Contributions à l'inférence statistique dans les modèles de régression partiellement linéaires additifs

Khalid Chokri

Université Pierre et Marie Curie

Thèse de doctorat de l'université paris 6

Discipline : Mathématiques

Option : Statistique

Sujet de la Thèse

Contributions à l'inférence statistique dans les modèles de régression partiellement linéaires additifs

présentée par

Khalid Chokri

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS 6
Soutenue le 21 novembre 2014

Composition du jury

Directeur de thèse :	Djamal Louani	Université de Reims
Rapporteurs :	Philippe Vieu	Université Paul Sabatier, Toulouse 3
	André Mas	Université Montpellier 2
Examinateurs :	Gérad Biau	Université Paris 6
	Michel Broniatowski	Université Paris 6
	Sophie Dabo-Niang	Université Charles De Gaulle, Lille 3

Laboratoire de Statistique Théorique et Appliquée - UFR 920

Institut de Mathématiques de Jussieu
4 place Jussieu
75 252 Paris cedex 05

École doctorale Paris centre Case 188
4 place Jussieu
75 252 Paris cedex 05

Contributions à l'inférence statistique dans les modèles de régression partiellement linéaires additifs

Résumé

Les modèles de régression paramétrique fournissent de puissants outils pour la modélisation des données lorsque celles-ci s'y prêtent bien. Cependant, ces modèles peuvent être la source d'importants biais lorsqu'ils ne sont pas adéquats. Pour éliminer ces biais de modélisation, des méthodes non paramétriques ont été introduites permettant aux données elles mêmes de construire le modèle. Ces méthodes présentent, dans le cas multivarié, un handicap connu sous l'appellation de fléau de la dimension où la vitesse de convergence des estimateurs est une fonction décroissante de la dimension des covariables. L'idée est alors de combiner une partie linéaire avec une partie non-linéaire, ce qui aurait comme effet de réduire l'impact du fléau de la dimension. Néanmoins l'estimation non-paramétrique de la partie non-linéaire, lorsque celle-ci est multivariée, est soumise à la même contrainte de détérioration de sa vitesse de convergence. Pour pallier ce problème, la réponse adéquate est l'introduction d'une structure additive de la partie non-linéaire de son estimation par des méthodes appropriées. Cela permet alors de définir des modèles de régression partièllement linéaires et additifs. L'objet de la thèse est d'établir des résultats asymptotiques relatifs aux divers paramètres de ce modèle (consistance, vitesses de convergence, normalité asymptotique et loi du logarithme itéré) et de construire aussi des tests d'hypothèses relatives à la structure du modèle, comme l'additivité de la partie non-linéaire, et à ses paramètres.

Mots-clefs : Modèle additif, test d'additivité, normalité asymptotique, fléau de la dimension, loi du logarithme itéré, estimateur à noyau, intégration marginale, modèle additif partiellement linéaire.

Contributions to the statistical inference in partially linear additive regression model

Abstract

Parametric regression models provide powerful tools for analyzing practical data when the models are correctly specified, but may suffer from large modelling biases when structures of the models are misspecified. As an alternative, nonparametric smoothing methods eases the concerns on modelling biases. However, nonparametric models are hampered by the so-called curse of dimensionality in multivariate settings. One of the methods for attenuating this difficulty is to model covariate effects via a partially linear structure, a combination of linear and nonlinear parts. To reduce the dimension impact in the estimation of the nonlinear part of the partially linear regression model, we introduce an additive structure of this part which induces, finally, a partially linear additive model. Our aim in this work is to establish some limit results pertaining to various parameters of the model (consistency, rate of convergence, asymptotic normality and iterated logarithm law) and to construct some hypotheses testing procedures related to the model structure, as the additivity of the nonlinear part, and to its parameters.

Keywords: Additive model, additivity test, asymptotic normality, curse of dimensionality, iterated logarithm law, kernel estimator, marginal integration, partially linear additive model.

Liste des travaux

[1] Bouzebda, S. and Chokri, K. (2014). Statistical tests in the partially linear additive regression models. *Statistical Methodology*. **19**, 4–24.

[2] Bouzebda, S., Chokri, K. and Louani, D. (2014). Some uniform consistency results in the partially linear additive model components estimation. *Comm. Statist. Theory Methods*. **43**. Accepted.

[3] Chokri, K. and Louani, D. (2013). Additivity test on the nonlinear part in partially linear models. *C. R. Math. Sci. Paris, Ser.I* **351**, 143–148.

[4] Chokri, K. and Louani, D. (2011). Asymptotic results for the linear parameter estimate in partially linear additive regression model. *C. R. Math. Sci. Paris, Ser.I* **349**, 1105–1109.

[5] Chokri, K. and Louani, D. Some asymptotic results with an additivity test in the partially linear regression model. *Journal Of Nonparametric Statistics.* Submitted.

Table des matières

Notations

$\widehat{\boldsymbol{\beta}}$	Estimateur des moindres carrés du paramètre $\boldsymbol{\beta}$.
$\mathbb{E}(Y\|\mathbf{X})$	Espérence conditionnelle de Y sachant X.
$\hat{m}_n^{NW}$	Estimateur de Nadaraya-Watson de la fonction de régression m.
$\mathbf{g}_{Y\|\mathbf{X}}$	Densité conditionnelle de Y sachant $\mathbf{X}$.
MSE	Erreur quadratique moyenne.
$\zeta_l,\ \eta_l$	$l^{ème}$ composante additive.
$\mathcal{X}^2_{(p)}$	Loi khi-deux à p degrés de liberté.
Var	Variance.
X_{-l}	Vecteur aléatoir dépourvu de la $l^{ème}$ composante.
ϕ^{-1}	Quantile de la loi normale centrée réduite.
$\mathbb{1}_A$	Fonction indicatrice de A.
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$\mathcal{C}^k$	Ensemble des fonctions k-fois continuement dérivable.
$\mathcal{N}(0,1)$	Loi normale centrée réduite.
$\mathcal{M}$	Classe des fonctions additives.
$\log_{10}$	logarithme décimal.
$\mathcal{O},\ o$	Notations de landau.
m_{add}	Fonction de régression additive.
$ess\sup$	Borne supérieure essentielle.
$\mathcal{M}_{n,n}(\mathbb{R})$	Ensemble des matices réelle $n \times n$.
$i.i.d$	Indépendante et identiquement distribuée.
$a.s.$	Almost surely.

Chapitre 1

Introduction

1.1 Estimation de la régression

La régression est une branche de la statistique mathématique qui développe des techniques et des modèles pour décrire la relation d'une variable par rapport à une ou plusieurs autres variables. En particulier, la relation « cause et effet »... Nous exposons dans les deux sections suivantes les méthodes d'estimation, à savoir, estimation paramétrique et estimation non paramétrique.

1.1.1 Estimation paramétrique

L'estimation paramétrique est l'approche classique pour estimer une fonction de régression. Ici, on suppose que la structure de la fonction de régression est connue et ne dépend que d'un nombre fini de paramètres, qu'on estime à partir d'un échantillon donné. Considérons le modèle linéaire suivant :

$$Y = \boldsymbol{\beta}_0 + \sum_{i=1}^{d} \boldsymbol{\beta}_i X_i + \varepsilon, \qquad (X_1, \ldots, X_d) \in \mathbb{R}^d,$$

où les X_i sont appelées covariables, ε représente l'erreur du modèle, souvent supposée independante de $\mathbf{X}$ et de loi normale centrée et de variance finie σ^2, et, $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ désignent les paramètres du modèle, supposés inconnus. La fonction de régression linéaire est définie par

$$m(X_1, \ldots, X_d) = \boldsymbol{\beta}_0 + \sum_{i=1}^{d} \boldsymbol{\beta}_i X_i, \qquad (X_1, \ldots, X_d) \in \mathbb{R}^d.$$

L'estimateur du vecteur inconnu $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ est obtenu par l'utilisation de la méthode des moindres carrés, donné par

$$(\hat{\boldsymbol{\beta}}_0, \ldots, \hat{\boldsymbol{\beta}}_d) = \arg \min_{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{j=1}^{n} \left| Y_j - \boldsymbol{\beta}_0 - \sum_{i=1}^{d} \boldsymbol{\beta}_i X_{ij} \right|^2 \right\},$$

où X_{ij} désigne la $i^{ème}$ composante de la covariable X_j. Ensuite, on défini l'estimateur de la régression linéaire par

$$\hat{m}_n(X_1, \ldots, X_d) = \hat{\boldsymbol{\beta}}_0 + \sum_{i=1}^{d} \hat{\boldsymbol{\beta}}_i X_i, \qquad (X_1, \ldots, X_d) \in \mathbb{R}^d.$$

Une littérature abondante s'est développée sur le sujet, on pourra consulter, entre autres,Prakasa Rao (1983), Seber (1977), Draper and Smith (1981), Farebrother (1988), Wolverton and Wagner (1969), Győrfi (1978), Devroye (1976).
Les modèles de régression paramétrique sont, en général, faciles à interpréter et ils fournissent de puissants outils pour la modélisation des données lorsque le modèle est bien spécifié. Cependant, ces modèles peuvent être la source d'importants biais lorsqu'ils sont inadéquats ou mal spécifiés. Afin de pallier ce problème, des mèthodes non paramétriques ont été introduites pour construire des modèles à partir des données constituant ainsi une alternative générale et intéressante.

1.1.2 Estimation nonparamétrique

Soient $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ des vecteurs aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, indépendants et identiquement distribués. La fonction de régression, lorsqu'elle existe et finie, est définie par

$$m(\mathbf{x}) = \mathbb{E}(Y | \mathbf{X} = \mathbf{x}).$$

L'un des estimateurs les plus utilisés pour la régression est celui Nadaraya et Watson (1964) défini, pour $\mathbf{x} \in \mathbb{R}^d$, par

$$\hat{m}_n^{NW}(\mathbf{x}) := \begin{cases} \frac{\sum_{i=1}^{n} Y_i K(\frac{\mathbf{x}-\mathbf{X}_i}{h_n})}{\sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{h_n})}, & \text{si } \sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}) \neq 0, \\ \frac{1}{n} \sum_{i=1}^{n} Y_i, & \text{si } \sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}) = 0. \end{cases}$$

où h_n est le paramètre de lissage, appelé aussi fenêtre d'estimation. une suite de nombres strictement positifs telle que $h_n \to 0 \quad$ lorsque $\quad n \to \infty$, et la fonction réelle K est ici le

noyau de convolution de Parzen-Rosenblatt. Cela se justifie par le fait que la fonction de régression s'écrit sous la forme

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}(Y|\mathbf{X}=\mathbf{x}) = \int y\mathbf{g}_{Y|\mathbf{X}}(y|\mathbf{x})dy \\ &= \int y\frac{\mathbf{g}_{\mathbf{X},Y}(\mathbf{x},y)}{\mathbf{g}_{\mathbf{X}}(\mathbf{x})}dy, \end{aligned}$$

où $\mathbf{g}_{\mathbf{X}}$ désigne la densité marginale de la covariable $\mathbf{X}$, $\mathbf{g}_{\mathbf{X},Y}$ est la densité jointe du couple $(\mathbf{X}, Y)$, et $\mathbf{g}_{Y|\mathbf{X}}$ représente la densité conditionnelle de Y sachant $\mathbf{X}$. L'estimateur de Nadaraya-Watson est obtenu après estimation des densités par la méthode du noyau. Rappelons qu'un estimateur de la densité par la méthode du noyau est donné, pour tout $\mathbf{x} \in \mathbb{R}^d$, par

$$\mathbf{g}_n(\mathbf{x}) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right). \tag{1.1}$$

Les comportements asymptotiques des estimateurs non paramétriques de la densité, de la fonction de régression ou d'autres fonctionnelles, comprenant la consistance et les lois limites, sont décrits dans de nombreux ouvrages et articles. Outre les méthodes d'estimation, la nature des données, comme leurs structures de dépendance (mélange et ergodicité) ou les méthodes d'observation (censure, troncature, données fonctionnelles ou temps continu), a aussi motivé de nombreuses recherche. Nous renvoyons, entre autres, aux travaux de Vapnik and Červonenkis (1971), Gasser and Müller (1979) Priestley and Chao (1972), Collomb (1980) pour l'estimation de la régression à l'aide de la méhode du noyau. Le livre de Prakasa Rao (1983) donne un aperçu assez général des méthodes et des résultats dans le cas réel et vectoriel. Nous citons aussi comme références importantes dans ce domaine Silverman (1986), Bosq and Lecoutre (1987), Fan and Gijbels (1995), Bosq (1996), Delecroix and Rosa (1996) et Györfi and Kohler (2002).

Le résultat suivant, qui donne la consistance de l'estimateur de Nadaraya-Watson avec une vitesse de convergence et l'ordre du biais, revêt un caractère important pour l'estimation de la fonction de régression.

Proposition 1.1. *Supposons que K est un noyau d'ordre k défini sur $\mathbb{R}^d$, borné et à support compact, Y est bornée, $h_n \to 0$, $nh_n^d \to \infty$, lorsque $n \to \infty$, $\mathbf{g}_{\mathbf{X}}$, $\mathbf{g}_{\mathbf{X},Y}$ sont de classe $\mathcal{C}^k$. On a alors*

$$\begin{aligned} &\mathbb{E}[\hat{m}_n^{NW}(\mathbf{x})] - m(\mathbf{x}) = \mathcal{O}\left(h_n^k\right) \\ &\mathbb{E}[(\hat{m}_n^{NW}(\mathbf{x}) - \mathbb{E}[\hat{m}_n^{NW}(\mathbf{x})])^2] = \mathcal{O}\left(\frac{1}{nh_n^d}\right). \end{aligned}$$

En effet, on remarque d'après la Proposition 1.1 que la consistance et le biais de l'estimateur $\hat{m}_n^{NW}$ dépendent du paramètre de lissage h_n. Un choix optimal de la fenêtre dans le terme du biais donne un mauvais résultat pour la variance, et vice versa ; c'est ce qu'on appelle "Tradeoff". Il s'agit alors de considérer un équilibre entre le terme de la variance et celui du biais pour le choix optimal du paramètre de lissage. Naturellement, on utilise ici le critère de l'erreur quadratique moyenne, noté "MSE", pour ce choix. D'autres critères de choix du paramètre de lissage, comme le critère de validation croisée, sont aussi utilisés.

Rappelons que l'erreur quadratique moyenne de l'estimateur $\hat{m}_n^{NW}$ est donnée par

$$\begin{aligned} MSE_{\mathbf{x}}[\hat{m}_n^{NW}] &= \mathbb{E}[(\hat{m}_n^{NW}(\mathbf{x}) - m(\mathbf{x}))^2] \\ &= \mathbb{E}[\hat{m}_n^{NW}(\mathbf{x}) - \mathbb{E}[\hat{m}_n^{NW}(\mathbf{x})]]^2 + [\mathbb{E}(\hat{m}_n^{NW}(\mathbf{x}) - m(\mathbf{x}))]^2 \\ &= \mathcal{O}\left(\frac{1}{nh_n^d}\right) + \mathcal{O}\left(h_n^{2k}\right). \end{aligned}$$

La fenêtre optimale ($h_n^{opt} = \mathcal{O}(n^{-\frac{1}{2k+d}})$), qui minimise l'erreur quadratique moyenne, nous permet de donner la vitesse suivante

$$n^{\frac{2k}{2k+d}}\mathbb{E}[(\hat{m}_n^{NW}(\mathbf{x}) - m(\mathbf{x}))^2] = \mathcal{O}(1).$$

Lorsque la dimension devient importante, cette vitesse se détériore rapidement. Ce phénomène causé par la dimension de la covariable $\mathbf{X}$ est connu sous l'appellation de « Fléau de la dimension ».

L'activité humaine d'aujourd'hui fournit des données qui sont chaque jour plus nombreuses et de plus grande dimension. Dans tous les secteurs d'activité, les bases de données n'arrêtent de grossir et le nombre de paramètres observés n'arrête de grandir. Cela conduit inéluctablement à des problèmes liés à leur dimension dans toute étude ou analyse de données. Par ailleurs, le traitement manuel de ces données n'etant pas envisageable, il est d'un intérêt majeur de pouvoir traiter correctement de façon automatique de telles données et de développer des méthodes appropriées pour cela. Dans un futur proche, on aura des données de plus grande taille et en nombre encore plus important que les données déjà disponibles.

1.2 Le fléau de la dimension : Définition et solutions

Le fléau de la dimension "Curse of dimensionality" fait référence au handicap des méthodes locales de lissage dans le cas des données multivariées. À cause de la parcimonie des données dans les espaces multidimensionnels, le comportement des estimateurs non paramétriques se détériore rapidement quand la dimension augmente. Ce

phénomène a été découvert en 1957 par le mathématicien américain Richard Bellman dans son ouvrage intitulé “Dynamic programming”, et fut un argument en faveur de la programmation dynamique. En effet, pour un échantillon uniformément réparti sur un pavé $[0,1]^d$ et une fenêtre h_n de longueur 0.1, le pourcentage des données qui se trouvent dans chaque hypercube d'arrête $1/10$ est de $10^{(2-d)}\%$, on remarque clairement une décroissance significative en fonction de la dimension de ce pourcentage. Silverman (1986) a abordé le problème en exploitant le nombre requis d'un échantillon dans le cadre de l'approximation d'une distribution gaussienne avec des noyaux gaussiens fixes. Et il a également montré que la taille de l'échantillon, avec une erreur qui ne dépasse pas 10%, croît exponentiellement en fonction de la dimension (voir Figure 1.1), et il a donné l'approximation suivante

$$\log_{10} N(d) = 0.6(d - 0.25),$$

avec $N(d)$ désignant la taille de l'échantillon et d étant la dimension de l'espace.

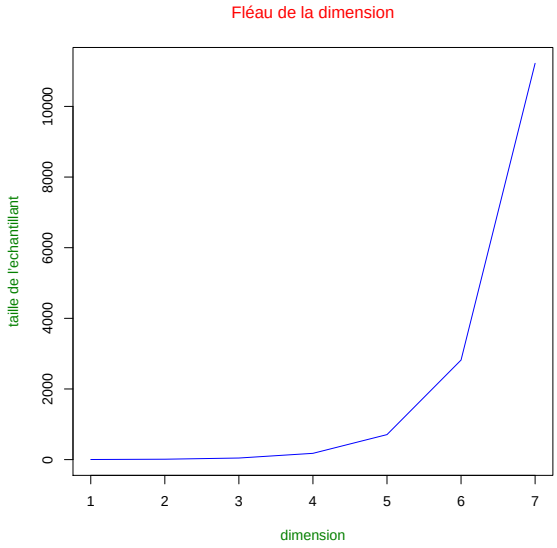

FIG. 1.1 – Taille d'échantillon requis pour approcher une distribution gaussienne avec des noyaux gaussiens fixes, et une erreur approximative d'environ 10%, par rapport à la dimension de l'espace.

Dans la pratique, si pour tout ensemble de données, cet ensemble ne grandit pas de manière exponentielle avec la dimension de l'espace alors cette dimension sera dite petite. Inversement, la dimension sera qualifiée de grande. Naturellement, la question qu'on peut poser est la suivante : Quelle est la limite entre petite dimension et grande dimension ?

Scott and Thompson (1983) ont trouvé un autre moyen pour aborder le concept de la grande dimension en introduisant le phenomène de l'espace vide "The empty space phenomenon". Le volume d'une sphère de rayon r en dimension d est donné par

$$V(d) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} r^d.$$

Dans le cas d'une hypersphère unité, ce volume peut être défini d'une manière équivalente à l'aide de la formule de récurrence suivante

$$V(d) = \frac{2\pi}{d} V(d-2),$$

où $V(1) = 2$ et $V(2) = \pi$. La simulation ci-dessous (figure 1.2) montre comment le volume de l'hypersphère unité décroît en fonction de la dimension de l'espace. On remarque que ce volume se rapproche de plus en plus de la valeur zéro dès qu'on dépasse la dimension 20.

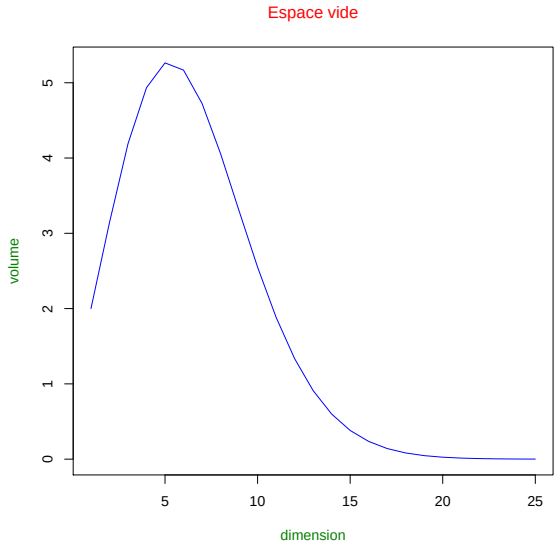

FIG. 1.2 – Volume de l'hypersphère unité en fonction de la dimension de l'espace.

Pour conclure, et afin de mieux appréhender ce concept, il est souhaitable de comparer le volume d'une hypersphère de rayon 0.9 à une autre hypersphère plus naturelle pour nous. On va choisir comme exemple l'hypersphère unité. Le graphe suivant (Figure 1.3) donne le rapport des volumes et son évolution en fonction de la dimension de l'espace.

Les valeurs observées dans la Figure 1.3 montrent que, pour une dimension supérieure à 20, tout le volume de l'hypersphère de rayon 0.9 est contenu dans seulement 10% de celui de l'hypersphère unité.

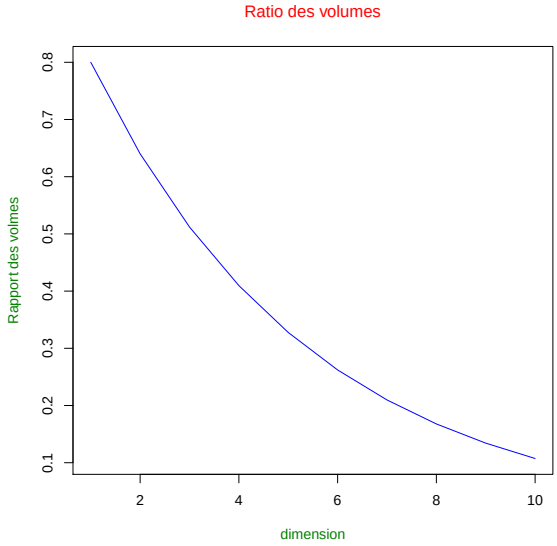

FIG. 1.3 – Rapport des Volumes entre l'hypersphère de rayon 0.9 et l'hypersphère unité en fonction de la dimension de l'espace.

1.2.1 Le modèle semi-paramétrique

Soient $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ et $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ deux échantillons de variables explicatives à valeurs respectives dans $\mathbb{R}^d$ et $\mathbb{R}^p$. Ici, les vecteurs $((\mathbf{X}_1, \mathbf{Z}_1), \ldots, (\mathbf{X}_n, \mathbf{Z}_n))$ sont des répliques indépendantes du couple $(\mathbf{X}, \mathbf{Z})$. Soit $(Y_1, \ldots, Y_n)$ un échantillon de variables expliquées, appelées aussi réponse, à valeurs dans $\mathbb{R}$. Le modèle semi-paramétrique est défini par

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + m(\mathbf{X}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.2}$$

où $\boldsymbol{\beta}$ est un paramètre inconnu p-dimensionnel, $\mathbf{Z}^\top$ représente la transposée du vecteur $\mathbf{Z}$, m est une fonction nonlinéaire à plusieurs variables à valeurs dans $\mathbb{R}$ et $\varepsilon_1, \ldots, \varepsilon_n$ est un échantillon aléatoire de moyenne nulle et de variance fini σ^2. Nous supposerons par la suite une condition supplémentaire d'indépendance de la varaible ε_i par rapport au vecteur aléatoire $(\mathbf{X}_i, Z_i)$.

Estimation du paramètre $\boldsymbol{\beta}$

Pour tout $i \in \{1, \ldots, n\}$, on a

$$\mathbb{E}[Y_i|\mathbf{X}_i] = \mathbb{E}[\mathbf{Z}_i^T|\mathbf{X}_i]\boldsymbol{\beta} + m(\mathbf{X}_i).$$

Il ressort alors

$$Y_i - \mathbb{E}[Y_i|\mathbf{X}_i] = (\mathbf{Z}_i - \mathbb{E}[\mathbf{Z}_i|\mathbf{X}_i])^T\boldsymbol{\beta} + \varepsilon_i.$$

En utilisant la méthode des moindres carrés, on en déduit que

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}(\mathbf{Z}_i - \mathbb{E}[\mathbf{Z}_i|\mathbf{X}_i])(\mathbf{Z}_i - \mathbb{E}[\mathbf{Z}_i|\mathbf{X}_i])^T\right)^{-1}\left(\sum_{i=1}^{n}(\mathbf{Z}_i - \mathbb{E}[\mathbf{Z}_i|\mathbf{X}_i])(Y_i - \mathbb{E}[Y_i|\mathbf{X}_i])^T\right).$$

Naturellement, les quantités $\mathbb{E}[\mathbf{Z}_i|\mathbf{X}_i]$ et $\mathbb{E}[Y_i|\mathbf{X}_i]$ sont inconnues. Soient $\hat{\mathbf{Z}}_i$ et $\hat{Y}_i$ leurs estimateurs respectifs obtenus par la méthode du noyau. Par conséquent, l'estimateur de $\boldsymbol{\beta}$, peut s'écrire sous la forme suivante

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n}(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)^T\right)^{-1}\left(\sum_{i=1}^{n}(\mathbf{Z}_i - \hat{\mathbf{Z}}_i)(Y_i - \hat{Y}_i)^T\right).$$

Estimation de la partie nonparamétrique du modèle

Le modèle (1.2) peut s'écrire d'une manière équivalente sous la forme suivante

$$Y_i - \mathbf{Z}_i^\top\boldsymbol{\beta} = m(\mathbf{X}_i) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

Par conséquent, un estimateur de Nadaraya-Watson de la fonction m est donné par

$$\widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) = \sum_{i=1}^{n}\frac{Y_i - \mathbf{Z}_i^\top\widehat{\boldsymbol{\beta}}}{nh_n^d\mathbf{g}_n(\mathbf{x})}K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right).$$

L'intérêt principal du modèle partiellement linéaire est qu'il permet de distinguer les relations linéaires et les relations non linéaires du modèle. L'idée est de prendre en compte l'a priori, que l'on a quant à la linéarité de certaines relations afin de réduire le coût de l'estimation qu'aurait un modèle non paramétrique, tout en gardant l'effet sous-jacente au modèle nonparamétrique pour expliquer les autres relations. Le modèle semi-paramétrique a été introduit pour la première fois en (1985) par Green et Yandell. Mais c'est juste après l'étude faite en (1986) par Engle, Granger, Rice et Weiss, sur la relation entre la vente de l'électricité et les conditions météorologiques, que ce modèle a commencé à intéresser les chercheurs. On consultera, entre autres, à ce sujet, Green (1985), Green and Yandell (1985), Eubank (1985), Heckman (1986, 1988), Speckman (1988), Horng Shiau and Wahba (1988), Wahba (1986) et Mammen et al. (2011).

1.2.2 Le modèle additif

Le modèle partiellement linéaire permet de modéliser un très grand nombre de phénomènes. Sa flexibilité par rapport au modèle linéaire standard et ces résultats satisfaisants sur l'estimation statistique ont contribué à son succès. Néanmoins, la partie nonparamétrique du modèle partiellement linéaire posera encore des problèmes, en particulier, pour la vitesse de convergence des estimateurs, dûs au fléau de la dimension. Pour pallier ce problème, on considère une structure additive de la fonction de regression m. On pose

$$Y_i = \sum_{l=1}^{d} m_l(X_{il}) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

où X_{il} est la $l^{\text{ième}}$ composante du vecteur $\mathbf{X}_i$ et m_l sont des fonctions réelles univariées inconnues. Compte tenu du problème d'identifiabilité, par la suite, on supposera que

$$\mathbb{E}(\mathbf{X}_{il}) = 0, \quad l = 1, \ldots, d, \quad i = 1, \ldots, n.$$

Stone (1985) a montré que l'introduction du modèle additif nous permet de résoudre le problème de la dimension dans le sens que l'on obtient des vitesses similaires à celles de la régression univariée.

De nombreuses méthodes ont été proposées pour estimer les composantes additves et donner les propriétés asymptotiques de leur estimateurs. Il s'agit notamment de la méthode algorithmique du Backfitting ordinaire employée par Buja et al. (1989), dont les proprétés théoriques ont été étudiées plus tard par Opsomer and Ruppert (1997). Mammen et al. (1999) ont proposé une nouvelle procedure du Backfitting dite « Smooth Backfitting »pour laquelle ils ont développé toute une théorie asymptotique. Cette procedure est ensuite reprise par Nielsen and Sperlich (2005) qui ont développé ses aspects pratiques. D'autres variantes de la méthode du Backfitting ont fait l'objet de plusieurs travaux, on peut citer ceux de Hastie and Tibshirani (1990) et Mammen and Park (2006). Comme alternative aux méthodes du Backfitting, la méthode d'intégration marginale a été étudiée par de nombreuses auteurs parmis lesquels nous citons Newey (1994), Tjøstheim and Auestad (1994), Linton and Nielsen (1995) et Fan et al. (1998). Sur la base de cette méthode d'estimation, le travail de cette thèse s'articule sur l'étude des propriétés asymptotiques des estimateurs du paramètre linéaire, de la variance des erreurs et des composantes additives du modèle additif partiellement linéaire ainsi que la construction de tests d'hypothèses relatives aux paramètres de ce modèle et à la structure d'additivité supposée de la composante non linéaire. Nos résultats s'inscrivent dans l'utilisation de la méthode d'intégration marginale pour

l'étude des modèles additifs partiellement linéaires et viennent en complément des travaux disponibles dans la litérature sur les modèles additifs de regression et les modèles partiellement linéaires.

Dans le paragraphe suivant, nous présentons le principe de cette méthode pour la bonne compréhension de notre étude.

1.3 La méthode d'intégration marginale

Dans ce paragraphe, on introduit la méthode d'intégration marginale pour estimer les composantes additives m_l, $1 \leq l \leq d$. Cette méthode est moins compliquée que l'algorithme du backfitting. Ce dernier s'est avéré moins facile à étendre aux études thèoriques sur les propriétés asymptotiques de l'estimateur associé, malgré l'aisance de l'implémentation de cet algorithem itératif, le plus adopté en pratique pour l'estimation des composantes additives. Pour plus de détails sur le sujet, on consultera, par exemple, les travaux de Mammen et al. (1999) et Lee et al. (2010).

1.3.1 Détail de la méthode d'intégration marginale

En premier lieu, considérons que $d \geq 2$ et que la fonction de régression m admet la forme additive suivante, pour tout $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$m(\mathbf{x}) = \sum_{l=1}^{d} m_l(x_l).$$

Soit q_l une densité réelle univariée, pour tout $1 \leq l \leq d$. Posons $\mathbf{q}(\mathbf{x}) = \prod_{l=1}^{d} q_l(x_l)$ et $\mathbf{q}_{-l}(\mathbf{x}_{-l}) = \prod_{j=1, j\neq l}^{d} q_j(x_j)$, où $\mathbf{x}_{-l} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d)$, des densités d'intégration définies respectivement sur $\mathbb{R}^{d-1}$ et $\mathbb{R}^d$, par conéquent, la fonction de régression m peut s'écrire sous la forme, pour tout $\mathbf{x} \in \mathbb{R}^d$,

$$m(\mathbf{x}) = \sum_{l=1}^{d} \zeta_l(x_l) + \int_{\mathbb{R}^d} m(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}, \tag{1.3}$$

où

$$\zeta_l(x_l) = \int_{\mathbb{R}^{d-1}} m(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} m(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{1.4}$$

Remark 1.1. Par intégration de la fonction de régression m, on en déduit que

$$\int_{\mathbb{R}^{d-1}} m(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} = m_l(\mathbf{x}_l) + \sum_{i\neq l}^{d} \int_{\mathbb{R}} m_i(x_i)q_i(x_i)dx_i.$$

Remarquons, ensuite, qu'on a l'égalité

$$\int_{\mathbb{R}^d} m(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x} = \sum_{l=1}^{d} \int_{\mathbb{R}} m_l(x_l)q_l(x_l)dx_l.$$

Compte tenu des deux équations précédentes, nous pouvons écrire

$$\begin{aligned} & \int_{\mathbb{R}^{d-1}} m(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} m(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x} \\ = \; & m_l(x_l) + \sum_{i\neq l}^{d} \int_{\mathbb{R}} m_i(x_i)q_i(x_i)dx_i - \sum_{l=1}^{d} \int_{\mathbb{R}} m_l(x_l)q_l(x_l)dx_l \\ = \; & m_l(x_l) - \int_{\mathbb{R}} m_l(x_l)q_l(x_l)dx_l. \end{aligned}$$

Par conséquent, il est facile de voir que

$$\zeta_l(x_l) = m_l(x_l) - \int_{\mathbb{R}} m_l(x_l)q_l(x_l)dx_l. \tag{1.5}$$

D'où l'équation (1.3).

Remark 1.2. En cohérence avec l'équation (1.5), on observe que, pour tout $l = 1, \ldots, d$, ζ_l et m_l sont des composantes additives de la fonction de régression m qui sont égales à une constante près.

Remark 1.3. En tenant compte de la condition d'identifiabilité $\mathbb{E}m_l(x_l) = 0$, pour tout $l = 1, \ldots, d$, et dans le cas où les densités d'intégration q_l sont égales aux fonctions densité f_l de la variable X_l, on obtient $\zeta_l = m_l$. Ce cas de figure n'est généralement pas vérifié, car la fonction densité f n'est pas connue ; ce qui implique que $\zeta_l \neq m_l$.

Par intégration de la fonction de régression multivariée, on peut en déduire, pour tout $l = 1, \ldots, d$, que

$$\widehat{\zeta}_l^{\boldsymbol{\beta}}(x_l) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{1.6}$$

Par conséquent, un estimateur de la fonction de régression m, dans le cadre additif, est donné par

$$\widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{l=1}^{d} \widehat{\zeta}_l^{\boldsymbol{\beta}}(x_l) + \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{z})\mathbf{q}(\mathbf{z})d\mathbf{z}. \tag{1.7}$$

1.4 Résultats existants

Dans ce paragraphe, nous allons donner quelques résultats scindés en deux sous-sections. La première est relative au modèle semi-paramétrique tandis que la deuxième concerne le modèle additif. Ces résultats seront présentés sous forme de théorèmes.

1.4.1 Le cadre semi-paramétrique

Supposons que l'on observe une suite de copies indépendantes $(Y^1, \mathbf{X}^1, \mathbf{Z}^1), \cdots, (Y^n, \mathbf{X}^n, \mathbf{Z}^n)$, du vecteur aléatoire $(Y, \mathbf{X}, \mathbf{Z})$, où $\mathbf{X} = (X_1, \cdots, X_p)^\top \in \mathbb{R}^p$ et $\mathbf{Z} = (Z_1, \cdots, Z_p)^\top \in \mathbb{R}^d$. On considère le modèle additif partièllement linéaire suivant

$$Y = \mathbf{X}^\top \boldsymbol{\beta} + m_1(Z_1) + \cdots + m_d(Z_d) + \varepsilon.$$

Pour des raisons d'identifiabilité des composantes additives, nous posons la contrainte $\mathbb{E}[m_j(Z_j)] = 0$, $1 \leq j \leq d$. Ici la variable ε est supposée indépendante des variables $\mathbf{X}$ et $\mathbf{Z}$. Soit $\widehat{\boldsymbol{\beta}}$ un estimateur de $\boldsymbol{\beta}$ définit par

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \widetilde{\mathbf{X}}^i \widetilde{\mathbf{X}}^{i\top} \right)^{-1} \left(\sum_{i=1}^{n} \widetilde{\mathbf{X}}^i \widetilde{Y}^i \right)$$

où $\widetilde{\mathbf{X}}^i = \mathbf{X}^i - \widehat{m}_{\mathbf{X}}^{add}(\mathbf{Z}^i)$ et $\widetilde{Y}^i = Y^i - \widehat{m}_Y^{add}(\mathbf{Z}^i)$. Les quantités $\widehat{m}_{\mathbf{X}}^{add}(\mathbf{Z}^i)$ et $\widehat{m}_Y^{add}(\mathbf{Z}^i)$ sont pris ici comme des estimateurs obtenus à l'aide de la méthode du Backfitting. On pose $\eta(\mathbf{Z}) = \Pi(\mathbb{E}(\mathbf{X}|\mathbf{Z})|\mathcal{H})$, où $\Pi(\cdot|\mathcal{H})$ est un opérateur de projection sur l'espace $\mathcal{H}$ des fonctions additives m de carré intégrable tel que $\mathbb{E}m(\mathbf{X}) = 0$. Alors, sous certaines conditions de régularité, Mammem *et al.* (2011) établissent le résultat suivant.

Theorem 1.1. Mammen et al. (2011)
Pour C > 0 et $1 \leq j \leq p$, on suppose que $\mathbb{E}[\exp(|X_j - \mathbb{E}(X_j|Z)|Z)] < C$ p.s.. Si $h_j = \mathcal{O}(n^{-\alpha})$, avec $1/5 \leq \alpha < 1/2$, alors on a

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathrm{var}(\varepsilon)\left[E(\mathbf{X} - \eta(\mathbf{Z}))(\mathbf{X} - \eta(\mathbf{Z}))^\top\right]^{-1}\right),$$

Ici, loi de la variabe ε n'est pas nécessairement connue. Suivant le travail de Mammen et al. (2011), la borne inférieure de l'information de Fisher pour estimer le paramètre $\boldsymbol{\beta}$ est donnée par

$$\mathcal{I}_{inf} = \mathcal{I}_g . \mathbb{E}[\mathbf{X} - \eta(\mathbf{Z})][\mathbf{X} - \eta(\mathbf{Z})]^\top,$$

où $\mathcal{I}_g = \int (g')^2/g < \infty$ et g est la densité de la variable ε. Comme la variance asymptotique de l'estimateur $\widehat{\boldsymbol{\beta}}$ est plus grande que la borne inférieure de l'information de

Fisher $\mathcal{I}_{inf}$, on a $\text{var}(\varepsilon) \geq \mathcal{I}_g^{-1}$. Dans le cas d'erreurs Gaussiennes, le résultat obtenu est $\text{var}(\varepsilon) = \mathcal{I}_g^{-1}$. Ainsi, l'estimateur $\widehat{\boldsymbol{\beta}}$ atteint la borne inférieure de l'eficacité semiparamétrique.

Soit $\{(Y_i, X_{i1}, \ldots, X_{ip}, T_i)\}_{i=1}^n$ une suite i.i.d. de $n(p+2)$ répliques aléatoires indépendantes de même loi que $(Y, X_1, \ldots, X_p, T)$, avec Y une variable aléatoire qui dépend simultanément d'une variable explicative réelle X_j $(j = 1, \ldots, p)$ et une variable explicative fonctionnelle T. On considère le modèle semi-fonctionnel partiellement linéaire suivant

$$Y_i = \sum_{j=1}^{p} X_{ij}\boldsymbol{\beta}_j + m(T_i) + \varepsilon_i \quad \forall i = 1, \ldots, n.$$

Les estimateurs de $\boldsymbol{\beta}$ et m sont obtenus successivement à l'aide de la méthode des moindres carrés et celle de Nadaraya-Watson, et seront notés par $\widehat{\boldsymbol{\beta}}_h$ et $\widehat{m}_h$. Sous certaines conditions de régularité (voir Aneiros-Pérez and Vieu (2008)), on a la consistance de l'estimateur de la partie non paramétrique, la normalité asymptotique du vecteur $\widehat{\boldsymbol{\beta}}_h$ ainsi que la loi du logarithme itéré de ces composantes.

Theorem 1.2. Aneiros-Pérez and Vieu (2008)
Sous les hypothèses $nh^{4\alpha} \to 0$, $\left(n^{\frac{1}{4}-\frac{1}{r}}\phi(h)\right)^{-1} \log n \to 0$, $n\phi(h)^{\frac{\varepsilon a(r-2)}{r}-1} = \mathcal{O}(1)$ et $\left(n^{1-\frac{\theta(a+r)}{r(a+1)}}\phi(h)\right)^{-2} \log n = \mathcal{O}(1)$ quand $n \to \infty$ ($\alpha > 0$, $0 \leq \varepsilon \leq 1$, $a > 9/2$, $r > 4$ et $\theta > 2$), on obtient

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

où Σ la matrice variance-covariance.
Si en plus $\{Y_i, X_{i1}, X_{ip}, T_i\}$ est strictement stationnaire, on a

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} \left|\widehat{\boldsymbol{\beta}}_{hj} - \boldsymbol{\beta}_j\right| = (\Sigma_{jj})^{1/2}$$

et

$$\sup_{t\in\mathcal{C}} |\widehat{m}_h(t) - m(t)| = \mathcal{O}(h^\alpha) + \mathcal{O}\left(\sqrt{\frac{\log n}{n\phi(h)}}\right) \quad a.s.$$

Les mêmes vitesses de convergence sont obtenues dans un cadre des variables indépendantes par Aneiros-Pérez and Vieu (2006).

Le résultat suivant qui traite le modèle partiellement linéaire avec des covariables indépendantes, est dû à Härdle and Liang (2007). Ils ont établi, la normalité asymptotique d'un estimateur du paramètre du modèle basé sur la méthode du noyau.

Theorem 1.3. Härdle and Liang (2007)
On suppose que (i) $\sup_{0\leq \mathbf{x}\leq 1}\mathbb{E}(\|\mathbf{Z}\|^3|\mathbf{x}) < \infty$ et $\Sigma = Cov\{\mathbf{Z}-\mathbb{E}(\mathbf{Z}|\mathbf{X})\}$ est une matrice définie positive. (ii) $m(\mathbf{x})$ et $\mathbb{E}(\mathbf{z}_{ij}|\mathbf{x})$ sont lipschitz et continues. (iii) la fenêtre $h_n \approx \lambda n^{-1/5}$ pour un $0 < \lambda < \infty$. Alors

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{KR} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\Sigma^{-1}\right).$$

Dans le cadre des séries chronologiques, Gao (1995) s'est intéressé au modèle auto-régressif stationnaire partiellement linéaire suivant

$$Y_t = \beta Y_{t-1} + m(Y_{t-2}, \ldots, Y_{t-p}) + \varepsilon_t,\ t = p+1, \ldots, T,$$

où β est un paramètre inconnu, m est une fonction inconnue sur $\mathbb{R}^{p-1}$, ε_t des erreurs aléatoires indépendantes et identiquement distribuées et les ε_t sont indépendantes des Y_s pour tout $s = 1, \ldots, p$. Sous certaines conditions de régularité (voir Gao (1995), Théorème 1), on obtient la normalité asymptotique des estimateurs du paramètre β ainsi que de la variance du modèle et exposé dans le théorème suivant

Theorem 1.4. Gao (1995)

On suppose que $\mathbb{E}(\varepsilon_1) = 0$ et $\mathbb{E}(\varepsilon_1^2) = \sigma^2 < \infty$. Lorsque $T \to \infty$, on obtient

$$\sqrt{T}(\widehat{\beta}_T - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2\sigma_p^{-2}),$$

avec $\sigma_p^2 = \lim_{T\to\infty}\frac{1}{T}\sum_{t=p+1}^{T}\mathbb{E}((Y_t - \mathbb{E}(Y_{t-1}|X_t))^2|\Omega_{t-1})$.

Supposons de plus que $\mathbb{E}(\varepsilon^4) < \infty$. Quand $T \to \infty$, on a

$$\sqrt{T}(\widehat{\sigma}_T^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Var(\varepsilon_1^2),$$

où $\widehat{\sigma}_T^2 = \frac{1}{T}\sum_{t=p+1}^{T}(\widehat{Y}_t - \widehat{\boldsymbol{\beta}}_T\widehat{Y}_{t-1})^2$, $\widehat{Y}_t = Y_t - \widehat{m}_{1T}(X_t)$, *et* $\widehat{Y}_{t-1} = Y_{t-1} - \widehat{m}_{2T}(X_t)$.

Les estimateurs $\widehat{\beta}_T$ et $\widehat{\sigma}_T^2$ sont obtenus à l'aide de la méthode des moindres carrés.

1.4.2 Le cadre additif

Soit $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ un processus stationnaire α-mélangeant. On considère le modèle additif suivant

$$Y_i = m_1(X_{1i}) + m_2(X_{2i}) + m_{34}(X_{3i}, X_{4i}) + \varepsilon_i,$$

avec (X_1, X_3) un vecteur de variables aléatoires continues dans $R^{p_1+p_2}$ et (X_2, X_4) un vecteur de variables aléatoires discrètes dans $R^{q_1+q_2}$. Sous certaines conditions de

régularité, Camlong-Viot et al. (2006) ont montré que l' estimateur des composantes additives obtenu par la méthode d'intégration marginale, qui dépend à la fois des covariables continues et discrètes, converge avec le même vitesse que dans le cas continu.

Theorem 1.5. Camlong-Viot et al. (2006)

- Cas discret :

$$\sqrt{nh_n^{p_1}}\left(\widehat{\eta}_1(x_1)-\eta_1(x_1)\right)\xrightarrow{\mathcal{L}}\mathcal{N}(b(x_1),v^2(x_1)),$$

où

$$b(x_1)=\frac{1}{k!}\sum_{j=1}^{p_1}\int u_j^k K(u)du\left[(-1)^k\frac{\partial^k m_1}{\partial x_{1j}^k}(x_1)+\int m_1(z_1)\frac{\partial^k q_1}{\partial z_{1j}^k}(z_1)dz_1\right],$$

et

$$\begin{aligned}v^2(x_1) &= \int K^2(u)du\int\int\int\left[\sigma_0^2(x_1,x_2,x_3,x_4)+m^2(x_1,x_2,x_3,x_4)\right]\\ &\quad\times\frac{[q_2(x_2)q_{34}(x_3,x_4)]^2}{\mathbf{g}(x_1,x_2,x_3,x_4)}\mu(dx_2)dx_3\mu(dx_4).\end{aligned}$$

- Cas discret – continu :

$$\sqrt{nh_n^{p_2}}\left(\widehat{\eta}_{34}(x_3,x_4)-\eta_{34}(x_3,x_4)\right)\xrightarrow{\mathcal{L}}\mathcal{N}(b(x_3,x_4),v^2(x_3,x_4)),$$

où

$$\begin{aligned}&b(x_3,x_4)\\ &= \frac{1}{k!}\sum_{j=1}^{p_2}\int u_j^k K(u)du\left[(-1)^k\frac{\partial^k m_{34}}{\partial x_{3j}^k}(x_3,x_4)+\int m_{34}(z_3,z_4)\frac{\partial^k q_{34}}{\partial z_{3j}^k}(z_3,z_4)dz_3\mu(dz_4)\right],\end{aligned}$$

et

$$\begin{aligned}v^2(x_3,x_4) &= \mathbf{g}_4(x_4)\int K^2(u)du\int\int\left[\sigma_0^2(x_1,x_2,x_3,x_4)+m^2(x_1,x_2,x_3,x_4)\right]\\ &\quad\times\frac{[q_1(x_1)q_2(x_2)]^2}{\mathbf{g}(x_1,x_2,x_3,x_4)}dx_1\mu(dx_2),\end{aligned}$$

avec $\sigma_0^2(\cdot)=Var(Y|\mathbf{X}=\cdot)$, $\mathbf{q}(x)=q_1(x_1)q_2(x_2)q_{34}(x_3,x_4)$ est une densité d'intégration, K est un noyau et μ est une mesure discrète.

Dans le cas discret, le résultat obtenu généralise celui donné par le Théorème 1 de Fan et al. (1998) dans le cas de variables dépendantes. Dans le cas discret-continu, on remarque que la variance asymptotique de l'estimateur de l'intégration marginale

souffre seulement de la dimension des variables continues. Cela veut dire que la dimension des variables discrètes n'affecte en aucun cas la vitesse de convergence de la variance asymptotique.

La modélisation de type GARCH, trouvant son application notamment dans le domaine de la finance, a été étudiée par Mammen et al. (2002). Les auteurs considèrent le modèle additif suivant

$$\mathbb{E}(Y|\mathbf{X}) = \mathbb{E}(Y) + \sum_{j=1}^{J} m_j(X_j),$$

où, pour des raisons d'identifiabilité, ils supposent que $\mathbb{E}[m_j(X_j)] = 0$. La particularité de ce modèle est que, pour un paramètre β_0 et pour $j \geq 2$, les fonctions m_j sont liées de la façon paramétrique suivante

$$m_j(\cdot) = \beta_0^{j-1} m_1(\cdot).$$

Soit $\widehat{\beta}_{LS}$, l'estimateur des moindres carrés de β, défini par

$$\widehat{\beta}_{LS} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} \omega_j(X_{ik}) \{\widehat{m}_j(X_{ik}) - \beta^{j-1} \widehat{m}_1(X_{ik})\}^2,$$

où les ω_j sont des fonctions de poinds et $\widehat{m}_j$, $\widehat{m}_1$ sonts deux estimateurs respectifs des fonctions m_j et m_1. Sous certaines conditions de régularité, les auteurs obtiennent le résultat suivant

Theorem 1.6. Mammen et al. (2002)
Supposons le paramètre de lissage de la forme $h_n \sim n^{-1/5}$. Alors,

$$n^{1/2}(\widehat{\beta}_{LS} - \beta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{LS}),$$

où la variance asymptotique est donnée par $\Sigma_{LS} = \mathbb{E}[\{\sigma^2(\mathbf{X})\mathcal{H}_1^2(\mathbf{X}, \beta_0) + \mathcal{H}_2^2(\mathbf{X}, \beta_0))\}/D_{LS}^2]$, $(\mathcal{H}_i^2(\mathbf{X}, \beta_0))_{i=1,2}$ sont deux fonctions de $\mathbf{X}$ et β_0, $\sigma^2(\mathbf{X})$ est la variance conditionnelle de Y sachant $\mathbf{X}$ et $D_{LS} = \sum_{j=1}^{J} \sum_{k=1}^{k} \{(j-1)\beta_0^{j-2}\}^2 \mathbf{E}\{\omega_j(X_k) m_1^2(X_k)\}$.

Ce résultat peut être étendu au cas où les composantes additives peuvent s'écrire sous la forme générale suivante

$$m_j(\cdot) = F_j\{m_1(\cdot), \beta_0\}.$$

Dans ce cas et de façon similaire, l'estimateur des moindres carrés de β est alors défini par

$$\widehat{\beta}_{LS} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} \omega_j(X_{ik}) \{\widehat{m}_j(X_{ik}) - F_j\{\widehat{m}_1(X_{ik}), \beta\}\}^2.$$

Cependant, cette fois ci, le paramètre de lissage n'est plus de la forme $h_n \sim n^{-1/5}$. Cela est dû au bais de l'estimation non paramétrique induite par $\widehat{m}_j(X_{ik}) - F_j\{\widehat{m}_1(X_{ik}), \beta\}$. Pour remédier à cela, la fonction F_j peut être remplacée par une fonction plus appropriée $\widetilde{F}_j = F_j + \mathcal{O}_p(h_n^2)$. Une autre alternative, qui donne aussi les mêmes résultats, demande un sous-lissage avec un choix de la fenêtre de la forme $h_n = o(n^{-1/4})$. En effet, dans ce cas de figure, on aura un paramètre $\beta^*_{h_n} = \beta_0 + \mathcal{O}(h_n^2)$ qui donne la convergence en loi du terme $n^{1/2}(\widehat{\beta} - \beta^*_{h_n})$ comme dans le théorème ci-dessus. Par suite, il suffit de remplacer $\beta^*_{h_n}$ par β_0.

1.5 Contributions de la thèse

Soit $(\mathbf{X}_i, Y_i, \mathbf{Z}_i)_{i\geqslant 1}$ une suite de répliques indépendantes d'un vecteur aléatoire $(\mathbf{X}, Y, \mathbf{Z})$ à valeurs dans $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^p$.

Notons que le modèle semi-paramétrique défini auparavant par l'equation (1.2) peut s'écrire sous la forme suivante

$$Y - \mathbf{Z}^\top \boldsymbol{\beta} = m(\mathbf{X}) + \varepsilon. \tag{1.8}$$

En utilisant la méthode de Wand & Jones, un estimateur de la regression de la partie non paramétrique du modèle précédent est défini, pour tout $\mathbf{x} \in \mathbb{R}^d$, par

$$\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}}{n \mathbf{g}_n(\mathbf{X}_i)} \left(\prod_{l=1}^{d} \frac{1}{h_n} K_l \left(\frac{x_l - X_{il}}{h_n} \right) \right), \tag{1.9}$$

où x_l et X_{il} sont respectivement la $l^{\text{ième}}$ composante de $\mathbf{x}$ et $\mathbf{X}_i$, et K_l $(1 \leq l \leq d)$ sont des noyaux définis sur $\mathbb{R}$, $\mathbf{g}_n$ désigne l'estimateur de la densité marginale de la covariable $\mathbf{X}$ donnée par l'equation (1.1) et h_n est un paramètre de lissage qui tend vers zero avec une vitesse bien choisie qui va être donnée plus tard. Notons que $\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})$ dépend du paramètre inconnu $\boldsymbol{\beta}$ qu'on doit estimer. Suivant la méthode des moindres carrés, l'estimateur de $\boldsymbol{\beta}$ est de la forme

$$\widehat{\boldsymbol{\beta}} = [\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top]^{-1} \tilde{\mathbf{Z}}\tilde{Y}, \tag{1.10}$$

avec

$$\begin{array}{rcl} \tilde{Y} & = & \left[Y_i - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i) Y_j \right]^{\top}_{1\leq i\leq n}, \\ \tilde{\mathbf{Z}} & = & \left[\mathbf{Z}_i - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i) \mathbf{Z}_j \right]_{1\leq i\leq n}, \end{array} \tag{1.11}$$

W_{nj} est une fonction poids qui sera définie plus loin dans notre travail.

Pour réduire l'impact de la dimension sur la partie non paramétrique dans le modèle de la régression semi-paramétrique, on considère une structure additive de la fonction de régression m et on introduit le modèle additif partièllement linéaire suivant

$$Y = \mathbf{Z}^{\top}\boldsymbol{\beta} + \sum_{l=1}^{d} m_l^{\boldsymbol{\beta}}(X_l) + \varepsilon := \mathbf{Z}^{\top}\boldsymbol{\beta} + m_{add}^{\boldsymbol{\beta}}(\mathbf{X}) + \varepsilon, \tag{1.12}$$

où X_l est la $l^{\text{ième}}$ composante du vecteur $\mathbf{X}$ et $m_l^{\boldsymbol{\beta}}$ est une fonction réelle univariée sous une contrainte supplémentaire afin que le modèle soit identifiable, $\mathbb{E}m_l^{\boldsymbol{\beta}}(X_l) = 0$ pour tout $l \in \{1, \ldots, d\}$, ε est l'erreur du modèle telque $\mathbb{E}[\varepsilon|\mathbf{X}, \mathbf{Z}] = 0$ et supposée de variance finie σ_{ε}^2.

En utilisant la méthode d'integration marginale, l'estimateur de la fonction de régression additive est donné, pour tout $\mathbf{x} \in \mathbb{R}^d$, par

$$\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{z}) = \sum_{l=1}^{d} \widehat{\zeta}_l^{\widehat{\boldsymbol{\beta}}}(x_l) + \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{z})\mathbf{q}(\mathbf{z})d\mathbf{z} \tag{1.13}$$

et

$$\widehat{\zeta}_l^{\widehat{\boldsymbol{\beta}}}(x_l) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{1.14}$$

Dans la suite, nous présentons brièvement les principaux résultats établis dans nos travaux. Pour pouvoir donner ces résultats, quelques conditions de régularité sont nécessaires pour ce fait (voir plus loin dans les chapitres de cette thèse).

1.5.1 Quelques résultats asymptotiques sur le paramètre $\boldsymbol{\beta}$

Supposons qu'il existe un réel $r \geq 2$ tel que $\max_{1\leq i\leq n} E|\varepsilon_i|^r < \infty$ et que $\mathbb{E}[\varepsilon_i^2|\mathbf{X}_i, \mathbf{Z}_i] = \sigma_{\varepsilon}^2$ p.s., alors nous établissons la normalité asymptotique de l'estimateur $\widehat{\boldsymbol{\beta}}$ du paramètre du modèle

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon}^2 B^{-1}),$$

où B est une matrice définie positive qui peut être estimée par $\frac{1}{n}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\top}$, σ_{ε}^2 est la variance de la variable aléatoire ε, et $\xrightarrow{d}$ désigne la convergence en loi.

Sous les mêmes conditions, on établit la loi du logarithme itéré des composantes de $\widehat{\boldsymbol{\beta}}$. Presque sûrement on obtient

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log\log n}} \left|\widehat{\beta}_j - \beta_j\right| = (\sigma_{\varepsilon}^2 b^{jj})^{1/2},$$

avec $\widehat{\beta}_j$ et β_j désignent, respectivement, la $j^{\text{ième}}$ composante de $\widehat{\boldsymbol{\beta}}$ et $\boldsymbol{\beta}$ et b^{jk} désigne l'élément de la $j^{\text{ième}}$ ligne et la $k^{\text{ième}}$ colonne de B^{-1}.

Un test d'hypothèse est construit pour tester l'hypothèse nulle $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ contre une alternative $\mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Sous $\mathcal{H}_0$ on obtient

$$\frac{n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T B(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\hat{\sigma}_n^2} \xrightarrow{d} \chi^2_{(p)}$$

où $\chi^2_{(p)}$ désigne la loi χ^2 à p degrés de liberté.

1.5.2 Étude asymptotique de la variance du modèle

Supposons maintenant qu'il existe un réel $r \geq 4$ tel que $\max_{1\leq i,j\leq n} E|\varepsilon_i|^r < \infty$, alors

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma_{\varepsilon}^2) \xrightarrow{d} \mathcal{N}(0, Var\varepsilon_1^2) \quad a.s.$$

Si de plus $\max_{1\leq i\leq n} E|\varepsilon_i|^{r+1} < \infty$, alors une loi du logarithme itéré pour l'estimateur de la variance du modèle est donnée par

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log\log n}} \left|\hat{\sigma}_n^2 - \sigma_{\varepsilon}^2\right| = (Var\varepsilon_1^2)^{1/2} \quad a.s.$$

Sous l'hypothèse nulle $\mathcal{H}_0 : \sigma_{\varepsilon} = \sigma_0$

$$\frac{n(\hat{\sigma}_n^2 - \sigma_{\varepsilon}^2)^2}{Var\varepsilon_1^2} \xrightarrow{d} \chi^2,$$

où χ^2 est la loi khi-deux à 1 degré de liberté.

1.5.3 Consistance uniforme de la partie additive du modèle

Dans un premier temps, nous allons donner le comportement asymptotique des estimateurs des composantes additives dans le modèle additif partiellement linéaire.

Pour cela, soit $I^d = \prod_{l=1}^{d} I_l$ un compact de $\mathbb{R}^d$, on définit la fonction ϕ sur l'intervalle I_l par

$$\phi(u_l) = \int_{\mathbb{R}^{d-1}} \frac{H^{\boldsymbol{\beta}}(\mathbf{u})}{\mathbf{g}(\mathbf{u}_{-l}|u_l)} q_{-l}(\mathbf{u}_{-l}) d\mathbf{u}_{-l},$$

où $\mathbf{g}(\mathbf{u}_{-l}|u_l)$ est la densité conditionnelle de $\mathbf{X}_{-l}$ sachant $X_l = u_l$,

avec

$$H^{\boldsymbol{\beta}}(\mathbf{u}) = \mathbb{E}[(Y - \mathbf{Z}^\top \boldsymbol{\beta})^2 | \mathbf{X} = \mathbf{u}], \mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d.$$

Pour tout $1 \leq l \leq d$, on considère la quantité suivante

$$\sigma_l = \sigma_l^{\boldsymbol{\beta}} = \sup_{u_l \in I_l} \sqrt{\frac{\phi(u_l)}{\mathbf{g}_l(u_l)} \int_{\mathbb{R}} K_l^2(t) dt},$$

où $\mathbf{g}_l$ est la $l^{\text{ième}}$ densité marginale de $\mathbf{X}$. Alors on obtient

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{x_1 \in I_1} \left| \widehat{\eta}_1^{\widehat{\boldsymbol{\beta}}}(x_1) - \eta_1(x_1) \right| = \sigma_1 \quad a.s.$$

Le résultat suivant fournit une consistance uniforme forte de l'estimateur $\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}$ de la partie additive du modèle . Sous les mêmes conditions et en utilisant le dernier résultat, on a

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{\mathbf{x} \in I} \left| \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{add}(\mathbf{x}) \right| = \sum_{l=1}^{d} \sigma_l \quad a.s.$$

Comme conséquence immédiate du résultat précédent, on a la consistance uniforme avec vitesse de convergence de l'estimateur de la régression semi-paramétrique donnée dans le corollaire suivant

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{\mathbf{x},\mathbf{z} \in I \times \mathbb{R}^p} \left| \widehat{m}(\mathbf{x}, \mathbf{z}) - m(\mathbf{x}, \mathbf{z}) \right| = \sigma_\varepsilon \sum_{l=1}^{p} (b^{ll})^{1/2} + \sum_{l'=1}^{d} \sigma_{l'} \quad a.s.,$$

avec $\widehat{m}(\mathbf{x}, \mathbf{z}) = \mathbf{z}^\top \widehat{\boldsymbol{\beta}} + \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ et σ_ε^2 est la variance de ε.

1.5.4 Test d'additivité

Pour pouvoir définir l'hypothèse qui servira à construire notre test d'additivité, on introduit la classe de fonctions suivante

$$\mathcal{M} = \left\{ m_{add} : \quad m_{add}(\mathbf{X}) = \sum_{l=1}^{d} m_l(X_l), \quad \mathbb{E}[m_l(X_l)] = 0, \quad 1 \leq l \leq d \right\}.$$

Le test consiste à tester l'hypothèse nulle $\mathcal{H}_0$: $''m \in \mathcal{M}''$, contre une hypothèse alternative $\mathcal{H}_1$: $'' m \notin \mathcal{M}''$. Suivant les mêmes idées que celles utilisées par Härdle and Mammen (1993) pour des propos quelque peu différents, on considère la quantité

$$R_n := \int_{\mathbb{R}^d} \left[\frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \left(Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)\right) \right]^2 \frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}_n^2(\mathbf{x})} d\mathbf{x} \tag{1.15}$$

qui estime la quantité $R = \mathbb{E}[\mathbb{E}(Y - \mathbf{Z}^\top \boldsymbol{\beta} - m_{add}(\mathbf{X}))|\mathbf{X}]^2$ égale à zéro si et seulement si l'hypothèse nulle $\mathcal{H}_0$ est vraie. Dans la suite, nous établissons la normalité asymptotique de la statistique R_n qui nous permettra de construire la région de rejet de notre test.

Supposons qu'il existe un réel $\eta \geq d/(3(2k+1)-d)$ tel que $E(\varepsilon_1^{2(1+\eta)}) < \infty$ et que ε est indépendante de $(\mathbf{X}, \mathbf{Z})$. Alors, sous l'hypothèse nulle H_0, on a

$$\frac{\sqrt{nh_n^d} R_n - D n^{-1/2} h_n^{-d/2}}{\sqrt{V}} \xrightarrow{d} \mathcal{N}(0,1),$$

où

$$V := 2 \int (\sigma_\varepsilon^2)^2 \mathbf{g}^{-2}(\mathbf{u}) \mathbf{q}^2(\mathbf{u}) d\mathbf{u} \times \int \left[\int K(\mathbf{t}) K(\mathbf{t}-\mathbf{r}) d\mathbf{t} \right]^2 d\mathbf{r},$$

et

$$D := \int \sigma_\varepsilon^2 \mathbf{g}^{-1}(\mathbf{u}) \mathbf{q}(\mathbf{u}) d\mathbf{u} \times \int K^2(\mathbf{t}) d\mathbf{t}.$$

Comme D et V contiennent des quantités inconnues, alors ils doivent être estimés pour qu'on puisse effectuer le test. Le résultat suivant donne la normalité asymptotique quand D et V sont estimés.

Sous $\mathcal{H}_0$, on a

$$\frac{\sqrt{nh_n^d} R_n - \widehat{D}_n n^{-1/2} h_n^{-d/2}}{\sqrt{\widehat{V}_n}} \xrightarrow{d} \mathcal{N}(0,1),$$

avec

$$\widehat{V}_n := 2 \int (\widehat{\sigma}_n^2)^2 \mathbf{g}_n^{-2}(\mathbf{u}) \mathbf{q}^2(\mathbf{u}) d\mathbf{u} \int \left[\int K(\mathbf{t}) K(\mathbf{t}-\mathbf{r}) d\mathbf{t} \right]^2 d\mathbf{r}.$$

et

$$\widehat{D}_n := \int \widehat{\sigma}_n^2 \mathbf{g}_n^{-1}(\mathbf{u}) \mathbf{q}(\mathbf{u}) d\mathbf{u} \int K^2(\mathbf{t}) d\mathbf{t}.$$

La région de reget :

Au seuil α, la région asymptotique de rejet associée au test est donnée par

$$\mathcal{D} = \{R_n > n^{-1} h_n^{-d} \widehat{D}_n + n^{-1/2} h_n^{-d/2} \sqrt{\widehat{V}_n} \phi^{-1}(1-\alpha)\},$$

où ϕ est la fonction de distribution d'une variable aléatoire $\mathcal{N}(0,1)$.

À l'aide des simulations, on a montré que notre test est puissant et que la puissance du test est une fonction croissante qui atteint la valeur 1 pour un échantillon de taille $n = 1000$. Tandis que l'erreur de première espèce est considérablement proche de la valeur théorique de 5%.

Chapitre 2

Asymptotic results and additivity test in partially linear regression model

Ce chapitre est composé de deux publications en (2011) et (2013) aux comptes rendus de l'académie des sciences, mis en forme pour être inséré dans le présent manuscrit de thèse.

This chapter is devoted to the study of some asymptotic properties of the linear part of the partially linear regression model defined by

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{l=1}^{d} m_l(X_{il}) + \varepsilon_i, \quad 1 \leq i \leq n,$$

together with the construction of a test to test the additivity form of the nonlinear part of this model. Here, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^\top$ are vectors of explanatory variables, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)^\top$is a vector of unknown parameters, $m_1, \ldots, m_d$ are unknown univariate real functions, and ε_i are i.i.d. random errors with finite variances σ_ε^2. The nonparametric kernel technique combined with the marginal integration method are used throughout this chapter to estimate the functions $(m_l)_{1\leq l\leq d}$ and the least-square error criterion to estimate the parameter $\boldsymbol{\beta}$, we establish the asymptotic normality together with the iterated logarithm law of the estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. In the second step, we build an additivity test on the nonlinear part of the partially linear model.

2.1 Introduction

Let $(\mathbf{X}_i, Y_i, \mathbf{Z}_i)_{i\geqslant 1}$ be a sequence of i.i.d. copies of the $\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^p$-valued random vector $(\mathbf{X}, Y, \mathbf{Z})$. Denote by f its joint density function with respect to the Lesbegue measure and by $\mathbf{g}$ the marginal density associated to the random vector $\mathbf{X}$. Parametric regression models provide powerful tools for analyzing practical data when the models are correctly specified, but may suffer from large modelling biases when structures of the models are misspecified. As an alternative, nonparametric smoothing methods eases the concerns on modelling biases. However, nonparametric models are hampered by the so-called curse of dimensionality in multivariate settings, see Stone (1985) for details.

One of the methods for attenuating this difficulty is to model covariate effects via a partially linear structure, a combination of linear and nonlinear parts. This results in the partially linear regression models of the form

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m(\mathbf{X}) + \varepsilon, \tag{2.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown parameters, m is the nonlinear part of the model and ε is random error. Here, $\mathbf{Z}^\top$ stands as the transpose of the vector $\mathbf{Z}$.

The partially linear regression model has a broad applicability in the fields of biology, economics, education and social sciences among others. This model and various associated estimators, test statistics, and extensions have generated a substantial body of literature, which includes the works of Robinson (1988), Liang (2000), Aneiros-Pérez et al. (2004) and Chen and Wang (2010) together with that of Aneiros-Pérez and Vieu (2006) and Dabo-Niang and Guillas (2010) in the semi-functional modelling setting. Notice also the modelization of real data with such models have been carried out in a number of papers, see, for instance, Robinson (1988), Härdle et al. (2000) and Aneiros-Pérez and Vieu (2008).

To reduce the dimension impact of the nonparametric part in the partially linear regression model, we consider the additive structure of the regression function m and introduce the following model, for any $1 \leq i \leq n$

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{l=1}^{d} m_l(X_{il}) + \varepsilon_i := \mathbf{Z}_i^\top \boldsymbol{\beta} + m_{add}(\mathbf{X}_i) + \varepsilon_i, \tag{2.2}$$

where X_{il} is the l-th component of the vector $\mathbf{X}_i$ and m_l is a real univariate function and ε_i are i.i.d. random errors with $\mathbb{E}[\varepsilon_i|\mathbf{X}_i, \mathbf{Z}_i] = 0$. Subsequently, we have to estimate

the unknown quantities, that is, the vector parameters $\boldsymbol{\beta}$ and the univariate functions $m_l, 1 \leq l \leq d$, as well.

2.2 Presentation of estimators

On the basis of an n-sample drawn from the random vector $(\mathbf{X}, Y, \mathbf{Z})$, we first define the kernel estimator of the marginal density $\mathbf{g}$, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\mathbf{g}_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right),$$

where K is a kernel, i.e., a non-negative function defined on $\mathbb{R}^d$ and integrating to 1, and h_n is a smoothing parameter tending to zero with a suitable rate that will be given later on. Notice that the model (2.1) may be written also as

$$Y - \mathbf{Z}^\top \boldsymbol{\beta} = m(\mathbf{X}) + \varepsilon. \tag{2.3}$$

On the basis of the model (2.3), following the usual Wand & Jones method, the regression estimator involving the nonparametric part of the model may be defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}}{n\mathbf{g}_n(\mathbf{X}_i)} \left(\prod_{l=1}^{d} \frac{1}{h_n} K_l\left(\frac{x_l - X_{il}}{h_n}\right) \right), \tag{2.4}$$

where x_l and X_{il} are the l-th component of $\mathbf{x}$ and $\mathbf{X}_i$ respectively, and K_l $(1 \leq l \leq d)$ are kernels defined on $\mathbb{R}$. Note that $\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})$ depend on the unknown parameter $\boldsymbol{\beta}$ which needs to be estimated. Considering the model (2.3), the function m clearly depends on the parameter $\boldsymbol{\beta}$ and its additive structure may be written as

$$m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \mu + \sum_{l=1}^{d} m_l^{\boldsymbol{\beta}}(x_l), \tag{2.5}$$

where model identifiability considerations impose that $Em_l^{\boldsymbol{\beta}}(X_l) = 0$, $1 \leq l \leq d$.

Various methods have been proposed in the literature to estimate the additive components of the regression model including the marginal integration method and the backfitting algorithms, we refer to Newey (1994) and Linton and Nielsen (1995), as well as Hastie and Tibshirani (1986), Sperlich et al. (1999) and the references therein for an account of results on these topics. The linear part of the semi-parametric models has been investigated in a number of works. Our aim in this chapter is to introduce

an estimate of the parameter $\boldsymbol{\beta}$ and to study its asymptotic properties related to the normality and the iterated logarithm law. Moreover, while the nonlinear part is estimated using the nonparametric kernel method combined with the marginal integration technique, we construct a test to test the additivity shape of the nonlinear part of the partially linear model.

Our main results in this chapter establish the asymptotic normality of the estimate $\widehat{\boldsymbol{\beta}}$ of the parameter $\boldsymbol{\beta}$ together with the iterated logarithm law for every component of $\widehat{\boldsymbol{\beta}}$. Furthermore, we establish the asymptotic normality of the testing statistics allowing to test the additivity form of the nonlinear part in the model (2.2). The test statistic is built on the basis of the estimation of the following quantity

$$\mathbb{E}[\mathbb{E}(Y - \mathbf{Z}^\top \boldsymbol{\beta} - m_{add}(\mathbf{X}))|\mathbf{X}]^2.$$

In order to set out our estimates and our testing procedure, introduce first some further notations. For any $1 \leq l \leq d$, set $\mathbf{x}_{-l} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d)$, $\mathbf{q}_{-l}(\mathbf{x}_{-l}) = \prod_{k=1, k\neq l}^{d} q_k(x_k)$ and $\mathbf{q}(\mathbf{x}) = \prod_{l=1}^{d} q_l(x_l)$, where q_l, $1 \leq l \leq d$, are univariate densities functions. Following the marginal integration method, the additive regression function estimator is given, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{l=1}^{d} \widehat{\zeta}_l^{\boldsymbol{\beta}}(x_l) + \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{z})\mathbf{q}(\mathbf{z})d\mathbf{z}, \tag{2.6}$$

where

$$\widehat{\zeta}_l^{\boldsymbol{\beta}}(x_l) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{2.7}$$

Here, q_{-l}, $1 \leq l \leq d$, and $\mathbf{q}$ stand as densities functions and $\widehat{\zeta}_l^{\boldsymbol{\beta}}$ is the estimate of the l-th component of the additive regression function which still depends on the parameter $\boldsymbol{\beta}$. Therefore, one has to estimate the vector parameter $\boldsymbol{\beta}$ to have ready estimates. As a first step in the modeling procedure, we begin by the vector parameter $\boldsymbol{\beta}$ estimation.

2.2.1 Estimation of the parameters $\boldsymbol{\beta}$ and σ_ε^2

While considering the partially linear additive regression model

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m_{add}(\mathbf{X}) + \varepsilon, \tag{2.8}$$

Making use of the statements (2.4), (2.6)-(2.8), and considering the least square error criterion, it follows that

$$\widehat{\boldsymbol{\beta}} = [\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top]^{-1}\tilde{\mathbf{Z}}\tilde{Y}, \tag{2.9}$$

where

$$\begin{array}{rcl} \tilde{Y} & = & \left[Y_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i) Y_j \right]_{1\leq i\leq n}^\top, \\ \tilde{\mathbf{Z}} & = & \left[\mathbf{Z}_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i) \mathbf{Z}_j \right]_{1\leq i\leq n}, \end{array} \tag{2.10}$$

$$W_{nj}(\mathbf{X}_i) = \frac{U_{nj}(\mathbf{X}_i)}{n\mathbf{g}_n(\mathbf{X}_j)} \tag{2.11}$$

and

$$U_{nj}(\mathbf{X}_i) = \sum_{l=1}^d \frac{1}{h_n} K_l\left(\frac{X_{il} - X_{jl}}{h_n}\right) D_l - (d-1) \int_{\mathbb{R}^d} \prod_{k=1}^d \frac{1}{h_n} K_k\left(\frac{x_k - X_{jk}}{h_n}\right) \mathbf{q}(\mathbf{x}) d\mathbf{x}$$

with

$$D_l = \int_{\mathbb{R}^{d-1}} \prod_{k=1, k\neq l} \frac{1}{h_n} K_k\left(\frac{x_k - X_{jk}}{h_n}\right) \mathbf{q}_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l}.$$

Therefore, the parameter σ_ε^2 is estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top (\tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i^\top)^{-1} \tilde{\mathbf{Z}}_i \tilde{Y}_i)^2. \tag{2.12}$$

Furthermore, the estimates of the regression function and the additive components are defined by

$$\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) = \sum_{l=1}^d \widehat{\zeta}_l^{\widehat{\boldsymbol{\beta}}}(x_l) + \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{z}) \mathbf{q}(\mathbf{z}) d\mathbf{z}$$

and

$$\widehat{\zeta}_l^{\widehat{\boldsymbol{\beta}}}(x_l) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) \mathbf{q}_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) \mathbf{q}(\mathbf{x}) d\mathbf{x}.$$

2.2.2 Construction of the test statistic

To set the hypothesis to be tested, introduce the following class of function

$$\mathcal{M} = \left\{ m_{add} : \quad m_{add}(\mathbf{X}) = \sum_{l=1}^d m_l(X_l), \quad \mathbb{E}[m_l(X_l)] = 0, \quad 1 \leq l \leq d \right\},$$

Thus, we have to test the null hypothesis $\mathcal{H}_0 : \, ''m \in \mathcal{M}\,''$, versus the alternative hypothesis is $\mathcal{H}_1 : \, ''\, m \notin \mathcal{M}\,''$. Notice that the theoretical deviation from the null to the alternative hypotheses is given by the quantity

$$R = \mathbb{E}[\mathbb{E}(Y - \mathbf{Z}^\top \boldsymbol{\beta} - m_{add}(\mathbf{X}))|\mathbf{X}]^2$$

and the test statistic

$$R_n := \int_{\mathbb{R}^d} \left[\frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \left(Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}} - \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)\right) \right]^2 \frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}_n^2(\mathbf{x})} d\mathbf{x}$$

follows as the estimate of R. In the sequel, we establish the asymptotic normality of the statistic R_n suitably normalized which allows to build the rejection region of the test.

2.3 Main results

The assumptions involving the distribution function of $\mathbf{X}$, the regression function m, the kernels K_l, $1 \leq l \leq d$ and the smoothing parameters are gathered together hereafter for easy reference. The first part of these conditions is devoted to the regression function m and the density $\mathbf{g}$. In the sequel, I^d is a compact subset of $\mathbb{R}^d$.

(G.1) m is k-times continuously differentiable.

(G.2) The marginal density $\mathbf{g}$ is bounded from below on its support.

(G.3) The marginal density $\mathbf{g}$ is uniformly continuous on its support.

(G.4) The marginal density $\mathbf{g}$ has $k+1$ continuous derivatives.

Throughout, the following hypothesis is considered upon the sequence of bandwidths $(h_n)_{n\geq 1}$.

(H.1) $h_n = \vartheta_1 \left(\dfrac{\log n}{n}\right)^{1/(2k+1)}$ for $0 < \vartheta_1 < \infty$ and $2k+1 \geq d$.

Set now, for any $\mathbf{x} \in \mathbb{R}^d$, $K(\mathbf{x}) := \prod_{l=1}^d K_l(x_l)$. The kernels are assumed to satisfy the following conditions

(K.1) For any $1 \leq l \leq d$, K_l is bounded, Lipschitz continuous and integrating to one.

(K.2) For any $1 \leq l \leq d$, $K_l(u) = 0$ for u $\notin [-\lambda/2, \lambda/2]$, for some $0 < \lambda < \infty$.

(K.3) K is a kernel of order k.

Consider also the following assumptions upon the random variables Y and $\mathbf{Z}$.

(M.1) Y and $\mathbf{Z}$ are bounded.

(M.2) $\mathbf{Z}$ is with mean zero.

The assumptions on the weight functions q_l, $1 \leq l \leq d$, are listed hereafter

(Q.1) For any $1 \leq l \leq d$, q_l has $k+1$ continuous and bounded derivatives.

(Q.2) The support of the function $\mathbf{q}$ is included in the support of the density $\mathbf{g}$.

2.3.1 Comments on hypotheses.

The most part of hypotheses are needed in considering estimation of the nonlinear part of the model to build up an estimate of the parameter $\boldsymbol{\beta}$. Indeed, the proofs needs to use the uniform convergence of the additive regression estimate $\hat{m}_{add}^{\boldsymbol{\beta}}$ to m_{add} stated in Camlong-Viot (2001).

Remark 2.1. The limiting behavior of the estimate $\mathbf{g}_n$, for suitable choices of the bandwidth h_n, has been studied in a number of works over the last decades. For an overview of the literature on the subject together with statistical applications we refer to Devroye and Lugosi (2001), Wand and Jones (1995), Scott (1992), Bosq and Lecoutre (1987), Devroye and Györfi (1985) and Prakasa Rao (1983). More particularly, Parzen (1962) established, under some assumptions upon the kernel K, that $\mathbf{g}_n$ is asymptotically an unbiased and consistent estimator of $\mathbf{g}$ at any of its continuity points, whenever $h_n \to 0$ and $nh_n^d \to \infty$ with n. Under some additional conditions on $\mathbf{g}$ and h_n, he also obtained the asymptotic normality. It is noteworthy to notice, in our framework, that condition $nh_n^d \to \infty$ is satisfied whenever $2k+1 \geq d$. Therefore, we have to suppose that $2k+1 \geq d$ to reach our results.

2.3.2 Theorems

Theorem 2.1 and Theorem 2.2 hereafter give asymptotic properties of $\widehat{\boldsymbol{\beta}}$. Let $\xrightarrow{d}$ denotes the convergence in distribution.

Theorem 2.1. Assume that assumptions [G.1-4], [H.1], [K.1-3], [M.1-2], [Q.1-2] hold. In addition, suppose that $\max_{1\leq i\leq n} \mathbb{E}|\varepsilon_i|^r < \infty$ for some $r \geq 2$ and $\mathbb{E}[\varepsilon_i^2|\mathbf{X}_i, \mathbf{Z}_i] = \sigma_\varepsilon^2$ a.s.. Then, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 B^{-1}),$$

where B is a positive definite matrix that can be consistently estimated by $\frac{1}{n}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top$ and σ_ε^2 is the variance of the random variable ε.

Theorem 2.2. Under assumptions of Theorem 2.1, for some $r > 2$, we have

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2\log\log n}} \left|\widehat{\beta}_j - \beta_j\right| = (\sigma_\varepsilon^2 b^{jj})^{1/2} \quad a.s.,$$

where $\widehat{\beta}_j$ and β_j denote the j-th components of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ respectively and b^{jk} denotes the j-th row and k-th rank element of B^{-1}.

The following theorem gives the asymptotic normality related to the testing statistic R_n.

Theorem 2.3. Assume that assumptions [G.1-4], [H.1], [K.1-3], [M.1-2], [Q.1-2] hold true. In addition, suppose that $\mathbb{E}(\varepsilon_1^{2(1+\eta)}) < \infty$, for some $\eta \geq d/(3(2k+1)-d)$, $2k-1 \geq d$ and ε is independent with $(\mathbf{X}, \mathbf{Z})$. Then, under the null hypothesis H_0, we have

$$\frac{\sqrt{nh_n^d}R_n - Dn^{-1/2}h_n^{-d/2}}{\sqrt{V}} \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$D := \int \sigma_\varepsilon^2 \mathbf{g}^{-1}(\mathbf{u})\mathbf{q}(\mathbf{u})d\mathbf{u} \times \int K^2(\mathbf{t})d\mathbf{t},$$

and

$$V := 2\int (\sigma_\varepsilon^2)^2 \mathbf{g}^{-2}(\mathbf{u})\mathbf{q}^2(\mathbf{u})d\mathbf{u} \times \int \left[\int K(\mathbf{t})K(\mathbf{t}-\mathbf{r})d\mathbf{t}\right]^2 d\mathbf{r}.$$

Since D and V include unknown quantities, we need to estimate them to perform the test. The following corollary gives the asymptotic normality when D and V are estimated.

Corollary 2.1. Assume that the assumptions of Theorem 2.1 are verified. Under H_0, we have

$$\frac{\sqrt{nh_n^d}R_n - \widehat{D}_n n^{-1/2}h_n^{-d/2}}{\sqrt{\widehat{V}_n}} \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\widehat{D}_n := \int \widehat{\sigma}_n^2 \mathbf{g}_n^{-1}(\mathbf{u})\mathbf{q}(\mathbf{u})d\mathbf{u} \int K^2(\mathbf{t})d\mathbf{t},$$

and

$$\widehat{V}_n := 2\int (\widehat{\sigma}_n^2)^2 \mathbf{g}_n^{-2}(\mathbf{u})\mathbf{q}^2(\mathbf{u})d\mathbf{u} \int \left[\int K(\mathbf{t})K(\mathbf{t}-\mathbf{r})d\mathbf{t}\right]^2 d\mathbf{r}.$$

2.4 Comments and concluding remarks

Remark 2.2. According to Theorem 2.1, the asymptotic variance is given by $\sigma_\varepsilon^2 B^{-1}$ where

$$B = \mathbb{E}\left[\left(\mathbf{Z} - \sum_{l=1}^{d} W^l(X_l)\right)\left(\mathbf{Z} - \sum_{l=1}^{d} W^l(X_l)\right)^\top\right],$$

and

$$W^l(X_l) \;:=\; \int_{\mathbb{R}^{d-1}} \mathbb{E}[\mathbf{Z}|\mathbf{X} = (X_l, \mathbf{x}_{-l})]\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l} - \left(\frac{d-1}{d}\right)\int_{\mathbb{R}^d} \mathbb{E}[\mathbf{Z}|\mathbf{X} = \mathbf{x}]\mathbf{q}(\mathbf{x})d\mathbf{x}.$$

Since B is an unknown quantity, it may be consistently estimated by

$$\widehat{B} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathbf{Z}}_i \tilde{\mathbf{Z}}_i^\top.$$

where $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\mathbf{Z}_j$. Indeed, considering Theorem 4.1, it may be stated similarly that

$$\sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \to \sum_{l=1}^{d} W^l(X_l) \quad a.s. \tag{2.13}$$

Therefore, it follows that $\widetilde{B} \to B \quad a.s.$

Remark 2.3. This chapter is a first step to various studies and extensions related to the partially linear model when the nonlinear part is assumed to be with an additive structure and estimated using the marginal integration method. In forthcoming works, properties of estimates of the additive components of the regression function are investigated considering the exact rate of pointwise and uniform strong consistencies. Such results enable one to derive 100%-confidence bands, in the spirit of the work of Deheuvels and Mason (2004), for the estimated function parameters. Notice that the asymptotic normality of these estimates will be considered and applied to build the usual $(1-\alpha)$-confidence bands for the underlying parameters. Further studies may deal with tests related to the parameters $\boldsymbol{\beta}$ and m and the model goodness-of-fit. Natural extensions may assume some dependency structure upon the data, as the mixing structure for example, or/and to consider functional data in the nonlinear part of the model.

Remark 2.4. For a given significance level α, the asymptotic rejection region associated to the testing procedure is given by

$$\mathcal{D} = \{R_n > n^{-1}h_n^{-d}\widehat{D}_n + n^{-1/2}h_n^{-d/2}\sqrt{\widehat{V}_n}\phi^{-1}(1-\alpha)\},$$

where ϕ is the distribution function of the $\mathcal{N}(0,1)$ random variable.

2.5 Semiparametric Efficiency

Let $\mathcal{A}$ be the space of all additive functions m such that $m(\mathbf{x}) = m_1(x_1) + \cdots + m_d(x_d)$ with $\mathbb{E}m_i(X_i) = 0$ and $\mathbb{E}m(\mathbf{X})^2 < \infty$. The logarithm of the joint density of $(Y, \mathbf{X}, \mathbf{Z})$ is given by $\ell(\beta, m; (y, \mathbf{x}, \mathbf{z})) = \log g_\varepsilon(y - \mathbf{x}^\top\beta - m(\mathbf{z}))$, where g_ε is the density of ε supposed have a finite variance σ_ε^2 and $\mathbb{E}(\varepsilon|\mathbf{X}, \mathbf{Z}) = 0$, a.s. Following Mammen et al. (2011), suppose that the mapping $\beta \rightarrowtail m_\beta$ is Fréchet differentiable function from $\mathbb{R}^p$ to $\mathcal{A}$. Then, for any fixed value (β^0, m_{β^0}), each finite-dimensional submodel $\{(\beta, m_\beta) : \beta\}$ has the score function

$$\begin{aligned} d\ell(\beta, m_\beta)/d\beta|_{\beta=\beta^0} &= \partial\ell(\beta, m_{\beta^0})/\partial\beta|_{\beta=\beta^0} + \partial\ell(\beta^0, m)/\partial m|_{m=m_{\beta^0}}(\delta) \\ &= g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)\mathbf{Z} + g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)\delta(\mathbf{X}), \end{aligned}$$

where $\delta = \partial m_\beta/\partial\beta|_{\beta=\beta^0} \in \mathcal{A}$ and $\partial\ell/\partial m$ denotes the Fréchet derivative of ℓ with respect to m. The Fisher information matrix for estimating the parameter β in each submodel $\{(\beta, m_\beta) : \beta\}$ is given by

$$\begin{aligned} \mathcal{I}(\delta) &:= \mathbb{E}\left[(g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon))^2 (\mathbf{Z} + \delta(\mathbf{X}))(\mathbf{Z} + \delta(\mathbf{X}))^\top\right] \\ &= \mathbb{E}[g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)]^2\mathbb{E}\left[(\mathbf{Z} + \delta(\mathbf{X}))(\mathbf{Z} + \delta(\mathbf{X}))^\top\right]. \end{aligned}$$

The least favorable direction δ^* that minimizes $\mathcal{I}(\delta)$ over the additive space $\mathcal{A}$ is the solution of the following integral equation : for all $\delta \in \mathcal{A}$

$$\begin{aligned} 0 &= \mathbb{E}[g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)]^2\mathbb{E}[\mathbf{Z} + \delta^*(\mathbf{X})]\delta(\mathbf{X}) \\ &= \mathbb{E}[g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)]^2\mathbb{E}\left[(\mathbb{E}(\mathbf{Z}|\mathbf{X}) + \delta^*(\mathbf{X}))\delta(\mathbf{X})\right]. \end{aligned}$$

Therefore

$$\arg\min_\delta \mathcal{I}(\delta) = -\Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X} = \cdot)),$$

here $\Pi_{\mathcal{A}}(\cdot)$ denotes the projection operator onto the additive space $\mathcal{A}$. Thus, the Fisher information matrix bound for estimating β is given by

$$\mathcal{I}(\delta^*) = \mathbb{E}[g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)]^2\mathbb{E}\left[(\mathbf{Z} - \Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X})))(\mathbf{Z} - \Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X})))^\top\right].$$

Whenever the errors in the model (2.2) are assumed to be Gaussian, the least square estimator is equal to the maximum likelihood estimator and we have

$$\mathbb{E}[g'_\varepsilon(\epsilon)/g_\varepsilon(\epsilon)]^2 = \int g'^2_\varepsilon(x)/g_\varepsilon(x)dx = \frac{1}{\sigma_\varepsilon^2}.$$

Thus

$$\mathcal{I}(\delta^*) = \frac{1}{\sigma_\varepsilon^2}\mathbb{E}\left[(\mathbf{Z} - \Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X})))\,(\mathbf{Z} - \Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X})))^\top\right].$$

Recall that $\tilde{\mathbf{Z}} = \mathbf{Z} - \sum_{j=1}^n W_{nj}(\mathbf{X})\mathbf{Z}_j := \mathbf{Z} - \sum_{l=1}^d \widetilde{W}_n^l(X_l)$, where $\widetilde{W}_n^l(X_{il}) = \frac{1}{n}\sum_{j=1}^n \frac{1}{\mathbf{g}_n(\mathbf{X}_j)}\left[\frac{1}{h_n}K\left(\frac{X_{il}-X_{jl}}{h_n}\right)D_l - (d-1)/d\int_{\mathbb{R}^d}\prod_{k=1}^d \frac{1}{h_n}K_k\Big(\frac{x_k-X_{jk}}{h_n}\Big)\mathbf{q}(\mathbf{x})d\mathbf{x}\right]\mathbf{Z}_j$ and $D_l = \int_{\mathbb{R}^{d-1}}\prod_{k=1,k\neq l}\frac{1}{h_n}K_k\left(\frac{x_k-X_{jk}}{h_n}\right)\mathbf{q}_{-l}(\mathbf{x}_{-l})d\mathbf{x}_{-l}$. It is worth noticing that, our estimator $\widehat{\beta}$ is semiparametrically efficient since $\Pi_{\mathcal{A}}(\mathbb{E}(\mathbf{Z}|\mathbf{X}))$ can be estimated by $\sum_{l=1}^d \widetilde{W}_n^l(X_l)$. It would be of interest to provide a complete investigation of the semiparametric efficiency question in our framework which requires nontrivial mathematics, that goes well beyond the scope of the present work.

2.6 Simulations

Hereafter we display some numerical results that show how the proposed test perform when testing the additivity against the nonadditivity of the model. Considering the following models

1. $\mathrm{Y} = \mathbf{Z}^\top\boldsymbol{\beta} + X_1 + X_2 + CX_1X_2$,
2. $\mathrm{Y} = \mathbf{Z}^\top\boldsymbol{\beta} + \sin(\pi X_1) + \sin(\pi X_2) + C\sin(\pi X_1)\sin(\pi X_2)$,
3. $\mathrm{Y} = \mathbf{Z}^\top\boldsymbol{\beta} + \exp(X_1) + \exp(X_2) + C\exp(X_1)\exp(X_2)$,

where X_1 and X_2 are $\mathcal{N}(0,1)$ random variables, C is a positive constant taking various values from the additivity ($C = 0$) to the far away position inside the nonadditivity. The deterministic vector $\boldsymbol{\beta}$ is chosen as the vector $(0.5, 1)$ while $\mathbf{Z}$ is taken as a gaussian random vector. In our simulations, samples of sizes $n = 100$, $n = 500$ and $n = 1000$ have been drawn following the scheme that has been described and $m = 500$ replicates have been considered to estimate the first kind error risk α, taken to be 5%, together with the power of the test. The obtained results

are displayed in the following table.

Model 1			
n	100	500	1000
C=0	0.309	0.102	0.0553
C=0.1	0.357	0.853	1
C=0.5	0.420	0.969	1
C=1	0.703	1	1
C=2	0.896	1	1

Model 2			
n	100	500	1000
C=0	0.246	0.091	0.0519
C=0.1	0.257	0.762	0.952
C=0.5	0.215	0.819	1
C=1	0.562	1	1
C=2	0.829	1	1

Model 3			
n	100	500	1000
C=0	0.512	0.113	0.0592
C=0.1	0.507	0.743	0.968
C=0.5	0.487	0.785	0.994
C=1	0.729	1	1
C=2	0.764	1	1

It is clear that the power of the test increases to reach the value 1 for large sample sizes while the first kind error risk estimate is very close to the theoretical value 5% when the sample size is $n = 1000$.

2.7 Proof

The proof section is split up into two parts. The first part gives some intermediate lemmas while the second one displays proofs of our theorems.

2.7.1 Some intermediate lemmas

First recall the following lemma which is needed in proving our results.

Lemma 2.1. (Liang (2000), Lemma A.1) Let $V_1, \ldots, V_n$ be independent random variables with zero means and finite variances, i.e., $\sup_{1\leq k\leq n}\mathbb{E}|V_k|^r \leq C$ $(for\ some\ r \geq 2)$. Assume that $(a_{ki})_{1\leq i,k\leq n}$, is a sequence such that $\max_{1\leq i,k\leq n}|a_{ki}| \leq n^{-p_1}$ for some $0 < p_1 < 1$ and $\sum_{k=1}^{n} a_{ki} = \mathcal{O}(n^{p_2})$ for $p_2 \geq \max(0, 2/r - p_1)$. Then, for $s = (p_1 - p_2)/2$,

$$\max_{1\leq i\leq n}\left|\sum_{k=1}^{n} a_{ki}V_k\right| = \mathcal{O}\Big(n^{-s}\log n\Big) \quad a.s.$$

Hereafter we prove some technical lemmas that are steps in establishing the main results of this chapter.

Lemma 2.2. Assume that conditions [G.2-3], [H.1], [K.1-2], [Q.1-2] are satisfied. Suppose, for some $r \geq 2$, that $\sup_{1\leq i\leq n}\mathbb{E}\,|\varepsilon_i|^r < \infty$. Then, we have

$$\begin{aligned}
&(i) \quad \max_{1\leq i\leq n}\max_{1\leq j\leq n}\Big|W_{nj}(\mathbf{X}_i)\Big| = \mathcal{O}\left(\frac{1}{nh_n}\right) \quad a.s.,\\
&(ii) \quad \max_{1\leq i\leq n}\left|\sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\right| = \mathcal{O}(1) \quad a.s.,\\
&(iii) \quad \max_{1\leq i\leq n}\left|\sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\varepsilon_j\right| = \mathcal{O}(n^{\frac{-k}{2k+1}}\log n) \quad a.s.
\end{aligned}$$

Proof : • **Part**(i) Recall that,

$$\begin{aligned}
&W_{nj}(\mathbf{X}_i)\\
=\ &\sum_{l=1}^{d}\frac{1}{n\mathbf{g}_n(\mathbf{X}_j)}\left[\frac{1}{h_n}K_l\Big(\frac{X_{il}-X_{jl}}{h_n}\Big)\int_{\mathbb{R}^{d-1}}\prod_{k=1,k\neq l}^{d}\frac{1}{h_n}K_k\Big(\frac{z_k-X_{jk}}{h_n}\Big)\mathbf{q}_{-l}(\mathbf{z}_{-l})d\mathbf{z}_{-l}\right.\\
&\left.-\Big(\frac{d-1}{d}\Big)\int_{\mathbb{R}^d}\prod_{k=1}^{d}\frac{1}{h_n}K_k\Big(\frac{z_k-X_{jk}}{h_n}\Big)\mathbf{q}(\mathbf{z})d\mathbf{z}\right].
\end{aligned}$$

Consider first the case where the covariate density $\mathbf{g}$ is known. Towards this end, replace $W_{nj}(\mathbf{X}_i)$ by

$$\begin{aligned}
&W^{\mathbf{g}}_{nj}(\mathbf{X}_i)\\
=\ &\sum_{l=1}^{d}\frac{1}{n\mathbf{g}(\mathbf{X}_j)}\left[\frac{1}{h_n}K_l\Big(\frac{X_{il}-X_{jl}}{h_n}\Big)\times\int_{\mathbb{R}^{d-1}}\prod_{k=1,k\neq l}^{d}\frac{1}{h_n}K_k\Big(\frac{z_k-X_{jk}}{h_n}\Big)\mathbf{q}_{-l}(\mathbf{z}_{-l})d\mathbf{z}_{-l}\right.\\
&\left.-\Big(\frac{d-1}{d}\Big)\int_{\mathbb{R}^d}\prod_{k=1}^{d}\frac{1}{h_n}K_k\Big(\frac{z_k-X_{jk}}{h_n}\Big)\mathbf{q}(\mathbf{z})d\mathbf{z}\right].
\end{aligned}$$

Since by condition (K.2) the kernels K_k's are compactly supported, making use of Bochner's Theorem, it follows, for any $j \in \{1, \ldots, n\}$ and any $k \in \{1, \ldots, d\}$, that we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{1}{h_n} K_k\Big(\frac{z_k - X_{jk}}{h_n}\Big) q_k(z_k) dz_k = q_k(X_{jk}).$$

Similarly, we obtain

$$\lim_{n\to\infty} \int_{\mathbb{R}^{d-1}} \prod_{k=1,k\neq l}^{d} \frac{1}{h_n} K_k\Big(\frac{z_k - X_{jk}}{h_n}\Big) \mathbf{q}_{-l}(\mathbf{z}_{-l}) d\mathbf{z}_{-l} = \prod_{k=1,k\neq l}^{d} q_k(X_{jk})$$

and also

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \prod_{k=1}^{d} \frac{1}{h_n} K_k\Big(\frac{z_k - X_{jk}}{h_n}\Big) \mathbf{q}(\mathbf{z}) d\mathbf{z} = \prod_{k=1}^{d} q_k(X_{jk}) = \mathbf{q}(X_j).$$

Therefore, under Condition (G.2) and the fact that the kernels K_k's are bounded, we have

$$\sum_{l=1}^{d} \frac{1}{n\mathbf{g}(\mathbf{X}_j)} \left(\frac{1}{h_n} K_l\Big(\frac{X_{il} - X_{jl}}{h_n}\Big) \prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \Big(\frac{d-1}{d}\Big) \prod_{k=1}^{d} q_k(X_{jk}) \right) = \mathcal{O}\Big((nh_n)^{-1}\Big).$$

Now, consider the case where the marginal density function $\mathbf{g}$ is unknown and estimated by $\mathbf{g_n}$. Subsequently, observe that the decomposition

$$\begin{aligned}
&\sum_{l=1}^{d} \frac{1}{n\mathbf{g}_n(\mathbf{X}_j)} \left(\frac{1}{h_n} K_l\Big(\frac{X_{il} - X_{jl}}{h_n}\Big) \prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \Big(\frac{d-1}{d}\Big) \prod_{k=1}^{d} q_k(X_{jk}) \right) \\
= &\sum_{l=1}^{d} \left[\frac{1}{n\mathbf{g}(\mathbf{X}_j)} \left(\frac{1}{h_n} K_l\Big(\frac{X_{il} - X_{jl}}{h_n}\Big) \prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \Big(\frac{d-1}{d}\Big) \prod_{k=1}^{d} q_k(X_{jk}) \right) \right. \\
&\left. - \frac{\mathbf{g}_n(\mathbf{X}_j) - \mathbf{g}(\mathbf{X}_j)}{n\mathbf{g}_n(\mathbf{X}_j)\mathbf{g}(\mathbf{X}_j)} \left(\frac{1}{h_n} K_l\Big(\frac{X_{il} - X_{jl}}{h_n}\Big) \prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \Big(\frac{d-1}{d}\Big) \prod_{k=1}^{d} q_k(X_{jk}) \right) \right]
\end{aligned} \tag{2.14}$$

holds true. Since by (G.3) the density $\mathbf{g}$ is uniformly continuous on its support and the bandwidth h_n satisfies the condition (H.1), which clearly implies the well-known Csörgo-Révész-Stute conditions, following Stute (1984), we have

$$\sup_{\mathbf{x}\in I^d} |\mathbf{g}_n(\mathbf{x}) - \mathbf{g}(\mathbf{x})| = \mathcal{O}\left(\sqrt{\frac{\log h_n^{-d}}{nh_n^d}} \right) \quad a.s. \tag{2.15}$$

Consequently, it follows from the statement (2.14) that

$$\sum_{l=1}^{d} \frac{1}{n\mathbf{g}_n(\mathbf{X}_j)}\left(\frac{1}{h_n}K_l\Big(\frac{X_{il}-X_{jl}}{h_n}\Big)\prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \Big(\frac{d-1}{d}\Big)\prod_{k=1}^{d} q_k(X_{jk})\right) \quad (2.16)$$
$$= \mathcal{O}\left(\frac{1}{nh}\right) + \mathcal{O}\left(\frac{1}{nh_n}\sqrt{\frac{\log h_n^{-d}}{nh_n^d}}\right) \ a.s.$$

This concludes the proof of Part (i) of Lemma 2.2.

• **Part**(ii) Observe that, for any $1 \leq l \leq d$, one may write

$$\mathbf{g}(x_1, \cdots, x_d) = g_{X_l}(x_l) \times g_{(X_1,\cdots,X_{l-1},X_{l+1},\cdots,X_d|X_l=x_l)}(x_1, \cdots, x_{l-1}, x_{l+1}, \cdots, x_d),$$

where $g_{(X_1,\cdots,X_{l-1},X_{l+1},\cdots,X_d|X_l=x_l)}(x_1, \cdots, x_{l-1}, x_{l+1}, \cdots, x_d)$ is the conditional density of the random vector $(X_1, \cdots, X_{l-1}, X_{l+1}, \cdots, X_d)$ given $X_l = x_l$ and g_{X_l} is the marginal density of X_l. Moreover, using the strong law of large numbers, it follows, for any $1 \leq l \leq d$ and any $x_l \in \mathbb{R}$, that

$$\lim_{n\to\infty} \frac{1}{nh_n}\sum_{j=1}^{n} \frac{1}{g_{X_{jl}}(X_{jl})}K_l\left(\frac{x_l - X_{jl}}{h_n}\right) = 1 \ \ a.s.$$

Therefore, making use of Condition (G.2), we obtain for any $1 \leq l \leq d$

$$\sum_{j=1}^{n} \frac{1}{n\mathbf{g}(\mathbf{X}_j)}\left[\frac{1}{h_n}K_l\left(\frac{x_l - X_{jl}}{h_n}\right)\prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \left(\frac{d-1}{d}\right)\prod_{k=1}^{d} q_k(X_{jk})\right] = \mathcal{O}(1) \quad a.s.$$

Proceeding similarly as in Part (i) and using the decomposition (2.14), we obtain

$$\sum_{j=1}^{n} \frac{1}{n\mathbf{g}_n(\mathbf{X}_j)}\left[\frac{1}{h_n}K_l\left(\frac{X_{il} - X_{jl}}{h_n}\right)\prod_{k=1,k\neq l}^{d} q_k(X_{jk}) - \left(\frac{d-1}{d}\right)\prod_{k=1}^{d} q_k(X_{jk})\right]$$
$$= \mathcal{O}(1) + \mathcal{O}\left(\sqrt{\frac{\log h_n^{-d}}{nh_n^d}}\right) \ a.s.$$

This states the result.

• **Part** (iii). Using Lemma 2.1 and the parts (i) and (ii) of Lemma 2.2, the proof of (iii) follows straightforwardly as we take $a_{ji} = W_{nj}(\mathbf{X}_i)$ and $V_j = \varepsilon_j$. □

To set the next lemma, we need to introduce further notations. In this respect, for any $1 \leq i \leq n$ and $1 \leq l \leq p$, set $\tilde{m}_{add}(\mathbf{X}_i) = m_{add}(\mathbf{X}_i) - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)m_{add}(\mathbf{X}_j)$.

Lemma 2.3. Under assumptions of Theorem 2.1, we have

$$\max_{1\leq i\leq n}\left|\tilde{m}_{add}(\mathbf{X}_i)\right| = \mathcal{O}\left(n^{\frac{-k}{2k+1}}\log n\right) \quad a.s.$$

Proof : Considering the statements (2.4), (2.6)-(2.8), and (2.11), it is easily seen that

$$\begin{aligned}
\max_{1\leq i\leq n}|\tilde{m}_{add}(\mathbf{X}_i)| &= \max_{1\leq i\leq n}\left|m_{add}(\mathbf{X}_i)-\sum_{j=1}^{n}m_{add}(\mathbf{X}_j)W_{nj}(\mathbf{X}_i)\right| \\
&= \max_{1\leq i\leq n}\left|m_{add}(\mathbf{X}_i)-\sum_{j=1}^{n}(Y-\mathbf{Z}^\top\boldsymbol{\beta})W_{nj}(\mathbf{X}_i)+\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\epsilon_i\right| \\
&\leq \max_{1\leq i\leq n}\left|m_{add}(\mathbf{X}_i)-\hat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)\right|+\max_{1\leq i\leq n}\left|\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\epsilon_j\right|. \qquad (2.17)
\end{aligned}$$

Making use of Lemma 1 due to Camlong-Viot (2001) where all the conditions related to the mixing setting considered there are relaxed, it follows, under hypotheses [G.1-2 ;4], [H.1], [K.1-3], [M.1] and [Q.1-2], that

$$\max_{1\leq i\leq n}|m_{add}(\mathbf{X}_i)-\hat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)| = \mathcal{O}\left(\sqrt{\frac{\log n}{nh_n}}\right) \quad a.s. \qquad (2.18)$$

Therefore, considering the statement (2.17) and using Lemma 2.2 (iii), we obtain the result. □

Lemma 2.4. Let $V_1,\ldots,V_n$ be independent random variables with mean 0 such that $\max_{1\leq k\leq n}\mathbb{E}|V_k|^\alpha<\infty$ for some $\alpha>2$. Suppose that $\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}Var(V_k)>0$, then we have

$$\limsup_{n\to\infty}\frac{|S_n|}{(2s_n^2\log\log s_n^2)^{1/2}}=1 \qquad a.s.,$$

where $S_n=\sum_{k=1}^{n}V_k$ and $s_n^2=\sum_{k=1}^{n}Var(V_k)$

Proof See, for instance, Stout (1974) Corollary 5.2.3. □

2.7.2 Proof of Theorems

Proof of Theorem 2.1

To state our results, let

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^\top = B \qquad a.s., \qquad (2.19)$$

where B is a $p \times p$-positive definite matrix. Recall that $\tilde{Y} = \tilde{\mathbf{Z}}^\top \boldsymbol{\beta} + \tilde{m}_{add}(\mathbf{X}) + \tilde{\varepsilon}$, $\widehat{\boldsymbol{\beta}} = [\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top]^{-1}\tilde{\mathbf{Z}}\tilde{Y}$ and $\tilde{\varepsilon}_i = \varepsilon_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j$. Thus,

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \left(n^{-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top\right)^{-1}\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \tilde{\mathbf{Z}}_i \tilde{m}_{add}(\mathbf{X}_i)\right. \\ &\quad \left. - \sum_{i=1}^n \tilde{\mathbf{Z}}_i \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j + \sum_{i=1}^n \tilde{\mathbf{Z}}_i \varepsilon_i\right). \qquad (2.20)\end{aligned}$$

Given that $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\mathbf{Z}_j$, for any $1 \leq i \leq n$, the first term in the right side of the equation (2.20) may be written as

$$\sum_{i=1}^n \tilde{\mathbf{Z}}_i \tilde{m}_{add}(\mathbf{X}_i) = \sum_{i=1}^n \mathbf{Z}_i \tilde{m}_{add}(\mathbf{X}_i) - \sum_{i=1}^n \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \tilde{m}_{add}(\mathbf{X}_i). \qquad (2.21)$$

Since, by Lemma 2.3, $\max_{1\leq i\leq n} |\tilde{m}_{add}(\mathbf{X}_i)| = \mathcal{O}\left(n^{(-k/2k+1)}\log n\right)$ almost surely. The classical L.I.L implies that $\sum_{i=1}^n \mathbf{Z}_i = \mathcal{O}(\sqrt{n\log\log n})$ $a.s.$ Thus, we have

$$\sum_{i=1}^n \mathbf{Z}_i \tilde{m}_{add}(\mathbf{X}_i) = o(\sqrt{n}) \quad a.s. \qquad (2.22)$$

Considering the second term of the statement (2.21). Observe that

$$\sum_{i=1}^n \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \tilde{m}_{add}(\mathbf{X}_i) \leq n \max_{1\leq\ i\leq n} \tilde{m}_{add}(\mathbf{X}_i) \max_{1\leq\ i\leq n} \max_{1\leq\ j\leq n} W_{nj}(\mathbf{X}_i) \sum_{j=1}^n \mathbf{Z}_j.$$

Since, by Lemma 2.2, $\max_{1\leq\ i\leq n} \max_{1\leq j\leq n} W_{nj}(\mathbf{X}_i) = \mathcal{O}\left((nh_n)^{-1}\right)$ a.s. From Lemma 2.3, $\max_{1\leq i\leq n} \tilde{m}_{add}(\mathbf{X}_i) = \mathcal{O}\left(n^{\frac{-k}{2k+1}}\log n\right)$ $a.s.$ and we have $\sum_{i=1}^n \mathbf{Z}_i = \mathcal{O}(\sqrt{n\log\log n})$ $a.s.$. It follows then that

$$\sum_{i=1}^n \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \tilde{m}_{add}(\mathbf{X}_i) = o(\sqrt{n}) \quad a.s. \qquad (2.23)$$

Combining the statements (2.22) and (2.23), one can show that

$$\sum_{i=1}^n \tilde{\mathbf{Z}}_i \tilde{m}_{add}(\mathbf{X}_i) = o(\sqrt{n}) \quad a.s. \qquad (2.24)$$

While handling the second term of the second part of the statement (2.20), we have

$$\begin{aligned}&\sum_{i=1}^n \tilde{\mathbf{Z}}_i \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j \\ &= \sum_{i=1}^n \left(\mathbf{Z}_i - \sum_{q=1}^n W_{nq}(\mathbf{X}_i)\mathbf{Z}_q\right) \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j \\ &= \sum_{i=1}^n \mathbf{Z}_i \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j - \sum_{i=1}^n \sum_{q=1}^n W_{nq}(\mathbf{X}_i)\mathbf{Z}_q \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\varepsilon_j. \qquad (2.25)\end{aligned}$$

Again since, by Lemma 2.2, $\max_{1\leq i\leq n}\sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\varepsilon_j = \mathcal{O}\Big(n^{-(k/2k+1)}\log n\Big)$ almost surely. We have $\sum_{i=1}^{n}\mathbf{Z}_i = \mathcal{O}(\sqrt{n\log\log n})$ a.s.. Thus we obtain

$$\sum_{i=1}^{n}\mathbf{Z}_i\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\varepsilon_j = o(\sqrt{n}) \quad a.s. \tag{2.26}$$

Proceeding similarly as before, and using assumption [H.1] combining with the fact that $\max_{1\leq i\leq n}\max_{1\leq q\leq n} W_{nq}(\mathbf{X}_i) = \mathcal{O}((nh_n)^{-1})$ a.s., we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{q=1}^{n}W_{nq}(\mathbf{X}_i)\mathbf{Z}_q\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\varepsilon_j &\leq n\max_{1\leq i\leq n}\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\varepsilon_j\max_{1\leq i\leq n}\max_{1\leq q\leq n}W_{nq}(\mathbf{X}_i)\sum_{q=1}^{n}\mathbf{Z}_q\\ &= o(\sqrt{n}) \quad a.s. \end{aligned}\tag{2.27}$$

Therefore, combining the statements (2.25)-(2.27), we obtain

$$\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\varepsilon_j = o(\sqrt{n}) \quad a.s. \tag{2.28}$$

Combining the statements (2.20), (2.24), (2.19) and (2.28) , it follows that

$$\sqrt{n}\Big(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\Big) = \Big(B^{-1}+o(1)\Big)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\varepsilon_i + o(1)\right) \quad a.s. \tag{2.29}$$

Recall that $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i - \sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\mathbf{Z}_j$. We have $\sum_{j=1}^{n}W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \to \sum_{l=1}^{d}W^l(X_l)$ $a.s.$, see the statement (2.13). Thus, by the central limit theorem, the random sequence $(1/\sqrt{n})\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\varepsilon_i$ converges in distribution to a Gaussian distribution with mean zero and covariance matrix $\sigma_\varepsilon^2 B$. Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0,\sigma_\varepsilon^2 B^{-1}). \tag{2.30}$$

This completes the proof of Theorem 2.1. □

Proof of Theorem 2.2

Recall that $B^{-1} = (b^{ij})_{ij}$ and set $\mathbf{b}^j = (b^{j1},\ldots,b^{jp})^\top$. From the statement (2.29), it's clear that the behavior of $(\widehat{\beta}-\beta)_j$ is the same as $\frac{1}{n}\sum_{i=1}^{n}\mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i$.
Moreover, for any $1\leq i\leq n$, we have $\mathbb{E}\mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i = \mathbf{b}^{j\top}\mathbb{E}[\tilde{\mathbf{Z}}_i\mathbb{E}(\varepsilon_i|\mathbf{X}_i,\mathbf{Z}_i)] = 0$ and

$$\mathbb{E}\Big|\mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i\Big|^r \leq C\max_{1\leq i\leq n}\mathbb{E}\Big|\varepsilon_i\Big|^r < \infty,$$

since the random vector $\mathbf{Z}$ is bounded. Furthermore, observe that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Var(\mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i) &= \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i)^2\\ &= \sigma_\varepsilon^2(b^{j1},\ldots,b^{jp})\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i{}^\top)(b^{j1},\ldots,b^{jp})^\top\\ &= \sigma_\varepsilon^2(b^{j1},\ldots,b^{jp})B(b^{j1},\ldots,b^{jp})^\top \ = \sigma_\varepsilon^2 b^{jj} > 0.\end{aligned}$$

Taking then $V_i = \mathbf{b}^{j\top}\tilde{\mathbf{Z}}_i\varepsilon_i$ and $s_n^2 = n\sigma_\varepsilon^2 b^{jj}$ in Lemma 2.4, we obtain

$$\limsup_{n\to\infty}\left|\left(\frac{1}{2n\log\log n}\right)^{1/2}\sum_{i=1}^{n}\mathbf{b}^{jT}\tilde{\mathbf{Z}}_i\varepsilon_i\right| = \left(\sigma_\varepsilon^2 b^{jj}\right)^{1/2} \quad a.s. \tag{2.31}$$

Combining then the statements (2.29) and (2.31), the result follows. □

Proof of Theorem 2.3

For the sake of simplicity, let $C > 0$ denote a constant which may have different values at each appearance in the sequel. Considering the quantity

$$R_n := \int_{\mathbb{R}^d}\left[\frac{1}{nh_n^d}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})+m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)+\varepsilon_i\right)\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}_n^2(\mathbf{x})}d\mathbf{x}$$

together with the decomposition

$$\frac{1}{\mathbf{g}_n} = \frac{1}{\mathbf{g}} - \frac{\mathbf{g}_n-\mathbf{g}}{\mathbf{g}\mathbf{g}_n},$$

it follows that

$$\begin{aligned}&R_n\\ &:= \int_{\mathbb{R}^d}\left[\frac{1}{nh_n^d}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})+m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)+\varepsilon_i\right)\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}\\ &+\int_{\mathbb{R}^d}\left[\frac{1}{nh_n^d}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})+m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)+\varepsilon_i\right)\right]^2\\ &\quad\times\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})\mathbf{g}_n^2(\mathbf{x})}(\mathbf{g}^2(\mathbf{x})-\mathbf{g}_n^2(\mathbf{x}))d\mathbf{x}.\end{aligned}$$

Moreover, observe that

$$\mathbf{g}^2(\mathbf{x})-\mathbf{g}_n^2(\mathbf{x}) = -\Big(\mathbf{g}(\mathbf{x})-\mathbf{g}_n(\mathbf{x})\Big)^2+2\mathbf{g}(\mathbf{x})\Big(\mathbf{g}(\mathbf{x})-\mathbf{g}_n(\mathbf{x})\Big). \tag{2.32}$$

Making use of the statement (2.15), we obtain

$$R_n := R_n^1\Big(1+\mathcal{O}\Big(\sqrt{\frac{\log h_n^{-d}}{nh_n^d}}\Big)\Big), \quad a.s.,$$

where,

$$\begin{aligned} & R_n^1 \\ &:= \int_{\mathbb{R}^d}\left[\frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\Big(\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})+m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)+\varepsilon_i\Big)\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}. \end{aligned}$$

Observe now that R_n^1 may be written as

$$R_n^1 := R_{n,1}^1+R_{n,2}^1+R_{n,3}^1+2R_{n,4}^1+2R_{n,5}^1+2R_{n,6}^1+R_{n,7}^1+2R_{n,8}^1+2R_{n,9}^1,$$

with

$$\begin{aligned} R_{n,1}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\varepsilon_i^2\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,2}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\frac{\varepsilon_i\varepsilon_j\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,3}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\left[\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,4}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) \\ &\qquad \times\Big(m_{add}(\mathbf{X}_j)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_j)\Big)\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,5}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\frac{\varepsilon_i\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,6}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\frac{\varepsilon_j\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,7}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\left[\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\Big(m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)\Big)\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,8}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\Big(m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)\Big)\frac{\varepsilon_i\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}, \\ R_{n,9}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\Big(m_{add}(\mathbf{X}_i)-\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{X}_i)\Big)\frac{\varepsilon_j\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}. \end{aligned}$$

Therefore, to prove the theorem, it suffices to establish the following assertions

$$nh_n^d R_{n,1}^1 = D + o_p(n^{1/2}h_n^{d/2}), \tag{2.33}$$

$$\sqrt{nh_n^d}\frac{R_{n,2}^1}{\sqrt{V}} \xrightarrow{d} \mathcal{N}(0,1), \tag{2.34}$$

$$R_{n,\tau}^1 = o_p(n^{-1/2}h^{-d/2}) \quad \text{for} \quad \tau \in \{3,...,9\}. \tag{2.35}$$

Theorem, we have

Proof of (2.33) : Considering first the expectation of $R_{n,1}^1$ and making use of Fubini's

$$\begin{aligned}
\mathbb{E}[R_{n,1}^1] &= \mathbb{E}\left[\int_{\mathbb{R}^d}(\frac{1}{nh_n^d})^2\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\varepsilon_i^2\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}\right] \\
&= \frac{1}{nh_n^{2d}}\int_{\mathbb{R}^d}\mathbb{E}\left[\varepsilon_1^2K^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right]\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x} \\
&= \frac{1}{nh_n^{2d}}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_1^2]\mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right]\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x} \\
&= \frac{1}{nh_n^{2d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\sigma_\varepsilon^2K^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mathbf{g}(\mathbf{u})\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{u}d\mathbf{x},
\end{aligned}$$

where σ_ε^2 is the variance of ε. Taking $\mathbf{p} = \dfrac{\mathbf{q}}{\mathbf{g}^2}$ and making the change of variable $\mathbf{s} = \dfrac{\mathbf{x}-\mathbf{u}}{h_n}$, it follows that

$$\mathbb{E}[R_{n,1}^1] = \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\sigma_\varepsilon^2K^2(\mathbf{s})\,\mathbf{g}(\mathbf{u})\mathbf{p}(\mathbf{u}+h_n\mathbf{s})d\mathbf{s}d\mathbf{u}.$$

By Taylor series expansion of the function $\mathbf{p}$, we obtain

$$\mathbb{E}[R_{n,1}^1] = \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\sigma_\varepsilon^2\mathbf{g}(\mathbf{u})\mathbf{p}(\mathbf{u})d\mathbf{u}\int_{\mathbb{R}^d}K^2(\mathbf{v})\,d\mathbf{v} + o(1).$$

To evaluate the variance of $R_{n,1}^1$, observe that

$$\begin{aligned}
&Var[R_{n,1}^1] \\
&= \frac{1}{n^4h_n^{4d}}Var\left[\sum_{i=1}^n\int_{\mathbb{R}^d}\varepsilon_i^2K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\right] \\
&= \frac{1}{n^4h_n^{4d}}\sum_{i=1}^n\mathbb{E}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\varepsilon_i^4K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K^2\left(\frac{\mathbf{y}-\mathbf{X}_i}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{y})d\mathbf{x}d\mathbf{y}\right] \\
&\quad -\frac{1}{n^4h_n^{4d}}\sum_{i=1}^n\left(\mathbb{E}\left[\int_{\mathbb{R}^d}\varepsilon_i^2K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\right]\mathbb{E}\left[\int_{\mathbb{R}^d}\varepsilon_i^2K^2\left(\frac{\mathbf{y}-\mathbf{X}_i}{h_n}\right)\mathbf{p}(\mathbf{y})d\mathbf{y}\right]\right) \\
&= \frac{1}{n^3h_n^{4d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_1^4]K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)K^2\left(\frac{\mathbf{y}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{y})\mathbf{g}(\mathbf{v}) \\
&\quad \times d\mathbf{x}d\mathbf{y}d\mathbf{v} - \frac{1}{n^3h_n^{4d}}\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_1^2]K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\mathbf{g}(\mathbf{v})d\mathbf{v}\right]^2.
\end{aligned}$$

Taking $\mathbf{s} = \dfrac{\mathbf{x}-\mathbf{v}}{h_n}$ and $\mathbf{t} = \dfrac{\mathbf{y}-\mathbf{v}}{h_n}$ and then integrating, we obtain

$$\begin{aligned} Var[R^1_{n,1}] &= \frac{1}{n^3 h_n^{2d}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{E}[\varepsilon_1^4] K^2(\mathbf{s})\, K^2(\mathbf{t})\, \mathbf{p}(\mathbf{s}h_n+\mathbf{v})\mathbf{p}(\mathbf{t}h_n+\mathbf{v})\mathbf{g}(\mathbf{v}) \\ &\quad \times d\mathbf{s}d\mathbf{t}d\mathbf{v} - \frac{1}{n^3 h_n^{2d}} \left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \sigma_\varepsilon^2 K^2(\mathbf{s})\, \mathbf{p}(\mathbf{s}h_n+\mathbf{v})\mathbf{g}(\mathbf{v}) d\mathbf{s}d\mathbf{v}\right]^2 \\ &= \mathcal{O}\Big(n^{-3}h_n^{-2d}\Big). \end{aligned}$$

Straightforwardly, by an immediate application of Tchebychev's inequality, it follows that

$$R^1_{n,1} = n^{-1}h_n^{-d}D + o_p(n^{-1/2}h_n^{-d/2}).$$

Proof of (2.34) : Observe that one may write

$$\begin{aligned} \sqrt{nh_n^d}R^1_{n,2} &= \int_{\mathbb{R}^d} \sqrt{\frac{1}{n^3h_n^{3d}}} \sum_{i\neq j}^{n} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right) \varepsilon_i\varepsilon_j \mathbf{p}(\mathbf{x})d\mathbf{x} \\ &= 2\sum_{1\le i<j\le n} \sqrt{\frac{1}{n^3h_n^{3d}}} \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right) \varepsilon_i\varepsilon_j \mathbf{p}(\mathbf{x})d\mathbf{x} \\ &= 2\sum_{1\le i<j\le n} \mathcal{H}_n(\zeta_i,\zeta_j), \end{aligned}$$

where $\zeta_i = (\mathbf{X}_i, \varepsilon_i)$ and

$$\mathcal{H}_n(\zeta_i,\zeta_j) = \sqrt{\frac{1}{n^3h_n^{3d}}} \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right) \varepsilon_i\varepsilon_j \mathbf{p}(\mathbf{x})d\mathbf{x}.$$

Therefore, the asymptotic normality of $\sqrt{nh_n^d}R^1_{n,2}$ follows by establishing the result for the following generalized U-statistics

$$U_n = \sum_{1\le i<j\le n} \mathcal{H}_n(\zeta_i,\zeta_j)$$

It suffices then to state the following statements

(**a**) $\mathbb{E}[\mathcal{H}_n(\zeta_i,\zeta_j)] = 0$,

(**b**) $n^3\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\right] = \frac{V}{2} + o(1)$,

(**c**) $n^3\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\mathbb{1}_{\mathcal{H}_n(\zeta_i,\zeta_j)>N}\right] = o(1)$ holds for any $N>0$ (Lindeberg's condition).

Proof of (**a**) : Notice that

$$\begin{aligned}
&\mathbb{E}[\mathcal{H}_n(\zeta_i,\zeta_j)]\\
=\;& \mathbb{E}\left[\sqrt{\frac{1}{n^3h_n^{3d}}}\int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\varepsilon_i\varepsilon_j\mathbf{p}(\mathbf{x})d\mathbf{x}\right]\\
=\;& \sqrt{\frac{1}{n^3h_n^{3d}}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_j]\mathbf{g}(\mathbf{u_1})\mathbf{g}(\mathbf{u_2})K\left(\frac{\mathbf{x}-\mathbf{u_1}}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{u_2}}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{u_1}d\mathbf{u_2}d\mathbf{x}\\
=\;& \sqrt{\frac{1}{n^3h_n^{3d}}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\mathbb{E}[\varepsilon_1])^2\mathbf{g}(\mathbf{u_1})\mathbf{g}(\mathbf{u_2})K\left(\frac{\mathbf{x}-\mathbf{u_1}}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{u_2}}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{u_1}d\mathbf{u_2}d\mathbf{x}.
\end{aligned}$$

Since $\mathbb{E}[\varepsilon_1]=0$, then we have

$$\mathbb{E}[\mathcal{H}_n(\zeta_i,\zeta_j)] \;=\; 0. \tag{2.36}$$

Proof of (**b**) : To calculate the variance of the U-statistic, observe that

$$\begin{aligned}
&\mathbb{E}[\mathcal{H}_n^2(\zeta_i,\zeta_j)]\\
=\;& \frac{1}{n^3h_n^{3d}}\mathbb{E}\left[\int_{\mathbb{R}^d}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\varepsilon_i\varepsilon_j\mathbf{p}(\mathbf{x})d\mathbf{x}\right]^2\\
=\;& \frac{1}{n^3h_n^{3d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}\left[\varepsilon_i^2\varepsilon_j^2K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)K\left(\frac{\mathbf{y}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{y}-\mathbf{X}_j}{h_n}\right)\right]\\
&\times\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{y})d\mathbf{x}d\mathbf{y}\\
=\;& \frac{1}{n^3h_n^{3d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_i^2]\mathbb{E}[\varepsilon_j^2]K\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)K\left(\frac{\mathbf{y}-\mathbf{u}}{h_n}\right)\\
&\times K\left(\frac{\mathbf{y}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{y})d\mathbf{u}d\mathbf{v}d\mathbf{x}d\mathbf{y}\\
=\;& \frac{1}{n^3h_n^{3d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\mathbb{E}[\varepsilon_1^2])^2K\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)K\left(\frac{\mathbf{y}-\mathbf{u}}{h_n}\right)\\
&\times K\left(\frac{\mathbf{y}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{y})d\mathbf{u}d\mathbf{v}d\mathbf{x}d\mathbf{y}.
\end{aligned}$$

By the changes of variables $\mathbf{s}=\dfrac{\mathbf{x}-\mathbf{u}}{h_n}$, $\mathbf{t}=\dfrac{\mathbf{y}-\mathbf{u}}{h_n}$, $\mathbf{z}=\dfrac{\mathbf{v}-\mathbf{u}}{h_n}$, we obtain

$$\begin{aligned}
\mathbb{E}[\mathcal{H}_n^2(\zeta_i,\zeta_j)] \;=\;& \frac{1}{n^3}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}[\varepsilon_1^2]^2\mathbf{g}(\mathbf{u})\\
&\times\mathbf{g}(\mathbf{u}+\mathbf{z}h_n)K(\mathbf{s})K(\mathbf{s}-\mathbf{z})K(\mathbf{t})K(\mathbf{t}-\mathbf{z})\mathbf{p}(\mathbf{u}+\mathbf{s}h_n)\\
&\times\mathbf{p}(\mathbf{u}+\mathbf{t}h_n)d\mathbf{s}d\mathbf{t}d\mathbf{z}d\mathbf{u}.
\end{aligned}$$

Therefore, using assumptions (Q.1), (G.4) and the fact that $\sigma_\varepsilon^2 > 0$, it follows that

$$\begin{aligned}\mathbb{E}[\mathcal{H}_n^2(\zeta_i,\zeta_j)] &= \frac{1}{n^3}\int_{\mathbb{R}^d}(\sigma_\varepsilon^2)^2\mathbf{g}^2(\mathbf{u})\mathbf{p}^2(\mathbf{u})\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\times K\left(\mathbf{t}\right)K\left(\mathbf{t}-\mathbf{z}\right)d\mathbf{t}\right]^2 d\mathbf{z}+o(n^{-3})\\ &= n^{-3}\frac{V}{2}+o(n^{-3}). \end{aligned} \tag{2.37}$$

Proof of (c) : By hölder inequality, whenever $\frac{1}{\delta}+\frac{1}{\gamma}=1$, we have

$$\begin{aligned}\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\mathbb{1}_{\mathcal{H}_n^2(\zeta_i,\zeta_j)>N}\right] &\leq \left[\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}\right]^{1/\delta}\left[\mathbb{E}\mathbb{1}_{\mathcal{H}_n(\zeta_i,\zeta_j)>N}\right]^{1/\gamma}\\ &\leq \left[\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}\right]^{1/\delta}\left[P\Big(\mathcal{H}_n(\zeta_i,\zeta_j)>N\Big)\right]^{1/\gamma}.\end{aligned}$$

Making use of Markov's inequality, it follows then that

$$\begin{aligned}\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\mathbb{1}_{\mathcal{H}_n^2(\zeta_i,\zeta_j)>N}\right] &\leq \left[\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}\right]^{1/\delta}\left[\frac{\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}}{N^{2\delta}}\right]^{1/\gamma}\\ &\leq \frac{\left[\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}\right]^{1/\delta+1/\gamma}}{N^{2\delta/\gamma}}\\ &\leq \frac{\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}}{N^{2\delta/\gamma}}.\end{aligned} \tag{2.38}$$

Moreover, observe that

$$\begin{aligned}&\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta}\\ =\;& \frac{1}{n^{3\delta}h_n^{3\delta d}}\mathbb{E}\left|\int_{\mathbb{R}^d}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\varepsilon_i\varepsilon_j\mathbf{p}(\mathbf{x})d\mathbf{x}\right|^{2\delta}\\ \leq\;& \frac{1}{n^{3\delta}h_n^{3\delta d}}(\mathbb{E}|\varepsilon_1|^{2\delta})^2\mathbb{E}\left[\int_{\mathbb{R}^d}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\right]^{2\delta}\\ \leq\;& \frac{1}{n^{3\delta}h_n^{3\delta d}}(\mathbb{E}|\varepsilon_1|^{2\delta})^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}K\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\right]^{2\delta}\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})d\mathbf{u}d\mathbf{v}\\ \leq\;& \frac{1}{n^{3\delta}h_n^{3\delta d}}(\mathbb{E}[\varepsilon_1^{2\delta}])^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}K^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mathbf{p}(\mathbf{x})d\mathbf{x}\right]^{\delta}\\ &\times\left[\int_{\mathbb{R}^d}K^2\left(\frac{\mathbf{y}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{y})d\mathbf{y}\right]^{\delta}\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})d\mathbf{u}d\mathbf{v}.\end{aligned}$$

Then, using the following changes of variables $\mathbf{s}=\dfrac{\mathbf{x}-\mathbf{u}}{h_n}$, $\mathbf{t}=\dfrac{\mathbf{y}-\mathbf{v}}{h_n}$, we obtain

$$\begin{aligned}\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta} &\leq \frac{1}{n^{3\delta}h_n^{\delta d}}(\mathbb{E}[\varepsilon_1^{2\delta}])^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})\left[\int_{\mathbb{R}^d}K^2\left(\mathbf{s}\right)\mathbf{p}(\mathbf{s}h_n+\mathbf{u})d\mathbf{s}\right]^{\delta}\\ &\quad\times\left[\int_{\mathbb{R}^d}K^2\left(\mathbf{t}\right)\mathbf{p}(\mathbf{t}h_n+\mathbf{v})d\mathbf{t}\right]^{\delta}d\mathbf{u}d\mathbf{v}.\end{aligned}$$

Also, by the continuity of the function $\mathbf{p}$, we readily obtain

$$\begin{aligned}\mathbb{E}|\mathcal{H}_n(\zeta_i,\zeta_j)|^{2\delta} &= \frac{1}{n^{3\delta}h_n^{\delta d}}(\mathbb{E}[\varepsilon_1^{2\delta}])^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{g}(\mathbf{u})\mathbf{g}(\mathbf{v})\left[\int_{\mathbb{R}^d}K^2(\mathbf{s})\,\mathbf{p}(\mathbf{u})d\mathbf{s}\right]^{\delta}\\ &\quad\times\left[\int_{\mathbb{R}^d}K^2(\mathbf{t})\,\mathbf{p}(\mathbf{v})d\mathbf{t}\right]^{\delta}d\mathbf{u}d\mathbf{v}+o\left(n^{-3\delta}h_n^{-\delta d}\right)\\ &= \mathcal{O}\left(n^{-3\delta}h_n^{-\delta d}\right). \qquad (2.39)\end{aligned}$$

Since $\delta>1$, one may choose $\delta=1+\eta$, for any $\eta>d/(3(2k+1)-d)$. Combining then the statements (2.38) and (2.39), we obtain

$$\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\mathbb{1}_{\mathcal{H}_n^2(\zeta_i,\zeta_j)>N}\right] = \mathcal{O}\left(n^{-3(\eta+1)}h_n^{-(\eta+1)d}\right).$$

Thus,

$$\begin{aligned}n^3\mathbb{E}\left[\mathcal{H}_n^2(\zeta_i,\zeta_j)\mathbb{1}_{\mathcal{H}_n^2(\zeta_i,\zeta_j)>N}\right] &= \mathcal{O}\left(n^{-3\eta}h_n^{-(\eta+1)d}\right)\\ &= o(1).\end{aligned}$$

Finally, the asymptotic normality of the statistic

$$\sqrt{nh_n^d}R_{n,2}^1=2U_n=2\sum_{1\le i<j\le n}\mathcal{H}_n(\zeta_i,\zeta_j).$$

follows by combining the statements (**a**), (**b**) and (**c**). **Proof of (2.35)** : To complete the proof of our result, we have to show that

$$R_{n,\tau}^1=o_p(n^{-1/2}h^{-d/2}) \text{ for any } \tau\in\{3,\cdots,9\}.$$

– **Case** $\tau=3$: Observe that

$$\begin{aligned}|R_{n,3}^1| &\le \max_{1\le i\le n}\|\mathbf{Z}_i^\top\|^2\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|^2\int_{\mathbb{R}^d}\left[\sum_{i=1}^n\frac{1}{nh_n^d}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right]^2\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}\\ &\le \max_{1\le i\le n}\|\mathbf{Z}_i^\top\|^2\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|^2\int_{\mathbb{R}^d}\mathbf{g}_n^2(\mathbf{x})\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}\\ &\le \max_{1\le i\le n}\|\mathbf{Z}_i^\top\|^2\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|^2\left[\int_{\mathbb{R}^d}\left[\mathbf{g}_n^2(\mathbf{x})-\mathbf{g}^2(\mathbf{x})\right]\frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}+\int_{\mathbb{R}^d}\mathbf{q}(\mathbf{x})d\mathbf{x}\right].\end{aligned}$$

Therefore, under assumptions (H.1) and (M.1) in combination with the decomposition (2.32), the statements (2.15) and the result of Theorem 2.2, it follows that

$$R_{n,3}^1 = o(n^{-1/2}h_n^{-d/2}),\ a.s.$$

– **Case** $\tau = 4$: By the assumption (M.1), there exists a positive constant C such that

$$\begin{aligned} |R_{n,4}^1| &\leq \frac{1}{n^2 h_n^{2d}} \left[\max_{1\leq i\leq n} \|\mathbf{Z}_i^\top\| \|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \sup_{\mathbf{u}\in I^d} |m_{add}(\mathbf{u}) - \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{u})| \right. \\ &\quad \left. \times \int_{\mathbb{R}^d} \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right) \frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})} d\mathbf{x} \right] \\ &\leq C\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \sup_{\mathbf{u}\in I^d} |m_{add}(\mathbf{u}) - \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{u})| \int_{\mathbb{R}^d} \mathbf{g}_n^2(\mathbf{x}) \frac{\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})} d\mathbf{x}. \end{aligned}$$

In view of Theorem 4.2, we have

$$\sup_{\mathbf{u}\in I^d} |m_{add}(\mathbf{u}) - \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{u})| = \mathcal{O}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n}}\right), \ a.s. \tag{2.40}$$

Combining the statements (2.15) and (2.40) together with Theorem 2.2, it follows that

$$\sqrt{nh_n^d} R_{n,4}^1 = \mathcal{O}\left(\sqrt{\frac{h_n^d \log h_n^{-1} \log\log n}{nh_n}}\right), \quad a.s.$$

– **Case** $\tau = 5$: Notice that, for any $i \geq 1$, X_i is independent of ε, then

$$\begin{aligned} &\mathbb{E}|R_{n,5}^1| \\ \leq\ & \mathbb{E}\left[\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \max_{1\leq i\leq n} \|\mathbf{Z}_i^\top\| \frac{1}{n^2 h_n^{2d}} \int_{\mathbb{R}^d} \sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) |\varepsilon_i| \, \mathbf{p}(\mathbf{x}) d\mathbf{x} \right] \\ \leq\ & ess\sup \|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \max_{1\leq i\leq n} \|\mathbf{Z}_i^\top\| \frac{1}{nh_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E}|\varepsilon_1| K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right) \mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v}) d\mathbf{x} d\mathbf{v}. \end{aligned}$$

Since $\mathbb{E}[\varepsilon_1^2] < \infty$, it follows by the Cauchy-Schwarz inequality that $\mathbb{E}|\varepsilon_1| < \infty$. Moreover, since the kernel K is bounded and because of the assumption (M.1), then there exists a positive constant C such that

$$\mathbb{E}|R_{n,5}^1| \leq Cess\sup \|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \frac{1}{nh_n^{2d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right) \mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v}) d\mathbf{x} d\mathbf{v}.$$

Considering now the change of variable $\mathbf{r} = \dfrac{\mathbf{x}-\mathbf{v}}{h_n}$, we obtain

$$\mathbb{E}|R_{n,5}^1| \leq Cess\sup \|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(\mathbf{r}) \, \mathbf{p}(\mathbf{r}h_n + \mathbf{v})\mathbf{g}(\mathbf{v}) d\mathbf{r} d\mathbf{v}.$$

Making use of Theorem 2.2, it is easily seen that

$$\sqrt{nh_n^d}\mathbb{E}R_{n,5}^1 = \mathcal{O}\left(\sqrt{\frac{\log\log n}{n^2h_n^d}}\right). \tag{2.41}$$

Moreover, we have

$$\begin{aligned}
&\mathbb{E}[(R_{n,5}^1)^2] \\
&= \frac{1}{n^4h_n^{4d}}\mathbb{E}\left[\left(\int_{\mathbb{R}^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\frac{\varepsilon_i\mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})}d\mathbf{x}\right)^2\right] \\
&\leq ess\sup\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|^2\max_{1\leq i\leq n}\|\mathbf{Z}_i^\top\|^2\left[\frac{\mathbb{E}[\varepsilon_1^2]}{n^3h_n^{4d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\times K^4\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{v})\ \mathbf{p}^2(\mathbf{x})d\mathbf{v}d\mathbf{x}\right. \\
&\quad\left.+\frac{(n-1)(\mathbb{E}|\varepsilon_1|)^2}{n^3h_n^{4d}}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v})d\mathbf{v}d\mathbf{x}\right)^2\right].
\end{aligned}$$

Integrating by substitution with $\mathbf{u}=\frac{\mathbf{x}-\mathbf{v}}{h_n}$, we obtain

$$\begin{aligned}
&\mathbb{E}[(R_{n,5}^1)^2] \\
&\leq ess\sup\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|^2\max_{1\leq i\leq n}\|\mathbf{Z}_i^\top\|^2\left[\frac{\mathbb{E}[\varepsilon_1^2]}{n^3h_n^{3d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K^4(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}^2(h_n\mathbf{u}+\mathbf{v})d\mathbf{v}d\mathbf{u}\right. \\
&\quad\left.+\frac{(n-1)(\mathbb{E}|\varepsilon_1|)^2}{n^3h_n^{2d}}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K^2(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}(h_n\mathbf{u}+\mathbf{v})d\mathbf{v}d\mathbf{u}\right)^2\right].
\end{aligned}$$

Using the assumption (K.1) together with Theorem 2.2 and the statements (2.41), it follows that

$$nh_n^dVar[R_{n,5}^1] = \mathcal{O}\left(\frac{\log h_n^{-1}}{n^2h^d}\right).$$

We complete the proof by making use of Tchebychev's inequality.

– **Case** $\tau=6$: We have

$$\begin{aligned}
R_{n,6}^1 &= \frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\mathbf{Z}_i^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\varepsilon_j\mathbf{p}(\mathbf{x})d\mathbf{x} \\
&\leq \|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|\max_{1\leq i\leq n}\|\mathbf{Z}_i^\top\|\frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\varepsilon_j\mathbf{p}(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Since $\mathbb{E}[\varepsilon^2]<\infty$, in view of the classical law of iterated logarithm, we have

$$\sum_{j=1}^n\varepsilon_j=\mathcal{O}\left(\sqrt{n\log\log n}\right),\quad a.s. \tag{2.42}$$

Furthermore, combining Theorem 2.2 with the statement (2.42), together with the assumption (M.1) and the fact that the kernel K is bounded, it follows that there exists a positive constant C such that

$$\begin{aligned} |R^1_{n,6}| &\leq C \times \frac{1}{nh_n^d} \times \sqrt{\frac{\log\log n}{n}} \times \sqrt{n \log\log n} \int_{\mathbb{R}^d} \mathbf{g}_n(x)\mathbf{p}(\mathbf{x})d\mathbf{x}, \quad a.s. \\ &\leq C \times \frac{\log\log n}{nh_n^d} \int_{\mathbb{R}^d} \mathbf{g}_n(x)\mathbf{p}(\mathbf{x})d\mathbf{x}, \quad a.s. \end{aligned}$$

Therefore, through the statement (2.15), we obtain

$$\sqrt{nh_n^d} R^1_{n,6} = \mathcal{O}\left(\frac{\log\log n}{\sqrt{nh_n^d}}\right), \quad a.s.$$

– **Case** $\tau = 7$: Observe, that

$$\begin{aligned} \mathbb{E}|R^1_{n,7}| &\leq \mathbb{E}\left[\sup_{\mathbf{u}\in I^d}(m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u}))^2 \int_{\mathbb{R}^d} \left[\frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right]^2 \mathbf{p}(\mathbf{x})d\mathbf{x}\right] \\ &\leq ess\sup(m_{add} - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add})^2 \left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{nh_n^{2d}} K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{v})\mathbf{p}(\mathbf{x})d\mathbf{v}d\mathbf{x}\right. \\ &\quad \left. + \frac{n-1}{nh_n^{2d}} \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{v})d\mathbf{v}\right)^2 \mathbf{p}(\mathbf{x})d\mathbf{x}\right]. \end{aligned}$$

Taking $\mathbf{u} = \dfrac{\mathbf{x}-\mathbf{v}}{h_n}$ and considering assumptions upon the functions $\mathbf{g}$ and $\mathbf{p}$, we obtain

$$\begin{aligned} \mathbb{E}|R^1_{n,7}| &\leq ess\sup(m_{add} - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add})^2 \left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{1}{nh_n^d} K^2(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}(h_n\mathbf{u}+\mathbf{v})d\mathbf{u}d\mathbf{v}\right. \\ &\quad \left. + \frac{n-1}{nh_n^d} \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} K(\mathbf{u})\,\mathbf{g}(\mathbf{v})d\mathbf{v}\right)^2 \mathbf{p}(h_n\mathbf{u}+\mathbf{v})d\mathbf{u}\right]. \end{aligned}$$

In view of the statement (2.40), it follows that

$$\sqrt{nh_n^d}\,\mathbb{E}|R^1_{n,7}| = \mathcal{O}\left(\frac{\log h_n^{-1}}{\sqrt{nh_n^{d+2}}}\right). \tag{2.43}$$

Moreover, we have

$$\begin{aligned}
&\mathbb{E}[(R^1_{n,7})^2] \\
= \;& \mathbb{E}\left[\frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\left(\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(m_{add}(\mathbf{X}_i)-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{X}_i)\right)\right)^2\mathbf{p}(\mathbf{x})d\mathbf{x}\right]^2 \\
\leq \;& ess\sup(m_{add}-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add})^4\left[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{1}{n^3h_n^{3d}}K^4(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}^2(h_n\mathbf{u}+\mathbf{v})d\mathbf{v}d\mathbf{u}\right. \\
& \left.+\frac{n^3-1}{n^3h_n^{2d}}\Big(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K^2(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}(h_n\mathbf{u}+\mathbf{v})d\mathbf{v}d\mathbf{u}\Big)^2\right].
\end{aligned}$$

Making use of the statements (2.40) and (2.43), the assumption (K.1) and the continuity of the function $\mathbf{p}$, it follows that

$$nh_n^dVar[R^1_{n,7}] \;=\; \mathcal{O}\left(\frac{\log^2 h_n^{-1}}{nh^{d+2}}\right).$$

Thus, by the Tchebychev's inequality, we achieve the proof.

– **Case** $\tau = 8$: Since $\mathbb{E}|\varepsilon_1| < \infty$ and the kernel K is bounded, then there exists a finite positive constant C such that

$$\begin{aligned}
&\mathbb{E}|R^1_{n,8}| \\
\leq \;& \mathbb{E}\left[\max_{1\leq i\leq n}\left|m_{add}(\mathbf{X}_i)-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(X_i)\right|\frac{1}{n^2h_n^{2d}}\int_{\mathbb{R}^d}\sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\varepsilon_i|\mathbf{p}(\mathbf{x})d\mathbf{x}\right] \\
\leq \;& ess\sup\left|m_{add}(\mathbf{u})-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u})\right|\frac{1}{nh_n^{2d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{E}|\varepsilon_1|K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v})d\mathbf{x}d\mathbf{v} \\
\leq \;& C\;ess\sup\left|m_{add}(\mathbf{u})-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u})\right|\frac{1}{nh_n^{2d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v})d\mathbf{x}d\mathbf{v}.
\end{aligned}$$

Considering now the change of variable $\mathbf{u} = \dfrac{\mathbf{x}-\mathbf{v}}{h_n}$, we obtain

$$\mathbb{E}|R^1_{n,8}| \;\leq\; C\;ess\sup\left|m_{add}-\widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}\right|\frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}K(\mathbf{u})\,\mathbf{p}(\mathbf{u}h_n+\mathbf{v})\mathbf{g}(\mathbf{v})d\mathbf{u}d\mathbf{v}.$$

Combining then the statement (2.40), the assumption, (K.1) and the continuity of the function $\mathbf{p}$, it follows that

$$\sqrt{nh_n^d}\mathbb{E}|R^1_{n,8}| = \mathcal{O}\left(\sqrt{\frac{\log h_n^{-1}}{n^2h_n^{d+1}}}\right). \tag{2.44}$$

Moreover, since, for any $i \geq 1$, X_i is independent of ε, we have

$$\begin{aligned}
&\mathbb{E}[(R^1_{n,8})^2] \\
&= \frac{1}{n^4 h_n^{4d}} \mathbb{E}\left[\int_{\mathbb{R}^d} \sum_{i=1}^n K^2\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(m_{add}(\mathbf{X}_i) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{X}_i)\right)\frac{\varepsilon_i \mathbf{q}(\mathbf{x})}{\mathbf{g}^2(\mathbf{x})} d\mathbf{x}\right]^2 \\
&\leq ess\sup\left|m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u})\right|^2 \left[\frac{\mathbb{E}[\varepsilon_1^2]}{n^3 h_n^{4d}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \times K^4\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{g}(\mathbf{v})\,\mathbf{p}^2(\mathbf{x}) d\mathbf{v} d\mathbf{x}\right. \\
&\quad \left. + \frac{(n-1)(\mathbb{E}|\varepsilon_1|)^2}{n^3 h_n^{4d}}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)\mathbf{p}(\mathbf{x})\mathbf{g}(\mathbf{v}) d\mathbf{v} d\mathbf{x}\right)^2\right].
\end{aligned}$$

Integrating by substitution with $\mathbf{u} = \frac{\mathbf{x}-\mathbf{v}}{h_n}$, we obtain

$$\begin{aligned}
&\mathbb{E}[(R^1_{n,8})^2] \\
&\leq ess\sup\left|m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u})\right|^2 \left[\frac{\mathbb{E}[\varepsilon_1^2]}{n^3 h_n^{3d}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K^4(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}^2(h_n\mathbf{u}+\mathbf{v}) d\mathbf{v} d\mathbf{u}\right. \\
&\quad \left. + \frac{(n-1)(\mathbb{E}|\varepsilon_1|)^2}{n^3 h_n^{2d}}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K^2(\mathbf{u})\,\mathbf{g}(\mathbf{v})\mathbf{p}(h_n\mathbf{u}+\mathbf{v}) d\mathbf{v} d\mathbf{u}\right)^2\right].
\end{aligned}$$

Using the assumption (K.1) together with the statements (2.40) and (2.44), it follows that

$$nh_n^d Var[R^1_{n,8}] = \mathcal{O}\left(\frac{\log h_n^{-1}}{n^2 h^{d+1}}\right).$$

We complete the proof by making use of Tchebychev's inequality.

– **Case** $\tau = 9$: Next, arguing similarly as above and reminding that

$$\begin{aligned}
R^1_{n,9} &= \frac{1}{n^2 h_n^{2d}} \int_{\mathbb{R}^d} \sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right) \\
&\quad \left(m_{add}(\mathbf{X}_i) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{X}_i)\right)\varepsilon_j \mathbf{p}(\mathbf{x}) d\mathbf{x},
\end{aligned}$$

we have

$$\begin{aligned}
|R^1_{n,9}| &\leq \sup_{\mathbf{u}\in I^d}\left|m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u})\right| \frac{1}{n^2 h_n^{2d}} \int_{\mathbb{R}^d} \sum_{i\neq j}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \\
&\quad K\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\varepsilon_j \mathbf{p}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Though, by the statement (2.42) and the fact that K is bounded, it follows that there exists a finite positive constant C such that

$$\begin{aligned} & |R^1_{n,9}| \\ \leq\ & C \sup_{\mathbf{u}\in I^d} \left| m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u}) \right| \frac{\sqrt{n \log\log n}}{n^2 h_n^{2d}} \int_{\mathbb{R}^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \mathbf{p}(\mathbf{x}) d\mathbf{x}, \quad a.s. \\ \leq\ & C \sup_{\mathbf{u}\in I^d} \left| m_{add}(\mathbf{u}) - \widehat{m}^{\widehat{\boldsymbol{\beta}}}_{add}(\mathbf{u}) \right| \frac{\sqrt{n \log\log n}}{n h_n^{d}} \int_{\mathbb{R}^d} \mathbf{g}_n(\mathbf{x}) \mathbf{p}(\mathbf{x}) d\mathbf{x}, \quad a.s. \end{aligned}$$

Making use of the assumption (H.1) combined with the statements (2.15) and (2.40) we readily obtain

$$\sqrt{nh_n^d} R^1_{n,9} \ = \ \mathcal{O}\left(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+1}}}\right) \times \mathcal{O}\left(\sqrt{\log\log n}\right), \quad a.s.$$

This achieves the proof of the statement (2.35).

Finally, by combining the statements (2.33), (2.34) and (2.35), we readily complete the proof of Theorem 2.3. □

Proof of the corollary 2.1 : Observe that

$$\frac{nh_n^d R_n - \widehat{D}_n}{\sqrt{nh_n^d \widehat{V}_n}} = \frac{\sqrt{V}}{\sqrt{\widehat{V}_n}} \left[\frac{nh_n^d R_n - D}{\sqrt{nh_n^d V}} + \frac{D - \widehat{D}_n}{\sqrt{nh_n^d V}} \right]. \tag{2.45}$$

In view of Theorem 2.1, we have

$$\frac{nh_n^d R_n - D}{\sqrt{nh_n^d V}} \xrightarrow{d} \mathcal{N}(0,1), \tag{2.46}$$

therefore, to prove Corollary 2.2, it suffices to check the following statements

(**d**) $\widehat{V}_n - V = o(1)$

(**e**) $D - \widehat{D}_n = o(h^{-d/2})$

Proof of (**d**) : Observe that

$$\begin{aligned}
&\widehat{V}_n - V \\
:= \;& 2\int (\widehat{\sigma}_n^2)^2 \mathbf{g}_n^{-2}(\mathbf{u})\mathbf{q}^2(\mathbf{u})d\mathbf{u} \times \int \left[\int K^2(\mathbf{t})K(\mathbf{t}-\mathbf{r})d\mathbf{t}\right]^2 d\mathbf{r} \\
& -2\int (\sigma_\varepsilon^2)^2 \mathbf{g}^{-2}(\mathbf{u})\mathbf{q}^2(\mathbf{u})d\mathbf{u} \times \int \left[\int K^2(\mathbf{t})K(\mathbf{t}-\mathbf{r})d\mathbf{t}\right]^2 d\mathbf{r} \\
= \;& 2\int \left[(\widehat{\sigma}_n^2)^2 \mathbf{g}_n^{-2}(\mathbf{u}) - (\sigma_\varepsilon^2)^2 \mathbf{g}^{-2}(\mathbf{u})\right] \mathbf{q}^2(\mathbf{u})d\mathbf{u} \times \int \left[\int K^2(\mathbf{t})K(\mathbf{t}-\mathbf{r})d\mathbf{t}\right]^2 d\mathbf{r}.
\end{aligned} \tag{2.47}$$

Note that

$$\begin{aligned}
(\widehat{\sigma}_n^2)^2 \mathbf{g}_n^{-2}(\mathbf{u}) - (\sigma_\varepsilon^2)^2 \mathbf{g}^{-2}(\mathbf{u}) &= \frac{(\widehat{\sigma}_n^2)^2}{(\sigma_\varepsilon^2)^2}.(\sigma_\varepsilon^2)^2 \left[\frac{1}{\mathbf{g}_n^2(\mathbf{u})} - \frac{1}{\mathbf{g}^2(\mathbf{u})}\right] + \left[(\widehat{\sigma}_n^2)^2 - (\sigma_\varepsilon^2)^2\right] \frac{1}{\mathbf{g}^2(\mathbf{u})} \\
&= \frac{(\widehat{\sigma}_n^2)^2}{(\sigma_\varepsilon^2)^2}.(\sigma_\varepsilon^2)^2 \left[\frac{\mathbf{g}^2(\mathbf{u}) - \mathbf{g}_n^2(\mathbf{u})}{\mathbf{g}_n^2(\mathbf{u})\mathbf{g}^2(\mathbf{u})}\right] + \left[(\widehat{\sigma}_n^2)^2 - (\sigma_\varepsilon^2)^2\right] \frac{1}{\mathbf{g}^2(\mathbf{u})}.
\end{aligned}$$

Also

$$\frac{\mathbf{g}^2(\mathbf{u}) - \mathbf{g}_n^2(\mathbf{u})}{\mathbf{g}_n^2(\mathbf{u})\mathbf{g}^2(\mathbf{u})} = \frac{-\Big(\mathbf{g}(\mathbf{x}) - \mathbf{g}_n(\mathbf{x})\Big)^2 + 2\mathbf{g}(\mathbf{x})\Big(\mathbf{g}(\mathbf{x}) - \mathbf{g}_n(\mathbf{x})\Big)}{\mathbf{g}_n^2(\mathbf{u})\mathbf{g}^2(\mathbf{u})}.$$

Making use of the statement (2.15) and the assumption (G.2), we obtain

$$\frac{\mathbf{g}^2(\mathbf{u}) - \mathbf{g}_n^2(\mathbf{u})}{\mathbf{g}_n^2(\mathbf{u})\mathbf{g}^2(\mathbf{u})} \to 0, \; a.s. \tag{2.48}$$

Since from the model (2.2) we have

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i) + \varepsilon_i + (m_{add}(\mathbf{X}_i) - \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i)),$$

it is clear that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 &= \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2 + \frac{1}{n}\sum_{i=1}^n (m_{add}(\mathbf{X}_i) - \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i))^2 \\
&\quad + \frac{2}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})(m_{add}(\mathbf{X}_i) - \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i)).
\end{aligned}$$

Observe by the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n |\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta}||m_{add}(\mathbf{X}_i) - \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i)| \\
\leq \;& \left[\frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2\right]^{\frac{1}{2}} \left[\frac{1}{n}\sum_{i=1}^n (m_{add}(\mathbf{X}_i) - \widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{X}_i))^2\right]^{\frac{1}{2}}.
\end{aligned}$$

Therefore, using the statement (2.18), it is easily seen that

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \sigma_\varepsilon^2, a.s. \tag{2.49}$$

Moreover, it is easy to see, again by Cauchy-Schwarz inequality, that

$$\begin{aligned}
|\widehat{\sigma}_n^2 - \sigma_\varepsilon^2| &\leq \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\mathbf{Z}}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)^2 + \frac{2}{n}\sum_{i=1}^{n}|\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta}||\tilde{\mathbf{Z}}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2 - \sigma_\varepsilon^2\right| \\
&\leq \max_{1\leq i\leq n}\|\tilde{\mathbf{Z}}_i\|^2\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 + 2\max_{1\leq i\leq n}\|\tilde{\mathbf{Z}}_i\|\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2\right]^{\frac{1}{2}} \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2 - \sigma_\varepsilon^2\right|.
\end{aligned}$$

Note that since the random vector $\mathbf{Z}$ is bounded, making use of Lemma 2.2 (ii), it follows that $\tilde{\mathbf{Z}}_i$ is also bounded for any $1 \leq i \leq n$. Therefore, considering the statement (2.49) together with Theorem 2.2, it is clear that

$$\widehat{\sigma}_n^2 - \sigma_\varepsilon^2 = o(1) \quad a.s. \tag{2.50}$$

Combining the statements (2.48) and (2.50), we obtain

$$(\widehat{\sigma}_n^2)^2\mathbf{g}_n^{-2}(\mathbf{u}) - (\sigma_\varepsilon^2)^2\mathbf{g}^{-2}(\mathbf{u}) = o(1).$$

Therefore, by the statement (2.47), we achieve the proof of (**d**).

Proof of (**e**) : The proof is similar to that of (**d**), it suffices to replace $\mathbf{g}_n^{-2}$ in the expression of D_n by $\mathbf{g}_n^{-1}$.

The proof of the corollary is completed while combining the statements (2.45), (2.46), (**d**), (**e**) and making use of Slutzky's Theorem. □

Chapitre 3

Estimation and tests in the partially linear additive regression model

Ce chapitre est composé d'un article publié en (2014) dans la revue Statistical Methodology, mis en forme pour être inséré dans le présent manuscrit de thèse.

In present chapter, we focus on the partially linear additive model defined by

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{\ell=1}^{d} m_\ell(X_{i\ell}) + \varepsilon_i,\ 1 \leq i \leq n,$$

where $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{ip})^\top$ and $\mathbf{X}_i = (X_{i,1}, \ldots, X_{id})^\top$ are vectors of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$is a vector of unknown parameters, $m_1, \ldots, m_d$ are unknown univariate real functions, and ε_i are i.i.d. random errors with finite variances σ_ε^2. More precisely, we first consider the problem of testing the null hypothesis $\mathcal{H}_0^{\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. The second aim of this paper is to propose a test for the null hypothesis $\mathcal{H}_0^{\sigma} : \sigma_\varepsilon^2 = \sigma_0^2$, in the partially linear additive regression models. Under the null hypotheses, the proposed test statistics are shown to have standard chi-squared distributions asymptotically.

3.1 Introduction

Regression analysis has proved to be a flexible tool and provided a powerful statistical modeling framework in a variety of applied and theoretical contexts. Parametric

regression models provide useful tools for analyzing practical data when the models are correctly specified, but may suffer from large modeling biases when structures of the models are misspecified which is the case in many practical problems. In the latter case, it is pertinent to proceed with nonparametric regression modeling of the data. However, nonparametric models are hampered by the so-called curse of dimensionality in multivariate settings, see Stone (1985, 1986), Fan Fan and Gijbels (1996) and Härdle (1990) among others. One of the methods for attenuating this difficulty is to model covariate effects via a partially linear structure, first introduced by Engle et al. (1986), are special semiparametric models. Since the partially linear models contain both parametric and nonparametric components, they are more flexible than the standard linear models when it is believed that the response depends on some variables in linear relationship but is nonlinearly related to some other particular independent variables, for several examples concerning with practical problems involving partial linear models Härdle et al. (2000). To be more precise, the partially linear regression models are defined as follows

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m(\mathbf{X}) + \varepsilon, \tag{3.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown parameters, m is the nonlinear part of the model and ε is the modelling error. Here and in the sequel, $\mathbf{Z}^\top$ stands as the transpose of the vector $\mathbf{Z}$. The partially linear regression model has a broad applicability in the fields of biology, economics, education and social sciences. This model and various associated estimators, test statistics, and generalizations have generated a substantial body of literature, which includes the works of of Rice (1986), Chen (1988), Robinson (1988), Chen and Shiau (1991), Eubank and Speckman (1990) Donald and Newey (1994), Shi and Li (1995a,b), Bhattacharya and Zhao (1997), Hamilton and Truong (1997), Liang et al. (2008), Shen et al. (2011), Mammen et al. (2011) and the reference therein. To reduce the dimension impact of the nonparametric part in the partially linear regression model, we consider the additive structure of the regression function m and introduce the following model

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{\ell=1}^{d} m_\ell(\mathrm{X}_{i\ell}) + \varepsilon_i, \tag{3.2}$$

where $X_{i\ell}$ is the ℓ-th component of the vector $\mathbf{X}_i$ and m_ℓ is a real univariate function and ε_i are i.i.d. random errors with $\mathbb{E}[\varepsilon_i|\mathbf{X}_i, \mathbf{Z}_i] = 0$ and $\mathbb{E}[\varepsilon_i^2|\mathbf{X}_i, \mathbf{Z}_i] = \sigma_\varepsilon^2$ almost surely. In many case we are interested to finding out the impact of the covariates $\mathbf{Z}$ on the response Y, in this chapter we construct a statistical test, under model (3.2), of the

null hypothesis

$$\mathcal{H}_0^{\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{versus} \quad \mathcal{H}_1^{\boldsymbol{\beta}} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

Also, we will be interested in testing the null hypothesis

$$\mathcal{H}_0^{\sigma} : \sigma_\varepsilon^2 = \sigma_0^2 \quad \text{versus} \quad \mathcal{H}_1^{\sigma} : \sigma_\varepsilon^2 \neq \sigma_0^2.$$

The remainder of the present chapter is organized as follows. In Section 3.2, we briefly describe the estimation procedure which play a central role in the construction the statistical tests. In Section 3.3, we propose statistical tests for checking the null hypotheses $\mathcal{H}_0^{\boldsymbol{\beta}}$ and $\mathcal{H}_0^{\sigma}$. The asymptotic distributions of the test statistics under the null hypotheses are derived in Section 3.4. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to the appendix.

3.2 Estimation procedures

We consider a sequence $\{\mathbf{X}_i, Y_i, \mathbf{Z}_i : i \geq 1\}$ be independent and identically distributed random replicæ of the random vector $(\mathbf{X}, Y, \mathbf{Z}) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^p$. We denote the joint density function of $(\mathbf{X}, Y, \mathbf{Z})$ by f with respect to the Lesbegue measure and denote by

$$g(\mathbf{x}) = \int_{\mathbb{R}^{p+1}} f(\mathbf{x}, y, \mathbf{z}) dy d\mathbf{z},$$

the density of $\mathbf{X}$. For each $n \geq 1$, and for each choice of the bandwidth $h_n > 0$, we define the kernel estimator of the marginal density $\mathbf{g}$, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\mathbf{g}_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right),$$

where K is a kernel function, i.e., a non-negative function defined on $\mathbb{R}^d$ and integrating to 1. Notice that the model (3.1) may be written also as

$$Y - \mathbf{Z}^\top \boldsymbol{\beta} = m(\mathbf{X}) + \varepsilon. \tag{3.3}$$

On the basis of the model (3.3), following the usual Wand & Jones method, the regression estimator involving the nonparametric part of the model may be defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}}{n\mathbf{g}_n(\mathbf{X}_i)} \left(\prod_{\ell=1}^{d} \frac{1}{h_n} K_\ell \left(\frac{x_\ell - X_{i\ell}}{h_n} \right) \right), \tag{3.4}$$

where x_ℓ and $X_{i\ell}$ are the ℓ-th component of $\mathbf{x}$ and $\mathbf{X}_i$ respectively, and K_ℓ $(1 \leq \ell \leq d)$ are kernel functions defined on $\mathbb{R}$. Note that $\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})$ depend on the unknown parameter $\boldsymbol{\beta}$ which needs to be estimated. Considering the model (3.3), the function m clearly depends on the parameter $\boldsymbol{\beta}$ and its additive structure may be written as

$$m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \mu + \sum_{\ell=1}^{d} m_\ell^{\boldsymbol{\beta}}(x_\ell), \tag{3.5}$$

where model identifiability considerations impose that $\mathbb{E}m_\ell^{\boldsymbol{\beta}}(X_\ell) = 0$, $1 \leq \ell \leq d$. Various methods have been proposed in the literature to estimate the additive components of the regression model including the marginal integration method [see, e.g., Newey (1994), Tjøstheim and Auestad (1994) and Linton and Nielsen (1995)] and the back-fitting algorithms [cf. to Hastie and Tibshirani (1986), Sperlich et al. (1999) and the references therein for an account of results on these topics]. The linear part of the semi-parametric models has been investigated in a number of works. Our approach combines the marginal integration method to estimate m_ℓ, $\ell = 1, \ldots, d$, and the least square error criterion to estimate the parameter $\boldsymbol{\beta}$. We need to introduce some further notations. For any $1 \leq \ell \leq d$, set $\mathbf{x}_{-\ell} = (x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_d)$,

$$\mathbf{q}_{-\ell}(\mathbf{x}_{-\ell}) = \prod_{j=1, j\neq\ell}^{d} q_j(x_j)$$

and

$$\mathbf{q}(\mathbf{x}) = \prod_{l=1}^{d} q_\ell(x_\ell),$$

where q_ℓ, $1 \leq \ell \leq d$, are known univariate density functions. Following the marginal integration method, the additive regression function estimator is given, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{\ell=1}^{d} \widehat{\xi}_\ell^{\boldsymbol{\beta}}(x_\ell) + \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}, \tag{3.6}$$

where

$$\widehat{\xi}_\ell^{\boldsymbol{\beta}}(x_\ell) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}_{-\ell}(\mathbf{x}_{-\ell})d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{3.7}$$

Here, $\widehat{\xi}_\ell^{\boldsymbol{\beta}}$ is the estimate of the ℓ-th component of the additive regression function which still depends on the parameter $\boldsymbol{\beta}$. Therefore, one has to estimate the vector parameter

$\boldsymbol{\beta}$ to have ready estimates. Now, we give the estimation procedure estimation of $\boldsymbol{\beta}$. While considering the partially linear additive regression model

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m_{add}(\mathbf{X}) + \varepsilon, \tag{3.8}$$

Making use of the statements (3.4), (3.6)-(3.8), and considering the least square error criterion, it follows that

$$\widehat{\boldsymbol{\beta}} = [\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top]^{-1}\widetilde{\mathbf{Z}}\tilde{Y}, \tag{3.9}$$

where

$$\begin{array}{rcl} \tilde{Y} & = & \left[Y_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i)Y_j \right]_{1\leq i\leq n}^\top, \\ \widetilde{\mathbf{Z}} & = & \left[\mathbf{Z}_i - \sum_{j=1}^n W_{nj}(\mathbf{X}_i)\mathbf{Z}_j \right]_{1\leq i\leq n}, \end{array} \tag{3.10}$$

$$W_{nj}(\mathbf{X}_i) = \frac{U_{nj}(\mathbf{X}_i)}{n\mathbf{g}_n(\mathbf{X}_j)} \tag{3.11}$$

and

$$U_{nj}(\mathbf{X}_i) = \sum_{\ell=1}^d \frac{1}{h_n} K_\ell\left(\frac{X_{i\ell} - X_{j\ell}}{h_n}\right) D_\ell - (d-1)\int_{\mathbb{R}^d} \prod_{k=1}^d \frac{1}{h_n} K_k\left(\frac{x_k - X_{jk}}{h_n}\right) \mathbf{q}(\mathbf{x})d\mathbf{x},$$

where

$$D_\ell = \int_{\mathbb{R}^{d-1}} \prod_{k=1, k\neq\ell} \frac{1}{h_n} K_k\left(\frac{x_k - X_{jk}}{h_n}\right) \mathbf{q}_{-\ell}(\mathbf{x}_{-\ell})d\mathbf{x}_{-\ell}.$$

Finally, the estimates of the regression function and the additive components are defined by

$$\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) = \sum_{\ell=1}^d \widehat{\xi}_\ell^{\widehat{\boldsymbol{\beta}}}(x_\ell) + \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x} \tag{3.12}$$

and

$$\widehat{\xi}_\ell^{\widehat{\boldsymbol{\beta}}}(x_\ell) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})\mathbf{q}_{-\ell}(\mathbf{x}_{-\ell})d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})\mathbf{q}(\mathbf{x})d\mathbf{x}. \tag{3.13}$$

3.3 Statistical tests

In order to test the null hypothesis $\mathcal{H}_0^{\boldsymbol{\beta}}$, through the Wald-type statistic, we propose the following statistic

$$R_n := \frac{n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top B(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\widehat{\sigma}_n^2}, \tag{3.14}$$

where

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2. \tag{3.15}$$

Here, B is a $p\times p$-positive definite matrix. Under some regularity conditions, we prove, in Theorem 3.3, that the limit law of the test statistic R_n a χ^2-distribution with p degrees of freedom. An application of Theorem 3.3, leads to reject the null hypothesis $\mathcal{H}_0^{\boldsymbol{\beta}}$, whenever the value of the statistic R_n exceeds $z_{1-\alpha}$, namely, the $(1-\alpha)$-quantile of the χ^2 law with p degrees of freedom. The corresponding test is then, asymptotically of level α, when $n\to\infty$. Then the confidence region associated to our test, for a given level significance α, is formulated by the following ellipsoid

$$\mathcal{R} = \{\mathbf{b}\in\mathbb{R}^p : n(\widehat{\sigma}_n^2)^{-1}(\widehat{\boldsymbol{\beta}}-\mathbf{b})^\top B(\widehat{\boldsymbol{\beta}}-\mathbf{b})\le z_\alpha\}.$$

Our second test concern the variance σ_ε^2 of the model (3.2), we test the null hypothesis

$$\mathcal{H}_0^\sigma : \sigma_\varepsilon^2 = \sigma_0^2$$

versus

$$\mathcal{H}_1^\sigma : \sigma_\varepsilon^2 \neq \sigma_0^2.$$

To test the null hypothesis $\mathcal{H}_0^\sigma$, we propose the following statistic

$$T_n := \frac{n(\widehat{\sigma}_n^2-\sigma_\varepsilon^2)^2}{V_n}, \tag{3.16}$$

where

$$V_n := \frac{1}{n}\sum_{i=1}^{n}\left((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \widehat{\sigma}_n^2\right)^2.$$

In Corollary 3.1, we show that T_n follows asymptotically a χ^2-distribution. Therefore, the confidence interval, for a given level significance α, is given by

$$\mathcal{T} = \left\{\sigma\in\mathbb{R} : n(\widehat{\sigma}_n^2-\sigma^2)^2 \le V_n z_\alpha\right\},$$

where z_α is a χ^2 quantile of order α.

3.4 Main results

Remind that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^\top = B \qquad a.s.,$$

where B is a $p \times p$-positive definite matrix. Further assumptions involving the distribution function of $\mathbf{X}$, the regression function m, the kernels K_ℓ, $1 \leq \ell \leq d$ and the smoothing parameters are gathered together hereafter for easy reference. The first part of these conditions is devoted to the regression function m and the density $\mathbf{g}$. In the sequel, I^d denotes a compact subset of $\mathbb{R}^d$. We make use of the following conditions.

(G.1) m is k-times continuously differentiable.

(G.2) The marginal density $\mathbf{g}$ is strictly positive on the support I^d of the function $\mathbf{q}$.

(G.3) The marginal density $\mathbf{g}$ is uniformly continuous on its support.

(G.4) The marginal density $\mathbf{g}$ has $k+1$ continuous derivatives.

Throughout, the following hypothesis is considered upon the sequence of bandwidths $(h_n)_{n\geq 1}$.

(H.1) $h_n = \vartheta_1 \left(\dfrac{\log n}{n}\right)^{1/(2k+1)}$ for $0 < \vartheta_1 < \infty$ and $2k+1 \geq d$.

Set now, for any $\mathbf{x} \in \mathbb{R}^d$, $K(\mathbf{x}) := \prod_{l=1}^{d} K_\ell(x_\ell)$. The kernels are assumed to satisfy the following conditions.

(K.1) For any $1 \leq \ell \leq d$, K_ℓ is bounded, Lipschitz continuous and integrating to one.

(K.2) For any $1 \leq \ell \leq d$, $K_\ell(u) = 0$ for u $\notin [-\lambda/2, \lambda/2]$, for some $0 < \lambda < \infty$.

(K.3) K is a kernel of order k.

Consider also the following assumptions upon the random variables Y and $\mathbf{Z}$.

(M.1) Y and $\mathbf{Z}$ are bounded.

(M.2) $\mathbf{Z}$ is with mean zero.

The assumptions on the weight functions q_ℓ, $1 \leq \ell \leq d$, needed for our analysis are the following.

(Q.1) For any $1 \leq \ell \leq d$, q_ℓ has $k+1$ continuous and bounded derivatives.

(Q.2) The support of the function $\mathbf{q}$ is included in the support of the density $\mathbf{g}$.

3.4.1 Comments on hypotheses.

To establish the consistency of the parameter $\boldsymbol{\beta}$, the most part of hypotheses are needed for this aim, see, for instance, Chokri and Louani (2011) for more details.
In the sequel, "$\xrightarrow{d}$" denotes the convergence in distribution.
The main results to be proved here may now be stated precisely as follows.

Theorem 3.1. Assume that assumptions (G.1-4), (H.1), (K.1-3), (M.1-2) and (Q.1-2) hold. In addition we suppose that $\max_{1\leq i\leq n}\mathbb{E}|\varepsilon_i|^r < \infty$ for some $r \geq 4$. Then, as $n \to \infty$,

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2) \xrightarrow{d} \mathcal{N}(0, Var\varepsilon_1^2),$$

The proof of Theorem 3.1 is postponed to the Appendix.

Theorem 3.2. Under assumptions of Theorem 3.1. In addition suppose that, for some $r \geq 4$, $\max_{1\leq i\leq n}\mathbb{E}|\varepsilon_i|^{r+1} < \infty$. Thus we have, as $n \to \infty$,

$$\limsup_{n\to\infty}\sqrt{\frac{n}{2\log\log n}}\left|\widehat{\sigma}_n^2 - \sigma_\varepsilon^2\right| = (Var\varepsilon_1^2)^{1/2} \quad a.s.$$

The proof of Theorem 3.2 is postponed to the Appendix.

Theorem 3.3. Assume that assumptions (G.1-4), (H.1), (K.1-3), (M.1-2) and (Q.1-2) hold true. In addition, suppose that $\max_{1\leq i\leq n}\mathbb{E}|\varepsilon_i|^r < \infty$ for some $r \geq 2$. Then, we have, under the null hypothesis $\mathcal{H}_0^{\boldsymbol{\beta}}$,

$$R_n \xrightarrow{d} \chi^2_{(p)},$$

where $\mathcal{X}^2(p)$ denotes the Chi-square distribution with p degrees of freedom.

The proof of Theorem 3.3 is postponed to the Appendix.

Theorem 3.4. Assume that assumptions of Theorem 3.1 hold. Then, under the null hypothesis $\mathcal{H}_0^{\sigma}$,

$$S_n := \frac{n(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)^2}{V} \xrightarrow{d} \chi^2,$$

where $V := Var\varepsilon_1^2$ and $\chi^2 := \chi^2(1)$ denotes the χ^2-distribution with one degree of freedom.

The proof of Theorem 3.4 is postponed to the Appendix.
Notice that when the quantity V is unknown, the previous Theorem doesn't permit to perform our test. To overcome this problem, it suffices to replace V by its estimate. Then, the corollary below is a new version of Theorem 3.4 when V is estimated by V_n.

Corollary 3.1. Assume that assumptions of Theorem 3.1 hold. Under the null hypothesis $\mathcal{H}_0^{\sigma}$, we have

$$T_n \xrightarrow{d} \mathcal{X}^2.$$

The proof of Corollary 3.1 is captured in the forthcoming appendix.

Remark 3.1. In the present paper, we are mainly concerned with performing statistical tests. Towards this aim, we have used the marginal integration technique to *profile* the nonparametric part and then minimize a *profiled* squared error criterion to estimate the parameter $\boldsymbol{\beta}$. In the paper by Mammen et al. (2011), the authors were interested in the investigation of the efficiency gains when the nonparametric part of the model has an additive structure. For this end, they make use of a smooth backfitting technique to deal with the additive nonparametric part in order to provide semi-parametric efficient estimators for $\boldsymbol{\beta}$. We mention that the estimated profile likelihood based on the Gaussian error model investigated in Section 3.2 of Mammen et al. (2011), coincides with the least square error estimators.

3.5 Simulations

In this section, series of experiments are conducted in order to examine the performance of the proposed statistical tests, defined in (3.14) and (3.16). More precisely, we have undertaken numerical illustrations regarding the power of these statistical tests in finite sample situations. The computing program codes are implemented in `R`. The following three models were considered in the simulation study

$$\begin{aligned}
\text{Model I}: \quad Y &= \mathbf{Z}^\top \boldsymbol{\beta} + X_1 + X_2, \\
\text{Model II}: \quad Y &= \mathbf{Z}^\top \boldsymbol{\beta} + \sin(\pi X_1) + \sin(\pi X_2), \\
\text{Model III}: \quad Y &= \mathbf{Z}^\top \boldsymbol{\beta} + \exp(X_1) + \exp(X_2),
\end{aligned}$$

where X_1, X_2 and the error are assumed to standard normal random variables. The deterministic vector $\boldsymbol{\beta}$ (respectively σ_ε^2) is chosen to take different values from near to far from the null hypothesis while $\mathbf{Z}$ is taken as a gaussian random vector. In our simulations, samples of sizes were $n = 25$, $n = 50$, $n = 100$, $n = 500$ and $n = 1000$ have been drawn following the scheme that has been described and $m = 1000$ replicates have been considered for each scenario. The first kind error risks were $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$. The obtained results are displayed in the following tables.

TAB. 3.1 – Power estimate of 0.01 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model I.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.11	.101	.098	.091	.084	.086
$(1.1, 2.1)^\top$	.159	.231	.337	.446	.779	.866
$(1.5, 2.5)^\top$	.724	.946	.999	1	1	1
$(2, 3)^\top$	.978	1	1	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

TAB. 3.2 – Power estimate of 0.01 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model II.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.111	.112	.132	.148	.104	.171
$(1.1, 2.1)^\top$	.129	.178	.246	.448	.796	.977
$(1.5, 2.5)^\top$	.41	.745	.977	1	1	1
$(2, 3)^\top$	.791	.983	1	1	1	1
$(3, 4)^\top$	.98	1	1	1	1	1

TAB. 3.3 – Power estimate of 0.01 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model III.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.078	.071	.085	.100	.117	.093
$(1.1, 2.1)^\top$	.091	.100	.099	.123	.128	.169
$(1.5, 2.5)^\top$	.193	.182	.294	.430	.719	.941
$(2, 3)^\top$	.379	.456	.671	.874	.999	1
$(3, 4)^\top$	.746	.886	.985	1	1	1

TAB. 3.4 – Power estimate of 0.05 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model I.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.204	.178	.220	.206	.203	.208
$(1.1, 2.1)^\top$	.267	.327	.381	.454	.801	1
$(1.5, 2.5)^\top$	.700	.917	.993	1	1	1
$(2, 3)^\top$	.966	.999	1	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

TAB. 3.5 – Power estimate of 0.05 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model II.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 1.01)^\top$	.213	.246	.214	.232	.225	.273
$(1.1, 2.1)^\top$	.313	.403	.499	.679	.914	.983
$(1.5, 2.5)^\top$	.842	.986	.999	1	1	1
$(2, 3)^\top$	.993	1	1	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

TAB. 3.6 – Power estimate of 0.05 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model III.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.202	.188	.194	.217	.234	.219
$(1.1, 2.1)^\top$	.194	.196	.205	.250	.273	.334
$(1.5, 2.5)^\top$	.305	.328	.458	.574	.840	.982
$(2, 3)^\top$	.522	.626	.812	.951	1	1
$(3, 4)^\top$	.862	.946	.997	1	1	1

TAB. 3.7 – Power estimate of 0.10 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model I.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.273	.275	.309	.315	.310	.303
$(1.1, 2.1)^\top$	.344	.393	.478	.524	.901	1
$(1.5, 2.5)^\top$	.781	.950	.996	1	1	1
$(2, 3)^\top$	.983	1	1	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

TAB. 3.8 – Power estimate of 0.10 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model II.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.285	.310	.292	.309	.324	.346
$(1.1, 2.1)^\top$	.442	.573	.681	.801	.972	1
$(1.5, 2.5)^\top$	.903	1	1	1	1	1
$(2, 3)^\top$	1	1	1	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

TAB. 3.9 – Power estimate of 0.10 tests for R_n against alternatives for different values of $\boldsymbol{\beta}$ based on 1000 replications : Model III.

$\boldsymbol{\beta}$	n					
	25	50	100	200	500	1000
$(1.01, 2.01)^\top$	.291	.289	.296	.298	.327	.339
$(1.1, 2.1)^\top$	.231	.245	.293	.325	.409	.455
$(1.5, 2.5)^\top$	.430	.466	.532	.624	.956	1
$(2, 3)^\top$	.624	.730	.885	1	1	1
$(3, 4)^\top$	1	1	1	1	1	1

Notice that, as in any other inferential context, the greater the sample size is, the better the power of the tests, studied here, is. Simple inspection of the results reported in the preceding tables allows to deduce that for large values of the sample size n, the empirical powers of the considered tests are all close to 1, in particular for $n = 1000$. We observe that even for moderate sample sizes ($n = 25$, $n = 50$) the power of the test is close to 1 for the value of $\alpha = 0.10$ and $\boldsymbol{\beta} = (3, 4)^\top$. This can be explained naturally by the fact that the value of $\boldsymbol{\beta}$ is rather far from the null hypothesis. Thus the modification of the test R_n may become necessary if the sample size is small. The results reported in the Tables 3.1-3.9 lead to the conclusion that the power of R_n is close to 1 for large sample sizes or in the situation when we are far from the null hypothesis. From Table 3.10, similar conclusions are valid for the test T_n where we have considered only the Model II, which leads to more satisfactory results comparing with the two other models. Finally, we conclude that the illustration of simulation studies display that the proposed tests methods are effective in all three models for R_n and in the Model II for T_n. In order to extract methodological recommendations for the use of the proposed statistics in this work, it will be interesting to conduct extensive Monte Carlo experiments to compare our procedures with other alternatives presented in the literature, but this would go well beyond the scope of the present paper.

TAB. 3.10 – Power estimate of the test T_n against alternatives for different values of σ_ε^2 based on 1000 replications : Model II.

α	σ_ε^2	n					
		25	50	100	200	500	1000
	1.01	.104	.197	.314	.347	.398	.339
	1.1	.121	.188	.257	.316	.582	.659
0.01	1.5	.159	.161	.304	.431	.778	.891
	2	.337	.357	.530	.782	.983	1
	3	.590	.694	.875	.988	1	1
	4	.746	.833	.961	1	1	1
	5	.851	.909	.986	1	1	1
	1.01	.274	.402	.478	.528	.502	.537
	1.1	.267	.287	.469	.476	.519	.675
	1.5	.293	.319	.413	.557	.846	.899
0.05	2	.408	.480	.619	.838	.998	1
	3	.651	.739	.913	.988	1	1
	4	.999	1	1	1	1	1
	5	1	1	1	1	1	1
	1.01	.375	.530	.590	.571	.570	.580
	1.1	.348	.461	.504	.506	.534	.874
0.1	1.5	.364	.449	.542	.601	.789	.968
	2	.477	.518	.675	.974	1	1
	3	.843	.950	.998	1	1	1
	4	1	1	1	1	1	1
	5	1	1	1	1	1	1

Conclusion

Testing the hypotheses $\mathcal{H}_0^{\boldsymbol{\beta}}$ and $\mathcal{H}_0^{\sigma}$ is an important step in practice. Towards this aim, the present paper gives two statistical tests in the framework of the partially linear additive model. The limiting distributions under the null hypotheses of the proposed test statistics are derived, and their properties are examined by Monte Carlo simulations. The simulation studies display that the test methods are effective in particular for large sample sizes.

3.6 Appendix

To unburden our notation a bit and for simplicity, C denotes a generic finite positive constant which may have different values at each appearance throughout the sequel.

Proof of Theorem 3.1

Keeping in mind the equation (3.15), one can see the following decomposition of $\widehat{\sigma}_n^2$

$$\begin{aligned}\widehat{\sigma}_n^2 &= \frac{1}{n}(\tilde{Y} - \widetilde{\mathbf{Z}}^\top(\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top)^{-1}\widetilde{\mathbf{Z}}\tilde{Y})^\top(\tilde{Y} - \widetilde{\mathbf{Z}}^\top(\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top)^{-1}\widetilde{\mathbf{Z}}\tilde{Y}) \\ &= \frac{1}{n}(\tilde{Y} - P_{\widetilde{\mathbf{Z}}}\tilde{Y})^\top(\tilde{Y} - P_{\widetilde{\mathbf{Z}}}\tilde{Y}),\end{aligned}$$

where

$$P_{\widetilde{\mathbf{Z}}} = \widetilde{\mathbf{Z}}^\top(\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top)^{-1}\widetilde{\mathbf{Z}}.$$

Notice that, it's easy to show that $P_{\widetilde{\mathbf{Z}}}$ is an idempotent operator, i.e., $P_{\widetilde{\mathbf{Z}}}P_{\widetilde{\mathbf{Z}}} = P_{\widetilde{\mathbf{Z}}}$, fulfilling $P_{\widetilde{\mathbf{Z}}}^\top = P_{\widetilde{\mathbf{Z}}}$. It follows that

$$\begin{aligned}&\sqrt{n}(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2) \\ &= \frac{1}{\sqrt{n}}(\tilde{Y}^\top\tilde{Y} - \tilde{Y}^\top P_{\widetilde{\mathbf{Z}}}\tilde{Y} - \tilde{Y}^\top P_{\widetilde{\mathbf{Z}}}^\top\tilde{Y} + \tilde{Y}^\top P_{\widetilde{\mathbf{Z}}}^\top P_{\widetilde{\mathbf{Z}}}\tilde{Y}) - \sqrt{n}\sigma_\varepsilon^2 \\ &= \frac{1}{\sqrt{n}}\tilde{Y}^\top(\mathcal{I}_n - P_{\widetilde{\mathbf{Z}}})\tilde{Y} - \sqrt{n}\sigma_\varepsilon^2,\end{aligned}$$

where $\mathcal{I}_n$ denotes the identity matrix of order n. Using the fact that $P_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Z}}^\top = \widetilde{\mathbf{Z}}^\top$ and $\widetilde{\mathbf{Z}}P_{\widetilde{\mathbf{Z}}} = \widetilde{\mathbf{Z}}$, we infer that

$$(\mathcal{I}_n - P_{\widetilde{\mathbf{Z}}})\widetilde{\mathbf{Z}}^\top = \mathbf{0}_{n\times p} \quad \text{and} \quad \widetilde{\mathbf{Z}}(\mathcal{I}_n - P_{\widetilde{\mathbf{Z}}}) = \mathbf{0}_{p\times n},$$

where $\mathbf{0}_{n\times p}$ and $\mathbf{0}_{p\times n}$ are the null $n\times p$, respectively $p\times n$, matrices. This, in turn, implies that

$$\begin{aligned}
&\sqrt{n}(\widehat{\sigma}_n^2-\sigma_\varepsilon^2)\\
&= \frac{1}{\sqrt{n}}\varepsilon^\top\varepsilon-\sqrt{n}\sigma_\varepsilon^2-\frac{1}{\sqrt{n}}\varepsilon^\top P_{\widetilde{\mathbf{Z}}}\varepsilon+\frac{1}{\sqrt{n}}\widehat{M}_\varepsilon^\top(\mathcal{I}_n-P_{\widetilde{\mathbf{Z}}})\widehat{M}_\varepsilon+\frac{2}{\sqrt{n}}\widehat{M}_\varepsilon^\top(\mathcal{I}_n-P_{\widetilde{\mathbf{Z}}})\varepsilon\\
&:= \sqrt{n}\left[(I_1-\sigma_\varepsilon^2)-I_2+I_3+2I_4\right], \qquad (3.17)
\end{aligned}$$

where

$$\widehat{M}_\varepsilon := (\widetilde{m}_{add}(\mathbf{X}_i)-\widehat{\varepsilon}_i)_{1\le i\le n},$$

$$\widetilde{m}_{add}(\mathbf{X}_i)=m_{add}(\mathbf{X}_i)-\sum_{j=1}^n W_{nj}(\mathbf{X}_i)m_{add}(\mathbf{X}_j)$$

and

$$\widehat{\varepsilon}_i := \sum_{k=1}^n W_{nk}(\mathbf{X}_i)\varepsilon_k.$$

Recall I_3 from the statement (3.17). We have the following bound

$$\begin{aligned}
|I_3| &\le n^{-1}\max_{1\le i\le n}(|\widetilde{m}_{add}(\mathbf{X}_i)|^2+|\sum_{k=1}^n W_{nk}(\mathbf{X}_i)\varepsilon_k|^2)||\mathcal{I}_n-P_{\widetilde{\mathbf{Z}}}||_{\mathcal{M}_{n,n}(\mathbb{R})}\\
&\le n^{-1}C\max_{1\le i\le n}(|\widetilde{m}_{add}(\mathbf{X}_i)|^2+|\sum_{k=1}^n W_{nk}(\mathbf{X}_i)\varepsilon_k|^2),
\end{aligned}$$

where $||\mathcal{I}_n-P_{\widetilde{\mathbf{Z}}}||_{\mathcal{M}_{n,n}(\mathbb{R})}\le C$ and $\mathcal{M}_{n,n}(\mathbb{R})$ indicate the set of $n\times n$ real matrices. In view of Lemma 2.2 and Lemma 2.3 given bellow, we have, almost surely,

$$\max_{1\le i\le n}|\widetilde{m}_{add}(\mathbf{X}_i)|=\mathcal{O}\left(n^{\frac{-k}{2k+1}}\log n\right) \quad\text{and}\quad \max_{1\le i\le n}|\sum_{k=1}^n W_{nk}(\mathbf{X}_i)\varepsilon_k|=\mathcal{O}\left(n^{\frac{-k}{2k+1}}\log n\right),$$

which implies that

$$|I_3|=n^{-1}\mathcal{O}\left(n^{\frac{-2k}{2k+1}}\log^2 n\right)=o(n^{-1/2}) \ \ a.s. \qquad (3.18)$$

Now, we claim that, almost surely,

$$I_2=o(n^{-1/2}),$$

where

$$I_2=\frac{1}{n}\varepsilon^\top\widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top)^{-1}\widetilde{\mathbf{Z}}^\top\varepsilon$$

$\widetilde{\mathbf{Z}}$ as defined in (3.10). Notice that, for any $i=1,\dots,n$ and $1\le\ell\le p$, we have

$$\mathbb{E}[\tilde{Z}_{i\ell}\varepsilon_i]=\mathbb{E}[\tilde{Z}_{i\ell}\mathbb{E}(\varepsilon_i|X_{i\ell},Z_{i\ell})]=0 \ \text{ and } \ \mathbb{E}\left|\tilde{Z}_{i\ell}\varepsilon_i\right|^r\le C\max_{1\le i\le n}\mathbb{E}\left|\varepsilon_i\right|^r<\infty,$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(\tilde{Z}_{i\ell}\varepsilon_i)^2 = \sigma_\varepsilon^2 b^{\ell\ell} > 0,$$

where $b^{\ell\ell}$ the $(\ell\ell)$-th element of B. Then, we infer from Lemma 2.4, by taking $V_i = \tilde{Z}_{i\ell}\varepsilon_i$ and $s_n^2 = n\sigma_\varepsilon^2 b^{\ell\ell}$, that

$$\sum_{i=1}^{n} \tilde{Z}_{i\ell}\varepsilon_i = \mathcal{O}\Big(n \log\log n\Big)^{1/2} \quad a.s. \tag{3.19}$$

In addition, we have

$$\lim_{n\to\infty} \frac{1}{n}\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top = B \ a.s.$$

Using the last equation in connection with (3.19), one finds

$$I_2 = o(n^{-1/2}) \ \ a.s. \tag{3.20}$$

Recall I_4 from the statement (3.17). We have the chain of inequalities

$$\begin{aligned}
|I_4| &\leq \frac{C}{n}\left|\sum_{i=1}^{n} \widetilde{m}_{add}(\mathbf{X}_i)\varepsilon_i - \sum_{i=1}^{n}\sum_{k=1}^{n} W_{nk}(\mathbf{X}_i)\varepsilon_k\varepsilon_i\right| \\
&\leq \frac{C}{n}\left|\sum_{i=1}^{n} \widetilde{m}_{add}(\mathbf{X}_i)\varepsilon_i - \sum_{i=1}^{n} W_{ni}(\mathbf{X}_i)\varepsilon_i^2 - \sum_{i=1}^{n}\sum_{k\neq i}^{n} W_{nk}(\mathbf{X}_i)\varepsilon_k\varepsilon_i\right| \\
&\leq C\left(\frac{1}{n}\left|\sum_{i=1}^{n} \widetilde{m}_{add}(\mathbf{X}_i)\varepsilon_i\right| + \frac{1}{n}\left|\sum_{i=1}^{n} W_{ni}(\mathbf{X}_i)\varepsilon_i^2\right| + \frac{1}{n}\left|\sum_{i=1}^{n}\sum_{k\neq i}^{n} W_{nk}(\mathbf{X}_i)\varepsilon_k\varepsilon_i\right|\right).
\end{aligned} \tag{3.21}$$

We evaluate the first term in the right side of the last inequality. One can see that

$$\sum_{i=1}^{n} \widetilde{m}_{add}(\mathbf{X}_i)\varepsilon_i \ \leq \ \max_{1\leq i\leq n} \widetilde{m}_{add}(\mathbf{X}_i)\sum_{i=1}^{n}\varepsilon_i.$$

Making use of the Lemma 2.3, we have, almost surely,

$$\max_{1\leq i\leq n} \widetilde{m}_{add}(\mathbf{X}_i) = \mathcal{O}\Big(n^{(-k/2k+1)}\log n\Big),$$

which, combined with the law of the iterated logarithm, we conclude that

$$\sum_{i=1}^{n} \widetilde{m}_{add}(\mathbf{X}_i)\varepsilon_i = o(n^{1/2}) \ \ a.s. \tag{3.22}$$

Once more, observe that

$$\sum_{i=1}^{n} W_{ni}(\mathbf{X}_i)\varepsilon_i^2 \leq \max_{1\leq i\leq n} W_{ni}(\mathbf{X}_i) \sum_{i=1}^{n} \varepsilon_i^2.$$

An application of Lemma 2.2 (i), gives

$$\max_{1\leq i\leq n} W_{ni}(\mathbf{X}_i) = \mathcal{O}(n^{-2k/2k+1}(\log n)^{-1/2k+1}) \ a.s.$$

This when combined with Lemma 2.4, implies

$$\sum_{i=1}^{n} W_{ni}(\mathbf{X}_i)\varepsilon_i^2 = o(n^{1/2}) \quad a.s. \tag{3.23}$$

By using similar arguments, we obtain

$$\begin{aligned} \sum_{i=1}^{n}\sum_{k\neq i}^{n} W_{nk}(\mathbf{X}_i)\varepsilon_k\varepsilon_i &\leq \mathcal{O}\Big(n^{(-k/2k+1)}\log n\Big)\sum_{i=1}^{n}\varepsilon_i \\ &= o(n^{1/2}) \quad a.s. \end{aligned} \tag{3.24}$$

Combining the statements (3.22), (3.23), (3.24), we conclude that

$$I_4 = o(n^{-1/2}) \quad a.s. \tag{3.25}$$

The central limit theorem gives

$$\sqrt{n}(I_1 - \sigma_\varepsilon^2) \xrightarrow{d} \mathcal{N}(0, Var\varepsilon_1^2) \quad a.s. \tag{3.26}$$

The proof of Theorem 3.1 is completed by combining the statements (3.17)-(3.18), (3.20), (3.25) and (3.26). □

Proof of theorem 3.2

By using the law of the iterated logarithm, we get

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{1/2} |I_1 - \sigma_\varepsilon^2| = \left(Var\varepsilon_1^2\right)^{1/2} \qquad a.s. \tag{3.27}$$

By the same arguments of the proof of Theorem 3.1 combined with the statement (3.27), the results of Theorem 3.2 follows. □

Proof of Theorem 3.3

Since B is positive definite matrix, then, from Theorem 2.1 of Chokri and Louani (2011), we have

$$R_n := \frac{n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top B(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \xrightarrow{d} \mathcal{X}^2_{(p)}. \tag{3.28}$$

Since from the model (3.8) we have

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i) + \varepsilon_i + (m_{add}(\mathbf{X}_i) - \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)),$$

it is clear that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2 + \frac{1}{n}\sum_{i=1}^{n}(m_{add}(\mathbf{X}_i) - \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i))^2 \\ &\quad + \frac{2}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})(m_{add}(\mathbf{X}_i) - \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)).\end{aligned}$$

Observe by the Cauchy-Schwarz inequality that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}|\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta}||m_{add}(\mathbf{X}_i) - \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)| \\ \leq &\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2\right]^{\frac{1}{2}}\left[\frac{1}{n}\sum_{i=1}^{n}(m_{add}(\mathbf{X}_i) - \widehat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i))^2\right]^{\frac{1}{2}}.\end{aligned}$$

Making use of Lemma 1 due to Camlong-Viot (2001) where all the conditions related to the mixing setting considered there are relaxed, it follows, under hypotheses [G.1-2 ;4], [H.1], [K.1-3], [M.1] and [Q.1-2], that

$$\max_{1\leq i\leq n}|m_{add}(\mathbf{X}_i) - \hat{m}^{\boldsymbol{\beta}}_{add}(\mathbf{X}_i)| = \mathcal{O}\left(\sqrt{\frac{\log n}{nh_n}}\right). \tag{3.29}$$

Therefore, using the statement (3.29), it is easily seen that

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i\boldsymbol{\beta})^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \sigma^2_\varepsilon, a.s. \tag{3.30}$$

Moreover, it is easy to see, again by Cauchy-Schwarz inequality, that

$$\begin{aligned}
|\widehat{\sigma}_n^2 - \sigma_\varepsilon^2| &\leq \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\mathbf{Z}}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)^2 + \frac{2}{n}\sum_{i=1}^{n}|\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta}||\tilde{\mathbf{Z}}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2 - \sigma_\varepsilon^2\right| \\
&\leq \max_{1\leq i\leq n}\|\tilde{\mathbf{Z}}_i\|^2\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 + 2\max_{1\leq i\leq n}\|\tilde{\mathbf{Z}}_i\|\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2\right]^{\frac{1}{2}} \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\mathbf{Z}}_i^\top\boldsymbol{\beta})^2 - \sigma_\varepsilon^2\right|.
\end{aligned}$$

Note that since the random vector $\mathbf{Z}$ is bounded, making use of Lemma 2.2 (ii), it follows that $\tilde{\mathbf{Z}}_i$ is also bounded for any $1 \leq i \leq n$. An application of Theorem 2.2 of Chokri and Louani (2011), gives

$$\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\| = o(1) \quad a.s. \tag{3.31}$$

Therefore, considering the statement (3.30) and (3.31) together, it is clear that

$$\widehat{\sigma}_n^2 - \sigma_\varepsilon^2 = o(1) \quad a.s. \tag{3.32}$$

Combining the statements (3.28), (3.32) and the Slutsky's Theorem, the proof of Theorem 3.3 follows. □

Proof of Theorem 3.4

The proof is a consequence of Theorem 3.1. □

Proof of Corollary 3.1

First, note that a straightforward calculus gives

$$\begin{aligned}
&V_n(\varepsilon_1^2) - Var\varepsilon_1^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \widehat{\sigma}_n^2\Big)^2 - \mathbb{E}\Big(\varepsilon_1^2 - E\varepsilon_1^2\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2 - (\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)\Big)^2 - \mathbb{E}\Big(\varepsilon_1^2 - \sigma_\varepsilon^2\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2\Big)^2 + (\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)^2 \\
&\quad -2\frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2\Big)(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2) - \mathbb{E}\Big(\varepsilon_1^2 - \sigma_\varepsilon^2\Big)^2.
\end{aligned}$$

In view of the statement (3.32) and the fact that

$$\frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2\Big) = o(1) \quad a.s.$$

and

$$(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)^2 = o(1) \quad a.s.,$$

we infer that

$$\begin{aligned}
&V_n(\varepsilon_1^2) - Var\varepsilon_1^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2\Big)^2 - \mathbb{E}\Big(\varepsilon_1^2 - \sigma_\varepsilon^2\Big)^2 + o(1) \quad a.s.
\end{aligned}$$

By the law of large numbers, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\Big((\tilde{Y}_i - \widetilde{\mathbf{Z}}_i^\top(\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i^\top)^{-1}\widetilde{\mathbf{Z}}_i\tilde{Y}_i)^2 - \sigma_\varepsilon^2\Big)^2 = \mathbb{E}\Big(\varepsilon_1^2 - \sigma_\varepsilon^2\Big)^2 + o(1) \quad a.s.$$

This, in turn, implies

$$V_n(\varepsilon_1^2) - Var\varepsilon_1^2 = o(1) \quad a.s. \tag{3.33}$$

Finally, we have

$$\frac{n(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)^2}{V_n(\varepsilon_1^2)} = \frac{n(\widehat{\sigma}_n^2 - \sigma_\varepsilon^2)^2}{Var\varepsilon_1^2} \times \frac{Var\varepsilon_1^2}{V_n(\varepsilon_1^2)}.$$

It suffice to combine the Theorem 3.4 and the statement (3.33) to achieve the proof of Corollary 3.1. □

Chapitre 4

Some uniform consistency results in the partially linear additive model components estimation

Ce chapitre est composé d'un article publié en (2014) dans la revue Comm. Statist. Theory Methods, mis en forme pour être inséré dans le présent manuscrit de thèse.

In this chapter, we are mainly concerned with the partially linear additive model defined, for a measurable function $\psi : \mathbb{R}^q \to \mathbb{R}$, by

$$\psi(\mathbf{Y}_i) := \mathcal{Y}_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \sum_{\ell=1}^{d} m_\ell(X_{\ell,i}) + \varepsilon_i \quad \text{for} \quad 1 \leq i \leq n,$$

where $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{ip})^\top$ and $\mathbf{X}_i = (X_{1,i}, \ldots, X_{id})^\top$ are vectors of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$is a vector of unknown parameters, $m_1, \ldots, m_d$ are unknown univariate real functions, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent random errors with mean zero, finite variances σ_ε and $\mathbb{E}(\varepsilon|\mathbf{X}, \mathbf{Z}) = 0$ a.s. We establish exact rates of strong uniform consistency of the nonlinear additive components of the model estimated by the marginal integration device with the kernel method. Our proofs are based upon the modern empirical process theory in the spirit of the works of Einmahl and Mason (2000) and Deheuvels and Mason (2004) relative to uniform deviations of nonparametric kernel-type estimators.

AMS Subject Classifications : 62G20, 62G32, 62J05, 60G07, 60F17, 60F15, 62G08.
Keywords : Additive model ; Regression function ; Marginal integration ; Non-parametric Estimation ; Density estimation ; Regression estimation ; Strongly consistent ; Kernel estimation ; Empirical processes ; Functional estimation.

4.1 Introduction

Let $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be a $\mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^p$-valued random vector and $\psi : \mathbb{R}^q \to \mathbb{R}$ be a measurable function. We assume that the relation between the response variables and the covariates can be represented by the partially linear model, that is

$$\psi(\mathbf{Y}) = \mathcal{Y} = m_0 + \mathbf{Z}^\top \boldsymbol{\beta} + m(\mathbf{X}) + \varepsilon, \tag{4.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector column of unknown parameters, m is the nonlinear part of the model and ε is the modeling error and m_0 is a constant. Here and elsewhere, the transpose of a vector $\mathbf{V}$ will be denoted by $\mathbf{V}^\top$. Model (4.1) compromises between the linear model and the fully nonparametric model, the linear component $\mathbf{Z}^\top \boldsymbol{\beta}$ provides a simple summary of covariate effects which is estimated with the best rates of convergence, while the smooth baseline component m is included to insure more model flexibility. We mention that the nonparametric isotonic regression model, which is a special case of the model (4.1) without the linear part, was first proposed by Brunk (1958) and has received extensive attention since the middle of the last century in the statistical literature. Since the introduction of (4.1) by Engle et al. (1986), for $\psi(\mathbf{y}) = y$, the partially linear model has been widely used in various fields, see for example Speckman (1988), Liang et al. (1999), Severini and Staniswalis (1994) and the references therein. The model (4.1), for $\psi(\mathbf{y}) = y$, and various associated estimators, test statistics, and generalizations have generated a substantial body of literature, which includes the works of Rice (1986), Chen (1988), Robinson (1988), Chen and Shiau (1991), Eubank and Speckman (1990) Donald and Newey (1994). Shi and Li (1995a,b) and Hamilton and Truong (1997). A comprehensive literature review of the estimation and application of model (4.1), for $\psi(\mathbf{y}) = y$, can be found in the monograph Härdle et al. (2000). Parametric regression models provide powerful tools for analyzing practical data when the models are correctly specified, but may suffer from large modeling biases when the structures of the models are misspecified. As an alternative, nonparametric smoothing methods ease the concerns on modeling biases. However, it is well known that unrestricted multivariate nonparametric regression models are subject to the *curse of dimensionality*, and fail to take advantage of the flexibility structure

in modeling phenomena with *moderate* set of data. The papers by Stone (1985, 1986) proposed the additive model regression, which allows easier interpretation of the contribution of each explanatory variable and reduction of the computational requirement. Hence, additive regression models circumvent the *curse of dimensionality* that afflicts the estimation of fully nonparametric regression models. There is a huge literature on these models and their applications. It is not the purpose of this paper to survey this extensive literature, the interested reader may refer to Fan and Gijbels (1996), Härdle (1990) and the references therein for further details.
To reduce the dimension impact of the nonparametric part in the partially linear regression model (4.1), we consider the partially linear additive model that puts an additive structure to the nonparametric function m

$$\psi(\mathbf{Y}) = m_0 + \mathbf{Z}^\top \boldsymbol{\beta} + \sum_{\ell=1}^{d} m_\ell(X_\ell) + \varepsilon, \tag{4.2}$$

where X_ℓ is the ℓ-th component of the vector $\mathbf{X}$ and m_ℓ is a real univariate function. Manzan and Zerom (2005) showed that the estimator $\boldsymbol{\beta}$ which is designed to exploit additivity is asymptotically more efficient than an estimator that ignores the additive structure, see for further details Chamberlain (1992). We may refer also to the paper by Mammen et al. (2011), that analyzes efficiency gains in semiparametric models from imposing additional structure on the nonparametric component.

The primary goal of the present paper is concerned with the exact rate of strong uniform consistency of estimators of additive regression model components in the same spirit, as the results on the weak consistency investigated in the work Debbarh (2008) and the related results established in Deheuvels and Mason (2004) and Einmahl and Mason (2000, 2005) for kernel-type functionals estimates. Our results allow us to build simultaneous almost certainty bands (100% confidence bands) for the components that we estimate.

The layout of the article is as follows. In the forthcoming section, we first present the model and the estimators that we consider throughout this paper. In Section 4.3, we state the major assumptions and give the main theoretical results. In Section 4.4, we give some concluding remarks and possible developments. To avoid interrupting the flow of the presentation, the mathematical developments are relegated to Section 4.5. A few technical results are given in the appendix.

4.2 Presentation of estimators

Let $(\mathbf{X}_i, \mathbf{Y}_i, \mathbf{Z}_i)_{i\geqslant 1}$ be a sequence of independent random replicæ of a random vector $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Denote by $f_{\mathbf{XYZ}}$ its joint density function with respect to the Lebesgue measure and by $f_{\mathbf{X}}$, $f_{\mathbf{Y}}$ and $f_{\mathbf{Z}}$ the corresponding marginals. For any $\mathbf{x} \in \mathbb{R}^d$, we define the kernel estimator of the marginal density $f_{\mathbf{X}}$ by

$$f_n(\mathbf{x}) = \frac{1}{na_n^d} \sum_{i=1}^{n} \mathbb{K}\left(\frac{\mathbf{x} - \mathbf{X}_i}{a_n}\right),$$

where $\mathbb{K}$ is a non-negative function defined on $\mathbb{R}^d$ and integrating to 1 and a_n is a smoothing parameter tending to zero with a suitable rate that will be given later on. To avoid unnecessary complexity, we assume $m_0 = 0$. Notice that the model (4.1) may be rewritten as follows

$$\psi(\mathbf{Y}) - \mathbf{Z}^\top \boldsymbol{\beta} = m(\mathbf{X}) + \varepsilon. \tag{4.3}$$

On the basis of the representation (4.3), using the internal estimator, the regression estimator involving the nonparametric part of the model may be defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{i=1}^{n} \frac{\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta}}{nh_n^d f_n(\mathbf{X}_i)} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \tag{4.4}$$

where K is a kernel function defined on $\mathbb{R}^d$ that is bounded, non-negative and integrated to 1, and h_n is the smoothing parameter converging to zero with a specific rate that we precise below in condition (H.2). See Remark 4.1 below for a discussion on the definition of this and other related estimators. Note that the estimator $\widehat{m}_n^{\boldsymbol{\beta}}$ depends on the unknown parameter $\boldsymbol{\beta}$ which needs to be estimated.
Keeping in mind the model defined in (4.3), the function m depends naturally on the parameter $\boldsymbol{\beta}$ and its additive structure may be written as follows

$$m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{\ell=1}^{d} m_\ell^{\boldsymbol{\beta}}(x_\ell), \tag{4.5}$$

where x_ℓ is the ℓ-th component of $\mathbf{x}$. For identifiability of the additive component functions m_ℓ, we put the constraints

$$\mathbb{E} m_\ell^{\boldsymbol{\beta}}(X_\ell) = 0 \quad \text{for} \quad 1 \leq \ell \leq d.$$

Various methods have been proposed in the literature to estimate the additive components of the regression model including the marginal integration method, which

consists in integrating the regression function m with respect to a suitable density function, [see, e.g., Newey (1994), Tjøstheim and Auestad (1994), Linton and Nielsen (1995), Camlong-Viot et al. (2000) and Camlong-Viot et al. (2006)] and the back-fitting algorithms [cf. to Hastie and Tibshirani (1986), Sperlich et al. (1999) and the references therein for an account of the results on these topics]. The back-fitting asymptotic properties are difficult to analyze, however, the works Opsomer and Ruppert (1997) and Mammen et al. (1999) have made an important progress in the development of the asymptotic theory of back-fitting methodology. In contrast, the asymptotic theory of the marginal integration method is relatively easy to establish which has attracted much attention among statisticians and econometricians. Motivated by these properties, our approach based on the marginal integration method to estimate the linear parameter $\boldsymbol{\beta}$ uses the least square criterion. It results then that our estimates take the form which depends only on the kernels, the smoothing parameters and the integration density, avoiding to consider any choice procedure of the weights leading to minimization issues. Notice that estimators built up considering unknown weight quantities are studied in several papers. Obviously, in such cases, optimization procedures are needed to obtain efficient estimates. The resulting optimal weights are always functions depending on unknown parameters that must be estimated. This fact was illustrated, for instance, by the work Fan et al. (1998).

Throughout this paper, the marginal integration method is used to build the estimates. To be more precise, we first need to introduce some notation and definitions. For any $1 \leq \ell \leq d$, set $\mathbf{x}_{-\ell} = (x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_d)$,

$$q_{-\ell}(\mathbf{x}_{-\ell}) = \prod_{j \neq \ell}^{d} q_j(x_j) \quad \text{and} \quad q(\mathbf{x}) = \prod_{\ell=1}^{d} q_\ell(x_\ell),$$

where q_ℓ, $1 \leq \ell \leq d$, are known univariate density functions. In the following, all the integrals related with continuous variables will be taken with respect to Lebesgue measure. Following the marginal integration procedure, the additive regression function is given, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{\ell=1}^{d} \eta_\ell^{\boldsymbol{\beta}}(x_\ell) + \int_{\mathbb{R}^d} m_{add}^{\boldsymbol{\beta}}(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \tag{4.6}$$

where

$$\eta_\ell^{\boldsymbol{\beta}}(x_\ell) = \int_{\mathbb{R}^{d-1}} m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} m_{add}^{\boldsymbol{\beta}}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \tag{4.7}$$

Here, $\eta_\ell^{\boldsymbol{\beta}}$ is the ℓ-th component of the additive regression function which still depends on the parameter $\boldsymbol{\beta}$. To unburden our notation a bit, $m_{add}^{\boldsymbol{\beta}}$ will be denoted by m_{add} and $\eta_\ell^{\boldsymbol{\beta}}$ by η_ℓ. Now, for a given parameter $\boldsymbol{\beta}$, using the marginal integration method, we define the estimator of m_{add} as follows

$$\widehat{m}_{add}^{\boldsymbol{\beta}}(\mathbf{x}) = \sum_{\ell=1}^{d} \widehat{\eta}_\ell^{\boldsymbol{\beta}}(x_\ell) + \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \tag{4.8}$$

where

$$\widehat{\eta}_\ell^{\boldsymbol{\beta}}(x_\ell) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \tag{4.9}$$

Either, in what follows, $\widehat{m}_{add}^{\boldsymbol{\beta}}$ will be denoted by $\widehat{m}_{add}$, $\widehat{m}_n^{\boldsymbol{\beta}}$ by $\widehat{m}_n$ and $\widehat{\eta}_\ell^{\boldsymbol{\beta}}$, the estimate of the ℓ-th component of the additive regression function with fixed parameter $\boldsymbol{\beta}$, by $\widehat{\eta}_\ell$. Therefore, we shall estimate the vector parameter $\boldsymbol{\beta}$ to get a ready estimate of m_{add}. We close this section by the following remark.

Remark 4.1. Let us recall general kernel-type estimator of regression function defined, for $\mathbf{x} \in \mathbb{R}^d$, by

$$\widehat{m}_{n;h_n}(\mathbf{x}, \psi) := \frac{\sum_{i=1}^n \psi(\mathbf{Y}_i) K((\mathbf{x} - \mathbf{X}_i)/h_n)}{\sum_{i=1}^n \mathbb{K}((\mathbf{x} - \mathbf{X}_i)/h_n)}. \tag{4.10}$$

By setting $\psi(y) = y$ into (4.10) we get the classical Nadaraya-Watson kernel regression function estimator of $m(\mathbf{x}) := \mathbb{E}(Y \mid \mathbf{X} = \mathbf{x})$ given by

$$\widehat{m}_{n;h_n}(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i K((\mathbf{x} - \mathbf{X}_i)/h_n)}{\sum_{i=1}^n \mathbb{K}((\mathbf{x} - \mathbf{X}_i)/h_n)}. \tag{4.11}$$

We define the *internal estimator* at some predefined point $\mathbf{x}$ by

$$\widehat{m}_{n;h_n}^{\text{Int}}(\mathbf{x}) := \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \frac{Y_i}{f_n(\mathbf{X}_i)}. \tag{4.12}$$

For more details on the estimators (4.11) and (4.12), the interested reader may refer, e.g., to Wand and Jones (1995). Linton and Jacho-Chávez (2010) pointed out that Mack and Müller (1989) were the first to propose $\widehat{m}_{n;h_n}^{\text{Int}}(\mathbf{x})$, for $d = 1$, with a view to the estimation of derivatives by computing the derivative of the regression, which has a simpler form than the derivative of the Nadaraya-Watson smoother. The term "internal" stands for the fact that the factor $f_n^{-1}(\mathbf{X}_i)$ is internal to the summation, while the estimator $\widehat{m}_{n;h_n}(\mathbf{x})$ has the factor

$$f_n^{-1}(\mathbf{x}) = \left(\frac{1}{na_n^d} \sum_{i=1}^n \mathbb{K}\left(\frac{\mathbf{x} - \mathbf{X}_i}{a_n}\right)\right)^{-1}$$

externally to the summation. Jones et al. (1994) considered various versions of kernel-type regression estimators, especially the Nadaraya-Watson estimator and the local linear estimator. They established the equivalence between the local linear estimator and the internal estimator. Linton and Jacho-Chávez (2010) and Shen and Xie (2013) indicated that the internal estimators are particularly adequate for the additive non-parametric regression model, since,

$$\int_{\mathbb{R}^{d-1}} \widehat{m}^{\text{Int}}_{n;h_n}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} = \frac{1}{nh^d} \sum_{i=1}^{n} \frac{Y_i}{f_n(\mathbf{X}_i)} \int_{\mathbb{R}^{d-1}} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} \tag{4.13}$$

If $q_{-\ell}(\cdot)$ is chosen smooth enough, and the kernel K such that the convolution

$$\int_{\mathbb{R}^{d-1}} K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell}$$

can be very closely approximated by

$$K_\ell\left(\frac{x_\ell - X_{i,\ell}}{h_n}\right) q_{-\ell}(\mathbf{X}_{-\ell}).$$

Then the last integral in (4.13) can be approximated as follows

$$\int_{\mathbb{R}^{d-1}} \widehat{m}^{\text{Int}}_{n;h_n}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} \approx \frac{1}{nh_n} \sum_{i=1}^{n} \frac{Y_i}{f_n(\mathbf{X}_i)} K_\ell\left(\frac{x_\ell - X_{i,\ell}}{h_n}\right) q_{-\ell}(\mathbf{X}_{-\ell}). \tag{4.14}$$

Notice that the estimator in (4.12) has been used also by Hengartner and Sperlich (2005) in the context of estimating additive models. The authors showed that it has some additional theoretical advantages over the use of Nadaraya-Watson estimator. More precisely, it is possible to obtain asymptotic normality of $\widehat{\eta}_\ell(x_\ell)$ at the optimal rate for one-dimensional nonparametric regression without assuming additional smoothness. They also mention the computational attractiveness and better performance of the internal estimator (compared to its classical counterpart), in particular, when the covatiates are correlated and nonuniformly distributed. Indeed, the internality of the factor $f^{-1}(\mathbf{X}_i)$ in the summation play an instrumental role in the approximation (4.14), that simplifies very much the computation of the estimator $\widehat{\eta}_\ell(x_\ell)$.

4.2.1 Estimation of the parameter $\boldsymbol{\beta}$

Consider the partially linear additive regression model

$$\psi(\mathbf{Y}) = \mathbf{Z}^\top \boldsymbol{\beta} + m_{add}(\mathbf{X}) + \varepsilon.$$

In the sequel, we will use the notation, for $\ell = 1, \ldots, d$,

$$K_\ell(u_\ell) = \int_{\mathbb{R}^{d-1}} K(\mathbf{u}) d\mathbf{u}_{-\ell}, \quad \text{and} \quad K_{-\ell}(\mathbf{u}_{-\ell}) = \int_{\mathbb{R}} K(\mathbf{u}) du_\ell.$$

Following Chokri and Louani (2011), the estimator of $\boldsymbol{\beta}$ is defined by

$$\widehat{\boldsymbol{\beta}} = [\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^\top]^{-1}\widetilde{\mathbf{Z}}\widetilde{Y}, \tag{4.15}$$

where

$$\widetilde{Y} = \left[\psi(\mathbf{Y}_i) - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\psi(\mathbf{Y}_j)\right]_{1\le i\le n}^{\top}, \quad \widetilde{\mathbf{Z}} = \left[\mathbf{Z}_i - \sum_{j=1}^{n} W_{nj}(\mathbf{X}_i)\mathbf{Z}_j\right]_{1\le i\le n} \tag{4.16}$$

$$W_{nj}(\mathbf{X}_i) = \frac{U_{nj}(\mathbf{X}_i)}{nf_n(\mathbf{X}_j)},$$

and

$$U_{nj}(\mathbf{X}_i) = \sum_{\ell=1}^{d} \frac{1}{h_n} K_\ell\left(\frac{X_{\ell,i} - X_{j,\ell}}{h_n}\right) D_{-\ell} - (d-1)\int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{\mathbf{z} - \mathbf{X}_j}{h_n}\right) q(\mathbf{z}) d\mathbf{z},$$

where

$$D_{-\ell} = \int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K\left(\frac{\mathbf{z}_{-\ell} - \mathbf{X}_{-\ell,j}}{h_n}\right) q_{-\ell}(\mathbf{z}_{-\ell}) d\mathbf{z}_{-\ell}.$$

For more details on the estimation of the parameter $\boldsymbol{\beta}$, see the paper Robinson (1988). Finally, to estimate the regression function and the additive components, we use the following

$$\widehat{m}_{add}(\mathbf{x}) = \sum_{\ell=1}^{d} \widehat{\eta}_\ell(x_\ell) + \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \tag{4.17}$$

and

$$\widehat{\eta}_\ell(x_\ell) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}. \tag{4.18}$$

4.3 Main results

To state our results, we consider an additional assumption on the model structure. In this respect we suppose that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \widetilde{\mathbf{Z}}_i^\top \widetilde{\mathbf{Z}}_i = B \qquad a.s., \tag{4.19}$$

where B is a $p \times p$-positive definite matrix. Throughout this paper, we assume that $\psi \in \mathcal{F}$ which will denote a pointwise measurable VC subgraph class of $\mathbb{R}^q$-valued functions. Let

$$I^d = \prod_{\ell=1}^{d} I_\ell = \prod_{\ell=1}^{d} [a_\ell, c_\ell]$$

and

$$J^d = \prod_{\ell=1}^{d} J_\ell = \prod_{\ell=1}^{d} [a'_\ell, c'_\ell]$$

be two fixed pavements of $\mathbb{R}^d$ such that $a'_\ell < a_\ell < c_\ell < c'_\ell$, for $1 \leq \ell \leq d$.
Further assumptions involving the density function of $\mathbf{X}$, the regression function m, the kernel functions K_ℓ, for $1 \leq \ell \leq d$, K and the smoothing parameters are gathered together hereafter for easy reference and reader's convenience. The first part of this set of assumptions concern the density function $f_{\mathbf{X}}$ and the regression function m.

(G.1) m is k-times continuously differentiable and there exists a constant $0 < \mathfrak{C} < \infty$ such that

$$\sup_{\mathbf{x} \in I} \left| \frac{\partial^k m(\mathbf{x})}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} \right| \leq \mathfrak{C}, \quad k_1, \dots, k_d \geq 0, \quad k_1 + \dots + k_d = k;$$

(G.2) The density $f_{\mathbf{X}}$ is continuous and bounded away from 0 on $\mathbb{R}^d$;

(G.3) The density $f_{\mathbf{X}}$ is k'-continuously differentiable on its support and $k' > kd$.

In the sequel, $(h_n)_{n \geq 1}$ and $(a_n)_{n \geq 1}$ denote sequences of positive constants satisfying the following conditions.

(H.1) $h_n = \vartheta_1 \left(\dfrac{\sqrt{\log n}}{n} \right)^{1/(2k+1)}$ for $0 < \vartheta_1 < \infty$;

(H.2) (i) $a_n \to 0$, (ii) $n a_n^d \to \infty$, (iii) $\dfrac{\log(1/a_n^d)}{\log \log n} \to \infty$ and $\dfrac{h_n}{a_n^d} \dfrac{\log n}{\log(1/h_n)} \to 0$ as $n \to \infty$.

We suppose that the kernel functions $\mathbb{K}$ and K are bounded, Lipschitz continuous functions. We say that the kernel L is of order s if

$$\int_{\mathbb{R}^d} L(\mathbf{x}) d\mathbf{x} = 1,$$
$$\int_{\mathbb{R}^d} x_1^{k_1} \dots x_d^{k_d} L(\mathbf{x}) d\mathbf{x} = 0, \quad k_1, \dots, k_d \geq 0, \quad k_1 + \dots + k_d = 1, \dots, s-1,$$
$$\int_{\mathbb{R}^d} |x_1^{k_1} \dots x_d^{k_d}| L(\mathbf{x}) d\mathbf{x} < \infty, \quad k_1, \dots, k_d \geq 0, \quad k_1 + \dots + k_d = s.$$

In addition, we assume that $\mathbb{K}$ and K satisfy the following additional assumptions.

(K.1) $\mathbb{K}$ and K are of order k' and k respectively ;

(K.2) $K(\mathbf{x}) = 0$, for $\mathbf{x} \notin [-\varrho/2, \varrho/2]^d$, for some $0 < \varrho < \infty$;

(K.3) K is of bounded variation on $\mathbb{R}^d$.

Also, we consider the following assumptions upon the random variables Y and $\mathbf{Z}$.

(M.1) The class of functions $\mathcal{F}$ is bounded ;

(M.2) $\mathbf{Z}\mathbb{1}_{\{\mathbf{X}\in I^d\}}$ is bounded.

Finally, to state our results, we will work under the following assumptions on the density functions q_ℓ, for $1 \leq \ell \leq d$.

(Q.1) q_ℓ is bounded and continuous, for all $1 \leq \ell \leq d$;

(Q.2) q_ℓ has $k+1$ continuous and bounded derivatives, with compact support $\mathscr{C}_\ell \subset I_\ell$ for all $1 \leq \ell \leq d$.

Now, we define a function ϕ on the interval I_ℓ as

$$\phi(u_\ell) = \int_{\mathbb{R}^{d-1}} \frac{H^{\boldsymbol{\beta}}(\mathbf{u})}{f(\mathbf{u}_{-\ell}|u_\ell)} q_{-\ell}(\mathbf{u}_{-\ell}) d\mathbf{u}_{-\ell}, \tag{4.20}$$

where, for fixed $\psi \in \mathcal{F}$,

$$H^{\boldsymbol{\beta}}(\mathbf{u}) = \mathbb{E}[(\psi(\mathbf{Y}) - \mathbf{Z}^\top \boldsymbol{\beta})^2 | \mathbf{X} = \mathbf{u}], \mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d. \tag{4.21}$$

Here, $f(\mathbf{u}_{-\ell}|u_\ell)$ denotes the conditional density of $\mathbf{X}_{-\ell}$ given $X_\ell = u_\ell$ which we assume to be bounded away from zero on the support of $f_{\mathbf{X}}$, this requirement is used in order to make ϕ well defined. For $1 \leq \ell \leq d$, consider the following quantity,

$$\sigma_\ell = \sigma_\ell^{\boldsymbol{\beta}} = \sup_{u_\ell \in I_\ell} \sqrt{\frac{\phi(u_\ell)}{f_\ell(u_\ell)} \int_{\mathbb{R}} K_\ell^2(t) dt}, \tag{4.22}$$

where f_ℓ is the ℓ-th marginal density of $\mathbf{X}$.

The main result, concerning the additive component estimate $\widehat{\eta}_\ell$, to be proved here may now be stated precisely as follows.

Theorem 4.1. Assume that the conditions (G.1)-(G.3), (H.1)-(H.2), (K.1)-(K.3), (M.1)-(M.2) and (Q.1)-(Q.2) are satisfied. Then, we have for fixed $\psi \in \mathcal{F}$,

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{x_1 \in I_1} |\widehat{\eta}_1(x_1) - \eta_1(x_1)| = \sigma_1 \quad a.s. \tag{4.23}$$

The proof of Theorem 4.1 is postponed until §4.5 .
The following result handles the uniform deviation of the estimate $\widehat{m}_{add}$ with respect to m_{add} and provides the uniform strong consistency with rate.

Theorem 4.2. Under assumptions of Theorem 4.1, we have

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{2\log(1/h_n)}}\sup_{\mathbf{x}\in I^d}\left|\widehat{m}_{add}(\mathbf{x})-m_{add}(\mathbf{x})\right|=\sum_{\ell=1}^{d}\sigma_\ell\quad a.s. \tag{4.24}$$

The proof of Theorem 4.1 is postponed until §4.5.
As a consequence of Theorem 4.1 and 4.2, we have the following corollary.

Corollary 4.1. Under assumptions of Theorem 4.1, we have

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{2\log(1/h_n)}}\sup_{\mathbf{x},\mathbf{z}\in I^d\times\mathbb{R}^p}\left|\widehat{m}(\mathbf{x},\mathbf{z})-m(\mathbf{x},\mathbf{z})\right|=\sigma_\varepsilon\sum_{\ell=1}^{p}(b^{\ell\ell})^{1/2}+\sum_{\ell'=1}^{d}\sigma_{\ell'}\quad a.s., \tag{4.25}$$

where

$$\widehat{m}(\mathbf{x},\mathbf{z})=\mathbf{z}^\top\widehat{\boldsymbol{\beta}}+\widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}),$$

σ_ε^2 is the variance of ε and $b^{\ell\ell}, \ell=1,\ldots,p$, are the diagonal entries of matrix B^{-1}.

The proof of Corollary 4.1 is postponed until §4.5.

Remark 4.2. Notice that the condition (H.2) is satisfied for the choice

$$a_n=\vartheta_2\left(\frac{\sqrt{\log n}}{n}\right)^{1/(2k'+d)}\quad\text{for}\quad 0<\vartheta_2<\infty.$$

The limiting behavior of $f_n(\cdot)$, for appropriate choices of the bandwidth a_n, has been studied by a large number of statisticians over many decades. For good sources of references to research literature in this area along with statistical applications consult Devroye and Lugosi (2001), Devroye and Györfi (1985), Bosq and Lecoutre (1987), Scott (1992), Wand and Jones (1995) and Prakasa Rao (1983). In particular, under our assumptions, the condition that $a_n\to 0$ together with $na_n^d\to\infty$ is necessary and sufficient for the convergence in probability of $f_n(\mathbf{x})$ towards the limit $f(\mathbf{x})$, independently of $\mathbf{x}\in\mathbb{R}^d$ and the density $f(\cdot)$. The convergence result in (4.34) below, should minimally require (H.2)(iii) to hold almost surely which illustrates the sharpness of the condition (H.2). Finally, if we replace (H.2)(iii) by $na_n^d/\log n\to\infty$, the results of the present work hold in probability.

Remark 4.3. Note that the condition (M.1) may be replaced by more general hypotheses upon moments of $\mathbf{Y}$ as in Einmahl and Mason (2000, 2005). That is

(M.1)′ The class $\mathcal{F}$ has a finite valued measurable envelope function

$$F(\mathbf{y}) \geq \sup_{\psi \in \mathcal{F}} |\psi(\mathbf{y})|, \quad \mathbf{y} \in \mathbb{R}^q,$$

such that, for some $s > 2$,

$$\sup_{\mathbf{x} \in I^d} \mathbb{E}(|F(\mathbf{Y})|^s | \mathbf{X} = \mathbf{x}) < \infty,$$

or a more general form as in Deheuvels (2011)

(M.1)″ We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$, and such that, for some $s > 2$, ultimately as $x \uparrow \infty$,

$$(i)\ x^{-s}\mathcal{M}(x) \downarrow; (ii)\ \ x^{-1}\mathcal{M}(x) \uparrow . \tag{4.26}$$

For each $t \geq \mathcal{M}(0)$, we define $\mathcal{M}^{inv}(t) \geq 0$ by $\mathcal{M}(\mathcal{M}^{inv}(t)) = t$. We assume further that :

$$\sup_{\mathbf{x} \in I^d} \mathbb{E}\left(\mathcal{M}\left(|F(\mathbf{Y})|\right) | \mathbf{X} = \mathbf{x}\right) < \infty.$$

The boundedness assumption on the support can be replaced by a finite moment assumption (M.1)′, but this will add much extra complexity to the proofs. We will need also that the sequence $\{h_n\}_{n \geq 1}$ satisfies

$$h_n^{-1} \leq (n/\log(1/h_n))^{1-2/s}.$$

Under the last additional condition and replacing condition (M.1) by (M.1)′ all our results remain valid, see Section 4.6.1. For more details we may refer to Einmahl and Mason (2000), Deheuvels and Mason (2004) and Deheuvels (2011). The introduction of the function ψ in our setting motivated by Remark 1.2 of Deheuvels and Mason (2004) or Remark 1.1 of Deheuvels (2011).

Remark 4.4. Notice that the conditions (G.1), (G.3) and (K.1) are classical in the nonparametric estimation procedures. In particular, by imposing the condition (K.1), the kernel function exploits the smoothness of the density function or the regression function. It is well known that the best obtainable rate of convergence of the kernel estimator, in the AMISE sense, is of order $n^{-4/5}$, in the univariate case. If we loose the condition that the kernel function K must be a density, the convergence rate could be faster. Indeed, the convergence rate can be made arbitrarily close to the parametric n^{-1}

as the order increases. In fact, Chacón et al. (2007) showed that the parametric rate n^{-1} can be attained by the use of superkernels, and that superkernel density estimators automatically adapt to the unknown degree of smoothness of the density. The main drawback of higher-order kernels in this situation is the negative contributions of the kernel may make the the estimated density not a density itself. The interested reader may refer to, e.g., Jones et al. (1995), Jones and Signorini (1997) and Jones (1995).

4.3.1 Confidence bands

Our results allow to construct simultaneous confidence bands for the true component $\eta_1(x_1)$ of the additive regression function. Towards this end, we infer from Theorem 4.1 that, for any $0 < \varepsilon < 1$ and suitable chosen data-dependent functions $L_n(x_1) > 0$, we have, as $n \to \infty$,

$$\mathbb{P}\left(\eta_1(x_1) \in [\widehat{\eta}_1(x_1) - (1+\varepsilon)L_n(x_1), \widehat{\eta}_1(x_1) + (1+\varepsilon)L_n(x_1)], \forall x_1 \in I_1\right) \to 1, \quad (4.27)$$

and

$$\mathbb{P}\left(\eta_1(x_1) \in [\widehat{\eta}_1(x_1) - (1-\varepsilon)L_n(x_1), \widehat{\eta}_1(x_1) + (1-\varepsilon)L_n(x_1)], \forall x_1 \in I_1\right) \to 1. \quad (4.28)$$

Whenever the statements (4.27) and (4.28) hold jointly, we can say that the intervals

$$[A_{n,1}(x_1), B_{n,1}(x_1)] = \left[\widehat{\eta}_1(x_1) - L_n(x_1), \widehat{\eta}_1(x_1) + L_n(x_1)\right],$$

provide asymptotic simultaneous optimal confidence bands (at an asymptotic confidence level of 100%) for $\eta_1(x_1)$ over $x_1 \in I_1$. In practice, the plot of $A_{n,1}(x_1)$ and $B_{n,1}(x_1)$ over $x_1 \in I_1$ provides useful visual information on the unknown values of η_1 on I_1 (refer to Derzko and Deheuvels (2002) for a biomedical example).
To construct $L_n(x_1)$, for $x_1 \in I_1$, it suffices to choose an appropriate estimator $\sigma_{n,1}$ of σ_1, and to take then

$$L_n(x_1) = \left\{ \frac{2\log(1/h_n)}{nh_n} \times \sigma_{1,n}^2(x_1) \right\}^{1/2}.$$

Considering the statements (4.20)-(4.22), it is natural to take the following estimate of σ_1^2

$$\begin{aligned} \sigma_{1,n}^2(x_1) &= \frac{\int_{\mathbb{R}} K_1^2(u)du}{nh_n} \sum_{i=1}^{n} (\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}})^2 K\left(\frac{x_1 - X_{1,i}}{h_n}\right) \\ &\quad \times \int_{\mathbb{R}^{d-1}} \frac{K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right)}{h_n^{d-1} f_n(\mathbf{x})^2} q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}. \end{aligned}$$

Notice that the confidence bands of the additive regression function $m_{add}(\mathbf{x})$ may be deduced similarly from Theorem 4.2. Indeed, the asymptotic confidence bands for $m_{add}(\mathbf{x})$, for $\mathbf{x}$ in I^d, is given by

$$[\mathbf{A}_n(\mathbf{x}), \mathbf{B}_n(\mathbf{x})] = \left[\sum_{i=1}^{d} A_{n,i}(x_i) + \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{z})q(\mathbf{z})d\mathbf{z}, \sum_{i=1}^{d} B_{n,i}(x_i) + \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{z})q(\mathbf{z})d\mathbf{z}\right].$$

Remark 4.5. Since the confidence bands depend upon h_n, an optimal choice of h_n with respect to some criterion will improve them. Notice that the paper Deheuvels and Mason (2004) consider local plug-in type estimators $\widehat{h}_n = \widehat{h}_n(\mathbf{x})$ of h that satisfy the condition

$$\mathbb{P}\left(a_n \leq \widehat{h}_n(\mathbf{x}) \leq b_n : \mathbf{x} \in I^d\right) \to 1,$$

where $a_n = c_1 h_n$ and $b_n = c_2 h_n$, with $c_1 < c_2$, or fulfill, for any $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{\mathbf{x}\in I^d} \left|\frac{\widehat{h}_n(\mathbf{x})}{h_n} - \varphi(\mathbf{x})\right| > \varepsilon\right) \to 0, \tag{4.29}$$

where φ is an appropriate continuous function on I^d. This issue together with the related uniform in bandwidth problem will be investigated in a future work.

4.4 Concluding remarks

We have addressed the problem of derivation of uniform results for nonparamptric kernel-type estimators of the components of additive regression. Toward this aim, we have used the methodology of proof proposed in Einmahl and Mason (2000). In general, the stochastic component of the error between m_n and m is a greater problem than the deterministic one, i.e., the bias. For bounding the stochastic part, i.e., in order to find almost sure bounds of the stochastic process, we especially need assumptions about the kernels with a recurrent assumption being right-continuity and bounded variation in addition of general assumptions about the density. The deterministic component requires more assumptions about the smoothness of m and the density function f.
There is an increasing interest in obtaining so-called uniform in bandwidth results for nonparametric estimators depending on a bandwidth sequence. The first paper that focused on obtaining uniform in bandwidth results for the kernel density estimator making use of empirical process techniques were Einmahl and Mason (2005). After this seminal paper many other works investigated these uniform in bandwidth problems in different contexts, such as the local polynomial regression [Dony et al. (2006) and Blondin (2007)], the local uniform empirical process [Varron (2006)], the conditional

U-statistics [Dony and Mason (2008)], the estimation of integral functionals of the density [Giné and Mason (2008)], Shannon's entropy [Bouzebda and Elhattab (2009, 2011)], the kernel distribution function estimators and the smoothed empirical process [Mason and Swanepoel (2010) and Chacón and Rodríguez-Casal (2010)], the kernel-type estimators of copula derivatives [Bouzebda (2012)]. It will be interesting to enrich our results presented here by an additional uniformity in term of h_n in the supremum appearing in all our theorems, which requires non trivial mathematics, this would go well beyond the scope of the present paper.

4.5 Proofs

This section is devoted to the proofs of our results. Throughout the proofs of our results, we shall assume, for the sake of notational convenience, but without loss of generality, that our kernel K has support contained in $[-1/2, 1/2]$.

Proof of Theorem 4.1.

Recall that

$$\widehat{\eta}_\ell(x_\ell) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{x}) q(\mathbf{x}) d\mathbf{x},$$

where

$$\widehat{m}_n(\mathbf{x}) = \sum_{i=1}^n \frac{\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}}{n h_n^d f_n(\mathbf{X}_i)} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right).$$

We first decompose $\{\widehat{\eta}_\ell(x_\ell) - \eta_\ell(x_\ell)\}$, for $1 \leq \ell \leq d$, into the sum of two components, by writing

$$\begin{aligned} \widehat{\eta}_\ell(x_\ell) - \eta_\ell(x_\ell) &= \widehat{\eta}_\ell^{\widehat{\boldsymbol{\beta}}}(x_\ell) - \widehat{\eta}_\ell^{\boldsymbol{\beta}}(x_\ell) + \widehat{\eta}_\ell^{\boldsymbol{\beta}}(x_\ell) - \eta_\ell(x_\ell) \\ &:= \mathcal{D}_1(x_\ell) + \mathcal{D}_2(x_\ell). \end{aligned} \tag{4.30}$$

Let, for $1 \leq \ell \leq d$,

$$\widehat{\widehat{\eta}}_\ell(x_\ell) := \widehat{\widehat{\eta}}_\ell^{\boldsymbol{\beta}}(x_\ell) := \int_{\mathbb{R}^{d-1}} \widehat{\widehat{m}}_n(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widehat{\widehat{m}}_n(\mathbf{x}) q(\mathbf{x}) d\mathbf{x},$$

where

$$\widehat{\widehat{m}}_n(\mathbf{x}) := \widehat{\widehat{m}}_n^{\boldsymbol{\beta}}(\mathbf{x}) := \sum_{i=1}^n \frac{\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta}}{n h_n^d f_{\mathbf{X}}(\mathbf{X}_i)} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right).$$

There is no loss of generality to display our proofs only for $\ell = 1$.

Part (I). Considering the second term $\mathcal{D}_2$, observe that one may write

$$\begin{aligned}\mathcal{D}_2(x_1) &= \widehat{\eta}_1(x_1) - \eta_1(x_1) \\ &= [\widehat{\eta}_1(x_1) - \widehat{\widehat{\eta}}_1(x_1)] + [\widehat{\widehat{\eta}}_1(x_1) - \mathbb{E}(\widehat{\widehat{\eta}}_1(x_1))] + [\mathbb{E}(\widehat{\widehat{\eta}}_1(x_1)) - \eta_1(x_1)] \\ &=: T_1(x_1) + T_2(x_1) + T_3(x_1). \end{aligned} \tag{4.31}$$

Consider the first term $T_1(x_1)$. Under condition (Q.2), we infer that

$$\begin{aligned}\Big|T_1(x_1)\Big| &= \Big|\widehat{\widehat{\eta}}_1(x_1) - \widehat{\eta}_1(x_1)\Big| \\ &= \left|\int_{\mathbb{R}^{d-1}} \Big[\widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x})\Big]\, q_{-1}(\mathbf{x}_{-1})d\mathbf{x}_{-1} - \int_{\mathbb{R}^d} \Big[\widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x})\Big]\, q(\mathbf{x})d\mathbf{x}\right| \\ &\leq \int_{\mathbb{R}^{d-1}} \sup_{\mathbf{x}\in I}\Big|\widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x})\Big|\, q_{-1}(\mathbf{x}_{-1})d\mathbf{x}_{-1} \\ &\quad + \int_{\mathbb{R}^d} \sup_{\mathbf{x}\in I}\Big|\widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x})\Big|\, q(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{4.32}$$

Recall the following elementary observation

$$\frac{1}{f_n} = \frac{1}{f_{\mathbf{X}}} + \frac{(f_{\mathbf{X}} - f_n)}{f_{\mathbf{X}} f_n}.$$

From the last equation we can infer the following

$$\widehat{m}_n(\mathbf{x}) = \widehat{\widehat{m}}_n(\mathbf{x}) + \frac{1}{nh_n^d}\sum_{i=1}^n (\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta}) K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \frac{f_{\mathbf{X}}(\mathbf{X}_i) - f_n(\mathbf{X}_i)}{f_{\mathbf{X}}(\mathbf{X}_i) f_n(\mathbf{X}_i)}.$$

This, in turn, implies that

$$\begin{aligned}&\Big|\widehat{m}_n(\mathbf{x}) - \widehat{\widehat{m}}_n(\mathbf{x})\Big| \\ &\leq \frac{1}{nh_n^d}\sum_{i=1}^n |(\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta})| K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \times \\ &\quad \frac{\max_{1\leq i\leq n} |f_{\mathbf{X}}(\mathbf{X}_i) - f_n(\mathbf{X}_i)|}{|f_{\mathbf{X}}(\mathbf{X}_i) f_n(\mathbf{X}_i)|}. \end{aligned} \tag{4.33}$$

According to Ango-Nze and Rios (2000) or Stute (1984), under conditions (G.2-3), (H.2) and (K.1), it holds, as $n \to \infty$, that

$$\sup_{\mathbf{x}\in I^d} |f_n(\mathbf{x}) - f(\mathbf{x})| = \mathcal{O}\left(\sqrt{\frac{\log n}{na_n^d}}\right) \quad a.s. \tag{4.34}$$

Since the kernels K_ℓ, for $1 \leq \ell \leq d$, are compactly supported, making use of Bochner Lemma [see, e.g., Theorem A1 of Parzen (1962)], it follows, for each $1 \leq j \leq n$ and $1 \leq \ell \leq d$, that

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{1}{h_n} K_\ell \left(\frac{x_\ell - X_{j,\ell}}{h_n} \right) q_\ell(x_\ell) dx_\ell = q_\ell(X_{j,\ell}).$$

Similar arguments yield also

$$\lim_{n\to\infty} \int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K_{-1} \left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,j}}{h_n} \right) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} = q_{-1}(\mathbf{X}_{-1,j}), \tag{4.35}$$

and

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \frac{1}{h_n^d} K \left(\frac{\mathbf{x} - \mathbf{X}_j}{h_n} \right) q(\mathbf{x}) d\mathbf{x} = q(\mathbf{X}_j). \tag{4.36}$$

Finally, combining the statements (4.33)-(4.36) and making use of assumptions (M.1)-(M.2) and (Q.1), we obtain readily that

$$\int_{\mathbb{R}^{d-1}} \sup_{\mathbf{x}\in I^d} \left| \widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x}) \right| q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} = \mathcal{O}\left(\sqrt{\frac{\log n}{n a_n^d}} \right) \quad a.s., \tag{4.37}$$

and

$$\int_{\mathbb{R}^d} \sup_{\mathbf{x}\in I^d} \left| \widehat{\widehat{m}}_n(\mathbf{x}) - \widehat{m}_n(\mathbf{x}) \right| q(\mathbf{x}) d\mathbf{x} = \mathcal{O}\left(\sqrt{\frac{\log n}{n a_n^d}} \right) \quad a.s. \tag{4.38}$$

Hence, one concludes, from (4.32), (4.37) in connection with (4.38), that

$$\sup_{x_1\in I_1} \left| \widehat{\widehat{\eta}}_1(x_1) - \widehat{\eta}_1(x_1) \right| = \mathcal{O}\left(\sqrt{\frac{\log n}{n a_n^d}} \right) \quad a.s. \tag{4.39}$$

By using the last equation in combination with the assumptions (H.1) and (H.2), we obtain

$$\sqrt{\frac{n h_n}{2\log(1/h_n)}} \sup_{x_1\in I_1} |T_1(x_1)| = o(1) \quad a.s. \tag{4.40}$$

Now, we investigate the behavior of the term $T_3(x_1)$ in (4.31). Recall that

$$T_3(x_1) = \mathbb{E}(\widehat{\widehat{\eta}}_1(x_1)) - \eta_1(x_1).$$

By Fubini's Theorem, it is plain that

$$\begin{aligned} &\left| \mathbb{E}(\widehat{\widehat{\eta}}_1(x_1)) - \eta_1(x_1) \right| \\ &= \left| \int_{\mathbb{R}^{d-1}} (m(\mathbf{x}) - \mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))) q_{-\ell}(\mathbf{x}_{-\mathbf{l}}) d\mathbf{x}_{-\mathbf{l}} - \int_{\mathbb{R}^d} (m(\mathbf{x}) \right. \\ &\quad \left. - \mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))) q(\mathbf{x}) d\mathbf{x} \right|. \end{aligned} \tag{4.41}$$

Note that

$$\begin{aligned}\left|m(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right| &= \left|m(\mathbf{x})-\mathbb{E}\left(\sum_{i=1}^{n}\frac{\psi(\mathbf{Y}_i)-\mathbf{Z}_i^{\top}\boldsymbol{\beta}}{nf_{\mathbf{X}}(\mathbf{X}_i)}\frac{1}{h_n^d}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right)\right| \\ &= \left|m(\mathbf{x})-\int_{\mathbb{R}^d}m(\mathbf{u})\frac{1}{h_n^d}K\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)d\mathbf{u}\right|.\end{aligned}$$

By a Taylor series expansion of order k and a change of variables in connection with a straightforward application of Lebesgue dominated convergence theorem, for $\mathbf{x}\in I^d$ and $0<\theta<1$, we have

$$\begin{aligned}&\left|m(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right| \\ &= \left|m(\mathbf{x})-\int_{\mathbb{R}^d}m(\mathbf{v}\mathbf{h}\theta+\mathbf{x})K(\mathbf{v})\,d\mathbf{v}\right| \\ &\leq \int_{\mathbb{R}^d}|m(\mathbf{x})-m(\mathbf{v}\mathbf{h}\theta+\mathbf{x})|\,K(\mathbf{v})\,d\mathbf{v} \\ &= \frac{1}{k!}\int_{\mathbb{R}^d}\sum_{k_1+\cdots+k_d=k}|h_n^{k_1}v_1^{k_1}\ldots h_n^{k_d}v_d^{k_d}|\left|\frac{\partial^k m(\mathbf{v}\mathbf{h}\theta+\mathbf{x})}{\partial v_1^{k_1}\ldots\partial v_d^{k_d}}\right|K(\mathbf{v})\,d\mathbf{v} \\ &\leq \frac{1}{k!}\sup_{\mathbf{x}\in I^d}\left|\frac{\partial^k m(\mathbf{x})}{\partial v_1^{k_1}\ldots\partial v_d^{k_d}}\right|\sum_{k_1+\cdots+k_d=k}h_n^{k_1}\ldots h_n^{k_d}\int_{\mathbb{R}^d}|v_1^{k_1}\ldots v_d^{k_d}|K(\mathbf{v})d\mathbf{v}. \quad (4.42)\end{aligned}$$

Keeping in mind the statement (4.41), it follows that

$$\begin{aligned}&\sup_{x_1\in I_1}|T_3(x_1)| \\ &\leq 2\sup_{\mathbf{x}\in I^d}|m(\mathbf{x})-\mathbb{E}(\widehat{m}_n(\mathbf{x}))| \\ &\leq \frac{2}{k!}\sup_{\mathbf{x}\in I^d}\left|\frac{\partial^k m(\mathbf{x})}{\partial v_1^{k_1}\ldots\partial v_d^{k_d}}\right|\sum_{k_1+\cdots+k_d=k}h_n^{k}\int_{\mathbb{R}^d}|v_1^{k_1}\ldots v_d^{k_d}|K(\mathbf{v})d\mathbf{v}. \quad (4.43)\end{aligned}$$

Now, under conditions (H.1), (G.1) and (K.1), we infer that

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}}\sup_{x_1\in I_1}|T_3(x_1)|=o(1)\quad a.s. \quad (4.44)$$

For $n\geq 1$, we consider the process defined by

$$\sqrt{n}\zeta_n(x_1)=nh_n\int_{\mathbb{R}^{d-1}}\left(\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right)q_{-1}(\mathbf{x}_{-1})d\mathbf{x}_{-1}. \quad (4.45)$$

Making use of the Fubini's Theorem, we obtain readily that

$$\begin{aligned} nh_n\left(\widehat{\widehat{\eta}}_1(x_1)-\mathbb{E}(\widehat{\widehat{\eta}}_1(x_1))\right) &= nh_n\int_{\mathbb{R}^{d-1}}\left(\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right)q_{-1}(\mathbf{x}_{-1})d\mathbf{x}_{-1} \\ &\quad -nh_n\int_{\mathbb{R}^d}\left(\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right)q(\mathbf{x})d\mathbf{x} \\ &= \sqrt{n}\zeta_n(x_1)-\int_{\mathbb{R}}\sqrt{n}\zeta_n(x_1)q_1(x_1)dx_1. \end{aligned} \tag{4.46}$$

We have, therefore, by (4.46),

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}}T_2(x_1)=\frac{\zeta_n(x_1)}{\sqrt{2h_n\log(1/h_n)}}-\int_{\mathbb{R}}\frac{\zeta_n(x_1)}{\sqrt{2h_n\log(1/h_n)}}q_1(x_1)dx_1. \tag{4.47}$$

Recall that $f_1=f_{X_1}$ is the marginal density of $\mathbf{X}$ associated to the first component X_1 of $\mathbf{X}$. Set

$$\widehat{\zeta}^{\boldsymbol{\beta}}_{n,1}(x_1)=\frac{1}{nh_n}\sum_{i=1}^{n}\frac{\widetilde{\widetilde{Y}}_{i,n}}{f_1(X_{1,i})}K_1\left(\frac{x_1-X_{1,i}}{h_n}\right),$$

and

$$\widetilde{\widetilde{Y}}_{i,n}=(\psi(\mathbf{Y}_i)-\mathbf{Z}_i^\top\boldsymbol{\beta})\int_{\mathbb{R}^{d-1}}\frac{1}{h_n^{d-1}}K_{-1}\left(\frac{\mathbf{x}_{-1}-\mathbf{X}_{-1,i}}{h_n}\right)\frac{q_{-1}(\mathbf{x}_{-\mathbf{1}})}{f(\mathbf{X}_{-1,i}|X_{1,i})}d\mathbf{x}_{-1},$$

where $\mathbf{X}_i=(X_{1,i},\ldots,X_{i,d})$ and $\mathbf{X}_{i,-1}=(X_{i,2},\ldots,X_{i,d})$. For the sake of notation simplicity, $\widehat{\zeta}^{\boldsymbol{\beta}}_{n,1}(x_1)$ will be denoted by $\widehat{\zeta}_{n,1}(x_1)$. Recalling the definition (4.45) of ζ_n, then

$$\begin{aligned} &\int_{\mathbb{R}}\frac{|\zeta_n(x_1)|}{\sqrt{2h_n\log(1/h_n)}}q_1(x_1)dx_1 \\ &= \sqrt{\frac{nh_n}{2\log(1/h_n)}}\left|\int_{\mathbb{R}^d}\left(\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right)q(\mathbf{x})d\mathbf{x}\right| \\ &= \sqrt{\frac{nh_n}{2\log(1/h_n)}}\left|\int_{\mathbb{R}}\left(\widehat{\zeta}_{n,1}(x_1)-\mathbb{E}(\widehat{\zeta}_{n,1}(x_1))\right)q_1(x_1)dx_1\right|. \end{aligned} \tag{4.48}$$

Considering the Cauchy-Schwartz inequality combined with the condition (Q.2), it follows that

$$\begin{aligned} \int_{\mathbb{R}}\frac{|\zeta_n(x_1)|}{\sqrt{2h_n\log(1/h_n)}}q_1(x_1)dx_1 &\leq \sqrt{\frac{nh_n}{2\log(1/h_n)}}\left[\int_{\mathscr{C}_1}\left(\widehat{\zeta}_{n,1}(x_1)-\mathbb{E}(\widehat{\zeta}_{n,1}(x_1))\right)^2dx_1\right]^{1/2} \\ &\quad\times\left[\int_{\mathscr{C}_1}q_1^2(x_1)dx_1\right]^{1/2}, \end{aligned} \tag{4.49}$$

where $\mathscr{C}_1$ is defined in (Q.2). Making use of Camlong-Viot et al. (2000)'s results, we have

$$\mathrm{Var}(\widehat{\zeta}_{n,1}(x_1)) = \mathbb{E}\left(\widehat{\zeta}_{n,1}(x_1) - \mathbb{E}(\widehat{\zeta}_{n,1}(x_1))\right)^2 = \mathcal{O}\left(1/nh_n\right),$$

which, combined with Fubini's Theorem, implies that

$$\begin{aligned}\int_{\mathscr{C}_1} \mathrm{Var}(\widehat{\zeta}_{n,1}(x_1))dx_1 &= \int_{\mathscr{C}_1} \mathbb{E}\left(\widehat{\zeta}_{n,1}(x_1) - \mathbb{E}(\widehat{\zeta}_{n,1}(x_1))\right)^2 dx_1 \\ &= \mathbb{E}\left(\int_{\mathscr{C}_1} \left(\widehat{\zeta}_{n,1}(x_1) - \mathbb{E}(\widehat{\zeta}_{n,1}(x_1))\right)^2 dx_1\right) \\ &= \mathcal{O}\left(1/nh_n\right). \end{aligned} \tag{4.50}$$

Combining now the statements (4.49) and (4.50), we conclude that

$$\int_{\mathbb{R}} \frac{|\zeta_n(x_1)|}{\sqrt{2h_n \log(1/h_n)}} q_1(x_1)dx_1 = \mathcal{O}\left(\sqrt{\frac{1}{\log(1/h_n)}}\right) \quad a.s. \tag{4.51}$$

The condition (H.1) on the smoothing parameter h_n gives

$$\int_{\mathbb{R}} \frac{|\zeta_n(x_1)|}{\sqrt{2h_n \log(1/h_n)}} q_1(x_1)dx_1 = o(1) \quad a.s., \tag{4.52}$$

which, by the statement (4.47), implies that

$$\begin{aligned}&\sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{x_1\in I_1} \left|T_2(x_1)\right| \\ &= \sup_{x_1\in I_1} \left| \frac{\zeta_n(x_1)}{\sqrt{2h_n \log(1/h_n)}} - \int_{\mathbb{R}} \frac{\zeta_n(x_1)}{\sqrt{2h_n \log(1/h_n)}} q_1(x_1)dx_1 \right| \\ &= \sup_{x_1\in I_1} \frac{|\zeta_n(x_1)|}{\sqrt{2h_n \log(1/h_n)}} + o(1) \quad a.s. \end{aligned} \tag{4.53}$$

Making use of the Proposition 4.1 given below, we obtain

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{x_1\in I_1} |T_2(x_1)| = \sigma_1 \quad a.s. \tag{4.54}$$

Thus, combining the statements (4.40), (4.44) and (4.54), we conclude that

$$\lim_{n\to\infty} \sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{x_1\in I_1} |\mathcal{D}_2(x_1)| = \sigma_1 \quad a.s. \tag{4.55}$$

Part (II). Consider now the first term $\mathcal{D}_1$ of the statement (4.30) and observe that

$$\begin{aligned}
\mathcal{D}_1(x_1) &= \widehat{\eta}_1^{\widehat{\boldsymbol{\beta}}}(x_1) - \widehat{\eta}_1^{\boldsymbol{\beta}}(x_1) \\
&= \int_{\mathbb{R}^{d-1}} \left(\widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) - \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\right) q_{-1}(\mathbf{x_{-1}}) d\mathbf{x_{-1}} - \int_{\mathbb{R}^d} \left(\widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) - \widehat{m}_n^{\boldsymbol{\beta}}(\mathbf{x})\right) q(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^{d-1}} \sum_{i=1}^{n} \frac{\mathbf{Z}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{n h_n^d f_n(\mathbf{X}_i)} K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right) K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right) q_{-1}(\mathbf{x_{-1}}) d\mathbf{x_{-1}} \\
&\quad - \Big(\frac{d-1}{d}\Big) \int_{\mathbb{R}^d} \sum_{i=1}^{n} \frac{\mathbf{Z}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{n h_n^d f_n(\mathbf{X}_i)} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) q(\mathbf{x}) d\mathbf{x} \\
&= \sum_{i=1}^{n} \mathbf{Z}_i^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \frac{1}{n f_n(\mathbf{X}_i)} \times \\
&\quad \left[\frac{1}{h_n} K_1\Big(\frac{\mathbf{x}_1 - X_{1,i}}{h_n}\Big) \int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right) q_{-l}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}\right. \\
&\quad \left. - \Big(\frac{d-1}{d}\Big) \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) q(\mathbf{x}) d\mathbf{x}\right]. \qquad (4.56)
\end{aligned}$$

For $1 \leq i \leq n$, set

$$\begin{aligned}
W_{ni}^1(\mathbf{x}) &= \frac{1}{n f_n(\mathbf{X}_i)} \left[\frac{1}{h_n} K_1\Big(\frac{\mathbf{x}_1 - X_{1,i}}{h_n}\Big) \int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right) q_{-l}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}\right. \\
&\quad \left. - \Big(\frac{d-1}{d}\Big) \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) q(\mathbf{x}) d\mathbf{x}\right].
\end{aligned}$$

Consider first the case where the covariate density $f_{\mathbf{X}}$ is known. Towards this end, replace $W_{ni}^1(\mathbf{x})$ by

$$\begin{aligned}
W_{ni}^{1,f}(\mathbf{x}) &= \frac{1}{n f_{\mathbf{X}}(\mathbf{X}_i)} \left[\frac{1}{h_n} K_1\Big(\frac{\mathbf{x}_1 - X_{1,i}}{h_n}\Big) \int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right) q_{-l}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}\right. \\
&\quad \left. - \Big(\frac{d-1}{d}\Big) \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) q(\mathbf{x}) d\mathbf{x}\right].
\end{aligned}$$

Observe, for any $1 \leq \ell \leq d$, that one may write

$$f_{\mathbf{X}}(x_1, \ldots, x_d) = f_\ell(x_\ell) \times f(x_1, \ldots, x_{\ell-1}, x_{\ell+1}, \ldots, x_d | x_\ell).$$

Notice that, for any $x_1 \in I_1$, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{h_n f_1(X_{1,i})} K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right)\right] &= \int_{\mathbb{R}} \frac{1}{h_n f_1(y)} K_1\left(\frac{x_1 - y}{h_n}\right) f_1(y) dy \\
&= \int_{\mathbb{R}} \frac{1}{h_n} K_1\left(\frac{x_1 - y}{h_n}\right) dy = 1
\end{aligned}$$

Thus, we have

$$\lim_{n\to\infty}\frac{1}{nh_n}\sum_{i=1}^{n}\frac{1}{f_1(X_{1,i})}K_1\left(\frac{x_1-X_{1,i}}{h_n}\right)=1 \quad a.s.$$

Therefore, considering the condition (G.2) and combining the statements (4.35) and (4.36), we obtain

$$\begin{aligned}\sum_{i=1}^{n}\frac{1}{nf_{\mathbf{X}}(\mathbf{X}_i)}&\left[\frac{1}{h_n}K_1\left(\frac{x_1-X_{1,i}}{h_n}\right)q_{-1}(\mathbf{X}_{-1,i})-\left(\frac{d-1}{d}\right)q(\mathbf{X}_i)\right]\\&=\mathcal{O}(1),\quad a.s.\end{aligned}\tag{4.57}$$

Investigating the case where the covariate density function $f_{\mathbf{X}}$ is unknown and estimated by f_n, observe that we have the following decomposition

$$\begin{aligned}&\frac{1}{nf_n(\mathbf{X}_i)}\left(\frac{1}{h_n}K_1\Big(\frac{x_1-X_{1,i}}{h_n}\Big)q_{-1}(\mathbf{X}_{-1,i})-\Big(\frac{d-1}{d}\Big)q(\mathbf{X}_i)\right)\\&=\frac{1}{nf_{\mathbf{X}}(\mathbf{X}_i)}\left(\frac{1}{h_n}K_1\Big(\frac{x_1-X_{1,i}}{h_n}\Big)q_{-1}(\mathbf{X}_{-1,i})-\Big(\frac{d-1}{d}\Big)q(\mathbf{X}_i)\right)\\&-\frac{f_n(\mathbf{X}_i)-f_{\mathbf{X}}(\mathbf{X}_i)}{nf_n(\mathbf{X}_i)f_{\mathbf{X}}(\mathbf{X}_i)}\times\\&\left(\frac{1}{h_n}K_1\Big(\frac{x_1-X_{1,i}}{h_n}\Big)q_{-1}(\mathbf{X}_{-1,i})-\Big(\frac{d-1}{d}\Big)q(\mathbf{X}_i)\right).\end{aligned}\tag{4.58}$$

Thus it follows, from condition (G.2), the statements (4.34), (4.57) and (4.58), that

$$\begin{aligned}&\sum_{i=1}^{n}\frac{1}{nf_n(\mathbf{X}_i)}\left(\frac{1}{h_n}K_1\Big(\frac{x_1-X_{1,i}}{h_n}\Big)q_{-1}(\mathbf{X}_{-1,i})-\Big(\frac{d-1}{d}\Big)q(\mathbf{X}_i)\right)\\&=\sum_{i=1}^{n}\frac{1}{nf(\mathbf{X}_i)}\left(\frac{1}{h_n}K_1\Big(\frac{x_1-X_{1,i}}{h_n}\Big)q_{-1}(\mathbf{X}_{-1,i})-\Big(\frac{d-1}{d}\Big)q(\mathbf{X}_i)\right)+\mathcal{O}\left(\sqrt{\frac{\log n}{na_n^d}}\right)\\&=\mathcal{O}(1)+\mathcal{O}\left(\sqrt{\frac{\log n}{na_n^d}}\right)\ a.s.\end{aligned}\tag{4.59}$$

Notice that it is stated in Chokri and Louani (2011), for $1\leq\ell\leq p$, that

$$\limsup_{n\to\infty}\sqrt{\frac{n}{2\log\log n}}\left|\widehat{\beta}_\ell-\beta_\ell\right|=(\sigma_\varepsilon^2 b^{\ell\ell})^{1/2}\quad a.s.,\tag{4.60}$$

where we recall that σ_ε^2 is the variance of ε. It is easy to see that the condition (H.1) implies that

$$\frac{\log\log n}{\log(1/h_n)}\to 0,\quad \text{as}\quad n\to\infty.\tag{4.61}$$

Therefore, making use of the condition (4.61), it follows, for any $1 \leq \ell \leq p$, that

$$\limsup_{n\to\infty} \sqrt{\frac{nh_n}{\log(1/h_n)}} \left|\widehat{\beta}_\ell - \beta_\ell\right| = 0 \quad a.s. \tag{4.62}$$

Using now the condition (M.2) and the statements (4.56), (4.59) and (4.62), we conclude that

$$\sup_{x_1 \in I_1} \mathcal{D}_1(x_1) = o\left(\sqrt{\frac{\log(1/h_n)}{nh_n}}\right) \quad a.s. \tag{4.63}$$

It suffices then to combine the statements (4.30), (4.55) and (4.63) to achieve the proof. □

Proof of Theorem 4.2.

Keep in mind the following definitions

$$\widehat{m}_{add}(\mathbf{x}) = \sum_{\ell=1}^{d} \widehat{\eta}_\ell(x_\ell) + \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{x}) q(\mathbf{x}) d\mathbf{x},$$

and

$$m_{add}(\mathbf{x}) = \sum_{\ell=1}^{d} \eta_\ell(x_\ell) + \int_{\mathbb{R}^d} m_{add}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}.$$

Considering the statements (4.17) and (4.6), it is obvious that

$$\begin{aligned}
&\sqrt{\frac{nh_n}{\log(1/h_n)}} \left|\widehat{m}_{add}(\mathbf{x}) - m_{add}(\mathbf{x}) - \sum_{\ell=1}^{d} \sigma_\ell\right| \\
&= \sqrt{\frac{nh_n}{\log(1/h_n)}} \left| \sum_{\ell=1}^{d} (\widehat{\eta}_\ell(x_\ell) - \eta_\ell(x_\ell) - \sigma_\ell) \right. \\
&\quad \left. + \int_{\mathbb{R}^d} (\widehat{m}_n(\mathbf{x}) - m_{add}(\mathbf{x}))\, q(\mathbf{x}) d\mathbf{x} \right| \\
&\leq \sqrt{\frac{nh_n}{\log(1/h_n)}} \sum_{\ell=1}^{d} \left|\widehat{\eta}_\ell(x_\ell) - \eta_\ell(x_\ell) - \sigma_\ell\right| \\
&\quad + \sqrt{\frac{nh_n}{\log(1/h_n)}} \int_{\mathbb{R}^d} \left|\widehat{m}_n(\mathbf{x}) - m_{add}(\mathbf{x})\right| q(\mathbf{x}) d\mathbf{x},
\end{aligned}$$

which, in turn, implies that

$$\begin{aligned}
&\sqrt{\frac{nh_n}{\log(1/h_n)}}\sup_{\mathbf{x}\in I^d}\left|\widehat{m}_{add}(\mathbf{x})-m_{add}(\mathbf{x})-\sum_{\ell=1}^{d}\sigma_\ell\right| \\
&\leq \sum_{\ell=1}^{d}\sqrt{\frac{nh_n}{\log(1/h_n)}}\sup_{x\in I_\ell}\left|\widehat{\eta}_\ell(x_\ell)-\eta_\ell(x_\ell)-\sigma_\ell\right| \\
&\quad +\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\left|\widehat{m}_n(\mathbf{x})-m_{add}(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x}. \qquad (4.64)
\end{aligned}$$

Recall that by Theorem 4.1, we have

$$\sum_{\ell=1}^{d}\sqrt{\frac{nh_n}{\log(1/h_n)}}\sup_{x\in I_\ell}\left|\widehat{\eta}_\ell(x_\ell)-\eta_\ell(x_\ell)-\sigma_\ell\right|=o(1)\quad a.s. \qquad (4.65)$$

For the second term of equation (4.64), observe that

$$\begin{aligned}
&\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\left|\widehat{m}_n(\mathbf{x})-m_{add}(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x} \\
&\leq \sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\left|\widehat{m}(\mathbf{x})-\widehat{m}_n(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x} \\
&\quad +\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\left|\widehat{m}_n(\mathbf{x})-m_{add}(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x}. \qquad (4.66)
\end{aligned}$$

Recall that

$$\widehat{m}_n(\mathbf{x})=\sum_{i=1}^{n}\frac{\psi(\mathbf{Y}_i)-\mathbf{Z}_i^\top\widehat{\boldsymbol{\beta}}}{nh_n^d f_n(\mathbf{X}_i)}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right).$$

Making use of conditions (M.1)-(M.2), there exists a positive constant Λ such that

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left|\widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})-\widehat{m}_n(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x} \\
&= \int_{\mathbb{R}^d}\left|\sum_{i=1}^{n}\frac{\mathbf{Z}_i^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{nh_n^d f_n(\mathbf{X}_i)}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right|q(\mathbf{x})d\mathbf{x} \\
&\leq \int_{\mathbb{R}^d}\{\max_{1\leq i\leq n}\|\mathbf{Z}_i^\top\|\}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|\left|\sum_{i=1}^{n}\frac{1}{nh_n^d f_n(\mathbf{X}_i)}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right|q(\mathbf{x})d\mathbf{x} \\
&\leq \Lambda\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|. \qquad (4.67)
\end{aligned}$$

We infer, in turn, from the last inequality and (4.60), that

$$\limsup_{n\to\infty}\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\left|\widehat{m}_n^{\widehat{\boldsymbol{\beta}}}(\mathbf{x})-\widehat{m}_n(\mathbf{x})\right|q(\mathbf{x})d\mathbf{x}=0\quad a.s. \qquad (4.68)$$

It is straightforward to see that

$$\begin{aligned}\int_{\mathbb{R}^d}\Big|\widehat{m}_n(\mathbf{x})-m_{add}(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x} \;\leq\; & \int_{\mathbb{R}^d}\Big|\widehat{m}_n(\mathbf{x})-\widehat{\widehat{m}}_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}\\ & +\int_{\mathbb{R}^d}\Big|\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}\widehat{\widehat{m}}_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}\\ & +\int_{\mathbb{R}^d}\Big|\mathbb{E}\widehat{\widehat{m}}_n(\mathbf{x})-m_{add}(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}.\end{aligned}$$

Under (H.2) and by using equation (4.38), we have

$$\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\Big|\widehat{m}_n(\mathbf{x})-\widehat{\widehat{m}}_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}=o(1)\quad a.s.$$

By combining (4.48) and (4.52), we get

$$\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\Big|\widehat{\widehat{m}}_n(\mathbf{x})-\mathbb{E}\widehat{\widehat{m}}_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}=o(1)\quad a.s.$$

Notice that the condition (H.1) implies that

$$\frac{nh_n^{2k+1}}{\log(1/h_n)}\to 0,\quad \text{as}\quad n\to\infty$$

which permits to infer from equation (4.42) that

$$\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\Big|\mathbb{E}\widehat{\widehat{m}}_n(\mathbf{x})-m_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}=o(1)\quad a.s.,$$

therefore

$$\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\Big|\widehat{m}_n(\mathbf{x})-m_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}=o(1)\quad a.s. \tag{4.69}$$

By (4.66) in connection with (4.67) and (4.69), we conclude that

$$\sqrt{\frac{nh_n}{\log(1/h_n)}}\int_{\mathbb{R}^d}\Big|\widehat{m}_n(\mathbf{x})-m_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}=o(1). \tag{4.70}$$

Finally, combining (4.64), (4.65) with (4.70) to achieve the proof of Theorem 4.2. □

Proof of Corollary 4.1.

We have, for all $(\mathbf{x}, \mathbf{z}) \in I^d \times \mathbb{R}^p$,

$$\begin{aligned} |\widehat{m}(\mathbf{x}, \mathbf{z}) - m(\mathbf{x}, \mathbf{z})| &= \left|\mathbf{z}^\top \widehat{\boldsymbol{\beta}} + \widehat{m}_{add}(\mathbf{x}) - \mathbf{z}^\top \boldsymbol{\beta} - m_{add}(\mathbf{x})\right| \\ &= |\mathscr{T}_1 + \mathscr{T}_2| \\ &\leq |\mathscr{T}_1| + |\mathscr{T}_2|, \end{aligned}$$

where, $T_1 = \mathbf{z}^\top \widehat{\boldsymbol{\beta}} - \mathbf{z}^\top \boldsymbol{\beta}$ and $\mathscr{T}_2 = \widehat{m}_{add}^{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{add}(\mathbf{x})$. According to the result of Chokri and Louani (2011) and using conditions (M.1)-(M.2), we show that

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}} |\mathbf{z}^\top \widehat{\boldsymbol{\beta}} - \mathbf{z}^\top \boldsymbol{\beta}| = \sigma_\varepsilon \sum_{\ell=1}^{p} (b^{\ell\ell})^{1/2} + o(1) \quad a.s. \tag{4.71}$$

By combining the last equation and Theorem 4.2, we obtain the desired result. □

4.6 Appendix

As indicated in the introduction, a part of our proof will be based on the results of Deheuvels and Mason (2004) and Einmahl and Mason (2000, 2005). Note that the notations of this section are similar to that used in Deheuvels and Mason (2004) and changes have been made in order to adopt it to our setting. We now impose some slightly more general assumptions on the kernel $K(\cdot)$ than that of our theorems. Here we gather together some basic facts that we need for the proofs. Let $(\mathscr{X}, \mathscr{A})$ be a measurable space. Throughout this section, we assume that on the basic probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we have independent $(\mathscr{F}, \mathscr{A})$-measurable variables $\mathbf{X}_i : \Omega \to \mathscr{X}$, $1 \leq i \leq n$, with common distribution function $\mathbb{F}$.

(C.1) Let $\mathscr{G}$ be a pointwise measurable class of functions from $\mathscr{X}$ to $\mathbb{R}$, that is, there exists a countable subclass $\mathscr{G}_0$ of $\mathscr{G}$ such that we can find for any function $g \in \mathscr{G}$ a sequence of functions $\{g_m : m \geq 1\}$ in $\mathscr{G}_0$ for which

$$g_m \longrightarrow g.$$

This measurability assumption is imposed to avoid using outer probability measures in all of statements which is discussed in van der Vaart and Wellner (1996). Consider the class of functions

$$\mathcal{K} := \Big\{K((\mathbf{x} - \cdot)/h) : h > 0,\ \mathbf{x} \in \mathbb{R}^d\Big\}.$$

The class of functions $\mathcal{K}$ fulfills the condition (C.1) whenever $\mathbb{K}$ is right continuous, that is,

(K.4) K is right continuous on $\mathbb{R}^d$, i.e., for any $\mathbf{t} = (t_1, \ldots, t_d)$, we have

$$K(t_1, \ldots, t_d) = \lim_{\varepsilon_1 \downarrow 0, \ldots, \varepsilon_d \downarrow 0} K(t_1 + \varepsilon_1, \ldots, t_d + \varepsilon_d).$$

which is implied by our conditions (refer to the papers by Deheuvels and Mason (2004) and Einmahl and Mason (2000, 2005)) .

(C.2) Let $\mathbb{G}$ be a finite-measurable function fulfilling for all $x \in \mathscr{X}$,

$$\mathbb{G}(\mathbf{x}) \geq \sup_{g \in \mathscr{G}} |g(\mathbf{x})|.$$

For $\varepsilon > 0$, set

$$N(\varepsilon, \mathscr{G}) = \sup_{Q} N(\varepsilon \sqrt{Q(\mathbb{G}^2)}, \mathscr{G}, d_Q),$$

where the supremum is taken over all probability measures Q on $(\mathscr{X}, \mathscr{A})$. Here, d_Q denotes the $L_2(Q)$-metric and $N(\varepsilon \sqrt{Q(\mathbb{G}^2)}, \mathscr{G}, d_Q)$ is the minimal number of balls $\{g : d_Q(g, g') < \varepsilon\}$ of d_Q-radius ε needed to cover $\mathscr{G}$. We assume that $\mathscr{G}$ satisfies the following uniform entropy condition.

(C.3) For some $C > 0$ and $\nu > 0$,

$$N(\varepsilon, \mathscr{G}) \leq C\varepsilon^{-\nu}, 0 < \varepsilon < 1. \tag{4.72}$$

Observe that the last condition is satisfied whenever (K.3) holds true, i.e., K is of bounded variation on $\mathbb{R}^d$ (in the sense of Hardy and Kauser, see, e.g. Clarkson and Adams (1933), Vituškin (1955) and Hobson (1958)), meaning that $K(d\mathbf{t})$ defines a totally bounded Lebesgue-Stieltjes signed measure on $\mathbb{R}^d$, we may refer to Deheuvels (2011) or (Bouzebda, 2012, pp. 60-61) for more references on the preceding conditions. In the sequel the conditions (C.1-3) will be denoted by $(\mathscr{E})$.
Let α_n be the multivariate empirical processus based upon $(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i)_{i \geq 1}$ defined, for $g \in \mathscr{G}$, by

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(g(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i) - \mathbb{E}[g(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i)] \right),$$

and set for any class $\mathscr{G}$ of such functions g

$$\left\| \sqrt{n}\alpha_n \right\|_{\mathscr{G}} = \sup_{g \in \mathscr{G}} \left| \sqrt{n}\alpha(g) \right|.$$

For any real valued function ϑ defined on a real set B, consider the notation

$$||\vartheta||_B = \sup_{x \in B} |\vartheta(x)| := ||\vartheta||.$$

Denote by $\lfloor u \rfloor$ the integer part of u. Recalling that $I_1 = [a_1, c_1]$, let $0 < \gamma < 1$ be a fixed number and set, for $n \geq 1$,

$$x_{1,j} = a_1 + j\gamma h_n,\ 0 \leq j \leq \left\lfloor \frac{c_1 - a_1}{\gamma h_n} \right\rfloor := l_n. \tag{4.73}$$

For $\mathbf{X}_i = (X_{1,i}, \ldots, X_{i,d}),\ 1 \leq i \leq n$, we introduce the following quantity

$$g_n^{x_1,\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i) = (\psi(\mathbf{Y}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta}) G(\mathbf{X}_i) K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right), \tag{4.74}$$

where

$$G(\mathbf{X}_i) = \frac{1}{h_n^{d-1} f(\mathbf{X}_i)} \int_{\mathbb{R}^{d-1}} K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{X}_{-1,i}}{h_n}\right) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}. \tag{4.75}$$

Set, for $n \geq 1$, $0 \leq j \leq l_n$, and fixed $\psi \in \mathcal{F}$,

$$\mathscr{G}_n = \left\{ g_n^{x_{1,j},\boldsymbol{\beta}} : 0 \leq j \leq l_n, \boldsymbol{\beta} \in \mathbb{R}^p \right\}.$$

Obviously, for each $0 \leq j \leq l_n$, $x_1 \in I_1$, and any $\boldsymbol{\beta} \in \mathbb{R}^p$, we have

$$||g_n^{x_{1,j},\boldsymbol{\beta}}|| + ||g_n^{x_1,\boldsymbol{\beta}}|| \leq 2\kappa,$$

where κ is a positive constant. Recall the definition of $\zeta_n(\cdot)$ in (4.45), and keep in mind that

$$\begin{aligned}
\zeta_n(x_1) &= nh_n \int_{\mathbb{R}^{d-1}} \left(\widehat{\widehat{m}}_n(\mathbf{x}) - \mathbb{E}(\widehat{\widehat{m}}_n(\mathbf{x}))\right) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} \\
&= \sum_{i=1}^{n} \left(g_n^{x_1,\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i) - \mathbb{E}[g_n^{x_1,\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i)]\right) \\
&= \sqrt{n}\alpha_n(g_n^{x_1,\boldsymbol{\beta}}).
\end{aligned}$$

Furthermore, let $\varepsilon_1, \ldots, \varepsilon_n$ be a sequence of independent Rademacher random variables, independent of $\mathbf{X}_1, \ldots, \mathbf{X}_n$. As said before, to prove Proposition 4.1 below, we shall require some standard fact, which we recall for easy reference and completeness. We need the following inequality, which is essentially due to Talagrand (1994) [see also Ledoux (9597)].

Fact 1. Let $\mathscr{G}$ be a pointwise measurable class of functions satisfying for some $0 < M < \infty$, such that

$$\|g\|_\infty \leq M, \quad g \in \mathscr{G}.$$

Then we have for all $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{ \max_{1 \leq m \leq n} \| \sqrt{m}\alpha_m \|_{\mathscr{G}} \geq A_1 \left(\mathbb{E} \left\| \sum_{i=i}^{n} \varepsilon_i g(\mathbf{X}_i) \right\|_{\mathscr{G}} + t \right) \right\} \\
&\qquad \leq 2 \left\{ \exp\left(-\frac{A_2 t^2}{n\sigma_{\mathscr{G}}^2}\right) + \exp\left(-\frac{A_2 t}{M}\right) \right\},
\end{aligned}$$

where

$$\sigma_{\mathscr{G}}^2 = \sup_{g\in\mathscr{G}} Var(g(\mathbf{X}))$$

and A_1, A_2 are universal constants.
The following fact is due to Einmahl and Mason (2000).

Fact 2. Let $\mathscr{G}$ be a pointwise measurable class of bounded functions fulfilling the assumptions above. In addition, for some constants $\tau > 0$, $\nu > 0$, $C_0 > 1$ and $0 < \sigma \leq 1/8C_0$ and a function $\mathbb{G}$ defined above, assume the following conditions

(A.1) $\mathbb{E}[\mathbb{G}^2(X)] \leq \tau^2$;

(A.2) $N(\varepsilon, \mathscr{G}) < C_0\varepsilon^{-\nu}, \quad 0 < \varepsilon < 1$;

(A.3) $\sigma_0^2 := \sup_{g\in\mathscr{G}} \mathbb{E}[g^2(X)] \leq \sigma^2$;

(A.4) $\sup_{g\in\mathscr{G}} \|g\|_\infty \leq (1/2\sqrt{\nu+1})\sqrt{n\sigma^2/\log(\tau\vee 1/\sigma)}$.

Then, for a universal constant $A_3 > 0$, we have

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(X_i)\right\|_{\mathscr{G}} \leq A_3\sqrt{\nu n\sigma^2 \log(\tau\vee 1/\sigma)}.$$

We need the following fact, given for example in the work (Pollard, 1984, Appendix B).

Fact 3. (Bernstein's inequality) Let $Z_1, \ldots, Z_n$ be independent random variables with common mean 0 and common finite variance $0 < \sigma^2 < \infty$. Assume further that for some constant $\mathcal{C} > 0$, $|Z_i| < \mathcal{C}$, $1 \leq i \leq n$. Then, for any $t > 0$, we have

$$\mathbb{P}(Z_1, \ldots, Z_n \geq t\sqrt{n}) \leq \exp\left(\frac{t^2}{2\sigma^2 + (2/3)\mathcal{C}n^{-1/2}t}\right).$$

The following proposition plays the central role in the proofs of the preceding section.

Proposition 4.1. Suppose that the assumptions (G.1)-(G.3), (H.1)-(H.2), (K.1)-(K.3), (M.1)-(M.2) and (Q.1)-(Q.2) are satisfied. Then, as $n \to \infty$, we have

$$\sup_{x_1\in I_1} \frac{|\alpha_n(g_n^{x_1,\boldsymbol{\beta}})|}{\sqrt{2h_n\log(1/h_n)}} \to \sigma_1 \quad a.s.$$

The proofs of Proposition 4.1 will be given in two steps, the upper bound, and then the lower bound as in Deheuvels and Mason (2004) and Einmahl and Mason (2000, 2005).

Proof of Proposition 4.1 : Upper bound part

Let σ_1 be as in the statement (4.22). In this part we will prove, for any $\varepsilon > 0$, that

$$\limsup_{n\to\infty} \sup_{x_1\in I_1} \frac{|\alpha_n(g_n^{x_1,\boldsymbol{\beta}})|}{\sqrt{2h_n \log(1/h_n)}} \leq \sigma_1(1+\varepsilon) \quad a.s. \tag{4.76}$$

The proof of (4.76) will be split up into two parts : the discretization and the oscillation.

• **Discretisation**

Investigating the behaviour of the process α_n on the grid defined in the statement (4.73), we have

$$\mathrm{Var}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right) = \mathbb{E}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right)^2 - \left[\mathbb{E}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right)\right]^2$$

Note that

$$\mathbb{E}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right) = \left(\mathbb{E}(\psi(\mathbf{Y}_i)-\mathbf{Z}_i^\top\boldsymbol{\beta})G(\mathbf{X}_i)K_1\left(\frac{x_{1,j}-X_{1,i}}{h_n}\right)\right).$$

Making use of conditions (M.1) and (M.2) combined with the statement (4.75) lead to

$$\mathbb{E}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right) = \mathcal{O}(h_n).$$

Therefore,

$$\begin{aligned}
&\mathrm{Var}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i,\psi(\mathbf{Y}_i),\mathbf{Z}_i)\right)\\
&= \mathbb{E}\left((\psi(\mathbf{Y}_i)-\mathbf{Z}_i^\top\boldsymbol{\beta})^2G^2(\mathbf{X}_i)K_1^2\left(\frac{x_{1,j}-X_{1,i}}{h_n}\right)\right)+o(h_n)\\
&= \mathbb{E}\left(\mathbb{E}(\psi(\mathbf{Y}_i)-\mathbf{Z}_i^\top\boldsymbol{\beta})^2|\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\left(\frac{x_{1,j}-X_{1,i}}{h_n}\right)\right)+o(h_n)\\
&= \mathbb{E}\left(H^{\boldsymbol{\beta}}(\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\left(\frac{x_{1,j}-X_{1,i}}{h_n}\right)\right)+o(h_n).
\end{aligned} \tag{4.77}$$

Observe now that

$$\begin{aligned}
&\mathbb{E}\left(H^{\boldsymbol{\beta}}(\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\left(\frac{x_{1,j}-X_{1,i}}{h_n}\right)\right)\\
&= \int_{\mathbb{R}^d} H^{\boldsymbol{\beta}}(\mathbf{u})\left(\frac{1}{h_n^{d-1}f(\mathbf{u})}\int_{\mathbb{R}^{d-1}}K_{-1}\left(\frac{\mathbf{x}_{-1}-\mathbf{u}_{-1}}{h_n}\right)q_{-1}(\mathbf{x}_{-1})d\mathbf{x}_{-1}\right)^2\times\\
&\quad\times K_1^2\left(\frac{x_{1,j}-u_1}{h_n}\right)f(\mathbf{u})d\mathbf{u}.
\end{aligned}$$

Let

$$\mathbf{v} = (v_1, \ldots, v_d) = \left(\frac{x_1 - u_1}{h_n}, \ldots, \frac{x_d - u_d}{h_n} \right).$$

By Taylor series expansion and using the assumptions (K.1) and (Q.2), we find

$$\begin{aligned}
&\int_{\mathbb{R}^{d-1}} \frac{1}{h_n^{d-1}} K_{-1}\left(\frac{\mathbf{x}_{-1} - \mathbf{u}_{-1}}{h_n} \right) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} \\
&= \int_{\mathbb{R}^{d-1}} K_{-1}(\mathbf{v}_{-1}) q_{-1}(\mathbf{v}_{-1}\mathbf{h}_{-1} + \mathbf{u}_{-1}) d\mathbf{v}_{-1} \\
&= \int_{\mathbb{R}^{d-1}} K_{-1}(\mathbf{v}_{-1}) \Big[q_{-1}(\mathbf{u}_{-1}) \\
&\quad + \sum_{k_2+\ldots+k_d=k} v_2^{k_2} \ldots v_d^{k_d} h_n^{k_2} \ldots h_n^{k_d} \frac{\partial^k q_{-1}}{\partial v_2^{k_1} \ldots \partial v_d^{k_d}} (\mathbf{v}_{-1}\mathbf{h}_{-1}\theta + \mathbf{u}_{-1}) \Big] d\mathbf{v}_{-1} \\
&= q_{-1}(\mathbf{u}_{-1}) + o(1), \qquad (4.78)
\end{aligned}$$

where $\mathbf{h}_{-1} = (\underbrace{h_n, \ldots, h_n}_{(d-1)\text{-times}})^\top$ and $0 < \theta < 1$. Then, we deduce

$$\begin{aligned}
&\mathbb{E}\left(H^{\boldsymbol{\beta}}(\mathbf{X}_i) G^2(\mathbf{X}_i) K_1^2 \left(\frac{x_{1,j} - X_{1,i}}{h_n} \right) \right) \\
&\quad = \int_{\mathbb{R}^d} \frac{H^{\boldsymbol{\beta}}(\mathbf{u})}{f(\mathbf{u})} q_{-1}^2(\mathbf{u}_{-1}) K_1^2 \left(\frac{x_{1,j} - u_1}{h_n} \right) d\mathbf{u} + o(h_n).
\end{aligned}$$

Thus, from (4.20), (4.22) and (4.77), we obtain

$$\mathrm{Var}\left(g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i) \right) \leq \sigma_1^2 h_n + o(h_n). \qquad (4.79)$$

Since $||g_n^{x_{1,j},\boldsymbol{\beta}}|| \leq \kappa$, an application Bernstein's inequality, given in Fact 3, with $\mathcal{C} = \kappa$, to the variable

$$Z_i = g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i), \quad \text{for} \quad i = 1, \ldots, n,$$

for any given $\tau > 0$ and for all large n, we have

$$\begin{aligned}
&\mathbb{P}\left\{ \max_{0 \leq j \leq l_n} \frac{|\alpha_n(g_n^{x_{1,j},\boldsymbol{\beta}})|}{\sqrt{2h_n \log(1/h_n)}} > \sigma_1(1+\tau) \right\} \\
&\leq 2(l_n + 1) \exp\left(- \frac{2\sigma_1^2(1+\tau)^2 \log(1/h_n)}{2\sigma_1^2 h_n + (\kappa\sigma_1/3\sqrt{n})\sqrt{2h_n \log(1/h_n)}} \right) \\
&\leq 2(l_n + 1) h_n^{1+\tau/2} \\
&= \mathcal{O}(h_n^{\tau/2}),
\end{aligned}$$

where we have used (4.73), i.e., $l_n = \mathcal{O}(h_n^{-1})$. Note that from assumption (H.1), which implies the statement (4.61), we can infer that

$$\sum_{n=1}^{\infty} h_n^{\tau/2} < \infty,$$

as it was mentioned in Einmahl and Mason (2000). Therefore we conclude that

$$\limsup_{n\to\infty} \max_{0\leq j\leq l_n} \frac{|\alpha_n(g_n^{x_{1,j},\boldsymbol{\beta}})|}{\sqrt{2h_n \log(1/h_n)}} \leq \sigma_1(1+\tau) \quad a.s., \tag{4.80}$$

by a standard argument based upon the Borel-Cantelli Lemma. □

• **Oscillation**

Next, we study the behavior of the process α_n between the grid points $x_{1,j}, x_{1,j+1}$ with $1 \leq j \leq l_n$ and $\boldsymbol{\beta} \in \mathbb{R}^p$. Set now

$$\mathcal{G}'_{n,j} = \left\{ g_n^{x_{1,j},\boldsymbol{\beta}} - g_n^{x_1,\boldsymbol{\beta}},\ x_{1,j} \leq x_1 \leq x_{1,j+1} \right\},$$

and

$$\mathcal{G}'_n = \bigcup_j \mathcal{G}'_{n,j}.$$

We claim that $\mathcal{G}'_n \subseteq \mathcal{G}'$, where $\mathcal{G}'$ satisfies conditions $(\mathscr{E})$. By the fact that the function $\mathbb{K}$ is of bounded variations, it follows that

$$\left\{ K_1\Big(\frac{x-\cdot}{h}\Big) : x \in I_1 \subset \mathbb{R}, h > 0 \right\}$$

is a bounded VC class of measurable functions, refer to van der Vaart and Wellner (1996). Now, consider the class

$$\mathbf{F} = \left\{ b(\psi(\mathbf{v}) - \mathbf{w}^\top \boldsymbol{\beta}) K_1\Big(\frac{x-u}{h}\Big) : x \in I_1 \subset \mathbb{R}, h > 0, |b| \leq D, \boldsymbol{\beta} \in \mathbb{R}^p \right\},$$

where $D > 0$ is the bound of the function $\mathbb{G}(\mathbf{u})$. Following Deheuvels and Mason (2004), one may show that $\mathbf{F}$ fulfills the conditions $(\mathscr{E})$. An easy argument shows now that the class of functions of $(u, v, \mathbf{w}) \in \mathbb{R}^2 \times \mathbb{R}^p$ defined by

$$\begin{aligned} G' \ = \ & \left\{ b(\psi(\mathbf{v}) - \mathbf{w}^\top \boldsymbol{\beta}) K_1\Big(\frac{x-u}{h}\Big) - b'(\psi(\mathbf{v}) - \mathbf{w}^\top \boldsymbol{\beta}') K_1\Big(\frac{x'-u}{h'}\Big) \right. \\ & \left. : x, x' \in I_1, h, h' > 0, |b|, |b'| \leq D, \boldsymbol{\beta}, \boldsymbol{\beta}' \in \mathbb{R}^p \right\} \end{aligned}$$

fulfills conditions ($\mathscr{E}$), refer to refer to van der Vaart and Wellner (1996). As measurable envelope function $\mathbb{G}$ for G', we can take

$$\mathbb{G}(u, \psi(\mathbf{v}), \mathbf{w}) = 2C_{\psi\mathbf{w}}||K||_\infty,$$

where $C_{\psi\mathbf{w}}$ depends only upon ψ and $\mathbf{w}$. Since $\mathcal{G}'_n \subset G'$, the claim is proved. □

Lemma 4.1. There exists a positive absolute constant B, such that for any $\epsilon > 0$, one may find a constant γ_ϵ satisfying the condition (4.73) such that $0 < \gamma < \gamma_\epsilon$. Then, we have with probability one

$$\limsup_{n\to\infty} \|n^{1/2}\alpha_n\|_{\mathcal{G}'_n} \leq B\sqrt{\epsilon n h_n \log(1/h_n)}. \tag{4.81}$$

Proof. Uniformly over $g \in \mathcal{G}'_{n,j} \subset \mathcal{G}'_n$, $||g|| \leq \kappa$. Moreover, by similar arguments as those used in the proof of (4.77), we have

$$\sigma^2_{\mathcal{G}'_n} = \sup_{g\in\mathcal{G}'_n} \mathrm{Var}(g(\mathbf{X}, \psi(\mathbf{Y}), \mathbf{Z})) \leq 4h_n\sigma_1^2. \tag{4.82}$$

Therefore, by Fact 1, for any $t > 0$, we have for suitable finite constants $A_1, A_2 > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{||n^{1/2}\alpha_n||_{\mathcal{G}'_n} \geq A_1\left(\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i g(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i)\right\|_{\mathcal{G}'_n} + t\right)\right\} \\
&\quad\leq\ 2\left\{\exp\left(-\frac{A_2t^2}{n\sigma^2_{\mathcal{G}'_n}}\right) + \exp\left(-\frac{A_2t}{\kappa}\right)\right\}.
\end{aligned} \tag{4.83}$$

Next, using the statement (4.82) combined with Fact 2, we obtain

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g_n^{x_1,\boldsymbol{\beta}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i)\right\|_{\mathcal{G}'_n} \leq A_3\sqrt{\nu n h_n \log(1/h_n)}, \tag{4.84}$$

where A_3 is an absolute constant. Thus, using the statements (4.84) and (4.83) we obtain the following bound

$$\begin{aligned}
&\mathbb{P}\left\{||n^{1/2}\alpha_n||_{\mathcal{G}'_n} \geq 2\left\{A_1A_3\sqrt{\nu n h_n \log(1/h_n)}\right\}\right\} \\
&\quad\leq\ 2\left\{\exp\left(-\frac{A_2A_3^2\nu\log(1/h_n)}{4\sigma_1^2}\right) + \exp\left(-\frac{A_2A_3\sqrt{\nu n h_n\log(1/h_n)}}{\kappa}\right)\right\} \\
&\quad\leq\ \mathcal{O}(h_n^{\tau'}),
\end{aligned} \tag{4.85}$$

where $\tau' = A_2 A_3^2/4\sigma_1^2 > 0$. Taking $B = 2A_1A_3\sqrt{\nu\epsilon}$ in the statement (4.81), we complete the proof of Lemma 4.1. □

For $1 \leq j \leq l_n$ we have $\mathcal{G}'_{n,j} \subseteq \mathcal{G}_n$ and then

$$\frac{\max_{0\leq j\leq l} ||n^{1/2}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_n\log(1/h_n)}} \leq \frac{||n^{1/2}\alpha_n||_{\mathcal{G}'_n}}{\sqrt{2nh_n\log(1/h_n)}}. \tag{4.86}$$

Using the statement (4.85) in connection with the inequality (4.86) and choosing $A = \frac{B}{\sigma_1\sqrt{2}}$, we obtain the following

$$\mathbb{P}\left\{\frac{\max_{1\leq j\leq l_n} ||n^{1/2}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_n\log(1/h_n)}} \geq \sigma_1 A\sqrt{\epsilon}\right\} = \mathcal{O}(h_n^{\tau'}). \tag{4.87}$$

□

Conclusion : Combining the statements (4.80) and (4.87), we conclude that there exists an absolute constant $A > 0$, such that

$$\begin{aligned}
&\mathbb{P}\left\{\frac{\sup_{x_1\in I_1}|\alpha_n(g_n^{x_1,\boldsymbol{\beta}})|}{\sqrt{2h_n\log(1/h_n)}} > (1+\tau+A\sqrt{\epsilon})\sigma_1\right\} \\
&\leq \mathbb{P}\left\{\max_{1\leq j\leq l_n}\frac{|\alpha_n(g_n^{x_{1,j},\boldsymbol{\beta}})|}{\sqrt{2h_n\log(1/h_n)}} > (1+\tau)\sigma_1\right\} \\
&\quad +\mathbb{P}\left\{\max_{1\leq j\leq l_n}\frac{||n^{\frac{1}{2}}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_n\log(1/h_n)}} > A\sqrt{\epsilon}\sigma_1\right\}.
\end{aligned}$$

Since for any $\varepsilon > 0$, we can choose $\tau > 0$ and $\epsilon > 0$ small enough such that $\tau + A\sqrt{\epsilon} < \varepsilon$. Then, we have

$$\mathbb{P}\left\{\frac{\sup_{x_1\in I_1}|\alpha_n(g_n^{x_1,\boldsymbol{\beta}})|}{\sqrt{2h_n\log(1/h_n)}} > (1+\varepsilon)\sigma_1\right\} = \mathcal{O}(h_n^{(\tau/2)\wedge\tau'}). \tag{4.88}$$

Noting that from assumption (H.1) and recall the statement (4.61), we can infer that

$$\sum_{n=1}^{\infty} h_n^{(\tau/2)\wedge\tau'} < \infty.$$

By using the Borel-Cantelli Lemma combined with the statement (4.88), we obtain

$$\limsup_{n\to\infty}\sup_{x_1\in I_1}\frac{|\alpha_n(g_n^{x_1,\boldsymbol{\beta}})|}{\sqrt{2h_n\log(1/h_n)}} \leq \sigma_1(1+\tau) \quad a.s. \tag{4.89}$$

□

Proof of Proposition 4.1 : Lower bound part

Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ be a sequence of i.i.d. random vectors taking values in $\mathbb{R}^m$ with common distribution $\mathbb{H}$. For each $n \geq 1$, consider the empirical distribution function, defined, for any $\mathbf{s} \in \mathbb{R}^m$, by

$$\mathbb{H}_n(\mathbf{s}) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{\boldsymbol{\xi}_i \leq \mathbf{s}\}},$$

where as usual $\mathbf{t} \leq \mathbf{s}$ means that each component of $\mathbf{t}$ is less than or equal to the corresponding component of $\mathbf{s}$. For any measurable real valued function g defined on $\mathbb{R}^m$, set

$$\mathbb{H}_n(g) = \int_{\mathbb{R}^m} g(\mathbf{s})\, d\mathbb{H}_n(\mathbf{s}), \quad \mu(g) = \mathbb{E}(g(\boldsymbol{\xi})) \text{ and } \overline{\sigma}(g) = \sqrt{\mathrm{Var}(g(\boldsymbol{\xi}))}.$$

Let $\{b_n : n \geq 1\}$ denote a sequence of positive constants converging to zero and satisfying the condition

$$\log(1/b_n)/\log\log n \to \infty. \tag{4.90}$$

Consider a sequence

$$\mathcal{G}_n = \{g_i^{(n)};\ i = 1, \ldots, k_n\}$$

of sets of real-valued measurable functions on $\mathbb{R}^m$, satisfying, whenever $g_i^{(n)} \in \mathcal{G}_n$, the conditions

(R.1)

$$\mathbb{P}\left\{g_i^{(n)}(\boldsymbol{\xi}) \neq 0,\ g_j^{(n)}(\boldsymbol{\xi}) \neq 0\right\} = 0 \quad \forall 1 \leq i \neq j \leq k_n,$$
$$\sum_{i=1}^{k_n} \mathbb{P}\left\{g_i^{(n)}(\boldsymbol{\xi}) \neq 0\right\} \leq 1/2.$$

Furthermore, assume that

(R.2) For some $0 < r < \infty$, $b_n k_n \to r$ as $n \to \infty$;

(R.3) For some $-\infty < \mu_1, \mu_2 < \infty$, uniformly in $1 \leq i \leq k_n$, for any n large enough,

$$b_n \mu_1 \leq \mu(g_i^{(n)}) \leq b_n \mu_2; \tag{4.91}$$

(R.4) For some $0 < \sigma_1 < \sigma_2 < \infty$, uniformly in $1 \leq i \leq k_n$, for any n large enough,

$$\sigma_1 \sqrt{b_n} \leq \sigma(g_i^{(n)}) \leq \sigma_2 \sqrt{b_n}; \tag{4.92}$$

(R.5) For some $0 < \kappa' < \infty$, uniformly in $1 \leq i \leq k_n$, for any n large enough, $\|g_i^{(n)}\| \leq \kappa'$.

The following lemma due to Einmahl and Mason (2000) is the main tool to prove our result. We will work only in the "+" version case, the arguments for the "−" case can be obtained similarly.

Lemma 4.2. Under the above assumptions, with probability one for each $0 < \varepsilon < 1$, there exists an N_ε such that for each $n \geq N_\varepsilon$,

$$\max_{1\leq i\leq k_n} \frac{n^{1/2}\{\mathbb{H}_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\bar{\sigma}(g_i^{(n)})\sqrt{2\log(1/b_n)}} \geq (1-\varepsilon). \tag{4.93}$$

Proof. See Proposition 2 of Einmahl and Mason (2000). □

For any $\epsilon > 0$, select a sub-interval $\mathcal{I}_1 = [\mathcal{A}_1, \mathcal{C}_1]$ of $I_1 = [a_1, c_1]$, such that

$$\mathbb{P}\{X_1 \in \mathcal{I}_1\} \leq \frac{1}{2}$$

and

$$\inf_{u_1\in\mathcal{I}_1} \sqrt{\frac{\phi(u_1)}{f_1(u_1)}} \left[\int_{\mathbb{R}} K_1^2\right]^{1/2} > \sigma_1(1-\varepsilon/2), \tag{4.94}$$

where ϕ and σ_1 are as defined, respectively, in (4.20) and (4.22). Our aim is to verify conditions allowing to apply Lemma 4.2. In the interval $\mathcal{I}_1$, select the points

$$x_{1,j} = \mathcal{A}_1 + 2jh_n, \text{ for } 1 \leq j \leq \left\lfloor \frac{\mathcal{C}_1 - \mathcal{A}_1}{2h_n} \right\rfloor - 1 := k_n.$$

For each $x_{1,j}$, with $1 \leq j \leq k_n$, define the function

$$g_j^{(n)}(\mathbf{x},\mathbf{y},\mathbf{z}) := g_n^{x_{1,j},\boldsymbol{\beta}}(\mathbf{x},\mathbf{y},\mathbf{z}) := (\psi(\mathbf{y}) - \mathbf{z}^\top\boldsymbol{\beta})G(\mathbf{x})K_1\Big(\frac{x_{1,j}-x_1}{h_n}\Big). \tag{4.95}$$

Notice that, for each $1 \leq j \leq k_n$,

$$||g_j^{(n)}|| \leq \kappa := \kappa'.$$

Along this proof, we choose $a_n = h_n$, then, as $n \to \infty$,

$$h_n k_n \to \left[\frac{\mathcal{C}_1 - \mathcal{A}_1}{2}\right] := r.$$

Since $K_1(s) = 0$ for $s \notin [-1/2, 1/2]$, observe that

$$g_i^{(n)}(\mathbf{X},\mathbf{Y},\mathbf{Z}) \neq 0 \Longleftrightarrow \left|\frac{x_{1,i} - X_{1,i}}{h_n}\right| \leq \frac{1}{2},$$

which implies that

$$|x_{1,j} - X_{1,i}| \;=\; |x_{1,j} - x_{1,i} + x_{1,i} - X_{1,i}| \geq 2h_n - \frac{h_n}{2}.$$

Thus, for all $1 \leq i \neq j \leq k_n$, we have

$$\mathbb{P}\Big\{\, g_i^{(n)}(\mathbf{X},\mathbf{Y},\mathbf{Z}) \neq 0 \text{ and } g_j^{(n)}(\mathbf{X},\mathbf{Y},\mathbf{Z}) \neq 0\Big\} = 0.$$

Now, for any $1 \leq i \leq k_n$, set

$$\bar{\sigma}(g_i^{(n)}) := \mathrm{Var}(g_i^{(n)}(\mathbf{X},\mathbf{Y},\mathbf{Z})).$$

Then, combining the statements (4.77), (4.79) and (4.94), we obtain

$$\sigma_1^2(1-\varepsilon)h_n \leq \bar{\sigma}(g_i^{(n)}) \leq \sigma_1^2(1+\varepsilon)h_n. \tag{4.96}$$

This implies that (4.92) holds true. Similarly, one may verify that the condition (4.91) is satisfied. Therefore, all the assumptions required in Lemma 4.2 are satisfied. Note that (4.96) implies that, for each $1 \leq i \leq k_n$,

$$\begin{aligned}\bar{\sigma}(g_i^{(n)})(1-\varepsilon) &\geq \sigma_1(1-\varepsilon)\sqrt{1-\varepsilon}\sqrt{h_n} \\ &= \sigma_1(1-\varepsilon)^{2/3}\sqrt{h_n}.\end{aligned}$$

So, with probability one, for each $0 \leq \varepsilon \leq 1$ and n large enough, we have

$$\max_{1\leq i\leq k_n} \frac{n^{1/2}\{\mathbb{H}_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sqrt{2h_n\log(1/h_n)}} \geq \sigma_1(1-\varepsilon)^{2/3}.$$

Taking $(1-\epsilon)^{3/2} = 1 - \frac{\epsilon'}{2}$, and using the inequality,

$$\frac{\sup_{x_1\in I_1}\alpha_n(g_n^{x_1,\boldsymbol{\beta}})}{\sqrt{2h_n\log(1/h_n)}} \geq \max_{1\leq i\leq k_n} \frac{n^{1/2}(\mathbb{H}_n(g_i^n) - \mu(g_i^{(n)}))}{\sqrt{2h_n\log(1/h_n)}},$$

we readily obtain

$$\liminf_{n\to\infty} \frac{\sup_{x_1\in I_1}\alpha_n(g_n^{x_1,\boldsymbol{\beta}})}{\sqrt{2h_n\log(1/h_n)}} \geq \sigma_1\left(1 - \frac{\epsilon'}{2}\right) \quad a.s. \tag{4.97}$$

This allows us to obtain the lower bound and to complete the proof of Proposition 4.1. □

4.6.1 The unbounded case

In this section we will briefly discuss the unbounded case, where (M.1) is replaced by (M.1)′ and the sequence h_n satisfies in addition to (H.1) the following

$$h_n^{-1} \leq (n/\log(1/h_n))^{1-2/s}.$$

We begin with a truncation argument as in Einmahl and Mason (2000). Let

$$\underline{g_n^{x_1,\boldsymbol{\beta}}}(\mathbf{X}_i, \psi(\mathbf{Y}_i), \mathbf{Z}_i) = (\underline{\psi(\mathbf{Y}_i)} - \mathbf{Z}_i^\top \boldsymbol{\beta}) G(\mathbf{X}_i) K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right), \tag{4.98}$$

and

$$\mu_n(x_1) = n\mathbb{E}\left(\overline{\psi(\mathbf{Y}_i)} G(\mathbf{X}_i) K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right)\right), \tag{4.99}$$

where

$$\underline{\psi(\mathbf{Y}_i)} := \psi(\mathbf{Y}_i)\mathbb{1}_{\{|\psi(\mathbf{Y}_i)|<n^{1/s}\}} \quad \text{and} \quad \overline{\psi(\mathbf{Y}_i)} := \psi(\mathbf{Y}_i)\mathbb{1}_{\{|\psi(\mathbf{Y}_i)|\geq n^{1/s}\}}.$$

It is easy to see that

$$\alpha_n(g_n^{x_1,\boldsymbol{\beta}}) - \alpha_n(\underline{g_n^{x_1,\boldsymbol{\beta}}}) = \sum_{i=1}^n \overline{\psi(\mathbf{Y}_i)} G(\mathbf{X}_i) K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right) - \mu_n(x_1). \tag{4.100}$$

Follwing the proof Lemma 1 of Einmahl and Mason (2000), we have

$$\sup_{x_1 \in I_1} \frac{|\mu_n(x_1)|}{\sqrt{nh_n \log(1/h_n)}} = 0, \tag{4.101}$$

and

$$\left|\sum_{i=1}^n \overline{\psi(\mathbf{Y}_i)} G(\mathbf{X}_i) K_1\left(\frac{x_1 - X_{1,i}}{h_n}\right)\right| < \infty. \tag{4.102}$$

When combined with (4.101), (4.102) implies, almost surely

$$\sup_{x_1 \in I_1} \frac{|\alpha_n(g_n^{x_1,\boldsymbol{\beta}}) - \alpha_n(\underline{g_n^{x_1,\boldsymbol{\beta}}})|}{\sqrt{nh_n \log(1/h_n)}} = 0. \tag{4.103}$$

Let, for fixed $\psi \in \mathcal{F}$,

$$\underline{\mathcal{G}_n} = \{\underline{g_n^{x_1,\boldsymbol{\beta}}} : x_1 \in I_1\}.$$

Following the proof of Proposition 4.1, we obtain the following result.

Proposition 4.2. Suppose that the assumptions (G.1)-(G.3), (H.1)-(H.2), (K.1)-(K.3), (M.1)′-(M.2) and (Q.1)-(Q.2) are satisfied. Then, as $n \to \infty$, we have

$$\sup_{x_1 \in I_1} \frac{|\alpha_n(\underline{g_n^{x_1,\boldsymbol{\beta}}})|}{\sqrt{2h_n \log(1/h_n)}} \to \sigma_1 \quad a.s,$$

provided that

$$h_n^{-1} \leq (n/\log(1/h_n))^{1-2/s}.$$

Chapitre 5

Perspectives de recherche

Les perspectives de recherche prolongeant les études effectuées dans le cadre de cette thèse sont diverses. Les résultats obtenus sur le modèle additif partiellemnt linéaire donnent la possibilité d'explorer d'autres alternatives et cela à plusieurs niveaux. Il serait intéressant d'étendre notre travail aux données présentant une structure dépendance faible des covariables (cas de processus mélangeants). Ce même modèle peut être envisager dans le cas de données incomplètes (données censurées). Dans ce travail, les erreurs de modélisation sont supposées homoscédastiques. Plusieurs phénomènes pourrait être modéliser en considérant une hypothèse d'hétéroscédasticité des erreurs. Il serait alors intéressant d'étudier les propriétés de ce modèle dans ce cas de figure. Enfin, en guise de perspective, nous présentons brièvement certains travaux en cours de préparation.

5.1 Les modèles semi-paramétriques : cas de données ergodiques

Dans un premier stade, on va considérer un modèle partiellement linéaire mais avec des covariables ergodiques, ensuite supposer une structure additive de la fonction de régression non-parammétrique, pour aboutir à un modèle additive partiellement linéaire dans le cadre ergodique.

5.2 Les modèles semi-fonctionnels partiellement linéaires

Dans ce travail, on va exploiter le modèle additif partiellement linéaire défini auparavant par

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + \sum_{l=1}^{d} m_l(X_l) + \varepsilon,$$

mais cette fois-ci les X_l sont des variables explicatives fonctionnelles.
Nous envisageons également d'introduire le modèle semi-fonctionnel généralisé partiellement linéaire défini par

$$Y = \mathbf{Z}^\top \boldsymbol{\beta} + m(\varphi(X)) + \varepsilon,$$

avec φ est une fonction multivariée.

5.3 Les modèles fonctionnels partiellement linéaires

C'est un autre thème principal de recherche que nous aimerions développer dans le futur en considérant un modèle de regression sous la forme

$$Y = \int_0^T Z(t)\rho(t)dt + \int_0^T m(X(t))dt + \varepsilon,$$

où $Z(\cdot), X(\cdot)$ sont des variables fonctionnelles définies sur $[0, T]$ à valeurs dans un espace de Helbert, ρ est un opérateur linéaire inconnu et m est une fonction de régression non-paramétrique inconnue. Pour le cas de la regression linéaire fonctionnelle on peut consulter Mas and Pumo (2009).

Bibliographie

Aneiros-Pérez, G., González-Manteiga, W., and Vieu, P. (2004). Estimation and testing in a partial linear regression model under long-memory dependence. *Bernoulli*, 10(1) :49–78.

Aneiros-Pérez, G. and Vieu, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.*, 76(11) :1102–1110.

Aneiros-Pérez, G. and Vieu, P. (2008). Nonparametric time series prediction : a semi-functional partial linear modeling. *J. Multivariate Anal.*, 99(5) :834–857.

Ango-Nze, P. and Rios, R. (2000). Density estimation in L^∞ norm for mixing processes. *J. Statist. Plann. Inference*, 83(1) :75–90.

Bhattacharya, P. K. and Zhao, P.-L. (1997). Semiparametric inference in a partial linear model. *Ann. Statist.*, 25(1) :244–262.

Blondin, D. (2007). Rates of strong uniform consistency for local least squares kernel regression estimators. *Statist. Probab. Lett.*, 77(14) :1526–1534.

Bosq, D. (1996). *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Estimation and prediction.

Bosq, D. and Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle*. Paris.

Bouzebda, S. (2012). On the strong approximation of bootstrapped empirical copula processes with applications. *Math. Methods Statist.*, 21(3) :153–188.

Bouzebda, S. and Elhattab, I. (2009). A strong consistency of a nonparametric estimate of entropy under random censorship. *C. R. Math. Acad. Sci. Paris*, 347(13-14) :821–826.

Bouzebda, S. and Elhattab, I. (2011). Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electron. J. Stat.*, 5 :440–459.

Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.*, 29 :437–454.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.*, 17(2) :453–555.

Camlong-Viot, C. (2001). Vers un test d'additivité en régression non paramétrique sous des conditions de mélange. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(9) :877–880.

Camlong-Viot, C., Rodríguez-Póo, J. M., and Vieu, P. (2006). Nonparametric and semiparametric estimation of additive models with both discrete and continuous variables under dependence. In *The art of semiparametrics*, Contrib. Statist., pages 155–178. Physica-Verlag/Springer, Heidelberg.

Camlong-Viot, C., Sarda, P., and Vieu, P. (2000). Additive time series : the kernel integration method. *Math. Methods Statist.*, 9(4) :358–375.

Chacón, J. E., Montanero, J., and Nogales, A. G. (2007). A note on kernel density estimation at a parametric rate. *J. Nonparametr. Stat.*, 19(1) :13–21.

Chacón, J. E. and Rodríguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator. *Statist. Probab. Lett.*, 80(17-18) :1414–1419.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, 60(3) :567–596.

Chen, G. and Wang, Z. (2010). The multivariate partially linear model with B-spline. *Chinese J. Appl. Probab. Statist.*, 26(2) :138–150.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, 16(1) :136–146.

Chen, H. and Shiau, J. J. H. (1991). A two-stage spline smoothing method for partially linear models. *J. Statist. Plann. Inference*, 27(2) :187–201.

Chokri, K. and Louani, D. (2011). Aymptotic results for the linear parameter estimate in partially linear additive regression model. *C. R. Math. Acad. Sci. Paris*, 349(19-20) :1105–1109.

Clarkson, J. A. and Adams, C. R. (1933). On definitions of bounded variation for functions of two variables. *Trans. Amer. Math. Soc.*, 35(4) :824–854.

Collomb, G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau : quelques propriétés de convergence ponctuelle. In *Nonparametric asymptotic statistics (Proc. Conf., Rouen, 1979) (French)*, volume 821 of *Lecture Notes in Math.*, pages 159–175. Springer, Berlin.

Dabo-Niang, S. and Guillas, S. (2010). Functional semiparametric partially linear model with autoregressive errors. *J. Multivariate Anal.*, 101(2) :307–315.

Debbarh, M. (2008). Some uniform limit results in additive regression model. *Comm. Statist. Theory Methods*, 37(18-20) :3090–3114.

Deheuvels, P. (2011). One bootstrap suffices to generate sharp uniform bounds in functional estimation. *Kybernetika (Prague)*, 47(6) :855–865.

Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, 7(3) :225–277.

Delecroix, M. and Rosa, A. C. (1996). Nonparametric estimation of a regression function and its derivatives under an ergodic hypothesis. *J. Nonparametr. Statist.*, 6(4) :367–382.

Derzko, G. and Deheuvels, P. (2002). Estimation non-paramétrique de la régression dichotomique—application biomédicale. *C. R. Math. Acad. Sci. Paris*, 334(1) :59–63.

Devroye, L. and Györfi, L. (1985). *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics : Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L_1 view.

Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York.

Devroye, L. P. (1976). On the convergence of statistical search. *IEEE Trans. Systems, Man Cybernet.*, SMC-6(1) :46–56.

Donald, S. G. and Newey, W. K. (1994). Series estimation of semilinear models. *J. Multivariate Anal.*, 50(1) :30–40.

Dony, J., Einmahl, U., and Mason, D. M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.*, 35 :105–120.

Dony, J. and Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, 14(4) :1108–1133.

Draper, N. R. and Smith, H. (1981). *Applied regression analysis*. John Wiley & Sons Inc., New York, second edition. Wiley Series in Probability and Mathematical Statistics.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13(1) :1–37.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3) :1380–1403.

Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, G. H. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, 81(394) :310–320.

Eubank, R. L. (1985). Diagnostics for smoothing splines. *J. Roy. Statist. Soc. Ser. B*, 47(2) :332–341.

Eubank, R. L. and Speckman, P. (1990). Curve fitting by polynomial-trigonometric regression. *Biometrika*, 77(1) :1–9.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting : variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B*, 57(2) :371–394.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Ann. Statist.*, 26(3) :943–971.

Farebrother, R. W. (1988). *Linear least squares computations*, volume 91 of *Statistics : Textbooks and Monographs*. Marcel Dekker Inc., New York.

Gao, J. T. (1995). Asymptotic properties of some estimators for partly linear stationary autoregressive models. *Comm. Statist. Theory Methods*, 24(8) :2011–2026.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, volume 757 of *Lecture Notes in Math.*, pages 23–68. Springer, Berlin.

Giné, E. and Mason, D. M. (2008). Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.*, 35(4) :739–761.

Green, P. and Yandell, B. (1985). Semi-parametric generalized linear models. *Technical Report*, 2847.

Green, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika*, 72(3) :527–537.

Győrfi, L. (1978). On the rate of convergence of nearest neighbor rules. *IEEE Trans. Inform. Theory*, 24(4) :509–512.

Györfi, L. and Kohler, M. (2002). Nonparametric regression estimation. In *Principles of nonparametric learning (Udine, 2001)*, volume 434 of *CISM Courses and Lectures*, pages 57–112. Springer, Vienna.

Hamilton, S. A. and Truong, Y. K. (1997). Local linear estimation in partly linear models. *J. Multivariate Anal.*, 60(1) :1–19.

Härdle, W. (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Härdle, W. and Liang, H. (2007). Partially linear models. In *Statistical methods for biostatistics and related fields*, pages 87–103. Springer, Berlin.

Härdle, W., Liang, H., and Gao, J. (2000). *Partially linear models*. Contributions to Statistics. Physica-Verlag, Heidelberg.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4) :1926–1947.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.*, 1(3) :297–318. With discussion.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B*, 48(2) :244–248.

Heckman, N. E. (1988). Minimax estimates in a semiparametric model. *J. Amer. Statist. Assoc.*, 83(404) :1090–1096.

Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.*, 95(2) :246–272.

Hobson, E. W. (1958). *The theory of functions of a real variable and the theory of Fourier's series. Vol. I.* Dover Publications Inc., New York, N.Y.

Horng Shiau, J.-J. and Wahba, G. (1988). Rates of convergence of some estimators for a semiparametric model. *Comm. Statist. Simulation Comput.*, 17(4) :1117–1133.

Jones, M. C. (1995). On higher order kernels. *J. Nonparametr. Statist.*, 5(2) :215–221.

Jones, M. C., Davies, S. J., and Park, B. U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.*, 89(427) :825–832.

Jones, M. C., Linton, O., and Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika*, 82(2) :327–338.

Jones, M. C. and Signorini, D. F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Statist. Assoc.*, 92(439) :1063–1073.

Ledoux, M. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1 :63–87 (electronic).

Lee, Y. K., Mammen, E., and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.*, 38(5) :2857–2883.

Liang, H. (2000). Asymptotic normality of parametric part in partially linear models with measurement error in the nonparametric part. *J. Statist. Plann. Inference*, 86(1) :51–62.

Liang, H., Härdle, W., and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Ann. Statist.*, 27(5) :1519–1535.

Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T., and Hauser, R. (2008). Additive partial linear models with measurement errors. *Biometrika*, 95(3) :667–678.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1) :93–100.

Linton, O. B. and Jacho-Chávez, D. T. (2010). On internally corrected and symmetrized kernel estimators for nonparametric regression. *TEST*, 19(1) :166–186.

Mack, Y. P. and Müller, H.-G. (1989). Derivative estimation in nonparametric regression with random predictor variable. *Sankhyā Ser. A*, 51(1) :59–72.

Mammen, E., Carroll, R. J., and Härdle, W. (2002). Estimation in an additive model when the components are linked parametrically. *Econometric Theory*, 18(4) :886–912.

Mammen, E., Linton, O., and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.*, 27(5) :1443–1490.

Mammen, E. and Park, B. U. (2006). A simple smooth backfitting method for additive models. *Ann. Statist.*, 34(5) :2252–2271.

Mammen, E., Yu, K., and Park, B. U. (2011). Semi-parametric regression : efficiency gains from modeling the nonparametric part. *Bernoulli*, 17(2) :736–748.

Manzan, S. and Zerom, D. (2005). Kernel estimation of a partially linear additive model. *Statist. Probab. Lett.*, 72(4) :313–322.

Mas, A. and Pumo, B. (2009). Functional linear regression with derivatives. *J. Nonparametr. Stat.*, 21(1) :19–40.

Mason, D. M. and Swanepoel, J. W. H. (2010). A general result on the uniform in bandwidth consistency of kernel-type function estimator. *TEST, in press.*

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2) :233–253.

Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(1) :43–61.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.*, 25(1) :186–211.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33 :1065–1076.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York.

Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation*. Probability and Mathematical Statistics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.

Priestley, M. B. and Chao, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B*, 34 :385–392.

Rice, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.*, 4(4) :203–208.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56(4) :931–954.

Scott, D. and Thompson, J. (1983). Probability density estimation in higher dimensions. *Computer Science and Statistics*, pages 173–179.

Scott, D. W. (1992). *Multivariate density estimation*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York. Theory, practice, and visualization, A Wiley-Interscience Publication.

Seber, G. A. F. (1977). *Linear regression analysis*. John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, 89(426) :501–511.

Shen, C.-W., Tsou, T.-S., and Balakrishnan, N. (2011). Robust likelihood inference for regression parameters in partially linear models. *Comput. Statist. Data Anal.*, 55(4) :1696–1714.

Shen, J. and Xie, Y. (2013). Strong consistency of the internal estimator of nonparametric regression with dependent data. *Statist. Probab. Lett.*, 83(8) :1915–1925.

Shi, P. D. and Li, G. Y. (1995a). Asymptotic normality of the M-estimators for parametric components in partly linear models. *Northeast. Math. J.*, 11(2) :127–138.

Shi, P. D. and Li, G. Y. (1995b). A note on the convergent rates of M-estimates for a partly linear model. *Statistics*, 26(1) :27–47.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B*, 50(3) :413–436.

Sperlich, S., Linton, O. B., and Härdle, W. (1999). Integration and backfitting methods in additive models – finite sample properties and comparison. *Test*, 8(2) :419–458.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, 13(2) :689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, 14(2) :590–606.

Stout, W. F. (1974). *Almost sure convergence*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London. Probability and Mathematical Statistics, Vol. 24.

Stute, W. (1984). The oscillation behavior of empirical processes : the multivariate case. *Ann. Probab.*, 12(2) :361–379.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1) :28–76.

Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series : projections. *J. Amer. Statist. Assoc.*, 89(428) :1398–1409.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Vapnik, V. N. and Červonenkis, A. J. (1971). The uniform convergence of frequencies of the appearance of events to their probabilities. *Teor. Verojatnost. i Primenen.*, 16 :264–279.

Varron, D. (2006). Some uniform in bandwidth functional results for the tail uniform empirical and quantile processes. *Ann. I.S.U.P.*, 50(1-2) :83–102.

Vituškin, A. G. (1955). *O mnogomernyh variaciyah*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow.

Wahba, G. (1986). Partial spline modelling of the tropopause and other discontinuities. In *Function estimates (Arcata, Calif., 1985)*, volume 59 of *Contemp. Math.*, pages 125–135. Amer. Math. Soc., Providence, RI.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Wolverton, C. T. and Wagner, T. J. (1969). Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Information Theory*, IT-15 :258–265.