Contribution to the study of local polynomial M-estimators

Camille Sabbah

Thèse de doctorat
de l'Université Pierre et Marie Curie

Ecole doctorale de Sciences mathématiques de Paris-Centre (ED 386)

Spécialité : Statistique Mathématique

Présentée par : M. Camille Sabbah

Pour obtenir le grade de

Docteur de l'Université
Pierre et Marie Curie

Sujet de la thèse : Contribution à l'étude des M-estimateurs polynômes locaux

Soutenue publiquement le ../../.... devant le jury composé de :

- M. Emmanuel Guerre, Directeur de thèse
- M. Gérard Biau, Examinateur
- M. Pascal Lavergne, Examinateur
- M. Enno Mammen, Rapporteur
- M. Valentin Patilea, Rapporteur
- M. Wolfgang K. Härdle, Rapporteur

Table des matières

Liste de notations

X désigne une variable aléatoire explicative dans $\mathbb{R}^d$.

Y désigne une variable aléatoire réponse dans $\mathbb{R}$.

$f(\cdot) : \mathbb{R}^d \to \mathbb{R}$ désigne la densité de X.

Q_α désigne le quantile d'ordre α dans $]0, 1[$ de X.

$f(\cdot, \cdot) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ désigne la densité du couple (Y, X).

$f(\cdot|x)$ désigne la densité conditionnelle de Y sachant $X = x$.

$F(\cdot|x)$ désigne la fonction de distribution conditionnelle de Y sachant $X = x$.

$Q(\alpha|x)$ désigne le quantile conditionnel d'ordre α dans $]0, 1[$ de Y sachant $X = x$.

$\mathbb{I}(A)$ désigne la fonction indicatrice de A, c'est-à-dire que $\mathbb{I}(A) = 1$ si A est vrai et 0 sinon.

$(X_i, Y_i)_{i=1}^n$ désigne un échantillon de n variables indépendantes et identiquement distribuées ayant la loi du couple (X, Y).

Soient $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ deux suites réelles telles que pour tout $n \geq 1$, $b_n \neq 0$.

$a_n = O(b_n)$ si et seulement si la suite $(a_n/b_n)_{n\geq 1}$ est bornée.

$a_n = o(b_n)$ si et seulement si $a_n/b_n \to 0$ lorsque $n \to \infty$.

Soient A une variable aléatoire et $(A_n)_{n\geq 1}$ et $(B_n)_{n\geq 1}$ deux suites de variables aléatoires réelles telles que pour tout $n \geq 1$, on ait $B_n \neq 0$ presque sûrement.

$A_n = O_{\mathbb{P}}(B_n)$ si et seulement si pour tout $\epsilon > 0$, $\exists N$, $\exists M$, tels que $\mathbb{P}(|A_n/B_n| \geq M) < \epsilon$ pour tout $n \geq N$.

$A_n = o_{\mathbb{P}}(B_n)$ si et seulement si pour tout $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|A_n/B_n| \geq \epsilon) = 0$.

$A_n \Rightarrow A$ si et seulement si $\mathbb{P}(A_n \leq x) \to \mathbb{P}(A \leq x) = F(x)$ pour tout point de continuité de $F(\cdot)$.

$N(\mu, \sigma^2)$ désigne la variable aléatoire gaussienne d'espérance μ et de variance σ^2.

Pour tout réel z, $\lfloor z \rfloor$ désigne le plus grand entier inférieur où égal à z.

Chapitre 1

Introduction

Les méthodes de régression sont très utilisées en statistiques. Ces modèles permettent d'étudier l'effet d'une variable aléatoire (v.a.), dite explicative, sur une autre variable aléatoire, dite observée. Parmi ces modèles, le plus simple semble être le modèle linéaire. La régression linéaire suppose que la v.a. observée Y est une fonction affine de la v.a. explicative X. En d'autres termes, le modèle linéaire suppose l'existence de deux paramètres a, b et d'une variable ε dite d'erreur tels que $Y = a + bX + \varepsilon$.

Le premier modèle de régression statistique a été la régression quantile. Ce modèle est issu des travaux sur la forme elliptique de la Terre de Boscovitch (voir Stigler (1986)) au milieu du 18ème siècle. En comparaison, il faudra attendre près d'un demi siècle pour voir les premiers travaux sur la méthode des moindres carrés initiée par Legendre. Par la suite, la méthode de Legendre, étudiée et améliorée par Gauss, sera la plus utilisée des deux. Une présentation historique de ces résultats statistiques est disponible dans Stigler (1986). Plusieurs raisons peuvent expliquer pourquoi la méthode des moindres carrés fût préférée. Une première explication possible tient dans l'optimalité de l'estimateur des moindres carrés au sens de l'efficacité de la borne de Cramer-Rao lorsque le terme d'erreur est normalement distribué. Une seconde explication possible est que l'estimateur des moindres carrés a une forme explicite. Cela signifie que son étude peut être menée directement tandis que l'estimateur obtenu par minimisation du contraste

L_1 n'a pas d'expression directe.

Néanmoins, outre les propriétés bien connues de robustesse de la médiane empirique, cette dernière est parfois plus efficace que l'estimateur des moindres carrés, notamment lorsque l'on cherche à estimer le paramètre de centralité d'une varaible aléatoire suivant la loi de Laplace. De plus Bahadur en 1966 fût le premier à exprimer l'estimateur du quantile comme une somme de v.a. (liée à la fonction de répartition empirique) moyennant un terme d'erreur relativement petit. Cette expression, dite de Bahadur, permet d'étudier cet estimateur avec les méthodes classiques des processus empiriques. Cela remédie partiellement au fait que cet estimateur n'ait pas forcément de forme explicite.

Dans une première partie, nous présentons quelques résultats concernant l'estimation de quantiles. Après avoir défini les quantiles, nous donnons les propriétés asymptotiques d'estimateurs de quantiles ainsi que leur représentation de Bahadur. Puis nous présentons la régression quantile introduite par Koenker et Basset. Nous donnons quelques avantages tirés de l'utilisation de ce type de modèle de régression par rapport notamment aux modèles de régression classiques. Nous étendons ensuite ce modèle de régression au cadre non-paramétrique en présentant essentiellement le résultat de Chaudhuri (1991).

Dans la deuxième partie de l'introduction, nous étendons les résultats de la première en donnant une brève introduction à la théorie de la M-estimation. En effet, l'estimation des quantiles peut être vue comme un cas particulier de M-estimation. Nous présentons donc quelques résultats asymptotiques des M-estimateurs dans un cadre paramétrique puis non-paramétrique.

Enfin nous terminons cette introduction en précisant quelques apports de cette thèse. L'objet de cette thèse est d'établir des résultats asymptotiques pour l'estimateur du quantile conditionnel par la méthode des polynômes locaux ainsi qu'à la généralisation de ces résultats pour les M-estimateurs. Nous étudions ces estimateurs et plus particulièrement leur représentation de Bahadur et leur biais. Nous donnons en outre un résultat sur les intervalles de confiance uniformes construits à partir de cette représentation pour le quantile conditionnel et ses dérivées.

1.1 Quantile et régression quantile

1.1.1 Estimation des quantiles

a) Définitions du quantile

Soit X une variable aléatoire réelle de fonction de répartition F et soit α dans l'intervalle ouvert $]0,1[$. On définit Q_α le quantile d'ordre α de X comme étant un réel (non nécessairement unique) tel que la probabilité qu'une réalisation de X soit inférieure ou égale à Q_α soit α.

Remarquons que les propriétés d'une fonction de répartition $F(\cdot)$, notamment sa continuité à droite et l'existence d'une limite à gauche en tout point (*càdlàg*), implique que le quantile Q_α vérifie

$$Q_\alpha = \inf_{t\in\mathbb{R}} \left\{F(t) \geq \alpha\right\}, \tag{1.1.1}$$

pour tout α dans $]0,1[$, où l'on adopte la convention $\inf\{\emptyset\} = +\infty$. Le quantile Q_α d'ordre α dans $]0,1[$ d'une variable aléatoire X peut être définit de plusieurs façons. Par exemple, les quantiles vérifiant (1.1.1) doivent vérifier

$$F(Q_\alpha -) \leq \alpha \leq F(Q_\alpha), \tag{1.1.2}$$

où $F(x-) = \lim_{\epsilon\to 0^+} F(x-\epsilon)$. Notons que Q_α est définie de façon unique par $Q_\alpha = F^{-1}(\alpha)$ si X a une densité $f(\cdot)$ continue et strictement positive, ce qu'on supposera dans la suite.

Une méthode utilisée pour caractériser les quantiles d'une distribution est liée au domaine plus vaste de la M-estimation. En effet, remarquons que Q_α peut être vu comme le réel minimisant la quantité $\mathcal{L}(q) = \mathbb{E}\left[\ell_\alpha(X-q)\right]$, par rapport à la variable q et où $\ell_\alpha(t) = |t| + (2\alpha-1)t$. Montrons cette propriété : On a

$$\begin{aligned}
\mathcal{L}(q) &= \int \left\{|x-q| + (2\alpha-1)(x-q)\right\} f(x)dx \\
&= 2\int_{-\infty}^{\infty} \left\{\alpha(x-q)\mathbb{I}\left(x>q\right) + (1-\alpha)(q-x)\mathbb{I}\left(x<q\right)\right\} f(x)dx \\
&= 2\alpha\int_q^{\infty} (x-q)f(x)dx + 2(1-\alpha)\int_{-\infty}^{q} (q-x)f(x)dx
\end{aligned}$$

$$= 2\alpha \int_{-\infty}^{\infty} xf(x)dx - 2\alpha q \int_{-\infty}^{\infty} f(x)dx - 2\int_{-\infty}^{q} xf(x)dx + 2q\int_{-\infty}^{q} f(x)dx.$$

La condition du premier ordre de $\mathcal{L}(q)$ par rapport à q peut donc s'écrire :

$$0 = -2\alpha + 2\int_{-\infty}^{q} f(x)dx.$$

Observons que $\mathcal{L}(q)$ est strictement convexe puis que $\mathcal{L}^{(2)}(q) = 2f(q) > 0$. Cela implique donc que le quantile d'ordre α de la distribution X est l'unique minimum de $\mathcal{L}(q)$. Nous noterons alors

$$Q_\alpha = \operatorname*{Arg\,Min}_{q\in\mathbb{R}} \mathcal{L}(q).$$

Une motivation statistique pour l'estimation des quantiles est de construire des intervalles de confiance. On se donne un niveau α dans $]0,1[$ et une variable réelle X. Un intervalle de confiance I_α est un intervalle de $\mathbb{R}$ tel qu'une réalisation de X appartienne à I_α avec une probabilité $1-\alpha$. Une façon naturelle de construire un tel intervalle est de calculer les quantiles d'ordre $1-\alpha/2$ et $\alpha/2$ de la variable X, $Q_{1-\alpha/2}$ et $Q_{\alpha/2}$. On a alors pour $I_\alpha = \left[Q_{\alpha/2}, Q_{1-\alpha/2}\right]$,

$$\mathbb{P}\left(X \in I_\alpha\right) = 1 - \alpha.$$

A partir d'un échantillon, il existe plusieurs façons d'estimer les quantiles $Q_{1-\alpha}$ et Q_α pour α fixé. Ces estimateurs de quantiles permettent alors de construire des intervalles de confiance approximatifs, et ainsi par exemple de déterminer les données pouvant apparaître "hors-norme". Ces données seront considérées comme telles si elles sont en dehors de l'intervalle de confiance estimé pour un niveau α de référence comme 10%, 5% ou 1%.

b) Propriétés asymptotiques

On considère une suite de variables aléatoires réelles $X_1, X_2, \ldots$ indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition commune F. Soit $0 < \alpha < 1$. Sans autres hypothèses sur F, on définit Q_α, le quantile d'ordre α de la fonction de répartition F par la relation (1.1.2). On définit maintenant la fonction de répartition empirique associée à la fonction F. Soit donc

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(X_i \le x),$$

où $\mathbb{I}(\cdot)$ désigne la fonction indicatrice. Soit $\widehat{Q}_\alpha$, le quantile d'ordre α de la fonction de répartition empirique F_n définit par la relation

$$\widehat{Q}_\alpha := \inf_{x\in\mathbb{R}}\left\{F_n(x) \ge \alpha\right\}. \tag{1.1.3}$$

La question est naturellement de savoir comment se comporte asymptotiquement la quantité $\widehat{Q}_\alpha - Q_\alpha$. En fait, $\widehat{Q}_\alpha$ hérite d'un certain nombre de propriétés de la fonction de répartition empirique F_n. Les propriétés générales des fonctions de répartition, notamment le fait que la loi de la variable aléatoire $nF_n(x)$ soit une binomiale $\mathcal{B}(n, F(x))$ permettent alors d'utiliser des outils tels que les inégalités de type Hoeffding et ainsi d'avoir par exemple le résultat suivant,

Théorème 1 *Soit $0 < \alpha < 1$. Si F est continue dans un voisinage de Q_α, alors pour tout $\epsilon > 0$*

$$\mathbb{P}\left(\left|\widehat{Q}_\alpha - Q_\alpha\right| > \epsilon\right) \le 2\exp\left(-2nt_\epsilon\right),$$

où $t_\epsilon = \min(F(Q_\alpha+\epsilon)-\alpha, \alpha - F(Q_\alpha-\epsilon))$.

Ce résultat disponible dans le livre de Serfling (1980) donne la convergence presque complète de $\widehat{Q}_\alpha$ vers Q_α et donc la convergence presque sûre. Une façon de prouver ce théorème est de composer par F_n^{-1} dans la probabilité, ayant au préalable décomposé la valeur absolue. Autrement dit, il s'agit de remarquer que

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{Q}_\alpha - Q_\alpha\right| > \epsilon\right) &\le \mathbb{P}\left(\widehat{Q}_\alpha > Q_\alpha + \epsilon\right) + \mathbb{P}\left(\widehat{Q}_\alpha < Q_\alpha - \epsilon\right) \\ &= \mathbb{P}\left(F_n(Q_\alpha - \epsilon) \ge \alpha\right) + \mathbb{P}\left(F_n(Q_\alpha+\epsilon) < \alpha\right),\end{aligned}$$

grâce aux propriétés de F_n, puis ayant correctement re-normalisé cette expression, on applique l'inégalité de Hoeffding aux variables binomiales $nF_n(Q_\alpha+\epsilon)$ et $nF_n(Q_\alpha-\epsilon)$.

Nous donnons aussi un résultat sur la densité de $\widehat{Q}_\alpha$. En effet, si X admet une densité f, alors il en est de même pour $\widehat{Q}_\alpha$ et cette densité est calculable. Elle peut se déduire de la fonction de répartition de $\widehat{Q}_\alpha$ que nous notons G_n. On a la relation

$$\begin{aligned} G_n(t) &= \mathbb{P}\left(\widehat{Q}_\alpha \leq t\right) = \mathbb{P}\left(nF_n(t) \geq n\alpha\right) = \mathbb{P}\left(\mathcal{B}(n, F(t)) \geq n\alpha\right) \\ &= \sum_{i=m}^{n} \begin{pmatrix} n \\ i \end{pmatrix} (F(t))^i \, (1 - F(t))^{n-i}, \end{aligned}$$

où $\mathcal{B}(n, F(t))$ désigne la loi binomiale de paramètres n et $F(t)$ et où

$$m = \begin{cases} n\alpha & \text{si } n\alpha \text{ est un entier} \\ \lfloor n\alpha \rfloor + 1 & \text{si } n\alpha \text{ n'est pas un entier.} \end{cases}$$

On en déduit par simple dérivation que la densité g_n de $\widehat{Q}_\alpha$ est

$$g_n(t) = n \begin{pmatrix} n-1 \\ m-1 \end{pmatrix} (F(t))^{m-1} \, (1 - F(t))^{n-m} \, f(t),$$

où $m = n\alpha$ si $n\alpha$ est un entier et $m = \lfloor n\alpha \rfloor + 1$ si $n\alpha$ n'est pas un entier.

Pour mieux comprendre l'interdépendance entre le quantile empirique et la fonction de répartition empirique, Bahadur (1966) a développé un outil permettant d'écrire la différence du quantile empirique avec le quantile comme une somme de variables aléatoires (précisément une fonction de la distribution empirique) moyennant une faible erreur.

La représentation de Bahadur a été un sujet de recherche important en statistique depuis son introduction dans l'article de Bahadur (1966). Initialement, cette représentation à été construite pour différents types d'estimateurs de quantiles de distributions par Bahadur (1966). Elle a été ensuite affinée et adaptée dans plusieurs contextes plus généraux que celui du quantile de distribution comme par exemple les modèles des M-estimateurs.

Avant de présenter le théorème principal de Bahadur (1966) nous donnons la définition d'une suite de variables aléatoires bornée presque sûrement.

Définition 1 *Soient* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires et* $(a_n)_{n\geq 1}$ *une suite de réels strictement positifs. On a* $X_n = O_{p.s.}(a_n)$ *lorsque* $n \to \infty$ *si et seulement si il existe un ensemble* Ω *tel que* $\mathbb{P}(\Omega) = 1$ *et tel que pour tout* $\omega \in \Omega$*, il existe une constante* $B(\omega)$ *telle que*

$$|X_n(\omega)| \leq B(\omega) a_n,$$

pour n *assez grand.*

Nous présentons maintenant le théorème principal de Bahadur (1966).

Théorème 2 (Bahadur (1966)) *Supposons que* F *est deux fois dérivable, de dérivée seconde continue et bornée dans un voisinage de* Q_α*, et telle que* $f(Q_\alpha) := F'(Q_\alpha) > 0$*. Alors si on définit*

$$R_n(\alpha) := \widehat{Q}_\alpha - Q_\alpha + \frac{F_n(Q_\alpha) - \alpha}{f(Q_\alpha)}, \tag{1.1.4}$$

on a

$$R_n(\alpha) = O_{p.s.}\left(\frac{\log^{3/4}(n)}{n^{3/4}}\right),$$

lorsque $n \to \infty$*.*

L'expression de $R_n(\alpha)$ dans (1.1.4) montre l'importance de l'hypothèse $f(Q_\alpha) > 0$. Notons en outre que ce théorème, et la représentation de Bahadur en général, permet d'avoir la normalité asymptotique de $\sqrt{n}(\widehat{Q}_\alpha - Q_\alpha)$ par une simple étude du processus empirique $F_n(Q_\alpha)$. En fait, de façon plus générale, la représentation de Bahadur pour les quantiles empiriques permet de ramener leur étude à celle des propriétés de la fonction de répartition empirique. Il s'ensuit immédiatement le corollaire suivant :

Corollaire 1 *Sous les hypothèses du théorème 2, nous avons*

$$\sqrt{n}\left(\widehat{Q}_\alpha - Q_\alpha\right) \Rightarrow N\left(0, \frac{\alpha(1-\alpha)}{f^2(Q_\alpha)}\right).$$

Le corollaire 1 se prouve par une directe application du théorème de la limite centrale en constatant au préalable que

$$\sqrt{n}R_n(\alpha) = o_{p.s.}(1),$$

grâce au théorème 2.

Nous présentons maintenant quelques résultats concernant l'étude précise du comportement de la variable aléatoire $R_n(\alpha)$ définie en (1.1.4). Le premier à avoir améliorer le résultat du théorème 2 est Kiefer (1967) en donnant l'ordre exact optimal du reste $R_n(\alpha)$.

Théorème 3 (Kiefer (1967)) *Considérons $0 < \alpha < 1$ et $R_n(\alpha)$ définit en (1.1.4). Supposons f bornée dans un voisinage de Q_α et $f(Q_\alpha) > 0$. Alors*

$$\limsup_{n\to\infty} \pm \frac{f(Q_\alpha)R_n(\alpha)}{n^{-3/4}(\log\log(n))^{3/4}} = \frac{2^{5/4}\left[\alpha(1-\alpha)\right]^{1/4}}{3^{3/4}}.$$

Les apports de Kiefer (1967) sont multiples. En premier lieu, il donne le comportement asymptotique exact de $R_n(\alpha)$. En second lieu, il diminue les hypothèses sur F qui est simplement supposée dérivable, de dérivée bornée dans un voisinage de Q_α au lieu de deux fois continûment dérivable. Dans le même article, on trouve la distribution limite de $R_n(\alpha)$. Soient Φ et ϕ, respectivement la fonction de distribution et la densité d'une variable aléatoire ayant pour loi une $N(0,1)$.

Théorème 4 (Kiefer (1967)) *Sous les mêmes hypothèses que celles du théorème 3,*

$$\lim_{n\to\infty} \mathbb{P}\left(n^{3/4}f(Q_\alpha)R_n(\alpha) \leq t\right) = \frac{2}{\left[\alpha(1-\alpha)\right]^{1/2}} \int_0^\infty \Phi\left(\frac{t}{u^{1/2}}\right)\phi\left(\frac{u}{\left[\alpha(1-\alpha)\right]^{1/2}}\right)du.$$

De nombreux autres travaux ont affiné ces résultats. En particulier, Kiefer (1970) a étudié le processus

$$R_n^* := \sup_{0<\alpha<1} f(Q_\alpha)\left|R_n(\alpha)\right|, \tag{1.1.5}$$

et a prouvé

Théorème 5 *Sous les mêmes hypothèses que celles du théorème 3,*

$$\limsup_{n\to\infty} \frac{n^{3/4} R_n^*}{(\log n)^{1/2} (\log\log n)^{1/4}} = 2^{-1/4}.$$

Ce dernier théorème est particulièrement utile car il donne une idée générale du comportement asymptotique de $R_n(\alpha)$ pour les pires valeurs de α.

1.1.2 Régression quantile

Un vaste domaine des statistiques est celui de la régression. Ce domaine étudie l'effet d'une ou plusieurs variables aléatoires dites variables explicatives sur une autre variable aléatoire dite variable réponse via une fonction de régression inconnue. Le modèle de régression quantile a été introduit par Koenker et Bassett (1978) comme une extension de l'estimation de quantile. Il y est exposé que l'un des interêts de cette méthode d'estimation est notamment la robustesse de l'estimateur du quantile conditionnel par rapport à l'estimateur de la régression. Par rapport a la régression linéaire classique, cette méthode permet aussi de mieux analyser la dispersion de la variable dépendante autour d'une tendance centrale, comme par exemple la médiane, en fonction des variables explicatives retenues.

La régression quantile a de nombreuses applications pratiques. Par exemple l'élaboration de tables de références en médecine plus précises que celles obtenue par la régression classique, voir Cole et Green (1992) et Royston et Altman (1994). La régression quantile est aussi utilisée dans le cadre de l'analyse de survie voir Koenker et Geiling (2001). Il existe aussi des applications en finance comme par exemples celles découlant de la notion de "valeur sous risque" (value at risk), voir Basset et Chen (2001). La régression quantile peut servir d'outil pour la détection d'hétéroscédasticité, voir Koenker et Basset (1982) et Wilcox et Keselman (2006). Powell (1986) a montré comment, dans un modèle à données censurées, on peut utiliser les quantiles conditionnels pour aussi tester l'hétéroscedasticité et la symétrie des distributions des erreurs. Härdle et Gasser (1984), Huber (1981) donnent des exemples montrant la robustesse d'estimateurs des quantiles conditionnels par rapport à d'autres estimateurs, notamment ceux de la régression. Honda (2000) et Shi (1995) étendent, parmi d'autres, cette approche aux cas de données dépendantes.

Un des avantages de la régression quantile par rapport à la régression classique consiste dans le fait qu'estimer les quantiles fournit naturellement des bandes de confiance au lieu d'une droite de régression. Le modèle de Konker et Basset (1978) suppose que la fonction de quantile $Q(\alpha|x)$ de Y sachant $X = x$ vérifie

$$Q(\alpha|x) = x^T \beta(\alpha),$$

où ici, x et $\beta(\alpha)$ sont dans $\mathbb{R}^d$ et $\beta(\alpha)$ est inconnu.

Disposant d'un n-échantillon (Y_1, X_1), $\cdots$, (Y_n, X_n) de même loi que (Y, X) dans $\mathbb{R} \times \mathbb{R}^d$, Koenker et Basset (1978) propose d'estimer $\beta(\alpha)$ par

$$\widehat{\beta}(\alpha) = \operatorname*{Arg\,Min}_{b \in \mathbb{R}^d} \sum_{i=1}^{n} \ell_\alpha(Y_i - X_i^T b) = \operatorname*{Arg\,Min}_{b \in \mathbb{R}^d} \mathcal{L}_n(b; \alpha), \tag{1.1.6}$$

où $\ell_\alpha(t) = |t| + (2\alpha - 1)t$, t dans $\mathbb{R}$.

Les avantages tirés de l'utilisation du quantile conditionnel plutôt que l'espérance conditionnelle pour expliciter le comportement d'une variable aléatoire Y sachant une variable explicative X est donné un grand nombre d'articles.

Exposons dans un premier temps certains des avantages de la régression quantile par rapport à d'autres méthodes d'estimation.

a) Deux propriétés importantes de l'approche quantile

Robustesse Les statistiques robustes ont connu un intérêt grandissant depuis le papier fondateurs de Huber (1967). Les problématiques de la robustesse proviennent d'une observation de l'estimateur de la moyenne. En effet la moyenne empirique est grandement affectée par la présence de données aberrantes (outliers). L'estimation du quantile ou de la régression quantile offre dans cette optique une alternative robuste d'estimation. Plus précisément, la moyenne empirique peut être rendue aussi grande que l'on veut en augmentant la valeur d'une seule variable tandis que la valeur de l'estimateur quantile changera peu, comme la fonction de répartition empirique a partir de laquelle il est définit.

Plus formellement, on mesure communément la robustesse d'une procédure statistique via la fonction d'influence, voir Hampel (1974) pour les premiers résultats et Huber (1981) pour une revue de la littérature. Soit θ un paramètre lié a une distribution F d'une v.a. X. En fait, on peut voir θ comme une fonction de la distribution F, soit $\theta = \theta(F)$. Par exemple le paramètre θ peut être l'espérance de X soit $\theta = \int x dF(x)$. On considère une fonction contaminée F_ε par le remplacement d'une petite quantité de masse ε de F par une masse équivalente concentrée en y soit

$$F_\varepsilon = \varepsilon \delta_y + (1 - \varepsilon) F,$$

ou δ_y désigne la masse de Dirac en y. On exprime la fonction d'influence de θ en F comme étant

$$IF_\theta(y, F) = \lim_{\varepsilon \to 0} \frac{\theta(F_\varepsilon) - \theta(F)}{\varepsilon}.$$

Par exemple pour l'espérance,

$$\theta(F_\varepsilon) = \int y dF_\varepsilon = \varepsilon y + (1 - \varepsilon)\theta(F),$$

ce qui donne pour la fonction d'influence

$$IF_\theta(y, F) = y - \theta(F).$$

Prenons le cas de la médiane : $\theta(F) = F^{-1}(1/2)$. Alors nous avons $\theta(F_\varepsilon) = F_\varepsilon^{-1}(1/2)$ et donc

$$IF_\theta(y, F) = \frac{\text{sgn}\left(y - F_\varepsilon^{-1}(1/2)\right)}{f(F^{-1}(1/2))},$$

ou f est la dérivée de F et sgn la fonction signe.

Comparons maintenant ces deux fonctions d'influence. En premier lieu, constatons que la fonction d'influence de l'espérance est proportionnelle à y. Donc même une petite contamination de F en un point y assez loin de $\theta(F)$ crée une grande différence avec la vraie valeur en F. En contrepartie, la fonction d'influence de la médiane d'une contamination en y est bornée par la constante $f(F^{-1}(1/2))^{-1}$ appelée fonction de sparsité à la médiane. L'estimation de la médiane est clairement une procédure plus robuste. En effet tant que le signe de la quantité $y - F_\varepsilon^{-1}(1/2)$ reste inchangé, la valeur de y ne change pas la fonction d'influence. En bref, la fonction d'influence de l'espérance est grandement affectée par la valeur de y tandis que celle de la médiane ne l'est pas, à condition que $y - F_\varepsilon^{-1}(1/2)$ ne change pas de signe. Ces résultats s'étendent aussi au cas de la régression quantile, voir Koenker (2005, Théorème 2.4).

Invariance La régression quantile bénéficie d'une importante propriété qui la différencie de la régression linéaire classique. Cette propriété est l'invariance par transformation croissante, comme rappelée par Koenker et Geiling (2001).

Soit Y dans $\mathbb{R}$ et X dans $\mathbb{R}^d$ deux v.a. telles que la quantile d'ordre α dans $[0,1]$ de Y sachant X, noté $Q_Y(\alpha|X)$ vérifie

$$Q_Y(\alpha|X) = X^\top \beta(\alpha), \text{ avec } \beta(\alpha) = \underset{\beta\in\mathbb{R}^d}{\operatorname{Arg\,Min}}\, \mathbb{E}\left[\ell_\alpha\left(Y - X^\top\beta\right)\right] = \underset{\beta\in\mathbb{R}^d}{\operatorname{Arg\,Min}}\, \mathcal{L}(b;\alpha).$$

Considérons maintenant la variable $Y^* = cY$ pour un $c > 0$. Grâce aux propriétés de $\mathcal{L}(b;\alpha)$ nous avons

$$\beta^*(\alpha) = \underset{\beta\in\mathbb{R}^d}{\operatorname{Arg\,Min}}\, \mathbb{E}\left[\ell_\alpha\left(Y^* - X^\top\beta\right)\right],$$

vérifie $\beta^*(\alpha) = c\beta(\alpha)$. Lorsque $c < 0$, nous avons $\beta^*(\alpha) = c\beta(1-\alpha)$. On en déduit que pour $\alpha = 1/2$, c'est-à-dire dans le cadre de la médiane conditionnelle, le paramètre $\beta(1/2)$ est invariant par changement d'échelle dans le sens ou $\beta^*(1/2) = c\beta(1/2)$ pour tout c dans $\mathbb{R}$.

Maintenant si la variable Y subit une translation dans le sens ou il existe un vecteur γ tel que

$$Y^* = Y + X^\top\gamma,$$

on a alors

$$\beta^*(\alpha) = \beta(\alpha) + \gamma.$$

Le coefficient de la régression linéaire vérifie les mêmes propriétés d'invariance, mais l'approche quantile permet également de considérer des transformations non linéaires monotones.

Plaçons nous dans un cadre non plus seulement affine. Soit une fonction $h(\cdot)$ croissante. Le quantile de $h(Y)$ se calcule facilement grâce à la transformation $h(\cdot)$, puisque

$$Q_{h(Y)}(\alpha|x) = h(Q_Y(\alpha|x)).$$

Cette propriété se démontre par la croissance de $h(\cdot)$ et les propriétés de la fonction de répartition conditionnelle. L'espérance conditionnelle ne vérifie pas cette propriété. En effet, on a en général

$$\mathbb{E}\left[h(Y)|X\right] \neq h(\mathbb{E}\left[Y|X\right]).$$

L'invariance peut être utile pour changer les échelles (par exemple de température) dans les données de la variable réponse. La propriété d'invariance est très utile dans les cadres qui traitent de données censurées, voir Powell (1986) dans le cas d'une censure déterministe. Le modèle est aussi étudié dans le cadre de censure aléatoire, voir Honore, Powell et Khan (2000).

Donnons un exemple simple pour illustrer cette propriété. Considérons le modèle de régression linéaire logarithmique issu d'un couple de v.a. (X, Y)

$$\log(Y) = X^\top \beta + \varepsilon,$$

avec ε et X indépendantes et $\mathbb{E}\left[\varepsilon\right] = 0$. On a alors

$$X^\top \beta = \mathbb{E}\left[\log Y|X\right],$$

mais en général $\exp\left(X^\top \beta\right) \neq \mathbb{E}\left[Y|X\right]$. Dans le cadre des quantiles de régression, nous avons

$$X^\top \beta = Q_{\log Y}(\alpha|X) = \log\left(Q_Y(\alpha|X)\right),$$

de telle façon que $\exp\left(X^\top \beta\right)$ est la quantile conditionnel de Y sachant X.

b) Consistance

Dans cette partie, nous présentons quelques résultats de convergence de l'estimateur du quantile de régression paramétrique. Les premiers résultats ainsi que leur introduction sont dus à Koenker et Basset (1978).

Pour plus de clarté, rappelons les hypothèses du modèle de régression quantile linéaire. On suppose que le quantile d'ordre α dans $(0,1)$ de Y sachant $X = x$, noté $Q(\alpha|x)$, est donné par la relation

$$Q(\alpha|x) = x^T \beta(\alpha).$$

On considère un estimateur de $\beta(\alpha)$ définit par la relation (1.1.6), que l'on note $\widehat{\beta}_n(\alpha)$. El Bantli et Hallin (1999) font les hypothèses suivantes :

i. Pour tout $\varepsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} n^{1/2} \left(n^{-1} \sum_{i=1}^{n} F_{Y|X}\left(X_i^T \beta(\alpha) - \varepsilon\right) - \alpha \right) &= \infty, \\ \lim_{n\to\infty} n^{1/2} \left(\alpha - n^{-1} \sum_{i=1}^{n} F_{Y|X}\left(X_i^T \beta(\alpha) + \varepsilon\right) \right) &= \infty. \end{aligned}$$

ii. Il existe $d > 0$ tel que

$$\liminf_{n\to\infty} \inf_{\|u\|=1} n^{-1} \sum_{i=1}^{n} \mathbb{I}\left(\left|X_i^T u\right| < d\right) = 0, \quad p.s.$$

iii. Il existe $D > 0$ tel que

$$\limsup_{n\to\infty} \sup_{\|u\|=1} n^{-1} \sum_{i=1}^{n} \left(X_i^T u\right)^2 \leq D, \quad p.s.$$

On a alors le résultat de consistance suivant :

Théorème 6 *Les hypothèses i, ii et iii ci-dessus sont nécessaires et suffisantes pour avoir*

$$\lim_{n\to\infty} \widehat{\beta}_n(\alpha) = \beta(\alpha) \quad p.s.$$

La première hypothèse suppose intuitivement que la fonction de répartition conditionnelle n'est pas constante dans un voisinage du quantile. La deuxième hypothèse assure la bonne répartition des observations issues de la v.a. X. Notons enfin que la convergence de la matrice $n^{-1} \sum_{i=1}^{n} X_i^T X_i$ vers une matrice définie positive implique la troisième hypothèse.

c) Normalité Asymptotique

Présentons un résultat de normalité asymptotique. Les conditions de normalité asymptotique sont plus fortes que celles faites pour la consistance. On suppose :

i. La distribution conditionnelle $F_{Y|X}(\cdot|\cdot)$ de Y sachant X est absolument continue et possède une densité $f(\cdot|\cdot)$ strictement positive et finie dans un voisinage de $Q(\alpha|x)$.

ii. Il existe une matrice définie positive D telle que $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n X_i^T X_i = D$ presque sûrement.

iii. $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n f(Q(\alpha|X_i)|X_i)X_i^T X_i = D(\alpha)$ presque sûrement avec $D(\alpha)$ matrice définie positive.

iv. $\max_{i=1}^n \|X_i\|/n^{1/2} \to_{n\to\infty} 0$ presque sûrement.

On a alors ce résultat :

Théorème 7 *Sous les hypothèses ci-dessus,*

$$n^{1/2}\left(\widehat{\beta}_n - \beta(\alpha)\right) \Rightarrow N\left(0, \alpha(1-\alpha)D(\alpha)^{-1}DD(\alpha)^{-1}\right).$$

Nous donnons une preuve de ce résultat due à Knight (1998). L'un des interêts de cette preuve est ici d'introduire certain raisonnements utiles pour l'élaboration de la représentation de Bahadur.

Preuve du théorème 7. On rappelle que $\ell_\alpha(t) = |t| + (2\alpha-1)t$. On étudie le comportement de $n^{1/2}(\widehat{\beta}_n - \beta(\alpha))$ à partir de la fonction

$$\Delta\mathcal{L}_n(b;\alpha) = \sum_{i=1}^n \left\{\ell_\alpha\left(Y_i - X_i^T\beta(\alpha) - n^{-1/2}X_i^T b\right) - \ell_\alpha\left(Y_i - X_i^T\beta(\alpha)\right)\right\}.$$

Remarquons que $b \mapsto \Delta\mathcal{L}_n(b;\alpha)$ est convexe et atteint son minimum en $n^{1/2}(\widehat{\beta}_n - \beta(\alpha))$. La fonction $\ell_\alpha(t)$ étant absolument continue, nous avons la décomposition suivante de $\Delta\mathcal{L}_n(b;\alpha)$:

$$\Delta\mathcal{L}_n(b;\alpha) = -\frac{1}{n^{1/2}}\sum_{i=1}^n \left(\alpha - \mathbb{I}\left(Y_i - X_i^T\beta(\alpha)\right)\right) X_i^T b$$

$$
\begin{aligned}
& + \sum_{i=1}^{n} \int_0^{n^{-1/2} X_i^T b} \left[\mathbb{I}\left(Y_i \leq X_i^T \beta(\alpha) + u\right) - \mathbb{I}\left(Y_i - X_i^T \beta(\alpha) \leq 0\right) \right] du \\
= \quad & \Delta \mathcal{L}_{n,1}(b;\alpha) + \Delta \mathcal{L}_{n,2}(b;\alpha).
\end{aligned}
$$

D'après le théorème de la limite centrale multivarié et l'hypothèse ii, on a la convergence en loi suivante : $\Delta \mathcal{L}_{n,1}(b;\alpha) \Rightarrow -b^T V$ ou V a pour loi une $N(0, \alpha(1-\alpha)D)$. Le deuxième terme $\Delta \mathcal{L}_{n,2}(b;\alpha)$ peut s'écrire

$$
\Delta \mathcal{L}_{n,2}(b;\alpha) = \mathbb{E}\left[\Delta \mathcal{L}_{n,2}(b;\alpha)|X\right] + \left(\Delta \mathcal{L}_{n,2}(b;\alpha) - \mathbb{E}\left[\Delta \mathcal{L}_{n,2}(b;\alpha)|X\right]\right).
$$

Un développement de Taylor de $F_{Y|X}(\cdot)$ autour du quantile d'ordre α nous assure que

$$
\mathbb{E}\left[\Delta \mathcal{L}_{n,2}(b;\alpha)|X\right] = n^{-1} \sum_{i=1}^{n} \int_0^{X_i^T b} f(Q(\alpha|X_i)|X_i) u du + o(1).
$$

Donc d'après iii, on a $\mathbb{E}\left[\Delta \mathcal{L}_{n,2}(b;\alpha)|X\right] \to b^T D(\alpha) b/2$. De plus nous avons $\mathrm{Var}(\Delta \mathcal{L}_{n,2}(b;\alpha)) \leq C \max_{i=1}^{n} (X_i^T b)/n^{1/2}$ donc d'après iv, on a $\Delta \mathcal{L}_n(b;\alpha) \Rightarrow -b^T V + b^T D(\alpha) b/2$.

D'après Hjørt et Pollard (1993), la convexité de la fonction limite $b \mapsto -b^T V + b^T D(\alpha) b/2$ assure l'unicité du minimum et donc

$$
n^{1/2}\left(\widehat{\beta}_n - \beta(\alpha)\right) = \operatorname{Arg}\underset{b}{\operatorname{Min}}\, G_\alpha(b) \Rightarrow D(\alpha)^{-1} V = \operatorname{Arg}\underset{b}{\operatorname{Min}}(-b^T V + b^T D(\alpha) b/2). \square
$$

1.1.3 Estimation non-paramétrique de quantiles conditionnels

Comme extension naturelle aux problématiques initiées par Koenker et Bassett (1978), nous présentons maintenant quelques résultats concernant l'estimation non-paramétrique de quantiles conditionnels. De fait, on suppose désormais que les fonctionnelles à estimer appartiennent à des espaces de dimension infinie. La méthode des polynômes locaux est introduite dans Stone (1977, 1982) ou encore Cleveland (1979). Leur résultats sont le plus souvent appliqués à l'estimation de la régression et de ses dérivées. Parmi les articles traitant de la méthode polynômes locaux, on trouve ceux de Tsybakov (1986), Hastie et Loader (1993), Ruppert et Wand (1994) parmi d'autres. Parmi ceux-ci, Hastie et Loader (1993) donne une bonne revue de la littérature existante et les avantages de la méthode des polynômes locaux. Fan et Gijbels (1996) donne aussi beaucoup de résultats sur la méthode des polynômes locaux.

He et Shi (1994) estiment des quantiles conditionnels par B-splines. Yu et Jones (1998) ont étudié dans leur papier une régression quantile avec la méthode de la constante locale en utilisant deux méthodes de caractérisation des quantiles conditionnels. La première méthode est celle exposée dans Koenker et Basset (1978) et l'autre méthode consiste à définir comme estimateur du quantile conditionnel l'inverse de la fonction de répartition conditionnelle. Ils comparent dans ce papier les deux méthodes, et montrent leur équivalence asymptotique. Chaudhuri (1991) a établi une représentation de Bahadur non-paramétrique du quantile conditionnel.

Avant de présenter le résultat de Chaudhui (1991), nous définissons la notion de régularité en introduisant les classes de Hölder. Dans la suite, $\|\cdot\|$ désigne la norme euclidienne usuelle. Si $u = (u_1, \cdots, u_d)^T$ est un vecteur tel que $u_i \in \mathbb{N}$ pour tout $1 \leq i \leq d$, on définit l'opérateur différentiel D^u par

$$D^u = \frac{\partial^{[u]}}{\partial x_1^{u_1} \cdots \partial x_d^{u_d}},$$

où $[u] = u_1 + \cdots + u_d$.

Définition 2 *Soient deux réels $s, L > 0$. On définit la classe de Hölder $C(L, s)$ sur un sous-ensemble A de $\mathbb{R}^d$ comme étant l'ensemble des fonctions $f : A \to \mathbb{R}$ qui soient dérivables $\lfloor s \rfloor$ fois et telles que*

$$|D^u f(z) - D^u f(z')| \leq L \|z - z'\|^{s - \lfloor s \rfloor},$$

pour tout (z, z') dans A^2 et tout $[u] = \lfloor s \rfloor$.

L'article de Chaudhuri (1991) développe une représentation de Bahadur avec la méthode polynômes locaux utilisant des noyau uniformes. Soit $0 < \alpha < 1$ fixé. On dispose d'observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ i.i.d. où les Y_i sont des variables réelles et les X_i appartiennent à $\mathbb{R}^d$. Chaudhuri (1991) considère le modèle de régression quantile non-paramétrique

$$Y_i = Q(\alpha | X_i) + \epsilon_i, \quad 0 \leq i \leq n, \tag{1.1.7}$$

les ϵ_i étant indépendantes des X_i pour tout i et telles que $F_\epsilon(0) = \alpha$ où F_ϵ désigne la fonction de répartition commune des ϵ_i. Cette hypothèse sur les résidus $\epsilon_i = \epsilon_i(\alpha)$ traduit que le quantile d'ordre α de la variable Y sachant $X = x$ est égale à $Q(\alpha|x)$.

Les hypothèses et notations faites par Chaudhuri sont les suivantes. On considère le problème de l'estimation de la fonction de quantile au point 0 de $\mathbb{R}^d$. Pour cela, on définit un voisinage V de 0 dans $\mathbb{R}^d$. On suppose que $Q(\alpha|x)$ appartient à la classe de Hölder $C(L, p+\gamma)$ sur V avec $p \in \mathbb{N}$, γ dans $]0,1]$ et $s = p + \gamma$.

On considère ensuite une suite $(h_n)_{n\geq 1}$ de réels strictement positifs tels que $h_n = an^{-1/(2s+d)}$, $a > 0$. Soit C_n le cube $[-ah_n, ah_n]^d$. Il est clair qu'il existe n_0 tel que pour tout $n \geq n_0$, $C_n \subset V$. Dans la suite nous supposerons que n est suffisamment grand pour que C_n est inclus dans V. Soit S_n l'ensemble des indices des X_i tels que $X_i \in C_n$

$$S_n = \{i \ ; \ \ X_i \in C_n, 1 \leq i \leq n\}$$

et soit $N_n = \mathrm{Card}(S_n)$. Soit A l'ensemble des vecteurs u de dimension d à entrées entières tels que $[u] \leq p$ et $s(A)$ le cardinal de A. Soit $\beta_n = (\beta_{n,u})_{u\in A}$ le vecteur de dimension $s(A)$ tel que

$$\beta_{n,u} = D^u Q(\alpha|0) h_n^{[u]} [u]^{-1}. \tag{1.1.8}$$

Pour x dans $\mathbb{R}^d$, soit

$$Q^*(\alpha|x) = \sum_{u\in A} \beta_{n,u} h_n^{[u]} x^u, \tag{1.1.9}$$

où par convention, $x^u = \Pi_{i=1}^d x_i^{u_i}$ si $u = (u_1, \cdots, u_d) \in A$. Il est clair que d'après l'équation (1.1.8), le polynôme donné en (1.1.9) est le polynôme de Taylor du quantile conditionnel à l'ordre p.

Nous sommes en mesure de définir l'estimateur proposé par Chaudhuri (1991). Soit donc $\widehat{\beta}_n$ une des solutions du problème de minimisation

$$\operatorname*{Arg\,Min}_{\beta\in\mathbb{R}^{s(A)}} \sum_{i=1}^n \ell_\alpha \left(Y_i - \sum_{u\in A} \beta_u h_n^{-[u]} X_i^u \right) \mathbb{I}\left(X_i \in C_n\right),$$

où la fonction $\ell_\alpha(\cdot)$ est définie sur $\mathbb{R}$ et telle que $\ell_\alpha(t) = |t| + (2\alpha - 1)t$. Définissons ensuite le réel $X_i(h_n, u) = h_n^{-[u]} X_i^u$ et le vecteur $X_i(h_n, A) = (X_i(h_n, u))_{u \in A}$. Soit w la densité des erreurs dans le modèle (1.1.7). Soit

$$w_{h_n}(x) = w(h_n x) \left(\int_{[-1,1]^d} w(h_n t) dt \right)^{-1},$$

et $\mathbf{Q}_n$ la matrice carrée symétrique d'ordre $s(A)$: $\mathbf{Q}_n = (\int_{[-1,1]^d} x^u x^v w_{h_n}(x) dx)_{1 \leq [u],[v] \leq s}$. Soit $\beta_n = (\beta_{n,u})_{u \in A}$ avec $\beta_{n,u} = D^u Q(\alpha|0) h_n^{[u]} (u!)^{-1}$ et $u! = \Pi_{i=1}^d (u_i!)$, u dans A. Nous pouvons maintenant présenter la représentation de Bahadur de Chaudhuri (1991).

Théorème 8 Chaudhuri (1991). *Supposons que la densité w de X est continue en 0 et strictement positive sur V, et que les erreurs ε_i possèdent une densité Hölderienne continue de paramètre η. Alors*

$$\widehat{\beta} - \beta_n = (N_n)^{-1} f(0)^{-1} (\mathbf{Q}_n)^{-1} \sum_{i \in S_n} X_i(h_n, A) \left(\alpha - \mathbb{I}(Y_i \leq Q^*(\alpha|X_i))\right) + R_n,$$

avec

$$\begin{aligned} R_n &= O_{p.s.} \left(\frac{\log^{1/2}(n)}{n^{s/(2s+d)}} \right)^{1+\eta} \quad si \;\; \eta \in (0, 1/2), \\ R_n &= O_{p.s.} \left(\frac{\log^{1/2}(n)}{n^{s/(2s+d)}} \right)^{3/2}, \quad si \;\; \eta \in [1/2, 1], \end{aligned}$$

lorsque n tend vers l'infini.

1.2 M-estimateurs

Les M-estimateurs ont été introduits par Huber (1964). Dans cet article, Huber propose de généraliser l'estimateur du maximum de vraisemblance en le considérant comme un problème de minimisation d'un certaine fonction.

On considère un échantillon i.i.d. de n variable aléatoire réelles $X_1, \ldots, X_n$ issu d'une variable X. Huber se place dans un cadre général d'estimateurs $\widehat{\theta}$ définis par

$$\widehat{\theta} = \operatorname*{Arg\,Min}_{\theta \in \mathbb{R}} \sum_{i=1}^{n} \rho\left(X_i, \theta\right), \tag{1.2.10}$$

où la fonction $\rho : \mathbb{R}^2 \to \mathbb{R}$ est supposée mesurable. Différents choix de la fonction ρ conduisent à différents estimateurs de fonctionnelles de la loi de X.

Intuitivement, on peut voir les M-estimateurs comme une généralisation d'une des définitions de l'espérance. En effet, l'espérance $\mathbb{E}\left[X\right]$ d'une variable aléatoire X peut-être définie comme la solution du problème de minimisation suivant,

$$\mathbb{E}\left[X\right] = \operatorname*{Arg\,Min}_{t \in \mathbb{R}} \mathbb{E}\left[(X-t)^2\right].$$

La loi des grands nombre nous assure que $\sum_{i=1}^{n}(X_i - t)^2/n$ est un estimateur de $\mathbb{E}\left[X-t\right]^2$. Il est donc naturel de proposer d'estimer la moyenne par

$$\widehat{\mathbb{E}}X = \operatorname*{Arg\,Min}_{t \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} (X_i - t)^2.$$

Nous pouvons étendre cette procédure à d'autres fonctions ρ dans (1.2.10). Deux exemples classiques sont ceux de la médiane et de la fonction de vraisemblance. Pour la médiane d'une variable aléatoire X, notée $Q_{1/2}$, nous avons la propriété suivante :

$$Q_{1/2} = \operatorname*{Arg\,Min}_{t \in \mathbb{R}} \mathbb{E}\left[|X-t|\right],$$

cf (1.2.10). Alors, l'estimateur de la médiane de X obtenu en remplaçant ρ par la fonction valeur absolue dans l'équation (1.2.10) nous donne le M-estimateur du quantile.

Dans son papier de 1964, Huber propose de séparer l'étude des M-estimateurs selon que la fonction ρ est convexe ou non. Nous présentons différents résultats selon ces cas.

Dans les deux cas, convexe et non convexe, la fonction ρ est supposée vérifier pour tout y et t dans $\mathbb{R}$, $\rho(y;t) = \rho(y-t)$.

1.2.1 M-estimateurs paramétriques

Un M-estimateur est donc une statistique définie comme un argument du minimum d'une certaine fonction ρ. Si la fonction ρ vérifie un certain nombre de conditions de régularité (continuité, dérivabilité, etc.), le problème de minimisation peut devenir la vérification d'une condition du premier ordre, autrement dit l'annulation d'une dérivée.

a) Consistance et normalité asymptotique

Fonction de perte convexe ρ est supposée convexe non constante, continue sur $\mathbb{R}$ et telle que $\lim_{|t|\to\infty} \rho(t) = \infty$. On considère $X_1, \cdots, X_n$, n variables i.i.d. ayant pour distribution F. On note $\mathbb{X} = (X_1, \cdots, X_n)$. Soit $\mathcal{Q}_n(t) = \sum_{i=1}^n \rho(X_i - t)$ et

$$\{T_n(\mathbb{X})\} = \left\{ t \in \mathbb{R} \ / \ \mathcal{Q}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \rho(X_i - t) = \min_{x\in\mathbb{R}} Q_n(x) \right\}.$$

Enfin, on note $T_n(\mathbb{X})$ le milieu de $\{T_n(\mathbb{X})\}$.

Lemme 1 (Huber (1964)). *La fonction $\mathcal{Q}_n(\cdot)$ est convexe, l'ensemble $\{T_n(\mathbb{X})\}$ n'est pas l'ensemble vide, et est convexe et compact. De plus, si $\rho(\cdot)$ est strictement convexe, alors* $\mathrm{Card}\{T_n(\mathbb{X})\} = 1$.

Ce premier résultat de Huber (1964) indique premièrement l'existence de solutions au problème de minimisation donné par (1.2.10). En second lieu nous pouvons constater les liens de convexité (stricte ou pas) entre les fonctions $\rho(\cdot)$ et $\mathcal{Q}_n(\cdot)$. C'est ce lien qui permet d'avoir l'existence des arguments du minimum, la stricte convexité impliquant l'unicité de la solution.

Lemme 2 (Huber (1964)). *Soit ϕ la dérivée presque partout de ρ et $\lambda(t) = \mathbb{E}[\phi(X - t)]$. S'il existe un réel c tel que $\lambda(t) > 0$ pour tout $t > c$ et $\lambda(t) < 0$ pour tout $t < c$, alors $T_n(\mathbb{X})$ tend vers c presque sûrement.*

Ce deuxième lemme nous donne la consistance des M-estimateurs. La condition sur $\lambda(\cdot)$ indique que ce résultat de consistance reste valable même si la fonction $\lambda(\cdot)$ est discontinue là ou $\mathbb{E}[\rho(Y - \cdot)]$ atteint son minimum.

Nous présentons maintenant un résultat de normalité asymptotique.

Lemme 3 *Nous reprenons les notations du Lemme 2. Supposons que $\lambda(c) = 0$, et que $\lambda(\cdot)$ est continuement dérivable avec $\lambda'(c) < 0$. Si de plus $\mathbb{E}\left[\phi^2(X-t)\right]$ existe pour tout t et est continue en t dans un voisinage de c, on a*

$$n^{1/2}(T_n(\mathbb{X}) - c) \Rightarrow N\left(0, \frac{\mathbb{E}\left[\phi^2(X-c)\right]}{(\lambda'(c))^2}\right).$$

Constatons que les hypothèses sont plus restrictives que celles faites pour la consistance. La fonction $\lambda(\cdot)$ est supposée continue et une hypothèse sur le moment d'ordre 2 de $\phi(X-\cdot)$ est faite pour l'application du théorème de la limite centrale. Dans les applications, il sera intéressant de comparer la variance de la loi normale asymptotique avec d'autres techniques d'estimations. Un bonne revue de la littérature est disponible dans l'article de Koltchinskii (1997).

Fonction de perte quelconque Dans le cas général, il existe plusieurs méthode d'étude des M-estimateurs. Nous présentons quelques résultats de consistance des M-estimateurs.

Considérons une fonction $M(\cdot)$ déterministe définie sur Θ, un espace métrique compact, admettant un minimum en θ_0. Considérons maintenant une suite de fonctions $M_n(\cdot)$, dépendant de l'échantillon $(X_1, \cdots, X_n)$, telle que $M_n(\cdot)$ converge simplement vers $M(\cdot)$ en probabilité. Cette situation inclue par exemple les cas où $M(\theta) = \mathbb{E}\left[m_\theta(X_i)\right]$ et ou $M_n(\theta) = \sum_{i=1}^n m_\theta(X_i)/n$. La question est de savoir à quelles conditions la suites des arguments du minimum $\widehat{\theta}_n$ des fonctions $M_n(\cdot)$ converge vers θ_0 en probabilité. Autrement dit si

$$\widehat{\theta}_n = \operatorname*{Arg\,Min}_{t\in\Theta} M_n(t),$$

à quelles conditions a-t-on $\widehat{\theta}_n \to \theta_0$ en probabilité ? Nous considérerons les estimateurs $\widehat{\theta}$ vérifiant

$$M_n(\widehat{\theta}) \leq \inf_{\theta\in\Theta} M_n(\theta) + o_{\mathbb{P}}(1), \tag{1.2.11}$$

définition légèrement plus faible que le strict minimum de $M_n(\cdot)$. Il est facilement démontrable que la converge simple de $M_n(\cdot)$ vers $M(\cdot)$ n'est en général pas suffisante pour avoir la convergence de $\widehat{\theta}_n$ vers θ_0. Nous supposerons ici que $M_n(\cdot)$ converge vers $M(\cdot)$ uniformément sur Θ.

Cette hypothèse est suffisante pour avoir la consistance, mais pas nécessaire. Néanmoins sa vérification dans certains modèles est relativement facile notamment si Θ est compact. Le théorème ci-dessous semble très connu et se trouve aussi dans van der Vaart (1998).

Théorème 9 *Soit $M_n(\cdot)$ une suite de fonction aléatoires et $M(\cdot)$ une fonction déterministe telles que pour tout $\varepsilon > 0$,*

$$\begin{aligned} \sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| &\stackrel{\mathbb{P}}{\to} 0 \quad , \\ \inf_{\theta,|\theta-\theta_0|\geq\varepsilon} M(\theta) &> M(\theta_0). \end{aligned}$$

Alors toute suite $\widehat{\theta}$ vérifiant (1.2.11) converge en probabilité vers θ_0.

Preuve du Théorème 9. L'équation (1.2.11) nous donne que $\widehat{\theta}$ vérifie $M_n(\widehat{\theta}) \leq M_n(\theta_0) + o_{\mathbb{P}}(1)$. Ceci et la convergence uniforme de $M_n(\cdot)$ vers $M(\cdot)$ en probabilité nous assure que $M_n(\widehat{\theta}) \leq M(\theta_0) + o_{\mathbb{P}}(1)$. Nous avons donc

$$\begin{aligned} 0 \leq M(\widehat{\theta}) - M(\theta_0) &\leq M(\widehat{\theta}) - M_n(\widehat{\theta}) + o_{\mathbb{P}}(1) \\ &\leq \sup_{\theta\in\Theta} |M(\theta) - M_n(\theta)| + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1). \end{aligned} \tag{1.2.12}$$

D'après la deuxième hypothèse du théorème, pour tout $\varepsilon > 0$, il existe un $\eta > 0$ tel que $M(\theta_0) < M(\theta) + \eta$ et donc

$$\mathbb{P}\left(\left|\widehat{\theta} - \theta_0\right| \geq \varepsilon\right) \leq \mathbb{P}\left(M(\widehat{\theta}) - M(\theta_0) \geq \eta\right).$$

Ceci termine la preuve par (1.2.12). □

Dans le cas général, il existe plusieurs méthode d'étude des M-estimateurs. Nous présentons quelques résultats de normalité asymptotique pour les M-estimateurs.

Supposons que l'on dispose d'une suite de v.a. $\widehat{\theta}_n$ convergeant en probabilité vers un paramètre θ_0. Supposons de plus que $\widehat{\theta}_n$ soit caractérisé par l'annulation d'une fonction score

$(\widehat{R}(\widehat{\theta}_n) = 0)$ telle que

$$\begin{aligned} R(\theta_0) &= \mathbb{E}\left[\phi_{\theta_0}(X)\right] = 0, \\ \widehat{R}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\phi_\theta(X_i). \end{aligned}$$

Théorème 10 *Supposons que la fonction $t \mapsto \psi_\theta(t)$ est deux fois continuement dérivable dans un voisinage de θ_0 pour tout t avec pour dérivées $\dot{\psi}_\theta(x)$ et $\ddot{\psi}_\theta(x)$ telles que $\left|\ddot{\psi}_\theta(x)\right| \le \ddot{\psi}(x)$ pour une fonction $\ddot{\psi}(\cdot)$ admettant un moment d'ordre 1. De plus supposons que $\mathbb{E}\left[\ddot{\psi}_{\theta_0}(X)^2\right] < \infty$, $\mathbb{E}\left[\left|\dot{\psi}_{\theta_0}(X)\right|\right] < \infty$ et $\mathbb{E}\left[\dot{\psi}_{\theta_0}(X)\right] \neq 0$. Alors $n^{1/2}(\widehat{\theta}_n - \theta_0)$ converge en loi vers une variable normale centrée et de variance $\mathbb{E}\left[\psi_{\theta_0}(X)^2\right]/\mathbb{E}\left[\dot{\psi}_{\theta_0}(X)\right]^2$.*

b) Représentation de Bahadur pour une fonction de perte convexe

A ce stade, il est intéressant de développer une représentation de Bahadur pour de tels estimateurs car cette représentation permet d'approximer les M-estimateurs par une somme de variables aléatoires et ainsi de ramener leur étude à celle des sommes de variables aléatoires.

Nous présentons une représentation de Bahadur pour les M-estimateurs définis par rapport à une fonction ρ convexe due à Niemiro (1992). Niemiro fournit un tel résultat dans un cadre multivarié, en ne supposant plus que la fonction ρ est telle que $\rho(z;t) = \rho(z-t)$. Nous supposons donc que la variable X est à valeurs dans $\mathbb{R}^d$. Soit α^* un réel tel que $Q(\alpha^*) = \min_{t\in\mathbb{R}} Q(t) = \min_{t\in\mathbb{R}} \mathbb{E}\left[\rho(X,t)\right]$ et soit ϕ la dérivée presque partout de ρ par rapport à t. Dans toutes les hypothèses qui suivent, $|\cdot|$ désigne la norme euclidienne usuelle. Niemiro fait les hypothèses suivantes :

i. La fonction $\rho(z,t)$ est convexe en t pour tout z dans le support de X.

ii. La fonction $Q(t) = \mathbb{E}\left[\rho(X;t)\right]$ existe et est finie pour tout t dans $\mathbb{R}$.

iii. α^* existe et est unique.

iv. Il existe un r tel que $\mathbb{E}\left[|\phi(X;t)|^r\right] < \infty$ pour tout t dans un voisinage de α^*.

v. $\mathbb{E}\left[\exp(a|\phi(X;t))|\right] < \infty$ pour tout t dans un voisinage de α^* et un certain $a > 0$.

vi. $Q(\cdot)$ est deux fois différentiable en α^* et $Q''(\alpha^*)$ est définie positive.

vii. $|Q'(t) - Q''(\alpha^*)(t - \alpha^*)| = O(|t - \alpha^*|^{3/2+s/2})$, lorsque $t \to \alpha^*$.

viii. $\mathbb{E}\left[|\phi(X;t) - \phi(X;\alpha^*)|^2\right] = O(|t - \alpha^*|^{1+s})$, lorsque $t \to \alpha^*$.

ix. Il existe un r tel que $\mathbb{E}\left[|\phi(X;t)|^r\right] = O(1)$.

On a alors le résultat suivant :

Théorème 11 (Niemiro (Theorem 5, 1992)). *Si les hypothèses ci-dessus sont vérifiées pour un s dans $[0,1[$ et pour un $r > [8 + d(1+s)]/(1-s)$, alors on a presque sûrement lorsque n tend vers l'infini,*

$$n^{1/2}(T_n(\mathbb{X}) - \alpha^*) = -(Q''(\alpha^*))^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^{n} \phi(X_i;\alpha^*) + O\left(n^{-(1+s)/4} \log^{1/2}(n)(\log\log(n))^{(1+s)/4}\right).$$

Remarquons que le théorème 2 de Bahadur (1966) est un cas particulier de celui-ci avec $s = 0$ et $d = 1$ (cas univarié).

Parmi les hypothèses faites par Niemiro (1992), c'est l'hypothèse viii. qui donne l'ordre exact dans la représentation de Bahadur. le paramètre s est supposé appartenir à $[0,1[$, la valeur 0 correspondant à l'existence du discontinuité de ϕ en α^*. Lorsque s tend vers 1, cela correspond au cas le plus régulier d'une fonction ϕ continue et Lipschitzienne dans une voisinage de α^*.

On déduit du théorème 11 un résultat de normalité asymptotique :

Corollaire 2 *Sous les hypothèses du théorème 11, on a*

$$n^{1/2}\left(T_n(\mathbb{X}) - \alpha^*\right) \Rightarrow N\left(0, Q''(\alpha^*)^{-1}\mathrm{Var}(\phi(X;\alpha^*))Q''(\alpha^*)^{-1}\right).$$

c) Représentation de Bahadur pour un fonction de perte quelconque

Nous présentons essentiellement le résultat dû à He et Shao (1996). Les auteurs étudient directement la fonction score (dérivée presque partout de la fonction de perte ρ). Cette démarche est aussi utilisée dans Jurečkovà (1985) et dans Jurečkovà et Sen (1987). Dans le papier de He et Shao (1996), les auteurs supposent les variables indépendantes mais pas nécessairement identiquement distribuées. Nous présentons leur résultat dans un cadre i.i.d..

Soit α^* un réel tel que la fonction $t \mapsto \mathbb{E}\left[\phi(X,t)\right]$ s'annule en α^*, pour une certaine fonction ϕ mesurable. Dans toute la suite, la norme $|\cdot|$ désigne la norme infinie. Considérons, sans supposer de forme particulière pour la fonction mesurable ϕ, les estimateurs T_n définis par

$$\sum_{i=1}^{n} \phi(X_i; T_n) = o(\delta_n), \tag{1.2.13}$$

où δ_n est une suite réelle et les X_i des réalisations i.i.d d'une variable aléatoire X. Notons

$$u(z,t,d) = \sup_{|t_1 - t| \leq d} |\phi(z;t_1) - \phi(z;t)|.$$

Les auteurs supposent :

i. Pour tout t dans Θ, ouvert de $\mathbb{R}^m$, $m \geq 1$, la fonction $\phi(z;t)$ est mesurable.

ii. Il existe un α^* tel que $\mathbb{E}\left[\phi(X;\alpha^*)\right] = 0$ et $|T_n - \alpha^*| \to 0$ presque sûrement lorsque n tend vers l'infini.

iii. Il existe $r > 0$, $d_0 > 0$, $a > 0$, tels que $\mathbb{E}\left[u^2(X,t,d)\right] \leq ad^r$ pour tout $|t - \alpha^*| \leq d_0$ et $d \leq d_0$.

iv. Il existe $0 < \beta \leq b$ et $\beta_1 > 0$ tels que $\mathbb{E}\left[u^{2+b}(X,\alpha^*,d)\right] \leq a^{2+\beta_1} d^{(2+\beta)r/2}$, pour tout $d \leq d_0$.

v. Il existe une suite $(s_n)_{n\geq 1}$, telle que $s_n \to \infty$, $s_n \leq an$ pour tout n et telle que

$$\limsup_{n\to\infty} \left(s_n \log\log(n)\right)^{-1/2} \left| \sum_{i=1}^{n} \phi(X_i;\alpha^*) \right| \leq 2,$$

presque sûrement.

vi. $|\mathbb{E}[\phi(X;T_n)]| \geq c_n|T_n - \alpha^*|$ presque sûrement pour une certaine suite déterministe $(c_n)_{n\geq 1}$ strictement positive.

vii. Il existe une matrice non dégénérée $\mathbf{D}_n$ et une suite $(b_n)_{n\geq 1}$ strictement positive telles que

$$|n\mathbb{E}[\phi(X;T_n)] - \mathbf{D}_n(T_n - \alpha^*)| \leq b_n,$$

presque sûrement lorsque n tend vers l'infini.

He et Shao ont alors prouvé :

Théorème 12 *Supposons que toutes les hypothèses ci-dessus sont vérifiées avec* $n/c_n = O(1)$, $|\mathbf{D}_n| = O(n^{-1})$, *et* $b_n = O(n^{1/2-r/4}(\log\log(n))^{r/2})$. *Alors n'importe quelle* T_n *satisfaisant (1.2.13) avec* $\delta_n = O(n^{1/2-r/4})$ *peut s'écrire*

$$T_n - \alpha^* = -\mathbf{D}_n^{-1}\sum_{i=1}^{n}\phi(X_i;\alpha^*) + O(n^{-(1/2+r/4)}(\log\log(n))^{1/2+r/4}),$$

presque sûrement lorsque n *tend vers l'infini.*

Dans ce théorème, le paramètre r est à rapprocher du paramètre s du théorème 5 de Niemiro (1992). En effet, ce paramètre indique le degré de régularité de la fonction score et détermine l'ordre du reste dans la représentation de Bahadur.

1.2.2 Estimation non-paramétrique des M-estimateurs

Hong (2003) a établi une représentation de Bahadur presque sûre pour les M-estimateurs non-paramétriques en utilisant la méthode des polynômes locaux. Le modèle non-paramétrique est

$$Y_i = m(X_i) + \varepsilon_i \quad i = 1, \cdots, n,$$

où les ε_i, $i = 1, \cdots, n$, sont i.i.d. et la fonction m est inconnue. Ce résultat généralise celui de Chaudhuri (1991) établi pour les quantiles bien que le cadre ne soit pas tout à fait le même dans la mesure ou Hong (2003) utilise le lissage à l'aide de noyaux. $m(x)$ est ici l'argument du minimum de la quantité $\mathbb{E}\left[\rho(Y,t)|X=x\right]$ par rapport à la variable t. Hong (2003) suppose que pour tout y dans $\mathbb{R}$, la fonction de perte $\rho(y,\theta)$ est absolument continue en θ, c'est-à-dire qu'il existe une fonction $\phi(y,\theta)$ telle que

$$\rho(y,\theta) = \rho(y,0) + \int_0^\theta \phi(y,t)dt,$$

et

$$\mathbb{E}\left[\phi(Y,m(x))|X=x\right] = 0.$$

Ce modèle englobe entre autre la régression quantile, ainsi que le modèle défini par rapport à la fonction de Huber (1967)

$$\rho_k(u) = \left(\frac{u^2}{2}\right)\mathbb{I}(|u|<k) + \left(k|u| - \frac{k^2}{2}\right)\mathbb{I}(|u|\geq k).$$

Soit $\widehat{\beta}_n(x) = (\widehat{m}_0(x,p,h),\ldots,p!\widehat{m}_p(x,p,h))$ l'argument du minimum de la quantité

$$\sum_{i=1}^n \rho(Y_i,\sum_{j=0}^p (X_i-x)^j t_j)K_h(X_i-x),$$

par rapport à la variable $t=(t_0,\cdots,t_p)$ dans $\mathbb{R}^{p+1}$ et où $K(\cdot)$ est un noyau de probabilité assez régulier. Soit

$$\beta_p(x) = (\beta_{p,0}(x),\ldots,\beta_{p,p}(x)) = \left(m(x),\cdots,\frac{m^{(p)}(x)}{p!}\right)^T,$$

de façon à définir le terme linéaire dans la représentation de Bahadur

$$\beta^*_{n,r}(x) = \frac{1}{nh}\sum_{i=1}^n K^*_{n,r}\left(\frac{X_i-x}{h}\right)\phi\left(Y_i,\sum_{j=0}^p (X_i-x)^j\beta_{p,j}(x)\right),$$

où $K^*_{n,r}$ est un noyau re-normalisé. Ce terme sera explicité plus loin. Pour l'étude du biais, la fonction $m(\cdot)$ est supposée $p+2$ fois dérivable et telle que $m^{(p+2)}(\cdot)$ soit continue. L'estimation étant faite ponctuellement autour d'un point x du support de X supposé compact. Il existe $C>0$ tel que pour tout t assez petit,

$$\mathbb{E}\left[(\phi(Y,t+a)-\phi(Y,a))^2|X=u\right] \leq C|t|,$$

uniformément en (a,u) dans un voisinage de $(m(x),x)$, et

$$|\phi(Y,u+t)-\phi(Y,u)| \leq C,$$

pour tout y, u dans $\mathbb{R}$. De plus, on note $G(t,u) = \mathbb{E}\left[\phi(Y,t)|X=u\right]$ et on pose $G_i(t,u) = (\partial^i/\partial t^i)G(t,u)$ pour $i=1,2$. On suppose $G_1(m(x),x) > 0$ et $G_2(t,u)$ est continue dans un voisinage de $(m(x),x)$. Pour rapprocher le résultat de Hong (2003) à celui de Chaudhuri (1991), observons que la variable explicative est supposée appartenir à un espace de dimension 1 dans l'article de Hong (2003). Donc pour reprendre les notations de l'article de Chaudhuri (1991), constatons que $A=(1,\cdots,s)$ et le résulat de Hong (2003) s'écrit

Théorème 13 Hong (2003). *On suppose que pour un certain réel $s \geq 0$, la fenêtre h vérifie $h \to 0$, $nh \to \infty$ et $nh^{2(p+s)+3} = O(1)$ lorsque n tend vers l'infini. Alors pour tout r dans A*

$$h^r\left(\widehat{m}(x,p,h) - m^{(r)}(x)\right) = \beta^*_{n,r}(x) + R_r(x,p,h),$$

et où le terme de reste $R_r(x,p,h)$ vérifie

$$R_r(x,p,h) = O_{\mathbb{P}}\left((nh)^{-\lambda(s)}\right),$$

avec

$$\lambda(s) = \min\left(\frac{p+1}{p+s+1}, \frac{3(p+1)+2s}{4(p+s+1)}\right).$$

1.2.3 Extensions : Approche uniforme

Une des questions qui peuvent se poser ensuite est de savoir si la représentation de Bahadur pour les M-estimateurs localement polynomiaux est toujours valable pour les "pires" valeurs des paramètres en jeu, que sont par exemple la valeur de la variable de conditionnement ou encore la taille de la fenêtre. Kong, Linton et Xia (2008) proposent une représentation de Bahadur uniforme par rapport au paramètre de conditionnement pour les M-estimateurs localement polynomiaux.

L'article de Kong, Linton et Xia (2008) utilise des notations proches de celui de Hong (2003). Néanmoins, ces auteurs développent une représentation de Bahadur uniforme par rapport à la variable explicative X supposée dans leur article multi-dimensionnelle de dimension d. Nous présentons leur résultat dans un cadre i.i.d. bien que leur résultats soient présentés dans le contexte plus général de données corrélées.

Considérons une fonction de perte $\rho(\cdot;\cdot)$. Les auteurs se proposent d'estimer une fonction de régression multivariée

$$m(x_1,\dots,x_d) = \operatorname*{Arg\,Min}_t \mathbb{E}\left[\rho(Y;t)|X=(x_1,\dots,x_d)^\top\right],$$

à partir d'un échantillon $(X_i,Y_i)_{1\le i\le n}$. Soient $\varepsilon_i = Y_i - m(X_i)$, $i=1,\dots,n$. Les auteurs font les hypothèses suivantes : On définit pour tout $M>2$, λ_2 dans $]0,1[$ et λ_1 dans $]\lambda_2,(1+\lambda_2)/2]$, les suites

$$\begin{aligned} d_n &= (nh^d/\log(n))^{-(\lambda_1+\lambda_2)/2}(nh^d\log(n))^{1/2} \\ r(n) &= (nh^d/\log(n))^{(1-\lambda_2)/2} \\ M_n &= M(nh^d/\log(n))^{-\lambda_1}. \end{aligned}$$

i. Pour tout y et t dans $\mathbb{R}$, $\rho(y;t)=\rho(y-t)$. La fonction $\rho(\cdot)$ est absolument continue. La densité des v.a. ϵ_i est supposée bornée et satisfait $\mathbb{E}\left[\phi(\varepsilon_i)|X_i\right]=0$ où $\phi(\cdot)$ est la dérivée presque partout de $\rho(\cdot)$. De plus, il existe $\nu_1>2$ tel que $\mathbb{E}\left[|\phi(\varepsilon_i)|^{\nu_1}\right]<\infty$.

ii. $\phi(\cdot)$ est une fonction lipshitzienne sur les intervalles $(-\infty,a_0]$, $(a_0,a_1]$, $\cdots$, (a_D,∞), où les a_j, $j=0,\cdots,D$ sont les points de discontinuité de ϕ.

iii. Le noyau $K(\cdot)$ est supposé à support compact et satisfaisant à une condition de Lipschitz sur son support.

iv. La densité $f(\cdot)$ des X_i est supposée bornée, dérivable et de dérivée bornée.

v. La fonction $m(\cdot)$ est supposée appartenir à la classe de Hölder $\mathcal{C}(L,s)$.

vi. La densité $f(\cdot|\cdot)$ est supposée bornée.

vii. La fenêtre h tend vers 0. De plus

$$\begin{aligned} nh^d/\log(n) &\to \infty, \\ nh^{d+(s+2)/\lambda_2}/\log(n) &= O(1), \\ n^{-1}r(n)^{\nu_2/2}d_n M_n &\to \infty, \end{aligned}$$

pour un certain $2 < \nu_2 < \nu_1$.

On a alors,

Théorème 14 Kong, Linton et Xia (2008). *Soit $s \geq 0$ tel que les hypothèses ci-dessus soient vérifiées avec $\lambda_2 = (p+1)/2(p+s+1)$. Alors, pour tout compact D de $\mathbb{R}^d$ et pour tout r dans $\mathbb{N}^d$ tel que $0 \leq |r| \leq p$, on a*

$$h^r \left(\widehat{m}(x,p,h) - m^{(r)}(x)\right) - \beta^*_{n,r}(x) = R_r(x,p,h),$$

et où le terme de reste $R_r(x,p,h)$ vérifie

$$\sup_{x \in D} |R_r(x,p,h)| = O_{p.s.}\left((nh)^{-\lambda(s)}\right),$$

avec

$$\lambda(s) = \min\left(\frac{p+1}{p+s+1}, \frac{3(p+1)+2s}{4(p+s+1)}\right).$$

Moyennant un terme logarithmique dans les hypothèses du théorème ci-dessus, les auteurs arrivent à la même conclusion que celle du théorème de Hong (2003) uniformément en x dans n'importe quel compact de $\mathbb{R}^d$.

1.3 Apports de la thèse

1.3.1 Présentation de la deuxième partie

L'estimation localement polynomiale (LP) robuste d'une fonction de régression et de ses dérivées a été un sujet grandement étudié dans de nombreux articles de recherche. Voir en particulier Tsybakov (1986) pour l'estimation ponctuelle optimale et Fan and Gijbels (1996) pour une revue de la littérature. Le modèle étudié dans papier présenté dans la deuxième partie est proche de ceux étudiés dans les articles de Truong (1989), Chauduri (1991) and Kong, Linton and Xia (2009).

Truong (1989) à montré que l'estimation de la médiane conditionnelle peut atteindre les vitesses d'estimation optimales décrites par Stone (1982) au sens minimax. Chaudhuri (1991) a établi une représentation de Bahadur pour l'estimateur LP de la fonction de quantile conditionnel pour un paramètre de fenêtre préalablement défini et pour une variable explicative fixée. L'article de Kong et al. (2009) développe une représentation de Bahadur pour une classe plus générale de M-estimateurs qui de plus est uniforme par rapport à la variable explicative. Dans la deuxième partie de ce mémoire, nous étendons ces résultats en donnant une représentation de Bahadur pour l'estimateur LP du quantiles conditionnel, uniforme par rapport au niveau du quantile, la variable explicative et la fenêtre d'estimation.

Dans un premier temps, nous étudions le biais l'estimateur LP du quantile conditionnel. La plupart des résultats dans ce domaine supposent que l'ordre p de l'estimateur LP doit être inférieur ou égal à l'ordre de régularité de la fonction à estimer. Nous donnons un résultat sous l'hypothèse que l'ordre de l'estimateur LP est plus grand que l'ordre de régularité de la fonction à estimer. Nous montrons que les estimateurs des dérivées supposées existantes convergent aux vitesses optimales de Stone (1982). Pour ce qui est des dérivées "non-existantes", c'est-à-dire lorsque l'ordre de l'estimateur LP est plus grand que l'ordre de régularité de la fonction à estimer, il est intéressant de constater que notre estimateur peut diverger.

Dans un deuxième temps, nous donnons une représentation de Bahadur, puis nous utilisons cette représentation dans différents cadres. En effet, grâce à notre représentation de Bahadur uniforme, nous montrons que nos estimateurs convergent aux vitesses optimales décrites par Stone (1982) aux sens L_m, $0 < m \leq \infty$. Ces résultats peuvent être combinées aux procédures décrites dans Li and Racine (2008) pour la construction d'une fenêtre dépendant des données.

Le théorème de la limite centrale (TLC) pour le quantile conditionnel fait intervenir dans le terme de variance la quantité $f\left(Q(\alpha|x)|x\right)^{-1}$ appelée fontion de sparsité et où $Q(\alpha|x)$ et $f(\cdot|x)$

sont respectivement le quantile conditionnel de Y sachant $X = x$ et la densité conditionnelle de Y sachant $X = x$. Cette quantité mesure la concentration de la distribution de Y sachant $X = x$ dans le voisinage du quantile d'ordre α. En un mot, plus la concentration est grande, plus la densité est grande et plus la variance est petite. Pour appliquer statistiquement le TLC à l'estimateur du quantile conditionnel et/ou ses dérivées, nous devons avoir un estimateur de la fonction de sparsité. Nous proposons un estimateur de cette quantité basé sur l'estimation de la fonction de quantile conditionnel. En effet, nous savons que $f(Q(\alpha|x)|x)$ n'est autre que la dérivée de la fonction de quantile par rapport à la variable α. En renforçant nos hypothèses de régularité sur la fonction de quantile, nous obtenons un estimateur de la fonction de sparsité.

1.3.2 Présentation de la troisième partie

La troisième partie consiste en un article à rapprocher des résultats récents de Hong (2003) and Kong et al. (2009). Hong (2003) considère une fonction de perte général et propose une représentation de Bahadur pour l'estimateur LP ponctuelle par rapport a la variable explicative et à la fenêtre d'estimation. Kong et al. (2009) ont proposé d'étendre ce résultat en fournissant une représentation de Bahadur uniforme par rapport à la variable explicative.ainsi que dans un cadre de données non-indépendantes.

Dans notre travail, nous étendons les résultats de Kong et al. (2009) et Hong (2003) en donnant une représentation de Bahadur pour le M-estimateur LP qui soit uniforme par rapport à la variable explicative et uniforme par rapport à la fenêtre h dans un intervalle $\left[\underline{h}, \overline{h}\right]$. Cette dernière propriété est par exemple utile pour des procédures de validation croisée afin d'estimer la fenêtre optimale à partir des données. Une autre application possible de ce type de résultats est la construction de procédures de tests dans les cadres de Horowitz et Spokoiny (2001) et Guerre and Lavergne (2005). Ces procédures combinent des statistiques non-paramétriques avec différentes fenêtres pour créer des tests d'ajustement de fonctions de régression.

1.3.3 Présentation de la quatrième partie

En tirant avantage des résultats de la deuxième partie, nous donnons la loi limite de la déviation maximale de l'estimateur de la fonction de quantile conditionnelle et de ses dérivées. Ce résultat étend celui de Härdle et Song (2009) puisqu'il est donné dans le cadre plus général de LP. Nous présentons aussi une méthode d'estimation de la variance pour rendre statistiquement applicables nos résultats.

Bibliographie

[1] BAHADUR R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577-580.

[2] BASSETT G .W. & H. CHEN (2001). Portfolio style : return-based attribution using quantile régression. *Emp. Econ.* **26**, 293-305.

[3] CHAUDHURI P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19**, 760-777.

[4] CLEVELAND W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journ. Amer. Statist. Assoc.* **74**, 829-836.

[5] COLE T.J. & P.J. GREEN (1992). Smoothing reference centile curves : the LMS method and penalized likelihood. *Statist. Med.* **11**, 1305-1319.

[6] EL BANTLI, F. & M. HALLIN (1999). L1-estimation in linear models with heterogeneous white noise. *Statist. Proba. lett.* **45**, 305-315.

[7] FAN, J. & I. GIJBELS (1996) *Local polynomial modelling and its applications.* Monographs on Statist. Appl. Proba. 66. Chapman & Hall.

[8] GUERRE, E. & P. LAVERGNE (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840-870.

[9] HAMPEL F. (1974). The influence curve and its role in robust estimation. *Journ. Amer. Statist. Assoc.* **69** 383-393.

[10] Härdle W. & T. Gasser (1984). Robust nonparametric function fitting. *Journ. Roy. Statist. Soc. Ser. B* **46**, 42-51.

[11] Härdle W. & S. Song (2009). Confidence bands in quantile regression. Forthcoming in *Econometric Theory.* Published online by Cambridge University Press 04 Nov 2009.

[12] Hastie T. & C. Loader (1993). Local regression : Automatic kernel carpentry (with discussion). *Statist. Sci.* **8**, 139-143.

[13] He X. & Q. Shao (1996) A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24**, 2608-2630.

[14] He X. & P.D. Shi (1994) Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journ. Nonparam. Statist.* **3**, 299-308.

[15] Hjort N. & D. Pollard (1993). Asymptotics for minimisers of convex processes. *Preprint*, available at `http://www.stat.yale.edu/~pollard/Papers/`.

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

[17] Hong S.Y. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M-regression. *Journ. Nonparam. Statist.* **15**, 237-251.

[18] Honore B., J. Powell & S. Khan (2000). Quantile regression under random censoring. *preprint.*

[19] Horowitz, J.L. & V.G. Spokoiny (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599-631.

[20] Huber P.J. (1964) Robust estimation of a location parameter. *Ann. Math. Statist.* **37**, 73-101.

[21] Huber P.J. (1967). Robust regression, asymptotics, conjectures and Monte-Carlo. *Ann. Statist.* **1**, 799-821.

[22] Huber P.J. (1981). *Robust statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1981.

[23] Jurečkovà, J. (1985). Representation of M-estimators with the second-order asymptotic distribution. *Statist. Decis.* **3**, 263-276.

[24] Jurečkovà, J. & P.K. Sen (1987). A second-order asymptotic distributional representation of *M*-estimators with discontinuous score functions. *Ann. Probab.* **15**, 814-823.

[25] Kiefer J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38**, 1323-1342.

[26] Kiefer J. (1970). Deviations between the sample quantile process abd the sample df. *Proc. Conference on Nonparametric Techniques in Statistical Inference,* Cambridge University Press, 299-319.

[27] Knight K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755-770.

[28] Koenker R. (2005) *Quantile regression.* Cambridge University Press.

[29] Koenker R. & G. Basset (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[30] Koenker R. & G. Basset (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**, 43-61.

[31] Koenker R. & O. Geiling (2001). Reappraising medfly longetivity : a quantile regression approach. *Journ. Amer. Statist. Assoc.* **96**, 458-468.

[32] Koltchinskii V.I. (1997). *M*-estimation, convextiy and quantiles. *Ann. Statist.* **25**, 435-477.

[33] Kong, E., O. Linton & Y. Xia (2009). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Forthcoming in *Econometric Theory.*

[34] Li Q. & J. S. Racine (2008). Nonparametric Estimation of Conditional CDF and Quantile Functions With Mixed Categorical and Continuous Data. *Journ. Busi. Eco. Statist.* **26**, 423-434.

[35] Niemiro W. (1992). Asymptotics for *M*-estimators defined by convex minimization. *Ann. Statist.* **20**, 1514-1533.

[36] Powell J.L. (1986). Censored regression quantiles. *Journ. Eco.* **32**, 143-155.

[37] Royston P. & D.G. Altman (1994). Regression using fractional polynomials of continuous covariates : parsimonious parametric modelling (with discussion). *Appl. Statist.* **43** 429-467.

[38] Ruppert D. & M.P. Wand (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** (3) 1346-1370.

[39] Serfling R.J. (1980) *Approximation theorems of mathematical statistics.* John Wiley & Sons, New York, U.S.A.

[40] Shi P. (1995). Asymptotic behaviour of nonparametric conditional quantile estimates for time series. *Canad. Journ. Statist.* Vol. 23, No 2, 161-169.

[41] Stigler S.M. (1986). *The history of statistics : the measurements of uncertainty before 1900.* Cambridge, MA, Harvard University Press.

[42] Stone C.J. (1977). Consistent nonparametric regression. With discussion and a reply by the author. *Ann. Statist.* **5**, 595-645.

[43] Stone C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

[44] Truong Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** (2), 606-617.

[45] Tsybakov A.B. (1986). Robust reconstruction of functions by a local approximation method. *Prob. Inform. Transm.* **22** (2), 133-146.

[46] van der Vaart A. W. (1998). *Asymptotic statistics.* Cambridge University Press.

[47] WILCOX R.R. & H.J. KESELMAN (2006) Detecting heteroscedasticity in a simple regression model via quantile regression slopes *Journ. Statist. Comput. Simu.* **76** (8), 705-712.

[48] YU K. & M.C. JONES (1998). Local linear quantile regression. *Journ. Amer. Statist. Assoc.* **93**, 228-237.

Chapter 2

LP estimators of the conditional quantile function

Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function [1]

Emmanuel Guerre [2]

Camille Sabbah [3]

Abstract

This paper investigates the bias and the 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. Financial Support from the Department of Economics, Queen Mary University of London, is gratefully acknowledged.

2. Department of Economics, Queen Mary University of London. E-mail: `e.guerre@qmul.ac.uk`

3. Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 175 rue du Chevaleret, Paris, France. E-mail : `camille.sabbah@univ-paris1.fr`

2.1 Introduction

Consider independent and identically observations $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ where Y is a real random variable and X is a random vector of dimension d. Define, for α in $(0,1)$, the loss function

$$\ell_\alpha(q) = |q| + (2\alpha-1)q = 2q\left(\alpha - \mathbb{I}(q \le 0)\right),\ q \text{ in } \mathbb{R}, \tag{2.1.1}$$

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

$$Q\left(\alpha\,|x\right) = \arg\inf_{q\in\mathbb{R}} \mathbb{E}\left[\ell_\alpha(Y-q)\,|X=x\right] \tag{2.1.2}$$

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 $\mathbf{b} = (b_0,\ldots,b_p)^T$,

$$\widehat{\mathbf{b}}(\alpha;h,x) = \arg\min_{\mathbf{b}\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \ell_\alpha\left(Y_i - b_0 - b_1\left(X_i - x\right) - \cdots - \frac{b_p}{p!}\left(X_i-x\right)^p\right) K\left(\frac{X_i - x}{h}\right). \tag{2.1.3}$$

In the expression above, $p!$ is the factorial $p\times(p-1)\times\cdots\times 1$, $K(\cdot)$ is a kernel function and h is a smoothing parameter which goes to 0 with the sample size. 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 (2.1.3) instead of the loss function $\ell_\alpha(\cdot)$. A Taylor expansion

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

suggests that $\widehat{b}_1(\alpha;h,x),\ldots,\widehat{b}_p(\alpha;h,x)$ estimate the partial derivatives $\partial Q(\alpha|x)/\partial x$ provided $Q(\alpha|x)$ is smooth enough. 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 $d>1$.

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), Chauduri (1991) and Kong, Linton and Xia (2009). 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. Chauduri (1991) obtained a Bahadur representation for the local polynomial quantile estimators when the kernel function $K(\cdot)$ of (2.1.3) is uniform. In few words, a Bahadur expansion 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. The Bahadur representation of Chaudhuri (1991) is pointwise, that is holds for some prescribed x and α and a given deterministic bandwidth $h \to 0$. As explained and illustrated in Kong et al. (2009), 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. (2009) derived a uniform Bahadur representation for robust local polynomial estimators. Here uniformity is with respect to the location variable x and Kong et al. (2009) mostly focus on the study of the remainder term under the difficult framework of dependent observations. In this work, we extend the scope of uniformity to the quantile level α and the bandwidth h. We study the bias term and 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 $p < s$ can be easily dealt with by ignoring derivatives of order higher than $p + 1$, we focus in the more interesting case where $p \geq s$, which has apparently not been considered in the statistical and econometric literature. As shown in Corollary 1, a local polynomial quantile estimator with $p \geq s$ 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 $\widehat{b}_v(\alpha; h, x)$, $v = s + 1, \ldots, p$ can diverge and Proposition 1 describes a simple example where it indeed happens. Such finding contrasts with standard regression models where standard t-tests can be used to remove a useless covariate. In the local polynomial setup, a high value of the vth t-statistic may also correspond to a non smooth quantile function in which case a lower degree $p < v$ could have been used. This shows that relying on standard interpretation of t-statistics is misleading in the context of local methods.

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 Lipshitz quantile functions, or Chauduri (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 $[\underline{h}, \overline{h}]$. Proposition 2 shows that a random bandwidth performs as well as its deterministic equivalent counterpart with respect to convergence rates of the uniform norm $\sup_x \left|\widehat{Q}_h(\alpha|x) - Q(\alpha|x)\right|$. 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.

A fifth contribution exploits uniformity with respect to the quantile order α. Uniformity in α is important *per se* for quantiles due to graphical representations that uses several values of α when plotting $x \mapsto Q(\alpha|x)$ to better illustrate the dependence relation between Y and X. Proposition 3 extends this scope of applications by considering estimation of the conditional quantile density function

$$q(\alpha|x) = \frac{\partial Q(\alpha|x)}{\partial \alpha} = \frac{1}{f\left(Q(\alpha|x)|x\right)}. \tag{2.1.4}$$

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 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 a 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 $Q(\alpha|x) = F^{-1}(\alpha|x)$. Marmer and Shneyerov (2008) proceed similarly. Xiang (1995) proposes the estimator

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

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 (2.1.3) 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.

2.2 Main assumptions and notations

The definition (2.1.3) of $\widehat{Q}_h(\alpha|x)$ assumes that the covariate X is univariate. In the multivariate case, we use a multivariate kernel function $K(z) = K(z_1, \ldots, z_d)$ but we restrict to an univariate bandwidth for the sake of simplicity. The univariate polynomial expansion $b_0 + b_1 (X_i - x) + \cdots + b_p (X_i - x)^p / p!$ is replaced by a multivariate counterpart as defined now. Let $\mathbb{N}$ be the set of natural integer numbers. For $\mathbf{v} = (v_1, \ldots, v_d)$ let $|\mathbf{v}| = v_1 + \cdots + v_d$ 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 $\mathbb{R}^P$,

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

and $\mathbf{v}! = \Pi_{i=1}^d v_i!$. 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 (2.1.3) is

$$\begin{aligned} \widehat{\mathbf{b}}(\alpha; h, x) &= \arg\min_{\mathbf{b}\in\mathbb{R}^P} \mathcal{L}_n(\mathbf{b}; \alpha, h, x) \text{ with} \\ \mathcal{L}_n(\mathbf{b}; \alpha, h, x) &= \frac{1}{nh^d} \sum_{i=1}^n \ell_\alpha \left(Y_i - \mathbf{U}(X_i - x)^T \mathbf{b} \right) K \left(\frac{X_i - x}{h} \right). \end{aligned} \tag{2.2.1}$$

As in the univariate case, the entry $\widehat{b}_{\mathbf{0}}(\alpha; h, x) = \widehat{Q}_h(\alpha|x)$ 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. A convenient way to assume its existence is to suppose that $Q(\cdot|\cdot)$ is in a suitable modification of the standard Hölder class considered in Chauduri (1991). 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 $\mathcal{X}$,

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

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 $\left\{z \in \mathbb{R}^d : \|z\| \leq 1\right\}$.

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* $\mathbb{R}^d$*, with compact support* $\mathcal{K}$ *and satisfies* $\int K(z)dz = 1$*. For some* $\underline{K} > 0$*,* $K(z) \geq \underline{K}\, \mathbb{I}\left(z \in \mathcal{B}\left(0,1\right)\right)$*. The bandwidth is in* $[\underline{h}_n, \overline{h}_n]$ *with* $0 < \underline{h}_n \leq \overline{h}_n < \infty$*,* $\lim_{n\to\infty} \overline{h}_n = 0$ *and* $\lim_{n\to\infty}(\log n)/(n\underline{h}_n^d) = 0$*.*

Assumption X is standard. Assumption F ensures uniqueness of the conditional quantile function $Q(\alpha|x) = F^{-1}(\alpha|x)$ in (2.1.2) and existence of the quantile density function (2.1.4). Assumption K allows for a wide range of smoothing parameters $h \to 0$ in $[\underline{h}_n, \overline{h}_n]$. 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) assume $h^d \geq C(\log n)/n$ to obtain uniform consistency so that Assumption K is fairly general.

2.3 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 $\widehat{\mathbf{b}}(\alpha; h, x)$ of (2.2.1) is an estimator of $\mathbf{b}^*(\alpha; h, x)$ with

$$\mathbf{b}^*(\alpha; h, x) = \arg\min_{\mathbf{b}\in\mathbb{R}^P} \mathbb{E}\left[\ell_\alpha\left(Y - \mathbf{U}\left(X - x\right)^T \mathbf{b}\right) K\left(\frac{X - x}{h}\right)\right]. \tag{2.3.1}$$

In particular, the conditional quantile estimator $\widehat{Q}_h(\alpha|x) = \widehat{b}_{\mathbf{0}}(\alpha; h, x)$ is an estimator of $Q_h^*(\alpha|x) = b_{\mathbf{0}}^*(\alpha; h, x)$ which 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

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

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 $b_{\mathbf{v}}\left(\alpha|x\right)$ 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* $\overline{h}$ *is small enough, there is a constant* C *such that for all* $|\mathbf{v}| \leq \lfloor s \rfloor$ *and* n *large enough,*

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

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

$$b^*_{\mathbf{v}}(\alpha;h,x) = O\left(h^{s-|\mathbf{v}|}\right) \quad \text{for } |\mathbf{v}| \geq s \tag{2.3.2}$$

uniformly in $(\alpha,h,x) \in [\underline{\alpha},\overline{\alpha}] \times [\underline{h},\overline{h}] \times \mathcal{X}_0$. See also Loader (1999, Theorem 4.2) which gives a less precise $b^*_{\mathbf{v}}(\alpha;h,x) = o\left(h^{-|\mathbf{v}|}\right)$. 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

$$Y = m(X) + \varepsilon, \quad m(x) = \begin{cases} |x|^{1/2} & \text{if } x \geq 0 \\ -|x|^{1/2} & \text{if } x < 0 \end{cases}, \tag{2.3.3}$$

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, $Q(\alpha|x) = \Phi^{-1}(\alpha) + m(x)$ 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 $x' \to 0$. The next Proposition uses the behavior of $m(\cdot)$ at $x = 0$ to show that the rate given in (2.3.2) is sharp.

Proposition 1 *Suppose that (X, Y) satisfies (2.3.3). Let $\mathbf{b}^*(\alpha; h, x) = (b_0^*(\alpha; h, x), b_1^*(\alpha; h, x))^T$ from (2.3.1) 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 $h^{-1/2}$,*

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

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 Taylor expansion of order p for x' in the neighborhood of x, $Q(\alpha|x') \simeq \mathbf{U}(x' - x)^T \mathbf{b}_p(\alpha|x)$. Consider the following counterpart of the Taylor approximation,

$$Q^*(x'; \alpha, h, x) = \mathbf{U}(x' - x)^T \mathbf{b}^*(\alpha, h, x) \tag{2.3.4}$$

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

$$\mathbf{S}_i(\alpha; h, x) = 2\left\{\mathbb{I}\left(Y_i \leq Q^*(X_i; \alpha, h, x)\right) - \alpha\right\} \mathbf{U}\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right), \tag{2.3.5}$$

$$\mathbf{J}_i(\alpha; h, x) = 2f\left(Q^*(X_i; \alpha, h, x) | X_i\right) \mathbf{U}\left(\frac{X_i - x}{h}\right) \mathbf{U}\left(\frac{X_i - x}{h}\right)^T K\left(\frac{X_i - x}{h}\right). \tag{2.3.6}$$

Since

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

and (2.1.1) gives

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

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

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

almost everywhere. Although the criterion function of (2.2.1) is not twice differentiable, 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

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

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

$$\beta_n(\alpha; h, x) = -\left(\frac{1}{nh^d} \sum_{i=1}^{n} \mathbf{J}_i(\alpha; h, x)\right)^{-1} \frac{1}{\left(nh^d\right)^{1/2}} \sum_{i=1}^{n} \mathbf{S}_i(\alpha; h, x). \tag{2.3.7}$$

It follows from Lemma A.1 in Appendix A that $\beta_n(\alpha; h, x)$ is asymptotically centered with $\beta_n(\alpha; h, x) = O_{\mathbb{P}}(1)$. Our Bahadur representation result studies the error term

$$\mathbf{E}_n(\alpha; h, x) = \left(nh^d\right)^{1/2} \mathbf{H}\left(\widehat{\mathbf{b}}(\alpha; h, x) - \mathbf{b}^*(\alpha; h, x)\right) - \beta_n(\alpha; h, x). \tag{2.3.8}$$

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) and 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

$$\mathbb{L}_{1n}(\beta; \alpha, h, x) \quad = \quad nh^d \left\{\mathcal{L}_n\left(\mathbf{b}^*(\alpha; h, x) + \frac{H^{-1}\beta}{\left(nh^d\right)^{1/2}}; \alpha, h, x\right) - \mathcal{L}_n\left(\mathbf{b}^*(\alpha; h, x); \alpha, h, x\right)\right\}$$

$$= \sum_{i=1}^{n} \left\{ \ell_\alpha \left(Y_i - Q^*(X_i;\alpha,h,x) - \frac{\mathbf{U}\left(\frac{X_i-x}{h}\right)^T}{(nh^d)^{1/2}}\beta \right) - \ell_\alpha\left(Y_i - Q^*(X_i;\alpha,h,x)\right) \right\} \times K\left(\frac{X_i-x}{h}\right),$$

which is such that

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

It then follows from (2.3.8) that

$$\begin{aligned} \mathbf{E}_n(\alpha;h,x) &= \arg\min_\epsilon \mathbb{L}_n\left(\beta_n(\alpha;h,x),\epsilon;\alpha;h,x\right) \text{ where} \\ \mathbb{L}_n\left(\beta,\epsilon;\alpha;h,x\right) &= \mathbb{L}_{1n}(\beta+\epsilon;\alpha,h,x) - \mathbb{L}_{1n}(\beta;\alpha,h,x). \end{aligned} \quad (2.3.9)$$

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

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

$$\mathbb{L}_n^0\left(\beta,\epsilon;\alpha;h,x\right) = \frac{1}{(nh^d)^{1/2}} \sum_{i=1}^{n} \mathbf{S}_i(\alpha;h,x)^T\epsilon + \frac{1}{2}\epsilon^T\left(\frac{1}{nh^d}\sum_{i=1}^{n} \mathbf{J}_i(\alpha;h,x)\right)(\epsilon+2\beta), \quad (2.3.10)$$

and $\mathbb{R}_n$ is a remainder term. Indeed, as noted in Fan et al. (1995) 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 which plays here the role of the Bernstein inequality used in the pointwise framework of Hong (2003).

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

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

In the case where the lower and upper bandwidths $\underline{h}$ and $\overline{h}$ have the same order, Theorem 2 gives uniformly in h in $[\underline{h}, \overline{h}]$, 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) obtain 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. (2009) 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 $[\underline{\alpha}, \overline{\alpha}]$,*

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 (a.s.) 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 a.s. 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 (2.3.3), 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.

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 $\lim_{n\to\infty}(\log n)/(nh_n^d) = 0$. 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 $|\mathbf{v}| \le \lfloor s \rfloor$,*

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

In particular, if the exact order of $\widehat{h}_n$ is $(\log(n)/n)^{1/(2s+d)}$ in probability,

$$\sup_{x\in\mathcal{X}_0}\left|\widehat{b}_{\mathbf{v}}(\alpha;\widehat{h},x) - b_{\mathbf{v}}(\alpha|x)\right|$$

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 (2.1.4). The considered estimator of $q(\alpha|x)$ is a conditional version of the convolution estimator of Parzen (1979),

$$\widehat{q}(\alpha|x) = \frac{1}{h_q}\int \widehat{Q}_h(a|x)dK_q\left(\frac{a-\alpha}{h_q}\right) = \frac{1}{h_q}\int \widehat{Q}_h(\alpha + h_q t|x)dK_q(t), \qquad (2.3.11)$$

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 $dK_k(t) = K_q'(t)dt$, substituting in (2.3.11) 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 (2.3.11). If $K_q(\cdot)$ is a linear combination of Dirac masses at t_j with weights κ_j, $j = 1, \ldots, J$, 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

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

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 $\mathcal{X}$, $\alpha \in [\underline{\alpha}, \overline{\alpha}] \mapsto q(\alpha|x)$ is $\lfloor s \rfloor$th differentiable;

iii. For each x in $\mathcal{X}$ and all $(\alpha, \alpha') \in [\underline{\alpha}, \overline{\alpha}]^2$

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

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

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

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 holds with $h = O(h_q)$, $h_q \to 0$ and $(\log n)/(nh^d) \to 0$. Then for any x in $\mathcal{X}_0$ and α in $(\underline{\alpha}, \overline{\alpha})$,*

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

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 $(\log^{3/4} n)\left(nh^{d+2}\right)^{-1/4} O_{\mathbb{P}}\left(\left(nh^{d+1}\right)^{-1/2}\right)$ 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 $nh^{d+2} \to \infty$, 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 $nh^{d+2} \to \infty$ 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 (2.1.4), coincides with the optimal rate for pointwise estimation of $f(y|x)$.

2.4 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. (2009). 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.

2.5 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, $\|\mathbf{A}\| = \sup_{\mathbf{u}\in\mathcal{B}(0,1)} \|\mathbf{Au}\| = \sup_{\mathbf{u}\in\mathcal{B}(0,1)} |\mathbf{u}^T\mathbf{Au}|$ is the largest eigenvalue in absolute value of $\mathbf{A}$. This norm is such that $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$ for any matrix or vector $\mathbf{B}$. Denote by $\|\cdot\|_\infty$ the uniform norm, $\|f(\cdot|\cdot)\|_\infty = \sup_{(x,y)\in\mathbb{R}^d\times\mathbb{R}} |f(y|x)|$. 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 (2.3.4), (2.3.5) and (2.3.6). We abbreviate $\underline{h}_n$ and $\overline{h}_n$ into $\underline{h}$ and $\overline{h}$. Define

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

where $\mathcal{X}_0$ is as in Assumption X and $[\underline{\alpha}, \overline{\alpha}] \subset (0,1)$ 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 (2.2.1), define

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

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 $\widehat{\mathbf{B}}(\theta) = \mathbf{H}\widehat{\mathbf{b}}(\theta)$ 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 $\overline{h}$ 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 $\overline{h}$ 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*

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

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

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

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 *and* x' *and* $\mathcal{X}$,

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

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 $\mathcal{X}_0 \times \mathcal{K}$,

$$Q(\alpha|x+hz) = \sum_{0\leq|\mathbf{v}|\leq\lfloor s\rfloor} \frac{b_{\mathbf{v}}(\alpha|x)}{\mathbf{v}!}(hz)^{\mathbf{v}} + \sum_{|\mathbf{v}|=\lfloor s\rfloor} \frac{(hz)^{\mathbf{v}}}{\mathbf{v}!}\left(b_{\mathbf{v}}(\alpha|x+thz) - b_{\mathbf{v}}(\alpha|x)\right)$$

$$= \mathbf{U}(z)^T \mathbf{H}\mathbf{b}(\alpha|x) + \epsilon(\theta, z). \tag{A.3}$$

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 $\mathbf{b}(\alpha|x) = (b_{\mathbf{v}}(\alpha|x), |\mathbf{v}| \leq \lfloor s \rfloor, 0, \ldots, 0)^T \in \mathbb{R}^P$. Since $Q(\cdot|\cdot) \in \mathcal{C}(L, s)$,

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

Let

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

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

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

A Taylor expansion with integral remainder gives

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

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

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

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 $\overline{h}$ small enough give that uniformly in θ in Θ^0 and $\mathbf{A}$ in $\mathbb{R}^P$,

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

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

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

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

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

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

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

uniformly in θ in Θ^0. This ends the proof of the Theorem and also establishes (2.3.2) since $\mathbf{b}(\alpha|x) = (b_{\mathbf{v}}(\alpha|x), |\mathbf{v}| \leq \lfloor s \rfloor, 0, \ldots, 0)^T$. □

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 (2.3.3) is such that

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

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

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

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

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

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 (2.3.10), where

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

Define also

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

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

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

which are such that

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

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,*

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

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,*

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

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 (2.3.7), 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

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

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

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

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,

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

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

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

$$= \frac{C}{t_n} \frac{\log^{3/4} n}{(n\underline{h}^d)^{1/4}} \left(\left(C_\eta + \frac{t_n}{\log^{1/2} n} \right)^{1/2} + \left(\frac{\log n}{n\underline{h}^d} \right)^{1/4} \left(C_\eta + \frac{2t_n}{\log^{1/2} n} \right)^2 \right).$$

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

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

Step 2: $\sup_{\theta\in\Theta^1} \|\mathbf{E}_n(\theta)\|$ *and* $\sup_{(\epsilon,\theta)\in\mathcal{B}(0,t_n)\times\Theta^1} |\mathbb{R}_n(\beta_n(\theta),\epsilon;\theta)|$. 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 $\mathbb{L}_n = \mathbb{L}_n^0 + \mathbb{R}_n$

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

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

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

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

$$\left\{ \sup_{\theta\in\Theta^1} \|\mathbf{E}_n(\theta)\| \geq t_n \right\} \subset \bigcup_{\theta\in\Theta^1} \left\{ \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0\left(\beta_n(\theta),\epsilon;\theta\right) - \sup_{\|\epsilon\|=t_n} \left|\mathbb{R}_n\left(\beta_n(\theta),\epsilon;\theta\right)\right| \leq 0 \right\}$$

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

Consider first $\inf_{\theta\in\Theta^1} \inf_{\|\epsilon\|=t_n} \mathbb{L}_n^0 \left(\beta_n(\theta), \epsilon; \theta\right)$. The definition (2.3.10) of $\mathbb{L}_n^0$ gives, for any ϵ with $\|\epsilon\| = t_n$,

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

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

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

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

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)), (2.3.5) and $h^d \geq C(\log n)/n$ give

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

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 $\overline{h} = Ch_n$. The condition on h_n ensures that $\underline{h}$ and $\overline{h}$ satisfy Assumption K for all $C > 1$. Theorems 1 and 2 give, for all $C > 1$,

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

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

A.6 Proof of Proposition 3

Substituting (2.3.8) in (2.3.11) yields

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

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

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

Hence it remains to show that

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

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

The two next steps establish these two equalities.

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

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

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

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

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

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

Step 2: proof of (A.15). Let $\theta_t = (\alpha + h_q t, h, x)$, $\theta = \theta_0$. Since $\int dK_q(t) = 0$, (2.3.7) gives

$$\begin{aligned}
&\frac{1}{h_q^{1/2}}\int \beta_n(\theta_t) dK_q(t) = \frac{1}{h_q^{1/2}}\int (\beta_n(\theta_t) - \beta_n(\theta)) dK_q(t) \\
&= \frac{1}{h_q^{1/2}}\int \left\{\left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(\theta)\right)^{-1} - \left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(\theta_t)\right)^{-1}\right\} \frac{\sum_{i=1}^n \mathbf{S}_i(\theta) dK_q(t)}{(nh^d)^{1/2}} \quad \text{(A.16)} \\
&\quad + \frac{1}{h_q^{1/2}}\int \left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(\theta_t)\right)^{-1} \frac{1}{(nh^d)^{1/2}}\sum_{i=1}^n \{\mathbf{S}_i(\theta) - \mathbf{S}_i(\theta_t)\} dK_q(t). \quad \text{(A.17)}
\end{aligned}$$

Since $\mathbf{A} \mapsto \mathbf{A}^{-1}$ is Lipshitz over the set of semi-definite positive matrices $\mathbf{A}$ with smallest eigenvalue bounded from below by $\underline{\gamma}$, Lemma A.2, (2.3.6) and Assumption F yield that (A.16) satisfies

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

The definition (2.3.4) 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,

$$\frac{1}{h_q^{1/2}}\int \frac{1}{nh^d}\sum_{i=1}^n |Q^*(X_i;\theta_t) - Q^*(X_i;\theta)| \, \mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right) |dK_q(t)|$$

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

This together $\sum_{i=1}^{n} \mathbf{S}_i(\theta)/(nh^d)^{1/2} = O_{\mathbb{P}}(1)$ 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, $\mathbb{E}[\mathbf{S}_i(\theta_t)] = 0$, (2.3.5), $Q^*(X;\theta_t) = Q^*(X;\theta) + O(h_q)$ 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

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

2.6 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

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

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

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

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

Proof of (i). It is sufficient to show that $\mathbf{B}^*(\theta) = \arg\min_{\mathbf{B}\in\mathbb{R}^P}\widetilde{\mathcal{L}}(\mathbf{B};\theta)$ exists and is unique. Note that $\mathbf{B}\mapsto\widetilde{\mathcal{L}}(\mathbf{B};\theta)$ is convex by (B.1) because $\ell_\alpha(\cdot)$ is convex. Since $\lim_{|t|\to+\infty}\ell_\alpha(t) = +\infty$ and $\mathbf{U}(z)^T\mathbf{B}$ diverges almost everywhere when $\|\mathbf{B}\|$ diverges, equation (B.1) implies that $\lim_{\|\mathbf{B}\|\to\infty}\widetilde{\mathcal{L}}(\mathbf{B};\theta) = \infty$. Hence $\widetilde{\mathcal{L}}(\mathbf{B};\theta)$ has a minimum. We show that this minimum is unique by showing that $\mathbf{B}\mapsto\widetilde{\mathcal{L}}(\mathbf{B};\theta)$ is strictly convex for all θ in Θ^0. We compute the first and second $\mathbf{B}$-derivatives of $\widetilde{\mathcal{L}}(\mathbf{B};\theta)$. Equation (2.1.1) gives that for almost all $\mathbf{B}$,

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

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

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

$$\begin{aligned} &= 2\int F\left(\mathbf{U}(z)^T\mathbf{B}|x+hz\right)f(x+hz)\mathbf{U}(z)K(z)dz \\ &\quad -2\alpha\int f(x+hz)\,\mathbf{U}(z)K(z)\,dz. \end{aligned} \tag{B.2}$$

Applying again the Dominated Convergence Theorem yields that

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

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

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

Hence $\widetilde{\mathcal{L}}^{(2)}(\cdot;\theta)$ 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 $\widetilde{\mathcal{L}}(\mathbf{B};\theta)$ achieves it minimum for a unique $\mathbf{B}^*(\theta)$.

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

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

see (B.2), so that (A.1) is proved. In particular, observe that $\mathbf{B}^*(\alpha;0,x)$ is the unique solution of $\widetilde{\mathcal{L}}^{(1)}(\mathbf{B};\alpha,0,x)=0$. 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 $\mathbf{B}_0^T(\alpha|x)=(Q(\alpha|x),0,\ldots,0)$ 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 $\widetilde{\Theta}^0$ and give bounds for $\mathbf{B}^*(\theta)$, $\partial\widetilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta);\theta)/\partial\theta^T$ and $\widetilde{\mathcal{L}}^{(2)}(\mathbf{B}^*(\theta);\theta)$. As shown above, $\mathbf{B}\mapsto\widetilde{\mathcal{L}}^{(1)}(\mathbf{B};\theta)$ is continuously differentiable and $\widetilde{\mathcal{L}}^{(2)}(\mathbf{B};\theta)$ 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 $\widetilde{\Theta}^0$ provided $\overline{h}$ is small enough. Hence the Dominated Convergence Theorem and (B.2) yield that $\widetilde{\mathcal{L}}^{(1)}(\mathbf{B};\theta)$ is continuously differentiable in θ over $\widetilde{\Theta}^0$. Then the Implicit Function Theorem (see e.g. Zeidler (1985), p.130) and the first-order condition $\widetilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta);\theta) = 0$ yields that $\mathbf{B}^*(\theta)$ is continuously differentiable in θ over $\widetilde{\Theta}^0$, with

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

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

$$\begin{aligned} \lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \|\mathbf{B}^*(\theta) - \mathbf{B}^*(\alpha;0,x)\| &= 0, \\ \lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\| \frac{\partial \widetilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta);\theta)}{\partial\theta^T} - \frac{\partial \widetilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\alpha;0,x);\alpha,0,x)}{\partial\theta^T} \right\| &= 0. \end{aligned} \tag{B.7}$$

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 $\overline{h}$ is small enough. Then (B.3) and (B.4) give that uniformly in θ in Θ^0,

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

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

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\| \frac{\partial \mathbf{B}^*(\theta)}{\partial\theta^T} \right\| \le C \left\| \left(\int_{\mathcal{B}(0,1)} \mathbf{U}(z)\mathbf{U}(z)^T dz \right)^{-1} \right\| \lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\| \frac{\partial \widetilde{\mathcal{L}}^{(1)}(\mathbf{B}^*(\theta);\theta)}{\partial\theta^T} \right\| \le C. \tag{B.8}$$

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 $h \ge \underline{h}$,

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

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

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

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 $\overline{h}$ 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 $\mathcal{X}$. □

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

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

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

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

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

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

The Cauchy-Schwarz inequality give

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

Hence Assumption F and (B.9) give

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

Then, under Assumptions K and X,

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

Define

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

The next lemma studies coverings of $\mathcal{F}$ with brackets $\left[\underline{R},\overline{R}\right]$. Recall that the bracket

$$
\left[\underline{R},\overline{R}\right]=\left[\underline{R}(X,Y),\overline{R}(X,Y)\right]
$$

is the set of random variables $r=r(X,Y)$ such that $\underline{R}\le r\le\overline{R}$ 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*

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

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

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

ii. *Let τ in $(0,1)$ be a bracket length. There is an set of brackets*

$$\mathcal{I}_\tau = \left\{\left[\underline{R}_{j,\tau}, \overline{R}_{j,\tau}\right], 1 \leq j \leq e^{H(\tau)}\right\},$$

such that

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

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

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

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

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

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

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

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

$$R(\beta;\epsilon,\theta) \;=\; \widetilde{R}(\beta+\epsilon;\theta)-\widetilde{R}(\beta;\theta). \tag{B.12}$$

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[\left|R(\beta,\epsilon;\theta)-\mathbb{E}\left[R(\beta,\epsilon;\theta)\right]\right|^k\right]\\
&\quad=\mathbb{E}\left[\left|\widetilde{R}(\beta+\epsilon;\theta)-\mathbb{E}\left[\widetilde{R}(\beta+\epsilon;\theta)\right]-\left(\widetilde{R}(\beta;\theta)-\mathbb{E}\left[\widetilde{R}(\beta;\theta)\right]\right)\right|^{k-2}\right.\\
&\qquad\left.\times\left|R(\beta,\epsilon;\theta)-\mathbb{E}\left[R(\beta,\epsilon;\theta)\right]\right|^2\right]\\
&\quad\le\left(2\times\frac{\overline{w}}{2}\right)^{k-2}\operatorname{Var}\left(R(\beta,\epsilon;\theta)\right)\le\overline{w}^{k-2}\overline{\sigma}^2.
\end{aligned}$$

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

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

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

$$H(t) \;\le\; C\log\left(\frac{nt}{\tau}\right). \tag{B.15}$$

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

$$\widetilde{\mathcal{F}}_{t_\beta}\subset\bigcup_{1\le j\le e^{H_1(\tau)}}\left[\underline{R}_j^1,\overline{R}_j^1\right],\quad \widetilde{\mathcal{F}}_{t_\beta+t_\epsilon}\subset\bigcup_{1\le j\le e^{H_2(\tau)}}\left[\underline{R}_j^2,\overline{R}_j^2\right],$$

$H_1(\tau)\le H_2(\tau)=H(\tau,t_\beta+t_\epsilon)$. Consider a $R(\beta,\epsilon;\theta)$ in $\mathcal{F}$. Since $\widetilde{R}(\beta;\theta)\in\left[\underline{R}_{j_1}^1,\overline{R}_{j_1}^1\right]$ and $\widetilde{R}(\beta+\epsilon;\theta)\in\left[\underline{R}_{j_2}^2,\overline{R}_{j_2}^2\right]$ for some j_1 and j_2, (B.12) implies that $R(\beta,\epsilon;\theta)\in\left[\underline{R}_{j_2}^2-\overline{R}_{j_1}^1,\overline{R}_{j_2}^2-\underline{R}_{j_1}^1\right]$.

Hence these $e^{H'(\tau)}$ brackets form a covering of $\mathcal{F}$ with, using (B.14) and (B.15),

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

Step 2: Preliminary results for the construction of a covering of $\widetilde{\mathcal{F}}_t$. 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

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

Under Assumption K

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

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

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

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

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

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

$\rho-$(i) $\quad 0 \leq \underline{\rho}(q,\delta) \leq \rho(q,\eta) \leq \overline{\rho}(q,\delta) \leq 1$ for all (q,δ), with$\underline{\rho}(q,\delta) = \rho(q,\eta) = \overline{\rho}(q,\delta)$ if$(q,\delta) \in \mathbb{R}^2 \setminus D_\eta$,

$\rho-$(ii) $\quad \sup_{(q,\delta)\in D_\eta}\left(\left|\frac{\partial\underline{\rho}(q,\delta)}{\partial q}\right| + \left|\frac{\partial\underline{\rho}(q,\delta)}{\partial \delta}\right| + \left|\frac{\partial\overline{\rho}(q,\delta)}{\partial q}\right| + \left|\frac{\partial\overline{\rho}(q,\delta)}{\partial \delta}\right|\right) \leq C\eta^{-1/2}$,

$\rho-$(iii) $\quad D \subset D' = \left\{(q,\delta) \in \mathbb{R}^2; |q| \leq C\eta^{-1/2} \text{ or } |q-\delta| \leq C\eta^{-1/2}\right\}$.

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

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

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

$$\underline{R}(\beta,\theta) \leq \widetilde{R}(\beta,\theta) \leq \overline{R}(\beta,\theta). \tag{B.16}$$

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

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

Hence Step 1, ρ-(i,ii), (B.9) and the Taylor inequality give for all (β,θ), (β',θ') in $\mathcal{B}(0,t)\times\Theta^1$,

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

Arguing symmetrically gives

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
& +C\frac{(1+t)^2}{n\underline{h}^{2p+2+d}}\left(\eta^2+\eta\right) \\
\leq \quad & \frac{8h_j^d\left\|\beta\right\|^2}{nh_j^d}\int K^2(z)\left\|\mathbf{U}(z)\right\|^2 \\
& \times\left[\int\int_0^1 \mathbb{I}\left(\left(y-Q^*(x_j+h_jz;\theta_j),v\delta\left(\beta_j,\theta_j\right)\right)\in D\right)dvf(y|x_j+h_jz)dy\right]f(x_j+h_jz)dz \\
& +C\frac{(1+t)^2}{n\underline{h}^{2p+2+d}}\left(\eta^2+\eta\right) \\
& \leq C\frac{(1+t)^2}{n\underline{h}^{2p+2+d}}\left(\eta^2+\eta+\eta^{1/2}\right).
\end{aligned}
$$

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

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

Hence (B.14) holds if η satisfies

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

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

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

It then follows for n large enough

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

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

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

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

Let $H(\cdot)$, $\overline{\sigma}$ and $\overline{w}$ be as in Lemma B.2. Recall that $t_\beta + t_\epsilon \geq 1$ and that $\overline{\sigma} < 1 \leq n(t_\beta + t_\epsilon)$ 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

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

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

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

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

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

Substituting gives

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

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

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

The definition (2.3.6) of $\mathbf{J}(\theta)$ gives

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

Define

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

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

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

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

$$|r(\beta;\theta)| \leq \frac{\overline{w}'}{2}, \quad \overline{w}' \asymp \frac{(t_\beta+t_\epsilon)^3}{\left(n\underline{h}^d\right)^{3/2}}.$$

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

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

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

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

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

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

$$\leq \quad C\frac{t_\epsilon\left(t_\epsilon+t_\beta\right)^2}{\left(n\underline{h}^d\right)^{1/2}},$$

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

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

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,

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

Hence for all θ in Θ^1,

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

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

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

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

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

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

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

endequation* 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,

$$\mathbf{u}^T M(\theta)\mathbf{u} = \frac{C}{h^d}\mathbb{E} \sum_{0\leq|\mathbf{v}_1|,|\mathbf{v}_2|\leq p} u_{\mathbf{v}_1} u_{\mathbf{v}_2} \left(\frac{X-x}{h}\right)^{\mathbf{v}_1+\mathbf{v}_2} K\left(\frac{X-x}{h}\right)$$

$$= C \sum_{0\leq|\mathbf{v}_1|,|\mathbf{v}_2|\leq p} u_{\mathbf{v}_1} u_{\mathbf{v}_2} \int z^{\mathbf{v}_1+\mathbf{v}_2} K(z) f(x+hz)dz = C \int \left(\sum_{0\leq|\mathbf{v}|\leq p} u_{\mathbf{v}} z^{\mathbf{v}}\right)^2 K(z) f(x+hz)dz$$

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

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 $\inf_{\theta\in\Theta^1} \underline{\gamma}_n(\theta) \geq \inf_{\theta\in\Theta^1} \inf_{\|\mathbf{u}\|=1} \mathbf{u}^T \mathbb{M}_n(\theta)\mathbf{u} \geq \underline{\gamma} + o_{\mathbb{P}}(1)$. □

B.5 Proof of Lemma A.3

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

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

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

$$\left|\frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}}\right| \leq \overline{w}'', \quad \overline{w}'' \asymp \left(n\underline{h}^d\right)^{-1/2},$$

$$\mathrm{Var}\left(\frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}}\right) \leq \mathbb{E}\left[\left(\frac{\mathbf{S}_{\mathbf{v},i}(\theta)}{(nh^d)^{1/2}}\right)^2\right] \leq \mathbb{E}\left[\left(\frac{\left(\frac{X_i-x}{h}\right)^{\mathbf{v}} K\left(\frac{X_i-x}{h}\right)}{(nh^d)^{1/2}}\right)^2\right] = \frac{h^d}{nh^d}\int \left(z^{\mathbf{v}} K(z)\right)^2$$

$$\leq (\overline{\sigma}'')^2, \quad \overline{\sigma}'' \asymp n^{-1/2}.$$

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

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

The Markov inequality then shows that the Lemma is proved. □

Bibliography

[1] CHAUDURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics*, **19**, 760-777.

[2] CHOW, Y.S. & H. TEICHER (2003). *Probability theory. Independence, interchangeability, martingales*. Third edition, Springer.

[3] EINMAHL, U. & D.M. MASON (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, **33**, 1380-1403.

[4] FAN, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87**, 998-1004.

[5] FAN, J. & I. GIJBELS (1996). *Local polynomial modeling and its applications.* Chapman & Hall/CRC.

[6] 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.

[7] GUERRE, E., I. PERRIGNE & Q. VUONG (2000). Optimal nonparametric estimation of first price auctions. *Econometrica*, **68**, 525-574.

[8] GUERRE, E., I. PERRIGNE & Q. VUONG (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, **77**, 1193-1227.

[9] HJORT N. & POLLARD D. (1993). Asymptotics for minimisers of convex processes. Unpublished manuscript, `http://www.stat.yale.edu/~Pollard/Papers/`.

[10] Hong, S.Y. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M-regression. *Journal of Nonparametric Statistics*, **15**, 237-251.

[11] Kong, E., O. Linton & Y. Xia (2009). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Forthcoming in *Econometric Theory*.

[12] Lee, K.L. & E.R. Lee (2008). Kernel methods for estimating derivatives of conditional quantiles. *Journal of the Korean Statistical Society*, **37**, 365-373.

[13] 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.

[14] Loader, C. (1999). *Local regression and likelihood*. Springer.

[15] Marmer, V. & A. Shneyerov (2008). Quantile-based nonparametric inference for first-price auctions. Document paper.

[16] 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.

[17] Parzen, E. (1979). Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, **74**, 105-121.

[18] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10**, 1040-1053.

[19] Tsybakov A.B. (1986). Robust reconstruction of functions by a local approximation method. *Problems of Information Transmission* **22** (2), 133-146.

[20] Truong, Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *The Annals of Statistics*, **17**, 606-617.

[21] van de Geer, S. (1999). *Empirical processes in M-estimation*. Cambridge University Press.

[22] van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge University Press.

[23] White, H. (1994). *Estimation, inference and specification analysis. Econometric Society Monographs*. Cambridge University Press.

[24] Xiang, X. (1995). Estimation of conditional quantile density function. *Journal of Nonparametric Statistics*, **4**, 309-316.

[25] Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications, I. Fixed-Point Theorems*. Springer Verlag.

Chapter 3

Bahadur representation for LP M-estimators

Uniform bias study and Bahadur representation for local polynomial M-estimators

Camille Sabbah [1]

Abstract

This paper investigates the bias and the Bahadur representation of a local polynomial M-estimator. The bias and Bahadur remainder term are studied uniformly with respect to the the covariates and the smoothing parameter. The order of the local polynomial M-estimator can be higher that the differentiability order of the function to be estimated. Applications of the results deal with global optimal consistency rates of the local polynomial M-estimator and performance of random bandwidths.

Keywords: Bahadur representation; Local polynomial estimation; M-estimator; Uniform consistency.

1. Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 175 rue du Chevaleret, Paris, France. E-mail : camille.sabbah@univ-paris1.fr

3.1 Introduction

Consider independent and identically distributed (i.i.d.) observations (X,Y), (X_i,Y_i), $i=1,\dots,n$, where Y is a real variable and X is d-dimensional. Let $\rho(\cdot)$ be a loss function and suppose that there is an unique $m(x)$ satisfying

$$\mathbb{E}\left[\rho(Y-m(x))|X=x\right]=\min_t \mathbb{E}\left[\rho(Y-t)|X=x\right]. \qquad (3.1.1)$$

The minimization problem in (3.1.1) is a generalization of the classical least square case $\rho(t)=t^2$, which gives that $m(x)=\mathbb{E}\left[Y|X=x\right]$. Another interesting case is when $\rho(t)=|t|$ which gives that $m(x)$ is the conditional median of Y. In a robust parametric estimation framework, Huber (1964) has proposed the choice of various loss functions $\rho(\cdot)$ which can be less sensitive to outliers than the standard square loss function. See Huber (1981) for an overview. In the present work we are interested in the nonparametric local polynomial (LP) estimation of $m(\cdot)$. Namely $m(\cdot)$ is not supposed to belong to any parametric class of functions and we aim to estimate $m(\cdot)$ and its derivatives. LP extension of (3.1.1) have been widely studied, see for example Cleveland (1979), Tsybakov (1986), Müller (1987), Chaudhuri (1991), Fan (1993), Hastie and Loader (1993), Fan, Hu and Truong (1994) and Ruppert and Wand (1994) among others. Fan and Gijbels (1996) gave an good review of LP methods and discuss their advantages over classical kernel smoothing methodology.

In the univariate case, LP M-estimation builds on the following pth order Taylor expansion

$$m(x)\approx m(x_0)+(x-x_0)m^{(1)}(x_0)+\cdots+\frac{(x-x_0)^p}{p!}m^{(p)}(x_0),$$

where $p! = 1\times 2\times\cdots\times p$, and aims to estimate all the derivatives $m^{(j)}(x_0)$, $j=0,...,p$ based on an i.i.d. sample (X_i,Y_i), $i=1,\dots,n$. Let $K(\cdot)$ be a kernel function and consider a bandwidth parameter $h=h(n)$ which goes to 0 when n diverges. The vector of derivatives $(m(x),m^{(1)}(x),\dots,m^{(p)}(x))^\top$ is estimated using $\widehat{\mathbf{b}}(h,x)$ which achieves the minimum of

$$\sum_{i=1}^n \rho\left(Y_i-b_0-(X_i-x)b_1-\cdots-\frac{(X_i-x)^p}{p!}b_p\right)K\left(\frac{X_i-x}{h}\right),$$

w.r.t. $\mathbf{b} = (b_0, \dots b_p)^\top$, the so-called LP M-estimator of $m(x)$ and its derivatives.

When the square loss function $\rho(t) = t^2$ is used the resulting least squares estimator $\widehat{\mathbf{b}}(h, x)$ is a linear combination of Y_i, $i = 1, \dots, n$. This important property facilitates the theoretical study of the LP least squares estimator. However linearity is lost in many other cases as for the popular case $\rho(t) = |t|$ which corresponds to LP estimation of the conditional median. In the parametric setup, Bahadur (1966) first provided a representation of the median estimator as a sum of i.i.d. variables up to a small remainder term. From this so-called Bahadur representation, the study of the behavior of the sample quantile is reduced to the study of a sum of i.i.d. variables. Extension of the Bahadur representation to the LP M-estimation setup has been widely studied. This representation is useful to derive for example asymptotic normality see e.g. Bhattacharya and Gangopadhyay (1990), Chaudhuri (1991), Welsh (1996), Hong (2003) and Kong, Linton and Xia (2009) among others.

The present paper is perhaps more related to the works of Hong (2003) and Kong et al. (2009). Hong (2003) considered general loss functions and proposes Bahadur representation for the LP M-estimator that is pointwise and with a deterministic bandwidth. Kong et al. (2009) proposes to extend this result by providing a Bahadur representation for the LP M-estimator in the difficult framework of data-dependence. Furthermore, their representation holds uniformly with respect to the explanatory variable x over a compact subset of the range of X. Kong et al. (2009) motivate the interest of their uniform results with many applications including the important case of semiparametric plug-in estimation.

In the present work, we go one step further studying the bias term and the Bahadur representation for the LP M-estimator uniformly with respect to the explanatory variable and the bandwidth parameter. Hence a potentially interesting development of the paper is a Bahadur representation which holds uniformly with respect to the bandwidth. Such a result can be useful to study cross validation bandwidth choices and the properties of the resulting local polynomial estimators. Another potential interesting application is adaptive testing with M-estimators, extending so the procedures of Horowitz et Spokoiny (2001) and Guerre and Lavergne (2005) which combines nonparametric kernel statistics associated with different bandwidth to obtain a lack of fit test for the regression function.

A first contribution given in Theorem 1 below deals with the study of the bias of LP M-estimators. In the LP bias study, most of authors has considered that the order p of the LP M-estimator is equal to $s-1$ where s is the order of differentiability of $x \mapsto m(x)$. Considering this amounts to assume that s is known, which is not possible in most of cases. The case where $p < s$ can be dealt with by ignoring derivatives of order higher than $p+1$. In this paper we consider the more interesting case $p \geq s$. As shown in Corollary 1, assuming that $p \geq s$ still allows to estimate $m(x)$ with the optimal rate $n^{-s/(2s+1)}$ of Stone (1982). This suggests that LP M-estimators using high order p should be preferred since they allow to estimate in an optimal way a wider range of smooth M-estimate functions. An interesting consequence of our bias study is that the additional LP coefficients $\widehat{b}_v(h,x)$, $v = s, \ldots, p$ can diverge and Proposition 1 describes a simple example where it indeed happens. Fortunately this does not affect the nonparametric M-estimator of the existing derivatives $\widehat{b}_v(h,x), v = 0, \ldots, s-1$.

A distinct group of contributions involves our uniform study of the Bahadur remainder term, namely Theorem 2, which 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 LP M-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.

A fourth contribution uses the fact that Theorems 1 and 2 hold uniformly with respect to h in an interval $\left[\underline{h}, \overline{h}\right]$. Proposition 2 shows that a random bandwidth performs as well as its deterministic equivalent counterpart with respect to convergence rates of the uniform norm $\sup_x \left\|\widehat{b}_v(h,x) - m^{(v)}(x)\right\|$.

The rest of the paper in organized as follows. In Section 3.2 are given the main Assumptions and notations for the estimator together a short discussion on the Assumptions. We give our main results in Section 3.3 including the bias study and the Bahadur representation. Finally the proofs of main results are gathered in Appendix A and the proofs of intermediary results are gathered in Appendix B.

3.2 Main Assumptions and Notations

Let us now extend our LP M-estimation strategy to the case where X is d-dimensional. LP M-estimators of $m(x)$ builds on Taylor expansions

$$Q(x') = \sum_{|\mathbf{v}|\leq p} b_{\mathbf{v}} \frac{(x'-x)^{\mathbf{v}}}{\mathbf{v}!},$$

where $\mathbf{v} = (v_1, \ldots, v_d)$ is a d-dimensional vector with integer numbers entries, $|\mathbf{v}| = v_1 + \cdots + v_d$, $x^{\mathbf{v}} = \prod_{j=1}^{d} x_j^{v_j}$ and $\mathbf{v}! = \prod_{j=1}^{d}(v_j!)$. Let P be the number of coefficients of $Q(\cdot)$, namely $P = \text{Card}\left\{\mathbf{v} \in \mathbb{N}^d, |\mathbf{v}| \leq p\right\}$ where $\mathbb{N}$ stands for the set of natural integer numbers, and stack the $b_{\mathbf{v}}$'s and $(x'-x)^{\mathbf{v}}/\mathbf{v}!$ in P-dimensional vectors $\mathbf{b}$ and $\mathbf{U}(x'-x)$ using lexicographic order so that $Q(x') = \mathbf{U}(x'-x)^{\top}\mathbf{b}$.

The LP M-estimator of the Taylor coefficients $\widehat{\mathbf{b}}(h,x)$ achieves the minimum over the set $\mathbb{R}^P$ of P-dimensional vectors of

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

In particular, the first entry $\widehat{b}_{0,\ldots,0}(h,x) = \widehat{b}_0(h,x)$ of $\widehat{\mathbf{b}}(h,x)$ is an estimator of $m(x)$. More generally $\widehat{b}_{\mathbf{v}}(h,x)$ is an estimator of the $\mathbf{v}$-th partial derivatives

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

provided this quantity exists. To ensure that this quantity exists, we will suppose, as in Chaudhuri (1991), that the map $m(\cdot)$ belongs to a slightly modified version of the Hölder class of functions. Let $s > 0$ be a strictly real number that will be an order of smoothness. Define the lowest integer part $\lfloor s \rfloor$ of the real number s as the unique integer number such that $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$. The function $m(\cdot)$ is said to be in the Hölder smoothness class $\mathcal{C}(L, s)$ if

– $m(\cdot)$ is $\lfloor s \rfloor$-times differentiable over the support of X.

– $|b_{\mathbf{v}}(x) - b_{\mathbf{v}}(x')| \leq L \|x - x'\|^{s - \lfloor s \rfloor}$, for $L > 0$, all $\mathbf{v} = (v_1, \ldots, v_d)$ in $\mathbb{N}^d$ with $|\mathbf{v}| = \lfloor s \rfloor$, and all x and x' in the support of X,

and where $\| \cdot \|$ stands for the Euclidean norm. Define $\mathbf{b}(x) = (b_{\mathbf{v}}(x), \mathbf{v} \leq \lfloor s \rfloor, 0, \ldots, 0)^\top$ the vector of derivatives of $m(\cdot)$ at x up to order $\lfloor s \rfloor$ completed with 0's to achieve dimension P.

In the following we use the fact that a function of bounded variation on $\mathbb{R}$ can be represented as the difference of two nondecreasing functions on $\mathbb{R}$, see Kolmogorov and Fomin (1969, p.331). More specifically, we shall assume that the differential $-r(\cdot)$ of the function $\rho(\cdot)$ has bounded variations so that it can be written at the difference of nondecreasing functions,

$$r(\cdot) = r_1(\cdot) - r_2(\cdot). \tag{3.2.1}$$

We will focus on those x which are in an inner subset $\mathcal{X}_0$ of the support $\mathcal{X}$ of X. Let $\mathcal{B}(0,1)$ be the closed ball $\left\{ z \in \mathbb{R}^d : \|z\| \leq 1 \right\}$. Define for all x in the range of X, the functions $t \in \mathbb{R} \mapsto R(t|x) = \mathbb{E}[\rho(Y - t)|X = x]$ and $R^{(j)}(t|x) = \partial^j R(t|x)/\partial t^j$, $j = 1, 2, 3$ provided these quantities exist. Positive constants are denoted by the generic letter C and may vary from line to line and $\mathbb{I}(\cdot)$ stands for the indicator function. Our main Assumptions are as follows.

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 R *i. For any x in $\mathcal{X}$, the function $t \in \mathbb{R} \mapsto R(t|x)$ has a unique minimum achieved for $t = m(x)$. There is an increasing continuous function $t \in \mathbb{R} \mapsto \psi(t)$ such that $\psi(0) = 0$ and satisfying for all x in $\mathcal{X}$, $R(t|x) - R(m(x)|x) \geq \psi(|t - m(x)|)$, for all t in $\mathbb{R}$.*

ii. There is an increasing continuous subadditive function $t \in \mathbb{R} \mapsto \Psi(t)$ such that $\Psi(0) = 0$, $\lim_{t \to \infty} \Psi(t) = \infty$, $m(x)$ achieves the minimum over $\mathbb{R}$ of $t \mapsto \mathbb{E}\left[\Psi\left(|Y - t|\right)|X = x\right]$, $\Psi\left(|\cdot|\right) \leq \rho(\cdot)$ and $\Psi(|\cdot|) \leq C\,|\cdot|$ over $\mathbb{R}$.

iii. The map $(t, x) \in \mathbb{R} \times \mathcal{X} \mapsto R(t|x)$ is three times continuously differentiable. The second order partial derivative $R^{(2)}(t|x)$ is such that there is $C > 0$ such that for all x in $\mathcal{X}$

$$\inf_{t \in [m(x) - C, m(x) + C]} R^{(2)}(t|x) \geq C > 0.$$

Assumption L *i. The loss function $\rho(\cdot) \geq 0$ is continuous and piecewise differentiable over $\mathbb{R}$ with derivative $-r(\cdot)$. Furthermore, $r(\cdot)$ is piecewise continuous and of bounded variation on $\mathbb{R}$.*

ii. For some $a > 1$ and for all x in $\mathcal{X}$,

$$\mathbb{E}\left[(r_j(Y-u-t) - r_j(Y-u))^2 \,|X=x\right] \leq C \max\left(|t|, |t|^a\right)\ , \quad j=1,2,$$

for all (u,t) in $\mathbb{R}^2$ and where $r_1(\cdot)$ and $r_2(\cdot)$ are as in (3.2.1).

iii. There is some positive constant $\nu > 2$ such that

$$\sup_{x\in\mathcal{X}} \mathbb{E}\left[\sup_{|t|\leq C} |r(Y-t)|^\nu \,|X=x\right] < \infty \ \ \text{and} \ \ \sup_{x\in\mathcal{X}} \mathbb{E}\left[|\rho(Y)|^\nu |X=x\right] < \infty.$$

Assumption M *The function $x \in \mathcal{X} \mapsto m(x)$ is in the Hölder smoothness class $\mathcal{C}(L,s)$.*

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 $\underline{K} > 0$, $K(z) \geq \underline{K}\ \mathbb{I}\left(z \in \mathcal{B}(0,1)\right)$. The bandwidth h is in $\left[\underline{h}, \overline{h}\right]$ with $0 < \underline{h} \leq \overline{h} < \infty$, $\lim_{n\to\infty} \overline{h} = 0$, $\lim_{n\to\infty} \log n / \left(n\underline{h}^d\right) = 0$ and $\liminf_{n\to\infty} n^{1-2/\nu}\underline{h}^d / \log n > 0$, where ν is as in Assumption L-(iii).*

Assumption X is standard. Assumption K allows for a wide range of smoothing parameters $h \to 0$ in $\left[\underline{h}, \overline{h}\right]$. When $\nu = \infty$, Assumption K is standard. When $2 < \nu < \infty$ we compare our assumptions on the lower bound for the bandwidth with the one made in Dony and Mason (2008) since this paper also provides uniform in bandwidth type results. The order of our lower bound $\underline{h}$ is $\left(\log n / n^{1-2/\nu}\right)^{1/d}$ while the order of the corresponding $\underline{h}$ in Dony and Mason (2008, Theorem 2) is $\left(\log n / n^{1-2/\nu}\right)^{1/d} \log^{-2/\nu} n$. Our order of $\underline{h}$ is then slightly larger than the one of Dony and Mason (2008) in their Theorem 2 but only differs to a logarithmic term, thus Assumption K is fairly general. The differentiability and local convexity properties assumed for $t \mapsto R(t|x)$ are standard and can be checked provided the conditional density $f(\cdot|\cdot)$ of Y given X is smooth enough. Indeed

$$R(t|x) = \int \rho(y-t) f(y|x) dy,$$

under the conditions of the Lebesgue Dominated Convergence Theorem, and the pointwise differentiability of $\rho(\cdot)$ will ensure that $t \mapsto R(t|x)$ has a first order derivative

$$R^{(1)}(t|x) = \int \rho'(y-t)f(y|x)dy.$$

Assumption R-(iii) follows from the fact that $m(x)$ is the unique minimum of $t \mapsto R(t|x)$. A celebrated example that satisfies Assumption R-(iii) is the case where $\rho(t) = |t|$ so that $m(x)$ is a conditional quantile function. In this case

$$R^{(1)}(t|x) = 2F(t|x) - 1,$$

where $t \mapsto F(t|x)$ is the cumulative distribution of Y given X and $R^{(2)}(t|x)$ exists provided the p.d.f. $f(t|x)$ is continuous. Hence assuming existence and continuity of $t \mapsto R^{(2)}(t|x)$ is a weak assumption. Assumption R also includes some lower bound conditions that involves some functions $\psi(\cdot)$ and $\Psi(\cdot)$ in R-(i), (ii). These are used to show that the LP estimator bias term studied in Theorem 1 below asymptotically vanishes and to obtain the consistency of $\widehat{\mathbf{b}}(h,x)$. Any other conditions that ensure these two requirements can be used instead. A more important group of differentiability conditions for $\rho(\cdot)$ are gathered in Assumption L. Assumption L-(i) allows for a wide class of loss functions $\rho(\cdot)$ that can have kinks including the absolute value $|\cdot|$ and the Huber's loss function

$$\rho_k(t) = \frac{t^2}{2}\mathbb{I}(|t| \leq k) + \left(k|t| - \frac{k^2}{2}\right)\mathbb{I}(|t| > k),$$

$k > 0$. Assumption L-(ii) is a typical condition for the derivative $r(\cdot)$ of such loss functions $\rho(\cdot)$. Proposition 2.1 in Hong (2003) ensures that a condition as L-(ii) holds provided that $y \mapsto f(y|x)$ is continuously differentiable and $r(\cdot)$ is continuously differentiable over each interval (a_j, a_{j+1}) where the a_j's are the finite discontinuity points of $r(\cdot)$. Such requirements holds for $\rho(t) = |t|$ and the Huber's loss function. The fact that the $r_j(\cdot)$'s satisfy

$$\mathbb{E}\left[(r_j(Y-u-t) - r_j(Y-u))^2 \,|X=x\right] \leq C|t|\,, \quad j = 1, 2,$$

for small values of t is a characteristic feature of such piecewise differentiable loss functions and the upper bound $C|t|$ cannot be improved by $C|t|^2$, which would hold for the smoother function

$\rho(t) = t^2$. Assumption L-(iii) is a standard condition that could be used to obtain asymptotic normality of $\sqrt{nh^d}h^{|\mathbf{v}|}\left(\widehat{b}_{\mathbf{v}}(h,x) - b_{\mathbf{v}}(x)\right)$. The part of L-(iii) that deals with $r(\cdot)$ obviously holds for $\rho(t) = |t|$ in which case ν can be taken as large as wanted. The part that deals with $\rho(\cdot)$ is used to obtain consistency of $\widehat{\mathbf{b}}(h,x)$.

3.3 Bias study and Bahadur representation

Observe that the conditional expectation of $\mathcal{L}_n(\mathbf{b};h,x)$ given the X_i's is always finite, and that its mean is $\mathcal{L}(\mathbf{b};h,x) = \mathbb{E}\left[\mathcal{L}_n(\mathbf{b};h,x)\right]$. Classical parametric estimation theory as detailed in van der Vaart (1998) suggests that $\widehat{\mathbf{b}}(h,x)$ should be viewed as an estimator of

$$\mathbf{b}^*(h,x) = \operatorname*{Arg\,Min}_{\mathbf{b}\in\mathbb{R}^P} \mathcal{L}(\mathbf{b};h,x) \ , \tag{3.3.1}$$

instead of $\mathbf{b}(x)$. Hence the quantity $\mathbf{b}^*(h,x) - \mathbf{b}(x)$ can be viewed as a bias term. The study of the bias term $\mathbf{b}^*(h,x) - \mathbf{b}(x)$ can be done considering the first order condition (F.O.C.)

$$\frac{\partial}{\partial \mathbf{b}^\top}\mathcal{L}(\mathbf{b}^*(h,x);h,x) = 0,$$

and the Implicit Functions Theorem. In order to state our first result we define the standardization matrix

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

using the lexicographic order for the $h^{|\mathbf{v}|}$'s as for $\mathbf{b}$. We now state our first result on the bias term.

Theorem 1 *Assume that $\lfloor s \rfloor \le p$. Under Assumptions and K, L, M, R and X*

$$\sup_{(h,x)\in\left[\underline{h},\overline{h}\right]\times\mathcal{X}_0}\left\|\frac{\mathbf{H}(\mathbf{b}^*(h,x)-\mathbf{b}(x))}{h^s}\right\| \le CL,$$

for some $C > 0$.

It follows that

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

uniformly provided $|\mathbf{v}| \le \lfloor s \rfloor$. It is worth noting that the the order p of the LP M-estimator does not influence the bias order $h^{s-|\mathbf{v}|}$. This bias order is better than the bias order $h^{p-|\mathbf{v}|}$, $|\mathbf{v}| \le p$, that would be achieved by suboptimal LP M-estimators of lower order $p < \lfloor s \rfloor$.

Furthermore Theorem 1 also gives the order of the coefficients $b_{\mathbf{v}}^{*}(h,x)$ with $|\mathbf{v}| > \lfloor s \rfloor$ which correspond to partial derivatives that may not exist. Indeed Theorem 1 and definition of $\mathbf{H}$ imply that

$$b_{\mathbf{v}}^{*}(h,x) = O\left(h^{s-|\mathbf{v}|}\right) \quad \text{for } |\mathbf{v}| \ge s \tag{3.3.2}$$

uniformly in $(h,x) \in \left[\underline{h}, \overline{h}\right] \times \mathcal{X}_0$. See also Loader (1999, Theorem 4.2) which gives a less precise $b_{\mathbf{v}}^{*}(h,x) = o\left(h^{-|\mathbf{v}|}\right)$. Hence the higher order polynomial coefficients $b_{\mathbf{v}}^{*}(h,x)$, $|\mathbf{v}| > s$, may diverge when $h > 0$. That this may be indeed the case can be seen on a simple regression example. Consider a loss function $\rho(\cdot)$ and a M-function $m(\cdot)$ satisfying respectively Assumption L and (3.1.1) so that

$$\operatorname*{Arg\,Min}_{t\in\mathbb{R}} \mathbb{E}\left[\rho(Y-t)|X=x\right] = m(x), \quad m(x) = \operatorname{sgn}(x)|x|^{1/2}, \tag{3.3.3}$$

where the $\mathcal{U}([-1,1])$ random variable X and ε are independent and where $\operatorname{sgn}(x) = \mathbb{I}(X \ge 0) - \mathbb{I}(x < 0)$. In this example, $m(\cdot)$ is at best in an Hölder class $\mathcal{C}(L, 1/2)$ since, for L large enough,

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

an inequality that cannot be improved by increasing the exponent $1/2$ as seen by taking $x = 0$ and $x' \to 0$. The next Proposition uses this example to show that the rate given in (3.3.2) is sharp due to the behavior of $m(\cdot)$ at $x = 0$.

Proposition 1 *Suppose that (X,Y) satisfies the model (3.3.3) and let the vector $\mathbf{b}^{*}(h,x) = \left(b_0^{*}(h,x), b_1^{*}(h,x)\right)^{\top}$ from (3.3.1) be given by a LP procedure of order 1. Then under Assumptions K, L, M, R, X, and $\int zK(z)dz = 0$, $b_0^{*}(h,0) = m(0) + O(h^{1/2})$ and $b_1^{*}(h,0)$ diverges with the* *exact rate* $h^{-1/2}$,

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

We now introduce some additional notations specific to Theorem 2. Finding the Bahadur expansion of $\mathbf{H}(\widehat{\mathbf{b}}(h,x) - \mathbf{b}^*(h,x))$ builds on

$$\begin{aligned} m^*(x';h,x) &= \mathbf{U}(x'-x)^\top \mathbf{b}^*(h,x), &(3.3.4)\\ \mathbf{S}(X_i,Y_i;h,x) &= K\left(\frac{X_i-x}{h}\right) r\left(Y_i - m^*(X_i;h,x)\right)\mathbf{H}^{-1}\mathbf{U}\left(X_i-x\right), &(3.3.5)\\ \mathbf{J}(X_i;h,x) &= K\left(\frac{X_i-x}{h}\right) R^{(2)}\left(m^*(X_i;h,x)|X_i\right) \\ &\quad \times \mathbf{H}^{-1}\mathbf{U}(X_i-x)\mathbf{U}(X_i-x)^\top\mathbf{H}^{-1}, &(3.3.6) \end{aligned}$$

abbreviating these variables in $\mathbf{S}_i(h,x)$ and $\mathbf{J}_i(h,x)$, and where $R^{(2)}(\cdot|\cdot)$ is as in Assumption R-(ii). This part builds on a local study of $\mathcal{L}_n(\mathbf{b};h,x)$ based on the change of variables

$$\beta = (nh^d)^{1/2}\mathbf{H}(\mathbf{b} - \mathbf{b}^*(h,x)) \quad \text{so that} \quad \mathbf{b} = \mathbf{b}^*(h,x) + \frac{\mathbf{H}^{-1}\beta}{(nh^d)^{1/2}}.$$

Hence

$$\widehat{\beta}_n(h,x) = (nh^d)^{1/2}\mathbf{H}\left(\widehat{\mathbf{b}}(h,x) - \mathbf{b}^*(h,x)\right),$$

achieves the minimum over $\mathbb{R}^P$ of

$$\mathbb{L}_{1n}(\beta;h,x) = nh^d\left(\mathcal{L}_n\left(\mathbf{b}^*(h,x) + \frac{1}{(nh^d)^{1/2}}\mathbf{H}^{-1}\beta;h,x\right) - \mathcal{L}_n\left(\mathbf{b}^*(h,x);h,x\right)\right)$$

It is known that an approximation of $\widehat{\beta}_n(h,x)$ is given by

$$\beta_n(h,x) = -\left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(h,x)\right)^{-1}\frac{1}{(nh^d)^{1/2}}\sum_{i=1}^n \mathbf{S}_i(h,x), \qquad (3.3.7)$$

see for example van der Vaart (1998). It will be shown that $\beta_n(h,x)$ is well defined uniformly in (h,x) in $\left[\underline{h},\overline{h}\right] \times \mathcal{X}_0$ with probability which can be arbitrarily large.

The Bahadur representation studies the error in the approximation of $\widehat{\beta}_n(h,x)$ by $\beta_n(h,x)$. Namely, the purpose of Theorem 2 is to give an uniform convergence rate of the quantity

$$\begin{aligned}\mathbf{E}_n(h,x) &= \widehat{\beta}_n(h,x)-\beta_n(h,x)\\ &= (nh^d)^{1/2}\mathbf{H}\left(\widehat{\mathbf{b}}(h,x)-\mathbf{b}^*(h,x)\right)\\ &\quad+\left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(h,x)\right)^{-1}\frac{1}{(nh^d)^{1/2}}\sum_{i=1}^n \mathbf{S}_i(h,x). \qquad (3.3.8)\end{aligned}$$

Observe that since $\widehat{\beta}_n(h,x)$ achieves the minimum over $\mathbb{R}^P$ of $\beta \mapsto \mathbb{L}_{1n}(\beta;h,x)$ for all (h,x) in $\left[\underline{h},\overline{h}\right]\times\mathcal{X}_0$, then $\mathbf{E}_n(h,x)$ can be viewed as the minimizer w.r.t. ε over $\mathbb{R}^P$ of

$$\mathbb{L}_{1n}(\beta_n(h,x)+\varepsilon;h,x)-\mathbb{L}_{1n}(\beta_n(h,x);h,x).$$

Define

$$\mathbb{L}_n(\beta,\varepsilon;h,x)=\mathbb{L}_{1n}(\beta+\varepsilon;h,x)-\mathbb{L}_{1n}(\beta;h,x).$$

Now observe that one can write

$$\begin{aligned}\mathbb{L}_n(\beta,\varepsilon;h,x) &= \left(\frac{1}{(nh^d)^{1/2}}\sum_{i=1}^n \mathbf{S}_i(h,x)\right)^{\top}\varepsilon+\frac{1}{2nh^d}\varepsilon^{\top}\left(\sum_{i=1}^n \mathbf{J}_i(h,x)\right)(\varepsilon+2\beta)\\ &\quad+\mathbb{R}_n(\beta,\varepsilon;h,x)\\ &= \mathbb{L}_n^0(\beta,\varepsilon;h,x)+\mathbb{R}_n(\beta,\varepsilon;h,x). \qquad (3.3.9)\end{aligned}$$

As explained in Fan and Gijbels (1996), the behavior of $\mathbf{E}_n(h,x)$ is closely related to the one of $\mathbb{R}_n(\beta,\varepsilon;h,x)$. Indeed, the proof of the next Theorem involves the study of the uniform behavior of $\mathbb{R}_n(\beta,\varepsilon;h,x)$, using a maximal inequality under bracketing entropy conditions instead of the classical Bernstein inequality used in the pointwise case e.g. in Hong (2003). The next theorem gives the order of $\mathbf{E}_n(h,x)$ uniformly with respect to h and x.

Theorem 2 *Under Assumptions K, L, M, R and X and*

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

The following Corollary is a combination of Theorems 1 and 2 with a optimal choice of the order of the bandwidth.

Corollary 1 *Assume that $m(\cdot)$ is in $\mathcal{C}(L,s)$ for some $\lfloor s \rfloor \le p$. Suppose that Assumptions K, L, M, R and X hold. Then for all partial derivative order $\mathbf{v}$ with $|\mathbf{v}| \le \lfloor s \rfloor$*

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

ii. $\sup_{x \in \mathcal{X}_0} \left|\widehat{b}_{\mathbf{v}}(h,x) - b_{\mathbf{v}}(x)\right| = O_{\mathbb{P}}\left(\frac{\log n}{n}\right)^{\frac{s-|\mathbf{v}|}{2s+d}}$ *if h is a.p. to $\left(\frac{\log n}{n}\right)^{\frac{1}{2s+d}}$.*

Since the $\widehat{b}_{\mathbf{v}}(h,x)$'s are estimators of the $b_{\mathbf{v}}(x)$, 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 the use of data-driven bandwidths.

Proposition 2 *Consider a random bandwidth $\widehat{h}$ such that $\widehat{h} = O_{\mathbb{P}}(h)$ and $1/\widehat{h} = O_{\mathbb{P}}(1/h)$ where h is a deterministic sequence satisfying $h = o(1)$ and $\limsup_{n\to\infty} \log(n)/(nh^d) < \infty$. Suppose that Assumptions K, L, M, R and X hold and that $m(\cdot)$ is in $\mathcal{C}(L,s)$. Then for any $\mathbf{v}$ with $|\mathbf{v}| \le \lfloor s \rfloor$,*

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

In particular, if the exact order of $\widehat{h}$ is $(\log(n)/n)^{1/(2s+d)}$ in probability,

$$\sup_{x \in \mathcal{X}_0} \left|\widehat{b}_{\mathbf{v}}(\widehat{h},x) - b_{\mathbf{v}}(x)\right|$$

has the optimal order $(\log(n)/n)^{(s-|\mathbf{v}|)/(2s+d)}$ of Corollary 1-(ii). It is likely that an L_k version of Proposition 2 holds but it is slightly longer to prove.

3.4 Appendix A: Proofs of main results

Appendix A groups the proofs of Theorems 1 and 2, Propositions 1 and 2 and Corollary 1. The proofs of intermediary results are gathered in Appendix B.

Let $\mathcal{B}(z,t)$ be the closed ball with center z and radius t. Depending on the context $\mathcal{B}(z,t)$ will be a subset of $\mathbb{R}^d$, $\mathbb{R}^P$ or $\mathbb{R}^{P+d+1}$ without further indications. 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, $\|\mathbf{A}\| = \sup_{\mathbf{u}\in\mathcal{B}(0,1)} \|\mathbf{A}\mathbf{u}\| = \sup_{\mathbf{u}\in\mathcal{B}(0,1)} \left|\mathbf{u}^\top \mathbf{A}\mathbf{u}\right|$ 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. $\|f(\cdot)\|_\infty = \sup_{x\in\mathbb{R}^d} f(x)$. We also use $K_h(z) = K(z/h)$. 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.

In the proofs of our main and intermediary results, we use the abbreviation $\theta = (h,x)$ for notational convenience. Define

$$\Theta^0 = \left[0, \overline{h}\right] \times \mathcal{X},\ \Theta = \left[\underline{h}, \overline{h}\right] \times \mathcal{X}_0,$$

where $\mathcal{X}_0$ is as in Assumption X. 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,z) in $\mathcal{X}_0 \times \left[\underline{h}, \overline{h}\right] \times \mathcal{K}$, $x + hz$ is in $\mathcal{X}$ under Assumption K provided $\overline{h}$ is small enough.

The following Lemma is used in the proofs of both Theorems 1 and 2. Its proof is given in Appendix B with the proofs of other intermediary results.

Lemma A.1 *Under Assumptions K, L, M, R and X we have for $\overline{h}$ 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* $\int R^{(1)}(\mathbf{U}(z)^\top \mathbf{B}^*(\theta)|x+hz) f(x+hz) K(z) \mathbf{U}(z)^\top dz = 0$.

iii. $\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\|\mathbf{H}\left(\mathbf{b}^*(\theta) - (m(x), 0, \ldots, 0)^\top\right)\right\| = 0$.

iv. The map $\theta \mapsto m^*(x';\theta) = \mathbf{U}(x'-x)^\top \mathbf{b}^*(\theta)$ *is continuously differentiable over Θ, and for all x' in $\mathcal{X}$ and (θ_1,θ_2) in Θ^2,*

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

v. $\lim_{\overline{h}\to 0}\sup_{\theta\in\Theta}\sup_{x'\in\mathcal{X}} K_h(x'-x)\left|m^*(x';\theta)-m(x)\right| = 0.$

A.1 Proof of Theorem 1

Recall that $\mathbf{b}(x)$ is the vector of derivatives of $m(\cdot)$ at x up to order $\lfloor s\rfloor$ completed with 0's to achieve dimension P. Observe that since $m(\cdot)$ is in $\mathcal{C}(L,s)$ by Assumption M, we have for some $C>0$ that

$$m(x+hz)-\mathbf{U}(z)^\top\mathbf{H}\mathbf{b}(x) = V_m(z,\theta), \quad \text{with} \quad \sup_{\theta\in\Theta}\sup_{z\in\mathcal{K}}\left|\frac{V_m(z,\theta)}{h^s}\right| \leq CL. \tag{A.1}$$

Lemma A.1-(ii) yields that

$$\int R^{(1)}(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)|x+hz)f(x+hz)K(z)\mathbf{U}(z)dz = 0. \tag{A.2}$$

Assumption R-(iii) yields that the map $t\in\mathbb{R}\mapsto R(t|x)$ is continuously differentiable for all x in $\mathcal{X}$ and Assumption R-(i) yields that it achieves its unique minimum at $m(x)$ w.r.t. t in $\mathbb{R}$, so that $R^{(1)}(m(x)|x)=0$. Then for $\overline{h}$ small enough we have $R^{(1)}(m(x+hz)|x+hz)=0$ for all θ in Θ and

$$\int R^{(1)}(m(x+hz)|x+hz)f(x+hz)K(z)\mathbf{U}(z)dz = 0.$$

It then follows from (A.2) that

$$\int \left(R^{(1)}(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)|x+hz) - R^{(1)}(m(x+hz)|x+hz)\right) f(x+hz)K(z)\mathbf{U}(z)^\top dz = 0. \tag{A.3}$$

Define

$$I(z,\theta) = \int_0^1 R^{(2)}\left(m(x+hz)+t\left(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)-m(x+hz)\right)|x+hz\right)dt, \tag{A.4}$$

so that

$$R^{(1)}(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)|x+hz) - R^{(1)}(m(x+hz)|x+hz) = \left(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)-m(x+hz)\right)I(\theta,z).$$

This and (A.3) then yield that

$$\int \left(\mathbf{U}(z)^\top\mathbf{B}^*(\theta)-m(x+hz)\right)I(\theta,z)f(x+hz)K(z)\mathbf{U}(z)dz = 0. \tag{A.5}$$

Continuity of $(t,x) \in \mathbb{R} \times \mathcal{X} \mapsto R^{(2)}(t|x)$ under Assumption R-(iii), continuity of $x \in \mathcal{X} \mapsto m(x)$ under Assumption M, Lemma A.1-(iii) and the dominated Convergence Theorem yield that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta} \sup_{z\in\mathcal{K}} \left| I(\theta,z) - R^{(2)}(m(x)|x) \right| = 0,$$

so that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta} \left\| \int \mathbf{U}(z)\mathbf{U}(z)^\top K(z) \left(f(x+hz)I(\theta,z) - f(x)R^{(2)}(m(x)|x) \right) dz \right\| = 0. \tag{A.6}$$

Now since $\inf_{x\in\mathcal{X}} f(x)R^{(2)}(m(x)|x) \geq C > 0$ under Assumptions R-(iii) and X we have

$$\inf_{x\in\mathcal{X}} f(x)R^{(2)}(m(x)|x) \int \mathbf{U}(z)\mathbf{U}(z)^\top K(z)dz \succ C \int \mathbf{U}(z)\mathbf{U}(z)^\top K(z)dz,$$

where the last symmetric matrix is positive definite under Assumption K on the kernel. It then follows form this and (A.6) that the symmetric matrix

$$\int \mathbf{U}(z)\mathbf{U}(z)^\top K(z) f(x+hz) I(\theta,z)dz,$$

is positive definite for $\overline{h}$ small enough and so has an inverse. Now observe that (A.1) and (A.5) yield that

$$\left(\int \mathbf{U}(z)\mathbf{U}(z)^\top f(x+hz)K(z)I(\theta,z)dz \right) \mathbf{H}\left(\mathbf{b}^*(\theta) - \mathbf{b}(x)\right) = \int f(x+hz)K(z)V_m(z,\theta)\mathbf{U}(z)dz.$$

Then (A.1) and (A.6) yield that

$$\begin{aligned} \sup_{\theta\in\Theta} \left\| \frac{\mathbf{H}\left(\mathbf{b}^*(\theta) - \mathbf{b}(x)\right)}{h^s} \right\| &\leq \sup_{\theta\in\Theta} \left\| \int \mathbf{U}(z)\mathbf{U}(z)^\top f(x+hz)K(z)I(\theta,z)dz \right\|^{-1} \\ &\quad \times \sup_{\theta\in\Theta} \left\| \int f(x+hz)K(z)\frac{V_m(z,\theta)}{h^s}\mathbf{U}(z)dz \right\| \\ &\leq C \left\| \int_{\mathcal{B}(0,1)} \mathbf{U}(z)\mathbf{U}(z)^\top dz \right\|^{-1} \sup_{\theta\in\Theta} \sup_{z\in\mathcal{K}} \left| \frac{V_m(z,\theta)}{h^s} \right| \leq CL, \end{aligned} \tag{A.7}$$

which ends the proof. □

A.2 Proof of Proposition 1

Recall that $\mathbf{U}(z) = (1, z)^\top$ and that (A.5) yields that at $x = 0$,

$$\int \left(\mathbf{U}(z)^\top \mathbf{B}^*(h, 0) - m(hz)\right) I(h, 0, z) f(hz) K(z) \mathbf{U}(z) dz = 0,$$

where $I(\theta, z)$ is defined in (A.4). Then (A.6) and assumptions K and X yield that

$$\begin{pmatrix} b_0^*(h,0) \\ hb_1^*(h,0) \end{pmatrix} = \left(\int \mathbf{U}(z)\mathbf{U}^\top(z) K(z) dz\right)^{-1} \int m(hz)\mathbf{U}(z)K(z)dz(1 + o(1)).$$

Recall that $m(z) = \text{sgn}(z)|z|^{1/2}$, z in $[-1, 1]$ so that the equation above gives

$$\begin{pmatrix} b_0(h,0) \\ hb_1(h,0) \end{pmatrix} = (1 + o(1))\, h^{1/2} \begin{bmatrix} \int m(z)K(z)dz \\ \frac{\int |z|^{3/2}K(z)dz}{\int z^2 K(z)dz} \end{bmatrix}. \square$$

A.3 Proof of Theorem 2

Let

$$\begin{aligned} \mathbf{R}^0(\beta, \varepsilon; \theta) &= \mathbf{R}^0(X, Y; \beta, \varepsilon, \theta), \ \ \mathbf{R}^1(\beta, \varepsilon; \theta) = \mathbf{R}^1(X, Y; \beta, \varepsilon, \theta) \ \text{ and} \\ \mathbf{R}^2(\beta, \varepsilon; \theta) &= \mathbf{R}^2(X; \beta, \varepsilon, \theta), \end{aligned}$$

be such that

$$\begin{aligned} \mathbf{R}^0(\beta, \varepsilon; \theta) &= \Bigg\{\rho\left(Y - m^*(X;\theta) - \frac{\mathbf{U}\left(\frac{X-x}{h}\right)^\top \beta}{(nh^d)^{1/2}} - \frac{\mathbf{U}\left(\frac{X-x}{h}\right)^\top \varepsilon}{(nh^d)^{1/2}}\right) && \text{(A.8)} \\ &\quad -\rho\left(Y - m^*(X;\theta) - \frac{\mathbf{U}\left(\frac{X-x}{h}\right)^\top \beta}{(nh^d)^{1/2}}\right)\Bigg\} K_h(X - x) - \frac{\mathbf{S}(\theta)^\top}{(nh^d)^{1/2}}\varepsilon, \\ \mathbf{R}^1(\beta, \varepsilon; \theta) &= \mathbf{R}^0(\beta, \varepsilon; \theta) - \mathbb{E}\left[\mathbf{R}^0(\beta, \varepsilon; \theta)\,|X\right], && \text{(A.9)} \\ \mathbf{R}^2(\beta, \varepsilon; \theta) &= \mathbb{E}\left[\mathbf{R}^0(\beta, \varepsilon; \theta)\,|X\right] - \frac{1}{2nh^d}\varepsilon^\top \mathbf{J}(\theta)(\varepsilon + 2\beta). && \text{(A.10)} \end{aligned}$$

Let $\mathbb{R}_n^0(\beta, \varepsilon; \theta) = \sum_{i=1}^n \mathbf{R}^0(X_i, Y_i; \beta, \varepsilon, \theta)$, $\mathbb{R}_n^1(\beta, \varepsilon; \theta) = \sum_{i=1}^n \mathbf{R}^1(X_i, Y_i; \beta, \varepsilon, \theta)$ and $\mathbb{R}_n^2(\beta, \varepsilon; \theta) = \sum_{i=1}^n \mathbf{R}^2(X_i; \beta, \varepsilon, \theta)$, so that $\mathbb{R}_n(\beta, \varepsilon; \theta) = \mathbb{R}_n^1(\beta, \varepsilon; \theta) + \mathbb{R}_n^2(\beta, \varepsilon; \theta)$ where $\mathbb{R}_n(\beta, \varepsilon; \theta)$ is as in (3.3.9).

In order to prove the Theorem, we first need an uniform consistency result.

Proposition A.1 *Under Assumptions K, L, R and X,*

$$\sup_{\theta\in\Theta}\left\|\mathbf{H}\left(\widehat{\mathbf{b}}(\theta)-\mathbf{b}^*(\theta)\right)\right\| = o_{\mathbb{P}}(1).$$

Proposition A.2 deals with $\mathbb{R}_n^1(\beta,\varepsilon;\theta)$ and Proposition A.3 studies $\mathbb{R}_n^2(\beta,\varepsilon;\theta)$.

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

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

Proposition A.3 *Consider two real numbers $t_\beta, t_\varepsilon > 0$ which may depend upon on n with $t_\beta \geq 1$, $t_\varepsilon \geq 1/n$, $t_\varepsilon + t_\beta = o\left(n\underline{h}^d\right)^{1/2}$ and $t_\varepsilon(t_\beta+t_\varepsilon)^2 = o\left(n^{3/2}\underline{h}^d\right)$. Then under Assumptions K, L, R, X and for n large enough,*

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

We also need the two following Lemmas to control the term $\beta_n(\theta)$ defined in (3.3.7).

Lemma A.2 *let $\underline{\lambda}_n(\theta)$ be the smallest eigenvalue of the positive symmetric matrix*

$$\frac{1}{nh^d}\sum_{i=1}^{n}\mathbf{J}_i(\theta).$$

Under Assumptions K, L, R and X, there is a $\underline{\lambda} > 0$ such that, $\inf_{\theta\in\Theta}\underline{\lambda}_n(\theta) \geq \underline{\lambda} + o_{\mathbb{P}}(1)$.

Lemma A.3 *Under Assumptions K, L, R and X*

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

Recall that (3.3.7) yields that

$$\beta_n(\theta) = -\left(\frac{1}{nh^d}\sum_{i=1}^n \mathbf{J}_i(\theta)\right)^{-1} \frac{1}{(nh^d)^{1/2}}\sum_{i=1}^n \mathbf{S}_i(\theta).$$

Then Lemma A.2 provides existence of $\beta_n(\theta)$ by proving that the matrix $\sum_{i=1}^n \mathbf{J}_i(\theta)/nh^d$ is definite positive and thus has an inverse with a probability that can be arbitrarily large and Lemma A.2 together with A.3 yield that

$$\sup_{\theta\in\Theta} \|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2} n\right). \tag{A.11}$$

We now return to the proof of Theorem 2. Recall that by (3.3.8),

$$\frac{\mathbf{E}_n(\theta)}{(nh^d)^{1/2}} = -\frac{\beta_n(\theta)}{(nh^d)^{1/2}} + \mathbf{H}\left(\widehat{\mathbf{b}}(\theta) - \mathbf{b}^*(\theta)\right).$$

Then (A.11) and Proposition A.1 yield that

$$\sup_{\theta\in\Theta} \frac{\|\mathbf{E}_n(\theta)\|}{(nh^d)^{1/2}} = O_{\mathbb{P}}\left(\frac{\log n}{n\underline{h}^d}\right)^{1/2} + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1),$$

under Assumption K which yields that $\log n/(n\underline{h}^d) = o(1)$. Then there is a sequence $\eta_n = o(1)$ such that

$$\sup_{\theta\in\Theta} \|\mathbf{E}_n(\theta)\| = o_{\mathbb{P}}\left(\eta_n \left(n\underline{h}^d\right)^{1/2}\right). \tag{A.12}$$

In what follows

$$t_{n,j} = 2^j \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{1/4}}, \quad j \geq J, \tag{A.13}$$

where j and J are integer numbers. In the sequel, $t_{n,j}$ will play the role of t_ε whereas t_β will be chosen such that $t_\beta \asymp \log^{1/2} n$. Let $\overline{J}$ be the unique integer number such that

$$2^{\overline{J}+1} \leq \eta_n \frac{\left(n\underline{h}^d\right)^{3/4}}{\log^{3/4} n} \text{ and } 2^{\overline{J}+2} > \eta_n \frac{\left(n\underline{h}^d\right)^{3/4}}{\log^{3/4} n}. \tag{A.14}$$

This and Assumption K then yield that

$$\sup_{j\geq J} \frac{\left(t_\beta + t_{n,j}\right)^{1/2}}{t_{n,j}} \leq \sup_{j\geq J} \frac{\left(n\underline{h}^d\right)^{1/4}}{2^j \log^{3/4} n}\left(C\log^{1/4} n + 2^{j/2}\frac{\log^{3/8} n}{\left(n\underline{h}^d\right)^{1/8}}\right) \leq \sup_{j\geq J} \frac{C}{2^{J/2}}\frac{\left(n\underline{h}^d\right)^{1/4}}{\log^{1/2} n}$$

$$
\begin{aligned}
&= O\left(\frac{\left(n\underline{h}^d\right)^{1/4}}{\log^{1/2} n}\right), \\
\sup_{J\le j\le \overline{J}+1} \frac{t_{n,j}(t_\beta+t_{n,j})^2}{n^{3/2}\underline{h}^d} &\le \sup_{J\le j\le \overline{J}+1} 2^j \frac{\log^{7/4} n}{n^{1/2}\left(n\underline{h}^d\right)^{5/4}} + 2^{3j}\frac{\log^{9/4} n}{n^{1/2}\left(n\underline{h}^d\right)^{7/4}} \\
&= 2^{\overline{J}+1}\frac{\log^{7/4} n}{n^{1/2}\left(n\underline{h}^d\right)^{5/4}}\left(1+2^{2(\overline{J}+1)}\frac{\log^{1/2} n}{\left(n\underline{h}^d\right)^{1/2}}\right) \\
&\le \eta_n \frac{\log n}{n^{1/2}\left(n\underline{h}^d\right)^{1/2}} + \eta_n^3 \underline{h}^{d/2} = o(1), \\
\sup_{J\le j\le \overline{J}+1} \frac{t_\beta+t_{n,j}}{\left(n\underline{h}^d\right)^{1/2}} &= \sup_{J\le j\le \overline{J}+1} 2^j \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} + C\frac{\log^{1/2} n}{\left(n\underline{h}^d\right)^{1/2}} = 2^{\overline{J}+1}\frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} + C\frac{\log^{1/2} n}{\left(n\underline{h}^d\right)^{1/2}} \\
&\le C\left(\eta_n + \frac{\log^{1/2} n}{\left(n\underline{h}^d\right)^{1/2}}\right) = o(1).
\end{aligned}
$$

Hence these choices of t_β and $t_{n,j}$ satisfy the conditions of Propositions A.2 and A.3 uniformly with respect to $J \le j \le \overline{J}+1$.

Let η be an arbitrary positive real number. Define for some $C_\eta > 0$, the set

$$
B_n = \left\{\sup_{\theta\in\Theta} \|\mathbf{E}_n(\theta)\| \ge \eta_n \left(n\underline{h}^d\right)^{1/2}\right\} \bigcup \left\{\sup_{\theta\in\Theta} \|\beta_n(\theta)\| \ge C_\eta \log^{1/2} n\right\} \bigcup \left\{\inf_{\theta\in\Theta} \lambda_n(\theta) \le \frac{\underline{\lambda}}{2}\right\}, \tag{A.15}
$$

where $\lambda_n(\theta)$ and $\underline{\lambda}$ are as in Lemma A.2. For any $\eta > 0$, C_η can be chosen to ensure that

$$
\begin{aligned}
\mathbb{P}(B_n) \;\le\; & \mathbb{P}\left\{\sup_{\theta\in\Theta} \|\mathbf{E}_n(\theta)\| \ge \eta_n \left(n\underline{h}^d\right)^{1/2}\right\} + \mathbb{P}\left\{\sup_{\theta\in\Theta} \|\beta_n(\theta)\| \ge C_\eta \log^{1/2} n\right\} \\
& + \mathbb{P}\left\{\inf_{\theta\in\Theta} \lambda_n(\theta) \le \frac{\underline{\lambda}}{2}\right\} \le \frac{\eta}{2},
\end{aligned}
$$

for n large enough by (A.12), (A.11) and Lemma A.2.

Now recall that $\mathbf{E}_n(\theta)$ achieves the minimum other $\mathbb{R}^P$ of $\varepsilon \mapsto \mathbb{L}_n(\beta_n(\theta), \varepsilon; \theta)$. Consider a subset A of $\mathbb{R}^P$ such that 0 does not belong to A. Then since $\mathbb{L}_n(\beta, 0, \theta) = 0$, for all β in $\mathbb{R}^P$ and θ in Θ, we have

$$
\{\mathbf{E}_n(\theta) \in A\} \subset \left\{\inf_{\varepsilon\in A} \mathbb{L}_n(\beta_n(\theta), \varepsilon; \theta) \le \mathbb{L}_n(\beta_n(\theta), 0; \theta)\right\} = \left\{\inf_{\varepsilon\in A} \mathbb{L}_n(\beta_n(\theta), \varepsilon; \theta) \le 0\right\}.
$$

This and the fact that $\{\sup_{\theta\in\Theta}\|\mathbf{E}_n(\theta)\| \geq t_{n,J}\} = \bigcup_{\theta\in\Theta}\{\|\mathbf{E}_n(\theta)\| \geq t_{n,J}\}$, gives that

$$\left\{\sup_{\theta\in\Theta}\|\mathbf{E}_n(\theta)\| \geq t_{n,J}\right\} \subset \underbrace{\left(\left\{\inf_{\theta\in\Theta}\inf_{t_{n,J}\leq\|\varepsilon\|\leq\eta_n\left(n\underline{h}^d\right)^{1/2}}\mathbb{L}_n(\beta_n(\theta),\varepsilon;\theta)\leq 0\right\}\bigcap B_n^c\right)}_{A_n}\bigcup B_n, \quad \text{(A.16)}$$

where B_n^c is the complement of B_n defined in (A.15) and $t_{n,J}$ is as in (A.13). We now study the event A_n. Define the event

$$\widetilde{B}_n = \left\{\sup_{\theta\in\Theta}\|\beta_n(\theta)\| \leq C_\eta \log^{1/2} n\right\} \cap \left\{\inf_{\theta\in\Theta}\lambda_n(\theta) \geq \frac{\lambda}{2}\right\},$$

and the set

$$\mathcal{B}_{n,j} = \left\{\varepsilon \in \mathbb{R}^P,\ t_{n,j} \leq \|\varepsilon\| \leq t_{n,j+1}\right\},$$

so that

$$A_n = \widetilde{B}_n \cap \bigcup_{j=J}^{\overline{J}}\left\{\inf_{\theta\in\Theta}\inf_{\varepsilon\in\mathcal{B}_{n,j}}\mathbb{L}_n(\beta_n(\theta),\varepsilon;\theta)\leq 0\right\}.$$

Recall that $\lambda_n(\theta)$ is the smallest eigenvalue of the matrix $\sum_{i=1}^n \mathbf{J}_i(\theta)$. Then since $\mathbb{L}_n(\beta,\varepsilon;\theta) = \mathbb{L}_n^0(\beta,\varepsilon;\theta) + \mathbb{R}_n(\beta,\varepsilon;\theta)$ by (3.3.9) and $\mathbb{L}_n^0(\beta_n(\theta),\varepsilon;\theta) = \varepsilon^\top\sum_{i=1}^n \mathbf{J}_i(\theta)\varepsilon/2$ we have

$$\begin{aligned}
A_n &\subset \widetilde{B}_n \cap \bigcup_{j=J}^{\overline{J}}\left\{\inf_{\theta\in\Theta}\inf_{\varepsilon\in\mathcal{B}_{n,j}}\mathbb{L}_n^0(\beta_n(\theta),\varepsilon;\theta) \leq \sup_{\theta\in\Theta}\sup_{\varepsilon\in\mathcal{B}_{n,j}}\mathbb{R}_n(\beta_n(\theta),\varepsilon;\theta)\right\} \\
&\subset \widetilde{B}_n \cap \bigcup_{j=J}^{\overline{J}}\left\{\inf_{\theta\in\Theta}\frac{\lambda_n(\theta)}{2}\inf_{\varepsilon\in\mathcal{B}_{n,j}}\|\varepsilon\|^2 \leq \sup_{\theta\in\Theta}\sup_{\varepsilon\in\mathcal{B}_{n,j}}\mathbb{R}_n(\beta_n(\theta),\varepsilon;\theta)\right\} \\
&= \widetilde{B}_n \cap \bigcup_{j=J}^{\overline{J}}\left\{t_{n,j}^2\inf_{\theta\in\Theta}\frac{\lambda_n(\theta)}{2} \leq \sup_{\theta\in\Theta}\sup_{\varepsilon\in\mathcal{B}_{n,j}}\mathbb{R}_n(\beta_n(\theta),\varepsilon;\theta)\right\} \\
&\subset \bigcup_{j=J}^{\overline{J}}\left\{t_{n,j}^2\frac{\lambda}{4} \leq \sup_{\theta\in\Theta}\sup_{\varepsilon\in\mathcal{B}_{n,j}}\sup_{\beta\in\mathcal{B}(0,C_\eta\log^{1/2}n)}\mathbb{R}_n(\beta,\varepsilon;\theta)\right\} = \widetilde{A}_n, \qquad \text{(A.17)}
\end{aligned}$$

where C_η is as in (A.16). We now study the event $\widetilde{A}_n$. For that, observe that Propositions A.2 and A.3 together with the Markov inequality and the fact that $\mathcal{B}_{n,j}$ is a subset of $\mathcal{B}(0, t_{n,j+1})$

yield that

$$\begin{aligned}
\mathbb{P}\left(\widetilde{A}_n\right) &\leq \sum_{j=J}^{\overline{J}} \mathbb{P}\left(t_{n,j}^2 \frac{\lambda}{4} \leq \sup_{\theta\in\Theta} \sup_{\varepsilon\in\mathcal{B}_{n,j}} \sup_{\beta\in\mathcal{B}(0,C_\eta \log^{1/2} n)} \mathbb{R}_n(\beta,\varepsilon;\theta)\right) \\
&\leq \sum_{j=J}^{\overline{J}} t_{n,j}^{-2} \frac{4}{\lambda} \mathbb{E}\left[\sup_{\theta\in\Theta} \sup_{\varepsilon\in\mathcal{B}_{n,j}} \sup_{\beta\in\mathcal{B}(0,C_\eta \log^{1/2} n)} \left|\mathbb{R}_n(\beta,\varepsilon;\theta)\right|\right] \\
&\leq C \sum_{j=J}^{\overline{J}} t_{n,j}^{-2} \left(\frac{t_{n,j+1}(C_\eta \log^{1/2} n + t_{n,j+1})^{1/2}}{\left(n\underline{h}^d\right)^{1/4}} \log^{1/2} n + \frac{t_{n,j+1}(C_\eta \log^{1/2} n + t_{n,j+1})^2}{\left(n\underline{h}^d\right)^{1/2}}\right) \\
&\leq C \sum_{j=J}^{\infty} \left(2^{-j/2} \frac{\log^{1/4} n}{\left(n\underline{h}^d\right)^{1/4}} + 2^{-j} C_\eta^{1/2} + 2^{-2j} \frac{\log^{1/4} n}{\left(n\underline{h}^d\right)^{1/4}} C_\eta^2\right) + C \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} \sum_{j=J}^{\overline{J}} 2^j \\
&\leq C C_\eta^{1/2} 2^{-J} (1+o(1)) + \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} \sum_{j=J}^{\overline{J}} 2^j,
\end{aligned}$$

under Assumption K. Now observe that

$$\frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} \sum_{j=J}^{\overline{J}} 2^j = \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} \left(2^{\overline{J}+1} - 2^J\right) \leq C\eta_n - 2^J \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{3/4}} = o(1),$$

by definition (A.14) of $\overline{J}$ and under Assumption K. Then (A.17) yields that $\limsup_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}\left(\widetilde{A}_n\right) \leq C C_\eta^{1/2} 2^{-J/2}$. (A.16) and (A.16) yield that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta} \|\mathbf{E}_n(\theta)\| \geq t_{n,J}\right) \leq \limsup_{n\to\infty} \mathbb{P}(B_n) + C C_\eta^{1/2} 2^{-J/2} \leq \frac{\eta}{2} + C C_\eta^{1/2} 2^{-J/2}.$$

The definition (A.13) of $t_{n,J}$ and (A.16) then give that for each η there is a J such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta} \|\mathbf{E}_n(\theta)\| \geq 2^J \frac{\log^{3/4} n}{\left(n\underline{h}^d\right)^{1/4}}\right) \leq \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

Hence the Theorem is proved. □

A.4 Proof of Corollary 1

Part (i) follows from Theorems 1 and 2 and the triangular inequality, together with Lemmas A.2 and A.3 which gives $\sup_{\theta\in\Theta}\|\beta_n(\theta)\| = O_{\mathbb{P}}\left(\log^{1/2} n\right)$, so that

$$\left(\int_{\mathcal{X}_0}\|\beta_n(\theta)\|\,dx\right)^{1/k} = O_{\mathbb{P}}(1).$$

We now prove the latter. Lemma A.2 and the Hölder inequality give

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

Since $\mathbb{E}[\mathbf{S}_i(\theta)] = 0$ by Lemma A.1-(ii), the Burkholder inequality gives (see Chow and Teicher (2003), Theorem 1, p. 414)

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

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

A.5 Proof of Proposition 2

Let $\underline{h} = h/C$ and $\overline{h} = Ch$. The condition on h ensures that $\underline{h}$ and $\overline{h}$ satisfy Assumption K for all $C > 1$. Theorems 1 and 2 give, for all $C > 1$,

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\widehat{b}_{\mathbf{v}}(\widehat{h},x) - b_{\mathbf{v}}(x)\right| &= \underline{h}^{-|\mathbf{v}|}O_{\mathbb{P}}\left(\underline{h}^s + \left(\frac{\log n}{n\underline{h}^d}\right)^{1/2}\right)\\ &= h^{-|\mathbf{v}|}O_{\mathbb{P}}\left(h^s + \left(\frac{\log n}{nh^d}\right)^{1/2}\right)\\ &= \widehat{h}_n^{-|\mathbf{v}|}O_{\mathbb{P}}\left(\widehat{h}^s + \left(\frac{\log n}{n\widehat{h}^d}\right)^{1/2}\right).\end{aligned}$$

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

3.5 Appendix B: Proofs of intermediary results

B.1 Proof of Lemma A.1.

Proof of (i). Observe that

$$\begin{aligned}
\mathcal{L}(\mathbf{b}; h, x) &= \frac{1}{h^d}\int \mathbb{E}\left[\rho(Y - \mathbf{U}(X-x)^\top \mathbf{b})|X = u\right] K\left(\frac{u-x}{h}\right) f(u)du \\
&= \int \mathbb{E}\left[\rho(Y - \mathbf{U}(hz)^\top \mathbf{b})|X = x+hz\right] K(z)\, f(x+hz)dz \\
&= \int R\left(\mathbf{U}(z)^\top \mathbf{H}\mathbf{b}|x+hz\right) f(x+hz)K(z)dz,
\end{aligned} \tag{B.1}$$

using the change of variables $u = x + hz$ and $\mathbf{U}(hz) = \mathbf{H}\mathbf{U}(z)$. Consider $\mathbf{m}_x = (m(x), 0, \ldots, 0)$ in $\mathbb{R}^P$, which is such that $\mathbf{U}(z)^\top \mathbf{H}\mathbf{m}_x = m(x)$. Define

$$\mathcal{L}'(\mathbf{b};\theta) = \mathcal{L}(\mathbf{b};\theta) - \int R(m(x+hz)|x+hz)f(x+hz)K(z),$$

and observe that $\mathbf{b}^*(\theta)$ satisfies

$$\mathbf{b}^*(\theta) = \operatorname*{Arg\,Min}_{\mathbf{b}\in\mathbb{R}^P} \mathcal{L}'(\mathbf{b};\theta).$$

Observe that definition of $\mathcal{L}(\mathbf{b};\theta)$ and the fact that $\int K(z)dz = 1$ under Assumption K yield that

$$\mathcal{L}'(\mathbf{b};\theta) = \int \left[R(\mathbf{U}(z)^\top \mathbf{H}\mathbf{b}|x+hz) - R(m(x+hz)|x+hz)\right] f(x+hz)K(z)dz,$$

showing that $\mathcal{L}'(\mathbf{b};\theta)$ is well defined for $h = 0$. Then Assumptions M, R-(iii) and X and the Lebesgue Theorem give

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} |\mathcal{L}'(\mathbf{m}_x;\theta)| = 0. \tag{B.2}$$

Now for $\overline{h}$ small enough, Assumptions R-(i) and X give that for all θ in Θ^0,

$$\begin{aligned}\mathcal{L}'(\mathbf{b};\theta) &\geq C\int \psi\left(\left|\mathbf{U}(z)^\top \mathbf{H}\mathbf{b} - m(x+hz)\right|\right) f(x+hz)K(z)dz \\ &\geq C\int \psi\left(\left|\mathbf{U}(z)^\top \mathbf{H}\mathbf{b} - m(x)\right| - |m(x)-m(x+hz)|\right) K(z)dz, \qquad \text{(B.3)}\end{aligned}$$

where the last bound comes from the triangular inequality. In a first step we prove existence of $\mathbf{b}^*(\theta)$ for all θ in Θ^0 and in a second step we prove that $\mathbf{b}^*(\theta)$ is unique for all θ in Θ^0.

Step 1. To show existence of $\mathbf{b}^*(\theta)$ we first show that the set $\{\mathbf{b}^*(\theta), \theta \in \Theta^0\}$ is bounded. To show that we will prove that $\mathcal{L}'(\mathbf{b};\theta)$ can be made as large as we want by increasing $\|\mathbf{b}\|$ uniformly in θ in Θ^0. It will then follow from this, (B.2) and the fact that $\sup_{x\in\mathcal{X}} |m(x)| \leq C$ under Assumption M that $\{\mathbf{b}^*(\theta), \theta \in \Theta^0\}$ is bounded.

We study the lower bound in (B.3). Set $\mathbf{B} = \mathbf{H}\mathbf{b}$. Since $\mathbf{B} \mapsto \sup_{z\in\mathcal{B}(0,1)} |\mathbf{U}(z)^\top \mathbf{B}|$ is a norm over $\mathbb{R}^P$, equivalence of norms yields that there is a constant $C_1 > 0$ such that for all $\mathbf{B}$ in $\mathbb{R}^P$,

$$\sup_{z\in\mathcal{B}(0,1)} |\mathbf{U}(z)^\top \mathbf{B}| \geq C_1 \|\mathbf{B}\|. \qquad \text{(B.4)}$$

Since for all $\mathbf{B}$ in $\mathbb{R}^P$, $z \in \mathcal{B}(0,1) \mapsto |\mathbf{U}(z)^\top \mathbf{B}|$ is continuous over the compact $\mathcal{B}(0,1)$ in $\mathbb{R}^d$, there is a $z_{\mathbf{B}}$ in $\mathcal{B}(0,1)$ such that $\max_{z\in\mathcal{B}(0,1)} |\mathbf{U}(z)^\top \mathbf{B}| = |\mathbf{U}(z_{\mathbf{B}})^\top \mathbf{B}|$. Define also

$$N_{\mathbf{B}} = \left\{ z \in \mathcal{B}(0,1);\ \|\mathbf{U}(z) - \mathbf{U}(z_{\mathbf{B}})\| \leq \frac{C_1}{2} \right\},$$

The triangular inequality then yields that

$$\inf_{z\in N_{\mathbf{B}}} \left|\mathbf{U}(z)^\top \mathbf{B}\right| \geq \left|\mathbf{U}\left(z_{\mathbf{B}}\right)^\top \mathbf{B}\right| - \sup_{z\in N_{\mathbf{B}}} \left|\left(\mathbf{U}(z_{\mathbf{B}}) - \mathbf{U}(z)\right)^\top \mathbf{B}\right| \geq \frac{C_1}{2}\|\mathbf{B}\|.$$

Then since $N_{\mathbf{B}}$ is a subset of $\mathcal{B}(0,1)$ which is a subset of $\mathcal{K}$, since $\psi(\cdot)$ is an increasing function under Assumption R-(i), and since $m(\cdot)$ is continuous under Assumption M, for all $\mathbf{B}$ outside the ball $\mathcal{B}\left(0, 3\sup_{x\in\mathcal{X}} |m(x)|/C_1\right)$, the lower bound in (B.3) is greater than

$$\begin{aligned}&\inf_{z\in N_{\mathbf{B}}} \psi\left(\left|\mathbf{U}(z)^\top \mathbf{B} - m(x)\right| - |m(x)-m(x+hz)|\right) \int_{z\in N_{\mathbf{B}}} K(z)dz \\ &\geq \psi\left(\inf_{z\in N_{\mathbf{B}}} \left|\mathbf{U}(z)^\top \mathbf{B}\right| - \left(\sup_{x\in\mathcal{X}} |m(x)| + \sup_{z\in\mathcal{K}} |m(x)-m(x+hz)|\right)\right) \int_{z\in N_{\mathbf{B}}} K(z)dz \\ &\geq \psi\left(\|\mathbf{B}\|\right) \underline{K} \int_{z\in N_{\mathbf{B}}} dz, \qquad \text{(B.5)}\end{aligned}$$

where $\underline{K}$ is as in Assumption K. We now prove that

$$\inf_{z_{\mathbf{B}}\in\mathcal{B}(0,1)}\int_{z\in N_{\mathbf{B}}} dz \geq C > 0. \tag{B.6}$$

Since

$$z \in \mathcal{B}(0,1) \mapsto \|\mathbf{U}(z)\| = \left(1 + \sum_{1\leq|\mathbf{v}|\leq p}(z^{\mathbf{v}})^2\right)^{1/2},$$

$z \in \mathcal{B}(0,1) \mapsto \|\mathbf{U}(z)\|$ is continuously differentiable over the compact $\mathcal{B}(0,1)$ and for all z and $z_{\mathbf{B}}$ in $\mathcal{B}(0,1)$,

$$\|\mathbf{U}(z) - \mathbf{U}(z_{\mathbf{B}})\| \leq C_2 \|z - z_{\mathbf{B}}\|. \tag{B.7}$$

It then follows by definition of $N_{\mathbf{B}}$ that for all $z_{\mathbf{B}}$ in $\mathcal{B}(0,1)$,

$$\mathcal{B}\left(z_{\mathbf{B}}, \frac{C_1}{2C_2}\right) \cap \mathcal{B}(0,1) \subset N_{\mathbf{B}}.$$

Then

$$\inf_{z_{\mathbf{B}}\in\mathcal{B}(0,1)}\int_{N_{\mathbf{B}}} dz \geq \inf_{z_{\mathbf{B}}\in\mathcal{B}(0,1)}\int_{\mathcal{B}\left(z_{\mathbf{B}},\frac{C_1}{2C_2}\right)\cap\mathcal{B}(0,1)} dz.$$

Now take a large enough C_2 in (B.7) such that $C_3 = C_1/(2C_2) < 1/2$. Then with the convention that $(\|z_{\mathbf{B}}\| - C_3/2)\, z_{\mathbf{B}}/\|z_{\mathbf{B}}\| = 0$ when $z_{\mathbf{B}} = 0$

$$\mathcal{B}\left(\frac{\|z_{\mathbf{B}}\| - C_3/2}{\|z_{\mathbf{B}}\|} z_{\mathbf{B}}, \frac{C_3}{2}\right) \subset \mathcal{B}\left(z_{\mathbf{B}}, C_3\right) \cap \mathcal{B}(0,1).$$

It follows that

$$\inf_{z_{\mathbf{B}}\in\mathcal{B}(0,1)}\int_{\mathcal{B}(z_{\mathbf{B}},C_3)\cap\mathcal{B}(0,1)} dz \geq \int_{\mathcal{B}(0,C_3/2)} dz > 0.$$

This, (B.5) and the fact that $\psi(\cdot)$ is increasing then yield that there exists $C' > 0$ and $C'' > 0$ such for all θ in Θ^0,

$$\inf_{\mathbf{b};\|\mathbf{b}\|\geq C''} \mathcal{L}'(\mathbf{b};\theta) \geq C'.$$

Now observe that (B.2) yields that for $\overline{h}$ small enough and all θ in Θ^0, $\|\mathbf{b}^*(\theta)\| \leq C''$. Hence the set $\{\mathbf{b}^*(\theta), \theta \in \Theta^0\}$ is bounded.

Now since $\mathcal{L}'(\mathbf{m}_x;\theta) \geq \mathcal{L}'(\mathbf{b}^*(\theta);\theta)$, it follows from (B.2) and (B.3) that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \int \psi\left(\left|\mathbf{U}(z)^\top \mathbf{H}\mathbf{b}^*(\theta) - m(x+hz)\right|\right) K(z)dz = 0.$$

Hence $\mathbf{H}\mathbf{b}^*(\theta) = \mathbf{B}^*(\theta)$ gives that for any $\epsilon > 0$

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \int_{\mathcal{K}} \mathbb{I}\left(\left|\mathbf{U}(z)^\top \mathbf{B}^*(\theta) - m(x+hz)\right| \geq \epsilon\right) dz = 0. \tag{B.8}$$

Let $\varphi(\cdot)$ be an integrable function that vanishes outside $\mathcal{K}$ and such that $\int_{\mathcal{K}} \varphi(z)dz = 1$ and $\int_{\mathcal{K}} z^i \varphi(z)dz = 0$ for all $1 \leq |i| \leq p$. Then (B.8), boundedness of $\{\mathbf{B}^*(\theta), \theta \in \Theta^0\}$ and continuity of $m(\cdot)$, compactness of $\mathcal{X}_0$ and $\mathcal{K}$ yield that, uniformly in θ in Θ^0,

$$\begin{aligned} B_0^*(\theta) &= \int_{\mathcal{K}} \mathbf{U}(z)^\top \mathbf{B}^*(\theta)\varphi(z)dz \\ &= \int_{\mathcal{K}} m(x+hz)\varphi(z)dz + \int_{\mathcal{K}} \left(\mathbf{U}(z)^\top \mathbf{B}^*(\theta) - m(x+hz)\right)\varphi(z)dz = m(x)(1+o(1)). \end{aligned}$$

Consider a function $\varphi(\cdot)$ such that $\int_{\mathcal{K}} z^i \varphi(z)dz = 0$, $i \neq j$ for $|i| \leq p$ and $\int_{\mathcal{K}} z^j \varphi(z)dz = 1$. Arguing as above gives that $B_j^*(\theta) = \int_{\mathcal{K}} \mathbf{U}(z)^\top \mathbf{B}^*(\theta)\varphi(z)dz = m(x)\int_{\mathcal{K}} \varphi(z)dz + o(1) = o(1)$ uniformly in θ in Θ^0. It then follows that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \|\mathbf{B}^*(\theta) - \mathbf{m}_x\| = 0. \tag{B.9}$$

Step 2. We now prove that $\mathbf{B}^*(\theta)$ is uniquely defined for $\overline{h}$ small enough. Define

$$\overline{\mathcal{L}}(\mathbf{B};\theta) = \mathcal{L}(\mathbf{b};\theta), \tag{B.10}$$

where $\mathcal{L}(\mathbf{b};\theta)$ is as in (B.1), and its Hessian matrix

$$\mathbf{M}(\mathbf{B};\theta) = \frac{\partial^2}{\partial \mathbf{B}^\top \partial \mathbf{B}} \overline{\mathcal{L}}(\mathbf{B};\theta) = \int R^{(2)}(\mathbf{U}(z)^\top \mathbf{B}|x+hz)\mathbf{U}(z)\mathbf{U}(z)^\top K(z)f(x+hz)dz. \tag{B.11}$$

Under Assumptions K, R-(iii) and X

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} |\mathbf{M}(\mathbf{B};\theta) - \mathbf{M}(\mathbf{B};0,x)| = 0,$$

where $\mathbf{M}(\mathbf{B};0,x) = \int R^{(2)}(\mathbf{U}(z)^\top \mathbf{B}|x)\mathbf{U}(z)\mathbf{U}(z)^\top K(z)f(x)dz$, uniformly in $\mathbf{B}$ in any compact set. Under Assumption R-(iii) $\mathbf{M}(\mathbf{B};0,x)$ is a positive definite matrix for all $\mathbf{B}$ in $\mathcal{B}(\mathbf{m}_x, C)$, and all x in $\mathcal{X}_0$. It then follows that the function

$$\mathbf{B} \in \mathcal{B}(\mathbf{m}_x, C) \mapsto \overline{\mathcal{L}}(\mathbf{B};\theta),$$

is strictly convex for all x in $\mathcal{X}_0$ provided $\overline{h}$ is small enough. Now since all candidates $\mathbf{B}^*(\theta)$ tends to $\mathbf{m}_x$ when $\overline{h}$ tends to 0 by (B.9) so that they all belong to $\mathcal{B}(\mathbf{m}_x, C)$ for $\overline{h}$ small enough, $\mathbf{B}^*(\theta)$ is uniquely defined and so is $\mathbf{b}^*(\theta)$, for all θ in Θ^0.

Proof of (ii). Recall that $\mathbf{B}^*(\theta) = \mathbf{H}\mathbf{b}^*(\theta)$ achieves the minimum over $\mathbb{R}^P$ of $\mathbf{B} \mapsto \overline{\mathcal{L}}(\mathbf{B};\theta)$, where

$$\overline{\mathcal{L}}(\mathbf{B};\theta) = \int R\left(\mathbf{U}(z)^\top \mathbf{B}|x+hz\right) K(z)f(x+hz)dz,$$

is defined in (B.10). Observe that $\mathbf{B} \in \mathbb{R}^P \mapsto \overline{\mathcal{L}}(\mathbf{B};\theta)$ is continuously differentiable by Assumption R-(iii) and the Lebesgue Theorem. Then $(\partial/\partial \mathbf{B}^\top)\overline{\mathcal{L}}(\mathbf{B};\theta) = \int R^{(1)}(\mathbf{U}(z)^\top \mathbf{B}|x+hz)K(z)f(x+hz)\mathbf{U}(z)dz$, and $\mathbf{B}^*(\theta)$ satisfies the first order condition

$$\frac{\partial}{\partial \mathbf{B}^\top}\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta) = \int R^{(1)}(\mathbf{U}(z)^\top \mathbf{B}^*(\theta)|x+hz)K(z)f(x+hz)dz = 0,$$

which gives (ii).

Proof of (iii). (iii) directly follows from (B.9).

Proof of (iv). Recall that by (B.11), $\mathbf{B} \in \mathbb{R}^P \mapsto \overline{\mathcal{L}}(\mathbf{B};\theta)$ has a second order derivative $\mathbf{M}(\mathbf{B};\theta)$ and has an inverse in a neighborhood of $\mathbf{B}^*(\theta)$. Observe also that for $\overline{h}$ small enough, and for all $\mathbf{B}$ in $\mathbb{R}^P$, $\theta \in \Theta^0 \mapsto (\partial/\partial \mathbf{B}^\top)\overline{\mathcal{L}}(\mathbf{B};\theta)$ is continuously differentiable under Assumptions R-(iii) and X and by the Lebesgue Theorem. Then the Implicit Function Theorem (see e.g. Zeidler (1985), p.130) yields that $\theta \mapsto \mathbf{B}^*(\theta)$ is continuously differentiable over Θ^0, and that

$$\frac{\partial}{\partial \theta^\top}\mathbf{B}^*(\theta) = -\left[\mathbf{M}(\mathbf{B}^*(\theta);\theta)\right]^{-1} \frac{\partial}{\partial \theta^\top}\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta). \tag{B.12}$$

It then follows by uniform continuity of the functions above over the compact Θ^0 that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left| \frac{\partial}{\partial \theta^\top}\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta) - \frac{\partial}{\partial \theta^\top}\overline{\mathcal{L}}(\mathbf{B}^*(0,x);0,x) \right| = 0. \tag{B.13}$$

We will bound the θ-derivative of $\mathbf{B}^*(\theta)$. Observe that by (B.9) and Assumption K,

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \sup_{z\in\mathcal{K}} \left\| \mathbf{U}(z)^\top \left(\mathbf{B}^*(\theta) - \mathbf{m}_x\right)\right\| = 0.$$

Then since $\mathbf{U}(z)^\top \mathbf{m}_x = m(x)$ by definition of $\mathbf{U}(\cdot)$, Assumptions K, R-(iii) and X yield that

$$\mathbf{M}(\mathbf{B}^*(\theta);\theta) = \int R^{(2)}(\mathbf{U}(z)^\top \mathbf{B}^*(\theta)|x+hz)\mathbf{U}(z)\mathbf{U}(z)^\top K(z) f(x+hz)dz \succ C\underline{K}\int_{\mathcal{B}(0,1)} \mathbf{U}(z)\mathbf{U}(z)^\top dz,$$

for all θ in Θ^0 provided $\overline{h}$ is small enough. Since the RHS is positive definite, (B.12) and (B.13) yield that

$$\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\| \frac{\partial}{\partial\theta^\top}\mathbf{B}^*(\theta)\right\| \leq C \left\| \left(\int_{\mathcal{B}(0,1)} \mathbf{U}(z)\mathbf{U}(z)^\top dz\right)^{-1}\right\| \lim_{\overline{h}\to 0} \sup_{\theta\in\Theta^0} \left\| \frac{\partial}{\partial\theta^\top}\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta)\right\| \leq C.$$

Now recall $m^*(x';\theta) = \mathbf{U}(x'-x)\mathbf{b}^*(\theta) = \mathbf{U}((x'-x)/h)\mathbf{B}^*(\theta)$ by (3.3.4) and observe that

$$\sup_{(x',\theta)\in\mathcal{X}\times\Theta} \left\| \mathbf{U}\left(\frac{x'-x}{h}\right)\right\| \leq \frac{C}{\underline{h}^p} \quad \text{and} \quad \sup_{(x',\theta)\in\mathcal{X}\times\Theta} \left\| \frac{\partial}{\partial(h,x)^\top}\mathbf{U}\left(\frac{x'-x}{h}\right)\right\| \leq \frac{C}{\underline{h}^{p+1}}.$$

Thus uniformly in (x',θ) in $\mathcal{X}\times\Theta$,

$$\left\| \frac{\partial}{\partial\theta^\top} m^*(x';\theta)\right\| \leq \left\| \mathbf{U}\left(\frac{x'-x}{h}\right)\right\| \left\| \frac{\partial}{\partial\theta^\top}\mathbf{B}^*(\theta)\right\| + \left\| \frac{\partial}{\partial\theta^\top}\mathbf{U}\left(\frac{x'-x}{h}\right)\right\| \|\mathbf{B}^*(\theta)\| \leq \frac{C}{\underline{h}^{p+1}},$$

uniformly in θ in Θ, provided $\overline{h}$ is less than 1. The Taylor inequality shows that (iv) is proved.

Proof of (v). Lemma A.1-(iii) gives

$$\begin{aligned}
&\lim_{\overline{h}\to 0} \sup_{\theta\in\Theta} \sup_{x'\in\mathcal{X}} K_h(x'-x)\left|m^*(x';\theta) - m(x)\right| \\
&= \lim_{\overline{h}\to 0} \sup_{(\theta,x')\in\Theta\times\mathcal{X}} K_h(x'-x)\left| \mathbf{U}\left(\frac{x'-x}{h}\right)^\top (\mathbf{B}^*(\theta) - \mathbf{m}_x)\right| \\
&\leq \lim_{\overline{h}\to 0} \sup_{\theta\in\Theta} \|\mathbf{B}^*(\theta) - \mathbf{m}_x\| \sup_{z\in\mathcal{K}} K(z)\left\|\mathbf{U}(z)\right\| = 0,
\end{aligned}$$

where $\mathbf{m}_x = (m(x), 0, \ldots, 0)^\top$ in $\mathbb{R}^P$ and under Assumption K. □

B.2 Proof of Proposition A.1

Let $\mathbf{B} = \mathbf{H}\mathbf{b}$ and $\widehat{\mathcal{L}}_n(\mathbf{B};\theta) = \mathcal{L}_n(\mathbf{b};\theta)$, so that $\widehat{\mathbf{B}}(\theta) = \mathbf{H}\widehat{\mathbf{b}}(\theta)$ achieves the minimum over $\mathbb{R}^P$ of $\mathbf{B} \mapsto \widehat{\mathcal{L}}_n(\mathbf{B};\theta)$.

In a first step, we prove that for all $t > 0$, we have

$$\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{B};\theta)\right]\right| = o_{\mathbb{P}}(1). \tag{B.14}$$

In a second step we prove that the set $\left\{\widehat{\mathbf{B}}(\theta), \theta \in \Theta\right\}$ is bounded and in a third step we prove the Proposition.

Step 1 : $\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{B};\theta)\right]\right| = o_{\mathbb{P}}(1)$. Since $\Theta\times\mathcal{B}(0,t)$ is a compact subset of $\mathbb{R}^{P+d+1}$, for any positive sequence $\eta_n = o\left(\underline{h}^{p+2d+1}\right)$ there is a finite number $V_n = V_n(\underline{h})$ such that $\Theta\times\mathcal{B}(0,t)$ is a subset of $\bigcup_{1\le k\le V_n}\mathcal{B}((\theta_k,\mathbf{B}_k),\eta_n)$. Now observe that

$$\begin{aligned}
\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{B};\theta)\right]\right| \le\ & \underbrace{\max_{1\le k\le V_n}\left|\widehat{\mathcal{L}}_n(\mathbf{B}_k;\theta_k) - \mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{B}_k;\theta_k)\right]\right|}_{A_1} \\
& + \underbrace{\max_{1\le k\le V_n}\sup_{(\theta,\mathbf{B})\in\mathcal{B}((\theta_k,\mathbf{B}_k),\eta_n)}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \widehat{\mathcal{L}}_n(\mathbf{B}_k;\theta_k)\right|}_{A_2} \\
& + \underbrace{\max_{1\le k\le V_n}\sup_{(\theta,\mathbf{B})\in\mathcal{B}((\theta_k,\mathbf{B}_k),\eta_n)}\left|\mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \widehat{\mathcal{L}}_n(\mathbf{B}_k;\theta_k)\right]\right|}_{A_3}.
\end{aligned}$$

For the rest of this part of the proof observe that Assumption L-(iii) and the convex inequality yield that for any fixed $C > 0$

$$\begin{aligned}
\sup_{x\in\mathcal{X}}\mathbb{E}\left[\sup_{|z|\le C}|\rho(Y-z)|^\nu|X=x\right] &= \sup_{x\in\mathcal{X}}\mathbb{E}\left[\sup_{|z|\le C}\left|\rho(Y) + \int_0^z r(Y-u)du\right|^\nu|X=x\right] \\
&\le \sup_{x\in\mathcal{X}}\mathbb{E}\left[\sup_{|z|\le C}2^{\nu-1}\left(|\rho(Y)|^\nu + |z|^\nu\int_0^z|r(Y-u)|^\nu du\right)|X=x\right] \\
&\le C\sup_{x\in\mathcal{X}}\mathbb{E}\left[|\rho(Y)|^\nu|X=x\right] + C\sup_{x\in\mathcal{X}}\mathbb{E}\left[\sup_{|z|\le C}|r(Y-z)|^\nu|X=x\right] < \infty.
\end{aligned} \tag{B.15}$$

We now return to the control of A_1, A_2 and A_3.

Control of A_2 and A_3. For all $(\mathbf{B},\theta)$ in $\mathbb{R}^P \times \Theta$, let

$$\ell(\mathbf{B};\theta) = \frac{1}{h^d}\rho\left(Y - \mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{B}\right) K\left(\frac{X-x}{h}\right).$$

First observe that by definition of $\widehat{\mathcal{L}}_n$,

$$\begin{aligned}
\mathbb{E}\left[A_2\right] &= \mathbb{E}\left[\max_{1\le k\le V_n} \sup_{(\theta,\mathbf{B})\in\mathcal{B}((\theta_k,\mathbf{B}_k),\eta_n)} \left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \widehat{\mathcal{L}}_n(\mathbf{B}_k;\theta_k)\right|\right] \\
&\le \mathbb{E}\left[\sup_{\substack{((\theta,\mathbf{B}),(\theta',\mathbf{B}'))\in(\Theta\times\mathcal{B}(0,t))^2 \\ \|(\theta,\mathbf{B})-(\theta',\mathbf{B}')\|\le\eta_n}} \left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \widehat{\mathcal{L}}_n(\mathbf{B}';\theta')\right|\right] \\
&\le \mathbb{E}\left[\sup_{\substack{((\theta,\mathbf{B}),(\theta',\mathbf{B}'))\in(\Theta\times\mathcal{B}(0,t))^2 \\ \|(\theta,\mathbf{B})-(\theta',\mathbf{B}')\|\le\eta_n}} \left|\ell(\mathbf{B};\theta) - \ell(\mathbf{B}';\theta')\right|\right].
\end{aligned}$$

Now observe that $|\ell(\mathbf{B};\theta) - \ell(\mathbf{B}';\theta')|$ is less than

$$\begin{aligned}
&\underbrace{\left|\frac{1}{h'^d}K\left(\frac{X-x'}{h'}\right)\left(\rho\left(Y - \mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{B}\right) - \rho\left(Y - \mathbf{U}\left(\frac{X-x'}{h'}\right)^\top \mathbf{B}'\right)\right)\right|}_{\widehat{A}_1} \\
&+ \underbrace{\left|\frac{1}{h^d}\left(K\left(\frac{X-x}{h}\right) - K\left(\frac{X-x'}{h'}\right)\right)\rho\left(Y - \mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{B}\right)\right|}_{\widehat{A}_2} \\
&+ \underbrace{\left|\frac{1}{h^d} - \frac{1}{h'^d}\right|\left|\rho\left(Y - \mathbf{U}\left(\frac{X-x'}{h'}\right)^\top \mathbf{B}'\right)\right| K\left(\frac{X-x'}{h'}\right)}_{\widehat{A}_3}, \qquad \text{(B.16)}
\end{aligned}$$

so that

$$\mathbb{E}\left[A_2\right] \le \sum_{j=1}^{3} \mathbb{E}\left[\sup_{\substack{((\theta,\mathbf{B}),(\theta',\mathbf{B}'))\in(\Theta\times\mathcal{B}(0,t))^2 \\ \|(\theta,\mathbf{B})-(\theta',\mathbf{B}')\|\le\eta_n}} \widehat{A}_j\right].$$

We now control the three terms in the equation above. We will only consider $\widehat{A}_1$ since the two other term can be studied similarly. First observe that for all $(\theta, \mathbf{B})$, $(\theta', \mathbf{B}')$ in $\Theta \times \mathcal{B}(0,t)$ with $\|(\theta, \mathbf{B}) - (\theta', \mathbf{B}')\| \leq \eta_n$, we have

$$\left| \mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{B} - \mathbf{U}\left(\frac{X-x'}{h'}\right)^\top \mathbf{B}' \right| \leq C\frac{\eta_n}{\underline{h}^{p+1}} \leq C,$$

since $\eta_n = o\left(\underline{h}^{p+2d+1}\right)$. Then for all $(\theta, \mathbf{B})$, $(\theta', \mathbf{B}')$ in $\Theta \times \mathcal{B}(0,t)$, such that $\|(\theta, \mathbf{B}) - (\theta', \mathbf{B}')\| \leq \eta_n$, we have

$$\widehat{A}_1 \leq \frac{1}{h'^d} K\left(\frac{X-x'}{h'}\right) \left| \int_{\mathbf{U}\left(\frac{X-x'}{h'}\right)^\top \mathbf{B}'}^{\mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{B}} |r\left(Y-u\right)| \, du \right| \leq C\frac{\eta_n}{\underline{h}^{p+d+1}} \sup_{|z|\leq C} |r(Y-z)|,$$

since $K(\cdot)$ is bounded. Hence the Hölder inequality and Assumption L-(iii) give

$$\begin{aligned} \mathbb{E}\left[\sup_{\substack{((\theta,\mathbf{B}),(\theta',\mathbf{B}'))\in(\Theta\times\mathcal{B}(0,t))^2 \\ \|(\theta,\mathbf{B})-(\theta',\mathbf{B}')\|\leq\eta_n}} \widehat{A}_1 \right] &\leq C\frac{\eta_n}{\underline{h}^{p+d+1}} \mathbb{E}\left[\sup_{|z|\leq C} |r(Y-z)|\right] \\ &\leq C\frac{\eta_n}{\underline{h}^{p+d+1}} \mathbb{E}^{1/\nu}\left[\sup_{|z|\leq C} |r(Y-z)|^\nu \, |X=x\right] = O\left(\frac{\eta_n}{\underline{h}^{p+d+1}}\right). \end{aligned}$$

Now arguing similarly for $\widehat{A}_2$ and $\widehat{A}_3$ using (B.15) yields that

$$\mathbb{E}\left[A_2\right] \leq \sum_{j=1}^3 \mathbb{E}\left[\sup_{\substack{((\theta,\mathbf{B}),(\theta',\mathbf{B}'))\in(\Theta\times\mathcal{B}(0,t))^2 \\ \|(\theta,\mathbf{B})-(\theta',\mathbf{B}')\|\leq\eta_n}} \widehat{A}_j \right] = O\left(\eta_n\left(\frac{1}{\underline{h}^{p+d+1}} + \frac{1}{\underline{h}^{2d+1}} + \frac{1}{\underline{h}^{d+1}} + \frac{1}{\underline{h}^2}\right)\right) = o(1),$$

by definition of η_n. The Markov inequality yields that $A_2 = o_{\mathbb{P}}(1)$. Finally observe that $A_3 \leq \mathbb{E}\left[A_2\right] = o(1)$.

Control of A_1. Define for all $1 \leq k \leq V_n$,

$$\rho_{i,k} = \rho\left(Y_i - \mathbf{U}\left(\frac{X_i - x_k}{h_k}\right)^\top \mathbf{B}_k\right) \text{ and } T_i = \mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right)\rho\left(Y_i - \mathbf{U}\left(\frac{X_i - x}{h}\right)^\top \mathbf{B}\right). \tag{B.17}$$

Now observe that A_1 is less than

$$\underbrace{\max_{1\le k\le V_n}\left|\frac{1}{nh_k^d}\sum_{i=1}^n\left(\rho_{i,k}K_{h_k}\left(X_i-x_k\right)\mathbb{I}\left(\left|\rho_{i,k}\right|\le un^{1/\nu}\right)-\mathbb{E}\left[\rho_{i,k}K_{h_k}\left(X_i-x_k\right)\mathbb{I}\left(\left|\rho_{i,k}\right|\le un^{1/\nu}\right)\right]\right)\right|}_{\widetilde{A}_1}$$
$$+\underbrace{\max_{1\le k\le V_n}\left|\frac{1}{nh_k^d}\sum_{i=1}^n\rho_{i,k}K\left(\frac{X_i-x_k}{h_k}\right)\mathbb{I}\left(\left|\rho_{i,k}\right|\ge un^{1/\nu}\right)\right|}_{\widetilde{A}_2}$$
$$+\underbrace{\max_{1\le k\le V_n}\left|\mathbb{E}\left[\frac{1}{h_k^d}\rho_{i,k}K\left(\frac{X_i-x_k}{h_k}\right)\mathbb{I}\left(\left|\rho_{i,k}\right|\ge un^{1/\nu}\right)\right]\right|}_{\widetilde{A}_3},$$

where u is a positive real number.

We first consider $\widetilde{A}_2$. Observe that

$$\begin{aligned}\widetilde{A}_2 &\le \sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\frac{1}{nh^d}\sum_{i=1}^n\left|T_i\right|\mathbb{I}\left(\left|T_i\right|\ge un^{1/\nu}\right)K\left(\frac{X_i-x}{h}\right)\\ &\le \max_{1\le i\le n}\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|T_i\right|\mathbb{I}\left(\left|T_i\right|\ge un^{1/\nu}\right)\sup_{\theta\in\Theta}\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right).\end{aligned} \tag{B.18}$$

Now observe that

$$\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top\mathbf{B}\right|\mathbb{I}\left(\frac{X_i-x}{h}\in\mathcal{K}\right)\le\sup_{z\in\mathcal{K}}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|\mathbf{U}\left(z\right)^\top\mathbf{B}\right|\le C,$$

so that by definition (B.17) of T_i

$$\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|T_i\right|\le\sup_{|z|\le C}\left|\rho(Y-z)\right|. \tag{B.19}$$

Now observe that

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le i\le n}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\sup_{\theta\in\Theta}\left|T_i\right|\mathbb{I}\left(\left|T_i\right|\ge Cun^{1/\nu}\right)\neq 0\right)\\ &\quad=\mathbb{P}\left(\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\left|T_i\right|\ge un^{1/\nu}\text{ for some }1\le i\le n\right)\end{aligned}$$

$$\leq \sum_{i=1}^{n} \mathbb{P}\left(\sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} |T_i| \geq un^{1/\nu} \right) \leq n\mathbb{P}\left(\sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} |T_i| \geq un^{1/\nu} \right)$$
$$\leq n \frac{\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)} |T_i|^{\nu}\right]}{nu^{\nu}} = \frac{\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)} |T_i|^{\nu}\right]}{u^{\nu}}. \tag{B.20}$$

Now observe that by (B.15) and (B.19) we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} |T_i|^{\nu} \right] &= \mathbb{E}\left[\mathbb{E}\left[\sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} \left| \rho\left(Y - \mathbf{U}\left(\frac{X-x}{h}\right)^{\top} \mathbf{B} \right) \mathbb{I}\left(\frac{X-x}{h} \in \mathcal{K} \right) \right|^{\nu} \right] |X \right] \\
&\leq \sup_{x\in\mathcal{X}} \mathbb{E}\left[\sup_{|z|\leq C} |\rho(Y-z)|^{\nu} |X=x \right] \leq C.
\end{aligned}$$

It then follows from (B.20) that with a probability that can be arbitrarily large by increasing u, we have that for all $1 \leq i \leq n$, $\sup_{\mathbf{B}\in\mathcal{B}(0,t)} \sup_{\theta\in\Theta} |T_i| \, \mathbb{I}\left(|T_i| \geq un^{1/\nu}\right) = 0$. Then (B.18) yields that $\widetilde{A}_2 = 0$ with a probability than can be made arbitrarily large by increasing u.

We now consider $\widetilde{A}_3$. Definition (B.17) of T_i yields that

$$\begin{aligned}
\widetilde{A}_3 &= \max_{1\leq k\leq V_n} \left| \mathbb{E}\left[\frac{1}{h_k^d} K\left(\frac{X_i - x_k}{h_k} \right) \rho_{i,k} \mathbb{I}\left(|\rho_{i,k}| \geq un^{1/\nu} \right) \right] \right| \\
&\leq \sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} \mathbb{E}\left[\frac{1}{h^d} K\left(\frac{X_i - x}{h} \right) |T_i| \, \mathbb{I}\left(|T_i| \geq un^{1/\nu} \right) \right],
\end{aligned}$$

for any $1 \leq i \leq n$. Then

$$\begin{aligned}
\widetilde{A}_3 &\leq \sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} \mathbb{E}\left[\frac{1}{h^d} K\left(\frac{X_i - x}{h} \right) \mathbb{E}\left[|T_i| \, \mathbb{I}\left(|T_i| \geq un^{1/\nu} \right) |X \right] \right] \\
&\leq \sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} \mathbb{E}\left[\frac{1}{h^d} K\left(\frac{X_i - x}{h} \right) \frac{\mathbb{E}\left[|T_i|^{\nu} |X \right]}{u^{\nu-1} n^{(\nu-1)/\nu}} \right] \\
&\leq Cu^{1-\nu} n^{(1-\nu)/\nu} \sup_{x'\in\mathcal{X}} \sup_{\theta\in\Theta} \sup_{\mathbf{B}\in\mathcal{B}(0,t)} \mathbb{E}\left[|T_i|^{\nu} |X = x' \right] \\
&\leq Cu^{1-\nu} n^{(1-\nu)/\nu} \sup_{x'\in\mathcal{X}} \mathbb{E}\left[\sup_{|z|\leq C} |\rho(Y-z)|^{\nu} |X=x' \right] \leq Cn^{(1-\nu)/\nu},
\end{aligned}$$

where $\mathbb{P}(\cdot|X) = \mathbb{E}\left[\mathbb{I}(\cdot)|X\right]$ and by (B.15). Now since $\nu > 2$, $\widetilde{A}_3 = o\left(\log^{1/2} n / \left(n\underline{h}^d \right)^{1/2} \right)$ under Assumption K.

We now consider $\widetilde{A}_1$. Define $\overline{T}_{i,k} = \rho_{i,k}\mathbb{I}\left(|\rho_{i,k}| \leq un^{1/\nu}\right)$. Observe that the Bonferroni inequality yields that

$$\mathbb{P}\left(\widetilde{A}_1 \geq t\right) \leq V_n \max_{1\leq k\leq V_n} \mathbb{P}\left(\left|\frac{1}{nh_k^d}\sum_{i=1}^n\left(\overline{T}_{i,k}K\left(\frac{X_i - x_k}{h_k}\right) - \mathbb{E}\left[\overline{T}_{i,k}K\left(\frac{X_i - x_k}{h_k}\right)\right]\right)\right| \geq t\right). \tag{B.21}$$

Observe that for all $1 \leq i \leq n$,

$$\left|\frac{1}{h_k^d}\overline{T}_{i,k}K\left(\frac{X_i - x_k}{h_k}\right)\right| \leq C\frac{n^{1/\nu}}{\underline{h}^d},$$

uniformly in $1 \leq k \leq V_n$. Observe also that the definition (B.17) of T_i yields that

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{h_k^d}\overline{T}_{i,k}K\left(\frac{X_i - x_k}{h_k}\right)\right) &\leq \mathbb{E}\left[\left(\frac{1}{h_k^d}\overline{T}_{i,k}K\left(\frac{X_i - x_k}{h_k}\right)\right)^2\right]\\ &\leq \sup_{\theta\in\Theta}\sup_{\mathbf{B}\in\mathcal{B}(0,t)}\mathbb{E}\left[\frac{1}{h^{2d}}T_i^2K\left(\frac{X_i - x}{h}\right)^2\right] \leq \frac{C}{\underline{h}^d},\end{aligned}$$

by (B.19) and (B.15). Then the Bernstein inequality, (B.21) and $V_n = O\left(\underline{h}^{-(p+2d+1)}\right)^{P+d+1}$ yield that

$$\begin{aligned}\mathbb{P}\left(\widetilde{A}_1 \geq t\right) &\leq CV_n\exp\left(-C\frac{n^2t^2}{\frac{n}{\underline{h}^d} + \frac{n^{1+1/\nu}}{\underline{h}^d}t}\right) \leq C\exp\left(-C\frac{t^2}{\frac{1}{n\underline{h}^d} + \frac{n^{1/\nu}}{n\underline{h}^d}t} - C\log(\underline{h})\right)\\ &\leq C\exp\left(-Ct^2\frac{n\underline{h}^d}{n^{1/\nu}}\left(1 - \frac{n^{1/\nu}\log(\underline{h})}{t^2n\underline{h}^d}\right)\right) = o(1),\end{aligned}$$

under Assumption K. This ends the proof of (B.14). Furthermore observe that since there exists C such that $\sup_{x\in\mathcal{X}}|m(x)| \leq C$, (B.14) gives that

$$\sup_{\theta\in\Theta}\left|\widehat{\mathcal{L}}_n(\mathbf{m}_x;\theta) - \mathbb{E}\left[\widehat{\mathcal{L}}_n(\mathbf{m}_x;\theta)\right]\right| = o_{\mathbb{P}}(1), \tag{B.22}$$

where $\mathbf{m}_x = (m(x), 0, \ldots, 0)^\top$ in $\mathbb{R}^P$.

Now observe that (B.15) still holds with $\Psi(|\cdot|)$ instead of $\rho(\cdot)$ under Assumption R-(ii). Furthermore $\Psi(|\cdot|)$ is a positive subadditive continuous function and $\Psi(|\cdot|) \leq |\cdot|$ under Assumption

R-(ii). Then for all (z, z') in $\mathbb{R}^2$,

$$|\Psi(|z|) - \Psi(|z'|)| \leq \Psi(|z - z'|) \leq C\,|z - z'|\,.$$

Then changing $\rho(\cdot)$ into $\Psi(|\cdot|)$ in (B.16) and arguing as in Step 1 under Assumption K, M, R and X yield

$$\sup_{\theta\in\Theta}\left|\frac{1}{nh^d}\sum_{i=1}^{n}\left(\Psi\left(|Y_i - m(x)|\right)K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left[\Psi\left(|Y_i - m(x)|\right)K\left(\frac{X_i - x}{h}\right)\right]\right)\right| = o_{\mathbb{P}}(1).$$

Observe that

$$\sup_{\theta\in\Theta}\left|\frac{1}{h^d}\mathbb{E}\left[\Psi\left(|Y - m(x)|\right)K\left(\frac{X - x}{h}\right)\right] - f(x)\mathbb{E}\left[\Psi\left(|Y - m(x)|\right)|X = x\right]\right| = o(1),$$

under Assumption R-(ii). This gives

$$\sup_{\theta\in\Theta}\left|f(x)\mathbb{E}\left[\Psi\left(|Y - m(x)|\right)|X = x\right] - \frac{1}{nh^d}\sum_{i=1}^{n}\Psi\left(|Y_i - m(x)|\right)K\left(\frac{X_i - x}{h}\right)\right| = o_{\mathbb{P}}(1). \quad \text{(B.23)}$$

Step 2. We now show that the set $\left\{\widehat{\mathbf{B}}(\theta), \theta\in\Theta\right\}$ is bounded with a large probability. To show that we will use the fact that $\widehat{\mathbf{B}}(\theta)$ achieves the minimum over $\mathbb{R}^P$ of $\mathbf{B}\mapsto\widehat{\mathcal{L}}_n(\mathbf{B};\theta)$, and we will show that $\widehat{\mathcal{L}}_n(\mathbf{B};\theta)$ can be made as large as desired by increasing $\|\mathbf{B}\|$. Indeed, we will show using (B.22) that

$$\inf_{\|\mathbf{B}\|\geq C}\widehat{\mathcal{L}}_n(\mathbf{B};\theta) > \widehat{\mathcal{L}}_n(\mathbf{m}_x;\theta),$$

uniformly in θ in Θ with a probability that can be arbitrarily large. This will prove that $\sup_{\theta\in\Theta}\left\|\widehat{\mathbf{B}}(\theta)\right\| \leq C(1 + o_{\mathbb{P}}(1))$.

We now return to the proof. Using the fact that $\Psi(\cdot)$ is increasing and subadditive over $\mathbb{R}$ under Assumption R-(ii) we have

$$\begin{aligned}
\widehat{\mathcal{L}}_n(\mathbf{B};\theta) &\geq \frac{1}{nh^d}\sum_{i=1}^{n}\Psi\left(\left|\left(\mathbf{U}\left(\frac{X_i - x}{h}\right)^{\top}\mathbf{B} - m(x)\right) - (Y_i - m(x))\right|\right)K\left(\frac{X_i - x}{h}\right)\\
&\geq \frac{1}{nh^d}\sum_{i=1}^{n}\Psi\left(\left|\mathbf{U}\left(\frac{X_i - x}{h}\right)^{\top}\mathbf{B} - m(x)\right| - |Y_i - m(x)|\right)K\left(\frac{X_i - x}{h}\right)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{C}{nh^d}\sum_{i=1}^n\left(\Psi\left(\left|\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top\mathbf{B}-m(x)\right|\right)-\Psi\left(|Y_i-m(x)|\right)\right)K\left(\frac{X_i-x}{h}\right)\\
&\geq \frac{C}{nh^d}\sum_{i=1}^n\Psi\left(\left|\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top\mathbf{B}-m(x)\right|\right)K\left(\frac{X_i-x}{h}\right)\\
&\quad -(1+o_{\mathbb{P}}(1))f(x)\mathbb{E}\left[\Psi(|Y-m(x)|)|X=x\right],
\end{aligned}$$

uniformly in θ in Θ, by the triangular inequality and (B.23). Now observe that

$$\sup_{x\in\mathcal{X}} f(x)\mathbb{E}\left[\Psi(|Y-m(x)|)|X=x\right]\leq C,$$

under Assumptions L and R. Then

$$\widehat{\mathcal{L}}_n(\mathbf{B};\theta)\geq C(1+o_{\mathbb{P}}(1))\left(-C+\frac{1}{nh^d}\sum_{i=1}^n\Psi\left(\left|\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top\mathbf{B}-m(x)\right|\right)K\left(\frac{X_i-x}{h}\right)\right),$$

uniformly in θ in Θ. We study the lower bound in the equation above. Using similar arguments as those lying between (B.4) and (B.6) yields that for all $\|\mathbf{B}\|\geq C$

$$\begin{aligned}
\widehat{\mathcal{L}}_n(\mathbf{B};\theta) &\geq C(1+o_{\mathbb{P}}(1))\left(\inf_{z\in N_{\mathbf{B}}}\Psi\left(\left|\mathbf{U}(z)^\top\mathbf{B}\right|-|m(x)|\right)\left[\frac{\underline{K}}{nh^d}\sum_{i=1}^n\mathbb{I}\left(\frac{X_i-x}{h}\in N_{\mathbf{B}}\right)\right]-C\right)\\
&\geq C(1+o_{\mathbb{P}}(1))\left(\Psi\left(C\|\mathbf{B}\|\right)\left[\frac{C}{nh^d}\sum_{i=1}^n\mathbb{I}\left(\frac{X_i-x}{h}\in N_{\mathbf{B}}\right)\right]-C\right). \qquad \text{(B.24)}
\end{aligned}$$

Now observe that Einmahl and Mason (2005, Theorem 1) yields that

$$\sup_{\theta\in\Theta}\left|\frac{1}{h^d}\mathbb{P}\left(\frac{X_i-x}{h}\in N_{\mathbf{B}}\right)-\frac{1}{nh^d}\sum_{i=1}^n\mathbb{I}\left(\frac{X_i-x}{h}\in N_{\mathbf{B}}\right)\right|=o_{\mathbb{P}}(1),$$

under Assumptions K and X. Now observe that arguing as for (B.6)

$$\begin{aligned}
\frac{1}{h^d}\mathbb{P}\left(\frac{X_i-x}{h}\in N_{\mathbf{B}}\right) &= \frac{1}{h^d}\int\mathbb{I}\left(\frac{z-x}{h}\in N_{\mathbf{B}}\right)f(z)dz=\int\mathbb{I}\left(z\in N_{\mathbf{B}}\right)f(x+hz)dz\\
&\geq C\int\mathbb{I}\left(z\in N_{\mathbf{B}}\right)dz\geq\inf_{z_{\mathbf{B}}\in\mathcal{B}(0,1)}\int_{z\in N_{\mathbf{B}}}dz\geq C.
\end{aligned}$$

This and (B.24) yield that uniformly in θ in Θ and for all $\mathbf{B}$, such that $\|\mathbf{B}\|\geq C$,

$$\widehat{\mathcal{L}}_n(\mathbf{B};\theta)\geq C'(1+o_{\mathbb{P}}(1))\Psi\left(C\|\mathbf{B}\|\right).$$

Now since $t \in \mathbb{R}_+ \mapsto \Psi(t)$ is increasing under Assumption R-(ii)

$$\inf_{\mathbf{B}, \|\mathbf{B}\| \geq C} \widehat{\mathcal{L}}_n(\mathbf{B};\theta) \geq C'(1+o_{\mathbb{P}}(1)) > \widehat{\mathcal{L}}_n(\mathbf{m}_x;\theta),$$

uniformly in θ in Θ by (B.22). This ensures that $\sup_{\theta\in\Theta} \left\|\widehat{\mathbf{B}}(\theta)\right\| \leq C(1+o_{\mathbb{P}}(1))$.

Step 3. Observe that (B.9) yields that $\sup_{\theta\in\Theta} \|\mathbf{B}^*(\theta)\| \leq C$ since

$$\sup_{\theta\in\Theta} \|\mathbf{H}\mathbf{m}_x\| = \sup_{x\in\mathcal{X}} |m(x)| \leq C,$$

by continuity of $m(\cdot)$ and compactness of $\mathcal{X}$. Now observe that (B.14) yields that

$$\begin{aligned}
\sup_{\theta\in\Theta} \left|\widehat{\mathcal{L}}_n\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right)\right| &= O_{\mathbb{P}}\left(\sup_{\theta\in\Theta}\sup_{\|\mathbf{B}\|\leq C}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \overline{\mathcal{L}}(\mathbf{B};\theta)\right|\right) = o_{\mathbb{P}}(1), \\
\sup_{\theta\in\Theta} \left|\widehat{\mathcal{L}}_n(\mathbf{B}^*(\theta);\theta) - \overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta)\right| &= O_{\mathbb{P}}\left(\sup_{\theta\in\Theta}\sup_{\|\mathbf{B}\|\leq C}\left|\widehat{\mathcal{L}}_n(\mathbf{B};\theta) - \overline{\mathcal{L}}(\mathbf{B};\theta)\right|\right) = o_{\mathbb{P}}(1),
\end{aligned}$$

where $\overline{\mathcal{L}}$ is as in (B.10). Recall that

$$\widehat{\mathbf{B}}(\theta) = \operatorname*{Arg\,Min}_{\mathbf{B}\in\mathbb{R}^P} \widehat{\mathcal{L}}_n(\mathbf{B};\theta) \quad\text{and}\quad \mathbf{B}^*(\theta) = \operatorname*{Arg\,Min}_{\mathbf{B}\in\mathbb{R}^P} \overline{\mathcal{L}}(\mathbf{B};\theta),$$

so that $\overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \overline{\mathcal{L}}\left(\mathbf{B}^*(\theta);\theta\right) \geq 0$ and $\widehat{\mathcal{L}}_n\left(\widehat{\mathbf{B}}(\theta);\theta\right) \leq \widehat{\mathcal{L}}_n\left(\mathbf{B}^*(\theta);\theta\right)$. Then

$$\begin{aligned}
0 &\leq \overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \overline{\mathcal{L}}\left(\mathbf{B}^*(\theta);\theta\right) \\
&= \overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \widehat{\mathcal{L}}_n\left(\widehat{\mathbf{B}}(\theta);\theta\right) + \widehat{\mathcal{L}}_n\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \widehat{\mathcal{L}}_n(\mathbf{B}^*(\theta);\theta) \\
&\leq \overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \widehat{\mathcal{L}}_n\left(\widehat{\mathbf{B}}(\theta);\theta\right) + \widehat{\mathcal{L}}_n\left(\mathbf{B}^*(\theta);\theta\right) - \widehat{\mathcal{L}}_n(\mathbf{B}^*(\theta);\theta).
\end{aligned}$$

This gives that

$$\sup_{\theta\in\Theta}\left|\overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right) - \overline{\mathcal{L}}\left(\mathbf{B}^*(\theta);\theta\right)\right| \leq 2\sup_{\theta\in\Theta}\sup_{\|\mathbf{B}\|\leq C}\left|\overline{\mathcal{L}}\left(\mathbf{B};\theta\right) - \widehat{\mathcal{L}}_n\left(\mathbf{B};\theta\right)\right| = o_{\mathbb{P}}(1). \tag{B.25}$$

Now observe that for all θ in Θ, $\mathbf{B}^*(\theta) = \operatorname*{Arg\,Min}_{\mathbf{B}\in\mathbb{R}^P} \overline{\mathcal{L}}(\mathbf{B};\theta)$, $\lim_{\|\mathbf{B}\|\to\infty} \overline{\mathcal{L}}(\mathbf{B};\theta) = \infty$ and that $\sup_{\theta\in\Theta}\|\mathbf{B}^*(\theta)\| \leq C$, so that there is C and for all θ in Θ, $\inf_{\mathbf{B};\|\mathbf{B}-\mathbf{B}^*(\theta)\|\geq C} \overline{\mathcal{L}}(\mathbf{B};\theta) \geq$

$\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta)+1$. We now consider $\mathbf{B}\in\mathcal{B}(\mathbf{B}^*(\theta),C)\mapsto\overline{\mathcal{L}}(\mathbf{B};\theta)$. Observe that $\mathbf{B}\mapsto\overline{\mathcal{L}}(\mathbf{B};\theta)$ is continuous over the compact $\mathcal{B}(\mathbf{B}^*(\theta),C)$ for all θ in Θ and that $\overline{\mathcal{L}}(\mathbf{B};\theta)$ attains its unique absolute minimum at $\mathbf{B}^*(\theta)$ by Lemma A.1-(i). Then there exists for every $\varepsilon>0$ a number $\eta>0$ such that $\overline{\mathcal{L}}(\mathbf{B};\theta)>\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta)+\eta$ for every $\mathbf{B}$ in the compact set $\{\mathbf{B};\|\mathbf{B}-\mathbf{B}^*(\theta)\|\geq\varepsilon\}\cap\mathcal{B}(\mathbf{B}^*(\theta),C)$ uniformly in θ in Θ[1] Then

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\widehat{\mathbf{B}}(\theta)-\mathbf{B}^*(\theta)\right\|\geq\varepsilon\right)\leq\mathbb{P}\left(\sup_{\theta\in\Theta}\left|\overline{\mathcal{L}}\left(\widehat{\mathbf{B}}(\theta);\theta\right)-\overline{\mathcal{L}}(\mathbf{B}^*(\theta);\theta)\right|\geq\eta\right).$$

The Lemma is then proved in view of (B.25). □

B.3 Proof of Proposition A.2

Recall that $\mathbf{H}^{-1}\mathbf{U}\left(X-x\right)=\mathbf{U}\left((X-x)/h\right)$ and define for all $\mathbf{b}$ in $\mathbb{R}^P$

$$\delta(\mathbf{b},\theta)=\delta(X;\mathbf{b},\theta)=\frac{1}{(nh^d)^{1/2}}\mathbf{U}\left(\frac{X-x}{h}\right)^{\top}\mathbf{b}. \tag{B.26}$$

We first need these two Lemmas which are established after the proof of the Proposition.

Lemma B.1 *Consider two real numbers t_β, $t_\varepsilon>0$ which may depend upon on n with $t_\beta+t_\varepsilon=o\left(n\underline{h}^d\right)^{1/2}$. Then under Assumptions K, L, R and X*

$$\sup_{(\theta,\varepsilon,\beta)\in\Theta\times\mathcal{B}(0,t_\varepsilon)\times\mathcal{B}(0,t_\beta)}\operatorname{Var}\left(\mathbf{R}^0(\beta,\varepsilon;\theta)\right)\leq C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)}{n\left(n\underline{h}^d\right)^{1/2}}.$$

The next Lemma studies coverings of the set

$$\mathcal{F}=\mathcal{F}(t_\beta,t_\varepsilon,\Theta)=\left\{\mathbf{R}^0\left(\beta,\varepsilon,\theta\right),(\beta,\varepsilon;\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta\right\}$$

with brackets $\left[\underline{R},\overline{R}\right]$. Recall that the bracket $\left[\underline{R},\overline{R}\right]=\left[\underline{R}(X,Y),\overline{R}(X,Y)\right]$ is the set of random variables $q=q(X,Y)$ such that $\underline{R}\leq q\leq\overline{R}$.

1. Indeed, if it was not true, it would be possible to find a sequence (B_n,θ_n) with $\|B_n\|\geq C'$ and θ_n in Θ such that $\overline{\mathcal{L}}(B_n,\theta_n)-\overline{\mathcal{L}}(B^*(\theta_n);\theta_n)=o(1)$. This would contradict the fact that $\overline{\mathcal{L}}(B_n,\theta_n)\geq\overline{\mathcal{L}}(B^*(\theta_n);\theta_n)+1$.

Lemma B.2 *Let $\mathcal{B}(0,t_\beta)$ and $\mathcal{B}(0,t_\varepsilon)$ be compact balls of $\mathbb{R}^P$ with radius t_β and t_ε that my depend upon n. Assume that $t_\varepsilon + t_\beta = o\left(n\underline{h}^d\right)^{1/2}$. Then under Assumptions K, L, R and X and and for $\underline{h}$ small enough*

i. There are some positive numbers $\overline{\sigma}^2$ and $\overline{w}$, with

$$\overline{\sigma}^2 \asymp \frac{t_\varepsilon^2(t_\varepsilon + t_\beta)}{n\left(n\underline{h}^d\right)^{1/2}} \quad \text{and} \quad \overline{w} \asymp \frac{t_\beta + t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}},$$

such that for any integer number $k \geq 2$, (β,ε) in $\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)$ and θ in Θ,

$$\mathbb{E}\left[\left|\mathbf{R}^0(\beta,\varepsilon;\theta)\right|^k\right] \leq \frac{k!}{2}\overline{w}^{k-2}\overline{\sigma}^2.$$

ii. Let τ in $(0,1)$ be a bracket length. There is a set of brackets

$$\mathcal{I}_\tau = \left\{[\underline{R}_j(X,Y),\overline{R}_j(X,Y)], 1 \leq j \leq e^{H(\tau)}\right\}$$

such that

- *$\mathcal{F}$ is a subset of $\bigcup_{j=1}^{e^{H(\tau)}}[\underline{R}_j(X,Y),\overline{R}_j(X,Y)]$.*
- *For any integer number $k \geq 2$ and all $1 \leq j \leq e^{H(\tau)}$*

$$\mathbb{E}\left[\left|\underline{R}_j(X,Y) - \overline{R}_j(X,Y)\right|^k\right] \leq \frac{k!}{2}\overline{w}^{k-2}\tau^2.$$

- *There is a constant C such that $H(\tau) \leq C\log\left((t_\beta + t_\varepsilon)n/\tau\right)$.*

We now return to the proof of the Proposition. For that write $\mathbb{X} = (X_1,\ldots,X_n)$. First note that by (A.9),

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

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

$$\mathbb{E}\left[\sup_{(\beta,\varepsilon;\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta}\left|\mathbb{R}_n^0(\beta,\varepsilon;\theta) - \mathbb{E}\left[\mathbb{R}_n^0(\beta,\varepsilon;\theta)\right]\right|\right] \leq C\left(\int_0^{\overline{\sigma}}\sqrt{nH(u)}du + (\overline{w}+\overline{\sigma})\,H(\overline{\sigma})\right).$$

Since $\overline{\sigma} < 1$, Lemma B.2 gives, for all u in $(0, \overline{\sigma}]$, that $H(u) \leq C\log((t_\beta + t_\varepsilon)n/u)$. This together with the Cauchy-Schwarz inequality gives

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

The order for $\overline{\sigma}$ given in Lemma B.2, assumptions on t_β and t_ε in the Proposition and Assumption K give $\log((t_\varepsilon + t_\beta)n/\overline{\sigma}) \leq C\log n$. Then, substituting and using the orders of $\overline{\sigma}$ and $\overline{w}$ gives

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

Thus the result follows since $(t_\beta + t_\varepsilon)^{1/2}/t_\varepsilon = O\left(\left(n\underline{h}^d\right)^{1/4}/\log^{1/2} n\right)$ as assumed in the Proposition. □

Proof of Lemma B.1 Observe that Assumption L-(i), (A.8) and (B.26) yield that

$$\mathbf{R}^0(\beta,\varepsilon;\theta) = K_h(X-x)\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}[r(Y-m^*(X;\theta)-t) - r(Y-m^*(X;\theta))]dt. \qquad \text{(B.27)}$$

The Cauchy-Schwarz inequality gives

$$\begin{aligned}\mathbf{R}^0(\beta,\varepsilon;\theta)^2 &= K_h^2(X-x)\left[\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}[r(Y-m^*(X;\theta)-t)-r(Y-m^*(X;\theta))]dt\right]^2\\ &\leq K_h^2(X-x)\,|\delta(\varepsilon,\theta)|\\ &\quad\times\left|\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}[r(Y-m^*(X;\theta)-t)-r(Y-m^*(X;\theta))]^2dt\right|.\end{aligned}$$

Then taking conditional expectation yields that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{R}^0(\beta,\varepsilon;\theta)^2\,|X\right]\\ &\quad\leq CK_h^2(X-x)\,|\delta(\varepsilon,\theta)|\left|\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}\mathbb{E}\left[(r(Y-m^*(X;\theta)-t)-r(Y-m^*(X;\theta)))^2|X\right]dt\right|.\end{aligned}$$

Then since $r(\cdot)=r_1(\cdot)-r_2(\cdot)$ by (3.2.1), the convex inequality and Assumption L-(ii) give

$$\mathbb{E}\left[\mathbf{R}^0(\beta,\varepsilon;\theta)^2\,|X\right]\leq CK_h^2(X-x)\delta(\varepsilon,\theta)^2\left(|\delta(\beta,\theta)|+|\delta(\varepsilon,\theta)|+[|\delta(\beta,\theta)|+|\delta(\varepsilon,\theta)|]^a\right),\quad\text{(B.28)}$$

where the last bound comes from the fact that $|t|\leq|\delta(\beta,\theta)|+|\delta(\varepsilon,\theta)|$ and where a is as in Assumption L-(ii). Recall definition (B.26) of $\delta(\mathbf{b},\theta)=\delta(X;\mathbf{b},\theta)$ and observe that

$$\begin{aligned}&\sup_{(\theta,\beta,\varepsilon)\in\Theta\times\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)}\sup_{x'\in\mathcal{X}}K\left(\frac{x'-x}{h}\right)(|\delta(x';\beta,\theta)|+|\delta(x';\varepsilon,\theta)|)\\ &\quad=\sup_{(\theta,\beta,\varepsilon)\in\Theta\times\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)}\sup_{x'\in\mathcal{X}}\frac{1}{(nh^d)^{1/2}}K\left(\frac{x'-x}{h}\right)\left(\left|\mathbf{U}\left(\frac{x'-x}{h}\right)^\top\beta\right|+\left|\mathbf{U}\left(\frac{x'-x}{h}\right)^\top\varepsilon\right|\right)\\ &\quad\leq\sup_{(x',\theta)\in\Theta\times\mathcal{X}}\frac{1}{(nh^d)^{1/2}}K\left(\frac{x'-x}{h}\right)\left\|\mathbf{U}\left(\frac{x'-x}{h}\right)\right\|\sup_{(\beta,\varepsilon)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)}(\|\beta\|+\|\varepsilon\|)\\ &\quad\leq\sup_{z\in\mathcal{K}}K(z)\,\|\mathbf{U}(z)\|\,\frac{t_\beta+t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}}\leq C\frac{t_\beta+t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}}.\end{aligned}\quad\text{(B.29)}$$

Now since $t_\beta+t_\varepsilon=o\left(n\underline{h}^d\right)^{1/2}$ and $a>1$, (B.28) yields that for n large enough

$$\mathbb{E}\left[\mathbf{R}^0(\beta,\varepsilon;\theta)^2\,|X\right]\leq CK^2\left(\frac{X-x}{h}\right)\delta(\varepsilon,\theta)^2\left(|\delta(\beta,\theta)|+|\delta(\varepsilon,\theta)|\right).$$

Assumptions K, X and (B.29) then yield that

$$\begin{aligned}\operatorname{Var}\left(\mathbf{R}^0(\beta,\varepsilon;\theta)\right) &\leq \mathbb{E}\left[\mathbf{R}^0\left(\beta,\varepsilon;\theta\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{R}^0\left(\beta,\varepsilon;\theta\right)^2|X\right]\right] \\ &\leq \frac{C}{n(nh^d)^{1/2}}\left\|f\left(\cdot\right)\right\|_\infty \int K\left(z\right)^2\left\|\mathbf{U}\left(z\right)\right\|^3 dz\left\|\varepsilon\right\|^2\left(\left\|\beta\right\|+\left\|\varepsilon\right\|\right). \quad \text{(B.30)}\end{aligned}$$

Then under Assumption K on the bandwidth, we have that uniformly in θ in Θ and (ε,β) in $\mathcal{B}(0,t_\varepsilon)\times\mathcal{B}(0,t_\beta)$,

$$\operatorname{Var}\left(\mathbf{R}^0(\beta,\varepsilon;\theta)\right) \leq C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)}{n\left(n\underline{h}^d\right)^{1/2}}.\square$$

Proof of Lemma B.2 *Proof of (i).* (B.30) gives that for any $(\beta,\varepsilon;\theta)$ in $\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta$,

$$\mathbb{E}\left[\mathbf{R}^0(\beta,\varepsilon;\theta)^2\right] \leq C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)}{n\left(n\underline{h}^d\right)^{1/2}}.$$

Observe also that (B.29) together with the fact that $t_\varepsilon+t_\beta=o\left(n\underline{h}^d\right)^{1/2}$ yield that

$$\sup_{(\beta,\varepsilon;\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta}\left|\mathbf{R}^0(\beta,\varepsilon;\theta)\right| \leq C\frac{t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}}.$$

Then for any integer number $k\geq 2$, and $(\beta,\varepsilon,\theta)$ in $\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta$,

$$\begin{aligned}\mathbb{E}\left[\left|\mathbf{R}^0(\beta,\varepsilon;\theta)\right|^k\right] &\leq \left(C\frac{t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}}\right)^{k-2}\mathbb{E}\left[\mathbf{R}^0(\beta,\varepsilon;\theta)^2\right] \leq \left(C\frac{t_\varepsilon}{\left(n\underline{h}^d\right)^{1/2}}\right)^{k-2}\left(C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)}{n\left(n\underline{h}^d\right)^{1/2}}\right) \\ &\leq \frac{k!}{2}\overline{w}^{k-2}\overline{\sigma}^2,\end{aligned}$$

with $\overline{\sigma}$ and $\overline{w}$ are as in the Lemma.

Proof of (ii). Define for all $\mathbf{b}$ in $\mathbb{R}^P$

$$\widetilde{\mathbf{R}}^0(\mathbf{b};\theta) = \widetilde{\mathbf{R}}^0(X,Y;\mathbf{b},\theta) = K_h(X-x)\int_0^{\delta(\mathbf{b},\theta)}\left[r(Y-m^*(X;\theta)-u)-r(Y-m^*(X;\theta))\right]du,$$

and observe that (B.27) and the fact that $\delta(\beta,\theta)+\delta(\varepsilon,\theta)=\delta(\beta+\varepsilon,\theta)$ by (B.26) give

$$\begin{aligned}\mathbf{R}^0(\beta,\varepsilon;\theta) &= K_h(X-x)\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}\left[r(Y-m^*(X;\theta)-u)-r(Y-m^*(X;\theta))\right]du \\ &= \widetilde{\mathbf{R}}^0(\beta+\varepsilon;\theta)-\widetilde{\mathbf{R}}^0(\beta;\theta). \quad \text{(B.31)}\end{aligned}$$

Let $\widetilde{\mathcal{F}}_{t_\beta}$ and $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$ be respectively the sets

$$\left\{\widetilde{\mathbf{R}}^0(\mathbf{b};\theta),(\mathbf{b},\theta)\in\mathcal{B}(0,t_\beta)\times\Theta\right\} \text{ and } \left\{\widetilde{\mathbf{R}}^0(\mathbf{b};\theta),(\mathbf{b},\theta)\in\mathcal{B}(0,t_\beta+t_\varepsilon)\times\Theta\right\}.$$

The proof is divided in three steps.

Step 1 : Coverings of $\mathcal{F}$, $\widetilde{\mathcal{F}}_{t_\beta}$, and $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$. Observe that definition (B.26) of $\delta(\mathbf{b},\theta)$ yields that $\mathbf{b}\mapsto\widetilde{\mathbf{R}}^0(\mathbf{b};\theta)$ is linear for all θ in Θ. Then since $t_\varepsilon>0$, the cardinality of the sets of τ-brackets covering $\widetilde{\mathcal{F}}_{t_\beta}$ is less than the one needed to cover $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$. Suppose then that we have constructed sets

$$\widetilde{\mathcal{I}}_\tau(t_\beta+t_\varepsilon)=\bigcup_{j\le e^{H(\tau)/2}}[\underline{\mathbf{R}}_{1,j}(X,Y),\overline{\mathbf{R}}_{1,j}(X,Y)] \text{ and } \widetilde{\mathcal{I}}_\tau(t_\beta)=\bigcup_{j\le e^{H(\tau)/2}}[\underline{\mathbf{R}}_{2,j}(X,Y),\overline{\mathbf{R}}_{2,j}(X,Y)],$$

of brackets covering respectively $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$ and $\widetilde{\mathcal{F}}_{t_\beta}$ where $H(\cdot)$ is as in the Lemma. Suppose furthermore that for all integer numbers $i=1,2$, $k\ge2$ and $j\le e^{H(\tau)/2}$,

$$\mathbb{E}\left[|\underline{\mathbf{R}}_{i,j}(X,Y)-\overline{\mathbf{R}}_{i,j}(X,Y)|^k\right]\le\frac{k!}{4}\tau^2\left(\frac{\overline{w}}{2}\right)^{k-2}. \tag{B.32}$$

Then, in view of (B.31), for any $\mathbf{R}^0$ in $\mathcal{F}$, there exists $j_1\le e^{H(\tau)/2}$ and $j_2\le e^{H(\tau)/2}$ such that

$$\underline{\mathbf{R}}_{1,j_1}(X,Y)-\overline{\mathbf{R}}_{2,j_2}(X,Y)\le\mathbf{R}^0\le\overline{\mathbf{R}}_{1,j_1}(X,Y)-\underline{\mathbf{R}}_{2,j_2}(X,Y).$$

Then the covering $\mathcal{I}_\tau$ of $\mathcal{F}$ is a subset of

$$\bigcup_{j_1\le e^{H(\tau)/2}}\bigcup_{j_1\le e^{H(\tau)/2}}\left[\underline{\mathbf{R}}_{1,j_1}(X,Y)-\overline{\mathbf{R}}_{2,j_2}(X,Y);\overline{\mathbf{R}}_{1,j_1}(X,Y)-\underline{\mathbf{R}}_{2,j_2}(X,Y)\right].$$

Observe that the number of brackets of $\mathcal{I}(\tau)$ is then less than $e^{H(\tau)}$ as requested in the Lemma. Note also that since $k\ge2$, the convex inequality yields that

$$\begin{aligned}&\mathbb{E}\left[\left|\overline{\mathbf{R}}_{1,j_1}(X,Y)-\underline{\mathbf{R}}_{2,j_2}(X,Y)-\left(\underline{\mathbf{R}}_{1,j_1}(X,Y)-\overline{\mathbf{R}}_{2,j_2}(X,Y)\right)\right|^k\right]\\ &\quad\le\ 2^{k-1}\left(\mathbb{E}\left[\left|\overline{\mathbf{R}}_{1,j_1}(X,Y)-\underline{\mathbf{R}}_{1,j_1}(X,Y)\right|^k\right]+\mathbb{E}\left[\left|\overline{\mathbf{R}}_{2,j_2}(X,Y)-\underline{\mathbf{R}}_{2,j_2}(X,Y)\right|^k\right]\right)\le\frac{k!}{2}\tau^2\overline{w}^{k-2},\end{aligned}$$

as requested in the Lemma. Finally observe that $\widetilde{\mathcal{F}}_{t_\beta}$ is a subset of $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$, so that it is sufficient to construct a set of brackets covering $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$ and satisfying the conditions above to achieve the proof. We now construct the set of brackets $\widetilde{\mathcal{I}}_\tau(t_\beta+t_\varepsilon)$ covering $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$.

Step 2: Preliminary results for the construction of brackets covering $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$. Consider $\eta>0$ and set $t=t_\beta+t_\varepsilon$. We begin by constructing brackets for the sets of functions

$$\{m^*(X;\theta),\theta\in\Theta\}\,,\ \{K_h(X-x),\theta\in\Theta\}\ \text{and}\ \{\delta(\mathbf{b},\theta),\mathbf{b}\in\mathcal{B}(0,t),\theta\in\Theta\}\,,$$

to get brackets for $\widetilde{\mathcal{F}}_t$. Observe first that there exists a covering of $\mathcal{B}(0,t)\times\Theta$ with N balls $\mathcal{B}\left((\mathbf{b}_j,\theta_j),\eta\right)$, with centers $(\mathbf{b}_j,\theta_j)$ and radius η, such that

$$N\leq\max\left(1,\frac{Ct^P}{\eta^{P+d+2}}\right), \tag{B.33}$$

see van de Geer (1999, p.20). For the rest of this part of the proof, we consider any $(\mathbf{b},\theta)$ in $\mathcal{B}((\mathbf{b}_j,\theta_j),\eta)$ for some integer number j in $[1,N]$. Lemma A.1-(iii) gives that

$$|m^*(X;\theta)|\leq\sup_{\theta\in\Theta}\sup_{z\in\mathcal{X}}|m^*(z;\theta)|\leq\sup_{\theta\in\Theta}\sup_{z\in\mathcal{X}}\left|\mathbf{U}(z-x)^\top(\mathbf{b}^*(\theta)-\mathbf{m}_x)\right|+\sup_{x\in\mathcal{X}_0}|m(x)|=m^*_\infty<\infty,$$

and Lemma A.1-(iv) gives that $|m^*(X;\theta)-m^*(X;\theta_j)|\leq C\underline{h}^{-(p+1)}\|\theta-\theta_j\|\leq C\underline{h}^{-(p+1)}\eta$. Now define $\underline{m}_j=\max(-m^*_\infty,m^*(X;\theta_j)-C\underline{h}^{-(p+1)}\eta)$ and $\overline{m}_j=\min(m^*_\infty,m^*(X;\theta_j)+C\underline{h}^{-(p+1)}\eta)$. Then

$$m^*(X;\theta)\leq m^*(X;\theta_j)+|m^*(X;\theta)-m^*(X;\theta_j)|\leq m^*(X;\theta_j)+C\underline{h}^{-(p+1)}\eta,$$

so that $m^*(X;\theta)\leq m^*_\infty$ implies $m^*(X;\theta)\leq\overline{m}_j$. Arguing symmetrically gives $m^*(X;\theta)\geq\underline{m}_j$. Hence by construction of the θ_j's, $\{m^*(X;\theta),\theta\in\Theta\}$ is a subset of $\bigcup_{j\leq N}[\underline{m}_j,\overline{m}_j]$. Observe also that

$$\max_{1\leq j\leq N}\sup_{z\in\mathcal{X}}|\overline{m}_j-\underline{m}_j|\leq C\underline{h}^{-(p+1)}\eta.$$

We now consider consider $K_h(X-x)$. Assumption K gives for $\underline{h}\leq1$,

$$\left|K\left(\frac{X-x}{h}\right)-K\left(\frac{X-x_j}{h_j}\right)\right|\ \leq\ L\left\|\frac{X-x}{h}-\frac{X-x_j}{h_j}\right\|$$

$$\begin{aligned} &\leq L\left\|\frac{x-x_j}{h_j}\right\| + L\left\|X-x\right\|\left|\frac{1}{h}-\frac{1}{h_j}\right| \\ &\leq C\left(\frac{1}{\underline{h}}\left\|x-x_j\right\| + \frac{1}{\underline{h}^2}\left|h-h_j\right|\right) \leq \frac{C}{\underline{h}^2}\left\|\theta-\theta_j\right\| \leq \frac{C}{\underline{h}^2}\eta. \end{aligned}$$

Define $\underline{K}_j = \max(0, K((X-x_j)/h_j) - C\eta/\underline{h}^2)$ and $\overline{K}_j = \min(||K||_\infty, K((X-x_j)/h_j) + C\eta/\underline{h}^2)$. Then

$$K_h\left(X-x\right) \leq K_{h_j}\left(X-x_j\right) + \left|K_h\left(X-x\right) - K_{h_j}\left(X-x_j\right)\right| \leq K_{h_j}\left(X-x_j\right) + C\underline{h}^{-2}\eta,$$

so that $K_h(X-x) \leq \|K\|_\infty$ implies $K_h(X-x) \leq \overline{K}_j$. Arguing symmetrically gives $K(X-x) \geq \underline{K}_j \geq 0$. Hence by construction of the θ_j's, $\{K_h(X-x), \theta \in \Theta\}$ is a subset of $\bigcup_{j\leq N}[\underline{K}_j, \overline{K}_j]$. Observe also that

$$\max_{1\leq j\leq N}\sup_{z\in\mathcal{X}}|\overline{K}_j - \underline{K}_j| \leq C\underline{h}^{-2}\eta.$$

We now consider the set of functions $\{\delta(\mathbf{b},\theta), \theta \in \Theta, \mathbf{b} \in \mathcal{B}(0,t)\}$. Recall that

$$\delta(\mathbf{b},\theta) = \frac{1}{(nh^d)^{1/2}}\mathbf{U}\left(\frac{X-x}{h}\right)^\top \mathbf{b},$$

so that, under Assumption K and X, we have

$$\sup_{\theta\in\Theta}\sup_{\mathbf{b}\in\mathcal{B}(0,t)}|\delta(\mathbf{b},\theta)| = \|\delta(\mathbf{b},\theta)\|_\infty = C\frac{t}{\underline{h}^p\left(n\underline{h}^d\right)^{1/2}}.$$

Define $\mathbf{U}_j = \mathbf{U}(X-x_j)$, $\mathbf{U} = \mathbf{U}(X-x)$, $\mathbf{H}_j^{-1} = \mathbf{H}(h_j)^{-1}$, $\delta = \delta(\mathbf{b},\theta)$ and $\delta_j = \delta(\mathbf{b}_j,\theta_j)$ so that

$$\begin{aligned} |\delta-\delta_j| &= \left|\mathbf{U}^\top\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}}\left(\mathbf{b}_j-\mathbf{b}\right) + \left(\mathbf{U}_j-\mathbf{U}\right)^\top\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}}\mathbf{b}_j + \mathbf{U}_j^\top\left(\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} - \frac{\mathbf{H}_j^{-1}}{(nh_j^d)^{1/2}}\right)\mathbf{b}_j\right| \\ &\leq \sup_{(z,z')\in\mathcal{X}^2}\left\|\mathbf{U}\left(z-z'\right)\right\|\left\|\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}}\right\|\left\|\mathbf{b}-\mathbf{b}_j\right\| \\ &\quad + \sup_{(z,z')\in\mathcal{X}^2}\left\|\mathbf{U}\left(x-z\right) - \mathbf{U}\left(x_j-z'\right)\right\|\left\|\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}}\right\|\left\|\mathbf{b}_j\right\| \\ &\quad + \sup_{(z,z')\in\mathcal{X}^2}\left\|\mathbf{U}\left(z-z'\right)\right\|\left\|\mathbf{b}_j\right\|\left\|\frac{\mathbf{H}^{-1}}{(nh^d)^{1/2}} - \frac{\mathbf{H}_j^{-1}}{(nh_j^d)^{1/2}}\right\| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{C(1+t)}{\underline{h}^{p}\left(n\underline{h}^{d}\right)^{1/2}}\left(\|\mathbf{b}-\mathbf{b}_j\|+\|x-x_j\|+\frac{1}{\underline{h}}\left|h-h_j\right|\right)\\
&\leq \frac{C(1+t)}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\left(\|\mathbf{b}-\mathbf{b}_j\|+\|\theta-\theta_j\|\right)\leq C\frac{1+t}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\eta,
\end{aligned}$$

provided $\underline{h}$ is less than 1. Define for $j \leq N$

$$\begin{aligned}
\underline{\delta}_j &= \max\left(-\|\delta(\mathbf{b},\theta)\|_\infty, \delta_j - C\frac{1+t}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\eta\right)\\
\overline{\delta}_j &= \min\left(\|\delta(\mathbf{b},\theta)\|_\infty, \delta_j + C\frac{1+t}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\eta\right).
\end{aligned}$$

Then

$$\delta\left(\mathbf{b},\theta\right)\leq\delta\left(\mathbf{b}_j,\theta_j\right)+\left|\delta\left(\mathbf{b},\theta\right)-\delta\left(\mathbf{b}_j,\theta_j\right)\right|\leq\delta_j+C\frac{(1+t)}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\eta,$$

so that $\delta\left(\mathbf{b},\theta\right)\leq\|\delta\left(\mathbf{b},\theta\right)\|_\infty$ implies $\delta\left(\mathbf{b},\theta\right)\leq\overline{\delta}_j$. Arguing symmetrically gives $\delta\left(\mathbf{b},\theta\right)\geq\underline{\delta}_j$. Hence by construction of the θ_j's, $\{\delta\left(\mathbf{b},\theta\right),(\mathbf{b},\theta)\in\mathcal{B}(0,t)\times\Theta\}$ is a subset of $\bigcup_{j\leq N}\left[\underline{\delta}_j,\overline{\delta}_j\right]$. Observe also that

$$\max_{1\leq j\leq N}\sup_{z\in\mathcal{X}}|\overline{\delta}_j-\underline{\delta}_j|\leq C\frac{1+t}{\underline{h}^{p+1}\left(n\underline{h}^{d}\right)^{1/2}}\eta. \tag{B.34}$$

Step 3 : Construction of brackets covering $\widetilde{\mathcal{F}}_{t_\beta+t_\varepsilon}$. For the rest of the proof, $(\mathbf{b},\theta)$ is an element of $\mathcal{B}(0,t)\times\Theta$ and the integer number $1\leq j\leq N$ is chosen such that $\|(\mathbf{b},\theta)-(\mathbf{b}_j,\theta_j)\|\leq\eta$. For such choice, Step 2 gives

$$\underline{m}_j\leq m^*(X;\theta)\leq\overline{m}_j,\ 0\leq\underline{K}_j\leq K_h(X-x)\leq\overline{K}_j\ \text{ and }\ \underline{\delta}_j\leq\delta(\mathbf{b},\theta)\leq\overline{\delta}_j.$$

Now recall that (B.27) gives

$$\widetilde{\mathbf{R}}^0(\mathbf{b};\theta)=K_h\left(X-x\right)\left(\int_0^{\delta(\mathbf{b},\theta)}\left[r\left(Y-m^*\left(X;\theta\right)-u\right)-r\left(Y-m^*\left(X;\theta\right)\right)\right]du\right).$$

Assumption L-(i) yields that $r(\cdot)$ is bounded in variation so that there exists $r_1(\cdot)$ and $r_2(\cdot)$, two non-decreasing functions such that $r(\cdot)=r_1(\cdot)-r_2(\cdot)$. Then

$$\widetilde{\mathbf{R}}^0(\mathbf{b};\theta) \;=\; K_h\left(X-x\right)\left(\int_0^{\delta(\mathbf{b},\theta)}\left[r_1(Y-m^*(X;\theta)-u)-r_1(Y-m^*(X;\theta))\right]du\right)$$

$$
\begin{aligned}
&+K_h\left(X-x\right)\left(\int_0^{\delta(\mathbf{b},\theta)}\left[r_2(Y-m^*(X;\theta))-r_2(Y-m^*(X;\theta)-u)\right]du\right)\\
=\;& \widetilde{R}_1(\mathbf{b};\theta)+\widetilde{R}_2(\mathbf{b};\theta).
\end{aligned}
$$

Because the study of the two integrals above is quite similar, we will construct brackets only for $\widetilde{R}_2(\mathbf{b};\theta)$. We will furthermore distinguish three cases : $\underline{\delta}_j \geq 0$, $\underline{\delta}_j < 0 < \overline{\delta}_j$ and $\overline{\delta}_j \leq 0$. Since the cases $\underline{\delta}_j \geq 0$ and $\overline{\delta}_j \leq 0$ are symmetric, we will focus on the cases $\underline{\delta}_j \geq 0$ and $\underline{\delta}_j < 0 < \overline{\delta}_j$.

Case 1 : $\underline{\delta}_j \geq 0$. First observe that the function

$$
u \in \mathbb{R}_+ \mapsto \mathbf{r}_2(u) = r_2(Y-m^*(X;\theta)) - r_2(Y-m^*(X;\theta)-u)
$$

is a non-decreasing and non-negative. We first bound this function by two functions over $\mathbb{R}_+$ since we have supposed $\underline{\delta}_j \geq 0$. Observe that construction in Step 2 yields

$$
\underline{\mathbf{r}}_{2,j}(u) = r_2\left(Y-\overline{m}_j\right) - r_2\left(Y-\underline{m}_j-u\right) \leq \mathbf{r}_2(u) \leq r_2\left(Y-\underline{m}_j\right) - r_2\left(Y-\overline{m}_j-u\right) = \overline{\mathbf{r}}_{2,j}(u),
$$

for all u in $\mathbb{R}_+$, since $u \in \mathbb{R}_+ \mapsto \mathbf{r}_2(u)$ is non-decreasing. Now since $0 \leq \underline{K}_j \leq K_h(X-x) \leq \overline{K}_j$, we have

$$
\underline{K}_j \underline{\mathbf{r}}_{2,j}(u) \leq K_h(X-x)\mathbf{r}_2(u) \leq \overline{K}_j \overline{\mathbf{r}}_{2,j}(u),
$$

for all u in $\mathbb{R}_+$ since $u \in \mathbb{R}_+ \mapsto \mathbf{r}_2(u)$ is non-negative. Now because $0 \leq \underline{\delta}_j \leq \delta(\mathbf{b},\theta) \leq \overline{\delta}_j$ we have

$$
\underline{\mathcal{R}}'_j(Y,X) = \underline{K}_j \int_0^{\underline{\delta}_j} \underline{\mathbf{r}}_{2,j}(u)du \leq \widetilde{R}_2(\mathbf{b},\theta) \leq \overline{K}_j \int_0^{\overline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du = \overline{\mathcal{R}}'_j(Y,X). \tag{B.35}
$$

Define $\underline{\mathcal{R}}_j(X,Y) = \max(-\overline{w}/4, \underline{\mathcal{R}}'_j(Y,X))$, $\overline{\mathcal{R}}_j(X,Y) = \min(\overline{w}/4, \overline{\mathcal{R}}'_j(Y,X))$ and where $\overline{w}$ is as in the Lemma. Then $\underline{\mathcal{R}}_j(Y,X) \leq \widetilde{R}_2(\mathbf{b};\theta) \leq \overline{\mathcal{R}}_j(Y,X)$, $-\overline{w}/4 \leq \underline{\mathcal{R}}_j(X,Y) \leq \overline{\mathcal{R}}_j(X,Y) \leq \overline{w}/4$ and

$$
\max_{j\leq N}\left|\overline{\mathcal{R}}_j(X,Y) - \underline{\mathcal{R}}_j(X,Y)\right| \leq \frac{\overline{w}}{2}. \tag{B.36}
$$

We now study the L^2-length of the brackets $\left[\underline{\mathcal{R}}_j(Y,X), \overline{\mathcal{R}}_j(Y,X)\right]$. For that observe that

$$
\overline{\mathcal{R}}_j(Y,X) - \underline{\mathcal{R}}_j(Y,X) \leq \overline{\mathcal{R}}'_j(Y,X) - \underline{\mathcal{R}}'_j(Y,X) = \left(\overline{K}_j \int_0^{\overline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du - \underline{K}_j \int_0^{\underline{\delta}_j} \underline{\mathbf{r}}_{2,j}(u)du\right).
$$

Convex and Cauchy-Schwarz inequalities yield that

$$\begin{aligned}
&\mathbb{E}\left[\overline{K}_j \int_0^{\overline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du - \underline{K}_j \int_0^{\underline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du\right]^2 \\
&= \mathbb{E}\left[\left(\overline{K}_j - \underline{K}_j\right) \int_0^{\overline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du + \underline{K}_j \int_{\overline{\delta}_j}^{\underline{\delta}_j} \overline{\mathbf{r}}_{2,j}(u)du + \underline{K}_j \int_0^{\underline{\delta}_j} \left(\overline{\mathbf{r}}_{2,j}(u) - \underline{\mathbf{r}}_{2,j}(u)\right) du\right]^2 \\
&\leq C\mathbb{E}\left[\left|\underline{K}_j - \overline{K}_j\right|^2 \overline{\delta}_j \int_0^{\overline{\delta}_j} \mathbb{E}\left[\overline{\mathbf{r}}_{2,j}(u)^2|X\right] du\right] + C\mathbb{E}\left[\underline{K}_j^2 \left|\overline{\delta}_j - \underline{\delta}_j\right| \int_{\overline{\delta}_j}^{\underline{\delta}_j} \mathbb{E}\left[\overline{\mathbf{r}}_{2,j}(u)^2|X\right] du\right] \\
&\quad + C\mathbb{E}\left[\underline{K}_j^2 \underline{\delta}_j \int_0^{\underline{\delta}_j} \mathbb{E}\left[\left(\overline{\mathbf{r}}_{2,j}(u) - \underline{\mathbf{r}}_{2,j}(u)\right)^2 |X\right] du\right] \\
&= V_1 + V_2 + V_3.
\end{aligned}$$

We now study the V_i's, $i = 1, 2, 3$. We only bound V_1 since the bounds for V_2 and V_3 are similarly obtained. Observe that Step 2 yields that

$$\begin{aligned}
V_1 &\leq \sup_{z \in \mathcal{X}} \left|\overline{K}_j(z) - \underline{K}_j(z)\right|^2 \times \|\delta(\mathbf{b}, \theta)\|_\infty \times \mathbb{E}\left[\int_0^{\overline{\delta}_j} \mathbb{E}\left[\overline{\mathbf{r}}_{2,j}(u)^2|X\right] du\right] \\
&\leq C\left(\frac{\eta}{\underline{h}^2}\right)^2 \times \frac{t}{\underline{h}^p (n\underline{h}^d)^{1/2}} \times \mathbb{E}\left[\int_0^{\overline{\delta}_j} \mathbb{E}\left[\overline{\mathbf{r}}_{2,j}(u)^2|X\right] du\right]. \qquad \text{(B.37)}
\end{aligned}$$

Now observe that Assumption L-(ii) and the convex inequality yield that for all $u \geq 0$,

$$\begin{aligned}
\mathbb{E}\left[\overline{\mathbf{r}}_{2,j}(u)^2|X\right] &= \mathbb{E}\left[\left(r_2\left(Y - \underline{m}_j\right) - r_2\left(Y - \overline{m}_j - u\right)\right)^2 |X\right] \\
&\leq \mathbb{E}\left[\left(r_2(Y - \underline{m}_j - (\overline{m}_j - \underline{m}_j) - u) - r_2(Y - \underline{m}_j)\right)^2 |X\right] \\
&\leq C\left(\max\left(\left|\overline{m}_j - \underline{m}_j\right|, \left|\overline{m}_j - \underline{m}_j\right|^a\right) + \max\left(|u|, |u|^a\right)\right).
\end{aligned}$$

This, Step 2 and (B.37) then yield that

$$V_1 \leq C\eta^2 \frac{t^2}{\underline{h}^{2p+4}\left(n\underline{h}^d\right)} \left[\max\left(\frac{\eta}{\underline{h}^{p+1}}, \left(\frac{\eta}{\underline{h}^{p+1}}\right)^a\right) + \max\left(\frac{t}{\underline{h}^p\left(n\underline{h}^d\right)^{1/2}}, \left(\frac{t}{\underline{h}^p\left(n\underline{h}^d\right)^{1/2}}\right)^a\right)\right].$$

Arguing similarly for V_2 and V_3 yields that there is positive constants D_1, D_2 and D_3 not depending upon n or η such that

$$\mathbb{E}\left[\left(\overline{\mathcal{R}}_j(X, Y) - \underline{\mathcal{R}}_j(X, Y)\right)^2\right] \leq C\frac{(1+t)^{D_1}}{\left(n\underline{h}^d\right)^{D_2} \underline{h}^{D_3}} \max\left(\eta, \eta^3, \eta^{2+a}\right).$$

This and (B.36) give,

$$\begin{aligned}\mathbb{E}\left[\left|\underline{\mathcal{R}}_{1,j}(X,Y)-\overline{\mathcal{R}}_{1,j}(X,Y)\right|^k\right] &\le \mathbb{E}\left[\left(\overline{\mathcal{R}}_{1,j}(X,Y)-\underline{\mathcal{R}}_{1,j}(X,Y)\right)^2\right]\left(\frac{\overline{w}}{2}\right)^{k-2}\\ &\le Ck!\frac{(1+t)^{D_1}}{\left(n\underline{h}^d\right)^{D_2}\underline{h}^{D_3}}\max\left(\eta,\eta^3,\eta^{2+a}\right)\left(\frac{\overline{w}}{2}\right)^{k-2}.\end{aligned}$$

Case 2 : $\underline{\delta}_j < 0 < \overline{\delta}_j$. Define

$$\begin{aligned}\overline{\mathbf{r}}_{2,j}(u) &= \left(r_2\left(Y-\underline{m}_j\right)-r_2\left(Y-\overline{m}_j-u\right)\right)\mathbb{I}\left(u\ge\underline{m}_j-\overline{m}_j\right),\\ \underline{\mathbf{r}}_{2,j}(u) &= \left(r_2\left(Y-\overline{m}_j\right)-r_2\left(Y-\underline{m}_j-u\right)\right)\mathbb{I}\left(u\le\overline{m}_j-\underline{m}_j\right).\end{aligned}$$

These two functions satisfy

$$\underline{\mathbf{r}}_{2,j}(u)\le\mathbf{r}_2(u)\le\overline{\mathbf{r}}_{2,j}(u),\ \text{for all } u \text{ in } \mathbb{R},$$

and furthermore, $\underline{\mathbf{r}}_{2,j}(\cdot)\le 0$ and $\overline{\mathbf{r}}_{2,j}(\cdot)\ge 0$. Then by construction we have that

$$\begin{aligned}\underline{\mathcal{R}}'_j(Y,X) &= \overline{K}_j\int_0^{\underline{\delta}_j}\overline{\mathbf{r}}_{2,j}(u)du+\overline{K}_j\int_0^{\overline{\delta}_j}\underline{\mathbf{r}}_{2,j}(u)du\le\widetilde{R}_2(\mathbf{b};\theta)\\ &\le \overline{K}_j\int_0^{\overline{\delta}_j}\overline{\mathbf{r}}_{2,j}(u)du+\overline{K}_j\int_0^{\underline{\delta}_j}\underline{\mathbf{r}}_{2,j}(u)du=\overline{\mathcal{R}}'_j(Y,X).\end{aligned}$$

Then construction (B.35) of the brackets gives

$$\underline{\mathcal{R}}_j(Y,X)\le\widetilde{R}_2(\mathbf{b};\theta)\le\overline{\mathcal{R}}_j(Y,X),$$

where $\underline{\mathcal{R}}_j(X,Y)=\max(-\overline{w}/4,\underline{\mathcal{R}}'_j(Y,X))$, $\overline{\mathcal{R}}_j(X,Y)=\min(\overline{w}/4,\overline{\mathcal{R}}'_j(Y,X))$ and where $\overline{w}$ is as in the Lemma. This gives that $-\overline{w}/4\le\underline{\mathcal{R}}_j(X,Y)\le\overline{\mathcal{R}}_j(X,Y)\le\overline{w}/4$ and

$$\max_{j\le N}\left|\overline{\mathcal{R}}_j(X,Y)-\underline{\mathcal{R}}_j(X,Y)\right|\le\frac{\overline{w}}{2}. \tag{B.38}$$

The study of the L^2-length of the brackets $\left[\underline{\mathcal{R}}_j(Y,X),\overline{\mathcal{R}}_j(Y,X)\right]$ is similar as the on made in the first case so we have

$$\mathbb{E}\left[\left(\overline{\mathcal{R}}_j(X,Y)-\underline{\mathcal{R}}_j(X,Y)\right)^2\right]\le C\frac{(1+t)^{D_1}}{\left(n\underline{h}^d\right)^{D_2}\underline{h}^{D_3}}\max\left(\eta,\eta^3,\eta^{2+a}\right).$$

This and (B.38) give

$$\begin{aligned}\mathbb{E}\left[\left|\underline{\mathcal{R}}_{1,j}(X,Y)-\overline{\mathcal{R}}_{1,j}(X,Y)\right|^{k}\right] &\leq \mathbb{E}\left[\overline{\mathcal{R}}_{1,j}(X,Y)-\underline{\mathcal{R}}_{1,j}(X,Y)\right]^{2}\left(\frac{\overline{w}}{2}\right)^{k-2} \\ &\leq Ck!\frac{(1+t)^{D_1}}{\left(n\underline{h}^{d}\right)^{D_2}\underline{h}^{D_3}}\max\left(\eta,\eta^{3},\eta^{2+a}\right)\left(\frac{\overline{w}}{2}\right)^{k-2}.\end{aligned}$$

We now return to the construction of the brackets for the set $\left\{\widetilde{\mathbf{R}}^{0}(\mathbf{b},\theta),\theta\in\Theta\right\}$. Cases 1 and 2 yield that there is a collection of brackets $\left[\underline{\mathbf{R}}_{j}(X,Y),\overline{\mathbf{R}}_{j}(X,Y)\right]$, $0\leq j\leq N$, such that $\mathcal{F}_t$ is a subset of $\bigcup_{j\leq N}\left[\underline{\mathbf{R}}_{j}(X,Y),\overline{\mathbf{R}}_{j}(X,Y)\right]$ and satisfying

$$\max_{0\leq j\leq N}\mathbb{E}\left[\left|\underline{\mathbf{R}}_{j}(X,Y)-\overline{\mathbf{R}}_{j}(X,Y)\right|^{k}\right]\leq Ck!\frac{(1+t)^{D_1}}{\left(n\underline{h}^{d}\right)^{D_2}\underline{h}^{D_3}}\max\left(\eta,\eta^{3},\eta^{2+a}\right)\left(\frac{\overline{w}}{2}\right)^{k-2}.$$

This upper bound is smaller than $k!\tau^{2}(\overline{w}/2)^{k-2}/4$, as requested in (B.32), if

$$\eta\leq C\min\left(\frac{\underline{h}^{D_3}\left(n\underline{h}^{d}\right)^{D_2}}{(1+t)^{D_1}}\tau^{2},\left(\frac{\underline{h}^{D_3}\left(n\underline{h}^{d}\right)^{D_2}}{(1+t)^{D_1}}\right)^{1/3}\tau^{2/3},\left(\frac{\underline{h}^{D_3}\left(n\underline{h}^{d}\right)^{D_2}}{(1+t)^{D_1}}\right)^{1/(2+a)}\tau^{2/(2+a)}\right).$$

Recall now that $\tau<1$ and that Assumption K gives that $n\underline{h}^{d}$ is bounded away from 0. This gives, from (B.33), that there is C such that the number of balls

$$e^{H(\tau)/2}\leq\max\left(1,Ct^{P}\left(\frac{(1+t)^{D_1}}{\left(n\underline{h}^{d}\right)^{D_2}\underline{h}^{D_3}}\min\left(\tau^{2},\tau^{2/3},\tau^{2/(2+a)}\right)\right)^{-(P+d+2)}\right).$$

Then

$$H(\tau)\leq C\log\left(\max\left(1,\frac{Cn^{C}(1+t)^{P+d+2}}{\tau^{(P+d+2)\min(2,2/3,2/(2+a))}}\right)\right)=C\max\left(0,\log\left(\frac{Cn^{C}t^{P/2(P+d+2)}(1+t)}{\tau}\right)\right).$$

And we have

$$\log\left(\frac{Cn^{C}t^{P/2(P+d+2)}(1+t)}{\tau}\right)\leq C\log\left(\frac{(tn)^{C}}{\tau}\right)\leq C\log\left(\frac{tn}{\tau}\right),$$

which gives the bound of $H(\tau)$ stated in the Lemma. □

B.4 Proof of Proposition A.3

The integral expression (B.27) of $\mathbf{R}^0(\beta,\varepsilon;\theta)$ and definition (B.26) of $\delta(\mathbf{b};\theta)$ give

$$\begin{aligned}\mathbf{R}^2(\beta,\varepsilon;\theta) &= K_h(X-x)\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}\left[R^{(1)}(m^*(X;\theta)+u|X)-R^{(1)}(m^*(X;\theta)|X)\right]du\\ &\quad -\frac{1}{2nh^d}\varepsilon^\top\mathbf{J}(X;\theta)(\varepsilon+2\beta).\end{aligned}$$

Then the definition (3.3.6) of $\mathbf{J}(X;\theta)$ gives

$$\mathbf{R}^2(\beta,\varepsilon;\theta)=K_h(X-x)\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}u\int_0^1\left[R^{(2)}(m^*(X;\theta)+vu|X)-R^{(2)}(m^*(X;\theta)|X)\right]dvdu,$$

by Assumption R-(iii) which gives continuous differentiability of $R^{(1)}(\cdot|\cdot)$. Observe that for all x in $\mathcal{X}$, the map $t\mapsto R^{(2)}(t|x)$ is continuously differentiable under Assumption R-(iii) so that it satisfies the Lipschitz condition over any compacts. Then Lemma A.1-(v) and (B.29) together with the fact $t_\beta+t_\varepsilon=o\left(n\underline{h}^d\right)^{1/2}$ yield that

$$\left|\mathbf{R}^2(\beta,\varepsilon;\theta)\right|\le CK_h(X-x)\left|\int_{\delta(\beta,\theta)}^{\delta(\beta,\theta)+\delta(\varepsilon,\theta)}u^2du\right|\le CK_h(X-x)|\delta(\varepsilon,\theta)|(|\delta(\beta,\theta)|+|\delta(\varepsilon,\theta)|)^2,$$

since $|u|\le|\delta(\varepsilon,\theta)|+|\delta(\beta,\theta)|$. Then arguing as in (B.29) yields that

$$\sup_{(\beta,\varepsilon;\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta}\left|\mathbf{R}^2(\beta,\varepsilon;\theta)\right|\le C\frac{t_\varepsilon(t_\beta+t_\varepsilon)^2}{\left(n\underline{h}^d\right)^{3/2}}=\overline{w}_2.$$

We now bound the variance of $\mathbf{R}^2(\beta,\varepsilon;\theta)$. Observe that the Cauchy-Schwarz inequality yields that

$$\begin{aligned}\operatorname{Var}\left(\mathbf{R}^2(\beta,\varepsilon;\theta)\right) &\le \operatorname{Var}\left(K_h(X-x)|\delta(\varepsilon,\theta)|(\delta(\beta,\theta)^2+\delta(\varepsilon,\theta)^2)\right)\\ &\le 2\mathbb{E}\left[K_h(X-x)^2\delta(\varepsilon,\theta)^2\left(|\delta(\beta,\theta)|^4+|\delta(\varepsilon,\theta)|^4\right)\right]\\ &\le \frac{1}{n(nh^d)^2}\int K(u)^2\|\mathbf{U}(u)\|^2\|\varepsilon\|^2\left(\|\mathbf{U}(u)\|^4\|\beta\|^4+\|\mathbf{U}(u)\|^4\|\varepsilon\|^4\right)\\ &\quad\times f(x+hu)du\\ &\le C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)^4}{n\left(n\underline{h}^d\right)^2}. \end{aligned}\tag{B.39}$$

Then there is a $\overline{\sigma}_2$ such that

$$\sup_{\theta\in\Theta}\sup_{(\beta,\varepsilon)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)}\operatorname{Var}\left(\mathbf{R}^2(\beta,\varepsilon;\theta)\right)\leq C\frac{t_\varepsilon^2(t_\beta+t_\varepsilon)^4}{n\left(n\underline{h}^d\right)^2}=\overline{\sigma}_2^2.$$

This gives that for any integer number k,

$$\mathbb{E}\left[\left|K_h(X-x)|\delta(\varepsilon,\theta)|(|\delta(\beta,\theta)|^2+|\delta(\varepsilon,\theta)|^2)\right|^k\right]\leq\frac{k!}{2}\overline{\sigma}_2^2\overline{w}_2^{k-2}.$$

Consider $0<\tau<1$ and note $e^{H_2(\tau)}$ the number of brackets with entries in

$$\mathcal{G}=\left\{K_h(X-x)|\delta(\varepsilon,\theta)|(|\delta(\beta,\theta)|^2+|\delta(\varepsilon,\theta)|^2),(\beta,\varepsilon,\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta\right\}$$

and L^2-length τ^2 needed to cover $\mathcal{G}$. Then constructing brackets as in Lemma B.2, with

$$|\delta(\varepsilon,\theta)|\left(|\delta(\beta,\theta)|^2+|\delta(\varepsilon,\theta)|^2\right)$$

instead of $\delta(\mathbf{b},\theta)$ (see (B.34)) yields that there is C such that $H_2(\tau)\leq C\log((t_\beta+t_\varepsilon)n/\tau)$. Recall that $\overline{\sigma}_2\leq 1$ as assumed in Proposition A.3. Arguing then as in the proof of Proposition A.2 yields that

$$\mathbb{E}\left[\sup_{(\beta,\varepsilon,\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta}\left|\mathbb{R}_n^2(\beta,\varepsilon;\theta)-\mathbb{E}\left[\mathbb{R}_n^2(\beta,\varepsilon;\theta)\right]\right|\right]$$
$$\leq n^{1/2}\overline{\sigma}_2\log^{1/2}\left(\frac{(t_\beta+t_\varepsilon)n}{\overline{\sigma}_2}\right)+(\overline{\sigma}_2+\overline{w}_2)\log\left(\frac{(t_\beta+t_\varepsilon)n}{\overline{\sigma}_2}\right),$$

and $\log((t_\beta+t_\varepsilon)n/\overline{\sigma}_2)\leq C\log n$. Substituting then in the equation above $\overline{\sigma}_2$ and $\overline{w}_2$ by their exact orders and using $n\underline{h}^d/\log n\geq C$ under Assumption K yield that

$$\mathbb{E}\left[\sup_{(\beta,\varepsilon,\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta}\left|\mathbb{R}_n^2(\beta,\varepsilon;\theta)-\mathbb{E}\left[\mathbb{R}_n^2(\beta,\varepsilon;\theta)\right]\right|\right]$$
$$\leq C\frac{t_\varepsilon(t_\beta+t_\varepsilon)^2\log^{1/2}n}{n\underline{h}^d}\left(1+\frac{\log^{1/2}n}{n^{1/2}}+\frac{\log^{1/2}n}{\left(n\underline{h}^d\right)^{1/2}}\right)\leq C\frac{t_\varepsilon(t_\beta+t_\varepsilon)^2}{n\underline{h}^d}\log^{1/2}n,\quad\text{(B.40)}$$

under Assumption K.

The Cauchy-Schwarz and the convex inequalities together with (B.39) yield that

$$\sup_{(\beta,\varepsilon,\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta} \mathbb{E}\left[K_h(X-x)|\delta(\varepsilon,\theta)|(\delta(\beta,\theta)^2+\delta(\varepsilon,\theta)^2)\right]$$

$$\leq \sqrt{2} \sup_{(\beta,\varepsilon,\theta)\in\mathcal{B}(0,t_\beta)\times\mathcal{B}(0,t_\varepsilon)\times\Theta} \mathbb{E}^{1/2}\left[K_h(X-x)^2\delta(\varepsilon,\theta)^2(\delta(\beta,\theta)^4+\delta(\varepsilon,\theta)^4)\right] \leq C\frac{t_\varepsilon(t_\beta+t_\varepsilon)^2}{n\left(n\underline{h}^d\right)^{1/2}}.$$

Finally observe that (B.40), the equation above and Assumption K yield that

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

B.5 Proof of Lemma A.2

Observe that Lemma A.1-(v) and Assumptions K and R-(iii) yield that

$$\mathbf{J}_i(\theta)\succ C\mathbf{M}(X_i;\theta),\ \text{where}\ \mathbf{M}(X_i;\theta)=K_h(X_i-x)\,\mathbf{U}\left(\frac{X_i-x}{h}\right)\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top,$$

for $\overline{h}$ small enough. Hence for all θ in Θ

$$\frac{1}{nh^d}\sum_{i=1}^n\mathbf{J}_i(\theta)\succ\frac{C}{nh^d}\sum_{i=1}^n\mathbf{M}(X_i;\theta)=\overline{\mathbf{M}}_n(\theta). \tag{B.41}$$

Observe now that the entries of $\overline{\mathbf{M}}_n(\theta)$ are

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

Then changing $\delta(\mathbf{b},\theta)$ into $((X-x)/h)^{\mathbf{v}_1+\mathbf{v}_2}K((X-x)/h)$ in (B.34) and arguing as in Lemma B.2 and Proposition A.2 yield that

$$\max_{\mathbf{v}_1,\mathbf{v}_2}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{nh^d}\sum_{i=1}^n\left(\left(\frac{X_i-x}{h}\right)^{\mathbf{v}_1+\mathbf{v}_2}K\left(\frac{X_i-x}{h}\right)-\mathbb{E}\left[\left(\frac{X_i-x}{h}\right)^{\mathbf{v}_1+\mathbf{v}_2}K\left(\frac{X_i-x}{h}\right)\right]\right)\right|\right\}=o_{\mathbb{P}}(1),$$

where $\max_{\mathbf{v}_1,\mathbf{v}_2}$ is the maximum over the $\mathbf{v}_1$, $\mathbf{v}_2$'s with $0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p$. Let $\mathbf{M}(\theta)$ be the matrix with entries

$$\frac{1}{h^d}\mathbb{E}\left[\left(\frac{X-x}{h}\right)^{\mathbf{v}_1+\mathbf{v}_2} K\left(\frac{X-x}{h}\right)\right],\ 0 \leq |\mathbf{v}_1|, |\mathbf{v}_2| \leq p.$$

It follows that

$$\sup_{\theta\in\Theta}\left\|\frac{1}{nh^d}\sum_{i=1}^{n}\mathbf{M}\left(X_i;\theta\right)-\mathbf{M}\left(\theta\right)\right\| = o_{\mathbb{P}}\left(1\right). \tag{B.42}$$

By definition of $\mathbf{M}(\theta)$, for all $\mathbf{z}$ in $\mathbb{R}^P$,

$$\begin{aligned}
\mathbf{z}^\top\mathbf{M}(\theta)\mathbf{z} &= \frac{1}{h^d}\mathbb{E}\left[\sum_{0\leq|\mathbf{v}_1|,|\mathbf{v}_2|\leq p}\mathbf{z}_{\mathbf{v}_1}\mathbf{z}_{\mathbf{v}_2}\left(\frac{X-x}{h}\right)^{\mathbf{v}_1+\mathbf{v}_2} K\left(\frac{X-x}{h}\right)\right] \\
&= \sum_{0\leq|\mathbf{v}_1|,|\mathbf{v}_2|\leq p}\mathbf{z}_{\mathbf{v}_1}\mathbf{z}_{\mathbf{v}_2}\int u^{\mathbf{v}_1+\mathbf{v}_2}K(u)f(x+hu)du \\
&= \int\left(\sum_{0\leq|\mathbf{v}|\leq p}\mathbf{z}_{\mathbf{v}}u^{\mathbf{v}}\right)^2 K(u)f(x+hu)du.
\end{aligned}$$

Assumption X gives that there exists C in $\mathbb{R}$ such that, $f(x+hu) \geq C > 0$ for all x in $\mathcal{X}_0$, $\overline{h}$ small enough and all u in $\mathcal{K}$. Recall also that $K(z) \geq \underline{K}\mathbb{I}(z \in \mathcal{B}(0,1))$ by Assumption K. This gives for all x in $\mathcal{X}_0$, h in $\left[\underline{h}, \overline{h}\right]$, $\overline{h}$ small enough,

$$\mathbf{z}^\top\mathbf{M}(\theta)\mathbf{z} \geq C\int_{\mathcal{B}(0,1)}\left(\sum_{0\leq|\mathbf{v}|\leq p}\mathbf{z}_{\mathbf{v}}u^{\mathbf{v}}\right)^2 du \geq C\left\|\mathbf{z}\right\|^2,$$

since $(\int_{\mathcal{B}(0,1)}(\sum_{0\leq|\mathbf{v}|\leq p}\mathbf{z}_{\mathbf{v}}u^{\mathbf{v}})^2du)^{1/2}$ is a squared norm in $\mathbf{z}$ and by norm equivalence. Hence the lower bound above, (B.41) and (B.42) yield that there exists $\underline{\lambda} > 0$ such that,

$$\inf_{\theta\in\Theta}\underline{\lambda}_n(\theta) \geq \inf_{\theta\in\Theta}\inf_{\|\mathbf{z}\|=1}\mathbf{z}^\top\left(\mathbf{M}\left(\theta\right)+o(1)\right)\mathbf{z} \geq \underline{\lambda}+o_{\mathbb{P}}(1).\square$$

B.6 Proof of Lemma A.3

Definition (3.3.5) of $\mathbf{S}_i(\theta)$ and the F.O.C. in Lemma A.1-(ii) yield that $\mathbb{E}\left[\mathbf{S}_i(\theta)\right] = 0$, for all θ in Θ. Define

$$S_{i,k}(\theta) = \frac{1}{nh^d} K\left(\frac{X_i - x}{h}\right)\left(\frac{X_i - x}{h}\right)^k r(Y_i - m^*(X_i;\theta)),$$

for $1 \leq i \leq n$ and $|k| \leq p$. It is sufficient to show that for all $|k| \leq p$,

$$\sup_{\theta\in\Theta}\left|\sum_{i=1}^n S_{i,k}(\theta)\right| = O_{\mathbb{P}}\left(\frac{\log n}{n\underline{h}^d}\right)^{1/2},$$

to prove the Lemma since

$$\sup_{\theta\in\Theta}\left\|\sum_{i=1}^n \mathbf{S}_i(\theta)\right\| \leq \max_{|k|\leq p}\sup_{\theta\in\Theta}\left|\sum_{i=1}^n S_{i,k}(\theta)\right|.$$

Define

$$N_i = r(Y_i - m^*(X_i;\theta))\mathbb{I}\left(\frac{X_i - x}{h} \in \mathcal{K}\right), \text{ and } K_{\mathbf{k}}(z) = K(z)z^k.$$

Now define

$$\begin{aligned} V_1 &= \sum_{i=1}^n \frac{1}{nh^d} K_{\mathbf{k}}\left(\frac{X_i - x}{h}\right)\left[N_i\mathbb{I}\left(|N_i| \leq tn^{1/\nu}\right) + tn^{1/\nu}\mathrm{sgn}(N_i)\mathbb{I}\left(|N_i| \geq tn^{1/\nu}\right)\right], \\ V_2 &= \sum_{i=1}^n \frac{1}{nh^d} K_{\mathbf{k}}\left(\frac{X_i - x}{h}\right)\left(N_i - tn^{1/\nu}\mathrm{sgn}(N_i)\right)\mathbb{I}\left(|N_i| \geq tn^{1/\nu}\right), \end{aligned}$$

where ν is as in Assumption L-(iii) and where t is a strictly positive real number, so that

$$\sum_{i=1}^n S_{i,k}(\theta) = \sum_{i=1}^n \left(S_{i,k}(\theta) - \mathbb{E}\left[S_{i,k}(\theta)\right]\right) = \left(V_1 - \mathbb{E}\left[V_1\right]\right) + V_2 - \mathbb{E}\left[V_2\right],$$

since $\mathbb{E}\left[S_{i,k}(\theta)\right] = 0$. We will bound separately $V_1 - \mathbb{E}\left[V_1\right]$, V_2 and $\mathbb{E}\left[V_2\right]$. We first bound V_2. First observe that

$$\begin{aligned} \sup_{\theta\in\Theta}|V_2| &\leq \sup_{\theta\in\Theta}\sum_{i=1}^n \frac{2}{nh^d} K_{\mathbf{k}}\left(\frac{X_i - x}{h}\right)|N_i|\mathbb{I}\left(|N_i| \geq tn^{1/\nu}\right) \\ &\leq \max_{1\leq i\leq n}\sup_{\theta\in\Theta}|N_i|\mathbb{I}\left(|N_i| \geq tn^{1/\nu}\right)\left(\sup_{\theta\in\Theta}\sum_{i=1}^n \frac{2}{nh^d} K_{\mathbf{k}}\left(\frac{X_i - x}{h}\right)\right). \qquad \text{(B.43)} \end{aligned}$$

Now observe that

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq i\leq n}\sup_{\theta\in\Theta}|N_i|\,\mathbb{I}\left(|N_i|\geq tn^{1/\nu}\right)\neq 0\right) &= \mathbb{P}\left(\sup_{\theta\in\Theta}|N_i|\geq tn^{1/\nu}\text{ for some }1\leq i\leq n\right)\\
&\leq \sum_{i=1}^{n}\mathbb{P}\left(\sup_{\theta\in\Theta}|N_i|\geq tn^{1/\nu}\right)\leq n\mathbb{P}\left(\sup_{\theta\in\Theta}|N_i|\geq tn^{1/\nu}\right)\\
&\leq n\frac{\mathbb{E}\left[\sup_{\theta\in\Theta}|N_i|^{\nu}\right]}{nt^{\nu}}=\frac{\mathbb{E}\left[\sup_{\theta\in\Theta}|N_i|^{\nu}\right]}{t^{\nu}}. \qquad \text{(B.44)}
\end{aligned}$$

We now bound $\mathbb{E}\left[\sup_{\theta\in\Theta}|N_i|^{\nu}\right]$. Observe that Lemma A.1-(v) yields that for $\overline{h}$ small enough we have

$$\sup_{\theta\in\Theta}\sup_{x'\in\mathcal{X}}\mathbb{I}\left(\frac{x'-x}{h}\in\mathcal{K}\right)r\left(Y-m^*(x';\theta)\right)\leq\sup_{\theta\in\Theta}\sup_{x'\in\mathcal{X}}\mathbb{I}\left(\frac{x'-x}{h}\in\mathcal{K}\right)\sup_{|t|\leq C}r\left(Y-t\right),$$

since $\sup_{x\in\mathcal{X}}|m(x)|\leq C$. Now recall the definition of N_i so that for $\overline{h}$ small enough we can apply Assumption L-(iii) and have

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}|N_i|^{\nu}\right] &= \mathbb{E}\left[\sup_{\theta\in\Theta}\mathbb{I}\left(\frac{X-x}{h}\in\mathcal{K}\right)|r(Y-m^*(X;\theta))|^{\nu}\right]\\
&\leq \mathbb{E}\left[\sup_{\theta\in\Theta}\mathbb{I}\left(\frac{X-x}{h}\in\mathcal{K}\right)\sup_{|t|\leq C}|r(Y-t)|^{\nu}\right]\\
&\leq \sup_{x\in\mathcal{X}}\mathbb{E}\left[\sup_{|t|\leq C}|r(Y-t)|^{\nu}\,|X=x\right]\leq C. \qquad \text{(B.45)}
\end{aligned}$$

It then follows from this and (B.44) that with a probability that can be arbitrarily large by increasing t, we have that for all $1\leq i\leq n$, $\sup_{\theta\in\Theta}|N_i|\,\mathbb{I}\left(|N_i|\geq tn^{1/\nu}\right)=0$. It then follows from (B.43) that $\sup_{\theta\in\Theta}|V_2|=0$ with a probability that can be made arbitrarily large by increasing t.

We now bound $\mathbb{E}[V_2]$. For that observe that (B.45) and the Hölder inequality yield that $\sup_{\theta\in\Theta}|\mathbb{E}[V_2]|$ is less than

$$\begin{aligned}
&\sup_{\theta\in\Theta}\frac{2}{h^d}\mathbb{E}\left[\left|K_{\mathbf{k}}\left(\frac{X_i-x}{h}\right)\right||N_i|\,\mathbb{I}\left(|N_i|\geq tn^{1/\nu}\right)\right]\\
&=\sup_{\theta\in\Theta}\frac{2}{h^d}\mathbb{E}\left[\left|K_{\mathbf{k}}\left(\frac{X_i-x}{h}\right)\right||r(Y-m^*(X_i;\theta))|\,\mathbb{I}\left(|r(Y-m^*(X_i;\theta))|\geq tn^{1/\nu}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{\theta \in \Theta} 2 \int |K_{\mathbf{k}}(u)| f(x+hu) \\
&\quad \times \mathbb{E}\left[|r(Y - m^*(x+hu;\theta))| \mathbb{I}\left(|r(Y - m^*(x+hu;\theta))| \geq t n^{1/\nu} \right) |X = x+hu \right] du \\
&\leq C \sup_{\theta \in \Theta} \sup_{u \in \mathcal{K}} \mathbb{E}\left[|r(Y - m^*(x+hu;\theta))| \mathbb{I}\left(|r(Y - m^*(x+hu;\theta))| \geq t n^{1/\nu} \right) |X = x+hu \right].
\end{aligned}
$$

Now observe that Lemma A.1-(iii) yields that

$$
\lim_{\overline{h} \to 0} \sup_{\theta \in \Theta} \sup_{u \in \mathcal{K}} |m^*(x+hu;\theta) - m(x)| = \lim_{\overline{h} \to 0} \sup_{\theta \in \Theta} \sup_{u \in \mathcal{K}} \left| \mathbf{U}(u)^\top \mathbf{H} \left(\mathbf{b}^*(\theta) - \mathbf{m}_x \right) \right| = 0,
$$

where $\mathbf{m}_x = (m(x), 0, \ldots, 0)^\top$ in $\mathbb{R}^P$. This, the Hölder inequality and $\sup_{x \in \mathcal{X}} |m(x)| \leq C$ then yield that

$$
\begin{aligned}
\sup_{\theta \in \Theta} |\mathbb{E}[V_2]| &\leq C \sup_{x' \in \mathcal{X}} \mathbb{E}\left[\sup_{|t| \leq C} |r(Y-t)| \mathbb{I}\left(\sup_{|t| \leq C} |r(Y-t)| \geq t n^{1/\nu} \right) |X = x' \right] \\
&\leq C \sup_{x' \in \mathcal{X}} \mathbb{E}^{1/\nu}\left[\sup_{|t| \leq C} |r(Y-t)|^\nu |X = x' \right] \\
&\quad \times \mathbb{P}\left(\sup_{|t| \leq C} |r(Y-t)| \geq t n^{1/\nu} |X = x' \right)^{(\nu-1)/\nu} \\
&\leq C \frac{\sup_{x' \in \mathcal{X}} \mathbb{E}\left[\sup_{|t| \leq C} |r(Y-t)|^\nu |X = x' \right]}{t^{\nu-1} n^{(\nu-1)/\nu}} = O\left(n^{(1-\nu)/\nu} \right),
\end{aligned}
$$

under Assumption L-(iii). Then Assumption K on the bandwidth yields that

$$
\sup_{\theta \in \Theta} \mathbb{E}[V_2] = O\left(\frac{\log n}{n \underline{h}^d} \right)^{1/2}.
$$

We finally bound $V_1 - \mathbb{E}[V_1]$. We will argue as in Lemma B.2 and Proposition A.2 to bound $\mathbb{E}\left[\sup_{\theta \in \Theta_0} |V_1 - \mathbb{E}(V_1)| \right]$. For that, observe that

$$
V_1 = \sum_{i=1}^n \left(V_{1,i} - \mathbb{E}[V_{1,i}] \right),
$$

where the $V_{1,i}$'s are i.i.d. random variables. Observe that there is C such that

$$
\sup_{\theta \in \Theta} |V_{i,1}| \leq C t \frac{n^{1/\nu}}{n \underline{h}^d} = \overline{w}_{V_1}.
$$

Observe also that $\mathrm{Var}(V_{i,1})$ is less than

$$\begin{aligned}
\mathbb{E}\left[V_{i,1}^2\right] &= \frac{1}{(nh^d)^2}\mathbb{E}^2\left[\left|K_{\mathbf{k}}\left(\frac{X_i-x}{h}\right)\right|^2\left(N_i\mathbb{I}\left(|N_i|\le tn^{1/\nu}\right)+tn^{1/\nu}\mathrm{sgn}(N_i)\mathbb{I}\left(|N_i|\ge tn^{1/\nu}\right)\right)^2\right]\\
&\le \frac{C}{(nh^d)^2}\mathbb{E}\left[\left|K_{\mathbf{k}}\left(\frac{X_i-x}{h}\right)\right|^2 N_i^2\right]\\
&= \frac{C}{n(nh^d)}\int |K_{\mathbf{k}}(u)|^2 f(x+hu)\mathbb{E}\left[r(Y-m^*(x+hu;\theta))^2|X=x+hu\right]du.
\end{aligned}$$

Now observe that by Lemma A.1-(v) yields that

$$\sup_{\theta\in\Theta}\sup_{x'\in\mathcal{X}} K_h(x'-x)\left|m^*(x';\theta)-m(x)\right| \to_{n\to\infty} 0.$$

Assumption L-(iii) then yields that

$$\sup_{\theta\in\Theta}\mathrm{Var}(V_{i,1}) \le \frac{C}{n\left(n\underline{h}^d\right)} = \overline{\sigma}_{V_1}^2.$$

It then follows that $\mathbb{E}\left[|V_{i,1}|^k\right] \le k!\overline{\sigma}_{V_1}^2\overline{w}_{V_1}^{k-2}/2$ for all integer number k. Consider $0<\tau<1$ and define $e^{H_{V_1}(\tau)}$ the number of brackets with entries in $\{V_{i,1}=V_{i,1}(\theta),\theta\in\Theta\}$ and L^2-length τ^2 needed to cover $\{V_{i,1}(\theta),\theta\in\Theta\}$. Now observe that

$$V_{1,i} = \frac{1}{nh^d}K\left(\frac{X_i-x}{h}\right)T(N_i),$$

where $T(z)=z\mathbb{I}\left(|z|\le tn^{1/\nu}\right)+tn^{1/\nu}\mathrm{sgn}(z)\mathbb{I}\left(|z|\ge tn^{1/\nu}\right)$ is such that $|T(z)-T(z')|\le|z-z'|$ for all (z,z') in $\mathbb{R}^2$. Then arguing as in Lemma B.2 yields that $H_{V_1}(\tau)\le C\log(n/\tau)$. Note that $\overline{\sigma}_{V_1}\le 1$ for n large enough, so that as in Proposition A.2, we have to bound the quantity

$$n^{1/2}\int_0^{\overline{\sigma}_{V_1}} H_{V_1}^{1/2}(u)du + \left(\overline{\sigma}_{V_1}+\overline{w}_{V_1}\right)H_{V_1}(\overline{\sigma}_{V_1}).$$

Applying the Cauchy-Schwarz inequality to the integral in the equation above and substituting $\overline{\sigma}_{V_1}$ and $\overline{w}_{V_1}$ by their exact orders give that this quantity is less than

$$C\frac{\log^{1/2}n}{\left(n\underline{h}^d\right)^{1/2}}\left(1+\frac{\log^{1/2}n}{n^{1/2}}+\frac{tn^{1/\nu}\log^{1/2}n}{\left(n\underline{h}^d\right)^{1/2}}\right)\le C\frac{\log^{1/2}n}{\left(n\underline{h}^d\right)^{1/2}},$$

under Assumption K on the bandwidth. Then

$$\sup_{\theta \in \Theta} |V_1| = O_{\mathbb{P}} \left(\frac{\log n}{n \underline{h}^d} \right)^{1/2},$$

by the Markov inequality. □

Bibliography

[1] BHATTACHARYA, P. K. & A. GANGOPADHYAY (1990). Kernel and nearest neighbor estimation of a conditional quantile. *The Annals of Statistics*, **18**, 1400-1415.

[2] CHOW, Y.S. & H. TEICHER (2003). *Probability theory. Independence, interchangeability, martingales*. Third edition, Springer.

[3] CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.

[4] DONY, J. & D.M. MASON (2008) Uniform in bandwidth consistency of conditional U-statistics *Bernoulli*, **14**(4), 1108-1133.

[5] EINMAHL, U. & D.M. MASON (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, **33**, 1380-1403.

[6] FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, **21**, 196-216.

[7] FAN, J. & I. GIJBELS (1996). *Local polynomial modeling and its applications*. Chapman & Hall/CRC.

[8] FAN, J., HU, T.C. & Y.K. TRUONG (1994) Robust nonparametric function estimation. *Scandinavian Journal of Statistics*, **21**, 433-446.

[9] GUERRE, E. & P. LAVERGNE (2005). Data-driven rate-optimal specification testing in regression models. *The Annals of Statistics*, **33**, 840-870.

[10] Hastie, T. & C. Loader (1993). Local regression: Automatic kernel carpentry (with discussion). *Statistical Science*, **8**, 139-143.

[11] Hong, S.Y. (2003). Bahadur representation and its applications for local polynomial estimates in nonparametric M-regression. *Journal of Nonparametric Statistics*, **15**, 237-251.

[12] Horowitz, J.L. & V.G. Spokoiny (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599-631.

[13] Huber P.J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, **35**, 73-101.

[14] Huber P.J. (1981). *Robust Statistics.* Wiley.

[15] Kolmogorov, A.N. & S.V. Fomin (1969). *Introductory Real Analysis* New York: Dover Publications.

[16] Kong, E., O. Linton & Y. Xia (2009). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. Forthcoming in *Econometric Theory.*

[17] Loader, C. (1999). *Local regression and likelihood.* Springer.

[18] 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.

[19] Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82**, 231-238.

[20] Ruppert, D. & M. P. Wand (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, **21**, 1346-1370.

[21] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10**, 1040-1053.

[22] Tsybakov, A.B. (1986). Robust reconstruction of functions by the local-approximation method. *Problemy Peredachi Informatsii*, **22**, 69-84.

[23] van de Geer, S. (1999). *Empirical processes in M-estimation.* Cambridge University Press.

[24] van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press.

[25] Welsh, A.H. (1996) Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica*, **6**, 347-366.

[26] Zeidler, E. (1985). *Nonlinear Functional Analysis and its Applications, I Fixed-Point Theorems.* Springer Verlag.

Chapter 4

Maximal deviations for the LP quantile estimator

Uniform Confidence Bands for Local Polynomial Quantile Estimators

Camille Sabbah [1]

Abstract

This paper deals with uniform consistency and uniform confidence bands for the quantile function and its derivatives. We describe a kernel local polynomial estimator of quantile function and give uniform consistency. Furthermore, we derive its maximal deviation limit distribution using an approximation in the spirit of Bickel and Rosenblatt (1973).

Keywords: Uniform confidence bands; conditional quantile estimation.

1. Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 175 rue du Chevaleret, Paris, France. E-mail : `camille.sabbah@univ-paris1.fr`

4.1 Introduction

Consider independent and identically distributed (i.i.d.) variables (X_i, Y_i), $i = 1, \dots, n$ having the same distribution than (X, Y) where X and Y are real random variables. Define for α in $(0, 1)$ the conditional quantile function of Y given X, $Q(\alpha|x) = F^{-1}(\alpha|x)$, where $F(\cdot|\cdot)$ is the cumulative distribution function of Y given X. It is well-known that $Q(\alpha|x) = \arg\min_t \mathbb{E}\left[\ell_\alpha(Y - t)|X = x\right]$ where $\ell_\alpha(t) = |t| + (2\alpha - 1)t$. The LP estimator $\widehat{\mathbf{b}}(\alpha, x)$ of order p of $Q(\alpha|x)$ is

$$\arg\min_{\mathbf{b}} \sum_{i=1}^{n} \ell_\alpha \left(Y_i - \sum_{j=0}^{p} b_j (X_i - x)^j \right) K\left(\frac{X_i - x}{h}\right),$$

where $\mathbf{b} = (b_0, \dots, b_p)^\top$, h is a bandwidth parameter and $K(\cdot)$ a kernel function. In particular, the first coordinate $\widehat{Q}(\alpha|x) = \widehat{b}_0(\alpha, x)$ of the vector $\widehat{\mathbf{b}}(\alpha, x)$ is an estimator of $Q(\alpha|x)$. Similarly $\widehat{b}_j(\alpha, x)$, the jth coordinate of the vector $\widehat{\mathbf{b}}(\alpha, x)$ is an estimator of $Q_j(\alpha|x) = Q^{(j)}(\alpha|x)/j!$ provided this quantity exists.

The main contribution of the present paper is to give the law of the maximal deviation of the LP conditional quantile estimator. Namely for some deterministic diverging sequences $\{a_n\}$, $\{b_n\}$, and a map $(\alpha, x) \mapsto r_j(\alpha, x)$ we have for all t and when n diverges

$$\mathbb{P}\left(a_n \left(\sup_{x \in \mathcal{X}_0} r_j(\alpha, x)^{-1/2} \left| \widehat{b}_j(\alpha, x) - Q_j(\alpha|x) \right| - b_n \right) < t \right) \to \exp\left(-2\exp(-t)\right), \qquad (4.1.1)$$

for all $0 \leq j \leq p$ such that the derivatives exists and where $\mathcal{X}_0$ is an inner compact subset of the support of X. An interesting feature of this result compared with the vast majority of existing results is that the order of the LP estimator may differ from the actual differentiability order of $Q(\alpha|x)$. This also extends Härdle and Song (2009) which is limited to LP estimators of order 0 and estimation of the quantile function. The papers is organized as follows. Section 2 gives our main notations and assumptions and Section 3 states our main results. Proofs are gathered in two Appendices.

4.2 Notations and Assumptions

Define the lowest integer part $\lfloor s \rfloor$ of s as the unique integer number with $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$. For the rest of the paper let $[\underline{\alpha}, \overline{\alpha}]$ be a compact subset of $(0, 1)$. $\|(x, y)\|$ stands for the Euclidean norm and $\mathbb{I}(\cdot)$ for the indicator function. Positive constants are denoted by the generic letter C and may vary from line to line. Our main assumptions are as follows.

Assumption X *The distribution of X has a cumulative distribution function $F(\cdot)$ twice continuously differentiable over the compact support $\mathcal{X}$ of X and $F'(\cdot) = f(\cdot) \geq C > 0$.*

Assumption F *The cumulative distribution function $F(\cdot|x)$ of Y given $X = x$ is continuously differentiable over $\mathbb{R}$ with derivative $f(\cdot|x) \geq C > 0$ for all x in $\mathcal{X}$. The map $(x, y) \in \mathcal{X} \times \mathbb{R} \mapsto f(y|x)$ is Lipschitz.*

Assumption K *The nonnegative kernel function $K(\cdot)$ has a compact support $\mathcal{K} = [a, b]$ with $\int K(z)dz = 1$. For some $\underline{K} > 0$, $K(\cdot) \geq \underline{K}\mathbb{I}(|\cdot| \leq 1)$. The kernel $K(\cdot)$ is continuously differentiable over $\mathcal{K}$, and $K(a) = K(b) = 0$.*

Assumption Q *For some real number $s > 1$, $(\alpha, x) \mapsto Q(\alpha|x)$ is $\lfloor s+1 \rfloor$-times continuously differentiable. The maps $\alpha \in [\underline{\alpha}, \overline{\alpha}] \mapsto \partial^{\lfloor s+1 \rfloor} Q(\alpha|x)/\partial\alpha^{\lfloor s+1 \rfloor}$ and $x \in \mathcal{X} \mapsto \partial^{\lfloor s+1 \rfloor} Q(\alpha|x)/\partial\alpha^{\lfloor s+1 \rfloor}$ are Hölder continuous with exponent $s - \lfloor s \rfloor$ respectively for all x in $\mathcal{X}$ and for all α in $[\underline{\alpha}, \overline{\alpha}]$.*

Assumption H *The order p of the LP estimator is such that $p \geq \lfloor s \rfloor$ for s as in Assumption Q. The bandwidth parameter h is such that $h \to 0$, $nh \to \infty$, $\log^5 n/(nh) \to 0$, $(nh)^{1/2} h^s \log^{1/2} n \to 0$ and $\limsup \log n/(nh^{s+1})^{1/2} < \infty$ when n diverges.*

Assumption F implies in particular that the conditional quantile $Q(\alpha|x)$ is uniquely defined for all α. Assumptions K and X are standard. For $s = 2$ Assumption H is similar to Assumption (A2) of Härdle and Song (2009) since it allows the bandwidth to belong to the interval $\left[C\left(\left(\log^2 n\right)/n\right)^{1/3}, \widetilde{h}_n\right]$, for any sequence $\widetilde{h}_n = o(n \log n)^{-1/5}$. Finally Assumption Q is common in nonparametric statistics.

4.3 Main Results

In the sequel α will be any element of $[\underline{\alpha}, \overline{\alpha}]$. Define the $(p+1)\times(p+1)$ standardization diagonal matrix $\mathbf{H} = \text{Diag}\,(h^v, v \in \mathbb{N}, v \leq p)$. Our limit distribution result depends upon

$$\begin{aligned}\mathbf{Q}_p &= \left[\int K'(u)^2 u^{i+j} du - \frac{\{i(i-1)+j(j-1)\}}{2}\int K(u)u^{i+j-2}du\right]_{0\leq i,j\leq p},\\ \mathbf{T}_p &= \left(\int K^2(u)u^{i+j}du\right)_{0\leq i,j\leq p},\ C_j = \frac{\left(\mathbf{N}_p^{-1}\mathbf{Q}_p\mathbf{N}_p^{-1}\right)_{j+1,j+1}}{\left(\mathbf{N}_p^{-1}\mathbf{T}_p\mathbf{N}_p^{-1}\right)_{j+1,j+1}}.\end{aligned} \tag{4.3.2}$$

Theorem 1 *Assume that Assumptions F, H, K, Q and X are verified. Then (4.1.1) holds for all $j = 0, \ldots, \lfloor s \rfloor$ with*

$$\begin{aligned}a_n &= (-2\log h)^{1/2},\ \ b_n = (-2\log h)^{1/2} + (-2\log h)^{-1/2}\log(C_j/(2\pi)),\\ r_j(\alpha,x) &= \frac{\alpha(1-\alpha)}{f^2(Q(\alpha|x)|x)f(x)}(\mathbf{N}_p^{-1}\mathbf{T}_p\mathbf{N}_p^{-1})_{j,j}.\end{aligned}$$

Note that $r_j(\alpha, x)$ involves the sparsity function $q(\alpha|x) = f(Q(\alpha|x)|x)^{-1}$ and $f(x)$. Our feasible quantile confidence bound relies on an estimator of $q(\alpha|x)$, $\widehat{q}(\alpha|x)$ which can be found in Guerre and Sabbah (2009, Proposition 3). Let $\widehat{q}(\alpha|x) = \frac{1}{h}\int \widehat{Q}(\alpha + ht|x)dK_q(t)$, where $K_q(\cdot)$ is a signed measure over $\mathbb{R}$ with compact support $\mathcal{K}_q$ and satisfying

$$\int t dK_q(t) = 1,\ \ \int t^j dK_q(t) = 0,\ \ j = 0, 2, \ldots, \lfloor s \rfloor,\ \ \int |dK_q(t)| < \infty.$$

Now define $\widehat{f}(x) = \sum_{i=1}^n K((X_i - x)/h)/(nh)$ and

$$\widehat{\mathbf{V}}(\alpha, x) = \frac{\alpha(1-\alpha)\widehat{q}(\alpha|x)^2}{\widehat{f}(x)}\mathbf{N}_p^{-1}\mathbf{T}_p\mathbf{N}_p^{-1}. \tag{4.3.3}$$

Theorem 2 *Under Assumptions F, H, K, Q and X, for any $0 < \lambda < 1$, a $(1-\lambda)100\%$ confidence band for $x \mapsto Q^{(j)}(\alpha|x)$, $j = 0, \ldots, \lfloor s \rfloor$ is given by the collection of all maps belonging to the set of functions $\mathcal{Q}_j$*

$$\left\{\mathcal{Q}_j;\ \sup_{x\in\mathcal{X}_0}\left[\left|j!\,\widehat{b}_j(\alpha,x) - \mathcal{Q}_j(\alpha,x)\right|\left(\widehat{\mathbf{V}}(\alpha,x)\right)_{j,j}^{-1/2}\right] \leq L_{\lambda,j}\right\},$$

where for all $j = 0, \ldots, \lfloor s \rfloor$,

$$L_{\lambda,j} = j! \left(nh^{2j+1}\right)^{-1/2} (-2\log h) \left(1 + (-2\log h)^{-1} \left[a_\lambda + \log\left(C_j^{1/2}/(2\pi)\right)\right]\right),$$

and $a_\lambda = -\log(-\log(1-\lambda)/2)$.

4.4 Appendix A: Proofs of main results

For the rest of the paper we define $\theta = (\alpha, x)$ for notational convenience.

A.1 Proof of Theorem 1

In order to give the limit distribution of the maximal deviation of the quantity

$$\sup_{x\in\mathcal{X}_0} \left|\widehat{b}_j(\alpha, x) - Q_j(\alpha|x)\right|,$$

we need a Bahadur expansion of the process $\widehat{\mathbf{b}}(\alpha, x) - \mathbf{b}(\alpha|x)$ where

$$\mathbf{b}(\alpha|x) = (Q(\alpha|x), Q_1(\alpha|x), \ldots, Q_{\lfloor s \rfloor}(\alpha|x), 0, \ldots 0)^\top.$$

Define $\mathbf{N}_p = \left(\int K(u)u^{i+j}du\right)_{0\le i,j\le p}$ and $\mathbf{J}(\theta) = 2f(Q(\alpha|x)|x)f(x)\mathbf{N}_p$. Furthermore define for any integer number $p \ge \lfloor s \rfloor$ the vectors in $\mathbb{R}^{p+1}$, $\mathbf{U}(u) = (1, u, \ldots, u^p)^\top$, $K_{\mathbf{U}}(u) = K(u)\mathbf{U}(u)$.

Lemma A.1 *Under Assumptions F, H, K, Q and X*

$$\begin{aligned}&(nh)^{1/2}\mathbf{H}\left(\widehat{\mathbf{b}}(\alpha, x) - \mathbf{b}(\alpha|x)\right)\\ &\quad= \frac{2\mathbf{J}^{-1}(\alpha, x)}{(nh)^{1/2}} \sum_{i=1}^n \{\alpha - \mathbb{I}(Y_i \le Q(\alpha|X_i))\} K_{\mathbf{U}}\left(\frac{X_i - x}{h}\right) + \mathbf{r}_n(\alpha, x),\end{aligned}$$

with $\sup_{x\in\mathcal{X}_0} \|\mathbf{r}_n(\theta)\| = O_{\mathbb{P}}((nh)^{1/2}h^s + h\log^{1/2} n + (\log n)^{3/4}/(nh)^{1/4})$.

Define $m(z) = \alpha(1-\alpha)f(z)$, z in $\mathcal{X}_0$ and observe that there exists $0 < \underline{m} \le \overline{m} < \infty$ such that

$$\underline{m} \le \inf_{(\alpha,z)\in[\underline{\alpha},\overline{\alpha}]\times\mathcal{X}_0} m(z) \le \sup_{(\alpha,z)\in[\underline{\alpha},\overline{\alpha}]\times\mathcal{X}_0} m(z) \le \overline{m},$$

under Assumption X. Define for all $0 \leq j \leq p$, $K_j(u) = K(u)u^j$ and

$$Y_{n,j}(x) = \frac{m(x)^{-1/2}}{(nh)^{1/2}} \sum_{i=1}^{n} \{\mathbb{I}\left(Y_i \leq Q(\alpha|X_i)\right) - \alpha\} K_j\left(\frac{X_i - x}{h}\right).$$

In a first step we will write the process $Y_{n,j}(x)$ with the uniform empirical process. In a second step we will compare the process $Y_{n,j}(x)$ with the process

$$Z_j(x) = h^{-1/2} \int K_j\left(\frac{z-x}{h}\right) dW(z),$$

where $W(\cdot)$ is a Wiener process defined on the support of X, and we prove that for all $j = 0, \ldots, p$, $Y_{n,j}(x)$ has the same limit distribution than the one of $Z_j(x)$. This will end the proof arguing as in Claeskens and van Keilegom (2003, proof of Theorem 2.2).

Step 1 : $Y_{n,j}(x)$. Since $K(\cdot)$ is continuously differentiable and has compact support,

$$K_j\left(\frac{X_i - x}{h}\right) = -\int \mathbb{I}(X_i \leq z) dK_j\left(\frac{z-x}{h}\right),$$

and because $F(q(\alpha|X)|X) = \alpha$, we can rewrite $Y_{n,j}(x)$ as

$$\begin{aligned}
&\frac{m(x)^{-1/2}}{(nh)^{1/2}} \sum_{i=1}^{n} \{\alpha - \mathbb{I}(F(Y_i|X_i) \leq \alpha)\} \int \mathbb{I}(X_i \leq z) dK_j\left(\frac{z-x}{h}\right) \\
&= \frac{m(x)^{-1/2}}{(nh)^{1/2}} \sum_{i=1}^{n} \int \{\alpha - \mathbb{I}(F(Y_i|X_i) \leq \alpha)\} \mathbb{I}(F(X_i) \leq F(z)) dK_j\left(\frac{z-x}{h}\right).
\end{aligned}$$

Rosenblatt (1952) yields that $(F(X_i), F(Y_i|X_i))$ is uniformly distributed over $[0,1]^2$ so that there exists a sequence of i.i.d. random vectors $(U_i, V_i)_{i \geq 1}$ uniformly distributed on $[0,1]^2$ such that

$$Y_{n,j}(x) = \frac{m(x)^{-1/2}}{(nh)^{1/2}} \sum_{i=1}^{n} \int \{\alpha - \mathbb{I}(V_i \leq \alpha)\} \mathbb{I}(U_i \leq F(z)) dK_j\left(\frac{z-x}{h}\right).$$

Define $\mathbb{U}_n(\cdot,\cdot)$ the empirical process

$$\mathbb{U}_n(u,v) = n^{-1/2} \sum_{i=1}^{n} \{\mathbb{I}(U_i \leq z, V_i \leq v) - uv\}, \ \ (u,v) \ \text{ in } \ [0,1]^2.$$

Then $Y_{n,j}(x)$ can be written w.r.t. $\mathbb{U}_n(\cdot,\cdot)$ as

$$\begin{aligned}
&\frac{m(x)^{-1/2}}{h^{1/2}} \int \{\alpha \mathbb{U}_n(F(z),1) - \mathbb{U}_n(F(z),\alpha)\}\, dK_j\left(\frac{z-x}{h}\right) \\
&+ \frac{m(x)^{-1/2}}{h^{1/2}} \int \{\alpha F(z) - \alpha F(z)\}\, dK_j\left(\frac{z-x}{h}\right) \\
&= \frac{m(x)^{-1/2}}{h^{1/2}} \int \{\alpha \mathbb{U}_n(F(z),1) - \mathbb{U}_n(F(z),\alpha)\}\, dK_j\left(\frac{z-x}{h}\right). \qquad \text{(A.1)}
\end{aligned}$$

Step 2 : $Y_{n,j}(x)$ and $Z_j(x)$ have the same limit distribution. First recall that the result of Tusnàdy (1977, Theroem 1) yields that there exists a sequence of Brownian bridges $(B_n(\cdot,\cdot))_{n\geq 1}$ defined on $[0,1]^2$ such that

$$\sup_{(u,v)\in[0,1]^2} |\mathbb{U}_n(u,v) - B_n(u,v)| = O_{\mathbb{P}}\left(\frac{\log^2 n}{n^{1/2}}\right).$$

Now take $Y_{n,j}(x)$ as in (A.1) and observe that

$$\begin{aligned}
&\sup_{x\in\mathcal{X}_0}\left|\frac{h^{-1/2}}{m(x)^{1/2}} \int |\alpha\left(\mathbb{U}_n(F(z),1) - B_n(F(z),1)\right) - \mathbb{U}_n(F(z),\alpha) + B_n(F(z),\alpha)|\right. \\
&\qquad \left. \times dK_j\left(\frac{z-x}{h}\right)\right| \\
&\leq \sup_{x\in\mathcal{X}_0}\left|\frac{h^{-1/2}}{m(x)^{1/2}} \int (1+\alpha) \sup_{(u,v)\in[0,1]^2} |\mathbb{U}_n(u,v) - B_n(u,v)|\, dK_j\left(\frac{z-x}{h}\right)\right| \\
&\leq \sup_{x\in\mathcal{X}_0}\left|\frac{1}{m(x)^{1/2}} \int dK_j\left(\frac{z-x}{h}\right)\right| \times O_{\mathbb{P}}\left(\frac{\log^2 n}{(nh)^{1/2}}\right).
\end{aligned}$$

The change of variable $z = x + hu$ in the integral above yields that

$$\sup_{x\in\mathcal{X}_0}\left|\frac{1}{m(x)^{1/2}} \int dK_j(u)\right| \leq \left(\inf_{x\in\mathcal{X}_0} m(x)^{1/2}\right)^{-1} \int \left|K_j'(u)\right| du.$$

Since $K(\cdot)$ is continuously differentiable and have a compact support under Assumption K, and since $\inf_{x\in\mathcal{X}_0} m(x) \geq C > 0$, we have that

$$\sup_{x\in\mathcal{X}_0}\left|\frac{1}{m(x)^{1/2}} \int dK_j\left(\frac{z-x}{h}\right)\right| = C.$$

It follows that

$$Y_{n,j}(x) = \frac{h^{-1/2}}{m(x)^{1/2}} \int \{\alpha B_n(F(z),1) - B_n(F(z),\alpha)\} \, dK_j\left(\frac{z-x}{h}\right) + \varepsilon_1(\theta), \tag{A.2}$$

with $\sup_{x\in\mathcal{X}_0} |\varepsilon_1(\theta)| = o_{\mathbb{P}}((-\log h)^{-1/2})$ under Assumption H.

Define now a sequence of bivariate Wiener processes $(W_n(\cdot,\cdot))_n \geq 1$ on $[0,1]^2$ such that $B_n(u,v) = W_n(u,v) - uvW_n(1,1)$ for all (u,v) in $[0,1]^2$. Equation (A.2) then yields that

$$\begin{aligned} Y_{n,j}(x) &= \frac{h^{-1/2}}{m(x)^{1/2}} \int (\alpha W_n(F(z),1) - W_n(F(z),\alpha)) \, dK_j\left(\frac{z-x}{h}\right) \\ &\quad + W_n(1,1)\frac{h^{-1/2}}{m(x)^{1/2}} \int (\alpha F(z) - \alpha F(z)) \, K_j\left(\frac{z-x}{h}\right) + \varepsilon_1(\theta) \\ &= \frac{h^{-1/2}}{m(x)^{1/2}} \int (\alpha W_n(F(z),1) - W_n(F(z),\alpha)) \, dK_j\left(\frac{z-x}{h}\right) \\ &\quad + \varepsilon_1(\theta) \end{aligned} \tag{A.3}$$

since $W_n(1,1)$ is almost surely finite. We now rewrite the integral in (A.3). For that observe that since $W_n(u,0) = W_n(0,t) = 0$ for all (u,t) in $[0,1]^2$ we have

$$\alpha W_n(F(z),1) - W_n(F(z),\alpha) = \int\int [\alpha - \mathbb{I}(v \leq \alpha)] \, \mathbb{I}(u \leq F(z)) dW_n(u,v).$$

Then (A.3) and Fubini's Theorem yield that $Y_{n,j}(x) - \varepsilon_1(\theta)$ can be written as

$$\begin{aligned} &\frac{h^{-1/2}}{m(x)^{1/2}} \int \left(\int\int [\alpha - \mathbb{I}(v \leq \alpha)] \, \mathbb{I}(u \leq F(z)) dW_n(u,v) \right) dK_j\left(\frac{z-x}{h}\right) \\ &= \frac{h^{-1/2}}{m(x)^{1/2}} \int\int\int [\alpha - \mathbb{I}(v \leq \alpha)] \, \mathbb{I}(F^{-1}(u) \leq z) dK_j\left(\frac{z-x}{h}\right) dW_n(u,v) \\ &= \frac{h^{-1/2}}{m(x)^{1/2}} \int\int \{\mathbb{I}(v \leq \alpha) - \alpha\} \, K_j\left(\frac{F^{-1}(u)-x}{h}\right) dW_n(u,v) \end{aligned} \tag{A.4}$$

Observe that $Y_{n,j}(x) - \varepsilon_1(\theta)$ is a centered Gaussian processes. Define $r(\cdot,\cdot)$ the covariance function of the process $Y_{n,j}(x) - \varepsilon_1(\theta)$.Then

$$r(x_1,x_2) = \frac{h^{-1}}{(m(x_1)m(x_2))^{1/2}} \int\int \{\mathbb{I}(v \leq \alpha) - \alpha\}^2 \, K_j\left(\frac{F(u)^{-1}-x_1}{h}\right)$$

$$
\begin{aligned}
& \times K_j\left(\frac{F(u)^{-1}-x_2}{h}\right) dv du \\
= \; & \frac{h^{-1}}{(m(x_1)m(x_2))^{1/2}} \int \left[\int \{\mathbb{I}(v \le \alpha)-\alpha\}^2 dv\right] K_j\left(\frac{z-x_1}{h}\right) \\
& \times K_j\left(\frac{z-x_2}{h}\right) f(z) dz \\
= \; & \frac{h^{-1}}{(m(x_1)m(x_2))^{1/2}} \int \alpha(1-\alpha) f(z) K_j\left(\frac{z-x_1}{h}\right) K_j\left(\frac{z-x_2}{h}\right) dz \\
= \; & \frac{h^{-1}}{(m(x_1)m(x_2))^{1/2}} \int m(z) K_j\left(\frac{z-x_1}{h}\right) K_j\left(\frac{z-x_2}{h}\right) dz \\
= \; & r_1(x_1, x_2),
\end{aligned}
$$

where $r_1(\cdot,\cdot)$ is the covariance function of the centered Gaussian process

$$
x \in \mathcal{X} \mapsto h^{-1/2} \int \left(\frac{m(z)}{m(x)}\right)^{1/2} K_j\left(\frac{z-x}{h}\right) dW(z),
$$

and where $W(\cdot)$ is a Wiener process defined on $\mathcal{X}$. It then follows from (A.4) that $Y_{n,j}(x)-\varepsilon_1(\theta)$ have the same distribution than

$$
x \in \mathcal{X} \mapsto h^{-1/2} \int \left(\frac{m(z)}{m(x)}\right)^{1/2} K_j\left(\frac{z-x}{h}\right) dW(z).
$$

It remains to show that

$$
A = \sup_{x \in \mathcal{X}} \left| h^{-1/2} \int \left(\left[\frac{m(z)}{m(x)}\right]^{1/2} - 1\right) K_j\left(\frac{z-x}{h}\right) dW(z)\right| = o_{\mathbb{P}}((-\log h)^{1/2}), \tag{A.5}
$$

to achieve the proof. The change of variable $z = x + hu$ and integration by parts in (A.5) under Assumption K and X yield that

$$
\begin{aligned}
& \left| h^{-1/2} \int \left(\left[\frac{m(z)}{m(x)}\right]^{1/2} - 1\right) K_j\left(\frac{z-x}{h}\right) dW(z)\right| \\
& \quad \le \left| h^{-1/2} \int W(x+hu) \left(\left[\frac{f(x+hu)}{f(x)}\right]^{1/2} - 1\right) K_j'(u)\, du \right. \\
& \quad \left. + \frac{h^{1/2}}{2} \int W(x+hu) \frac{f'(x+hu)}{(f(x)f(x+hu))^{1/2}} K_j(u)\, du \right|.
\end{aligned}
$$

Since $f(\cdot)$ is continuously differentiable, bounded away from 0 under Assumption X and since $\sup_{u\in\mathcal{X}_0}|W(u)| = O_{\mathbb{P}}(1)$, Assumption K yields that the second integral in the inequality above is a $O_{\mathbb{P}}(h^{1/2})$ uniformly in x in $\mathcal{X}_0$. To control the first integral in the inequality above, observe continuous differentiability of $f(\cdot)$ under Assumption X and compactness of $\mathcal{X}_0$ yield that $\sup_{x\in\mathcal{X}_0}|[f(x+hu)/f(x)]^{1/2}-1| \le Ch|u|$. Hence $A \le \left|\int CK_j'(u)\,|u|du\right| O_{\mathbb{P}}\left(h^{1/2}\right)$, since $\sup_{x\in\mathcal{X}_0}|W(x)| = O_{\mathbb{P}}(1)$, which gives the result under Assumption K.

A.2 Proof of Theorem 2

The Theorem is proved if we show that

$$\sup_{x\in\mathcal{X}_0}\left\|\widehat{\mathbf{V}}(\theta) - \frac{\alpha(1-\alpha)q(\alpha|x)^2}{f(x)}\mathbf{N}_p^{-1}\mathbf{T}_p\mathbf{N}_p^{-1}\right\| = o_{\mathbb{P}}\left(\log n\right)^{-1/2}.$$

First observe that Einmahl and Mason (2005, Theorem 1) under Assumptions K, H and X yields that

$$\sup_{x\in\mathcal{X}_0}\left|f(x) - \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\right| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{(nh)^{1/2}} + h\right).$$

Then since $\inf_{x\in\mathcal{X}_0} f(x) \ge C > 0$ under Assumption X and $\inf_{x\in\mathcal{X}_0} q(\alpha|x) \ge C > 0$ under Assumption F, the Theorem is proved if we show that

$$\sup_{x\in\mathcal{X}_0}|\widehat{q}(\alpha|x) - q(\alpha|x)| = o_{\mathbb{P}}\left(\log n\right)^{-1/2}.$$

Guerre and Sabbah (2009, Lemmas A.2, A.3 and Theorems 1 and 2) under Assumptions F, H, K, Q and X yields that

$$\sup_{(\alpha,x)\in[\underline{\alpha},\overline{\alpha}]\times\mathcal{X}_0}\left|\widehat{Q}(\alpha|x) - Q(\alpha|x)\right| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{(nh)^{1/2}} + h^{s+1}\right)$$

Now recall that $\widehat{q}(\alpha|x) = \frac{1}{h}\int \widehat{Q}(\alpha+ht|x)dK_q(t)$. Then

$$\begin{aligned}
&\sup_{x\in\mathcal{X}_0}\left|\widehat{q}(\alpha|x) - \frac{1}{h}\int Q(\alpha+ht|x)dK_q(t)\right| \\
&\quad\le \frac{1}{h}\sup_{(\alpha,x)\in[\underline{\alpha},\overline{\alpha}]\times\mathcal{X}_0}\left|\widehat{Q}(\alpha|x) - Q(\alpha|x)\right|\int |dK_q(t)| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{h(nh)^{1/2}} + h^{s}\right).
\end{aligned}$$

Now since

$$\sup_{\substack{t\in\mathcal{K}_q \\ (\alpha,x)\in[\underline{\alpha},\overline{\alpha}]\times\mathcal{X}_0}} |Q(\alpha+ht)-Q(\alpha|x)-htq(\alpha|x)| = O(h),$$

under Assumptions H and Q, and since $\int dK_q(t)=0$ and $\int tdK_q(t)=1$, we have

$$\sup_{x\in\mathcal{X}_0}|\widehat{q}(\alpha|x)-q(\alpha|x)| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{h(nh)^{1/2}}+h^s+h\right) = o_{\mathbb{P}}\left(\log n\right)^{-1/2},$$

under Assumption H.

4.5 Appendix B: Proof of intermediary results

B.1 Proof of Lemma A.1

The vector $\widehat{\mathbf{b}}(\alpha,x)$ is in a first step an estimator of

$$\mathbf{b}^*(\alpha,x) = \arg\min_{\mathbf{b}\in\mathbb{R}^{p+1}}\mathbb{E}\left[\ell_\alpha\left(Y-\mathbf{U}(X-x)^\top\mathbf{b}\right)K\left(\frac{X-x}{h}\right)\right].$$

Under Assumptions F, H, K, Q and X and provided h is small enough we have by Guerre and Sabbah (2009, Theorem 1),

$$\sup_{x\in\mathcal{X}_0}\left|\frac{\mathbf{H}\left(\mathbf{b}^*(\theta)-\mathbf{b}(\alpha|x)\right)}{h^s}\right| \leq C. \tag{B.1}$$

We now present the Bahadur representation given in Guerre and Sabbah (2009). In order to do that, define $Q^*(X_i;\theta)=\mathbf{U}(X_i-x)^\top\mathbf{b}^*(\theta)$ and

$$\begin{aligned}
\mathbf{S}_n(\theta) &= \frac{2}{(nh)^{1/2}}\sum_{i=1}^n\left\{\mathbb{I}\left(Y_i\leq Q^*(X_i;\theta)\right)-\alpha\right\}K_{\mathbf{U}}\left(\frac{X_i-x}{h}\right),\\
\mathbf{J}_n(\theta) &= \frac{2}{nh}\sum_{i=1}^n f\left(Q^*(X_i;\theta)|X_i\right)K_{\mathbf{U}}\left(\frac{X_i-x}{h}\right)\mathbf{U}\left(\frac{X_i-x}{h}\right)^\top,\\
\mathbf{E}_n(\theta) &= (nh)^{1/2}\mathbf{H}\left(\widehat{\mathbf{b}}(\theta)-\mathbf{b}^*(\theta)\right)+\frac{\mathbf{S}_n(\theta)}{\mathbf{J}_n(\theta)}.
\end{aligned}$$

Under Assumptions F, H, K, Q and X, Guerre and Sabbah (2009, Theorem 2) yields that

$$\sup_{x\in\mathcal{X}_0}\|\mathbf{E}_n(\theta)\| = O_{\mathbb{P}}\left(\frac{\log^3 n}{nh}\right)^{1/4}. \tag{B.2}$$

Lemma A.2 *Under Assumptions F, H, K and X and uniformly in x in $\mathcal{X}_0$,*

$$\left\|\mathbf{S}_n(\theta) - \frac{2}{(nh)^{1/2}}\sum_{i=1}^{n}\mathbb{I}(Y_i \leq Q(\alpha|X_i))K_{\mathbf{U}}\left(\frac{X_i - x}{h}\right)\right\| = O_{\mathbb{P}}\left((nh)^{1/2}h^s\right),$$

and $\sup_{x\in\mathcal{X}_0}\|\mathbf{J}(\theta) - \mathbf{J}_n(\theta)\| = O_{\mathbb{P}}\left(h + (\log n)^{1/2}/(nh)^{1/2}\right)$.

We now return to the Proof of Lemma A.1. We consider the Bahadur representation in equation (B.2) so that $(nh)^{1/2}\mathbf{H}(\widehat{\mathbf{b}}(\theta) - \mathbf{b}(\theta))$ can be written as

$$\frac{2\mathbf{J}(\theta)^{-1}}{(nh)^{1/2}}\sum_{i=1}^{n}\left\{\alpha - \mathbb{I}(Y_i \leq Q(\alpha|X_i))\right\}K_{\mathbf{U}}\left(\frac{X_i - x}{h}\right) + \mathbf{r}_n(\theta),$$

with $\mathbf{r}_n(\theta) = \sum_{j=1}^{4}\mathbf{r}_{n,j}(\theta)$, and where

$$\begin{aligned}
\mathbf{J}(\theta)\mathbf{r}_{n,1}(\theta) &= \frac{1}{(nh)^{1/2}}\sum_{i=1}^{n}\left\{\mathbb{I}(Y_i \leq Q(\alpha|X_i)) - \alpha\right\}K_{\mathbf{U}}\left(\frac{X_i - x}{h}\right) - \mathbf{S}_n(\theta),\\
\mathbf{r}_{n,2}(\theta) &= (nh)^{1/2}\mathbf{H}\left(\mathbf{b}^*(\theta) - \mathbf{b}(\alpha|x)\right), \quad \mathbf{r}_{n,3}(\theta) = \frac{\mathbf{J}_n(\theta) - \mathbf{J}(\theta)}{\mathbf{J}(\theta)\mathbf{J}_n(\theta)}\mathbf{S}_n(\theta),
\end{aligned}$$

and $\mathbf{r}_{n,4}(\theta) = \mathbf{E}_n(\theta)$. We now bound the $\mathbf{r}_{n,j}(\theta)$ for $j = 1, 2, 3, 4$, uniformly in x in $\mathcal{X}_0$ showing that $\max_{1\leq j\leq 4}\sup_{x\in\mathcal{X}_0}\|\mathbf{r}_{n,j}(\theta)\| = o_{\mathbb{P}}(\log n)^{-1/2}$. Lemma A.2 and the fact $\inf_{x\in\mathcal{X}_0}\|\mathbf{J}(\theta)\| \geq C > 0$ yield that $\mathbf{r}_{n,1}(\theta)$ is a $O_{\mathbb{P}}((nh)^{1/2}h^s)$ uniformly in x in $\mathcal{X}_0$. Equation (B.1) yields that $\mathbf{r}_{n,2}(\theta)$ is a $O(h^s)$ uniformly in x in $\mathcal{X}_0$. Now observe that Guerre and Sabbah (2009, Lemma A.3) together with the fact that Guerre and Sabbah (2009, Lemma A.2) yields that $\inf_{x\in\mathcal{X}_0}\|\mathbf{J}_n(\theta)\| \geq C > 0$ with a probability that can be arbitrarily large yield that $\mathbf{r}_{n,3}(\theta)$ is a $O_{\mathbb{P}}(\log n/(nh)^{1/2} + h\log^{1/2} n)$ uniformly in x in $\mathcal{X}_0$. Finally (B.2) yields that $\sup_{x\in\mathcal{X}_0}\|\mathbf{E}_n(\theta)\| = O_{\mathbb{P}}\left(\log^3 n/(nh)\right)^{1/4}$. Then

$$(nh)^{1/2}\,\mathbf{H}\left(\widehat{\mathbf{b}}(\theta) - \mathbf{b}(\alpha|x)\right) = -\frac{\mathbf{J}(\theta)^{-1}}{(nh)^{1/2}}\sum_{i=1}^{n}\left\{\mathbb{I}(Y_i \leq Q(\alpha|X_i)) - \alpha\right\} + \mathbf{r}_n(\theta),$$

with

$$\sup_{x\in\mathcal{X}_0} \|\mathbf{r}_n(\theta)\| = O_{\mathbb{P}}\left((nh)^{1/2}h^s + \frac{\log n}{(nh)^{1/2}} + h\log^{1/2} n + \frac{\log^{3/4} n}{(nh)^{1/4}}\right).$$

The result then follows since $(\log n)/(nh)^{1/2} = o\left((\log^{3/4} n)/(nh)^{1/4}\right)$ under Assumption H.

B.2 Proof of Lemma A.2

For simplicity of notations, define $Q^*(u) = Q^*(u;\theta)$ and $Q(u) = Q(\alpha|u)$. A Taylor expansion of $Q(\cdot)$ in a neighborhood of x can be written as

$$Q(x+hz) = \mathbf{U}(hz)\mathbf{b}(\theta) + R_Q(\theta, z),$$

where $\mathbf{b}(\alpha, x)$ is as in Lemma A.1. This gives that

$$\begin{aligned}
&\sup_{(x,z)\in\mathcal{X}_0\times\mathcal{K}} |Q^*(x+hz) - Q(x+hz)| \\
&\quad= \sup_{(x,z)\in\mathcal{X}_0\times\mathcal{K}} |\mathbf{U}(hz)\mathbf{b}^*(\alpha,h,x) - \mathbf{U}(hz)\mathbf{b}(\alpha,x) - R_Q(\theta)| \\
&\quad\le \sup_{z\in\mathcal{K}}\|\mathbf{U}(z)\| \sup_{x\in\mathcal{X}_0} \|\mathbf{H}\left(\mathbf{b}^*(\alpha,h,x) - \mathbf{b}(\alpha,x)\right)| + \sup_{(x,z)\in\mathcal{X}_0\times\mathcal{K}} |R_Q(\theta,z)| \\
&\quad= O(h^s),
\end{aligned} \tag{B.3}$$

by equation (B.1) and under Assumption Q. Define

$$A_{n,j}(\theta) = \sum_{i=1}^n \left\{\mathbb{I}\left(Y_i \le Q^*(X_i;\theta)\right) - \mathbb{I}\left(Y_i \le Q(\alpha|X_i)\right)\right\} K_j\left(\frac{X_i - x}{h}\right).$$

Define $W_i = Y_i - Q(\alpha|X_i)$ and observe that (B.3) yields that

$$\begin{aligned}
|A_{n,j}(\theta)| &\le nh^{s+1}\frac{1}{nh^{s+1}}\sum_{i=1}^n \mathbb{I}\left(|Y_i - Q(\alpha|X_i)| \le Ch^s\right)\left|K_j\left(\frac{X_i-x}{h}\right)\right| \\
&= nh^{s+1}\frac{1}{nh^{s+1}}\sum_{i=1}^n \mathbb{I}\left(|W_i| \le Ch^s\right)\left|K_j\left(\frac{X_i-x}{h}\right)\right| \\
&= nh^{s+1}\widetilde{f}(x,0),
\end{aligned} \tag{B.4}$$

where $\widetilde{f}(x,0)$ is a kernel estimator of the density of the couple (X_i, W_i) at $(x,0)$. Observe that

$$\sup_{x\in\mathcal{X}_0}\left|\widetilde{f}(x,0)\right| \leq \sup_{x\in\mathcal{X}_0}\left|\widetilde{f}(x,0)-\mathbb{E}\left[\widetilde{f}(x,0)\right]\right| + \sup_{x\in\mathcal{X}_0}\left|\mathbb{E}\left[\widetilde{f}(x,0)\right]\right|. \tag{B.5}$$

We now bound the two terms in the RHS of (B.5). Einmahl and Mason (2005, Theorem 1) yields that

$$\sup_{x\in\mathcal{X}_0}\left|\widetilde{f}(x,0)-\mathbb{E}\left[\widetilde{f}(x,0)\right]\right| = O_{\mathbb{P}}\left(\frac{\log^{1/2} n}{(nh^{s+1})^{1/2}}\right), \tag{B.6}$$

under Assumptions F, K and X. The change of variables $z = x + hu$ yields that $\mathbb{E}\left[\widetilde{f}(x,0)\right]$ can be written as

$$\int\int \mathbb{I}\left(|y - Q(\alpha|x+hu)| \leq Ch^s\right)|K_j(u)|\, f(y|x+hu)f(x+hu)dydu\,.$$

It then follows from Assumptions F, K and X that $\sup_{x\in\mathcal{X}_0}\left|\mathbb{E}\left[\widetilde{f}(x,0)\right]\right| \leq C$. This, Assumption H and (B.6) yield that the RHS in (B.5) is a $O_{\mathbb{P}}(1)$ which together with (B.4) proves the first part of the Lemma.

We now prove the second part of the Lemma. For that first note that

$$\sup_{x\in\mathcal{X}_0}\|\mathbf{J}_n(\theta)-\mathbf{J}(\theta)\| \leq \sup_{x\in\mathcal{X}_0}\|\mathbf{J}_n(\theta)-\mathbb{E}\left[\mathbf{J}_n(\theta)\right]\| + \sup_{x\in\mathcal{X}_0}\|\mathbb{E}\left[\mathbf{J}_n(\theta)\right]-\mathbf{J}(\theta)\|\,.$$

We control the first term in the RHS of the above equation. Observe that $\mathbf{J}_n(\theta)$ is the matrix with entries

$$J_{v_1,v_2}(\theta) = \sum_{i=1}^{n}\frac{2}{nh}K_{v_1+v_2}\left(\frac{X_i-x}{h}\right)f\left(Q^*(X_i;\theta)|X_i\right) = \sum_{i=1}^{n}J_{v_1,v_2,i}(\theta),$$

where $0 \leq v_1, v_2 \leq p$. Using Einmahl and Mason (2005, Theorem 1) yields that

$$\sup_{x\in\mathcal{X}_0}\|J_{v_1,v_2}(\theta)-\mathbb{E}\left[J_{v_1,v_2}(\theta)\right]\| \leq C\left(\frac{\log n}{nh}\right)^{1/2},$$

for all $0 \leq v_1, v_2 \leq p$, under Assumptions F, H, K and X. The Lemma is then proved if we show that $\sup_{x\in\mathcal{X}_0}\|\mathbb{E}\left[\mathbf{J}_n(\theta)\right]-\mathbf{J}(\theta)\| = O(h)$. (B.3) and the change of variables $z = x + hu$ then yield that the generic entries of $\|\mathbf{J}_n(\theta) - \mathbf{J}(\theta)\|$ are less than

$$\begin{aligned}
&\int \Bigg(\sup_{(x,u)\in\mathcal{X}_0\times\mathcal{K}} \left| f\left(Q^*(x+hu;\theta)|x+hu\right)\left(f(x+hu)-f(x)\right)\right| \\
&\quad + \sup_{(x,u)\in\mathcal{X}_0\times\mathcal{K}} \left| f(x)\left(f\left(Q^*(x+hu;\theta)|x+hu\right) - f(Q(\alpha|x)|x) f(x)\right)\right| \Bigg) \\
&\quad \times \left|K_{v_1+v_2}(u)\right| du \\
&\quad \le C \int \Bigg(\sup_{(x,u)\in\mathcal{X}_0\times\mathcal{K}} \left| f(x+hu) - f(x)\right| \\
&\quad + \sup_{(x,u)\in\mathcal{X}_0\times\mathcal{K}} \left| f\left(Q^*(x+hu;\theta)|x+hu\right) - f(Q(\alpha|x)|x)\right| \Bigg) \left|K_{v_1+v_2}(u)\right| du \\
&\quad = O(h),
\end{aligned}$$

under Assumption F, X and equation (B.1).

Bibliography

[1] Bickel P.J. & M. Rosenblatt (1973). On some global measures of the deviation of density function estimates. *Ann. Statist.* **1** 1071-1095.

[2] Claeskens G. & van Keilegom I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852-1884.

[3] Einmahl U. & D.M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **3** 1380-1403.

[4] Guerre, E. & C. Sabbah (2009). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Available at :* `econpapers.repec.org/paper/qmwqmwecw/wp648.htm`

[5] Härdle W. & S. Song (2009). Confidence bands in quantile regression. Forthcoming in *Econometric Theory.* Published online by Cambridge University Press 04 Nov 2009.

[6] Rosenblatt M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470-472.

[7] Tusnàdy G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. *Period. Math. Hungar.* **8** 53-55.