



**HAL**  
open science

## Of nonparametric testing in regression

Samuel Maistre

► **To cite this version:**

Samuel Maistre. Of nonparametric testing in regression. Statistics [math.ST]. Université de Rennes, 2014. English. NNT: 2014REN1S057 . tel-01097407

**HAL Id: tel-01097407**

**<https://theses.hal.science/tel-01097407>**

Submitted on 19 Dec 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1  
*sous le sceau de l'Université Européenne de Bretagne*  
pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

*Mention : Mathématiques et applications*

**Ecole doctorale Matisse**

présentée par

**Samuel MAISTRE**

préparée à l'IRMAR (UMR CNRS 6625) et au CREST  
Institut de recherche mathématiques de Rennes  
Centre de recherche en économie et statistique  
École Nationale de la Statistique et de l'Analyse de l'Information

---

**Des tests  
non paramétriques  
en régression**

**Thèse soutenue à l'Ensaï  
le 12 septembre 2014**

devant le jury composé de :

**Patrice BERTAIL**

Professeur, Université Paris X / rapporteur

**Bernard DELYON**

Professeur, Université Rennes 1 / examinateur

**Jan JOHANNES**

Professeur, Université Catholique de Louvain /  
examineur

**Alois KNEIP**

Professeur, Université de Bonn / examinateur

**Pascal LAVERGNE**

Professeur, Toulouse School of Economics / co-  
directeur de thèse

**Valentin PATILEA**

Professeur, ENSAI / directeur de thèse



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Les tests d'adéquation à un modèle . . . . .	5
1.2	Test d'adéquation à un modèle paramétrique . . . . .	6
1.2.1	Approche par processus empirique . . . . .	7
1.2.2	Approche par lissage à noyau . . . . .	8
1.2.3	Comparaison visuelle des approches . . . . .	9
1.3	Les tests de significativité en régression non paramétrique . . . . .	10
1.4	La régression quantile . . . . .	13
1.4.1	Introduction . . . . .	13
1.4.2	Les tests d'adéquation à un modèle pour la régression quantile	14
1.5	Les données fonctionnelles en régression . . . . .	15
1.5.1	Considérations simples pour les variables fonctionnelles . . . . .	15
1.5.2	Bases fonctionnelles . . . . .	15
1.5.3	Les modèles linéaires fonctionnels . . . . .	17
1.6	Le modèle single-index . . . . .	19
<b>2</b>	<b>Test de significativité en régression non paramétrique</b>	<b>25</b>
2.1	Introduction . . . . .	28
2.2	Testing Framework and Procedure . . . . .	29
2.2.1	Testing Principle . . . . .	29
2.2.2	The Test . . . . .	30
2.3	Theoretical Properties . . . . .	32
2.3.1	Assumptions . . . . .	32
2.3.2	Asymptotic Analysis . . . . .	33
2.3.3	Bootstrap Critical Values . . . . .	34
2.4	Monte Carlo Study . . . . .	35
2.5	Conclusion . . . . .	36
2.6	Proofs . . . . .	37
2.6.1	Proof of Theorem 2.6 . . . . .	37
2.6.2	Proof of Corollary 2.7 . . . . .	42
2.6.3	Proof of Theorem 2.8 . . . . .	43
<b>3</b>	<b>Test d'adéquation pour la régression quantile</b>	<b>67</b>
3.1	Introduction . . . . .	70
3.2	Lack-of-Fit Test for Quantile Regression . . . . .	71
3.2.1	Principle and Test . . . . .	71

3.2.2	Behavior Under the Null Hypothesis . . . . .	73
3.2.3	Behavior under Local Alternatives . . . . .	74
3.2.4	Bootstrap Critical Values . . . . .	75
3.3	Numerical Evidence . . . . .	76
3.3.1	Small Sample Performances . . . . .	76
3.3.2	Empirical Illustration . . . . .	77
3.4	Proofs . . . . .	78
3.4.1	Proof of Theorem 3.5 . . . . .	78
3.4.2	Proof of Theorem 3.6 . . . . .	80
3.4.3	Proof of Theorem 3.7 . . . . .	86
<b>4</b>	<b>Test d'adéquation pour la régression de variables fonctionnelles</b>	<b>95</b>
4.1	Introduction . . . . .	98
4.2	The method . . . . .	100
4.2.1	The idea behind the testing method . . . . .	101
4.2.2	The test statistics . . . . .	102
4.3	Asymptotic theory . . . . .	103
4.3.1	The asymptotic critical values . . . . .	103
4.3.2	The consistency of the test . . . . .	104
4.3.3	Goodness-of-fit test . . . . .	105
4.3.4	Bootstrap critical values . . . . .	106
4.4	The error in covariates case . . . . .	106
4.5	Empirical evidence . . . . .	109
4.5.1	The scalar response case . . . . .	109
4.5.2	The case of functional response . . . . .	111
4.5.3	Real data application . . . . .	114
4.6	Technical results and proofs . . . . .	117
<b>5</b>	<b>Test d'adéquation pour les modèles "single-index"</b>	<b>127</b>
5.1	Introduction . . . . .	130
5.2	Single-index models . . . . .	131
5.3	A general approach for testing nonparametric significance . . . . .	132
5.4	Single-index assumptions checks . . . . .	133
5.4.1	Testing SIM for mean regression . . . . .	134
5.4.2	Testing SIM for the conditional law . . . . .	135
5.4.3	Asymptotic results . . . . .	136
5.5	Empirical evidence . . . . .	137
5.6	Conclusions and further extensions . . . . .	140
5.7	Appendix 1: assumptions and proofs . . . . .	144
5.8	Appendix 2: technical lemmas . . . . .	153
	<b>Bibliographie</b>	<b>167</b>

## Remerciements

Je remercie tout d'abord les membres du jury, en commençant par mes deux rapporteurs, les professeurs Patrice Bertail et Holger Dette, qui m'ont fait l'honneur de relire mon manuscrit et dont les relectures attentives m'ont permis de corriger quelques erreurs, et de me donner d'ores et déjà des pistes pour la suite. Je remercie aussi les autres membres du jury, les professeurs Bernard Delyon, Alois Kneip et Jan Johannes, d'avoir accepté de venir à ma soutenance. Enfin, je remercie mes directeurs de thèse, Pascal Lavergne et Valentin Patilea, qui m'ont supporté dans tous les sens du terme durant ces quatre dernières années.

J'aimerais également remercier tous mes collègues de l'Ensaï, les économistes, les informaticiens et bien sûr les statisticiens, en particulier ceux avec qui j'ai partagé mon bureau, Daniel et Cyril ; l'ensemble de l'équipe de statistique de Rennes, Agro-campus, ENS, Inria, Rennes 1 et Rennes 2, notamment François et Quentin avec qui j'ai partagé plus que la thématique de la réduction de la dimension. Je remercie aussi l'équipe bruxelloise de l'ULB pour leur soutien ainsi que l'ensemble du personnel de l'Ensaï.

Sur un plan plus personnel, je remercie ceux qui m'ont permis de conserver un semblant de vie sociale avec par ordre chronologique mes anciens pseudo-colocs Axel, Marine, Béranger, Nico ; les internes des promos 2012 et 2013 avec une mention toute particulière à Luc et Marie-Cécile puis Renaud et Florine ; les doctorants, Émilie, Julie, Cyril, Mélisande, Émeline et Gaspar (ainsi que Cécile la stagiaire) ; Salima, Noëlle, Thomas et ses collègues de l'équipe Visages ; ceux qui sont venus me voir plus ponctuellement, Clélie, Adrienne, François, Nidal, Cyril, Romain, Steven, Thomas, Julien, Fanny, Pépèl, Benjamin, Dimitri et Samuel (malgré la porte ouverte), Jeanne et Damien, mais également ceux que j'ai vu en dehors de Rennes, je pense bien sûr à tous les cajistes parisiens, aux ex-grenoblois(es), aux toulousains et aux martins en général, ainsi qu'aux Lugduno-villardo-chavannais, avec une mention spéciale à la classe de 87.

Pour terminer, je remercie ma famille, tout ceux qui ont suivi de près ou de loin mon parcours, et en particulier mes parents, ma sœur Lucie (et ses amis caennais, mâconnais ou lyonnais), Xavier, mon frère Alexandre, Élise, Juliette et Hélène.



# 1 Introduction





## 1.1 Les tests d'adéquation à un modèle

Une partie très importante de la statistique consiste à faire une hypothèse de modélisation puis à estimer le modèle. Le nombre d'applications de ce paradigme est infini. Ne prenons que quelques exemples :

- si l'on observe un phénomène qui est amené à se reproduire, on peut chercher à poser un modèle sur le phénomène pour prédire l'avenir. En sachant le temps qu'il fait aujourd'hui (ou plus précisément si l'on connaît les précipitations, la pression atmosphérique, l'hygrométrie, les images satellites, etc.) on peut essayer de prédire le temps qu'il fera demain en modélisant l'effet de ces variables sur le temps observé ;
- si l'on veut connaître les habitudes de consommation (consommation énergétique, production de déchets, ...) de la population française, on peut sonder une partie de la population et modéliser leur comportement en fonction des caractéristiques du foyer (composition, revenus, logement, ...), puis inférer un comportement sur l'ensemble de la population (à condition bien sûr de connaître les caractéristiques de tous les foyers) ;
- si l'on veut connaître l'effet d'un engrais sur le rendement d'une culture, on peut choisir entre :
  - une modélisation monotone (par exemple linéaire) : plus on met d'engrais et plus la production est élevée ;
  - une modélisation en forme de cloche (par exemple quadratique) : passé un certain seuil, l'ajout d'engrais diminue la production.

Dans ce deuxième cas, la modélisation permet d'estimer une quantité d'engrais optimale au sens du rendement.

En général, on dispose d'une part d'une variable réponse que l'on notera  $Y$ , qui pourra être scalaire, vectorielle, voire dans un espace plus général. Ici, si rien n'est précisé,  $Y$  sera scalaire. D'autre part, on est en présence de covariables explicatives que l'on notera  $X$  et qui appartiennent à un espace  $\mathcal{X}$ , par exemple  $\mathbb{R}^p$  ou  $L^2([0, 1])$ . On cherche donc à modéliser  $Y$  en fonction de  $X$ . Communément, on pose

$$Y = r(X) + \varepsilon$$

avec  $r$  appartenant à un certain espace de fonctions  $E$  de dimension finie (dans le cas paramétrique) ou infinie, par exemple :

- $E = \{(w, z) \mapsto m(w)\}$  – dans ce cas  $X = (W, Z)$  – pour la significativité de variables ;
- $E = \{x \mapsto g(x, \theta) : \theta \in \Theta\}$  avec  $g$  fixée pour un modèle paramétrique ;
- $E = \{x \mapsto m(v(x, \theta)) : \theta \in \Theta, m : \mathbb{R} \rightarrow \mathbb{R}\}$  avec  $v(\cdot, \cdot)$  connue, typiquement  $v(x, \theta) = x'\theta$ , pour un modèle single-index.

On suppose également une certaine « centralité » de la variable  $\varepsilon$  vis à vis de  $X$ . Dans le cas de la régression, c'est  $\mathbb{E}[\varepsilon | X] = 0$ . Lorsque la modélisation est paramétrique et pour  $Y \in \mathbb{R}$  et  $X \in \mathbb{R}^p$ , on utilise par exemple le modèle linéaire général

$$E = \left\{ x \mapsto a + x'b : (a, b)' \in \Theta \subset \mathbb{R}^{p+1}, x \in \mathbb{R}^p \right\}$$

et pour  $Y \in \{0, 1\}$  le modèle de régression logistique

$$E = \left\{ x \mapsto \frac{1}{1 + \exp\{-a - x'b\}} : (a, b)' \in \Theta \subset \mathbb{R}^{p+1}, x \in \mathbb{R}^p \right\}.$$

Sur la base d'un échantillon de taille  $n \in \mathbb{N}^*$  de ce couple de variables, que l'on notera  $\mathcal{Z}_n = \{(X_i, Y_i), 1 \leq i \leq n\}$ , on approxime la fonction  $r(\cdot)$  en estimant les paramètres de dimension finie  $\theta$  ou infinie  $m$ .

Une étape importante de la modélisation est de vérifier si l'hypothèse du modèle utilisé est bien valide. En reprenant l'exemple du rendement agricole en fonction de la quantité d'engrais, on peut se demander si une modélisation linéaire explique bien la réalité ou si les causes qui sous-tendent le phénomène sont plus complexes. On cherche donc dans un cadre très général à tester

$$(H_0) : r \in E \quad \text{contre} \quad (H_1) : r \notin E.$$

Notons dès à présent que pour la régression, si l'on pose

$$\tilde{r} = \arg \min_{r \in E} \mathbb{E} \left[ (Y - r(X))^2 | X \right]$$

et  $U = Y - \tilde{r}(X)$ , cela s'écrit comme un test de moment conditionnel

$$(H_0) : \mathbb{E}[U | X] = 0 \text{ p.s.} \quad \text{contre} \quad (H_1) : \mathbb{P}\{\mathbb{E}[U | X] = 0\} < 1.$$

La plupart du temps, on a besoin d'estimer le modèle pour faire le test : pour le paramètre de dimension finie  $\theta$ , on pourra en général compter sur un estimateur  $\hat{\theta}_n$  tel que  $\hat{\theta}_n - \theta = O_{\mathbb{P}}(n^{-1/2})$ .

Nous allons maintenant présenter dans le cadre de la régression paramétrique les deux approches principales, approches qui sont d'autre part à la base des statistiques étudiées dans cette thèse : l'approche par processus empirique et l'approche par lissage à noyau.

## 1.2 Un cas d'école : le test d'adéquation à un modèle paramétrique

Prenons comme exemple simple un modèle linéaire simple avec  $(X, Y) \in \mathbb{R}^2$  un couple de variables aléatoires. On définit l'espérance linéaire de  $Y$  sachant  $X$ , comme  $\beta_0 + \beta_1 X$ , où

$$(\beta_0, \beta_1)' = \arg \min_{a, b} \mathbb{E} \left[ (Y - a - bX)^2 \right]$$

et on peut ensuite définir  $U = Y - \beta_0 + \beta_1 X$ . L'hypothèse paramétrique de linéarité est

$$\mathbb{E}[U \mid X] = 0 \text{ p.s.} \quad (1.1)$$

Ce modèle est très simple et sera utilisé en guise d'illustration graphique, mais de la même manière, on peut à partir de n'importe quel modèle paramétrique définir le paramètre

$$\theta = \arg \min_{\beta \in \Theta} \mathbb{E} \left[ (Y - g(X, \beta))^2 \right],$$

sous certaines conditions (sur  $g(\cdot)$  et  $\Theta$ ) permettant son identifiabilité, et  $U = Y - g(X, \theta)$ , puis tester l'hypothèse (1.1).

### 1.2.1 Approche par processus empirique

Nous présentons deux statistiques utilisant l'approche par processus empirique : Bierens (1982) et Stute et al. (1998b). Dans son article de 1982, Bierens propose d'utiliser la statistique  $nI_n^B(\hat{\theta}_n)$  avec

$$\begin{aligned} I_n^B(\hat{\theta}_n) &= \int_{N_0} \left| \frac{1}{n} \sum_{j=1}^n (Y_j - g(X_j, \hat{\theta}_n)) \exp\{it' \phi(X_j)\} \right|^2 dt \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{U}_i \hat{U}_j \int_{N_0} \exp\{it'(\phi(X_i) - \phi(X_j))\} dt \end{aligned}$$

où  $\hat{U}_j = Y_j - g(X_j, \hat{\theta}_n)$ ,  $\phi(\cdot)$  est une fonction bijective bornée de  $\mathbb{R}^p \rightarrow \mathbb{R}^p$  et  $N_0$  est un voisinage de l'origine choisi arbitrairement petit.

La statistique proposée par Stute et al. (1998b) se base sur le processus de somme cumulée des résidus

$$\tilde{R}_n(x, \hat{\theta}_n) = n^{-1/2} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} [Y_i - g(X_i, \hat{\theta}_n)]$$

et sur une transformation linéaire  $T$ , puis à construire une statistique de Cramér-von Mises sur  $T_n \tilde{R}_n$ , où  $T_n$  est une version empirique de  $T$ . Remarquons que si l'on prend  $T$  comme l'identité<sup>1</sup>, la statistique de Cramér-von Mises basée sur  $T_n \tilde{R}_n$  n'est

<sup>1</sup>Le choix donné dans l'article permet d'obtenir le fait que  $T \tilde{R}_n$  soit pivotale, ou encore mieux, une martingale. Mais quitte à estimer les valeurs critiques numériquement, le choix présenté permet simplifier les choses dans le cadre de cette présentation.

autre que

$$\begin{aligned} I_n^{(STZ, CVM)}(\hat{\theta}_n) &= \frac{1}{n} \sum_{k=1}^n \tilde{R}_n^2(X_k, \hat{\theta}_n) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{U}_i \hat{U}_j \left( \sum_{k=1}^n \mathbf{1}_{\{X_i \leq X_k\}} \mathbf{1}_{\{X_j \leq X_k\}} \right) \end{aligned}$$

qui ressemble beaucoup à la statistique de Bierens, modulo le choix de la fonction  $\psi_n$  dans la statistique générale de la forme

$$I_n(\hat{\theta}_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{U}_i \hat{U}_j \psi_n(X_i, X_j).$$

À partir de ces statistiques, on montre en général que du fait de la vitesse de convergence de  $\hat{\theta}_n$  vers  $\theta$ , on a  $I_n(\hat{\theta}_n)$  qui se comporte comme  $I_n(\theta)$  lorsque  $n \rightarrow +\infty$ . On cherche ensuite à donner un caractère pivotale à  $I_n(\theta)$  ou à estimer numériquement les valeurs critiques.

Nous allons voir dans la section suivante que d'autres statistiques proposées dans la littérature peuvent se mettre sous cette forme, en faisant dépendre  $\psi_n$  d'un paramètre de lissage.

### 1.2.2 Approche par lissage à noyau

L'une des premières occurrences de l'utilisation du lissage à noyau dans le but de construire un test non paramétrique d'adéquation à un modèle de régression paramétrique se trouve dans l'article de Härdle et Mammen (1993).

Pour vérifier l'hypothèse (1.1) sur la base de l'échantillon  $\mathcal{Z}_n$ , les auteurs proposent de comparer la modélisation linéaire à un lissage non paramétrique, respectivement en bleu pointillé et rouge plein sur la Figure (1.1). Une approche raisonnable serait de se baser sur l'aire comprise entre les deux courbes. Härdle et Mammen proposent la statistique suivante :

$$\begin{aligned} T_n^{HM}(\theta) &= nh^{p/2} \int (\hat{m}_h(x) - \mathcal{K}_{h,n} g(x, \theta))^2 \pi(x) dx \\ &= nh^{p/2} \int \left( \frac{\sum_{i=1}^n (Y_i - g(X_i, \theta)) K((x - X_i)/h)}{\sum_{i=1}^n K((x - X_i)/h)} \right)^2 \pi(x) dx. \end{aligned}$$

Pour

$$\pi(x) = \hat{f}^2(x) = \left( \frac{1}{nh^p} \sum_{i=1}^n K((x - X_i)/h) \right)^2,$$

on a<sup>2</sup> en posant  $\hat{U}_i = Y_i - g(X_i, \hat{\theta})$

$$\begin{aligned} T_n^{HM}(\hat{\theta}) &= \frac{h^{p/2}}{n} \int \left( \sum_{i=1}^n (Y_i - g(X_i, \hat{\theta})) h^{-p} K((x - X_i)/h) \right)^2 dx \\ &= \frac{h^{p/2}}{n} \sum_{i=1}^n \sum_{j=1}^n \hat{U}_i \hat{U}_j \int h^{-2d} K((x - X_i)/h) K((x - X_j)/h) dx \\ &= \frac{h^{p/2}}{n} \sum_{i=1}^n \sum_{j=1}^n \hat{U}_i \hat{U}_j h^{-p} \tilde{K}((X_i - X_j)/h) \end{aligned}$$

où  $\tilde{K}(y) = \int K(x - y) K(x) dx$ . En particulier notons que  $\tilde{K} = K * K$  pour un noyau symétrique, et que la convolée d'un noyau standard par lui-même peut donner un autre noyau standard : gaussien  $\rightarrow$  gaussien, uniforme  $\rightarrow$  triangle. En reprenant un noyau quelconque pour  $\tilde{K}$  et en montrant que la quantité  $nh^{p/2} \sum_{i=1}^n \hat{U}_i^2 h^{-p} \tilde{K}(0)$  joue un rôle parasite dans la distribution asymptotique, on peut la retirer et on obtient à un facteur près la statistique proposée par Zheng (1996) :

$$\begin{aligned} T_n^{Zh}(\hat{\theta}) &= \frac{h^{p/2}}{n-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \hat{U}_i \hat{U}_j h^{-p} \tilde{K}((X_j - X_i)/h) \\ &= \frac{1}{n-1} \left( n T_n^{HM}(\hat{\theta}) - h^{p/2} \sum_{i=1}^n \hat{U}_i^2 h^{-p} \tilde{K}(0) \right). \end{aligned}$$

Cette dernière est à la base des développements de Fan et Li (1996) et Lavergne et Vuong (2000) pour tester la significativité de variables en régression non paramétrique.

### 1.2.3 Comparaison visuelle des approches

Pour illustrer graphiquement à quoi correspondent le processus empirique  $\tilde{R}_n$  et le lissage, on génère  $n = 150$  observations indépendantes de  $X \sim \mathcal{U}([0, 1])$  et  $\varepsilon \sim \mathcal{N}(0, 1)$  indépendant de  $X$ , et on calcule les  $Y_i^{(j)}$ ,  $j = 1, 2, 3$ , correspondant aux trois modèles

$$Y_i^{(1)} = 1 - 2X_i + \varepsilon_i, \tag{1.2}$$

$$Y_i^{(2)} = 4X_i - 6X_i^2 + \varepsilon_i, \tag{1.3}$$

$$Y_i^{(3)} = \frac{\pi}{3} \sin(2\pi X_i) + \varepsilon_i, \tag{1.4}$$

<sup>2</sup>Ce choix pour  $\pi$  permet de simplifier l'écriture de la statistique en supprimant les dénominateurs. Dans l'article, aucun choix n'est préconisé, même dans l'étude empirique.

avec  $1 \leq i \leq n$ , ceux-ci ayant été calibrés afin que l'ajustement linéaire soit le même dans les trois cas :

$$\arg \min_{a,b} \mathbb{E} \left[ \left( Y^{(j)} - a - bX \right)^2 \right] = (1, -2)'.$$

On calcule sur les trois échantillons  $\left\{ (X_i, Y_i^{(j)}), 1 \leq i \leq 150 \right\}$  avec  $j \in \{1, 2, 3\}$  l'ajustement linéaire d'une part et un lissage non paramétrique à noyau d'autre part.

On représente sur la première ligne de la Figure 1.1 les nuages de points, droites de régression obtenues en bleu pointillé et le lissage en rouge. Pour le lissage, la courbe est

$$x \mapsto \frac{\sum_{k=1}^n Y_k^{(j)} K\left(\frac{x - X_k}{h}\right)}{\sum_{k=1}^n K\left(\frac{x - X_k}{h}\right)}, \quad j = 1, 2, 3$$

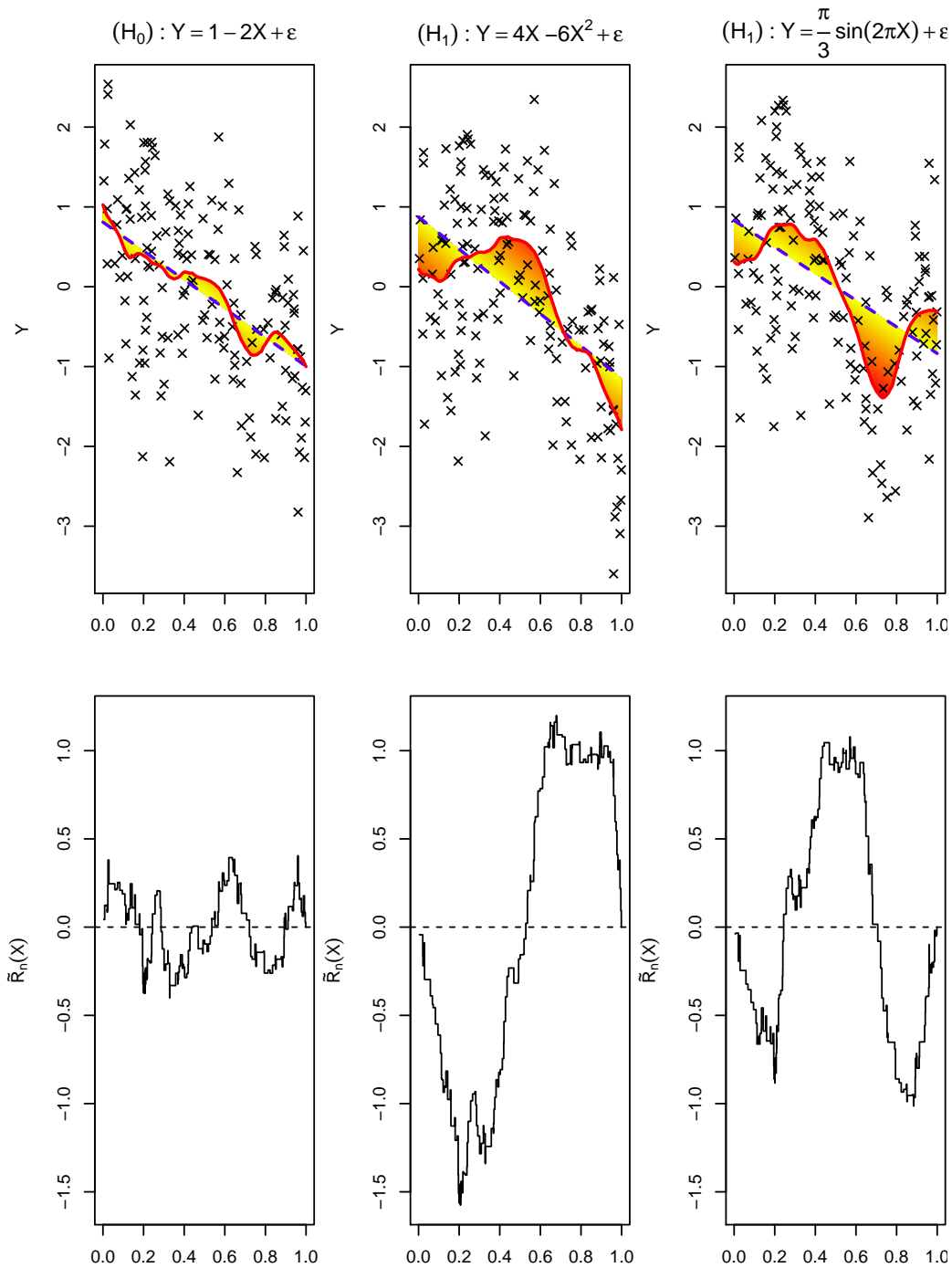
avec  $K(x) = \frac{3}{4}(1 - x^2) \mathbf{1}_{[-1,1]}(x)$   
 et  $h = \frac{7}{2}n^{-1/5}$ ,

$7/2$  étant l'écart-type de  $X$ . Pour le modèle (1.2) représenté sur la partie la plus à gauche, l'hypothèse de linéarité est vérifiée et le lissage est très proche de la droite. Dans les deux autres cas (alternatives quadratique et sinusoidale) le lissage s'éloigne de la droite et une bonne mesure de l'éloignement pourrait être l'aire colorée. On représente ici un dégradé afin de laisser à penser que plus le lissage s'éloigne de la courbe et plus on a envie de rejeter la linéarité. En effet, on a vu que la statistique de Härdle et Mammen (1993) utilise en quelque sorte l'intégrale du carré de la distance entre la courbe et la droite.

Sur la deuxième ligne, on représente le processus empirique utilisé par Stute et al. (1998b). Que l'on considère une statistique de Kolmogorov-Smirnov (amplitude maximale du processus) ou de Cramér-von Mises (aire sous le carré du processus), on voit bien que sa valeur sera plus grande pour les deux et troisième cas que pour le premier.

### 1.3 Les tests de significativité en régression non paramétrique

En régression paramétrique, il peut paraître facile de tester la significativité de variables : il suffit de tester la nullité de tous les paramètres qui concernent l'ensemble des sous-variables dont on veut tester le non-effet. La tâche se complique en régression non paramétrique puisque, cela va sans dire, nous n'avons pas de paramètres



**FIGURE 1.1** – Illustration des deux approches de test : lissage sur la première et processus empirique sur la seconde.

En colonne, trois cas : hypothèse nulle, alternatives quadratique et sinusoidale.



sur lesquels faire des tests. Mais on a vu dans la section précédente plusieurs idées permettant de résoudre ce problème : si l'on est en présence de deux variables seulement et que l'on veut tester la significativité de la seconde, rien ne nous empêche de calculer les résidus dans un modèle contenant seulement la première variable, puis de construire des processus empiriques ou des quantités lissées dépendant de ces résidus. Ce cas particulier de test est traité dans le Chapitre 2.

Lorsque l'on veut tester la significativité d'une variable  $X$  dans la régression de  $Y$  sachant  $W \in \mathbb{R}^p$  et  $X \in \mathbb{R}^q$ , on a donc le problème

$$(H_0) : \mathbb{E}[Y | W, X] = \mathbb{E}[Y | W]$$

qui s'écrit également

$$(H_0) : \mathbb{E}[U | W, X] = 0$$

avec  $U = Y - \mathbb{E}[Y | W]$ . Il paraît donc essentiel d'avoir des estimations de  $U$  sur les observations. Il existe bien sûr de nombreuses techniques non paramétriques qui permettent de les obtenir : lissage à noyau, plus proches voisins et bases de fonctions (splines, ondelettes, Fourier). Jusqu'à maintenant, les statistiques proposées pour l'estimation des résidus se sont concentrées sur le lissage à noyau : Delgado et González Manteiga (2001) pour les processus empiriques, et Fan et Li (1996) et Lavergne et Vuong (2000) pour une deuxième utilisation du lissage à noyau. On construit donc les statistiques avec

$$\begin{aligned} \hat{U}_i &= Y_i - \frac{\sum_{\substack{k=1 \\ k \neq i}}^n Y_k L\left(\frac{W_i - W_k}{g}\right)}{\sum_{\substack{k=1 \\ k \neq i}}^n L\left(\frac{W_i - W_k}{g}\right)} \\ &= \frac{\sum_{\substack{k=1 \\ k \neq i}}^n (Y_i - Y_k) L\left(\frac{W_i - W_k}{g}\right)}{\sum_{\substack{k=1 \\ k \neq i}}^n L\left(\frac{W_i - W_k}{g}\right)}. \end{aligned}$$

Afin de s'affranchir des problèmes posés par le dénominateur, les trois statistiques citées plus haut travaillent sur les quantités  $\hat{U}_i \hat{f}_i$  avec

$$\hat{f}_i = (n-1)^{-1} \sum_{\substack{k=1 \\ k \neq i}}^n g^{-p} L\left(\frac{W_i - W_k}{g}\right)$$

l'estimation de la densité de  $W$  prise en  $W_i$ , ce qui donne

$$\hat{U}_i \hat{f}_i = (n-1)^{-1} \sum_{\substack{k=1 \\ k \neq i}}^n (Y_i - Y_k) g^{-p} L\left(\frac{W_i - W_k}{g}\right).$$

La statistique de Fan et Li (1996) est exactement la reprise de celle de Zheng (1996), en remplaçant les  $\hat{U}_i$  par  $\hat{U}_i \hat{f}_i$  :

$$\begin{aligned} T_n^{FL} &= \frac{h^{p/2}}{n-1} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \hat{U}_i \hat{f}_i \hat{U}_j \hat{f}_j h^{-p} K((X_j - X_i)/h) \\ &= \frac{h^{p/2}}{(n-1)^3} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j}}^n (Y_i - Y_k)(Y_l - Y_i) L_{g,ik} L_{g,jl} h^{-p} K((X_j - X_i)/h) \end{aligned}$$

avec  $L_{g,ik} = g^{-p} L\left(\frac{W_i - W_k}{g}\right)$ .

La statistique de Lavergne et Vuong est la reprise de celle de Fan et Li (1996) en remplaçant la quadruple somme par une somme sur les arrangements  $(i, j, k, l)$  de quatre éléments parmi  $n$ .

## 1.4 La régression quantile

### 1.4.1 Introduction

La régression quantile consiste à modéliser un quantile d'ordre  $\tau \in ]0, 1[$  fixé de la variable  $Y$  conditionnellement à  $X$ . En rappelant que le quantile conditionnel de  $Y$  sachant  $X$  est défini par

$$\begin{aligned} Q_\tau(Y | x) &= \inf \{y : F_Y(y | X = x) \geq \tau\}, \\ \text{où } F_Y(y | X = x) &= \mathbb{P}(Y \leq y | X = x), \end{aligned}$$

une régression quantile paramétrique consiste à supposer que pour une fonction  $g(\cdot, \cdot)$  fixée, il existe un  $\theta \in \Theta$  tel que  $Q_\tau(Y | x) = g(x, \theta)$ . La condition dite de « centralité » de la variable  $\varepsilon$  vis à vis de  $X$  dans l'égalité  $Y = g(x, \theta) + \varepsilon$  est ici  $Q_\tau(\varepsilon | X) = 0$ . La fonction « check »

$$\begin{aligned} \rho_\tau(t) &= (\tau - 1)t \mathbf{1}\{t < 0\} + \tau t \mathbf{1}\{t \geq 0\} \\ &= t(\tau - \mathbf{1}\{t < 0\}) \end{aligned}$$

joue un rôle important sans le cadre de ce type de modélisation. En effet, peut aussi définir le quantile d'ordre  $\tau$  de  $Y$  par

$$\arg \min_a \mathbb{E}[\rho_\tau(Y - a)]$$

et si la régression classique peut être estimée en utilisant les moindres carrés, ici  $\theta$  peut être estimé en minimisant  $L(\theta) = \sum_{i=1}^n \rho_\tau(Y_i - g(X_i, \theta))$ , en se basant sur un échantillon i.i.d.  $\{(X_i, Y_i), 1 \leq i \leq n\}$  du couple  $(X, Y)$ .

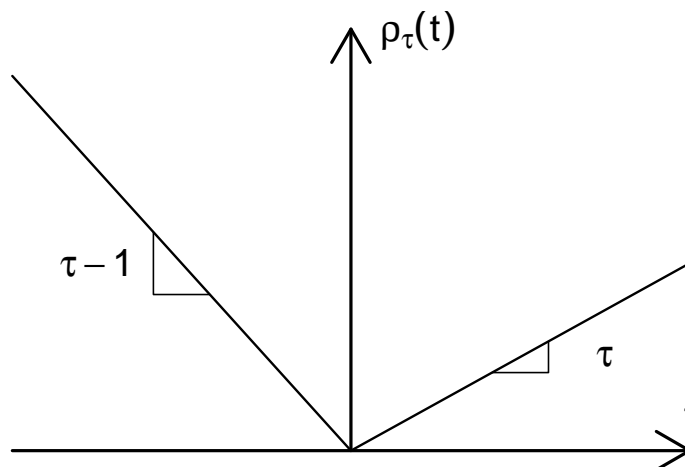


FIGURE 1.2 – Fonction « check ».

On représente cette fonction sur la Figure 1.2. C'est la fonction affine par morceaux de pente  $\tau - 1$  pour  $x < 0$  et de pente  $\tau$  pour  $x > 0$ , ayant 0 pour ordonnée à l'origine. Elle vaut donc  $|x|/2$  pour  $\tau = 1/2$ , c'est à dire pour la régression médiane. La fonction à optimiser  $L$  est donc dérivable presque partout sauf en  $\theta$  tel qu'il existe  $i$  avec  $Y_i = g(X_i, \theta)$ . En dérivant cette fonction partout ailleurs, on obtient

$$L'(\theta) = (\tau - 1) \sum_{i=1}^n \mathbf{1}\{Y_i - g(X_i, \theta) < 0\} + \tau \sum_{i=1}^n \mathbf{1}\{Y_i - g(X_i, \theta) \geq 0\}.$$

L'optimum est donc atteint lorsqu'une proportion  $1 - \tau$  des résidus  $Y_i - g(X_i, \theta)$  est négative et l'autre partie positive (en proportion  $\tau$ ).

### 1.4.2 Les tests d'adéquation à un modèle pour la régression quantile

Comme pour la régression « moyenne », on peut tester l'adéquation de l'échantillon au modèle paramétrique envisagé, ou la significativité de variables. Dans la littérature, citons He et Zhu (2003) pour les processus empiriques et Zheng (1998) pour le lissage à noyau.

## 1.5 Les données fonctionnelles en régression

### 1.5.1 Considérations simples pour les variables fonctionnelles

Considérons un échantillon  $\{X_i, 1 \leq i \leq n\}$  de fonctions définies sur un intervalle fermé, que l'on supposera sans perte de généralité être l'intervalle  $[0, 1]$ . Chaque  $X_i$  ne peut être observé que sur une grille de cet intervalle. La grille peut être différente pour chacune des observations mais la plupart du temps c'est la même. Une grille particulière est la suivante :  $\left\{t_k = \frac{2k-1}{2n_T}, 1 \leq k \leq n_T\right\}$ . Elle permet en effet d'estimer facilement les intégrales par la méthode du point milieu. En effet,

$$\begin{aligned} \int_0^1 X(t) dt &= \sum_{k=1}^{n_T} \int_{(k-1)/n_T}^{k/n_T} X(t) dt \\ &\simeq \left(\frac{k}{n_T} - \frac{k-1}{n_T}\right) \sum_{k=1}^{n_T} X\left(\left(\frac{k}{n_T} + \frac{k-1}{n_T}\right)/2\right) \\ &\simeq \frac{1}{n_T} \sum_{k=1}^{n_T} X\left(\frac{2k-1}{2n_T}\right) \end{aligned}$$

et l'estimation de l'intégrale est donc la moyenne des  $X(t_k)$ . L'erreur commise est du type

$$\begin{aligned} \int_0^1 X(t) dt - \frac{1}{n_T} \sum_{k=1}^{n_T} X\left(\frac{2k-1}{2n_T}\right) &= \frac{1}{24n_T^3} \sum_{k=1}^{n_T} X''(\xi_k) \\ \xi_k &\in \left[\frac{k-1}{n_T}, \frac{k}{n_T}\right] \\ &\leq \frac{1}{24n_T^2} \sup_{t \in [0,1]} |X''(t)| \end{aligned}$$

lorsque  $X$  est deux fois différentiable<sup>3</sup>. On utilisera par la suite la notation  $X^{(n_T)} = (X(t_k))_{1 \leq k \leq n_T}$ .

### 1.5.2 Bases fonctionnelles

Une autre manière de représenter les fonctions est d'utiliser les projections sur une base fonctionnelle particulière : on considère que la fonction  $X$  est une combinaison

<sup>3</sup>Lorsque les trajectoires de  $X$  ne sont pas différentiables, l'erreur aura un ordre différent. Par exemple, lorsque  $X$  est un mouvement brownien sur  $[0, 1]$  on peut facilement montrer que la variance de la différence entre son intégrale et l'estimation proposée est de l'ordre de  $n_T^{-1}$ .

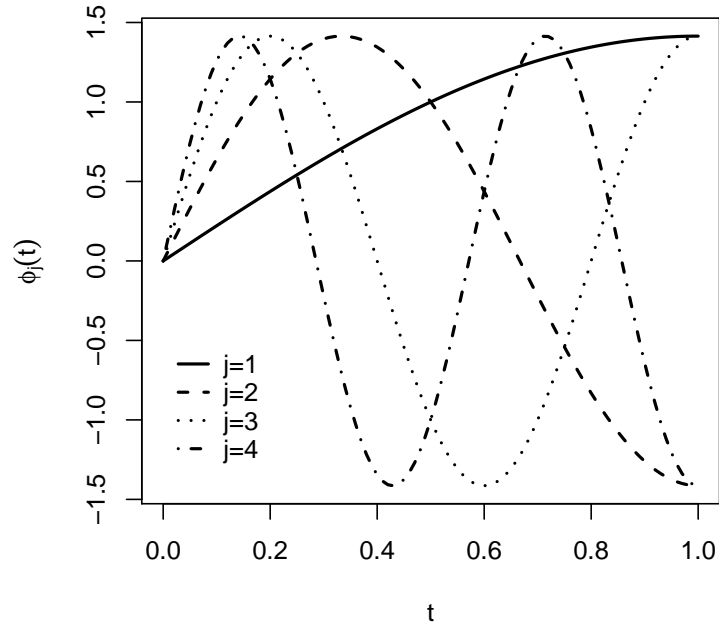


FIGURE 1.3 – Premiers éléments de la base de Fourier.

linéaire des éléments de la base  $X(t) = \sum_{j=1}^{\infty} c_j \phi_j(t)$  avec  $\{\phi_j, j \geq 1\}$  les éléments connus de la base. Pour des raisons évidentes d'utilisation de cette décomposition dans les calculs ultérieurs, on tronque cette somme à partir d'un indice  $J$  ce qui donne  $X(t) = \sum_{j=1}^J c_j \phi_j(t)$ . Lorsque la base est orthogonale, l'estimation des paramètres se fait naturellement en calculant les produits scalaires  $c_j = \langle X, \phi_j \rangle$ , et si l'on prend le design  $X^{(n_T)}$  et la matrice  $\Phi = (\phi_j(t_k))_{1 \leq j \leq J, 1 \leq k \leq n_T}$  cela donne  $\hat{c} = n_T^{-1} \Phi X^{(n_T)} \in \mathbb{R}^J$ .

### 1.5.2.1 Base de Fourier

La famille  $\{\sqrt{2} \sin((j-1/2)\pi t), j \geq 1\}$  est la base de Fourier sur  $[0, 1]$ . On vérifie sans peine que comme cette base est orthogonale, la matrice

$$\Phi^{(n_T)} = \left( \sqrt{2} \sin((j-1/2)\pi t_k) \right)_{1 \leq j \leq J, 1 \leq k \leq n_T}$$

est telle que  $n_T^{-1} \Phi^{(n_T)} \Phi^{(n_T)'} \xrightarrow{n_T \rightarrow \infty} \mathbf{I}_J$ . On représente les quatre premiers éléments de cette base en Figure 1.3.

### 1.5.2.2 Bases de composantes principales

Pour un échantillon de fonctions  $\{X_i, 1 \leq i \leq n\}$ , on définit

$$\forall i \in \{1, \dots, n\} \quad \tilde{X}_i = X_i - \bar{X}$$

$$\text{où } \bar{X} = n^{-1} \sum_{i=1}^n X_i.$$

Si l'on avait comme contrainte de représenter chacun de ces  $\tilde{X}_i$  par  $c_i g$  où  $c_i \in \mathbb{R}$ , avec pour objectif de minimiser  $\sum_{i=1}^n \int (\tilde{X}_i(t) - c_i g(t))^2$  sous la contrainte que  $\int g^2(t) = 1$ , on obtient que  $c_i = \langle X_i, g \rangle$  et que  $g$  est la première fonction propre de la fonction de covariance empirique  $v(s, t) = n^{-1} \sum_{i=1}^n \tilde{X}_i(s) \tilde{X}_i(t)$ .

Plus généralement, on définit la base de composantes principales par l'ensemble des solutions de l'équation  $\int v(s, t) \xi(t) dt = \langle v(s, \cdot), \xi \rangle = \rho \xi(s)$ . Pour la base particulière considérée précédemment,

$$\int v(s, t) \xi(t) dt = n^{-1} \sum_{i=1}^n \tilde{X}_i(s) \int \tilde{X}_i(t) \xi(t) dt$$

$$\simeq n^{-1} \sum_{i=1}^n \tilde{X}_i(s) n_T^{-1} \sum_{k=1}^{n_T} \tilde{X}_i(t_k) \xi(t_k)$$

et l'estimation du vecteur  $\left( \int v(\cdot, t) \xi(t) dt \right)^{(n_T)}$  n'est autre que

$$n_T^{-1} \mathbf{X}^{(n_T)} \mathbf{X}^{(n_T)'} \xi^{(n_T)}$$

$$\text{où } \mathbf{X}^{(n_T)} = (X_{ik})_{1 \leq i \leq n, 1 \leq k \leq n_T}$$

avec  $X_{ik} = X_i(t_k)$ . On peut donc chercher les vecteurs propres de la matrice  $\mathbf{X}^{(n_T)} \mathbf{X}^{(n_T)'}$  puisque l'équation approchée en termes vectoriels devient

$$\mathbf{X}^{(n_T)} \mathbf{X}^{(n_T)'} \xi^{(n_T)} = n_T \rho \xi^{(n_T)}.$$

## 1.5.3 Les modèles linéaires fonctionnels

### 1.5.3.1 Le modèle linéaire fonctionnel pour $Y$ scalaire

Pour un modèle linéaire avec  $X \in \mathbb{R}^p$ , on utilise  $Y = a + X'\beta + \varepsilon$  avec  $a \in \mathbb{R}$  et  $\beta \in \mathbb{R}^p$ , i.e.  $Y = a + \langle X, \beta \rangle + \varepsilon$  avec  $\langle \cdot, \cdot \rangle$  le produit scalaire usuel sur  $\mathbb{R}^p$ . Il est donc naturel de définir le modèle linéaire lorsque  $X$  est fonctionnel (on prendra  $X \in L^2([0, 1])$  de la manière suivante :

$$\begin{aligned} Y &= a + \langle X, b \rangle + \varepsilon \\ &= a + \int_0^1 X(t) b(t) dt + \varepsilon \end{aligned}$$

avec  $\beta \in L^2([0, 1])$ . Encore une fois, l'estimation des paramètres  $a$  et  $b$  est facilitée lorsque l'on utilise une grille d'observation régulière. En effet, une estimation<sup>4</sup> de  $b^{(n_T)}$  est  $\hat{b}^{(n_T)} = n_T \widehat{\tilde{\mathbf{X}}^{(n_T)}}^+ \tilde{\mathbf{y}}$  avec  $\tilde{y}_i = Y_i - \bar{Y}$  et où  $\widehat{\tilde{\mathbf{X}}^{(n_T)}}^+$  est une version approchée de la matrice inverse de Moore-Penrose<sup>5</sup> de  $\tilde{\mathbf{X}}^{(n_T)}$ . Pour en obtenir une version simple, on peut utiliser la décomposition en valeurs singulières  $\tilde{\mathbf{X}}^{(n_T)} = \mathbf{U}\mathbf{D}\mathbf{V}'$  avec  $\mathbf{U} \in \mathbb{R}^{n \times n \wedge n_T}$  et  $\mathbf{V} \in \mathbb{R}^{n_T \times n \wedge n_T}$ . Si pour  $r \in \{1, \dots, n \wedge n_T\}$ , on note  $\mathbf{U}_r$  et  $\mathbf{V}_r$  les matrices composées respectivement des  $r$  premières colonnes de  $\mathbf{U}$  et  $\mathbf{V}$  et  $\mathbf{D}_r$  la matrice diagonale contenant les  $r$  premières valeurs singulières de  $\tilde{\mathbf{X}}^{(n_T)}$ , alors on peut prendre  $\hat{b}_{(r)}^{(n_T)} = \mathbf{V}_r \mathbf{D}_r^{-1} \mathbf{U}_r' \tilde{\mathbf{y}}^{(n_T)}$ . Notons que

$$\begin{aligned} \hat{b}_{(n \wedge n_T)}^{(n_T)} &= \mathbf{V}\mathbf{D}^{-1}\mathbf{U}'\tilde{\mathbf{y}} \\ &= \begin{cases} \tilde{\mathbf{X}}^{(n_T)'} \left( \tilde{\mathbf{X}}^{(n_T)} \tilde{\mathbf{X}}^{(n_T)'} \right)^{-1} \tilde{\mathbf{y}} & \text{si } n \leq n_T \\ \left( \tilde{\mathbf{X}}^{(n_T)'} \tilde{\mathbf{X}}^{(n_T)} \right)^{-1} \tilde{\mathbf{X}}^{(n_T)'} \tilde{\mathbf{y}} & \text{si } n \geq n_T \end{cases} \\ &= \widehat{\tilde{\mathbf{X}}^{(n_T)}}^+ \tilde{\mathbf{y}}. \end{aligned}$$

Pour l'estimation de  $a$ , on utilise  $\hat{a} = \bar{Y} - \langle \bar{X}, \hat{b}_{(n \wedge n_T)}^{(n_T)} \rangle$ .

### 1.5.3.2 Le modèle linéaire fonctionnel pour $Y$ fonctionnel

Lorsque  $X$  et  $Y$  sont des variables aléatoires fonctionnelle ( $X \in L^2([c_X, d_X])$  et  $Y \in L^2([c_Y, d_Y])$ ), il est naturel d'utiliser un paramètre qui est une fonction de deux variables :  $\xi : [c_X, d_X] \times [c_Y, d_Y] \rightarrow \mathbb{R}$  et d'utiliser le modèle

$$Y(t) = \mu(t) + \int X(s) \xi(s, t) ds + \varepsilon(t).$$

On prendra sans perte de généralité  $[c_X, d_X] = [c_Y, d_Y] = [0, 1]$ . De la même manière que lorsque  $Y$  est scalaire, on peut estimer  $\xi$  en utilisant

$$\hat{\xi} = \widehat{\tilde{\mathbf{X}}^{(n_T)}}^+ \tilde{\mathbf{Y}}^{(n_T)}$$

avec  $\tilde{Y}_{ik}^{(n_T)} = Y_i(t_k) - \bar{Y}(t_k)$ . On estime ensuite  $\hat{\mu}^{(n_T)} = \bar{Y}^{(n_T)} - \bar{X}^{(n_T)} \hat{\xi}$ .

<sup>4</sup>voir Ramsay et Silverman (2005)

<sup>5</sup>L'inverse de Moore-Penrose d'une matrice  $\mathbf{X}$  est la matrice  $\mathbf{X}^+$  telle que  $\mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X}$ ,  $\mathbf{X}^+\mathbf{X}\mathbf{X}^+ = \mathbf{X}^+$ , et  $\mathbf{X}\mathbf{X}^+$  et  $\mathbf{X}^+\mathbf{X}$  sont symétriques. (Moore (1920))

### 1.5.3.3 Le modèle “concurrent” pour $Y$ fonctionnel

Lorsque  $X$  et  $Y$  vivent dans le même espace, il est possible de considérer le modèle

$$Y(t) = a(t) + X(t)b(t) + \varepsilon(t)$$

Dans ce cas, on peut considérer l'estimation de  $a$  et de  $b$  pour chaque  $t$  indépendamment

$$\hat{b}(t) = \frac{\sum_{i=1}^n (X_i(t) - \bar{X}_n(t)) (Y_i(t) - \bar{Y}_n(t))}{\sum_{i=1}^n (X_i(t) - \bar{X}_n(t))^2};$$

$$\hat{a}(t) = \bar{Y}_n(t) - \bar{X}_n(t) \hat{b}(t).$$

## 1.6 Le modèle single-index

Si la modélisation paramétrique offre des avantages en terme d'estimation (dans le sens où les vitesses de convergence des estimateurs sont rapides) et en terme d'interprétation (comparaison des signes et des magnitudes des paramètres) elle a le fâcheux inconvénient d'être dépendante d'une hypothèse forte de validité du modèle : quels sont le sens et l'utilité du calcul d'une espérance linéaire lorsque la réalité est quadratique ? On a vu que l'on peut tester l'hypothèse paramétrique pour se prémunir d'un risque d'erreur de spécification du modèle. Mais que faire alors si l'on rejette l'hypothèse ? La modélisation qui ne fait aucune hypothèse (si ce n'est le fait que les observations proviennent du même phénomène observé plusieurs fois de manière indépendante) est la statistique non paramétrique. En régression, on écrit simplement  $Y = r(X) + \varepsilon$  avec  $r(x) = \mathbb{E}[Y | X]$ , ou  $\mathbb{E}[\varepsilon | X]$ , ce qui est équivalent. Le calcul de cette espérance conditionnelle souffre de ce qu'on appelle le « fléau de la dimension » : pour une classe de fonctions de régularité  $s$  et une variable  $X$  de dimension  $p$ , la vitesse minimax de convergence d'un estimateur  $r$  de  $\hat{r}$  est  $n^{s/(p+2s)}$ . Si la régularité  $s$  augmente, on s'approche d'une vitesse paramétrique  $n^{1/2}$ , mais si c'est la dimension  $p$  qui augmente, alors cette vitesse est de plus en plus lente.

Il existe plusieurs modélisations que l'on appelle semi-paramétriques et qui permettent de s'affranchir du fléau de la dimension tout en gardant une certaine flexibilité. Le modèle single-index suppose que

$$Y = r(v(x, \theta)) + \varepsilon$$

avec l'index  $v(\cdot, \cdot)$  fixé de  $\mathcal{X} \times \Theta$  dans  $\mathbb{R}$ .

On ne considérera ici que le cas  $v(x, \theta) = x'\theta$ , qui conserve une hypothèse de linéarité puisqu'on suppose que la variable  $Y$  ne dépend que d'une combinaison linéaire des



composantes de  $X$ . Mais il généralise encore le modèle linéaire dit généralisé (GLM) puisqu'on suppose dans les deux cas que  $Y = r(X'\theta) + \varepsilon$  avec  $r$  fixé dans le cas GLM, mais inconnu dans le cas single-index, ce qui le rend plus flexible. Citons parmi les autres modèles semi-paramétrique le modèle additif généralisé où l'on suppose que l'espérance de  $Y$  sachant  $X$  est une somme de fonctions ne dépendant que d'une seule composante de  $X$  :  $Y = \sum_{j=1}^p m_j(X_j) + \varepsilon$ .

Quoi qu'il en soit, puisque un modèle single-index est une restriction, cette hypothèse peut se tester. En effet, pour qu'il soit valide, on a besoin qu'il existe  $\theta \in \mathbb{R}^p$ , unique à un facteur d'échelle près tel que

$$\mathbb{E}[Y | X] = \mathbb{E}[Y | X'\theta].$$

Autrement dit, l'hypothèse testée est donc qu'il existe un vecteur  $\theta$  et une fonction  $g(\cdot)$  tels que

$$\mathbb{E}[Y - g(X'\theta) | X] = 0 \quad \text{p.s.} \quad (1.5)$$

En reprenant la formulation de Xia et al. (2004), si l'on note

$$g_\theta(v) = \mathbb{E}[Y | X'\theta = v] \quad \text{et} \quad \theta_0 = \arg \min_{\theta: \|\theta\|=1} \mathbb{E}[(Y - g_\theta(X'\theta))^2]$$

alors l'hypothèse testée est

$$\mathbb{E}[Y - g_{\theta_0}(X'\theta_0) | X] = 0 \quad \text{p.s.}$$

Xia et al. (2004) proposent une statistique de type Cramér-von Mises en remarquant que l'hypothèse (1.5) est équivalente

$$\mathbb{E}[(Y - g_{\theta_0}(X'\theta_0)) \mathbf{1}(X < x)] \equiv 0$$

où  $X < x$  est l'ordre de produit, c'est à dire que chaque composante de  $X$  est strictement inférieur à la composante de  $x$  correspondante.

$$S_n(x) = n^{-1/2} \sum_{j=1}^n (Y_j - \hat{Y}_j) \mathbf{1}(X_j < x)$$

L'estimation  $\hat{Y}_j = \hat{g}_{\hat{\theta}}(X_j'\hat{\theta})$  est dans cet article donnée par un estimateur linéaire local

$$\hat{g}_\theta(v) = \frac{\sum_{i=1}^n W_{n,h}(X_i'\theta - v) Y_i}{\sum_{i=1}^n W_{n,h}(X_i'\theta - v)},$$

où

$$\begin{aligned} W_{n,h}(X_i'\theta - v) &= s_{n,\theta,2}(v) n^{-1} K_h(X_i'\theta - v) - s_{n,\theta,1}(v) K_h(X_i'\theta - v) \frac{X_i'\theta - v}{h}, \\ s_{n,\theta,k}(v) &= n^{-1} \sum_{j=1}^n K_h(X_j'\theta - v) \left( \frac{X_j'\theta - v}{h} \right)^k, \quad k = 1, 2 \end{aligned}$$

avec  $K(\cdot)$  un noyau,  $K_h(\cdot) = h^{-1}K(\cdot/h)$  et  $h > 0$  une fenêtre. Parmi les statistiques basées sur le processus, on retrouve une l'idée développée par Stute et al. (1998b).

La statistique de Stute et Zhu (2005) est de la forme

$$\hat{T}_n = n^{-1/2} \sum_{j=1}^n \hat{U}_j W(X_j)$$

où  $W(\cdot)$  sera défini plus bas. Les  $\hat{U}_j = Y_j - \hat{\psi}_n^{(j)}(\hat{V}_j)$  sont obtenus par un lissage non paramétrique

$$\hat{\psi}_n^{(j)}(v) = n^{-1} \sum_{\substack{i=1 \\ i \neq j}}^n K_h(v - \hat{V}_i)$$

où  $\hat{V}_i = F_n(X_i' \hat{\theta})$  avec  $F_n$  la fonction de répartition empirique de  $\{X_i' \hat{\theta}, 1 \leq i \leq n\}$  et  $\hat{\theta}$  un estimateur de  $\theta$  (à déf.).

Si le but est de détecter une alternative du type

$$Y_i = g(X_i' \theta) + n^{-1/2} s(X_i) + \varepsilon_i$$

alors ils montrent que le choix optimal pour  $W$  est

$$W_0(x) \propto s(x).$$

Ils considèrent alors les cas où  $s(\cdot)$  est effectivement connue et utilisent la théorie maximin dans le cas où  $s(x) = \sum_{j=1}^d \gamma_j s_j(x)$  avec les  $s_j$  connus, mais puisque les tests considérés dans la thèse sont non paramétriques, on se concentrera uniquement sur les solutions proposées dans le cas de tests omnibus. Dans ce cadre, Stute et Zhu (2005) proposent d'utiliser  $W(\gamma, x) = \exp\{iX'\gamma\}$ . La quantité  $\hat{T}_n$  considérée dépend donc de  $\gamma$ , et les auteurs proposent d'utiliser la statistique  $\tilde{T}_n = \sup_{\gamma} \hat{T}_n(\gamma)$ . On remarque que cette statistique est le pendant de type Kolmogorov-Smirnov de la statistique de Bierens (1984) vue comme une statistique de Cramér-von Mises. Bien entendu, les résidus considérés ne sont pas du même type, puisque pour le modèle SIM cela nécessite d'utiliser des techniques non paramétriques, ici un lissage à noyau, alors que Bierens (1984) se plaçait dans le cadre d'un modèle paramétrique.

Pour les statistiques utilisant le lissage à noyau, on a l'article de Fan et Li (1996) qui considère le cas single-index comme un cas particulier du test de significativité. En effet, si l'on prend  $U^{Sig} = Y - \mathbb{E}[Y | W]$  et  $U^{SIM} = Y - g_{\theta_0}(X'\theta_0)$  respectivement dans le cas du test de significativité de  $W$  dans la régression de  $Y$  sur  $X = (W, Z)$  et du test du single-index, le test s'écrit

$$\mathbb{E}[U | X] = 0 \text{ p.s.}$$

D'autre part, l'article de Chen et Van Keilegom (2009) considère le problème très général des test de modèles paramétriques et semi-paramétriques lorsque  $Y$  est vectoriel. Pour le single-index, cela signifie que chacune des composantes de  $Y$  est expliquée par le même indice  $X'\theta$ .

On trouve également un article d'Escanciano et Song(2010), mais qui n'est pas omnibus à proprement parler, puisque l'hypothèse alternative considérée est

$$H_1 : \mathbb{E}[(Y_i - g(X_i'\theta)) b(X)] \neq 0 \quad \forall \theta$$

où

$$b(\cdot) = -\frac{1}{f_X(\cdot)} \frac{\partial f_X}{\partial x_1}(\cdot),$$

$f_X(\cdot)$  étant la densité de  $X$ , c'est à dire que le test se concentre sur les dérivées moyennées. En effet, l'hypothèse alternative pour un test omnibus serait qu'il existe une fonction  $b(\cdot)$  telle que l'on ait  $H_1$ .

## Présentation des articles

La suite de ce manuscrit est rédigée sous forme d'articles indépendants les uns des autres ou presque – on trouve des références croisées – et destinés à être publiés dans des revues. Chacun constitue en l'étude d'une statistique de test qui aura toujours la même forme mais qui sera appliquée à des cas de nature variée. Par conséquent les techniques utilisées dans les preuves seront très diverses.

On commencera par présenter un test de significativité en régression dans le Chapitre 2, puis dans le cas des tests d'adéquation pour un modèle paramétrique en régression quantile au Chapitre 3, un test d'effet lorsque la variable explicative, et éventuellement la variable réponse, sont de nature fonctionnelle. Ce test sera abordé dans le cadre du Chapitre 4 et les techniques utilisées en simulation feront un usage répété des concepts développés en Section 1.5. Enfin, on abordera dans le Chapitre 5 le test d'adéquation à un modèle de régression à direction révélatrice unique, que ce soit pour l'espérance conditionnelle ou pour la loi conditionnelle.



## **2 Test de significativité en régression non paramétrique**

Ce chapitre a été déposé sur la plate-forme d'enregistrement électronique arXiv. Il porte l'identification arXiv : 1403.7063 [math.ST].



## **Abstract**

We consider testing the significance of a subset of covariates in a nonparametric regression. These covariates can be continuous and/or discrete. We propose a new kernel-based test that smoothes only over the covariates appearing under the null hypothesis, so that the curse of dimensionality is mitigated. The test statistic is asymptotically pivotal and the rate of which the test detects local alternatives depends only on the dimension of the covariates under the null hypothesis. We show the validity of wild bootstrap for the test. In small samples, our test is competitive compared to existing procedures.



## 2.1 Introduction

Testing the significance of covariates is common in applied regression analysis. Sound parametric inference hinges on the correct functional specification of the regression function, but the likelihood of misspecification in a parametric framework cannot be ignored, especially as applied researchers tend to choose functional forms on the basis of parsimony and tractability. Significance testing in a nonparametric framework has therefore obvious appeal as it requires much less restrictive assumptions. Fan (1996), Fan and Li (1996), Racine (1997), Chen and Fan (1999), Lavergne and Vuong (2000), Ait-Sahalia et al. (2001), and Delgado and González Manteiga (2001) proposed tests of significance for continuous variables in nonparametric regression models. Delgado (1993), Dette and Neumeyer (2001), Lavergne (2001), Neumeyer and Dette (2003), Racine et al. (2006) focused on significance of discrete variables. Volgushev et al. (2013) considered significance testing in nonparametric quantile regression. For each test, one needs first to estimate the model without the covariates under test, that is under the null hypothesis. The result is then used to check the significance of extra covariates. Two competing approaches are then possible. In the “smoothing approach,” one regresses the residuals onto the whole set of covariates nonparametrically, while in the “empirical process approach” one uses the empirical process of residuals marked by a function of all covariates.

In this work, we adopt an hybrid approach to develop a new significance test of a subset of covariates in a nonparametric regression. Our new test has three specific features. First, it does not require smoothing with respect to the covariates under test as in the “empirical process approach.” This allows to mitigate the curse of dimensionality that appears with nonparametric smoothing, hence improving the power properties of the test. Our simulation results show that indeed our test is more powerful than competitors under a wide spectrum of alternatives. Second, the test statistic is asymptotically pivotal as in the “smoothing approach,” while wild bootstrap can be used to obtain small samples critical values of the test. This yields a test whose level is well controlled by bootstrapping, as shown in simulations. Third, our test equally applies whether the covariates under test are continuous or discrete, showing that there is no need of a specific tailored procedure for each situation.

The paper is organized as follows. In Section 2, we present our testing procedure. In Section 3, we study its asymptotic properties under a sequence of local alternatives and we establish the validity of wild bootstrap. In Section 4, we compare the small sample behavior of our test to some existing procedures. Section 5 gathers our proofs.

## 2.2 Testing Framework and Procedure

### 2.2.1 Testing Principle

We want to assess the significance of  $X \in \mathbb{R}^q$  in the nonparametric regression of  $Y \in \mathbb{R}$  on  $W \in \mathbb{R}^p$  and  $X$ . Formally, this corresponds to the null hypothesis

$$H_0 : \mathbb{E}[Y | W, X] = \mathbb{E}[Y | W] \quad \text{a.s.}$$

which is equivalent to

$$H_0 : \mathbb{E}[u | W, X] = 0 \quad \text{a.s.} \quad (2.1)$$

where  $u = Y - \mathbb{E}[Y | W]$ . The corresponding alternative hypothesis is

$$H_1 : \mathbb{P}\{\mathbb{E}[u | W, X] = 0\} < 1.$$

The following result is the cornerstone of our approach. It characterizes the null hypothesis  $H_0$  using a suitable unconditional moment equation.

**Lemma 2.1.** *Let  $(W_1, X_1, u_1)$  and  $(W_2, X_2, u_2)$  be two independent draws of variables  $(W, X, u)$ ,  $\nu(W)$  a strictly positive function on the support of  $W$  such that  $\mathbb{E}[u^2 \nu^2(W)] < \infty$ , and  $K(\cdot)$  and  $\psi(\cdot)$  even functions with (almost everywhere) positive Fourier integrable transforms. Define*

$$I(h) = \mathbb{E}\left[u_1 u_2 \nu(W_1) \nu(W_2) h^{-p} K((W_1 - W_2)/h) \psi(X_1 - X_2)\right].$$

Then for any  $h > 0$ ,

$$\mathbb{E}[u | W, X] = 0 \quad \text{a.s.} \Leftrightarrow I(h) = 0.$$

*Proof.* Let  $\langle \cdot, \cdot \rangle$  denote the standard inner product. Using Fourier Inversion Theorem, change of variables, and elementary properties of conditional expectation,

$$\begin{aligned} I(h) &= \mathbb{E}\left[u_1 u_2 \nu(W_1) \nu(W_2) \int_{\mathbb{R}^p} e^{2\pi i \langle t, W_1 - W_2 \rangle} \mathcal{F}[K](th) dt \right. \\ &\quad \left. \times \int_{\mathbb{R}^q} e^{2\pi i \langle s, X_1 - X_2 \rangle} \mathcal{F}[\psi](s) ds \right] \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left| \mathbb{E}\left[\mathbb{E}[u | W, X] \nu(W) e^{2\pi i \langle t, W \rangle + \langle s, X \rangle} \right] \right|^2 \mathcal{F}[K](th) \mathcal{F}[\psi](s) dt ds. \end{aligned}$$

Since the Fourier transforms  $\mathcal{F}[K]$  and  $\mathcal{F}[\psi]$  are strictly positive,  $I(h) = 0$  iff

$$\mathbb{E}\left[\mathbb{E}[u | W, X] \nu(W) e^{2\pi i \langle t, W \rangle + \langle s, X \rangle} \right] = 0 \quad \forall t, s.$$

But this is equivalent to  $\mathbb{E}[u | W, X] \nu(W) = 0$  a.s., which by our assumption on  $\nu(\cdot)$  is equivalent to  $H_0$ .  $\square$

## 2.2.2 The Test

Lemma 2.1 holds whether the covariates  $W$  and  $X$  are continuous or discrete. For now, we assume  $W$  is continuously distributed, and we later comment on how to modify our procedure in the case where some of its components are discrete. We however do not restrict  $X$  to be continuous. Since it is sufficient to test whether  $I(h) = 0$  for any arbitrary  $h$ , we can choose  $h$  to obtain desirable properties. So we consider a sequence of  $h$  decreasing to zero when the sample size increases, which is one of the ingredient that allows to obtain a tractable asymptotic distribution for the test statistic.

Assume we have at hand a random sample  $(Y_i, W_i, X_i)$ ,  $1 \leq i \leq n$ , from  $(Y, W, X)$ . In what follows,  $f(\cdot)$  denotes the density of  $W$ ,  $r(\cdot) = \mathbb{E}[Y | W = \cdot]$ ,  $u = Y - r(W)$ , and  $\hat{f}_i$ ,  $\hat{r}_i$ ,  $\hat{u}_i$  respectively denote  $f(W_i)$ ,  $r(W_i)$ , and  $Y_i - r(W_i)$ . Since nonparametric estimation should be entertained to approximate  $u_i$ , we consider usual kernel estimators based on kernel  $L(\cdot)$  and bandwidth  $g$ . With  $L_{nik} = \frac{1}{g^p} L\left(\frac{W_i - W_k}{g}\right)$ , let

$$\begin{aligned}\hat{f}_i &= (n-1)^{-1} \sum_{k \neq i, k=1}^n L_{nik}, \\ \hat{r}_i &= \frac{1}{\hat{f}_i} \frac{1}{(n-1)} \sum_{k \neq i, k=1}^n Y_k L_{nik} \\ \text{so that } \hat{u}_i &= Y_i - \hat{r}_i = \frac{1}{\hat{f}_i} \frac{1}{(n-1)} \sum_{k \neq i, k=1}^n (Y_i - Y_k) L_{nik}.\end{aligned}$$

Denote by  $n^{(m)}$  the number of arrangements of  $m$  distinct elements among  $n$ , and by  $[1/n^{(m)}] \sum_a$ , the average over these arrangements. In order to avoid random denominators, we choose  $\nu(W) = f(W)$ , which fulfills the assumption of Lemma 2.1. Then we can estimate  $I(h)$  by the second-order U-statistic

$$\begin{aligned}\hat{I}_n &= \frac{1}{n^{(2)}} \sum_a \hat{u}_i \hat{f}_i \hat{u}_j \hat{f}_j K_{nij} \psi_{ij} \\ &= \frac{1}{n^{(2)} (n-1)^2} \sum_a \sum_{k \neq i} \sum_{l \neq j} (Y_i - Y_k) (Y_j - Y_l) L_{nik} L_{njl} K_{nij} \psi_{ij},\end{aligned}$$

with  $K_{nij} = \frac{1}{h^p} K\left(\frac{W_i - W_j}{h}\right)$  and  $\psi_{ij} = \psi(X_i - X_j)$ . We also consider the alternative statistic

$$\tilde{I}_n = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k) (Y_j - Y_l) L_{nik} L_{njl} K_{nij} \psi_{ij}.$$

It is clear that  $\tilde{I}_n$  is obtained from  $\hat{I}_n$  by removing asymptotically negligible ‘‘diagonal’’ terms. Under the null hypothesis, both statistics will have the same asymptotic normal distribution, but removing diagonal terms reduces the bias of the statistic

under  $H_0$ . Our statistics  $\tilde{I}_n$  and  $\hat{I}_n$  are respectively similar to the ones of Fan and Li (1996) and Lavergne and Vuong (2000), with the fundamental difference that there is no smoothing relative to the covariates  $X$ . Indeed these authors used a multidimensional smoothing kernel over  $(W, X)$ , that is  $h^{-(p+q)}\tilde{K}((W_i - W_j)/h, (X_i - X_j)/h)$ , while we use  $K_{nij}\psi_{ij}$ . We will show that  $nh^{p/2}I_n \xrightarrow{d} \mathcal{N}(0, \omega^2)$  under  $H_0$  whereas  $nh^{p/2}I_n \xrightarrow{p} \infty$  under  $H_1$ , for  $I_n$  being either  $\tilde{I}_n$  or  $\hat{I}_n$ . By contrast, the statistics of Fan and Li (1996) and Lavergne and Vuong (2000) exhibit a  $nh^{(p+q)/2}$  rate of convergence. The alternative test of Delgado and González Manteiga (2001) uses the kernel residuals  $\hat{u}_i$  and the empirical process approach of Stute (1997). This avoids extra smoothing, but at the cost of a test statistic with a non pivotal asymptotic law under  $H_0$ . Hence, our proposal is an hybrid approach that combines the advantages of existing procedures, namely smoothing only for the variables  $W$  appearing under the null hypothesis but with an asymptotic normal distribution for the statistic. Given a consistent estimator  $\omega_n^2$  of  $\omega^2$ , as provided in the next section, we obtain an asymptotic  $\alpha$ -level test of  $H_0$  as

$$\text{Reject } H_0 \text{ if } nh^{p/2}I_n/\omega_n > z_{1-\alpha},$$

where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ -th quantile of the standard normal distribution. In small samples, we will show the validity of a wild bootstrap scheme to obtain critical values.

The test applies whether  $X$  is continuous or has some discrete components. The procedure is also easily adapted to some discrete components of  $W$ . In that case, one would replace kernel smoothing by cells' indicators for the discrete components, so that for  $W$  composed of continuous  $W_c$  of dimension  $p_c$  and discrete  $W_d$ , one would use  $h^{-p_c}K\left(\frac{W_{ic}-W_{jc}}{h}\right)\mathbb{I}(W_{id}=W_{jd})$  instead of  $K_{nij}$ . It would also be possible to smooth on the discrete components, as proposed by Racine and Li (2004). To obtain scale invariance, we recommend that observations on covariates should be scaled, say by their sample standard deviation as is customary in nonparametric estimation. It is equally important to scale the  $X_i$  before they are used as arguments of  $\psi(\cdot)$  to preserve such invariance.

The outcome of the test may depend on the choice of the kernels  $K(\cdot)$  and  $L(\cdot)$ , while this influence is expected to be limited as it is in nonparametric estimation. The choice of the function  $\psi(\cdot)$  might be more important, but our simulations reveal that it is not. From our theoretical study, this function, as well as  $K(\cdot)$  should possess an almost everywhere positive and integrable Fourier transform. This is true for (products of) the triangular, normal, Laplace, and logistic densities, see Johnson et al. (1995), and for a Student density, see Hurst (1995). Alternatively, one can choose  $\psi(x)$  as a univariate density applied to some transformation of  $x$ , such as its norm. This yields  $\psi(x) = g(\|x\|)$  where  $g(\cdot)$  is any of the above univariate densities. This is the form we will consider in our simulations to study the influence of  $\psi(\cdot)$ .

## 2.3 Theoretical Properties

We here give the asymptotic properties of our test statistics under  $H_0$  and some local alternatives. To do so in a compact way, we consider the sequence of hypotheses

$$H_{1n} : \mathbb{E}[Y | W, X] = r(W) + \delta_n d(W, X), \quad n \geq 1,$$

where  $d(\cdot)$  is a fixed integrable function. Since  $r(W) = \mathbb{E}[Y | W]$ , our setup imposes  $\mathbb{E}[d(W, X) | W] = 0$ . The null hypothesis corresponds to the case  $\delta_n \equiv 0$ , while considering a sequence  $\delta_n \rightarrow 0$  yields local Pitman-like alternatives.

### 2.3.1 Assumptions

We begin by some useful definitions.

**Definition 2.2. (i)**  $\mathcal{U}^p$  is the class of integrable uniformly continuous functions from  $\mathbb{R}^p$  to  $\mathbb{R}$ ;

**(ii)**  $\mathcal{D}_s^p$  is the class of  $m$ -times differentiable functions from  $\mathbb{R}^p$  to  $\mathbb{R}$ , with derivatives of order  $\lfloor s \rfloor$  that are uniformly Lipschitz continuous of order  $s - \lfloor s \rfloor$ , where  $\lfloor s \rfloor$  denotes the integer such that  $\lfloor s \rfloor \leq s < \lfloor s \rfloor + 1$ .

Note that a function belonging to  $\mathcal{U}^p$  is necessarily bounded.

**Definition 2.3.**  $\mathcal{K}_m^p$ ,  $m \geq 2$ , is the class of even integrable functions  $K : \mathbb{R}^p \rightarrow \mathbb{R}$  with compact support satisfying  $\int K(t) dt = 1$  and, if  $t = (t_1, \dots, t_p)$ ,

$$\int_{\mathbb{R}^p} t_1^{\alpha_1} \dots t_p^{\alpha_p} K(t) dt = 0 \quad \text{for } 0 < \sum_{i=1}^p \alpha_i \leq m - 1, \alpha_i \in \mathbb{N} \quad \forall i$$

This definition of higher-order kernels is standard in nonparametric estimation. The compact support assumption is made for simplicity and could be relaxed at the expense of technical conditions on the rate of decrease of the kernels at infinity, see e.g. Definition 1 in Fan and Li (1996). In particular, the gaussian kernel could be allowed for. We are now ready to list our assumptions.

**Assumption 2.4.** (i) For any  $x \in \mathbb{R}^q$  in the support of  $X$ , the vector  $W$  admits a conditional density given  $X = x$  with respect to the Lebesgue measure in  $\mathbb{R}^p$ , denoted by  $\pi(\cdot | x)$ . Moreover,  $\mathbb{E}[Y^8] < \infty$ . (ii) The observations  $(W_i, X_i, Y_i)$ ,  $i = 1, \dots, n$  are independent and identically distributed as  $(W, X, Y)$ .

The existence of the conditional density given  $X = x$  for all  $x \in \mathbb{R}^q$  in the support of  $X$  implies that  $W$  admits a density with respect to the Lebesgue measure on  $\mathbb{R}^p$ . As noted above, our results easily generalizes to some discrete components of  $W$ , but for the sake of simplicity we do not formally consider this in our theoretical analysis.

- Assumption 2.5.** (i)  $f(\cdot)$  and  $r(\cdot)f(\cdot)$  belong to  $\mathcal{U}^p \cap \mathcal{D}_s^p$ ,  $s \geq 2$ ;
- (ii)  $\mathbb{E}[u^2 | W = \cdot]f(\cdot)$ ,  $\mathbb{E}[u^4 | W = \cdot]f^4(\cdot)$  belong to  $\mathcal{U}^p$ ;
- (iii) the function  $\psi(\cdot)$  is bounded and has a almost everywhere positive and integrable Fourier transform;
- (iv)  $K(\cdot) \in \mathcal{K}_2^p$  and has an almost everywhere positive and integrable Fourier transform, while  $L(\cdot) \in \mathcal{K}_{[s]}^p$  and is of bounded variation;
- (v) let  $\sigma^2(w, x) = \mathbb{E}[u^2 | W = w, X = x]$ , then  $\sigma^2(\cdot, x)f^2(\cdot)\pi(\cdot | x)$  belongs to  $\mathcal{U}^p$  for any  $x$  in the support of  $X$ , has integrable Fourier transform, and  $\mathbb{E}[\sigma^4(W, X)f^4(W)\pi(W | X)] < \infty$ ;
- (vi)  $\mathbb{E}[d^2(W, X) | W = \cdot]f^2(\cdot)$  belongs to  $\mathcal{U}^p$ ,  $d(\cdot, x)f(\cdot)\pi(\cdot | x)$  is integrable and squared integrable for any  $x$  in the support of  $X$ , and  $\mathbb{E}[d^2(W, X)f^2(W)\pi(W | X)] < \infty$ .

Standard regularity conditions are assumed for various functions. A higher-order kernel  $L(\cdot)$  is used in conjunction with the differentiability conditions in (i) to ensure that the bias in nonparametric estimation is small enough.

### 2.3.2 Asymptotic Analysis

The following result characterizes the behavior of our statistics under the null hypothesis and a sequence of local alternatives.

**Theorem 2.6.** Let  $I_n$  be any of the statistics  $\hat{I}_n$  or  $\tilde{I}_n$ . Under Assumptions 2.4 and 2.5, and if as  $n \rightarrow \infty$  (i)  $g, h \rightarrow 0$ , (ii)  $n^{7/8}g^p / \ln n, nh^p \rightarrow \infty$ , (iii)  $nh^{p/2}g^{2s} \rightarrow 0$ , and (iv)  $h/g \rightarrow 0$  if  $I_n = \tilde{I}_n$  or  $h/g^2 \rightarrow 0$  if  $I_n = \hat{I}_n$ , then

(i) If  $\delta_n^2 nh^{p/2} \rightarrow C$  with  $0 \leq C < \infty$ ,  $nh^{p/2}I_n \xrightarrow{d} \mathcal{N}(C\mu, \omega^2)$  where

$$\mu = \mathbb{E} \left[ \int d(w, X_1) d(w, X_2) f^2(w) \pi(w | X_1) \pi(w | X_2) \psi(X_1 - X_2) dw \right] > 0$$

$$\text{and } \omega^2 = 2 \int K^2(s) ds$$

$$\mathbb{E} \left[ \int \sigma^2(w, X_1) \sigma^2(w, X_2) f^4(w) \pi(w | X_1) \pi(w | X_2) \psi^2(X_1 - X_2) dw \right].$$

(ii) If  $\delta_n^2 nh^{p/2} \rightarrow \infty$ ,  $nh^{p/2}I_n \xrightarrow{p} \infty$ .

The rate of convergence of the test statistic depends only on the dimension of  $W$ , the covariates present under the null hypothesis, but not on the dimension of  $X$ , the covariates under test. Similarly, the rate of local alternatives that are detected

by the test depends only on the dimension of  $W$ . As shown in the simulations, this yields some gain in power compared to competing “smoothing” tests. Conditions (i) to (iv) together require that  $s > p/2$  for  $I_n = \tilde{I}_n$  and  $s > p/4$  for  $I_n = \hat{I}_n$ , so removing diagonal terms in  $\hat{I}_n$  allows to weaken the restrictions on the bandwidths. Condition (ii) could be slightly weakened to  $ng^p \rightarrow \infty$  at the price of handling high order  $U$ -statistics in the proofs, but allows for a shorter argument based on empirical processes, see Lemma 2.14 in the proofs section.

To estimate  $\omega^2$ , we can either mimic Lavergne and Vuong (2000) to consider

$$\tilde{\omega}_n^2 = \frac{2h^p}{n^{(6)}} \sum_a (Y_i - Y_k)(Y_i - Y_{k'}) (Y_j - Y_l)(Y_j - Y_{l'}) L_{nik} L_{nik'} L_{njl} L_{njl'} K_{nij}^2 \psi_{ij}^2,$$

or generalize the variance estimator of Fan and Li (1996) as

$$\hat{\omega}_n^2 = \frac{2h^p}{n^{(2)}} \sum_a \hat{u}_i^2 \hat{f}_i^2 \hat{u}_j^2 \hat{f}_j^2 K_{nij}^2 \psi_{ij}^2.$$

The first one is consistent for  $\omega^2$  under both the null and alternative hypothesis, but the latter is faster to compute.

**Corollary 2.7.** *Let  $I_n$  be any of the statistics  $\hat{I}_n$  or  $\tilde{I}_n$  and let  $\omega_n$  denote any of  $\hat{\omega}_n$  or  $\tilde{\omega}_n$ . Under the assumptions of Theorem 2.6, the test that rejects  $H_0$  when  $nh^{p/2}I_n/\omega_n > z_{1-\alpha}$  is of asymptotic level  $\alpha$  under  $H_0$  and is consistent under the sequence of local alternatives  $H_{1n}$  provided  $\delta_n^2 nh^{p/2} \rightarrow \infty$ .*

### 2.3.3 Bootstrap Critical Values

It is known that asymptotic theory may be inaccurate for small and moderate samples when using smoothing methods. Hence, as in e.g. Härdle and Mammen (1993) or Delgado and González Manteiga (2001), we consider a wild bootstrap procedure to approximate the quantiles of our test statistic. Resamples are obtained from  $Y_i^* = \hat{r}_i + u_i^*$ , where  $u_i^* = \eta_i \hat{u}_i$  and  $\eta_i$  are i.i.d. variables independent of the initial sample with  $\mathbb{E}\eta_i = 0$  and  $\mathbb{E}\eta_i^2 = \mathbb{E}\eta_i^3 = 1$ ,  $1 \leq i \leq n$ . The  $\eta_i$  could for instance follow the two-point law of Mammen (1993). With at hand a bootstrap sample  $(Y_i^*, W_i, X_i)$ ,  $1 \leq i \leq n$ , we obtain a bootstrapped statistic  $nh^{p/2}I_n^*/\omega_n^*$  with bootstrapped observations  $Y_i^*$  in place of original observations  $Y_i$ . When the scheme is repeated many times, the bootstrap critical value  $z_{1-\alpha,n}^*$  at level  $\alpha$  is the empirical  $(1 - \alpha)$ -th quantile of the bootstrapped test statistics. The asymptotic validity of this bootstrap procedure is guaranteed by the following result.

**Theorem 2.8.** *Suppose Assumptions 2.4, 2.5, and Conditions (i) to (iii) of Theorem 2.6 hold. Moreover, assume  $\inf_{w \in \mathcal{S}_W} f(w) > 0$  and  $h/g^2 \rightarrow 0$ . Then for  $I_n^*$  equal to any of  $\hat{I}_n^*$  and  $\tilde{I}_n^*$ ,*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left[ nh^{p/2}I_n^*/\omega_n^* \leq z \mid Y_1, W_1, X_1, \dots, Y_n, W_n, X_n \right] - \Phi(z) \right| \xrightarrow{p} 0,$$

where  $\Phi(\cdot)$  is the standard normal distribution function.

## 2.4 Monte Carlo Study

We investigated the small sample behavior of our test and studied its performances relative to alternative tests. We generated data through

$$Y = (W'\theta)^3 - W'\theta + \delta d(X) + \varepsilon$$

where  $W$  follow a two-dimensional standard normal,  $X$  independently follows a  $q$ -variate standard normal,  $\varepsilon \sim \mathcal{N}(0, 4)$ , and we set  $\theta = (1, -1)'/\sqrt{2}$ . The null hypothesis corresponds to  $\delta = 0$ , and we considered various forms for  $d(\cdot)$  to investigate power. We only considered the test based on  $\tilde{I}_n$ , labelled LMP, as preliminary simulation results showed that it had similar or better performances than the test based on  $\hat{I}_n$ . We compared it to the test of Lavergne and Vuong (2000, hereafter LV), and the test of Delgado and Gonzalez-Manteiga (2001, hereafter DGM). The statistic for the latter test is the Cramer-von-Mises statistic

$$\sum_{i=1}^n \left[ \sum_{j=1}^n \hat{u}_j \hat{f}_j \mathbf{1}\{W_j \leq W_i\} \mathbf{1}\{X_j \leq X_i\} \right]^2,$$

and critical values are obtained by wild bootstrapping as for our own statistic. To compute bootstrap critical values, we used 199 bootstrap replications and the two-point distribution

$$\mathbb{P}\left(\eta_i = \frac{1 - \sqrt{5}}{2}\right) = \frac{5 + \sqrt{5}}{10}, \quad \mathbb{P}\left(\eta_i = \frac{1 + \sqrt{5}}{2}\right) = \frac{5 - \sqrt{5}}{10}.$$

For all tests, each time a kernel appears, we used the Epanechnikov kernel applied to the norm of its argument  $u$ , that is  $0.75 (1 - \|u\|^2) \mathbf{1}\{\|u\| < 1\}$ . The bandwidth parameters are set to  $g = n^{-1/6}$  and  $h = cn^{-2.1/6}$ , and we let  $c$  vary to investigate the sensitivity of our results to the smoothing parameter's choice. To study the influence of  $\psi(\cdot)$  on our test, we considered  $\psi(x) = l(\|x\|)$ , where  $l(\cdot)$  is a triangular or normal density, each with a second moment equal to one.

Figure 2.1 reports the empirical level of the various tests for  $n = 100$  based on 5000 replications when we let  $c$  and  $q$  vary. For our test, bootstrapping yields more accurate rejection levels than the asymptotic normal critical values for any bandwidth factor  $c$  and dimension  $q$ . The choice of  $\psi(\cdot)$  does not influence the results. The empirical level of LV test is much more sensitive to the bandwidth and the dimension. The empirical level of the DGM test is close to the nominal one for a low dimension  $q$ , but decreases with increasing  $q$ .

To investigate power, we considered different forms of alternatives as specified by  $d(\cdot)$ . We first focus on a quadratic alternative, where  $d(X) = (X'\beta - 1)^2/\sqrt{2}$ , with  $\beta = (1, 1, 1, \dots)'/\sqrt{q}$ . Figure 2.2 reports power curves of the different tests for the quadratic alternative,  $n = 100$ , and a nominal level of 10% based on 2000 replications. We also report the power of a Fisher test based on a linear specification



in the components of  $X$ . The power of our test, as well as the one of LV test, increases when the bandwidth factor  $c$  increases. This is in line with theoretical findings, though we may expect this relationship to revert for very large bandwidths. Our test always dominates LV test, as well as the Fisher test and DGM test, for any choice of  $c$  and any dimension  $q$ . The power of all tests decreases when the dimension  $q$  increases, but the more notable degradation is for the DGM test. In Figure 2.3, we let  $n$  vary for a fixed dimension  $q = 5$ . The power of all tests improve, but our main qualitative findings are not affected. It is noteworthy that the power advantage of our test compared to LV test become more pronounced as  $n$  increases. In Figure 2.4, we considered a linear alternative  $d(X) = X'\beta$  and a sine alternative,  $d(X) = \sin(2X'\beta)$ . Our main findings remain unchanged. For a linear alternative, the Fisher test is most powerful as expected. Compared to this benchmark, the loss of power when using our test is moderate for a large enough bandwidth factors  $c$ . For a sine alternative, our test is more powerful than the Fisher test for  $c = 2$  or  $4$ .

We also considered the case of a discrete  $X$ . We generated data following

$$Y = (W'\theta)^3 - W'\theta + \delta d(W) \mathbf{1}\{X = 1\} + \varepsilon$$

where  $W$  and  $\varepsilon$  are generated as before, and  $X$  is Bernoulli with probability of success  $p = 0.6$ . We compared our test to two competitors. The test proposed by Lavergne (2001) is similar to our test with the main difference that  $\psi(\cdot)$  is the indicator function, i.e.  $\psi(X_i - X_j) = \mathbf{1}\{X_i = X_j\}$ . The test of Neumeyer et Dette (2003, hereafter ND) is similar in spirit to the DGM test. The details of the simulations are similar to above. Figures 2.5 and 2.6 report our results. Bootstrapping our test and Lavergne's test yield accurate rejection levels, while the asymptotic tests and the ND test underrejects. Under a quadratic alternative, the power of our test is comparable to the one of the ND test for a large enough bandwidth factor  $c$ . Under a sine alternative, our test outperforms ND test in all cases.

## 2.5 Conclusion

We have developed a testing procedure for the significance of covariates in a nonparametric regression. Smoothing is entertained only for the covariates under the null hypothesis. The resulting test statistic is asymptotically pivotal, and wild bootstrap can be used to obtain critical values in small and moderate samples. The test is versatile, as it applies whether the covariates under test are continuous and/or discrete. Simulations reveal that our test outperforms its competitors in many situations, and especially when the dimension of covariates is large.

## 2.6 Proofs

We here provide the proofs of the main results. Technical lemmas are relegated to the Appendix.

In the following, for any integrable function  $\delta(X)$ , let  $\mathcal{F}_X[\delta](u) = \mathbb{E}[e^{-2\pi i \langle X, u \rangle} \delta(X)]$ ,  $u \in \mathbb{R}^q$ . Moreover, for any index set  $I$  not containing  $i$  with cardinality  $|I|$ , define

$$\widehat{f}_i^I = (n - |I| - 1)^{-1} \sum_{k \neq i, k \notin I} L_{nik},$$

consistent with  $\widehat{f}_i$  that corresponds to the case where  $I$  is the empty set.

### 2.6.1 Proof of Theorem 2.6

We first consider the case  $I_n = \widetilde{I}_n$ . Next, we study the difference between  $\widetilde{I}_n$  and  $\widehat{I}_n$  and hence deduce the result for  $I_n = \widehat{I}_n$ .

**Case  $I_n = \widetilde{I}_n$ .** Consider the decomposition

$$\begin{aligned} I_n &= \frac{1}{n^{(4)}} \sum_a (u_i - u_k)(u_j - u_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad + \frac{2}{n^{(4)}} \sum_a (u_i - u_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad + \frac{1}{n^{(4)}} \sum_a (r_i - r_k)(r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &= I_1 + 2I_2 + I_3, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i u_j f_i f_j K_{nij} \psi_{ij} + \frac{2(n-2)}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j f_j K_{nij} \psi_{ij} \\ &\quad + \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j (\widehat{f}_j^i - f_j) K_{nij} \psi_{ij} - \frac{2}{n^{(3)}} \sum_a u_i f_i u_l L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{2}{n^{(3)}} \sum_a u_i (\widehat{f}_i^{j,l} - f_i) u_l L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(4)}} \sum_a u_k u_l L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a u_i u_j L_{nik} L_{njk} K_{nij} \psi_{ij} \\ &= \frac{n-2}{n-3} [I_{0n} + 2I_{1,1} + I_{1,2}] - 2I_{1,3} - 2I_{1,4} + I_{1,5} - I_{1,6}, \end{aligned}$$

and

$$\begin{aligned} I_2 &= \frac{1}{n^{(3)}} \sum_a u_i f_i (r_j - r_l) L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(3)}} \sum_a u_i (\widehat{f}_i^{j,l} - f_i) (r_j - r_l) L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a u_k (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} = I_{2,1} + I_{2,2} - I_{2,3}. \end{aligned}$$

In Proposition 2.9 we prove that, under  $H_0$ ,  $I_{0n}$  is asymptotically centered Gaussian with variance  $\omega^2$ , while in Proposition 2.11 we prove that, under  $H_{1n}$ ,  $I_{0n}$  is asymptotically Gaussian with mean  $\mu$  and variance  $\omega^2$  provided  $\delta_n^2 n h^{p/2}$  converges to some positive real number. In Propositions 2.12 and 2.13 we show that all remaining terms in the decomposition of  $I_n$  are asymptotically negligible.

**Proposition 2.9.** *Under the conditions of Theorem 2.6,  $nh^{p/2} I_{0n} \xrightarrow{d} \mathcal{N}(0, \omega^2)$  under  $H_0$ .*

*Proof.* Let us define the martingale array  $\{S_{n,m}, \mathcal{F}_{n,m}, 1 \leq m \leq n, n \geq 1\}$  where  $S_{n,1} = 0$ , and

$$S_{n,m} = \sum_{i=1}^m G_{n,i} \quad \text{with} \quad G_{n,i} = \frac{2h^{p/2}}{n-1} u_i f_i \sum_{j=1}^{i-1} u_j f_j K_{nij} \psi_{ij}, \quad 2 \leq i, m \leq n,$$

and  $\mathcal{F}_{n,m}$  is the  $\sigma$ -field generated by  $\{W_1, \dots, W_n, X_1, \dots, X_n, Y_1, \dots, Y_m\}$ . Thus  $nh^{p/2} I_{0n} = S_{n,n}$ . Also define

$$V_n^2 = \sum_{i=2}^n E \left[ G_{n,i}^2 \mid \mathcal{F}_{n,i-1} \right] = \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sigma_i^2 f_i^2 \left( \sum_{j=1}^{i-1} u_j f_j K_{nij} \psi_{ij} \right)^2$$

where  $\sigma_i^2 = \sigma^2(W_i, X_i)$ . We can decompose  $V_n^2$  as

$$\begin{aligned} V_n^2 &= \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sigma_i^2 f_i^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} u_j f_j u_k f_k K_{nij} K_{nik} \psi_{ij} \psi_{ik} \\ &= \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \sigma_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \\ &\quad + \frac{8h^p}{(n-1)^2} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \sigma_i^2 f_i^2 u_j f_j u_k f_k K_{nij} K_{nik} \psi_{ij} \psi_{ik} = A_n + B_n. \end{aligned}$$

The result follows from the Central Limit Theorem for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). The conditions required for Corollary 3.1 of Hall and Heyde (1980), among which  $V_n^2 \xrightarrow{p} \omega^2$ , are checked in Lemma 2.10 below. Its proof is provided in the Appendix.

**Lemma 2.10.** *Under the conditions of Proposition 2.9,*

1.  $A_n \xrightarrow{p} \omega^2$ ,
2.  $B_n \xrightarrow{p} 0$ ,
3. the martingale difference array  $\{G_{n,i}, \mathcal{F}_{n,i}, 1 \leq i \leq n\}$  satisfies the Lindeberg condition

$$\forall \varepsilon > 0, \quad \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] \xrightarrow{p} 0.$$

□

**Proposition 2.11.** *Under the conditions of Theorem 2.6 and  $H_{1n}$ , if  $\delta_n^2 nh^{p/2} \rightarrow C$  with  $0 < C < \infty$ ,  $nh^{p/2} I_{0n} \xrightarrow{d} \mathcal{N}(C\mu, \omega^2)$ .*

*Proof.* Let  $\varepsilon_i = Y_i - \mathbb{E}[Y_i \mid W_i, X_i]$  and let us decompose

$$\begin{aligned} nh^{p/2} I_{0n} &= \frac{h^{p/2}}{n-1} \sum_{i=1}^n \sum_{j \neq i} u_i f_i u_j f_j K_{nij} \psi_{ij} \\ &= \frac{h^{p/2}}{n-1} \sum_{i=1}^n \sum_{j \neq i} (\delta_n d_i + \varepsilon_i) f_i (\delta_n d_j + \varepsilon_j) f_j K_{nij} \psi_{ij} \\ &= \frac{h^{p/2}}{n-1} \sum_{i=1}^n \sum_{j \neq i} \varepsilon_i f_i \varepsilon_j f_j K_{nij} \psi_{ij} + \frac{\delta_n h^{p/2}}{n-1} \sum_{i=1}^n \sum_{j \neq i} d_i f_i (\delta_n d_j + 2\varepsilon_j) f_j K_{nij} \psi_{ij} \\ &= C_{0n} + C_n. \end{aligned}$$

By Proposition 2.9,  $C_{0n} \xrightarrow{d} \mathcal{N}(0, \omega^2)$ . As for  $C_n$ , we have

$$\mathbb{E}[C_n] = \delta_n^2 nh^{p/2} \mathbb{E}[d_i f_i d_j f_j K_{nij} \psi_{ij}] = \delta_n^2 nh^{p/2} \mu_n.$$

By repeated application of Fubini's Theorem, Fourier Inverse formula, Dominated Convergence Theorem, and Parseval's identity, we obtain

$$\begin{aligned}
\mu_n &= \mathbb{E} [d_1 f_2 d_1 f_2 K_{n12} \psi_{12}] \\
&= \mathbb{E} \left[ \iint d(w_1, X_1) d(w_2, X_2) f(w_1) f(w_2) \right. \\
&\quad \left. \times f(w_1|X_1) f(w_2|X_2) h^{-p} K\left(\frac{w_1 - w_2}{h}\right) dw_1 dw_2 \psi(X_1 - X_2) \right] \\
&= \mathbb{E} \left[ \int \mathcal{F} [d(\cdot, X_1) f(\cdot) \pi(\cdot | X_1)](t) \right. \\
&\quad \left. \times \mathcal{F} [d(\cdot, X_2) f(\cdot) \pi(\cdot | X_2)](-t) \mathcal{F} [K](ht) dt \psi(X_1 - X_2) \right] \\
&\rightarrow \mathbb{E} \left[ \left[ \int \mathcal{F} [d(\cdot, X_1) f(\cdot) \pi(\cdot | X_1)](t) \right. \right. \\
&\quad \left. \left. \times \mathcal{F} [d(\cdot, X_2) f(\cdot) \pi(\cdot | X_2)](-t) dt \right] \psi(X_1 - X_2) \right] \\
&= \mathbb{E} \left[ \int d(w, X_1) d(w, X_2) f^2(w) \pi(w | X_1) \pi(w | X_2) \psi(X_1 - X_2) dw \right] \\
&= \int \left[ \int \mathcal{F}_X [d(w, \cdot) \pi(w | \cdot)](u) \right. \\
&\quad \left. \times \mathcal{F}_X [d(w, \cdot) \pi(w | \cdot)](-u) \mathcal{F}[\psi](u) du \right] f^2(w) dw \\
&= \iint |\mathcal{F}_X [d(w, \cdot) \pi(w | \cdot)](u)|^2 \mathcal{F}[\psi](u) f^2(w) dudw = \mu.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{Var} [C_n] &\leq \frac{4\delta_n^4 h^p}{(n-1)^2} \sum_a \mathbb{E} [d_i^2 f_i^2 d_k d_l f_k f_l K_{nik} K_{nil} \psi_{ik} \psi_{il}] \\
&\quad + \frac{2\delta_n^4 h^p}{(n-1)^2} \sum_a \mathbb{E} [d_i^2 f_i^2 d_k^2 f_k^2 K_{nik}^2 \psi_{ik}^2] \\
&\quad + \frac{4\delta_n^2 h^p}{(n-1)^2} \sum_a \mathbb{E} [d_i f_i d_j f_j \varepsilon_k^2 f_k^2 K_{nik} K_{nj k} \psi_{ik} \psi_{jk}] \\
&\quad + \frac{4\delta_n^2 h^p}{(n-1)^2} \sum_a \mathbb{E} [d_i^2 f_i^2 \varepsilon_k^2 f_k^2 K_{nik}^2 \psi_{ik}^2] \\
&= O(\delta_n^4 n h^p) + O(\delta_n^4) + O(\delta_n^2 n h^p) + O(\delta_n^2).
\end{aligned}$$

Therefore  $C_n = C\mu_n + O_p(\delta_n n^{1/2} h^{p/2}) \xrightarrow{p} C\mu$ , and the desired result follows.  $\square$

**Proposition 2.12.** *Under the conditions of Theorem 2.6,*

- (i)  $nh^{p/2}I_{1,3} = \delta_n \sqrt{nh^{p/2}} O_p(1) + o_p(1)$ ,
- (ii)  $nh^{p/2}I_{1,5} = o_p(1)$ ,
- (iii)  $nh^{p/2}I_{1,6} = \delta_n^2 nh^{p/2} o_p(1) + o_p(1)$ ,
- (iv)  $nh^{p/2}I_{2,1} = \delta_n \sqrt{nh^{p/2}} o_p(1) + \delta_n \sqrt{nh^{p/2}} g^s O_p(1) + o_p(1)$ ,
- (v)  $nh^{p/2}I_{2,3} = o_p(1)$ ,
- (vi)  $nh^{p/2}I_3 = nh^{p/2} O_p(g^{2s}) + o_p(1)$ .

**Proposition 2.13.** *Under the conditions of Theorem 2.6,*

- (i)  $nh^{p/2}I_{1,1} = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{nh^{p/2}} o_p(1) + o_p(1)$ ,
- (ii)  $nh^{p/2}I_{1,2} = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{nh^{p/2}} o_p(1) + o_p(1)$ ,
- (iii)  $nh^{p/2}I_{1,4} = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{nh^{p/2}} o_p(1) + (ng^p)^{-1/2} o_p(1) + o_p(1)$ ,
- (iv)  $nh^{p/2}I_{2,2} = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{nh^{p/2}} o_p(1) + o_p(1)$ .

The proofs of the above propositions follow the ones in Lavergne and Vuong (2000). For illustration, we provide in the Appendix the proofs of the first statements of each proposition.

**Case**  $I_n = \hat{I}_n$ . We have the following decomposition

$$n^{(4)} \tilde{I}_n = n(n-1)^3 \hat{I}_n - n^{(3)} V_{1n} - 2n^{(3)} V_{2n} + n^{(2)} V_{3n} \quad (2.2)$$

$$\begin{aligned} \text{where } V_{1n} &= \frac{1}{n^{(3)}} \sum_a (Y_i - Y_k)(Y_j - Y_k) L_{nik} L_{nj k} K_{nij} \psi_{ij}, \\ V_{2n} &= \frac{1}{n^{(3)}} \sum_a (Y_i - Y_j)(Y_j - Y_k) L_{nij} L_{nj k} K_{nij} \psi_{ij}, \\ \text{and } V_{3n} &= \frac{1}{n^{(2)}} \sum_a (Y_i - Y_j)^2 L_{nij}^2 K_{nij} \psi_{ij}. \end{aligned}$$

Hence, to show that  $\hat{I}_n$  has the same asymptotic distribution as  $\tilde{I}_n$ , it is sufficient to investigate the behavior of  $V_{1n}$  to  $V_{3n}$ . Using  $Y_i = r_i + u_i$ , it is straightforward to see that the dominating terms in  $V_{1n}$ ,  $V_{2n}$  and  $V_{3n}$  are

$$\begin{aligned} V_{13} &= \frac{1}{n^{(3)}} \sum_a (r_i - r_k)(r_j - r_k) L_{nik} L_{nj k} K_{nij} \psi_{ij}, \\ V_{23} &= \frac{1}{n^{(3)}} \sum_a (r_i - r_j)(r_j - r_k) L_{nij} L_{nj k} K_{nij} \psi_{ij}, \end{aligned}$$

$$V_{33} = \frac{1}{n^{(2)}} \sum_a (r_i - r_j)^2 L_{nij}^2 K_{nij} \psi_{ij},$$

respectively. Now

$$\begin{aligned} \mathbb{E}[|V_{13}|] &= \mathbb{E}[|(r_i - r_k)(r_j - r_k) L_{nik} L_{nj k} K_{nij}|] \\ &= O(g^{-p}) \mathbb{E}[|r_i - r_k| \mathbf{L}_{nik} \mathbb{E}[|r_j - r_k| \mathbf{K}_{nij} | Z_i, Z_k]] = O(g^{-p}), \\ \mathbb{E}[|V_{23}|] &= \mathbb{E}[|(r_i - r_j)(r_j - r_k) L_{nij} L_{nj k} K_{nij}|] \\ &= \mathbb{E}[\mathbb{E}[|r_j - r_k| \mathbf{L}_{nj k} | Z_j] |r_i - r_j| \mathbf{L}_{nij} \mathbf{K}_{nij}] \\ &= o(1) \mathbb{E}[|r_i - r_j| \mathbf{L}_{nij} \mathbf{K}_{nij}] = o(g^{-p}) \\ \mathbb{E}[|V_{33}|] &= \mathbb{E}[(r_i - r_j)^2 L_{nij}^2 | K_{nij}|] \\ &= O(g^{-2p}) \mathbb{E}[(r_i - r_j)^2 \mathbf{K}_{nij}] = o(g^{-2p}). \end{aligned}$$

It then follows that  $nh^{p/2}(\tilde{I}_n - \hat{I}_n) = O_p(h^{p/2}g^{-p})$  which is negligible if  $h/g^2 \rightarrow 0$ . The asymptotic irrelevance of the above diagonal terms thus require more restrictive relationships between the bandwidths  $h$  and  $g$ . For the sake of comparison, recall that Fan and Li (1996) impose  $h^{(p+q)}g^{-2p} \rightarrow 0$  while Lavergne and Vuong (2000) require only  $h^{p+q}g^{-p} \rightarrow 0$ . Since we do not smooth the covariates  $X$ , we are able to further relax the restriction between the two bandwidths.

## 2.6.2 Proof of Corollary 2.7

It suffices to prove  $\omega_n^2 - \omega^2 = o_p(1)$  with  $\omega_n^2$  any of  $\hat{\omega}_n^2$  or  $\tilde{\omega}_n^2$ . First we consider the case  $\omega_n^2 = \hat{\omega}_n^2$ . A direct approach would consist in replacing the definition of  $\hat{u}_i \hat{f}_i$  and  $\hat{u}_j \hat{f}_j$ , writing  $\hat{\omega}_n^2$  as a  $U$ -statistic of order 6, and studying its mean and variance. A shorter approach is based on empirical process tools. The price to pay is the stronger condition  $n^{7/8}g^p/\ln n \rightarrow \infty$  instead of  $ng^p \rightarrow \infty$ . Let  $\Delta \hat{f}_i = \hat{f}_i - f_i$ ,  $\Delta \hat{r}_i \hat{f}_i = \hat{r}_i \hat{f}_i - r_i f_i$ , and write

$$\hat{u}_i \hat{f}_i = u_i f_i + Y_i \Delta \hat{f}_i - \Delta \hat{r}_i \hat{f}_i. \quad (2.3)$$

**Lemma 2.14.** *Under Assumption 2.4, if  $r(\cdot)f(\cdot) \in \mathcal{U}^p$ ,  $L(\cdot)$  is a function of bounded variation,  $g \rightarrow 0$ , and  $n^{7/8}g^p/\ln n \rightarrow \infty$ , then*

$$\sup_{1 \leq i \leq n} \{|\Delta \hat{r}_i \hat{f}_i| + |\Delta \hat{f}_i|\} = o_p(1).$$

The proof relies on the uniform convergence of empirical processes and is provided in the Appendix. Now proceed as follows: square Equation (2.3), replace  $\hat{u}_i^2 \hat{f}_i^2$  in the definition of  $\hat{\omega}_n^2$ , and use Lemma 2.14 to deduce that

$$\hat{\omega}_n^2 = \frac{2h^p}{n^{(2)}} \sum_{a(2)} u_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 + o_p(1).$$

Elementary calculations of mean and variance yield

$$\frac{2h^p}{n^{(2)}} \sum_{a^{(2)}} u_i^2 f_i^2 u_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 - \omega^2 = o_p(1),$$

and thus  $\tilde{\omega}_n^2 - \omega^2 = o_p(1)$ .

To deal with  $\tilde{\omega}_n^2$ , note that  $\tilde{\omega}_n^2 - \hat{\omega}_n^2$  consists of “diagonal” terms plus a term which is  $O(n^{-1}\tilde{\omega}_n^2)$ . By tedious but rather straightforward calculations, one can check that such diagonal terms are each of the form  $n^{-1}g^{-p}$  times a  $U$ -statistic which is bounded in probability. Hence  $\tilde{\omega}_n^2 - \hat{\omega}_n^2 = o_p(1)$ .

### 2.6.3 Proof of Theorem 2.8

Let  $\bar{Z}$  denote the sample  $(Y_i, W_i, X_i)$ ,  $1 \leq i \leq n$ . Since the limit distribution is continuous, it suffices to prove the result pointwise by Polya’s theorem. Hence we show that  $\forall t \in \mathbb{R}$ ,  $\mathbb{P} \left[ nh^{p/2} I_n^* / \omega_n^* \leq t \mid \bar{Z} \right] - \Phi(t) = o_p(1)$ .

First, we consider the case  $I_n^* = \tilde{I}_n$ . Consider

$$\begin{aligned} I_{n,LV}^* &= \frac{1}{n^{(4)}} \sum_a (\eta_i \hat{u}_i - \eta_k \hat{u}_k) (\eta_j \hat{u}_j - \eta_l \hat{u}_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad + \frac{2}{n^{(4)}} \sum_a (\eta_i \hat{u}_i - \eta_k \hat{u}_k) (\hat{r}_j - \hat{r}_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad + \frac{1}{n^{(4)}} \sum_a (\hat{r}_i - \hat{r}_k) (\hat{r}_j - \hat{r}_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &= I_1^* + 2I_2^* + I_3^* \end{aligned}$$

where we can further decompose

$$\begin{aligned} I_1^* &= \frac{1}{n^{(4)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{2}{n^{(4)}} \sum_a \eta_j \hat{u}_j \eta_k \hat{u}_k L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad + \frac{1}{n^{(4)}} \sum_a \eta_k \hat{u}_k \eta_l \hat{u}_l L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &= I_{1,1}^* + I_{1,2}^* + I_{1,3}^* \end{aligned}$$



with

$$\begin{aligned}
I_{1,1}^* &= \frac{(n-1)^2}{(n-3)(n-4)} \times \frac{1}{n^{(2)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j \hat{f}_i \hat{f}_j K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n-4} \times \frac{1}{n^{(3)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{nij} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{n-4} \times \frac{1}{n^{(3)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nik} L_{nj k} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{(n-3)(n-4)} \times \frac{1}{n^{(2)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j L_{nij}^2 K_{nij} \psi_{ij} \\
&= I_{0n}^* - \frac{2}{n-4} I_{1,1,1}^* - \frac{1}{n-4} I_{1,1,2}^* - \frac{1}{(n-3)(n-4)} I_{1,1,3}^*.
\end{aligned}$$

Now let  $D_n^* = \tilde{I}_n^* - I_{0n}^*$  and write

$$\begin{aligned}
\mathbb{P} \left( \frac{nh^{p/2} \tilde{I}_n^*}{\tilde{\omega}_n^*} \leq t \mid \bar{Z} \right) &= \mathbb{P} \left( \frac{nh^{p/2} (I_{0n}^* + D_n^*)}{\tilde{\omega}_n^*} \leq t \mid \bar{Z} \right) \\
&= \mathbb{P} \left( \frac{nh^{p/2} I_{0n}^*}{\hat{\omega}_n} + \frac{nh^{p/2} D_n^*}{\hat{\omega}_n} \right. \\
&\quad \left. + \frac{nh^{p/2} (I_{0n}^* + D_n^*)}{\hat{\omega}_n} \left( \frac{\tilde{\omega}_n}{\hat{\omega}_n} - 1 \right) \leq t \mid \bar{Z} \right).
\end{aligned}$$

It thus suffices to prove that

$$\mathbb{P} \left( \frac{nh^{p/2} I_{0n}^*}{\hat{\omega}_{n,FL}} \leq t \mid \bar{Z} \right) - \Phi(t) \xrightarrow{p} 0 \quad \forall t \in \mathbb{R},$$

$$\frac{nh^{p/2} D_n^*}{\hat{\omega}_{n,FL}} = o_p(1), \quad \text{and} \quad \frac{nh^{p/2} (I_{0n}^* + D_n^*)}{\hat{\omega}_{n,FL}} \left( \frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,LV}^*} - 1 \right) = o_p(1). \quad (2.4)$$

The first result is stated below.

**Proposition 2.15.** *Under the conditions of Theorem 2.8, conditionally on the observed sample, the statistic  $nh^{p/2} I_{0n}^* / \hat{\omega}_{n,FL}$  converges in law to a standard normal distribution.*

*Proof.* We proceed as in the proof of Proposition 2.9 and check the conditions for a CLT for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). Define the martingale array  $\{S_{n,m}^*, \mathcal{F}_{n,m}^*, 1 \leq m \leq n, n \geq 1\}$  where  $\mathcal{F}_{n,m}^*$  is the  $\sigma$ -field generated by  $\{\bar{Z}, \eta_1, \dots, \eta_m\}$ ,  $S_{n,1}^* = 0$ , and  $S_{n,m}^* = \sum_{i=1}^m G_{n,i}^*$  with

$$G_{n,i}^* = \frac{2h^{p/2}}{n-1} \eta_i \hat{u}_i \sum_{j=1}^{i-1} \eta_j \hat{u}_j \hat{f}_i \hat{f}_j K_{nij} \psi_{ij}.$$

Then

$$I_{0n}^* = \frac{(n-1)^2}{(n-3)(n-4)} \times \frac{1}{n^{(2)}} \sum_a \eta_i \hat{u}_i \eta_j \hat{u}_j \hat{f}_i \hat{f}_j K_{nij} \psi_{ij} = \frac{(n-1)^2}{(n-3)(n-4)} S_{n,n}^*.$$

Now consider

$$\begin{aligned} V_n^{2*} &= \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^{2*} \mid \mathcal{F}_{n,i-1}^* \right] \\ &= \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} \hat{u}_i^2 \eta_j \eta_k \hat{u}_j \hat{u}_k \hat{f}_i^2 \hat{f}_j \hat{f}_k K_{nij} K_{nik} \psi_{ij} \psi_{ik} \\ &= \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \hat{u}_i^2 \eta_j^2 \hat{u}_j^2 \hat{f}_i^2 \hat{f}_j^2 K_{nij}^2 \psi_{ij}^2 \\ &\quad + \frac{8h^p}{(n-1)^2} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \hat{u}_i^2 \eta_j \eta_k \hat{u}_j \hat{u}_k \hat{f}_i^2 \hat{f}_j \hat{f}_k K_{nij} K_{nik} \psi_{ij} \psi_{ik} \\ &= A_n^* + B_n^*. \end{aligned}$$

Note that  $\mathbb{E} \left[ A_n^* \mid \bar{Z} \right] = [n/(n-1)] \mathbb{E} [\hat{\omega}_n^2]$  and that

$$\begin{aligned} \text{Var} \left[ \tilde{A}_n^* \mid \bar{Z} \right] &\leq \frac{16h^{2p} \mathbb{E} [\eta^4]}{(n-1)^4} \sum_{i=2}^n \sum_{i'=2}^n \sum_{j=1}^{i \wedge i'-1} \hat{u}_i^2 \hat{u}_{i'}^2 \hat{u}_j^4 \hat{f}_i^2 \hat{f}_{i'}^2 \hat{f}_j^4 K_{nij}^2 K_{ni'j}^2 \psi_{ij}^2 \psi_{i'j}^2 \\ &\leq \frac{16h^{2p} \mathbb{E} [\eta^4]}{(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \hat{u}_i^4 \hat{u}_j^4 \hat{f}_i^4 \hat{f}_j^4 K_{nij}^4 \psi_{ij}^4 \\ &\quad + \frac{32h^{2p} \mathbb{E} [\eta^4]}{(n-1)^4} \sum_{i=3}^n \sum_{i'=2}^{i-1} \sum_{j=1}^{i'-1} \hat{u}_i^2 \hat{u}_{i'}^2 \hat{u}_j^4 \hat{f}_i^2 \hat{f}_{i'}^2 \hat{f}_j^4 K_{nij}^2 K_{ni'j}^2 \psi_{ij}^2 \psi_{i'j}^2 \\ &= Q_{1n} + Q_{2n}. \end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E} \left[ B_n^{*2} \mid \bar{Z} \right] &= \frac{64h^{2p}}{(n-1)^4} \sum_{i=3}^n \sum_{i'=3}^n \sum_{j=2}^{i \wedge i' - 1} \sum_{k=1}^{j-1} \hat{u}_i^2 \hat{u}_{i'}^2 \hat{u}_j^2 \hat{u}_k^2 \hat{f}_i^2 \hat{f}_{i'}^2 \hat{f}_j^2 \hat{f}_k^2 K_{nij} K_{ni'j} \\
&\quad \times K_{nik} K_{ni'k} \psi_{ij} \psi_{i'j} \psi_{ik} \psi_{i'k} \\
&= \frac{64h^{2p}}{(n-1)^4} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \hat{u}_i^4 \hat{u}_j^2 \hat{u}_k^2 \hat{f}_i^4 \hat{f}_j^2 \hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2 \psi_{ik}^2 \\
&\quad + \frac{128h^{2p}}{(n-1)^4} \sum_{i=4}^n \sum_{i'=3}^{i-1} \sum_{j=2}^{i'-1} \sum_{k=1}^{j-1} \hat{u}_i^2 \hat{u}_{i'}^2 \hat{u}_j^2 \hat{u}_k^2 \hat{f}_i^2 \hat{f}_{i'}^2 \hat{f}_j^2 \hat{f}_k^2 K_{nij} K_{ni'j} \\
&\quad \times K_{nik} K_{ni'k} \psi_{ij} \psi_{i'j} \psi_{ik} \psi_{i'k} \\
&= Q_{3n} + Q_{4n}.
\end{aligned}$$

Finally the Lindeberg condition involves

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E} \left[ G_{n,i}^{2*} I \left( |G_{n,i}^*| > \varepsilon \right) \mid \mathcal{F}_{n,i-1}^* \right] \\
&\leq \frac{1}{\varepsilon^4} \sum_{i=1}^n \mathbb{E} \left[ G_{n,i}^{4*} \mid \mathcal{F}_{n,i-1}^* \right] \\
&\leq \frac{16h^{2p} \mathbb{E} [\eta^4]}{\varepsilon^4 (n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} \hat{u}_i^4 \hat{u}_j^2 \hat{u}_k^2 \hat{f}_i^4 \hat{f}_j^2 \hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2 \psi_{ik}^2 \\
&\leq \frac{16h^{2p} \mathbb{E} [\eta^4]}{\varepsilon^4 (n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \hat{u}_i^4 \hat{u}_j^4 \hat{f}_i^4 \hat{f}_j^4 K_{nij}^4 \psi_{ij}^4 \\
&\quad + \frac{32h^{2p} \mathbb{E} [\eta^4]}{\varepsilon^4 (n-1)^4} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \hat{u}_i^4 \hat{u}_j^2 \hat{u}_k^2 \hat{f}_i^4 \hat{f}_j^2 \hat{f}_k^2 K_{nij}^2 K_{nik}^2 \psi_{ij}^2 \psi_{ik}^2 \\
&= Q_{5n} + Q_{6n}.
\end{aligned}$$

It thus suffices to show that  $Q_{jn} = o_p(1)$ ,  $j = 1, \dots, 6$ . Now, there exist positive random variables  $\tilde{\gamma}_{1n}$  and  $\tilde{\gamma}_{2n}$  such that  $\tilde{\gamma}_{1n} + \tilde{\gamma}_{2n} = o_p(1)$  and

$$\hat{u}_i^{2k} \hat{f}_i^{2k} \leq 3^{2k-1} \left( u_i^{2k} f_i^{2k} + Y_i^{2k} f_i^{2k} \tilde{\gamma}_{1n}^{2k} + \tilde{\gamma}_{2n}^{2k} \right) \quad \forall 1 \leq i \leq n \text{ and } \forall k = 1, 2 \in \{1, 2\}.$$

Indeed,  $\hat{u}_i \hat{f}_i = u_i f_i + Y_i f_i f_i^{-1} (\hat{f}_i - f_i) + [\hat{r}_i \hat{f}_i - r_i f_i] = u_i f_i + Y_i f_i \gamma_{1i} - \gamma_{2i}$ , where  $\sup_{1 \leq i \leq n} |\gamma_{ji}| \leq \tilde{\gamma}_j$  and  $\tilde{\gamma}_j = o_p(1)$  by Lemma 2.14. Hence

$$\hat{u}_i^2 \hat{f}_i^2 \leq 3 \left( u_i^2 f_i^2 + Y_i^2 f_i^2 \tilde{\gamma}_{1n}^2 + \tilde{\gamma}_{2n}^2 \right).$$

The inequality for  $k = 2$  is obtained similarly. Using these inequalities, one can bound the expectations of  $|Q_{1n}|$  to  $|Q_{6n}|$  and thus show that  $|Q_{1n}| + \dots + |Q_{6n}| = o_p(1)$ .  $\square$

Next we show (2.4). First we need the following.

**Proposition 2.16.** *Under the conditions of Theorem 2.8, we have  $\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,FL}^*} \xrightarrow{p} 1$  and  $\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,LV}^*} \xrightarrow{p} 1$ .*

The proof uses the following result, which is proved in the Appendix.

**Lemma 2.17.** *Under the conditions of Theorem 2.8,  $\sup_{1 \leq i \leq n} |\hat{u}_i^* f_i - u_i^* f_i| = o_p(1)$ , where  $u_i^* = \eta_i \hat{u}_i$  and*

$$\hat{u}_i^* = Y_i^* - \frac{\sum_{k \neq i} Y_k^* L_{nik}}{\sum_{k \neq i} L_{nik}}.$$

*Proof.* Using Lemma 2.17, we have

$$\hat{\omega}_{n,FL}^{*2} = \omega_n^{*2} + o_p(1)$$

where  $\omega_n^{*2} = \frac{2h^p}{n^{(2)}} \sum_a u_i^{*2} u_j^{*2} \hat{f}_i^2 \hat{f}_j^2 K_{nij}^2 \psi_{ij}^2$ . Notice that  $\mathbb{E}[\omega_n^{*2} | \bar{Z}] = \hat{\omega}_{n,FL}^2$  and that

$$\text{Var}(\omega_n^{*2} - \hat{\omega}_{n,FL}^2) = \text{Var}(\mathbb{E}[\omega_n^{*2} - \hat{\omega}_{n,FL}^2 | \bar{Z}]) + \mathbb{E}[\text{Var}(\omega_n^{*2} | \bar{Z})]$$

where the first term is zero and

$$\text{Var}(\omega_n^{*2} | \bar{Z}) = \frac{8h^{2p} \text{Var}(\eta^2)}{\{n^{(2)}\}^2} \sum_a \hat{u}_i^4 \hat{u}_j^4 \hat{f}_i^4 \hat{f}_j^4 K_{nij}^4 \psi_{ij}^4.$$

Then,

$$\frac{\hat{\omega}_{n,FL}}{\hat{\omega}_{n,FL}^*} = 1 + \frac{\hat{\omega}_{n,FL} - \hat{\omega}_{n,FL}^*}{\hat{\omega}_{n,FL}^*} = 1 + \frac{o_p(1)}{\omega^2 [1 + o_p(1)]} = 1 + o_p(1).$$

Since  $\hat{\omega}_{n,LV}^* - \hat{\omega}_{n,FL}^*$  contains only diagonal terms, we deduce  $\hat{\omega}_{n,FL} / \hat{\omega}_{n,LV}^* \xrightarrow{p} 1$ .  $\square$

We next have to bound  $D_n^* = I_{n,LV}^* - I_{0n}^*$ . For this, let us decompose

$$\hat{r}_i - \hat{r}_k = (\hat{r}_i - r_i) - (\hat{r}_k - r_k) + (r_i - r_k)$$

and replace all such differences appearing in the definition of  $D_n^*$ . First, let us look

at  $I_3^*$  which does not contain any bootstrap variable  $\eta$ . We obtain

$$\begin{aligned}
I_3^* &= \frac{1}{n^{(4)}} \sum_a (\hat{r}_i - \hat{r}_k) (\hat{r}_j - \hat{r}_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= \frac{1}{n^{(4)}} \sum_a (r_i - r_k) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a (\hat{r}_i - r_i) (\hat{r}_j - r_j) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a (\hat{r}_k - r_k) (\hat{r}_l - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{2}{n^{(4)}} \sum_a (\hat{r}_i - r_i) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n^{(4)}} \sum_a (\hat{r}_k - r_k) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{2}{n^{(4)}} \sum_a (\hat{r}_k - r_k) (\hat{r}_j - r_j) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{3,1}^* + I_{3,2}^* + I_{3,3}^* + 2I_{3,4}^* - 2I_{3,5}^* - 2I_{3,6}^*.
\end{aligned}$$

Next, use the fact that

$$\begin{aligned}
\hat{r}_i - r_i &= (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} (Y_{i'} - r_i) L_{nii'} \\
&= (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} (r_{i'} - r_i) L_{nii'} + (n-1)^{-1} \hat{f}_i^{-1} \sum_{i' \neq i} u_{i'} L_{nii'} \quad (2.5)
\end{aligned}$$

and further replace terms like  $\hat{r}_i - r_i$ . Among the terms  $I_{3,1}^*$  to  $I_{3,6}^*$ , the term  $I_{3,1}^*$  could be easily handled with existing results in Lavergne and Vuong (2000). Namely  $nh^{p/2} I_{3,1}^* = nh^{p/2} O_p(g^{2s}) + o_p(1)$  by Proposition 7 of Lavergne and Vuong (2000). For the other five terms we have to control the density estimates appearing in the denominators. For this purpose, let us introduce the notation  $\Delta (f_i^I)^{-1} = (\hat{f}_i^I)^{-1} - f_i^{-1}$  and write

$$\frac{n-|I|}{n-1} \times \hat{f}_i^{-1} = \left( \frac{(n-|I|) \hat{f}_i^I}{(n-1) \hat{f}_i} - 1 \right) (\hat{f}_i^I)^{-1} + (\hat{f}_i^I)^{-1} = \frac{\sum_{k \in I} L_{nik}}{(n-1) \hat{f}_i \hat{f}_i^I} + \Delta (f_i^I)^{-1} + f_i^{-1}. \quad (2.6)$$

Then, we obtain for instance

$$\begin{aligned}
I_{3,5}^* &= \frac{1}{n^{(4)}} \sum_a (\hat{r}_k - r_k) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} f_k^{-1} (r_{k'} - r_k) (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \Delta \left( f_k^{i,j,l,k'} \right)^{-1} (r_{k'} - r_k) (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{(n-1)n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \left( \hat{f}_k \hat{f}_k^{i,j,l,k'} \right)^{-1} (L_{nik} + L_{njc} + L_{nlk} + L_{nk'k}) \\
&\quad \quad \quad \times (r_{k'} - r_k) (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} f_k^{-1} u_{k'} (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \Delta f_k^{-1} u_{k'} (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{(n-1)n^{(5)}} \sum_{a(4)} \sum_{k' \neq k} \left( \hat{f}_k \hat{f}_k^{i,j,l,k'} \right)^{-1} (L_{nik} + L_{njc} + L_{nlk} + L_{nk'k}) \\
&\quad \quad \quad \times u_{k'} (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{3,5,1}^* + I_{3,5,2}^* + I_{3,5,3}^* + I_{3,5,4}^* + I_{3,5,5}^* + I_{3,5,6}^*.
\end{aligned}$$

Next, if we consider for instance  $I_{3,5,1}^*$  that contains only terms like  $f_i^{-1}$  appearing from the decomposition 2.6, we obtain

$$\begin{aligned}
I_{3,5,1}^* &= \frac{1}{n^{(5)}} \sum_{a(5)} f_k^{-1} (r_{k'} - r_k) (r_j - r_l) L_{nkk'} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} (r_i - r_k) (r_j - r_l) L_{nik}^2 L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} (r_j - r_k) (r_j - r_l) L_{njc} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_{a(4)} f_k^{-1} (r_l - r_k) (r_j - r_l) L_{nlk} L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{3,5,1,1}^* + I_{3,5,1,2}^* + I_{3,5,1,3}^* + I_{3,5,1,4}^*
\end{aligned}$$

where the terms  $I_{3,5,1,2}^*$  to  $I_{3,5,1,4}^*$  are called ‘‘diagonal terms’’. Such terms require more restrictions on the bandwidths. next, the terms with containing terms like  $\Delta \left( f_i^I \right)^{-1}$  produced by the decomposition (2.6) can be treated like in the Propositions 8 to 11 of Lavergne and Vuong (2000). Finally, given that  $I$  is finite and with fixed cardinal

$$(n-1)^{-1} \hat{f}_i^{-1} \left( \hat{f}_i^I \right)^{-1} \sum_{k \in I} L_{nik} = O_p \left( n^{-1} g^{-p} \right) = o_p(1)$$

given that  $\|f^{-1}\|_\infty < \infty$ . Therefore the terms of  $I_3^*$  containing

$$(n-1)^{-1} \hat{f}_i^{-1} \left( \hat{f}_i^I \right)^{-1} \sum_{k \in I} L_{nik}$$

can be easily handled by taking absolute values. Now let us investigate the diagonal term  $I_{3,5,1,2}^*$ . We have

$$\begin{aligned} \mathbb{E} \left[ |I_{3,5,1,2}^*| \right] &= O(n^{-1}) \mathbb{E} \left[ f_k^{-1} |r_j - r_k| |r_j - r_l| |L_{njk}| |L_{nik}| |L_{njl}| |K_{nij}| \right] \\ &= O(n^{-1} g^{-p}) \mathbb{E} \left[ f_k^{-1} |r_j - r_k| |r_j - r_l| |L_{njk}| |L_{njl}| |K_{nij}| \right] \\ &= O(n^{-1} g^{-p}) \mathbb{E} \left[ f_k^{-1} |r_j - r_k| |L_{njk}| \mathbb{E} [|r_j - r_l| |L_{njl}| |Z_j]| |K_{nij}| \right] \\ &= o(n^{-1} g^{-p}) \mathbb{E} \left[ f_k^{-1} |r_j - r_k| |L_{njk}| |K_{nij}| \right] \\ &= o(n^{-1} g^{-p}). \end{aligned}$$

To prove that the term  $I_{3,5,1,2}^* = o_p(nh^{p/2})$  it suffices to prove  $\mathbb{E} \left[ |I_{3,5,1,2}^*| \right] = o(nh^{p/2})$  and this latter rate is implied by the condition  $h/g^2 = o(1)$ . This additional condition on the bandwidths is not surprising as the bootstrapped statistic introduced “diagonal” terms as in Fan and Li (1996) which indeed require the condition  $h/g^2 \rightarrow 0$ .

Let us now consider a term in the decomposition of  $D_n^*$  that involve bootstrap variables  $\eta$ , namely we investigate  $I_2^*$ . The arguments for the other terms are similar. Consider

$$\begin{aligned} I_2^* &= \frac{1}{n^{(4)}} \sum_a (\eta_i \hat{u}_i - \eta_k \hat{u}_k) (\hat{r}_j - \hat{r}_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &= \frac{1}{n^{(4)}} \sum_a \eta_i \hat{u}_i (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(4)}} \sum_a \eta_i \hat{u}_i (\hat{r}_j - r_j) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a \eta_i \hat{u}_i (\hat{r}_l - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} - \frac{1}{n^{(4)}} \sum_a \eta_k \hat{u}_k (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &\quad - \frac{1}{n^{(4)}} \sum_a \eta_k \hat{u}_k (\hat{r}_j - r_j) L_{nik} L_{njl} K_{nij} \psi_{ij} + \frac{1}{n^{(4)}} \sum_a \eta_k \hat{u}_k (\hat{r}_l - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\ &= I_{2,1}^* + I_{2,2}^* - I_{2,3}^* - I_{2,4}^* - I_{2,5}^* + I_{2,6}^*. \end{aligned}$$

Next it suffices to use the fact that

$$\hat{u}_i = u_i - \hat{f}_i^{-1} \sum_{i' \neq i} u_{i'} L_{ni i'} + \hat{f}_i^{-1} \sum_{i' \neq i} (r_i - r_{i'}) L_{ni i'}.$$

For instance, using this identity with  $I_{2,1}^*$  we can write

$$\begin{aligned}
I_{2,1}^* &= \frac{1}{n^{(4)}} \sum_a \eta_i u_i (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{(n-1)n^{(4)}} \sum_a \sum_{i' \neq i} \hat{f}_i^{-1} \eta_i u_{i'} (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{(n-1)n^{(4)}} \sum_a \sum_{i' \neq i} \hat{f}_i^{-1} \eta_i (r_i - r_{i'}) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= \frac{1}{n^{(3)}} \sum_a \eta_i u_i f_i (r_j - r_l) L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a \eta_i u_i (r_j - r_l) \Delta f_i^{j,l} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{(n-1)n^{(4)}} \sum_a \sum_{i' \neq i} f_i^{-1} \eta_i u_{i'} (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{n^{(5)}} \sum_a \Delta \left( f_i^{j,k,l,i'} \right)^{-1} \eta_i u_{i'} (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad - \frac{1}{(n-1)n^{(4)}} \sum_a \sum_{i' \neq i} \left( \hat{f}_i \hat{f}_i^{j,k,l,i'} \right)^{-1} \eta_i u_{i'} (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{(n-1)n^{(4)}} \sum_a \sum_{i' \neq i} f_i^{-1} \eta_i (r_i - r_{i'}) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_a \Delta \left( f_i^{j,k,l,i'} \right)^{-1} \eta_i (r_i - r_{i'}) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&\quad + \frac{1}{n^{(5)}} \sum_a \left( \hat{f}_i \hat{f}_i^{j,k,l,i'} \right)^{-1} \eta_i (r_i - r_{i'}) (r_j - r_l) L_{nik} L_{njl} K_{nij} \psi_{ij} \\
&= I_{2,1,1}^* + I_{2,1,2}^* + I_{2,1,3}^* + I_{2,1,4}^* + I_{2,1,5}^* + I_{2,1,6}^* + I_{2,1,7}^* + I_{2,1,8}^*
\end{aligned}$$

Handling one problem at a time, let us notice that  $I_{2,1,1}^*$  is a zero-mean  $U$ -statistic of order three with kernel  $H_n \left( Z_i^*, Z_j^*, Z_l^* \right) = \eta_i u_i f_i (r_j - r_l) L_{njl} K_{nij} \psi_{ij}$  where  $Z_i^* = (Y_i, W_i, X_i, \eta_i)$ . Using the Hoeffding decomposition of term  $I_{2,1,1}^*$  in degenerate  $U$ -statistics, it is easy to check that the third and second order projections are small. For the first order degenerate  $U$ -statistic it suffices to note that  $\mathbb{E} \left[ H_n \mid Z_j^* \right] = \mathbb{E} \left[ H_n \mid Z_l^* \right] = 0$  and  $\mathbb{E} \left[ H_n \mid Z_i^* \right] = \eta_i u_i f_i \mathbb{E} \left[ (r_j - r_l) L_{njl} K_{nij} \psi_{ij} \mid Z_i \right]$  so that

$$\begin{aligned}
\mathbb{E} \left[ \mathbb{E}^2 \left[ H_n \mid Z_i^* \right] \right] &= \mathbb{E} \left[ \eta_i^2 u_i^2 f_i^2 \mathbb{E}^2 \left[ (r_j - r_l) L_{njl} K_{nij} \psi_{ij} \mid Z_i \right] \right] \\
&= \mathbb{E} \left[ u_i^2 f_i^2 \mathbb{E}^2 \left[ (r_j - r_l) L_{njl} K_{nij} \psi_{ij} \mid Z_i \right] \right]
\end{aligned}$$

which, given that  $\|\psi\|_\infty < \infty$ , is similar to the term  $\xi_1$  bounded in the proof of Proposition 5 of Lavergne and Vuong (2000).



Finally, let us briefly consider the case  $I_n^* = \tilde{I}_n$ . Like in the decomposition (2.2), we have

$$n(n-1)^3 I_{n,FL}^* = n^{(4)} I_{n,LV}^* + n^{(3)} V_{1n}^* + 2n^{(3)} V_{2n}^* - n^{(2)} V_{3n}^*$$

where  $\forall j \in \{1, 2, 3\}$ , the  $V_{jn}^*$ s are obtained by replacing the  $Y_i$ s by the  $Y_i^*$ s in the  $V_{jn}$ s. All these terms could be handled by arguments similar to the ones detailed above. The proof of Theorem 2.8 is now complete.

## Appendix (not for publication)

We here provide proofs of technical lemmas and additional details for the proofs in the manuscript. We define  $Z_i = (Y_i, W_i, X_i)$ ,  $\|\psi\|_\infty = \sup_{x \in \mathbb{R}^q} |\psi(x)|$ ,

$$\mathbf{K}_{nij} = |K_{nij}| = \frac{1}{h^p} \left| K \left( \frac{W_i - W_j}{h} \right) \right|, \quad \text{and} \quad \mathbf{L}_{nij} = |L_{nij}| = \frac{1}{g^p} \left| L \left( \frac{W_i - W_j}{g} \right) \right|.$$

*Proof of Lemma 2.10.* 1. We have

$$\mathbb{E}[A_n] = \frac{4h^p}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \right] = \frac{2nh^p}{n-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \right],$$

and

$$\begin{aligned} \text{Var}[A_n] &\leq \frac{64h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{j'=1}^{j-1} \mathbb{E} \left[ \sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_{j'}^2 f_{j'}^2 K_{nij}^2 K_{nij'}^2 \right] \\ &\quad + \frac{32h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^n \sum_{i'=1}^{i-1} \sum_{j=2}^{i'-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j^4 f_j^4 K_{nij}^2 K_{ni'j}^2 \right] \\ &\quad + \frac{16h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} \left[ \sigma_i^4 f_i^4 u_j^4 f_j^4 K_{nij}^4 \right] \\ &= O(n^{-1}) \mathbb{E} \left[ \sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij} \mathbf{K}_{nik} \right] \\ &\quad + O(n^{-1}) \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j^4 f_j^4 \mathbf{K}_{nij} \mathbf{K}_{ni'j} \right] \\ &\quad + O(n^{-2} h^{-p}) \mathbb{E} \left[ \sigma_i^4 f_i^4 u_j^4 f_j^4 \mathbf{K}_{nij} \right] \\ &= O(n^{-1}) + O(n^{-2} h^{-p}). \end{aligned}$$

Deduce that  $\text{Var}[A_n] \rightarrow 0$ , and hence remains to show that  $\mathbb{E}[A_n] \rightarrow \omega^2$ . We have

$$h^p \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \right] = \mathbb{E} \left[ \int \varphi_{X_i}(t) \varphi_{X_j}(-t) \mathcal{F}[K^2](ht) \psi^2(X_i - X_j) dt \right]$$

where  $\varphi_x(t) = \mathcal{F}[\sigma^2(\cdot, x) f^2(\cdot) \pi(\cdot | x)](t)$ . Let us note that

$$\begin{aligned} \mathbb{E} \left[ \int |\varphi_{X_i}(t) \varphi_{X_j}(-t)| \psi^2(X_i - X_j) dt \right] &\leq \|\psi\|_\infty \mathbb{E} \left[ \int |\varphi_X(t)|^2 dt \right] \\ &= \|\psi\|_\infty \mathbb{E} \left[ \sigma^4(W, X) f^4(W) \pi(W | X) \right], \end{aligned}$$

by Plancherel Theorem. Moreover,  $\mathcal{F}[K^2](ht)$  is bounded and converges pointwise to  $\int K^2(s) ds$  as  $h \rightarrow 0$ . Then by Lebesgue's dominated convergence theorem,

$$h^p \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \right] \rightarrow \mathbb{E} \left[ \int \varphi_{X_i}(t) \varphi_{X_j}(-t) \psi^2(X_i - X_j) dt \right] \int K^2(s) ds = \omega^2,$$

by Parseval's Theorem.

2. By elementary calculations,

$$\begin{aligned}
\mathbb{E} \left[ B_n^2 \right] &= \frac{64h^{2p}}{(n-1)^4} \sum_{i=3}^n \sum_{i'=3}^n \sum_{j=2}^{i-1} \sum_{j'=2}^{i'-1} \sum_{k=1}^{j-1} \sum_{k'=1}^{j'-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 u_j f_j u_{j'} f_{j'} u_k f_k u_{k'} f_{k'} \right. \\
&\quad \left. \times K_{nij} K_{ni'j'} K_{nik} K_{ni'k'} \psi_{ij} \psi_{i'j'} \psi_{ik} \psi_{i'k'} \right] \\
&\leq \frac{64h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^n \sum_{i'=3}^n \sum_{j=2}^{i \wedge i'-1} \sum_{k=1}^{j-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij} K_{ni'j} K_{nik} K_{ni'k} \right] \\
&= \frac{64h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \mathbb{E} \left[ \sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij}^2 K_{nik}^2 \right] \\
&\quad + \frac{128h^{2p} \|\psi\|_\infty^4}{(n-1)^4} \sum_{i=3}^n \sum_{i'=3}^{i-1} \sum_{j=2}^{i'-1} \sum_{k=1}^{j-1} \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 K_{nij} K_{ni'j} K_{nik} K_{ni'k} \right] \\
&= O(n^{-1}) \mathbb{E} \left[ \sigma_i^4 f_i^4 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij} \mathbf{K}_{nik} \right] \\
&\quad + O(h^p) \mathbb{E} \left[ \sigma_i^2 f_i^2 \sigma_{i'}^2 f_{i'}^2 \sigma_j^2 f_j^2 \sigma_k^2 f_k^2 \mathbf{K}_{nij} \mathbf{K}_{ni'j} \mathbf{K}_{nik} \right] \\
&= O(n^{-1}) + O(h^p) = o(1).
\end{aligned}$$

3. We have  $\forall \varepsilon > 0$ ,  $\forall n \geq 1$ , and  $1 < i \leq n$ ,

$$\begin{aligned}
\mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] &\leq \mathbb{E}^{1/2} \left[ G_{n,i}^4 \mid \mathcal{F}_{n,i-1} \right] \mathbb{E}^{1/2} \left[ I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] \\
&\leq \frac{\mathbb{E} \left[ G_{n,i}^4 \mid \mathcal{F}_{n,i-1} \right]}{\varepsilon^2}.
\end{aligned}$$

Then

$$\begin{aligned}
\sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] &\leq \frac{1}{\varepsilon^2} \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^4 \mid \mathcal{F}_{n,i-1} \right] \\
&\leq \frac{1}{\varepsilon^2} \frac{16h^{2p}}{(n-1)^4} \sum_{i=2}^n \mathbb{E} \left[ u_i^4 f_i^4 \mid W_i, X_i \right] \\
&\quad \times \left( \sum_{j=1}^{i-1} u_j K_{nij} \psi_{ij} \right)^4 \\
&\leq \frac{1}{\varepsilon^2} \frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^n \left( \sum_{j=1}^{i-1} u_j K_{nij} \psi_{ij} \right)^4,
\end{aligned}$$

where  $\kappa_4$  is any constant that bounds  $\mathbb{E}[u^4 f^4 \mid W, X]$ . The last expression that

multiplies  $\varepsilon^{-2}$  is positive and has expectation

$$\begin{aligned}
& \frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^n \sum_{j_1=1}^{i-1} \sum_{j_2=1}^{i-1} \sum_{j_3=1}^{i-1} \sum_{j_4=1}^{i-1} \mathbb{E} [u_{j_1} f_{j_1} u_{j_2} f_{j_2} u_{j_3} f_{j_3} u_{j_4} f_{j_4} \\
& \quad \times K_{ni j_1} K_{ni j_2} K_{ni j_3} K_{ni j_4} \psi_{ij_1} \psi_{ij_2} \psi_{ij_3} \psi_{ij_4}] \\
&= \frac{96\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=3}^n \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \mathbb{E} [u_j^2 f_j^2 u_k^2 f_k^2 K_{ni j}^2 K_{ni k}^2 \psi_{ij}^2 \psi_{ik}^2] \\
& \quad + \frac{16\kappa_4 h^{2p}}{(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} [u_j^4 f_j^4 K_{ni j}^4 \psi_{ij}^4] \\
&= O(n^{-1}) \mathbb{E} [u_j^2 f_j^2 u_k^2 f_k^2 \mathbf{K}_{ni j} \mathbf{K}_{ni k}] + O(n^{-2} h^{-p}) \mathbb{E} [u_j^4 f_j^4 \mathbf{K}_{ni j}] \\
&= O(n^{-1}) + O(n^{-2} h^{-p}).
\end{aligned}$$

The desired result follows.  $\square$

The following result, known as Bochner's Lemma (see Theorem 1.1.1. of Bochner (1955)) will be repeatedly use in the following. We recall it for the sake of completeness.

**Lemma 2.18.** *For any function  $l(\cdot) \in \mathcal{U}^p$  and any integrable kernel  $K(\cdot)$ ,*

$$\sup_{x \in \mathbb{R}^p} \left| \int l(y) \frac{1}{h^p} K\left(\frac{x-y}{h}\right) dy - l(x) \int K(u) du \right| \rightarrow 0.$$

In the following we provide the proofs for rates for the remaining terms in the decomposition of  $I_n$ , see Propositions 2.12 and 2.13. For this purpose, we use the following a decomposition for  $U$ -statistics that can be found in Lavergne and Vuong (2000): if  $U_n = (1/n^{(m)}) \sum_a H_n(Z_{i_1}, \dots, Z_{i_m})$ , then

$$\begin{aligned}
\mathbb{E} [U_n^2] &= \left( \frac{1}{n^{(m)}} \right)^2 \sum_{c=0}^m \frac{n^{(2m-c)}}{c!} \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1, \Delta_2) \\
&= \sum_{c=0}^m O(n^{-c}) \sum_{|\Delta_1|=c=|\Delta_2|}^{(c)} I(\Delta_1, \Delta_2),
\end{aligned}$$

where  $\sum^{(c)}$  denotes summation over sets  $\Delta_1$  and  $\Delta_2$  of ordered positions of length  $c$ ,

$$I(\Delta_1, \Delta_2) = \mathbb{E} [H_n(Z_{i_1}, \dots, Z_{i_m}) H_n(Z_{j_1}, \dots, Z_{j_m})]$$

and the  $i$ 's position in  $\Delta_1$  coincide with the  $j$ 's position in  $\Delta_2$  and are pairwise distinct otherwise. Now, we will bound  $\mathbb{E}[U_n^2]$  using the  $\xi_c = \sum^{(c)} I(\Delta_1, \Delta_2)$  and the fact that by Cauchy's inequality,

$$\begin{aligned}
I^2(\Delta_1, \Delta_2) &= \mathbb{E}^2 [\mathbb{E} [H_n(Z_{i_1}, \dots, Z_{i_m}) | Z_c] \mathbb{E} [H_n(Z_{j_1}, \dots, Z_{j_m}) | Z_c]] \\
&\leq \mathbb{E} [\mathbb{E}^2 [H_n(Z_{i_1}, \dots, Z_{i_m}) | Z_c]] \mathbb{E} [\mathbb{E}^2 [H_n(Z_{j_1}, \dots, Z_{j_m}) | Z_c]]
\end{aligned}$$

where  $Z_c$  denotes the common  $Z_i$ 's.

*Proof of Proposition 2.12.* After bounding the  $\psi_{ij}$ 's by  $\|\psi\|_\infty$  the arguments are very similar to those used in Lavergne and Vuong (2000). We prove only the first statement.

(i)  $I_{1,3}$  is a U-statistic with kernel  $H_n(Z_i, Z_j, Z_l) = u_i f_i u_l L_{njl} K_{nij} \psi_{ij}$ . We need to bound the  $\xi_c$ ,  $c = 0, 1, 2, 3$ .

1.  $\mathbb{E}[H_n] = 0$ , thus  $\xi_0 = 0$ .

2.  $\xi_1 = O(\delta_n^2)$ . Indeed, we have  $\mathbb{E}[H_n | Z_l] = \delta_n u_l \mathbb{E}[d_i f_i L_{njl} K_{nij} \psi_{ij} | Z_l]$  and  $\mathbb{E}[H_n | Z_i] = 0 = \mathbb{E}[H_n | Z_j]$ . Then

$$\begin{aligned} \mathbb{E} \left[ \mathbb{E}^2 [H_n | Z_l] \right] &\leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E} \left[ u_l^2 \mathbb{E}^2 [d_i f_i L_{njl} K_{nij} | Z_l] \right] \\ &= O(\delta_n^2) \mathbb{E} \left[ u_l^2 \mathbb{E}^2 [L_{njl} d_j f_j^2 | Z_l] \right] = O(\delta_n^2). \end{aligned}$$

3.  $\xi_2 = O(g^{-p})$ . Indeed, we have

$$\begin{aligned} \mathbb{E}[H_n | Z_i, Z_j] &= u_i f_i K_{nij} \psi_{ij} \mathbb{E}[u_l L_{njl} | Z_j] = 0, \\ \mathbb{E}[H_n | Z_i, Z_l] &= u_i f_i u_l \mathbb{E}[L_{njl} K_{nij} \psi_{ij} | Z_i, Z_l], \\ \mathbb{E}[H_n | Z_j, Z_l] &= u_l L_{njl} \mathbb{E}[u_i f_i K_{nij} \psi_{ij} | Z_j] \\ &= \delta_n u_l L_{njl} \mathbb{E}[d_i f_i K_{nij} \psi_{ij} | Z_j]. \end{aligned}$$

By successive applications of Lemma 2.18,

$$\begin{aligned} \mathbb{E} \left[ \mathbb{E}^2 (H_n | Z_i, Z_l) \right] &\leq \|\psi\|_\infty^2 \mathbb{E} \left[ u_i^2 f_i^2 u_l^2 \mathbb{E} [L_{njl} K_{nij} | Z_i, Z_l] \mathbb{E} [L_{njl} K_{nij'} | Z_i, Z_l] \right] \\ &= O(g^{-p}) \mathbb{E} \left[ u_i^2 f_i^2 u_l^2 \mathbb{E} [\mathbf{L}_{njl} \mathbf{K}_{nij} | Z_i, Z_l] \mathbb{E} [\mathbf{K}_{nij'} | Z_i, Z_l] \right] \\ &= O(g^{-p}) \mathbb{E} \left[ u_i^2 f_i^3 u_l^2 \mathbf{L}_{njl} \mathbf{K}_{nij} \right] = O(g^{-p}), \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left[ \mathbb{E}^2 [H_n | Z_j, Z_l] \right] &\leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E} \left[ u_l^2 L_{njl}^2 \mathbb{E}^2 [d_i f_i K_{nij} | Z_j] \right] \\ &\leq O(\delta_n^2) \mathbb{E} \left[ u_l^2 L_{njl}^2 d_j^2 f_j^4 \right] \\ &= O(\delta_n^2) O(g^{-p}) \mathbb{E} \left[ u_l^2 \mathbf{L}_{njl} d_j^2 f_j^4 \right] = O(g^{-p}). \end{aligned}$$

4.  $\xi_3 = O(g^{-p} h^{-p})$ , as  $\mathbb{E}[H_n^2]$  equals

$$\mathbb{E} \left[ u_i^2 u_l^2 f_i^2 L_{njl}^2 K_{nij}^2 \psi_{ij}^2 \right] = O(g^{-p} h^{-p}) \mathbb{E} \left[ u_i^2 u_l^2 f_i^2 \mathbf{L}_{njl} \mathbf{K}_{nij} \right] = O(g^{-p} h^{-p}).$$

Collecting results,  $\mathbb{E} \left[ \left( n h^{p/2} I_{1,3} \right)^2 \right] = O(\delta_n^2 n h^p) + O(h^p/g^p) + O(n^{-1} g^{-p}) = o(1)$ .  $\square$

*Proof of Proposition 2.13.* As in Proposition 2.12, we only prove the first statement. We will use the following lemma, which is similar to Lemma 2 of Lavergne and Vuong (2000), and whose proof is then omitted.

**Lemma 2.19.** *Let  $\Delta f_i^j = \widehat{f}_i^j - f_i$ . If  $f(\cdot) \in \mathcal{U}^p$  and  $ng^p \rightarrow \infty$ , then*

$$\mathbb{E} \left[ \Delta^2 f_i^j \mid Z_i, Z_j, Z_{i'}, Z_{j'} \right] = o(1)$$

and

$$\mathbb{E} \left[ \Delta^2 f_i^{j,l} \mid Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'} \right] = o(1)$$

uniformly in the indices.

(i) Let us denote  $\Delta f_i^j = \widehat{f}_i^j - f_i$ . We have  $I_{1,1} = (1/n^{(2)}) \sum_a u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij}$  so that

$$\mathbb{E} \left[ I_{1,1}^2 \right] = \left( \frac{1}{n^{(2)}} \right)^2 \left[ \sum_a u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij} \right] \left[ \sum_a u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'} \psi_{i'j'} \right], \quad (2.7)$$

where the first (respectively the second) sum is taken over all arrangements of different indices  $i$  and  $j$  (respectively different indices  $i'$  and  $j'$ ). Let  $\overline{W}$  denote the sample of  $W_i$ ,  $1 \leq i \leq n$ , and let  $\lambda_n = E \left[ \Delta^2 f_i^j \mid Z_i, Z_j, Z_{i'}, Z_{j'} \right]$ . By Lemma 2.19,  $\lambda_n = o(1)$  uniformly in the indices. By Equation (2.7),  $\mathbb{E} \left[ I_{1,1}^2 \right]$  is equal to a normalized sum over four indices. This sum could split in three sums of the following types.

1. All indices are different, that is a sum of  $n^{(4)}$  terms. Each term in the sum can be bounded as follows:

$$\begin{aligned} & \mathbb{E} \left[ u_i \Delta f_i^j u_j f_j K_{nij} \psi_{ij} u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'} \psi_{i'j'} \right] \\ & \leq \|\psi\|_\infty^2 \delta_n^4 \mathbb{E} \left[ \Delta f_i^j f_j \Delta f_{i'}^{j'} f_{j'} \mathbb{E} \left[ d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} \mid \overline{W} \right] \right] \\ & \leq \|\psi\|_\infty^2 \delta_n^4 \mathbb{E} \left[ f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} \mathbb{E} \left[ \Delta f_i^j \Delta f_{i'}^{j'} \mid Z_i, Z_j, Z_{i'}, Z_{j'} \right] \right] \\ & \leq O(\delta_n^4 \lambda_n) \mathbb{E} \left[ f_j f_{j'} d_i d_j d_{i'} d_{j'} K_{nij} K_{ni'j'} \right] = O(\delta_n^4 \lambda_n). \end{aligned}$$

2. One index is common to  $\{i, j\}$  and  $\{i', j'\}$ , that is a sum of  $4n^{(3)}$  terms. For each of such terms we can write

$$\begin{aligned} (i' = i) \quad & \mathbb{E} \left[ u_i^2 \Delta f_i^j u_j f_j K_{nij} \psi_{ij} \Delta f_i^{j'} u_{j'} f_{j'} K_{ni'j'} \psi_{i'j'} \right] \\ & \leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E} \left[ \Delta f_i^j f_j \Delta f_i^{j'} f_{j'} E \left[ u_i^2 d_j d_{j'} K_{nij} K_{ni'j'} \mid \overline{W} \right] \right] \\ & \leq O(\delta_n^2 \lambda_n) \mathbb{E} \left[ f_j f_{j'} u_i^2 d_j d_{j'} K_{nij} K_{ni'j'} \right] = O(\delta_n^2 \lambda_n), \end{aligned}$$

$$\begin{aligned} (j' = j) \quad & \mathbb{E} \left[ u_i \Delta f_i^j u_j^2 f_j^2 K_{nij} \psi_{ij} u_{i'} \Delta f_{i'}^j K_{ni'j} \psi_{i'j} \right] \\ & \leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E} \left[ \Delta f_i^j f_j^2 \Delta f_{i'}^j \mathbb{E} \left[ d_i u_j^2 d_{i'} K_{nij} K_{ni'j} \mid \overline{W} \right] \right] \\ & \leq O(\delta_n^2 \lambda_n) \mathbb{E} \left[ f_j^2 d_i u_j^2 d_{i'} K_{nij} K_{ni'j} \right] = O(\delta_n^2 \lambda_n), \end{aligned}$$

$$\begin{aligned} (i' = j) \quad & \mathbb{E} \left[ u_i \Delta f_i^j u_j^2 f_j K_{nij} \psi_{ij} \Delta f_j^{j'} u_{j'} f_{j'} K_{njj'} \psi_{jj'} \right] \\ & \leq \|\psi\|_\infty^2 \delta_n^2 \mathbb{E} \left[ \Delta f_i^j f_j \Delta f_j^{j'} f_{j'} E \left[ d_i u_j^2 d_{j'} K_{nij} K_{njj'} \mid \overline{W} \right] \right] \\ & \leq O(\delta_n^2 \lambda_n) \mathbb{E} \left[ f_j f_{j'} d_i u_j^2 d_{j'} K_{nij} K_{njj'} \right] = O(\delta_n^2 \lambda_n). \end{aligned}$$

The case  $j' = i$  is similar to  $i' = j$ .

3. Two indices in common to  $\{i, j\}$  and  $\{i', j'\}$ , that is a sum of  $2n^{(2)}$  terms. For each term in the sum we can write

$$\mathbb{E} \left[ u_i^2 u_j^2 (\Delta f_i^j)^2 f_j^2 K_{nij}^2 \psi_{ij}^2 \right] = O(\lambda_n h^{-p})$$

and

$$\mathbb{E} \left[ u_i^2 u_j^2 \Delta f_i^j \Delta f_j^i f_i f_j K_{nij}^2 \psi_{ij}^2 \right] = O(\lambda_n h^{-p}).$$

Therefore,  $\mathbb{E} \left[ (nh^{p/2} I_{1,1})^2 \right] = \delta_n^4 n^2 h^p O(\lambda_n) + \delta_n^2 n h^p O(\lambda_n) + O(\lambda_n) = O(\lambda_n)$ . The result then follows from Lemma 2.19.  $\square$

*Proof of Lemma 2.14.* We only prove the result for  $\Delta \hat{r}_i \hat{f}_i$ , as the reasoning is similar for  $\Delta \hat{f}_i$ . We have

$$\begin{aligned} \Delta \hat{r}_i \hat{f}_i &= \frac{1}{(n-1)g^p} \sum_{k \neq i} \left\{ Y_k L((W_i - W_k)g^{-1}) - \mathbb{E} \left[ Y L((W_i - W)g^{-1}) \right] \right\} \\ &\quad + \mathbb{E} \left[ r(W)g^{-p} L((W_i - W)g^{-1}) \right] - r(W_i)f(W_i) \\ &= \Delta_{1i} + \Delta_{2i}. \end{aligned}$$

The uniform continuity of  $r(\cdot)f(\cdot)$  implies  $\sup_i |\Delta_{2i}| = o_p(1)$  by Lemma 2.18. For  $\sup_i |\Delta_{1i}|$ , we use empirical process tools. Let us introduce some notation. Let  $\mathcal{G}$  be a class of functions of the observations with envelope function  $G$  and let

$$J(\delta, \mathcal{G}, L^2) = \sup_Q \int_0^\delta \sqrt{1 + \ln N(\varepsilon \|G\|_2, \mathcal{G}, L^2(Q))} d\varepsilon, \quad 0 < \delta \leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions  $Q$  on the space of the observations, and  $\|G\|_2$  denotes the norm of  $G$  in  $L^2(Q)$ . Let  $Z_1, \dots, Z_n$  be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma(Z_i), \quad \gamma \in \mathcal{G}$$

be the empirical process indexed by  $\mathcal{G}$ . If the covering number  $N(\varepsilon, \mathcal{G}, L^2(Q))$  is of polynomial order in  $1/\varepsilon$ , there exists a constant  $c > 0$  such that  $J(\delta, \mathcal{G}, L^2) \leq c\delta \sqrt{\ln(1/\delta)}$  for  $0 < \delta < 1/2$ . Now if  $\mathbb{E}\gamma^2 < \delta^2 \mathbb{E}G^2$  for every  $\gamma$  and some  $0 < \delta < 1$ , and  $\mathbb{E}G^{(4v-2)/(v-1)} < \infty$  for some  $v > 1$ , under mild additional measurability conditions, Theorem 3.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n \gamma| = J(\delta, \mathcal{G}, L^2) \left( 1 + \frac{J(\delta^{1/v}, \mathcal{G}, L^2) \|G\|_{(4v-2)/(v-1)}^{2-1/v}}{\delta^2 \sqrt{n} \|G\|_2^{2-1/v}} \right)^{v/(2v-1)} \|G\|_2 O_p(1),$$

(2.8)

where  $\|G\|_2^2 = \mathbb{E}G^2$  and the  $O_p(1)$  term is independent of  $n$ . Note that the family  $\mathcal{G}$  could change with  $n$ , as soon as the envelope is the same for all  $n$ . We apply this result to the family of functions  $\mathcal{G} = \{YL((W - w)/g) : w \in \mathbb{R}^p\}$  for a sequence  $g$  that converges to zero and the envelope  $G(Y, W) = Y \sup_{w \in \mathbb{R}^p} L(w)$ . Its entropy number is of polynomial order in  $1/\varepsilon$ , independently of  $n$ , as  $L(\cdot)$  is of bounded variation, see for instance van der Vaart and Wellner (1996). Now for any  $\gamma \in \mathcal{G}$ ,  $\mathbb{E}\gamma^2(Y, W) \leq Cg^p\mathbb{E}G^2(Y, W)$ , for some constant  $C$ . Let  $\delta = g^{3p/7}$ , so that  $\mathbb{E}\gamma^2(Y, W) \leq C'\delta^2\mathbb{E}G^2(Y, W)$ , for some constant  $C'$  and  $v = 3/2$ , which corresponds to  $\mathbb{E}G^8 < \infty$  that is guaranteed by our assumptions. The bound in (2.8) thus yields

$$\sup_{\mathcal{G}} \left| \frac{1}{g^p \sqrt{n}} \mathbb{G}_n \gamma \right| = \frac{\ln^{1/2}(n)}{g^{4p/7} \sqrt{n}} \left[ 1 + n^{-1/2} g^{-4p/7} \ln^{1/2}(n) \right]^{3/4} O_p(1),$$

where the  $O_p(1)$  term is independent of  $n$ . Since  $n^{7/8}g^p/\ln n \rightarrow \infty$ , the expected result follows.  $\square$

*Proof of Lemma 2.17.* We have

$$\begin{aligned} \hat{u}_i^* \hat{f}_i &= \frac{1}{n-1} \sum_{k \neq i} (Y_i^* - Y_k^*) L_{nik} \\ &= u_i^* \hat{f}_i - \frac{1}{n-1} \sum_{k \neq i} u_k^* L_{nik} + \frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik} \end{aligned}$$

where

$$\begin{aligned} \frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik} &= \frac{1}{n-1} \sum_{k \neq i} (r_i - r_k) L_{nik} + (\hat{r}_i - r_i) \hat{f}_i \\ &\quad - \frac{1}{(n-1)^2 \hat{f}_k} \sum_{k \neq i} \sum_{k' \neq k} (r_{k'} - r_k) L_{nkk'} L_{nik} \\ &\quad - \frac{1}{(n-1)^2 \hat{f}_k} \sum_{k \neq i} \sum_{k' \neq k} u_{k'} L_{nkk'} L_{nik}. \end{aligned}$$

By Lemma 2.14 and the fact that  $f(\cdot)$  is bounded away from zero, deduce that  $\sup_i |\hat{r}_i - r_i| = o_p(1)$ . From this and applying several times the arguments in the proof of Lemma 2.14 we obtain

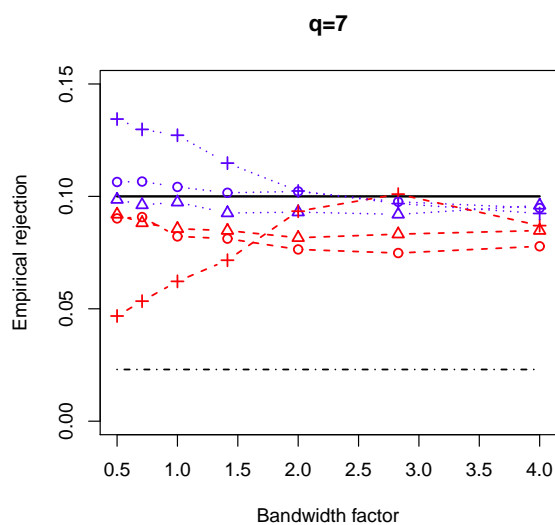
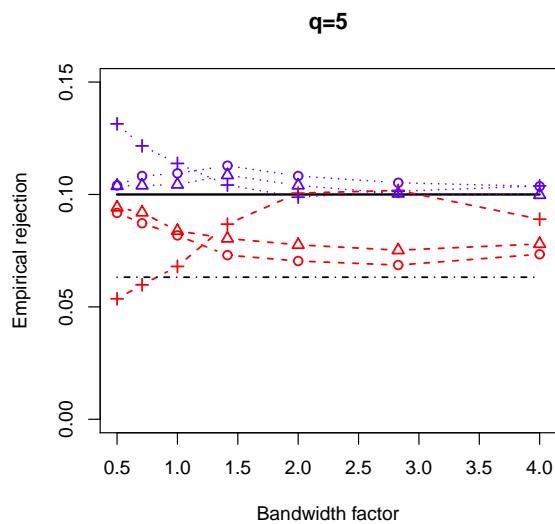
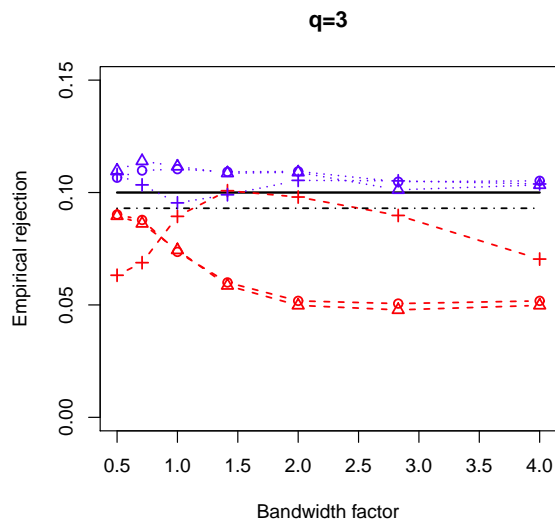
$$\frac{1}{n-1} \sum_{k \neq i} (\hat{r}_i - \hat{r}_k) L_{nik} = o_p(1).$$

On the other hand,

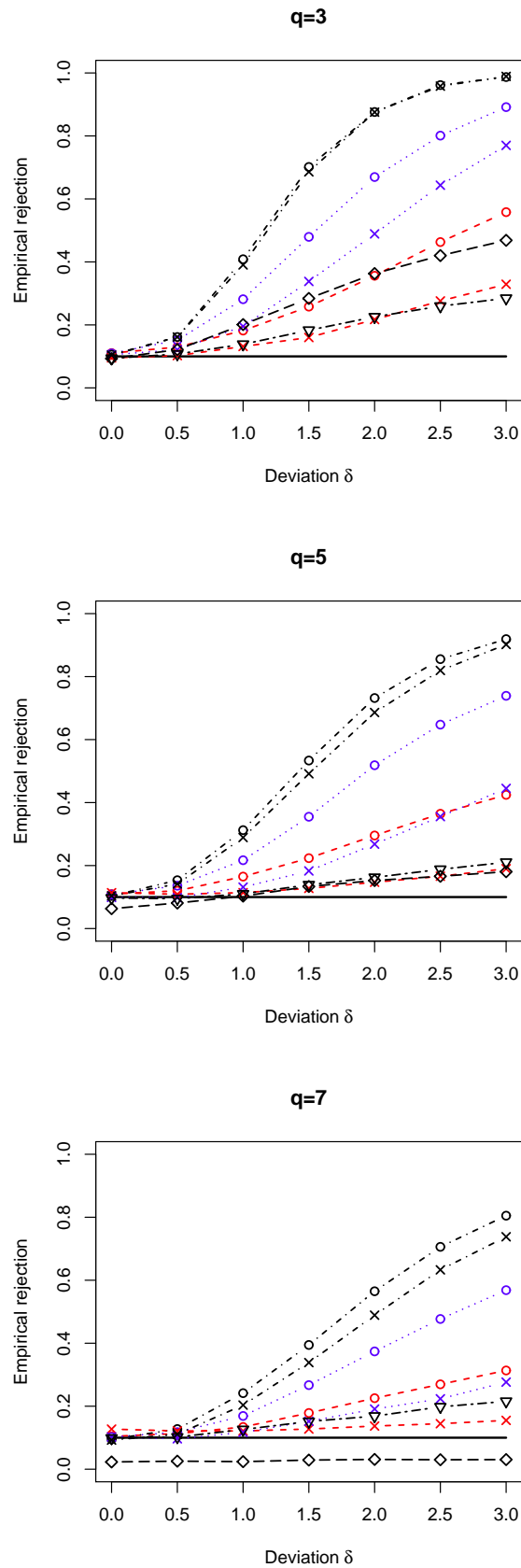
$$\begin{aligned} \left| \frac{1}{n-1} \sum_{k \neq i} u_k^* L_{nik} \right| &\leq \left| \frac{1}{n-1} \sum_{k \neq i} \eta_k u_k L_{nik} \right| + \frac{\sup_j |\hat{r}_j - r_j|}{n-1} \sum_{k \neq i} |\eta_k| \mathbf{L}_{nik} \\ &= o_p(1), \end{aligned}$$



where we used again the arguments for  $\Delta_{1i}$  in the proof of Lemma 2.14 (here with  $\eta_k u_k$  and  $|\eta_k|$  in the place of  $Y_k$ ) to derive the last rate.  $\square$

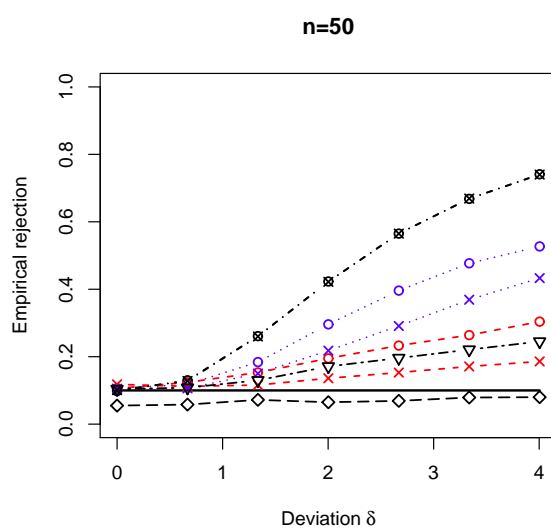


- - - Theoretical level
- △- LMP with triangle  $\psi$
- LMP with gaussian  $\psi$
- △· Bootstrap LMP with triangle  $\psi$
- Bootstrap LMP with gaussian  $\psi$
- +- LV
- +· Bootstrap LV
- Bootstrap DGM



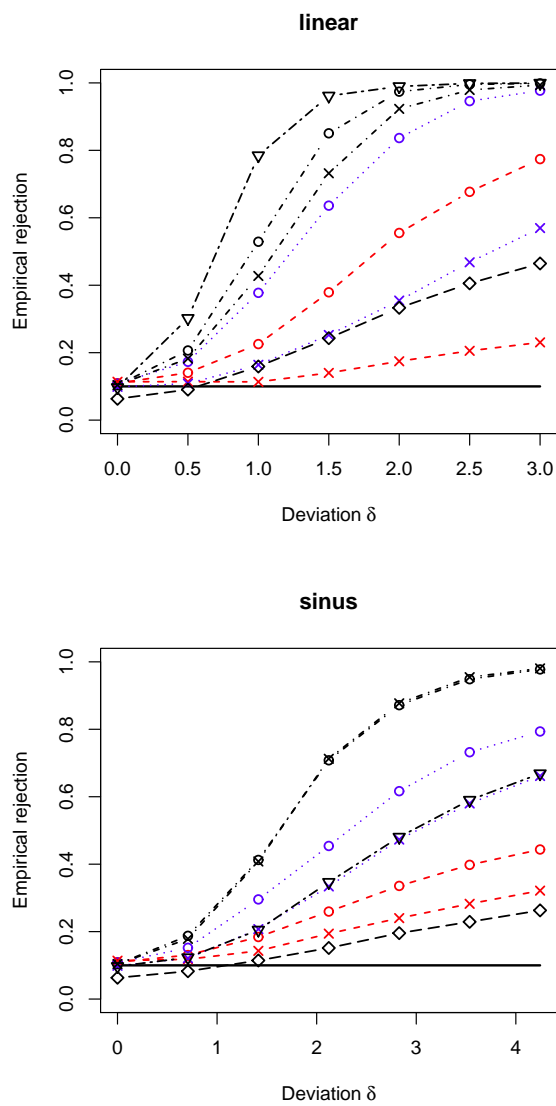
includegraphics[scale=0.6,angle=270]grapheContPowerLegend

**Figure 2.2:** Empirical power curves for a quadratic alternative,  $n = 100$

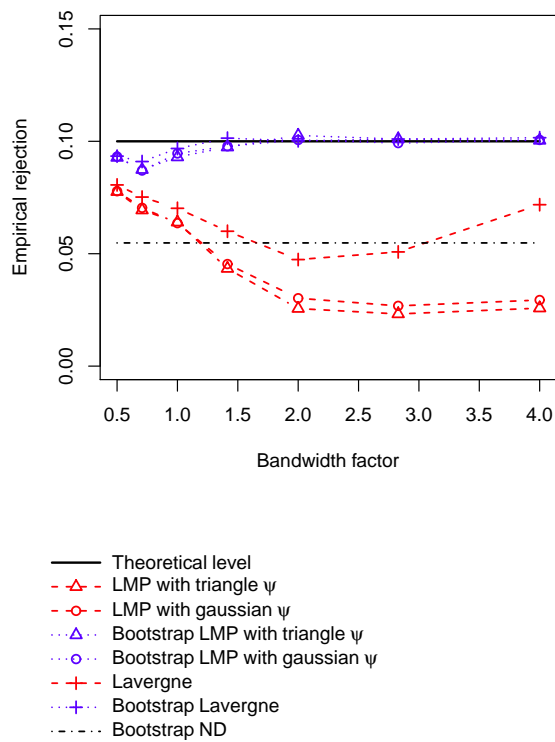


includegraphics[scale=0.6,angle=270]grapheContPowern200quad

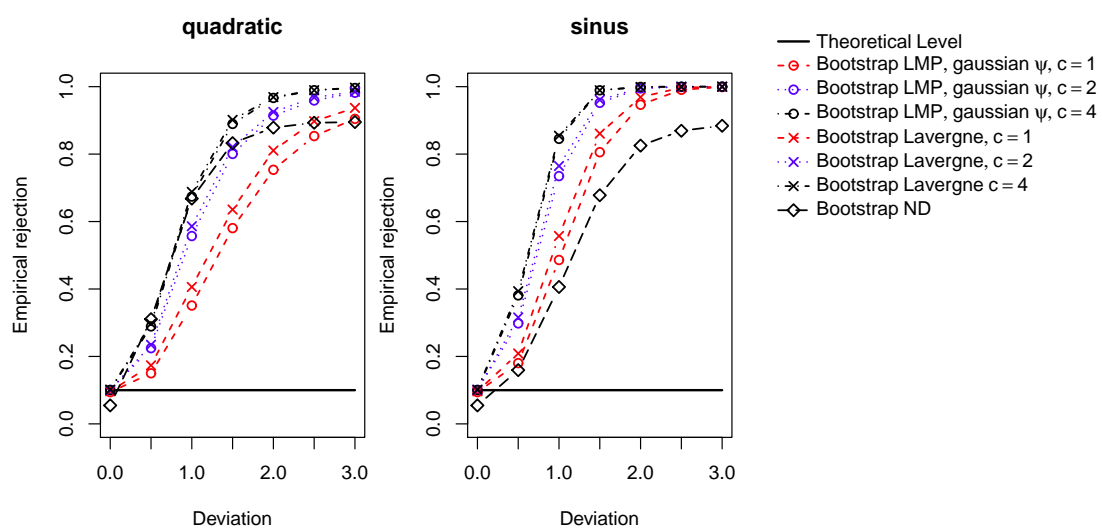
**Figure 2.3:** Empirical power curves for a quadratic alternative,  $q = 5$



**Figure 2.4:** Empirical power curves for linear and sine alternative,  $n = 100$  and  $q = 5$



**Figure 2.5:** Empirical rejection under  $H_0$  as a function of the bandwidth,  $X$  Bernoulli and  $n = 100$



**Figure 2.6:** Empirical power curves,  $X$  Bernoulli and  $n = 100$

# 3 Test d'adéquation pour la régression quantile

Ce chapitre a été déposé sur la plate-forme d'enregistrement électronique arXiv. Il porte l'identification arXiv : 1404.0216 [math.ST].





## **Abstract**

We address the issue of lack-of-fit testing for a parametric quantile regression. We propose a simple test that involves one-dimensional kernel smoothing, so that the rate at which it detects local alternatives is independent of the number of covariates. The test has asymptotically gaussian critical values, and wild bootstrap can be applied to obtain more accurate ones in small samples. Our procedure appears to be competitive with existing ones in simulations. We illustrate the usefulness of our test on birthweight data.

Keywords: Quantile regression, Omnibus test, Smoothing.

MSC2000: Primary 62G10

### 3.1 Introduction

Quantile regression, as introduced by Koenker and Bassett (1978), has emerged as an alternative to mean regression. It allows for a richer data analysis by exploring the effect of covariates at different quantiles of the conditional distribution of the variable of interest. Parametric quantile regression generalizes usual regression and is particularly valuable if variables have asymmetric distributions or heavy tails. Koenker's monograph (2005) and the review of Yu et al. (2003) detail the theory and practice of quantile regression.

As in any statistical modeling exercise, it is crucial to check the fit of a parametric quantile model. There has been a large effort devoted to testing of the fit of parametric mean regressions, however only few lack-of-fit tests of parametric quantile regressions. He and Zhu (2003) extend the approach of Stute (1997) and is based on a vector-weighted cumulative summed process of the residuals. Bierens and Ginther (2002) generalize the integrated conditional moment test of Bierens and Ploberger (1997) to quantile regression. In both cases, the limit distribution of the test statistic is a non-linear functional of a Gaussian process, so that implementation may require rather involved computations to obtain critical values. Zheng (1998) use kernel smoothing over the design space, to obtain an asymptotically pivotal test statistic. Horowitz and Spokoiny (2002) extend such an approach and propose an adaptive procedure to choose the smoothing parameter. As in any multidimensional nonparametric problem, the curse of dimensionality may be detrimental to the performances of the test, see e.g. Lavergne and Patilea (2012) for illustrations.

In this paper, we introduce a new testing methodology that avoids multidimensional smoothing, but still yield an omnibus test. Our test has three specific features. First, it does not require smoothing with respect to all covariates under test. This allows to mitigate the curse of dimensionality that appears with nonparametric smoothing, hence improving the power properties of the test. Second, the test statistic is asymptotically pivotal, while wild bootstrap can be used to obtain small samples critical values of the test. This yields a test whose level is well controlled by bootstrapping, as shown in simulations. Third, our test equally applies whether some of the covariates are discrete.

The paper is organized as follows. In Section 2, we present our testing procedure, we study its asymptotic behavior under the null hypothesis and under a sequence of local alternatives, and we establish the validity of wild bootstrap. In Section 3, we compare the small sample behavior of our test to some existing procedures, and we illustrate its use on birthweight data. Section 3 concludes. Section 4 gathers our technical proofs.

## 3.2 Lack-of-Fit Test for Quantile Regression

### 3.2.1 Principle and Test

Consider modeling the quantile of a real random variable  $Y$  conditional upon covariates  $Z \in \mathbb{R}^q$ ,  $q \geq 1$ . We assume that  $Z = (W, X)'$ , where  $W$  is continuous and admits a density with respect to the Lebesgue measure, while  $X$  may include both continuous and discrete variables. Formally, if  $F(\cdot | z)$  denotes the conditional distribution of  $Y$  given  $Z = z$ , the  $\tau$ -th conditional quantile is  $Q_\tau(z) = \inf\{y : F(y | z) \geq \tau\}$ . Assuming  $F(\cdot | z)$  is absolutely continuous for almost all  $z$ , this is equivalent to  $F(Q_\tau(z) | z) = \tau$ . The parametric quantile regression model of interest posits that the conditional  $\tau$ -th quantile of  $Y$  is given by  $g(Z; \beta_0)$ , where  $g(\cdot; \beta)$  is known up to the parameter vector  $\beta \in B \subset \mathbb{R}^p$ , that is,

$$Y = g(Z; \beta_0) + \varepsilon, \quad F(g(Z; \beta_0) | Z) = \tau. \quad (3.1)$$

The validity of our parametric quantile regression is thus equivalent to

$$H_0 : \exists \beta_0 \in B : F(g(Z; \beta_0) | Z) - \tau = \mathbb{E}\{\mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau | Z\} = 0 \text{ a.s.} \quad (3.2)$$

Hence testing the the correct specification of our parametric quantile regression models reduces to testing a zero conditional mean hypothesis. The alternative hypothesis is then

$$H_1 : \mathbb{P}[\mathbb{E}\{\mathbb{I}\{Y \leq g(Z; \beta)\} - \tau | Z\} = 0] < 1 \quad \text{for any } \beta \in B.$$

The key element of our testing approach is the following lemma. See also Lavergne et al. (2014) for a related result. First let us introduce some notation. Hereafter, if  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  is an integrable function,  $\mathcal{F}[g]$  denotes its Fourier transform, that is

$$\mathcal{F}[g](t) = \int_{\mathbb{R}^k} \exp(-2\pi i t' u) g(u) du.$$

**Lemma 3.1.** *Let  $(W_1, X_1, U_1)$  and  $(W_2, X_2, U_2)$  be two independent draws of vector  $(W, X, u)$ , and  $K(\cdot)$  and  $\psi(\cdot)$  even functions with (almost everywhere) positive Fourier integrable transforms. Define*

$$I(h) = \mathbb{E}\left[U_1 U_2 h^{-p} K((W_1 - W_2)/h) \psi(X_1 - X_2)\right].$$

Then for any  $h > 0$ ,  $\mathbb{E}[U | W, X] = 0$  a.s.  $\Leftrightarrow I(h) = 0$ .

**Proof.** Let  $\langle \cdot, \cdot \rangle$  denote the standard inner product and  $\mathcal{F}[K]$  be the Fourier transform of  $K(\cdot)$ . Using Fourier Inversion Theorem, change of variables, and elementary properties of conditional expectation,

$$\begin{aligned} I(h) &= \mathbb{E}\left[U_1 U_2 \int_{\mathbb{R}^p} e^{2\pi i \langle t, W_1 - W_2 \rangle} \mathcal{F}[K](th) dt \int_{\mathbb{R}^q} e^{2\pi i \langle s, X_1 - X_2 \rangle} \mathcal{F}[\psi](s) ds\right] \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left|\mathbb{E}\left[U | W, X\right] e^{2\pi i \{\langle t, W \rangle + \langle s, X \rangle\}}\right|^2 \mathcal{F}[K](th) \mathcal{F}[\psi](s) dt ds. \end{aligned}$$

Since the Fourier transforms  $\mathcal{F}[K]$  and  $\mathcal{F}[\psi]$  are strictly positive,  $I(h) = 0$  iff

$$\mathbb{E} \left[ \mathbb{E}[U \mid W, X] e^{2\pi i\{t, W\} + \{s, X\}} \right] = 0 \quad \forall t, s \Leftrightarrow \mathbb{E}[U \mid W, X] = 0 \quad \text{a.s.} \quad \blacksquare$$

From the above results, it is sufficient to test whether  $I(h) = 0$  for any arbitrary  $h$ . We chose to consider a sequence of  $h$  decreasing to zero when the sample size increases, which is one of the ingredient that allows to obtain a tractable asymptotic distribution for the test statistic. Assume we have at hand a random sample  $(Y_i, W_i, X_i)$ ,  $1 \leq i \leq n$ , from  $(Y, W, X)$ . Then we can estimate  $I(h)$  by the second-order U-statistic

$$I_n(\beta_0) = I_n(\beta_0; h) = \frac{1}{n(n-1)} \sum_{1 \leq j \neq i \leq n} U_i(\beta_0) U_j(\beta_0) \frac{1}{h} K_h(W_i - W_j) \psi(X_i - X_j)$$

where  $U_i(\beta) = \mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau$  and  $K_h(\cdot) = K(\cdot/h)$ .

For estimating  $\beta_0$ , we follow Koenker and Bassett (1978), who showed that under (3.1) a consistent estimator of  $\beta_0$  is obtained by minimizing

$$\arg \min_{\beta} \sum_{i=1}^n \rho_{\tau}(Y_i - g(Z_i; \beta)), \quad (3.3)$$

where  $\rho_{\tau}(e) = (\tau - \mathbb{I}(e < 0)) e$  is the so-called check function. While this is not a differentiable optimization problem, it is convex and tractable, see e.g. Koenker (2005) for some computational algorithms. Let us define

$$T_n = nh^{1/2} \frac{I_n(\hat{\beta})}{v_n} \quad \text{where} \quad v_n^2 = \frac{2\tau^2(1-\tau)^2}{n(n-1)} \sum_{j \neq i} h^{-1} K_h^2(W_i - W_j) \psi^2(X_i - X_j). \quad (3.4)$$

An asymptotic  $\alpha$ -level test of  $H_0$  is then

Reject  $H_0$  if  $T_n \geq z_{\alpha}$ , where  $z_{\alpha}$  is the  $(1 - \alpha)$ -quantile of the standard normal distribution.

Our test statistic is very similar to the one proposed by Zheng (1998), but the latter uses smoothing on all components of  $Z$  while we smooth only on the first component  $W$ .

The statistic  $v_n^2$  is the variance of  $nh^{1/2}I_n(\beta_0)$  conditional on the  $Z_i$  under  $H_0$ . In general,  $v_n^2$  does not consistently estimate the conditional variance of  $nh^{1/2}I_n(\beta)$  under the alternative hypothesis. In some cases  $v_n^2$  overestimates this conditional variance (this is certainly the case for misspecified median regression model because  $\tau(1 - \tau)$  attains the maximum value at  $\tau = 1/2$ ), so that the test may suffer some power loss. In a mean regression context, Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) proposed to use a nonparametric estimator of the conditional variance. This might be adapted to quantile regression, but in simulations our test appears to be well-behaved and more powerful than competitors, so we decided in favor of the simplest estimator  $v_n^2$ .

### 3.2.2 Behavior Under the Null Hypothesis

To derive the asymptotic properties of our lack-of-fit test, we introduce our set of assumptions on the data-generating process, the parametric model (3.1), the functions  $K(\cdot)$  and  $\psi(\cdot)$ , and the bandwidth  $h$ .

**Assumption 3.2.** (a) The random vectors  $(\varepsilon_1, Z_1)', \dots, (\varepsilon_n, Z_n)'$  are independent copies of the random vector  $(\varepsilon, Z)' \in \mathbb{R}^{1+q}$ . The conditional  $\tau$ th quantile of  $\varepsilon$  given  $Z = (W, X)'$  is equal to zero.

(b) The variable  $W$  admits an absolutely continuous density with the respect of the Lebesgue measure on the real line.

(c) The conditional density  $f_\varepsilon(\cdot | z)$  of  $\varepsilon$  given  $Z = z$  is uniformly bounded. There exists  $a > 0$  such that  $f_\varepsilon(\cdot | z)$  is differentiable on  $(-a, a)$  for any  $z$  with  $|f'_\varepsilon(0 | z)| \leq C\infty$ . Moreover, the derivatives  $f'_\varepsilon(\cdot | z)$  satisfy a uniform Hölder continuity condition, that is there exist positive constants  $C_2$  and  $c$  independent of  $z$  such that  $\forall |u_1|, |u_2| \leq a$ ,  $|f'_\varepsilon(u_1 | z) - f'_\varepsilon(u_2 | z)| \leq C_2 |u_1 - u_2|^c$ .

**Assumption 3.3.** (a) The parameter space  $B$  is a compact convex subset of  $\mathbb{R}^p$ .  $\beta_0$  is the unique solution of  $\min_B \mathbb{E}[\rho_\tau(Y - g(Z, \beta))]$  and is an interior point of  $B$ .

(b) The matrix

$$\mathbb{E} \left[ f_\varepsilon(0 | Z) \frac{\partial}{\partial \beta} g(Z; \beta_0) \frac{\partial}{\partial \beta'} g'(Z; \beta_0) \right]$$

is finite and nonsingular.

(c) There exists functions  $A(\cdot)$ ,  $B(\cdot)$ , and  $D(\cdot)$ , with  $\mathbb{E}[A^4(Z)]$ ,  $\mathbb{E}[B^2(Z)] < \infty$ , and  $\mathbb{E}[D^4(Z)]$ , such that

$$\left\| \frac{\partial}{\partial \beta} g(z; \beta) \right\| \leq A(z), \quad \left\| \frac{\partial}{\partial \beta} g(z; \beta) \frac{\partial}{\partial \beta'} g'(z; \beta) \right\| \leq D(z) \quad \text{for any } \beta,$$

$$\left\| \frac{\partial}{\partial \beta} g(z; \beta_1) - \frac{\partial}{\partial \beta} g(z; \beta_2) \right\| \leq B(z) \|\beta_1 - \beta_2\| \quad \text{for any } z, \beta_1, \beta_2.$$

(d) The class of functions  $\{g(Z; \beta) : \beta \in B\}$  is a Vapnik-Červonenkis (VC) class.

**Assumption 3.4.** (a) The function  $K(\cdot)$  is a bounded symmetric univariate density of bounded variation with positive Fourier transform.

(b) The function  $\psi(\cdot)$  is a bounded symmetric multivariate function with positive Fourier transform.

(c)  $h \rightarrow 0$  and  $n^\alpha h^2 \rightarrow \infty$  for some  $\alpha \in (0, 1)$  as  $n \rightarrow \infty$ .

Our assumptions combine standard assumptions for parametric quantile regression estimation and specific ones for our lack-of-fit test. Among the latter, the conditions on the error term  $\varepsilon$  impose neither independence of  $\varepsilon$  and  $Z$ , nor a specific form of dependence such as  $\varepsilon = s(Z)e$  with  $e$  independent of  $Z$  as in He and Zhu (2003). Assumption 3.3(d) is a mild technical condition that guarantees suitable uniform rates of convergence for some  $U$ -processes appearing in the proofs. This condition is satisfied for many parametric models, for instance when  $g(Z, \beta) = q(Z'\beta)$  with  $q : \mathbb{R} \rightarrow \mathbb{R}$  monotone or of bounded variation, see e.g. van der Vaart and Wellner (1996, Section 2.6). Also, if there is  $\beta \in B$  such that  $g(Z, \beta)$  is squared integrable, then Assumption 3.3(d) follows from 3.3(c). Assumptions on  $K(\cdot)$  allows for the use of a triangular, normal, logistic, Student (including Cauchy), or Laplace densities. For  $\psi(\cdot)$ , one can choose e.g.  $\psi(x) = \exp(-\|x\|^2)$ , or any multivariate extension of the aforementioned densities. Restrictions on the bandwidth are compatible with optimal choices for regression estimation, see e.g. Härdle and Marron (1985), and for regression checks, see Guerre and Lavergne (2002) and Horowitz and Spokoiny (2002). The following theorem states the asymptotic validity of our test.

**Theorem 3.5.** *Under the Assumptions 3.2 to 3.4, the test based on  $T_n$  has asymptotic level  $\alpha$  under  $H_0$ .*

### 3.2.3 Behavior under Local Alternatives

We now investigate the behavior of our test when  $H_0$  does not hold, and specifically we consider a sequence of local alternatives of the form

$$H_{1n} : Y = g(Z; \beta_0) + r_n \delta(Z) + \varepsilon, \quad F(g(Z; \beta_0) | Z) = \tau, \quad (3.5)$$

where  $r_n$ ,  $n \geq 1$ , is a sequence of real numbers tending to zero and  $\delta(Z)$  is a real-valued function satisfying

$$\mathbb{E} \left[ f_\varepsilon(0 | Z) \delta(Z) \frac{\partial}{\partial \beta} g(Z; \beta_0) \right] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(Z)] < \infty. \quad (3.6)$$

This condition ensures that our sequence of models (3.5) does not belong to the null hypothesis  $H_0$ . We do not impose any smoothness restriction on the function  $\delta(\cdot)$  as is frequent in this kind of analysis, see e.g. Zheng (1998). As shown in Lemma 3.8 in the Proofs section,  $\hat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2} + r_n^2)$  under  $H_{1n}$ . Our next result states that these local alternatives can be detected whenever  $r_n^2 n h^{1/2} \rightarrow \infty$ . Hence our test *does not suffer from the curse of dimensionality* against local alternatives, since its power is unaffected by the number of regressors.

**Theorem 3.6.** *Under Assumptions 3.2 to 3.4, the test based on  $T_n$  is consistent against the sequence of alternatives  $H_{1n}$  with  $\delta(Z)$  satisfying (3.6) if  $r_n^2 n h^{1/2} \rightarrow \infty$ .*

### 3.2.4 Bootstrap Critical Values

The asymptotic approximation of the behavior of  $T_n$  may not be satisfactory in small samples as is customary in smoothing-based lack-of-fit tests. This motivates the use of bootstrapping for obtaining critical values. The distribution of  $T_n$  depends weakly on the distribution of the error term  $\varepsilon$ , because  $\mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau$  under  $H_0$  is a Bernoulli random variable irrespective of the particular distribution of  $\varepsilon$ . The same phenomenon is noted by Horowitz and Spokoiny (2002) for their test statistic. Their proposal is thus to naively (or nonparametrically) bootstrap from the empirical distribution of the residuals. This is a valid bootstrap procedure when errors are identically distributed, and it remains asymptotically valid for non identically distributed errors. A first possibility is thus to adopt naive residual bootstrap for our test. Alternatively, He and Zhu (2003) note that one could use any continuous distribution with the  $\tau$ -th quantile equal to 0. This constitutes a second possibility. While asymptotically valid, these two methods do not account for potential heteroscedastic errors. Thus a third possibility is the wild bootstrap method for quantile regression introduced by Feng et al. (2011). The wild bootstrap procedure for our test works as follows.

1. Let  $\hat{\varepsilon}_i = Y_i - g(Z_i; \hat{\beta})$ ,  $1 \leq i \leq n$ , and  $w_1, \dots, w_n$  be bootstrap weights generated independently from a two-point mass distribution with probabilities  $1 - \tau$  and  $\tau$  at  $2(1 - \tau)$  and  $-2\tau$ . Compute  $\varepsilon_i^* = w_i |\hat{\varepsilon}_i|$  and  $Y_i^* = g(Z_i; \hat{\beta}) + \varepsilon_i^*$  for each  $i = 1, \dots, n$ .
2. Use the bootstrap data set  $\{Y_i^*, Z_i : i = 1, \dots, n\}$  to compute the estimator  $\hat{\beta}^*$ , the new  $U_i^*(\hat{\beta}^*) = \mathbb{I}\{Y_i^* \leq g(Z_i; \hat{\beta}^*)\} - \tau$ , and the new test statistic  $T_n^*$ .
3. Repeat Steps 1 et 2 many times, and estimate the  $\alpha$ -level critical value  $z_\alpha^*$  by the  $(1 - \alpha)$ -th quantile of the empirical distribution of  $T_n^*$ .

The bootstrap test then rejects  $H_0$  if  $T_n \geq z_\alpha^*$ . Alternatively, one could resample residuals in Step 1 by naive bootstrap, or obtain  $\varepsilon_i^*$  by random draws from e.g. a uniform law on the interval  $[-\tau, 1 - \tau]$ . The following theorem yields the asymptotic validity of the bootstrap test.

**Theorem 3.7.** *Under the conditions of Theorem 3.5,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(T_n^* \leq t \mid Y_1, Z_1, \dots, Y_n, Z_n) - \Phi(t)| \xrightarrow{p} 0,$$

where  $\Phi(\cdot)$  is the standard normal distribution function.



### 3.3 Numerical Evidence

#### 3.3.1 Small Sample Performances

We investigated the performances of our procedure for testing lack-of-fit of a linear median regression for two setups considered by He and Zhu (2003), namely

$$Y = 1 + W + X + \delta (W^2 + WX + X^2) + \varepsilon, \quad (3.7)$$

$$Y = \delta \log(1 + W^2 + X^2) + \varepsilon, \quad (3.8)$$

where  $W$  follows a standard normal, and  $X$  independently follows a binomial of size 5 and probability of success 0.5. For the error term, we considered the three distributions  $\mathcal{N}(0, 1)$ ,  $\log \mathcal{N}(0, 1) - 1$  and  $\mathcal{N}(0, (1 + W^2)/2)$ .

For implementation, we chose  $\psi(\cdot)$  as the standard normal density and  $K(\cdot)$  as triangle density with variance one. We set  $\delta = 0$  in Model (3.7) to evaluate the comparative performances of the three possible bootstrapping procedures. Figure 3.1 reports our results based on 5000 replications for a sample size of  $n = 100$  at nominal level 10%, when the bandwidth is  $h = cn^{-1/5}$  with  $c$  varying. The three bootstrap methods yield accurate levels for any bandwidth choice when errors are identically distributed, while the use of asymptotic critical values yield large underrejection. In the heteroscedastic case, however, only the wild bootstrap yield an empirical level close to 10%, while the use of naive or uniform bootstrap results in a severely oversized test.

Next, we investigated the power of our test for Models (3.7) and (3.8) with either standard gaussian or heteroscedastic gaussian errors. We compared our test to the one proposed by He and Zhu (2003, hereafter HZ), based on

$$\max_{\|\mathbf{a}\|=1} n^{-1} \sum_{i=1}^n (\mathbf{a}' \mathbf{R}_n(X_i))^2$$

where  $\mathbf{R}_n(\mathbf{t}) = n^{-1/2} \sum_{j=1}^n (\tau - \mathbb{I}[U_j(\hat{\beta}) < 0]) Z_j \mathbb{I}(Z_j \leq \mathbf{t})$ .

We also computed the statistic proposed by Zheng (1998), which in our setup writes

$$\frac{h}{\hat{\sigma}(n-1)} \sum_{j \neq i} U_i(\hat{\beta}) U_j(\hat{\beta}) h^{-q} \tilde{K} \left( \frac{W_i - W_j}{h}, \frac{X_i - X_j}{h} \right)$$

where  $\hat{\sigma}^2 = \frac{2\tau^2(1-\tau)^2}{n(n-1)} \sum_{j \neq i} h^{-q} \tilde{K}^2 \left( \frac{W_i - W_j}{h}, \frac{X_i - X_j}{h} \right)$ , and  $\tilde{K}$  is a triangle kernel applied to the norm of its argument. We apply the wild bootstrap procedure to compute the critical values of all tests. Figure 3.2 reports power curves of the different tests as a function of  $\delta$  based on 2500 replications. For the linear Model (3.7), all tests perform almost similarly. Our test is a bit more powerful, especially for a larger bandwidth, which was expected given our theoretical analysis. For the nonlinear Model (3.8), the power advantage of our test is more pronounced. Its power can be as large as twice the power of the test by He and Zhu (2003).

### 3.3.2 Empirical Illustration

We studied some parametric quantile models for children birthweight using data analyzed by Abrevaya (2001) and Koenker and Hallock (2001), who gave a detailed data description. We focused on median regression and the 10th percentile quantile regression. Models are estimated and tested on a subsample of 1168 smoking college graduate mothers. We first analyzed the simple model considered by He and Zhu (2003), which is linear in weight gain during pregnancy (WTGAIN), average number of cigarettes per day (CIGAR), and age (AGE). When implementing our test, we chose age as the  $W$  variable, and we standardize all explanatory variables. Other details are identical to what was done in simulations. For both quantiles, HZ test does not reject this specification. Our test does not reject the linear median regression at 10% level, but detects misspecification for the lower decile regression when  $c = 2$ .

Since the more detailed analysis of Abrevaya (2001) and Koenker and Hallock (2001) suggests that birthweight is quadratic in age, we then considered this variation. None of the tests detects a misspecified model. Finally, we considered a more complete model similar to Abrevaya (2001), where we added the explanatory binary variables BOY (1 if child is male), BLACK (1 if mother is black), MARRIED (1 if married), and NOVISIT (1 if no prenatal visit during the pregnancy). HZ test does not reject the model at either quantiles. Our test however indicates a misspecified median regression model at 10% level, while it does not reject the model for the lower decile. Our limited empirical exercise suggests that our new test, beside existing procedures such as the test by He and Zhu (2003), is a valuable addition to the practitioner toolbox.

### 3.4 Proofs

We first recall some definitions. For the definition of a VC-class, we refer to Section 2.6.2 of van der Vaart and Wellner (1996). Next, let  $\mathcal{G}$  be a class of real-valued functions on a set  $\mathcal{S}$ . We call  $\mathcal{G}$  an *Euclidean*( $c, d$ ) family of functions, or simply *Euclidean*, for the envelope  $G$  if there exists positive constants  $c$  and  $d$  with the following properties: if  $0 < \epsilon \leq 1$  and  $\lambda$  is a measure for which  $\int G^2 d\lambda < \infty$ , then there are functions  $g_1, \dots, g_j$  in  $\mathcal{G}$  such that (i)  $j \leq c\epsilon^{-d}$ ; and (ii) for each  $g$  in  $\mathcal{G}$  there is an  $g_i$  with  $\int |g - g_i|^2 d\lambda \leq \epsilon^2 \int G^2 d\lambda$ . The constants  $c$  and  $d$  must not depend on  $\lambda$ . See e.g. Nolan and Pollard (1987) or Sherman (1994). Recall that if  $\mathcal{F}$  is a VC-class of functions then the class  $\{\mathbb{I}\{f \geq 0\} : f \in \mathcal{F}\}$  is Euclidean for the envelope  $F \equiv 1$ , see van der Vaart and Wellner (1996) Lemma 2.6.18(iii) and Theorem 2.6.7 or Pakes and Pollard (1989). Below, we shall use this property with the VC-classes of functions of  $\{\varepsilon + g(Z, \beta_0) - g(Z, \beta) : \beta \in B\}$  and  $\{\varepsilon + g(Z, \beta_0) + r_n \delta(Z) - g(Z, \beta) : \beta \in B\}$ .

In the following,  $F_\varepsilon(\cdot | x)$  is the conditional distribution function of  $\varepsilon$  given  $Z = x$ ; that means  $F_\varepsilon(0 | \cdot) \equiv \tau$ . Below  $C, C_1, C_2, \dots$  denote constants, not necessarily the same as before and possibly changing from line to line.

#### 3.4.1 Proof of Theorem 3.5

*Proof.* First, we prove that if  $\mathcal{H}_0$  holds

$$n\sqrt{h} \{W_n(\hat{\beta}) - W_n(\beta_0)\} = o_{\mathbb{P}}(1). \quad (3.1)$$

Let us introduce some simplifying notation:

$$G_i(\beta, \beta_0) = g(Z_i; \beta) - g(Z_i; \beta_0), \quad \psi_{ij} = \psi(X_i - X_j), \quad K_{h,ij} = K_h(W_i - W_j). \quad (3.2)$$

Under  $\mathcal{H}_0$

$$\begin{aligned} W_n(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau] [\mathbb{I}\{Y_j \leq g(Z_j; \beta)\} - \tau] K_{h,ij} \psi_{ij} \\ &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(0 | Z_i)] \\ &\quad \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0)\} - F_\varepsilon(0 | Z_j)] K_{h,ij} \psi_{ij}. \end{aligned}$$

By a Taylor expansion, decompose

$$F_\varepsilon(0 | Z_i) = F_\varepsilon(G_i(\beta, \beta_0) | Z_i) - f_\varepsilon(0 | Z_i) g'(Z_i; \beta_0) (\beta - \beta_0) + O_{\mathbb{P}}(\|\beta - \beta_0\|^2).$$

We can write  $W_n(\beta) - W_n(\beta_0) = \{W_{1n}^0(\beta) - W_{1n}^0(\beta_0)\} + 2W_{2n}^0(\beta) + W_{3n}^0(\beta) + R_n^0$  where

$$W_{1n}^0(\beta) = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) | Z_i)] \\ \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0)\} - F_\varepsilon(G_j(\beta, \beta_0) | Z_j)] K_{h,ij} \psi_{ij}$$

$W_{2n}^0(\beta) = (\beta - \beta_0)' \widetilde{W}_{2n}^0(\beta)$  with

$$\widetilde{W}_{2n}^0(\beta) = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) | Z_i)] \\ \times f_\varepsilon(0 | Z_j) \dot{g}(Z_j; \beta_0) K_{h,ij} \psi_{ij},$$

$W_{3n}^0(\beta) = (\beta - \beta_0)' \widetilde{W}_{3n}^0(\beta - \beta_0)$  with

$$\widetilde{W}_{3n}^0 = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon(0 | Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_j; \beta_0) f_\varepsilon(0 | Z_j) K_{h,ij} \psi_{ij} = O_{\mathbb{P}}(1).$$

The rate of  $\widetilde{W}_{3n}^0$  follows simply by computing its mean and variance. By Assumption 3.2(c) and Assumption 3.3(c) it is easy to check that  $|R_n^0| \leq \|\beta - \beta_0\|^2 O_{\mathbb{P}}(1)$ . For deriving the order of  $\widetilde{W}_{2n}^0$ , apply Hoeffding decomposition and write  $h\widetilde{W}_{2n}^0(\beta) = V_n^2(\beta) + V_n^1(\beta)$  with  $V_n^1, V_n^2$  degenerate  $U$ -processes of order 1 and 2, respectively. In view of Assumptions 3.3(d) and 3.4(a), apply Corollary 4 of Sherman (1994) and deduce that  $V_n^2(\beta) = O_{\mathbb{P}}(n^{-1})$  uniformly in  $\beta$  (and  $h$ ). Next, if  $\dot{g}^{(l)}$  denotes the  $l$ th component of the vector of first-order derivatives  $\dot{g}$ ,  $1 \leq l \leq p$ , and

$$\pi^{(l)}(Z_i) = \mathbb{E} [f_\varepsilon(0 | Z_j) \dot{g}^{(l)}(Z_j; \beta_0) h^{-3/4} K_{h,ij} \psi_{ij} | Z_i]$$

we can rewrite the  $l$ th component of the vector  $V_n^1(\beta)$  as

$$\frac{h^{3/4}}{n} \sum_{i=1}^n [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) | Z_i)] \pi^{(l)}(Z_i).$$

By Hölder inequality, Assumption 3.2(c), Assumption 3.3(c) and a change of variables,

$$\begin{aligned} |\pi^{(l)}(X_i)| &\leq \mathbb{E} [f_\varepsilon(0 | Z_j) |\dot{g}^{(l)}(Z_j; \beta_0)| h^{-3/4} K_{h,ij} |\psi_{ij}| | Z_i] \\ &\leq C_1 \mathbb{E}^{1/4} [A^4(Z_j)] \mathbb{E}^{3/4} [h^{-1} K_{h,ij}^{4/3} | Z_i] \\ &\leq C_2, \end{aligned}$$

for any  $1 \leq l \leq p$ . Now, by Corollary 4 of Sherman (1994),  $h^{-3/4} V_n^1(\beta) = O_{\mathbb{P}}(n^{-1/2})$  uniformly in  $\beta$ . Deduce that

$$\sup_{\beta} |W_{2n}^0(\beta)| \leq \|\beta - \beta_0\| O_{\mathbb{P}}(h^{-1} n^{-1} + h^{-1/4} n^{-1/2}).$$

Finally, by Lemma 1 of Zheng (1998), for any  $\alpha \in (0, 1)$

$$\sup_{\beta} |W_{1n}^0(\beta) - W_{1n}^0(\beta_0)| = O_{\mathbb{P}}\left(h^{-1}n^{-1-\alpha/4}\right)$$

uniformly over  $O_{\mathbb{P}}\left(n^{-1/2}\right)$  neighborhoods of  $\beta_0$ . Gathering the results and using Lemma 3.8 with  $\delta(\cdot) \equiv 0$  we obtain (3.1). It remains to check that  $nh^{1/2}W_n(\beta_0)/v_n$  converges in law to a standard normal distribution. This result easily follows as a particular case of Lemma 3.9 below.  $\square$

### 3.4.2 Proof of Theorem 3.6

First, we derive the behavior of  $\hat{\beta}$ , the estimator of  $\beta_0$  under the sequence of local alternatives  $H_{1n}$ .

**Lemma 3.8.** *Suppose that Assumptions 3.2, 3.3 hold, let  $\delta(\cdot)$  be a function such that Condition (3.6) holds, and let  $r_n$ ,  $n \geq 1$  be a sequence of real numbers such that  $r_n \rightarrow 0$ . If  $\hat{\beta} = \arg \min_{\beta \in B} \Gamma_n(\beta)$  with  $\Gamma_n(\beta) = \sum_{i=1}^n \rho_{\tau}(Y_i - g(Z_i; \beta))$ , then under  $\mathcal{H}_0$ ,  $\hat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$  and under  $H_{1n}$  defined in (3.5),  $\hat{\beta} - \beta_n = O_{\mathbb{P}}(n^{-1/2})$  where*

$$\beta_n = \beta_0 - r_n^2 [\mathbb{E}[f_{\varepsilon}(0 | Z) \dot{g}(Z; \beta_0) \dot{g}'(Z; \beta_0)]]^{-1} \mathbb{E}[f'_{\varepsilon}(0 | Z) \delta^2(Z) \dot{g}(Z; \beta_0)].$$

*Proof.* It is easy to check that

$$|\rho_{\tau}(a - b) - \rho_{\tau}(a)| \leq |b| \max(\tau, 1 - \tau) \leq |b|. \quad (3.3)$$

Combine this with the Mean Value Theorem and Assumption 3.3(c) to check the conditions of Lemma 2.13 of Pakes and Pollard (1989) and to derive the Euclidean property for an integrable envelope for the family of functions

$$\{(y, z) \mapsto \rho_{\tau}(y - g(z; \beta)) : \beta \in B\}.$$

Next, we study the consistency of  $\hat{\beta}$  under  $H_0$ . By the uniform law of large numbers,  $\sup_{\beta} |n^{-1} \Gamma_n(\beta) - \mathbb{E}[\rho_{\tau}(Y - g(Z; \beta))]| \rightarrow 0$ , in probability (use for instance Lemma 2.8 of Pakes and Pollard 1989). This uniform convergence, the identification condition in Assumption 3.3(a), the continuity of  $g(z; \cdot)$  for any  $z$ , and usual arguments used for proving consistency of argmax estimators, allow to deduce  $\hat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$ . To obtain the consistency under the local alternatives approaching  $H_0$ , it suffices to prove  $\sup_{\beta \in B} |\Delta_n(\beta)| \rightarrow 0$  in probability, where

$$\Delta_n(\beta) = \frac{1}{n} \sum_{i=1}^n \{\rho_{\tau}(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) - \rho_{\tau}(l(\varepsilon_i, Z_i; \beta))\}$$

and  $l(u, z; \beta) = u + g(z; \beta_0) - g(z; \beta)$ . By inequality (3.3),

$$|\Delta_n(\beta)| \leq \frac{|r_n|}{n} \sum_{i=1}^n |\delta(Z_i)|.$$

Consequently,  $\Delta_n(\beta) = o_{\mathbb{P}}(1)$  uniformly over  $\beta \in B$ , and thus the consistency follows.

Define  $\psi_\tau(e) = \tau - \mathbb{I}(e < 0)$  as the derivative of  $\rho_\tau$ . To obtain the rate of convergence of  $\widehat{\beta}$  under  $\mathcal{H}_{1n}$  (in particular under  $H_0$  by taking  $r_n \equiv 0$ ) consider the empirical process

$$\begin{aligned} \nu_n(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\psi_\tau(Y_i - g(Z_i; \beta)) - \mathbb{E}[\psi_\tau(Y_i - g(Z_i; \beta)) \mid Z_i]\} \dot{g}(Z_i; \beta) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\psi_\tau(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) \\ &\quad - \mathbb{E}[\psi_\tau(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) \mid Z_i]\} \dot{g}(Z_i; \beta) \end{aligned}$$

indexed by  $\beta$ . First, let us notice that

$$\nu_n(\beta) - \nu_n(\beta_0) = o_{\mathbb{P}}(1) \tag{3.4}$$

uniformly over  $o_{\mathbb{P}}(1)$  neighborhoods of  $\beta_0$ , as a consequence of Corollary 8 of Sherman (1994). Indeed, by Lemma 2.13 of Pakes and Pollard (1989), the class of functions  $\{\dot{g}(\cdot; \beta) : \beta \in B\}$  is Euclidean for a squared integrable envelope. Next, by the VC-class property of the regression functions  $\{g(\cdot; \beta), \beta \in B\}$ , the class of functions  $\{(u, z) \mapsto \psi_\tau(l(u, z; \beta) + r_n \delta(z)) : \beta \in B\}$  is *Euclidean*( $c, d$ ) for a constant envelope. See Lemma 2.12 of Pakes and Pollard (1989). Moreover, the constants  $c$  and  $d$  can be taken independent of  $n$ , see, for instance, the proof of Lemma 2.6.18(v) of van der Vaart and Wellner (1996). Finally, by repeated applications of the Mean Value Theorem and Assumptions 3.2(c) and 3.3(c), for any  $z, \beta_1, \beta_2$  we have

$$\begin{aligned} &|\mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta_1) + r_n \delta(z))] - \mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta_2) + r_n \delta(z))]| \tag{3.5} \\ &\leq |F_\varepsilon(g(z; \beta_1) - g(z; \beta_0) - r_n \delta(z) \mid z) - F_\varepsilon(g(z; \beta_2) - g(z; \beta_0) - r_n \delta(z) \mid z)| \\ &\leq f_\varepsilon(v_n \mid z) |g(z; \beta_1) - g(z; \beta_2)| \\ &\leq CA(z) \|\beta_1 - \beta_2\| \end{aligned}$$

for some  $v_n$  between  $g(z; \beta_1) - g(z; \beta_0) - r_n \delta(z)$  and  $g(z; \beta_2) - g(z; \beta_0) - r_n \delta(z)$ . By Pakes and Pollard (1989, Lemma 2.13), the class of functions

$$\{z \mapsto \mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta) + r_n \delta(z))] : \beta \in B\}$$

is *Euclidean*( $c, d$ ) for an envelope with a finite fourth moment, with  $c$  and  $d$  independent of  $n$ . Deduce that the empirical process  $\nu_n(\beta)$ ,  $\beta \in B$ , is indexed by a class of functions that is Euclidean for a squared integrable envelope. Finally, condition (ii) of Corollary 8 of Sherman (1994), can be checked from inequalities like in (3.5) and conditions on  $|\dot{g}(z; \beta) - \dot{g}(z; \beta_0)|$ .

On the other hand, because  $\widehat{\beta}$  minimizes  $\Gamma_n(\beta)$  defined in (3.3) over  $\beta$ , the directional derivative of  $\Gamma_n(\beta)$  at  $\widehat{\beta}$  along any direction  $\gamma$  (with  $\|\gamma\| = 1$ ) is nonnegative. That

is

$$\begin{aligned}
0 &\leq \lim_{t \rightarrow 0} t^{-1} \left[ \Gamma_n(\hat{\beta} + t\gamma) - \Gamma_n(\hat{\beta}) \right] & (3.6) \\
&= - \sum_{\{Y_i \neq g(Z_i; \hat{\beta})\}} \psi_\tau \left( Y_i - g(Z_i; \hat{\beta}) \right) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
&\quad + \lim_{t \rightarrow 0} \sum_{\{Y_i = g(Z_i; \hat{\beta})\}} t^{-1} \rho_\tau \left( g(Z_i; \hat{\beta}) - g(Z_i; \hat{\beta} + t\gamma) \right) \\
&= - \sum_{\{Y_i \neq g(Z_i; \hat{\beta})\}} \psi_\tau \left( Y_i - g(Z_i; \hat{\beta}) \right) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
&\quad - \sum_{\{Y_i = g(Z_i; \hat{\beta})\}} \psi_\tau \left( -\gamma' \dot{g}(Z_i; \hat{\beta}) \right) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
&= -D_{1n}(\hat{\beta}) - D_{2n}(\hat{\beta}).
\end{aligned}$$

By Assumption 3.3,  $|D_{2n}(\hat{\beta})|$  is bounded by  $\sum_{\{Y_i = g(Z_i; \hat{\beta})\}} A(Z_i)$ . As, for any  $x$ , the error term  $u$  has a continuous law given  $Z = z$ , the number of observations with  $Y_i = g(Z_i; \hat{\beta})$  is bounded in probability as the sample size tends to infinity. On the other hand, the moment condition on  $A(\cdot)$  implies that  $\max_{1 \leq i \leq n} A(Z_i) = o_{\mathbb{P}}(n^{1/2})$ . As  $\gamma$  is an arbitrary direction, it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau \left( Y_i - g(Z_i; \hat{\beta}) \right) \dot{g}(Z_i; \hat{\beta}) = o_{\mathbb{P}}(1). \quad (3.7)$$

Finally, since  $\hat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$  and  $\tau = F_\varepsilon(0 \mid Z_i)$ , deduce that

$$\begin{aligned}
\nu_n(\beta_0) &= \nu_n(\hat{\beta}) + o_{\mathbb{P}}(1) && \text{[by (3.4)]} \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[ \psi_\tau \left( Y_i - g(Z_i; \hat{\beta}) \right) \mid Z_i \right] \dot{g}(Z_i; \hat{\beta}) + o_{\mathbb{P}}(1) && \text{[by (3.7)]} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ F_\varepsilon \left( g(Z_i; \hat{\beta}) - g(Z_i; \beta_0) - r_n \delta(Z_i) \mid Z_i \right) - \tau \right] \dot{g}(Z_i; \hat{\beta}) + o_{\mathbb{P}}(1) \\
&= \left\{ \frac{1}{n} \sum_{i=1}^n f_\varepsilon(0 \mid Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_i; \beta_0) \right\} \sqrt{n} (\hat{\beta} - \beta_0) \\
&\quad - r_n \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n f_\varepsilon(0 \mid Z_i) \delta(Z_i) \dot{g}(Z_i; \beta_0) \right\} \\
&\quad + r_n^2 \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n f_\varepsilon'(0 \mid Z_i) \delta^2(Z_i) \dot{g}(Z_i; \beta_0) \right\} \\
&\quad + o_{\mathbb{P}}(\sqrt{n} \|\hat{\beta} - \beta_0\|) + o_{\mathbb{P}}(r_n^2 \sqrt{n}),
\end{aligned}$$

where the last equality is based on a local expansions of  $F_\varepsilon(\cdot \mid z)$  and  $g(z; \cdot)$ . By the law of large numbers, the central limit theorem and the fact that  $\nu_n(\beta_0) = O_{\mathbb{P}}(1)$

and the random vector  $f_u(0 | Z)\delta(Z)\dot{g}(Z; \beta_0)$  has zero mean, we obtain

$$\mathbb{E}[f_\varepsilon(0 | Z)\dot{g}(Z; \beta_0)\dot{g}'(Z; \beta_0)]\sqrt{n}(\hat{\beta} - \beta_0) + r_n^2\sqrt{n}\mathbb{E}[f'_\varepsilon(0 | Z)\delta^2(Z)\dot{g}(Z; \beta_0)] = O_{\mathbb{P}}(1)$$

from which the result follows.  $\square$

Lemma 3.8 shows in particular that under  $\mathcal{H}_{1n}$ ,  $\hat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2} + r_n^2)$ . To our best knowledge, this result on the behavior of  $\hat{\beta}$  under the local alternatives is new. He and Zhu (2003) only considered the case  $r_n = n^{-1/2}$  while Zheng (1998) assumed  $\hat{\beta} - \beta^* = O_{\mathbb{P}}(n^{-1/2})$  under  $\mathcal{H}_{1n}$ , for some fixed  $\beta^*$ . Our Lemma 3.8 indicates that such  $\sqrt{n}$ -convergence assumptions on the local alternatives may be too restrictive. Below, we improve the point (C) in the Theorem of Zheng (1998) also because we can take into account the rates of convergence of  $\hat{\beta}$  under the alternatives slower than  $O_{\mathbb{P}}(n^{-1/2})$ .

In the case of a fixed deviation from the null hypothesis, that is  $r_n \equiv 1$ , the tools used for proving Theorem 3.6 could be easily adapted to show the  $\sqrt{n}$ -convergence of  $\hat{\beta}$  to  $\beta^*$  that minimizes the map  $\beta \mapsto \mathbb{E}[\rho_\tau(Y - g(Z, \beta))] = \mathbb{E}[\rho_\tau(g(Z, \beta_0) + \delta(Z) + \varepsilon - g(Z, \beta))]$ . The consistency of the test is then a consequence of the fact that  $nh^{1/2}I_n(\beta^*)$  tends to infinity.

Let  $\delta_i = \delta(Z_i)$  and let  $G_i(\beta, \beta_0)$  and  $K_{h,ij}$  be defined as in equation (3.2). Under  $\mathcal{H}_{1n}$

$$\begin{aligned} W_n(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau] [\mathbb{I}\{Y_j \leq g(Z_j; \beta)\} - \tau] K_{h,ij} \psi_{ij} \\ &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon(0 | Z_i)] \\ &\quad \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n \delta_j\} - F_\varepsilon(0 | Z_j)] K_{h,ij} \psi_{ij}. \end{aligned}$$

Let us decompose

$$\begin{aligned} F_\varepsilon(0 | Z_i) &= F_\varepsilon(G_i(\beta, \beta_0) - r_n \delta_i | Z_i) - f_\varepsilon(0 | Z_i) \{g'(Z_i; \beta_0)(\beta - \beta_0) - r_n \delta_i\} \\ &\quad - 2^{-1} r_n^2 f'_\varepsilon(0 | Z_i) \delta_i^2 + O_{\mathbb{P}}(\|\beta - \beta_0\|^2 + r_n \|\beta - \beta_0\|) + o_{\mathbb{P}}(r_n^2). \end{aligned}$$

We can write

$$W_n(\beta) = W_{1n}(\beta) + 2[W_{2n}(\beta) + W_{3n}(\beta) + W_{4n}(\beta)] + W_{5n}(\beta) + 2W_{6n}(\beta) + W_{7n} + R_n$$

where

$$\begin{aligned} W_{1n}(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n \delta_i | Z_i)] \\ &\quad \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n \delta_j\} - F_\varepsilon(G_j(\beta, \beta_0) - r_n \delta_j | Z_j)] K_{h,ij} \psi_{ij} \end{aligned}$$



$W_{2n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{2n}(\beta)$  with

$$\begin{aligned} \widetilde{W}_{2n}(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n \delta_i | Z_i)] \\ &\quad \times f_\varepsilon(0 | Z_j) \dot{g}(Z_j; \beta_0) K_{h,ij} \psi_{ij}, \end{aligned}$$

$$\begin{aligned} W_{3n}(\beta) &= \frac{r_n h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n \delta_i | Z_i)] \\ &\quad \times f_\varepsilon(0 | Z_j) \delta_j K_{h,ij} \psi_{ij}, \end{aligned}$$

$$\begin{aligned} W_{4n}(\beta) &= \frac{r_n^2 h^{-1}}{2n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n \delta_i | Z_i)] \\ &\quad \times f'_\varepsilon(0 | Z_j) \delta_j^2 K_{h,ij} \psi_{ij}, \end{aligned}$$

$W_{5n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{5n}(\beta - \beta_0)$  with

$$\widetilde{W}_{5n} = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon(0 | Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_j; \beta_0) f_\varepsilon(0 | Z_j) K_{h,ij} \psi_{ij} = O_{\mathbb{P}}(1),$$

$W_{6n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{6n}$  with

$$\widetilde{W}_{6n} = \frac{r_n h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon(0 | Z_i) \delta_i f_\varepsilon(0 | Z_j) \dot{g}(X_j; \beta_0) K_{h,ij} \psi_{ij} = O_{\mathbb{P}}(r_n),$$

$$W_{7n} = \frac{r_n^2 h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon(0 | Z_i) \delta(X_i) f_\varepsilon(0 | Z_j) \delta(Z_j) K_{h,ij} \psi_{ij} = C_1 r_n^2 + o_{\mathbb{P}}(r_n^2)$$

with  $C_1 > 0$  and  $R_n$  a reminder term that is negligible because of the properties of  $f'_\varepsilon$  and  $\dot{g}$ . Note that the  $U$ -statistics  $\widetilde{W}_{5n}$ ,  $\widetilde{W}_{6n}$  and  $W_{7n}$  depend only on the  $X_i$ . Their orders are obtained from elementary calculations of mean and variance.

Next, we can write  $W_{1n}(\beta) = \{W_{1n}(\beta) - W_{1n}(\beta_0)\} + W_{1n}(\beta_0)$ . As  $W_{1n}(\beta_0)$  is centered, its order in probability is given by the variance. We have

$$\begin{aligned} \text{Var}(W_{1n}(\beta_0) | Z_1, \dots, Z_n) &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j} F_\varepsilon(-r_n \delta_i | Z_i) \\ &\quad \times [1 - F_\varepsilon(-r_n \delta_i | Z_i)] F_\varepsilon(-r_n \delta_j | Z_j) \\ &\quad \times [1 - F_\varepsilon(-r_n \delta_j | Z_j)] h^{-2} K_{h,ij}^2 \psi_{ij}(\mu) \\ &\leq \frac{h^{-1}}{16n(n-1)} \left[ \frac{1}{n(n-1)} \sum_{i \neq j} h^{-1} K_{h,ij}^2 \psi_{ij} \right] \end{aligned}$$

The expectation of the last  $U$ -statistic in the display converges to a constant while the variance tends to zero. As  $W_{1n}(\beta_0)$  is of zero conditional mean given the  $Z_i$ , deduce that the variance of  $W_{1n}(\beta_0)$  is bounded by  $Cn^{-2}h^{-1}$ . By Chebyshev's inequality,  $W_{1n}(\beta_0) = o_{\mathbb{P}}(r_n^2)$ , provided that  $r_n^2nh^{1/2} \rightarrow \infty$ . Next, let

$$\begin{aligned} H_{1n}(Z_i, Z_j, \beta) &= [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n\delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n\delta_i | Z_i)] \\ &\quad \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n\delta_j\} - F_\varepsilon(G_j(\beta, \beta_0) - r_n\delta_j | Z_j)] K_{h,ij}\psi_{ij}, \quad \beta \in B. \end{aligned}$$

By the arguments used for Lemma 3.8 above, the class of functions  $\{H_{1n}(\cdot, \cdot, \beta) : \beta \in B\}$  is *Euclidean*( $c, d$ ) for an envelope with a finite fourth moment, with  $c$  and  $d$  independent of  $n$ . Now, we can use equation (A.11) of Zheng (1998) and his Lemma 1 with the condition (ii) replaced by  $\mathbb{E}[H_{1n}(\cdot, \beta) - H_{1n}(\cdot, \beta_0)]^2 \leq \Lambda \|\beta - \beta_0\|$ . By a close inspection of the proof of Zheng's Lemma 1, see his equations (A.2) to (A.5), it is obvious to adapt his conclusion and to deduce that in our setup for any  $0 < \alpha < 1$

$$W_{1n}(\beta) - W_{1n}(\beta_0) = O_{\mathbb{P}}\left(n^{-1}h^{-1} \|\beta - \beta_0\|^{\alpha/2}\right) = O_{\mathbb{P}}\left(n^{-1}h^{-1} \{r_n + n^{-1/4}\}^\alpha\right)$$

uniformly over  $O_{\mathbb{P}}(r_n^2 + n^{-1/2})$  neighborhoods of  $\beta_0$ . Thus, when  $n^{1/2}r_n^2 \rightarrow \infty$ , we have

$$W_{1n}(\hat{\beta}) - W_{1n}(\beta_0) = O_{\mathbb{P}}\left(n^{-1}h^{-1}r_n^\alpha\right) = O_{\mathbb{P}}\left(n^{-1/2}\right) = o_{\mathbb{P}}\left(r_n^2\right),$$

whereas in the case where  $n^{1/2}r_n^2$  is bounded, use  $nh^{1/2}r_n^2 \rightarrow \infty$  and take  $\alpha$  sufficiently close to one to obtain

$$W_{1n}(\hat{\beta}) - W_{1n}(\beta_0) = O_{\mathbb{P}}\left(n^{-1-\alpha/4}h^{-1}\right) = o_{\mathbb{P}}\left(r_n^2\right).$$

The remaining terms  $W_{2n}$ ,  $W_{3n}$  and  $W_{4n}$  can be treated in the following way. By Hoeffding's decomposition

$$r_n^{-1}hW_{3n}(\beta) = U_n^2(\beta) + U_n^1(\beta)$$

with  $U_n^1$ ,  $U_n^2$  degenerate  $U$ -processes of order 1 and 2, respectively. In view of Assumption 3.3(d) and the fact that  $K(\cdot)$  is bounded, apply Corollary 4 of Sherman (1994) to deduce that  $U_n^2(\beta) = O_{\mathbb{P}}(n^{-1})$  uniformly in  $\beta$ . If  $K_{h,ij}(\theta) = K_h((X_i - X_j)'\theta)$  and

$$\xi(Z_i) = \mathbb{E}\left[\mathbb{E}\left\{f_\varepsilon(0 | Z_j)\delta(Z_j) | Z_j'\theta\right\} h^{-3/4}K_{h,ij}\psi_{ij} | Z_i\right]$$

we can write

$$U_n^1(\beta) = \frac{h^{3/4}}{n} \sum_i [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n\delta_i\} - F_\varepsilon(G_i(\beta, \beta_0) - r_n\delta_i | Z_i)] \xi(Z_i).$$

By Hölder inequality, Assumption 3.2(c) and a change of variables,

$$|\xi(Z_i)| \leq \mathbb{E}^{1/4}[\delta^4(Z_j)] \mathbb{E}^{3/4}[h^{-1}K_{h,ij}^{4/3} | Z_i] \leq C,$$

for some  $C > 0$ . Now, by Corollary 4 of Sherman (1994),  $h^{-3/4}U_n^1(\beta) = O_{\mathbb{P}}(n^{-1/2})$  uniformly in  $\beta$ . As  $nh^{1/2}r_n^2 \rightarrow \infty$ , deduce that

$$\sup_{\beta} |W_{3n}(\beta)| = O_{\mathbb{P}}(r_n h^{-1} n^{-1} + r_n h^{-1/4} n^{-1/2}) = o_{\mathbb{P}}(r_n^2).$$

By similar arguments,  $\sup_{\beta} |W_{4n}(\beta)| = o_{\mathbb{P}}(r_n^2)$  (here apply Hölder inequality with  $p = q = 2$ ) and  $W_{3n}$ ,  $\sup_{\beta} |\widetilde{W}_{2n}(\beta)| = O_{\mathbb{P}}(h^{-1} n^{-1} + h^{-1/4} n^{-1/2})$ , and thus

$$\sup_{\beta} |W_{2n}(\beta)| = O_{\mathbb{P}}(r_n^2 + n^{-1/2}) O_{\mathbb{P}}(h^{-1} n^{-1} + h^{-1/4} n^{-1/2}) = o_{\mathbb{P}}(r_n^2).$$

Collecting results, under  $\mathcal{H}_{1n}$ ,  $T_n \geq Cnh^{1/2}r_n^2\{1 + o_{\mathbb{P}}(1)\}$  or some constants  $C > 0$ . Now, the proof is complete.

### 3.4.3 Proof of Theorem 3.7

Let  $W_n^*(\beta)$  be the statistic obtained after replacing  $U_i(\beta)$  with  $U_i^*(\beta) = \mathbb{I}\{Y_i^* \leq g(Z_i; \beta)\} - \tau$  in the formula of  $W_n(\beta)$ . The proof of the bootstrap procedure consistency follows the steps of the proof of Theorem 3.5, but requires several specific ingredients: (a) the convergence in law of  $nh^{1/2}W_n^*(\hat{\beta})/v_n$  *conditionally* upon the original sample; and (b) the  $O_{\mathbb{P}}(n^{-1/2})$  rate for  $\hat{\beta}^* - \hat{\beta}$ , and the negligibility of  $W_n^*(\hat{\beta}^*) - W_n^*(\hat{\beta})$  given the original sample. If  $S_{1n}^*$  and  $S_{2n}^*$  denote bootstrapped statistics,  $S_{1n}^*$  is bounded in probability given the sample if

$$\lim_{M \rightarrow \infty} \mathbb{P}[|S_{1n}^*| > M \mid Y_1, Z_1, \dots, Y_n, Z_n] = o_p(1).$$

while  $S_{2n}^*$  is asymptotically negligible given the sample if

$$\forall \epsilon > 0, \quad \mathbb{P}[|S_{2n}^*| > \epsilon \mid Y_1, Z_1, \dots, Y_n, Z_n] = o_p(1).$$

The asymptotic normality of  $nh^{1/2}W_n^*(\hat{\beta})/v_n$  given the sample is obtained below from a martingale central limit theorem as stated in Hall and Heyde (1980).

**Lemma 3.9.** *Under the assumptions of Theorem 3.7,*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(nh^{1/2}W_n^*(\hat{\beta})/v_n \leq t \mid Y_1, Z_1, \dots, Y_n, Z_n\right) - \Phi(t) \right| \rightarrow 0, \quad \text{in probability.}$$

*Proof.* The proof is based on the Central limit Theorem (CLT) for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). Recall that  $U_i^*(\hat{\beta}) = \mathbb{I}\{Y_i^* \leq g(Z_i; \hat{\beta})\} - \tau$ . Define the martingale array  $\{S_{n,m}^*, \mathcal{F}_{n,m}^*, 1 \leq m \leq n, n \geq 1\}$  where  $S_{n,1}^* = 0$  and  $S_{n,m}^* = \sum_{i=2}^m G_{n,i}^*$  with

$$G_{n,i}^* = \frac{2h^{-1/2}}{n-1} U_i^*(\hat{\beta}) \sum_{j=1}^{i-1} U_j^*(\hat{\beta}) K_{h,ij} \psi_{ij},$$

and  $\mathcal{F}_{n,m}^*$  is the  $\sigma$ -field generated by  $\{\bar{Z}, \eta_1, \dots, \eta_m\}$  where the sample is denoted by  $\bar{Z} = \{Y_1, \dots, Y_n, Z_1, \dots, Z_n\}$ . Thus  $nh^{1/2}W_n^*(\hat{\beta}) = S_{n,n}^*$ . Next define

$$\begin{aligned}
V_n^{2*} &= \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^{2*} \mid \mathcal{F}_{n,i-1}^* \right] \\
&= \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} U_j^*(\hat{\beta}) U_k^*(\hat{\beta}) K_{h,ij} K_{h,ik} \psi_{ij} \psi_{ik} \\
&= \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} U_j^{*2}(\hat{\beta}) K_{h,ij}^2 \psi_{ij}^2 \\
&\quad + \frac{8h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} U_j^*(\hat{\beta}) U_k^*(\hat{\beta}) K_{h,ij} K_{h,ik} \psi_{ij} \psi_{ik} \\
&= A_n^* + B_n^*.
\end{aligned}$$

Recall that

$$v_n^2 = \frac{2h^{-1}\tau^2(1-\tau)^2}{n(n-1)} \sum_{j \neq i} K_{h,ij}^2 \psi_{ij}^2$$

and by standard calculations of the means and variance it could be shown to tend to a positive constant. Next, note that

$$\mathbb{E} \left[ A_n^* \mid \bar{Z} \right] = \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} \left[ U_j^{*2}(\hat{\beta}) \mid \bar{Z} \right] K_{h,ij}^2 \psi_{ij}^2 = \frac{n}{n-1} v_n^2.$$

Moreover,

$$\begin{aligned}
\mathbb{E} \left[ \text{Var} \left( A_n^* \mid \bar{Z} \right) \right] &= \frac{16\tau^2(1-\tau)^2}{h^2(n-1)^4} \\
&\quad \times \sum_{i=2}^n \sum_{i'=2}^n \sum_{j=1}^{i \wedge i'-1} \mathbb{E} \left[ \mathbb{E} \left[ U_j^{*4}(\hat{\beta}) - \tau^2(1-\tau)^2 \mid \bar{Z} \right] K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2 \right] \\
&= \frac{16\tau^4(1-\tau)^4 \{ \tau(1-\tau)(1-3\tau(1-\tau)) - 1 \}}{h^2(n-1)^4} \\
&\quad \times \sum_{i=2}^n \sum_{i'=2}^n \sum_{j=1}^{i \wedge i'-1} \mathbb{E} \left[ K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2 \right] \\
&= \frac{32\tau^4(1-\tau)^4 (\tau(1-\tau)(1-3\tau(1-\tau)) - 1)}{h^2(n-1)^4} \\
&\quad \times \sum_{i=3}^n \sum_{i'=2}^{i-1} \sum_{j=1}^{i'-1} \mathbb{E} \left[ K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2 \right] \\
&\quad + \frac{16\tau^4(1-\tau)^4 (\tau(1-\tau)(1-3\tau(1-\tau)) - 1)}{h^2(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} \left[ K_{h,ij}^4 \psi_{ij}^4 \right] \\
&= O(n^{-1}) + O(n^{-2}h^{-1})
\end{aligned}$$

because  $\psi_{ij}$ ,  $\mathbb{E} \left[ h^{-1} K_{h,ij}^4 \right]$  and  $\mathbb{E} \left[ h^{-2} K_{h,ij}^2 K_{h,i'j}^2 \right]$  are bounded for all pairwise distinct indexes  $i, i'$  and  $j$ . Deduce that  $A_n^*/v_n^2 \rightarrow 1$  in probability. On the other hand,

$$\mathbb{E} \left[ B_n^{*2} \right] = \frac{8\tau^4(1-\tau)^4}{h^2(n-1)^4} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \mathbb{E} \left[ K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2 \right] = O(n^{-1})$$

so that  $V_n^{2*}/v_n^2 \rightarrow 1$  in probability. To use the CLT it remains to check the Lindeberg condition. For any  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{E} \left[ \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^{*2} \mathbb{I}(G_{n,i}^{*2} > \epsilon) \mid \mathcal{F}_{n,i-1}^* \right] \right] &\leq \epsilon^{-4} \mathbb{E} \left[ \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^{*4} \mid \mathcal{F}_{n,i-1}^* \right] \right] \\ &\leq \frac{16\tau^3(1-\tau)^3 \{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} \mathbb{E} \left[ K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2 \right] \\ &\leq \frac{32\tau^3(1-\tau)^3 \{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \mathbb{E} \left[ K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2 \right] \\ &+ \frac{16\tau^3(1-\tau)^3 \{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E} \left[ K_{h,ij}^4 \psi_{ij}^4 \right] \\ &= O(n^{-1}) + O(n^{-2}h). \end{aligned}$$

Eventually, applying the CLT for martingale arrays along the subsequences of  $V_n^{2*}$  that converge almost surely to the limit of  $v_n^2$  and subsequences for which the Lindeberg condition is satisfied almost surely, the result follows.  $\square$

To obtain the  $O_{\mathbb{P}}(n^{-1/2})$  rate for  $\widehat{\beta}^* - \widehat{\beta}$ , and the negligibility of  $W_n^*(\widehat{\beta}^*) - W_n^*(\widehat{\beta})$  given the original sample, we use a conditional version of the moment inequality for  $U$ -processes proved by Sherman (1994). Before stating this new result that has its own interest let us introduce some more notation: for  $k$  a positive integer let  $(n)_k = n(n-1)\dots(n-k+1)$  and let  $\mathbf{i}_k^n = (i_1, \dots, i_k)$  be a  $k$ -tuple of distinct integers from the set  $\{1, \dots, n\}$ . Similarly,  $\mathbf{i}_k^{2n} = (i_1, \dots, i_k)$  denotes a  $k$ -tuples of distinct integers from  $\{1, \dots, 2n\}$ . Moreover, a function  $g$  on  $\mathcal{S}^k$  is called degenerate if for each  $i = 1, \dots, k$ , and all  $s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_k \in \mathcal{S}$ ,  $\mathbb{E}[g(s_1, \dots, s_{i-1}, \mathcal{S}, s_{i+1}, \dots, s_k)] = 0$ .

**Lemma 3.10.** *Let  $k$  be a positive integer and  $\mathcal{G}$  a degenerate class of real-valued functions on  $\mathbb{R}^{1+q} \times \dots \times \mathbb{R}^{1+q}$ . Suppose  $\mathcal{G}$  is Euclidean(c,d) for a squared integrable envelope and some  $c, d > 0$ . Fix  $z_1, \dots, z_n \in \mathbb{R}^q$  and let  $u_1, \dots, u_n, u_{n+1}, \dots, u_{2n}$  be independent copies of the random variable  $u$ . For  $i = 1, \dots, n$ , let  $v_i = (u_i, z_i)$  and  $v_{n+i} = (u_{n+i}, z_i)$ . Define  $g_{\mathbf{i}_k^n}(u_{i_1}, \dots, u_{i_k}) = g(v_{i_1}, \dots, v_{i_k})$  and define  $g_{\mathbf{i}_k^{2n}}$  similarly. Suppose that for any  $k$ -tuple  $\mathbf{i}_k^n$ , the function  $g_{\mathbf{i}_k^n}$  is degenerate as a function of  $u_i$  variables (necessarily the same property holds also for any  $k$ -tuple  $\mathbf{i}_k^{2n}$ ). Let*

$$U_{n,z_1,\dots,z_n}^k(g) = (n)_k^{-1} \sum_{\mathbf{i}_k^n} g_{\mathbf{i}_k^n}(u_{i_1}, \dots, u_{i_k}), \quad U_{2n,z_1,\dots,z_n}^k(g) = (2n)_k^{-1} \sum_{\mathbf{i}_k^{2n}} g_{\mathbf{i}_k^{2n}}(u_{i_1}, \dots, u_{i_k}).$$

Then for any  $\alpha \in (0, 1)$ , there exists a constant  $\Lambda$  depending only on  $\alpha$  and  $k$  (and independent of  $n$  and the sequence  $z_1, \dots, z_n$ ) such that

$$\mathbb{E} \left[ \sup_{\mathcal{G}} |n^{k/2} U_{n, z_1, \dots, z_n}^k(g)| \right] \leq \Lambda \mathbb{E}^{1/2} \left[ \sup_{\mathcal{G}} \{U_{2n, z_1, \dots, z_n}^k(g^2)\}^\alpha \right].$$

*Proof.* We sketch the steps of the proof that follows the lines of the proof of the Main Corollary in Sherman (1994). For the sake of simplicity, we only consider the case of Euclidean families for a constant envelope. Fix  $n$  and  $z_1, \dots, z_n$  arbitrarily.

*i) Symmetrization inequality.* For each  $g \in \mathcal{G}$  define  $\tilde{g}(\mathbf{i}_k^n)$  as a sum of  $2^k$  terms, each having the form

$$(-1)^r g_{\mathbf{i}_k^n}(u_{i_1}^*, \dots, u_{i_k}^*)$$

with  $u_{i_j}^*$  equal to either  $u_{i_j}$  or  $u_{n+i_j}$  where  $i_j$  ranges over the set  $\{1, \dots, n\}$ , and  $r$  is the number of elements  $u_{i_1}^*, \dots, u_{i_k}^*$  belonging to  $\{u_{n+1}, \dots, u_{2n}\}$ . Independently, take a sample  $\sigma_1, \dots, \sigma_n$  of Rademacher random variables, that is symmetric variables on the two points set  $\{-1, 1\}$ . Let  $\Phi$  be a convex function on  $[0, \infty)$ . Then

$$\mathbb{E} \Phi \left( \sup_{\mathcal{G}} \left| \sum_{\mathbf{i}_k^n} g_{\mathbf{i}_k^n}(u_{i_1}, \dots, u_{i_k}) \right| \right) \leq \mathbb{E} \Phi \left( \sup_{\mathcal{G}} \left| \sum_{\mathbf{i}_k^n} \sigma_{i_1} \dots \sigma_{i_k} \tilde{g}(\mathbf{i}_k^n) \right| \right). \quad (3.8)$$

The proof of this inequality is omitted as it can be derived with only formal changes from the proof of Sherman (1994)'s symmetrization inequality. It can be also be derived from the lines of de la Peña and Giné (1999), Theorem 3.5.3 (see also Remark 3.5.4 of de la Peña and Giné).

*ii) Maximal inequality.* The following arguments are similar to those in Sherman (1994), section 5. Define the stochastic process

$$Z(g) = n^{k/2} \sum_{\mathbf{i}_k^n} \sigma_{i_1} \dots \sigma_{i_k} \tilde{g}(\mathbf{i}_k^n), \quad g \in \mathcal{G}$$

and the pseudo-metric  $d_{U_{2n}^k}(g_1, g_2) = [U_{2n, z_1, \dots, z_n}^k(|g_1 - g_2|^2)]^{1/2}$ . Finally, let us remark that for each  $g$ , by Cauchy-Schwarz inequality and the definitions of  $\tilde{g}(\mathbf{i}_k^n)$  and  $g_{\mathbf{i}_k^{2n}}$  we have

$$\sum_{\mathbf{i}_k^n} \tilde{g}(\mathbf{i}_k^n)^2 \leq 2^k \sum_{\mathbf{i}_k^{2n}} g_{\mathbf{i}_k^{2n}}^2(u_{i_1}, \dots, u_{i_k}) = 2^k (2n)_k U_{2n, z_1, \dots, z_n}^k(g^2)$$

which is the counterpart of inequality (5) of Sherman (1994). Now, we have all the ingredients to continue exactly as in the proof of Sherman's maximal inequality and to deduce that for any positive integer  $m$

$$\mathbb{E} \left[ \sup_{\mathcal{G}} |n^{k/2} U_{n, z_1, \dots, z_n}^k(g)| \right] \leq \Gamma \mathbb{E} \left[ \int_0^{\delta_n^k} [D(x, d_{U_{2n}^k}, \mathcal{G})]^{1/2m} dx \right]$$

where  $D(\epsilon, d_{U_{2n}^k}, \mathcal{G})$  are the *packing numbers* of the set  $\mathcal{G}$  with respect to the pseudometric  $d_{U_{2n}^k}$ ,  $\delta_n^k = \sup_{\mathcal{G}} \sqrt{U_{2n, z_1, \dots, z_n}^k(g^2)}$  and  $\Gamma$  is a constant depending only on  $m$  and  $k$ .

*iii) Moment inequality for Euclidean families.* If  $\mathcal{G}$  is *Euclidean*( $c, d$ ) for a constant envelope equal to one, then the packing number  $D(\epsilon, d_{U_{2n}^k}, \mathcal{G})$  is bounded by  $c\epsilon^{-d}$ . To check this, apply the definition of an Euclidean family for  $\mathcal{G}$  with  $\mu$  the measure that places mass  $(2n)_k^{-1}$  at each of the  $(2n)_k$  pairs  $(v_i, v_j)$ ,  $1 \leq i \neq j \leq 2n$ . Finally, our result follows using the arguments of the Main Corollary of Sherman (1994).  $\square$

To establish the rate of  $\hat{\beta}^* - \hat{\beta}$  given the sample, it suffices to consider a simplified version of our Lemma 3.8. By Lemma 3.10,  $\sup_{\beta} |n^{-1} \Gamma_n^*(\beta) - \mathbb{E}[\rho_{\tau}(Y - g(Z; \beta)) \mid \bar{Z}]|$  is asymptotically negligible given the sample  $\bar{Z} = \{Y_1, \dots, Y_n, Z_1, \dots, Z_n\}$ . Reconsidering the arguments for the consistency of argmax estimators along almost surely convergent subsequences depending on  $\bar{Z}$ , deduce that  $\hat{\beta}^* - \hat{\beta}$  is asymptotically negligible given the sample  $\bar{Z}$ . Next, define the empirical process

$$\nu_n^*(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi_{\tau}(Y_i^* - g(Z_i; \beta)) - \mathbb{E}[\psi_{\tau}(Y_i - g(Z_i; \beta)) \mid \bar{Z}] \right\} \dot{g}(Z_i; \beta)$$

indexed by  $\beta$ . Lemma 3.10 guarantees that  $\sup_{\beta} |\nu_n^*(\beta)|$ , and in particular  $\nu_n^*(\hat{\beta}^*) - \nu_n^*(\hat{\beta})$ , are bounded in probability given the sample. Proceeding like in (3.6), that is using the directional derivative of  $\Gamma_n^*(\beta)$  at  $\hat{\beta}^*$  along any direction  $\gamma$ , deduce

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\tau}(Y_i^* - g(Z_i; \hat{\beta}^*)) \dot{g}(Z_i; \hat{\beta}^*)$$

is bounded in probability given the sample (conditional negligibility could be also derived but boundedness given the sample suffices for the present purpose). Since for all  $i$ ,

$$\mathbb{E}[\psi_{\tau}(Y_i^* - g(Z_i; \hat{\beta}^*)) \mid \bar{Z}] = F_{\epsilon^*}(g(Z_i; \hat{\beta}^*) - g(Z_i; \hat{\beta}) \mid \bar{Z}) - \tau,$$

and for any sample  $\bar{Z}$ , the distribution function  $F_{\epsilon^*}(\cdot \mid \bar{Z})$  is that of the uniform law on  $[-\tau, 1 - \tau]$ , the boundedness of  $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$  follows by a Taylor expansion of  $F_{\epsilon^*}(\cdot \mid \bar{Z})$  around the origin, exactly like in the proof of Lemma 3.8 in the case  $r_n = 0$ . The case of the wild bootstrap and linear quantile regression follows as a consequence of Theorem 1 of Feng et al. (2011). The arguments of Theorem 1 of Feng et al. (2011) could be adapted to nonlinear models using a linearization like in the proof of Lemma 3.8. The details are omitted.

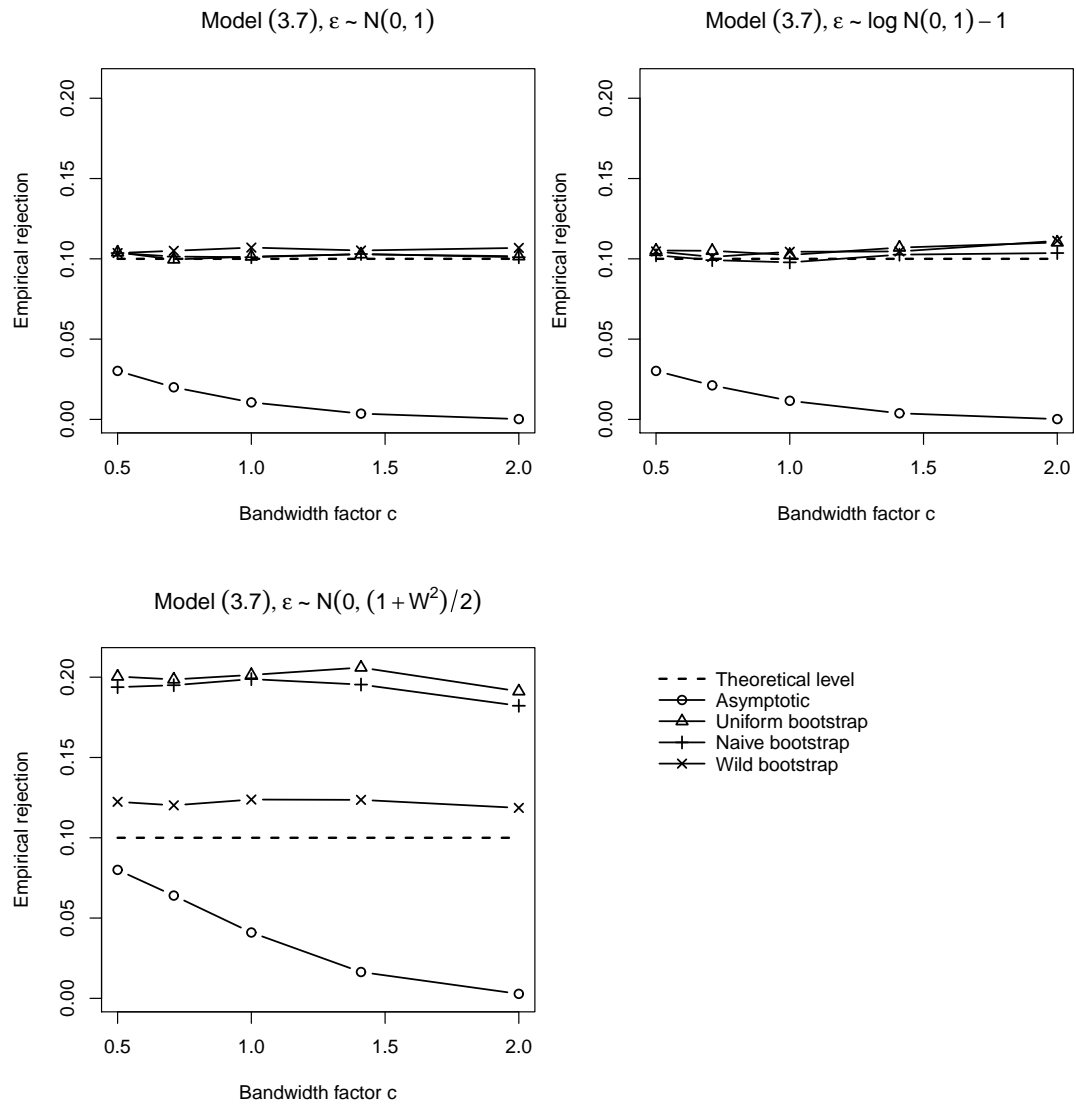
Finally, using Lemma 3.10, derive conditional versions of Lemma 1 of Zheng (1998) and of Corollary 4 of Sherman in the case of constant envelopes. Combine these results with the fact that  $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$  is bounded in probability given the sample and follow the lines of the proof of Theorem 3.5 above to deduce that for any  $\epsilon > 0$

$$\mathbb{P}\left(nh^{1/2} \left| W_n^*(\hat{\beta}^*) - W_n^*(\hat{\beta}) \right| > \epsilon \mid Y_1, Z_1, \dots, Y_n, Z_n\right) \rightarrow 0, \quad \text{in probability.}$$

**Table 3.1:** Application: estimation results and tests p-values

	$\tau = 0.5$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.1$
CIGAR	-5.35 (2.28)	-7.53 (4)	-5.05 (2.3)	-8.36 (3.53)	-5.07 (2.36)	-8.07 (3.25)
WTGAIN	8.09 (1.33)	14.73 (0.75)	7.69 (1.32)	14.96 (1.2)	8.31 (1.31)	15.91 (1.4)
AGE	-9.34 (3.82)	-5.13 (4.47)	43.6 (50.59)	133.67 (30.11)	78.59 (45.85)	117.62 (48.42)
AGESQ			-0.84 (0.81)	-2.23 (0.5)	-1.38 (0.72)	-1.94 (0.82)
BOY					137.22 (34.35)	-5.22 (47.33)
BLACK					-177.78 (75.09)	-124.18 (69.17)
MARRIED					21.62 (48.39)	41.75 (54.66)
NOVISIT					-211.62 (406.72)	-275.15 (112.5)
HZ	0.347	0.227	0.266	0.356	0.272	0.135
Our test $c=1$	0.791	0.165	0.738	0.942	0.068	0.972
Our test $c=2$	0.704	0.044	0.741	0.968	0.078	0.796





**Figure 3.1:** Empirical rejections under  $H_0$  as a function of the bandwidth,  $n = 100$

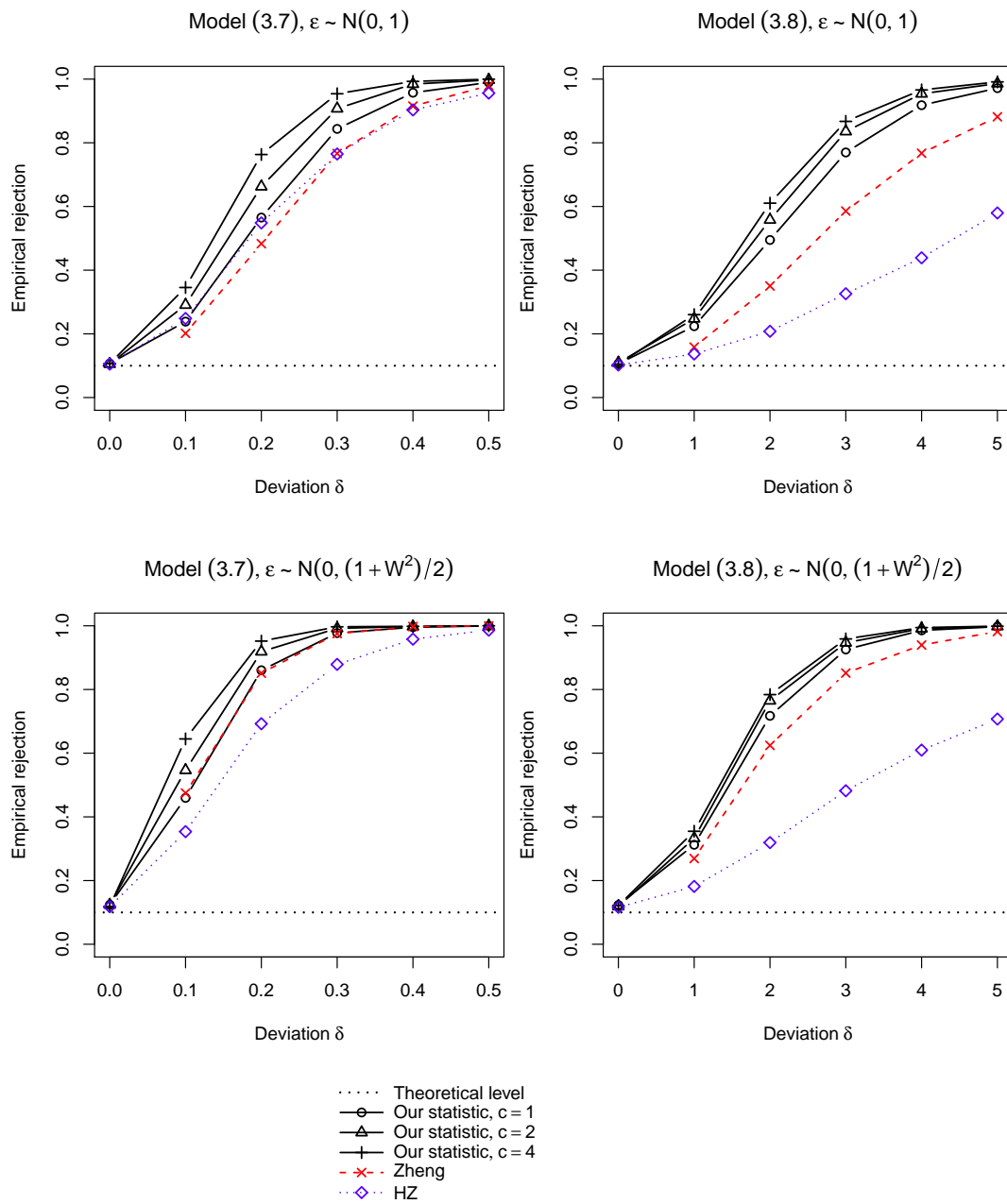


Figure 3.2: Power curves for models (3.7) and (3.8),  $n = 100$ .



## **4 Test d'adéquation pour la régression de variables fonctionnelles**



## Abstract

Regression models with a response variable taking values in a Hilbert space and hybrid covariates are considered. This means two sets of regressors are allowed, one of finite dimension and a second one functional with values in a Hilbert space. The problem we address is the test of the effect of the functional covariates. This problem occurs for instance when checking the goodness-of-fit of some regression models for functional data against general alternatives. The significance test for functional regressors in nonparametric regression with hybrid covariates and scalar or functional responses is another example where the core problem is the test on the effect of functional covariates. We propose a new test based on kernel smoothing. The test statistic is asymptotically standard normal under the null hypothesis provided the smoothing parameter tends to zero at a suitable rate. The one-sided test is consistent against any fixed alternative and detects local alternatives à la Pitman approaching the null hypothesis. In particular we show that neither the dimension of the outcome nor the dimension of the functional covariates influences the theoretical power of the test against such local alternatives. Simulation experiments and a real data application illustrate the performance of the new test with finite samples.

## 4.1 Introduction

Let  $(\mathcal{H}_1, \langle \cdot, \cdot \rangle_{\mathcal{H}_1})$  and  $(\mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_2})$  denote two possibly different Hilbert spaces. The main examples of Hilbert spaces we have in mind are  $\mathbb{R}^p$ , for some  $p \geq 1$ , and  $L^2[0, 1]$ , the space of squared integrable real-valued functions defined on the unit interval.

Consider the random variables  $U \in \mathcal{H}_1$  and  $W \in \mathcal{H}_2$  and let  $Z$  be a column random vector in  $\mathbb{R}^q$ ,  $q \geq 0$ . By convention,  $q = 0$  means that  $Z$  is a constant. Let  $(U_i, Z_i, W_i)$ ,  $1 \leq i \leq n$ , denote a sample of independent copies of  $(U, Z, W)$ . The statistical problem we consider is the test of the hypothesis

$$\mathbb{E}[U \mid Z, W] = 0 \quad \text{a.s.}, \quad (4.1)$$

against a general alternative like  $\mathbb{P}(\mathbb{E}[U \mid Z, W] = 0) < 1$ . This type of problem occurs in many model check problems.

Consider the random variables  $U \in \mathcal{H}_1$ ,  $\widetilde{W} \in \mathcal{H}_2$ . For illustration, suppose that  $U$  is centered. Consider the problem of *testing the effect of the functional variable  $\widetilde{W}$* , that is testing the condition  $\mathbb{E}[U \mid \widetilde{W}] = 0$ . Patilea et al. (2012b) proposed a test procedure based on projections into finite dimension subspaces of  $\mathcal{H}_2$ . Their test statistic is somehow related to a Kolmogorov-Smirnov statistic in a finite dimension space with the dimension growing with the sample size. Here we propose an alternative route that avoids optimization in high dimension. Let  $Z = \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2}$  where  $\phi_1$  is an element of an orthonormal basis of  $\mathcal{H}_2$ . Suppose that  $Z$  admits a density with respect to the Lebesgue measure. The basis of  $\mathcal{H}_2$  could be the one given by the functional principal components which in general has to be estimated from the data. In such a case, the sample of  $Z'_i$ s has to be estimated too. Let  $W = \widetilde{W} - \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$ . Then, testing  $\mathbb{E}[U \mid \widetilde{W}] = 0$  is nothing but testing condition (4.1).

Aneiros-Pérez and Vieu (2006) introduced the *semiparametric functional partially linear models* as an extension of the partially linear model to functional data. Such model writes as

$$Y = Z^\top \beta + m(W) + U, \quad \text{with} \quad \mathbb{E}[U \mid Z, W] = 0 \quad \text{a.s.},$$

where  $Y$  is a scalar response and  $Z$  is a  $q$ -dimension vector of random covariates,  $W$  is a random variable taking values in a functional space, typically  $L^2[0, 1]$ . The column vector of  $q$  coefficients  $\beta$  and the function  $m(\cdot)$  have to be estimated. Before estimating  $m(\cdot)$  nonparametrically, one should first check the significance of the variable  $W$  which means exactly testing condition (4.1). In this example, the variable  $U$  is not observed and the sample  $U_1, \dots, U_n$  could be estimated by the residuals of the linear fit of  $Y$  given  $Z$ . The estimation error for the sample of  $U$  is of rate  $O_{\mathbb{P}}(n^{-1/2})$  and could be easily proved to be negligible for our test.

Other examples of regression model checks that lead to a problem like (4.1) are the functional linear regression with scalar or functional responses, quadratic functional regression, generalized functional regression, *etc.* See for instance Horváth and Kokoszka (2012) for a recent panorama on the functional regression models. In such situations one has to estimate the sample  $U_1, \dots, U_n$  from the functional regression model considered. The estimation error is in general larger than the parametric rate  $O_{\mathbb{P}}(n^{-1/2})$ , but one can still show that, under reasonable conditions, it remains negligible for the test purposes. See Patilea et al. (2012b) for a related framework.

Another example, related to the problem of testing the effect of a functional variable, is the *variable selection in functional nonparametric regression with functional responses*. Regression models for functional responses are now widely used, see for instance Faraway (1997). Two situations were studied: finite and infinite dimension covariates; see Ramsay and Silverman (2005), Ferraty et al. (2011), Ferraty et al. (2012). Consider the hybrid case with both finite and infinite dimension covariates. An important question is the significance of the functional covariates. In a more formal way, let  $Y \in \mathcal{H}_1$  be the response and let  $Z \in \mathbb{R}^q$  and  $W \in \mathcal{H}_2$  denote the covariates. Then the problem is to test the equality

$$\mathbb{E}[Y \mid Z, W] = \mathbb{E}[Y \mid Z].$$

Let  $U = Y - \mathbb{E}[Y \mid Z]$ . Then the problem becomes to test whether  $\mathbb{E}[U \mid Z, W] = 0$  almost surely, that is the condition (4.1). Again the sample of the variable  $U$  is not observed and has to be estimated by the residuals of the nonparametric regression of  $Y$  given  $Z$ . See also Lavergne et al. (2014) for a related procedure.

As a last example where a condition like (4.1) occurs consider the problem of testing the independence between a random variable  $Y$  and a functional spaced valued variable  $\widetilde{W}$ . Without loss of generality, one could suppose that  $Y$  takes values in the unit interval. Define  $U(t) = \mathbf{1}\{Y \leq t\} - \mathbb{P}(Y \leq t)$ ,  $t \in [0, 1]$ , that is centered and belongs to  $L^2[0, 1]$ . The independence between  $Y$  and  $\widetilde{W}$  is equivalent to the condition  $\mathbb{E}[U \mid \widetilde{W}] = 0$ . Conditional independence of  $Y$  and a functional random variable given some finite random vector  $Z$  could be also tested. It suffices to define  $U(t)$  by centering with the conditional probability of the event  $\{Y \leq t\}$  given  $Z$  and to check a condition like (4.1).

To our best knowledge the statistical problem we address in this work was very little investigated in full generality. Chiou and Müller (2007) and Kokoszka et al. (2008) investigated the problem of goodness-of-fit with functional responses. Chiou and Müller (2007) considered plots of functional principal components (FPC) scores of the response and the covariate. They also used residuals versus fitted values FPC scores plots. However, such two dimension plots could not capture all types of effects of the covariate on the response. Kokoszka et al. (2008) used the response and covariate FPC scores to build a test statistic with  $\chi^2$  distribution under the null hypothesis of no linear effect. See also the textbook Horváth and Kokoszka (2012). Again, by construction, such tests cannot detect any nonlinear alternative. The



goodness-of-fit or no-effect against nonparametric alternatives has been recently explored in functional data context. In the case of scalar response, Delsol et al. (2011) proposed a testing procedure adapted from the approach of Härdle and Mammen (1993). Their procedure involves smoothing in the functional space and requires quite restrictive conditions. Patilea et al. (2012a) and García-Portugués et al. (2012) proposed alternative nonparametric goodness-of-fit tests for scalar response and functional covariate using projections of the covariate. Patilea et al. (2012b) extended the idea to functional responses and seems to be the only contribution allowing for functional responses. Such projection-based methods are less restrictive and perform well in applications. However, they require a search for the most suitable projection and this may involve optimization in high dimension.

The paper is organized as follows. In section 4.2 we introduce our testing approach, while in section 4.3 we provide the asymptotic analysis. The asymptotically standard normal critical values and the consistency of the test are derived. The application to goodness-of-fit tests of functional data models is discussed. The extension to the case of estimated covariates is presented in section 4.4. This allows in particular for an estimated basis in the infinite-dimensional space of the functional covariate. Section 4.5 presents some empirical evidence on the performances of our test and comparisons with existing procedures. The proofs and some technical lemmas are relegated to the Appendix.

## 4.2 The method

Let us first introduce some notation. Let  $\{\phi_1, \phi_2, \dots\}$  be some orthonormal basis of  $\mathcal{H}_2$  that for the moment is supposed to be fixed. In section 4.4 we consider the case of a data-driven basis. For simplicity and without any loss of generality in the following, assume hereafter that  $\mathbb{E}(W) = 0$ . Then we can decompose  $W = \sum_{k \geq 1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$  and the norm of  $W$  satisfies the relationship  $\|W\|_{\mathcal{H}_2}^2 = \sum_{k \geq 1} \langle W, \phi_k \rangle_{\mathcal{H}_2}^2$ . Let us note that

$$\mathbb{E}(U \mid Z, W) = \mathbb{E}(U \mid Z, \langle W, \phi_1 \rangle_{\mathcal{H}_2}, \langle W, \phi_2 \rangle_{\mathcal{H}_2}, \dots).$$

Next, for any positive integer  $p$ , let

$$W_p = (\langle W, \phi_1 \rangle_{\mathcal{H}_2}, \dots, \langle W, \phi_p \rangle_{\mathcal{H}_2})^\top.$$

For a function  $l$ , let  $\mathcal{F}[l]$  denote the Fourier Transform of  $l$ . Let  $K$  be a multivariate kernel defined on  $\mathbb{R}^q$  such that  $\mathcal{F}[K] > 0$  and  $\varphi(s) = \exp(-\|s\|^2/2)$ ,  $\forall s \in \mathbb{R}^p$ , where here  $\|\cdot\|$  is the Euclidean norm in  $\mathbb{R}^p$ . Many univariate kernels satisfy the positive Fourier Transform condition, for instance the gaussian, triangle, Student and logistic densities. To obtain a multivariate kernel with positive Fourier Transform it suffice to consider a multiplicative kernel with positive Fourier Transform univariate kernels.

### 4.2.1 The idea behind the testing method

The new procedure proposed below is motivated by the following facts. First, for any positive function  $\omega(\cdot)$  and any  $h > 0$  and  $p$  positive integer, if  $(U_1, Z_1, W_1)$  and  $(U_2, Z_2, W_2)$  are independent copies of  $(U, Z, W)$ , by the Inverse Fourier Transform formula,

$$\begin{aligned} I_p(h) &= \mathbb{E} \left[ \langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) h^{-q} K((Z_1 - Z_2)/h) \varphi(W_{1,p} - W_{2,p}) \right] \\ &= \mathbb{E} \left[ \langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) \int_{\mathbb{R}^q} e^{2\pi i t^\top (Z_1 - Z_2)} \mathcal{F}[K](th) dt \right. \\ &\quad \left. \times \int_{\mathbb{R}^p} e^{2\pi i s^\top (W_{1,p} - W_{2,p})} \mathcal{F}[\varphi](s) ds \right] \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left\| \mathbb{E} \left[ \mathbb{E}[U | Z, W_p] \omega(Z) e^{-2\pi i \{t^\top Z + s^\top W_p\}} \right] \right\|_{\mathcal{H}_1}^2 \mathcal{F}[K](th) \mathcal{F}[\varphi](s) dt ds. \end{aligned}$$

By the properties of the Fourier Transform and the conditions  $\mathcal{F}[\varphi], \mathcal{F}[K] > 0$  (and  $\omega > 0$ ), for any  $h > 0$  and  $p$  the real number  $I_p(h)$  is nonnegative and

$$\mathbb{E}(U | Z, W_p) = 0 \text{ a.s.} \Leftrightarrow I_p(h) = 0.$$

Second, by a martingale convergence argument with respect to  $p$ , it follows that

$$\mathbb{E}(U | Z, W) = 0 \text{ a.s.} \Leftrightarrow \mathbb{E}(U | Z, W_p) = 0 \text{ a.s.} \quad \forall p \in \{1, 2, \dots\}.$$

These intuitions are formalized in the following fundamental lemma, up to some technical modification. In the following  $a_1, a_2, \dots$  is a fixed sequence of positive real numbers. For any sequences  $u = \{u_k\}_{k \geq 1}$ ,  $v = \{v_k\}_{k \geq 1}$ , let

$$\langle u, v \rangle_{\mathcal{A}} = \sum_{k \geq 1} a_k u_k v_k \quad \text{and} \quad \|u\|_{\mathcal{A}}^2 = \sum_{k \geq 1} a_k u_k^2, \quad (4.1)$$

whenever the series converge.

**Lemma 4.1.** *Assume that  $\omega(\cdot) > 0$ ,  $\mathcal{F}[K] > 0$ ,  $\mathcal{F}[K]$  is integrable,  $\mathbb{E}(\|W\|_{\mathcal{H}_2}^2) < \infty$  and  $\mathbb{E}(\omega^2(Z) \|U\|_{\mathcal{H}_1}^2) < \infty$ . Assume that  $\sum_{k \geq 1} a_k < \infty$  and let*

$$I(h) = \mathbb{E} \left[ \langle U_1, U_2 \rangle_{\mathcal{H}_1} \omega(Z_1) \omega(Z_2) h^{-q} K((Z_1 - Z_2)/h) \exp(-\|W_1 - W_2\|_{\mathcal{A}}^2/2) \right], \quad h > 0.$$

*Then, for any  $h > 0$  we have*

$$\mathbb{E}(U | Z, W) = 0 \text{ a.s.} \quad \Leftrightarrow \quad I(h) = 0.$$

The reason for introducing a sequence  $\{a_k\}$  with convergent partial sums is technical. It allows for an inverse Fourier Transform formula in infinite-dimensional Hilbert spaces. In the remark following Theorem 4.2 we argue that considering the weighted norm  $\|\cdot\|_{\mathcal{A}}$  is not restrictive.

The idea behind the new approach we propose is to build a test statistic using an approximation of  $I(h)$ . Moreover, we will let  $h$  tend to zero in order to obtain an asymptotically pivotal test statistic with standard gaussian critical values. A convenient choice of the function  $\omega(\cdot)$  will allow to simplify this task. As explained below, in many examples one could simply take  $\omega(\cdot) \equiv 1$ .

## 4.2.2 The test statistics

To estimate  $I(h)$  using the i.i.d. sample  $(U_i, Z_i, W_i)$ ,  $1 \leq i \leq n$ , we consider the  $U$ -statistic

$$I_n(h) = \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij},$$

where

$$K_{ij}(h) = K((Z_i - Z_j)/h), \quad \varphi_{ij} = \exp(-\|W_i - W_j\|_{\mathcal{A}}^2/2). \quad (4.2)$$

The variance of  $I_n(h)$  could be estimated by

$$v_n^2(h) = \frac{2}{n^2(n-1)^2 h^{2q}} \sum_{1 \leq i \neq j \leq n} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}^2 K_{ij}^2(h) \varphi_{ij}^2.$$

Then, the test statistic is

$$T_n = \frac{I_n(h)}{v_n(h)}. \quad (4.3)$$

When the  $U_i \omega(Z_i)$ 's need to be estimated, the test statistics becomes

$$\widehat{T}_n = \frac{\widehat{I}_n(h)}{\widehat{v}_n(h)}, \quad (4.4)$$

where

$$\widehat{I}_n(h) = \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i \omega(Z_i)}, \widehat{U_j \omega(Z_j)} \right\rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij},$$

$$\widehat{v}_n^2(h) = \frac{2}{n^2(n-1)^2 h^{2q}} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i \omega(Z_i)}, \widehat{U_j \omega(Z_j)} \right\rangle_{\mathcal{H}_1}^2 K_{ij}^2(h) \varphi_{ij}^2.$$

and the  $\widehat{U_i \omega(Z_i)}$  are some estimates of the  $U_i \omega(Z_i)$ 's.

In the example on testing the effect of a functional variable the  $U_i$  are supposed observed so that  $T_n$  could be used. For the semiparametric functional partially linear models, to build  $\widehat{T}_n$  it is convenient to take the  $\omega(Z_i)$  constant equal to 1 while the  $\widehat{U}_i$  will be the residuals of the linear model with response  $Y$  and covariate vector  $Z \in \mathbb{R}^q$ . In the other examples of functional regression models mentioned above (functional linear regression with scalar or functional responses, quadratic functional regression, generalized functional regression, *etc.*), it is convenient to set all  $\omega(Z_i)$  equal to 1 and take the  $\widehat{U}_i$ 's to be the residuals of the functional regression model. Below we will provide an example of argument for showing that, under suitable assumptions, replacing the  $U_i$ 's by the  $\widehat{U}_i$ 's does not change the asymptotic behavior of our test

statistics. Next, for variable selection in functional nonparametric regression with functional responses one can use  $\widehat{T}_n$  and a convenient choice is  $\omega(\cdot)$  equal to the density of  $Z$  and

$$\widehat{U_i\omega}(Z_i) = \frac{1}{n-1} \sum_{k \neq i} (Y_i - Y_k) \frac{1}{g^q} L_{ik}(g),$$

where  $L$  is another kernel,  $L_{ik}(g) = L((Z_i - Z_k)/g)$  and  $g$  is a bandwidth converging to zero at a suitable rate. Showing that the estimation error of the  $U_i\omega(Z_i)$ 's is negligible for the testing purpose requires more complicated technical assumptions but could be obtained along the lines of the results of Lavergne et al. (2014). However, such an investigation is left for future work. Finally, for testing the independence between a  $[0, 1]$ -valued random variable  $Y$  and a  $L^2[0, 1]$ -valued random variable, one could take  $\omega(\cdot) \equiv 1$  and define  $\widehat{U}_i = \mathbf{1}\{Y_i \leq t\} - n^{-1} \sum_{j=1}^n \mathbf{1}\{Y_j \leq t\}$ .

## 4.3 Asymptotic theory

In this section we investigate the asymptotic properties of  $T_n$  under the null hypothesis (4.1) and under a sequence of alternative hypothesis. When the  $U_i\omega(X_i)$ 's have to be estimated, the idea is to show that the difference  $\widehat{T}_n - T_n$  is asymptotically negligible under suitable model assumptions. This aspect is investigated in section 4.3.3 below.

### 4.3.1 The asymptotic critical values

Under mild technical conditions we show that the test statistic is asymptotically standard normal under the null hypothesis  $\mathbb{E}[U | Z, W] = 0$  a.s.

#### Assumption D.

- (a) The random vectors  $(U_1, Z_1, W_1), \dots, (U_n, Z_n, W_n)$  are independent draws from the random vector  $(U, Z, W) \in \mathcal{H}_1 \times \mathbb{R}^q \times \mathcal{H}_2$  that satisfies  $\mathbb{E}\|U\omega(Z)\|_{\mathcal{H}_1}^4 < \infty$ .
- (b) (i) The vector  $Z$  admits a density  $f_Z$  that is either bounded or satisfies  $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t) dt < \infty$  for some  $a \in (0, 1]$ .  
(ii) The functional covariate satisfies  $\mathbb{E}[\|W\|_{\mathcal{H}_2}^2] < \infty$ .  
(iii) The norm  $\|\cdot\|_{\mathcal{A}}$  is defined like in equation (4.1) with a positive sequence  $\{a_k\}$  such that  $\sum_{k \geq 1} a_k < \infty$ .
- (c)  $\exists \underline{\sigma}^2, C > 0$  and  $\nu > 2$  such that:
  - (i)  $0 < \underline{\sigma}^2 \leq \mathbb{E} \left[ \langle U_1\omega(Z_1), U_2\omega(Z_2) \rangle_{\mathcal{H}_1}^2 \mid (Z_1, W_1), (Z_2, W_2) \right]$  almost surely;
  - (ii)  $\mathbb{E} \left[ \|U\omega(Z)\|_{\mathcal{H}_1}^\nu \mid Z, W \right] \leq C$  almost surely.

**Assumption K.**

- (a) The kernel  $K$  is multiplicative kernel in  $\mathbb{R}^q$ , that is  $K(z_1, \dots, z_q) = \prod_{k=1}^q \tilde{K}(z_k)$  where  $\tilde{K}$  is a symmetric density non increasing on  $[0, \infty)$ . The density  $\tilde{K}$  is differentiable except for a finite set of points and the derivative of  $\tilde{K}$  is integrable. Moreover the Fourier Transform  $\mathcal{F}[\tilde{K}]$  is positive and integrable.
- (b)  $h \rightarrow 0$  and  $nh^q / \ln n \rightarrow \infty$ .

**Theorem 4.2.** Under the Assumptions D and K the test statistic  $T_n$  converges in law to a standard normal, provided the hypothesis (4.1) is true. Consequently, the test given by  $\mathbb{I}(T_n \geq z_{1-\alpha})$ , with  $z_\alpha$  the  $(1 - \alpha)$ -quantile of the standard normal distribution, has asymptotic level  $\alpha$ .

*Remark 1.* Let us comment on Assumption D-(b)-(ii,iii). Suppose that the functional covariate satisfies  $\mathbb{E}[\langle W, \phi_k \rangle_{\mathcal{H}_2}^2] \sim k^{-\beta}$ ,  $\forall k \geq 1$ , for some  $\beta > 0$ . If  $\beta > 2$  one could use directly  $\|\cdot\|_{\mathcal{H}_2}$  instead of  $\|\cdot\|_{\mathcal{A}}$  to build the test. Indeed, taking  $a_k \sim k^{-\beta/2}$  and replacing  $W$  by  $\check{W} = \sum_{k \geq 1} b_k^{-1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$ , with  $b_k = a_k^{1/2}$ , one would have  $\mathbb{E}(U | Z, W) = \mathbb{E}(U | Z, \check{W})$ ,  $\check{W} \in \mathcal{H}_2$  and  $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}} = \|W_i - W_j\|_{\mathcal{H}_2}$ . In the case  $\beta = 2$ , one could take  $a_k \sim k^{-1} (\ln k)^{-(1+\varepsilon)}$ , for some  $\varepsilon > 0$ , and replace  $W$  by  $\check{W} = \sum_{k \geq 1} a_k^{-1/2} (\ln k)^{-(1+\varepsilon)} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$ . In this case one still has  $\mathbb{E}(U | Z, W) = \mathbb{E}(U | Z, \check{W})$  and  $\check{W} \in \mathcal{H}_2$ . However,  $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$  and  $\|W_i - W_j\|_{\mathcal{H}_2}$  are no longer equal but in general remain close. Our simulation experiments reveal that in many situations where  $\beta = 2$  one could confidently use  $\|W_i - W_j\|_{\mathcal{H}_2}$  instead of  $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$  to build the test.

Finally, with suitable choices, our setup covers also the range  $0 < \beta < 2$ . When  $1 < \beta < 2$ , one can set  $a_k \sim k^{-1} \ln^{-(1+\varepsilon)} k$  and work with  $\|W_i - W_j\|_{\mathcal{A}}$ . For the case  $0 < \beta \leq 1$  one could transform  $W$  in  $\check{W} = \sum_{k \geq 1} b_k^{-1} \langle W, \phi_k \rangle_{\mathcal{H}_2} \phi_k$  with  $b_k = k^{(1-\beta)/2} \ln^{(1+\varepsilon)/2} k$ , and take  $a_k \sim k^{\beta-2} \ln^{-2(1+\varepsilon)} k$ . The test is then built with  $\|\check{W}_i - \check{W}_j\|_{\mathcal{A}}$ .

In summary, Assumption-(D)-(ii,iii) represent mild conditions that are satisfied directly, or after simple modifications of the covariate  $W$ , in most situations.  $\square$

**4.3.2 The consistency of the test**

Let  $(U_i^0, Z_i, W_i)$ ,  $i \geq 1$ , i.i.d. such that  $\mathbb{E}(U_i^0 | Z_i, W_i) = 0$  almost surely. Here we show that our test is consistent against any fixed alternative and detect Pitman alternatives

$$H_{1n} : U_{in} = U_i^0 + r_n \delta(Z_i, W_i), \quad 1 \leq i \leq n, \quad n \geq 1,$$

with probability tending to 1, provided that the rate of decrease of the sequence  $\{r_n\}$  satisfies some conditions. These conditions are the same as for nonparametric checks of parametric regression models with finite dimension covariates.

**Theorem 4.3.** *Under the Assumptions D holds for the  $(U_i^0, Z_i, W_i)$ 's and Assumption K holds too. Suppose that  $\delta(\cdot, \cdot)$  and  $r_n$  defining the sequence of  $H_{1n}$  satisfy the conditions  $\mathbb{E}[\delta(Z, W)] = 0$ ,  $0 < \mathbb{E}[\|\delta(Z, W)\omega(Z)\|_{\mathcal{H}_1}^4] < \infty$  and  $r_n^2 nh^{q/2} \rightarrow \infty$ . Then the test based on  $T_n$  is consistent against the sequence of alternatives  $H_{1n}$ .*

The zero mean condition for  $\delta(\cdot)$  keeps  $U_{in}$  of zero mean under the alternative hypotheses  $H_{1n}$ . The proof is based on standard arguments and is relegated to the appendix.

### 4.3.3 Goodness-of-fit test

In this section we provide some guidelines on how our test could be used for testing the goodness-of-fit of functional data models. The detailed investigation of specific situations depend on the model and could not be considered in a unified framework.

In many situations, the  $U_i$ 's are not observed and one has to replace them by some  $\hat{U}_i$  obtained as residuals of some models. In this case one cannot build  $T_n$  and has to work with the statistic  $\hat{T}_n$  defined in equation (4.4) instead. In section 4.5 we use some simulation experiments to show that our test could still perform well in such situations, especially with a bootstrap correction, as described in the following, when the sample size is not large enough.

From the theoretical point of view, one shall expect that the asymptotic standard normal critical values are still valid and the test is still consistent, provided that the difference  $\hat{U}_i - U_i$  could be controlled in a suitable way. Indeed, using the notation from section (4.2.2) and considering the simple case where  $\omega(\cdot) \equiv 1$ , we can write

$$\begin{aligned} \hat{I}_n(h) &= I_n(h) \\ &\quad + \frac{2}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i, \hat{U}_j - U_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &\quad + \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle \hat{U}_i - U_i, \hat{U}_j - U_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &\stackrel{def}{=} I_n(h) + R_{1n}(h) + R_{2n}(h). \end{aligned}$$

Next, one has to control  $\hat{U}_i - U_i$  and hence  $R_{1n}(h)$  and  $R_{2n}(h)$  and this strongly depends on the specific model considered. Many functional data models would allow to show that  $R_{1n}(h)$  and  $R_{2n}(h)$  are negligible under reasonable conditions in the model (regularity conditions on the model parameter and the functional covariate  $W$ ) and for suitable rates of the bandwidth. For instance, Patilea et al. (2012a)

investigated in detail the case of linear model with scalar responses. Their investigation could be adapted to our test and obtain similar conclusions. In the case of a functional linear model with  $L^2[0, 1]$  responses and finite and infinite dimension covariates one would observe a sample of  $(Y, Z, W)$  where

$$Y(t) = Z^\top \beta + \langle \xi(t, \cdot), W \rangle_{\mathcal{H}_2} + U(t), \quad t \in [0, 1].$$

Since  $\beta$  is expected to be estimated at parametric rate, the control of  $\widehat{U}_i - U_i$  would depend on the conditions on the rate of convergence of  $\widehat{\xi}(\cdot, \cdot)$ , the estimate of  $\xi(\cdot, \cdot)$ . Under suitable but mild conditions, one could expect the rate of  $R_{1n}(h)$  to be of order  $n^{-1}h^{-q/2}$  times the norm of  $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$ , while the rate of  $R_{2n}(h)$  to given by the square of the norm of  $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$ . Meanwhile, the rate of  $I_n(h)$  is  $O_{\mathbb{P}}(n^{-1}h^{-q/2})$ . The required restrictions on the bandwidth to preserve the asymptotic standard normal critical values follow. Let us point out that slower rates for the norm of  $\widehat{\xi}(\cdot, \cdot) - \xi(\cdot, \cdot)$  will require faster decreases for  $h$ , and this will result in a loss of power against sequences of local alternatives.

#### 4.3.4 Bootstrap critical values

To correct the finite sample critical values let us propose a simple wild bootstrap procedure. The bootstrap sample, denoted by  $U_i^*$ ,  $1 \leq i \leq n$ , is defined as  $U_i^* = \zeta_i U_i$ ,  $1 \leq i \leq n$ , where  $\zeta_i$ ,  $1 \leq i \leq n$  are independent random variables following the two-points distribution proposed by Mammen (1993). That means  $\zeta_i = -(\sqrt{5}-1)/2$  with probability  $(\sqrt{5}+1)/(2\sqrt{5})$  and  $\zeta_i = (\sqrt{5}+1)/2$  with probability  $(\sqrt{5}-1)/(2\sqrt{5})$ . A bootstrap test statistic  $T_n^*$  is built from a bootstrap sample as was the original test statistic. When this scheme is repeated many times, the bootstrap critical value  $z_{1-\alpha, n}^*$  at level  $\alpha$  is the empirical  $(1-\alpha)$ -th quantile of the bootstrapped test statistics. The asymptotic validity of this bootstrap procedure is guaranteed by the following result. It states that the bootstrap critical values are asymptotically standard normal under the null hypothesis and under the alternatives like in section 4.3.2. The proof could be obtained by rather standard modifications of the proof of Theorem 4.2 and hence will be omitted.

**Theorem 4.4.** *Suppose that the conditions of Theorem 4.3 hold true, in particular in the case  $r_n \equiv 0$ . Then*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(T_n^* \leq x \mid U_1, Z_1, W_1, \dots, U_n, Z_n, W_n) - \mathbb{P}(T_n \leq x)| \rightarrow 0, \quad \text{in probability.}$$

#### 4.4 The error in covariates case

In this section we show that our testing procedure extends to the case where the covariates are observed with error. In some applications, the observations  $Z_i$  and  $W_i$  are not directly observed but could be estimated by some  $\widehat{Z}_i$  and  $\widehat{W}_i$  computed

from the data. To better illustrate the methodology, let us focus on the test for the effect of a functional variable. For this reason, in this section let us take  $q = 1$ ,  $Z = \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2}$  and  $W = \widetilde{W} - \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$ , where  $\widetilde{W} \in \mathcal{H}_2$  and  $\phi_1, \phi_2, \dots$  are the elements of an orthonormal basis in  $\mathcal{H}_2$ .

In functional data analysis where usually  $\mathcal{H}_2 = L^2[0, 1]$  the choice of the basis is a key point. The statistician would likely prefer a basis allowing an accurate representation of  $\widetilde{W}$  with a minimal number of basis elements. A commonly used basis is given by the eigenfunctions of the covariance operator  $\mathcal{K}$  that is defined by  $(\mathcal{K}v)(\cdot) = \int \mathcal{K}(\cdot, s)v(s)ds$ ,  $v \in L^2[0, 1]$ , where  $\widetilde{W}$  is supposed to satisfy  $\int \mathbb{E}(\widetilde{W}^2(t))dt < \infty$  and

$$\mathcal{K}(t, s) = \mathbb{E}[\{\widetilde{W}(t) - \mathbb{E}(\widetilde{W}(t))\}\{\widetilde{W}(s) - \mathbb{E}(\widetilde{W}(s))\}]$$

is supposed positive definite. Let  $\theta_1, \theta_2, \dots$  denote the eigenvalues of  $\mathcal{K}$  and let  $\mathcal{R} = \{\phi_1, \phi_2, \dots\}$  be the corresponding basis of eigenfunctions that are usually called the functional principal components (FPC). The FPC orthonormal basis provide optimal (with respect to the mean-squared error) low-dimension representations of  $\widetilde{W}$ . See, for instance, Ramsay and Silverman (2005). In most of the applications the FPC are unknown and has to be estimated from

$$(\widehat{\mathcal{K}}v)(t) = \int_{[0,1]} \widehat{\mathcal{K}}(t, s)v(s)ds, \quad t \in [0, 1],$$

where

$$\widehat{\mathcal{K}}(t, s) = n^{-1} \sum_{i=1}^n \left\{ \widetilde{W}_i(t) - n^{-1} \sum_{j=1}^n \widetilde{W}_j(t) \right\} \left\{ \widetilde{W}_i(s) - n^{-1} \sum_{j=1}^n \widetilde{W}_j(s) \right\}. \quad (4.1)$$

Let  $\widehat{\theta}_1, \widehat{\theta}_2, \dots \geq 0$  denote the eigenvalues of  $\widehat{\mathcal{K}}$  and let  $\widehat{\phi}_1, \widehat{\phi}_2, \dots$  be the corresponding basis of eigenfunctions, that is the estimated FPC. For identification purposes, we adopt the usual condition  $\langle \phi_j, \widehat{\phi}_j \rangle \geq 0, \forall j$ . Now, we can define  $\widehat{Z}_i = \langle \widetilde{W}_i, \widehat{\phi}_1 \rangle_{\mathcal{H}_2}$  and  $\widehat{W}_i = \widetilde{W}_i - \langle \widetilde{W}_i, \widehat{\phi}_1 \rangle_{\mathcal{H}_2} \widehat{\phi}_1$ , the estimates of  $Z_i$  and  $W_i$ .

Having in mind such types of situations, herein we will suppose that

$$\|\widehat{Z}_i - Z_i\| + \|\widehat{W}_i - W_i\|_{\mathcal{H}_2} \leq \Gamma_i \Delta_n, \quad 1 \leq i \leq n, \quad (4.2)$$

where  $\Gamma_i$  are independent copies of some random variable  $\Gamma$  that depend on  $X$ , and  $\Delta_n$  depend on the data but could be taken the same for all  $i$ . For  $\Delta_n$  and  $\Gamma$  we will suppose

$$\Delta_n = O_{\mathbb{P}}(n^{-1/2}) \quad \text{and} \quad \exists a > 0 \quad \text{such that} \quad \mathbb{E} \exp(a\Gamma) < \infty. \quad (4.3)$$

Clearly, alternative conditions on the rate of  $\Delta_n$  and the moments of  $\Gamma$  could be considered, resulting in alternative conditions on the bandwidths in the statements below. As it will be explained below, the conditions (4.3) are convenient for the example of  $\widehat{Z}_i$  and  $\widehat{W}_i$  obtained from estimated FPC basis.



Let us introduce some notation

$$\widehat{K}_{ij}(h) = K((\widehat{Z}_i - \widehat{Z}_j)/h), \quad \widehat{\varphi}_{ij} = \exp(-\|\widehat{W}_i - \widehat{W}_j\|_{\mathcal{A}}^2/2). \quad (4.4)$$

Let

$$\widetilde{T}_n = \frac{\widetilde{I}_n(h)}{\widetilde{v}_n(h)}.$$

where  $\widetilde{I}_n(h)$  and  $\widetilde{v}_n(h)$  are defined as  $I_n(h)$  and  $v_n(h)$  in section 4.2.2 but with  $\widehat{Z}_i$  and  $\widehat{W}_i$  instead of  $Z_i$  and  $W_i$ .

**Theorem 4.5.** *Suppose that  $q = 1$ , the Assumptions D-(a), D-(b)-(ii, iii), K-(a) are met and conditions (4.2) and (4.3) hold true. Assume one of the following conditions is met:*

1.  $nh^4/\ln^2 n \rightarrow \infty$  and  $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t) dt < \infty$  for some  $a \in (0, 1]$ ;
2.  $nh^2/\ln^2 n \rightarrow \infty$  and  $f_Z$  is bounded.

*Then the Theorems 4.2, 4.3 and 4.4 remain valid with the test statistic  $T_n$  replaced by  $\widetilde{T}_n$ .*

The proof of Theorem 4.5 is a direct consequence of Lemma 4.7 in the Appendix and is hence omitted.

Let us revisit the problem of the test for the effect of a functional variable, where  $Z = \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2}$  and  $W = \widetilde{W} - \langle \widetilde{W}, \phi_1 \rangle_{\mathcal{H}_2} \phi_1$ . The conditions on the random variable  $Z$  required in Lemma 4.7 are mild conditions satisfied in the common examples of functional covariates considered in the literature. Concerning condition (4.2), consider the operator norm  $\|\mathcal{K}\|_S$  defined by

$$\|\mathcal{K}\|_S^2 = \int \int \sigma^2(t, s) dt ds.$$

Under Assumptions D-(a), D-(b)-(ii), the empirical covariance operator satisfies

$$\|\widehat{\mathcal{K}} - \mathcal{K}\|_S = O_{\mathbb{P}}(1/\sqrt{n}),$$

see for instance Bosq (2000) or Horváth and Kokoszka (2012). On the other hand, suppose that  $\theta_1$ , the eigenvalue associated to  $\phi_1$ , is different from all the others eigenvalues of the operator  $\mathcal{K}$ . By Lemma 4.3 in Bosq (2000) or Lemma 2.3 in Horváth and Kokoszka (2012), and the fact that the spectral norm of the operator  $\widehat{\mathcal{K}} - \mathcal{K}$  is smaller or equal to  $\|\widehat{\mathcal{K}} - \mathcal{K}\|_S$ ,

$$\|\widehat{\phi}_1 - \phi_1\|^2 \leq \frac{8}{\varsigma^2} \|\widehat{\mathcal{K}} - \mathcal{K}\|_S^2,$$

where  $\varsigma$  is the distance between  $\theta_1$  and the set  $\{\theta_2, \theta_3, \dots\}$  of all the other eigenvalues of  $\mathcal{K}$ . Here the eigenvalues of  $\mathcal{K}$  are not necessarily ordered,  $\theta_1$  could be any eigenvalue separated from all the others. Deduce that condition (4.2) is guaranteed for instance if there exists  $a > 0$  such that

$$\mathbb{E} \exp(a \|\widetilde{W}\|_{\mathcal{H}_2}) < \infty.$$

The exponential moment condition is met if, for instance,  $\widetilde{W}$  is a mean-zero Gaussian process defined on the unit interval with  $\sup_{t \in [0,1]} \mathbb{E}[\widetilde{W}^2(t)] < \infty$ ; see chapter A.2 in van der Vaart and Wellner (1996). Moreover, in general, a moment restriction on  $\widetilde{W}$  is not restrictive for significance testing. Indeed, if  $\widetilde{W}$  does not satisfy such a condition, it suffices to transform  $\widetilde{W}$  into some variable  $\in \mathcal{H}_2$  such that  $\widetilde{V}$  generates the same  $\sigma$ -field and  $\widetilde{V}$  satisfies the required moment condition.

## 4.5 Empirical evidence

In this section we illustrate the empirical performances of our testing procedure. For that purpose, we consider both scalar and functional responses  $Y$ . We used an Epanechnikov kernel in our applications, that is  $K(x) = 0.75(1-x^2)\mathbf{1}\{|x| < 1\}$ . We calculated  $\varphi_{ij}$  and  $\widehat{\varphi}_{ij}$  in two ways: with the norm in the Hilbert space  $L^2[0, 1]$  of the covariate and with the norm  $\|\cdot\|_{\mathcal{H}}$  proposed in Remark 1 for the case  $\beta = 2$ .

Below  $\langle \cdot, \cdot \rangle$  is the usual inner product on  $L^2[0, 1]$ , that is  $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ . Let  $\widehat{\mathcal{K}}$  be the empirical covariance operator defined in (4.1) and let  $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \dots \geq 0$  be its eigenvalues and  $\widehat{\phi}_1, \widehat{\phi}_2, \dots$  the corresponding eigenfunctions.

### 4.5.1 The scalar response case

We simulate data samples of size  $n = 40$  using the models

$$Y_i = a + \langle X_i, b \rangle + \delta \langle X_i, b \rangle^2 + U_i, \quad (4.1)$$

$$Y_i = \left( \lambda_k^{-1} \langle X_i, e_k \rangle^2 - 1 \right) + U_i, \quad 1 \leq i \leq n, \quad (4.2)$$

where  $X_i$  is a Wiener process,  $U_i$  are independent centered normal variables with variance  $\sigma^2 = 1/16$ ,

$$a = 0 \quad \text{and} \quad b(t) = \sin^3(2\pi t^3), \quad t \in [0, 1].$$

Moreover,

$$e_k(t) = \sqrt{2} \sin((k-1/2)\pi t), \quad t \in [0, 1],$$

and  $\lambda_k = (k-1/2)^{-2}\pi^{-2}$  and  $k$  is some fixed positive integer. The null hypothesis corresponds to  $\delta = 0$  while nonnegative  $\delta$ 's yield quadratic alternatives.

We then estimate  $b$  using the functional principal component approach, see see, e.g., Ramsay and Silverman (2005) and Horváth and Kokoszka (2012). The first five principal components of the  $X_i$ s are used so that  $b$  is estimated by

$$\hat{b}(t) = \sum_{j=1}^5 \hat{b}_j \hat{\phi}_j(t),$$

where  $\hat{b}_j = \hat{\theta}_j^{-1} \hat{g}_j$ ,  $\hat{g}_j = \langle \hat{g}, \hat{\phi}_j \rangle$  with

$$\hat{g}(t) = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n) (X_i(t) - \bar{X}_n(t))$$

and  $a$  by  $\hat{a} = \bar{Y}_n - \langle \bar{X}_n, \hat{b} \rangle$ . The test statistics are built with  $q = 1$ ,

$$\hat{U}_i = Y_i - \hat{a} - \langle X_i, \hat{b} \rangle, \quad \hat{Z}_i = \frac{\tilde{Z}_i}{\sqrt{n^{-1} \sum_{j=1}^n \tilde{Z}_j^2}} \quad \text{and} \quad \hat{W}_i = \frac{\tilde{W}_i}{\sqrt{n^{-1} \sum_{j=1}^n \|\tilde{W}_j\|_{\mathcal{H}_2}^2}},$$

where  $\tilde{Z}_i = \langle X_i, \hat{\phi}_1 \rangle$  and  $\tilde{W}_i = X_i - \langle X_i, \hat{\phi}_1 \rangle \hat{\phi}_1$ .

First, we investigate the accuracy of the asymptotic critical values and the effectiveness of the bootstrap correction, with 199 bootstrap samples, for level  $\alpha = 10\%$ . Several bandwidths are considered, that is  $h = cn^{-1/5}$  with  $c \in \{2^{k/2}, k = -2, -1, 0, 1, 2\}$ . The results of 5000 replications are plotted in the left panel of Figure 4.1. The normal critical values are quite inaccurate, while the bootstrap corrections are very effective, whatever the considered bandwidth is. The differences between the results for the statistics defined with  $\|\cdot\|_{\mathcal{H}_2}$  and those for the statistics defined with  $\|\cdot\|_{\mathcal{A}}$  are imperceptible.

Next, we compare our test to the one introduced by Patilea et al. (2012a) (hereafter PSSa) based on projections. The test statistic of PSSa is

$$T_n^{PSSa} = \frac{Q_n(\hat{\gamma}_n; \hat{a}, \hat{b})}{\hat{v}_n(\hat{\gamma}_n; \hat{a}, \hat{b})}$$

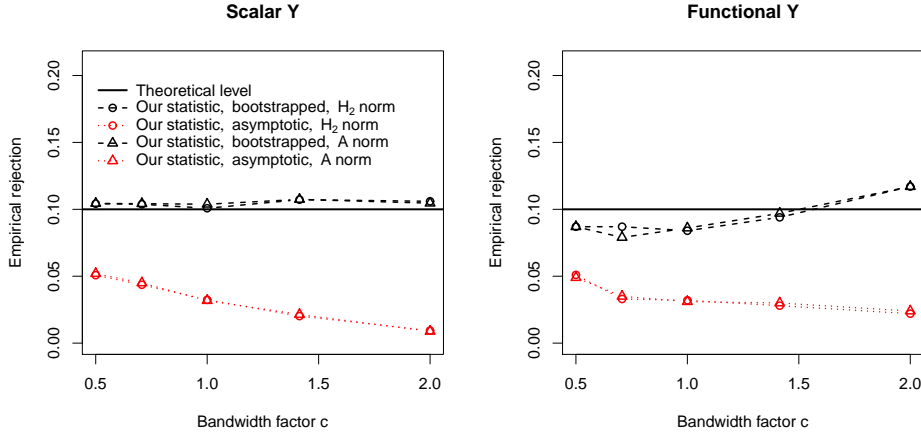
where

$$Q_n(\gamma; \hat{a}, \hat{b}) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \hat{U}_i \hat{U}_j \frac{1}{h} K(h^{-1} \langle X_i - X_j, \gamma \rangle), \quad \gamma \in \mathbb{R}^p,$$

and  $\hat{v}_n^2(\gamma; \hat{a}, \hat{b})$  is an estimation of the variance of  $nh^{1/2}Q_n(\gamma; \hat{a}, \hat{b})$ . Here and in the following, the vector  $\gamma = (\gamma_1, \dots, \gamma_p)^\top$  is identified with  $\sum_{k=1}^p \gamma_k \phi_k \in L^2[0, 1]$ . The value of  $p$  is chosen by the statistician. The direction  $\hat{\gamma}_n$  is selected as

$$\hat{\gamma}_n = \arg \max_{\gamma \in B_p} \left[ nh^{1/2}Q_n(\gamma, \hat{a}, \hat{b}) / \hat{v}_n(\gamma, \hat{a}, \hat{b}) - \alpha_n \mathbf{1} \{ \gamma \neq \gamma_0 \} \right],$$

where  $B_p \subset \mathcal{S}^p = \{ \gamma \in \mathbb{R}^p : \|\gamma\| = 1 \}$  is a set of positive Lebesgue measure on  $\mathcal{S}^p$  and  $\gamma_0$  is a privileged direction chosen by the statistician and  $\alpha_n$  is a penalty term.



**Figure 4.1:** Empirical rejection for scalar (left panel) and functional (right panel)  $Y$  under the null hypothesis.

Here we follow PSSa and we take  $p = 3$  and  $B_3$  as a set of 1200 points on  $\mathcal{S}^3$ ,  $\gamma_0 = (1, 1, 1) / \sqrt{3}$  and  $\alpha_n = 3$ .

The results are presented on Figure 4.2 the null hypothesis (5000 replications) and several alternatives (2500 replications) defined by some positive values of  $\delta$ . The PSSa statistic is computed with wild bootstrap critical values. The rejection rate for the bootstrap version of our test appears to be better than that based on asymptotic critical values for each considered alternative. Moreover, the results obtained with  $\|\cdot\|_{\mathcal{H}_2}$  are better than those obtained with  $\|\cdot\|_{\mathcal{A}}$ . The PSS1 outperforms our test in terms of power for the setups (4.1) and (4.2) with  $k = 2$ . This could be explained by the nature of the PSS1 statistic which by construction is powerful against such alternatives. When considering the setup (4.2) with  $k = 4$  the power is deteriorates drastically for all the tests. The fourth coordinate  $\langle X_i, e_4 \rangle$  being independent of the first three involved in the PSS1 statistic, the empirical power of that statistic is practically equal to the level for any sample size. The empirical power of our statistic improves with the sample size and so confirms the asymptotic results. The plateau for the empirical rejection curves for our test could be explained by an inflated variance on the alternatives, but its level increases with the sample size.

### 4.5.2 The case of functional response

Three models with functional  $Y$  are considered:

$$Y_i(t) = \delta \times \beta(t) X_i(t) + \epsilon_i(t) \quad (4.3)$$

$$Y_i(t) = \delta \times H(B_i(t)) + \epsilon_i(t) \quad (4.4)$$

$$Y_i(t) = \delta \times \lambda_k^{-1/2} \langle B_i, e_k \rangle + \epsilon_i(t) \quad (4.5)$$

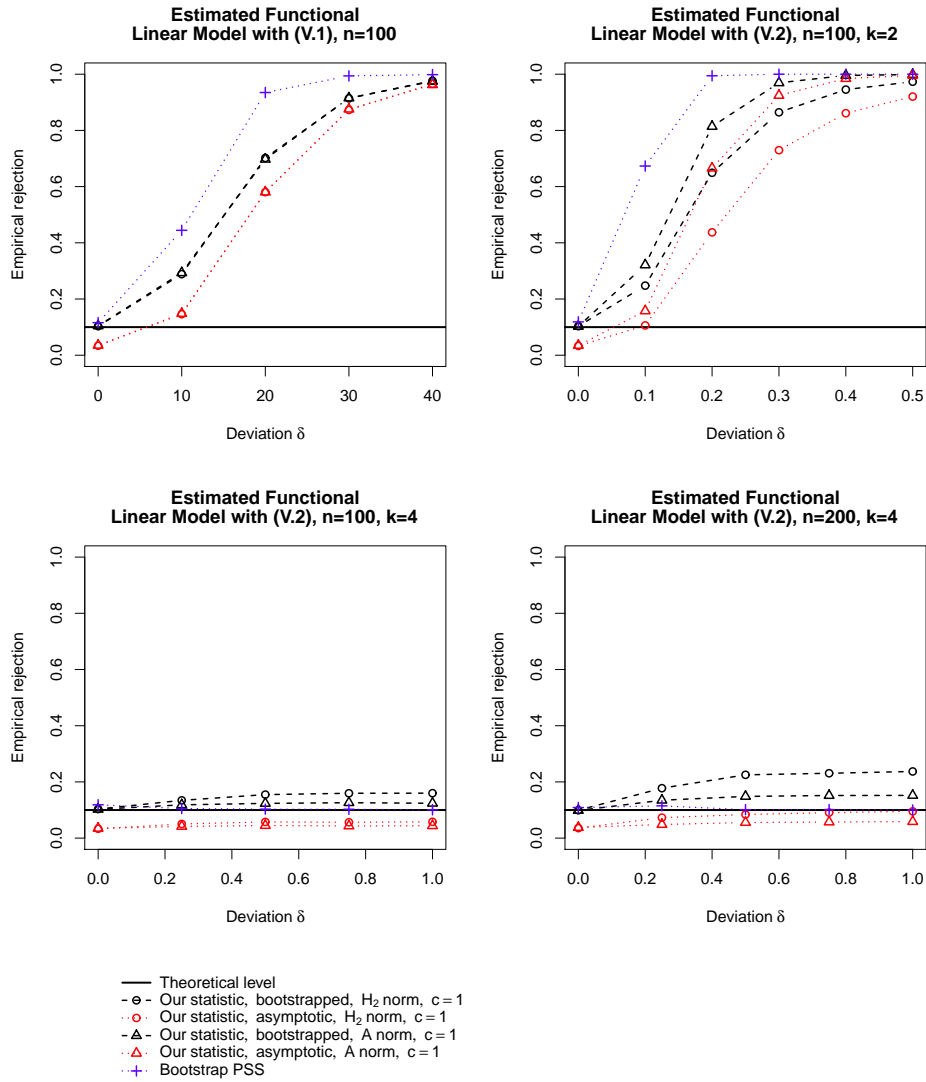


Figure 4.2: Empirical rejection for scalar  $Y$ .

$1 \leq i \leq n$ , where  $X_i$  and  $\epsilon_i$  are independent Brownian bridges,  $B_i$  is a Brownian motion,

$$\beta(t) = \exp\{-4(t - 0.3)^2\}, \quad t \in [0, 1],$$

$e_k(\cdot)$  and  $\lambda_k$  are defined as in the case of scalar response for some fixed  $k \geq 1$ , and  $H(x) = x^2 - 1$ ,  $x \in \mathbb{R}$ . We consider  $q = 1$  and the  $\widehat{Z}_i$ ,  $\widehat{W}_i$  and  $\widehat{\varphi}_{ij}$  are built like in the case of a scalar response.

We compare our test with the one considered by Patilea et al. (2012b) (hereafter PSSb). Their statistic, let us call it  $T_n^{PSSb}$ , which is a variant of  $T_n^{PSSa}$  above defined with a different  $Q_n$ . That is, in the definition of  $Q_n$  the product  $\widehat{U}_i \widehat{U}_j$  is replaced by the scalar product  $\langle \widehat{U}_i, \widehat{U}_j \rangle$  and  $K(h^{-1} \langle X_i - X_j, \gamma \rangle)$  by

$$K_h\left(h^{-1}[F_{\gamma,n}(\langle X_i, \gamma \rangle) - F_{\gamma,n}(\langle X_j, \gamma \rangle)]\right)$$

where  $F_{\gamma,n}$  is the empirical c.d.f. of the sample  $\langle X_1, \gamma \rangle, \dots, \langle X_n, \gamma \rangle$ ,  $\gamma \in B_p \subset \mathbb{R}^p$ . Following PSS2, in this case we take  $p = 3$ ,  $B_3$  as a set of 1200 points on  $\mathcal{S}^3$ ,  $\gamma_0 = (1, 1, 1)/\sqrt{3}$  and  $\alpha_n = 2$ . Moreover, since here we test for the effect,  $\widehat{U}_i$  are nothing but the observations  $Y_i$ .

We also compare our test with the test of Kokoszka et al. (2008) (hereafter KMSZ) based on the eigenvalues  $(\widehat{\gamma}_k)_k$  and  $(\widehat{\lambda}_k)_k$  and eigenvectors  $(\widehat{u}_k)_k$  and  $(\widehat{v}_k)_k$ ,  $1 \leq k \leq n$  of the respective empirical operators

$$\Gamma_n x = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle X_i, \quad \Lambda_n x = \frac{1}{n} \sum_{i=1}^n \langle Y_i, x \rangle Y_i,$$

and also

$$\Delta_n x = \frac{1}{n} \sum_{i=1}^n \langle X_i, x \rangle Y_i,$$

the test statistic being

$$T_n(\tilde{p}, \tilde{q}) = n \sum_{k=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} \frac{\langle \Delta_n \widehat{v}_k, \widehat{u}_j \rangle^2}{\widehat{\gamma}_k \widehat{\lambda}_j}.$$

This statistic is asymptotically  $\chi^2(\tilde{p}\tilde{q})$  distributed when there is no linear effect of  $X$  on  $Y$ . We test the “no effect” model on the three setups (4.3), (4.4) and (4.5) using  $\widehat{U}_i = Y_i - \bar{Y}_n$ . For this we consider the cases  $\tilde{p} = 1$ ,  $\tilde{q} = 6$  and  $\tilde{p} = 2$ ,  $\tilde{q} = 6$ .

Again, we investigate the accuracy of the asymptotic critical values and the bootstrap correction, following the same steps as in the case of scalar  $Y$ , this time for 1000 replications under the null hypothesis. We present the results in the right panel of Figure 4.1. The conclusions are similar to those of the scalar case, that is the asymptotic critical values are rather inaccurate with  $n = 40$ . The bootstrap correction is quite effective, whatever the considered bandwidth is. The empirical power results for positive deviations  $\delta$  for the three models considered are presented

in Figure 4.3. They are based on a number of 500 replications of the experiment. The results obtained with  $\|\cdot\|_{\mathcal{H}_2}$  are again preferable. One can see that KMSZ and PSS2 perform very well for the concurrent alternative. However, for a linear alternative with  $k = 4$ , the bootstrap version of our test seems to be the best choice. The good performance of the PSS2 with samples of size  $n = 40$  could be explained by a correlation between  $\langle B_i, \hat{\phi}_1 \rangle, \dots, \langle B_i, \hat{\phi}_3 \rangle$  and  $\langle B_i, \hat{\phi}_4 \rangle$  which approximates  $\langle B_i, e_4 \rangle$ . This correlation vanishes when  $n$  increases resulting in a loss of power for PSS2 test. In this experiment we also studied the effect of larger dimension  $q$  with  $n = 40$  and the concurrent alternative, equation (4.3), and quadratic alternative, equation (4.4). The results presented in Figure 4.3 reveals a drastic decrease of power. A possible explanation is that when the first components  $\langle X_i, \hat{\phi}_1 \rangle$  carry enough information on the covariate, the price to pay in terms of power for smoothing in higher dimension could be too high, so that it may be preferable to consider  $q = 1$ .

### 4.5.3 Real data application

The approach proposed in this paper is applied to check the goodness-of-fit of several models for the Canadian weather dataset. This dataset is studied in Ramsay and Silverman (2005) and is included in the R package `fda` (<http://www.r-project.org>). The data consist of the daily mean temperature and rain registered in 35 weather stations in Canada. A curve is available for each station, describing the rainfall for each day of the year. This is the functional response. The same type of curve with the temperature is used as functional predictor. Several regression models with functional covariate and functional response have been studied in Ramsay and Silverman (2005), and illustrated with the Canadian weather dataset. The purpose here is to assess the validity of each of the following three models

$$Y_{ij}(t) = \mu(t) + \varepsilon_{ij}(t), \quad (4.6)$$

$$Y_{ij}(t) = \mu(t) + \alpha_j(t) + \varepsilon_{ij}(t), \quad (4.7)$$

$$Y_{ij}(t) = \mu(t) + \alpha_j(t) + \int X_{ij}(s) \xi(s, t) ds + \varepsilon_{ij}(t), \quad (4.8)$$

where  $\sum_{j=1}^J \alpha_j(\cdot) \equiv 0$  to ensure identification of models (4.7) and (4.8). The stations are classified in four climatic zones (Atlantic, Pacific, Continental, Arctic) and  $Y_{ij}(t)$  represents the logarithm of the rainfall at the station  $i$  of the climate zone  $j$  on day  $t$ ,  $X_{ij}(t)$  is the temperature at the same station on day  $t$  of the year. Since each observation  $Y_{ij}$  is observed for the same time design, we just use

$$\hat{U}_{ij}(\cdot) = Y_{ij}(\cdot) - \bar{Y}_n^{(i)}(\cdot) \quad \text{and} \quad \hat{U}_{ij}(\cdot) = Y_{ij}(\cdot) - \bar{Y}_{\cdot j}^{(i)}(\cdot)$$

for models (4.6) and (4.7) respectively. Here we use the notation

$$\bar{A}_n^{(i)} = (n-1)^{-1} \left( -A_{ij} + \sum_{j=1}^J \sum_{k=1}^{n_j} A_{kj} \right)$$

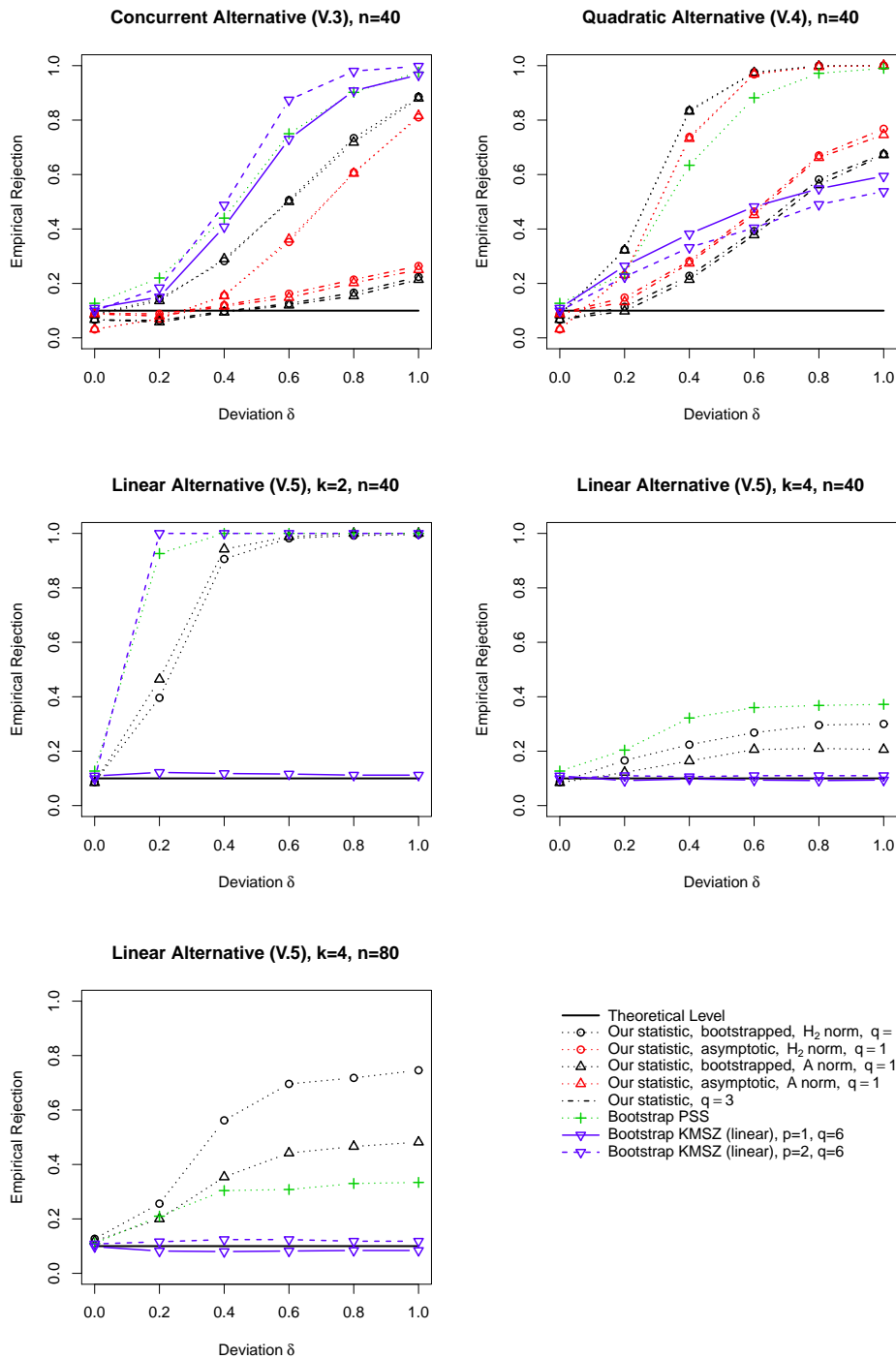


Figure 4.3: Empirical rejection for functional  $Y$ .



Model	$q = 1$	$q = 3$
(4.6)	0	0
(4.7)	0	0.009
(4.8), $K = 12$	0.016	0.403
(4.8), $K = 13$	0.023	0.736
(4.8), $K = 14$	0.015	0.723

**Table 4.1:** Bootstrap  $p$ -values for modeling Canadian Weather data three different ways and for different smoothing dimension  $q \in \{1, 2\}$  and 999 bootstrap samples.

and  $\bar{A}_{\cdot j}^{(i)} = (n_j - 1)^{-1} \sum_{k \in \{1, \dots, n_j\} \setminus \{i\}} A_{kj}$  represent respectively the leave-one out overall mean and the class  $j$  mean for the variable  $A$  and the observation  $i$ . For the model (4.8), let us notice that

$$\bar{Y}_{\cdot j}^{(i)}(t) = \mu(t) + \alpha_j(t) + \int \bar{X}_{\cdot j}^{(i)}(s) \xi(s, t) ds + \bar{\varepsilon}_{\cdot j}^{(i)}(t)$$

and then

$$\tilde{Y}_{ij}(t) = \int \tilde{X}_{ij} \xi(s, t) ds + \tilde{\varepsilon}_{ij}$$

where  $\tilde{A}_{ij} = A_{ij} - A_{\cdot j}^{(i)}$ . Thus we construct the functional principal components based on

$$\left\{ \tilde{X}_{ij}(\cdot), j \in \{1, \dots, J\}, i \in \{1, \dots, n_j\} \right\}$$

which leads to  $\tilde{X}_{ij}(\cdot) \simeq \sum_{\kappa=1}^K \lambda_{\kappa} c_{ij\kappa} v_{\kappa}(\cdot)$  (where  $\sum_{j=1}^J \sum_{i=1}^{n_j} c_{ij\kappa}^2 = 1$ ,  $\|v_{\kappa}(\cdot)\|^2 = 1$  and  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ ) and

$$\hat{U}_{ij}(t) = \tilde{Y}_{ij}(t) - \sum_{\kappa=1}^K c_{ij\kappa} \left( -c_{ij\kappa} \tilde{Y}_{ij}(t) + \sum_{l=1}^J \sum_{k=1}^{n_l} c_{kl\kappa} \tilde{Y}_{kl}(t) \right).$$

All this leave-one-out feature is used to avoid overfitting and for the choice of  $K = 13$ , we used the one that minimizes  $\sum_{j=1}^J \sum_{i=1}^{n_j} \|\hat{U}_{ij}(\cdot)\|_{\mathcal{H}_2}^2$ . We also consider the effect of this choice considering also  $K = 12$  and  $K = 14$ .

On one hand, we choose not to project the response variable before the test process, because some of the link between  $Y$  and  $X$  could be in the truncated part of  $Y$ . On the other hand, reducing the dimension for  $X$  is compulsory to solve the infinite dimension inverse problem. We consider the smoothing dimensions  $q = 1$  and  $q = 3$ , with  $h = n^{-1/(q+4)}$  for the test. Only the norm  $\|\cdot\|_{\mathcal{H}_2}$  was used for the functional covariates. Our test rejects all the models when using  $q = 1$ . Meanwhile the model (4.8) is not rejected with  $q = 3$ . This could be explained by a possible lack of power due to smoothing in higher dimension.

## 4.6 Technical results and proofs

*Proof of Lemma 4.1.* The implication from left to right is obvious. For the reverse one, let us consider  $l^2 \subset \mathbb{R}^\infty$  the space of real valued, square integrable sequences  $x = (x_1, x_2, \dots)$ , endowed with the scalar product  $\langle x, y \rangle_2 = \sum_{k=1}^\infty x_k y_k$ . Since any  $w \in \mathcal{H}_2$  can be decomposed  $w = \sum_{k \geq 1} \langle w, \phi_k \rangle_{\mathcal{H}_2} \phi_k$ , where  $\{\phi_1, \phi_2, \dots\}$  is the orthonormal basis considered in  $\mathcal{H}_2$ , we shall use the usual identification between  $\mathcal{H}_2$  and  $l^2$  given by the isomorphism  $w \in \mathcal{H}_2 \mapsto (\langle w, \phi_1 \rangle_{\mathcal{H}_2}, \langle w, \phi_2 \rangle_{\mathcal{H}_2}, \dots) \in l^2$ . Denote  $W_{12} = W_1 - W_2$ .

Next, consider the linear operator  $Q$  from  $\mathcal{H}_2$  into  $\mathcal{H}_2$  defined by

$$Q\phi_k = a_k \phi_k, \quad k \geq 1.$$

The condition that the series  $\sum_{k \geq 1} a_k$  is convergent means that the trace of the operator  $Q$  is finite. Now, since  $\mathbb{E}[\|W_{12}\|_{\mathcal{H}_2}^2] < \infty$ , there exists a set of events  $N$  such that  $\mathbb{P}(N) = 1$  and for any  $\omega \in N$ ,  $W_{12}(\omega) \in l^2$  and hence  $Q(W_{12}(\omega)) \in l^2$ . By classical results in mathematical analysis in infinite-dimensional Hilbert spaces, see for instance Theorem 1.12 in Da Prato (2006), there exists a (unique) probability measure  $\mu_Q$  on  $\mathcal{H}_2$  endowed with the Borel  $\sigma$ -field such that for any  $\omega \in N$ ,

$$\begin{aligned} \exp(-\|W_{12}(\omega)\|_{\mathcal{H}_2}^2/2) &= \exp(-\langle Q(W_{12}(\omega)), W_{12}(\omega) \rangle_{\mathcal{H}_2}/2) \\ &= \int_{\mathbb{R}^\infty} \exp(i \langle W_{12}(\omega), x \rangle_2) \mu_Q(x) \\ &= \int_{l^2} \exp(i \langle W_{12}(\omega), x \rangle_2) \mu_Q(x). \end{aligned}$$

The last equality expresses the fact that the probability measure  $\mu_Q$  concentrates on  $l^2$ . Using this identity for each  $\omega \in N$ , the inverse Fourier transform for  $h^{-q}K((Z_1 - Z_2)/h)$ , the Fubini theorem and a change of variables we can write

$$\begin{aligned} I(h) &= \int_{l^2} \int_{\mathbb{R}^q} \mathbb{E} \left[ \left\langle V_1 e^{i\{t^\top Z_1 + \langle x, W_1 \rangle_2\}}, V_2 e^{-i\{t^\top Z_2 + \langle x, W_2 \rangle_2\}} \right\rangle_{\mathcal{H}_1} \right] \mathcal{F}[K](th) dt d\mu_Q(x) \\ &= \int_{l^2} \int_{\mathbb{R}^q} \left\| \mathbb{E} \left[ V_1 e^{i\{t^\top Z_1 + \langle x, W_1 \rangle_2\}} \right] \right\|_{\mathcal{H}_1}^2 \mathcal{F}[K](th) dt d\mu_Q(x), \end{aligned}$$

where  $V_j = \mathbb{E}[U_j | Z_j, W_j] \omega(Z_j)$ ,  $j = 1, 2$ . Deduce that

$$\mathbb{E} \left[ \mathbb{E}[U | Z, W] \omega(Z) e^{i\{t^\top Z + \langle x, W \rangle_{\mathcal{H}_2}\}} \right] = 0, \quad \forall t \in \mathbb{R}^q, x \in l^2.$$

By the uniqueness of the Fourier Theorem in Hilbert spaces, see for instance Proposition 1.7 of Da Prato (2006), it follows that  $\mathbb{E}[U | Z, W] = 0$ . Now, the proof is complete.  $\square$

**Lemma 4.6.** *Suppose that Assumptions D-(a) and K are met.*

(a)

$$\sup_{t \in \mathbb{R}^q} \frac{1}{nh^q} \sum_{i=1}^n K^k((t - Z_i)/h) = O_{\mathbb{P}} \left( \sqrt{\frac{\ln n}{nh^q}} \right) + o(h^{-q/2}),$$

for  $k = 1$  or  $k = 2$ .

(b) Let  $0 < \gamma_1, \gamma_2$  i.i.d. random variables such that  $\mathbb{E}[\mathbb{E}^4(\gamma_1 \mid Z_1)] < \infty$ . Then  $\mathbb{E}[\gamma_1 \gamma_2 h^{-q} K^2((Z_1 - Z_2)/h)]$  converges to a positive constant as  $h \rightarrow 0$ .

*Proof of Lemma 4.6.* (a) We only consider the case  $k = 1$ , the case  $k = 2$  is very similar. By Theorem 2.1 of van der Vaart and Wellner (2011),

$$\sup_{t \in \mathbb{R}^q} \left| \frac{1}{n} \sum_{i=1}^n \{K((t - Z_i)/h) - \mathbb{E}[K((t - Z)/h)]\} \right| = O_{\mathbb{P}} \left( \sqrt{\frac{h^q \ln n}{n}} \right) \quad (4.1)$$

Indeed, let  $\mathcal{G}$  be a class of functions of the observations with envelope function  $G$ , that here will be supposed bounded, and let

$$J(\delta, \mathcal{G}, L^2) = \sup_Q \int_0^\delta \sqrt{1 + \ln N(\varepsilon \|G\|_{Q,2}, \mathcal{G}, L^2(Q))} d\varepsilon, \quad 0 < \delta \leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions  $Q$  on the space of the observations, and  $\|G\|_{Q,2}$  denotes the norm of  $G$  in  $L^2(Q)$ . Let  $Z_1, \dots, Z_n$  be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g(Z_i) - \mathbb{E}[g(Z)]\}, \quad \gamma \in \mathcal{G}$$

be the empirical process indexed by  $\mathcal{G}$ . If the covering number  $N(\varepsilon, \mathcal{G}, L^2(Q))$  is of polynomial order in  $1/\varepsilon$ , there exists a constant  $c > 0$  such that  $J(\delta, \mathcal{G}, L^2) \leq c\delta \sqrt{\ln(1/\delta)}$  for  $0 < \delta < 1/2$ . Now if  $\mathbb{E}g^2 < \delta^2 \mathbb{E}G^2$  for every  $\gamma$  and some  $0 < \delta < 1$ , Theorem 2.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n g| = J(\delta, \mathcal{G}, L^2) \left( 1 + \frac{J(\delta, \mathcal{G}, L^2)}{\delta^2 \sqrt{n} \|G\|_2} \right) \|G\|_2 O_p(1), \quad (4.2)$$

where  $\|G\|_2^2 = \mathbb{E}G^2$  and the  $O_p(1)$  term is independent of  $n$ . Note that the family  $\mathcal{G}$  could change with  $n$ , as soon as the envelope is the same for all  $n$ . We can thus apply this result to the family of functions  $\mathcal{G} = \{K((t - \cdot)/h) : t \in \mathbb{R}^q\}$  for a sequence  $h$  that converges to zero, the envelope  $G(\cdot) \equiv K(0)$ , and  $\delta = h^{q/2}$ . Its entropy number is of polynomial order in  $1/\varepsilon$ , independently of  $n$ , as  $K(\cdot)$  is of bounded variation. Thus the rate in (4.1) follows.

On the other hand, if  $|\mathcal{F}[f_Z](u)|^{2-a}$  is integrable for some  $a \in (0, 1]$ , by the properties of the Fourier and inverse Fourier transforms, Fubini theorem and the Cauchy-Schwarz inequality, for any  $t \in \mathbb{R}$ ,

$$\begin{aligned}
\mathbb{E}[h^{-q}K((t-Z)/h)] &= \left| (2\pi)^{-q/2} \mathbb{E} \int_{\mathbb{R}^q} \exp\{iu^\top t\} \exp\{-iu^\top Z\} \mathcal{F}[K](hu) du \right| \\
&= \left| \int_{\mathbb{R}^q} \exp\{iu^\top t\} \mathcal{F}[f_Z](u) \mathcal{F}[K](hu) du \right| \\
&\leq \left[ \int_{\mathbb{R}^q} |\mathcal{F}[f_Z](u)|^{2-a} du \right]^{\frac{1}{2-a}} \left[ \int_{\mathbb{R}^q} |\mathcal{F}[K](hu)|^{(2-a)/(1-a)} du \right]^{\frac{1-a}{2-a}} \\
&\leq C \left[ h^{-q} \int_{\mathbb{R}^q} |\mathcal{F}[K](u)| du \right]^{\frac{1-a}{2-a}} \\
&= Ch^{-q(1-a)/(2-a)} \\
&= o(h^{-q/2}), \tag{4.3}
\end{aligned}$$

for some constant  $C$  independent of  $t$ . Alternatively, if the density  $f_Z$  is bounded, by a change of variables we can write

$$\mathbb{E}[h^{-q}K((t-Z)/h)] = \int_{\mathbb{R}^q} K(u) f_Z(t-uh) du \leq C', \tag{4.4}$$

for some constant  $C'$  independent of  $t$ . From equations (4.12), (4.1), (4.3) and (4.4)

$$\sup_{t \in \mathbb{R}^q} \frac{1}{nh^q} \sum_{i=1}^n K^k((t-Z_i)/h) = O_{\mathbb{P}} \left( \sqrt{\frac{\ln n}{nh^q}} \right) + o(h^{-q/2}), \tag{4.5}$$

for  $k = 1$  or  $k = 2$ .

(b) Let  $e(z) = \mathbb{E}[\gamma_1 | Z_1 = z]$ . If  $f_Z$  satisfies the condition  $\int_{\mathbb{R}^q} |\mathcal{F}[f_Z]|^{2-a}(t) dt < \infty$  for some  $a \in (0, 1]$ , then  $\int_{\mathbb{R}^q} f_Z^2 < \infty$  and hence by Cauchy-Schwarz inequality

$$\int_{\mathbb{R}^q} f_Z^2 e^2 \leq \left( \int_{\mathbb{R}^q} f_Z^2 \right)^{1/2} \left( \int_{\mathbb{R}^q} f_Z e^4 \right)^{1/2} < \infty.$$

Using Fubini Theorem, the inverse Fourier Transform formula and Parseval identity,

we can write

$$\begin{aligned}
\mathbb{E}[\gamma_1 \gamma_2 h^{-q} K((Z_1 - Z_2)/h)] &= \mathbb{E}[h^{-q} e(Z_1) e(Z_2) K((Z_1 - Z_2)/h)] \\
&= (2\pi)^{-\frac{q}{2}} \mathbb{E} \int_{\mathbb{R}^n} e(Z_1) \exp\{iu^\top Z_1\} \\
&\quad \times e(Z_2) \exp\{-iu^\top Z_2\} \mathcal{F}[K](hu) du \\
&= \int_{\mathbb{R}^n} |\mathcal{F}[f_Z e](u)|^2 |\mathcal{F}[K]|(hu) du \\
&\rightarrow \int_{\mathbb{R}^n} |\mathcal{F}[f_Z e](u)|^2 du \\
&= \int_{\mathbb{R}^n} f_Z^2(u) e^2(u) du,
\end{aligned}$$

where for the limit we use the Dominated Convergence Theorem. If  $f_Z$  is bounded, we can use a change of variables like for equation (4.4) and again the Dominated Convergence Theorem to obtain the same strictly positive and finite limit.  $\square$

*Proof of Theorem 4.2.* The proof is based on the Central Limit Theorem 5.1 of de Jong (1987). Let

$$\Omega_{ij} = \frac{1}{n(n-1)h^q} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij}, \quad 1 \leq i \neq j \leq n,$$

and  $\Omega_{ii} = 0$ ,  $1 \leq i \leq n$ . Let  $\Omega(n) = \sum_{i \neq j} \Omega_{ij}$  and  $\sigma(n)^2 = 2 \sum_{i \neq j} \sigma_{ij}^2$  where

$$\sigma_{ij}^2 = \mathbb{E}[\Omega_{ij}^2 \mid (Z_i, W_i), (Z_j, W_j)] = \frac{V_{ij}^2 K_{ij}^2(h) \varphi_{ij}^2}{n^2(n-1)^2 h^{2q}}$$

and

$$V_{ij}^2 = \mathbb{E}[\langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)]$$

Consider the following conditions:

1. there exists a sequence of real numbers  $k_n$  such that

$$k_n^2 \sigma(n)^{-2} \max_{1 \leq i \leq n} \sum_{1 \leq j \leq n} \sigma_{ij}^2 = o_{\mathbb{P}}(1) \quad (4.6)$$

and

$$\max_{1 \leq i \neq j \leq n} \sigma_{ij}^{-2} \mathbb{E} \left[ \Omega_{ij}^2 \mathbf{1}_{\{|\Omega_{ij}| > k_n \sigma_{ij}\}} \mid (Z_i, W_i), (Z_j, W_j) \right] = o_{\mathbb{P}}(1); \quad (4.7)$$

- 2.

$$\sigma(n)^{-2} \max_{1 \leq i \leq n} \mu_i^2 = o_{\mathbb{P}}(1), \quad (4.8)$$

where  $\mu_1, \dots, \mu_n$  are the eigenvalues of the matrix  $(\sigma_{ij})$ .

If these conditions hold true, using the characterization of the convergence in probability based on almost surely convergence subsequences, Theorem 5.1 of de Jong (1987) applied conditionally on the covariates implies that for any  $t \in \mathbb{R}$ ,

$$\mathbb{P}\left(\sigma(n)^{-1}\Omega(n) \leq t \mid (Z_1, W_1), \dots, (Z_n, W_n)\right) - \Phi(t) = o_{\mathbb{P}}(1).$$

By the dominated convergence theorem,  $\sigma(n)^{-1}\Omega(n)$  converges to in law to a standard normal distribution. Hence, it remains to check conditions (4.6) to (4.8).

First, let us bound from below  $\sigma(n)$ . By Assumption D-(c)-(i),  $V_{ij}^2 \geq \underline{\sigma}^2$  almost surely, so that

$$\sigma(n)^2 \geq \frac{\underline{\sigma}^2}{n^2(n-1)^2h^{2q}} \sum_{i \neq j} K_{ij}^2(h) \varphi_{ij}^2 \geq \frac{\underline{\sigma}^2}{n^2(n-1)^2h^{2q}} \sum_{i \neq j} K_{ij}^2(h) \lambda_i^2 \lambda_j^2 = \underline{\sigma}^2 \tau(n)^2,$$

where  $\lambda_i = \exp(-2\|W_i\|_{\mathcal{A}}^2)$ . By standard calculations, the variance of  $n(n-1)h^q \tau(n)^2$  tends to zero. By Lemma 4.6-(b) the expectation of  $n(n-1)h^q \tau(n)^2$  tends to a positive constant. Deduce that

$$\sigma(n)^{-2} = O_{\mathbb{P}}(n^2 h^q). \quad (4.9)$$

Next, note that by Hölder inequality and Assumption D-(c)-(ii),

$$V_{ij}^2 \leq \mathbb{E}[\|U_i \omega(Z_i)\|_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)] \mathbb{E}[\|U_j \omega(Z_j)\|_{\mathcal{H}_1}^2 \mid (Z_i, W_i), (Z_j, W_j)] \leq C^{4/\nu},$$

almost surely. Deduce from this and Lemma 4.6-(a) that

$$\max_{1 \leq i \leq n} \sum_{1 \leq j \leq n} \sigma_{ij}^2 = o_{\mathbb{P}}(n^{-2} h^{-q}). \quad (4.10)$$

Then condition (4.6) follows from (4.9) and (4.10) for some suitable sequence  $k_n \rightarrow \infty$ .

Next, let us note that

$$\frac{\Omega_{ij}^2}{\sigma_{ij}^2} \leq \underline{\sigma}^{-2} \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}^2,$$

so that for any  $i$  and  $j$ , by Hölder inequality, Markov inequality and Assumption D-(c)-(ii),

$$\begin{aligned} \sigma_{ij}^{-2} \mathbb{E} \left[ \Omega_{ij}^2 \mathbf{1}_{\{\Omega_{ij} > k_n \sigma_{ij}\}} \mid (Z_i, W_i), (Z_j, W_j) \right] &\leq \underline{\sigma}^{-2} \mathbb{E} \left[ \langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}^2 \right. \\ &\quad \left. \times \mathbf{1}_{\{|\langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}| > \underline{\sigma} k_n\}} \mid (Z_i, W_i), (Z_j, W_j) \right] \\ &\leq \underline{\sigma}^{-2} \mathbb{E}^{2/\nu} [\|U_i \omega(Z_i)\|_{\mathcal{H}_1}^\nu \|U_j \omega(Z_j)\|_{\mathcal{H}_1}^\nu \mid (Z_i, W_i), (Z_j, W_j)] \\ &\quad \times \mathbb{P}^{(\nu-2)/\nu} [|\langle U_i \omega(Z_i), U_j \omega(Z_j) \rangle_{\mathcal{H}_1}| > \underline{\sigma} k_n \mid (Z_i, W_i), (Z_j, W_j)] \\ &\leq \underline{\sigma}^{-2} C^{4/\nu} [C^{2/\nu} (\underline{\sigma} k_n)^{-1}]^{(\nu-2)/\nu}, \end{aligned}$$

almost surely. Thus condition (4.7) holds true for any  $k_n \rightarrow \infty$ .

To check condition (4.8), let  $\mathcal{K}$  be the matrix with elements

$$\mathcal{K}_{ij} = V_{ij} \varphi_{ij} K((Z_i - Z_j)/h) / [n(n-1)h^q], \quad i \neq j, \quad \text{and} \quad \mathcal{K}_{ii} = 0, \quad (4.11)$$

and  $\|\mathcal{K}\|_2$  is the spectral norm of  $\mathcal{K}$ . By definition,  $\|\mathcal{K}\|_2 = \sup_{u \in \mathbb{R}^n, u \neq 0} \|\mathcal{K}u\| / \|u\|$  and  $|u' \mathcal{K} w| \leq \|\mathcal{K}\|_2 \|u\| \|w\|$  for any  $u, w \in \mathbb{R}^n$ . By Cauchy-Schwarz inequality, for any  $u \in \mathbb{R}^n$ ,

$$\begin{aligned} \|\mathcal{K}u\|^2 &= \sum_{i=1}^n \left( \sum_{j=1, j \neq i}^n V_{ij} \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} u_j \right)^2 \\ &\leq C^{4/\nu} \sum_{i=1}^n \left( \sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} \right) \sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q n(n-1)} u_j^2 \\ &\leq C^{4/\nu} \|u\|^2 n^{-2} \left[ \max_{1 \leq i \leq n} \left( \sum_{j=1, j \neq i}^n \frac{K((Z_i - Z_j)/nh)}{h^q (n-1)} \right) \right]^2, \end{aligned} \quad (4.12)$$

for some constant  $c > 0$ . By Lemma 4.6-(a) deduce that

$$\max_{1 \leq i \leq n} \mu_i^2 = \frac{1}{n^2} \left[ O_{\mathbb{P}} \left( \frac{\ln n}{nh^q} \right) + o(h^{-q}) \right] = o_{\mathbb{P}} \left( \frac{1}{n^2 h^q} \right).$$

Condition (4.8) follows from this and the rate (4.9). Now the proof is complete.  $\square$

*Proof of Theorem 4.3.* Let us simplify the notation and denote  $\omega_i = \omega(Z_i)$  and  $\delta_i = \delta(Z_i, W_i)$ . Next let us decompose

$$\begin{aligned} I_n(h) &= \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i^0 \omega_i, U_j^0 \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &+ \frac{r_n}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle U_i^0 \omega_i, \delta_j \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &+ \frac{r_n^2}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \langle \delta_i \omega_i, \delta_j \omega_j \rangle_{\mathcal{H}_1} K_{ij}(h) \varphi_{ij} \\ &= I_{0n} + 2I_{1n} + I_{2n}. \end{aligned}$$

The rate of  $I_{0n}$  is given by Theorem 4.2, so that it remains to investigate the rates of  $I_{1n}$  and  $I_{2n}$  and to bound in probability  $v_n^2(h)$ . By standard calculations,

$$\mathbb{E}(I_{0n}) = \mathbb{E}(I_{1n}) = 0, \quad \text{Var}(I_{0n}) = O_{\mathbb{P}}(n^{-2}h^{-q}) \quad \text{and} \quad \text{Var}(I_{1n}) = O_{\mathbb{P}}(r_n^2 n^{-1}h^{-q}).$$

Moreover,

$$\mathbb{E}(I_{2n}) = r_n^2 \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} h^{-q} K_{12}(h) \varphi_{12}] \quad \text{and} \quad \text{Var}(I_{1n}) = O_{\mathbb{P}}(r_n^4 n^{-1}h^{-q}).$$

By dominated convergence we have

$$\begin{aligned}
\mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} h^{-q} K_{12}(h) \varphi_{12}] &= \mathbb{E}[\mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \mid Z_1, Z_2] h^{-q} K_{12}(h)] \\
&= (2\pi)^{-\frac{q}{2}} \mathbb{E} \left[ \int_{\mathbb{R}^q} \mathbb{E}[\langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \mid Z_1, Z_2] \right. \\
&\quad \times \exp\{iu^\top Z_1\} \\
&\quad \times \exp\{-iu^\top Z_2\} \mathcal{F}[K](hu) du \left. \right] \\
&\rightarrow (2\pi)^{-\frac{q}{2}} \int_{\mathbb{R}^q} \mathbb{E} \left[ \langle \delta_1 \omega_1, \delta_2 \omega_2 \rangle_{\mathcal{H}_1} \varphi_{12} \right. \\
&\quad \times \exp\{iu^\top Z_1\} \exp\{-iu^\top Z_2\} \left. \right] du
\end{aligned}$$

By arguments as used in the proof of Lemma 4.1 the expectation of  $I_{2n}$  could be shown to be strictly positive. Since  $r_n^2 n h^{q/2} \rightarrow \infty$ , the result follows.  $\square$

**Lemma 4.7.** *Suppose the assumptions of Theorem 4.5 hold true. Then, for  $l = 1$  and  $l = 2$ ,*

$$\max_{1 \leq i \leq n} \frac{1}{nh} \sum_{j=1}^n \left| [\widehat{K}_{ij}(h) \widehat{\varphi}_{ij}]^l - [K_{ij}(h) \varphi_{ij}]^l \right| = o_{\mathbb{P}}(1) \quad (4.13)$$

and

$$\frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \left\{ [\widehat{K}_{ij}(h) \widehat{\varphi}_{ij}]^2 - [K_{ij}(h) \varphi_{ij}]^2 \right\} = o_{\mathbb{P}}(1). \quad (4.14)$$

*Proof of Lemma 4.7.* Assume that  $nh^4 / \ln^2 n \rightarrow \infty$ . By the Lipschitz property of the kernel and of the  $\varphi(\cdot)$  function, the bound (4.2) and conditions (4.3), for  $l = 1$  and  $l = 2$ ,

$$\begin{aligned}
\max_{i,j} h^{-1} \left| [\widehat{K}_{ij}(h) \widehat{\varphi}_{ij}]^l - [K_{ij}(h) \varphi_{ij}]^l \right| &\leq Ch^{-2} \Delta_n \max_{1 \leq i \leq n} |\Gamma_i| \\
&= O_{\mathbb{P}}(n^{-1/2} h^{-2} \ln n) = o_{\mathbb{P}}(1).
\end{aligned}$$

The rates (4.13) and (4.14) follow.

If conditions at point (2) of Theorem 4.5 are met, the arguments are of different nature. Since  $K$  is bounded, it suffices to consider the case  $l = 1$ . First, note that the conditions of point (3) involve that  $f_Z$  is bounded. Next, since the kernel  $K$  is of bounded univariate kernels, let  $K_1$  and  $K_2$  be non decreasing bounded functions such that  $K = K_1 - K_2$  and denote  $K_{1h} = K_1(\cdot/h)$ . Clearly, it is sufficient to prove (4.13) for  $K_1$ , similar arguments apply for  $K_2$  and hence we get the results for  $K$ . For simpler writings let us assume that  $K$  is differentiable and let  $K_1(x) =$



$\int_{-\infty}^x [K'(t)]^+ dt = \int_{-\infty}^{\min\{x,0\}} [K'(t)]^+ dt$ , and  $K_2(x) = K_1(x) - K(x) = \int_{-\infty}^x [K'(t)]^- dt = \int_0^{\max\{x,0\}} [K'(t)]^- dt$ ,  $x \in \mathbb{R}$ . Here  $[K']^+$  (resp.  $[K']^-$ ) denotes the positive (resp. negative) part of  $K'$ . The general case where the kernel is differentiable except for a finite set of points can be handled with obvious modifications. Let  $b_n \downarrow 0$  such that  $b_n \sqrt{n} / \ln n \rightarrow \infty$  and define the event

$$\mathcal{E}_{1n} = \left\{ \max_{1 \leq i \leq n} [\|\widehat{Z}_i - Z_i\| + \|\widehat{W}_i - W_i\|_{\mathcal{H}_2}] \leq b_n \right\}, \quad (4.15)$$

so that  $\mathbb{P}(\mathcal{E}_{1n}^c) \rightarrow 0$ . Since  $|\exp(-t^2/2) - \exp(-s^2/2)| \leq |t - s|$ , on the set  $\mathcal{E}_{1n}$ ,  $\forall i, j$

$$\begin{aligned} & -b_n K_{1h}(Z_i - Z_j + 2b_n) - [K_{1h}(Z_i - Z_j - 2b_n) - K_{1h}(Z_i - Z_j)] \varphi_{ij} \\ & \leq \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) - K_{1h}(Z_i - Z_j) \right| \varphi_{ij} - K_{1h}(\widehat{Z}_i - \widehat{Z}_j) |\varphi_{ij} - \widehat{\varphi}_{ij}| \\ & \leq \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) \widehat{\varphi}_{ij} - K_{1h}(Z_i - Z_j) \varphi_{ij} \right|. \end{aligned}$$

Similarly

$$\begin{aligned} & \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) \widehat{\varphi}_{ij} - K_{1h}(Z_i - Z_j) \varphi_{ij} \right| \\ & \leq b_n K_{1h}(Z_i - Z_j + 2b_n) + [K_{1h}(Z_i - Z_j + 2b_n) - K_{1h}(Z_i - Z_j)] \varphi_{ij}. \end{aligned}$$

We focus on the second inequality, the first one can be handled similarly. Moreover, since  $b_n/h \rightarrow 0$ , we only have to bound  $[K_{1h}(Z_i - Z_j + 2b_n) - K_{1h}(Z_i - Z_j)] \varphi_{ij}$ . To justify (4.13) we can write

$$\begin{aligned} & \frac{1}{nh} \sum_{j=1}^n \left| K_{1h}(\widehat{Z}_i - \widehat{Z}_j) \widehat{\varphi}_{ij} - K_{1h}(Z_i - Z_j) \varphi_{ij} \right| \\ & \leq \frac{b_n}{h} \mathbb{E}[K_{1h}(Z_i - Z + 2b_n) \mid Z_i] \\ & \quad + \frac{b_n}{nh} \sum_{j=1}^n [K_{1h}(Z_i - Z_j + 2b_n) - \mathbb{E}[K_{1h}(Z_i - Z_j + 2b_n) \mid Z_i]] \\ & \quad + \frac{1}{nh} \sum_{j=1}^n \{K_{1h}(Z_i - Z_j + 2b_n) \varphi_{ij} - \mathbb{E}[K_{1h}(Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i]\} \\ & \quad - \frac{1}{nh} \sum_{j=1}^n \{K_{1h}(Z_i - Z_j) \varphi_{ij} - \mathbb{E}[K_{1h}(Z_i - Z_j) \varphi_{ij} \mid Z_i, W_i]\} \\ & \quad + \mathbb{E}[h^{-1} K_{1h}(Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i] - \mathbb{E}[h^{-1} K_{1h}(Z_i - Z_j) \varphi_{ij} \mid Z_i, W_i] \\ & = A_{0n,i} + A_{1n,i} + A_{2n,i} - A_{3n,i} + A_{4n,i}. \end{aligned}$$

Since  $f_Z$  is supposed bounded, by a simple change of variable, we get

$$\max_{1 \leq i \leq n} A_{0n,i} = O(b_n) = o(1).$$

The terms  $A_{1n,i}$ ,  $A_{2n,i}$  and  $A_{3n,i}$  could be treated similarly, hence we only investigate  $A_{2n,i}$ . First note that, since  $\varphi_{ij} \leq 1$ , the function  $K_1$  is bounded and integrable and  $f_Z$  is bounded, for all  $j$  we have

$$\text{Var}(K_{1h}(Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i) \leq Ch, \quad \forall i,$$

for some constant  $C$  independent of  $n$  and  $Z_i, W_i$ . Using this conditional variance bound and applying Bernstein inequality<sup>1</sup> conditionally on the  $Z_i, W_i$ 's, for any  $t > 0$

$$\begin{aligned} & \mathbb{P} \left[ \max_{1 \leq i \leq n} \left| \frac{1}{n} \sum_{j=1}^n \{K_{1h}(Z_i - Z_j + 2b_n) - \mathbb{E}[K_{1h}(Z_i - Z_j + 2b_n) \mid X_i, W_i]\} \right| > th \right] \\ & \leq \sum_{i=1}^n \mathbb{E} \left[ \mathbb{P} \left[ \left| \frac{1}{n} \sum_{j=1}^n K_{1h}(Z_i - Z_j + 2b_n) \right. \right. \right. \\ & \quad \left. \left. \left. - \mathbb{E}[K_{1h}(Z_i - Z_j + 2b_n) \mid X_i, W_i] \right| > th \mid X_i, W_i \right] \right] \\ & \leq 2n \exp \left( -\frac{t^2}{2} \frac{nh^2}{Ch + thM/3} \right) \leq 2 \exp \left( [\ln n] \left[ 1 - \frac{t^2}{2} \frac{nh/\ln n}{C + tM/3} \right] \right) \rightarrow 0, \end{aligned}$$

since  $nh/\ln n \rightarrow \infty$  under the conditions of point (2) or those of point (3) (here  $M$  is any constant that bounds  $K_1$ ). Deduce that  $\max_{1 \leq i \leq n} A_{2n,i} = o_{\mathbb{P}}(1)$ .

To complete the proof of (4.13) it remains to investigate the convergence of  $A_{4n,i}$  uniformly with respect to  $i$ . First, since  $\varphi_{ij} \leq 1$ ,

$$\begin{aligned} & \left| \mathbb{E}[h^{-1}K_{1h}(Z_i - Z_j + 2b_n) \varphi_{ij} \mid Z_i, W_i] - \mathbb{E}[h^{-1}K_{1h}(Z_i - Z_j) \varphi_{ij} \mid Z_i, W_i] \right| \\ & \leq \mathbb{E} \left[ \left| h^{-1}K_{1h}(Z_i - Z_j + 2b_n) - h^{-1}K_{1h}(Z_i - Z_j) \right| \mid Z_i \right]. \end{aligned}$$

If the conditions of point (2) are met, suppose that the sequence  $b_n$  used for the definition of the set  $\mathcal{E}_{1n}$  in equation (4.15) is such that  $b_n/h \rightarrow 0$  and  $b_n\sqrt{n}/\ln n \rightarrow \infty$ . By obvious changes of variables, using the uniform bound for  $f_Z$ , the sign and

<sup>1</sup>Recall that Bernstein inequality states that if  $W_1, \dots, W_n$  are i.i.d. centered random variables of variance  $\sigma^2$  taking values in the interval  $[-M, M]$ , then for any  $s > 0$

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n W_i \right| > s \right) \leq 2 \exp \left( -\frac{ns^2}{2[\sigma^2 + Ms/3]} \right).$$

the integrability of  $K'$ , for each  $1 \leq i \leq n$ ,

$$\begin{aligned}
& \mathbb{E} \left[ \left| h^{-1} K_{1h}(Z_i - Z_j + 2b_n) - h^{-1} K_{1h}(Z_i - Z_j) \right| \mid Z_i \right] \\
&= \int_{\mathbb{R}} [K_1(u + 2b_n h^{-1}) - K_1(u)] f_Z(Z_i - uh) du \\
&\leq \int_{-\infty}^0 \left[ \int_{-2b_n h^{-1}}^0 [K'(u + 2b_n h^{-1} + v)]^+ dv \right] f_Z(Z_i - uh) du \\
&\leq C \int_{-2b_n h^{-1}}^0 \left[ \int_{-\infty}^0 [K'(u + 2b_n h^{-1} + v)]^+ du \right] dv \\
&\leq C \int_{-2b_n h^{-1}}^0 \left[ \int_{-\infty}^0 |K'(u)| du \right] dv \rightarrow 0. \quad (4.16)
\end{aligned}$$

In the case of  $K_2$ , the inequalities in the last display become

$$\begin{aligned}
& \mathbb{E} \left[ \left| h^{-1} K_{2h}(Z_i - Z_j + 2b_n) - h^{-1} K_{2h}(Z_i - Z_j) \right| \mid Z_i \right] \\
&= \int_{\mathbb{R}} [K_2(u) - K_2(u - 2b_n h^{-1})] f_Z(Z_i + 2b_n h^{-1} - uh) du \\
&\leq \int_0^{\infty} \left[ \int_0^{2b_n h^{-1}} [K'(u - 2b_n h^{-1} + v)]^- dv \right] f_Z(Z_i + 2b_n h^{-1} - uh) du \\
&\leq C \int_0^{2b_n h^{-1}} \left[ \int_0^{\infty} [K'(u - 2b_n h^{-1} + v)]^- du \right] dv \\
&\leq C \int_0^{2b_n h^{-1}} \left[ \int_0^{\infty} |K'(u)| du \right] dv \rightarrow 0.
\end{aligned}$$

Now the arguments are complete for justifying (4.13). The rate in (4.14) could be easily derived from the rate in (4.13).  $\square$

## **5 Test d'adéquation pour les modèles "single-index"**



## Abstract

Semiparametric single-index assumptions are convenient and widely used dimension reduction approaches that represent a compromise between the parametric and fully nonparametric models for regressions or conditional laws. In a mean regression setup, the SIM assumption means that the conditional expectation of the response given the vector of covariates is the same as the conditional expectation of the response given a scalar projection of the covariate vector. In a conditional distribution modeling, under the SIM assumption the conditional law of a response given the covariate vector coincides with the conditional law given a linear combination of the covariates. Several estimation techniques for single-index models are available and commonly used in applications. However, the problem of testing the goodness-of-fit seems less explored and the existing proposals still have some major drawbacks. In this paper, a novel kernel-based approach for testing SIM assumptions is introduced. The covariate vector needs not have a density and only the index estimated under the SIM assumption is used in kernel smoothing. Hence the effect of high-dimensional covariates is mitigated while asymptotic normality of the test statistic is obtained. Irrespective of the fixed dimension of the covariate vector, the new test detects local alternatives approaching the null hypothesis slower than  $n^{-1/2}h^{-1/4}$ , where  $h$  is the bandwidth used to build the test statistic and  $n$  is the sample size. A wild bootstrap procedure is proposed for finite sample corrections of the asymptotic critical values. The small sample performances of our test compared to existing procedures are illustrated through simulations.

**Keywords.** Single-index regression, conditional law, lack-of-fit test, kernel smoothing,  $U$ -statistics

## 5.1 Introduction

Semiparametric single index models (SIM) are widely used tools for statistical modeling. The paradigm of such models is based on the assumption that the information contained in a vector of conditioning random variables is equivalent, in some sense, to the information contained in some index, that is usually a linear combination of the vector components. This assumption underlies most of the statistical parametric models including covariates, but allows for more general semiparametric modeling. The most common semiparametric SIM are those for the mean regression. See Powell et al. (1989), Ichimura (1993), Härdle et al. (1993), see also Horowitz (2009) for a recent review. In such models, the index and the conditional mean given the index are unknown. SIM for quantile regression were considered recently, see Kong and Xia (2012). A more restrictive, but still of significant interest, class of models is obtained by imposing the single-index paradigm to the conditional distribution of response variable given a vector of covariates. In these cases the index and the conditional law of the response given the index are unknown. The famous Cox proportional hazard model, see Cox (1972), is a particular case of SIM for conditional laws. See Delecroix et al. (2003), Hall and Yao (2005), Chiang and Huang (2012) for more general situations.

The large amount of interest for SIM could be explained by the fact that the single-index assumption is very often the first intermediate step from a parametric framework towards a fully nonparametric paradigm. Then an important question is whether this dimension reduction compromise is good enough to capture the relevant information contained in the covariate vector. A possible way to answer is to build a statistical test of the single-index assumption against general alternatives. Several tests of the goodness-of-fit of single-index mean regression models have proposed in the literature. See Fan and Li (1996), Xia et al. (2004), Stute and Zhu (2005), Chen and Van Keilegom (2009), Escanciano and Song (2010) and the references therein. The problem of testing SIM models for conditional distribution in full generality seems open.

In this paper we propose a new and quite simple kernel smoothing-based approach for testing single-index assumptions. We focus on mean regression and conditional law models. The approach is inspired by the remark that, up to some error in covariates, the single-index assumption check could be interpreted as a test of significance in nonparametric regression. Next, the single-index assumption could be conveniently reformulated as an equivalent unconditional moment condition. Finally, a kernel based test statistic could be used to test the unconditional moment condition. The smoothing based goodness-of-fit test approach allows to make the error in covariates negligible and thus to obtain a pivotal asymptotic law under the null hypothesis. The covariate vector needs not have a density, discrete covariables are allowed. Only the index estimated under the SIM assumption is used in kernel smoothing and this fact mitigates the effect of high-dimensional covariates. Meanwhile the asymptotical critical values are given by the quantiles of the normal law. Irrespective of the fixed

dimension of the covariate vector, the new test detects local alternatives approaching the null hypothesis slower than  $n^{-1/2}h^{-1/4}$ , where  $h$  is the bandwidth used to build the test statistic and  $n$  is the sample size.

The paper is organized as follows. In Section 5.2, we recall general considerations on single-index models. In Section 5.3, we present a general approach of testing nonparametric significance and in Section 5.4 we apply it to single-index hypotheses for mean regression as well as for conditional law. In Section 5.5 we introduce a wild bootstrap procedure to correct the asymptotic critical values with small samples and illustrate the performance of our test by an empirical study. Technical results and proofs are relegated to the appendix.

## 5.2 Single-index models

Let  $Y \in \mathbb{R}^d$ ,  $d \geq 1$ , denote the random response vector and let  $X \in \mathbb{R}^p$ ,  $p \geq 1$ , be the random column vector of covariates. The data consists of independent copies of  $(Y', X')'$ . For mean regression the single-index assumption means that there exists a column parameter vector  $\beta_0 \in \mathbb{R}^p$  such that

$$\mathbb{E}[Y | X] = \mathbb{E}[Y | X'\beta_0]. \quad (5.1)$$

Only the direction given by  $\beta_0$  is identified, so that an additional identification condition accompanies the model assumption, as for instance  $\|\beta_0\| = 1$  and an arbitrary component is set positive, or an arbitrary component is set to 1. The scalar product  $X'\beta_0$  is the so-called index. The direction  $\beta_0$  and the nonparametric univariate regression  $\mathbb{E}[Y | X'\beta_0]$  have to be estimated. See Hristache et al. (2001), Delecroix et al. (2006), Horowitz (2009), Xia et al. (2011) and the references therein for a panorama of the existing estimation procedures.

When applying the single-index paradigm to conditional laws of  $Y$  given  $X$ , one supposes

$$Y \perp X | X'\beta_0. \quad (5.2)$$

In this case the direction defined by  $\beta_0$  and the conditional law of the response  $Y$  given the index  $X'\beta_0$  have to be estimated. See Delecroix et al. (2003), Hall and Yao (2005) and Chiang and Huang (2012) for the available estimation approaches.

There are several model check approaches for SIM for mean regressions. Xia et al. (2004) use an empirical process-based statistic related to that of Stute et al. (1998a). Fan and Li (1996) use a kernel smoothing-based quadratic form to a wide range of situations, including single-index. Our test statistics are somehow close to that of Fan and Li (1996). Chen and Van Keilegom (2009) use an empirical likelihood test for multi-dimensional  $Y$  in a parametric or semiparametric modeling, the single-index mean regression is presented as a particular case but without getting into the details.



In this paper we propose an alternative model check approach that is able to detect any departure from the single-index assumption, both for mean regressions and conditional law models. It is inspired by a general approach for testing nonparametric significance that is presented in the following section.

### 5.3 A general approach for testing nonparametric significance

Let  $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$  be a Hilbert space. The examples we have in mind corresponds to  $\mathcal{H} = \mathbb{R}^d$ , for some  $d \geq 1$ , or  $\mathcal{H} = L^2[0, 1]$ . Consider  $U \in \mathcal{H}$ ,  $Z \in \mathbb{R}^q$  et  $W \in \mathbb{R}^r$  and let  $(U_i, Z_i, W_i)$ ,  $1 \leq i \leq n$  denote an independent sample of  $U$ ,  $Z$  and  $W$ . Consider the problem of testing the equality

$$\mathbb{E}[U \mid Z, W] = 0 \quad \text{p.s.} \quad (5.3)$$

against the nonparametric alternative  $\mathbb{P}(\mathbb{E}[U \mid Z, W] = 0) < 1$ . Several testing procedures against nonparametric alternatives, including the single-index assumptions check, lead to this type of problem.

Let us introduce some notation: for any real-valued, univariate or multivariate function  $l$ , let  $\mathcal{F}[l]$  denote the Fourier Transform of  $l$ . Let  $K$  be a multivariate kernel  $\mathbb{R}^q$  such that  $\mathcal{F}[K] > 0$  and let  $\phi(s) = \exp(-\|s\|^2/2)$ ,  $\forall s \in \mathbb{R}^r$ . The kernel  $K$  could be a multiplicative kernel with univariate kernels with positive Fourier Transform. Many univariate kernels have this property: gaussian, triangle, Student, logistic, etc.

Our approach is based on the following remark; see also Lavergne et al. (2014). Let  $w(\cdot) > 0$  be some weight function. For any  $h > 0$ , let

$$\begin{aligned} I(h) &= \mathbb{E} \left[ \langle U_1, U_2 \rangle_{\mathcal{H}} w(Z_1)w(Z_2)h^{-q}K((Z_1 - Z_2)/h)\phi(W_1 - W_2) \right] \\ &= \mathbb{E} \left[ \langle U_1, U_2 \rangle_{\mathcal{H}} w(Z_1)w(Z_2) \right. \\ &\quad \times \left. \int_{\mathbb{R}^q} e^{2\pi i v'(Z_1 - Z_2)} \mathcal{F}[K](vh)dv \int_{\mathbb{R}^r} e^{2\pi i s'(W_1 - W_2)} \mathcal{F}[\phi](s)ds \right] \\ &= \int_{\mathbb{R}^q} \int_{\mathbb{R}^r} \left\| \mathbb{E} \left[ \mathbb{E}[U \mid Z, W]w(Z)e^{-i\{v'Z+s'W\}} \right] \right\|_{\mathcal{H}}^2 \mathcal{F}[K](vh)\mathcal{F}[\phi](s)dt ds. \end{aligned}$$

Since  $\mathcal{F}[\phi], \mathcal{F}[K] > 0$ , and  $w(\cdot) > 0$ , the following equivalence holds true:  $\forall h > 0$ ,

$$\mathbb{E}[U \mid Z, W] = 0 \quad \text{p.s.} \quad \Leftrightarrow \quad I(h) = 0.$$

To check condition (5.3) the idea is to build a sample based approximation of  $I(h)$ , to suitably normalize it and to let  $h$  to decrease to zero. A convenient choice of  $w(\cdot)$  could avoid handling denominators close to zero.

In many situations the sample of the variable  $Uw(Z)$  is not observed and has to be estimated inside the model. Then, an estimate of  $I(h)$  is given by the  $U$ -statistic

$$I_n(h) = \frac{1}{n(n-1)h^q} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i w(Z_i)}, \widehat{U_j w(Z_j)} \right\rangle_{\mathcal{H}} K_{ij}(h) \phi_{ij},$$

where

$$K_{ij}(h) = K((Z_i - Z_j)/h), \quad \phi_{ij} = \exp(-\|W_i - W_j\|^2/2).$$

The variance of  $I_n(h)$  could be estimated by

$$v_n^2(h) = \frac{2}{n^2(n-1)^2 h^{2q}} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i w(Z_i)}, \widehat{U_j w(Z_j)} \right\rangle_{\mathcal{H}}^2 K_{ij}^2(h) \phi_{ij}^2.$$

Then the test statistic is

$$T_n = \frac{I_n(h)}{v_n(h)}.$$

Under mild technical conditions and provided that  $h$  converges to zero at a suitable rate,  $T_n$  converges in law to a standard normal distribution provided that condition (5.3) holds true. Hence, a one-sided test with standard normal critical values could be defined; see Lavergne et al. (2014). One could also show  $T_n$  tends to infinity in probability if  $\mathbb{P}(\mathbb{E}[U | Z, W] = 0) < 1$ . Making  $h$  to decrease to zero at suitable rate allows to render negligible the effect of the errors  $\widehat{U_i w(Z_i)} - U_i w(Z_i)$ . On the other hand, the test detects Pitman alternative hypotheses like

$$H_{1n} : \mathbb{E}(U | Z, W) = r_n \delta(Z, W), \quad n \geq 1, \quad (5.4)$$

as soon as  $r_n^2 n h^{q/2} \rightarrow \infty$ .

## 5.4 Single-index assumptions checks

In this section we extend the approach described in section (5.3) to test single-index assumptions like (5.1) and (5.2). In this case, with the notation from section 5.3,

$$q = 1, \quad r = p - 1, \quad Z = Z(\beta) \quad \text{and} \quad W = W(\beta)$$

where, for  $\beta \in \mathcal{B} \subset \mathbb{R}^p$ ,

$$Z(\beta) = X' \beta \quad \text{et} \quad W(\beta) = X' \mathbf{A}(\beta)$$

with  $\mathbf{A}(\beta)$  a  $p \times (p-1)$  matrix with real entries such that the  $p \times p$  matrix  $(\beta \ \mathbf{A}(\beta))$  is orthogonal. The orthogonality is not necessary, invertibility suffices, but orthogonality is expected to lead to better finite sample properties for the tests.

An additional challenge will come from the fact that the sample of the covariates  $Z$  and  $W$  depend on estimator of the single-index direction  $\beta_0$ . Again, the kernel smoothing and a suitable choice of  $h$  allows to render this effect negligible and preserve a pivotal asymptotic law under the null hypothesis.

### 5.4.1 Testing SIM for mean regression

To simplify the presentation, let us focus on the case of a univariate response, that is  $d = 1$ . At the end, it will be quite clear how the case  $d > 1$  could be handled. To restate the single-index condition (5.1), let  $\mathcal{H} = \mathbb{R}$ ,  $Uw(Z) = U(\beta_0)w(Z; \beta_0)$  where

$$U(\beta)w(Z; \beta) = \{Y - \mathbb{E}[Y \mid Z(\beta)]\}f_\beta(Z(\beta)).$$

Here  $f_\beta(\cdot)$  denotes the density of  $X'\beta$  that is supposed to exist, at least for some  $\beta$ . Let

$$\widehat{U_i w(Z_i)}(\beta) = \frac{1}{n-1} \sum_{k \neq i} (Y_i - Y_k) \frac{1}{g} L_{ik}(\beta, g), \quad (5.5)$$

where  $L$  is a univariate kernel,  $L_{ik}(\beta, g) = L((Z_i(\beta) - Z_k(\beta))/g)$  and  $g$  is a bandwidth converging to zero at some suitable rate described in a following section. Let  $\hat{\beta}$  be some estimator of the index direction and consider

$$I_n^{\{m\}}(\hat{\beta}) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \widehat{U_i w(Z_i)}(\hat{\beta}) \widehat{U_j w(Z_j)}(\hat{\beta}) K_{ij}(\hat{\beta}, h) \phi(W_i(\hat{\beta}) - W_j(\hat{\beta})),$$

where  $K_{ij}(\hat{\beta}, h) = K((Z_i(\hat{\beta}) - Z_j(\hat{\beta}))/h)$ . The variance of  $I_n^{\{m\}}(\hat{\beta})$  could be estimated by

$$\hat{\omega}_n^{\{m\}}(\hat{\beta})^2 = \frac{2}{n^2(n-1)^2 h^2} \sum_{1 \leq i \neq j \leq n} \left[ \widehat{U_i w(Z_i)}(\hat{\beta}) \widehat{U_j w(Z_j)}(\hat{\beta}) \right]^2 K_{ij}^2(\hat{\beta}, h) \phi^2(W_i(\hat{\beta}) - W_j(\hat{\beta})).$$

The test statistic is then

$$T_n^{\{m\}}(\hat{\beta}) = \frac{I_n^{\{m\}}(\hat{\beta})}{\hat{\omega}_n^{\{m\}}(\hat{\beta})}.$$

Let us point out that only smoothing with the  $X_i'\hat{\beta}$ 's is required in order to build this statistic.

In section 5.4.3 we show that whenever  $\hat{\beta} - \beta^* = O_{\mathbb{P}}(n^{-1/2})$ , for some  $\beta^*$  that could depend on  $n$ ,

$$I_n^{\{m\}}(\hat{\beta}) - I_n^{\{m\}}(\beta^*) = o_{\mathbb{P}}(I_n^{\{m\}}(\beta^*)) \quad \text{and} \quad \hat{\omega}_n^{\{m\}}(\hat{\beta}) - \hat{\omega}_n^{\{m\}}(\beta^*) = o_{\mathbb{P}}(\hat{\omega}_n^{\{m\}}(\beta^*)), \quad (5.6)$$

provided some mild technical conditions hold true. Under the null hypothesis (5.1) one expects to have  $\beta^* = \beta_0$ . Then  $T_n^{\{m\}}(\hat{\beta})$  has an asymptotic standard normal law under the single-index assumption as soon as  $T_n^{\{m\}}(\beta_0)$  is standard normal asymptotically distributed. Sufficient conditions for guaranteeing the asymptotic normality of  $T_n^{\{m\}}(\beta_0)$  when (5.1) holds true are provided in Lavergne et al. (2014).

When the SIM (5.1) is wrong, even asymptotically, in general a semiparametric estimator  $\hat{\beta}$  converges at the rate  $O_{\mathbb{P}}(n^{-1/2})$  to some *pseudo-true* value  $\beta^* \in \mathcal{B}$  that depends on the estimation procedure; see Delecroix et al. (1999) for some general theoretical results. Then the asymptotic equivalence (5.6) and the results of Lavergne et al. (2014) imply that a test based on  $T_n^{\{m\}}(\hat{\beta})$  would reject the null hypothesis with probability tending to 1, in just the way the test based on  $T_n^{\{m\}}(\beta^*)$  would do. The case of Pitman alternatives requires a longer investigation since the conclusion depends on the estimation method and the properties of the deviation from the null hypothesis. Such a detailed investigation is beyond our present scope. Let us, however, briefly describe what would happen in the case where the index  $\beta_0$  was estimated through a semiparametric least-squares procedure as introduced by Ichimura (1993). Let  $r_{\beta}(s) = \mathbb{E}[Y \mid X'\beta = s]$  and

$$\nabla_{\beta} \mathbb{E}(Y \mid X'\beta_0) = \left. \frac{\partial}{\partial \beta} r_{\beta}(X'\beta) \right|_{\beta=\beta_0}.$$

Let  $\delta(X)$  satisfy  $\mathbb{E}[\delta(X) \mid X'\beta_0] = 0$  and  $\mathbb{E}[\delta(X) \nabla_{\beta} \mathbb{E}(Y \mid X'\beta_0) \tau(X)] = 0$  where  $\tau(\cdot)$  is a trimming function required in theory to keep the denominators appearing in kernel smoothing away from zero. See, for instance, Delecroix et al. (2006) for detailed discussion on the role of the trimming. Consider the sequence of alternatives

$$\mathbb{E}(Y \mid X) = \mathbb{E}(Y \mid X'\beta_0) + r_n \delta(X), \quad n \geq 1,$$

with  $r_n \rightarrow 0$ . Then it can be proved that  $\hat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$ , and hence  $T_n^{\{m\}}(\hat{\beta})$  allows to detect such local alternatives as soon as  $r_n^2 n h^{1/2} \rightarrow \infty$ .

### 5.4.2 Testing SIM for the conditional law

In order to test the single-index condition (5.2) for the conditional law of an univariate  $Y$  given  $X$ , let  $\mathcal{H} = L^2[0, 1]$  et

$$U(t; \beta) w(Z; \beta) = \{\mathbf{1}\{\Phi(Y) \leq t\} - \mathbb{P}[\Phi(Y) \leq t \mid Z(\beta)]\} f_{\beta}(Z(\beta)), \quad t \in [0, 1], \beta \in \mathcal{B},$$

where  $\Phi$  is some distribution function on the real line, for instance a normal distribution function or the marginal distribution function of  $Y$ . In the latter case, in general the distribution is unknown but could be estimated by the empirical distribution function. The case of multivariate  $Y$  could be also considered after obvious modifications and for the sake of simplicity will not be investigated herein.

Let

$$\widehat{U_i w(Z_i)}(\beta)(t) = \frac{1}{n-1} \sum_{k \neq i} (\mathbf{1}\{\Phi(Y_i) \leq t\} - \mathbf{1}\{\Phi(Y_k) \leq t\}) \frac{1}{g} L_{ik}(\beta, g), \quad t \in [0, 1]. \quad (5.7)$$

Let  $\tilde{\beta}$  be some estimator of  $\beta_0$  and consider

$$I_n^{\{l\}}(\tilde{\beta}) = \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i w}(Z_i)(\tilde{\beta}), \widehat{U_j w}(Z_j)(\tilde{\beta}) \right\rangle_{L^2} K_{ij}(\tilde{\beta}, h) \phi(W_i(\tilde{\beta}) - W_j(\tilde{\beta})),$$

where for any  $u(\cdot)$  and  $v(\cdot)$  squared integrable functions defined on the unit interval,

$$\langle u, v \rangle_{L^2} = \int_0^1 u(t)v(t)dt.$$

The variance of  $I_n^{\{l\}}(\tilde{\beta})$  could be estimated by

$$\hat{\omega}_n^{\{l\}}(\tilde{\beta})^2 = \frac{2}{n^2(n-1)^2 h^2} \sum_{1 \leq i \neq j \leq n} \left\langle \widehat{U_i w}(Z_i)(\tilde{\beta}), \widehat{U_j w}(Z_j)(\tilde{\beta}) \right\rangle_{L^2}^2 K_{ij}^2(\tilde{\beta}, h) \phi^2(W_i(\tilde{\beta}) - W_j(\tilde{\beta})). \quad (5.8)$$

The test statistic is then

$$T_n^{\{l\}}(\tilde{\beta}) = \frac{I_n^{\{l\}}(\tilde{\beta})}{\hat{\omega}_n^{\{l\}}(\tilde{\beta})}.$$

In section 5.4.3 we show that, under suitable technical conditions, whenever  $\tilde{\beta} - \beta^\# = O_{\mathbb{P}}(n^{-1/2})$ ,

$$I_n^{\{l\}}(\tilde{\beta}) - I_n^{\{l\}}(\beta^\#) = o_{\mathbb{P}}(I_n^{\{l\}}(\beta^\#)) \quad \text{and} \quad \hat{\omega}_n^{\{l\}}(\tilde{\beta}) - \hat{\omega}_n^{\{l\}}(\beta^\#) = o_{\mathbb{P}}(\hat{\omega}_n^{\{l\}}(\beta^\#)). \quad (5.9)$$

Under the null hypothesis (5.2) one expects to have  $\beta^\# = \beta_0$ . Then the asymptotic normality of  $T_n^{\{l\}}(\beta_0)$ , proved in Proposition 5.2 below, implies that the asymptotic one-sided test based on  $T_n^{\{l\}}(\tilde{\beta})$  has standard normal critical values.

If the single-index assumption fails and the alternative is fixed, like in the case of mean regression, one expects  $\tilde{\beta} - \beta^* = O_{\mathbb{P}}(n^{-1/2})$  for some *pseudo-true* value  $\beta^* \in \mathcal{B}$  that depends on the estimation procedure. Then  $T_n^{\{l\}}(\tilde{\beta})$  would detect the alternative with probability tending to 1. Concerning the case of local alternatives, let  $\delta(X, t)$  and  $r_n \rightarrow 0$  such that

$$\mathbb{P}[\Phi(Y) \leq t \mid X] = \mathbb{P}[\Phi(Y) \leq t \mid X' \beta_0] + r_n \delta(X, t), \quad t \in [0, 1],$$

is a conditional distribution function. Suitable orthogonality conditions for the function  $\delta(X, t)$  would yield  $\tilde{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$  and hence  $T_n^{\{l\}}(\tilde{\beta})$  allows to detect such local alternatives as soon as  $r_n^2 n h^{1/2} \rightarrow \infty$ .

### 5.4.3 Asymptotic results

In this section we formally state the results that guarantee the asymptotic equivalences (5.6) and (5.9). Let  $\widehat{U_i w}(Z_i)(\beta)$  be defined as in (5.5) or (5.7). Let  $I_n(\beta)$  (resp.  $\hat{\omega}_n(\beta)^2$ ) denote any of  $I_n^{\{m\}}(\beta)$  or  $I_n^{\{l\}}(\beta)$  (resp.  $\hat{\omega}_n^{\{m\}}(\beta)^2$  or  $\hat{\omega}_n^{\{l\}}(\beta)^2$ ).

**Proposition 5.1.** *Suppose the conditions in Assumption 5.3 are met. If  $\beta_n$  is an estimator such that  $\beta_n - \bar{\beta} = O_{\mathbb{P}}(n^{-1/2})$ , then*

$$I_n(\beta_n) - I_n(\bar{\beta}) = o_{\mathbb{P}}(I_n(\bar{\beta})) \quad \text{and} \quad \hat{\omega}_n(\beta_n) - \hat{\omega}_n(\bar{\beta}) = o_{\mathbb{P}}(\hat{\omega}_n(\bar{\beta})).$$

As mentioned above, the asymptotic behavior of  $I_n(\bar{\beta})$  in the case of mean regression was investigated by Lavergne et al. (2014). The case where  $U_i w(Z_i)(\beta)$  is a stochastic process seems less explored and is hence considered in the following proposition. Let  $\hat{\omega}_n(\beta_0)$  be a variance estimator defined as in equation (5.8) with  $\bar{\beta}$  replaced by  $\beta_0$ .

**Proposition 5.2.** *Suppose the conditions in Assumption 5.3 are met and the null hypothesis (5.2) holds true. Then  $nh^{1/2}I_n^{\{\cup\}}(\beta_0)/\hat{\omega}_n^{\{\cup\}}(\beta_0) \rightarrow \mathcal{N}(0, 1)$  in law under  $H_0$ , and*

$$\begin{aligned} \hat{\omega}_n^{\{\cup\}}(\beta_0) \rightarrow \omega^2(\beta_0) &= 2 \int K^2(u) du \times \int \int \Gamma^2(s, t) ds dt \\ &\times \mathbb{E} \left[ \int f_{\beta_0}^4(z) \phi^2(W_1(\beta_0) - W_2(\beta_0)) \pi_{\beta_0}(z | W_1(\beta_0)) \pi_{\beta_0}(z | W_2(\beta_0)) dz \right], \end{aligned}$$

where  $\pi_{\beta_0}(\cdot, w)$  is the conditional density of  $Z(\beta_0)$  knowing that  $W(\beta_0) = w$ , and

$$\Gamma(s, t) = \mathbb{E}[\epsilon(s)\epsilon(t)] \quad t, s \in [0, 1],$$

and  $\epsilon(t) = \mathbf{1}\{\Phi(Y) \leq t\} - \mathbb{P}[\Phi(Y) \leq t | X'\beta_0]$ .

## 5.5 Empirical evidence

For conditional mean, we simulate our data using the following model

$$Y_i = X_i'\beta + 4 \exp\{-(X_i'\beta)^2\} + \delta \sqrt{X_i'X_i} + \sigma \varepsilon_i \quad (5.10)$$

where  $X_i = (X_{i1}, \dots, X_{ip})'$  follows a standard normal  $p$ -variate law, the parameter  $\beta_0 = (1, 1, 0, \dots, 0)'$  and  $\sigma = 0.3$ . For  $\varepsilon_i$ , we consider two cases: a standard univariate normal law independent of the  $X_i$ 's and a centered log-normal heteroscedastic setup

$$\varepsilon_i = (\log \mathcal{N}(0, 1) - \sqrt{e}) \times \sqrt{(1 + X_{i2}^2)/2}.$$

The model (5.10) was proposed by Xia et al. (2004) and investigated only in the case of a homoscedastic noise.

To estimate the parameter  $\beta$ , we consider the approach of Delecroix et al. (2006), that is

$$\tilde{\beta} = \arg \min_{\beta: \beta_1 > 0} \sum_{i=1}^n \left( Y_i - \frac{\sum_{k \neq i} Y_k \tilde{L}_{ik}(\beta)}{\sum_{k \neq i} \tilde{L}_{ik}(\beta)} \right)^2, \quad (5.11)$$

where

$$\tilde{L}_{ik}(\beta) = L\left((\tilde{X}_i - \tilde{X}_k)' \beta\right), \quad \tilde{X}_i = \frac{X_i}{\sqrt{n^{-1} \sum_{k=1}^n (X_k - \bar{X})^2}} \quad \text{and} \quad \bar{X} = n^{-1} \sum_{k=1}^n X_k.$$

Then the estimator is defined as  $\hat{\beta} = \tilde{\beta} / \|\tilde{\beta}\|$  and the bandwidth  $g$  is equal to  $\|\tilde{\beta}\|^{-1}$ .

To improve the asymptotic critical values with small samples, we propose the following bootstrap procedure:

(i) Define

$$\hat{m}_i = \frac{\sum_{k \neq i} Y_k \tilde{L}_{ik}(\tilde{\beta})}{\sum_{k \neq i} \tilde{L}_{ik}(\tilde{\beta})}.$$

(ii) For  $b \in \{1, \dots, B\}$

(a) let  $Y_i^{*,b} = \hat{m}_i + \eta_i (Y_i - \hat{m}_i)$ , where the  $\eta_i$ s are independent variables with the two-point distribution

$$\mathbb{P}[\eta_i = (1 - \sqrt{5})/2] = (5 + \sqrt{5})/10, \quad \mathbb{P}[\eta_i = (1 + \sqrt{5})/2] = (5 - \sqrt{5})/10.$$

(b) define

$$\tilde{\beta}^{*,b} = \arg \min_{\beta: \beta_1 > 0} \sum_{i=1}^n \left( Y_i^* - \frac{\sum_{k \neq i} Y_k^* \tilde{L}_{ik}(\beta)}{\sum_{k \neq i} \tilde{L}_{ik}(\beta)} \right)^2$$

$$\text{and } \hat{\beta}^{*,b} = \tilde{\beta}^{*,b} / \|\tilde{\beta}^{*,b}\| \quad \text{and } g^{*,b} = \|\tilde{\beta}^{*,b}\|^{-1}.$$

(iii) Define  $T_n^{\{m\},*,b}$  as  $T_n^{\{m\}}$  where the  $Y_i$ s are replaced by the  $Y_i^{*,b}$ s,  $\hat{\beta}$  by  $\hat{\beta}^{*,b}$ , and the bandwidth  $g$  by  $g^{*,b}$ . The bandwidth  $h$  does not change. Repeat Step (iii)  $B$  times. Compute the empirical quantiles of  $T_n^{\{m\},*,b}$  using the  $B$  bootstrap values.

In our experiments the bootstrap correction is used with  $B = 499$  bootstrap samples. The level is fixed as  $\alpha = 10\%$ . We considered  $L(\cdot) = K(\cdot)$  and equal to the standard gaussian density. With this choice no numerical problem occurred due to denominators too close to zero and therefore we did not consider any trimming in equation (5.11) and its bootstrap version.

First, we investigated the influence of the bandwidth  $h$  on the level. Several bandwidths were considered, that is  $h = c \times n^{-2/9}$  with  $c \in \{2^{k/2} : k = -2, -1, 0, 1, 2\}$ . The results on empirical rejection rates for the model defined in equation (5.10) with  $\delta = 0$  (that is on the null hypothesis) and  $n = 100$  are presented in Figure 5.1. The results are based on 500 replications, with homoscedastic noise and  $p = 2$ ,  $p = 4$ , and with heteroscedastic log-normal noise and  $p = 4$ . The normal critical values are quite inaccurate, while the bootstrap correction seems to overreject slightly,

particularly for a large bandwidth  $h$ . For the third case with heteroscedastic noise, the test rejects too often. However, for larger sample sizes, this drawback is mitigated, as could be seen from the fourth plot in Figure 5.1 where we considered the heteroscedastic noise with  $p = 4$  and  $n = 200$ .

Next, we studied the behavior of our statistic under the null hypothesis (500 replications) and several alternatives (250 replications) defined by some positive value of  $\delta$ . We only considered the statistics with bandwidth factor  $c = 1$  and compared it to the statistic introduced by Xia et al. (2004). The results are presented in Figure (5.2). Xia et al. (2004)'s test performs better for  $p = 2$ , while our test shows better performance for  $p = 4$ . It appears that the greater  $p$  is, the more advantageous it will be to use our test statistic.

For conditional law, we simulate our data using the following mixture model

$$Y = (1 - \delta)\mathcal{N}(X'\beta, 0.09) + \delta\mathcal{N}(\|X\|, 0.09) \quad (5.12)$$

where  $X_i = (X_{i1}, X_{i2})'$  has a standard normal bivariate law and  $\beta_0 = (1, 1)'/\sqrt{2}$ . We apply the test statistic  $I_n^{\{l\}}$  based on the quantities  $\widehat{U_i w(Z_i)}(\beta)(t)$  introduced in (5.7). Here the events  $\{\Phi(Y_i) \leq t\}$  are defined with  $\Phi(\cdot)$  equal to the empirical distribution function of the  $Y_i$ 's. In this case an event  $\{\Phi(Y_i) \leq t\}$  is determined by the rank of  $Y_i$  in the sample of the response variable. To estimate the index parameter  $\beta$  we use the approach of Delecroix et al. (2003) that we adapt to our particular choice of  $\Phi(\cdot)$ . More precisely, let

$$(\tilde{\beta}, \tilde{g}_Y) = \arg \min_{(\beta, g_Y): \beta_1 > 0} \sum_{i=1}^n \log \frac{\sum_{j \neq i} g_Y^{-1} L((R_i - R_j)/(ng_Y)) \tilde{L}_{ij}(\beta)}{\sum_{j \neq i} \tilde{L}_{ij}(\beta)},$$

where  $R_i \in \{1, \dots, n\}$  is the index of  $Y_i$  in the order statistics  $\{Y_{(1)}, \dots, Y_{(n)}\}$ , that is  $Y_{(R_i)} = Y_i$ ,  $1 \leq i \leq n$ . The aim is to estimate  $\beta$  and  $g$  simultaneously, using again  $\hat{\beta} = \tilde{\beta}/\|\tilde{\beta}\|$  and the bandwidth  $g = \|\tilde{\beta}\|^{-1}$ .

For this test statistic, the bootstrap procedure considered is:

- (i) For  $b \in \{1, \dots, B\}$  let  $\widehat{U_i w(Z_i)}^*(\beta)(t) = \eta_i \times \widehat{U_i w(Z_i)}(\beta)(t)$  where the  $\eta_i$ 's are independent variables with the two-point distribution defined above.
- (ii) Define  $T_n^{\{l\}*b}$  as  $T_n^{\{l\}}$  where the  $\widehat{U_i w(Z_i)}(\beta)(t)$ 's are replaced by their bootstrap counterparts  $\widehat{U_i w(Z_i)}^*(\beta)(t)$ 's. Repeat Step (i)  $B$  times. Compute the empirical quantiles of  $T_n^{\{l\}*b}$ .

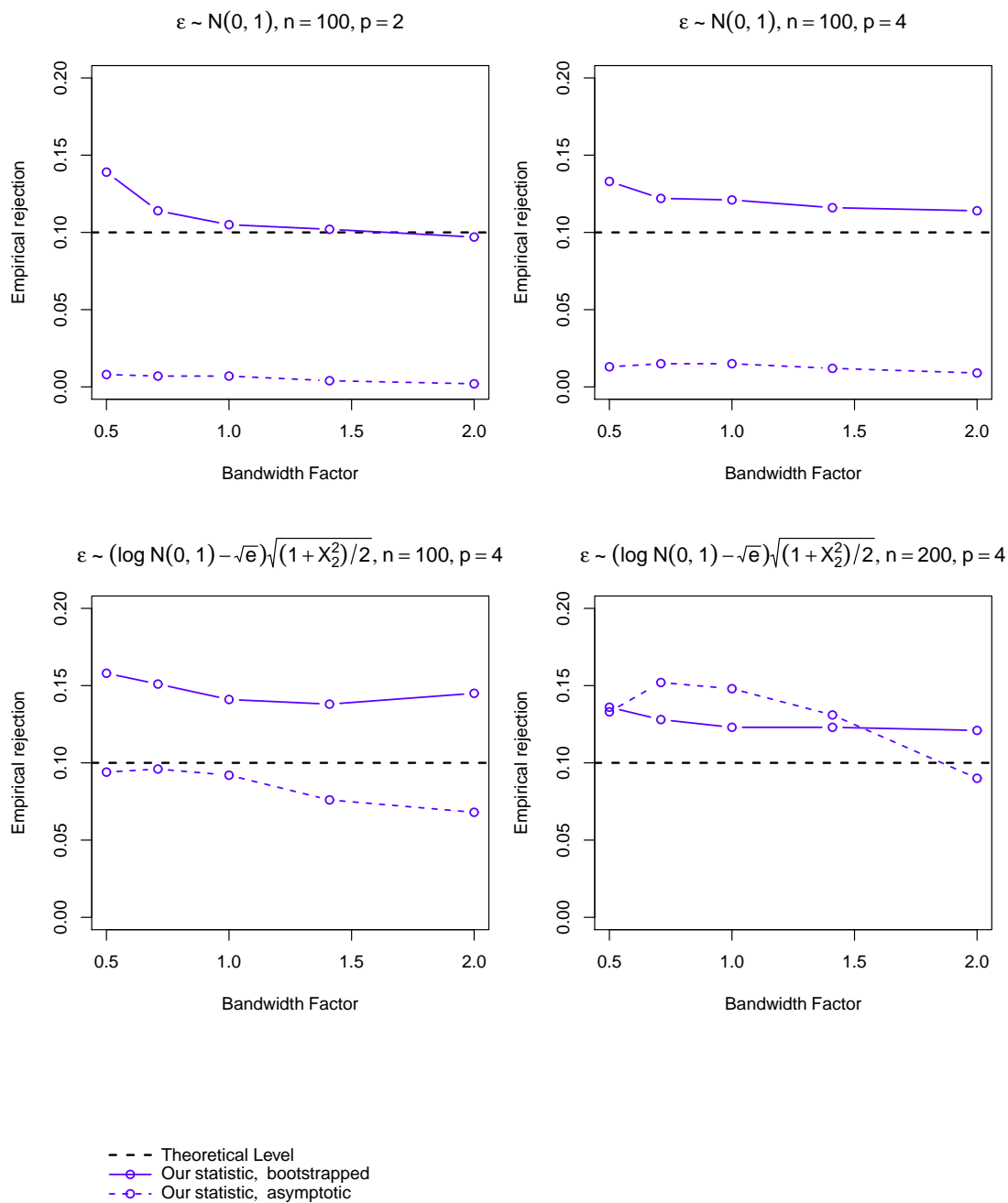
We study the influence of bandwidth  $h$  on empirical rejection under  $H_0$  on the left part of Figure 5.3, where  $h = c \times n^{-2/9}$  with  $c \in \{2^{k/2} : k = -2, -1, 0, 1, 2\}$ , with 1000 replications and 199 bootstrap steps. Because  $\beta$  is not reestimated in the bootstrap procedure, we do not correct the estimation bias and the two rejection rate are very similar. However, they are not far from the theoretical level.



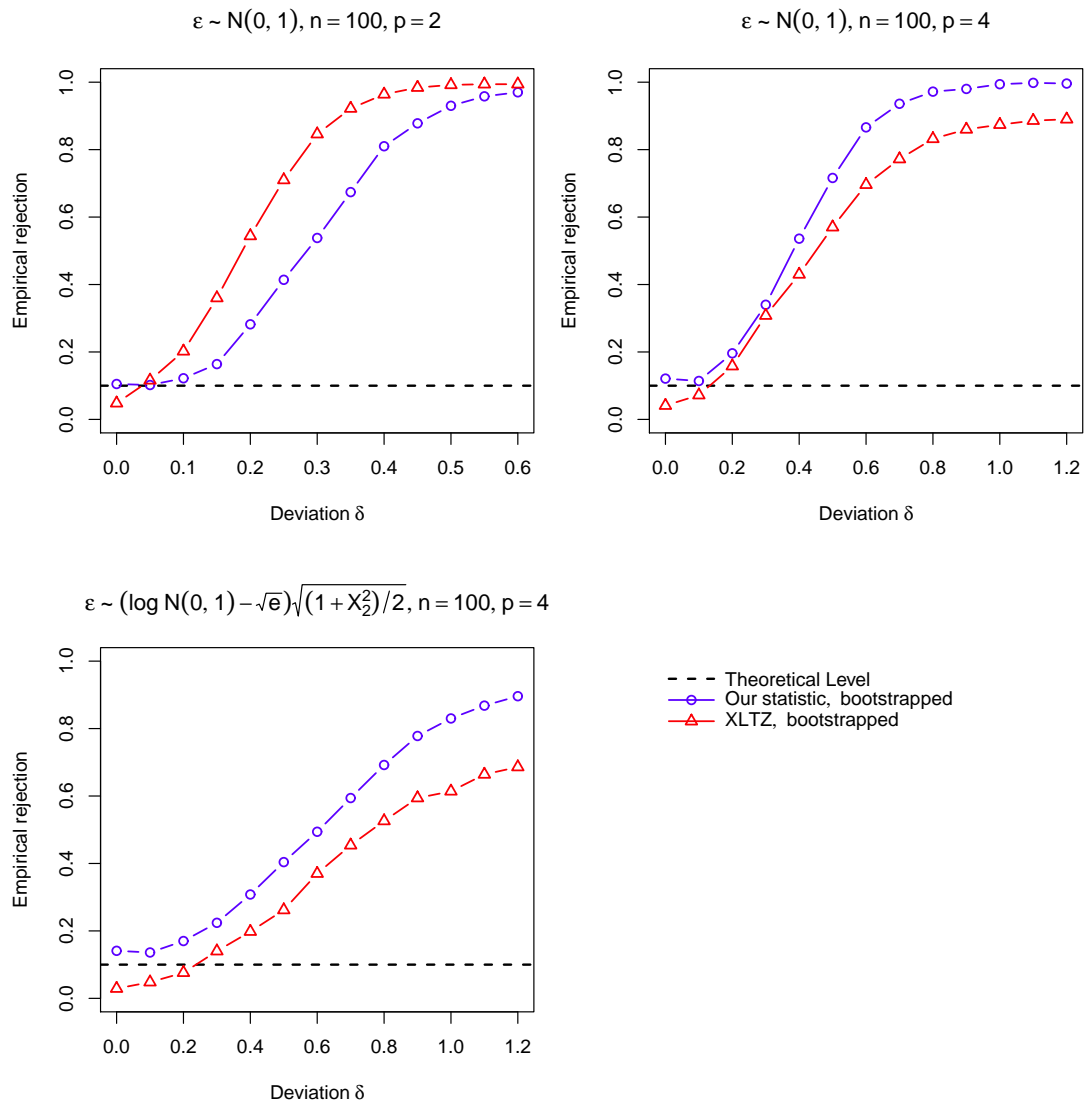
We also investigate the empirical rejection rate for different values of the mixture proportion  $\delta$  in the model (5.12). The results are presented in the right panel of Figure 5.3. We used 1000 replications for  $\delta = 0$ , 500 replications otherwise, and 199 bootstrap steps. For  $\delta$  from 0.1 to 0.2, the empirical rejection rate decreases, but it resumes its rise after  $\delta = 0.3$ . On the basis of the simulation results, we could explain this through the estimate of the variance of  $I_n^{(l)}(\tilde{\beta})$ . A small deviation  $\delta > 0$  in the model (5.12) induces less variance for  $Y$  and for the estimator of  $\beta_0$ . As a consequence, the tail of the estimator of the variance of  $I_n^{(l)}(\tilde{\beta})$  is lighter and this produces more power in the case  $\delta = 0.1$ . When the deviation  $\delta$  slightly increases beyond  $\delta = 0.1$ , the variance of  $I_n^{(l)}(\tilde{\beta})$  becomes too important and locally we observe a loss of power. When  $\delta$  increases more, the increase of the variance of  $I_n^{(l)}(\tilde{\beta})$  is dominated by the increase  $I_n^{(l)}(\tilde{\beta})$  and the test has more power.

## 5.6 Conclusions and further extensions

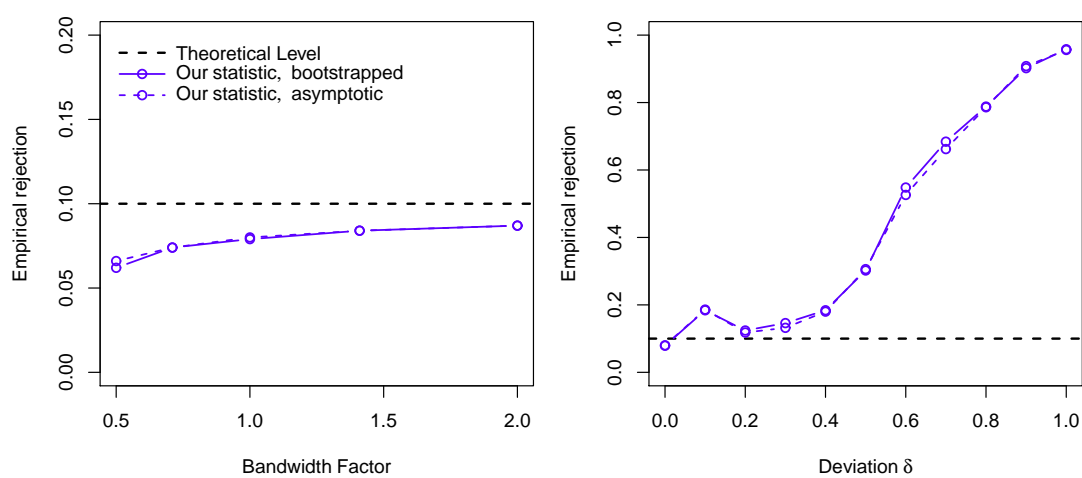
We have constructed new smoothing-based test procedures for SIM hypotheses for mean regression and for conditional law. Smoothing is only used on the estimated index, and the corresponding test statistics are asymptotically standard normal. A quite effective wild bootstrap procedure allows to correct the critical values with small samples. For simplicity we focused on univariate responses but, with obvious adjustments, our approach also applies to the case of multivariate responses. See Picone and Butler (2000) and Chen and Van Keilegom (2009) for more general situations with multivariate responses where our test methodology applies. Moreover, our statistics directly generalize to test multiple index against fully nonparametric alternatives. It suffices to consider the general methodology presented in section 5.3 with  $q$  equal to the number of indices. Some other possible extensions that would require additional, though quite straightforward, investigation are the goodness-of-fit checks of index quantile regressions, see Kong and Xia (2012), and the functional index models, see Chen et al. (2011). Such extensions are left for future work.



**Figure 5.1:** Empirical rejections under  $H_0$  as a function of the bandwidth



**Figure 5.2:** Power curves for model (5.10),  $n = 100$



**Figure 5.3:** Empirical rejections under  $H_0$  and  $H_1$  for conditional law,  $n = 200$ . On the left part,  $h = c \times n^{-2/9}$  with varying  $c$ . On the right part,  $Y = (1 - \delta)\mathcal{N}(X'\beta, 0.09) + \delta\mathcal{N}(\|X\|, 0.09)$ .

## 5.7 Appendix 1: assumptions and proofs

Let  $\mathcal{H}$  be the real line or the Hilbert space of squared integrable functions defined on  $[0, 1]$ . Let  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  and  $\|\cdot\|_{\mathcal{H}}$  denote the associated inner product and norm. For an observation  $(Y_i, X_i)'$ ,  $Y_i \in \mathbb{R}$  and  $X_i \in \mathbb{R}^p$ , let  $Y_i(t) \equiv Y_i$  or  $Y_i(t) = \mathbf{1}\{Y_i \leq \Phi^{-1}(t)\}$ , and for any  $\beta$  in the parameter set  $\mathcal{B} \subset \mathbb{R}^p$ , let  $r_i(t; \beta) = \mathbb{E}[Y_i(t) | Z_i(\beta)]$ ,  $t \in [0, 1]$ . Thus,  $Y_i(\cdot)$  is an element of  $\mathcal{H}$ .

Let  $(\epsilon_i(\cdot), X_i)'$ ,  $1 \leq i \leq n$ , be random variables such that  $\epsilon_i(\cdot) \in \mathcal{H}$  and  $X_i \in \mathbb{R}^p$ . Let  $\bar{\beta}$  be some element in the parameter set  $\mathcal{B}$ . Consider  $r_i(t; \bar{\beta})$  that depends only on  $Z_i(\bar{\beta}) = X_i' \bar{\beta}$  and  $\delta(X_i, t)$  be such that  $\mathbb{E}[\delta(X_i, t) | Z_i(\bar{\beta})] = 0$ ,  $t \in [0, 1]$ . Define

$$Y_{ni}(t) = r_i(t; \bar{\beta}) + r_n \delta(X_i, t) + \epsilon_i(t), \quad t \in [0, 1], 1 \leq i \leq n,$$

where  $r_n$ ,  $n \geq 1$ , is some bounded sequence of real numbers. In particular that means  $\mathbb{E}[Y_{ni}(\cdot) | Z_i(\bar{\beta})] = r_i(\cdot; \bar{\beta})$ . The case of a null sequence  $(r_n)$  corresponds to the null hypothesis, while a sequence tending to zero corresponds to Pitman alternatives.

**Assumption 5.3.** a) The random variables  $(\epsilon_i(\cdot), X_i)'$ ,  $1 \leq i \leq n$ , are independent copies of  $\epsilon(\cdot) \in \mathcal{H}$  and  $X \in \mathbb{R}^p$ . Moreover,  $X' \bar{\beta}$  admits a bounded density  $f_{\bar{\beta}}$ .

b)  $\mathbb{E}[\exp(\rho \|X_i\|)] < \infty$  for some  $\rho > 0$  and  $\mathbb{E}[\sup_t |r_i(t; \bar{\beta}) + \epsilon_i(t)|^a] < \infty$  for some  $a > 8$ . Moreover,  $\mathbb{E}[\|\epsilon_i(\cdot)\|_{\mathcal{H}}^2 | X_i]$  is bounded.

c) For any  $t \in [0, 1]$ , the map  $v \mapsto \mathbb{E}[Y_{ni}(t) | Z_i(\bar{\beta}) = v]$  is twice differentiable. The second derivative  $r_i''(\cdot; \bar{\beta})$  is uniformly Lipschitz (that is the Lipschitz constant independent of  $t$ ) and uniformly bounded, while the first derivative satisfies  $\mathbb{E}[\|r_i'(\cdot; \bar{\beta})\|_{\mathcal{H}}^4] < \infty$ .

d) The function  $f_{\bar{\beta}}(\cdot)$  is uniformly Lipschitz.

e) The function  $\delta(\cdot, \cdot)$  is bounded.

f) The kernels  $K$  and  $L$  are symmetric integrable functions, differentiable except at most a finite set of points and  $L'$  is Lipschitz continuous. Moreover,  $\int_{\mathbb{R}} |L(t)| dt = \int |K(t)| dt = 1$  and  $\int_{\mathbb{R}} (|L'(t)| + |K'(t)|) dt < \infty$ . The map  $v \mapsto |L'(v)|/v$  is bounded in a neighborhood of the origin,  $v^2 K(v) \rightarrow 0$  if  $v \rightarrow \infty$ , and  $\int v^2 \{|L(v)| + |K(v)|\} dv < \infty$ . Moreover, the Fourier Transform  $\mathcal{F}[K]$  is positive on the real line.

g) The bandwidths satisfy the conditions  $g, h \rightarrow 0$ ,  $h/g^2 \rightarrow 0$ ,  $nh^{1/2}g^4 \rightarrow 0$  and  $r_n^2 nh^{1/2} \rightarrow \infty$ . Moreover,  $g = n^{-\gamma}$  with  $\gamma \in (1/5, 1/4)$  and thus  $nh^2 \rightarrow \infty$ .

*Proof Proposition 5.1.* First let us remark that for any  $(a_n)$ , a sequence divergent to infinity,

$$\mathbb{P}\left(\max_{1 \leq i \leq n} \|X_i\| > a_n \ln n\right) \rightarrow 0 \quad \text{and} \quad \mathbb{P}\left(\|\beta_n - \bar{\beta}\| > a_n n^{-1/2}\right) \rightarrow 0. \quad (5.1)$$

Moreover, at least for  $\beta$  in a fixed but small enough neighborhood of  $\bar{\beta}$ , the matrix  $\mathbf{A}(\beta)$  could be built such that the norm of each of the  $p-1$  columns of  $\mathbf{A}(\beta) - \mathbf{A}(\bar{\beta})$  is bounded by  $c\|\beta_n - \bar{\beta}\|$  with  $c$  a constant independent of  $\beta$ . Indeed, one could consider  $p-1$  independent vectors which completed by any  $\beta$  close to  $\bar{\beta}$  form a basis. Then one could use the Gram-Schmidt procedure to orthonormalize the basis. By construction, the norm of any columns of  $\mathbf{A}(\beta) - \mathbf{A}(\bar{\beta})$  is bounded by  $c\|\beta_n - \bar{\beta}\|$  for some  $c$  depending only on the initial  $p-1$  independent vectors. All these facts show that we can reduce the parameter set to  $\mathcal{B}_n$ ,  $n \geq 1$ , a sequence of balls centered at  $\bar{\beta}$  of radius converging to zero. Consider the set of elementary events

$$\mathcal{E}_n = \left\{ \max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_n} [\|Z_i(\beta) - Z_i(\bar{\beta})\| + \|W_i(\beta) - W_i(\bar{\beta})\|] \leq b_n \right\}, \quad (5.2)$$

where  $b_n$  is a sequence such that  $b_n \rightarrow 0$ . The equation (5.1) indicates that the sequences  $\mathcal{B}_n$  and  $b_n$  could be taken such that the radius of  $\mathcal{B}_n$  converges to zero slower than  $n^{-1/2}$  and faster than  $b_n$ , and  $b_n n^{1/2} / \ln n \rightarrow \infty$ . Then  $\mathbb{P}(\beta_n \in \mathcal{B}_n) \rightarrow 1$  and  $\mathbb{P}(\mathcal{E}_n^c)$  decreases to zero faster than any negative power of the sample size  $n$ . Hence, in the following it will suffice to prove the statements on the set  $\{\beta_n \in \mathcal{B}_n\} \cap \mathcal{E}_n$ .

We will focus on  $I_n(\beta_n)$  since the arguments for  $\hat{\omega}(\beta_n)$  are similar and much simpler. Hereafter, by abuse, we write  $Y_i(t)$  instead of  $Y_{ni}(t)$  even when  $r_n \neq 0$ . To prove that  $I_n(\beta_n) - I_n(\bar{\beta}) = o_{\mathbb{P}}(I_n(\bar{\beta}))$  we will show below that  $I_n(\beta_n) - I_n(\bar{\beta}) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$ . This shows that  $I_n(\beta_n)$  is negligible compared to  $I_n(\bar{\beta})$  both on the null and alternative hypotheses. Indeed, under the null hypothesis,  $r_n \equiv 0$ ,  $\bar{\beta} = \beta_0$  and  $nh^{1/2}I_n(\beta_0)$  is asymptotically centered normal distributed, while on the alternative the  $I_n(\beta)$  is driven by a term of order  $r_n^2$ .

In the following  $C, C', \dots$  denote constants that may have different values from line to line. Let us simplify notation and write

$$\hat{V}_i(\beta) = \widehat{U_i w(Z_i)}(\beta)$$

and

$$L_{ij}(\beta) = L_{ij}(\beta, g), \quad K_{ij}(\beta) = K_{ij}(\beta, h), \quad \phi_{ij}(\beta) = \phi(W_i(\beta) - W_j(\beta)). \quad (5.3)$$

Then,

$$\begin{aligned} I_n(\beta) - I_n(\bar{\beta}) &= \frac{1}{n(n-1)h} \sum_{i \neq j} [\langle \hat{V}_i(\beta), \hat{V}_j(\beta) \rangle_{\mathcal{H}} - \langle \hat{V}_i(\bar{\beta}), \hat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}}] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &+ \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \hat{V}_i(\bar{\beta}), \hat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}} [K_{ij}(\beta) \phi_{ij}(\beta) - K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta})] \\ &+ \frac{1}{n(n-1)h} \sum_{i \neq j} [\langle \hat{V}_i(\beta), \hat{V}_j(\beta) \rangle_{\mathcal{H}} - \langle \hat{V}_i(\bar{\beta}), \hat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}}] \\ &\quad \times [K_{ij}(\beta) \phi_{ij}(\beta) - K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta})] \\ &= D_{n1}(\beta) + D_{n2}(\beta) + D_{n3}(\beta). \end{aligned}$$

Let us investigate the uniform rates of  $D_{n1}$  and  $D_{n2}$ , the term  $D_{n3}$  being uniformly smaller. We can write

$$\begin{aligned} D_{n1}(\beta) &= \frac{2}{n(n-1)h} \sum_{i \neq j} \langle \widehat{V}_i(\beta) - \widehat{V}_i(\bar{\beta}), \widehat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}} K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &\quad + \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \widehat{V}_i(\beta) - \widehat{V}_i(\bar{\beta}), \widehat{V}_j(\beta) - \widehat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}} K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &= 2D_{n11}(\beta) + D_{n12}(\beta). \end{aligned}$$

Moreover,

$$\begin{aligned} \widehat{V}_i(\bar{\beta})(t) &= \frac{1}{n-1} \sum_{k \neq i} [Y_i(t) - Y_k(t)] \frac{1}{g} L_{ik}(\bar{\beta}) \\ &= [Y_i(t) - r_i(t; \bar{\beta})] f_{\bar{\beta}}(X'_i \bar{\beta}) \\ &\quad + [Y_i(t) - r_i(t; \bar{\beta})] \left[ \frac{1}{n-1} \sum_{k \neq i} \frac{1}{g} L_{ik}(\bar{\beta}) - f_{\bar{\beta}}(X'_i \bar{\beta}) \right] \\ &\quad + \frac{1}{n-1} \sum_{k \neq i} \{r_i(t; \bar{\beta}) - r_k(t; \bar{\beta})\} \frac{1}{g} L_{ik}(\bar{\beta}) \\ &\quad - \frac{1}{n-1} \sum_{k \neq i} \{Y_k(t) - r_k(t; \bar{\beta})\} \frac{1}{g} L_{ik}(\bar{\beta}) \\ &= [Y_i(t) - r_i(t; \bar{\beta})] f_{\bar{\beta}}(X'_i \bar{\beta}) + [Y_i(t) - r_i(t; \bar{\beta})] R_{1,ni} + R_{2,ni}(t) - R_{3,ni}(t), \end{aligned}$$

where, by Lemma 5.6

$$\sup_{1 \leq i \leq n} |R_{1,ni}| = O_{\mathbb{P}}(g + n^{-1/2} g^{-1/2} \ln^{1/2} n),$$

and, Lemma 5.5 yields

$$\sup_{1 \leq i \leq n} \sup_{t \in [0,1]} |R_{3,ni}(t)| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n).$$

A representation of  $R_{2,ni}(t)$  is provided in Lemma 5.7. On the other hand,

$$\widehat{V}_i(\beta)(t) - \widehat{V}_i(\bar{\beta})(t) = \frac{1}{n-1} \sum_{k \neq i} [Y_i(t) - Y_k(t)] \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right].$$

**Uniform bounds for  $D_{n1}$ .**

**The rate of  $D_{n11}$ .** Since  $Y_i(t) = r_i(t; \bar{\beta}) + r_n \delta(X_i, t) + \epsilon_i(t)$ , with  $\mathbb{E}[\epsilon_i(t) | X_i] = 0$ , we have  $D_{n11}(\beta) = D_{n111}(\beta) + R_{n11}(\beta)$  with

$$\begin{aligned} D_{n111}(\beta) &= \frac{1}{n(n-1)^2 h} \sum_{i \neq j \neq k} \langle Y_i(\cdot) - Y_k(\cdot), Y_j(\cdot) - r_j(\cdot; \bar{\beta}) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &= \frac{1}{n(n-1)^2 h} \sum_{i \neq j \neq k} \langle Y_i(\cdot) - Y_k(\cdot), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}), \end{aligned}$$

and  $R_{n11}(\beta) = D_{n11}(\beta) - D_{n111}(\beta)$ . We decompose

$$\begin{aligned} D_{n111}(\beta) &= \frac{1}{n(n-1)^2 h} \sum_{i \neq j \neq k} \langle \epsilon_i(\cdot) - \epsilon_k(\cdot), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &+ \frac{1}{n(n-1)^2 h} \sum_{i \neq j \neq k} \langle r_i(\cdot; \bar{\beta}) - r_k(\cdot; \bar{\beta}), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &+ \frac{r_n}{n(n-1)^2 h} \sum_{i \neq j \neq k} \langle \delta(X_i, t) - \delta(X_k, t), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &= D_{n1111}(\beta) + D_{n1112}(\beta) + r_n D_{n1113}(\beta). \end{aligned}$$

The quantity  $ghD_{n1111}(\beta)$  could be decomposed in a sum of degenerate  $U$ -process of order 3 and another one of order 2 indexed by  $\beta$ . To bound them we use the maximal inequality of Sherman (1994). Since  $nh^2, ng^4 \rightarrow \infty$ , deduce that the degenerate  $U$ -process of order 3 is of uniform rate

$$n^{-3/2} O_{\mathbb{P}}(h^{\alpha/2} \{b_n^2 g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1} h^{-1/2}),$$

over any sequence of balls centered at  $\bar{\beta}$  with radius decreasing to zero faster than  $b_n$ , where  $b_n$  is a sequence such that  $b_n n^{1/2} / \ln n \rightarrow \infty$  and  $\alpha$  could be a number in the interval  $(0, 1)$  arbitrarily close to 1. The details on how the maximal inequality of Sherman (1994) applies are provided below for deriving the uniform rate of  $D_{n12}$ . To bound the right-hand side term in that maximal inequality we use the fact that



$\mathbb{E}(\|\epsilon\|_{\mathcal{H}}^2 \mid X)$  and  $f_{\bar{\beta}}(X'\bar{\beta})$  are bounded and the uniform bounds (5.6), (5.4) and (5.5) from Lemma 5.8 in the Appendix. Using very similar arguments, the degenerate  $U$ -process of order 2 in the decomposition of  $ghD_{n111}(\beta)$  could be shown to be of uniform rate

$$n^{-1}O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2})$$

provided that  $nh^2, ng^4 \rightarrow \infty$  and  $\alpha$  is sufficiently close to 1. Next, for  $ngD_{n112}(\beta)$ , that is centered, use the Hoeffding decomposition and the regularity of the function  $v \mapsto \mathbb{E}[Y(t) \mid Z(\bar{\beta}) = v]$ . For the degenerate  $U$ -processes of order 3 and 2 in the Hoeffding decomposition of  $D_{n112}(\beta)$  we apply the maximal inequality of Sherman (1994) as previously. Deduce the respective uniform rates over  $\mathcal{B}_n$

$$g^2n^{-3/2}O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}),$$

and

$$g^2n^{-1}O_{\mathbb{P}}(h^{\alpha/2}\{b_n^2g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}).$$

It remains the  $U$ -process of order 1. Using again the bounds from Lemma 5.8, deduce the uniform rate over  $\mathcal{B}_n$

$$g^2n^{-1/2}O_{\mathbb{P}}(h^{\alpha}\{b_n^2g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}).$$

Deduce  $D_{n112}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2})$ . For  $ghD_{n113}(\beta)$  the arguments are similar, but without the  $g^2$  factor, and yield the uniform rate

$$n^{-1/2}O_{\mathbb{P}}(h^{\alpha}\{b_n^2g^{-1}\}^{\alpha/2}) = gh \times o_{\mathbb{P}}(n^{-1}h^{-1/2}) = o_{\mathbb{P}}(n^{-1/2}h^{-1/4}),$$

provided  $nh^2, ng^4 \rightarrow \infty$  and  $\alpha$  is sufficiently close to 1. Deduce that  $D_{n111}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$ .

For  $R_{n11}(\beta)$  we can write

$$\begin{aligned} R_{n11}(\beta) &= \frac{1}{n(n-1)^2h} \sum_{i \neq j} \langle Y_i(\cdot) - Y_j(\cdot), Y_j(\cdot) - r_j(t; \bar{\beta}) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\ &\quad \times \left[ \frac{1}{g} L_{ij}(\beta) - \frac{1}{g} L_{ij}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &+ \frac{1}{n(n-1)^2h} \sum_{i \neq j, i \neq k} \langle Y_i(\cdot) - Y_k(\cdot), Y_j(\cdot) - r_j(t; \bar{\beta}) \rangle_{\mathcal{H}} R_{1,nj} \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &+ \frac{1}{n(n-1)^2h} \sum_{i \neq j, i \neq k} \langle Y_i(\cdot) - Y_k(\cdot), R_{2,nj}(\cdot) + R_{3,nj}(\cdot) \rangle_{\mathcal{H}} \\ &\quad \times \left[ \frac{1}{g} L_{ik}(\beta) - \frac{1}{g} L_{ik}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\ &= R_{n111}(\beta) + R_{n112}(\beta) + R_{n113}(\beta). \end{aligned}$$

We only investigate  $R_{n111}(\beta)$ , the terms  $R_{n112}(\beta)$  and  $R_{n113}(\beta)$  are uniformly smaller compared to  $D_{n111}(\beta)$ . We can write

$$\begin{aligned}
R_{n111}(\beta) &= \frac{1}{n-1} \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \epsilon_i(\cdot) - \epsilon_j(\cdot), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\
&\quad \times \left[ \frac{1}{g} L_{ij}(\beta) - \frac{1}{g} L_{ij}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&+ \frac{1}{n-1} \frac{1}{n(n-1)h} \sum_{i \neq j} \langle r_i(\cdot; \bar{\beta}) - r_j(\cdot; \bar{\beta}), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\
&\quad \times \left[ \frac{1}{g} L_{ij}(\beta) - \frac{1}{g} L_{ij}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&+ \frac{1}{n-1} \frac{r_n}{n(n-1)h} \sum_{i \neq j} \langle \delta(X_i, t) - \delta(X_j, t), \epsilon_j(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_j \bar{\beta}) \\
&\quad \times \left[ \frac{1}{g} L_{ij}(\beta) - \frac{1}{g} L_{ij}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \\
&= R_{n1111}(\beta) + R_{n1112}(\beta) + r_n R_{n1113}(\beta).
\end{aligned}$$

The leading term in  $R_{n1111}(\beta)$  is

$$\frac{1}{n-1} \frac{1}{n(n-1)h} \sum_{i \neq j} \|\epsilon_j(\cdot)\|_{\mathcal{H}}^2 f_{\bar{\beta}}(X'_j \bar{\beta}) \left[ \frac{1}{g} L_{ij}(\beta) - \frac{1}{g} L_{ij}(\bar{\beta}) \right] K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}).$$

Use the boundedness of  $\mathbb{E}[\|\epsilon_j(\cdot)\|_{\mathcal{H}}^2 \mid X_j]$  and  $f_{\bar{\beta}}(X'_j \bar{\beta})$ , and Lemma 5.8 to deduce that  $R_{n1111}(\beta_n) = o_{\mathbb{P}}(n^{-1})$ . Gathering facts deduce that

$$D_{n11}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2).$$

**The rate of  $D_{n12}$ .** We have

$$\widehat{V}_i(\beta)(t) - \widehat{V}_i(\bar{\beta})(t) = Y_i(t) \Delta_{1,ni}(\beta) + \Delta_{2,ni}(\beta)$$

with  $\Delta_{1,ni}(\beta)$  and  $\Delta_{2,ni}(\beta)$  independent of  $t$  and

$$\sup_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_n} \{|\Delta_{1,ni}| + |\Delta_{2,ni}|\} = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n + b_n);$$

see Lemma 5.5. Replacing and taking absolute values, deduce

$$D_{n12}(\beta_n) = O_{\mathbb{P}}(n^{-1}g^{-1} \ln n + n^{-1} \ln^2 n) = o_{\mathbb{P}}(n^{-1}h^{-1/2}),$$

since  $g^{-1}h^{1/2} \rightarrow 0$  and  $h \ln^4 n \rightarrow 0$ .

Gathering facts deduce that

$$D_{n1}(\beta_n) = D_{n11}(\beta_n) + D_{n12}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2).$$

**Uniform bounds for  $D_{n2}$ .**

We have

$$\begin{aligned}
D_{n2}(\beta) &= \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \widehat{V}_i(\bar{\beta}), \widehat{V}_j(\bar{\beta}) \rangle_{\mathcal{H}} [K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta})] \\
&= \frac{1}{n(n-1)h} \sum_{i \neq j} \langle [Y_i(t) - r_i(t; \bar{\beta})]f_{\bar{\beta}}(X'_i\bar{\beta}), [Y_j(t) - r_j(t; \bar{\beta})]f_{\bar{\beta}}(X'_j\bar{\beta}) \rangle_{\mathcal{H}} \\
&\quad \times [K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta})] \\
&\quad + \text{terms of smaller rate} \\
&= D_{n21}(\beta) + \text{terms of smaller rate.}
\end{aligned}$$

Recall that by construction,

$$\mathbb{E}[Y_i(t) \mid X_i] = r_i(t; \bar{\beta}) + r_n \delta(X_i, t),$$

so that

$$[Y_i(t) - r_i(t; \bar{\beta})]f_{\bar{\beta}}(X'_i\bar{\beta}) = [\epsilon_i(t) + r_n \delta(X_i, t)]f_{\bar{\beta}}(X'_i\bar{\beta}),$$

with  $\mathbb{E}[\epsilon_i(t) \mid X_i] = 0, \forall t \in [0, 1]$ . Thus

$$\begin{aligned}
D_{n21}(\beta) &= \frac{1}{n(n-1)h} \sum_{i \neq j} \langle \epsilon_i(t), \epsilon_j(t) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_i\bar{\beta})f_{\bar{\beta}}(X'_j\bar{\beta}) \\
&\quad \times [K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta})] \\
&\quad + \frac{2r_n}{n(n-1)h} \sum_{i \neq j} \langle \epsilon_i(t), \delta(X_j, t) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_i\bar{\beta})f_{\bar{\beta}}(X'_j\bar{\beta}) \\
&\quad \times [K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta})] \\
&\quad + \frac{r_n^2}{n(n-1)h} \sum_{i \neq j} \langle \delta(X_i, t), \delta(X_j, t) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X'_i\bar{\beta})f_{\bar{\beta}}(X'_j\bar{\beta}) \\
&\quad \times [K_{ij}(\beta)\phi_{ij}(\beta) - K_{ij}(\bar{\beta})\phi_{ij}(\bar{\beta})] \\
&= D_{n211}(\beta) + 2r_n D_{n212}(\beta) + r_n^2 D_{n213}(\beta).
\end{aligned}$$

The term  $D_{n211}(\cdot)$  is a degenerate  $U$ -process of order 2, indexed by  $\beta$ . Consider the family of functions

$$\mathcal{F}_n = \{h(\cdot, \cdot; \beta) : \beta \in \mathcal{B}_n\} \tag{5.4}$$

with

$$h((x_1, \epsilon_1), (x_2, \epsilon_2); \beta) = \langle \epsilon_1(\cdot), \epsilon_2(\cdot) \rangle_{\mathcal{H}} f_{\bar{\beta}}(x'_1\bar{\beta})f_{\bar{\beta}}(x'_2\bar{\beta})[K_{12}(\beta)\phi_{12}(\beta) - K_{12}(\bar{\beta})\phi_{12}(\bar{\beta})].$$

It is quite easy to see that  $\mathcal{F}_n$  is a VC class, or Euclidean in the terminology of Sherman (1994), for a squared integrable envelope  $H(\cdot)$ , with some  $A$  and  $V$  independent of  $n$ . (Recall that the  $\delta$ -covering number of an Euclidean class of function

is bounded by  $A\delta^{-V}$ .) Since  $\mathbb{E}(\|\epsilon_1(\cdot)\|_{\mathcal{H}}^2 \mid X_1)$  and  $f_{\bar{\beta}}(X_1'\bar{\beta})$  are bounded, and the kernel  $K$  is bounded, by Lemma 5.8 deduce that

$$\mathbb{E} \left[ \sup_{\beta \in \mathcal{B}_n} h(\cdot, \cdot; \beta)^2 \right] \leq Ch^{1/2}b_n$$

for some constant  $C > 0$  independent on  $n$  and  $\bar{\beta}$ . See Lemma 5.8 below. Applying the Main Corollary of Sherman (1994) with  $k = 2$ ,  $p = 1$ , deduce that<sup>2</sup>

$$\sup_{\beta} |hD_{n211}(\beta)| \leq \frac{C'}{n} (b_n h^{1/2})^{\alpha/2} = n^{-1} h^{1/2} \times O(n^{-\alpha/4} h^{\alpha/4 - 1/2})$$

for  $0 < \alpha < 1$ . Since  $\alpha$  could be arbitrarily close to 1 and  $b_n$  could be any sequence such that  $b_n n^{1/2} \ln^{-1} n \rightarrow \infty$  and  $nh^{3/2} \rightarrow 0$ , deduce that

$$D_{n211}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2}).$$

For the uniform rate of the centered  $U$ -process  $D_{n212}(\cdot)$ , use the Hoeffding decomposition. The degenerate  $U$ -process of order 1 in this decomposition could be handled with the arguments used for  $D_{n211}(\cdot)$  and shown to be of uniform rate  $o_{\mathbb{P}}(n^{-1/2})$ . The degenerate  $U$ -process of order 2 in the decomposition is

$$D_{n212,1}(\beta) = \frac{1}{n} \sum_{i \neq j} \langle \epsilon_i(t), \gamma_i(\beta, t; h) \rangle_{\mathcal{H}} f_{\bar{\beta}}(X_i'\bar{\beta})$$

where

$$\gamma_i(\beta, t; h) = \mathbb{E} \left\{ \delta(X_j, t) f_{\bar{\beta}}(X_j'\bar{\beta}) h^{-1} \left[ K_{ij}(\beta) \phi_{ij}(\beta) - K_{ij}(\bar{\beta}) \phi_{ij}(\bar{\beta}) \right] \mid X_i \right\}.$$

Since  $f_{\bar{\beta}}$  and  $\delta(X, \cdot)$  are supposed bounded, arguments as used for Lemma 5.8 allow to show that  $\mathbb{E}[\sup_{\beta} \gamma_i^2(\beta, t)] = o(1)$ . Deduce that  $D_{n212,1}(\beta_n) = o_{\mathbb{P}}(n^{-1/2})$ . Gathering facts,  $r_n D_{n212}(\beta_n)$  is negligible compared to  $r_n^2$ . By the same arguments,  $D_{n213}(\beta_n) = o_{\mathbb{P}}(1)$  so that we can conclude that  $D_{n2}(\beta_n) = o_{\mathbb{P}}(n^{-1}h^{-1/2} + r_n^2)$ .  $\square$

*Proof of Proposition 5.2.* Let us consider the simplified notation from equation (5.3) and further simplify in the case  $\beta = \beta_0$  and write

$$L_{ij} = L_{ij}(\beta_0, g), \quad K_{ij} = K_{ij}(\beta_0, h), \quad \text{and} \quad \phi_{ij} = \phi(W_i(\beta_0) - W_j(\beta_0)). \quad (5.5)$$

<sup>2</sup>Let us point out that the rate could be improved if one tracks the dependence of the constants appearing in Sherman's result on the  $\delta$ -covering number of  $\mathcal{F}_n$ . This covering number decreases with  $n$  as the parameter set  $\mathcal{B}_n$  shrinks to  $\bar{\beta}$ . For our purposes we do not need this refinement.

Notice that

$$\begin{aligned}
I_n^{\{U\}}(\beta_0) &= \frac{1}{n(n-1)h} \sum_{1 \leq i \neq j \leq n} \left\{ \langle (r_i - \tilde{r}_i)(\cdot; \beta_0), (r_j - \tilde{r}_j)(\cdot; \beta_0) \rangle_{L^2} \right. \\
&\quad + \langle \epsilon_i(\cdot), \epsilon_j(\cdot) \rangle_{L^2} \\
&\quad + \langle \tilde{\epsilon}_i(\cdot), \tilde{\epsilon}_j(\cdot) \rangle_{L^2} \\
&\quad + 2 \langle \epsilon_i(\cdot), (r_j - \tilde{r}_j)(\cdot; \beta_0) \rangle_{L^2} \\
&\quad - 2 \langle \tilde{\epsilon}_i(\cdot), (r_j - \tilde{r}_j)(\cdot; \beta_0) \rangle_{L^2} \\
&\quad \left. - 2 \langle \epsilon_i(\cdot), \tilde{\epsilon}_j(\cdot) \rangle_{L^2} \right\} \hat{f}_{\beta_0, i} \hat{f}_{\beta_0, j} K_{ij} \phi_{ij} \\
&= I_1(\beta_0) + I_2(\beta_0) + I_3(\beta_0) + 2I_4(\beta_0) - 2I_5(\beta_0) - 2I_6(\beta_0)
\end{aligned}$$

with

$$\begin{aligned}
\hat{f}_{\beta, i} &= \frac{1}{(n-1)g} \sum_{k \neq i} L_{ik}(\beta), \quad r_i(t; \beta) = \mathbb{P}[Y_i \leq \Phi^{-1}(t) \mid X'_i \beta], \\
\tilde{r}_i(t; \beta) &= \frac{1}{(n-1)g \hat{f}_{\beta, i}} \sum_{k \neq i} r_k(t; \beta) L_{ik}(\beta)
\end{aligned}$$

and  $\tilde{\epsilon}_i(\cdot)$  is defined as  $\tilde{r}_i(t; \beta)$  by replacing  $r_i(t; \beta)$  by  $\epsilon_i(\cdot)$ . This decomposition of  $I_n^{\{U\}}(\beta_0)$  is given by the identity

$$\widehat{U_i \omega}(Z_i)(\cdot; \beta_0) = [r_i(\cdot; \beta_0) - \tilde{r}_i(\cdot; \beta_0) + \epsilon_i(\cdot) - \tilde{\epsilon}_i(\cdot)] \hat{f}_{\beta_0, i}.$$

The terms  $I_1(\beta_0)$  and  $I_3(\beta_0)$  are treated in Lemmas 5.10 and 5.11 in Section 5.8. For  $I_2(\beta_0)$ , let us introduce

$$\omega_n^2(\beta) = \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j \neq i} \int \int \Gamma^2(s, t) ds dt \hat{f}_{\beta, i}^2 \hat{f}_{\beta, j}^2 K_{ij}^2(\beta) \phi_{ij}^2(\beta).$$

Proposition 5.4 below ensures that  $nh^{1/2}\omega_n^{-1}(\beta_0)I_2(\beta_0) \rightarrow \mathcal{N}(0, 1)$  in law. The terms  $I_4(\beta_0)$ ,  $I_5(\beta_0)$  and  $I_6(\beta_0)$  can be shown to be negligible in a similar way as  $I_1(\beta_0)$  and  $I_3(\beta_0)$ . Lemma 5.13 shows that  $\omega_n^2(\beta_0) \rightarrow \omega^2(\beta_0)$  in probability with  $\omega^2(\beta_0) > 0$  and thus  $I_j(\beta_0)/\omega_n(\beta_0)$  is of the same order as  $I_j(\beta_0)$  for  $j \in \{1, 3, 4, 5, 6\}$ . Finally, it is easy to check that  $\omega_n(\beta_0) - \hat{\omega}_n^{\{U\}}(\beta_0) = o_{\mathbb{P}}(1)$ . Then the result of the proposition follows.  $\square$

**Proposition 5.4.** *Under the conditions of Proposition 5.2,*

$$nh^{1/2}\omega_n^{-1}(\beta_0)I_2(\beta_0) \rightarrow \mathcal{N}(0, 1) \quad \text{in law.}$$

*Proof.*  $\{S_{n,m}, \mathcal{F}_{n,m}, 1 \leq m \leq n, n \geq 1\}$  is a martingale array with  $S_{n,1} = 0$  and

$$S_{n,m}(\beta_0) = \sum_{i=1}^m G_{n,i}(\beta_0)$$

with

$$G_{n,i}(\beta_0) = \frac{2h^{p/2}}{\omega_n(n-1)h} \left\langle \epsilon_i(\cdot) \hat{f}_{\beta_0,i}, \sum_{j=1}^{i-1} \epsilon_j(\cdot) \hat{f}_{\beta_0,j} K_{ij} \phi_{ij} \right\rangle_{L^2}$$

and  $\mathcal{F}_{n,m}$  is the  $\sigma$ -field generated by  $\{X_1, \dots, X_n, Y_1, \dots, Y_m\}$ . Thus

$$nh^{1/2}\omega_n^{-1}(\beta_0) I_2(\beta_0) = S_{n,n}(\beta_0).$$

Next, define

$$\begin{aligned} V_n^2(\beta_0) &= \sum_{i=2}^n \mathbb{E} [G_{n,i}^2(\beta_0) \mid \mathcal{F}_{n,i-1}] \\ &= \frac{4}{\omega_n^2(n-1)^2 h} \sum_{i=2}^n \int \int \Gamma(s, t) \hat{f}_{\beta_0,i}^2 \left( \sum_{j=1}^{i-1} \epsilon_j(s) \hat{f}_{\beta_0,j} K_{ij} \phi_{ij} \right) \\ &\quad \times \left( \sum_{k=1}^{i-1} \epsilon_k(t) \hat{f}_{\beta_0,k} K_{ik} \phi_{ik} \right) ds dt \end{aligned}$$

and decompose

$$\begin{aligned} V_n^2(\beta_0) &= \frac{4}{\omega_n^2(n-1)^2 h} \sum_{i=2}^n \sum_{j=1}^{i-1} \int \int \Gamma(s, t) \hat{f}_{\beta_0,i}^2 \epsilon_j(s) \epsilon_j(t) \hat{f}_{\beta_0,j}^2 K_{ij}^2 \phi_{ij}^2 ds dt \\ &\quad + \frac{8}{\omega_n^2(n-1)^2 h} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \int \int \Gamma(s, t) \hat{f}_{\beta_0,i}^2 \epsilon_j(s) \epsilon_k(t) \hat{f}_{\beta_0,j} \hat{f}_{\beta_0,k} K_{ij}^2 \phi_{ij}^2 ds dt \\ &= A_n(\beta_0) + B_n(\beta_0). \quad (5.6) \end{aligned}$$

From Lemma 5.12, we have that the martingale array satisfies Corollary 3.1 of Hall and Heyde (1980) and the result follows.  $\square$

## 5.8 Appendix 2: technical lemmas

In the following results the kernels  $L$  and  $K$  are supposed to satisfy the conditions of Assumption 5.3-(f).

**Lemma 5.5.** *Assume that  $\mathbb{E}[\exp(a\|X\|)] < \infty$  for some  $a > 0$ . Consider that  $g \rightarrow 0$  and  $ng^{4/3}/\ln n \rightarrow \infty$ . For any  $t \in [0, 1]$  let  $Y_k(t)$ ,  $1 \leq k \leq n$ , be an i.i.d. random variables like in the proof of Proposition 5.1 such that  $\mathbb{E}[\sup_t |Y_k(t)|^a] < \infty$  for some  $a > 8$ . Moreover, assume that the maps  $v \mapsto \mathbb{E}[|Y_k(t)| \mid X'\beta = v]f_{\bar{\beta}}(v)$ ,  $v \in \mathbb{R}$ ,  $t \in [0, 1]$ , are uniformly Lipschitz (the Lipschitz constant does not depend on  $t$ ). Then*

$$\max_{1 \leq i \leq n} \sup_{t \in [0, 1]} \sup_{\beta \in \mathcal{B}_n} \left| \frac{1}{n-1} \sum_{k \neq i} Y_k(t) \frac{1}{g} [L_{ik}(\beta) - L_{ik}(\bar{\beta})] \right| = O_{\mathbb{P}}(n^{-1/2}g^{-1/2} \ln^{1/2} n + b_n).$$

Moreover,

$$\max_{1 \leq i \leq n} \sup_{t \in [0, 1]} \sup_{\beta \in \mathcal{B}_n} \left| \frac{1}{n-1} \sum_{k \neq i} \{Y_k(t) - \mathbb{E}[Y_k(t) \mid X_k'\bar{\beta}]\} \frac{1}{g} L_{ik}(\bar{\beta}) \right| = O_{\mathbb{P}}(n^{-1/2}g^{-1/2} \ln^{1/2} n).$$

*Proof of Lemma 5.5.* Recall that  $Y_i(t) \equiv Y_i$  (in the case of SIM for mean regression) or  $Y_i(t) = \mathbf{1}\{Y_i \leq \Phi^{-1}(t)\}$  (for the case of single-index assumption on the conditional law), and  $r_i(t; \bar{\beta}) = \mathbb{E}[Y_i(t) \mid Z(\bar{\beta})]$ ,  $t \in [0, 1]$ . For any  $t \in [0, 1]$  we decompose

$$\begin{aligned} & \frac{1}{ng} \sum_{k \neq i} Y_k(t) L_{ik}(\beta) \\ &= \frac{1}{ng} \sum_{k=1}^n \{Y_k(t) L((X_i - X_k)'\beta/g) - \mathbb{E}[Y(t) L((X_i - X)'\beta/g) \mid X_i]\} \\ & \quad + \mathbb{E}[Y(t)g^{-1} L((X_i - X)'\beta/g) \mid X_i] - n^{-1}g^{-1}L(0)Y_i(t) \\ & = \Sigma_{1ni}(\beta, t) + \Sigma_{2i}(\beta, t) - n^{-1}g^{-1}L(0)Y_i(t). \end{aligned}$$

The moment condition on  $Y$  guarantees that

$$\max_{1 \leq i \leq n} \sup_t |Y_i(t)| = o_{\mathbb{P}}(n^b)$$

for some  $0 < b < 1/8$ . This and the fact that  $ng^{4/3}/\ln n \rightarrow \infty$  make that

$$\max_{1 \leq i \leq n} \sup_t n^{-1}g^{-1}|Y_i(t)| = o_{\mathbb{P}}(n^{-1/2}g^{-1/2} \ln^{1/2} n).$$

On the other hand, by Lemma 5.9,

$$\max_{1 \leq i \leq n} \sup_{t \in [0, 1]} \sup_{\beta \in \mathcal{B}_n} \left| \Sigma_{2i}(\beta, t) - \Sigma_{2i}(\bar{\beta}, t) \right| = O_{\mathbb{P}}(b_n).$$

It remains to uniformly bound  $\Sigma_{ni}(\beta, t)$  and for this purpose we use empirical process tools. Let us introduce some notation. Let  $\mathcal{G}$  be a class of functions of the observations with envelope function  $G$  and let

$$J(\delta, \mathcal{G}, L^2) = \sup_Q \int_0^\delta \sqrt{1 + \ln N(\varepsilon \|G\|_2, \mathcal{G}, L^2(Q))} d\varepsilon, \quad 0 < \delta \leq 1,$$

denote the uniform entropy integral, where the supremum is taken over all finitely discrete probability distributions  $Q$  on the space of the observations, and  $\|G\|_2$  denotes the norm of  $G$  in  $L^2(Q)$ . Let  $Z_1, \dots, Z_n$  be a sample of independent observations and let

$$\mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma(Z_i), \quad \gamma \in \mathcal{G},$$

be the empirical process indexed by  $\mathcal{G}$ . If the covering number  $N(\varepsilon, \mathcal{G}, L^2(Q))$  is of polynomial order in  $1/\varepsilon$ , there exists a constant  $c > 0$  such that  $J(\delta, \mathcal{G}, L^2) \leq c\delta\sqrt{\ln(1/\delta)}$  for  $0 < \delta < 1/2$ . Now if  $\mathbb{E}\gamma^2 < \delta^2\mathbb{E}G^2$  for every  $\gamma$  and some  $0 < \delta < 1$ , and  $\mathbb{E}G^{(4v-2)/(v-1)} < \infty$  for some  $v > 1$ , under mild additional measurability conditions that are satisfied in our context, Theorem 3.1 of van der Vaart and Wellner (2011) implies

$$\sup_{\mathcal{G}} |\mathbb{G}_n \gamma| = J(\delta, \mathcal{G}, L^2) \left( 1 + \frac{J(\delta^{1/v}, \mathcal{G}, L^2) \|G\|_{(4v-2)/(v-1)}^{2-1/v}}{\delta^2 \sqrt{n} \|G\|_2^{2-1/v}} \right)^{v/(2v-1)} \|G\|_2 O_{\mathbb{P}}(1), \quad (5.1)$$

where  $\|G\|_2^2 = \mathbb{E}G^2$  and the  $O_{\mathbb{P}}(1)$  term is independent of  $n$ . Note that the family  $\mathcal{G}$  could change with  $n$ , as soon as the envelope is the same for all  $n$ . We apply this result to the family of functions  $\mathcal{G} = \{\gamma(\cdot; \beta, w, t) - \gamma(\cdot; \bar{\beta}, w, t) : t \in [0, 1], \beta \in \mathcal{B}, w \in \mathbb{R}\}$  where

$$\gamma(Y, X; \beta, w, t) = Y(t)L((X'\beta - w)g^{-1})$$

for a sequence  $g$  that converges to zero and the envelope

$$G(Y, X) = \sup_{t \in [0, 1]} |Y(t)| \sup_{w \in \mathbb{R}} L(w).$$

Its entropy number is of polynomial order in  $1/\varepsilon$ , independently of  $n$ , as  $L(\cdot)$  is of bounded variation and the families of indicator functions have polynomial complexity, see for instance van der Vaart (1998). Now for any  $\gamma \in \mathcal{G}$ ,  $\mathbb{E}\gamma^2 \leq Cg\mathbb{E}G^2$ , for some constant  $C$ . Let  $\delta = g^{1/2}$ , so that  $\mathbb{E}\gamma^2 \leq C'\delta^2\mathbb{E}G^2$ , for some constant  $C'$  and  $v = 3/2$ , which corresponds to  $\mathbb{E}G^8 < \infty$  that is guaranteed by our assumptions. Thus the bound in (5.1) yields

$$\sup_{\mathcal{G}} \left| \frac{1}{g\sqrt{n}} \mathbb{G}_n \gamma \right| = \frac{\ln^{1/2}(n)}{\sqrt{ng}} \left[ 1 + n^{-1/2} g^{-2/3} \ln^{1/2}(n) \right]^{3/4} O_{\mathbb{P}}(1),$$

where the  $O_{\mathbb{P}}(1)$  term is independent of  $n$ . Since  $ng^{4/3}/\ln n \rightarrow \infty$ ,

$$\max_{1 \leq i \leq n} \sup_{t \in [0, 1]} \sup_{\beta \in \mathcal{B}_n} |\Sigma_{1ni}(\beta, t) - \Sigma_{1ni}(\bar{\beta}, t)| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n).$$

The second part of the statement is now obvious. □



**Lemma 5.6.** *Assume that the density  $f_{\bar{\beta}}(\cdot)$  is Lipschitz. Then*

$$\max_{1 \leq i \leq n} \left| \frac{1}{n-1} \sum_{k \neq i} \frac{1}{g} L_{ik}(\bar{\beta}) - f_{\bar{\beta}}(X'_i \bar{\beta}) \right| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n + g).$$

*Proof of Lemma 5.6.* We can write

$$\begin{aligned} \frac{1}{n-1} \sum_{k \neq i} \frac{1}{g} L_{ik}(\bar{\beta}) - f_{\bar{\beta}}(X'_i \bar{\beta}) &= \frac{1}{n} \sum_{k=1}^n \left\{ g^{-1} L_{ik}(\bar{\beta}) - \mathbb{E}[g^{-1} L_{ik}(\bar{\beta}) \mid X_i] \right\} \\ &\quad + \mathbb{E}[g^{-1} L_{ik}(\bar{\beta}) \mid X_i] - f_{\bar{\beta}}(X'_i \bar{\beta}) + O(n^{-1} g^{-1}). \end{aligned}$$

By the empirical process arguments used in Lemma 5.5, the sum on the right-hand side of the display is of rate  $O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n)$  uniformly with respect to  $i$ . The Lipschitz property of  $f_{\bar{\beta}}$  and the fact that  $\int |vL(v)|dv < \infty$  guarantee that

$$\max_{1 \leq i \leq n} |\mathbb{E}[g^{-1} L_{ik}(\bar{\beta}) \mid X_i] - f_{\bar{\beta}}(X'_i \bar{\beta})| \leq Cg$$

for some constant  $C$ . □

**Lemma 5.7.** *For any  $t \in [0, 1]$  let  $Y_k(t)$ ,  $1 \leq k \leq n$ , be an independent sample from a random variable  $Y(t)$  defined like in the proof of Proposition 5.1. Let  $r(v; t, \bar{\beta}) = \mathbb{E}[Y(t) \mid X' \bar{\beta} = v]$ ,  $v \in \mathbb{R}$ , and assume that  $r(\cdot; t, \bar{\beta})$  is twice differentiable and the second derivative is bounded by a constant independent of  $t$ . If  $r'(v; t, \bar{\beta})$  is the first derivative of  $r(\cdot; t, \bar{\beta})$ , then, for any  $t \in [0, 1]$ ,*

$$\frac{1}{n-1} \sum_{k \neq i} \{r(X'_i \bar{\beta}; t, \bar{\beta}) - r(X'_k \bar{\beta}; t, \bar{\beta})\} \frac{1}{g} L_{ik}(\bar{\beta}) = r'(X'_i \bar{\beta}; t, \bar{\beta}) g D_{1,ni} + g^2 D_{1,ni}(t),$$

where

$$\max_{1 \leq i \leq n} |D_{1,ni}| = n^{-1/2} g^{-1/2} \ln^{1/2} n$$

and

$$\max_{1 \leq i \leq n} \sup_{t \in [0,1]} |D_{1,ni}(t)| = O_{\mathbb{P}}(1).$$

*Proof of Lemma 5.7.* By Taylor expansion

$$\begin{aligned} &\frac{1}{n-1} \sum_{k \neq i} \{r(X'_i \bar{\beta}; t, \bar{\beta}) - r(X'_k \bar{\beta}; t, \bar{\beta})\} \frac{1}{g} L_{ik}(\bar{\beta}) \\ &= r'(X'_i \bar{\beta}; t, \bar{\beta}) \frac{1}{n} \sum_{k=1}^n (X_i - X_k)' \bar{\beta} \frac{1}{g} L_{ik}(\bar{\beta}) \\ &\quad + \frac{1}{n} \sum_{k=1}^n r''(x_{ik}(t); t, \bar{\beta}) [(X_i - X_k)' \bar{\beta}]^2 \frac{1}{g} L_{ik}(\bar{\beta}), \end{aligned}$$

where  $r''$  stands for the second derivative with respect to  $v$  and  $x_{ik}(t)$  is a point between  $X'_i\bar{\beta}$  and  $X'_k\bar{\beta}$ . Since  $L(\cdot)$  is symmetric, by the empirical process arguments as in Lemma 5.5

$$\max_{1 \leq i \leq n} \left| \frac{1}{n} \sum_{k=1}^n \frac{(X_i - X_k)' \bar{\beta}}{g} \frac{1}{g} L_{ik}(\bar{\beta}) \right| = O_{\mathbb{P}}(n^{-1/2} g^{-1/2} \ln^{1/2} n).$$

The result follows taking absolute values in the last sum in the last display, using the boundedness of  $r''$  and the fact that

$$\max_{1 \leq i \leq n} \left| \frac{1}{n} \sum_{k=1}^n \frac{[(X_i - X_k)' \bar{\beta}]^2}{g^2} \frac{1}{g} L_{ik}(\bar{\beta}) - f_{\bar{\beta}}(X'_i \bar{\beta}) \int_{\mathbb{R}} v^2 |L(v)| dv \right| = o_{\mathbb{P}}(1).$$

□

**Lemma 5.8.** *Assume that  $\mathbb{E}[\exp(a\|X\|)] < \infty$  for some  $a > 0$ . Moreover the kernels  $K$  and  $L$  are of bounded variation, differentiable except at most a finite set of points, and  $\int_{\mathbb{R}} |uK(u)| du < \infty$ . Let  $\mathcal{B}_n$  be a subset in the parameter space such that the event defined in equation (5.2) with  $b_n \rightarrow 0$  and  $b_n n^{1/2} / \ln n \rightarrow \infty$  has probability tending to 1. Let*

$$K_{12}(\beta) = K((X_1 - X_2)' \beta / h), \quad L_{12}(\beta) = L((X_1 - X_2)' \beta / g)$$

and  $\phi(\beta) = \phi((X_1 - X_2)' \mathbf{A}(\beta))$ . If the density  $f_{\bar{\beta}}$  is Lipschitz with constant  $C_{1,\bar{\beta}}$ , then there exists a constant  $C$  depending only on  $K$ ,  $L$ ,  $\|f_{\bar{\beta}}\|_{\infty}$  and  $C_{1,\bar{\beta}}$  such that

$$\mathbb{P} \left\{ \mathbb{E} \left[ \sup_{b \in \mathcal{B}_n} \left| K_{12}(\beta) \phi_{12}(\beta) - K_{12}(\bar{\beta}) \phi_{12}(\bar{\beta}) \right| \mid X_1 \right] \leq C b_n h^{1/2} \right\} \rightarrow 1, \quad (5.2)$$

$$\mathbb{E} \left[ \sup_{b \in \mathcal{B}_n} \left| K_{12}(\beta) \phi_{12}(\beta) - K_{12}(\bar{\beta}) \phi_{12}(\bar{\beta}) \right| \right] \leq C b_n h^{1/2}, \quad (5.3)$$

$$\mathbb{P} \left\{ \mathbb{E} \left[ \sup_{b \in \mathcal{B}_n} \left| L_{12}(\beta) - L_{12}(\bar{\beta}) \right|^2 \mid X_1 \right] \leq C b_n g^{-1} \right\} \rightarrow 1 \quad (5.4)$$

$$\mathbb{P} \left\{ \mathbb{E} \left[ \sup_{b \in \mathcal{B}_n} \left| L_{13}(\beta) - L_{13}(\bar{\beta}) \right|^2 \mid K_{12}(\bar{\beta}) \right|^2 \mid X_2, X_3 \right] \leq C h b_n g^{-1} \right\} \rightarrow 1, \quad (5.5)$$

and

$$\mathbb{E} \left[ \sup_{b \in \mathcal{B}_n} \left| L_{13}(\beta) - L_{13}(\bar{\beta}) \right|^2 \mid K_{12}(\bar{\beta}) \right|^2 \phi_{12}^2(\bar{\beta}) \right] \leq C h b_n g^{-1}, \quad (5.6)$$

In Lemma 5.8 we provide different bounds for  $L(\cdot)$  and  $K(\cdot)$  because the bandwidths  $g$  and  $h$  have to satisfy the condition  $h/g^2 \rightarrow 0$ . Hence we need less restrictive conditions on the range of  $h$  if we want to allow for a larger domain for the pair  $(g, h)$ .

*Proof of Lemma 5.8.* Since the kernel  $K$  is of bounded univariate kernels, let  $K_1$  and  $K_2$  non decreasing bounded functions such that  $K = K_1 - K_2$  and denote  $K_{1h} = K_1(\cdot/h)$ . Clearly, it is sufficient to prove the result with  $K_1$ , similar arguments apply for  $K_2$  and hence we get the results for  $K$ . For simpler writings we assume that  $K$  is differentiable and let  $K_1(x) = \int_{-\infty}^x [K'(t)]^+ dt$  and  $K_2(x) = \int_{-\infty}^x [K'(t)]^- dt$ ,  $x \in \mathbb{R}$ , Here  $[K']^+$  (resp.  $[K']^-$ ) denotes the positive (resp. negative) part of  $K'$ . The general case where a finite set of nondifferentiability is allowed can be handled with obvious modifications. Let  $K_{1h}(t) = K_1(t/h)$  and recall that  $Z_i(\beta) = X'_i\beta$ . Note that  $|\exp(-t^2) - \exp(-s^2)| \leq \sqrt{2}|t - s|$ . For any  $\beta \in \mathcal{B}_n$  and an elementary event in the set  $\mathcal{C}_n = \{\max_{1 \leq i \leq n} \|X_i\| \leq c \log n\} \subset \mathcal{E}_n$  for some large constant  $c$ ,

$$\begin{aligned} & \left| K_{1h}(Z_1(\beta) - Z_2(\beta)) \phi_{12}(\beta) - K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta})) \phi_{12}(\bar{\beta}) \right| \\ & \leq \sqrt{2} b_n K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) + 2b_n) \\ & \quad + [K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) + 2b_n) - K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) - 2b_n)] \phi_{12}(\bar{\beta}). \end{aligned}$$

The upper bound on the left-hand side is uniform with respect to  $\beta$ . By a suitable change of variable and since the density  $f_{\bar{\beta}}$  is bounded, it is easy to check that

$$\mathbb{E} \left[ K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) + 2b_n) \mid Z_1(\bar{\beta}) \right]$$

is bounded by a constant times  $hb_n$ . Next, note that since  $nh \rightarrow \infty$ , there exists a constant  $C'$  independent of  $n$  such that on the set  $\mathcal{C}_n$  we have  $|Z_1(\bar{\beta}) - Z_2(\bar{\beta}) \pm 2b_n|/h \leq C'h^{-1/2}$ . Then, applying twice a change of variables and using the Lipschitz property of  $f_{\bar{\beta}}$ , on the set  $\mathcal{C}_n$ ,

$$\begin{aligned} & \mathbb{E} \left[ \left| K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) + 2b_n) - K_{1h}(Z_1(\bar{\beta}) - Z_2(\bar{\beta}) - 2b_n) \right| \mathbf{1}\{\mathcal{C}_n\} \mid Z_1(\bar{\beta}) \right] \\ & \leq h \int_{[-C'/h^{1/2}, C'/h^{1/2}]} K_1(u) \left| f_{\bar{\beta}}(2b_n + Z_1(\bar{\beta}) - uh) - f_{\bar{\beta}}(-2b_n + Z_1(\bar{\beta}) - uh) \right| du \\ & \leq h \times \sup_{t \in \mathbb{R}} \left| f_{\bar{\beta}}(2b_n + t) - f_{\bar{\beta}}(-2b_n + t) \right| \int_{[-C'/h^{1/2}, C'/h^{1/2}]} K_1(u) du \\ & \leq Ch^{1/2}b_n, \end{aligned}$$

for some constant  $C > 0$ . Since by a suitable choice of  $c$  the probability of  $\mathbf{1}\{\mathcal{C}_n\}$  given  $Z_1(\bar{\beta})$  could be made smaller than any fixed negative power of  $n$ , and the probability of the event  $\{|Z_1(\bar{\beta})| \leq c \log n\}$  could be also made very small, the

bound in the last display implies the statement (5.2). For the statement (5.3) it suffices to take expectation.

For the bound in equation (5.4), recall that  $L(t) = L(|t|)$  for any  $t \in \mathbb{R}$  so that we can consider only nonnegative  $t$ . Moreover, without loss of generality we can consider  $L$  nonnegative and decreasing on  $[0, \infty)$ , otherwise, since  $L$  is of bounded variation, it could be written as the difference of two nonnegative decreasing functions on  $[0, \infty)$ . Moreover, let  $Z_{13}(\beta) = |Z_1(\beta) - Z_3(\beta)|$  and  $L_{g,13}(\beta) = L(Z_{13}(\beta)/g)$ . We split the problem in two cases:  $Z_{13}(\beta) \leq Z_{13}(\bar{\beta})$  and  $Z_{13}(\beta) > Z_{13}(\bar{\beta})$ . Then, for  $\beta \in \mathcal{B}_n$  and on the set  $\mathcal{C}_n$  we have

$$\begin{aligned} & \left| L_{g,13}(\beta) - L_{g,13}(\bar{\beta}) \right| \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta})\} \\ & \leq [L(0) - L_{g,13}(\bar{\beta})] \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \leq 2b_n\} \\ & + [L_{g,13}(\beta) - L_{g,13}(\bar{\beta})] \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \geq 2b_n\} \\ & \leq Cb_n^2g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\ & + \left[ L((Z_{13}(\bar{\beta}) - 2b_n)/g) - L(Z_{13}(\bar{\beta})/g) \right] \mathbf{1}\{Z_{13}(\beta) \leq Z_{13}(\bar{\beta}), Z_{13}(\bar{\beta}) \geq 2b_n\} \\ & = Cb_n^2g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} + A_n \end{aligned}$$

and

$$\begin{aligned} & \left| L_{g,13}(\beta) - L_{g,13}(\bar{\beta}) \right| \mathbf{1}\{Z_{13}(\beta) > Z_{13}(\bar{\beta})\} \\ & \leq \left[ L(Z_{13}(\bar{\beta})/g) - L((Z_{13}(\bar{\beta}) + 2b_n)/g) \right] \mathbf{1}\{Z_{13}(\beta) > Z_{13}(\bar{\beta})\} = B_n, \end{aligned}$$

for some constant  $C$ . Let us notice that

$$\begin{aligned} A_n + B_n & \leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)] \mathbf{1}\{Z_{13}(\bar{\beta}) \geq 2b_n\} \\ & + [L([Z_{13}(\bar{\beta})]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)] \mathbf{1}\{0 \leq Z_{13}(\bar{\beta}) \leq 2b_n\} \\ & \leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)] \mathbf{1}\{Z_{13}(\bar{\beta}) \geq 2b_n\} \\ & + Cb_n^2g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\ & \leq [L([Z_{13}(\bar{\beta}) - 2b_n]/g) - L([Z_{13}(\bar{\beta}) + 2b_n]/g)] \\ & + 2Cb_n^2g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\} \\ & = D_n + 2Cb_n^2g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\}. \end{aligned}$$

On the other hand,  $0 \leq D_n \leq 4b_n g^{-1} |L'(\tilde{Z})|$  where  $\tilde{Z}$  is some value such that  $|\tilde{Z} - Z_{13}(\bar{\beta})| \leq 2b_n g^{-1}$ . Since, for some constant  $c$ ,  $|L'(v)| \leq c|v|$  in a neighborhood of the origin,

$$D_n \leq 4b_n g^{-1} |L'(Z_{13}(\bar{\beta}))| + C'b_n^2 g^{-2},$$

for some constant  $C'$ . Since  $L'$  is bounded, deduce that  $|L_{g,13}(\beta) - L_{g,13}(\bar{\beta})|^2$  is bounded by  $Cb_n^2 g^{-1} |g^{-1} L'(Z_{13}(\bar{\beta}))| + o(b_n^2 g^{-1})$  for some constant  $C$ . Take conditional expectation given  $X_1$ , that is the same with the conditional expectation given  $Z_1(\beta)$ , and deduce the bound in equation (5.4).

On the set of events  $\mathcal{C}_n$ ,

$$\sup_{\beta \in \mathcal{B}_n} |L_{13}(\beta) - L_{13}(\bar{\beta})| |K_{12}(\bar{\beta})| \leq \{D_n + 3Cb_n^2 g^{-2} \mathbf{1}\{Z_{13}(\bar{\beta}) \leq 2b_n\}\} |K_{12}(\bar{\beta})|.$$

Take conditional expectation and use standard change of variables to derive the bound in equation (5.5). Take expectation and remember that  $\phi_{12}(\bar{\beta})$  is bounded to derive the moment bound in equation (5.6).  $\square$

**Lemma 5.9.** *Under the conditions of Lemma 5.5*

$$\sup_{t \in [0,1]} \sup_{\beta \in \mathcal{B}_n} \max_{1 \leq i \leq n} |\Sigma_{2i}(\beta, t) - \Sigma_{2i}(\bar{\beta}, t)| = O_{\mathbb{P}}(b_n).$$

*Proof of Lemma 5.9.* We can write

$$\begin{aligned} & |\Sigma_{2i}(\beta, t) - \Sigma_{2i}(\bar{\beta}, t)| \\ & \leq \mathbb{E} \left[ |Y(t)| \left| g^{-1} L((X_i - X)' \beta / g) - g^{-1} L((X_i - X)' \bar{\beta} / g) \right| \mid X_i \right] \\ & = \mathbb{E} \left[ \mathbb{E} \{ |Y(t)| X \} g^{-1} \left| L((X_i - X)' \beta / g) - L((X_i - X)' \bar{\beta} / g) \right| \mid X_i \right]. \end{aligned}$$

Now, we can apply the monotonicity argument we used in Lemma 5.8 and deduce the bound.  $\square$

**Lemma 5.10.** *Under the conditions of Proposition 5.2,  $I_1(\beta_0) = o_{\mathbb{P}}(n^{-1}h^{-1/2})$ .*

*Proof of Lemma 5.10.* With the notation defined in equation (5.5) we have

$$I_1(\beta_0) = \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^n \sum_{\substack{j, k \neq i \\ l \neq j}} \langle (r_i - r_k)(\cdot; \beta_0), (r_j - r_l)(\cdot; \beta_0) \rangle_{L^2} L_{ik} L_{jl} K_{ij} \phi_{ij}$$

and if we denote by  $I_{1,1}(\beta_0)$  the term where  $i, j, k$  and  $l$  are all different, then

$$\begin{aligned} \mathbb{E}[I_{1,1}(\beta_0)] &= \frac{(n-2)(n-3)}{(n-1)^2 g^2 h} \mathbb{E} \left[ \langle \mathbb{E}[(r_i - r_k)(\cdot; \beta_0) L_{ik} \mid Z_i(\beta_0)], \right. \\ & \quad \left. \mathbb{E}[(r_j - r_l)(\cdot; \beta_0) L_{jl} \mid Z_j(\beta_0)] \rangle_{L^2} K_{ij} \phi_{ij} \right] \\ &= O(g^4) \end{aligned}$$

as soon as  $g^{-1} \mathbb{E}[(r_i - r_k)(t; \beta_0) L_{ik}(\beta_0) \mid Z_i(\beta_0)] = O(g^2) D(t; Z_i(\beta_0))$  with  $D(\cdot)$  bounded, which is guaranteed by Assumption 5.3-(c). When  $i, j, k$  and  $l$  take no more than 3 different values, the number of terms is reduced by a factor  $n$ , and thus we have that  $\mathbb{E}[I_{1,2}(\beta_0)] = O(n^{-1}g^{-1}) = o(n^{-1}h^{-1/2})$ . Similar reasoning can be applied to prove that  $\mathbb{E}[I_1^2(\beta_0)] = o(n^{-2}h^{-1})$ . See also Proposition A.1. in Fan and Li (1996).  $\square$

**Lemma 5.11.** *Under the conditions of Proposition 5.2,  $I_3(\beta_0) = o_{\mathbb{P}}(n^{-1}h^{-1/2})$ .*

*Proof of Lemma 5.11.* Write

$$\begin{aligned}
I_3(\beta_0) &= \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j} \langle \epsilon_k(\cdot), \epsilon_l(\cdot) \rangle_{L^2} L_{ik} L_{jl} K_{ij} \phi_{ij} \\
&= \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j, k} \langle \epsilon_k(\cdot), \epsilon_l(\cdot) \rangle_{L^2} L_{ik} L_{jl} K_{ij} \phi_{ij} \\
&\quad + \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i, j} \|\epsilon_k(\cdot)\|_{L^2}^2 L_{ik} L_{ji} K_{ij} \phi_{ij} \\
&\quad + \frac{1}{n(n-1)^3 g^2 h} \sum_{i=1}^n \sum_{j \neq i} \|\epsilon_j(\cdot)\|_{L^2}^2 L_{ij} L_{ji} K_{ij} \phi_{ij} \\
&= I_{3,1}(\beta_0) + I_{3,2}(\beta_0) + I_{3,3}(\beta_0).
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}[I_{3,1}(\beta_0)] &= \frac{1}{(n-1)^2 g^2 h} \mathbb{E} \left[ \langle \epsilon_1(\cdot), \epsilon_2(\cdot) \rangle_{L^2} L_{12}^2 K_{12} \phi_{12} \right] \\
&= O(n^{-2} g^{-2}) \mathbb{E} \left[ \left| \langle \epsilon_1(\cdot), \epsilon_2(\cdot) \rangle_{L^2} h^{-1} K_{12} \right| \right] \\
&= O(n^{-2} g^{-2}),
\end{aligned}$$

$\mathbb{E}[I_{3,2}(\beta_0)] = O(n^{-1} g^{-1})$  and  $\mathbb{E}[I_{3,3}(\beta_0)] = O(n^{-2} g^{-2})$ , thus we have  $\mathbb{E}[I_3(\beta_0)] = o(n^{-1} h^{-1/2})$ . By quite straightforward but tedious calculations, it can be proved that  $\mathbb{E}[I_3^2(\beta_0)] = o(n^{-2} h^{-1})$  and the rate of  $I_3(\beta_0)$  follows.  $\square$

**Lemma 5.12.** *Let  $A_n(\beta_0)$  and  $B_n(\beta_0)$  be defined as in equation (5.6). Under the conditions of Proposition 5.2,  $A_n(\beta_0) \rightarrow 1$  and  $B_n(\beta_0) \rightarrow 0$  in probability, and*

$$\forall \varepsilon > 0, \quad \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] \rightarrow 0, \quad \text{in probability.}$$

*Proof of Lemma 5.12.* First, we have

$$\begin{aligned}
\mathbb{E}[A_n(\beta_0)] &= \mathbb{E}[\mathbb{E}[A_n(\beta_0) \mid X_1, \dots, X_n]] \\
&= \mathbb{E} \left[ \frac{2n}{\omega_n^2(\beta_0)(n-1)h} \int \int \Gamma(s, t) \hat{f}_{\beta_0, i}^2 \mathbb{E}[\epsilon_j(s) \epsilon_j(t)] \hat{f}_{\beta_0, j}^2 K_{ij}^2 \phi_{ij}^2 ds dt \right] \\
&= \frac{n}{n-1} \xrightarrow{n \rightarrow \infty} 1.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{Var}(A_n(\beta_0)) &\leq \frac{64 \|\phi\|_\infty^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{j'=1}^{j-1} \mathbb{E} \left[ \omega_n^{-2}(\beta_0) \hat{f}_{\beta_0,i}^4 \hat{f}_{\beta_0,j}^2 \hat{f}_{\beta_0,j'}^2 K_{ij}^2 K_{ij'}^2 \right] \\
&\quad \times \int \int \int \int \Gamma^2(s,t) \Gamma^2(u,v) ds dt du dv \\
&+ \frac{32 \|\phi\|_\infty^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{i'=2}^{i-1} \sum_{j=1}^{i'-1} \mathbb{E} \left[ \omega_n^{-2}(\beta_0) \hat{f}_{\beta_0,i}^2 \hat{f}_{\beta_0,i'}^2 \hat{f}_{\beta_0,j}^4 K_{ij}^2 K_{i'j}^2 \right] \\
&\quad \times \int \int \int \int \Gamma(s,t) \Gamma(u,v) \mathfrak{G}(s,t,u,v) ds dt du dv \\
&+ \frac{16 \|\phi\|_\infty^4}{(n-1)^4 h^2} \sum_{i=3}^n \sum_{i'=2}^{i-1} \sum_{j=1}^{i'-1} \mathbb{E} \left[ \omega_n^{-2}(\beta_0) \hat{f}_{\beta_0,i}^4 \hat{f}_{\beta_0,j}^4 K_{ij}^4 \right] \\
&\quad \times \int \int \int \int \Gamma(s,t) \Gamma(u,v) \mathfrak{G}(s,t,u,v) ds dt du dv \\
&= o(n^{-1}h^{-1/2}),
\end{aligned}$$

where  $\mathfrak{G}(s,t,u,v) = \mathbb{E}[\epsilon(s)\epsilon(t)\epsilon(u)\epsilon(v)]$ . The decomposition of  $\mathbb{E}[B_n^2]$  involves the same type of terms and is therefore also of rate  $o(n^{-1}h^{-1/2})$ . For the Lindeberg condition, we have  $\forall \varepsilon > 0, \forall n \geq 1$  and  $1 < i \leq n$

$$\mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] \leq \frac{\mathbb{E} \left[ G_{n,i}^4 \mid \mathcal{F}_{n,i-1} \right]}{\varepsilon^2}.$$

Then

$$\begin{aligned}
&\sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^2 I(|G_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right] \\
&\leq \frac{1}{\varepsilon^2} \sum_{i=2}^n \mathbb{E} \left[ G_{n,i}^4 \mid \mathcal{F}_{n,i-1} \right] \\
&\leq \frac{1}{\varepsilon^2} \frac{16}{(n-1)^4 h^2} \sum_{i=2}^n \int \int \int \int \mathfrak{G}(s_1, s_2, s_3, s_4) \hat{f}_{\beta_0,i}^4 \\
&\quad \times \prod_{k=1}^4 \sum_{j_k=1}^{i-1} \epsilon_{j_k}(s_k) \hat{f}_{\beta_0,j_k} K_{ij_k} \phi_{ij_k} ds_k.
\end{aligned}$$

The expectation of the last majorant is of rate

$$\begin{aligned}
& O(n^{-1}) \int \int \int \int \mathfrak{G}(s_1, s_2, s_3, s_4) \Gamma(s_1, s_2) \Gamma(s_3, s_4) ds_1 ds_2 ds_3 ds_4 \\
& \quad \times \mathbb{E} \left[ \hat{f}_{\beta_0, i}^4 \hat{f}_{\beta_0, j}^2 \hat{f}_{\beta_0, j'}^2 h^{-1} K_{ij}^2 h^{-1} K_{ij'}^2 \phi_{ij}^2 \phi_{ij'}^2 \right] \\
& \quad + O(n^{-2} h^{-1}) \sum_{i=2}^n \int \int \int \int \mathfrak{G}^2(s_1, s_2, s_3, s_4) ds_1 ds_2 ds_3 ds_4 \\
& \quad \times \mathbb{E} \left[ \hat{f}_{\beta_0, i}^4 \hat{f}_{\beta_0, j}^4 h^{-1} K_{ij}^4 \phi_{ij}^4 \right] \\
& = o(n^{-1} h^{-1/2}).
\end{aligned}$$

□

**Lemma 5.13.** *Under the conditions of Proposition 5.2,  $\omega_n^2(\beta_0) \rightarrow \omega^2(\beta_0) > 0$ , in probability.*

*Proof of Lemma 5.13.* We have

$$\mathbb{E} \left[ \omega_n^2(\beta_0) \right] = 2 \mathbb{E} \left[ \hat{f}_{\beta, i} \hat{f}_{\beta, j} h^{-1} K_{ij}^2(\beta) \phi_{ij}^2(\beta) \right] \times \int \int \Gamma^2(s, t) ds dt.$$

On the other hand,

$$\begin{aligned}
& \mathbb{E} \left[ \hat{f}_{\beta, i} \hat{f}_{\beta, j} h^{-1} K_{ij}^2 \phi_{ij}^2 \right] \\
& = \frac{1}{g^2 h} \mathbb{E} \left[ \sum_{k \neq i} \sum_{l \neq j} \sum_{k' \neq i} \sum_{l' \neq j} L_{ik} L_{jl} L_{ik'} L_{j'l'} h^{-1} K_{ij}^2 \phi_{ij}^2 \right] \\
& = \frac{1}{g^2 h (n-1)^2} \mathbb{E} \left[ \sum_{k \neq i} \sum_{l \neq j} \sum_{k' \neq i} \sum_{l' \neq j} L_{ik} L_{jl} L_{ik'} L_{j'l'} h^{-1} K_{ij}^2 \phi_{ij}^2 \right] \\
& = \frac{1}{g^2 h (n-1)^4} \mathbb{E} \left[ \sum_{k \neq i} \sum_{l \neq j} \sum_{k' \neq i} \sum_{l' \neq j} L_{ik} L_{jl} L_{ik'} L_{j'l'} h^{-1} K_{ij}^2 \phi_{ij}^2 \right] \\
& \quad + o(n^{-1} h^{-1/2}) \\
& = \frac{(n-1)^3}{(n-2)(n-3)(n-4)} \tilde{\omega}_n^2(\beta_0) + o(n^{-1} h^{-1/2})
\end{aligned}$$



where

$$\begin{aligned}
\tilde{\omega}_n^2(\beta_0) &= \mathbb{E} \left[ \int \int \int \int \frac{1}{g} L\left(\frac{z_i - z_k}{g}\right) \frac{1}{g} L\left(\frac{z_j - z_l}{g}\right) \right. \\
&\quad \times \frac{1}{g} L\left(\frac{z_i - z_{k'}}{g}\right) \frac{1}{g} L\left(\frac{z_j - z_{l'}}{g}\right) \frac{1}{h} K^2\left(\frac{z_i - z_j}{h}\right) \phi_{ij} \\
&\quad \times f_{\beta_0}(z_k) f_{\beta_0}(z_l) f_{\beta_0}(z_{k'}) f_{\beta_0}(z_{l'}) \\
&\quad \left. \times \pi_{\beta_0}(z_i | W_i(\beta_0)) \pi_{\beta_0}(z_j | W_j(\beta_0)) dz_i dz_j dz_k dz_l dz_{k'} dz_{l'} \right] \\
&= \mathbb{E} \left[ \int \int \int \int f_{\beta_0}(z_i - gs_1) f_{\beta_0}(z_i - gs_2) f_{\beta_0}(z_j - gt_1) \right. \\
&\quad \times f_{\beta_0}(z_j - gt_2) \pi_{\beta_0}(z_i | W_i(\beta_0)) \pi_{\beta_0}(z_j | W_j(\beta_0)) \phi_{ij} L(s_1) \\
&\quad \left. \times L(t_1) L(s_2) L(t_2) \frac{1}{h} K^2\left(\frac{z_i - z_j}{h}\right) dz_i dz_j ds_1 dt_1 ds_2 dt_2 \right] \\
&= \mathbb{E} \left[ \int \int \int \int f_{\beta_0}(z_i - gs_1) f_{\beta_0}(z_i - gs_2) f_{\beta_0}(z_i - gu - gt_1) \right. \\
&\quad \times f_{\beta_0}(z_i - gu - gt_2) \pi_{\beta_0}(z_i | W_i(\beta_0)) \pi_{\beta_0}(z_i - gu | W_j(\beta_0)) \\
&\quad \left. \times \phi_{ij} L(s_1) L(t_1) L(s_2) L(t_2) K^2(u) dz_i ds_1 dt_1 ds_2 dt_2 du \right] \\
&\rightarrow \mathbb{E} \left[ \int f_{\beta_0}^4(z) \pi_{\beta_0}(z | W_i(\beta_0)) \pi_{\beta_0}(z | W_j(\beta_0)) \phi_{ij} dz \right] \times \int K^2(u) du
\end{aligned}$$

where the limit is obtained by standard arguments, using uniform continuity of  $f_{\beta_0}(\cdot)$  and  $\pi_{\beta_0}(\cdot | w)$ .  $\square$

# Conclusion générale et perspectives

Nous avons développé dans cette thèse un nouveau type de statistique pour les tests non paramétriques en régression. La statistique générale, à la frontière entre l'approche utilisant les processus empiriques et celle utilisant le lissage à noyau. En général, on a cherché à réduire la dimension sur laquelle le lissage est effectué, mais tout en la gardant au moins égale à 1, ce qui nous a permis de conserver des valeurs critiques asymptotiquement gaussiennes. Mais dans chaque cas, on aurait pu augmenter la dimension et conserver les mêmes résultats. Le choix de cette dimension est donc laissé à la discrétion du praticien. En effet, dans certains cas, celui-ci peut avoir la connaissance du fait que si violation de l'hypothèse il y a, cela a des chances de l'être à cause d'une ou de plusieurs variables en particulier.

Un des points sur lesquels nous avons peu insisté et qui mériterait d'être approfondi est le choix de la fenêtre. Les articles de Horowitz et Spokoiny (2001 et 2002) donnent une solution pour les modèles paramétriques en régression classique et en régression quantile en reprenant les articles de Zheng (1996 et 1998). Leur idée est de calculer une statistique de test utilisant le lissage à noyau pour une grille de fenêtres et de prendre le maximum de ces statistiques. Le but de cette procédure est d'adapter la statistique de test à la régularité de la fonction  $m$ , et rien n'empêche d'utiliser ce type de technique pour notre statistique. Le principal inconvénient est que cela rend l'utilisation du bootstrap obligatoire pour obtenir les valeurs critiques, et ce quelle que soit la taille de l'échantillon utilisé.

Citons d'autre part l'article de Guerre et Lavergne (2005) et Gao et Gijbels (2008).

On donne dans la Table 5.1 un aperçu des cas envisagés au cours de cette thèse, en croisant :

- le type de régression – classique ou quantile ;
- le type de test – modèle paramétrique, semi-paramétrique ou significativité ;
- la dimension des variables – finie ou infinie ;

et on voit qu'il reste encore beaucoup de cas à envisager. Enfin, comme cela a été mentionné dans le 4, les techniques qui y sont développées ouvrent le champ des tests d'indépendance entre deux variables, ainsi que d'indépendance conditionnelle sachant une troisième variable.

	Régression classique		Régression quantile	
	Dim. finie	Dim. inf.	Dim. finie	Dim. inf.
Modèle paramétrique		4	3	
Modèle semi-paramétrique	5			
Significativité	2			

**TABLE 5.1** – Chemins parcourus et restant à parcourir. La case en haut à gauche est la plus abordée dans la littérature. Les numéros correspondent aux chapitres de la thèse.

# Bibliographie

## A

---

- Jason Abrevaya. The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics*, 26(1) :247–257, 2001.
- Yacine Ait-Sahalia, Peter J. Bickel, and Thomas M. Stoker. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105(2) :363 – 412, 2001.
- Germán Aneiros-Pérez and Philippe Vieu. Semi-functional partial linear regression. *Statistics & Probability Letters*, 76(11) :1102–1110, 2006.

## B

---

- Herman J. Bierens. Consistent model specification tests. *Journal of Econometrics*, 20(1) :105–134, 1982.
- Herman J. Bierens. Model specification testing of time series regressions. *Journal of Econometrics*, 26(3) :323–353, 1984.
- Herman J. Bierens and Donna K. Ginther. Integrated conditional moment testing of quantile regression models. In Bernd Fitzenberger, Roger Koenker, and José A.F. Machado, editors, *Economic Applications of Quantile Regression*, Studies in Empirical Economics, pages 307–324. Physica-Verlag HD, 2002. ISBN 978-3-7908-2502-2.
- Herman J. Bierens and Werner Ploberger. Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65(5) :1129–1151, 1997.
- Salomon Bochner. *Harmonic analysis and the theory of probability*. University of California Press, Berkeley and Los Angeles, 1955.
- Denis Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. ISBN 0-387-95052-4. Theory and applications.

## C

---

- Dong Chen, Peter Hall, and Hans-Georg Müller. Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics*, 39(3) :1720–1747, 2011.

- Song Xi Chen and Ingrid Van Keilegom. A goodness-of-fit test for parametric and semi-parametric models in multiresponse regression. *Bernoulli*, 15(4) :955–976, 2009.
- Xiaohong Chen and Yanqin Fan. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics*, 91(2) :373 – 401, 1999.
- C.-T. Chiang and M.-Y. Huang. New estimation and inference procedures for a single-index conditional distribution model. *Journal of Multivariate Analysis*, 111 :271–285, 2012.
- Jeng-Min Chiou and Hans-Georg Müller. Diagnostics for functional regression via residual processes. *Computational Statistics & Data Analysis*, 51(10) :4849–4863, 2007.
- D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B.*, 34 :187–220, 1972.

## D

- Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Universitext. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-29020-9 ; 3-540-29020-6. Revised and extended from the 2001 original by Da Prato.
- Peter de Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75(2) :261–277, 1987.
- Víctor H. de la Peña and Evarist Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. ISBN 0-387-98616-2. From dependence to independence, Randomly stopped processes.  $U$ -statistics and processes. Martingales and beyond.
- M. Delecroix, M. Hristache, and V. Patilea. Optimal smoothing in semiparametric index approximation of regression functions. CREST Working Paper 9952, 1999.
- M. Delecroix, W. Härdle, and M. Hristache. Efficient estimation in conditional single-index regression. *Journal of Multivariate Analysis*, 86 :213–226, 2003.
- M. Delecroix, M. Hristache, and V. Patilea. On semiparametric  $m$ -estimation in single-index regression. *Journal of Statistical Planning and Inference*, 136 :730–769, 2006.
- Miguel A. Delgado. Testing the equality of nonparametric regression curves. *Statistics & Probability Letters*, 17(3) :199–204, 1993.
- Miguel A. Delgado and Wenceslao González Manteiga. Significance testing in non-parametric regression based on the bootstrap. *The Annals of Statistics*, 29(5) : 1469–1507, 2001.
- Laurent Delsol, Frédéric Ferraty, and Philippe Vieu. Structural test in regression on functional variables. *Journal of Multivariate Analysis*, 102(3) :422–447, 2011.

Holger Dette and Natalie Neumeyer. Nonparametric analysis of covariance. *The Annals of Statistics*, 29(5) :1361–1400, 2001.

## E

---

Juan Carlos Escanciano and Kyungchul Song. Testing single-index restrictions with a focus on average derivatives. *Journal of Econometrics*, 156(2) :377–391, 2010.

## F

---

Jianqing Fan. Test of significance based on wavelet thresholding and Neyman’s truncation. *Journal of the American Statistical Association*, 91(434) :674–688, 1996.

Yanqin Fan and Qi Li. Consistent model specification tests : Omitted variables and semiparametric functional forms. *Econometrica*, 64(4) :865–90, 1996.

Julian J. Faraway. Regression analysis for a functional response. *Technometrics*, 39(3) :254–261, 1997.

Xingdong Feng, Xuming He, and Jianhua Hu. Wild bootstrap for quantile regression. *Biometrika*, 98(4) :995–999, 2011.

Frédéric Ferraty, Ali Laksaci, Amel Tadj, and Philippe Vieu. Kernel regression with functional response. *Electronic Journal of Statistics*, 5 :159–171, 2011.

Frédéric Ferraty, Ingrid Van Keilegom, and Philippe Vieu. Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109 :10–28, 2012.

## G

---

Jiti Gao and Irène Gijbels. Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, 103(484) :1584–1594, 2008.

Eduardo García-Portugués, Wenceslao González-Manteiga, and Manuel Febrero-Bande. A goodness-of-fit test for the functional linear model with scalar response. arXiv :1205.6167 [stat.ME], 2012.

Emmanuel Guerre and Pascal Lavergne. Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, 18(5) :1139–1171, 2002.

Emmanuel Guerre and Pascal Lavergne. Data-driven rate-optimal specification testing in regression models. *The Annals of Statistics*, 33(2) :pp. 840–870, 2005.

---

**H**

---

- P. Hall and Q. Yao. Approximating conditional distribution functions using dimension reduction. *The Annals of Statistics*, 33 :977–1454, 2005.
- Peter Hall and Christopher C. Heyde. *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. ISBN 0-12-319350-8. Probability and Mathematical Statistics.
- Härdle, P. W., Hall, and H. Ichimura. Optimal smoothing in single-index models. *The Annals of Statistics*, 21 :157–178, 1993.
- Wolfgang Härdle and Enno Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4) :1926–1947, 1993.
- Wolfgang Härdle and James Stephen Marron. Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 13(4) :1465–1481, 1985.
- Xuming He and Li-Xing Zhu. A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, 98(464) :1013–1022, 2003.
- J. L. Horowitz. *Semiparametric and Nonparametric Methods in Econometrics*. Springer Series in Statistics. Springer : New-York., 2009.
- Joel L. Horowitz and Vladimir G. Spokoiny. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3) :599–631, 2001.
- Joel L. Horowitz and Vladimir G. Spokoiny. An adaptive, rate-optimal test of linearity for median regression models. *Journal of the American Statistical Association*, 97(459) :822–835, 2002.
- Lajos Horváth and Piotr Kokoszka. *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York, 2012. ISBN 978-1-4614-3654-6.
- M. Hristache, A. Juditsky, and V. Spokoiny. Direct estimation of the index coefficient in a single-index model. *The Annals of Statistics*, 29 :595–917, 2001.
- Simon Hurst. The characteristic function of the student t distribution. Technical report, Center for Financial Mathematics, Canberra, 1995.

**I**

---

- H. Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58 :71–120, 1993.

**J**

---

- Norman L. Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions*. Wiley :New-York, 1995.

## K

---

- Roger Koenker. *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-60827-5 ; 0-521-60827-9.
- Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica*, 46(1) : 33–50, 1978.
- Roger Koenker and Kevin F. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15 :143–156, 2001.
- Piotr Kokoszka, Inga Maslova, Jan Sojka, and Lie Zhu. Testing for lack of dependence in the functional linear model. *The Canadian Journal of Statistics*, 36(2) : 207–222, 2008.
- E. Kong and Y. Xia. A single-index quantile regression model and its estimation. *Econometric Theory*, 28 :730–768, 2012.

## L

---

- Pascal Lavergne. An equality test across nonparametric regressions. *Journal of Econometrics*, 103(1-2) :307–344, 2001. Studies in estimation and testing.
- Pascal Lavergne and Valentin Patilea. One for all and all for one : regression checks with many regressors. *Journal of Business J. Bus. Econom. Statist. Economic Statistics*, 30(1) :41–52, 2012.
- Pascal Lavergne and Quang Vuong. Nonparametric significance testing. *Econometric Theory*, 16(4) :576–601, 2000.
- Pascal Lavergne, Samuel Maistre, and Valentin Patilea. A significance test for covariates in nonparametric regression. arXiv :1403.7063 [math.ST], 2014.

## M

---

- Enno Mammen. Bootstrap and wild bootstrap for high-dimensional linear models. *The Annals of Statistics*, 21(1) :255–285, 1993.
- Eliakim H. Moore. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26 :394–395, 1920.

## N

---

- Natalie Neumeyer and Holger Dette. Nonparametric comparison of regression curves : An empirical process approach. *The Annals of Statistics*, 31(3) :880–920, 2003.



Natalie Neumeier and Holger Dette. Nonparametric comparison of regression curves : an empirical process approach. *The Annals of Statistics*, 31(3) :880–920, 2003.

Deborah Nolan and David Pollard.  $U$ -processes : rates of convergence. *The Annals of Statistics*, 15(2) :780–799, 1987.

## P

---

Ariél Pakes and David Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5) :1027–1057, 1989.

Valentin Patilea, Cesar Sanchez-Sellero, and Matthieu Saumard. Nonparametric testing for no-effect with functional responses and functional covariates. arXiv :1209.2085 [math.ST], 2012a.

Valentin Patilea, Cesar Sanchez-Sellero, and Matthieu Saumard. Projection-based nonparametric goodness-of-fit testing with functional covariates. arXiv :1205.5578 [math.ST], 2012b.

G.A. Picone and J.S. Butler. Semiparametric estimation of multiple equation models. *Econometric Theory*, 26 :551–575, 2000.

J. L. Powell, J. H. Stock, and T. M. Stocker. Semiparametric estimation of index coefficients. *Econometrica*, 57 :1403–1430, 1989.

## R

---

Jeffrey S. Racine. Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics*, 15(3) :pp. 369–378, 1997.

Jeffrey S. Racine and Qi Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1) :99 – 130, 2004.

Jeffrey S. Racine, Jeffrey Hart, and Qi Li. Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews*, 25(4) :523–544, 2006.

James O. Ramsay and Bernard W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005. ISBN 978-0387-40080-8 ; 0-387-40080-X.

## S

---

Robert P. Sherman. Maximal inequalities for degenerate  $U$ -processes with applications to optimization estimators. *The Annals of Statistics*, 22(1) :439–459, 1994.

W. Stute, W. Gonzalez Manteiga, and M. Presedo Quindimil. Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441) :pp. 141–149, 1998a.

Winfried Stute. Nonparametric model checks for regression. *The Annals of Statistics*, 25(2) :613–641, 1997.

Winfried Stute and Li-Xing Zhu. Nonparametric checks for single-index models. *The Annals of Statistics*, 33(3) :pp. 1048–1083, 2005.

Winfried Stute, Silke Thies, and Li-Xing Zhu. Model checks for regression : an innovation process approach. *The Annals of Statistics*, 26(5) :1916–1934, 1998b.

## V

---

Aad W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9 ; 0-521-78450-6.

Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. With applications to statistics.

Aad W. van der Vaart and Jon A. Wellner. A local maximal inequality under uniform entropy. *Electronic Journal of Statistics*, 5 :192–203, 2011.

Stanislav Volgushev, Melanie Birke, Holger Dette, and Natalie Neumeyer. Significance testing in quantile regression. *Electronic Journal of Statistics*, 7 :105–145, 2013.

## X

---

C. Xia, W. Härdle, and L. Zhu. The efm approach for single-index models. *The Annals of Statistics*, 39 :1658–1688, 2011.

Yingcun Xia, Wai Keung Li, Howell Tong, and Dixin Zhang. A goodness-of-fit test for single-index models. *Statistica Sinica*, 14 :1–39, 2004.

## Y

---

Keming Yu, Zudi Lu, and Julian Stander. Quantile regression : applications and current research areas. *The Statistician*, 52(3) :331–350, 2003.

## Z

---

John Xu Zheng. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2) :263–289, 1996.

John Xu Zheng. A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, 14(1) :123–138, 1998.





**RÉSUMÉ :** Dans cette thèse, nous étudions des tests du type :

$$(H_0) : \mathbb{E}[U | X] = 0 \text{ p.s.} \quad \text{contre} \quad (H_1) : \mathbb{P}\{\mathbb{E}[U | X] = 0\} < 1$$

où  $U$  est le résidu de la modélisation d'une variable  $Y$  en fonction de  $X$ . Dans ce cadre et pour plusieurs cas particuliers – significativité de variables, régression quantile, données fonctionnelles, modèle single-index –, nous proposons une statistique de test permettant d'obtenir des valeurs critiques issues d'une loi asymptotique pivotale. Dans chaque cas, nous donnons également une méthode de bootstrap appropriée pour les échantillons de petite taille.

Nous montrons la consistance envers des alternatives locales – ou à la Pitman – des tests proposés, lorsque ce type d'alternative ne tend pas trop vite vers l'hypothèse nulle. À chaque fois, nous vérifions à partir de simulations sous l'hypothèse nulle et sous une séquence d'hypothèses alternatives que les résultats théoriques sont en accord avec la pratique.

**Mots-clés :** tests non paramétriques ; tests omnibus ; significativité de variables ; régression quantile ; données fonctionnelles ; bootstrap sauvage.

**SUMMARY :** In this thesis, we study test statistics of the form :

$$(H_0) : \mathbb{E}[U | X] = 0 \text{ p.s.} \quad \text{contre} \quad (H_1) : \mathbb{P}\{\mathbb{E}[U | X] = 0\} < 1$$

where  $U$  is the residual of some  $Y$  modeling with respect to covariates  $X$ . In this setup and for several particular cases – significance, quantile regression, functional data, single-index model –, we introduce test statistics that have pivotal asymptotic critical values. For each case, we also give a suitable bootstrap procedure for small samples.

We prove the consistency against local – or Pitman – alternatives for the proposed test statistics, when such an alternative does not get close to the null hypothesis too fast. Simulation studies are used to check the effectiveness of the theoretical results in applications.

**Keywords :** nonparametric testing ; omnibus tests ; significance ; quantile regression ; functional data ; wild bootstrap.