Bernoulli **21**(3), 2015, 1435–1466
DOI: 10.3150/14-BEJ610

Local bilinear multiple-output quantile/depth regression

MARC HALLIN[1,2,*], ZUDI LU[3], DAVY PAINDAVEINE[1,**] and MIROSLAV ŠIMAN[4]

[1]*E.C.A.R.E.S., Université libre de Bruxelles CP114/04, 50 Avenue F.D. Roosevelt, B-1050 Brussels, Belgium. E-mail:* [*]*mhallin@ulb.ac.be;* [**]*dpaindav@ulb.ac.be; url: http://homepages.ulb.ac.be/~dpaindav*
[2]*ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA*
[3]*Mathematical Sciences & Statistical Sciences Research Institute, University of Southampton, Building 54, Highfield, Southampton SO17 1BJ, UK. E-mail: zudilu@gmail.com;*
url: https://sites.google.com/site/zudiluwebsite/
[4]*Department of Stochastic Informatics, Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, CZ-182 08 Prague 8, Czech Republic. E-mail: siman@utia.cas.cz*

A new quantile regression concept, based on a directional version of Koenker and Bassett's traditional single-output one, has been introduced in [*Ann. Statist.* (2010) **38** 635–669] for multiple-output location/linear regression problems. The polyhedral contours provided by the empirical counterpart of that concept, however, cannot adapt to unknown nonlinear and/or heteroskedastic dependencies. This paper therefore introduces local constant and local linear (actually, bilinear) versions of those contours, which both allow to asymptotically recover the conditional halfspace depth contours that completely characterize the response's conditional distributions. Bahadur representation and asymptotic normality results are established. Illustrations are provided both on simulated and real data.

Keywords: conditional depth; growth chart; halfspace depth; local bilinear regression; multivariate quantile; quantile regression; regression depth

1. Introduction

1.1. Quantile/depth contours: From multivariate location to multiple-output regression

A multiple-output extension of Koenker and Bassett's celebrated concept of regression quantiles was recently proposed in Hallin, Paindaveine, and Šiman [18] (hereafter HPŠ). That extension provides regions that are enjoying, at population level, a double interpretation in terms of quantile and halfspace depth regions. In the empirical case, those regions are limited by polyhedral contours which can be computed via parametric linear programming techniques.

Those results establish a strong and quite fruitful link between two seemingly unrelated statistical worlds – on one hand the typically one-dimensional concept of quantiles, deeply rooted into the strong ordering features of the real line and L_1 optimality, with linear programming algorithms, and traditional central-limit asymptotics; the intrinsically multivariate concept of depth on the other hand, with geometric characterizations, computationally intensive combinatorial algorithms, and nonstandard asymptotics. From their relation to depth, quantile hyperplanes and regions inherit a variety of geometric properties – connectedness, nestedness, convexity, affine-equivariance ... while, via its relation to quantiles, depth accedes to L_1 optimality, feasible linear programming algorithms, and tractable asymptotics.

The HPŠ approach, however, is focused on the case of i.i.d. m-variate observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, and the quantile/depth contours they propose provide a consistent reconstruction of the corresponding population contours in $\mathbb{R}^m$ – call them *unconditional* or *location* contours. In the presence of covariates $\mathbf{X}_1, \ldots, \mathbf{X}_n$, with $\mathbf{X}_i = (1, \mathbf{W}_i')'$, the objective of the statistical analysis is a study of the influence of the covariate(s) $\mathbf{W}$ on the response $\mathbf{Y}$, that is, a study of the distribution of $\mathbf{Y}$ conditional on $\mathbf{W}$. The contours of interest, thus, are the collection of the population *conditional quantile/depth contours* of $\mathbf{Y}$, indexed by the values $\mathbf{w} \in \mathbb{R}^{p-1}$ of $\mathbf{W}$ – that is, the collection of location ($p = 1$) quantile/depth contours associated with the conditional (on $\mathbf{W} = \mathbf{w}$) distributions of $\mathbf{Y}$.

An apparently simple solution would consist in introducing the covariate values $\mathbf{w}$ into the linear equations that characterize (via the minimization of an L_1 criterion) the HPŠ contours. The resulting regions and contours, unfortunately, in general carry little information about conditional distributions, and rather produce some averaged (over the covariate space) quantile/depth contours – the only exception being the overly restrictive case of a linear regression relation between the response and the covariates, under which, for some $\mathfrak{b} \in \mathbb{R}^p$, the distribution of $\mathbf{Y} - (\mathbf{1}, \mathbf{w}')\mathfrak{b}$ conditional on $\mathbf{W} = \mathbf{w}$ does not depend on $\mathbf{w} \in \mathbb{R}^{p-1}$.

This problem is not specific to the multiple-output context and, in the traditional single-output setting, it has motivated weighted, local polynomial and nearest-neighbor versions of quantile regression, among others. We refer to [43–45] for conceptual insight and practical information, to [4,9,17,19,28,47] for some recent asymptotic results, and to [2,6,14,16,22–24,38] for some less recent ones.

Our objective in this paper is to extend those local estimation ideas to the HPŠ concept of multiple-output regression quantiles. Since local constant and local linear methods have been shown to perform extremely well in the single-output single-regressor case (Yu and Jones [44]), we will concentrate on local constant and local *bilinear* approaches – in the multiple-output context, indeed, it turns out that the adequate extensions of locally linear procedures are of a *bilinear* nature. Just as in the single-output case, the local methods we propose in this paper do not require any a priori knowledge of any trend and – see [30] for details – asymptotically characterize the conditional distributions of $\mathbf{Y}$ given $\mathbf{W} = \mathbf{w}$ for any $\mathbf{w} \in \mathbb{R}^{p-1}$. The final result is thus much more informative on the dependence of $\mathbf{Y}$ on the covariates than any standard linear or local polynomial mean regression.

It should be clear, however, that our methods, as well as other local nonparametric methods, do not escape the curse of dimensionality, and will run into problems in the presence of high-dimensional regressors. It follows indeed from the asymptotic results of Section 5 and, more particularly, from the rates in Theorem 5.2, that consistency rates are affected by p but not by m.

Growth chart applications (with $(p - 1) = 1$) do not suffer this drawback, as only univariate kernels are involved. Growth charts (reference curves, percentile curves) have been used for a long time by practitioners in order to assess the impact of regressors on the quantiles of some given univariate variable of interest, and several methods have been developed (see, e.g., [3,8, 40,42], and the references therein), including single-response quantile regression (see [15,41]). Much less results are available in the multiple-output case, with a recent proposal by Wei [39], who defines a new concept of dynamic multiple-output regression contours generalizing single-output proposals by [4], [25] and [40]. These contours, however, do not have the nature and interpretation of (conditional) depth contours. They enjoy interesting conditional coverage probability properties (without any "minimal volume" or "maximal accuracy" features, though) but rely on a sequential conditioning of response components, and crucially depend on the order adopted for that conditioning. Their empirical versions are equivariant under marginal location-scale transformations of the response, but they are neither affine- nor rotation-equivariant. Our methodology, which is based on entirely different principles, appears as a natural alternative (see [32] for a real-data example of bivariate growth charts based on the methods we are describing here), yielding affine-equivariant regression contours with well-accepted conditional depth interpretation; moreover, we provide consistency and asymptotic distributional results.

1.2. Motivating examples

As a motivating example, we generated $n = 999$ points from the model

$$(Y_1, Y_2) = \left(W, W^2\right) + \left(1 + \frac{3}{2}\left(\sin\left(\frac{\pi}{2}W\right)\right)^2\right)\boldsymbol{\varepsilon},$$

with $W \sim U([-2, 2])$ independent of the bivariate standard normal vector $\boldsymbol{\varepsilon}$. In Figure 1, we are plotting the $\tau = 0.2$ and $\tau = 0.4$ HPŠ regression quantile/depth contours obtained by using the covariate vector $\mathbf{X} = (1, W)'$ (Figure 1(a)) and the covariate vector $\mathbf{X} = (1, W, W^2)'$ (Figure 1(b)) in the equations of the quantile/depth hyperplanes of the (global) HPŠ mehod. More precisely, these figures provide the intersections of the HPŠ contours with hyperplanes orthogonal to the w-axis at fixed w-values $-1.89, -1.83, -1.77, \ldots, 1.89$.

Clearly, the results are very poor: Figure 1(a) neither reveals the parabolic trend, nor the periodic heteroskedasticity pattern in the data. Although it is obtained by fitting the "true" regression function, Figure 1(b), while doing much better with the trend, still fails to catch heteroskedasticity correctly. Instead of providing genuine conditional quantile/depth contours, the "global" HPŠ methodology produces some averaged (over the w values) contours.

In contrast, the contours obtained from the local constant and local bilinear methods proposed in this paper – without exploiting any a priori knowledge of the actual regression function – exhibit a very good agreement with the population contours (see Figure 1(c)–(e) to which we refer for details); both the parabolic trend and the periodic heteroskedascticity features now are picked up quite satisfactorily. Note that, compared to the local constant approach, the local bilinear one does better, as expected, close to the boundary of the regressor space (in particular, the local constant approach is missing the decay of the conditional scale when w converges to -2).

Similar comments remain valid for smaller sample sizes; see Figure 3, based on a sample of $n = 499$ data points.

A second example is contrasting a homoskedastic setup and a heteroskedastic one. More specifically, we generated $n = 999$ points from the homoskedastic model $(Y_1, Y_2) = (W, W^2) + \boldsymbol{\varepsilon}$ and from the heteroskedastic one $(Y_1, Y_2) = (W, W^2) + (1 + W^2)\boldsymbol{\varepsilon}$, where $W \sim U([-2, 2])$

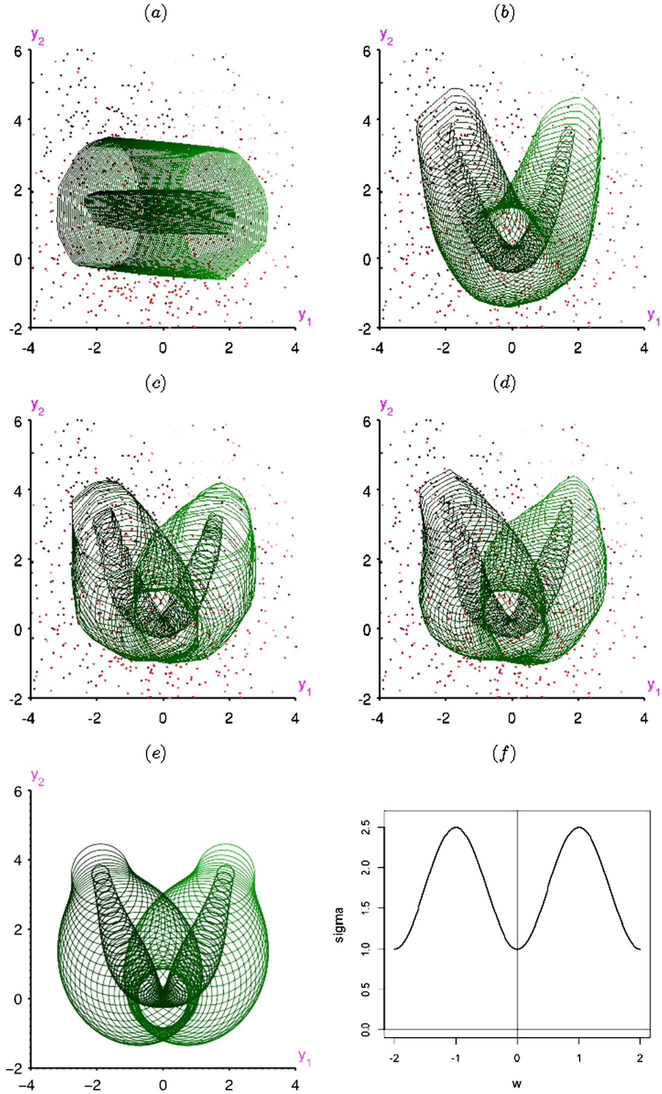

Figure 1. For $n = 999$ points following the model $(Y_1, Y_2) = (W, W^2) + (1 + \frac{3}{2}(\sin(\frac{\pi}{2}W))^2)\boldsymbol{\varepsilon}$, where $W \sim U([-2, 2])$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)^2$ are independent, the plots above show the intersections, with hyperplanes orthogonal to the w-axis at fixed w-values $-1.89, -1.83, -1.77, \ldots, 1.89$, of (a) the HPŠ regression quantile regions with the single random regressor W, (b) the HPŠ regression quantile regions with random regressors W and W^2, and (c)–(d) the proposed local constant and local bilinear regression quantile regions (in each case, $\tau = 0.2$ and $\tau = 0.4$ are considered). For the sake of comparison, the corresponding population (conditional) halfspace depth regions are provided in (e). The conditional scale function $w \mapsto 1 + \frac{3}{2}(\sin(\frac{\pi}{2}w))^2$ is plotted in (f). Local methods use a Gaussian kernel and bandwidth value $H = 0.37$, and 360 equispaced directions $\mathbf{u} \in \mathcal{S}^1$ were used to obtain results in (d).

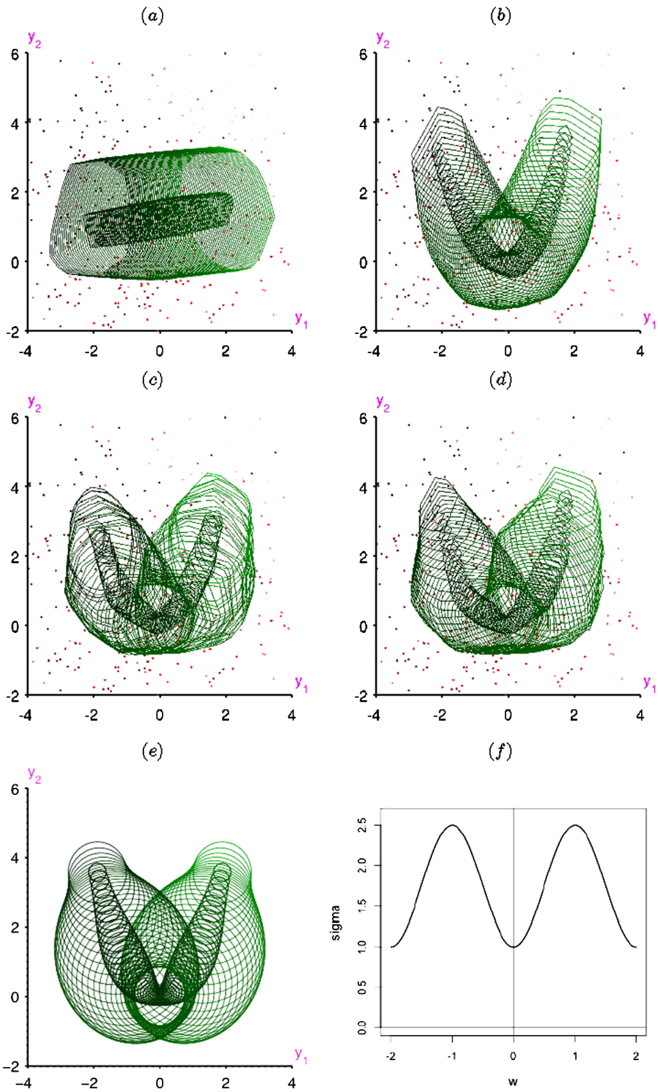

Figure 2. For $n = 499$ points following the model $(Y_1, Y_2) = (W, W^2) + (1 + \frac{3}{2}(\sin(\frac{\pi}{2}W))^2)\boldsymbol{\varepsilon}$, where $W \sim U([-2, 2])$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1)^2$ are independent, the plots above show the intersections, with hyperplanes orthogonal to the w-axis at fixed w-values $-1.89, -1.83, -1.77, \ldots, 1.89$, of (a) the HPŠ regression quantile regions with the single random regressor W, (b) the HPŠ regression quantile regions with random regressors W and W^2, and (c)–(d) the proposed local constant and local bilinear regression quantile regions (in each case, $\tau = 0.2$ and $\tau = 0.4$ are considered). For the sake of comparison, the corresponding population (conditional) halfspace depth regions are provided in (e). The conditional scale function $w \mapsto 1 + \frac{3}{2}(\sin(\frac{\pi}{2}w))^2$ is plotted in (f). Local methods use a Gaussian kernel and bandwidth value $H = 0.37$, and 360 equispaced directions $\mathbf{u} \in \mathcal{S}^1$ were used to obtain results in (d).

and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1/4)^2$ are mutually independent. As above, the intersections of the resulting contours with hyperplanes orthogonal to the w-axis at fixed w-values are provided. Figure 3 shows those intersections for the local constant and local bilinear quantile contours associated with $w \in \{-1.89, -1.83, -1.77, \ldots, 1.89\}$, for $\tau = 0.2$ and $\tau = 0.4$. As in the previous example, those sample contours approximate their population counterparts (shown in Figure 3(e) and (f)) remarkably well. In particular, the inner regions mimic the trend faithfully even for quite extreme regressor values. Again, the local bilinear method seems to provide a much better boundary behavior than its local constant counterpart; in the heteroskedastic case, the latter indeed severely underestimates the conditional scale for extreme values of W.

1.3. Relation to the depth and multivariate quantile literature

As already explained, this work is lying at the intersection of two distinct, if not unrelated, strands of the statistical literature – namely (i) statistical depth and (ii) multivariate quantiles. Under both strands, definitions have been proposed for *unconditional* concepts, that is, for statistical models that do not involve covariates. When covariates are present, the focus is shifted from unconditional features to *conditional* ones. The main objective, indeed, now is the analysis of the dependence of a response $\mathbf{Y}$ on a set of covariates $\mathbf{X}$, that is, a study of the distributions of $\mathbf{Y}$ conditional on the values $\mathbf{x}$ of $\mathbf{X}$ – in its broadest sense, the *regression problem* – and various attempts have been made to propose *regression* versions of (unconditional) depth or quantile concepts, respectively.

Now, if a study of the dependence on $\mathbf{x}$ of the distributions of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$ is the main objective, *conditional* depth and *conditional* (multivariate) quantiles, associated with the distributions of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$, are or should be the concepts of interest. Not all definitions of *regression depth* or (multiple-output) regression quantiles are meeting that requirement, though. Nor do they all preserve, conditionally on $\mathbf{X} = \mathbf{x}$, the distinctive properties of a depth/quantile concept. In contrast with this, the concept we are proposing in this paper, being the conditional version of the unconditional HPŠ concept, enjoys all the properties that are expected from a conditional depth/quantile concept, while fully characterizing the conditional distributions of $\mathbf{Y}$.

1.3.1. *Regression depth*

An excellent summary of depth-related problems is provided in Serfling [37], which further clarifies the nature of depth by placing it in the broader perspective of the so-called *DOQR paradigm*, relating Depth to the companion concepts of Outlyingness, Quantiles, and Ranks. To the best of our knowledge, this paradigm never has been considered in a conditional (regression) context, but it seems quite desirable that any *regression depth* concept should similarly be placed, conditionally, in the same DOQR perspective.

The celebrated *regression depth* concept by Rousseeuw and Hubert [35], for instance, does not bear any direct relation to conditional depth and the DOQR paradigm. Rather than the depth of a point in the observation space, that concept aims at defining, via *non-fits* and *breakdown values*, the depth of a (single-output) regression hyperplane. A multiple-output version is considered in Bern and Eppstein [1]. Similarly, an elegant general theory has been developed by Mizera [33]

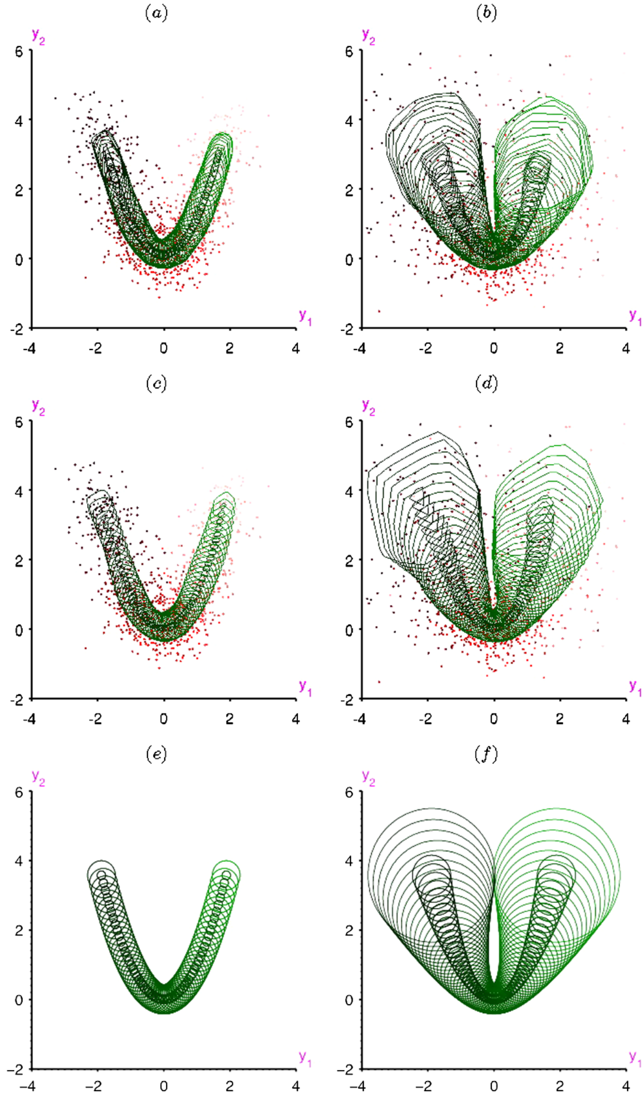

Figure 3. Local multiple-output quantile regression with Gaussian kernel and ad-hoc bandwidth $H = 0.37$: cuts through $w \in \{-1.89, -1.83, -1.77, \ldots, 1.89\}$ for $\tau = 0.2$ and $\tau = 0.4$ corresponding to $n = 999$ random points drawn from a homoskedastic model $(Y_1, Y_2) = (W, W^2) + \boldsymbol{\varepsilon}$ ((a), (c)) or a heteroskedastic model $(Y_1, Y_2) = (W, W^2) + (1 + W^2)\boldsymbol{\varepsilon}$ ((b), (d)), where $W \sim U([-2, 2])$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, 1/4)^2$ are independent. The plots are showing the intersections, with hyperplanes orthogonal to the w-axis at fixed w-values, of the contours obtained either from the local constant method ((a), (b)) or the local bilinear one ((c), (d)). Color scaling of the points (resp., the intersections) mimics their regressor values, whose higher values are indicated by lighter red (resp., lighter green). For the sake of comparison, the population (conditional) halfspace depth regions are provided in (e) and (f). A color version of this figure is more readable, and can be found in the on-line edition of the paper.

who, in the context of a general parametric model, defines the depth of a parameter value. Again, the approach and, despite the terminology, the concept, is of a different nature, unrelated to any conditional depth. Extensions to a nonparametric regression setting, moreover, seem problematic.

Kong and Mizera [29] propose an approach to unconditional depth, based on projection quantiles, which provides an approximation to the unconditional halfspace depth contours – see [18] and [29]. Although an application to bivariate growth charts is briefly described, in which a local smoothing, based on regression spline techniques, of their unconditional concept is performed (little details are provided), the regression setting is only briefly touched there. In particular, no asymptotic analysis of the type we are providing in Section 5 is made available.

1.3.2. *Multivariate regression quantiles*

Turning to conditional multivariate or multiple-output regression quantile issues, much work has been devoted to the notion of *spatial* regression quantiles; see, essentially, Chakraborty [5] for linear and Cheng and De Gooijer [7] for nonparametric regression. Despite a strong depth flavor, those spatial quantiles and spatial regression quantiles, however, intrinsically fail to be affine-equivariant; Chakraborty [5] defines affine-equivariant spatial quantiles for linear regression via a transformation–retransformation device, but, to the best of our knowledge, there exists no affine-equivariant version of spatial quantiles for general nonparametric regression.

For the sake of completeness, one also should mention here the closely related literature on growth charts described at the end of Section 1.1, which, besides a lack of affine-invariance, essentially fails, in the multiple-output case, to address the conditional nature of the regression quantile concept it is dealing with.

1.4. Outline of the paper

The rest of this paper is organized as follows. Section 2 defines the (population) conditional regression quantile/depth regions and contours we would like to estimate in the sequel. This estimation will make use of (empirical) *weighted* multiple-output regression quantiles, which we introduce in Section 3. Section 4 explains how these weighted quantiles lead to local constant (Section 4.2) and local bilinear (Section 4.3) depth contours. Section 5 provides asymptotic results (Bahadur representation and asymptotic normality) both for the local constant and local bilinear cases. Section 6 deals with the practical problem of bandwidth selection. In Section 7, the usefulness and applicability of the proposed methods are illustrated on real data. Finally, the Appendix collects proofs of asymptotic results.

2. Conditional multiple-output quantile/depth contours

Denote by $(\mathbf{X}_i', \mathbf{Y}_i')' = (X_{i1}, \dots, X_{ip}, Y_{i1}, \dots, Y_{im})'$, $i = 1, \dots, n$, an observed n-tuple of independent copies of $(\mathbf{X}', \mathbf{Y}')'$, where $\mathbf{Y} := (Y_1, \dots, Y_m)'$ is an m-dimensional response and $\mathbf{X} := (1, \mathbf{W}')'$ a p-dimensional random vector of covariates. For any $\tau \in (0, 1)$ and any direction $\mathbf{u}$ in the unit sphere $\mathcal{S}^{m-1}$ of the m-dimensional space of the response $\mathbf{Y}$, the HPŠ concept produces a hyperplane $\pi_{\tau\mathbf{u}}$ ($\pi_{\tau\mathbf{u}}^{(n)}$ in the empirical case) which is defined as the classical Koenker and Bassett regression quantile hyperplane of order τ once $(\mathbf{0}'_{p-1}, \mathbf{u}')'$ has been chosen as the "vertical direction" in the computation of the relevant L_1 deviations.

More specifically, decompose $\mathbf{y} \in \mathbb{R}^m$ into $(\mathbf{u}'\mathbf{y})\mathbf{u} + \mathbf{\Gamma_u}(\mathbf{\Gamma}'_\mathbf{u}\mathbf{y})$, where $\mathbf{\Gamma_u}$ is such that $(\mathbf{u}, \mathbf{\Gamma_u})$ is an $m \times m$ orthogonal matrix; then the *directional quantile hyperplanes* $\pi_{\tau\mathbf{u}}$ and $\pi^{(n)}_{\tau\mathbf{u}}$ are the hyperplanes with equations

$$\mathbf{u}'\mathbf{y} - \mathbf{c}'_\tau \mathbf{\Gamma}'_\mathbf{u}\mathbf{y} - \mathbf{a}'_\tau\big(1, \mathbf{w}'\big)' = 0 \quad \text{and} \quad \mathbf{u}'\mathbf{y} - \mathbf{c}^{(n)\prime}_\tau \mathbf{\Gamma}'_\mathbf{u}\mathbf{y} - \mathbf{a}^{(n)\prime}_\tau\big(1, \mathbf{w}'\big)' = 0 \tag{2.1}$$

($\mathbf{w} \in \mathbb{R}^{p-1}$) minimizing, with respect to $\mathbf{c} \in \mathbb{R}^{m-1}$ and $\mathbf{a} \in \mathbb{R}^p$,

$$\mathrm{E}\big[\rho_\tau\big(\mathbf{u}'\mathbf{Y} - \mathbf{c}'\mathbf{\Gamma}'_\mathbf{u}\mathbf{Y} - \mathbf{a}'\mathbf{X}\big)\big] \quad \text{and} \quad \sum_{i=1}^n \rho_\tau\big(\mathbf{u}'\mathbf{Y}_i - \mathbf{c}'\mathbf{\Gamma}'_\mathbf{u}\mathbf{Y}_i - \mathbf{a}'\mathbf{X}_i\big), \tag{2.2}$$

respectively, where $\zeta \mapsto \rho_\tau(\zeta)$, with

$$\rho_\tau(\zeta) := \zeta\big(\tau - I[\zeta < 0]\big) = \max\big\{(\tau-1)\zeta, \tau\zeta\big\} = \big(|\zeta| + (2\tau-1)\zeta\big)/2, \qquad \zeta \in \mathbb{R} \tag{2.3}$$

as usual denotes the well-known τ-quantile *check function*. HPŠ moreover show that $\pi_{\tau\mathbf{u}}$ and $\pi^{(n)}_{\tau\mathbf{u}}$ can equivalently be defined, in a more symmetric way, as the hyperplanes with equations

$$\mathbf{b}'_\tau\mathbf{y} - \mathbf{a}'_\tau\big(1, \mathbf{w}'\big)' = 0 \quad \text{and} \quad \mathbf{b}^{(n)\prime}_\tau\mathbf{y} - \mathbf{a}^{(n)\prime}_\tau\big(1, \mathbf{w}'\big)' = 0, \tag{2.4}$$

minimizing, with respect to $\mathbf{b} \in \mathbb{R}^m$ satisfying $\mathbf{b}'\mathbf{u} = 1$ and $\mathbf{a} \in \mathbb{R}^p$, the L_1 criteria

$$\mathrm{E}\big[\rho_\tau\big(\mathbf{b}'\mathbf{Y} - \mathbf{a}'\mathbf{X}\big)\big] \quad \text{and} \quad \sum_{i=1}^n \rho_\tau\big(\mathbf{b}'\mathbf{Y}_i - \mathbf{a}'\mathbf{X}_i\big), \tag{2.5}$$

respectively.

For $p = 1$, the multiple-output regression model reduces to a multivariate location one: $\mathbf{a}_\tau$ and $\mathbf{a}^{(n)}_\tau$ reduce to scalars, a_τ and $a^{(n)}_\tau$, while the equations describing $\pi_{\tau\mathbf{u}}$ and $\pi^{(n)}_{\tau\mathbf{u}}$ take the simpler forms

$$\mathbf{u}'\mathbf{y} - \mathbf{c}'_\tau\mathbf{\Gamma}'_\mathbf{u}\mathbf{y} - a_\tau = 0 \quad \text{and} \quad \mathbf{u}'\mathbf{y} - \mathbf{c}^{(n)\prime}_\tau\mathbf{\Gamma}'_\mathbf{u}\mathbf{y} - a^{(n)}_\tau = 0, \tag{2.6}$$

respectively. Those *location* quantile hyperplanes $\pi_{\tau\mathbf{u}}$ and $\pi^{(n)}_{\tau\mathbf{u}}$ are studied in detail in HPŠ, where it is shown that their fixed-τ collections characterize regions and contours that actually coincide with the Tukey halfspace depth ones. Consistency, asymptotic normality and Bahadur-type representation results for the $\pi^{(n)}_{\tau\mathbf{u}}$'s are also provided there, together with a linear programming method for their computation.

The objective here is an analysis of the distribution of $\mathbf{Y}$ conditional on $\mathbf{W}$, that is, of the dependence of $\mathbf{Y}$ on $\mathbf{W}$ – in strong contrast with traditional regression, where investigation is limited to the mean of $\mathbf{Y}$ conditional on $\mathbf{W}$. The relevant quantile hyperplanes, depth regions and contours of interest are the location quantile/depth hyperplanes/regions/contours associated (in the sense of HPŠ) with the m-dimensional distributions of $\mathbf{Y}$ conditional on $\mathbf{W}$ – more precisely, with the distributions $\mathrm{P}^{\mathbf{Y}|\mathbf{W}=\mathbf{w}_0}$ of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}_0$ ($\mathbf{w}_0 \in \mathbb{R}^{p-1}$). We now carefully define these objects – call them $\mathbf{w}_0$-*conditional* $\boldsymbol{\tau}$-*quantile or depth hyperplanes, regions and contours*.

Let $\tau \in (0,1)$ and $\mathbf{u} \in \mathcal{S}^{m-1} := \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{u}\| = 1\}$ (the unit sphere in $\mathbb{R}^m$), and write $\boldsymbol{\tau} := \tau\mathbf{u}$. Denoting by $\mathbf{w}_0$ some fixed point of $\mathbb{R}^{p-1}$ at which the marginal density $f^{\mathbf{W}}$ of $\mathbf{W}$ does not vanish (in order for the distribution of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}_0$ to make sense), define the *extended* and *restricted* $\mathbf{w}_0$-*conditional* $\boldsymbol{\tau}$-*quantile hyperplanes of* $\mathbf{Y}$ as the $(m+p-2)$-dimensional and $(m-1)$-dimensional hyperplanes

$$\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{y} - a_{\boldsymbol{\tau};\mathbf{w}_0} = 0\} \tag{2.7}$$

and

$$\pi_{\boldsymbol{\tau};\mathbf{w}_0} := \{(\mathbf{w}'_0, \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{y} - a_{\boldsymbol{\tau};\mathbf{w}_0} = 0\}, \tag{2.8}$$

respectively, where $a_{\boldsymbol{\tau};\mathbf{w}_0}$ and $\mathbf{b}_{\boldsymbol{\tau};\mathbf{w}_0}$ minimize

$$\Psi_{\tau;\mathbf{w}_0}(a, \mathbf{b}) := \mathrm{E}[\rho_\tau(\mathbf{b}'\mathbf{Y} - a) \mid \mathbf{W} = \mathbf{w}_0] \qquad \text{subject to } \mathbf{b}'\mathbf{u} = 1, \tag{2.9}$$

with the check function ρ_τ defined in (2.3). Comparing (2.9) with (2.5) immediately shows that $\pi_{\boldsymbol{\tau};\mathbf{w}_0}$ is the $(m-1)$-dimensional (location) $\boldsymbol{\tau}$-quantile hyperplane of $\mathbf{Y}$ associated with the distribution of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}_0$. Of course, $\pi_{\boldsymbol{\tau};\mathbf{w}_0}$ is also the intersection of $\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$ with the m-dimensional hyperplane $C_{\mathbf{w}_0} := \{(\mathbf{w}'_0, \mathbf{y}')' \mid \mathbf{y} \in \mathbb{R}^m\}$. This, and the fact that $\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$ is "parallel to the space of covariates" (in the sense that if $(\mathbf{w}'_0, \mathbf{y}'_0)' \in \boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$, then $(\mathbf{w}', \mathbf{y}'_0)' \in \boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$ for all $\mathbf{w}$), fully characterizes $\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$.

Associated with $\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$ are the extended *upper and lower* $\mathbf{w}_0$-*conditional* $\boldsymbol{\tau}$-*quantile halfspaces*

$$\mathbf{H}^+_{\boldsymbol{\tau};\mathbf{w}_0} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{y} - a_{\boldsymbol{\tau};\mathbf{w}_0} \geq 0\}$$

and

$$\mathbf{H}^-_{\boldsymbol{\tau};\mathbf{w}_0} := \{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{y} - a_{\boldsymbol{\tau};\mathbf{w}_0} < 0\},$$

with the extended (cylindrical) $\mathbf{w}_0$-*conditional quantile/depth regions*

$$\mathbf{R}_{\mathbf{w}_0}(\tau) := \bigcap_{\mathbf{u} \in \mathcal{S}^{m-1}} \{\mathbf{H}^+_{\tau\mathbf{u};\mathbf{w}_0}\} \tag{2.10}$$

and their boundaries $\partial\mathbf{R}_{\mathbf{w}_0}(\tau)$, the extended $\mathbf{w}_0$-*conditional quantile/depth contours*. The intersections of those extended regions $\mathbf{R}_{\mathbf{w}_0}(\tau)$ (resp., contours $\partial\mathbf{R}_{\mathbf{w}_0}(\tau)$) with $C_{\mathbf{w}_0}$ are the restricted $\mathbf{w}_0$-conditional quantile/depth regions $R_{\mathbf{w}_0}(\tau)$ (resp., contours $\partial R_{\mathbf{w}_0}(\tau)$), that is, the location HPŠ regions (resp., contours) for $\mathbf{Y}$, conditional on $\mathbf{W} = \mathbf{w}_0$. It follows from HPŠ that those regions are compact, convex, and nested. As a consequence, the regions $\mathbf{R}_{\mathbf{w}_0}(\tau)$ also are closed, convex, and nested.

Finally, define the *nonparametric* τ-*quantile/depth regions* as

$$\mathbf{R}(\tau) := \bigcup_{\mathbf{w}_0 \in \mathbb{R}^{p-1}} R_{\mathbf{w}_0}(\tau) = \bigcup_{\mathbf{w}_0 \in \mathbb{R}^{p-1}} (\mathbf{R}_{\mathbf{w}_0}(\tau) \cap C_{\mathbf{w}_0})$$

and write $\partial\mathbf{R}(\tau)$ for their boundaries. The regions $\mathbf{R}(\tau)$ are still closed and nested but they adapt to the general dependence of $\mathbf{Y}$ on $\mathbf{W}$: in particular, $\partial\mathbf{R}(\tau)$, for any τ, goes through *all* corresponding $\partial R_{\mathbf{w}_0}(\tau)$'s, $\mathbf{w}_0 \in \mathbb{R}^{p-1}$. Consequently, the regions $\mathbf{R}(\tau)$ in general are no longer convex.

The fixed-$\mathbf{w}_0$ collection (over $\tau \in (0, 1/2)$) of all $\mathbf{w}_0$-conditional location quantile/depth contours $\partial R_{\mathbf{w}_0}(\tau)$ (which, by construction, are the intersections of $\partial\mathbf{R}(\tau)$ with the "vertical hyperplanes" $C_{\mathbf{w}_0}$) will be called a $\mathbf{w}_0$-*quantile/depth cut* or $\mathbf{w}_0$-*cut*. Such cuts are of crucial interest, since they characterize the distribution of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}_0$, hence provide a full description of the dependence of the response $\mathbf{Y}$ on the regressors $\mathbf{W}$. Note that the nonparametric contours $\partial\mathbf{R}(\tau)$, via the location depth interpretation, for fixed $\mathbf{w}_0$, of the $\partial R_{\mathbf{w}_0}(\tau)$'s, inherit a most interesting interpretation as "regression depth contours". Clearly, this concept of regression depth, that defines regression depth of any point $(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{m+p-1}$, is not of the same nature as the regression depth concept proposed in [35], that defines the depth of any regression "fit" (i.e., of any regression hyperplane).

3. Weighted multiple-output empirical quantile regression

Under the assumption of absolute continuity, the number of observations, in a sample of size n, belonging to $C_{\mathbf{w}_0}$ clearly is (a.s.) zero, which implies that no empirical version of the conditional regression hyperplanes (2.7) or (2.8) can be constructed. If nonparametric τ-quantile/depth regions or contours, or simply some selected cuts, are to be estimated, local smoothing techniques have to be considered. Those techniques typically involve weighted versions, with sequences $\boldsymbol{\omega}^{(n)}_{\mathbf{w}_0} = (\omega^{(n)}_{\mathbf{w}_0,i}, i = 1, \dots, n)$ of weights, of the empirical quantile regression hyperplanes developed in HPŠ. In this section, we provide general definitions and basic results for such weighted concepts, under fixed sample size n and weights ω_i; see [21] for another approach combining weights with halfspace depth. In Section 4, we will consider the data-driven weights to be used in the local approach.

Consider a sample of size n, with observations $(\mathbf{X}'_i, \mathbf{Y}'_i)' = ((1, \mathbf{W}'_i), \mathbf{Y}'_i)'$, $i = 1, \dots, n$, along with n nonnegative weights ω_i satisfying (without any loss of generality) $\sum_{i=1}^n \omega_i = n$ ($\omega_i \equiv 1$ then yields the unweighted case). The definitions of HPŠ extend, *mutatis mutandis*, quite straightforwardly, into the following weighted versions. The coefficients $\mathbf{a}^{(n)}_{\boldsymbol{\tau};\omega} \in \mathbb{R}^p$ and $\mathbf{b}^{(n)}_{\boldsymbol{\tau};\omega} \in \mathbb{R}^m$ of the *weighted empirical* $\boldsymbol{\tau}$*-quantile hyperplane*

$$\boldsymbol{\pi}^{(n)}_{\boldsymbol{\tau};\omega} := \left\{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{y} - \mathbf{a}^{(n)\prime}_{\boldsymbol{\tau};\omega}(1, \mathbf{w}')' = 0\right\} \tag{3.1}$$

(an $(m + p - 2)$-dimensional hyperplane) are defined as the minimizers of

$$\Psi^{(n)}_{\tau;\omega}(\mathbf{a}, \mathbf{b}) := \frac{1}{n}\sum_{i=1}^n \omega_i \rho_\tau\left(\mathbf{b}'\mathbf{Y}_i - \mathbf{a}'\mathbf{X}_i\right) \qquad \text{subject to } \mathbf{b}'\mathbf{u} = 1. \tag{3.2}$$

As usual in the empirical case, the solution may not be unique, but the minimizers always form a convex set. When substituted for the $\boldsymbol{\pi}_{\boldsymbol{\tau};\mathbf{w}_0}$'s in the definitions of upper and lower conditional

$\boldsymbol{\tau}$-quantile halfspaces, those $\boldsymbol{\pi}^{(n)}_{\boldsymbol{\tau};\omega}$'s also characterize upper and lower weighted $\boldsymbol{\tau}$-quantile halfspaces $\mathbf{H}^{(n)+}_{\boldsymbol{\tau};\omega}$ and $\mathbf{H}^{(n)-}_{\boldsymbol{\tau};\omega}$, with *weighted τ-quantile/depth regions and contours*

$$\mathbf{R}^{(n)}_{\omega}(\tau) := \bigcap_{\mathbf{u}\in\mathcal{S}^{m-1}} \{\mathbf{H}^{(n)+}_{\tau\mathbf{u};\omega}\} \quad \text{and} \quad \partial\mathbf{R}^{(n)}_{\omega}(\tau),$$

respectively. Note that the objective function in (3.2) rewrites as

$$\Psi^{(n)}_{\tau;\omega}(\mathbf{a},\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\big(\mathbf{b}'\mathbf{Y}_{i;\omega} - \mathbf{a}'\mathbf{X}_{i;\omega}\big),$$

with $\mathbf{X}_{i;\omega} := \omega_i\mathbf{X}_i$ and $\mathbf{Y}_{i;\omega} := \omega_i\mathbf{Y}_i$. As an important consequence, the weighted quantile/depth hyperplanes, contours and regions can be computed in the same way as their non-weighted counterparts because the corresponding algorithm in [34] allows to have $(\mathbf{X}_i)_1 \neq 1$. Due to quantile crossing, however, and contrary to the population regions and contours defined in the previous section, the $\mathbf{R}^{(n)}_{\omega}(\tau)$'s need not be nested for $p \geq 2$; if nestedness is required, one may rather consider the regions $\mathbf{R}^{(n)}_{\omega\cap}(\tau) := \bigcap_{0<t\leq\tau}\{\mathbf{R}^{(n)}_{\omega}(t)\}$.

The necessary sample subgradient conditions for $(\mathbf{a}^{(n)\prime}_{\boldsymbol{\tau};\omega}, \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau};\omega})'$ can be derived as in the unweighted case. They state in particular that

$$\frac{1}{n}\sum_{i=1}^{n}\omega_i I\big[\mathbf{b}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{Y}_i - \mathbf{a}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{X}_i < 0\big] \leq \tau \leq \frac{1}{n}\sum_{i=1}^{n}\omega_i I\big[\mathbf{b}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{Y}_i - \mathbf{a}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{X}_i \leq 0\big],$$

which controls the probability contents of $\mathbf{H}^{(n)-}_{\boldsymbol{\tau};\omega}$ with respect to the distribution putting probability mass ω_i/n on $(\mathbf{W}'_i, \mathbf{Y}'_i)'$, $i = 1, \ldots, n$. The width of this interval depends only on the weights ω_i associated with those data points $(\mathbf{W}'_i, \mathbf{Y}'_i)'$ that belong to $\boldsymbol{\pi}^{(n)}_{\boldsymbol{\tau};\omega}$. Another consequence worth mentioning is that there always exists a $\boldsymbol{\pi}^{(n)}_{\tau\mathbf{u};\omega}$ hyperplane containing at least $(m + p - 1)$ data points of the form $(\mathbf{W}_i, \mathbf{Y}_i)$. With probability one, thus, the intersection defining the regions $\mathbf{R}^{(n)}_{\omega}(\tau)$ is finite.

Note that, unlike the extended conditional quantile hyperplanes (2.7), the weighted empirical quantile hyperplanes (3.1) involve an unrestricted coefficient $\mathbf{a} \in \mathbb{R}^p$. As a consequence, $\boldsymbol{\pi}^{(n)}_{\boldsymbol{\tau};\omega}$ is not necessarily parallel to the space of covariates (as defined in page 1444). That degree of freedom will be exploited in the local linear approach described in Section 4.3 (in an augmented regressor space, though, which makes it bilinear rather than linear). If we impose the additional constraint $\mathbf{a} = (a_1, 0, \ldots, 0)'$ in (3.1) and (3.2), we obtain hyperplanes of the form

$$\boldsymbol{\pi}^{(n)}_{\boldsymbol{\tau};\omega} := \big\{(\mathbf{w}', \mathbf{y}')' \in \mathbb{R}^{p-1} \times \mathbb{R}^m \mid \mathbf{b}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{y} - a^{(n)}_{1;\boldsymbol{\tau};\omega} = 0\big\}. \tag{3.3}$$

The corresponding minimization problem yields hyperplanes that are parallel to the space of covariates, hence "horizontal" cylindrical regions and contours, to be considered in the local constant approach of Section 4.2.

Finally, it should be pointed out that (**y** and/or **w**)-affine-invariant weights $\omega_i := \omega(\mathbf{w}_i, \mathbf{y}_i)$ yield weighted quantile/depth hyperplanes, regions, and contours with good (**y** and/or **w**)-affine-equivariance properties.

4. Local quantile/depth regression

4.1. From weighted to local quantile/depth regression

The weighted quantiles of Section 3 have an interest on their own. They can be used for handling multiple identical observations (allowing, for instance, for bootstrap procedures), or for down-weighting observations that are suspected to be outliers or leverage points. Above all, weighted regression quantiles allow for a nonparametric approach to regression quantiles that will take care of the drawbacks of the unweighted approach of HPŠ (see the example considered in Section 1.2). In particular, adequate sequences of weights will allow to estimate the conditional contours described in Section 2, thus extending to the multiple-output case the *local constant* and *local linear* approaches to quantile regression proposed, for example, by [43,44] in the single-output context.

The basic idea is very standard: in order to estimate $\mathbf{w}_0$-conditional quantile/depth hyperplanes, regions or contours, we will consider weighted quantile/depth hyperplanes, regions or contours, with sequences of weights $\omega_i^{(n)} := \omega_{\mathbf{w}_0}^{(n)}(\mathbf{W}_i)$ based on weight functions of the form

$$\mathbf{w} \mapsto \omega_{\mathbf{w}_0}^{(n)}(\mathbf{w}) := h_n^{-p+1} K\big(h_n^{-1}(\mathbf{w} - \mathbf{w}_0)\big), \tag{4.1}$$

where h_n is a sequence of positive bandwidths and K a nonnegative kernel function over $\mathbb{R}^{p-1}$. The literature proposes a variety of possible kernels, and there is no compelling reason for not considering the most usual, such as the rectangular (uniform), Epanechnikov or (spherical) Gaussian ones.

Since we typically intend, for any fixed $\tau \in (0, 1)$, to compute by means of parametric programming the directional quantile hyperplanes for all $\mathbf{u} \in \mathcal{S}^{m-1}$, we should use the same weights for all of them. This is why we only consider **u**-independent bandwidths. However, exact computation of all quantiles (for each fixed τ) is possible in the local constant case, but not in the local bilinear one. In the latter case, depth contours will be approximated by sampling the unit sphere (in Figures 1 and 2, for instance, 360 directions were sampled uniformly over the unit circle), which of course would allow **u**-dependent bandwidths if desired.

4.2. Local constant quantile/depth contours

The above weighting scheme can be applied in the computation of the weighted cylindrical regions generated by the hyperplanes in (3.3); more precisely, these cylindrical regions, with edges parallel to the space of covariates, are obtained by computing the intersection (over all **u**'s, for fixed τ) of the upper quantile halfspaces associated with the quantile hyperplanes in (3.3); see Figure 4(a).

The intersection with the $\mathbf{w} = \mathbf{w}_0$ hyperplane of these cylindrical regions yields a local constant estimate, $\partial \hat{R}_{\mathbf{w}_0}^{(n)\mathrm{const}}(\tau)$ say, of the corresponding population $\mathbf{w}_0$-cut $\partial R_{\mathbf{w}_0}(\tau)$; see Section 5

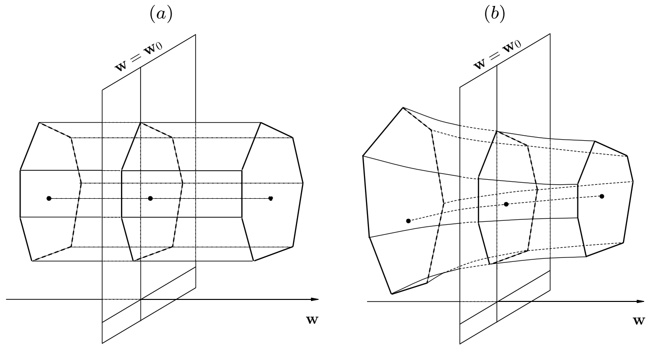

Figure 4. Construction of (a) the local constant and (b) the local bilinear τ-quantile regions as described in Sections 4.2 and 4.3.

for asymptotic results. Of course, the resulting local constant τ-quantile/depth contours, namely

$$\partial\hat{\mathbf{R}}^{(n)\mathrm{const}}(\tau) := \bigcup_{\mathbf{w}_0 \in \mathbb{R}^{p-1}} \partial\hat{R}^{(n)\mathrm{const}}_{\mathbf{w}_0}(\tau),$$

are not (globally) cylindrical, but rather adapt to the underlying possibly nonlinear and/or heteroskedastic dependence structures.

This approach, which constitutes a generalization of the local constant approach adopted elsewhere for single-output regression, has many advantages. The main one is parsimony: each quantile hyperplane involved in the construction only entails m parameters, which is strictly less than in the local bilinear approach of the next section. On the other hand, the local constant approach does not provide any information on, nor does take any advantage of, the behavior of $\mathbf{w}$-cuts for $\mathbf{w}$ values in the neighborhood of $\mathbf{w}_0$, and its boundary performances are likely to be poor. These two reasons, in traditional contexts, have motivated the development of local linear and local polynomial methods; see [10] for a classical reference. Local linear methods were successfully used in single-output quantile regression ([43–45,47]). Considering them in the present context, thus, is a quite natural idea.

4.3. Local bilinear quantile/depth contours

Assume that the distribution of $(\mathbf{W}', \mathbf{Y}')'$ is smooth enough that the coefficients of $\mathbf{w}$-conditional quantile hyperplanes are differentiable with respect to $\mathbf{w}$. Getting back to the first characterization (2.1) and (2.2) of quantile hyperplanes, the (restricted) $\mathbf{w}_0$-conditional $\boldsymbol{\tau}$-quantile hyperplane of $\mathbf{Y}$ defined in (2.8) and (2.9) has equation (in $\mathbf{y}$ – of course, in $\mathbf{w}$, we just have $\mathbf{w} = \mathbf{w}_0$)

$$\mathbf{u}'\mathbf{y} - \left(a_{\boldsymbol{\tau};\mathbf{w}_0}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\right)\begin{pmatrix} 1 \\ \boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y} \end{pmatrix} = 0. \tag{4.2}$$

The same hyperplane equation, relative to a point $\mathbf{w}$ in the neighborhood of $\mathbf{w}_0$, takes the form

$$\begin{aligned}\mathbf{u}'\mathbf{y} &- \big(a_{\boldsymbol{\tau};\mathbf{w}_0}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\big)\begin{pmatrix}1\\ \boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y}\end{pmatrix} \\ &- (\mathbf{w}-\mathbf{w}_0)'\big(\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0}, \dot{\mathbf{c}}'_{\boldsymbol{\tau};\mathbf{w}_0}\big)\begin{pmatrix}1\\ \boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y}\end{pmatrix} + \mathrm{o}\big(\|\mathbf{w}-\mathbf{w}_0\|\big) = 0,\end{aligned} \tag{4.3}$$

where $\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0}$ stands for the gradient of $\mathbf{w} \mapsto a_{\boldsymbol{\tau};\mathbf{w}}$ and $\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}}$ for the Jacobian matrix of $\mathbf{w} \mapsto \mathbf{c}_{\boldsymbol{\tau};\mathbf{w}}$, respectively, both taken at $\mathbf{w} = \mathbf{w}_0$. In order to express this equation into the equivalent quantile formulation in (2.4) and (2.5), note that we have $\mathbf{b}_{\boldsymbol{\tau};\mathbf{w}_0} = \mathbf{u} - \boldsymbol{\Gamma}_{\mathbf{u}}\mathbf{c}_{\boldsymbol{\tau};\mathbf{w}_0}$, which entails $\dot{\mathbf{b}}_{\boldsymbol{\tau};\mathbf{w}_0} = -\boldsymbol{\Gamma}_{\mathbf{u}}\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}_0}$, where $\dot{\mathbf{b}}_{\boldsymbol{\tau};\mathbf{w}_0}$ is the Jacobian matrix of $\mathbf{w} \mapsto \mathbf{b}_{\boldsymbol{\tau};\mathbf{w}}$ at $\mathbf{w} = \mathbf{w}_0$. Neglecting the $\mathrm{o}(\|\mathbf{w}-\mathbf{w}_0\|)$ term, (4.3) then rewrites, after some algebra, as

$$\begin{aligned}&\big(\mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0} - \mathbf{w}'_0\dot{\mathbf{b}}'_{\boldsymbol{\tau};\mathbf{w}_0}\big)\mathbf{y} \\ &- \big(a_{\boldsymbol{\tau};\mathbf{w}_0} - \mathbf{w}'_0\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0}, \dot{\mathbf{a}}'_{\boldsymbol{\tau};\mathbf{w}_0}, -(\operatorname{vec}\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}_0})'\big)\begin{pmatrix}1\\ \mathbf{w}\\ \mathbf{w}\otimes\big(\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y}\big)\end{pmatrix} = 0.\end{aligned} \tag{4.4}$$

Letting $\bar{\mathbf{x}} := (1, \bar{\mathbf{w}}')' := (1, \mathbf{w}', (\mathbf{w}\otimes\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y})')'$, the latter equation is of the form $\boldsymbol{\beta}'_{\boldsymbol{\tau}}\mathbf{y} - \boldsymbol{\alpha}'_{\boldsymbol{\tau}}(1, \bar{\mathbf{w}}')' = 0$, with $\boldsymbol{\beta}'_{\boldsymbol{\tau}}\mathbf{u} = (\mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0} - \mathbf{w}'_0\dot{\mathbf{b}}'_{\boldsymbol{\tau};\mathbf{w}_0})\mathbf{u} = \mathbf{b}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{u} = 1$ since $\dot{\mathbf{b}}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{u} = -\dot{\mathbf{c}}'_{\boldsymbol{\tau};\mathbf{w}_0}\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{u} = 0$. Comparing with (2.4), this suggests a local linear approach based on weighted quantile hyperplanes (in the mp-dimensional regressor-response space associated with the augmented regressor $\bar{\mathbf{x}}$, that is, the $(\bar{\mathbf{w}}', \mathbf{y}')'$-space), yielding weighted empirical quantile hyperplanes with equations

$$\boldsymbol{\beta}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{y} - \boldsymbol{\alpha}^{(n)\prime}_{\boldsymbol{\tau};\omega}\big(1, \bar{\mathbf{w}}'\big)' = 0, \tag{4.5}$$

based on the same sequences of weights $\omega_i^{(n)} := \omega^{(n)}_{\mathbf{w}_0}(\mathbf{W}_i)$, $i = 1, \dots, n$, as in Section 4.1. Interpretation of the results, however, is easier from (4.3) than from (4.4). The left-hand side of (4.3) indeed splits naturally into two parts of independent interest: (i) the first one, made of the first two terms, yields the equation of the $\mathbf{w}_0$-conditional $\boldsymbol{\tau}$-quantile hyperplane of $\mathbf{Y}$, hence provides the required information for constructing the empirical $\mathbf{w}_0$-cuts, whereas (ii) the second part (the third term) provides the linear (linear with respect to $(\mathbf{w}-\mathbf{w}_0)$; actually, bilinear in $(\mathbf{w}-\mathbf{w}_0)$ and $\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y}$) correction required for a small perturbation $(\mathbf{w}-\mathbf{w}_0)$ of the value of the conditioning variable. Therefore, the important quantities to be recovered from $\boldsymbol{\alpha}^{(n)}_{\boldsymbol{\tau};\omega}$ and $\boldsymbol{\beta}^{(n)}_{\boldsymbol{\tau};\omega}$ are estimations of these two parts, which are easily obtained by

(i) letting $\mathbf{w} = \mathbf{w}_0$ in (4.5), which yields the equation

$$\boldsymbol{\beta}^{(n)\prime}_{\boldsymbol{\tau};\omega}\mathbf{y} - \boldsymbol{\alpha}^{(n)\prime}_{\boldsymbol{\tau};\omega}\big(1, \mathbf{w}'_0, \big(\mathbf{w}_0\otimes\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{y}\big)'\big)' = 0$$

of an empirical hyperplane providing an estimate of the two first terms in (4.3), namely, the $\mathbf{w}_0$-conditional $\boldsymbol{\tau}$-quantile hyperplane;

(ii) subtracting the latter equation from (4.5), which provides the bilinear correction term.

The bilinear nature of the local approximation in (ii) is easily explained by the fact that, in general, unless the $\mathbf{w}_0$-conditional and $\mathbf{w}$-conditional $\boldsymbol{\tau}$-quantile hyperplanes are parallel to each other, no higher-dimensional hyperplane can run through both (for instance, two mutually skew non-intersecting straight lines in $\mathbb{R}^3$ do not span a plane). Omitting the additional $\mathbf{W} \otimes (\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Y})$ regressors (in (i) above) may result in inconsistent estimators of the $\mathbf{w}_0$-conditional $\boldsymbol{\tau}$-quantile hyperplanes. The resulting regions in $\mathbb{R}^{m+p-1}$, are not polyhedral anymore, but delimited by ruled quadrics (hyperbolic paraboloids for $m = 2$ and $p - 1 = 1$), the intersections of which with the $\mathbf{w} = \mathbf{w}_0$ hyperplane yield polyhedral estimated $\mathbf{w}_0$-cuts; see Figure 4(b).

The local bilinear approach is more informative than the local constant one, and should be more reliable at boundary points; the price to be paid is an increase of the covariate space dimension (due to the presence of the regressors $\mathbf{W}$ and $\mathbf{W} \otimes (\boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Y})$ in (4.5)), hence of the number of free parameters (mp instead of m for the local constant method). Note however that the smoothing features of the problem, namely the dimension of kernels, remains unaffected ($p - 1$, irrespective of m).

5. Asymptotics

Throughout this section, we fix $\mathbf{w}_0$ and $\boldsymbol{\tau} = \tau\mathbf{u}$, hence also $a_{\boldsymbol{\tau};\mathbf{w}_0}$ and $\mathbf{c}_{\boldsymbol{\tau};\mathbf{w}_0}$, and write, for simplicity, $Y_{\mathbf{u}} := \mathbf{u}'\mathbf{Y}$ and $\mathbf{Y}_{\mathbf{u}}^{\perp} := \boldsymbol{\Gamma}'_{\mathbf{u}}\mathbf{Y}$. Asymptotic results require some regularity assumptions on the density f, the kernel K, and the bandwidth h_n.

Assumption (A1).

(i) *The n-tuple* $(\mathbf{W}'_i, \mathbf{Y}'_i)'$, $i = 1, \ldots, n$ *is an i.i.d. sample from* $(\mathbf{W}', \mathbf{Y}')'$.

(ii) *The density* $\mathbf{w} \mapsto f^{\mathbf{W}}(\mathbf{w})$ *of* $\mathbf{W}$ *is continuous and strictly positive at* $\mathbf{w}_0$.

(iii) *For any* $\mathbf{t} \in \mathbb{R}^{m-1}$, *there exist a neighborhood* $B_{\mathbf{t}}$ *of* $a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{t}$ *and a neighborhood* $\mathbf{B}_{\mathbf{t}}(\mathbf{w}_0)$ *of* $\mathbf{w}_0$ *such that* $s \mapsto f^{Y_{\mathbf{u}}|\mathbf{Y}_{\mathbf{u}}^{\perp}=\mathbf{t},\mathbf{W}=\mathbf{w}}(s)$ *is continuous over* $s \in B_{\mathbf{t}}$, *uniformly in* $\mathbf{w} \in \mathbf{B}_{\mathbf{t}}(\mathbf{w}_0)$, *and* $\mathbf{w} \mapsto f^{Y_{\mathbf{u}}|\mathbf{Y}_{\mathbf{u}}^{\perp}=\mathbf{t},\mathbf{W}=\mathbf{w}}(s)$ *is continuous over* $\mathbf{w} \in \mathbf{B}_{\mathbf{t}}(\mathbf{w}_0)$ *for all* $s \in B_{\mathbf{t}}$.

(iv) *The density* $f^{\mathbf{Y}_{\mathbf{u}}^{\perp}|\mathbf{W}=\mathbf{w}}(\mathbf{t})$ *of* $\mathbf{Y}_{\mathbf{u}}^{\perp}$ *conditional on* $\mathbf{W} = \mathbf{w}$ *is continuous with respect to* $\mathbf{w}$ *over a neighborhood of* $\mathbf{w}_0$, *except perhaps for a set of* $\mathbf{t}$ *values of* $f^{\mathbf{Y}_{\mathbf{u}}^{\perp}}$*-measure zero.*

(v) *The* $m \times m$ *matrix*

$$\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} := \int_{\mathbb{R}^{m-1}} \begin{pmatrix} 1 & \mathbf{t}' \\ \mathbf{t} & \mathbf{t}\mathbf{t}' \end{pmatrix} f^{Y_{\mathbf{u}}|\mathbf{Y}_{\mathbf{u}}^{\perp}=\mathbf{t},\mathbf{W}=\mathbf{w}_0}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{t}\big) f^{\mathbf{Y}_{\mathbf{u}}^{\perp}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t}$$

is finite and positive definite.

Assumption (A2). *The kernel function* K

(i) *is a compactly supported bounded probability density over* $\mathbb{R}^{p-1}$ *such that*

(ii) $\int_{\mathbb{R}^{p-1}} \mathbf{w}K(\mathbf{w})\,\mathrm{d}\mathbf{w} = \mathbf{0}$ *and* $\boldsymbol{\mu}_2^K := \int_{\mathbb{R}^{p-1}} \mathbf{w}\mathbf{w}'K(\mathbf{w})\,\mathrm{d}\mathbf{w}$ *is positive definite.*

Assumption (A3). *The bandwidth* h_n *is such that* $\lim_{n\to\infty} h_n = 0$ *and* $\lim_{n\to\infty} nh_n^{p-1} = \infty$.

The conditions we are imposing in Assumption (A1) are quite mild. For example, Assumption (A1)(ii) is the same as Condition (A)(iii) in [11] and Assumption (A1)(i) in [17]; Assumption (A1)(iii)–(v) are similar to Condition (A)(i, iv) in [11] and Condition (A1)(ii) in [17], where the existence and positive-definiteness ensure the invertibility of $\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0}$ in Theorem 5.1.

Assumptions (A2) and (A3) on the kernel function and the bandwidth also are quite standard in the nonparametric literature. For example, any compactly supported symmetric density function satisfies Assumption (A2). The compact support of K in Assumption (A2) is only a technical assumption to simplify the proof of theorems. In practice, Gaussian kernels can be considered; indeed, at the cost of more involved proof, the compact support assumption in Theorems 5.1 and 5.2 can be replaced with the assumption that both $C_0^K := \int_{\mathbb{R}^{p-1}} K^2(\mathbf{w})\,\mathrm{d}\mathbf{w}$ and $C_2^K := \int_{\mathbb{R}^{p-1}} \mathbf{w}\mathbf{w}' K^2(\mathbf{w})\,\mathrm{d}\mathbf{w}$ are finite. As for Assumption (A3), it is the usual one in the i.i.d. setting; see Section 6 for a discussion.

Let $\boldsymbol{\mathcal{X}}^c_{\mathbf{u}} := (1, \mathbf{Y}^{\perp\prime}_{\mathbf{u}})'$ and $\boldsymbol{\mathcal{X}}^\ell_{\mathbf{u}} := (1, \mathbf{Y}^{\perp\prime}_{\mathbf{u}})' \otimes (1, (\mathbf{W} - \mathbf{w}_0)')'$, where the superscript c and ℓ stand for the local constant and local bilinear cases, respectively. For $(\mathbf{W}, \mathbf{Y}) = (\mathbf{W}_i, \mathbf{Y}_i)$, we use the notation $Y_{i\mathbf{u}}$, $\mathbf{Y}^\perp_{i\mathbf{u}}$, $\boldsymbol{\mathcal{X}}^c_{i\mathbf{u}}$, $\boldsymbol{\mathcal{X}}^\ell_{i\mathbf{u}}$, etc. in an obvious way.

Referring to (4.2) for the notation, the parameter of interest for the local constant case is $\boldsymbol{\theta}^c = \boldsymbol{\theta}^c_{\boldsymbol{\tau};\mathbf{w}_0} := (a_{\boldsymbol{\tau};\mathbf{w}_0}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0})'$, whereas, in the local bilinear case (see (4.3)), we rather have to estimate

$$\boldsymbol{\theta}^\ell = \boldsymbol{\theta}^\ell_{\boldsymbol{\tau};\mathbf{w}_0} := \operatorname{vec}\begin{pmatrix} a_{\boldsymbol{\tau};\mathbf{w}_0} & \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0} \\ \dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0} & \dot{\mathbf{c}}'_{\boldsymbol{\tau};\mathbf{w}_0} \end{pmatrix}. \tag{5.1}$$

The local constant and local bilinear methods described in the previous sections provide estimators of the form $\hat{\boldsymbol{\theta}}^{c(n)} := (\hat{a}, \hat{\mathbf{c}}')'$ and

$$\hat{\boldsymbol{\theta}}^{\ell(n)} := \operatorname{vec}\begin{pmatrix} \hat{a} & \hat{\mathbf{c}}' \\ \hat{\dot{\mathbf{a}}} & \hat{\dot{\mathbf{c}}}' \end{pmatrix} \tag{5.2}$$

(we should actually discriminate between $(\hat{a}, \hat{\mathbf{c}}') = (\hat{a}^c, \hat{\mathbf{c}}^{c\prime})$ and $(\hat{a}, \hat{\mathbf{c}}') = (\hat{a}^\ell, \hat{\mathbf{c}}^{\ell\prime})$, but will not do so in order to avoid making the notation too heavy); those estimators are defined as the corresponding minimizer $\boldsymbol{\theta}^r$ of

$$\sum_{i=1}^n K_h(\mathbf{W}_i - \mathbf{w}_0)\rho_\tau\big(Y_{i\mathbf{u}} - \boldsymbol{\theta}^{r\prime}\boldsymbol{\mathcal{X}}^r_{i\mathbf{u}}\big), \qquad r = c, \ell. \tag{5.3}$$

The following result provides Bahadur representations for $\hat{\boldsymbol{\theta}}^{c(n)}$ and $\hat{\boldsymbol{\theta}}^{\ell(n)}$.

Theorem 5.1 (Bahadur representations). *Let Assumptions* (A1), (A2)(i) *and* (A3) *hold, assume that* $\mathbf{w} \mapsto (a_{\boldsymbol{\tau};\mathbf{w}}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}})'$ *is continuously differentiable at* $\mathbf{w}_0$, *and write* $\psi_\tau(y) := \tau - I[y < 0]$. *Then, as* $n \to \infty$,

$$\begin{aligned}&\sqrt{nh_n^{p-1}}\mathbf{M}^r_h\big(\hat{\boldsymbol{\theta}}^{r(n)} - \boldsymbol{\theta}^r\big)\\ &\quad= \frac{\boldsymbol{\eta}^r_{\boldsymbol{\tau};\mathbf{w}_0}}{\sqrt{nh_n^{p-1}}}\sum_{i=1}^n K\left(\frac{\mathbf{W}_i - \mathbf{w}_0}{h_n}\right)\psi_\tau\big(Z^r_{i\mathbf{u}}(\boldsymbol{\theta})\big)\big(\mathbf{M}^r_h\big)^{-1}\boldsymbol{\mathcal{X}}^r_{i\mathbf{u}} + o_{\mathrm{P}}(1),\end{aligned} \tag{5.4}$$

where $Z^r_{i\mathbf{u}}(\boldsymbol{\vartheta}) := Y_{i\mathbf{u}} - \boldsymbol{\vartheta}'\boldsymbol{\mathcal{X}}^r_{i\mathbf{u}}$ $(r = c, \ell)$, $\mathbf{M}^c_h := \mathbf{I}_m$, $\mathbf{M}^\ell_h := \mathbf{I}_m \otimes \mathrm{diag}(1, h_n\mathbf{I}_{p-1})$,

$$\boldsymbol{\eta}^c_{\boldsymbol{\tau};\mathbf{w}_0} := \big(f^{\mathbf{W}}(\mathbf{w}_0)\big)^{-1}\mathbf{G}^{-1}_{\boldsymbol{\tau};\mathbf{w}_0} \quad \textit{and} \quad \boldsymbol{\eta}^\ell_{\boldsymbol{\tau};\mathbf{w}_0} := \big(f^{\mathbf{W}}(\mathbf{w}_0)\big)^{-1}\mathbf{G}^{-1}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathrm{diag}\big(1, \big(\boldsymbol{\mu}^K_2\big)^{-1}\big),$$

with $\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0}$ *defined in Assumption* (A1)(v) (*the result for the local constant case does not require* (A2)(ii)).

This result, along with Assumption (A4) below, entails the asymptotic normality of $\hat{\boldsymbol{\theta}}^{r(n)}$, $r = c, \ell$. That assumption deals with the existence, at $\mathbf{w} = \mathbf{w}_0$, of the second derivatives of $\mathbf{w} \mapsto (a_{\boldsymbol{\tau};\mathbf{w}}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}})'$. With $\mathbf{c}_{\boldsymbol{\tau};\mathbf{w}} =: (c_{\boldsymbol{\tau};\mathbf{w},1}, \ldots, c_{\boldsymbol{\tau};\mathbf{w},m-1})'$, denote by $\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}}$ and $\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},j}$ the $(p-1) \times 1$ vectors of first derivatives and by $\ddot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}}$ and $\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},j}$ the $(p-1) \times (p-1)$ matrices of second derivatives (when they exist) of $\mathbf{w} \mapsto a_{\boldsymbol{\tau};\mathbf{w}}$ and $\mathbf{w} \mapsto c_{\boldsymbol{\tau};\mathbf{w},j}$, respectively (recall that $\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}}$ and $\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}} = (\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},1}, \ldots, \dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},m-1})'$ were already defined in page 1449). Finally, write $\ddot{\mathbf{c}}'_{\boldsymbol{\tau};\mathbf{w}}$ for the $(p-1) \times (m-1)(p-1)$ matrix $(\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},1}, \ldots, \ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},m-1})$.

Assumption (A4).

(i) *The function* $\mathbf{w} \mapsto (a_{\boldsymbol{\tau};\mathbf{w}}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}})'$ *is twice continuously differentiable at* $\mathbf{w} = \mathbf{w}_0$, *that is,* $\ddot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}}$ *and* $\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}}$ *exist in a neighborhood of* $\mathbf{w}_0$ *and are continuous with respect to* $\mathbf{w}$ *at* $\mathbf{w}_0$.

(ii) *The function* $\mathbf{w} \mapsto f^{\mathbf{W}}(\mathbf{w})$ *is continuously differentiable at* $\mathbf{w} = \mathbf{w}_0$, *that is, the* $(p-1) \times 1$ *vector of first derivatives of* $f^{\mathbf{W}}$, $\dot{f}^{\mathbf{W}}(\mathbf{w})$, *exists in a neighborhood of* $\mathbf{w}_0$ *and is continuous with respect to* $\mathbf{w}$ *at* $\mathbf{w}_0$.

The following matrices are involved in the asymptotic bias and variance expressions of the asymptotic normality result in Theorem 5.2 below. Define

$$\boldsymbol{\Sigma}^c_{\mathbf{w}} := \tau(1-\tau)f^{\mathbf{W}}(\mathbf{w})C^K_0\boldsymbol{\eta}^c_{\boldsymbol{\tau};\mathbf{w}}\bigg[\int_{\mathbb{R}^{m-1}} f^{\mathbf{Y}^\perp_{\mathbf{u}}|\mathbf{W}=\mathbf{w}}(\mathbf{t})\begin{pmatrix} 1 & \mathbf{t}' \\ \mathbf{t} & \mathbf{t}\mathbf{t}' \end{pmatrix} \mathrm{d}\mathbf{t}\bigg]\boldsymbol{\eta}^c_{\boldsymbol{\tau};\mathbf{w}}, \tag{5.5}$$

$$\begin{aligned}\boldsymbol{\Sigma}^\ell_{\mathbf{w}} &:= \tau(1-\tau)f^{\mathbf{W}}(\mathbf{w})\boldsymbol{\eta}^\ell_{\boldsymbol{\tau};\mathbf{w}} \\ &\quad \times \bigg[\int_{\mathbb{R}^{m-1}} f^{\mathbf{Y}^\perp_{\mathbf{u}}|\mathbf{W}=\mathbf{w}}(\mathbf{t})\begin{pmatrix} 1 & \mathbf{t}' \\ \mathbf{t} & \mathbf{t}\mathbf{t}' \end{pmatrix} \mathrm{d}\mathbf{t} \otimes \mathrm{diag}\big(C^K_0, C^K_2\big)\bigg]\boldsymbol{\eta}^\ell_{\boldsymbol{\tau};\mathbf{w}},\end{aligned} \tag{5.6}$$

and, for $r = c, \ell$,

$$\begin{aligned}\mathbf{B}^r_{\mathbf{w}} &:= f^{\mathbf{W}}(\mathbf{w})\boldsymbol{\eta}^r_{\boldsymbol{\tau};\mathbf{w}} \\ &\quad \times \int_{\mathbb{R}^{m-1}} f^{Y_{\mathbf{u}}|\mathbf{Y}^\perp_{\mathbf{u}}=\mathbf{t},\mathbf{W}=\mathbf{w}}\big(a_{\boldsymbol{\tau};\mathbf{w}} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}}\mathbf{t}\big) f^{\mathbf{Y}^\perp_{\mathbf{u}}|\mathbf{W}=\mathbf{w}}(\mathbf{t})\begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix} \otimes \bigg[\mathbf{B}^r_{\mathbf{w};0}\begin{pmatrix} 1 \\ \mathbf{t} \end{pmatrix}\bigg] \mathrm{d}\mathbf{t},\end{aligned} \tag{5.7}$$

where (putting $\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},0} := \ddot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}}$) $\mathbf{B}^c_{\mathbf{w};0}$ is the $1 \times m$ matrix with jth entry

$$B^c_{\mathbf{w};0,j} := \mathrm{tr}\bigg[\bigg(\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},j-1} + 2\frac{\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},j-1}(\dot{f}^{\mathbf{W}}(\mathbf{w}))'}{f^{\mathbf{W}}(\mathbf{w})}\bigg)\boldsymbol{\mu}^K_2\bigg], \qquad j = 1, \ldots, m,$$

and $\mathbf{B}^{\ell}_{\mathbf{w};0}$ denotes the $p \times m$ matrix with (i, j)th entry

$$B^{\ell}_{\mathbf{w};0,ij} := \mathrm{tr}\left[\ddot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w},j-1} \int_{\mathbb{R}^{p-1}} w_{i-1}\mathbf{w}\mathbf{w}' K(\mathbf{w})\,\mathrm{d}\mathbf{w}\right], \qquad i = 1, \dots, p,\ j = 1, \dots, m;$$

here, we wrote $\mathbf{w} = (w_1, w_2, \dots, w_{p-1})'$, $w_0 = 1$. We then have:

Theorem 5.2 (Asymptotic normality). *Let Assumptions (A1)–(A4) hold. Then, for $r = c, \ell$,*

$$\sqrt{nh_n^{p-1}}\mathbf{M}^r_h\left(\hat{\boldsymbol{\theta}}^{r(n)} - \boldsymbol{\theta}^r - \frac{h^2}{2}\mathbf{B}^r_{\mathbf{w}_0}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}^r_{\mathbf{w}_0}\big), \tag{5.8}$$

as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution (the result for the local bilinear case does not require Assumption (A4)(ii)).

Remark 5.1. The local bilinear fitting has an expression of bias that is independent of $\dot{f}^{\mathbf{W}}$. In contrast, the local constant fitting has a large bias at the regions where the derivative of $f^{\mathbf{W}}$ is large, that is, it does not adapt to highly-skewed designs (see [10,12]). Another important advantage of local bilinear fitting over the local constant approach is its much better boundary behavior. This advantage often has been emphasized in the usual regression settings when the regressors take values on a compact subset of $\mathbb{R}^{p-1}$. For example, considering a univariate random regressor W $(p = 2)$ with bounded support ([0, 1], say), it can be proved, using an argument similar to the one developed in the corresponding proof in [10], that asymptotic normality (with the same rate) still holds at boundary points of the form ch_n, where $c \in \mathbb{R}^+_0$, with asymptotic bias and variances of the same form as in the local bilinear ($r = \ell$) versions of (5.7) and (5.6), with $p = 2$, $\mathbf{w}_0$ replaced by $w_0 = 0^+$, and $\int_{\mathbb{R}^{p-1}}$ by $\int_{-c}^{\infty}$; see, for example, page 666 of [17].

Remark 5.2. In practice, we may be concerned with the estimation of the quantile regression functions at different $\boldsymbol{\tau}$'s simultaneously. Restricting to the estimation of $(\boldsymbol{\theta}'_{\boldsymbol{\tau}_1;\mathbf{w}_0}, \boldsymbol{\theta}'_{\boldsymbol{\tau}_2;\mathbf{w}_0})'$, it can be shown by proceeding as in the proof of Theorem 5.2 that $(\hat{\boldsymbol{\theta}}'_{\boldsymbol{\tau}_1;\mathbf{w}_0}, \hat{\boldsymbol{\theta}}'_{\boldsymbol{\tau}_2;\mathbf{w}_0})'$ is asymptotically normal with a block-diagonal asymptotic covariance matrix, that is, $\hat{\boldsymbol{\theta}}_{\boldsymbol{\tau}_1;\mathbf{w}_0}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\tau}_2;\mathbf{w}_0}$ are asymptotically independent for $\boldsymbol{\tau}_1 \neq \boldsymbol{\tau}_2$.

6. Bandwidth selection

While the choice of a kernel, as usual, has little impact on the final result, selecting the bandwidth h is more delicate. A full plug-in estimator in principle could be derived from the asymptotic normality result of Theorem 5.2, along the same lines as, for instance, in Zhang and Lee [46], who do it for mean regression. Such an approach, however, requires the estimation of several conditional densities, hence raises further problems, besides being computationally quite heavy, certainly when several values of τ are to be considered. A simpler heuristic rule is thus preferable; the one we are describing here is adapted from [45], where it is proposed in the context of single-output quantile regression.

Without loss of generality, we restrict to $p-1=1$ for notational simplicity, writing W and w for $\mathbf{W}$ and $\mathbf{w}$, h for h_n and $\hat{\boldsymbol{\theta}}_h = (\hat{a}^h_{\boldsymbol{\tau};w_0}, \hat{\mathbf{c}}^{h\prime}_{\boldsymbol{\tau};w_0})'$ for the estimator of $\boldsymbol{\theta} = (a_{\boldsymbol{\tau};w_0}, \mathbf{c}'_{\boldsymbol{\tau};w_0})'$ associated with bandwidth h, respectively. Throughout, the kernel K is some symmetric density function, such as the standard normal one. The objective is to minimize, with respect to h, the asymptotic mean square error which, in view of Theorem 5.2 with $p-1=1$, after some straightforward algebra takes the form

$$MSE(h) = \mathrm{E}(\hat{\boldsymbol{\theta}}_h - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}_h - \boldsymbol{\theta}) \approx \frac{1}{4} h^4 B^2_{\boldsymbol{\tau}} + \frac{1}{nh} V_{\boldsymbol{\tau}}, \tag{6.1}$$

with

$$B^2_{\boldsymbol{\tau}} := \big(\mu^K_2\big)^2 \left(\ddot{a}^2_{\boldsymbol{\tau};w_0} + \sum_{j=1}^{m-1} \ddot{\mathbf{c}}^2_{\boldsymbol{\tau};w_0,j} \right) \quad \text{and} \quad V_{\boldsymbol{\tau}} := \frac{\tau(1-\tau)C^K_0}{f^W(w_0)} \operatorname{tr}\big(\mathbf{G}^{-1}_{\boldsymbol{\tau};w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau};w_0}\big),$$

where $\ddot{\mathbf{c}}_{\boldsymbol{\tau};w_0,j}$ is the second-order derivative with respect to w_0 of the jth component of $\mathbf{c}_{\boldsymbol{\tau};w_0}$, $\mathbf{G}_{\boldsymbol{\tau};w_0}$ is defined in Assumption (A1)(v), and

$$\mathbf{G}_{w_0} := \int_{\mathbb{R}^{m-1}} f^{\mathbf{Y}^\perp_{\mathbf{u}}|W=w_0}(\mathbf{t}) \begin{pmatrix} 1 & \mathbf{t}' \\ \mathbf{t} & \mathbf{t}\mathbf{t}' \end{pmatrix} \mathrm{d}\mathbf{t}.$$

The minimizer $h_{\boldsymbol{\tau}}$ of (6.1) satisfies

$$h^5_{\boldsymbol{\tau}} = \frac{V_{\boldsymbol{\tau}}}{nB^2_{\boldsymbol{\tau}}} = \frac{\tau(1-\tau)C^K_0 \operatorname{tr}[\mathbf{G}^{-1}_{\boldsymbol{\tau};w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau};w_0}]}{n(\mu^K_2)^2 f^W(w_0)(\ddot{a}^2_{\boldsymbol{\tau};w_0} + \sum_{j=1}^{m-1} \ddot{\mathbf{c}}^2_{\boldsymbol{\tau},w_0,j})}, \tag{6.2}$$

so that for any $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$,

$$\left(\frac{h_{\boldsymbol{\tau}_1}}{h_{\boldsymbol{\tau}_2}}\right)^5 = \frac{\tau_1(1-\tau_1)}{\tau_2(1-\tau_2)} \frac{(\ddot{a}^2_{\boldsymbol{\tau}_2,w_0} + \sum_{j=1}^{m-1} \ddot{\mathbf{c}}^2_{\boldsymbol{\tau}_2,w_0,j}) \operatorname{tr}(\mathbf{G}^{-1}_{\boldsymbol{\tau}_1,w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau}_1,w_0})}{(\ddot{a}^2_{\boldsymbol{\tau}_1,w_0} + \sum_{j=1}^{m-1} \ddot{\mathbf{c}}^2_{\boldsymbol{\tau}_1,w_0,j}) \operatorname{tr}(\mathbf{G}^{-1}_{\boldsymbol{\tau}_2,w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau}_2,w_0})}. \tag{6.3}$$

As in [45], we assume that $\ddot{a}_{\tau\mathbf{u},w_0}$ and $\ddot{\mathbf{c}}_{\tau\mathbf{u},w_0}$ do not depend on τ (an assumption we do not make on $a_{\tau\mathbf{u},w_0}$ and $\mathbf{c}_{\tau\mathbf{u},w_0}$). If $f^{Y_{\mathbf{u}}|\mathbf{Y}^\perp_{\mathbf{u}}=\mathbf{t},W=w_0}$ were a normal density with mean $\mu_{\mathbf{t},w_0}$ and variance $\sigma^2_{\mathbf{t},w_0}$, denoting by ϕ and Φ the standard normal density and distribution functions, respectively, we would have $f^{Y_{\mathbf{u}}|\mathbf{Y}^\perp_{\mathbf{u}}=\mathbf{t},W=w_0}(a_{\boldsymbol{\tau};w_0} + \mathbf{c}'_{\boldsymbol{\tau};w_0}\mathbf{t}) = \sigma^{-1}_{\mathbf{t},w_0}\phi(\Phi^{-1}(\tau))$, hence

$$\mathbf{G}_{\boldsymbol{\tau};w_0} = \phi\big(\Phi^{-1}(1/2)\big) \int_{\mathbb{R}^{m-1}} \sigma^{-1}_{\mathbf{t},w_0} f^{\mathbf{Y}^\perp_{\mathbf{u}}|W=w_0}(\mathbf{t}) \begin{pmatrix} 1 & \mathbf{t}' \\ \mathbf{t} & \mathbf{t}\mathbf{t}' \end{pmatrix} \mathrm{d}\mathbf{t}$$

and

$$\frac{\operatorname{tr}(\mathbf{G}^{-1}_{\boldsymbol{\tau}_1,w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau}_1,w_0})}{\operatorname{tr}(\mathbf{G}^{-1}_{\boldsymbol{\tau}_2,w_0} \mathbf{G}_{w_0} \mathbf{G}^{-1}_{\boldsymbol{\tau}_2,w_0})} = \left[\frac{\phi(\Phi^{-1}(\tau_2))}{\phi(\Phi^{-1}(\tau_1))}\right]^2.$$

If we further assume that $\sigma^2_{\mathbf{t},w_0} = \sigma^2_{w_0}$, (6.2) for $\tau = 1/2$ takes the form

$$h^5_{\mathbf{u}/2} = \frac{\pi}{2}\left(\frac{C^K_0}{n(\mu^K_2)^2}\frac{\mathrm{tr}(\mathbf{G}^{-1}_{w_0})\sigma^2_{w_0}}{f^W(w_0)(\ddot{a}^2_{\mathbf{u}/2;w_0} + \sum_{j=1}^{m-1}\ddot{\mathbf{c}}^2_{\mathbf{u}/2;w_0,j})}\right), \tag{6.4}$$

while (6.3) yields

$$\left(\frac{h_{\boldsymbol{\tau}_1}}{h_{\boldsymbol{\tau}_2}}\right)^5 = \frac{\tau_1(1-\tau_1)}{\tau_2(1-\tau_2)}\frac{(\phi(\Phi^{-1}(\tau_2)))^2}{(\phi(\Phi^{-1}(\tau_1)))^2} \tag{6.5}$$

hence, for $\boldsymbol{\tau}_2 = \mathbf{u}/2$, $h^5_{\boldsymbol{\tau}} = (2/\pi)\tau(1-\tau)(\phi(\Phi^{-1}(\tau)))^{-2}h^5_{\mathbf{u}/2}$.

This latter expression still is not readily implementable. However, (6.4) bears a strong relation to the optimal bandwidth value h_{FZ} obtained by Fan and Zhang in Theorem 1 of [13] for the estimation of the conditional mean in the varying-coefficient linear regression model $Y_{\mathbf{u}} = a(W) + \mathbf{c}(W)'\mathbf{Y}^{\perp}_{\mathbf{u}} + \epsilon_{\mathbf{u}}$ with $\mathrm{Var}(\epsilon_{\mathbf{u}} \mid W = w_0) = \sigma^2_{w_0}$, namely

$$h^5_{\mathrm{FZ}} = \frac{C^K_0}{n(\mu^K_2)^2}\frac{\mathrm{tr}(\mathbf{G}^{-1}_{w_0})\sigma^2_{w_0}}{f^W(w_0)(\ddot{a}^2_{w_0} + \sum_{j=1}^{m-1}\ddot{\mathbf{c}}^2_{w_0,j})} = (2/\pi)h^5_{\mathbf{u}/2}.$$

We therefore propose, for $\boldsymbol{\tau} = \tau\mathbf{u}$, the bandwidth $h_{\boldsymbol{\tau}}$ provided by

$$h^5_{\boldsymbol{\tau}} = \tau(1-\tau)\big(\phi\big(\Phi^{-1}(\tau)\big)\big)^{-2}h^5_{\mathrm{FZ}}, \tag{6.6}$$

where, for the selection of h_{FZ}, we may rely, for instance, on the plug-in rule developed by [46].

This rule (6.6) can be regarded as the combination of a plug-in strategy and a rule-of-thumb: plug-in strategy in the selection of h_{FZ} but rule-of-thumb for the dependence on τ. It furthermore implies that the selected $h_{\boldsymbol{\tau}}$ has the same $n^{-1/7}$ rate of convergence as h_{FZ} (see [46]).

7. A real data example

In order to illustrate the data-analytic power of the proposed method, we consider the “body girth measurement” dataset from [20], that was already investigated in HPŠ. The dataset consists of joint measurements of nine skeletal and twelve body girth dimensions, along with weight, height, and age, in a group of 247 young men and 260 young women. As in HPŠ, we discard the male observations, we restrict to the calf maximum girth (Y_1) and the thigh maximum girth (Y_2) for the response, and use a single random regressor W (weight, height, age, or BMI). Figures 5 and 6 provide cuts – for the same w- and τ-values as in HPŠ – obtained from the proposed local constant and local bilinear approaches, respectively.

These cuts confirm most of the global analysis conducted in HPŠ and moreover reveal some interesting new features. For instance,

(a) for the dependence on weight, the local bilinear approach confirms the positive trend in location, the increase in dispersion, and the evolution of “principal directions” (as weight increases, the first “principal direction” rotates from horizontal to vertical), and it further

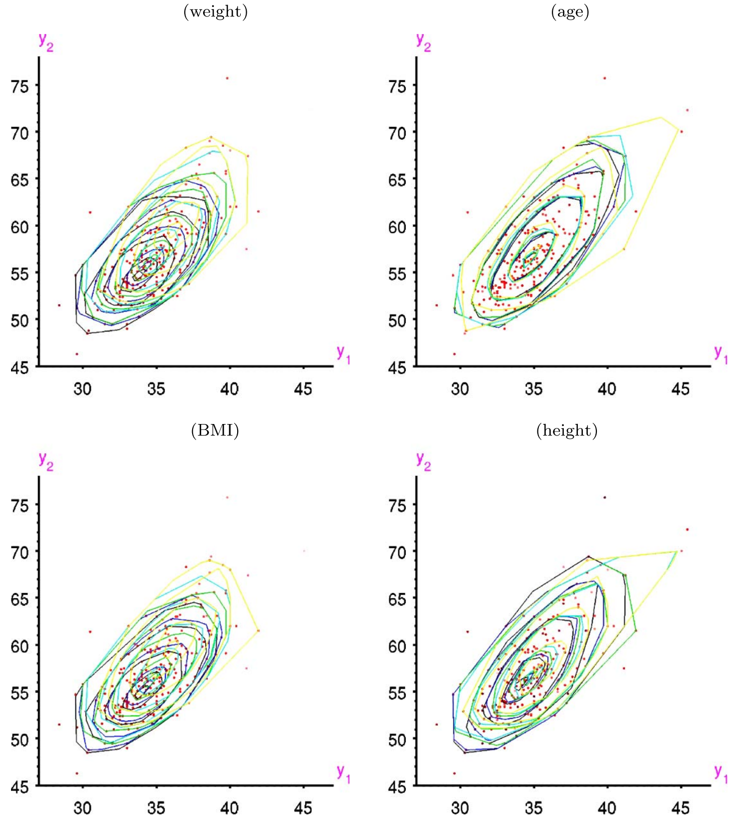

Figure 5. Four empirical (local constant) regression quantile plots from the body girth measurements dataset (women subsample; see [20]). Throughout, the bivariate response $(Y_1, Y_2)'$ involves calf maximum girth (Y_1) and thigh maximum girth (Y_2), while a single random regressor is used: weight, age, BMI, or height. The plots are providing, for $\tau = 0.01$, 0.03, 0.10, 0.25, and 0.40, the cuts of the local constant regression τ-quantile contours, at the empirical p-quantiles of the regressors, for $p = 0.10$ (black), 0.30 (blue), 0.50 (green), 0.70 (cyan) and 0.90 (yellow). The $n = 260$ data points are shown in red (the lighter the red color, the higher the regressor value). The results are based on a Gaussian kernel and the bandwidth $H = 3\sigma_w n^{-1/5}$, where σ_w stands for the empirical standard deviation of the regressor (the corresponding cuts obtained from linear regression are provided in Figure 7 of HPŠ). A color version of this figure is more readable, and can be found in the on-line edition of the paper.

indicates that high weights give rise to simultaneously large extreme values in Y_1 and Y_2. The differences, for low and high values of the covariate (weight), between the contours resulting from the local bilinear and local constant approaches illustrate the sensitivity of the latter to boundary effect;

(b) for the dependence on age, the local regression quantile regions, parallel to their global HPŠ counterparts, do indicate that the location and the first principal direction (along the main bisector) are constant over age. Still as in HPŠ, the local approaches confirm that

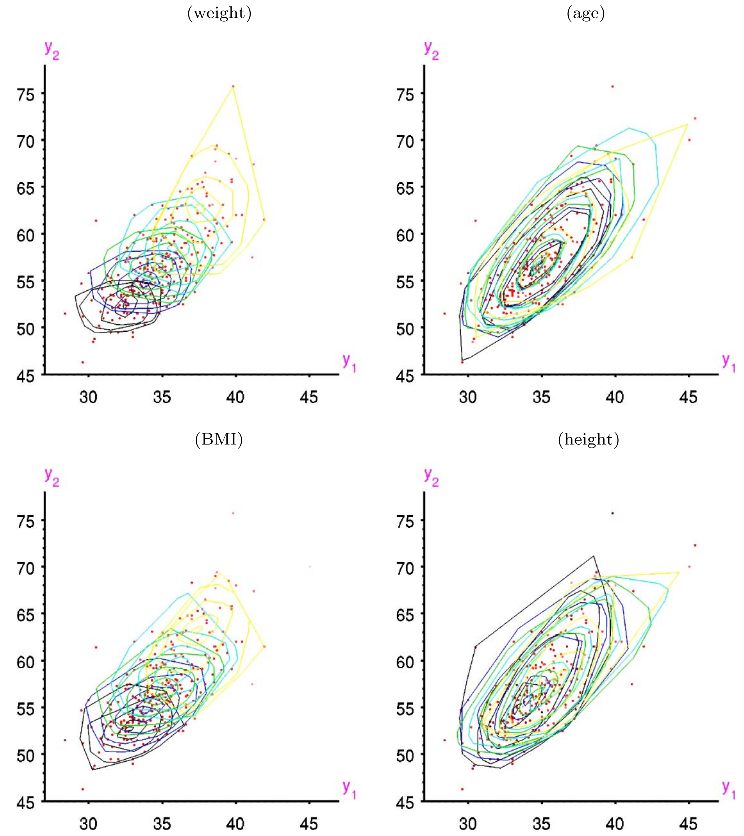

Figure 6. Same quantities as in Figure 5, here obtained from the local bilinear approach, with the same kernel and bandwidth as in Figure 5 (the computation was based on 360 equispaced directions $\mathbf{u} \in \mathcal{S}^1$). A color version of this figure is more readable, and can be found in the on-line edition of the paper.

the shapes of outer contours vary quite significantly with age, indicating an increasing (with age) simultaneous variability of both calf and thigh girth largest values. Now, compared to HPŠ, the local bilinear approach further shows that *young* women present a large simultaneous variability of both calf and thigh girth *smallest* values;

(c) for the dependence on height, the local methods confirm the regression effect specific to inner contours. The local bilinear approach further shows that there is also a regression effect for outer contours that, as height increases, get more widespread in the direction **u** (corresponding to simultaneously large values of both responses).

Limited as it is, this short application demonstrates how the local quantile regression analysis proposed here complements and refines the findings obtained from the global approach introduced in HPŠ by revealing the possible non-linear, heteroskedastic, skewness . . . features of the distributions of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}$. We refer to [32] for a further application, in the context of bivariate growth charts.

We conclude this section with a brief discussion of the computational aspects of the proposed methods. In principle, any quantile regression/linear programming/convex optimization solver can be used for that purpose. The exact local constant quantile/depth contours can be computed for any w_0 via a weighted version of the HPŠ algorithm – see Paindaveine and Šiman [34] for a detailed description of its Matlab implementation and its computation cost. The local bilinear contours, for given w_0, are determined by considering a fixed number M of directions; their computation then is as demanding as M times the standard simple-output quantile regression with the same number of regressors; see Koenker [26] for computational and algorithmic details.

8. Conclusion

In this paper, we propose a definition of regression depth as the conditional depth of an m-dimensional response conditional on a p-dimensional covariate. We also propose local constant and local bilinear methods for the estimation of conditional depth contours, and establish the consistency and asymptotic normality of the estimators. As a descriptive tool, the resulting contours provide a powerful data-analytic tool, while our asymptotic results guarantee that, for n large enough, those contours are able to detect any covariate-dependent feature of the conditional distributions of the response. An important domain of application for such methods is in the analysis of multiple output growth charts, where current practice is essentially restricted to a marginal approach that neglects all information related to *joint* conditional features.

Appendix: Proofs of asymptotic results

We actually restrict to the local bilinear case (proofs for the local constant case are entirely similar). The proofs rely on several lemmas, and require some further notation.

Referring to (5.1) and (5.2), define

$$\boldsymbol{\theta}^{\ell} = \operatorname{vec}\begin{pmatrix} a_{\boldsymbol{\tau};\mathbf{w}_0} & \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0} \\ \dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0} & \dot{\mathbf{c}}'_{\boldsymbol{\tau};\mathbf{w}_0} \end{pmatrix} =: \operatorname{vec}\begin{pmatrix} \boldsymbol{\varpi}'_{\mathbf{w}_0} \\ \dot{\boldsymbol{\varpi}}'_{\mathbf{w}_0} \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\theta}}^{\ell(n)} = \operatorname{vec}\begin{pmatrix} \hat{a} & \hat{\mathbf{c}}' \\ \hat{\dot{\mathbf{a}}} & \hat{\dot{\mathbf{c}}}' \end{pmatrix} =: \operatorname{vec}\begin{pmatrix} \widehat{\boldsymbol{\varpi}}'_{\mathbf{w}_0} \\ \widehat{\dot{\boldsymbol{\varpi}}}'_{\mathbf{w}_0} \end{pmatrix}.$$

Denote by $\boldsymbol{\varpi}_1 = (a_1, \mathbf{c}'_1)'$ and $\widetilde{\boldsymbol{\varpi}}_1 = (\tilde{a}_1, \tilde{\mathbf{c}}'_1)'$ two arbitrary vectors of $\mathbb{R}^m$, by $\boldsymbol{\varpi}_2 = (\mathbf{a}_2, \mathbf{c}'_2)'$ and $\widetilde{\boldsymbol{\varpi}}_2 = (\tilde{\mathbf{a}}_2, \tilde{\mathbf{c}}'_2)'$ two arbitrary $m \times (p-1)$ matrices. Let $H_n := \sqrt{nh_n^{p-1}}$ and put

$$\begin{aligned} \boldsymbol{\varphi}^{(n)} &:= H_n \mathbf{M}_h^{\ell} \operatorname{vec}\begin{pmatrix} (\widehat{\boldsymbol{\varpi}}_{\mathbf{w}_0} - \boldsymbol{\varpi}_{\mathbf{w}_0})' \\ (\widehat{\dot{\boldsymbol{\varpi}}}_{\mathbf{w}_0} - \dot{\boldsymbol{\varpi}}_{\mathbf{w}_0})' \end{pmatrix}, \\ \boldsymbol{\varphi} &:= H_n \mathbf{M}_h^{\ell} \operatorname{vec}\begin{pmatrix} (\boldsymbol{\varpi}_1 - \boldsymbol{\varpi}_{\mathbf{w}_0})' \\ (\boldsymbol{\varpi}_2 - \dot{\boldsymbol{\varpi}}_{\mathbf{w}_0})' \end{pmatrix}, \qquad \text{(A.1)} \\ \tilde{\boldsymbol{\varphi}} &:= H_n \mathbf{M}_h^{\ell} \operatorname{vec}\begin{pmatrix} (\widetilde{\boldsymbol{\varpi}}_1 - \boldsymbol{\varpi}_{\mathbf{w}_0})' \\ (\widetilde{\boldsymbol{\varpi}}_2 - \dot{\boldsymbol{\varpi}}_{\mathbf{w}_0})' \end{pmatrix}, \end{aligned}$$

and note that $\boldsymbol{\varphi}^{(n)} = \sqrt{nh_n^{p-1}}\mathbf{M}_h^\ell(\hat{\boldsymbol{\theta}}^{\ell(n)} - \boldsymbol{\theta}^\ell)$. Define $\mathbf{W}_{hi} := (\mathbf{W}_i - \mathbf{w}_0)/h_n$, $K_{hi} := K(\mathbf{W}_{hi})$ and $\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^\ell := (\mathbf{M}_h^\ell)^{-1}\boldsymbol{\mathcal{X}}_{i\mathbf{u}}^\ell = (1, \mathbf{Y}_{i\mathbf{u}}^{\perp\prime})' \otimes (1, \mathbf{W}_{hi}')'$.

Let $Z_{i\mathbf{u}}^\ell = Z_{i\mathbf{u}}^\ell(\boldsymbol{\theta}^\ell) := Y_{i\mathbf{u}} - \boldsymbol{\theta}^{\ell\prime}\boldsymbol{\mathcal{X}}_{i\mathbf{u}}^\ell$ as in Theorem 5.1, and define

$$T_{ni} := h_n\dot{\mathbf{a}}_{\boldsymbol{\tau};\mathbf{w}_0}'\mathbf{W}_{hi} + h_n(\operatorname{vec}\dot{\mathbf{c}}_{\boldsymbol{\tau};\mathbf{w}_0})'\big(\mathbf{Y}_{i\mathbf{u}}^\perp \otimes \mathbf{W}_{hi}\big),$$
$$Z_{ni}^*(\boldsymbol{\varphi}) := Z_{i\mathbf{u}}^\ell - H_n^{-1}\boldsymbol{\varphi}'\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^\ell \quad \text{and} \quad U_{ni} = U_{ni}(\boldsymbol{\varphi}) := T_{ni} + H_n^{-1}\boldsymbol{\varphi}'\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^\ell$$

(note that the latter two quantities depend on the choice of $\boldsymbol{\varpi}_1$ and $\boldsymbol{\varpi}_2$). The following identities will be useful in the sequel:

$$Z_{i\mathbf{u}}^\ell = Y_{i\mathbf{u}} - \big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}_{\boldsymbol{\tau};\mathbf{w}_0}'\mathbf{Y}_{i\mathbf{u}}^\perp\big) - T_{ni}, \tag{A.2}$$

$$\begin{aligned} Z_{ni}^*(\boldsymbol{\varphi}) &= Y_{i\mathbf{u}} - \big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}_{\boldsymbol{\tau};\mathbf{w}_0}'\mathbf{Y}_{i\mathbf{u}}^\perp\big) - U_{ni}(\boldsymbol{\varphi}) \\ &= Y_{i\mathbf{u}} - \big(\operatorname{vec}(\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2)'\big)'\boldsymbol{\mathcal{X}}_{i\mathbf{u}}^\ell. \end{aligned} \tag{A.3}$$

Let C be a generic constant whose value may vary from line to line. Since K is a bounded density with a bounded support, we have, whenever $K_{hi} > 0$,

$$\|\mathbf{W}_{hi}\| \le C \quad \text{and} \quad \big\|\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^\ell\big\| \le C\big(1 + \big\|\mathbf{Y}_{i\mathbf{u}}^\perp\big\|\big), \tag{A.4}$$

and, when moreover $\|\boldsymbol{\varphi}\| \le M$,

$$|T_{ni}| \le Ch_n\big(1 + \big\|\mathbf{Y}_{i\mathbf{u}}^\perp\big\|\big) \quad \text{and} \quad |U_{ni}| \le C\big(h_n + H_n^{-1}\big)\big(1 + \big\|\mathbf{Y}_{i\mathbf{u}}^\perp\big\|\big). \tag{A.5}$$

It follows from the definition of $\hat{\boldsymbol{\theta}}^{\ell(n)}$ as the argmin of (5.3) that

$$\boldsymbol{\varphi}^{(n)} = \underset{\boldsymbol{\varphi}\in\mathbb{R}^{mp}}{\arg\min} \sum_{i=1}^n K_{hi}\rho_\tau\big(Z_{ni}^*(\boldsymbol{\varphi})\big). \tag{A.6}$$

Recalling that $\psi_\tau(y) := \tau - I[y < 0]$, define

$$\mathbf{V}_n(\boldsymbol{\varphi}) := H_n^{-1}\sum_{i=1}^n K_{hi}\psi_\tau\big(Z_{ni}^*(\boldsymbol{\varphi})\big)\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^\ell. \tag{A.7}$$

In order to prove Theorem 5.1, we need the following lemma.

Lemma A.1. *Let $\mathbf{V}_n(\cdot) : \mathbb{R}^{mp} \to \mathbb{R}^{mp}$ be a sequence of functions that satisfies the following two properties*:

(i) *for all $\lambda \ge 1$ and all $\boldsymbol{\psi} \in \mathbb{R}^{mp}$, $-\boldsymbol{\psi}'\mathbf{V}_n(\lambda\boldsymbol{\psi}) \ge -\boldsymbol{\psi}'\mathbf{V}_n(\boldsymbol{\psi})$ a.s.*;
(ii) *there exist a $p \times p$ positive definite matrix $\mathbf{D}$ and a sequence of mp-dimensional random vectors $\mathbf{A}_n$ satisfying $\|\mathbf{A}_n\| = O_P(1)$ such that, for all $M > 0$, $\sup_{\|\boldsymbol{\psi}\|\le M}\|\mathbf{V}_n(\boldsymbol{\psi}) + (\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D})\boldsymbol{\psi} - \mathbf{A}_n\| = o_P(1)$, where $\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0}$ is given in Assumption (A1)(v).*

Then, if $\boldsymbol{\psi}_n$ *is such that* $\|\mathbf{V}_n(\boldsymbol{\psi}_n)\| = o_P(1)$, *it holds that* $\|\boldsymbol{\psi}_n\| = O_P(1)$ *and*

$$\boldsymbol{\psi}_n = (\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D})^{-1}\mathbf{A}_n + o_P(1). \tag{A.8}$$

Proof. The proof follows along the same lines as in page 809 of [27]; details are left to the reader. □

The proof of Theorem 5.1 consists in checking that the assumptions of Lemma A.1 hold for $\mathbf{V}_n$ defined in (A.7); we use the following lemma.

Lemma A.2. *Under Assumptions (A1)–(A3), for any* $(\boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}})$ *such that* $\max(\|\boldsymbol{\varphi}\|, \|\tilde{\boldsymbol{\varphi}}\|) \le M$, *and n large enough,*

$$\begin{aligned}\mathrm{E}\big[K_{hi}\big|\psi_\tau\big(Z^*_{ni}(\boldsymbol{\varphi})\big) - \psi_\tau\big(Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big)\big|\big] &\le C\mathrm{E}\big[K_{hi} I\big[\big|Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big| < CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big]\big] \\ &\le Ch_n^{p-1}H_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\end{aligned} \tag{A.9}$$

and

$$\begin{aligned}\mathrm{E}\big[K^2_{hi}\big|\psi_\tau\big(Z^*_{ni}(\boldsymbol{\varphi})\big) - \psi_\tau\big(Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big)\big|^2\big] &\le C\mathrm{E}\big[K^2_{hi} I\big[\big|Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big| < CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big]\big] \\ &\le Ch_n^{p-1}H_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|.\end{aligned} \tag{A.10}$$

Proof. The claim, in this lemma, is similar to that of Lemma A.3 in [17], which essentially follows from the same argument as in the time series case (cf. [31]). Details, however, are quite different. It follows from (A.4) that

$$\begin{aligned}K_{hi}\big|\psi_\tau\big(Z^*_{ni}(\boldsymbol{\varphi})\big) - \psi_\tau\big(Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big)\big| &= K_{hi}\big|I\big[Z^*_{ni}(\boldsymbol{\varphi}) < 0\big] - I\big[Z^*_{ni}(\tilde{\boldsymbol{\varphi}}) < 0\big]\big| \\ &= K_{hi}\big|I\big[Z^*_{ni}(\tilde{\boldsymbol{\varphi}}) < H_n^{-1}(\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}})'\boldsymbol{\mathcal{X}}^{\ell}_{hi\mathbf{u}}\big] - I\big[Z^*_{ni}(\tilde{\boldsymbol{\varphi}}) < 0\big]\big| \\ &\le K_{hi} I\big[\big|Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big| < CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)\big].\end{aligned}$$

Hence, from (A.3) and the mean value theorem, we obtain

$$\begin{aligned}&\mathrm{E}\big[K_{hi}\big|\psi_\tau\big(Z^*_{ni}(\boldsymbol{\varphi})\big) - \psi_\tau\big(Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big)\big|\big] \\ &\quad\le \mathrm{E}\big[K_{hi} I\big[\big|Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big| < CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)\big]\big] \\ &\quad= \mathrm{E}\big[K_{hi}\mathrm{P}\big[\big|Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big| < CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)|\mathbf{Y}^{\perp}_{i\mathbf{u}}, \mathbf{W}_i\big]\big] \\ &\quad= \mathrm{E}\big[K_{hi} F^{Y_\mathbf{u}|(\mathbf{Y}^{\perp}_\mathbf{u},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^{\perp}_{i\mathbf{u}} + U_{ni}(\tilde{\boldsymbol{\varphi}}) + CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)\big)\big] \\ &\qquad - \mathrm{E}\big[K_{hi} F^{Y_\mathbf{u}|(\mathbf{Y}^{\perp}_\mathbf{u},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^{\perp}_{i\mathbf{u}} + U_{ni}(\tilde{\boldsymbol{\varphi}}) - CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)\big)\big] \\ &\quad\le \mathrm{E}\big[K_{hi}\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big) f^{Y_\mathbf{u}|(\mathbf{Y}^{\perp}_\mathbf{u},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^{\perp}_{i\mathbf{u}} + U_{ni}(\tilde{\boldsymbol{\varphi}}) \\ &\qquad\qquad + \lambda CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\big(1 + \big\|\mathbf{Y}^{\perp}_{i\mathbf{u}}\big\|\big)\big)\big] \\ &\qquad \times 2CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\end{aligned}$$

for some $\lambda \in (-1, 1)$. Assumptions (A1)–(A3), together with (A.5), therefore yield that, for $\boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}} \in \{\boldsymbol{\varphi} : \|\boldsymbol{\varphi}\| \le M\}$ and n large enough,

$$\begin{aligned}
&\mathrm{E}\big[K_{hi}\big|\psi_\tau\big(Z^*_{ni}(\boldsymbol{\varphi})\big) - \psi_\tau\big(Z^*_{ni}(\tilde{\boldsymbol{\varphi}})\big)\big|\big] \\
&\quad \le CH_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\|\mathrm{E}\bigg[K_{hi}\int_{\mathbb{R}^{m-1}}\big(1 + \|\mathbf{t}\|\big) f^{Y_{\mathbf{u}}|(\mathbf{Y}^\perp_{\mathbf{u}}=\mathbf{t},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{t}\big) f^{\mathbf{Y}^\perp_{\mathbf{u}}|\mathbf{W}}(\mathbf{t})\,\mathrm{d}\mathbf{t}\bigg] \\
&\quad = Ch_n^{p-1}H_n^{-1}\|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}\| f^{\mathbf{W}}(\mathbf{w}_0) \\
&\qquad \times \int_{\mathbb{R}^{m-1}}\big(1 + \|\mathbf{t}\|\big) f^{Y_{\mathbf{u}}|(\mathbf{Y}^\perp_{\mathbf{u}}=\mathbf{t},\mathbf{W}=\mathbf{w}_0)}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{t}\big) f^{\mathbf{Y}^\perp_{\mathbf{u}}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t},
\end{aligned}$$

which establishes (A.9); (A.10) follows along similar lines. □

Lemma A.3. *Under Assumptions (A1)–(A3), we have that, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\varphi}\|\le M}\big\|\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0}) - \mathrm{E}\big[\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0})\big]\big\| = o_{\mathrm{P}}(1). \tag{A.11}$$

Proof. The proof of this lemma is quite similar, in view of Lemma A.2, to that of Lemma A.4 in [17]. Details are therefore omitted. □

Lemma A.4. *Under Assumptions (A1)–(A3), we have that, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\varphi}\|\le M}\big\|\mathrm{E}\big[\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0})\big] + (\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D})\boldsymbol{\varphi}\big\| = o(1), \tag{A.12}$$

where $\mathbf{D} = f^{\mathbf{W}}(\mathbf{w}_0)\,\mathrm{diag}(1, \boldsymbol{\mu}_2^K)$.

Proof. Note that $\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0}) = H_n^{-1}\sum_{i=1}^n K_{hi}[\psi_\tau(Z^*_{ni}(\boldsymbol{\varphi})) - \psi_\tau(Z^\ell_{i\mathbf{u}})]\boldsymbol{\mathcal{X}}^\ell_{hi\mathbf{u}}$. It follows from (A.2) and (A.3) that

$$\begin{aligned}
\mathrm{E}\big[\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0})\big] &= nH_n^{-1}\mathrm{E}\big[K_{hi}\big(I\big[Z^\ell_{i\mathbf{u}} < 0\big] - I\big[Z^*_{ni}(\boldsymbol{\varphi}) < 0\big]\big)\boldsymbol{\mathcal{X}}^\ell_{hi\mathbf{u}}\big] \\
&= H_n h_n^{-(p-1)}\mathrm{E}\big[K_{hi}\big(F^{Y_{\mathbf{u}}|(\mathbf{Y}^\perp_{\mathbf{u}},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^\perp_{i\mathbf{u}} + T_{ni}\big) \\
&\qquad - F^{Y_{\mathbf{u}}|(\mathbf{Y}^\perp_{\mathbf{u}},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^\perp_{i\mathbf{u}} + U_{ni}\big)\big)\boldsymbol{\mathcal{X}}^\ell_{hi\mathbf{u}}\big].
\end{aligned}$$

Then, similar to the proof of Lemma A.2, by the mean value theorem, since $U_{ni} - T_{ni} = H_n^{-1}\boldsymbol{\mathcal{X}}^{\ell\prime}_{hi\mathbf{u}}\boldsymbol{\varphi}$, there exists $\xi \in (0, 1)$ such that

$$\begin{aligned}
&\sup_{\|\boldsymbol{\varphi}\|\le M}\big\|\mathrm{E}\big[\mathbf{V}_n(\boldsymbol{\varphi}) - \mathbf{V}_n(\mathbf{0})\big] + (\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D})\boldsymbol{\varphi}\big\| \\
&\quad = \sup_{\|\boldsymbol{\varphi}\|\le M}\big\|(\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D})\boldsymbol{\varphi}
\end{aligned}$$

$$
\begin{aligned}
&\quad - h_n^{-(p-1)} \mathrm{E}\big[K_{hi} f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp} + T_{ni} + \xi H_n^{-1} \boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\boldsymbol{\varphi}\big)\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell}\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\boldsymbol{\varphi}\big]\big\| \\
&= \sup_{\|\boldsymbol{\varphi}\| \le M} \big\|\big\{(\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D}) - h_n^{-(p-1)} \mathrm{E}\big[K_{hi} f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp}\big)\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell}\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\big]\big\}\boldsymbol{\varphi} \\
&\qquad\quad - h_n^{-(p-1)} \mathrm{E}\big[K_{hi}\big(f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp} + T_{ni} + \xi H_n^{-1} \boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\boldsymbol{\varphi}\big) \\
&\qquad\qquad\qquad - f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp}\big)\big)\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell}\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\boldsymbol{\varphi}\big]\big\| \\
&\le C\big\|(\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0} \otimes \mathbf{D}) - h_n^{-(p-1)} \mathrm{E}\big[K_{hi} f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp}\big)\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell}\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\big]\big\| \\
&\quad + C \sup_{\|\boldsymbol{\varphi}\| \le M} h_n^{-(p-1)} \mathrm{E}\big[K_{hi}\big|f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp} + T_{ni} + \xi H_n^{-1} \boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\boldsymbol{\varphi}\big) \\
&\qquad\qquad\qquad - f^{Y_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{i\mathbf{u}}^{\perp}\big)\big|\big\|\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell}\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell\prime}\big\|\big] = \mathrm{o}(1),
\end{aligned}
$$

where we used Assumptions (A1) and (A2), together with (A.5). □

Lemma A.5. *Let Assumptions (A2) and (A3) hold. Then the random vector $\boldsymbol{\varphi}^{(n)}$ defined in (A.1) satisfies $\|\mathbf{V}_n(\boldsymbol{\varphi}^{(n)})\| = \mathrm{o}_{\mathrm{P}}(1)$.*

Proof. The proof follows from a argument similar to that of Lemma A.2 on page 836 of [36]. □

Lemma A.6. *Under Assumptions (A1)–(A3), for any $\mathbf{d} \in \mathbb{R}^{mp}$,*

$$
\begin{aligned}
&\lim_{n\to\infty} \mathrm{E}\big[\big\{\mathbf{d}'\big(\mathbf{V}_n(\mathbf{0}) - \mathrm{E}\big[\mathbf{V}_n(\mathbf{0})\big]\big)\big\}^2\big] \\
&\quad = \tau(1-\tau) f^{\mathbf{W}}(\mathbf{w}_0) \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}^{m-1}} \big(\big[\big(1, \mathbf{t}'\big) \otimes \big(1, \mathbf{w}'\big)\big]\mathbf{d}\big)^2 f^{\mathbf{Y}_{\mathbf{u}}^{\perp}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t}) K^2(\mathbf{w})\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{w}.
\end{aligned}
$$

Proof. Set $\tilde{v}_i = K_{hi}\psi_\tau(Z_{i\mathbf{u}}^{\ell})\mathbf{d}'\boldsymbol{\mathcal{X}}_{hi\mathbf{u}}^{\ell} = K_{hi}\psi_\tau(Z_{i\mathbf{u}}^{\ell})[(1, \mathbf{Y}_{i\mathbf{u}}^{\perp\prime}) \otimes (1, \mathbf{W}_{hi}')]\mathbf{d}$. A simple calculation yields

$$
\mathrm{E}\big[\big\{\mathbf{d}'\big(\mathbf{V}_n(\mathbf{0}) - \mathrm{E}\big[\mathbf{V}_n(\mathbf{0})\big]\big)\big\}^2\big] = H_n^{-2} n \operatorname{Var}[\tilde{v}_1] = h_n^{-(p-1)} \operatorname{Var}[\tilde{v}_1]. \tag{A.13}
$$

Note that, for $k = 1, 2$,

$$
\begin{aligned}
&\lim_{n\to\infty} h_n^{-(p-1)} \mathrm{E}\big[K_{h1}^k I\big[Z_{1\mathbf{u}}^{\ell} < 0\big]\big(\mathbf{d}'\boldsymbol{\mathcal{X}}_{h1\mathbf{u}}^{\ell}\big)^k\big] \\
&\quad = \lim_{n\to\infty} h_n^{-(p-1)} \mathrm{E}\big[K_{h1}^k F^{\mathbf{Y}_{\mathbf{u}}|(\mathbf{Y}_{\mathbf{u}}^{\perp},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0} + \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}_{1\mathbf{u}}^{\perp} + T_{n1}\big)\big(\mathbf{d}'\boldsymbol{\mathcal{X}}_{h1\mathbf{u}}^{\ell}\big)^k\big] \\
&\quad = \tau f^{\mathbf{W}}(\mathbf{w}_0) \int_{\mathbb{R}^{p-1}} \int_{\mathbb{R}^{m-1}} K^k(\mathbf{w})\big(\big[\big(1, \mathbf{t}'\big) \otimes \big(1, \mathbf{w}'\big)\big]\mathbf{d}\big)^k f^{\mathbf{Y}_{\mathbf{u}}^{\perp}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{w},
\end{aligned}
$$

which leads to

$$\begin{aligned}
&\lim_{n\to\infty} h_n^{-(p-1)}\mathrm{E}[\tilde{v}_1] \\
&\quad = \lim_{n\to\infty} h_n^{-(p-1)}\mathrm{E}\big[K_{h1}\big(\tau - I\big[Z^{\ell}_{1\mathbf{u}} < 0\big]\big)\big(\mathbf{d}'\boldsymbol{\mathcal{X}}^{\ell}_{h1\mathbf{u}}\big)\big] \\
&\quad = (\tau-\tau) f^{\mathbf{W}}(\mathbf{w}_0)\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}^{m-1}} K(\mathbf{w})\big(\big[\big(1,\mathbf{t}'\big)\otimes\big(1,\mathbf{w}'\big)\big]\mathbf{d}\big) f^{\mathbf{Y}^{\perp}_{\mathbf{u}}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{w} = 0
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{n\to\infty} h_n^{-(p-1)}\mathrm{E}\big[\tilde{v}_1^2\big] \\
&\quad = \lim_{n\to\infty} h_n^{-(p-1)}\mathrm{E}\big[K^2_{h1}\big(\tau^2 - 2\tau I\big[Z^{\ell}_{1\mathbf{u}} < 0\big] + I\big[Z^{\ell}_{1\mathbf{u}} < 0\big]\big)\big(\mathbf{d}'\boldsymbol{\mathcal{X}}^{\ell}_{h1\mathbf{u}}\big)^2\big] \\
&\quad = \tau(1-\tau) f^{\mathbf{W}}(\mathbf{w}_0)\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}^{m-1}} K^2(\mathbf{w})\big(\big[\big(1,\mathbf{t}'\big)\otimes\big(1,\mathbf{w}'\big)\big]\mathbf{d}\big)^2 f^{\mathbf{Y}^{\perp}_{\mathbf{u}}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{w}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} h_n^{-(p-1)}\,\mathrm{Var}[\tilde{v}_1] \\
&\quad = \lim_{n\to\infty} \big(h_n^{-(p-1)}\mathrm{E}\big[\tilde{v}_1^2\big] - h_n^{-(p-1)}\big(\mathrm{E}[\tilde{v}_1]\big)^2\big) \\
&\quad = \tau(1-\tau) f^{\mathbf{W}}(\mathbf{w}_0)\int_{\mathbb{R}^{p-1}}\int_{\mathbb{R}^{m-1}} K^2(\mathbf{w})\big(\big[\big(1,\mathbf{t}'\big)\otimes\big(1,\mathbf{w}'\big)\big]\mathbf{d}\big)^2 f^{\mathbf{Y}^{\perp}_{\mathbf{u}}|\mathbf{W}=\mathbf{w}_0}(\mathbf{t})\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{w},
\end{aligned}$$

which, together with (A.13), establishes the result. □

Proof of Theorem 5.1. The proof consists in checking that the conditions of Lemma A.1 are satisfied. Lemmas A.3 and A.4 entail that Lemma A.1(ii) holds, with $\mathbf{D} = f^{\mathbf{W}}(\mathbf{w}_0)\,\mathrm{diag}(1, \boldsymbol{\mu}_2^K)$ (yielding $(\mathbf{G}_{\boldsymbol{\tau};\mathbf{w}_0}\otimes\mathbf{D})^{-1} = \boldsymbol{\eta}^{\ell}_{\boldsymbol{\tau};\mathbf{w}_0}$) and $\mathbf{A}_n = \mathbf{V}_n(\mathbf{0}) = H_n^{-1}\sum_{i=1}^n K_{hi}\psi_\tau(Z^{\ell}_{i\mathbf{u}})\boldsymbol{\mathcal{X}}^{\ell}_{hi\mathbf{u}}$, which, by Lemma A.6, is $\mathrm{O_P}(1)$. As for Lemma A.1(ii), the fact that

$$\lambda \mapsto -\boldsymbol{\varphi}'\mathbf{V}_n(\lambda\boldsymbol{\varphi}) = H_n^{-1}\sum_{i=1}^n K_{hi}\psi_\tau\big(Z^{\ell}_{i\mathbf{u}} - \lambda H_n^{-1}\boldsymbol{\varphi}'\boldsymbol{\mathcal{X}}^{\ell}_{hi\mathbf{u}}\big)\big(-\boldsymbol{\varphi}'\boldsymbol{\mathcal{X}}^{\ell}_{hi\mathbf{u}}\big)$$

is non-decreasing directly follows from the fact $y\mapsto\psi_\tau(y)$ is non-decreasing. Since (Lemma A.5 and Assumptions (A2) and (A3)) $\|\mathbf{V}_n(\boldsymbol{\varphi}^{(n)})\|$ is $\mathrm{o_P}(1)$, Lemma A.1 applies, which concludes the proof. □

Proof of Theorem 5.2. On the basis of the Bahadur representation of Theorem 5.1, the asymptotic normality of $\hat{\boldsymbol{\theta}}^{\ell(n)}$ follows exactly as in the corresponding proofs for usual nonparametric regression in the i.i.d. case (see, e.g., [10]), yielding the asymptotic normality with the bias (i.e., the expectation) of the first term on the right-hand side of (5.4) as

$$
\begin{aligned}
&\mathrm{E}\left[\frac{\eta^{\ell}_{\boldsymbol{\tau};\mathbf{w}_0}}{\sqrt{nh_n^{p-1}}}\sum_{i=1}^{n} K_{h1}\psi_\tau\big(Z^{\ell}_{i\mathbf{u}}\big)\boldsymbol{\mathcal{X}}^{\ell}_{hi\mathbf{u}}\right] \\
&\quad= \frac{\eta^{\ell}_{\boldsymbol{\tau};\mathbf{w}_0}}{\sqrt{nh_n^{p-1}}} n\mathrm{E}\big[K_{h1}\psi_\tau\big(Z^{\ell}_{1\mathbf{u}}\big)\boldsymbol{\mathcal{X}}^{\ell}_{h1\mathbf{u}}\big] \\
&\quad= \eta^{\ell}_{\boldsymbol{\tau};\mathbf{w}_0}\sqrt{nh_n^{p-1}}\,h_n^{-(p-1)}\mathrm{E}\big[K_{h1}\big(F^{Y_{\mathbf{u}}|(\mathbf{Y}^{\perp}_{\mathbf{u}},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{W}}+\mathbf{c}'_{\boldsymbol{\tau};\mathbf{W}}\mathbf{Y}^{\perp}_{1\mathbf{u}}\big) \\
&\qquad\qquad - F^{Y_{\mathbf{u}}|(\mathbf{Y}^{\perp}_{\mathbf{u}},\mathbf{W})}\big(a_{\boldsymbol{\tau};\mathbf{w}_0}+\mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}_0}\mathbf{Y}^{\perp}_{1\mathbf{u}}+T_{n1}\big)\big)\boldsymbol{\mathcal{X}}^{\ell}_{h1\mathbf{u}}\big] \\
&\quad= \sqrt{nh_n^{p-1}}\left(\frac{h_n^2}{2}\mathbf{B}^{\ell}_{\mathbf{w}_0}+\mathrm{o}\big(h_n^2\big)\right),
\end{aligned}
$$

where the last equality is derived from a first-order Taylor expansion of $y \mapsto F^{Y_{\mathbf{u}}|(\mathbf{Y}^{\perp}_{\mathbf{u}},\mathbf{X})}(y)$ and a second-order Taylor expansion of $\mathbf{w} \mapsto (a_{\boldsymbol{\tau};\mathbf{w}}, \mathbf{c}'_{\boldsymbol{\tau};\mathbf{w}})'$ at $\mathbf{w} = \mathbf{w}_0$ (these expansions exist in view of Assumptions (A1) and (A4)). The $\mathrm{o}(h_n^2)$ term is taken care of by Assumption (A3). The asymptotic variance of the theorem readily follows from Lemma A.6. Details are omitted. □

Acknowledgements

The research of Marc Hallin and Davy Paindaveine is supported by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). Davy Paindaveine moreover benefits from an A.R.C. contract of the Communauté Française de Belgique. The research of Marc Hallin and Zudi Lu is supported by a Discovery Grant, and Zudi Lu benefits from a Future Fellowship Grant, both of the Australian Research Council. Miroslav Šiman acknowledges the support of Project 1M06047 of the Ministry of Education, Youth and Sports of the Czech Republic. Marc Hallin and Davy Paindaveine are members of ECORE, the association between CORE and ECARES. Marc Hallin is also member of the Académie Royale de Belgique, and a fellow of CentER (Tilburg University).

References

[1] Bern, M. and Eppstein, D. (2002). Multivariate regression depth. *Discrete Comput. Geom.* **28** 1–17. MR1904006

[2] Bhattacharya, P.K. and Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415. MR1062716

[3] Bonellie, S.R. and Raab, G.M. (1996). A comparison of different approaches for fitting centile curves to birthweight data. *Stat. Med.* **15** 2657–2667.

[4] Cai, Z. and Xu, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Amer. Statist. Assoc.* **104** 371–383. MR2504383

[5] Chakraborty, B. (2003). On multivariate quantile regression. *J. Statist. Plann. Inference* **110** 109–132. MR1944636

[6] Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777. MR1105843

[7] Cheng, Y. and De Gooijer, J.G. (2007). On the uth geometric conditional quantile. *J. Statist. Plann. Inference* **137** 1914–1930. MR2323873

[8] Cole, T.J. (1988). Fitting smoothed centile curves to reference data. *J. Roy. Statist. Soc. Ser. A* **151** 385–418.

[9] El Ghouch, A. and Van Keilegom, I. (2009). Local linear quantile regression with dependent censored data. *Statist. Sinica* **19** 1621–1640. MR2589201

[10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. London: Chapman and Hall. MR1383587

[11] Fan, J., Hu, T.C. and Truong, Y.K. (1994). Robust non-parametric function estimation. *Scand. J. Stat.* **21** 433–446. MR1310087

[12] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods. Springer Series in Statistics*. New York: Springer. MR1964455

[13] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354

[14] Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. Statist. Plann. Inference* **117** 207–223. MR2004656

[15] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference curves based on non-parametric quantile regression. *Stat. Med.* **21** 3119–3135.

[16] Giloni, A. and Simonoff, J.S. (2005). The conditional breakdown properties of least absolute value local polynomial estimators. *J. Nonparametr. Stat.* **17** 15–30. MR2112685

[17] Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli* **15** 659–686. MR2555194

[18] Hallin, M., Paindaveine, D. and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670

[19] Härdle, W.K. and Song, S. (2010). Confidence bands in quantile regression. *Econometric Theory* **26** 1180–1200. MR2660296

[20] Heinz, G., Peterson, L.J., Johnson, R.W. and Kerk, C.J. (2003). Exploring relationships in body dimensions. *J. Statist. Education* **11**. Available at www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

[21] Hlubinka, D., Kotík, L. and Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika* (*Prague*) **46** 125–148. MR2666899

[22] Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470. MR1794246

[23] Honda, T. (2004). Quantile regression in varying coefficient models. *J. Statist. Plann. Inference* **121** 113–125. MR2027718

[24] Ioannides, D.A. (2004). Fixed design regression quantiles for time series. *Statist. Probab. Lett.* **68** 235–245. MR2083492

[25] Kim, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35** 92–108. MR2332270

[26] Koenker, R. (2005). *Quantile Regression. Econometric Society Monographs* **38**. Cambridge: Cambridge Univ. Press. MR2268657

[27] Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12** 793–813. MR1421404

[28] Kong, E., Linton, O. and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* **26** 1529–1564. MR2684794

[29] Kong, L. and Mizera, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statist. Sinica* **22** 1589–1610. MR3027100

[30] Kong, L. and Zuo, Y. (2010). Smooth depth contours characterize the underlying distribution. *J. Multivariate Anal.* **101** 2222–2226. MR2671212

[31] Lu, Z., Hui, Y.V. and Zhao, Q. (1998). Local linear quantile regression under dependence: Bahadur representation and application. Discussion paper, Dept. Management Sciences, City Univ. Hong Kong.

[32] McKeague, I.W., López-Pintado, S., Hallin, M. and Šiman, M. (2011). Analyzing growth trajectories. *J. Developmental Origins of Health and Disease* **2** 322–329.

[33] Mizera, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736. MR1969447

[34] Paindaveine, D. and Šiman, M. (2012). Computing multiple-output regression quantile regions. *Comput. Statist. Data Anal.* **56** 840–853. MR2888729

[35] Rousseeuw, P.J. and Hubert, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–433. With discussion and a reply by the authors and Stefan Van Aelst. MR1702314

[36] Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838. MR0600964

[37] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. MR2738875

[38] Wang, F.T. and Scott, D.W. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Statist. Assoc.* **89** 65–76. MR1266287

[39] Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.* **103** 397–409. MR2420242

[40] Wei, Y. and He, X. (2006). Conditional growth charts. *Ann. Statist.* **34** 2069–2131. With discussions and a rejoinder by the authors. MR2291494

[41] Wei, Y., Pere, A., Koenker, R. and He, X. (2006). Quantile regression methods for reference growth charts. *Stat. Med.* **25** 1369–1382. MR2226792

[42] Wright, E.M. and Royston, P. (1997). A comparison of statistical methods for age-related reference intervals. *J. Roy. Statist. Soc. Ser. A* **160** 47–69.

[43] Yu, K. and Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimators. *Comput. Statist. Data Anal.* **25** 159–166.

[44] Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93** 228–237. MR1614628

[45] Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scand. J. Stat.* **31** 333–346. MR2087829

[46] Zhang, W. and Lee, S.-Y. (2000). Variable bandwidth selection in varying-coefficient models. *J. Multivariate Anal.* **74** 116–134. MR1790616

[47] Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444

Received August 2012 and revised May 2013