UNIVERSITÉ DE LILLE.

École Doctorale MADIS

Thèse de doctorat

Mention : Mathématiques et leurs interactions

présentée par

Alaa Ali Hassan

Analyse des données spatiales: applications à la santé des populations.

Spatial data analysis: applications to population health.

---

dirigée par Sophie Dabo-Niang et Dharini Pathmanathan.

Soutenue le 3 décembre 2021 devant le jury composé de :

Pr Sophie Dabo-Niang	Université de Lille	Directrice
Dr Dharini Pathmanathan	University of Malaya (Malaysia)	Co-encadrante
Pr Azzouz Dermoune	Université de Lille	Président du jury
Dr Shahid Ullah	Flinders University (Australia)	Examinateur
Pr Ibrahim Bin Mohamed	University of Malaya (Malaysia)	Examinateur
MCF Ali Amad	Université de Lille	Examinateur
Pr Raquel Menezes	University of Minho (Portugal)	Rapportrice
MCF Yousri Slaoui	Université de Poitiers	Rapporteur
Dr Marielle Wathelet	Université de Lille	Invitée
Dr Zaynab Salloum	Université Libanaise (Liban)	Invitée

Résumé

Ces dernières années, les études sur les prévisions démographiques se sont considérablement développées. L'un des objectifs de la démographie est d'analyser et de prévoir statistiquement les taux de mortalité et de fécondité sans se fier aux opinions subjectives des experts. Par conséquent, pour identifier les caractéristiques de la dynamique de mortalité d'une population, de nombreux modèles ont été développés depuis l'introduction du célèbre modèle proposé par Lee et Carter (1992). De nombreuses recherches disponibles dans la littérature tendent à se concentrer sur la perspective des séries temporelles de la prévision des taux de mortalité. Le manque d'études dans le cadre spatial a suscité notre intérêt pour l'étude des taux de mortalité dans le cadre spatial. L'extension du modèle de Lee-Carter en incorporant l'idée de l'analyse des données fonctionnelles (FDA) a inspiré la première partie de cette thèse où le concept de FDA a été appliqué au cadre d'analyse démographique spatiale. Nous étudions l'existence d'une autocorrélation spatiale pour des données de mortalité des pays voisins. Une méthode fonctionnelle spatiale de composantes principales est proposée pour révéler les modèles spatiaux en considérant directement l'information spatiale. Une statistique fonctionnelle du I de Moran est introduite. Cette statistique aide à déterminer l'autocorrélation spatiale dans les données fonctionnelles par la mise en œuvre de l'ACP spatio-fonctionnelle. Cette statistique fonctionnelle du I de Moran est la première de son genre dans le cadre des données fonctionnelles.

La deuxième partie de cette thèse étudie l'impact du système VigilanS (programme de prévention des tentatives de suicide en France) sur la récidive suicidaire où les données de ce système (âge, sexe, adresse, historique des tentatives de suicide, séjour à l'hôpital etc.) sont cartographiées sur la carte de la région Nord-Pas-de-Calais tout en construisant des modèles de prédiction spatiale. Les risques de tentatives de suicide sont étudiés à l'aide de modèles probit spatiaux. Nous proposons un modèle probit partiellement linéaire pour les données spatialement dépendantes. Ce modèle n'a pas été étudié dans la littérature d'un point de vue théorique et cette partie comble cette lacune en abordant un modèle d'erreur autorégressive spatiale (SAE) où la structure de dépendance spatiale est intégrée dans un terme de perturbation du modèle étudié. Une méthode d'estimation semi-paramétrique est obtenue en combinant l'approche de la méthode des moments généralisée et la méthode de vraisemblance pondérée. Nous avons examiné l'utilisation de ce modèle de régression probit spatial ainsi que d'autres modèles existants dans la littérature pour étudier les récidives de tentatives suicidaires des patients impliqués dans le système VigilanS.

Cette thèse met en évidence l'importance des statistiques spatiales dans l'analyse des problèmes démographiques et de suicide. Il est également intéressant de voir comment les données fonctionnelles peuvent être utilisées comme un outil dans le domaine de la démographie, notamment pour capturer l'autocorrélation spatiale dans les taux de mortalité

lorsque l'espace est concerné. En outre, l'utilisation de modèles de régression spatiale pour étudier les récidives de tentatives de suicide, met en évidence l'impact des localisations voisines sur l'acte suicidaire.

Mots clés: Statistiques spatiales, Autocorrélation spatiale, I de Moran, Matrice de poids, Démographie, Mortalité, Analyse des données fonctionnelles, Analyse fonctionnelle en composantes principales, Statistiques non paramétriques, Récidive du suicide, Modèle probit.

Abstract

In recent years, studies in demographic forecasting have grown significantly. One of the goals of demography is to statistically analyse and predict mortality and fertility rates without relying on subjective opinions of experts. Therefore, to identify the characteristics of the mortality dynamics of a population, many models were developed since the introduction of the famous model proposed by Lee and Carter (1992). Many research available in the literature tend to focus on the time series perspective of forecasting mortality rates. Lack of studies from the spatial framework sparked our interest in investigating the mortality rates from the spatial framework. The extension of the Lee-Carter (1992) model by incorporating the idea of functional data analysis (FDA) inspired the first part of this thesis where the FDA concept was applied to the spatial demographic analysis framework. We investigate the existence of spatial autocorrelation in mortality data of neighbouring countries. A functional spatial principal component method is proposed to reveal spatial patterns by directly considering spatial information. A functional Moran's I statistic is introduced. This statistic aids in determining the spatial autocorrelation in functional data through the implementation of the spatio-functional PCA. This functional Moran's I statistic is the first of its kind in the functional data framework.

The second part of this thesis investigates the impact of the VigilanS system (program to prevent suicide reattempts in France) on suicide recidivism where the data from this system (patient's age, sex, address, history of suicide attempts, hospital stay etc.) are mapped on the map of the Nord-Pas-de-Calais region while constructing spatial prediction models. The risks of suicide attempts are mapped with the help of spatial probit models. We propose a partially linear probit model for spatially dependent data. This model has not been investigated in the literature from a theoretical point of view and this part fills that gap by addressing a spatial autoregressive error (SAE) model where the spatial dependence structure is integrated in a disturbance term of the studied model. A semi-parametric estimation method is obtained by combining the generalized method of moments approach and the weighted likelihood method. We examined the use of this spatial probit regression model as well as other existing models in the literature to study the suicide relapses of patients involved in the VigilanS system.

Keywords: Spatial statistics, Spatial autocorrelation, Moran's I, Weight matrix, Demography, Mortality, Functional data Analysis, Functional principal component analysis (FPCA), Non-parametric statistics, Suicide relapses, Probit model.

Dedication

'I would like to dedicate this work to GOD Almighty for giving me this opportunity to do this thesis which would not have been possible otherwise.

This work is also dedicated to my father who always prays for me to achieve this work, to the soul of my mother "Hadia", to my brother, my aunt and my husband.'

Acknowledgments

However, a Thank You is only appreciable if it is really sincere, so I will try to choose the right words to thank my God who gave me the chance to meet more than remarkable people during my life.

The first and biggest thank you goes to no one else than my thesis supevisor Sophie Dabo-Niang, who has gone far and beyond the call of duty from time to time, to push me forward, keep me in the game, and simply was there all the time despite her various responsibilities and her busy schedule. We share this journey together from the master's study till the third year. Thank you for having given me your confidence by accepting to supervise my thesis but also for having given me the chance to live enriching experiences. I hope that we will have the opportunity to collaborate during the next years.

I am very grateful to Dr Dharini Pathmanathan for her constant guidance and supervision, as well as for providing necessary information regarding the project. Her constant guidance and willingness to share her vast knowledge helped me to complete the assigned tasks on time. I would like to thank you for pushing me forward and for being there all the time despite your various responsibilities, your busy schedule, and the time lag between France and Malaysia.

I would like to thank the members of the jury who agreed to judge this work and to be the rapporteurs of my thesis and who agreed to examine my project and to be part of this jury. My thanks and appreciations also go to Marielle and Zeinab.

I thank the doctoral students of the laboratory Ahmed Salem, Zied, Fahriat, Salsabil, Solange, Yoba, Mamadou, Feryel and Christelle.
My friends with whom I spent beautiful moments and unforgettable memories, thank you for being my family in France. I love you all very much: Fatima, Zaynab, Doha, Hussein,Jana, Kawthar, Meryem, Batoul, Zeinab, Joelle, Abir and all the girls.
Thank you to my cousin Sarah and her husband Maysam who have always been here for me and to hear my news.
I also thank Ghadir and Zakariya and their little girls Sara and Tasnim for being my family in Lille.
I would like to thank all my family: téta, my cousins, my uncles and aunts especially Leyla, Hayat and Hanaa and my aunt Khalida who helped me a lot since my master's degree, which was my first year in France.
And never forget the family of my husband, and the prayers of his mother and sisters.
My sincere thanks to my best friends: Sawsan and Waed.

My dear family,

thank you for always being at my side and encouraging me.
Dear husband, because of you I didn't feel any stress or depression and I have make all my efforts to finish my PhD as soon as possible in order to join together.
Dear Amina, I thank you from the bottom of my heart for every detail, your unconditional love, the happiness that you have brought to our family and your prayers.
Dear brother, I would like to extend my sincere thanks to you for being always by my side, for your friendship, your unconditional support, and your encouragement.
Baba, thanks to God and to you I am today at this stage, you have sacrificed everything for your children me and my brother, sparing neither health nor effort. You gave me a magnificent model of hard work and perseverance. I am indebted to an education that I am proud.
Mama, I love you and I know that you are always proud of us.

I end by saying that everything I have done and am doing is in the eyes of God. Thank you God for all the gifts, for all the graces, for all that you give us. Thank you God.

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
	and 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

Introduction

This thesis is motivated by two real population health problems such as mortality forecasting and the mapping of suicidal relapses.

The first part of this thesis investigates mortality from a spatial perspective. In recent years, studies of demographic forecasts have grown significantly. One of the goals in demographic research is to analyse and predict mortality and fertility rates in a purely statistical way without relying on the subjective opinions of experts. Therefore, to identify the characteristics of the mortality dynamics of a population, many models have been developed since the introduction of the famous model proposed by Lee and Carter (1992). Many researches available in the literature tend to focus on time series forecasting of mortality rates. However, this research sparked the interest in investigating mortality from the spatial framework since the main theme of this thesis is spatial statistics. The extension of the Lee-Carter (1992) model by Hyndman and Ullah (2007) using functional data analysis (FDA) inspired the application of the FDA concept in the spatial framework.
In this thesis, we investigate the existence of spatial autocorrelation in mortality data of neighbouring countries. The spatial principal component analysis by Jombart et al. (2008) aimed to reveal spatial patterns by directly considering spatial information. We constructed a functional Moran's I statistic based on this, which will aid in determining spatial autocorrelation in functional data through the implementation of the spatio-functional PCA. This functional Moran's I statistic is the first of its kind in the functional data framework. The data of mortality rates for 28 European countries from the Human Mortality Database (HMD) for ages 0 to 110 (where ages above 100 are grouped as 100 +) were used to investigate the spatial dependency at a fixed time, where years 1990, 2000 and 2010 were examined separately).

The second part investigates the area of spatial modelling by considering a partially linear probit model for spatially dependent data. Semi-parametric binary models have not been investigated from a theoretical point of view and this part fills that gap by addressing a spatial autoregressive error (SAE) model when the spatial dependence structure is integrated in a disturbance term of the studied model. A semi-parametric estimation method is obtained by combining the generalized method of moments approach and the weighted likelihood method.
In France, the VigilanS healthcare system is an effort to support those who have attempted suicide in various regions. It was established in Nord-Pas-de-Calais in February 2015. This programme to monitor and prevent recidivism of suicide attempts is executed via phone calls by teams of professionals who are specialized in this type of remote care. This six-month programme is managed by the adult psychiatry department, directed by Professor Guillaume Vaiva in the Lille University Hospital. Posthospital support is offered to those patients who attempted suicide.

The patients receive a resource card with a dedicated contact number and are called back for follow up after 6 months. Those with a history of suicide reattempts are contacted from the 10th to 21st day, after the suicide attempts. The VigilanS healthcare database of the Lille University Hospital contains the data of more than 23000 patients which include age, sex, address, history of recidivism of suicide attempts, hospital stay (date, duration of hospitalization, establishment), recidivism of suicide attempt at the first enrolment (mode, contact, alcohol consumption). The aim of this study is to investigate recidivism of suicide attempts for all patients at the initial point of enrollment and after 6 months. It is of interest to study the spatial dependency of the occurrences of recidivism of suicide attempts. Furthermore, the risks of suicide attempts are mapped with the help of the spatial probit models. The impact of the VigilanS system on recidivism of suicide attempts is investigated, where the data from this system (patient's age, sex, address, history of suicide attempts, hospital stay etc.) are mapped on the map of Nord Pas de Calais while constructing spatial prediction models. The factors of suicide relapse among these patients based on the regions they belong to were studied.

Chapter 1 discusses the fundamental concepts and a state of art of the models and methods used. In Chapter 2 of this thesis, we examined the idea of FPCA which is an important dimensionality tool reduction technique on functional data with infinite dimensionality. We investigated the spatial autocorrelation of mortality rates for 28 European countries, with data from the Human Mortality Database (HMD) using spatial associations in the context of functional areal data. This motivated to the development of a functional Moran's I statistic which is the first of its kind in the functional data analysis framework, which can detect spatial autocorrelation and spatial PCA for areal data.
Chapter 3 focuses on partially linear spatial probit models. Here we are interested in various patterns of spatial data by assuming conditional spatial heteroscedasticity, non-identically distributed observations, and a linear process for disturbances. The estimation procedure involves the combination of a weighted likelihood and a generalised method of moments.
Chapter 4 is mainly an application of the models studied in Chapter 3. Here, we investigate the impact of the Vigilans system (programme to prevent suicide reattempts in France) on suicide recidivism where the data from this system (patient's age, sex, address, history of suicide attempts, hospital stay etc.) are mapped on the map of the Nord-Pas-de-Calais region. This will aid in making spatial prediction models. The probit regression model was used to study the suicide relapses and then the risk of suicide attempts was mapped by using the spatial probit and spatial Poisson models.
Chapter 5 provides the concluding remarks as well as some significant contributions of this research. Suggestions on extending research works related to this research are also included in this chapter.

Chapter 1

State of art and general concepts

This chapter discusses the methods used in formulating the main outcomes of this research. Spatial data analysis and spatial modelling are the main components of this thesis. We implement the use of functional data analysis in the field of demography with respect to mortality models to investigate spatial dependence of mortality rates of neighbouring countries. The second part of this study focuses on examining the theoretical foundations and application of spatial econometric models.

1.1 Functional data analysis

Functional data analysis (FDA) has been gaining importance in analysis of data involving surfaces, multidimensional objects, shapes or more complex mathematical objects of infinite dimension (Cardot et al., 2003) . Ramsay and Silverman (2005) gave an excellent overview of the concepts, foundation and applications of FDA. Some reputable works in this field include those of Bosq (2000), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Hsing and Eubank (2015).

Functional data is applied widely in various fields such as medicine, meteorology, hydrology, genetics (genetic sequence) and so on. FDA deals with the analysis and theory of data that are in the form of functions, rather than by vectors of $\mathbb{R}^n$. In other words, each observed variable has functional values rather than real values.
A functional random variable is a random variable which takes its values in a vector space of infinite dimension, and the functional data is a realization of a functional variable. The functional data are considered as observations of infinite-dimensional stochastic processes. Let I be a compact interval of $\mathbb{R}^d$ with $d \in \mathbb{N}$, we observe N independant realizations $X^{(1)}, \ldots, X^{(n)}, \ldots, X^{(N)}$ of an underlying stochastic process $X = (X_t : t \in I)$, where $X^{(n)}$ is a continuous function on I. Random functions can be viewed as random elements taking values in a Hilbert space $L^2(I)$. The associated inner product is defined as:

$$\langle f, g \rangle = \int_I f(t)g(t)dt, \quad f, g \in L^2(I). \tag{1.1}$$

The mean function $\mu : I \mapsto \mathbb{R}$ and the covariance function $c : I \times I \mapsto \mathbb{R}$ are defined as:

$$\mu(t) = \mathbb{E}(X(t)), \quad c(s,t) = \mathrm{Cov}[X(s), X(t)], \quad s, t \in I. \tag{1.2}$$

The statistical methods for multivariate data encounter difficulties involving high dimension. This is mainly because handling infinite dimensions of functional variables (curves, shapes, and so on) can be complicated. Besides that, the dependency between observations when considering time-series of spatial functional objects may be difficult to manage. Therefore, it is necessary to develop statistical methods for visualization and modeling to handle such data. During the last two decades, various exploration and modeling techniques have been proposed for functional variables. As for regression models, there are essentially two popular approaches: the parametric (Ramsay & Silverman, 2005) and nonparametric models (Ferraty & Vieu, 2006).

Hyndman and Ullah (2007) proposed the combination of ideas from FDA, nonparametric smoothing and robust statistics to form a methodology which is a significant contribution to the field of demography especially mortality modeling. This idea is widely applicable to any functional time series data, particulary age-specific mortality and fertility. A nonparametric smoothing method was used to smooth the data and the fitted curves were decomposed via a basis function expansion. This method is discussed further in Section 1.2.2.

1.1.1 From discrete data to functions

The first step in FDA is to convert a discrete set of measurements (the observed data points) into either a rough or smooth curve.
Let $\Phi = \{\phi_j(\cdot) : j \in \mathbb{N}\}$ be an infinite basis of $L^2(I)$. The elements of Φ are linearly independent. Every element of $L^2(I)$ can be written as a linear combination of the elements of Φ. The realization $X^{(n)}$ of the stochastic process X is decomposed into:

$$X^{(n)}(\cdot) = \sum_{j\geq 1} c_{j,n}\phi_j(\cdot), \tag{1.3}$$

where $\{c_{j,n}\}_{j\geq 1}$ is an infinite set of coefficients. The basis expansion is used to approximate the realization $X^{(n)}$ by its projection on the span of a finite basis functions $\Phi_J = \{\phi_j(\cdot) : 1 \leq j \leq J\}$, a finite subset of Φ and $\{c_{j,n}\}_{1\leq j\leq J}$ a subset of $\{c_{j,n}\}_{j\geq 1}$:

$$X^{(n)}(\cdot) \approx \sum_{j=1}^{J} c_{j,n}\phi_j(\cdot), \tag{1.4}$$

$X^{(n)}$ can be summarized by a J-dimensional vector. The functional data smoothing methods imply a decomposition of each of the process realizations into a common basis of function, such as the Fourier bases or splines bases. The Fourier basis system is commonly used for periodic data while the B-spline basis system is preferable for nonperiodic data (Ramsay et al., 2009). Splines are polynomial segments joined end-to-end and forced to be smoothed at the joint.

Let $\{t_i, X_i : i = 1 \ldots n\}$ be a set of observations X_i at design points t_i, modeled by the relation $X_i = s(t_i) + \epsilon_i$ where $s(t)$ is a smooth curve and ϵ_i are iid. The $s(t)$ that minimizes the residual sum of squares (RSS) plus roughness penalty is the cubic smoothing spline fit to the data:

$$\sum_{i=1}^{n} (x(t_i) - s(t_i))^2 + \lambda \int \left(s''(t)\right)^2 dt, \tag{1.5}$$

where $\lambda \geq 0$ is the smoothing parameter which governs the trade-off between smoothness and goodness of fit.

1.1.2 Functional principal component analysis

Functional principal component analysis (FPCA) is the extension of the multivariate principal component analysis in the functional framework. As in the classical case, the FPCA corresponds to an optimal linear representation of a set of functional data in a finite dimensional space. This is to reduce the dimensionality of the data using FPCA and then identify the main sources of variability. The point behind FPCA which is a dimension reduction method, is transforming the sampled curves so that only a low-dimensional space represents the patterns of variability of the curves. The general approaches are presented by Ramsay and Silverman (2005) and Ferraty and Vieu (2006).

If we have n functional observations in L^2, $X^{(1)}, \ldots, X^{(n)}$, we thus look for j functions of L^2, $\phi_1, \ldots, \phi_k$, orthogonal, and such that the projection of $X^{(i)}$ onto the vector space generated by the ϕ_j generates the minimum loss possible. The FPCA is performed by searching the spectrum of a compact operator. This operator is defined from the covariance function given by :

$$\begin{aligned} c(s,t) &= \mathrm{Cov}[X(s), X(t)], \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i(s) - \mu(s))(X_i(t) - \mu(t)), \end{aligned} \tag{1.6}$$

where μ denotes the average function of X_i. FPCA identifies principal components explaining the variability of $\{X_i\}$ by computing the eigenfunctions corresponding to the ordered eigenvalues (from largest to smallest) of an empirical covariance operator. Therefore, performing the PCA of the X_i amounts to searching for the eigenvalues of the operator $\Gamma f(t) = L^2(I) \longmapsto L^2(I)$ defined by:

$$\Gamma f(t) = \langle C(\cdot, t), f(\cdot)\rangle, \quad t \in I, f \in L^2(I), \tag{1.7}$$

where Γ is a positive, linear and self-adjoint operator in L^2 (Horváth & Kokoszka, 2012). In particular, it is a compact operator and has a finite trace.

There exists a complete orthonormal basis $\{\phi_j\}_{j\geq 1}$ and a sequence of real numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ such that:

$$\Gamma\phi_j = \lambda_j\phi_j, \quad \text{and} \quad \lambda_j \to 0 \quad \text{as} \quad j \to \infty, \tag{1.8}$$

where $\{\lambda_j\}_{j\geq 1}$ is the set of eigenvalues of the covariance operator Γ associated to $\{\phi_j\}_{j\geq 1}$ the set of its eigenfunctions. The eigenfunctions associated to the eigenvalues are then the $\{\phi_j\}$. It can be shown that the eigenfunction associated with the largest eigenvalue, ϕ_1 , is a solution of the following constrained optimization problem:

$$\max_{\|\phi\|_2=1} \langle \Gamma\phi, \phi\rangle \tag{1.9}$$

The process X admits the Karhunen-Loève representation:

$$X(t) = \mu(t) + \sum_{j\geq 1} \mathfrak{c}_j\phi_j(t), \quad t \in I, \tag{1.10}$$

with $\mathfrak{c}_j = \langle X - \mu, \phi_j\rangle$, $\mathbb{E}(\mathfrak{c}_j) = 0$, $\mathrm{cov}(\mathfrak{c}_j, \mathfrak{c}_l) = \lambda_j 1_{j=l}$ and the $\{\phi_j\}_{j\geq 1}$ are the FPCA basis.

Hence, $X^{(n)}$ is approximated by truncating the infinite sum at the first J terms:

$$X^{(n)}(t) \approx \mu(t) + \sum_{j=1}^{J} \mathfrak{c}_{j,n}\phi_j(t), t \in I \quad \text{with} \quad \mathfrak{c}_{j,n} = \langle X^{(n)} - \mu, \phi_j \rangle. \tag{1.11}$$

The concept of FPCA in FDA is the crux of the analysis of mortality data. The classical FPCA was performed to find new functions that reveal the most important variations in the curve data with the absence of the spatial structure. After that, this idea was extended to the spatial framework with the implementation of the FPCA on areal spatial data.

1.2 Mortality modeling

1.2.1 The Lee-Carter model

Mortality, fertility, and migration are processes which can change the population. The twentieth century witnessed the improvement in mortality especially in developed countries. Mortality forecasting has been gaining interest and this led to the development of stochastic mortality forecasting approaches over time which involve extrapolation and time series methods. The Lee-Carter (LC) model (Lee & Carter, 1992) is one of the most prominent contributions to the study of mortality models which opened the path to various innovations of its kind. The Lee-Carter model is given as follows:

$$\log(m_{xt}) = a_x + b_x k_t + \epsilon_{x,t} \tag{1.12}$$

where

- $m_{x,t}$ is the age-specific death rate for age x and year t;
- a_x is the average age of the log mortality rates across years;
- b_x is a deviation in mortality due to changes in the time index (k_t);
- k_t is the mortality index in the year t; it describes the evolution of the level of mortality over time.
- the error term $\epsilon_{x,t}$ reflects age-specific historical influences not captured by the model.It is normally distributed with mean 0 and variance σ_ϵ^2.

Three unboserved parameters a_x, b_x and k_t in the single equation 1.12 means that the LC model is over-parametrized and therefore two normalization constraints are imposed:

$$\sum k_t = 0, \quad \sum b_x = 1.$$

The parameters b_x and k_t are obtained by singular value decomposition (SVD). Lee and Carter (1992) fitted the parameter k_t using standard ARIMA models. The random walk model with drift given as follows:

$$k_t = k_{t-1} + d + e_t,$$

was found to be suitable to describe k_t, where d is the drift parameter which reflects the average annual change and e_t is an uncorrelated error.

Some of the commonly used extensions of the LC model consist of the Renshaw and Haberman (2006) model which used the generalised linear model approach, the Lee and Miller (2001) which involves re-estimation of the time component according to the observed life expectancy at birth, and the Booth et al. (2002) model which adjusts the time component with respect to the age distribution of deaths.Wiśniowski et al. (2015) extended the LC model using the Bayesian framework. Hyndman and Ullah (2007) applied smoothing techniques in increasing the LC variants. The HU model extended the LC model by incorporating FDA, nonparametric smoothing and robust statistics.

1.2.2 The Hyndman and Ullah (2007) method

The HU model is more robust than the LC model when we consider the variation explanation, outlier identification and forecast accuracy. The HU method extends the LC method by :

- smoothing the log mortality rates before modeling,
- using the FPCA,
- using more than one principal component for forecasting.

Hyndman and Booth (2008) used functional data models with time series coefficients to model age-specific mortality and fertility rates. Hyndman et al. (2013) proposed a method for non-divergent or coherent forecasting of mortality rates for two or more subpopulations, based on functional principal components models of simple and interpretable functions of rates. In a nutshell, various approaches in forecasting mortality rates were discovered using functional data analysis as a foundation.

The HU method is summarized in three steps:

1. Before modeling, Hyndman and Ullah (2007) proposed to smooth the log death rates using penalized regression splines. Assume that there is a smooth continuous underlying function $S_t(x)$ which is observed with error at discrete ages.

$$Y_t(x_i) = S_t(x_i) + \sigma_t(x_i)\epsilon_{t,i}; \quad t = 1 \dots n; \quad i = 1 \dots p \tag{1.13}$$

   where $Y_t(x_i)$ denotes the log of observed death rates for age x_i at year t, $S_t(x_i)$ is the derived smooth function of x, $\sigma_t(x_i)$ is a noise component enables the amount of noise to vary with x in year t, hence correcting the assumption of homoscedastic error in the LC model, and the $\epsilon_{t,i}$ is an independent and identically distributed standard normal random variable.

   Estimation of the smoothing function $S_t(x)$ is done through x, while $y_t(x)$ by a nonparametric smoothing method $\forall t$.

2. The FPCA on the time series was applied to find the main sources of variability. A larger number of principal components were used to capture additional dimensions of changes in mortality rates. The FPCA is expressed as :

$$S_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x), \tag{1.14}$$

   where $S_t(x)$ is the derived smooth function of x, $\mu(x)$ is the mean function estimated by $\hat{\mu}(x) = \frac{1}{n}\sum_1^n S_t(x)$, $\{\phi_k(x)\} = \{\phi_1(x), \dots, \phi_k(x)\}$ is the set of first K functional principal components (orthonormal basis function), $\{\beta_k\} = \{\beta_1, \dots, \beta_k\}$ is the set of uncorrelated principal component scores, $e_t(x)$ is the residual function with mean zero and and K is the number of principal components used.

   The bivariate surface $S_t(x) - \mu(x)$ is approximated through the sum of products of the orthonormal basis functions $\{\phi_k(x)\}$ of age x and coefficients $\{\beta_k(x)\}$ of time t. The basis functions of equation (1.14) can be obtained by principal components and the uncorrelated coefficients are produced with PCA.

3. We are interested to forecast the future functional data h-years ahead. Therefore, the main idea is to use the obtained principal component scores to predict new values of $\hat{\beta}_{n,k,h}$, h-years ahead conditioned to the fixed set of functional principal components $\phi_k(x)$. The estimated coefficients are then used to approximate the functional data using equation (1.14). Hence, the h-step ahead forecast of $Y_{n+h}(x)$ is obtained by:

$$\hat{Y}_{n,h}(x) = E[Y_{n+h}(x)] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n,k,h}\phi_k(x); \tag{1.15}$$

where $\hat{\beta}_{n,k,h}$ is the h-step ahead forcast of $\beta_{n,k,h}$. The coefficicients, $\hat{\beta}_{t,k}$ and $\hat{\beta}_{t,l}$ are uncorrelated for $k \neq l$ so the univariate time series model such as ARIMA is used to predict the time-varying scores $\{\beta_{t,k}\}$.

The HU method can explain more variation of the demographic dynamics when we have data of high quality. However, its performance is comparable to the LC model when limited and scarce data sets such as Chinese data sets are used (Fang & Härdle, 2015).

1.3 Spatial data analysis

Statistics for spatial data was first explored in the 1960s in geology and meteorology. The main feature of spatial data analysis is to find a correlation structure between data observed at a given location and that available at neighbouring locations. These kind of data are widely available in economics, epidemiology, agriculture, environmental science and so on.
The past five decades witnessed the development of several methods in analysing spatial data and estimating values of a property of interest for the unsampled locations, from the available sample data. Data dependency is one of the practical considerations that influence the available techniques used in the spatial data modeling. Often, spatial data are dependent and a spatial model is tailored to handle this aspect. The realisation of the importance of spatial data analysis led to the development of statistical models to capture the spatial patterns (Cressie, 2015). A wide range of models and methods have been developed by considering the spatial dependence structure, mainly in the context of geostatistics, lattice data, and point patterns (Cressie, 2015). In the framework of geostatistics, the spatial location is valued in a continuous set of $\mathbb{R}^N$, $N \geq 2$. When compared to geostatistical and lattice data, spatial point patterns occurs when the locations of the available data are random. It is not always easy to distinguish the following three types of data:

1. **Geostatistical data**
   - The spatial set of interest $\mathcal{S} \subset \mathbb{R}^N$, $N \geq 2$ is a fixed subset of the plane of positive area (2-D) or volume (3-D).
   - A spatial process (collection of random variables observed at spatial points) $Y = \{Y(s),\ s \in \mathcal{S}\}$ is of interest.
2. **Lattice data**
   - The spatial process $Y = \{Y(s),\ s \in \mathcal{S}\}$ of interest is defined on a spatial fixed regular or irregular lattice $\mathcal{S}$ of $\mathbb{R}^N$.
   - This type of process includes extensions to well know time-series process.
3. **Point patterns**
   - The spatial locations $s \in \mathcal{S} \in\subset \mathbb{R}^N$ where the process $Y = \{Y(s),\ s \in \mathcal{S}\}$ is defined are random.
   - This type of process is an extension of the usual point processes.

The major distinction between time series and spatial data is the absence of an orderly relationship such as past, present and future i.e. the time axis is unidirectional. Past events may have an influence on the future while the reverse is not true. Time series models are not applicable to spatial data directly since the natural order in the time domain does not exist in the spatial context.

In this thesis, we are interested in the spatial lattice processes. For spatial lattice data, the locations form a lattice set. For lattice data, the seminal work by Besag (1974) is of great significance. The underlying idea is that a lattice has a neighbourhood structure and observations recorded on a lattice point are conditionally independent of the remaining lattice points given

the observation in the neighbouring lattice points (Kauermann et al., 2012). A comprehensive collection of applications and theory in this field are available in Rue and Held (2005) as well as Gaetan and Guyon (2010). Similar to other spatial processes, the exploration of the spatial correlation structure is the first step in the context of the lattice. The spatial weight matrix is a basic correlation tool in spatial econometrics describing the connectivity between different locations. The spatial weight matrices take different forms.

1.3.1 Specification of the spatial weight matrix

Spatial autocorrelation and its modeling are essentially based on the spatial weight matrices. In the spatial econometrics's literature, spatial dependency between spatial units is defined via the spatial weight matrix. W_n is a $n \times n$ non-stochastic weight matrix, it describes the spatial interactions between the n spatial units.
Formally, W_n is a positive $n \times n$ matrix with zero on the diagonal:

$$W_n = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,j} & \cdots & w_1,n \\ w_{2,1} & 0 & \cdots & w_{2,j} & \cdots & w_2,n \\ \vdots & \vdots & \ddots & \cdots & \cdots & \cdots \\ w_{i,1} & w_{i,2} & \vdots & 0 & \cdots & w_{i,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \\ w_{n,1} & w_{n,2} & \vdots & \vdots & \vdots & 0 \end{bmatrix}$$

where $w_{i,j}$ is the spatial weight between locations i and j.

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, Pinkse & Slade, 1998). More specifically, the weight matrices are classified into two groups: weights based on distance and weights based on boundaries. For weights based on distance, spatial weight matrices are constructed by using the distance d_{ij} between each pair of spatial units (regions, cities, centroids, ...) i and j. Weights based on distance are as follows:

- *k-Nearest Neighbour weights*
  $w_{ij} = \begin{cases} 1 & \text{if } j \in N_k(i), \\ 0 & \text{Otherwise,} \end{cases}$
  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 Euclidean 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, and it should be able to guarantee that each region has at least one neighbour.

- *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 and 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. The blocks are defined according to some specific criterion.

For weight matrices based on boundaries, the spatial contiguity is often used to specify neighbouring locations that share a common boundary. Various spatial contiguities are available in the literature. In the classical case of a regular square grid layout, the options of contiguity are referred to as the *Rook contiguity* (only common boundaries), *Bishop contiguity* (with only common vertices) and *Queen contiguity* (both boundaries and vertices). The contiguity weights are given as follows:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous,} \\ 0 & \text{otherwise.} \end{cases}$$

In general, the last equation can be rewritten as:
$w_{ij} = \begin{cases} 1 & \ell_{ij} > 0, \\ 0 & \ell_{ij} = 0, \end{cases}$
where ℓ_{ij} denotes the length of shared boundary.

1.3.2 The Moran's index

The basic principle of spatial data analysis is the idea that values of variables in nearby locations are closely related compared to those locations that are far apart.
The Moran's index is a measure of spatial autocorrelation. It was introduced first by Moran (1948, 1950) and developped by Cliff and Ord (1973, 1981). Moran's I is a correlation coefficient used to measure the overall spatial correlation in a data set and is bounded by 1 and -1. The spatial autocorrelation in terms of Moran's I can be classified as follows:

- positive autocorrelation occurs when Moran's I is close to $+1$.
  The spatial correlation is positive when similar values cluster together on a map.

- negative autocorrelation occurs when Moran's I is close to -1.
  The spatial correlation is negative when dissimilar values cluster together on a map.

- a Moran's I value of 0 denotes the absence spatial autocorrelation.

The results of this test is interpreted in the context of a null hypothesis which assumes that the random of distributed among locations.
Suppose that, in at n locations $\{s_1, \ldots, s_n\}$, we observe $y_1 = y(s_1); \ldots; y_n = y(s_n)$. We also suppose that spatial weights w_{ij} , the weights between each pair of spatial units s_i and s_j, which satisfy:

$$w_{ij} \geq 0 \quad \text{for any } i \neq j, \qquad w_{ii} = 0 \text{ for any } i.$$

The Moran's I index is given by:

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{1.16}$$

where n is the number of spatial units indexed by i and j, $\bar{y}$ is the mean of y and w_{ij} is a matrix of spatial weights with zeroes on the diagonal and W is the sum of all w_{ij}.

In this thesis, the Moran's I is extended to the functional data analysis context which was discussed in Section 1.1. This aims to consider the spatial dependency in the PCA to analyze the degree of spatial autocorrelation among observations in the geographic space with the idea of spatial principal component analysis, sPCA (Jombart et al., 2008) which highlights spatial patterns by the direct consideration of spatial information.

1.3.3 Functional PCA for spatial data

There are many functional data sets taking the form : $X_s(t), s \in S$ and $t \in [0, T]$. where s is a spatial location and $[0, T]$ is the interval of time. In such data, the spatial dependence are often not taken into account. Note the famous example of the Canadian temperature data (Ramsay & Silverman, 2005) available at 35 locations, the curves have very similar characteristics when they are quite close. The research on spatial functional data is growing, and there are few research on the fundamental properties of spatially distributed functional data. Delicado et al. (2010) provided a useful approach which involves the integration of classical types of spatial data structures (geostatistical data, point patterns, and areal data) with functional data. FDA approaches were applied to PCA in a spatial framework especially in geostatistical and point patterns((Li & Guan, 2014); (Liu et al., 2014); (Hörmann & Kokoszka, 2011);(Liu et al., 2014); (Illian et al., 2006)). Gromenko et al. (2012, 2017) use a geostatistical framework for spatially indexed functional data to solve space physics problems. Kuenzer et al. (2020) proposed a dimension reduction technique suitable for functional data, indexed by spatial locations on a grid. So far, there is no work on FPCA on spatial data in the the lattice framework.
The idea of the HU model in equation (1.13) motivated the use of functional data analysis in our study. The versatility of functional data analysis is appealing because it can be adapted to many areas of mortality forecasting. The smoothed data are ready for conversion to functional objects and after that various approaches can be implemented so these data can be investigated from different angles. This thesis investigates mortality from a spatial point of view instead of the usual approaches which give emphasis to the time aspect of mortality models. It is of interest to investigate the existence of spatial autocorrelation among neighbouring countries by adapting dimensionality reduction techniques using FPCA.
The sPCA by Jombart et al. (2008) aimed to reveal spatial patterns by directly considering

spatial information. The sPCA analyses a matrix of relative allele frequencies $\mathbf{X}$ where the spatial information is stored inside a spatial weighting matrix L, and defines the following function to measure both the spatial structure and variability in $\mathbf{x}$:

$$\mathcal{C}(\mathbf{x}) = \text{var}(\mathbf{x})\ I(\mathbf{x}) = \frac{1}{n}\mathbf{x}^T L\mathbf{x}. \tag{1.17}$$

Interestingly, spatial autocorrelation can efficiently be captured from functional data using the newly derived functional Moran's I statistic in this thesis. This statistic is the first of its kind in the functional data framework. 28 European countries whose mortality data are available on HMD, for a fixed period since it is of interest to investigate the existence of spatial autocorrelation in mortality data of neighbouring countries are considered. Once this approach successfully captures the spatial autocorrelation, it will be of great help to use this idea to construct spatial predictive models to predict mortality rates for neighbouring countries whose data are unavailable on HMD.

1.3.4 Spatial econometric models

Modelling spatially dependent data requires correlation between random variables in one location with those in neighbouring locations (Pinkse & Slade, 1998). In this thesis, the lattice type data is examined. Statistical models for lattice data are linked to nearest neighbours to express the fact that data are nearby. Two popular spatial dependence models for lattice data are the spatial autoregressive (SAR) dependent variable model and the spatial autoregressive error model (SAE, where the model error is an SAR), which extend the regression in a time series to spatial data.

Let (Y, X) be a random vector observed at n locations where $\{s_1, \ldots, s_n\}$ in an irregularly spaced, countable lattice $\mathcal{I} \subset \mathbb{R}^k,\ k \geq 2$ such that $||s_i - s_j|| \geq d_0$, with $d_0 > 0$. Suppose that $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^T$ is the sample response and $\mathbf{X}_n$ the $n \times p$ matrix of explanatory variables observations with elements $X_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, p$.
To explain why an observation located in a specific location depends on observations made at other locations, it is necessary to determine the three types of interaction effects:

- Endogenous interaction effects among dependent variables. The variable Y_i at spatial units i depends on Y_j at spatial units j. This model is named as the spatial lag model or spatial autoregressive (SAR) model (Cliff & Ord, 1973),where the interaction effect is denoted by the spatial lag $W_n\mathbf{Y}_n$.
- Exogenous interaction effects among independent variables where the variable Y_i at a spatial unit i depends on the independent explanatory variables X_j at spatial units j.
- Correlated effects, where similar unobserved characteristics result in similar behavior. Here, the interaction is among the error terms. This model is known as the spatial autoregressive error (SAE) model (or spatial error model; SEM).

In practice, a population that contains these three types of interactions jointly does not exist. Researchers always focuses on models with one interaction such as the SAR model, the SAE model, or a model with two interactions. These kind of models are used when the spatial autocorrelation affects the response and the error terms. According to the terminology developed by Lesage (2008), we refer to this model as the spatial autocorrelation (SAC) model:

$$\begin{aligned}\mathbf{Y}_n &= \lambda_0 W_n \mathbf{Y}_n + \mathbf{X}_n \beta_0 + \mathbf{U}_n;\\ \mathbf{U}_n &= \gamma_0 W_n \mathbf{U}_n + \varepsilon_n,\\ \varepsilon_n &\sim N(0, \sigma_0^2 I_n),\end{aligned} \tag{1.18}$$

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, which 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. $W_n \mathbf{U}_n$ represent the interaction effects among the disturbance terms of the different spatial units. However, the SAR model is a SAC model with $\gamma_0 = 0$ and SAE model is a SAC model with $\lambda_0 = 0$.

From a theoretical point of view, various linear spatial regression SAR and SAE models as well as their identification and estimation methods, e.g., two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood (ML) or quasi-maximum likelihood (QML) and the generalized method of moments (GMM), have been developed and summarized by many authors such as Anselin (2013), Kelejian and Prucha (1998), Kelejian and Prucha (1999), Conley (1999), Cressie (2015), Case (1993), L.-F. Lee (2004), L.-f. Lee (2007), Lin and Lee (2010), Zheng and Zhu (2012), Malikov and Sun (2017), Garthoff and Otto (2017), Yang and Lee (2017).

1.3.5 Semi-parametric modeling

A semi-parametric model is an alternative to fully parametric models when there exists a non-linear relationship between the discrete binary variable and some explanatory variables. This type of model is known as the partially linear probit model for spatially dependent data. A triangular array setting is used to cover various patterns of spatial data. Allowing various spatial dependencies, we assume the existence of conditional spatial heteroscedasticity, non-identically distributed observations, and a linear process for disturbances.

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^*:

$$Y_{in}^* = X_{in}^T \beta_0 + g_0(Z_{in}) + U_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{1.19}$$

with

$$Y_{in} = \mathbb{I}\left(Y_{in}^* \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{1.20}$$

where $\mathbb{I}(\cdot)$ is the indicator function; X and Z are explanatory random variables taking values in the two compact 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 $\Theta_\beta \subset \mathbb{R}^p$; and $g_0(\cdot)$ 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 $\mathbb{R}$ and C a positive constant. In model (3.1), β_0 and $g_0(\cdot)$ are constant over i (and n). Assume that the disturbance term U_{in}

in (3.2) is modelled by the following spatial autoregressive process (SAR):

$$U_{in} = \lambda_0 \sum_{j=1}^{n} W_{ijn} U_{jn} + \varepsilon_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{1.21}$$

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 weight matrix W_n, which 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 (Pinkse & Slade, 1998).

The estimation procedure is a combination of a weighted likelihood (Staniswalis, 1989) and a generalised method of moments (Pinkse & Slade, 1998). The procedure first fixes the parametric components of the model and then estimates the non-parametric part using weighted likelihood; the obtained estimate is then used to construct a GMM (Generalised Method of Moments) parametric component estimate. The consistency and asymptotic distribution of the estimators are established under sufficient conditions. Some numerical results are provided to investigate the finite sample performance of the estimators.

In Chapter 3 of this thesis, a semi-parametric estimation method combining the GMM approach and the weighted likelihood method is proposed. The parametric components of the model are first fixed and the non-linear components are estimated by the weighted likelihood approach (Staniswalis, 1989). The estimator which is obtained depends on the values at which the parametric components are fixed. This is used to construct a GMM estimator (Pinkse & Slade, 1998) of these components.

In chapter 4, spatial probit regression models are applied to study suicidal relapses to shed light on the impact of neighbouring locations on suicide cases by investigating the impact of the VigilanS system on suicide relapses of patients involved in the VigilanS system.

The VigilanS healthcare system is an effort to support those who have attempted suicide in various regions. It was established in Nord-Pas-de-Calais in February 2015. This program to monitor and prevent recidivism of suicide attempts is executed via phone calls by teams of professionals who are specialized in this type of remote care. In the Lille University Hospital, this six-month programme is managed by the adult psychiatry department under Professor Guillaume Vaiva. Posthospital support is offered to those patients who attempted suicide. When these patients are discharged from the hospital, they receive a resource card with a dedicated contact number and are called back for follow-up after 6 months. Those with a history of suicide reattempts are contacted between the 10th to 21st day after being discharged from the hospital.

We aim to investigate the impact of spatial dependency on suicide recidivism in Nord-pas-de-Calais and examine the effect of non-linear explanatory variables such as median revenue, unemployment rates etc by applying several frameworks of the probit regression to model the suicide recidivism, especially after 6 months from the entry of VigilanS system.

Chapter 2

Exploring spatial patterns of mortality in Europe using functional spatial principal components for areal data

Abstract

We examined the spatial autocorrelation of mortality rates for 28 European countries, with data from the Human Mortality Database (HMD) using spatial associations in the context of functional areal data. We developed a functional Moran's I statistic which is the first of its kind in the functional data analysis framework to determine spatial autocorrelation and spatial PCA for areal data. These data were converted to functions before performing the classical and spatial PCA. Results showed the existence of spatial autocorrelation between neighbouring countries (using K-nearest neighbours (KNN) and contiguity neighbours) based on the functional Moran's I statistic applied on the functional PCA approximation. However, no strong correlation was displayed when the scores of the classical PCA which ignored spatial information, was considered. This work proved the existence of spatial dependency in mortality rates of neighbouring countries in Europe and showed that the functional Moran's I statistic is a powerful tool in measuring spatial dependency.

Keywords: functional principal component analysis, Moran's I, spatial autocorrelation, KNN, contiguity, mortality.

2.1 Introduction

Mortality rates depict the population health and economic status of a country. Spatial demographic models play a major role in monitoring spatial dependence of mortality. Demographers, social scientists, economists, and many others have been extensively studying mortality patterns. Most researches available in the literature are inclined to mortality models with emphasis on temporal forecasting methods. Hence, we were motivated in investigating mortality from the spatial framework.

The Lee-Carter (LC) model (Lee & Carter, 1992) is one of the most prominent contributions to the study of mortality models which opened the path to various innovations of its kind. The

Hyndman and Ullah (2007) (HU) model extended the LC model by incorporating functional data analysis (FDA), nonparametric smoothing and robust statistics. Hyndman et al. (2013) applied coherent forecasting of mortality rates for two or more subpopulations based on functional principal components models of simple and interpretable functions of rates to sex-specific data for Sweden and state-specific data for Australia. Greco and Scalone (2013) combined mortality modelling techniques with the Bayesian approach in forecasting mortality rates by age and sex for provincial areas in Italy. Based on a study on county level mortality rates of the United States of America, Raymer et al. (2018) argued that the mortality rate of a certain county may be associated with the features of its neighbouring counties beyond its own features. S. Wang and Ren (2019) explored the spatial distribution patterns and economic determinants of China by calculating the four indexes (lifespan expectancy at birth, infant mortality rate, under-5 mortality rate and crude mortality rate) at county level in China and illustrated the spatial distribution of these patterns.

Principal component analysis (PCA) is of the essence in FDA. The combination of these elements is of great significance in enhancing spatial demographic modelling. Delicado et al. (2010) provided a useful approach which involves the integration of classic types of spatial data structures (geostatistical data, point patterns, and areal data) with functional data. FDA approaches were applied to PCA in a spatial framework (Li and Guan (2014); Liu et al. (2014)). Kuenzer et al. (2020) proposed a dimension reduction technique suitable for functional data, indexed by spatial locations on a grid. So far, there is no work on functional PCA (FPCA) on spatial data in the areal framework.
For a fixed country, Hyndman and Ullah (2007) applied FPCA to decompose smoothed functional time series into a set of functional principal components and their principal component scores. FPCA was applied to find the main sources of variability. Léger and Mazzuco (2021) showed that a functional framework can be informative where it allows clustering of complete mortality profiles without losing sight of the role played by single components where the changes of age-specific mortality in low-mortality countries in the last few decades with functional clustering were investigated. Léger and Mazzuco (2021) suggested three different methods of functional clustering of mortality profiles (seen as curves over age, which can be observed for every country and every year) by: a two-stage method based on spline coefficients, a distance-based method through FPCA and a model-based method.

In this paper, we use the functional data analysis approach since it works with smooth curves rather than scalar data. We aim to investigate the spatial relationship of mortality of neighbouring countries in Europe by employing a more specific form of principal component analysis developed to reduce multidimensionality in geo-referenced genetic data. This form, known as the spatial principal component analysis (sPCA) was introduced by Jombart et al. (2008) and was formulated to investigate the spatial pattern of genetic variability using allelic frequency data of individuals or populations. sPCA is effective in revealing spatial connections in mortality data compared to the classical PCA. In our study, we consider 28 European countries whose mortality data are available on the Human Mortality Database (HMD) for a fixed period since it is of interest to investigate the existence of spatial autocorrelation in mortality data of neighbouring countries. The idea of sPCA by Jombart et al. (2008) aimed to reveal spatial patterns by directly considering spatial information. We constructed a functional Moran's

I statistic which will aid in determining spatial autocorrelation in functional data through the implementation on the spatio-functional PCA. Our functional Moran's I statistic is the first of its kind from the functional data framework as no work has been done so far from this perspective.

Section 2.2 gives a brief description of the data used, Section 2.3 explains the methodology used, Section 2.4 discusses the results and Section 2.5 concludes.

2.2 Description of data

We consider mortality rates for 28 European countries (Figure 2.1) available on HMD for ages 0 to 110 (where ages above 100 are grouped as 100+) to investigate the spatial dependency at a fixed time, where years 1990, 2000 and 2010 were investigated separately. Lee and Carter (1992) modeled the logs of the age-specific death rates as a linear function of an unobserved period-specific intensity index, with parameters depending on age. We followed suit but had to translate the mortality rates by adding a constant (the smallest value of death rates of the 28 countries studied) to each death rate value before taking the natural logarithm. This step is crucial to avoid taking logarithms of zeros because the death rates for certain age groups from Luxembourg were zero. Each colour indicates a country in the log of death rates plots for males and females (Figures 2.2(a) and 2.3(a)).

Figure 2.1: The map of European countries with mortality data from HMD.
Note: Data for the red shaded region is not available on HMD.

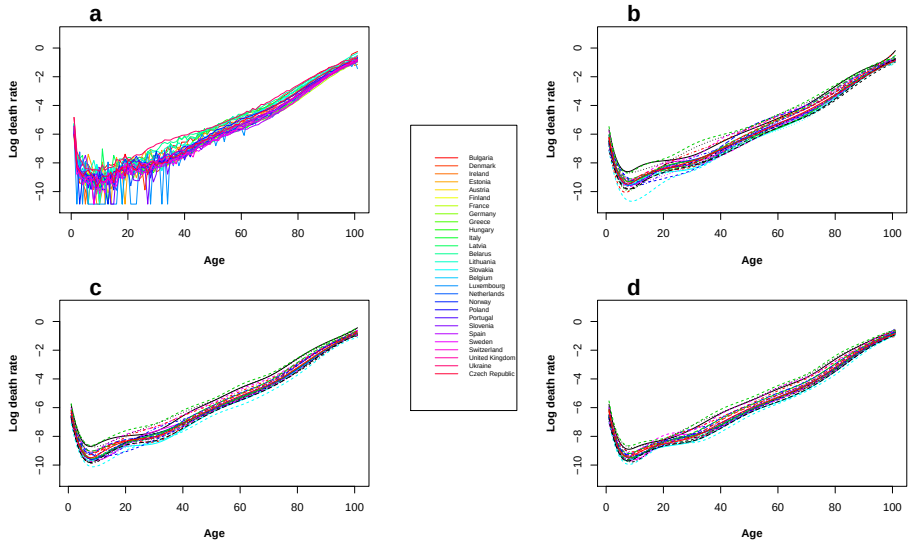


Figure 2.2: The female log death rates in 2010 for 28 European countries: (a) observed, (b) smoothed, (c) reconstructed using the matrix multiplication of the scores and principal components based on contiguity weights, (d) reconstructed using the matrix multiplication of the scores and principal components based on KNN weights.

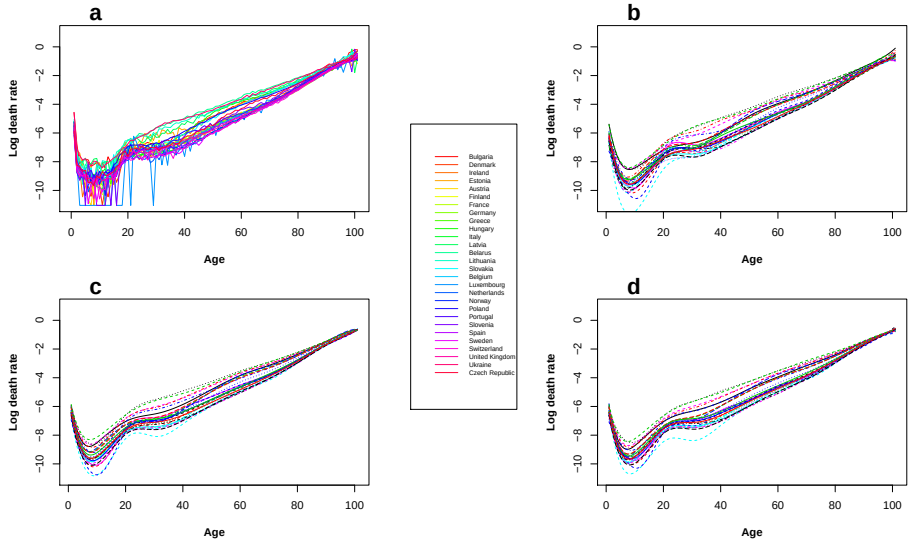


Figure 2.3: The male log death rates in 2010 for 28 European countries: (a) observed, (b) smoothed, (c) reconstructed using the matrix multiplication of the scores and principal components based on contiguity weights, (d) reconstructed using the matrix multiplication of the scores and principal components based on KNN weights.

2.3 Methodology

2.3.1 Functional principal component analysis on areal spatial data

Consider n spatial locations i, one observed discrete measurement $Y_{i,x,t}$ taken at time t of location $i \in \mathcal{I} \subset \mathbb{Z}^2$, $\mathcal{I}$ a lattice region V, for a given $x \in \mathcal{X} = [0, T] \subset \mathbb{N}$. In our setting x is an age between 0 and $T = 100$ and $Y_{i,x,t}$ is a mortality rate observed at a year $t \in \mathcal{D} \subset \mathbb{N}$ for a country of

location i. We assume that for a given t these measurements points $Y_{i,x,t}$ are noisy observations of a smooth areal stochastic functional process $\{X_i\}_{i\in\mathcal{I}}$:

$$Y_{i,x,t} = \mu_t(x) + S_{i,t}(x) + \epsilon_{i,x,t} = X_{i,t}(x) + \epsilon_{i,x,t} \tag{2.1}$$

where μ_t is the mean function at time t (year). The n functions $S_{i,t}(.)$ are centered spatio-temporal squared integrable functional random variables on the space-time domain $\mathcal{I}\times\mathcal{D}$, namely $S_i(.)$ is valued in the Hilbert space $L^2(\mathcal{X})$ endowed with the inner product $\langle f, g\rangle = \int_{\mathcal{X}} f(x)g(x)dx$, for f, g in $L^2(\mathcal{X})$. The unobserved variables $\{\epsilon_{i,x,t},\ i = 1, ..., n\}$ are i.i.d with zero mean Gaussian measurement errors and variance σ^2.

We are interested in a functional PCA study where the classical PCA is replaced with its spatial counterpart, to consider spatial autocorrelation on the variable of interest in the sampling locations. This autocorrelation may be quantified by a weight matrix depending on the neighbour locations.

Let us consider in the following, the measurements of a given time t, ignoring the temporal distribution of different years. Let us fix the time t and delete the subscript t in equation (2.1), and consider $S_{i,t}(x) = S_i(x)$ as a spatial functional variable and postulate a Karhunen-Loève expansion (Ash & Gardner, 1975):

$$S_i(x) = \sum_{k=1}^{\infty} \beta_{k,i}\phi_k(x), \tag{2.2}$$

where ϕ_k's are the orthonormal eigenfunctions (functional principal components, FPC) and $\beta_{k,i}$ are auto-correlated scores. In practice, the sum is truncated to a finite integer, K which is to be chosen.

To compute the FPCs, let us express the sample data $(S_i)_{i=1,...,n}$ by means of a basis expansion:

$$S_i(x) = \sum_{m=1}^{\infty} c_{i,m}B_m(x) \approx \sum_{m=1}^{p} c_{i,m}B_m(x), \quad x \in \mathcal{X}, \tag{2.3}$$

where $B_m(.)$ is some collection of basis functions, $c_{i,m} = \langle S_i, B_m\rangle$ have zero-mean. In practice, the first p functions are used where a sufficiently large p is good for approximation.

Ramsay and Silverman (2005) presented two main basis systems for building functions. The Fourier basis system is commonly used for periodic data while the B-spline basis system is preferable for nonperiodic data (Ramsay et al., 2009). For the log death rates data, the B-spline basis system provides a more flexible basis. The selection of the number of basis functions is vital where a large number can lead to overfitting while a small number may cause underfitting. The smoothing degree depends on the aim of the analysis. An important point we took note in setting the number of basis functions is the selection of a number less than the number of countries studied. The selection was made using step-by-step cross validation. In our case, we fit the data using 13 basis functions. The smoothing of the curves using B-spline was performed using the `fda` (Ramsay et al., 2020) package in the R software.

Estimation of the principal components and functional Moran's I statistic

Let us extend the well know Moran's statistic to the functional context. This aims to take into account the spatial dependency in the principal component analysis to analyze the degree of

spatial autocorrelation among observations in the geographic space $\mathcal{I}$ (Jombart et al., 2008).

Let $W = (W_{ij})$ be a weighted spatial matrix where W_{ij} is the neighbouring relation between locations i and j. Let W be standardised where the rows sum to one.
The functional Moran's index of the n row vector $\{S_i(x)\}_{i=1,...,n}$ is introduced:

$$I_n(\mathbf{S}(x)) = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} S_i(x) S_j(x)}{\sum_{i=1}^n S_i(x)^2} = \frac{C_n(\mathbf{S}(x))}{\sigma_n(\mathbf{S}(x))} \ , \tag{2.4}$$

where

$$\begin{aligned} C_n(\mathbf{S}(x)) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n W_{ij} S_i(x) S_j(x) \\ &\approx \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^p \sum_{l=1}^p W_{ij} c_{i,m} c_{j,l} B_m(x) B_l(y) \\ &= \frac{1}{n} \mathbf{B}(x)^\top \mathbf{X}^\top W \mathbf{X} \mathbf{B}(x), \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \sigma_n(\mathbf{S}(x)) &= \frac{1}{n} \sum_{i=1}^n S_i(x)^2 \\ &\approx \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^p \sum_{l=1}^p c_{i,m} c_{i,l} B_m(x) B_l(y) \\ &= \frac{1}{n} \mathbf{B}(x)^\top \mathbf{X}^\top \mathbf{X} \mathbf{B}(x). \end{aligned} \tag{2.6}$$

$\mathbf{X}$ is the $n \times p$ matrix composed of the scores $(c_{i,m})_{i=1,...,n;m=1,...p}$ of S_i, $\mathbf{B}(x)$ is the $p \times 1$ vector of components $B_m(x)$, $m = 1, ..., p$, $\mathbf{S}(x)$ is the $p \times 1$ vector of functions $S_i(x)$.

The trace functional Moran's index is then introduced as:

$$I_n(\mathbf{S}) = \int_0^T I_n(\mathbf{S}(x)) dx. \tag{2.7}$$

The classical univariate Moran's index (Eckardt & Mateu, 2020; Jombart et al., 2008) of a n raw vector $\mathbf{X}_m$ of components $\{c_{i,m}\}_{i=1,...,n}$ is

$$\tilde{I}(\mathbf{X}_m) = \frac{\mathbf{X}_m^\top W \mathbf{X}_m}{\mathbf{X}_m^\top \mathbf{X}_m}.$$

Let

$$V(\mathbf{X}_m) \ = \frac{1}{n} (\mathbf{X}_m^\top \mathbf{X}_m) \tilde{I}(\mathbf{X}_m) = \ \frac{1}{n} \ \mathbf{X}_m^\top W \mathbf{X}_m.$$

It is highly positive when $\mathbf{X}_m$ has a large variance and shows a global spatial structure and is negative in a situation with high variance and gives a local structure.

The purpose of the functional areal spatial principal component (FASPCA) proposed here is based on scaled $\mathbb{R}^p$ vectors $\mathbf{u}$ (loadings) ($\|\mathbf{u}\|$=1) such that the n raw vectors $\chi = \mathbf{X}\mathbf{u}$ are scattered and spatially autocorrelated. In other words, this aims to find the extreme values (Jombart et al., 2008) of

$$C(\mathbf{u}) = V(\mathbf{X}\mathbf{u}) = \frac{1}{n} \mathbf{u}^\top \mathbf{X}^\top W \mathbf{X} \mathbf{u}. \tag{2.8}$$

The solutions (Jombart et al., 2008) are the eigenvectors $\mathbf{u}_k$ of $\frac{1}{2n}\mathbf{X}^\top(W+W^\top)\mathbf{X}$ associated with the largest and smallest eigenvalues $\alpha_k = var(\chi_k)\tilde{I}(\chi_k)$ (where $\chi_k = \mathbf{X}\mathbf{u}_k$, $var(\chi_k)$ the variance of χ_k). Note that some eigenvalues α_k may be negative since $\tilde{I}(\chi_k)$ is not always positive.
By the help of orthonormal vectors $\mathbf{u}_k$ and their eigen-values α_k, we introduce the estimated functional loading (eigen-function), $\hat{\phi}_k(x)$ of the functional spatial areal PCA.
In fact, approximating $\mathbf{X}$ by

$$\mathbf{X} \approx \widehat{\mathbf{X}} = \sum_{k=1}^{K} \chi_k \mathbf{u}_k^\top,$$

based on K (sufficiently large) relevant scores χ_k corresponding to the K largest (in absolute values) eigen-values, lead to

$$\mathbf{S}(x) \approx \widehat{\mathbf{X}}\mathbf{B}(x) = \sum_{k=1}^{K} \chi_k \mathbf{u}_k^\top \mathbf{B}(x).$$

The functional spatial PCA is then obtained by letting the estimated eigen-functions as $\hat{\phi}_k(x) = \mathbf{u}_k^\top \mathbf{B}(x)$ and the n-row functional scores $\hat{\beta}_k = \langle \mathbf{S}(.), \hat{\phi}_k(.) \rangle$.

Then the FASPCA decomposition is obtained using equation (2.3) where the orthonormality of the vectors $\mathbf{u}_k$ and the functions B_m gives:

$$S_i(x) \approx \sum_{k=1}^{K} \hat{\beta}_{k,i}\hat{\phi}_k(x), \tag{2.9}$$

$$X_i(x) \approx \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{k,i}\hat{\phi}_k(x), \tag{2.10}$$

where $\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} X_i(x)$, is the empirical mean with $\hat{\beta}_{k,i} = \int_0^T S_i(x)\hat{\phi}_k(x)dx$.

2.3.2 Implemention of the functional Moran's I statistic on spatial weight matrices

This study aims at examining if spatial autocorrelation exists in the mortality rates for 28 European countries with data from HMD. The Moran's I statistic was extended and applied to the functional context. To ensure robustness, we used spatial weight matrices such as the k-nearest neighbour (KNN) and contiguity matrices. We constructed a KNN matrix of the 28 countries studied, belonging to the set of the five nearest neighbours of each other. For the contiguity weight matrix, we built a list of neighbours based on these 28 countries with contiguous boundaries where the single shared boundary point meets the contiguity condition. The weighted spatial matrix W_{ij} can be classified into weights based on distance and weights based on boundaries. We use the distance d_{ij} between each pair of spatial units (regions, cities, centroids, ...); i and j to construct spatial weight matrices for weights based on distance. The k-nearest neighbour weights are given as

$$w_{ij} = \begin{cases} 1 & \text{if } j \in N_k(i); \\ 0 & \text{otherwise;} \end{cases}$$

where $N_k(i)$ is the set of the k closest units or regions to i for $k \in \{1, \ldots, n-1\}$.

For weights based on boundaries, we often use spatial contiguity to specify neighbouring locations in the sense of sharing a common border. There are various types of spatial contiguities but the classical cases are those known as the Rook's contiguity (where two cells of a matrix which share a common boundary are neighbours), the Bishop's contiguity (where two cells of a matrix share a common vertice) and the Queen's contiguity (neighbours by either the Rook's or the Bishop's contiguity). The contiguity weights are given as

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous;} \\ 0 & \text{otherwise.} \end{cases}$$

A Shapiro-Wilk test to assess multivariate normality on the log of death rates data is required to investigate if the data fulfills the normality assumption. If the data fulfils this assumption, we can proceed with computing the functional Moran's I statistic for the KNN- and contiguity-based neighbourhoods to show spatial autocorrelation of log death rates. If the data violates the normality assumption, we are required to run a Monte Carlo simulation on the Moran's I statistic. It is essential to investigate the global autocorrelation which measures the degree of clustering as well as the local indicators which allows the decomposition of the Moran's I global indicator into the contribution of each observation. The `spdep` (Bivand & Wong, 2018) package in the R software was used to aid the use of spatial weights as well as the Moran's I statistic to measure the spatial autocorrelation.

For the first part of the study, the classical functional PCA is performed on the smoothed data (Figures 2.2(b) and 2.3(b)). The classical functional PCA does not account for spatial information. Hence, the application of FASPCA enhanced with the implementation of the multivariate (non-functional) spatial pca (sPCA) by Jombart et al. (2008). sPCA which complements PCA, was introduced to explicitly include spatial information in the analysis of genetic variation for investigating spatial genetic structures (Jombart et al., 2008). We employed FASPCA to reduce the dimensionality of data in our study since spatial information is a vital aspect in studying spatial autocorrelation. The PC scores of FASPCA consist of two types of patterns which are defined as global and local structures (Jombart et al., 2008). A global pattern differentiates between two spatial groups or a cline (or any intermediate state) while local scores retrieve stronger genetic differences among neighbours than among random pairs of entities (Jombart et al., 2008). The global pattern corresponds to positive spatial autocorrelation while the local pattern corresponds to negative spatial autocorrelation (Jombart et al., 2008). FASPCA was implemented using the `fda` (Ramsay et al., 2020), `adegenet` (Jombart, 2008), `ade4` (Bougeard & Dray, 2018; Chessel et al., 2004; Dray & Dufour, 2007; Dray et al., 2007) and `adespatial` (Dray et al., 2019) packages from the R software.

2.4 Results and discussion

The log death rates data for the 28 countries involving males and females described previously are analysed with the proposed FASPCA. First, the data are smoothed for convertion to functional objects. Figures 2.2(b) and 2.3(b) represent the smoothed log death rates data using B-splines for the male and female populations in year 2010. The smoothed data for years 1990 and 2000 portray similar characteristics to those in Figures 2.2(b) and 2.3(b).

We then investigate the presence of spatial autocorrelation among the data. The log death rates data for the 28 countries involving males and females do not satisfy the normality assumption. We performed the Shapiro-Wilk test to assess multivariate normality (`mvnormtest` R package (Jarek, 2012)) on the log death rates data and found that this data violated the normality assumption. Hence, permutation tests for the Moran's I statistics are calculated for these data using 999 random permutations of the log death rates for all cases studied based on the KNN and contiguity weighting schemes. Table 2.1 shows the existence of significant spatial autocorrelation for female and male log death rates for years 1990, 2000 and 2010 for the 28 countries. In this table, a classical Moran's I index is calculated considering for each year and gender, the raw data matrix as a panel dataset. Positive spatial autocorrelation indicates that locations nearby tend to be similar on map where high values tend to be near high values and low values near low values. On the other hand, when geographic values are dissimilar, the map shows negative spatial autocorrelation. The Moran's I statistics reported values close to +1 for both males and females in years 1990, 2000 and 2010. This suggests that neighbouring locations have strong positive autocorrelation in mortality for both KNN and contiguity neighbours. The Moran's I index reported in Table 1 are somehow aggregations of the functional index (equation (2.4)) like the functional trace index defined in equation (2.7).

Table 2.1: Moran's test for spatial autocorrelation based on log of death rates for females and males in 28 countries in Europe based on KNN and contiguity weights.

	Female			Male		
Year	1990	2000	2010	1990	2000	2010
KNN	0.9857***	0.9835***	0.9842***	0.9770***	0.9814***	0.9831***
Contiguity	0.9846***	0.9756***	0.9734***	0.9782***	0.9802***	0.9699***

Note: *** p-value < 0.001

Figure 2.4 illustrates the smoothed functional Moran statistics for ages 0 to 100+ using B-splines for the female and male log death rates data based on the KNN and contiguity weights. Figure 2.4 also highlights that the functional Moran's I statistics showed comparable behaviour for the KNN and contiguity weights for all three years. For females, higher spatial dependency is visible from approximately 20 years of age to 80 years of age. The similar attributes are observed in the log death rates of males. Generally, for ages approximately 80 years and above, spatial autocorrelation decreases because the number of people who live past 80 is low. The life expectancy for the countries in Europe is between 70 to 80 years of age, from 1990 to 2010 (European Union Data, 2020). This contributed to the high spatial dependency (approximately 0.6) for these ages. We also observed that the spatial dependency increased from 1990 to 2010. The formation of the European Union on 1 November 1993 could be the possible reason for the spatial dependence of mortality rates in these countries to increase over time as one of the objectives of the European Union is towards improving the quality of life of the population through cross-border cooperation, especially in healthcare.

We performed the classical FPCA to find new functions that reveal the most important types of variation in the curve data with the absence of the spatial structure. Table 2.2 gives the results for this FPCA on the male and female log death rates data for the aforesaid years and an autocorrelation Moran's test on the scores. The first PC reveals autocorrelation for both genders

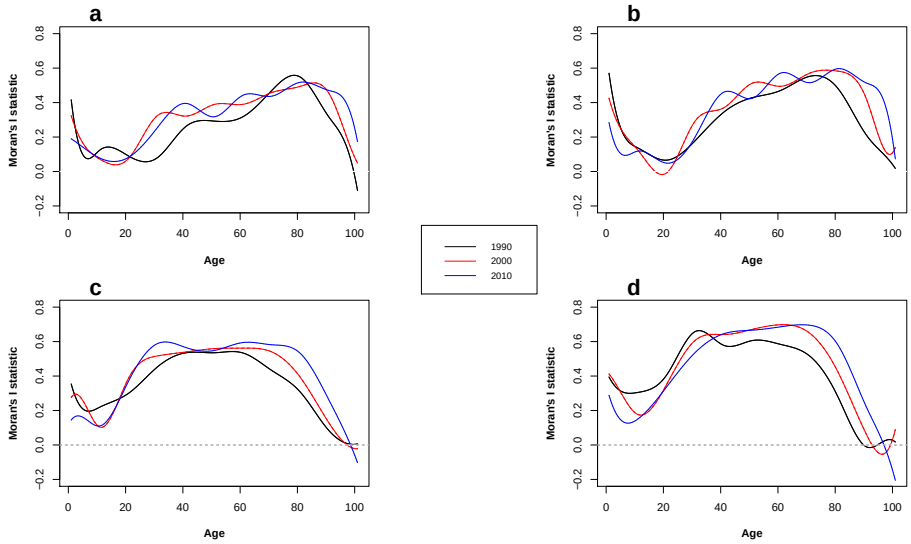


Figure 2.4: Smoothed functional Moran's I statistics curves for log death rates from ages 0 to 100+ of: (a) females using KNN, (b) females using contiguity, (c) males using KNN, and (d) males using contiguity, weight matrices for years 1990, 2000 and 2010.

for the three years studied. Table 2.2 also reveals that the first PC alone accounts to more than 99.5% of the total variability of the data for all cases where we can see the data alone on one dimension. This does not give a clear picture of the data due to the absence of the spatial factor in the method.

We then proceeded to perform FASPCA for KNN and contiguity neighbours for the basis functions of the male and female log death rates data for all three years. Figure 2.5 gives a picture of the global structures (positive spatial autocorrelation, where the log of death rates are similar at neighbouring locations) and local structures (negative spatial autocorrelation, where the log of death rates tend to be dissimilar at neighbouring locations) to be retained for female and male data using KNN weights (Figure 2.6 for contiguity weights). Then, we ran the FASPCA described in Section 2.3 for the cases studied by considering the top three positive and top two negative eigen-values (Table 2.2). The percentage of variability explained by the functional principal components of FASPCA are given in Table 2.2.

The functional Moran's I statistics calculated based on these functional principal components show significant spatial autocorrelation for the principal components reported in Table 2.2 for KNN and contiguity weights (refer to p-values). Spatial autocorrelation can effectively be detected from the spatial principal components of the functional data which explains more than 95% of the percentage of variability based on the KNN and contiguity weights involving males and females. We reconstructed the data for each case by using the top two positive and top one negative PCs (Figures 2.2(c,d) and 2.3(c,d)). These PCs were mapped onto geographic spaces (representing 28 European countries generated using the `maps` package (Becker et al., 2018) and the `rgdal` package (Bivand et al., 2019) where the black and white squares of the variable size represent positive and negative scores of the PCs respectively. The large black squares are well differentiated from the large white squares, while the small squares are less differentiated. The area of the square is proportional to the absolute value of the score. These graphical representations are applied to the significant PCs of each case for KNN and contiguity neighbours.

Table 2.2: Moran's test on principal components using FPCA and FASPCA, based on KNN and contiguity weight matrices for females and males of 28 European countries in 2010.

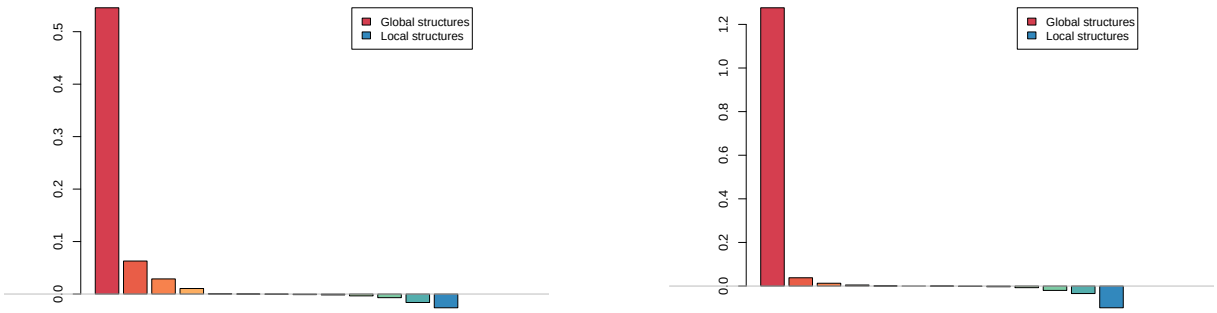
	Female		Male	
	Moran's I	variability (%)	Moran's I	variability (%)
Classical FPCA				
1^{st} score	0.4973^{***}	99.81	0.5482^{***}	99.67
2^{nd} score	0.0402	0.08	-0.0873	0.14
3^{rd} score	0.2305^{**}	0.04	0.2687^{**}	0.08
4^{th} score	0.1448^{*}	0.03	-0.0548	0.04
Total		99.96		99.93
FASPCA (KNN (3,2))				
1^{st} score positive	0.5400^{***}	77.74	0.6021^{***}	85.14
2^{nd} score positive	0.3442^{**}	8.93	0.3239^{**}	2.54
3^{rd} score positive	0.1636^{*}	4.09	0.1821^{*}	0.86
2^{nd} score negative	-0.1775^{*}	2.29	$-0.1575^{\dagger}$	2.29
1^{st} score negative	-0.1030	3.72	$-0.1516^{\dagger}$	6.65
Total		96.77		97.48
FASPCA (Contiguity (3,2))				
1^{st} score positive	0.5041^{**}	67.15	0.6315^{***}	83.41
2^{nd} score positive	0.3857^{**}	14.37	$0.1831^{\dagger}$	1.87
3^{rd} score positive	0.3890^{**}	5.89	0.2537^{*}	1.78
2^{nd} score negative	-0.1546	1.64	-0.2023	1.60
1^{st} score negative	-0.3316^{*}	5.99	$-0.2544^{\dagger}$	8.61
Total		95.04		97.27

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Figure 2.7 displays the projection of the log death rates scores for females in year 2010 using the KNN neighbours of the three spatial PCs (top two positive and top one negative PC scores) onto the geographical map. The first positive PC (Figure 2.7(a)) shows spatial connectivity between the states split into two clusters, one in the west and one in the east. The projection for the second positive PC (Figure 2.7(b)) also shows spatial connectivity where two clusters are formed (northern and southern regions). The first negative PC (Figure 2.7(c)) does not seem to display a particular spatial pattern. Similar patterns are observed for cases involving the contiguity neighbours (Figure 2.8). This outcome is anticipated because the negative principal components are associated with local structures that highlight dissimilarities on the geographical map at neighbouring locations. In general, spatial patterns are noticeable for the first and second PCs involving male and female mortality data for years 1990, 2000 and 2010. These characteristics appear to be similar for the KNN and contiguity neighbours.

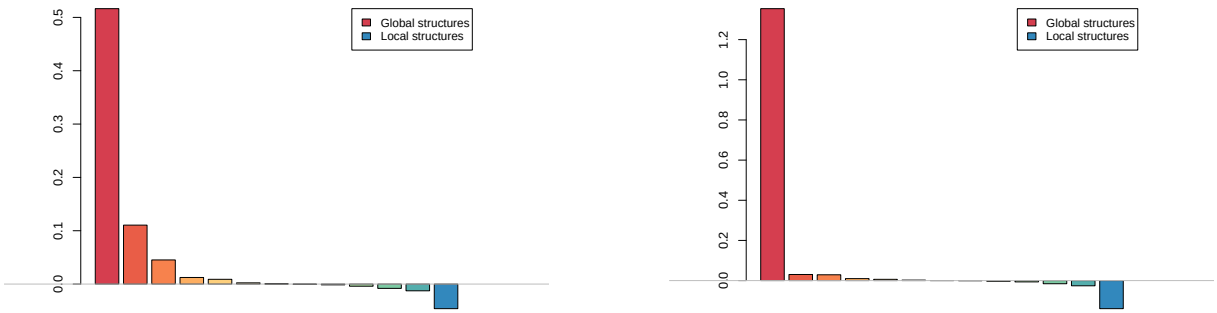
The scores of the first PC from the classical FPCA were displayed on the geographical map (Figure 2.9). This FPCA does not consider spatial autocorrelation and as mentioned earlier, the first PC alone accounts to more than 99.5% of the total variability (Table 2.2) indicating its failure to cluster the locations based on spatial dependency. Hence, it is vital to consider the spatial aspect when performing PCA to identify spatially dependent locations on a geographical map. Since this is done on functional data, the functional Moran's I statistic will be the best tool to efficiently assess the spatial autocorrelation in functional data. We found that all the scores

belonged to the same group. This outcome was expected since the FPCA does not consider spatial autocorrelation and as mentioned earlier the first PC alone accounts to more than 99.5% of the total variability (Table 2.2).



(a) (b)

Figure 2.5: Eigenvalues based on FASPCA using the KNN weight matrices for (a) females and (b) males from 28 European countries in 2010.



(a) (b)

Figure 2.6: Eigenvalues based on FASPCA using the contiguity weight matrices for (a) females and (b) males from 28 European countries in 2010.

2.5 Concluding remarks

We extended the Moran's I statistic in the context of functional data analysis and applied it to the mortality data of 28 countries in Europe. The objective step in preparing the data to be used in the FDA context is by data smoothing. After converting the data into functions, spatial functional PCA was employed to find new functions that reveal the most important type of variation in the curve. The data was analysed from classical and spatial perspectives. Interestingly, we found the existence of spatial dependency in the mortality rates of neighbouring countries via the KNN and contiguity neighbourhood approaches. Our newly introduced functional Moran's I statistic is proved to be efficient in identifying the existence of spatial dependency of mortality rates of neighbouring countries effectively using functional principal components, which are outcomes of dimensionality reduction. In our future work, this idea of spatial dependency

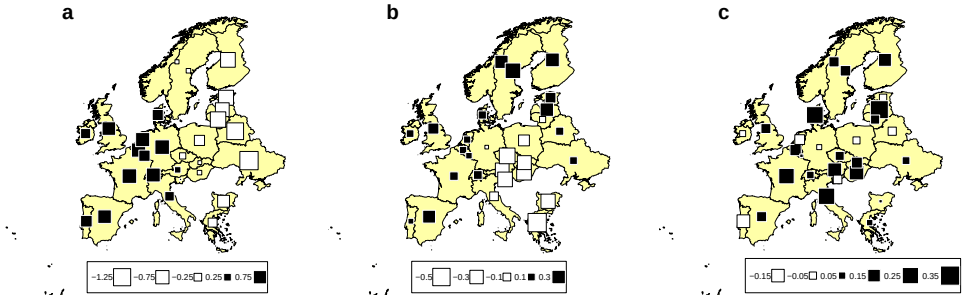


Figure 2.7: The scores of the (a) first positive, (b) second positive, and (c) first negative eigenvalues of the FASPCA based on KNN weights for females from 28 European countries in 2010.

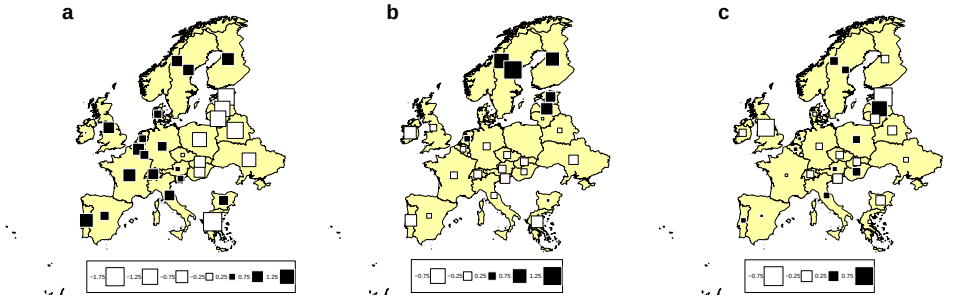


Figure 2.8: The scores of the (a) first positive, (b) second positive, and (c) first negative eigenvalues of the FASPCA based on contiguity weights for females from 28 European countries in 2010.

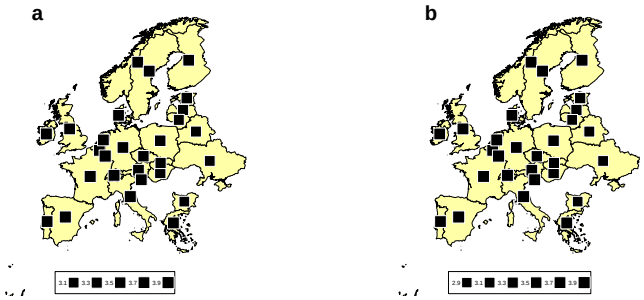


Figure 2.9: The scores of the first eigenvalues based on the classical FPCA for (a) females and (b) males from 28 European countries in 2010.

in mortality rates will be extended towards constructing a spatial predictive model to predict mortality rates for neighbouring countries with limited or no data. It is also of interest to further modify this work to the spatio-temporal framework.

2.6 Appendix

Table 2.3: Moran's test on principal components using FPCA and FASPCA, using KNN and contiguity weight matrices for females and males of 28 European countries for 1990.

	Female		Male	
	Moran's I	variability (%)	Moran's I	variability (%)
Classical FPCA				
1^{st} score	0.4135^{***}	99.79	0.5233^{***}	99.73
2^{nd} score	-0.1203	0.08	$-0.1342^{\dagger}$	0.11
3^{rd} score	-0.0019	0.05	0.3420^{***}	0.06
4^{th} score	0.4431^{***}	0.03	-0.0271	0.03
Total		99.95		99.93
FASPCA (KNN (3,2))				
1^{st} score positive	0.4905^{***}	65.46	0.5527^{***}	81.18
2^{nd} score positive	0.4211^{**}	11.77	0.3575^{***}	5.33
3^{rd} score positive	$0.1067^{\dagger}$	1.64	$0.0957^{\dagger}$	0.88
2^{nd} score negative	-0.1026	4.51	-0.2093^{**}	2.98
1^{st} score negative	-0.1324	10.66	-0.1655^{*}	7.78
Total		94.04		98.15
FASPCA (Contiguity (3,2))				
1^{st} score positive	0.5723^{***}	58.16	0.6376^{***}	80.64
2^{nd} score positive	0.5975^{***}	13.96	0.3917^{*}	6.38
3^{rd} score positive	$0.2018^{\dagger}$	6.52	$0.2248^{\dagger}$	2.90
2^{nd} score negative	-0.0948	2.26	-0.1105	2.32
1^{st} score negative	-0.3174^{*}	12.30	$-0.2637^{\dagger}$	3.71
Total		93.2		95.95

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2.4: Moran's test on principal components using FPCA and FASPCA, using KNN and contiguity weight matrices for females and males of 28 European countries for 2000.

	Female		Male	
	Moran's I	variability (%)	Moran's I	variability (%)
Classical FPCA				
1^{st} score	0.4710^{***}	99.75	0.5418^{***}	99.69
2^{nd} score	-0.1141	0.12	0.1414^{*}	0.15
3^{rd} score	0.0070	0.06	0.4498^{***}	0.07
4^{th} score	0.3234^{***}	0.02	-0.1972^{*}	0.03
Total		99.95		99.94
FASPCA (KNN (3,2))				
1^{st} score positive	0.4786^{***}	71.66	0.5645^{***}	83.74
2^{nd} score positive	0.4584^{***}	6.79	0.2613^{**}	9.24
3^{rd} score positive	$0.0968^{\dagger}$	1.69	$0.1128^{\dagger}$	0.66
2^{nd} score negative	-0.1553^{*}	3.94	-0.1790^{*}	1.78
1^{st} score negative	-0.2072^{*}	13.27	-0.1563^{*}	3.10
Total		97.35		98.52
FASPCA (Contiguity (3,2))				
1^{st} score positive	0.5182^{**}	72.7	0.6758^{***}	87.16
2^{nd} score positive	0.3055^{*}	7.53	0.1626	5.56
3^{rd} score positive	0.2524^{*}	3.78	0.2606^{*}	1.91
2^{nd} score negative	$-0.1558^{\dagger}$	4.32	-0.0977	1.11
1^{st} score negative	-0.1479	6.30	-0.3445^{*}	1.94
Total		94.63		97.68

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

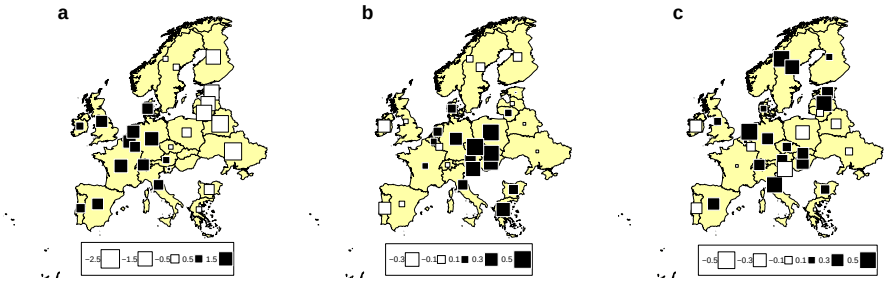


Figure 2.10: The scores of the (a) first positive, (b) second positive, and (c) first negative eigenvalues based on KNN weights for males from 28 European countries in 2010.

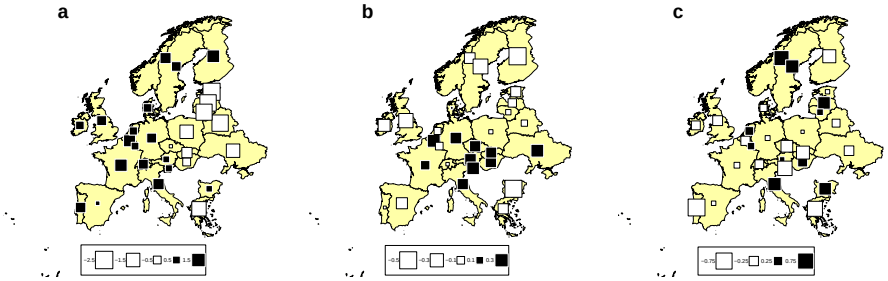


Figure 2.11: The scores of the (a) first positive, (b) second positive, and (c) first negative eigenvalues based on contiguity weights for males from 28 European countries in 2010.

Chapter 3

Partially linear spatial probit models

Abstract

A partially linear probit model for spatially dependent data is considered. A triangular array setting is used to cover various patterns of spatial data. Conditional spatial heteroscedasticity and non-identically distributed observations and a linear process for disturbances are assumed, allowing various spatial dependencies. The estimation procedure is a combination of a weighted likelihood and a generalized method of moments. The procedure first fixes the parametric components of the model and then estimates the non-parametric part using weighted likelihood; the obtained estimate is then used to construct a GMM (Generalised Method of Moments) parametric component estimate. The consistency and asymptotic distribution of the estimators are established under sufficient conditions. Some numerical results are provided to investigate the finite sample performance of the estimators.

Keywords: Binary choice model, GMM, non-parametric statistics, spatial econometrics, spatial statistics.

Introduction

Agriculture, economics, environmental sciences, urban systems, and epidemiology activities often utilize spatially dependent data. Therefore, modelling such activities requires one to find a type of correlation between some random variables in one location with other variables in neighbouring locations; see for instance Pinkse and Slade (1998). This is a significant feature of spatial data analysis. Spatial/Econometrics statistics provides tools to perform such modelling. Many studies on spatial effects in statistics and econometrics using many diverse models have been published; see Cressie (2015), Anselin (2010), Anselin (2013) and Arbia (2006) for a review.

Two main methods of incorporating a spatially dependent structure (see for instance Cressie, 2015) can essentially be distinguished as between 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. In image analysis, remote sensing from satellites, agriculture etc., data are often received as a regular lattice and identified as the centroids of square pixels, whereas a mapping often forms an irregular lattice. Basically, statistical models for lattice data

are linked to nearest neighbours to express the fact that data are nearby.
Two popular spatial dependence models have received substantial attention for lattice data, the spatial autoregressive (SAR) dependent variable model and the spatial autoregressive error model (SAE, where the model error is an SAR), which extend the regression in a time series setting to spatial one.
From a theoretical point of view, various linear spatial regression SAR and SAE models as well as their identification and estimation methods, e.g., two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood (ML) or quasi-maximum likelihood (QML) and the generalized method of moments (GMM), have been developed and summarized by many authors such as Anselin (2013), Kelejian and Prucha (1998), Kelejian and Prucha (1999), Conley (1999), Cressie (2015), Case (1993), L.-F. Lee (2004), L.-f. Lee (2007), Lin and Lee (2010), Zheng and Zhu (2012), Malikov and Sun (2017), Garthoff and Otto (2017), Yang and Lee (2017).

Introducing nonlinearity into the field of spatial linear lattice models has attracted less attention; see Robinson (2011), who generalised kernel regression estimation to spatial lattice data. Su (2012) proposed a semi-parametric GMM estimation for some semi-parametric SAR models. Extending these models and methods to discrete choice spatial models has seen less attention where only a few researches are concerned with this topic in recent years. This may be as noted 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 spatially discrete choice models often requires solving a very computationally demanding problem of n-dimensional integration, where n is the sample size.
For linear models, many discrete choice models are fully linear and utilize a continuous latent variable; see Smirnov (2010), H. Wang et al. (2013) and Martinetti and Geniaux (2017), who proposed pseudo-ML methods, and Pinkse and Slade (1998) who studied a method based on the GMM approach. Also, others methodologies of estimation are used such as the EM algorithm (McMillen, 1992) and Gibbs sampling approach (LeSage, 2000).

When the relationship between the discrete choice variable and some explanatory variables is not linear, a semi-parametric model may be an alternative to fully parametric models. This type of model is known in the 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 special cases of the famous generalised additive models (Hastie & Tibshirani, 1990) and have received substantial attention in the literature, and various estimation methods have been explored (see for instance Carroll et al., 1997; Hunsberger, 1994; Severini & Staniswalis, 1994).
To the best of our knowledge, semi-parametric spatial choice models have not yet been investigated from a theoretical point of view. To fill this gap, this work addresses an SAE spatial probit model for when the spatial dependence structure is integrated in a disturbance term of the studied model.
We propose a semi-parametric estimation method combining the GMM approach and the weighted likelihood method. The method consists of first fixing the parametric components of the model and non-parametrically estimating the non-linear component by weighted likelihood (Staniswalis, 1989). The obtained estimator depending on the values at which the parametric components are fixed is used to construct a GMM estimator (Pinkse & Slade, 1998) of these components.
The remainder of this paper is organised as follows. In Section 3.1, we introduce the studied spatial model and the estimation procedure. Section 3.2 is devoted to hypotheses and asymptotic results, while Section 3.3 reports a discussion and computation of the estimates. Section 3.4 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.

3.1 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^*:

$$Y_{in}^* = X_{in}^T \beta_0 + g_0(Z_{in}) + U_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{3.1}$$

with

$$Y_{in} = \mathbb{I}\left(Y_{in}^* \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{3.2}$$

where $\mathbb{I}(\cdot)$ is the indicator function; X and Z are explanatory random variables taking values in the two compact 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 $\Theta_\beta \subset \mathbb{R}^p$; and $g_0(\cdot)$ 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 $\mathbb{R}$ and C a positive constant. In model (3.1), β_0 and $g_0(\cdot)$ are constant over i (and n). Assume that the disturbance term U_{in} in (3.2) is modelled by the following spatial autoregressive process (SAR):

$$U_{in} = \lambda_0 \sum_{j=1}^{n} w_{ijn} U_{jn} + \varepsilon_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{3.3}$$

where, we assume that, for all $n = 1, 2, \ldots$, $\{\varepsilon_{in},\ 1 \leq i \leq n\}$ is independent of $\{X_{in},\ 1 \leq i \leq n\}$ and $\{Z_{in},\ 1 \leq i \leq n\}$, and $\{X_{in},\ 1 \leq i \leq n\}$ is independent of $\{Z_{in},\ 1 \leq i \leq n\}$.
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 weight matrix W_n, which 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 and Slade (1998) for additional details. The $n \times n$ matrix $(I_n - \lambda_0 W_n)$ is assumed to be non-singular 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; $\mathbb{E}(\varepsilon_{in}) = 0$ and $\mathbb{E}(\varepsilon_{in}^2) = 1$ for $i = 1, \ldots, n$ $n = 1, 2, \ldots$. Note that one can rewrite (3.3) as

$$U_n = (I_n - \lambda_0 W_n)^{-1} \varepsilon_n, \quad n = 1, 2, \ldots \tag{3.4}$$

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

$$V_n(\lambda_0) \equiv \mathrm{Var}(U_n) = (I_n - \lambda_0 W_n)^{-1} \left\{(I_n - \lambda_0 W_n)^T\right\}^{-1}, \quad n = 1, 2, \ldots \tag{3.5}$$

This matrix allows one 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 dependencies and heteroscedasticity depend on the neighbourhood structure established by the spatial weight ma-

trix W_n.

The elements w_{ijn} of W_n are usually considered as inversely proportional to the distance between spatial units i and j with respect to some metric Pinkse and Slade (physical distance, social network or economic distance, see for instance 1998). The matrices W_n are usually classified into two groups: *Weights Based on Distance* and *Weights Based on Boundaries*. For Weights Based on Distance, the distance d_{ij} between each pair of spatial units (regions, cities, centroids,...) i and j are basically considered.

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

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

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).

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

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.

Before proceeding further, let us give some particular cases of the model.

If one considers 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 viewed as a special case of classical generalised partially linear models (e.g. Severini & Staniswalis, 1994) or the classical generalised additive model (Hastie & Tibshirani, 1990). Several approaches for estimating this particular model have been developed; among these methods, we cite that of Severini and Staniswalis (1994) based on the concept of the generalised profile likelihood Severini and Wong (e.g 1992). This approach consists of first fixing the parametric parameter β and non-parametrically estimating $g_0(\cdot)$ using the weighted likelihood method. This last estimate is then used to construct a profile likelihood to estimate β_0.

When $g_0 \equiv 0$ (or is an affine function), that is, without a non-parametric 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 involves an n-dimensional normal integral; thus, when n is high, the computation or asymptotic properties of the estimates may present difficulties (e.g. Poirier & Ruud, 1988). Various approaches have been proposed to addressed this difficulty; among these approaches, we cite the following:

- Feasible Maximum Likelihood approach: this approach consists of replacing the true likelihood function by a pseudo-likelihood function constructed via marginal likelihood functions. Smirnov (2010) proposed a pseudo-likelihood function obtained by replacing $V_n(\lambda_0)$

by some diagonal matrix with the diagonal elements of $V_n(\lambda_0)$. Alternatively, H. Wang et al. (2013) proposed to divide the observations by pairwise groups, where the latter are assumed to be independent with a bivariate normal distribution in each group, and estimate β_0 and λ_0 by maximizing the likelihood of these groups. Recently Martinetti and Geniaux (2017) proposed a pseudo-likelihood function defined as an approximation of the likelihood function where the latter is inspired by some univariate conditioning procedure.

- Generalised Method of Moments (GMM) approach used by Pinkse and Slade (1998). These authors used the generalized residuals defined by $\tilde{U}_{in}(\beta,\lambda) = \mathbb{E}\left(U_{in}|Y_{in},\beta,\lambda\right),\ 1 \leq i \leq n,\ n = 1,2,\ldots$ with some instrumental variables to construct moment equations to define the 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 non-parametric estimator of the smooth function $g_0(\cdot)$.
We give asymptotic results according to *increasing domain* asymptotic. This consists of a sampling structure whereby new observations are added at the edges (boundary points) compare to the *infill* asymptotic, which consists of a sampling structure whereby new observations are added in-between existing observations. A typical example of an increasing domain is lattice data. An infill asymptotic is appropriate when the spatial locations are in a bounded domain.

3.1.1 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 non-parametric component using a weighted likelihood. The obtained estimate is then used to construct generalised residuals, where the latter are combined with the instrumental variables to propose GMM parametric estimates. This approach will be described as follows:

By equation (3.2), we have

$$\mathbb{E}_0\left(Y_{in}|X_{in},Z_{in}\right) = \Phi\left(\left(v_{in}(\lambda_0)\right)^{-1}\left(X_{in}^T\beta_0 + g_0(Z_{in})\right)\right), \quad 1 \leq i \leq n, \quad n = 1,2,\ldots \tag{3.6}$$

where $\mathbb{E}_0$ 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 $(v_{in}(\lambda_0))^2 = V_{iin}(\lambda_0),\ 1 \leq i \leq n,\ n = 1,2,\cdots$ 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} for $1 \leq i \leq n,\ n = 1,2,\ldots$, as

$$H(\eta;\beta,\lambda,z) = \mathbb{E}_0\left(\mathcal{L}\left(\Phi\left(\left(v_{in}(\lambda)\right)^{-1}\left(\eta + X_{in}^T\beta\right)\right);Y_{in}\right)\middle|\, Z_{in} = z\right), \tag{3.7}$$

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

$$\frac{\partial}{\partial\eta}H(\eta;\beta,\lambda,z) = 0. \tag{3.8}$$

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 the GMM estimates of β_0 and λ_0 as in Pinkse and Slade, 1998. For that, we define the generalised residuals, replacing $g_0(Z_{in})$ in (3.1) by $g_{\beta,\lambda}(Z_{in})$:

$$\begin{aligned}\tilde{U}_{in}(\beta,\lambda,g_{\beta,\lambda}) &= \mathbb{E}\left(U_{in}|Y_{in},\beta,\lambda\right) \\ &= \frac{\phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\left(Y_{in}-\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\right)}{\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\left(1-\Phi\left(G_{in}(\beta,\lambda,g_{\beta,\lambda})\right)\right)},\end{aligned} \tag{3.9}$$

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 notational simplicity, we write $\theta = (\beta^T,\lambda)^T \in \Theta \equiv \Theta_\beta \times \Theta_\lambda$ when possible.
Note that in (3.9), the generalised residual $\tilde{U}_{in}(\cdot\,,\cdot)$ is calculated by conditioning only on Y_{in} and 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 generalised residuals, but it allows one to avoid a complex computation; see Poirier and Ruud (1988) for additional details. To address this loss of efficiency, let us follow Pinkse and Slade (1998)'s procedure, which consists of employing some instrumental variables to create some moment conditions, and use a random matrix to define a criterion function. Both the instrumental variables and the random matrix permit one to consider more information about the spatial dependences and heteroscedasticity characterizing the dataset. Let us now discuss the details of the estimation procedure. Let

$$S_n(\theta,g_\theta) = n^{-1}\xi_n^T\tilde{U}_n(\theta,g_\theta), \tag{3.10}$$

where $\tilde{U}_n(\theta,g_\theta)$ is an $n\times 1$ vector, composed of $\tilde{U}_{in}(\theta,g_\theta)$, $1\le i\le n$ and ξ_n is an $n\times q$ matrix of instrumental variables, whose ith row is given 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 instrumental variables on the parameters for notational simplicity. The GMM approach consists of minimising the following sample criterion function:

$$Q_n(\theta,g_\theta) = S_n^T(\theta,g_\theta)M_nS_n(\theta,g_\theta), \tag{3.11}$$

where M_n is some positive-definite $q\times q$ weight matrix that may depend on the sample information. The choice of the instrumental variables and weight matrix characterizes the difference between GMM estimator and all pseudo-maximum likelihood estimators. For instance, if one takes

$$\xi_{in}(\theta,g_\theta) = \frac{\partial G_{in}(\theta,\eta_i)}{\partial\theta} + \frac{\partial G_{in}(\theta,\eta_i)}{\partial\eta}\frac{\partial g_\theta}{\partial\theta}(Z_{in}), \tag{3.12}$$

with $\eta_i = g_\theta(Z_{in})$, $G_{in}(\theta,\eta_i) = (v_{in}(\lambda))^{-1}\left(X_{in}^T\beta+\eta_i\right)$, and $M_n = I_q$ with $q = p+1$, then the GMM estimator of θ is equal to a pseudo-maximum profile likelihood estimator of θ, accounting only for the spatial heteroscedasticity.
Now, let

$$S(\theta,g_\theta) = \lim_{n\to\infty}\mathbb{E}_0\left(S_n(\theta,g_\theta)\right), \tag{3.13}$$

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.

Clearly, $g_\theta(\cdot)$ is not available in practice. However, we need to estimate it, particularly by an asymptotically efficient estimate. By (3.8) 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

$$\sum_{i=1}^{n} \frac{\partial}{\partial \eta} \mathcal{L}\left(\Phi\left(G_{in}(\theta, \eta)\right); Y_{in}\right) K\left(\frac{z - Z_{in}}{b_n}\right) = 0, \tag{3.14}$$

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 (3.11) by the estimator $\hat{g}_\theta(\cdot)$ permits one to obtain the GMM estimator $\hat{\theta}$ of θ as

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta} Q_n(\theta, \hat{g}_\theta). \tag{3.15}$$

A classical inconvenience of the estimator $\hat{g}_\theta(z)$ proposed in (3.14) is that the bias of $\hat{g}_\theta(z)$ is high for z near the boundary of $\mathcal{Z}$. Of course, this bias will affect the estimator of θ given in (3.15) when some of the observations Z_{in} are near the boundary of $\mathcal{Z}$. A 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 using only observations associated with 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 small sample sizes.

3.2 Large sample properties

We now turn to the asymptotic properties of the estimators derived in the previous section: $\hat{\theta}^T = (\hat{\beta}^T, \hat{\lambda})$ and $\hat{g}_{\hat{\theta}}(\cdot)$. Let us use the following notation: $\frac{d}{d\theta} S(\theta, g_\theta)$ means that we differentiate $S(.,.)$ with respect to θ, and $\frac{\partial}{\partial \theta} S(\theta, g_\theta)$ is the partial derivative of $S(\cdot, \cdot)$ w.r.t the first variable. The partial derivative of $S_n(\theta, g)$ w.r.t g, for any function $v \in \mathcal{G}$, is

$$\frac{\partial S_n}{\partial g}(\theta, g)(v) = n^{-1} \sum_{i=1}^{n} \xi_{in} \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta_i) v(Z_{in}).$$

Without ambiguity, $\|a\|$ denotes $\sup_t |a(t)|$ when a is a function, $\left(\sum a_i^2\right)^{1/2}$ when a is a vector, and $\left(\sum\sum a_{ij}^2\right)^{1/2}$ when a is a matrix.
Let the following matrices be needed in the asymptotic variance-covariance matrix of $\hat{\theta}$:

$$B_1(\theta_0) = \lim_{n \to \infty} \mathbb{E}_0\left(n S_n\left(\theta_0, g_0\right) S_n^T\left(\theta_0, g_0\right)\right), \; B_2(\theta_0) = \left\{ \frac{d}{d\theta} S^T\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0} \right\} M \left\{ \frac{d}{d\theta} S\left(\theta, g_\theta\right)\Big|_{\theta=\theta_0} \right\},$$

with

$$\frac{d}{d\theta} S\left(\theta, g_\theta\right) = \frac{\partial S}{\partial \theta}\left(\theta, g_\theta\right) + \frac{\partial S}{\partial g}\left(\theta, g_\theta\right) \frac{\partial}{\partial \theta} g_\theta, \tag{3.16}$$

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

$$\sup_{\theta \in \Theta, z \in \mathcal{Z}} \left| \frac{\partial}{\partial \eta} H(g(z); \theta, z) \right| \leq \gamma \qquad \Longrightarrow \qquad \sup_{\theta \in \Theta, z \in \mathcal{Z}} |g(z) - g_\theta(z)| \leq \varepsilon. \tag{3.17}$$

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

$$\liminf_{n \to \infty} \inf_{z \in \mathcal{Z}} \frac{1}{n} \sum_{i=1}^{n} f_{in}(z) > 0. \tag{3.18}$$

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, $l = 1, \ldots, n$,

$$|\mathrm{Cov}_0 (\psi(T_{in}), \psi(T_{ln}))| \leq \{\mathrm{Var}_0 (\psi(T_{in})) \, \mathrm{Var}_0 (\psi(T_{ln}))\}^{1/2} \, \alpha_{iln}, \tag{3.19}$$

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

$$\left|\mathrm{Cov}_0 \left(\xi_{itn} \tilde{U}_{in}(\theta, g_\theta), \xi_{jsn} \tilde{U}_{jn}(\theta, g_\theta)\right)\right| \leq \left\{\mathrm{Var}_0 \left(\xi_{itn} \tilde{U}_{in}(\theta, g_\theta)\right) \mathrm{Var}_0 \left(\xi_{jsn} \tilde{U}_{jn}(\theta, g_\theta)\right)\right\}^{1/2} \alpha_{ijn}, \tag{3.20}$$

for all $\theta \in \Theta$, $i, j = 1, \ldots, n$, $n = 1, 2, \ldots$ and for any $s, t = 1, \ldots, q$,

and

$$\left|\mathrm{Cov}_0 \left(\xi_{in}^{(2)}(\theta_0, \eta_i^0), \xi_{jn}^{(2)}(\theta_0, \eta_j^0)\right)\right| \leq \left\{\mathrm{Var}_0 \left(\xi_{in}^{(2)}(\theta_0, \eta_i^0)\right) \mathrm{Var}_0 \left(\xi_{jn}^{(2)}(\theta_0, \eta_j^0)\right)\right\}^{1/2} \alpha_{ijn}, \tag{3.21}$$

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 $\|w\| = 1$.

In addition, assume that there is a decreasing (to 0) positive function $\varphi(\cdot)$ such that the "mixing" numbers verify $\alpha_{ijn} = O\left(\varphi\left(\|s_i - s_j\|\right)\right)$, $r^2 \varphi(r r^*)/\varphi(r^*) = o(1)$, as $r \to 0$, for all fixed $r^* > 0$, where s_i and s_j are spatial coordinates associated with observations i and j, respectively.

Assumption A4. The kernel K satisfies $\int K(u) du = 1$. It is Lipschitzian, 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 instrumental 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 instrumental 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 the diagonal 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 1. *Assumption A1 ensures the smoothness of $H(.; ., .)$ around its extrema point $g_\theta(.)$; see Severini and Staniswalis (1994). Assumption A2 is a decay of the 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 (3.18) 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, it is a particular case of the Assumption A in Pinkse et al. (2007) and may be verified by mixing random variables, see Pinkse et al. (2007) for more details. Note that the processes that we use are not assumed stationary; this allows for greater generalizability and the dependence structure to change with the sample size n (see Pinkse and Slade (1998) for more discussion). Conditions (3.19), (3.20) and (3.21) are not restrictive. When the regressors and instrumental variables are deterministic, conditions (3.19) and (3.20) are equivalent to $|Cov_0(Y_{in}, Y_{ln})| \leq \alpha_{iln}$. The condition on $\varphi(\cdot)$ is satisfied when the latter tends to zero at a polynomial rate, i.e., $\varphi(t) = O(t^{-\tau})$, for all $\tau > 2$, as in the case of mixing random variables.*

Assumption A6 requires that the instruments and explanatory variables 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; however, it is reasonable in most applications. Condition A8 is necessary to ensure the 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. This has been considered by Pinkse and Slade (1998). Assumption A10 is classic (Pinkse and Slade (1998)) and required in the proof of Theorem 3.2.2. Those authors noted that in their model (without a non-parametric component), when the autoregressive parameter $\lambda_0 = 0$, $B_2(\theta_0)$ is not invertible, regardless of the choice of M_n. This is also the case in our context because for each $g_\theta(z)$ solution of (3.8), $\theta \in \Theta$ 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

$$\frac{\partial g_\theta}{\partial \lambda}(z) = \frac{v_{jn}^{'}(\lambda)}{v_{jn}(\lambda)} \frac{E\left(\Gamma_{jn}(\theta, g_\theta(z))\left(X_{jn}^T\beta + g_\theta(z)\right)\middle| Z_{jn} = z\right)}{E\left(\Gamma_{jn}(\theta, g_\theta(z))\middle| Z_{jn} = z\right)}$$
$$= \frac{v_{jn}^{'}(\lambda)}{v_{jn}(\lambda)}\left(g_\theta(z) - \beta^T \frac{\partial g_\theta}{\partial \beta}(z)\right),$$

where $v_{jn}^{'}(\lambda) = \frac{d}{d\lambda} v_{jn}(\lambda) = v_{jn}(\lambda)\left[W_n S_n^{-1}(\lambda) V_n(\lambda)\right]_{jj}$,

$$\Gamma_{jn}(\cdot) = \Lambda^{'}(G_{jn}(\cdot))\left[Y_{jn} - \Phi(G_{jn}(\cdot))\right] - \Lambda\left(G_{jn}(\cdot)\right)\phi\left(G_{jn}(\cdot)\right)$$

and $\Lambda(\cdot) = \phi(\cdot)/(1 - \Phi(\cdot))\Phi(\cdot)$. *However*

$$\frac{\partial g_\theta}{\partial \lambda}(z)\bigg|_{\lambda=0} = 0 \qquad \textit{because} \qquad v_{jn}^{'}(0) = 0,$$

then $B_2(\theta_0)$ *will be singular when* $\lambda_0 = 0$.

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 addresses the question of convergence to a normal distribution of the parametric component when it is properly standardised.

Theorem 3.2.1. *Under Assumptions A1-A10, we have*

$$\hat{\theta} - \theta_0 = o_p(1).$$

Corollary 3.2.1. *If the assumptions of Theorem 3.2.1 are satisfied, then we have*

$$\left\|\hat{g}_{\hat{\theta}} - g_0\right\| = o_p(1).$$

Proof of Corollary 3.2.1 Note that

$$\left\|\hat{g}_{\hat{\theta}} - g_0\right\| \leq \left\|\hat{g}_{\hat{\theta}} - g_{\hat{\theta}}\right\| + \left\|g_{\hat{\theta}} - g_0\right\|$$
$$\leq \sup_\theta \left\|\hat{g}_\theta - g_\theta\right\| + \sup_\theta \left\|\frac{\partial g_\theta}{\partial \theta}\right\| \left\|\hat{\theta} - \theta_0\right\| = o_p(1),$$

since, by the assumptions of Theorem 3.2.1, $\sup_\theta \left\|\hat{g}_\theta - g_\theta\right\| = o_p(1)$ and $\sup_\theta \left\|\frac{\partial g_\theta}{\partial \theta}\right\| < \infty$. The following gives an asymptotic normality result of $\hat{\theta}$.

Theorem 3.2.2. *Under assumptions A1-A10, we have*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}\left(0, \Omega(\theta_0)\right)$$

Remark 2. *In practice, the previous asymptotic normality result can be used to construct asymptotic confidence intervals and build hypothesis tests when a consistent estimate of the asymptotic covariance matrix* $\Omega(\theta_0)$ *is available. To estimate this matrix, let us follow the idea of Pinkse*

and Slade (1998) and define the estimator

$$\Omega_n(\hat{\theta}) = \left\{B_{2n}(\hat{\theta})\right\}^{-1} \left\{ \frac{d}{d\theta} S_n^T(\theta, \hat{g}_\theta)\Big|_{\theta=\hat{\theta}} \right\} M_n B_{1n}(\hat{\theta}) M_n \left\{ \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta)\Big|_{\theta=\hat{\theta}} \right\} \left\{B_{2n}(\hat{\theta})\right\}^{-1},$$

with

$$B_{1n}(\theta) = nS_n(\theta, \hat{g}_\theta) S_n^T(\theta, \hat{g}_\theta) \qquad \text{and} \qquad B_{2n}(\theta) = \left\{ \frac{d}{d\theta} S_n^T(\theta, \hat{g}_\theta) \right\} M_n \left\{ \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta) \right\}.$$

The consistency of $\Omega_n(\hat{\theta})$ will be based on that of $B_{1n}(\hat{\theta})$ and $B_{2n}(\hat{\theta})$, the estimators of $B_1(\theta_0)$ and $B_2(\theta_0)$, respectively. Note that the consistency of $B_{2n}(\hat{\theta})$ is relatively easy to establish. On the other hand, that of $B_{1n}(\hat{\theta})$ asks for additional assumptions and an adaption of the proof of Theorem 3 of Pinkse and Slade (1998, p.134) to our case; this is of interest to future research.

3.3 Computation of the estimates

The aim of this section is to outline in detail 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 $\hat{g}_\theta(z)$, which will play a crucial role in what follows.

3.3.1 Computation of the estimate of the non-parametric component

An iterative method is needed to compute the $\hat{g}_\theta(z)$ solution of (3.14) 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-hand side of (3.14), which can be rewritten as

$$\psi(\eta; \theta, z) = \sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-1} \Lambda\left(G_{in}(\theta, \eta)\right) \left[Y_{in} - \Phi(G_{in}(\theta, \eta))\right] K\left(\frac{z - Z_{in}}{b_n}\right). \tag{3.22}$$

Consider the Fisher information:

$$\begin{aligned} \Psi(\eta_\theta; \theta, z) &= E_0\left(\frac{\partial}{\partial \eta} \psi(\eta; \theta, z)\Big|_{\eta=\eta_\theta} \Big| \{(X_{in}, Z_{in}),\ 1 \leq\ i \leq n,\ n = 1, \ldots\} \right) \\ &= -\sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-2} \Lambda\left(G_{in}(\theta, \eta_\theta)\right) \phi\left(G_{in}(\theta, \eta_\theta)\right) K\left(\frac{z - Z_{in}}{b_n}\right) + \\ &\quad \sum_{i=1}^{n} \left[v_{in}(\lambda)\right]^{-2} \Lambda'\left(G_{in}(\theta, \eta_\theta)\right) \left[\Phi\left(G_{in}(\theta_0, \eta_0)\right) - \Phi\left(G_{in}(\theta, \eta_\theta)\right)\right] K\left(\frac{z - Z_{in}}{b_n}\right) \end{aligned} \tag{3.23}$$

Note that the second term in the RHS (Right Hand Side) of (3.24) is negligible when θ is near the true parameter θ_0.

Because $\psi(\eta; \theta, z) = 0$ for $\eta = \hat{g}_\theta(z)$, an initial estimate $\tilde{\eta}$ can be updated to $\eta^\dagger$ using Fisher's scoring method:

$$\eta^\dagger = \tilde{\eta} - \frac{\psi(\tilde{\eta}; \theta, z)}{\Psi(\tilde{\eta}; \theta, z)}. \tag{3.24}$$

The iteration procedure (3.24) requests some starting value $\tilde{\eta} = \tilde{\eta}_0$ to ensure convergence of the algorithm. To this end, let us adapt the approach of Severini and Staniswalis (1994), which

consists of supposing 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, (3.24) can be updated using the following initial value:

$$\eta_0^{\dagger} = \tilde{\eta}_0 - \frac{\psi(\tilde{\eta}_0; \theta, z)}{\Psi(\tilde{\eta}_0; \theta, z)} = \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-1} \Lambda(C_{in})\phi(C_{in}) \left[C_{in} - [v_{in}(\lambda)]^{-1} X_{in}^T\beta\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Lambda(C_{in})\phi(C_{in}) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

where $C_{in} = \Phi^{-1}(Y_{in})$, $i = 1, \ldots, n$, is computed using a slight adjustment because $Y_{in} \in \{0, 1\}$. With this initial value, the algorithm iterates until convergence.

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 semi-parametric model. Because cross-validation may be very time consuming, which is true in the case of our model, we adapt the following approach used in Severini and Staniswalis (1994) to achieve greater 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 $\mathbb{E}(R_{in}|Z_{in} = z)$ to have similar smoothness properties as $g_0(.)$, the optimal bandwidth b_n is that of the non-parametric regression of the $\{R_{in}\}_{i=1,\cdots,n}$ on $\{Z_{in}\}_{i=1,\cdots,n}$, chosen by applying any non-parametric regression bandwidth selection method. For that, we use the cross-validation method in the *np* R Package.

3.3.2 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 instrumental variables ξ_n and the weight matrix M_n. The choices of these instrumental variables and weight matrix M_n are as follows.
Because $\psi(\hat{g}_\theta(z); \theta, z) = 0$, if we differentiate the latter with respect to β and λ, we have

$$\frac{\partial}{\partial\beta}\hat{g}_\theta(z) = -\frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) X_{in} K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

and

$$\begin{aligned}\frac{\partial}{\partial\lambda}\hat{g}_\theta(z) =& \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-1} v_{in}'(\lambda)\Delta_{in}(\theta, z)\left[X_{in}^T\beta + \hat{g}_\theta(z)\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)} \\ &+ \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} v_{in}'(\lambda)\Lambda\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\left[Y_{in} - \Phi\left(G_{in}(\theta, \hat{g}_\theta(z))\right)\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)},\end{aligned}$$

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 instrumental 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, one can use $M_n = I_q$ with $q = p + 1$ as in Pinkse and Slade (1998). 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 for the spatial heteroscedasticity. Another empirical choice could be the idea of continuous updating GMM estimator (One step GMM) used in Pinkse et al. (2006):

$$M_n(\theta) = \left\{ n^{-1} \sum_{i,j=1}^{n} \delta_{ij} \xi_{ni} \xi_{jn}^T \tilde{U}_{in}(\theta, \hat{g}_\theta) \tilde{U}_{jn}(\theta, \hat{g}_\theta) \right\}^{-1} \tag{3.25}$$

with the 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_{nij} such that the nearer location i is to location j, the larger τ_{ij} is. For instance, we expect to have more efficient estimators with this matrix.

3.4 Finite sample properties

In this section, we study the performance of the proposed model based on some numerical results, which highlight the importance of accounting for both the spatial dependence and the partial linearity. Random datasets from the following spatial semi-parametric models are generated and first we investigate the estimation quality of the proposed procedure which accounts both the spatial dependence and the partial linearity. The influences of the spatial dependence and the partial linearity are investigated by comparing the behavior of our model to that of the non-spatial partially linear probit (NSPLP) model and the fully linear SAE probit (LSAEP) model, respectively. The *GAM* and *ProbitSpatial* (Martinetti & Geniaux, 2016) R packages will be used to provide the estimates associated to NSPLP and LSAEP models respectively. We generate observations from the following spatial latent partial linear model:

$$\begin{aligned} Y_{in}^* &= \beta_1 X_{in}^{(1)} + \beta_2 X_{in}^{(2)} + g(Z_{in}) + U_{in}; \qquad Y_{in} = \mathbb{I}(Y_{in}^* > 0),\ i = 1, \ldots, n \\ U_n &= (I_n - \lambda W_n)^{-1} \varepsilon_n \end{aligned}$$

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 and with elements constructed in such way that each location has at least 6 neighbors. 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 $[-0.25, 0.25]$. Here, we use the non-linear function $g(t) = t + 2\cos(0.5\pi t)$ and parameters $\beta_1 = -1$, $\beta_2 = 1$.

Different spatial dependence parameters λ; 0.2 (weak spatial dependence) 0.5 and 0.8 (strong spatial dependence) are considered. Finally, the sample size effect is observed by considering n equals to 200, 400 and 800 with 300 replications of each simulation.

Our estimation procedure is applied with a Gaussian kernel $K(t) = (2\pi^{-1/2})\exp(-t^2/2)$ and optimal bandwidth b_n selected by Severini and Staniswalis (1994)'s approach detailed previously. We consider the trivial instrumental variables and two choices of matrix $M_n = I_n$ which leads to the pseudo-maximum profile likelihood estimators (named PLSP 1) and a second choice M_n given in (3.25) with components $\tau_{ij} = w_{nij}$, the estimates obtained with this matrix choice are denoted PLSP 2. The second choice of the weight matrix allows to incorporate more information about the spatial dependance.
The results are given in Table 3.1, the columns titles Mean, Median and SD give the average, median and standard deviation, respectively, over these 300 replications associated with each estimation method.

In one hand, when we compare the estimators (PLSP 1 and PLSP2) based on our approach (PLSPM) with those based on the LSAEP model, we notice that the latter yields more biased estimators of the coefficients β_1 and β_2. It makes sense that ignoring the partial linearity (see also Figure 3.1) weakens the quality of the estimation of the coefficients β_1 and β_2.

On the other hand, note that the LSAEP and PLSP 1 estimates are similar in case of low spatial dependence ($\lambda = 0.2$) compare to large spatial dependence ($\lambda = 0.8$) framework. It makes sense that ignoring a high spatial dependence does not allow a model that does not account any spatial structure to find consistent estimates of the coefficients β_1 and β_2 and the smooth function $g(\cdot)$ (see Figure 3.1) .
Note that the second choice of the weight matrix (estimates PLSP 2) allowed to improve the efficiency of the proposed estimates particularly in case of high spatial dependence (see PLSP 2 estimates in case of $\lambda = 0.8$). In contrast, it is less appropriate in case of low spatial dependance. However, one may think of testing the intensity of the spatial dependence before applying the proposed model with a non identity weight matrix, using for instance Moran's test.

Discussion

In this manuscript, we have proposed a spatial semi-parametric probit model for identifying risk factors at onset and with spatial heterogeneity. The parameters involved in the models are estimated using weighted likelihood and generalised method of moment methods. A technique based on dependent random arrays facilitates the estimation and derivation of asymptotic properties, which otherwise would have been difficult to perform due to the complexity introduced by the spatial dependence to the model and high-dimensional integration required by a full maximum likelihood approach. Moreover, the technique yields consistent estimates through proper choices of the bandwidth, weight matrix, and instrumental variables. The proposed models provide a general framework and tools for researchers and practitioners when addressing binary semi-parametric choice models in the presence of spatial correlation. Although they provide significant contributions to the body of knowledge, additional investigations need to be done.
As indicated previously, weights are used to improve the efficiency and convergence of the GMM procedure. For instance, the finite sample properties section shown that the kind of weight matrix defined in 3.25 with elements τ_{ij} may improve the efficiency of the proposed estimator but is less appropriate in case of weak spatial dependence. Then, it would be interesting to develop other choices of weights τ_{ij} toward achieving a better performance. Another topic of future research is to allow some spatial dependency in the covariates (SAR models) and the response (endogenous models) for more generality.

Table 3.1: The mean, median and standard deviation (SD) of the parameters β_1, β_2, and λ estimates, over the 300 replications

λ	n	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
			Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.20	200	PLSP 1	-1.06	-1.00	0.40	1.05	0.98	0.26	0.12	0.00	0.40
		PLSP 2	-1.06	-1.07	0.28	1.06	1.04	0.19	0.24	0.16	0.43
		LSAEP	-0.65	-0.65	0.20	0.67	0.67	0.10	-0.14	0.01	0.5
		NSPLP	-1.02	-1.00	0.22	1.02	1.00	0.11			
	400	PLSP 1	-1.01	-0.99	0.23	1.01	0.99	0.15	0.05	0,00	0.32
		PLSP 2	-1.08	-1.06	0.22	1.06	1.05	0.15	0.21	0.08	0.40
		LSAEP	-0.64	-0.66	0.15	0.66	0.66	0.06	-0.02	0.09	0.37
		NSPLP	-1.02	-1.00	0.22	1.02	1.00	0.11			
	800	PLSP 1	-0.99	-1.01	0.16	0.99	0.98	0.09	0.05	0.00	0.23
		PLSP 2	-1.06	-1.06	0.21	1.06	1.04	0.13	0.27	0.24	0.42
		LSAEP	-0.62	-0.62	0.12	0.65	0.64	0.05	0.01	0.05	0.29
		NSPLP	-1.01	-1.00	0.16	0.98	0.99	0.07			
0.50	200	PLSP 1	-1.10	-1.04	0.42	1.08	1.00	0.34	0.24	0.01	0.43
		PLSP 2	-1.06	-1.06	0.32	1.12	1.09	0.24	0.33	0.49	0.45
		LSAEP	-0.62	-0.62	0.12	0.65	0.64	0.05	0.01	0.05	0.29
		NSPLP	-1.00	-1.00	0.30	0.98	0.97	0.16			
	400	PLSP 1	-1.04	-1.01	0.30	1.04	0.98	0.23	0.23	0.01	0.36
		PLSP 2	-1.03	-1.01	0.25	1.06	1.03	0.18	0.33	0.42	0.42
		LSAEP	-0.62	-0.61	0.17	0.65	0.64	0.08	0.15	0.27	0.37
		NSPLP	-0.96	-0.94	0.24	0.97	0.97	0.11			
	800	PLSP 1	-0.96	-0.94	0.16	1,00	0.97	0.13	0.24	0.06	0.29
		PLSP 2	-1.02	-1.00	0.18	1.05	1.00	0.15	0.36	0.47	0.40
		LSAEP	-0.62	-0.60	0.12	0.65	0.65	0.05	0.27	0.30	0.19
		NSPLP	-0.98	-0.98	0.15	0.97	0.96	0.07			
0.80	200	PLSP 1	-1.11	-1.03	0.53	1.12	1,00	0.4	0.54	0.79	0.41
		PLSP 2	-0.99	-1.01	0.31	0.99	0.95	0.23	0.45	0.65	0.44
		LSAEP	-0.67	-0.67	0.26	0.65	0.65	0.12	0.47	0.54	0.24
		NSPLP	-0.86	-0.87	0.30	0.85	0.84	0.15			
	400	PLSP 1	-1.03	-0.97	0.36	1.06	0.95	0.35	0.52	0.70	0.39
		PLSP 2	-0.97	-0.93	0.26	0.98	0.96	0.19	0.54	0.74	0.39
		LSAEP	-0.62	-0.62	0.19	0.67	0.66	0.08	0.56	0.57	0.11
		NSPLP	-0.81	-0.81	0.21	0.82	0.81	0.11			
	800	PLSP 1	-0.97	-0.95	0.26	1.00	0.92	0.27	0.49	0.60	0.38
		PLSP 2	-1,00	-0.97	0.23	1,00	0.97	0.20	0.57	0.76	0.39
		LSAEP	-0.63	-0.61	0.13	0.67	0.66	0.06	0.60	0.60	0.07
		NSPLP	-0.80	-0.81	0.15	0.83	0.83	0.08			

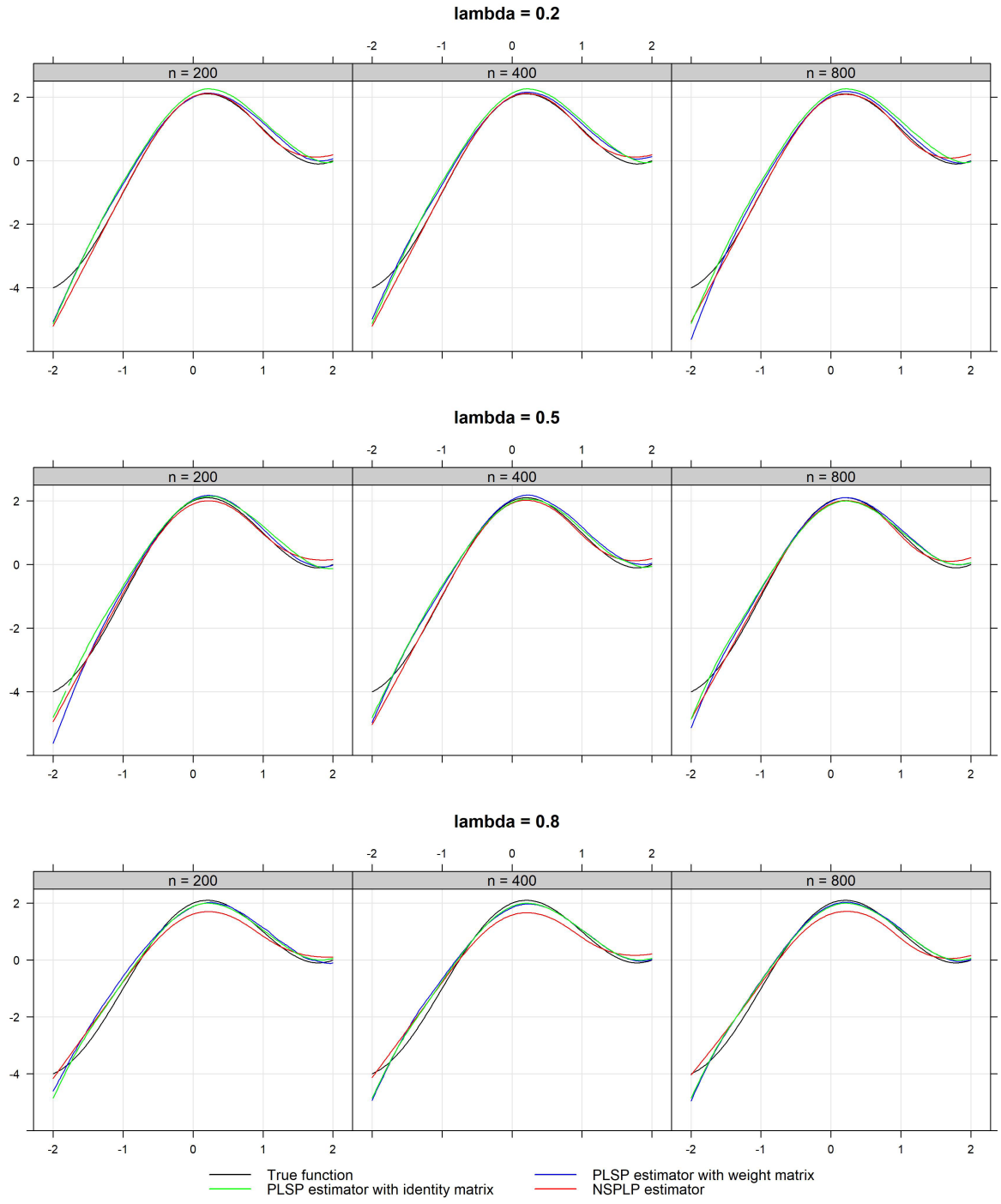


Figure 3.1: The true function $g(\cdot)$ and the average of its estimates, over the 300 replications

3.5 Appendix

Proposition 3.5.1. *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 (3.8) and (3.14), respectively, satisfy*

1. *for all $i, j = 0, 1, 2,\ i + j \leq 2$,*

$$\frac{\partial^{i+j}}{\partial \theta_l^i \partial \theta_r^j} g_\theta(z) \qquad \text{and} \qquad \frac{\partial^{i+j}}{\partial \theta_l^i \partial \theta_r^j} \hat{g}_\theta(z) \qquad \text{exist and are finite for all } 1 \leq l, r \leq p+1.$$

2. $\sup_{\theta \in \Theta} \|\hat{g}_\theta - g_\theta\|$, $\sup_{\theta \in \Theta} \max_{j=1,\ldots,p+1} \left\| \frac{\partial}{\partial \theta_j} (\hat{g}_\theta - g_\theta) \right\|$ *and* $\sup_{\theta \in \Theta} \max_{1 \leq i,j \leq p+1} \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\hat{g}_\theta - g_\theta) \right\|$,

are all order $o_p(1)$ as $n \to \infty$.

Without loss of generality, the proof of this proposition is ensured by Lemma 3.5.2 in the univariate case i.e., Θ, $\mathcal{Z} \subset \mathbb{R}$.

The following lemma is useful in the proof of Lemma 3.5.2. It is an extension of Lemma 8 in Severini and Wong (1992) to spatially dependent data.

Lemma 3.5.1. *Let $\zeta_\theta(Y_i)$ denote a scalar function of Y_{in}, $i = 1, \ldots, n,\ n = 1, 2, \ldots$, depending on a scalar parameter $\theta \in \Theta$, and for $j = 0, 1, 2$, let*

$$\zeta_\theta^{(j)}(Y_{in}) = \frac{\partial^j}{\partial \theta^j} \zeta_\theta(Y_{in}), \qquad i = 1, \ldots, n,\ n = 1, 2, \ldots$$

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 $\sup_\theta \sup_{1 \leq i \leq n, n} \left| \zeta_\theta^{(j)}(Y_{in}) \right| < \infty$ *for $j = 0, \ldots, 3$.*

H.2 *For all $\theta \in \Theta$, $j = 0, 1, 2$, and $1 \leq i, l \leq n$:*

$$\left| \text{Cov}\left(K_{in}(z), K_{ln}(z)\right) \right| \leq \left\{ \text{Var}(K_{in}(z)) \text{Var}(K_{in}(z)) \right\}^{1/2} \varphi\left(\|s_i - s_l\|\right), \tag{3.26}$$

$$\left| \text{Cov}\left(\zeta_\theta^{(j)}(Y_{in}) K_{in}(z), \zeta_\theta^{(j)}(Y_{ln}) K_{ln}(z) \right) \right| \leq \left\{ \text{Var}\left(\zeta_\theta^{(j)}(Y_{in}) K_{in}(z) \right) \text{Var}\left(\zeta_\theta^{(j)}(Y_{ln}) K_{ln}(z) \right) \right\}^{1/2} \varphi\left(\|s_i - s_l\|\right), \tag{3.27}$$

with $K_{in}(z) = K\left((z - Z_{in})/b\right)$.

Let $m_\theta(z) = \mathbb{E}\left(\zeta_\theta(Y_{in}) | Z_{in} = z\right)$ for $z \in \mathcal{Z}$, and assume that $\frac{\partial^j}{\partial \theta^j} m_\theta(\cdot)$ 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 $j = 0, 1, 2$.

Lemma 3.5.1 generalizes Lemma 8 in Severini and Wong (1992) to spatially dependent data.

Proof of Lemma 3.5.1

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 similar to this case and thus 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

$$\sup_\theta \sup_z |\widehat{v}_\theta(z) - v_\theta(z)| = o_p(1) \tag{3.28}$$

and

$$\sup_z \left|\widehat{f}(z) - \bar{f}(z)\right| = o_p(1) \tag{3.29}$$

We give the proof of (3.28), and that of (3.29) is similar.

Asymptotic behavior of $|\widehat{v}_\theta(z) - v_\theta(z)|$

Let us first consider the bias $|\mathbb{E}(\widehat{v}_\theta(z)) - v_\theta(z)|$. We have

$$\begin{aligned}\mathbb{E}(\widehat{v}_\theta(z)) &= (nb^d)^{-1}\sum_{i=1}^{n}\int K\left(\frac{z-u}{b}\right)m_\theta(u)f_i(u)du \\ &= b^{-d}\int v_\theta(u)K\left(\frac{z-u}{b}\right)du; \\ &= \int v_\theta(z-bu)K(u)du\end{aligned}$$

thus,

$$\mathbb{E}(\widehat{v}_\theta(z)) - v_\theta(z) = \int \left(v_\theta(z-bu) - v_\theta(z)\right)K(u)du = o(1)$$

by Assumption A4, the continuity of $f_i(\cdot)$ (see A2) and $m_\theta(\cdot)$, and the compactness of $\mathcal{Z}$. Clearly, the bias term does not depend on θ or z.
Let us now treat $|\widehat{v}_\theta(z) - \mathbb{E}(\widehat{v}_\theta(z))|$. 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

$$\begin{aligned}\mathrm{Var}\left(\zeta_\theta(Y_{in})K_{in}(z)\right) &\leq \mathbb{E}\left(\zeta_\theta^2(Y_{in})K_{in}^2(z)\right)\\ &\leq C\mathbb{E}\left(K_{in}^2(z)\right) = Cb^d\sum_{i=1}^{n}\int K^2(u)f_i(z-ub)du\\ &= Cb^d\sup_u|K(u)|^2\int f_i(z-ub)du = Cb^d\sup_u|K(u)|^2, \qquad (3.30)\end{aligned}$$

because $\zeta_\theta(Y_{in})$ 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 the compactness of $\mathcal{Z}$). Then, we have

$$\mathbf{S}_n = O\left((nb^d)^{-1}\right). \qquad (3.31)$$

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 being the sequence of integers going to ∞, and let $\bar{D}_n$ denote the complement of D_n in the set of locations $\{s_i,\, i=1,...,n\}$.
On the 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

$$\begin{aligned}|A| &= \left|\mathbb{E}\left(\zeta_\theta(Y_{in})K_{in}(z)\zeta_\theta(Y_{jn})K_{jn}(z)\right)\right|\\ &\leq C\left|\int K\left(\frac{z-u}{b}\right)K\left(\frac{z-v}{b}\right)f_{i,j}(u,v)dudv\right|\\ &\leq C\,b^{2d}\left|\int K(u)K(v)f_{i,j}(z-bu,z-bv)dudv\right|\\ &\leq Cb^{2d}\left(\sup_u|K(u)|\right)^2\left|\int f_{i,j}(z-bu,z-bv)dudv\right| = Cb^{2d},\end{aligned}$$

by Assumption **H.1**, $\sup_u|K(u)|<\infty$ (Assumption A4 and the compactness of $\mathcal{Z}$), with $f_{i,j}$ being the joint density (Assumption A2 and the compactness of $\mathcal{Z}$).
Note that the second term B is

$$B = \mathbb{E}\left(\zeta_\theta(Y_{in})K_{in}(z)\right)\mathbb{E}\left(\zeta_\theta(Y_{jn})K_{jn}(z)\right).$$

Using similar arguments as above, we have $|B| \leq Cb^{2d}$ by Assumptions A2 and A4, the compactness of $\mathcal{Z}$ and the continuity of $m_\theta(\cdot)$. Thus, we have

$$\mathbf{R}_n^{(1)} \leq Cn^{-2} \sum_{i,j \in D_n} \leq C \frac{c_n^2 - \rho^2}{n} = O\left(\frac{c_n^2}{n}\right). \tag{3.32}$$

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 (3.30), 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

$$\mathbf{R}_n^{(2)} \leq C(n\, b^d)^{-1} \sum_{i > c_n/\rho} i\varphi(i\rho). \tag{3.33}$$

Thus, we derive the following result:

$$\mathbf{R}_n = \mathbf{R}_n^{(1)} + \mathbf{R}_n^{(2)} = O\left(n^{-1}\left\{c_n^2 + b^{-d} \sum_{i > c_n/\rho} i\varphi(i\rho)\right\}\right). \tag{3.34}$$

The following steps of the proof are inspired by the proof of Lemma 8 in Severini and Wong (1992) (p. 1800–1801). Let

$$\tilde{v}_\theta(z) = \frac{1}{n} b^{-d} \sum_{i=1}^{n} \left\{\zeta_\theta(Y_{in})K_{in}(z) - \mathbb{E}\left(\zeta_\theta(Y_{in})K_{in}(z)\right)\right\}.$$

For some $\epsilon > 0$, Markov's inequality yields

$$\mathbb{P}\left(|\tilde{v}_\theta(z)| > \epsilon\right) \leq \frac{\mathbf{R}_n + \mathbf{S}_n}{\epsilon^2}. \tag{3.35}$$

Now, let θ_1 and θ_2 be two elements in Θ; because $\mathbb{E}\left(\sup_{\theta, 1 \leq i \leq n, n} |\zeta_\theta^{(1)}(Y_{in})|\right) < \infty$ (by **H.1**), there exists a 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 $\sup_{1 \leq i \leq n,\, n} \mathbb{E}\left(|W_{in}^{(1)}|\right) < \infty$ 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 exists a 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 $\sup_{1 \leq i \leq n,\, n} \mathbb{E}\left(|W_i^{(2)}|\right) < \infty$ and

$$\sup_\theta \left|\tilde{v}_\theta(z^{(2)}) - \tilde{v}_\theta(z^{(1)})\right| \leq C \frac{\|z^{(2)} - z^{(1)}\|}{b^{d+1}} \frac{1}{n} \sum_{i=1}^{n} W_{in}^{(2)},$$

because $K(\cdot)$ is Lipschitzian (see Assumption **H.2**).
Hence, there exists a random triangular array $\{W_{in},\ 1 \leq i \leq n,\ n = 1, 2 \ldots\}$ such that $\sup_{1 \leq i \leq n,\, n} \mathbb{E}\left(|W_{in}|\right) <$

∞ and

$$\sup_{\|z^{(2)}-z^{(1)}\|<\delta_1} \sup_{|\theta_2-\theta_1|<\delta_2} \left|\tilde{v}_{\theta_2}(z^{(2)}) - \tilde{v}_{\theta_1}(z^{(1)})\right| \leq C\left(b^{-d}\delta_2 + b^{-(d+1)}\delta_1\right)\frac{1}{n}\sum_{i=1}^{n} W_{in},$$

for some $\delta_1 > 0$, $\delta_2 > 0$ and large n.

Because $\mathcal{Z}$ 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 $\delta > 0$.
In addition, because Θ is compact, one can cover it by $r_n = \lfloor \gamma_n b^{-d} \rfloor$ finite intervals of centers θ_i with the same half length $\delta_2 = O(1/r_n)$.
With these coverings, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{\theta,z}|\tilde{v}_\theta(z)| > \epsilon\right) &\leq \mathbb{P}\left(\max_{j\leq r_n}\max_{k\leq l_n}\left|\tilde{v}_{\theta_j}(z^{(k)})\right| > \epsilon/2\right) \\
&\qquad + \mathbb{P}\left(\sup_{\|z^{(2)}-z^{(1)}\|<\delta_1}\sup_{|\theta_2-\theta_1|<\delta_2}\left|\tilde{v}_{\theta_2}(z^{(2)}) - \tilde{v}_{\theta_1}(z^{(1)})\right| > \epsilon/2\right) \\
&\leq \quad r_n\, l_n\, \mathbb{P}\left(|\tilde{v}_\theta(z)| > \varepsilon/2\right) + Cb^{-d}\left(\delta_2 + \delta_1 b^{-1}\right) \\
&= \; C\, r_n\, l_n(\mathbf{S}_n + \mathbf{R}_n) + Cb^{-d}\left(\delta_2 + \delta_1 b^{-1}\right) \\
&:= \; I^{(1)} + I^{(2)} + I^{(3)},
\end{aligned}$$

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 by the fact that $\varphi(t) \to 0$ as $t \to \infty$ by Assumption A3. This yields the proof. □

Lemma 3.5.2. *For each $\theta \in \Theta$ and $z \in \mathcal{Z}$, let*

$$H(\eta;\theta,z) = \mathbb{E}_0\left(h_{in}^{\theta,\eta}(Y_{in}|X_{in},Z_{in})|Z_{in} = z\right),\ 1 \leq i \leq n,\ n = 1,2,\ldots$$

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 the 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 $j_1+\cdots+j_{\tilde{p}}+r \leq 6$, 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 all 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

$$\sup_z\sup_\theta\sup_\eta\left|\frac{\partial^j}{\partial\theta^j}H^{(k)}(\eta;\theta,z)\right| < \infty, \tag{3.36}$$

for $j = 0,1,2$ and $k = 2,3,4$ such that $j+k \leq 4$, 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 $j = 0,1,2$,

$$\sup_\theta\sup_z\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{g}_\theta(z) - g_\theta(z)\right)\right| = o_p(1). \tag{3.37}$$

The assumptions used in the previous lemma are satisfied under the conditions used in the main results. **Condition I** is needed to ensure the 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 (3.1). **Condition S** allows integrals to be interchanged with differentiation; this will be combined with the implicit function theorem (see Saaty & Bram, 2012) to ensure the 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 (3.36) 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 3.5.2

The proof of this lemma is similar to that of Lemma 5 in Severini and Wong (1992). Let us follow similar lines as in the proof of Lemma 3.5.1 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 3.5.1 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 $\theta \in \Theta,\ z \in \mathcal{Z}$, (3.26) is satisfied and (3.27) 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 clear that **H.1'** is verified, and **H.2'** is also satisfied by Assumption A3 (in particular, conditions (3.19)).
Using the results of Lemma 3.5.1, we have the following for all $j = 0, 1, 2$:

$$\sup_{\theta,\eta,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(1)}(\eta;\theta,z) - H^{(1)}(\eta;\theta,z)\right)\right| = o_p(1), \tag{3.38}$$

$$\sup_{\theta,\eta,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(2)}(\eta;\theta,z) - H^{(2)}(\eta;\theta,z)\right)\right| = o_p(1), \tag{3.39}$$

$$\sup_{\theta,\eta,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(3)}(\eta;\theta,z) - H^{(3)}(\eta;\theta,z)\right)\right| = o_p(1), \tag{3.40}$$

$$\sup_{\theta,\eta,z}\left|\frac{\partial^j}{\partial\theta^j}\left(\widehat{H}_n^{(4)}(\eta;\theta,z) - H^{(4)}(\eta;\theta,z)\right)\right| = o_p(1). \tag{3.41}$$

Under Assumption A1, for any $\epsilon > 0$, there exists $\gamma > 0$ such that

$$\begin{aligned} P\left(\sup_{\theta,z}|\widehat{g}_\theta(z) - g_\theta(z)| > \epsilon\right) &\le P\left(\sup_{\theta,z}|H^{(1)}(\theta, \widehat{g}_\theta(z), z)| > \gamma\right) \\ &= P\left(\sup_{\theta,z}|\widehat{H}^{(1)}(\widehat{g}_\theta(z);\theta,z) - H^{(1)}(\widehat{g}_\theta(z);\theta,z)| > \gamma\right) \\ &\le P\left(\sup_{\theta,z,\eta}|\widehat{H}^{(1)}(\eta;\theta,z) - H^{(1)}(\eta;\theta,z)| > \gamma\right). \end{aligned}$$

Hence,

$$\sup_{\theta,z}|\widehat{g}_\theta(z) - g_\theta(z)| = o_p(1) \tag{3.42}$$

The remainder of the proof is very similar to that of Lemma 5 in Severini and Wong (1992) (p. 1798–1799); for the sake 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

$$\inf_{\theta} \inf_{z} \inf_{|\eta - g_\theta(z)| \le \epsilon} \left| H^{(2)}(\eta; \theta, z) \right| > 0. \tag{3.43}$$

Because $g_\theta(z)$ 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

$$\begin{aligned} 0 &= \widehat{H}^{(1)}(\widehat{g}_\theta(z); \theta, z) - H^{(1)}(g_\theta(z); \theta, z) \\ &= \widehat{H}^{(1)}(\widehat{g}_\theta(z); \theta, z) - H^{(1)}(\widehat{g}_\theta(z); \theta, z) + H^{(1)}(\widehat{g}_\theta(z); \theta, z) - H^{(1)}(g_\theta(z); \theta, z) \\ &= r_n(\theta, z) + d_n(\theta, z) \left(\widehat{g}_\theta(z) - g_\theta(z) \right), \end{aligned} \tag{3.44}$$

for each θ, z, where

$$r_n(\theta, z) = \widehat{H}^{(1)}(\widehat{g}_\theta(z); \theta, z) - H^{(1)}(\widehat{g}_\theta(z); \theta, z) \qquad \text{and} \quad d_n(\theta, z) = \int_0^1 H^{(2)}(t g_\theta(z) + (1-t)\widehat{g}_\theta(z); \theta, z) dt.$$

Note that by (3.43) and $\sup_\theta \|\widehat{g}_\theta - g_\theta\| = o_p(1)$, we have

$$\liminf \inf_{z} \inf_{\theta} \left| \widehat{H}^{(2)}(\widehat{g}_\theta(z); \theta, z) \right| > 0 \qquad \text{and} \qquad \liminf \inf_{z} \inf_{\theta} |d_n(\theta, z)| > 0 \qquad \text{as} \quad n \to \infty. \tag{3.45}$$

Because

$$\widehat{H}^{(1)}(\widehat{g}_\theta(z); \theta, z) = 0,$$

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 (3.45), (3.38), and (3.39) that

$$\sup_\theta \sup_z \left| \frac{\partial \widehat{g}_\theta}{\partial \theta}(z) \right| = O_p(1).$$

Similarly, we have

$$\sup_\theta \sup_z \left| \frac{\partial^j \widehat{g}_\theta}{\partial \theta^j}(z) \right| = O_p(1), \qquad j = 0, 1, 2. \tag{3.46}$$

Then, (3.46) and (3.38)–(3.41) yield

$$\sup_\theta \sup_z \left| \frac{\partial^j}{\partial \theta^j} r_n(\theta, z) \right| = o_p(1), \qquad \text{and} \qquad \sup_\theta \sup_z \left| \frac{\partial^j}{\partial \theta^j} d_n(\theta, z) \right| = O_p(1), \qquad j = 0, 1, 2. \tag{3.47}$$

Now, differentiating (3.44) with respect to θ yields

$$\frac{\partial r_n}{\partial \theta}(\theta, z) + (\widehat{g}_\theta(z) - g_\theta(z)) \frac{\partial d_n}{\partial \theta}(\theta, z) + d_n(\theta, z) \left(\frac{\partial \widehat{g}_\theta}{\partial \theta}(z) - \frac{\partial g_\theta}{\partial \theta}(z) \right) = 0. \tag{3.48}$$

Then, by (3.38)–(3.47),

$$\sup_\theta \sup_z \left| \frac{\partial \widehat{g}_\theta}{\partial \theta}(z) - \frac{\partial g_\theta}{\partial \theta}(z) \right| = o_p(1).$$

On can similarly obtain

$$\sup_{\theta}\sup_{z}\left|\frac{\partial^2\widehat{g}_\theta}{\partial\theta^2}(z)-\frac{\partial^2 g_\theta}{\partial\theta^2}(z)\right|=o_p(1).$$

This completes the proof. □

Proof of Theorem 3.2.1

By Lemmas 3.5.3 and 3.5.4, Q_n converges to Q in probability uniformly, i.e.,

$$\sup_{\theta\in\Theta}|Q_n(\theta,g_\theta)-Q(\theta,g_\theta)|=o_p(1). \tag{3.49}$$

This result allows one to obtain

$$\left|Q(\hat{\theta},g_{\hat{\theta}})-Q(\theta_0,g_0)\right|=o_p(1). \tag{3.50}$$

Indeed, using $|\sup a-\sup b|\leq\sup|a-b|$, 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 3.5.5, (3.49) and $\sup_\theta Q(\theta,g_\theta)=Q(\theta_0,g_0)$ (see Assumption A8).

By Assumption A8, we have for a given $\theta\in\Theta$ that there exists $\varepsilon>0$ and an open neighbourhood N_θ such that

$$\inf_{\theta_1\in N_\theta}|Q(\theta_1,g_{\theta_1})-Q(\theta_0,g_0)|>\varepsilon. \tag{3.51}$$

This and (3.50) imply that

$$\mathbb{P}_0\left(\hat{\theta}\in N_\theta\right)\leq\mathbb{P}_0\left(\left|Q(\hat{\theta},g_{\hat{\theta}})-Q(\theta_0,g_0)\right|>\varepsilon\right)\to 0,\ \text{as } n\to\infty. \tag{3.52}$$

Let N_0 be an open neighbourhood 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 neighbourhood N_θ satisfies (3.51)). By the compactness of Θ_0, let $\{N_{\theta_1},\ldots,N_{\theta_r}\}$ be a finite sub-covering; then,

$$\mathbb{P}_0\left(\hat{\theta}\notin N_0\right)=\mathbb{P}_0\left(\hat{\theta}\in\Theta_0\right)\leq\sum_{j=1}^{r}\mathbb{P}_0\left(\hat{\theta}\in N_{\theta_j}\right)\to 0,\ \text{ as } n\to\infty,$$

by (3.52). Therefore, we can conclude that

$$\hat{\theta}-\theta_0=o_p(1),\qquad\text{as}\qquad n\to\infty.$$

This yields the proof of Theorem 3.2.1. □

Lemmas 3.5.3-3.5.5

We use the following notation:

$$\eta_i = g(Z_{in}); \quad \tilde{U}_{in} = \tilde{U}_{in}(\theta, \eta_i); \qquad \Phi_{in} = \Phi(G_{in}(\theta, g_\theta)); \qquad \Lambda_{in} = \Lambda(G_{in}(\theta, g_\theta)),$$

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 3.5.3. *Under Assumptions A3, A6 and A9, we have for all $\theta \in \Theta$,*

$$S_n(\theta, g_\theta) - S(\theta, g_\theta) = o_p(1). \tag{3.53}$$

In addition, we have

$$Q_n(\theta, g_\theta) - Q(\theta, g_\theta) = o_p(1), \tag{3.54}$$

if $M_n - M = o_p(1)$.

Note that if Assumption A10 is satisfied, then $M_n - M = o_p(1)$.

Proof of Lemma 3.5.3

Let us start with the proof of (3.53). We 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_i is the $q \times 1$ vector representing the ith row in the matrix of instrumental variables. By definition (see (3.13)), we have $\mathbb{E}_0(S_n(\theta, g_\theta)) - S(\theta, g_\theta) = o(1)$. Then, it suffices to show that

$$S_n(\theta, g_\theta) - \mathbb{E}_0(S_n(\theta, g_\theta)) = o_p(1). \tag{3.55}$$

Indeed (omitting the $(\theta, g_\theta)-$arguments to simplify the notation), we have

$$\begin{aligned}
\mathbb{E}_0\left(\|S_n - \mathbb{E}_0(S_n)\|^2\right) &= n^{-2} \sum_{i,j=1}^{n} \mathbb{E}_0\left(\left(\xi_{in}\tilde{U}_{in} - \mathbb{E}_0(\xi_{in}\tilde{U}_{in})\right)^T \left(\xi_{jn}\tilde{U}_{jn} - \mathbb{E}_0(\xi_{jn}\tilde{U}_{jn})\right)\right) \\
&\overset{(3.20)}{\leq} n^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} \sum_{t=1}^{q} \left\{\mathrm{Var}_0\left(\xi_{itn}\tilde{U}_{in}\right) \mathrm{Var}_0\left(\xi_{jtn}\tilde{U}_{jn}\right)\right\}^{1/2} \\
&\leq C n^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} = O\left(n^{-1} \sum_{s=1}^{\sqrt{n}} s\varphi(s)\right) = o(1),
\end{aligned}$$

because $\mathrm{Var}_0(\xi_{itn}\tilde{U}_{in})$ is bounded uniformly on θ, i, and $t = 1, \ldots, q$ (by Assumption A6) and because $\varphi(s) \to$ as $s \to +\infty$ (by assumption A3). This completes the proof of (3.55) and thus

that of (3.53).
The proof of (3.54) is made straightforward by combining (3.53) with Assumption A10. □

Lemma 3.5.4. *Under Assumptions A6-A9, we have $S_n(\cdot, g_\cdot) - S(\cdot, g_\cdot)$ is stochastically equicontinuous on Θ.*
In addition, if $M_n - M = o_p(1)$, then we have $Q_n(\cdot, g_\cdot) - Q(\cdot, g_\cdot)$ is also stochastically equicontinuous on Θ.

Proof of Lemma 3.5.4

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 $S_n(\cdot, g_\cdot)$ is stochastically equicontinuous on θ because $S(\cdot, g_\cdot)$ is continuous by Assumption A8. It suffices to show that (Andrews, 1992) for each $\theta_1, \theta_2 \in \Theta$:

$$\|S_n(\theta_1, g_{\theta_1}) - S_n(\theta_2, g_{\theta_2})\| = O_p\left(\|\theta_1 - \theta_2\|\right). \tag{3.56}$$

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 3.5.1, we have that $\sup_{i,n} \|\xi_{in}\|$ is bounded and $\sup_\theta \left\| \frac{\partial g_\theta}{\partial \theta} \right\|$ is finite, respectively. Then, we have to show that

$$n^{-1} \sum_{i=1}^{n} \sup_{\theta,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| + \sup_{\theta,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| = O_p(1); \tag{3.57}$$

This is equivalent to

$$\sup_{\theta,\eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta, \eta) \right\| = O_p(1), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{3.58}$$

and

$$\sup_{\theta,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| = O_p(1), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{3.59}$$

Let us prove (3.58) in the following. The proof of (3.59) follows the same lines and is thus omitted.

Proof of (3.58):
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

$$a_{in}(\theta) := (v_{in}(\lambda))^{-1} \qquad \text{and} \qquad b_{in}(\theta,\eta) := X_{in}^T\beta + \eta, \qquad 1 \le i \le n,\ n = 1,2,\ldots, \tag{3.60}$$

with $\theta^T = (\beta^T, \lambda)$. We have

$$\begin{aligned}\frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta,\eta) &= \Big\{\Lambda'(G_{in}(\theta,\eta))(Y_{in} - \Phi(G_{in}(\theta,\eta))) \\ &\quad - \Lambda(G_{in}(\theta,\eta))\phi(G_{in}(\theta,\eta))\Big\}\frac{\partial G_{in}}{\partial \theta}(\theta,\eta)\end{aligned} \tag{3.61}$$

where $\Lambda'(\cdot)$ denotes the derivative of $\Lambda(\cdot)$.

Let us first establish that

$$\sup_{t\in\mathcal{M}, y\in\{0,1\}} \left|\Lambda'(t)(y-\Phi(t)) - \phi(t)\Lambda(t)\right| < \infty, \tag{3.62}$$

which is equivalent to showing that $\Lambda'(t)$ and $\phi(t)\Lambda(t)$ are bounded uniformly in $t \in \mathcal{M}$ (the definition of $\mathcal{M}$ is given in **A.1**). Because $\phi'(t) = -t\phi(t)$, we can rewrite $\Lambda'(t)$ as

$$\Lambda'(t) = \frac{1}{\Phi(t)}\left\{\frac{\phi(t)}{1-\Phi(t)}\left(\frac{\phi(t)}{1-\Phi(t)} - t\right)\right\} - \frac{\phi^2(t)}{\Phi^2(t)(1-\Phi(t))}. \tag{3.63}$$

Notice that $\Lambda(\cdot)$ and $\Lambda'(\cdot)$ may be unbounded only at $\pm\infty$, and because $\mathcal{M}$ is a compact subset of $\mathbb{R}$, these functions are bounded on $\mathbb{R}$. This establishes (3.62).
We remark that

$$\left\|\frac{\partial G_{in}(\theta,\eta)}{\partial\theta}\right\| \le \left\|\frac{\partial a_{in}(\theta)}{\partial\theta}\right\| |b_{in}(\theta,\eta)| + \left\|\frac{\partial b_{in}(\theta,\eta)}{\partial\theta}\right\| |a_{in}(\theta)|. \tag{3.64}$$

Then, $\left\|\frac{\partial G_{in}(\theta,\eta)}{\partial\theta}\right\|$ is bounded uniformly in i,n,θ,η by Assumptions A6 and A9 and the compactness of Θ (see assumption A7). This completes the proof of (3.58); hence, (3.56) is proved. □

Lemma 3.5.5. *Under the assumptions of Proposition 3.5.1 and Assumptions A6 and A9, we have*

$$\sup_{\theta\in\Theta} \|S_n(\theta,\hat{g}_\theta) - S_n(\theta,g_\theta)\| = o_p(1). \tag{3.65}$$

If in addition $M_n - M = o_p(1)$, then we have

$$\sup_{\theta\in\Theta} |Q_n(\theta,\hat{g}_\theta) - Q_n(\theta,g_\theta)| = o_p(1). \tag{3.66}$$

Proof of Lemma 3.5.5

Let us prove (3.65). For each $\theta \in \Theta$

$$\begin{aligned}
\|S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta)\| &= n^{-1} \left\| \sum_{i=1}^{n} \xi_i \left(\tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \right) \right\| \\
&\leq n^{-1} \sum_{i=1}^{n} \sup_{i,n} \|\xi_{in}\| \left| \tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_i(\theta, g_\theta) \right| \\
&\leq n^{-1} \sum_{i=1}^{n} \sup_{i,n} \|\xi_{in}\| \sup_{\theta,\eta} \left| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right| \sup_{\theta} \|\hat{g}_\theta - g_\theta\| \\
&= o_p(1),
\end{aligned}$$

because $\sup_{i,n} \|\xi_{in}\| = O_p(1)$ (by Assumption A6), $\sup_\theta \|\hat{g}_\theta - g_\theta\| = o_p(1)$ (see Proposition 3.5.1) 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 3.5.4).
The proof of (3.66) is made trivial by combining (3.65) with Assumption A10. □

Proof of Theorem 3.2.2

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

$$\hat{\theta} - \theta_0 = - \left\{ \frac{d^2}{d\theta d\theta^T} Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta^*} \right\}^{-1} \left\{ \frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} \right\}, \tag{3.67}$$

for some θ^* between θ_0 and $\hat{\theta}$.
First, we would like to replace $\hat{g}_\theta(.)$ in (3.67) 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 the same behavior as a function of θ in a neighbour of θ_0. In other words,

$$\sup_\theta \left\| \frac{d^2}{d\theta d\theta^T} Q_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta) \right\| = o_p(1) \tag{3.68}$$

and

$$\frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} - \frac{d}{d\theta} Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} = o_p(1). \tag{3.69}$$

We remark that (3.68) is equivalent to

$$\sup_\theta \left\| \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta) - \frac{d}{d\theta} S_n(\theta, g_\theta) \right\| = o_p(1) \tag{3.70}$$

and

$$\sup_\theta \left\| \frac{d^2}{d\theta d\theta^T} S_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T} S_n(\theta, g_\theta) \right\| = o_p(1) \tag{3.71}$$

by (3.11) (because $M_n - M = o_p(1)$ thanks to Assumption A10) and

$$\sup_{\theta} \|S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta)\| = o_p(1)$$

(see Lemma 3.5.5).
Then, (3.70) and (3.71) follow immediately from Lemma 3.5.8.
To prove (3.69), we have the following Taylor 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 3.5.8 below (see (3.89)). Therefore, we obtain

$$\begin{aligned}\frac{d}{d\theta}Q_n(\theta, \hat{g}_\theta)\Big|_{\theta=\theta_0} - \frac{d}{d\theta}Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} &= \frac{d}{d\theta}\frac{\partial Q_n}{\partial g}(\theta, g_\theta)\Big|_{\theta=\theta_0}(\hat{g}_0 - g_0) \\ &\quad + \frac{\partial Q_n}{\partial g}(\theta_0, g_0)(\hat{g}_0' - g_0') + \frac{d}{d\theta} r_n(\theta)\Big|_{\theta=\theta_0}, \\ &= o_p(1)\end{aligned}$$

by Lemma 3.5.7, where $g_0'(.) = \frac{g_\theta}{\partial\theta^T}(.)\Big|_{\theta=\theta_0}$.
Consequently, we obtain

$$\hat{\theta} - \theta_0 = -\left\{\frac{d^2}{d\theta d\theta^T}Q_n(\theta, g_\theta)\Big|_{\theta=\theta^*}\right\}^{-1}\left\{\frac{d}{d\theta}Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0}\right\} + o_p(1) \tag{3.72}$$

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 (3.72) by its limit $B_2(\theta_0)$.
Let us consider the first- and second-order differentials of $Q_n(\theta, g_\theta)$ with respect to θ:

$$\frac{d}{d\theta}Q_n(\theta, g_\theta) = 2S_n^T(\theta, g_\theta)M_n\left\{\frac{\partial S_n}{\partial\theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta)g_\theta'\right\} \tag{3.73}$$

with g_θ' being a $1 \times \tilde{p}$ ($\tilde{p} = p + 1$) matrix given by $\frac{\partial g_\theta}{\partial\theta^T}$ and

$$\begin{aligned}\frac{d^2}{d\theta d\theta^T}Q_n(\theta, g_\theta) &= 2\left\{\frac{\partial S_n}{\partial\theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta)g_\theta'\right\}^T M_n\left\{\frac{\partial S_n}{\partial\theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta)g_\theta'\right\} \\ &\quad + 2S_n^T(\theta, g_\theta)M_n\frac{d}{d\theta^T}\left\{\frac{\partial S_n}{\partial\theta}(\theta, g_\theta) + \frac{\partial S_n}{\partial g}(\theta, g_\theta)g_\theta'\right\}\end{aligned} \tag{3.74}$$

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

$$S_n(\theta^*, g_{\theta^*}) = S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) + S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),$$

because $S(\theta_0, g_0) = 0$ and by Lemmas 3.5.3-3.5.4,

$$S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),$$

and because θ^* lies between $\hat{\theta}$ and θ_0, by Lemma 3.5.4

$$S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) = o_p(1).$$

Using similar arguments as in the proof of (3.58) in Lemma 3.5.4 using Assumption A9 to ensure the boundedness when differentiating twice with respect to θ, we have

$$\left\| \frac{d}{d\theta^T}\frac{\partial S_n}{\partial \theta}(\theta, g_\theta) \right\| = O_p(1) \qquad \text{and} \qquad \left\| \frac{d}{d\theta^T}\frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta^{'} \right\| = O_p(1). \tag{3.75}$$

Then, we can ignore the second term in the right-hand side of (3.74) at $\theta = \theta^*$. Hence, by Lemma 3.5.6 and $\theta^* - \theta_0 = o_p(1)$ (thanks to Theorem 3.2.1), 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

$$\begin{aligned}\left.\frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta)\right|_{\theta=\theta^*} &= 2\left\{\frac{\partial S}{\partial \theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0) g_0^{'}\right\}^T M \left\{\frac{\partial S}{\partial \theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0) g_0^{'}\right\} + o_p(1) \\ &= 2\, B_2(\theta_0) + o_p(1).\end{aligned}$$

We remark that

$$\left.\frac{d}{d\theta} Q_n(\theta, g_\theta)\right|_{\theta=\theta_0} = 2 S_n^T(\theta_0, g_0) M_n \left\{\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_0^{'}\right\}.$$

Then, by (3.79) (see the proof of Lemma 3.5.6), 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 obtain

$$\frac{d}{d\theta}Q_n(\theta, g_\theta)\Big|_{\theta=\theta_0} = 2S_n^T(\theta_0, g_0)M\left\{\frac{\partial S}{\partial \theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0^{'}\right\} + o_p(1).$$

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 be shown 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 $\|w\| = 1$,

$$\begin{aligned} A_n &= w^T\left\{\mathbb{E}_0\left(nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)\right)\right\}^{-1/2}\sqrt{n}S_n(\theta_0, g_0) \\ &= n^{-1/2}\sum_{i=1}^{n} B_{in}, \end{aligned}$$

with

$$B_{in} = w^T\left\{\mathbb{E}_0\left(nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)\right)\right\}^{-1/2}\xi_{in}\tilde{U}_{in}(\theta_0, g_0).$$

By the 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 $\mathcal{G}_{n1}, \ldots, \mathcal{G}_{nJ}$, $1 \leq J < \infty$, 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, as 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 $\mathcal{F} = \left\{f : \forall t \in \mathbb{R} f(t) = t \text{ or } \exists \upsilon \in \mathbb{R} : \forall t \in \mathbb{R} f(t) = e^{\iota \upsilon t}\right\}$, 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

$$\sigma_{jtn}^2=\mathbb{E}_0(A_{jtn}^2),\qquad\text{and}\qquad\gamma_{nj}^2=\sum_{t=1}^{m_{jn}}\sigma_{jtn}^2.$$

Assumption C. For some $\tau>1$

$$\mathbb{E}_0\left(|A_{jtn}|^{2\tau}\right)=o\left(\sigma_{jtn}^2\gamma_{jn}^{2\tau-2}\right),\ j=1,\ldots,J,\ t=1,\ldots,m_{jn}.$$

If assumptions $A-C$ hold, then by Theorem 1 in Pinkse et al. (2007), we have $A_n\longrightarrow\mathcal{N}(0,1)$. Thus, to complete the proof, we have to check these assumptions in our context.

Assumption A: This holds under (3.20) (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 $(m_{1n}+m_{2n})c_n=O(n)$. Because $\varphi(\cdot)$ is a decreasing function (Assumption A3), $\alpha_{jn}=O(\varphi(\sqrt{c_n}))$ 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 is the limit of
$\mathbb{E}_0\left(nS_n(\theta_0,g_0)S_n^T(\theta_0,g_0)\right)$. 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 because ξ_{in} is bounded uniformly on i and n by Assumption A6, as is $\tilde{U}_{in}(\theta_0,g_0)$. Then, for all $j=1,\ldots,J$ and $t=1,\ldots,m_{nj}$,

$$\sigma_{jtn}=\left\{n^{-1}\mathbb{E}_0\left(\sum_{i\in\mathcal{G}_{jtn}}B_{in}\right)\right\}^{1/2}=O\left(n^{-1/2}\text{Card}(\mathcal{G}_{jtn})\right)$$

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 $j = 2, \ldots, J$

$$\frac{\gamma_{jn}}{\gamma_{1n}} = O\left(\frac{m_{jn} \max_{t \le m_{jn}} \text{Card}(\mathcal{G}_{jtn})}{m_{1n} \max_{t \le m_{1n}} \text{Card}(\mathcal{G}_{1tn})}\right) = O\left(\frac{m_{jn}}{m_{1n}}\right) \to 0 \text{ as } n \to \infty,$$

because $m_{jn}/m_{1n} = o(1)$ for all $j = 2, \ldots, J$ as $n \to \infty$.
Assumption C : By an easy calculation, we can show that

$$\frac{\mathbb{E}_0\left(|A_{jtn}|^{2\tau}\right)}{\sigma_{jtn}^2 \gamma_{jn}^{2\tau-2}} = O(m_{jn}^{2-2\tau}) \to 0 \text{ as } n \to \infty.$$

Lemma 3.5.6. *Under the assumptions of Theorem 3.2.2 and for any $\tilde{\theta}$ such that $\tilde{\theta} - \theta_0 = o_p(1)$, we have*

$$\frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) = o_p(1) \tag{3.76}$$

and

$$\frac{\partial S_n}{\partial g}(\tilde{\theta}, g_{\tilde{\theta}}) g'_{\tilde{\theta}} - \frac{\partial S}{\partial g}(\theta_0, g_0) g'_0 = o_p(1), \tag{3.77}$$

with $g'_{\tilde{\theta}}(.) = \frac{g_\theta}{\partial \theta^T}(.)\Big|_{\theta=\tilde{\theta}}$.

Proof of Lemma 3.5.6

To prove (3.76), 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

$$w^T \left\{ \frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right\} = o_p(1) \tag{3.78}$$

and

$$w^T \left\{ \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) \right\} = o_p(1). \tag{3.79}$$

The proof of (3.78) is similar to that of (3.56), using the fact that

$$\sup_{\theta,\eta} \left\| \frac{\partial^2 \tilde{U}_i}{\partial \theta \partial \theta^T}(\theta, \eta) \right\| \quad \text{and} \quad \sup_{\theta,\eta} \left\| \frac{\partial^2 \tilde{U}_i}{\partial \theta \partial \eta}(\theta, \eta) \right\|$$

are bounded uniformly on i and n, and $\tilde{\theta} - \theta_0 = o_p(1)$.
Now, let us prove (3.79). By the definition of $S(\cdot\,, \cdot)$ (see 3.13)

$$\lim_{n\to\infty} \mathbb{E}_0 \left(\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right) = \frac{\partial S}{\partial \theta}(\theta_0, g_0).$$

Thus, it suffices to prove that

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - w^T \mathbb{E}_0 \left(\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right) = o_p(1). \tag{3.80}$$

Let

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) = n^{-1} w^T \xi_{in} \frac{\partial \tilde{U}_{in}}{\partial \theta}(\theta_0, \eta_i^0), = \Delta_{n1} - \Delta_{n2}, \tag{3.81}$$

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 (3.80) is then reduced to proving

$$\mathbb{E}_0\left(\|\Delta_{n1}\|^2\right) = o(1) \qquad \text{and} \qquad \mathbb{E}_0\left(\|\Delta_{n2} - \mathbb{E}_0(\Delta_{n2})\|^2\right) = o(1). \tag{3.82}$$

This last part is trivial because $\xi_{in}^{(1)}$ and $\xi_{in}^{(2)}$ are bounded uniformly on i and n (see Assumption A6 and the compactness of Θ, $\mathcal{X}$, and $\mathcal{Z}$) and by use of the mixing condition (3.20) and (3.21) in Assumption A3. This completes the proof of (3.76).

To prove (3.77), we remark that

$$\frac{\partial S_n}{\partial g}(\tilde{\theta}, g_{\tilde{\theta}}) g_{\tilde{\theta}}^{'} - \frac{\partial S}{\partial g}(\theta_0, g_0) g_0^{'} = \left\{\frac{\partial S_n}{\partial g}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial g}(\theta_0, g_0)\right\} g_{\tilde{\theta}}^{'} + \frac{\partial S}{\partial g}(\theta_0, g_0)\left(g_{\tilde{\theta}}^{'} - g_0^{'}\right). \tag{3.83}$$

Consider the second term on the right-hand side in (3.83), where we remark that because $\left\|\frac{\partial S}{\partial g}(\theta_0, g_0)\right\|$ 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 on the right-hand side in (3.83), because $g_{\tilde{\theta}}^{'} = O_p(1)$ by Proposition 3.5.1, using similar arguments as when proving (3.76) permits one 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 (3.77). □

Lemma 3.5.7. *Under the assumptions of Theorem 3.2.2, 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 3.5.7

To prove (i), and we 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 (3.5.3) and $S(\theta_0, g_0) = 0$, we obtain

$$S_n(\theta_0, g_0) = S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1). \tag{3.84}$$

In addition, we have

$$\begin{aligned}\left\|\frac{\partial^2 S_n}{\partial \theta \partial g}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\| &= n^{-1}\left\|\sum \xi_{in}\frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\| \\ &\leq n^{-1}\sum \sup_{i,n}\|\xi_{in}\| \sup_{\eta}\left\|\frac{\partial^2 \tilde{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta)\right\| \|\hat{g}_0 - g_0\| \\ &= o_p(1),\end{aligned} \tag{3.85}$$

because ξ_i is bounded uniformly on i, n and θ (Assumption A6), $\|\hat{g}_0 - g_0\| = o_p(1)$ by Proposition 3.5.1, 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 (3.85), we obtain

$$\begin{aligned}\left\|\frac{\partial^2 S_n}{\partial g^2}(\theta_0, g_0)(\hat{g}_0 - g_0)g_0^{'}\right\| &= n^{-1}\left\|\sum \xi_i\frac{\partial^2 U_{in}}{\partial \eta^2}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))g_0^{'}(Z_{in})\right\| \\ &= o_p(1),\end{aligned} \tag{3.86}$$

$$\left\| \frac{\partial S_n}{\partial g}(\theta_0, g_0)(\hat{g}_0 - g_0)g_0^{'} \right\| = n^{-1} \left\| \sum \xi_{in} \frac{\partial U_{in}}{\partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))g_0^{'}(Z_{in}) \right\| = o_p(1), \tag{3.87}$$

and

$$\left\| \frac{\partial S_n}{\partial \theta}(\theta_0, g_0)(\hat{g}_0 - g_0) \right\| = n^{-1} \left\| \sum \xi_{in} \frac{\partial U_{in}}{\partial \theta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in})) \right\| = o_p(1). \tag{3.88}$$

Combining (3.84)-(3.88) with Assumption A10 permits one 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 3.5.8. *Under the assumptions of Theorem 3.2.2, 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 3.5.8

By applying Taylor's theorem to $\tilde{U}_i(\theta, \cdot)$ for each $\theta \in \Theta$, we obtain

$$\begin{aligned} S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) &= n^{-1} \sum_{i=1}^{n} \xi_{in} \left(\tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \right) \\ &= n^{-1} \sum_{i=1}^{n} \xi_{in} \left(\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}) \right) \\ &\quad \times \int_0^1 \frac{\partial \tilde{U}_{in}}{\partial \eta} \left(\theta, g_\theta(Z_{in}) + t \left(\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}) \right) \right) dt \\ &:= r_n^{(1)}(\theta). \end{aligned}$$

Because the instrumental 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 3.5.1, it suffices to show that

$$\sup_{\theta,\eta} \sup_i \left\| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| = O_p(1) \tag{3.89}$$

$$\sup_{\theta,\eta} \sup_i \left\| \frac{\partial}{\partial \theta} \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| = O_p(1) \qquad \text{and} \qquad \sup_{\theta,\eta} \sup_i \left\| \frac{d^2}{\partial \theta \partial \theta^T} \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| = O_p(1). \tag{3.90}$$

Equation (3.89) is already proved in the proof of Lemma 3.5.4 (see (3.59)). The proof of (3.90)

can be established in a similar manner and is thus omitted. □

Chapter 4

Application of spatial models to investigate suicide recidivism in Nord-Pas-de-Calais.

4.1 Introduction

According to the World Health Organization (WHO), more than 800000 people die from suicide every year, including 10000 in France, preventing suicide is a major public health issue. In 2013, the World Health Assembly adopted the first-ever Mental Health Action Plan of the World Health Organization. This plan aimed to reduce the rates of suicide by 10% from 2013 to 2020. The understanding of risk factors involved in suicidal behavior is crucial for the development of effective prevention plans. Interestingly, one of the most robust risk factors for death by suicide is a history of previous suicide attempts, as a substantial number of patients who attempt a suicide ultimately die by suicide.

Suicide is a serious global public health issue and is among the leading causes of death worldwide. In 2019, more than one in every 100 deaths (1.3%) were the result of suicide. A previous suicide attempt is one of the most important contributory factors of future suicide (Hawton and Heeringen, 2009) as well as death by suicide (Brent et al., 1996). Hulten et al. (2001) investigated the repetition of attempted suicide among young people aged 15–19 years in some European countries and identified relevant factors associated with repeated suicidal behaviour. Many studies have proven that postcard or telephone follow-up helped to prevent reattempts (Beautrais et al., 2010; Carter et al., 2005; Cedereke et al., 2002; Evans et al., 2005; Exbrayat et al., 2017; Guillaume et al., 2006; Motto & Bostrom, 2001; Vaiva et al., 2011). Exbrayat et al. (2017) verified that a protocol of early telephone follow-up after attempted suicide is helpful in preventing reattempts.

In France, the VigilanS healthcare system is an effort to support those who have attempted suicide in various regions. It was established in the Nord-Pas-de-Calais region in February 2015. This programme to monitor and prevent recidivism of suicide attempts is executed via phone calls by teams of professionals who are specialized in this type of remote care. At Lille University Hospital, this six-month programme is managed by the adult psychiatry department under Professor Guillaume Vaiva. Posthospital support is offered to those patients who attempted suicide.

A patient discharged from the hospital is given a resource card which includes the single regional call number. The patient is also given an information note indicating the terms and

conditions of the system as well as his or her right to object. At the same time, the partner center notifies the VigilanS secretariat of the patient's discharge and entry into the surveillance system by sending an encrypted fax or email, or via a Web interface. The VigilanS secretariat opens a follow-up file and sends an information letter to the care partners with the contact number for health professionals (dedicated line).

- Non-primary suicidal patients are called by telephone between days 10 and 21 after their discharge from the hospital by the follow-up team.
- For all suicidal persons benefiting from the system, a telephone contact is scheduled at the end of the sixth month following discharge from hospital with a view to a clinical assessment and evaluation with a proposal for ending the monitoring. If necessary, the monitoring can be extended. For this call, the patient is informed 1 week before by a letter.
- In VigilanS, the suicidal person leaves the hospital with a resource card containing the toll-free number of the contact unit, which he or she can call at any time during opening hours. Many incoming calls are received; these are generally long calls from patients in need of help.
- The sending of personalized postcards; once a month for 4 months, can be decided by the contactor after any phone call and for unreachable patients.

The aim of this chapter is to:

1. identify the recidivism risk factors in the Nord-Pas-de-Calais region;
2. model the suicide recidivism by applying several frameworks of the probit regression model;
3. investigate the impact of non-linear explanatory variables (exogenous determinants) such as median revenue, unemployment rates, worker rates and graduation (high school graduates) rate.

The spatial autocorrelation factor as well as the non-linear explanatory variables are then incorporated to the partially linear spatial probit model (Ahmed et al., 2020) to improve the accuracy of predicting suicide recidivism in Nord-pas-de-Calais. This model can aid in developing strategies to combat suicide death in a population.

4.2 Data description

The data consists of information on 34000 cases of suicide attempts from January 2015 to May 2021 for over 20000 patients. Among the suicide attempts, 31.3% entries show reattempts (22.95% of the patients). The reattempt cases will be called as feature recidivism and treated as our target variable.

This 6-month follow up study collected data at the point of recruitment, 10 days after and 6 months after recruitment. The database includes the data of more than 23000 patients from different establishments, with information on sociodemographic characteristics such as date of birth (age), gender, address, and native country. This data set also contains information on alcohol consumption before the suicide attempt; history of suicide attempts; suicide methods (overdose

of drugs, drug poisoning (medicaments), drowning, wounding (lesions), hanging, voluntary drug intoxication (VDI), phlebotomy, jumping from a great height, combination of several methods and others); duration of hospitalisation; information concerning calls (answered and missed calls, and the number of these calls) that will be made on the 10th day as well as 6 months after the suicide attempt; and if there has been contact with the patient's family and friends or a professional.

The data at day 10 is based on the responses of the patients to the questionnaire on the 10th day. It contains variables such as the date of call, the duration of call, and the responses (the options are in parentheses) to the following questions (classified as categorical variables):

1. how does the patient feel? (Better, Worse, No changes)
2. does support from family, professional (psychiatrist) and social domain help to overcome the temptation to commit suicide? (Yes/ No)
3. is there psychological monitoring of patient? (Yes/ No)
4. does the patient need help? (Yes/ No)
5. is the patient suffering and needs help? (Suffering and needs help/ suffering but does not need help/ in difficulty/ no issues)
6. effect of postcard intervention (Helpful/ Not helpful)
7. available support (Friends, family, VigilanS professional, attending psychiatrist, doctor, Psychiatric Medical Centre (inpatient treatment), Hospital Emergency (outpatient treatment), Others)

On the other hand, the 6-month data is based on the patient's responses to the 6-month questionnaire. The variables are related to the date and duration of the call and the patient's responses (the options are in parentheses) to the following questions:

1. display of risky behaviour since the attempt (Yes/ No);
2. existence of suicidal thoughts since the attempt (Yes/ No);
3. continuous monitoring on patients (Yes/ No);
4. effect of postcard intervention (Helpful/ Not helpful);
5. the total number of lifetime suicide attempts.

REC denotes recidivism at the point of recruitment. A target variable called Recidivism after 6 months denoted by **REC6** was created for this data. **REC6** is a binary variable which tells if the patient attempted suicide after the 6-month-call. It was created using the following logic:

- for a patient with single entry: REC = No,
- for a patient with multiple entries: REC = Yes for all but the last entry No.

Then, based on the date of call made at the 6th month, the duration between two successive acts was calculated and classified as follows for the post 6-month study:

- if REC = Yes and duration is more than 6 months, REC6 = Yes,
- if REC = Yes and duration is lesser than 6 months, REC6 = No,
- if REC = No, then REC6 = No.

4.2.1 Data for exogenous determinants

Apart from the data collected by VigilanS, it was of interest to examine the impact of exogenous determinants such as the income, the manual worker rate and the level of education of the population in the Nord-Pas-de-Calais region. The data was retrieved from website `atlasante.fr`.

4.2.2 Predictors examined

The predictors examined at initial point of recruitment and 6 months after recruitment are given in the following table:

Table 4.1: Description of the variables.

Predictors	
Age	age of the patient.
Gender	male or female.
Total SA	the total number of suicide attempts.
Alcohol	alcohol consumption by patients before the suicide attempt.
Companion	Patient accompanied to hospital after the first suicide attempt.
VSA	Using violent means of suicide such as firearm, wounding, hanging, VDI, wrist-cutting.
Hospitalisation	Hospitalisation of patient after the suicide attempt.
Contact	Family and friends OR Professional OR both OR none (to attend calls from VigilanS).
Impact domain	Affected by family, social or professional surroundings - one of the factors mentioned above has an impact, - two of the factors mentioned above have an impact, - three of the factors mentioned above have an impact, - none of the factors mentioned above have an impact.
Follow up after 6 months	If the patient is continued to be monitored after 6 months.
Psychological monitoring D10	Follow-up with psychiatrist.
Native country	France, Outside France, No information.
VSA Yes: Hospitalisation No	combination of VSA and No Hospitalisation
Median income	Median of income in Euros.
Unemployment rate	Rate in % of share of unemployed in the 15-64 age group.
Worker rate	Rate in % of share of manual workers in the 15-64 age group.
Graduation rate	Rate in % of share of high school graduates.

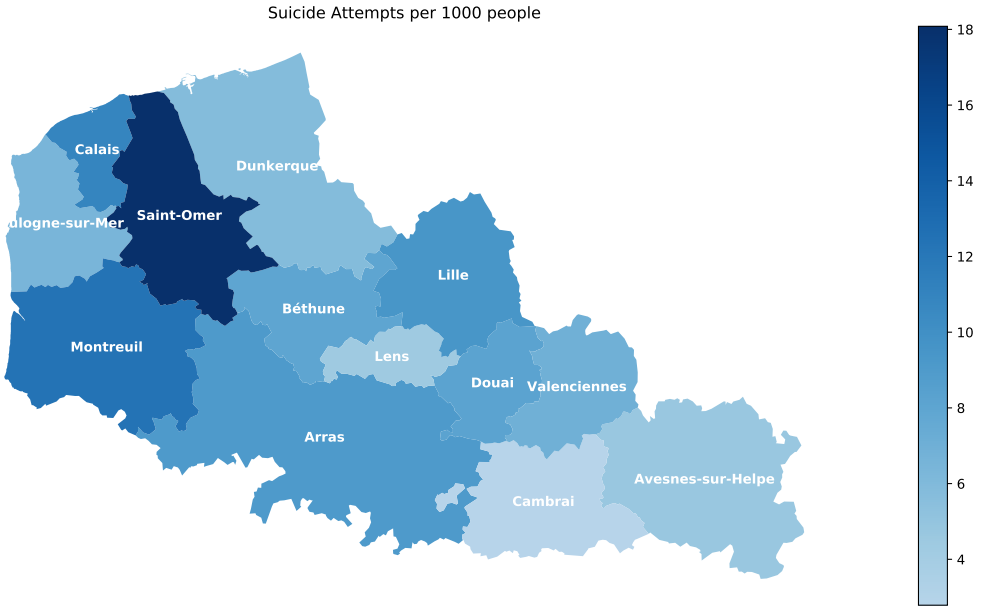


Figure 4.1: Map of suicide attempts per 1000 people

4.3 Methods

Modeling the suicide attempt by taking into account potential spatial dependency requires to find a correlation structure between data observed at a given location and that available at neighboring locations. In spatial econometrics, spatial dependency is usually modeled by using a spatial linear process defined by a spatial weight matrix which is a $n \times n$ non-stochastic weight matrix, W_n describing the spatial interactions between n spatial units. For a lattice data, the spatial autoregressive (SAR) dependent variable model and the spatial autoregressive error model (SAE).

Our goal is to model suicide recidivism for patients enrolled in the VigilanS programme. The response variable is a binary variable so we use the probit regression model. The integration space-dependent correlation to the model (if exists) further enhances the ability of the model by increasing its accuracy. Space-dependent correlation is detected when data observed between neighbouring locations form a correlation structure. The inclusion of the spatial index based on the location of the address reported for each patient lead to the use of the spatial probit model.

Suppose we have a sample of n observations collected from points in a 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 $d > 0$ for each other; i.e $\forall$ $s_i, s_j \in \mathcal{I}$: $\|s_i - s_j\| \geq d$. In this section, to facilitate the notation, we denote i for individual in location s_i. The variables Y_i are binary responses ($Y_i = 1$ correspond to recidivism while 0 is no recidivism). 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 two alternatives for each observation is based on a latent dependent variable Y_i^* via the following spatial autoregressive regression:

$$\begin{aligned} \mathbf{Y}_n^* &= \lambda_0 W_n \mathbf{Y}_n^* + \mathbf{X}_n \beta_0 + \varepsilon_n, \qquad \varepsilon_n \sim N(0, I_n), \\ Y_i &= \mathbb{I}\left(Y_i^* \geq 0\right), \qquad i = 1, \ldots, n. \end{aligned} \tag{4.1}$$

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 1. 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

$$V_n(\lambda_0) = \mathrm{Var}\left(\mathbf{Y}_n^* | \mathbf{X}_n\right) = (I_n - \lambda_0 W_n)^{-1} \left\{ (I_n - \lambda_0 W_n)' \right\}^{-1}.$$

The `ProbitSpatial` R package was used to provide the estimates of β_0 and λ_0 in the following where the results are based on the spatial weight matrix W_n with K-nearest neighbour (KNN) weights.

The SAR probit model will be compared with the classical binary probit model (does not consider spatial dependence). We will present the numerical results for the spatial probit SAR (model 4.1) and the following basic non-spatial binary probit model:

$$\begin{aligned} \mathbf{Y}_n^* &= \mathbf{X}_n \beta_0 + \varepsilon_n, \qquad & \varepsilon_n &\sim N(0, I_n), \\ Y_i &= \mathbb{I}\left(Y_i^* \geq 0\right), \qquad & i &= 1, \ldots, n. \end{aligned} \tag{4.2}$$

4.3.1 Partially linear spatial probit 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^*:

$$Y_{in}^* = X_{in}^T \beta_0 + g_0(Z_{in}) + U_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{4.3}$$

with

$$Y_{in} = \mathbb{I}\left(Y_{in}^* \geq 0\right), \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{4.4}$$

where $\mathbb{I}(\cdot)$ is the indicator function; X and Z are explanatory random variables taking values in the two compact 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 $\Theta_\beta \subset \mathbb{R}^p$; and $g_0(\cdot)$ 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 $\mathbb{R}$ and C a positive constant. In model (4.3), β_0 and $g_0(\cdot)$ are constant over i (and n). Assume that the disturbance term U_{in} in (4.4) is modelled by the following spatial autoregressive process (SAR):

$$U_{in} = \lambda_0 \sum_{j=1}^{n} W_{ijn} U_{jn} + \varepsilon_{in}, \qquad 1 \leq i \leq n,\ n = 1, 2, \ldots \tag{4.5}$$

where λ_0 is the autoregressive parameter, valued in the compact subset $\Theta_\lambda \subset \mathbb{R}$, $W_{ijn}, j = 1, \ldots, n$ are the elements in the i–th row of a non-stochastic $n \times n$ spatial weight matrix W_n.

4.4 Results and discussion

Bivariate analysis (Tables 4.8, 4.9, 4.10, 4.11) showed statistically significant differences in the potential predictors of suicide recidivism. P-value less than 0.05 indicates statistical significant difference in the category studied.

Females consist about 61% of the total patients enrolled with those in the adolescent phase as well as in their 20s contributing the highest number of cases. On the other hand, for males, patients in their 40s and 50s contribute to the highest number of suicide cases.

This study also shows that alcohol consumption prior to a suicide attempt contributes to suicide recidivism (more than 50% of the patients, Table 4.2). Patients aged 40 and above have higher tendency in consuming alcohol before the act of committing suicide (Figure 4.5). Alcohol consumption among males (60%) is higher than that of females (40%) based on findings.

Patients use various methods for attempting suicide where VDI was used by almost 80% of females and 70% of males. Males tend to resort to violent means of committing suicide such as hanging, using firearms, wounding, VDI and wrist cutting. Some patients even adopted more than one method of suicide. Firearm (86%), wounding (67%) and hanging (73%) methods are more common among men (Table 4.8). Women were more likely to use VDI (64%) and wrist cutting (55%) (Table 4.8).

Note that 62% of those who were recruited were accompanied by someone to enroll in the VigilanS programme (Tabel 4.2). Our data revealed that those that suicide recidivism was more common among those who were unaccompanied during their visit to enroll for the VigilanS programme.

The chi-squared test was performed to investigate the relationship between two qualitative variables. Statistical significance for the test was defined as $p<0.05$. Tabel 4.2 shows the predictors that have high association with recidivism. The chi-square test shows a strong dependence between recidivism and age with other essential variables such as alcohol consumption, which is a factor contributing to recidivism, and with other variables.

Table 4.2: Factors of suicide recidivism and methods used.

Variables	Total Sample N= 1364	recidivism N= 417	No recidivism N=947	t/χ^2	*df*	p-value
Gender				3.53		0.06
Male (%)	433 (31.74)	117 (28.06)	316 (33.37)			
Female (%)	931 (68.26%)	300 (71.94)	631 (66.63)			
Age (mean±SD)	42.58 ± 15.38	42.87 ± 13.35	42.45 ± 16.19	-0.51	953.42	0.6
Alcohol (Yes)	689 (50.51)	212 (50.83)	477 (50.37)	0.01	1	0.8
Companion (Yes)	984 (62.14)	275 (65.95)	709	11.02	1	**0.0008**
TotalSA	7.396 ± 13.59	9.66 ± 16.17	6.4 ± 12.15	-3.68	631.43	**0.00025**
Hospitalisation (Yes)	593	172	421	1.08	1	0.2
Methods				36.10	8	**1.68 e^{-5}**
Firearm	5	0	5			
Other ways	24	0	24			
VDI	1075	334	741			
Wounding	12	2	10			
Drowning	11	4	7			
Hanging	31	3	28			
Phlebo	112	52	60			
Several ways	82	19	63			
Jump	12	3	9			
Native country				30.4	2	**2.5e^{-7}**
Not France	32	11	21			
France	1047	357	690			
Not available	285	49	236			
Contact				30.11	3	**1.3e^{-6}**
Entourage	126	20	106			
Professional	320	121	199			
Ent/prof	259	97	162			
No contact	659	179	480			
Impact domain				8.88	3	**0.03**
One domain	535	179	356			
Two domains	209	59	150			
Three domains	40	5	35			
No impact	580	174	406			

4.4.1 Models to predict suicide recidivism

Models predicting suicide recidivism over 6 months were developed. To determine the strongest predictors of suicide recidivism in Northern France, the data of the patients at the point of recruitment was modelled using a binary regression model (see Appendix section 4.6) for the table of predictors examined). The best predictors of suicide recidivism at the point of recruitment are marked with asterisks (Table 4.12).

The probit regression model

The classical probit regression model (equation 4.2) predicting possible recidivism 6 months after the current attempt was validated.

$$\begin{aligned}\text{Residivism} \sim \beta_0 &+ \beta_1 \text{Age} + \beta_2 \text{Gender} + \beta_3 \text{TotalSA} + \beta_4 \text{Alcohol} + \beta_5 \text{Hospitalisation} + \beta_6 \text{VSA} \\ &+ \beta_7 \text{Companion} + \beta_8 \text{Native country} + \beta_9 \text{Contact} + \beta_{10} \text{Impact domains} \\ &+ \beta_{11} \text{Psychological monitoring D10} + \beta_{12} \text{Follow up 6m.}\end{aligned}$$

Table 4.3: Probit model.

Variable	Estimate	Std Error	z-value	p-value	
Intercept	-0.8353	0.1856	-4.5010	$6.75e^{-06}$	***
Age	0.0036	0.0025	1.4230	0.154642	
Gender(Male)	-0.0954	0.0849	-1.1240	0.261224	
TotalSA	0.0060	0.0028	2.1470	0.031813	*
Alcohol (Yes)	0.0337	0.0791	0.4270	0.669584	
Companion(Yes)	-0.2280	0.0851	-2.6800	0.007361	**
VSA (Yes)	-0.5960	0.1574	-3.7860	0.000153	***
Hospitalisation (No)	-0.1124	0.0867	-1.2960	0.194817	
Contact (Ent & Prof)	0.2663	0.1012	2.6320	0.008485	**
Contact (Entourage)	-0.2919	0.1521	-1.9190	0.055001	†
Contact (Professional)	0.1431	0.0944	1.5160	0.129458	
One impact domain	0.0939	0.0837	1.1210	0.262124	
Two impact domains	-0.0334	0.1135	-0.2940	0.768647	
Three impact domains	-0.5364	0.2716	-1.9750	0.048281	*
Follow up after 6 months (Yes)	0.3308	0.0908	3.6420	0.000271	***
Psychological monitoring D10 (Yes)	0.4353	0.0884	4.9270	$8.37e^{-07}$	***
Native country (Not.France)	-0.5562	0.1022	-5.4420	$5.28e^{-08}$	***
Native country (Not Available)	0.0712	0.2406	0.2960	0.767191	
VSA (Yes):Hospitalisation (NH)	0.8353	0.2002	4.1730	$3.01e^{-05}$	***
AUC	0.6954				
AIC	1569.2900				

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
The classification in parenthesis refer to the dummy for the respective predictor.

This model treats recidivism after 6 months as the response variable, which tells if the patient re-attempted suicide 6 months after the initial attempt or not. A p-value <0.05 indicates that the predictor variable is statistically significant to predict suicide recidivism based on this model.

This model resulted in an AUC of 69.54%. It was of interest to study if spatial effects are of significance to predicting recidivism.
This model treats Recidivism after 6 months as the response variable, which informs the probability of the re-attempted suicide after 6 months of the current attempt. Male patients as well as those who were accompanied have lower probability of recidivism after 6 months. Missing hospitalisation after a violent suicide attempt increases the risk of reattempting suicide after 6 months.
One of the most prominent risk factors for death by suicide is a history of previous suicide attempts as a substantial number of patients who attempt suicide ultimately die committing suicide. If VigilanS can establish contact with the patient's family member, the chances of recidivism after 6 months are lower. The influence of more than one surrounding domain such as family, professional and social reduces the probability of recidivism after 6 months.
The risk of suicide recidivism is lower for those who were accompanied by someone such as family or friends to the hospital as well as those who were hospitalised. The severity of the previous suicidal act such as a violent act instils fear in the patient to reattempt suicide. Those patients with strong family support tend not to be victims of suicide recidivism.
When contact with a professional i.e. psychiatrist or psychologist is established, it indicates that the patient is in need of great help. Family and people around (entourage) play a vital role in giving support to the patient apart from consulting a professional to reduce suicide reattempts. The duration of the VigilanS programme is 6 months and the decision of the clinician to prolong the duration indicates that the patients need more attention and require professional help. This also applies to the cases with follow-up after 10 days.

The spatial probit model, SAR

A Moran's I test was conducted on the residuals for the aforesaid model to investigate the existence of spatial autocorrelation based on spatial weight matrix created from the location coordinates of the address of each patient. Since the Moran's I test appeared to be significant with I = 0.1497 and p-value = 0.001, spatial effects were included to the probit regression model. The p-value of ρ suggests that spatial effects are significant in this model and the AUC increases to 71.25%.
The SAR and SEM models (see in the Appendix section 4.6 Tables 4.12 and 4.13) give similar results. The inclusion of spatial effects indicated by ρ (see Table 4.4) increased the AUC to 71.25%.

Table 4.4: Results for the SAR probit model

Predictors	Estimate	Std Error	p-value	
Intercept	-0.6567	12.3006	$4.53e^{-04}$	***
Age	0.0032	2.1130	0.1460	
Gender (Male)	-0.0621	0.9706	0.3245	
TotalSA	0.0054	2.9662	0.0850	†
Alcohol (Yes)	0.0981	0.4347	0.5097	
Companion (Yes)	-0.1626	4.5991	0.0320	*
VSA (Yes)	-0.4942	16.2366	$5.59e^{-05}$	***
Hospitalisation (No)	-0.1013	2.3747	0.1233	
Contact (Ent & Prof)	0.1926	5.4671	0.0194	*
Contact (Entourage)	-0.3487	3.5624	0.0591	†
Contact (Professional)	0.1081	5.0848	0.0241	*
One impact domain	0.0698	0.2995	0.5842	
Two impact domains	-0.0104	0.1127	0.7371	
Three impact domains	-0.3460	2.7501	0.0972	†
Follow up after 6 months (Yes)	0.2819	10.0414	0.0015	**
Psychological monitoring D10 (Yes)	0.4377	24.1588	$8.87e^{-07}$	***
Native country (Not Available)	-0.4843	25.6858	$4.02e^{-07}$	***
Native country (Not.France)	0.1162	0.0810	0.7759	
VSA (Yes): Hospitalisation (No)	0.6192	18.5235	$1.68e^{-07}$	***
ρ	0.4524	73.7243	$8.98e^{-18}$	***
AUC	0.7125			
AIC	1558.7180			

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4.5: The Moran's test

Variables	Moran's I
Median revenue	0.2048 ***
Unemployment rate	0.1598 ***
Worker rate	0.3028 ***
Graduation rate	0.3457 ***
1^{st} score PCA	0.2476 ***
1^{st} score positive (sPCA)	0.2048 ***

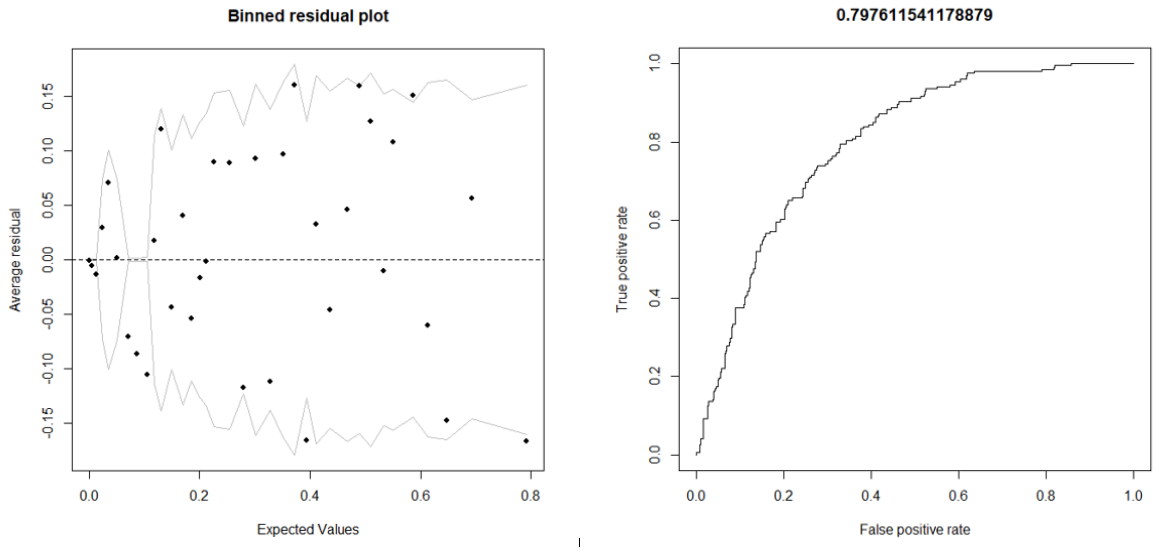
Note: *** $p < 0.001$

The partially linear spatial probit model (PLSPM)

In attempting to further enhance the performance of the prediction model, features with non-linear explanatory variables (exogenous determinants) were then added to the spatial probit regression model: median revenue (in Euros), unemployment rates, worker rates and graduation rates. We extract the smooth terms by fitting a GAM model (see Appendix Table 4.15) and use them as additive terms in the spatial probit model.

This study shows that significant predictors of suicide reattempts based on the best predictive model i.e the partially linear spatial probit model after a 6-month follow-up are gender, total number of suicide attempts by patient, if patient was accompanied to hospital, hospitalisation, contact with family, friends or professionals (psychiatrists), 10-day psychological monitoring, 3-impact domain (affected by family, social and professional surroundings), monitoring of patient for 6-months, patients with unidentified native countries, and finally the combination of violent methods of suicide attempt as well as hospitalisation.

The Moran's I test for each of these variables showed the existence of spatial autocorrelation (Table 4.5). The spatial component for this model is undoubtedly significant (p-value of ρ) resembling the earlier model. All the other predictors used are similar to the earlier model. This partially linear spatial probit model has an AUC of 79.76%. The AUC curve for this model is given in Figure 4.2a. A binned residual plot (Figure 4.2b) was also constructed for the partially linear spatial model to assess the fit of this model. Only a few points lie outside the confidence limits but no systematic pattern is detected in the plot. Therefore, this plot indicates that the partially linear spatial probit model is a good fit to the data compared to some of the investigated models from the probit regression model paradigm.



(a) (b)

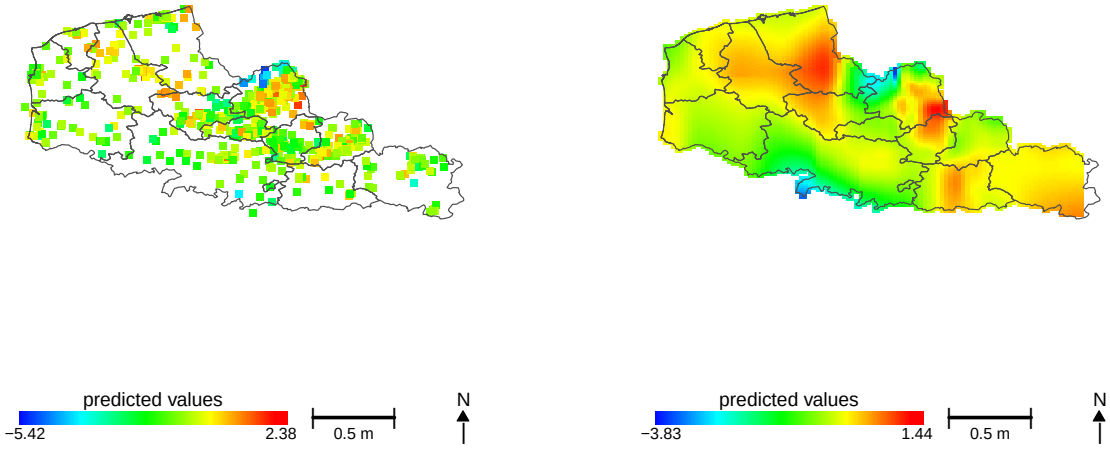
Figure 4.2: Binned and AUC plots of PLSPM

In addition to the results in Table 4.6, we provide the log odds ratio of the prediction of suicide attempts at sample locations (Figures 4.3a and 4.4a) and the prediction maps (Figures 4.3b and 4.4b) for all the Nord pas de calais region, based on the PLSPM model and GAM model. Figure 4.3b highlighted the communes (Saint Omer, Dunkerque, Lille and Douai) where the suicide attempts is high (red and yellow). This map seems to overestimate the suicide attempts compare to the prediction given by the best model (PLSPM) (Figure 4.4b).

Table 4.6: Partially linear spatial probit model

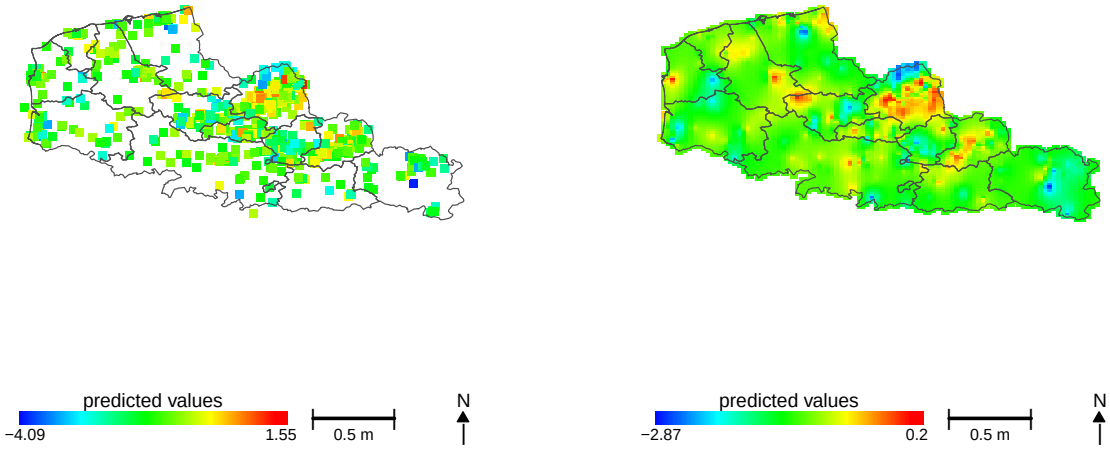
Predictors	Estimate	Std Error	p-value	
Linear Effects				
Intercept	-0.8513	27.35512	$1.69e^{-07}$	***
Gender (male)	-0.1712	5.34287	$2.08e^{-02}$	†
TotalSA	0.0069	4.78931	$2.86e^{-02}$	*
Companion(Yes)	-0.1569	1.43605	$2.31e^{-01}$	
VSA (Yes)	-0.8185	23.22975	$1.44e^{-06}$	***
Hospitalisation (No)	-0.2300	6.62259	$1.01e^{-02}$	*
Contact (Ent/Prof)	0.5579	19.94233	$7.98e^{-06}$	***
Contact (Entourage)	-0.1413	0.31537	$5.74e^{-01}$	
Contact (Professional)	0.2485	5.74343	$1.66e^{-02}$	*
Psychological monitoring D10 (Yes)	0.3926	13.63203	$2.22e^{-04}$	***
One impact domain	0.1668	1.58415	$2.08e^{-01}$	
Two impact domains	-0.0655	0.07427	$7.85e^{-01}$	
Three impact domains	-0.6247	6.25429	$1.24e^{-02}$	*
Follow up after 6 months (Yes)	0.3159	9.16583	$2.47e^{-03}$	**
Native country (Not Available)	-0.6148	19.81851	$8.52e^{-06}$	***
Native country (Not.France)	-0.0784	0.07734	$7.81e^{-01}$	
VSA (Yes): Hospitalisation (No)	1.2267	34.48394	$4.30e^{-09}$	***
ρ	0.3954	27.37676	$1.67e^{-07}$	***
Non-Linear Effects	edf	Chi.sq	p-value	
Median revenue	1.1178	48.3874	$3.50e^{-12}$	***
Unemployment rate	1.1348	57.3218	$3.70e^{-14}$	***
Worker rate	1.1405	46.2091	$1.06e^{-11}$	***
Graduation rate	1.1451	53.3024	$2.86e^{-13}$	***
Age	1.1549	46.4688	$9.31e^{-12}$	***
AUC	0.7976			
AIC	1386.848			

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$



(a) (b)

Figure 4.3: (a) Log of odd ratios of the prediction of suicide attempts at sample locations and (b) the prediction map in all the Nord-Pas-de-Calais region using the GAM model.



(a) (b)

Figure 4.4: (a) Log of odd ratios of the prediction of suicide attempts at sample locations and (b) the prediction map in all the Nord-Pas-de-Calais region using the PLSPM model.

PLSPM with deprivation index

We calculate the index of deprivation for each commune by extracting the first component of the principal component analysis (PCA) of 4 available variables namely, worker rate, graduation rate, unemployment rate and median revenue. We conduct spatial PCA and the first positive component is also used as a spatial index of deprivation. The Moran's test on these components is well significant (Table 4.5). From the GAM model (Table 4.16 in Appendix), we use this index as an additive term in this spatial probit model (Table 4.7).

Table 4.7: The PLSPM model with deprivation index.

Predictors	Estimate	Std Error	p-value
Linear Effects			
Intercept	-0.8201	21.4205	$3.69e^{-06}$
Gender (Male)	-0.1426	4.4011	$3.59e^{-02}$
TotalSA	0.0084	5.5623	$1.84e^{-02}$
Companion (Yes)	-0.1768	3.3318	$6.80e^{-02}$
VSA (Yes)	-0.6851	20.8754	$4.90e^{-06}$
Hospitalisation (Yes)	-0.1521	4.3604	$3.68e^{-02}$
Contact (Ent/Prof)	0.3968	12.0644	$5.14e^{-04}$
Contact (Entourage)	-0.1732	1.5308	$2.16e^{-01}$
Contact (Professional)	0.1919	3.4248	$6.42e^{-02}$
Psychological monitoring D10 (Yes)	0.4425	15.8963	$6.69e^{-05}$
One impact domain	0.1098	0.9976	$3.18e^{-01}$
Two impact domains	-0.1186	0.9021	$3.42e^{-01}$
Three impact domains	-0.6323	7.0471	$7.94e^{-03}$
Follow up after 6 months (Yes)	0.3794	11.8171	$5.87e^{-04}$
Native country (Not Available)	-0.5419	18.7576	$1.48e^{-05}$
Native country (Not France)	-0.0293	0.2395	$6.25e^{-01}$
VSA (Yes):Hospitalisation (No)	1.0426	30.8615	$2.77e^{-08}$
ρ	0.5027	54.4303	$1.61e^{-13}$
Non-Linear Effects	edf	Chi.sq	p-value
Age	1.1830	35.8953	$2.08e^{-09}$
spatial index deprivation	1.1555	39.5736	$3.16e^{-10}$
AUC	0.7533		

4.5 Conclusion

The application of the partially linear spatial probit model on the suicide recidivism data clearly showed that the inclusion of the spatial effect significantly improved the prediction ability of the model. This model also identified significant predictors of suicide recidivism as well as exogenous determinants which can contribute towards suicide reattempts. Multiple suicide attempters must be given more attention and the VigilanS programme is working towards significantly lowering suicide cases in Northern France. This statistical analysis can work hand in hand with the

VigilanS programme towards achieving its goal. Socio-familial isolation is considered a fairly important risk factor of suicide recidivism attempt. Moreover, a patient committing a violent act of suicide and was not hospitalised tends to repeat the act.

4.6 Appendix

Table 4.8: Chi-square analysis for gender from the original data.

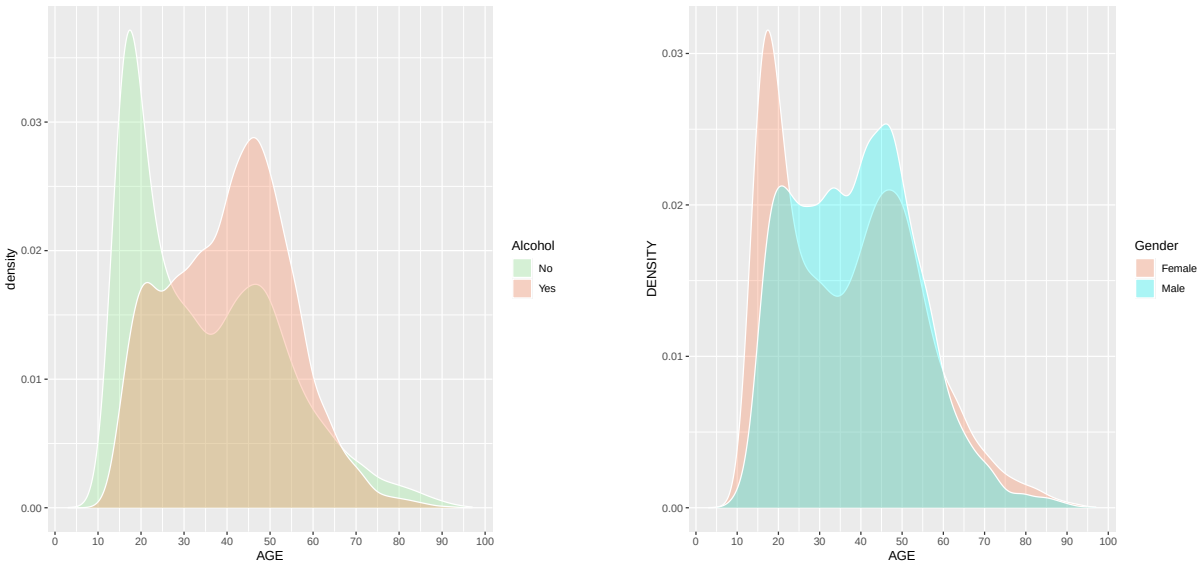
Predictors	P-value	Classes	Gender	
			Female	Male
Alcohol	$<2.2e^{-16}$	No	0.6864	0.3136
		Yes	0.5177	0.4823
Methods	$<2.2e^{-16}$	Firearm	0.1379	0.8621
		Other ways	0.4675	0.5325
		IDV	0.6454	0.3546
		Wounding	0.3223	0.6777
		Drowning	0.5556	0.4444
		Hanging	0.2648	0.7352
		Phlebo	0.5655	0.4345
		Several ways	0.5584	0.4416
		Jump	0.5031	0.4969
Companion	$<2.2e^{-16}$	No	0.5310	0.4690
		Yes	0.6332	0.3668
Age Groups	$<2.2e^{-16}$	≤ 24	0.7111	0.2889
		25 - 40	0.5322	0.4678
		40 - 54	0.5683	0.4317
		55 +	0.6275	0.3725

Table 4.9: Chi-square analysis for suicide recidivism from the original data.

Predictors	P-value	Classes	Recidive	
			Yes	No
Alcohol	$<2.2e^{-16}$	No	0.7885	0.2115
		Yes	0.7497	0.2502
Companion	$3.25e^{-13}$	No	0.2869	0.7131
		Yes	0.4072	0.5928

Table 4.10: Chi-square analysis (methods used to suicide vs alcohol).

Predictors	P-value	Classes	Alcohol	
			Yes	No
Methods	$6.40e^{-10}$	Firearm	0.6136	0.3864
		Other ways	0.5590	0.4410
		VDI	0.4613	0.5387
		Wounding	0.5289	0.4711
		Drowning	0.4000	0.6000
		Hanging	0.4962	0.5038
		Phlebo	0.4671	0.5329
		Several ways	0.4424	0.5576
		Jump	0.3834	0.6166



(a) (b)

Figure 4.5: Age vs alcohol and sex.

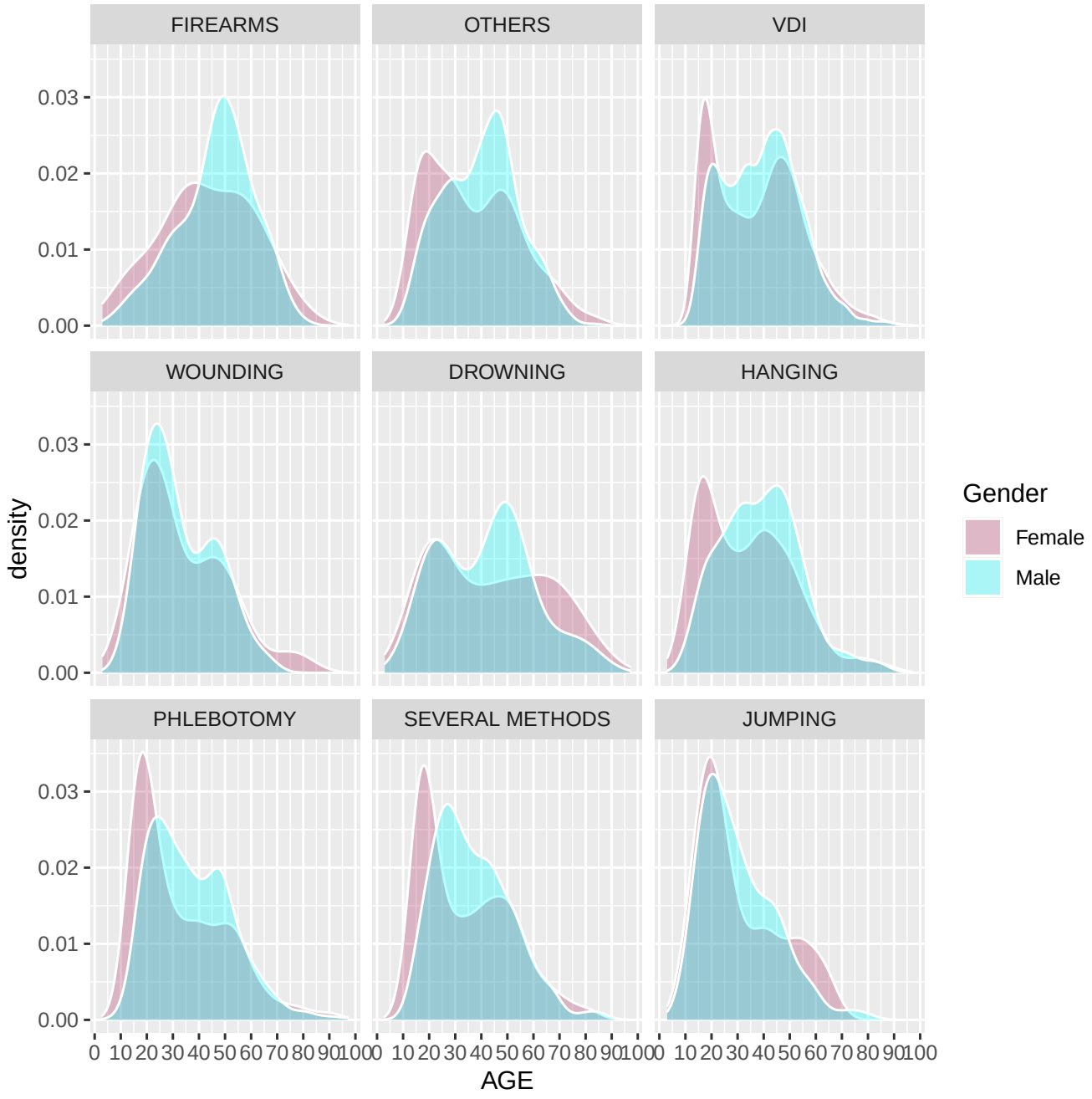


Figure 4.6: Age, gender and methods used in suicide.

Table 4.11: Chi-square analysis for age groups in the orginal data.

Predictors	P-value	Classes	Age groups			
			≤ 24	25-40	40-54	55+
Alcohol	$<2.2e^{-16}$	No	0.3804	0.2274	0.2433	0.1489
		Yes	0.1511	0.2803	0.3970	0.1716
Methods	$<2.2e^{-16}$	Firearm	0.0805	0.2299	0.3793	0.3103
		Other ways	0.2271	0.2811	0.3286	0.1632
		VDI	0.2664	0.2459	0.3248	0.1629
		Wounding	0.3306	0.3306	0.2397	0.0992
		Drowning	0.2870	0.1389	0.2778	0.2963
		Hanging	0.2352	0.2933	0.3324	0.1390
		Phlebo	0.3611	0.2680	0.2271	0.1438
		Several ways	0.3258	0.2851	0.2572	0.1318
		Jump	0.4497	0.2516	0.2013	0.0975
Companion	$<2.2e^{-16}$	No	0.1809	0.2760	0.3729	0.1702
		Yes	0.3043	0.2441	0.2957	0.1559
Recidivism	$<2.2e^{-16}$	No	0.2922	0.2536	0.2942	0.1600
		Yes	0.2477	0.2565	0.3724	0.1235
10 day call	$6.32e^{-11}$	No	0.2836	0.2604	0.3090	0.1469
		Yes	0.2433	0.2277	0.3327	0.1963
6 month call	$8.383e^{-08}$	No	0.2839	0.2584	0.3076	0.1501
		Yes	0.2092	0.2135	0.3605	0.2168

Table 4.12: Model for suicide recidivism based on the original data.

Variable	Estimate	Std Error	z-value	p-value	
Intercept	-0.1693	0.0294	-5.7500	$8.91e^{-09}$	***
Age	-0.0019	0.0005	-3.9270	$8.59e^{-05}$	***
Primo (Yes)	-0.5693	0.0189	-30.0620	$<2e^{-16}$	***
Alcohol (Yes)	0.0948	0.0153	6.1790	$6.45e^{-10}$	***
Companion (Yes)	-0.1262	0.0163	-7.7370	$1.02e^{-14}$	***
VSA (Yes)	-0.1605	0.0196	-8.2040	$2.32e^{-16}$	***
Hospitalisation (No)	-0.0217	0.0156	-1.3910	0.1640	
Native.country (Not.France)	-0.0678	0.0415	-1.6340	0.1020	
Native.country (Not Available)	-0.5122	0.0188	-27.3040	$<2e^{-16}$	***
Days since Prev. SA (>1 Year)	-0.0240	0.0272	-0.8840	0.3770	
Days since Prev.SA (1Year)	0.3277	0.0300	10.9060	$<2e^{-16}$	***
Days since.Prev.SA (6Months)	0.5710	0.0215	26.6170	$<2e^{-16}$	***
AUC	0.7235				
AIC	37927.71				

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4.13: The SEM model after 6 months.

Variable	Estimate	Std Error	p-value	
Intercept	-0.897245842	16.14462191	$5.87e^{-05}$	***
Age	0.003935661	1.40424956	$2.36e^{-01}$	
Gender(male)	-0.096120316	0.2099344	$6.47e^{-01}$	
TotalSA	0.006516742	2.01090302	$1.56e^{-01}$	
Alcohol(Yes)	0.037510691	0.07331228	$7.87e^{-01}$	
Companion(Yes)	-0.241803841	7.21854051	$7.22e^{-03}$	***
TS.Violent(Yes)	-0.632613562	18.48920955	$1.71e^{-05}$	***
Hospitalisation(No)	-0.117829442	2.34092106	$1.26e^{-01}$	
Contact (Entourage & Professional)	0.288350319	9.02851656	$2.66e^{-03}$	**
Contact (Entourage)	-0.310575309	2.61167453	$1.06e^{-01}$	
Contact(Professional)	0.151338502	3.17268511	$7.49e^{-02}$	†
Impact.Domains (1)	0.094252792	0.70474297	$4.01e^{-01}$	
Impact.Domains(2)	-0.044340637	0.26630705	$6.06e^{-01}$	
Impact.Domains (3)	-0.561442638	4.93141408	$2.64e^{-02}$	*
Follow up after 6 months (Yes)	0.35109305	9.28905535	$2.31e^{-03}$	**
Psychological monitoring D10 (Yes)	0.460645284	17.51353636	$2.85e^{-05}$	***
Native.region(Not Available)	-0.597281804	24.50062316	$7.43e^{-07}$	***
Native.country(Not.France)	0.082105223	0.10788625	$7.43e^{-01}$	
TS.Violent(Yes):Hospitalised(NH)	0.889339097	24.20113291	$8.68e^{-07}$	***
rho	0.528275364	76.80227304	$1.89e^{-18}$	***
AUC	0.6955			

Note: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4.14: The variables used in the 10-day and 6-month questionnaires.

Predictors	Classes	Propotions (%)
Evolution of initial discomfort	unfavorable	6
	favorable	66
	no change	28
Impact of suicid attempt	No	48
	Yes	52
Exit compromise	monitoring	72
	not followed	19
	no news	9
Psychological Monitoring in Progress	No	29
	Yes	71
Need of help	No	30
	Yes	70
Type of interview	crise avec ES	1
	crise sans ES	4
	in difficulty	57
	TVB	38
Sending postcard	No	51
	Yes	49
Ressources indentified	Friends	6
	Others	5
	CMP	10
	Family	13
	Treating Medcin	16
	Number Vigilans	31
	Attending psychiatrist	9
	Emergency	10
Recdivism before 6 month	No	78
	Yes	16
	NA	7
Recdivism after 6 month	No	94
	Yes	6
suicide risk	No	73
	Yes	21
	NA	6
suicidal thoughts	No	73
	Yes	20
	NA	7
continued monitoring	No	80
	Yes	20
send postcard	No	77
	Yes	23

Table 4.15: The results of the GAM model (AUC=79.68%, AIC=1470.28).

Predictors	Estimate	Std Error	p-value	
Intercept	-0.7983	0.1511	$1.26e^{-07}$	***
Gender(Male)	-0.1615	0.0952	$8.98e^{-02}$	†
TotalSA	0.0065	0.0030	$3.12e^{-02}$	*
Companion (Yes)	-0.1579	0.0927	$8.84e^{-02}$	†
VSA (Yes)	-0.7792	0.1820	$1.85e^{-05}$	***
Hospitalised (No)	-0.2144	0.0967	$2.66e^{-02}$	*
Contact(Ent & Profes)	0.5100	0.1151	$9.40e^{-06}$	***
Contact (Entourage)	-0.1525	0.1671	$3.61e^{-01}$	
Contact (Professional)	0.2304	0.1045	$2.74e^{-02}$	*
Psychological monitoring D10 (Yes)	0.3812	0.0980	$9.96e^{-05}$	***
One Impact domain	0.1562	0.0946	$9.86e^{-02}$	†
Two Impact domains	-0.0608	0.1273	$6.33e^{-01}$	
Three Impact domains	-0.6051	0.2886	$3.60e^{-02}$	*
Follow up after 6 months (Yes)	0.3079	0.1012	$2.34e^{-03}$	**
Native region (Not Available)	-0.5815	0.1219	$1.83e^{-06}$	***
Native region (Not.France)	-0.0699	0.2639	$7.91e^{-01}$	
VSA (Yes):Hospitalised (No)	1.1571	0.2274	$3.61e^{-07}$	***
Smooth Terms	edf	Chi.SQ	p-value	
s(Age)	3.3860	40.8800	$<2e^{-16}$	***
s(Median revenue)	8.9230	29.5500	$5.42e^{-04}$	***
s(Graduation rate)	12.8090	23.6400	$4.88e^{-02}$	*
s(Worker rate)	14.1820	20.4800	$1.72e^{-01}$	
s(Unemployment rate)	6.6330	23.6000	$1.95e^{-03}$	**

Table 4.16: The results of the GAM model with depreviation index (AUC=73.24%, AIC=1524.27).

Predictors	Estimate	Std Error	p-value	
Intercept	-0.71132	0.139536	$3.44e^{-07}$	***
Gender (Male)	-0.15403	0.085915	0.072998	+
TotalSA	0.007136	0.002848	0.012222	*
Companion (Yes)	-0.19466	0.086297	0.024091	*
VSA (Yes)	-0.61327	0.161846	0.000151	***
Hospitalised (No)	-0.15422	0.089068	0.083373	+
Contact (Ent & Prof) l	0.333436	0.104012	0.001347	**
Contact (Entourage)	-0.17706	0.156075	0.256613	
Contact (Professional)	0.186255	0.095993	0.052344	+
Psychological monitoring D10 (Yes)	0.401598	0.089802	$7.75e^{-06}$	***
One impact domain	0.13121	0.084663	0.121191	
Two impact domains	-0.06951	0.11508	0.545826	
Three impact domains	-0.56176	0.273615	0.040062	*
Follow up after 6 months (Yes)	0.328319	0.092329	0.000377	***
Native region (Not Available)	-0.52811	0.106015	$6.31e^{-07}$	***
Native region (Not.France)	0.019803	0.247085	0.936119	
VSA (Yes):Hospitalised (No)	0.957272	0.204028	$2.71e^{-06}$	***
Smooth Terms	edf	Chi.sq	p-value	
s(depreviation index)	3.77	16.8	0.00316	**
s(Age)	3.383	39.94	$<2e^{-16}$	***

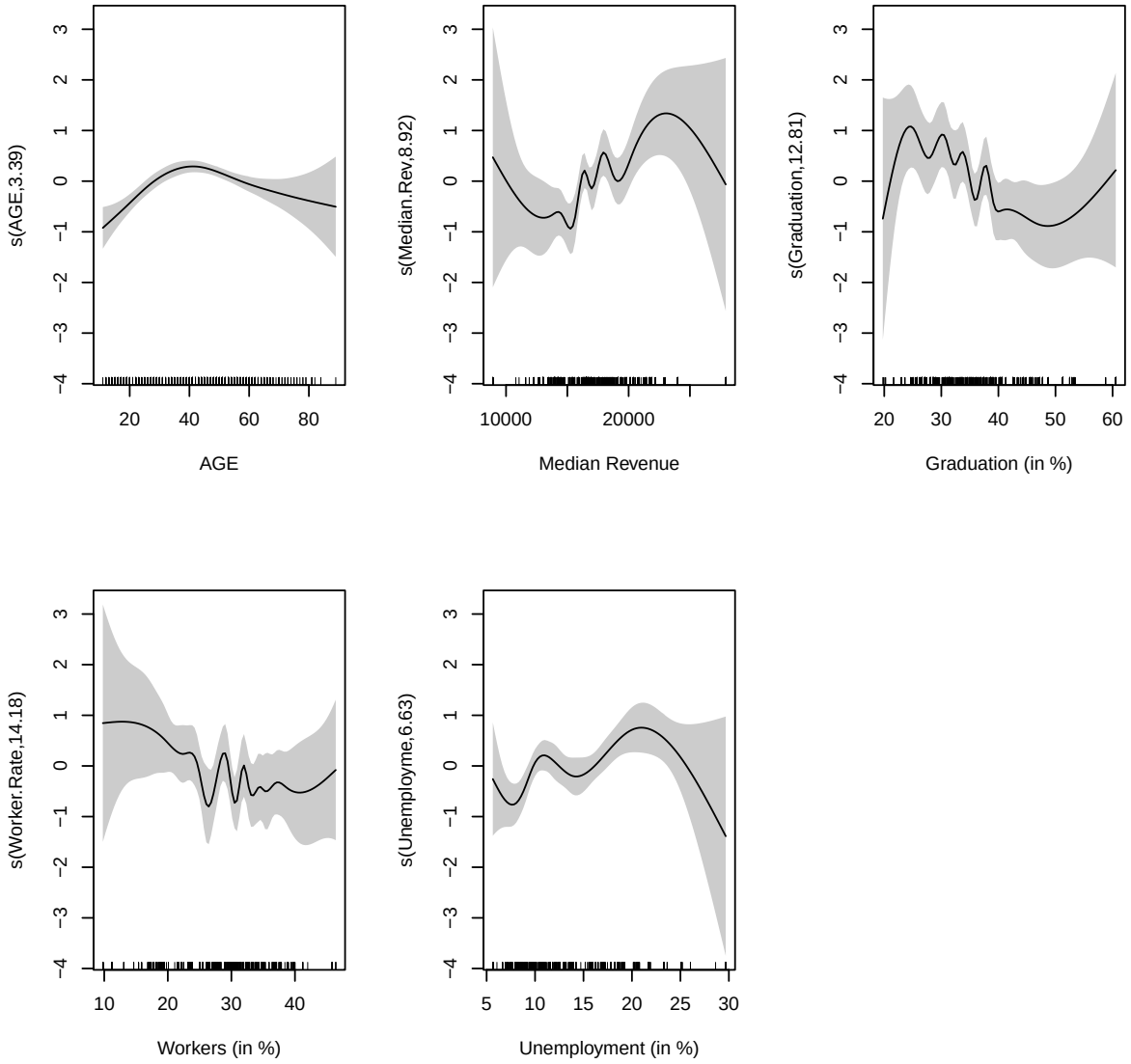


Figure 4.7: Smoothed terms

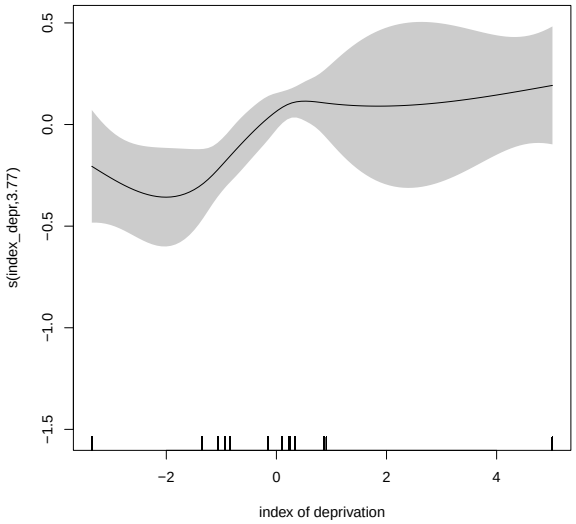


Figure 4.8: Smoothed deprivation index.

Chapter 5

Concluding remarks and future research

5.1 Concluding remarks

The main theme of this thesis is spatial analysis. Spatial analysis can identify and solve complex location-oriented problems. Spatial analysis provides important insights in analysing contents of a map where characteristics of places and the relationships between these places are investigated. This thesis is motivated by two real population health problems such as mortality modelling and the mapping of suicidal relapses. The first application deals with data of functional and spatial nature. Functional Data Analysis (FDA) is dedicated to analyze this kind of data.

FDA is an approach in statistics which encompasses the statistical methodology of data expressed in the form of functions. Data represented in the form of functions are then used in statistical modelling and prediction information can be retrieved from such data with the aid of some multivariable statistical concepts. In chapter 3 of this thesis, the study of the use of spatial statistics on functional data in demography was done. Mortality was investigated from a spatial perspective where the log of death rates data of 28 European countries were converted to functional data before performing spatial analysis on these data. The aim was to construct a tool to detect spatial autocorrelation in functional data. This led to the discovery of the functional Moran's I statistic. This statistic has the ability and potential to be used to detect spatial relationships involving functional data in various settings besides demography.

The second part of this thesis focuses on spatial modelling by considering a partially linear probit model for spatially dependent data. A deep dive on semi-parametric binary models from the theoretical perspective was conducted in Chapter 4. The combination of the generalized method of moments approach as well as the weighted likelihood method led to a semi-parametric estimation method. A spatial autoregressive error (SAE) model when the spatial dependence structure is integrated in a disturbance term of the studied model was addressed. This model was then applied to the suicide recidivism data collected by the VigilanS healthcare system in France. It was interesting to observe the existence of spatial autocorrelation of suicide recidivism cases in Nord Pas de Calais. The partially linear spatial autocorrelation model proved to be a good fit in modelling suicide recidivism by considering spatial autocorrelation and partially linear functions which were obtained based on potential exogenous determinants of suicide such as median income, graduation rates, unemployment rates and worker rates of populations in the

locations studied. Hence, spatial-dependent correlation can be integrated with models from a regression framework. This opens the path to solving complex space-related issues where no separate spatial analysis is required to investigate and predict the importance of spatial auto-correlation since this model is versatile in handling this issue.

5.2 Future research

Future work in the field of demography can shed some light on constructing spatial predictive models to predict mortality rates for neighbouring countries with limited or no data. This idea can further be extended to the spatio-temporal framework where mortality of a certain location can be predicted at a certain time.

The second part of the thesis and its application in investigating suicide recidivism, the choice of the instrumental variables for a more efficient regression estimator is a hot point in particular for real data.
A future work related to the project entitled: "Impact of the COVID-19 pandemic context on suicidal behaviors and their management", proposes to study the impact of the pandemic context on the incidence of suicide attempts and deaths by suicide and on suicidal ideation (based on content published on social networks) during the different periods that marked the pandemic (over a period of 3 years beginning 1 year before the pandemic).
A population which is vulnerable to suicidal behaviour i.e. that with high number of suicide history will be investigated. Finally, as the pandemic context is likely to influence the management of suicidal behaviour, a qualitative exploration of the difficulties encountered will be carried out. The consequences of the COVID-19 pandemic and its associated measures on the prevalence of psychiatric troubles but also on the key factors associated with suicide (e.g. social isolation, precariousness) as well as the reduction of access to psychiatric care during the health crisis indicate the fear of increased suicidal behaviors which has not been evaluated in France. The objectives of the project include:

- quantitative investigation of the impact of the Covid19 pandemic and health measures on suicidal behaviours (suicide attempts and deaths by suicide) in the general population, and in a particularly vulnerable population (people with a history of suicide attempts);
- extensively study the content posted on social networks to examine the impact of the pandemic and health measures on suicidal thoughts;
- qualitatively analyse the difficulties of dealing with attempted suicides in the pandemic context.

The first part of the analysis will concern VigilanS activity data for the population of four French regions with sufficient history to participate in the study: Nord-Pas-de-Calais, Normandy, Brittany and Languedoc-Roussillon. The VigilanS activity data will be studied over a 3-year period: from March 17, 2019 (one year before the first lockdown) to March 17, 2022. In second phase, the analyses will focus on the subgroup of people with a history of suicide attempt (cohort of patients included in the VigilanS March 17, 2019, until March 17, 2022). A descriptive phase will allow us to map suicide attempts on one hand and suicide mortality on the other hand, on the

French territory and in accordance to time. This will allow us to study the spatial changes in the incidence of suicide attempts and deaths by suicide before, during and after the pandemic period.

In order to study the impact of territorial factors related to the pandemic context by taking into account the proven risk factors of suicidal behaviors, we will employ recurrent multivariate models (by taking geography into account), multilevel models and spatial econometric and statistical specifications to map ecological risks. Moreover, joint modeling of individual and ecological risk factors avoid problems of ecological and individual inference and thus substantially reduce uncertainty about the interpretation of the respective role of each level of risk factors.

List of Figures

List of Tables

Bibliography

Ahmed, M.-S., Dabo-Niang, S., Genin, M., & Hassan, A. A. (2020). Partially Linear Spatial Probit Models. *Annales de l'ISUP*. https://hal.inria.fr/hal-03133818

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, *8*(02), 241–257.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in regional science*, *89*(1), 3–25.

Anselin, L. (2013). *Spatial econometrics: Methods and models* (Vol. 4). Springer Science & Business Media.

Arbia, G. (2006). *Spatial econometrics: Statistical foundations and applications to regional convergence*. Springer Science & Business Media.

Ash, R., & Gardner, M. F. (1975). Topics in stochastic processes, acad. *Press, New York*.

Beautrais, A. L., Gibb, S. J., Faulkner, A., Fergusson, D. M., & Mulder, R. T. (2010). Postcard intervention for repeat self-harm: Randomised controlled trial. *The British Journal of Psychiatry*, *197*(1), 55–60.

Becker, R. A., version by Ray Brownrigg. Enhancements by Thomas P Minka, A. R. W. R., & Deckmyn., A. (2018). *Maps: Draw geographical maps* [R package version 3.3.0]. https://CRAN.R-project.org/package=maps

Bernstein, S. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, *97*(1), 1–59.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 192–225.

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*(2, 3), 131–154.

Bivand, R., Keitt, T., & Rowlingson, B. (2019). *Rgdal: Bindings for the 'geospatial' data abstraction library* [R package version 1.4-8]. https://CRAN.R-project.org/package=rgdal

Bivand, R., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, *27*(3), 716–748. https://doi.org/10.1007/s11749-018-0599-x

Booth, H., Maindonald, J., & Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population studies*, *56*(3), 325–336.

Bosq, D. (2000). *Linear processes in function spaces* (Vol. 149). Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-1154-9

Bougeard, S., & Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software*, *86*(1), 1–17. https://doi.org/10.18637/jss.v086.i01

Brent, D. A., Bridge, J., Johnson, B. A., & Connolly, J. (1996). Suicidal behavior runs in families: A controlled family study of adolescent suicide victims. *Archives of general psychiatry*, *53*(12), 1145–1152.

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*(10), 1185–1199. https://doi.org/10.1080/0266476032000107187

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

Carter, G. L., Clover, K., Whyte, I. M., Dawson, A. H., & Este, C. D. (2005). Postcards from the edge project: Randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. *Bmj*, *331*(7520), 805.

Case, A. (1993). Spatial patterns in household demand. *Econometrica*, *52*, 285–307.

Cedereke, M., Monti, K., & Öjehagen, A. (2002). Telephone contact with patients in the year after a suicide attempt: Does it affect treatment attendance and outcome? a randomised controlled study. *European Psychiatry*, *17*(2), 82–91.

Chessel, D., Dufour, A.-B., & Thioulouse, J. (2004). The ade4 package – I: One-table methods. *R News*, *4*(1), 5–10. https://cran.r-project.org/doc/Rnews/

Cliff, A., & Ord, K. (1973). Spatial autocorrelation. *London: Pion Ltd.*

Cliff, A., & Ord, K. (1981). *Spatial processes: Models & applications*. Taylor & Francis.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, *92*(1), 1–45.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data. *J-01152 (MTM2006-09920)*.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2019). *Adespatial: Multivariate multiscale spatial analysis* [R package version 0.3-7]. https://CRAN.R-project.org/package=adespatial

Dray, S., & Dufour, A.-.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04

Dray, S., Dufour, A.-B., & Chessel, D. (2007). The ade4 package – II: Two-table and K-table methods. *R News*, *7*(2), 47–52. https://cran.r-project.org/doc/Rnews/

Eckardt, M., & Mateu, J. (2020). Partial and semi-partial statistics of spatial associations for multivariate areal data. *Geographical Analysis*.

Evans, J., Evans, M., Morgan, H. G., Hayward, A., & Gunnell, D. (2005). Crisis card following self-harm: 12-month follow-up of a randomised controlled trial. *The British Journal of Psychiatry*, *187*(2), 186–187.

Exbrayat, S., Coudrot, C., Gourdon, X., Gay, A., Sevos, J., Pellet, J., Trombert-Paviot, B., & Massoubre, C. (2017). Effect of telephone follow-up on repeated suicide attempt in patients discharged from an emergency psychiatry department: A controlled study. *BMC psychiatry*, *17*(1), 1–6.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66* (Vol. 66). CRC Press.

Fang, L., & Härdle, W. K. (2015). Stochastic population analysis: A functional data approach. *Available at SSRN 2630301*.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis* [Theory and practice]. Springer, New York.

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. *Advances in spatial econometrics* (pp. 145–168). Springer.

Gaetan, C., & Guyon, X. (2010). *Spatial statistics and modeling* (Vol. 90). Springer.

Garthoff, R., & Otto, P. (2017). Control charts for multivariate spatial autoregressive models. *AStA Adv. Stat. Anal.*, *101*(1), 67–94. https://doi.org/10.1007/s10182-016-0276-x

Greco, F., & Scalone, F. (2013). A space-time extension of the lee-carter model in a hierarchical bayesian.

Gromenko, O., Kokoszka, P., Zhu, L., & Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Annals of Applied Statistics - ANN APPL STAT*.

Gromenko, O., Kokoszka, P., Zhu, L., & Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*.

Guillaume, V., Francois, D., Philippe, M., Daniel, M., Alain, P., Christian, L., & Michel, G. (2006). Effect of telephone contact on further suicide attempts in patients discharged from an emergency department: Randomized controlled study. *BMJ*, *332*, 1241.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC Press.

Hörmann, S., & Kokoszka, P. (2011). Consistency of the mean and the principal components of spatially distributed functional data. *Bernoulli*.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer, New York. https://doi.org/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.

Hulten, A., Jiang, G.-X., Wasserman, D., Hawton, K., Hjelmeland, H., De Leo, D., Ostamo, A., Salander-Renberg, E., & Schmidtke, A. (2001). Repetition of attempted

suicide among teenagers in europe: Frequency, timing and risk factors. *European Child & Adolescent Psychiatry*, *10*(3), 161–169.

Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *Journal of the American Statistical Association*, *89*(428), 1354–1365.

Hyndman, R., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, *24*(3), 323–342.

Hyndman, R., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography*, *50*(1), 261–283.

Hyndman, R., & Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*.

Illian, J., Benson, E., Crawford, J., & Staines, H. (2006). Principal component analysis for spatial point processes — assessing the appropriateness of the approach in an ecological context. https://doi.org/10.1007/0-387-31144-0_7

Jarek, S. (2012). *Mvnormtest: Normality test for multivariate variables* [R package version 0.1-9]. https://CRAN.R-project.org/package=mvnormtest

Jombart, T. (2008). Adegenet: A r package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., Dufour, A.-B., & Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*.

Kauermann, G., Haupt, H., & Kaufmann, N. (2012). A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. *Statistical Modelling*, *12*(5), 419–440.

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*(1), 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*(2), 509–533. https://doi.org/10.1111/1468-2354.00027

Kuenzer, T., Hörmann, S., & Kokoszka, P. (2020). Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*.

Lee, & Carter, L. R. (1992). Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*(6), 1899–1925. https://doi.org/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*(2), 489–514. https://doi.org/10.1016/j.jeconom.2005.10.004

Lee, & Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, *38*(4), 537–549.

Léger, A.-E., & Mazzuco, S. (2021). What can we learn from the functional clustering of mortality data? an application to the human mortality database. *European Journal of Population*, 1–30.

Lesage, J. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, *32*(1), 19–35.

Li, Y., & Guan, Y. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2014.885434

Lin, X., & Lee, L.-f. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *J. Econometrics*, *157*(1), 34–52. https://doi.org/10.1016/j.jeconom.2009.10.035

Liu, C., Ray, S., & Hooker, G. (2014). Functional principal components analysis of spatially correlated data. *Statistics and Computing*.

Malikov, E., & Sun, Y. (2017). Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *J. Econometrics*, *199*(1), 12–34. https://doi.org/10.1016/j.jeconom.2017.02.005

Martinetti, D., & Geniaux, G. (2016). Probitspatial: Probit with spatial dependence, sar and sem models. *CRAN*. https://CRAN.R-project.org/package=ProbitSpatial

Martinetti, D., & Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, *64*, 30–45.

Mátyás, L. (1999). *Generalized method of moments estimation* (Vol. 5). Cambridge University Press.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, *32*(3), 335–348.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B (Methodological)*, *10*(2), 243–251.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17–23.

Motto, J. A., & Bostrom, A. G. (2001). A randomized controlled trial of postcrisis suicide prevention. *Psychiatric services*, *52*(6), 828–833.

Pinkse, J., Shen, L., & Slade, M. (2007). A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics*, *140*(1), 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*(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*(1), 125–154.

Poirier, D. J., & Ruud, P. A. (1988). Probit with dependent observations. *The Review of Economic Studies*, *55*(4), 593–614.

Ramsay, J. O., Graves, S., & Hooker, G. (2020). *Fda: Functional data analysis* [R package version 5.1.5.1]. https://CRAN.R-project.org/package=fda

Ramsay, J., Hooker, G., & Graves, S. (2009). Functional data analysis with r and matlab: Springer science & business media.

Ramsay, J., & Silverman, B. (2005). *Functional data analysis* (2ed). Springer.

Raymer, J., Willekens, F., & Rogers, A. (2018). Spatial demography: A unifying core and agenda for further research. *Population, Space and Place*. https://doi.org/10.1002/psp.2179

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and economics*, *38*(3), 556–570.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, *165*(1), 5–19.

Rue, H., & Held, L. (2005). *Gaussian markov random fields: Theory and applications*. CRC press.

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*(426), 501–511.

Severini, T. A., & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of statistics*, 1768–1802.

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional science and urban economics*, *40*(5), 292–298.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association*, *84*(405), 276–283.

Su, L. (2012). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, *167*(2), 543–560.

Vaiva, G., Walter, M., Al Arab, A. S., Courtet, P., Bellivier, F., Demarty, A. L., Duhem, S., Ducrocq, F., Goldstein, P., & Libersa, C. (2011). Algos: The development of a randomized controlled trial testing a case management algorithm designed to reduce suicide risk among suicide attempters. *BMC psychiatry*, *11*(1), 1–7.

Wang, H., Iglesias, E. M., & Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, *172*(1), 77–89.

Wang, S., & Ren, Z. (2019). Spatial variations and macroeconomic determinants of life expectancy and mortality rate in china: A county-level study based on spatial analysis models. *International Journal of Public Health*. https://doi.org/10.1007/s00038-019-01251-y

Wiśniowski, A., Smith, P. W., Bijak, J., Raymer, J., & Forster, J. J. (2015). Bayesian population forecasting: Extending the lee-carter method. *Demography*, *52*(3), 1035–1059.

Yang, K., & Lee, L.-f. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *J. Econometrics*, *196*(1), 196–214. https://doi.org/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*(1), 29–56. https://doi.org/10.1007/s13171-012-0009-5