(5x)-1} \exp(-\exp(3x)y), \qquad \text{with } X \sim U(0, 1).$$

Model 1 is inspired by a model considered in Fan and Gijbels (1996), and Model 2 by a heteroscedastic regression model in Zhang and Mei (2008). Figure 3a, b shows a sample of size $n = 100$ of, respectively, Model 1 and Model 2, together with the true expectiles curves, shown for ω values 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. Notice that in the homoscedastic Model 1, the expectile curves are parallel, whereas this is not the case in the heteroscedastic Model 2. Figure 8a



(a) *Model 1.* **(b)** *Model 2.*

Fig. 3 Scatterplot and true expectile curves for Models 1 and 2

depicts the scatterplot of a sample of size $n = 100$ from Model 3, with the true expectile curves.

For all models, we take samples of size $n = 100$, unless differently indicated. For each sample, we calculate the local linear expectile regression estimate $\widehat{\beta}_0(x)$, for each point in a grid of 200 equispaced grid-values denoted by $\{x_1, \ldots, x_{200}\}$ on the domain (a, b) of the variable X. For the local linear method, we use a Gaussian kernel $K(u) = (\sqrt{2\pi})^{-1} e^{-\frac{1}{2}u^2}$ and take $k_0(\cdot)$ the indicator function on $[-2.8, 2.8]$ for Model 1, and on $[0.1, 0.9]$ for Models 2 and 3.

For each of the simulation models, we investigate the performance of the bandwidth selection methods, discussed in Sect. 5. Under a location-scale model, we can consider two situations: when the distribution of ϵ is known or not. So for the bandwidth selectors in Sects. 5.3.1, 5.3.2 and 5.3.3 we could consider these two situations. When we know the error distribution the quantile $\gamma(\omega)$ in (25) is known and the bandwidths from Sects. 5.3.1 and 5.3.2 coincide. We present simulation results that investigate the impact of the error distribution to be known or not for the rule-of-thumb (ROT) type of bandwidth selectors in Sects. 5.3.1 and 5.3.2, and the quantile-based bandwidth selector in Sect. 5.3.3. A summary of the investigated bandwidth selectors is presented in Table 1, together with the abbreviations used when presenting the results. For the data driven choice for h_{mean} in the quantile-based method in Sects. 5.3.3 and 5.4, we used the plug-in bandwidth selector for mean regression from Fan and Gijbels (1996) that is implemented in the R package `locpol` under the command `pluginBw`.

We draw 100 samples of the indicated sizes for each model. For each sample, we calculate the local linear expectile regression estimator $\widehat{\tau}_\omega(\cdot)$ for $\tau_\omega(\cdot)$ using the specified bandwidth selector/method. We present results for five values of ω: 0.1, 0.3, 0.5, 0.7 and 0.9. For each value of ω and each method (see Table 1), we compute for each estimator $\widehat{\tau}_\omega(\cdot)$ the approximate integrated square error (AISE):

Table 1 Different data-driven bandwidth selectors and methods

Method	Described in section	Distribution of ϵ	Data-driven bandwidth	Abbreviation
General rule-of-thumb	5.2		$\check{h}^{[1]}_{\text{opt}}$ in (24)	GenROT
Location-scale based ROT without one-one mapping	5.3.1	Unknown	$\check{h}^{[2]}_{\text{opt}}$ in (26)	LSROTWithout
Location-scale based ROT with one-one mapping	5.3.2	Unknown	$\check{h}^{[3]}_{\text{opt}}$ in (28)	LSROTWith
Location-scale-based ROT		Known	$\check{h}^{[2]}_{\text{opt}}$ in (26) But with (25)	LSROT
Location-scale quantile-based	5.3.3	Unknown	$\check{h}^{[4]}_{\text{opt}}$	LSQBased
Location-scale quantile-based		Known	$\check{h}^{[5]}_{\text{opt}}$	LSQBasedKnown
Approach of Efron (1991)	5.4	Unknown	$\check{h}^{[4]}_{\text{opt}}$	LSEfron

AISE $= \frac{b-a}{200} \sum_{j=1}^{200} \left(\widehat{\tau}_\omega(x_j) - \tau_\omega(x_j)\right)^2$. For the method in Sect. 5.4, we present results for $\alpha \in \{0.1945, 0.3680, 0.5000, 0.6320, 0.8055\}$, since these α values correspond to the theoretical ω values 0.1, 0.3, 0.5, 0.7 and 0.9, respectively (knowing the distribution of ϵ).

For each bandwidth selector/method in Table 1, we present a boxplot of the AISE-values. Furthermore, we depict three representatives of the 100 estimated curves as follows. We order the 100 values of AISE, and depict the estimates corresponding to the 0.05th percentile, the 0.50th percentile (i.e. the median) and the 0.95th percentile of the AISE-values. These are in the sequel called the three representative estimated curves.

6.2 Simulation results

6.2.1 Performances of practical bandwidth selectors

We first compare the qualities of the different data-driven bandwidth selectors in Sects. 5.2 and 5.3 with their (approximate) theoretical counterparts (denoted by h_{opt}) in, respectively, (22) and (30), where the value of the theoretical optimal h_{mean} is as can be found in Fan and Gijbels (Fan and Gijbels 1996, see expression (3.21) on page 68). We take $\omega = 0.3$ for this illustration. To give an idea about the behaviour of the practical bandwidth selectors with increasing sample size, we here also consider three sample sizes $n = 100$, $n = 500$ and $n = 1\,000$. For each simulated sample, we calculate the different data-driven bandwidth selectors.

We investigate the performances of the various practical bandwidth selectors for Models 1—3. For Models 1 and 2, which are location-scale models, we can compare the bandwidth estimates with the theoretical optimal bandwidths h_{opt}. Model 3, however, is not a location-scale model, and hence for this model we have no theoretical optimal value to compare with. Due to space limitations, we only present here results for Model 2, for sample sizes $n = 100$ and $n = 1\,000$. Similar results for Model 1 (for sample sizes $n = 100$, $n = 500$ and $n = 1\,000$), and for Model 3 (for sample sizes $n = 100$ and $n = 1\,000$) are provided in Section S3.1 in the Supplementary Material.

For all bandwidth selectors, we present density estimates of these data-driven bandwidth selectors based on their 100 realizations for the 100 simulated samples. For a clear graphical presentation, we present on the horizontal axis the values of $\widehat{h} - h_{\text{opt}}$ with $\widehat{h}$ the considered data-driven bandwidth selector (for $\check{h}^{[k]}_{\text{opt}}$ for $k = 1, 2, 3, 4, 5$) and h_{opt} the respective optimal bandwidth selector.

Figures 4 and 5 depict kernel density estimates of $\check{h}^{[k]}_{\text{opt}} - h_{\text{opt}}$ for $k = 1, 2, 3$, for Model 2, for, respectively, the sample sizes $n = 100$ and $n = 1\,000$. For the purpose of visual comparison, the range of the vertical and horizontal axes is kept the same for the two plots, and we indicate with a vertical line the position of the point zero. Firstly, the density estimates of the three ROT bandwidths selectors GenROT, LSROTWith and LSROT are quite comparable. Remarkable is that using knowledge of the error distribution (in LSROT) or not makes little difference. Secondly,

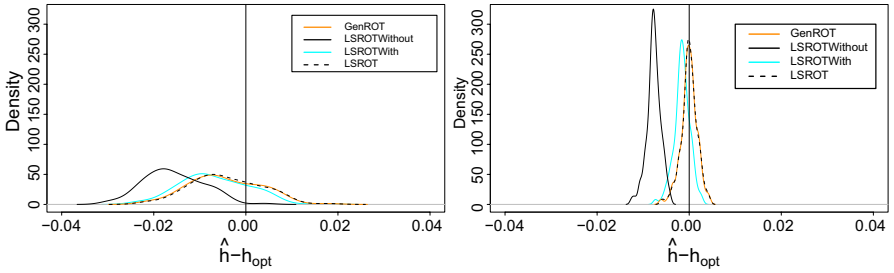


Fig. 4 Model 2. Kernel density estimates of the three ROT bandwidth selectors in Sects. 5.2 and 5.3, for estimation of $\tau_{0.3}(\cdot)$. The vertical lines indicates the zero position. Sample size $n = 100$ (left) and $n = 1\,000$ (right)

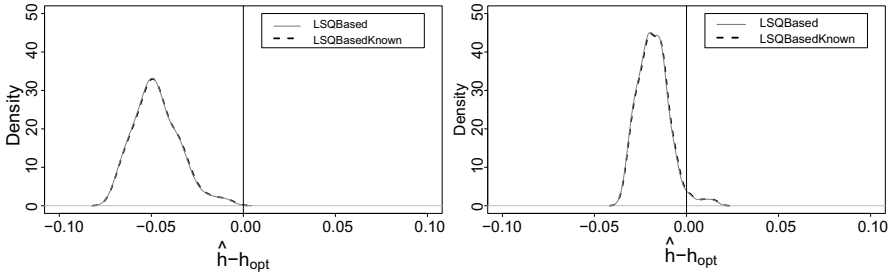


Fig. 5 Model 2. Kernel density estimates of the quantile-based bandwidth selector in Sect. 5.3.3, with or without assuming knowledge of the distribution of ϵ, for estimation of $\tau_{0.3}(\cdot)$. The vertical lines indicates the zero position. Sample size $n = 100$ (left) and $n = 1000$ (right)

the bandwidth selector LSROTWithout is further away from the theoretical bandwidth h_{opt}. Thirdly, all bandwidth selectors improve with increasing n, but the convergence of them to h_{opt} is moderately slow: for sample size $n = 100$, the difference $\widehat{h} - h_{\text{opt}}$ is, for GenROT, LSROTWith and LSROT, concentrated around approximately -0.008, and for $n = 1\,000$ this mode position has shifted closer to zero.

What is of course crucial is to see whether the bandwidth selectors lead to a good performance for the expectile estimator $\widehat{\tau}_{\omega}(\cdot)$. We investigated this and present results regarding this aspect for Models 2 and 3 in, respectively, Sects. 6.2.2 and 6.2.3. Similar results for Model 1 are discussed in Section S3.2 in the Supplementary Material.

6.2.2 Simulation results for Model 2

Boxplots of the AISE-values of the local linear expectile regression estimates for Model 2 for all methods in Table 1, for the considered values of ω, are provided in Fig. 6. In general the more extreme expectiles (i.e. for $\omega = 0.1$ and $\omega = 0.9$) are (a bit) less well estimated. The (red) dot in each boxplot presents the mean of the AISE-values. The performances of the local linear estimation method with any of

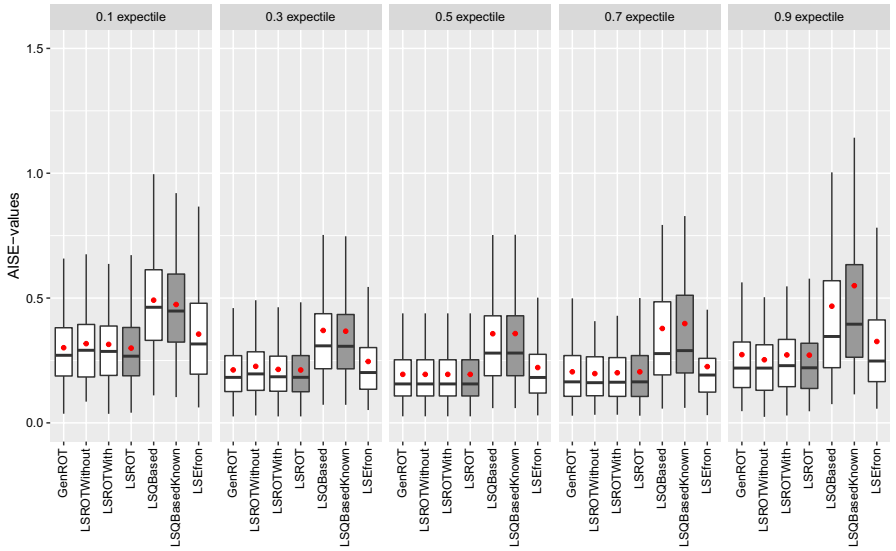


Fig. 6 Model 2. Boxplots of the AISE-values from 100 simulated samples of size $n = 100$, using the different methods listed in Table 1. Grey-filled boxplots are for the cases when we assume the error distribution to be known

the rule-of-thumb (ROT) bandwidth selectors are quite comparable (see the first four boxplots). Furthermore, having to estimate the error distribution (compare the grey-filled boxplots with the appropriate non-filled boxplots), only has little impact. From the overall slightly larger AISE-values (compared to those in Figure S.6), we see that the estimation task for Model 2 is a bit more difficult than for Model 1 (See Section S3.2). Furthermore, for Model 2 there is a clear superior performance of the bandwidth selection procedures in Sects. 5.2, 5.3.1 and 5.3.2. The quantile-based bandwidth selectors in Sect. 5.3.3 perform the worst among all methods, even in case the distribution of the error ϵ is known. The LSEfron method performs better for this simulation model, but it performs less than the bandwidth selectors in Sects. 5.2, 5.3.1, and 5.3.2. In Table S.1 in the Supplementary Material, we provide 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. From Table S.1, it can be seen that the local linear expectile estimator using the LSEfron implementation approach has a (too) high computational cost. Therefore, it is not included in our further summary of simulation results.

Figure 7 depicts the three representatives of the estimated curves for the expectile curve $\tau_{0.3}(\cdot)$ under Model 2, using the bandwidth selectors GenROT and LSQBased. The estimated curve with the GenROT bandwidth appears as slightly smoother.

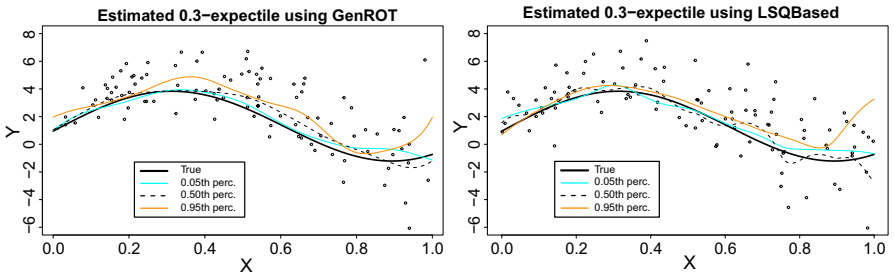


Fig. 7 Model 2. True expectile curve $\tau_{0.3}(\cdot)$ (in black) and three representative local linear estimates, based on samples of size $n = 100$: 0.05th AISE-percentile (light-grey; color blue), 0.5th AISE-percentile (dashed line), 0.95th AISE-percentile (grey; color ochre yellow), using, respectively, the bandwidth selection method GenROT (left panel) and the LSQBased method (right panel) (color figure online)

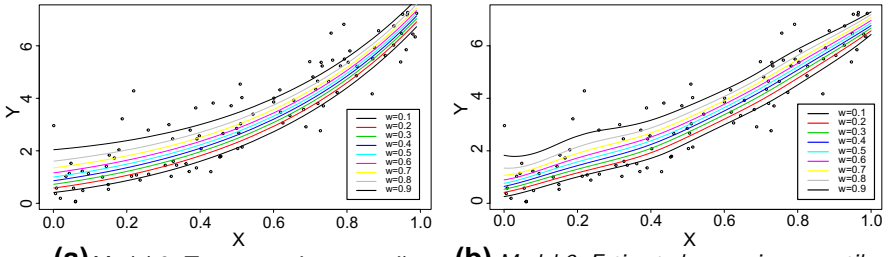


Fig. 8 Model 3. **a**. True expectile curves; and **b** estimated expectile curves, based on a sample of size $n = 100$, using the GenROT bandwidth selector

6.2.3 Simulation results for Model 3

A sample of size $n = 100$ together with the GenROT data-driven bandwidth implementation of the expectile curves is depicted in Fig. 8b. We can observe that the estimated expectile curves seem to be a bit less 'regular' than the true curves.

Since Model 3 is a non-location-scale model, we can use this model to investigate the loss in efficiency when using location-scale based data-driven bandwidth selectors (obviously all without knowing the error distribution). See Fig. 9 for boxplots of the AISE-values. Among the methods that are exploiting (wrongly) a location-scale modelling setting, the two ROT type methods perform still quite well, as well as the quantile-based method LSQBased. However, the latter method is performing a bit less, with slightly higher mean AISE-values and a slightly larger variance. The performance of the GenROT method which is theoretically the only appropriate one is comparable to these for the other ROT type of methods, in particular that of LSROTWith. In Figure S.8 in the Supplementary Material, the reader can find a graphical presentation of the true $\tau_{0.3}(\cdot)$ expectile curve together with the three representative estimates, for the four methods for which boxplots are shown in Fig. 9.

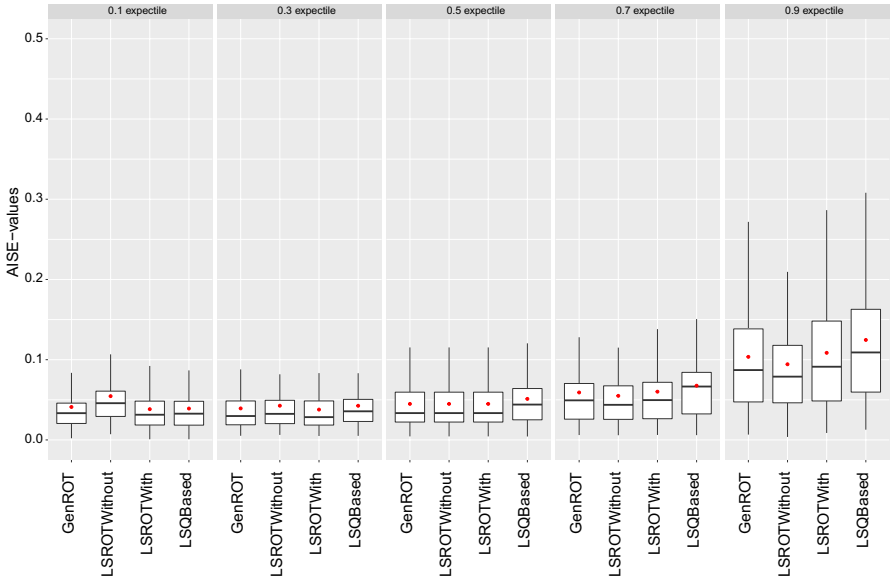


Fig. 9 Model 3. Boxplots of the AISE-values from 100 simulated samples of size $n = 100$, using methods listed in Table 1

Keeping in mind the results for the three simulation models, we recommend to use any of the rule-of-thumb bandwidth selectors.

7 Real data illustration

We apply the studied method to real data on the Head Circumference of Dutch Boys. The data are coming from the Fourth Dutch Growth Study, see Fredriks et al. (2000), which is a cross-sectional study that measures growth and development of the Dutch boys population between the ages 0 and 21 years and contains 7 040 observations. The X values are the square root of the age, and the Y values are the head circumference (in cm). There are 1 000 different observations for X. The data are to be found in the R package `gamlss`. Since for a real data set, one does not know whether a location-scale modelling background would be appropriate or not, one would prefer to work with the GenROT bandwidth selector in Sect. 5.2. For comparison purpose, we include the estimates using the LSQBased method.

Table 2 Head circumference data. Data-driven bandwidth values

ω	0.1	0.3	0.5	0.7	0.9
GenROT	0.1059	0.0961	0.0944	0.0956	0.1052
LSQBased	0.0951	0.0787	0.0741	0.0742	0.0760

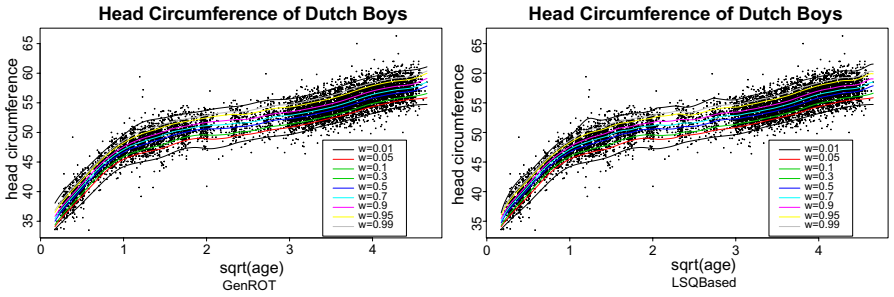


Fig. 10 Head circumference data. Estimated expectile regression curves for $\tau_\omega(\cdot)$ for ω taking values 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95 and 0.99. Estimated expectiles using the GenROT (left panel) and the LSQBased (right panel) bandwidth selectors

Scatterplots of the data together with the estimated expectile regression curves for ω equal to 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95 and 0.99 are shown in Fig. 10. In Table 2, we give the values of the data-driven bandwidths, for five considered ω values. Bandwidth values from the GenROT method are mostly larger than the values from the LSQBased bandwidth selector. The estimated expectile curves appear as quite parallel, which might indicate that a location-scale model could possibly be appropriate for the modelling. The data clearly show heteroscedasticity.

From the estimated expectile curves, one can get some insights. For example, if we focus on Dutch boys of 1 and 4 years old, we see that the estimated 0.3-expectile for the Dutch boys of 1 year old equals 46.64. So the average distance from the data Y_i (the head circumference) below 46.64 to 46.64 is 30%. In comparison, for Dutch boys of 4 years old, the estimated 0.3-expectile is 50.57, which is an increase of about 8.5% when compared to the group of 1 year old boys. So the individual differences are increasing when passing from one group to the other.

Additional real data applications are provided in Section S4 of the Supplementary Material.

8 Further discussion and conclusion

This paper contributes with a detailed study of local polynomial expectile regression. The unique solution to the optimization problem is found by an iterative procedure, which results into consecutively solving reweighted least squares problems. As such the way to proceed is similar as for expectile estimation in the linear case, introduced by Newey and Powell (1987). This is due to the fact to local polynomial fitting can conveniently be viewed as a weighted least squares problem. The above arguments are also at the heart of the fact that the iterative algorithm converges to effectively the minimizer of (12).

Although we do not establish a formal theoretical result for the approximate solution $\boldsymbol{\beta}^{(t)}$ (with t a fixed (but random) number of iterations), we would like to mention that we expect that such a result could be proven formally following the ideas

provided in Chen and Shao (1993). See Theorem 3 in that paper, which states such a result for the iterative weighted least squares estimators in a linear model context. A key result to get to this is Theorem 1 in Chen and Shao (1993) which constitutes a kind of i.i.d. representation of the concerned estimator. From the proof of Theorem 1 (see e.g. (A.3)), it is seen that such an i.i.d. representation is also valid in our context. Mimicking reasonings as in Chen and Shao (1993) would enable to show that an asymptotic normality result continues to hold for the approximate estimator $\boldsymbol{\beta}^{(t)}$, for any fixed number of iterations t, under the working conditions of a starting vector $\boldsymbol{\beta}^{(0)}$ as in (17), and under the assumption that the error term ϵ in a regression model has a symmetric distribution.

In this paper, we also deal with the important bandwidth selection issue. We provide a general rule-of-thumb (ROT) bandwidth selector and also discuss special cases of it when one is in the framework of a location-scale model. Furthermore, we also discuss a quantile-based bandwidth selector that exploits the relationship between quantiles and expectiles. Our detailed study shows that the ROT bandwidth selectors, although based intermediately on some rough approximations of unknown quantities, lead to very good finite-sample performance of the local polynomial expectile regression estimator. We therefore recommend in general to use either the general ROT bandwidth selector (GenROT), or in case of a location-scale model, the bandwidth selector LSROTWith.

As mentioned above, the bandwidth selectors were derived making some rough approximations. Of course, one could further improve these approximations. As an example, one could estimate conditional expectations with appropriate estimates for *conditional* expectations. This, however, would be at the cost of introducing extra smoothing/bandwidth parameters. In future research, one could thus work towards more sophisticated practical bandwidth selection rules, which will very likely improve upon their convergence rates towards the theoretical optimal bandwidths. It is not expected though that this would lead to a significant improvement of the finite-sample performance of the local polynomial expectile regression estimator, which is already very good.

Appendix

A.1 Proof of Theorem 1

The proof of this theorem is similar in setup as the one provided by Fan et al. (1994) to study nonparametric regression based on i.i.d. observations. The main idea of the proof is to approximate the quantity to be minimized in (11) by a quadratic function whose minimizer is asymptotically normal, and then to show that $(\widehat{\tau}_\omega(x), \widehat{\tau}_\omega^{(1)}(x), \cdots, \widehat{\tau}_\omega^{(p)}(x))^{\mathrm{T}}$ lies close enough to that minimizer to share the latter's asymptotic behaviour. The convexity lemma (Pollard 1991) plays a role in the above approximation. We give the details of the proof below.

Recall that, for x a given point, $\beta_0 = \tau_\omega(x), \beta_1 = \tau_\omega^{(1)}(x), \cdots, \beta_p = \frac{\tau_\omega^{(p)}(x)}{p!}$ and $\widehat{\beta}_0 = \widehat{\tau}_\omega(x), \widehat{\beta}_1 = \widehat{\tau}_\omega^{(1)}(x), \cdots, \widehat{\beta}_p = \frac{\widehat{\tau}_\omega^{(p)}(x)}{p!}$ with $(\widehat{\beta}_0, \cdots, \widehat{\beta}_p)$ minimizing

$$\sum_{i=1}^{n} Q_\omega\left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j\right) K\left(\frac{X_i - x}{h}\right).$$

Let

$$\begin{aligned}
K_i &= K\left(\tfrac{X_i - x}{h}\right) \\
\mathbf{Z}_i &= \left(1, \tfrac{X_i - x}{h}, \left(\tfrac{X_i - x}{h}\right)^2, \ldots, \left(\tfrac{X_i - x}{h}\right)^p\right)^{\mathrm{T}} \\
Y_i^* &= Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \tfrac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p \\
\widehat{\boldsymbol{\theta}} &= \sqrt{nh}\left(\widehat{\beta}_0 - \tau_\omega(x), \cdots, h^p\left(\widehat{\beta}_p - \tfrac{\tau_\omega^{(p)}(x)}{p!}\right)\right)^{\mathrm{T}}.
\end{aligned}$$

For $(\theta_0, \ldots, \theta_p)^{\mathrm{T}} = \boldsymbol{\theta} \in \mathbb{R}^{p+1}$, $\widehat{\boldsymbol{\theta}}$ minimizes the function

$$G_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}} \frac{\mathbf{Z}_i}{\sqrt{nh}}\right) - Q_\omega(Y_i^*) \right] K_i.$$

Note that the function $G_n(\boldsymbol{\theta})$ is convex in $\boldsymbol{\theta}$ (the second derivative is ≥ 0 for all $\boldsymbol{\theta}$). It is sufficient to prove that this function converges pointwise to its conditional expectation, since it follows from the convexity lemma of Pollard (1991) that the convergence is also uniform on any compact set of $\boldsymbol{\theta}$.

We next approximate $G_n(\cdot)$ by a quadratic function whose minimizing value has an asymptotic normal distribution. Two terms contribute to the approximation. One is a quadratic function obtained via a Taylor expansion of the expected value, and the other term is random and linear in $\boldsymbol{\theta}$. Write

$$G_n(\boldsymbol{\theta}) = \mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] - \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^{n} L_\omega(Y_i^*)\mathbf{Z}_i K_i - \mathrm{E}_{Y|X}\left[L_\omega(Y_i^*)|X_i\right]\mathbf{Z}_i K_i\right)^{\mathrm{T}} \boldsymbol{\theta} + R_n(\boldsymbol{\theta}) \tag{A.1}$$

with

$$R_n(\boldsymbol{\theta}) = G_n(\boldsymbol{\theta}) - \mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] + \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^{n} L_\omega(Y_i^*)\mathbf{Z}_i K_i - \mathrm{E}_{Y|X}[L_\omega(Y_i^*)|X_i]\mathbf{Z}_i K_i\right)^{\mathrm{T}} \boldsymbol{\theta}. \tag{A.2}$$

Let M be a real number such that the interval $[-M, M]$ contains the support of K. By Taylor expansion,

$$\tau_\omega(X_i) = \tau_\omega(x) + \tau_\omega^{(1)}(x)(X_i - x) + \cdots + \frac{\tau_\omega^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + \xi_{n,i} \quad \text{for} \quad |X_i - x| \leq Mh$$

with $\xi_{n,i} = o_P\left(|X_i - x|^{p+1}\right) = o_P(h^{p+1})$ holds uniformly as $X_i \to x$, i.e. $\max_{\{i:|X_i - x| \leq Mh\}} ||\xi_{n,i}||_\infty = o_P(h^{p+1})$ since $\tau_\omega(.)$ has a continuous $(p+2)$th derivative.

We have

$$
\begin{aligned}
&\mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] \\
&= \mathrm{E}_{Y|X}\left[\sum_{i=1}^{n}\left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\boldsymbol{Z}_i}{\sqrt{nh}}\right) - Q_\omega(Y_i^*)\right]K_i\middle|\mathcal{X}\right] \\
&= \sum_{i=1}^{n}\left[\mathrm{E}_{Y|X}\left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\boldsymbol{Z}_i}{\sqrt{nh}}\right)\middle|X_i\right] - \mathrm{E}_{Y|X}\left[Q_\omega(Y_i^*)|X_i\right]\right]K_i \\
&= \sum_{i=1}^{n}\left(\mathrm{E}_{Y|X}\left[Q_\omega\left(Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p - \boldsymbol{\theta}^{\mathrm{T}}\frac{\boldsymbol{Z}_i}{\sqrt{nh}}\right)\middle|X_i\right]\right. \\
&\quad \left.-\mathrm{E}_{Y|X}\left[Q_\omega\left(Y_i - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p\right)\middle|X_i\right]\right)K_i \\
&= \sum_{i=1}^{n}\left[\varphi\left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p - \boldsymbol{\theta}^{\mathrm{T}}\frac{\boldsymbol{Z}_i}{\sqrt{nh}}\middle|X_i\right)\right. \\
&\quad \left.-\varphi\left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p\middle|X_i\right)\right]K_i \\
&= -\sum_{i=1}^{n}\varphi^{(1)}\left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p\middle|X_i\right)\frac{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_i}{\sqrt{nh}}K_i \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\varphi^{(2)}\left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p\middle|X_i\right) \\
&\quad \frac{(\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_i)^2}{nh}K_i(1 + o_P(1)).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\varphi^{(2)}\left(\tau_\omega(X_i) - \tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p\middle|X_i\right) \\
&\quad = \varphi^{(2)}\left(\tau_\omega(x) + \tau_\omega^{(1)}(x)(X_i - x) + \cdots + \frac{\tau_\omega^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + \xi_{n,i}\right. \\
&\quad \left.-\tau_\omega(x) - \tau_\omega^{(1)}(x)(X_i - x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(X_i - x)^p)\middle|X_i\right) \\
&\quad = \varphi^{(2)}\left(\frac{\tau_\omega^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + \xi_{n,i}\middle|X_i\right) \\
&\quad = \varphi^{(2)}\left(0|X_i\right) + O_P(h^{p+1}).
\end{aligned}
$$

It follows that

$$\mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] = \frac{-2}{\sqrt{nh}} \sum_{i=1}^{n} \mathrm{E}_{Y|X}\left[L_\omega\left(Y_i^*|X_i\right)\right](\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_i)K_i + \frac{1}{2nh}\boldsymbol{\theta}^{\mathrm{T}}\left(\sum_{i=1}^{n} K_i\gamma(\omega, X_i)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}\right)\boldsymbol{\theta}(1 + o_P(1)).$$

Thus, we have

$$\frac{1}{nh}\sum_{i=1}^{n} K_i\gamma(\omega, X_i)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} = \frac{1}{nh}\sum_{i=1}^{n} K_i\gamma(\omega, X_i)\begin{pmatrix} 1 & \frac{X_i-x}{h} & \left(\frac{X_i-x}{h}\right)^2 & \cdots & \left(\frac{X_i-x}{h}\right)^p \\ \frac{X_i-x}{h} & \left(\frac{X_i-x}{h}\right)^2 & \left(\frac{X_i-x}{h}\right)^3 & \cdots & \left(\frac{X_i-x}{h}\right)^{p+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \left(\frac{X_i-x}{h}\right)^p & \left(\frac{X_i-x}{h}\right)^{p+1} & \left(\frac{X_i-x}{h}\right)^{p+2} & \cdots & \left(\frac{X_i-x}{h}\right)^{2p} \end{pmatrix}.$$

Denoting $\widetilde{S}_{n,j} = \frac{1}{nh}\sum_{i=1}^{n}\gamma(\omega, X_i)\left(\frac{X_i-x}{h}\right)^j K\left(\frac{X_i-x}{h}\right)$, for $j = 0, 1, \ldots, 2p$, it follows from the fact that K has bounded support (see e.g. Fan and Gijbels 1996) that

$$\widetilde{S}_{n,j} = \mathrm{E}_X[\widetilde{S}_{n,j}] + O_P\left(\sqrt{\mathrm{Var}_X(\widetilde{S}_{n,j})}\right)$$

$$\mathrm{E}_X[\widetilde{S}_{n,j}] = \frac{n}{nh}\int \gamma(\omega, v)f_X(v)\left(\frac{v-x}{h}\right)^j K\left(\frac{v-x}{h}\right)\mathrm{d}v = \int \gamma(\omega, x+uh)f_X(x+uh)u^j K(u)\mathrm{d}u$$

$$= f_X(x)\gamma(\omega, x)\mu_j + o(1),$$

where the last equality comes from the dominated convergence theorem where we assumed that $h \to 0$ and $f_X(.)$ is continuous in a neighbourhood of x.

A similar argument leads to

$$\mathrm{Var}_X[\widetilde{S}_{n,j}] = \mathrm{E}_X[\widetilde{S}_{n,j}^2] - \mathrm{E}_X[\widetilde{S}_{n,j}]^2 \le \frac{n}{n^2h^2}\int\left(\gamma(\omega, v)\left(\frac{v-x}{h}\right)^j K\left(\frac{v-x}{h}\right)\right)^2 f_X(v)\mathrm{d}v$$

$$= \frac{1}{nh}\int\left(\gamma(\omega, x+uh)u^j K(u)\right)^2 f_X(x+uh)\mathrm{d}u = o(1).$$

With this result and the definition of the matrix $\mathbf{S}$, we have

$$\frac{1}{nh}\sum_{i=1}^{n} K_i\gamma(\omega, X_i)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} = \gamma(\omega, x)f_X(x)\mathbf{S} + o_P(1).$$

We then obtain that

$$\begin{aligned}\mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] =& \frac{-2}{\sqrt{nh}}\sum_{i=1}^{n}\mathrm{E}_{Y|X}\left[L_\omega\left(Y_i^*|X_i\right)\right](\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)K_i \\ &+ \frac{1}{2nh}\boldsymbol{\theta}^{\mathrm{T}}\left(\sum_{i=1}^{n}K_i\gamma(\omega,X_i)\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)\boldsymbol{\theta}(1+o_P(1)) \\ =& \frac{-2}{\sqrt{nh}}\sum_{i=1}^{n}\mathrm{E}_{Y|X}\left[L_\omega\left(Y_i^*|X_i\right)\right](\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)K_i + \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta}(1+o_P(1)).\end{aligned}$$

Next we show that $R_n(\boldsymbol{\theta}) = o_P(1)$ (for the definition of $R_n(\boldsymbol{\theta})$ see (A.2)). We start by rewriting and approximating this quantity as follows:

$$\begin{aligned}&R_n(\boldsymbol{\theta}) \\ &= G_n(\boldsymbol{\theta}) - \mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] + \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^{n}L_\omega(Y_i^*)\mathbf{Z}_iK_i - \mathrm{E}_{Y|X}[L_\omega(Y_i^*)|X_i]\mathbf{Z}_iK_i\right)^{\mathrm{T}}\boldsymbol{\theta} \\ &= \sum_{i=1}^{n}\left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\mathbf{Z}_i}{\sqrt{nh}}\right) - Q_\omega(Y_i^*)\right]K_i \\ &\quad + \frac{2}{\sqrt{nh}}\sum_{i=1}^{n}\mathrm{E}_{Y|X}\left[L_\omega\left(Y_i^*|X_i\right)\right](\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)K_i - \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta}(1+o_P(1)) \\ &\quad + \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^{n}L_\omega(Y_i^*)\mathbf{Z}_iK_i - \mathrm{E}_{Y|X}[L_\omega(Y_i^*)|X_i]\mathbf{Z}_iK_i\right)^{\mathrm{T}}\boldsymbol{\theta} \\ &= \sum_{i=1}^{n}\left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\mathbf{Z}_i}{\sqrt{nh}}\right) - Q_\omega(Y_i^*)\right]K_i - \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta}(1+o_P(1)) \\ &\quad + \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^{n}L_\omega(Y_i^*)\mathbf{Z}_iK_i\right)^{\mathrm{T}}\boldsymbol{\theta} \\ &= \sum_{i=1}^{n}\left[Q_\omega\left(Y_i^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\mathbf{Z}_i}{\sqrt{nh}}\right) - Q_\omega(Y_i^*) + \frac{2}{\sqrt{nh}}L_\omega(Y_i^*)\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i\right]K_i - \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta}(1+o_P(1)).\end{aligned}$$

By using Assumption (A3), we obtain

$$\begin{aligned}&\mathrm{Var}_{X,Y}[R_n(\boldsymbol{\theta})] \le n\mathrm{E}_{Y,X}\left[\left(Q_\omega\left(Y_1^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{\mathbf{Z}_1}{\sqrt{nh}}\right) - Q_\omega(Y_1^*) + \frac{2}{\sqrt{nh}}\left(L_\omega(Y_1^*)\mathbf{Z}_1\right)^{\mathrm{T}}\boldsymbol{\theta}\right)^2\right]K_1^2 \\ &\le n\int\int\left(Q_\omega\left(y_v^* - \boldsymbol{\theta}^{\mathrm{T}}\frac{z}{\sqrt{nh}}\right) - Q_\omega(y_v^*) + \frac{2}{\sqrt{nh}}\left(L_\omega(y_v^*)z\right)^{\mathrm{T}}\boldsymbol{\theta}\right)^2 \\ &H(y|v)\mathrm{d}y K^2\left(\frac{v-x}{h}\right)f_X(v)\mathrm{d}v \\ &= o\left(n\int\left(\boldsymbol{\theta}^{\mathrm{T}}\frac{z}{\sqrt{nh}}\right)^2 K^2\left(\frac{v-x}{h}\right)f_X(v)\mathrm{d}v\right) = o(1)\end{aligned}$$

with $\mathbf{z} = \left(1, \frac{v-x}{h}, \left(\frac{v-x}{h}\right)^2, \cdots, \left(\frac{v-x}{h}\right)^p\right)^{\mathrm{T}}$ and $y_v^* = y - \tau_\omega(x) - \tau_\omega^{(1)}(x)(v-x) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(v-x)^p$.

It follows from the definition of $R_n(\boldsymbol{\theta})$ (in (A.2)) that for any $\boldsymbol{\theta} \in \mathbf{R}^{p+1}$, $\omega \in (0,1)$ and $x \in \mathbb{R}^{p+1}$, $\mathrm{E}_{X,Y}[R_n(\boldsymbol{\theta})] = 0$. Therefore $R_n(\boldsymbol{\theta}) = o_P(1)$. Indeed, for any constant $\epsilon > 0$ and by the inequality of Chebyshev,

$$P\left[\left|R_n(\boldsymbol{\theta})\right| > \epsilon\right] = P\left[\left|R_n(\boldsymbol{\theta}) - \mathrm{E}_{X,Y}\left[R_n(\boldsymbol{\theta})\right]\right| > \epsilon\right] \leq \frac{1}{\epsilon^2}\mathrm{Var}_{X,Y}\left[R_n(\boldsymbol{\theta})\right] = o(1).$$

For the quantity $G_n(\boldsymbol{\theta})$ in (A.1), we thus obtain

$$\begin{aligned}
G_n(\boldsymbol{\theta}) &= \mathrm{E}_{Y|X}[G_n(\boldsymbol{\theta})|\mathcal{X}] - \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^n L_\omega(Y_i^*)\mathbf{Z}_i K_i - \mathrm{E}_{Y|X}\left[L_\omega(Y_i^*)|X_i\right]\mathbf{Z}_i K_i\right)^{\mathrm{T}}\boldsymbol{\theta} + R_n(\boldsymbol{\theta}) \\
&= \frac{-2}{\sqrt{nh}}\sum_{i=1}^n \mathrm{E}_{Y|X}\left[L_\omega\left(Y_i^*|X_i\right)\right](\boldsymbol{\theta}^{\mathrm{T}}\mathbf{Z}_i)K_i + \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta}(1+o_P(1)) \\
&\quad - \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^n L_\omega(Y_i^*)\mathbf{Z}_i K_i - \mathrm{E}_{Y|X}\left[L_\omega(Y_i^*)|X_i\right]\mathbf{Z}_i K_i\right)^{\mathrm{T}}\boldsymbol{\theta} + R_n(\boldsymbol{\theta}) \\
&= \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta} - \frac{2}{\sqrt{nh}}\left(\sum_{i=1}^n L_\omega(Y_i^*)\mathbf{Z}_i K_i\right)^{\mathrm{T}}\boldsymbol{\theta} + r_n(\boldsymbol{\theta}) \\
&= \frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(\omega,x)f_X(x)\mathbf{S}\boldsymbol{\theta} + \mathbf{W}_n^{\mathrm{T}}\boldsymbol{\theta} + r_n(\boldsymbol{\theta})
\end{aligned}$$

with $r_n(\boldsymbol{\theta}) = o_P(1)$ for each fixed $\boldsymbol{\theta}$ and

$$\mathbf{W}_n = \frac{-2}{\sqrt{nh}}\left(\sum_{i=1}^n L_\omega(Y_i^*)\mathbf{Z}_i K_i\right).$$

It easy to see that $\mathbf{W}_n$ has a bounded second moment and hence is stochastically bounded. For $c > 0$ and by Assumption (A1) ($\varphi(t|z)$ is bounded), we have, with $z_u = \left(1, u, u^2, \cdots, u^p\right)^{\mathrm{T}}$, $y_u^* = y - \tau_\omega(x) - \tau_\omega^{(1)}(x)(hu) - \cdots - \frac{\tau_\omega^{(p)}(x)}{p!}(hu)^p$, using Hölder's inequality and Assumption (A3),

$$
\begin{aligned}
\mathrm{E}_{X,Y}[\mathbf{W}_n\mathbf{W}_n^{\mathrm{T}}] =& \frac{4}{nh}\mathrm{E}_{X,Y}\left[\left(\sum_{i=1}^{n} L_w(Y_i^*)\mathbf{Z}_i K_i\right)\left(\sum_{i=1}^{n} L_w(Y_i^*)\mathbf{Z}_i K_i\right)^{\mathrm{T}}\right] \\
\leq& \frac{4c}{nh}\mathrm{E}_{X,Y}\left[\left(\sum_{i=1}^{n} L_w(Y_i^*)^2 K_i^2 \mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)^{\mathrm{T}}\right] \\
=& \frac{4c}{h}\int\int L_w(y_v^*)^2 zz^{\mathrm{T}} f_{Y|X}(y|v)\mathrm{d}y K^2\left(\frac{v-x}{h}\right) f_X(v)\mathrm{d}v \\
=& \frac{4c}{h}\int\int L_w(y-\tau_\omega(x)-\tau_\omega^{(1)}(x)(v-x)-\cdots-\frac{\tau_\omega^{(p)}(x)}{p!}(v-x)^p)^2 zz^{\mathrm{T}} \\
& f_{Y|X}(y|v)\mathrm{d}y K^2\left(\frac{v-x}{h}\right) f_X(v)\mathrm{d}v \\
=& 4c\int\int L_w(y_u^*)^2 f_{Y|X}(y|u)\mathrm{d}y z_u z_u^{\mathrm{T}} K^2(u) f_X(u+xh)\mathrm{d}u \\
=& 4c\int\int \left|L_w(y_u^*)\right|^2 f_{Y|X}(y|u)\mathrm{d}y z_u z_u^{\mathrm{T}} K^2(u) f_X(u+xh)\mathrm{d}u \\
\leq& 4c\int\left(\int \left|L_w(y_u^*)\right|^{2+\delta} f_{Y|X}(y|u)\mathrm{d}y\right)^{\frac{2}{2+\delta}} z_u z_u^{\mathrm{T}} K^2(u) f_X(u+xh)\mathrm{d}u \\
=& O\left(\mathrm{E}_X\left[(K)^2 z_u z_u^{\mathrm{T}}\right]\right) = O(1)
\end{aligned}
$$

which also implies that $\mathrm{E}_{Y,X}[\boldsymbol{W}_n] = O(1)$ as a result of Jensen's inequality.

Note that

$$G_n(\boldsymbol{\theta}) - \mathbf{W}_n^{\mathrm{T}}\boldsymbol{\theta}$$

is a convex function of $\boldsymbol{\theta}$ which converges in probability to the convex function $\frac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}\gamma(w,x)f_X(x)\mathbf{S}\boldsymbol{\theta}$.

By the convexity lemma, Pollard (1991), for any compact subset $\Lambda \in \mathbb{R}^{p+1}$

$$\sum_{\boldsymbol{\theta}\in\Lambda} |r_n(\boldsymbol{\theta})| = o_P(1).$$

So the quadratic approximation to the convex function $G_n(\boldsymbol{\theta})$ holds uniformly for $\boldsymbol{\theta}$ in any compact set. Then, using the convexity assumption again, the minimizer $\widehat{\boldsymbol{\theta}}$ of $G_n(\boldsymbol{\theta})$ converges in probability to the minimizer of the quadratic function $-(\gamma(\omega,x)f_X(x)\mathbf{S})^{-1}\mathbf{W}_n$

$$\widehat{\boldsymbol{\theta}} + (\gamma(\omega,x)f_X(x)\mathbf{S})^{-1}\mathbf{W}_n = o_P(1).$$

In matrix notation, we have

$$
\sqrt{nh}\begin{pmatrix} \widehat{\beta}_0 - \tau_\omega(x) \\ h(\widehat{\beta}_1 - \tau_\omega^{(1)}(x)) \\ \vdots \\ h^p\left(\widehat{\beta}_p - \frac{\tau_\omega^{(p)}(x)}{p!}\right) \end{pmatrix} - \frac{2}{\sqrt{nh}\gamma(\omega,x)f_X(x)}\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix}^{-1}\left(\sum_{i=1}^{n} L_\omega(Y_i^*)\mathbf{Z}_i K_i\right) = o_P(1)
$$

with

$$\mathbf{W}_n^* = \left(\sum_{i=1}^{n} L_\omega(Y_i^*) \mathbf{Z}_i K_i \right) = \sum_{i=1}^{n} L_\omega(Y_i^*) \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \\ \vdots \\ \left(\frac{X_i - x}{h} \right)^p \end{pmatrix} K\left(\frac{X_i - x}{h} \right)$$

and hence

$$\mathbf{S}^{-1}\mathbf{W}_n^* = \sum_{i=1}^{n} L_\omega(Y_i^*) \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \\ \vdots \\ \left(\frac{X_i - x}{h} \right)^p \end{pmatrix} K\left(\frac{X_i - x}{h} \right).$$

So

$$\sqrt{nh} \begin{pmatrix} \widehat{\beta}_0 - \tau_\omega(x) \\ h(\widehat{\beta}_1 - \tau_\omega^{(1)}(x)) \\ \vdots \\ h^p \left(\widehat{\beta}_p - \frac{\tau_\omega^{(p)}(x)}{p!} \right) \end{pmatrix}$$

$$- \frac{2}{\sqrt{nh}\gamma(\omega, x) f_X(x)} \sum_{i=1}^{n} L_\omega(Y_i^*) \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_p & \mu_{p+1} & \cdots & \mu_{2p} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \frac{X_i - x}{h} \\ \vdots \\ \left(\frac{X_i - x}{h} \right)^p \end{pmatrix} K\left(\frac{X_i - x}{h} \right) = o_P(1). \tag{A.3}$$

The $(j+1)$th component (for $j = 0, 1, \ldots, p$) of the above equality is

$$\sqrt{nh} \left(h^j \left(\frac{\widehat{\tau}_\omega^{(j)}(x)}{j!} - \frac{\tau_\omega^{(j)}(x)}{j!} \right) - \frac{2}{nh\gamma(\omega, x) f_X(x)} \sum_{i=1}^{n} L_\omega(Y_i^*) (\mathbf{S}^{-1}\mathbf{Z}_i)_{j+1} K_i \right) = o_P(1)$$

$$\sqrt{nh} \left(h^j \left(\frac{\widehat{\tau}_\omega^{(j)}(x)}{j!} - \frac{\tau_\omega^{(j)}(x)}{j!} \right) - V_{n,j} \right) = o_P(1),$$

denoting $V_{n,j} = \frac{U_{n,j}}{\gamma(\omega,x) f_X(x) \det(\mathbf{S})}$ and $U_{n,j} = 2(nh)^{-1} \sum_{i=1}^{n} L_\omega(Y_i^*) (\mathrm{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1} K_i$, where $\det(\mathbf{S})$ is the determinant of $\mathbf{S}$ and $\mathrm{adj}(\mathbf{S})$ is the adjugate matrix of $\mathbf{S}$.

Equivalently, for any $\epsilon > 0$, we have

$$\mathrm{E}_{Y,X} \left[\mathbf{P} \left(\sqrt{nh} \left| h^j \left(\frac{\widehat{\tau}_\omega^{(j)}(x)}{j!} - \frac{\tau_\omega^{(j)}(x)}{j!} \right) - V_{n,j} \right| \geq \epsilon \middle| \mathcal{X} \right) \right] = o_P(1).$$

This implies that

$$\mathbf{P}\left(\sqrt{nh}\left|h^j\left(\frac{\widehat{\tau}_\omega^{(j)}(x)}{j!}-\frac{\tau_\omega^{(j)}(x)}{j!}\right)-V_{n,j}\right|\geq\epsilon\middle|\mathcal{X}\right)=o_P(1).$$

Hence, the conditional asymptotic normality follows from that of $U_{n,j}$, which is established with the help of Lemmas 1 and 2 stated in Section A.2. The proofs of the lemmas are provided in Section S6 of the Supplementary Material part. □

A.2 Two lemmas

Lemma 1 *Under the assumptions of Theorem 1, we have*

$$\mathrm{E}_{Y|X}[U_{n,j}|\mathcal{X}]=dh^{p+1}(1+o_P(1)) \qquad \text{and} \qquad \mathrm{Var}_{Y|X}[U_{n,j}|\mathcal{X}]=\frac{v^2}{nh}(1+o_P(1))$$

where $U_{n,j}=2(nh)^{-1}\sum_{i=1}^n L_\omega(Y_i^*)(\mathrm{adj}(\mathbf{S})\mathbf{Z}_i)_{j+1}K_i$,

$$\begin{aligned}d=&\frac{1}{(p+1)!}\tau_\omega^{(p+1)}(x)\gamma(\omega,x)(\mathrm{adj}(\mathbf{S})\mathbf{c}_p)_{j+1}f_X(x)+\frac{1}{(p+2)!}\tau_\omega^{(p+2)}(x)h\gamma(\omega,x)(\mathrm{adj}(\mathbf{S})\widetilde{\mathbf{c}}_p)_{j+1}f_X(x)\\&+\frac{1}{(p+1)!}\tau_\omega^{(p+1)}(x)h\gamma(\omega,x)(\mathrm{adj}(\mathbf{S})\widetilde{\mathbf{c}}_p)_{j+1}f_X^{(1)}(x)\end{aligned} \tag{A.4}$$

$$\text{and} \qquad v^2=f_X(x)\int\left(2L_\omega(y-\tau_\omega(x))\right)^2 f_{Y|X}(y|x)\mathrm{d}y\int(\mathrm{adj}(\mathbf{S})\mathbf{z}_v)_{j+1}^2K^2(v)\mathrm{d}v \tag{A.5}$$

with $\mathbf{z}_v=\left(1,v,v^2,\cdots,v^p\right)^{\mathrm{T}}$.

Lemma 2 *Under Assumptions* (A1)—(A5), *we have*

$$P\left[\sqrt{nh}\frac{(U_{n,j}-dh^{p+1})}{v}\leq t|\mathcal{X}\right]=\boldsymbol{\Phi}(t)+o_P(1),$$

with d and v define as in (A.4) *and* (A.5), *respectively.*

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-021-00799-y.

Acknowledgements The authors are grateful to an Associate Editor and two reviewers for the very valuable comments which led to an improvement of the paper. The authors gratefully acknowledge support of Research Grant FWO G0D6619N from the Flemish Science Foundation and of GOA/12/014 and C16/20/002 projects from the Research Fund KU Leuven.

References

Bellini, F., D Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance, 23*(6), 487–506.

Bellini, F., Klar, B., Müller, A., Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics, 54,* 41–48.

Breckling, J., Chambers, R. (1988). M-quantiles. *Biometrika, 75*(4), 761–771.

Chen, J., Shao, J. (1993). Iterative weighted least squares estimators. *The Annals of Statistics, 21*(2), 1071–1092.

De Rossi, G., Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics, 152,* 179–185.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica, 1,* 93–125.

Fan, J., Gijbels, I. (1995). Adaptive order polynomial fitting: bandwidth robustification and bias reduction. *Journal of Computational and Graphical Statistics, 4*(3), 213–227.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications. Number 66 in monographs on statistics and applied probability series*. London: Chapman & Hall.

Fan, J., Hu, T., Truong, Y. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics, 21*(4), 433–446.

Fredriks, A., van Buuren, S., Burgmeijer, R., Meulmeester, J., Beuker, R., Brugman, E., et al. (2000). Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Research, 47*(3), 316–323.

Gijbels, I., Karim, R., Verhasselt, A. (2019). On quantile-based asymmetric family of distributions: Properties and inference. *International Statistical Review, 87*(3), 471–504.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

Huber, P., Ronchetti, E. (2009). *Robust statistics* (2nd ed.). New Jersey: Wiley.

Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters, 20*(2), 149–153.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge: Cambridge University Press.

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50.

Krätschmer, V., Zähle, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics, 44*(2), 425–454.

Newey, W., Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica, 55*(4), 819–847.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory, 7*(2), 186–199.

Remillard, B., Abdous, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics, 47,* 371–384.

Schnabel, S., Eilers, P. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis, 53*(12), 4168–4177.

Schulze Waltrup, L., Sobotka, F., Kneib, T., Kauermann, G. (2015). Expectile and quantile regression-David and Goliath? *Statistical Modelling, 15*(5), 433–456.

Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics, 6*(2), 231–252.

Wand, M., Jones, M. (1995). *Kernel smoothing*. London: Chapman and Hall.

Wolke, R., Schwetlick, H. (1988). Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons. *SIAM Journal on Scientific and Statistical Computing, 9*(5), 907–921.

Yang, Y., Zou, H. (2015). Nonparametric multiple expectile regression via er-boost. *Journal of Statistical Computation and Simulation, 85*(7), 1442–1458.

Yao, Q., Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics, 6*(2), 273–292.

Yu, K., Jones, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association, 93*(441), 228–237.

Zhang, L., Mei, C. (2008). Testing heteroscedasticity in nonparametric regression models based on residual analysis. *Applied Mathematics, 23,* 265–272.

Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance, 26*(4), 901–918.