Contribution to spatial statistics and functional data analysis

Mohamed-Salem Ahmed

UNIVERSITÉ DE LILLE, FRANCE.

École Doctorale SESAM

Thése de doctorat

Mention : Mathématiques et Applications

présentée par

Mohamed-Salem Ahmed

Contribution à la statistique spatiale et l'analyse de données fonctionnelles

---

Contribution to spatial statistics and functional data analysis.

---

dirigée par Sophie Dabo-Niang et Mohamed Attouch.

Soutenue le 12 décembre 2017 devant le jury composé de :

M^{r} Mohammed Attouch	Université Djillali Liabes (Algérie)	Co-directeur de thèse
M^{me} Laurence Broze	Université de Lille	Examinatrice
M^{me} Sophie Dabo-Niang	Université de Lille	Directrice de thèse
M^{r} Ramón Giraldo	Université Nationale de Colombie (Colombie)	Rapporteur
M^{me} Sana Louhichi	Université de Grenoble-Alpes	Rapportrice
M^{r} Cristian Preda	Université de Lille	Examinateur
M^{r} Nourddine Rhomari	Université Mohammed Premier (Maroc)	Examinateur
M^{r} Anuj Srivastava	Université de Floride (Etats-Unis)	Rapporteur

Résumé en français

Contribution à la statistique spatiale et l'analyse de données fonctionnelles

Ce mémoire de thèse porte sur la statistique inférentielle des données spatiales et/ou fonctionnelles. En effet, nous nous sommes intéressés à l'estimation de paramètres inconnus de certains modèles à partir d'échantillons obtenus par un processus d'échantillonnage aléatoire ou non (stratifié), composés de variables indépendantes ou spatialement dépendantes. La spécificité des méthodes proposées réside dans le fait qu'elles tiennent compte de la nature de l'échantillon étudié (échantillon stratifié ou composé de données spatiales dépendantes).

Tout d'abord, nous étudions des données à valeurs dans un espace de dimension infinie ou dites "données fonctionnelles". Dans un premier temps, nous étudions les modèles de choix binaires fonctionnels dans un contexte d'échantillonnage par stratification endogène (*échantillonnage Cas-Témoin* ou *échantillonnage basé sur le choix*). La spécificité de cette étude réside sur le fait que la méthode proposée prend en considération le schéma d'échantillonnage. Nous décrivons une fonction de vraisemblance conditionnelle sous l'échantillonnage considérée et une stratégie de réduction de dimension afin d'introduire une estimation du modèle par vraisemblance conditionnelle. Nous étudions les propriétés asymptotiques des estimateurs proposées ainsi que leurs applications à des données simulées et réelles. Nous nous sommes ensuite intéressés à un modèle linéaire fonctionnel spatial auto-régressif. La particularité du modèle réside dans la nature fonctionnelle de la variable explicative et la structure de la dépendance spatiale des variables de l'échantillon considéré. La procédure d'estimation que nous proposons consiste à réduire la dimension infinie de la variable explicative fonctionnelle et à maximiser une quasi-vraisemblance associée au modèle. Nous établissons la consistance, la normalité asymptotique et les performances numériques des estimateurs proposés.
Dans la deuxième partie du mémoire, nous abordons des problèmes de régression et prédiction de variables dépendantes à valeurs réelles. Nous commençons par généraliser la méthode de k-plus proches voisins (k-nearest neighbors; k-NN) afin de prédire un processus spatial en des sites non-observés, en présence de co-variables spatiaux. La spécificité du prédicteur proposé est qu'il tient compte d'une hétérogénéité au niveau de la co-variable utilisée. Nous établissons la convergence presque complète avec vitesse du prédicteur et donnons des résultats numériques à l'aide de données simulées et environnementales. Nous généralisons ensuite le modèle probit partiellement linéaire pour données indépendantes à des données spatiales. Nous utilisons un processus spatial linéaire pour modéliser les perturbations du processus considéré, permettant ainsi plus de flexibilité et d'englober plusieurs types de dépendances spatiales. Nous proposons une approche d'estimation semi-paramétrique basée sur une vraisemblance pondérée et la méthode des moments généralisées et en étudions les propriétés asymptotiques et performances numériques. Une étude sur la détection des facteurs de risque de cancer VADS (voies aéro-digestives supérieures) dans la région Nord de France à l'aide de modèles spatiaux à choix binaire termine notre contribution.

Mots-Clefs

Modèle à choix binaire, Analyses de données fonctionnelles, Échantillonnage basé sur le choix, Échantillonnage Cas-Témoin, Modèle linéaire fonctionnel, Processus auto-régressif spatial, Quasi-maximum de vraisemblance, Statistique Non-paramétrique, Régression,

Prédiction, k-plus proches voisins, Estimateur à Noyau, Processus spatial, Econométrie spatiale, Estimation Semi-paramétrique, Méthodes des moments généralisées.

Abstract

Contribution to spatial statistics and functional data analysis

This thesis is about statistical inference for spatial and/or functional data. Indeed, we are interested in estimation of unknown parameters of some models from random or non-random (stratified) samples composed of independent or spatially dependent variables. The specificity of the proposed methods lies in the fact that they take into consideration the considered sample nature (stratified or spatial sample).

We begin by studying data valued in a space of infinite dimension or so-called "functional data". First, we study a functional binary choice model explored in a case-control or choice-based sample design context. The specificity of this study is that the proposed method takes into account the sampling scheme. We describe a conditional likelihood function under the sampling distribution and a reduction of dimension strategy to define a feasible conditional maximum likelihood estimator of the model. Asymptotic properties of the proposed estimates as well as their application to simulated and real data are given. Secondly, we explore a functional linear autoregressive spatial model whose particularity is on the functional nature of the explanatory variable and the structure of the spatial dependence. The estimation procedure consists of reducing the infinite dimension of the functional variable and maximizing a quasi-likelihood function. We establish the consistency and asymptotic normality of the estimator. The usefulness of the methodology is illustrated via simulations and an application to some real data.

In the second part of the thesis, we address some estimation and prediction problems of real random spatial variables. We start by generalizing the k-nearest neighbors method, namely k-NN, to predict a spatial process at non-observed locations using some covariates. The specificity of the proposed k-NN predictor lies in the fact that it is flexible and allows a number of heterogeneity in the covariate. We establish the almost complete convergence with rates of the spatial predictor whose performance is ensured by an application over simulated and environmental data. In addition, we generalize the partially linear probit model of independent data to the spatial case. We use a linear process for disturbances allowing various spatial dependencies and propose a semiparametric estimation approach based on weighted likelihood and generalized method of moments methods. We establish the consistency and asymptotic distribution of the proposed estimators and investigate the finite sample performance of the estimators on simulated data. We end by an application of spatial binary choice models to identify UADT (Upper aerodigestive tract) cancer risk factors in the north region of France which displays the highest rates of such cancer incidence and mortality of the country.

Keywords

Binary choice model, Functional data analysis, Choice-based sampling, Case-control, Functional Linear Model, Spatial Autoregressive Process, Quasi-maximum likelihood estimator, Nonparametric statistics, Regression, Prediction, k-nearest neighbors, Kernel estimate, Spatial process, Spatial econometrics, Semi-parametric estimation, Generalized method of moments.

Remerciements

Je tiens tout d'abord à exprimer toute ma reconnaissance à Sophie Dabo-Niang, ma directrice de thèse, qui m'a tellement appris, et ce depuis ma dernière année de master, durant laquelle j'ai réalisé mon premier travail de recherche. Pendant ces derniìeres années, elle a su être à l'écoute et a été très disponible, malgré ses diverses responsabilités et son emploi du temps chargé. Je la remercie également de m'avoir accordé sa confiance en acceptant d'encadrer ma thèse mais aussi de m'avoir donné la chance de vivre des expériences enrichissantes. J'espère que nous aurons l'opportunité de collaborer durant les prochaines années.

Je remercie vivement Mohammed Kadi Attouch d'avoir tout fait pour que je puisse m'inscrire en thèse, mais aussi de m'avoir encadré durant ma dernière année de master et d'avoir co-dirigé ma thèse. Merci à Mohammed pour sa bonne humeur, ses nombreux conseils et pour son accueil lors de mon dernier séjour à Sidi Bel Abbés (Algérie).

Je remercie vivement Ramón Giraldo, Sana Louhichi et Anuj Srivastava d'avoir accepté de rapporter ce travail de thèse. Je suis très honoré qu'ils aient pris le temps d'expertiser mon travail. Je souhaite également témoigner ma reconnaissance à Laurence Broze, Cristian Preda et Nourddine Rhomari pour leur participation dans le jury de soutenance.

Je remercie vivement Fateh Chebana et Taha B. M. J. Ouarda de m'avoir permis de travailler à INRS et l'intérêt qu'ils ont porté à notre projet de collaboration. C'est grâce à eux que j'ai fait mon premier pas dans le monde des "applications".

J'adresse un remerciement aux doctorants de UFR MIME de Lille 3 (Aladji, Emad, Florian, Mamadou, Sara, Walid, Zied,...) avec qui nous avons discuter "Mathématiques" mais aussi de nos différentes cultures et de petites histoires du quotidien. J'ai également une pensée pour tous les membres du laboratoire LEM (Aboubacar, Baba, Olivier, Luc, Camille, Ophélie, Aurore,...). Il me faut également remercier l'ensemble des personnes présentes à l'UFR MIME pour leur gentillesse et disponibilité.

Je tiens aussi à remercier tous les membres du département de mathématiques de la faculté de sciences exactes de l'université Djillali Liabes (Algérie), et en particulier Ali Laksaci avec qui j'ai découvert l'utilité de la statistique.

Je souhaite également remercier les réviseurs anonymes de nos articles publiés ou en révision ainsi que les personnes présentes lors de mes exposés. Grâce à leurs remarques et suggestions, j'ai pu me poser de nouvelles questions et ainsi améliorer le travail présenté dans ce manuscrit.

Pour terminer, je remercie très sincèrement toutes les personnes qui n'ont jamais arrêté de demander de mes nouvelles durant ces années de thèse.

Contents

Notations

$\mathbb{N}$	set of natural numbers: $0, 1, 2 \ldots$
$\mathbb{N}^*$	set of non-zero natural numbers: $1, 2 \ldots$
$\mathbb{Z}$	set of integers: $\ldots, -1, 0, 1, \ldots$
$\mathbb{R}$	set of real numbers: $]-\infty, +\infty[$
$\mathbb{R}_+$	set of real positives numbers: $[0, +\infty[$
$\mathbb{R}^d$	euclidian space of dimension d
$\lfloor \cdot \rfloor$	integer part
$\lvert \cdot \rvert$	absolute value if the argument is number or determinant if the argument is matrix
$\lVert \cdot \rVert$	norm such that: if the argument is a vector $x \in \mathbb{R}^d$: $\lVert x \rVert = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$ if the argument is a matrix A: $\lVert A \rVert = \sqrt{\sum \sum a_{ij}^2}$ if the argument is a function f: $\lVert f \rVert = \sup \lvert f(x) \rvert$
$x^{'}$ or x^T	transpose of vector or matrix x
$\mathrm{tr}(\cdot)$	trace of matrix
$\otimes$	Kronecker product
$\bar{A}$ (or A^{c})	complement of set A
$A \cup B$	union of A and B
$A \subset B$	A is included in B
$A \cap B$	intersection of A and B
$A \setminus B$	set of elements of A that are not included in B
$\mathrm{Card}(A)$	cardinality of A
$\varnothing$	empty set
$\mathrm{dist}(A, B)$	euclidian distance between A and B
$\mathbb{I}(\cdot)$ (or $\mathbb{I}_A(\cdot)$)	indicator function (of set A)
$L^2(\mathcal{T})$	space of square-integrable functions in interval $\mathcal{T}$
$\sigma(\ldots)$	σ-algebra generated by $(\ldots)$
$(\Omega, \mathcal{A}, P)$	probability space Ω : nonempty set $\mathcal{A}$:σ-algebra of subset of Ω P: probability measure on $\mathcal{A}$
i.i.d	independent and identically distributed
$\mathcal{N}(0, 1)$	normal distribution
$u_n = O(v_n)$	a constant c exists such that $u_n \leq c v_n$
$u_n = o(v_n)$	$\frac{u_n}{v_n} \to 0$ as $n \to \infty$
■	end of a proof

Chapter 1

General introduction

This thesis was supported by a PhD scholarship of ministry of higher education and scientific research of Mauritania and a scholarship from the French government (Campus-France).

This thesis is about statistical inference for spatial and/or functional data. Indeed, we are interested in modelization of unknown parameters of some population from random or non-random (stratified) samples composed of independent or spatially dependent variables.

In one hand, the samples used in statistical inference are basically of random nature. Non-random samples are useful in a number of situations. In fact, in choice models, the dependent variable is discrete and the partitioner prefer to have a sample with all possible values of the dependent variable, especially when one or more outcomes occur infrequently in the population but are important to determine some key parameters of the model. This can be ensured by the concept of case-control or choice-based sampling that consists to stratify the population with respect to the values of the categorical response. By stratifying the population with respect to the responses, one can gather information on those infrequent outcomes at a much lower cost than would be incurred by simply increasing the size of a random sample.

On the other hand, in a number of disciplines, such as environmental sciences, economics, hydrology, medical studies, neuroimaging and genomics, mining industry,..., data are available at several spatial locations (geographical, voxels,...). Spatial statistics embody a suite of methods for analyzing spatial data and for instance estimating the values of a property of interest at non-sampled locations, from available sample data points using spatial correlation tools. Among the practical considerations that influence the available techniques used in the spatial data modeling, is the data dependency. In fact, spatial data are often dependent and a spatial model must be able to handle this aspect.

Nowadays, modern technology has facilitated the monitoring of very large time and/or spatial datasets, particularly functional data. The last two decades have seen an emergence of new area in statistics as functional data analysis (FDA). FDA is concerned with data objects, such as curves, shapes, images or a more complex mathematical object, thought as smooth realizations of a stochastic process. Since this last is an infinite dimensional object, functional data are part of Big Data. In these contexts, developing models and methods able to account features of datasets of interest as non-random nature, high or infinite dimension and spatial correlation of sample, seems to be essential.

To this aim, the thesis is composed of 7 chapters. The first concerns fundamental concepts and a state of art on the models and methods used. The following two chapters give our contributions on functional data analysis. Namely, we propose a functional choice model in a context of stratified sampling and a functional autoregressive spatial model, respectively. The fifth and sixth chapters concern two spatial regression models (a kernel nearest neighbor method and a partially linear model, respectively) involving real-valued processes. The last chapter is an applied one on upper aerodigestive cancer risk factors in the north of France. Some conclusions and perspectives end the document.

Written and oral communications

Works and publications

- Binary Functional Linear Models Under Choice-Based Sampling (In collaboration with M.K.Attouch and S. Dabo-Niang), *Econometrics and Statistics*, 2017. In press, available on-line `https://doi.org/10.1016/j.ecosta.2017.07.001`.
- Nonparametric Prediction By k-Nearest Neighbour Method For Spatial Data (In collaboration with M.K. Attouch, S. Dabo-Niang and M. N'diaye). *Journal de la Société Française de Statistique*, 2017. In revision.
- Functional Linear Spatial Autoregressive Models (In collaboration with L. Broze, S. Dabo-Niang and Z. Gharbi), 2017. Submitted.
- Partially Linear Spatial Probit Models (In collaboration with S. Dabo-Niang). To submit
- Outlier Detection in Functional Framwork, Application To Temperature Data (In collaboration with F. Chebana, S. Dabo-Niang, and T.B.M.J. Ouarda). In progress.
- Identification of the determinants of UADT cancers incidence in French Northern Region (In collaboration with S. Dabo-Niang, E. Darwich and J. Foncel). In progress.
- Sur l'estimation de la fonction de régression par la méthode des k plus proches voisins dans le cas de données spatiales (In French) [On estimation of regression function by k nearest neighbors method in case of spatial data]. Master thesis, 2013.

Seminars and Conferences

- 61st World Statistics Congress, Marrakech, Morocco, "Partially Linear Spatial Probit Models". July 2017.
- $7^{\text{ièmes}}$ Rencontres des Jeunes Statisticiens, Porquorelles, France, "Partially Linear Spatial Probit Models". Avril 2017.
- CIMPA school at Gasten Berger university , Saint–Louis, Senegal, "Functional Binary Choice Models Under Choice-Based Sampling". Avril 2016.
- $15^{\text{ièmes}}$Forum de jeunes mathématicien-n-e-s, Lille, France, "Functional Binary Choice Models Under Choice-Based Sampling". Novembre 2015.
- $47^{\text{ièmes}}$ Journées de Statistique de la Société française de Statistique, Lille, France, "Nonparametric Prediction By k-Nearest Neighbor Method For Spatial Data". Juin 2015.

Participation in conferences, seminars and workshops

- "Thursday's" Econometrics and Statistical seminars, LEM UMR 9221 CNRS (before laboratory EQUIPPE), Lille (France), 2014-2017
- Lille research workshop on Statistics and econometrics, Lille (France), May 2016.
- Symposium "Statistical Methods for Recurrent Data workshop", Lille (France). December 2016.
- Rencontre des jeunes chercheurs africains en France organized by the SFdS, Paris (France). November 2014.

Chapter 2

State of art and general concepts

Contents

This chapter gives a general introduction and state of art on the main contributions of this thesis on the three fundamental concepts; Functional Data Analysis (FDA), Choice-based sampling and Spatial models.

2.1 Functional data analysis

The last two decades saw the emergence of a new branch of statistics named Functional Data Analysis (FDA), popularized by the monographs of Ramsay & Silverman (2005), Bosq (2000), Ferraty & Vieu (2006), Horváth & Kokoszka (2012) and Hsing & Eubank (2015). This fields deals with curves, shapes or more complex mathematical objects of infinite dimension (see Cardot et al., 2003, for image processing). In fact, functional data (FD) are considered as observations of stochastic processes of infinite dimension, this makes FD as part of Big Data. Such data occur in many areas such as medicine (growth curves), hydrology (flows), genetics (genetic sequence), among others. The collection of such massive data is facilitated by technological advances (recordings capacity,...).
To treat such data, it is necessary to develop statistical methods (visualization, modeling,...) able to handle large or infinite dimension of data, since statistical methods for multivariate data have difficulties to deal with with large dimension. This may be due to strong collinearity between variables or infinity of solutions of a system of equations with a number of unknowns more than the number of equations. This is usually the case when the dimension of the covariates is larger than the number of observations.

As for multivariate statistics, various exploration and modeling techniques adapted to the nature of considered data have been proposed for functional variables during the last two decades. When considering regression models, one can basically distinguish two

popular approaches, parametric (Ramsay & Silverman, 2005) and nonparametric models Ferraty & Vieu (2006).
In the context of this thesis, we will focus on parametric models with a functional covariate, particularly functional linear models with scalar response and functional binary choice models.

In the following, we briefly introduce generalized functional linear models as they regroup our two functional models of interest.

2.1.1 Generalized functional linear models

Functional linear models for scalar response were originally introduced by Hastie & Mallows (1993), while binary functional models were considered in James & Hastie (2001). More recently, Müller & Stadtmüller (2005) and Cardot & Sarda (2005) introduced the famous generalized functional linear models, which can be viewed as a generalization of the previous two models.
Consider that we have a sample of n i.i.d observations $(Y_i, \{X_i(t),\, t \in \mathcal{T}\})\,,\ i = 1, \ldots, n$, where Y is a real response variable (it may take values in $\{0, 1\}$) and $\{X(t), t \in \mathcal{T}\}$ is a random function that corresponds to a stochastic process on the interval $\mathcal{T} \subset \mathbb{R}$, taking values in the space $\mathcal{X} \subset L^2(\mathcal{T})$ of square integrable functions in $\mathcal{T}$. We are interested in describing the relation between the response variable Y and the explanatory random function $X(\cdot)$. We assume that this relation is given by a regression problem

$$E\left(Y|\left\{X(t),\, t \in \mathcal{T}\right\}\right) = \Phi\left(\eta^*\right) \qquad \text{and} \qquad \mathrm{Var}\left(Y|\left\{X(t),\, t \in \mathcal{T}\right\}\right) = \tilde{\sigma}^2(\eta^*), \tag{2.1}$$

where the linear predictor η^* is defined by:

$$\eta^* = \alpha^* + \int_{\mathcal{T}} X(t)\theta^*(t)dt.$$

The link function $\Phi(\cdot)$ is some strictly increasing cumulative distribution function and $\tilde{\sigma}^2(\cdot) = \sigma^2(\Phi(\cdot))$ with $\sigma^2(\cdot)$ is some positive function. In case of linear models, $\Phi(\cdot)$ is defined by the identity function and $\sigma^2(\cdot)$ is defined by a constant function. For binary models, $\sigma^2(t) = t(1-t)$.
The parameters of interest to be estimated are the intercept α^* and the parameter function $\theta^*(\cdot)$, assumed to belong to the space of functions $L^2(\mathcal{T})$.
Assume that these n observations satisfy the quasi-likelihood functional model:

$$Y_i = \Phi(\eta_i^*) + U_i\tilde{\sigma}^2(\eta^*), \qquad i = 1, \ldots, n, \tag{2.2}$$

where the error term satisfies $E\left(U_i|\left\{X_i(t),\, t \in \mathcal{T}\right\}\right) = 0$ and $E\left(U_i^2\right) = 1$, $i = 1, \ldots, n$.
The infinite dimension of the functional variable $X(\cdot)$ is always the first problem of the estimation procedures. Two very popular approaches of dimension reduction are used in generalized linear models with explanatory random functions. On the one hand, we have the Penalized Likelihood Method (Cardot & Sarda, 2005). It consists of projecting the parameter function $\theta^*(.)$ into a finite-dimensional space, spanned by a spline basis and maximizing the pseudo conditional log-likelihood function obtained by replacing the parameter function with its projector, adding a penalty that controls the degree of smoothness $\theta^*(.)$.
On the other hand, we have the approach used by Müller & Stadtmüller (2005). It is based on a truncation strategy that consists of projecting the functional explanatory variable and parameter function into a space of functions generated by a basis of functions with a dimension that increases asymptotically as the sample size tends towards infinity.

In this dissertation, we use the second approach, recalled in the following.

In fact, the approach of Müller & Stadtmüller (2005), is based on some truncation strategy which is motivated by the following considerations. Let $\{\varphi_j,\, j = 1, 2, \ldots\}$ be an orthonormal basis of the functional space $L^2(\mathcal{T})$, usually a Fourier or spline basis or a basis constructed from the eigenfunctions of the covariance operator of covariate $X(\cdot)$. One can rewrite $X(t)$ and $\theta^*(t)$ as follows:

$$X(t) = \sum_{j\geq 1} \varepsilon_j \varphi_j(t) \qquad \text{and} \qquad \theta^*(t) = \sum_{j\geq 1} \theta_j^* \varphi_j(t),$$

where the real random variables ε_j and the coefficients θ_j^* are given by

$$\varepsilon_j = \int_{\mathcal{T}} X(t)\varphi_j(t)dt \qquad \text{and} \qquad \theta_j^* = \int_{\mathcal{T}} \theta^*(t)\varphi_j(t)dt.$$

By the orthonormality of the basis $\{\varphi_j,\, j = 1, 2, \ldots\}$, we have

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j\geq 1} \theta_j^* \varepsilon_j.$$

Let p_n be a positive sequence of integers that increases asymptotically as $n \to \infty$, and let us consider the following decomposition:

$$\eta^* = \alpha^* + \sum_{j=1}^{\infty} \theta_j^* \varepsilon_j, \qquad \tilde{\eta}^* = \alpha^* + \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j, \qquad \eta^* - \tilde{\eta}^* = \sum_{j=p_n+1}^{\infty} \theta_j^* \varepsilon_j.$$

The truncation strategy introduced by Müller & Stadtmüller (2005) is based on the following approximation, under the assumption that $\|\Phi'\| < C$:

$$E\left((\Phi(\eta^*) - \Phi(\tilde{\eta}^*))^2\right) \leq CE\left((\eta^* - \tilde{\eta}^*)^2\right). \tag{2.3}$$

Similar to the expectation approximation function, the error of approximation in term of variance $E\left((\tilde{\sigma}^2(\eta^*) - \tilde{\sigma}^2(\tilde{\eta}^*))^2\right)$ is also majored by the same term if $\|\sigma^{2'}(\cdot)\| < C$. Assuming that the right-hand side of (2.3) vanishes asymptotically as $n \to \infty$, one may instead of (2.2), work with the approximated sequence of models:

$$Y_i^{(p_n)} = \Phi(\tilde{\eta}_i^*) + U_i \tilde{\sigma}^2(\tilde{\eta}_i^*), \qquad i = 1, \ldots, n, \tag{2.4}$$

where $\tilde{\eta}_i^* = \alpha^* + \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j^{(i)}$, with $\varepsilon_j^{(i)} = \int_{\mathcal{T}} X_i(t)\varphi_j(t)dt$.

The parameters of interest in the truncated model are then the intercept, α^*, and the first p_n coefficients of the parameter function, $\theta_1^*, \ldots, \theta_{p_n}^*$. For simplicity, let $\theta^\dagger = (\alpha^*, \theta_1^*, \ldots, \theta_{p_n}^*)^T$. Therefore an estimator of the $1 \times p_n$ vector $\theta^\dagger$ can be defined by solving the gradient associated to the quasi-likelihood function of truncated models (2.4), defined by:

$$\Delta(\theta) = \sum_{i=1}^{n} \frac{\Phi'(\tilde{\eta}_i)}{\tilde{\sigma}^2(\tilde{\eta}_i)} (Y_i - \Phi(\tilde{\eta}_i))\varepsilon^{(i)} = 0, \tag{2.5}$$

with respect to $\theta \in \mathbb{R}^{p_n+1}$, where $\tilde{\eta}_i = \theta^T \varepsilon^{(i)}$ and $\varepsilon^{(i)} = (\varepsilon_1^{(i)}, \ldots, \varepsilon_{p_n}^{(i)})^T$ with $\varepsilon_j^{(i)} = \int X_i(t)\varphi_j(t)dt$ and $\varepsilon_0^{(i)} = 1$.

Consequently, let the solution of (2.5) be $\widehat{\theta^\dagger} = (\hat{\alpha}, \hat{\theta}_1, \ldots, \hat{\theta}_{p_n})^T$. Therefore, an estimator of the intercept α^* is $\hat{\alpha}$, and that of the truncated parameter function is given by

$$\hat{\theta}(t) = \sum_{j=1}^{p_n} \hat{\theta}_j \varphi_j(t).$$

Müller & Stadtmüller (2005) established the asymptotic normality of $\widehat{\theta^\dagger}$ under some assumptions, as well as that of the distance between $\hat{\theta}(\cdot)$ and $\theta^*(\cdot)$ with respect to some $L^2(\mathcal{T})$ metric defined via a generalized covariance operator.

2.2 Choice-based sampling

Choice-based sampling (CBS) or case-control sampling design is of particular interest in the context of choice models. It consists of stratifying a population of interest with respect to the values of the categorical response. By this stratification, one can gather information on infrequent outcomes at a much lower cost than would be incurred by simply increasing the size of using a random sample. Equivalently, for any given sampling budget, one can increase the efficiency of predictions and parameter estimates using a suitably designed response-based sample. Such sampling designs have been independently investigated by econometricians (using the term choice-based sampling) who study choice behaviour and biostatisticians (with the term case-control) who are interested in rare diseases.

In biostatistics, case-control designs are useful for identifying the impact of several factors on the occurrence of a particular disease. The response is often binary (having the disease or not), but there may be more than two categorical responses. In such design, separate samples of cases (diseased individuals) and controls (individuals without the disease) are selected, unlike in a prospective study design, in which a sample of individuals is chosen and followed through time until their responses are recorded, figure 2.1 illustrates the difference between these two studies compared to the timeline. In the case of rare diseases, even large studies may produce only a few diseased individuals and little information about the hazard. In that case, the researcher might wish to oversample the rare disease of interest to increase the accuracy of his analysis. Therefore, compared with case-control studies, prospective studies are disadvantageous in terms of time and cost. For a general overview of medical case-control studies, see Keogh & Cox (2014), for instance. Case-control studies are also used in political science (King & Zeng, 2001) and sociology (Xie & Manski, 1989).

Originally, choice-based sampling was used by econometricians. They were interested in exploring the relationships between the choices made by an individual and several explanatory variables. The choice of transportation mode is the most popular example; see, for instance, Manski & McFadden (1981). In this case, it may be advantageous (simpler and less expensive) to apply choice-based sampling by selecting, for instance, separate samples of individuals from bus terminals, train stations and car parks rather than to take a single sample from the entire population. The reasons for using such stratified samples have been discussed extensively in several econometric papers such as Manski & Lerman (1977), Manski & McFadden (1981), Cosslett (1981), Imbens (1992), and Cosslett (2013).

Let us describe the choice-based sampling process. Let us follow the notations of Imbens (1992). Consider that in a given population we observe a discrete random variable Y with values in $C = \{1, 2, \ldots, M\}$ and a continuous or discrete random vector X valued in $\mathcal{X} \subset \mathbb{R}^p$, and assume that the joint density of these two variables is

$$f(i, x) = P(i|x, \theta) \cdot r(x),\ x \in \mathcal{X},\ i \in C \tag{2.6}$$

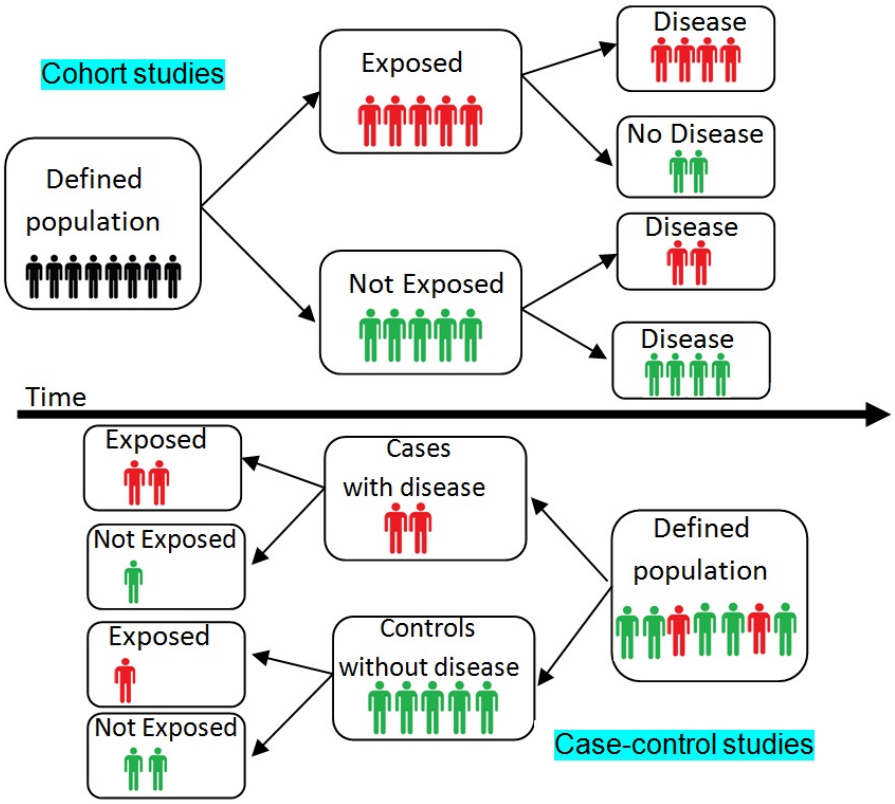

Figure 2.1: A graphical comparison between prospective and retrospective studies

for $i \in C = \{1, 2, \ldots, M\}$, $x \in \mathcal{X} \subset \mathbb{R}^p$, and $\theta \in \Theta \subset \mathbb{R}^k$. $P(\cdot|\cdot, \cdot)$ is a known function and θ is an unknown parameter vector. The distribution function of X will be denoted by $R(.)$ while the density is $r(.)$. The partitioner is interested in the parameter θ of the conditional probabilities but one might also be interested in $Q(i)$, the marginal probability or population share of choice $Y = i$. Even if one is not interested in $Q(i)$ itself, it is useful to define it explicitly. This will make easier to incorporate prior information about the marginal probability and such prior information (in particular for rare choices) is often a motivation for sampling according to some choice-based manner. Let the true value of θ be θ^* and the corresponding $Q(i)$ be $Q^*(i)$:

$$Q^*(i) = \int_{\mathcal{X}} P(i|x, \theta^*) dR(x), \; x \in \mathcal{X}, \; i \in C.$$

More generally, assume that the population of interest is divided, according to the values of the discrete variable Y, into S stratas, $\mathcal{J}(s) = \{(i, x), i \in C, \; x \in \mathcal{X}\}$ for $s = 1, 2, \ldots, S$ and we let H_s be the probability with which one will draw from stratum $\mathcal{J}(s)$. The probabilities H_s satisfy $\sum_{s=1}^{S} H_s = 1, H_s > 0$. We assume the sample is chosen as follows: *select an observation by first drawing a stratum* $s \in \{1, 2, \ldots, S\}$ *with probability* H_s *and then draw randomly an observation from* $\mathcal{J}(s)$. If $S = M$ and $\mathcal{J}(s) = \{(s, x), \; x \in \mathcal{X}\}$, the sampling is known as pure choice-based sampling. Note that H_s is not always known to the investigator, let H_s^* denote the true values corresponding to above model.
To look at the effect of the sampling process on the marginals probabilities H_s and $Q(i)$, let $H(i)$ and Q_s be:

$$Q_s = \sum_{i \in \mathcal{J}(s)} Q(i), \qquad H(i) = Q(i) \sum_{s|i \in \mathcal{J}(s)} \frac{H_s}{Q_s}.$$

If there is no s such that $i \in \mathcal{J}(s)$, then $H(i) = 0$. $H(i)$ is the marginal probability of choice i induced by the choice-based sampling. In the population, the marginal probability of choice i is $Q(i)$, but with the sampling scheme, one has to multiply it by the sum of the bias factors H_s/Q_s. Q_s is the marginal probability with which an observation randomly drawn from the population is in $\mathcal{J}(s)$. Note that we have in the case of pure choice-based sampling, $H(i) = H_i$ and $Q_i = Q(i)$ for $i \in C$.

Under the sampling process, the joint probability of stratum, the covariate and the response, is the product of the marginal probability of H_s, and the conditional density of Y and X given the stratum s. The latter is

$$g(i, x|s) = \frac{f(i,x)}{\sum_{j\in\mathcal{J}(s)} \int_{\mathcal{X}} f(j,y)dy} = \frac{P(i|x,\,\theta)\cdot r(x)}{\sum_{j\in\mathcal{J}(s)} \int_{\mathcal{X}} P(j|y,\,\theta)dR(y)dy},\ x \in \mathcal{X},\ i \in C \tag{2.7}$$

and the joint density can be written as

$$g(s, i, x) = H_s \frac{P(i|x,\,\theta)\cdot r(x)}{\sum_{j\in\mathcal{J}(s)} \int_{\mathcal{X}} P(j|y,\,\theta)dR(y)dy} = H_s \frac{P(i|x,\,\theta)\cdot r(x)}{\sum_{j\in\mathcal{J}(s)} Q(j)}, \tag{2.8}$$

for $i \in C, s \in \{1, 2, \ldots, S\}$, and $x \in \mathcal{X}$. Consequently, the conditional probability of Y given X in the sample is

$$g(i|x) = \frac{P(i|x,\theta)H(i)/Q(i)}{\sum_{j=1}^{M} P(j|x,\theta)H(j)/Q(j)}. \tag{2.9}$$

Several procedures have been proposed to estimate the parameter of interest θ^* by using a sample of N i.i.d observations $\{(s_n, i_n, x_n),\ n = 1, \ldots, N\}$ drawn through the above CBS process. For instance, Manski & McFadden (1981) proposed maximizing a conditional likelihood function as a function of (2.9) given knowledge on $Q^*(i)$ and H_s^*. They proved the consistency and asymptotic normality of their estimator. Cosslett (1981) proposed a pseudo maximum likelihood estimator. He considers the likelihood function based on the density (2.7). He investigated asymptotic properties of the estimator as well as its asymptotic normality. The estimator proposed by Cosslett (1981) is efficient in the class of asymptotically unbiased estimator but it is very hard to compute. Imbens (1992) proposed a generalized method of moments estimate. He defined a criterion function based on the gradient's equations associated to the logarithm of a likelihood function based on (2.8). He showed the efficiency of its estimator and its flexible computation.

2.2.1 Functional data with choice-based sampling

In functional data analysis, to the best of our knowledge, only two works deal with survey sampling techniques. In fact, Cardot et al. (2010) generalized the functional principal components analysis (FPCA) to functional objects collected through survey sampling techniques. They proposed estimators of the eigen-elements (of the variance-covariance operator) as well as their variance and proved under some assumptions that these estimators are asymptotically unbiased and consistent. Cardot & Josserand (2011) propose estimators of the mean and variance functions of FDA objects based on Horvitz-Thompson estimator under stratified sampling. They investigated under some assumptions on the sampling design, the uniform consistency of the proposed estimators and stated a functional central limit theorem and deduced asymptotic confidence bands.
Despite many potential applications, no work has been done on functional choice models in the context of case-control or choice-based sampling. Note that, one work (Fan et al., 2014)

addresses a functional logit model applied to case-control data. These authors propose a functional logit model to test the associations between a dichotomous trait and multiple genetic variants in a region using several covariates. However, they do not take into account in the estimation procedure the case-control nature of the data. Consequently, their model is similar to a classical generalized functional linear model for random sampling.

2.2.2 Contribution in choice-based sampling

In the first part of this thesis, we are interested in a functional binary choice model when the data are from a choice-based sampling design. In fact, we present in Chapter 3, a model in which the response is binary, the explanatory variable is functional, and the sample is obtained by a pure choice-based sampling process. We propose a conditional likelihood function under the sampling distribution and a dimensional reduction of the space of the explanatory random function based on a Karhunen–Loève expansion (Müller & Stadtmüller, 2005) to define a feasible estimator (Manski & McFadden, 1981) of the proposed model. Several asymptotic properties are given. A simulation study and an application to kneading data are used to compare the proposed method with the ordinary maximum likelihood method, which ignores the nature of the sampling.

2.3 Spatial data

Agriculture, economics, environmental sciences, urban systems and epidemiology activities are often located in space. Therefore, modeling such activities requires to find a correlation structure between data observed at a given location and that available at neighboring locations. This is a significant feature of spatial data analysis. Three main ways of incorporating a spatial dependence structure (see for instance Cressie, 1993) can be distinguished, basically for geostatistics data, lattice data and point patterns. In the domain of geostatistics, the spatial location is valued in a continuous set of $\mathbb{R}^N$, $N \geq 2$. For spatial lattice data, the locations form a lattice set. Compare to geostatistical and lattice data, spatial point patterns occurs when the locations where the data are available are random. It is not always easy to distinguish these three types of data. Several methods for analyzing spatial data and estimating the values of a property of interest at non-sampled locations, from available sample data points have been proposed during the last fifty years. Among the practical considerations that influence the available techniques used in the spatial data modeling, is the data dependency. In fact, spatial data are often dependent and a spatial model must be able to handle this aspect.

Originally developed for the mining industry, spatial statistics were first developed for spatial prediction of geological resources over two or three dimensional areas. Indeed, a crucial need in many scientific disciplines where spatial data are available is the prediction of a variable at an unobserved location, using observations available at other locations. Hence, as for time-dependent data, there is a need to measure the dependence between neighboring locations. The main difference between spatial and time series data is the absence of an order relation like the notions of past, present and future: the axis of time is unidirectional. Indeed, past events may have an influence on the future while the reverse is not true. Thus time series models cannot be directly applied to spatial data. Differences and similarities between spatial and time series data are highlighted in Tjøstheim (1987).

In this dissertation, we are interest to regression models for geostatistical and lattice data in a *fixed-design* context, where the latter means that locations at which the spatial phenomenon is recorded, are selected non-randomly. For the estimators we propose, the asymptotic results are given according to *Increasing domain* asymptotic (for more detail

see Gaetan & Guyon, 2008, Chapter 5). It consists of a sampling structure where new observations are added at the edges (boundary points) compare to *infill* asymptotic, that consists of a sampling structure where new observations are added in between existing ones. Infill asymptotic is appropriate when the spatial locations are in a bounded domain.

In the following, we introduce briefly some spatial models related to the contributions of this thesis, with the corresponding estimation methods.

2.3.1 Spatial parametric estimation

We consider that at n spatial locations $\{s_1, s_2, \ldots, s_n\}$ satisfying $\|s_i - s_j\| > \rho$ with $\rho > 0$, observations of a random vector (Y, X) are available. Let X be an explanatory random variable taking values in $\mathbb{R}^p$ and Y is an univariate response variable of interest. When the latter takes binary values in $\{0, 1\}$ (we will be concern with binary choice-models, in particular, Probit models will be considered in this dissertation), we assume that it is associated to a latent dependent variable Y^*, then the observations will be:

$$Y_{s_i} = \mathbb{I}\left(Y^*_{s_i} \geq 0\right), \qquad i = 1, \ldots, n. \tag{2.10}$$

To facilitate the notation, we will denote in this section i for individual in location s_i and adopt the notations: $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^T$, and $\mathbf{X}_n$ the $n \times p$ matrix of explanatory variables with elements X_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, p$.

In spatial econometrics's literature, the spatial dependency is usually modeled by using a spatial linear process defined through a spatial weight matrix. Basically, the latter is a $n \times n$ non-stochastic weight matrix W_n, that allows to describe the spatial interactions between the n spatial units. The elements $w_{ij} = w_{ij,n}$ of this matrix are usually considered as inversely proportional to the distance between spatial units i and j with respect to some metric (physical distance, social network or economic distance, see for instance Pinkse & Slade, 1998). More precisely, the matrices W_n can be classified into two groups: *Weights Based on Distance* and *Weights Based on Boundaries*. For Weights Based on Distance, one way to construct spatial weight matrices is to use the distance d_{ij} between each pair of spatial units (regions, cities, centroids,...) i and j.

- *k-Nearest Neighbor weights*
  $w_{ij} = \begin{cases} 1 & \texttt{if } j \in N_k(i), \\ 0 & \texttt{Otherwise} \end{cases}$ where $N_k(i)$ is the set of the k closest units or regions to i for $k \in \{1, ..., n-1\}$

- *Radial Distance weights*
  $w_{ij} = \begin{cases} 1 & \texttt{if } 0 \leq d_{ij} \leq \delta \\ 0 & \texttt{if } d_{ij} > \delta \end{cases}$, where d_{ij} is the euclidian distance between units i and j, and δ is a critical distance (*threshold distance* or *bandwidth*) cut-off after which spatial effects are considered to be negligible, it should be able to guarantee that each region has at least one neighbor.

- *Power Distance Decay weights*
  $w_{ij} = \begin{cases} d_{ij}^{-\alpha} & \texttt{if } 0 \leq d_{ij} \leq \delta \\ 0 & \texttt{if } d_{ij} > \delta \end{cases}$, where α is any positive exponent, typically $\alpha = 1$ or $\alpha = 2$.

- *Exponential Distance Decay weights*
  $w_{ij} = \begin{cases} \exp(-\alpha d_{ij}) & \texttt{if } 0 \leq d_{ij} \leq \delta \\ 0 & \texttt{if } d_{ij} > \delta \end{cases}$

- *Double-Power Distance weights*
  $w_{ij} = \begin{cases} [1-(d_{ij}/\delta)]^k & \texttt{if } 0 \leq d_{ij} \leq \delta \\ 0 & \texttt{if } d_{ij} > \delta \end{cases}$, with k is a positive integer, typically $k = 2$, $k = 3$ or $k = 4$.

- *Cliff-Ord weights*
  Cliff & Ord (1973) suggested to use the length of the common border between contiguous regions, weighted by a distance function:

  $$w_{ij} = d_{ij}^{-a} D_{ij}^{b}$$

  where D_{ij} is the share of common boundary between i and j, a and b are parameters estimated from data or chosen a priori.

- *Block structure*
  In this case $w_{ij} = 1$ for all i and j in the same block. And the blocks are defined according to some specific criterion.

For Weights Based on Boundaries, spatial contiguity is often used to specify neighboring location in the sense of sharing a common border. There are different type of spatial contiguity but the classical cases are those referred to *Rook contiguity* (with only common boundaries), *Bishop contiguity* (with only common vertices) and *Queen contiguity* (with both Rook and Bishop contiguity).

$w_{ij} = \begin{cases} 1 & \texttt{if } i \texttt{ and } j \texttt{ are contiguity} \\ 0 & \texttt{Otherwise} \end{cases}$

In general, we can rewrite the last equation as:

$w_{ij} = \begin{cases} 1 & \ell_{ij} > 0 \\ 0 & \ell_{ij} = 0 \end{cases}$,

with ℓ_{ij} denotes the length of shared boundary.

By using one of these spatial weight matrices, one can mainly distinguish three different types of interaction effects that may explain why an observation associated with a specific location may be dependent on observations at other locations:

- Endogenous interaction effects, where the variable Y (or Y^* for the latent model) at some spatial unit depends on values of Y taken by other spatial units. This means that the interaction is among the dependent variable, it is the so-called spatial autoregressive (SAR) model (Cliff & Ord, 1973).

- Exogenous interaction effects, where the variable Y (or Y^*) at some spatial unit depends on independent explanatory variables at other spatial units. This means the the interaction is among the explanatory variable.

- Correlated effects, where similar unobserved characteristics result in similar behavior. This means that the interaction is among the error terms, this model is called spatial autoregressive error (SAE) model (or spatial error model; SEM). One motivation might be some spatial heterogeneity.

Note that in practice, it is rare or maybe impossible to find a population that contains these three types of interaction together. In fact, researchers have always been focused on models that contain one interaction, whether SAR model, SAE model, or a model with two kind of interaction. The latter model may be used when the spatial autocorrelation

can affect both response and error terms. According to the terminology developed by LeSage (2008), we refer to this model by spatial autocorrelation (SAC) model:

$$\mathbf{Y}_n = \lambda_0 W_n \mathbf{Y}_n + \mathbf{X}_n \beta_0 + \mathbf{U}_n; \qquad \mathbf{U}_n = \gamma_0 W_n \mathbf{U}_n + \varepsilon_n, \qquad \varepsilon_n \sim N(0, \sigma_0^2 I_n),$$

where $\mathbf{U}_n = (U_1, \ldots, U_n)^T$ and $\varepsilon_n = (\varepsilon_1, \ldots, \varepsilon_n)^T$. The coefficients λ_0 and γ_0 are scalar autoregressive parameters indicating the degree of spatial dependence, β_0 is a $p \times 1$ vector of parameters. $W_n \mathbf{Y}_n$ is the spatial lag, it denotes the endogenous interaction effects among the dependent variables, i.e. for each observation Y_i, the corresponding element in $W_n \mathbf{Y}_n$ gives weighted sum of Y_j, $j \neq i$, with weights given by the relative connectivity from j to i. $W_n \mathbf{U}_n$ is the interaction effects among the disturbance terms of the different spatial units. However, SAR model is a SAC model with $\gamma_0 = 0$ and SAE model is a SAC model with $\lambda_0 = 0$.

2.3.1.1 Estimation in SAR models with continuous response variable

SAR models for real-valued data and their identification and estimation methods have been developed by two stage least squares (2SLS) (Kelejian & Prucha, 1998; Lee, 2007), maximum likelihood (ML) (Ord, 1975) and generalized method of moments (GMM) (Smirnov & Anselin, 2001), among others. The identification and estimation of SAR models by quasi-maximum likelihood (QML) are limited. Lee (2004) and more recently Yang & Lee (2017), proposed quasi-maximum likelihood estimators for a SAR model with a spatial dependency structure based on a spatial weights matrix. The quasi-maximum likelihood estimator (QMLE) is appropriate when the disturbances in the considered model are not normally distributed. In the literature on SAR models for real-valued data, the QMLE and maximum likelihood estimator (MLE) are proved to be computationally challenging, consistent with rates of convergence depending on the spatial weights matrix of the considered model (Lee, 2004; Yang & Lee, 2017). We recall the principle of this method as it will be explored in Chapter 4.
Let us first recall the definition of SAR models:

$$\mathbf{Y}_n = \lambda_0 W_n \mathbf{Y}_n + \mathbf{X}_n \beta_0 + \mathbf{U}_n; \qquad \mathbf{U}_n \sim N(0, \sigma_0^2 I_n). \tag{2.11}$$

Let $S_n(\lambda) = I_n - \lambda W_n$, $\theta = (\beta', \lambda, \sigma^2)'$,$\delta = (\beta', \lambda)'$, and $\mathbf{U}_n(\delta) = \mathbf{Y}_n - \mathbf{X}_n\beta - \lambda W_n \mathbf{Y}_n$. Thus $\mathbf{U}_n = \mathbf{U}_n(\delta_0)$. Therefore the logarithm of the quasi-likelihood function of (2.11) is

$$L_n(\theta) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\mathbf{U}_n'(\delta)\mathbf{U}_n(\delta). \tag{2.12}$$

For a fixed λ, (2.12) is maximized at

$$\hat{\beta}_{n,\lambda} = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n' S_n(\lambda)\mathbf{Y}_n,$$

and

$$\begin{aligned} \hat{\sigma}^2_{n,\lambda} &= \frac{1}{n}\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\beta}_{n,\lambda}\right)'\left(S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n\hat{\beta}_{n,\lambda}\right) \\ &= \frac{1}{n}\mathbf{Y}_n' S_n'(\lambda)\left(I_n - \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\right)S_n(\lambda)\mathbf{Y}_n. \end{aligned}$$

Then the concentrated log-quasi-likelihood function of λ is:

$$L_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}^2_{n,\lambda} + \ln|S_n(\lambda)|.$$

The estimator of λ_0 is $\hat{\lambda}_n$, which maximizes $L_n(\lambda)$, and those of the vector β_0 and variance σ_0^2 are respectively $\hat{\beta}_{n,\hat{\lambda}_n}$, $\hat{\sigma}^2_{n,\hat{\lambda}_n}$.

Lee (2004) investigated the identifications of the parameters β_0, λ_0, and σ_0^2 and the asymptotic properties of the estimators, under some assumptions. He discussed the rates of asymptotic normality, and show that the latter depends on the structure of the spatial weight matrix W_n, particularly when it is constructed such that each units is influenced by few neighboring units. He proved the asymptotic normality of the estimators with optimal rates $\sqrt{n}$.

2.3.1.2 Estimation in SAE models with binary response variable

Extending SAR or SAE models to binary response variable has attracted less attention, only a few number of papers were concerned with this topic over the recent years. This may be, as pointed out by Fleming (2004), due to the "added complexity that spatial dependence introduces into discrete choice models". We recall here some estimation approaches of SAE probit models. Let us recall the definition this type of models, that is:

$$\begin{aligned} \mathbf{Y}_n^* &= \mathbf{X}_n\beta_0 + \mathbf{U}_n; \qquad \mathbf{U}_n = \gamma_0 W_n \mathbf{U}_n + \varepsilon_n, \qquad \varepsilon_n \sim N(0, \sigma_0^2 I_n), \\ Y_i &= \mathbb{I}\left(Y_i^* \geq 0\right), \qquad i = 1, \ldots, n. \end{aligned} \tag{2.13}$$

Assume that the $n \times n$ matrix $(I_n - \lambda_0 W_n)$ is nonsingular for all n, therefore the variance-covariance matrix of $\mathbf{U}_n$ is

$$V_n(\lambda_0) = \mathrm{Var}(U_n) = (I_n - \lambda_0 W_n)^{-1} \left\{ (I_n - \lambda_0 W_n)' \right\}^{-1}.$$

The structure of $V_n(\lambda_0)$ provides the major difficulty of estimating the parameters by a full ML since it requires solving a very computationally demanding problem of n-dimensional integration. Some authors have proposed a feasible maximum likelihood approach which consists of replacing the true likelihood function by a pseudo likelihood function constructed via marginal likelihood functions. Smirnov (2010) proposes a pseudo likelihood function obtained by replacing $V_n(\lambda_0)$ by some diagonal matrix obtained with the diagonal elements of $V_n(\lambda_0)$. Alternatively, Wang et al. (2013) proposed to divide the observations by pairwise groups where the latter are assumed to be independent with bivariate normal distribution in each group and estimated β_0 and λ_0 by maximizing the likelihood of these groups. Other approach is the GMM method used by Pinkse & Slade (1998) which will be recalled in the following as it will be explored in Chapter 6.
By equation (2.13), we have

$$E_0\left(Y_i | X_i\right) = \Phi\left(\left(v_i(\lambda_0)\right)^{-1} X_i' \beta_0 \right), \quad i = 1, \ldots, n, \tag{2.14}$$

where E_0 denotes the expectation under the true parameters (i.e β_0, γ_0), $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution and $(v_i(\gamma_0))^2 = V_{ii}(\gamma_0)$, $i = 1, \cdots, n$ are the diagonal elements of $V_n(\gamma_0)$.
Pinkse & Slade (1998) defined the generalized residuals as:

$$\tilde{U}_i(\theta) = E\left(U_i | Y_i, \theta\right) = \frac{\phi\left(G_i(\theta)\right)\left(Y_i - \Phi\left(G_i(\theta)\right)\right)}{\Phi\left(G_i(\theta)\right)\left(1 - \Phi\left(G_i(\theta)\right)\right)}, \qquad \theta = (\beta', \gamma)', \tag{2.15}$$

where $\phi(\cdot)$ is the density of the standard normal distribution and $G_i(\theta) = \left(v_i(\gamma)\right)^{-1} X_i'\beta$. Note that in (2.15), the generalized residual $\tilde{U}_i(\cdot)$ is calculated by conditioning only on Y_i not on the entire sample $\{Y_i,\, i = 1, 2, \ldots, n\}$ or a subset of it. This of course will

influence the efficiency of the estimators of θ obtained by these generalized residuals, but it allows to avoid a complex computation, see Poirier & Ruud (1988) for more details. To address this loss of efficiency, Pinkse & Slade (1998)'s procedure consists of employing some instrumental variables in order to create some moments conditions and using some random matrix to define a criterion function. Both the instrumental variables and the random matrix permit to take into account more informations about the spatial dependence and heteroscedasticity in the dataset of interest.

Let us now detail the GMM estimation procedure. Let

$$S_n(\theta) = n^{-1}\xi_n^{'}\tilde{\mathbf{U}}_n(\theta), \tag{2.16}$$

where $\tilde{\mathbf{U}}_n(\theta)$ is the $n\times 1$ vector, composed of $\tilde{U}_i(\theta),\ i=1,\ldots,n$ and ξ_n is a $n\times q$ matrix of instrumental variables. The GMM approach consists of minimizing the following sample criterion function,

$$Q_n(\theta) = S_n^{'}(\theta)M_nS_n(\theta),$$

where M_n is some positive-definite $q\times q$ weight matrix that may depend on sample information. Therefore, the GMM estimator $\hat{\theta}$ of θ_0 verifies

$$\hat{\theta} = \text{argmin}_\theta Q_n(\theta).$$

Under some assumptions, Pinkse & Slade (1998) proved the identification of the parameters β_0 and γ_0, and the asymptotic normality of $\hat{\theta}$. They proposed also a consistence estimator of the variance-covariance matrix of $\hat{\theta}$.

Extending these spatial lattice models to functional data is far from being trivial. Ruiz-Medina (2011) and Ruiz-Medina (2012) considered a spatial unilateral autoregressive Hilbertian (SARH(1)) processes where the autoregressive part is given in terms of three functional random components located in three points defining the boundary between some notions of past and future.

2.3.2 Spatial nonparametric estimation

In this thesis, we are also interest to semi and non parametric regression estimation and prediction. To state an art on these type of modeling, let $\{Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}) \in \mathbb{R}^d\times\mathbb{R}\ ,\ \mathbf{i} = (i_1,\ldots,i_N) \in \mathbb{N}^N\}$ $(d\geq 1)$ be a spatial process defined over some probability space $(\Omega,\mathcal{A},\mathbb{P})$, $N\in\mathbb{N}^*$. We assume that the process is observable in $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i}\in\mathbb{N}^N\ :\ 1\leq i_r\leq n_r\ r=1,\ldots,N\}$, $\mathbf{n}=(n_1,\ldots,n_N)\in\mathbb{N}^N$, and $\hat{\mathbf{n}} = n_1\times\ldots\times n_N$, we write $\mathbf{n}\to\infty$ if $\min\{n_r\}\to+\infty$, $n_k/n_i\leq C,\ \forall\ 1\leq k,i\leq N$. We assume that the relation between the two process $(X_{\mathbf{i}},\ \mathbf{i}\in\mathbb{N}^N)$ and $(Y_{\mathbf{i}},\ \mathbf{i}\in\mathbb{N}^N)$ is described by the regression function $r(\cdot) = E\left(Y_{\mathbf{i}}|X_{\mathbf{i}}=\cdot\right)$ where the latter is assumed to be independent of $\mathbf{i}$. We recall in the following different approaches of estimation which allow to estimate the function $r(\cdot)$ in case of geostatistical or lattice data.

2.3.2.1 Kernel estimator of the regression function for spatial data

Compare to parametric modeling, the literature on nonparametric spatial regression is not extensive. Since the seminal work of Tran (1990) on spatial kernel density estimation, a number of papers has been devoted to spatial nonparametric regression and prediction, using particularly kernel methods.

The kernel method was introduced independently by Nadaraya (1964) and Watson (1964) to estimate a regression function from i.i.d observations by a weighted average

of the response variable values. One of the first results on the kernel spatial regression estimation was developed by Lu & Chen (2004). They extended the Nadaraya-Watson estimator or $r(.)$ to spatial data:

$$\hat{r}_{\mathrm{NW}}(x) = \begin{cases} \dfrac{g_{\mathbf{n}}(x)}{f_{\mathbf{n}}(x)} & \text{if } f_{\mathbf{n}}(x) \neq 0, \\ \dfrac{1}{\hat{\mathbf{n}}} \displaystyle\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} & \text{otherwise.} \end{cases}$$

where

$$g_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^{d}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\left(\frac{x - X_{\mathbf{i}}}{h_{\mathbf{n}}}\right) Y_{\mathbf{i}} \qquad \text{and} \qquad f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^{d}} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\left(\frac{x - X_{\mathbf{i}}}{h_{\mathbf{n}}}\right).$$

with $K : \mathbb{R}^d \longrightarrow \mathbb{R}^+$ is a kernel and $h_{\mathbf{n}}$ is a bandwidth such that $h_{\mathbf{n}} \to 0$ as $\mathbf{n} \to +\infty$.

Lu & Chen (2004) established the weak consistency with rate of this estimator under mixing conditions on the considered spatial process in an isotropic framework. Nonparametric prediction of spatial process was first considered in Biau & Cadre (2004). They started by studying the previous regression estimator and proposed a spatial predictor. They gave the uniform almost sure convergence and an asymptotic normality result of the proposed predictor. Carbon et al. (2007) considered a nonparametric autoregressive models for a prediction purpose. These authors proposed a regression model $g(x) = E\left(\psi(X_{\mathbf{i}})|(X_{\mathbf{i}_1}, \ldots, X_{\mathbf{i}_l}) = x\right)$ for $x \in \mathbb{R}^{l \times d}$ and $\psi(\cdot)$ is a real values continuous function, $X_{\mathbf{i}_1}, \ldots, X_{\mathbf{i}_l}$ denote observations at neighbor locations of $\mathbf{i}$. A kernel estimator of $g(x)$ is investigated and the uniform convergence over compacts subsets under spatial mixing condition in addition to some general assumptions, is established. Optimal rates of L^{∞} convergence are also given. A kernel robust estimate of a spatial regression model has been investigated by Gheriballah et al. (2010). They gave an almost complete convergence and an asymptotic normality result of the proposed estimator under some general mixing conditions, as well as a robust procedure to select the smoothing parameter adapted to the spatial structure of the data. All these mentioned papers consider stationary processes. Robinson (2011) employed a basic triangular arrays setting in order to estimate nonparametrically a regression function. This allows to account various kind of spatial variables, in particular non stationary processes. Instead of mixing conditions, a (possibly nonstationary) linear process is assumed for disturbances, and a conditional heteroscedasticity is allowed, as well as non-identically distributed observations. Under sufficient conditions, consistency and asymptotic normality results are obtained.

Recently, Dabo-Niang et al. (2016) proposed a new spatial predictor of a locally identically distributed spatial process. The proposed predictor depends on two kernels in order to control both the distance between observations and that between spatial locations. These authors investigated the uniform almost complete consistency and the asymptotic normality of the kernel predictor under some spatial mixing condition.

2.3.2.2 k-Nearest neighbor method in nonparametric regression

k-Nearest Neighbor method is an alternative to kernel method. The first contribution on k-Nearest Neighbor regression comes back to Stone (1977) who proposed to estimate a regression function from i.i.d observations as the average of the k values of the response variable associate to explanatory variable whose values are the k-Nearest Neighbor to the estimation point. After that, Collomb (1980) proposed to weighting Stone (1977)'e estimate like as the kernel estimator by using a kernel function and call it by k-nearest

neighbor kernel estimator. He investigated the convergence in probability and almost complete for i.i.d observations.

Compare to kernel method, k-Nearest Neighbor (k-NN) method is not very popular in the context of spatial processes. To the best of our knowledge, Li & Tran (2009) is the unique contribution that generalizes the k-NN method to spatial regression. They proposed to estimate the spatial regression function $r(x)$ by:

$$\hat{r}_{\mathbf{n}}(x) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathbb{I}_{V_1}\left(\left(X_{\mathbf{i}} - x\right)/H_{\mathbf{n}}\right) Y_{\mathbf{i}}}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \mathbb{I}_{V_1}\left(\left(X_{\mathbf{i}} - x\right)/H_{\mathbf{n}}\right)}$$

where V_1 is the unit sphere in $\mathbb{R}^d$ and $H_{\mathbf{n}}$ is the distance between x and its $k_{\mathbf{n}}$th nearest neighbors among $X_{\mathbf{i}}$'s, $k_{\mathbf{n}}$ is a fixed integer sequence satisfying $k_{\mathbf{n}} \to \infty$ as $\mathbf{n} \to \infty$. The asymptotic normality of this estimator is obtained under general mixing assumptions.

2.3.2.3 Contribution to spatial nonparametric regression and prediction for real-valued processes

In Chapters 5 and 6 we are interested in nonparametric regression estimation or/and predict spatial process. Firstly, in Chapter 5, we generalize the k-NN method to predict a spatial process at non-observed locations. In this contribution, we consider spatial processes (with a kind of local stationarity) composed by a scalar spatial response variable to predict at some locations and a multivariate spatial covariate. The k-NN kernel predictor proposed is a combination of the k-NN method which consists to introduce a random bandwidth as the kth lower distance between the covariate at the prediction point and covariate's observations and the idea on the predictor proposed by Dabo-Niang et al. (2016) which integrates two kernels one to control distance between observation and the other to control distance between locations. We establish under spatial mixing condition and some general assumptions, the almost complete convergence with rates of the k-NN kernel predictor. Numerical results are given with simulated and environmental data to compare the performance of the spatial k-NN predictor with that of the spatial kernel proposed by Dabo-Niang et al. (2016).

In Chapter 6, we consider a semiparametric model and use the spatial kernel method in a context similar to that considered by Robinson (2011) to estimate a nonparametric component of a partially linear spatial probit model. In this work, the kernel method is integrated in a spatial semiparametric estimation procedure of the proposed model. Precisely, we combine a weighted likelihood method (Staniswalis, 1989) and a generalized method of moments (Pinkse & Slade, 1998) to estimate the model. We first fixe the parametric components of the model and estimate the nonparametric part using a spatial weighted likelihood based on a kernel, the obtained estimate is then used to construct a GMM estimate of the parametric component. Consistency and asymptotic distribution of the estimators are established under sufficient conditions. Some simulated experiments are provided to investigate the finite sample performance of the estimators.

In Chapter 7, we consider an application of spatial binary choice models to identify UADT (Upper aerodigestive) cancer risk factors in the northen region of France which displays the highest rates of such cancer incidence and mortality of the country. This region is characterized by a high proportion of unemployment, social aids and limited resources. Most of the UADT cancers can be considered as preventable, as they are linked to behavior (tobacco, alcohol), environmental factors (industries), and socio-economic conditions or more rarely to genetic familial factors. Local, regional and systemic recurrences are common and often associated with secondary and/or new primaries due to the same risk factors. Those factors are well identified in the literature and we aim to highlight some potential spatial heterogeneity of their distributions.

2.3.2.4 Contribution to spatial functional linear models

Chapter 4 is an contribution to functional linear autoregressive spatial model where the explanatory variable takes values in a function space while the response process is real-valued and spatially autocorrelated. The specificity of the model is due to the functional nature of the explanatory variable and the structure of a spatial weight matrix that defines the spatial dependency between neighbors. The estimation procedure consists of reducing the infinite dimension of the functional explanatory variable and maximizing the quasi-likelihood function. We establish the consistency and asymptotic normality of the estimator. The performance of the methodology is illustrated by simulated data and an application to real data.

Chapter 3

Binary functional linear models under choice-based sampling

Contents

Résumé en français

Les modèles de choix binaire fonctionnels ont été développés pour prédire une variable réponse binaire en fonction d'une variable explicative fonctionnelle ou pour faire la discrimination et la classification des processus stochastiques ou des données fonctionnelles. Dans ce chapitre, on s'intéresse à généraliser ce type de modèles dans un cadre d'échantillonnage non aléatoire, en particulier lorsque les données sont collectées par un processus de stratification endogène. L'intérêt de cette dernière est qu'elle permet de s'adapter à la structure de la population lorsque certains choix (valeurs de la variable réponse) sont rarement choisis, contrairement à l'échantillonnage aléatoire où tous les items de la population ont la même probabilité d'être choisi. La stratification endogène est connue sous le nom *échantillonnage basé sur le choix ou Choice based sampling* (CBS) en économétrie quand on s'intéresse par exemple aux choix de certains consommateurs/utilisateurs ou *échantillonnage Cas-Témoin* en épidémiologie quand on étudie les maladies rares.

Supposons que dans une population donnée, nous observons une variable réponse binaire Y à valeurs dans $\{0,1\}$ et une fonction aléatoire $\{X(t), t \in \mathcal{T}\}$ à valeurs dans $\mathcal{X} \subset L^2(\mathcal{T})$, correspondant à un processus stochastique carré intégrable dans un intervalle $\mathcal{T} \subset \mathbb{R}$. Nous nous intéressons à décrire la relation entre la variable réponse Y et la

fonction explicative $X(\cdot)$. On suppose que cette relation est donnée par le problème de régression à choix binaire suivant

$$E\left(Y|X\right) = P\left(Y=1|X,\alpha^*,\theta^*(\cdot)\right) = \Phi\left(\alpha^* + \int_{\mathcal{T}} X(t)\theta^*(t)dt\right), \tag{3.1}$$

où $\Phi(\cdot)$ est la fonction de lien supposée connue, α^* le terme constant, et $\theta^*(\cdot)$ le paramètre fonctionnel sont inconnus. Notre objectif est d'estimer ces paramètres dans un contexte d'échantillonnage endogène.
Soit $Q^* = P\left(Y=1\right)$, la proportion des individus associés à $Y=1$ dans la population étudiée. On suppose que la population est divisée selon les valeurs de la variable réponse Y en deux strates $\mathcal{J}(0) = \{(0,X),\ X\in\mathcal{X}\}$ et $\mathcal{J}(1) = \{(1,X),\ X\in\mathcal{X}\}$. Soit $0 < H^* < 1$, la probabilité avec laquelle on tire dans la strate $\mathcal{J}(1)$. Le CBS consiste à échantillonner de la manière suivante :*on choisit d'abord une strate* $i\in\{0,1\}$ *avec une probabilité* $H(i)$ *(*$H(1)=H^*$*), ensuite on tire une observation* $(Y=i,X)$ *d'une manière aléatoire dans la strate* $\mathcal{J}(i)$ *choisie.*
Sous cet échantillonnage, la densité conditionnelle de Y sachant $X=x$ est définie par

$$g(i|x) = \frac{P\left(Y=i|x,\alpha^*,\theta^*(\cdot)\right)H(i)/Q(i)}{\sum_{j=0}^{1} P\left(Y=j|x,\alpha^*,\theta^*(\cdot)\right)H(j)/Q(j)}, \quad x\in\mathcal{X},\ i\in\{0,1\}, \tag{3.2}$$

où $Q(i) = P\left(Y=i\right)$. Ainsi, l'espérance induite par la distribution de cet échantillonnage, est définie par (voir, e.g., Cosslett, 2013)

$$E_s(\cdot) = H(0)E(\cdot|Y=0) + H(1)E(\cdot|Y=1).$$

Par conséquence, l'espérance conditionnelle de Y sachant X sous le CBS est donnée par

$$E_s\left(Y|X\right) = g(1|X) = \mu\left(\alpha^* + \int_{\mathcal{T}} \theta^*(t)X(t)dt\right),$$

où

$$\mu(\cdot) = \frac{\Phi(\cdot)H^*/Q^*}{\Lambda(\cdot)} \qquad \text{with} \qquad \Lambda(\cdot) = \frac{H^*}{Q^*}\Phi(\cdot) + \frac{1-H^*}{1-Q^*}\left(1-\Phi(\cdot)\right).$$

Supposons que nous disposons de N observations indépendantes
$\left(Y_n = i_n, \{X_n(t),\ t\in\mathcal{T}\}\right),\ n=1,\ldots,N$, tirées par le processus d'échantillonnage précédant. En se basant sur (3.2), le logarithme de la fonction de la vraisemblance conditionnelle est défini par

$$L(\alpha,\theta(\cdot)) = \sum_{n=1}^{N} \log\left(\frac{P\left(Y_n|X_n,\alpha,\theta(\cdot)\right)H(i_n)/Q(i_n)}{\sum_{j=0}^{1} P\left(Y_n=j|X_n,\alpha,\theta(\cdot)\right)H(j)/Q(j)}\right). \tag{3.3}$$

Dans le cas où la variable explicative X est à valeurs réelles, Manski & McFadden (1981) ont défini des estimateurs du maximum de vraisemblance pour α^* et le vecteur de paramètres θ^*, en maximisant l'équivalent de (3.3). En revanche, dans un cadre fonctionnel, on devrait adresser le problème de la dimension infinie de l'espace de la fonction explicative $X(\cdot)$. Pour pallier à ce problème, nous suivons la stratégie de troncature proposée par Müller & Stadtmüller (2005). Cette approche consiste à projeter la fonction explicative et le paramètre fonctionnel dans un espace de fonctions engendré par une base de fonctions dont la dimension croit asymptotiquement avec la taille de l'échantillon N. En effet, soit $\{\varphi_j,\ j=1,2,\ldots\}$ une base orthonormale de $L^2(\mathcal{T})$. On peut récrire $X(t)$ et $\theta^*(t)$ comme suit :

$$X(t) = \sum_{j\geq 1} \varepsilon_j\varphi_j(t), \qquad \text{et} \qquad \theta^*(t) = \sum_{j\geq 1} \theta^*_j\varphi_j(t),$$

où les variables aléatoire réelles ε_j et les coefficients θ_j^* sont définis par

$$\varepsilon_j = \int_{\mathcal{T}} X(t)\varphi_j(t)dt \qquad \text{et} \qquad \theta_j^* = \int_{\mathcal{T}} \theta^*(t)\varphi_j(t)dt.$$

Nous avons alors :

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j\geq 1} \theta_j^*\varepsilon_j.$$

Soit p_N une suite positive et d'entiers naturels, qui croît asymptotiquement avec $N \to \infty$. On considère la décomposition

$$\eta^* = \alpha^* + \sum_{j=1}^{\infty} \theta_j^*\varepsilon_j, \qquad \tilde{\eta}^* = \alpha^* + \sum_{j=1}^{p_N} \theta_j^*\varepsilon_j, \qquad \eta^* - \tilde{\eta}^* = \sum_{j=p_N+1}^{\infty} \theta_j^*\varepsilon_j.$$

L'idée de l'approche de troncature utilisée consiste à remplacer $E(Y|X)$ par $\Phi(\tilde{\eta}^*)$. Ceci implique que nous allons nous intéresser à estimer p_N+1 paramètres plutôt qu'un nombre infini de paramètres. Les paramètres à estimer dans le modèle tronqué sont donc α^*, et les p_N-premiers coefficients de la fonction $\theta_1^*, \ldots, \theta_{p_N}^*$. Par simplicité, nous notons $\theta^\dagger = (\alpha^*, \theta_1^*, \ldots, \theta_{p_N}^*)^T$. Donc, le logarithme de la vraisemblance conditionnelle tronquée est obtenu en remplaçant $\Phi(\eta_n)$ par $\Phi(\tilde{\eta}_n)$ dans (3.3), soit

$$\tilde{L}_{p_N}(\theta) = \sum_{n=1}^{N} Y_n \log \frac{H^*\Phi(\tilde{\eta}_n)}{Q^*\Lambda(\tilde{\eta}_n)} + (1-Y_n)\log \frac{(1-H^*)(1-\Phi(\tilde{\eta}_n))}{(1-Q^*)\Lambda(\tilde{\eta}_n)},\ \theta \in \Theta \subset \mathbb{R}^{p_N+1}, \quad (3.4)$$

où $\tilde{\eta}_n = \sum_{j=0}^{p_N} \theta_j \varepsilon_j^{(n)}$, avec $\varepsilon_j^{(n)} = \int X_n(t)\varphi_j(t)dt$ et $\varepsilon_0^{(n)} = 1$.
Le vecteur de paramètre $\theta^\dagger$ est ainsi estimé par

$$\widehat{\theta^\dagger} = (\hat{\alpha}, \hat{\theta}_1, \ldots, \hat{\theta}_{p_N})^T = \operatorname{argmax}\left\{\tilde{L}_{p_N}(\theta),\ \theta \in \Theta\right\}.$$

Par conséquent, l'estimateur du terme constant α^* est $\hat{\alpha}$, et celui du paramètre fonctionnel est

$$\hat{\theta}(t) = \sum_{j=1}^{p_N} \hat{\theta}_j \varphi_j(t).$$

Dans la suite, nous étudions le comportement asymptotique de l'estimateur proposé, notamment la normalité asymptotique. Sous des conditions similaires à celles utilisées par Müller & Stadtmüller (2005) mais adaptées au contexte d'échantillonnage utilisé. Plus précisément, nous montrons que

$$\frac{N(\widehat{\theta^\dagger} - \theta^\dagger)^T \Delta_{p_N} (\widehat{\theta^\dagger} - \theta^\dagger) - (p_N+1)}{\sqrt{2(p_N+1)}} \to \mathcal{N}(0,1),$$

où Δ_{p_N} est une $(p_N+1)\times(p_N+1)$ matrice définie par

$$\Delta_{p_N} = E_s\left(\frac{\mu'^2(\tilde{\eta}^*)}{\sigma^2(\mu(\tilde{\eta}^*))}\varepsilon\varepsilon^T\right),$$

ε et $\tilde{\eta}^*$ sont des copies génériques $\varepsilon^{(n)} = (\varepsilon_0^{(n)}, \varepsilon_1^{(n)}, \ldots, \varepsilon_{p_N}^{(n)})^T$ et $\tilde{\eta}_n^* = \theta^{\dagger T}\varepsilon^{(n)}$, respectivement, et $\sigma^2(t) = t(1-t)$.
Si le modèle (3.1) ne contient pas de terme constant ($\alpha^* = 0$), nous concluons que

$$\frac{N d_G^2\left(\hat{\theta}(\cdot), \theta^*(\cdot)\right) - p_N}{\sqrt{2p_N}} \to \mathcal{N}(0,1),$$

où $d_G(\cdot,\cdot)$ est une métrique dans $L^2(\mathcal{T})$, définie par

$$d_G^2(f,g) = \int\int (f(t)-g(t))\, G(t,v)\, (f(v)-g(v))\, dtdv,\ f,g \in L^2(\mathcal{T}),$$

où

$$G(t,v) = \frac{H^*(1-H^*)}{Q^*(1-Q^*)} E\left(\frac{\Phi'^2(\eta^*)}{\sigma^2(\Phi(\eta^*))\Lambda(\eta^*)} X(t)X(v)\right),\ t,v \in \mathcal{T},$$

et $\eta^* = \int_{\mathcal{T}} X(t)\theta^*(t)dt$.

Des résultats numériques montrant la performance de l'estimateur proposé par rapport à l'estimateur classique qui ignore la nature de l'échantillonnage sont présentés. Il s'agit de résultats issus de données simulées et réelles sur la résistance de farines utilisées pour la fabrication de biscuits.

The results of this chapter are in collaboration with Mohamed Kadi Attouch (University Djilalli Liabes, Algeria), Sophie Dabo-Niang (University of Lille) and are published in *Journal of Econometrics and Statistics*.

3.1 Introduction

Choice models are characterized by the feature that the dependent variable is discrete instead of continuous. Examples include having a given disease or not, participation decisions, and transport choices made by individuals. In the context of choice models, the main idea in case-control or choice-based sampling design is to stratify the population with respect to the values of the categorical response. It often occurs that one or more outcomes occur infrequently in the population but are important for determining some key parameters of the model. By stratifying the population with respect to the responses, one can gather information on those infrequent outcomes at a much lower cost than would be incurred by simply increasing the size of a random sample. Equivalently, for any given sampling budget, one can increase the efficiency of predictions and parameter estimates using a suitably designed response-based sample. Such sampling designs have been independently investigated by econometricians who study choice behaviour and biostatisticians who are interested in rare diseases.
In biostatistics, case-control designs are useful for identifying the impact of several factors on the occurrence of a particular disease. The response is often binary (having the disease or not), but there may be more than two categorical responses. In a case-control study, separate samples of cases (diseased individuals) and controls (individuals without the disease) are selected, unlike in a prospective study design, in which a sample of individuals is chosen and followed through time until their responses are recorded. In the case of rare diseases, even large studies may produce only a few diseased individuals and little information about the hazard. In that case, the researcher might wish to oversample the rare disease of interest to increase the accuracy of his analysis. Therefore, compared with case-control studies, prospective studies are disadvantageous in terms of time and cost. For a general overview of medical case-control studies, see Keogh & Cox (2014), for instance. Case-control studies are also used in political science (King & Zeng, 2001) and sociology (Xie & Manski, 1989).

Originally, choice-based sampling was used by econometricians. They were interested in exploring the relationships between the choices made by an individual and several explanatory variables. The choice of transportation mode is the most popular example; see, for instance, Manski & McFadden (1981). In this case, it may be advantageous

(simpler and less expensive) to apply choice-based sampling by selecting, for instance, separate samples of individuals from bus terminals, train stations and car parks rather than to take a single sample from the entire population. The reasons for using such stratified samples have been discussed extensively in several econometric papers such as Manski & McFadden (1981) and Cosslett (1981). In this work, we consider a binary choice model with a functional covariate in the context of *case-control or choice-based sampling*.

Functional data analysis (FDA) was widely popularized by Ramsay & Silverman (2005). Since its introduction, considerable work has been done on the representation, exploration and modelling of functional data. Nonparametric methods have also been developed for functional data, and an overview is available in Ferraty & Vieu (2006). Moreover, a number of reference textbooks addressing functional data analysis, such as Bosq (2000), Ramsay & Silverman (2005), and Horváth & Kokoszka (2012), already exist. FDA is thus an active research topic with potential applications in a large number of fields.

The objective of the present paper is to propose a binary functional linear model adapted for case-control sampled data. All available information concerning the sampling design is used to obtain a finer estimation and understanding of the phenomenon of interest. Several types of functional linear models have been developed over the years, therein serving different purposes. Among all functional linear models that have been introduced, the most studied is perhaps the functional linear model for scalar responses, originally introduced by Hastie & Mallows (1993). Functional linear models have also been generalized by Müller & Stadtmüller (2005), Cardot & Sarda (2005), and Escabias et al. (2007), and more recently, functional generalized additive models (see McLean et al., 2014) have also been developed. Furthermore, related models, such as functional linear discriminant analysis, are considered in James & Hastie (2001) and are applied in many fields, e.g., image processing Cardot et al. (2003), medicine Ratcliffe et al. (2002), genetics Müller et al. (2008), ecology Bel et al. (2011), and marketing Sood et al. (2009). All of these applications show that there is increasing interest in the application of functional linear models and their generalization for practical purposes.

To the best of our knowledge, despite many potential applications, no work has been done on binary choice functional linear models for case-control or choice-based sampling studies that consider the method of sampling the data. However, some work does exist (see Cardot et al., 2010) on functional principal component analysis adapted to certain types of sampling data. Cardot & Josserand (2011) also propose consistent estimators of mean and variance functions based on the Horvitz-Thompson estimator. Note that one work (Fan et al., 2014) exists that addresses a functional logit model applied to case-control data. These authors propose a functional logit model to test the associations between a dichotomous trait and multiple genetic variants in a region using several covariates. However, they do not consider the case-control nature of their data. Consequently, their model is similar to a classical generalized functional linear model in the case of random sampling. Several authors have proposed consistent methods for estimating the parameters of interest in a *choice-based sampling* model when the explanatory variables take real values; see, for instance, Manski & Lerman (1977), Manski & McFadden (1981), Cosslett (1981), Imbens (1992), and Cosslett (2013).

Our goal is to generalize, in a functional framework, the conditional maximum likelihood method suggested by Manski & McFadden (1981) to estimate a binary functional linear model in the context of choice-based sampling. We adapt the approach of Müller & Stadtmüller (2005) to reduce the infinite dimension of the space of the explanatory random function using a Karhunen–Loève expansion. Notably, as for real-valued covariates, the improvements that can be expected to be achieved using the functional design framework are mostly related to the performance of the constant parameter estimation rather

than the functional parameter estimation performance. We present a way to improve the accuracy of a traditional binary functional regression model applied to such sample data.

The remainder of the chapter is organized as follows. In Section 3.2, we introduce the design and the model under choice-based sampling, and we discuss the usual approach to estimating a binary functional model in such a case. We then present our proposed method of integrating the sampling design into the estimation process. In Section 3.3, we present asymptotic results, whereas Section 3.4 reports a simulation case study and an application to kneading data to illustrate the performance of the proposed estimators. Finally, the last section presents the proofs of our main results.

3.2 Conditional maximum likelihood estimator with a functional covariate

We assume that in a given population, we observe a binary random variable Y that takes values in $\{0,1\}$ and a random function $\{X(t), t \in \mathcal{T}\}$ that corresponds to a square-integrable stochastic process on the interval $\mathcal{T} \subset \mathbb{R}$. Suppose that the process $\{X(t), t \in \mathcal{T}\}$ takes values in some space $\mathcal{X} \subset L^2(\mathcal{T})$, where $L^2(\mathcal{T})$ is the space of square-integrable functions in $\mathcal{T}$. We are interested in describing the relation between the response variable Y and the explanatory random function $X(\cdot)$. We assume that this relation is given by a binary choice regression problem and that the expectation of Y given $X(\cdot)$ is defined as

$$E\left(Y|X\right) = P\left(Y=1|X,\alpha^*,\theta^*(\cdot)\right) = \Phi\left(\alpha^* + \int_{\mathcal{T}} X(t)\theta^*(t)dt\right), \tag{3.5}$$

where the link function $\Phi(\cdot)$ is some strictly increasing cumulative distribution function. The parameters of interest are the constant intercept α^* in a compact subset of $\mathbb{R}$ and the parameter function $\theta^*(\cdot)$, which is assumed to belong to the space of functions $L^2(\mathcal{T})$.
Let $Q^* = P\left(Y=1\right)$ be the share of individuals such that $Y=1$ in the considered population. We assume that the population is divided, according to the values of the response variable Y, into two strata $\mathcal{J}(0) = \{(0,X),\ X \in \mathcal{X}\}$ and $\mathcal{J}(1) = \{(1,X),\ X \in \mathcal{X}\}$, and we let $0 < H^* < 1$ be the probability with which we will draw from stratum $\mathcal{J}(1)$. We assume that we sample this population as follows: *We select an observation by first drawing a stratum $i \in \{0,1\}$ with probability $H(i)$ ($H(1) = H^*$) and then drawing an observation $(Y=i,X)$ at random from $\mathcal{J}(i)$.*
This type of sampling is known in the econometric or biostatistics literature as pure choice-based sampling or case-control sampling. This sampling process allows the structure of the population to be considered when one of the values of the response variable Y has a small probability of being observed compared with the random sampling case, in which all values have the same probability of being chosen. Under this sampling process, the conditional density of Y given $X=x$ is

$$g(i|x) = \frac{P\left(Y=i|x,\alpha^*,\theta^*(\cdot)\right)H(i)/Q(i)}{\sum_{j=0}^{1} P\left(Y=j|x,\alpha^*,\theta^*(\cdot)\right)H(j)/Q(j)}, \quad x \in \mathcal{X},\ i \in \{0,1\}, \tag{3.6}$$

where $Q(i) = P\left(Y=i\right)$. The expectation value with respect to the distribution under Choice-Based Sampling (CBS) is defined by (see, e.g., Cosslett, 2013)

$$E_s(\cdot) = H(0)E(\cdot|Y=0) + H(1)E(\cdot|Y=1).$$

Note that $E_s(\cdot)$ is different from the expectation value $E(\cdot)$ under the population distribution. Consequently, the expectation of Y given X under CBS is given by

$$E_s\left(Y|X\right) = g(1|X) = \mu\left(\alpha^* + \int_{\mathcal{T}} \theta^*(t)X(t)dt\right),$$

where

$$\mu(\cdot) = \frac{\Phi(\cdot)H^*/Q^*}{\Lambda(\cdot)} \qquad \text{with} \qquad \Lambda(\cdot) = \frac{H^*}{Q^*}\Phi(\cdot) + \frac{1-H^*}{1-Q^*}\left(1-\Phi(\cdot)\right).$$

Our objective is to perform estimation using observations following the same law as (Y, X), the intercept parameter α^* and the parameter function $\theta^*(\cdot)$ when the sampling process is the CBS process defined above and when we assume that we have prior information providing knowledge on Q^* and H^*.
Let us assume that $E\left(X(t)\right) = 0,\ \forall t \in \mathcal{T}$, which will be needed to ensure identification of the intercept.
Let Γ denote the covariance operator of the $\mathcal{X}$-valued random function:

$$\Gamma x(t) = \int_{\mathcal{T}} E(X(t)X(v))x(v)dv,\ x \in \mathcal{X},\ t \in \mathcal{T}.$$

The operator Γ is a linear integral operator whose integral kernel is

$$K(t,v) = E\left(X(t)X(v)\right), \qquad \text{for all } t, v \in \mathcal{T}. \tag{3.7}$$

It is a compact self-adjoint Hilbert-Schmidt operator because

$$\int |K(t,v)|^2\, dtdv \leq \left(E\left(\int X^2(t)dt\right)\right)^2 < \infty;$$

thus, it can be diagonalized (see, e.g., Conway, 2013, p.47).
In addition to the previous assumption regarding the expectation value of the random function, the following assumptions are necessary to ensure the identification of our model.

(H1) The eigenvalues of Γ are nonzero.

(H2) The link function $\Phi(\cdot)$ is monotonic and invertible and has two continuous bounded derivatives with $\|\Phi'\| = \sup_t |\Phi^{'}(t)| < C$ and $\|\Phi^{''}\| < C$ for some constant $C > 0$, and there exists a $\delta > 0$ such that for all $x \in \mathcal{X}$, $\theta(\cdot) \in L^2(\mathcal{T})$ and $\alpha \in \mathbb{R}$,

$$\left(1 - \Phi\left(\alpha + \int_{\mathcal{T}} x(t)\theta(t)dt\right)\right)\Phi\left(\alpha + \int_{\mathcal{T}} x(t)\theta(t)dt\right) > \delta.$$

Assumptions (H1) and (H2) allow us to ensure the identification of our model (see, e.g., Cardot & Sarda, 2005, p.27). Assumption (H2) is similar to assumption $(M1)$ in Müller & Stadtmüller (2005), where it is assumed that the link function is monotonic and invertible and has first and second bounded derivatives and that the conditional variance of the response variable is bounded away from 0.

3.2.1 Infeasible maximum likelihood estimate

We assume that we have a sample of N independent observations
$\left(Y_n = i_n, \{X_n(t),\, t \in \mathcal{T}\}\right),\ n = 1, \ldots, N$, following the same law as (Y, X) and drawn through the CBS process. Then, based on the conditional density (3.6), the conditional log-likelihood function is defined as

$$L(\alpha, \theta(\cdot)) = \sum_{n=1}^{N} \log\left(\frac{P\left(Y_n | X_n, \alpha, \theta(\cdot)\right) H(i_n)/Q(i_n)}{\sum_{j=0}^{1} P\left(Y_n = j | X_n, \alpha, \theta(\cdot)\right) H(j)/Q(j)}\right). \tag{3.8}$$

For the case in which the explanatory variable X takes real values, Manski & McFadden (1981) have maximized (3.8) to find the maximum likelihood estimate of the intercept α^* and the vector of estimates of the parameter θ^* in (3.6).

This method is usually referred to as the conditional maximum likelihood estimator. In our functional context, our aim is to estimate α^* and the parameter function $\theta^*(\cdot)$ by maximizing (3.8) on α and $\theta(\cdot)$. However, this cannot be done before we address the difficulty posed by the infinite dimensionality of the explanatory random function. This can be achieved using one of two very popular approaches used in generalized linear models with explanatory random functions. On the one hand, we have the Penalized Likelihood Method (Cardot & Sarda, 2005), which consists of projecting the parameter function into a finite-dimensional space spanned by a spline basis and then maximizing the pseudo conditional log-likelihood function obtained by replacing the parameter function $\theta(\cdot)$ in (3.8) with its projector, adding a penalty that controls the degree of smoothness of the parameter function. On the other hand, we have the second approach, used by Müller & Stadtmüller (2005). It is based on a truncation strategy that consists of projecting the functional explanatory variable and parameter function into a space of functions generated by a basis of functions with a dimension that increases asymptotically as the sample size tends towards infinity. We shall adapt the strategy of this second approach to resolve the infinite dimensionality problem of the functional space in the context of CBS. This method will be called the *truncated conditional likelihood method*.

3.2.2 Truncated conditional likelihood method

Analogously to Müller & Stadtmüller (2005), the truncation strategy is motivated by the following considerations. Let $\{\varphi_j,\, j = 1, 2, \ldots\}$ be an orthonormal basis of the functional space $L^2(\mathcal{T})$, usually a Fourier or spline basis or a basis constructed from the eigenfunctions of the covariance operator Γ. In our numerical experiments, this last basis will be used. We can rewrite $X(t)$ and $\theta^*(t)$ as follows:

$$X(t) = \sum_{j\geq 1} \varepsilon_j \varphi_j(t),\ \theta^*(t) = \sum_{j\geq 1} \theta_j^* \varphi_j(t),$$

where the real random variables ε_j and the coefficients θ_j^* are given by

$$\varepsilon_j = \int_{\mathcal{T}} X(t)\varphi_j(t)dt \qquad \text{and} \qquad \theta_j^* = \int_{\mathcal{T}} \theta^*(t)\varphi_j(t)dt.$$

By the orthonormality of the basis $\{\varphi_j,\, j = 1, 2, \ldots\}$, we have

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j\geq 1} \theta_j^* \varepsilon_j.$$

Let p_N be a positive sequence of integers that increases asymptotically as $N \to \infty$, and let us consider the following decomposition:

$$\eta^* = \alpha^* + \sum_{j=1}^{\infty} \theta_j^* \varepsilon_j, \qquad \tilde{\eta}^* = \alpha^* + \sum_{j=1}^{p_N} \theta_j^* \varepsilon_j, \qquad \eta^* - \tilde{\eta}^* = \sum_{j=p_N+1}^{\infty} \theta_j^* \varepsilon_j.$$

The truncation strategy introduced by Müller & Stadtmüller (2005) is based on the following approximation, under the assumption that $\|\Phi'\| < C$ (see assumption (H2)):

$$E\left((\Phi(\eta^*) - \Phi(\tilde{\eta}^*))^2\right) \leq CE\left((\eta^* - \tilde{\eta}^*)^2\right). \tag{3.9}$$

If the right-hand side of (3.9) is asymptotically negligible, then, with the help of (H2), we can truncate $E(Y|X)$ and replace it with $\Phi(\tilde{\eta}^*)$. This is usually the case when we consider the eigenbasis of the variance-covariance operator:

$$E\left((\eta^* - \tilde{\eta}^*)^2\right) = \sum_{j>p_N} {\theta_j^*}^2 E\left(\varepsilon_j^2\right) = \sum_{j>p_N} {\theta_j^*}^2 \lambda_j,$$

where the λ_j are the eigenvalues. We then need to estimate only p_N+1 parameters rather than an infinite number of parameters, under the assumption that the right-hand side of (3.9) vanishes asymptotically as $N \to \infty$.
Now, the parameters of interest in this truncated model are the intercept, α^*, and the first p_N coefficients of the parameter function, $\theta_1^*, \ldots, \theta_{p_N}^*$. For simplicity, let $\theta^\dagger = (\alpha^*, \theta_1^*, \ldots, \theta_{p_N}^*)^T$. The parameter $\theta^\dagger$ takes values in a compact subset $\Theta \subset \mathbb{R}^{p_N+1}$, as α^* takes values in a compact subset of $\mathbb{R}$ and $\theta^*(\cdot) \in L^2(\mathcal{T})$ by assumption. Then, the truncated conditional log-likelihood function is obtained by replacing $\Phi(\eta_n)$ with $\Phi(\tilde{\eta}_n)$ in (3.8). The corresponding feasible conditional likelihood is

$$\tilde{L}_{p_N}(\theta) = \sum_{n=1}^{N} Y_n \log \frac{H^*\Phi(\tilde{\eta}_n)}{Q^*\Lambda(\tilde{\eta}_n)} + (1-Y_n)\log\frac{(1-H^*)\,(1-\Phi(\tilde{\eta}_n))}{(1-Q^*)\,\Lambda(\tilde{\eta}_n)},\ \theta \in \Theta, \quad (3.10)$$

where $\tilde{\eta}_n = \sum_{j=0}^{p_N} \theta_j \varepsilon_j^{(n)}$, with $\varepsilon_j^{(n)} = \int X_n(t)\varphi_j(t)dt$ and $\varepsilon_0^{(n)} = 1$.
Then, $\theta^\dagger$ is estimated as

$$\widehat{\theta^\dagger} = (\hat{\alpha}, \hat{\theta}_1, \ldots, \hat{\theta}_{p_N})^T = \operatorname{argmax}\left\{\tilde{L}_{p_N}(\theta),\ \theta \in \Theta\right\}.$$

Therefore, the estimator of the intercept α^* is $\hat{\alpha}$, and that of the truncated parameter function is given by

$$\hat{\theta}(t) = \sum_{j=1}^{p_N} \hat{\theta}_j \varphi_j(t).$$

We define the $(p_N+1)\times(p_N+1)$ matrix

$$\Delta_{p_N} = E_s\left(\frac{\mu'^2(\tilde{\eta}^*)}{\sigma^2\left(\mu(\tilde{\eta}^*)\right)}\varepsilon\varepsilon^T\right),$$

where ε and $\tilde{\eta}^*$ are generic copies of $\varepsilon^{(n)} = (\varepsilon_0^{(n)}, \varepsilon_1^{(n)}, \ldots, \varepsilon_{p_N}^{(n)})^T$ and $\tilde{\eta}_n^* = \theta^{\dagger T}\varepsilon^{(n)}$, respectively, and $\sigma^2(t) = t(1-t)$. This matrix is seen as an asymptotic Hessian matrix of the pseudo likelihood function (3.10) and will be used to establish an asymptotic normality result for the proposed estimator. In practice, this matrix can be replaced with an adequate empirical version.

Remark 3.1. *Let us investigate the effect of considering the sampling scheme on the asymptotic Hessian matrix Δ_{p_N}. Using the definitions of E_s, $\mu(\cdot)$, and $\Lambda(\cdot)$ together with the truncation strategy, one can show through simple computations that*

$$\Delta_{p_N} \approx \frac{H^*(1-H^*)}{Q^*(1-Q^*)} E\left(\frac{\Phi'^2(\tilde{\eta}^*)}{\sigma^2(\Phi(\tilde{\eta}^*))\Lambda(\tilde{\eta}^*)}\varepsilon\varepsilon^T\right), \quad (3.11)$$

where "$\approx$" indicates that the term $\Phi(\eta^*)$ has been replaced with $\Phi(\tilde{\eta}^*)$. If the sampling process is ignored ($H^* = Q^*$), then the right-hand side of (3.11) will be $E\left(\frac{\Phi'^2(\tilde{\eta}^*)}{\sigma^2(\Phi(\tilde{\eta}^*))}\varepsilon\varepsilon^T\right)$. This is exactly the asymptotic Hessian matrix given in Müller & Stadtmüller (2005), p.8.

In the following section, we present the assumptions and the consistency results regarding $\hat{\alpha}$ and $\hat{\theta}(\cdot)$.

3.3 Assumptions and results

In addition to the previous hypotheses, we need to consider the following assumptions.

(H3) The integer p_N satisfies $p_N \to \infty$ and $N^{-1/4} p_N \to 0$ as $N \to \infty$.

(H4) We have

$$\sum_{r_1,r_2,r_3,r_4=0}^{p_N} E_s\left(\frac{\mu'^4(\tilde{\eta}^*)}{\sigma^4\left(\mu(\tilde{\eta}^*)\right)}\varepsilon_{r_1}\varepsilon_{r_2}\varepsilon_{r_3}\varepsilon_{r_4}\right)\kappa_{r_1r_2}\kappa_{r_3r_4} = o(N/p_N^2),$$

where the $\kappa_{kl},\ k,l = 0,\ldots,p_N+1$, are the elements of $\Xi_{p_N} = \Delta_{p_N}^{-1}$.

(H5) We assume that

$$\sum_{r_1,\ldots,r_8=0}^{p_N} E_s\left(\frac{\mu'^4(\tilde{\eta}^*)}{\sigma^4\left(\mu(\tilde{\eta}^*)\right)}\varepsilon_{r_1}\varepsilon_{r_3}\varepsilon_{r_5}\varepsilon_{r_7}\right) \times E_s\left(\frac{\mu'^4(\tilde{\eta}^*)}{\sigma^4\left(\mu(\tilde{\eta}^*)\right)}\varepsilon_{r_2}\varepsilon_{r_4}\varepsilon_{r_6}\varepsilon_{r_8}\right)\kappa_{r_1r_2}\kappa_{r_3r_4}\kappa_{r_5r_6}\kappa_{r_7r_8} = o(N^2p_N^2).$$

Assumptions (H4) and (H5) are technical assumptions required to establish the following proof of asymptotic normality; they are similar to assumptions (M.3) and (M.4) in Müller & Stadtmüller (2005). Assumption (H4) will then be used in the proof of (3.24), and (H5) is needed to prove (3.23) in the Appendix. Hypothesis (H3) concerns the convergence of p_N; for more details on the utility of these assumptions, see Müller & Stadtmüller (2005). Under these assumptions, we prove the asymptotic normality of $\widehat{\theta^\dagger}$ as follows.

Theorem 3.1. *Under assumptions (H1)-(H5), the estimator $\widehat{\theta^\dagger}$ converges in probability to $\theta^\dagger$ and satisfies*

$$\frac{N(\widehat{\theta^\dagger}-\theta^\dagger)^T\Delta_{p_N}(\widehat{\theta^\dagger}-\theta^\dagger)-(p_N+1)}{\sqrt{2(p_N+1)}} \to \mathcal{N}(0,1). \tag{3.12}$$

The previous result confirms only the consistency of the p_N+1 parameter vector estimator $\widehat{\theta^\dagger}$. In the case in which one is interested in investigating the convergence of the parameter function $\hat{\theta}(\cdot)$ (such as in a model without an intercept), the following procedure can be used (Müller & Stadtmüller, 2005). Let $G(\cdot,\cdot)$ denote the integral kernel, defined as

$$G(t,v) = \frac{H^*(1-H^*)}{Q^*(1-Q^*)}E\left(\frac{\Phi'^2(\eta^*)}{\sigma^2(\Phi(\eta^*))\Lambda(\eta^*)}X(t)X(v)\right),\ t,v\in\mathcal{T},$$

with $\eta^* = \int_{\mathcal{T}} X(t)\theta^*(t)dt$ and let A_G be the Hilbert-Schmidt operator associated with G. Consider φ_j^G, $j=1,2,\ldots$, the eigenbasis of the operator A_G, and the eigenvalues λ_j^G associated with this eigenbasis. The estimated parameter function $\hat{\theta}(\cdot)$ and the parameter function $\theta^*(\cdot)$ can be expressed in this eigenbasis as

$$\theta^*(t) = \sum_{j\geq 1}\theta^*_{\varphi_j^G}\varphi_j^G(t) \quad \text{and} \quad \hat{\theta}(t) = \sum_{j=1}^{p_N}\hat{\theta}_{\varphi_j^G}\varphi_j^G(t),$$

where the $\hat{\theta}_{\varphi_j^G}$ are obtained as above using the eigenbasis of A_G.
Let $d_G(\cdot,\cdot)$ denote the metric defined in the $L^2(\mathcal{T})$ space through the operator A_G, and let it be defined by

$$d_G^2(f,g) = \int\int (f(t)-g(t))\,G(t,v)\,(f(v)-g(v))\,dtdv,\ f,g\in L^2(\mathcal{T}).$$

Then, the distance between $\hat{\theta}(\cdot)$ and $\theta^*(\cdot)$ under this metric is given by

$$\begin{aligned} d_G^2\left(\hat{\theta}(\cdot),\theta^*(\cdot)\right) &= \sum_{j=1}^{p_N}\lambda_j^G\left(\hat{\theta}_{\varphi_j^G}-\theta^*_{\varphi_j^G}\right)^2+\sum_{j>p_N}\lambda_j^G\left(\theta^*_{\varphi_j^G}\right)^2 \\ &= \left(\hat{\theta}_{\varphi^G}-\theta^*_{\varphi^G}\right)^T\Delta_{p_N}^G\left(\hat{\theta}_{\varphi^G}-\theta^*_{\varphi^G}\right)+\sum_{j>p_N}\lambda_j^G\left(\theta^*_{\varphi_j^G}\right)^2, \end{aligned}$$

where

$$\hat{\theta}_{\varphi^G}=\left(\hat{\theta}_{\varphi_1^G},\ldots,\hat{\theta}_{\varphi_{p_N}^G}\right)^T,\qquad \theta^*_{\varphi^G}=\left(\theta^*_{\varphi_1^G},\ldots,\theta^*_{\varphi_{p_N}^G}\right)^T,$$

and the diagonal matrix $\Delta_{p_N}^G$ is equal to Δ_{p_N} if one replaces $\{\varphi_j,\, j=1,2,\ldots\}$ with $\{\varphi_j^G,\, j=1,2,\ldots\}$. Let the following condition hold, which is related to the contribution of the oscillation of the functional covariate to the L^2 norm of the parameter function $\theta^*(\cdot)$:

$$\sum_{j>p_N}E\left((\varepsilon_j^G)^2\right)\left(\int\theta^*(t)\varphi_j^G(t)dt\right)^2=o(\sqrt{p_N}/N), \tag{3.13}$$

where $\varepsilon_j^G=\int_{\mathcal{T}}X(t)\varphi_j^G(t)dt$. The following corollary provides an asymptotic normality result for the parameter function estimate $\hat{\theta}(\cdot)$ obtained using the distance $d_G(.,.)$ in the case without an intercept.

Corollary 3.1. *Under the conditions of Theorem 3.1 and if $\theta^*(\cdot)$ satisfies (3.13), then as $N\to\infty$, we have*

$$\frac{Nd_G^2\left(\hat{\theta}(\cdot),\theta^*(\cdot)\right)-p_N}{\sqrt{2p_N}}\to\mathcal{N}(0,1).$$

The following result is derived from Theorem 3.1 and provides a confidence band for $\theta^*(\cdot)$.

Corollary 3.2. *Let the eigenelements of the matrix $\Delta_{p_N}^{(-1)}$ be denoted by $(v^{(1)},\lambda_1),\ldots,(v^{(p_N)},\lambda_{p_N})$, where $\Delta_{p_N}^{(-1)}$ denotes the $p_N\times p_N$ sub-matrix of Δ_{p_N} obtained by removing the first row/column. Let*

$$v^{(k)}=(v_1^{(k)},\ldots,v_{p_N}^{(k)})^T,\qquad \omega_k(t)=\sum_{j=1}^{p_N}v_j^{(k)}\varphi_j(t),\ k=1,\ldots,p_N;$$

then, for large N and p_N, an approximate $(1-\rho)$ simultaneous confidence band is determined, under the conditions of Theorem 3.1, as follows:

$$\hat{\theta}(t)\pm\sqrt{c(\rho)\sum_{k=1}^{p_N}\frac{\omega_k^2(t)}{\lambda_k}},$$

where $c(\rho)=\left(p_N+z_{1-\rho}\sqrt{2p_N}\right)/N$ and $z_{1-\rho}$ is the $(1-\rho)\%$ quantile of the standard normal distribution, with $0<\rho<1$.

Under assumptions similar to those used in Müller & Stadtmüller (2005) but adapted to our context of CBS, we show above that the proposed conditional maximum likelihood (CML) estimator of the binary functional choice model has the same asymptotic properties as those of the ordinary maximum likelihood (OML) estimator used in the random sampling context. However, the two estimators are distinct. The following theorem proves that in the considered choice-based or case-control sampling process, the OML method (Müller & Stadtmüller, 2005) does not yield consistent estimates of $\theta^\dagger=(\alpha^*,\theta_1^*,\ldots,\theta_{p_N}^*)^T$ under (H1), (H2) and $E\left(\varepsilon\Phi'(\tilde{\eta}^*)\right)\neq 0$.

Theorem 3.2. *Under choice-based sampling ($Q^* \neq H^*$), if assumptions (H1), (H2) and $E\left(\varepsilon\Phi'(\tilde{\eta}^*)\right) \neq 0$ are satisfied, then the OML estimator of the parameter vector $\theta^\dagger$, which is given by*

$$\widehat{\theta^\dagger}_{\mathrm{RS}} = \operatorname{argmax}\left\{L^{\mathrm{RS}}_{p_N}(\theta),\, \theta \in \Theta\right\},$$

where

$$L^{\mathrm{RS}}_{p_N}(\theta) = \sum_{n=1}^{N} Y_n \log(\Phi(\tilde{\eta}_n)) + (1 - Y_n)\log(1 - \Phi(\tilde{\eta}_n)),$$

is inconsistent. $L^{\mathrm{RS}}_{p_N}(\cdot)$ is the truncated ordinary conditional likelihood function of Y given X adapted for random sampling; it does not account for the sampling scheme.

Remark 3.2.

1. *Note that the condition $E\left(\varepsilon\Phi'(\tilde{\eta}^*)\right) \neq 0$ is related to the parameter vector $\theta^\dagger$, the link function $\Phi(\cdot)$, and the joint distribution of the first p_N components of the functional covariate. It is always satisfied in a model with an intercept (equation (3.5) with $\alpha^* \neq 0$) because in this case, $\varepsilon_0 = 1$ and the link function is monotonic and invertible by assumption (H2).*

2. *Although we can offer no general analysis of this inconsistency question of the OML estimator, let us characterize it for the logit model. Let us prove that as for binary logit models for CBS with real-valued covariates, the inconsistency of the OML estimator is related to the intercept estimator, which has a bias depending on H^* and Q^* (see, e.g., Manski & Lerman, 1977, p.1986).*
   For $x \in \mathcal{X}$, $Y \in \{0, 1\}$, $\alpha \in \mathbb{R}$, $\theta(\cdot) \in L^2(\mathcal{T})$ and $\eta = \alpha + \int \theta(t)x(t)dt$, let $\mathcal{L}(\eta; Y)$ denote the conditional log-likelihood function of Y given $X = x$ when the sampling scheme is ignored:

$$\mathcal{L}(\eta; Y) = Y \log(\Phi(\eta)) + (1 - Y)\log(1 - \Phi(\eta)).$$

   Let $\mathcal{F}(\eta)$ denote the expectation value of $\mathcal{L}(\eta; Y)$ under the true parameters α^* and $\theta^*(\cdot)$ ($\eta^* = \alpha^* + \int \theta^*(t)x(t)dt$):

$$\mathcal{F}(\eta) = E_s\left(\mathcal{L}(\eta; Y)|\eta^*\right).$$

   Let $\delta = \log(H^*/Q^*)$ and $\bar{\delta} = \log((1 - H^*)/(1 - Q^*))$; now, using the definition of $E_s(\cdot)$, we obtain

$$\begin{aligned}
\mathcal{F}(\eta) &= \sum_{j=0}^{1}\left\{j\log\left(\frac{e^\eta}{1+e^\eta}\right) + (1-j)\log\left(\frac{1}{1+e^\eta}\right)\right\} \\
&\qquad \times\left\{j\, e^{\delta}\frac{e^{\eta^*}}{1+e^{\eta^*}} + (1-j)e^{\bar{\delta}}\frac{1}{1+e^{\eta^*}}\right\} \\
&= \frac{e^{\bar{\delta}} + e^{\eta^*+\delta}}{1+e^{\eta^*}}\sum_{j=0}^{1}\left\{j\log\left(\frac{e^\eta}{1+e^\eta}\right) + (1-j)\log\left(\frac{1}{1+e^\eta}\right)\right\} \\
&\qquad \times\left\{j\frac{e^{\eta^*+\delta-\bar{\delta}}}{1+e^{\eta^*+\delta-\bar{\delta}}} + \frac{1-j}{1+e^{\eta^*+\delta-\bar{\delta}}}\right\} \\
&= \frac{e^{\bar{\delta}} + e^{\eta^*+\delta}}{1+e^{\eta^*}} E\left(\mathcal{L}(\eta; Y)|\eta^* + (\delta - \bar{\delta})\right).
\end{aligned}$$

   For every $x \in \mathcal{X}$, $\mathcal{F}(\eta)$ has the same maximum with respect to η as does $E\left(\mathcal{L}(\eta; Y)|\eta^* + (\delta - \bar{\delta})\right)$. The latter is maximized at $\eta = \eta^* + (\delta - \bar{\delta})$, which is

equivalent to $\alpha = \alpha^* + (\delta - \bar{\delta})$ *and* $\theta(\cdot) = \theta^*(\cdot)$*, under assumption (H1). Thus, for logit models, as in the case of real-valued covariates, an estimation procedure based on the log-likelihood* $\mathcal{L}(\cdot; Y)$ *yields a consistent estimator of the parameter function and a biased estimate of the intercept, with a bias equal to* $\delta - \bar{\delta} = \log((1-Q^*)H^*/(1-H^*)Q^*)$.

3. *In addition, note that, even if* $\widehat{\theta^\dagger}_{\mathrm{RS}}$ *is consistent (this may be the case if* $E(\varepsilon\Phi'(\tilde{\eta}^*)) = 0$*),*

$$\frac{N(\widehat{\theta^\dagger}_{\mathrm{RS}} - \theta^\dagger)^T \tilde{\Delta}_{p_N}(\widehat{\theta^\dagger}_{\mathrm{RS}} - \theta^\dagger) - (p_N + 1)}{\sqrt{2(p_N + 1)}},$$

where $\tilde{\Delta}_{p_N} = E\left(\frac{\Phi'^2(\tilde{\eta}^*)}{\sigma^2(\Phi(\tilde{\eta}^*))}\varepsilon\varepsilon^T\right)$ *(the Hessian matrix of Müller & Stadtmüller, 2005, Theorem 4.1) does not have an asymptotic standard normal distribution. Let* $U_{\mathrm{RS}}(\theta)$ *denote the gradient of* $L^{\mathrm{RS}}_{p_N}(\theta)$*, defined as*

$$U_{\mathrm{RS}}(\theta) = \frac{\partial}{\partial\theta}L^{\mathrm{RS}}_{p_N}(\theta) = \sum_{n=1}^{N}\frac{\Phi'(\tilde{\eta}_n)}{\sigma^2((\Phi(\tilde{\eta}_n))}(Y_n - \Phi(\tilde{\eta}_n))\varepsilon^{(n)}.$$

Let $J^{\mathrm{RS}}_{\theta^\dagger}$ *denote the Hessian matrix of* $L^{\mathrm{RS}}_{p_N}(\theta)$ *at* $\theta = \theta^\dagger$*, that is,*

$$J^{\mathrm{RS}}_{\theta^\dagger} = \frac{\partial}{\partial\theta^T}U_{\mathrm{RS}}(\theta)\Big|_{\theta^\dagger} = R_{RS} - D^T_{RS}D_{\mathrm{RS}},$$

where

$$D^T_{RS}D_{RS} = \sum_{n=1}^{N}\frac{\Phi'^2(\tilde{\eta}^*_n)}{\sigma^2\left(\Phi(\tilde{\eta}^*_n)\right)}\varepsilon^{(n)}\varepsilon^{(n)T}$$

and

$$R_{RS} = \sum_{n=1}^{N}(Y_n - \Phi(\tilde{\eta}^*_n))\left\{\frac{\Phi''(\tilde{\eta}^*_n)}{\sigma^2\left(\Phi(\tilde{\eta}^*_n)\right)} - \frac{\Phi'^2(\tilde{\eta}^*_n)\sigma^{2\prime}\left(\Phi(\tilde{\eta}^*_n)\right)}{\sigma^4\left(\Phi(\tilde{\eta}^*_n)\right)}\right\}\varepsilon^{(n)}\varepsilon^{(n)T}.$$

The proof of the asymptotic normality of $\widehat{\theta^\dagger}_{\mathrm{RS}}$ *(see, Müller & Stadtmüller, 2005, Theorem 4.1) is based on the fact that* R_{RS} *will eventually be negligible; however, this does not occur under CBS. Indeed, using the matrix norm* $\|M\|_2 = (\sum_{k,l} m^2_{kl})^{1/2}$*, one can easily show that*

$$E_s\left(\left\|\frac{R_{\mathrm{RS}}}{N}\right\|^2_2\right) = O\left(\frac{p^2_N}{N}\right) + \left|\frac{H^*}{Q^*} - \frac{1-H^*}{1-Q^*}\right| O\left(p^2_N\right); \tag{3.14}$$

therefore, the asymptotic normality of $\widehat{\theta^\dagger}_{\mathrm{RS}}$ *remains valid in the CBS context if the second term on the right-hand side of (3.14) is null. However, this cannot hold since* $Q^* \neq H^*$ *in our context.*

The following section investigates the numerical performance of the proposed methodology.

3.4 Numerical experiments

In this section, we study the performance of the proposed model based on some numerical results, which highlight the importance of considering the method of sampling the data. We first describe the estimation procedure for the investigated model. We conduct some

simulations and compare the proposed CML method with the OML method. An application to real data is also considered.
We consider the model defined in (3.5), and using the first twenty functions of the Fourier basis $\{\varphi_j(t) \equiv \sqrt{2}\sin(j\pi t),\ t \in [0,1]\}$, we generate the explanatory pseudo-random function

$$X(t) = \sum_{j=1}^{20} \varepsilon_j \varphi_j(t), \tag{3.15}$$

where $\varepsilon_j \sim \mathcal{N}(0, 1/j)$ for $j \geq 1$. We define the parameter function as $\theta(t) = \sum_{j=1}^{20} \theta_j \varphi_j(t)$, with $\theta_j = 0$ for $j > 4$. The intercept α and the first three coefficients θ_1, θ_2, and θ_3 will be chosen for each of the three following models for different proportions of $Y_n = 1$ in the population: $Q^* = 0.10$, 0.70, and 0.85.

- Logit model: $\Phi(t) = \exp(t)/(1 + \exp(t))$.
- Probit model: $\Phi(\cdot)$ is the standard normal distribution function.
- C-loglog model: $\Phi(t) = \exp(-\exp(-t))$.

For each model, we generate a population $(Y_n, X_n)_{n=1,\ldots m}$ of size $m = 3,000$, where X_n is generated as described in (3.15) and the response variable $Y_n = 0$, 1 is a pseudo-Bernoulli random variable with probability $\Phi\left(\alpha + \int X_n(t)\theta(t)dt\right)$; we also calculate the associated proportion of $Y_n = 1$, that is, Q^*. Then, we draw three stratified samples of sizes $N = 100$, 200, and 400 using same-share sampling for the two strata, $H^* = 0.5$. This means that in the obtained stratified sample, the number of individuals with response $Y_n = 1$ (called cases) is $N/2$, equal to the number of individuals with response $Y_n = 0$ (controls). Note that fixing H^* and varying Q^* allows us to observe the influence of the ratio H^*/Q^* on the proposed CML method.
Recall that the truncation strategy used in this paper requires an appropriate choice of orthonormal basis. This basis can be chosen to be a fixed orthonormal basis, such as the Fourier basis; alternatively, it can be constructed by estimating the eigenfunctions of the covariance kernel (3.7) and applying functional principal component analysis (FPCA) to the explanatory random functions X_n. This FPCA should consider the approach used in selecting the stratified sample.
The stratified FPCA method used here is similar to that of Cardot et al. (2010) in a non-random sampling context; this FPCA method helps to construct a more efficient eigenbasis compared with classical FPCA (Cardot et al., 2010, p.84). More precisely, the FPCA applied in our CBS framework can be regarded as an FPCA applied to a stratified sample, where the latter is built by independently drawing two samples of size $N/2$ each through random sampling (without replacement) from the two strata defined by the response variable. We apply the CML method using the eigenfunctions obtained from this stratified FPCA. These eigenfunctions are those of the integral operator associated with the integral kernel defined by the variance-covariance function of X, which is estimated for each $t, v \in [0,1]$ as follows:

$$\hat{K}(t,v) = \frac{1}{N-1} \sum_{n=1}^{N} X_n(t) X_n(v) \left(\frac{Q^*}{H^*} \mathbb{I}(Y_n = 1) + \frac{1-Q^*}{1-H^*} \mathbb{I}(Y_n = 0) \right). \tag{3.16}$$

Note that when the OML method is applied, the eigenfunctions will be chosen through a classic FPCA, which is equivalent to using (3.16) with $H^* = Q^*$.
Another key step is the choice of the number p of eigenfunctions used in the truncation strategy. We will consider the Akaike Information Criterion (AIC) based on (3.10). Müller

& Stadtmüller (2005) discussed the consistency of the choice of p using AIC in the case of random sampling. We think that this criterion remains consistent in the CBS context; this could be theoretically investigated in the future. Note that we use a pre-selected p based on the cumulative inertia. Indeed, we focus on the selection of p (using AIC) from among those associated with cumulative inertia values lower than some threshold (here, 95%).

As a measure of accuracy of the parameter function, (see, e.g., Escabias et al., 2007) the usual Integrated Mean Square Error

$$\text{IMSE} = \int_0^1 \left(\theta(t) - \hat{\theta}(t)\right)^2 dt, \tag{3.17}$$

is considered to compare the two estimation strategies: CML and OML. Other alternative approaches to choosing p (e.g., correct classification rate and variance of the estimated parameter function) have been tested. They give similar results for the two estimation methods but are less stable than AIC.

The studied models are replicated 200 times, and the results are presented in Tables 3.1, 3.2 and 3.3. In each table, the columns titled PCs, α and IMSE give the averages over these 200 replications (with the standard deviation in brackets) of the number of eigenfunctions p, the intercept estimate $\hat{\alpha}$, and the associated IMSE defined in (3.17), respectively. The p-val column represents the p-value associated with a Wilcoxon-Mann-Whitney test with the following alternative hypothesis: *The mean IMSE associated with the estimate obtained using the OML method is greater than that associated with the CML method.* Note that for the logit model (Table 3.1 and Figure 3.1), the CML and OML methods yield very similar results in terms of the IMSE for both small and large proportions Q^* of events $(Y_n = 1)$ in the sample. The OML method yields a biased intercept estimate compared with the CML method, whereas the slope estimates are similar. These findings illustrate the comments given in Section 3.3 regarding the bias of the intercept and the consistency of the slope in the logit case when the OML approach is used. As stated before, this is a well-known phenomenon for the case in which the explanatory variables take real values, and it is still valid in our functional context (see Table 3.1 and Figure 3.1). For the probit model (Table 3.2 and Figure 3.2), the OML method also yields a biased intercept estimate compared with the CML method. However, the performance of the CML estimator of $\theta^*(\cdot)$ is superior to that of the OML estimator, particularly for a low or high number of cases in the sample ($Q^* = 0.10$ or 0.85) and a large sample size. High performance of the CML estimator of $\theta^*(\cdot)$ is observed when the C-loglog model is used (Table 3.3 and Figure 3.3), particularly when $Q^* = 0.10$ or 0.85. This can be explained by the fact that the C-loglog distribution is better adapted to extreme values than the logit and probit distributions, which are symmetric around 0.

For the different models, the strategy used to choose p yields (on average) values close to the true parameter $p = 3$ (see the columns titled PCs in Tables 3.1-3.3).

3.4.1 Empirical power simulations

This section is dedicated to testing for no regression effect using the asymptotic results of Section 3.3. We consider the null hypothesis $\text{H}_0 : \alpha^* = 0$ and $\theta_i^* = 0,\ i = 1, 2, \ldots$, for the case of a logit model with $Q^* \approx 0.70$. The rejection region, derived from Theorem 3.1, is $|Z| > z_{0.95}$, where Z is the test statistic defined by the left-hand side of (3.12) (under H_0) and $z_{0.95}$ is the 95% quantile of a standard normal distribution. The empirical power is calculated as the proportion of cases in which H_0 is rejected over 500 replications of two stratified samples of sizes $N = 100$ and 400. The power is a function of $\delta \in [0, 2]$, where this parameter serves in each replication to generate the logit model with parameter

Table 3.1: Logit Model

Parameters	N	OML			CML			p-val
		α	IMSE	PCs	α	IMSE	PCs	
$Q^* = 0.10$	100	-0.17	0.85	2.30	-1.97	0.92	2.27	0.88
$\alpha = -2$		(0.12)	(0.41)	(0.46)	(0.10)	(0.45)	(0.45)	
$\theta_1 = -0.5$	200	-0.17	0.50	2.56	-1.99	0.50	2.56	0.62
$\theta_2 = -0.7$		(0.07)	(0.38)	(0.50)	(0.08)	(0.38)	(0.50)	
$\theta_3 = -0.9$	400	-0.16	0.13	2.97	-2.00	0.14	2.97	0.76
		(0.05)	(0.16)	(0.16)	(0.06)	(0.13)	(0.16)	
$Q^* = 0.70$	100	0.32	0.47	2.03	1.21	0.47	2.02	0.30
$\alpha = 1.2$		(0.15)	(0.29)	(0.18)	(0.14)	(0.35)	(0.15)	
$\theta_1 = 1.3$	200	0.31	0.34	2.10	1.19	0.33	2.12	0.41
$\theta_2 = 0.7$		(0.10)	(0.25)	(0.30)	(0.11)	(0.11)	(0.32)	
$\theta_3 = 0.4$	400	1.00	0.28	2.25	1.21	0.27	2.26	0.49
		(0.10)	(0.17)	(0.44)	(0.10)	(0.15)	(0.44)	
$Q^* = 0.85$	100	0.49	0.94	2.21	2.28	0.92	2.18	0.50
$\alpha = 2.3$		(0.17)	(0.43)	(0.41)	(0.17)	(0.41)	(0.38)	
$\theta_1 = 1$	200	0.48	0.63	2.46	2.27	0.65	2.42	0.59
$\theta_2 = -0.7$		(0.11)	(0.39)	(0.50)	(0.12)	(0.37)	(0.50)	
$\theta_3 = 0.9$	400	0.51	0.29	2.81	2.30	0.29	2.83	0.65
		(0.39)	(0.28)	(0.39)	(0.10)	(0.27)	(0.41)	

Table 3.2: Probit Model

Parameters	N	OML			CML			p-val
		α	IMSE	PCs	α	IMSE	PCs	
$Q^* = 0.10$	100	-0.36	0.50	2.85	-1.49	0.41	2.93	0.06
$\alpha = -1.5$		(0.13)	(0.39)	(0.48)	(0.09)	(0.32)	(0.54)	
$\theta_1 = -0.5$	200	-0.36	0.24	3.08	-1.51	0.17	3.13	10^{-3}
$\theta_2 = -0.7$		(0.08)	(0.20)	(0.28)	(0.07)	(0.12)	(0.34)	
$\theta_3 = -0.9$	400	-0.36	0.15	3.11	-1.51	0.08	3.19	10^{-7}
		(0.06)	(0.11)	(0.32)	(0.05)	(0.06)	(0.39)	
$Q^* = 0.70$	100	0.52	0.50	2.12	1.01	0.46	2.11	0.36
$\alpha = 1$		(0.12)	(0.43)	(0.33)	(0.18)	(0.33)	(0.31)	
$\theta_1 = 1.5$	200	0.52	0.26	2.24	1.00	0.27	2.22	0.59
$\theta_2 = 0.8$		(0.12)	(0.17)	(0.43)	(0.11)	(0.17)	(0.42)	
$\theta_3 = 0.4$	400	0.53	0.17	2.43	1.01	0.16	2.44	0.13
		(0.08)	(0.10)	(0.50)	(0.08)	(0.10)	(0.50)	
$Q^* = 0.85$	100	1.13	0.86	2.48	2.05	0.80	2.50	0.16
$\alpha = 2$		(0.24)	(0.52)	(0.50)	(0.23)	(0.49)	(0.51)	
$\theta_1 = 1.6$	200	1.14	0.44	2.76	2.02	0.37	2.72	0.04
$\theta_2 = -0.7$		(0.20)	(0.34)	(0.43)	(0.15)	(0.30)	(0.45)	
$\theta_3 = 0.9$	400	1.14	0.18	3.01	2.01	0.15	3.04	0.13
		(0.13)	(0.16)	(0.09)	(0.12)	(0.11)	(0.20)	

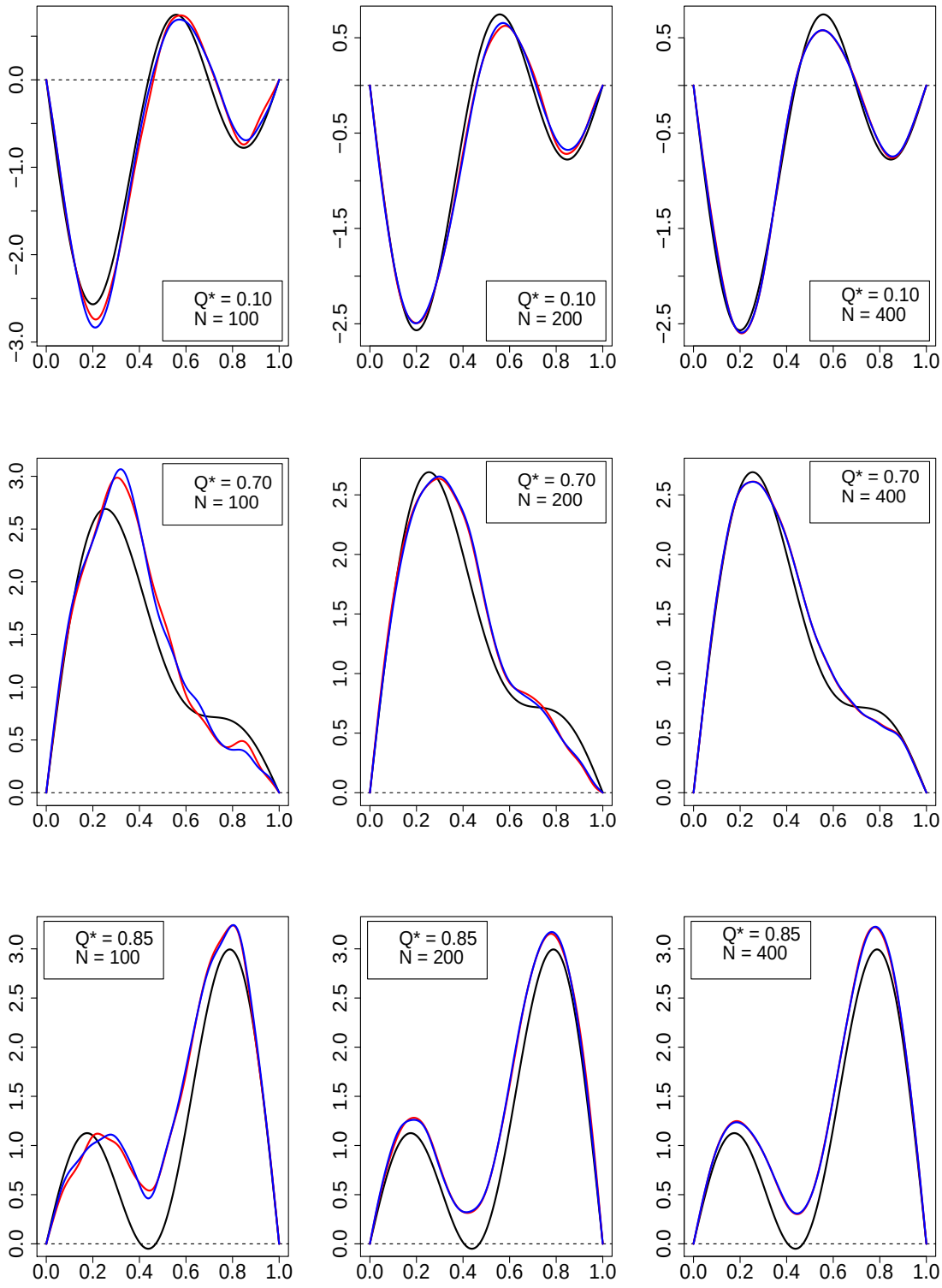

Figure 3.1: Graphs of the simulated parameter function $\theta(\cdot)$ (black curve) and the means (using 200 replications) of its estimates obtained using the OML method (blue curve) and the CML method (red curve) for the logit model.

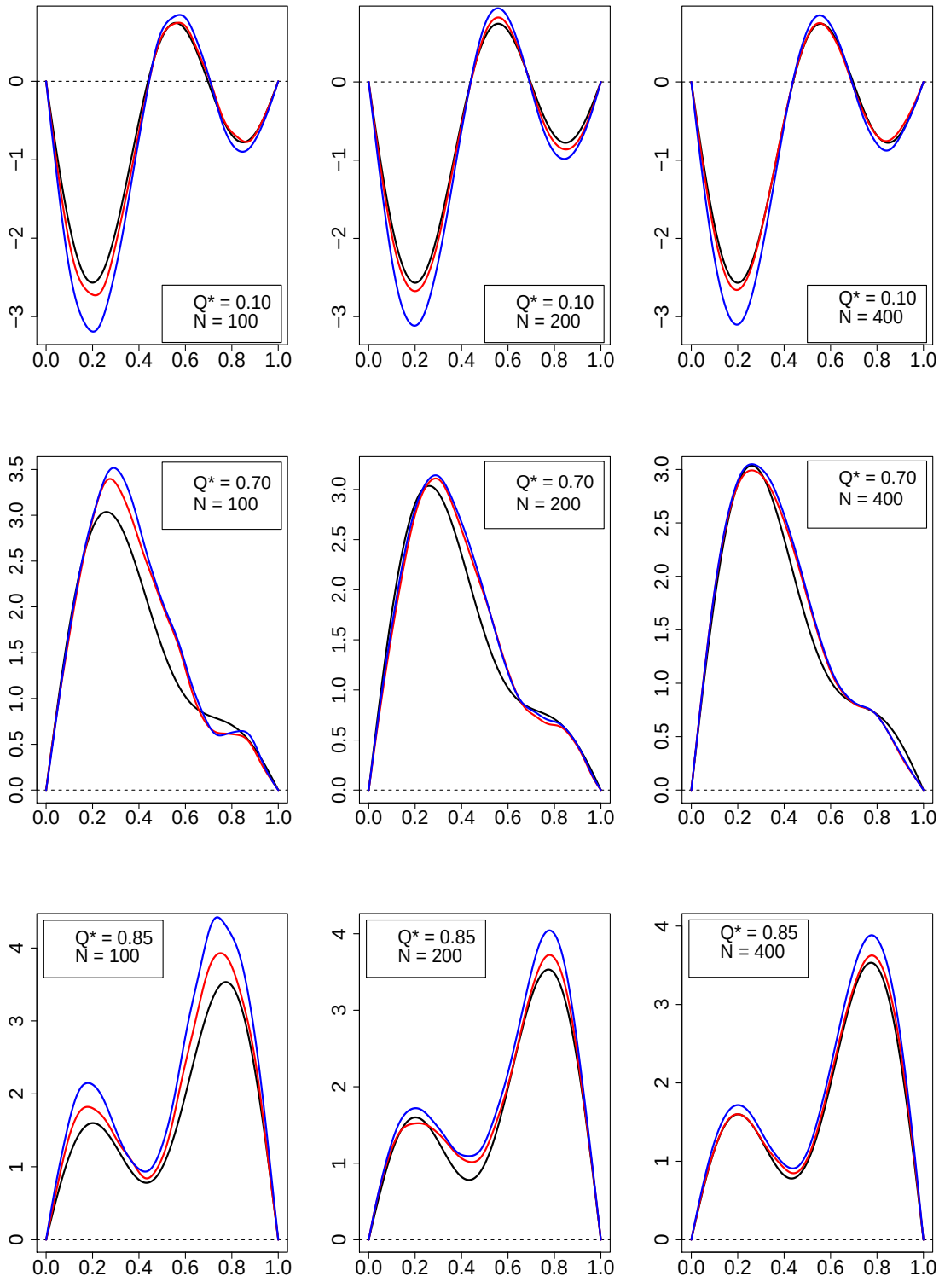

Figure 3.2: Graphs of the simulated parameter function $\theta(\cdot)$ (black curve) and the means (using 200 replications) of its estimates obtained using the OML method (blue curve) and the CML method (red curve) for the probit model.

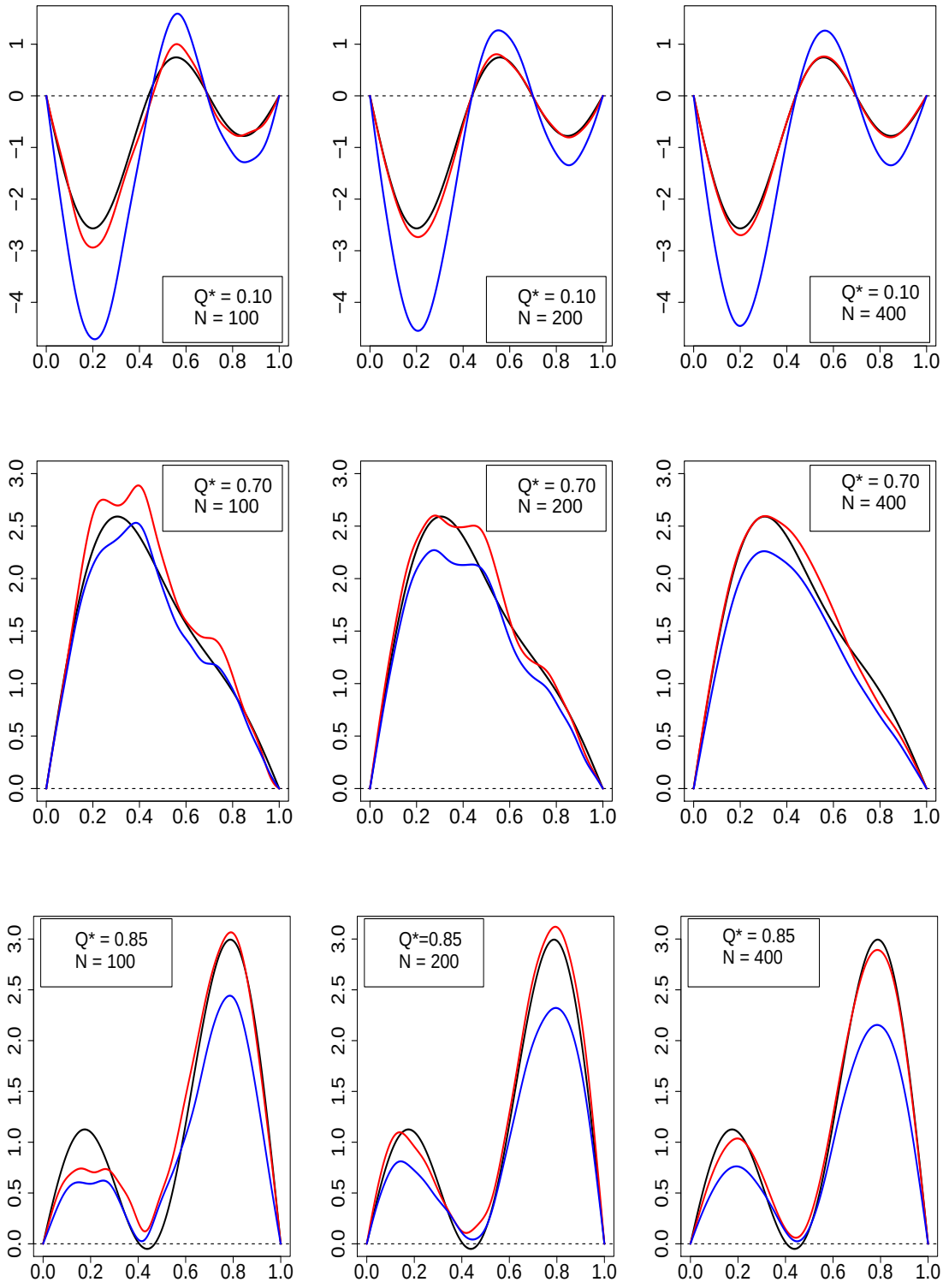

Figure 3.3: Graphs of the simulated parameter function $\theta(\cdot)$ (black curve) and the means (using 200 replications) of its estimates obtained using the OML method (blue curve) and the CML method (red curve) for the C-loglog model.

Table 3.3: C-loglog Model

Parameters	N	OML			CML			p-val
		α	IMSE	PCs	α	IMSE	PCs	
$Q^* = 0.10$	100	-0.07	1.64	2.95	-1.00	0.37	2.97	10^{-27}
$\alpha = -1$		(0.12)	(1.22)	(0.48)	(0.08)	(0.31)	(0.55)	
$\theta_1 = -0.5$	200	-0.07	1.01	3.05	-1.00	0.12	3.14	10^{-42}
$\theta_2 = -0.7$		(0.12)	(0.62)	(0.22)	(0.06)	(0.09)	(0.35)	
$\theta_3 = -0.9$	400	-0.08	0.85	3.07	-1.00	0.07	3.16	10^{-45}
		(0.06)	(0.41)	(0.27)	(0.05)	(0.07)	(0.38)	
$Q^* = 0.70$	100	0.84	0.26	2.02	1.50	0.31	2.05	0.77
$\alpha = 1.5$		(0.14)	(0.16)	(0.15)	(0.16)	(0.27)	(0.21)	
$\theta_1 = 1.6$	200	0.85	0.17	2.25	1.49	0.16	2.16	0.02
$\theta_2 = 0.5$		(0.10)	(0.10)	(0.60)	(0.13)	(0.11)	(0.52)	
$\theta_3 = 0.2$	400	0.85	0.13	2.12	1.51	0.11	2.15	10^{-5}
		(0.08)	(0.06)	(0.36)	(0.09)	(0.07)	(0.40)	
$Q^* = 0.85$	100	0.76	0.74	2.30	2.29	0.85	2.31	0.86
$\alpha = 2.3$		(0.09)	(0.37)	(0.46)	(0.15)	(0.50)	(0.47)	
$\theta_1 = 1$	200	0.76	0.42	2.65	2.29	0.41	2.66	0.35
$\theta_2 = -0.7$		(0.09)	(0.33)	(0.49)	(0.15)	(0.31)	(0.47)	
$\theta_3 = 0.9$	400	0.76	0.27	2.93	2.30	0.16	2.96	10^{-9}
		(0.06)	(0.16)	(0.25)	(0.09)	(0.14)	(0.18)	

$\alpha^* = 1.2\delta$ and the first three coefficients of $\theta^*(\cdot)$ are given by $\theta_1^* = 1.3\delta$, $\theta_2^* = 0.7\delta$, and $\theta_3^* = 0.4\delta$. The two estimation procedures, CML and OML, are used. The results are presented in Figure 3.4. Figure 3.5 shows the histograms of the 500 values of Z obtained in replications with $\theta^\dagger = (1, 1.3, 0.7, 0.4)^T$ for the two estimates of $\widehat{\theta^\dagger}$ obtained via the CML and OML methods (see Remark 3.2). This figure shows that the asymptotic distribution of Z in the OML case is not a standard normal distribution, unlike in the CML case. One can conclude that the test based on the CML method is more powerful than that based on the OML method. In addition, for both methods, the power of the test is influenced by the sample size.

3.4.2 Application to kneading data

Here, we compare our methodology with the OML method on kneading data. Let us consider the quality of cookies from curves representing the resistance (density) of the dough for 90 flours. For a given flour, the resistance of the dough is recorded at 241 equi-spaced time points during the first 480 seconds of a kneading process. The cookie quality associated with each flour is observed; 50 flours are assessed as being of good quality, and 40 are assessed as being of poor quality. This dataset comes from a French agro-industry company. A good interpretation of cookie quality based on flour resistance may allow agro-industry companies to avoid the use of certain flours that could threaten their cookie quality.

The resistance curves can be regarded as sample paths of a square-integrable stochastic process $\{X(t), t \in [0, 480]\}$ (left panel in Figure 3.6) and are smoothed using cubic B-spline functions with 16 knots:

$\{10, 42, 84, 88, 108, 134, 148, 200, 216, 284, 286, 328, 334, 380, 388, 478\}$ (right panel in Figure 3.6); see Preda et al. (2007) for more details on this smoothing. The poor-quality

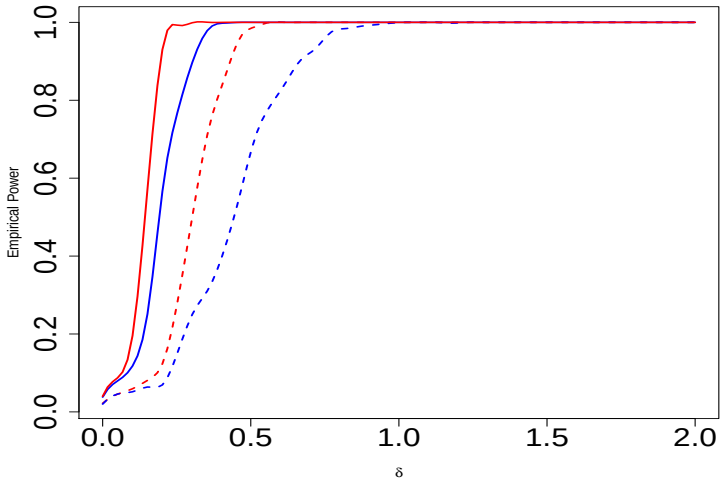

Figure 3.4: Graphs of the empirical power of the significance test for a logit model. The curves represent the results for the CML (red) and OML (blue) methods with sample sizes of 100 (dashed) and 400 (solid). A total of 500 replications, the Fourier basis and a fixed $p = 3$ are used.

cookies are regarded as cases, $Y = 1$, whereas the good-quality cookies are the controls, $Y = 0$. Let $S = \{(y_i, x_i),\ i = 1, ..., 90\}$ denote this kneading dataset composed of the 90 resistance curves and corresponding quality of cookies. The aim of this case study is to illustrate our methodology on this dataset rather than selecting a specific cookie quality prediction method. For that, we adopt the following strategy. The sample S is considered as the flour population of the considered French agro-industry company, with the proportion of cases $Q^* = 40/90 = 44\%$. In this population, we draw a CBS sample S^{cbs} of size $N = 60$ with a proportion of cases equal to $H^* = 25\%$. With this sample, three models (Logit, Probit, and C-loglog) with an intercept are estimated using the OML and CML methods, each with the corresponding FPCA approach detailed above. The AIC is used to choose the number of eigenfunctions p. With the obtained estimates $\hat{\alpha}$ and $\hat{\theta}(\cdot)$, we compute the average squared error (ASE) on the remainder of the population (a set S^r of 30 observations):

$$\text{ASE} = \frac{1}{30} \sum_{(y_i, x_i) \in S^r} \left(y_i - \Phi \left(\hat{\alpha} + \int_0^{480} x_i(t) \hat{\theta}(t) dt \right) \right)^2 .$$

We replicate this methodology 100 times and report the results in Table 3.4 and Figure 3.7. The rows of Table 3.4 titled Intercept, PCs, and ASE give the averages over the 100 replications (with the standard deviation in brackets) of the intercept estimate $\hat{\alpha}$, the number of eigenfunctions p, and the ASE.
For the three models, the results show differences between the OML and CML methods, particularly in terms of the intercept estimate. The ASE associated with the CML estimate is lower than that associated with the OML method. The C-loglog model seems to better fit the data; it gives a smoother estimated parameter function (lower number of eigenfunctions, approximately 3) and lower average squared error by the CML method. These findings illustrate the differences mentioned above regarding the OML and CML methods.

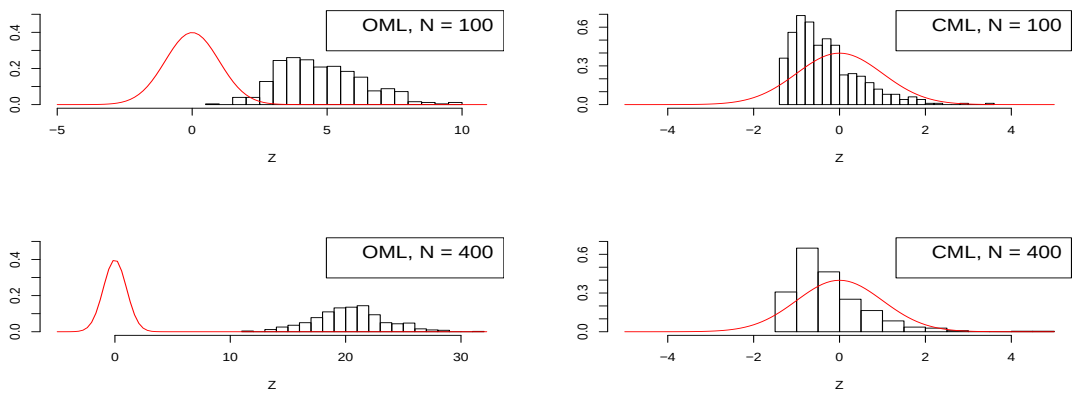

Figure 3.5: Graphs of the histograms associated with a logit model for 500 values of Z (defined on the right-hand side of (3.12)) obtained with 500 replications with $\theta^{\dagger} = (1, 1.3, 0.7, 0.4)^T$ for the two estimates of $\widehat{\theta^{\dagger}}$ obtained using the CML and OML methods. The red curve represents the density of a standard normal distribution.

Table 3.4: Results over 100 replications with a CBS sample of size $N = 60$ drawn in S, $Q^* = 44\%$, $H^* = 25\%$.

	Logit Model		Probit Model		C-loglog Model	
	OML	CML	OML	CML	OML	CML
Intercept	-9.13	-8.08	-1.68	-1.14	-0.35	0.20
	(9.52)	(9.4)	(1.57)	(1.32)	(0.26)	(0.26)
PCs	5.74	5.53	3.96	4.08	3.44	3.17
	(4.03)	(3.75)	(3.27)	(3.00)	(2.27)	(1.86)
ASE	0.18	0.15	0.13	0.10	0.10	0.07
	(0.07)	(0.07)	(0.06)	(0.05)	(0.04)	(0.03)

3.5 Conclusion

In this work, we propose a functional binary choice model for use in analyzing a sample obtained via a choice-based sampling process. A conditional maximum likelihood method (Manski & McFadden, 1981) and a truncation strategy (Müller & Stadtmüller, 2005) are combined to obtain estimators of the intercept and the parameter function. The novel aspect of the proposed method is that it considers both the functional nature of the covariate and the particular sampling design. It is shown that our estimator is asymptotically normal. After studying the theoretical behaviour of the proposed methodology, we consider its practical use. The presented numerical study shows that our method performs better than the ordinary maximum likelihood method under choice-based sampling. According to the numerical results, the proposed estimation method yields significantly more accurate estimates for the intercept and the parameter function, particularly for a probit or C-loglog model. Consequently, one can see the proposed methodology as a good alternative to the classical maximum likelihood method for estimating a binary functional choice model under choice-based or case-control sampling.
In future work, we would like to apply the proposed method to investigate the association between genetic variants (genotypes) and phenotypes (see Fan et al., 2014); these authors find that generalized functional linear models are a good tool for addressing this type of

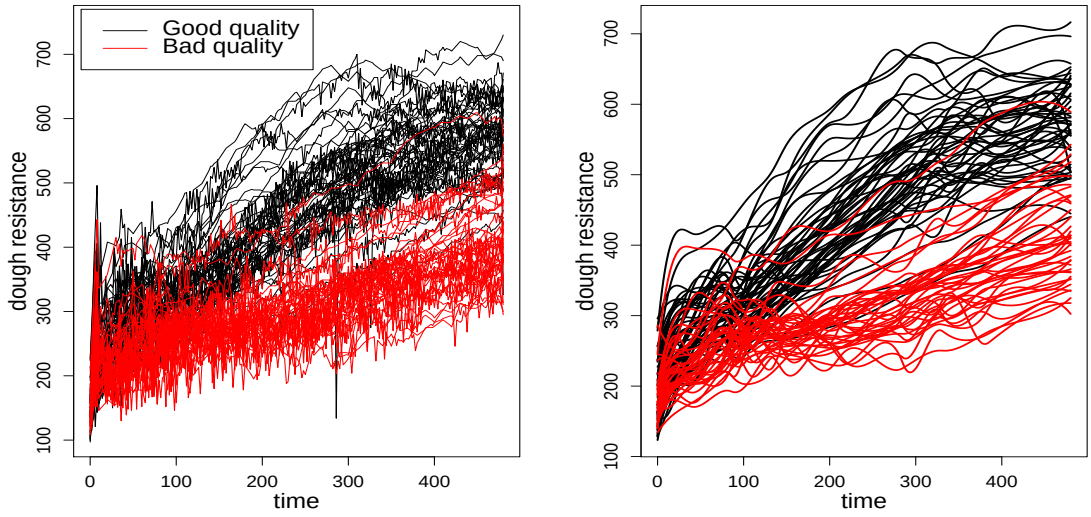

Figure 3.6: Kneading data for 90 flours observed over 480 s. Left: observed data. Right: smoothed data.

problem. We could address a model where one does not have knowledge of the size of the cases in the population. Moreover, the consideration of spatio-functional random covariates or functional space-time series, which are not currently included in our approach, could be investigated in the future.

3.6 Appendix

Let us present some preliminary lemmas.

Lemma 3.1. *(Rao, 1973, p.59)*
Let $g(s,\beta)$ be a real-valued function over a space $\mathcal{S}\times\Theta$ such that g is integrable with respect to a measure ν over $\mathcal{S}$ and $g(s,\beta)\geq 0$ for all $s\in S$ and $\beta\in\Theta$. Let β^* be an element of Θ such that $g(s,\beta^*)>0$ for almost every $s\in S$ and $\int_{\mathcal{S}}(g(s,\beta^*)-g(s,\beta))d\nu(s)\geq 0$ for all $\beta\in\Theta$. Then, the expression

$$\int_{\mathcal{S}} g(s,\beta^*)\log g(s,\beta)d\nu(s),$$

attains its maximum at $\beta=\beta^*$.

Lemma 3.2. *(Amemiya, 1973, Lemma 3, p.1002)*
Let $f_N(s,\beta)$, $N=1,2,\ldots$, be a sequence of measurable functions on a measurable space $\mathcal{S}$, where for each $s\in\mathcal{S}$, $f_N(s,\beta)$ is a continuous function of $\beta\in\Theta$, where Θ is compact. Then, there exists a sequence of measurable functions $\hat{\beta}_N(s)$, $N=1,2,\ldots$, such that $\hat{\beta}_N(s)=\sup_{\beta\in\Theta} f_N(s,\beta)$ for all $s\in\mathcal{S}$ and $N=1,2,\ldots$. Furthermore, if, for almost every $s\in\mathcal{S}$, $f_N(s,\beta)$ uniformly converges to $f(\beta)$ for all $\beta\in\Theta$ and if $f(\beta)$ has a unique maximum at $\beta^*\in\Theta$, then $\hat{\beta}_N(s)$ converges to β^* for almost every $s\in\mathcal{S}$.

Let us adopt the following notation (see Section 3.2.2):

$$\eta=\alpha+\int_{\mathcal{T}} X(t)\theta(t)dt=\alpha+\sum_{i=1}^{\infty}\theta_i\varepsilon_i,\qquad \eta^*=\alpha^*+\sum_{i=1}^{\infty}\theta_i^*\varepsilon_i,$$

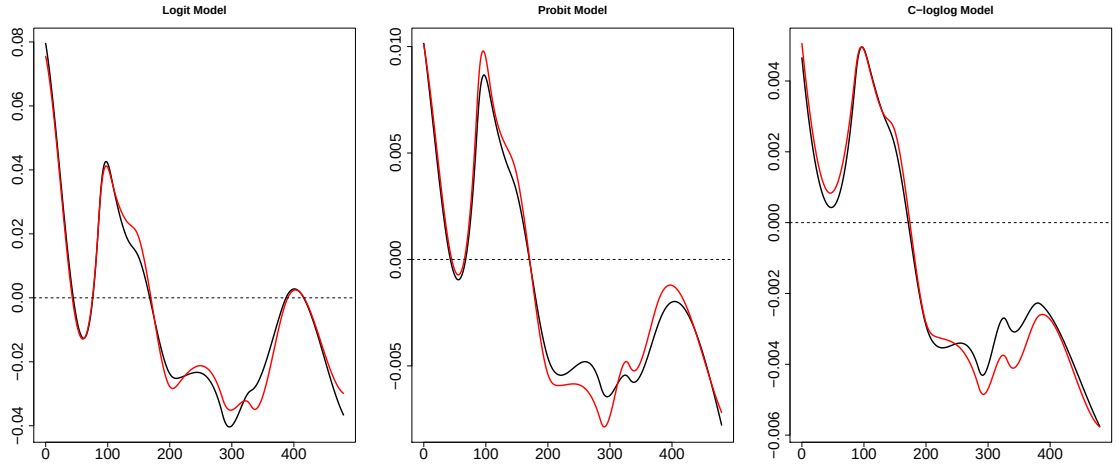

Figure 3.7: Graphs of the average (over 100 replications of a CBS sample of size $N = 60$ with $Q^* = 44\%$ and $H^* = 25\%$) parameter function estimates obtained using the OML method (black curve) and CML method (red curve).

for all $\theta(\cdot) \in L^2(\mathcal{T})$, $X \in \mathcal{X}$.

$$\tilde{\eta} = \sum_{i=0}^{p_N} \theta_i \varepsilon_i, \qquad \tilde{\eta}^* = \alpha^* + \sum_{i=1}^{p_N} \theta_i^* \varepsilon_i \qquad \text{for all} \qquad \theta \in \Theta.$$

The convergence in probability of $\widehat{\theta^\dagger}$ to $\theta^\dagger$ can be deduced from the following Lemma.

Lemma 3.3. *Under assumptions (H1) and (H2), we have*

$$\widehat{\theta^\dagger} - \theta^\dagger = o_p(1).$$

Proof of Lemma 3.3

Let

$$F(\theta(\cdot), X) = E_s\left(\{Y \log\left(\mu(\tilde{\eta})\right) + (1 - Y) \log\left(1 - \mu(\tilde{\eta})\right)\} | X\right).$$

We utilize the definition of E_s,

$$\begin{aligned} F(\theta(\cdot), X) &= \mu(\eta^*) \log\left(\mu(\tilde{\eta}) + (1 - \mu(\eta^*)) \log\left(1 - \mu(\tilde{\eta})\right)\right. \\ &= \tilde{F}(\theta, \varepsilon) + (\mu(\eta^*) - \mu(\tilde{\eta}^*)) \log\left(\frac{\mu(\tilde{\eta})}{1 - \mu(\tilde{\eta})}\right), \end{aligned}$$

with

$$\tilde{F}(\theta, \varepsilon) = \mu(\tilde{\eta}^*) \log\left(\mu(\tilde{\eta}) + (1 - \mu(\tilde{\eta}^*)) \log\left(1 - \mu(\tilde{\eta})\right)\right.$$

Using assumption (H2) and the approximation of the truncation strategy, we deduce that

$$E_s\left((\mu(\eta^*) - \mu(\tilde{\eta}^*))^2\right) = o(1), \tag{3.18}$$

and that $\log\left(\frac{\mu(\tilde{\eta})}{1-\mu(\tilde{\eta})}\right)$ is bounded uniformly on $\mathcal{X}$ and Θ. Therefore,

$$E_s(F(\theta(\cdot), X)) = E_s(\tilde{F}(\theta, \varepsilon)) + o(1). \tag{3.19}$$

Under CBS, the observations $(Y_n, \{X_n(t), t \in \mathcal{T}\})$, $n = 1, \ldots, N$, are i.i.d with a distribution characterized by E_s. However, by the law of large numbers, we have, for all $\theta \in \mathbb{R}^{p_N+1}$,

$$\frac{1}{N}\tilde{L}_{p_N}(\theta) - E_s\left(F(\theta(\cdot), X)\right) = o_p(1), \tag{3.20}$$

where $\tilde{L}_{p_N}$ is the truncated log-likelihood defined in (3.10).
Therefore,

$$\frac{1}{N}\tilde{L}_{p_N}(\theta) - E_s(\tilde{F}(\theta, \varepsilon)) = o_p(1). \tag{3.21}$$

The function $\tilde{F}(\theta, \varepsilon)$ can be regarded as the conditional expectation of some binary random variable $\tilde{Y}$ given ε such that $E_s(\tilde{Y}|\varepsilon) = \mu(\tilde{\eta}^*)$. By Lemma 3.1 and under assumption $(H2)$, $E_s(F(\theta, \varepsilon))$ attains its unique maximum over $\theta \in \Theta$ at $\theta = \theta^\dagger$. Hence, by (3.21) and Lemma 3.2, $\widehat{\theta^\dagger}$ converges in probability to $\theta^\dagger$.

■

Proof of Theorem 3.1. This proof is an adaptation of the proof of Theorem 4.1 of Müller & Stadtmüller (2005). Our notation is also similar to that of these authors. For completeness, we present the main steps of the proofs. Let $\|M\|_2^2 = (\sum_{k,l} m_{kl}^2)$ denote the matrix norm considered here. Using $\mu(\cdot)$ and η, we rewrite the pseudo-likelihood (3.10) as

$$\tilde{L}_{P_N}(\theta) = \sum_{n=1}^{N} Y_n \log(\mu(\tilde{\eta}_n)) + (1 - Y_n)\log(1 - \mu(\tilde{\eta}_n)).$$

We will denote by $U(\theta)$ the gradient of this function, defined as

$$U(\theta) = \frac{\partial}{\partial\theta}\tilde{L}_{p_N}(\theta) = \sum_{n=1}^{N} \frac{\mu^{'}(\tilde{\eta}_n)}{\tilde{\sigma}^2(\tilde{\eta}_n)}(Y_n - \mu(\tilde{\eta}_n))\varepsilon^{(n)}, \tag{3.22}$$

with $\tilde{\sigma}(\cdot) = \sigma(\mu(\cdot))$; by definition, $U(\widehat{\theta^\dagger}) = 0$. Let $J_{\theta^\dagger}$ denote the Hessian matrix at $\theta^\dagger$, that is,

$$\begin{aligned}
J_{\theta^\dagger} &= \frac{\partial}{\partial\theta^T}U(\theta)\Big|_{\theta^\dagger} = \sum_{n=1}^{N} \frac{\partial}{\partial\tilde{\eta}} \left\{ \frac{\mu^{'}(\tilde{\eta}_n)}{\tilde{\sigma}^2(\tilde{\eta}_n)}(Y_n - \mu(\tilde{\eta}_n))\varepsilon^{(n)} \right\}\Big|_{\tilde{\eta}_n^*} \frac{\partial}{\partial\theta}\tilde{\eta}_n\Big|_{\theta^\dagger} \\
&= -\sum_{n=1}^{N} \frac{\mu^{'2}(\tilde{\eta}_n^*)}{\tilde{\sigma}^2(\tilde{\eta}_n^*)}\varepsilon^{(n)}\varepsilon^{(n)T} \\
&\qquad + \sum_{n=1}^{N}(Y_n - \mu(\tilde{\eta}_n^*)) \left\{ \frac{\mu^{''}(\tilde{\eta}_n^*)}{\tilde{\sigma}^2(\tilde{\eta}_n^*)} - \frac{\mu^{'}(\tilde{\eta}_n^*)\tilde{\sigma}^{2'}(\tilde{\eta}_n^*)}{\tilde{\sigma}^4(\tilde{\eta}_n^*)} \right\} \varepsilon^{(n)}\varepsilon^{(n)T} \\
&= -D^T D + R,
\end{aligned}$$

where $D = \left(\mu^{'}(\tilde{\eta}_n^*)\varepsilon_k^{(n)}/\tilde{\sigma}(\tilde{\eta}_n^*)\right)_{1\le n\le N, 0\le k\le p_N}$. As in Müller & Stadtmüller (2005), we would like to show that the term R can be neglected. Now, we apply a Taylor expansion to $U(\cdot)$ for $\widetilde{\theta^\dagger}$ between $\theta^\dagger$ and $\widehat{\theta^\dagger}$, obtaining

$$U(\theta^\dagger) = \left\{D^T D + (J_{\theta^\dagger} - J_{\widetilde{\theta^\dagger}}) - (J_{\theta^\dagger} + D^T D)\right\}(\widehat{\theta^\dagger} - \theta^\dagger).$$

Then, we have

$$\begin{aligned}
\sqrt{N}(\widehat{\theta^\dagger} - \theta^\dagger) &= \left\{ I_{p_N+1} + \left(\frac{D^T D}{N}\right)^{-1} \left(\frac{J_{\theta^\dagger} - J_{\widetilde{\theta^\dagger}}}{N}\right) \right. \\
&\qquad \left. - \left(\frac{D^T D}{N}\right)^{-1} \left(\frac{J_{\theta^\dagger} + D^T D}{N}\right) \right\}^{-1} \left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}.
\end{aligned}$$

By assumption (H2), we have $\|\mu^{(r)}\| < C$, $r = 1, 2$, $\tilde{\sigma}^2(\cdot) < C$ and $\tilde{\sigma}(\cdot) > \delta$, and thus,

$$E_s\left(\left\|\frac{J_{\theta^\dagger} + D^T D}{N}\right\|_2^2\right) = O\left(\frac{p_N^2}{N}\right).$$

Therefore, (H3) implies that

$$\left\|\left(\frac{D^T D}{N}\right)^{-1} \frac{J_{\theta^\dagger} + D^T D}{N} \left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}\right\|_2 = o_p(1).$$

Since $\widetilde{\theta^\dagger}$ converges in probability to $\theta^\dagger$ by Lemma 3.3, we have

$$\left\|\left(\frac{D^T D}{N}\right)^{-1} \frac{J_{\theta^\dagger} - J_{\widetilde{\theta^\dagger}}}{N} \left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}\right\|_2 = o_p(1).$$

Then, it follows that as $N \to \infty$,

$$\left\|\sqrt{N}(\widehat{\theta^\dagger} - \theta^\dagger) - \left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}\right\|_2 = o_p(1).$$

Now, the asymptotic distribution of $\sqrt{N}(\widehat{\theta^\dagger} - \theta^\dagger)$ is seen as that of

$$\left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}.$$

Let us define the $(p_N + 1)$ vector $\mathcal{Z}_N$ and the $(p_N + 1) \times (p_N + 1)$ matrix Ψ_N as follows:

$$\mathcal{Z}_N = \Xi_{p_N}^{1/2} \frac{D^T e}{\sqrt{N}}; \qquad \Psi_N = \Delta_{p_N}^{1/2} \left(\frac{D^T D}{N}\right)^{-1} \Delta_{p_N}^{1/2},$$

where $e_n = (Y_n - \mu(\tilde{\eta}_n^*))/\tilde{\sigma}(\tilde{\eta}_n^*)$, $n = 1, \ldots, N$ are the components of the vector e.
We consider the following decomposition:

$$\begin{aligned}
&\left\{\left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}\right\}^T \Delta_{p_N} \left\{\left(\frac{D^T D}{N}\right)^{-1} \frac{U(\theta^\dagger)}{\sqrt{N}}\right\} \\
&\qquad = \mathcal{Z}_N^T \Psi_N^2 \mathcal{Z}_N \\
&\qquad = \mathcal{Z}_N^T \mathcal{Z}_N + 2\mathcal{Z}_N^T (\Psi_N - I_{p_N+1}) \mathcal{Z}_N \\
&\qquad\qquad + \mathcal{Z}_N^T (\Psi_N - I_{p_N+1})(\Psi_N - I_{p_N+1}) \mathcal{Z}_N \\
&\qquad \equiv F_N + G_N + H_N.
\end{aligned}$$

We have, by (H4)-(H5) and Proposition 7.1 in Müller & Stadtmüller (2005),

$$\left(\mathcal{Z}_N^T \mathcal{Z}_N - (p_N + 1)\right) / \sqrt{2p_N} \to \mathcal{N}(0, 1). \tag{3.23}$$

Thus, we deduce that $\left|\mathcal{Z}_N^T \mathcal{Z}_N\right| = O_p(p_N)$, and by (H3) and (H4),

$$\|\Psi_N - I_{p_N+1}\|_2 = o_p(1/\sqrt{p_N}). \tag{3.24}$$

Then, by (H3)-(H4) and using similar arguments as in Lemma 7.2 of Müller & Stadtmüller (2005), we have

$$|G_N| \le \left|\mathcal{Z}_N^T \mathcal{Z}_N\right| \|\Psi_N - I_{p_N+1}\|_2 = o_p(\sqrt{p_N}), \quad |H_N| \le \left|\mathcal{Z}_N^T \mathcal{Z}_N\right| \|\Psi_N - I_{p_N+1}\|_2^2 = o_p(1).$$

This completes the proof of Theorem 3.1.

■

Proof of Corollary 3.2. This proof is similar to that of Corollary 4.3 of Müller & Stadtmüller (2005) and is thus omitted.

■

Proof of Theorem 3.2 : By the law of large numbers, we have, for each $\theta \in \Theta$,

$$\frac{1}{N} L_{p_N}^{RS}(\theta) - E_s\left(Y \log\left(\frac{\Phi(\tilde{\eta})}{1-\Phi(\tilde{\eta})}\right) + \log\left(1-\Phi(\tilde{\eta})\right)\right) = o_p(1).$$

Through simple computations, one can prove that

$$\begin{aligned} & E_s\left(Y \log\left(\frac{\Phi(\tilde{\eta})}{1-\Phi(\tilde{\eta})}\right) + \log\left(1-\Phi(\tilde{\eta})\right)\right) \\ & \quad = E\left(\frac{H^*}{Q^*}\Phi(\tilde{\eta}^*)\log(\Phi(\tilde{\eta})) + \frac{1-H^*}{1-Q^*}(1-\Phi(\tilde{\eta}^*))\log(1-\Phi(\tilde{\eta}))\right) \\ & \qquad + E\left((\Phi(\eta^*)-\Phi(\tilde{\eta}^*))\log\left(\frac{(\Phi(\tilde{\eta}))^{H^*/Q^*}}{(1-\Phi(\tilde{\eta}))^{(1-H^*)/(1-Q^*)}}\right)\right) \qquad (3.25) \\ & \quad = F_{RS}(\theta) + o(1), \qquad (3.26) \end{aligned}$$

where

$$F_{RS}(\theta) \equiv E\left(\frac{H^*}{Q^*}\Phi(\tilde{\eta}^*)\log(\Phi(\tilde{\eta})) + \frac{1-H^*}{1-Q^*}(1-\Phi(\tilde{\eta}^*))\log(1-\Phi(\tilde{\eta}))\right),$$

and where the second term on the right-hand side of (3.25) is of order $o(1)$ by applying a truncation strategy (similar to (3.9)) and assumption (H2). Now, as in Lemma 3.3, the consistency of $\widehat{\theta^\dagger}_{\mathrm{RS}}$ when estimating $\theta^\dagger$ requires that

$$\frac{\partial F_{RS}}{\partial \theta}(\theta^\dagger) = E\left(\left(\frac{H^*}{Q^*} - \frac{1-H^*}{1-Q^*}\right)\Phi'(\tilde{\eta}^*)\varepsilon\right) = 0. \qquad (3.27)$$

This is not possible since $H^* \neq Q^*$ and $E(\varepsilon\Phi^{'}(\tilde{\eta}^*)) \neq 0$ by assumption. Consequently, (3.27) (first-order condition) will not be satisfied. Therefore, by Lemma 3.2, $\widehat{\theta^\dagger}_{RS}$ converges in probability to some $\tilde{\theta} \neq \theta^\dagger$ that effectively maximizes $F_{RS}(\cdot)$. This concludes the proof of the theorem.

■

Chapter 4

Functional linear spatial autoregressive models

Contents

Résumé en français

Précédemment, nous nous sommes intéressés à un modèle à choix binaire fonctionnel dans un cadre d'échantillonnage non aléatoire, dans ce chapitre nous allons considérer un modèle linéaire fonctionnel dans un carde spatial. Nous supposons que nous observons une variable réponse Y à valeurs dans $\mathbb{R}$ et une fonction aléatoire $\{X(t),\ t \in \mathcal{T}\}$ à valeurs dans $\mathcal{X} \subset L^2(\mathcal{T})$ en n unités spatiales situées dans une région $\mathcal{I}_n$ de cardinal n, incluse dans une région de type lattice, dénombrable $\mathcal{I} \subset \mathbb{R}^N$. Contrairement au Chapitre 3, nous supposons que ces observations ont été collectées via un processus d'échantillonnage aléatoire et sont spatialement dépendantes. Nous considérons que la structure de cette dépendance spatiale entre les n sites est décrite par une matrice de poids déterministe W_n (matrice d'interactions) $n \times n$, dont l'élément w_{ijn} est généralement défini en fonction de la distance entre les sites i et j par rapport à une certaine métrique (distance physique, lien sociale ou distance économique, voir par exemple Pinkse & Slade (1998)). Dans la suite, nous supposons que les n observations (X_i, Y_i), $i = 1, ..., n$, suivent le modèle linéaire fonctionnel autorégressif spatial suivant :

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ijn} Y_j + \int_{\mathcal{T}} X_i(t)\theta^*(t)dt + U_i, \qquad i = 1, \ldots, n,\ \ n = 1, 2, \ldots, \tag{4.1}$$

où le paramètre λ_0 d'autocorrélation spatiale au niveau de la variable réponse, $\theta^*(\cdot)$ le paramètre fonctionnel, sont inconnus. Les termes d'erreurs $\{U_i,\ i=1,\ldots,n,\ n=1,2,\ldots\}$ sont supposées centrées indépendantes et identiquement distribuées avec $E(U_i^2)=\sigma_0^2$. Comme dans le chapitre précédent, nous nous intéressons à l'estimation des paramètres λ_0, $\theta^*(.)$ et σ_0^2. Nous supposons que les éléments de la matrice de poids spatial W_n vérifient $w_{ij}=O(h_n^{-1})$ uniformément á i,j, oú $h_n=o(n)$.

Soit $\mathbf{X}_n(\theta^*(.))$ le vecteur $(n\times 1)$ dont le i-éme élément est $\int_{\mathcal{T}} X_i(t)\theta^*(t)dt$, nous pouvons alors ré-écrire (4.1) sous la forme matricielle suivante :

$$S_n\mathbf{Y}_n=\mathbf{X}_n(\theta^*(.))+\mathbf{U}_n\,,\qquad n=1,2,\ldots$$

où $S_n=(I_n-\lambda_0 W_n)$, $\mathbf{Y}_n$ et $\mathbf{U}_n$ sont deux vecteurs dont les éléments sont Y_i et U_i, $i=1,\ldots,n$ respectivement. Par conséquence, le logarithme de la fonction de quasi vraisemblance conditionnelle est défini par

$$\begin{aligned}L_n(\lambda,\theta(.),\sigma^2)=&-\frac{n}{2}\ln\sigma^2-\frac{n}{2}\ln(2\pi)+\ln|S_n(\lambda)|\\&-\frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n-\mathbf{X}_n(\theta(.))\right]'\left[S_n(\lambda)\mathbf{Y}_n-\mathbf{X}_n(\theta(.))\right],\end{aligned}\tag{4.2}$$

avec $S_n(\lambda)=I_n-\lambda W_n$.
Contrairement au cadre fonctionnel, des travaux d'estimation de ce type de modèles existent dans le cas où la variable explicative X est à valeurs réelles (voir par exemple Lee (2004), qui ont défini des estimateurs du quasi maximum de vraisemblance pour λ_0, le vecteur de paramètres θ^* et σ_0^2, en maximisant l'équivalent de (4.2)).
Dans le cadre du modèle fonctionnel (4.1) considéré, nous proposons une méthode d'estimation qui étend les travaux existants dans le cadre réel à l'aide d'une réduction de la dimension infinie de l'espace de la variable explicative fonctionnelle $X(\cdot)$. Nous utilisons la technique de troncature utilisée dans le chapitre précédent. Soit $\{\varphi_j,\ j=1,2,\ldots\}$ une base orthonormale de $L^2(\mathcal{T})$. On peut récrire $X(t)$ et $\theta^*(t)$ comme suit :

$$X(t)=\sum_{j\geq 1}\varepsilon_j\varphi_j(t)\qquad\text{et}\qquad\theta^*(t)=\sum_{j\geq 1}\theta_j^*\varphi_j(t),$$

où les variables aléatoires réelles ε_j et les coefficients θ_j^* sont définis par

$$\varepsilon_j=\int_{\mathcal{T}}X(t)\varphi_j(t)dt\qquad\text{et}\qquad\theta_j^*=\int_{\mathcal{T}}\theta^*(t)\varphi_j(t)dt.$$

Nous avons alors :

$$\int_{\mathcal{T}}X(t)\theta^*(t)dt=\sum_{j\geq 1}\theta_j^*\varepsilon_j=\sum_{j=1}^{p_n}\theta_j^*\varepsilon_j+\sum_{j=p_n+1}^{\infty}\theta_j^*\varepsilon_j,\tag{4.3}$$

où p_n est une suite d'entiers naturels, qui croît asymptotiquement avec n. Nous reprenons l'idée d'approximation utilisée dans le chapitre 3 avec une fonction de lien identité ($\Phi(t)=t$). Le logarithme de la quasi vraisemblance conditionnelle tronquée est obtenu en approchant la partie à gauche de (4.3) par le premier terme de la partie à droite de la décomposition précédente. Par conséquent, nous avons

$$\begin{aligned}\tilde{L}_n(\lambda,\theta,\sigma^2)\;=\;&-\frac{n}{2}\ln\sigma^2-\frac{n}{2}\ln(2\pi)+\ln|S_n(\lambda)|\\&-\frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n-\xi_{p_n}\theta\right]'\left[S_n(\lambda)\mathbf{Y}_n-\xi_{p_n}\theta\right].\end{aligned}\tag{4.4}$$

où ξ_{p_n} est une $n \times p_n$ matrice dont les éléments

$$\xi_{p_n ij} = \int_{\mathcal{T}} X_i(t)\varphi_j(t)dt\,, \quad j = 1, \ldots, p_n, \quad i = 1, \ldots, n.$$

Pour un λ fixé, (4.4) est maximisée par

$$\hat{\theta}_{n,\lambda} = (\xi'_{p_n}\xi_{p_n})^{-1}\xi'_{p_n}S_n(\lambda)\mathbf{Y}_n = (\hat{\theta}_{nj,\lambda})_{j=1,\ldots,p_n}$$

et

$$\begin{aligned}\hat{\sigma}^2_{n,\lambda} &= \frac{1}{n}\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda}\right)'\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda}\right)\\ &= \frac{1}{n}\mathbf{Y}'_n S'_n(\lambda)M_n S_n(\lambda)\mathbf{Y}_n,\end{aligned}$$

où $M_n = I_n - \xi_{p_n}(\xi'_{p_n}\xi_{p_n})^{-1}\xi'_{p_n}$.
Par conséquent, le logarithme de la quasi vraisemblance conditionnelle tronquée correspondant à λ est défini par

$$\tilde{L}_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}^2_{n,\lambda} + \ln|S_n(\lambda)|.$$

Ainsi, l'estimateur de λ_0 est le paramètre $\hat{\lambda}_n$ qui maximise $\tilde{L}_n(\lambda)$, et ceux du vecteur θ^* est la variance σ_0^2 sont $\hat{\theta}_{n,\hat{\lambda}_n}$ et $\hat{\sigma}^2_{n,\hat{\lambda}_n}$ respectivement. On déduit de ces derniers l'estimateur du paramètre fonctionnel

$$\hat{\theta}_n(t) = \sum_{j=1}^{p_n}\hat{\theta}_{nj,\hat{\lambda}_n}\varphi_j(t).$$

Dans la suite, nous étudions l'identification des paramètres étudiés et les comportements asymptotiques des estimateurs proposés, notamment la normalité asymptotique. Sous des conditions non restrictives, nous montrons que les paramètres λ_0, σ_0^2 et le paramètre fonctionnel $\theta^*(\cdot)$, sont identifiables et que

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, s^2_\lambda) \qquad \text{et} \qquad \sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) \to \mathcal{N}(0, s^2_\sigma),$$

avec

$$s^2_\lambda = \lim_{n\to\infty}\frac{s_n^2 h_n}{n}\left\{\frac{h_n}{n}\left[\Delta_n + \sigma_0^2\mathrm{tr}(G_n(G'_n + G_n))\right]\right\}^{-2};\ s^2_\sigma = \mu_4 - \sigma_0^4 + 4s^2_\lambda\lim_{n\to\infty}h_n\left[\frac{\mathrm{tr}(G_n)}{n}\right]^2,$$

où

$$\begin{aligned}s_n^2 &= \sigma_0^2\left[\theta^{*\prime}\Gamma_{p_n}\theta^* + \sigma_0^2\right]\mathrm{tr}\left(G_n(G'_n + G_n)\right) + \left[3\sigma_0^2\theta^{*\prime}\Gamma_{p_n}\theta^* + \sigma_0^4 - \mu_4\right]\frac{1}{n}\mathrm{tr}^2(G_n)\\ &\quad + \left[\mu_4 - 3\sigma_0^4 - \sigma_0^2\theta^{*\prime}\Gamma_{p_n}\theta^*\right]\sum_{i=1}^{n}G_{ii}^2,\end{aligned}$$

tel que $G_n = S_n^{-1}W_n$, $\Gamma_{p_n} = E\left(\frac{1}{n}\xi'_{p_n}\xi_{p_n}\right)$, $\mu_4 = E(U^4)$, et $\theta^* = (\theta_1^*, \ldots, \theta_{p_n}^*)'$.
Nous déduisons de ces résultats que

$$\frac{n\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right)'\Gamma_{p_n}\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right) - p_n}{\sqrt{2p_n}} \to \mathcal{N}(0, \sigma_0^4),$$

et

$$\frac{nd^2\left(\hat{\theta}_n(\cdot), \theta^*(\cdot)\right) - p_n}{\sqrt{2p_n}} \to \mathcal{N}(0, \sigma_0^4),$$

où $d(\cdot,\cdot)$ est une métrique dans $L^2(\mathcal{T})$, définie par

$$d^2(f,g) = \int\int (f(t)-g(t))\, E(X(t)X(v))\, (f(v)-g(v))\, dtdv,\ f,g \in L^2(\mathcal{T}).$$

Des résultats numériques basés sur des données simulées et de concentration d'ozone au Sud-Est des États-Unis montrent la performance du modèle proposé ainsi que l'utilité de prendre en consideration la dependence spatiale.

The results of this chapter are in collaboration with Laurence Broze (University of Lille), Sophie Dabo-Niang (University of Lille) and Zied Gharbi (University of Lille). A related paper is submitted as a book chapter.

4.1 Introduction

This work addresses two research areas: spatial statistics and functional data analysis. Spatial functional random variables are becoming more common in statistical analyses due to the availability of high-frequency spatial data and new mathematical strategies to address such statistical objects.

Many fields, such as urban systems, agriculture, environmental science and economics, often consider spatially dependent data. Therefore, modeling spatial dependency in statistical inferences (estimation of the spatial distribution, regression, prediction, ...) is a significant feature of spatial data analysis. Spatial statistics provide tools to solve such modelling. Various spatial models and methods have been proposed, particularly within the scope of geostatistics or lattice data. Most of the spatial modeling methods are parametric and concern non-functional data. Several types of functional linear models for independent data have been developed for different purposes. The most studied model is perhaps the functional linear model for scalar response, originally introduced by Hastie & Mallows (1993). Estimation and prediction problems for this model and some of its generalizations have been reported mainly for independent data (see, e.g., Crambes et al. (2009), Comte & Johannes (2012), Cai & Yuan (2012), Cuevas (2014)). Some research exists on functional spatial linear prediction using kriging methods (see, e.g., Nerini et al. (2010), Giraldo et al. (2010), Giraldo et al. (2011), Horváth & Kokoszka (2012), Giraldo (2014), Bohorquez et al. (2016), Bohorquez et al. (2017),...), highlighting the interest in considering spatial linear functional models.

Complex issues arise in spatial econometrics (statistical techniques to address economic modeling), many of which are neither clearly defined nor completely resolved but form the basis for current research. Among the practical considerations that influence the available techniques used in spatial data modeling, particularly in econometrics, is data dependency. In fact, spatial data are often dependent, and a spatial model must be able to account for this characteristic. Linear spatial models, which are common in geostatistical modeling, generally impose a dependency structure model based on linear covariance relationships between spatial locations. However, under many circumstances, the spatial index does not vary continuously over a subset of $\mathbb{R}^N$, $N \geq 2$ and may be of the lattice type, the baseline of this current work. This is, for instance, the case in a number of problems. In images analysis, remote sensing form satellites, agriculture and so one, data are often received as regular lattice and identified as the centroids of square pixels, whereas a mapping forms often an irregular lattice. Basically, statistical models for lattice data are linked to nearest neighbors to express the fact that data are nearby.

We are concerned here about spatial functional models for lattice data. One of the well-known spatial lattice models is the spatial autoregressive model (SAR) of Cliff &

Ord (1973), which extends regression in time series to spatial data. This model has been extensively studied and extended in several ways in the case of real-valued data, compare to the functional framework. Ruiz-Medina (2011) and Ruiz-Medina (2012) considered a spatial unilateral autoregressive Hilbertian (SARH(1)) processes where the autoregressive part is given in terms of three functional random components located in three points defining the boundary between some notions of past and future.

The structure of SAR model for real-valued data and its identification and estimation by the two stage least squares (2SLS), the three stage least squares (3SLS), the maximum likelihood (ML) and the generalized method of moments (GMM) estimation methods have been developed and summarized by many authors, such as Anselin (1988), Case (1993), Kelejian & Prucha (1998), Kelejian & Prucha (1999), Lee (2007), Lin & Lee (2010), Zheng & Zhu (2012), Malikov & Sun (2017), Garthoff & Otto (2017),... The identification and estimation of such SAR models by the quasi-maximum likelihood (QML) are limited. Lee (2004) and more recently Yang & Lee (2017), proposed the quasi-maximum likelihood estimator for the SAR model with a spatial dependency structure based on a spatial weights matrix. The quasi-maximum likelihood estimator (QMLE) is appropriate when the disturbances in the considered model are not normally distributed. In the literature on SAR models for real-valued data, the QMLE and maximum likelihood estimator (MLE) are proved to be computationally challenging, consistent with rates of convergence depending on the spatial weights matrix of the considered model (Lee, 2004; Yang & Lee, 2017). These last works considered real-valued random responses and deterministic or random real-valued covariates and investigated the asymptotic properties of the QMLE estimator under some disturbance specifications.

The present work considers an estimation of a spatial functional linear model with a random functional covariate and a real-valued response using spatial autoregression on the response based on a weight matrix. We investigate parameter identification and asymptotic properties of the QMLE estimator using the so-called *increasing domain asymptotics*. We provide identification conditions combining identification in the classical SAR model and identification in the functional linear model. Monte Carlo experiments illustrate the performance of the QML estimation.

The rest of this chapter is organized as follows. In Section 4.2, we provide the functional SAR (FSAR) and its quasi-likelihood estimator (QML). In Section 4.3, we state the consistency and asymptotic normality of the estimator. To check the performance of the estimator, numerical results are reported in Section 4.4 using different spatial scenarios, where each unit is influenced by neighboring units. Proofs and technical lemmas are given in the Appendix.

4.2 Model

We consider that at n spatial units located on $\mathcal{I}_n$, a finite subset of cardinal n of a regular or irregularly spaced, countable lattice $\mathcal{I} \subset \mathbb{R}^N$, we observe a real-valued random variable Y considered as the *response variable* and a functional covariate $\{X(t), t \in \mathcal{T}\}$, a square-integrable stochastic process on the interval $\mathcal{T} \subset \mathbb{R}$. Assume that the process $\{X(t), t \in \mathcal{T}\}$ takes values in space $\mathcal{X} \subset L^2(\mathcal{T})$, where $L^2(\mathcal{T})$ is the space of square-integrable functions in $\mathcal{T}$. The spatial dependency structure between these n spatial units is described by an $n \times n$ non-stochastic spatial weights matrix W_n that depends on n. The elements $w_{ij} = w_{ijn}$ of this matrix are usually considered as inversely proportional to the distance between spatial units i and j with respect to some metric, see Chapter 2. Since the weight matrix changes with n, we consider these observations as triangular array observations. This is required to conduct an asymptotic study of the following model that describes the

relationship between the response variable Y and the covariate function $X(.)$ (Robinson, 2011).
Here, we assume that the relationship between Y and X is modelled by the following functional spatial autoregressive model (FSAR) with endogenous interactions:

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \int_{\mathcal{T}} X_i(t)\theta^*(t)dt + U_i, \qquad i = 1, \ldots, n, \;\; n = 1, 2, \ldots \tag{4.5}$$

where the autoregressive parameter λ_0 is in compact space Λ, $\theta^*(\cdot)$ is a parameter function assumed to belong to the space of functions $L^2(\mathcal{T})$, and $(w_{ij})_{j=1,\ldots,n}$ is the i-th row of W_n. Assume that $w_{ij} = O(h_n^{-1})$ uniformly in all i, j, where the rate sequence h_n can be bounded or divergent, such as $h_n = o(n)$. This kind of matrix can be obtained by Nearest Neighbor weights.
In practice, it is common, but not necessary, to *row normalize* the spatial weight matrix. The row-standardization helps the interpretation and the comparison of weight matrices and parameters λ, it allows $0 \leq w_{ij} \leq 1$ and $-1 \leq \lambda \leq 1$. In this way, the spatially-lagged variables are equal to a weighted average of the neighboring values.

The disturbances $\{U_i,\ i = 1, \ldots, n,\ n = 1, 2, \ldots\}$ are assumed to be independent random variables such that $E(U_i) = 0$, $E(U_i^2) = \sigma_0^2$. They are also independent of $\{X_i(t), t \in \mathcal{T},\ i = 1, \ldots, n,\ n = 1, 2 \ldots\}$.
We are interested in estimating the unknown true parameters λ_0, $\theta^*(.)$ and σ_0^2. Let $\mathbf{X}_n(\theta^*(.))$ be the $n \times 1$ vector of i-th element $\int_{\mathcal{T}} X_i(t)\theta^*(t)dt$; then, one can rewrite (4.5) as

$$S_n \mathbf{Y}_n = \mathbf{X}_n(\theta^*(.)) + \mathbf{U}_n\,, \qquad n = 1, 2, \ldots$$

where $S_n = (I_n - \lambda_0 W_n)$, $\mathbf{Y}_n$ and $\mathbf{U}_n$ are two $n \times 1$ vectors of elements Y_i and U_i, $i = 1, \ldots, n$ respectively, and I_n denotes the $n \times n$ identity matrix.
Let $S_n(\lambda) = I_n - \lambda W_n$, so the conditional log-likelihood function of the vector $\mathbf{Y}_n$ given $\{X_i(t), t \in \mathcal{T},\ i = 1, \ldots, n,\ n = 1, 2 \ldots\}$ is given by:

$$\begin{aligned} L_n(\lambda, \theta(.), \sigma^2) = &-\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| \\ &- \frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\theta(.))\right]' \left[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\theta(.))\right], \end{aligned} \tag{4.6}$$

where A' denotes the transpose of matrix A.
The quasi-maximum likelihood estimates of λ_0, $\theta^*(\cdot)$ and σ_0^2 are the parameters λ, $\theta(\cdot)$, and σ^2 that maximize (4.6). But this likelihood cannot be maximized without addressing the difficulty produced by the infinite dimensionality of the explanatory random function. To solve this problem, we use the dimension reduction technique described in Chapter 3.

4.2.1 Truncated conditional likelihood method

Let $\{\varphi_j,\ j = 1, 2, \ldots\}$ be an orthonormal basis of the functional space $L^2(\mathcal{T})$, usually a Fourier or a Spline basis or a basis constructed by the eigenfunctions of the covariance operator Γ, where the operator is defined by:

$$\Gamma x(t) = \int_{\mathcal{T}} E(X(t)X(s))x(s)ds\,, \qquad x \in \mathcal{X},\, t \in \mathcal{T}. \tag{4.7}$$

Using an expansion on this orthonormal basis, we can write $X(.)$ and $\theta^*(.)$ in as follows:

$$X(t) = \sum_{j \geq 1} \varepsilon_j \varphi_j(t) \qquad \text{and} \qquad \theta^*(t) = \sum_{j \geq 1} \theta_j^* \varphi_j(t) \qquad \text{for all } \; t \in \mathcal{T},$$

where the real random variables ε_j and the coefficients θ_j^* are given by

$$\varepsilon_j = \int_{\mathcal{T}} X(t)\varphi_j(t)dt \qquad \text{and} \qquad \theta_j^* = \int_{\mathcal{T}} \theta^*(t)\varphi_j(t)dt.$$

Let p_n be a positive sequence of integers that increase asymptotically as $n \to \infty$; by the orthonormality of the basis $\{\varphi_j,\ j = 1, 2, \ldots\}$, we can consider the following decomposition

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j=1}^{\infty} \theta_j^*\varepsilon_j = \sum_{j=1}^{p_n} \theta_j^*\varepsilon_j + \sum_{j=p_n+1}^{\infty} \theta_j^*\varepsilon_j. \tag{4.8}$$

The truncation strategy introduced by Müller & Stadtmüller (2005) consists of approximating the left-hand side in (4.8) using only the first term of the right-hand side. This is possible when the approximation error vanishes asymptotically, where this error is controlled by a square expectation of the second term on the right-hand side of (4.8). In particular, the approximation error vanishes asymptotically when one considers the eigenbasis of the variance-covariance operator Γ by remarking that

$$E\left(\sum_{j=p_n+1}^{\infty} \theta_j^*\varepsilon_j\right)^2 = \sum_{j=p_n+1}^{\infty} {\theta_j^*}^2 E\left(\varepsilon_j^2\right) = \sum_{j=p_n+1}^{\infty} {\theta_j^*}^2 \delta_j$$

where $\delta_j, j = 1, 2 \ldots$ are the eigenvalues. Under this truncation strategy, $\mathbf{X}_n(\theta^*(.))$ may be approximated by $\xi_{p_n}\theta^*$, where $\theta^* = (\theta_1^*, \ldots, \theta_{p_n}^*)'$ and ξ_{p_n} is an $n \times p_n$ matrix of the (i, j)-th element given by

$$\varepsilon_j^{(i)} = \int_{\mathcal{T}} X_i(t)\varphi_j(t)dt\,, \qquad i = 1, \ldots, n\ \ j = 1, \ldots, p_n.$$

Now, the truncated conditional log-likelihood function can be obtained by replacing (4.6) $\mathbf{X}_n(\theta(.))$ with $\xi_{p_n}\theta$ for all $\theta(.) \in L^2(\mathcal{T})$ and $\theta \in \mathbb{R}^{p_n}$. The corresponding and feasible log conditional likelihood is

$$\begin{aligned} \tilde{L}_n(\lambda, \theta, \sigma^2) &= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| \\ &\quad -\frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right]'\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right]. \end{aligned} \tag{4.9}$$

For a fixed λ, (4.9) is maximized at

$$\hat{\theta}_{n,\lambda} = (\xi_{p_n}'\xi_{p_n})^{-1}\xi_{p_n}' S_n(\lambda)\mathbf{Y}_n = (\hat{\theta}_{nj,\lambda})_{j=1,\ldots,p_n}$$

and

$$\begin{aligned} \hat{\sigma}_{n,\lambda}^2 &= \frac{1}{n}\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda}\right)'\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda}\right) \\ &= \frac{1}{n}\mathbf{Y}_n' S_n'(\lambda) M_n S_n(\lambda)\mathbf{Y}_n, \end{aligned}$$

where $M_n = I_n - \xi_{p_n}(\xi_{p_n}'\xi_{p_n})^{-1}\xi_{p_n}'$.
The concentrated truncated conditional log-likelihood function of λ is:

$$\tilde{L}_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}_{n,\lambda}^2 + \ln|S_n(\lambda)|.$$

Then the estimator of λ_0 is $\hat{\lambda}_n$, which maximizes $\tilde{L}_n(\lambda)$, and those of the vector θ^* and variance σ_0^2 are, respectively, $\hat{\theta}_{n,\hat{\lambda}_n}$, $\hat{\sigma}^2_{n,\hat{\lambda}_n}$. The corresponding estimator of the function parameter $\theta^*(.)$ is:

$$\hat{\theta}_n(t) = \sum_{j=1}^{p_n} \hat{\theta}_{nj,\hat{\lambda}_n}\varphi_j(t).$$

The estimation of the model is given, we focus on the asymptotics results in the next section.
For that purpose, we need to define some asymptotic method. As mentioned in Chapter 2, there are two main asymptotic methods in the spatial literature: increasing domain and infill asymptotics (see Cressie, 1993, p. 480). In the following, we consider increasing domain asymptotics.

4.3 Assumptions and results

Let us first state some combining condition assumptions related to the spatial dependency structure and assumptions on the functional nature of the data.
Let $I_n + \lambda_0 G_n = S_n^{-1}$ where $G_n = W_n S_n^{-1}$, $B_n(\lambda) = S_n(\lambda)S_n^{-1} = I_n + (\lambda_0 - \lambda)G_n$ for all $\lambda \in \Lambda$ and $A_n(\lambda) = B_n^{'}(\lambda)B_n(\lambda)$.
We assume that

Assumption 1

i. The matrix S_n, is nonsingular.
ii. The sequences of matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.
iii. The matrices $\{S_n^{-1}(\lambda)\}$ are uniformly bounded in either row or column sums and uniformly bounded in λ in compact parameter space Λ. The true λ_0 is in the interior of Λ.

Assumption 2 The sequence p_n satisfies $p_n \to \infty$ and $p_n n^{-1/4} \to 0$ as $n \to \infty$, and

$$\text{i.} \quad p_n \sum_{r_1, r_2 > p_n} E\left(\varepsilon_{r_1}\varepsilon_{r_2}\right) = o(1)$$

$$\text{ii.} \quad \sum_{r_1,\ldots,r_4 > p_n} E\left(\varepsilon_{r_1}\ldots\varepsilon_{r_4}\right) = o(1)$$

$$\text{iii.} \quad \sqrt{n}\sum_{s=1}^{p_n} \sum_{r_1, r_2 > p_n} E\left(\varepsilon_s\varepsilon_{r_1}\right) E\left(\varepsilon_s\varepsilon_{r_2}\right) = o(1).$$

Remark 4.1. ***Assumption 1****-i ensures that* $\mathbf{Y}_n$ *has mean* $S_n^{-1}\mathbf{X}_n(\theta^*(\cdot))$ *and variance* $\sigma_0^2 S_n^{-1}S_n^{'-1}$. *The uniform boundedness of* W_n *and* S_n^{-1} *in* ***Assumption 1****-ii enables the control of the degree of spatial correlation and plays an important role in the asymptotic properties of the estimators. By easy computation, one can show under this assumption that the matrix* $G_n = W_n S_n^{-1}$ *is uniformly bounded in both row and column sums together with elements of order* h_n^{-1}. *Consequently, the matrix* $A_n(\lambda) = B_n^{'}(\lambda)B_n(\lambda)$ *has a trace of order* n *uniformly in* $\lambda \in \Lambda$ *by the compactness condition of* Λ *in* ***Assumption 1****-iii.* ***Assumption 1****-iii makes it possible to address the nonlinearity of* $ln|S_n(\lambda)|$ *as a function of* λ *in (4.9). For more detail and a discussion of* ***Assumption 1****, see Lee (2004).* ***Assumption 2*** *considers the rate of convergence of* p_n *with respect to* n. *Condition iii of Assumption 2 is satisfied when one consider the eigenbasis, since in this case* $E(\varepsilon_r\varepsilon_s) = 0$, *for* $s \neq r$.

To obtain the identifiability of λ_0, $\theta^* = (\theta_1^*, \ldots, \theta_{p_n}^*)^{'}$, and σ_0^2 in the truncated model, remark that

$$\begin{aligned} E\left(\tilde{L}_n(\lambda, \theta, \sigma^2)\right) &= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| \\ &\quad -\frac{1}{2\sigma^2}E\left(\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right]^{'}\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right]\right). \end{aligned}$$

We have

$$\begin{aligned}
&E\left(\left[S_n(\lambda)\mathbf{Y}_n-\xi_{p_n}\theta\right]'\left[S_n(\lambda)\mathbf{Y}_n-\xi_{p_n}\theta\right]\right)\\
&\quad=E\left(\left[B_n(\lambda)\mathbf{X}_n(\theta^*(.))-\xi_{p_n}\theta\right]'\left[B_n(\lambda)\mathbf{X}_n(\theta^*(.))-\xi_{p_n}\theta\right]\right)+\sigma_0^2\mathrm{tr}\left(A_n(\lambda)\right)\\
&\quad=E\left(\left[B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta\right]'\left[B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta\right]\right)+E\left(R_n'A_n(\lambda)R_n\right)\\
&\qquad+\sigma_0^2\mathrm{tr}\left(A_n(\lambda)\right)+2E\left(\left[B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta\right]'B_n(\lambda)\mathbf{R}_n\right),
\end{aligned}$$

where $\mathbf{R}_n=(R_1,\ldots,R_n)'$ with $R_i=\sum_{j>p_n}\theta_j^*\varepsilon_j^{(i)}$. Let R denote the generic copy of $R_i, i=1,\ldots,n$, where $E(R)=0$.
We then have

$$E\left(\theta^{*'}\xi_{p_n}'B_n(\lambda)\mathbf{R}_n\right)=\mathrm{tr}\left(B_n(\lambda)\right)\epsilon_{n1},\quad\text{where}\ \ \epsilon_{n1}=\sum_{r=1}^{p_n}\sum_{s>p_n}\theta_r\theta_s^*E(\varepsilon_r\varepsilon_s),$$

$$E\left(\theta'\xi_{p_n}'A_n(\lambda)\mathbf{R}_n\right)=\mathrm{tr}\left(A_n(\lambda)\right)\epsilon_{n2},\quad\text{where}\ \epsilon_{n2}=\sum_{r=1}^{p_n}\sum_{s>p_n}\theta_r^*\theta_s^*E(\varepsilon_r\varepsilon_s),$$

$$E\left(\mathbf{R}_n'A_n(\lambda)\mathbf{R}_n\right)=\mathrm{tr}(A_n(\lambda))\epsilon_{n3},\quad\text{where}\ \epsilon_{n3}=E(R^2).$$

Note that $\epsilon_{n1},\epsilon_{n2}$ and ϵ_{n3} are of order $o(1)$ by **Assumption 2**, and they are independent of λ. In addition, ϵ_{n1} and ϵ_{n2} are null if one uses the eigenbasis.
Consequently,

$$\begin{aligned}
E\left(\tilde{L}_n(\lambda,\theta,\sigma^2)\right)&=-\frac{1}{2\sigma^2}E\left(\left(B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta\right)'\left(B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta\right)\right)\\
&\quad+\ln|S_n(\lambda)|-\frac{n}{2}\left(\ln\sigma^2+\ln 2\pi\right)-\frac{\sigma_0^2}{2\sigma^2}\mathrm{tr}\left(A_n(\lambda)\right)\\
&\quad+\epsilon_{n1}\mathrm{tr}\left(B_n(\lambda)\right)+\epsilon_{n4}\mathrm{tr}\left(A_n(\lambda)\right),
\end{aligned}\tag{4.10}$$

with $\epsilon_{n4}:=\epsilon_{n2}+\epsilon_{n3}$. Note that the terms that contain ϵ_{n1} and ϵ_{n4} are negligible with respect to the others by **Assumption 2**.
For fixed λ, the expectation $E\left(\tilde{L}_n(\lambda,\theta,\sigma^2)\right)$ is maximum with respect to θ and σ^2 at

$$\begin{aligned}
\theta_{n,\lambda}^*&=\frac{1}{n}\Gamma_{p_n}^{-1}E\left(\xi_{p_n}'B_n(\lambda)\xi_{p_n}\right)\theta^*\\
&=\theta^*+(\lambda_0-\lambda)\Gamma_{p_n}^{-1}\frac{1}{n}E\left(\xi_{p_n}'G_n\xi_{p_n}\right)\theta^*=\theta^*+(\lambda_0-\lambda)\theta^*\frac{1}{n}\mathrm{tr}\left(G_n\right)
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{n,\lambda}^{*2}&=\frac{1}{n}E\left(\left[B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta_{n,\lambda}^*\right]'\left[B_n(\lambda)\xi_{p_n}\theta^*-\xi_{p_n}\theta_{n,\lambda}^*\right]\right)+\frac{\sigma_0^2}{n}\mathrm{tr}\left(A_n(\lambda)\right)\\
&=(\lambda_0-\lambda)^2\frac{1}{n}\Delta_n+\frac{\sigma_0^2}{n}\mathrm{tr}\left(A_n(\lambda)\right),
\end{aligned}\tag{4.11}$$

with $\Delta_n=n\left(\mathrm{tr}\left(\frac{G_n'G_n}{n}\right)-\mathrm{tr}^2\left(\frac{G_n}{n}\right)\right)\theta^{*'}\Gamma_{p_n}\theta^*$ since

$$E\left(\xi_{p_n}'G_n'G_n\xi_{p_n}\right)=\mathrm{tr}(G_n'G_n)\Gamma_{p_n}\ \text{ and }\ E\left(\xi_{p_n}'G_n\xi_{p_n}\right)=\mathrm{tr}(G_n)\Gamma_{p_n},$$

where $\Gamma_{p_n}=E\left(\frac{1}{n}\xi_{p_n}'\xi_{p_n}\right)$ is assumed to be positive definite. This is the case when the eigenbasis is considered in the truncation strategy.

Based on these results, it is clear that $\theta^*_{n,\lambda_0} = \theta^*$ and $\sigma^{*2}_{n,\lambda_0} = \sigma_0^2$. Hence, the identifiability of θ^* and σ_0^2 depends on that of λ_0. Note that

$$\begin{aligned} Q_n(\lambda) &= E\left(\tilde{L}_n\left(\lambda, \theta^*_\lambda, \sigma^{*2}_{n,\lambda}\right)\right) \\ &= \ln|S_n(\lambda)| - \frac{n}{2}\ln\sigma^{*2}_{n,\lambda} - \frac{n}{2}\left(1 + \ln(2\pi)\right) + \epsilon_{n1}\mathrm{tr}\left(B_n(\lambda)\right) + \epsilon_{n4}\mathrm{tr}\left(A_n(\lambda)\right). \end{aligned}$$

Therefore, proving the identifiability of λ_0 is equivalent to showing that λ_0 maximizes $Q_n(\lambda)$. This will be proved before addressing the consistency of the estimators.

We will need to compose some additional assumptions

Assumption 3 Let $\lim_{n\to\infty} \frac{1}{n}\Delta_n = c$, where (a) $c > 0$; (b) $c = 0$. Under the latter condition,

$$\lim_{n\to\infty} \frac{h_n}{n}\left\{\ln\left|\sigma_0^2 S_n^{-1} S_n'^{-1}\right| - \ln\left|\sigma^2_{n,\lambda} S_n^{-1}(\lambda) S_n'^{-1}(\lambda)\right|\right\} \neq 0,$$

whenever $\lambda \neq \lambda_0$, with $\sigma^2_{n,\lambda} = \frac{\sigma_0^2}{n}\mathrm{tr}(A_n(\lambda))$.

Assumption 4 U_i, $i = 1, \ldots, n$ in $\mathbf{U}_n = (U_1, \ldots, U_n)'$ are i.i.d. with mean zero and variance σ_0^2. The moment $E\left(|U_i|^{4+\delta}\right)$ exists for some $\delta > 0$. Let $\mu_4 = E(U_i^4)$.

Remark 4.2. ***Assumption 3*** *enables the identification of λ_0 according to the boundless of h_n. It is similar to that used in Lee (2004) in the case of multivariate deterministic covariates. This assumption ensures that $tr^2(G_n/n)$ is dominated by $tr(G_n'G_n/n)$, which is the case when $h_n \to \infty$, as under* ***Assumption 1****, $tr(G_n'G_n)$ and $tr(G_n)$ are of order $O(n/h_n)$. Situation (b) is related to the existence of a unique variance of $\mathbf{Y}_n$.* ***Assumption 4*** *characterizes the properties of the disturbance term.*

Under assumptions similar to those used in Lee (2004) but adapted to the functional context, we show that the proposed QMLE estimator has the same asymptotic properties as those in the context of independent data(see e.g. Müller & Stadtmüller, 2005) and the spatial model with real-valued covariates (see e.g. Lee, 2004). The following theorems give the identification, consistency and asymptotic normality results of the autoregressive, functional and variance parameters estimates.

Theorem 4.1. *Under* ***Assumptions 1-4*** *and $h_n^4 = O(n)$ for divergent h_n, the QMLE $\hat{\lambda}_n$ derived from the maximization of $\tilde{L}_n(\lambda)$ is consistent and satisfies*

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, s_\lambda^2),$$

with $s_\lambda^2 = \lim_{n\to\infty} \frac{s_n^2 h_n}{n}\left\{\frac{h_n}{n}\left[\Delta_n + \sigma_0^2 tr(G_n(G_n' + G_n))\right]\right\}^{-2}$, *where*

$$\begin{aligned} s_n^2 &= \sigma_0^2\left[\theta^{*\prime}\Gamma_{p_n}\theta^* + \sigma_0^2\right] tr\left(G_n(G_n' + G_n)\right) + \left[3\sigma_0^2\theta^{*\prime}\Gamma_{p_n}\theta^* + \sigma_0^4 - \mu_4\right]\frac{1}{n}tr^2(G_n) \\ &\quad + \left[\mu_4 - 3\sigma_0^4 - \sigma_0^2\theta^{*\prime}\Gamma_{p_n}\theta^*\right]\sum_{i=1}^n G_{ii}^2. \end{aligned} \tag{4.12}$$

Note that, when h_n is divergent, the last two terms in (4.12) are negligible.

Theorem 4.2. *Under assumptions of Theorem 4.1, $\hat{\sigma}_n^2$ is a consistent estimator of σ_0^2 and satisfies*

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) \to \mathcal{N}(0, s_\sigma^2),$$

with

$$s_\sigma^2 = \mu_4 - \sigma_0^4 + 4s_\lambda^2 \lim_{n\to\infty} h_n \left[\frac{tr(G_n)}{n}\right]^2.$$

When h_n is divergent, s_σ^2 will be reduced to $\mu_4 - \sigma_0^4$.

The following assumptions are needed to ensure the asymptotic property of the parameter function estimator. They are similar to ones used in Müller & Stadtmüller (2005).

Assumption 5 We have

$$\sum_{r_1,r_2,r_3,r_4=0}^{p_n} E\left(\varepsilon_{r_1}\varepsilon_{r_2}\varepsilon_{r_3}\varepsilon_{r_4}\right)\nu_{r_1r_2}\nu_{r_3r_4} = o(n/p_n^2),$$

where the $\nu_{kl},\ k,l = 1,\ldots,p_n$, are the elements of $\Gamma_{p_n}^{-1}$.

Assumption 6 We assume that

$$\sum_{r_1,\ldots,r_8=0}^{p_n} E\left(\varepsilon_{r_1}\varepsilon_{r_3}\varepsilon_{r_5}\varepsilon_{r_7}\right) E\left(\varepsilon_{r_2}\varepsilon_{r_4}\varepsilon_{r_6}\varepsilon_{r_8}\right)\nu_{r_1r_2}\nu_{r_3r_4}\nu_{r_5r_6}\nu_{r_7r_8} = o(n^2p_n^2).$$

The asymptotic normality of the parameter function estimator is given in the following theorem

Theorem 4.3. *Under **Assumptions 1-6**, we have*

$$\frac{n\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right)' \Gamma_{p_n}\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right) - p_n}{\sqrt{2p_n}} \to \mathcal{N}(0, \sigma_0^4).$$

Moreover, if

$$\sum_{j>p_n} E\left(\varepsilon_j^2\right)\left(\int_{\mathcal{T}} \theta^*(t)\varphi_j(t)dt\right)^2 = o(\sqrt{p_n}/n), \tag{4.13}$$

where here $\{\varphi_j,\ j = 1,2,\ldots\}$ is the eigenbasis associated to the variance-covariance operator Γ, we have

$$\frac{nd^2\left(\hat{\theta}_n(\cdot), \theta^*(\cdot)\right) - p_n}{\sqrt{2p_n}} \to \mathcal{N}(0, \sigma_0^4), \tag{4.14}$$

where $d^2(\cdot,\cdot)$ denotes the metric defined in $L^2(\mathcal{T})$ through operator Γ, and defined by

$$d^2(f,g) = \int_{\mathcal{T}}\int_{\mathcal{T}} (f(t) - g(t))\, E(X(t)X(s))\, (f(s) - g(s))\, dtds,$$

for all $f, g \in L^2(\mathcal{T})$.

Now that we have checked the theoretical behavior of the estimator, we study its practical features through numerical results. We investigate the numerical performance of the proposed methodology based on some simulations and an application to ozone concentrations.

4.4 Numerical experiments

In this section, we study the performance of the proposed model based on numerical results that highlight the importance of truncation of the functional covariate and the spatial nature of the data. We first describe the estimation procedure for the investigated model.

Recall that the truncation strategy requires an appropriate selection of orthonormal basis. This basis can be chosen to be a fixed orthonormal basis, such as the Fourier basis, or it can be constructed by estimating the eigenfunctions of the covariance kernel (4.7) and applying functional principal component analysis (FPCA) to the explanatory random functions X_i. We use the eigenfunctions obtained from the FPCA to construct the expansion basis in this numerical section. The eigenfunctions are those of the integral operator associated with the integral kernel defined by the variance-covariance function of X, which is estimated for each $t, v \in [0, 1]$ as follows:

$$\hat{K}(t, v) = \frac{1}{n-1} \sum_{i=1}^{n} X_i(t) X_i(v). \tag{4.15}$$

A key step is the choice of the number p of functions used in the truncation strategy; we consider the average squared error (ASE),

$$\text{ASE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \tag{4.16}$$

the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The choice of p using AIC is consistent in the setting of functional linear models, see Müller & Stadtmüller (2005) for more details. Notice that we use a pre-selected p based on the cumulative inertia. We focus on the selection of p from among those associated with cumulative inertia values lower than some threshold (here 95%).
As measure of accuracy of the parameter function, (see Escabias et al., 2007) the usual integrated mean square error,

$$\text{IMSE} = \int_0^1 \left(\theta^*(t) - \hat{\theta}_n(t) \right)^2 dt, \tag{4.17}$$

is considered to compare the three choice strategies for p, namely, ASE, AIC and BIC.

4.4.1 Monte Carlo simulations

The main objective of the Monte Carlo Simulation is to investigate the finite sample behavior of the QMLEs of $\hat{\theta}_n(.)$, $\hat{\lambda}_n$ and $\hat{\sigma}_n^2$. We consider two spatial scenarios (see Su, 2004) in a data grid with 60×60 locations, where we randomly allocate n spatial units.

- **Scenario 1**: The spatial weight matrix W_n is constructed by taking the k neighbors of each unit using kNN method (k nearest neighbors algorithm). Let us take $k = \{4, 8\}$.

- **Scenario 2**: We consider a number of districts r (block or group) with m members in each district, where the units of the same district have the same weight. As in Case (1993), we can define the spatial weight matrix as block diagonal $W_n = I_r \otimes B_m$, where $\otimes$ is the Kronecker product, $B_m = (l_m l_m^{'} - I_m)/(m-1)$, and l_m is an m vector of 1 .

The simulations are performed based on the following data:

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \int_{\mathcal{T}} X_i(t)\theta^*(t)dt + U_i \tag{4.18}$$

where $U_i \sim \mathcal{N}(0, \sigma_0^2)$.

We generate the functional covariate as in Müller & Stadtmüller (2005) using the Fourier orthonormal basis $\{\varphi_j(t) = \sqrt{2}\sin(j\pi t),\ t \in [0,1], j = 1, 2, ...\}$. Let us use the first twenty functions of this basis to generate the explanatory random function

$$X(t) = \sum_{j=1}^{20} \varepsilon_j \varphi_j(t), \tag{4.19}$$

where $\varepsilon_j \sim \mathcal{N}(0, 1/j)$ for $j \geq 1$. We define the parameter function as $\theta^*(t) = \sum_{j=1}^{20} \theta_j^* \varphi_j(t)$, with $\theta_j^* = 0$ for $j > 3$, $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)' = (1, 1/2, 1/3)'$. With this parameter function and $\sigma_0^2 = 1$, different samples are generated using different values of the autoregressive parameter $\lambda_0 = 0.2; 0.4; 0.6$; and 0.8.

We apply the truncation strategy to reduce the infinite dimensionality of our model $Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j^{(i)} + \sum_{j=p_n+1}^{\infty} \theta_j^* \varepsilon_j^{(i)} + U_i,\ i = 1, \ldots, n,\ \ n = 1, 2, \ldots$ to a p_n-finite linear approximation and compute the quasi-likelihood estimator. The parameters λ_0, σ_0^2 and $\theta_1^*, ..., \theta_{p_n}^*$ are estimated by solving the score equations defined in Section 4.3. Different sample sizes, $n = \{100, 200, 400\}$, are tested for the first scenario; for the second, we take $r = \{10, 20, 30\}$ and $m = \{5, 10, 15\}$, with sample size $n = m \times r$.
The studied models are replicated 200 times, and the results of Scenario 1 are presented in Tables 4.1 and 4.2, respectively, for $k = 4$ and $k = 8$. For Scenario 2, the results are reported in Tables 4.3 to 4.6. Each table represents a specific model. In each table, the rows λ, σ^2, IMSE and PCs give the averages over these replications (with the standard deviation in brackets) of the autoregressive parameter estimate $\hat{\lambda}_n$, the standard deviation parameter $\hat{\sigma}_n^2$, the associated IMSE defined in (4.17) and the number p of eigenfunctions (used in the truncation), respectively. For the different models, the strategies used to select p yield (on average) values close to the true parameter of $p = 3$, especially for ASE and AIC and large sample sizes (see the columns titled PCs in Tables 4.1-4.6). The parameter function estimates are in given in Figures 4.2-4.3.
For all the models, the three methods used to select p and two spatial scenarios, the performance of the parameter function and the variance estimates varies with the sample size.

A larger IMSE (the smallest is in bold) of order 0.2 is noted for sample size $n = 100$ and $k = 8$.
The methods using the ASE and AIC criteria outperform the other methods. The spatial structure, namely, the number of neighbors k (Scenario 1) and the number of observations m in each district (Scenario 2), has a slight impact on the performance of the spatial parameter estimator $\hat{\lambda}_n$. Better results are obtained for lower values, namely, $k = 4$ and $m = 5$, since the weights are more important in these cases. For a fixed value of k or m, the performance varies with sample size.

Table 4.1: Estimation of parameters with $n = \{100, 200, 400\}$, $k = 4$

		n = 100			n = 200			n = 400		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
$\lambda_0 = 0.2$	λ	.1783	.1786	.1800	.2045	.2046	.2043	.1955	.1955	.1956
		(.1150)	(.1160)	(.1144)	(.0727)	(.0727)	(.0731)	(.0493)	(.0494)	(.0495)
	σ^2	.9669	.9732	.9878	.9835	.9858	.9913	.9829	.9834	.9870
		(.1438)	(.1465)	(.1511)	(.1036)	(.1040)	(.1055)	.0710	(.0711)	(.0710)
	IMSE	**.1584**	.1996	.2595	**.0796**	.1141	.1489	**.0337**	.0478	.0860
		(.1499)	(.1332)	(.1339)	(.0654)	(.0709)	(.0747)	(.0325)	(.0476)	(.0564)
	PCs	2.920	2.170	1.715	2.965	2.445	2.115	2.990	2.785	2.415
		(.2720)	(.6349)	(.3637)	(.1842)	(.5463)	(.5226)	(.0997)	(.4119)	(.5139)
$\lambda_0 = 0.4$	λ	.3952	.3969	.3979	.3992	.3996	.3997	.3945	.3947	.3947
		(.0987)	(.1428)	(.0997)	(.0581)	(.0984)	(.0580)	(.0449)	(.0447)	(.0448)
	σ^2	.9573	.9609	.9786	.9786	.9798	.9865	.9885	.9888	.9929
		(.1432)	(.1428)	(.1503)	(.0983)	(.0984)	(.1002)	(.0723)	(.0725)	(.0448)
	IMSE	**.1778**	.2063	.2786	**.0880**	.1067	.1536	**.0399**	.0507	.0977
		(.1680)	(.1574)	(.1528)	(.0830)	(.0794)	(.0908)	(.0365)	(.0464)	(.0629)
	PCs	2.850	2.285	1.720	2.865	2.520	2.125	2.950	2.790	2.360
		(.3850	(.6753)	(.6662)	(.3426)	(.5108)	(.5926)	(.2185)	(.4083)	(.5309)
$\lambda_0 = 0.6$	λ	.5859	.5877	.5884	.5975	.5990	.5988	.5979	.5984	.5985
		(.0725)	(.0722)	(.0731)	(.0452)	(.0458)	(.0455)	(.0365)	(.0366)	(.0368)
	σ^2	.9623	.9605	.9773	.9872	.9835	.9916	.9981	.9970	1.0009
		(.1357)	(.1335)	(.1387)	(.0965)	(.0947)	(.0964)	(.0743)	(.0741)	(.0744)
	IMSE	**.1568**	.1770	.2428	**.1080**	.1092	.1642	.0508	**.0506**	.0912
		(.1248)	(.1191)	(.1272)	(.0844)	(.0747)	(.0942)	(.0497)	(.0462)	(.0527)
	PCs	2.680	2.275	1.710	2.680	2.525	2.070	2.845	2.800	2.410
		(.6160)	(.6175)	(.6387)	(.5560)	(.5393)	(.5889)	(.3764)	(.4010)	(.5032)
$\lambda_0 = 0.8$	λ	.7863	.7889	.7884	.7929	.7940	.7938	.7990	.7997	.7998
		(.0468)	(.0461)	(.0470)	(.0312)	(.0312)	(.0313)	(.0192)	(.0190)	(.0191)
	σ^2	.9814	.9632	.9788	.9978	.9875	.9953	.9986	.9892	.9927
		(.1519)	(.1482)	(.1536)	(.0971)	(.0952)	(.0966)	(.0741)	(.0689)	(.0696)
	IMSE	.2303	**.1976**	.2422	.1326	**.1085**	.1624	.0932	**.0520**	.0898
		(.1469)	(.1329)	(.1281)	(.1177)	(.0809)	(.0937)	(.1007)	(.0468)	(.0520)
	PCs	2.295	2.330	1.845	2.465	2.470	2.035	2.535	2.765	2.390
		(.8007)	(.6428)	(.6581)	(.7151)	(.539)	(.5525)	(.6488)	(.4251)	(.4991)

4.4.2 Real data application

The goal is to forecast ground-level ozone concentrations using observations from stations within the Southeastern and Southwestern of United States over a span of 48 hours in the summer of 2015. The data are collected from monitoring stations (agencies) across the United States and are available at `https://www.epa.gov/outdoor-air-quality-data`. We are given the ozone concentration for 106 stations for every hour from 12am July 19 to 11pm July 20, 2015 (that is, 48 hours). We use linear interpolation to estimate the missing values.

We organize the original space-time series into a set of daily functional data to apply the functional methodology.

Let us consider at each station a response variable Y as the ozone concentration at 12pm on July 20 and a covariate function $\{X(t), t \in [0, 23]\}$ corresponding to the 24 records of ozone concentrations from 12pm on July 19 to 11am July 20. Figure 4.4 presents the geographical positions of the 106 stations (red points) and the curves of the ozone concentration from 12pm July 19 to 11am July 20.

To highlight the performance of the spatial FSARM model, we compare with the usual functional linear model (FLM), that does not take into account any spatial structure in the estimation procedure.

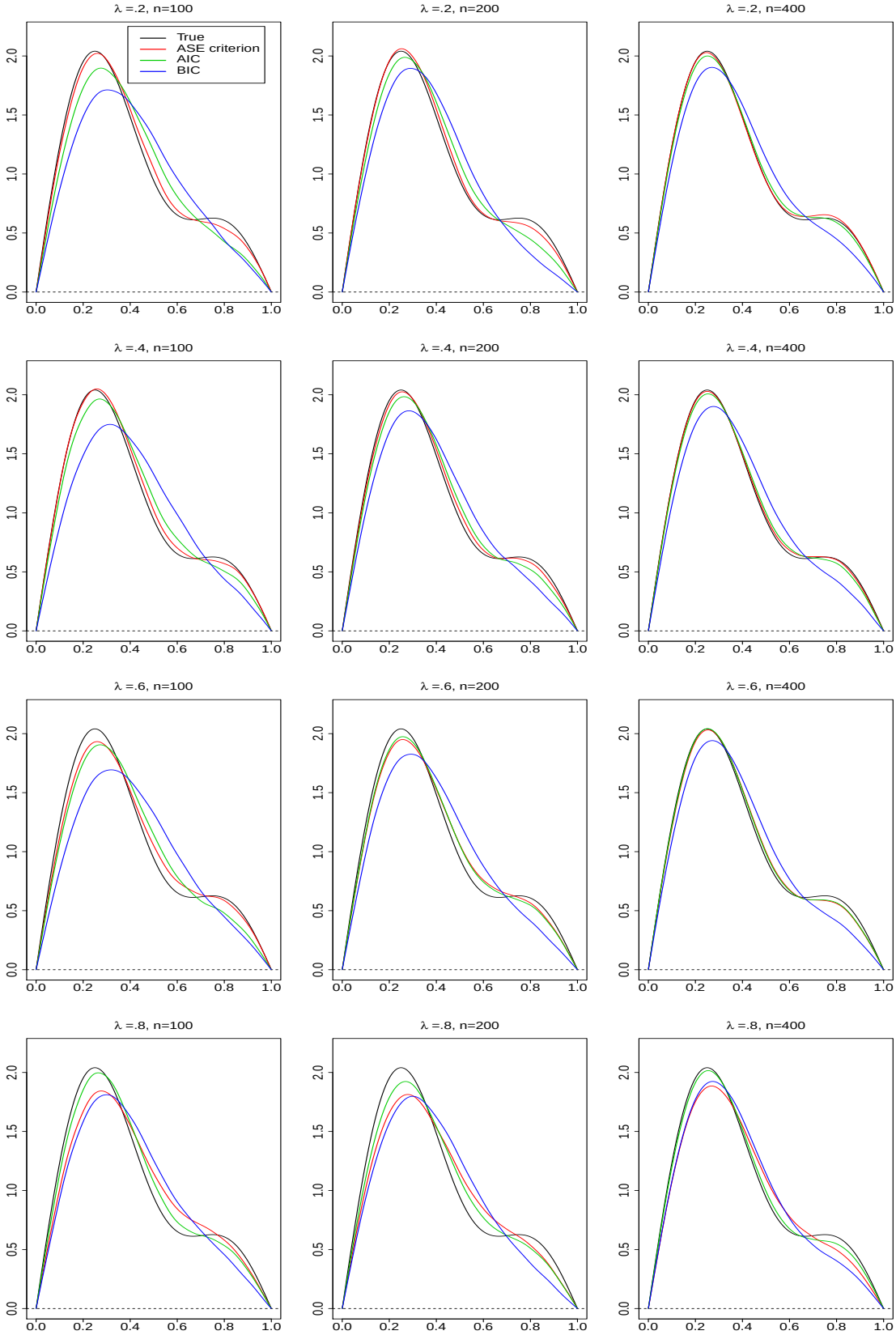

Figure 4.1: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria and $k = 4$.

Table 4.2: Estimation of parameters with $n = \{100, 200, 400\}$, $k = 8$

		n=100			n=200			n=400		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
$\lambda_0 = 0.2$	λ	.1711	.1709	.1690	.1876	.1875	.1886	.1912	.1912	.1916
		(.1604)	(.1614)	(.1439)	(.1031)	(.1037)	(.1036)	(.0800)	(.0799)	(.0801)
	σ^2	.9656	.9706	.9892	.9781	.9797	.9860	.9833	.9839	.9871
		(.1364)	(.1385)	(.1439)	(.0995)	(.1000)	(.1010)	(.0687)	(.0688)	(.0690)
	IMSE	**.1612**	.1920	.2480	**.0866**	.1116	.1517	**.0394**	.0548	.0881
		(.1731)	(.1693)	(.1560)	(.0795)	(.0840)	(.0955)	(.0409)	(.0484)	(.0476)
	PCs	2.925	2.275	1.705	2.950	2.540	2.190	2.980	2.725	2.395
		(.2641)	(.6256)	(.1496)	(.2185)	(.5290)	(.5964)	(.1404)	(.4476)	(.4901)
$\lambda_0 = 0.4$	λ	.3803	.3809	.3811	.3859	.3861	.3859	.3881	.3880	.3877
		(.1416)	(.1416)	(.1413)	(.0822)	(.0822)	(.0834)	(.0705)	(.0727)	(.0710)
	σ^2	.9593	.9638	.9782	.9787	.9800	.9871	.9945	.0727	.9985
		(.1438)	(.1456)	(.1501)	(.1019)	(.1024)	(.1048)	(.0725)	(.0518)	(.0724)
	IMSE	**.1541**	.1821	.2359	**.0828**	.1066	.1490	**.0426**	.0518	.0895
		(.1111)	(.1114)	(.1274)	(.0718)	(.0801)	(.0863)	(.0389)	(.0457)	(.0556)
	PCs	2.855	2.180	1.730	2.925	2.555	2.165	2.940	2.800	2.445
		(.3669)	(.6632)	(.6237)	(.2641)	(.5554)	(.1240)	(.2381)	(.4010)	(.5180)
$\lambda_0 = 0.6$	λ	.5758	.5791	.5794	.5924	.5933	.5935	.5940	.5947	.5944
		(.1061)	(.1060)	(.1072)	(.0671)	(.0672)	(.0675)	(.0496)	(.0495)	(.0494)
	σ^2	.9719	.9680	.9844	.9792	.9790	.9868	.9932	.9921	.9950
		(.1419)	(.1398)	(.1072)	(.0982)	(.0994)	(.1020)	(.0757)	(.0749)	(.0494)
	IMSE	.2024	**.2024**	.2628	**.0939**	.1026	.1540	**.0477**	.0485	.9950
		(.1581)	(.1421)	(.1414)	(.0868)	(.0739)	(.0864)	(.0476)	(.0463)	(.0755)
	PCs	2.600	2.290	1.760	2.780	2.530	2.110	2.855	2.795	2.465
		(.6497)	(.6542)	(.6743)	(.4612)	(.5201)	(.0864)	(.3669)	(.4047)	(.5000)
$\lambda_0 = 0.8$	λ	.7741	.7777	.7771	.7890	.7909	.7905	.7941	.7950	.7950
		(.0630)	(.0630)	(.0633)	(.0410)	(.0411)	(.0412)	(.0321)	(.0318)	(.0321)
	σ^2	.9852	.9686	.9840	1.0069	.9984	1.0071	.9957	.9889	.9925
		(.1439)	(.1374)	(.1403)	(.1037)	(.1022)	(.1044)	(.0745)	(.0720)	(.0536)
	IMSE	.2076	**.1989**	.2516	.1199	**.1027**	.1609	.0811	**.0492**	.0880
		(.1378)	(.1277)	(.1403)	(.1040)	(.0695)	(.0886)	(.0970)	(.0476)	(.0536)
	PCs	2.245	2.200	1.720	2.545	2.505	2.035	2.615	2.775	2.405
		(.7798)	(.6725)	(.6509)	(.6858)	(.5398)	(.5616)	(.6315)	(.4186)	(.5022)

The observations $(\{X_i(t), t \in [0, 23]\}, Y_i), i = 1, \ldots, 106$, are then used to estimate, on one hand, the parameter function and hypothetical intercept using the FLM methodology and, on the other hand, the parameter function and the autoregressive parameter using the FSARM methodology developed here. Even though the variance is estimated by the two methods, we do present it here but focus on the covariate and autoregressive parameters. We describe the spatial dependence between the stations using a 106×106 spatial weight matrix W_n. We follow the idea of Pinkse & Slade (1998) to define the elements of W_n by:

$$w_{ij} = \begin{cases} \dfrac{1}{1 + d_{ij}} & \text{if } d_{ij} < \rho \\ 0 & \text{otherwise,} \end{cases}$$

where d_{ij} is the euclidian distance between station i and station j, and ρ is some cut-off distance chosen such that each station has at least four neighbors. Other weight matrices have been tested, but we choose to present the results corresponding to this matrix.
Note that FPCA is used to smooth the curves before we reduce the spatial dimensions of the functional covariate using the eigenbasis, as explained above (see Figure 4.5). The AIC is used to select the number of eigenfunctions. For the two models, we have the same optimal number of eigenfunctions $p = 3$. Table 4.7 and Figure 4.6 give the estimation results of the FLM and FSARM. Note that the curves obtained by the two estimation

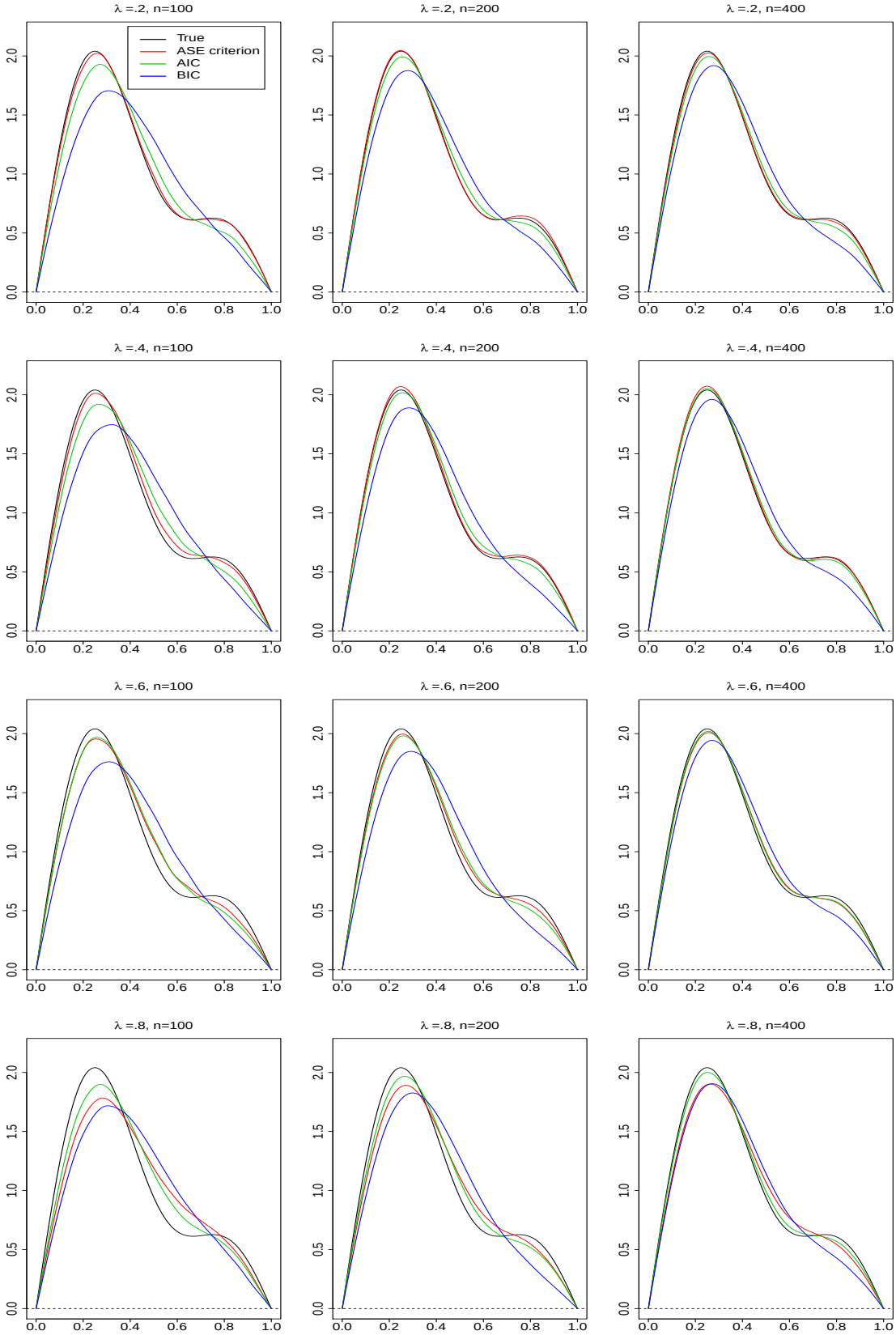

Figure 4.2: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria and $k = 8$.

Table 4.3: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.2$.

		m=5			m=10			m=15		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
r = 10	λ	.1457	.1474	.1483	.1828	.1842	.1838	.1734	.1734	.1746
		(.1687)	(.1700)	(.1685)	(.1466)	(.1476)	(.1468)	(.1476)	(.1479)	(1483)
	σ^2	.9090	.9209	.9463	.9583	.9627	.9759	.9810	.9836	.9947
		(.1897)	(.1941)	(.2047)	(.1340)	(.1353)	(.1382)	(.1076)	(.1086)	(.1108)
	IMSE	**.3347**	.3603	.3778	**.1655**	.1925	.2412	**.1109**	.1430	.1973
		(.2848)	(.2541)	(.2267)	(.1515)	(.1328)	(.1251)	(.1076)	(.1103)	(.1115)
	PCs	2.900	1.94	1.505	2.930	2.275	1.860	2.945	2.425	1.940
		(.3170)	(.7611)	(.6497)	(.2747)	(.6010)	(.6577)	(.2286)	(.5883)	(.6232)
r = 20	λ	.1794	.1796	.1788	.1850	.1851	.1853	.1917	.1919	.1914
		(.0934)	(.0938)	(.0940)	(.1079)	(.1079)	(.1073)	(.1027)	(.1023)	(.1026)
	σ^2	.9413	.9450	.9602	.9748	.9768	.9841	.9832	.9840	.9892
		(.1429)	(.1436)	(.1498)	(.1014)	(.1018)	(.1045)	(.0809)	(.1023)	(.0823)
	IMSE	**.1767**	.2133	.2686	**.0725**	.1032	.1507	**.0528**	.0709	.1164
		(.1676)	(.1620)	(.1614)	(.0666)	(.0693)	(.0874)	(.0456)	(.0561)	(.0612)
	PCs	2.920	2.285	1.805	2.970	2.545	2.140	2.9850	2.690	2.280
		(.2720)	(.6900)	(.7138)	(.1710)	(.5092)	(.5585)	(.1219)	(4637)	(.5225)
r = 30	λ	.1990	.1985	.1988	.1942	.1941	.1943	.1890	.1890	.1832
		(.0853)	(.0860)	(.0869)	(.0762)	(.0816)	(.0762)	(.0867)	(.0866)	(.0867)
	σ^2	.9668	.9692	.9797	.9927	.9938	.9986	.9900	.9904	.9930
		(.1152)	(.1156)	(.0869)	(.0816)	(.0816)	(.0829)	(.0639)	(.0638)	(.0643)
	IMSE	**.1112**	.1446	.1991	**.0555**	.0755	.1143	**.0330**	.0452	.0755
		(.1047)	(.1088)	(.1130)	(.0615)	(.0651)	(.0680)	(.0298)	(.0452)	(.0643)
	PCs	2.920	2.455	1.990	2.980	2.6500	2.2750	2.9900	2.8100	2.5150
		(.2720)	(.5653)	(.6340)	(.1404)	(.4782)	(.5299)	(.0997)	(.3933)	(.5010)

Table 4.4: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.4$.

		m = 5			m = 10			m = 15		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
r = 10	λ	.3590	.3613	.3619	.3746	.3756	.3751	.3739	.3751	.3748
		(.1106)	(.1134)	(.1130)	(.1184)	(.1190)	(.1186)	(.1107)	(.1110)	(.1117)
	σ^2	.9175	.9271	.9487	.9642	.9682	.9845	.9862	.9890	.9999
		(.1891)	(.1906)	(.1943)	(.1375)	(.1399)	(.1412)	(.1245)	(.1252)	(.1276)
	IMSE	**.3812**	.4078	.4122	**.1702**	.2024	.2702	**.1057**	.1387	.1976
		(.3904)	(.3665)	(.3247)	(.1452)	(.1443)	(.1459)	(.0883)	(.0856)	(.1102)
	PCs	2.7300	1.9150	1.5150	2.8100	2.2200	1.7000	2.8950	2.3950	1.9300
		(.5464)	(.7816)	(.3263)	(.4414)	(.6811)	(.6650)	(.3073)	(.5750)	(.6140)
r = 20	λ	.3873	3.883	.3887	.3733	.3737	.3736	.3829	.3830	.3829
		(.0749)	(.0751)	(.0749)	(.0857)	(.0861)	(.0864)	(.0777)	(.0777	(.0775)
	σ^2	.9587	.9618	.9769	.9875	.9890	.9966	.9914	.9921	.9967
		(.1353)	(.1341)	(.1402)	(.1043)	(.1047)	(.1070)	(.0868)	(.0870)	(.0881)
	IMSE	**.1700**	.1980	.2573	**.0853**	.1070	.1563	**.0570**	.0734	.1157
		(.1368)	(.1240)	(.1271)	(.0681)	(.0696)	(.0838)	(.0455)	(.0539)	(.0699)
	PCs	2.780	2.275	1.795	2.905	2.530	2.115	2.90	2.670	2.300
		(.4825)	(.6414)	(.1271)	(.1277)	(.5296)	(.5861)	(.3008)	(.4714)	(.5582)
r = 30	λ	.3943	3952	.3950	.3867	.3867	.3871	.3910	.3911	.3912
		(.0670)	(.0671)	(.0676)	(.0675)	(.0675)	(.0677)	(.0647)	(.0649)	(.0654)
	σ^2	.9706	.9718	.9832	.9857	.9863	.9913	.9870	.9873	.9906
		(.1178)	(.1176)	(.1228)	(.0840)	(.0843)	(.0854)	(.0674)	(.0676)	(.0680)
	IMSE	**.1150**	.1343	.1951	**.0577**	.0722	.1122	**.0374**	.0461	.0830
		(.0903)	(.0861)	(.1100)	(.0604)	(.0687)	(.0652)	(.0343)	(.0470)	(.0512)
	PCs	2.810	2.395	0.915	2.895	2.690	2.290	2.960	2.830	2.465
		(.4181)	(.5750)	(.6162)	(.3073)	(.4848)	(.5169)	(.1965)	(.3897)	(.5100)

Table 4.5: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.6$.

		m = 5			m = 10			m = 15		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
r = 10	λ	.5867	.5895	.5903	.5815	.5843	.5849	.5736	.5746	.5747
		(.0746)	(.0752)	(.0744)	(.0838)	(.0840)	(.0835)	(.0998)	(.1002)	(.1009)
	σ^2	.9536	.9524	.9746	.9617	.9573	.9718	.9752	.9736	.9823
		(.2158)	(.2105)	(.2150)	(.1464)	(.1463)	(.1522)	(.1121)	(.1100)	(.1115)
	IMSE	**.3911**	.4201	.4261	**.1919**	.2053	.2598	**.1354**	.1441	.1922
		(.3470)	(.3498)	(.3227)	(.1480)	(.1489)	(.1418)	(.1075)	(.0988)	(.1142)
	PCs	2.454	2.025	1.640	2.5750	2.2750	1.8150	2.6700	2.3700	1.9900
		(.6558)	(.7598)	(.6948)	(.6375)	(.6256)	(.6656)	(.5501)	(.5698)	(.6179)
r = 20	λ	.5875	.5899	.5899	.5865	.5884	.5887	.5851	.5860	.5860
		(.0491)	(.0493)	(.0493)	(.0571)	(.0575)	(.0574)	(.0574)	(.0580)	(.0582)
	σ^2	.9666	.9580	.9732	.9838	.9784	.9866	.9810	.9785	.9829
		(.1403)	(.1323)	(.1385)	(.1053)	(.1005)	(.1019)	(.0791)	(.0772)	(.0582)
	IMSE	.2148	**.2138**	.2629	.1129	**.1074**	.1615	.0723	**.0685**	.1062
		(.1745)	(.1755)	(.1652)	(.0932)	(.0790)	(.0893)	(.0677)	(.0517)	(.0605)
	PCs	2.495	2.300	1.800	2.640	2.5500	2.0950	2.7450	2.6800	2.3300
		(.6873)	(.6650)	(.6725)	(.5934)	(.5375)	(.5724)	(.4911)	(.4676)	(.5220)
r = 30	λ	.5948	.5964	.5958	.5879	.5885	.5883	.5886	.5888	.5888
		(.0425)	(.0421)	(.0428)	(.0443)	(.0445)	(.0444)	(.0479)	(.0479)	(.0481)
	σ^2	.9846	.9798	.9899	.9965	.9953	1.0009	.9964	.9956	.9994
		(.1100)	(.1077)	(.1102)	(.0803)	(.0806)	(.0822)	(.0682)	(.0683)	(.0481)
	IMSE	**.1293**	.1404	.1920	**.0684**	.0689	.1184	.0439	**.0402**	.0814
		(.0988	(.0910)	(.1035)	(.0592)	(.0555)	(.0798)	(.0538)	(.0418)	(.0527)
	PCs	2.630	2.420	1.995	2.8050	2.7150	2.2900	2.8900	2.8850	2.4800
		(.5698)	(.5790)	(.5802)	(.3972)	(.4525)	(.5723)	(.3442)	(.3198)	(.5009)

Table 4.6: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.8$.

		m = 5			m = 10			m = 15		
		ASE	AIC	BIC	ASE	AIC	BIC	ASE	AIC	BIC
r = 10	λ	.7883	.7905	.7900	.7921	.7941	.7941	.7834	.7857	.7856
		(.0474)	(.0468)	(.0474)	(.0407)	(.0404)	(.0405)	(.0432)	(.0430)	(.0430)
	σ^2	.9461	.9349	.9596	.9682	.9549	.9703	.9917	.9782	.9883
		(.2330)	(.2326)	(.2436)	(.1353)	(.1324)	(.1379)	(.1132)	(.1073)	(.1098)
	IMSE	**.3333**	.3607	.3814	.1946	**.1890**	.2367	.1635	**.1405**	.1928
		(.2556)	(.2545)	(.2152)	(.1303)	(.1239)	(.1224)	(.1248)	(.1028)	(.1132)
	PCs	2.265	1.950	1.515	2.340	2.275	1.785	2.415	2.420	1.975
		(.7860)	(.7749)	(.6723)	(.7464)	(.6335)	(.6088)	(.7454)	(.5703)	(.6215)
r = 20	λ	.7955	.7968	.7968	.7943	.7959	.7960	.7945	.7956	.7957
		(.0297)	(.0292)	(.0296)	(.0307)	(.0302)	(.0302)	(.0285)	(.0281)	(.0280)
	σ^2	.9782	.9713	.9871	.9957	.9821	.9887	.9951	.9823	.9872
		(.1512)	(.1527)	(.1575)	(.1096)	(.1025)	(.1055)	(.0890)	(.0835)	(.0848)
	IMSE	.1890	**.1883**	.2532	.1340	**.1006**	.1449	.1120	**.0737**	.1164
		(.1541)	(.1390)	(.1445)	(.1104)	(.0645)	(.0802)	(.1114)	(.0628)	(.0676)
	PCs	2.470	2.250	1.735	2.430	2.570 0	2.200	2.515	2.715	2.300
		(.7153)	(.6706)	(.6534)	(.7265)	(.5162)	(.5931)	(.7158)	(.4525)	(.5399)
r = 30	λ	.7938	.7947	.7946	.7948	.7957	.7957	.7951	.7959	.7959
		(.0240)	(.0238)	(.0240)	(.0214)	(.0211)	(.0212)	(.0223)	(.0224)	(.0224)
	σ^2	.9946	.9838	.9949	1.0017	.9905	.9954	1.0027	.9932	.9965
		(.1199)	(.1149)	(.1201)	(.0873)	(.0854)	(.0866)	(.0731)	(.0700)	(.0707)
	IMSE	.1572	**.1310**	.1909	.0962	**.0638**	.1074	.0871	**.0489**	.0869
		(.1366)	(.1056)	(.1207)	(.0982)	(.0532)	(.0630)	(.0923)	(.0442)	(.0481)
	PCs	24450	2.4400	1.9550	2.5500	2.7600	2.3450	2.5650	2.8100	2.4450
		(.7414)	(.5815)	(.5956)	(.6555)	(.4397)	(.5454)	(.6307)	(.3933)	(.4982)

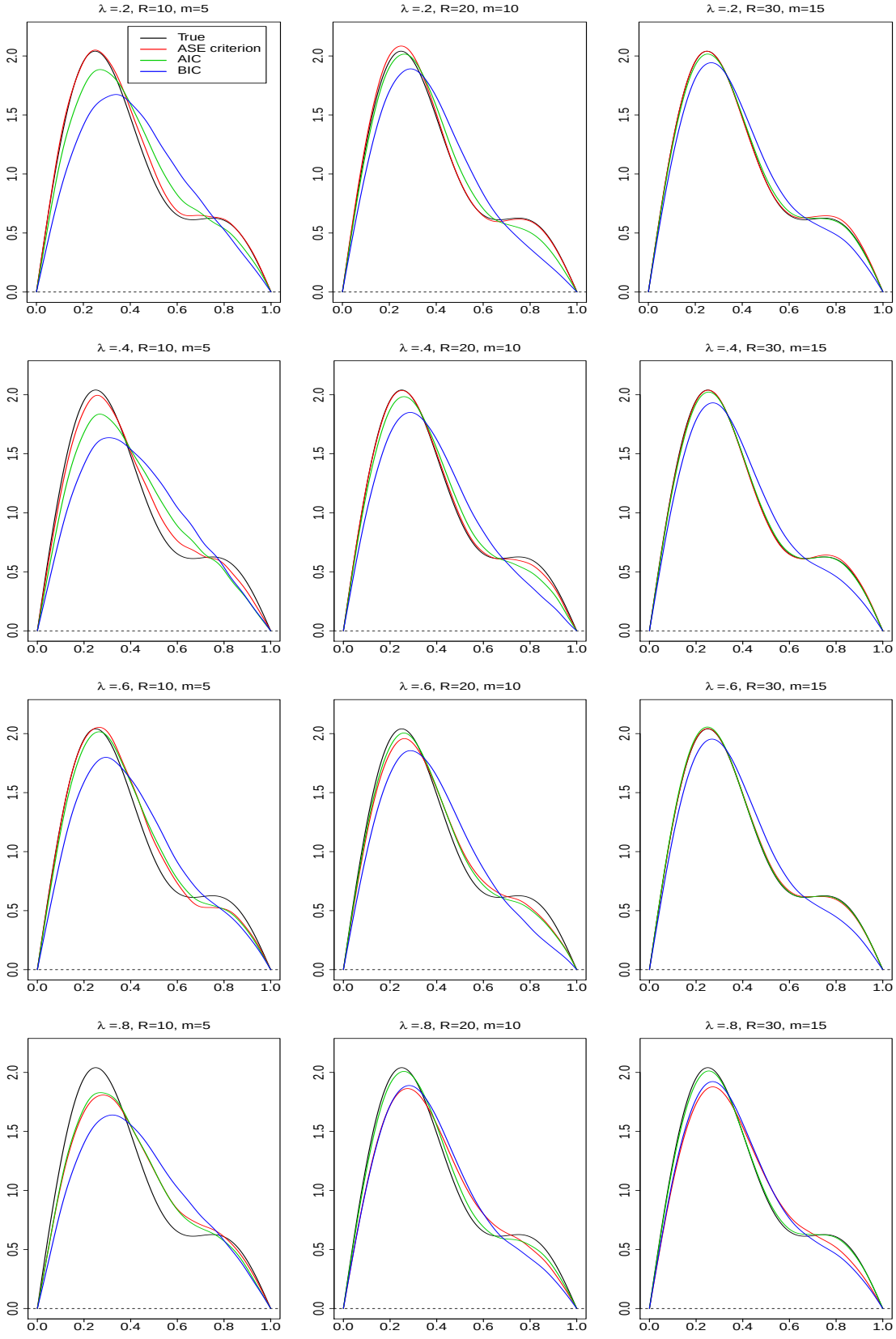

Figure 4.3: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria in Scenario 2 for different values of r and m.

methods are similar, with small differences around 12pm and 7pm. The FLM gives an intercept estimate close to zero, while with FSARM, we have a spatial structure with an estimated autoregressive parameter close to 0.2.
Now, let us consider the following problem of prediction. At a given station s_0, we aim to predict the ozone concentration every hour, from 12am to 11pm, on July 20, 2015. For this aim, assume that at s_0, we observe only the 24 records of ozone concentration from 12am to 11pm on July 19, 2015 and we would like to predict the ozone concentration of the following day, that is, from 12am to 11pm on July 20, 2015. To obtain these predictions, we proceed as follows.

1. For the prediction at 12am July 20, 2015, we estimate the parameters of FLM or FSARM where the 105 observations (X_i, Y_i) are: $\{X_i(t), t \in \{0, \dots, 23\}\}$, the ozone concentrations from 12am to 11pm on July 19, and Y_i is the ozone concentration at 12am, July 20, at station i. The obtained estimated model is used to predict the ozone concentration at 12am July 20 at station s_0 (not contained in the sample), using the covariate $\{X_{s_0}(t), t \in \{0, \dots, 23\}\}$ composed of the ozone concentrations from 12am to 11pm on July 20. Let $\hat{Y}_{s_0}^{(1)}$ denote this prediction.

2. For the prediction at 1am July 20, 2015, let $X_i(t), t \in \{0, \dots, 23\}$ be the ozone concentrations from 1am July 19 to 12pm July 20 and Y_i be the ozone concentration at 1am July 20, 2015 at station i. Use these observations to estimate the parameters of FLM or FSARM, and use them to predict the ozone concentration of station s_0 at 1am July 20 using $X_{s_0}(t), t \in \{0, \dots, 23\}$, where the first 23 records are the real ozone concentrations from 1am to 11pm July 20 and $X_{s_0}(23) = \hat{Y}_{s_0}^{(1)}$. Let $\hat{Y}_{s_0}^{(2)}$ denote the obtained prediction.

... Repeat the above steps to obtain predictions from 2am to 11pm, July 20, 2015.

We randomly select 4 stations among the 106 and apply the prediction procedure. Figure 4.7 presents the prediction results; the true values are in black, while the predictions are in red for the FSARM model and in blue for the FLM (with no spatial structure) model. FSARM achieves some improvements, particularly around 12pm, when the ozone concentration is higher.

Table 4.7: Estimated parameters for FLM and FSARLM.

	PCs	Autoregressive parameter	Intercept
FSARLM	3	0.19	
FLM	3		0.006

4.5 Conclusion

This work proposes a spatial functional linear regression function for functional random field covariates. Our main theoretical contribution was to study the consistency and asymptotic normality of the estimator. One can see the proposed methodology as an extension of the real-valued SAR model to functional data. More precisely, it is apparent that the proposed estimation approach based on a truncation technique is particularly well adapted to spatial regression estimation for functional data in the presence of spatial dependence. This good behavior is observed both from an asymptotic point of view and from a numerical study. This work offers interesting perspectives for investigation. Future work will be tied to generalized functional linear spatial models (see, for instance Kelejian

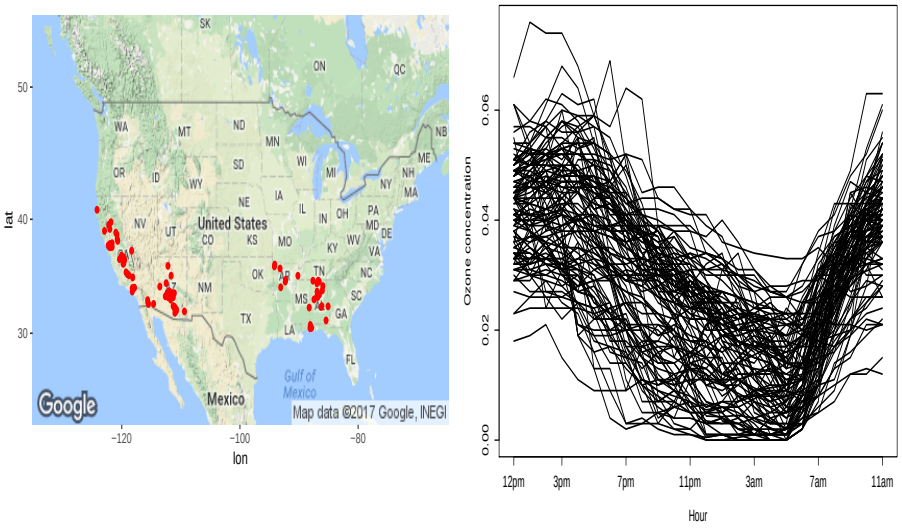

Figure 4.4: Locations and areas of the 106 stations (left panel) and corresponding ozone concentration curves from 12pm, July 19 to 11am, July 20 (right panel).

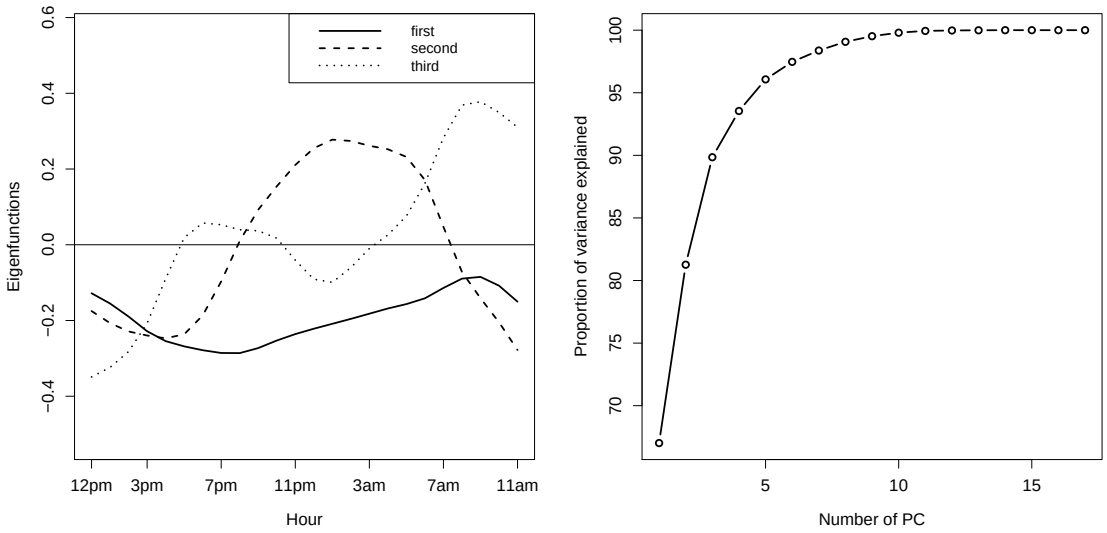

Figure 4.5: The three first eigenfunctions (left panel) and the proportion of explained variance (right panel).

& Prucha, 1998; Müller & Stadtmüller, 2005). Also, an adaptation of this method to issues using different covariates (functional and non-functional) with or without a spatial weight matrix with correlated errors could be developed. The application of the proposed regression estimator to additional real data, will be investigated.

4.6 Appendix

We start by showing the identifiability of the parameter λ_0 and the consistency of the estimator $\hat{\lambda}_n$ when the sequence h_n is bounded or not bounded. This is given in the

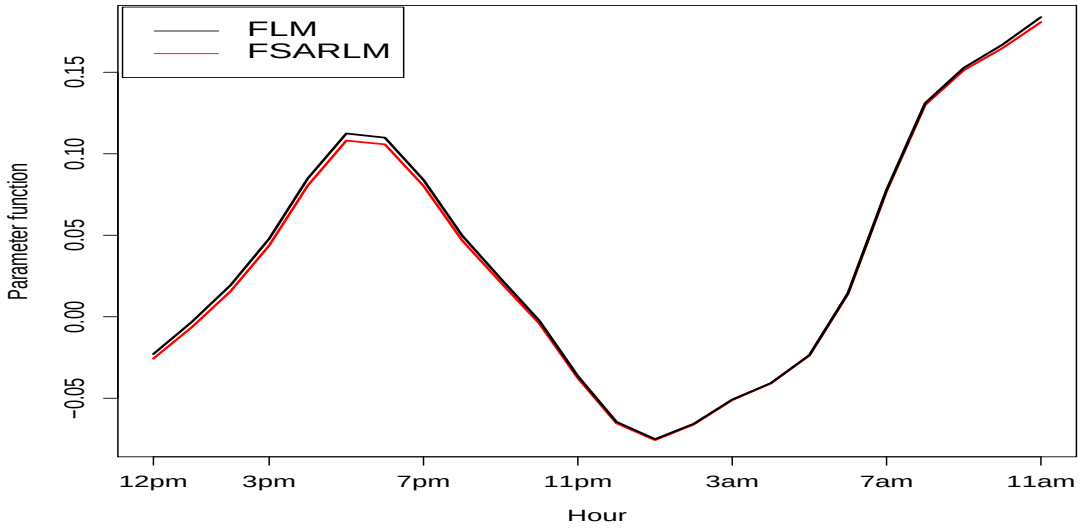

Figure 4.6: Estimated parameter functions.

following proposition

Proposition 4.1. *Assume* ***Assumptions 1-3****.*

(i) If the sequence $\{h_n\}$ *is bounded,* λ_0 *is identifiable and* $\hat{\lambda}_n$ *is consistent.*

(ii) If the sequence $\{h_n\}$ *is divergent,* λ_0 *is identifiable and* $\hat{\lambda}_n$ *is consistent.*

Proof of Proposition 4.1

Proof of (i). Let us first establish the identifiability. Proving identification of λ_0 is equivalent to showing that the concentrated likelihood function $Q_n(\lambda)$ is maximum at λ_0. This can be done by checking the following uniqueness condition:

$$\text{for any } \epsilon > 0 \qquad \limsup_{n\to\infty} \max_{\lambda\in\bar{N}_\epsilon(\lambda_0)} \frac{1}{n}\{Q_n(\lambda) - Q_n(\lambda_0)\} < 0$$

where $\bar{N}_\epsilon(\lambda_0)$ is the complement of an open neighbourhood of λ_0 in Λ with diameter ϵ.

Let us prove that $Q_{n,0}(\lambda) - Q_{n,0}(\lambda_0) \leq 0, \qquad$ for all $\lambda \in \Lambda$,

$$\text{where} \qquad Q_{n,0}(\lambda) = -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\sigma^2_{n,\lambda} + \ln|S_n(\lambda)|,$$

with

$$\sigma^2_{n,\lambda} = \frac{\sigma_0^2}{n}\text{tr}\left(A_n(\lambda)\right) = \sigma_0^2\left\{1 + 2(\lambda_0-\lambda)\frac{1}{n}\text{tr}(G_n) + (\lambda_0-\lambda)^2\frac{1}{n}\text{tr}(G_nG_n')\right\}.$$

Recall that the log-likelihood function of an SAR process without covariate ($\theta^*(t) = 0, \forall t \in \mathcal{T}$), $\mathbf{Y}_n = \lambda_0 W_n \mathbf{Y}_n + \mathbf{U}_n$, $\mathbf{V}_n \sim \mathcal{N}(0, \sigma_0^2 I_n)$ is

$$L_{n,0}(\lambda,\sigma^2) = \frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\mathbf{Y}_n'S_n'(\lambda)S_n(\lambda)Y_n.$$

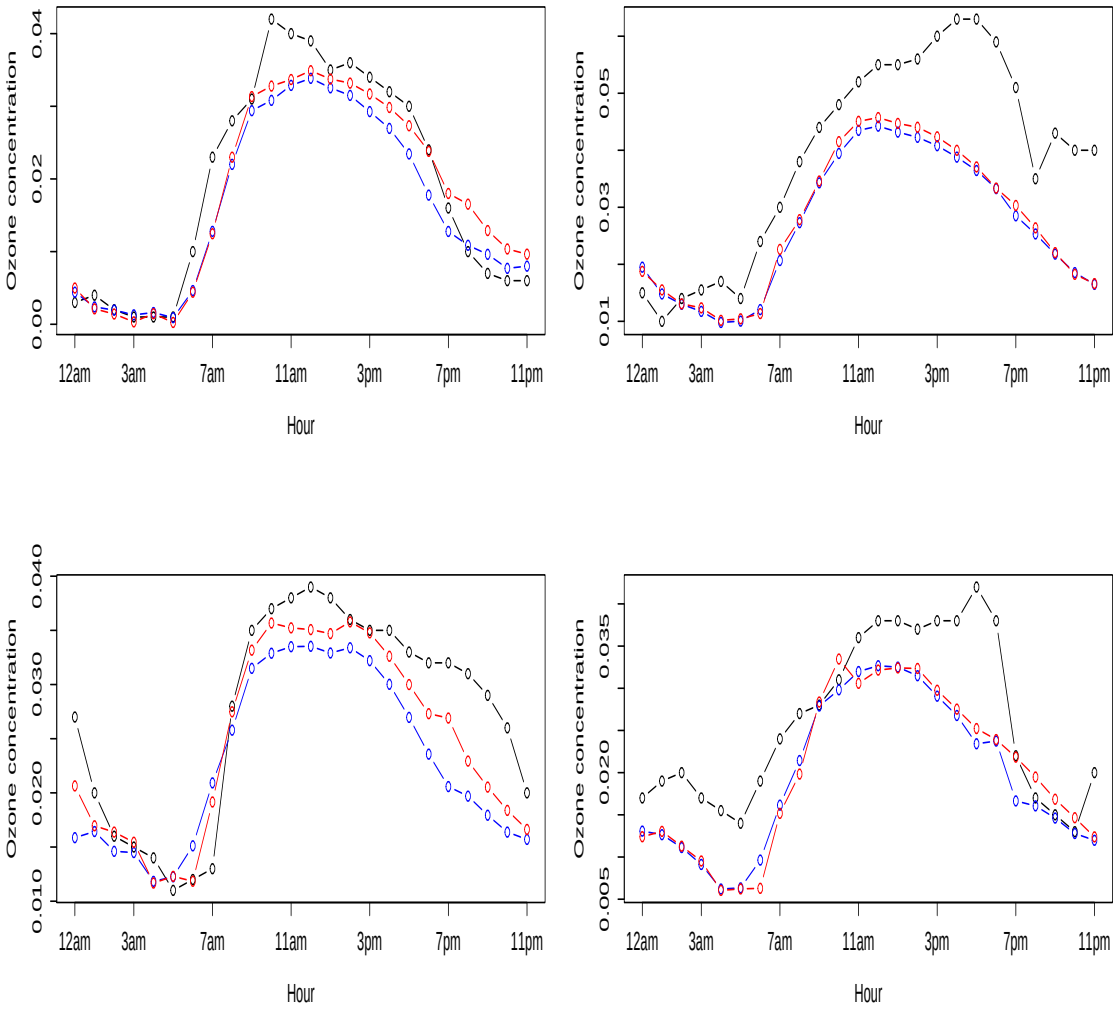

Figure 4.7: Ozone concentration (black curves) at four stations selected randomly from the 106 stations and their predictions obtained using the FSAR model (red curves) and FLM (blue curves).

It is easy to see that $Q_{n,0}(\lambda) = \max_{\sigma^2} E_0(L_{n,0}(\lambda, \sigma^2))$, where E_0 is the expectation under this SAR process. By Jensen's inequality, $Q_{n,0}(\lambda) \leq E_0(L_{n,0}(\lambda_0, \sigma_0^2)) = Q_{n,0}(\lambda_0)$ for all λ. This implies that

$$Q_{n,0}(\lambda) - Q_{n,0}(\lambda_0) \leq 0, \qquad \text{for all } \lambda \in \Lambda.$$

Let us prove that $\frac{1}{n}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = O(1)$, for λ_1 and λ_2 in Λ.

By the mean value theorem, $\frac{1}{n}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = \frac{1}{n}\text{tr}(W_n S_n^{-1}(\bar{\lambda}_n))(\lambda_2 - \lambda_1)$, where $\bar{\lambda}_n$ lies between λ_1 and λ_2. By the uniform boundedness of **Assumption 1**-iii, $\text{tr}(W_n S_n^{-1}(\bar{\lambda}_n)) = O(n/h_n)$. Thus, $\frac{1}{n}\ln|S_n(\lambda)|$ is uniformly equicontinuous in λ in Λ. As Λ is a bounded set, $\frac{1}{n}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = O(1)$ uniformly on λ_1 and λ_2.

Let us prove that $\sigma^2_{n,\lambda}$ is uniformly bounded away from zero on Λ.

Suppose that $\sigma^2_{n,\lambda}$ is not uniformly bounded away from zero on Λ. Then there would exist a sequence $\{\lambda_n\}$ in Λ such that $\lim_{n\to\infty}\sigma^2_{n,\lambda_n}=0$. Since we have $Q_{n,0}(\lambda)-Q_{n,0}(\lambda_0)\leq 0$ for all λ and $\frac{1}{n}(\ln|S_n(\lambda_0)|-\ln|S_n(\lambda)|)=O(1)$ uniformly on Λ, then $-\frac{1}{2}\ln\sigma^2_{n,\lambda}\leq -\frac{1}{2}\ln\sigma^2_0-\frac{1}{n}(\ln|S_n(\lambda_0)|-\ln|S_n(\lambda)|)=O(1)$. That is, $-\frac{1}{2}\ln\sigma^2_{n,\lambda}$ is bounded, and this is a contradiction with the previous statement. Therefore, $\sigma^2_{n,\lambda}$ must be bounded away from zero uniformly on Λ.

Let us prove the uniform equicontinuity of $Q_n(\lambda)$.

We have to show that $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous on Λ. The parameter $\sigma^{*2}_{n,\lambda}$ (see (4.11)) is uniformly bounded on Λ because it is a quadratic form of λ, and its components $\frac{1}{n}\Delta_n$, $\frac{1}{n}\mathrm{tr}(G_n)$ and $\frac{1}{n}\mathrm{tr}(G_nG_n')$ are bounded by **Assumption 1** (i-ii). The uniform continuity of $\ln\sigma^{*2}_{n,\lambda}$ on Λ then follows because $1/\sigma^{*2}_{n,\lambda}$ is uniformly bounded on Λ since $\sigma^{*2}_{n,\lambda}\geq\sigma^2_{n,\lambda}$ for all $\lambda\in\Lambda$ by **Assumption 3**. Hence, $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous.

Let us prove uniqueness of the maximum λ_0.
Remark that

$$\frac{1}{n}\left(Q_n(\lambda)-Q_n(\lambda_0)\right)$$
$$=\frac{1}{n}\left(Q_{n,0}(\lambda)-Q_{n,0}(\lambda_0)\right)-\frac{1}{2}\left(\ln\sigma^{*2}_{n,\lambda}-\ln\sigma^2_{n,\lambda}\right)+o(1).$$

Now, assume that the uniqueness does not hold. Then, there would exist $\epsilon>0$ and a sequence $\{\lambda_n\}$ in $\bar{N}_\epsilon(\lambda_0)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\left\{Q_n(\lambda_n)-Q_n(\lambda_0)\right\}=0.$$

Because $\bar{N}_\epsilon(\lambda_0)$ is a compact set, there exists a convergent subsequence λ_{n_m} of λ_n. Let λ_+ be the limit of this subsequence in Λ.
Now, as $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous in λ,

$$\lim_{n_m\to\infty}\frac{1}{n_m}\left\{Q_{n_m}(\lambda_+)-Q_{n_m}(\lambda_0)\right\}=0.$$

This is possible only if

$$\lim_{n_m\to\infty}\frac{1}{n_m}\left\{Q_{n_m,0}(\lambda_+)-Q_{n_m,0}(\lambda_0)\right\}=0 \text{ and } \lim_{n_m\to\infty}\sigma^{*2}_{n_m,\lambda_+}-\sigma^2_{n_m,\lambda_+}=0.$$

Since $Q_{n,0}(\lambda)-Q_{n,0}(\lambda_0)\leq 0$ and $-(\ln\sigma^{*2}_{n,\lambda}-\ln\sigma^2_{n,\lambda})\leq 0$ for all $\lambda\in\Lambda$, the fact that $\lim_{n_m\to\infty}\sigma^{*2}_{n_m,\lambda_+}-\sigma^2_{n_m,\lambda_+}=0$ is in contradiction with the above statement under **Assumption 3**(a). Under **Assumption 3**(b), the contradiction comes from $\lim_{n\to\infty}\frac{1}{n}\left\{Q_{n,0}(\lambda)-Q_{n,0}(\lambda_0)\right\}=0$. Indeed, under **Assumption 3**(b), we have

$$\lim_{n\to\infty}\left\{\frac{1}{n}\left(\ln|S_n(\lambda)|-\ln|S_n|\right)+\frac{1}{2}\left(\ln\sigma^2_{n,\lambda}-\ln\sigma^2_0\right)\right\}$$
$$=\lim_{n\to\infty}\frac{1}{n}\left\{Q_{n,0}(\lambda)-Q_{n,0}(\lambda_0)\right\}\neq 0 \qquad \text{for all } \lambda\neq\lambda_0.$$

Now to finish the proof of (i), it remains to show the convergence in probability of $\tilde{L}_n(\lambda)$ to $Q_n(\lambda)$ uniformly on λ in Λ.
Let us prove that

$$\sup_{\lambda\in\Lambda}\frac{1}{n}\left|\tilde{L}_n(\lambda)-Q_n(\lambda)\right|=o_p(1). \tag{4.20}$$

By definition, for each $\lambda \in \Lambda$

$$\frac{1}{n}\left(\tilde{L}_n(\lambda) - Q_n(\lambda)\right) = -\frac{1}{2}\left(\ln\hat{\sigma}^2_{n,\lambda} - \ln\sigma^{*2}_{n,\lambda}\right) + o(1).$$

We will show that, for all $\lambda \in \Lambda$

$$\hat{\sigma}^2_{n,\lambda} - \sigma^{*2}_{n,\lambda} = o_p(1). \tag{4.21}$$

Equation (4.21) combined with the fact that $\sigma^{*2}_{n,\lambda}$ is bounded away from zero uniformly on Λ implies that $\hat{\sigma}^2_{n,\lambda}$ is bounded away from zero uniformly on Λ in probability. Hence,

$$\ln\hat{\sigma}^2_{n,\lambda} - \ln\sigma^{*2}_{n,\lambda} = o_p(1), \qquad \text{uniformly on} \quad \Lambda.$$

Let us prove in the following that $\hat{\sigma}^2_{n,\lambda} - \sigma^{*2}_{n,\lambda} = o_p(1)$.
Let

$$\begin{aligned} M_n S_n(\lambda)\mathbf{Y}_n &= M_n S_n(\lambda) S_n^{-1}\left(\mathbf{X}_n(\theta^*(\cdot)) + \mathbf{U}_n\right)) \\ &= M_n \mathbf{R}_n(\theta^*(\cdot)) + (\lambda_0 - \lambda) M_n G_n \xi_{p_n}\theta^* + M_n S_n(\lambda) S_n^{-1}\mathbf{U}_n, \end{aligned}$$

where $\mathbf{R}_n(\theta^*(\cdot)) = B_n(\lambda)\mathbf{R}_n$.
Note that

$$\begin{aligned} \hat{\sigma}^2_{n,\lambda} - \sigma^{*2}_{n,\lambda} &= \frac{1}{n}\mathbf{Y}'_n S'_n(\lambda) M_n S_n(\lambda)\mathbf{Y}_n - \sigma^{*2}_{n,\lambda} \\ &= (\lambda_0 - \lambda)^2 H_{n0} + 2(\lambda_0 - \lambda) H_{1n}(\lambda) + H_{n2}(\lambda) - \sigma^2_{n,\lambda} \\ &\qquad + H_{n3}(\lambda) + H_{n4}(\lambda), \end{aligned} \tag{4.22}$$

where

$$\begin{aligned} H_{n0} &= \theta^{*\prime}\left\{\frac{\xi'_{p_n} G'_n G_n \xi_{p_n}}{n} - \operatorname{tr}\left(\frac{G'_n G_n}{n}\right)\Gamma_{p_n}\right\}\theta^* \\ &\qquad -\theta^{*\prime}\left\{\frac{\xi'_{p_n} G'_n \xi_{p_n}}{n}\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\frac{\xi'_{p_n} G_n \xi_{p_n}}{n} - \operatorname{tr}^2\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right\}\theta^*, \end{aligned}$$

and

$$\begin{aligned} H_{n1}(\lambda) &= \frac{1}{n}\left(G_n\xi_{p_n}\theta^*\right)' M_n B_n(\lambda)\mathbf{U}_n, \\ H_{n2}(\lambda) &= \frac{1}{n}\mathbf{U}'_n B'_n(\lambda) M_n B_n(\lambda)\mathbf{U}_n, \\ H_{n3}(\lambda) &= \frac{2}{n}\mathbf{R}'_n B'_n(\lambda) M_n\left(2(\lambda_0 - \lambda) G_n\xi_{p_n}\theta^* + B_n(\lambda)\mathbf{U}_n\right), \\ H_{n4}(\lambda) &= \frac{1}{n}\mathbf{R}'_n B'_n(\lambda) M_n B_n(\lambda)\mathbf{R}_n. \end{aligned}$$

Note that the parameter function $\theta^*(\cdot)$ is square integrable; therefore, $\|\theta^*\|_2 < \infty$. Then, by Lemma 4.1 and 4.2,

$$H_{n0} = O_p\left(\frac{p_n}{h_n\sqrt{n}}\right). \tag{4.23}$$

Also, Lemma 4.3 implies that $H_{n3}(\lambda)$ and $H_{n4}(\lambda)$ are of order $o_p(1)$ uniformly on λ in Λ. In the following, we show that $H_{n1}(\lambda)$ and $H_{n2}(\lambda) - \sigma^2_{n,\lambda}$ are all of order $o_p(1)$ for all $\lambda \in \Lambda$.

Proof of $H_{n1}(\lambda)$:
Note that

$$\begin{aligned} E\left(\left\|\mathbf{U}_n^{'} G_n \xi_{p_n}\right\|^2\right) &= \sum_{r=1}^{p_n} E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} U_i G_{ij} \varepsilon_r^{(j)}\right)^2 \\ &= \sigma_0^2 \sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}^2 \sum_{r=1}^{p_n} E\left(\varepsilon_r^2\right) = O\left(\|G_n\|^2\right), \end{aligned}$$

since $\sum_{r=1}^{p_n} E\left(\varepsilon_r^2\right) < E\left(\int X^2(t)dt\right) < \infty$. Therefore,

$$\xi_{p_n}^{'} \mathbf{U}_n = O_p(\sqrt{n}) \qquad \text{and} \qquad \mathbf{U}_n^{'} G_n \xi_{p_n} = O_p\left(\sqrt{\frac{n}{h_n}}\right), \tag{4.24}$$

by **Assumption 1**-ii. In addition, by Lemma 4.1, we have

$$\left|\xi_{p_n}^{'} G_n^{'} \xi_{p_n} \left(\xi_{p_n}^{'} \xi_{p_n}\right)^{-1} \xi_{p_n}^{'} \mathbf{U}_n\right| = O_p\left(\frac{p_n\sqrt{n}}{h_n}\right),$$

and

$$\left|\xi_{p_n}^{'} G_n^{'} \xi_{p_n} \left(\xi_{p_n}^{'} \xi_{p_n}\right)^{-1} \xi_{p_n}^{'} G_n \mathbf{U}_n\right| = O_p\left(p_n\sqrt{\frac{n}{h_n^3}}\right).$$

Then, for each $\lambda \in \Lambda$, we may conclude that

$$\begin{aligned} H_{n1}(\lambda) &= \frac{1}{n}\left(G_n \xi_{p_n} \theta^*\right)^{'} M_n \mathbf{U}_n + (\lambda_0 - \lambda)\frac{1}{n}\left(G_n \xi_{p_n} \theta^*\right)^{'} M_n G_n \mathbf{U}_n \\ &= O_p\left(\frac{p_n + \sqrt{h_n}}{h_n\sqrt{n}}\right), \end{aligned}$$

hence the results follows by **Assumption 2**.

■

Proof of $H_{n2}(\lambda)$:
For each $\lambda \in \Lambda$, we have

$$H_{n2}(\lambda) - \sigma_{n,\lambda}^2 = \frac{1}{n}\mathbf{U}_n^{'} A_n(\lambda)\mathbf{U}_n - \frac{\sigma_0^2}{n}\mathrm{tr}\left(A_n(\lambda)\right) - T_n(\lambda),$$

with

$$T_n(\lambda) = \frac{1}{n}\mathbf{U}_n^{'} B_n^{'}(\lambda)\xi_{p_n}\left(\xi_{p_n}^{'}\xi_{p_n}\right)^{-1}\xi_{p_n}^{'} B_n(\lambda)\mathbf{U}_n.$$

Similar to (4.24), we have

$$T_n(\lambda) = O_p\left(\frac{p_n\|B_n(\lambda)\|^2}{n^2}\right) = O_p\left(\frac{p_n}{n}\right),$$

since $\|B_n(\lambda)\|^2 = O(n)$ uniformly on λ. We have also,

$$E\left(\frac{1}{n}\mathbf{U}_n^{'} A_n(\lambda)\mathbf{U}_n\right) = \sigma_{n,\lambda}^2$$

and

$$\begin{aligned} \mathrm{Var}\left(\mathbf{U}_n^{'} A_n(\lambda)\mathbf{U}_n\right) &= (\mu_4 - 3\sigma_0^2)\sum_{i=1}^{n} A_{ii}^2(\lambda) + \sigma_0^4\left[\|A_n(\lambda)\|^2 + \mathrm{tr}(A_n^2(\lambda))\right] \\ &= O\left(\|A_n(\lambda)\|^2\right), \end{aligned}$$

with the symmetry of $A_n(\lambda) = B_n^{'}(\lambda)B_n(\lambda)$. Consequently,

$$\frac{1}{n}\mathbf{U}_n^{'}A_n(\lambda)\mathbf{U}_n - \frac{\sigma_0^2}{n}\mathrm{tr}\left(A_n(\lambda)\right) = O_p\left(\frac{\|A_n(\lambda)\|}{n}\right) = O_p(n^{-1/2}),$$

since $\|A_n(\lambda)\| = O(n^{1/2})$ uniformly on λ. This yields the proof of $H_{n2}(\lambda)$ and therefore that of (i).

■

Proof of (ii):
We start to show the following convergence

$$\frac{h_n}{n}\left\{\left(\tilde{L}_n(\lambda) - \tilde{L}_n(\lambda_0)\right) - \left(Q_n(\lambda) - Q_n(\lambda_0)\right)\right\} = o_p(1).$$

Recall that,

$$\tilde{L}_n(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}_{n,\lambda}^2 + \ln|S_n(\lambda)|,$$

$$\hat{\sigma}_{n,\lambda}^2 = \frac{1}{n}\mathbf{Y}_n^{'}S_n^{'}(\lambda)M_nS_n(\lambda)\mathbf{Y}_n,$$

and

$$\sigma_{n,\lambda}^{*2} = \frac{1}{n}(\lambda_0 - \lambda)^2\Delta_n + \frac{\sigma_0^2}{n}\mathrm{tr}\left(A_n(\lambda)\right).$$

Then, we have

$$\begin{aligned}
&\frac{h_n}{n}\left\{\left(\tilde{L}_n(\lambda) - \tilde{L}_n(\lambda_0)\right) - \left(Q_n(\lambda) - Q_n(\lambda_0)\right)\right\} \\
&\qquad = -\frac{h_n}{2}\left\{\left(\ln\hat{\sigma}_{n,\lambda}^2 - \ln\sigma_{n,\lambda}^{*2}\right) - \left(\ln\hat{\sigma}_{n,\lambda_0}^2 - \ln\sigma_{n,\lambda_0}^{*2}\right)\right\} + o(1), \\
&\qquad = -\frac{h_n}{2}\frac{\partial\left(\ln\hat{\sigma}_{n,\lambda_n}^2 - \ln\sigma_{n,\lambda_n}^{*2}\right)}{\partial\lambda}(\lambda - \lambda_0) + o(1),
\end{aligned}$$

since $\mathrm{tr}(B_n(\lambda) - B_n(\lambda_0))$ and $\mathrm{tr}(A_n(\lambda) - A_n(\lambda_0))$ are of order $O(\frac{n}{h_n})$, $\epsilon_{n1}, \epsilon_{n4}$ are of order $o(1)$, and λ_n lies between λ and λ_0.
Note that

$$\frac{\partial\hat{\sigma}_{n,\lambda}^2}{\partial\lambda} = -\frac{2}{n}\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda)\mathbf{Y}_n,$$

and

$$\frac{\partial\sigma_{n,\lambda}^{*2}}{\partial\lambda} = \frac{2}{n}\left[(\lambda - \lambda_0)\Delta_n - \sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda)\right)\right].$$

This implies that

$$\begin{aligned}
&\frac{h_n}{n}\left\{\left(\tilde{L}_n(\lambda) - \tilde{L}_n(\lambda_0)\right) - \left(Q_n(\lambda) - Q_n(\lambda_0)\right)\right\} \\
&\qquad = \frac{h_n}{n}\frac{1}{\hat{\sigma}_{n,\lambda_n}^2}\left\{\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda_n)\mathbf{Y}_n\right. \\
&\qquad \left. -\frac{\hat{\sigma}_{n,\lambda_n}^2}{\sigma_{n,\lambda_n}^{*2}}\left[(\lambda_0 - \lambda_n)\Delta_n + \sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda_n)\right)\right]\right\} \\
&\qquad = \frac{h_n}{n}\frac{1}{\hat{\sigma}_{n,\lambda_n}^2}\left\{\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda)\mathbf{Y}_n - \left[(\lambda_0 - \lambda_n)\Delta_n + \sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda_n)\right)\right]\right. \\
&\qquad\qquad \left. -\frac{\hat{\sigma}_{n,\lambda_n}^2 - \sigma_{n,\lambda_n}^{*2}}{\sigma_{n,\lambda_n}^{*2}}\left[(\lambda_0 - \lambda_n)\Delta_n + \sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda_n)\right)\right]\right\}(\lambda - \lambda_0).
\end{aligned}$$

By noting that $B_n(\lambda) = I_n + (\lambda_0 - \lambda)G_n$ and let $\mathbf{V}_n = \xi_{p_n}\theta^*$, we have

$$\begin{aligned}\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda)\mathbf{Y}_n &= (\lambda_0-\lambda)\left[\mathbf{V}_n+\mathbf{R}_n+\mathbf{U}_n\right]^{'}G_n^{'}M_nG_n\left[\mathbf{V}_n+\mathbf{R}_n+\mathbf{U}_n\right] \\ &\quad +\left[\mathbf{V}_n+\mathbf{R}_n+\mathbf{U}_n\right]^{'}G_n^{'}M_n\left[\mathbf{R}_n+\mathbf{U}_n\right] \\ &= (\lambda_0-\lambda)\left[\mathbf{V}_n^{'}G_n^{'}M_nG_n\left[\mathbf{V}_n+2\mathbf{U}_n\right]+\mathbf{U}_n^{'}G_n^{'}M_nG_n\mathbf{U}_n\right] \\ &\quad +\mathbf{U}_n^{'}M_nG_n\left[\mathbf{V}_n+\mathbf{U}_n\right]+\mathbf{R}_n^{'}M_nG_n\left[\mathbf{V}_n+\mathbf{R}_n+\mathbf{U}_n\right] \\ &\quad +2(\lambda_0-\lambda)\mathbf{R}_n^{'}G_n^{'}M_nG_n\left[\mathbf{V}_n+\mathbf{R}_n+\mathbf{U}_n\right].\end{aligned}$$

We have

$$\frac{h_n}{n}\left(\mathbf{V}_n^{'}G_n^{'}M_nG_n\mathbf{V}_n-\Delta_n\right)=h_nH_{n0}=O_p\left(\frac{p_n}{\sqrt{n}}\right). \tag{4.25}$$

By the proof of $H_{n1}(\lambda)$, we have

$$\sqrt{\frac{h_n}{n}}\mathbf{V}_n^{'}G_n^{'}M_n\left[I_n+(\lambda_0-\lambda)G_n\right]\mathbf{U}_n=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right). \tag{4.26}$$

By the proof of $H_{n2}(\lambda)$, we have

$$\sqrt{\frac{h_n}{n}}\left[\mathbf{U}_n^{'}G_n^{'}M_n\mathbf{U}_n-\sigma_0^2\mathrm{tr}(G_n)\right]=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right) \qquad \text{and}$$

$$\sqrt{\frac{h_n}{n}}\left[\mathbf{U}_n^{'}G_n^{'}M_nG_n\mathbf{U}_n-\sigma_0^2\mathrm{tr}(G_n^{'}G_n)\right]=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right). \tag{4.27}$$

Therefore, by Lemma 4.3, we may write

$$\begin{aligned}\sqrt{\frac{h_n}{n}}\left\{\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda_n)\mathbf{Y}_n-(\lambda_0-\lambda_n)\Delta_n-\sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda_n)\right)\right\} \\ =O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right).\end{aligned} \tag{4.28}$$

Note that when h_n is unbounded, we have

$$\sigma_{n,\lambda}^2=\sigma_0^2+o(1),$$

since $\mathrm{tr}(G_n)$ and $\mathrm{tr}(G_n^{'}G_n)$ are of order $O(n/h_n)$. Thus, $1/\sigma_{n,\lambda}^{*2}=O(1)$ uniformly in λ, because $\sigma_{n,\lambda}^{*2}\geq\sigma_{n,\lambda}^2$ and $\sigma_0^2>0$. However, we have also $1/\hat{\sigma}_{n,\lambda}^2=O_p(1)$ by (4.21).

Now, note that under **Assumption 1** (ii-iii), Δ_n and $\mathrm{tr}(G_n^{'}B_n(\lambda))$ are of order $O(n/h_n)$ and using (4.21) and (4.28), we conclude

$$\frac{h_n}{n}\left\{\left(\tilde{L}_n(\lambda)-\tilde{L}_n(\lambda_0)\right)-\left(Q_n(\lambda)-Q_n(\lambda_0)\right)\right\}=o_p(1), \tag{4.29}$$

uniformly in $\lambda\in\Lambda$, since $p_n^2=o(n)$ by **Assumption 2**.

Let us proof the uniform equicontinuity of $\dfrac{h_n}{n}\left[Q_n(\lambda)-Q_n(\lambda_0)\right]$.

Recall that

$$\frac{h_n}{n}\left[Q_n(\lambda)-Q_n(\lambda_0)\right]=-\frac{h_n}{2}\left(\ln\sigma_{n,\lambda}^{*2}-\ln\sigma_0^2\right)+\frac{h_n}{2}\left(\ln|S_n(\lambda)|-\ln|S_n|\right)+o(1).$$

Since $\mathrm{tr}(A_n(\lambda)) - n = 2(\lambda_0 - \lambda)\mathrm{tr}(G_n) + (\lambda_0 - \lambda)^2\mathrm{tr}(G_n'G_n)$, we have

$$\begin{aligned} h_n(\sigma^{*2}_{n,\lambda} - \sigma_0^2) &= (\lambda_0 - \lambda)^2 \frac{h_n}{n}\Delta_n + \sigma_2^2 \frac{h_n}{n}(\mathrm{tr}(A_n(\lambda)) - n) \\ &= (\lambda_0 - \lambda)^2 \frac{h_n}{n}\Delta_n + 2\sigma_2^2 \frac{h_n}{n}(\lambda_0 - \lambda)\mathrm{tr}(G_n) \\ &\quad + \sigma_2^2 \frac{h_n}{n}(\lambda_0 - \lambda)^2 \mathrm{tr}(G_n'G_n), \end{aligned}$$

is uniformly equicontinuous in $\lambda \in \Lambda$ by **Assumption 1**. By the mean value theorem,

$$h_n\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma_0^2\right) = \frac{h_n}{\tilde{\sigma}^2_{n,\lambda}}(\sigma^{*2}_{n,\lambda} - \sigma_0^2),$$

where $\tilde{\sigma}^2_{n,\lambda}$ lies between σ_0^2 and $\sigma^{*2}_{n,\lambda}$. Consequently, it is uniformly bounded from above. Hence, $h_n\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma_0^2\right)$ is uniformly equicontinuous on Λ.
Then, the function

$$\frac{h_n}{n}\left(\ln|S_n(\lambda) - \ln|S_n|\right) = \frac{h_n}{n}\mathrm{tr}(W_n S_n^{-1}(\tilde{\lambda}_n))(\lambda - \lambda_0),$$

is uniformly equicontinuous on Λ because $\mathrm{tr}(W_n S_n^{-1}(\lambda)) = O(n/h_n)$ uniformly on λ by **Assumption 1**.
In conclusion, $\dfrac{h_n}{n}\left(Q_n(\lambda) - Q_n(\lambda_0)\right)$ is uniformly equicontinuous on Λ.

Let us prove uniqueness of the maximum λ_0.

Let

$$D_n(\lambda) = -\frac{h_n}{2}\left(\ln\sigma^2_{n,\lambda} - \ln\sigma_0^2\right) + \frac{h_n}{n}\left(\ln|S_n(\lambda)| - \ln|S_n|\right).$$

Then,

$$\frac{h_n}{n}\left(Q_n(\lambda) - Q_n(\lambda_0)\right) = D_n(\lambda) - \frac{h_n}{2}\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma^2_{n,\lambda}\right).$$

We have by the Taylor expansion,

$$h_n\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma^2_{n,\lambda}\right) = \frac{\sigma^{*2}_{n,\lambda} - \sigma^2_{n,\lambda}}{\tilde{\sigma}^2_{n,\lambda}} = \frac{(\lambda - \lambda_0)^2}{\tilde{\sigma}^2_{n,\lambda}}\frac{h_n}{n}\Delta_n,$$

where $\tilde{\sigma}^2_{n,\lambda}$ lies between $\sigma^{*2}_{n,\lambda}$ and $\sigma^2_{n,\lambda}$. Since $\sigma^{*2}_{n,\lambda} \geq \sigma^2_{n,\lambda}$ for all $\lambda \in \Lambda$, it follows

$$h_n\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma^2_{n,\lambda}\right) \geq \frac{(\lambda - \lambda_0)^2}{\sigma^{*2}_{n,\lambda}}\frac{h_n}{n}\Delta_n.$$

As h_n is unbounded and under **Assumption 1**, $\sigma^{*2}_{n,\lambda} - \sigma^2_{n,\lambda} = o(1)$ uniformly on Λ. Thus, $\lim_{n\to\infty}\sigma^{*2}_{n,\lambda} = \sigma_0^2$.
Therefore, under **Assumption 3** (a),

$$\begin{aligned} -\lim_{n\to\infty} h_n\left(\ln\sigma^{*2}_{n,\lambda} - \ln\sigma^2_{n,\lambda}\right) &\leq -\lim_{n\to\infty}\frac{(\lambda - \lambda_0)^2}{\sigma^{*2}_{n,\lambda}}\frac{h_n}{n}\Delta_n \\ &= -\frac{(\lambda - \lambda_0)^2}{\sigma_0^2}\lim_{n\to\infty}\frac{h_n}{n}\Delta_n < 0, \end{aligned}$$

for any $\lambda \neq \lambda_0$. Furthermore, under **Assumption 3** (b), $D_n(\lambda) < 0$, if $\lambda \neq \lambda_0$.
In conclusion, for a certain rank, we have $\frac{h_n}{n}\left(Q_n(\lambda) - Q_n(\lambda_0)\right) < 0$, when $\lambda \neq \lambda_0$.
The proof of (ii) follows from the uniform convergence (4.29) and the identification uniqueness condition.

■

Proof of Theorem 4.1

Identification and consistency of $\hat{\lambda}_n$ are given by Proposition 4.1. Let us now focus on the asymptotic normality of $\hat{\lambda}_n$.
Consider the first and second order derivatives of the concentrated log likelihood $\tilde{L}_n(\lambda)$:

$$\frac{\partial \tilde{L}_n(\lambda)}{\partial \lambda} = \frac{1}{\hat{\sigma}^2_{n,\lambda}} \mathbf{Y}_n^{'} W_n^{'} M_n S_n(\lambda) \mathbf{Y}_n - \mathrm{tr}\left(W_n S_n^{-1}(\lambda)\right),$$

and

$$\begin{aligned}\frac{\partial^2 \tilde{L}_n(\lambda)}{\partial \lambda^2} &= \frac{2}{n\hat{\sigma}^4_{n,\lambda}} \left[\mathbf{Y}_n^{'} W_n^{'} M_n S_n(\lambda) \mathbf{Y}_n\right]^2 \\ &\quad - \frac{1}{\hat{\sigma}^2_{n,\lambda}} \mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n - \mathrm{tr}\left(\left[W_n S_n^{-1}(\lambda)\right]^2\right).\end{aligned}$$

By (4.26) and Lemma 4.3, we have

$$\frac{h_n}{n} \mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n = \frac{h_n}{n} \mathbf{V}_n^{'} G_n^{'} M_n G_n \mathbf{V}_n + \frac{h_n}{n} \mathbf{U}_n^{'} G_n^{'} M_n G_n \mathbf{U}_n + o_p(1), \tag{4.30}$$

and

$$\begin{aligned}\frac{h_n}{n} \mathbf{Y}_n^{'} W_n^{'} M_n S_n(\lambda) \mathbf{Y}_n &= \frac{h_n}{n} \mathbf{U}_n^{'} G_n^{'} M_n \mathbf{U}_n + (\lambda_0 - \lambda) \frac{h_n}{n} \mathbf{V}_n^{'} G_n^{'} M_n G_n \mathbf{V}_n \\ &\quad + (\lambda_0 - \lambda) \frac{h_n}{n} \mathbf{U}_n^{'} G_n^{'} M_n G_n \mathbf{U}_n + o_p(1) \\ &= O_p(1),\end{aligned}$$

by (4.28) and since under **Assumption 1**, Δ_n and $\mathrm{tr}(G_n B_n(\lambda))$ are of order $O_p(n/h_n)$, uniformly in λ.
From (4.21), we proved that $\hat{\sigma}^2_{n,\lambda} = \sigma^{*2}_{n,\lambda} + o_p(1)$. Thus, we have

$$\begin{aligned}\frac{h_n}{n} \frac{\partial^2 \tilde{L}_n(\lambda)}{\partial \lambda^2} &= -\frac{1}{\sigma^{*2}_{n,\lambda}} \left[\frac{h_n}{n} \mathbf{V}_n^{'} G_n^{'} M_n G_n \mathbf{V}_n + \frac{h_n}{n} \mathbf{U}_n^{'} G_n^{'} M_n G_n \mathbf{U}_n\right] \\ &\quad - \frac{h_n}{n} \mathrm{tr}\left(\left[W_n S_n^{-1}(\lambda)\right]^2\right) + o_p(1),\end{aligned}$$

uniformly on Λ. For any $\tilde{\lambda}_n$ that converges in probability to λ_0, one can easily show that

$$\sigma^{*2}_{n,\tilde{\lambda}_n} - \sigma^{*2}_{n,\lambda_0} = o_p(1),$$

and as $\sigma^{*2}_{n,\lambda} \geq \sigma_0^2 > 0$ uniformly on Λ, we can conclude by the Taylor expansion

$$\begin{aligned}\frac{h_n}{n}\left[\frac{\partial^2 \tilde{L}_n(\tilde{\lambda}_n)}{\partial \lambda^2} - \frac{\partial^2 \tilde{L}_n(\lambda_0)}{\partial \lambda^2}\right] &= \frac{h_n}{n}\left[\mathrm{tr}\left(W_n S_n^{-1}(\tilde{\lambda}_n)\right)^2 - \mathrm{tr}\left(G_n^2\right)\right] + o_p(1) \\ &= -2(\tilde{\lambda}_n - \lambda_0) \frac{h_n}{n} \mathrm{tr}\left(G_n^3(\bar{\lambda}_n)\right) + o_p(1) \\ &= o_p(1),\end{aligned}$$

as under **Assumption 1**, $\mathrm{tr}\left(G_n^3(\lambda)\right)$ is of order $O(n/h_n)$ uniformly on Λ.

Finally, using (4.25), (4.27), and the fact that $\sigma^{*2}_{n,\lambda_0} = \sigma^2_0$, we have

$$\frac{h_n}{n}\frac{\partial^2 \tilde{L}_n(\lambda_0)}{\partial \lambda^2} = -\frac{1}{\sigma^2_0}\frac{h_n}{n}\Delta_n - \frac{h_n}{n}\left[\mathrm{tr}(G_n^{'}G_n) + \mathrm{tr}\left(G_n^2\right)\right] + o_p(1). \tag{4.31}$$

Let us now prove the asymptotic normality of $\sqrt{\frac{h_n}{n}}\frac{\partial \tilde{L}_n(\lambda_0)}{\partial \lambda}$.
Using the results of Lemma 4.3, we have

$$\sqrt{\frac{h_n}{n}}\mathbf{Y}_n^{'}W_n^{'}M_nS_n\mathbf{Y}_n = \sqrt{\frac{h_n}{n}}\left[\mathbf{V}_n^{'} + \mathbf{U}_n^{'}\right]G_n^{'}M_n\mathbf{U}_n + o_p(1), \tag{4.32}$$

and

$$\hat{\sigma}^2_{n,\lambda_0} = \frac{1}{n}\mathbf{Y}_n^{'}S_n^{'}M_nS_n\mathbf{Y}_n = \frac{1}{n}\mathbf{U}_n^{'}M_n\mathbf{U}_n + o_p(1).$$

It follows that

$$\sqrt{\frac{h_n}{n}}\frac{\partial \tilde{L}_n(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}^2_{n,\lambda_0}}\sqrt{\frac{h_n}{n}}\left[\mathbf{V}_n^{'}G_n^{'}M_n\mathbf{U}_n + \mathbf{U}_n^{'}C_n^{'}M_n\mathbf{U}_n\right] + o_p(1),$$

where $C_n = G_n - \mathrm{tr}(\frac{G_n}{n})I_n$. Using (4.24), we have

$$\sqrt{\frac{h_n}{n}}\mathbf{U}_n^{'}C_n^{'}\xi_{p_n}(\xi_{p_n}^{'}\xi_{p_n})^{-1}\xi_{p_n}^{'}\mathbf{U}_n = O_p\left(\frac{p_n}{\sqrt{n}}\right), \tag{4.33}$$

since under **Assumption 1**, the matrix C_n is uniformly bounded in both row and column sums, and $C_{ij} = O(1/h_n)$ uniformly in i and j.
Consider the following decomposition

$$\begin{aligned}
\xi_{p_n}^{'}G_n^{'}\xi_{p_n}(\xi_{p_n}^{'}\xi_{p_n})^{-1}\xi_{p_n}^{'}\mathbf{U}_n &= \left[\frac{\xi_{p_n}^{'}G_n^{'}\xi_{p_n}}{n} - \mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right]\left[\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right]^{-1}\xi_{p_n}^{'}\mathbf{U}_n \\
&\quad -\mathrm{tr}\left(\frac{G_n}{n}\right)\left[\frac{\xi_{p_n}^{'}\xi_{p_n}}{n} - \Gamma_{p_n}\right]\left[\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right]^{-1}\xi_{p_n}^{'}\mathbf{U}_n + \mathrm{tr}\left(\frac{G_n}{n}\right)\xi_{p_n}^{'}\mathbf{U}_n \\
&= \mathrm{tr}\left(\frac{G_n}{n}\right)\xi_{p_n}^{'}\mathbf{U}_n + O_p\left(\frac{p_n^2}{h_n}\right),
\end{aligned}$$

by (4.24) and Lemma 4.1. Thus

$$\sqrt{\frac{h_n}{n}}\mathbf{V}_n^{'}G_n^{'}\xi_{p_n}(\xi_{p_n}^{'}\xi_{p_n})^{-1}\xi_{p_n}^{'}\mathbf{U}_n = \frac{\sqrt{h_n}}{n}\mathrm{tr}\left(G_n\right)\frac{\mathbf{V}_n^{'}\mathbf{U}_n}{\sqrt{n}} + O_p\left(\frac{p_n^2}{\sqrt{nh_n}}\right). \tag{4.34}$$

Consequently, (4.33) and (4.34) imply

$$\sqrt{\frac{h_n}{n}}\frac{\partial \tilde{L}_n(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}^2_{n,\lambda_0}}\sqrt{\frac{h_n}{n}}\left[\mathbf{V}_n^{'}D_n^{'}\mathbf{U}_n + \mathbf{U}_n^{'}C_n^{'}\mathbf{U}_n\right] + o_p(1),$$

with $D_n = G_n + \mathrm{tr}(\frac{G_n}{n})I_n$.
Let $G_n^s = (G_n + G_n^{'})/2$, $C_n^s = (C_n + C_n^{'})/2$, and $D_n^s = (D_n + D_n^{'})/2$. These matrices satisfy $C_{ij}^s = D_{ij}^s = G_{ij}^s$ for all $i \neq j$.

Now, because $\mathrm{tr}(C_n) = 0$, one can consider the decomposition

$$\mathbf{V}_n^{'} D_n^{'} \mathbf{U}_n + \mathbf{U}_n^{'} C_n^{'} \mathbf{U}_n = \sum_{i=1}^{n} Z_{ni}, \tag{4.35}$$

with

$$Z_{ni} = D_{ii} U_i V_i + C_{ii}(U_i^2 - \sigma_0^2) + 2U_i \sum_{j=1}^{i-1} G_{ij}^s T_j,$$

where $T_i = V_i + U_i$, $i = 1, \ldots, n$. It is easy to show that

$$\begin{aligned} \sum_{i=1}^{n} E\left(Z_{ni}^2\right) &= \sigma_0^2 \left[E(V^2) + \sigma_0^2\right] \mathrm{tr}\left(G_n(G_n^{'} + G_n)\right) + \left[3\sigma_0^2 E(V^2) + \sigma_0^4 - \mu_4\right] \frac{1}{n}\mathrm{tr}^2(G_n) \\ &\quad + \left[\mu_4 - 3\sigma_0^4 - \sigma_0^2 E(V^2)\right] \sum_{i=1}^{n} G_{ii}^2. \end{aligned}$$

Finally, let

$$s_Z^2 = \lim_{n\to\infty} \frac{h_n}{n} \sum_{i=1}^{n} E\left(Z_{ni}^2\right) \qquad \text{and} \qquad \tilde{Z}_{ni} = \sqrt{\frac{h_n}{n}} \frac{Z_{ni}}{s_Z}.$$

Note that condition C.1 in Lemma 4.5 implies that $\left\{\tilde{Z}_{ni},\, i = 1, \ldots, n\ n = 1, 2, \ldots\right\}$ form a triangular array of martingale differences sequences. According to Kelejian & Prucha (Theorem A.1, 2001, p.240) and under conditions C.2 and C.3 in Lemma 4.5, we have

$$\sqrt{\frac{h_n}{n}} \frac{\partial \tilde{L}_n(\lambda_0)}{\partial \lambda} = \frac{s_Z}{\hat{\sigma}_{n,\lambda_0}^2} \sum_{i=1}^{n} \tilde{Z}_{ni} + o_p(1) \to \mathcal{N}\left(0, \frac{s_Z^2}{\sigma_0^4}\right). \tag{4.36}$$

Finally, using (4.31) and (4.36) we can conclude by the Taylor expansion, that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, s_\lambda^2), \tag{4.37}$$

where

$$s_\lambda^2 = \lim_{n\to\infty} s_Z^2 \left\{\frac{h_n}{n}\left[\Delta_n + \sigma_0^2 \mathrm{tr}(G_n(G_n^{'} + G_n))\right]\right\}^{-2}.$$

This concludes the proof of Theorem 4.1.

■

Proof of Theorem 4.2

Let us consider the decomposition $S_n(\hat{\lambda}_n) = S_n + (\lambda_0 - \hat{\lambda}_n) W_n$ and note that

$$\begin{aligned} \hat{\sigma}_{n,\hat{\lambda}_n}^2 &= \frac{1}{n} \mathbf{Y}_n^{'} S_n^{'}(\hat{\lambda}_n) M_n S_n(\hat{\lambda}_n) \mathbf{Y}_n \\ &= \frac{1}{n} \mathbf{Y}_n^{'} S_n^{'} M_n S_n \mathbf{Y}_n + 2(\lambda_0 - \hat{\lambda}_n) \frac{1}{n} \mathbf{Y}_n^{'} W_n^{'} M_n S_n \mathbf{Y}_n \\ &\qquad + (\lambda_0 - \hat{\lambda}_n)^2 \frac{1}{n} \mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n. \end{aligned}$$

Lemma 4.3 and (4.33) imply that

$$\frac{1}{n} \mathbf{Y}_n^{'} S_n^{'} M_n S_n \mathbf{Y}_n = \frac{1}{n} \mathbf{U}_n^{'} \mathbf{U}_n + o_p(1).$$

Thus

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) &= \sqrt{\frac{n}{h_n}}(\lambda_0 - \hat{\lambda}_n)^2 \frac{\sqrt{h_n}}{n} \mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n \\ &\quad -2\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)\frac{\sqrt{h_n}}{n}\mathbf{Y}_n^{'} W_n^{'} M_n S_n \mathbf{Y}_n + \frac{1}{\sqrt{n}}(\mathbf{U}_n^{'}\mathbf{U}_n - n\sigma_0^2).\end{aligned}$$

Note that (4.26), (4.32) and (4.33) imply

$$\frac{\sqrt{h_n}}{n}\mathbf{Y}_n^{'} W_n^{'} M_n S_n \mathbf{Y}_n = \frac{\sqrt{h_n}}{n}\mathrm{tr}(G_n) + o_p(1) = O_p\left(\frac{1}{\sqrt{h_n}}\right). \tag{4.38}$$

By (4.25), (4.27) and (4.30), we have

$$\frac{\sqrt{h_n}}{n}\mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n = \frac{\sqrt{h_n}}{n}\Delta_n + \sigma_0^2\frac{\sqrt{h_n}}{n}\mathrm{tr}\left(G_n G_n^{'}\right) + o_p(1) = O_p\left(\frac{1}{\sqrt{h_n}}\right).$$

Consequently, the asymptotic normality of $\hat{\lambda}_n$ implies

$$\sqrt{\frac{n}{h_n}}(\lambda_0 - \hat{\lambda}_n)^2 \frac{\sqrt{h_n}}{n}\mathbf{Y}_n^{'} W_n^{'} M_n W_n \mathbf{Y}_n = o_p(1).$$

If $\lim_{n\to\infty} h_n = \infty$, (4.38) will be of order $o_p(1)$. Hence

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) = \frac{1}{\sqrt{n}}(\mathbf{U}_n^{'}\mathbf{U}_n - n\sigma_0^2) + o_p(1) \to \mathcal{N}(0, \mu_4 - \sigma_0^4).$$

Otherwise, we have

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) &= \frac{1}{\sqrt{n}}(\mathbf{U}_n^{'}\mathbf{U}_n - n\sigma_0^2) \\ &\quad -2\frac{\sqrt{h_n}}{n}\mathrm{tr}(G_n)\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_p(1).\end{aligned} \tag{4.39}$$

By the asymptotic normality proof of $\hat{\lambda}_n$ (see (4.31) and (4.35)), one can conclude

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) = -\delta_n\sqrt{\frac{h_n}{n}}\sum_{i=1}^{n} Z_{ni} + o_p(1),$$

where

$$\delta_n = \frac{n}{h_n}\left[\Delta_n + \sigma_0^2\mathrm{tr}\left(G_n(G_n^{'} + G_n)\right)\right]^{-1}.$$

Therefore, on can rewrite (4.39) as

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) = 2\delta_n\frac{\sqrt{h_n}}{n}\mathrm{tr}(G_n)\sqrt{\frac{n}{h_n}}\sum_{i=1}^{n} Z_{ni}^{\dagger} + o_p(1), \tag{4.40}$$

where

$$Z_{ni}^{\dagger} = D_{ii}U_iV_i + \tilde{C}_{ii}(U_i^2 - \sigma_0^2) + 2U_i\sum_{j=1}^{i-1} G_{ij}^s T_j,$$

where $\tilde{C}_{ii} = C_{ii} + \frac{n}{2\delta_n\mathrm{tr}(G_n)}$, $\tilde{C}_{ii}$ is bounded uniformly in i, when h_n is bounded.
It is easy to show that

$$\sum_{i=1}^{n} E\left(Z_{ni}^{\dagger 2}\right) = \sum_{i=1}^{n} E\left(Z_{ni}^2\right) + n(\mu_4 - \sigma_0^4)\left[\frac{n}{2\delta_n\mathrm{tr}(G_n)}\right]^2.$$

Let

$$s^2_{Z^\dagger} = \lim_{n\to\infty} \frac{h_n}{n} \sum_{i=1}^{n} E\left(Z^{\dagger 2}_{ni}\right) \qquad \text{and} \qquad \tilde{Z}^\dagger_{ni} = \sqrt{\frac{h_n}{n}} \frac{Z^\dagger_{ni}}{s_{Z^\dagger}}.$$

Note that conditions C.1-C.3 in Lemma 4.5 hold when Z_{ni} and $\tilde{Z}_{ni}$ are replaced by $Z^\dagger_{ni}$ and $\tilde{Z}^\dagger_{ni}$ respectively. Therefore, Kelejian & Prucha (Theorem A.1, 2001, p.240) implies that

$$\sum_{i=1}^{n} \tilde{Z}^\dagger_{ni} \to \mathcal{N}(0,1). \tag{4.41}$$

Finally, by (4.40) and (4.41), we have

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma^2_0) \to \mathcal{N}(0, s^2_\sigma),$$

where

$$s^2_\sigma = \lim_{n\to\infty} h_n s^2_{Z^\dagger} \left[\frac{2\delta_n \mathrm{tr}(G_n)}{n}\right]^2 = \mu_4 - \sigma^4_0 + 4s^2_\lambda \lim_{n\to\infty} h_n \left[\frac{\mathrm{tr}(G_n)}{n}\right]^2.$$

This finishes the proof.

■

Proof of Theorem 4.3

Recall that $S_n(\lambda)S_n^{-1} = I_n + (\lambda_0 - \lambda)G$, for all $\lambda \in \Lambda$, and

$$\hat{\theta}_{n,\hat{\lambda}_n} = (\xi'_{p_n}\xi_{p_n})^{-1}\xi'_{p_n} S_n(\hat{\lambda}_n)\mathbf{Y}_n. \tag{4.42}$$

By Lemma 4.3, we have

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right) &= \sqrt{n}(\lambda_0 - \hat{\lambda}_n)\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}G_n\xi_{p_n}}{n}\theta^* + \frac{\xi'_{p_n}G_n\mathbf{U}_n}{n}\right] \\ &\quad + \left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right] + o_p(1).\end{aligned}$$

By Lemma 4.1, we have

$$\begin{aligned}\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\frac{\xi'_{p_n}G_n\xi_{p_n}}{n} &= \left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}G_n\xi_{p_n}}{n} - \mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right] \\ &\quad -\mathrm{tr}\left(\frac{G_n}{n}\right)\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\xi_{p_n}}{n} - \Gamma_{p_n}\right] + \mathrm{tr}\left(\frac{G_n}{n}\right)I_{p_n} \\ &= \mathrm{tr}\left(\frac{G_n}{n}\right)I_{p_n} + O_p\left(\frac{p^2_n}{h_n\sqrt{n}}\right).\end{aligned}$$

The asymptotic normality result of $\hat{\lambda}_n$ and (4.24), imply that

$$\sqrt{n}(\lambda_0 - \hat{\lambda}_n)\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\frac{\xi'_{p_n}G_n\mathbf{U}_n}{n} = O_p\left(\frac{p_n}{\sqrt{nh_n}}\right).$$

Hence,

$$\sqrt{n}\left(\hat{\theta}_{n,\hat{\lambda}_n} - \theta^*\right) = \left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right] + \sqrt{n}(\lambda_0 - \hat{\lambda}_n)\mathrm{tr}\left(\frac{G_n}{n}\right)\theta^* + o_p(1).$$

Therefore,

$$
\begin{aligned}
& n\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right)^{'}\Gamma_{p_n}\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right) \\
&\quad=\left\{\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\mathbf{U}_n}{\sqrt{n}}\right]\right\}^{'}\Gamma_{p_n}\left\{\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\mathbf{U}_n}{\sqrt{n}}\right]\right\} \\
&\qquad+2\sqrt{n}(\lambda_0-\hat{\lambda}_n)\mathrm{tr}\left(\frac{G_n}{n}\right)\theta^{*'}\Gamma_{p_n}\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\mathbf{U}_n}{\sqrt{n}}\right] \\
&\qquad+n(\lambda_0-\hat{\lambda}_n)^2\mathrm{tr}^2\left(\frac{G_n}{n}\right)\theta^{*'}\Gamma_{p_n}\theta^*+o_p(1). \qquad (4.43)
\end{aligned}
$$

Consider the last two terms in (4.43), we have by the asymptotic normality of $\hat{\lambda}_n$

$$
n(\lambda_0-\hat{\lambda}_n)^2\mathrm{tr}^2\left(\frac{G_n}{n}\right)\theta^{*'}\Gamma_{p_n}\theta^*=O_p\left(\frac{1}{h_n}\right). \qquad (4.44)
$$

In addition, by (4.24) and Lemma 4.1, we have

$$
\sqrt{n}(\lambda_0-\hat{\lambda}_n)\mathrm{tr}\left(\frac{G_n}{n}\right)\theta^{*'}\Gamma_{p_n}\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\mathbf{U}_n}{\sqrt{n}}\right]=O_p\left(\frac{1}{\sqrt{h_n}}\right). \qquad (4.45)
$$

Let us now give the asymptotic distribution of the first term in (4.43). Let

$$
\Psi_n=\Gamma^{\frac{1}{2}}\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\Gamma^{\frac{1}{2}},\qquad \mathcal{X}_n=\Gamma_{p_n}^{-\frac{1}{2}}\frac{\xi_{p_n}^{'}\tilde{\mathbf{U}}_n}{\sqrt{n}},\quad \text{with}\quad \tilde{\mathbf{U}}_n=\sigma_0^{-1}\mathbf{U}_n,
$$

and consider the following decomposition

$$
\begin{aligned}
\left\{\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\tilde{\mathbf{U}}_n}{\sqrt{n}}\right]\right\}^{'}\Gamma_{p_n}\left\{\left(\frac{\xi_{p_n}^{'}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi_{p_n}^{'}\tilde{\mathbf{U}}_n}{\sqrt{n}}\right]\right\} &= \mathcal{X}_n^{'}\Psi_n^2\mathcal{X}_n \\
&= \mathcal{X}_n^{'}\mathcal{X}_n-2\mathcal{X}_n^{'}(I_{p_n}-\Psi_n)\mathcal{X}_n \\
&\quad+\mathcal{X}_n^{'}(I_{p_n}-\Psi_n)^2\mathcal{X}_n. \qquad (4.46)
\end{aligned}
$$

We have, by **Assumptions 2**, **5**, **6** and Proposition 7.1 of Müller & Stadtmüller (2005),

$$
\frac{\mathcal{X}_n^{'}\mathcal{X}_n-p_n}{\sqrt{2p_n}}\to\mathcal{N}(0,1).
$$

Thus, we deduce by (4.24) and Lemma 4.4, that

$$
\mathcal{X}_n^{'}(I_{p_n}-\Psi_n)\mathcal{X}_n=o_p(\sqrt{p_n})\qquad\text{and}\qquad\mathcal{X}_n^{'}(I_{p_n}-\Psi_n)^2\mathcal{X}_n=o_p(\sqrt{p_n}).
$$

Therefore,

$$
\frac{n\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right)^{'}\Gamma_{p_n}\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right)-p_n}{\sqrt{2p_n}} = \sigma_0^2\frac{\mathcal{X}_n^{'}\mathcal{X}_n-p_n}{\sqrt{2p_n}}+O_p\left(\frac{1}{\sqrt{h_np_n}}\right)\to\mathcal{N}(0,\sigma_0^4),
$$

by (4.43), (4.44) and (4.45). This yields (4.14) and completes the proof of Theorem 4.3.

■

Lemma 4.1. *Assume that $E(\varepsilon_i^4)$ is finite, where $\varepsilon_i = \int X(t)\varphi_i(t)dt$. Under **Assumption 1**, we have*

$$\frac{\xi_{p_n}' G_n \xi_{p_n}}{n} - tr\left(\frac{G_n}{n}\right)\Gamma_{p_n} = O_p\left(\frac{p_n + \sqrt{h_n}}{h_n\sqrt{n}}\right),$$

and

$$\left\|\frac{\xi_{p_n}' G_n \xi_{p_n}}{n}\right\| = O_p\left(\frac{1}{h_n}\left[1 + \frac{p_n + \sqrt{h_n}}{\sqrt{n}}\right]\right).$$

Proof of Lemma 4.1

Note that $E(\varepsilon_r\varepsilon_s)^2 \leq E(\varepsilon_r^2)E(\varepsilon_s^2)$, and $E(\varepsilon_s^2)$ is finite since $X(\cdot)$ is square integrable. Since $E(\varepsilon_s^4)$ is finite, $E(\varepsilon_r^2\varepsilon_s^2)$ is also finite.
Note that

$$\begin{aligned}
&E\left(\left\|\xi_{p_n}' G_n \xi_{p_n} - E\left(\xi_{p_n}' G_n \xi_{p_n}\right)\right\|^2\right) \\
&\quad= \sum_{\substack{i_1=1\\ j_1=1}}^{n}\sum_{\substack{i_2=1\\ j_2=1}}^{n}\sum_{r=1}^{p_n}\sum_{s=1}^{p_n} G_{i_1 j_1} G_{i_2 j_2}\left[E\left(\varepsilon_s^{(i_1)}\varepsilon_r^{(j_1)}\varepsilon_s^{(i_2)}\varepsilon_r^{(j_2)}\right)\right. \\
&\qquad\qquad \left. -E\left(\varepsilon_s^{(i_1)}\varepsilon_r^{(j_1)}\right)E\left(\varepsilon_s^{(i_2)}\varepsilon_r^{(j_2)}\right)\right] \\
&\quad= \sum_{i=1}^{n} G_{ii}^2 \sum_{r=1}^{p_n}\sum_{s=1}^{p_n} \mathrm{Cov}\left(\varepsilon_r^2, \varepsilon_s^2\right) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}^2 \sum_{r=1}^{p_n}\sum_{s=1}^{p_n} E\left(\varepsilon_s^2\right)E\left(\varepsilon_r^2\right) \\
&\qquad\qquad + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}G_{ji} \sum_{r=1}^{p_n}\sum_{s=1}^{p_n} E\left(\varepsilon_s\varepsilon_r\right)E\left(\varepsilon_s\varepsilon_r\right) \\
&\quad= O\left(p_n^2 \sum_{i=1}^{n} G_{ii}^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}G_{ji}\right) \\
&\quad= O\left(p_n^2\frac{n}{h_n^2} + \|G_n\|^2 + \left|\mathrm{tr}\left(G_n^2\right)\right|\right) = O\left(\frac{n}{h_n^2}(p_n^2 + h_n)\right),
\end{aligned}$$

since $\|G_n\|^2$ and $\left|\mathrm{tr}\left(G_n^2\right)\right|$ are of order $O(n/h_n)$ by **Assumption 1**-ii. This concludes the proof.

■

Lemma 4.2. *Assume that $E(\varepsilon_i^4)$ is finite, where $\varepsilon_i = \int X(t)\varphi_i(t)dt$. Under **Assumption 1**, we have*

$$\frac{\xi_{p_n}' G_n' \xi_{p_n}}{n}\left[\frac{\xi_{p_n}'\xi_{p_n}}{n}\right]^{-1}\frac{\xi_{p_n}' G_n \xi_{p_n}}{n} - tr^2\left(\frac{G_n}{n}\right)\Gamma_{p_n} = O_p\left(\frac{p_n}{h_n^2\sqrt{n}}\left[1 + \frac{p_n^2}{\sqrt{n}}\right]\right).$$

Proof of Lemma 4.2

Note that

$$\begin{aligned}
&\frac{\xi_{p_n}^{'} G_n^{'} \xi_{p_n}}{n} \left[\frac{\xi_{p_n}^{'} \xi_{p_n}}{n}\right]^{-1} \frac{\xi_{p_n}^{'} G_n \xi_{p_n}}{n} - \mathrm{tr}^2\left(\frac{G_n}{n}\right)\Gamma_{p_n} \\
&= \left[\frac{\xi_{p_n}^{'} G_n^{'} \xi_{p_n}}{n} - \mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right]\left[\frac{\xi_{p_n}^{'} \xi_{p_n}}{n}\right]^{-1}\left[\frac{\xi_{p_n}^{'} G_n \xi_{p_n}}{n} - \mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right] \\
&\qquad + 2\mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\left[\frac{\xi_{p_n}^{'} \xi_{p_n}}{n}\right]^{-1}\left[\frac{\xi_{p_n}^{'} G_n \xi_{p_n}}{n} - \mathrm{tr}\left(\frac{G_n}{n}\right)\Gamma_{p_n}\right] \\
&\qquad + \mathrm{tr}^2\left(\frac{G_n}{n}\right)\Gamma_{p_n}\left[\frac{\xi_{p_n}^{'} \xi_{p_n}}{n}\right]^{-1}\left[\Gamma_{p_n} - \frac{\xi_{p_n}^{'} \xi_{p_n}}{n}\right] \\
&= O_p\left(\frac{p_n}{h_n^2\sqrt{n}}\left[1 + \frac{p_n^2}{\sqrt{n}}\right]\right),
\end{aligned}$$

by Lemma 4.1.

■

Lemma 4.3. *Under **Assumptions 1-2**, we have*

$$\sqrt{\frac{h_n}{n}}\mathbf{U}_n^{'} G_n^{'} M_n G_n \mathbf{R}_n = o_p(1), \tag{4.47}$$

$$\sqrt{\frac{h_n}{n}}\mathbf{R}_n^{'} M_n G_n \xi_{p_n} = o_p(1), \tag{4.48}$$

$$\sqrt{\frac{h_n}{n}}\mathbf{R}_n^{'} G_n^{'} M_n G_n \mathbf{R}_n = o_p(1). \tag{4.49}$$

Proof of Lemma 4.3

Let

$$\pi_{n1} = \sum_{r=1}^{p_n} E\left(R^2\varepsilon_r^2\right) \qquad \text{and} \qquad \pi_{n2} = \sum_{r=1}^{p_n} E\left(R\varepsilon_r\right)^2.$$

Consider (4.47), and note that by **Assumption 1**,

$$E\left(\left\|\mathbf{R}_n^{'} G_n \xi_{p_n}\right\|^2\right) = O\left(\frac{n}{h_n^2}\left[h_n E(R^2) + \pi_{n1} + n\pi_{n2}\right]\right), \tag{4.50}$$

$$E\left(\left\|\mathbf{R}_n^{'} \xi_{p_n}\right\|^2\right) = O\left(n\pi_{n1}\right), \text{ and } E\left(\left[\mathbf{R}_n^{'}\mathbf{U}_n\right]^2\right) = O\left(nE(R^2)\right). \tag{4.51}$$

Thus

$$\begin{aligned}
\mathbf{U}_n^{'} G_n^{'} M_n G_n \mathbf{R}_n &= \mathbf{U}_n^{'} G_n^{'} G_n \mathbf{R}_n - \mathbf{U}_n^{'} G_n^{'} \xi_{p_n}\left(\xi_{p_n}^{'}\xi_{p_n}\right)\xi_{p_n}^{'} G_n \mathbf{R}_n \\
&= o_p\left(\sqrt{\frac{n}{h_n}}\right) + O_p\left(\frac{p_n}{h_n}\sqrt{h_n E(R^2) + \pi_{n1} + n\pi_{n2}}\right),
\end{aligned}$$

by (4.24), (4.50), and (4.51).
Let us treat (4.48),

$$\begin{aligned}
\mathbf{R}_n^{'} G_n^{'} M_n G_n \xi_{p_n} &= \mathbf{R}_n^{'} G_n^{'} G_n \xi_{p_n} - \mathbf{R}_n^{'} G_n^{'} \xi_{p_n}\left(\xi_{p_n}^{'}\xi_{p_n}\right)\xi_{p_n}^{'} G_n \xi_{p_n} \\
&= O_p\left(\frac{\sqrt{n}}{h_n}\left[1 + \frac{p_n}{h_n}\right]\sqrt{h_n E(R^2) + \pi_{n1} + n\pi_{n2}}\right).
\end{aligned}$$

Finally, considering (4.49), we have

$$\begin{aligned}\mathbf{R}_n^{'}G_n^{'}M_nG_n\mathbf{R}_n &= \mathbf{R}_n^{'}G_n^{'}G_n\mathbf{R}_n - \mathbf{R}_n^{'}G_n^{'}\xi_{p_n}\left(\xi_{p_n}^{'}\xi_{p_n}\right)\xi_{p_n}^{'}G_n\mathbf{R}_n \\ &= O_p\left(\frac{p_n}{h_n^2}\left[h_nE(R^2)+\pi_{n1}+n\pi_{n2}\right]\right).\end{aligned}$$

Therefore the proof follows from **Assumption 2**.

■

Lemma 4.4. *Under **Assumptions 2** and **5**, we have*

$$\|\Psi_n - I_{p_n}\|_2 = O_p(p_n^{-1}).$$

For the proof of this lemma, see Müller & Stadtmüller (Lemma 7.2, 2005, p.28).

■

The following lemma gives conditions under which a martingale central limit theorem can be applicable to the triangular array of martingale difference sequences $\{Z_{ni},\, 1 \le i \le n,\, n \in \mathbb{N}\}$, for more of details see Kelejian & Prucha (Theorem A.1, 2001, p.240).

Lemma 4.5. *Under assumptions of Theorem 4.1, we have*

C.1. The random variables $\{Z_{ni},\, 1 \le i \le n,\, n \in \mathbb{N}\}$ form a triangular array of martingale difference sequence w.r.t the filtrations $(\mathcal{F}_{n,i}) = \sigma\left\{\varepsilon_r^{(j)}, U_j,\, 1 \le j \le i,\, 1 \le r \le p_n\right\}(1 \le i \le n,\, n \in \mathbb{N})$.

C.2. Conditional normalization condition:

$$\sum_{i=1}^{n} E\left(\tilde{Z}_{ni}^2 \middle| \mathcal{F}_{n,i-1}\right) \to 1, \qquad \text{in probability as } n \to \infty.$$

C.3. There exists a constant $\delta > 0$:

$$\sum_{i=1}^{n} E\left(\left|\tilde{Z}_{ni}\right|^{2+\delta}\right) \to 0, \qquad n \to \infty.$$

(Lyapunov condition if $\delta = 2$).

Proof of Lemma 4.5

Proof of C.1 This is immediate, because $E\left(Z_{ni}|\mathcal{F}_{n,i-1}\right) = 0$.

■

Proof of C.2
For each $i = 1, \ldots, n$, let

$$Q_{ni} = \sum_{j=1}^{i-1} G_{ij}^s T_j.$$

We have

$$E\left(Z_{ni}^2 \middle| \mathcal{F}_{n,i-1}\right) = \sigma_0^2 E(V^2) D_{ii}^2 + (\mu_4 - \sigma_0^4) C_{ii}^2 + 4\sigma_0^2 Q_{ni}^2,$$

hence

$$E\left(\sum_{i=1}^{n} E\left(Z_{ni}^2 \middle| \mathcal{F}_{n,i-1}\right)\right) = \sigma_0^2 E(V^2)\sum_{i=1}^{n} D_{ii}^2 + (\mu_4 - \sigma_0^4)\sum_{i=1}^{n} C_{ii}^2 + 2\sigma_0^2 E(T^2)\sum_{i=1}^{n}\sum_{j=1}^{i-1} G_{ij}^{s2}.$$

By definition of $\tilde{Z}_{ni}$,

$$E\left(\sum_{i=1}^{n} E\left(\tilde{Z}_{ni}^2 \middle| \mathcal{F}_{n,i-1}\right)\right) = 1 + o(1).$$

Remark that

$$\mathrm{Var}\left(\sum_{i=1}^{n} E\left(Z_{ni}^2 \middle| \mathcal{F}_{n,i-1}\right)\right) = 16\sigma_0^4 \mathrm{Var}\left(\sum_{i=1}^{n} Q_{ni}^2\right), \tag{4.52}$$

when U_i is normally distributed. Otherwise, result (4.55) remains valid.

Let us consider $\mathrm{Var}\left(\sum_{i=1}^{n} Q_{ni}^2\right)$. First, we have

$$\sum_{i=1}^{n} E\left(Q_{ni}^2\right) = E(T^2)\sum_{i=1}^{n}\sum_{j=1}^{i-1} G_{ij}^{s2}. \tag{4.53}$$

Let for all $1 \le i \le j \le n$,

$$\begin{aligned} E\left(Q_{ni}^2 Q_{nj}^2\right) &= \sum_{k_1,k_2=1}^{i-1}\sum_{r_1,r_2=1}^{j-1} G_{ik_1}^s G_{ik_2}^s G_{jr_1}^s G_{jr_2}^s E\left(T_{k_1}T_{k_2}T_{r_1}T_{r_2}\right) \\ &= \sum_{k_1,k_2=1}^{i-1}\sum_{r_1,r_2=1}^{i-1} G_{ik_1}^s G_{ik_2}^s G_{jr_1}^s G_{jr_2}^s E\left(T_{k_1}T_{k_2}T_{r_1}T_{r_2}\right) \\ &\quad + \left[\sum_{k_1,k_2=1}^{i-1} G_{ik_1}^s G_{ik_2}^s E\left(T_{k_1}T_{k_2}\right)\right] \times \left[\sum_{r_1,r_2=i}^{j-1} G_{jr_1}^s G_{jr_2}^s E\left(T_{r_1}T_{r_2}\right)\right] \\ &= E\left(T^4\right)\sum_{k=1}^{i-1} G_{ik}^{s2}G_{jk}^{s2} + E(T^2)^2\sum_{k=1}^{i-1}\sum_{r=i}^{j-1} G_{ik}^{s2}G_{jr}^{s2} \\ &\quad + E(T^2)^2\sum_{k\neq r=1}^{i-1}\left[G_{ik}^{s2}G_{jr}^{s2} + 2G_{ik}^s G_{ir}^s G_{jk}^s G_{jr}^s\right]. \end{aligned}$$

Then, we have

$$\begin{aligned} E\left(\left[\sum_{i=1}^{n} Q_{ni}^2\right]^2\right) &= E\left(T^4\right)\sum_{i=1}^{n}\sum_{k=1}^{i-1} G_{ik}^{s4} + 3E(T^2)^2\sum_{i=1}^{n}\sum_{k\neq r=1}^{i-1} G_{ik}^{s2}G_{ir}^{s2} \\ &\quad + 2E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1}\left[G_{ik}^{s2}G_{jr}^{s2} + 2G_{ik}^s G_{ir}^s G_{jk}^s G_{jr}^s\right] \\ &\quad + 2E\left(T^4\right)\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1} G_{ik}^{s2}G_{jk}^{s2} + 2E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}\sum_{r=i}^{j-1} G_{ik}^{s2}G_{jr}^{s2}. \end{aligned}$$

We can rewrite (4.53) as

$$\left[2E(T^2)^2\right]^{-1}\left[E\left(\sum_{i=1}^{n}Q_{ni}^2\right)\right]^2 = \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}G_{ik}^{s2}G_{jk}^{s2}+\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1}G_{ik}^{s2}G_{jr}^{s2} + \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}\sum_{r=i}^{j-1}G_{ik}^{s2}G_{jr}^{s2}.$$

Therefore, we have

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n}Q_{ni}^2\right) &= E\left(T^4\right)\sum_{i=1}^{n}\sum_{k=1}^{i-1}G_{ik}^{s4}+3E(T^2)^2\sum_{i=1}^{n}\sum_{k\neq r=1}^{i-1}G_{ik}^{s2}G_{ir}^{s2}\\ &\quad+\left[2E\left(T^4\right)-2E(T^2)^2\right]\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}G_{ik}^{s2}G_{jk}^{s2}\\ &\quad+4E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1}G_{ik}^{s}G_{ir}^{s}G_{jk}^{s}G_{jr}^{s}\\ &= O\left[\frac{n}{h_n^2}\left(E\left(T^4\right)+h_nE(T^2)^2\right)\right]. \end{aligned} \tag{4.54}$$

Then, by (4.52) and (4.54), we have

$$\mathrm{Var}\left(\sum_{i=1}^{n}E\left(\tilde{Z}_{ni}^2\middle|\mathcal{F}_{n,i-1}\right)\right)=O\left(\frac{E\left(T^4\right),+h_nE(T^2)^2}{n}\right)=o(1) \tag{4.55}$$

since $E(T^4)=O(E(V^4))=O(p_n^2)$ and $E(T^2)=O(E(V^2))=O(1)$. Hence the result follows.

■

Proof of C.3

For any positive constants p and q such that $\frac{1}{p}+\frac{1}{q}=1$,

$$\begin{aligned}|Z_{ni}| &\leq |D_{ii}||V_iU_i|+|C_{ii}||U_i^2-\sigma_0^2|+2|U_i|\sum_{j=1}^{i-1}|G_{ij}^s||T_j|\\ &\leq |D_{ii}|^{\frac{1}{p}}|D_{ii}|^{\frac{1}{q}}|V_iU_i|+|C_{ii}|^{\frac{1}{p}}|C_{ii}|^{\frac{1}{q}}|U_i^2-\sigma_0^2|\\ &\quad+\sum_{j=1}^{i-1}|G_{ij}^s|^{\frac{1}{p}}|G_{ij}^s|^{\frac{1}{q}}2|T_j||U_i|.\end{aligned}$$

Holder's inequality for inner products applied to the last term, implies that

$$\begin{aligned}|Z_{ni}|^q &\leq \left\{\left[(|D_{ii}|^{\frac{1}{p}})^p+(|C_{ii}|^{\frac{1}{p}})^p+\sum_{j=1}^{i-1}(|G_{ij}^s|^{\frac{1}{p}})^p\right]^{\frac{1}{p}}\left[(|D_{ii}|^{\frac{1}{q}}|V_iU_i|)^q\right.\right.\\ &\quad\left.\left.+(|C_{ii}|^{\frac{1}{q}}|U_i^2-\sigma_0^2|)^q+\sum_{j=1}^{i-1}(|G_{ij}^s|^{\frac{1}{q}}2|T_j||U_i|)^q\right]^{\frac{1}{q}}\right\}^q\\ &= \left[|D_{ii}|+|C_{ii}|+\sum_{j=1}^{i-1}|G_{ij}^s|\right]^{\frac{q}{p}}\left[|D_{ii}||V_iU_i|^q+|C_{ii}||U_i^2-\sigma_0^2|^q+2^q|U_i|^q\sum_{j=1}^{i-1}|G_{ij}^s||T_j|^q\right]\\ &= O\left(|D_{ii}||V_iU_i|^q+|C_{ii}||U_i^2-\sigma_0^2|^q+2^q|U_i|^q\sum_{j=1}^{i-1}|G_{ij}^s||T_j|^q\right)\end{aligned}$$

since under **Assumption 1**, D_{ii} and C_{ii} are of order $O(1/h_n)$ and G_n is uniformly bounded in row sums.
Let $q = 2 + \delta$, and note that

$$\sum_{i=1}^{n} E\left(\left|\widetilde{Z}_{ni}\right|^{2+\delta}\right) = O\left(\frac{h_n^{\frac{\delta}{2}}}{n^{\frac{\delta}{2}}}\left[E\left(U^{4+2\delta}\right) + h_n E\left(|T|^{2+\delta}\right)\right]\right). \tag{4.56}$$

Let $\delta = 2$, then (4.56) is of order $O\left(\frac{h_n^2 p_n^2}{n}\right)$, since $E(T^4) = O\left(p_n^2\right)$ and $E(U^8)$ is finite. This yields the proof as by assumption $h_n^4 = O(n)$ (when h_n is divergent) and $p_n^4 = o(n)$.

■

Chapter 5

Spatial prediction by k-NN method k-nearest neighbor kernel method

Contents

Résumé en français

Dans ce chapitre, nous proposons un prédicteur spatial non-paramétrique à partir de la fonction de régression d'un processus spatial par la méthode des k-plus proches voisins. La spécificité du prédicteur proposé est qu'elle utilise une fenêtre de lissage aléatoire adaptée à une éventuelle hétérogénéité au niveau des réalisations de la variable explicative spatiale utilisée. L'approche proposée dans ce chapitre généralise celle classique de k-plus proches voisins (Collomb, 1980) aux données spatiales et est une alternative à l'approche par noyau étudiée dans Dabo-Niang et al. (2016).

Nous considérons le processus spatial $\{Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}) \in \mathbb{R}^d \times \mathbb{R}\ ,\ \mathbf{i} \in \mathbb{N}^N\}$ $(d \geq 1)$ définit sur l'espace de probabilité $(\Omega, \mathcal{A}, P)$, $N \in \mathbb{N}^*$. On suppose que ce processus est observable sur l'ensemble spatial discret $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} \in \mathbb{N}^N \ : \ 1 \leq i_r \leq n_r \ r = 1, \ldots, N\}$ avec $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{N}^N$ et $\hat{\mathbf{n}} = n_1 \times \ldots \times n_N$. On suppose que $\mathbf{n} \to \infty$ équivaut à $\min\{n_r\} \to +\infty$ et $n_k/n_i \leq C$, pour $1 \leq k, i \leq N$ où C est une constante positive. Nous admettrons que la relation entre les processus $(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ et $(Y_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$, soit définie par le modèle de régression suivant

$$Y_{\mathbf{i}} = r(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N, \tag{5.1}$$

où la fonction de régression $r(\cdot) = E\left(Y_{\mathbf{i}} | X_{\mathbf{i}} = \cdot\right)$ est supposée indépendante de $\mathbf{i}$, le bruit $(\varepsilon_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ est centré, $\alpha-$mélangeant et indépendant de $(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$.

Nous nous intéressons à prédire le processus $(Y_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^N)$ en des sites non observés, particulièrement en un site $\mathbf{s_0} \in \mathcal{I}_{\mathbf{n}}$ en exploitant l'information sur $X_{\mathbf{s_0}}$ et les observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}}$, où $\mathcal{O}_{\mathbf{n}} \subset \mathcal{I}_{\mathbf{n}}$ est l'ensemble spatial dans lequel le processus $(Z_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^N)$ est observé, avec $\mathbf{s_0} \notin \mathcal{O}_{\mathbf{n}}$ et $\text{Card}(\mathcal{O}_{\mathbf{n}})$ tend vers l'infini quand $\mathbf{n} \to \infty$. Pour atteindre cet objectif, la littérature suppose généralement que le processus considéré est strictement stationnaire. Dans ce travail, nous supposons que les variables $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}}$ sont localement identiquement distribuées (voir Klemelä, 2008). Avec cette hypothèse, nous pouvons imaginer que s'il existe assez de sites dans $\mathcal{O}_{\mathbf{n}}$ proches de $\mathbf{s_0}$ où des données sont disponibles, alors les observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}}$ pourront être utilisées pour prédire $Y_{\mathbf{s_0}}$. Nous supposons que $(Y_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ est intégrable et que $(X_{\mathbf{s_0}}, Y_{\mathbf{s_0}})$ a la même distribution que (X, Y) et que (X, Y) et $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ admettent des densités inconnues par rapport à la mesure de Lebesgue. Soient f et $f_{X,Y}$, les densités respectives de X et (X, Y).

Nous définissons le prédicteur de $Y_{\mathbf{s_0}}$ en combinant le principe de la méthode des k-plus proches voisins et le prédicteur spatiale proposé par Dabo-Niang et al. (2016) :

$$\widehat{Y}_{\mathbf{s_0}} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{X_{\mathbf{s_0}}-X_{\mathbf{i}}}{H_{\mathbf{n},X_{\mathbf{s_0}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{X_{\mathbf{s_0}}-X_{\mathbf{i}}}{H_{\mathbf{n},X_{\mathbf{s_0}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}, \tag{5.2}$$

si le dénominateur est non nul, sinon $\widehat{Y}_{\mathbf{s_0}}$ est la moyenne empirique. Notons que, K_1 et K_2 sont deux noyaux de $\mathbb{R}^d$ à $\mathbb{R}_+$ et de $\mathbb{R}$ à $\mathbb{R}_+$ respectivement, $\dfrac{\mathbf{i}}{\mathbf{n}} = \left(\dfrac{i_1}{n_1}, \ldots, \dfrac{i_N}{n_N}\right)$,

$H_{\mathbf{n},X_{\mathbf{s_0}}} = \min\left\{h \in \mathbb{R}_+^* \mid \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} \mathbb{I}\left(\|X_{\mathbf{i}} - X_{\mathbf{s_0}}\| < h\right) = k(\mathbf{n})\right\}$ et

$h_{\mathbf{n},\mathbf{s_0}} = \min\left\{h \in \mathbb{R}_+^* \mid \sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} \mathbb{I}\left(\left\|\frac{\mathbf{i}-\mathbf{s_0}}{\mathbf{n}}\right\| < h\right) = k_{\mathbf{n}}^1\right\}$ où $k_{\mathbf{n}}^1$, $k(\mathbf{n})$ sont deux suites d'entiers positives. Notons que la fenêtre aléatoire de lissage $H_{\mathbf{n},X_{\mathbf{s_0}}}$ est une variable aléatoire positive qui dépend de $X_{\mathbf{s_0}}$ et des observations $\{X_{\mathbf{i}},\ \mathbf{i} \in \mathcal{O}_{\mathbf{n}}\}$.
Le principal avantage de ce prédicteur comparé à celui par noyau

$$\widehat{Y}_{\mathbf{s_0}}^{NW} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right) K_2\left(\frac{X_{\mathbf{s_0}}-X_{\mathbf{i}}}{h_{\mathbf{n}}}\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\rho_{\mathbf{n}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right) K_2\left(\frac{X_{\mathbf{s_0}}-X_{\mathbf{i}}}{h_{\mathbf{n}}}\right)}, \tag{5.3}$$

où la fenêtre de lissage $h_{\mathbf{n}}$ (voir Dabo-Niang et al., 2016) est déterministe, réside sur le fait que $H_{\mathbf{n},X_{\mathbf{s_0}}}$ permet au prédicteur de s'adapter à la structure locale des observations, particulièrement s'il existe une certaine hétérogénéité des données. De plus le prédicteur proposé $\widehat{Y}_{\mathbf{s_0}}$ est plus facile à mettre en ?uvre que celui par noyau $\widehat{Y}_{\mathbf{s_0}}^{NW}$. En effet, il est plus facile de choisir le paramètre du nombre de voisins $k(\mathbf{n})$ qui prend ses valeurs dans un sous-ensemble discret que la fenêtre $h_{\mathbf{n}}$ qui est dans $\mathbb{R}_+$.

Sous certaines hypothèses, nous obtenons des résultats de la consistance du prédicteur proposé. Plus particulièrement, la convergence presque complète est obtenue avec vitesse. Nous montrons que

$$\left|\widehat{Y}_{\mathbf{s_0}} - Y_{\mathbf{s_0}}\right| \underset{\mathbf{n}\to\infty}{\longrightarrow} 0 \quad a.co.$$

Une étude numérique où nous comparons les performances du prédicteur proposé $\widehat{Y}_{\mathbf{s_0}}$ et celui par noyau; $\widehat{Y}_{\mathbf{s_0}}^{NW}$ de Dabo-Niang et al. (2016) est menée sur des données simulées et réelles.

The results of this chapter are in collaboration with Mohamed Kadi Attouch (University of Djilalli Liabes, Algeria), Sophie Dabo-Niang (University of Lille) and Mamadou N'diaye (University of Dakar, Senegal). A related paper is in revision.

5.1 Introduction

Spatio-temporal data naturally arise in many fields such as environmental sciences, geophysics, soil science, oceanography, econometrics, epidemiology, forestry, image processing and many others in which the data of interest are collected across space. The literature on spatio-temporal models is relatively abundant, see for example the monograph of Cressie & Wikle (2015).

Complex issues arise in spatial analysis, many of which are neither clearly defined nor completely resolved, but form the basis for current researches. Among the practical considerations that influence the available techniques used in the spatial data modeling, is the data dependency. In fact, spatial data are often dependent and a spatial model must be able to handle this aspect. Notice that the linear models for spatial data only capture global linear relationships between spatial locations. However, in many circumstances the spatial dependency is not linear. It is for example, the classical case where one deals with the spatial pattern of extreme events such as in the economic analysis of poverty.

Then in such situations, it is more appropriate to use a nonlinear spatial dependence measure by using for instance the strong mixing coefficients concept (see Tran, 1990). The literature on nonparametric estimation techniques, which incorporate nonlinear spatial dependency is not extensive compare to that of linear dependence. For an overview on results and applications considering spatial dependent data for density, regression estimation or prediction, we highlight the following works: Lu & Chen (2004), Hallin et al. (2004), Biau & Cadre (2004), Carbon et al. (2007), Dabo-Niang & Yao (2007), Menezes et al. (2010), El Machkouri & Stoica (2010), Wang & Wang (2009), Ternynck (2014). Other authors deal with the spatial quantile regression estimation Hallin et al. (2009), Abdi et al. (2010) and Dabo-Niang et al. (2012).

The k-Nearest Neighbor (k-NN) kernel estimator is a weighted average of response variables in the neighborhood of the value of covariate. The k-NN kernel estimate has a significant advantage over the classical kernel estimate. The specificity of the k-NN estimator liens in the fact that it is flexible to all sort of presence of heterogeneity in used covariate which allows to account the local structure of the data. This consists in the choice of an appropriate number of neighbors, using a random bandwidth adapted to the local structure of the data and permitting to learn more on the local data dependency. Another advantage of the k-NN method is in the nature of the smoothing parameter. Indeed, in the classical kernel method, the smoothing parameter is the bandwidth h_n, which is a real positive number and in k-NN method, the smoothing parameter takes k_n its values in discrete set. The use of this method is very recent in the case of spatial data. Li & Tran (2009) proposed a regression estimator of spatial data based on the k nearest neighbors method. They proved an asymptotic normality result of their estimator in the case of multivariate data.

The lack of spatio-nonparametric techniques motivates this work. Namely, we are interested in asymptotic properties of nonparametric prediction for spatial processes using k-nearest neighbors method. The originality of the suggested present predictor lies in the fact that it depends on two kernels, one of which controls the distance between observations using random bandwidth and the other controls the spatial dependence structure. This idea has been presented in Menezes et al. (2010), Dabo-Niang et al. (2016), Ternynck (2014) in the context of kernel prediction problem for multivariate or functional spatial

data.

The present work extends the previous results in the case of k-nearest neighbors non-parametric prediction in the context of multivariate spatial data. We derive a double nearest neighbors selection method of the classical k-nearest neighbors one (see Li & Tran, 2009) and we study some asymptotic results of such predictor and give some numerical results.

The outline of the rest of this chapter is as follows. In Section 5.2, we introduce the model and define the predictor. Section 5.3 is dedicated to the almost complete convergence[1] whereas Section 5.4 gives some simulations and application to real data, to illustrate the performance of the proposed predictor. Section 5.5 is devoted to some conclusions. Finally, the proofs of some lemmas and the main results are postponed to the last Section.

5.2 Model and construction of predictor

Let $\{Z_{\mathbf{i}} = (X_{\mathbf{i}}, Y_{\mathbf{i}}) \in \mathbb{R}^d \times \mathbb{R}\ ,\ \mathbf{i} \in \mathbb{N}^N\}$ $(d \geq 1)$ be a spatial process defined over some probability space $(\Omega, \mathcal{A}, P)$, $N \in \mathbb{N}^*$. We assume that the process is observable in $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} \in \mathbb{N}^N\ :\ 1 \leq i_r \leq n_r\ r = 1, \ldots, N\}$, $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{N}^N$, and $\hat{\mathbf{n}} = n_1 \times \ldots \times n_N$, we write $\mathbf{n} \to \infty$ if $\min\{n_r\} \to +\infty$, $n_k/n_i \leq C,\ \forall\ 1 \leq k, i \leq N$. Let $\|\cdot\|$ denote the Euclidian norm in $\mathbb{R}^N$ or in $\mathbb{R}^d$. We assume that the relation between these two process $(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ and $(Y_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ is described by this following model:

$$Y_{\mathbf{i}} = r(X_{\mathbf{i}}) + \varepsilon_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N \tag{5.4}$$

where

$$r(\cdot) = E\left(Y_{\mathbf{i}} | X_{\mathbf{i}} = \cdot\right) \tag{5.5}$$

is assumed to be independent of $\mathbf{i}$, the noise $(\epsilon_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ is centered, α-mixing (see Section 5.3 for a description of this condition) and independent of $(X_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$. We are interested in predicting the spatial process $(Y_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^N)$ in some unobserved locations and particularly at an unobserved site $\mathbf{s_0} \in \mathcal{I}_{\mathbf{n}}$ under the information that can be drawn on $X_{\mathbf{s_0}}$ and observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{O}_{\mathbf{n}}}$, where $\mathcal{O}_{\mathbf{n}}$ is the observed spatial set of finite cardinality tending to ∞ as $\mathbf{n} \to \infty$ and contained in $\mathcal{I}_{\mathbf{n}}$, with $\mathbf{s_0} \notin \mathcal{O}_{\mathbf{n}}$. In the following proposed predictor, we integrate information that might be drawn from the structure of the spatial dependence between the considered site $\mathbf{s_0}$ and all sites in $\mathcal{O}_{\mathbf{n}}$. To achieve this objective, we do not suppose as usual a strict stationarity assumption. We assume that the observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{O}_{\mathbf{n}}}$ are locally identically distributed (given in assumption (H7), see Dabo-Niang et al., 2016; Klemelä, 2008, for more detail). Indeed, we say that a substantial number of observations $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ has a distribution close to that of $(X_{\mathbf{s_0}}, Y_{\mathbf{s_0}})$. In such case, one may imagine that if there is enough sites $\mathbf{i}$ closed to $\mathbf{s_0}$, then sequence $(X_{\mathbf{i}}, Y_{\mathbf{i}})_{\mathbf{i} \in \mathcal{O}_{\mathbf{n}}}$ may be used to predict $Y_{\mathbf{s_0}}$. Assume that $(Y_{\mathbf{i}},\ \mathbf{i} \in \mathbb{N}^N)$ is integrable and that $(X_{\mathbf{s_0}}, Y_{\mathbf{s_0}})$ has the same distribution as that of some pair (X, Y). We assume that (X, Y) and $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ have unknown continuous densities with respect to Lebesgue measure and let $f_{X,Y}$ and f be the densities of (X, Y) and X respectively.

A predictor of $Y_{\mathbf{s_0}}$ could be defined by combining the principle of k-NN method using a

[1] Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real random variables. We say that z_n converges almost completely (a.co.) toward zero if, and only if, $\forall \epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of the almost complete convergence of z_n to zero is of order u_n (with $u_n \to 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists \epsilon > 0$ such that $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

random bandwidth depending on the observations and the kernel weight (see Dabo-Niang et al., 2016), as follows:

$$\widehat{Y}_{\mathbf{s_0}} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} Y_\mathbf{i} K_1\left(\frac{X_{\mathbf{s_0}}-X_\mathbf{i}}{H_{\mathbf{n},X_{\mathbf{s_0}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} K_1\left(\frac{X_{\mathbf{s_0}}-X_\mathbf{i}}{H_{\mathbf{n},X_{\mathbf{s_0}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}, \tag{5.6}$$

if the denominator is not null otherwise the predictor is equal to the empirical mean. Here, K_1 and K_2 are two kernels from $\mathbb{R}^d$ and $\mathbb{R}$ to $\mathbb{R}_+$ respectively, $\frac{\mathbf{i}}{\mathbf{n}} = \left(\frac{i_1}{n_1},\cdots,\frac{i_N}{n_N}\right)$, and $H_{\mathbf{n},X_{\mathbf{s_0}}} = \min\left\{h\in\mathbb{R}_+^* \mid \sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} \mathbb{I}(\|X_\mathbf{i}-X_{\mathbf{s_0}}\|<h) = k(\mathbf{n})\right\}$ and

$h_{\mathbf{n},\mathbf{s_0}} = \min\left\{h\in\mathbb{R}_+^* \mid \sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} \mathbb{I}\left(\left\|\frac{\mathbf{i}-\mathbf{s_0}}{\mathbf{n}}\right\|<h\right) = k_\mathbf{n}^1\right\}$ where $k_\mathbf{n}^1$, $k(\mathbf{n})$ are positive integers sequences. The random bandwidth $H_{\mathbf{n},X_{\mathbf{s_0}}}$ is a positive random variable which depends on $X_{\mathbf{s_0}}$ and the observations $\{X_\mathbf{i},\, \mathbf{i}\in\mathcal{O}_\mathbf{n}\}$.

The main advantage of using this predictor compare to the fully kernel method may be the fact that $H_{\mathbf{n},X_{\mathbf{s_0}}}$ depends on $X_{\mathbf{s_0}}$ allows the predictor to be adapted to a local structure of the observations, particularly if these are heterogeneous (see Burba et al., 2009). In addition, the k-NN method is easy to be computed, in fact it is easier to choose the smoothing parameters $k_\mathbf{n}^1$ and $k(\mathbf{n})$ which take their values in a discrete subset than the bandwidths used in the following kernel counterpart of (5.6) (see Dabo-Niang et al., 2016).

$$\widehat{Y}_{\mathbf{s_0}}^{NW} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} Y_\mathbf{i} K_1\left(\frac{X_{\mathbf{s_0}}-X_\mathbf{i}}{h_\mathbf{n}}\right) K_2\left(\rho_\mathbf{n}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{O}_\mathbf{n}} K_1\left(\frac{X_{\mathbf{s_0}}-X_\mathbf{i}}{h_\mathbf{n}}\right) K_2\left(\rho_\mathbf{n}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}, \tag{5.7}$$

the bandwidths $h_\mathbf{n}$, $\rho_\mathbf{n}$ are non random.

5.3 Assumptions and results

To account for spatial dependency, we assume that the process $\{Z_\mathbf{i} = (X_\mathbf{i}, Y_\mathbf{i}) \in \mathbb{R}^d\times\mathbb{R}\ ,\ \mathbf{i}\in\mathbb{N}^N\}$ satisfies a mixing condition defined as follows: there exists a function $\varphi(x)\searrow 0$ as $x\to\infty$, such that

$$\begin{aligned}\alpha\left(\sigma\left(S\right),\sigma\left(S'\right)\right) &= \sup\left\{\left|P(A\cap B)-P\left(A\right)P\left(B\right)\right|,\quad A\in\sigma\left(S\right),\quad B\in\sigma\left(S'\right)\right\}\\ &\leq \psi\left(\mathrm{Card}(S),\mathrm{Card}(S')\right)\varphi\left(\mathrm{dist}(S,S')\right)\end{aligned} \tag{5.8}$$

where S and S' are two finite sets of sites, $\sigma\left(S\right)=\{Z_\mathbf{i},\mathbf{i}\in S\}$ and $\sigma\left(S'\right)=\{Z_\mathbf{i},\mathbf{i}\in S'\}$ are σ-fields generated by the $Z_\mathbf{i}$'s, and $\psi(\cdot)$ is a positive symmetric function nondecreasing in each variable. We recall that the process is said to be strongly mixing if $\psi(\cdot)=1$ (see Doukhan, 1994). As usual, we will assume that $\varphi(i)$ verifies :

$$\varphi(t)\leq Ct^{-\theta}\qquad,\ \theta>0\,,\ t\in\mathbb{R}_+^*. \tag{5.9}$$

for some $C>0$ (i.e. $\varphi\left(t\right)$ tends to zero at a polynomial rate).
Before stating the main results, the following set of assumptions are listed and all along the chapter, we fix a compact subset D in $\mathbb{R}^d$ and when no confusion is possible, we will denote by C, a strictly positive generic constant.

(H1) f and $r(\cdot)$ are continuous Lipschitz functions in D. In addition, $\inf_{x\in D} f(x) > 0$.

(H2) The density $f_{X_{\mathbf{i}}X_{\mathbf{j}}}$ of $(X_{\mathbf{i}}, X_{\mathbf{j}})$ is bounded in D and $\left|f_{X_{\mathbf{i}}X_{\mathbf{j}}}(u,v) - f_{X_{\mathbf{i}}}(u)f_{X_{\mathbf{j}}}(v)\right| \leq C$ for all $\mathbf{i} \neq \mathbf{j}$ and $(u,v) \in D \times D$.

(H3) $k(\mathbf{n}) \sim \hat{\mathbf{n}}^{\gamma}$ and $k^1_{\mathbf{n}} \sim \hat{\mathbf{n}}^{\tilde{\gamma}}$, where $\gamma,\ \tilde{\gamma} \in]\frac{1}{2}, 1[$, $\gamma < \tilde{\gamma}$ and $\gamma + \tilde{\gamma} > 3/2$.

(H4) (i) The kernel K_1 is bounded, of compact support and

$$\forall u \in \mathbb{R}^d,\ K_1(u) \leq K_1(tu) \quad \forall\, t \in]0,1[. \tag{5.10}$$

(ii) K_2 is a bounded nonnegative function, and there exist constants C_1, C_2 and ρ such that

$$C_1 \mathbb{I}\,(t \leq \rho) \leq K_2(t) \leq C_2 \mathbb{I}\,(t \leq \rho) \qquad \forall\, t \in \mathbb{R}_+, \tag{5.11}$$

with $0 < C_1 \leq C_2 < \infty, \rho > 0$.

(H5) $\forall n, m \in \mathbb{N} \quad \psi(n,m) \leq C \min(n,m)$ and
$\theta > N(sd(3-\gamma-\tilde{\gamma}) + 2s(3-\gamma) + 2d)/(1 - s(2-\gamma-\tilde{\gamma}))$ where $2 < s < 1/(2-\gamma-\tilde{\gamma})$.

(H6) $\forall n, m \in \mathbb{N} \quad \psi(n,m) \leq C(n+m+1)^{\tilde{\beta}}$, $\tilde{\beta} \geq 1$ and
$\theta > N\left(s(d(3-\gamma-\tilde{\gamma}) + (7 + 2\tilde{\beta} - 3\gamma - \tilde{\gamma})) + 2(d+1)\right)/(1 - s(2-\gamma-\tilde{\gamma}))$ where $2 < s < 1/(2-\gamma-\tilde{\gamma})$,

(H7) The densities $f_{\mathbf{i}}$ and $f_{X_{\mathbf{i}},Y_{\mathbf{i}}}$ of $X_{\mathbf{i}}$ and $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ are such that

$$\sup_{x\in D, \|\frac{\mathbf{i}-\mathbf{s_0}}{\mathbf{n}}\| < h_{\mathbf{n},\mathbf{s_0}}} |f_{\mathbf{i}}(x) - f(x)| = o(1) \qquad \text{and} \qquad \sup_{x\in D, \|\frac{\mathbf{i}-\mathbf{s_0}}{\mathbf{n}}\| < h_{\mathbf{n},\mathbf{s_0}}} |g_{\mathbf{i}}(x) - g(x)| = o(1),$$

as $\mathbf{n} \to \infty$ with $g_{\mathbf{i}}(x) = \int y f_{X_{\mathbf{i}},Y_{\mathbf{i}}}(x,y)dy$.
The conditional density $f_{Y_{\mathbf{i}},Y_{\mathbf{j}}|X_{\mathbf{i}},X_{\mathbf{j}}}$ of $(Y_{\mathbf{i}}, Y_{\mathbf{j}})$ given $(X_{\mathbf{i}}, X_{\mathbf{j}})$ and the conditional density $f_{Y_{\mathbf{i}}|X_{\mathbf{j}}}$ of $Y_{\mathbf{i}}$ given $X_{\mathbf{j}}$ exists and

$$f_{Y_{\mathbf{i}},Y_{\mathbf{j}}|X_{\mathbf{i}},X_{\mathbf{j}}}(y,t|u,v) < C \qquad \text{and} \qquad f_{Y_{\mathbf{i}}|X_{\mathbf{j}}}(y|u) < C,$$

for all $y, t, u, v, \mathbf{i}, \mathbf{j}$.

Remark 5.1.

1. *More generally one can extend* $\widehat{Y}_{\mathbf{s_0}}$ *using* $K_2\left(h^{-1}_{\mathbf{n},\mathbf{s}}\left(\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right)\right)$ *(instead of* $K_2\left(h^{-1}_{\mathbf{n},\mathbf{s_0}}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)$*) where sites* $\mathbf{i}$ *and* $\mathbf{s_0}$ *are not normalized and* $K_2(\cdot)$ *is a kernel on* $\mathbb{R}^N$*, that is*

$$\widehat{Y}_{\mathbf{s_0}} = \frac{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} Y_{\mathbf{i}} K_1\left(\frac{X_{\mathbf{s_0}} - X_{\mathbf{i}}}{H_{\mathbf{n},\mathbf{s_0}}}\right) K_2\left(h^{-1}_{\mathbf{n},\mathbf{s}}\left(\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right)\right)}{\sum_{\mathbf{i}\in\mathcal{O}_{\mathbf{n}}} K_1\left(\frac{X_{\mathbf{s_0}} - X_{\mathbf{i}}}{H_{\mathbf{n},\mathbf{s_0}}}\right) K_2\left(h^{-1}_{\mathbf{n},\mathbf{s}}\left(\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right)\right)}. \tag{5.12}$$

2. *In assumption (H1), we assume that* f *is lipschitzian, this is necessary particularly in the bias term (see the proof of condition* $(L1)$ *in Lemmas 5.3 and 5.4) and it allows with assumption (H7) to specify the rate of convergence in Corollary 5.1.*

3. *Assumption (H2) is very basic in nonparametric estimation of spatial dependent data, it allows to control the local dependence between the observations (see, e.g Carbon et al., 2007).*

4. *The condition on $k(\mathbf{n})$ in assumption (H3) is similar of the condition of number of the nearest neighbors assumed by Muller & Dippon (2011) in the case of dependent functional data. The condition on $k_{\mathbf{n}}^1$ is same as the condition assumed by Dabo-Niang et al. (2016) on the number of the neighbors of the predict site.*

5. *Condition (5.10) on the kernel K_1 is required in the proofs of Lemma 5.3 and Lemma 5.4, for more details on this kernel, see Collomb (1980). Condition (5.11) on the kernel K_2 allows the simplicity and brevity of our proofs (see Dabo-Niang et al., 2016). It is satisfied, for instance, by several kernels with compact support such as triangular (Bartlett), biweight, triweight, Epanechnikov, Parzen kernels.*

6. *Assumptions (H5) and (H6) are very standard to handle the strong mixing dependence (seee Neaderhouser, 1980; Rosenblatt, 1985; Takahata, 1983; Dabo-Niang et al., 2016). They appear (in the calculations when studying the asymptotic behavior of the estimator) in the particular case where the mixing coefficient is such that θ tends to zero at a polynomial rate (see Neaderhouser, 1980; Rosenblatt, 1985, for some examples). Each of these conditions is related to a specific case of mixing in the spatial context and are used respectively in Neaderhouser (1980) and Takahata (1983).*

The following theorem gives an almost complete convergence of the predictor.

Theorem 5.1. *Under assumptions (H1)-(H4), (H7) and (H5) or (H6), as $\mathbf{n} \to \infty$, we have*

$$\widehat{Y}_{\mathbf{s_0}} - Y_{\mathbf{s_0}} = o(1) \quad a.co. \tag{5.13}$$

If $r(\cdot)$ is lipschitzian we can obtain the rate of almost complete convergence stated in the following Corollary.

Corollary 5.1. *Under assumptions (H1)-(H4), (H7) and (H5) or (H6), as $\mathbf{n} \to \infty$, we have*

$$\widehat{Y}_{\mathbf{s_0}} - Y_{\mathbf{s_0}} = O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d} + \left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^1 k(\mathbf{n})}\right)^{1/2}\right) \quad a.co. \tag{5.14}$$

The results of Theorem 5.1 and Corollary 5.1 can be proved easily from the asymptotic results (stated respectively in Lemmas 5.1 and 5.2) of the following function

$$r_{\mathrm{kNN}}(x) = \begin{cases} \dfrac{g_{\mathbf{n}}(x)}{f_{\mathbf{n}}(x)} & \text{if } f_{\mathbf{n}}(x) \neq 0; \\ \overline{Y}, & \text{the empirical mean, otherwise,} \end{cases}$$

with

$$g_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n},\mathbf{s_0}}^N H_{\mathbf{n},x}^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x - X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right) Y_{\mathbf{i}},$$

and

$$f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n},\mathbf{s}}^N H_{\mathbf{n},x}^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x - X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right).$$

Lemma 5.1. *Under assumptions (H1)-(H4), (H7) and (H5) or (H6), we have*

$$\sup_{x\in D} |r_{\mathrm{kNN}}(x) - r(x)| \underset{\mathbf{n}\to\infty}{\longrightarrow} 0 \quad a.co. \tag{5.15}$$

Lemma 5.2. *Under assumptions (H1)-(H4), (H7) and (H5) or (H6), as* $\mathbf{n} \to \infty$*, we have*

$$\sup_{x\in D} |r_{\mathrm{kNN}}(x) - r(x)| = O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d} + \left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^1 k(\mathbf{n})}\right)^{1/2}\right) \quad a.co. \tag{5.16}$$

The proofs will be postponed in the last section. Since the proofs of Theorem 5.1 and Corollary 5.1 come directly from that of Lemmas 5.1 and 5.2, they will be omitted. The main difficulty in the proofs of Lemmas 5.1 and 5.2, comes from randomness of the window $H_{\mathbf{n},x}$. Then, we do not have in the numerator and denominator of $r_{\mathrm{kNN}}(x)$ sums of identically distributed variables. The idea is to frame sensibly $H_{\mathbf{n},x}$ by two non-random bandwidths.

Now, that we have checked the theoretical behavior of our predictor, we study the practical features through some simulations as well as an application to a multivariate soil data set related to heavy metal contamination in the Swiss Jura.

5.4 Numerical experiments

5.4.1 Simulation dataset

In order to evaluate the efficiency of the k-NN prediction for a set of spatial data, we use the average of mean absolute errors (MAE) to compare the prediction by k-NN method and that by kernel of Dabo-Niang et al. (2016) using simulated data based on observations $(X_{i,j}, Y_{i,j})$, $1 \le i \le n_1, 1 \le j \le n_2$ such that $\forall i, j$:

$$X_{i,j} = A_{i,j} \times U_{i,j} \times T_{i,j} + (1 - A_{i,j}) \times (6 + U_{i,j} \times Z_{i,j})$$

and

$$Y_{i,j} = r(X_{i,j}) + \varepsilon_{i,j},$$

where

$$r(x) = x^2.$$

Let the $A_{i,j}$ be independent Bernoulli random variables with parameter 0.5, $T = (T_{i,j})_{1\le i\le n_1, 1\le j\le n_2} = \mathrm{GRF}(0, 5, 3)$, $Z = (Z_{i,j})_{1\le i\le n_1, 1\le j\le n_2} = \mathrm{GRF}(0, \sigma, 3)$ and $\varepsilon = (\varepsilon_{i,j})_{1\le i\le n_1, 1\le j\le n_2} = \mathrm{GRF}(0, 0.1, 3)$, where we denote by $\mathrm{GRF}(\mu, \sigma^2, s)$ a stationary Gaussian random field with mean μ and covariance function defined by $C(h) = \sigma^2 \exp\left(-(\|h\|/s)^2\right)$, $h \in \mathbb{R}^2, s > 0, \sigma > 0$. The process $U = (U_{i,j})_{1\le i\le n_1, 1\le j\le n_2}$ allows to control the local dependence between the sites and is:

$$U_{i,j} = \frac{1}{n_1 \times n_2} \sum_{t,m} \exp\left(-\|(i,j) - (m,t)\|/a\right), \; a > 0,$$

the greater a is, weaker is the spatial dependency. Accordingly, we provide simulation results obtained with different values of a; $a = 5$, 10, 20, different grid sizes $n_1 = 25$, $n_2 = 25$ and $n_1 = 35$, $n_2 = 30$ and two variance parameters σ^2 ($\sigma = 5$ and 0.1). The model is replicated 50 times. We take kernels

$$K_1(x) = 0.75(1 - x^2)\mathbb{I}(|x| < 1),$$

and

$$K_2(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{if } |x| < 0.5; \\ 2(1 - |x|)^3 & \text{if } 0.5 \le |x| \le 1; \\ 0 & \text{otherwise,} \end{cases}$$

satisfying assumption (H4). The smoothing parameters are computed using the cross-validation procedure as used in Dabo-Niang et al. (2016) using the mean absolute error

$$\text{MAE} = \frac{1}{n_1 \times n_2} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \left| Y_{\mathbf{i}} - \widehat{Y}_{\mathbf{i}} \right| \text{ with } \widehat{Y}_{\mathbf{i}} = \tilde{Y}_{\mathbf{i}} \text{ or } \tilde{Y}_{\mathbf{i}}^{NW},$$

where

$$\tilde{Y}_{\mathbf{i}} = \frac{\sum_{\mathbf{j} \neq \mathbf{i}} K_1\left(\frac{X_{\mathbf{i}} - X_{\mathbf{j}}}{H_{\mathbf{n},X_{\mathbf{i}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{i}}^{-1} \left\|\frac{\mathbf{i}-\mathbf{j}}{\mathbf{n}}\right\|\right) Y_{\mathbf{j}}}{\sum_{\mathbf{j} \neq \mathbf{i}} K_1\left(\frac{X_{\mathbf{i}} - X_{\mathbf{j}}}{H_{\mathbf{n},X_{\mathbf{i}}}}\right) K_2\left(h_{\mathbf{n},\mathbf{i}}^{-1} \left\|\frac{\mathbf{i}-\mathbf{j}}{\mathbf{n}}\right\|\right)}$$

and

$$\tilde{Y}_{\mathbf{i}}^{NW} = \frac{\sum_{\mathbf{j} \neq \mathbf{i}} K_1\left(\frac{X_{\mathbf{i}} - X_{\mathbf{j}}}{h_{\mathbf{n}}}\right) K_2\left(\rho_{\mathbf{n}}^{-1} \left\|\frac{\mathbf{i}-\mathbf{j}}{\mathbf{n}}\right\|\right) Y_{\mathbf{j}}}{\sum_{\mathbf{j} \neq \mathbf{i}} K_1\left(\frac{X_{\mathbf{i}} - X_{\mathbf{j}}}{h_{\mathbf{n}}}\right) K_2\left(\rho_{\mathbf{n}}^{-1} \left\|\frac{\mathbf{i}-\mathbf{j}}{\mathbf{n}}\right\|\right)}.$$

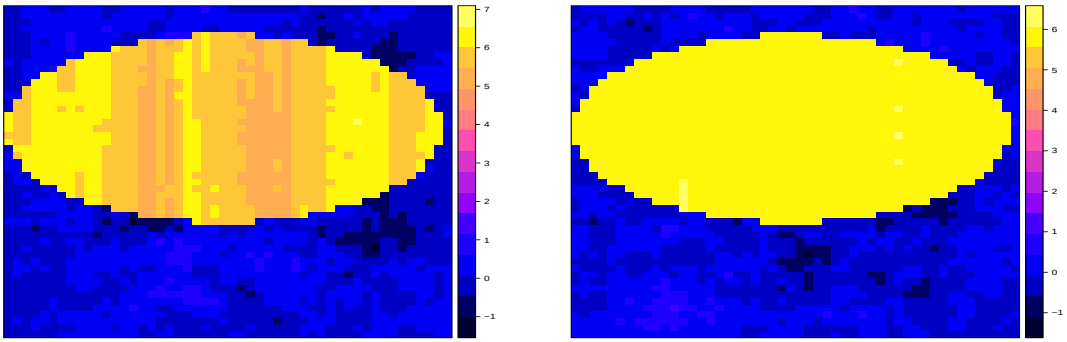

Figure 5.1: Two realizations of the random field with $a = 10$, $\sigma = 5$ (left) and $\sigma = .1$ (right) over a grid 50×50.

Table 5.1: Results of simulations

$n_1 \times n_2$	σ	a	AMAEKE	AMAEkNN	p-value
25×25	5	5	0.258 (0.0484)	0.230 (0.0156)	1.48×10^{-4}
		10	0.375 (0.1081)	0.317 (0.0498)	4.74×10^{-4}
		20	0.435 (0.1043)	0.415 (0.0865)	1.49×10^{-1}
	0.1	5	0.175 (0.0097)	0.159 (0.0066)	3.27×10^{-16}
		10	0.313 (0.0280)	0.212 (0.0105)	2.00×10^{-33}
		20	0.478 (0.0895)	0.278 (0.0277)	3.99×10^{-22}
35×30	5	5	0.239 (0.00062)	0.229 (0.00013)	5.40×10^{-3}
		10	0.316 (0.00284)	0.269 (0.00021)	4.96×10^{-8}
		20	0.360 (0.00318)	0.315 (0.00097)	2.93×10^{-6}
	0.1	5	0.176 (0.00003)	0.167 (0.00002)	4.26×10^{-16}
		10	0.287 (0.00034)	0.214 (0.00009)	2.20×10^{-16}
		20	0.372 (0.00644)	0.264 (0.00021)	4.86×10^{-13}

The table 5.1 gives the average of the mean absolute errors of the both methods, in brackets we have the corresponding standard deviations over the 50 replications. The column entitled p-value, gives for each considered case, the p-value of a paired t-test

Table 5.2: Three considered cases

Case	Primary variable	Secondary variables
1	**Cd**	**Ni**, **Zn**
2	**Cu**	**Pb**, **Ni**, **Zn**
3	**Pb**	**Cu**, **Ni**, **Zn**

performing in order to determine if the mean of MAEKP (mean absolute error of the kernel Prediction) is significantly greater than that of MAEkNN (mean absolute error of our prediction). We notice that the k-NN method performs better than the kernel method in all cases of the spatial dependency parameter a and standard deviation parameter σ. In particular, k-NN method is more efficient than kernel method with very small p-value (less than 4.86×10^{-13}) when the deviation is small, which highlight that k-NN method is more adapted to a local data structure.

5.4.2 A real dataset

In this part, we focus on how the prediction by k-NN method will behave through the famous Jura data set https://sites.google.com/site/goovaertspierre/pierregoovaertswebsite/download/jura-data. This data were collected by Swiss Federal Institute of Technology at Lausanne and studied by several authors (see Atteia et al., 1994; Goovaerts, 1998). It concerns seven potentially toxic metals (Cadmium **Cd**, Cobalt **Co**, Chromium **Cr**, Copper **Cu**, Nickel **Ni**, lead **Pb** and zinc **Zn**) of a 14.5 Km^2 region in the Swiss Jura. All metal concentrations were measured at 359 locations, these locations are divided in two subsets. The first (prediction data set) presents the training sample, composed of 259 locations whereas the second (validation data set) presents the testing sample, composed of 100 locations, will be used to check results provided by predictors. The two subset are represented in Figure 5.2. In order to compare the prediction by k-NN

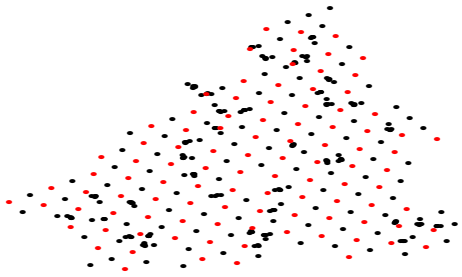

Figure 5.2: Locations considered in the studied region of Swiss Jura, training sample (black points), testing sample (red question mark points).

method and that by kernel introduced by Dabo-Niang et al. (2016). We keep the results on the Jura dataset obtained by Dabo-Niang et al. (2016), and we apply k-NN method to the three cases considered by these authors where the covariates $X_\mathbf{i}$'s and the responses $Y_\mathbf{i}$'s are respectively the secondary variables and the primary variables presented in Table 5.2. Table 5.3 gives the mean absolute error of prediction and shows that k-NN method performs compare to the parametric methods (Ordinary Cokriging, Revisited Cokriging (cov), Revisited Cokriging (corr)) and kernel method of Dabo-Niang et al. (2016) in cases 1 and 3. In case 2, k-NN and kernel methods give similar results with a slight better performance of the kernel method.

Table 5.3: The mean absolute error of prediction for different parametric and nonparametric methods on the three considered cases.

Method	Case 1	Case 2	Case 3
Ordinary Cokriging	0.51	7.90	10.80
Revisited Cokriging (cov)	0.52	7.80	10.70
Revisited Cokriging (corr)	0.52	7.40	10.60
Kernel Method	0.42	**7.02**	11.02
k-NN Method	**0.40**	7.12	**10.51**
K_1	Silverman	Gaussian	Silverman
K_2	Biweight	Parzen	Parzen

5.5 Conclusion

In this work, we used a k-NN method to define a nonparametric spatial predictor for real-valued spatial processes. In one hand, we generalize the classical k-NN kernel method to predict a spatial process at non-observed locations. The proposed predictor combines two kernels to controls distances between observation and locations and uses a bandwidth as the kth lower distance between covariate's point of prediction and covariate's observations. This idea allowed more flexibility to account some heterogeneity in the covariate. We established almost complete convergence with rates of the predictor. The proposed predictor is applied to a prediction problem through an environmental data set. The numerical results show that k-NN kernel method outperforms kernel methods, particularly in presence of a local spatial heterogeneity data structure. This is well known in the case of non-spatial data. One can then see the proposed methodology as a good alternative to the classical k-NN approach for spatial data of Li & Tran (2009) that does not take into account the proximity between locations.

5.6 Appendix

We start to introduce these followings technical lemmas that will permit us to handle the difficulties induced by the random bandwidth $H_{\mathbf{n},x}$ in the expression of the function $r_{\mathrm{kNN}}(x)$. These technical lemmas represent adaptation of the results given in Collomb (1980) (for independent multivariate data) and their generalized version by Burba et al. (2009), Kudraszow & Vieu (2013) (for independent functional data).

Technical Lemmas

For $x \in D$, we define

$$c_{\mathbf{n}}(H_{\mathbf{n},x}) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} Y_{\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}$$

and

$$\forall\, \mathbf{n} \in (\mathbb{N}^*)^N \qquad v_{\mathbf{n}} = \left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d} + \left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^1 k(\mathbf{n})}\right)^{1/2}.$$

For all $\beta \in]0,1[$ and $x \in D$, let

$$D_{\mathbf{n}}^-(\beta_{\mathbf{n}},x) = \left(\frac{k(\mathbf{n})}{cf(x)\hat{\mathbf{n}}}\right)^{1/d}\beta^{1/2d},\ D_{\mathbf{n}}^+(\beta_{\mathbf{n}},x) = \left(\frac{k(\mathbf{n})}{cf(x)\hat{\mathbf{n}}}\right)^{1/d}\beta^{-1/2d} \tag{5.17}$$

where c is the bulk of the unit sphere of $\mathbb{R}^d$. It is clair that

$$\forall\, \mathbf{n} \in (\mathbb{N}^*)^N, \forall x \in D \qquad D_{\mathbf{n}}^-(\beta,x) \leq D_{\mathbf{n}}^+(\beta,x).$$

Lemma 5.3. *If the following conditions are verified:*

(L_1) $\mathbb{I}\left(D_{\mathbf{n}}^{-}(\beta,x) \leq H_{\mathbf{n},x} \leq D_{\mathbf{n}}^{+}(\beta,x),\, \forall x \in D\right) \longrightarrow 1 \quad a.co.$

$$(L_2)\ \sup_{x\in D}\left|\frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)} - \beta\right| \longrightarrow 0 \quad a.co.$$

$$(L_3)\ \sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{-}(\beta,x)\right) - r(x)\right| \longrightarrow 0 \quad a.co.,\ \sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{+}(\beta,x)\right) - r(x)\right| \longrightarrow 0 \quad a.co.,$$

then we have $\sup_{x\in D}|c_{\mathbf{n}}\left(H_{\mathbf{n},x}\right) - r(x)| \longrightarrow 0 \qquad a.co.$

Lemma 5.4. *Under the following conditions:*

(L_1) $\mathbb{I}\left(D_{\mathbf{n}}^{-}(\beta,x) \leq H_{\mathbf{n},x} \leq D_{\mathbf{n}}^{+}(\beta,x),\, \forall x \in D\right) \longrightarrow 1 \quad a.co.$

$$(L_2')\ \sup_{x\in D}\left|\frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)} - \beta\right| = O(v_{\mathbf{n}}) \qquad a.co.$$

$$(L_3')\ \sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{-}(\beta,x)\right) - r(x)\right| = O(v_{\mathbf{n}}) \qquad a.co,$$
$$\sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{+}(\beta,x)\right) - r(x)\right| = O(v_{\mathbf{n}}) \qquad a.co,$$

we have, $\sup_{x\in D}|c_{\mathbf{n}}\left(H_{\mathbf{n},x}\right) - r(x)| = O(v_{\mathbf{n}}) \qquad a.co.$

Proof of Lemma 5.3

The proof of this lemma is a particular case of that of Lemma 5.4 when taking $v_{\mathbf{n}} = 1$ and $C = 1$, it is then omitted.

■

Proof of Lemma 5.4

We give the proof when the random variables(r.v.) $Y_{\mathbf{i}}$ is positive. For any real valued random variable $Y_{\mathbf{i}}$, same arguments as bellow can be used by considering $Y_{\mathbf{i}} = Y_{\mathbf{i}}^{+} - Y_{\mathbf{i}}^{-}$ where $Y_{\mathbf{i}}^{-} = -\min(Y_{\mathbf{i}}, 0)$ and $Y_{\mathbf{i}}^{+} = \max(Y_{\mathbf{i}}, 0)$. For all $\mathbf{n} \in (\mathbb{N}^*)^N$, $\beta \in]0,1[$ and $x \in D$, let

$$C_{\mathbf{n}}^{-}(\beta,x) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} Y_{\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)},$$

$$C_{\mathbf{n}}^{+}(\beta,x) = \frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} Y_{\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}.$$

From (L_2') and (L_3'), it follows that

$$\sup_{x\in D}\left|C_{\mathbf{n}}^{-}(\beta,x) - r(x)\beta\right| = O(v_{\mathbf{n}}) \quad a.co. \tag{5.18}$$

$$\sup_{x\in D}\left|C_{\mathbf{n}}^{+}(\beta,x) - \frac{r(x)}{\beta}\right| = O(v_{\mathbf{n}})\ a.co. \tag{5.19}$$

For all $\varepsilon > 0$, let

$$T_{\mathbf{n}}(\varepsilon) = \left\{\sup_{x\in D}|c_{\mathbf{n}}(H_{\mathbf{n},x}) - r(x)| \leq \varepsilon v_{\mathbf{n}}\right\}$$

$$S_{\mathbf{n}}^{-}(\varepsilon,\beta) = \left\{\sup_{x\in D}\left|C_{\mathbf{n}}^{-}(\beta,x) - r(x)\right| \leq \varepsilon v_{\mathbf{n}}\right\},$$

$$S_{\mathbf{n}}^{+}(\varepsilon,\beta)=\{\sup_{x\in D}\left|C_{\mathbf{n}}^{+}(\beta,x)-r(x)\right|\leq\varepsilon v_{\mathbf{n}}\},$$

and

$$S_{\mathbf{n}}(\beta)=\{C_{\mathbf{n}}^{-}(\beta,x)\leq c_{\mathbf{n}}(H_{\mathbf{n},x})\leq C_{\mathbf{n}}^{+}(\beta,x),\ \forall x\in D\}.$$

It is clear that

$$\forall\,\varepsilon>0,\ \ \forall\,\beta\in]0,1[,\ \ S_{\mathbf{n}}^{-}(\varepsilon,\beta)\cap S_{\mathbf{n}}(\beta)\cap S_{\mathbf{n}}^{+}(\varepsilon,\beta)\subset T_{\mathbf{n}}(\varepsilon). \tag{5.20}$$

As $v_{\mathbf{n}}\longrightarrow 0$, there exists $C>0$ such that $\forall\mathbf{n}\in(\mathbb{N}^{*})^{N},\quad v_{\mathbf{n}}<C$.
For all $\varepsilon>0$ and $x\in D$, such that

$$0<\varepsilon<\varepsilon_{0}=\frac{3\inf_{x\in D}r(x)}{2C},\quad \beta=\beta_{\mathbf{n},\varepsilon,x}=1-\frac{\varepsilon v_{\mathbf{n}}}{3r(x)} \tag{5.21}$$

let

$$G_{\mathbf{n}}^{-}(\varepsilon)=\left\{\sup_{x\in D}\left|C_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)-\beta_{\mathbf{n},\varepsilon,x}r(x)\right|\leq\frac{\varepsilon v_{\mathbf{n}}}{3}\right\}$$

$$G_{\mathbf{n}}^{+}(\varepsilon)=\left\{\sup_{x\in D}\left|C_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x)-\frac{r(x)}{\beta_{\mathbf{n},\varepsilon,x}}\right|\leq\frac{\varepsilon v_{\mathbf{n}}}{3}\right\}$$

and

$$G_{\mathbf{n}}(\varepsilon)=\{D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)\leq H_{\mathbf{n},x}\leq D_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x),\ \forall x\in D\}.$$

Under (5.21), we have by a simple calculation

$$G_{\mathbf{n}}^{-}(\varepsilon)\subset S_{\mathbf{n}}^{-}(\varepsilon,\beta_{\mathbf{n},\varepsilon,x})\quad\text{and}\quad G_{\mathbf{n}}^{+}(\varepsilon)\subset S_{\mathbf{n}}^{+}(\varepsilon,\beta_{\mathbf{n},\varepsilon,x}).$$

So, under condition (ii) in (H4) and $D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)\leq H_{\mathbf{n},x}\leq D_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x)$, for all $x\in D$, we have

$$K_{2}\left(\frac{x-X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right)\ =\ K_{2}\left(\frac{D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)}{H_{\mathbf{n},x}}\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)}\right)\geq K_{2}\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)}\right)$$

and

$$K_{2}\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x)}\right)\ =\ K_{2}\left(\frac{H_{\mathbf{n},x}}{D_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x)}\frac{x-X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right)\geq K_{2}\left(\frac{x-X_{\mathbf{i}}}{H_{\mathbf{n},x}}\right).$$

It is clear that for all $x\in D$, $C_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)\leq c_{\mathbf{n}}(H_{\mathbf{n},x})\leq C_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x)$, thus $G_{\mathbf{n}}(\varepsilon)\subset S_{\mathbf{n}}(\beta_{\mathbf{n},\varepsilon,x})$. Then, $\forall\varepsilon\in]0,\varepsilon_{0}[$, we have

$$\overline{T}_{\mathbf{n}}(\varepsilon)\subset\overline{S_{\mathbf{n}}^{-}}(\varepsilon)\cup\overline{G}_{\mathbf{n}}(\varepsilon)\cup\overline{S_{\mathbf{n}}^{+}}(\varepsilon).$$

Then, we can write

$$\begin{aligned}
&P\left(\sup_{x\in D}\mid c_{\mathbf{n}}(H_{\mathbf{n},x})-r(x)\mid>\varepsilon v_{\mathbf{n}}\right)\leq\\
&P\left(\sup_{x\in D}\mid C_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x})-r(x)\beta_{\mathbf{n},\varepsilon,x}\mid>\frac{\varepsilon v_{\mathbf{n}}}{3}\right)\\
&\qquad+P\left(\sup_{x\in D}\mid C_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x})-\frac{r(x)}{\beta_{\mathbf{n},\varepsilon,x}}\mid>\frac{\varepsilon v_{\mathbf{n}}}{3}\right)\\
&\qquad+P\left(\mathbb{I}\left(D_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x},x)\leq H_{\mathbf{n},x}\leq D_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x},x),\ \forall x\in D\right)=0\right).
\end{aligned}$$

From (5.18), there exist $\varepsilon_{1}\in]0,\varepsilon_{0}[$, such that

$$\sum_{\mathbf{n}\in(\mathbb{N}^{*})^{N}}P\left(\sup_{x\in D}\mid C_{\mathbf{n}}^{-}(\beta_{\mathbf{n},\varepsilon,x})-r(x)\beta_{\mathbf{n},\varepsilon,x}\mid>\frac{\varepsilon_{1}v_{\mathbf{n}}}{3}\right)<\infty.$$

If (5.19) holds, then $\exists\varepsilon_{1}\in]0,\varepsilon_{0}[$, such that

$$\sum_{\mathbf{n}\in(\mathbb{N}^{*})^{N}}P\left(\sup_{x\in D}\mid C_{\mathbf{n}}^{+}(\beta_{\mathbf{n},\varepsilon,x})-\frac{r(x)}{\beta_{\mathbf{n},\varepsilon,x}}\mid>\frac{\varepsilon_{1}v_{\mathbf{n}}}{3}\right)<\infty$$

and by (L_1)

$$\sum_{\mathbf{n}\in(\mathbb{N}^*)^N} P\left(\mathbb{I}\left(D_{\mathbf{n}}^-(\beta_{\mathbf{n},\varepsilon,x},x)\leq H_{\mathbf{n},x}\leq D_{\mathbf{n}}^+(\beta_{\mathbf{n},\varepsilon,x},x),\,\forall x\in D\right)=0\right)<\infty.$$

Then, there exists $\varepsilon_1\in]0,\varepsilon_0[$, such that

$$\sum_{\mathbf{n}\in(\mathbb{N}^*)^N} P\left(\sup_{x\in D}\mid c_{\mathbf{n}}(H_{\mathbf{n},x})-r(x)\mid>\varepsilon_1 v_{\mathbf{n}}\right)<\infty,$$

this completes the proof.

■

Proofs of Lemma 5.1 and Lemma 5.2

Since the proof of Lemma 5.1 is based on the result of Lemma 5.3, it is sufficient to check conditions (L_1), (L_2) and (L_3). For the proof of Lemma 5.2, it suffices to check conditions (L_2') and (L_3'). To check the condition (L_1) we will need to use the following two lemmas.

■

Lemma 5.5. *(Ibragimov & Linnik (1971) or Deo (1973))*

i) We assume that the condition (5.8) of the dependence is satisfied. We denote by $\mathcal{L}_r(\mathcal{F})$ the class $\mathcal{F}-$mesurable of random variables X satisfying $\|X\|_r:=\left(E\left(|X|^r\right)\right)^{1/r}<\infty$. Let $X\in\mathcal{L}_r(\mathcal{B}(E))$ and $Y\in\mathcal{L}_s(\mathcal{B}(E'))$. Let $1\leq r,s,t\leq\infty$ such that $\frac{1}{r}+\frac{1}{s}+\frac{1}{t}=1$. Then

$$\left|E(XY)-E(X)E(Y)\right|\leq \|X\|_r\|Y\|_s\left\{\psi\left(\mathrm{Card(E)},\mathrm{Card(E}^{'}\right)\right)\varphi\left(\mathrm{dist}(E,E^{'})\right)\Big\}^{1/t}. \tag{5.22}$$

ii) For random variables bounded with probability 1, we have

$$\left|E(XY)-E(X)E(Y)\right|\leq C\psi\left(\mathrm{Card(E)},\mathrm{Card(E}^{'}\right)\right)\varphi\left(\mathrm{dist}(E,E^{'})\right). \tag{5.23}$$

Lemma 5.6. *Under assumptions of Theorem 5.1, we have*

$$S_{\mathbf{n}}+R_{\mathbf{n}}=O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right)$$

where

$$\Lambda_{\mathbf{i}}=\mathbb{I}_{B(x,D_{\mathbf{n}})},\quad \mathbf{i}\in\mathcal{I}_{\mathbf{n}},\quad \delta_{\mathbf{n}}=P\left(\|X-x\|<D_{\mathbf{n}}\right),\quad D_{\mathbf{n}}^d=O\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)$$

$$S_{\mathbf{n}}=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\mathrm{Var}\left(\Lambda_{\mathbf{i}}\right)\ \text{and}\quad R_{\mathbf{n}}=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\sum_{\substack{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}\\ \mathbf{j}\neq\mathbf{i}}}\left|\mathrm{Cov}\left(\Lambda_{\mathbf{i}},\Lambda_{\mathbf{j}}\right)\right|$$

where $B(x,\varepsilon)$ denote the closed ball of $\mathbb{R}^d$ with center x and radius ε.

Proof of Lemma 5.6

Let $\delta_{\mathbf{n},\mathbf{i}}=P\left(\|X_{\mathbf{i}}-x\|<D_{\mathbf{n}}\right)$, we can easily deduced that

$$S_{\mathbf{n}}=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}Var\left(\Lambda_{\mathbf{i}}\right)=\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\delta_{\mathbf{n},\mathbf{i}}(1-\delta_{\mathbf{n},\mathbf{i}})=O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right)$$

by the following results. Under assumption (H7), we have

$$|\delta_{\mathbf{n},\mathbf{i}} - \delta_{\mathbf{n}}| = o\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right), \tag{5.24}$$

in addition, under (H1), it is easy to sees that

$$\delta_{\mathbf{n}} = cf(x)D_{\mathbf{n}}^d + O\left(D_{\mathbf{n}}^{d+1}\right). \tag{5.25}$$

For $R_{\mathbf{n}}$, note that by (H2) and for each $\mathbf{j} \neq \mathbf{i}$

$$\begin{aligned} |\mathrm{Cov}\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)| &= |P\left(\|X_{\mathbf{i}} - x\| < D_{\mathbf{n}}, \|X_{\mathbf{j}} - x\| < D_{\mathbf{n}}\right) \\ &\qquad - P\left(\|X_{\mathbf{i}} - x\| < D_{\mathbf{n}}\right) P\left(\|X_{\mathbf{j}} - x\| < D_{\mathbf{n}}\right)| \\ &\leq \int_{B(x,D_{\mathbf{n}})\times B(x,D_{\mathbf{n}})} \left|f_{X_{\mathbf{i}}X_{\mathbf{j}}}(u,v) - f_{\mathbf{i}}(u)f_{\mathbf{j}}(v)\right| dudv \\ &\leq CD_{\mathbf{n}}^{2d} \leq C\delta_{\mathbf{n}}^2 \end{aligned}$$

since by (5.25)

$$\frac{D_{\mathbf{n}}^d}{\delta_{\mathbf{n}}} \to \frac{1}{cf(x)}, \qquad \text{as} \quad \mathbf{n} \to \infty.$$

Using Lemma 5.5, we can write for $r = s = 2/\lambda$ with $\lambda = 1/2$

$$\begin{aligned} |\mathrm{Cov}\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)| &\leq C\left(E\left(\Lambda_{\mathbf{i}}^{(2/\lambda)}\right)\right)^{\lambda/2}\left(E\left(\Lambda_{\mathbf{j}}^{(2/\lambda)}\right)\right)^{\lambda/2}\left(\psi(1,1)\varphi\left(\|\mathbf{i}-\mathbf{j}\|\right)\right)^{1-\lambda} \\ &\leq C\delta_{\mathbf{n}}^{\lambda}\varphi\left(\|\mathbf{i}-\mathbf{j}\|\right)^{1-\lambda}. \end{aligned}$$

Let $q_{\mathbf{n}}$ be a sequence of real numbers defined as $q_{\mathbf{n}}^N = O\left(\frac{\hat{\mathbf{n}}}{k(\mathbf{n})}\right)$, $S = \{\mathbf{i},\mathbf{j} \in \mathcal{I}_{\mathbf{n}},\ 0 < \|\mathbf{i}-\mathbf{j}\| \leq q_{\mathbf{n}}\}$ and S^c its complementary in $\mathcal{I}_{\mathbf{n}}$ and write

$$R_{\mathbf{n}} = \sum_{\mathbf{i},\mathbf{j}\in S} |\mathrm{Cov}\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)| + \sum_{\mathbf{i},\mathbf{j}\in S^c} |\mathrm{Cov}\left(\Lambda_{\mathbf{i}}, \Lambda_{\mathbf{j}}\right)| \leq R_{\mathbf{n}}^{(1)} + R_{\mathbf{n}}^{(2)}$$

where $R_{\mathbf{n}}^{(1)} = \sum_{\mathbf{i},\mathbf{j}\in S} C\delta_{\mathbf{n}}^2$ and $R_{\mathbf{n}}^{(2)} = \sum_{\mathbf{i},\mathbf{j}\in S^c} C\delta_{\mathbf{n}}^{\lambda}\varphi\left(\|\mathbf{i}-\mathbf{j}\|\right)^{1-\lambda}$.
Clearly, according to the definitions of $q_{\mathbf{n}}$ and S, and the equation (4.44)

$$R_{\mathbf{n}}^{(1)} = C\delta_{\mathbf{n}}^2 \sum_{\mathbf{i},\mathbf{j}\in S} 1 \leq C\delta_{\mathbf{n}}^2\hat{\mathbf{n}}q_{\mathbf{n}}^N \leq C\hat{\mathbf{n}}\delta_{\mathbf{n}}\frac{\delta_{\mathbf{n}}\hat{\mathbf{n}}}{k(\mathbf{n})} \leq C\hat{\mathbf{n}}\delta_{\mathbf{n}}$$

since by (5.25), $\frac{\delta_{\mathbf{n}}\hat{\mathbf{n}}}{k(\mathbf{n})} \to C$ as $\mathbf{n} \to \infty$. Then, we have $R_{\mathbf{n}}^{(1)} = O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right)$.
In addition, by (5.9), we get

$$\begin{aligned} R_{\mathbf{n}}^{(2)} &= C\delta_{\mathbf{n}}^{\lambda} \sum_{\mathbf{i},\mathbf{j}\in S^c} \varphi\left(\|\mathbf{i}-\mathbf{j}\|\right)^{1-\lambda} = C\delta_{\mathbf{n}}^{\lambda}\hat{\mathbf{n}} \sum_{\|\mathbf{i}\|\geq q_{\mathbf{n}}} \varphi\left(\|\mathbf{i}\|\right)^{1-\lambda} \\ &= C\delta_{\mathbf{n}}\hat{\mathbf{n}}\delta_{\mathbf{n}}^{\lambda-1} \sum_{\|\mathbf{i}\|\geq q_{\mathbf{n}}} \varphi\left(\|\mathbf{i}\|\right)^{1-\lambda} \leq C\delta_{\mathbf{n}}\hat{\mathbf{n}}\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{\lambda-1} \sum_{\|\mathbf{i}\|\geq q_{\mathbf{n}}} \varphi\left(\|\mathbf{i}\|\right)^{1-\lambda} \\ &\leq C\delta_{\mathbf{n}}\hat{\mathbf{n}} \sum_{\|\mathbf{i}\|\geq q_{\mathbf{n}}} \|\mathbf{i}\|^{(N-\theta)(1-\lambda)} \leq C\delta_{\mathbf{n}}\hat{\mathbf{n}}. \end{aligned}$$

This last implies that $R_{\mathbf{n}}^{(2)} = O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right)$. Finally, the result follows:

$$R_{\mathbf{n}} = O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right) \text{ and } S_{\mathbf{n}} + R_{\mathbf{n}} = O\left(\hat{\mathbf{n}}\delta_{\mathbf{n}}\right).$$

■

Verification of (L_1)

Let $\varepsilon_{\mathbf{n}} = 0.5\varepsilon_0 \left(k(\mathbf{n})/\hat{\mathbf{n}}\right)^{1/d}$ with $\varepsilon_0 > 0$ and let $N_{\varepsilon_{\mathbf{n}}} = O(\varepsilon_{\mathbf{n}}^{-d})$ be a positive integer. Since D is compact, one can cover it by $N_{\varepsilon_{\mathbf{n}}}$ closed balls in $\mathbb{R}^d$ of centers $x_i \in D,\ i = 1, \ldots, N_{\varepsilon_{\mathbf{n}}}$ and reduis $\varepsilon_{\mathbf{n}}$. Let us show that

$$\mathbb{I}\left(D_{\mathbf{n}}^-(\beta, x) \leq H_{\mathbf{n},x} \leq D_{\mathbf{n}}^+(\beta, x),\, \forall x \in D\right) \longrightarrow 1 \qquad a.co.$$

which can be written as, $\forall\, \eta > 0$,

$$\sum_{\mathbf{n} \in \mathbb{N}^{*N}} P\left(\left|\mathbb{I}\left(D_{\mathbf{n}}^-(\beta, x) \leq H_{\mathbf{n},x} \leq D_{\mathbf{n}}^+(\beta, x),\, \forall x \in D\right) - 1\right| > \eta\right) < \infty.$$

We have

$$\begin{aligned}
&P\left(\left|\mathbb{I}\left(D_{\mathbf{n}}^-(\beta, x) \leq H_{\mathbf{n},x} \leq D_{\mathbf{n}}^+(\beta, x),\, \forall x \in D\right) - 1\right| > \eta\right) \\
&\leq P\left(\inf_{x \in D} H_{\mathbf{n},x} - D_{\mathbf{n}}^-(\beta, x) < 0\right) + P\left(\sup_{x \in D} H_{\mathbf{n},x} - D_{\mathbf{n}}^+(\beta, x) > 0\right) \\
&\leq P\left(\min_{1 \leq i \leq N_{\varepsilon_{\mathbf{n}}}} H_{\mathbf{n},x_i} - D_{\mathbf{n}}^-(\beta, x_i) < 2\varepsilon_{\mathbf{n}}\right) + P\left(\max_{1 \leq i \leq N_{\varepsilon_{\mathbf{n}}}} H_{\mathbf{n},x_i} - D_{\mathbf{n}}^+(\beta, x_i) > -2\varepsilon_{\mathbf{n}}\right) \\
&\leq N_{\varepsilon_{\mathbf{n}}} \max_{1 \leq i \leq N_{\varepsilon_{\mathbf{n}}}} P\left(H_{\mathbf{n},x_i} < D_{\mathbf{n}}^-(\beta, x_i) + 2\varepsilon_{\mathbf{n}}\right) \\
&\qquad + N_{\varepsilon_{\mathbf{n}}} \max_{1 \leq i \leq N_{\varepsilon_{\mathbf{n}}}} P\left(H_{\mathbf{n},x_i} > D_{\mathbf{n}}^+(\beta, x_i) - 2\varepsilon_{\mathbf{n}}\right), \qquad (5.26)
\end{aligned}$$

Let us evaluate the first term of right-hand side of (5.26), without ambiguity we ignore the i index in x_i. Remark that

$$\begin{aligned}
P\left(H_{\mathbf{n},x} < D_{\mathbf{n}}^-(\beta, x) + 2\varepsilon_{\mathbf{n}}\right) &\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \mathbb{I}_{B(x, D_{\mathbf{n}}^-(\beta,x) + 2\varepsilon_{\mathbf{n}})}(X_{\mathbf{i}}) > k(\mathbf{n})\right) \qquad (5.27) \\
&\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \xi_{\mathbf{i}} > k(\mathbf{n}) - \hat{\mathbf{n}}\delta_{\mathbf{n}}\right) \qquad (5.28) \\
&\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \xi_{\mathbf{i}} > Ck(\mathbf{n})(1 - \beta^{1/2})\right) = P_{1,\mathbf{n}} \qquad (5.29)
\end{aligned}$$

where $\xi_{\mathbf{i}} = \Lambda_{\mathbf{i}} - \delta_{\mathbf{n},\mathbf{i}}$ is centered, $\Lambda_{\mathbf{i}}$ is defined in Lemma 5.8 when we replace $D_{\mathbf{n}}$ by $D_{\mathbf{n}}^- + 2\varepsilon_{\mathbf{n}}$. From (5.27), we get (5.28) by (5.25) while result (5.28) permits to get (5.29) by the help of the following. In fact, according to the definition of $D_{\mathbf{n}}^-$ in (5.17), when we replace $D_{\mathbf{n}}$ by $D_{\mathbf{n}}^- + 2\varepsilon_{\mathbf{n}}$ in (5.25), we get

$$\hat{\mathbf{n}}\delta_{\mathbf{n}} - k(\mathbf{n})\left(\varepsilon_0(cf(x))^d + \beta^{1/2d}\right)^d = o(k(\mathbf{n})) \qquad (5.30)$$

then we have for all $\varepsilon_1 > 0$,

$$k(\mathbf{n}) - \hat{\mathbf{n}}\delta_{\mathbf{n}} > k(\mathbf{n})\left(1 - \left(\varepsilon_0(cf(x))^d + \beta^{1/2d}\right)^d - \varepsilon_1\right).$$

Then, for ε_1 and ε_0 very small such that $1 - \left(\varepsilon_0(cf(x))^d + \beta^{1/2d}\right)^d - \varepsilon_1 > 0$, we can find some constant $C > 0$ such that

$$k(\mathbf{n}) - \hat{\mathbf{n}}\delta_{\mathbf{n}} > Ck(\mathbf{n})(1 - \beta^{1/2}). \qquad (5.31)$$

For the second term in the right-hand side of (5.26),

$$\begin{aligned}
P\left(H_{\mathbf{n},x} > D_{\mathbf{n}}^+(\beta, x) - 2\varepsilon_{\mathbf{n}}\right) &\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} (1 - \mathbb{I}_{B(x, D_{\mathbf{n}}^+(\beta,x) - 2\varepsilon_{\mathbf{n}})}(X_{\mathbf{i}})) > \hat{\mathbf{n}} - k(\mathbf{n})\right) \qquad (5.32) \\
&\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}} > \hat{\mathbf{n}}\delta_{\mathbf{n}} - k(\mathbf{n})\right) \qquad (5.33) \\
&\leq P\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}} > Ck(\mathbf{n})\left(\beta^{-1/2} - 1\right)\right) = P_{2,\mathbf{n}} \qquad (5.34)
\end{aligned}$$

with $\Delta_{\mathbf{i}} = \delta_{\mathbf{n},\mathbf{i}} - \Lambda_{\mathbf{i}}$ is centered, $\Lambda_{\mathbf{i}}$ is defined in Lemma 5.8 replacing $D_{\mathbf{n}}$ by $D^+_{\mathbf{n}} - 2\varepsilon_{\mathbf{n}}$. Result (5.33) is obtained by (5.25). To prove (5.34) remark that by $D^+_{\mathbf{n}}$ in (5.17), when replacing $D_{\mathbf{n}}$ by $D^+_{\mathbf{n}} - 2\varepsilon_{\mathbf{n}}$ in (5.25), we get

$$\hat{\mathbf{n}}\delta_{\mathbf{n}} - k(\mathbf{n})\left(\beta^{-1/2d} - \varepsilon_0(cf(x))^d\right)^d = o(k(\mathbf{n})). \tag{5.35}$$

Thus for all $\varepsilon_2 > 0$, it is easy to see that

$$\hat{\mathbf{n}}\delta_{\mathbf{n}} - k(\mathbf{n}) > k(\mathbf{n})\left(\left(\beta^{-1/2d} - \varepsilon_0(cf(x))^d\right)^d - 1 - \varepsilon_2\right),$$

so for ε_2 and ε_0 small enough such that $\left(\left(\beta^{-1/2d} - \varepsilon_0(cf(x))^d\right)^d - 1 - \varepsilon_2\right) > 0$, there is $C > 0$ such that

$$\hat{\mathbf{n}}\sigma_{\mathbf{n}} - k(\mathbf{n}) > Ck(\mathbf{n})\left(\beta^{-1/2} - 1\right) \tag{5.36}$$

Now, it suffices to prove that

$$\sum_{\mathbf{n}\in\mathbb{N}^{*N}} N_{\varepsilon_{\mathbf{n}}} P_{1,\mathbf{n}} < \infty \quad \text{and} \quad \sum_{\mathbf{n}\in\mathbb{N}^{*N}} N_{\varepsilon_{\mathbf{n}}} P_{2,\mathbf{n}} < \infty.$$

Let us consider $P_{1,\mathbf{n}}$

This proof is based on the classical spatial block decomposition of the sum $\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \xi_{\mathbf{i}}$ similarly to Tran (1990). Without loss of generality, we assume that $n_l = 2bq_l$, $l = 1,\dots,N$, then this decomposition can be presented as follows

$$\begin{aligned}
U(1,\mathbf{n},\mathbf{j}) &= \sum_{\substack{i_l=2j_lb+1,\\ k=1,\dots,N.}}^{(2j_l+1)b} \xi_{\mathbf{i}} \\
U(2,\mathbf{n},\mathbf{j}) &= \sum_{\substack{i_l=2j_lb+1,\\ l=1,\dots,N-1.}}^{(2j_l+1)b} \sum_{i_N=(2j_N+1)b+1,}^{2(j_N+1)b} \xi_{\mathbf{i}} \\
U(3,\mathbf{n},\mathbf{j}) &= \sum_{\substack{i_l=2j_lb+1,\\ l=1,\dots,N-2.}}^{(2j_l+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1,}^{2(j_{N-1}+1)b} \sum_{i_N=2j_Nb+1,}^{(2j_N+1)b} \xi_{\mathbf{i}} \\
U(4,\mathbf{n},\mathbf{j}) &= \sum_{\substack{i_l=2j_lb+1,\\ l=1,\dots,N-2.}}^{(2j_l+1)b} \sum_{i_{N-1}=(2j_{N-1}+1)b+1,}^{2(j_{N-1}+1)b} \sum_{i_N=(2j_N+1)b+1,}^{2(j_N+1)b} \xi_{\mathbf{i}} \\
&\dots
\end{aligned}$$

Note that

$$U(2^{N-1},\mathbf{n},\mathbf{j}) = \sum_{\substack{i_l=(2j_l+1)b+1,\\ l=1,\dots,N-1.}}^{2(j_l+1)b} \sum_{i_N=2j_Nb+1,}^{(2j_N+1)b} \xi_{\mathbf{i}}$$

and that

$$U(2^N,\mathbf{n},\mathbf{j}) = \sum_{\substack{i_l=(2j_l+1)b+1,\\ l=1,\dots,N.}}^{2(j_l+1)b} \xi_{\mathbf{i}}.$$

For each integer $1 \le l \le 2^N$, let

$$T(\mathbf{n},l) = \sum_{\substack{j_l=0\\ l=1,\dots,N}}^{q_l-1} U(l,\mathbf{n},\mathbf{j}).$$

Therefore, we have

$$\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \xi_{\mathbf{i}} = \sum_{l=1}^{2^N} T(\mathbf{n},l). \tag{5.37}$$

Replacing (5.37) on the expression of $P_{1,\mathbf{n}}$, it follows that

$$\begin{aligned} P_{1,\mathbf{n}} &= P\left(\sum_{l=1}^{2^N} T(\mathbf{n},l) > Ck(\mathbf{n})(1-\sqrt{\beta})\right) \\ &\leq 2^N P\left(\mid T(\mathbf{n},1)\mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^N}\right). \end{aligned}$$

We enumerate in an arbitrary manner the $\hat{q} = q_1 \times \ldots \times q_N$ terms $U(1,\mathbf{n},\mathbf{j})$ of the sum $T(\mathbf{n},1)$ and denote them $W_1,\ldots,W_{\hat{q}}$. Notice that, $U(1,\mathbf{n},\mathbf{j})$ is measurable with respect to the field generated by the $Z_{\mathbf{i}}$ with $\mathbf{i}\in\mathbf{I}(\mathbf{n},\mathbf{j}) = \{\mathbf{i}\in\mathcal{I}_{\mathbf{n}} \mid 2j_l b+1 \leq i_l \leq (2j_l+1)b,\ l=1,\ldots,N\}$, the set $\mathbf{I}(\mathbf{n},\mathbf{j})$ contains b^N sites and $\text{dist}(\mathbf{I}(\mathbf{n},\mathbf{j}),\mathbf{I}(\mathbf{n},\mathbf{j}^{'})) > b$. In addition, we have $\mid W_l \mid \leq b^N$.
According to Lemma 4.5 of Carbon et al. (1997), we can find a sequence of independent random variables $W_1^*,\ldots,W_{\hat{q}}^*$ where W_l has same distribution as W_l^* and:

$$\sum_{l=1}^{\hat{q}} E(\mid W_l - W_l^* \mid) \leq 4\hat{q}b^N\psi((\hat{q}-1)b^N,b^N)\varphi(b).$$

Then, we can write

$$\begin{aligned} P_{1,\mathbf{n}} &\leq 2^N P\left(\mid T(\mathbf{n},1)\mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^N}\right) \\ &\leq 2^N P\left(\mid \sum_{l=1}^{\hat{q}} W_l \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^N}\right) \\ &\leq 2^N P\left(\sum_{l=1}^{\hat{q}} \mid W_l - W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) \\ &\quad + 2^N P\left(\sum_{l=1}^{\hat{q}} \mid W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right). \end{aligned}$$

Noting that $P_{11,\mathbf{n}} = P\left(\sum_{l=1}^{\hat{q}} \mid W_l - W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right)$
and $P_{12,\mathbf{n}} = P\left(\sum_{l=1}^{\hat{q}} \mid W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right)$.
It suffices to show that $\sum_{\mathbf{n}\in\mathbb{N}^{*N}} P_{11,\mathbf{n}} < \infty$ and $\sum_{\mathbf{n}\in\mathbb{N}^{*N}} P_{12,\mathbf{n}} < \infty$.

Consider first $P_{11,\mathbf{n}}$

Using Markov's inequality, we get

$$\begin{aligned} P_{11,\mathbf{n}} &= P\left(\sum_{l=1}^{\hat{q}} \mid W_l - W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) \\ &\leq \frac{2^{N+3}}{Ck(\mathbf{n})(1-\sqrt{\beta})}\hat{q}b^N\psi((\hat{q}-1)b^N,b^N)\varphi(b) \\ &\leq \frac{C}{k(\mathbf{n})(1-\sqrt{\beta})}\hat{\mathbf{n}}\psi((\hat{q}-1)b^N,b^N)\varphi(b). \end{aligned}$$

Let

$$b^N = O\left(\left(\frac{\hat{\mathbf{n}}^{1+\tilde{\beta}}}{k(\mathbf{n})}\right)^{2(1-s(2-\gamma-\tilde{\gamma}))/a}\right) \tag{5.38}$$

with $a = s(d+1)(3-\gamma-\tilde{\gamma}) + s(5-2\gamma) + 2d - 1$. Under the assumption on the function $\psi(n,m)$, we distinguish the following two cases:

Case 1

$$\psi(n,m) \leq C \min(n,m)$$

and $\theta > N(sd(3-\gamma-\tilde{\gamma}) + 2s(3-\gamma) + 2d)/(1-s(2-\gamma-\tilde{\gamma}))$ where $2 < s < 1/(2-\gamma-\tilde{\gamma})$. In this case, we have

$$P_{11,\mathbf{n}} \leq \frac{C}{k(\hat{\mathbf{n}})}\hat{\mathbf{n}}b^N\varphi(b). \leq C\frac{\hat{\mathbf{n}}}{k(\hat{\mathbf{n}})}b^{N-\theta}$$

Then by using (5.38) and the definition of $N_{\varepsilon_\mathbf{n}}$, we find that

$$N_{\varepsilon_\mathbf{n}} P_{11,\mathbf{n}} \leq C\hat{\mathbf{n}}^{-2\beta}.$$

This shows that $\sum_{\mathbf{n}\in\mathbb{N}^{*N}} N_{\varepsilon_\mathbf{n}} P_{11,\mathbf{n}} < \infty$.

■

Case 2

$$\psi(n,m) \leq C(n+m+1)^{\tilde{\beta}}$$

and $\theta > N\left(s(d(3-\gamma-\tilde{\gamma}) + (7+2\tilde{\beta}-3\gamma-\tilde{\gamma})) + 2(d+1)\right)/(1-s(2-\gamma-\tilde{\gamma}))$. In this case, we have

$$P_{11,\mathbf{n}} \leq \frac{C}{k(\mathbf{n})}\hat{\mathbf{n}}(\hat{q}b^N)^{\tilde{\beta}}\varphi(b) \leq \frac{C}{k(\mathbf{n})}\hat{\mathbf{n}}^{(\tilde{\beta}+1)}b^{-\theta}.$$

As in Case 1, we have

$$N_{\varepsilon_\mathbf{n}} P_{11,\mathbf{n}} \leq C\hat{\mathbf{n}}^{-\beta}.$$

Then, it follows that $\sum_{\mathbf{n}\in\mathbb{N}^N} N_{\varepsilon_\mathbf{n}} P_{12,\mathbf{n}} < \infty$.

Consider $P_{12,\mathbf{n}}$

Applying Markov's ineguality, we have for each $t > 0$:

$$\begin{aligned} P_{12,\mathbf{n}} &= P\left(\sum_{l=1}^{\hat{q}} \mid W_l^* \mid > \frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) \\ &\leq \exp\left(-t\frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) E\left(\exp\left(t\sum_{l=1}^{\hat{q}} W_l^*\right)\right) \\ &\leq \exp\left(-t\frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) \prod_{l=1}^{\hat{q}} E\left(\exp\left(tW_l^*\right)\right) \end{aligned}$$

since the variables $W_1^*, \ldots, W_{\hat{q}}^*$ are independent.

Let $r > 0$, for $t = \dfrac{r\log(\hat{\mathbf{n}})}{k(\mathbf{n})}$, $l = 1, \ldots, \hat{q}$ and using (5.38), we can easily get

$$\begin{aligned} t \mid W_l^* \mid &\leq \frac{r\log(\hat{\mathbf{n}})}{k(\mathbf{n})}b^N \leq C\frac{\log(\hat{\mathbf{n}})}{k(\mathbf{n})}\left(\frac{\hat{\mathbf{n}}^{\tilde{\beta}+1}}{k(\mathbf{n})}\right)^{2(1-s(2-\gamma-\tilde{\gamma}))/a} \\ &\leq C\frac{\log(\hat{\mathbf{n}})}{\hat{\mathbf{n}}^{\beta/a}}, \end{aligned}$$

where $\beta = a\gamma - 2(1+\tilde{\beta}-\gamma)(1-s(2-\gamma-\tilde{\gamma})) > 0$, we then have $t \mid W_l^* \mid < 1$ for $\mathbf{n}$ large enough. So, $\exp(tW_l^*) \leq 1 + tW_l^* + t^2(W_l^*)^2$ then

$$E\left(\exp\left(tW_l^*\right)\right) \leq 1 + E\left(t^2(W_l^*)^2\right) \leq \exp\left(E\left(t^2(W_l^*)^2\right)\right).$$

Therefore,

$$\prod_{l=1}^{\hat{q}} E\left(\exp\left(tW_l^*\right)\right) \leq \exp\left(t^2 \sum_{l=1}^{\hat{q}} E\left((W_l^*)^2\right)\right).$$

As W_l^* and W_l have the same distribution, we have

$$\sum_{l=1}^{\hat{q}} E\left((W_l^*)^2\right) = Var\left(\sum_{l=1}^{\hat{q}} W_l^*\right) = Var\left(\sum_{l=1}^{\hat{q}} W_l\right) \leq S_{\mathbf{n}} + R_{\mathbf{n}}.$$

From Lemma 5.6, we obtain

$$\begin{aligned}
\prod_{l=1}^{\hat{q}} E\left(\exp\left(tW_l^*\right)\right) &\leq \exp\left(Ct^2 k(\mathbf{n})\left(\sqrt{\beta} + o(1)\right)\right) \\
&\leq \exp\left(Cr^2\left(\sqrt{\beta} + o(1)\right)\frac{\log(\hat{\mathbf{n}})^2}{k(\mathbf{n})}\right) \longrightarrow 1\ ,\ \mathbf{n} \to \infty,
\end{aligned}$$

because $\log(\hat{\mathbf{n}})^2/k(\mathbf{n}) \to 0$ as $\mathbf{n} \to \infty$. Then, we deduce that

$$\begin{aligned}
P_{12,\mathbf{n}} &\leq C\exp\left(-t\frac{Ck(\mathbf{n})(1-\sqrt{\beta})}{2^{N+1}}\right) \\
&\leq C\exp\left(-\frac{rC(1-\sqrt{\beta})}{2^{N+1}}\log(\hat{\mathbf{n}})\right) \leq C\hat{\mathbf{n}}^{-\frac{rC(1-\sqrt{\beta})}{2^{N+1}}}.
\end{aligned}$$

Then, we deduce that

$$N_{\varepsilon_{\mathbf{n}}} P_{12,\mathbf{n}} < C\hat{\mathbf{n}}^{1-\gamma-\frac{rC(1-\sqrt{\beta})}{2^{N+1}}}$$

Therefore, for some $r > 0$ such that $\gamma + \dfrac{rC(1-\sqrt{\beta})}{2^{N+1}} > 2$, we get

$$\sum_{\mathbf{n}\in\mathbb{N}^N} N_{\varepsilon_{\mathbf{n}}} P_{12,\mathbf{n}} < \infty.$$

Combining the two results, we get $\displaystyle\sum_{\mathbf{n}\in\mathbb{N}^N} N_{\varepsilon_{\mathbf{n}}} P_{1,\mathbf{n}} < \infty$.

Using same calculations as for $P_{1,\mathbf{n}}$, we have $\sum_{\mathbf{n}\in\mathbb{N}^N} N_{\varepsilon_{\mathbf{n}}} P_{2,\mathbf{n}} < \infty$.

■

Now the check of conditions (L_2), (L_3), (L_2') and (L_3') is based on Theorem 3.1 in Dabo-Niang et al. (2016), we need to show that $D_{\mathbf{n}}^-(\beta,x)$, $D_{\mathbf{n}}^+(\beta,x)$ satisfy assumptions (H6) and (H7) used by these authors for all $(\beta,x) \in]0,1[\times D$. This is proved in the following lemmas where without ambiguity D_n will denote $D_{\mathbf{n}}^-(\beta,x)$ or $D_{\mathbf{n}}^+(\beta,x)$.

Lemma 5.7. *Under assumption (H5) on function $\psi(.)$, we have*

$$\hat{\mathbf{n}} D_{\mathbf{n}}^{d\theta_0} h_{\mathbf{n},\mathbf{s}}^{N\theta_1} \log(\hat{\mathbf{n}})^{-\theta_2} u_{\mathbf{n}}^{-\theta_3} \to \infty$$

with

$$\theta_0 = \frac{s(\theta + N(d+2))}{\theta - N(s(d+4)+2d)}; \qquad \theta_1 = \frac{s(\theta+Nd)}{\theta - N(s(d+4)+2d)},$$
$$\theta_2 = \frac{s(\theta - N(d+2))}{\theta - N(s(d+4)+2d)}; \qquad \theta_3 = \frac{2(\theta + N(d+s))}{\theta - N(s(d+4)+2d)},$$

and $u_{\mathbf{n}} = \displaystyle\prod_{i=1}^{N} (\log(\log(n_i)))^{1+\varepsilon} \log(n_i)$ *for all* $\varepsilon > 0$.

Proof of Lemma 5.7

We have

$$\begin{aligned}\hat{\mathbf{n}} D_{\mathbf{n}}^{d\theta_0} h_{\mathbf{n},\mathbf{s}}^{N\theta_1} \log(\hat{\mathbf{n}})^{-\theta_2} u_{\mathbf{n}}^{-\theta_3} &\geq C\hat{\mathbf{n}} \left(\frac{k_{\mathbf{n}}^1}{\hat{\mathbf{n}}}\right)^{\theta_0} \left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{\theta_1} \log(\hat{\mathbf{n}})^{-\theta_2} u_{\mathbf{n}}^{-\theta_3} \\ &\geq C\frac{\hat{\mathbf{n}}^{1+(\gamma-1)\theta_0+(\tilde{\gamma}-1)\theta_1}}{\log(\hat{\mathbf{n}})^{\theta_2} u_{\mathbf{n}}^{\theta_3}}.\end{aligned}$$

Note that $u_{\mathbf{n}} \leq \log(n_j)^{N(2+\varepsilon)} \Rightarrow \frac{1}{u_{\mathbf{n}}^{\theta_3}} \geq \frac{1}{\log(n_j)^{(2+\varepsilon)N\theta_3}}$, where $n_j = \max_{k=1,\ldots,N} n_k$, and

$$\log(\hat{\mathbf{n}}) \leq C\log(n_j) \Rightarrow \frac{1}{\log(\hat{\mathbf{n}})^{\theta_2}} \geq C\frac{1}{\log(n_j)^{\theta_2}}.$$

Since $\frac{n_k}{n_i} \leq C,\ \forall\, 1 \leq k, i \leq N$, we deduce that $\hat{\mathbf{n}} \geq Cn_j^N$. Indeed, a simple calculation gives

$$\begin{aligned}&\hat{\mathbf{n}} D_{\mathbf{n}}^{d\theta_0} h_{\mathbf{n},\mathbf{s}}^{N\theta_1} \log(\hat{\mathbf{n}})^{-\theta_2} u_{\mathbf{n}}^{-\theta_3} \\ &\quad\geq C\left[\frac{\hat{\mathbf{n}}^{\theta(1-s(2-\gamma-\tilde{\gamma}))-N(s(d(3-\gamma-\tilde{\gamma})+2(3-\gamma))+2d)}}{\log(n_j)^{s(\theta-N(d+2))+2N(2+\varepsilon)(\theta+N(d+s))}}\right]^{1/(\theta-N(s(d+4)+2d))} \\ &\quad\geq C\left[\frac{n_j^{N(\theta(1-s(2-\gamma-\tilde{\gamma}))-N(s(d(3-\gamma-\tilde{\gamma})+2(3-\gamma))+2d))}}{\log(n_j)^{s(\theta-N(d+2))+2N(2+\varepsilon)(\theta+N(d+s))}}\right]^{1/(\theta-N(s(d+4)+2d))} \\ &\quad\to +\infty.\end{aligned}$$

■

Lemma 5.8. *Under assumption (H6) on $\psi(.)$, we have*

$$\hat{\mathbf{n}} D_{\mathbf{n}}^{d\theta_0'} h_{\mathbf{n},\mathbf{s}}^{N\theta_1'} \log(\hat{\mathbf{n}})^{-\theta_2'} u_{\mathbf{n}}^{-\theta_3'} \to \infty$$

with

$$\theta_0' = \frac{s(\theta+N(d+3))}{\theta-N(s(d+3++2\tilde{\beta})+2(d+1))};\quad \theta_1' = \frac{s(\theta+N(d+1))}{\theta-N(s(d+3++2\tilde{\beta})+2(d+1))}$$

$$\theta_2' = \frac{s(\theta-N(d+1))}{\theta-N(s(d+3++2\tilde{\beta})+2(d+1))};\quad \theta_3' = \frac{2(\theta+N(s+d+1))}{\theta-N(s(d+3+2\tilde{\beta})+2(d+1))}.$$

Proof of Lemma 5.8

The proof is very similar to the one of Lemma 5.7 and is omitted.

■

Verification of (L_2)

Let

$$f_{\mathbf{n}}\left(x, D_{\mathbf{n}}^-(\beta,x)\right) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n},\mathbf{s_0}}^N \left(D_{\mathbf{n}}^-(\beta,x)\right)^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^-(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s}-\mathbf{i}}{\mathbf{n}}\right\|\right),$$

and

$$f_{\mathbf{n}}\left(x, D_{\mathbf{n}}^+(\beta,x)\right) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n},\mathbf{s_0}}^N \left(D_{\mathbf{n}}^+(\beta,x)\right)^d} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}} K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^+(\beta,x)}\right) K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right).$$

Under the hypotheses of Lemma 5.1 and the results of Lemma 5.7 and Lemma 5.8, (see Dabo-Niang et al., 2016), we have

$$\sup_{x\in D}\left|f_{\mathbf{n}}\left(x, D_{\mathbf{n}}^-(\beta,x)\right) - f(x)\right| \longrightarrow 0 \qquad a.co.$$

$$\sup_{x\in D}\left|f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{+}(\beta,x)\right)-f(x)\right|\longrightarrow 0 \qquad a.co,$$

then,

$$\sup_{x\in D}\left|\frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s}\neq\mathbf{i}}K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right)K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}}K_1\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right)K_2\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)}-\beta\right|=\beta\sup_{x\in D}\left|\frac{f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{-}(\beta,x)\right)}{f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{+}(\beta,x)\right)}-1\right|\to 0 \quad a.co.$$

■

Verification of (L_3)

Under assumptions of Lemma 5.1, Lemmas 5.7 and 5.8, it follows that (see Dabo-Niang et al., 2016)

$$\sup_{x\in D}|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{-}(\beta,x)\right)-r(x)|\to 0 \qquad a.co \text{ and } \sup_{x\in D}|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{+}(\beta,x)\right)-r(x)|\to 0 \quad a.co.$$

■

Proof of Lemma 5.2

The proof of this lemma is based on the results of Lemma 5.4. It suffices to check the conditions (L_2') and (L_3'). Clearly, similar arguments as those involved to to prove (L_2) and (L_3) can be used to obtain the requested conditions.

Verification of (L_2')

Under assumptions of corollary 5.1 and Lemmas 5.7, 5.8, we have

$$\begin{aligned}\sup_{x\in D}\left|f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{-}(\beta,x)\right)-f(x)\right| &= O\left(D_{\mathbf{n}}^{-}(\beta,x)\right)+O\left(\left(\frac{\log(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(D_{\mathbf{n}}^{-}(\beta,x))^{d}h_{\mathbf{n},\mathbf{s_0}}^{N}}\right)^{1/2}\right) a.co.\\ &= O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d}+\left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^{1}k(\mathbf{n})}\right)^{1/2}\right) a.co.,\end{aligned}$$

$$\begin{aligned}\sup_{x\in D}\left|f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{+}(\beta,x)\right)-f(x)\right| &= O\left(D_{\mathbf{n}}^{+}(\beta,x)\right)+O\left(\left(\frac{\log(\hat{\mathbf{n}})}{\hat{\mathbf{n}}(D_{\mathbf{n}}^{+}(\beta,x))^{d}h_{\mathbf{n},\mathbf{s_0}}^{N}}\right)^{1/2}\right) a.co.\\ &= O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d}+\left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^{1}k(\mathbf{n})}\right)^{1/2}\right). a.co.\end{aligned}$$

Then, we deduce that

$$\begin{aligned}&\sup_{x\in D}\left|\frac{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}}K_1\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)K_2\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{-}(\beta,x)}\right)}{\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}},\mathbf{s_0}\neq\mathbf{i}}K_1\left(h_{\mathbf{n},\mathbf{s_0}}^{-1}\left\|\frac{\mathbf{s_0}-\mathbf{i}}{\mathbf{n}}\right\|\right)K_2\left(\frac{x-X_{\mathbf{i}}}{D_{\mathbf{n}}^{+}(\beta,x)}\right)}-\beta\right|\\ &\qquad=\beta\sup_{x\in D}\left|\frac{f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{-}(\beta,x)\right)}{f_{\mathbf{n}}\left(x,D_{\mathbf{n}}^{+}(\beta,x)\right)}-1\right|=O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d}+\left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^{1}k(\mathbf{n})}\right)^{1/2}\right) a.co.\end{aligned}$$

■

Verification of (L_3')

It is relatively easy to deduce from Lemmas 5.7 and 5.8 (Dabo-Niang et al., 2016), that

$$\sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{-}(\beta,x)\right)-r(x)\right|=O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d}+\left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^{1}k(\mathbf{n})}\right)^{1/2}\right) a.co.$$

$$\sup_{x\in D}\left|c_{\mathbf{n}}\left(D_{\mathbf{n}}^{+}(\beta,x)\right)-r(x)\right|=O\left(\left(\frac{k(\mathbf{n})}{\hat{\mathbf{n}}}\right)^{1/d}+\left(\frac{\hat{\mathbf{n}}\log(\hat{\mathbf{n}})}{k_{\mathbf{n}}^{1}k(\mathbf{n})}\right)^{1/2}\right)\ a.co.$$

■

Chapter 6

Partially linear spatial probit models

Contents

Resumé français

Dans ce chapitre, nous étendons l'étude du modèle à choix binaire considéré dans la première contribution au cas où les données sont de nature spatiale mais de dimension finie. Nous considérons un modèle probit spatial partialement linéaire. Ce modèle semi-paramétrique est estimé en combinant une technique non-paramétrique et la méthode des moments généralisés.

Nous supposons que nous disposons d'un vecteur aléatoire (Y, X, Z) observé en n sites spatiaux notés $\{s_1, s_2, \ldots, s_n\}$ tels que $\|s_i - s_j\| > \rho$ avec $\rho > 0$, où X et Z sont des variables explicatives à valeurs dans les sous compacts $\mathcal{X} \subset \mathbb{R}^p (p \geq 1)$ et $\mathcal{Z} \subset \mathbb{R}^d (d \geq 1)$ respectivement, $Y \in \{0, 1\}$. On considère les observations $(Y_{s_i}, X_{s_i}, Z_{s_i})_{i=1,\ldots,n}$ comme des *triangular arrays* $(Y_{in}, X_{in}, Z_{in})_{i=1,\ldots,n}$ (Robinson, 2011) et écrivons le modèle à l'aide d'une variable latente Y^*_{in} :

$$Y^*_{in} = X^T_{in}\beta_0 + g_0(Z_{in}) + U_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.1}$$

avec

$$Y_{in} = \mathbb{I}\left(Y^*_{in} \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.2}$$

Le paramètre β_0 est un vecteur $(p \times 1)$ inconnu, dans un sous ensemble compact $\Theta_\beta \subset \mathbb{R}^p$, $g_0(\cdot)$ est une fonction inconnue dans l'espace de fonctions
$\mathcal{G} = \left\{g \in C^2(\mathcal{Z}) : \|g\| = \sup_{z \in \mathcal{Z}} |g(z)| < C\right\}$ où $C^2(\mathcal{Z})$ est l'espace des fonctions deux fois différentiables de $\mathcal{Z}$ à $\mathbb{R}$. Nous supposons que β_0 et $g_0(\cdot)$ sont indépendants de i (et n) et que le terme d'erreur U_{in} dans (6.2) suit un processus autoregressive spatial :

$$U_{in} = \lambda_0 \sum_{j=1}^{n} w_{ijn} U_{jn} + \varepsilon_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots, \tag{6.3}$$

où λ_0 est un paramètre autoregressif, à valeurs dans un sous ensemble compact $\Theta_\lambda \subset \mathbb{R}$, w_{ijn}, $j = 1, ..., n$, sont les éléments de la i-ème ligne d'une matrice $(n \times n)$ de poids spatial W_n. La matrice $(I_n - \lambda_0 W_n)$ est supposée inversible pour tout n, les variables $\{\varepsilon_{in},\ 1 \leq i \leq n\}$ sont des gaussiennes standards, indépendantes et identiquement distribuées. Nous supposons que, pour tout $n = 1, 2, \ldots$, la suite $\{\varepsilon_{in},\ 1 \leq i \leq n\}$ est indépendante des suites $\{X_{in},\ 1 \leq i \leq n\}$ et $\{Z_{in},\ 1 \leq i \leq n\}$, et que $\{X_{in},\ 1 \leq i \leq n\}$ est indépendante de $\{Z_{in},\ 1 \leq i \leq n\}$.
Nous pouvons alors ré-écrire l'equation (6.3) sous la forme matricielle suivante :

$$U_n = (I_n - \lambda_0 W_n)^{-1} \varepsilon_n, \quad n = 1, 2, \ldots$$

où $U_n = (U_{n1}, \ldots, U_{nn})^T$ et $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$.
Par conséquent, la matrice de variance-covariance de U_n est

$$V_n(\lambda_0) = \mathrm{Var}(U_n) = (I_n - \lambda_0 W_n)^{-1} \left\{ (I_n - \lambda_0 W_n)^T \right\}^{-1}, \quad n = 1, 2, \ldots$$

Cette matrice permet de décrire les dépendances spatiales croisées entre les observations. La méthode d'estimation que nous proposons dans la suite reste valable pour toute sorte de dépendance spatiale au niveau des erreurs, avec une matrice de variance-covariance dépendante d'un certain paramètre λ_0.

Notre objectif est d'estimer le modèle (6.1) à savoir les paramètres β_0 et λ_0 et la fonction $g_0(\cdot)$ à l'aide les n observations (X_{in}, Y_{in}, Z_{in}), $i = 1, ..., n$ et d'une approche semi paramétrique. Dans ce but, remarquons que d'après (6.2)

$$E_0 \left(Y_{in} | X_{in}, Z_{in} \right) = \Phi \left((v_{in}(\lambda_0))^{-1} \left(X_{in}^T \beta_0 + g_0(Z_{in}) \right) \right), \quad 1 =, \ldots n, \tag{6.4}$$

où E_0 indique l'espérance sous les vrais paramètres (c-à-d β_0, λ_0 et $g_0(\cdot)$), $\Phi(\cdot)$ est la fonction de répartition d'une loi normale standard et $(v_{in}(\lambda_0))^2 = V_{iin}(\lambda_0),\ 1 \leq i \leq n,\ n = 1, 2, \cdots$, sont les éléments de la diagonale de la matrice de variance-covariance $V_n(\lambda_0)$.
Pour chaque $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$, $z \in \mathcal{Z}$ et $\eta \in \mathbb{R}$, définissons l'espérance conditionnelle par rapport à Z_{in} du logarithme de la fonction de vraisemblance de Y_{in} pour $1 \leq i \leq n,\ n = 1, 2, \ldots$:

$$H(\eta; \beta, \lambda, z) = E_0 \left(\mathcal{L} \left(\Phi \left((v_{in}(\lambda))^{-1} \left(\eta + X_{in}^T \beta \right) \right); Y_{in} \right) \middle| \, Z_{in} = z \right),$$

avec $\mathcal{L}(u; v) = \log \left(u^v (1-u)^{1-v} \right)$. Nous considérons que $H(\eta; \beta, \lambda, z)$ est indépendante de i (et n).
Pour tout $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$ et $z \in \mathcal{Z}$ fixés, nous noterons par $g_{\beta,\lambda}(z)$ la solution par rapport à η de

$$\frac{\partial}{\partial \eta} H(\eta; \beta, \lambda, z) = 0. \tag{6.5}$$

Cette solution vérifie $g_{\beta_0,\lambda_0}(z) = g_0(z)$ pour tout $z \in \mathcal{Z}$.
Pour obtenir des estimateurs par méthode des moments généralisés (GMM) (Pinkse & Slade, 1998) de β_0 et λ_0, nous exploitons la fonction $g_{\beta,\lambda}(\cdot)$. Nous définissons les résidus généralisés, en remplaçant $g_0(Z_{in})$ dans (6.1) par $g_{\beta,\lambda}(Z_{in})$:

$$\begin{aligned} \tilde{U}_{in}(\beta, \lambda, g_{\beta,\lambda}) &= E \left(U_{in} | Y_{in}, \beta, \lambda \right) \\ &= \frac{\phi \left(G_{in}(\beta, \lambda, g_{\beta,\lambda}) \right) \left(Y_{in} - \Phi \left(G_{in}(\beta, \lambda, g_{\beta,\lambda}) \right) \right)}{\Phi \left(G_{in}(\beta, \lambda, g_{\beta,\lambda}) \right) \left(1 - \Phi \left(G_{in}(\beta, \lambda, g_{\beta,\lambda}) \right) \right)}, \end{aligned} \tag{6.6}$$

où $\phi(\cdot)$ est la densité d'une normale standard et $G_{in}(\beta, \lambda, g_{\beta,\lambda}) = (v_{in}(\lambda))^{-1} \left(X_{in}^T \beta + g_{\beta,\lambda}(Z_{in}) \right)$.
Pour simplifier les notations, nous écrirons quand c'est possible $\theta = (\beta^T, \lambda)^T \in \Theta = \Theta_\beta \times \Theta_\lambda$.
Le résidu généralisé $\tilde{U}_{in}(\cdot\,, \cdot)$ est calculé en conditionnant uniquement par rapport à Y_{in} et non l'échantillon entier du variable binaire $\{Y_{in},\ i = 1, 2, \ldots, n,\ n = 1, \ldots\}$ ou un sous-ensemble de celui-ci. Cela influencera l'efficacité des estimateurs de θ obtenu par ces résidus généralisés, mais permet d'éviter un calcul complexe, voir Poirier & Ruud (1988) pour plus de détails. Pour remédier à cette perte d'efficacité, nous utilisons des variables instrumentales (Pinkse & Slade, 1998), et une matrice aléatoire pour définir une fonction critère. Les variables instrumentales et la matrice de poids aléatoire permettent de prendre en compte plus d'informations sur la dépendance et

l'hétéroscédasticité spatiales présentes dans les données.
Soit

$$S_n(\theta, g_\theta) = n^{-1}\xi_n^T \tilde{U}_n(\theta, g_\theta), \tag{6.7}$$

où $\tilde{U}_n(\theta, g_\theta)$ est le vecteur $(n \times 1)$ composé des éléments $\tilde{U}_{in}(\theta, g_\theta)$, $1 \leq i \leq n$ et ξ_n est une matrice $n \times q$ des variables instrumentales dont la $i$ème ligne est notée ξ_{in}. Nous permettons à cette dernière de dépendre de $g_\theta(\cdot)$ et θ dans l'éventualité que ceci permette d'aider à intégrer plus d'informations sur θ. Nous supposons que ξ_{in} est $\sigma(X_{in}, Z_{in})$ mesurable, $i = 1, \ldots, n,\ n = 1, 2, \ldots$.

Par conséquent, la fonction $S_n(\theta, g_\theta)$ qui peut être vue comme équations de moment, est combinée à une matrice stochastique de poids M_n, $q \times q$, semi-définie positive et qui peut dépendre de l'échantillon, pour définir la fonction suivante,

$$Q_n(\theta, g_\theta) = S_n^T(\theta, g_\theta) M_n S_n(\theta, g_\theta). \tag{6.8}$$

L'estimateur GMM de θ proposé doit minimiser $Q_n(\theta, g_\theta)$ par rapport á θ. Cependant la fonction $g_\theta(\cdot)$ n'est pas connue et doit être remplacée par un estimateur asymptotiquement efficient. D'après (6.5), pour $\theta^T = (\beta^T, \lambda) \in \Theta$ fixé, un estimateur de $g_\theta(z)$, $z \in \mathcal{Z}$ est donné par $\hat{g}_\theta(z)$, solution par rapport η de

$$\sum_{i=1}^{n} \frac{\partial}{\partial \eta} \mathcal{L}\left(\Phi\left(G_{in}(\theta, \eta)\right); Y_{in}\right) K\left(\frac{z - Z_{in}}{b_n}\right) = 0, \tag{6.9}$$

où $K(\cdot)$ est un noyau de $\mathbb{R}^d$ à $\mathbb{R}_+$ et b_n est une fenêtre de lissage.
En remplaçant $g_\theta(\cdot)$ dans (6.8) par l'estimateur $\hat{g}_\theta(\cdot)$, nous obtenons l'estimateur de θ suivant

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} Q_n(\theta, \hat{g}_\theta). \tag{6.10}$$

On en déduit l'estimateur $\hat{g}_{\hat{\theta}}(\cdot)$ de $g_0(\cdot)$.
Dans la suite, nous donnons le comportement asymptotique des estimateurs proposés, notamment la consistance et la normalité asymptotique de $\hat{\theta}$, ainsi que la consistance de $\hat{g}_{\hat{\theta}}(\cdot)$. Ces résultats asymptotiques sont présentés dans un cadre *Increasing domain*. Sous des conditions sur l'identification, la dépendance spatiale, le noyau $K(\cdot)$ et la fenêtre de lissage b_n, nous montrons que

$$\hat{\theta} - \theta_0 = o_p(1), \qquad \left\|\hat{g}_{\hat{\theta}} - g_0\right\| = o_p(1),$$

et

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}\left(0, \Omega(\theta_0)\right),$$

où

$$\Omega(\theta_0) = \left\{B_2(\theta_0)\right\}^{-1} \left\{ \frac{d}{d\theta} S^T(\theta, g_\theta)\Big|_{\theta=\theta_0} \right\} M B_1(\theta_0) M \left\{ \frac{d}{d\theta} S(\theta, g_\theta)\Big|_{\theta=\theta_0} \right\} \left\{B_2(\theta_0)\right\}^{-1},$$

avec

$$B_1(\theta_0) = \lim_{n\to\infty} E_0\left(n S_n(\theta_0, g_0) S_n^T(\theta_0, g_0)\right),$$

$$B_2(\theta_0) = \left\{ \frac{d}{d\theta} S^T(\theta, g_\theta)\Big|_{\theta=\theta_0} \right\} M \left\{ \frac{d}{d\theta} S(\theta, g_\theta)\Big|_{\theta=\theta_0} \right\},$$

tel que

$$\frac{d}{d\theta} S(\theta, g_\theta) = \frac{\partial S}{\partial \theta}(\theta, g_\theta) + \frac{\partial S}{\partial g}(\theta, g_\theta) \frac{\partial}{\partial \theta} g_\theta, \tag{6.11}$$

et

$$S(\theta, g_\theta) = \lim_{n\to\infty} E_0\left(S_n(\theta, g_\theta)\right), \tag{6.12}$$

où M est une matrice déterministe semi-définie positive, et est la limite de la matrice M_n.

Pour une application du résultat de la normalité asymptotique, un estimateur de la matrice de variance covariance $\Omega(\theta_0)$ est nécessaire. Nous proposons l'estimateur suivant

$$\Omega_n(\hat{\theta}) = \left\{B_{2n}(\hat{\theta})\right\}^{-1} \left\{ \frac{d}{d\theta} S_n^T(\theta, \hat{g}_\theta)\Big|_{\theta=\hat{\theta}} \right\} M_n B_{1n}(\hat{\theta}) M_n \left\{ \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta)\Big|_{\theta=\hat{\theta}} \right\} \left\{B_{2n}(\hat{\theta})\right\}^{-1},$$

où

$$B_{1n}(\theta) = nS_n(\theta, \hat{g}_\theta)S_n^T(\theta, \hat{g}_\theta) \qquad \text{and} \qquad B_{2n}(\theta) = \left\{ \frac{d}{d\theta} S_n^T(\theta, \hat{g}_\theta) \right\} M_n \left\{ \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta) \right\}.$$

Dans la suite de ce chapitre, nous introduisons plus en détail le modèle proposé, son estimation, les résultats asymptotiques, une procédure permettant de mettre en oeuvre l'approche proposée. Nous donnons également des résultats numériques sur données simulées pour étudier la performance des estimateurs proposés.

The results of this chapter are obtained in collaboration with Sophie Dabo-Niang (University of Lille).

6.1 Introduction

Agriculture, economics, environmental sciences, urban systems, epidemiology activities are often located in space. Therefore, modeling such activities requires to find a kind of correlation between some random variables in one location with others at neighboring locations, see for instance Pinkse & Slade (1998). This is a significant feature of spatial data analysis. Spatial/Econometrics statistics provides tools to solve such modelling. A lot of studies on spatial effects in statistics and econometrics in many divers models have been published; see Cressie (1993), Anselin (1988) and Arbia (2006) for a review.

Two main ways of incorporating the spatial dependence structure (see for instance Cressie, 1993) can be distinguished basically for geostatistics and lattice data. In the domain of geostatistics, the spatial location is valued in a continuous set of $\mathbb{R}^N$, $N \geq 2$. However, for many activities, the spatial index or location does not vary continuously and may be of the lattice type, the baseline of this current work. This is, for instance, the case in a number of problems. In images analysis, remote sensing form satellites, agriculture and so one, data are often received as regular lattice and identified as the centroids of square pixels, whereas a mapping forms often an irregular lattice. Basically, statistical models for lattice data are linked to nearest neighbors to express the fact that data are nearby. Two popular spatial dependence models have received a lot of attention in lattice data: the spatial autoregressive (SAR) dependent variable model and the spatial autoregressive error model (SAE, where the model error is a SAR), which extend regression in time series to spatial data.

In a in theoretical point of view, various linear spatial regression SAR and SAE models, their identification and estimation methods by the two stage least squares (2SLS), the three stage least squares (3SLS), the maximum likelihood (ML) or quasi-maximum likelihood (QML) and the generalized method of moments (GMM) methods have been developed and summarized by many authors, such as Anselin (1988), Cressie (1993), Kelejian & Prucha (1998), Kelejian & Prucha (1999), Conley (1999), Lee (2004), Lee (2007), Lin & Lee (2010), Zheng & Zhu (2012), Malikov & Sun (2017), Garthoff & Otto (2017), Yang & Lee (2017). Nonlinearity into the field of spatial linear lattice models have less attention, see for instance Robinson (2011) who generalized the kernel regression estimation to spatial lattice data. Su (2012) proposed a semiparametric GMM estimation for some semiparametric SAR models. Extending these models and methods to discrete choice spatial models have less attention, only a few number of papers were concerned in recent years. This may be, as pointed out by Fleming (2004) (see also Smirnov (2010) and Billé (2014)), due to the "added complexity that spatial dependence introduces into discrete choice models". Estimating the model parameters with a full ML approach in spatial discrete choice models, often requires solving a very computationally demanding problem of n-dimensional integration, where n is the sample size.

As for linear models many discrete choice models are fully linear and make use of a continuous latent variable, see for instance Smirnov (2010) and Wang et al. (2013) that proposed pseudo ML methods and Pinkse & Slade (1998) who studied a method based on GMM approach.
When the relationship between the discrete choice variable and some explanatory variables is not linear, then a semiparametric model may be an alternative to fully parametric models. This kind of models is known in literature as *partially linear choice spatial models* and is the baseline of this current work. When the data are independent, these choice models can be viewed as particular

cases of the famous generalized additive models (Hastie & Tibshirani, 1990) and have received a lot of attention in the literature, various methods of estimation have been explored (see for instance Hunsberger, 1994; Severini & Staniswalis, 1994; Carroll et al., 1997).

To the best of our knowledge, semiparametric spatial choice models, have not yet been investigated in a theoretical point of view. To fill in this gap, this work addresses a SAE spatial probit model when the spatial dependence structure is integrated in a disturbance term of the studied model.

We propose a semiparametric estimation method combining the GMM approach and the weighted likelihood method. It consists to first fix the parametric components of the model and estimate nonparametrically the nonlinear component by weighted likelihood (Staniswalis, 1989). The obtained estimator depending on the values at which the parametric components were fixed is used to construct a GMM estimator (Pinkse & Slade, 1998) of these components.

The remainder of the paper is organized as follows. In Section 6.2, we introduce the studied spatial model and the estimation procedure. Section 6.3 is devoted to hypotheses and asymptotic results, whereas Section 6.4 reports some discussions and computation of the estimates. Section 6.5 gives some numerical results based on simulated data to illustrate the performance of the proposed estimators. The last section presents the proofs of the main results.

6.2 Model

We consider that at n spatial locations $\{s_1, s_2, \ldots, s_n\}$ satisfying $\|s_i - s_j\| > \rho$ with $\rho > 0$, observations of a random vector (Y, X, Z) are available. Assume that these observations are considered as triangular arrays (Robinson, 2011) and follow the partially linear model of a latent dependent variable Y^*:

$$Y_{in}^* = X_{in}^T \beta_0 + g_0(Z_{in}) + U_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.13}$$

with

$$Y_{in} = \mathbb{I}\left(Y_{in}^* \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.14}$$

where X and Z are explanatory random variables taking values in two compacts subsets $\mathcal{X} \subset \mathbb{R}^p$ $(p \geq 1)$ and $\mathcal{Z} \subset \mathbb{R}^d$ $(d \geq 1)$ respectively. The parameter β_0 is an unknown $p \times 1$ vector that belongs to a compact subset $\Theta_\beta \subset \mathbb{R}^p$, $g_0(\cdot)$ is an unknown smooth function valued in the space of functions $\mathcal{G} = \left\{g \in C^2(\mathcal{Z}) : \|g\| = \sup_{z \in \mathcal{Z}} |g(z)| < C\right\}$ with $C^2(\mathcal{Z})$ the space of twice differentiable functions from $\mathcal{Z}$ to $\mathbb{R}$, C a positive constant. In model (6.13), β_0 and $g_0(\cdot)$ are constant over i (and n). Assume that the term of disturbance U_{in} in (6.13) is modeled by the following spatial autoregressive process (SAR) :

$$U_{in} = \lambda_0 \sum_{j=1}^{n} w_{ijn} U_{jn} + \varepsilon_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.15}$$

where λ_0 is the autoregressive parameter, valued in the compact subset $\Theta_\lambda \subset \mathbb{R}$, w_{ijn}, $j = 1, ..., n$ are the elements in the i–th row of a non stochastic $n \times n$ spatial weights matrix W_n, that contains the information on the spatial relationship between observations. This spatial weight matrix is usually constructed as a function of the distances (with respect to some metric) between locations, see Pinkse & Slade (1998) for more of details. The $n \times n$ matrix $(I_n - \lambda_0 W_n)$ is assumed to be nonsingular for all n where I_n denotes the $n \times n$ identity matrix, and $\{\varepsilon_{in},\ 1 \leq i \leq n\}$ are assumed to be independent random Gaussian variables; $E(\varepsilon_{in}) = 0$ and $E(\varepsilon_{in}^2) = 1$ for $i = 1, \ldots, n$ $n = 1, 2, \ldots$. Note that one can rewrite (6.15) as:

$$U_n = \left(I_n - \lambda_0 W_n\right)^{-1} \varepsilon_n, \quad n = 1, 2, \ldots$$

where $U_n = (U_{n1}, \ldots, U_{nn})^T$ and $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$. Therefore the variance-covariance matrix of U_n is

$$V_n(\lambda_0) = \mathrm{Var}(U_n) = \left(I_n - \lambda_0 W_n\right)^{-1} \left\{\left(I_n - \lambda_0 W_n\right)^T\right\}^{-1}, \quad n = 1, 2, \ldots$$

This matrix allows to describe the cross sectional spatial dependencies between the n observations. Furthermore, the fact that the diagonal elements of $V_n(\lambda_0)$ depend on λ_0 and particularly on i and

n allows some spatial heteroscedasticity. These spatial dependence and heretoscedasticity depend on the neighborhood structure established by the spatial weights matrix W_n.

Before going further, let us give some particular cases of the model.
If one consider i.i.d observations, that is $V_n(\lambda_0) = \sigma^2 I_n$, with σ depending on λ_0, the obtained model may be seen as a particularly case of the classical generalized partially linear models (e.g Severini & Staniswalis, 1994) or the classical generalized additive model (Hastie & Tibshirani, 1990). Several approaches of estimating this particular model have been developed, among others, we cite that of Severini & Staniswalis (1994), based on the concept of generalized profile likelihood (e.g Severini & Wong, 1992). This approach consists to first fix the parametric parameter β and estimate nonparametrically $g_0(\cdot)$ by using the weighted likelihood method. This last estimate is then used to construct a profile likelihood to estimate β_0.
When $g_0(\cdot) = 0$ (or is an affine function), that is without a nonparametric component, several approaches have been developed to estimate the parameters β_0 and λ_0. The basic difficulty encountered is that the likelihood function of this model involve a n dimensional normal integral, thus when n is high, the computation or asymptotic properties of the estimates may be difficult (e.g Poirier & Ruud, 1988). Various approaches have been proposed to addressed this difficulty, among these we cite:

- Feasible Maximum Likelihood approach: it consists of replacing the true likelihood function by a pseudo likelihood function constructed via marginal likelihood functions. Smirnov (2010) proposes a pseudo likelihood function obtained by replacing $V_n(\lambda_0)$ by some diagonal matrix by the diagonal elements of $V_n(\lambda_0)$. Alternatively, Wang et al. (2013) proposed to divide the observations by pairwise groups where the latter are assumed to be independent with bivariate normal distribution in each group and estimate β_0 and λ_0 by maximizing the likelihood of these groups.
- GMM approach used by Pinkse & Slade (1998). These authors used the generalized residuals defined by $\tilde{U}_{in}(\beta, \lambda) = E\left(U_{in}|Y_{in}, \beta, \lambda\right), i = 1, \ldots, n,\ n = 1, 2, \ldots$ with some instrumentals variables to construct moments equations to define GMM estimators of β_0 and λ_0.

In what follows, using the n observations (X_{in}, Y_{in}, Z_{in}), $i = 1, ..., n$, we propose parametric estimators of β_0, λ_0 and a nonparametric estimator of the smooth function $g_0(\cdot)$.
To this aim, assume that, for all $n = 1, 2, \ldots$, $\{\varepsilon_{in},\ i = 1 \ldots, n\}$ is independent of $\{X_{in},\ i = 1, \ldots, n\}$ and $\{Z_{in},\ i = 1, \ldots, n\}$, and $\{X_{in},\ i =, 1 \ldots, n\}$ is independent of $\{Z_{in},\ i = 1, \ldots, n\}$. We give asymptotic results according to *Increasing domain* asymptotic.

6.2.1 Estimation procedure

We propose an estimation procedure based on a combination of a weighted likelihood method and a generalized method of moments. We first fix the parametric components β and λ of the model and estimate the nonparametric component using a weighted likelihood. The obtained estimate is then used to construct generalized residuals where the latter are combined to instrumentals variables to propose GMM parametric estimates. This approach will be described as follow.

By equation (6.14) we have

$$E_0\left(Y_{in}|X_{in}, Z_{in}\right) = \Phi\left(\left(v_{in}(\lambda_0)\right)^{-1}\left(X_{in}^T\beta_0 + g_0(Z_{in})\right)\right),\ i = 1, \ldots, n,\ n = 1, 2, \ldots \tag{6.16}$$

where E_0 denotes the expectation under the true parameters (i.e β_0, λ_0 and $g_0(\cdot)$), $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, and $(v_{in}(\lambda_0))^2 = V_{iin}(\lambda_0),\ 1 \leq i \leq n,\ n = 1, 2, \ldots$ are the diagonal elements of $V_n(\lambda_0)$.
For each $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$, $z \in \mathcal{Z}$ and $\eta \in \mathbb{R}$, we define the conditional expectation on Z_{in} of the log-Likelihood of Y_{in} given (X_{in}, Z_{in}) for $1 \leq i \leq n,\ n = 1, 2, \ldots$, as

$$H(\eta; \beta, \lambda, z) = E_0\left(\mathcal{L}\left(\Phi\left(\left(v_{in}(\lambda)\right)^{-1}\left(\eta + X_{in}^T\beta\right)\right); Y_{in}\right)\Big|\, Z_{in} = z\right),$$

with $\mathcal{L}(u; v) = \log\left(u^v(1-u)^{1-v}\right)$. Note that $H(\eta; \beta, \lambda, z)$ is assumed to be constant over i (and n). For each fixed $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$ and $z \in \mathcal{Z}$, $g_{\beta,\lambda}(z)$ denotes the solution in η of

$$\frac{\partial}{\partial \eta} H(\eta; \beta, \lambda, z) = 0. \tag{6.17}$$

Then, we have $g_{\beta_0,\lambda_0}(z) = g_0(z)$ for all $z \in \mathcal{Z}$.

Now using $g_{\beta,\lambda}(\cdot)$, we construct GMM estimates of β_0 and λ_0 as Pinkse & Slade (1998). For that, we define the generalized residuals, replacing $g_0(Z_{in})$ in (6.13) by $g_{\beta,\lambda}(Z_{in})$:

$$\begin{aligned} \tilde{U}_{in}(\beta,\lambda,g_{\beta,\lambda}) &= E\left(U_{in}|Y_{in},\beta,\lambda\right) \\ &= \frac{\phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\left(Y_{in}-\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\right)}{\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\left(1-\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\right)}, \end{aligned} \tag{6.18}$$

where $\phi(\cdot)$ is the density of the standard normal distribution and
$G_{in}(\beta,\lambda,g_{\beta,\lambda}) = (v_{ni}(\lambda))^{-1}\left(X_{in}^T\beta + g_{\beta,\lambda}(Z_{in})\right)$.
For simplicity of notation, we write when it is possible $\theta = (\beta^T,\lambda)^T \in \Theta = \Theta_\beta \times \Theta_\lambda$.
Note that in (6.18), the generalized residual $\tilde{U}_{in}(\cdot\,,\cdot)$ is calculated by conditioning only on Y_{in} not on the entire sample $\{Y_{in},\, i = 1,2,\ldots,n,\, n = 1,\ldots\}$ or a subset of it. This of course will influence the efficiency of the estimators of θ obtained by these generalized residuals, but it allows to avoid a complex computation, see Poirier & Ruud (1988) for more details. To address this loss of efficiency, let us follow Pinkse & Slade (1998)'s procedure that consists of employing some instrumentals variables in order to create some moments conditions, and use some random matrix to define a criterion function. Both the instrumentals variables and the random matrix permit to take into account more informations about the spatial dependence and heteroscedasticity in the dataset. Let us now detail the estimation procedure.
Let

$$S_n(\theta,g_\theta) = n^{-1}\xi_n^T\tilde{U}_n(\theta,g_\theta), \tag{6.19}$$

where $\tilde{U}_n(\theta,g_\theta)$ is the $n\times 1$ vector, composed of $\tilde{U}_{in}(\theta,g_\theta)$, $1 \le i \le n$ and ξ_n is a $n\times q$ matrix of instrumentals variables whose ith row is denoted by the $1\times q$ random vector ξ_{in}. The latter may depend on $g_\theta(\cdot)$ and θ. We assume that ξ_{in} is $\sigma(X_{in},Z_{in})$ measurable for each $i = 1,\ldots,n,\ n = 1,2,\ldots$. We suppress the possible dependence of the instrumentals variables on the parameters for notational simplicity. The GMM approach consists to minimize the following sample criterion function,

$$Q_n(\theta,g_\theta) = S_n^T(\theta,g_\theta)M_nS_n(\theta,g_\theta), \tag{6.20}$$

where M_n is some positive-definite $q\times q$ weight matrix that may depend on sample information. The choice of the instrumentals variables and weight matrix characterizes the difference between GMM estimator and all pseudo maximum likelihood estimators. For instance, if one takes

$$\xi_{in}(\theta,g_\theta) = \frac{\partial G_{in}(\theta,\eta_i)}{\partial\theta} + \frac{\partial G_{in}(\theta,\eta_i)}{\partial\eta}\,\frac{\partial g_\theta}{\partial\theta}(Z_{in}), \tag{6.21}$$

with $\eta_i = g_\theta(Z_{in})$, $G_{in}(\theta,\eta_i) = (v_{in}(\lambda))^{-1}\left(X_{in}^T\beta+\eta_i\right)$, $M_n = I_q$ with $q = p+1$, then the GMM estimator of θ is equal to a pseudo maximum profile likelihood estimator of θ, accounting only the spatial heretoscedasticity.
Now, let

$$S(\theta,g_\theta) = \lim_{n\to\infty} E_0\left(S_n(\theta,g_\theta)\right), \tag{6.22}$$

and

$$Q(\theta,g_\theta) = S^T(\theta,g_\theta)MS(\theta,g_\theta),$$

where M, the limit of the sequence M_n, is a nonrandom positive definite matrix. The functions $S_n(\cdot,\cdot)$ and $Q_n(\cdot,\cdot)$ are viewed as empirical counterparts of $S(\cdot,\cdot)$ and $Q(\cdot,\cdot)$ respectively.
It is clear that $g_\theta(\cdot)$ is not available in practice. However, we need to estimate it, particularly by an asymptotically efficient estimate. By (6.17) and for fixed $\theta^T = (\beta^T,\lambda) \in \Theta$ an estimator of $g_\theta(z)$, for $z \in \mathcal{Z}$ can be given by $\hat{g}_\theta(z)$ which denotes the solution in η of

$$\sum_{i=1}^{n}\frac{\partial}{\partial\eta}\mathcal{L}\left(\Phi\left(G_{in}(\theta,\eta)\right);Y_{in}\right)K\left(\frac{z-Z_{in}}{b_n}\right) = 0 \tag{6.23}$$

where $K(\cdot)$ is a kernel from $\mathbb{R}^d$ to $\mathbb{R}_+$ and b_n is a bandwidth depending on n.

Now, replacing $g_\theta(\cdot)$ in (6.20) by the estimator $\hat{g}_\theta(\cdot)$ permits to obtain the GMM estimator $\hat{\theta}$ of θ as

$$\hat{\theta} = \text{argmin}_{\theta\in\Theta} Q_n(\theta, \hat{g}_\theta). \tag{6.24}$$

A classical inconvenience of the estimator $\hat{g}_\theta(z)$ proposed in (6.23) is that the bias of $\hat{g}_\theta(z)$ is high for z near the boundary of $\mathcal{Z}$. Of course this bias will effect the estimator of θ given in (6.24) when some of observations Z_{in} are near the boundary of $\mathcal{Z}$. Local linear method, or more generally, the local polynomial method (Fan & Gijbels, 1996) can be used to reduce this bias. Another alternative is to use *trimming* (Severini & Staniswalis, 1994) in which the function $S_n(\theta, g_\theta)$ is computed by using only observations associated to Z_{in} that are away from the boundary. The advantage of this approach is that the theoretical results can be presented in a clear form but it is less tractable from a practical point of view in particular for low sample sizes.

6.3 Large sample properties

We now turn to the asymptotic properties of the estimators derived in previous section; $\hat{\theta}^T = (\hat{\beta}^T, \hat{\lambda})$ and $\hat{g}_{\hat{\theta}}(\cdot)$. Let us use the following notations: $\frac{d}{d\theta}S(\theta, g_\theta)$ means that we differentiate $S(.,.)$ with respect to θ and $\frac{\partial}{\partial\theta}S(\theta, g_\theta)$ is the partial derivative of $S(\cdot,\cdot)$ with respect to the first variable. The partial derivative of $S_n(\theta, g)$ with respect to g, for any function $v \in \mathcal{G}$ is

$$\frac{\partial S_n}{\partial g}(\theta, g)(v) = n^{-1}\sum_{i=1}^{n} \xi_{in}\frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta_i)v(Z_{in}).$$

Let the following matrices needed in the asymptotic variance-covariance matrix of $\hat{\theta}$:

$$B_1(\theta_0) = \lim_{n\to\infty} E_0\left(nS_n\left(\theta_0, g_0\right) S_n^T\left(\theta_0, g_0\right)\right),$$

$$B_2(\theta_0) = \left\{\frac{d}{d\theta}S^T\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0}\right\} M \left\{\frac{d}{d\theta}S\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0}\right\},$$

with

$$\frac{d}{d\theta}S\left(\theta, g_\theta\right) = \frac{\partial S}{\partial\theta}\left(\theta, g_\theta\right) + \frac{\partial S}{\partial g}\left(\theta, g_\theta\right)\frac{\partial}{\partial\theta}g_\theta, \tag{6.25}$$

and

$$\Omega(\theta_0) = \left\{B_2(\theta_0)\right\}^{-1}\left\{\frac{d}{d\theta}S^T\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0}\right\} M B_1(\theta_0) M \left\{\frac{d}{d\theta}S\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0}\right\}\left\{B_2(\theta_0)\right\}^{-1}.$$

The following assumptions are required to establish the asymptotic results.

Assumption A1. (Smoothing condition). For each fixed $\theta \in \Theta$ and $z \in \mathcal{Z}$, let $g_\theta(z)$ denote the unique solution with respect to η of

$$\frac{\partial}{\partial\eta}H(\eta; \theta, z) = 0.$$

For any $\varepsilon > 0$ and $g \in \mathcal{G}$, there exists $\gamma > 0$ such that

$$\sup_{\theta\in\Theta, z\in\mathcal{Z}}\left|\frac{\partial}{\partial\eta}H(g(z); \theta, z)\right| \leq \gamma \qquad \Longrightarrow \qquad \sup_{\theta\in\Theta, z\in\mathcal{Z}}|g(z) - g_\theta(z)| \leq \varepsilon. \tag{6.26}$$

Assumption A2. (Local dependence). The density $f_{in}(\cdot)$ of Z_{in} exists, is continuous on $\mathcal{Z}$ uniformly on i and n and satisfies

$$\liminf_{n\to\infty}\ \inf_{z\in\mathcal{Z}}\frac{1}{n}\sum_{i=1}^{n} f_{in}(z)\ > 0. \tag{6.27}$$

The joint probability density $f_{ijn}(.,.)$ of (Z_{in}, Z_{jn}) exists and is bounded on $\mathcal{Z} \times \mathcal{Z}$, uniformly on $i \neq j$ and n.

Assumption A3. (Spatial dependence). Let $h_{in}^{\theta,\eta_i}(\cdot|\cdot,\cdot)$ denote the conditional log likelihood function of Y_{in} given (X_{in}, Z_{in}), where $\eta_i = g(Z_{in})$. Let T_{in} be the vector (Y_{in}, X_{in}, Z_{in}), $i = 1, \ldots, n$, $n = 1, 2 \ldots$, $\tilde{p} = p + 1$, and assume that for all i, $l = 1, \ldots, n$,

$$|\mathrm{Cov}_0\left(\psi(T_{in}), \psi(T_{ln})\right)| \leq \left\{\mathrm{Var}_0\left(\psi(T_{in})\right) \mathrm{Var}_0\left(\psi(T_{ln})\right)\right\}^{1/2} \alpha_{iln}, \tag{6.28}$$

with

$$\psi(T_{in}) = K\left(\frac{z - Z_{in}}{b_n}\right) \quad \text{or} \quad \psi(T_{in}) = K\left(\frac{z - Z_{in}}{b_n}\right) \frac{\partial^{j_1+\cdots+j_{\tilde{p}}+r}}{\partial\theta_1^{j_1} \cdots \partial\theta_{\tilde{p}}^{j_{\tilde{p}}} \partial\eta^r} h_{in}^{\theta,\eta}(Y_{in}|\, X_{in},\, Z_{in} = z),$$

for all $z \in \mathcal{Z}$, $\theta \in \Theta, \eta = g(z)$ with $g \in \mathcal{G}$, and for all nonnegative integers $j_1, \ldots, j_{\tilde{p}} = 0, 1, 2$ and $r = 0, \ldots, 4$, such that $j_1 + \cdots + j_{\tilde{p}} + r \leq 6$.

We assume that

$$\begin{aligned} &\left|\mathrm{Cov}_0\left(\xi_{itn}\tilde{U}_{in}(\theta, g_\theta), \xi_{jsn}\tilde{U}_{jn}(\theta, g_\theta)\right)\right| \\ &\qquad \leq \left\{\mathrm{Var}_0\left(\xi_{itn}\tilde{U}_{in}(\theta, g_\theta)\right) \mathrm{Var}_0\left(\xi_{jsn}\tilde{U}_{jn}(\theta, g_\theta)\right)\right\}^{1/2} \alpha_{ijn}, \end{aligned} \tag{6.29}$$

for all $\theta \in \Theta$, $i, j = 1, \ldots, n$, $n = 1, 2, \ldots$ and for any $s, t = 1, \ldots, q$,

and

$$\begin{aligned} &\left|\mathrm{Cov}_0\left(\xi_{in}^{(2)}(\theta_0, \eta_i^0), \xi_{jn}^{(2)}(\theta_0, \eta_j^0)\right)\right| \\ &\qquad \leq \left\{\mathrm{Var}_0\left(\xi_{in}^{(2)}(\theta_0, \eta_i^0)\right) \mathrm{Var}_0\left(\xi_{jn}^{(2)}(\theta_0, \eta_j^0)\right)\right\}^{1/2} \alpha_{ijn}, \end{aligned} \tag{6.30}$$

with

$$\xi_{in}^{(2)}(\theta_0, \eta_i^0) = w^T \xi_i \Lambda\left(G_{in}(\theta_0, \eta_i^0)\right) \phi\left(G_{in}(\theta_0, \eta_i^0)\right) \frac{\partial G_{in}}{\partial\theta}(\theta_0, \eta_i^0),$$

where $\eta_i^0 = g_0(Z_i)$ for each $w \in \mathbb{R}^q$ such that $\|w\| = 1$, and $\mathrm{Var}_0(\cdot)$ denotes the variance under the trues parameters.

In addition, assume that there is a decreasing (to 0) positive function $\varphi(\cdot)$ such that $\alpha_{ijn} = O\left(\varphi\left(\|s_i - s_j\|\right)\right)$, $r^2\varphi(rr^*)/\varphi(r^*) = o(1)$, as $r \to 0$, for all fixed $r^* > 0$, where s_i and s_j are spatial coordinates associated to observations i and j respectively.

Assumption A4. The kernel K satisfies $\int K(u)du = 1$. It is Lipshitzian, i.e there is a positive constant C such that

$$|K(u) - K(v)| \leq C\|u - v\| \qquad \text{for all } \; u, v \in \mathbb{R}^d.$$

Assumption A5. The bandwidth b_n satisfies $b_n \to 0$ and $nb_n^{3d+1} \to \infty$ as $n \to \infty$.

Assumption A6. The instrumentals variables satisfy $\sup_{i,\,n} \|\xi_{in}\| = O_p(1)$ where ξ_{in} is the i-th column of the $n \times q$ matrix of instrumentals variables ξ_n.

Assumption A7. $\theta^T = (\beta^T, \lambda)$ takes values in a compact and convex set $\Theta = \Theta_\beta \times \Theta_\lambda \subset \mathbb{R}^p \times \mathbb{R}$ and $\theta_0^T = (\beta_0^T, \lambda_0)$ is in the interior of Θ.

Assumption A8. $S(\cdot,\cdot)$ is continuous on both arguments θ and g, and $Q(\cdot, g_\cdot)$ attains a unique minimum over Θ at θ_0.

Assumption A9. The square root of diagonal's elements of $V_n(\lambda)$ are twice continuous differentiable functions with respect to λ and $\sup_{\lambda\in\Theta_\lambda} \left|v_{in}^{-1}(\lambda) + \frac{d}{d\lambda}v_{in}(\lambda) + \frac{d^2}{d\lambda^2}v_{in}(\lambda)\right| < \infty$, uniformly on i and n.

Assumption A10. $B_1(\theta_0)$ and $B_2(\theta_0)$ are positive definite matrices, and $M_n - M = o_p(1)$.

Remark 6.1. *Assumption A1 ensures smoothness of $H(.;.,.)$ around its extrema point $g_\theta(.)$, see Severini & Staniswalis (1994). Assumption A2 is a decay of local independence condition of the covariates Z_{in}, meaning that these variables are not identically distributed, a similar condition can be find in Robinson (2011). Condition (6.27) generalizes the classical assumption $\inf_z f(z) > 0$ used in the case of estimating the density function $f(\cdot)$ with identically distributed or stationary random variables. This assumption has been used in Robinson (2011) (**Assumption A7**(x), p. 8). Assumption A3 describes the spatial dependence structure. The processes we use are not assumed stationary, this allows more generally and the dependence structure to change with the sample size n (see Pinkse & Slade (1998) for more discussion). Conditions (6.28),(6.29) and (6.30) are not restrictives. When the regressors and instrumentals variables are deterministic then conditions (6.28) and (6.29) are equivalent to $|Cov_0(Y_{in}, Y_{ln})| \leq \alpha_{iln}$. The condition on $\varphi(\cdot)$ is satisfied when this last tends to zero at a polynomial rate, i.e $\varphi(t) = O(t^{-\tau})$, for all $\tau > 2$ as in case of mixing random variables.*
Assumption A6 requires that the instruments to be bounded uniformly on i and n. In addition, when the instruments depend on θ and $g(\cdot)$, they are also uniformly bounded with respect to these parameters. The compactness condition in Assumption A7 is standard and the convexity is somewhat unusual, but is reasonable in most applications. Condition A8 is necessary to ensure identification of the true parameters θ_0. Assumption A9 requires the standard deviations of the errors to be uniformly bounded away from zero with bounded derivatives. it has been considered by Pinkse & Slade (1998). Assumption A10 is classic (Pinkse & Slade, 1998) required in the proof of Theorem 6.2. These last authors noted that in their model (without a nonparametric component) when the autoregressive parameter $\lambda_0 = 0$, then $B_2(\theta_0)$ is not invertible, regardless to the choice of M_n. This is also the case in our context, since for each $g_\theta(z)$ solution of (6.17), $\theta \in \Theta$ and $z \in \mathcal{Z}$, we have

$$\frac{\partial g_\theta}{\partial \beta}(z) = -\frac{E\left(\Gamma_{jn}(\theta, g_\theta(z))X_{jn} \middle| Z_{jn} = z\right)}{E\left(\Gamma_{jn}(\theta, g_\theta(z)) \middle| Z_{jn} = z\right)},$$

and

$$\begin{aligned}\frac{\partial g_\theta}{\partial \lambda}(z) &= \frac{v_{jn}^{'}(\lambda)}{v_{jn}(\lambda)} \frac{E\left(\Gamma_{jn}(\theta, g_\theta(z))\left(X_{jn}^T\beta + g_\theta(z)\right) \middle| Z_{jn} = z\right)}{E\left(\Gamma_{jn}(\theta, g_\theta(z)) \middle| Z_{jn} = z\right)} \\ &= \frac{v_{jn}^{'}(\lambda)}{v_{jn}(\lambda)} \left(g_\theta(z) - \beta^T \frac{\partial g_\theta}{\partial \beta}(z)\right),\end{aligned}$$

where $v_{jn}^{'}(\lambda) = \frac{d}{d\lambda} v_{jn}(\lambda) = v_{jn}(\lambda)\left[W_n S_n^{-1}(\lambda) V_n(\lambda)\right]_{jj}$,

$$\Gamma_{jn}(\cdot) = \Lambda^{'}\left(G_{jn}(\cdot)\right)\left[Y_{jn} - \Phi(G_{jn}(\cdot))\right] - \Lambda\left(G_{jn}(\cdot)\right)\phi\left(G_{jn}(\cdot)\right)$$

and $\Lambda(\cdot) = \phi(\cdot)/(1 - \Phi(\cdot))\Phi(\cdot)$. *However,* $v_{jn}^{'}(\lambda_0) = 0$ *if* $\lambda_0 = 0$, *then* $B_2(\theta_0)$ *will be singular.*

With these assumptions in place, we are able to give some asymptotic results. The weak consistencies of the proposed estimators are given in the following two results. The first theorem and corollary below establish the consistency of our estimators, whereas the second theorem deals with the question of convergence to normal distribution of the parametric component when it is properly standardized.

Theorem 6.1. *Under assumptions A1-A10, we have*

$$\hat{\theta} - \theta_0 = o_p(1).$$

Corollary 6.1. *If the assumptions of Theorem 6.1 are satisfied, then we have*

$$\left\|\hat{g}_{\hat{\theta}} - g_0\right\| = o_p(1).$$

Proof of Corollary 6.1 Note that

$$\begin{aligned}\left\|\hat{g}_{\hat{\theta}} - g_0\right\| &\leq \|\hat{g}_{\hat{\theta}} - g_{\hat{\theta}}\| + \|g_{\hat{\theta}} - g_0\| \\ &\leq \sup_\theta \|\hat{g}_\theta - g_\theta\| + \sup_\theta \left\|\frac{\partial g_\theta}{\partial \theta}\right\| \|\hat{\theta} - \theta_0\| = o_p(1),\end{aligned}$$

since, by the assumptions of Theorem 6.1, $\sup_\theta \|\hat{g}_\theta - g_\theta\| = o_p(1)$ and $\sup_\theta \left\|\frac{\partial g_\theta}{\partial \theta}\right\| < \infty$.
The following gives an asymptotic normality result of $\hat{\theta}$.

Theorem 6.2. *Under assumptions A1-A10, we have*

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \to \mathcal{N}\left(0, \Omega(\theta_0)\right)$$

Remark 6.2. *In practice, the previous results can be used to construct asymptotic confidence intervals and make hypotheses tests when a consistent estimation of the asymptotic covariance matrix $\Omega(\theta_0)$ is founded. We follow the idea of Pinkse & Slade (1998) to define some estimator of this matrix without establishing its consistence. Let $\Omega(\theta_0)$ be estimated by*

$$\Omega_n(\hat{\theta}) = \left\{B_{2n}(\hat{\theta})\right\}^{-1} \left\{ \frac{d}{d\theta} S_n^T\left(\theta, \hat{g}_\theta\right)\Big|_{\theta=\hat{\theta}} \right\} M_n B_{1n}(\hat{\theta}) M_n \left\{ \frac{d}{d\theta} S_n\left(\theta, \hat{g}_\theta\right)\Big|_{\theta=\hat{\theta}} \right\} \left\{B_{2n}(\hat{\theta})\right\}^{-1},$$

with

$$B_{1n}(\theta) = nS_n(\theta, \hat{g}_\theta) S_n^T(\theta, \hat{g}_\theta) \qquad \text{and} \qquad B_{2n}(\theta) = \left\{ \frac{d}{d\theta} S_n^T\left(\theta, \hat{g}_\theta\right) \right\} M_n \left\{ \frac{d}{d\theta} S_n\left(\theta, \hat{g}_\theta\right) \right\}.$$

The consistency of $\Omega_n(\hat{\theta})$ is based on theirs of $B_{1n}(\hat{\theta})$ and $B_{2n}(\hat{\theta})$ as estimators of $B_1(\theta_0)$ and $B_2(\theta_0)$ respectively. Note that the consistence of $B_{2n}(\hat{\theta})$ is relative easily to be established while that of $B_{1n}(\hat{\theta})$ asked some adaption of the proof of Theorem 3 of (Pinkse & Slade, 1998, p.134) in our case.

6.4 Computation of the estimates

The aim of this section is to outline in details how the regression parameters β, the spatial autocorrelation parameter λ and the non-linear function g_θ can be estimated. We begin with the computation of $\hat{g}_\theta(z)$ that will play a crucial role in what follows.

6.4.1 Computation of the estimate the nonparametric component

An iterative method is needed to compute $\hat{g}_\theta(z)$ solution of (6.23) for each fixed $\theta \in \Theta$ and $z \in \mathcal{Z}$. For fixed $\theta^T = (\beta, \lambda) \in \Theta$ and $z \in \mathcal{Z}$, let $\eta_\theta = g_\theta(z)$ and $\psi(\eta; \theta, z)$ denote the left hide side in (6.23), that can be rewritten as

$$\psi(\eta; \theta, z) = \sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-1} \Lambda\left(G_{in}(\theta, \eta)\right) \left[Y_{in} - \Phi(G_{in}(\theta, \eta))\right] K\left(\frac{z - Z_{in}}{b_n}\right).$$

Consider the fisher information:

$$\begin{aligned}
\Psi(\eta_\theta; \theta, z) &= E_0\left(\frac{\partial}{\partial \eta} \psi(\eta; \theta, z)\Big|_{\eta=\eta_\theta} \Big| \{(X_{in}, Z_{in}),\, 1 \leq\ i \leq n,\, n = 1, \ldots\} \right) \\
&= -\sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-2} \Lambda\left(G_{in}(\theta, \eta_\theta)\right) \phi\left(G_{in}(\theta, \eta_\theta)\right) K\left(\frac{z - Z_{in}}{b_n}\right) \\
&\quad + \sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-2} \Lambda'\left(G_{in}(\theta, \eta_\theta)\right) \\
&\quad \times \left[\Phi\left(G_{in}(\theta_0, \eta_0)\right) - \Phi\left(G_{in}(\theta, \eta_\theta)\right)\right] K\left(\frac{z - Z_{in}}{b_n}\right). \qquad (6.31)
\end{aligned}$$

Note that the second term in the right hand-side in (6.31) is negligible when θ is near to the true parameter θ_0.
Since $\psi(\eta; \theta, z) = 0$ for $\eta = \hat{g}_\theta(z)$, an initial estimate $\tilde{\eta}$ can be updated to $\eta^\dagger$ using Fisher's scoring method:

$$\eta^\dagger = \tilde{\eta} - \frac{\psi(\tilde{\eta}; \theta, z)}{\Psi(\tilde{\eta}; \theta, z)}. \qquad (6.32)$$

The iterated procedure (6.32) requests some starting value $\tilde{\eta} = \tilde{\eta}_0$ in order to ensure convergence of the algorithm. For that, let us adapt the approach of Severini & Staniswalis (1994), that consists

to suppose that for fixed $\theta \in \Theta$, there exists a $\tilde{\eta}_0$ satisfying $G_{in}(\theta, \tilde{\eta}_0) = \Phi^{-1}(Y_{in})$ for $i = 1, \ldots, n$. Knowing that $G_{in}(\theta, \tilde{\eta}_0) = (v_{in}(\lambda))^{-1}\left(X_{ni}^T\beta + \tilde{\eta}_0\right)$, we have $\tilde{\eta}_0 = v_{in}(\lambda)\Phi^{-1}(Y_{in}) - X_{in}^T\beta$, then (6.32) can be updated using the following first value:

$$\eta_0^{\dagger} = \tilde{\eta}_0 - \frac{\psi(\tilde{\eta_0}; \theta, z)}{\Psi(\tilde{\eta_0}; \theta, z)} = \frac{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-1} \Lambda(C_{in})\phi(C_{in}) \left[C_{in} - \left[v_{in}(\lambda)\right]^{-1} X_{in}^T\beta\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} \Lambda(C_{in})\phi(C_{in}) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

where $C_{in} = \Phi^{-1}(Y_{in})$, $i = 1, \ldots, n$ is computed using a slight adjustment since $Y_{in} \in \{0, 1\}$. With this first value, the algorithm is iterates until convergence.

6.4.1.1 Selection of the bandwidth

A critical step (in non or semi parametric models) is the choice of the bandwidth parameter b_n which is usually selected by applying some cross validation approach. The latter was adapted by Su (2012) in the case of a spatial semiparametric model. Since cross-validation may be very time consuming, it is particularly the case of our model, we adapt the following approach used in Severini & Staniswalis (1994) for more flexibility:

1. Consider the linear regression of C_{in} on X_{in}, $i = 1, \ldots, n$ without an intercept term, and let $R_{1n}, \ldots, R_{nn}$ denote the corresponding residuals.
2. Since we expect $E(R_{in}|Z_{in} = z)$ to have similar smoothness properties as $g_0(.)$, the optimal bandwidth b_n is that of the nonparametric regression of the $\{R_{in}\}_{i=1,\cdots,n}$ on $\{Z_{in}\}_{i=1,\cdots,n}$, chosen by applying any nonparametric regression bandwidth selection method. For that, we use the cross-validation method by *np* R Package.

6.4.2 Computation of $\hat{\theta}$

The parametric component β and the spatial autoregressive parameter λ are computed as mentioned above, by a GMM approach based on some instrumentals variables ξ_n and the weight matrix M_n. The choice of these instrumentals variables and weight matrix M_n are as follow.
Since $\psi(\hat{g}_\theta(z); \theta, z) = 0$, if we differentiate the latter with respect to β and λ, we have

$$\frac{\partial}{\partial\beta}\hat{g}_\theta(z) = -\frac{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} \Delta_{in}(\theta, z) X_{in} K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

and

$$\begin{aligned}\frac{\partial}{\partial\lambda}\hat{g}_\theta(z) &= \frac{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-1} v_{in}^{'}(\lambda)\Delta_{in}(\theta, z)\left[X_{in}^T\beta + \hat{g}_\theta(z)\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)} \\ &+ \frac{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} v_{in}^{'}(\lambda)\Lambda\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\left[Y_{in} - \Phi\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n \left[v_{in}(\lambda)\right]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)},\end{aligned}$$

with

$$\Delta_{in}(\theta, z) = \Lambda^{'}\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\left[Y_{in} - \Phi\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\right] - \Lambda\left(G_{ni}(\theta, \hat{g}_\theta(z))\right)\phi\left(G_{in}(\theta, \hat{g}_\theta(z))\right).$$

Then, the previous result is used to define the following instrumentals variables

$$\xi_{in}(\theta, \hat{g}_\theta) = \frac{\partial G_{in}(\theta, \hat{\eta}_i)}{\partial\theta} + \frac{\partial G_{in}(\theta, \hat{\eta}_i)}{\partial\eta}\frac{\partial}{\partial\theta}\hat{g}_\theta(Z_{in}),$$

with $\hat{\eta}_i = \hat{g}_\theta(Z_{in})$.
For the weight matrix, we use (as in Pinkse & Slade, 1998) $M_n = I_q$ with $q = p + 1$. Then the obtained GMM estimator of θ with this choice of M_n is equal to the pseudo profile maximum likelihood estimator of θ, accounting only the spatial heretoscedasticity.
The final step is to plug the GMM estimator $\hat{\theta}$ to have $\hat{g}_{\hat{\theta}}$.

6.5 Simulation study

In this section, we study the performance of the proposed model based on some numerical results, which highlight the importance of considering the spatial dependence and the partial linearity. We simulated some semiparametric models and estimate them by our proposed method, the one that does not account the spatial dependence (using the same estimation procedure above, based on Partially linear probit model (PLPM)) and by a full linear SAE probit (LSAEP) method. The latter method can account the spatial dependence but ignores the partial linearity. *ProbitSpatial* R package (Martinetti & Geniaux, 2016) was be used to provide estimates for the LSAEP model. We generate observations from the following spatial latent partial linear model:

$$Y^*_{in} = \beta_1 X^{(1)}_{in} + \beta_2 X^{(2)}_{in} + g(Z_{in}) + U_{in}; \qquad Y_{in} = \mathbb{I}(Y^*_{in} \geq 0),\ i = 1, \ldots, n \tag{6.33}$$

$$U_n = (I_n - \lambda W_n)^{-1} \varepsilon_n \tag{6.34}$$

where $U_n \sim \mathcal{N}(0, I_n)$ and W_n is the spatial weight matrix associated to n locations chosen randomly in a 60×60 regular grid, based on 6 nearest neighbors of each unit. In order to observe the effect of partial linearity when we compare our estimation procedure to that based on LSAEP models, we will consider the two following cases:

Case 1: The explanatory variables $X^{(1)}$ and $X^{(2)}$ are generated as pseudo $\mathcal{B}(0.7)$ and $\mathcal{U}[-2, 2]$ respectively, and the other explanatory variable Z is equal to the sum of 48 independent random variables, each uniformly distributed over $[-0.25, 0.25]$. We use here the non-linear function $g(t) = t + 2\cos(0.5\pi t)$.

Case 2: The explanatory variables $X^{(1)}$, $X^{(2)}$ and Z are generated as pseudo $\mathcal{N}(0, 1)$ and we considerer the linear function $g(t) = 1 + 0.5t$.

We take $\beta_1 = -1$, $\beta_2 = 1$ and different values of the spatial parameter λ; that is $\lambda \in \{0.2, 0.5, 0.8\}$. The bandwidth b_n is selected by using Severini & Staniswalis (1994)'s approach detailed previously with $C_{ni} = \Phi^{-1}\left(0.9Y_{ni} + 0.1(1 - Y_{ni})\right), i = 1, \ldots, n$. A Gaussian kernel will be considered; $K(t) = (2\pi^{-1/2})\exp(-t^2/2)$. As mentioned above, the instrumentals variables are the trivial choice and the weight matrix $M_n = I_3$ is the identity one.

The two studied cases are replicated 200 times for a sample size $n = 200$, the results are presented in Tables 6.1, 6.2. In each table, the columns titled Mean, Median and SD give the average, the median and the standard deviation over these 200 replications associated to each method of estimation, respectively.

Firstly, when we compare the estimators based on our approach (PLSPM) with those based on LSAEP model, we notice that the latter yields a more biased estimators of the coefficients β_1 and β_2 particularly in Case 1. It makes sense that ignoring the partial linearity (see also Figure 6.1) weakens the quality of estimation of the coefficients β_1 and β_2. While in Case 2, these two approaches yield similar results in term of consistency but our approach seems to be less efficient.

Secondly, note that for the two cases (Table 6.1 and Table 6.2) LSAEP and PLPM estimates are similar in case of low spatial dependence ($\lambda = 0.2$). However, this is not the case for large spatial dependence ($\lambda = 0.8$) framework where in this case the estimation procedure based on PLPM models yields inconsistent estimates of the parameters β_1 and β_2 and the smooth function $g(\cdot)$ (see the right panel in Figure 6.1). It makes sense that considering the spatial dependence allows to find consistent estimates of the coefficients β_1, β_2 and the smooth function $g(\cdot)$.
Note that our approach is less efficient, this can be observed when observing the differences between the mean and the median (or the high values of the standard deviation) associated to our estimators in tables 6.1-6.2. However this is eventually due to the use of GMM approach with the trivial choice of weight matrix $M_n = I_n$. In addition, when estimating the spatial parameter λ, our procedure yields biased estimators, this may be related to the considered choice of instrumentals variables. Better choices of the weight matrix and the instrumentals variables have to be investigated in future research. [ht!]

Table 6.1: Case 1 with $n = 200$ and 200 replications.

λ	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.20	PLSPM	-1.08	-1.00	0.53	1.07	0.99	0.33	0.09	0.00	0.29
	LSAEP	-0.67	-0.69	0.25	0.67	0.66	0.11	-0.04	0.02	0.36
	PLPM	-0.98	-0.99	0.32	0.98	0.96	0.15			
0.50	PLSPM	-1.13	-0.96	0.67	1.08	0.98	0.40	0.27	0.10	0.37
	LSAEP	-0.65	-0.64	0.24	0.66	0.65	0.12	0.20	0.26	0.29
	PLPM	-0.90	-0.88	0.30	0.90	0.89	0.15			
0.80	PLSPM	-1.12	-0.86	0.86	1.08	0.89	0.55	0.53	0.71	0.39
	LSAEP	-0.57	-0.56	0.25	0.61	0.60	0.12	0.60	0.61	0.10
	PLPM	-0.65	-0.66	0.25	0.65	0.63	0.13			

$\lambda = 0.2$ $\lambda = 0.5$ $\lambda = 0.8$

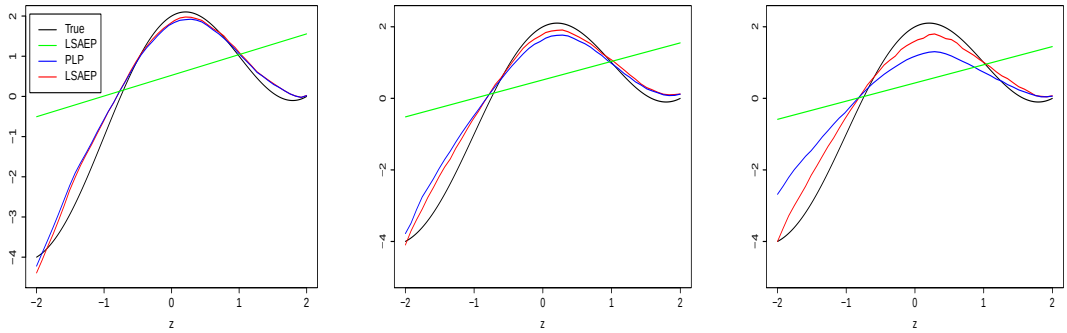

Figure 6.1: Case 1 with $n = 200$ and 200 replications.

Table 6.2: Case 2 with $n = 200$ and 200 replications.

λ	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.20	PLSPM	-1.12	-1.05	0.32	1.13	1.06	0.30	0.26	0.05	0.31
	LSAEP	-1.08	-1.06	0.19	1.09	1.07	0.20	0.02	0.17	0.47
	PLPM	-1.00	-0.99	0.20	0.99	0.98	0.14			
0.50	PLSPM	-1.08	-1.03	0.37	1.06	0.99	0.31	0.30	0.18	0.31
	LSAEP	-1.06	-1.06	0.21	1.05	1.01	0.19	0.40	0.48	0.29
	PLPM	-0.95	-0.94	0.21	0.93	0.91	0.18			
0.80	PLSPM	-1.02	-0.91	0.44	1.01	0.86	0.43	0.56	0.68	0.35
	LSAEP	-0.88	-0.87	0.19	0.87	0.86	0.20	0.72	0.73	0.09
	PLPM	-0.66	-0.65	0.15	0.66	0.65	0.16			

Discussion

In this manuscript, we have proposed a spatial semiparametric probit models for identifying risk factors at onset and spatial heterogeneity. Parameters involved in the models are estimated using weighted likelihood and generalized method of moments methods. The technique based on dependent random arrays facilitates the estimation and derivation of asymptotic properties which, otherwise would have been difficult due to the complexity added by the spatial dependence introduces into the model and high dimensional integration required by a full maximum likelihood approach. Moreover, it yields consistent estimates through proper choices of bandwidth, weight matrix, instrumentals variables. The proposed models provide a general framework and tools for researchers and practitioners when dealing with binary semiparametric choice models in the presence of spatial correlation. Though they provide significant contribution to the body of knowledge,

more need to be done to the best of our knowledge.
As indicated, the weights are used to improve efficiency and convergence. It would be interesting to develop criteria for choices of optimal weights leading to better performance. For instance, it may be improved by choosing for instance a weight matrix M_n as a consistent estimator $B_{1n}(\hat{\theta})$ of the matrix $B_1(\theta_0)$. Another empirical choice could be the idea of continuous updating GMM estimator (One step GMM) used in Pinkse et al. (2006):

$$M_n(\theta) = n^{-1} \sum_{i,j=1}^{n} \delta_{ij} \xi_{ni} \xi_{jn}^T \tilde{U}_{in}(\theta, \hat{g}_\theta) \tilde{U}_{jn}(\theta, \hat{g}_\theta)$$

with weights

$$\delta_{ij} = \frac{\sum_{r=1}^{n} \tau_{ri}\tau_{rj}}{\left[\sum_{r=1}^{n} \tau_{ri}^2 \sum_{r=1}^{n} \tau_{rj}^2\right]^{1/2}} \qquad \text{for } i,j = 1,\dots,n,$$

where τ_{ij} is a number depending on w_{ijn}. The nearer location i is to j, the larger is τ_{ij}.
Another topic of future research is allowing some spatial dependency on the covariates (SAR models) and the response (endogenous models) for more generality. These topics will be of interest in future research.

6.6 Appendix

Proposition 6.1. *Under assumptions A1-A6, for $\theta \in \Theta$ and $z \in \mathcal{Z}$, the functions $g_\theta(z)$ and $\hat{g}_\theta(z)$, solutions of (6.17) and (6.23) respectively, satisfy*

1. *for all $i,j = 0,1,2,\ i+j \le 2$,*

$$\frac{\partial^{i+j}}{\partial\theta_l^i \partial\theta_r^j} g_\theta(z) \qquad \text{and} \qquad \frac{\partial^{i+j}}{\partial\theta_l^i \partial\theta_r^j} \hat{g}_\theta(z) \qquad \textit{exist and are finite for all } 1 \le l,r \le p+1.$$

2. $\sup_{\theta\in\Theta} \|\hat{g}_\theta - g_\theta\|$, $\sup_{\theta\in\Theta} \max_{j=1,\dots,p+1} \left\| \frac{\partial}{\partial\theta_j}(\hat{g}_\theta - g_\theta) \right\|$ *and* $\sup_{\theta\in\Theta} \max_{1\le i,j\le p+1} \left\| \frac{\partial^2}{\partial\theta_i\partial\theta_j}(\hat{g}_\theta - g_\theta) \right\|$,

are all of order $o_p(1)$ as $n \to \infty$.

Without loss of generality, the proof of this proposition is ensured by Lemma 6.2 in the univariate case i.e Θ, $\mathcal{Z} \subset \mathbb{R}$.

The following lemma is useful in the proof of Lemma 6.2. It is an extension of Lemma 8 in Severini & Wong (1992) to spatial dependent data.

Lemma 6.1. *Let $\zeta_\theta(Y_{in})$ denote a scalar function of Y_{in}, $i = 1,\dots,n,\ n = 1,2,\dots$ depending on a scalar parameter $\theta \in \Theta$ and let for $j = 0,1,2$*

$$\zeta_\theta^{(j)}(Y_{in}) = \frac{\partial^j}{\partial\theta^j} \zeta_\theta(Y_{in}), \qquad i = 1,\dots,n,\ n = 1,2,\dots$$

Let $f_i(\cdot)$ denote the density of Z_{in} (given in assumption A2) and let $\bar{f}(z) = \frac{1}{n}\sum_{i=1}^{n} f_i(z)$. Assume that:

H.1 $\sup_\theta \sup_{1\le i\le n,n} \left|\zeta_\theta^{(j)}(Y_{in})\right| < \infty$ *for $j = 0,\dots,3$.*

H.2 *For all $\theta \in \Theta$, $j = 0,1,2$, and $1 \le i,l \le n$:*

$$\left|\mathrm{Cov}\left(K_{in}(z), K_{ln}(z)\right)\right| \le \left\{\mathrm{Var}(K_{in}(z))\mathrm{Var}(K_{in}(z))\right\}^{1/2} \varphi\left(\|s_i - s_l\|\right), \tag{6.35}$$

$$\left|\mathrm{Cov}\left(\zeta_\theta^{(j)}(Y_{in})K_{in}(z), \zeta_\theta^{(j)}(Y_{ln})K_{ln}(z)\right)\right| \le \left\{\mathrm{Var}\left(\zeta_\theta^{(j)}(Y_{in})K_{in}(z)\right)\mathrm{Var}\left(\zeta_\theta^{(j)}(Y_{ln})K_{ln}(z)\right)\right\}^{1/2} \varphi\left(\|s_i - s_l\|\right), \tag{6.36}$$

with $K_{in}(z) = K\left((z - Z_{in})/b\right)$.

Let $m_\theta(z) = E\left(\zeta_\theta(Y_{in})|Z_{in} = z\right)$, *for* $z \in \mathcal{Z}$, *and assume that* $\dfrac{\partial^j}{\partial\theta^j} m_\theta(\cdot)$ *is continuous on* $\mathcal{Z}$, $j = 0, 1, 2$.

For each fixed $\theta \in \Theta$ *and* $z \in \mathcal{Z}$, *let the kernel estimator* $\widehat{m}_\theta(z)$ *of* $m_\theta(z)$ *be defined by*

$$\widehat{m}_\theta(z) = \frac{\sum_{i=1}^n \zeta_\theta(Y_{in}) K_{in}(z)}{\sum_{i=1}^n K_{in}(z)}.$$

If assumptions A2, A4, and A5 are satisfied, then

$$\sup_{\theta\in\Theta}\sup_{z\in\mathcal{Z}} \left| \frac{\partial^j}{\partial\theta^j}\widehat{m}_\theta(z) - \frac{\partial^j}{\partial\theta^j} m_\theta(z) \right| = o_p(1),$$

for $j = 0, 1, 2$.

Proof of Lemma 6.1

We give the proof in the case where $j = 0$, corresponding to the study of the uniform consistency of the kernel estimator of the regression function of $\zeta_\theta(Y_{in})$ on Z_{in}. The other cases are similarly to this last and then are omitted.
Let

$$\widehat{v}_\theta(z) = \frac{1}{nb^d}\sum_{i=1}^n \zeta_\theta(Y_{in}) K_{in}(z); \qquad \widehat{f}(z) = \frac{1}{nb^d}\sum_{i=1}^n K_{in}(z),$$

$$v_\theta(z) = m_\theta(z)\bar{f}(z).$$

We have to show that

$$\sup_\theta \sup_z |\widehat{v}_\theta(z) - v_\theta(z)| = o_p(1) \tag{6.37}$$

and

$$\sup_z \left|\widehat{f}(z) - \bar{f}(z)\right| = o_p(1) \tag{6.38}$$

We give the proof of (6.37), that of (6.38) is similar.

Asymptotic behavior of $|\widehat{v}_\theta(z) - v_\theta(z)|$

Let us first consider the bias $|E(\widehat{v}_\theta(z)) - v_\theta(z)|$. We have

$$\begin{aligned} E(\widehat{v}_\theta(z)) &= (nb^d)^{-1}\sum_{i=1}^n \int K\left(\frac{z-u}{b}\right) m_\theta(u) f_i(u) du \\ &= b^{-d}\int v_\theta(u) K\left(\frac{z-u}{b}\right) du, \\ &= \int v_\theta(z - bu) K(u) du \end{aligned}$$

thus

$$E(\widehat{v}_\theta(z)) - v_\theta(z) = \int \left(v_\theta(z - bu) - v_\theta(z)\right) K(u) du = o(1)$$

by assumption A4, the continuity of $f_i(\cdot)$ (see A2) and $m_\theta(\cdot)$, and the compactness of $\mathcal{Z}$. It is clear that the bias term does not depend on θ and z.
Let us now treat $|\widehat{v}_\theta(z) - E(\widehat{v}_\theta(z))|$. Consider the sum of variances

$$\mathbf{S}_n = (nb^d)^{-2}\sum_{i=1}^n \mathrm{Var}\left(\zeta_\theta(Y_{in}) K_{in}(z)\right).$$

We have

$$\begin{aligned}\mathrm{Var}\left(\zeta_\theta(Y_{in})K_{in}(z)\right) &\leq E\left(\zeta_\theta^2(Y_{in})K_{in}^2(z)\right)\\ &\leq CE\left(K_{in}^2(z)\right) = Cb^d\int K^2(u)f_i(z-ub)du\\ &= Cb^d\sup_u|K(u)|^2\int f_i(z-ub)du = Cb^d\sup_u|K(u)|^2, \qquad (6.39)\end{aligned}$$

since $\zeta_\theta(Y_{in})$ is bounded uniformly on i and θ by assumption **H.1**, $\int f_i(z-ub)du \leq C$ (see assumption A2) and $\sup_u|K(u)|^2 < \infty$ (see assumption A4 and compactness of $\mathcal{Z}$). Then, we have

$$\mathbf{S}_n = O\left((nb^d)^{-1}\right). \qquad (6.40)$$

Now consider the covariance term

$$\mathbf{R}_n = (nb^d)^{-2}\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n \mathrm{Cov}\left(\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z)\right).$$

Let us partition the spatial locations of the observations using

$$D_n = \{1 \leq i,j \leq n \,:\, \rho < \|s_i - s_j\| \leq c_n\}$$

with c_n is sequence of integers going to ∞ and let $\bar{D}_n$ denote the complement of D_n in the set of locations $\{s_i,\, i = 1, ..., n\}$.
In one hand, let

$$\mathbf{R}_n^{(1)} = (n\,b^d)^{-2}\sum_{i,j\in D_n}\left|\mathrm{Cov}\left(\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z)\right)\right| = (n\,b^d)^{-2}\sum_{i,j\in D_n}|A-B|,$$

with

$$\begin{aligned}|A| &= \left|E\left(\zeta_\theta(Y_{in})K_{in}(z)\zeta_\theta(Y_{jn})K_{jn}(z)\right)\right|\\ &\leq C\left|\int K\left(\frac{z-u}{b}\right)K\left(\frac{z-v}{b}\right)f_{i,j}(u,v)dudv\right|\\ &\leq C\,b^{2d}\left|\int K(u)K(v)f_{i,j}(z-bu, z-bv)dudv\right|\\ &\leq Cb^{2d}\left(\sup_u|K(u)|\right)^2\left|\int f_{i,j}(z-bu, z-bv)dudv\right| = Cb^{2d},\end{aligned}$$

by assumption **H.1**, $\sup_u|K(u)| < \infty$ (assumption A4 and compactness of $\mathcal{Z}$), with $f_{i,j}$ the joint density (assumption A2 and compactness of $\mathcal{Z}$).
Note that the second term B is

$$B = E\left(\zeta_\theta(Y_{in})K_{in}(z)\right)E\left(\zeta_\theta(Y_{jn})K_{jn}(z)\right)$$

Using similar arguments as above, we have $|B| \leq Cb^{2d}$ by assumptions A2, A4, compactness of $\mathcal{Z}$ and continuity of $m_\theta(\cdot)$. Thus, we have

$$\mathbf{R}_n^{(1)} \leq \sum_{i,j\in D_n} Cn^{-2} \leq C\frac{c_n^2-\rho^2}{n} = O\left(\frac{c_n^2}{n}\right). \qquad (6.41)$$

On the other hand, let

$$\mathbf{R}_n^{(2)} = (n\,b^d)^{-2}\sum_{i,j\in \bar{D}_n}\left|\mathrm{Cov}\left(\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z)\right)\right|.$$

By assumption **H.2** combined with (6.39), we have for all $\theta \in \Theta$ and $i,j = 1,\ldots,n$,

$$\left|\mathrm{Cov}\left(\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z)\right)\right| \leq C\,b^d\varphi(\|s_i - s_j\|).$$

Then, we have

$$\mathbf{R}_n^{(2)} \leq C(n\,b^d)^{-1} \sum_{i>c_n/\rho} i\varphi(i\rho).$$

Thus, we derive the following result

$$\mathbf{R}_n = \mathbf{R}_n^{(1)} + \mathbf{R}_n^{(2)} = O\left(n^{-1}\left\{c_n^2 + b^{-d}\sum_{i>c_n/\rho} i\varphi(i\rho)\right\}\right).$$

The following steps of the proof are inspired by the proof of Lemma 8 in (Severini & Wong, 1992, p. 1800–1801) . Let

$$\tilde{v}_\theta(z) = \frac{1}{n}b^{-d}\sum_{i=1}^{n}\left\{\zeta_\theta(Y_{in})K_{in}(z) - E\left(\zeta_\theta(Y_{in})K_{in}(z)\right)\right\}.$$

For some $\epsilon > 0$, Markov's inequality yields

$$P\left(|\tilde{v}_\theta(z)| > \epsilon\right) \quad \leq \quad \frac{\mathbf{R}_n + \mathbf{S}_n}{\epsilon^2}. \tag{6.42}$$

Now, let θ_1 and θ_2 two elements in Θ, since $E\left(\sup_{\theta,1\leq i\leq n,n}|\zeta_\theta^{(1)}(Y_{in})|\right) < \infty$ (by **H.1**), there exist random triangular array (see Severini & Wong, 1992, p.1801) $\left\{W_{in}^{(1)},\, 1\leq i\leq n,\, n=1,2\ldots\right\}$ not depending on θ_1 and θ_2 such that $\sup_{1\leq i\leq n,\,n} E\left(|W_{in}^{(1)}|\right) < \infty$ and

$$\sup_z|\tilde{v}_{\theta_1}(z) - \tilde{v}_{\theta_2}(z)| \leq \sup_z|K(z)|\,\frac{|\theta_2-\theta_1|}{b^d}\,\frac{1}{n}\sum_{i=1}^{n}W_{in}^{(1)}.$$

Similarly, for all $z^{(1)}$ and $z^{(2)}$ in $\mathcal{Z}$, there exist random triangular array $\left\{W_{in}^{(2)},\, 1\leq i\leq n,\, n=1,2\ldots\right\}$ not depending on $z^{(1)}$ and $z^{(2)}$, such that $\sup_{1\leq i\leq n,\,n} E\left(|W_i^{(2)}|\right) < \infty$ and

$$\sup_\theta\left|\tilde{v}_\theta(z^{(2)}) - \tilde{v}_\theta(z^{(1)})\right| \leq C\frac{\|z^{(2)}-z^{(1)}\|}{b^{d+1}}\,\frac{1}{n}\sum_{i=1}^{n}W_{in}^{(2)},$$

since $K(\cdot)$ is Lipschitzian (see assumption **H.2**).
Hence, there exist random triangular array $\{W_{in},\, 1\leq i\leq n,\, n=1,2\ldots\}$ such that $\sup_{1\leq i\leq n,\,n} E\left(|W_{in}|\right) < \infty$ and

$$\sup_{\|z^{(2)}-z^{(1)}\|<\delta_1}\;\sup_{|\theta_2-\theta_1|<\delta_2}\left|\tilde{v}_{\theta_2}(z^{(2)}) - \tilde{v}_{\theta_1}(z^{(1)})\right| \quad \leq \quad C\left(b^{-d}\delta_2 + b^{-(d+1)}\delta_1\right)\frac{1}{n}\sum_{i=1}^{n}W_{in},$$

for some $\delta_1 > 0$, $\delta_2 > 0$ and large n.

As $\mathcal{Z}$ is compact, one can define a real number $\delta_1 > 0$, an integer l_n such that $l_n\delta_1 < C$ with $l_n = \lfloor\gamma_n b^{-(d+1)}\rfloor$ and

$$\mathcal{Z} \subset \bigcup_{j=1}^{l_n} B(z^{(j)}, \delta_1),$$

where $B(z,\delta)$ is the closed ball in $\mathbb{R}^d$ with center z and radius $\delta > 0$.
Also as Θ is compact, one can cover it by $r_n = \lfloor\gamma_n b^{-d}\rfloor$ finite intervals of centers θ_i with same half length $\delta_2 = O(1/r_n)$.

With these covering, we have

$$\begin{aligned} P\left(\sup_{\theta,z}|\tilde{v}_\theta(z)|>\epsilon\right) &\le P\left(\max_{j\le r_n}\max_{k\le l_n}\left|\tilde{v}_{\theta_j}(z^{(k)})\right|>\epsilon/2\right) \\ &\quad + P\left(\sup_{\|z^{(2)}-z^{(1)}\|<\delta_1}\sup_{|\theta_2-\theta_1|<\delta_2}\left|\tilde{v}_{\theta_2}(z^{(2)})-\tilde{v}_{\theta_1}(z^{(1)})\right|>\epsilon/2\right) \\ &\le \quad r_n\, l_n\, P\left(|\tilde{v}_\theta(z)|>\varepsilon/2\right)+Cb^{-d}\left(\delta_2+\delta_1 b^{-1}\right) \\ &= \; C\, r_n\, l_n(\mathbf{S}_n+\mathbf{R}_n)+Cb^{-d}\left(\delta_2+\delta_1 b^{-1}\right) \\ &= \; I^{(1)}+I^{(2)}+I^{(3)}, \end{aligned}$$

where

$$I^{(1)}=O\left(\frac{\gamma_n^2}{nb^{2d+1}}\left(c_n^2+b^{-d}\sum_{i>c_n/\rho}i\varphi(i\rho)\right)\right);\qquad I^{(2)}=O\left(\gamma_n^{-1}\right);\qquad I^{(3)}=O\left(\frac{\gamma_n^2}{nb^{3d+1}}\right).$$

If we take $c_n=o(b^{-d/2})$ and $\gamma_n^2=o(nb^{3d+1})$, then $I^{(1)},I^{(2)}$ and $I^{(3)}$ are all of order $o(1)$ by assumption A5 and the fact that $\varphi(t)\to 0$ as $t\to\infty$ by assumption A3. This yields the proof.

■

Lemma 6.2. *Let for each $\theta\in\Theta$ and $z\in\mathcal{Z}$*

$$H(\eta;\theta,z)=E_0\left(h_{in}^{\theta,\eta}(Y_{in}|X_{in},Z_{in})|Z_{in}=z\right),\ 1\le i\le n,\ n=1,2,\ldots$$

where $\eta=g(z)$, $g\in\mathcal{G}$ and $h_{in}^{\theta,\eta}(\cdot|\cdot,\cdot)$ is defined in assumption A3.

***Condition I**: For fixed but arbitrary $\theta_1\in\Theta$ and $\eta_1\in\Pi$ with $\Pi=g_0(\mathcal{Z})$, let*

$$\vartheta(\theta,\eta)=\int h_{in}^{\theta,\eta}(y|x\,,z)\exp(h_{in}^{\theta_1,\eta_1}(y|x\,,z))dy,\qquad \theta\in\Theta,\ \eta\in\Pi,(x,z)\in\mathcal{Z}\times\mathcal{Z}$$

where $\{\exp(h_{in}^{\theta,\eta}(y|x\,,z)),\theta\in\Theta,\ \eta\in\Pi\}$ denotes the family of conditional density functions (indexed by parameters θ and η) of Y_{in} given $(X_{in},Z_{in})=(x,z)\in\mathcal{X}\times\mathcal{Z}$. For each $\theta\neq\theta_1$, assume that

$$\vartheta(\theta,\eta)<\vartheta(\theta_1,\eta_1).$$

***Condition S**: Let $\tilde{p}=p+1$ and for all nonnegative integers $j_1,\ldots,j_{\tilde{p}}=0,1,2$ and $r=0,\ldots,4$, such that $j_1+\cdots+j_{\tilde{p}}+r\le 6$, assume that the derivative*

$$\frac{\partial^{j_1+\cdots+j_{\tilde{p}}+r}h_{in}^{\theta,\eta}}{\partial\theta_1^{j_1}\cdots\partial\theta_{\tilde{p}}^{j_{\tilde{p}}}\partial\eta^r}(y|x\,,z),$$

exists for almost y and that

$$E_0\left(\sup_{i,\,n}\sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}}\left|\frac{\partial^{j_1+\cdots+j_{\tilde{p}}+r}h_{in}^{\theta,\eta_i}}{\partial\theta_1^{j_1}\cdots\partial\theta_{\tilde{p}}^{j_{\tilde{p}}}\partial\eta^r}(Y_{in}|X_{in}\,,Z_{in})\right|^2\right)<\infty,\qquad\text{with}\qquad \eta_i=g(Z_{in}).$$

Assume that

$$\sup_z\sup_\theta\sup_\eta\left|\frac{\partial^j}{\partial\theta^j}H^{(k)}(\eta;\theta,z)\right|<\infty, \tag{6.43}$$

for $j=0,1,2$ and $k=2,3,4$ such that $j+k\le 4$; with

$$H^{(k)}(\eta;\theta,z)=\frac{\partial^k}{\partial\eta^k}H(\eta;\theta,z).$$

Let

$$\widehat{H}(\eta;\theta,z)=\frac{\sum_{i=1}^{n}h_{in}^{\theta,\eta}(Y_{in}|X_{in},z)K_{in}(z)}{\sum_{i=1}^{n}K_{in}(z)},$$

then $\widehat{g}_\theta(z)$ *is a solution of* $\widehat{H}^{(1)}(\eta;\theta,z)=0$ *with respect to* η *for each fixed* $\theta\in\Theta$ *and* $z\in\mathcal{Z}$.
If we assume that assumptions A1-A6 are satisfied, then we have for all $j=0,1,2$,

$$\sup_{\theta}\sup_{z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{g}_\theta(z)-g_\theta(z)\right)\right|=o_p(1). \tag{6.44}$$

The assumptions used in the previous lemma are satisfied under the conditions used in the main results. In fact, **Condition I** is need to ensure identifiability of the arbitrary parameter θ_1 (it plays the role of the true parameter θ_0). This condition is verified when $\theta_1=\theta_0$, by the identifiability of our model (6.13). **Condition S** allows interchanging integrals with differentiation, it will be combined to the implicit function theorem (see Saaty & Bram, 2012) to ensure differentiability of $\hat{g}_\theta(z)$ with respect to θ.
Knowing that $\Phi(\cdot)$ is a smooth function on $\mathbb{R}$ and $h_{in}^{\theta,\eta}(\cdot|\cdot\,,\cdot)$ is

$$h_{in}^{\theta,\eta_i}(Y_{in}|X_{in}\,,Z_{in})=Y_{in}\log\left(\frac{\Phi(G_{in}(\theta,\eta_i))}{1-\Phi(G_{in}(\theta,\eta_i))}\right)-\log\left(1-\Phi(G_{in}(\theta,\eta_i))\right),$$

Condition S and assumption (6.43) are satisfied under the continuity condition of $\Phi(\cdot)$ and $\phi(\cdot)$, assumption A9 and the compactness of $\mathcal{X}$ and $\mathcal{Z}$.

Proof of Lemma 6.2

The proof of this lemma is similar to that of Lemma 5 in Severini & Wong (1992). Let us follow similar lines as in the proof of Lemma 6.1 above, replacing $\zeta_\theta^{(j)}(Y_{in})$ by

$$\zeta_{\theta,\eta}^{(j,k)}(Y_{in},X_{in})=\frac{\partial^j}{\partial\theta^j}\frac{\partial^k}{\partial\eta^k}h_{in}^{\theta,\eta}(Y_{in}|X_{in}\,,z).$$

and assumptions **H.1** and **H.2** in Lemma 6.1 by the following

H.1' $\sup_{\theta}\sup_{\eta}\sup i,n\left|\zeta_{\theta,\eta}^{(j,k)}(Y_{in},X_{in})\right|<\infty$, for $j=0,\ldots,3,\ k=0,\ldots,5$

H.2' For all $k=0,\ldots,4,\ j=0,1,2$ and $\theta\in\Theta,\ z\in\mathcal{Z}$, (6.35) is satisfied and (6.36) holds with $\zeta_\theta^{(j)}(Y_{in})$ replaced by $\zeta_{\theta,\eta}^{(j,k)}(Y_{in},X_{in})$.

Under the conditions used in the lemma, it is easy to see that **H.1'** is verified, **H.2'** is also satisfied by assumption A3 (in particular, conditions (6.28)).
Using the results of Lemma 6.1, we have for all $j=0,1,2$.

$$\sup_{\theta,\,\eta,\,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(1)}(\eta;\theta,z)-H^{(1)}(\eta;\theta,z)\right)\right|=o_p(1), \tag{6.45}$$

$$\sup_{\theta,\,\eta,\,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(2)}(\eta;\theta,z)-H^{(2)}(\eta;\theta,z)\right)\right|=o_p(1), \tag{6.46}$$

$$\sup_{\theta,\,\eta,\,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(3)}(\eta;\theta,z)-H^{(3)}(\eta;\theta,z)\right)\right|=o_p(1), \tag{6.47}$$

$$\sup_{\theta,\,\eta,\,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(4)}(\eta;\theta,z)-H^{(4)}(\eta;\theta,z)\right)\right|=o_p(1). \tag{6.48}$$

Under assumption A1, we have for any $\epsilon>0$, there exists $\gamma>0$ such that

$$\begin{aligned}
P\left(\sup_{\theta,z}|\widehat{g}_\theta(z)-g_\theta(z)|>\epsilon\right) &\leq P\left(\sup_{\theta,z}|H^{(1)}(\widehat{g}_\theta(z);\theta,z)|>\gamma\right)\\
&= P\left(\sup_{\theta,z}|\widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z)-H^{(1)}(\widehat{g}_\theta(z);\theta,z)|>\gamma\right)\\
&\leq P\left(\sup_{\theta,z,\eta}|\widehat{H}^{(1)}(\eta;\theta,z)-H^{(1)}(\eta;\theta,z)|>\gamma\right).
\end{aligned}$$

Hence

$$\sup_{\theta,z} |\widehat{g}_\theta(z) - g_\theta(z)| = o_p(1) \tag{6.49}$$

The rest of the proof is very similar to that of Lemma 5 in (Severini & Wong, 1992, p. 1798–1799) , for seek of completeness, we present the details.
We have by **Condition I**,

$$\inf_\theta \inf_z -H^{(2)}(g_\theta(z);\theta,z) > 0.$$

In addition, by **Condition S**, for every $\delta > 0$, there exists $\epsilon > 0$, such that

$$\sup_\theta \sup_z \sup_{\eta_1,\eta_2:|\eta_1-\eta_2|\le\epsilon} \left| H^{(2)}(\eta_2;\theta,z) - H^{(2)}(\eta_1;\theta,z) \right| < \delta.$$

Hence, there exists $\epsilon > 0$, such that

$$\inf_\theta \inf_z \inf_{|\eta - g_\theta(z)|\le\epsilon} \left| H^{(2)}(\eta;\theta,z) \right| > 0. \tag{6.50}$$

Since $g_\theta(z)$ and $\widehat{g}_\theta(z)$ satisfy

$$H^{(1)}(g_\theta(z);\theta,z) = 0 \qquad \text{and} \qquad \widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) = 0,$$

respectively for each θ and z, it follows that

$$\begin{aligned} 0 &= \widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) - H^{(1)}(g_\theta(z);\theta,z) \\ &= \widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) - H^{(1)}(\widehat{g}_\theta(z);\theta,z) + H^{(1)}(\widehat{g}_\theta(z);\theta,z) - H^{(1)}(g_\theta(z);\theta,z) \\ &= r_n(\theta,z) + d_n(\theta,z)\left(\widehat{g}_\theta(z) - g_\theta(z)\right), \end{aligned} \tag{6.51}$$

for each θ, z, where

$$r_n(\theta,z) = \widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) - H^{(1)}(\widehat{g}_\theta(z);\theta,z),$$

and

$$d_n(\theta,z) = \int_0^1 H^{(2)}(tg_\theta(z) + (1-t)\widehat{g}_\theta(z);\theta,z)dt.$$

Note that, by (6.50) and $\sup_\theta \|\widehat{g}_\theta - g_\theta\| = o_p(1)$, we have

$$\liminf_z \inf_\theta \left| \widehat{H}^{(2)}(\widehat{g}_\theta(z);\theta,z) \right| > 0 \text{ and} \quad \liminf_z \inf_\theta |d_n(\theta,z)| > 0 \qquad \text{as} \quad n \to \infty. \tag{6.52}$$

Since,

$$\widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) = 0,$$

for all θ, z, we have

$$\widehat{H}^{(2)}(\widehat{g}_\theta(z);\theta,z)\frac{\partial \widehat{g}_\theta}{\partial\theta}(z) + \frac{\partial \widehat{H}^{(1)}}{\partial\theta}(\widehat{g}_\theta(z);\theta,z) = 0.$$

Then, we can deduce from (6.52), (6.45), and (6.46),

$$\sup_\theta \sup_z \left| \frac{\partial \widehat{g}_\theta}{\partial\theta}(z) \right| = O_p(1).$$

Similarly, we have

$$\sup_\theta \sup_z \left| \frac{\partial^j \widehat{g}_\theta}{\partial\theta^j}(z) \right| = O_p(1), \qquad j = 0,1,2. \tag{6.53}$$

Then, (6.53), (6.45)–(6.48) yield

$$\sup_\theta \sup_z \left| \frac{\partial^j}{\partial\theta^j} r_n(\theta,z) \right| = o_p(1), \qquad \text{and} \qquad \sup_\theta \sup_z \left| \frac{\partial^j}{\partial\theta^j} d_n(\theta,z) \right| = O_p(1), \quad j = 0,1,2. \tag{6.54}$$

Now differentiating (6.51) with respect to θ yields

$$\frac{\partial r_n}{\partial\theta}(\theta,z) + (\widehat{g}_\theta(z) - g_\theta(z))\frac{\partial d_n}{\partial\theta}(\theta,z) + d_n(\theta,z)\left(\frac{\partial \widehat{g}_\theta}{\partial\theta}(z) - \frac{\partial g_\theta}{\partial\theta}(z)\right) = 0,$$

then, by (6.45)–(6.54),

$$\sup_{\theta}\sup_{z}\left|\frac{\partial\widehat{g}_{\theta}}{\partial\theta}(z)-\frac{\partial g_{\theta}}{\partial\theta}(z)\right|=o_p(1).$$

On can obtain similarly,

$$\sup_{\theta}\sup_{z}\left|\frac{\partial^2\widehat{g}_{\theta}}{\partial\theta^2}(z)-\frac{\partial^2 g_{\theta}}{\partial\theta^2}(z)\right|=o_p(1).$$

This finishes the proof.

■

Proof of Theorem 6.1

By Lemmas 6.3 and 6.4, Q_n converges to Q in probability, uniformly, i.e

$$\sup_{\theta\in\Theta}|Q_n(\theta,g_\theta)-Q(\theta,g_\theta)|=o_p(1). \tag{6.55}$$

This result allows to have

$$\left|Q(\hat{\theta},g_{\hat{\theta}})-Q(\theta_0,g_0)\right|=o_p(1). \tag{6.56}$$

Indeed, with the help of $|\sup a-\sup b|\leq\sup|a-b|$, we have

$$\begin{aligned}
\left|Q(\hat{\theta},g_{\hat{\theta}})-Q(\theta_0,g_0)\right| &\leq \left|Q_n(\hat{\theta},\hat{g}_{\hat{\theta}})-Q(\hat{\theta},g_{\hat{\theta}})\right|+\left|Q_n(\hat{\theta},\hat{g}_{\hat{\theta}})-Q(\theta_0,g_0)\right| \\
&\leq \sup_{\theta}|Q_n(\theta,\hat{g}_\theta)-Q(\theta,g_\theta)|+\left|\sup_{\theta}Q_n(\theta,\hat{g}_\theta)-\sup_{\theta}Q(\theta,g_\theta)\right| \\
&\leq 2\sup_{\theta}|Q_n(\theta,\hat{g}_\theta)-Q(\theta,g_\theta)| \\
&\leq 2\sup_{\theta}|Q_n(\theta,\hat{g}_\theta)-Q_n(\theta,g_\theta)|+2\sup_{\theta}|Q_n(\theta,g_\theta)-Q(\theta,g_\theta)| \\
&= o_p(1),
\end{aligned}$$

by Lemma 6.5, (6.55) and $\sup_\theta Q(\theta,g_\theta)=Q(\theta_0,g_0)$ (see assumption A8).

By assumption A8, we have for a given $\theta\in\Theta$, there exists $\varepsilon>0$ and an open neighborhood N_θ such that

$$\inf_{\theta_1\in N_\theta}|Q(\theta_1,g_{\theta_1})-Q(\theta_0,g_0)|>\varepsilon. \tag{6.57}$$

This and (6.56) imply that

$$P_0\left(\hat{\theta}\in N_\theta\right)\leq P_0\left(\left|Q(\hat{\theta},g_{\hat{\theta}})-Q(\theta_0,g_0)\right|>\varepsilon\right)\to 0,\ \text{as } n\to\infty. \tag{6.58}$$

Let N_0 be an open neighborhood of θ_0 and consider the compact set $\Theta_0=\Theta\setminus N_0$. Let $\{N_\theta:\theta\in\Theta,\theta\neq\theta_0\}$ denote the open covering of Θ_0 by the procedure given above (each neighborhood N_θ satisfies (6.57)). By the compactness of Θ_0, let $\{N_{\theta_1},\ldots,N_{\theta_r}\}$ be a finite sub-covering, then

$$P_0\left(\hat{\theta}\notin N_0\right)=P_0\left(\hat{\theta}\in\Theta_0\right)\leq\sum_{j=1}^{r}P_0\left(\hat{\theta}\in N_{\theta_j}\right)\to 0,\ \text{ as } n\to\infty,$$

by (6.58). Therefore, we can conclude that

$$\hat{\theta}-\theta_0=o_p(1),\qquad\text{as}\qquad n\to\infty.$$

This yields proof of Theorem 6.1.

■

Proof of Corollary 6.1

Note that

$$\begin{aligned}\left\|\hat{g}_{\hat{\theta}} - g_0\right\| &\leq \|\hat{g}_{\hat{\theta}} - g_{\hat{\theta}}\| + \|g_{\hat{\theta}} - g_0\| \\ &\leq \sup_{\theta} \|\hat{g}_\theta - g_\theta\| + \sup_{\theta} \left\| \frac{\partial g_\theta}{\partial \theta} \right\| \|\hat{\theta} - \theta_0\| = o_p(1),\end{aligned}$$

since, by Proposition 6.1, $\sup_\theta \|\hat{g}_\theta - g_\theta\| = o_p(1)$ and $\sup_\theta \left\| \frac{\partial g_\theta}{\partial \theta} \right\| < \infty$.

■

Lemmas 6.3-6.5

Let the following notations:

$$\eta_i = g(Z_{in}); \quad \tilde{U}_{in} = \tilde{U}_{in}(\theta, \eta_i); \qquad \Phi_{in} = \Phi(G_{in}(\theta, g_\theta)); \qquad \Lambda_{in} = \Lambda(G_{in}(\theta, g_\theta)),$$

for all $\theta \in \Theta$, $1 \leq i \leq n$, $n = 1, 2, \ldots$, with $\Lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)(1 - \Phi(\cdot))$.
The partial derivatives of $S_n(\theta, g)$ with respect to g of order $s = 1, 2, \ldots$, for any functions $v_1, \ldots, v_s$ in $\mathcal{G}$, are given by

$$\frac{\partial^s S_n}{\partial g^s}(\theta, g)(v_1, \cdots, v_s) = n^{-1} \sum_{i=1}^{n} \xi_{in} \frac{\partial^s \tilde{U}_{in}}{\partial \eta^s}(\theta, \eta_i) v_1(Z_{in}) \cdots v_s(Z_{in}).$$

Lemma 6.3. *Under assumptions A3, A6 and A9, we have for all $\theta \in \Theta$,*

$$S_n(\theta, g_\theta) - S(\theta, g_\theta) = o_p(1). \tag{6.59}$$

In addition we have

$$Q_n(\theta, g_\theta) - Q(\theta, g_\theta) = o_p(1), \tag{6.60}$$

if $M_n - M = o_p(1)$.

Note that if assumption A10 is satisfied, then $M_n - M = o_p(1)$.

Proof of Lemma 6.3

Let us start with the proof of (6.59). Remark that

$$S_n(\theta, g_\theta) = n^{-1} \xi_n^T \tilde{U}_n(\theta, g_\theta) = n^{-1} \sum_{i=1}^{n} \xi_{in} \tilde{U}_{in}(\theta, g_\theta),$$

where ξ_{in} is the $q \times 1$ vector representing the ith row in the matrix of instrumentals variables. By definition (see (6.22)), we have $E_0(S_n(\theta, g_\theta)) - S(\theta, g_\theta) = o(1)$. Then, it suffices to show that

$$S_n(\theta, g_\theta) - E_0(S_n(\theta, g_\theta)) = o_p(1). \tag{6.61}$$

Indeed (omitting the $(\theta, g_\theta)-$arguments to simplify the notation), we have

$$\begin{aligned} E_0\left(\|S_n - E_0(S_n)\|^2\right) &= n^{-2} \sum_{i,j=1}^{n} E_0\left(\left(\xi_{in}\tilde{U}_{in} - E_0(\xi_{in}\tilde{U}_{in})\right)^T \left(\xi_{jn}\tilde{U}_{jn} - E_0(\xi_{jn}\tilde{U}_{jn})\right)\right) \\ &\overset{(6.29)}{\leq} n^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} \sum_{t=1}^{q} \left\{\mathrm{Var}_0\left(\xi_{itn}\tilde{U}_{in}\right) \mathrm{Var}_0\left(\xi_{jtn}\tilde{U}_{jn}\right)\right\}^{1/2} \\ &\leq Cn^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} = O\left(n^{-1} \sum_{s=1}^{\sqrt{n}} s\varphi(s)\right) = o(1), \end{aligned}$$

since $\mathrm{Var}_0(\xi_{itn}\tilde{U}_{in})$ is bounded uniformly on θ, i, and $t = 1, \ldots, q$ (by assumption A6) and because $\varphi(s) \to 0$ as $s \to +\infty$ (by assumption A3). This finishes the proof of (6.61) and then that of (6.59). The proof of (6.60) is straightforward by combining (6.59) with assumption A10.

■

Lemma 6.4. *Under assumptions A6-A9, we have $S_n(\cdot, g.) - S(\cdot, g.)$ is stochastically equicontinuous on Θ.*
In addition, if $M_n - M = o_p(1)$, then we have $Q_n(\cdot, g.) - Q(\cdot, g.)$ is also stochastically equicontinuous on Θ.

Proof of Lemma 6.4

Stochastic equicontinuity in Θ can be obtained by proving that $S_n(\theta, g_\theta)$ satisfies a stochastic Lipschitz-type condition on θ (see Mátyás, 1999, p. 17).
Let us show that $S_n(\cdot, g.)$ is stochastically equicontinuous on θ since $S(\cdot, g.)$ is continuous by assumption A8. It suffices to show that (Andrews, 1992) for each $\theta_1, \theta_2 \in \Theta$,

$$\|S_n(\theta_1, g_{\theta_1}) - S_n(\theta_2, g_{\theta_2})\| = O_p\left(\|\theta_1 - \theta_2\|\right). \tag{6.62}$$

Indeed, for $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned}
\|S_n(\theta_1, g_{\theta_1}) - S_n(\theta_2, g_{\theta_2})\| &\leq n^{-1} \sup_{i,\,n} \|\xi_{in}\| \sum_{i=1}^{n} \left|\tilde{U}_{in}(\theta_1, g_{\theta_1}) - \tilde{U}_{in}(\theta_2, g_{\theta_2})\right| \\
&\leq n^{-1} \sup_{i,\,n} \|\xi_{in}\| \sum_{i=1}^{n} \left\{ \sup_{\theta,\,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| \|\theta_1 - \theta_2\| \right. \\
&\qquad \left. + \sup_{\theta,\,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| \|g_{\theta_1} - g_{\theta_2}\| \right\} \\
&\leq n^{-1} \sup_{i,\,n} \|\xi_{in}\| \sum_{i=1}^{n} \left\{ \sup_{\theta,\,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| \right. \\
&\qquad \left. + \sup_{\theta} \left\| \frac{\partial g_\theta}{\partial \theta} \right\| \sup_{\theta,\,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| \right\} \|\theta_1 - \theta_2\|.
\end{aligned}$$

By assumption A6 and Proposition 6.1, we have respectively, $\sup_{i,\,n} \|\xi_{in}\|$ is bounded and $\sup_\theta \left\| \frac{\partial g_\theta}{\partial \theta} \right\|$ is finite. Then, we have to show that

$$n^{-1} \sum_{i=1}^{n} \sup_{\theta,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| + \sup_{\theta,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| = O_p(1);$$

This last is equivalent to

$$\sup_{\theta,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| = O_p(1), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.63}$$

and

$$\sup_{\theta,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| = O_p(1), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{6.64}$$

Let us prove (6.63) in the following. The proof of (6.64) follows the same lines and is then omitted.

Proof of (6.63):
Recall that

$$\Lambda(t) = \frac{\phi(t)}{\Phi(t)(1 - \Phi(t))}.$$

By definition, we have

$$\tilde{U}_{in}(\theta, \eta) = \Lambda(G_{in}(\theta, \eta)) \left(Y_{in} - \Phi(G_{in}(\theta, \eta))\right),$$

with $G_{in}(\theta, \eta) = a_{in}(\theta) b_{in}(\theta, \eta)$ where $a_{in}(\cdot)$ and $b_{in}(\cdot)$ are defined by

$$a_{in}(\theta) = (v_{in}(\lambda))^{-1} \qquad \text{and} \qquad b_{in}(\theta, \eta) = X_{in}^T \beta + \eta, \qquad 1 \leq i \leq n,$$

with $\theta^T = (\beta^T, \lambda)$. We have

$$\frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta,\eta) = \Big\{\Lambda^{'}(G_{in}(\theta,\eta))(Y_{in} - \Phi(G_{in}(\theta,\eta))) - \Lambda(G_{in}(\theta,\eta))\phi(G_{in}(\theta,\eta))\Big\} \frac{\partial G_{in}}{\partial \theta}(\theta,\eta) \tag{6.65}$$

where $\Lambda^{'}(\cdot)$ denotes the derivative of $\Lambda(\cdot)$.

Let us first establish that

$$\sup_{t\in\mathcal{M}, y\in\{0,1\}} \left|\Lambda^{'}(t)(y - \Phi(t)) - \phi(t)\Lambda(t)\right| < \infty, \tag{6.66}$$

which is equivalent to show that $\Lambda^{'}(t)$ and $\phi(t)\Lambda(t)$ are bounded uniformly in $t \in \mathcal{M}$ where the latter denotes the compact subset of $\mathbb{R}$ such that $(v_{ni}(\lambda))^{-1}\left(x^T\beta + g(z)\right) \in \mathcal{M}$, for all $i = 1,\ldots,n,\ n \in \mathbb{N}^*, \lambda \in \Theta_\lambda,\ \beta \in \Theta_\beta,\ x \in \mathcal{X}, z \in \mathcal{Z},\ g \in \mathcal{G}$. Since $\phi^{'}(t) = -t\phi(t)$, we can rewrite $\Lambda^{'}(t)$ as

$$\Lambda^{'}(t) = \frac{1}{\Phi(t)}\left\{\frac{\phi(t)}{1-\Phi(t)}\left(\frac{\phi(t)}{1-\Phi(t)} - t\right)\right\} - \frac{\phi^2(t)}{\Phi^2(t)(1-\Phi(t))}. \tag{6.67}$$

Notice that $\Lambda(\cdot)$ and $\Lambda^{'}(\cdot)$ may be unbounded only at $\pm\infty$ and since $\mathcal{M}$ is a compact subset of $\mathbb{R}$, these functions are bounded on $\mathbb{R}$. This establishes (6.66).
Remark that,

$$\left\|\frac{\partial G_{in}(\theta,\eta)}{\partial \theta}\right\| \leq \left\|\frac{\partial a_{in}(\theta)}{\partial \theta}\right\| |b_{in}(\theta,\eta)| + \left\|\frac{\partial b_{in}(\theta,\eta)}{\partial \theta}\right\| |a_{in}(\theta)|, \tag{6.68}$$

then $\left\|\frac{\partial G_{in}(\theta,\eta)}{\partial \theta}\right\|$ is bounded uniformly in i, n, θ, η by A6, A9 and the compactness of Θ (see assumption A7). This finishes the proof of (6.63), hence (6.62) is proved.

■

Lemma 6.5. *Under assumptions of Proposition 6.1, A6 and A9, we have*

$$\sup_{\theta\in\Theta} \|S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta)\| = o_p(1). \tag{6.69}$$

If in addition $M_n - M = o_p(1)$, then we have

$$\sup_{\theta\in\Theta} |Q_n(\theta, \hat{g}_\theta) - Q_n(\theta, g_\theta)| = o_p(1). \tag{6.70}$$

Proof of Lemma 6.5

Let us prove (6.69). For each $\theta \in \Theta$

$$\begin{aligned}
\|S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta)\| &= n^{-1}\left\|\sum_{i=1}^n \xi_{in}\left(\tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta)\right)\right\| \\
&\leq n^{-1}\sum_{i=1}^n \sup_{i,n}\|\xi_{in}\|\left|\tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_i(\theta, g_\theta)\right| \\
&\leq n^{-1}\sum_{i=1}^n \sup_{i,n}\|\xi_{in}\| \sup_{\theta,\eta}\left|\frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta,\eta)\right| \sup_{\theta}\|\hat{g}_\theta - g_\theta\| \\
&= o_p(1),
\end{aligned}$$

since $\sup_{i,n}\|\xi_{in}\| = O_p(1)$ (by assumption A6), $\sup_\theta \|\hat{g}_\theta - g_\theta\| = o_p(1)$ (see Proposition 6.1) and $\sup_{\theta,\eta}\left|\frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta,\eta)\right| = O_p(1)$ uniformly on i and n (see the proof of Lemma 6.4).
The proof of (6.70) is trivial by combining (6.69) with assumption A10.

■

Proof of Theorem 6.2

Recall that $\frac{d}{d\theta}Q_n(\theta, g_\theta)$ denotes differentiation with respect to θ while $\frac{\partial}{\partial\theta}Q_n(\theta, g_\theta)$ denotes the partial derivative with respect to θ.
Using a Taylor's series expansion and the fact that

$$\frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\hat{\theta}} = 0,$$

we have

$$\hat{\theta} - \theta_0 = -\left\{ \frac{d^2}{d\theta d\theta^T}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta^*} \right\}^{-1} \left\{ \frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} \right\}, \tag{6.71}$$

for some θ^* between θ_0 and $\hat{\theta}$.
First, we would like to replace $\hat{g}_\theta(.)$ in (6.71) with $g_\theta(.)$. For this, let us show that $\frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)$ (resp. $\frac{d^2}{d\theta d\theta^T}Q_n(\theta, \hat{g}_\theta)$) and $\frac{d}{d\theta}Q_n(\theta, g_\theta)$ (resp. $\frac{d^2}{d\theta d\theta^T}Q_n(\theta, g_\theta)$) have same behavior, as function of θ in a neighbor of θ_0. In other words, that is

$$\sup_\theta \left\| \frac{d^2}{d\theta d\theta^T}Q_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T}Q_n(\theta, g_\theta) \right\| = o_p(1) \tag{6.72}$$

and

$$\frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} - \frac{d}{d\theta}Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} = o_p(1). \tag{6.73}$$

Remark that (6.72) is equivalent to

$$\sup_\theta \left\| \frac{d}{d\theta}S_n(\theta, \hat{g}_\theta) - \frac{d}{d\theta}S_n(\theta, g_\theta) \right\| = o_p(1) \tag{6.74}$$

and

$$\sup_\theta \left\| \frac{d^2}{d\theta d\theta^T}S_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T}S_n(\theta, g_\theta) \right\| = o_p(1), \tag{6.75}$$

by (6.20) (since $M_n - M = o_p(1)$, thanks to assumption A10) and

$$\sup_\theta \|S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta)\| = o_p(1)$$

(see Lemma 6.5). Then, (6.74) and (6.75) follow immediately from Lemma 6.8.
To prove (6.73), let the following Taylor's expansion

$$\frac{d}{d\theta}\left(Q_n(\theta, \hat{g}_\theta) - Q_n(\theta, g_\theta)\right) = \frac{d}{d\theta}\left(\frac{\partial Q_n}{\partial g}(\theta, g_\theta)(\hat{g}_\theta - g_\theta) + \tilde{r}_n(\theta) \right),$$

where

$$\tilde{r}_n(\theta) = \int_0^1 \frac{\partial^2 Q_n}{\partial g^2}(\theta, g_\theta + t(\hat{g}_\theta - g_\theta))(\hat{g}_\theta - g_\theta)^2\, dt.$$

We have

$$\frac{d}{d\theta}\tilde{r}_n(\theta)\Big|_{\theta=\theta_0} = o_p(1),$$

using similar arguments as for the terms $\frac{d^j}{d\theta^j}r_n^{(1)}(\theta)$ for $j = 0, 1$ and $\frac{d^2}{d\theta d\theta^T}r_n^{(1)}(\theta)$ in Lemma 6.8 below (see (6.91)). Therefore, we get

$$\begin{aligned}
\frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} - \frac{d}{d\theta}Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} &= \frac{d}{d\theta}\frac{\partial Q_n}{\partial g}(\theta, g_\theta)\Big|_{\theta=\theta_0}(\hat{g}_0 - g_0) \\
&\quad + \frac{\partial Q_n}{\partial g}(\theta_0, g_0)(\hat{g}_0' - g_0') + \frac{d}{d\theta}r_n(\theta)\Big|_{\theta=\theta_0}, \\
&= o_p(1)
\end{aligned}$$

by Lemma 6.7, where $g_0^{'}(.) = \frac{g_\theta}{\partial\theta^T}(.)\Big|_{\theta=\theta_0}$.
Consequently, we get

$$\hat{\theta} - \theta_0 = -\left\{ \frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta)\Big|_{\theta=\theta^*} \right\}^{-1} \left\{ \frac{d}{d\theta} Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} \right\} + o_p(1) \tag{6.76}$$

where θ^* is between $\hat{\theta}$ and θ_0.
Let us show that for each θ^* lying between θ_0 and $\hat{\theta}$:

$$\frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta)\Big|_{\theta=\theta^*} = 2\, B_2(\theta_0) + o_p(1),$$

to replace the Hessian matrix in the right-hand side of (6.76) by its limit $B_2(\theta_0)$.
Let us consider the first and second order differential of $Q_n(\theta, g_\theta)$ with respect to θ:

$$\frac{d}{d\theta} Q_n(\theta, g_\theta) = 2 S_n^T(\theta, g_\theta) M_n \left\{ \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\}$$

with $g_\theta^{'}$ a $1 \times \tilde{p}$ ($\tilde{p} = p + 1$) matrix given by $\frac{\partial g_\theta}{\partial \theta^T}$ and

$$\begin{aligned} \frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta) &= 2 \left\{ \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\}^T M_n \left\{ \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\} \\ &\quad + 2 S_n^T(\theta, g_\theta) M_n \frac{d}{d\theta^T} \left\{ \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\} \end{aligned} \tag{6.77}$$

with

$$\frac{d}{d\theta^T} \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) = \frac{\partial^2 S_n}{\partial \theta \partial \theta^T}(\theta, g_\theta) + \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta, g_\theta) g_\theta^{'},$$

$$\frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) = \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta, g_\theta) + \frac{\partial^2 S_n}{\partial g^2}(\theta, g_\theta) \frac{\partial g_\theta}{\partial \theta}.$$

Note that

$$S_n(\theta^*, g_{\theta^*}) = S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) + S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),$$

since $S(\theta_0, g_0) = 0$ and by Lemmas 6.3-6.4,

$$S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),$$

and as θ^* lies between $\hat{\theta}$ and θ_0, by Lemma 6.4

$$S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) = o_p(1).$$

Using similar arguments as in the proof of (6.63) in Lemma 6.4 by using A9 in order to ensure the boundedness when differentiating twice with respect to θ, we have

$$\left\| \frac{d}{d\theta^T} \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) \right\| = O_p(1) \qquad \text{and} \qquad \left\| \frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\| = O_p(1).$$

Then we can ignore the second term in the right hand in (6.77) at $\theta = \theta^*$. Hence, by Lemma 6.6 and $\theta^* - \theta_0 = o_p(1)$ (thanks to Theorem 6.1), we have

$$\frac{\partial S_n}{\partial \theta}(\theta^*, g_{\theta^*}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) = o_p(1)$$

and

$$\frac{\partial S_n}{\partial g}(\theta^*, g_{\theta^*}) g_{\theta^*}^{'} - \frac{\partial S}{\partial g}(\theta_0, g_0) g_0^{'} = o_p(1),$$

with $g_{\theta^*}^{'} = \left.\frac{g_\theta}{\partial\theta^T}\right|_{\theta=\theta^*}$.
In addition, if $M_n - M = o_p(1)$, we deduce that

$$\begin{aligned}\left.\frac{d^2}{d\theta d\theta^T}Q_n(\theta, g_\theta)\right|_{\theta=\theta^*} &= 2\left\{\frac{\partial S}{\partial\theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'}\right\}^T M\left\{\frac{\partial S}{\partial\theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'}\right\} + o_p(1)\\ &= 2\,B_2(\theta_0) + o_p(1).\end{aligned}$$

Remark that

$$\left.\frac{d}{d\theta}Q_n(\theta, g_\theta)\right|_{\theta=\theta_0} = 2S_n^T(\theta_0, g_0)M_n\left\{\frac{\partial S_n}{\partial\theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0)g_0^{'}\right\}.$$

Then by (6.81) (see the proof of Lemma 6.6), we have

$$\frac{\partial S_n}{\partial\theta}(\theta_0, g_0) - \frac{\partial S}{\partial\theta}(\theta_0, g_0) = o_p(1) \qquad \text{and} \qquad \frac{\partial S_n}{\partial g}(\theta_0, g_0)g_0^{'} - \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'} = o_p(1).$$

Consequently, we get

$$\left.\frac{d}{d\theta}Q_n(\theta, g_\theta)\right|_{\theta=\theta_0} = 2S_n^T(\theta_0, g_0)M\left\{\frac{\partial S}{\partial\theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'}\right\} + o_p(1).$$

Then we have

$$\hat\theta - \theta_0 = -\left\{B_2(\theta_0)\right\}^{-1}\left\{\frac{\partial S}{\partial\theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'}\right\}^T M\, S_n(\theta_0, g_0) + o_p(1).$$

To end the proof, it remains to show that

$$\sqrt{n}B_1(\theta_0)^{-1/2}S_n(\theta_0, g_0) \longrightarrow \mathcal{N}(0, \mathbb{I}_q).$$

Consider, for all $w \in \mathbb{R}^q$ such that $\|w\| = 1$,

$$\begin{aligned}A_n &= w^T\left\{E_0\left(nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)\right)\right\}^{-1/2}\sqrt{n}S_n(\theta_0, g_0)\\ &= n^{-1/2}\sum_{i=1}^n B_{in},\end{aligned}$$

with

$$B_{in} = w^T\left\{E_0\left(nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)\right)\right\}^{-1/2}\xi_{in}\tilde U_{in}(\theta_0, g_0).$$

By Cramer-Wold device, it suffices to show that A_n converges asymptotically to a standard normal distribution, for all $w \in \mathbb{R}^q$ such that $\|w\| = 1$.
To prove this, we will use the central theorem limit (CTL) proposed by Pinkse et al. (2007). These authors used an idea of Bernstein (1927), based on partitioning the observations into J groups $\mathcal{G}_{n1}, \ldots, \mathcal{G}_{nJ}$, $1 \le J < \infty$, which are divided up into mutually exclusive subgroups $\mathcal{G}_{j1n}, \ldots, \mathcal{G}_{jm_{jn}n}$, $j = 1, \ldots, J$. Each observation belongs to one subgroup and its membership can vary with the sample size n and so can the number of subgroups m_{jn} in group j. We assume that the partition is constructed such that

$$m_{jn}/m_{1n} = o(1) \qquad j = 2, \ldots, J$$

and

$$\text{Card}(\mathcal{G}_{irn}) = O\left(\text{Card}(\mathcal{G}_{jtn})\right), \qquad \forall\, i, j = 1, \ldots, J,\ r = 1, \ldots, m_{in}\ , t = 1, \ldots, m_{jn}.$$

Partial sums over elements in groups and subgroups are denoted by A_{nj} and A_{jtn} ,$j = 1, \ldots, J$ and $t = 1, \ldots, m_{jn}$, respectively. Thus, we have

$$A_n = \sum_{j=1}^J A_{jn} = \sum_{j=1}^J\sum_{t=1}^{m_{jn}} A_{jtn}, \qquad A_{jtn} = n^{-1/2}\sum_{i\in\mathcal{G}_{jtn}} B_{in}.$$

Let us recall in the following, the assumptions under which the CTL of Pinkse et al. (2007) holds.
Assumption A. For any $j = 1, \ldots, J$, let $\mathcal{G}^*$, $\mathcal{G}^{**} \subset \mathcal{G}_{jn}$ be any sets for which

$$\forall t = 1, \ldots, m_{jn} \; : \; \mathcal{G}^* \cap \mathcal{G}_{jtn} \neq \emptyset \qquad \Rightarrow \quad \mathcal{G}^{**} \cap \mathcal{G}_{jtn} = \emptyset.$$

Then for any function f in $\mathcal{F} = \{f : \forall t \in \mathbb{R} f(t) = t \text{ or } \exists \upsilon \in \mathbb{R} : \forall t \in \mathbb{R} f(t) = e^{\iota \upsilon t}\}$, where ι is the imaginary number

$$\left| \text{Cov} \left(f \left(\sum_{i \in \mathcal{G}^*} B_{in} \right), f \left(\sum_{i \in \mathcal{G}^{**}} B_{in} \right) \right) \right| \leq \left\{ \text{Var} \left(f \left(\sum_{i \in \mathcal{G}^*} B_{in} \right) \right) \text{Var} \left(f \left(\sum_{i \in \mathcal{G}^{**}} B_{in} \right) \right) \right\}^{1/2} \alpha_{jn},$$

for some mixing numbers α_{jn} with

$$\lim_{n \to \infty} \sum_{j=1}^{J} m_{jn}^2 \alpha_{jn} = 0.$$

Assumption B.

$$\lim_{n \to \infty} \max_{t \leq m_{jn}} \frac{\sigma_{jtn}}{\gamma_{jn}} = 0, \; j = 1, \ldots, J, \qquad \lim_{n \to \infty} \frac{\gamma_{jn}}{\gamma_{1n}} = 0, \; j = 2, \ldots, J,$$

where

$$\sigma_{jtn}^2 = E_0(A_{jtn}^2), \qquad \text{and} \qquad \gamma_{nj}^2 = \sum_{t=1}^{m_{jn}} \sigma_{jtn}^2.$$

Assumption C. For some $\tau > 1$

$$E_0 \left(|A_{jtn}|^{2\tau} \right) = o \left(\sigma_{jtn}^2 \gamma_{jn}^{2\tau - 2} \right), \; j = 1, \ldots, J, \; t = 1, \ldots, m_{jn}.$$

If assumptions $A - C$ hold, then by Theorem 1 in Pinkse et al. (2007), we have $A_n \longrightarrow \mathcal{N}(0, 1)$. So to finish the proof, we have to check these assumptions in our context.

Assumption A: It holds under (6.29) (assumption A3).
Let us choose for instance $J = 2$ groups, each with m_{1n}, m_{2n} subgroups such that $m_{2n} = o(m_{1n})$. Each subgroup is viewed as an area of size $O(\sqrt{c_n} \times \sqrt{c_n})$ such that $(m_{1n} + m_{2n}) c_n = O(n)$. Since $\varphi(\cdot)$ is a decreasing function (assumption A3), then $\alpha_{jn} = O(\varphi(\sqrt{c_n}))$ for $j = 1, 2$. The sequence c_n must be such that $c_n = O(n^{-\nu + 1/2})$ for some $0 < \nu < 1/2$ and $n^{\nu + 1/2} \varphi(\sqrt{c_n}) \to 0$ as $n \to \infty$
If for instance $\varphi(t) = O(t^{-\iota})$, then $n^{\nu + 1/2} \varphi(\sqrt{c_n}) = O(n^{\iota(\nu - 1/4) + (1 + \nu)/2})$, this tends to 0 for each $\iota > 2(1 + \nu)/(1 - 4\nu)$.
Assumption B : By assumption A10, $B_1(\theta_0)$ is positive definite and by definition it is the limit of $E_0 \left(n S_n(\theta_0, g_0) S_n^T(\theta_0, g_0) \right)$, then for sufficiently large n the last matrix is positive definite and its inverse is $O(1)$. Therefore B_{in} is bounded uniformly on i and n, since ξ_{in} is bounded uniformly on i and n by assumption A6 and so is $\tilde{U}_{in}(\theta_0, g_0)$. Then, for all $j = 1, \ldots, J$ and $t = 1, \ldots, m_{nj}$

$$\sigma_{jtn} = \left\{ n^{-1} E_0 \left(\sum_{i \in \mathcal{G}_{jtn}} B_{in} \right) \right\}^{1/2} = O \left(n^{-1/2} \text{Card}(\mathcal{G}_{jtn}) \right)$$

and

$$\gamma_{jn} = O \left(\frac{m_{jn}}{\sqrt{n}} \max_{t \leq m_{jn}} \text{Card}(\mathcal{G}_{jtn}) \right).$$

Therefore,

$$\frac{\sigma_{jtn}}{\gamma_{jn}} = O(1/m_{jn}) \to 0 \text{ as } n \to \infty,$$

for all $j = 1, \ldots, J$ and $t = 1, \ldots, m_{jn}$.
Now consider the second limit in assumption B, we have for all $j = 2, \ldots, J$

$$\frac{\gamma_{jn}}{\gamma_{1n}} = O\left(\frac{m_{jn} \max_{t \leq m_{jn}} \text{Card}(\mathcal{G}_{jtn})}{m_{1n} \max_{t \leq m_{1n}} \text{Card}(\mathcal{G}_{1tn})}\right) = O\left(\frac{m_{jn}}{m_{1n}}\right) \to 0 \text{ as } n \to \infty,$$

since $m_{jn}/m_{1n} = o(1)$ for all $j = 2, \ldots, J$ as $n \to \infty$.
Assumption C : By easy calculation, we can show that

$$\frac{E_0\left(|A_{jtn}|^{2\tau}\right)}{\sigma_{jtn}^2 \gamma_{jn}^{2\tau-2}} = O(m_{jn}^{2-2\tau}) \to 0 \text{ as } n \to \infty.$$

■

Lemma 6.6. *Under assumptions of Theorem 6.2 and for any $\tilde{\theta}$ such that $\tilde{\theta} - \theta_0 = o_p(1)$, we have*

$$\frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) = o_p(1) \tag{6.78}$$

and

$$\frac{\partial S_n}{\partial g}(\tilde{\theta}, g_{\tilde{\theta}}) g'_{\tilde{\theta}} - \frac{\partial S}{\partial g}(\theta_0, g_0) g'_0 = o_p(1), \tag{6.79}$$

with $g'_{\tilde{\theta}}(.) = \left.\frac{g_\theta}{\partial \theta^T}(.)\right|_{\theta=\tilde{\theta}}$.

Proof of Lemma 6.6

To prove (6.78), we need to show that for all $w \in \mathbb{R}^q$ with $\|w\| = 1$,

$$w^T \left\{\frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0)\right\} = o_p(1)$$

which is equivalent to

$$w^T \left\{\frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S_n}{\partial \theta}(\theta_0, g_0)\right\} = o_p(1) \tag{6.80}$$

and

$$w^T \left\{\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - \frac{\partial S}{\partial \theta}(\theta_0, g_0)\right\} = o_p(1). \tag{6.81}$$

The proof of (6.80) is similar to that of (6.62), using the fact that

$$\sup_{\theta, \eta} \left\| \frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \theta^T}(\theta, \eta) \right\| \quad \text{and} \quad \sup_{\theta, \eta} \left\| \frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \eta}(\theta, \eta) \right\|$$

are bounded uniformly on i and n, and $\tilde{\theta} - \theta_0 = o_p(1)$..
Now, let us prove (6.81). By definition of $S(\cdot\,, \cdot)$ (see 6.22)

$$\lim_{n \to \infty} E_0\left(\frac{\partial S_n}{\partial \theta}(\theta_0, g_0)\right) = \frac{\partial S}{\partial \theta}(\theta_0, g_0).$$

Thus it suffices to prove that

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - w^T E_0\left(\frac{\partial S_n}{\partial \theta}(\theta_0, g_0)\right) = o_p(1). \tag{6.82}$$

Let

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) = n^{-1} w^T \xi_{in} \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta_0, \eta_i^0), = \Delta_{n1} - \Delta_{n2}, \tag{6.83}$$

where

$$\Delta_{n1} = n^{-1} \sum_{i=1}^{n} \xi_{in}^{(1)}(\theta_0, \eta_i^0) \left(Y_{in} - \Phi\left(G_{in}(\theta_0, \eta_i^0)\right)\right) \quad \text{and} \quad \Delta_{n2} = n^{-1} \sum_{i=1}^{n} \xi_{in}^{(2)}(\theta_0, \eta_i^0),$$

with

$$\xi_{in}^{(1)}(\theta_0,\eta_i^0) = w^T\xi_i\Lambda'\left(G_{in}(\theta_0,\eta_i^0)\right)\frac{\partial G_i}{\partial\theta}(\theta_0,\eta_i^0),$$

$$\xi_{in}^{(2)}(\theta_0,\eta_i^0) = w^T\xi_{in}\Lambda\left(G_{in}(\theta_0,\eta_i^0)\right)\phi\left(G_{in}(\theta_0,\eta_i^0)\right)\frac{\partial G_{in}}{\partial\theta}(\theta_0,\eta_i^0),$$

and $\eta_i^0 = g_0(Z_{in})$.
The proof of (6.82) is then reduced to prove

$$E_0\left(\|\Delta_{n1}\|^2\right) = o(1) \qquad \text{and} \qquad E_0\left(\|\Delta_{n2} - E_0(\Delta_{n2})\|^2\right) = o(1). \tag{6.84}$$

This last is trivial since $\xi_{in}^{(1)}$ and $\xi_{in}^{(2)}$ are bounded uniformly on i and n (see assumption A6 and compactness of Θ, $\mathcal{X}$, and $\mathcal{Z}$), and by use of the mixing condition (6.29) and (6.30) in Assumption A3). This finishes the proof of (6.78).

To prove (6.79), remark that

$$\frac{\partial S_n}{\partial g}(\tilde\theta, g_{\tilde\theta})g_{\tilde\theta}^{'} - \frac{\partial S}{\partial g}(\theta_0,g_0)g_0^{'} = \left\{\frac{\partial S_n}{\partial g}(\tilde\theta, g_{\tilde\theta}) - \frac{\partial S}{\partial g}(\theta_0,g_0)\right\}g_{\tilde\theta}^{'} + \frac{\partial S}{\partial g}(\theta_0,g_0)\left(g_{\tilde\theta}^{'} - g_0^{'}\right). \tag{6.85}$$

Consider the second term in the right hand in (6.85) and remark that since $\left\|\frac{\partial S}{\partial g}(\theta_0,g_0)\right\|$ and $\sup_\theta\sup_z\left\|\frac{\partial g_\theta(z)}{\partial\theta\partial\theta^T}\right\|$ are finite and $\tilde\theta - \theta_0 = o_p(1)$,

$$\frac{\partial S}{\partial g}(\theta_0,g_0)\left(g_{\tilde\theta}^{'} - g_0^{'}\right) = (\tilde\theta-\theta_0)\,O\left(\left\|\frac{\partial S}{\partial g}(\theta_0,g_0)\right\|\sup_\theta\sup_z\left\|\frac{\partial g_\theta(z)}{\partial\theta\partial\theta^T}\right\|\right) = o_p(1).$$

For the first term in the right hand in (6.85), since $g_{\tilde\theta}^{'} = O_p(1)$ by Proposition 6.1, using similar arguments as to prove (6.78), permits to obtain

$$\frac{\partial S_n}{\partial g}(\tilde\theta, g_{\tilde\theta}) - \frac{\partial S}{\partial g}(\theta_0,g_0) = o_p(1).$$

This yields the proof of (6.79).

■

Lemma 6.7. *Under assumptions of Theorem 6.2, we have*

$$(i) \qquad \frac{d}{d\theta}\frac{\partial Q_n}{\partial g}(\theta,g_\theta)\Big|_{\theta=\theta_0}(\hat g_0 - g_0) = o_p(1)$$

$$(ii) \qquad \frac{\partial Q_n}{\partial g}(\theta,g_\theta)\Big|_{\theta=\theta_0}(\hat g_0^{'} - g_0^{'}) = o_p(1),$$

where

$$\hat g_0^{'}(.) = \frac{\partial\hat g_\theta}{\partial\theta}(.)\Big|_{\theta=\theta_0} \qquad \text{and} \qquad g_0^{'}(.) = \frac{\partial g_\theta}{\partial\theta}(.)\Big|_{\theta=\theta_0}.$$

Proof of Lemma 6.7

To prove (i), note that

$$\begin{aligned}\frac{d}{d\theta}\frac{\partial Q_n}{\partial g}(\theta,g_\theta) &= 2\frac{d}{d\theta}\left\{S_n^T(\theta,g_\theta)M_n\frac{\partial S_n}{\partial g}(\theta,g_\theta)\right\}\\ &= 2\frac{d}{d\theta}S_n^T(\theta,g_\theta)M_n\frac{\partial S_n}{\partial g}(\theta,g_\theta) + 2S_n^T(\theta,g_\theta)M_n\frac{d}{d\theta}\frac{\partial S_n}{\partial g}(\theta,g_\theta).\end{aligned}$$

One can easily see that

$$\frac{d}{d\theta}S_n(\theta, g_\theta) = \frac{\partial S_n}{\partial \theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta)g_\theta^{'}$$

and

$$\frac{d}{d\theta}\frac{\partial S_n}{\partial g}(\theta, g_\theta) = \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta, g_\theta) + \frac{\partial^2 S_n}{\partial g^2}(\theta, g_\theta)g_\theta^{'}.$$

Therefore, we have

$$\begin{aligned}
&\frac{d}{d\theta}\ \frac{\partial Q_n}{\partial g}(\theta, g_\theta)\Big|_{\theta=\theta_0}(\hat{g}_0 - g_0) = \\
&2S_n^T(\theta_0, g_0)M_n\left\{\frac{\partial^2 S_n}{\partial \theta \partial g}(\theta_0, g_0) + \frac{\partial^2 S_n}{\partial g^2}(\theta_0, g_0)g_0^{'}\right\}(\hat{g}_0 - g_0) \\
&\quad + 2\frac{\partial S_n}{\partial g}(\theta_0, g_0)M_n\left\{\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0)g_\theta^{'}\right\}(\hat{g}_0 - g_0).
\end{aligned}$$

By Lemma (6.3) and $S(\theta_0, g_0) = 0$, we get

$$S_n(\theta_0, g_0) = S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1). \tag{6.86}$$

In addition, we have

$$\begin{aligned}
\left\|\frac{\partial^2 S_n}{\partial \theta \partial g}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\| &= n^{-1}\left\|\sum \xi_{in}\frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\| \\
&\leq n^{-1}\sum \sup_{i,n}\|\xi_{in}\| \sup_{\eta}\left\|\frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta)\right\|\|\hat{g}_0 - g_0\| \\
&= o_p(1),
\end{aligned} \tag{6.87}$$

since ξ_i is bounded uniformly on i, n and θ (assumption A6), $\|\hat{g}_0 - g_0\| = o_p(1)$ by Proposition 6.1, and

$$\sup_{i,\,n}\sup_{\eta}\left\|\frac{\partial^2 U_{in}}{\partial \theta \partial \eta}(\theta_0, \eta)\right\| < \infty.$$

Using similar arguments as in the proof of (6.87), we obtain

$$\begin{aligned}
\left\|\frac{\partial^2 S_n}{\partial g^2}(\theta_0, g_0)(\hat{g}_0 - g_0)g_0^{'}\right\| &= n^{-1}\left\|\sum \xi_i \frac{\partial^2 U_{in}}{\partial \eta^2}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))g_0^{'}(Z_{in})\right\| \\
&= o_p(1),
\end{aligned} \tag{6.88}$$

$$\begin{aligned}
\left\|\frac{\partial S_n}{\partial g}(\theta_0, g_0)(\hat{g}_0 - g_0)g_0^{'}\right\| &= n^{-1}\left\|\sum \xi_{in}\frac{\partial U_{in}}{\partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))g_0^{'}(Z_{in})\right\| \\
&= o_p(1),
\end{aligned} \tag{6.89}$$

and

$$\begin{aligned}
\left\|\frac{\partial S_n}{\partial \theta}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\| &= n^{-1}\left\|\sum \xi_{in}\frac{\partial U_{in}}{\partial \theta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\| \\
&= o_p(1).
\end{aligned} \tag{6.90}$$

Combining (6.86)-(6.90) with assumption A10, permits to have

$$\frac{d}{d\theta}\ \frac{\partial Q_n}{\partial g}(\theta, g_\theta)\Big|_{\theta=\theta_0}(\hat{g}_0 - g_0) = o_p(1).$$

This yields the proof of (i).
The proof of (ii) follows along similar lines as (i) and hence is omitted.

■

Lemma 6.8. *Under assumptions of Theorem 6.2, we have*

$$S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) = r_n^{(1)}(\theta),$$

where

$$\sup_\theta \left\| \frac{\partial}{\partial \theta} r_n^{(1)}(\theta) \right\| = o_p(1), \qquad \text{and} \qquad \sup_\theta \left\| \frac{\partial^2}{\partial \theta \partial \theta^T} r_n^{(1)}(\theta) \right\| = o_p(1)$$

Proof of Lemma 6.8

By applying Taylor's Theorem to $\tilde{U}_i(\theta, \cdot)$ for each $\theta \in \Theta$, we get

$$\begin{aligned} S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) &= n^{-1} \sum_{i=1}^n \xi_{in} \left(\tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \right) \\ &= n^{-1} \sum_{i=1}^n \xi_{in} \left(\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}) \right) \\ &\qquad \times \int_0^1 \frac{\partial \tilde{U}_{in}}{\partial \eta} \left(\theta, g_\theta(Z_{in}) + t \left(\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}) \right) \right) dt \\ &= r_n^{(1)}(\theta). \end{aligned}$$

Since the instrumentals variables are bounded uniformly on i, n, and θ (assumption A6), $\sup_{\theta \in \Theta} \|\hat{g}_\theta - g_\theta\|$, $\sup_{\theta \in \Theta} \max_{j=1,\ldots,p+1} \left\| \frac{\partial}{\partial \theta_j} (\hat{g}_\theta - g_\theta) \right\|$ and $\sup_{\theta \in \Theta} \max_{1 \le i,j \le p+1} \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\hat{g}_\theta - g_\theta) \right\|$ are all of order $o_p(1)$ by Proposition 6.1, it suffices to show

$$\sup_{\theta, \eta} \sup_i \left\| \frac{\partial \tilde{U}_{in}}{\partial \eta} (\theta, \eta) \right\| = O_p(1) \tag{6.91}$$

$$\sup_{\theta, \eta} \sup_i \left\| \frac{\partial}{\partial \theta} \frac{\partial \tilde{U}_{in}}{\partial \eta} (\theta, \eta) \right\| = O_p(1) \qquad \text{and} \qquad \sup_{\theta, \eta} \sup_i \left\| \frac{d^2}{\partial \theta \partial \theta^T} \frac{\partial \tilde{U}_{in}}{\partial \eta} (\theta, \eta) \right\| = O_p(1). \tag{6.92}$$

Equation (6.91) is already proved in the proof of Lemma 6.4 (see (6.64)). The proof of (6.92) can be established in a similar manner and is omitted.

■

Chapter 7

Identification of the determinants of UADT cancers incidence in French Northern Region

Contents

Résumé en français

Ce chapitre concerne un travail de nature appliquée, sur l'analyse des facteurs de risque des cancers VADS (voies aéro-digestives supérieures) dans le nord de la France. Selon les statistiques officielles, l'incidence pour ce type de cancer est plus élevée dans le nord que le reste de la France. L'identification des facteurs de risque des cancers VADS devient alors nécessaire dans la région Nord-Pas-De-Calais.

The results of this chapter are obtained in collaboration S. Dabo-Niang (University of Lille), E. Darwich (University of Lille), and J. Foncel (University of Lille).

7.1 Introduction

This chapter is integrated in a project developed in Nord-Pas-de-Calais region where incidence and mortality rates are highest in France and Europe for upper aerodigestive tract (UADT) cancers. Some facts underpin this dramatic regional situation. According to official statistics, the frequency of UADT cancers in the Nord Pas-de-Calais region is one of the highest in France, Europe and perhaps even in the world. The data of the National Federation of Regional Health Observatories shows that the Nord Pas-de-Calais region has the highest excess mortality (in both men and women) by UADT cancers. Another source of information is the "UADT Cancer Registry" which ran from January 1984 to December 1996. This was, in fact, a record of all new cases of UADT cancers reported by otorhinolaryngology, stomatologists, surgeons and radiotherapists specialists in the two departments Nord and Pas-de-Calais. On this 13-year period, 19 024 new cancers were reported, approximately 1,500 new cases per year. In terms of incidence, this represented 70.7

per 100,000 men and 5.2 per 100,000 women. By standardizing the population to the European population, these became respectively 89,5 and 5,5 ; standardizing on the world population; the figures became 67.3 and 4.0. This excess is still relevant in 2005, since the Cancer Registry of Lille and its region has recorded an incidence of lip-mouth-pharynx cancers locally twice higher than the value recorded in national territory. These elements are described in the SIRIC ONCOLille project which has been labeled by INCa under the Cancer 2 plan. SIRIC ONCOLille has developed 5 integrated programs. Two of these programs focus on the regional incidence of these "avoidable" cancers and access to Clinical Research. They relate to UADT, esophageal and hepatic tumors. Program 1 develops an econometric database around the "Cancer Registry of Lille and its Region", integrating humanities and social sciences, clinical and biological data. This database is transversal to all programs, and Program 1 uses it to study the incidence aspects of UADT cancers in the region. Program 2 focuses on the access to care for these patients through the study of socio-economic reluctance to care and through the adaptation of clinical research models to this population. This program has a very strong interaction with Program 1. Three other programs are based on understanding the mechanisms of tumor recurrence associating other tumor models (prostate and melanoma). Program 3 develops modeling of patient follow-up based on the database (Program 1) and socio-professional reintegration projects. Program 4 is totally devoted to the integration of the model of tumor dormancy in the field of solid tumors and finally, Program 5 develops the concepts of image-based treatments to improve the treatment of the initial tumor and its possible recurrences. Through these 5 programs, SIRIC ONCOLille will lead to the implementation of a unique database in its fields, the adaptation of clinical Research program and socio-professional reintegration, understanding or explaining biological phenomena of recurrence (dormancy) and finally, improvement of loco-regional and systemic support strategies . Based on the last 8 years of very strong interactions between all actors of research and care, strongly supported by regional policies, SIRIC ONCOLille is an opportunity to implement a real integrated research site in Oncology in France. This work is part of the SIRIC Program 1 and aims to identify the determinants of the incidence of upper aero-digestive tract (UADT) cancers in the population of Nord Pas de Calais (NPDC). Since UADT cancers are preventable cancers, primary prevention must play a predominant role. Given the marked over-incidences in the region, it is important to undertake new public health actions to improve prevention practices in the population. Our objective is to provide tangible elements to clarify future prevention and public health actions.
In epidemiological studies, there is an abundant literature on the causes of oral cancers (which are part of the UADT cancers). Many international publications have been particularly interested in the relationship between socioeconomic inequalities and oral cancer risk. They revealed the key role of tobacco, alcohol and unfavorable living conditions. The meta-analysis of Conway et al. (2008) based on case-control studies provides an overall synthesis of the results of the literature and provides interesting insights for future research. Overall, this work has found that the oral cancer risk is significantly linked to an unfavorable socio-economic status resulting from the strong social inequalities. The different dimensions of socio-economic status (income, education level, employment level) are all important indirect factors for cancer risk. These relationships are rather stable in the various analyzes, even when the direct risk factors (tobacco, alcohol, food, sexual history, etc.) were taken into account. However, these direct risk factors are related to the inequalities themselves and may explain part of the impact of socio-economic status (there are more smokers in deprived populations, for example). In general, the interactions between socio-economic status and these risk behaviors themselves are complex. For example, consumption of alcohol and cigarettes has been reported as a mechanism for coping with stress associated with poverty. For example, strong tendency for cigarette smoking and alcohol has been reported as a mechanism for coping with stress associated with poverty. Thus, socio-economic circumstances can play a role in the etiology of the disease by being not only potentially a cause itself, but also a "cause of causes". In France, a new research line on social inequalities and cancers has been developing for several years. Its main interests are the UADT and lung cancers, which are frequent in the French population with a highly unequal social distribution (Faggiano et al., 1997). However, few studies on social inequalities have been carried out on these cancers in France. A notable exception is the ICARE project, whose data collection was completed in 2005. Highlighting the role played by occupational exposure as a risk factor for these cancers reinforces the hypothesis of the multiple determinants of cancers linked to social inequality. All these analyzes show (this is particularly true in the NPDC region) that the usual determinants are far from explaining the total cancer risk

for the populations and that it is necessary to better identify the relationships between the factors and their impact on cancer risk. In particular, levels of tobacco and alcohol consumption in the NPDC region alone cannot explain the excess of UADT cancers. It is, therefore, necessary to seek new associations between socioeconomic status and new risk factors.
Among the suggested approaches, there are various associations that might exist between socioeconomic statuses and stress (due to particular working conditions or social pressure), the access to the healthcare system and health information, cognitive abilities and risk behaviors, exposure to adverse environmental factors, etc.
In view of the situation described above, it is urgent to understand the causes of the high incidence of UADT cancers in NPDC region. As mentioned above, "classical" risk factors such as alcohol and tobacco consumption are not sufficient to explain the important social disparities in these cancers. Experts agree that even with comparable levels of social status and tobacco/alcohol consumption profiles, the impact is still much higher in NPDC than in other French regions. There are, therefore, other relevant factors to explain the risk of UADT cancers and a study of individual data can address this problem.
The main objective of the present work is to determine the risk factors in NPDC for UADT cancers (which are little or not identified in the literature) and to assess their share of attributable risk, taking into account the direct and indirect factors already known of the risk (tobacco, alcohol, social status, etc.). Our approach is global and innovative since existed studies generally focus on specific factors and assess the risks associated with each factor independently. Here, the aim is to identify a fairly large set of relevant factors, whether already known or not, by integrating some potential spatial heterogeneity, the interactions between multiple exposures to risk, aspects that are rarely addressed in the literature (Blair et al., 1999). Among the risk factors that are little or not identified in the NPDC region, but with a potentially significant impact, are environmental factor (geographical) Personal hygiene

7.2 Database

To collect data, a case-control study was conducted in the NPDC region. The principle of the case-control study is to compare the frequency of exposure to various risk factors between two populations, one consisting of patients with the disease of interest, the other subjects without that disease. For this purpose, two samples were made in the NPDC region, one composed of healthy individuals with no apparent signs or symptoms of UADT cancers, the other composed of patients medically treated for UADT cancers. The same questionnaire, consisting of several question modules, was proposed for patients and healthy individuals. Two are specifically derived from the DEREDIA protocol and applied to patients and healthy individuals:

1. Module 1 (hetero questionnaire of DEREDIA protocol): sociodemographic, socioprofessional, and socioeconomic indicators such as sex, age, position, lifestyle, occupational activity, annual income, last diploma obtained, family history of chronic diseases, etc.
2. Module 2 (DEREDIA self-reported questionnaire): socio-cognitive and emotional factors related to health (subjective perceptions of health, feeling of control, emotional state, difficulties and strategies of emotional regulation, social incentives, sources of medical information).
3. Module 3: part of the information in the DEREDIA protocol (primary localization of the pathology, TNM stage at initial diagnosis, primary medical and surgical history, current symptoms and treatments, dates and methods of entering the care path, exposure to certain risk factors including tobacco and alcohol) will be partially applied to healthy individuals (exposure to certain risk factors, medical and surgical history).
4. Module 4: The focus here is on some primary risk factors (oral hygiene, exposure to pollution and occupational hazards, sexual practices, nutrition, etc.).

The data gathered in these four modules are used to provide information about some of the different risk dimensions we want to test in the whole project: environmental pollution and occupational exposure, sexual practices, hygiene of life, physical and hereditary characteristics of subjects, cognitive perception of risks. In addition, we have a wide range of information on the factors already identified (alcohol, tobacco consumptions,...).

These questionnaires were applied to individuals (in the six medical centers of the region (Center Oscar Lambret, Center Hospitalier Régional Universitaire de Lille-Hôpital Claude Huriez, Center Médical Specialité du Littoral, Boulogne-sur-Mer Hospital Center, Lens Hospital, La Louvière-Lille)) whose sampling protocol had to be defined.
Of the 600 patients scheduled for inclusion (it lasted about 3 years in the protocol DEREDIA), only 90 patients were eventually retained. For healthy individuals, we have chosen a survey company (IPSOS) to constitute a sample of approximately 348 individuals and to collect the desired information. The chosen strategy consists of sampling by relatively fine strata to allow, on the one hand, grouping between cases and controls in the epidemiological approach and on the other hand, to be able to estimate, without bias, the distribution of risk (conditionally to risk factors) in the econometric approach. In this last case, it is also necessary to make an exhaustive census of the information available in the Nord-Pas-de-Calais population on the distribution of risk on the one hand (aggregate measure of incidence), and on the distribution of individual characteristics on the other hand (INSEE census).
The obtained case-control dataset, composed of 348 controls and 90 cases are used in the following to identify risk factors by econometric spatial binary choice models. In this contribution, we decided to not take into account in the estimation procedure the non-random feature of the sample because of computation complexity of spatial variance-covariance matrix involved in basic binary spatial models incorporating the case-control feature of the sample.

7.3 Applying spatial binary choice models to identify UADT risk factors

Suppose we have a sample of n observations collected from points of the region of interest, located on an irregularly spaced, countable lattice $\mathcal{I} \subset \mathbb{R}^N$, $N \geq 2$. Let $(Y_{s_i}, X_{s_i})_{i=1,\ldots,n}$ be a sequence of spatially dependent observations at these spatial n points denoted $s_i \in \mathbb{R}^N$ drawn from lattice $\mathcal{I}$. Assume that all sites in $\mathcal{I}$ are located at distances of at least $\rho > 0$ for each other; i.e $\forall\ s_i, s_j \in \mathcal{I}$: $\|s_i - s_j\| \geq \rho$. To facilitate the notation, we will denote in this section i for individual in location s_i. The variables Y_i are binary responses ($Y_i = 1$ correspond to cases while 0 is for controls), let $\mathbf{X}_n$ be a $n \times p$ matrix of p exogenous discrete or continuous random variables with elements X_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, p$. Suppose also that the two alternatives for each observation is based on a latent dependent variable Y_i^* via the following spatial autoregressive regression:

$$\begin{aligned} \mathbf{Y}_n^* &= \lambda_0 W_n \mathbf{Y}_n^* + \mathbf{X}_n \beta_0 + \varepsilon_n, \qquad \varepsilon_n \sim N(0, I_n), \\ Y_i &= \mathbb{I}\left(Y_i^* \geq 0\right), \qquad i = 1, \ldots, n. \end{aligned} \tag{7.1}$$

where the coefficient λ_0 is a scalar autoregressive parameter indicating the degree of spatial dependence, β_0 is a $p \times 1$ vector of parameters. W_n is a spatial weight matrix described by one of previous methods given in Chapter 2. Assume that the $n \times n$ matrix $(I_n - \lambda_0 W_n)$ is nonsingular for all n, therefore the variance-covariance matrix of the latent dependent vector of variables $\mathbf{Y}_n^*$ is

$$V_n(\lambda_0) = \mathrm{Var}\left(\mathbf{Y}_n^* | \mathbf{X}_n\right) = \left(I_n - \lambda_0 W_n\right)^{-1} \left\{ \left(I_n - \lambda_0 W_n\right)' \right\}^{-1}.$$

The structure of $V_n(\lambda_0)$ provides the major difficulty of estimating the parameters by a full ML since it requires solving a very computationally demanding problem of n-dimensional integration. As cited in Chapter 2, the GMM (Pinkse & Slade, 1998) or a pseudo maximum likelihood method (Smirnov, 2010) can be used to address this difficulty of estimation. Also, others methodologies of estimation are emerged like, EM algorithm (McMillen, 1992) and Gibbs sampling approach (LeSage, 2000). Recently, Calabrese & Elkink (2014) compare the almost popular methodologies of estimation used in binary choice spatial models. Several R packages permit to estimate binary choice models. Note that in the following, *ProbitSpatial* R package will be used to provide the estimates of parameters β_0 and λ_0.

After applying several models as the SAR, SAE or SAC probit or logit model, we opte to present the previous SAR probit model since the obtained results on the UADT database seem

more realistic and convincing. Indeed, in the following, the SAR probit model will be compared with a basic binary probit model that does not take into account the spatial dependence. Indeed, we will present the numerical results of the spatial probit SAR model (7.1) and the following basic non-spatial binary probit model:

$$\begin{aligned} \mathbf{Y}_n^* &= \mathbf{X}_n\beta_0 + \varepsilon_n, \qquad \varepsilon_n \sim N(0, I_n), \\ Y_i &= \mathbb{I}\left(Y_i^* \geq 0\right), \qquad i = 1, \ldots, n. \end{aligned}$$

The following results are based on spatial weight matrix W_n is such that

$$w_{ij} = \begin{cases} \dfrac{1}{1+d_{ij}} & \text{if } d_{ij} < \rho \\ 0 & \text{otherwise,} \end{cases}$$

with d_{ij} the Euclidean distance between station i and station j, and ρ is some cut-off distance chosen such that each station has at least three neighbors. Other weight matrices have been tested.

7.3.1 Description of exogenous variables and results

We present here the factors already identified in the literature (alcohol, tobacco consumptions, gender, BMI; body mass index) and others; hygiene of life and income characteristics that we want to test with the following non-spatial general model are various variantes (including the spatial contrepart with matrix W_n defined above):

$$\begin{aligned} Y^* = \beta_0 &+ \text{NGADPD} \times (\alpha_1 + \beta_1 \times \mathbb{I}(\text{SA} = 1) \times \text{TSD}) + \text{NCSPD} \times (\alpha_2 + \beta_2 \times \mathbb{I}(\text{ST} = 1) \times \text{TSS}) \\ &+ \beta_3 \times \text{Sex} + \beta_4 \times \text{BMI} + \beta_5 \times \text{FruitVeg} + \beta_6 \times \text{VistDent} + \beta_7 \times \text{Income}, \end{aligned}$$

where,

NGADPD: corresponds to alcohol consumption, "NGADPD" is the number of glasses of alcohol drank per day, "TSD" presents how long the individual has stopped drinking alcohol in months, and "SA" equals one if the individual has stopped to drink. The term $\text{NGADPD} \times (\alpha_1 + \beta_1 \times \mathbb{I}(\text{SA} = 1) \times \text{TSD})$ in the previous model permits to introduce the effect of alcohol consumption by the two exogenous variables Alcohol1 = NGADPD and Alcohol2 = $\text{NGADPD} \times \mathbb{I}(\text{SA} = 1) \times \text{TSD}$. Note that there are some missing values. That of TSD are replaced by 0 while nine missing values of NGADPD are replaced by medians of two groups of individuals that have the same gender (Male or Female) and same disease status (Case or Control).

NCSPD: represents tobacco consumption. Similarly to alcohol, the tobacco consumption is presented by $\text{NCSPD} \times (\alpha_2 + \beta_2 \times \mathbb{I}(\text{ST} = 1) \times \text{TSS})$ where "NCSPD" is the number of cigarettes smoked per day, "TSS" presents how long the individual has stopped to smoke in months, and "ST" equals one if the individual has stopped to smoke. This allows to introduce the effect of consumption of tobacco by the two exogenous variables Tobacco1 = NCSPD and Tobacco2 = $\text{NCSPD} \times \mathbb{I}(\text{ST} = 1) \times \text{TSS}$. Missing values in TSS are replaced by 0.

Sex: represents the gender of each individual, "1" for Male and "0" for Female.

BMI: makes reference to the logarithm of body mass index (BMI).

FruitVeg: corresponds to fruit and vegetable consumption. It is a binary variable taking "1" if the individual eats often fruit or vegetables and "0" otherwise.

VistDent: corresponds to dentist visits. The binary variable VistDent equals "1" if the individual stayed two years (or more) without consulting a dentist and "0" otherwise.

Income: represents the net monthly income. It is a discrete variable with three modalities, "0" if the individual earns less that 1099, "1" if he/she earns between 1100 and 2199, and "3" if he/she earns more that 2200. Thirty-nine missing values are replaced by medians of two groups of individuals that have the same sex (Male or Female) and same disease status (Case or Control).

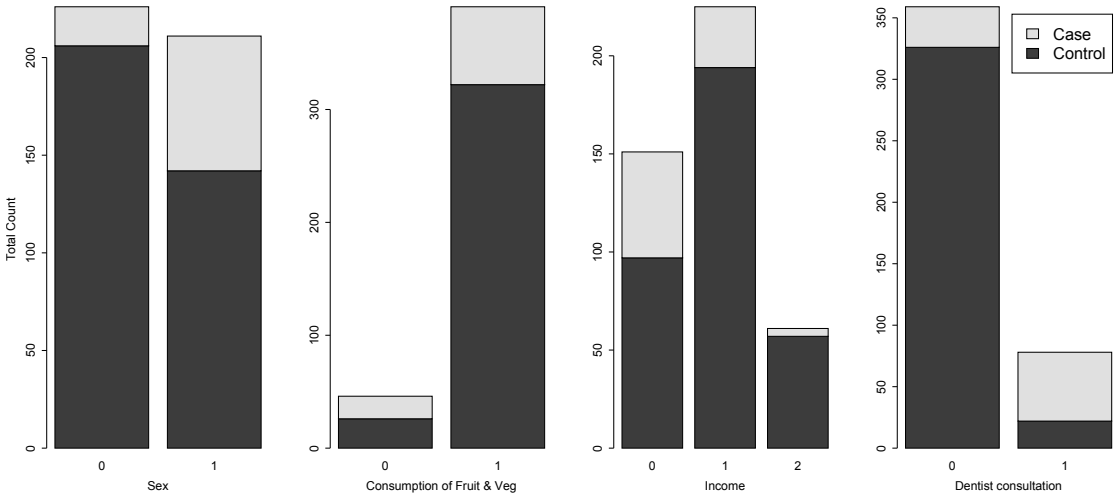

Figure 7.1: Gender, consumption of fruit and vegetables, income and dentist visits for case and control samples

Some descriptive statistics on the cases and controls samples are given in Table 7.1 and Figure 7.1.

Table 7.1: Tobacco gives the quantity in (gramme) taken per day while Alcohol represents the number of cigarettes smoked per day

	Response variable					
	Cases		Controls		Total	
Variables	Mean	Std	Mean	Std	Mean	Std
NCSPD (Tobacco)	20.31	12.71	7.44	11.65	10.06	12.95
BMI	23.68	5.02	26.30	4.91	25.77	5.04
NGADPD (Alcohol)	38.88	36.66	8.85	17.05	14.97	25.48

Using these covariates, we obtain the results in Tables 7.3-7.10 for eight different models described in Table 7.2. Among these eight models, there are various associations between socio-economic, gender, life style variables. In terms of AIC, the best model is Model 7 with the covariates gender, body mass index, tobacco and alcohol consumptions, dentist visit, and income. It has similar AIC as Model 8, Model 3 and Model 5. Comparing Table 7.3 and Table 7.5, we observe that the gender is correlated to dental hygiene. In fact adding dental visits into Model1 (containing the gender) changes the behavior (it becomes less significant) of this last covariate.
In view of the results, we may say that "classical" factors such as the gender (men have higher risk), IBM, alcohol and tobacco consumption are significant risk factors but not sufficient to explain the high incidence in these cancers in the Nord-Pas-De-Calais. The results show other relevant factors to explain the risk of UADT cancers. Indeed, we find also that economic characteristic (income), life style; dental hygiene are risk factors. In addition, there is some spatial dependency (the spatial parameter is significant for all models) explaining spatial disparities in these cancers.
In this sense our spatial approach is innovative since existed studies generally focus on specific factors and assess the risks associated with each factor independently and do not highlight some spatial dependency.
These results gives first more exhaustive ideas on risk factors for UADT cancer in Nord-Pas-de-Calais. However, they are to be taken with caution. The nature of sample (case-control) is not taken into account in the estimation procedure. In addition, the sample size of cases and the quality of the data has to be improved. Indeed, after various statistical analyses, we suspect some data (particularly, the alcohol, tobacco, fruit and vegetable consumptions) to be observed with errors. In a future work, these aspects have to be taken into account to improve the results before

being able to give a final conclusion on specific UADT risk factors in the Northern of France.

Table 7.2: Covariables in the different models

Models	Explanotary variables
Model 1	*Sex, IBM, Tobacco and Alcohol*
Model 2	*Sex, IBM, Tobacco, Alcohol and consumption of fruit and vegetable*
Model 3	*Sex, IBM, Tobacco, Alcohol and dentist visit*
Model 4	*Sex, IBM, Tobacco, Alcohol and income*
Model 5	*Sex, IBM, Tobacco, Alcohol, consumption of fruit and vegetable, and dentist visit*
Model 6	*Sex, IBM, Tobacco, Alcohol, consumption of fruit and vegetable, and income*
Model 7	*Sex, IBM, Tobacco, Alcohol, dentist visit, and income*
Model 8	*Sex, IBM, Tobacco, Alcohol, consumption of fruit and vegetable, dentist visit, and income*

Table 7.3: Model 1

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	3.0115	0.0519		3.5363	0.0245			
Sex	0.4677	0.0163	0.0909	0.4437	0.0247	0.0719	0.0520	0.1239
BMI	-1.5350	0.0018	-0.2577	-1.5509	0.0017	-0.2514	-0.1816	-0.4330
Tobacco1	0.0530	0.0000	0.0089	0.0544	0.0000	0.0088	0.0064	0.0152
Tobacco2	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000	0.0000	-0.0001
Alcohol1	0.1721	0.0000	0.0289	0.1718	0.0000	0.0278	0.0201	0.0480
Alcohol2	-0.0008	0.2345	-0.0001	-0.0009	0.2453	-0.0001	-0.0001	-0.0002
λ				0.4261	0.0182			
AIC	277.52			273.94				

Table 7.4: Model 2

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	2.9751	0.0548		3.4807	0.0270			
Sex	0.4380	0.0262	0.0841	0.4084	0.0419	0.0659	0.0480	0.1140
BMI	-1.4813	0.0027	-0.2480	-1.4884	0.0029	-0.2403	-0.1751	-0.4154
Tobacco1	0.0520	0.0000	0.0087	0.0535	0.0000	0.0086	0.0063	0.0149
Tobacco2	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000	0.0000	-0.0001
Alcohol1	0.1685	0.0000	0.0282	0.1683	0.0000	0.0272	0.0198	0.0470
Alcohol2	-0.0007	0.2674	-0.0001	-0.0008	0.2827	-0.0001	-0.0001	-0.0002
FruitVeg	-0.1633	0.3781	-0.0259	-0.1708	0.3605	-0.0276	-0.0201	-0.0477
λ				0.4283	0.0176			
AIC	278.77			275.14				

Table 7.5: Model 3

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	2.4240	0.1465		3.2337	0.0575			
Sex	0.3478	0.1024	0.0540	0.3481	0.1062	0.0462	0.0306	0.0769
BMI	-1.4068	0.0076	-0.1945	-1.5111	0.0048	-0.2008	-0.1330	-0.3337
Tobacco1	0.0458	0.0000	0.0063	0.0463	0.0000	0.0062	0.0041	0.0102
Tobacco2	-0.0001	0.0006	0.0000	-0.0001	0.0006	0.0000	0.0000	0.0000
Alcohol1	0.1402	0.0006	0.0194	0.1343	0.0011	0.0178	0.0118	0.0297
Alcohol2	-0.0009	0.2384	-0.0001	-0.0008	0.2822	-0.0001	-0.0001	-0.0002
VistDent	1.3854	0.0000	0.2933	1.4024	0.0000	0.1863	0.1234	0.3097
λ				0.4039	0.0316			
AIC	236.76			234.13				

Table 7.6: Model 4

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	3.3667	0.0297		3.9574	0.0121			
Sex	0.7268	0.0007	0.1431	0.7147	0.0011	0.1103	0.0625	0.1728
BMI	-1.5217	0.0018	-0.2417	-1.5761	0.0014	-0.2432	-0.1379	-0.3811
Tobacco1	0.0475	0.0000	0.0075	0.0481	0.0000	0.0074	0.0042	0.0116
Tobacco2	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000	0.0000	0.0000
Alcohol1	0.1499	0.0001	0.0238	0.1481	0.0002	0.0229	0.0130	0.0358
Alcohol2	-0.0008	0.2904	-0.0001	-0.0008	0.2882	-0.0001	-0.0001	-0.0002
Income0vs1	-0.7281	0.0003	-0.0920	-0.7168	0.0004	-0.1106	-0.0627	-0.1733
Income0vs1	-0.9075	0.0044	-0.1085	-0.8629	0.0067	-0.1332	-0.0755	-0.2087
lambda				0.3659	0.0485			
AIC	265.40			263.51				

Table 7.7: Model 5

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	2.3937	0.1518		3.1985	0.0607			
Sex	0.3272	0.1301	0.0504	0.3198	0.1455	0.0424	0.0288	0.0712
BMI	-1.3680	0.0101	-0.1889	-1.4617	0.0070	-0.1937	-0.1317	-0.3254
Tobacco1	0.0452	0.0000	0.0062	0.0456	0.0000	0.0060	0.0041	0.0102
Tobacco2	-0.0001	0.0008	0.0000	-0.0001	0.0008	0.0000	0.0000	0.0000
Alcohol1	0.1383	0.0007	0.0191	0.1320	0.0013	0.0175	0.0119	0.0294
Alcohol2	-0.0008	0.2582	-0.0001	-0.0008	0.3129	-0.0001	-0.0001	-0.0002
FruitVeg	-0.1131	0.5766	-0.0151	-0.1391	0.4986	-0.0184	-0.0125	-0.0310
VistDent	1.3809	0.0000	0.2920	1.3994	0.0000	0.1854	0.1261	0.3115
λ				0.4106	0.0294			
AIC	238.45			235.71				

Table 7.8: Model 6

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	3.3537	0.0305		3.9363	0.0127			
Sex	0.7168	0.0011	0.1408	0.6995	0.0018	0.1079	0.0616	0.1695
BMI	-1.5066	0.0022	-0.2393	-1.5538	0.0019	-0.2396	-0.1369	-0.3766
Tobacco1	0.0473	0.0000	0.0075	0.0479	0.0000	0.0074	0.0042	0.0116
Tobacco2	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000	0.0000	0.0000
Alcohol1	0.1490	0.0001	0.0237	0.1470	0.0002	0.0227	0.0130	0.0356
Alcohol2	-0.0007	0.3026	-0.0001	-0.0008	0.3050	-0.0001	-0.0001	-0.0002
FruitVeg	-0.0469	0.8086	-0.0073	-0.0641	0.7427	-0.0099	-0.0056	-0.0155
Income0vs1	-0.7236	0.0003	-0.0915	-0.7109	0.0004	-0.1096	-0.0626	-0.1723
Income0vs2	-0.8939	0.0059	-0.1073	-0.8440	0.0093	-0.1302	-0.0744	-0.2046
λ				0.3677	0.0474			
AIC	267.35			265.41				

Table 7.9: Model 7

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impacts	indirect impacts	total impacts
Intercept	2.7163	0.1020		3.5215	0.0373			
Sex	0.5978	0.0105	0.0965	0.6060	0.0109	0.0779	0.0417	0.1196
BMI	-1.3815	0.0080	-0.1834	-1.5024	0.0044	-0.1931	-0.1034	-0.2965
Tobacco1	0.0409	0.0000	0.0054	0.0409	0.0000	0.0053	0.0028	0.0081
Tobacco2	-0.0001	0.0025	0.0000	-0.0001	0.0025	0.0000	0.0000	0.0000
Alcohol1	0.1215	0.0033	0.0161	0.1164	0.0057	0.0150	0.0080	0.0230
Alcohol2	-0.0008	0.2699	-0.0001	-0.0009	0.2926	-0.0001	-0.0001	-0.0002
VistDent	1.3516	0.0000	0.2714	1.3711	0.0000	0.1762	0.0944	0.2706
Income0vs1	-0.6798	0.0017	-0.0737	-0.6714	0.0023	-0.0863	-0.0462	-0.1325
Income0vs2	-0.8054	0.0197	-0.0844	-0.7590	0.0282	-0.0976	-0.0522	-0.1498
λ				0.3522	0.0655			
AIC	229.36			227.97				

Table 7.10: Model 8

	Non-Spatial			Spatial				
Variables	estimates	p-values	marginal effects	estimates	p-values	direct impact	indirect impact	total impact
Intercept	2.7162	0.1025		3.5105	0.0383			
Sex	0.5977	0.0121	0.0964	0.5981	0.0141	0.0769	0.0413	0.1182
BMI	-1.3813	0.0087	-0.1834	-1.4911	0.0053	-0.1916	-0.1030	-0.2946
Tobacco1	0.0409	0.0000	0.0054	0.0408	0.0000	0.0052	0.0028	0.0081
Tobacco2	-0.0001	0.0026	0.0000	-0.0001	0.0027	0.0000	0.0000	0.0000
Alcohol1	0.1215	0.0034	0.0161	0.1160	0.0059	0.0149	0.0080	0.0229
Alcohol2	-0.0008	0.2717	-0.0001	-0.0008	0.3018	-0.0001	-0.0001	-0.0002
FruitVeg	-0.0005	0.9982	-0.0001	-0.0322	0.8802	-0.0041	-0.0022	-0.0064
VistDent	1.3516	0.0000	0.2714	1.3703	0.0000	0.1761	0.0946	0.2708
Income0vs1	-0.6797	0.0017	-0.0737	-0.6679	0.0024	-0.0858	-0.0461	-0.1320
Income0vs2	-0.8052	0.0223	-0.0844	-0.7490	0.0339	-0.0963	-0.0517	-0.1480
λ				0.3531	0.0650			
AIC	231.36			229.96				

Chapter 8

General conclusion and perspectives

Conclusion

In thesis we are interested in modelization of unknown parameters of some population from random or non-random (stratified) samples composed of independent or spatially dependent, multivariate or functional data.

We started by studying a functional binary choice model explored in a case-control or choice-based sample design context. We use a conditional likelihood function under the sampling distribution and some dimension reduction strategy to define a feasible conditional maximum likelihood estimator of the model. Large and small sample properties of the proposed estimators, able to account the sample scheme, were studied.

In continuity with the functional framework, we propose in a second contribution a functional linear autoregressive spatial model but with random sampling context. The dimension reduction method explored in the first chapter is combined to a quasi-maximum likelihood method to estimate the model. We establish the consistency and asymptotic normality of the proposed estimators and highlighted the influence of the spatial weight matrix structure of the estimation. The performance of the methodology is illustrated via simulations and an application to ozone concentration data.

The last two contributions concern regression models involving real-valued spatial processes. In one hand, we generalize the classical nearest neighbor method (k-NN) to predict a spatial process at non-observed locations. The proposed predictor combines two kernels to controls distances between observation and locations and uses a bandwidth as the kth lower distance between covariate's point of prediction and covariate's observations. This idea allowed more flexibility to account some heterogeneity in the covariate. We established almost complete convergence with rates of the predictor. The usefulness of the proposed spatial predictor compare with the spatial kernel predictor proposed in the literature is illustrated towards some numerical experiments.

On the other hand, we generalize the partially linear probit model for i.i.d data to spatial dependent data. In this contribution a linear process for disturbances is used which for more spatial dependences flexibility. Parameters involved in the models are estimated by introducing a semi-parametric estimation approach based on weighted likelihood and generalized method of moments methods. Consistency and asymptotic distribution of the estimators are established under sufficient conditions on the choices of bandwidth, the spatial dependence and some instrumentals variables. Some simulated experiments are provided to investigate the finite sample performance of the estimators. An applied chapter on detecting UADT cancer risk factors in the North of France ends the contributions.

In conclusion, this thesis contributes to several statistical issues for both theoretical and applied point of views.

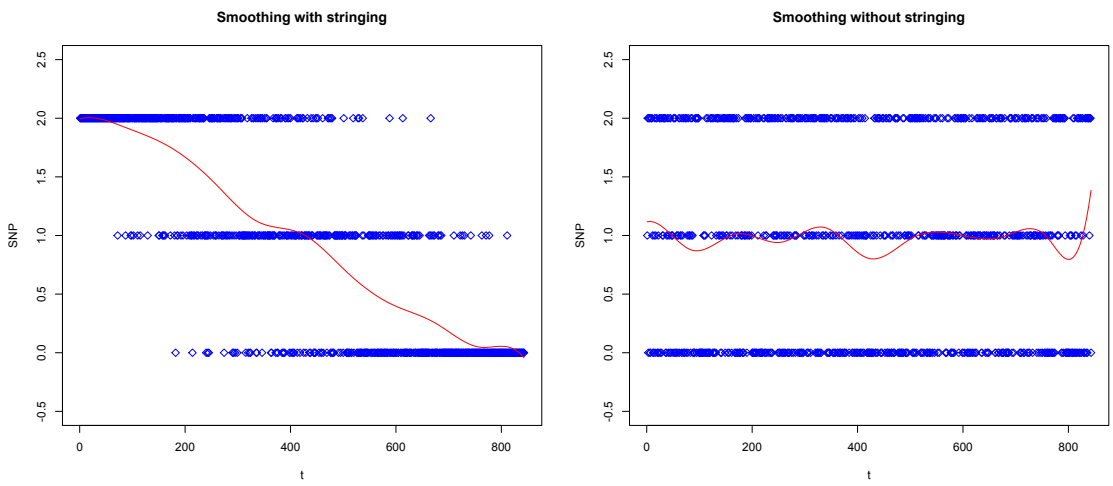

Figure 8.1: Genotypes are presented by the blue points while the red curve presented their smoothed associated function.

Perspectives

In our contributions, some issues and remarks appear, leading to some future research, some of them are developed hereafter.

- Firstly, we would like to apply the proposed method in Chapter 3 to investigate the association between genetic variants (genotypes) and phenotypes (see Fan et al., 2014). These authors found that generalized functional linear models are a good tool for addressing this type of problem. For that, we propose to use the stringing technique (Chen et al., 2011) to address the smoothing problem as the genotypes are defined by a random discrete function $g(t_i)(= 0; 1; 2)$, that is the number of minor alleles of a some individual at the ith-variant located at location t_i for $i = 1, \ldots, m$, when one considers m variants randomly selected. Figure 8.1 illustrates a genotype given by 843 Single Nucleotide Polymorphisms (SNP) smoothed with stringing (left panel) and without (right panel).

- Other improvement of the first contribution could be to address a more practical model where one does not have knowledge on the size of the cases in the population (the parameter Q).
- Our contribution in Chapter 6 is based a semiparametric GMM approach of estimation. However it is important to provide an optimal choice of the weight matrix M_n in order to have efficient GMM estimators. This consists to choose the weight matrix M_n as a consistent estimator $B_{1n}(\hat{\theta})$ of the matrix $B_1(\theta_0)$. Another empirical choice could be the idea of continuous updating GMM estimator (One step GMM) used in Pinkse et al. (2006):

$$M_n(\theta) = n^{-1} \sum_{i,j=1}^{n} \delta_{ij} \xi_{ni} \xi_{jn}^T \tilde{U}_{in}(\theta, \hat{g}_\theta) \tilde{U}_{jn}(\theta, \hat{g}_\theta)$$

with weights

$$\delta_{ij} = \frac{\sum_{r=1}^{n} \tau_{ri} \tau_{rj}}{\left[\sum_{r=1}^{n} \tau_{ri}^2 \sum_{r=1}^{n} \tau_{rj}^2\right]^{1/2}} \qquad \text{for } i, j = 1, \ldots, n,$$

where τ_{ij} is a number depending on w_{ijn}. The nearest the location i is to j, the larger is τ_{ij}.

- Another topic of future research could be to allow (in Chapters 6 and 4) some spatial dependency like the following latent SAR models:

$$Y_{in}^* = \lambda_0 \sum_{j=1}^{n} w_{ijn} Y_{in}^* + X_{in}^T \beta_0 + g_0(Z_{in}) + \varepsilon_{in}, \qquad 1 \leq i \leq n, \ n = 1, 2, \ldots$$

with

$$Y_{in} = \mathbb{I}\left(Y_{in}^{*} \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots$$

- In hydrology or climatology, partitioner is always interested to identify the unexpected events (flow) in some interval of times rather than the ordinary events, then he has to measure the extremality rather than the centrality. In the other hand, in FDA, statisticians are always interested to identify phenomenons more general than the extreme events such that the outliers observations. Febrero et al. (2008) identified two mains reasons why outliers may arise in functional data. First, outliers may be curves with gross errors such as measurement, recording, and typing mistakes. Secondly, outliers can be real data curves that are somehow suspicious or surprising in the sense that they do not follow the same pattern as that of the majority of the curves. When an outlier curve is detected through the first reason, the error associated should be identified and corrected if possible. However, when it is detected by the second reason and it is a *shift outlier* (see Hubert et al., 2015, for more details on the taxonomy of functional outliers) this curve can be viewed as realization of extreme events. However, it is very hard to identify extremes curves by using outliers detection procedure. The latter are based on the notion of depth function which consists to provide a center-outward ordering of curves without considering their position from the deepest curve. Thus, one need to define tools that allow to rank a collection of curves with respect to their extremality. Recently, Franco-Pereira & Romo (2014) defined some measures called *heperextremality* and *hypoextremality* that allow to reflect the *extremality* of a curve with respect to a collection of curves observations. These measures can provide natural ordering for sample curves (see also López-Pintado & Romo, 2011), but it is very hard to identify curves that are extreme in a short interval by using these tools. Therefore, in hydrology, we need to define new measures of extremality that allow to emphasize curves that are extreme in a short interval. This may be done by adapting the idea of extremality measures with the relation order proposed by Narisetty & Nair (2015). The latter defined an extremal depth and observed that it can penalize curves that are extreme in a short interval even if they are representative in the rest of their domain.
  We would like to propose new extremality measures. First steps are in the following.
  Consider a stochastic process X of a distribution P observed in the space of continuous functions in a compact interval $\mathcal{T}$ ($C(\mathcal{T})$). Without loss of generality, take the interval $\mathcal{T}$ to be $[0\,,\,1]$. Let $S = \{x_1(t), x_2(t), \ldots, x_n(t)\}$ be a collection of n observations from X and x a function in $C([0\,,\,1])$.
  López-Pintado & Romo (2011) defined the extremality of a curve x with respect a collection of functions $x_1(t), x_2(t), \ldots, x_n(t)$, as follows.
  The *hyperextremality*:

$$\mathrm{HEM}_n(x) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(x(t) \leq x_i(t)\,,\, t \in \mathcal{T}\right) \tag{8.1}$$

  The *hypoextremality*:

$$\mathrm{hEM}_n(x) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(x(t) \geq x_i(t)\,,\, t \in \mathcal{T}\right). \tag{8.2}$$

  The hyperextremality (resp. hypoextremality) of x is one minus the proportion of functions in the sample below (above) x. When the curves (the collection S) are very irregular, the hyperextremality and hypoextremality will not be good measures of extremality, since for a given curve we may not have many curves lying "totally" below or above it. Then Franco-Pereira & Romo (2014) replaced the indicator function by some Lebesgue measure that seems to be more appropriate when the curves are irregular.
  In a work in progress, we propose the following extremal measures.
  For each fixed $t \in [0\,,\,1]$, we define the pointwise *hyperextremality* of $x(t)$ with respect to S as

$$\mathrm{H}_x(t, S) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(x_i(t) < x(t)), \tag{8.3}$$

and similar we define the pointwise *hypoextremality* of $x(t)$ with respect to S as

$$\mathrm{h}_x(t,S) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(x_i(t) > x(t)). \tag{8.4}$$

One can note that the hyperextremality defined by (8.3) (resp. hypoextremality (8.4)) is evaluated at a fixed point $t \in [0\,,\,1]$ while the hyperextremality defined by (8.1) (resp(8.2)) is evaluated over the whole interval. Indeed, the indice of hyperextremality (8.3) reflects local heperextremality while (8.1) reflects global hyperextremality.
To extend the local heperextremality (resp. hypoextremality) to global heperextremality (resp. hypoextremality), we proceed similary as Narisetty & Nair (2015). We define the cumulative distribution function (CDF) of $\mathrm{EM}_x(t\,,\,S)$ (where EM_x denotes H_x or h_x without ambiguity) as

$$\Phi_{n,x}(r) = \int_0^1 \mathbb{I}(\mathrm{EM}_x(t,S) \leq r)dt\,, \qquad r \in [0\,,\,1] \tag{8.5}$$

This CDF will be called the H–CDF if $\mathrm{EM}_x = \mathrm{H}_x$ or h–CDF if $\mathrm{EM}_x = \mathrm{h}_x$. When $\mathcal{T}$ is not $[0\,;\,1]$, the right hand in (8.5) will be divided by the measure of Lebesgue of $\mathcal{T}$. Also, one can replace the indicator function in (8.5) by certain weight function that allows to emphasize or downweight certain time intervals.
The relation order proposed by Narisetty & Nair (2015) consists to say that a function x is considered as hyperextreme (resp. hypoextreme) with respect to the collection of functions S, if its associated H–CDF (resp. h–CDF), $\Phi_{n,x}(\cdot)$ has most of its mass close to zero. This explains that there are a lot of points in $[0\,,\,1]$ for which the second term on the right hand side in (8.3) (resp. (8.4)) is close to one.
This relation order is explained in the following.
For two functions x and y with corresponding H–CDFs (resp. h–CDFs) $\Phi_{n,x}(\cdot)$ and $\Phi_{n,y}(\cdot)$, let $0 \leq r_1 < r_2 < \cdots < r_M \leq 1$ be the ordered elements of their combined heperextremality (hypoextremality) levels. If $\Phi_{n,x}(r_1) < \Phi_{n,y}(r_1)$, we say that y is hyperextreme (or hypoextreme) than x and we denote that by $x \prec y$. Similarly, if $\Phi_{n,x}(r_1) > \Phi_{n,y}(r_1)$, then $x \succ y$. If $\Phi_{n,x}(r_1) = \Phi_{n,y}(r_1)$, we move to r_2 and make a similar comparison based on their values at r_2. The comparison is repeated until the tie is broken. If $\Phi_{n,x}(r_j) = \Phi_{n,y}(r_j)$ for all $j = 1, \ldots, M$, we say that x and y are equivalent on hyperextremality (or hypoextremality) and denote that by $x \sim y$.
Finally the Extremal measure of hyperextremality of a function x w.r.t S is defined as

$$\mathrm{EH}(x,S) = \frac{\mathrm{Card}\{i\,:\,x_i \preccurlyeq x\}}{n}, \tag{8.6}$$

where this relation order ($\preccurlyeq$) corresponds to H–CDFs.
Simillary, the Extremal measure of hypoextremality of a function x w.r.t S is defined as

$$\mathrm{Eh}(x,S) = \frac{\mathrm{Card}\{i\,:\,x_i \preccurlyeq x\}}{n}, \tag{8.7}$$

where ($\preccurlyeq$) corresponds to h–CDFs. Figure 8.2 (respectively Figure 8.3) gives the third hyperextreme (respectively hypoextreme) curves of hourly measurements of July air temperature over period of 29 years, from 1982 to 2010 in region of Abu Dhabi, United arab emirate.

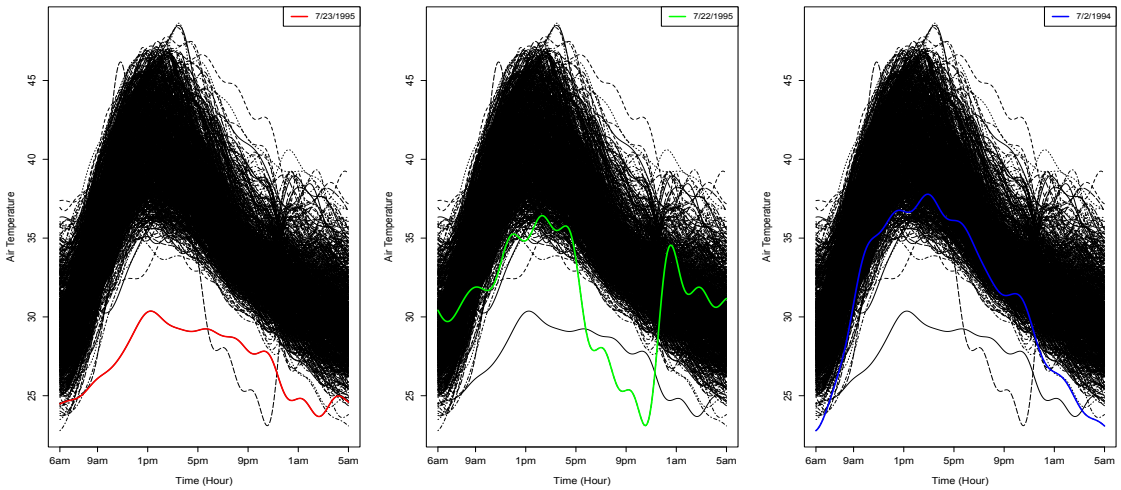

Figure 8.2: third hepoextreme curves.

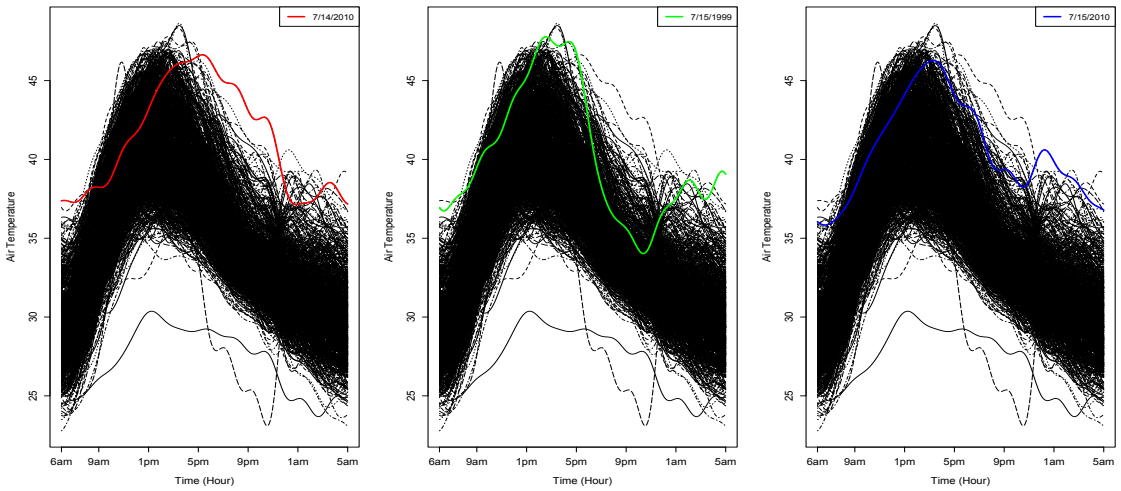

Figure 8.3: third heperextreme curves.

List of Figures

List of Tables

Bibliography

Abdi, A. O., Diop, A., Dabo-Niang, S., & Abdi, S. A. O. (2010). Estimation non paramétrique du mode conditionnel dans le cas spatial. *Comptes Rendus Mathematique*, *348*, 815–819.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, *41*, 997–1016. doi:10.2307/1914031.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, *8*, 241–257.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models* volume 4. Springer Science & Business Media.

Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*. Springer Science & Business Media.

Atteia, O., Dubois, J.-P., & Webster, R. (1994). Geostatistical analysis of soil contamination in the swiss jura. *Environmental Pollution*, *86*, 315–327.

Bel, L., Bar-Hen, A., Petit, R., & Cheddadi, R. (2011). Spatio-temporal functional regression on paleoecological data. *J. Appl. Stat.*, *38*, 695–704. doi:10.1080/02664760903563650.

Bernstein, S. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, *97*, 1–59.

Biau, G., & Cadre, B. (2004). Nonparametric spatial prediction. *Statistical Inference for Stochastic Processes*, *7*, 327–349.

Billé, A. G. (2014). Computational issues in the estimation of the spatial probit model: A comparison of various estimators. *The Review of Regional Studies*, *43*, 131–154.

Blair, A., Zahm, S. H., & Silverman, D. T. (1999). Occupational cancer among women: research status and methodologic considerations. *American journal of industrial medicine*, *36*, 6–17.

Bohorquez, M., Giraldo, R., & Mateu, J. (2016). Optimal sampling for spatial prediction of functional data. *Stat. Methods Appl.*, *25*, 39–54. doi:10.1007/s10260-015-0340-9.

Bohorquez, M., Giraldo, R., & Mateu, J. (2017). Multivariate functional random fields: prediction and optimal sampling. *Stochastic Environmental Research and Risk Assessment*, *31*, 53–70.

Bosq, D. (2000). *Linear processes in function spaces* volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. doi:10.1007/978-1-4612-1154-9.

Burba, F., Ferraty, F., & Vieu, P. (2009). k-nearest neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, *21*, 453–469.

Cai, T. T., & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.*, *107*, 1201–1216. doi:10.1080/01621459.2012.716337.

Calabrese, R., & Elkink, J. A. (2014). Estimators of binary spatial autoregressive models: A monte carlo study. *Journal of Regional Science*, *54*, 664–687.

Carbon, M., Francq, C., & Tran, L. T. (2007). Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference*, *137*, 778–798.

Carbon, M., Tran, L. T., & Wu, B. (1997). Kernel density estimation for random fields (density estimation for random fields). *Statistics & Probability Letters*, *36*, 115–125.

Cardot, H., Chaouch, M., Goga, C., & Labruère, C. (2010). Properties of design-based functional principal components analysis. *J. Statist. Plann. Inference*, *140*, 75–91. doi:10.1016/j.jspi.2009.06.012.

Cardot, H., Faivre, R., & Goulard, M. (2003). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *J. Appl. Stat.*, *30*, 1185–1199. doi:10.1080/0266476032000107187.

Cardot, H., & Josserand, E. (2011). Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, *98*, 107–118. doi:10.1093/biomet/asq070.

Cardot, H., & Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.*, *92*, 24–41. doi:10.1016/j.jmva.2003.08.008.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

Case, A. (1993). Spatial patterns in household demand. *Econometrica*, *52*, 285–307.

Chen, K., Chen, K., Müller, H.-G., & Wang, J.-L. (2011). Stringing high-dimensional data for functional analysis. *Journal of the American Statistical Association*, *106*, 275–284.

Cliff, A., & Ord, K. (1973). Spatial autocorrelation. *London: Pion Ltd*, .

Collomb, G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau: quelques propriétés de convergence ponctuelle. *Statistique non Paramétrique Asymptotique*, (pp. 159–175).

Comte, F., & Johannes, J. (2012). Adaptive functional linear regression. *Ann. Statist.*, *40*, 2765–2797. doi:10.1214/12-AOS1050.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, *92*, 1–45.

Conway, D. I., Petticrew, M., Marlborough, H., Berthiller, J., Hashibe, M., & Macpherson, L. (2008). Socioeconomic inequalities and oral cancer risk: A systematic review and meta-analysis of case-control studies. *International Journal of Cancer*, *122*, 2811–2819.

Conway, J. B. (2013). *A course in functional analysis* volume 96. Springer Science & Business Media.

Cosslett, S. R. (1981). Maximum likelihood estimator for choice-based samples. *Econometrica*, *49*, 1289–1316. doi:10.2307/1912755.

Cosslett, S. R. (2013). Efficient semiparametric estimation for endogenously stratified regression via smoothed likelihood. *J. Econometrics*, *177*, 116–129. doi:10.1016/j.jeconom.2013.07.003.

Crambes, C., Kneip, A., & Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.*, *37*, 35–72. doi:10.1214/07-AOS563.

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cressie, N. A. (1993). Statistics for spatial data, .

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, *147*, 1–23. doi:10.1016/j.jspi.2013.04.002.

Dabo-Niang, S., Kaid, Z., & Laksaci, A. (2012). Spatial conditional quantile regression: Weak consistency of a kernel estimate. *Rev. Roumaine Math. Pures Appl.*, *57*, 311–339.

Dabo-Niang, S., Ternynck, C., & Yao, A.-F. (2016). Nonparametric prediction of spatial multivariate data. *Journal of Nonparametric Statistics*, *28*, 428–458.

Dabo-Niang, S., & Yao, A. F. (2007). Kernel regression estimation for continuous spatial processes. *Mathematical Methods of Statistics*, *16*, 298–317.

Deo, C. M. (1973). A note on empirical processes of strong-mixing sequences. *The Annals of Probability*, (pp. 870–875).

Doukhan, P. (1994). *Mixing* volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York. doi:10.1007/978-1-4612-2642-0 properties and examples.

El Machkouri, M., & Stoica, R. (2010). Asymptotic normality of kernel estimates in a regression model for random fields. *Journal of Nonparametric Statistics*, *22*, 955–971.

Escabias, M., Aguilera, A. M., & Valderrama, M. J. (2007). Functional PLS logit regression model. *Comput. Statist. Data Anal.*, *51*, 4891–4902. doi:10.1016/j.csda.2006.08.011.

Faggiano, F., Partanen, T., Kogevinas, M., & Boffetta, P. (1997). Socioeconomic differences in cancer incidence and mortality. *IARC Scientific Publications*, (pp. 65–176).

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* volume 66. CRC Press.

Fan, R., Wang, Y., Mills, J. L., Carter, T. C., Lobach, I., Wilson, A. F., Bailey-Wilson, J. E., Weeks, D. E., & Xiong, M. (2014). Generalized functional linear models for gene-based case-control association studies. *Genetic epidemiology*, *38*, 622–637.

Febrero, M., Galeano, P., & González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, *19*, 331–345.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In *Advances in spatial econometrics* (pp. 145–168). Springer.

Franco-Pereira, L. R. E., Alba M., & Romo, J. (2014). Measures of extremality and a rank test for functional data. *https://arxiv.org/abs/1409.1816*, (pp. 1–20).

Gaetan, C., & Guyon, X. (2008). *Modélisation et statistique spatiales*. Springer.

Garthoff, R., & Otto, P. (2017). Control charts for multivariate spatial autoregressive models. *AStA Adv. Stat. Anal.*, *101*, 67–94. doi:10.1007/s10182-016-0276-x.

Gheriballah, A., Laksaci, A., & Rouane, R. (2010). Robust nonparametric estimation for spatial regression. *Journal of Statistical Planning and Inference*, *140*, 1656–1670.

Giraldo, R. (2014). Cokriging based on curves, prediction and estimation of the prediction variance. *InterStat*, *2*, 1–30.

Giraldo, R., Delicado, P., & Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *J. Agric. Biol. Environ. Stat.*, *15*, 66–82. doi:10.1007/s13253-009-0012-z.

Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environ. Ecol. Stat.*, *18*, 411–426. doi:10.1007/s10651-010-0143-y.

Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, *30*, 21–42.

Hallin, M., Lu, Z., Tran, L. T. et al. (2004). Local linear spatial regression. *The Annals of Statistics*, *32*, 2469–2500.

Hallin, M., Lu, Z., Yu, K. et al. (2009). Local linear spatial quantile regression. *Bernoulli*, *15*, 659–686.

Hastie, T., & Mallows, C. (1993). A statistical view of some chemometrics regression tools: Discussion. *Technometrics*, *35*, 140–143.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* volume 43. CRC Press.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York. doi:10.1007/978-1-4614-3655-3.

Hsing, T., & Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Hubert, M., Rousseeuw, P. J., & Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications*, *24*, 177–202.

Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *Journal of the American Statistical Association*, *89*, 1354–1365.

Ibragimov, I. A., & Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

Imbens, G. W. (1992). An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica*, *60*, 1187–1214. doi:10.2307/2951544.

James, G. M., & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *63*, 533–550. doi:10.1111/1467-9868.00297.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*, 99–121.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Econom. Rev.*, *40*, 509–533. doi:10.1111/1468-2354.00027.

Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, *104*, 219–257.

Keogh, R. H., & Cox, D. R. (2014). *Case-control studies* volume 4 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, Cambridge. doi:10.1017/CBO9781139094757.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, *9*, 137–163.

Klemelä, J. (2008). Density estimation with locally identically distributed data and with locally stationary data. *J. Time Ser. Anal.*, *29*, 125–141. doi:10.1111/j.1467-9892.2007.00547.x.

Kudraszow, N. L., & Vieu, P. (2013). Uniform consistency of knn regressors for functional variables. *Statistics & Probability Letters*, *83*, 1863–1870.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*, 1899–1925. doi:10.1111/j.1468-0262.2004.00558.x.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econometrics*, *137*, 489–514. doi:10.1016/j.jeconom.2005.10.004.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, *32*, 19–35.

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*, (pp. 19–44).

Li, J., & Tran, L. T. (2009). Nonparametric estimation of conditional expectation. *Journal of Statistical Planning and Inference*, *139*, 164–175.

Lin, X., & Lee, L.-f. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *J. Econometrics*, *157*, 34–52. doi:10.1016/j.jeconom.2009.10.035.

López-Pintado, S., & Romo, J. (2011). A half-region depth for functional data. *Computational Statistics & Data Analysis*, *55*, 1679–1695.

Lu, Z., & Chen, X. (2004). Spatial kernel regression estimation: weak consistency. *Statistics & probability letters*, *68*, 125–136.

Malikov, E., & Sun, Y. (2017). Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *J. Econometrics*, *199*, 12–34. doi:10.1016/j.jeconom.2017.02.005.

Manski, C. F., & Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica*, *45*, 1977–1988. doi:10.2307/1914121.

Manski, C. F., & McFadden, D. (1981). Alternative estimators and sample designs for discrete choice analysis. In *Structural analysis of discrete data with econometric applications* (pp. 51–111). Cambridge M, A: MIT Press.

Martinetti, D., & Geniaux, G. (2016). Probitspatial: Probit with spatial dependence, sar and sem models. *CRAN*, . URL: https://CRAN.R-project.org/package=ProbitSpatial.

Mátyás, L. (1999). *Generalized method of moments estimation* volume 5. Cambridge University Press.

McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., & Ruppert, D. (2014). Functional generalized additive models. *J. Comput. Graph. Statist.*, *23*, 249–269. doi:10.1080/10618600.2012.729985.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, *32*, 335–348.

Menezes, R., Garcia-Soidan, P., & Ferreira, C. (2010). Nonparametric spatial prediction under stochastic sampling design. *Journal of Nonparametric Statistics*, *22*, 363–377.

Müller, H.-G., Chiou, J.-M., & Leng, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC Bioinformatics*, *9*, 60. doi:10.1186/1471-2105-9-60.

Müller, H.-G., & Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*, *33*, 774–805. doi:10.1214/009053604000001156.

Muller, S., & Dippon, J. (2011). k-nn kernel estimate for nonparametric functional regression in time series analysis. *Fachbereich Mathematik, Fakultat Mathematik und Physik (Pfaffenwaldring 57)*, *14*, 2011.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*, 141.

Narisetty, N. N., & Nair, V. N. (2015). Extremal depth for functional data and applications. *Journal of the American Statistical Association*, (pp. 1–38).

Neaderhouser, C. C. (1980). Convergence of block spins defined by a random field. *Journal of Statistical Physics*, *22*, 673–684.

Nerini, D., Monestiez, P., & Manté, C. (2010). Cokriging for spatial functional data. *J. Multivariate Anal.*, *101*, 409–418. doi:10.1016/j.jmva.2009.03.005.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120–126.

Pinkse, J., Shen, L., & Slade, M. (2007). A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics*, *140*, 215–225.

Pinkse, J., Slade, M., & Shen, L. (2006). Dynamic spatial discrete choice using one-step gmm: an application to mine operating decisions. *Spatial Economic Analysis*, *1*, 53–99.

Pinkse, J., & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, *85*, 125–154.

Poirier, D. J., & Ruud, P. A. (1988). Probit with dependent observations. *The Review of Economic Studies*, *55*, 593–614.

Preda, C., Saporta, G., & Lévéder, C. (2007). PLS classification of functional data. *Comput. Statist.*, *22*, 223–235. doi:10.1007/s00180-007-0041-4.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics (2nd ed.). Springer, New York.

Rao, C. R. (1973). *Linear statistical inference and its applications*. (2nd ed.). John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.

Ratcliffe, S. J., Heller, G. Z., & Leader, L. R. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression. *Statistics in medicine*, *21*, 1115–1127.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, *165*, 5–19.

Rosenblatt, M. (1985). Stationary sequences and random fields, .

Ruiz-Medina, M. D. (2011). Spatial autoregressive and moving average hilbertian processes. *Journal of Multivariate Analysis*, *102*, 292–305.

Ruiz-Medina, M. D. (2012). Spatial functional prediction from spatial autoregressive hilbertian processes. *Environmetrics*, *23*, 119–128. URL: http://dx.doi.org/10.1002/env.1143. doi:10.1002/env.1143.

Saaty, T. L., & Bram, J. (2012). *Nonlinear mathematics*. Courier Corporation.

Severini, T. A., & Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association*, *89*, 501–511.

Severini, T. A., & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of statistics*, (pp. 1768–1802).

Smirnov, O., & Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics & Data Analysis*, *35*, 301–319.

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional science and urban economics*, *40*, 292–298.

Sood, A., James, G. M., & Tellis, G. J. (2009). Functional regression: A new model for predicting market penetration of new products. *Marketing Science*, *28*, 36–51.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association*, *84*, 276–283.

Stone, C. J. (1977). Consistent nonparametric regression. *The annals of statistics*, (pp. 595–620).

Su, L. (2004). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, *167*, 1899–1925.

Su, L. (2012). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, *167*, 543–560.

Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *64*, 445–456.

Ternynck, C. (2014). Spatial regression estimation for functional data with spatial dependency. *J. SFdS*, *155*, 138–160.

Tjøstheim, D. (1987). Spatial series and time series: similarities and differences. In *Spatial processes and spatial time series analysis, Proceedings of the 6th Franco-Belgian Meeting of Statisticians. Publications des Facultés Universitaires Saint-Louis, Brussels* (pp. 217–228).

Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, *34*, 37–53.

Wang, H., Iglesias, E. M., & Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, *172*, 77–89.

Wang, H., & Wang, J. (2009). Estimation of the trend function for spatio-temporal models. *Journal of Nonparametric Statistics*, *21*, 567–588.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, (pp. 359–372).

Xie, Y., & Manski, C. F. (1989). The logit model and response-based samples. *Sociological Methods & Research*, *17*, 283–302.

Yang, K., & Lee, L.-f. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *J. Econometrics*, *196*, 196–214. doi:10.1016/j.jeconom.2016.04.019.

Zheng, Y., & Zhu, J. (2012). On the asymptotics of maximum likelihood estimation for spatial linear models on a lattice. *Sankhya A*, *74*, 29–56. doi:10.1007/s13171-012-0009-5.