



Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function

Guerre, E; Sabbah, C

Cambridge University Press 2011
subject to the Cambridge Core terms of use

For additional information about this publication click this link.
<http://qmro.qmul.ac.uk/xmlui/handle/123456789/15806>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

UNIFORM BIAS STUDY AND BAHADUR REPRESENTATION
FOR LOCAL POLYNOMIAL ESTIMATORS OF THE CONDITIONAL QUANTILE FUNCTION¹

Emmanuel Guerre²

Camille Sabbah³

September 2014

Abstract

This paper investigates the bias and the weak Bahadur representation of a local polynomial estimator of the conditional quantile function and its derivatives. The bias and Bahadur remainder term are studied uniformly with respect to the quantile level, the covariates and the smoothing parameter. The order of the local polynomial estimator can be higher than the differentiability order of the conditional quantile function. Applications of the results deal with global optimal consistency rates of the local polynomial quantile estimator, performance of random bandwidths and estimation of the conditional quantile density function. The latter allows to obtain a simple estimator of the conditional quantile function of the private values in a first price sealed bids auctions under the independent private values paradigm and risk neutrality.

JEL Classification: Primary C14; Secondary C21.

Keywords: Bahadur representation; Conditional quantile function; Local polynomial estimation; Econometrics of Auctions.

¹ This paper was started and completed when both authors were at Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, which support is gratefully acknowledged. Financial Support from the Department of Economics, Queen Mary University of London, is also gratefully acknowledged. The authors would like to thank the participants of the Queen Mary Econometrics Reading Group, of the Berlin Quantile Regression Workshop, of the LSE Econometrics and Statistics Workshop as well as the Associate Editor and two anonymous referees whose careful readings, suggestions and comments have helped to improve the paper. All remaining errors are under our responsibility. This version corrects an error in the proof of Lemma B.2 which was pointed out by Zhongjun Qu but does not change the results of the published version.

² School of Economics and Finance, Queen Mary University of London, Great-Britain. e.guerre@qmul.ac.uk

³Laboratoire EQUIPPE, Université Lille 3, France. camille.sabbah@univ-lille3.fr

1. INTRODUCTION

The conditional quantile function is a powerful tool to represent the dependence between two variables. Let $Q(\alpha|x)$, α in $(0, 1)$, be the conditional quantile function of a univariate dependent variable Y given $X = x$, where X is the d dimensional covariate, $Q(\alpha|x) = \inf\{y : \mathbb{P}(Y \leq y|X = x) \geq \alpha\}$. Under fairly general conditions, the Lévy-Smirnov-Rosenblatt transformation ensures that there is a random variable A independent of X and uniform over $[0, 1]$ such that

$$(1.1) \quad Y = Q(A|X).$$

In other words, the knowledge of the conditional quantile function allows to compute the impact on Y of a shock on X for any given A . The conditional quantile function is also central in the identification of the impact of such shocks or of more general parameters in nonseparable models in microeconometrics, see Chesher (2003), Chernozhukov and Hansen (2005), Holderlein and Mammen (2007) and Imbens and Newey (2009) to mention just a few. See also Firpo, Fortin and Lemieux (2009) or Rothe (2010) for an unconditional point of view when evaluating distributional policy effects. Conditional quantile approaches can also be useful in industrial organization due to the important role played by increasing functions and the equivariance property of quantile function which states that $\Psi(Q(\alpha|x))$ is the conditional quantile function of $\Psi(Y)$ given X provided Ψ is an increasing transformation. See Haile, Hong and Shum (2003), Marmer and Shneyerov (2008) and below for the case of auctions. Echenique and Komunjer (2009) show the usefulness of a conditional quantile approach when analyzing general multiple equilibria economic models.

However, inference with the quantile representation (1.1) is potentially difficult due to non-separability. In a regression model $Y = m(X) + \varepsilon$ where X and ε are independent, the dependence between Y and X is summarized through the regression function $m(\cdot)$ and does not involve the unobserved noise ε . This contrasts with (1.1) where the random variable A may potentially change the shape of $x \mapsto Q(A|x)$. Hence, inference in (1.1) should not focus on a particular value of the quantile level α but should consider instead all α in an interval $[\underline{\alpha}, \bar{\alpha}]$ close enough to $[0, 1]$, as recommended for instance in the case of the more constrained quantile regression model analyzed in Koenker (2005). In practice, this often leads to consider graphical representations of the estimated curves $x \mapsto \widehat{Q}(\alpha|x)$ for various α . A natural norm for evaluating these estimated graphs is the uniform norm with respect to α and x , $\sup_{\alpha, x} \left| \widehat{Q}(\alpha|x) - Q(\alpha|x) \right|$.

The present paper contributes to this issue for local polynomial quantile estimators $\widehat{Q}_h(\alpha|x)$ which depends upon a bandwidth h . We study its bias uniformly in α and x and derive a uniform Bahadur representation for $\widehat{Q}_h(\alpha|x)$ and its derivatives which holds in probability, that is a

weak Bahadur representation. In few words, a Bahadur representation is an approximation of $\widehat{Q}_h(\alpha|x) - Q(\alpha|x)$ by a bias term plus a leading stochastic term up to remainder term with an explicit order. In our setup, uniformity is with respect to the level α , the bandwidth h , and the covariate x , implying that our Bahadur representation is an important step for the study of $\sup_{\alpha,x} \left| \widehat{Q}(\alpha|x) - Q(\alpha|x) \right|$, see Proposition 2 below. Various other interesting results also follow from our uniform results.

To be more specific, consider independent and identically observations $(X_1, Y_1), \dots, (X_n, Y_n)$ with the same distribution than (X, Y) . Define, for α in $(0, 1)$, the loss function

$$(1.2) \quad \ell_\alpha(q) = |q| + (2\alpha - 1)q = 2q(\alpha - \mathbb{I}(q \leq 0)), \quad q \text{ in } \mathbb{R},$$

where \mathbb{R} stands for the set of real numbers. It is well known that

$$(1.3) \quad Q(\alpha|x) = \arg \min_{q \in \mathbb{R}} \mathbb{E}[\ell_\alpha(Y - q) | X = x]$$

is the conditional quantile of Y given $X = x$. When $d = 1$, the local polynomial estimator of order p of $Q(\alpha|x)$ is $\widehat{Q}_h(\alpha|x) = \widehat{b}_0(\alpha; h, x)$ where, for $\mathbf{b} = (b_0, \dots, b_p)^T$,

$$(1.4) \quad \widehat{\mathbf{b}}(\alpha; h, x) = \arg \min_{\mathbf{b} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \ell_\alpha \left(Y_i - b_0 - b_1(X_i - x) - \dots - \frac{b_p}{p!} (X_i - x)^p \right) K \left(\frac{X_i - x}{h} \right).$$

In the expression above, $p!$ is the factorial $p \times (p-1) \times \dots \times 1$, $K(\cdot)$ is a kernel function and h is a smoothing parameter which goes to 0 with the sample size. As detailed in Section 2 and studied throughout the paper, the local polynomial estimator $\widehat{Q}_h(\alpha|x)$ has a natural extension which covers the multivariate case $d > 1$. As noted in Fan and Gijbels (1996, Chapter 5), the local polynomial estimator $\widehat{Q}_h(\alpha|x)$ is a modification of the Least Squares local polynomial estimator of a regression function which uses the square loss function in (1.4) instead of the loss function $\ell_\alpha(\cdot)$. A Taylor expansion

$$Q(\alpha|X_i) \simeq Q(\alpha|x) + \frac{\partial Q(\alpha|x)}{\partial x} (X_i - x) + \dots + \frac{1}{p!} \frac{\partial^p Q(\alpha|x)}{\partial x^p} (X_i - x)^p$$

suggests that $\widehat{b}_1(\alpha; h, x), \dots, \widehat{b}_p(\alpha; h, x)$ estimate the partial derivatives $\partial^r Q(\alpha|x)/\partial x^r$, $r = 1, \dots, p$, provided $Q(\alpha|x)$ is smooth enough.

Robust local polynomial estimation of a regression function and its derivatives, including quantile methods, has already been considered in many research articles. See in particular Tsybakov (1986) for optimal pointwise consistency rates, Fan (1992) for design adaptation, and Fan and Gijbels (1996) and Loader (1999) for a general overview. The present paper is perhaps more specifically related to Truong (1989), Chaudhuri (1991), Holderlein and Mammen (2009) and Kong, Linton and Xia (2010). Truong (1989) showed that local median estimators achieve

the global optimal rates of Stone (1982) with respect to L_m norms, $0 < m \leq \infty$, for conditional quantile function satisfying a Lipschitz condition. Chaudhuri (1991) obtained a strong (that is which holds in an almost sure sense) Bahadur representation for the local polynomial quantile estimators when the kernel function $K(\cdot)$ of (1.4) is uniform. Hong (2003) extended this result to local polynomial robust M-estimation and more general kernels. The Bahadur representation of Chaudhuri (1991) is pointwise, that is holds for some prescribed x and α and a given deterministic bandwidth $h \rightarrow 0$. As explained and illustrated in Kong et al. (2010), pointwise Bahadur representations are not sufficient for many applications including plug in estimation of conditional quantile functionals or marginal integration estimators. Hence Kong et al. (2010) derives a strong uniform Bahadur representation for robust local polynomial M-estimators for dependent observations. Here uniformity is with respect to the location variable x . For local polynomial quantile estimators of order $p = 1$, Holderlein and Mammen (2009) considers uniformity with respect to α and x but they just show that their remainder term is negligible in probability and does not obtain its order.

In this work, we study the bias term and obtain the order in probability of the Bahadur remainder term uniformly in α , h and x for local polynomial quantile estimators. A first contribution given in Theorem 1 below deals with the study of the bias of local polynomial quantile estimators. Most of the literature has focused on the case where the order p of the local polynomial is equal to the order of differentiability of $x \mapsto Q(\alpha|x)$, say s . This is somehow unrealistic since it amounts to assume that s is known. Since the case where $p \leq s$ can be easily dealt with by ignoring higher order derivatives, we focus in the more interesting case where $p > s$, which has apparently not been considered in the statistical and econometric literature. As shown in Corollary 1, a local polynomial quantile estimator with $p > s$ still allows to estimate $Q(\alpha|x)$ with the optimal rate $n^{-s/(2s+d)}$ of Stone (1982). This suggests that local polynomial estimators using high order p should be preferred since they allow to estimate in an optimal way a wider range of smooth conditional quantile functions. Another interesting conclusion of our bias study is that the additional local polynomial coefficients $\widehat{b}_v(\alpha; h, x)$, $v = s + 1, \dots, p$ can diverge and Proposition 1 describes a simple example where it indeed happens. Hence, in the local polynomial setup, a high value of $\widehat{b}_v(\alpha; h, x)$ may also correspond to a non smooth quantile function in which case a lower degree $p < v$ could have been used.

Our uniform study of the Bahadur remainder term, namely Theorem 2, is the second main contribution of the paper. A third contribution builds on the fact that Theorems 1 and 2 hold uniformly with respect to x in a compact inner subset of the support of X . Combining these results with a study of the stochastic part of the Bahadur representation allows us to show that

the local polynomial quantile estimator achieves the global optimal rates of Stone (1982) for the L_m and uniform norms provided the bandwidth goes to 0 with an appropriate rate. This result, stated in Corollary 1, is apparently new and extends Truong (1989) which is restricted to Lipschitz quantile functions, or Chaudhuri (1991) who considers pointwise optimality. A fourth contribution uses the fact that Theorems 1 and 2 hold uniformly with respect to h in an interval $[\underline{h}, \bar{h}]$. Proposition 2 shows that a random bandwidth performs as well as its deterministic equivalent counterpart with respect to convergence rates of the uniform norm $\sup_{\alpha, x} |\widehat{Q}_h(\alpha|x) - Q(\alpha|x)|$. Such a result gives a solid theoretical basis to Li and Racine (2008) suggestion of choosing the local polynomial bandwidth h via a simpler cross validation procedure for the conditional cumulative distribution function. As mentioned earlier, uniformity with respect to α and x is also useful for graphical representations of (1.1).

A fifth contribution also exploits uniformity with respect to the quantile order α . Proposition 3 considers estimation of the conditional quantile density function

$$(1.5) \quad q(\alpha|x) = \frac{\partial Q(\alpha|x)}{\partial \alpha} = \frac{1}{f(Q(\alpha|x)|x)}.$$

As argued in Parzen (1979), the quantile density function $q(\alpha|x)$ or its inverse $1/q(\alpha|x)$ is a renormalization of the density function $f(y|x)$ which is well suited for statistical explanatory analysis. The function $q(\alpha|x)$ is also crucial for quantile based statistical inference. Indeed, the asymptotic variance of $\widehat{Q}_h(\alpha|x)$ is proportional to

$$\frac{1}{nh} \frac{\alpha(1-\alpha)}{q^2(\alpha|x)f(x)}$$

where $f(\cdot)$ is the marginal density of X , see Fan and Gijbels (1996, p. 202). Hence estimating $q(\alpha|x)$ is useful to estimate the variance of $\widehat{Q}_h(\alpha|x)$. As noted in Guerre, Perrigne and Vuong (2009), the conditional quantile density function plays an important role in the identification of first-price sealed bids auction models. Under the independent private values paradigm and risk neutrality, the conditional quantile function of the private values $Q^v(\alpha|x)$ satisfies

$$Q^v(\alpha|x) = Q^b(\alpha|x) + \frac{\alpha q^b(\alpha|x)}{I-1},$$

where $Q^b(\alpha|x)$ and $q^b(\alpha|x)$ are the conditional quantile function and quantile density function of the bids. Hence estimating $Q^b(\alpha|x)$ and $q^b(\alpha|x)$ gives a straightforward estimation of the conditional quantile function of the private values $Q^v(\alpha|x)$ which is an alternative to the two steps approach of Guerre, Perrigne and Vuong (2000). See Haile et al. (2003) or Marmer and Shneyerov (2008) for a related estimation strategy.

There is however just a few references that address the estimation of $q(\alpha|x)$. For the related function $q(\alpha|x)\partial F(Q(\alpha|x)|x)\partial x$, Lee and Lee (2008) uses a composition approach which non-parametrically estimates $\partial F(y|x)/\partial x$, $f(y|x)$ and $Q(\alpha|x) = F^{-1}(\alpha|x)$. Haile et al. (2003) and Marmer and Shneyerov (2008) proceeds similarly. Xiang (1995) proposes the estimator

$$\frac{1}{h_q} \int \widehat{F}^{-1}(\alpha + h_q a|x) dK_q(a),$$

where $\widehat{F}(y|x)$ is a kernel estimator of the conditional cumulative distribution function, $K_q(\cdot)$ a probability distribution and h_q a smoothing parameter. As argued in Fan and Gijbels (1996), local polynomial estimators may have better design adaptation properties than kernel ones. Hence we propose to use the local polynomial $\widehat{Q}_h(\alpha|x)$ instead of the kernel $\widehat{F}^{-1}(\alpha|x)$. Thanks to uniformity with respect to α in Theorems 1 and 2, the resulting conditional quantile density function estimator $\widehat{q}(\alpha|x)$ has a simple Bahadur representation which facilitates the study of its consistency rate, see Proposition 3.

The rest of the paper is organized as follows. The next section groups our main assumptions and notations and explained in particular how to extend (1.4) to multivariate covariates. Section 3 exposes our main results and Section 4 concludes the paper. The proofs of our statements are gathered in two appendices.

2. MAIN ASSUMPTIONS AND NOTATIONS

The definition (1.4) of $\widehat{Q}_h(\alpha|x)$ assumes that the covariate X is univariate. In the multivariate case, we use a multivariate kernel function $K(z) = K(z_1, \dots, z_d)$ but we restrict to an univariate bandwidth for the sake of simplicity. The univariate polynomial expansion $b_0 + b_1(X_i - x) + \dots + b_p(X_i - x)^p/p!$ is replaced by a multivariate counterpart as defined now. Let \mathbb{N} be the set of natural integer numbers. For $\mathbf{v} = (v_1, \dots, v_d)$ let $|\mathbf{v}| = v_1 + \dots + v_d$ and let P be the number of \mathbf{v} 's with $|\mathbf{v}| \leq p$. Then a generic expression for multivariate polynomial function of order p is, for \mathbf{b} in \mathbb{R}^P ,

$$\mathbf{U}(z)^T \mathbf{b} = \sum_{\mathbf{v}; |\mathbf{v}| \leq p} b_{\mathbf{v}} \frac{z^{\mathbf{v}}}{\mathbf{v}!}, \text{ where } z^{\mathbf{v}} = z_1^{v_1} \times \dots \times z_d^{v_d}, \mathbf{U}(z)^T = \left(\frac{z^{\mathbf{v}}}{\mathbf{v}!}, |\mathbf{v}| \leq p \right),$$

and $\mathbf{v}! = \prod_{i=1}^d v_i!$. In the expression above, the vectors \mathbf{v} of \mathbb{N}^d are ordered according to the lexicographic order. The multivariate version of the local polynomial estimator (1.4) is

$$(2.1) \quad \widehat{\mathbf{b}}(\alpha; h, x) = \arg \min_{\mathbf{b} \in \mathbb{R}^P} \mathcal{L}_n(\mathbf{b}; \alpha, h, x) \text{ with}$$

$$\mathcal{L}_n(\mathbf{b}; \alpha, h, x) = \frac{1}{nh^d} \sum_{i=1}^n \ell_{\alpha} \left(Y_i - \mathbf{U}(X_i - x)^T \mathbf{b} \right) K \left(\frac{X_i - x}{h} \right).$$

As in the univariate case, the entry $\widehat{b}_0(\alpha; h, x) = \widehat{Q}_h(\alpha|x)$ of $\widehat{\mathbf{b}}(\alpha; h, x)$ is an estimator of $Q(\alpha|x)$. The entry $\widehat{b}_{\mathbf{v}}(\alpha; h, x)$ can be viewed as an estimator of the partial derivative

$$b_{\mathbf{v}}(\alpha|x) = \frac{\partial^{|\mathbf{v}|} Q(\alpha|x)}{\partial x_1^{v_1} \times \cdots \times \partial x_d^{v_d}}$$

provided this partial derivative exists. We shall consider later on the following Hölder class. Consider a subset $[\underline{\alpha}, \overline{\alpha}]$ of $(0, 1)$ over which $Q(\alpha|x)$ or its partial derivatives will be estimated. Let $\lfloor s \rfloor$ be the lowest integer part of s , i.e. $\lfloor s \rfloor$ is the unique integer number with $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$. Then $Q(\cdot|\cdot)$ is in $\mathcal{C}(L, s)$, $L, s > 0$, if

- (i) for all α in $[\underline{\alpha}, \overline{\alpha}]$, $x \mapsto Q(\alpha|x)$ is $\lfloor s \rfloor$ -th continuously differentiable over the support \mathcal{X} of X ;
- (ii) for all \mathbf{v} in \mathbb{N}^d with $|\mathbf{v}| = \lfloor s \rfloor$, all α in $[\underline{\alpha}, \overline{\alpha}]$, all x, x' in \mathcal{X} ,

$$|b_{\mathbf{v}}(\alpha|x) - b_{\mathbf{v}}(\alpha|x')| \leq L \|x - x'\|^{s - \lfloor s \rfloor}$$

where $\|\cdot\|$ stands for the Euclidean norm.

Since the estimators $\widehat{b}_{\mathbf{v}}(\alpha; h, x)$ of the partial derivatives $b_{\mathbf{v}}(\alpha|x)$ converge with different rates, we use the diagonal standardization matrix

$$\mathbf{H} = \mathbf{H}(h) = \text{Diag} \left(h^{|\mathbf{v}|}, \mathbf{v} \in \mathbb{N}^d, |\mathbf{v}| \leq p \right).$$

It is well known that local polynomial estimation techniques apply at the boundaries. However we will focus on those x which are in an inner subset \mathcal{X}_0 of the support \mathcal{X} of X to avoid technicalities. Our main assumptions are as follows. Let $\mathcal{B}(0, 1)$ be the closed unit ball $\{z \in \mathbb{R}^d : \|z\| \leq 1\}$.

Assumption X. *The distribution of X has a probability density function $f(\cdot)$ with respect to the Lebesgue measure, which is strictly positive and continuously differentiable over the compact support \mathcal{X} of X . The set \mathcal{X}_0 is a compact subset of the interior of \mathcal{X} .*

Assumption F. *The cumulative distribution function $F(\cdot|\cdot)$ of Y given X has a continuous probability density function $f(y|x)$ with respect to the Lebesgue measure, which is strictly positive for y in \mathbb{R} and x in \mathcal{X} . The partial derivative $\partial F(y|x)/\partial x$ is continuous over $\mathbb{R} \times \mathcal{X}$. There is a $L_0 > 0$, such that*

$$|f(y|x) - f(y|x')| \leq L_0 \|(x, y) - (x', y')\| \text{ for all } (x, y), (x', y') \text{ of } \mathcal{X} \times \mathbb{R}.$$

Assumption K. *The nonnegative kernel function $K(\cdot)$ is Lipschitz over \mathbb{R}^d , has a compact support \mathcal{K} and satisfies $\int K(z)dz = 1$. For some $\underline{K} > 0$, $K(z) \geq \underline{K} \mathbb{I}(z \in \mathcal{B}(0, 1))$. The bandwidth is in $[\underline{h}_n, \bar{h}_n]$ with $0 < \underline{h}_n \leq \bar{h}_n < \infty$, $\lim_{n \rightarrow \infty} \bar{h}_n = 0$ and $\lim_{n \rightarrow \infty} (\log n)/(n\underline{h}_n^d) = 0$.*

Assumption X is standard. Assumption F ensures uniqueness of the conditional quantile $Q(\alpha|x) = F^{-1}(\alpha|x)$ in (1.3) and existence of the quantile density function (1.5). Assumption K allows for a wide range of smoothing parameters $h \rightarrow 0$ in $[\underline{h}_n, \bar{h}_n]$. In the univariate case $d = 1$, Hong (2003) restricts to bandwidths $h = O(n^{-1/(2p+3)})$, a condition which is not imposed here, and Chauduri assumes that h has the exact order $n^{-1/(2p+d)}$. In the simpler context of univariate kernel regression, Einmahl and Mason (2005) assumes $h^d \geq C(\log n)/n$ to obtain uniform consistency so that Assumption K is fairly general.

3. BIAS STUDY AND BAHADUR REPRESENTATION

Applying standard parametric M -estimation theory as detailed in White (1994) or van der Vaart (1998) suggests that the local polynomial estimator $\hat{\mathbf{b}}(\alpha; h, x)$ of (2.1) is an estimator of $\mathbf{b}^*(\alpha; h, x)$ with

$$(3.1) \quad \mathbf{b}^*(\alpha; h, x) = \arg \min_{\mathbf{b} \in \mathbb{R}^p} \mathbb{E} \left[\ell_\alpha \left(Y - \mathbf{U}(X - x)^T \mathbf{b} \right) K \left(\frac{X - x}{h} \right) \right].$$

In particular, $Q_h^*(\alpha|x) = b_0^*(\alpha; h, x)$ may differ from the true conditional quantile $Q(\alpha|x)$ due to a bias term $Q_h^*(\alpha|x) - Q(\alpha|x)$. Studying this bias term can be done using the first-order condition

$$\frac{\partial}{\partial \mathbf{b}^T} \mathbb{E} \left[\ell_\alpha \left(Y - \mathbf{U}(X - x)^T \mathbf{b}^*(\alpha; h, x) \right) K \left(\frac{X - x}{h} \right) \right] = 0,$$

and the Implicit Functions Theorem. This approach gives in particular the order of the difference between $b_{\mathbf{v}}^*(\alpha; h, x)$ and the \mathbf{v} th partial derivative $b_{\mathbf{v}}(\alpha|x)$ of $Q(\alpha|x)$ provided the partial derivative exists.

Theorem 1. *Assume that $Q(\cdot|\cdot)$ is in a Hölder class $\mathcal{C}(L, s)$ with $\lfloor s \rfloor \leq p$. Then under Assumptions F, K and X and provided \bar{h} is small enough, there is a constant C such that for all $|\mathbf{v}| \leq \lfloor s \rfloor$ and n large enough,*

$$\sup_{(\alpha, h, x) \in [\underline{\alpha}, \bar{\alpha}] \times [\underline{h}, \bar{h}] \times \mathcal{X}_0} \left| \frac{b_{\mathbf{v}}^*(\alpha; h, x) - b_{\mathbf{v}}(\alpha|x)}{h^{s-|\mathbf{v}|}} \right| \leq CL.$$

It follows that $Q^*(\alpha|x) - Q(\alpha|x) = O(h^s)$ and more generally that

$$b_{\mathbf{v}}^*(\alpha; h, x) - b_{\mathbf{v}}(\alpha|x) = O\left(h^{s-|\mathbf{v}|}\right)$$

uniformly provided $|\mathbf{v}| \leq \lfloor s \rfloor$. Since $\lfloor s \rfloor \leq p$, the bias order $h^{s-|\mathbf{v}|}$ is not affected by the order p of the local polynomial estimator. This bias order is better than the bias order $h^{p-|\mathbf{v}|}$, $|\mathbf{v}| \leq p$, that would be achieved by suboptimal local polynomial estimators of lower order $p < \lfloor s \rfloor$.

The proof of Theorem 1 establishes a slightly stronger result since it also gives the order of the coefficients $b_{\mathbf{v}}^*(\alpha; h, x)$ with $|\mathbf{v}| > \lfloor s \rfloor$ which correspond to partial derivatives that may not exist. Indeed, equation (A.8) of the proof of Theorem 1 implies that

$$(3.2) \quad b_{\mathbf{v}}^*(\alpha; h, x) = O\left(h^{s-|\mathbf{v}|}\right) \quad \text{for } |\mathbf{v}| \geq s$$

uniformly in $(\alpha, h, x) \in [\underline{\alpha}, \bar{\alpha}] \times [\underline{h}, \bar{h}] \times \mathcal{X}_0$. See also Loader (1999, Theorem 4.2) which gives a less precise $b_{\mathbf{v}}^*(\alpha; h, x) = o(h^{-|\mathbf{v}|})$. Hence the higher order polynomial coefficients $b_{\mathbf{v}}^*(\alpha; h, x)$, $|\mathbf{v}| > s$, may diverge when $h > 0$. That this may be indeed the case can be seen on a simple regression example. Consider

$$(3.3) \quad Y = m(X) + \varepsilon, \quad m(x) = \begin{cases} |x|^{1/2} & \text{if } x \geq 0 \\ -|x|^{1/2} & \text{if } x < 0 \end{cases},$$

where the $\mathcal{U}([-1, 1])$ random variable X and the $\mathcal{N}(0, 1)$ ε are independent. Let $\Phi(\cdot)$ be the cumulative distribution function of the standard normal $\mathcal{N}(0, 1)$. In this example, $Q(\alpha|x) = \Phi^{-1}(\alpha) + m(x)$ inherits of the smoothness properties of the regression function $m(\cdot)$. Note that the differential of $m(\cdot)$ at $x = 0$ is infinite. It also follows that $Q(\alpha|x)$ is at best in an Hölder class $\mathcal{C}(L, 1/2)$ since, for L large enough,

$$|m(x) - m(x')| \leq L |x - x'|^{1/2} \quad \text{for all } (x, x') \in [-1, 1]^2,$$

an inequality that cannot be improved by increasing the exponent $1/2$ as seen by taking $x = 0$ and $x' \rightarrow 0$. The next Proposition uses the behavior of $m(\cdot)$ at $x = 0$ to show that the rate given in (3.2) is sharp.

Proposition 1. *Suppose that (X, Y) satisfies (3.3). Let $\mathbf{b}^*(\alpha; h, x) = (b_0^*(\alpha; h, x), b_1^*(\alpha; h, x))^T$ from (3.1) be given by a local polynomial procedure of order 1. Then under Assumption K and $\int zK(z)dz = 0$, $b_0^*(0.5; h, 0) = m(0) + O(h^{1/2})$ and $b_1^*(0.5; h, 0)$ diverges with the exact rate $h^{-1/2}$,*

$$\lim_{h \rightarrow 0} h^{1/2} b_1^*(0.5; h, 0) = \frac{\int |z|^{3/2} K(z) dz}{\int z^2 K(z) dz} \neq 0.$$

The divergence of $b_1^*(0.5; h, 0)$ implies that the estimator $\widehat{b}_1(0.5; h, 0)$ will diverge in probability. This recalls that observing a large $\widehat{b}_1(0.5; h, 0)$ is not an argument for claiming that a local polynomial estimator of order $p = 1$ should be used.

We now consider the stochastic terms $\widehat{Q}_h(\alpha|x) - Q_h^*(\alpha|x)$ and the rescaled

$$\mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right).$$

Let us first introduce some additional notations. Local polynomial estimation builds on a order p Taylor expansion of $Q(\alpha|x')$ with x' in the vicinity of x . This Taylor expansion can be written as $Q(\alpha|x') \simeq \mathbf{U}(x' - x)^T \mathbf{b}_p(\alpha|x)$ where $\mathbf{b}_p(\alpha|x)$ groups the partial derivatives of $Q(\alpha|x)$ with respect to x . Consider the following counterpart of the Taylor approximation,

$$(3.4) \quad Q^*(x'; \alpha, h, x) = \mathbf{U}(x' - x)^T \mathbf{b}^*(\alpha, h, x)$$

Define also $\mathbf{S}_i(\alpha; h, x) = \mathbf{S}(X_i, Y_i; \alpha, h, x)$ and $\mathbf{J}_i(\alpha; h, x) = \mathbf{J}(X_i; \alpha, h, x)$ with

$$(3.5) \quad \mathbf{S}_i(\alpha; h, x) = 2 \{ \mathbb{I}(Y_i \leq Q^*(X_i; \alpha, h, x)) - \alpha \} \mathbf{U} \left(\frac{X_i - x}{h} \right) K \left(\frac{X_i - x}{h} \right),$$

$$(3.6) \quad \mathbf{J}_i(\alpha; h, x) = 2f(Q^*(X_i; \alpha, h, x) | X_i) \mathbf{U} \left(\frac{X_i - x}{h} \right) \mathbf{U} \left(\frac{X_i - x}{h} \right)^T K \left(\frac{X_i - x}{h} \right).$$

Since

$$\mathbf{U}(X_i - x) = \mathbf{H} \mathbf{U} \left(\frac{X_i - x}{h} \right)$$

and (1.2) gives

$$\begin{aligned} & \frac{\partial \ell_\alpha}{\partial \mathbf{b}^T} \left(Y_i - \mathbf{U}(X_i - x)^T \mathbf{b} \right) K \left(\frac{X_i - x}{h} \right) \\ &= 2 \left\{ \mathbb{I} \left(Y_i \leq \mathbf{U}(X_i - x)^T \mathbf{b} \right) - \alpha \right\} \mathbf{U}(X_i - x) K \left(\frac{X_i - x}{h} \right) \end{aligned}$$

almost everywhere, the variables $\mathbf{S}_i(\alpha; h, x)$ satisfy

$$\frac{\partial \mathcal{L}_n}{\partial \mathbf{b}^T} (\mathbf{b}^*(\alpha, h, x); \alpha, h, x) = \frac{\mathbf{H}}{nh^d} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)$$

almost everywhere. Hence $\sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)$ can be viewed as a score function term whereas $\sum_{i=1}^n \mathbf{J}_i(\alpha; h, x)$ is actually similar to a second derivative of the objective function \mathcal{L}_n although it is not twice differentiable. Indeed, it can be shown that it admits a quadratic approximation with second-order derivatives

$$\mathbf{H} \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right) \mathbf{H}.$$

Classical results of White (1994) or van der Vaart (1998) for parametric estimation suggests that a candidate approximation for $\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x)$ is

$$\begin{aligned} & - \left(\mathbf{H} \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right) \mathbf{H} \right)^{-1} \frac{\mathbf{H}}{nh^d} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x) \\ & = -\mathbf{H}^{-1} \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right)^{-1} \frac{1}{nh^d} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x). \end{aligned}$$

Hence the rescaled $(nh^d)^{1/2} \mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right)$ is expected to be close to

$$(3.7) \quad \beta_n(\alpha; h, x) = - \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right)^{-1} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x).$$

$\sum_{i=1}^n \mathbf{J}_i(\alpha; h, x)/(nh^d)$ is similar to a Kernel regression estimator and obeys a Law of Large Numbers for triangular array which ensures that this matrix is asymptotically close to

$$2f(Q^*(x; \alpha, h, x) | x) \int \mathbf{U}(t) \mathbf{U}(t)^T K(t) dt.$$

Since this matrix is symmetric positive definite, the inverse in (3.7) exists with a probability tending to 1. The term $\sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)/(nh^d)^{1/2}$ has a similar kernel structure but with centered

$\mathbf{S}_i(\alpha; h, x)$, see (A.1) in Lemma A.1 of Appendix A. Hence $\sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)/(nh^d)^{1/2}$ satisfies a pointwise Central Limit Theorem, as $\beta_n(\alpha; h, x)$. Hence $(nh^d)^{1/2} \mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right)$ should also be asymptotically Gaussian provided the so called Bahadur error term

$$(3.8) \quad \mathbf{E}_n(\alpha; h, x) = (nh^d)^{1/2} \mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right) - \beta_n(\alpha; h, x).$$

is asymptotically negligible pointwisely. But transposing the various uniform results established in the Appendices for the leading term $\beta_n(\alpha; h, x)$ of the expansion of $(nh^d)^{1/2} \mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right)$ requests a uniform study of $\mathbf{E}_n(\alpha; h, x)$.

Techniques to study $\mathbf{E}_n(\alpha; h, x)$ for a fixed argument α , h and x are given in Hjort and Pollard (1993). See also Fan, Heckman and Wand (1995, p.143) or Fan and Gijbels (1996, p.210). In our uniform setup, obtaining an uniform order for $\mathbf{E}_n(\alpha; h, x)$ is performed using a

preliminary uniform study of a stochastic process we introduce now. Define first

$$\begin{aligned} & \mathbb{L}_{1n}(\beta; \alpha, h, x) \\ &= nh^d \left\{ \mathcal{L}_n \left(\mathbf{b}^*(\alpha; h, x) + \frac{H^{-1}\beta}{(nh^d)^{1/2}}; \alpha, h, x \right) - \mathcal{L}_n(\mathbf{b}^*(\alpha; h, x); \alpha, h, x) \right\} \\ &= \sum_{i=1}^n \left\{ \ell_\alpha \left(Y_i - Q^*(X_i; \alpha, h, x) - \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T}{(nh^d)^{1/2}} \beta \right) - \ell_\alpha(Y_i - Q^*(X_i; \alpha, h, x)) \right\} K \left(\frac{X_i - x}{h} \right), \end{aligned}$$

which is such that

$$(nh^d)^{1/2} \mathbf{H} \left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x) \right) = \arg \min_{\beta} \mathbb{L}_{1n}(\beta; \alpha, h, x).$$

It then follows from (3.8) that

$$\begin{aligned} \mathbf{E}_n(\alpha; h, x) &= \arg \min_{\epsilon} \mathbb{L}_n(\beta_n(\alpha; h, x), \epsilon; \alpha, h, x) \text{ where} \\ (3.9) \quad \mathbb{L}_n(\beta, \epsilon; \alpha, h, x) &= \mathbb{L}_{1n}(\beta + \epsilon; \alpha, h, x) - \mathbb{L}_{1n}(\beta; \alpha, h, x). \end{aligned}$$

Hence the stochastic process \mathbb{L}_n plays a central role in our analysis. Especially useful is the decomposition

$$\mathbb{L}_n(\beta, \epsilon; \alpha, h, x) = \mathbb{L}_n^0(\beta, \epsilon; \alpha, h, x) + \mathbb{R}_n(\beta, \epsilon; \alpha, h, x)$$

where \mathbb{L}_n^0 is the quadratic approximation of \mathbb{L}_n ,

$$\begin{aligned} \mathbb{L}_n^0(\beta, \epsilon; \alpha, h, x) &= \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)^T (\beta + \epsilon) + \frac{1}{2} (\beta + \epsilon)^T \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right) (\beta + \epsilon) \\ &\quad - \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)^T \beta + \frac{1}{2} \beta^T \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right) \beta \\ (3.10) \quad &= \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x)^T \epsilon + \frac{1}{2} \epsilon^T \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\alpha; h, x) \right) (\epsilon + 2\beta), \end{aligned}$$

and \mathbb{R}_n is a remainder term. As in the expression above (3.9) for $\mathbf{E}_n(\alpha; h, x)$, the variable β above in (3.10) will be taken equal to $\beta_n(\alpha; h, x)$ in the proof of Theorem 2 below. As noted in the quadratic approximation lemma of Fan et al. (1995, p.148) in the pointwise case, the order of $\mathbf{E}_n(\alpha; h, x)$ is driven by the order of \mathbb{R}_n . The proof of the next Theorem relies on an uniform study of \mathbb{R}_n based on a maximal inequality under bracketing entropy conditions from Massart (2007), see the proof of Proposition A.1. This maximal inequality plays here the role of the Bernstein inequality used in the pointwise framework of Hong (2003).

Theorem 2. *Under Assumptions F, K and X,*

$$\sup_{(\alpha, h, x) \in [\underline{\alpha}, \bar{\alpha}] \times [\underline{h}, \bar{h}] \times \mathcal{X}_0} \|\mathbf{E}_n(\alpha; h, x)\| = O_{\mathbb{P}} \left(\frac{\log^3(n)}{n\underline{h}^d} \right)^{1/4}.$$

In the case where the lower and upper bandwidths \underline{h} and \bar{h} have the same order, Theorem 2 gives uniformly in h in $[\underline{h}, \bar{h}]$, α and x ,

$$\widehat{Q}_h(\alpha|x) = Q_h^*(\alpha|x) + \frac{\mathbf{e}_0^T \beta_n(\alpha; h, x)}{(nh^d)^{1/2}} + O_{\mathbb{P}} \left(\frac{\log n}{nh^d} \right)^{3/4},$$

where \mathbf{e}_0 is the first vector of the canonical basis of \mathbb{R}^P , which first coordinate is equal to 1 and the other ones are equal to 0. For h of order $n^{-1/(2p+d)}$ as studied in Chauduri (1991, Theorem 3.2), the order of the remainder term is $n^{-3p/(2(2p+d))} \log^{3/4} n$ as found by this author. When $d = 1$, Hong (2003) obtains the better order $(\log \log n / (nh))^{-3/4}$ but his Bahadur representation only holds pointwisely in α and x . It can be conjectured that the order $(\log n / (nh^d))^{-3/4}$ is optimal for Bahadur expansion holding uniformly with respect to x .

For higher order partial derivatives, Theorem 2 yields

$$\widehat{b}_{\mathbf{v}}(\alpha; h, x) = b^*(\alpha; h, x) + \frac{\mathbf{e}_{\mathbf{v}}^T \beta_n(\alpha; h, x)}{(nh^d)^{1/2} h^{|\mathbf{v}|}} + \frac{1}{h^{|\mathbf{v}|}} O_{\mathbb{P}} \left(\frac{\log n}{nh^d} \right)^{3/4},$$

where the \mathbf{v} th entry of $\mathbf{e}_{\mathbf{v}}$ is 1 and the other are 0, see also Hong (2003) for a pointwise version of this expansion and Kong et al. (2010) for a version which is uniform with respect to x . Such expansion can be used to study the pointwise asymptotic normality of the local polynomial quantile estimator. Combining this Bahadur representation with the bias study of Theorem 1 gives a global rate result which is apparently new. The next Corollary extends the study of local medians in Truong (1989).

Corollary 1. *Assume that $Q(\alpha|x)$ is in $\mathcal{C}(L, s)$ for some $\lfloor s \rfloor \leq p$. Suppose that Assumptions F, K and X hold. Then for all partial derivative order \mathbf{v} with $|\mathbf{v}| \leq \lfloor s \rfloor$ and all α in $[\underline{\alpha}, \bar{\alpha}]$,*

- (i) $\left(\int_{\mathcal{X}_0} \left| \widehat{b}_{\mathbf{v}}(\alpha; h, x) - b_{\mathbf{v}}(\alpha|x) \right|^m dx \right)^{1/m} = O_{\mathbb{P}} \left(\frac{1}{n} \right)^{\frac{s-|\mathbf{v}|}{2s+d}}$ for any finite $m > 0$ provided h is asymptotically proportional to $n^{-\frac{1}{2s+d}}$;
- (ii) $\sup_{x \in \mathcal{X}_0} \left| \widehat{b}_{\mathbf{v}}(\alpha; h, x) - b_{\mathbf{v}}(\alpha|x) \right| = O_{\mathbb{P}} \left(\frac{\log n}{n} \right)^{\frac{s-|\mathbf{v}|}{2s+d}}$ if h is asymptotically proportional to $\left(\frac{\log n}{n} \right)^{\frac{1}{2s+d}}$.

Since the $b_{\mathbf{v}}(\alpha|x)$ are estimators of the partial derivatives of $m(x)$ in a regression model as (3.3), It follows from Stone (1982) that the global rates derived in Corollary 1 are optimal in a minimax sense.

A second application builds on the uniformity with respect to the bandwidth h of our Bahadur representation. The next Proposition allows for data-driven bandwidths. Observe that it also deals with the uniform norm $\sup_{(\alpha,x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}_0} \left| \widehat{Q}_h(\alpha|x) - Q(\alpha|x) \right|$ which evaluates the estimated curves $(\alpha, x) \mapsto \widehat{Q}_h(\alpha|x)$ used in empirical graphic illustrations of (1.1).

Proposition 2. *Consider a random bandwidth \widehat{h}_n such that $\widehat{h}_n = O_{\mathbb{P}}(h_n)$ and $1/\widehat{h}_n = O_{\mathbb{P}}(1/h_n)$ where h_n is a deterministic sequence satisfying $h_n = o(1)$ and $\lim_{n \rightarrow \infty} (\log n)/(nh_n^d) = 0$. Suppose that Assumption K, F and X hold and that $Q(\alpha|x)$ is in $\mathcal{C}(L, s)$. Then for any \mathbf{v} with $|\mathbf{v}| \leq \lfloor s \rfloor$,*

$$\sup_{(\alpha,x) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}_0} \left| \widehat{b}_{\mathbf{v}}(\alpha; \widehat{h}_n, x) - b_{\mathbf{v}}(\alpha|x) \right| = h_n^{-|\mathbf{v}|} O_{\mathbb{P}} \left(h_n^s + \left(\frac{\log n}{nh_n^d} \right)^{1/2} \right).$$

In particular if the exact order of \widehat{h}_n is $(\log(n)/n)^{1/(2s+d)}$ in probability, $\sup_{x \in \mathcal{X}_0} \left| \widehat{b}_{\mathbf{v}}(\alpha; \widehat{h}_n, x) - b_{\mathbf{v}}(\alpha|x) \right|$ has the optimal order $(\log(n)/n)^{(s-|\mathbf{v}|)/(2s+d)}$ of Corollary 1-(ii). It is likely that an L_m version of Proposition 2 holds but it is slightly longer to prove. Proposition 2 can be for instance fruitfully applied to cross-validated bandwidths for the conditional cumulative distribution as proposed by Li and Racine (2008).

Our last application builds on the fact that Theorems 1 and 2 hold uniformly with respect to the quantile order α . This application concerns estimation of the conditional quantile density function (1.5). The considered estimator of $q(\alpha|x)$ is a conditional version of the Parzen (1979) convolution estimator,

$$(3.11) \quad \widehat{q}(\alpha|x) = \frac{1}{h_q} \int \widehat{Q}_h(a|x) dK_q \left(\frac{a - \alpha}{h_q} \right) = \frac{1}{h_q} \int \widehat{Q}_h(\alpha + h_q t|x) dK_q(t),$$

see also Xiang (1995). In the expression above, $h_q > 0$ is a bandwidth and $K_q(\cdot)$ is a signed measure over \mathbb{R} such that

$$\int dK_q(t) = 0, \quad \int t dK_q(t) = 1.$$

In particular, if $K_q(\cdot)$ has a Lebesgue derivative $dK_q(t) = K'_q(t)dt$, substituting in (3.11) gives

$$\widehat{q}(\alpha|x) = \frac{1}{h_q} \int \widehat{Q}_h(\alpha + h_q t|x) K'_q(t) dt.$$

Computing these integrals may request intensive numerical steps so that the resulting estimator may be difficult to implement in practice. A more realistic estimator uses a discrete measure $K_q(\cdot)$ in (3.11). If $K_q(\cdot)$ is a linear combination of Dirac masses at t_j with weights κ_j , $j = 1, \dots, J$, the resulting estimator

$$\widehat{q}(\alpha|x) = \frac{1}{h_q} \sum_{j=1}^J \kappa_j \widehat{Q}_h(\alpha + h_q t_j|x), \quad \sum_{j=1}^J \kappa_j = 0 \text{ and } \sum_{j=1}^J t_j \kappa_j = 1,$$

may be indeed simpler to compute. Note that this includes the well known numerical derivatives

$$\frac{\widehat{Q}_h(\alpha + h_q|x) - \widehat{Q}_h(\alpha|x)}{h_q}, \quad \frac{\widehat{Q}_h(\alpha|x) - \widehat{Q}_h(\alpha - h_q|x)}{h_q} \quad \text{and} \quad \frac{\widehat{Q}_h(\alpha + h_q|x) - \widehat{Q}_h(\alpha - h_q|x)}{2h_q}.$$

To study the bias of $\widehat{q}(\alpha|x)$, we strengthen the definition of the smoothness class $\mathcal{C}(L, s)$ as follows. $Q(\alpha|x)$ is in $\mathcal{C}_q(L, s)$ if

- (i) $Q(\alpha|x)$ is in $\mathcal{C}(L, s + 1)$;
- (ii) For each x in \mathcal{X} , $\alpha \in [\underline{\alpha}, \bar{\alpha}] \mapsto q(\alpha|x)$ is $\lfloor s \rfloor$ th differentiable;
- (iii) For each x in \mathcal{X} and all $(\alpha, \alpha') \in [\underline{\alpha}, \bar{\alpha}]^2$

$$\left| \frac{\partial^{\lfloor s \rfloor} q(\alpha|x)}{\partial \alpha^{\lfloor s \rfloor}} - \frac{\partial^{\lfloor s \rfloor} q(\alpha'|x)}{\partial \alpha^{\lfloor s \rfloor}} \right| \leq L |\alpha - \alpha'|^{s - \lfloor s \rfloor}.$$

We shall assume in addition that $K_q(\cdot)$ has a compact support and satisfies the additional conditions

$$\int t^j dK_q(t) = 0, \quad j = 1, \dots, \lfloor s \rfloor, \quad \int |dK_q(t)| < \infty.$$

Proposition 3. *Assume that $Q(\alpha|x)$ is in $\mathcal{C}_q(L, s)$ and $\lfloor s + 1 \rfloor \leq p$. Suppose that Assumptions K , F and X hold with $h = O(h_q)$, $h_q \rightarrow 0$ and $(\log n)/(nh^d) \rightarrow 0$. Then for any x in \mathcal{X}_0 and α in $(\underline{\alpha}, \bar{\alpha})$,*

$$\widehat{q}(\alpha|x) = q(\alpha|x) + O_{\mathbb{P}} \left(h_q^s + \frac{1}{(nh^d h_q)^{1/2}} \right) + \frac{\log^{3/4} n}{(nh^d h_q^2)^{1/4}} O_{\mathbb{P}} \left(\frac{1}{(nh_q h^d)^{1/2}} \right).$$

Taking h_q and h of the same order is the optimal choice for the order of h in the expansion of Proposition 3. This gives

$$\widehat{q}(\alpha|x) = q(\alpha|x) + O_{\mathbb{P}} \left(h^s + \frac{1}{(nh^{d+1})^{1/2}} \right) + \frac{\log^{3/4} n}{(nh^{d+2})^{1/4}} O_{\mathbb{P}} \left(\frac{1}{(nh^{d+1})^{1/2}} \right).$$

The item $(\log^{3/4} n) (nh^{d+2})^{-1/4} O_{\mathbb{P}} \left((nh^{d+1})^{-1/2} \right)$ is given by the Bahadur error term $\mathbf{E}_n(\alpha; h, x)$ of Theorem 2. The other item, $O_{\mathbb{P}} \left(h^s + (nh^{d+1})^{-1/2} \right)$, can be viewed as a bias variance decomposition component. The latter is the leading term of the expansion provided $nh^{d+2} \rightarrow \infty$, a condition also used in Lee and Lee (2008) when $d = 1$. In this case, the optimal order for h is $n^{-1/(2s+d+1)}$, which is such that $nh^{d+2} \rightarrow \infty$ provided $s > 1/2$. In this case, the optimal rate for pointwise estimation of $q(\alpha|x)$ is $n^{-s/(2s+d+1)}$ which, as expected from (1.5), coincides with the optimal rate for pointwise estimation of $f(y|x)$.

4. FINAL REMARKS

This paper has investigated the bias and the Bahadur representation of a local polynomial estimator of the conditional quantile function and its derivatives. Compared to the existing literature, a distinctive feature is that the bias and Bahadur remainder term are studied uniformly with respect to the quantile level, the covariates and the smoothing parameter, extending so Chauduri (1991) and Kong et al. (2010). Our framework also considers the case where the order of the local polynomial estimator p is higher than the order of differentiability s of the conditional quantile function. An interesting consequence of our bias study is that using a local polynomial estimator of order $p \geq s$ does not affect its rate optimality.

Our uniform study of the bias and of the Bahadur remainder term are applied to derive the global rate optimality of the local polynomial estimators of the conditional quantile function and its derivatives with respect to L_m norms, $0 < m \leq \infty$ provided the bandwidth goes to 0 with an appropriate rate. This extends Truong (1989) who states a similar result for local medians and under a rather strong Lipschitz condition for the conditional quantile function. Another application deals with the performance of randomly selected bandwidths that are shown to perform as well as their deterministic equivalent in term of consistency rates in uniform norm. Our framework is flexible enough to be adapted to other global norms. This new result is especially useful in view of Li and Racine (2008) suggestion of implementing local polynomial quantile estimation with a data-driven bandwidth given by a cross validation criterion for the conditional cumulative distribution function. A last application to nonparametric estimation of the quantile density function can be useful for confidence intervals and in Econometrics of Auctions where the conditional quantile density function plays an important role.

Our uniform results can also be useful for other studies. For instance an issue far beyond the scope of the present paper is the choice of the local polynomial order p . Local polynomial quantile estimation can be implemented using a large p , possibly growing with the sample size. This would allow to estimate very smooth conditional quantile function with a small bias although it may inflate the asymptotic variance of the resulting estimator. Another approach would be to use a data-driven local polynomial order p . Such a problem is very close to the issue of choosing the order of the kernel when estimating a regression or a probability density function. The latter can be addressed following the recent adaptive approach of Goldenshluger and Lepski (2008,2009) which gives a data-driven choice of the kernel and bandwidth in the context of the continuous time white noise model. Our uniform Bahadur representation is a preliminary step that can be useful to extend their results to local polynomial quantile estimation.

REFERENCES

- [1] CHAUDURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics*, **19**, 760-777.
- [2] CHERNOZHUKOV, V. & C. HANSEN (2005). An IV model of quantile treatments effects. *Econometrica* **73**, 245-261.
- [3] CHESHER, A. (2003). Identification in nonseparable models. *Econometrica*, **71**, 1405-1441.
- [4] CHOW, Y.S. & H. TEICHER (2003). *Probability theory. Independence, interchangeability, martingales*. Third edition, Springer.
- [5] ECHENIQUE, F. & I. KOMUNJER (2009). Testing models with multiple equilibria by quantile methods. *Econometrica* **77**, 1281-1297.
- [6] EINMAHL, U. & D.M. MASON (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, **33**, 1380-1403.
- [7] FAN, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87**, 998-1004.
- [8] FAN, J. & I. GIJBELS (1996). *Local polynomial modeling and its applications*. Chapman & Hall/CRC.
- [9] FAN, J., N.E. HECKMAN & M.P. WAND (1995). Local polynomial kernel regression for generalized linear model and quasi-likelihood functions. *Journal of the American Statistical Association*, **90**, 141-151.
- [10] FIRPO, S., N. FORTIN & T. LEMIEUX (2009). Unconditional quantile regression. *Econometrica* **77**, 953-973.
- [11] GUERRE, E., I. PERRIGNE & Q. VUONG (2000). Optimal nonparametric estimation of first price auctions. *Econometrica*, **68**, 525-574.
- [12] GUERRE, E., I. PERRIGNE & Q. VUONG (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica* **77**, 1193-1227.
- [13] GOLDENSHLUGER, A. & O. LEPSKI (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14**, 1150-1190.
- [14] GOLDENSHLUGER, A. & O. LEPSKI (2009). Structural adaptation via \mathbb{L}_p -norm oracle inequalities. *Probability Theory and Related Fields* **143**, 41-71.
- [15] HAILE, P.A, H. HONG & M. SHUM (2003). Nonparametric tests for common values in first-price sealed-bid auctions. Cowles Foundation discussion paper.
- [16] HJORT N. & POLLARD D. (1993). Asymptotics for minimisers of convex processes. Unpublished manuscript, <http://www.stat.yale.edu/Pollard/Papers/>.
- [17] HOLDERLEIN, S. & E. MAMMEN (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* **75**, 1513-1518.
- [18] HOLDERLEIN, S. & E. MAMMEN (2009). Identification and estimation of local average derivatives in nonseparable models without monotonicity. *Econometrics Journal* **12**, 1-25.
- [19] HONG, S.Y. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M -regression. *Journal of Nonparametric Statistics*, **15**, 237-251.
- [20] IMBENS, G.W. & W.K. NEWEY (2009). Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica* **77**, 1481-1512.
- [21] KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press.

- [22] KONG, E., O. LINTON & Y. XIA (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* **26**, 1529-1564.
- [23] LEE, K.L. & E.R. LEE (2008). Kernel methods for estimating derivatives of conditional quantiles. *Journal of the Korean Statistical Society*, **37**, 365-373.
- [24] LI Q. & J. S. RACINE (2008). Nonparametric Estimation of Conditional CDF and Quantile Functions With Mixed Categorical and Continuous Data. *Journal of Business and Economic Statistics*, **26**, 423-434.
- [25] LOADER, C. (1999). *Local regression and likelihood*. Springer.
- [26] MARMER, V. & A. SHNEYEROV (2008). Quantile-based nonparametric inference for first-price auctions. Document paper.
- [27] MASSART, P. (2007). *Concentration inequalities and model selection. Lecture Notes in Mathematics*, **1896**. Ecole d'Eté de Probabilités de Saint Flour XXXIII-2003, Jean Picard Editor. Springer.
- [28] PARZEN, E. (1979). Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, **74**, 105-121.
- [29] ROTHE, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics* **155**, 56-70.
- [30] STONE, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10**, 1040-1053.
- [31] TSYBAKOV, A.B. (1986). Robust reconstruction of functions by the local-approximation method. *Problemy Peredachi Informatsii*, **22**, 69-84.
- [32] TRUONG, Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *The Annals of Statistics*, **17**, 606-617.
- [33] VAN DE GEER, S. (1999). *Empirical processes in M-estimation*. Cambridge University Press.
- [34] VAN DER VAART, A.W. (1998). *Asymptotic statistics*. Cambridge University Press.
- [35] WHITE, H. (1994). *Estimation, inference and specification analysis. Econometric Society Monographs*. Cambridge University Press.
- [36] XIANG, X. (1995). Estimation of conditional quantile density function. *Journal of Nonparametric Statistics*, **4**, 309-316.
- [37] ZEIDLER, E. (1985). *Nonlinear Functional Analysis and its Applications, I. Fixed-Point Theorems*. Springer Verlag.

APPENDIX A: PROOFS OF MAIN RESULTS

Appendix A groups the proofs of Theorems 1 and 2, Propositions 1, 2 and 3, and Corollary 1. The proofs of intermediary results used to prove these main results are grouped in Appendix B.

We first introduce some additional notations. Sequences $\{a_n\}$ and $\{b_n\}$ satisfy $a_n \asymp b_n$ if $|a_n|/C \leq |b_n| \leq C|a_n|$ for some $C > 0$ and n large enough. Recall that $\|\cdot\|$ is the Euclidean norm and $\mathcal{B}(0, 1) = \{z; \|z\| \leq 1\}$. Let \succ be the usual order for symmetric matrices, that is $\mathbf{A}_1 \succ \mathbf{A}_2$ if and only if $\mathbf{A}_1 - \mathbf{A}_2$ is a non-negative symmetric matrix. If \mathbf{A} is a symmetric matrix, $\|\mathbf{A}\| = \sup_{\mathbf{u} \in \mathcal{B}(0,1)} \|\mathbf{A}\mathbf{u}\| = \sup_{\mathbf{u} \in \mathcal{B}(0,1)} |\mathbf{u}^T \mathbf{A} \mathbf{u}|$ is the largest eigenvalue in absolute value of \mathbf{A} . This norm is such that $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ for any matrix or vector \mathbf{B} . Denote by $\|\cdot\|_\infty$ the uniform norm, i.e. $\|f(\cdot)\|_\infty = \sup_{(x,y) \in \mathbb{R}^d \times \mathbb{R}} |f(y|x)|$.

We use the abbreviation $\theta = (\alpha, h, x)$. In particular, $Q^*(x'; \theta)$, $\mathbf{S}_i(\theta)$ and $\mathbf{J}_i(\theta)$ stand for $Q^*(x'; \alpha, h, x)$, $\mathbf{S}(X_i, Y_i; \alpha, h, x)$ and $\mathbf{J}(X_i; \alpha, h, x)$, see equations (3.4), (3.5) and (3.6). We abbreviate \underline{h}_n and \bar{h}_n into \underline{h} and \bar{h} . Define

$$\Theta^0 = [\underline{\alpha}, \bar{\alpha}] \times [0, \bar{h}] \times \mathcal{X}_0, \quad \Theta^1 = [\underline{\alpha}, \bar{\alpha}] \times [\underline{h}, \bar{h}] \times \mathcal{X}_0,$$

where \mathcal{X}_0 is as in Assumption X and $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ is as in the definition of the smoothness class $\mathcal{C}(L, s)$. For $\mathcal{L}_n(\mathbf{b}; \alpha, h, x) = \mathcal{L}_n(\mathbf{b}; \theta)$ as in (2.1), define

$$\mathcal{L}(\mathbf{b}; \theta) = \mathbb{E}[\mathcal{L}_n(\mathbf{b}; \theta)] = \frac{1}{h^d} \mathbb{E} \left[\left\{ \ell_\alpha \left(Y - \mathbf{U}(X - x)^T \mathbf{b} \right) - \ell_\alpha(Y) \right\} K \left(\frac{X - x}{h} \right) \right].$$

We also use $K_h(z) = K(z/h)$. It is convenient to change \mathbf{b} into its standardization $\mathbf{B} = \mathbf{H}\mathbf{b}$ and to define $\widehat{\mathbf{B}}(\theta) = \widehat{\mathbf{H}}\widehat{\mathbf{b}}(\theta)$ and $\mathbf{B}^*(\theta) = \mathbf{H}\mathbf{b}^*(\theta)$. Absolute constants are denoted by the generic letter C and may vary from line to line.

The following argument is used systemically. Recall that \mathcal{X}_0 is an inner subset of the compact \mathcal{X} under Assumption X. Hence for any $(x, h) \in \mathcal{X}_0 \times \mathcal{K}$, $x + hz$ is in \mathcal{X} under Assumption K provided \bar{h} is small enough.

The next lemma is used in the proof of Theorems 1 and 2. Its proof is given in Appendix B with the proof of the other intermediary results.

Lemma A.1. *Under Assumption F, K and X, we have for \bar{h} small enough,*

- (i) $\mathbf{b}^*(\theta)$ exists and is unique for all θ in Θ^0 .
- (ii) $\mathbf{B}^*(\theta) = \mathbf{H}\mathbf{b}^*(\theta)$ satisfies

$$(A.1) \quad \mathbb{E}[\mathbf{S}_i(\theta)] = \int \left\{ F \left(\mathbf{U}(z)^T \mathbf{B}^*(\theta) | x + hz \right) - F(Q(\alpha | x + hz) | x + hz) \right\} f(x + hz) \mathbf{U}(z) K(z) dz = 0,$$

$$(A.2) \quad \lim_{\bar{h} \rightarrow 0} \sup_{\theta \in \Theta^0} \|\mathbf{B}^*(\theta) - \mathbf{B}^*(\alpha; 0, x)\| = 0,$$

$$\text{where } \mathbf{B}^*(\alpha; 0, x) = (Q(\alpha | x), 0, \dots, 0)^T.$$

- (iii) for all (x', θ_i) in $\mathcal{X} \times \Theta^1$, $i = 1, 2$,

$$|Q^*(x'; \theta_1) - Q^*(x'; \theta_2)| \leq C \underline{h}^{-p} (1 + \underline{h}^{-1}) \|\theta_1 - \theta_2\|.$$

- (iv) There exists C such that, for all θ in Θ^1 , all x' in \mathcal{X} and all x in \mathcal{X}_0 ,

$$f(Q^*(x'; \theta) | x') K \left(\frac{x - x'}{h} \right) \geq CK \left(\frac{x - x'}{h} \right).$$

A.1. Proof of Theorem 1. Since $Q(\cdot | \cdot)$ is in $\mathcal{C}(L, s)$, the Taylor-Lagrange Formula and Assumption K yield that there exists $t = t(h, x, z)$ in $(0, 1)$ such that for h small enough and all (x, z) in $\mathcal{X}_0 \times \mathcal{K}$,

$$(A.3) \quad \begin{aligned} Q(\alpha | x + hz) &= \sum_{0 \leq |\mathbf{v}| \leq [s]} \frac{b_{\mathbf{v}}(\alpha | x)}{\mathbf{v}!} (hz)^{\mathbf{v}} + \sum_{|\mathbf{v}| = [s]} \frac{(hz)^{\mathbf{v}}}{\mathbf{v}!} (b_{\mathbf{v}}(\alpha | x + thz) - b_{\mathbf{v}}(\alpha | x)) \\ &= \mathbf{U}(z)^T \mathbf{H}\mathbf{b}(\alpha | x) + \epsilon(\theta, z). \end{aligned}$$

In the equation above, $b_{\mathbf{v}}(\alpha|x)$ is the \mathbf{v} th partial derivatives of $Q(\alpha|x)$ with respect to x and $\mathbf{b}(\alpha|x) = (b_{\mathbf{v}}(\alpha|x), |\mathbf{v}| \leq \lfloor s \rfloor, 0, \dots, 0)^T \in \mathbb{R}^P$. Since $Q(\cdot) \in \mathcal{C}(L, s)$,

$$(A.4) \quad \lim_{\bar{h} \rightarrow 0} \sup_{(\theta, z) \in \Theta^1 \times \mathcal{K}} \left| \frac{\epsilon(\theta, z)}{h^s} \right| \leq CL.$$

Let

$$I(\theta, z) = \int_0^1 f(Q(\alpha|x + hz) + t(\mathbf{U}(z)^T \mathbf{B}^*(\theta) - Q(\alpha|x + hz))|x + hz) dt.$$

Assumptions F, K, X, $Q(\cdot) \in \mathcal{C}(L, s)$ and (A.2) give

$$(A.5) \quad \lim_{\bar{h} \rightarrow 0} \sup_{(\theta, z) \in \Theta^0 \times \mathcal{K}} |I(\theta, z) - f(Q(\alpha|x)|x)| = 0.$$

A Taylor expansion with integral remainder gives

$$F(\mathbf{U}(z)^T \mathbf{B}^*(\theta)|x + hz) - F(Q(\alpha|x + hz)|x + hz) = (\mathbf{U}(z)^T \mathbf{B}^*(\theta) - Q(\alpha|x + hz)) I(\theta, z).$$

Substituting in the first-order condition (A.1) yields

$$(A.6) \quad \int \mathbf{U}(z) (\mathbf{U}(z)^T \mathbf{B}^*(\theta) - Q(\alpha|x + hz)) I(\theta, z) f(x + hz) K(z) dz = 0.$$

We show that the matrix $\int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz$ has an inverse. Indeed, Assumptions K and X, (A.5) and \bar{h} small enough give that uniformly in θ in Θ^0 and \mathbf{A} in \mathbb{R}^P ,

$$\begin{aligned} \mathbf{A}^T \int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz \mathbf{A} &= \int \left\| \mathbf{U}(z)^T \mathbf{A} \right\|^2 I(\theta, z) f(x + hz) K(z) dz \\ &= (1 + o(1)) f(Q(\alpha|x)|x) \int \left\| \mathbf{U}(z)^T \mathbf{A} \right\|^2 K(z) dz \\ &\geq C \|\mathbf{A}\|^2, \end{aligned}$$

using the fact that $\mathbf{A} \mapsto \int \left\| \mathbf{U}(z)^T \mathbf{A} \right\|^2 K(z) dz$ is a square norm and norm equivalence over \mathbb{R}^P . It follows that $\int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz$ is strictly positive definite and has an inverse which satisfies, for n large enough

$$(A.7) \quad \sup_{\theta \in \Theta^0} \left\| \left[\int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz \right]^{-1} \right\| < \infty.$$

(A.6) and (A.3) give

$$\mathbf{Hb}^*(\theta) = \mathbf{Hb}(\alpha|x) + \left[\int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz \right]^{-1} \int \epsilon(\theta, z) I(\theta, z) f(x + hz) \mathbf{U}(z) K(z) dz.$$

It then follows from (A.4) and (A.7) that

$$\begin{aligned} &\| \mathbf{Hb}^*(\theta) - \mathbf{Hb}(\alpha|x) \| \\ &\leq \left\| \left[\int \mathbf{U}(z) \mathbf{U}(z)^T I(\theta, z) f(x + hz) K(z) dz \right]^{-1} \right\| \left\| \int \epsilon(\theta, z) I(\theta, z) f(x + hz) \mathbf{U}(z) K(z) dz \right\| \\ (A.8) \quad &\leq CLh^s \end{aligned}$$

uniformly in θ in Θ^0 . This ends the proof of the Theorem and also establishes (3.2) since $\mathbf{b}(\alpha|x) = (b_{\mathbf{v}}(\alpha|x), |\mathbf{v}| \leq \lfloor s \rfloor, 0, \dots, 0)^T$. \square

A.2. **Proof of Proposition 1.** Let $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$, $\Phi(t) = \int_{-\infty}^t \varphi(u)du$ be the p.d.f and c.d.f of the standard normal. The regression model (3.3) is such that

$$F(y|x) = \Phi(y - m(x)), \quad f(x) = \mathbb{I}(x \in [-1, 1]).$$

(A.2) gives that $\lim_{h \rightarrow 0} \max_{z \in \mathcal{K}} |\mathbf{U}(z)^T \mathbf{B}(0.5; h, 0)| = Q(0.5|0) = m(0) = 0$. Hence (A.6), (A.5) and Assumption K give

$$(1 + o(1)) \varphi(0) \int \mathbf{U}(z) (\mathbf{U}(z)^T \mathbf{B}(0.5; h, 0) - m(hz)) K(z) dz = 0.$$

Recall that $\mathbf{U}(z) = (1, z)^T$, so that the equation above gives

$$\begin{aligned} \begin{bmatrix} b_0(0.5; h, 0) \\ hb_1(0.5; h, 0) \end{bmatrix} &= (1 + o(1)) \left(\int \mathbf{U}(z) \mathbf{U}^T(z) K(z) dz \right)^{-1} \begin{bmatrix} h^{1/2} \int m(z) K(z) dz \\ h^{1/2} \int |z|^{3/2} K(z) dz \end{bmatrix} \\ &= (1 + o(1)) h^{1/2} \begin{bmatrix} \int m(z) K(z) dz \\ \frac{\int |z|^{3/2} K(z) dz}{\int z^2 K(z) dz} \end{bmatrix}. \square \end{aligned}$$

A.3. **Proof of Theorem 2.** We first state some intermediary results. The two following propositions deals with the remainder term $\mathbb{R}_n(\beta, \epsilon; \theta) = \sum_{i=1}^n \mathbf{R}_i(\beta, \epsilon; \theta)$ from (3.10), where

$$\begin{aligned} \mathbf{R}_i(\beta, \epsilon; \theta) &= \left\{ \ell_\alpha \left(Y_i - Q^*(X_i; \theta) - \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T (\beta + \epsilon)}{(nh^d)^{1/2}} \right) - \ell_\alpha \left(Y_i - Q^*(X_i; \theta) - \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T \beta}{(nh^d)^{1/2}} \right) \right\} K \left(\frac{X_i - x}{h} \right) \\ &\quad - \frac{1}{(nh^d)^{1/2}} \mathbf{S}_i(\theta)^T \epsilon - \frac{1}{2} \epsilon^T \left(\frac{1}{nh^d} \mathbf{J}_i(\theta) \right) (\epsilon + 2\beta). \end{aligned}$$

Define also

$$\begin{aligned} \text{(A.9)} \quad R_i(\beta, \epsilon; \theta) &= \mathbf{R}_i(\beta, \epsilon; \theta) + \frac{1}{2} \epsilon^T \left(\frac{1}{nh^d} \mathbf{J}_i(\theta) \right) (\epsilon + 2\beta) \\ &= \left\{ \ell_\alpha \left(Y_i - Q^*(X_i; \theta) - \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T (\beta + \epsilon)}{(nh^d)^{1/2}} \right) \right. \\ &\quad \left. - \ell_\alpha \left(Y_i - Q^*(X_i; \theta) - \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T \beta}{(nh^d)^{1/2}} \right) \right. \\ &\quad \left. - 2 \{ \mathbb{I}(Y_i \leq Q^*(X_i; \theta)) - \alpha \} \frac{\mathbf{U} \left(\frac{X_i - x}{h} \right)^T \epsilon}{(nh^d)^{1/2}} \right\} K \left(\frac{X_i - x}{h} \right), \end{aligned}$$

$$\text{(A.10)} \quad \mathbf{R}_i^1(\beta, \epsilon; \theta) = R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta) | X_i],$$

$$\text{(A.11)} \quad \mathbf{R}_i^2(\beta, \epsilon; \theta) = \mathbb{E}[R_i(\beta, \epsilon; \theta) | X_i] - \frac{1}{2} \epsilon^T \left(\frac{1}{nh^d} \mathbf{J}_i(\theta) \right) (\epsilon + 2\beta),$$

which are such that

$$\mathbb{R}_n(\beta, \epsilon; \theta) = \mathbb{R}_n^1(\beta, \epsilon; \theta) + \mathbb{R}_n^2(\beta, \epsilon; \theta), \quad \mathbb{R}_n^j(\beta, \epsilon; \theta) = \sum_{i=1}^n \mathbf{R}_i^j(\beta, \epsilon; \theta), \quad j = 1, 2.$$

Proposition A.1. Consider two real numbers $t_\beta, t_\epsilon > 0$ which may depend upon on n with $t_\beta \geq 1$, $t_\epsilon \geq 1/n$ and $(t_\beta + t_\epsilon)^{1/2}/t_\epsilon \leq O\left(\left(n\underline{h}^d\right)^{1/4}/\log^{1/2}n\right)$. Then, under Assumptions F, K and X and for n large enough,

$$\mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^1(\beta, \epsilon; \theta)| \right] \leq C \frac{\log^{1/2}n}{\left(n\underline{h}^d\right)^{1/4}} t_\epsilon (t_\beta + t_\epsilon)^{1/2}.$$

Proposition A.2. Consider two real numbers $t_\beta, t_\epsilon > 0$ which may depend upon on n with $t_\beta \geq 1$ and $t_\beta/t_\epsilon = O\left(n\underline{h}^d/\log^{1/2}n\right)$. Then, under Assumptions F, K and X and for n large enough,

$$\mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^2(\beta, \epsilon; \theta)| \right] \leq C \frac{t_\epsilon (t_\beta + t_\epsilon)^2}{\left(n\underline{h}^d\right)^{1/2}}.$$

The next lemma is used to bound the eigenvalues of $\sum_{i=1}^n \mathbf{J}_i(\theta)/(nh^d)$ from below. It implies in particular that all the $\beta_n(\theta)$ in (3.7), θ in Θ^1 , are well defined with a probability tending to 1. Let $\underline{\gamma}_n(\theta)$ be the smallest eigenvalue of the nonnegative symmetric matrix $\sum_{i=1}^n \mathbf{J}_i(\theta)/(nh)^d$.

Lemma A.2. Under Assumptions F, K and X, $\inf_{\theta \in \Theta^1} \underline{\gamma}_n(\theta) \geq \underline{\gamma} + o_{\mathbb{P}}(1)$ for some $\underline{\gamma} > 0$.

Lemma A.2 together Lemma A.3 below gives $\sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2}n\right)$.

Lemma A.3. Suppose that Assumptions F, K and X are satisfied. Then

$$\sup_{\theta \in \Theta^1} \left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\theta) \right\| = O_{\mathbb{P}}\left(\log^{1/2}n\right).$$

The rest of the proof of Theorem 2 is divided in two steps. In what follows

$$t_n = t \frac{\log^{3/4}n}{\left(n\underline{h}^d\right)^{1/4}}, \quad t > 0.$$

Under Assumption K, $(\log n)/(n\underline{h}^d) = o(1)$ so that $t_n = o\left(\log^{1/2}n\right)$. In the sequel, t_n will play the role of t_ϵ whereas t_β will be chosen such that $t_\beta \asymp \log^{1/2}n$. Hence

$$\begin{aligned} \frac{(t_\beta + t_\epsilon)^{1/2}}{t_\epsilon} &\asymp \frac{\left(n\underline{h}^d\right)^{1/4} \log^{1/4}n}{t \log^{3/4}n} = \frac{1}{t} O\left(\frac{\left(n\underline{h}^d\right)^{1/4}}{\log^{1/2}n}\right), \\ \frac{t_\beta}{t_\epsilon} &\asymp \frac{\left(n\underline{h}^d\right)^{1/4} \log^{1/2}n}{t \log^{3/4}n} = O\left(\frac{\left(n\underline{h}^d\right)^{1/4}}{\log n}\right)^{1/4} = o\left(\frac{n\underline{h}^d}{\log n} \times \log^{1/2}n\right) = o\left(\frac{n\underline{h}^d}{\log^{1/2}n}\right). \end{aligned}$$

Hence these choices of t_β and t_ϵ satisfy the conditions of Propositions A.1 and A.2 provided t is chosen large enough.

Step 1: order of $\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)|$. Consider $\eta > 0$ arbitrarily small. Let $\underline{\gamma}$ be as in Lemma A.2. Since Lemmas A.2 and A.3 give $\sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2}n\right)$, there is a C_η such that,

for n large enough,

$$\begin{aligned}
& \mathbb{P} \left(\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \geq \frac{\gamma t_n^2}{4} \right) \\
& \leq \mathbb{P} \left(\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \geq \frac{\gamma t_n^2}{4}, \sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| \leq C_\eta \log^{1/2} n \right) \\
& \quad + \mathbb{P} \left(\sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| > C_\eta \log^{1/2} n \right) \\
& \leq \mathbb{P} \left(\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, C_\eta \log^{1/2} n) \times \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta, \epsilon; \theta)| \geq \frac{\gamma t_n^2}{4} \right) + \eta.
\end{aligned}$$

Propositions A.1 and A.2, $\mathbb{R}_n = \mathbb{R}_n^1 + \mathbb{R}_n^2$ and the Markov inequality give

$$\begin{aligned}
& \mathbb{P} \left(\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, C_\eta \log^{1/2} n) \times \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta, \epsilon; \theta)| \geq \frac{\gamma t_n^2}{4} \right) \\
& \leq \frac{C}{t_n^2} \left(\frac{t_n (C_\eta \log^{1/2} n + t_n)^{1/2} \log^{1/2} n}{(nh^d)^{1/4}} + \frac{t_n (C_\eta \log^{1/2} n + t_n)^2}{(nh^d)^{1/2}} \right) \\
& = \frac{C \log^{3/4} n}{t_n (nh^d)^{1/4}} \left(\left(C_\eta + \frac{t_n}{\log^{1/2} n} \right)^{1/2} + \left(\frac{\log n}{nh^d} \right)^{1/4} \left(C_\eta + \frac{t_n}{\log^{1/2} n} \right)^2 \right).
\end{aligned}$$

The definition of t_n , $t_n = o(\log^{1/2} n)$ and Assumption K give

$$(A.12) \quad \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \geq \frac{\gamma t_n^2}{4} \right) = \eta + O\left(\frac{C_\eta^{1/2}}{t}\right) \text{ when } t \rightarrow \infty.$$

Step 2: $\sup_{\theta \in \Theta^1} \|\mathbf{E}_n(\theta)\|$. Consider $\tau_n \geq t_n$ and $\epsilon = \tau_n \mathbf{e}$, $\|\mathbf{e}\| = 1$ so that $\|\epsilon\| \geq t_n$. Since $\ell_\alpha(\cdot)$ is convex, $\epsilon \mapsto \mathbb{L}_n(\beta(\theta), \epsilon; \theta)$ is convex. This gives since $\mathbb{L}_n(\beta(\theta), 0; \theta) = 0$ and $\mathbb{L}_n = \mathbb{L}_n^0 + \mathbb{R}_n$

$$\begin{aligned}
\frac{t_n}{\tau_n} \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta) &= \frac{t_n}{\tau_n} \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta) + \left(1 - \frac{t_n}{\tau_n}\right) \mathbb{L}_n(\beta_n(\theta), 0; \theta) \\
&\geq \mathbb{L}_n\left(\beta_n(\theta), \frac{t_n}{\tau_n} \epsilon; \theta\right) = \mathbb{L}_n(\beta_n(\theta), t_n \mathbf{e}; \theta) \\
&\geq \mathbb{L}_n^0(\beta_n(\theta), t_n \mathbf{e}; \theta) + \mathbb{R}_n(\beta_n(\theta), t_n \mathbf{e}; \theta).
\end{aligned}$$

Hence $\mathbf{E}_n(\theta) = \arg \min_\epsilon \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta)$ and the latter inequality give

$$\begin{aligned}
\{\|\mathbf{E}_n(\theta)\| \geq t_n\} &\subset \left\{ \inf_{\epsilon; \|\epsilon\| \geq t_n} \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta) \leq \inf_{\epsilon; \|\epsilon\| < t_n} \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta) \right\} \\
&\subset \left\{ \inf_{\epsilon; \|\epsilon\| \geq t_n} \mathbb{L}_n(\beta_n(\theta), \epsilon; \theta) \leq \mathbb{L}_n(\beta_n(\theta), 0; \theta) = 0 \right\} \\
&\subset \left\{ \inf_{\mathbf{e}; \|\mathbf{e}\|=1} [\mathbb{L}_n^0(\beta_n(\theta), t_n \mathbf{e}; \theta) + \mathbb{R}_n(\beta_n(\theta), t_n \mathbf{e}; \theta)] \leq 0 \right\} \\
&\subset \left\{ \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0(\beta_n(\theta), \epsilon; \theta) - \sup_{\|\epsilon\|=t_n} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \leq 0 \right\}.
\end{aligned}$$

Since

$$\left\{ \sup_{\theta \in \Theta^1} \|\mathbf{E}_n(\theta)\| \geq t_n \right\} = \bigcup_{\theta \in \Theta^1} \{ \|\mathbf{E}_n(\theta)\| \geq t_n \},$$

this gives

$$\begin{aligned} \left\{ \sup_{\theta \in \Theta^1} \|\mathbf{E}_n(\theta)\| \geq t_n \right\} &\subset \bigcup_{\theta \in \Theta^1} \left\{ \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0(\beta_n(\theta), \epsilon; \theta) - \sup_{\|\epsilon\|=t_n} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \leq 0 \right\} \\ (A.13) \quad &\subset \left\{ \inf_{\theta \in \Theta^1} \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0(\beta_n(\theta), \epsilon; \theta) \leq \sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \right\}. \end{aligned}$$

Consider first $\inf_{\theta \in \Theta^1} \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0(\beta_n(\theta), \epsilon; \theta)$. The definition (3.10) of \mathbb{L}_n^0 gives, for any ϵ with $\|\epsilon\| = t_n$,

$$\mathbb{L}_n^0(\beta_n(\theta), \epsilon; \theta) = \frac{1}{2} \epsilon^T \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta) \right) \epsilon \geq \frac{1}{2} \underline{\gamma}_n(\theta) t_n^2.$$

Hence (A.13), Lemma A.2 and (A.12) give

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\theta \in \Theta^1} \|\mathbf{E}_n(\theta)\| \geq t_n \right) &\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \geq \frac{\underline{\gamma}_n(\theta) t_n^2}{2} \right) \\ &\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{(\epsilon, \theta) \in \mathcal{B}(0, t_n) \times \Theta^1} |\mathbb{R}_n(\beta_n(\theta), \epsilon; \theta)| \geq \frac{\underline{\gamma} t_n^2}{4} \right) \\ &= \eta + O \left(\frac{C_\eta^{1/2}}{t} \right) \text{ when } t \rightarrow \infty. \end{aligned}$$

Since the latter can be made arbitrarily small by taking η arbitrarily small and then t large enough, the Theorem is proved. \square

A.4. Proof of Corollary 1. Part (i) follows from Theorems 1 and 2 and the triangular inequality, together with

$$\left(\int_{\mathcal{X}_0} \|\beta_n(\alpha; h, x)\|^m dx \right)^{1/m} = O_{\mathbb{P}}(1).$$

We now prove the latter. Lemma A.2 and the Hölder inequality give, since \mathcal{X}_0 is compact,

$$\left(\int_{\mathcal{X}_0} \|\beta_n(\alpha; h, x)\|^m dx \right)^{1/m} = O_{\mathbb{P}} \left(\int_{\mathcal{X}_0} \left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\alpha; h, x) \right\|^{2[m]+2} dx \right)^{1/(2[m]+2)}.$$

Since $\mathbb{E}[\mathbf{S}_i(\theta)] = 0$, the Marcinkiewicz-Zygmund inequality (see Chow and Teicher, 2003), (3.5) and $h^d \geq C(\log n)/n$ give

$$\begin{aligned} \mathbb{E}^{1/(2[m]+2)} \left[\left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\theta) \right\|^{2[m]+2} \right] &\leq C \mathbb{E}^{1/(2[m]+2)} \left[\left(\frac{1}{nh^d} \sum_{i=1}^n \|\mathbf{S}_i(\theta)\|^2 \right)^{[m]+1} \right] \\ &\leq C \left(\frac{1}{(nh^d)^{[m]+1}} \sum_{i_1, \dots, i_{[m]+1}=1}^n \mathbb{E} \left[\mathbb{I} \left(\frac{X_{i_1} - x}{h} \in \mathcal{K} \right) \times \dots \times \mathbb{I} \left(\frac{X_{i_{[m]+1}} - x}{h} \in \mathcal{K} \right) \right] \right)^{1/2} = O(1), \end{aligned}$$

uniformly in x . Part (ii) similarly follows from Lemmas A.2 and A.3 which gives $\sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2} n\right)$. \square

A.5. Proof of Proposition 2. Let $\underline{h} = h_n/C$ and $\bar{h} = Ch_n$. The condition on h_n ensures that \underline{h} and \bar{h} satisfy Assumption K for all $C > 1$. Recall that Lemma A.2 together Lemma A.3 gives $\sup_{\theta \in \Theta^1} \|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2} n\right)$. Hence (3.8), Theorems 1 and 2 give, for all $C > 1$,

$$\begin{aligned} \sup_{(\alpha, x, h) \in [\underline{\alpha}, \bar{\alpha}] \times \mathcal{X}_0 \times [\underline{h}, \bar{h}]} \left| \widehat{b}_{\mathbf{v}}(\alpha; \widehat{h}, x) - b_{\mathbf{v}}(\alpha|x) \right| &= \underline{h}^{-|\mathbf{v}|} O_{\mathbb{P}}\left(\underline{h}^s + \left(\frac{\log n}{n\underline{h}^d}\right)^{1/2}\right) \\ &= h_n^{-|\mathbf{v}|} O_{\mathbb{P}}\left(h_n^s + \left(\frac{\log n}{nh_n^d}\right)^{1/2}\right). \end{aligned}$$

This ends the proof of the Proposition since $\liminf_{n \rightarrow \infty} \mathbb{P}\left(\widehat{h}_n \in [\underline{h}, \bar{h}]\right)$ can be made arbitrarily close to 1 by increasing C . \square

A.6. Proof of Proposition 3. Substituting (3.8) in (3.11) yields

$$\begin{aligned} \widehat{q}(\alpha|x) - q(\alpha|x) &= \frac{1}{h_q} \int Q(\alpha + h_q t|x) dK_q(t) - q(\alpha|x) \\ &\quad + \frac{1}{h_q} \int (Q^*(\alpha + h_q t|x) - Q(\alpha + h_q t|x)) dK_q(t) \\ &\quad + \int \frac{\mathbf{e}_0^T \beta_n(\alpha + h_q t; h, x)}{h_q (nh^d)^{1/2}} dK_q(t) + \int \frac{\mathbf{e}_0^T \mathbf{E}_n(\alpha + h_q t; h, x)}{h_q (nh^d)^{1/2}} dK_q(t). \end{aligned}$$

Theorems 1 and 2 with $h = O(h_q)$ and $h_q \rightarrow 0$ give

$$\begin{aligned} \frac{1}{h_q} \int (Q^*(\alpha + h_q t|x) - Q(\alpha + h_q t|x)) dK_q(t) &= O\left(\frac{h^{s+1}}{h_q} \int |dK_q(t)|\right) = O(h_q^s), \\ \int \frac{\mathbf{E}_n(\alpha + h_q t; h, x)}{h_q (nh^d)^{1/2}} dK_q(t) &= \frac{\log^{3/4} n}{(nh^d h_q^2)^{1/4}} O_{\mathbb{P}}\left(\frac{1}{(nh_q h^d)^{1/2}}\right). \end{aligned}$$

Hence it remains to show that

$$(A.14) \quad \frac{1}{h_q} \int Q(\alpha + h_q t|x) dK_q(t) - q(\alpha|x) = O(h_q^s),$$

$$(A.15) \quad \frac{1}{h_q^{1/2}} \int \beta_n(\alpha + h_q t; h, x) dK_q(t) = O_{\mathbb{P}}(1).$$

The two next steps establish these two equalities.

Step 1: proof of (A.14). Let $q^{(j)}(\alpha|x) = \partial^j q(\alpha|x)/\partial x^j$. Since $Q(\alpha|x) \in \mathcal{C}(L, s+1)$, the Taylor-Lagrange Formula gives, for some ω in $[0, 1]$,

$$Q(\alpha + h_q t|x) - Q(\alpha|x) = \sum_{j=0}^{\lfloor s \rfloor} \frac{q^{(j)}(\alpha|x)}{(j+1)!} (h_q t)^j + \frac{q^{(\lfloor s \rfloor)}(\alpha + \omega h_q t|x) - q^{(\lfloor s \rfloor)}(\alpha|x)}{(\lfloor s \rfloor + 1)!} (h_q t)^{\lfloor s \rfloor}.$$

The definition of the smoothness class $\mathcal{C}_q(L, s)$ gives

$$\left| q^{(\lfloor s \rfloor)}(\alpha + \omega h_q t | x) - q^{(\lfloor s \rfloor)}(\alpha | x) \right| \leq L |h_q t|^{s - \lfloor s \rfloor}.$$

Hence, since the support of $K_q(\cdot)$ is compact, $\int |dK_q(t)| < \infty$ and $\int dK_q(t) = 0$, $\int t dK_q(t) = 1$, $\int t^2 dK_q(t) = \dots = \int t^{\lfloor s \rfloor} dK_q(t) = 0$,

$$\begin{aligned} \frac{1}{h_q} \int Q(\alpha + h_q t | x) dK_q(t) &= \frac{Q(\alpha | x)}{h_q} \int dK_q(t) + q(\alpha | x) \int t dK_q(t) + \frac{h_q q^{(1)}(\alpha | x)}{2} \int t^2 dK_q(t) \\ &\quad + \dots + \frac{h_q^{\lfloor s \rfloor} q^{(\lfloor s \rfloor)}(\alpha | x)}{(\lfloor s \rfloor + 1)} \int t^{\lfloor s \rfloor} dK_q(t) + O(h^s) \\ &= q(\alpha | x) + O(h^s). \end{aligned}$$

Step 2: proof of (A.15). Let $\theta_t = (\alpha + h_q t, h, x)$, $\theta = \theta_0$. Since $\int dK_q(t) = 0$, (3.7) gives

$$\begin{aligned} \frac{1}{h_q^{1/2}} \int \beta_n(\theta_t) dK_q(t) &= \frac{1}{h_q^{1/2}} \int (\beta_n(\theta_t) - \beta_n(\theta)) dK_q(t) \\ \text{(A.16)} \quad &= \frac{1}{h_q^{1/2}} \int \left\{ \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta) \right)^{-1} - \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta_t) \right)^{-1} \right\} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\theta) dK_q(t) \end{aligned}$$

$$\text{(A.17)} \quad + \frac{1}{h_q^{1/2}} \int \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta_t) \right)^{-1} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} dK_q(t).$$

Since $\mathbf{A} \mapsto \mathbf{A}^{-1}$ is Lipschitz over the set of semi-definite positive matrices \mathbf{A} with smallest eigenvalue bounded from below by $\underline{\gamma}$, Lemmas A.2 and A.3, (3.6) and Assumption F yield that (A.16) satisfies

$$\begin{aligned} &\left\| \frac{1}{h_q^{1/2}} \int \left\{ \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta) \right)^{-1} - \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta_t) \right)^{-1} \right\} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\theta) dK_q(t) \right\| \\ &\leq \frac{O_{\mathbb{P}}(1)}{h_q^{1/2}} \int \left\| \frac{1}{nh^d} \sum_{i=1}^n \{\mathbf{J}_i(\theta_t) - \mathbf{J}_i(\theta)\} \right\| \left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_i(\theta) \right\| |dK_q(t)| \\ &\leq \frac{O_{\mathbb{P}}(\log n)^{1/2}}{h_q^{1/2}} \int \frac{1}{nh^d} \sum_{i=1}^n |Q^*(X_i; \theta_t) - Q^*(X_i; \theta)| \mathbb{I} \left(\frac{X_i - x}{h} \in \mathcal{K} \right) |dK_q(t)|. \end{aligned}$$

The definition (3.4) of $Q^*(X; \theta)$ and (A.8) give, since $Q(\alpha|x) \in \mathcal{C}(L, s+1)$ and because the support of K_q is compact,

$$\begin{aligned}
& \frac{1}{h_q^{1/2}} \int \frac{1}{nh^d} \sum_{i=1}^n |Q^*(X_i; \theta_t) - Q^*(X_i; \theta)| \mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right) |dK_q(t)| \\
&= \frac{1}{h_q^{1/2}} \int \frac{1}{nh^d} \sum_{i=1}^n \left| \mathbf{U}\left(\frac{X_i - x}{h}\right)^T (\mathbf{Hb}^*(\theta_t) - \mathbf{Hb}^*(\theta)) \right| \mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right) |dK_q(t)| \\
&\leq C \frac{1}{nh^d} \sum_{i=1}^n \mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right) \frac{1}{h_q^{1/2}} \int \|\mathbf{Hb}^*(\theta_t) - \mathbf{Hb}^*(\theta)\| |dK_q(t)| \\
&\leq O_{\mathbb{P}}(1) \frac{1}{h_q^{1/2}} \left(\int \|\mathbf{Hb}(\alpha + h_q t|x) - \mathbf{Hb}(\alpha|x)\| |dK_q(t)| + O(h^{s+1}) \right) \\
&\leq O_{\mathbb{P}}(1) \frac{1}{h_q^{1/2}} \left(\int |Q(\alpha + h_q t|x) - Q(\alpha|x)| |dK_q(t)| + O(h^{s+1} + h) \right) = O_{\mathbb{P}}\left(h_q^{1/2}\right).
\end{aligned}$$

This gives that the item in (A.16) is $O_{\mathbb{P}}\left(h_q^{1/2}\right) = o_{\mathbb{P}}(1)$.

For (A.17), Lemma A.2, $\mathbb{E}[\mathbf{S}_i(\theta_t)] = 0$, (3.5), $Q^*(X; \theta_t) = Q^*(X; \theta) + O(h_q)$ uniformly with respect to t in the support of K_q and $X \in x + h\mathcal{K}$ (as easily seen arguing as in the equation above) and Assumptions F, X give

$$\begin{aligned}
& \left\| \frac{1}{h_q^{1/2}} \int \left(\frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta_t) \right)^{-1} \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} dK_q(t) \right\| \\
&\leq \frac{O_{\mathbb{P}}(1)}{h_q^{1/2}} \int \left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} \right\| |dK_q(t)| \\
&= \frac{O_{\mathbb{P}}(1)}{h_q^{1/2}} \mathbb{E} \left[\int \left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} \right\| |dK_q(t)| \right] \\
&\leq \frac{O_{\mathbb{P}}(1)}{h_q^{1/2}} \int \mathbb{E}^{1/2} \left[\left\| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} \right\|^2 \right] |dK_q(t)| \\
&= \frac{O_{\mathbb{P}}(1)}{h_q^{1/2}} \int \text{Var}^{1/2} \left(\frac{1}{h^{d/2}} \{\mathbf{S}(\theta) - \mathbf{S}(\theta_t)\} \right) |dK_q(t)| \\
&= O_{\mathbb{P}}(1) \int \left[\int \left(\frac{1}{h_q} \int_{Q^*(x+hz;\theta)-Ch_q}^{Q^*(x+hz;\theta)+Ch_q} f(y|x+hz) dy \right) \mathbb{I}(z \in \mathcal{K}) f(x+hz) dz \right]^{1/2} |dK_q(t)| \\
&= O_{\mathbb{P}}(1). \square
\end{aligned}$$

APPENDIX B: PROOFS OF INTERMEDIARY RESULTS

B.1. Proof of Lemma A.1. Recall

$$\mathbf{U}(X-x)^T \mathbf{b} = \mathbf{U}(X-x)^T \mathbf{H}^{-1} \mathbf{B} = \mathbf{U} \left(\frac{X-x}{h} \right)^T \mathbf{B}$$

and define

$$\tilde{\mathcal{L}}(\mathbf{B}; \theta) = \mathcal{L}(\mathbf{b}; \theta) = \frac{1}{h^d} \mathbb{E} \left[\{ \ell_{\alpha}(Y - \mathbf{U}((X-x)/h)^T \mathbf{B}) - \ell_{\alpha}(Y) \} K_h(X-x) \right].$$

The change of variable $x_1 = x + hz$ gives

$$\begin{aligned} \tilde{\mathcal{L}}(\mathbf{B}; \theta) &= \frac{1}{h^d} \int \left[\int \left(\ell_\alpha(y - \mathbf{U} \left(\frac{x_1 - x}{h} \right)^T \mathbf{B}) - \ell_\alpha(y) \right) f(y|x_1) dy \right] f(x_1) K \left(\frac{x_1 - x}{h} \right) dx_1 \\ (B.1) \quad &= \int \left[\int \left(\ell_\alpha(y - \mathbf{U}(z)^T \mathbf{B}) - \ell_\alpha(y) \right) f(y|x + hz) dy \right] f(x + hz) K(z) dz, \end{aligned}$$

showing that $\tilde{\mathcal{L}}(\mathbf{B}; \theta)$ is also defined for $h = 0$.

Proof of (i). It is sufficient to show that $\mathbf{B}^*(\theta) = \arg \min_{\mathbf{B} \in \mathbb{R}^P} \tilde{\mathcal{L}}(\mathbf{B}; \theta)$ exists and is unique. Note that $\mathbf{B} \mapsto \tilde{\mathcal{L}}(\mathbf{B}; \theta)$ is convex by (B.1) because $\ell_\alpha(\cdot)$ is convex. Since $\lim_{|t| \rightarrow +\infty} \ell_\alpha(t) = +\infty$ and $\mathbf{U}(z)^T \mathbf{B}$ diverges almost everywhere when $\|\mathbf{B}\|$ diverges, (B.1) gives that $\lim_{\|\mathbf{B}\| \rightarrow +\infty} \tilde{\mathcal{L}}(\mathbf{B}; \theta) = +\infty$. Hence $\tilde{\mathcal{L}}(\mathbf{B}; \theta)$ has a minimum. We show that this minimum is unique by showing that $\mathbf{B} \mapsto \tilde{\mathcal{L}}(\mathbf{B}; \theta)$ is strictly convex for all θ in Θ^0 . We compute the first and second \mathbf{B} -derivatives of $\tilde{\mathcal{L}}(\mathbf{B}; \theta)$. Equation (1.2) gives that for almost all \mathbf{B} ,

$$\frac{\partial \ell_\alpha(y - \mathbf{U}(z)^T \mathbf{B})}{\partial \mathbf{B}^T} = 2 (\mathbb{I}(y \leq \mathbf{U}(z)^T \mathbf{B}) - \alpha) \mathbf{U}(z)$$

which is bounded for z in the compact \mathcal{K} . Assumptions F, K and X, the Lebesgue Dominated Convergence Theorem and (B.1) yield that

$$\begin{aligned} \tilde{\mathcal{L}}^{(1)}(\mathbf{B}; \theta) &= \frac{\partial \tilde{\mathcal{L}}(\mathbf{B}; \theta)}{\partial \mathbf{B}^T} = 2 \int \left(\int (\mathbb{I}(y \leq \mathbf{U}(z)^T \mathbf{B}) - \alpha) f(y|x + hz) dy \right) f(x + hz) \mathbf{U}(z) K(z) dz \\ (B.2) \quad &= 2 \int F(\mathbf{U}(z)^T \mathbf{B} | x + hz) f(x + hz) \mathbf{U}(z) K(z) dz - 2\alpha \int f(x + hz) \mathbf{U}(z) K(z) dz. \end{aligned}$$

Applying again the Dominated Convergence Theorem yields that

$$(B.3) \quad \tilde{\mathcal{L}}^{(2)}(\mathbf{B}; \theta) = \frac{\partial^2 \tilde{\mathcal{L}}(\mathbf{B}; \theta)}{\partial \mathbf{B}^T \partial \mathbf{B}} = 2 \int f(\mathbf{U}(z)^T \mathbf{B} | x + hz) f(x + hz) \mathbf{U}(z) \mathbf{U}(z)^T K(z) dz.$$

For all $\mathbf{A} \neq 0$ in \mathbb{R}^P , (B.3), Assumptions F, K X and $x \in \mathcal{X}_0$ give

$$\begin{aligned} \mathbf{A}^T \tilde{\mathcal{L}}^{(2)}(\mathbf{B}; \theta) \mathbf{A} &= 2 \int f(\mathbf{U}(z)^T \mathbf{B} | x + hz) f(x + hz) \mathbf{A}^T \mathbf{U}(z) \mathbf{U}(z)^T \mathbf{A} K(z) dz \\ (B.4) \quad &= 2 \int f(\mathbf{U}(z)^T \mathbf{B} | x + hz) f(x + hz) \left\| \mathbf{U}(z)^T \mathbf{A} \right\|^2 K(z) dz > 0. \end{aligned}$$

Hence $\tilde{\mathcal{L}}^{(2)}(\cdot; \theta)$ is a positive definite symmetric matrix for all θ in Θ^0 and \mathbf{B} in \mathbb{R}^P so that the strictly convex function $\tilde{\mathcal{L}}(\mathbf{B}; \theta)$ achieves its minimum for a unique $\mathbf{B}^*(\theta)$.

Proof of (ii). Consider a fixed \bar{h} to be chosen small enough, and let $\tilde{\Theta}^0$ be the corresponding Θ^0 , which is compact. The proof of (i) yields that $\mathbf{B}^*(\theta)$ is unique for all θ in $\tilde{\Theta}^0$ and is the unique solution of the first-order condition $\tilde{\mathcal{L}}^{(1)}(\mathbf{B}; \theta) = 0$, that is

$$(B.5) \quad \int F(\mathbf{U}(z)^T \mathbf{B} | x + hz) f(x + hz) \mathbf{U}(z) K(z) dz = \alpha \int f(x + hz) \mathbf{U}(z) K(z) dz,$$

see (B.2), so that (A.1) is proved. In particular, $\mathbf{B}^*(\alpha; 0, x)$ is the unique solution of $\tilde{\mathcal{L}}^{(1)}(\mathbf{B}; \alpha, 0, x) = 0$.

If $h = 0$, the first order condition (A.1) is equivalent to

$$\int F(\mathbf{U}(z)^T \mathbf{B}^*(\alpha; 0, x) | x) \mathbf{U}(z) K(z) dz = \alpha \int \mathbf{U}(z) K(z) dz.$$

Let $\mathbf{B}_0^T(\alpha|x) = (Q(\alpha|x), 0, \dots, 0)$ in \mathbb{R}^P . Since $\mathbf{U}(z)^T \mathbf{B}_0(\alpha|x) = Q(\alpha|x)$, $\mathbf{B}_0(\alpha|x)$ satisfies the first-order condition equation above. Hence $\mathbf{B}^*(\alpha; 0, x) = \mathbf{B}_0(\alpha|x)$ by uniqueness.

We now show that $\mathbf{B}^*(\theta)$ is continuously differentiable in θ over $\tilde{\Theta}^0$ and give bounds for $\mathbf{B}^*(\theta)$, $\partial \tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta); \theta) / \partial \theta^T$ and $\tilde{\mathcal{L}}^{(2)}(\mathbf{B}^*(\theta); \theta)$. As shown above, $\mathbf{B} \mapsto \tilde{\mathcal{L}}^{(1)}(\mathbf{B}; \theta)$ is continuously differentiable and $\tilde{\mathcal{L}}^{(2)}(\mathbf{B}; \theta)$ is a symmetric positive definite matrix for all \mathbf{B} in \mathbb{R}^P and so has an inverse. Assumptions F, K and X yield that $F(\mathbf{U}(z)^T \mathbf{B}|x + hz)$ and $f(x + hz)$ are bounded and have bounded θ -partial derivatives over $\tilde{\Theta}^0$ provided \bar{h} is small enough. Hence the Dominated Convergence Theorem and (B.2) yield that $\tilde{\mathcal{L}}^{(1)}(\mathbf{B}; \theta)$ is continuously differentiable in θ over $\tilde{\Theta}^0$. Then the Implicit Function Theorem (see e.g. Zeidler (1985), p.130) and the first-order condition $\tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta); \theta) = 0$ yields that $\mathbf{B}^*(\theta)$ is continuously differentiable in θ over $\tilde{\Theta}^0$, with

$$(B.6) \quad \frac{\partial \mathbf{B}^*(\theta)}{\partial \theta^T} = - \left[\tilde{\mathcal{L}}^{(2)}(\mathbf{B}^*(\theta); \theta) \right]^{-1} \frac{\partial \tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta); \theta)}{\partial \theta^T}.$$

Recall now that $\Theta^0 \subset \tilde{\Theta}^0$ when \bar{h} tends to 0. Hence continuity of $\mathbf{B}^*(\cdot)$, $\partial \tilde{\mathcal{L}}^{(1)}(\cdot, \cdot) / \partial \theta^T$ and compactness of $\tilde{\Theta}^0$ give

$$(B.7) \quad \begin{aligned} \limsup_{\bar{h} \rightarrow 0} \sup_{\theta \in \Theta^0} \|\mathbf{B}^*(\theta) - \mathbf{B}^*(\alpha; 0, x)\| &= 0, \\ \limsup_{\bar{h} \rightarrow 0} \sup_{\theta \in \Theta^0} \left\| \frac{\partial \tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta); \theta)}{\partial \theta^T} - \frac{\partial \tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\alpha; 0, x); \alpha, 0, x)}{\partial \theta^T} \right\| &= 0. \end{aligned}$$

Since the first limit is (A.2), (ii) is proved.

Proof of (iii). We bound the partial derivative (B.6). Observe that (A.2), the expression of $\mathbf{B}^*(\alpha; 0, x)$, the compactness of Θ^0 and Assumption F yield that there is a compact \mathcal{B} such that $\mathbf{B}^*(\theta)$ is in \mathcal{B} for all θ in Θ^0 , provided \bar{h} is small enough. Then (B.3) and (B.4) give that uniformly in θ in Θ^0 ,

$$\tilde{\mathcal{L}}^{(2)}(\mathbf{B}^*(\theta); \theta) \succ C \int_{\mathcal{B}(0,1)} \mathbf{U}(z) \mathbf{U}(z)^T dz.$$

Hence (B.6) and (B.7) give

$$(B.8) \quad \limsup_{\bar{h} \rightarrow 0} \sup_{\theta \in \Theta^0} \left\| \frac{\partial \mathbf{B}^*(\theta)}{\partial \theta^T} \right\| \leq C \left\| \left(\int_{\mathcal{B}(0,1)} \mathbf{U}(z) \mathbf{U}(z)^T dz \right)^{-1} \right\| \limsup_{\bar{h} \rightarrow 0} \sup_{\theta \in \Theta^0} \left\| \frac{\partial \tilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta); \theta)}{\partial \theta^T} \right\| \leq C.$$

Let us now return to the proof of (iii). The differentiability results above yield that $\theta \in \Theta^1 \mapsto Q^*(x'; \theta) = \mathbf{U}((x - x')/h)^T \mathbf{B}^*(\theta)$ is continuously differentiable in θ . We have for all x, x' in \mathcal{X} and $h \geq \underline{h}$,

$$\left\| \mathbf{U} \left(\frac{x - x'}{h} \right) \right\| \leq \frac{C}{\underline{h}^p}, \quad \left\| \frac{\partial}{\partial \theta^T} \mathbf{U} \left(\frac{x - x'}{h} \right) \right\| \leq \frac{C}{\underline{h}^{p+1}}.$$

Hence for \bar{h} small enough, (A.2) and (B.8) yield that for all θ in Θ^1 and x' in \mathcal{X} ,

$$\begin{aligned} \left\| \frac{\partial Q^*(x'; \theta)}{\partial \theta^T} \right\| &= \left\| \left[\frac{\partial}{\partial \theta^T} \mathbf{U} \left(\frac{x - x'}{h} \right)^T \right] \mathbf{B}^*(\theta) + \mathbf{U} \left(\frac{x - x'}{h} \right)^T \frac{\partial \mathbf{B}^*(\theta)}{\partial \theta^T} \right\| \\ &\leq \left\| \frac{\partial}{\partial \theta^T} \mathbf{U} \left(\frac{x - x'}{h} \right) \right\| \|\mathbf{B}^*(\theta)\| + \left\| \mathbf{U} \left(\frac{x - x'}{h} \right) \right\| \left\| \frac{\partial \mathbf{B}^*(\theta)}{\partial \theta^T} \right\| \leq C \underline{h}^{-p} (1 + \underline{h}^{-1}). \end{aligned}$$

The Taylor inequality shows that (iii) is proved.

Proof of (iv). The change of variable $x' = x + hz$ shows that it is sufficient to prove that, for all θ in Θ^0 and z in \mathcal{K} ,

$$f(Q^*(x + hz; \theta)|x + hz) \geq C \text{ with } f(Q^*(x + hz; \theta)|x + hz) = f(\mathbf{U}(z)^T \mathbf{B}^*(\theta)|x + hz),$$

which is true for \bar{h} small enough by (A.2) and under Assumption F which gives that $f(y|x) \geq C > 0$ for y in any compact subset of \mathbb{R} and any x in \mathcal{X}_0 . \square

B.2. Proof of Proposition A.1. The proof of the Proposition uses the two following Lemmas. In what follows, the stochastic processes $R(\cdot; \cdot)$, $\mathbf{R}^1(\cdot; \cdot)$ and $\mathbf{R}^2(\cdot; \cdot)$ have the same distribution than the $R_i(\cdot; \cdot)$, $\mathbf{R}^1(\cdot; \cdot)$ and $\mathbf{R}^2(\cdot; \cdot)$ in (A.9), (A.10) and (A.11). Define also

$$(B.9) \quad \delta(\beta, \theta) = \mathbf{U} \left(\frac{X - x}{h} \right)^T \frac{\beta}{(nh^d)^{1/2}}.$$

Lemma B.1. *Under Assumptions F, K and X, we have*

$$\text{Var}(R(\beta, \epsilon; \theta)) \leq C \frac{\|\epsilon\|^2 (\|\beta\| + \|\epsilon\|)}{n (nh^d)^{1/2}}.$$

Proof of Lemma B.1. Observe $\ell_\alpha(t) = 2 \int_0^t (\alpha - \mathbb{I}(z \leq 0)) dz$. Hence (A.9) and (B.9) yield

$$(B.10) \quad R(\beta, \epsilon; \theta) = 2K_h(X - x) \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} (\mathbb{I}(Y \leq Q^*(X; \theta) + t) - \mathbb{I}(Y \leq Q^*(X; \theta))) dt.$$

The Cauchy-Schwarz inequality give

$$\begin{aligned} R(\beta, \epsilon; \theta)^2 &= 4K_h(X - x)^2 \left(\int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} (\mathbb{I}(Y \leq Q^*(X; \theta) + t) - \mathbb{I}(Y \leq Q^*(X; \theta))) dt \right)^2 \\ &\leq 4K_h(X - x)^2 |\delta(\epsilon, \theta)| \left| \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} (\mathbb{I}(Y \leq Q^*(X; \theta) + t) - \mathbb{I}(Y \leq Q^*(X; \theta)))^2 dt \right| \\ &< 4K_h(X - x)^2 |\delta(\epsilon, \theta)| \left| \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} \mathbb{I}(|Y - Q^*(X; \theta)| < |t|) dt \right|. \end{aligned}$$

Hence Assumption F and (B.9) give

$$\begin{aligned} \mathbb{E}[R^2(\beta, \epsilon; \theta) | X] &\leq 4K_h(X - x)^2 |\delta(\epsilon, \theta)| \left| \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} \left\{ \int \mathbb{I}(|y - Q^*(X; \theta)| < |t|) f(y|X) dy \right\} dt \right| \\ &\leq 4K_h(X - x)^2 \|f(\cdot|\cdot)\|_\infty |\delta(\epsilon, \theta)| \left| 2 \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} |t| dt \right| \\ &\leq CK_h(X - x)^2 \delta(\epsilon, \theta)^2 (|\delta(\beta, \theta)| + |\delta(\epsilon, \theta)|) \\ &\leq C \frac{K^2 \left(\frac{X-x}{h} \right) \|\mathbf{U} \left(\frac{X-x}{h} \right)\|^3}{(nh^d)^{3/2}} \|\epsilon\|^2 (\|\beta\| + \|\epsilon\|). \end{aligned}$$

Then, under Assumptions K and X,

$$\begin{aligned}
\text{Var}(R(\beta, \epsilon; \theta)) &\leq \mathbb{E}[R^2(\beta, \epsilon; \theta)] = \mathbb{E}[\mathbb{E}[R^2(\beta, \epsilon; \theta) | X]] \\
&\leq \frac{C \|\epsilon\|^2 (\|\beta\| + \|\epsilon\|)}{(nh^d)^{3/2}} \int K^2\left(\frac{x' - x}{h}\right) \left\| \mathbf{U}\left(\frac{x' - x}{h}\right) \right\|^3 f_X(x') dx' \\
&\leq \frac{C \|\epsilon\|^2 (\|\beta\| + \|\epsilon\|)}{(nh^d)^{3/2}} h^d \int K^2(z) \|\mathbf{U}(z)\|^3 f_X(x + hz) dx' \leq C \frac{\|\epsilon\|^2 (\|\beta\| + \|\epsilon\|)}{n (nh^d)^{1/2}}. \square
\end{aligned}$$

Define

$$\mathcal{F} = \mathcal{F}(t_\beta, t_\epsilon, \Theta^1) = \{R(\beta, \epsilon; \theta), (\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1\}.$$

The next lemma studies coverings of \mathcal{F} with brackets $[\underline{R}, \overline{R}]$. Recall that the bracket $[\underline{R}, \overline{R}] = [\underline{R}(X, Y), \overline{R}(X, Y)]$ is the set of random variables $r = r(X, Y)$ such that $\underline{R} \leq r \leq \overline{R}$ almost surely.

Lemma B.2. *Under Assumptions F, K and X and if $t_\beta + t_\epsilon \geq 1$ and n is large enough,*

(i) *There are some $\overline{\sigma}^2$ and \overline{w} , with*

$$\overline{\sigma}^2 \asymp \frac{t_\epsilon^2(t_\epsilon + t_\beta)}{n(nh^d)^{1/2}}, \quad \overline{w} \asymp \frac{t_\beta + t_\epsilon}{(nh^d)^{1/2}},$$

such that for all integer number $k \geq 2$, $(\beta, \epsilon, \theta)$ in $\mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1$,

$$\mathbb{E} \left[|R(\beta, \epsilon; \theta) - \mathbb{E}[R(\beta, \epsilon; \theta)]|^k \right] \leq \frac{k!}{2} \overline{w}^{k-2} \overline{\sigma}^2.$$

(ii) *Let τ in $(0, 1)$ be a bracket length. There is an set of brackets $\mathcal{I}_\tau = \{[\underline{R}_{j,\tau}, \overline{R}_{j,\tau}], 1 \leq j \leq e^{H(\tau)}\}$ such that*

$$\mathcal{F} \subset \bigcup_{1 \leq j \leq e^{H(\tau)}} [\underline{R}_{j,\tau}, \overline{R}_{j,\tau}],$$

$$\mathbb{E} \left[|\underline{R}_{j,\tau} - \overline{R}_{j,\tau}|^k \right] \leq \frac{k!}{2} \overline{w}^{k-2} \tau^2 \text{ for all integer number } k \geq 2 \text{ and all } j \text{ in } [1, e^{H(\tau)}],$$

$$H(\tau) \leq C \log \left(\frac{n(t_\beta + t_\epsilon)}{\tau} \right) \text{ for all } \tau, t_\beta \text{ and } t_\epsilon.$$

Proof of Lemma B.2. Define for β in \mathbb{R}^P

$$\tilde{R}(\beta; \theta) = 2K_h(X - x) \int_0^{\delta(\beta, \theta)} (\mathbb{I}(Y \leq Q^*(X; \theta) + u) - \mathbb{I}(Y \leq Q^*(X; \theta))) du.$$

Let $\text{sgn}(t) = \mathbb{I}(t \geq 0) - \mathbb{I}(t < 0)$. Observe that $\tilde{R}(\beta; \theta) \geq 0$ with

$$\begin{aligned}
\tilde{R}(\beta; \theta) &= 2K_h(X - x) \int_0^{|\delta(\beta, \theta)|} |\mathbb{I}(Y \leq Q^*(X; \theta) + \text{sgn}(\delta(\beta, \theta))u) - \mathbb{I}(Y \leq Q^*(X; \theta))| du \\
&= 2K_h(X - x) |\delta(\beta, \theta)| \int_0^1 |\mathbb{I}(Y \leq Q^*(X; \theta) + \delta(\beta, \theta)v) - \mathbb{I}(Y \leq Q^*(X; \theta))| dv \\
\text{(B.11)} \quad &= 2K_h(X - x) |\delta(\beta, \theta)| \int_0^1 \mathbb{I}(Y - Q^*(X; \theta) \text{ lies between } 0 \text{ and } \delta(\beta, \theta)v) dv.
\end{aligned}$$

(B.10) and $\delta(\beta, \theta) + \delta(\epsilon, \theta) = \delta(\beta + \epsilon, \theta)$ give

$$\text{(B.12)} \quad R(\beta; \epsilon, \theta) = \tilde{R}(\beta + \epsilon; \theta) - \tilde{R}(\beta; \theta).$$

It also follows from (B.9) and Assumption K that for all β in $\mathcal{B}(0, t_\beta + t_\epsilon)$ and all θ in Θ^1

$$(B.13) \quad \left| \tilde{R}(\beta; \theta) \right| \leq 2 \left\| \mathbf{U} \left(\frac{X-x}{h} \right) \right\| K \left(\frac{X-x}{h} \right) \frac{\|\beta\|}{(nh^d)^{1/2}} \leq \frac{\bar{w}}{2}, \quad \bar{w} \asymp \frac{t_\beta + t_\epsilon}{(n\bar{h}^d)^{1/2}}.$$

Part (i) follows from Lemma B.1 and (B.12) which give

$$\begin{aligned} & \mathbb{E} \left[|R(\beta, \epsilon; \theta) - \mathbb{E}[R(\beta, \epsilon; \theta)]|^k \right] \\ &= \mathbb{E} \left[\left| \tilde{R}(\beta + \epsilon; \theta) - \mathbb{E}[\tilde{R}(\beta + \epsilon; \theta)] - \left(\tilde{R}(\beta; \theta) - \mathbb{E}[\tilde{R}(\beta; \theta)] \right) \right|^{k-2} |R(\beta, \epsilon; \theta) - \mathbb{E}[R(\beta, \epsilon; \theta)]|^2 \right] \\ &\leq \left(2 \times \frac{\bar{w}}{2} \right)^{k-2} \text{Var}(R(\beta, \epsilon; \theta)) \leq \bar{w}^{k-2} \bar{\sigma}^2. \end{aligned}$$

The proof of part (ii) will be divided in three steps. Let $\tilde{\mathcal{F}}_t$ be $\{\tilde{R}(\beta; \theta), (\beta, \theta) \in \mathcal{B}(0, t) \times \Theta^1\}$. For the sake of brevity we abbreviate $\underline{R}_{j,\tau}, \bar{R}_{j,\tau}$ into $\underline{R}_j, \bar{R}_j$.

Step 1 : Coverings of \mathcal{F} and $\tilde{\mathcal{F}}_t$, $t = t_\beta + t_\epsilon \geq 1$. We show in this step that it is sufficient to find a covering of $\tilde{\mathcal{F}}_t$ with $H(\tau) = H(\tau; t)$ brackets satisfying

$$(B.14) \quad \mathbb{E} \left[|\underline{R}_j - \bar{R}_j|^k \right] \leq \frac{k!}{8} \left(\frac{\bar{w}}{2} \right)^{k-2} \tau^2,$$

$$(B.15) \quad H(t) \leq C \log \left(\frac{nt}{\tau} \right).$$

Indeed, consider two such coverings of $\tilde{\mathcal{F}}_{t_\beta}$ and $\tilde{\mathcal{F}}_{t_\beta+t_\epsilon}$,

$$\tilde{\mathcal{F}}_{t_\beta} \subset \bigcup_{1 \leq j \leq e^{H_1(\tau)}} [\underline{R}_j^1, \bar{R}_j^1], \quad \tilde{\mathcal{F}}_{t_\beta+t_\epsilon} \subset \bigcup_{1 \leq j \leq e^{H_2(\tau)}} [\underline{R}_j^2, \bar{R}_j^2],$$

$H_1(\tau) \leq H_2(\tau) = H(\tau, t_\beta + t_\epsilon)$. Consider a $R(\beta, \epsilon; \theta)$ in \mathcal{F} . Since $\tilde{R}(\beta; \theta) \in [\underline{R}_{j_1}^1, \bar{R}_{j_1}^1]$ and $\tilde{R}(\beta + \epsilon; \theta) \in [\underline{R}_{j_2}^2, \bar{R}_{j_2}^2]$ for some j_1 and j_2 , (B.12) implies that $R(\beta, \epsilon; \theta) \in [\underline{R}_{j_2}^2 - \bar{R}_{j_1}^1, \bar{R}_{j_2}^2 - \underline{R}_{j_1}^1]$. Hence these $e^{H'(\tau)}$ brackets form a covering of \mathcal{F} with, using (B.14) and (B.15),

$$\begin{aligned} \mathbb{E} \left[\left| \bar{R}_{j_2}^2 - \underline{R}_{j_1}^1 - \left(\bar{R}_{j_2}^2 - \underline{R}_{j_1}^1 \right) \right|^k \right] &\leq 2^{k-1} \left(\mathbb{E} \left[\left| \bar{R}_{j_2}^2 - \underline{R}_{j_2}^2 \right|^k \right] + \mathbb{E} \left[\left| \bar{R}_{j_1}^1 - \underline{R}_{j_1}^1 \right|^k \right] \right) \\ &\leq 2^k \frac{k!}{8} \left(\frac{\bar{w}}{2} \right)^{k-2} \tau^2 = \frac{k!}{2} \bar{w}^{k-2} \tau^2, \\ H'(\tau) = H_1(\tau) + H_2(\tau) &\leq C \log \left(\frac{n(t_\beta + t_\epsilon)}{\tau} \right). \end{aligned}$$

Step 2: Preliminary results for the construction of a covering of $\tilde{\mathcal{F}}_t$. We bound the increments of $(\beta, \theta) \mapsto Q^*(X; \theta), K_h(X-x), \delta(\beta, \theta)$. Lemma A.1-(iii) gives that for all θ, θ' in Θ^1

$$|Q^*(X; \theta) - Q^*(X; \theta')| \leq C \bar{h}^{-p} (1 + \bar{h}^{-1}) \|\theta - \theta'\|.$$

Under Assumption K

$$\begin{aligned} \left| K \left(\frac{X-x}{h} \right) - K \left(\frac{X-x'}{h'} \right) \right| &\leq C \left(\left\| \frac{x-x'}{h'} \right\| + \|X-x\| \left| \frac{1}{h} - \frac{1}{h'} \right| \right) \\ &\leq C \left(\frac{1}{\bar{h}} \|x-x'\| + \frac{1}{\bar{h}^2} |h-h'| \right) \leq \frac{C}{\bar{h}^2} \|\theta - \theta'\|. \end{aligned}$$

For the increments of $\delta(\beta, \theta)$, define $\mathbf{U} = \mathbf{U}(X - x)$, $\mathbf{U}' = \mathbf{U}(X - x')$, $\mathbf{H}' = \mathbf{H}(h')$. This gives

$$\begin{aligned}
& |\delta(\beta, \theta) - \delta(\beta', \theta')| \\
&= \left| \mathbf{U}^T \frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} (\beta - \beta') + (\mathbf{U}' - \mathbf{U})^T \frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} \beta' + \mathbf{U}'^T \left(\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} - \frac{\mathbf{H}'^{-1}}{(nh'^d)^{1/2}} \right) \beta' \right| \\
&\leq C \left\| \frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} \right\| \|\beta - \beta'\| + \|x - x'\| \left\| \frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} \right\| \|\beta'\| + C \|\beta'\| \left\| \frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} - \frac{\mathbf{H}'^{-1}}{(nh'^d)^{1/2}} \right\| \\
&\leq \frac{C(1+t)}{\underline{h}^p (n\underline{h}^d)^{1/2}} \left(\|\beta - \beta'\| + \|x - x'\| + \frac{1}{\underline{h}} |h - h'| \right) \leq \frac{C(1+t)}{\underline{h}^{p+1} (n\underline{h}^d)^{1/2}} (\|\beta - \beta'\| + \|\theta - \theta'\|).
\end{aligned}$$

Step 3 : Construction of the covering of $\tilde{\mathcal{F}}_t$. Define

$$\begin{aligned}
\rho(q, \delta) &= |\mathbb{I}(q \leq \delta) - \mathbb{I}(q \leq 0)| = \mathbb{I}(q \in (0, \delta]) \mathbb{I}(\delta \geq 0) + \mathbb{I}(q \in [\delta, 0)) \mathbb{I}(\delta < 0), \\
r(q, \delta) &= \int_0^1 \rho(q, \delta v) dv.
\end{aligned}$$

Hence (B.11) shows

$$\tilde{R}(\beta; \theta) = 2K_h(X - x) |\delta(\beta, \theta)| r(Y - Q^*(X; \theta), \delta(\beta, \theta)).$$

For any $\eta > 0$, there exists functions $\underline{\rho}(q, \delta) = \underline{\rho}_\eta(q, \delta)$ and $\bar{\rho}(q, \delta) = \bar{\rho}_\eta(q, \delta)$ and an open set $D = D_\eta \subset \mathbb{R}^2$ such that

$$\begin{aligned}
\rho - \text{(i)} \quad & 0 \leq \underline{\rho}(q, \delta) \leq \rho(q, \eta) \leq \bar{\rho}(q, \delta) \leq 1 \text{ for all } (q, \delta), \text{ with } \underline{\rho}(q, \delta) = \rho(q, \eta) = \bar{\rho}(q, \delta) \text{ if } (q, \delta) \in \mathbb{R}^2 \setminus D_\eta, \\
\rho - \text{(ii)} \quad & \sup_{(q, \delta) \in D_\eta} \left(\left| \frac{\partial \underline{\rho}(q, \delta)}{\partial q} \right| + \left| \frac{\partial \underline{\rho}(q, \delta)}{\partial \delta} \right| + \left| \frac{\partial \bar{\rho}(q, \delta)}{\partial q} \right| + \left| \frac{\partial \bar{\rho}(q, \delta)}{\partial \delta} \right| \right) \leq C\eta^{-1/2}, \\
\rho - \text{(iii)} \quad & D \subset D' = \{(q, \delta) \in \mathbb{R}^2; |q| \leq C\eta^{-1/2} \text{ or } |q - \delta| \leq C\eta^{-1/2}\}.
\end{aligned}$$

Define $\underline{r}(q, \delta) = \int_0^1 \underline{\rho}(q, v\delta) dv$, $\bar{r}(q, \delta) = \int_0^1 \bar{\rho}(q, v\delta) dv$ and

$$\begin{aligned}
\underline{R}(\beta, \theta) &= 2K_h(X - x) |\delta(\beta, \theta)| \underline{r}(Y - Q^*(X; \theta), \delta(\beta, \theta)), \\
\bar{R}(\beta, \theta) &= 2K_h(X - x) |\delta(\beta, \theta)| \bar{r}(Y - Q^*(X; \theta), \delta(\beta, \theta)).
\end{aligned}$$

Since $K(\cdot) \geq 0$, ρ -(i) gives that these functions are such that

$$(B.16) \quad \underline{R}(\beta, \theta) \leq \tilde{R}(\beta, \theta) \leq \bar{R}(\beta, \theta).$$

We now bound $\underline{R}(\beta, \theta) - \underline{R}(\beta', \theta')$ and $\bar{R}(\beta, \theta) - \bar{R}(\beta', \theta')$. We have

$$\begin{aligned}
& |\underline{R}(\beta, \theta) - \underline{R}(\beta', \theta')| \leq 2|K_h(X - x) - K_{h'}(X - x')| |\delta(\beta, \theta)| \underline{r}(Y - Q^*(X; \theta), \delta(\beta, \theta)) \\
& \quad + 2K_{h'}(X - x') |\delta(\beta, \theta) - \delta(\beta', \theta')| \underline{r}(Y - Q^*(X; \theta), \delta(\beta, \theta)) \\
& \quad + 2K_{h'}(X - x') |\delta(\beta', \theta')| |\underline{r}(Y - Q^*(X; \theta), \delta(\beta, \theta)) - \underline{r}(Y - Q^*(X; \theta'), \delta(\beta', \theta'))|.
\end{aligned}$$

Hence Step 1, ρ -(i,ii), (B.9) and the Taylor inequality give for all $(\beta, \theta), (\beta', \theta')$ in $\mathcal{B}(0, t) \times \Theta^1$, provided n is large enough,

$$\begin{aligned} |\underline{R}(\beta, \theta) - \underline{R}(\beta', \theta')| &\leq C \left[\frac{t}{\underline{h}^p (n\underline{h}^d)^{1/2}} \frac{\|\theta - \theta'\|}{\underline{h}^2} + \frac{1+t}{\underline{h}^{p+1} (n\underline{h}^d)^{1/2}} (\|\theta - \theta'\| + \|\beta - \beta'\|) \right] \\ &\quad + C\eta^{-1/2} \left[\frac{\|\theta - \theta'\|}{\underline{h}^{p+1}} + \frac{1+t}{\underline{h}^{p+1} (n\underline{h}^d)^{1/2}} (\|\theta - \theta'\| + \|\beta - \beta'\|) \right] \\ &\leq C \frac{(1 + \eta^{-1/2})(1+t)}{\underline{h}^{p+2}} (\|\theta - \theta'\| + \|\beta - \beta'\|). \end{aligned}$$

Arguing symmetrically gives

$$|\overline{R}(\beta, \theta) - \overline{R}(\beta', \theta')| \leq C \frac{(1 + \eta^{-1/2})(1+t)}{\underline{h}^{p+2}} (\|\theta - \theta'\| + \|\beta - \beta'\|).$$

We now construct the brackets. Recall that there is a covering of $\mathcal{B}(0, t) \times \Theta^1$ with N balls $\mathcal{B}((\beta_j, \theta_j), \eta)$, $\theta_j = (\alpha_j, h_j, x_j)$, with center (β_j, θ_j) and radius η such that

$$(B.17) \quad N \leq \max \left(1, \frac{Ct^P}{\eta^{P+d+2}} \right),$$

see van de Geer (1999, p.20). Define

$$\underline{R}'_j = \underline{R}(\beta_j, \theta_j) - C\eta \frac{(1 + \eta^{-1/2})(1+t)}{\underline{h}^{p+2}}, \quad \overline{R}'_j = \overline{R}(\beta_j, \theta_j) + C\eta \frac{(1 + \eta^{-1/2})(1+t)}{\underline{h}^{p+2}},$$

$$(B.18) \quad \underline{R}_j = \max(0, \underline{R}'_j), \quad \overline{R}_j = \min\left(\frac{\overline{w}}{2}, \overline{R}'_j\right).$$

Bounding $\underline{R}(\beta, \theta) - \underline{R}_j$ and $\overline{R}(\beta, \theta) - \overline{R}_j$ for (β, θ) in $\mathcal{B}((\beta_j, \theta_j), \eta)$, (B.16) and (B.13) give

$$(B.19) \quad \underline{R}'_j \leq \underline{R}_j \leq \tilde{R}(\beta, \theta) \leq \overline{R}_j \leq \overline{R}'_j.$$

It then follows that $\{[\underline{R}_j, \overline{R}_j], j = 1, \dots, N\}$ is a covering of $\tilde{\mathcal{F}}_t$ with, since $0 \leq \underline{R}_j \leq \overline{R}_j \leq \overline{w}/2$,

$$(B.20) \quad |\overline{R}_j - \underline{R}_j| \leq \frac{\overline{w}}{2} \asymp C \frac{t}{(n\underline{h}^d)^{1/2}}.$$

We now bound $\mathbb{E} \left[(\overline{R}_j - \underline{R}_j)^2 \right]$ and $\mathbb{E} \left[|\overline{R}_j - \underline{R}_j|^k \right]$. (B.19), ρ -(i,iii), (B.9) and Assumptions F, K give

$$\begin{aligned}
\mathbb{E} \left[(\overline{R}_j - \underline{R}_j)^2 \right] &\leq \mathbb{E} \left[(\overline{R}'_j - \underline{R}'_j)^2 \right] \leq 2\mathbb{E} \left[(\overline{R}(\beta_j, \theta_j) - \underline{R}(\beta_j, \theta_j))^2 \right] + C\eta^2 \frac{(1 + \eta^{-1/2})^2 (1 + t)^2}{\underline{h}^{2(p+2)}} \\
&\leq 8\mathbb{E} \left[K_{h_j}^2 (X - x_j) \delta^2(\beta_j, \theta_j) (\overline{r}(Y - Q^*(X; \theta_j), \delta(\beta_j, \theta_j)) - \underline{r}(Y - Q^*(X; \theta_j), \delta(\beta_j, \theta_j)))^2 \right] \\
&\quad + C \frac{(1+t)^2}{\underline{h}^{2(p+2)}} (\eta^2 + \eta) \\
&\leq 8\mathbb{E} \left[K_{h_j}^2 (X - x_j) \delta^2(\beta_j, \theta_j) \int \left(\int_0^1 \mathbb{I}((y - Q^*(X; \theta_j), v\delta(\beta_j, \theta_j)) \in D) dv \right)^2 f(y|X) dy \right] \\
&\quad + C \frac{(1+t)^2}{\underline{h}^{2(p+2)}} (\eta^2 + \eta) \\
&\leq \frac{8h_j^d \|\beta\|^2}{nh_j^d} \int K^2(z) \|\mathbf{U}(z)\|^2 \\
&\quad \times \left[\int \int_0^1 \mathbb{I}((y - Q^*(x_j + h_j z; \theta_j), v\delta(\beta_j, \theta_j)) \in D) dv f(y|x_j + h_j z) dy \right] f(x_j + h_j z) dz \\
&\quad + C \frac{(1+t)^2}{\underline{h}^{2(p+2)}} (\eta^2 + \eta) \\
&\leq C \frac{(1+t)^2}{\underline{h}^{2(p+2)}} (\eta^2 + \eta + \eta^{1/2}).
\end{aligned}$$

This together with (B.20) give for any integer number $k \geq 2$

$$\mathbb{E} \left[|\overline{R}_j - \underline{R}_j|^k \right] \leq \left(\frac{\overline{w}}{2} \right)^{k-2} \mathbb{E} \left[(\overline{R}_j - \underline{R}_j)^2 \right] \leq \frac{k!}{8} \left(\frac{\overline{w}}{2} \right)^{k-2} \times C \frac{(1+t)^2}{\underline{h}^{2(p+2)}} (\eta^2 + \eta + \eta^{1/2}).$$

Hence (B.14) holds if η satisfies

$$\eta = \frac{C}{3} \min \left(\left(\frac{\underline{h}^{2(p+2)}}{(1+t)^2} \right)^{1/2} \tau, \frac{\underline{h}^{2(p+2)}}{(1+t)^2} \tau^2, \left(\frac{\underline{h}^{2(p+2)}}{(1+t)^2} \right)^2 \tau^4 \right).$$

Recall now that $\tau < 1$, $t \geq 1$ and that $\underline{h} \geq Cn^{-1/d}$ under Assumption K. The bound (B.17) for $N = \exp(H(\tau))$ gives taking η as above

$$\begin{aligned}
e^{H(\tau)} &\leq \max \left(1, \frac{Ct^P}{\min \left(\left(\frac{\underline{h}^{2(p+2)}}{(1+t)^2} \right)^{1/2} \tau, \frac{\underline{h}^{2(p+2)}}{(1+t)^2} \tau^2, \left(\frac{\underline{h}^{2(p+2)}}{(1+t)^2} \right)^2 \tau^4 \right)^{P+d+2}} \right) \\
&\leq \max \left(1, \frac{Ct^3 n^{(4p+4)/d}}{\tau} \right)^{P+d+2}.
\end{aligned}$$

It then follows for n large enough

$$H(\tau) \leq (P + d + 2) \max \left(0, \log \left(\frac{Ct^3 n^{(4p+4)/d}}{\tau} \right) \right) = C \left(3 \log t + \frac{4p+4}{d} \log n - \log \tau \right) \leq C \log \left(\frac{tn}{\tau} \right),$$

and (B.15) is proved. This ends the proof of the Lemma. \square

Let us now return to the proof of Proposition A.1. Define $\mathbb{X} = (X_1, \dots, X_n)$. The definition of \mathbb{R}_n^1 and (A.10) give

$$\begin{aligned}
& \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^1(\beta, \epsilon; \theta)| \right] \\
&= \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \left| \sum_{i=1}^n (R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta) | \mathbb{X}]) \right| \right] \\
&\leq \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \left| \sum_{i=1}^n (R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta)]) \right| \right] \\
&\quad + \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \left| \mathbb{E} \left[\sum_{i=1}^n (R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta)]) | \mathbb{X} \right] \right| \right] \\
&\leq 2\mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \left| \sum_{i=1}^n (R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta)]) \right| \right].
\end{aligned}$$

Let $H(\cdot)$, $\bar{\sigma}$ and \bar{w} be as in Lemma B.2. Recall that $t_\beta + t_\epsilon \geq 1$ and that $\bar{\sigma} < 1 \leq n(t_\beta + t_\epsilon)$ for n large enough under the assumptions for t_β and t_ϵ of the Proposition. It follows from Massart (2007, Theorem 6.8) that

$$\mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \left| \sum_{i=1}^n (R_i(\beta, \epsilon; \theta) - \mathbb{E}[R_i(\beta, \epsilon; \theta)]) \right| \right] \leq C \left(n^{1/2} \int_0^{\bar{\sigma}} H(u)^{1/2} du + (\bar{w} + \bar{\sigma}) H(\bar{\sigma}) \right).$$

Since $\bar{\sigma} < 1$, Lemma B.2 gives, for all u in $(0, \bar{\sigma}]$, $H(u) \leq C \log(n(t_\beta + t_\epsilon)/u)$. This gives

$$\begin{aligned}
n^{1/2} \int_0^{\bar{\sigma}} H^{1/2}(u) du &\leq (n\bar{\sigma})^{1/2} \left(\int_0^{\bar{\sigma}} H(u) du \right)^{1/2} \leq C(n\bar{\sigma})^{1/2} \left(\int_0^{\bar{\sigma}} \log \left(\frac{n(t_\beta + t_\epsilon)}{u} \right) du \right)^{1/2} \\
&= C(n\bar{\sigma})^{1/2} \left(\bar{\sigma} \left(\log \left(\frac{(t_\beta + t_\epsilon)n}{\bar{\sigma}} \right) + 1 \right) \right)^{1/2} \leq Cn^{1/2} \bar{\sigma} \log^{1/2} \left(\frac{(t_\beta + t_\epsilon)n}{\bar{\sigma}} \right).
\end{aligned}$$

The order for $\bar{\sigma}$ given in Lemma B.2, assumption on $t_\beta + t_\epsilon$ and Assumption K give

$$\log(n(t_\beta + t_\epsilon)/\bar{\sigma}) \leq C \log \left(\frac{n^{3/2} (nh^d)^{1/4} (t_\beta + t_\epsilon)^{1/2}}{t_\epsilon} \right) \leq C \log \left(\frac{n^{3/2} (nh^d)^{1/2}}{\log^{1/2} n} \right) \leq C \log n.$$

Substituting gives

$$\begin{aligned}
& \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^1(\beta, \epsilon; \theta)| \right] \leq C \left(n^{1/2} \bar{\sigma} \log^{1/2} n + (\bar{\sigma} + \bar{w}) \log n \right) \\
&\leq C \frac{t_\epsilon (t_\beta + t_\epsilon)^{1/2}}{(nh^d)^{1/4}} \log^{1/2} n \left(1 + \log^{1/2} n \left(\frac{1}{n^{1/2}} + \frac{(t_\beta + t_\epsilon)^{1/2}}{t_\epsilon (nh^d)^{1/4}} \right) \right) \leq C \frac{t_\epsilon (t_\beta + t_\epsilon)^{1/2}}{(nh^d)^{1/4}} \log^{1/2} n. \square
\end{aligned}$$

B.3. Proof of Proposition A.2. The proof of Proposition A.2 follows the same steps of the proof of Proposition A.1 and we only sketch it. The integral expression of $R(\beta, \epsilon; \theta)$ in (B.10) and the expression (A.11) of $\mathbf{R}^2(\beta, \epsilon; \theta)$ give

$$\mathbf{R}^2(\beta, \epsilon; \theta) = 2K_h(X - x) \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} (F(Q^*(X; \theta) + u|X) - F(Q^*(X; \theta)|X)) du - \frac{1}{2nh^d} \epsilon^T \mathbf{J}(\theta)(\epsilon + 2\beta).$$

The definition (3.6) of $\mathbf{J}(\theta)$ gives

$$\begin{aligned} \mathbf{R}^2(\beta, \epsilon; \theta) &= 2K_h(X-x) \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} (F(Q^*(X; \theta) + u|X) - F(Q^*(X; \theta)|X) - uf(Q^*(X; \theta)|X)) du \\ &= 2K_h(X-x) \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} u \left\{ \int_0^1 (f(Q^*(X; \theta) + vu|X) - f(Q^*(X; \theta)|X)) dv \right\} du. \end{aligned}$$

Define

$$r(\beta; \theta) = 2K_h(X-x) \int_0^{\delta(\beta, \theta)} u \left\{ \int_0^1 (f(Q^*(X; \theta) + vu|X) - f(Q^*(X; \theta)|X)) dv \right\} du$$

which is such that $\mathbf{R}^2(\beta, \epsilon; \theta) = r(\beta + \epsilon; \theta) - r(\beta; \theta)$. Since $|f(q+v|x) - f(q|x)| \leq L_0|v|$ under Assumption F, (B.9) gives

$$\begin{aligned} |\mathbf{R}^2(\beta, \epsilon; \theta)| &\leq K_h(X-x)L_0 \left| \int_{\delta(\beta, \theta)}^{\delta(\beta, \theta) + \delta(\epsilon, \theta)} u^2 du \right| \leq CK_h(X-x)|\delta(\epsilon, \theta)|(|\delta(\beta, \theta)| + |\delta(\epsilon, \theta)|)^2 \\ (B.21) \quad &\leq C \left\| \mathbf{U} \left(\frac{X-x}{h} \right) \right\|^3 K \left(\frac{X-x}{h} \right) \frac{\|\epsilon\|(\|\beta\| + \|\epsilon\|)^2}{(nh^d)^{3/2}}, \\ |r(\beta; \theta)| &\leq CK_h(X-x)|\delta(\beta, \theta)|^3 \leq C \left\| \mathbf{U} \left(\frac{X-x}{h} \right) \right\|^{3/2} K \left(\frac{X-x}{h} \right) \frac{\|\beta\|^3}{(nh^d)^{3/2}}. \end{aligned}$$

The latter inequality gives for all β in $\mathcal{B}(0, t_\beta + t_\epsilon)$ and all θ in Θ^1

$$|r(\beta; \theta)| \leq \frac{\bar{w}'}{2}, \quad \bar{w}' \asymp \frac{(t_\beta + t_\epsilon)^3}{(nh^d)^{3/2}}.$$

It follows from (B.21) that, for all (β, ϵ) in $\mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon)$,

$$\begin{aligned} \text{Var}(\mathbf{R}^2(\beta, \epsilon; \theta)) &\leq \mathbb{E}[\mathbf{R}^2(\beta, \epsilon; \theta)^2] \\ &\leq C \left(\frac{\|\epsilon\|(\|\beta\| + \|\epsilon\|)^2}{(nh^d)^{3/2}} \right)^2 \int \left\| \mathbf{U} \left(\frac{x'-x}{h} \right) \right\|^4 K^2 \left(\frac{x'-x}{h} \right) f(x') dx' \\ &\leq C \frac{\|\epsilon\|^2(\|\beta\| + \|\epsilon\|)^4}{(nh^d)^3} h^d \int \|\mathbf{U}(z)\|^4 K^2(z) dz \leq (\bar{\sigma}')^2, \quad \bar{\sigma}' \asymp \frac{t_\epsilon(t_\beta + t_\epsilon)^2}{n^{3/2}h^d}. \end{aligned}$$

Then constructing brackets as in Lemma B.2 and arguing as in the proof of Proposition A.1 give

$$\begin{aligned} &\mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^2(\beta, \epsilon; \theta) - \mathbb{E}[\mathbb{R}_n^2(\beta, \epsilon; \theta)]| \right] \\ &\leq n^{1/2} \bar{\sigma}' \log^{1/2} \left(\frac{n(t_\beta + t_\epsilon)}{\bar{\sigma}'} \right) + (\bar{\sigma}' + \bar{w}') \log \left(\frac{n(t_\beta + t_\epsilon)}{\bar{\sigma}'} \right). \end{aligned}$$

Since (B.21) yields for all $(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1$

$$\begin{aligned} |\mathbb{E}[\mathbb{R}_n^2(\beta, \epsilon; \theta)]| &= |n\mathbb{E}[\mathbf{R}^2(\beta, \epsilon; \theta)]| \leq Cn\mathbb{E} \left[\left\| \mathbf{U} \left(\frac{X-x}{h} \right) \right\|^3 K \left(\frac{X-x}{h} \right) \frac{\|\epsilon\|(\|\beta\| + \|\epsilon\|)^2}{(nh^d)^{3/2}} \right] \\ &\leq C \frac{t_\epsilon(t_\beta + t_\epsilon)^2}{(nh^d)^{1/2}}, \end{aligned}$$

substituting gives, using $t_\beta \geq 1$, $t_\beta/t_\epsilon = O\left(n\underline{h}^d/\log^{1/2}n\right)$ and Assumption K which ensures $\log\left(n^{5/2}\underline{h}^d/t_\epsilon\right) = O(\log n)$,

$$\begin{aligned} & \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} |\mathbb{R}_n^2(\beta, \epsilon; \theta)| \right] \\ & \leq \mathbb{E} \left[\sup_{(\beta, \epsilon, \theta) \in \mathcal{B}(0, t_\beta) \times \mathcal{B}(0, t_\epsilon) \times \Theta^1} \{ |\mathbb{R}_n^2(\beta, \epsilon; \theta) - \mathbb{E}[\mathbb{R}_n^2(\beta, \epsilon; \theta)]| + \mathbb{E}[\mathbb{R}_n^2(\beta, \epsilon; \theta)] \} \right] \\ & \leq C \frac{t_\epsilon(t_\beta + t_\epsilon)^2}{n\underline{h}^d} \left(1 + \frac{t_\beta + t_\epsilon}{t_\epsilon(n\underline{h}^d)^{1/2}} \right) \log^{1/2} \left(\frac{n^{5/2}\underline{h}^d}{t_\epsilon(t_\beta + t_\epsilon)} \right) + C \frac{t_\epsilon(t_\epsilon + t_\beta)^2}{(n\underline{h}^d)^{1/2}} \leq \frac{t_\epsilon(t_\epsilon + t_\beta)^2}{(n\underline{h}^d)^{1/2}}. \square \end{aligned}$$

B.4. Proof of Lemma A.2. Lemma A.1 (iv) and Assumptions K and F give that there is a $C > 0$ such that for all θ in Θ^1 and all i ,

$$\mathbf{J}_i(\theta) \succ C \mathbf{M}_i(\theta), \quad \mathbf{M}_i(\theta) = 2K_h(X_i - x) \mathbf{U} \left(\frac{X_i - x}{h} \right) \mathbf{U} \left(\frac{X_i - x}{h} \right)^T.$$

Hence for all θ in Θ^1 ,

$$(B.22) \quad \frac{1}{nh^d} \sum_{i=1}^n \mathbf{J}_i(\theta) \succ \frac{C}{nh^d} \sum_{i=1}^n \mathbf{M}_i(\theta) = \mathbb{M}_n(\theta).$$

The entries of $\mathbb{M}_n(\theta)$ write

$$\frac{C}{nh^d} \sum_{i=1}^n \left(\frac{X_i - x}{h} \right)^{\mathbf{v}_1 + \mathbf{v}_2} K \left(\frac{X_i - x}{h} \right), \quad 0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p.$$

Let $M(\theta)$ be the matrix with entries

$$\frac{C}{h^d} \mathbb{E} \left[\left(\frac{X - x}{h} \right)^{\mathbf{v}_1 + \mathbf{v}_2} K \left(\frac{X - x}{h} \right) \right], \quad 0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p.$$

Arguing as in the proof of Proposition A.1 for each of the entries of $\mathbb{M}_n(\theta)$ gives

$$\sup_{\theta \in \Theta^1} \|\mathbb{M}_n(\theta) - M(\theta)\| = o_{\mathbb{P}}(1).$$

Assumptions K, F and X give, for all \mathbf{u} in \mathbb{R}^P , all x in \mathcal{X}_0 and \underline{h} small enough,

$$\begin{aligned} \mathbf{u}^T M(\theta) \mathbf{u} &= \frac{C}{h^d} \mathbb{E} \sum_{0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p} u_{\mathbf{v}_1} u_{\mathbf{v}_2} \left(\frac{X - x}{h} \right)^{\mathbf{v}_1 + \mathbf{v}_2} K \left(\frac{X - x}{h} \right) \\ &= C \sum_{0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p} u_{\mathbf{v}_1} u_{\mathbf{v}_2} \int z^{\mathbf{v}_1 + \mathbf{v}_2} K(z) f(x + hz) dz = C \int \left(\sum_{0 \leq |\mathbf{v}| \leq p} u_{\mathbf{v}} z^{\mathbf{v}} \right)^2 K(z) f(x + hz) dz \\ &\geq C \int_{\mathcal{B}(0,1)} \left(\sum_{0 \leq |\mathbf{v}| \leq p} u_{\mathbf{v}} z^{\mathbf{v}} \right)^2 dz \geq C \|\mathbf{u}\|^2, \end{aligned}$$

where the last bound uses the fact that

$$\mathbf{u} \mapsto \left(\int_{\mathcal{B}(0,1)} \left(\sum_{0 \leq |\mathbf{v}| \leq p} u_{\mathbf{v}} z^{\mathbf{v}} \right)^2 dz \right)^{1/2}$$

is a norm and that norms over \mathbb{R}^P are equivalent. Hence (B.22) and $\|\mathbb{M}_n(\theta) - M(\theta)\| = o_{\mathbb{P}}(1)$ yield that there is a $\underline{\gamma} > 0$ such that $\inf_{\theta \in \Theta^1} \underline{\gamma}_n(\theta) \geq \inf_{\theta \in \Theta^1} \inf_{\|\mathbf{u}\|=1} \mathbf{u}^T \mathbb{M}_n(\theta) \mathbf{u} \geq \underline{\gamma} + o_{\mathbb{P}}(1)$. \square

B.5. Proof of Lemma A.3. The first order condition (A.1) implies that $\mathbb{E}[\mathbf{S}_i(\theta)] = 0$. Consider the \mathbf{v} coordinate of $\mathbf{S}_i(\theta)$,

$$\mathbf{S}_{\mathbf{v},i}(\theta) = 2 \{ \mathbb{I}(Y_i \leq Q^*(X_i; \theta)) - \alpha \} \left(\frac{X_i - x}{h} \right)^{\mathbf{v}} K \left(\frac{X_i - x}{h} \right).$$

Hence Assumptions K and X give, uniformly in $\theta \in \Theta^1$ and for all i ,

$$\begin{aligned} \left| \frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}} \right| &\leq \bar{w}'', \quad \bar{w}'' \asymp (nh^d)^{-1/2}, \\ \text{Var} \left(\frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}} \right) &\leq \mathbb{E} \left[\left(\frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}} \right)^2 \right] \leq \mathbb{E} \left[\left(\frac{\left(\frac{X_i - x}{h} \right)^{\mathbf{v}} K \left(\frac{X_i - x}{h} \right)}{(nh^d)^{1/2}} \right)^2 \right] = \frac{h^d}{nh^d} \int (z^{\mathbf{v}} K(z))^2 \\ &\leq (\bar{\sigma}'')^2, \quad \bar{\sigma}'' \asymp n^{-1/2}. \end{aligned}$$

Hence arguing as in the proof of Proposition A.1 gives, under Assumption K,

$$\mathbb{E} \left[\sup_{\theta \in \Theta^1} \left| \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^n \mathbf{S}_{\mathbf{v},i}(\theta) \right| \right] = O \left(n^{1/2} \bar{\sigma}'' \log^{1/2} n + (\bar{\sigma}'' + \bar{w}'') \log^{1/2} n \right) = O \left(\log^{1/2} n \right).$$

The Markov inequality then shows that the Lemma is proved. \square