Semiparametric quantile regression using family of quantile-based asymmetric densities

Non Peer-reviewed author version

Gijbels, Irene; KARIM, Rezaul & VERHASSELT, Anneleen (2021) Semiparametric quantile regression using family of quantile-based asymmetric densities. In: COMPUTATIONAL STATISTICS & DATA ANALYSIS, 157 (Art N° 107129).

DOI: 10.1016/j.csda.2020.107129
Handle: http://hdl.handle.net/1942/33968

Semiparametric quantile regression using family of quantile-based asymmetric densities

Irène Gijbels[a], Rezaul Karim[a,b,c], Anneleen Verhasselt [1c]

[a] *Department of Mathematics and Leuven Statistics Research Center (LStat), KU Leuven*
[b]*Department of Statistics, Jahangirnagar University*
[c]*Data Science Institute, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Universiteit Hasselt*

Abstract

Quantile regression is an important tool in data analysis. Linear regression, or more generally, parametric quantile regression imposes often too restrictive assumptions. Nonparametric regression avoids making distributional assumptions, but might have the disadvantage of not exploiting distributional modeling elements that might be brought in. A semiparametric approach towards estimating conditional quantile curves is proposed. It is based on a recently studied large family of asymmetric densities of which the location parameter is a quantile (and not a mean). Passing to conditional densities and exploiting local likelihood techniques in a multiparameter functional setting then leads to a semiparametric estimation procedure. For the local maximum likelihood estimators the asymptotic distributional properties are established, and it is discussed how to assess finite sample bias and variance. Due to the appealing semiparametric framework, one can discuss in detail the bandwidth selection issue, and provide several practical bandwidth selectors. The practical use of the semiparametric method is illustrated in the analysis of maximum winds speeds of hurricanes in the North Atlantic region, and of bone density data. A simulation study includes a comparison with nonparametric local linear quantile regression as well as an investigation of robustness against miss-specifying the parametric model part.

Keywords: Asymptotic distribution, Bandwidth selection, Local likelihood, Local polynomial fitting

[1]Corresponding author: Data Science Institute, Universiteit Hasselt, Agoralaan D, 3590 Diepenbeek, Belgium; anneleen.verhasselt@uhasselt.be; tel.+32 11268278.

1. Introduction

Classical regression focuses on estimation of the conditional mean function $\mathbb{E}(Y|\mathbf{X})$ of a response Y given a set of d covariates $\mathbf{X} = (X_1, ..., X_d)^T$. A vast literature dealing with estimation of $\mathbb{E}(Y|\mathbf{X})$ via parametric, semi-parametric or nonparametric approaches is available. In the context of non-parametric mean regression, [1] developed local polynomial estimators in an extended generalized linear models framework, using quasi-likelihood methods under standard regularity conditions. Background information on local polynomial modeling and detailed discussions on a variety of applications can be found in [2]. Maximum likelihood estimation is among the key tools in statistics, and it provides a unified method for constructing approximate confidence intervals for parameters. It requires the specification of a particular conditional density function for the response variable given the covariate vector, but the distributional assumption, when justified, allows to draw statistically more accurate conclusions. [3] established a general framework to use the maximum likelihood technique and extending its scope towards non-parametric estimation. Their technique of localizing a likelihood and employing local polynomial fitting, together with the outlined statistical inference steps is widely applicable.

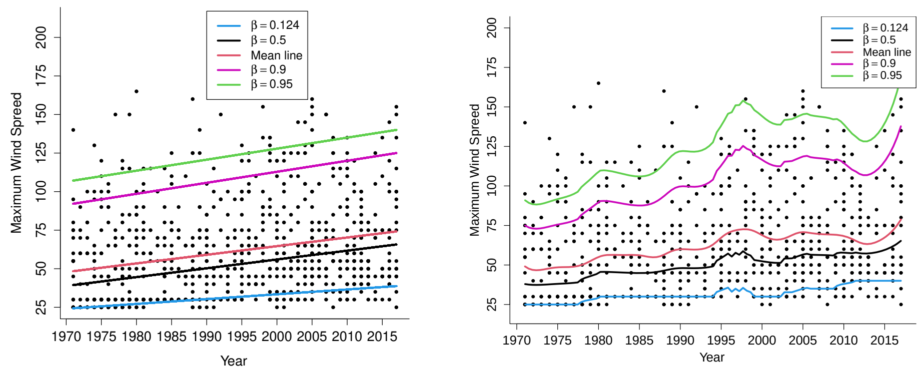
Conditional mean estimation focuses only on the average effect of the response Y given $\mathbf{X}$, and is not a very appropriate measure of central position in case of a skewed conditional response distribution. A conditional mean is just one characteristic of the conditional distribution, whereas the conditional quantile function fully characterizes it. Quantile curves are an important tool in, for example, environmental studies where upper quantiles of pollution levels are critical from a public health perspective.

The conditional quantile of order β (with $0 < \beta < 1$) of Y given $\mathbf{X} = \mathbf{x}$ is denoted and defined as $Q_\beta(Y \mid \mathbf{X} = \mathbf{x}) = \inf_y\{y : F_{Y|\mathbf{X}}(y \mid \mathbf{x}) \geqslant \beta\}$, with $F_{Y|\mathbf{X}}(\cdot \mid \mathbf{x})$ the cumulative distribution function of Y given $\mathbf{X} = \mathbf{x}$. See [4] and [5] for a comprehensive overview of the area of quantile regression. Throughout the paper we use the shorthand notation $q_\beta(\mathbf{x})$ for $Q_\beta(Y \mid \mathbf{X} = \mathbf{x})$. The conditional quantile of Y given $\mathbf{X} = \mathbf{x}$ coincides with the minimizer of $\mathbb{E}_{Y|\mathbf{X}}\{\rho_\beta(Y - a) \mid \mathbf{X} = \mathbf{x}\}$ with respect to a, where $\rho_\beta(u) = u(\beta - \mathbb{I}(u < 0))$ is the so-called check function. Henceforth $q_\beta(\mathbf{x})$ is such that $\mathbb{E}_{\mathbf{X},Y}\left[\rho_\beta\left(Y - q_\beta(\mathbf{X})\right)\right]$ is minimal.

For estimating a conditional quantile $q_\beta(\mathbf{x})$ one can rely on parametric, semiparametric or nonparametric approaches. In linear quantile regression $q_\beta(\mathbf{x})$ is modelled as a linear function, i.e. $q_\beta(\mathbf{x}) = \boldsymbol{\theta}^T\widetilde{\mathbf{x}}$, with $\widetilde{\mathbf{x}} = (1, \mathbf{x}^T)^T$ the $(d+1)$-dimensional column vector, and with $\boldsymbol{\theta}$ the column vector (of dimension $d+1$) of unknown regression coefficients (including an intercept). Parametric quantile regression estimation is performed by considering the empirical version of $\mathbb{E}_{\mathbf{X},Y}\left[\rho_\beta\left(Y - q_\beta(\mathbf{X})\right)\right]$, and estimated parameters are those for which this empirical quantity is minimized. See also Section 2. Linear or, more generally, parametric quantile regression can be insufficient, in case no appropriate form for the conditional quantiles can be put forward. Consider data on maximum wind speeds in hurricanes occurring in the North Atlantic region during the period 1971 and 2017. Figure 1(a) depicts the data together with some estimated linear regression curves. From this figure it appears as if the maximum wind speed (in knots per hour) of the strongest hurricanes in the North Atlantic region have increased over the whole period. Figure 1(b) (produced using the proposed semiparametric quantile estimation method) reveals however that also decreases are noticeable, even for several quantile curves, during that period. As discussed in Section 8 the semiparametric estimation method led to considerably smaller prediction errors than when using a nonparametric estimation method (see Figure 9). This example simply illustrates the specific merits of the semiparametric approach that is presented in this paper.

Nonparametric approaches towards quantile regression include the one of [6], who developed a local linear quantile estimation method, similar as for mean regression. Following-up on this, [7] further contributed to bandwidth selection for local linear quantile regression. They also introduce the local linear double-kernel smoothing method, in which estimation of the conditional cumulative distribution function, is followed by estimation of the conditional quantile function via an inversion technique. Such an inversion procedure involves the choice of two smoothing parameters. An extensive discussion on the two approaches can be found in [7].

A particular point of attention when estimating conditional quantile curves is that of the non-crossing property. By definition, for $0 < \beta_1 \leqslant \beta_2 < 1$, the conditional quantile curves satisfy $q_{\beta_1}(\mathbf{x}) \leqslant q_{\beta_2}(\mathbf{x})$, for all $\mathbf{x}$ in the domain of the random vector $\mathbf{X}$. Estimated conditional quantile curves should also (among others for interpretability reasons) satisfy this property. But, the local linear quantile estimation or the local linear double-kernel smoothing estimator do not necessarily satisfy this property unless extra precautions



(a) (b)

Figure 1: Maximum wind speeds of hurricanes in the North Atlantic region during the period 1971–2017. (a) Estimated linear quantile curves; (b) Estimated semiparametric quantile curves.

are taken (see [7]).

In this paper we contribute with an appealing semiparametric method to estimate conditional quantile curves. As a starting point we rely on a very broad family of asymmetric densities, recently studied in [8]. Within this large family of densities, with index-parameter α ($0 < \alpha < 1$), the location parameter (say μ) coincides with the αth quantile of the distribution. It therefore is called the quantile-based family of asymmetric densities (QBA densities). This family provides a very advantageous framework, since many probabilistic properties, and a detailed study of estimators and their behaviour (with explicit expressions for asymptotic variance-covariance matrices) were established in [8]. Moreover, the family is of a location-scale type. A density in the family is symmetric if and only if $\alpha = 0.5$. Based on this family we consider a class of asymmetric conditional densities that involves an unknown location function $\mu(\mathbf{x})$ and an unknown scale function $\phi(\mathbf{x})$. For a given member of the family of conditional densities (constituting the parametric component), one can produce a localized version of the log-likelihood, locally modelling both the unknown location and scale function (the nonparametric components) via polynomials (of possibly different degree). This results into a local polynomial likelihood type of problem, but under nonstandard working conditions (i.e. non-differentiability), due to the quantile-based setting. Only in case $\alpha = 0.5$ we are back to standard working conditions. The advantages of this particular semiparametric approach are:

- due to the QBA-based framework, the estimated quantile curves inherently satisfy the non-crossing property;
- due to the key ingredients of the framework, a detailed study of the estimation methods, including asymptotic distributional results, theoretical optimal bandwidths and data-driven bandwidth selectors, finite-sample assessments of bias and variance of the estimators, as well as construction of confidence intervals and bands can be provided in the generic setting (for all members of the large QBA family);
- when the asymmetric conditional density is well-specified the method largely outperforms the nonparametric local polynomial quantile estimation method;
- the nonparametric local polynomial quantile estimation context links up to a special case of the considered estimation method; and as a side product we also contribute to the area of nonparametric quantile regression.

In the present paper we restrict to a univariate covariate setting (i.e. $d = 1$) although extension to a multivariate setting is methodologically rather straightforward, as is briefly discussed in Section 3.3.

The paper is further organized as follows. After a very brief recall of parametric and nonparametric approaches to conditional quantile estimation in Section 2, we present our semiparametric local likelihood estimation type approach in Section 3. Section 4 contains the asymptotic results of the semiparametric local likelihood estimator. The important issue of bandwidth choice is discussed in Section 5. Section 6 is devoted to the adaptation of the method to account for some additional unknown parameter (function). The finite-sample performance of the semiparametric procedure is investigated via a simulation study in Section 7. Real data applications in Section 8 illustrate the use of the proposed semiparametric estimation method. Proofs of all theoretical results are provided in the Supplementary Material. This material also contains new asymptotic results for nonparametric local polynomial quantile regression and optimal bandwidth choice for it. These results fill some gap in the literature. The Supplementary material further presents some additional results from the simulation study. The discussed semiparametric estimation method is implemented in the R package `QBAsyDist` [9].

2. Parametric and nonparametric conditional quantile estimation

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample from (X, Y). In parametric settings one assumes that the βth conditional quantile of Y given $X = x$ (for arbitrary $0 < \beta < 1$) takes on a parametric form, for example, $q_\beta(x) = \theta_{\beta,0} + \theta_{\beta,1}x$ in linear quantile regression or $q_\beta(x) = \theta_{\beta,0} + \theta_{\beta,1}x + \cdots + \theta_{\beta,p}x^p$ in pth order polynomial quantile regression ($p \in \mathbb{N}$). Since $q_\beta(x)$ minimizes $\mathbb{E}_{X,Y}\left[\rho_\beta\left(Y - q_\beta(X)\right)\right]$, the parameter $\boldsymbol{\theta}_\beta = (\theta_{\beta,0}, ..., \theta_{\beta,p})^T$ in a parametric polynomial setting can be estimated by minimizing

$$\sum_{i=1}^{n} \rho_\beta(Y_i - \theta_{\beta,0} - \cdots - \theta_{\beta,p}X_i^p), \tag{1}$$

with the check function $\rho_\beta(u) = u(\beta - \mathbb{I}(u < 0))$ as in Section 1.

In a nonparametric setting, the functional form of $q_\beta(x)$ is completely unknown. Similar as in nonparametric mean regression, the βth quantile function $q_\beta(x)$ can be estimated using local modelling, and a locally kernel-weighted check loss function. In this case no distributional assumption is made on the conditional distribution of Y given X. See e.g. [6] and [7]. The idea of a local polynomial fit is to approximate the unknown βth quantile function $q_\beta(x_0)$, for x_0 given, by a pth order polynomial, i.e. for z in a neighbourhood of x_0

$$\begin{aligned} q_\beta(z) &\approx q_\beta(x_0) + q'_\beta(x_0)(z - x_0) + \cdots + \frac{1}{p!}q_\beta^{(p)}(x_0)(z - x_0)^p \\ &\equiv \theta_{\beta,0} + \theta_{\beta,1}(z - x_0) + \cdots + \theta_{\beta,p}(z - x_0)^p, \end{aligned} \tag{2}$$

where $q_\beta^{(v)}(x_0)$ denotes the vth derivative of the function q_β evaluated in the point x_0, and with $\theta_{\beta,v} = q_\beta^{(v)}(x_0)/v!$, for $v = 0, 1, \ldots, p$. For all observations X_i close to x_0, we can apply the Taylor approximation for $q_\beta(X_i)$ and obtain from (2) that

$$q_\beta(X_i) \approx \theta_{\beta,0} + \theta_{\beta,1}(X_i - x_0) + \cdots + \theta_{\beta,p}(X_i - x_0)^p \equiv \mathbf{X}_{i,p}^T\boldsymbol{\theta}_\beta,$$

where $\mathbf{X}_{i,p} = (1, X_i - x_0, \ldots, (X_i - x_0)^p)^T$ and $\boldsymbol{\theta}_\beta = (\theta_{\beta,0}, \ldots, \theta_{\beta,p})^T$. Since the Taylor expansion in (2) is only valid in a neighbourhood of x_0, the loss function in (1) needs to be localized as to contain only contributions from observations X_i that are close to x_0. This is done by a weighting factor $K_h(X_i - x_0)$, where $K_h(\cdot) = K(\cdot/h)/h$ is a rescaling of $K(\cdot)$, a compactly supported symmetric probability density, and $h > 0$ a bandwidth parameter determining the size of the neighbourhood. The unknown parameter $\boldsymbol{\theta}_\beta$ is estimated by

$$\widehat{\boldsymbol{\theta}}_\beta(x_0) = \arg \min_{\boldsymbol{\theta}_\beta \in \mathbb{R}^{(p+1)}} \sum_{i=1}^{n} \rho_\beta(Y_i - \mathbf{X}_{i,p}^T \boldsymbol{\theta}_\beta) K_h(X_i - x_0). \tag{3}$$

The local polynomial estimator $\widehat{q}_{\beta,v}(x_0)$ for $q_\beta^{(v)}(x_0)$, $v = 0, 1, \ldots, p$ is then given by $\widehat{q}_{\beta,v}(x_0) = v!\,\widehat{\theta}_{\beta,v}(x_0)$. Local linear estimation of the conditional quantile $q_\beta(x_0)$ corresponds to taking $p = 1$ in the above.

3. Semiparametric conditional quantile estimation

In this section we present the framework for the semiparametric estimation approach. We first review the quantile-based family of asymmetric densities which is a basic element.

3.1. Quantile-based family of asymmetric densities

Consider f a symmetric around 0 density. We assume f to be unimodal. We call f the reference symmetric density. Denote by F and F^{-1} the cumulative distribution function and the quantile function associated with f. The QBA family, indexed by a parameter α, $\alpha \in (0,1)$, with parameters $\mu \in I\!R$, and $\phi \in I\!R_+$, is then defined by

$$\widetilde{f}_\alpha(y; \mu, \phi) = \frac{2\alpha(1-\alpha)}{\phi} \begin{cases} f\big((1-\alpha)(\frac{\mu - y}{\phi})\big) & \text{if} \quad y \leqslant \mu \\ f\big(\alpha(\frac{y-\mu}{\phi})\big) & \text{if} \quad y > \mu. \end{cases} \tag{4}$$

Herein the index-parameter α controls the allocation of mass to the left and the right of the mode μ. The density $\widetilde{f}_\alpha(y; \mu, \phi)$ in (4) is symmetric if and only if $\alpha = 0.5$. The density is left-skewed (respectively right-skewed) if $\alpha > 0.5$ (respectively $\alpha < 0.5$). The QBA family constitutes a very broad family of asymmetric densities, including (new) asymmetric normal, Student-t, logistic and Laplace densities (see [8]). The important properties of (4) are:

(*i*) μ is the αth quantile of Y, i.e. $\int_{-\infty}^{\mu} \widetilde{f}_\alpha(y; \mu, \phi) dy = \alpha$ and $\int_{\mu}^{\infty} \widetilde{f}_\alpha(y; \mu, \phi) dy = 1 - \alpha$.

(*ii*) Mean and variance (if they exist) are given by:

$$\mathbb{E}(Y) = \mu + \frac{\phi(1-2\alpha)\mu_1}{\alpha(1-\alpha)} \tag{5}$$

$$\mathrm{Var}(Y) = \frac{\phi^2}{\alpha^2(1-\alpha)^2}[(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2], \tag{6}$$

where $\mu_r = 2\int_0^\infty s^r f(s)ds$ is a moment-type quantity of the reference density f.

(*iii*) Cumulative distribution function:

$$\widetilde{F}_\alpha(y;\mu,\phi) = \begin{cases} 2\alpha F\Big((1-\alpha)(\frac{y-\mu}{\phi})\Big) & \text{if} \quad y < \mu \\ 2\alpha - 1 + 2(1-\alpha)F\Big(\alpha(\frac{y-\mu}{\phi})\Big) & \text{if} \quad y \geqslant \mu. \end{cases}$$

(*iv*) The quantile function: for any β (with $0 < \beta < 1$), $\widetilde{F}_\alpha^{-1}(\beta) = \mu + \phi \cdot C_\alpha(\beta)$, where $C_\alpha(\beta) = \frac{1}{1-\alpha}F^{-1}\big(\frac{\beta}{2\alpha}\big)\mathbb{I}(\beta < \alpha) + \frac{1}{\alpha}F^{-1}\Big(\frac{1+\beta-2\alpha}{2(1-\alpha)}\Big)\mathbb{I}(\beta \geqslant \alpha)$.

Note that the quantity $C_\alpha(\beta)$ depends on β, the index-parameter α and the quantile function of the reference symmetric density f. For example, for f a symmetric Laplace density: $C_\alpha(\beta) = \frac{1}{1-\alpha}\ln\big(\frac{\beta}{\alpha}\big) - \frac{1}{\alpha}\ln\big(\frac{1-\beta}{1-\alpha}\big)$. In general, it holds that for $\beta = \alpha$, $C_\alpha(\alpha) = 0$ and consequently $\widetilde{F}_\alpha^{-1}(\alpha) = \mu$. Obviously, $C_\alpha(\beta)$ is an increasing function of β.

Given an i.i.d. sample $Y_1, \ldots, Y_n$ from $Y \sim \widetilde{f}_\alpha(\cdot;\mu,\phi)$, [8] discussed maximum likelihood estimation of the parameter vector (μ,ϕ), and established that the Fisher information matrix for the maximum likelihood estimator of (μ,ϕ) equals

$$\widetilde{\mathcal{I}}(\mu,\phi) = \begin{bmatrix} \frac{2\alpha(1-\alpha)\gamma_1}{\phi^2} & 0 \\ 0 & \frac{1}{\phi^2}(2\gamma_3 - 1) \end{bmatrix},$$

where $\gamma_\ell = \int s^{\ell-1}f(s)ds$, for $\ell = 1, 2, 3$ (γ_ℓ is assumed to be finite). For example, for f the symmetric Laplace density, we have $\gamma_1 = \gamma_2 = \dfrac{1}{2}$ and $\gamma_3 = 1$. For detailed results, see Proposition 3.2 in [8].

Since $\phi \in I\!R_+$ it is important to obtain an estimator that takes on only non-negative values. This is of particular importance when passing to the regression setting in which ϕ is a function of x. In order to automatically obtain non-negative estimators for ϕ (or $\phi(x)$ in the regression setting) it is advantageous to reparametrize the above QBA family as follows. Using the one-to-one transformation $(\theta_1, \theta_2) = (\mu, \ln\{\phi\})$ and denoting the resulting asymmetric density by $f_\alpha(\cdot;\theta_1,\theta_2)$ we get

$$f_\alpha(y;\theta_1,\theta_2) = \widetilde{f}_\alpha(y;\theta_1,\exp\{\theta_2\}) = \frac{2\alpha(1-\alpha)}{\exp\{\theta_2\}} \begin{cases} f\big((1-\alpha)(\frac{\theta_1-y}{\exp\{\theta_2\}})\big) & \text{if } y \leqslant \theta_1 \\ f\big(\alpha(\frac{y-\theta_1}{\exp\{\theta_2\}})\big) & \text{if } y > \theta_1. \end{cases} \tag{7}$$

The quantile function of a random variable Y with density $f_\alpha(\cdot;\theta_1,\theta_2)$ is then

$$F_\alpha^{-1}(\beta) = \theta_1 + \exp(\theta_2)\cdot C_\alpha(\beta).$$

The Fisher information matrix is calculated from the second order partial derivatives of the log-density. Denoting $v_1 = u_1$ and $v_2 = \exp\{u_2\}$, and applying the chain rule we get, for $r, s \in \{1,2\}$,

$$\begin{aligned} \frac{\partial}{\partial u_r}\ln f_\alpha(y;u_1,u_2) &= \frac{\partial}{\partial v_r}\ln \widetilde{f}_\alpha(y;v_1,v_2)\frac{\partial v_r}{\partial u_r} \\ \frac{\partial^2}{\partial u_r \partial u_s}\ln f_\alpha(y;u_1,u_2) &= \frac{\partial^2}{\partial v_r \partial v_s}\ln \widetilde{f}_\alpha(y;v_1,v_2)\frac{\partial v_r}{\partial u_r}\frac{\partial v_s}{\partial u_s}. \end{aligned} \tag{8}$$

Since

$$\frac{\partial v_1}{\partial u_1} = 1 \qquad \text{and} \qquad \frac{\partial v_2}{\partial u_2} = v_2,$$

the Fisher information matrix for the maximum likelihood estimators for (θ_1,θ_2) in (7) is

$$\mathcal{I}(\theta_1,\theta_2) = \begin{bmatrix} \frac{2\alpha(1-\alpha)\gamma_1}{\exp\{2\theta_2\}} & 0 \\ 0 & (2\gamma_3 - 1) \end{bmatrix}. \tag{9}$$

3.2. Semiparametric local polynomial maximum likelihood conditional quantile estimation

We now turn to the regression setting involving one covariate. For the conditional density of Y given $X = x$ we consider the density $f_{Y|X}(\cdot;\theta_1(x),\theta_2(x))$ in (7) and allow θ_1 and θ_2 to depend on x, for given index-parameter α. This leads to the conditional density

$$f_{Y|X,\alpha}(y;\theta_1(x),\theta_2(x)|x) = \frac{2\alpha(1-\alpha)}{\exp(\theta_2(x))} \begin{cases} f\big((1-\alpha)(\frac{\theta_1(x)-y}{\exp(\theta_2(x))})\big) & \text{if } y \leqslant \theta_1(x) \\ f\big(\alpha(\frac{y-\theta_1(x)}{\exp(\theta_2(x))})\big) & \text{if } y > \theta_1(x), \end{cases} \tag{10}$$

for which the βth quantile function is

$$q_\beta(x) = F^{-1}_{Y|X,\alpha}(\beta|x) = \theta_1(x) + \exp\{\theta_2(x)\} \cdot C_\alpha(\beta).$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample from (X, Y) with conditional density (10), where $\theta_1(x)$ and $\theta_2(x)$ are unknown real-valued functions. Given estimates $\widehat{\theta}_1(x)$ and $\widehat{\theta}_2(x)$ for $\theta_1(x)$ and $\theta_2(x)$, an estimator for the conditional quantile function is

$$\widehat{q}_\beta(x) = \widehat{\theta}_1(x) + \exp\left\{\widehat{\theta}_2(x)\right\} \cdot C_\alpha(\beta). \tag{11}$$

We estimate $\theta_1(x)$ and $\theta_2(x)$, starting from the localized conditional log-likelihood.

If $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are known, the conditional log-likelihood is

$$\sum_{i=1}^{n} \ln f_{Y|X,\alpha}(Y_i; \theta_1(X_i), \theta_2(X_i) \mid X = X_i)$$

Since $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are unknown, we use a local polynomial fitting technique. For given x_0, we approximate $\theta_r(z)$ (for $r \in \{1, 2\}$) with z in the neighbourhood of x_0, using a Taylor expansion of order p_r ($p_r \in I\!N$):

$$\theta_r(z) \approx \theta_r(x_0) + \theta_r'(x_0)(z - x_0) + \cdots + \frac{\theta_r^{(p_r)}(x_0)}{p_r!}(z - x_0)^{p_r} \equiv \sum_{j=0}^{p_r} \theta_{rj}(z - x_0)^j,$$

with $\theta_{rv} = \frac{\theta_r^{(v)}(x_0)}{v!}; v = 0, 1, \ldots, p_r$. For the conditional log-likelihood this means that only data points (X_i, Y_i) for which X_i is close to x_0 contribute to the localized version of it. Denoting $\ell(\theta_1(X_i), \theta_2(X_i); Y_i) =$ $\ln f_{Y|X,\alpha}(Y_i, \theta_1(X_i), \theta_2(X_i) \mid X_i)$, and using $\theta_r(X_i) \approx \sum_{j=0}^{p_r} \theta_{rj}(X_i - x_0)^j \equiv$ $\mathbf{X}^T_{i,p_r}\boldsymbol{\theta}_r$, where $\mathbf{X}_{i,p_r} = (1, (X_i - x_0), \cdots, (X_i - x_0)^{p_r})^T$ and $\boldsymbol{\theta}_r = (\theta_{r0}, \cdots, \theta_{rp_r})^T$, we get to the *local kernel-weighted conditional log-likelihood*

$$\mathcal{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; h, x_0) = \sum_{i=1}^{n} \ell(\mathbf{X}^T_{i,p_1}\boldsymbol{\theta}_1, \mathbf{X}^T_{i,p_2}\boldsymbol{\theta}_2; Y_i) K_h(X_i - x_0), \tag{12}$$

This local kernel-weighted conditional log-likelihood needs to be maximized with respect to $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, leading to the vector of estimators $(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0)) = \left((\widehat{\theta}_{10}(x_0), ..., \widehat{\theta}_{1p_1}(x_0))^T, (\widehat{\theta}_{20}(x_0), ..., \widehat{\theta}_{2p_2}(x_0))^T\right)$ defined as

$$(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0)) = \arg\max_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} \mathcal{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; h, x_0) \tag{13}$$

The estimator $\widehat{\theta}_r^{(v)}(x_0)$ for $\theta_r^{(v)}(x_0)$, $v = 0, 1, \ldots, p_r$, is given by $\widehat{\theta}_r^{(v)}(x_0) = v!\,\widehat{\theta}_{rv}(x_0)$. The whole function $\theta_r(\cdot)$ (for $r \in \{1, 2\}$) is estimated by considering a grid of x_0-values and solving maximization problem (13) for each point in the grid.

Illustrative example: the conditional asymmetric Laplace density. Taking as a symmetric reference density, the standard Laplace density, i.e. $f(s) = 0.5e^{-|s|}$, the contribution from a data point (X_i, Y_i) to the local conditional likelihood is

$$
\begin{aligned}
& \ell(\mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1, \mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2; Y_i) \\
= \; & \ln\left(\alpha(1-\alpha)\right) - \mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2 \\
& - \frac{1}{\exp[\mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2]} \Big[(1-\alpha)(\mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1 - Y_i)\mathbb{I}(Y_i - \mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1 \leqslant 0) \\
& \qquad\qquad + \alpha(Y_i - \mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1)\mathbb{I}(Y_i - \mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1 > 0) \Big] \\
= \; & \ln\left(\alpha(1-\alpha)\right) - \mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2 - \frac{1}{\exp[\mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2]} \rho_\alpha(Y_i - \mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1),
\end{aligned}
$$

since $\rho_\beta(u) = u(\beta - \mathbb{I}(u < 0)) = u\left[(1-\beta)\mathbb{I}(u \leqslant 0) + \beta\mathbb{I}(u > 0)\right]$.

A special situation occurs when we take $p_2 = 0$, and hence approximate $\theta_2(z)$ locally by a constant, and $\boldsymbol{\theta}_2 = \theta_{20}$. The solution to the maximization problem (13) is then

$$
\begin{cases}
\widehat{\boldsymbol{\theta}}_1(x_0) = \arg \min\limits_{\boldsymbol{\theta}_1 \in I\!R^{(p_1+1)}} \sum\limits_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1) K_h(X_i - x_0), \\
\widehat{\theta}_2(x_0) = \ln \left[\dfrac{\sum\limits_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{X}_{i,p_1}^T \widehat{\boldsymbol{\theta}}_1(x_0)) K_h(X_i - x_0)}{\sum\limits_{i=1}^{n} K_h(X_i - x_0)} \right].
\end{cases}
\tag{14}
$$

If we are in a setting of a constant scale function, i.e. $\theta_2(x) = \theta_2$, for all x, then an overall estimator for θ_2 is $\widehat{\theta}_2 = \sum\limits_{j=1}^{n_{\text{grid}}} \widehat{\theta}_2(x_j)/n_{\text{grid}}$; where n_{grid} is the total number of grid points for which the optimization problem is carried out.

Using the estimators $\widehat{\boldsymbol{\theta}}_1(x_0)$ and $\widehat{\boldsymbol{\theta}}_2(x_0)$, and substituting these estimators into (11), with the appropriate constant $C_\alpha(\beta) = \frac{1}{1-\alpha} \ln\left(\frac{\beta}{\alpha}\right) - \frac{1}{\alpha} \ln\left(\frac{1-\beta}{1-\alpha}\right)$ for the case of a symmetric Laplace reference density, we obtain the estimated βth quantile function. See further Section S1.

Remark 3.1. An important remark is that the minimization problem leading to the estimator $\widehat{\boldsymbol{\theta}}_1(x_0)$ in (14) coincides with the minimization problem in (3) provided $\beta = \alpha$ in the latter problem (or $\alpha = \beta$ in the former). In other words, when the scaling function is constant and known, the considered semiparametric estimator for the βth quantile $\boldsymbol{\theta}_1(x_0)$ coincides with the fully nonparametric αth order quantile estimator. This of course does not mean that the asymptotic properties of both estimators coïncide, since the model assumptions are different, and come into play when investigating asymptotic behaviour.

Remark 3.2. In the procedures described above, we take α fixed. Estimation of α or $\alpha(x)$ is however discussed in Section 6, and in data applications α (and $\alpha(x)$) is estimated.

3.3. Extension to the multivariate covariate setting

The methodology presented in Section 3.2 is rather straightforward to generalize to a multivariate setting. Consider a vector of covariates $\mathbf{X} = (X_1, \ldots, X_d)^T$. The task is to estimate the d-variate location and scale functions $\theta_1(\mathbf{x})$ and $\theta_2(\mathbf{x})$ (with $\mathbf{x} = (x_1, \cdots, x_d)^T$), based on an i.i.d. sample $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from $(\mathbf{X}, Y)$. Relying on the idea of local modelling, we can approximate the unknown d-variate functions locally using Taylor expansion for d-variate functions.

For simplicity of presentation, we only briefly discuss the extension of the methodology considering local linear fitting (i.e. $p_1 = p_2 = 1$) in this multidimensional covariate setting. For a given value $\mathbf{x}_0 = (x_{01}, \ldots, x_{0d})^T$, the function $\theta_r(\mathbf{z})$ can, for $\mathbf{z} = (z_1, \ldots, z_d)^T$ in a neighbourhood of $\mathbf{x}_0$, be approximated by

$$\theta_r(\mathbf{z}) \approx \theta_r(\mathbf{x}_0) + \sum_{j=1}^{d} \frac{\partial \theta_r(\mathbf{x})}{\partial x_j}\Big|_{\mathbf{x}=\mathbf{x}_0} (z_j - x_{0j}) \equiv \theta_{r0} + \sum_{j=1}^{d} \theta_{rj}(z_j - x_{0j}) \quad r \in \{1, 2\},$$

using a Taylor expansion up to order one for a d-variate function, and denoting $\theta_{r0} = \theta_r(\mathbf{x}_0)$ and $\theta_{rj} = \frac{\partial \theta_r(\mathbf{x})}{\partial x_j}\Big|_{\mathbf{x}=\mathbf{x}_0}$; for $j = 1, \ldots, d$. Applying this approximation to a datum $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ that is close to $\mathbf{x}_0$ leads to the approximation

$$\theta_r(\mathbf{X}_i) \approx \theta_{r0} + \sum_{j=1}^{d} \theta_{rj}(X_{ij} - x_{0j}) \equiv \mathbf{X}_{i;d}^T \boldsymbol{\theta}_r,$$

where we denoted $\mathbf{X}_{i;d} = (1, X_{i1} - x_{01}, \ldots, X_{id} - x_{0d})^T$ and $\boldsymbol{\theta}_r = (\theta_{r0}, \theta_{r1}, \ldots, \theta_{rd})^T$. Since this approximation is only valid for $\mathbf{X}_i$ in a neighbourhood of $\mathbf{x}_0$, this is accounted for by considering appropriate multivariate weights.

Let $K_d : I\!R^d \to I\!R$ be a d-variate non-negative kernel function satisfying $\int K_d(\mathbf{u})d\mathbf{u} = 1$ and $\int \mathbf{u}K_d(\mathbf{u})d\mathbf{u} = \mathbf{0}$. Furthermore, K_d is assumed to have compact support and $\int u_i u_j K_d(\mathbf{u})d\mathbf{u} = \delta_{ij}\mu_2(K_d)$, with $\mu_2(K_d) > 0$. The matrix of second componentwise moments of K_d is thus $\mu_2(K_d)\mathbf{I}_d$, where $\mathbf{I}_d$ is the identity matrix of dimension $d \times d$. A rescaled version is $K_{d,\mathbf{H}}(\boldsymbol{u}) = |\mathbf{H}|^{-1/2}K_d(\mathbf{H}^{-1/2}\boldsymbol{u})$, where $\mathbf{H}$ is a positive definite matrix of bandwidths with determinant $|\mathbf{H}|$.

The multivariate version of equation (12) for $p_1 = p_2 = 1$ is:

$$\mathcal{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \mathbf{H}, \mathbf{x}_0)) = \sum_{i=1}^{n} \ell(\mathbf{X}_{i;d}^T\boldsymbol{\theta}_1, \mathbf{X}_{i;d}^T\boldsymbol{\theta}_2; Y_i)K_{d,\mathbf{H}}(\mathbf{X}_i - \mathbf{x}_0),$$

which needs to be maximized with respect to the model parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, leading to $(\widehat{\boldsymbol{\theta}}_1(\mathbf{x}_0), \widehat{\boldsymbol{\theta}}_2(\mathbf{x}_0))$ satisfying

$$(\widehat{\boldsymbol{\theta}}_1(\mathbf{x}_0)), \widehat{\boldsymbol{\theta}}_2(\mathbf{x}_0))) = \arg\max_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} \sum_{i=1}^{n} \ell(\mathbf{X}_{i;d}^T\boldsymbol{\theta}_1, \mathbf{X}_{i;d}^T\boldsymbol{\theta}_2; Y_i)K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}_0),$$

and estimates of the d-variate function $\theta_r(\cdot)$ and its d first-order partial derivatives

$$\widehat{\theta}_r(\mathbf{x}_0) = \widehat{\theta}_{r0} \quad \text{and} \quad \widehat{\frac{\partial \theta_r(\mathbf{x})}{\partial x_j}}\Big|_{\mathbf{x}=\mathbf{x}_0} = \widehat{\theta}_{rj} \quad j = 1, \ldots, d; \qquad (r \in \{1, 2\}).$$

For practical use of the above method one needs to be aware of the problem of curse of dimensionality. For a given number of data points, local neighbourhoods in higher dimensions are obviously 'emptier'. Consequently, using local modelling techniques in high dimensions requires a large amount of data, to guarantee a sufficient amount of local data to warrant sufficient accuracy in the estimation task. For moderate to small data sets one likely needs to bring in some additional structure, such as an additive model structure on the unknown location and shape functions, i.e. modelling $\theta_r(x_0)$ as a sum of d unknown univariate functions: $\theta_r(x_0) = \theta_{r1}(x_{01}) + \ldots + \theta_{rd}(x_{0d})$, with now $\theta_{r1}(\cdot), \ldots, \theta_{rd}(\cdot)$ d unidimensional functions to be estimated. Investigating statistical inference with such an additive structure will be investigated in future research.

4. Asymptotic results

Before stating the results, recall that the quantile estimator in (11) is obtained from estimation of $\theta_1(x)$ and $\theta_2(x)$. The proposed local polynomial estimation procedure, not only allows to estimate $q_\beta(x)$ but also its derivatives (up to some order), based on these of $\theta_1(x)$ and $\theta_2(x)$. We present the asymptotic theory in its full generality, including derivative estimation. The main results are

- the asymptotic normality results for the estimators of $\theta_1(x)$ and $\theta_2(x)$ (and their derivatives) in Theorem 4.2;
- the asymptotic normality result for the estimator of $q_\beta(x)$ in Theorem 4.3.

A key step in achieving these is provided by Theorem 4.1.

We now come to investigate these asymptotic properties of the local polynomial maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0))$ in (13). The proposed semiparametric method is likelihood-based, relying on the conditional density $f_{Y|X,\alpha}(y; \theta_1(x), \theta_2(x)|x)$ in (10). Consequently assumptions related to the key quantities in this conditional density (the symmetric reference density f and the unknown location and scale functions), as well as assumptions on the density of X (i.e. the design density) are necessary.

4.1. Notations and assumptions

As in all likelihood-based methods, we need notations for the partial derivatives of first and second order of the log-likelihood function involved. We denote, for $r, s, t \in \{1, 2\}$,

$$\begin{aligned}
\psi_r(y; v_1(x), v_2(x)) &= \frac{\partial}{\partial u_r} \ln f_{Y|X,\alpha}(y; u_1, u_2 \mid x)\Big|_{(u_1,u_2)=(v_1(x),v_2(x))} \\
\psi_{rs}(y; v_1(x), v_2(x)) &= \frac{\partial^2}{\partial u_r \partial u_s} \ln f_{Y|X,\alpha}(y; u_1, u_2 \mid x)\Big|_{(u_1,u_2)=(v_1(x),v_2(x))} \\
\psi_{rst}(y; v_1(x), v_2(x)) &= \frac{\partial^3}{\partial u_r \partial u_s \partial u_t} \ln f_{Y|X,\alpha}(y; u_1, u_2 \mid x)\Big|_{(u_1,u_2)=(v_1(x),v_2(x))},
\end{aligned}$$

if these partial derivatives exist. Note that the conditional density $f_{Y|X,\alpha}(y; \theta_1(x), \theta_2(x)|x)$ in (10) is continuous everywhere, but it is not differentiable in $y = \theta_1(x)$, leading to non-differentiability of the log-likelihood at points $Y_i = \theta_1(X_i)$. This non-differentiability of the local log-likelihood function requires to rely on theoretical results for maximum likelihood estimation under *nonstandard conditions*. This goes back to the seminal work of [10]. In [8] asymptotic normality results of maximum likelihood estimators in the unconditional case of density (4) was established, involving a stochastic differentiability condition. The basic insight of such a stochastic differentiability condition is that smoothness with a non-differentiable objective function can be replaced by smoothness of the limit (or approximation) of the objective function. Such limiting objective functions are often expectations that are smoother and 'more' differentiable than their sample counterpart [e.g., 11].

In establishing our asymptotic results, we have been inspired by [1] in their study of local polynomial kernel regression in a random design setting of generalized linear models, and by [12] who deal with multiparameter likelihood models, but under a fixed design setting. Of essence here is to keep in mind three key issues when establishing the asymptotic behaviour of the local maximum likelihood estimators $(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0))$: (i) we are dealing with maximum likelihood estimation under non-standard conditions; (ii) the problem involves multiple parameter functions; (iii) we allow for random design (i.e. the X_i observations are random, with design density f_X).

For establishing the asymptotic behaviour of the local maximum likelihood estimators under non-standard conditions, we need the expected value of the score functions (related to the two parameters). Consider, for $r, s \in \{1, 2\}$,

$$\begin{aligned} \lambda_r(v_1(x), v_2(x)) &= \mathbb{E}_{Y|X}[\psi_r(Y; v_1(X), v_2(X)) \mid X = x] \\ \lambda_{rs}(v_1(x), v_2(x)) &= \frac{\partial}{\partial u_s}\lambda_r(u_1, u_2)\Big|_{(u_1,u_2)=(v_1(x),v_2(x))}. \end{aligned} \quad (15)$$

Under appropriate assumptions, [8] showed (see their Proposition 3.1) that the (unconditional) expected value of the (unconditional) score function under density (4) is zero. Keeping in mind (8) it then follows immediately that, for $r \in \{1, 2\}$,

$$\lambda_r(\theta_1(x), \theta_2(x)) = \mathbb{E}_{Y|X}\left[\psi_r(Y; \theta_1(X), \theta_2(X)) \mid X = x\right] = 0 \quad \text{for all } x. \quad (16)$$

From (9), and under appropriate assumptions (see below), we get to the Fisher information matrix

$$\mathcal{I}(\theta_1(x), \theta_2(x)) = \begin{bmatrix} \frac{2\alpha(1-\alpha)\gamma_1}{[\exp(\theta_2(x)]^2} & 0 \\ 0 & (2\gamma_3 - 1) \end{bmatrix}, \quad (17)$$

where, $\mathcal{I}_{rs}(\theta_1(x), \theta_2(x)) = \mathbb{E}_{Y|X}[\psi_r(Y; \theta_1(X), \theta_2(X))\psi_s(Y; \theta_1(X), \theta_2(X)) \mid X = x] = \mathbb{E}_{Y|X}[-\psi_{rs}(Y; \theta_1(X), \theta_2(X)) \mid X = x]$ for all x, and $r, s \in \{1, 2\}$,.

Remark 4.1. Note that from (17) it is clear that the function $\mathcal{I}_{rs}(u_1, u_2)$, is differentiable with respect to both arguments u_1 and u_2 (for finite scale parameter $\theta_2(x)$), and moreover that the partial derivative with respect to u_1 equals zero.

We now state the assumptions that are needed for the above outlined derivation and establishing the asymptotic behaviour of the local maximum likelihood estimators.

Assumptions.

(A1) The densities $f_{Y|X,\alpha}(y; \theta_1(x), \theta_2(x)|x)$ have a common support for all x. There exists an open subset $\mathring{\boldsymbol{\Theta}}$ of the parameter space $\boldsymbol{\Theta}$ containing the true parameters $(\theta_1^0(x), \theta_2^0(x))$ for all x.

(A2) The reference density $f(\cdot)$ satisfies $\int_0^\infty \big|\ln f(s)\big| f(s) ds < \infty$.

(A3) The reference symmetric density $f(s)$ is differentiable almost everywhere and satisfies $\gamma_\ell = \int_0^\infty s^{\ell-1} \cdot \frac{(f'(s))^2}{f(s)} ds < \infty \quad$ for $\ell = 1, 2, 3$.

(A4) The reference symmetric density $f(s)$ satisfies $\lim\limits_{s\to\infty} s f(s) = 0$ or $\int_0^\infty s f'(s) ds = -\frac{1}{2}$.

(A5) The function $\theta_r(\cdot)$ has a (p_r+1)th (respectively (p_r+2)nd) continuous derivative for p_r odd (respectively p_r even).

(A6) $\psi_{rs}(y; \theta_1, \theta_2) < 0$ for $\theta_r \in I\!R$ and y in the range of the response variable (with $r, s \in \{1, 2\}$).

(A7) The functions $f'_X(x_0)$, $\theta_r^{(p_r+2)}(x_0)$, $\psi_r(y; \theta_1(x_0), \theta_2(x_0))$, $\psi_{rs}(y; \theta_1(x_0), \theta_2(x_0))$, $\psi_{rst}(y; \theta_1(x_0), \theta_2(x_0))$ and $\frac{d}{dx_0}\mathcal{I}_{rr}(\theta_1(x_0), \theta_2(x_0))$ are continuous in x_0 (with $r, s, t \in \{1, 2\}$).

(A8) The kernel function $K(\cdot)$ is a symmetric probability density on $[-1, 1]$.

(A9) For each point x_∂ on the boundary of supp(f_X) (the support of f_X), there exist an interval $\mathcal{C}$ containing x_∂ having nonnull interior such that $\inf_{x\in\mathcal{C}} f_X(x) > 0$.

(A10) The bandwidth sequence $h = h_n$ satisfies: $h_n \to 0$ and $nh_n^3 \to \infty$.

Assumptions **(A1)—(A4)** concern conditions needed on the reference symmetric density f, whereas Assumptions **(A5)—(A10)** are needed for the local likelihood approach. Assumptions **(A2)—(A4)** are, for example, satisfied for f a standard normal, Student's -t, logistic and Laplace densities. See [8], including its Supplement, for details on this.

Remark 4.2. It is well-known that the asymptotic properties of local polynomial estimators for mean regression differ for points in the interior or at the boundary of the support of the domain of X (denoted by $\text{supp}(f_X)$). See [2]. With K supported on $[-1, 1]$, the support of $K_h(\cdot - x_0)$ is $\mathcal{E}_{x_0,h} = \{u : |u - x_0| \leqslant h\}$. We call x_0 an interior point of $\text{supp}(f_X(\cdot))$ if $\mathcal{E}_{x_0,h} \subseteq \text{supp}(f_X)$. Otherwise x_0 is called a boundary point. If $\text{supp}(f_X) = [b_1, b_2]$ then x_0 is a boundary point if and only if $x_0 = b_1 + \tau h$ or $x_0 = b_2 - \tau h$ for some $0 \leqslant \tau < 1$. Denoting $\mathcal{D}_{x_0;h} = \{u : x_0 - hu \in \text{supp}(f_X)\} \cap [-1, 1]$, we have that x_0 is an interior point if and only if $\mathcal{D}_{x_0;h} = [-1, 1]$. Both cases, these of interior points and of boundary points, are covered in the sequel by considering the measurable sets $\mathcal{A} = \mathcal{D}_{x_0;h} \subset \mathbb{R}$. Notably, if x_0 is an boundary point and hence of the form $x_0 = x_\partial + ch$ where x_∂ is a point on the boundary of $\text{supp}(f_X)$ and $c \in [-1, 1]$, then $\mathcal{A}$ is replaced by $\mathcal{D}_{x_0,h} \subset [-1, 1]$. For x_0 an interior point in $\text{supp}(f_X)$, $\mathcal{A} = [-1, 1]$.

For writing down the theoretical results we need some further notations. To ease the reading, we use notations that stay close to these of [12] and [1]. The jth moment of the kernel function, restricted to the domain $\mathcal{A}$, is denoted by $\nu_j(\mathcal{A}) = \int_{\mathcal{A}} u^j K(u)du$. Let $\boldsymbol{N}_{p_r p_s}(x_0; \mathcal{A})$, $\boldsymbol{T}_{p_r p_s}(x_0; \mathcal{A})$ and $\boldsymbol{Q}_{p_r p_s}(x_0; \mathcal{A})$ $(r, s \in \{1, 2\})$ be matrices of dimension $(p_r + 1) \times (p_s + 1)$ of which the $(k+1, l+1)$th entry equals $\nu_{k+l}(\mathcal{A})$, $\int_{\mathcal{A}} u^{k+l} K^2(u)du$ and $\nu_{k+l+1}(\mathcal{A})$ $(k = 0, ..., p_r; l = 0, ..., p_s)$. Let $\boldsymbol{M}_{v,p_s}(u; \mathcal{A})$ be the same as $\boldsymbol{N}_{p_r p_s}(x_0; \mathcal{A})$, but with the $(v + 1)$st column replaced by $(1, u, ..., u^{p_s})^T$. For $|\boldsymbol{N}_{p_r p_s}(x_0; \mathcal{A})| \neq 0$ define

$$K_{v,p_s}(u; \mathcal{A}) = v!\{|\boldsymbol{M}_{v,p_s}(u; \mathcal{A})|/|\boldsymbol{N}_{p_s p_s}(x_0; \mathcal{A})|\}K(u).$$

For notational simplification, we suppress in what follows the $\mathcal{A}$ notation, and simply write ν_j, $\boldsymbol{N}_{p_r p_s}(x_0)$, $\boldsymbol{M}_{v,p_s}(u)$ and $K_{v,p_s}(u)$. We also suppress the region of integration $\mathcal{A}$, and write short-handed $\int K_{v,p_s}^2(u)du$ instead of $\int_{\mathcal{A}} K_{v,p_s}^2(u; \mathcal{A})du$.

The kernel $K_{v,p_s}(u)$ is an equivalent kernel as defined by [13]. For several kernels K satisfying Assumption **(A9)**, the expressions for $K_{v,p_s}(u)$ for different value of v and p_s are tabulated in [2] (pp. 66). The concept of equivalent kernel is useful for giving concise expressions for the asymptotic distribution of $\widehat{\theta}_{rv}(x_0) = \widehat{\theta}_{rv}(x_0; p_r, h)$; $(r \in \{1,2\})$ for x_0. As mentioned in [1] $(-1)^v K_{v,p_s}(u)$ is an order (v,t) kernel, where $t = p_s + 1$ if $p_s - v$ is odd, and $t = p_s + 2$ if $p_s - v$ is even. Further, for $r \in \{1,2\}$, denote

$$\begin{aligned} \boldsymbol{H}_{p_r} &= \operatorname{diag}(1, h, \ldots, h^{p_r}) \\ \boldsymbol{\Sigma}_{x_0} &= f_X(x_0)\mathcal{I}(\theta_1(x_0), \theta_2(x_0)) \otimes \boldsymbol{N}(x_0) \quad (18) \\ \boldsymbol{\Gamma}_{x_0} &= f_X(x_0)\mathcal{I}(\theta_1(x_0), \theta_2(x_0)) \otimes \boldsymbol{T}(x_0) \quad (19) \\ \boldsymbol{\Lambda}_{x_0} &= \boldsymbol{D}(x_0) \otimes \boldsymbol{Q}(x_0) \quad (20) \end{aligned}$$

$$\boldsymbol{N}(x_0) = \begin{pmatrix} \boldsymbol{N}_{p_1p_1}(x_0) & \boldsymbol{N}_{p_1p_2}(x_0) \\ \boldsymbol{N}_{p_2p_1}(x_0) & \boldsymbol{N}_{p_2p_2}(x_0) \end{pmatrix} \quad \boldsymbol{T}(x_0) = \begin{pmatrix} \boldsymbol{T}_{p_1p_1}(x_0) & \boldsymbol{T}_{p_1p_2}(x_0) \\ \boldsymbol{T}_{p_2p_1}(x_0) & \boldsymbol{T}_{p_2p_2}(x_0) \end{pmatrix}$$

$$\boldsymbol{Q}(x_0) = \begin{pmatrix} \boldsymbol{Q}_{p_1p_1}(x_0) & \boldsymbol{Q}_{p_1p_2}(x_0) \\ \boldsymbol{Q}_{p_2p_1}(x_0) & \boldsymbol{Q}_{p_2p_2}(x_0) \end{pmatrix}$$

$$\boldsymbol{D}(x_0) = \begin{pmatrix} \frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{11}(\theta_1(x_0), \theta_2(x_0))\} & 0 \\ 0 & \frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{22}(\theta_1(x_0), \theta_2(x_0))\} \end{pmatrix}$$

where $\otimes$ denotes a generalized Kronecker product. For a $(r \times s)$ matrix $\boldsymbol{C} = (c_{ij})$ and a partitioned matrix $\boldsymbol{D}$ with submatrices $\boldsymbol{D}_{ij}$ $(i = 1, ..., r; j = 1, ..., s)$, the generalized Kronecker product $\boldsymbol{C} \otimes \boldsymbol{D}$ is defined as a partitioned matrix with submatrices $(c_{ij}\boldsymbol{D}_{ij}), i = 1, ..., r; j = 1, ..., s$. If all submatrices $\boldsymbol{D}_{ij}$ are identical (say to $\boldsymbol{D}$), then the generalized Kronecker product simplifies to the ordinary Kronecker product of the matrices $\boldsymbol{C}$ and $\boldsymbol{D}$. This happens when in (18), (19) and (20) we take $p_1 = p_2$.

4.2. Asymptotic results for local polynomial maximum likelihood conditional quantile estimators

A key step in deriving the asymptotic normality result for the local polynomial maximum likelihood estimator $\widehat{\boldsymbol{\theta}}(x_0) = (\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0))^T$ is studying first the asymptotic behaviour of the quantity $\boldsymbol{W}^n(x_0) = (\boldsymbol{W}_1^n(x_0)^T, \boldsymbol{W}_2^n(x_0)^T)^T$, where $\boldsymbol{W}_r^n(x_0)$, for $r \in \{1,2\}$, is a column vector of dimension $p_r + 1$, with $(k+1)$st component

$$W_{rk}^n(x_0) = \frac{1}{\sqrt{nh^{2k+1}}} \sum_{i=1}^{n} \psi_r(Y_i; \overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0)) K\{(X_i - x_0)/h\}(X_i - x_0)^k \quad (21)$$

$$\text{with } \overline{\theta}_r(X_i, x_0) = \sum_{j=0}^{p_r} \theta_{rj}(x_0)(X_i - x_0)^j; \qquad r \in \{1, 2\}. \tag{22}$$

Let $\mathbf{I}_{p_1+p_2+2}$ denote an identity matrix of order $(p_1 + p_2 + 2) \times (p_1 + p_2 + 2)$ and let $\mathbf{0}_{p_1+p_2+2}$ be a $(p_1 + p_2 + 2)$-dimensional column null vector, but yet in terms of $\boldsymbol{W}^n(x_0)$. Theorem 4.1 states the asymptotic normality result for the semiparametric local polynomial maximum likelihood estimators.

Theorem 4.1. Assume **(A1)—(A10)** hold. Then, for $n \to \infty$,

$$\begin{aligned}(\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Gamma}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1})^{-1/2}\Big\{\sqrt{nh}\,\Big(\boldsymbol{H}_{p_1}\big(\widehat{\boldsymbol{\theta}}_1(x_0) - \boldsymbol{\theta}_1(x_0)\big), \boldsymbol{H}_{p_2}\big(\widehat{\boldsymbol{\theta}}_2(x_0) - \boldsymbol{\theta}_2(x_0)\big)\Big)^T \\ - (\boldsymbol{\Sigma}_{x_0}^{-1} - h\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1})\mathbb{E}[\boldsymbol{W}^n(x_0)]\Big\} \\ \xrightarrow{D} \mathcal{N}_{p_1+p_2+2}(\mathbf{0}_{p_1+p_2+2}, \mathbf{I}_{p_1+p_2+2}).\end{aligned}$$

where $\boldsymbol{W}^n(x_0)$ is defined in (21) and x_0 an interior point of $\mathrm{supp}(f_X)$.

From Theorem 4.1 is becomes transparent that one needs to study first the quantities $\mathbb{E}[\boldsymbol{W}^n(x_0)]$, $\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ and $\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$. Asymptotic expressions for these quantities are established in Lemmas S2.1 and S2.2 in Section S2.1 of the Supplementary Material.

From Theorem 4.1, one can derive the asymptotic normality of the maximum likelihood estimator of the vth derivative $\theta_r^{(v)}(x_0)$ of $\theta_r(x_0)$. This result is stated in Theorem 4.2.

Theorem 4.2. Assume **(A1)—(A10)** hold. For x_0 given, we have

(i). if $p_r - v$ is odd ($v = 0, ..., p_r$ and $r \in \{1, 2\}$), then for $n \to \infty$,

$$\begin{aligned}&\sqrt{nh^{2v+1}}\left(\frac{\mathcal{I}_{rr}^{-1}(\theta_1(x_0), \theta_2(x_0))}{f_X(x_0)}\int K_{v,p_r}^2(u)du\right)^{-1/2} \\ &\times\left[\widehat{\theta}_r^{(v)}(x_0) - \theta_r^{(v)}(x_0) - h^{p_r-v+1}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\left\{\int u^{p_r+1}K_{v,p_r}(u)du\right\}(1 + O(h))\right] \\ &\xrightarrow{D} \mathcal{N}(0, 1);\end{aligned}$$

(ii). if $p_r - v$ even ($v = 0, 1, ..., p_r$ and $r \in \{1,2\}$), then for $n \to \infty$,

$$\sqrt{nh^{2v+1}} \left(\frac{\mathcal{I}_{rr}^{-1}(\theta_1(x_0), \theta_2(x_0))}{f_X(x_0)} \int K_{v,p_r}^2(u) du \right)^{-1/2} \Big(\widehat{\theta}_r^{(v)}(x_0; p_r, h) - \theta_r^{(v)}(x_0)$$
$$- \left[\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!} \int u^{p_r+2} K_{v,p_r}(u) du + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} \frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0), \theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0), \theta_2(x_0))} \right.$$
$$\left. \times \left\{ \int u^{p_r+2} K_{v,p_r}(u) du - v \int u^{p_r+1} K_{v-1,p_r}(u) du \right\} \right] h^{p_r - v + 2}(1 + O(h)) \Big)$$
$$\xrightarrow{D} \mathcal{N}(0,1).$$

The estimated βth quantile in x_0 can be obtained from (11). When $\beta = \alpha$, $C_\alpha(\alpha) = 0$ and $\widehat{q}_\alpha(x_0) = \widehat{\theta}_1(x_0)$. Therefore the asymptotic distribution of $\widehat{q}_\alpha(x_0)$, is that of $\widehat{\theta}_1(x_0)$, given in Theorem 4.2. In general, for any $\beta \in (0,1)$, the asymptotic distribution of $\widehat{q}_\beta(x_0)$ is given in Theorem 4.3.

Theorem 4.3. Assume **(A1)—(A10)** hold. For x_0 given, we have

$$\sqrt{nh} \left\{ \widehat{q}_\beta(x_0) - q_\beta(x_0) - \left[\text{ABias}[\widehat{\theta}_1(x_0)] + C_\alpha(\beta) \cdot e^{\theta_2(x_0)} \text{ABias}[\widehat{\theta}_2(x_0)] \right] \right\}$$
$$\xrightarrow{D} \mathcal{N}\left(0, \sigma_q^2(x_0)\right),$$

for $n \to \infty$, where for $r \in \{1,2\}$

(i). if p_r (for $r \in \{1,2\}$) is odd, then

$$\text{ABias}[\widehat{\theta}_r(x_0)] = h^{p_r+1} \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} \left\{ \int u^{p_r+1} K_{0,p_r}(u) du \right\} (1 + O(h)),$$

(ii). if p_r (for $r \in \{1,2\}$) is even, then

$$\text{ABias}[\widehat{\theta}_r(x_0)]$$
$$= \left[\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!} \int u^{p_r+2} K_{0,p_r}(u) du + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} \frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0), \theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0), \theta_2(x_0))} \right.$$
$$\left. \times \left\{ \int u^{p_r+2} K_{0,p_r}(u) du \right\} \right] h^{p_r+2}(1 + O(h)),$$

$$\sigma_q^2(x_0) = \frac{\mathcal{I}_{11}^{-1}(\theta_1(x_0), \theta_2(x_0))}{f_X(x_0)} \int K_{0,p_1}^2(u)du$$
$$+ C_\alpha^2(\beta)e^{2\theta_2(x_0)} \frac{\mathcal{I}_{22}^{-1}(\theta_1(x_0), \theta_2(x_0))}{f_X(x_0)} \int K_{0,p_2}^2(u)du.$$

5. Bandwidth selection

The semiparametric procedure in Section 3.2 involves the choice of a bandwidth parameter. Similarly, the nonparametric approach, briefly reviewed in Section 2, needs a bandwidth choice (see Section S1.2). Thanks to the asymptotic theory established in Section 4 for both estimation approaches, we are able to study this bandwidth selection issue. In this section we propose practical bandwidth selectors. The finite-sample performance of the proposed semiparametric procedure, including that of the data-driven bandwidth selectors of this section, is investigated in a simulation study in Section S3.1 of the Supplementary Material.

A standard way to obtain theoretical optimal local (respectively global) bandwidths is by looking at an asymptotic expression for the Mean Squared Error (respectively the Mean Integrated Squared Error) of an estimator $\widehat{\theta}(\cdot)$ for a target quantity $\theta(\cdot)$, defined as respectively $\text{MSE}\big(\widehat{\theta}(x_0)\big) = \mathbb{E}\big(\widehat{\theta}(x_0) - \theta(x_0)\big)^2$ and $\mathbb{E}\big\{\int \big(\widehat{\theta}(x) - \theta(x)\big)^2 w(x)dx\big\}$, with $w(\cdot)$ a given weight function. Since the Mean Squared Error of an estimator can be decomposed into the squared bias and the variance, an asymptotic expression for the former is obtained by squaring an asymptotic expression of the bias and adding an asymptotic expression for the variance. An asymptotic expression for the Mean Integrated Squared Error is obtained by considering weighted integrated versions of the latter asymptotic expression.

Since there are two unknown functions $\theta_1(\cdot)$ and $\theta_2(\cdot)$ the local (respectively global) performance of the estimation method is measured via $\text{MSE}\big(\widehat{\theta}_1(x_0)\big) + \text{MSE}\big(\widehat{\theta}_2(x_0)\big)$ (respectively the weighted integral version of this).

5.1. Theoretical optimal bandwidths

Using the asymptotic expressions for bias and variance provided in Theorem 4.2, restricting to $p_1 = p_2 = p$, we get to the following asymptotic expression of the MSE, denoted by AMSE, for estimation of the vth derivative of the two functions of interest, evaluated in an interior point x_0,

$$\begin{aligned}
&\text{AMSE}\{\widehat{\theta}_1^{(v)}(x_0), \widehat{\theta}_2^{(v)}(x_0)\} \\
&= \frac{1}{nh^{2v+1} f_X(x_0)} \left[\mathcal{I}_{11}^{-1}(\theta_1(x_0), \theta_2(x_0)) + \mathcal{I}_{22}^{-1}(\theta_1(x_0), \theta_2(x_0))\right] \int K_{v,p}^2(u)du \\
&\quad + h^{2(p-v+1)} \left\{\int u^{p+1} K_{v,p}(u)du\right\}^2 \sum_{r=1}^{2} \left\{\frac{\theta_r^{(p+1)}(x_0)}{(p+1)!}\right\}^2
\end{aligned}$$

Working with weighted integrated squared bias and variance and minimization of the expression with respect to h leads to an asymptotic optimal global bandwidth given by

$$h_{\alpha,\text{opt}}^{\text{AMISE}} = C_{v,p}(K) \left[\frac{\int \sum_{r=1}^{2} \mathcal{I}_{rr}^{-1}(\theta_1(x), \theta_2(x)) w(x)/f_X(x)dx}{\int \sum_{r=1}^{2} \left\{\theta_r^{(p+1)}(x)\right\}^2 w(x)dx}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}, \quad (23)$$

where the elements for the Fisher information are in (17), and where

$$C_{v,p}(K) = \left[\frac{\{(p+1)!\}^2 (2v+1) \int K_{v,p}^2(u)du}{2(p+1-v)\left\{\int u^{p+1} K_{v,p}(u)du\right\}^2}\right]^{\frac{1}{2p+3}}.$$

The constants $C_{v,p}(K)$ are easily calculated, and for some kernels K are listed in Table 3.2 in [2] (p. 67).

When the $\theta_2(\cdot)$ is known, and hence the criterion for choosing an optimal bandwidth is reduced to $\text{MSE}\big(\widehat{\theta}_1(x_0)\big)$, the asymptotic local and global bandwidths are

$$h_{\alpha,\text{opt}}^{\text{AMSE}}(x_0) = C_{v,p}(K) \left[\frac{\exp(2\theta_2(x_0))}{2\alpha(1-\alpha)\gamma_1 f_X(x_0)\left\{\theta_1^{(p+1)}(x_0)\right\}^2}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}, \quad (24)$$

and

$$h_{\alpha,\text{opt}}^{\text{AMISE}} = C_{v,p}(K) \left[\frac{\int \exp(2\theta_2(x)) w(x)/f_X(x)dx}{2\alpha(1-\alpha)\gamma_1 \int \left\{\theta_1^{(p+1)}(x)\right\}^2 w(x)dx}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}. \quad (25)$$

The asymptotically optimal bandwidths depends on unknown quantities such as the design density $f_X(\cdot)$, the function $\theta_2(\cdot)$ and the derivative function $\theta_r^{(p+1)}(\cdot); (r \in \{1,2\})$. We therefore next discuss practical bandwidth selection procedures.

5.2. Data-driven bandwidth selection

5.2.1. Rule of thumb bandwidth selector

The optimal bandwidth (23) depends on the unknown functions $f_X(\cdot)$, $\theta_2(\cdot)$ and $\theta_r^{(p+1)}(\cdot); (r \in \{1,2\})$. In mean regression, [14] proposed a rule of thumb for bandwidth selection. We adopt a similar procedure here. We start by estimating parametrically $\theta_r(x)$ by fitting globally a polynomial of order $p+3$. The resulting parametric fit is denoted as $\breve{\theta}_r(x) = \breve{\theta}_{r0} + \breve{\theta}_{r1}x + \cdots + \breve{\theta}_{r(p+3)}x^{p+3}$. Taking $w(x) = f_X(x)w_0(x)$ for some specific function w_0 in (23), and substituting the parametric pilot estimates $\breve{\theta}_1$ and $\breve{\theta}_2$ into (23), we obtain the following expression

$$C_{v,p}(K)\left[\left\{\frac{1}{\int\sum_{r=1}^{2}\left\{\breve{\theta}_r^{(p+1)}(x)\right\}^2 w_0(x)f_X(x)dx}\right\}\right.$$
$$\left.\times\left\{\frac{\int\exp\left(2\breve{\theta}_2(x)w_0(x)dx\right)}{2\alpha(1-\alpha)\gamma_1}+\frac{\int w_0(x)dx}{2\gamma_3-1}\right\}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}.$$

The quantity $\int\left\{\breve{\theta}_r^{(p+1)}(x)\right\}^2 w_0(x)f_X(x)dx$ can be estimated by $n^{-1}\sum_{i=1}^{n}\left\{\breve{\theta}_r^{(p+1)}(X_i)\right\}^2 w_0(X_i)$. The term $\int\exp\left(2\breve{\theta}_2(x)\right)w_0(x)dx$ is approximated by $\breve{\theta}_2$ the average of $\breve{\theta}_2(x_j)$ over the number of grid points, multiplied with $\int w_0(x)dx$. All together this leads to the following rule of thumb (ROT) bandwidth selector

$$\widehat{h}_\alpha^{\mathrm{ROT}} = C_{v,p}(K)\left[\left\{\frac{\exp\left(2\breve{\theta}_2\right)}{2\alpha(1-\alpha)\gamma_1}+\frac{1}{2\gamma_3-1}\right\}\left\{\frac{\int w_0(x)dx}{\sum_{r=1}^{2}\frac{1}{n}\sum_{i=1}^{n}\left\{\breve{\theta}_r^{(p+1)}(X_i)\right\}^2 w_0(X_i))}\right\}\right]^{\frac{1}{2p+3}},$$

with $\breve{\theta}_r^{(p+1)}(X_i) = (p+1)!\breve{\theta}_{r(p+1)} + (p+2)!\breve{\theta}_{r(p+2)}X_i + (p+3)!\breve{\theta}_{r(p+3)}X_i^2/2$.

5.2.2. Cross-validation bandwidth selector

An alternative approach is to rely on cross-validation. A cross-validated bandwidth is obtained via

$$\widehat{h}_{\alpha}^{\mathrm{CV}} = \arg\max_{h>0} \sum_{i=1}^{n} \ln f_{Y|X,\alpha}\left(Y_i; \widehat{\theta}_1^{[-i]}(X_i), \widehat{\theta}_2^{[-i]}(X_i)|X_i\right)$$

where $\widehat{\theta}_1^{[-i]}(X_i)$ and $\widehat{\theta}_2^{[-i]}(X_i)$ are the estimators for, respectively $\theta_1(X_i)$ and $\theta_2(X_i)$ based on the sample without the ith observation (X_i, Y_i).

5.2.3. Quantile-Mean bandwidth selector

Here we restrict to the case that $\theta_2(x)$ is known for all x. Another way of obtaining a bandwidth selector is by linking the approximated asymptotically optimal bandwidth for quantile curve and mean curve estimation (i.e. estimation of $\mathbb{E}(Y \mid X = x)$).

For local polynomial estimation of the conditional mean $\mathbb{E}(Y \mid X)$ the local and global theoretical optimal bandwidths are given by (see [2] [pp. 67-68])

$$h_{\mathrm{mean}}(x_0) = C_{v,p}(K) \left[\frac{\sigma_*^2(x_0)}{\{m^{(p+1)}(x_0)\}^2 f_X(x_0)} \right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}} \qquad (26)$$

$$h_{\mathrm{mean}} = C_{v,p}(K) \left[\frac{\int \sigma_*^2(x) w(x)/f_X(x) dx}{\int \{m^{(p+1)}(x)\}^2 w(x) dx} \right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}},$$

where $\sigma_*^2(x_0)$ is the conditional variance of Y given $X = x_0$. From (24) and (26), we obtain that

$$\frac{h_{\alpha,\mathrm{opt}}^{\mathrm{AMSE}}(x_0)}{h_{\mathrm{mean}}(x_0)} = \left[\frac{1}{2\alpha(1-\alpha)\gamma_1} \right]^{\frac{1}{2p+3}} \times \left[\frac{m^{(p+1)}(x_0)}{\theta_1^{(p+1)}(x_0)} \right]^{\frac{2}{2p+3}} \left[\frac{\exp\left(2(x_0)\right)}{\sigma_*^2(x_0)} \right]^{\frac{1}{2p+3}}. \qquad (27)$$

For the conditional density as in (10) we know that

$$\sigma_*^2(x_0) = \mathbb{V}\mathrm{ar}(Y|X = x_0) = \frac{\exp\left(2\theta_2(x_0)\right)}{\alpha^2(1-\alpha)^2} [(1-2\alpha)^2(\mu_2 - \mu_1^2) + \alpha(1-\alpha)\mu_2],$$

where $\mu_r = 2\int_0^\infty s^r f(s)ds$ $(r \in \{1, 2\})$. Therefore, we set,

$$\frac{\exp(2\theta_2(x_0))}{\sigma_*^2(x_0)} = \frac{\alpha^2(1-\alpha)^2}{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2}. \tag{28}$$

We then make the rough approximation that the $(p+1)$th derivative of the mean and of the quantile curves are equal, i.e. $m^{(p+1)}(x_0) = \theta_1^{(p+1)}(x_0)$. From (27) and (28) we find

$$\begin{aligned}\frac{h_{\alpha,\text{opt}}^{\text{AMSE}}(x_0)}{h_{\text{mean}}(x_0)} &= \left[\frac{1}{2\alpha(1-\alpha)\gamma_1}\right]^{\frac{1}{2p+3}}\left[\frac{\alpha^2(1-\alpha)^2}{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2}\right]^{\frac{1}{2p+3}}\\ &= \left[\frac{\alpha(1-\alpha)}{2\gamma_1\{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2\}}\right]^{\frac{1}{2p+3}}.\end{aligned}$$

This then leads to the optimal local and global bandwidths

$$h_{\alpha,\text{opt}}^{\text{QM}}(x_0) = h_{\text{mean}}(x_0)\left[\frac{\alpha(1-\alpha)}{2\gamma_1\{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2\}}\right]^{\frac{1}{2p+3}} \tag{29}$$

$$h_{\alpha,\text{opt}}^{\text{QM}} = h_{\text{mean}}\left[\frac{\alpha(1-\alpha)}{2\gamma_1\{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2\}}\right]^{\frac{1}{2p+3}}, \tag{30}$$

in which we replace $h_{\text{mean}}(x_0)$ or h_{mean} by a good data-driven (local or global) bandwidth selector for mean regression estimation. We refer to the resulting selector as a Quantile-Mean (QM) bandwidth selector.

In (29) and (30) we can use any available good performing data-driven bandwidth selectors for mean regression. For example, we can use the Plug-in bandwidth selector or the Cross Validation bandwidth selector that are available in the R package `locpol`.

6. Estimation of the index parameter/function

In the previous sections the index-parameter α was considered to be known and fixed. However, the proposed methodology in Section 3.2 remains valid when (i) α is constant but unknown; and (ii) α changes with the covariate value, i.e. $\alpha(x)$. In these cases we need to estimate α or $\alpha(x)$. Obviously, when we know that α is constant, we should exploit this in the estimation procedure. We discuss these estimation tasks in respectively Sections 6.1 and 6.2.

6.1. Estimation of α constant but unknown

A first estimator for the index parameter α is

$$\widehat{\alpha}^{(1)} = \frac{1}{n_{\text{grid}}} \sum_{j=1}^{n_{\text{grid}}} \widehat{\alpha}(x_j), \tag{31}$$

where, for a fixed point x_j in the grid $\{x_1, \ldots, x_{n_{\text{grid}}}\}$, the estimator $\widehat{\alpha}(x_j)$ is obtained by maximizing the local kernel-weighted conditional log-likelihood

$$\mathcal{L}_n(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2; h, x_j) = \sum_{i=1}^{n} \ell(\alpha, (\mathbf{X}_{i,p_1}^{(j)})^T \boldsymbol{\theta}_1, (\mathbf{X}_{i,p_2}^{(j)})^T \boldsymbol{\theta}_2; Y_i) K_h(X_i - x_j), \tag{32}$$

with respect to $(\alpha, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Herein $\mathbf{X}_{i,p}^{(j)} = (1, X_i - x_j, \ldots, (X_i - x_j)^p)^T$. We refer to this estimation method as Method 1, and the estimator in (31) as $\widehat{\alpha}^{(1)}$.

A second natural estimator for α is as follows. Based on data $(Y_1, X_1), \ldots, (Y_n, X_n)$ obtain a nonparametric estimator for the mean regression function $m(x) = \mathbb{E}(Y \mid X = x)$, for example by using a local linear fit. Using the nonparametric estimator $\widehat{m}$ for the mean function, we form the residuals $Y_i - \widehat{m}(X_i)$. Since the effect of the covariate is reduced by subtracting the estimated mean function, we can pretend to be in an unconditional (non-regression) setting and estimate the index-parameter via maximum likelihood estimation (as in [8]) We denote the resulting estimator by $\widehat{\alpha}^{(2)}$, and refer to this estimation method as Method 2.

An alternative to the second method is to consider a nonparametric estimator of the median of Y given $X = x$, denoted by $q_{0.5}(x)$. Denoting this nonparametric estimator by $\widehat{q}_{0.5}^{\text{NP}}(x)$, we obtain the residuals $Y_i - \widehat{q}_{0.5}^{\text{NP}}(X_i)$ and based on these pseudo observations, we estimate α using maximum likelihood estimation (in an unconditional setting). We denote this estimator by $\widehat{\alpha}^{(3)}$, and refer to the estimation method as Method 3.

In Section 7.3 we investigate the finite-sample performance of these three estimation methods. It is important to first make some remarks about the methods. A first important observation is that Methods 2 and 3 in fact do not exploit the parametric setting of model (10), whereas this is done with Method 1. We therefore expect Method 1 to be more efficient. Methods 2 and 3 involve a bandwidth parameter for the nonparametric mean and median estimation part. For these bandwidths one can use data-driven bandwidths that are available in the literature and in software. For Method 1 we would need a bandwidth in the local kernel-weighted conditional log-likelihood. Here no data-driven bandwidth selection method is available yet. Also the maximization problem in (32) involves now optimization over $(p_1 + p_2 + 3)$ parameters, which is one parameter more than in case α is known. Hence we expect a higher computational cost for Method 1.

6.2. Estimation of the unknown function $\alpha(x)$

When nothing is know about how the level of asymmetry of a distribution is affected by the covariate, it is advised to start by estimating $\alpha(x)$ nonparametrically. Since $\alpha(x) \in (0,1)$ a one-to-one transformation is needed before a local approximation of the function can be considered. A natural transformation is $\theta_3(x) = \ln(\alpha(x)/(1-\alpha(x)))$, which takes values in $\mathbb{R}$. Extending the framework of Section 3.2 is then achieved by using a Taylor approximation for $\theta_3(x_0)$, and adapting the local log-likelihood function in (12) to the local kernel-weighted conditional log-likelihood

$$\mathcal{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3; h, x_0) = \sum_{i=1}^{n} \ell(\mathbf{X}_{i,p_1}^T \boldsymbol{\theta}_1, \mathbf{X}_{i,p_2}^T \boldsymbol{\theta}_2, \mathbf{X}_{i,p_3}^T \boldsymbol{\theta}_3; Y_i) K_h(X_i - x_0), \quad (33)$$

where $\mathbf{X}_{i,p_r} = (1, (X_i - x_0), \cdots, (X_i - x_0)^{p_r})^T$, $\boldsymbol{\theta}_r = (\theta_{r0}, \cdots, \theta_{rp_r})^T$ with $\theta_{rv} = \frac{\theta_r^{(v)}(x_0)}{v!}; v = 0, 1, \ldots, p_r; (r = 1, 2, 3)$.

Maximizing (33) with respect to $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_2)$ gives the vector of estimators $(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0), \widehat{\boldsymbol{\theta}}_3(x_0)) = \Big((\widehat{\theta}_{10}(x_0), \ldots, \widehat{\theta}_{1p_1}(x_0))^T, (\widehat{\theta}_{20}(x_0), ..., \widehat{\theta}_{2p_2}(x_0))^T,$ $(\widehat{\theta}_{30}(x_0), \ldots, \widehat{\theta}_{3p_3}(x_0))^T\Big)$. For a given point x_0, we have

$$\widehat{\alpha}(x_0) = \frac{\exp[\widehat{\boldsymbol{\theta}}_3(x_0)]}{1 + \exp[\widehat{\boldsymbol{\theta}}_3(x_0)]},$$

and the estimated βth quantile curve at a point x_0 is

$$\widehat{q}_\beta(x_0) = \widehat{\theta}_1(x) + \exp\left\{\widehat{\theta}_2(x)\right\} \cdot C_{\widehat{\alpha}(x_0)}(\beta), \quad (34)$$

as opposed to (11).

As can be seen from Section 7.3 the methodology performs very well also in this setting of an unknown function $\alpha(x)$.

7. Simulation study

In this simulation study we investigate several aspects of the finite-sample performance of the proposed semiparametric methodology, including

(i) the quality of the data-driven bandwidth selectors in Section 5.2;

(ii) performance under miss-specification of the asymmetric model;

(iii) effect of estimation of the index-parameter α;

(iv) performance of the semiparametric estimator $\widehat{q}_\beta(x)$ in (11), including a comparison with nonparametric local linear quantile estimator.

Items (ii) and (iv) are discussed in Section 7.2, whereas Section 7.3 deals with item (iii). Results regarding item (i) and some further simulation results can be found in Section S3 of the Supplementary Material. See also Section 7.4.

7.1. Simulation models, details of implementation and performance criteria

We consider a regression model

$$Y = \theta_1(X) + 0.1\varepsilon, \tag{35}$$

where ε has an asymmetric density $f_{X,\alpha}$ belonging to the family (7) which is of a standard type, i.e. with $\theta_1 = \theta_2 = 0$. Note that in simulation model (35) Y given $X = x$ has density of the form (10) with $\theta_2(x) = \ln(0.1) = -2.3026$, is a known constant. Hence in this simulation study we focus only on the estimation of $\theta_1(\cdot)$. The design density is: $X \sim U(-2, 2)$. For the error ε we consider two cases: (a) an asymmetric normal distribution (AND) and (b) asymmetric Laplace distribution (ALaD). For the unknown function $\theta_1(x)$, we considered two examples: $\theta_1(x) = x + 2e^{-16x^2}$ and $\theta_1(x) = \sin(x^2) + x + 2e^{-16x^2}$. In the first example the function $\theta_1(x)$ is a linear function with a superimposed bump, whereas in the second example the linear part is replaced by a more variable sinus function. For brevity we mainly report on results for the second example. Results for the first example lead to similar conclusions. In all sections, with exception of Section 7.3, the index-parameter α is fixed (known) and equals $\alpha = 0.25$.

In the estimation procedures we use the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$, take $p_1 = 1$ in the local fitting part, and w the indicator function on $[-2, 2]$. The number of grid points on the interval $[-2, 2]$ for which we obtain $\widehat{\theta}_1(x)$ is

101. We draw 100 samples of size $n = 100$ from each simulation model. For each sample we obtain the estimate of $\theta_1(\cdot)$, using the semiparametric procedure, by considering a conditional density (10), with f a standard normal density, as well as f a standard symmetric Laplace density. Note that when the model is considered with ε distributed according to an AND (respectively ALad) density, and we use in the estimation method a conditional density with f a symmetric Laplace density (respectively a normal density), this allows us to investigate the impact of miss-specifying the parametric model in the semiparametric estimation procedure.

To assess the quality of the estimation, for each method, we calculate the squared error for each grid point (in $\{x_1, \ldots, x_{n_{\text{grid}}}\}$, with $n_{\text{grid}} = 101$) and sum all squared errors:

$$\text{AISE} = \sum_{j=1}^{n_{\text{grid}}} \left(\widehat{\theta}_1(x_j) - \theta_1(x_j)\right)^2, \tag{36}$$

of which the average (across samples) is a finite-sample approximation of the theoretical mean integrated squared error.

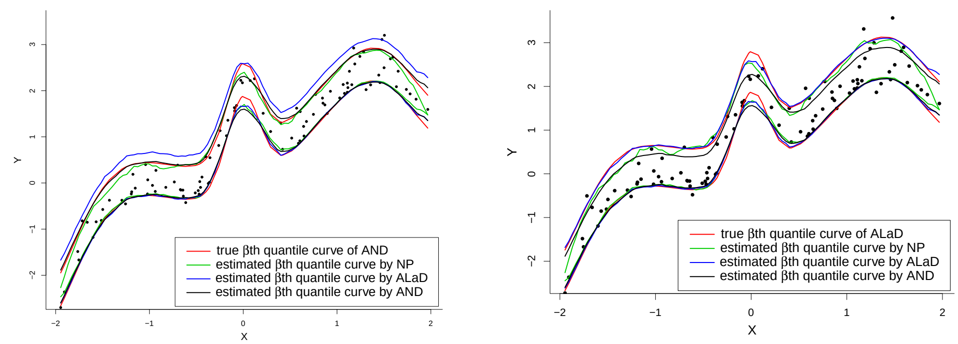
The semiparametric method of Section 3 is implemented in the R package `QBAsyDist` [9].

7.2. Performance of the estimators, including under model miss-specification

For α known we use the data driven-bandwidth of Section 5.2.3. For each simulated sample, h_{mean} is substituted by the Plug-in bandwidth selector for mean regression estimation (using the R package `locpol`). This results into estimated (data-driven) optimal bandwidths $\widehat{h}^{\text{QM}}_{\alpha,\text{ALaD}}$ and $\widehat{h}^{\text{QM}}_{\alpha,\text{AND}}$.

In Figures 2(a) and 2(b) we present the estimates for the βth quantile curves (means over all simulations), for $\beta = (0.10, 0.90)$. Figure 2(a) (respectively Figure 2(b)) refers to the case that the error distribution is AND (respectively ALaD). It is seen that all estimated conditional quantile curves under AND and ALaD (estimation) models are parallel, which was expected due to the constant θ_2. On the contrary, the nonparametrically estimated conditional quantile curves are not parallel (and might even cross each other).

The mean and standard error of the AISE-values are presented in Table 1. In each column the smallest mean AISE-value is indicted in bold. As can be seen, all AISE averages are smallest for the semiparametric method, provided the chosen estimation reference model is the correct one. Note however that even if the estimation model is wrongly specified, the semiparametric method



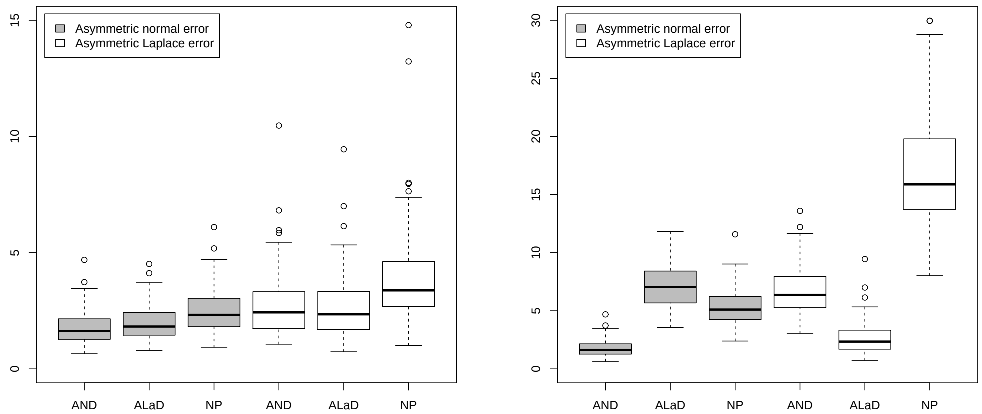
(a) (b)

Figure 2: Example 2. True $\beta = (0.10, 0.90)$th quantile functions of (a) AND with $\alpha = 0.25$; (b) ALaD with $\alpha = 0.25$, and its estimates using semiparametric ALaD and AND conditional densities and nonparametric approach.

Table 1: Example 2: Mean (standard error) of AISE-values across simulations for $\beta = (0.10, 0.25, 0.50, 0.75, 0.90)$. Abbreviations: AND using f is a standard normal density; ALaD: using f is a symmetric Laplace density; Nonp.: fully nonparametric quantile estimation.

Asymmetric Normal Error (AND)					
β	0.10	0.25	0.50	0.75	0.90
AND	**1.7476** (0.7043)	**1.7476** (0.7043)	**1.7476** (0.7043)	**1.7476** (0.7043)	**1.7476** (0.7043)
ALaD	1.9681 (0.7152)	1.9986 (0.7425)	1.9672 (0.7152)	2.4944 (0.9375)	7.1149 (1.7256)
Nonp.	2.5077 (0.9470)	1.9542 (0.7544)	2.2638 (0.7403)	3.4024 (1.0607)	5.3157 (1.4564)
Asymmetric Laplace Error (ALaD)					
AND	2.7357 (1.4162)	2.7134 (1.4061)	2.7368 (1.4164)	2.9210 (1.4249)	6.7044 (1.9965)
ALaD	**2.6904** (1.4055)	**2.6904** (1.4055)	**2.6904** (1.4055)	**2.6904** (1.4055)	**2.6904** (1.4055)
Nonp.	3.9767 (2.0488)	**2.6904** (1.4055)	3.6076 (1.8066)	8.4810 (4.1164)	17.4914 (6.8411)

often still outperforms the nonparametric method. This shows thus a certain robustness against model miss-specification. Figures 3(a) and 3(b) depict boxplots of the AISE values, from which it is clearly seen that the fully nonparametric estimation procedure has larger mean AISE-values and high variability compared to the semiparametric methods (even under model miss-specification).



(a) (b)

Figure 3: Example 2. Boxplot of AISE-values for estimated (a) 10% conditional quantile functions, for each; (b) 90% conditional quantile functions, for each.

7.3. Estimation of the index-parameter α

7.3.1. Simulation results for the case α is constant and unknown

We now include estimation of α in our simulations, applying the methods exposed in Section 6.1. For simplicity, we here consider a fixed bandwidth parameter h. We considered three values for h which are close to the theoretical optimal values of h and the average values of the data-driven bandwidths obtained in Section 7.2. See also Table S.1 in the Supplementary Material.

We consider regression model (35) with ε an Asymmetric Normal distribution (AND) with $\alpha = 0.25$. Our simulation results are based on the same 100 samples of size $n = 100$ as in Section 7.2. We consider two values for β, namely $\beta = \{0.25, 0.90\}$, and three fixed bandwidth values $h = \{0.075, 0.090, 0.095\}$. The simulation results for $h = 0.090$ are presented in Table 2, and these for other bandwidth values can be found in Table S.2 in Section S3.3.1 of the Supplementary Material. Tables 2 and S.2 list the mean and standard error of the AISE-values for estimation of $\theta_1(\cdot)$, when relying on one of the three methods for estimating α (indicated in column 1). Again, we consider the log-likelihood based on the true AND model, as well as that based on the miss-specified ALaD model. From Tables 2 and S.2 it can be seen that using Method 1 for estimation of α leads to the smallest mean AISE-values for estimation of $\theta_1(\cdot)$. When comparing the results, for Example 2, provided in Tables 2 and 1, it can be seen that the estimation of α has only a small effect on the AISE-values.

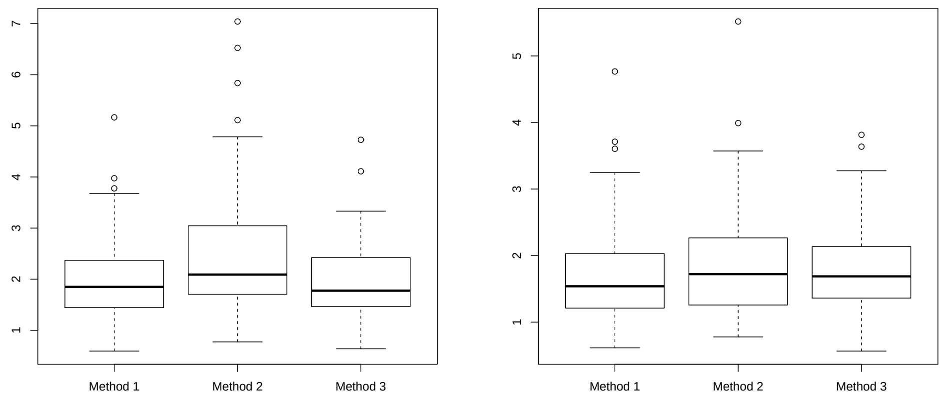
Table 2: Example 2 with Asymmetric Normal Error: Mean (standard error) of AISE-values based on 100 simulations, for $\beta = (0.25, 0.90)$. Using three different methods to estimate $\alpha : \widehat{\alpha}^{(j)}$ (Method j), $j = 1, 2, 3$.

Bandwidth $h = 0.090$					
		Example 1		Example 2	
Method	β	0.25	0.9	0.25	0.9
1	AND	1.9541 (0.7455)	2.9321 (1.7002)	1.6790 (0.7071)	1.8029 (0.7589)
	ALaD	2.8595 (1.4388)	3.1442 (2.0296)	2.8586 (1.4712)	4.2914 (1.5169)
2	AND	2.3858 (1.0460)	5.1145 (4.2681)	2.0427 (0.7601)	3.3649 (1.6695)
	ALaD	3.0394 (1.8936)	7.2960 (5.0737)	3.5570 (1.8193)	9.5770 (2.8737)
3	AND	1.9256 (0.7659)	3.0923 (3.5634)	1.9368 (0.7349)	2.9227 (0.9309)
	ALaD	2.8890 (1.6768)	5.3327 (2.5925)	2.9120 (1.4572)	5.8714 (2.0696)

Figure 4 presents boxplots of the AISE-values for estimation of $\theta_1(\cdot)$, in the left panel for Example 1, and in the right panel for Example 2. The boxplots confirm the conclusions from Tables 2 and S.2. For both examples, it is clearly seen that the smallest variation is found when using Method 1. The variations of AISE-values for estimation of $\theta_1(\cdot)$, are reasonably close when using either Method 1 or Method 3 (to estimate α).

For looking in more detail into the quality of estimation of the index-parameter α, we present in Figure 5 boxplots of the ASE-values for estimation of α: $\text{ASE}(\widehat{\alpha}^{(j)}) = (\widehat{\alpha}^{(j)} - \alpha)^2$ for Methods 1—3. It is seen that when using Method 1 for estimating α the resulting mean and variation of the ASE-values are smallest among the three methods (Methods 1—3). This was to be expected since only Method 1 exploits the parametric structure of model (10). The highest variation is observed when using Method 2 to estimate the index-parameter α. These conclusions hold for both simulation examples.

In Section S3.3.2 in the Supplementary Material we provide in Table S.3 the average computing time for estimation of $\theta_1(\cdot)$ when using the different methods for estimating α. As can be observed the computational cost is up to about a factor 4 higher than when using Methods 2 or 3. This is not surprising since the optimization problem involved has to be done for one parameter more, and the optimization problem needs to be solved for a grid



(a) (b)

Figure 4: Examples 1 and 2 with Asymmetric Normal Error and $\alpha = 0.25$: Boxplots of AISE values for estimation of $\theta_1(\cdot)$, for $\beta = 0.25$, using bandwidth $h = 0.090$ and three different methods to estimate α: $\widehat{\alpha}^{(j)}$ (Method j), $j = 1, 2, 3$. Results for (a) Example 1; and (b) Example 2.

of x_j-values.

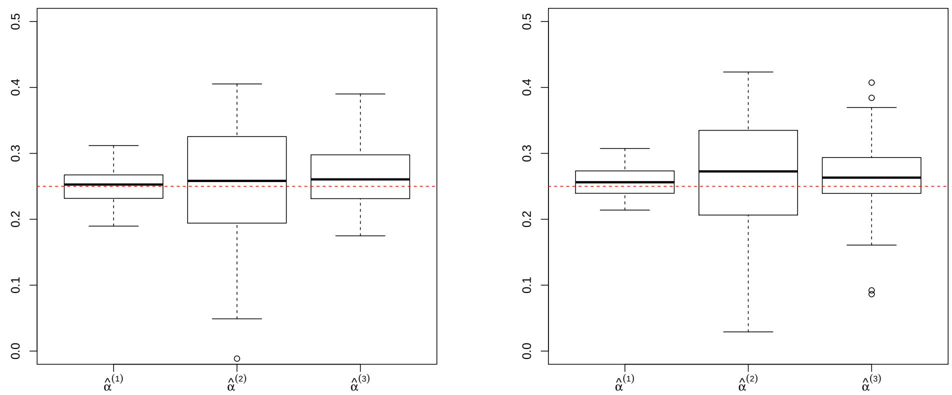
7.3.2. Simulation results for the case of an unknown function $\alpha(x)$

We next investigate the finite-sample performance of the method described in Section 6.2. Therefore, we consider an extension of Example 2 with the constant θ_2 and α in Example 2 replaced by

$$\theta_2(x) = 1 - 0.5x - 1.5x^2 \qquad \text{and} \qquad \alpha(x) = \frac{e^{1+3x}}{1 + e^{1+3x}}. \tag{37}$$

We use a cross-validated bandwidth $\widehat{h}^{\mathrm{CV}}$ for estimating $\theta_1(x)$, $\theta_2(x)$ and $\alpha(x)$, and subsequently the quantile function $q_\beta(x)$.

We simulated 100 samples for the simulation model in this extended Example 2, and this for sample sizes $n = 100$, and $n = 200$. We report on the quality of estimation of $\theta_1(\cdot)$, $\theta_2(\cdot)$, $\alpha(\cdot)$ and $q_\beta(\cdot)$ using an AISE-type criterion (36) adapted to each of these target functions. Table 3 presents mean and standard errors of the AISE-values for estimation of $\mu(\cdot) = \theta_1(\cdot)$, $\phi(\cdot) = \exp\{\theta_2(\cdot)\}$ and $\alpha(\cdot)$. Table 3 summarizes the AISE-values for estimation of the quantile function $q_\beta(\cdot)$, for five values of β. Table 3 also allows, to



(a) (b)

Figure 5: Examples 1 and 2 with Asymmetric Normal Error and $\alpha = 0.25$: Estimates of α based on 100 simulations for fixed bandwidth $h = 0.090$, using three different methods: $\widehat{\alpha}^{(j)}$ (Method j), $j = 1, 2, 3$. Results for (a) Example 1; and (b) Example 2.

a certain extent, the see the effect when having to estimate three unknown functions ($\theta_1(x)$, $\theta_2(x)$ and $\theta_3(x)$), instead of one function ($\theta_1(x)$). The function $\theta_1(x) = \mu(x)$ is the same as in Example 2. Note that the AISE-values for estimation of $\theta_1(x)$ are only slightly higher than in Table 1. Since estimation of the quantile function (see (34)) relies on estimation of $\theta_1(x)$, $\theta_2(x)$ and $\theta_3(x)$, the estimation errors are also higher in Table 4. Tables 3 and 4 also show the effect of the sample size on the finite-sample performance.

7.4. Further simulation results

In Section S3 in the Supplementary Material we investigate the quality of the data-driven bandwidth selector discussed in Section 5.2.3. The main finding is: although the data-driven Quantile-Mean based bandwidth selector is only a rough bandwidth selector, and tends to be larger than the theoretical optimal bandwidth, it overall produces good quality (quantile) estimators. Herein the quality of an estimator is measured according to a criterion that approximates the mean integrated squared error. See (36). In Section S3 we also present estimated quantiles curves together with confidence bands.

Table 3: Extended Example 2 ((37)): Mean (standard error) of AISE-values for estimation of $\mu(\cdot)$, $\phi(\cdot)$ and $\alpha(\cdot)$. Abbreviations: AND using f is a standard normal density and ALaD: using f is a symmetric Laplace density. Results are for sample sizes $n = 100$ and $n = 200$.

Asymmetric Normal Distribution (AND)				
	n	$\widehat{\mu}(\cdot)$	$\widehat{\phi}(\cdot)$	$\widehat{\alpha}(\cdot)$
AND	100	1.9591 (1.8421)	1.1251 (0.4732)	0.0452 (0.0412)
	200	1.6231 (1.2421)	1.0125 (0.3455)	0.0345 (0.0381)
ALaD	100	2.0140 (2.0214)	1.5498 (0.5155)	0.0421 (0.0515)
	200	1.7583 (1.1523)	1.3242 (0.4215)	0.0312 (0.0318)
Asymmetric Laplace Distribution (ALD)				
AND	100	2.1256 (1.9957)	1.2520 (0.5153)	0.0561 (0.0414)
	200	1.8453 (1.1245)	1.0245 (0.3215)	0.0351 (0.0341)
ALaD	100	1.8759 (1.4683)	1.1995 (0.4210)	0.0486 (0.0401)
	200	1.4125 (1.0065)	1.0152 (0.3154)	0.0394 (0.0315)

Table 4: Extended Example 2 ((37)): Mean (standard error) of AISE-values for estimation of $q_\beta(\cdot)$ across simulations for $\beta = (0.10, 0.25, 0.50, 0.75, 0.90)$. Abbreviations: AND using f is a standard normal density; ALaD: using f is a symmetric Laplace density; Nonp.: fully nonparametric quantile estimation. Results on the first row are for $n = 100$, on the second row for $n = 200$.

Asymmetric Normal Error (AND)					
β	0.10	0.25	0.50	0.75	0.90
AND	10.7369 (4.7021)	7.7369 (3.6793)	7.6561 (2.2415)	11.1543 (4.1552)	14.4535 (5.5340)
	6.8428 (3.0531)	4.5431 (2.7421)	3.1786 (1.7421)	5.2656 (2.1046)	7.1542(3.1765)
ALaD	11.5835 (5.0700)	8.1835 (3.9421)	8.5835 (2.5520)	14.4824 (4.7051)	19.2412 (7.0421)
	6.9420 (3.4432)	4.6421 (2.9814)	3.4312 (1.9931)	6.4881 (2.1635)	12.4586 (4.4610)
Nonp	15.2276 (8.8447)	9.4331 (4.2960)	9.9925 (3.1445)	17.4612 (6.1834)	20.4921 (13.3412)
	9.5923 (4.2185)	5.1415 (3.2561)	5.1521 (2.9431)	10.1592 (3.9813)	15.1488 (7.1541)
Asymmetric Laplace Error (ALD)					
AND	12.2601 (5.8241)	7.3701 (4.3942)	8.1862 (2.5869)	12.4815 (5.1565)	17.7625 (9.1543)
	7.4881 (3.8431)	2.8431 (2.1542)	3.1540 (1.6121)	7.1543 (3.1550)	11.1582 (5.1560)
ALaD	11.2172 (5.7481)	8.0291 (4.0281)	6.1876 (2.1492)	10.4982 (4.1875)	14.1434 (7.1574)
	7.0164 (3.7462)	2.0598 (2.0482)	2.3614 (1.5421)	5.1698 (2.1065)	8.1681 (4.0450)
Nonp	16.4260 (8.4361)	10.0060 (5.4441)	8.1983 (3.1986)	14.8541 (7.1422)	24.1544 (10.1572)
	10.1421 (4.4550)	3.5421 (2.7921)	4.0158 (1.9923)	8.1889 (4.3211)	14.1581 (6.1572)

8. Real data application: Maximum wind speed in hurricanes

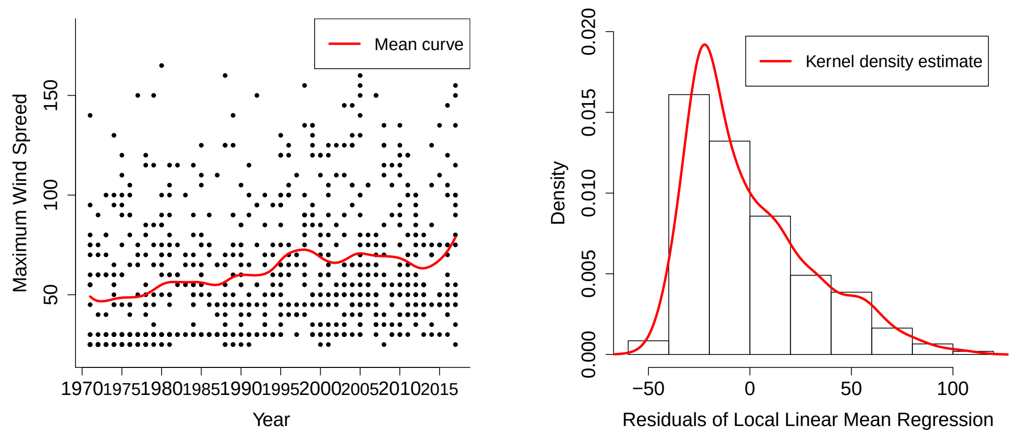
We illustrate the use of the proposed semiparametric estimation method on two real data examples.

8.1. Maximum wind speed in hurricanes

The National Hurricane Center (NHC) conducts a post-storm analysis of each tropical cyclone in its area of responsibility to determine the official assessment of the cyclone's history. These analyses lead to the North Atlantic hurricane database (or HURDAT). We consider this database of size $n = 764$ with response variable (Y) the maximum wind speed of a tropical hurricane and covariate (X) the year of its occurrence between 1971 to 2017. The winds were measured in knots (where one knot is equal to 1.15 miles per hour). The data are available for download at `https://www.nhc.noaa.gov/data/hurdat/` and via the R-package `HURDAT`. Part of this dataset (the period 1981–2006) have been analyzed in the literature, e.g. in [15, 16]. [17] fitted a linear quantile regression model and reported that the strongest tropical cyclones in the North Atlantic basin have gotten stronger over the last couple of decades. [15] used simultaneous linear quantile regression estimation in a context of semiparametric Bayesian analysis and reported that not only the upper tail of the intensity distribution but also the entire range of the intensity distribution has gotten stronger during the period 1981–2006. [16] used B-spline basis functions in nonparametric simultaneous quantile regression analysis and found an increasing pattern of the higher quantile curves during the periods 1987–1994 and 2002–2005 while a decreasing pattern was found prominent during 1994–2002.

We analyze the available data for the period 1971–2017 using the semiparametric method exposed in Section 3.2. A scatter plot of the data together with a local linear estimate $\widehat{m}$ (using a Gaussian kernel) of the mean function $\mathbb{E}(Y \mid X = x)$ is in Figure 6 (a). A density estimate based on the residuals $Y_i - \widehat{m}(X_i)$ is presented in Figure 6 (b). From this plot, it is clearly seen that the residual data are right skewed.

For real data, and without any prior knowledge, one does not know whether it is reasonable to assume α to not depend on the covariate x. We therefore start our analysis by estimating $\alpha(x)$ as described in Section 6.2. The results here are for f a Laplace density. The resulting estimated $\widehat{\alpha}(x)$ is depicted in Figure 7, as a scatterplot of $\widehat{\alpha}(X_i)$ and a smooth curve of these. Some summary statistics describing the variability of $\widehat{\alpha}(\cdot)$ over all values of



(a) (b)

Figure 6: Hurricane data. (a) Scatter plot of the data and estimated mean function $\widehat{m}$ using local linear fitting; (b) Kernel density estimate based on the residuals $Y_i - \widehat{m}(X_i)$.

X_i are provided in the first row of Table 5. From this it appears that it is reasonable to assume α to be a constant. We thus, in further analysis, work under this assumption.

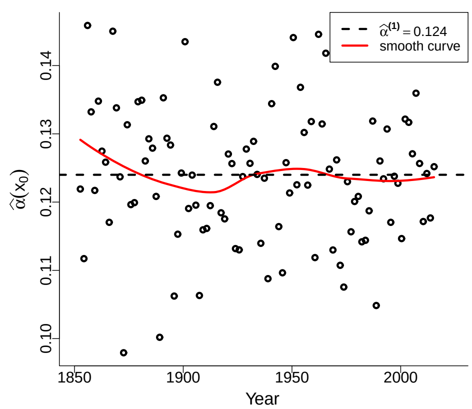


Figure 7: Hurricane data. Estimated $\widehat{\alpha}(x)$ (smooth curve) together with the estimate $\widehat{\alpha}^{(1)}$ assuming α is a constant.

In Section 6.1 we discussed three methods to estimate α. In the second column of Table 6 we present the estimated values for α for each of the

Table 5: Hurricane data and Bone density data. Summary statistics for estimated $\widehat{\alpha}(x)$.

	Min	First quartile	Second quartile	Mean	Third quartile	Max
Hurricane	0.0979	0.1171	0.1238	0.1338	0.1306	0.1459
Bone Density	0.2848	0.3287	0.3425	0.3556	0.3572	0.4291

three methods. All these values confirm what was already suspected from Figure 6 (b) (similar figure for Method 3, not presented here): there is a clear asymmetry present in the residual observations. Following the findings of Section 6.1 we use $\widehat{\alpha} = 0.1240$ in the next steps of our analysis.

Table 6: Hurricane data. The estimated $\widehat{\alpha}$, the selected distribution and the data-driven bandwidth values.

Method	$\widehat{\alpha}$	Selected Density	P-value	AIC-value	$\widehat{h}_{\alpha}^{\text{ROT}}$	$\widehat{h}_{\alpha}^{\text{CV}}$
1	0.1240	Asymm. Laplace	0.8895	7017.427	1.9507	2.0425
2	0.1650	Asymm. Laplace	0.9990	7130.5480	2.1811	2.1130
3	0.1453	Asymm. Laplace	0.9941	7110.4810	1.7203	1.8670

With the estimated α at hand, we seek for finding an appropriate (unconditional) asymmetric density to fit to the data. As candidate densities we consider asymmetric Laplace, normal, Student's-t and logistic densities ((7) with these listed choices for f). For each candidate density we test its appropriateness via the Kolmogorov Smirnov goodness-of-fit test (the P-value for the selected density is given in column 4 of Table 6). In the third column of Table 6 we list the selected density model, when using the specified method for estimating α. For all methods the same density was selected (among the four candidate asymmetric densities). For each method we provide for the selected density the AIC-value of the corresponding likelihood in column 5 of Table 6.

Note that for the selected conditional density, which is an Asymmetric Laplace density, there is a clear interpretation of the parameter functions involved. See [8]. Indeed, with α a constant, $\mu(x)$ is the mode of the conditional density, and the mean and the variance of the conditional density of

Y given $X = x$, are

$$\mathbb{E}(Y|x) = \mu(x) + \frac{\phi(x)\,[1-2\alpha]}{\alpha\,(1-\alpha)} \quad \text{and} \quad \mathrm{Var}(Y|x) = \frac{\phi^2(x)\,[1-2\alpha+2\alpha^2]}{\alpha^2\,(1-\alpha)^2}. \tag{38}$$

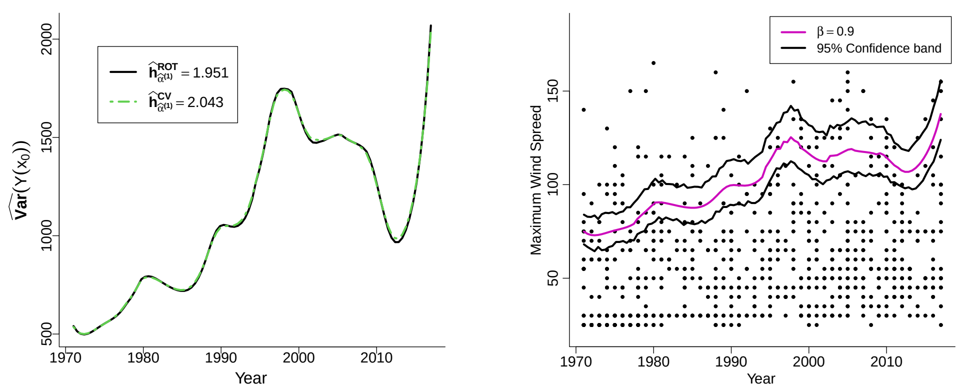
Furthermore we know that the mode $\mu(x)$ equals the conditional αth quantile of Y. Using the estimators for α, $\mu(\cdot)$ and $\phi(\cdot)$ we thus obtain subsequently estimates for this conditional mean and variance.

For estimating $\theta_1(\cdot)$ and $\theta_2(\cdot)$ we take $p_1 = p_2 = 1$. For Methods 2 and 3, we present the values for $\widehat{h}_\alpha^{\mathrm{ROT}}$ and $\widehat{h}_\alpha^{\mathrm{CV}}$ in columns 6 and 7 of Table 6. For Method 1, we used as bandwidths just the averages of the bandwidths for Methods 2 and 3. Using these bandwidths for Method 1 (to estimate α) we obtain the estimates for $\boldsymbol{\theta}_1(x_0)$ and $\boldsymbol{\theta}_2(x_0)$ via (13). An estimate for the βth quantile curve $\widehat{q}_\beta(x_0)$ is then obtained from (11).

Recall that Method 1 resulted in $\widehat{\alpha} = 0.1240$. The $\beta = (\widehat{\alpha}, 0.5, 0.90, 0.95)$th estimated conditional quantile curves $\widehat{q}_\beta(\cdot)$, using the bandwidth $\widehat{h}_\alpha^{\mathrm{ROT}}$, are depicted in Figure 1(b). Recall that for $\beta = 0.5$ we actually get the estimated median curve.

Further, we depict in Figure 8(a) the estimated conditional variance function which is under the selected model, just a rescaling of the estimated $(\widehat{\phi}(x))^2$, as can been seen from (38). Two estimates of the conditional variance are shown in Figure 8(a), using the Rule of Thumb bandwidth and the cross-validation bandwidth selectors of Section 5.2. The two estimates are almost indistinguishable, and indicate the non-constant and non-linear pattern of the variance function (and the scale function) over the years. Of interest is also to note the peak in estimated variance of the maximum wind speed of hurricanes in the second half of the nineties. This could be linked with a so-called super El Niño event, which began in the spring months of 1997.

Figure 8(b) presents the estimated 90%th quantile curve, together with an estimated 95% Bonferroni-type confidence band. Since for obvious reasons the interest mainly goes to hurricanes with very high maximum speeds, it is most interesting to look at a high quantile, such as the 90%th quantile. For constructing confidence bands we rely on the asymptotic normality result established in Theorem 4.3. This however also requires to estimate (asymptotic) bias and variance. There is an overall increasing trend in all presented quantile curves. However, the signs and intensities of the changes in maximum wind speed are not constant in the considered time period.



(a) (b)

Figure 8: Hurricane data. (a) Estimated conditional variance of the maximum wind speed, as a function of years. (b) Estimated β% conditional quantile curve (using $\widehat{h}_{\alpha}^{\mathrm{ROT}}$), for $\beta = 0.90$, together with 95% Bonferroni-type confidence band.

Figure 1(b) shows an increasing pattern of the higher quantile curves during the periods 1971–1997, 2001–2005 and 2012–2017 while a decreasing pattern is prominent during 1997–2001 and 2005–2012. These nonlinear patterns should be further investigated, possibly coupled with other meteorological or environmental phenomena.

For comparison purpose, we also estimate the quantile functions nonparametrically. See Section S1. The performances of the semiparametric method (using the three methods for estimating α) and nonparametric estimators of the quantile curve are evaluated via prediction errors calculated through a cross-validation method. We split the full sample into a training set (used for estimation) with approximately 80% of the observations (i.e. $n_e = 611$), and 20% (i.e. $n_{\mathrm{pred}} = n - n_e = 153$) is allocated to a test set for validation/evaluation. The observations are chosen randomly without replacement. Based on these splitted subsamples, we compute the predicted error defined as $\frac{1}{n_{\mathrm{pred}}} \sum_{i=1}^{n_{\mathrm{pred}}} \rho_\beta \left(Y_i - \widehat{q}_\beta(X_i) \right)$ for all three semiparametric methods (abbreviated as SP(Method j) in the plots), and the nonparametric quantile estimator. We repeat this $R = 99$ times and present the boxplots of the obtained prediction error values for all methods in Figure 9. It is

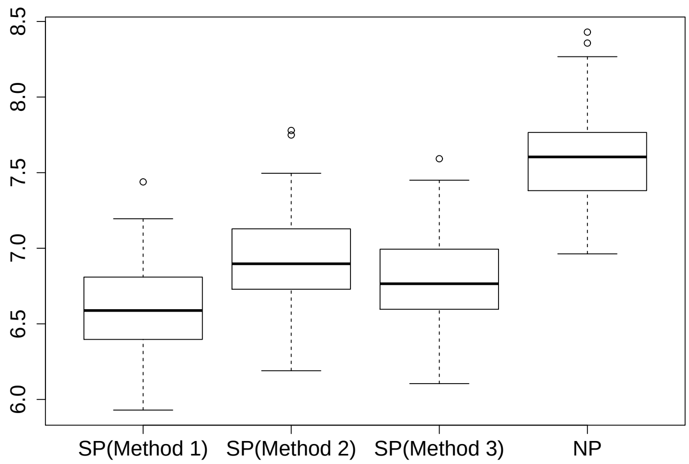


Figure 9: Hurricane data. The prediction error of the 90%th quantile curve obtained by using semiparametric (SP) and nonparametric (NP) approach.

clearly seen that the predictive performance for the semiparametric method is better than that of the nonparametric method, no matter which method for estimation of α was used.

8.2. Real data application: Bone density

We consider a dataset concerning the actual measurements of bone density (BMD) of $n = 485$ adolescents. These data were originally reported in [18], and are for example available via `https://web.stanford.edu/~hastie/ElemStatLearn/datasets/bone.data`, and were analysed in e.g. [19]. [20] used a quadratic programming method for estimating quantiles in a nonparametric quantile regression context.

Measuring the bone mass in children can help to understand the predisposition to suffer from osteoporosis in more advanced age. Osteoporosis involves the loss and a consequent weakening of the bone tissue and affects mostly women. Medical research proved that high peak bone density in early age reduces osteoporosis risk at later age. It is therefore of interest to investigate how the peak bone mass varies in children and young adults, and in particular when it achieves its maximal value. The response variable is the change in spinal bone mineral density value and the covariate the age of the adolescents.

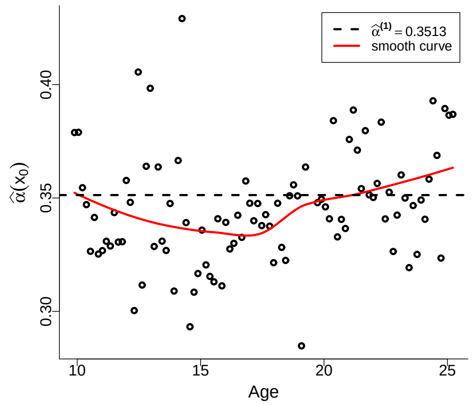


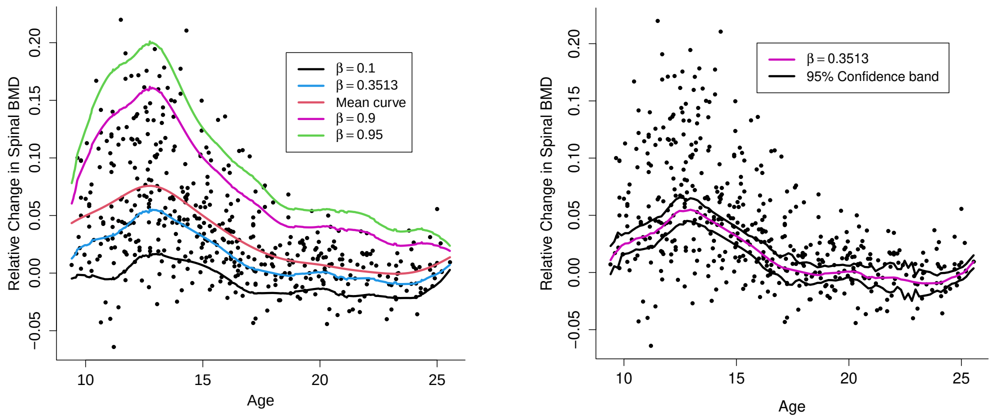
Figure 10: Bone Density data. Estimated $\widehat{\alpha}(x)$ (smooth curve) together with the estimate $\widehat{\alpha}^{(1)}$ assuming α is a constant.

We analyse the data using the semiparametric method of Section 3.2, in a similar fashion as for the Hurricane data. The estimate $\widehat{\alpha}(x)$ is depicted in Figure 10. Summary statistics about the values of this estimate are in the second row of Table 5. Also in this example, it seems reasonable to assume a constant α in further analysis. The results from the three methods for estimating α, as discussed in Section 6.1, are presented in Table 7. In our discussion further we focus on the semiparametric method using Method 1 to estimate α.

Figure 11(a) shows a scatter plot together with estimated quantile curves. Note the smaller variability in the data for higher ages. The residual data $Y_i - \widehat{m}(X_i)$ (not shown here) are slightly positively skewed, which indicates that the conditional error distribution is right skewed. This was also reported in [20].

Table 7: Bone density data. The estimated $\widehat{\alpha}$, the selected distribution and the data-driven bandwidth values.

Method	$\widehat{\alpha}$	Selected Density	P-value	AIC-value	$\widehat{h}_{\alpha}^{\mathrm{ROT}}$	$\widehat{h}_{\alpha}^{\mathrm{CV}}$
1	0.3513	Asymm. Laplace	0.9935	−1772.198	0.6224	0.890
2	0.4101	Asymm. Laplace	0.9990	−1772.198	0.6170	0.873
3	0.3724	Asymm. Laplace	0.9939	−1728.697	0.6277	0.894



(a) (b)

Figure 11: Bone density data. (a) Semiparametrically estimated βth quantile curves, using an ALaD density. (b) Estimated β% conditional quantile curve (using $\widehat{h}_{\alpha}^{\text{ROT}}$), for $\beta = \widehat{\alpha} = 0.3513$, together with 95% Bonferroni-type confidence band.

An appropriate asymmetric density for the residual data was found to be also an asymmetric Laplace density (with an associated P-value of 0.9935 for the Kolmogorov-Smirnow goodness-of-fit test, and an AIC-value of -1772.198, smallest amongst all considered asymmetric density models). Fitting an asymmetric Laplace density to the data, led to the maximum likelihood estimator $\widehat{\alpha}^{(1)} = 0.3513$. See also Table 7.

Using (10) with f the symmetric Laplace density and $\alpha = 0.3513$, we obtain the estimated location and scale function (with $p_1 = p_2 = 1$) from (13). For the Rule-of-Thumb bandwidth selector we here use the weight function $w_0(x) = \mathbb{I}_{[9.4,25.55]}(x)$, leading to $\widehat{h}_{\alpha}^{\text{ROT}} = 0.6224$. The cross-validation bandwidth selector for these data is $\widehat{h}_{\alpha}^{\text{CV}} = 0.890$. The estimated conditional quantile curves for orders $\beta = (\widehat{\alpha}, 0.50, 0.90, 0.95)$, using $\widehat{h}_{\alpha}^{\text{ROT}}$, are presented in the Figure 11(a). Note that the estimated quantile curves also clearly represent the reduced variability for higher ages. The estimated β quantile curve, with $\beta = 0.3513$, together with 95% Bonferroni-type confidence band is depicted in Figure 11(b). Recall that this quantile is nothing but the conditional mode of the density. Note that looking at the mode, the peak of relative change in bone density is appearing around the age of 12–13.

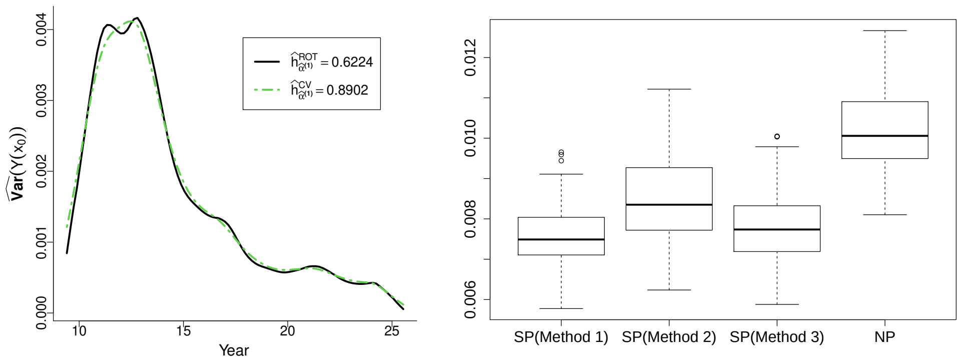


Figure 12: Bone density data. (a) Estimated conditional variance of the maximum wind speed, as a function of years. Right panel: Estimated conditional variance of the change in spinal bone mineral density. (b) The prediction error of the 90%th quantile curve obtained by using ALaD likelihood and nonparametric approach.

We estimated the function $\theta_2(\cdot)$, using either the Rule-of-Thumb bandwidth selector or the cross-validation bandwidth selector. Based on this estimate and the estimated value for α, we obtain the estimated conditional variance via (38). Figure 12 (a) presents the estimated conditional variance of change in spinal bone mineral density value, in function of its evolution with age. Figure 12(a) reveals the importance of the evolution in the early teenage years.

The cross-validated prediction error for estimation of the 0.90 quantile curve, for both the studied semiparametric method and the nonparametric method, are presented in Figure 12(b). Also here the conclusion is that the prediction error for the semiparametric method SP (using the various methods for estimating α), which better exploits the particular asymmetry (even if small) present in the data, has a lower estimated prediction error.

9. Further discussion

In this paper we study a semiparametric method to estimate regression quantiles. The key advantage of the method is that it originates from a family of asymmetric distributions. Hence asymmetry present in data can be exploited in this way. Local likelihood techniques are at the core of the method, which therefore involves a bandwidth parameter. As in all local modeling frameworks, the choice of the bandwidth needs to be studied, and underpinned by theoretical considerations. A starting point hereby is the study of asymptotic bias and variance, and obtaining optimal bandwidth parameters by balancing the squared bias and variance of estimators. In the paper several approaches to a practical bandwidth selector are explored. One approach is based on a rule of thumb kind of method; and another approach exploits the link between optimal bandwidths for mean and quantile estimation. The theoretical derivations in Section 5 eventually focuses on estimation of $\mu(x)$, and thus is somewhat limited. It would be of interest to investigate bandwidth selectors in general, considering in full estimation of $\mu()$, $\phi(x)$ and $\alpha(x)$. Nevertheless, all three practical bandwidth selectors represented in Section 5 perform, according to our extensive experiences, quite well.

The semiparametric framework involves the choice of the parametric component f (the symmetric reference density). In the real data applications, a set of candidates for f are considered, and via goodness-of-fit testing and an AIC-criterion one of these candidate densities is selected as the final parametric component. It would be of interest to study in detail the model selection issue that appears here.

The parameter α in the considered framework reflects the amount of asymmetry present in the data. The semiparametric estimation method can also deal with the cases that (i) α is constant and unknown, and (ii) $\alpha(x)$ is an unknown function. This is exposed in Section 6. The asymptotic properties of the semiparametric estimators are proven so far only for the case that α is a known constant. Establishing asymptotic theory for the more general settings would be very interesting (and tedious) and is part of future research.

Acknowledgements

The authors thank an Associate Editor and reviewers for their valuable comments which led to a considerable improvement of the manuscript. The authors gratefully acknowledge support from the Flemish Science Foundation (FWO research project G.0826.15N). The first and second authors acknowledge support of the GOA project GOA/12/014 of the Research Council KU Leuven. The third author is grateful for the support from the Flemish Science Foundation (FWO research grant 1518917N), and from the Special Research Fund (Bijzonder Onderzoeksfonds) of Hasselt University.

References

[1] J. Fan, N. E. Heckman, M. P. Wand, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, Journal of the American Statistical Association 90 (429) (1995) 141–150.

[2] J. Fan, I. Gijbels, Local Polynomial Modelling and Its applications, CRC Press, 1996.

[3] J. Fan, M. Farmen, I. Gijbels, Local maximum likelihood estimation and inference, Journal of the Royal Statistical Society: Series B 60 (3) (1998) 591–608.

[4] R. Koenker, G. Bassett Jr, Regression quantiles, Econometrica 46 (1) (1978) 33–50.

[5] R. Koenker, Quantile Regression, Cambridge University Press, 2005.

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

[7] K. Yu, M. Jones, Local linear quantile regression, Journal of the American statistical Association 93 (441) (1998) 228–237.

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

[9] R. Karim, I. Gijbels, A. Verhasselt, QBAsyDist: Asymmetric Distributions and Quantile Estimation, R-package version 0.1.2, 2019. URL `https://CRAN.R-project.org/package=QBAsyDist`

[10] P. Huber, The behavior of maximum likelihood estimates under nonstandard conditions, in: Proceedings of the fifth Berkeley symposium on Mathematical Statistics and Probability, Vol. 1, 1967, pp. 221–233.

[11] W. K. Newey, D. McFadden, Large sample estimation and hypothesis testing, Handbook of econometrics 4 (1994) 2111–2245.

[12] M. Aerts, G. Claeskens, Local polynomial estimation in multiparameter likelihood models, Journal of the American Statistical Association 92 (440) (1997) 1536–1545.

[13] T. Gasser, H.-G. Muller, V. Mammitzsch, Kernels for nonparametric curve estimation, Journal of the Royal Statistical Society: Series B 47 (2) (1985) 238–252.

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

[15] S. T. Tokdar, J. B. Kadane, et al., Simultaneous linear quantile regression: a semiparametric Bayesian approach, Bayesian Analysis 7 (1) (2012) 51–72.

[16] P. Das, S. Ghosal, Bayesian non-parametric simultaneous quantile regression for complete and grid data, Computational Statistics & Data Analysis 127 (2018) 172–186.

[17] J. B. Elsner, J. P. Kossin, T. H. Jagger, The increasing intensity of the strongest tropical cyclones, Nature 455 (7209) (2008) 92.

[18] L. K. Bachrach, T. Hastie, M.-C. Wang, B. Narasimhan, R. Marcus, Bone mineral acquisition in healthy asian, hispanic, black, and caucasian youth: a longitudinal study, The Journal of Clinical Endocrinology & Metabolism 84 (12) (1999) 4702–4712.

[19] T. Hastie, R. Tibshirani, J. Friedman, in: The Elements of Statistical Learning, Springer, 2009.

[20] I. Takeuchi, Q. V. Le, T. D. Sears, A. J. Smola, Nonparametric quantile estimation, Journal of Machine Learning Research 7 (2006) 1231–1264.

SUPPLEMENTARY MATERIAL

to "Semiparametric quantile regression using family of quantile-based asymmetric densities"

by
Irène Gijbels, Rezaul Karim and Anneleen Verhasselt

This supplement contains the following additional parts:

- Section S1: nonparametric local polynomial conditional quantile estimation;
- Section S2: proofs of Theorem 4.1, Lemmas S2.1 and S2.2 and of Theorems 4.2 and 4.3
- Section S3: additional results regarding the simulation study;
- Section S4: real data application - bone density data.

S1. Nonparametric conditional quantile estimators

S1.1. Asymptotic results for nonparametric conditional quantile estimators

As mentioned in Remark 3.1 the local polynomial maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_1(x_0)$ obtained in (14) under an asymmetric Laplace likelihood and the fully nonparametric estimator $\widehat{\boldsymbol{\theta}}_\beta(x_0)$ obtained in (3) are identical only for $\alpha = \beta$ due to using an exactly same loss function $\rho_\beta(\cdot)$ when the scaling function θ_2 is constant and known. Following the same lines of proofs as for Theorem 4.1 we can thus derive the asymptotic distribution of the fully nonparametric local pth polynomial estimator $\widehat{\boldsymbol{\theta}}_\beta(x_0)$, and hence subsequently that for the vth derivative of the βth conditional quantile function $q_\beta(\cdot)$.

It suffices to consider in our framework the log-likelihood function $\ell(u) = -u(\beta - \mathbb{I}(u \leqslant 0))$ with $\psi_1(u) = \mathbb{I}(u \leqslant 0) - \beta$. Assumption **(A5)** needs to be replaced by

(A5′) The function $q_\beta(\cdot)$ has a $(p+1)$th (respectively $(p+2)$nd) continuous derivative for p odd (respectively p even).

The mathematical derivation for proving the asymptotic normality result for the fully nonparametric conditional quantile estimator is similar to the proof of Theorem 4.1 but now the log-likelihood function only involves the unknown parameter $\boldsymbol{\theta}_\beta$. A main difference in this setting is in the calculation of (15), for which now

$$\begin{aligned}
\lambda_1(\theta_1(x)) &= \mathbb{E}_{Y|X}[\mathbb{I}(Y \leqslant \rho_\beta(X)) - \beta \mid X = x] = 0 \\
\lambda_{11}(\theta_1(x)) &= \frac{\partial}{\partial u}\mathbb{E}_{Y|X}[\beta - \mathbb{I}(Y \leqslant u) \mid X = x]\Big|_{u=q_\beta(x)} = -f_{Y|X}(q_\beta(x) \mid x).
\end{aligned}$$

This then leads to the following expressions for the crucial matrices

$$\begin{aligned}
\boldsymbol{\Sigma}_{x_0} &= \left(f_{Y|X}\left(q_\beta(x_0)|x_0\right) f_X(x_0)\right) \boldsymbol{N}_{pp}(x_0) \\
\boldsymbol{\Gamma}_{x_0} &= \left(\beta(1-\beta) f_X(x_0)\right) \boldsymbol{T}_{pp}(x_0) \\
\boldsymbol{\Lambda}_{x_0} &= \frac{d}{dx_0}\left(f_{Y|X}\left(q_\beta(x_0)|x_0\right) f_X(x_0)\right) \boldsymbol{Q}_{pp}(x_0),
\end{aligned}$$

where $f_{Y|X}(\cdot \mid x)$ is the (unknown) conditional density of the response Y given $X = x$, which is assumed to be continuous and for which we assume $f_{Y|X}\left(q_\beta(x_0)|x_0\right) > 0$. Theorem S1.1 states the asymptotic normal distribution result for the fully nonparametric local polynomial quantile estimator.

Theorem S1.1. Assume **(A5′)** and **(A6)**–**(A10)** hold. For given x_0 we have:

(i) If $p - v$ odd ($v = 0, \ldots, p$), then for $n \to \infty$,

$$\sqrt{nh^{2v+1}} \left(\frac{\beta(1-\beta)}{\left[f_{Y|X}\left(q_\beta(x_0)|x_0\right)\right]^2 f_X(x_0)} \int K_{v,p}^2(u)du \right)^{-1/2}$$
$$\times \left[\widehat{q}_{\beta,v}(x_0) - q_\beta^{(v)}(x_0) - h^{p-v+1} \frac{q_\beta^{(p+1)}(x_0)}{(p+1)!} \left\{ \int u^{p+1} K_{v,p}(u)du \right\} \{1 + O_P(h)\} \right]$$
$$\xrightarrow{D} \mathcal{N}(0,1);$$

(ii) If $p - v$ even ($v = 0, \ldots, p$), then for $n \to \infty$,

$$\sqrt{nh^{2v+1}} \left(\frac{\beta(1-\beta)}{\left[f_{Y|X}\left(q_\beta(x_0)|x_0\right)\right]^2 f_X(x_0)} \int K_{v,p}(u)^2 du \right)^{-1/2}$$
$$\times \left[\widehat{q}_{\beta,v}(x_0) - q_\beta^{(v)}(x_0) - \left(\frac{q_\beta^{(p+2)}(x_0)}{(p+2)!} \int u^{p+2} K_{v,p}(u)du \right.\right.$$
$$+ \frac{q_\beta^{(p+1)}(x_0)}{(p+1)!} \frac{\frac{d}{dx_0}\left(f_{Y|X}\left(q_\beta(x_0)|x_0\right) f_X(x_0)\right)}{f_{Y|X}\left(q_\beta(x_0)|x_0\right) f_X(x_0)} \left\{ \int u^{p+2} K_{v,p}(u)du \right.$$
$$\left.\left. - v \int u^{p+1} K_{v-1,p}(u)du \right\} \right) h^{p-v+2} \{1 + O_P(h)\} \Bigg] \xrightarrow{D} \mathcal{N}(0,1).$$

Note that this result extends the result of [1] who established asymptotic results for the nonparametric local linear βth conditional quantile estimator. Their results are obtained as a special case of Theorem S1.1 by taking $p = 1$.

Remark S1.1. An advantage of our semiparametric approach (even when taking the Laplace density), is that it exploits the asymmetry, which is not done in the existing nonparametric approach(es).

S1.2. Bandwidth selection in nonparametric local polynomial conditional quantile estimation

Recall that here we only have one target function, namely $\theta_1(\cdot) = q_\beta(\cdot)$.

S1.2.1. Theoretical optimal bandwidths

Here we use Theorem S1.1. From this theorem we have asymptotic expressions for bias and variance of the estimator, from which optimal theoretical local and global bandwidths

$$h_{\beta,\text{opt}}^{\text{AMSE}}(x_0) = C_{v,p}(K)\left[\frac{\beta(1-\beta)}{\left[f_{Y|X}\left(q_\beta(x_0)|x_0\right)\right]^2\{q_\beta^{(p+1)}(x_0)\}^2 f_X(x_0)}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}} \tag{S.1}$$

$$h_{\beta,\text{opt}}^{\text{AMISE}} = C_{v,p}(K)\left[\frac{\beta(1-\beta)\int[f_{Y|X}(q_\beta(x)|x)]^{-2}w(x)/f_X(x)dx}{\int\{q_\beta^{(p+1)}(x)\}^2 w(x)dx}\right]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}},$$

are obtained.

These asymptotically optimal bandwidths depend on unknown quantities: the design density $f_X(\cdot)$, the conditional density $f_{Y|X}(\cdot|x)$ and the derivative of the quantile function $q_\beta^{(p+1)}(\cdot)$. We next discuss practical bandwidth selectors for the nonparametric local polynomial quantile estimator $\widehat{q}_\beta(x)$.

S1.2.2. Data-driven bandwidth selection

Following a similar reasoning as for the Quantile-Mean type of bandwidth selector in Section 5.2 we discuss a data-driven bandwidth selection procedure for nonparametric local polynomial conditional quantile estimation. Such a procedure extends that of [2] for local linear quantile regression to local polynomial quantile regression estimation, and is a nice side-product of the asymptotic results provided in Theorem S1.1.

Using (S.1) and (26), for a local bandwidth selector, we find

$$\frac{h_{\beta,\text{opt}}^{\text{AMSE}}(x_0)}{h_{\text{mean}}(x_0)} = \left[\frac{\beta(1-\beta)}{\sigma_*^2(x_0)\left[f_{Y|X}\left(q_\beta(x_0)|x_0\right)\right]^2}\right]^{\frac{1}{2p+3}} \times \left[\frac{m^{(p+1)}(x_0)}{q_\beta^{(p+1)}(x_0)}\right]^{\frac{2}{2p+3}}.$$

As [2] proposed for local linear quantile estimation, we assume that $q_\beta^{(p+1)}(x_0)$ and $m^{(p+1)}(x_0)$ are approximately equal. Therefore, we can write the expression of $h_{\beta,\text{opt}}^{\text{AMSE}}(x_0)$ via data-driven bandwidth selector given by

$$h_{\beta,\text{opt}}^{\text{QM}}(x_0) = h_{\text{mean}}(x_0)\left[\frac{\beta(1-\beta)}{\sigma_*^2(x_0)\left[f_{Y|X}\left(q_\beta(x_0)|x_0\right)\right]^2}\right]^{\frac{1}{2p+3}}, \tag{S.2}$$

in which $h_{\text{mean}}(x_0)$ is to be replaced with a good data-driven bandwidth selector for mean regression. For the special case that the conditional density if $f_{Y|X}(\cdot|x)$ is a member of the quantile-based family of asymmetric densities, i.e., $f_{Y|X}(\cdot|x) = f_{Y|X,\alpha}(\cdot|x)$, this further reduces to

$$h_{\beta,\text{opt}}^{\text{QM}}(x_0) = h_{\text{mean}}(x_0)\left[\frac{\beta(1-\beta)}{4f^2(0)\left\{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2\right\}}\right]^{\frac{1}{2p+3}},$$

for the local bandwidth and

$$h_{\beta,\text{opt}}^{\text{QM}} = h_{\text{mean}} \left[\frac{\beta(1-\beta)}{4f^2(0)\left\{(1-2\alpha)^2(\mu_2-\mu_1^2)+\alpha(1-\alpha)\mu_2\right\}} \right]^{\frac{1}{2p+3}}, \tag{S.3}$$

for the global bandwidth.

In case of a general location-scale family of conditional densities $f_{Y|X}(\cdot|x)$, we can write $\sigma_*(x_0)f_{Y|X}\left(q_\beta(x_0)|x_0\right) = f_S(F_S^{-1}(\beta))$, where f_S and F_S are respectively the cumulative distribution and the density function of $S = \frac{Y-m(X)}{\sigma_*(x_0)}$. Thus (from equation (S.2)) we obtain

$$\begin{aligned} h_{\beta,\text{opt}}^{\text{QM}}(x_0) &= h_{\text{mean}}(x_0) \left[\frac{\beta(1-\beta)}{\left[f_S(F_S^{-1}(\beta))\right]^2} \right]^{\frac{1}{2p+3}} \\ \text{and } h_{\beta,\text{opt}}^{\text{QM}} &= h_{\text{mean}} \left[\frac{\beta(1-\beta)}{\left[f_S(F_S^{-1}(\beta))\right]^2} \right]^{\frac{1}{2p+3}}. \end{aligned}$$

Taking $f_S = \phi$ and $F_S = \Phi$, respectively the standard normal density and distribution function, this leads to

$$h_{\beta,\text{opt}}^{\text{QM}} = h_{\text{mean}} \left[\frac{\beta(1-\beta)}{\left[\phi(\Phi^{-1}(\beta))\right]^2} \right]^{\frac{1}{2p+3}}.$$

When $p = 1$ this corresponds to the data-driven bandwidth selector proposed by [2] for local linear quantile regression.

S2. Proofs of all theoretical results

S2.1. Necessary lemmas

A crucial quantity appearing in Theorem 4.1 is $\mathbb{E}[\boldsymbol{W}^n(x_0)]$. Lemma S2.1 provides an asymptotic expression for $\mathbb{E}[\boldsymbol{W}^n(x_0)]$.

Lemma S2.1. Under the assumptions of Theorem 4.1, the following holds for $\boldsymbol{W}^n(x_0)$

(i). for $0 \leqslant k \leqslant p_r$ (with $r \in \{1,2\}$), an asymptotic mean expression for the kth component of the mean vector $\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ is

$$\begin{aligned} \left(\mathbb{E}[\boldsymbol{W}_r^n(x_0)]\right)_k = \sqrt{nh} f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0)) \Big[& h^{p_r+1}\nu_{p_r+k}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} \\ & + h^{p_r+2}\nu_{p_r+k+1}\xi_{rr}(x_0)\Big] + o(\sqrt{n}h^{\min(p_1,p_2)+5/2}) \end{aligned}$$

$$\text{where} \quad \xi_{rr}(x_0) = \frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!} + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}.$$

(ii). the variance-covariance matrix converges to

$$\mathbb{C}\text{ov}[\boldsymbol{W}^n(x_0)] \to \boldsymbol{\Gamma}_{x_0} \quad \text{as} \quad n \to \infty,$$

where $\boldsymbol{\Gamma}_{x_0}$ is defined as in (19).

(iii). the asymptotic distribution:

$$\sqrt{n}\left(\boldsymbol{W}^n(x_0) - \mathbb{E}[\boldsymbol{W}^n(x_0)]\right) \xrightarrow{D} \mathcal{N}_{p_1+p_2+2}(\mathbf{0}_{p_1+p_2+2}, \boldsymbol{\Gamma}_{x_0}), \quad \text{as} \ \ n \to \infty.$$

Lemma S2.2 provides an asymptotic expression for $\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ and $\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$, which are appearing in the asymptotic distributional result in Theorem 4.1.

Lemma S2.2. Under the assumptions of Theorem 4.1, for $v = 0, 1, ..., p_r$ and $r \in \{1, 2\}$, we have

$$\begin{aligned}\left(\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]\right)_{v+1} \;=\;& (nh^{2p_r+3})^{1/2}\frac{1}{v!}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\int u^{p_r+1}K_{v,p_r}(u)du \\ &+(nh^{2p_r+5})^{1/2}\frac{1}{v!}\Bigg[\int u^{p_r+2}K_{v,p_r}(u)du\Bigg\{\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!} \\ &+\int u^{p_r+2}K_{v,p_r}(u)du\;\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} \\ &\quad \frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\Bigg\}\Bigg] \\ &+o((nh^{2\min(p_1,p_2)+5})^{1/2}),\end{aligned}$$

$$\begin{aligned}&\left(\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]\right)_{v+1} \\ &\quad= (nh^{2p_r+3})^{1/2}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\left\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\right\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))} \\ &\qquad\times\frac{1}{v!}\Big[v\int u^{p_r+1}K_{v-1,p_r}(u)du + \frac{1}{p_r!}\int u^{p_r+1}K_{p_r,p_r}(u)du \\ &\qquad\quad \int u^{p_r+1}K_{v,p_r}(u)du\Big] \\ &\qquad+(nh^{2p_r+5})^{1/2}\frac{\frac{d}{dx_0}\left\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\right\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\xi_{rr}(x_0) \\ &\quad\times\frac{1}{v!}\Big[v\int u^{p_r+2}K_{v-1,p_r}(u)du \\ &\quad+\frac{1}{p_r!}\int u^{p_r+2}K_{p_r,p_r}(u)du\int u^{p_r+2}K_{v,p_r}(u)du\Big] + o((nh^{2p_r+5})^{1/2}).\end{aligned}$$

The proofs of both lemmas can be found in Section S2.2.

S2.2. Proof of Theorem 4.1

The proof is along the same lines as the proof of the main theorem in [3]. We use the following notation:

$$\widetilde{\boldsymbol{\theta}}_r = \sqrt{nh}\begin{pmatrix} \widehat{\theta}_{r0} - \theta_{r0} \\ h(\widehat{\theta}_{r1} - \theta_{r1}) \\ \vdots \\ h^{p_r}(\widehat{\theta}_{rp_r} - \theta_{rp_r}) \end{pmatrix}, \ \boldsymbol{Z}_{ri} = \begin{pmatrix} 1 \\ \frac{X_i - x_0}{h} \\ \vdots \\ \left(\frac{X_i - x_0}{h}\right)^{p_r} \end{pmatrix} \text{ and } \boldsymbol{Z}_r = \begin{pmatrix} 1 \\ \frac{X - x_0}{h} \\ \vdots \\ \left(\frac{X - x_0}{h}\right)^{p_r} \end{pmatrix} \text{ for } r \in \{1, 2\}. \tag{S.4}$$

Then $\sum_{j=0}^{p_r} \widehat{\theta}_{rj}(X_i - x_0)^j = \overline{\theta}_r(X_i, x_0) + a_n \widetilde{\boldsymbol{\theta}}_r^T \boldsymbol{Z}_{ri}$, where $\overline{\theta}_r(X_i, x_0)$ is defined in (22) and $a_n = (nh)^{-1/2}$. As $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^T, \widehat{\boldsymbol{\theta}}_2^T)^T$ maximizes (13), we have that $\widehat{\widetilde{\boldsymbol{\theta}}} = (\widehat{\widetilde{\boldsymbol{\theta}}}_1^T, \widehat{\widetilde{\boldsymbol{\theta}}}_2^T)^T$ maximizes $\sum_{i=1}^n \ell(\overline{\theta}_1(X_i, x_0) + a_n \widetilde{\boldsymbol{\theta}}_1^T \boldsymbol{Z}_{1i}, \overline{\theta}_2(X_i, x_0) + a_n \widetilde{\boldsymbol{\theta}}_2^T \boldsymbol{Z}_{2i}; Y_i) K\{(X_i - x_0)/h\}$ as a function of $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^T, \widetilde{\boldsymbol{\theta}}_2^T)^T$. We study the asymptotic behaviour of $\widehat{\widetilde{\boldsymbol{\theta}}}$ by using the quadratic approximation lemma described in [4] (p. 210) applied to the maximization of the normalized function

$$\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0) = \sum_{i=1}^n \Big[\ell(\overline{\theta}_1(X_i, x_0) + a_n \widetilde{\boldsymbol{\theta}}_1^T \boldsymbol{Z}_{1i}, \overline{\theta}_2(X_i, x_0) + a_n \widetilde{\boldsymbol{\theta}}_2^T \boldsymbol{Z}_{2i}; Y_i)$$
$$- \ell(\overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0); Y_i)\Big] K\{(X_i - x_0)/h\}.$$

Note that $\widehat{\widetilde{\boldsymbol{\theta}}}$ maximizes $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0)$ with respect to $\widetilde{\boldsymbol{\theta}}$. We can easily prove (as in the proof of Theorem 3.3 in [5]) that under Assumption **(A2)** the Hessian matrix of $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0)$ is negative definite, indicating that $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0)$ is concave in $\widetilde{\boldsymbol{\theta}}$. Note that $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0)$ is differentiable on a event with probability 1 (as the derivative of the density with respect to $\theta_1(x_0)$ does not exist when $Y_i = \theta_1(x_0)$, an event with probablity 0).

Using a Taylor approximation of $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0)$ around $(\overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0))$, we find

$$\begin{aligned} \mathcal{L}_n(\widetilde{\boldsymbol{\theta}}; h, x_0) &= a_n \sum_{i=1}^n \sum_{r=1}^2 \psi_r(Y_i; \overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0)) \widetilde{\boldsymbol{\theta}}_r^T \boldsymbol{Z}_{ri} K\{(X_i - x_0)/h\} \\ &+ \frac{a_n^2}{2} \sum_{i=1}^n \sum_{r=1}^2 \sum_{s=1}^2 \psi_{rs}(Y_i; \overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0)) \widetilde{\boldsymbol{\theta}}_r^T \boldsymbol{Z}_{ri} \widetilde{\boldsymbol{\theta}}_s^T \boldsymbol{Z}_{si} K\{(X_i - x_0)/h\} \\ &+ \frac{a_n^3}{6} \sum_{i=1}^n \sum_{r=1}^2 \sum_{s=1}^2 \sum_{t=1}^2 \psi_{rst}(Y_i; \theta_1, \theta_2) \widetilde{\boldsymbol{\theta}}_r^T \boldsymbol{Z}_{ri} \widetilde{\boldsymbol{\theta}}_s^T \boldsymbol{Z}_{si} \widetilde{\boldsymbol{\theta}}_t^T \boldsymbol{Z}_{ti} K\{(X_i - x_0)/h\}, \end{aligned} \tag{S.5}$$

where $\|(\overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0)) - (\theta_1, \theta_2)\| < a_n \|(\widetilde{\boldsymbol{\theta}}_1^T \boldsymbol{Z}_{1i}, \widetilde{\boldsymbol{\theta}}_1^T \boldsymbol{Z}_{1i})\|$ and $\|.\|$ the Euclidean norm. Denote $\boldsymbol{A}_{rs}^n = a_n^2 \sum_{i=1}^n \psi_{rs}(Y_i; \overline{\theta}_1(X_i, x_0), \overline{\theta}_2(X_i, x_0)) \boldsymbol{Z}_{ri} \boldsymbol{Z}_{si}^T K\{(X_i - x_0)/h\}$, then the second term in (S.5) is $\frac{1}{2} \sum_{r=1}^2 \sum_{s=1}^2 \widetilde{\boldsymbol{\theta}}_r^T \boldsymbol{A}_{rs}^n \widetilde{\boldsymbol{\theta}}_s$. Further $(\boldsymbol{A}_{rs}^n)_{kl} = \mathbb{E}(\boldsymbol{A}_{rs}^n)_{kl} + O_P(\mathrm{Var}((\boldsymbol{A}_{rs}^n)_{kl})^{1/2})$ (for $k = 1, \ldots, p_r$ and $l = 1, \ldots, p_s$) and $\mathbb{E}\boldsymbol{A}_{rs}^n = \frac{1}{h}\mathbb{E}(\psi_{rs}(Y; \overline{\theta}_1(X, x_0), \overline{\theta}_2(X, x_0)) \boldsymbol{Z}_r \boldsymbol{Z}_s^T K\{(X - x_0)/h\})$. In a similar way as in the proof of Lemma S2.1, using a Taylor approximation of

$\psi_{rs}(y;\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))$ around $(\theta_1(x_0),\theta_2(x_0))$ and $f_X(x)$ around x_0, we find that

$$
\begin{aligned}
(\mathbb{E}(\boldsymbol{A}^n_{rs}))_{kl} &= -f_X(x_0)\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0))\nu_{k+l-2} - h\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0))\}\nu_{k+l-1} \\
&\quad +o(h).
\end{aligned}
$$

In a similar way, we find that $\mathrm{Var}((\boldsymbol{A}^n_{rs})_{kl}) = O(\frac{1}{nh})$. Therefore, using $\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0)) = 0$ for $r \neq s$ and Assumptions **(A7)** and**(A10)**, $\mathcal{L}_n(\widetilde{\boldsymbol{\theta}};h,x_0) = \boldsymbol{W}^n(x_0)^T\widetilde{\boldsymbol{\theta}} - \frac{1}{2}\widetilde{\boldsymbol{\theta}}^T(\boldsymbol{\Sigma}_{x_0}+h\boldsymbol{\Lambda}_{x_0})\widetilde{\boldsymbol{\theta}}+o_P(h)$. Further, we have that $\mathcal{L}'_n(\widetilde{\boldsymbol{\theta}};h,x_0) = \boldsymbol{W}^n(x_0) - (\boldsymbol{\Sigma}_{x_0}+h\boldsymbol{\Lambda}_{x_0})\widetilde{\boldsymbol{\theta}} + o_P(h)$ and $\mathcal{L}''_n(\widetilde{\boldsymbol{\theta}};h,x_0) = -(\boldsymbol{\Sigma}_{x_0}+h\boldsymbol{\Lambda}_{x_0}) + o_P(h)$, where the derivatives are w.r.t. $\widetilde{\boldsymbol{\theta}}$, and $\mathcal{L}'_n$ is the gradient vector and $\mathcal{L}''_n$ the Hessian matrix. By the quadratic approximation lemma, we have that $\hat{\widetilde{\boldsymbol{\theta}}} = (\boldsymbol{\Sigma}^{-1}_{x_0} - h\boldsymbol{\Sigma}^{-1}_{x_0}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}^{-1}_{x_0})\boldsymbol{W}^n(x_0) + o_P(h)$.

The asymptotic normality result of $\hat{\widetilde{\boldsymbol{\theta}}}$ follows from the asymptotic normality of $\boldsymbol{W}^n(x_0)$, in Lemma S2.1: for $n \to \infty$

$$
\begin{aligned}
&(\boldsymbol{\Sigma}^{-1}_{x_0}\boldsymbol{\Gamma}_{x_0}\boldsymbol{\Sigma}^{-1}_{x_0})^{-1/2}\Big\{\sqrt{nh}\left(\boldsymbol{H}_{p_1}\left(\widehat{\boldsymbol{\theta}}_1(x_0)-\boldsymbol{\theta}_1(x_0)\right), \boldsymbol{H}_{p_2}\left(\widehat{\boldsymbol{\theta}}_2(x_0)-\boldsymbol{\theta}_2(x_0)\right)\right)^T \\
&\qquad - (\boldsymbol{\Sigma}^{-1}_{x_0} - h\boldsymbol{\Sigma}^{-1}_{x_0}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}^{-1}_{x_0})\mathbb{E}[\boldsymbol{W}^n(x_0)]\Big\} \xrightarrow{D} \mathcal{N}_{p_1+p_2+2}(\mathbf{0}_{p_1+p_2+2}, \mathbf{I}_{p_1+p_2+2}).
\end{aligned}
$$

S2.3. Proof of Lemma S2.1

The proof of this lemma is inspired by the proof of Lemma 2 in [3].

(i). From the definition of $\boldsymbol{W}^n(x_0)$ with components $W^n_{rk}(x_0)$ given in (21), we write $\boldsymbol{W}^n(x_0)$ as a sum of independent random vectors:

$$
\boldsymbol{W}^n(x_0) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \begin{pmatrix} \boldsymbol{Y}^*_{1i} \\ \boldsymbol{Y}^*_{2i} \end{pmatrix},
$$

where $\boldsymbol{Y}^*_{ri} = \psi_r(Y_i;\overline{\theta}_1(X_i,x_0),\overline{\theta}_2(X_i,x_0))K\{(X_i-x_0)/h\}\boldsymbol{Z}_{ri}$ for $r \in \{1,2\}$, and $\boldsymbol{Z}_{ri}$ is as in (S.4). Therefore we have that the kth component of $\mathbb{E}\boldsymbol{Y}^*_{ri}$ equals

$$
\begin{aligned}
\mathbb{E}[(\boldsymbol{Y}^*_{ri})_k] &= \int\int \psi_r(y;\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))K\{(x-x_0)/h\}\left(\frac{x-x_0}{h}\right)^{k-1} f_{X,Y}(x,y)dx\,dy \\
&= \int K\{(x-x_0)/h\}\left(\frac{x-x_0}{h}\right)^{k-1} f_X(x)\left\{\int \psi_r(y;\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))f_{Y|X}(y|x)dy\right\}dx \\
&= \int K\{(x-x_0)/h\}\left(\frac{x-x_0}{h}\right)^{k-1} f_X(x)\lambda_r(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))dx \\
&= h\int K(z)z^{k-1}f_X(x_0+hz)\lambda_r(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0))dz.
\end{aligned}
$$

We use a first order Taylor approximation for $\lambda_r(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0))$ around $(\overline{\theta}_1(x_0,x_0),\overline{\theta}_2(x_0,x_0)) = (\theta_1(x_0),\theta_2(x_0))$, a (constant) Taylor approximation of $\lambda_{rs}(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0))$ (for $s \in \{1,2\}$), a Taylor approximation of $\theta_r(x) = \overline{\theta}_r(x,x_0)+$

$\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}(x-x_0)^{p_r+1} + \frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}(x-x_0)^{p_r+2} + o(h^{p_r+2})$, (16) and Assumption **(A7)**:

$$\begin{aligned}\lambda_r(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0)) &= -\sum_{s=1}^{2}\lambda_{rs}(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0)) \\ &\quad \times\left(\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}(hz)^{p_r+1}+\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}(hz)^{p_r+2}\right) \\ &\quad +o(h^{\min(p_1,p_2)+2}).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}[(\boldsymbol{Y}_{ri}^*)_k] &= -\sum_{s=1}^{2}\int K(z)z^{k-1}f_X(x_0+hz)\lambda_{rs}(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0)) \\ &\quad \times\left(\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}h^{p_r+2}z^{p_r+1}+\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}h^{p_r+3}z^{p_r+2}\right)dz+o(h^{\min(p_1,p_2)+3}).\end{aligned}$$

Finally using a first order Taylor expansion for $f_X(x_0+hz)\lambda_{rs}(\overline{\theta}_1(x_0+hz,x_0),\overline{\theta}_2(x_0+hz,x_0))$ around x_0, the fact that $\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0)) = -\lambda_{rs}(\overline{\theta}_1(x_0,x_0),\overline{\theta}_2(x_0,x_0))$, and Assumptions **(A7)**–**(A9)** we find that

$$\begin{aligned}\mathbb{E}[(\boldsymbol{Y}_{ri}^*)_k] &= \sum_{s=1}^{2}\Big(h^{p_r+2}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}f_X(x_0)\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0))\nu_{p_r+k} \\ &\quad +h^{p_r+3}f_X(x_0)\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0))\xi_{rs}(x_0)\nu_{p_r+k+1}\Big)+o(h^{\min(p_1,p_2)+3}),\end{aligned}$$

and since $\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0)) = 0$ for $r \neq s$

$$\begin{aligned}\mathbb{E}[(\boldsymbol{W}_r^n)_k] &= \sqrt{nh}f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\Big(h^{p_r+1}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\nu_{p_r+k} \\ &\quad +h^{p_r+2}\xi_{rr}(x_0)\nu_{p_r+k+1}\Big)+o(\sqrt{n}h^{\min(p_1,p_2)+5/2}).\end{aligned}$$

(ii). The covariance between the kth component of $\boldsymbol{W}_r^n(x_0)$ and the lth component of $\boldsymbol{W}_s^n(x_0)$ (for $k = 1,\ldots,p_r$ and $l = 1,\ldots,p_s$) can be calculated in a similar way, using $\mathbb{E}[(\boldsymbol{W}_r^n)_k] = O(\sqrt{n}h^{p_r+3/2})$, a (constant order) Taylor approximation of

$f_X(x)\psi_r(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))\psi_s(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))$ and Assumptions **(A7)**–**(A9)**:

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\boldsymbol{W}^n_{rk}(x_0)\boldsymbol{W}^n_{sl}(x_0)]\\
&\quad= \mathbb{E}[\boldsymbol{W}^n_{rk}(x_0)\boldsymbol{W}^n_{sl}(x_0)] - \mathbb{E}[\boldsymbol{W}^n_{rk}(x_0)]\mathbb{E}[\boldsymbol{W}^n_{sl}(x_0)]\\
&\quad= \frac{1}{nh}\mathbb{E}\left(\sum_{i=1}^n(\boldsymbol{Y}^*_{ri})_k\sum_{j=1}^n(\boldsymbol{Y}^*_{sj})_l\right) - \frac{1}{nh}\mathbb{E}\left(\sum_{i=1}^n(\boldsymbol{Y}^*_{ri})_k\right)\mathbb{E}\left(\sum_{j=1}^n(\boldsymbol{Y}^*_{sj})_l\right)\\
&\quad= \frac{1}{nh}\mathbb{E}\left(\sum_{i=1}^n(\boldsymbol{Y}^*_{ri})_k(\boldsymbol{Y}^*_{si})_l\right) + O(h^{2\min(p_1,p_2)+3})\\
&\quad= \frac{1}{h}\mathbb{E}\left((\boldsymbol{Y}^*_{ri})_k(\boldsymbol{Y}^*_{si})_l\right) + O(h^{2\min(p_1,p_2)+3})\\
&\quad= \frac{1}{h}\int\int\psi_r(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))\ \psi_s(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))K^2\{(x-x_0)/h\}\\
&\quad\quad\times\left(\frac{x-x_0}{h}\right)^{k+l-2} f_{X,Y}(x,y)dx\ dy + O(h^{2\min(p_1,p_2)+3})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{h}\int\int\psi_r(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))\psi_s(\overline{\theta}_1(x,x_0),\overline{\theta}_2(x,x_0))\\
&\quad\times K^2\{(x-x_0)/h\}\left(\frac{x-x_0}{h}\right)^{k+l-2} f_X(x)f_{Y|X}(y|x)dy\ dx + O(h^{2\min(p_1,p_2)+3})\\
&= f_X(x_0)\mathcal{I}_{rs}(\theta_1(x_0),\theta_2(x_0))\int K^2(z)z^{k+l-2}dz + O(h)\\
&= (\boldsymbol{\Gamma}_{x_0})_{kl} + O(h).
\end{aligned}$$

Therefore $\boldsymbol{\Gamma}_{x_0}^{-1/2}\mathbb{C}\mathrm{ov}[\boldsymbol{W}^n(x_0)] \to \mathbf{I}_{p_1+p_2+2}$, as $n\to\infty$.

(iii). In a similar way as in the proof of Lemma 1 in [3], by using a Cramer-Wold device and by checking Lyapunov's condition, we find that for $n\to\infty$,

$$\boldsymbol{\Gamma}_{x_0}^{-1/2}\sqrt{n}\left(\boldsymbol{W}^n(x_0) - \mathbb{E}[\boldsymbol{W}^n(x_0)]\right) \xrightarrow{D} \mathcal{N}_{p_1+p_2+2}(\mathbf{0}_{p_1+p_2+2},\mathbf{I}_{p_1+p_2+2}).$$

S2.4. Proof of Lemma S2.2

The proof of this lemma is inspired by the proof of Lemma 2 in [3].

From the Fisher information matrix in (17): $\mathcal{I}_{12}(\theta_1(x_0),\theta_2(x_0)) = \mathcal{I}_{21}(\theta_1(x_0),\theta_2(x_0)) = 0$ implies $[\mathcal{I}^{-1}(\theta_1(x_0),\theta_2(x_0))]_{rr} = \mathcal{I}^{-1}_{rr}(\theta_1(x_0),\theta_2(x_0))$ for $r\in\{1,2\}$. Therefore, $\boldsymbol{\Sigma}_{x_0}$, $\boldsymbol{\Lambda}_{x_0}$ and $\boldsymbol{\Gamma}_{x_0}$ are block diagonal matrices. Since, the inverse of a block diagonal matrix is again a block diagonal matrix, we obtain,

$$\boldsymbol{\Sigma}_{x_0}^{-1} = \begin{pmatrix}(\boldsymbol{\Sigma}_{x_0}^{-1})_{11} & 0\\ 0 & (\boldsymbol{\Sigma}_{x_0}^{-1})_{22}\end{pmatrix} \text{ with } (\boldsymbol{\Sigma}_{x_0}^{-1})_{rr} = \frac{\mathcal{I}^{-1}_{rr}(\theta_1(x_0),\theta_2(x_0))}{f_X(x_0)}\boldsymbol{N}^{-1}_{p_rp_r}(x_0),$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1} &= \begin{pmatrix}(\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1})_{11} & 0\\ 0 & (\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1})_{22}\end{pmatrix}\\
\text{with } (\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1})_{rr} &= \frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X^2(x_0)\mathcal{I}^2_{rr}(\theta_1(x_0),\theta_2(x_0))}\boldsymbol{N}^{-1}_{p_rp_r}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}^{-1}_{p_rp_r}(x_0).
\end{aligned}$$

We now want to find an asymptotic expression for the $(v+1)$th $(v=0,1,...,p_r)$ component of $\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ and $\boldsymbol{\Sigma}_{x_0}^{-1}\boldsymbol{\Lambda}_{x_0}\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$. It is noted that $(\boldsymbol{Q}_{p_rp_r}(x_0))_{k,l} = (\boldsymbol{N}_{p_rp_r}(x_0))_{k,l+1}$ $(k=1,..,p_r; l=0,1,...)$. In the proof of Lemma 2 of [3], it has been shown that

$$\begin{aligned}&\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{(v+1),j}\\&\quad=\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{v,j}+\left\{\sum_{k=1}^{p_r+1}\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{(v+1),k}\nu_{p_r+k}\right\}\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{(p_r+1),j}\end{aligned}$$

for $v=1,...,p_r$ and that

$$\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{1,j}=\left\{\sum_{k=1}^{p_r+1}\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{1,k}\nu_{p_r+k}\right\}\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{(p_r+1),j}.$$

Lemma 3 of [3] further states that, for $l=0,1,...,$

$$\int u^{p_r+l+1}K_{v,p_r}(u)du = v!\sum_{j=1}^{p_r+1}\{\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\}_{v+1,j}\nu_{p_r+j+l}. \tag{S.6}$$

Using (S.6), we find that, for $v=0,...,p_r$,

$$\begin{aligned}&\sum_{j=1}^{p_r+1}\left(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\right)_{(v+1),j}\nu_{p_r+j}\\&\quad=\frac{1}{(v-1)!}\int u^{p_r+1}K_{v-1,p_r}(u)du+\frac{1}{v!p_r!}\int u^{p_r+1}K_{p_r,p_r}(u)du\int u^{p_r+1}K_{v,p_r}(u)du.\end{aligned}$$

By using Lemma S2.1, we can easily find an asymptotic expression of the $(v+1)$th component of $\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ (for $v=0,1,...,p_r$; $r\in\{1,2\}$), which is

$$\begin{aligned}&\left(\boldsymbol{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]\right)_{v+1}\\&\quad=\frac{\mathcal{I}_{rr}^{-1}(\theta_1(x_0),\theta_2(x_0))}{f_X(x_0)}\sum_{k=1}^{p_r+1}\left(\boldsymbol{N}_{p_r,p_r}^{-1}(x_0)\right)_{(v+1),k}(\mathbb{E}[\boldsymbol{W}_r^n(x_0)])_k\\&\quad=\sqrt{nh}\sum_{k=1}^{p_r+1}\left(\boldsymbol{N}_{p_r,p_r}^{-1}(x_0)\right)_{(v+1),k}\left[h^{p_r+1}\nu_{p_r+k}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}+h^{p_r+2}\nu_{p_r+k+1}\xi_{rr}(x_0)\right]\\&\qquad+o((nh^{2\min(p_1,p_2)+5})^{1/2})\\&\quad=\sqrt{nh}\Bigg[h^{p_r+1}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{1}{v!}\int u^{p_r+1}K_{v,p_r}(u)du+h^{p_r+2}\frac{1}{v!}\int u^{p_r+2}K_{v,p_r}(u)du\Bigg\{\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}\\&\qquad+\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\Bigg\}\Bigg]+o((nh^{2\min(p_1,p_2)+5})^{1/2}\end{aligned}$$

$$= (nh^{2p_r+3})^{1/2}\frac{1}{v!}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\int u^{p_r+1}K_{v,p_r}(u)du + (nh^{2p_r+5})^{1/2}\frac{1}{v!}\Bigg[\int u^{p_r+2}K_{v,p_r}(u)du\Bigg\{\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}$$

$$+\int u^{p_r+2}K_{v,p_r}(u)du\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\Bigg\}\Bigg] + o((nh^{2\min(p_1,p_2)+5})^{1/2}.$$

Similarly, the asymptotic expression of the $(v+1)$th component of $\mathbf{\Sigma}_{x_0}^{-1}\mathbf{\Lambda}_{x_0}\mathbf{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]$ (for $v=0,1,...,p_r$; $r\in\{1,2\}$) can be written as

$$\Big(\mathbf{\Sigma}_{x_0}^{-1}\mathbf{\Lambda}_{x_0}\mathbf{\Sigma}_{x_0}^{-1}\mathbb{E}[\boldsymbol{W}_r^n(x_0)]\Big)_{v+1}$$
$$= \frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X^2(x_0)\mathcal{I}_{rr}^2(\theta_1(x_0),\theta_2(x_0))}\sum_{k=1}^{p_r+1}\big(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\big)_{(v+1),k}(\mathbb{E}[\boldsymbol{W}_r^n(x_0)])_k$$

$$= \sqrt{nh}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\sum_{k=1}^{p_r+1}\big(\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{Q}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\big)_{(v+1),k}$$
$$\times\Big[h^{p_r+1}\nu_{p_r+k}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!} + h^{p_r+2}\nu_{p_r+k+1}\xi_{rr}(x_0)\Big] + o((nh^{2p_r+5})^{1/2})$$
$$= (nh^{2p_r+3})^{1/2}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}$$
$$\times\frac{1}{v!}\Big[v\int u^{p_r+1}K_{v-1,p_r}(u)du + \frac{1}{p_r!}\int u^{p_r+1}K_{p_r,p_r}(u)du\int u^{p_r+1}K_{v,p_r}(u)du\Big]$$
$$+ (nh^{2p_r+5})^{1/2}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\xi_{rr}(x_0)$$
$$\times\frac{1}{v!}\Big[v\int u^{p_r+2}K_{v-1,p_r}(u)du + \frac{1}{p_r!}\int u^{p_r+2}K_{p_r,p_r}(u)du\int u^{p_r+2}K_{v,p_r}(u)du\Big]$$
$$+ o((nh^{2p_r+5})^{1/2}).$$

S2.5. Proof of Theorem 4.2

S2.5.1. Preliminaries

Before proving the asymptotic results more specifically for each $\widehat{\theta}_{rv}(x_0)$, we discuss the asymptotic bias in the foregoing general results. From Theorem 4.1, Lemmas S2.1 and S2.2 we have that the $(v+1)$th component of the asymptotic bias ($v=0,1,...,p_r$) is

$$\Big(\text{ABias}\Big\{\sqrt{nh}\,\Big(\boldsymbol{H}_{p_r}\big(\widehat{\boldsymbol{\theta}}_r(x_0)-\boldsymbol{\theta}_r(x_0)\big)\Big)\Big\}\Big)_{v+1}$$
$$= (nh^{2p_r+3})^{1/2}\frac{1}{v!}a_{1,v} + (nh^{2p_r+5})^{1/2}\frac{1}{v!}a_{2,v} + o((nh^{2p_r+5})^{1/2}),$$

$$
\begin{aligned}
\text{where } a_{1,v} &= \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\int u^{p_r+1}K_{v,p_r}(u)du \\
a_{2,v} &= \frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}\int u^{p_r+2}K_{v,p_r}(u)du \\
&\quad + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))} \\
&\quad \times \left\{\int u^{p_r+2}K_{v,p_r}(u)du \quad -v\int u^{p_r+1}K_{v-1,p_r}(u)du \right. \\
&\quad \left. -\frac{1}{p_r!}\int u^{p_r+1}K_{p_r,p_r}(u)du\int u^{p_r+1}K_{v,p_r}(u)du\right\}.
\end{aligned}
$$

S2.5.2. Proof of 4.2

Theorem 4.2 follows from Theorem 4.1, as the marginal distributions of the components of $\hat{\hat{\boldsymbol{\theta}}}$. The asymptotic covariance of $\hat{\hat{\boldsymbol{\theta}}}_r$ (for $r \in \{1,2\}$) can be written as

$$
\text{ACov}(\hat{\hat{\boldsymbol{\theta}}}_r) = \frac{1}{f_X(x_0)}\mathcal{I}_{rr}^{-1}(\theta_1(x_0),\theta_2(x_0))\boldsymbol{N}_{p_rp_r}^{-1}(x_0)\boldsymbol{T}_{p_rp_r}(x_0)\boldsymbol{N}_{p_rp_r}^{-1}(x_0).
$$

The asymptotic variance of the ML estimator $\hat{\theta}_r^{(v)}(x_0) = v!\hat{\theta}_{rv}(x_0)$ (for $v = 0,1,...,p_r; r \in \{1,2\}$) is $(v!)^2\text{AVar}(\hat{\theta}_{rv}(x_0))$. Note that $\text{AVar}(\hat{\theta}_{rv}(x_0))$ is the $(v+1,v+1)$ entry of the matrix $n^{-1}h^{-(2v+1)}\text{ACov}(\hat{\hat{\boldsymbol{\theta}}}_r)$. The $(v+1,v+1)$ entry of $\text{ACov}(\hat{\hat{\boldsymbol{\theta}}}_r)$ is

$$
\begin{aligned}
\left[\text{ACov}(\hat{\hat{\boldsymbol{\theta}}}_r)\right]_{v+1,v+1} &= \frac{1}{f_X(x_0)}\mathcal{I}_{rr}^{-1}(\theta_1(x_0),\theta_2(x_0))\sum_{k=1}^{p_r+1}\sum_{l=1}^{p_r+1}\frac{c_{v+1,k}c_{v+1,l}}{|\boldsymbol{N}_{p_rp_r}(x_0)|^2}\{\boldsymbol{T}_{p_rp_r}(x_0)\}_{k,l} \\
&= \frac{1}{f_X(x_0)}\mathcal{I}_{rr}^{-1}(\theta_1(x_0),\theta_2(x_0))\frac{1}{(v!)^2}\int K_{v,p_r}^2(u)du,
\end{aligned}
$$

where $c_{i,j}$ is the cofactor of $\{\boldsymbol{N}_{p_rp_r}(x_0)\}_{i,j}$. Hence

$$
\text{AVar}(\hat{\theta}_r^{(v)}(x_0)) = (v!)^2\text{AVar}(\hat{\theta}_{rv}(x_0)) = \frac{\mathcal{I}_{rr}^{-1}(\theta_1(x_0),\theta_2(x_0))}{nh^{(2v+1)}f_X(x_0)}\int K_{v,p_r}^2(u)du.
$$

The asymptotic bias of $\hat{\hat{\theta}}_{rv}(x_0)$ is given in (S.7). Note that for $p_r \geqslant v$, and $p_r - v$ even, we have from Lemma 4 in [3] that $\int u^{p_r+1}K_{v,p_r}(u)du = 0$. In that case, $a_{1,v} = 0$ and

$$
\begin{aligned}
a_{2,v} = \frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}\int u^{p_r+2}K_{v,p_r}(u)du + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))} \\
\times\left\{\int u^{p_r+2}K_{v,p_r}(u)du - v\int u^{p_r+1}K_{v-1,p_r}(u)du\right\}.
\end{aligned}
$$

Hence, the expressions of asymptotic bias and asymptotic variance of $\hat{\theta}_r^{(v)}(x_0;p_r,h)$ follow.

S2.6. Proof of Theorem 4.3

Theorem 4.1 gives the asymptotic joint multivariate normal distribution of $(\widehat{\boldsymbol{\theta}}_1(x_0), \widehat{\boldsymbol{\theta}}_2(x_0))$. By applying the multivariate delta method to $g(\widehat{\theta}_1(x_0), \widehat{\theta}_2(x_0)) = \widehat{q}_\beta(x_0)$ with $g(u, v) = u + C_\alpha(\beta) \cdot e^v$ and $\nabla g(u, v) = (1, C_\alpha(\beta) \cdot e^v)$, we find that, as $n \to \infty$,

$$\sqrt{nh\sigma_q^2}\left\{g\left(\widehat{\theta}_1(x_0), \widehat{\theta}_2(x_0)\right) - g\left(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0)]\right)\right\} \xrightarrow{D} \mathcal{N}(0, 1),$$

where $\sigma_q^2 = \nabla g(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0))\boldsymbol{A}\nabla g(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0)])^T$ and the asymptotic covariance matrix

$$\boldsymbol{A} = \begin{pmatrix} \text{AVar}[\widehat{\theta}_1(x_0)] & \text{ACov}\left(\widehat{\theta}_1(x_0), \widehat{\theta}_2(x_0)\right) \\ \text{ACov}\left(\widehat{\theta}_1(x_0), \widehat{\theta}_2(x_0)\right) & \text{AVar}[\widehat{\theta}_2(x_0)] \end{pmatrix} = \begin{pmatrix} \text{AVar}[\widehat{\theta}_1(x_0)] & 0 \\ 0 & \text{AVar}[\widehat{\theta}_2(x_0)] \end{pmatrix}.$$

Note that $\mathbb{E}[\widehat{\theta}_r(x_0)] = \theta_r(x_0) + \text{Bias}[\widehat{\theta}_r(x_0)]$ for $r \in \{1, 2\}$. From Theorem 4.2, we have that the asymptotic bias of $\widehat{\theta}_r(x_0)$ is

$$\text{ABias}[\widehat{\theta}_r(x_0)] = \begin{cases} h^{p_r+1}\frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\left\{\int u^{p_r+1}K_{0,p_r}(u)du\right\}(1 + O(h)) & \text{if} \quad p_r \text{ is odd} \\ \left[\frac{\theta_r^{(p_r+2)}(x_0)}{(p_r+2)!}\int u^{p_r+2}K_{0,p_r}(u)du + \frac{\theta_r^{(p_r+1)}(x_0)}{(p_r+1)!}\frac{\frac{d}{dx_0}\left\{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))\right\}}{f_X(x_0)\mathcal{I}_{rr}(\theta_1(x_0),\theta_2(x_0))}\right. & \\ \qquad \left.\times\left\{\int u^{p_r+2}K_{0,p_r}(u)du\right\}\right]h^{p_r+2}(1 + O(h)) & \text{if} \quad p_r \text{ is even,} \end{cases}$$

and the asymptotic variance of $\widehat{\theta}_r(x_0)$ is

$$\text{AVar}[\widehat{\theta}_r(x_0)] = \frac{\mathcal{I}_{rr}^{-1}(\theta_1(x_0), \theta_2(x_0))}{hf_X(x_0)}\int K_{0,p_r}^2(u)du.$$

Therefore, we have by Assumption **(A10)**

$$\begin{aligned} & g\left(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0)]\right) \\ & = \mathbb{E}[\widehat{\theta}_1(x_0)] + C_\alpha(\beta) \cdot e^{\mathbb{E}[\widehat{\theta}_2(x_0)]} \\ & = \theta_1(x_0) + \text{Bias}[\widehat{\theta}_1(x_0)] + C_\alpha(\beta) \cdot e^{\theta_2(x_0) + \text{Bias}[\widehat{\theta}_2(x_0)]} \\ & = \theta_1(x_0) + \text{Bias}[\widehat{\theta}_1(x_0)] + C_\alpha(\beta) \cdot e^{\theta_2(x_0)}\left\{1 + \sum_{k=1}^{\infty}\frac{\left(\text{Bias}[\widehat{\theta}_2(x_0)]\right)^k}{k!}\right\} \\ & = q_\beta(x_0) + \text{ABias}[\widehat{\theta}_1(x_0)] + C_\alpha(\beta) \cdot e^{\theta_2(x_0)}\text{ABias}[\widehat{\theta}_2(x_0)] \\ & \quad + o_P(\text{ABias}[\widehat{\theta}_1(x_0)] + \text{ABias}[\widehat{\theta}_2(x_0)]), \end{aligned}$$

$$
\begin{aligned}
&\nabla g(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0)) \boldsymbol{A} \nabla g(\mathbb{E}[\widehat{\theta}_1(x_0)], \mathbb{E}[\widehat{\theta}_2(x_0)])^T \\
&\quad = (1, C_\alpha(\beta) \cdot e^{\mathbb{E}[\widehat{\theta}_2(x_0)]}) \begin{pmatrix} \mathrm{AVar}[\widehat{\theta}_1(x_0)] & 0 \\ 0 & \mathrm{AVar}[\widehat{\theta}_2(x_0)] \end{pmatrix} \begin{pmatrix} 1 \\ C_\alpha(\beta) \cdot e^{\mathbb{E}[\widehat{\theta}_2(x_0)]} \end{pmatrix} \\
&\quad = \mathrm{AVar}[\widehat{\theta}_1(x_0)] + (C_\alpha(\beta))^2 \cdot \mathrm{AVar}[\widehat{\theta}_2(x_0)] \cdot e^{2\mathbb{E}[\widehat{\theta}_2(x_0)]} \\
&\quad = \mathrm{AVar}[\widehat{\theta}_1(x_0)] + (C_\alpha(\beta))^2 \cdot \mathrm{AVar}[\widehat{\theta}_2(x_0)] \cdot e^{2\theta_2(x_0)} e^{2\mathrm{Bias}[\widehat{\theta}_2(x_0)]} \\
&\quad = \mathrm{AVar}[\widehat{\theta}_1(x_0)] + (C_\alpha(\beta))^2 \cdot \mathrm{AVar}[\widehat{\theta}_2(x_0)] e^{2\theta_2(x_0)} + o(1) \\
&\quad = \frac{\mathcal{I}_{11}^{-1}(\theta_1(x_0), \theta_2(x_0))}{h f_X(x_0)} \int K_{0,p_1}^2(u) du + (C_\alpha(\beta))^2 e^{2\theta_2(x_0)} \frac{\mathcal{I}_{22}^{-1}(\theta_1(x_0), \theta_2(x_0))}{h f_X(x_0)} \int K_{0,p_2}^2(u) du + o(1).
\end{aligned}
$$

S3. Simulation study: additional results

S3.1. Performance of data-driven bandwidth selectors in Section 5.2

Note that for a symmetric Laplace as well as for a symmetric normal density the quantity γ_1 is 0.5. Consequently the expression for the theoretical optimal global bandwidth (25) is the same in both cases. From (30) the expressions for the Quantile-Mean based bandwidth selector for the AlaD and AND log-likelihood estimator of θ_1 are

$$
h_{\alpha,\mathrm{ALaD}}^{\mathrm{QM}} = h_{\mathrm{mean}} \left[\frac{\alpha(1-\alpha)}{\{(1-2\alpha)^2 + 2\alpha(1-\alpha)\}} \right]^{\frac{1}{2p_1+3}} \quad \text{for ALaD,} \tag{S.7}
$$

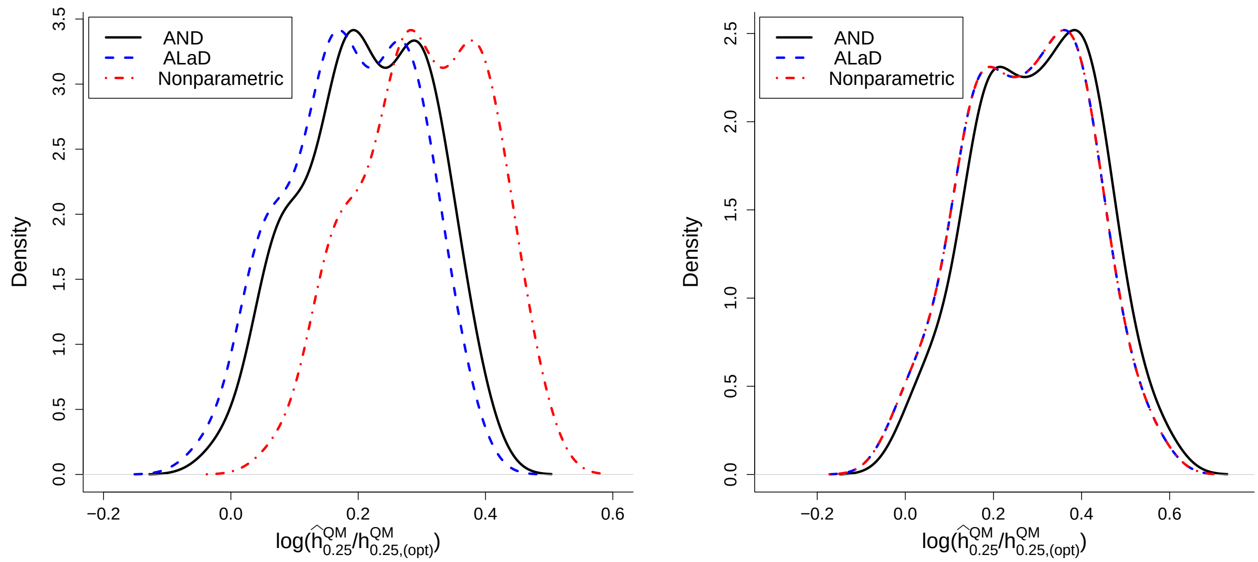
$$
\text{and} \quad h_{\alpha,\mathrm{AND}}^{\mathrm{QM}} = h_{\mathrm{mean}} \left[\frac{\alpha(1-\alpha)\pi}{\{(1-2\alpha)^2(\pi-2) + \alpha(1-\alpha)\pi\}} \right]^{\frac{1}{2p_1+3}} \quad \text{for AND.} \tag{S.8}
$$

It is important to mention that the bandwidth selector for the nonparametric approach, $h_{\beta,\mathrm{NP}}^{\mathrm{QM}}$ in (S.3), and $h_{\alpha,\mathrm{ALaD}}^{\mathrm{QM}}$ are identical for $\beta = \alpha$ due to Remark 3.1. Recall also that in case $\beta = \alpha$ the point estimates of the semiparametric conditional quantile function under the ALaD (estimation) model and the nonparameric quantile estimator are equal.

Table S.1: True global bandwidth and its average estimated values via data-driven bandwidth selectors.

Error distribution	$\theta_1(x)$	$h_{0.25,\mathrm{opt}}^{\mathrm{QM}}$	$\widehat{h}_{0.25,\mathrm{AND}}^{\mathrm{QM}}$	$\widehat{h}_{0.25,\mathrm{ALaD}}^{\mathrm{QM}}$	$\widehat{h}_{0.25,\mathrm{NP}}^{\mathrm{QM}}$
AND	$x + 2e^{-16x^2}$	0.0758	0.0923	0.0901	0.1010
	$\sin(x^2) + x + 2e^{-16x^2}$	0.0753	0.0933	0.0912	0.1021
ALaD	$x + 2e^{-16x^2}$	0.0758	0.1013	0.0990	0.0990
	$\sin(x^2) + x + 2e^{-16x^2}$	0.0753	0.1023	0.0999	0.0999

For each simulation example, the values of theoretical optimal bandwidths (S.7) and (S.8) are listed, together with the average of the estimated bandwidths over the 100 simulations, in



(a) (b)

Figure S.1: Example 2. Kernel density estimates of $\log(\widehat{h}^{\mathrm{QM}}_{0.25}/h^{\mathrm{QM}}_{0.25,\mathrm{opt}})$ obtained by data-driven selector under the true (a) AND; (b) ALaD likelihood.

Table S.1. It is clearly seen that the estimated bandwidths $\widehat{h}^{\mathrm{QM}}_{0.25,\ \mathrm{ALaD}}$ and $\widehat{h}^{\mathrm{QM}}_{0.25,\ \mathrm{NP}}$ are equal under the true asymmetric Laplace error distribution for $\alpha = \beta$, which is as expected. In case of a true asymmetric normal error distribution, $\widehat{h}^{\mathrm{QM}}_{0.25,\ \mathrm{NP}}$ is larger than $\widehat{h}^{\mathrm{QM}}_{0.25,\ \mathrm{AND}}$ and $\widehat{h}^{\mathrm{QM}}_{0.25,\ \mathrm{ALaD}}$ in each example, although the values are close. In Figures S.1(a) and (b) kernel density estimates of $\log(\widehat{h}^{\mathrm{QM}}_{0.25}/h^{\mathrm{QM}}_{0.25,\mathrm{opt}})$ obtained by using the three methods (two semiparametric estimation models, and nonparametric approach) for the simulation model with $\theta_1(x) = \sin(x^2)+x+2e^{-16x^2}$, under asymmetric normal and asymmetric Laplace error distributions are presented. The rough data-driven bandwidths tend to be often larger than the optimal bandwidth, but seem to lead to good quality estimators.

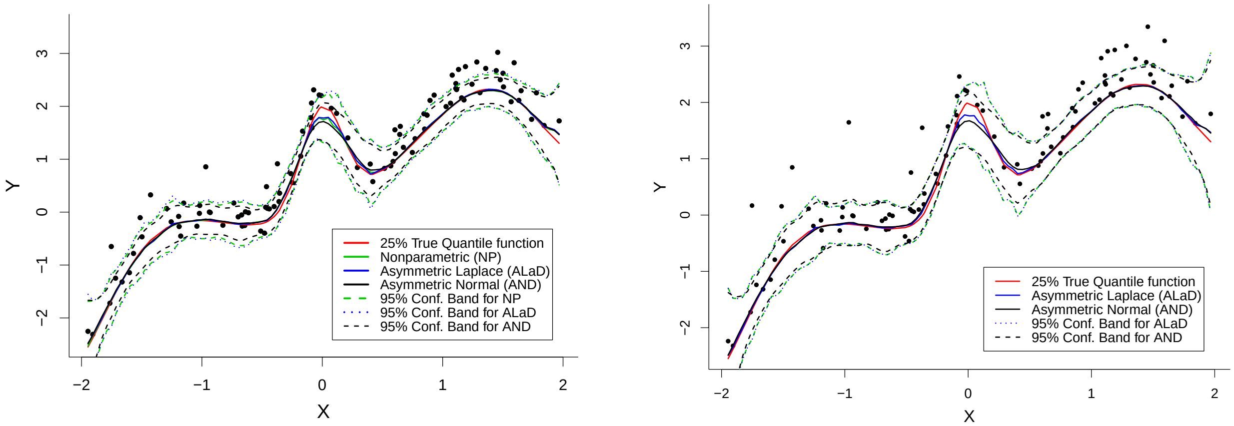
S3.2. Further simulation results on Example 2

The estimated 25% quantile curve (mean of 100 estimated curves based on 100 simulated samples) with its Bonferroni-type 95% confidence bands are presented in Figure S.2 (a) for AND likelihood and in Figure S.2(b) for ALaD likelihood. Note that the confidence band of the estimated curves is narrower compared to others when indeed working with the AND (simulation) model.

S3.3. Effect of estimating α: additional results

S3.3.1. Simulation results for other bandwidth values

Table S.2 complements Table 2 and presents results for bandwidths $h = 0.075$ and $h = 0.095$.



(a) (b)

Figure S.2: Example 2. 95% Bonferroni-type confidence bands, using semiparametric ALaD and AND conditional densities and nonpametric approach together with estimated 25% conditional quantile functions under (a) a AND model (true); (b) a ALaD model (true).

Table S.2: Examples 1 and 2 with Asymmetric Normal Error: Mean (standard error) of AISE values based on 100 simulations for $\beta = (0.25, 0.90)$. Using three different methods to estimate $\alpha : \widehat{\alpha}^{(j)}$ (Method j), $j = 1, 2, 3$.

Bandwidth $h = 0.075$					
		Example 1		Example 2	
Method	β	0.25	0.9	0.25	0.9
1	AND	1.8982(0.7553)	2.6641(0.6830)	1.7065(0.7310)	1.7862 (0.8501)
	ALaD	3.6166(1.8479)	5.7934(2.6830)	3.6535(1.8623)	6.2259(2.8831)
2	AND	2.3445(1.0657)	3.4019(0.9339)	2.1485(0.7719)	3.4019(1.7426)
	ALaD	3.5563(1.8817)	10.0399(2.8981)	4.2930(2.1692)	10.0399(6.1337)
3	AND	2.0001(0.7864)	3.1516(0.7959)	2.0075(0.7586)	2.9696(0.9339)
	ALaD	3.4065(1.7938)	5.7690(2.7700)	3.6297(1.8144)	6.4370(2.8981)
Bandwidth $h = 0.095$					
		Example 1		Example 2	
Method	β	0.25	0.9	0.25	0.9
1	AND	2.0165(0.7479)	2.9542(2.6578)	1.7434(0.7164)	2.0529(0.9416)
	ALaD	2.7566(1.4088)	5.0948(3.4930)	2.7721(1.4216)	5.1631(2.6049)
2	AND	2.4597(1.0521)	7.8926(4.0313)	2.0371(0.7597)	3.3829(1.6499)
	ALaD	2.9606(1.7902)	6.0273(5.8069)	3.4222(1.7765)	9.5293(5.8021)
3	AND	1.9322(0.7631)	3.1009(3.5245)	1.9452(0.7307)	2.9364(0.9318)
	ALaD	2.8046(1.6488)	5.7454(4.9430)	2.8625(1.4908)	5.8082(2.8680)

S3.3.2. Effect of estimating α: computational costs

We recorded the computation time for the semiparametric method, when using the different methods for estimating α. The average computation time (in seconds) of the resulting three semiparametric methods are presented in Table S.3. It is observed that the average computation times when using Methods 2 and 3 are very close. When using Method 1 for estimating α, calculations take a factor of about 4 to 5 times longer than for the two other methods.

Table S.3: Average computation time (in seconds) when using three different methods to estimate α : $\widehat{\alpha}^{(j)}$ (Method j), $j = 1, 2, 3$ for the sample sizes $n = 100$.

Bandwidth $h = 0.075$		Example 1		Example 2	
Method	β	0.25	0.9	0.25	0.9
1	AND	1.2236	1.0777	1.2584	2.03.54
	ALaD	2.3603	2.3182	2.3578	2.5937
2	AND	0.2586	0.2580	0.2722	0.2792
	ALaD	0.4670	0.4747	0.4871	0.4969
3	AND	0.4253	0.4857	0.9554	0.9965
	ALaD	0.4825	0.5455	0.4208	0.4720

Bandwidth $h = 0.090$		Example 1		Example 2	
Method	β	0.25	0.90	0.25	0.90
1	AND	1.1091	1.3036	1.1884	1.2661
	ALaD	2.4913	2.6766	2.7797	2.6444
2	AND	0.2672	0.2831	0.2498	0.2729
	ALaD	0.4607	0.5258	0.4716	0.5415
3	AND	0.2745	0.2915	0.2646	0.3060
	ALaD	0.4228	0.5215	0.4175	0.5502

Bandwidth $h = 0.055$		Example 1		Example 2	
Method	β	0.25	0.9	0.25	0.9
1	AND	1.2256	1.1914	1.1658	1.9713
	ALaD	2.3838	4.4875	2.2769	2.1900
2	AND	0.2708	0.2663	0.2617	0.2597
	ALaD	0.4787	0.5516	0.4604	0.4607
3	AND	0.3416	0.3451	0.3716	0.4616
	ALaD	0.4489	0.4555	0.3226	0.3941

References

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

[2] K. Yu, M. Jones, Local linear quantile regression, Journal of the American statistical Association 93 (441) (1998) 228–237.

[3] J. Fan, N. E. Heckman, M. P. Wand, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions, Journal of the American Statistical Association 90 (429) (1995) 141–150.

[4] J. Fan, I. Gijbels, Local Polynomial Modelling and Its applications, CRC Press, 1996.

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