UNIVERSITÉ DE LILLE, FRANCE.

École Doctorale SESAM

THÈSE DE DOCTORAT

Mention : Mathématiques et Applications

présentée par

Zied GHARBI

Contribution à l'économétrie spatiale et l'analyse de données fonctionnelles

---

Contribution to spatial econometric and functional data analysis.

Dirigée par Sophie DABO-NIANG

Co-dirigée par Laurence BROZE

Unité de recherche, LEM UMR 9221

Soutenue le 24 JUIN 2019

Le jury est composé de :

Mme Sophie DABO-NIANG	Université de Lille	Directrice
Mme Laurence BROZE	Université de Lille	Co-directrice
M. Ahmed EL GHINI	Université Mohammed V	Rapporteur
M. Abderrahim TAAMOUTI	Université de Durham	Rapporteur
M. Philippe BOUSSEMART	Université de Lille	Examinateur
Mme Raja CHAKIR	UMR Économie Publique INRA-AgroParisTech	Examinatrice

Résumé

Contribution à l'économétrie spatiale et l'analyse de données fonctionnelles

Ce mémoire de thèse touche deux champs de recherche importants en statistique inférentielle, notamment l'économétrie spatiale et l'analyse de données fonctionnelles. Plus précisément, nous nous sommes intéressés à l'analyse de données réelles spatiales ou spatio-fonctionnelles en étendant certaines méthodes inférentielles pour prendre en compte une éventuelle dépendance spatiale.

Nous avons d'abord considéré l'estimation d'un modèle autorégressif spatiale (SAR) ayant une variable dépendante fonctionnelle et une variable réponse réelle à l'aide d'observations sur une unité géographique donnée. Il s'agit d'un modèle de régression avec la spécificité que chaque observation de la variable indépendante collectée dans un emplacement géographique dépend d'observations de la même variable dans des emplacements voisines. Cette relation entre voisins est généralement mesurée par une matrice carrée nommée matrice de pondération spatiale et qui mesure l'effet d'interaction entre les unités spatiales voisines. Cette matrice est supposée exogène c'est-à-dire la métrique utilisée pour la construire ne dépend pas de mesure de la variable à expliquer du modèle. L'apport de cette thèse sur ce modèle réside dans le fait que la variable explicative est de nature fonctionnelle, à valeurs dans un espace de dimension infinie. Notre méthodologie d'estimation est basée sur une réduction de la dimension de la variable explicative fonctionnelle, par l'analyse en composantes principales fonctionnelles suivie d'une maximisation de la vraisemblance tronquée du modèle. Des propriétés asymptotiques des estimateurs, des illustrations des performances des estimateurs via une étude de Monte Carlo et une application à des données réelles environnementales ont été considérées.

Dans la deuxième contribution, nous reprenons le modèle SAR fonctionnel étudié dans la première partie en considérant une structure endogène de la matrice de pondération spatiale. Au lieu de se baser sur un critère géographique pour calculer les dépendances entre localisations voisines, nous calculons ces dernières via un processus endogène, c'est-à-dire qui dépend des variables à expliquer. Nous appliquons la même approche d'estimation à deux étapes décrite ci-dessus, nous étudions aussi les performances de l'estimateur proposé pour des échantillons à taille finie et discutons le cadre asymptotique.

Dans la troisième partie de cette contribution, nous nous intéressons à l'hétéroscédasticité dans les modèles partiellement linéaires pour variables exogènes réelles et variable réponse binaire. Nous proposons un modèle Probit spatial contenant une partie non-paramétrique. La dépendance spatiale est introduite au niveau des erreurs (perturbations) du modèle considéré. L'estimation des parties paramétrique et non paramétrique du modèle est récursive et consiste à fixer d'abord les composants paramétriques et à estimer la partie non paramétrique à l'aide de la méthode de vraisemblance pondérée puis utiliser cette dernière estimation pour construire un profil de la vraisemblance pour estimer la partie paramétrique. La performance de la méthode proposée est étudiée via une étude Monte-Carlo. La contribution finit par une étude empirique sur la relation entre la croissance économique et la qualité environnementale au Suède à l'aide d'outils de l'économétrie spatiale.

Mots-Clefs : Analyses de données fonctionnelles, Modèle linéaire fonctionnel, Processus auto-régressif spatial, Matrice de poids endogène, Quasi-maximum de vraisemblance, Statistique non-paramétrique, Régression, Estimateur à Noyau, Processus spatial, Econométrie spatiale, Estimateur semi-paramétrique, Hétéroscédasticité spatiale.

Abstract

Contribution to spatial econometric and functional data analysis.

This thesis covers two important fields of research in inferential statistics, namely spatial econometrics and functional data analysis. More precisely, we have focused on the analysis of real spatial or spatio-functional data by extending certain inferential methods to take into account a possible spatial dependence.

We first considered the estimation of a spatial autoregressive model (SAR) with a functional dependent variable and a real response variable using observations on a given geographical unit. This is a regression model with the specificity that each observation of the independent variable collected in a geographical location depends on observations of the same variable in neighboring locations. This relationship between neighbors is generally measured by a square matrix called the spatial weighting matrix, which measures the interaction effect between neighboring spatial units. This matrix is assumed to be exogenous, i.e. the metric used to construct it does not depend on the response variable of the model. The contribution of this thesis to this model lies in the fact that the explanatory variable is of a functional nature, with values in a space of infinite dimension. Our estimation methodology is based on a dimension reduction of the functional explanatory variable through functional principal component analysis followed by maximization of the truncated likelihood of the model. Asymptotic properties of the estimators, illustrations of the performance of the estimators via a Monte Carlo study and an application to real environmental data were considered.

In the second contribution, we use the functional SAR model studied in the first part by considering an endogenous structure of the spatial weighting matrix. Instead of using a geographical criterion to calculate the dependencies between neighboring locations, we calculate them via an endogenous process, i.e. one that depends on response variables. We apply the same two-step estimation approach described above and study the performance of the proposed estimator for finite or infinite-tending samples.

In the third part of this thesis we focus on heteroskedasticity in partially linear models for real exogenous variables and binary response variable. We propose a spatial Probit model containing a non-parametric part. Spatial dependence is introduced at the level of errors (perturbations) of the model considered. The estimation of the parametric and non-parametric parts of the model is recursive and consists of first setting the parametric parameters and estimating the non-parametric part using the weighted likelihood method and then using the latter estimate to construct a likelihood profile to estimate the parametric part. The performance of the proposed method is investigated via a Monte-Carlo study. An empirical study on the relationship between economic growth and environmental quality in Sweden using some spatial econometric tools finishes the document.

Keywords : Functional data analysis, Functional Linear Model, Spatial Autoregressive Process, Endogenous spatial weight matrix, Quasi-maximum likelihood estimator, Nonparametric statistics, Regression, Kernel estimate, Spatial process, Spatial econometrics, Semi-parametric estimation, Spatial heteroskedasticity.

Remerciement

Ces travaux de thèse ont été effectués au laboratoire LEM (Lille Économie et Management) CNRS 9221.

Je délivre mon expérience et je commence par un remerciement. Alors j'adresse un grand merci aux personnes qui ont cru en moi et qui m'ont encouragé pour arriver au bout de cette thèse.

Mes premiers remerciements vont spontanément à ma directrice de thèse Madame Sophie Dabo-Niang pour m'avoir guidé et encouragé tout au long de cette belle aventure. Un grand Merci pour ses conseils avisés et pour m'avoir encouragé dans cette voie. Ses conseils, sa disponibilité et son soutien tout au long de ces années m'ont permis de mener à bien ce travail. Qu'elle trouve ici toute ma reconnaissance. Je tiens également à remercier ma co-directrice de thèse, Madame Laurence Broze. Sa compétence et sa rigueur scientifique m'ont énormément aidé.

J'adresse ensuite mes remerciements aux membres du jury pour avoir accepté de juger mon travail et d'y avoir apporté leurs suggestions qui m'ont été d'une grande utilité: Messieurs Ahmed El Ghini, professeur à l'université Mohammed V, Abderrahim Taamouti professeur à l'université de Durham et Philippe Boussemart professeur à l'université de Lille, et madame Raja Chakir, professeur des universités à l' INRA-AgroParisTech.

Ma reconnaissance va ensuite au personnel du Laboratoire Lille Économie et Management pour m'avoir accueilli durant mes années de thèse et contribué à son bon déroulement. J'adresse mes remerciements à monsieur Hervé Leleu, directeur du Laboratoire Lille Économie et Management pour son accueil.

Je pense particulièrement à mes collègues et amis que j'ai rencontrés au laboratoire. Merci à tous les doctorants. Je n'oublierai jamais les bons moments que nous avons passés ensemble.

Enfin, mes dernières pensées vont à toute ma famille, aussi bien en France qu'en Tunisie. Je voudrais exprimer ma gratitude pour leurs présences, leurs encouragements et leurs aides sans lesquels je n'aurai pas pu atteindre cet objectif.

Contents

Notations

$\mathbb{N}$	set of natural numbers: $0, 1, 2\dots$
$\mathbb{N}^*$	set of non-zero natural numbers: $1, 2\dots$
$\mathbb{Z}$	set of integers: $\dots, -1, 0, 1, \dots$
$\mathbb{R}$	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 $\mathcal{T}$
$\sigma(\dots)$	σ-algebra generated by $(\dots)$
$(\Omega, \mathcal{A}, P)$	probability space
	Ω : nonempty set
	$\mathcal{A}$:σ-algebra of subset of Ω
	P: probability measure on $\mathcal{A}$
i.i.d	independent and identically distributed
$\mathcal{N}(0, 1)$	standard normal distribution
$u_n = O(v_n)$	a constant c exists such that $u_n \leq cv_n$
$u_n = o(v_n)$	$\frac{u_n}{v_n} \to 0$ as $n \to \infty$
$\blacksquare$	end of a proof

Chapter 1

General introduction

The present chapter provides a short English introduction to the topic of this dissertation and sums up its content. French readers are invited to read the summary in French at the beginning of each chapter.

In many area such as economic activities, epidemiology, climatology, ecology, environmental health, ..., data are geographically referenced. Proximity in space introduces correlations between the observations making standard statistical methods invalid. So taking into account spatial dependency, heterogeneity in model inference becomes of great importance and is a major interest of Spatial econometrics. This field of Econometrics has a huge set of concerns going from description, modeling and estimation. Basically, spatial data are observations of vector-valued variables at a finite number of points, but nowadays the technological progress produces tools capable of recording data with a fine grid and/or with high frequency. The results is a spatial dataset with very high dimension taking various forms like curves, shapes, images or a more complex mathematical object, named functional data.

A dynamic research area is combining functional data analysis (FDA) and spatial statistics to handle spatial data of functional nature. Various FDA methods to analyze curves or functions recorded at different locations have been developed within the geostatistics framework. Even if many potential applications are available in various domains, FDA methods on data distributed in a regular lattice are less developed.

Econometrics still considered as a powerful tool for decision-making and policy-making, in particular via the discrete choice models. It is mostly used technique to explain or predict choice of an individual or an agent by maximizing the utility of this choice. But the decision-makers can be inter-related based on spatial proximity, so decisions can be influenced through these interactions. In such situation, space's role become paramount.

This thesis cover some new features about outlined framework and it organized into 6 chapters, where the content is given as follows: Chapter 2 provides background information about the subjects of interest, models and estimation methods needed for the rest of the chapters. Our contribution starts by a spatial lag model with functional covariate given in Chapter 3 with a spatial weight matrix assumed to exogenous, whereas in the following Chapter 4 this assumption is relaxed to extend the model into a more general framework. Chapter 5 addresses the class of binary models in particular the Probit model. Within this framework, we propose a partially Probit spatial heteroskedasticity model. Chapter 6 consists of an empirical study applying some of the basic spatial econometric models to relate environment and economic growth. The last Chapter 7 summarizes the thesis and overviews futures works.

Written and oral communications

Publications

- Functional Linear Spatial Autoregressive Models. Ahmed, M. S., Broze, L., Dabo-Niang, S., & Gharbi, Z. (2018). In R. Giraldo, & J. Mateu (Eds.), *Geostatistical Functional Data Analysis: Theory and Methods*. Wiley. (Under press)

Communications

- CRoNoS, Limassol, Chypre, April 14-16 2019.
- 2nd conference on econometrics for the environment, Nador, Morocco, December 19-21 2018.
- 61st World Statistics Congress, Marrakech, Morocco, July 2017.
- Lebanese International Conference on Mathematics and Applications, Beirut, Lebanon, May 16-19 2017.
- Fifth Meeting on Statistics and Data Mining, Hammamet, Tunisia, March 14-17 2015.

Chapter 2

State of art and general concepts

Contents

This chapter offer a brief introduction of the fundamental concepts on which this thesis is based namely; Spatial econometric, Functional data analysis (FDA), Spatial Probit model, Partially linear model and Spatial heteroskedasticity.

2.1 Functional data analysis

For years, tools for data collection are in progress and for that there are a lot of fields that produce samples at a fine grid, for instance; high frequency data from monitoring equipment or optical tracking equipment (e.g handwriting data, but also for physiology, motor control, ...), electrical measurements (EKG, EEG and others) and spectral measurements (astronomy, materials, sciences). Therefore, it is obvious that these samples must be observed on a continuum that is not necessary related to temporal even thought it is usually the case. This kind of data can be classified within big dimensional data and classical statistics tools have difficulties to deal with it.

Growing interests in this category of data and related statistical techniques that can account for the infinite-dimensional nature of such data can be noted. Hence the emergence of a new branch of statistics named Functional Data Analysis (FDA) dealing with the data of high frequency but noisier and less frequent data can also be considered like weather data (temperature, precipitation, ...). Generally, the issues that arise regularly in FDA are data display and summarization, smoothing and interpolation, patterns in variability (principal component analysis) and regression (with functional predictors, outcomes, or both). This field was popularized by the monographies of Ramsay & Silverman (1997, 2005), Bosq (2000), Horváth & Kokoszka (2012) and Hsing & Eubank (2015). Various statistical tasks have been studied with functional data but the previous literature were concentrated around parametric models and methods which we addressing in this thesis, for the non-parametric models, one we refer to Ferraty & Vieu (2000, 2006).

Classical statistical methods for multivariate statistic have difficulties to deal with functional dependent data with mainly two reasons: first, the infinite dimension of the functional variables (curves, shapes, ...) to be handle in practice, the other reason is the dependency between observations when considering time-series of spatial functional objects that may be difficult to manage in a non-costly way. So adapted techniques have been proposed during the last two decades particularly in the parametric framework considering linear regression models. In the next section, we introduce the generalized functional linear model, the baselines for our proposed model in Chapters 3 and 4 with a functional co-variate.

2.1.1 Generalized functional linear models

The idea of functional linear models dates back to the 1990's. There are three different scenarios for the predictive relationships in a functional linear model:

1. Functional covariate, real-valued response;
2. Real-valued covariate, functional response;
3. Functional and non-functional covariates, functional response.

Various investigations focused on a functional response such as Ramsay & Silverman (1997), Faraway (1997), Cardot et al. (1999) and Fan & Zhang (2000). For situations where the response variable is real-valued and the predictor is functional, Hastie & Mallows (1993) introduced the functional linear model with scalar response variable while James & Hastie (2001) discussed performing regression where the response is binary and the predictor is functional. More recently, Muller & Stadtmuller (2005) and Cardot & Sarda (2005) proposed a generalized version of the functional linear model (GFLM) the baseline of the functional models we propose. An application to GFLM was investigated by James (2002).

Let $(Y_i,\ \{X_i(t),\ t \in \mathcal{T}\})$, $i = 1, ..., n$ be a n sample of i.i.d observations from a scalar response Y and a functional predictor $X(t)$, $t \in \mathcal{T}$ that is a random curve, namely a square integrable stochastic process on a real interval $\mathcal{T} \subset \mathbb{R}$ taking values in $\mathcal{X} \subset L^2(\mathcal{T})$.

In the GFLM model of interest, the linear predictor is obtained by forming the scalar product of the predictor function $X(t)$ with a smooth parameter function $\beta(\cdot)$ which is assumed to be squared integrable and belongs to $L^2(\mathcal{T})$. Assuming a link function $\Phi(.)$ that is monotone and twice continuously differentiable, the variance function $\sigma^2(\cdot)$ related to $\Phi(.)$ is strictly positive, let the linear predictor η be:

$$\eta = \alpha + \int \beta(t)X(t)dt(t) \tag{2.1}$$

with the conditional mean $\mu = \Phi(\eta)$, where $\mathbf{E}\left(Y|X_i(t),\, t \in \mathcal{T}\right) = \mu$ and $\mathbf{Var}\left(Y|X_i(t), t \in \mathcal{T}\right) = \tilde{\sigma}^2(\eta)$ for $\tilde{\sigma}^2(\eta) = \sigma^2(\mu) = \sigma^2(\Phi(\eta))$.

The distribution of Y can be specified within the exponential family. Using the sample, let:

$$Y_i = \Phi(\eta_i) + e_i, \quad i = 1, \ldots, n, \tag{2.2}$$

where the errors e_i are i.i.d, $\mathbf{E}(e_i|X_i) = 0$ and $\mathbf{Var}(e_i|X_i) = \tilde{\sigma}^2(\eta_i)$. It is more convenient to work with standardized errors and we can define $e'_i = e_i/\sigma(\mu) = e_i/\tilde{\sigma}(\eta_i)$, hence model (2.2) can be rewritten as follows:

$$Y_i = \Phi(\eta_i) + e'_i\tilde{\sigma}(\eta_i), \quad i = 1, \dots, n, \tag{2.3}$$

for which $\mathbf{E}(e'_i|X_i) = 0$, $\mathbf{E}(e'_i) = 0$ and $\mathbf{E}({e_i}'^2) = 1$.

Model (2.3) may be seen as the generalization of the GLM introduced by Nelder & Wedderburn (1972). Such models introduced in the non-functional literature are estimated using different well known methods such as likelihood or quasi-likelihood methods. These last are not directly applicable for GFLM because of the infinite dimensionality of the predictors. Basis expansions are basically used to reduce the dimension of the functional predictor (see Castro et al. (1986)). This expansion idea was used in several papers such as Cardot et al. (2003) who used functional principal component analysis. Marx & Eilers (1999) used a penalized P-spline approach while Cardot & Sarda (2005) used a penalized method with B-Splines.

In our contribution, we adopt the following basis expansion truncation strategy of Muller & Stadtmuller (2005) in the context of GFLM to reduce the dimension of the functional explanatory variable $X(.)$ and the corresponding parameter function $\beta(.)$. Denote by $\varphi_j(t)$, $j = 1, 2, \dots$ an orthonormal basis function of $L^2(\mathcal{T})$ commonly chosen as the Fourier or a basis formed by eingen-fucntions of the covariance operator of $X(.)$. The predictor process $X(.)$ and parameter function $\beta(.)$ can be expanded into:

$$X(t) = \sum_{j=1}^{\infty} X_j^\dagger \varphi_j(t) \quad \text{and} \quad \beta(t) = \sum_{j=1}^{\infty} \beta_j^* \varphi_j(t) \quad \text{for all } t \in T,$$

with $X_j^\dagger$ and coefficients β_j^* are given by $X_j^\dagger = \int_{\mathcal{T}} X(t)\varphi_j(t)dt$ and $\beta_j^* = \int_{\mathcal{T}} \beta(t)\varphi_j(t)dt$, respectively. Therefore from the orthonormality characteristic of the basis, it results that:

$$\int_T X(t)\beta(t)dt = \sum_{j=1}^{\infty} \beta_j^* X_j^\dagger \tag{2.4}$$

For a positive sequence of integer p_n, let the next decomposition:

$$\sum_{j=1}^{\infty} \beta_j^* X_j^\dagger = \sum_{j=1}^{p_n} \beta_j^* X_j^\dagger + \sum_{j=p_n+1}^{\infty} \beta_j^* X_j^\dagger \tag{2.5}$$

Then η in (2.1) can be expressed as:

$$\tilde{\eta} = U_{p_n} + V_{p_n} \quad \text{with} \quad U_{p_n} = \alpha + \sum_{j=1}^{p_n} \beta_j^* X_j^\dagger \quad \text{and} \quad V_{p_n} = \sum_{j=p_n+1}^{\infty} \beta_j^* X_j^\dagger.$$

Note that $\mathbf{E}(Y|X) = \Phi(U_{p_n} + V_{p_n})$, so by assuming that the term related to V_{p_n} vanishes asymptotically as $p_n \longmapsto \infty$, one may instead of (2.3) use the p_n-truncated model defined by:

$$Y_i^{(p_n)} = \Phi(\tilde{\eta}_i^*) + e'_i\tilde{\sigma}(\tilde{\eta}_i^*), \quad i = 1, \dots, n, \tag{2.6}$$

where $\tilde{\eta}_i^* = \alpha + \sum_{j=1}^{p_n} \beta_j^* X_j^{\dagger(i)}$, with $X_j^{\dagger(i)} = \int_{\mathcal{T}} X_i(t)\varphi_j(t)dt$. Now fixing the truncation level p_n, one can estimate the unknown parameter vector $\theta^T = (\alpha, \beta_1^*, \dots, \beta_{p_n}^*)$, where α is the intercept.

Let $\hat{\theta}^T = (\hat{\alpha}, \hat{\beta}_1^*, \ldots, \hat{\beta}_{p_n}^*)$ be the estimator of θ^T, then an estimator of the parameter function β is given by:

$$\hat{\beta}(t) = \sum_{j=1}^{p_n} \hat{\beta}_j \varphi_j(t). \tag{2.7}$$

A key point in the parameter vector θ^T estimation is the choice of the truncation parameter p_n. Several initiatives have been proposed in the functional literature. A usual one if Akaike information criterion (AIC) used by Muller & Stadtmuller (2005) in a context of generalized functional linear models with a functional covariate and a real-valued response. These authors also shown the consistency of AIC criterion under appropriate assumptions and discussed the usefulness of this criterion. For more details on the choice of the truncation parameter, we refer the reader to Muller & Stadtmuller (2005).

2.2 Spatial data

Statistics for spatial data was first studied in geology and meteorology fields during the 1960's. A growing interest may be noted since the seminal works of Krige and Matheron on spatial prediction. A main feature of spatial statistics is that, data are located in locations of at least bivariate dimension with spatial dependency between observed data at different locations. This kind of data is available in various fields such as urban systems, agriculture, economics, environmental science and economics. Moreover, examining and modeling the spatial patterns is an important task for statisticians, consequently a wide range of models and methods have been developed to incorporate the spatial dependence structure basically within the scope of geostatistics, lattice data and point patterns (see Cressie, 1993). In the geostatistics framework, the correlation between locations are expressed as a continuous function of distance (see Cressie, 1993, Chapter 2).

There are many case studies and techniques for such framework, particularly spatial parametric interpolation methods, namely kriging (e.g. Fedorov (1989), Berke (2004), Oliveros et al. (2010), Giraldo et al. (2011) and Bohorquez & Mateu (2016)) and non-parametric regression or prediction (e.g. Cortes-D et al. (2016), Dabo-Niang et al. (2010), Giraldo et al. (2010a) and Ternynck (2014)). In this framework, locations are in a continuous spatial set, compare to many domains such as; remote sensing from satellites, image analysis, weather patterns, agriculture among others, where data consist of counties or census tracts or in general are observed at regular lattice (regular spaced points in $\mathbb{R}^2$). Basically statistical models for such lattice data express the fact that observations are nearby when they tend to be alike. Compare to geostatistical and lattice data, spatial point patterns occurs when locations where data are available are random. It is not always easy to distinguish these three types of data:

Geostatistical data

- The spatial set of interest $\left\{\mathcal{S} \subset \mathbb{R}^N,\ N \geq 2\right\}$ is a fixed subset of the plane of positive area (2-D) or volume (3-D).
- And a spatial process (collection of random variables observed at spatial points) $Y = \{Y(s),\ s \in \mathcal{S}\}$ is of interest.

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 processes includes extension to well known time-series processes.

Point patterns

- The spatial locations $s \in \mathcal{S} \subset \mathbb{R}^N$ where the process $Y = \{Y(s),\ s \in \mathcal{S}\}$ is defined are random.

- This type of processes is an extension of usual point processes.

In this dissertation, we are interested in spatial lattice processes that are of great interest in many domains such as econometrics, particularly one is interested to predict the behavior of some locations knowing that of some neighbors. Such processes have some analogies with time series where probabilistic models are used to describe the relation between future and past values of the process for prediction purpose. However, times series models cannot be directly applied to spatial data since the natural order in the time domain does not exist in the spatial context. Differences and similarities between spatial and times series data are highlighted in Tjøstheim (1987).

As other types of spatial processes, exploration of spatial correlation structure is the first step in the lattice context. A basic correlation tool in spatial econometrics is the spatial weight matrix which describe the connectivity between different locations. This matrix takes different forms and play a crucial role in econometrics inference. Next, we provide more details on the construction of the spatial weight matrix and its influence on spatial models inference.

2.2.1 Specification of the Spatial Weight Matrix

In spatial econometrics literature, interdependence or/and interactions between spatial units is defined via the spatial weight matrix, denoted W_n in the following, with n is the sample size. The major weakness of spatial econometric models is that the spatial weight matrix must be specified, hence the risk of misspecification of this matrix is often arises. Furthermore, an incorrect specified structure could lead to wrong conclusion and cause bias in model estimation. Many investigations pointed out the critical dependency of econometrics model's estimation on the spatial weighting matrix choice, see for instance Mizruchi & Neuman (2008) and Farber et al. (2009).

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 weights between locations i and j. There are mainly three different ways of specifying the weights using either a binary or continuous or hybrid metrics, depending on the nature of the data. The reader may refer to Pinkse & Slade (1998) for these specifications, we remind some example afterwards.

Binary: This form is usually used to specify whether spatial units share a common border or not. So for a spatial unit i with a set of boundary denoted *bnd(i)*, the elements w_{ij} are defined by:

$$w_{ij} = \begin{cases} 1, & \text{if } bnd(i) \cap bnd(j) \neq \varnothing \\ 0, & \text{if } bnd(i) \cap bnd(j) = \varnothing \end{cases}$$

These weights are called *queen contiguity weights* and allow the possibility that spatial units share only a boundary point. This definition may be set more restrictive to require some *positive* portion of the boundary. Let l_{ij} denote the length of shared boundary between locations i and j, then:

$$w_{ij} = \begin{cases} 1, & \text{if } l_{ij} > 0 \\ 0, & \text{if } l_{ij} = 0 \end{cases}$$

This is called a *rook contiguity weight*. Another binary form can be obtained using distances by assuming some *threshold distance* d such that beyond d there is no direct spatial influence between location i and j. The corresponding weights are called radial distance weights:

$$w_{ij} = \begin{cases} 1, & \text{if } 0 \leq d_{ij} \leq d, \text{ with } d_{ij} \text{ being the distance between } i \text{ and } j \\ 0, & \text{if } d_{ij} > d \end{cases}$$

Continuous: In this category, the used metric is a function of the centroid distance[1] d_{ij} between each pair of spatial units (i, j). The elements of W_n are defined by the inverse of distance $w_{ij} = \frac{1}{d_{ij}}$. If one suppose that there is a diminishing effect with the distance, then one can assume that weights are *negative power functions* and defined as $w_{ij} = d_{ij}^{-\alpha}$, where α is a positive exponent. An alternative to the last, is the use of *negative exponential functions*, with $w_{ij} = e^{-\alpha d_{ij}}$.

Hybrid: In many situations, binary and continuous relations may be combined. An example is the one given by Cliff & Ord (1969) when studying Eire blood-group data, they found that the best structure of W_n to capture spatial autocorrelation is to combine *power distance* and *boundary shares* as:

$$w_{ij} = \frac{l_{ij} d_{ij}^{-\alpha}}{\sum_{k \neq i} l_{ik} d_{ik}^{-\alpha}}$$

where, $\alpha = 1$ and l_{ij} is the portion of shared border between spatial unit i and j. The number of neighbors of a spatial unit i may be fixed using the k-nearest neighbors method. Let the distance between a unit i and all other units $j \neq i$ be ranked as follows: $d_{ij(1)} \leq d_{ij(2)} \leq \cdots \leq d_{ij(n-1)}$. Let the set $N_k(i) = \{j(1), j(2), \cdots, j(k)\}$ contains the k units closest to i, then for $k = 1, \cdots, n-1$, the weight matrix is:

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

Numerical investigations (LeSage & Pace, 2014) confirm that inference results are usually robust to the choice of k.

[1] In real data, this geographical distance could be a linear distance or a travel time distance, or a combination of both to calculate a distance between two points (region's centroid or other spatial units such as cities).

Other weight matrix specifications are available in the literature within these three groups. In practice, it is common but not necessarily to row normalize W_n, this leads to $0 \leq w_{ij} \leq 1$ and $\sum_{i=1} \sum_i^n w_{ij} = 1$, so each element w_{ij} can be then interpreted as the *fraction* of influence on unit i attributable to j. Technically, in model inference, row normalization avoids numerical issues related to different scaling of variables and permits to have an intuitive interpretation of spatial autoregressive parameter in some autoregressive models (SAR models) discussed in the following. Note that row-normalization does not affect the relative weight neighbors effect on other units as much that it alters the magnitude of the collective impact.

As mentioned before, misspecification of W_n may have important impacts on the model inference. Then one may naturally ask "*How can we specify correctly the spatial weighting matrix ?*". Several solutions are proposed in the literature. Kooijman (1976) proposed a simple technique that select the spatial weight matrix that maximizes the well known Moran's coefficients. Another proposal by Holloway & Lapar (2007) relies on the Bayesian marginal likelihood approach to select neighborhoods. Kostov (2010) suggests using component-wise boosting algorithm to choose the appropriate spatial weighting matrix amongst a set of predetermined alternatives, extended later by Kostov (2013) into a two selection procedure. Recently, Ahrens & Bhattacharjee (2015) proposed a two-step lasso selection.

Historically, the structure of the weight matrix is based on geographic criteria which simplifies models inference. In this case, the weight matrix is assumed to be exogenous. In fact, geographic distance is not always appropriate to all situations. Tobler (1970) claimed that "the space is irregular and heterogeneous so the influence may be of any type across space". Spatial interactions may not be necessarily defined in a geographical point of view, sometimes economic or social metrics may be more realistic. This is in line with the idea of allotopy principal stated by Ancot & Molle (1982): often what happens in a location is related to other phenomena located in distinct and remote parts of the space. Sometimes the spatial weighting matrix may involve socio-economic indicators that entails endogeneity on the model of interest. As said by Anselin & Bera (1998) " weights should be chosen with great care to ensure their exogeneity, unless their endogeneity is considered explicitly in the model specification ". In case of spatial endogenous weighting, basic inference approaches usually fail.

Compare to exogenous weight matrices, the literature on endogenous context is limited. Baicker (2005) used the per capita income levels to define spatial weighting matrix to estimate the degree of influence of a state spending on other neighboring states spending levels. Ertur & Koch (2011) simply defined the weights w_{ij} as the average imports of country i coming from country j. Moreover Ho et al. (2013) proposed a slow growth model with a spatial autoregressive term and a spatial weighting matrix based on sum of trade flows between countries to examine the international spillover effect of economic growth. However, few attention is given to the problem of estimation when having endogenous weight. The reader can refer to Kelejian & Piras (2014), who proposed a 2SLS estimator for panel data with endogenous spatial matrix and Lee & Yu (2017), who estimate spatial panel data with a weight matrix based on a control function.

2.2.2 Spatial econometric models and their inference

Let (Y, X) be a random vector observed at n locations $\{s_1, \dots, s_n\}$ 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 $\mathbf{Y}_n = (Y_1, \dots, Y_n)^T$ is the sample response and $\mathbf{X}_n$ the $n \times p$ matrix of explanatory variables observations with elements X_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, p$. By using one of the spatial matrices introduced earlier, three different types of interaction effects are mainly considered:

- Endogenous interaction effects among dependent variables.
- Exogenous interaction effects among independent variables.
- Correlated effects, where similar unobserved characteristics result in similar behaviors.

Manski (1993) gathered these three interaction types into a full model defined by:

$$\mathbf{Y}_n = \rho W_n \mathbf{Y}_n + \mathbf{X}_n \beta + W_n \mathbf{X}_n^T \eta + \mu_n^v; \quad \mu_n^v = \lambda M_n \mu_n^v + \epsilon_n^v, \quad \epsilon_n^v \sim N(0, I\sigma^2) \tag{2.8}$$

where $\mu_n^v = (\mu_1, \ldots, \mu_n)^T$ and $\epsilon_n^v = (\epsilon_1, \ldots, \epsilon_n)^T$, $1 \leq \rho \leq 1$ and $1 \leq \lambda \leq 1$ are scalar autoregressive parameters indicating the degree of spatial dependence and η is a $p \times 1$ vector of parameters. W_n and M_n are spatial weighting matrices with element w_{ij} and m_{ij} respectively reflecting the relative degree of connections between units i and j.

In practical point of view, a population that contains jointly these different interactions is almost non-existent. In fact the interest of practitioners has been focused on modeling one or two of these interaction effects so the general Manski model is typically not used. An alternative is one of the particular cases of model (2.8) illustrated in Figure (2.1). If the interaction effects are:

i. endogenous, if Y_i at spatial units i depends on Y_j at spatial units j, the corresponding model is named spatial lag model or spatial autoregressive model (SAR) where the interaction effect is denoted by the spatial lag $W_n \mathbf{Y}_n$,

ii. exogenous, if Y_i at spatial units i depends depends on X_j at spatial units j,

iii. correlation effects, this means that the interaction is among the error terms and we use spatial autoregressive error (SAE) model (or spatial error model; SEM) with the interaction effects $W_n \mu_n^v$.

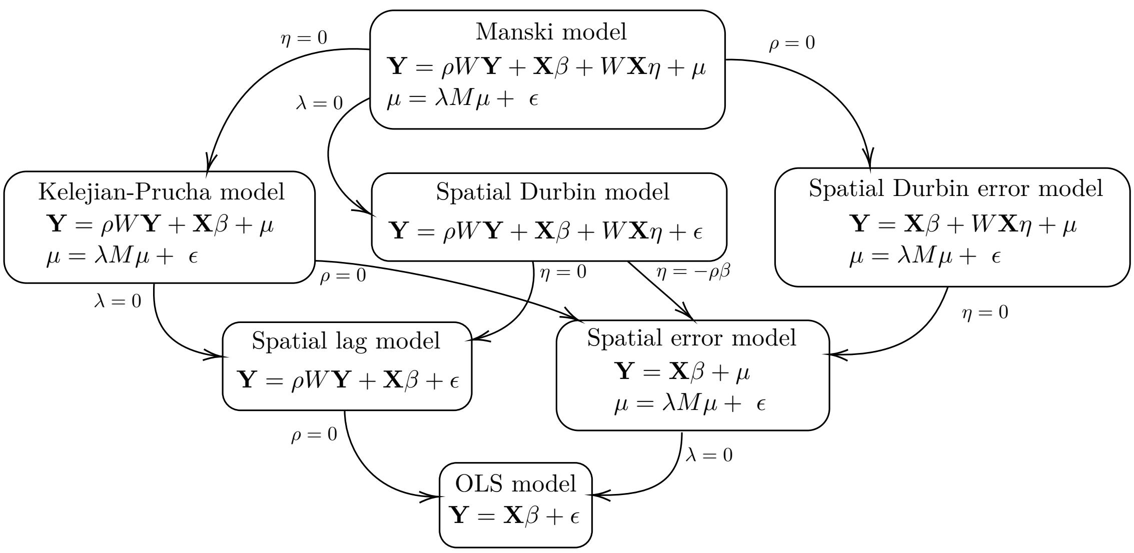

Figure 2.1: The relationships between different spatial dependence models.

A large number of papers investigates these models and their inference method as the two stage least squares (2SLS), the three stage least squares (3SLS), the maximum likelihood (ML) and the generalized method of moments (GMM), see among others Anselin (1988), Kelejian & Robinson (1993), Kelejian & Prucha (1998), LeSage & Pace (2009), Elhorst (2010), Case (1991), Arbia & Baltagi (2010), ... A particular interest is dedicated to spatial autoregressive model (SAR), useful in many domains such as econometrics ecology, social networks, ... It extends time series autoregressive models and the basic regression model by incorporating a spatial lag vector reflecting the average effect from neighboring locations to explain variation across the spatial set of interest. Intuitively, the value of a variable located at a given geographic point depends on the values of the same variable located in neighboring points. A SAR model may be obtained by taking $\eta = 0$ and $\lambda = 0$ in model (2.8). In this thesis, we are in some contributions interesting to SAR models and its maximum likelihood inference.

2.2.3 Estimation of basic SAR models

Inference for SAR models in a real-valued data context has been extensively studied in the literature. The first estimation method proposed was the maximum likelihood (ML) method introduced by Ord (1975) but this last has a computationally weakness when more than one spatial lag is used. Instrumental variable IV (or 2SLS) method (Anselin (1980) and Kelejian & Prucha (1998, 1999)) and the generalized method of moments (GMM) summarized by Lee (2007) and Lin & Lee (2010) are also proposed. It is obvious that ML estimator is performant and consistent when the model disturbances are normally distributed. When the disturbances are not normal, quasi-maximum likelihood method (QML) is the alternative that we investigate in some contributions. The identification and estimation of SAR models when using QML have not been studied enough. Lee (2004) provided a QML estimator for SAR with real-valued data and gave asymptotic proprieties. The next lines recall the principle of this method. Let the SAR model:

$$\mathbf{Y}_n = \lambda W_n \mathbf{Y}_n + \mathbf{X}_n \beta + \epsilon_n^v, \tag{2.9}$$

where ϵ_n^v is an $n \times 1$ vector of i.i.d elements with mean zero and variance σ^2. The parameter of interest to be estimated are the $k \times 1$ vector of regression β, the spatial parameter λ and σ^2. The logarithm of the quasi-likelihood function of (2.9) is:

$$L_n(\theta) = -\frac{n}{n}\ln(2\pi) - \frac{\pi}{2}\ln(\sigma^2) + \ln|\mathbf{S}_n(\lambda)| - \frac{1}{2\sigma^2}\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\lambda), \tag{2.10}$$

where $\theta = (\beta', \lambda, \sigma^2)'$, $\mathbf{S}_n(\lambda) = I_n - \lambda W_n$ with I_n is an $n \times n$ identity matrix, and $\epsilon_n(\beta,\lambda) = S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n\beta$. For a given λ, (2.10) is maximized at:

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

and

$$\hat{\sigma}^2_{n,\lambda} = \frac{1}{n}\left(\mathbf{S}_n(\lambda)\mathbf{Y}_n - \mathbf{X}\hat{\beta}_{n,\lambda}\right)'\left(\mathbf{S}_n(\lambda)\mathbf{Y}_n - \mathbf{X}\hat{\beta}_{n,\lambda}\right) \tag{2.11}$$

$$= \frac{1}{n}\mathbf{Y}_n'\mathbf{S}_n'(\lambda)M_n\mathbf{S}_n(\lambda)\mathbf{Y}_n, \tag{2.12}$$

where, $M_n = I_n - \mathbf{X}_n\left(\mathbf{X}_n'\mathbf{X}_n\right)^{-1}\mathbf{X}_n'$. By substituting $\hat{\beta}_{n,\lambda}$ and $\hat{\sigma}^2_{n,\lambda}$ into (2.10) one can derive the concentrated log-quasi-likelihood function of λ as:

$$L_n^c(\lambda) = -\frac{n}{2}\left(\ln(2\pi) + 1\right) - \frac{n}{2}\ln\hat{\sigma}^2_{n,\lambda} + \ln|\mathbf{S}_n(\lambda)|$$

The value $\hat{\lambda}_n$ which maximize $L_n^c(\lambda)$ is the estimator of λ, and thus the QMLEs of β and σ^2 are $\hat{\beta}_{n,\hat{\lambda}_n}$ and $\hat{\sigma}^2_{n,\hat{\lambda}_n}$, respectively.

Asymptotic results (consistency and asymptotic normality) may be found in Lee (2004) under some conditions. In particular, he proved that the rate of convergence of the above estimators is at $\sqrt{n}$ and may be slower for some parameter if the degree of spatial dependence grows with the sample size n. Note that a key condition to establish these asymptotic results is that the spatial weight matrix is strictly exogenous.

2.2.4 Estimation of SAR with endogenous spatial weight matrix

Inference methods mentioned earlier are established under the basic assumption of exogenous spatial weight matrix. As mentioned before, basic SAR inference methods cannot be used directly when the weight matrix is endogenous due in particular to some technical complications. This is actually one of the several problems emphasized by Pinkse & Slade (2010), waiting for good solutions. Recently Kelejian & Piras (2014) proposed to instrumenting the endogenous spatial weight matrix in case of spatial panel data while Shi & Lee (2018) proposed a QML estimator for spatial panel models with endogenous time varying spatial weights matrices. Furthermore Han & Lee (2016) proposed a Bayesian estimation approach accounting for endogenous spatial weight matrices. In a context of cross-sectional SAR model, Qu & Lee (2015) considered the case of endogenous spatial weight constructed by a univariate economic variable. Let us recall the estimating method proposed by Qu & Lee (2015) the baseline of some of our contributions (see Chapter 4).

In the SAR model (2.9) the components of the spatial weight matrix are $w_{ij} = h(d_{ij})$, with $h(.)$ a bounded function and d_{ij} a metric of geographic distance between spatial unit i and j. Instead of using geographic distance, one may use another metric in a endogenous setting. Let $Z_n = (z_{1,n}, \ldots, z_{n,n})'$ be an $n \times p$ matrix with $z_{i,n} = (z_{1,in}, \ldots, z_{p,in})'$ with $p \in \mathbb{N}$:

$$\mathbf{Z}_n = \mathbf{U}_n \Upsilon + \upsilon_n, \tag{2.13}$$

where $\mathbf{U}_n$ is $n \times k$ matrix with elements $\{u_{in}\}$, being bounded and deterministic, Υ is a $k \times p$ matrix of coefficients and $\upsilon_n = (\upsilon_{1,n}, \ldots, \upsilon_{n,n})'$ is an $n \times p$ matrix of disturbances with $\upsilon_{i,n} = (\upsilon_{1,in}, \ldots, \upsilon_{p,in})'$. Thus in model (2.9) the elements of W_n, can be constructed using $\mathbf{Z}_n$, that is $w_{ij,n} = h_{ij}(\mathbf{Z}_n)$. To apply QML method with exogenous weight matrix the covariance between ϵ_n, υ_n needs to be zero; $Cov(\epsilon_n, \upsilon_n) = 0$ but this is not the case if $w_{ij,n} = h_{ij}(\mathbf{Z}_n)$. Then, Qu & Lee (2015) proposed a solution by assuming that the errors terms ϵ_n and υ_n have a joint distribution: $(\epsilon_{in}, \upsilon'_{in}) \sim i.i.d(0, \Sigma_{\epsilon\upsilon})$, where $\Sigma_{\upsilon\epsilon} = \begin{pmatrix} \sigma_\epsilon^2 & \sigma'_{\epsilon\upsilon} \\ \sigma_{\epsilon\upsilon} & \Sigma_\upsilon \end{pmatrix}$ is a positive variance-covariance matrix, σ_ϵ^2 is a scalar variance, $\sigma_{\epsilon\upsilon} = (\sigma_{\epsilon\upsilon_1}, \ldots, \sigma_{\epsilon\upsilon_n})'$ is a p dimensional covariance vector, and Σ_υ is a $p \times p$ matrix. Based on some conditional moments the outcome equation (2.9) becomes:

$$\mathbf{Y}_n = \lambda \mathbf{Y}_n W_n \mathbf{X}_n \beta + (\mathbf{Z}_n - \mathbf{U}_n \Upsilon)\delta + \xi_n, \tag{2.14}$$

where ξ_n are i.i.d with $E(\xi_{i,n}|\upsilon_{i,n}) = 0$ and $E(\xi_{i,n}^2|\upsilon_{i,n}) = \sigma_\xi^2 I_n = \sigma_\epsilon^2 - \sigma'_{\epsilon\upsilon}\Sigma_\upsilon^{-1}\sigma_{\epsilon\upsilon}$ and $\delta = \Sigma_\upsilon^{-1}$. The term $(\mathbf{Z}_n - \mathbf{U}_n\Upsilon)$ is considered as a variable controlling the endogeneity of W_n. The log likelihood function based on the normal joint distribution of υ_n and ϵ_n can be written as:

$$\begin{aligned} \ln L_n(\theta) &= -n\ln(2\pi) - \tfrac{n}{2}\ln\sigma_\xi^2 + \ln|S_n(\lambda)| - \tfrac{n}{2}\ln|\Sigma_\epsilon| \\ &\quad -\tfrac{1}{2}\sum_{i=1}^{n}(Z_i' - U_i'\Upsilon)\Sigma_\epsilon^{-1}(Z_i - \Upsilon' U_i) \\ &\quad -\tfrac{1}{2\sigma_\xi}[S_n(\lambda)\mathbf{Y}_n - \psi_{p_n}\beta - (\mathbf{Z}_n - \mathbf{U}_n\Upsilon)\delta]' \\ &\quad \times[S_n(\lambda)\mathbf{Y}_n - \psi_{p_n}\beta - (\mathbf{Z}_n - \mathbf{U}_n\Upsilon)\delta], \end{aligned} \tag{2.15}$$

where $S_n(\lambda) = I_n - \lambda W_n$. The QMLE estimator is of θ is $\hat{\theta} = argmax_{\theta\in\Theta}\ln L_n(\theta)$, where $\theta = (\lambda, \beta', \mathrm{vec}(\Upsilon)', \sigma_\xi^2, \alpha', \delta')'$ with $\delta = \Sigma_\epsilon^{-1}\sigma_{\epsilon\upsilon}$, $\sigma_\xi^2 = \sigma_\epsilon^2 - \sigma_{\epsilon\upsilon}'\Sigma_\upsilon^{-1}\sigma_{\epsilon\upsilon}$ and α being a J-dimensional column vector of distinct elements in Σ_ϵ. Qu & Lee (2015) established the consistency and asymptotic normality of $\hat{\theta}$ under near-epoch dependence (NED), see Jenish & Prucha (2012) for more details.

2.3 Semi-parametric modeling

Semi-parametric models are alternatives to fully parametric or fully non-parametric models. They are more realistic in some situations and may characterize a non-linear relation and maintains the natural behavioral interpretation of the linear regression. In this thesis, we are interested in some contributions to model a relation between a response variable Y and independent variables $Z = (X, T)$ when the conditional expected of Y given Z is written as $g(X\alpha + f(T))$ where α is a vector of parameter, $f(.)$ is an unknown smooth function and $g(.)$ is known link function. Hence, we are face with a semi-parametric model.

Several semi-parametric models have been proposed in the literature mainly for i.i.d data, see for instance Green & Yandell (1985) that proposed an inference method for both parametric and non-parametric components using the penalized likelihood function. Hastie & Tibshirani (1990) adopted the same approach, together with the "backfitting algorithm" which gave proof of performance in the estimation of the purely non-parametric generalized additive model while Hunsberger (1994) used a weighted likelihood function. Severini & Staniswalis (1994) proposed a quasi-likelihood estimation of a semi-parametric model based on the generalized profile likelihood of Severini & Wong (1992). Non-linearity also can be found in the field of spatial econometric but not too developed. For instance Gao et al. (2006) proposed estimators for a spatial semi-parametric (partially linear) based on an additive marginal integration projection on the set of additive functions, and Robinson (2010) consider two different adaptive estimates for a spatial autoregressive model, containing non-stochastic explanatory variables. More recently Hoshino (2018) proposed a semi-parametric series generalized method of moments estimator for spatial autoregressive models. More examples and details will be exhibiting in Chapter 5.

2.3.1 Quasi-likelihood estimation of semi-parametric models

Quasi-likelihood method is a flexible and robust alternative to maximum likelihood method when exact information on the distribution is not available and only second moments are available. In this section, we recall the main lines of the quasi-likelihood estimation method to estimate the semi-parametric model proposed by Severini & Staniswalis (1994) and used in Chapter 5.

Consider $\{Y_i, X_i, T_i\}$, $i = 1, \ldots n$, $n \in \mathbb{N}$, a set of observations from a random vector $\{Y, X, T\}$, where $Y \subset \mathbb{R}$ is a response variable, $X \subset \mathbb{R}^p$, $p > 1$ are covariates and T takes values in $\mathcal{T} \subset \mathbb{R}^q$, $q > 1$. Let now $E(Y|X,T) = \mu(X,T)$ and $\mathrm{var}(Y|X,T) = \sigma^2 V(g(.))$, where $\mu(X,T) = g(f(T) + X\alpha)$, $f(.)$ is an unknown smooth function from $\mathbb{R}^q$ to $\mathbb{R}$ and α is $p \times 1$ unknown parameter that belong to a compact subset $\Theta_\alpha \subset \mathbb{R}^p$. So we can define the quasi-likelihood function by:

$$Q(g(.);Y) = \int_{g(.)}^{Y} \frac{s-Y}{V(s)} ds.$$

For a fixed α, the estimation of $f(.)$ is a non-parametric problem, the weight quasi-likelihood of Severini & Staniswalis (1994) may be used. Assuming that $K(.)$ is a kernel on $\mathbb{R}^d$ to $\mathbb{R}_+$ with a sequence of bandwidth $b > 0$ depending on n, the estimator $\hat{f}(.)$ of $f(.)$, is the solution in η of:

$$\sum_{i=1}^{n} K\left(\frac{t-T_i}{b}\right) \frac{\partial}{\partial\eta} Q\left(g(\eta + X_i\alpha); Y_i\right) = 0. \tag{2.16}$$

Given $\hat{f}(.)$, the estimator of α is a parametric problem and the estimate $\hat{\alpha}$ is obtained by solving:

$$\frac{\partial}{\partial\alpha} \sum_{i=1}^{n} Q(g(\hat{f}(T_i) + X_i\alpha); Y_i) = 0. \tag{2.17}$$

Then the final estimate of $f(.)$ is obtained by solving (2.16) using $\hat{\alpha}$ instead of α. In fact the estimator here suffers from high bias for t near the boundary of $\mathcal{T}$ but it can be reduced using the "trimming" approach which consists of using in (2.17), only observation T_i away from boundary. Hence, let $I_i = 1$, if $T_i \in \mathcal{T}_0$ and 0 otherwise, where $\mathcal{T}_0$ is a subset of t away from boundary of $\mathcal{T}$. Then $\hat{\alpha}$ can be obtained by solving:

$$\frac{\partial}{\partial\alpha} \sum_{i=1}^{n} I_i Q(g(\hat{f}(T_i) + X_i\alpha); Y_i) = 0. \tag{2.18}$$

Severini & Staniswalis (1994) give some examples where explicit forms of (2.16) and (2.18) give a closed solution for α and $f(.)$. In general, let us consider the next algorithm for computing the estimates. Let

$$\psi_1(\eta;\alpha,t) = \sum_i K\left(\frac{t-T_i}{b}\right) G(\eta + X_i\alpha)(Y_i - g(\eta + X_i\alpha)) \tag{2.19}$$

and

$$\begin{aligned} \psi_{2k}(\alpha, f_\alpha(.)) = \sum_i & I_i G(f_\alpha(T_i) + X_i\alpha)(X_{jk} + f'_{k\alpha}(T_i)) \\ & \times (Y_i - g(f_\alpha(T_i) + X_i\alpha)),\ k = 1, \dots p\,, \end{aligned} \tag{2.20}$$

where $G(.) = g'(.)/V(g(.))$, and $f_\alpha(t)$ is an arbitrary function from $\mathbb{R}^q$ to $\mathbb{R}$ for each α. Let the derivative function of α with respect to component α_k be $f'_{k\alpha} = \partial f_\alpha / \partial\alpha_k$. The estimation procedure of f and α follow the next steps:

1. For each t and α, obtain $\hat{f}_\alpha$ by solving $\psi_1(\eta;\alpha,t) = 0$ for η.

2. Solve $\psi_{2k} = 0$ for α and let the solution be the estimate $\hat{\alpha}$.

3. Estimate again $\hat{f}$ using $\psi_1(\eta;\hat{\alpha},t)$.

Since the estimate of α depends on f_α, an iterative approach using Fisher's scoring method (McCullagh & Nelder, 1989) is applied. Let

$$A_{ij} = E\left\{\frac{d}{d\alpha_i}\psi_{2i}(\alpha; f_\alpha)|T_1, \ldots, T_n; X_1, \ldots, X_n\right\}, \tag{2.21}$$

then an initial estimate $\tilde{\alpha}$ can be updated to $\tilde{\alpha}^\dagger$ using:

$$\tilde{\alpha}^\dagger = \tilde{\alpha} - \hat{A}(\tilde{\alpha})^{-1}\hat{B}(\tilde{\alpha}) \tag{2.22}$$

this iteration can be continued until convergence, with $\hat{A}(\tilde{\alpha})$ the $p \times p$ matrix with (i, j)'th element of A_{ij} and $\hat{B}$ the $p \times 1$ vector with k'th element given by $\psi_{2k}(\alpha; \hat{f}_\alpha)$. In the same spirit, since the estimate $\hat{f}$ depends on α, an iteration approach is needed to estimate f. So for a fixed t and α an initial estimate $\tilde{\eta}$ can be updated to $\tilde{\eta}^\dagger$ using:

$$\tilde{\eta}^\dagger = \tilde{\eta} - \frac{\psi_1(\eta; \alpha, t)}{E\left\{\frac{d}{d\eta}\psi_1(\eta; \alpha, t)|X_1, \ldots, X_n\right\}} \tag{2.23}$$

This iteration can be continued until convergence. The starting values $\tilde{\alpha}$ and $\tilde{\eta}$ can be obtained by applying the approach of McCullagh & Nelder (1989) (see Severini & Staniswalis (1994) for more details).

Assuming some regularity conditions Severini & Staniswalis (1994) proved asymptotic normality of the proposed estimators: Namely, let α_0 and σ_0^2 is the true parameters:

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \longrightarrow N(0, \sigma_0^2\Sigma_0) \quad \text{as} \quad n \longrightarrow \infty$$

Let f_0 denote the true parameter and $||\phi(t)|| = \sup_{s \in \mathcal{T}_0}|\phi(s)|$, then

$$||\hat{f}_{\hat{\alpha}} - f_0|| = o_p(n^{-1/4}) \quad \text{and} \quad \hat{\sigma}^2 = \sigma_0^2 + o_p(1) \quad \text{as} \quad n \longrightarrow \infty$$

where Σ_0 is a $p \times p$ matrix such that Σ_0^{-1} has (i, j)'th value of:

$$E_0\left\{I_1\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}Q(g(\hat{f}(T_1) + X_1\alpha); Y_1)\right\}$$

This approach presented in this subsection will be developed in Chapter 5 in the case of discrete choice models with $g(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Chapter 3

Functional linear spatial autoregressive models

Contents

Résumé en français

Dans ce chapitre nous proposons un modèle linéaire spatial fonctionnel, il s'agit d'un modèle autorégressif spatial (SAR) dont le prédicteur est de la forme fonctionnelle. Prenons des n unités spatiales localisées dans une région $\mathcal{D}_n \subset \mathcal{D} \subset \mathbb{R}^N$, où $\mathcal{D}$ est un ensemble dénombrable de type lattice. Dans chaque unité spatiale nous observons une variable de réponse Y de type réel et une variable explicative $\{X(t),\ t \in \mathcal{T}\}$ qui prend des valeurs dans un espace de dimension infinie $\mathcal{X} \subset L^2(\mathcal{T})$. La relation endogène entre Y et $X(t)$ selon le modèle SAR est donnée par :

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ijn} Y_j + \int_{\mathcal{T}} X_i(t)\theta^*(t)dt + U_i, \qquad i = 1, \ldots, n,\ \ n = 1, 2, \ldots, \tag{3.1}$$

où λ_0 est le paramètre d'autocorrélation spatiale et $\theta^*(\cdot)$ est une fonction de paramètre à estimer. Les termes d'erreurs $\{U_i,\ i = 1, \ldots, n,\ n = 1, 2, \ldots\}$ sont supposées centrées indépendantes et identiquement distribuées avec $E(U_i^2) = \sigma_0^2$.

L'ensemble $\{w_{ijn},\ i, j = 1, \ldots, n,\ n \in \mathbb{N}\}$ représente les éléments de la matrice de pondération spatiale W_n qui détermine la grandeur de connectivité entre les entités spatiales. Elle est une matrice déterministe qui est définie généralement en fonction de la distance physique entre unité i et j (voir Pinkse & Slade (1998) pour des exemples). Dans le cadre

de ce chapitre, cette matrice est supposée exogène contrairement au chapitre 4 où une structure endogène est prise en considération lors de l'estimation.

Soient Y_n et U_n deux vecteurs de dimension $n \times 1$ dont les éléments sont Y_i et U_i, $i = 1, \ldots, n$, le modèle (3.1) peut être récrit sous forme matricielle suivante :

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

où $S_n = (I_n - \lambda_0 W_n)$ avec I_n la matrice d'identité et $\mathbf{X}_n(\theta^*(.))$ est un vecteur $(n \times 1)$ dont le i-ème élément est $\int_{\mathcal{T}} X_i(t)\theta^*(t)dt$. Pour estimer les paramètres d'intérêts λ_0, σ_0^2 et le paramètre fonctionnel $\theta^*(.)$, nous chercherons les valeurs qui maximisent le logarithme de la fonction de quasi vraisemblance conditionnelle suivante :

$$\begin{aligned} L_n(\lambda_0, \theta^*(.), \sigma_0^2) = & -\frac{n}{2}\ln\sigma_0^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda_0)| \\ & - \frac{1}{2\sigma^2}\left[S_n(\lambda_0)\mathbf{Y}_n - \mathbf{X}_n(\theta^*(.))\right]' \left[S_n(\lambda_0)\mathbf{Y}_n - \mathbf{X}_n(\theta^*(.))\right], \end{aligned} \quad (3.2)$$

avec $S_n(\lambda_0) = I_n - \lambda_0 W_n$. Différentes méthodes d'estimation ont été proposées pour ce type de modèles dans le cas où la variable explicative X est à valeurs réelles notamment Lee (2004). Cet auteur a défini des estimateurs de quasi maximum de vraisemblance pour λ_0, le vecteur de paramètres θ^* et σ_0^2, en maximisant l'équivalent de (3.2). En revanche dans le cadre fonctionnel, il y a au moins deux difficultés principales : premièrement, nous n'observons pas la forme fonctionnelle de la covariable, deuxièmement la vraisemblance n'est pas calculable car la fonction de paramètres $\beta(.)$ et la covariable sont de dimension infinie.

Afin de contourner ces problèmes, nous adoptons la même technique que Muller & Stadtmuller (2005), en proposant des estimateurs à partir d'un modèle tronqué de (3.1). L'idée de cette approche est de projeter la fonction explicative et le paramètre fonctionnel dans un espace de fonctions engendré par une base de fonctions dont la dimension croit asymptotiquement avec la taille de l'échantillon n. Pour effectuer cette troncature, nous considérons une base orthonormale $\{\varphi_j,\ j = 1, 2, \ldots\}$ définie dans $L^2(\mathcal{T})$. On peut ainsi récrire $X(t)$ et $\theta^*(t)$ comme suit :

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

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

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

Nous obtenons ainsi :

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j \geq 1} \theta_j^* \varepsilon_j. \quad (3.3)$$

Pour une suite d'entiers naturels p_n qui croît asymptotiquement avec la taille d'échantillon n, Muller & Stadtmuller (2005) ont proposé une décomposition de la partie droite de (3.3) comme suit :

$$\sum_{j \geq 1} \theta_j^* \varepsilon_j = \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j + \sum_{j=p_n+1}^{\infty} \theta_j^* \varepsilon_j. \quad (3.4)$$

L'idée est d'approcher la somme infinie de la partie gauche de cette décomposition par celle finie de la partie droite de (3.4). Cela est possible à condition que l'erreur disparaisse asymptotiquement, or cette erreur est contrôlée par le carré de l'espérance mathématique du deuxième terme de la partie droite de cette décomposition :

$$E\left(\sum_{j=p_n+1}^{\infty}\theta_j^*\varepsilon_j\right)^2 = \sum_{j=p_n+1}^{\infty}\theta_j^*E(\varepsilon_j)^2. \tag{3.5}$$

En particulier, dans le cadre des bases orthonormales basées de fonctions propres, cette erreur tend vers 0 car $E(\varepsilon_j)^2 = \delta_j$, avec δ_j sont des valeurs propres. Donc, au lieu d'utiliser la variable explicative de dimension infinie, nous nous limiterons à une version de dimension p_n et le problème revient à estimer $p_n + 2$ paramètres. Le logarithme de la quasi vraisemblance conditionnelle tronquée au niveau p_n est donc

$$\begin{aligned}\tilde{L}_n(\lambda_0,\theta^*,\sigma_0^2) &= -\frac{n}{2}\ln\sigma_0^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda_0)| \\ &\quad -\frac{1}{2\sigma_0^2}\left[S_n(\lambda_0)\mathbf{Y}_n - \xi_{p_n}\theta^*\right]'\left[S_n(\lambda_0)\mathbf{Y}_n - \xi_{p_n}\theta^*\right].\end{aligned} \tag{3.6}$$

où $\theta^* \in \mathbb{R}^{p_n}$ est un vecteur de paramètres à estimer et ξ_{p_n} est une $n \times p_n$ matrice dont les éléments sont

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

Pour un λ_0 fixé, (3.6) est maximisée par :

$$\hat{\theta}_{n,\lambda_0} = (\xi_{p_n}'\xi_{p_n})^{-1}\xi_{p_n}'S_n(\lambda_0)\mathbf{Y}_n = (\hat{\theta}_{nj,\lambda})_{j=1,\ldots,p_n}$$

et

$$\begin{aligned}\hat{\sigma}_{n,\lambda_0}^2 &= \frac{1}{n}\left(S_n(\lambda_0)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda_0}\right)'\left(S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\hat{\theta}_{n,\lambda}\right) \\ &= \frac{1}{n}\mathbf{Y}_n'S_n'(\lambda_0)M_nS_n(\lambda_0)\mathbf{Y}_n,\end{aligned}$$

où $M_n = I_n - \xi_{p_n}(\xi_{p_n}'\xi_{p_n})^{-1}\xi_{p_n}'$. Par conséquent, en considérant les estimateurs $\hat{\theta}_{n,\lambda}$ et $\hat{\sigma}_{n,\lambda}^2$ de θ^* et σ_0^2 respectivement, le logarithme de la quasi vraisemblance conditionnelle tronquée, profilé par rapport à λ_0 est défini par :

$$\tilde{L}_n(\lambda_0) = -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\hat{\sigma}_{n,\lambda}^2 + \ln|S_n(\lambda_0)|.$$

Ainsi, la valeur $\hat{\lambda}_n$ qui maximise $\tilde{L}_n(\lambda_0)$ est l'estimateur de λ_0. On déduit ainsi un estimateur $\hat{\theta}_{n,\hat{\lambda}_n}$ du paramètre fonctionnel $\theta^*(t)$ comme suit :

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

Dans la suite, nous établissons les propriétés asymptotiques de l'estimateur QMLE ci-dessus proposé sous certaines conditions et hypothèses similaires à celles utilisés par Lee (2004) et adaptées au cadre fonctionnel. Nous montrons tout d'abord que les paramètres λ_0, σ_0^2 sont identifiables, consistant et vérifient

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

avec

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

où :

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

tel que $G_n = S_n^{-1} W_n$, $\Gamma_{p_n} = E\left(\frac{1}{n} \xi_{p_n}^{'} \xi_{p_n}\right)$, $\mu_4 = E(U^4)$, et $\theta^* = (\theta_1^*, \ldots, \theta_{p_n}^*)^{'}$. Nous en déduisons par la suite que le paramètre fonctionnel $\theta^*(\cdot)$, est également consistent et

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

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

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

Ces propriétés ainsi que le comportement pour échantillons à taille finie des estimateurs à travers une étude de simulations ont été prouvées. En outre, les résultats d'une application sur des données de concentration d'ozone au Sud-Est des États-Unis montrent la performance du modèle proposé ainsi que l'utilité de prendre en considération la dépendance spatiale.

Les résultats de ce chapitre sont en collaboration avec Laurence BROZE (Université de Lille), Sophie DABO-NIANG (Université de Lille) et Mohamed-Salem AHMED (Université de Lille). Ce travail est accepté pour publication en tant que chapitre de livre chez Wiley.

3.1 Introduction

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

Many fields, such as urban systems, agriculture, environmental science and economics, often consider spatially dependent data. Therefore, modeling spatial dependency in statistical inferences (estimation of the spatial distribution, regression, prediction, ...) is a significant feature of spatial data analysis. Spatial statistics provide tools to solve such modelling. Various spatial models and methods have been proposed, particularly within the scope of geostatistics or lattice data. Most of the spatial modeling methods are parametric and concern non-functional data.

Several types of functional linear models for independent data have been developed for different purposes. The most studied model is perhaps the functional linear model for scalar response, originally introduced by Hastie & Mallows (1993). Estimation and prediction problems for this model and some of its generalizations have been reported mainly for independent data (see, e.g., Crambes et al. (2009), Comte & Johannes (2012), Cai & Yuan (2012), Cuevas (2014)). Some research exists on functional spatial linear prediction using kriging methods (see, e.g., Nerini et al. (2010), Giraldo et al. (2010b), Giraldo et al. (2011), Horváth & Kokoszka (2012), Giraldo (2014), Bohorquez et al. (2016), Bohorquez et al. (2017), ...), highlighting the interest in considering spatial linear functional models.

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

We are concerned here about spatial functional models for lattice data. One of the well-known spatial lattice models is the spatial autoregressive model (SAR) of Cliff & Ord (1973), which extends regression in time series to spatial data. This model has been extensively studied and extended in several ways in the case of real-valued data, compared to the functional framework. Ruiz-Medina (2011) and Ruiz-Medina (2012) considered a spatial unilateral autoregressive Hilbertian (SARH(1)) processes where the autoregressive part is given in terms of three functional random components located in three points defining the boundary between some notions of past and future. Recently, Pineda-Ríos & Giraldo (2016) studied a functional linear model with real-valued response and a functional covariate, with SAR disturbances. Note that Tingting Huang (2019) considered after the proposition of our model a similar work.

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

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

The rest of this chapter is organized as follows. In Section 3.2, we provide the functional SAR (FSAR) and its quasi-likelihood estimator (QML). In Section 3.3, we state the consistency and asymptotic normality of the estimator. To check the performance of the estimator, numerical results are reported in Section 3.4 using different spatial scenarios, where each unit is influenced by neighboring units. Proofs and technical lemmas are given in the Appendix.

3.2 Model

We consider that at n spatial units located on $\mathcal{I}_n$, a finite subset of cardinal n of a regular or irregularly spaced, countable lattice $\mathcal{I} \subset \mathbb{R}^N$, we observe a real-valued random variable Y considered as the *response variable* and a functional covariate $\{X(t), t \in \mathcal{T}\}$, a square-integrable stochastic process on the interval $\mathcal{T} \subset \mathbb{R}$. Assume that the process $\{X(t), t \in \mathcal{T}\}$ takes values in space $\mathcal{X} \subset L^2(\mathcal{T})$, where $L^2(\mathcal{T})$ is the space of square-integrable functions in $\mathcal{T}$. The spatial dependency structure between these n spatial units is described by an $n \times n$ non-stochastic spatial weights matrix W_n that depends on n. The elements $w_{ij} = w_{ijn}$ of this matrix are usually considered as inversely proportional to the distance between spatial units i and j with respect to some metric; see Chapter 2 for more details. Since the weight matrix changes with n, we consider these observations as triangular array observations. This is required to conduct an asymptotic study of the following model that describes the relationship between the response variable Y and the covariate function $X(.)$ (Robinson, 2011).

There are mainly three different types of interaction effects that may explain why an observation associated with a specific location may be dependent on observations at other locations:

- Endogenous interaction effects, where the variable Y at some spatial unit depends on values of Y taken by other spatial units,
- Exogenous interaction effects, where the variable Y at some spatial unit depends on independent explanatory variables at other spatial units.
- Correlated effects, where similar unobserved characteristics result in similar behavior.

Here, we assume that the relationship between Y and X is modeled by the following functional spatial autoregressive model (FSAR) with endogenous interactions:

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \int_{\mathcal{T}} X_i(t)\theta^*(t)dt + U_i, \qquad i = 1, \dots, n, \; n = 1, 2, \dots \tag{3.7}$$

where the autoregressive parameter λ_0 is in compact space Λ, $\theta^*(\cdot)$ is a parameter function assumed to belong to the space of functions $L^2(\mathcal{T})$, and $(w_{ij})_{j=1,\dots,n}$ is the i-th row of W_n. Assume that $w_{ij} = O(h_n^{-1})$ uniformly in all i, j, where the rate sequence h_n can be bounded or divergent, such as $h_n = o(n)$. This kind of matrix can be obtained by Nearest Neighbor weights. In most cases it is convenient to *row normalize* the spatial weight matrix. The row-normalisation helps in the interpretation and the comparison of the parameters λ, it allows $-1 \leq \lambda \leq 1$. In this way, the spatially-lagged variables are equal to a weighted average of the neighboring values. In general, these matrices W_n can be classified into three groups: *Binary*, *Continuous* and *Hybrid*, see Chapter 2.

The disturbances $\{U_i,\ i = 1, \ldots, n,\ n = 1, 2, \ldots\}$ are assumed to be independent random variables such that $E(U_i) = 0$, $E(U_i^2) = \sigma_0^2$. They are also independent of $\{X_i(t), t \in \mathcal{T},\ i = 1, \ldots, n,\ n = 1, 2 \ldots\}$. We are interested in estimating the unknown true parameters λ_0, $\theta^*(.)$ and σ_0^2. Let $\mathbf{X}_n(\theta^*(.))$ be the $n \times 1$ vector of i-th element $\int_{\mathcal{T}} X_i(t)\theta^*(t)dt$; then, one can rewrite (3.7) as

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

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

$$\begin{aligned} L_n(\lambda, \theta(.), \sigma^2) = &-\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| \\ &- \frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\theta(.))\right]'\left[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\theta(.))\right], \end{aligned} \tag{3.8}$$

where A' denotes the transpose of matrix A.

The quasi-maximum likelihood estimates of λ_0, $\theta^*(\cdot)$ and σ_0^2 are the parameters λ, $\theta(\cdot)$, and σ^2 that maximize (3.8). But this likelihood cannot be maximized without addressing the difficulty produced by the infinite dimensionality of the explanatory random function. To solve this problem we project, as usual, the functional explanatory variable and parameter function into the space of the functions generated by a basis of functions with dimensions that increase asymptotically as the sample size tends to infinity. Several truncation techniques exist. Cardot et al. (1999) proposed the estimated eigenbasis of the sample; Cardot & Sarda (2005) considered a Spline basis, adding a penalty that controls the degree of smoothness of the parameter function. Muller & Stadtmuller (2005) proposed the use of any basis of functions that verifies some truncation criterion. We adapt the alternative proposed by Muller & Stadtmuller (2005) to solve the infinite dimension problem of the functional space. This method is denoted *truncated conditional likelihood method*.

3.2.1 Truncated conditional likelihood method

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

$$\Gamma x(t) = \int_{\mathcal{T}} E(X(t)X(s))x(s)ds\,, \qquad x \in \mathcal{X}, t \in \mathcal{T}. \tag{3.9}$$

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

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

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

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

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

$$\int_{\mathcal{T}} X(t)\theta^*(t)dt = \sum_{j=1}^{\infty} \theta_j^* \varepsilon_j = \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j + \sum_{j=p_n+1}^{\infty} \theta_j^* \varepsilon_j. \tag{3.10}$$

The truncation strategy introduced by Muller & Stadtmuller (2005) consists of approximating the left-hand side in (3.10) using only the first term of the right-hand side. This is possible when the approximation error vanishes asymptotically, where this error is controlled by a square expectation of the second term on the right-hand side of (3.10). In particular, the approximation error vanishes asymptotically when one considers the eigenbasis of the variance-covariance operator Γ by remarking that

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

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

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

Now, the truncated conditional log-likelihood function can be obtained by replacing (3.8) $\mathbf{X}_n(\theta(.))$ with $\xi_{p_n}\theta$ for all $\theta(.) \in L^2(\mathcal{T})$ and $\theta \in \mathbb{R}^{p_n}$. The corresponding and feasible log conditional likelihood is:

$$\begin{aligned}\tilde{L}_n(\lambda, \theta, \sigma^2) &= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) + \ln|S_n(\lambda)| \\ &\quad -\frac{1}{2\sigma^2}\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right]'\left[S_n(\lambda)\mathbf{Y}_n - \xi_{p_n}\theta\right].\end{aligned} \tag{3.11}$$

For a fixed λ, (3.11) is maximized at:

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

and

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

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

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

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

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

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

3.3 Assumptions and results

Let us first state some combining condition assumptions related to the spatial dependency structure and assumptions on the functional nature of the data.

Let $I_n + \lambda_0 G_n = S_n^{-1}$ where $G_n = W_n S_n^{-1}$, $B_n(\lambda) = S_n(\lambda) S_n^{-1} = I_n + (\lambda_0 - \lambda) G_n$ for all $\lambda \in \Lambda$ and $A_n(\lambda) = B_n^{'}(\lambda) B_n(\lambda)$.
We assume that

Assumption 1

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

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

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

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

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

Remark 3.1. ***Assumption 1**-i ensures that $\mathbf{Y}_n$ has mean $S_n^{-1}\mathbf{X}_n(\theta^*(\cdot))$ and variance $\sigma_0^2 S_n^{-1} S_n^{'-1}$. The uniform boundedness of W_n and S_n^{-1} in **Assumption 1**-ii enables the control of the degree of spatial correlation and plays an important role in the asymptotic properties of the estimators. By easy computation, one can show under this assumption that the matrix $G_n = W_n S_n^{-1}$ is uniformly bounded in both row and column sums together with elements of order h_n^{-1}. Consequently, the matrix $A_n(\lambda) = B_n^{'}(\lambda) B_n(\lambda)$ has a trace of order n uniformly in $\lambda \in \Lambda$ by the compactness condition of Λ in **Assumption 1**-iii. **Assumption 1**-iii makes it possible to address the nonlinearity of $ln|S_n(\lambda)|$ as a function of λ in (3.11). For more detail and a discussion of **Assumption 1**, see Lee (2004). **Assumption 2** considers the rate of convergence of p_n with respect to n. Condition iii of Assumption 2 is satisfied when one consider the eigenbasis, since in this case $E(\varepsilon_r \varepsilon_s) = 0$, for $s \neq r$.*

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

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

We have

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

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

We then have

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

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

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

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

$$\begin{aligned} E\left(\tilde{L}_n(\lambda,\theta,\sigma^2)\right) &= -\frac{1}{2\sigma^2}E\left((B_n(\lambda)\xi_{p_n}\theta^* - \xi_{p_n}\theta)'\left(B_n(\lambda)\xi_{p_n}\theta^* - \xi_{p_n}\theta\right)\right) \\ &\quad +\ln|S_n(\lambda)| - \frac{n}{2}\left(\ln\sigma^2 + \ln 2\pi\right) - \frac{\sigma_0^2}{2\sigma^2}\operatorname{tr}\left(A_n(\lambda)\right) \\ &\quad +\epsilon_{n1}\operatorname{tr}\left(B_n(\lambda)\right) + \epsilon_{n4}\operatorname{tr}\left(A_n(\lambda)\right), \end{aligned} \tag{3.12}$$

with $\epsilon_{n4} := \epsilon_{n2} + \epsilon_{n3}$. Note that the terms that contain ϵ_{n1} and ϵ_{n4} are negligible with respect to the others by **Assumption 2**.

For fixed λ, the expectation $E\left(\tilde{L}_n(\lambda,\theta,\sigma^2)\right)$ is maximum with respect to θ and σ^2 at:

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

and

$$\begin{aligned} \sigma_{n,\lambda}^{*2} &= \frac{1}{n}E\left(\left[B_n(\lambda)\xi_{p_n}\theta^* - \xi_{p_n}\theta_{n,\lambda}^*\right]'\left[B_n(\lambda)\xi_{p_n}\theta^* - \xi_{p_n}\theta_{n,\lambda}^*\right]\right) + \frac{\sigma_0^2}{n}\operatorname{tr}\left(A_n(\lambda)\right) \\ &= (\lambda_0-\lambda)^2\frac{1}{n}\Delta_n + \frac{\sigma_0^2}{n}\operatorname{tr}\left(A_n(\lambda)\right), \end{aligned} \tag{3.13}$$

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

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

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

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

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

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

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

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

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

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

Remark 3.2. ***Assumption 3*** *enables the identification of λ_0 according to the boundless of h_n. It is similar to that used in Lee (2004) in the case of multivariate deterministic covariates. This assumption ensures that $tr^2(G_n/n)$ is dominated by $tr(G_n' G_n/n)$, which is the case when $h_n \to \infty$, as under* ***Assumption 1****, $tr(G_n' G_n)$ and $tr(G_n)$ are of order $O(n/h_n)$. Situation (b) is related to the existence of a unique variance of $\mathbf{Y}_n$.* ***Assumption 4*** *characterizes the properties of the disturbance term.*

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

Theorem 3.1. *Under* ***Assumptions 1-4*** *and $h_n^4 = O(n)$ for divergent h_n, the QMLE $\hat{\lambda}_n$ derived from the maximization of $\tilde{L}_n(\lambda)$ is consistent and satisfies*

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

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

$$\begin{aligned} s_n^2 &= \sigma_0^2 \left[\theta^{*'}\Gamma_{p_n}\theta^* + \sigma_0^2\right] tr\left(G_n(G_n' + G_n)\right) + \left[3\sigma_0^2\theta^{*'}\Gamma_{p_n}\theta^* + \sigma_0^4 - \mu_4\right]\frac{1}{n}tr^2(G_n) \\ &\quad + \left[\mu_4 - 3\sigma_0^4 - \sigma_0^2\theta^{*'}\Gamma_{p_n}\theta^*\right]\sum_{i=1}^{n} G_{ii}^2. \end{aligned} \tag{3.14}$$

Note that, when h_n is divergent, the last two terms in (3.14) are negligible.

Theorem 3.2. *Under assumptions of Theorem 3.1, $\hat{\sigma}_n^2$ is a consistent estimator of σ_0^2 and satisfies*

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

with

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

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

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

Assumption 5 We have

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

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

Assumption 6 We assume that

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

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

Theorem 3.3. *Under **Assumptions 1-6**, we have*

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

Moreover, if

$$\sum_{j>p_n} E\left(\varepsilon_j^2\right)\left(\int_{\mathcal{T}}\theta^*(t)\varphi_j(t)dt\right)^2 = o(\sqrt{p_n}/n), \tag{3.15}$$

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

$$\frac{nd^2\left(\hat{\theta}_n(\cdot),\theta^*(\cdot)\right)-p_n}{\sqrt{2p_n}} \to \mathcal{N}(0,\sigma_0^4), \tag{3.16}$$

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

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

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

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

3.4 Numerical experiments

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

Recall that the truncation strategy requires an appropriate selection of orthonormal basis. This basis can be chosen to be a fixed orthonormal basis, such as the Fourier basis, or it can be constructed by estimating the eigenfunctions of the covariance kernel (3.9) and applying functional principal component analysis (FPCA) to the explanatory random functions X_i. We use the eigenfunctions obtained from the FPCA to construct the expansion basis in this numerical section. The eigenfunctions are those of the integral operator associated with the integral kernel defined by the variance-covariance function of X, which is estimated for each $t, v \in [0, 1]$ as follows:

$$\hat{K}(t, v) = \frac{1}{n-1} \sum_{i=1}^{n} X_i(t) X_i(v). \tag{3.17}$$

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

$$\text{ASE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \tag{3.18}$$

the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The choice of p using AIC is consistent in the setting of functional linear models, see Muller & Stadtmuller (2005) for more details. Notice that we use a pre-selected p based on the cumulative inertia. We focus on the selection of p from among those associated with cumulative inertia values lower than some threshold (here 95%).

As measure of accuracy of the parameter function, (see M. Escabias & Valderrama, 2007) the usual integrated mean square error,

$$\text{IMSE} = \int_0^1 \left(\theta^*(t) - \hat{\theta}_n(t) \right)^2 dt, \tag{3.19}$$

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

3.4.1 Monte Carlo simulations

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

- **Scenario 1**: The spatial weight matrix W_n is constructed by taking the k neighbors of each unit using kNN method (k nearest neighbors algorithm). Let us take $k = \{4, 8\}$.
- **Scenario 2**: We consider a number of districts r (block or group) with m members in each district, where the units of the same district have the same weight. As in Case (1993), we can define the spatial weight matrix as block diagonal $W_n = I_r \otimes B_m$, where $\otimes$ is the Kronecker product, $B_m = (l_m l_m^{'} - I_m)/(m-1)$, and l_m is an m vector of 1.

The simulations are performed based on the following data:

$$Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \int_{\mathcal{T}} X_i(t) \theta^*(t) dt + U_i \tag{3.20}$$

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

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

$$X(t) = \sum_{j=1}^{20} \varepsilon_j \varphi_j(t), \tag{3.21}$$

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

We apply the truncation strategy to reduce the infinite dimensionality of our model $Y_i = \lambda_0 \sum_{j=1}^{n} w_{ij} Y_j + \sum_{j=1}^{p_n} \theta_j^* \varepsilon_j^{(i)} + \sum_{j=p_n+1}^{\infty} \theta_j^* \varepsilon_j^{(i)} + U_i,\ i = 1, \ldots, n,\ \ n = 1, 2, \ldots$ to a p_n-finite linear approximation and compute the quasi-likelihood estimator. The parameters λ_0, σ_0^2 and $\theta_1^*, ..., \theta_{p_n}^*$ are estimated by solving the score equations defined in Section 3.3. Different sample sizes, $n = \{100, 200, 400\}$, are tested for the first scenario; for the second, we take $r = \{10, 20, 30\}$ and $m = \{5, 10, 15\}$, with sample size $n = m \times r$.

The studied models are replicated 200 times, and the results of Scenario 1 are presented in Tables 3.1 and 3.2, respectively, for $k = 4$ and $k = 8$. For Scenario 2, the results are reported in Tables 3.3 to 3.6. Each table represents a specific model. In each table, the rows λ, σ^2, IMSE and PCs give the averages over these replications (with the standard deviation in brackets) of the autoregressive parameter estimate $\hat{\lambda}_n$, the standard deviation parameter $\hat{\sigma}_n^2$, the associated IMSE defined in (3.19) and the number p of eigenfunctions (used in the truncation), respectively. For the different models, the strategies used to select p yield (on average) values close to the true parameter of $p = 3$, especially for ASE and AIC and large sample sizes (see the columns titled PCs in Tables 3.1-3.6). The parameter function estimates are in given in Figures 3.2-3.3. For all the models, the three methods used to select p and two spatial scenarios, the performance of the parameter function and the variance estimates varies with the sample size.

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

In all the cases considered, AIC criterion outperforms. This is in adequacy with Muller & Stadtmuller (2005)'s finite and large samples results.

Table 3.1: Estimation of parameters with $n = \{100, 200, 400\}$, $k = 4$

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

3.4.2 Real data application

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

We organize the original space-time series into a set of daily functional data to apply the functional methodology. Let us consider at each station a response variable Y as the ozone concentration at 12pm on July 20 and a covariate function $\{X(t), t \in [0, 23]\}$ corresponding to the 24 records of ozone concentrations from 12pm on July 19 to 11am July 20. Figure 3.4 presents the geographical positions of the 106 stations (red points) and the curves of the ozone concentration from 12pm July 19 to 11am July 20. To highlight the performance of the spatial FSARM model, we compare with the usual functional linear model (FLM), that does not take into account any spatial structure in the estimation procedure.

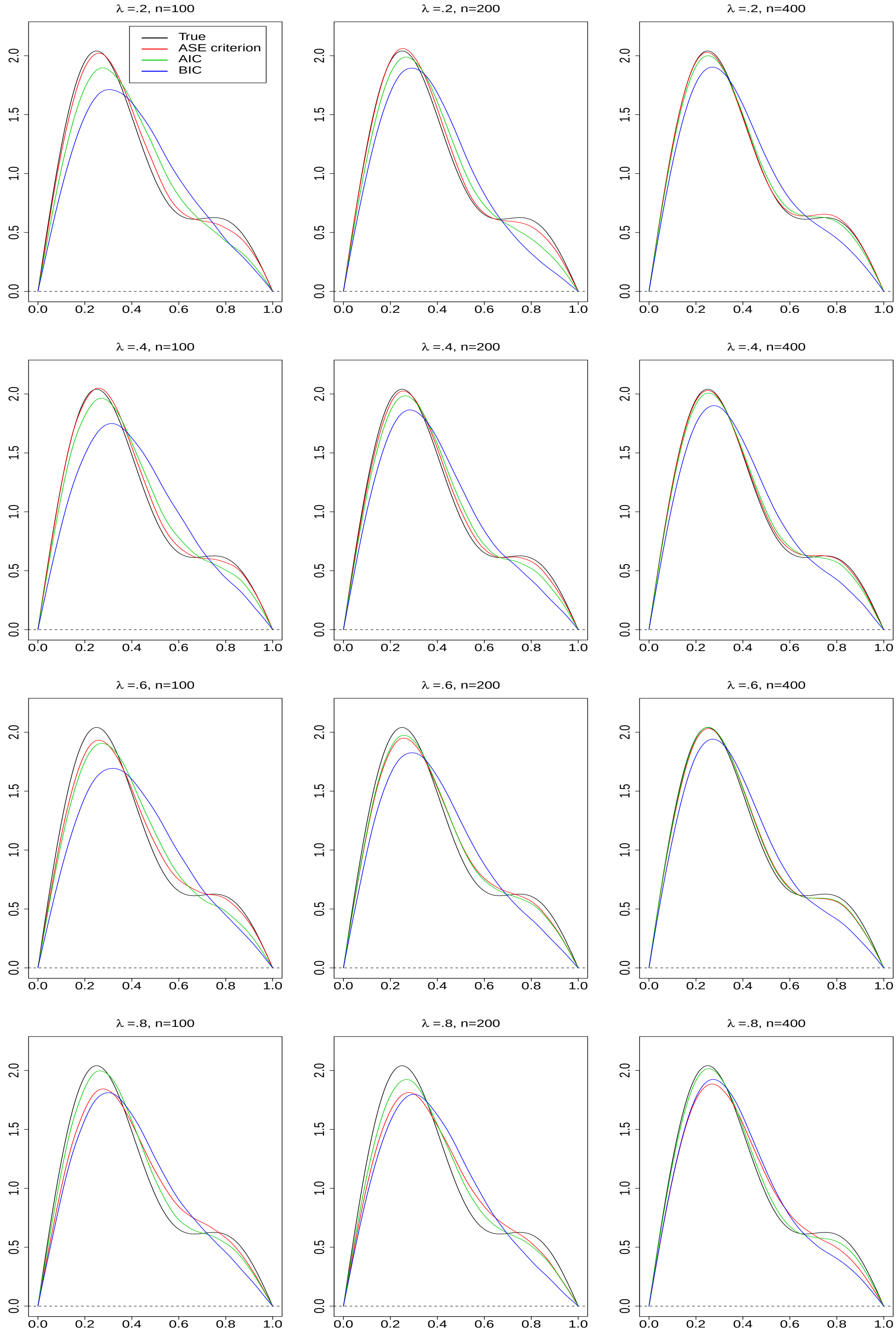

Figure 3.1: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria and $k = 4$.

Table 3.2: Estimation of parameters with $n = \{100, 200, 400\}$, $k = 8$

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

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

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

where d_{ij} is the euclidean distance between station i and station j, and ρ is some cut-off distance chosen such that each station has at least four neighbors. Other weight matrices have been tested, but we choose to present the results corresponding to this matrix.

Note that FPCA is used to smooth the curves before we reduce the spatial dimensions of the functional covariate using the eigenbasis, as explained above (see Figure 3.5). The AIC is used to select the number of eigenfunctions. For the two models, we have the same optimal number of eigenfunctions $p = 3$. Table 3.7 and Figure 3.6 give the estimation results of the FLM and FSARM. Note that the curves obtained by the two estimation

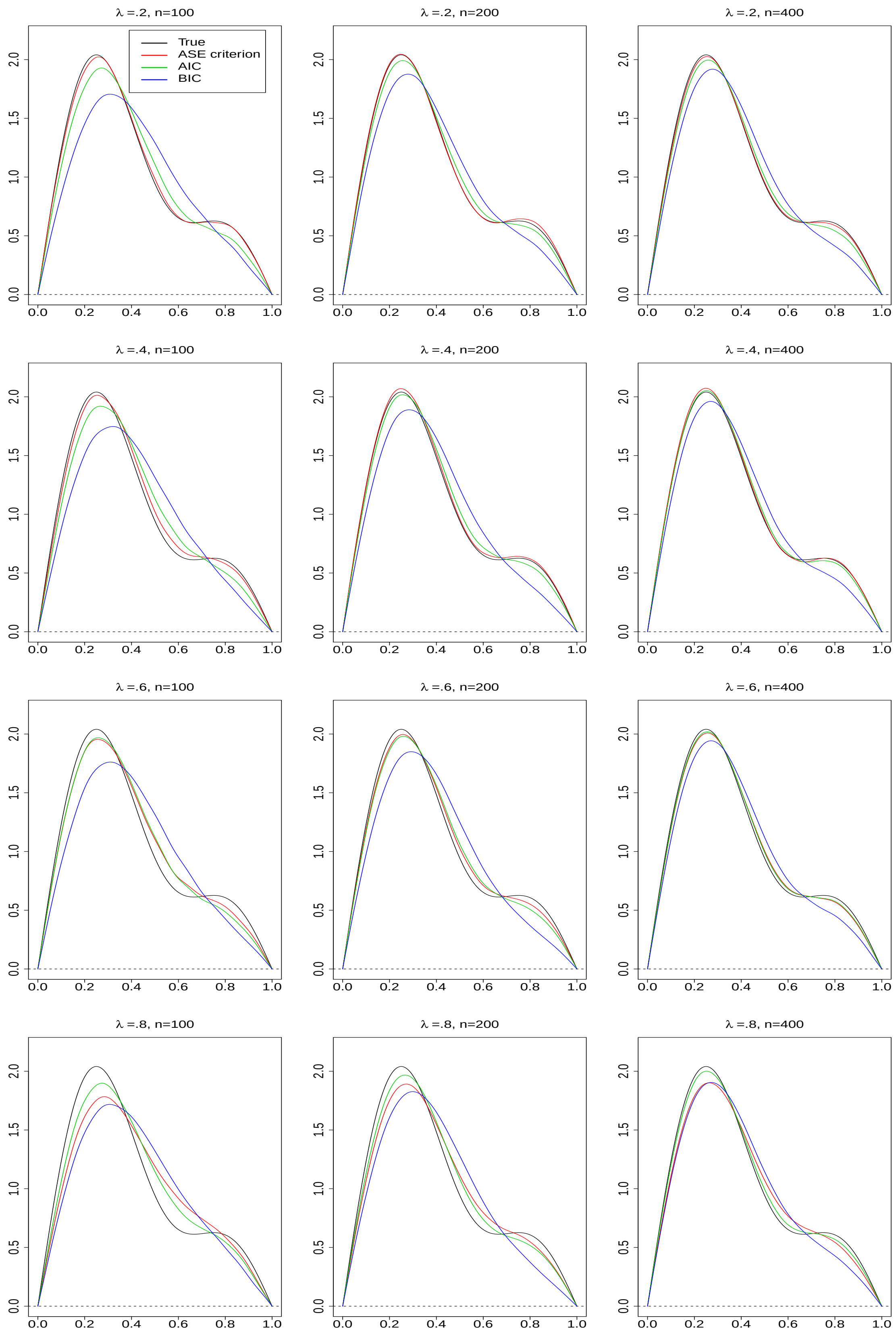

Figure 3.2: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria and $k = 8$.

Table 3.3: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.2$.

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

Table 3.4: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.4$.

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

Table 3.5: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.6$.

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

Table 3.6: Estimation of parameters associated to scenario 2 with $\lambda_0 = 0.8$.

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

methods are similar, with small differences around 12pm and 7pm. The FLM gives an intercept estimate close to zero, while with FSARM, we have a spatial structure with an estimated autoregressive parameter close to 0.2.

Now, let us consider the following problem of prediction. At a given station s_0, we aim to predict the ozone concentration every hour, from 12am to 11pm, on July 20, 2015. For this aim, assume that at s_0, we observe only the 24 records of ozone concentration from 12am to 11pm on July 19, 2015 and we would like to predict the ozone concentration of the following day, that is, from 12am to 11pm on July 20, 2015. To obtain these predictions, we proceed as follows.

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

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

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

We randomly select 4 stations among the 106 and apply the prediction procedure. Figure 3.7 presents the prediction results; the true values are in black, while the predictions are in red for the FSARM model and in blue for the FLM (with no spatial structure) model. FSARM achieves some improvements, particularly around 12pm, when the ozone concentration is higher.

Table 3.7: Estimated parameters for FLM and FSARLM.

	PCs	Autoregressive parameter	Intercept
FSARLM	3	0.19	
FLM	3		0.006

3.5 Conclusion

This chapter proposes a spatial functional linear regression function for functional random field covariates. Our main theoretical contribution was to study the consistency and asymptotic normality of the estimator. One can see the proposed methodology as an extension of the real-valued SAR model to functional data. More precisely, it is apparent that the proposed estimation approach based on a truncation technique is particularly well adapted to spatial regression estimation for functional data in the presence of spatial dependence. This good behavior is observed both from an asymptotic point of view and from a numerical study. This work offers interesting perspectives for investigation. Future work will be tied to generalized functional linear spatial models (see, for instance Kelejian & Prucha, 1998; Muller & Stadtmuller, 2005). Also, an adaptation of this method to issues using different covariates (functional and non-functional) with or without a spatial weight matrix with correlated errors could be developed. The application of the proposed regression estimator to additional real data, will be investigated.

3.6 Appendix

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

Proposition 3.1. *Assume* ***Assumptions 1-3****.*

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

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

Proof of Proposition 3.1

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

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

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

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

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

with

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

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

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

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

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

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

By the mean value theorem, $\frac{1}{n}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = \frac{1}{n}\text{tr}(W_nS_n^{-1}(\bar{\lambda}_n))(\lambda_2-\lambda_1)$, where $\bar{\lambda}_n$ lies between λ_1 and λ_2. By the uniform boundedness of **Assumption 1**-iii,

$\mathrm{tr}(W_n S_n^{-1}(\bar{\lambda}_n)) = O(n/h_n)$. Thus, $\frac{1}{n}\ln|S_n(\lambda)|$ is uniformly equicontinuous in λ in Λ. As Λ is a bounded set, $\frac{1}{n}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = O(1)$ uniformly on λ_1 and λ_2.

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

Suppose that $\sigma^2_{n,\lambda}$ is not uniformly bounded away from zero on Λ. Then there would exist a sequence $\{\lambda_n\}$ in Λ such that $\lim_{n\to\infty} \sigma^2_{n,\lambda_n} = 0$. Since we have $Q_{n,0}(\lambda) - Q_{n,0}(\lambda_0) \leq 0$ for all λ and $\frac{1}{n}(\ln|S_n(\lambda_0)| - \ln|S_n(\lambda)|) = O(1)$ uniformly on Λ, then $-\frac{1}{2}\ln\sigma^2_{n,\lambda} \leq -\frac{1}{2}\ln\sigma_0^2 - \frac{1}{n}(\ln|S_n(\lambda_0)| - \ln|S_n(\lambda)|) = O(1)$. That is, $-\frac{1}{2}\ln\sigma^2_{n,\lambda}$ is bounded, and this is a contradiction with the previous statement. Therefore, $\sigma^2_{n,\lambda}$ must be bounded away from zero uniformly on Λ.

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

We have to show that $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous on Λ. The parameter $\sigma^{*2}_{n,\lambda}$ (see (3.13)) is uniformly bounded on Λ because it is a quadratic form of λ, and its components $\frac{1}{n}\Delta_n$, $\frac{1}{n}\mathrm{tr}(G_n)$ and $\frac{1}{n}\mathrm{tr}(G_n G_n')$ are bounded by **Assumption 1** (i-ii). The uniform continuity of $\ln\sigma^{*2}_{n,\lambda}$ on Λ then follows because $1/\sigma^{*2}_{n,\lambda}$ is uniformly bounded on Λ since $\sigma^{*2}_{n,\lambda} \geq \sigma^2_{n,\lambda}$ for all $\lambda \in \Lambda$ by **Assumption 3**. Hence, $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous.

Let us prove uniqueness of the maximum λ_0. Remark that

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

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

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

Because $\bar{N}_\epsilon(\lambda_0)$ is a compact set, there exists a convergent subsequence λ_{n_m} of λ_n. Let λ_+ be the limit of this subsequence in Λ.

Now, as $\frac{1}{n}Q_n(\lambda)$ is uniformly equicontinuous in λ,

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

This is possible only if

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

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

$$\begin{aligned} &\lim_{n\to\infty}\left\{\frac{1}{n}\left(\ln|S_n(\lambda)| - \ln|S_n|\right) + \frac{1}{2}\left(\ln\sigma^2_{n,\lambda} - \ln\sigma_0^2\right)\right\} \\ &\qquad = \lim_{n\to\infty} \frac{1}{n}\left\{Q_{n,0}(\lambda) - Q_{n,0}(\lambda_0)\right\} \neq 0 \qquad \text{for all } \lambda \neq \lambda_0. \end{aligned}$$

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

Let us prove that

$$\sup_{\lambda\in\Lambda}\frac{1}{n}\left|\tilde{L}_n(\lambda)-Q_n(\lambda)\right| = o_p(1). \tag{3.22}$$

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

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

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

$$\hat{\sigma}^2_{n,\lambda}-\sigma^{*2}_{n,\lambda} = o_p(1). \tag{3.23}$$

Equation (3.23) combined with the fact that $\sigma^{*2}_{n,\lambda}$ is bounded away from zero uniformly on Λ implies that $\hat{\sigma}^2_{n,\lambda}$ is bounded away from zero uniformly on Λ in probability. Hence,

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

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

Let

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

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

Note that

$$\begin{aligned} \hat{\sigma}^2_{n,\lambda}-\sigma^{*2}_{n,\lambda} &= \frac{1}{n}\mathbf{Y}^{'}_nS^{'}_n(\lambda)M_nS_n(\lambda)\mathbf{Y}_n-\sigma^{*2}_{n,\lambda}\\ &= (\lambda_0-\lambda)^2H_{n0}+2(\lambda_0-\lambda)H_{1n}(\lambda)+H_{n2}(\lambda)-\sigma^2_{n,\lambda}\\ &\qquad +H_{n3}(\lambda)+H_{n4}(\lambda), \end{aligned} \tag{3.24}$$

where

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

and

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

Note that the parameter function $\theta^*(\cdot)$ is square integrable; therefore, $\|\theta^*\|_2<\infty$. Then, by Lemma 3.1 and 3.2,

$$H_{n0} = O_p\left(\frac{p_n}{h_n\sqrt{n}}\right). \tag{3.25}$$

Also, Lemma 3.3 implies that $H_{n3}(\lambda)$ and $H_{n4}(\lambda)$ are of order $o_p(1)$ uniformly on λ in Λ.

In the following, we show that $H_{n1}(\lambda)$ and $H_{n2}(\lambda) - \sigma^2_{n,\lambda}$ are all of order $o_p(1)$ for all $\lambda \in \Lambda$.

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

Note that

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

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

$$\xi_{p_n}' \mathbf{U}_n = O_p(\sqrt{n}) \qquad \text{and} \qquad \mathbf{U}_n' G_n \xi_{p_n} = O_p\left(\sqrt{\frac{n}{h_n}}\right), \tag{3.26}$$

by **Assumption 1**-ii. In addition, by Lemma 3.1, we have

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

and

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

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

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

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

■

Proof of $H_{n2}(\lambda)$:

For each $\lambda \in \Lambda$, we have

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

with

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

Similar to (3.26), we have

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

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

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

and

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

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

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

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

■

Proof of (ii):

We start to show the following convergence

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

Recall that,

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

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

and

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

Then, we have

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

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

Note that

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

and

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

This implies that

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

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

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

We have

$$\frac{h_n}{n}\left(\mathbf{V}_n^{'}G_n^{'}M_nG_n\mathbf{V}_n-\Delta_n\right)=h_nH_{n0}=O_p\left(\frac{p_n}{\sqrt{n}}\right). \tag{3.27}$$

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

$$\sqrt{\frac{h_n}{n}}\mathbf{V}_n^{'}G_n^{'}M_n\left[I_n+(\lambda_0-\lambda)G_n\right]\mathbf{U}_n=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right). \tag{3.28}$$

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

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

$$\sqrt{\frac{h_n}{n}}\left[\mathbf{U}_n^{'}G_n^{'}M_nG_n\mathbf{U}_n-\sigma_0^2\mathrm{tr}(G_n^{'}G_n)\right]=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right). \tag{3.29}$$

Therefore, by Lemma 3.3, we may write

$$\sqrt{\frac{h_n}{n}}\left\{\mathbf{Y}_n^{'}W_n^{'}M_nS_n(\lambda_n)\mathbf{Y}_n-(\lambda_0-\lambda_n)\Delta_n-\sigma_0^2\mathrm{tr}\left(G_n^{'}B_n(\lambda_n)\right)\right\}=O_p\left(1+\frac{p_n}{\sqrt{h_n}}\right). \tag{3.30}$$

Note that when h_n is unbounded, we have

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

since $\mathrm{tr}(G_n)$ and $\mathrm{tr}(G_n^{'}G_n)$ are of order $O(n/h_n)$. Thus, $1/\sigma^{*2}_{n,\lambda}=O(1)$ uniformly in λ, because $\sigma^{*2}_{n,\lambda}\geq\sigma^2_{n,\lambda}$ and $\sigma_0^2>0$. However, we have also $1/\hat{\sigma}^2_{n,\lambda}=O_p(1)$ by (3.23).

Now, note that under **Assumption 1** (ii-iii), Δ_n and $\mathrm{tr}(G_n' B_n(\lambda))$ are of order $O(n/h_n)$ and using (3.23) and (3.30), we conclude

$$\frac{h_n}{n}\left\{\left(\tilde{L}_n(\lambda)-\tilde{L}_n(\lambda_0)\right)-\left(Q_n(\lambda)-Q_n(\lambda_0)\right)\right\}=o_p(1), \tag{3.31}$$

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

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

Recall that

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

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

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

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

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

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

Then, the function

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

is uniformly equicontinuous on Λ because $\mathrm{tr}(W_nS_n^{-1}(\lambda)) = O(n/h_n)$ uniformly on λ by **Assumption 1**.

In conclusion, $\frac{h_n}{n}\left(Q_n(\lambda)-Q_n(\lambda_0)\right)$ is uniformly equicontinuous on Λ.

Let us prove uniqueness of the maximum λ_0.

Let

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

Then,

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

We have by the Taylor expansion,

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

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

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

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

Therefore, under **Assumption 3** (a),

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

for any $\lambda \neq \lambda_0$. Furthermore, under **Assumption 3** (b), $D_n(\lambda) < 0$, if $\lambda \neq \lambda_0$.

In conclusion, for a certain rank, we have $\frac{h_n}{n}\left(Q_n(\lambda) - Q_n(\lambda_0)\right) < 0$, when $\lambda \neq \lambda_0$.

The proof of (ii) follows from the uniform convergence (3.31) and the identification uniqueness condition.

■

Proof of Theorem 3.1

Identification and consistency of $\hat{\lambda}_n$ are given by Proposition 3.1. Let us now focus on the asymptotic normality of $\hat{\lambda}_n$.

Consider the first and second order derivatives of the concentrated log likelihood $\tilde{L}_n(\lambda)$:

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

and

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

By (3.28) and Lemma 3.3, we have

$$\frac{h_n}{n}\mathbf{Y}'_n W'_n M_n W_n \mathbf{Y}_n = \frac{h_n}{n}\mathbf{V}'_n G'_n M_n G_n \mathbf{V}_n + \frac{h_n}{n}\mathbf{U}'_n G'_n M_n G_n \mathbf{U}_n + o_p(1), \tag{3.32}$$

and

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

by (3.30) and since under **Assumption 1**, Δ_n and $\mathrm{tr}(G_n B_n(\lambda))$ are of order $O_p(n/h_n)$, uniformly in λ.

From (3.23), we proved that $\hat{\sigma}^2_{n,\lambda} = \sigma^{*2}_{n,\lambda} + o_p(1)$. Thus, we have

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

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

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

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

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

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

Finally, using (3.27), (3.29), and the fact that $\sigma^{*2}_{n,\lambda_0} = \sigma_0^2$, we have

$$\frac{h_n}{n}\frac{\partial^2 \tilde{L}_n(\lambda_0)}{\partial\lambda^2} = -\frac{1}{\sigma_0^2}\frac{h_n}{n}\Delta_n - \frac{h_n}{n}\left[\text{tr}(G_n' G_n) + \text{tr}\left(G_n^2\right)\right] + o_p(1). \tag{3.33}$$

Let us now prove the asymptotic normality of $\sqrt{\frac{h_n}{n}}\frac{\partial \tilde{L}_n(\lambda_0)}{\partial\lambda}$.

Using the results of Lemma 3.3, we have

$$\sqrt{\frac{h_n}{n}}\mathbf{Y}_n' W_n' M_n S_n \mathbf{Y}_n = \sqrt{\frac{h_n}{n}}\left[\mathbf{V}_n' + \mathbf{U}_n'\right] G_n' M_n \mathbf{U}_n + o_p(1), \tag{3.34}$$

and

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

It follows that

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

where $C_n = G_n - \text{tr}(\frac{G_n}{n})I_n$. Using (3.26), we have

$$\sqrt{\frac{h_n}{n}}\mathbf{U}_n' C_n' \xi_{p_n}(\xi_{p_n}'\xi_{p_n})^{-1}\xi_{p_n}'\mathbf{U}_n = O_p\left(\frac{p_n}{\sqrt{n}}\right), \tag{3.35}$$

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

Consider the following decomposition

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

by (3.26) and Lemma 3.1. Thus

$$\sqrt{\frac{h_n}{n}}\mathbf{V}_n^{'}G_n^{'}\xi_{p_n}(\xi_{p_n}^{'}\xi_{p_n})^{-1}\xi_{p_n}^{'}\mathbf{U}_n = \frac{\sqrt{h_n}}{n}\mathrm{tr}\left(G_n\right)\frac{\mathbf{V}_n^{'}\mathbf{U}_n}{\sqrt{n}} + O_p\left(\frac{p_n^2}{\sqrt{nh_n}}\right). \tag{3.36}$$

Consequently, (3.35) and (3.36) imply

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

with $D_n = G_n + \mathrm{tr}(\frac{G_n}{n})I_n$.

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

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

$$\mathbf{V}_n^{'}D_n^{'}\mathbf{U}_n + \mathbf{U}_n^{'}C_n^{'}\mathbf{U}_n = \sum_{i=1}^{n} Z_{ni}, \tag{3.37}$$

with

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

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

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

Finally, let

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

Note that condition C.1 in Lemma 3.5 implies that $\left\{\tilde{Z}_{ni},\, i = 1, \ldots, n\; n = 1, 2, \ldots\right\}$ form a triangular array of martingale differences sequences. According to Kelejian & Prucha (Theorem A.1, 2001, p.240) and under conditions C.2 and C.3 in Lemma 3.5, we have

$$\sqrt{\frac{h_n}{n}}\frac{\partial \tilde{L}_n(\lambda_0)}{\partial \lambda} = \frac{s_Z}{\hat{\sigma}^2_{n,\lambda_0}}\sum_{i=1}^{n}\tilde{Z}_{ni} + o_p(1) \to \mathcal{N}\left(0, \frac{s_Z^2}{\sigma_0^4}\right). \tag{3.38}$$

Finally, using (3.33) and (3.38) we can conclude by the Taylor expansion, that

$$\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, s_\lambda^2), \tag{3.39}$$

where

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

This concludes the proof of Theorem 3.1.

■

Proof of Theorem 3.2

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

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

Lemma 3.3 and (3.35) imply that

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

Thus

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma^2_0) &= \sqrt{\frac{n}{h_n}}(\lambda_0 - \hat{\lambda}_n)^2 \frac{\sqrt{h_n}}{n}\mathbf{Y}^{'}_n W^{'}_n M_n W_n \mathbf{Y}_n \\ &\quad -2\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0)\frac{\sqrt{h_n}}{n}\mathbf{Y}^{'}_n W^{'}_n M_n S_n \mathbf{Y}_n + \frac{1}{\sqrt{n}}(\mathbf{U}^{'}_n \mathbf{U}_n - n\sigma^2_0).\end{aligned}$$

Note that (3.28), (3.34) and (3.35) imply

$$\frac{\sqrt{h_n}}{n}\mathbf{Y}^{'}_n W^{'}_n M_n S_n \mathbf{Y}_n = \frac{\sqrt{h_n}}{n}\mathrm{tr}(G_n) + o_p(1) = O_p\left(\frac{1}{\sqrt{h_n}}\right). \tag{3.40}$$

By (3.27), (3.29) and (3.32), we have

$$\frac{\sqrt{h_n}}{n}\mathbf{Y}^{'}_n W^{'}_n M_n W_n \mathbf{Y}_n = \frac{\sqrt{h_n}}{n}\Delta_n + \sigma^2_0 \frac{\sqrt{h_n}}{n}\mathrm{tr}\left(G_n G^{'}_n\right) + o_p(1) = O_p\left(\frac{1}{\sqrt{h_n}}\right).$$

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

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

If $\lim_{n\to\infty} h_n = \infty$, (3.40) will be of order $o_p(1)$. Hence

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma^2_0) = \frac{1}{\sqrt{n}}(\mathbf{U}^{'}_n \mathbf{U}_n - n\sigma^2_0) + o_p(1) \to \mathcal{N}(0, \mu_4 - \sigma^4_0).$$

Otherwise, we have

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma^2_0) &= \frac{1}{\sqrt{n}}(\mathbf{U}^{'}_n \mathbf{U}_n - n\sigma^2_0) \\ &\quad -2\frac{\sqrt{h_n}}{n}\mathrm{tr}(G_n)\sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) + o_p(1).\end{aligned} \tag{3.41}$$

By the asymptotic normality proof of $\hat{\lambda}_n$ (see (3.33) and (3.37)), one can conclude

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

where

$$\delta_n = \frac{n}{h_n}\left[\Delta_n + \sigma^2_0 \mathrm{tr}\left(G_n(G^{'}_n + G_n)\right)\right]^{-1}.$$

Therefore, on can rewrite (3.41) as

$$\sqrt{n}(\hat{\sigma}^2_{n,\hat{\lambda}_n} - \sigma_0^2) = 2\delta_n \frac{\sqrt{h_n}}{n} \mathrm{tr}(G_n) \sqrt{\frac{n}{h_n}} \sum_{i=1}^{n} Z^{\dagger}_{ni} + o_p(1), \tag{3.42}$$

where

$$Z^{\dagger}_{ni} = D_{ii} U_i V_i + \tilde{C}_{ii}(U_i^2 - \sigma_0^2) + 2U_i \sum_{j=1}^{i-1} G^s_{ij} T_j,$$

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

It is easy to show that

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

Let

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

Note that conditions C.1-C.3 in Lemma 3.5 hold when Z_{ni} and $\tilde{Z}_{ni}$ are replaced by $Z^{\dagger}_{ni}$ and $\tilde{Z}^{\dagger}_{ni}$ respectively. Therefore, Kelejian & Prucha (Theorem A.1, 2001, p.240) implies that

$$\sum_{i=1}^{n} \tilde{Z}^{\dagger}_{ni} \to \mathcal{N}(0,1). \tag{3.43}$$

Finally, by (3.42) and (3.43), we have

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

where

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

This finishes the proof.

■

Proof of Theorem 3.3

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

$$\hat{\theta}_{n,\hat{\lambda}_n} = (\xi'_{p_n}\xi_{p_n})^{-1}\xi'_{p_n} S_n(\hat{\lambda}_n)\mathbf{Y}_n. \tag{3.44}$$

By Lemma 3.3, we have

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

By Lemma 3.1, we have

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

The asymptotic normality result of $\hat{\lambda}_n$ and (3.26), imply that

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

Hence,

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

Therefore,

$$\begin{aligned} & n\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right)'\Gamma_{p_n}\left(\hat{\theta}_{n,\hat{\lambda}_n}-\theta^*\right)\\ &= \left\{\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right]\right\}'\Gamma_{p_n}\left\{\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right]\right\}\\ &\quad +2\sqrt{n}(\lambda_0-\hat{\lambda}_n)\mathrm{tr}\left(\frac{G_n}{n}\right)\theta^{*\prime}\Gamma_{p_n}\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right]\\ &\quad +n(\lambda_0-\hat{\lambda}_n)^2\mathrm{tr}^2\left(\frac{G_n}{n}\right)\theta^{*\prime}\Gamma_{p_n}\theta^*+o_p(1).\end{aligned} \tag{3.45}$$

Consider the last two terms in (3.45), we have by the asymptotic normality of $\hat{\lambda}_n$

$$n(\lambda_0-\hat{\lambda}_n)^2\mathrm{tr}^2\left(\frac{G_n}{n}\right)\theta^{*\prime}\Gamma_{p_n}\theta^*=O_p\left(\frac{1}{h_n}\right). \tag{3.46}$$

In addition, by (3.26) and Lemma 3.1, we have

$$\sqrt{n}(\lambda_0-\hat{\lambda}_n)\mathrm{tr}\left(\frac{G_n}{n}\right)\theta^{*\prime}\Gamma_{p_n}\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\mathbf{U}_n}{\sqrt{n}}\right]=O_p\left(\frac{1}{\sqrt{h_n}}\right). \tag{3.47}$$

Let us now give the asymptotic distribution of the first term in (3.45). Let

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

and consider the following decomposition

$$\begin{aligned}\left\{\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\tilde{\mathbf{U}}_n}{\sqrt{n}}\right]\right\}'\Gamma_{p_n}\left\{\left(\frac{\xi'_{p_n}\xi_{p_n}}{n}\right)^{-1}\left[\frac{\xi'_{p_n}\tilde{\mathbf{U}}_n}{\sqrt{n}}\right]\right\} &= \mathcal{X}'_n\Psi_n^2\mathcal{X}_n\\ &= \mathcal{X}'_n\mathcal{X}_n-2\mathcal{X}'_n(I_{p_n}-\Psi_n)\mathcal{X}_n\\ &\quad +\mathcal{X}'_n(I_{p_n}-\Psi_n)^2\mathcal{X}_n.\end{aligned} \tag{3.48}$$

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

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

Thus, we deduce by (3.26) and Lemma 3.4, that

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

Therefore,

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

by (3.45), (3.46) and (3.47). This yields (3.16) and completes the proof of Theorem 3.3.

■

Lemma 3.1. *Assume that $E\left(\varepsilon_i^4\right)$ is finite, where $\varepsilon_i = \int X(t)\varphi_i(t)dt$. Under **Assumption 1**, we have*

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

and

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

Proof of Lemma 3.1

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

Note that

$$\begin{aligned}
&E\left(\left\|\xi_{p_n}'G_n\xi_{p_n} - E\left(\xi_{p_n}'G_n\xi_{p_n}\right)\right\|^2\right)\\
&= \sum_{\substack{i_1=1\\ j_1=1}}^{n}\sum_{\substack{i_2=1\\ j_2=1}}^{n}\sum_{r=1}^{p_n}\sum_{s=1}^{p_n} G_{i_1j_1}G_{i_2j_2}\left[E\left(\varepsilon_s^{(i_1)}\varepsilon_r^{(j_1)}\varepsilon_s^{(i_2)}\varepsilon_r^{(j_2)}\right)\right.\\
&\qquad\qquad \left. - E\left(\varepsilon_s^{(i_1)}\varepsilon_r^{(j_1)}\right)E\left(\varepsilon_s^{(i_2)}\varepsilon_r^{(j_2)}\right)\right]\\
&= \sum_{i=1}^{n}G_{ii}^2\sum_{r=1}^{p_n}\sum_{s=1}^{p_n}\mathrm{Cov}\left(\varepsilon_r^2,\varepsilon_s^2\right) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}^2\sum_{r=1}^{p_n}\sum_{s=1}^{p_n}E\left(\varepsilon_s^2\right)E\left(\varepsilon_r^2\right)\\
&\qquad + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}G_{ji}\sum_{r=1}^{p_n}\sum_{s=1}^{p_n}E\left(\varepsilon_s\varepsilon_r\right)E\left(\varepsilon_s\varepsilon_r\right)\\
&= O\left(p_n^2\sum_{i=1}^{n}G_{ii}^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}G_{ij}G_{ji}\right)\\
&= O\left(p_n^2\frac{n}{h_n^2} + \|G_n\|^2 + \left|\mathrm{tr}\left(G_n^2\right)\right|\right) = O\left(\frac{n}{h_n^2}(p_n^2 + h_n)\right),
\end{aligned}$$

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

■

Lemma 3.2. *Assume that $E\left(\varepsilon_i^4\right)$ is finite, where $\varepsilon_i = \int X(t)\varphi_i(t)dt$. Under **Assumption 1**, we have*

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

Proof of Lemma 3.2

Note that

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

by Lemma 3.1.

■

Lemma 3.3. *Under **Assumptions 1-2**, we have*

$$\sqrt{\frac{h_n}{n}}\mathbf{U}_n' G_n' M_n G_n \mathbf{R}_n = o_p(1), \tag{3.49}$$

$$\sqrt{\frac{h_n}{n}}\mathbf{R}_n' M_n G_n \xi_{p_n} = o_p(1), \tag{3.50}$$

$$\sqrt{\frac{h_n}{n}}\mathbf{R}_n' G_n' M_n G_n \mathbf{R}_n = o_p(1). \tag{3.51}$$

Proof of Lemma 3.3

Let

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

Consider (3.49), and note that by **Assumption 1**,

$$E\left(\left\|\mathbf{R}_n' G_n \xi_{p_n}\right\|^2\right) = O\left(\frac{n}{h_n^2}\left[h_n E(R^2) + \pi_{n1} + n\pi_{n2}\right]\right), \tag{3.52}$$

$$E\left(\left\|\mathbf{R}_n' \xi_{p_n}\right\|^2\right) = O\left(n\pi_{n1}\right), \text{ and } E\left(\left[\mathbf{R}_n' \mathbf{U}_n\right]^2\right) = O\left(nE(R^2)\right). \tag{3.53}$$

Thus

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

by (3.26), (3.52), and (3.53).

Let us treat (3.50),

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

Finally, considering (3.51), we have

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

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

■

Lemma 3.4. *Under* ***Assumptions 2*** *and* ***5****, we have*

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

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

■

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

Lemma 3.5. *Under assumptions of Theorem 3.1, we have*

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

C.2. Conditional normalization condition:

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

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

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

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

Proof of Lemma 3.5

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

■

Proof of C.2

For each $i=1,\ldots,n$, let

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

We have

$$E\left(Z^2_{ni}\middle|\mathcal{F}_{n,i-1}\right)=\sigma_0^2E(V^2)D^2_{ii}+(\mu_4-\sigma_0^4)C^2_{ii}+4\sigma_0^2Q^2_{ni},$$

hence

$$\begin{aligned}E\left(\sum_{i=1}^nE\left(Z^2_{ni}\middle|\mathcal{F}_{n,i-1}\right)\right) &= \sigma_0^2E(V^2)\sum_{i=1}^nD^2_{ii}+(\mu_4-\sigma_0^4)\sum_{i=1}^nC^2_{ii}\\ &\quad +2\sigma_0^2E(T^2)\sum_{i=1}^n\sum_{j=1}^{i-1}G^{s2}_{ij}.\end{aligned}$$

By definition of $\tilde{Z}_{ni}$,

$$E\left(\sum_{i=1}^nE\left(\widetilde{Z}^2_{ni}\middle|\mathcal{F}_{n,i-1}\right)\right)=1+o(1).$$

Remark that

$$\mathrm{Var}\left(\sum_{i=1}^nE\left(Z^2_{ni}\middle|\mathcal{F}_{n,i-1}\right)\right)=16\sigma_0^4\mathrm{Var}\left(\sum_{i=1}^nQ^2_{ni}\right), \tag{3.54}$$

when U_i is normally distributed. Otherwise, result (3.57) remains valid.

Let us consider $\mathrm{Var}\left(\sum_{i=1}^nQ^2_{ni}\right)$. First, we have

$$\sum_{i=1}^nE\left(Q^2_{ni}\right)=E(T^2)\sum_{i=1}^n\sum_{j=1}^{i-1}G^{s2}_{ij}. \tag{3.55}$$

Let for all $1\leq i\leq j\leq n$,

$$\begin{aligned}E\left(Q^2_{ni}Q^2_{nj}\right) &= \sum_{k_1,k_2=1}^{i-1}\sum_{r_1,r_2=1}^{j-1}G^s_{ik_1}G^s_{ik_2}G^s_{jr_1}G^s_{jr_2}E\left(T_{k_1}T_{k_2}T_{r_1}T_{r_2}\right)\\ &= \sum_{k_1,k_2=1}^{i-1}\sum_{r_1,r_2=1}^{i-1}G^s_{ik_1}G^s_{ik_2}G^s_{jr_1}G^s_{jr_2}E\left(T_{k_1}T_{k_2}T_{r_1}T_{r_2}\right)\\ &\quad+\left[\sum_{k_1,k_2=1}^{i-1}G^s_{ik_1}G^s_{ik_2}E\left(T_{k_1}T_{k_2}\right)\right]\times\left[\sum_{r_1,r_2=i}^{j-1}G^s_{jr_1}G^s_{jr_2}E\left(T_{r_1}T_{r_2}\right)\right]\\ &= E\left(T^4\right)\sum_{k=1}^{i-1}G^{s2}_{ik}G^{s2}_{jk}+E(T^2)^2\sum_{k=1}^{i-1}\sum_{r=i}^{j-1}G^{s2}_{ik}G^{s2}_{jr}\\ &\quad+E(T^2)^2\sum_{k\neq r=1}^{i-1}\left[G^{s2}_{ik}G^{s2}_{jr}+2G^s_{ik}G^s_{ir}G^s_{jk}G^s_{jr}\right].\end{aligned}$$

Then, we have

$$\begin{aligned}
E\left(\left[\sum_{i=1}^{n} Q_{ni}^2\right]^2\right) &= E\left(T^4\right)\sum_{i=1}^{n}\sum_{k=1}^{i-1} G_{ik}^{s4} + 3E(T^2)^2\sum_{i=1}^{n}\sum_{k\neq r=1}^{i-1} G_{ik}^{s2}G_{ir}^{s2} \\
&\quad +2E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1}\left[G_{ik}^{s2}G_{jr}^{s2}+2G_{ik}^{s}G_{ir}^{s}G_{jk}^{s}G_{jr}^{s}\right] \\
&\quad +2E\left(T^4\right)\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1} G_{ik}^{s2}G_{jk}^{s2} + 2E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}\sum_{r=i}^{j-1} G_{ik}^{s2}G_{jr}^{s2}.
\end{aligned}$$

We can rewrite (3.55) as

$$\begin{aligned}
\left[2E(T^2)^2\right]^{-1}\left[E\left(\sum_{i=1}^{n} Q_{ni}^2\right)\right]^2 &= \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1} G_{ik}^{s2}G_{jk}^{s2} + \sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1} G_{ik}^{s2}G_{jr}^{s2} \\
&\quad +\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1}\sum_{r=i}^{j-1} G_{ik}^{s2}G_{jr}^{s2}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} Q_{ni}^2\right) &= E\left(T^4\right)\sum_{i=1}^{n}\sum_{k=1}^{i-1} G_{ik}^{s4} + 3E(T^2)^2\sum_{i=1}^{n}\sum_{k\neq r=1}^{i-1} G_{ik}^{s2}G_{ir}^{s2} \\
&\quad +\left[2E\left(T^4\right) - 2E(T^2)^2\right]\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k=1}^{i-1} G_{ik}^{s2}G_{jk}^{s2} \\
&\quad +4E(T^2)^2\sum_{j=1}^{n}\sum_{i=1}^{j-1}\sum_{k\neq r=1}^{i-1} G_{ik}^{s}G_{ir}^{s}G_{jk}^{s}G_{jr}^{s} \\
&= O\left[\frac{n}{h_n^2}\left(E\left(T^4\right)+h_nE(T^2)^2\right)\right]. \qquad (3.56)
\end{aligned}$$

Then, by (3.54) and (3.56), we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} E\left(\tilde{Z}_{ni}^2\middle|\mathcal{F}_{n,i-1}\right)\right) = O\left(\frac{E\left(T^4\right), +h_nE(T^2)^2}{n}\right) = o(1) \qquad (3.57)$$

since $E(T^4) = O(E(V^4)) = O(p_n^2)$ and $E(T^2) = O(E(V^2)) = O(1)$. Hence the result follows.

■

Proof of C.3

For any positive constants p and q such that $\frac{1}{p}+\frac{1}{q}=1$,

$$\begin{aligned}
|Z_{ni}| &\leq |D_{ii}||V_iU_i| + |C_{ii}||U_i^2-\sigma_0^2| + 2|U_i|\sum_{j=1}^{i-1}|G_{ij}^s||T_j| \\
&\leq |D_{ii}|^{\frac{1}{p}}|D_{ii}|^{\frac{1}{q}}|V_iU_i| + |C_{ii}|^{\frac{1}{p}}|C_{ii}|^{\frac{1}{q}}|U_i^2-\sigma_0^2| \\
&\quad +\sum_{j=1}^{i-1}|G_{ij}^s|^{\frac{1}{p}}|G_{ij}^s|^{\frac{1}{q}}2|T_j||U_i|.
\end{aligned}$$

Holder's inequality for inner products applied to the last term, implies that

$$\begin{aligned}
|Z_{ni}|^q &\leq \Bigg\{ \left[(|D_{ii}|^{\frac{1}{p}})^p + (|C_{ii}|^{\frac{1}{p}})^p + \sum_{j=1}^{i-1} (|G^s_{ij}|^{\frac{1}{p}})^p \right]^{\frac{1}{p}} \Big[(|D_{ii}|^{\frac{1}{q}} |V_i U_i|)^q \\
&\qquad + (|C_{ii}|^{\frac{1}{q}} |U_i^2 - \sigma_0^2|)^q + \sum_{j=1}^{i-1} (|G^s_{ij}|^{\frac{1}{q}} 2|T_j||U_i|)^q \Big]^{\frac{1}{q}} \Bigg\}^q \\
&= \left[|D_{ii}| + |C_{ii}| + \sum_{j=1}^{i-1} |G^s_{ij}| \right]^{\frac{q}{p}} \left[|D_{ii}||V_i U_i|^q + |C_{ii}||U_i^2 - \sigma_0^2|^q + 2^q |U_i|^q \sum_{j=1}^{i-1} |G^s_{ij}||T_j|^q \right] \\
&= O\left(|D_{ii}||V_i U_i|^q + |C_{ii}||U_i^2 - \sigma_0^2|^q + 2^q |U_i|^q \sum_{j=1}^{i-1} |G^s_{ij}||T_j|^q \right)
\end{aligned}$$

since under **Assumption 1**, D_{ii} and C_{ii} are of order $O(1/h_n)$ and G_n is uniformly bounded in row sums.

Let $q = 2 + \delta$, and note that

$$\sum_{i=1}^{n} E\left(\left| \widetilde{Z}_{ni} \right|^{2+\delta} \right) = O\left(\frac{h_n^{\frac{\delta}{2}}}{n^{\frac{\delta}{2}}} \left[E\left(U^{4+2\delta} \right) + h_n E\left(|T|^{2+\delta} \right) \right] \right). \tag{3.58}$$

Let $\delta = 2$, then (3.58) is of order $O\left(\frac{h_n^2 p_n^2}{n}\right)$, since $E(T^4) = O\left(p_n^2\right)$ and $E(U^8)$ is finite. This yields the proof as by assumption $h_n^4 = O(n)$ (when h_n is divergent) and $p_n^4 = o(n)$.

■

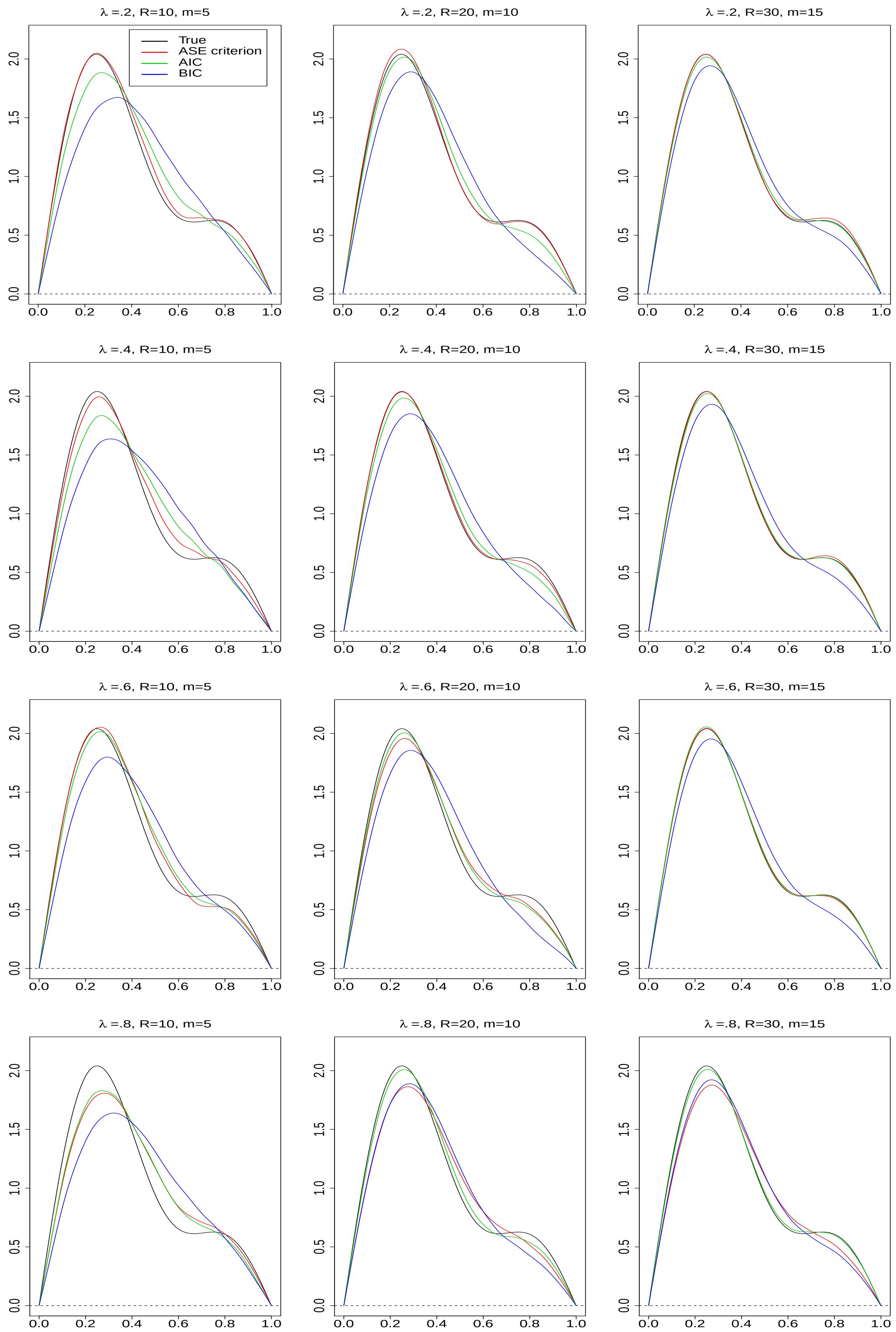

Figure 3.3: Estimated parameter function $\hat{\theta}_n(\cdot)$ with the different criteria in Scenario 2 for different values of r and m.

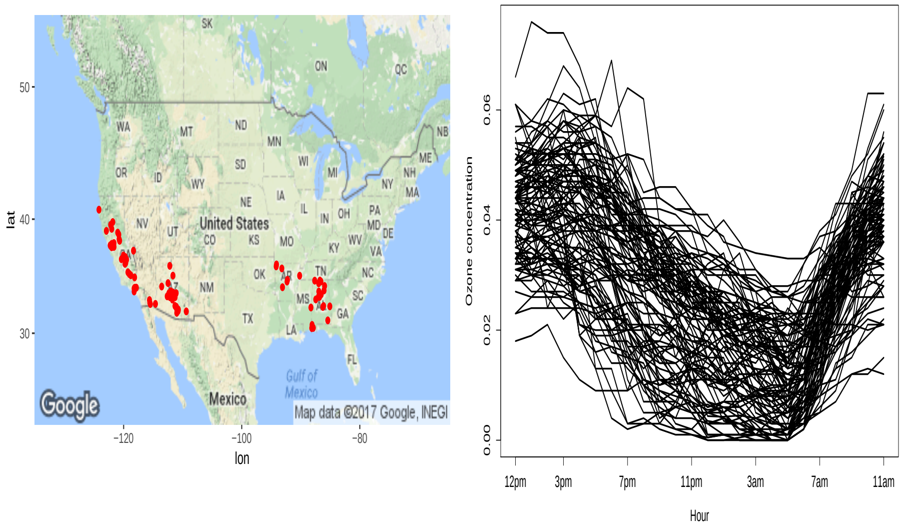

Figure 3.4: Locations and areas of the 106 stations (left panel) and corresponding ozone concentration curves from 12pm, July 19 to 11am, July 20 (right panel).

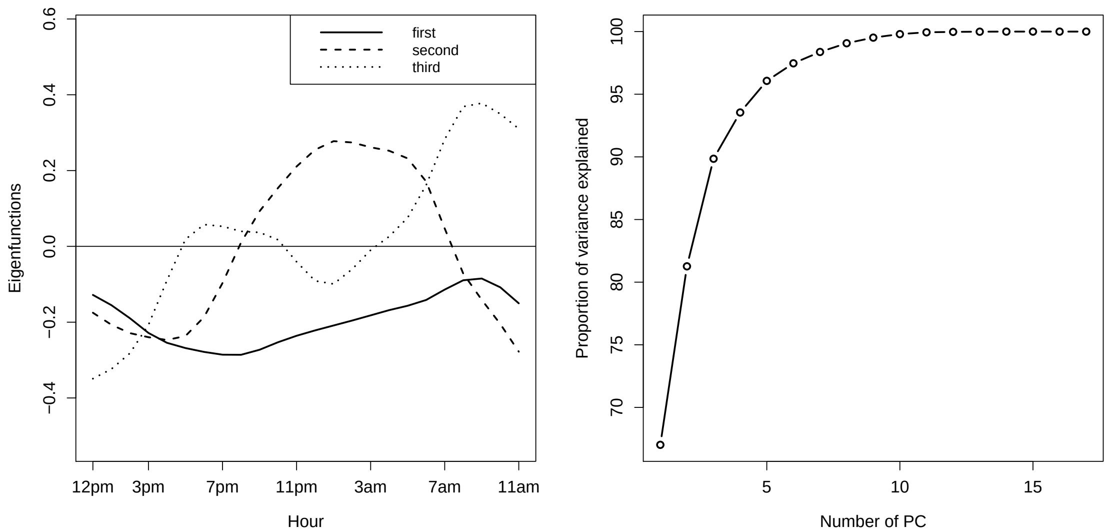

Figure 3.5: The three first eigenfunctions (left panel) and the proportion of explained variance (right panel).

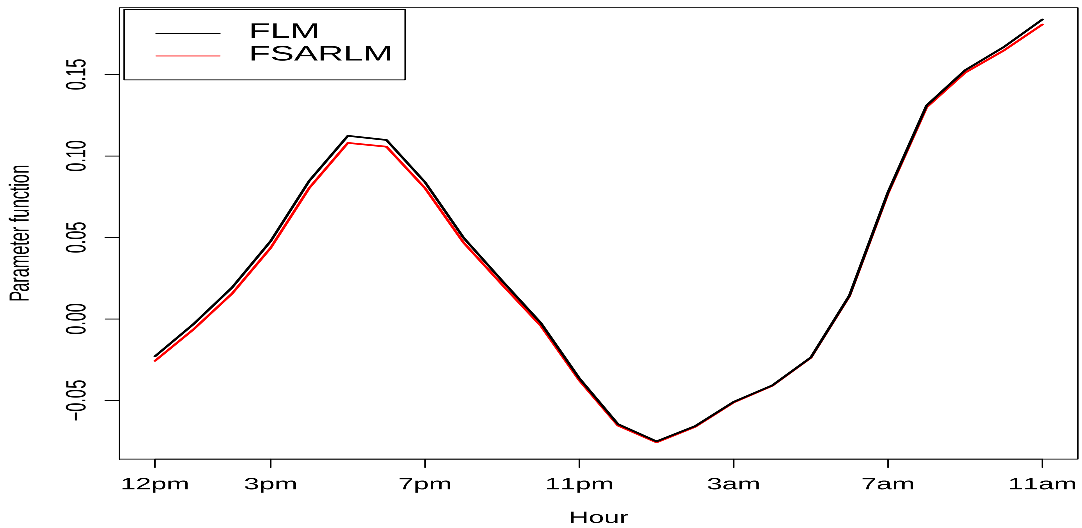

Figure 3.6: Estimated parameter functions.

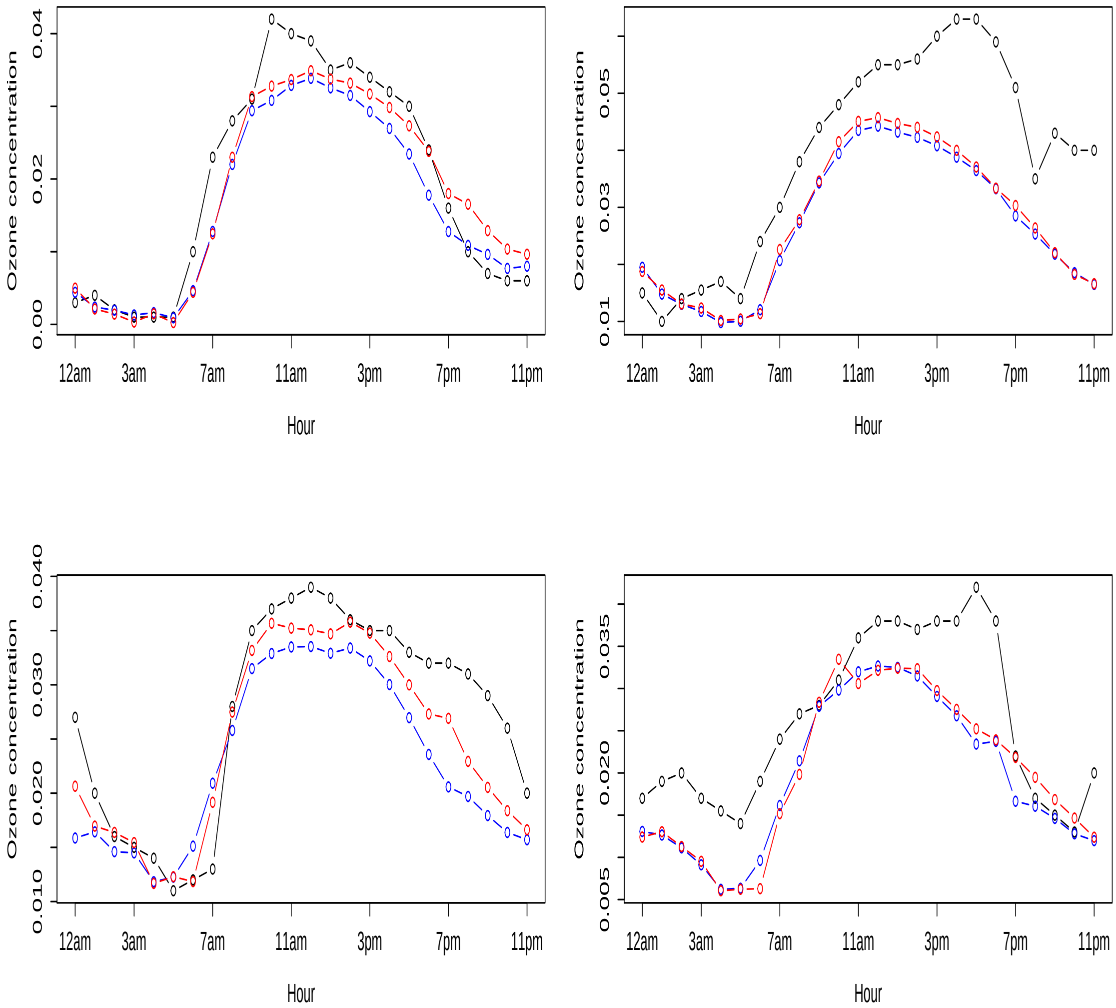

Figure 3.7: Ozone concentration (black curves) at four stations selected randomly from the 106 stations and their predictions obtained using the FSAR model (red curves) and FLM (blue curves).

Chapter 4

QMLE for Functional Spatial Autoregressive Models with Endogenous Weight Matrix

Contents

Resumé en français

Dans le Chapitre 3 nous avons proposé un estimateur de type quasi-vraisemblance pour un modèle fonctionnel autorégressif spatial sous l'hypothèse que la matrice de pondération nommée généralement W_n de dimension $n \times n$ est exogène contrairement à ce chapitre où nous relaxons cette hypothèse.

Soit $\{(\epsilon_{l(i),n}, \upsilon_{l(i),n}); l(i) \in D_n,\ n \in N\}$ un *triangular double array* de variables aléatoires définies dans un espace de probabilité $(\Omega; F; P)$. Ici $l(i)$ représente une unité spatiale appartenant à l'ensemble d'index fini $D_n \subset D$ où $D \in \mathbb{R}^d$, $d \geq 1$, avec D_n est de type lattice. Pour soucis de simplicité, nous utilisons $\epsilon_{i,n}$ et $\upsilon_{i,n}$ pour faire référence à $\epsilon_{l(i),n}$ et $\upsilon_{l(i),n}$ respectivement, et nous procéderons de même dans la suite. Supposons que dans chaque unité spatiale $l(i)$ nous observons sur D_n un processus linéaire Z en fonction d'une certaine co-variable U. Ces observations sont définies par

$$Z_n = U_n\Gamma + \epsilon_n, \tag{4.1}$$

où U_n est une matrice de dimension $n \times k$ d'éléments $\{u_{in}; l(i) \in D_n,\ i = 1, \dots, n,\ n \in N\}$ déterministes et bornés en valeur absolue pour tous i et n, Γ est un vecteur de coefficients de dimension $k \times p$ inconnu et $\epsilon_n = (\epsilon_{1,n}, \dots, \epsilon_{p,n})'$ est une matrice de dimension $n \times p$ de bruits blancs avec $\epsilon_{i,n} = (\epsilon_{1,in}, \dots \epsilon_{n,in})'$ un vecteur de taille p. Ainsi, $Z_n = (z_{1,n}, ..., z_{n,n})'$ est une matrice de dimension $n \times p$ avec $z_{i,n} = (z_{1,in}, \dots, z_{p,in})'$.

Pour n observations spatiales d'une *variable de réponse* Y réelle et d'une co-variable fonctionnelle $\{X(t),\, t \in T\}$, nous nous intéressons au modèle SAR fonctionnel ci-dessous :

$$Y_i = \lambda \sum_{j=1}^{n} w_{ij} Y_j + \int_T X_i(t)\beta(t)dt + \upsilon_i, \qquad i = 1, \dots, n \tag{4.2}$$

où λ est un scalaire dans un espace compact Λ, et $\{\upsilon_i,\ i = 1, \dots, n,\ n = 1, 2, \dots\}$ sont les termes d'erreur. Nous supposons que $\{X(t), t \in T\}$ est un processus stochastique qui prend ses valeurs dans $\mathcal{X} \subset L^2(T)$, avec $L^2(T)$ l'espace des fonctions carrées intégrables dans l'intervalle $T \subset \mathbb{R}$ et $\beta(.)$ est une fonction de paramètre qui appartient au même espace. La matrice de pondération spatiale $W_n = (w_{ij,n})$ est de dimension $n \times n$ et supposée non négative avec zéro au diagonale et avec $w_{ij,n} = h_{ij}(Z_n, d_{ij})$ pour $i, j = 1, ..., n; i \neq j$ où $h(.)$ est une fonction bornée. Soit $S_n(\lambda) = I_n - \lambda W_n$ ainsi le modèle (4.2) peut se re-écrire comme :

$$Y_n = S_n(\lambda) \left(\mathbf{X}_n \left(\beta(.) \right) + V_n \right),\ n = 1, 2, ... \tag{4.3}$$

où $\mathbf{X}_n(\beta(.))$ est le vecteur $(n \times 1)$ dont le $i$ème élément est $\int_T X_i(t)\beta(t)dt$, $\mathbf{Y}$ est le vecteur de dimension $n \times 1$ des éléments Y_i et $V_n = (\upsilon_{1,n}, \dots, \upsilon_{i,n})'$.

Nous supposons maintenant que les termes $\upsilon_{i,n}$ et $\epsilon_{i,n}$ ont une distribution jointe $(\upsilon_{i,n}, \epsilon'_{i,n}) \sim$ i.i.d $(0, \Sigma_{\upsilon\epsilon})$ où $\Sigma_{\upsilon\epsilon} = \begin{pmatrix} \sigma_\upsilon^2 & \sigma'_{\upsilon\epsilon} \\ \sigma_{\upsilon\epsilon} & \Sigma_\epsilon \end{pmatrix}$ est une matrice de covariance (définie positive), σ_υ^2 est la variance de V_n, le vecteur de covariance $\sigma_{\upsilon\epsilon} = (\sigma_{\upsilon\epsilon_1}, \dots, \sigma_{\upsilon\epsilon_p})'$ est de dimension p. Supposons que $\mathrm{Sup}_{i,n} E|\upsilon_{i,n}|^{4+\zeta}$ et $\sup_{i,n} E||\epsilon_{i,n}||^{4+\zeta}$ existent pour un certain $\zeta \neq 0$. En outre, nous posons $E(\upsilon_{i,n}|\epsilon_{i,n}) = \epsilon'_{i,n}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$ et $Var(\upsilon_{i,n}|\epsilon_{i,n}) = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$ où Σ_ϵ est une matrice de dimension $p \times p$.

La corrélation entre $\upsilon_{i,n}$ et $\epsilon_{i,n}$ est la source directe de l'endogénéité de la matrice de pondération spatiale W_n où réside l'originalité de ce travail. Dans le cas particulier où $\sigma_{\upsilon\epsilon} = 0$, W_n devient exogène comme dans le chapitre précédent. En se basant alors sur les moments conditionnels présentés nous pouvons définir $\xi_n = V_n - \epsilon_n\delta$ avec $\delta = \Sigma_\epsilon^{-1}\sigma'_{\upsilon\epsilon}$, le modèle (4.3) se transforme en :

$$Y_n = S_n(\lambda) \left(\mathbf{X}_n \left(\beta(.) \right) + (Z_n - U_n\Gamma)\delta + \xi_n \right),\ n = 1, 2, ... \tag{4.4}$$

où ξ_n sont des i.i.d avec $E(\xi_{i,n}|\epsilon_{i,n}) = 0$ et $E(\xi_{i,n}^2|\epsilon_{i,n}) = \sigma_\xi^2$ avec $\sigma_\xi^2 = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$. En particulier ξ_n n'est pas corrélée avec ϵ_n, on peut donc considérer la variable $(Z_n - U_n\Gamma)$ comme variable de contrôle de l'endogénéité.

Selon White (1982), nous définissons pour le modèle (4.4) le logarithme de la fonction de quasi-vraisemblance sous la spécification de la distribution normale comme suit :

$$\begin{aligned} \ln L_n(\theta) \;=\; & -n\ln(2\pi) - \tfrac{n}{2}\ln\sigma_\xi + \ln|S_n| - \tfrac{n}{2}\ln|\Sigma_\epsilon| \\ & -\tfrac{1}{2}\textstyle\sum_{i=1}^{n}(z'_{i,n} - u'_{in}\Gamma)\Sigma_\epsilon^{-1}(z_{i,n} - \Gamma' u_{in}) \\ & -\tfrac{1}{2\sigma_\xi}[S_n(\lambda)Y_n - \mathbf{X}_n(\beta(.)) - (Z_n - U_n\Gamma)\delta]' \\ & \times[S_n(\lambda)Y_n - \mathbf{X}_n(\beta(.)) - (Z_n - U_n\Gamma)\delta], \end{aligned} \tag{4.5}$$

où $\theta = (\lambda, \beta(.), \text{vect}(\Gamma), \sigma_\xi^2, \tau', \delta')^T$ est l'ensemble des paramètres à estimer en maximisant (4.5), τ est le vecteur de dimension notée J des éléments distincts de Σ_ϵ, avec $\delta = \Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$ et $\sigma_\xi^2 = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$. Les estimateurs QMLE de ces paramètres sont alors les valeurs qui maximisent la précédente vraisemblance. Dans le cadre non-fonctionnel, Qu & Lee (2015) ont défini des estimateurs de type quasi-maximum de vraisemblance pour θ en maximisant l'équivalent de (4.5).

Pour le modèle fonctionnel (4.4) considéré, nous proposons une méthode d'estimation qui étend le travail de Qu & Lee (2015) à l'aide d'une réduction de la dimension infinie de l'espace de la variable explicative fonctionnelle $X(.)$ en utilisant la technique de troncature du Chapitre 3. Soit $\Phi = (\phi_j(t))', j > 1$ une base orthonormale de $L^2(T)$, habituellement la base de Fourier est utilisée. La co-variable $X(t)$ et la fonction de paramètre $\beta(t)$ peuvent être ré-écrites à l'aide de la base :

$$X(t) = \sum_{j=1}^{\infty} \omega_j\phi_j(t) \quad \text{and} \quad \beta(t) = \sum_{j=1}^{\infty} \beta_j\phi_j(t) \quad \text{pour tout } t \in T,$$

où les variables aléatoires ω_j et les coefficients β_j sont données par $\omega_j = \int X(t)\phi_j(t)dt$ et $\beta_j = \int \beta(t)\phi_j(t)dt$, respectivement. Nous avons alors la décomposition suivante :

$$\int_T X(t)\beta(t)dt = \sum_{j=1}^{\infty} \beta_j\omega_j = \sum_{j=1}^{p_n} \beta_j\omega_j + \sum_{j=p_n+1}^{\infty} \beta_j\omega_j. \tag{4.6}$$

De manière similaire au Chapitre 3, le modèle (4.4) est approché en remplaçant $\mathbf{X}(\beta(.))$ dans (4.5) par $\psi_{p_n}\beta^\star$, avec $\beta^\star = (\beta_1, ..., \beta_{p_n})$ et ψ_{p_n} la matrice de dimension $n \times p_n$ d'éléments $\left\{\omega_j^{(i)} = \int_T X_i\phi_j(t)dt,\ i = 1, ..., n\ j = 1, ..., p_n\right\}$, afin d'obtenir la fonction de quasi vraisemblance calculable suivante :

$$\begin{aligned} \ln\tilde{L}_n(\theta) \quad = \quad & -n\ln(2\pi) - \tfrac{n}{2}\ln\sigma_\xi + \ln|S_n| - \tfrac{n}{2}\ln|\Sigma_\epsilon| \\ & -\tfrac{1}{2}\textstyle\sum_{i=1}^n (z'_{i,n} - u'_{in}\Gamma)\Sigma_\epsilon^{-1}(z_{i,n} - \Gamma' u_{in}) \\ & -\tfrac{1}{2\sigma_\xi}[S_n(\lambda)Y_n - \psi_{p_n}\beta^\star - (Z_n - U_n\Gamma)\delta]' \\ & \times[S_n(\lambda)Y_n - \psi_{p_n}\beta^\star - (Z_n - U_n\Gamma)\delta]. \end{aligned} \tag{4.7}$$

Les éléments du vecteur $\hat{\theta} = (\hat{\lambda}, \hat{\beta}^{\star\prime}, \text{vec}(\hat{\Gamma})', \hat{\sigma}_\xi^2, \hat{\tau}', \hat{\delta}')^T$ qui maximisent (4.7) donnent les estimateurs des paramètres λ, $\beta^\star$, $\text{vec}(\Gamma)$, σ_ξ^2, τ et δ respectivement avec $\delta = \Sigma_\epsilon^{-1}\sigma'_{\upsilon\epsilon}$ et $\sigma_\xi^2 = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$. L'estimateur du paramètre fonctionnel $\beta(t)$ est défini par :

$$\hat{\beta}(t) = \sum_{j=1}^{p_n} \hat{\beta}^\star\phi_j(t) \tag{4.8}$$

Le comportement pour échantillons à taille finie des estimateurs a été étudié à travers une étude de type Monte-Carlo.

4.1 Introduction

The last decades have seen an extensive development of statistical tools able to manage large quantities of data containing inherent spatial components. Spatial Functional Data Analysis (FDA) is a field of statistics for modeling such data available in diverse disciplines as environmental sciences, economics, agronomy, mining, forestry, among others.

In this contribution, we are interested to spatial FDA in some econometric problems. As said in the previous chapter, the term *functional data analysis* was popularized by Ramsay (1982) and Ramsay & Dalzell (1991), even the concept is older and dates back to Grenander (1950) and Rao (1958). FDA deals with the analysis of tightly spaced repeated measurements on same individuals or discrete observations of a phenomena that can be represented in a form of function such as curves, shapes, images. FDA spatial objects are thought as smooth realizations of a stochastic spatial process (see Brumback & Rice (1998)) and are widely present in many research areas e.g. chemiometrics (see Frank & Friedman (1993)), meteorology, speech analysis (see Hastie et al. (1995)), environment, biology among others (see e.g. Hastie & Mallows (1993)). An important FDA literature has been developed for representation, exploration and modeling parametrically or non-parametrically functional data, see for instance the monographs of Ramsay & Silverman (2005), Ferraty & Vieu (2006) and Horváth & Kokoszka (2012), among others.

Spatial statistics (Cressie, 1993) embodies a suite of methods for analyzing spatial data and for instance estimating the values of a property of interest at non-sampled locations, from available sample data points using spatial correlation tools. Spatial data analysis encompass various techniques for modeling correlation between observed variables located in space. There are three main types of spatial data, namely geostatistics data, lattice data and point patterns (see Cressie (1993)). Our main focus is the lattice context of *spatial autoregressive model* (SAR), useful in spatial Econometrics. Earlier developments in estimating and testing SAR models are established using the two stage least squares (2SL), the three stage least squares (3SLS), the maximum likelihood estimation (MLE) or quasi-maximum likelihood (QMLE) and the generalized method of moments (GMM) methods have been summarized in: Anselin (1988), Cressie (1993), Kelejian & Prucha (1998, 1999), Conley (1999), Lee (2004, 2007), Lin & Lee (2010), Zheng & Zhu (2012), Malikov & Sun (2017), Garthoff & Otto (2017), Yang & Lee (2017) and among others.

SAR model permits interdependence between spatial units via a well known spatial weight matrix (W_n) that can be defined in different ways. In fact, W_n is usually supposed to be exogenous and based only on geography criteria or spatial arrangement of the observations. For instance spatial units are considered neighbors when they share a border or when they are within a given distance. However, in many empirical applications the assumption of exogeneity may not be reasonable then elements of W_n may involve other distances, as "economic distance" or "socio-economic distance" like Cohen & Morrison Paul (2004) who used a weight matrix based on a technological proximity index or Conley & Ligon (2002) where an economic definition based on transport costs was used.

Estimation methods mentioned earlier fails when W_n is endogenous because of technical complications. This is one of the several problems emphasized by Pinkse & Slade (2010) that still under investigation. So far, very few works relaxing the assumption of an exogenous matrix exist; Kelejian & Piras (2014) proposed instrumenting the endogenous spatial weight matrix in a context of spatial panel data. Qu & Lee (2015) considered the case of endogenous spatial weight matrix constructed by some univariate economic variable. Later on, Lee & Yu (2017) extend this idea to panel data. For the best of our knowledge, in the context of spatial endogenous weight matrix, the data studied so far are real-valued.

Extending this context to FDA is the aim of this contribution. In fact, nowadays in many applied domains as; economic, environmental, hydrology,.., spatially correlated functional data are more and more available and there is a dynamic on developing statistical tools to analysis such data. Generally, FDA is more developed in the scope of geostatistics. Some limited works exist in the functional lattice context (see e.g Ruiz-Medina (2011), Ruiz-

Medina (2012), Pineda-Ríos & Giraldo (2016) and the references in the previous chapter) with exogenous weight matrices. The present work considers an estimation of the spatial autoregressive model with a random functional covariate and a real-valued response using spatial weight matrix assumed to be exogenous compare to Chapter 3.

The remainder of this chapter is organized as follow. In Section 4.2, we introduce the model specification of the outcome equation, the structure of the spatial weight matrix and give some motivations on the particular structure of the spatial weight matrix. Section 4.3 is devoted to the QMLE estimation method. In Section 4.4 we give the numerical experiments; a Monte Carlo simulation are provided to investigate the finite behavior of the estimator compared to the estimator under the exogenous assumption. The conclusion in Section 4.5 ends the chapter.

4.2 The model

In this section, we give the specification of a SAR model, the structure of the proposed spatial weight matrix and show how this induces endogeneity. We also motivate the use of endogeneity assumption.

4.2.1 Model specification

Similar to Jenish & Prucha (2009, 2012), we consider a spatial process located on (possibly) unevenly spaced lattice $D \subset \mathbb{R}^d$ with $d \geq 1$. There are two main asymptotic methods commonly used in the literature; increasing domain and infill asymptotic, see Cressie (1993). The Infill asymptotic is analogical to the term "infill drilling", where extra core samples are drilled between existing ones, so the sample region remains fixed and the sample data growth by sampling points arbitrarily dense in the given region. Under increasing domain asymptotic, more observations may be taken, so the sample region is expanded, specifically this arises when lattice data have a fixed spacing between neighbors (e.g., location of trees in a domain), it is more appropriate than infill asymptotic. Here we employ the increasing domain asymptotic method as Qu & Lee (2015) which is ensured by the next assumption.

Assumption 1. The lattice $D \in \mathbb{R}^d$, $d \geq 1$, is infinitely countable. The location $l : \{1, ..., n\} \longrightarrow D_n \subset D$ is a mapping of individual i to its location $l(i) \in D_n \subset \mathbb{R}^d$. All elements in D are located at distances of at least $d_0 > 0$ from each other, i.e., for $\forall l(i), l(j) \in D : d_{ij} \geq d_0$ where d_{ij} is the distance between individual i and individual j; w.l.o.g. we assume that $d_0 = 1$.

Consider $\{(\epsilon_{l(i),n}, \upsilon_{l(i),n}); l(i) \in D_n, n \in \mathbb{N}\}$ a triangular double array of real random variables defined on a probability space $(\Omega; F; P)$, where the index set $D_n \subset D$ is a finite set and satisfies *Assumption 1.* For simplicity, let $\epsilon_{i,n}$ and $\upsilon_{i,n}$ to refer to $\epsilon_{l(i),n}$ and $\upsilon_{l(i),n}$ respectively and we do the same for the rest of variables. We assume that for each location $l(i)$, we observe two variables Z and U in D_n related with a linear process defined by:

$$\mathbf{Z}_n = \mathbf{U}_n \Gamma + \epsilon_n \tag{4.9}$$

where $\mathbf{U}_n = (u_1, ..., u_n)'$ is an $n \times q$ matrix with its element $\{u_i;\ l(i) \in D_n,\ i = 1, \ldots, n,\ n \in N\}$ being deterministic and bounded in absolute value for all i and n, Γ is a $q \times d$ vector of coefficients to be estimated and ϵ_n is an i.i.d $n \times d$ matrix of disturbances with variance Σ_ϵ with the element $\epsilon_i = (\epsilon_{1,i}, \ldots, \epsilon_{d,i})'$ is a d-column vector.

$\mathbf{Z}_n = (Z_1, ..., Z_n)'$ is an $n \times d$ matrix, where $Z_i = (Z_{1,i}, \ldots, Z_{d,i})'$ may be a vector of economic variables such as GDP, consumption or economic growth rate, influencing interactions across spatial units.

Let $\mathbf{Y}_n$ is an $n \times 1$ *response vector of variable* and $\mathbf{X}_n(t) = \{X_1(t), ..., X_n(t) : t \in T\}$ be a random sample of observations or a sample paths of a functional variable $X(t)$. The following model gives the functional SAR model with endogenous reaction between the response and the covariate:

$$Y_i = \lambda \sum_{j=1}^{n} w_{ij} Y_j + \int_T X_i(t)\beta(t)dt + v_i, \qquad i = 1, \ldots, n \tag{4.10}$$

where λ is a scalar in a compact space Λ, $\beta(.)$ is a functional parameter that belongs to the space functions $L^2(T)$, where $L^2(T)$ is the space of square-integrable functions in T. Assume that the process $\{X(t), t \in T\}$ takes values in space $\mathcal{X} \subset L^2(T)$ and the errors $\{v_i,\ i = 1, \ldots, n,\ n \in \mathbb{N}\}$ are i.i.d. The particularity of the model is that $W_n = (w_{ij,n})$ is $n \times n$ non-negative matrix with zero diagonals where $w_{ij,n} = h_{ij}(\mathbf{Z}_n, d_{ij})$ for $i, j = 1, ..., n; i \neq j$: $h(.)$ is a bounded function. Let $S_n(\lambda) = I_n - \lambda W_n$ we can write (4.10) as:

$$\mathbf{Y}_n = S_n(\lambda)\left(\mathbf{X}_n(\beta(.)) + \mathbf{V}_n\right) \tag{4.11}$$

where $\mathbf{X}_n(\beta(.))$, $\mathbf{Y}_n$ are the $n \times 1$ vectors of i-th element $\int_T X_i(t)\beta(t)dt$ and Y_i, respectively and $\mathbf{V}_n = (v_1, \ldots, v_n)'$ is a $n \times 1$ vector of i.i.d variables.

4.2.2 Motivation

On the conventional SAR model of Cliff & Ord (1973) the spatial linkage presented by the weighted matrix is essentially physical based on geographic distance between spatial units, so variable such as incomes, behavior, expenditure, ..., of an individual in location $l(i)$ that may influence the values of the same variables at some neighbor location $l(j)$ are not taken into account.

To account other type of spatial interactions, the concept of spatial weight was extended to go beyond the geographical notion, then one can consider that locations are neighbors if they are similar economically or demographically. For instance, Case et al. (1993) said that spillovers effect may originate in locations that are not neighbors, for example expenditures in education in one location (state) are likely to have the most impact on labor markets of states with similar economic or demographic characteristics. So an alternative criteria is defined in the literature and elements of a spatial weight matrix based on an known function $g(.)$ may take the form $w_{ij,n} = g(Z_{i,n}, Z_{j,n})$, where $Z_{i,n}$ and $Z_{j,n}$ are observations on "meaningful" socio-economic characteristics, such as per capita income, or proportion in a given racial group as in Case et al. (1993). In a migration context, Smith & C. (2004) used migration flow while Crabbé & Vandenbussche (2008) in addition to physical distance, they constructed a spatial weight matrix based on the inverse of the trade share between states. Note that in this case *Assumption 1* holds and one can replace easily "physical distance" by economic one.

It is then not obvious to be limited to geographic proximity, an alternative may relies on socio-economic characteristics or combination of socio-economic and geographic factor. The spatial weight matrix W can be expressed as a simple weighting:

$$W = \alpha W^d + (1 - \alpha) W^e$$

where $0 \leq \alpha \leq 1$. W^d and W^e are respectively the weight matrix based on geographic and socio-economic metrics. To generate W^d, we may use one of the several existing methods. Some of them are recalled in the following:

- *k-Nearest Neighbor weights*

$$w_{ij}^d = \begin{cases} 1 & \text{if } j \in N_k(i) \\ 0 & \text{Otherwise} \end{cases},$$

  where $N_k(i)$ is the set of the k closest units or regions to i for $k = 1, ..., (n-1)$.

- *Inverse Distance weights*

  $w_{ij}^d = 1/d_{ij}$, d_{ij} is the euclidean distance between location i and j.

- *Contiguity weights*

$$w_{ij}^d = \begin{cases} 1 & \text{if } i \text{ and } j \text{ share borders} \\ 0 & \text{Otherwise} \end{cases}$$

For W^e we may consider one of the next cases:

- $w_{ij}^e = 1/|Z_{i,n} - Z_{j,n}|$,
- $w_{ij}^e = 1/(Z_{i,n} - Z_{j,n})^2$,
- $w_{ij}^e = 1/[1 + \log(Z_{i,n}/Z_{j,n})]$,

where the $Z_{i,n}$ are observations of a socio-economic variable.

Such structure of W impose endogeneity on the model and a specific estimation method must be carried out. In the next we show how this endogeneity is considered in the model.

4.2.3 Source of endogeneity

Assumption 2. The error terms $\upsilon_{i,n}$ and $\epsilon_{i,n}$ have a joint distribution: $(\upsilon_{i,n}, \epsilon'_{i,n}) \sim$ i.i.d $(0, \Sigma_{\upsilon\epsilon})$ where $\Sigma_{\upsilon\epsilon} = \begin{pmatrix} \sigma_\upsilon^2 & \sigma'_{\upsilon\epsilon} \\ \sigma_{\upsilon\epsilon} & \Sigma_\epsilon \end{pmatrix}$ is a covariance matrix (positive and definite), σ_υ^2 is a scalar variance of $\mathbf{V}_n$, $\sigma_{\upsilon\epsilon} = (\sigma_{\upsilon\epsilon_1}, \ldots, \sigma_{\upsilon\epsilon_p})$ is a covariance with p-dimensional vector and Σ_ϵ is a $d \times d$ matrix. The $\sup_{i,n} E|\upsilon_{i,n}|^{4+\zeta}$ and $\sup_{i,n} E||\epsilon_{i,n}||^{4+\zeta}$ exist for some $\zeta \neq 0$. In addition $E(\upsilon_{i,n}|\epsilon_{i,n}) = \epsilon'_i \Sigma_\epsilon^{-1} \sigma_{\upsilon\epsilon}$ and $\mathrm{Var}(\upsilon_{i,n}|\epsilon_{i,n}) = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon} \Sigma_\epsilon^{-1} \sigma_{\upsilon\epsilon}$.

The correlation between $\upsilon_{i,n}$ and $\epsilon_{i,n}$ is the direct source of the endogeneity of the spatial weight matrix W_n. In case where $\sigma_{\upsilon\epsilon} = 0$, W_n become strictly exogenous and conventional estimator of functional spatial autoregressive model can be applied[1], see previous chapter.

Based on the conditional moments presented in *Assumption 2*, let $\delta = \Sigma_\epsilon^{-1} \sigma_{\upsilon\epsilon}$, the scalar $\sigma_\xi^2 = \sigma_\upsilon^2 - \sigma'_{\upsilon\epsilon} \Sigma_\epsilon^{-1} \sigma_{\upsilon\epsilon}$ and denote by $\xi_n = \mathbf{V}_n - \epsilon_n \delta$. The outcome equation (4.11) can be transformed into:

$$\mathbf{Y}_n = S_n(\lambda)\left(\mathbf{X}_n(\beta(.)) + (\mathbf{Z}_n - \mathbf{U}_n \Gamma)\delta + \xi_n\right) \; i = 1, ..., n, \; n = 1, 2, ..., \tag{4.12}$$

where $\mathbf{X}_n(\beta(.))$ be the $n \times 1$ vector of i-th element $\int_{\mathcal{T}} X_i(t)\beta(t)dt$, $S_n(\lambda) = I_n - \lambda W_n$ and ξ_n are i.i.d with $E(\xi_{i,n}|\epsilon_{i,n}) = 0$ and $E(\xi_{i,n}^2|\epsilon_{i,n}) = \sigma_\xi^2$. In particular ξ_n are uncorrelated

[1] Quasi-maximum likelihood of Lee (2004)

with ϵ_n in the case where $(\upsilon_{i,n}, \epsilon_{i,n})$ has a joint normal distribution. We can consider the variable $(\mathbf{Z}_n - \mathbf{U}_n\Gamma)$ as a control variable of the endogeneity of W_n, so our estimation method will mainly rely on equation (4.12).

Note that *Assumption 2* is a general case with no restriction on the nature of the disturbances, only conditional moments is matter.

4.3 Estimation methodology

Assuming the i.i.d disturbances $(\upsilon_{i,n}, \epsilon'_{i,n}) \sim (0, \Sigma_{\upsilon\epsilon})$ with $\Sigma_{\upsilon\epsilon} = \begin{pmatrix} \sigma^2_\upsilon & \sigma'_{\upsilon\epsilon} \\ \sigma_{\upsilon\epsilon} & \Sigma_\epsilon \end{pmatrix}$ and according to White (1982), we can define the log quasi-likelihood function under the normal distribution specification as follow:

$$\begin{aligned} \ln L_n(\theta) &= -n\ln(2\pi) - \tfrac{n}{2}\ln\sigma^2_\xi + \ln|S_n(\lambda)| - \tfrac{n}{2}\ln|\Sigma_\epsilon| \\ &\quad -\tfrac{1}{2}\textstyle\sum_{i=1}^n (Z'_i - U'_i\Gamma)\Sigma_\epsilon^{-1}(Z_i - \Gamma' U_i) \\ &\quad -\tfrac{1}{2\sigma_\xi}[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\beta(.)) - (\mathbf{Z}_n - \mathbf{U}_n\Gamma)\delta]' \\ &\quad \times[S_n(\lambda)\mathbf{Y}_n - \mathbf{X}_n(\beta(.)) - (\mathbf{Z}_n - \mathbf{U}_n\Gamma)\delta], \end{aligned} \tag{4.13}$$

where $\theta = (\lambda, \beta(.), \text{vect}(\Gamma)', \sigma^2_\xi, \tau', \delta)'$ is the set of parameters to be estimated, by maximizing (4.13), with τ being a J-dimensional column vector of distinct elements in Σ_ϵ, $\delta = \Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$ and $\sigma^2_\xi = \sigma^2_\upsilon - \sigma'_{\upsilon\epsilon}\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$. So the QMLE($\theta$) is

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \ln L_n(\theta)$$

However, doing the maximization is not simple in practice, in fact we cannot handle the functional form of the sample paths of the functional covariate $\mathbf{X}_n$, as much we can observe each of them in a finite set of discrete time points. In addition the parameter $\beta(.)$ is a function. In order to solve the first task, there are different proposals in the literature that suggest to reconstruct the functional form of the sample paths by different smoothing methods. Some works, both smooth the functional covariate and handle the infinite dimension of the parameter, we refer among others to Cardot & Sarda (2005) who worked on identifiability and estimation of functional regression models via B-splines penalized likelihood. The idea is to express the functional explanatory vector of variables $\mathbf{X}_n$ and the parameter function $\beta(.)$ in a finite dimensional space belonging to the same space of sample paths. Many truncation methods using eigenbasis, Splines or Fourier basis can be considered, see e.g. Cardot et al. (1999), Cardot & Sarda (2005).

In this contribution, we use the method of Muller & Stadtmuller (2005) where one can project $\mathbf{X}_n$ and $\beta(t)$ in a finite dimensional space spanned by a basis of functions verifying some criteria. This approach is used in Chapter 3 and to refresh the reader's memory, we give some details on this proposed technique.

Truncation strategies

The truncation is a approximation technique to reduce the infinite dimension of the functional variables $\mathbf{X}_n$ and parameter function $\beta(.)$ by projecting them in a finite dimensional space generated by a basis of functions. The usual basis are trigonometric basis (Aguilera et al. (1995)), cubic spline basis (Aguilera et al. (1996)), wavelets basis (Amato et al. (2006)) and eigenfunctions basis used in the following.

Denote by $\Phi = (\phi_1(t), ...\phi_j(t))', j > 1$ a vector of basic functions of the space $L^2(T)$, in particular the eigenfunction basis of the covariance operator Ψ defined by:

$$\Psi_X(t) = \int_T E[X(t)X(s)]x(s)\mathrm{d}s, \quad x \in \mathcal{X}, \, t \in T. \tag{4.14}$$

Then the functional variable $X(.)$ and the parameter function $\beta(.)$ can be expanded into:

$$X(t) = \sum_{j=1}^{\infty} \omega_j \phi_j(t) \quad \text{and} \quad \beta(t) = \sum_{j=1}^{\infty} \beta_j \phi_j(t) \quad \text{for all } t \in T,$$

with w.r.'s ω_j and coefficients β_j are given by $\omega_j = \int X(t)\phi_j(t)dt$ and $\beta_j = \int \beta(t)\phi_j(t)dt$, respectively. Therefore from the orthogonality characteristic of the basis, it results immediately that:

$$\int_T X(t)\beta(t)dt = \sum_{j=1}^{\infty} \beta_j \omega_j \tag{4.15}$$

Let us consider the decomposition:

$$\sum_{j=1}^{\infty} \beta_j \omega_j = \sum_{j=1}^{p_n} \beta_j \omega_j + \sum_{j=p_n+1}^{\infty} \beta_j \omega_j. \tag{4.16}$$

The idea is to approximate model (4.11) with another one where the functional predictor is truncated at p_n variables. Similar as Muller & Stadtmuller (2005) the left hand side of (4.16) can be approximated only with the first term of the right hand side. A necessary condition is that the error vanishes asymptotically, this is the case when considering the eigenbasis of the variance-covariance operator Ψ:

$$E\left(\sum_{j=p_n+1}^{\infty} \beta_j \omega_j\right)^2 = \sum_{j=p_n+1}^{\infty} \beta_j^2 E(\omega_j^2) = \sum_{j=p_n+1}^{\infty} \beta_j^2 \kappa_j, \tag{4.17}$$

where $\kappa_j, j = 1, 2...$ are the eigenvalues. Then we can express $\mathbf{X}_n(\beta(.))$ by $\psi_{p_n}\beta^\star$, where $\beta^\star = (\beta_1, ..., \beta_{p_n})$ and ψ_{p_n} is a $n \times p_n$ matrix with (i, j) element given by

$$\omega_j^{(i)} = \int_T X_i \phi_j(t)dt, \quad i = 1, ..., n \; j = 1, ..., p_n \tag{4.18}$$

Replacing $\mathbf{X}_n(\beta(.))$ in (4.13) by $\psi_{p_n}\beta^\star$ permit to have what we call a feasible log likelihood, defined as:

$$\begin{aligned} \ln\tilde{L}_n(\theta) \;\; = \;\; & -n\ln(2\pi) - \tfrac{n}{2}\ln\sigma_\xi^2 + \ln|S_n(\lambda)| - \tfrac{n}{2}\ln|\Sigma_\epsilon| \\ & -\tfrac{1}{2}\textstyle\sum_{i=1}^n (Z_i' - U_i'\Gamma)\Sigma_\epsilon^{-1}(Z_i - \Gamma' U_i) \\ & -\tfrac{1}{2\sigma_\xi}[S_n(\lambda)\mathbf{Y}_n - \psi_{p_n}\beta^\star - (\mathbf{Z}_n - \mathbf{U}_n\Gamma)\delta]' \\ & \times[S_n(\lambda)\mathbf{Y}_n - \psi_{p_n}\beta^\star - (\mathbf{Z}_n - \mathbf{U}_n\Gamma)\delta], \end{aligned} \tag{4.19}$$

$\hat{\theta} = (\hat{\lambda}, \hat{\beta}^{\star\prime}, \mathrm{vec}(\hat{\Gamma})', \hat{\sigma}_\xi^2, \hat{\tau}', \hat{\delta}')'$ are the estimator that maximize (4.19) respectively to λ, $\beta^\star$, $\mathrm{vec}(\Gamma)$, σ_ξ^2, τ and δ where $\delta = \Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$ and $\sigma_\xi^2 = \sigma_\upsilon^2 - \sigma_{\upsilon\epsilon}'\Sigma_\epsilon^{-1}\sigma_{\upsilon\epsilon}$. Finally the estimator of the parameter function $\beta(t)$ is defined by

$$\hat{\beta}(t) = \sum_{j=1}^{p_n} \hat{\beta}_j^\star \phi_j(t) \tag{4.20}$$

A necessary condition to obtain $\hat{\theta}$ is that $\frac{\partial \ln\tilde{L}_n(\theta)}{\partial\theta} = 0$, the first derivatives are given in Appendix A.

4.4 Monte Carlo simulation

To evaluate numerically the finite behavior of the proposed QMLE estimator, a Monte Carlo simulation was done using different parameters. Here we propose various endogeneity degree and different spatial interactions.

4.4.1 Data generating process

We start by generating a bivariate normal distribution variables $(\upsilon_{i,n}, \epsilon_{i,n}) \sim N\left(0, \begin{pmatrix} \sigma_\upsilon & \rho \\ \rho & \Sigma_\epsilon \end{pmatrix}\right)$ as disturbances used in the outcome equation and the spatial interactions. Let $\sigma_\upsilon = \Sigma_\epsilon = 1$, so the data generating process (DGP) is:

$$Y_i = \lambda \sum_{j=1}^{n} w_{ij} Y_j + \int_T X_i(t)\beta(t)dt + V_i \tag{4.21}$$

Like Muller & Stadtmuller (2005), $X(t) = \sum_{j=1}^{20} \omega_j \phi_j(t)$ is an explanatory random function with $\omega_j \sim N(0, 1/j)$ for $j \geq 1$ and $(\phi_j)_{j=1,\dots,20}$ are the first twenty functions of the Fourier orthogonal basis defined by $\left\{\phi_j(t) = \sqrt{2} sin(j\pi t),\ t \in [0,1], j = 1, 2, \dots.\right\}$. We also define the parameter function as $\beta(t) = \sum_{j=1}^{20} \beta_j \phi_j(t)$ with $\beta_j = (1.2, 1.3, 1.5, 1.4, .5, -.4, -.4)$ for $j \leq 7$ and $\beta_j = 0$ for $j > 7$.

The spatial endogenous matrix W_n is constructed as follow:

1. Build a weighted matrix W_n^d based on geographic distance.
2. Construct W_n^e as $w_{ij}^e = 1/|Z_i - Z_j|$ if $i \neq j$ and $w_{ij}^e = 0$ if $i = j$, where elements of Z_i are generated by $\mathbf{Z}_n = \mathbf{U}_n \Gamma + \epsilon_n$ with $\Gamma = (\gamma_1, \gamma_2) = (1, 0.8)$ is a vector of parameters and $\mathbf{U}_n = (u_1, \dots, u_n)$, $u_i = (u_i^1, u_i^2)$ where $u_i^1, u_i^2 \sim N(0,1)$
3. $W_n = W_n^d \circ W_n^e$ where $\circ$ is the Hadamard product.
4. Row-normalize W_n.

To construct W^d we generate a data grid with dimension $D = 60 \times 60$ locations, where we chose randomly n spatial units. Using those random locations, we construct W^d by two distinct scenarios:

- The first one is such that W_{ij}^d is simply the inverse of euclidean distance between location i and j: $W_{ij}^d = \frac{1}{\text{dist}(i,j)}$.
- For the second one, we use the nearest neighbors algorithm (kNN) with 3 neighbors so each location i has at least 3 neighbors.

For different degree of endogeneity; weak, medium and strong are considered towards the coefficient of correlation $\rho = 0.2$, 0.5 and 0.8. We set the spatial coefficient $\lambda = 0.2$ and 0.8 to look over the affects of the spatial correlation on the estimation. A number of $N = 300$ replications was carried out for each setting considering with two different sample sizes; $n = 200, 400$. Recall that we use an auxiliary parameters p_n; for the strategy to fit the parameters function $\beta(t)$ and reduce the dimension of the functional covariate and parameter. The eigenbasis based on functional principal components analysis is used for the truncation. The choice of p_n has an importance in the estimation procedure. For this choice, we consider three criteria, ASE, AIC and BIC defined respectively:

$$\begin{aligned} ASE &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ AIC &= 2k - 2\text{Ln}(L) \\ BIC &= -2Ln(L) + Ln(n)k \end{aligned}$$

where k is the number of parameters, n is the sample size and L is the maximum of the likelihood function. To evaluate the efficiency of the parameter function $\hat{\beta}(t)$ and measure the accuracy of the candidates model considered by the choices strategies of p_n, we use the integrated mean squared error (IMSE) (M. Escabias & Valderrama, 2007):

$$IMSE = \int (\beta(t) - \hat{\beta}(t))^2 \text{d}t \qquad (4.22)$$

We also perform a comparison between the proposed estimator and that of the functional SAR model assuming W_n exogenous (the one considered in the previous chapter).

4.4.2 Results

The results are reported in Tables 4.1 to 4.6, where we return the mean of each estimator (the empirical standard-deviation is between brackets) based on the 300 replication using the two alternative form of W_n described in Section 4.4.1. EFSAR refer to the proposed model of Section 4.2 and the functional autoregressive spatial model with the conventional structure of W_n is named by FSAR. The last are estimated using the same truncation level p_n selected for EFSAR model regardless each time one of the criteria listed previously.

The combination of different values of the different parameters in addition to the two types of physical matrix W^d, we obtain 24 cases, almost in all of them we note small bias for the parameters of the proposed method.

- For the true parameter $\lambda = 0.2$ with $\rho = 0.2$, the estimated spatial parameter $\hat{\lambda}$ has a bias of order 33% with small size sample when using kNN method for W^d. A bias with the same magnitude affect $\hat{\lambda}$ in case of $\rho = 0.5$ and W^d constructed by distance.

- The proposed method has difficulties to estimate $\hat{\rho}$ in case of high spatial dependence ($\lambda = 0.8$) using geographic distance for W^d when $\rho = 0.2$ and $n = 200$.

The estimated parameter $\hat{\lambda}$ with the conventional FSAR model suffers more with bias than the proposed model in almost all situations particularly in case of high endogeneity where the bias exceed a 100% of magnitude, see for instance the case with high spatial dependency $\rho = 0.8$ and $\lambda = 0.2$ with W^d defined as the inverse of distance between spatial units.

Giving the empirical standard deviation of the estimators, we notice that for a fixed degree of endogeneity the standard deviation decreases when spatial dependence increases. Similarly, for a fixed degree of spatial dependence, the standard deviation decreases when endogeneity increases.

The estimator of the functional parameter $\hat{\beta}(.)$ based on the criteria ASE, AIC and BIC versus the true coefficient $\beta(.)$ are given in Figures 4.1 to 4.4 using kNN and geographic distance to construct W^d. The integrated mean square errors (IMSE) show the good quality of these functional parameter estimates. This quality increases with the sample size. We notice also that for the same degree of endogeneity, the IMSE is larger for high spatial dependence. Even more, the IMSE decreases with endogeneity for a fixed spatial parameter λ.

Comparing the tables, we notice that obtained proposed estimation procedure with the different specifications of W_n leads to different empirical results. This is not surprising in the literature since the choice of the spatial weight matrix is not guided by any known economic theory, see among others Marbuah & Amuakwa-Mensah (2017).

4.5 Conclusion

The work proposed in this chapter is an extension of the functional spatial autoregressive model given in Chapter 3, that overcome the issue of endogeneity of the spatial weight matrix in cross-sectional FSAR model. It covers the case where the entry of the spatial weight matrix are defined by a stochastic process. This process can depend of the dependent variable in FSAR model through a present correlation between error in the stochastic process and disturbance in the FSAR outcome equation. We employed the quasi-maximum likelihood method to estimates parameters combined with a dimensional reduction approach. We conducted also a Monte Carlo simulation to investigate the finite sample of the proposed method. Results show that the proposed method in Chapter 3 under exogenous matrix produce bias when true weight matrix is endogenous specially with high level of endogeneity. On the other hand our estimates have a good finite samples properties. For future researchers, we would like to applied this structure on a real situation.

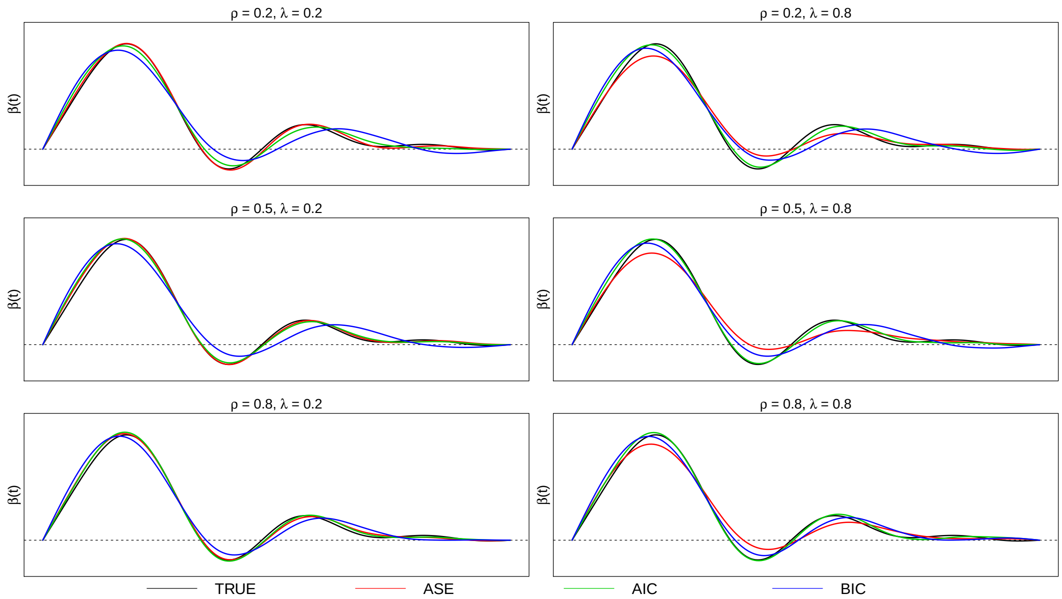

Figure 4.1: Estimation of $\hat{\beta}(.)$ with W^d based on distance and $n = 200$

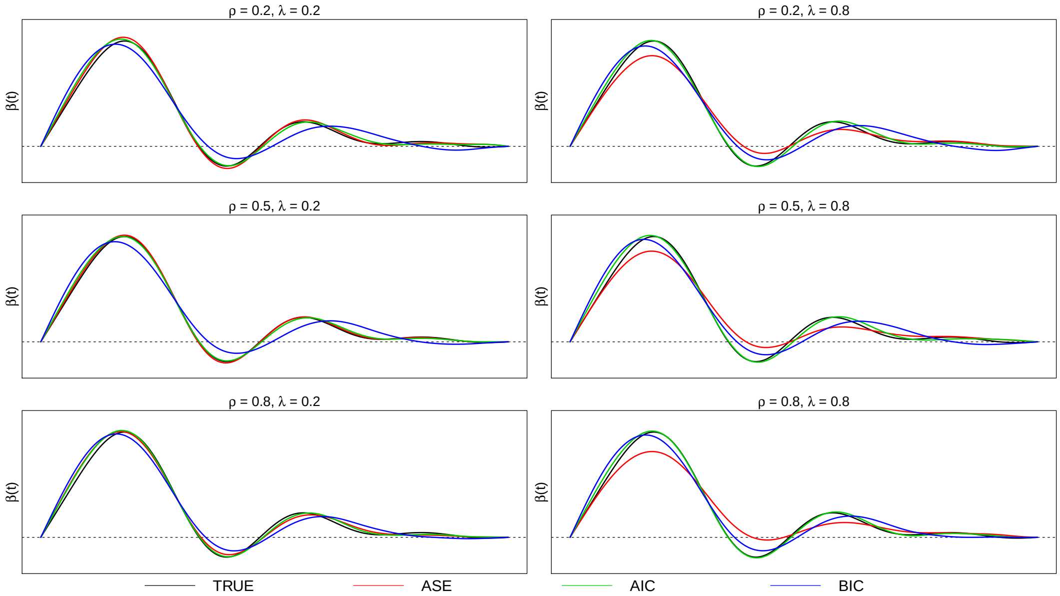

Figure 4.2: Estimation of $\hat{\beta}(.)$ with W^d based on kNN with 3 neighbors and $n = 200$

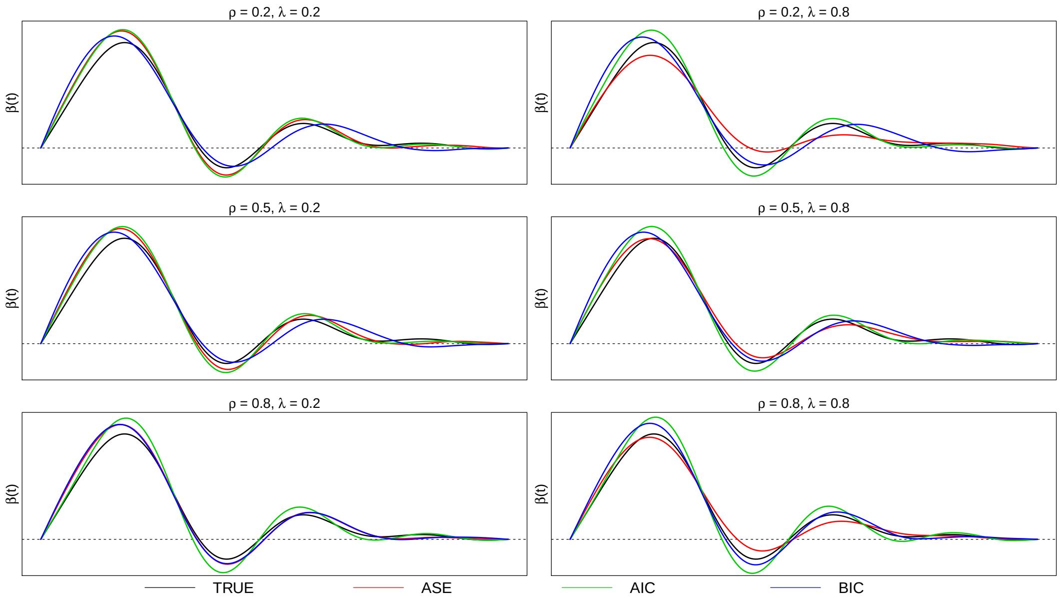

Figure 4.3: Estimation of $\hat{\beta}(.)$ with W^d based on distance and $n = 400$

Table 4.1: Estimation with weak endogeneity (n=200)

$\rho = 0.2$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.1981	0.2214	0.1993	0.2226	0.199	0.2222	0.2604	0.2643	0.2618	0.2656	0.2618	0.2656
	(0.0557)	(0.0584)	(0.0564)	(0.059)	(0.0575)	(0.06)	(0.0376)	(0.0385)	(0.0374)	(0.0383)	(0.0376)	(0.0383)
$\hat{\sigma}_\upsilon$	0.8964	0.8927	0.9008	0.8974	0.9129	0.9096	0.9209	0.9172	0.9244	0.9212	0.9373	0.9344
	(0.0188)	(0.0186)	(0.0202)	(0.0201)	(0.0218)	(0.0216)	(0.0244)	(0.0246)	(0.0251)	(0.0254)	(0.0286)	(0.0289)
$\hat{\Sigma}_\epsilon$	0.9505	–	0.9505	–	0.9505	–	0.9059	–	0.9059	–	0.9058	–
	(0.0002)		(0.0002)		(0.0218)		(0.0002)		(0.0002)		(0.0286)	
$\hat{\sigma}_\xi$	0.8487	–	0.8527	–	0.8648	–	0.8304	–	0.8332	–	0.8462	–
	(0.018)		(0.019)		(0.0204)		(0.0224)		(0.0229)		(0.0266)	
$\hat{\rho}$	0.2125	–	0.2134	–	0.2133	–	0.286	–	0.2872	–	0.287	–
	(0.0148)		(0.0148)		(0.0145)		(0.0137)		(0.014)		(0.0143)	
$\hat{\gamma}_1$	0.955	–	0.955	–	0.9548	–	1.0761	–	1.076	–	1.0762	–
	(0.0032)		(0.0032)		(0.0033)		(0.0042)		(0.0041)		(0.0043)	
$\hat{\gamma}_2$	0.7431	–	0.743	–	0.7431	–	0.7925	–	0.7925	–	0.7922	–
	(0.0035)		(0.0036)		(0.0036)		(0.0048)		(0.0049)		(0.0047)	
$\hat{\delta}$	0.2235	–	0.2246	–	0.2245	–	0.3157	–	0.317	–	0.3168	–
	(0.0156)		(0.0156)		(0.0153)		(0.0151)		(0.0154)		(0.0157)	
PCs	7	–	4	–	4	–	7	–	5	–	4	–
	(0.8253)		(1.1406)		(0.6297)		(0.8384)		(1.0978)		(0.7378)	
IMSE	0.0465	0.0491	0.0482	0.0489	**0.0414**	0.0419	0.0561	0.0617	0.0549	0.0572	**0.0469**	0.0483
	(0.0331)	(0.0347)	(0.0321)	(0.0333)	(0.0205)	(0.0209)	(0.0477)	(0.0517)	(0.0457)	(0.0493)	(0.0383)	(0.0409)
$\lambda = 0.8$												
$\hat{\lambda}$	0.7789	0.7812	0.7824	0.7845	0.7822	0.7844	0.8097	0.8071	0.8126	0.8100	0.8127	0.8101
	(0.0381)	(0.0374)	(0.0363)	(0.0356)	(0.0367)	(0.0359)	(0.0138)	(0.0137)	(0.0127)	(0.0127)	(0.0128)	(0.0128)
$\hat{\sigma}_\upsilon$	0.9467	0.9458	0.8869	0.8860	0.8995	0.8986	0.9861	0.9888	0.8982	0.9006	0.9123	0.9149
	(0.1238)	(0.1238)	(0.0219)	(0.0218)	(0.0246)	(0.0247)	(0.1829)	(0.1837)	(0.0281)	(0.0284)	(0.0288)	(0.0291)
$\hat{\Sigma}_\epsilon$	1.0530		1.0530		1.0530		1.0041		1.0041		1.0041	
	(0.0000)		(0.0000)		(0.0246)		(0.0001)		(0.0000)		(0.0288)	
$\hat{\sigma}_\xi$	0.9428	–	0.8835	–	0.8960	–	0.9524	–	0.8654	–	0.8794	–
	0.1236		0.0217		0.0245		0.1809		0.0277		0.0287	
$\hat{\rho}$	0.0599	–	0.0572	–	0.0576	–	0.1820	–	0.1809	–	0.1812	–
	0.0228		0.0168		0.0167		0.0268		0.0160		0.0160	
$\hat{\gamma}_1$	0.9363	–	0.9365	–	0.9365	–	1.0127	–	1.0123	–	1.0123	–
	0.0013		0.0010		0.0011		0.0040		0.0028		0.0028	
$\hat{\gamma}_2$	0.6421	–	0.6418	–	0.6418	–	0.8788	–	0.8796	–	0.8794	–
	0.0013		0.0010		0.0010		0.0041		0.0029		0.0030	
$\hat{\delta}$	0.0568	–	0.0544	–	0.0547	–	0.1812	–	0.1801	–	0.1805	–
	0.0216		0.0160		0.0159		0.0267		0.0159		0.0159	
PCs	4	–	4	–	4	–	7	–	5	–	4	–
	(1.6002)		(1.1329)		(0.6300)		(1.8471)		(1.1187)		(0.7688)	
IMSE	0.0749	0.0746	0.0494	0.0493	**0.0440**	0.0439	0.0904	0.0905	0.0580	0.0586	**0.0484**	0.0483
	(0.0643)	(0.0645)	(0.0335)	(0.0335)	(0.0253)	(0.0252)	(0.0789)	(0.0785)	(0.0452)	(0.0458)	(0.0327)	(0.0335)

Table 4.2: Estimation with medium endogeneity (n=200)

$\rho = 0.5$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.2660	0.3830	0.2670	0.3844	0.2668	0.3840	0.1995	0.2346	0.1999	0.2350	0.1996	0.2346
	(0.0579)	(0.0586)	(0.0588)	(0.0586)	(0.0597)	(0.0593)	(0.0347)	(0.0372)	(0.0347)	(0.0373)	(0.0345)	(0.0372)
$\hat{\sigma}_v$	1.0894	1.0579	1.0928	1.0621	1.1055	1.0759	1.0264	1.0057	1.0309	1.0115	1.0430	1.0253
	(0.0274)	(0.0301)	(0.0281)	(0.0299)	(0.0296)	(0.0310)	(0.0224)	(0.0239)	(0.0234)	(0.0252)	(0.0264)	(0.0279)
$\hat{\Sigma}_\epsilon$	1.0487	–	1.0487	–	1.0488	–	1.1376	–	1.1376	–	1.1375	–
	(0.0008)		(0.0008)		(0.0296)		(0.0012)		(0.0012)		(0.0264)	
$\hat{\sigma_\xi}$	0.8194	–	0.8199	–	0.8328	–	0.6655	–	0.6668	–	0.6788	–
	(0.0198)		(0.0190)		(0.0211)		(0.0155)		(0.0156)		(0.0184)	
$\hat{\rho}$	0.5319	–	0.5346	–	0.5344	–	0.6405	–	0.6434	–	0.6435	–
	(0.0187)		(0.0186)		(0.0193)		(0.0145)		(0.0146)		(0.0151)	
$\hat{\gamma}_1$	1.1124	–	1.1122	–	1.1123	–	1.0531	–	1.0534	–	1.0525	–
	(0.0088)		(0.0086)		(0.0088)		(0.0080)		(0.0080)		(0.0086)	
$\hat{\gamma}_2$	0.8864	–	0.8861	–	0.8860	–	0.8180	–	0.8181	–	0.8176	–
	(0.0076)		(0.0076)		(0.0077)		(0.0071)		(0.0074)		(0.0074)	
$\hat{\delta}$	0.5072	–	0.5097	–	0.5096	–	0.5631	–	0.5656	–	0.5657	–
	(0.0176)		(0.0175)		(0.0182)		(0.0126)		(0.0127)		(0.0132)	
PCs	7	–	5	–	4	–	7	–	7	–	4	–
	(1.0277)		(1.1213)		(0.6445)		(0.8663)		(1.1376)		(0.7469)	
IMSE	0.0471	0.0605	0.0470	0.0523	**0.0407**	0.0430	0.0426	0.0641	0.0422	0.0522	**0.0395**	0.0458
	(0.0306)	(0.0444)	(0.0317)	(0.0419)	(0.0223)	(0.0262)	(0.0323)	(0.0431)	(0.0321)	(0.0406)	(0.0280)	(0.0325)
$\lambda = 0.8$												
$\hat{\lambda}$	0.7791	0.7988	0.7820	0.8016	0.7821	0.8016	0.8055	0.8112	0.8089	0.8145	0.8089	0.8144
	0.0336	0.0378	0.0313	0.0358	0.0315	0.0362	(0.0168)	(0.018)	(0.0157)	(0.0171)	(0.0157)	(0.017)
$\hat{\sigma}_v$	1.0525	1.0315	0.9893	0.9677	0.9994	0.9794	1.1129	1.0972	1.0278	1.0119	1.0396	1.025
	(0.1260)	(0.1262)	(0.0257)	(0.0272)	(0.0270)	(0.0287)	(0.1648)	(0.1661)	(0.031)	(0.0324)	(0.0332)	(0.0342)
$\hat{\Sigma}_\epsilon$	1.0319	–	1.0322	–	1.0321	–	0.9285	–	0.9288	–	0.9288	–
	(0.0014)		(0.0010)		(0.0270)		(0.0014)		(0.001)		(0.0332)	
$\hat{\sigma_\xi}$	0.7080	–	0.6430	–	0.6538	–	0.7856	–	0.6995	–	0.7113	–
	(0.1237)		(0.0154)		(0.0176)		(0.1646)		(0.0208)		(0.0231)	
$\hat{\rho}$	0.5958	–	0.5977	–	0.5970	–	0.551	–	0.5521	–	0.552	–
	0.0237		0.0148		0.0155		(0.0203)		(0.014)		(0.0144)	
$\hat{\gamma}_1$	1.0630	–	1.0660	–	1.0656	–	0.9928	–	0.9959	–	0.9957	–
	(0.0126)		(0.0077)		(0.0083)		(0.0117)		(0.0076)		(0.0078)	
$\hat{\gamma}_2$	0.8315	–	0.8316	–	0.8313	–	0.8675	–	0.8695	–	0.8692	–
	(0.0110)		(0.0079)		(0.0085)		(0.0103)		(0.0069)		(0.0072)	
$\hat{\delta}$	0.5773	–	0.5791	–	0.5784	–	0.5934	–	0.5944	–	0.5944	–
	(0.0230)		(0.0144)		(0.0152)		(0.0218)		(0.0151)		(0.0155)	
PCs	4	–	5	–	4	–	7	–	5	–	4	–
	(1.6216)		(1.1139)		(0.7737)		(1.7543)		(1.0758)		(0.6996)	
IMSE	0.0726	0.0805	0.0415	0.0493	**0.0398**	0.0447	0.0834	0.0918	0.0484	0.0586	**0.0435**	0.0498
	(0.0702)	(0.0706)	(0.0288)	(0.0402)	(0.0250)	(0.0310)	(0.0747)	(0.0746)	(0.0373)	(0.0598)	(0.0305)	(0.0533)

Table 4.3: Estimation with strong endogeneity (n=200)

$\rho = 0.8$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.1912	0.4419	0.1921	0.4430	0.1917	0.4428	0.1910	0.2410	0.1914	0.2414	0.1910	0.2409
	(0.0431)	(0.0587)	(0.0429)	(0.0586)	(0.0426)	(0.0583)	(0.0248)	(0.0354)	(0.0249)	(0.0356)	(0.0249)	(0.0357)
$\hat{\sigma}_{\upsilon}$	0.9410	0.8443	0.9424	0.8450	0.9492	0.8548	0.9180	0.8894	0.9211	0.8922	0.9264	0.9003
	(0.0259)	(0.0344)	(0.0263)	(0.0348)	(0.0279)	(0.0356)	(0.0182)	(0.0214)	(0.0187)	(0.0220)	(0.0215)	(0.0243)
$\hat{\Sigma}_{\epsilon}$	1.0758	–	1.0758	–	1.0759	–	0.8585	–	0.8585	–	0.8585	–
	(0.0007)		(0.0007)		(0.0279)		(0.0008)		(0.0007)		(0.0215)	
$\hat{\sigma_{\xi}}$	0.3363	–	0.3334	–	0.3394	–	0.3799	–	0.3768	–	0.3828	–
	(0.0109)		(0.0077)		(0.0097)		(0.0132)		(0.0087)		(0.0107)	
$\hat{\rho}$	0.8064	–	0.8093	–	0.8098	–	0.6796	–	0.6835	–	0.6830	–
	(0.0172)		(0.0166)		(0.0172)		(0.0106)		(0.0100)		(0.0112)	
$\hat{\gamma}_1$	0.9748	–	0.9743	–	0.9741	–	0.9566	–	0.9564	–	0.9564	–
	(0.0102)		(0.0098)		(0.0104)		(0.0068)		(0.0066)		(0.0073)	
$\hat{\gamma}_2$	0.7850	–	0.7850	–	0.7848	–	0.7747	–	0.7748	–	0.7749	–
	(0.0089)		(0.0091)		(0.0095)		(0.0075)		(0.0070)		(0.0077)	
$\hat{\delta}$	0.7496	–	0.7522	–	0.7527	–	0.7916	–	0.7962	–	0.7956	–
	(0.0159)		(0.0153)		(0.0159)		(0.0123)		(0.0116)		(0.0129)	
PCs	7	–	7	–	4	–	7	–	7	–	4	–
	(1.0742)		(1.0400)		(0.8988)		(1.0438)		(1.0722)		(0.9217)	
IMSE	**0.0272**	0.0523	0.0278	0.0471	0.0287	0.0385	**0.0269**	0.0490	0.0285	0.0419	0.0315	0.0393
	(0.0188)	(0.0413)	(0.0183)	(0.0396)	(0.0171)	(0.0308)	(0.0162)	(0.0347)	(0.0165)	(0.0302)	(0.0151)	(0.0238)
$\lambda = 0.8$												
$\hat{\lambda}$	0.7850	0.8690	0.7881	0.8730	0.7879	0.8724	0.7968	0.8145	0.8005	0.8181	0.8004	0.8179
	(0.0240)	(0.0341)	(0.0222)	(0.0348)	(0.0228)	(0.0332)	(0.0131)	(0.0174)	(0.0101)	(0.0158)	(0.0100)	(0.0156)
$\hat{\sigma}_{\upsilon}$	0.7690	0.7069	0.7281	0.6652	0.7333	0.6722	1.0237	0.9783	0.9169	0.8701	0.9233	0.8793
	(0.0962)	(0.0963)	(0.0202)	(0.0292)	(0.0217)	(0.0299)	(0.2086)	(0.2096)	(0.0249)	(0.0357)	(0.0258)	(0.0362)
$\hat{\Sigma}_{\epsilon}$	0.8945	–	0.8945	–	0.8945	–	1.0090	–	1.0110	–	1.0107	–
	(0.0012)		(0.0010)		(0.0217)		(0.0043)		(0.0020)		(0.0258)	
$\hat{\sigma_{\xi}}$	0.3422	–	0.3013	–	0.3063	–	0.4666	–	0.3525	–	0.3592	–
	(0.0910)		(0.0075)		(0.0096)		(0.2073)		(0.0095)		(0.0113)	
$\hat{\rho}$	0.6177	–	0.6178	–	0.6179	–	0.7493	–	0.7553	–	0.7550	–
	(0.0173)		(0.0133)		(0.0138)		(0.0269)		(0.0136)		(0.0138)	
$\hat{\gamma}_1$	1.0388	–	1.0370	–	1.0372	–	0.9182	–	0.9293	–	0.9281	–
	(0.0152)		(0.0103)		(0.0110)		(0.0231)		(0.0077)		(0.0088)	
$\hat{\gamma}_2$	0.7839	–	0.7825	–	0.7827	–	0.8469	–	0.8514	–	0.8511	–
	(0.0129)		(0.0090)		(0.0093)		(0.0199)		(0.0094)		(0.0103)	
$\hat{\delta}$	0.6905	–	0.6906	–	0.6907	–	0.7426	–	0.7470	–	0.7469	–
	(0.0190)		(0.0143)		(0.0149)		(0.0263)		(0.0134)		(0.0136)	
PCs	7	–	7	–	4	–	7	–	7	–	4	–
	(1.5130)		(0.9926)		(0.9708)		(1.8997)		(1.0654)		(0.9124)	
IMSE	0.0487	0.0560	**0.0248**	0.0354	0.0285	0.0336	0.0839	0.0961	**0.0271**	0.0419	0.0303	0.0382
	(0.0574)	(0.0555)	(0.0143)	(0.0229)	(0.0140)	(0.0183)	(0.0907)	(0.0868)	(0.0154)	(0.0326)	(0.0144)	(0.0227)

Table 4.4: Estimation with weak endogeneity (n=400)

$\rho = 0.2$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.2239	0.2535	0.2248	0.2546	0.2254	0.255	0.183	0.1852	0.1836	0.1858	0.1835	0.1857
	(0.0495)	(0.0514)	(0.0497)	(0.0516)	(0.05)	(0.052)	(0.03)	(0.0303)	(0.0302)	(0.0305)	(0.03)	(0.0303)
$\hat{\sigma}_v$	0.9318	0.929	0.9303	0.9275	0.9387	0.936	1.0719	1.071	1.0711	1.0703	1.0816	1.0808
	(0.0124)	(0.0124)	(0.0109)	(0.011)	(0.0145)	(0.0146)	(0.0118)	(0.0118)	(0.0114)	(0.0114)	(0.0143)	(0.0143)
$\hat{\Sigma}_\epsilon$	1.0768		1.0768		1.0768		1.0684		1.0684		1.0684	
	(0.0000)		(0.0000)		(0.0145)		-0,0001		-0,0001		(0.0143)	
$\hat{\sigma_\xi}$	0.8889	–	0.887	–	0.8955	–	1.0358	–	1.0347	–	1.0451	–
	(0.0117)		(0.0098)		(0.0135)		(0.0119)		(0.0113)		(0.0141)	
$\hat{\rho}$	0.2147	–	0.2156	–	0.2153	–	0.196	–	0.197	–	0.1972	–
	(0.0088)		(0.009)		(0.0096)		(0.0081)		(0.0083)		(0.0091)	
$\hat{\gamma}_1$	0.9751	–	0.9751	–	0.9751	–	1.09	–	1.0902	–	1.09	–
	(0.002)		(0.0019)		(0.0021)		(0.0016)		(0.0016)		(0.0017)	
$\hat{\gamma}_2$	0.8374	–	0.8375	–	0.8372	–	0.7896	–	0.7896	–	0.7896	–
	(0.0016)		(0.0018)		(0.0019)		(0.0012)		(0.0013)		(0.0015)	
$\hat{\delta}$	0.1994	–	0.2002	–	0.1999	–	0.1835	–	0.1844	–	0.1846	–
	(0.0081)		(0.0083)		(0.0089)		(0.0075)		(0.0077)		(0.0085)	
PCs	7	–	7	–	4	–	7	–	5	–	4	–
	(0.9319)		(1.0729)		(0.7744)		(0.9186)		(1.0678)		(0.7007)	
IMSE	**0.0475**	0.0488	0.0592	0.059	0.0477	0.0477	0.0511	0.0529	0.0599	0.0600	**0.0461**	0.0458
	(0.0348)	(0.0363)	(0.0486)	(0.0488)	(0.0369)	(0.0368)	(0.0355)	(0.0366)	(0.0449)	(0.0443)	(0.0285)	(0.0284)
$\lambda = 0.8$												
$\hat{\lambda}$	0.8061	0.8076	0.8109	0.8123	0.8109	0.8123	0.8001	0.8006	0.8023	0.8027	0.8023	0.8028
	(0.0256)	(0.0252)	(0.0234)	(0.023)	(0.0234)	(0.023)	(0.0132)	(0.0129)	(0.0124)	(0.012)	(0.0125)	(0.0121)
$\hat{\sigma}_v$	1.0801	1.0796	0.9702	0.9697	0.9804	0.9799	1.1413	1.1402	1.0607	1.0595	1.0704	1.0693
	(0.1805)	(0.1802)	(0.0132)	(0.0132)	(0.0164)	(0.0164)	(0.155)	(0.1555)	(0.0179)	(0.0178)	(0.0197)	(0.0196)
$\hat{\Sigma}_\epsilon$	0.9088	–	0.9088	–	0.9088	–	1.1479	–	1.1479	–	1.1479	–
	(0.0000)		(0.0000)		(0.0164)		(0.0001)		(0.0000)		(0.0197)	
$\hat{\sigma_\xi}$	1.0773	–	0.9676	–	0.9779	–	1.0934	–	1.0126	–	1.0222	–
	(0.1799)		(0.0132)		(0.0164)		(0.1557)		(0.0172)		(0.0189)	
$\hat{\rho}$	0.0485	–	0.0482	–	0.0476	–	0.2339	–	0.235	–	0.2352	–
	(0.0157)		(0.0073)		(0.0079)		(0.0168)		(0.0081)		(0.0091)	
$\hat{\gamma}_1$	0.9964	–	0.9966	–	0.9966	–	1.1191	–	1.1182	–	1.1182	–
	(0.001)		(0.0005)		(0.0005)		(0.0035)		(0.0019)		(0.0021)	
$\hat{\gamma}_2$	0.8096	–	0.8094	–	0.8094	–	0.8055	–	0.8061	–	0.806	–
	(0.0008)		(0.0004)		(0.0005)		(0.0029)		(0.0016)		(0.0017)	
$\hat{\delta}$	0.0533	–	0.053	–	0.0524	–	0.2038	–	0.2048	–	0.2049	–
	(0.0173)		(0.0081)		(0.0087)		(0.0147)		(0.007)		(0.0079)	
PCs	3	–	7	–	4	–	7	–	5	–	4	–
	(1.7386)		(1.0438)		(0.7192)		(1.7454)		(1.0242)		(0.6767)	
IMSE	0.0936	0.0936	0.0597	0.0594	**0.0466**	0.0466	0.0856	0.0867	0.0546	0.0551	**0.0448**	0.0450
	(0.077)	(0.077)	(0.0439)	(0.0436)	(0.0317)	(0.0316)	(0.0732)	(0.0734)	(0.0401)	(0.041)	(0.0267)	(0.026)

Table 4.5: Estimation with medium endogeneity (n=400)

$\rho = 0.5$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.2694	0.3844	0.2703	0.3857	0.2701	0.3853	0.2334	0.2333	0.2342	0.2339	0.234	0.2338
	(0.0385)	(0.0425)	(0.0385)	(0.0422)	(0.0387)	(0.0425)	(0.0259)	(0.0284)	(0.0259)	(0.0283)	(0.0257)	(0.028)
$\hat{\sigma}_v$	1.0265	1.0028	1.0248	1.001	1.0328	1.0095	1.0219	1.0172	1.0201	1.0154	1.0281	1.0242
	(0.0172)	(0.0194)	(0.0166)	(0.0189)	(0.0189)	(0.021)	(0.0134)	(0.0142)	(0.0118)	(0.0124)	(0.0147)	(0.0154)
$\hat{\Sigma}_\epsilon$	0.9646	–	0.9646	–	0.9646	–	0.9148	–	0.9148	–	0.9148	–
	(0,0004)		(0,0004)		(0.0189)		(0,0002)		(0,0002)		(0.0147)	
$\hat{\sigma_\xi}$	0.8502	–	0.8471	–	0.8554	–	0.732	–	0.7279	–	0.7359	–
	(0.0125)		(0.0103)		(0.0131)		(0.0127)		(0.0096)		(0.0124)	
$\hat{\rho}$	0.4123	–	0.4138	–	0.4134	–	0.5149	–	0.517	–	0.5169	–
	(0.0116)		(0.0117)		(0.0122)		(0.0065)		(0.0063)		(0.0071)	
$\hat{\gamma}_1$	1.0141	–	1.0139	–	1.0141	–	0.9150	–	0.9151	–	0.9153	–
	(0.0041)		(0.004)		(0.0041)		(0.0033)		(0.0034)		(0.0039)	
$\hat{\gamma}_2$	0.6812	–	0.6812	–	0.6812	–	0.8214	–	0.8216	–	0.8214	–
	(0.0039)		(0.0038)		(0.0039)		(0.0034)		(0.0035)		(0.0037)	
$\hat{\delta}$	0.4274	–	0.429	–	0.4286	–	0.5629	–	0.5651	–	0.5651	–
	(0.0119)		(0.012)		(0.0125)		(0.0071)		(0.0069)		(0.0077)	
PCs	7	–	7	–	4	–	7	–	7	–	4	–
	(1.0013)		(1.0468)		(0.7609)		(1.0329)		(1.0619)		(0.8943)	
IMSE	**0.0450**	0.0504	0.0549	0.0549	0.0447	0.045	**0.0426**	0.0545	0.0558	0.0605	0.0458	0.0479
	(0.03)	(0.0368)	(0.0391)	(0.0419)	(0.027)	(0.031)	(0.03)	(0.0416)	(0.0422)	(0.0512)	(0.0332)	(0.0389)
$\lambda = 0.8$												
$\hat{\lambda}$	0.7994	0.8676	0.8014	0.8700	0.8013	0.8697	0.7934	0.8072	0.7955	0.8094	0.7954	0.8093
	(0.0262)	(0.0273)	(0.0256)	(0.0267)	(0.0256)	(0.0268)	(0.0114)	(0.0125)	(0.0101)	(0.0114)	(0.0103)	(0.0116)
$\hat{\sigma}_v$	1.0271	0.9906	0.99	0.9537	0.998	0.962	1.0476	1.0277	0.975	0.9549	0.9825	0.9629
	(0.079)	(0.0796)	(0.0175)	(0.0193)	(0.0188)	(0.0206)	(0.1161)	(0.1168)	(0.0152)	(0.0165)	(0.0182)	(0.0194)
$\hat{\Sigma}_\epsilon$	0.9498	–	0.9498	–	0.9498	–	1.0307	–	1.0306	–	1.0307	
	(0,0003)		(0,0003)		(0.0188)		(0,0003)		(0,0002)		(0.0182)	
$\hat{\sigma_\xi}$	0.8385	–	0.8011	–	0.8091	–	0.8043	–	0.7291	–	0.7368	–
	(0.0798)		(0.0095)		(0.0115)		(0.1172)		(0.0097)		(0.0133)	
$\hat{\rho}$	0.423	–	0.4234	–	0.4234	–	0.5005	–	0.5034	–	0.5032	–
	(0.0147)		(0.0123)		(0.0127)		(0.0158)		(0.0088)		(0.0091)	
$\hat{\gamma}_1$	1.0293	–	1.0295	–	1.0294	–	0.9741	–	0.9726	–	0.9726	–
	(0.0048)		(0.0041)		(0.0042)		(0.0074)		(0.0035)		(0.0038)	
$\hat{\gamma}_2$	0.7959	–	0.7952	–	0.7954	–	0.8596	–	0.8596	–	0.8596	–
	(0.0055)		(0.004)		(0.0042)		(0.006)		(0.0034)		(0.0038)	
$\hat{\delta}$	0.4453	–	0.4458	–	0.4457	–	0.4856	–	0.4884	–	0.4882	–
	(0.0154)		(0.0128)		(0.0133)		(0.0153)		(0.0085)		(0.0088)	
PCs	7	–	7	–	4	–	3	–	7	–	5	–
	(1.4398)		(1.0036)		(0.7649)		(1.621)		(1.0392)		(0.8685)	
IMSE	0.0575	0.0583	0.0504	0.0504	**0.0424**	0.0417	0.0772	0.0804	0.0561	0.0571	**0.0472**	0.0470
	(0.0487)	(0.0491)	(0.0387)	(0.0386)	(0.0289)	(0.0247)	(0.0665)	(0.066)	(0.045)	(0.0474)	(0.0414)	(0.0415)

Table 4.6: Estimation with strong endogeneity (n=400)

$\rho = 0.8$	Distance						kNN					
	ASE		AIC		BIC		ASE		BIC		BIC	
$\lambda = 0.2$	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR	EFSAR	FSAR
$\hat{\lambda}$	0.1618	0.3947	0.162	0.3956	0.1618	0.3953	0.165	0.2199	0.1654	0.2204	0.1653	0.2203
	(0.0302)	(0.0423)	(0.0302)	(0.0421)	(0.0306)	(0.042)	(0.0196)	(0.0282)	(0.0199)	(0.0281)	(0.02)	(0.0283)
$\hat{\sigma}_v$	0.9429	0.8793	0.941	0.8761	0.9446	0.8811	1.03	1.0079	1.0292	1.0055	1.0326	1.0107
	(0.0139)	(0.0205)	(0.013)	(0.0186)	(0.014)	(0.019)	(0.0115)	(0.013)	(0.0101)	(0.0114)	(0.012)	(0.0131)
$\hat{\Sigma}_\epsilon$	0.9692	–	0.9692	–	0.9692	–	1.0404	–	1.0405	–	1.0404	–
	(0,0005)		(0,0005)		(0.014)		(0,0009)		(0,0008)		(0.0120)	
$\hat{\sigma_\xi}$	0.3508	–	0.3457	–	0.3493	–	0.353	–	0.347	–	0.3507	–
	(0.0092)		(0.0039)		(0.0057)		(0.0098)		(0.0040)		(0.0058)	
$\hat{\rho}$	0.7575	–	0.7596	–	0.7595	–	0.8392	–	0.8425	–	0.8422	–
	(0.0082)		(0.0082)		(0.0084)		(0.0069)		(0.0058)		(0.0067)	
$\hat{\gamma}_1$	0.9932	–	0.9929	–	0.9931	–	0.9897	–	0.9902	–	0.9899	–
	(0.0064)		(0.0061)		(0.0067)		(0.0052)		(0.0047)		(0.0055)	
$\hat{\gamma}_2$	0.7886	–	0.7889	–	0.7887	–	0.7797	–	0.7803	–	0.7798	–
	(0.0059)		(0.0053)		(0.0061)		(0.0047)		(0.0042)		(0.0046)	
$\hat{\delta}$	0.7816	–	0.7837	–	0.7837	–	0.8066	–	0.8097	–	0.8095	–
	(0.0087)		(0.0087)		(0.0089)		(0.0067)		(0.0058)		(0.0065)	
PCs	7	–	7	–	5	–	7	–	7	–	5	–
	(1.1101)		(0.8263)		(1.0476)		(1.1047)		(0.7376)		(0.979)	
IMSE	**0.0328**	0.0526	0.0374	0.0549	0.0371	0.0462	**0.0316**	0.0539	0.0376	0.0544	0.0369	0.046
	(0.0175)	(0.039)	(0.0236)	(0.0456)	(0.0222)	(0.0388)	(0.0149)	(0.04)	(0.0224)	(0.0421)	(0.0208)	(0.0324)
$\lambda = 0.8$												
$\hat{\lambda}$	0.8022	0.9611	0.8042	0.9601	0.8042	0.9626	0.7928	0.8047	0.7947	0.8065	0.7947	0.8066
	(0.0148)	(0.0291)	(0.0118)	(0.0182)	(0.0122)	(0.024)	(0.0095)	(0.0126)	(0.0078)	(0.0113)	(0.0078)	(0.0113)
$\hat{\sigma}_v$	1.0968	0.8982	1.039	0.8411	1.0419	0.8445	1.0184	0.997	0.9448	0.922	0.9491	0.9277
	(0.1245)	(0.1245)	(0.023)	(0.0316)	(0.0244)	(0.0327)	(0.1356)	(0.1364)	(0.0123)	(0.0155)	(0.014)	(0.0172)
$\hat{\Sigma}_\epsilon$	1.1078	–	1.1083	–	1.1083	–	1.0788	–	1.0789	–	1.0789	–
	(0.0018)		(0.0011)		(0.0244)		-0,0007		-0,0003		(0.014)	
$\hat{\sigma_\xi}$	0.3811	–	0.3225	–	0.3257	–	0.452	–	0.3786	–	0.3827	–
	(0.1232)		(0.0037)		(0.0056)		(0.1308)		(0.0047)		(0.0066)	
$\hat{\rho}$	0.8902	–	0.891	–	0.8908	–	0.7815	–	0.7815	–	0.7817	–
	(0.0218)		(0.0139)		(0.0148)		(0.0163)		(0.0069)		(0.0076)	
$\hat{\gamma}_1$	0.9585	–	0.9577	–	0.9576	–	1.0037	–	1.0048	–	1.0049	–
	(0.0103)		(0.0068)		(0.0074)		(0.0096)		(0.0039)		(0.0045)	
$\hat{\gamma}_2$	0.8006	–	0.7964	–	0.7966	–	0.7988	–	0.7962	–	0.7962	–
	(0.0135)		(0.0072)		(0.0078)		(0.0095)		(0.0036)		(0.0044)	
$\hat{\delta}$	0.8035	–	0.8039	–	0.8038	–	0.7245	–	0.7244	–	0.7245	–
	(0.0193)		(0.0119)		(0.0127)		(0.0151)		(0.0064)		(0.007)	
PCs	7	–	7	–	5	–	7	–	7	–	5	–
	(1.6367)		(0.7021)		(1.0295)		(1.6747)		(0.7708)		(0.9624)	
IMSE	0.0635	0.069	**0.0384**	0.0494	0.039	0.0442	0.0725	0.0833	**0.0408**	0.0587	0.0409	0.0501
	(0.0644)	(0.0636)	(0.0244)	(0.0364)	(0.0228)	(0.0324)	(0.0665)	(0.0692)	(0.0303)	(0.052)	(0.029)	(0.0482)

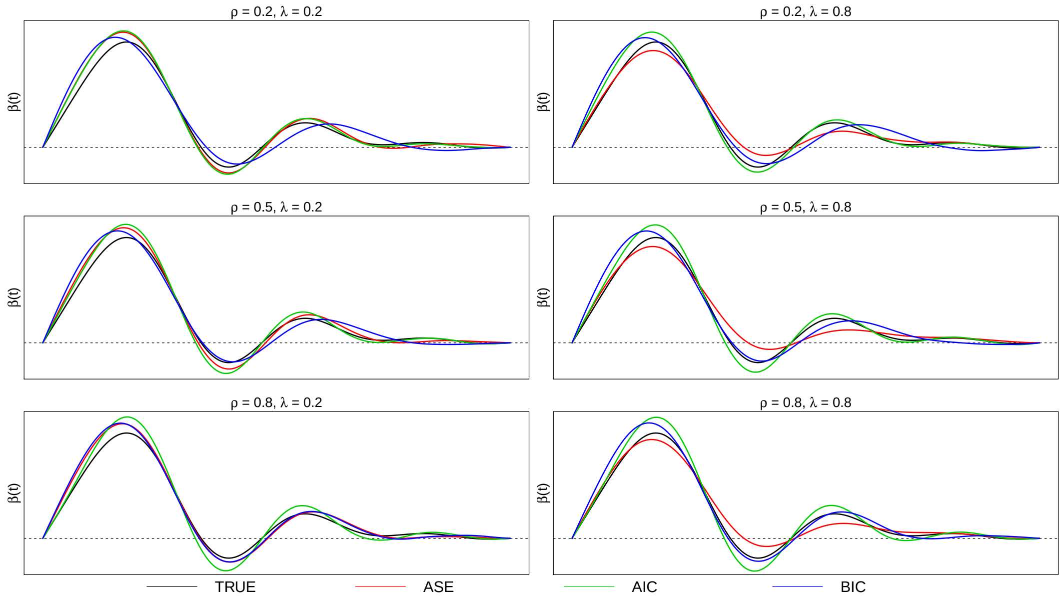

Figure 4.4: Estimation of $\hat{\beta}(.)$ with W^d based on kNN with 3 neighbors and $n = 400$

Appendix

Appendix A.

A.1. The first derivatives of the log-likelihood function $\ln\tilde{L}$ are:

- $$\frac{\partial \ln\tilde{L}_n(\theta)}{\partial\lambda} = \frac{1}{\sigma_\xi^2}(W_nY_n)'\xi_n(\theta) - tr[W_nS_n^{-1}];$$

- $$\frac{\partial \ln\tilde{L}(\theta)}{\partial\beta^\star} = \frac{1}{\sigma_\xi^2}\psi_{p_n}\xi_n(\theta);$$

- $$\frac{\partial \ln\tilde{L}(\theta)}{\partial \text{vec}(\Gamma)} = (\Sigma_\epsilon^{-1}\otimes U_n')\text{vec}(Z_n - U_n\Gamma) - \frac{1}{\sigma_\xi^2}\otimes(U_n'\xi_n(\theta));$$

- $$\frac{\partial \ln\tilde{L}(\theta)}{\partial\sigma_\xi^2} = -\frac{n}{2\sigma_\xi^2} + \frac{1}{2\sigma_\xi^4}\xi_n(\theta)'\xi_n(\theta);$$

- $$\frac{\partial \ln\tilde{L}(\theta)}{\partial\delta} = \frac{1}{\sigma_\xi^2}\xi_n(\theta)'\xi_n(\theta);$$

- $$\frac{\partial \ln\tilde{L}(\theta)}{\partial\tau} = -\frac{n}{2}\frac{\partial\ln|\Sigma_\epsilon|}{\partial\tau} - \frac{1}{2}\frac{\partial}{\partial\tau}tr[\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)],$$

where $\xi_n(\theta) = S_nY_n - \psi_{p_n}\beta^\star - \epsilon_n(\Gamma)$ and $\epsilon_n(\Gamma) = Z_n - U_n\Gamma$. Note that the J-dimensional vector $\frac{\partial\ln|\Sigma_\epsilon|}{\partial\tau}$ has the j-element of $\text{tr}(\Sigma_\epsilon^{-1}\frac{\partial\Sigma_\epsilon}{\tau_j})$ and $\frac{\partial}{\partial\tau}tr[\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)]$ has its jth element $-\text{tr}\Big(\Sigma_\epsilon^{-1}\frac{\partial\Sigma_\epsilon}{\partial\tau_j}\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)\Big)$ for $j = 1, \ldots, J$.

The expectation of the log-likelihood of (4.19) presented in Section 4.3 is:

$$\begin{aligned}\frac{1}{n}E\left(\mathrm{Ln}\tilde{L}_n(\theta)\right) &= -\ln(2\pi) - \frac{1}{2}(\sigma_\xi^2) - \frac{1}{n}\ln|\Sigma_\epsilon| + \frac{1}{n}E(\ln|S_n(\lambda)|) - \frac{1}{2}\mathrm{tr}(\Sigma_\epsilon^{-1}\Sigma_{\epsilon 0}) \\ & -\frac{1}{2n}\sum_{i=1}^{n} u_{in}'(\Gamma_0-\Gamma)\Sigma_\epsilon^{-1}(\Gamma_0-\Gamma)'u_{in} \\ & -\frac{1}{2n}\frac{\sigma_{\xi 0}^2}{\sigma_\xi^2}E[\mathrm{tr}(S_n^{-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{-1})] \\ & -\frac{1}{2\sigma_\xi^2}\Big((\lambda_0-\lambda),\ (\beta_0-\beta)',\ ((\Gamma-\Gamma_0)\delta)',\ (\delta_0-\delta)'\Big) \\ & H_{1n}\Big((\lambda_0-\lambda),\ (\beta_0-\beta)',\ ((\Gamma-\Gamma_0)\delta),\ (\delta_0-\delta)'\Big)'\end{aligned}$$

where $H_{1,n} = \frac{1}{n}E\left[\Big(G_n(\psi_{p_n}\beta_0+\epsilon_n\delta_0),\ \psi_{p_n}, U_n, \epsilon_n\Big)'\Big(G_n(\psi_{p_n}\beta_0+\epsilon_n\delta_0),\ \psi_{p_n}, U_n, \epsilon_n\Big)\right]$

A.2 For the hessian matrix $H_{\tilde{L}_n(\theta)}$ we present the second derivatives of the $\ln\tilde{L}_n(\theta)$:

- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\lambda} = -\mathrm{tr}[W_nS_n^{-1}(\lambda)]^2 - \frac{1}{\sigma_\xi^2}(W_nY_n)'W_nY_n$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\beta^\star} = -\frac{1}{\sigma_\xi^2}\psi_{p_n}'W_nY_n$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\mathrm{vec}(\Gamma)} = \frac{1}{\sigma_\xi^2}\delta\otimes(U_n'W_nY_n)$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}(W_nY_n)'\xi_n(\theta)$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\tau} = 0$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\lambda\partial\delta} = -\frac{1}{\sigma_\xi^2}\epsilon_n(\Gamma)'(W_nY_n)$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\beta^\star\partial\beta^{\star\prime}} = -\frac{1}{\sigma_\xi^2}\psi_{p_n}'\psi_{p_n}$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\beta^\star\partial\mathrm{vec}(\Gamma)'} = \frac{1}{\sigma_\xi^2}\otimes\delta(U_n'\psi_{p_n})$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\beta\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}\psi_{p_n}'\xi_n(\theta)$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\beta^\star\partial\tau'} = 0$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\beta^\star\partial\delta'} = -\frac{1}{\sigma_\xi^2}\psi_{p_n}'\epsilon_n(\Gamma)$
- $\frac{\partial^2 \ln\tilde{L}_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\mathrm{vec}(\Gamma)'} = -\Sigma_\epsilon^{-1}\otimes(U_n'U_n) - \frac{1}{\sigma_\xi^2}\delta\delta'\otimes(U_n'U_n)$

- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial \text{vec}(\Gamma)\partial\sigma_\xi^2} = \dfrac{1}{\sigma_\xi^4}\delta \otimes (U_n'\xi_n(\theta))$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial \text{vec}(\Gamma)\partial\tau'} = [I_{p_2} \otimes (U_n'\xi_n(\theta))]\dfrac{\partial \text{vec}(\Sigma_\epsilon^{-1})}{\partial\tau'}$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\delta\partial \text{vec}(\Gamma)'} = -\dfrac{1}{\sigma_\xi^2} I_{p_2} \otimes (U_n'\xi_n(\theta)) + \dfrac{1}{\sigma_\xi^2}\delta \otimes (U_n'\epsilon_n(\Gamma))$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\sigma_\xi^2\sigma_\xi^2} = \dfrac{n}{2\sigma_\xi^4} - \dfrac{1}{\sigma_\xi^6}\xi_n(\theta)'\xi_n(\theta)$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\sigma_\xi^2\partial\tau} = 0$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\sigma_\xi^2\partial\delta} = -\dfrac{1}{\sigma_\xi^2}\xi_n(\theta)'\xi_n(\theta)$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\tau\partial\tau'} = -\dfrac{n}{2}\dfrac{\partial^2 \ln|\Sigma_\epsilon|}{\partial\tau\partial\tau'} - \dfrac{1}{2}\dfrac{\partial^2}{\partial\tau\partial\tau'}\text{tr}\left[\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)\right]$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\tau\partial\delta'} = 0$
- $\dfrac{\partial^2 \ln \tilde{L}_n(\theta)}{\partial\delta\partial\delta'} = -\dfrac{1}{\sigma_\xi^2}\xi_n(\theta)'\xi_n(\theta)$

where $\dfrac{\partial^2 \ln|\Sigma_\epsilon|}{\partial\tau\partial\tau'}$ is a $J \times J$ matrix with the $(j,\ k)$th element of $\dfrac{\partial^2 \ln|\Sigma_\epsilon|}{\partial\tau_j\partial\tau_k} = -\text{t}\left(\Sigma_\epsilon^{-1}\dfrac{\partial\Sigma_\epsilon}{\partial\tau_k}\Sigma_\epsilon^{-1}\dfrac{\partial\Sigma_\epsilon}{\partial\tau_j}\right)$ and the $(j,\ k)$th element of $\dfrac{\partial^2}{\partial\tau\partial\tau'} = -\text{tr}\left[\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)\right]$ is
$\dfrac{\partial^2}{\partial\tau_j\partial\tau_k}\text{tr}\left[\Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)\right] = \text{tr}\left(\Sigma_\epsilon^{-1}\left(\dfrac{\partial\Sigma_\epsilon}{\partial\tau_k}\Sigma_\epsilon^{-1}\dfrac{\partial\Sigma_\epsilon}{\partial\tau_j} + \dfrac{\partial\Sigma_\epsilon}{\partial\tau_j}\Sigma_\epsilon^{-1}\dfrac{\partial\Sigma_\epsilon}{\partial\tau_k}\right) \times \Sigma_\epsilon^{-1}\epsilon_n(\Gamma)'\epsilon_n(\Gamma)\right)$
for $j,\ k = 1, \ldots, J$.

Thereupon we can define the information matrix of the QMLE $I_n = E\left(\dfrac{\partial^2 \ln \tilde{L}_n(\theta_0)}{\partial\theta\partial\theta'}\right)$ as:

$$I_n = \begin{pmatrix} I_{\lambda\lambda} & I'_{\lambda\beta} & I'_{\lambda \text{vec}(\Gamma)} & -E[tr(G_n)] & 0 & I'_{\lambda\delta} \\ * & \psi'_{p_n}\psi_{p_n} & \delta_0' \otimes (\psi_{p_n}U_{2n}) & 0 & 0 & 0 \\ * & * & I_{\text{vec}(\Gamma)\text{vec}(\Gamma')} & 0 & 0 & 0 \\ * & 0 & 0 & -\dfrac{n}{2\sigma_{\xi 0}^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{\tau\tau'} & 0 \\ * & 0 & 0 & 0 & 0 & -n\Sigma_{\epsilon 0} \end{pmatrix}$$

with
$I_{\lambda\lambda} = -\sigma_{\xi 0}^2 tr\left[E(G_n^2 + G_nG_n')\right] - E\left[(\psi_{p_n}\beta_0 + \epsilon_n\delta_0)'G_n'(\psi_{p_n}\beta_0 + \epsilon_n\delta_0)\right]$;
$I_{\lambda\beta} = -\psi_{p_n}E(G_n\psi_{p_n}\beta_0 + G_n\epsilon_n\delta_0)$;
$I_{\lambda \text{vec}(\Gamma)} = \delta_0 \otimes [U_n'E(G_n\psi_{p_n}\beta_0 + G_n\epsilon_n\delta_0)]$
$I_{\lambda\delta} = -E\left[\epsilon_n'G_n(\psi_{p_n}\beta_0 + \epsilon_n\delta_0)\right]$;

$$I_{\text{vec}(\Gamma)\text{vec}(\Gamma')} = -(\sigma_{\xi 0}^2 \Sigma_\epsilon^{-1} + \delta_0 \delta_0') \otimes (U_n' U_n) \ ;$$
$$I_\tau \tau' = -\frac{n\sigma_{\xi 0}^2}{2} tr \left(\Sigma_{\epsilon 0}^{-1} \frac{\partial \Sigma_{\epsilon 0}}{\partial \tau_k} \Sigma_{\epsilon 0}^{-1} \frac{\partial \Sigma_{\epsilon 0}}{\partial \tau_j} \right), \text{ for } j, k = 1, \ldots, J$$

Chapter 5

Partially linear Probit models with spatial heteroskedasticity

Contents

Résumé en français

Contrairement au Chapitres 3 et 4 où la dépendance spatiale est considérée au niveau de la variable indépendante du modèle, dans ce chapitre nous incorporons cette dernière au niveau de l'erreur afin de tenir compte d'une hétéroscédasticité spatiale entre les termes d'erreurs. De plus, le modèle proposé est plus flexible car il contient une partie non linéaire entre des variables explicatives et celle à expliquer. En effet, nous nous intéressons à un modèle partialement linéaire à choix binaire avec hétéroscédasticité spatiale et proposons une approche d'estimation récursive.

Nous supposons que nous disposons d'observations d'un vecteur aléatoire (Y, X, Z), collectées en des localisations spatiales $\{s_1, \ldots, s_n\}$ $(n \in \mathbb{N})$ appartenant à un treillis $\mathcal{I} \in \mathbb{R}^N$ $(N \geq 2)$ et situées à une distance minimale ρ les unes des autres, c.-à-d. $\forall\ s_i, s_j \in \mathcal{I}$, $\|s_i - s_j\| \geq \rho$ avec $\rho > 0$. On suppose que X et Z sont des variables explicatives à valeurs dans les sous-ensembles compacts $\mathcal{X} \subset \mathbb{R}^p (p \geq 1)$ et $\mathcal{Z} \subset \mathbb{R}^d (d \geq 1)$, respectivement, et considérons un modèle partiellement linéaire de la variable indépendante latente Y^* liée à Y :

$$Y^*_{s_i} = X^T_{s_i}\beta_0 + g_0(Z_{s_i}) + \sigma(s_i, \lambda_0)\varepsilon_{s_i}, \qquad i = 1, \ldots, n \tag{5.1}$$

avec

$$Y_{s_i} = \mathbb{I}\left(Y^*_{s_i} \geq 0\right), \qquad 1 \leq i \leq n, \tag{5.2}$$

où $\mathbb{I}$ est la fonction indicatrice. Le vecteur de paramètres β_0 $(p \times 1)$ appartient à un sous-ensemble compact $\Theta_\beta \subset \mathbb{R}^p$ et $g_0(\cdot)$ est une fonction inconnue à valeurs dans un espace de fonctions $\mathcal{G} = \{g \in C^2(\mathcal{Z}) : \|g\| = \sup_{z\in\mathcal{Z}} |g(z)| < C\}$, avec $C^2(\mathcal{Z})$ est l'espace de fonctions deux fois différentiables de $\mathcal{Z}$ à $\mathbb{R}$ et C est une constante positive. Les variables ε_{s_i} sont supposées indépendantes et identiquement distribuées, de distribution normale standard qu'on note $\Phi(\cdot)$. Par soucis de simplicité nous noterons dans la suite l'indice s_i par i, à la place de Y_{s_i}, X_{s_i} et Z_{s_i}, nous écrirons Y_{in}, X_{in} et Z_{in}, respectivement. La fonction $\sigma(\cdot, \cdot)$ est strictement positive et décrit une éventuelle hétéroscédasticité spatiale à l'aide de λ_0, un vecteur de paramètre $(q \times 1,\ q \geq 1)$ à estimer.

Notre objectif est d'estimer le modèle (5.1) avec l'approche de Severini & Staniswalis (1994) qui décompose le problème en deux étapes : l'estimation paramétrique de β_0 et λ_0 et suivie de celle non paramétrique de $g_0(\cdot)$. Pour ce faire, remarquons qu'à partir (5.2), l'espérance de Y_{in}, étant donné (X_{in}, Z_{in}), est définie par :

$$\mathbb{E}_0\left(Y_i | X_{in}, Z_{in}\right) = \Phi\left(\frac{X_{in}^T\beta_0 + g_0(Z_{in})}{\sigma(s_i, \lambda_0)}\right), \quad i = 1, \ldots, n \tag{5.3}$$

Ainsi pour chaque $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$, $z \in \mathcal{Z}$ et $\eta \in \mathbb{R}$, nous définissons l'espérance conditionnelle par rapport à Z_{in} du logarithme de la fonction de vraisemblance de Y_{in} $(1 \leq i \leq n,\ n = 1, 2, \ldots)$, comme :

$$H(\eta; \beta, \lambda, z) = \mathbb{E}_0\left(\mathcal{L}\left(\Phi\left(\frac{\eta + X_{in}^T\beta}{\sigma(s_i, \lambda)}\right); Y_{in}\right) \middle| Z_{in} = z\right), \tag{5.4}$$

avec $\mathcal{L}(u; v) = \log\left(u^v(1-u)^{1-v}\right)$ et nous considérons que $H(\eta; \beta, \lambda, z)$ est indépendante de i et de n. Pour tout $\beta \in \Theta_\beta$, $\lambda \in \Theta_\lambda$ et $z \in \mathcal{Z}$ fixé, soit $g_{\beta,\lambda}(z)$ la solution par rapport à η de

$$\frac{\partial}{\partial\eta} H(\eta; \beta, \lambda, z) = 0. \tag{5.5}$$

En pratique, $g_0(\cdot)$ n'est pas connue, on l'estime via la méthode de vraisemblance pondérée, à l'aide de (5.5), pour $\theta^T = (\beta^T, \lambda)$ fixé et $z \in \mathcal{Z}$. Soit $\hat{g}_\theta(\cdot)$ l'estimateur de $g_\theta(\cdot)$ ainsi obtenu, solution par rapport à η de :

$$\sum_{i=1}^{n} \frac{\partial}{\partial\eta}\mathcal{L}\left(\Phi\left(\frac{\eta + X_{in}^T\beta}{\sigma(s_i, \lambda)}\right); Y_{in}\right) K\left(\frac{z - Z_{in}}{b_n}\right) = 0, \tag{5.6}$$

où $K(\cdot)$ est un noyau de $\mathbb{R}^d$ à $\mathbb{R}_+$ et b_n une fenêtre de lissage qui dépend de n.

Après avoir estimé la partie non-paramétrique, nous utilisons $\hat{g}_\theta(\cdot)$ pour construire le profil de vraisemblance afin d'estimer θ^T en maximisant :

$$L_n(\theta \,|\, \hat{g}_\theta) = \sum_{i=1}^{n} Y_i \log\left(\Phi\left(\frac{X_{in}^T\beta + \hat{g}_\theta(Z_i)}{\sigma(s_i, \lambda)}\right)\right) + (1 - Y_i)\log\left(1 - \Phi\left(\frac{X_{in}^T\beta + \hat{g}_\theta(Z_i)}{\sigma(s_i, \lambda)}\right)\right).$$

L'estimateur de θ_0 est alors donné par :

$$\widehat{\theta} = \operatorname{argmax}_{\theta\in\Theta} L_n(\theta \,|\, \widehat{g}_\theta). \tag{5.7}$$

L'inconvénient de l'estimateur $\hat{g}_\theta(\cdot)$ via (5.6) est que le biais est plus important pour un z proche des limites de $\mathcal{Z}$, ce qui peut avoir un impact sur la qualité de l'estimateur $\hat{\theta}$ obtenu par (5.7). Pour remédier à ce problème, nous utilisons les observations Z_i éloignées des bords de $\mathcal{Z}$ pour maximiser $L_n(\theta \,|\, \hat{g}_\theta)$. On en déduit par la suite l'estimateur final $\hat{g}(\cdot) = \hat{g}_{\hat{\theta}}(\cdot)$. L'algorithme d'estimation est donné en Section 5.3.

Après avoir défini les estimateurs, nous donnons des indications pour établir leur comportement asymptotique sous certaines conditions :

$$\hat{\theta} \xrightarrow{P} \theta_0, \quad ||\hat{g}_{\hat{\theta}}(\cdot) - g_0(\cdot)|| = O_p(n^{-1/4}), \quad \text{avec} \quad n \to \infty \tag{5.8}$$

et

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \sigma_0^2 \Sigma_0^{-1}), \quad \text{où} \quad \Sigma_0 = E_0 \left\{ \frac{\partial^2}{\partial \theta_i \theta_j} L_n(\theta | \hat{g}_\theta(\cdot)) \right\}$$

Une étude numérique est menée sur des données simulées en Section 5.4 pour étudier la performance des estimateurs proposés pour échantillons de taille finie.

5.1 Introduction

When analyzing data in order to decide among a set of choices or even to classify objects, some usual statistical techniques are discrete choice models or qualitative choice models. Among these kind of models one may cite; binomial or multinomial choice models largely studied in the econometric literature. These models are particular cases of regression models where the response variable is qualitative and the model expressed as a linear combination of explicative variables.

This kind of models is of interest in this contribution. Namely, we focus on estimating the relationship between a qualitative dependent variable Y and a set of independent variables X given a spatial random sample of same distribution as (Y, X). The particular model considered is a spatial partial linear choice model with a non-linear component in the same spirit as the generalized additive model of Hastie & Tibshirani (1990), considered and extended in several directions by Severini & Staniswalis (1994), Carroll et al. (1997),Ruppert et al. (2003), Wong et al. (2014), among others.

A number of technical and practical investigations have been dedicated to handle spatial heterogeneity (see Anselin (1988, 1990), McMillen (1992) among others), spatial dependence or autocorrelation (Anselin (2002), Wang et al. (2013)) or spatial heteroskedasticity (Anselin & Rey (1991), Baltagi et al. (2003)) issues in regression models. The case of spatial heterogeneity in partial linear choice models is the baseline of this contribution.

Integrating spatial dependency in discrete choice models where agents must make decisions or choices, is an active area of research with applications in various areas, such as; energy, housing, agriculture, economics, environmental sciences, transportation, urban systems, ...

Several extensions of basic choice models and techniques are given in the literature since non-spatial approaches yield inconsistent estimates especially when estimating spatially correlated error terms discrete choice models. The case of binary discrete choice models with spatial error dependency is the center of a vast literature. To the best of our knowledge, the work of Case (1992) on neighbor influence on farmers attitude facing the use of technological tools, first proposed a consistent maximum likelihood estimate of a Probit model with spatial autoregressive errors (Spatial Error Probit model; SEPM). This model was investigated by Pinkse & Slade (1998) who used the generalized method of moments (GMM) estimation. The authors also gave a test to detect spatial correlation on the disturbances. McMillen (1992) proposed two estimation approaches of Probit model with spatial autocorrelation; the first one is based on EM algorithm while the second one uses weighted least squares method. More recently, a particular attention has been paid to Probit models with spatial errors, see for instance Fleming (2004),Beron & Vijverberg (2004), Holloway & Lapar (2007), Bhat & Sener (2009), Chakir & Parent (2009), Martinetti & Geniaux (2017) and among others.

In this chapter, we adopt the formulation of Harvey (1976) with a multiplicative heteroskedasticity function, to propose a partially linear Probit model (PLPSH) with a spatial disturbance variance. The estimation methodology is similar to the approach of Severini & Staniswalis (1994) based on the concept of "generalized profile likelihood" of Severini & Wong (1992). It consists on fixing first the parametric components and estimating the non-parametric one using "the weighted likelihood method" and then incorporating the obtained estimator to construct a "profile likelihood" to estimate the parametric components.

To this aims, the chapter is structured in the following way. We expose the proposed model in Section 5.2 and the estimation procedure and discuss asymptotic properties of the estimators whereas Section 5.3 gives the estimates computation steps. Section 5.4 yields the finite sample properties of the estimates.

5.2 Model

We consider that at n spatial locations $\{s_1, s_2, \ldots, s_n\}$ drawn from lattice $\mathcal{I} \in \mathbb{R}^N$ $(N \geq 2)$ and located at a minimum distance ρ from each other; i.e $\forall\ s_i, s_j \in \mathcal{I}$, $\|s_i - s_j\| \geq \rho$ with $\rho > 0$, observations of a random vector (Y, X, Z) are available. Assume that these observations follow the partially linear model of a latent dependent variable Y^*:

$$Y^*_{s_i} = X^T_{s_i}\beta_0 + g_0(Z_{s_i}) + \sigma(s_i, \lambda_0)\varepsilon_{s_i}, \qquad i = 1, \ldots, n, \tag{5.9}$$

with

$$Y_{s_i} = \mathbb{I}\left(Y^*_{s_i} \geq 0\right), \qquad 1 \leq i \leq n, \tag{5.10}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Let X and Z be 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 vector of unknown parameters β_0 $(p \times 1)$ 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 $\mathcal{G} = \{g \in C^2(\mathcal{Z}) : \|g\| = \sup_{z\in\mathcal{Z}} |g(z)| < C\}$, with $C^2(\mathcal{Z})$ the space of twice differentiable functions from $\mathcal{Z}$ to $\mathbb{R}$ and C a positive constant. Let ε_{s_i}, $i = 1 \ldots, n$ be i.i.d with marginal standard normal distribution $\Phi(\cdot)$ with $E(\varepsilon^2_{s_i}) = \sigma^2_0 = 1$. For simplicity, let us note in the following s_i by i when no confusion arises. Then, we write X_{in}, Y_{in} and Z_{in} for X_{s_i}, Y_{s_i} and Z_{s_i} respectively.

Following Harvey (1976), the multiplicative function of the disturbances variance $\sigma(\cdot,\cdot)$ is known as a non-zero positive function. This function describes the spatial heteroskedasticity at each location with respect to λ_0; a $q \times 1$ $(q \geq 1)$ vector to be estimated. We adopt the functional form of $\sigma(\cdot,\cdot)$ of Messner et al. (2014) who considered an extended logistic model. He suggested to use $\sigma(\mathbf{z}, \delta) = \exp(\mathbf{z}^T\delta)$ where δ is a vector of parameters and $\mathbf{z}$ is the vector of input variables. We consider that, in our subset of locations s_i, there is a set of k specific locations $b_i \in \mathcal{I}$ (as hot spots), $B = \{b_1, \ldots, b_k\}$ having impacts on the choice of individuals, namely:

$$\sigma(s_i, \lambda_0) = \exp\left(\lambda_0 \text{dist}(s_i, B)\right),$$

where $\lambda_0 \in \mathbb{R}$. The exponential function is used here as a simple way to ensure positive values for $\sigma(\cdot,\cdot)$.

The model (5.9) can be seen as a general case of several models. If one take $\sigma(\cdot,\cdot) = I_n$, where I_n is an identity vector, the model becomes the generalized partially linear model (e.g. Severini & Staniswalis, 1994) or the classical generalized additive model of (Hastie & Tibshirani, 1990). If $g_0(\cdot) = 0$, we obtain the regression model with a general multiplicative variance function presented in Harvey (1976). The model (5.9) is an alternative of the previously cited models by including the particular spatial heteroskedasticity considered or incorporating a non-linearity component for more flexibility. To estimate (5.9), we adopt the estimation approach of Severini & Staniswalis (1994), based on the concept of generalized profile likelihood (e.g Severini & Wong, 1992). That is a recursive approach consisting first on fixing the parametric parameter $\theta = (\beta^T, \lambda^T)^T$ and non-parametrically estimate $g_0(\cdot)$ using the weighted likelihood method. This last estimate is then used to construct a profile likelihood to estimate $\theta_0 = (\beta_0^T, \lambda_0^T)^T$.

In what follows, using the n observations (X_{in}, Y_{in}, Z_{in}), $i = 1, ..., n$, we give the parametric estimators of β_0, λ_0 and the non-parametric estimator of the smooth function $g_0(\cdot)$. To this end, we assume that, for all $n = 1, 2, \ldots$ and $1 \leq i \leq n$; $\{\varepsilon_{in}\}$ is independent of $\{X_{in}\}$ and $\{Z_{in}\}$, and $\{X_{in}\}$ is independent of $\{Z_{in}\}$. We give also 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 *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.

5.2.1 Estimation Procedure

As said before, the estimate procedure is based on the concept of a "generalized profile likelihood" (Severini & Staniswalis, 1994). The technique separates the estimating problem into two parts, which is an advantageous point in this approach. By equation (5.10), the conditional expectation value under the true parameters (i.e., β_0, λ_0 and $g_0(\cdot)$) of Y_{in} given (X_{in}, Z_{in}) is defined by:

$$\mathbb{E}_0\left(Y_i|X_{in}, Z_{in}\right) = \Phi\left(\frac{X_{in}^T\beta + g_0(Z_{in})}{\sigma(s_i, \lambda_0)}\right), \quad i = 1, \ldots, n \tag{5.11}$$

Thus 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 (5.11), for $1 \leq i \leq n$, $n = 1, 2, \ldots$, as

$$H(\eta; \beta, \lambda, z) = \mathbb{E}_0\left(\mathcal{L}\left(\Phi\left(\frac{\eta + X_{in}^T\beta}{\sigma(s_i, \lambda)}\right); Y_i\right)\middle| Z_{in} = z\right), \tag{5.12}$$

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

$$\frac{\partial}{\partial\eta} H(\eta; \beta, \lambda, z) = 0. \tag{5.13}$$

Since $g_\theta(\cdot)$ is not available in practice, we need to estimate it. Therefore by (5.13) and for fixed $\theta^T = (\beta^T, \lambda^T) \in \Theta$, the weighted likelihood method can estimate $g_\theta(z)$, for $z \in \mathcal{Z}$ by $\hat{g}_\theta(z)$, the solution in η of

$$\sum_{i=1}^{n} \frac{\partial}{\partial\eta}\mathcal{L}\left(\Phi\left(\frac{\eta + X_{in}^T\beta}{\sigma(s_i, \lambda)}\right); Y_{in}\right) K\left(\frac{z - Z_{in}}{b_n}\right) = 0, \tag{5.14}$$

where $K(\cdot)$ is a kernel from $\mathbb{R}^d$ to $\mathbb{R}_+$ and b_n is a bandwidth parameter.

Now, using $\hat{g}_\theta(\cdot)$, we construct the Profile likelihood estimates of β_0 and λ_0. This latter is given by:

$$L_n(\theta\,|\,\hat{g}_\theta) = \sum_{i=1}^{n} Y_i \log\left(\Phi\left(\frac{X_{in}^T\beta + \hat{g}_\theta(Z_i)}{\sigma(s_i,\lambda)}\right)\right) + (1-Y_i)\log\left(1-\Phi\left(\frac{X_{in}^T\beta + \hat{g}_\theta(Z_i)}{\sigma(s_i,\lambda)}\right)\right).$$

Hence, the estimator of θ_0 is given by

$$\widehat{\theta} = \text{argmax}_{\theta\in\Theta} L_n(\theta\,|\,\widehat{g}_\theta). \tag{5.15}$$

The estimate of g_0 is then $\widehat{g}_{\hat{\theta}}$.

A basic inconvenience of the estimator $\widehat{g}_\theta(z)$ proposed in (5.14) is that the bias of $\widehat{g}_\theta(z)$ is high for z near the boundary of $\mathcal{Z}$. Of course, this bias will affect the estimator of θ given in (5.15) 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 profile likelihood function $L_n(\theta\,|\,\widehat{g}_\theta)$ 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.

We give some indications to investigate infinite sample size properties of the estimators $\hat{\theta}$ and $\hat{g}_0(\cdot)$ based on the following assumptions.

Assumption A1. $\theta^T = (\beta^T, \lambda)$ takes values in a compact and convex set $\Theta = \Theta_\beta \times \Theta_\lambda \subset \mathbb{R}^p \times \mathbb{R}^q$ and $\theta_0^T = (\beta_0^T, \lambda_0)$ is in the interior of Θ.

Assumption A2. (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

$$\sup_{\theta\in\Theta, z\in\mathcal{Z}} \left|\frac{\partial}{\partial\eta} H(g(z);\theta,z)\right| \leq \gamma \quad\Longrightarrow\quad \sup_{\theta\in\Theta, z\in\mathcal{Z}} |g(z) - g_\theta(z)| \leq \varepsilon. \tag{5.16}$$

Assumption A3. For $\theta \in \Theta$ and $z \in \mathcal{Z}$, the functions $g_\theta(z)$ and $\hat{g}_\theta(z)$, the solutions of (5.13) and (5.14) respectively, satisfy

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

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

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

Assumption A4. (Local dependence). The density $f_{in}(\cdot)$ of Z_{in} exists, is continuous on $\mathcal{Z}$ uniformly on i and n and satisfies

$$\liminf_{n\to\infty} \inf_{z\in\mathcal{Z}} \frac{1}{n}\sum_{i=1}^{n} f_{in}(z) > 0. \tag{5.17}$$

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 A5. The kernel K satisfies $\int K(u)du = 1$. It is Lipshitzian, i.e there is a positive constant C such that

$$|K(u) - K(v)| \leq C\|u - v\| \qquad \text{for all} \;\; u, v \in \mathbb{R}^d.$$

Assumption A6. The bandwidth b_n satisfies $b_n \to 0$ and $nb_n^{3d+1} \to \infty$ as $n \to \infty$.

Assumption A7. The multiplicative function $\sigma(s_i, \lambda)$ are twice continuous differentiable functions with respect to λ and $\sup_{\lambda \in \Theta_\lambda} \left| \sigma^{-1}(s_i, \lambda) + \frac{d}{d\lambda}\sigma(s_i, \lambda) + \frac{d^2}{d\lambda^2}\sigma(s_i, \lambda) \right| < \infty$, uniformly on i and n.

Assumption A1 is standard and provides compactness condition. *Assumption A2* is similar to the one given by Severini & Staniswalis (1994) to ensure smoothness of $H(\cdot, \cdot, \cdot)$ around its extrema point $g_\theta(\cdot)$. *Assumption A3* is the non-parametric condition on $g_0(\cdot)$. Similar to assumption **A7** in Robinson (2011), *Assumption A4* is a generalization of the classical assumption, $\inf_z f(z) > 0$ in case of estimation probability density function $f(\cdot)$ with identically distributed or stationary random variables. *Assumption A7* requires the standard deviations of the errors to be uniformly bounded away from zero with bounded derivatives.

With assumptions A1 to A7 in place, one can give asymptotic results using similar lines as Severini & Wong (1992) and Ahmed (2017), Chapter 6;

$$||\hat{g}_{\hat{\theta}}(\cdot) - g_0(\cdot)|| = O_p(n^{-1/4}), \quad n \to \infty$$

with $||g|| = \sup_{s \in \mathcal{Z}} |g(s)|$,

$$\hat{\theta} \xrightarrow{P} \theta_0 \quad \text{as} \quad n \to \infty$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(0, \sigma_0^2 \Sigma_0^{-1}),$$

where,

$$\Sigma_0 = E_0 \left\{ \frac{\partial^2}{\partial \theta_i \theta_j} L_n(\theta | \hat{g}(\cdot)) \right\},$$

Σ_0 equal to the marginal Fisher information for θ.

5.3 Computation of the estimates

The aim of this section is to outline in detail how the regression parameters β, the spatial auto-correlation parameter λ and the non-linear function g_θ can be estimated. We begin with the computation of $\widehat{g}_\theta(z)$, which plays a crucial role in what follows. Let us first recall the estimation algorithm steps:

1. For each z and θ, $\hat{g}_\theta$ is obtained by solving (5.14) in η.

2. Estimate θ by solving $\frac{\partial}{\partial \theta_j} L_n(\theta | \hat{g}_\theta)$.

3. Estimate $g_0(\cdot)$ by solving again (5.14) in η using $\hat{\theta}$.

4. Repeat steps 2 and 3 until convergence.

The details of steps 2 and 3 are given in the following sections.

5.3.1 Computation of the non-parametric component estimate

An iterative method is needed to compute $\widehat{g}_\theta(z)$ solution of (5.14) for each fixed $\theta \in \Theta$ and $z \in \mathcal{Z}$. For fixed $\theta^T = (\beta^T, \lambda^T) \in \Theta$ and $z \in \mathcal{Z}$, let $\eta_\theta = g_\theta(z)$ and $\psi(\eta; \theta, z)$ denote the left-hand side of (5.14), which can be rewritten as

$$\psi(\eta; \theta, z) = \sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-1} \Lambda\left(G_i(\theta, \eta)\right) [Y_{in} - \Phi(G_i(\theta, \eta))] K\left(\frac{z - Z_{in}}{b_n}\right), \tag{5.18}$$

where

$$G_i(\theta, \eta) = \frac{X_{in}^T \beta + \eta}{\sigma(s_i, \lambda)} \quad \text{and} \quad \Lambda(\cdot) = \frac{\Phi'(\cdot)}{(1 - \Phi(\cdot))\Phi(\cdot)}.$$

Consider the Fisher information:

$$\begin{aligned} \varPsi(\eta_\theta; \theta, z) &= E_0\left(\frac{\partial}{\partial \eta} \psi(\eta; \theta, z)\Big|_{\eta = \eta_\theta} \Big| \{(X_{in}, Z_{in}),\ i = 1, \ldots, n\} \right) \\ &= -\sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-2} \Lambda\left(G_i(\theta, \eta_\theta)\right) \Phi'\left(G_i(\theta, \eta_\theta)\right) K\left(\frac{z - Z_{in}}{b_n}\right) + \\ &\quad + \sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-2} \Lambda'\left(G_i(\theta, \eta_\theta)\right) [\Phi\left(G_i(\theta_0, \eta_0)\right) - \\ &\quad \Phi\left(G_i(\theta, \eta_\theta)\right)] K\left(\frac{z - Z_{in}}{b_n}\right) \end{aligned} \tag{5.19}$$

Note that the second term in the right hand side of (5.19) is negligible when θ is near the true parameter θ_0. Because $\psi(\eta; \theta, z) = 0$ when $\eta = \widehat{g}_\theta(z)$, an initial estimate $\tilde{\eta}$ can be updated to $\eta^\dagger$ using Fisher's scoring method:

$$\eta^\dagger = \tilde{\eta} - \frac{\psi(\tilde{\eta}; \theta, z)}{\varPsi(\tilde{\eta}; \theta, z)}. \tag{5.20}$$

The iteration procedure (5.20) requests some starting value $\tilde{\eta} = \tilde{\eta}_0$ to ensure convergence of the algorithm. To this end, let us adopt the approach of Severini & Staniswalis (1994), which consists of supposing that for fixed $\theta \in \Theta$, there exists a $\tilde{\eta}_0$ satisfying $G_i(\theta, \tilde{\eta}_0) = \Phi^{-1}(Y_i)$ for $i = 1, \ldots, n$. Knowing that $G_i(\theta, \tilde{\eta}_0) = [\sigma(s_i, \lambda)]^{-1}\left(X_i^T \beta + \tilde{\eta}_0\right)$, we have $\tilde{\eta}_0 + X_i^T \beta = \sigma(s_i, \lambda) \Phi^{-1}(Y_i)$, $i = 1, ..., n$. Then, (5.20) can be updated using the following initial value:

$$\eta_0^\dagger = \tilde{\eta}_0 - \frac{\psi(\tilde{\eta}_0; \theta, z)}{\varPsi(\tilde{\eta}_0; \theta, z)} = \frac{\sum_{i=1}^n [\sigma(s_i, \lambda)]^{-1} \Lambda(C_i) \Phi'(C_i) \left[C_i - [\sigma(s_i, \lambda)]^{-1} X_i^T \beta\right] K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^n [\sigma(s_i, \lambda)]^{-2} \Lambda(C_i) \Phi'(C_i) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

where $C_i = \Phi^{-1}(Y_i)$, $i = 1, \ldots, n$, is computed using a slight adjustment because $Y_i \in \{0, 1\}$. 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, usually selected by applying some cross-validation approach. The latter was adopted by Su (2012) in the case of a spatial semi-parametric model. Because cross-validation may be very time consuming in some situations, as the partial linear model considered, we adopt the following approach of Severini & Staniswalis (1994) for more flexibility:

1. Consider the linear regression of C_i on X_i, $i = 1, \ldots, n$, without an intercept term, and let $R_1, \ldots, R_n$ denote the corresponding residuals.

2. Since we expect $\mathbb{E}(R_i|Z_{in} = z)$ to have similar smoothness properties as $g_0(\cdot)$, the optimal bandwidth b_n is that of the non-parametric regression of the $\{R_i\}_{i=1,\ldots,n}$ on $\{Z_{in}\}_{i=1,\cdots,n}$, chosen by applying any non-parametric regression bandwidth selection method. For this last regression bandwidth choice, we use the cross-validation method of *np* R Package.

5.3.2 Computation of $\hat{\theta}$

Estimation of θ is a purely parametric problem for a given $\hat{g}_\theta$, consisting to step 2 of the previous Algorithm 5.3. For that, we give the gradient and hessian associated with the maximization problem (5.15). The gradient is given by:

$$U(\theta \,|\, g_\theta(\cdot)) = \frac{\partial L_n}{\partial \theta}(\theta \,|\, \widehat{g}_\theta) + \frac{\partial L_n}{\partial g}(\theta \,|\, \widehat{g}_\theta)\frac{\partial \widehat{g}_\theta}{\partial \theta}, \qquad \forall \quad \theta \in \Theta. \tag{5.21}$$

The first term in (5.21) is composed of the following partial derivatives with respect to β and λ of the profile likelihood function $L_n(\cdot \,|\, \cdot)$:

$$\frac{\partial L_n}{\partial \beta}(\theta \,|\, \widehat{g}_\theta) = \sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-1} \Lambda\left(G_i(\theta, \widehat{g}_\theta(Z_i))\right) \left[Y_i - \Phi(G_i(\theta, \widehat{g}_\theta(Z_i)))\right] X_i, \tag{5.22}$$

and

$$\frac{\partial L_n}{\partial \lambda}(\theta \,|\, \widehat{g}_\theta) = -\sum_{i=1}^{n}[\sigma(s_i, \lambda)]^{-1} G_i(\theta, \widehat{g}_\theta(Z_i)) \Lambda\left(G_i(\theta, \widehat{g}_\theta(Z_i))\right) \times \left[Y_i - \Phi(G_i(\theta, \widehat{g}_\theta(Z_i)))\right] \frac{\partial \sigma}{\partial \lambda}(s_i, \lambda). \tag{5.23}$$

Concerning the second term in (5.21), the partial derivative with respect to $g(\cdot)$ of $L_n(\cdot \,|\, \cdot)$ is given by:

$$\frac{\partial L_n}{\partial g}(\theta \,|\, \widehat{g}_\theta) = \sum_{i=1}^{n}[\sigma(s_i, \lambda)]^{-1} \Lambda\left(G_i(\theta, \widehat{g}_\theta(Z_i))\right) \left[Y_i - \Phi(G_i(\theta, \widehat{g}_\theta(Z_i)))\right]. \tag{5.24}$$

Let us now consider the derivative with respect to θ of the estimator $\widehat{g}_\theta$. Because $\psi(\widehat{g}_\theta(z); \theta, z) = 0$, if one differentiate the latter with respect to β and λ, then:

$$\frac{\partial}{\partial \beta}\widehat{g}_\theta(z) = -\frac{\sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-2} \Delta_i(\theta, z) X_i K\left(\frac{z - Z_{in}}{b_n}\right)}{\sum_{i=1}^{n} [\sigma(s_i, \lambda)]^{-2} \Delta_i(\theta, z) K\left(\frac{z - Z_{in}}{b_n}\right)},$$

and

$$\frac{\partial}{\partial\lambda}\widehat{g}_\theta(z) = \frac{\sum_{i=1}^n [\sigma(s_i,\lambda)]^{-2}\,\Delta_i(\theta,z)\,(G_i(\theta,\widehat{g}_\theta(z)))\,K\left(\frac{z-Z_{in}}{b_n}\right)\frac{\partial\sigma}{\partial\lambda}(s_i,\lambda)}{\sum_{i=1}^n [\sigma(s_i,\lambda)]^{-2}\,\Delta_i(\theta,z)K\left(\frac{z-Z_{in}}{b_n}\right)} + \frac{\sum_{i=1}^n [\sigma(s_i,\lambda)]^{-2}\,\Lambda\left(G_i(\theta,\widehat{g}_\theta(z))\right)[Y_i-\Phi\left(G_i(\theta,\widehat{g}_\theta(z))\right)]\,K\left(\frac{z-Z_{in}}{b_n}\right)\frac{\partial\sigma}{\partial\lambda}(s_i,\lambda)}{\sum_{i=1}^n [\sigma(s_i,\lambda)]^{-2}\,\Delta_i(\theta,z)K\left(\frac{z-Z_{in}}{b_n}\right)},$$

with

$$\Delta_i(\theta,z) = \Lambda'\left(G_i(\theta,\widehat{g}_\theta(z))\right)[Y_i-\Phi\left(G_i(\theta,\widehat{g}_\theta(z))\right)] - \Lambda\left(G_i(\theta,\widehat{g}_\theta(z))\right)\Phi'\left(G_i(\theta,\widehat{g}_\theta(z))\right).$$

Let us now define the $(p+q)\times(p+q)$ hessian matrix associated with the maximization problem (5.15) by:

$$\begin{aligned}\Gamma(\theta_0\,|\,g_{\theta_0}(\cdot)) &= E_0\left(\frac{d}{d\theta}U(\theta\,|\,g_\theta(\cdot))\Big|_{\theta=\theta_0}\Bigg|\{(X_i,Z_{in}),\, i=1,\dots,n\}\right)\\ &= \begin{bmatrix}\Gamma_{[p,p]}(\theta_0\,|\,g_{\theta_0}(\cdot)) & \Gamma_{[p,q]}(\theta_0\,|\,g_{\theta_0}(\cdot))\\ \Gamma^T_{[p,q]}(\theta_0\,|\,g_{\theta_0}(\cdot)) & \Gamma_{[q,q]}(\theta_0\,|\,g_{\theta_0}(\cdot))\end{bmatrix},\end{aligned}$$

where the blocs matrices are given as follows. First the $p\times p$ matrix associated to differentiation w.r.t β is:

$$\begin{aligned}\Gamma_{[p,p]}(\theta_0\,|\,g_{\theta_0}(\cdot)) &= E_0\left(\frac{d^2}{d\beta d\beta^T}L_n(\theta\,|\,g_\theta(\cdot))\Big|_{\theta=\theta_0}\Bigg|\{(X_i,Z_{in}),\, i=1,\dots,n\}\right)\\ &= -\sum_{i=1}^n \xi_i^{(0)}\left\{X_i+\frac{\partial}{\partial\beta}g_\theta(Z_i)\Big|_{\theta=\theta_0}\right\}\left\{X_i+\frac{\partial}{\partial\beta}g_\theta(Z_i)\Big|_{\theta=\theta_0}\right\}^T,\end{aligned}$$

where

$$\xi_i^{(0)} = [\sigma(s_i,\lambda_0)]^{-2}\,\Lambda\left(G_i(\theta_0,g_0(Z_i))\right)\Phi'(G_i(\theta_0,g_0(Z_i))), \qquad i=1,\dots,n.$$

The $p\times q$ matrix associated to differentiation w.r.t β and λ is:

$$\begin{aligned}\Gamma_{[p,q]}(\theta_0\,|\,g_{\theta_0}(\cdot)) &= E_0\left(\frac{d^2}{d\beta d\lambda^T}L_n(\theta\,|\,g_\theta(\cdot))\Big|_{\theta=\theta_0}\Bigg|\{(X_i,Z_{in}),\, i=1,\dots,n\}\right)\\ &= -\sum_{i=1}^n \xi_i^{(0)}\left\{X_i+\frac{\partial}{\partial\beta}g_\theta(Z_i)\Big|_{\theta=\theta_0}\right\}^T\\ &\qquad\times\left\{\frac{\partial}{\partial\lambda}g_\theta(Z_i)\Big|_{\theta=\theta_0} - G_i(\theta_0,g_0(Z_i))\,\frac{\partial}{\partial\lambda}\sigma(s_i,\lambda)\Big|_{\lambda=\lambda_0}\right\}.\end{aligned}$$

The $q\times q$ matrix associated to differentiation w.r.t λ is:

$$\begin{aligned}\Gamma_{[q,q]}(\theta_0\,|\,g_{\theta_0}(\cdot)) &= E_0\left(\frac{d^2}{d\lambda d\lambda^T}L_n(\theta\,|\,g_\theta(\cdot))\Big|_{\theta=\theta_0}\Bigg|\{(X_i,Z_{in}),\, i=1,\dots,n\}\right)\\ &= -\sum_{i=1}^n \xi_i^{(0)}\left\{\frac{\partial}{\partial\lambda}g_\theta(Z_i)\Big|_{\theta=\theta_0} - G_i(\theta_0,g_0(Z_i))\,\frac{\partial}{\partial\lambda}\sigma(s_i,\lambda)\Big|_{\lambda=\lambda_0}\right\}^T\\ &\qquad\times\left\{\frac{\partial}{\partial\lambda}g_\theta(Z_i)\Big|_{\theta=\theta_0} - G_i(\theta_0,g_0(Z_i))\,\frac{\partial}{\partial\lambda}\sigma(s_i,\lambda)\Big|_{\lambda=\lambda_0}\right\}.\end{aligned}$$

This hessian matrix then permits to give the parametric component estimate.

5.4 Finite sample properties

In this section, we provide some simulation experiments to study the performance of the proposed model, which highlights the importance of considering spatial dependence and partial linearity. For that we use diverse schemes of parametrization to generate data using the model of interest (PLPSH). We provide comparisons with other models: (i) the fully linear Probit that accounts for some unknown heteroskedasticity (LPH), (ii) the partial linear Probit without heteroskedasticity (PLP) and (iii) the ordinary linear Probit (LP). *GPLM* R package is used to provide estimates for PLP model and for LPH model we use the *GLMX* R package. In case of the basic Probit (LP), we use *GLM* function of *STATS* R package.

The particularity of our model is that the dependency on the error terms counts on some locations $s_1, \ldots, s_n$, chosen here randomly in a 30×30 regular grid. We generate observations from the following spatial latent partial linear model:

$$Y_{in}^* = \beta_1 X_{in}^{(1)} + \beta_2 X_{in}^{(2)} + g_0(Z_{in}) + \sigma(s_i, \lambda_0)\varepsilon_{in},$$
$$Y_{in} = \mathbb{I}(Y_{in}^* > 0),\ i = 1, \ldots, n \tag{5.25}$$

where the (ε_{in}) are i.i.d standard Gaussian random variables. To describe the spatial relation between errors, we generate a set $B = 3$ of spots in the 30×30 regular grid and consider d_i as the minimum Euclidean distance between the location i and B spots, thereby $\sigma(s_i, \lambda_0) = \lambda_0 \times d_i$. To account a partial linearity effect, we consider the following three cases:

Case 1: The explanatory variables $X^{(1)}$ and $X^{(2)}$ are generated as pseudo $\mathcal{N}(0, 1)$ and $\mathcal{U}[-2, 2]$, respectively, and the 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 $g(t) = t + 2\cos(0.5\pi t)$.

Case 2: We consider the same variables as **Case 2** and replace the non-linear function by $g(t) = \cos(t)$.

Case 3: The explanatory variables $X^{(1)}$, $X^{(2)}$ and Z are generated as pseudo $\mathcal{N}(0, 1)$, and we consider the linear function $g(t) = 1 + 0.5t$.

For all cases we take $\beta_1 = -1$, $\beta_2 = 1$ and different values of the spatial parameter λ, that is, $\lambda \in \{0.2, 0.5, 0.9\}$. The bandwidth b_n is selected using Severini & Staniswalis (1994)'s approach detailed previously with $C_{ni} = \Phi^{-1}(0.9Y_{ni} + 0.1(1 - Y_{ni}))\,,\ i = 1, \ldots, n$. The Gaussian kernel is considered: $K(t) = (2\pi^{-1/2})\exp(-t^2/2)$. The Monte Carlo simulation is based on 200 replications where for each one we generate sample data with size 400 in the different cases.

Tables 5.1 to 5.3 reports the results; the columns named Mean, Median and SD report the average, the median, and the standards deviation, respectively of the replications. Overall we notice that in all cases, the estimates procedures based on the proposed method (PLPSH) of β_1, β_2 and λ outperform. However, the estimation of the heteroskedasticity parameter is not very stable regarding the standard deviations in all cases. Therefore, in all situations the proposed method gives performant non-parametric estimate presented with blue line in Figure 5.1, compared to the estimate using PLP method given by green line. The two models (PLP ignoring the spatial heteroskedasticity and PLPSH) estimates (parametric and non-parametric) are similar in case of weak heteroskedasticity ($\lambda = 0.2$). If λ increases the performance of PLP decreases which makes sense since this last ignore the heteroskedasticity produced by $\sigma(s_i, \lambda)$.

However regarding the method of LPH that ignores the non-linear part and accounts some unknown heteroskedasticity in the errors, the estimates of the parameters β_1 and β_1 are inconsistent in particular for **Case 1** with different spatial parameter λ because of its non linear component. For **Case 2** and **Case 3** the parametric estimators have less bias when $\lambda = 0.2$, small spatial heterogeneity compare to larger values of λ. When estimating the data generated process **Case 1** (with a non-linear function $g(\cdot)$) by a ordinary Probit modelization (PL), the parameters β_1 and β_2 are miss-estimated, compare to **Case 1** where the function $g(\cdot)$ is slightly linear and **Case 3** with a fully linear function $g(\cdot)$. As for LPH, the estimates are improved when considering a small spatial heterogeneity on the error terms.

Table 5.1: **Case 1**: $g(t) = t + 2\cos(0.5\pi \times t)$, $n = 400$ and 200 replications.

λ	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.2	PLPSH	-1.00	-0.93	0.32	0.99	0.97	0.30	0.15	0.14	0.59
	LPH	-0.53	-0.53	0.08	0.52	0.52	0.06	-	-	-
	PLP	-1.01	-0.99	0.16	1.00	0.98	0.14	-	-	-
	LP	-0.50	-0.50	0.07	0.49	0.49	0.06	-	-	-
0.5	PLPSH	-1.01	-0.93	0.36	0.99	0.94	0.33	0.44	0.43	0.61
	LPH	-0.48	-0.47	0.08	0.47	0.47	0.07	-	-	-
	PLP	-0.88	-0.88	0.16	0.87	0.86	0.14	-	-	-
	LP	-0.46	-0.46	0.08	0.45	0.44	0.07	-	-	-
0.9	PLPSH	-1.03	-0.97	0.37	1.02	0.93	0.36	0.84	0.83	0.69
	LPH	-0.44	-0.44	0.09	0.44	0.44	0.07	-	-	-
	PLP	-0.74	-0.73	0.18	0.73	0.73	0.16	-	-	-
	LP	-0.43	-0.43	0.08	0.42	0.42	0.07	-	-	-

Table 5.2: **Case 2**: $g(t) = \cos(t)$, $n = 400$ and 200 replications.

λ	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.2	PLPSH	-1.05	-1.01	0.33	1.04	1.01	0.30	0.20	0.21	0.61
	LPH	-0.87	-0.87	0.12	0.86	0.87	0.10	-	-	-
	PLP	-0.98	-0.98	0.13	0.98	0.97	0.12	-	-	-
	LP	-0.86	-0.85	0.12	0.85	0.84	0.09	-	-	-
0.5	PLPSH	-1.09	-1.02	0.41	1.10	1.01	0.42	0.55	0.54	0.61
	LPH	-0.76	-0.76	0.13	0.76	0.77	0.10	-	-	-
	PLP	-0.83	-0.83	0.14	0.84	0.83	0.12	-	-	-
	LP	-0.75	-0.75	0.12	0.75	0.76	0.10	-	-	-
0.9	PLPSH	-1.07	-1.01	0.39	1.08	1.00	0.44	0.91	0.88	0.65
	LPH	-0.64	-0.65	0.15	0.65	0.65	0.12	-	-	-
	PLP	-0.69	-0.70	0.15	0.70	0.69	0.14	-	-	-
	LP	-0.64	-0.64	0.14	0.64	0.64	0.11	-	-	-

5.5 Conclusion

In this chapter, we propose a semi-parametric model for a binary outcome considering **heteroskedasticity**. The proposed model can be considered as a framework for tools where the outcome might be a policy, a decision, a transition, or otherwise binary outcome with spatial heteroskedastic errors. For the model inference, we adopt the approach

Table 5.3: **Case 3**: $g(Z) = 1 + 0.5 \times t$, $n = 400$ and 200 replications.

λ	Methods	$\beta_1 = -1$			$\beta_2 = 1$			λ		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0.2	PLPSH	-1.08	-1.04	0.34	1.08	1.01	0.36	0.25	0.21	0.61
	LPH	-0.86	-0.86	0.12	0.86	0.85	0.12	-	-	-
	PLP	-0.98	-0.98	0.15	0.98	0.97	0.15	-	-	-
	LP	-0.84	-0.84	0.10	0.84	0.83	0.11	-	-	-
0.5	PLPSH	-1.08	-1.04	0.28	1.08	1.03	0.32	0.6	0.58	0.63
	LPH	-0.76	-0.76	0.11	0.76	0.75	0.13	-	-	-
	PLP	-0.85	-0.85	0.14	0.85	0.84	0.14	-	-	-
	LP	-0.75	-0.75	0.11	0.75	0.75	0.12	-	-	-
0.9	PLPSH	-1.06	-1.01	0.38	1.06	1.04	0.39	0.98	0.97	0.66
	LPH	-0.64	-0.64	0.14	0.64	0.64	0.14	-	-	-
	PLP	-0.69	-0.68	0.16	0.69	0.69	0.15	-	-	-
	LP	-0.63	-0.63	0.13	0.63	0.62	0.13	-	-	-

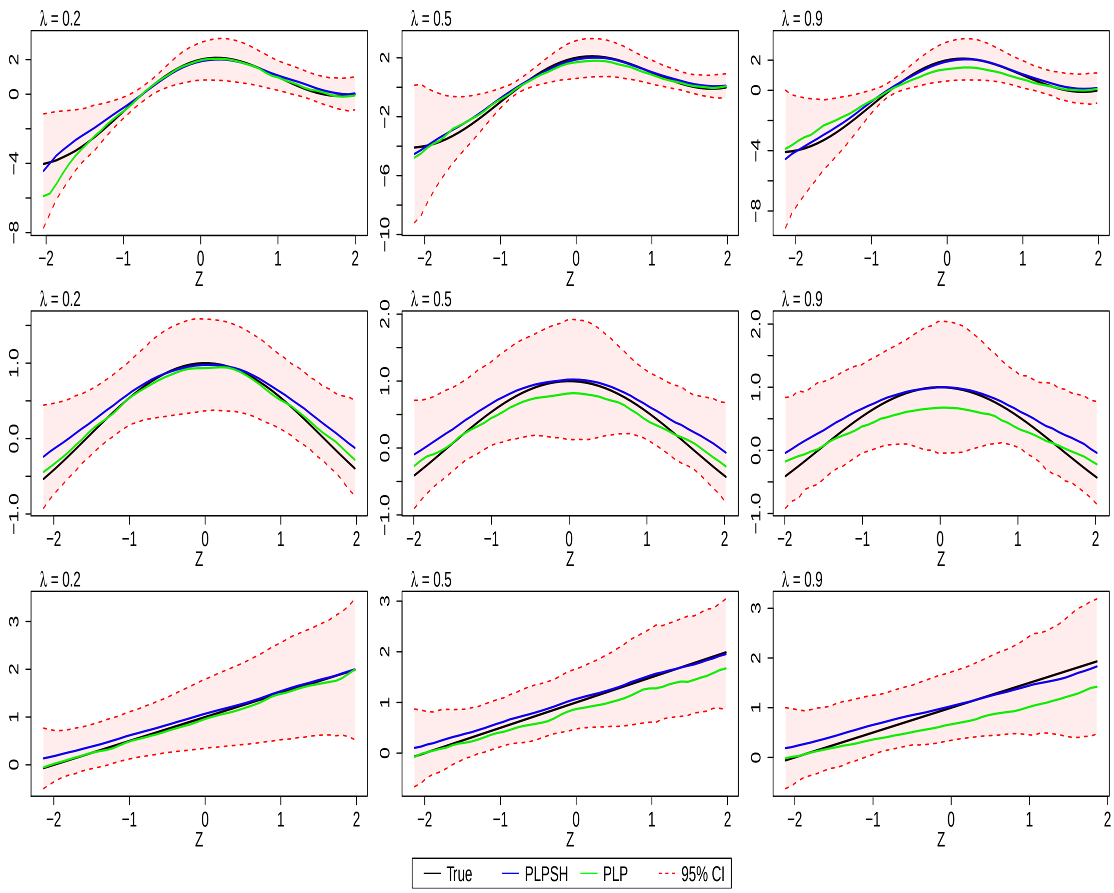

Figure 5.1: Top is the non-parametric estimate in **Case 1**, the middle is for **Case 2** while the bottom goes to **Case 3**.

of Severini & Staniswalis (1994) based on the concept of generalized profile likelihood. One of the advantage of this technique is that estimation procedure is divided into two problems; parametric and nonparametric parts. We outline some regularity conditions presented by Severini & Wong (1992) and some baseline about consistency and asymptomatic normality of the estimators. Finite sample study compares the performance of the proposed estimate to some existing techniques as the basic linear Probit, partially linear probit and the linear probit with unknown heteroskedasticity. Results show that for weak spatial heteroskedasticity, the proposed methodology and the partially linear method give good results for parametric component estimates while for large spatial heteroskedasticity the proposed methodology outperformed. This is not surprising since the non-spatial partially linear model does not account the spatial dependence incorporated in the data generated process compare to the proposed method. However the proposed heteroskedasticity parameter estimate suffers from large standard deviations, probably caused by the non-parametric estimation step that should be improved in a future work. Applications to many real data situations may be investigated in a number of fields like economics, political science, biostatistic, epidemiology, ...

Chapter 6

Spatial econometric models applied to model relation between environmental quality and economic growth

Contents

Résumé en français

Ce chapitre est de nature empirique et donne une application de certains modèles économétriques spatiaux vus dans les chapitres précédents pour modéliser la relation entre la qualité de l'environnement et le développement économique via la fameuse théorie de la courbe de Kuznet. Nous nous intéressons en particulier à des données d'emission de polluants et économiques sur les 290 municipalités de la Suède durant la période allant de 2005 à 2013. Ces données ont été traitées dans Marbuah & Amuakwa-Mensah (2017) qui ont utilisé une approche basée sur des modèles de panel SAR, SEM, Durbin, ...

L'approche utilisée ici est similaire à celle de ces auteurs, la différence réside au fait que nous avons utilisé en plus de la méthodologie de Marbuah & Amuakwa-Mensah (2017) une réduction de la dimension des co-variablse du modèle pour faire face à la forte colinéarité de ces variables. L'approche utilisé n'est pas fonctionnelle car les séries temporelles des 290 municipalités sont de courte période à savoir 2005-2013. Les résultats obtenus sont également comparés avec ceux trouvés par Marbuah & Amuakwa-Mensah (2017). Cependant, le travail de ce chapitre est largement motivé par l'approche de réduction de

dimension utilisée dans les Chapitres 3 et 4. Une perspective est d'étendre les approches fonctionnelles développées dans cette thèse avec des données du type utilisées dans ce chapitre sur une plus longue période.

6.1 Introduction

During the past decades, the remarkable economic development and the life quality improvement have not been without a dark side. In fact environmental degradation remains intimately connected to economic growth. Many authors tried to give theoretical and empirical evidence to the relation between economic growth and environmental degradation, and one of the most dominant approach is the Environmental Kuznets Curve Hypothesis (EKC). Seminal papers that studied on this relation were Grossman et al. (1991, 1993), Grossman & Krueger (1995) and Torras & Boyce (1998) among others, even thought the primary concept is from Kuznets (1955). The EKC's hypothesis describes the relationship between environmental quality (air quality, emissions, ...) and the income as an inverted U-shape function, that is at lower level of income, environmental degradation increases but once a threshold point of a high level of income is reached, improvement of the environmental quality arises. In fact, good economic growth and technological progress may offer great resources to finance investment in new technologies for a friendly environment.

As much as the EKC hypothesis was supported for many years, it received serious criticisms in theoretical or econometrical point of view (see Stern (2004), Kijima et al. (2010), Kaika & Zervas (2013), ...). The EKC approach is statistically not robust, and the U-shaped relation cannot hold for all pollutants emission. This weakness was subsequently confirmed by many authors. In addition, the approach is generally limited to single country studies, while some studies founded that a number of pollutants peaked at an income level closed to the world mean per capita income. This last implies that for a given country having income more than the world mean per capita income will improve the environmental quality, or the distribution of the income is very skewed to the left which means that this improvement does not have a major affect.

In EKC theory, usually the conducted studies used only historical times series and focus only on dependence in time but over-look the fact that pollutants emissions are spatially correlated. Omitting spatial interaction effects in emission in a country of world level may lead to inconsistent and insufficient estimates of the model of interest. Many authors pointed out that considering spatial effects in evaluating the impact of economic growth on environmental quality is strongly recommended (see among others Bockstael (1996), Goodchild et al. (2000) Giacomini & Granger (2004), Halkos & Tzeremes (2011), ...).

The aim of this chapter is to asses the relationship between income and pollutant emissions across all Sweden municipalities within the framework of spatial panel data using principal component analysis to handle the high correlation between the economic and environmental covariates of interest. In fact, the methodology of this work is related to Marbuah & Amuakwa-Mensah (2017), but is slightly different from a methodological point of view, in particular we shed light on the problem of collinearity of covariates of the models by using a dimension reduction method based on principal component analysis.

In Section 6.2, we give a state of art on the relationship between economic activity and environment within the EKC theory in a spatial context. We then present in Section 6.3 the data and the methodology used in this contribution, whereas Section 6.4 reports the model estimation results and compares them with those obtained by using Marbuah & Amuakwa-Mensah (2017)'s methodology. A conclusion and some perspectives are given in Section 6.5.

6.2 Economic growth and environmental quality and spatial econometrics

In this section, we give a brief survey of literature on the link between economic outcomes and environment using the hypothesis of EKC in a context of spatial data analysis. In a context of EKC hypothesis, many studies were in continuation with Grossman et al. (1991, 1993) when estimating the income level that promotes air quality, usually called the break point or the returning point. Grossman & Krueger (1995) founded a returning point varies between 4,772\$ and 5,965\$ level income in 1990. Moreover List & Gallet (1999) observed that reduction of the pollutants sulfur dioxide (SO_2), nitrogen oxides (NO_x) starts at an income level close to 9,000\$ and 21,000\$, respectively in 1987 while Dinda (2004) estimated that the turning point occur between 3,000\$ and 10,000\$ level income per capita. Panayotou (2003) reported in a study that several air pollutants such as sulfur dioxide (SO_2), nitrogen oxides (NO_x) and particulate matter (PM) have a turning point income ranging between 3,000\$ and 5,000\$ per capita. One can refer to more recent studies as Everett et al. (2010), Sulemana et al. (2017), among others.

Various researches doubted about the validity of the EKC hypothesis and the existence of the U-shaped relationship between environment quality and economic growth. Shafik (1994) showed that the curve of urban waste and carbon emission doesn't decline at a high income level but rises monotonically with income per capita. This was confirmed by Bradford et al. (2005), who claimed that EKC holds only for six pollutants from fourteen extracted from data used earlier by Grossman & Krueger (1995). Furthermore, Vincent (1997) tested the environmental Kuznets curve within a developing country, namely Malaysia and they founded that any of the six pollutants emission studied has decreased at a certain high income level. More recently Stern (2017), critically reviewed the EKC hypothesis and discussed a large game of innovative alternative approach.

We remark that studies which validated EKC hypothesis concerned mostly developed countries, that may lead to investigate strategies which recompense somehow the degradation of the environment quality. Cole (2004) claimed that in developing countries, high pollution may be caused by exportation in direction to countries with higher income. This can be a consequence of the EKC shape in developed countries since they export their pollution to other countries. Within the same spirit, Stern (2002) and Dinda (2004) argued that there is a leakage of dirty production from developed to developing countries.

All the works discussed above did not take into account some potential spatial interactions on pollution and economic growth even if pollutants emissions are correlated in space not only in time (Elsom, 1978; Deng et al., 2017; Wang et al., 2018). In fact, it is well know that if there are spatial interactions, not taking them into account may lead to biased and inconsistent models, as pointed out by Keene & Deller (2015), Rupasingha et al. (2004), Maddison (2006), Burnett et al. (2013a,b), Aklin (2016), Halkos & Tzeremes (2011),... The spatial econometrics modeling is then an alternative to basic EKC model when potential spatial correlations exist, see the seminal papers of Maddison (2006), Burnett et al. (2013a).

The basic technique of EKC model is to involve in the regression model with per capita pollutant emissions as response and income per capita, its squared value and a time trend as covariates. A more complex EKC model adds other covariates such as population density, social capital, income inequality. Collinearity that may occur between predictor are always untold even it is crucial from the econometric point of view at least for three major reasons:

- the coefficients may seem insignificant, even when a significant relationship exists between the predictors and the response variable,
- the strongly correlated predictor coefficients vary considerably from one sample to another,
- when variables of a model are highly correlated, the removal of one of them will have a considerable impact on the estimated coefficients of the others. The coefficients of strongly correlated terms may even have the wrong sign.

In this chapter, we attempt to model pollution-income relationship with the most popular emission; carbon dioxide ($\mathrm{CO_2}$), sulfur dioxide ($\mathrm{SO_2}$), nitrogen oxides ($\mathrm{NO_x}$), carbon monoxide (CO), particulate matter ($\mathrm{PM_{2.5}}$ and $\mathrm{PM_{10}}$) and total suspended particulates (TSP) by taking into account the correlation between covariates.

6.3 Spatial econometric models to model relation between environmental quality and economic growth in Sweden municipalities

6.3.1 Database

Database used in this study (provided by Marbuah & Amuakwa-Mensah (2017)) is in the format of a balanced panel data for 290 municipalities (located by latitude and longitude) in Sweden between 2005 and 2013. It contains 6 pollutants emission variables (SO_2, NO_x, CO, $PM_{2.5}$, PM_{10} and TSP) and the greenhouse gas (CO_2) measured in tonnes provided by the Swedish national emissions database. In the following when using the term pollutant emission, we mean one of these pollutants including $CO2$. In addition we have population density expressed in square kilometer for each municipality and the mean income per capita earned by residents aged 20 years and above in each municipality. Density and income data are provided by Statistic Sweden website[1] and used to calculate the income per capita. Table 6.3.1 reports the descriptive statistics and the definition of the variables.

6.3.2 Empirical methods

Recall that the objective of this work is to handle the correlation between economic and social covariates when studying the spatial effect in the analysis of emissions. As a first step one may plot the map of different emission variables to see if some clusters across space are present or not. But rather than visualizing only clusters, we can also have a quantitative spatial clustering measure. For that we can apply the famous spatial autocorrelation test based on Moran's I index. It was introduced first by Moran (1948, 1950) and later on by Cliff & Ord (1973, 1981) who suggested a new comprehensive formula of Moran's I statistic which is bounded by 1 and -1 as the basis correlation coefficient of Pearson. A positive Moran's I statistic indicates a positive spatial autocorrelation, a negative value means a negative spatial autocorrelation and if it tends to zero, there is absence of spatial autocorrelation. An alternative test is based on the Geary's C statistics of Geary (1954), less robust than Moran's I test. The results of the both tests are always interpreted in

[1] https://www.scb.se

Table 6.1: Descriptive statistics

Variable	Description	Mean	Std.dev	Min	Max	N
CO_2pc	Carbon dioxide per capita (tonnes)	6.464	13.591	0.3807	236.184	2610
SO_2pc	Sulfur dioxide per capita (tonnes)	0.0049	0.0112	0.00005	0.1054	2610
NO_xpc	Nitrogen oxides per capita (tonnes)	0.0264	0.0251	0.0013	0.2671	2610
$COpc$	Carbon monoxide per capita (tonnes)	0.0890	0.0417	0.0083	0.3057	2610
$PM_{2.5}pc$	Particulate matter per capita (< 2.5 micrometers; tonnes)	0.0043	0.0050	0.0002	0.0783	2610
$PM_{10}pc$	Particulate matter per capita (< 100 micrometers; tonnes)	0.0062	0.0059	0.0006	0.0837	2610
$TSPpc$	Total suspended particulate matter per capita (< 100 micrometers; tonnes)	0.0069	0.0069	0.0011	0.0974	2610
$Incomepc$	Real per capita mean income earned in municipality by residents aged 20 years and older (Swedish Krona, SEK 2014 prices)	17597	13900	308.7	86908	2610
$Popdens$	Total population density per sq. km	135.0	464.7	0.200	4917	2610

the context of a null hypothesis which assumes in our case that the analyzed pollutant emission is distributed randomly among different municipalities.

The Moran or Geary statistics is a global index of spatial autocorrelation in the sense that it provides summary about the overall spatial relationship over all locations here municipalities. Another useful visualization tools to asses how similar is an observation to its neighbors is the Moran scatter plot, first outlined in Anselin (1996). In this plot, the horizontal axis is the pollutant emission and the vertical axis is based on the average pollutant emission of neighbors (or spatial lag). The slope of the least squared regression line that best fits the obtained points in the Moran scatter plot is nothing than the Moran's I index. The Moran scatter plot is centred on point $(0, 0)$ since it takes variables as deviations from their means and points are distributed into four quadrants that portray different spatial association between a location i (located by latitude and longitude) and its neighbors. The upper-right and the lower-left quadrant represent a positive spatial association and we refer to them as respectively, high-high HH clustering which means in our case study that a municipality with high emission is surrounded by other municipalities with high emission too and low-low LL clustering which mean that a municipality with low emission is surrounded by other municipalities with low emission too. In contrast the upper-left and the lower-right correspond to a negative spatial association and we refer to them as respectively, low-high LH clustering which means that a municipality with low emission is surrounded by municipalities with high emission and high-low HL clustering meaning that a municipality with high emission is surrounded by municipalities with low emission. Compared to Moran's I index which is a clue of the global spatial dependence present in the dataset, Moran scatter plot gives more precision about spatial interactions between locations (here municipalities) by the classification of the spatial autocorrelation into four types (HH, LL, HL and LH). This feature makes the transition between *global* and *local* spatial autocorrelation. So we can consider that a local form of the global index indicates where clusters are located. This is know as local indicator of spatial association (LISA) or local Moran's I. This local index provides a local measure of similarity between a location and its neighbors and also permits to detect hot (cold) spots using Getis-Ord G_i*.

The Moran spatial dependency (auto-correlation) method and other dependence techniques rely on the concept of neighboring as pointed out in the first chapters. As usual, in our case study the spatial correlation of each municipality to its neighborhood is modeled by a spatial weight matrix constructed usually by geographic metric. It generally let W_n this weight matrix, it may be obtained in different ways as discussed in Chapter 2. The basic weight matrices are those based on contiguity, the inverse distances between spatial units or the one based on the $k-$nearest neighborhood algorithm (KNN). In this study, we ignore the contiguity type since Sweden has an island and for the KNN method we consider that each municipality has at least $k = 5$ neighbors (see Marbuah & Amuakwa-Mensah (2017) where several number of nearest neighborhood have been used by).

Following the conventional approach of spatial analysis, firstly we start by the classical ordinary least squares (OLS) and then testing for eventual spatial dependence among the error terms or the dependent variable. Specification of the classical OLS model is:

$$Y_{it} = \alpha + X_{it}\beta + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T \tag{6.1}$$

where i and t are the individual (municipality) and timed index respectively. Y_{it} is an $(n \times 1)$ dependent vector of variables and X_{it} is the $(n \times k)$ matrix of explanatory variables. The parameters of interest in (6.1) are the scalar α and the $(k \times 1)$ vector β. The EKC curve can be model via the model (6.1) by incorporating the squared of explanatory variables and other socio/economic variables. The resulting model in our case study is given by:

$$lnE = \alpha + \beta_1 \times lnIncpc + \beta_2 \times lnIncpc^2 + \beta_3 \times lnPopdens + \epsilon \tag{6.2}$$

where lnE is $n \times 1$ response vector of variables corresponding to the logarithm of the pollutant emission (CO_2, SO_2, NO_x, CO, $PM_{2.5}$, PM_{10} or TSP) with three $(n \times 1)$ vectors of independent variables; $lnIncpc$ is the log of income per capita, $lnIncpc^2$ is the square of the log income per capita and $lnPopdens$ is the log of population density. This model has been estimated in Marbuah & Amuakwa-Mensah (2017) without taking into account a potential collinearity between the independent variables. To avoid consequences in case of multicollinearity, we apply principal component analysis (PCA) on the three covariates (standardized) $lnIncpc$, $lnIncp^2$ and $lnPopdens$ and we create new factor variables formed by a linear combination of variables and their principal component coordinates, we retain only coordinates corresponds to high factor loading (correlation between variable and factors greater than 0.5). The results are two uncorrelated factors; $Factor_1$ highly correlated with $lnIncpc$ and $lnIncp^2$, and $Factor_2$ highly correlated with $lnPopdens$. Thus model (6.2) can be transformed into:

$$lnE = \alpha + \beta_1 \times Factor_1 + \beta_2 \times Factor_2 + \epsilon \tag{6.3}$$

Using the two spatial weight matrices discussed earlier, we apply the Lagrange multiplier (LM) of Anselin (1988) and robust-Lm (RLM) of Anselin (1996) tests on the residuals of the OLS model (6.3) to test the hypothesis absence of spatial spillover effects. If the tests fail to accept the null hypothesis of no spatial dependence, one of the 2 alternative hypotheses should be admitted:

1. Hypothesis 1a (Spatial lag hypothesis): value observed in a particular location is determined by the spatially average of values in neighbors locations.

2. Hypothesis 1b (Spatial interaction in the error hypothesis): error associated with any observation in a particular location is determined by the spatially average of the errors in neighbors locations.

Hypothesis 1a is covered by the spatial lag model or the spatial autoregressive model (SAR) given by:

$$lnE = \alpha + \lambda W \times lnE + \beta_1 \times Factor_1 + \beta_2 \times Factor_2 + \epsilon \tag{6.4}$$

where λ is the autoregressive parameter. Whereas Hypothesis 1b is covered by the spatial error model (SEM) given by:

$$lnE = \alpha + \beta_1 \times Factor_1 + \beta_2 \times Factor_2 + u, \quad u = \rho W \times u + \epsilon,\ \epsilon \sim N(0, \sigma^2) \tag{6.5}$$

where ρ is the autoregressive parameter. In the two case α, β_1 and β_2 are the scalar parameters and $W = W_n$ is the spatial weight matrix. Estimation of these two models (6.4) and (6.5) with OLS can lead to inconsistent estimates of the covariates parameters due to the spatial lagged variable. Alternative techniques are instrumental variable (IV), generalized method of moment (GMM), or the ML method which is more recommended in case where the distribution of the error is not specified. Maddison (2006) proposed that the SEM model can be approximated by SAR model with an additional spatial lag in the dependent variable. As a result, model (6.4) can be extended to the spatial Durbin (SDM) form which is written as:

$$\begin{aligned} lnE = \alpha + \lambda W \times lnE + \beta_1 \times Factor_1 + \beta_2 \times Factor_2 \\ + \theta_1 W \times Factor_1 + \theta_2 W \times Factor_2 + \epsilon \end{aligned} \tag{6.6}$$

where the scalars θ_1 and θ_2 are two more parameters to estimate. SAR, SEM and SDM are special cases of the general model of Manski (Manski, 1993) which model the three possible interaction effects at the same time (endogenous, exogenous and correlated effects). Technically, using the general model of Manski is the best strategy to test the different spatial dependence effects but due identification issue, LeSage & Pace (2009) suggested that the best option is SDM, that is excluding the spatially autocorrelated error term from the general model. Elhorst (2010) noted that one strength of SDM is that it produces unbiased coefficient and correct standard errors or t-values of the coefficient estimates wether the true data generation process is a spatial lag or spatial error model. Furthermore SDM requires no prior restriction on the magnitude of potential spatial spillover effects.

SDM can be simplified to the SAR or SEM model by canceling either the exogenous interaction effect and imposing $\theta = (\theta_1, \theta_2)' = \mathbf{0}$ or deleting the correlated effect and imposing $\theta = -\rho\beta$, where $\beta = (\beta_1, \beta_2)'$. If the LM test rejects the estimated OLS model (6.3) in favor to the SAR, or the SEM or both models, then the SDM should be estimated and Wald test may be performed to test $H_0: \ \theta = 0$ and $H_0: \ \theta = -\rho\beta$. If the test couldn't rejected, then SAR (Hypothesis 1a) or SEM (Hypothesis 1b) model fits better the data than SDM. To test the hypothesis whether spillover effects exist or not we may interpret directly the coefficient estimates but according to LeSage & Pace (2009) this may lead to erroneous conclusions. A partial derivative interpretation of the impact of the change of covariates on the dependent variable is more realistic for testing the spillover effects. Note that the SDM model (6.6) can be rewritten into:

$$lnE = (I_n - \lambda W)^{-1}(\iota_n \alpha + X\beta + WX\theta + \epsilon), \tag{6.7}$$

where $X = (Factor_1, Factor_2)$, $\beta = (\beta_1, \beta_2)'$, $\theta = (\theta_1, \theta_2)'$ and I_n is an identity matrix of order n. The partial derivatives of (6.7) with respect to the r'th explanatory variable is given by:

$$\frac{\partial lnE}{\partial x_r'} = (I_n - \lambda W)^{-1}(I_n \beta_r + W\theta_r) \tag{6.8}$$

The derivative (6.8) have three important properties that can be interpreted in our case study as follow: If an explanatory variable in municipality i changes, so will the emission in other municipalities j ($j \neq i$). The first effect (change on i) is called a *direct effect* while the second (changes on sites $j \neq i$) an *indirect effect.* the sum of these effects is denoted the *total effect.*

6.4 Results and discussions

Figure 6.1 is a mapping of the different emission pollutants in all municipalities. It is clearer that for all emission pollutant (CO_2, SO_2, NO_x, CO, $PM_{2.5}$, PM_{10} or TSP) there is spatial heterogeneity with high levels concentrated in municipalities located in the north part of the country. A global spatial dependency is given by both Moran's I and Geary's C indexes given in Table 6.2. The $p-$values between brackets show a significant positive spatial autocorrelation for all pollutants independently of the spatial weight matrix, namely $5-$KKN for the matrix based on 5 nearest neighbors and ID for the matrix constructed by the inverse of the euclidean distance between municipalities. This implies that municipalities with similar level of emission tend to be spatially clustered than being distributed randomly.

Table 6.2: Spatial autocorrelation tests, based on KNN (k=5) and inverse distance (ID)

	Moran's I stat.		Geary's C stat.	
Variables/Weight matrix	5-KNN	ID	5-KNN	ID
$lnCO_2pc$	0.2210(0000)	0.1976(0000)	0.7363(0000)	0.7811(0000)
$lnSO_2pc$	0.1790(0000)	0.1476(0000)	0.7859(0000)	0.8365(0000)
$lnNO_xpc$	0.3162(0000)	0.2563(0000)	0.6456(0000)	0.7243(0000)
$lnCOpc$	0.5964(0000)	0.5311(0000)	0.4292(0000)	0.4888(0000)
$lnPM_{2.5}pc$	0.4394(0000)	0.4192(0000)	0.5496(0000)	0.5808(0000)
$lnPM_{10}pc$	0.4424(0000)	0.3943(0000)	0.5488(0000)	0.6048(0000)
$lnTSPpc$	0.4568(0000)	0.4454(0000)	0.5096(0000)	0.5349(0000)

p-values are in parentheses.

These results are confirmed by a Moran scatter plot given in Figure 6.2. Examining this figure shows that the majority of the spatial association falls in the first and the third quadrants, this characterizes positive spatial autocorrelation. For additional analysis, we display cluster maps for emission pollutants averaged over $2005-2013$. Significant clusters associated to the four quadrant in the scatter plot are given in Figure 6.3 where the local Moran's I is plotted. Cluster maps in Figure 6.4 with different signification level, indicate patterns of spatial clustering for hot spots (HH clustering) and cold spots (LL clustering) for each pollutant. We notice also that a significant particular cluster exist for the CO and TSP pollutants in the north part of Sweden.

We now turn to the econometric analyses, starting by estimation the OLS model (6.3) for each pollutants. Results are reported in Table 6.3. Factor covariates are statistically significant for all pollutants except for the coefficient $Factor_1$ that regroups the income per capita and its square that are statistically insignificant for $PM_{2.5}$, $PM_{1.0}$ and TSP. This coefficient could be biased or inconsistent probably because of the presence of the spatial dependence then estimating a spatial lag model or error model would be an appropriate choice. In order to determine which specification is suitable, we consider the classic LM and robust-LM test on the least-squared residuals for the two spatial weight matrices. As shown in Table 6.3, both tests are statistically significant at a level of 1%, concluding that

Table 6.3: Pooled OLS regression and test for spatial dependency

Independent varibale	Dependent variables: $lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
Intercept	9.006***	1.490***	3.349***	4.800***	1.990***	2.303***	2.545***
	(0.103)	(0.103)	(0.106)	(0.045)	(0.009)	(0.081)	(0.086)
$Factor_1$	−0.001	−0.010***	0.002*	0.003***	0.000	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$Factor_2$	−0.252***	−0.153***	−0.238***	−0.358***	−0.394***	−0.300***	−0.382***
	(0.14)	(0.014)	(0.015)	(0.006)	(0.012)	(0.011)	(0.012)
Adjusted R^2	0.156	0.056	0.159	0.714	0.395	0.438	0.384
F-*Stat.*	241.6***	78.5***	249.1***	3255***	853.9***	1017***	814.5***
Spatial tests							
5-KNN							
Global Moran's I	0.057***	0.157***	0.214***	0.138***	0.117***	0.104***	0.178***
LM. no spatial error	23.924***	178.71***	331.01***	137.41***	99.537***	77.905***	229.070***
LM. no spatial lag	60.512***	190.04***	324.69***	98.639***	160.290***	125.28***	274.310***
RLM. no spatial error	91.100***	0.828	10.16***	45.295***	6.509*	2.819*	0.888
RLM. no spatial lag	127.69***	12.156***	3.839*	6.527*	67.261***	50.190***	46.134***
ID							
Global Moran's I	0.066***	0.119***	0.181***	0.115***	0.170***	0.134***	0.209***
LM. no spatial error	89.541***	293.28***	679.220***	274.660***	599.430***	369.65***	901.810***
LM. no spatial lag	129.700***	328.730***	528.050***	140.530***	474.120***	241.600***	720.840***
RLM. no spatial error	3.559*	8.7645***	152.760***	150.760***	153.330***	134.460***	222.850***
RLM. no spatial lag	43.714***	44.214***	$1.584^{0.2082}$	16.639***	28.019***	6.412***	41.889***

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

OLS model must be rejected in favor to both SAR and SEM models.

As SAR and SEM models are specials case of the general model of Manski (1993), it will be more reasonable to start with the most general model to test for spatial interaction. Following the strategy of LeSage & Pace (2009), we proceed to estimate the SDM first for panel data[2]. Estimation results are summarized on Table 6.6 for all pollutants. Note first that the spatial autoregressive parameter λ are highly statistically significant, confirming the presence of spatial dependence. The coefficients are around $0.227 - 0.503$ with 5-knn weight matrix and slightly higher with ID matrix $(0.253 - 0.704)$. After doing PCA, the two news variables $Factor_1$ and $Factor_2$ have been calculated using the PCA scores, see Table 6.4 for more details.

Table 6.4: PCA scores of variables and factor variables construction

	Dim1	Dim2	Dim3
$lnIncpc$	$\gamma_1 = \mathbf{0.9673}$	0.2507	−0.0380
$lnIncpc^2$	$\gamma_2 = \mathbf{0.9690}$	0.2441	0.0381914445
$lnPopdens$	−0.8020	$\gamma_3 = \mathbf{0.5973}$	0.0003

$$Factor_1 = \gamma_1 \times lnIncpc + \gamma_2 \times LnIncpc^2 + 0 \times lnPopdens.$$
$$Factor_2 = 0 \times lnIncpc + 0 \times LnIncpc^2 + \gamma_3 \times lnPopdens$$

The variable $Factor_1$ is a linear combination of $lnIncpc$ and its square while $Factor_2$ is linked to $lnPopdens$. In almost all spatial models estimated with these factor variables, the coefficient of the lag variable $Factor_2$ ($lag\ Factor_2$) is highly significant compare to the coefficients of $lnPopdens$ and $Lag\ lnPopdens$ in model (6.2) used in Marbuah & Amuakwa-Mensah (2017). The results are given in Table 6.7 using the proposed spatial

[2]in all models we account for municipality and year fixed effects

weights matrices. We notice also that coefficients of $lnIncomepc$, $lnIncomepc^2$ and their lagged variables ($lag\ lnIncomepc$ and $lag\ lnIncomepc^2$) reported in the same Table 6.7 are non-significant in many cases because perhaps of the high correlation between the variables (see Table 6.5). Based on the significant lagged variables of $Factor_1$ and $Factor_2$, we may consider that level of emission in a specific municipality affects the level of emission in neighboring municipalities.

To cover all possible spatial interactions, we perform a Wald test to examine weather the estimated nested SDM is reducible to SEM or SAR. Results are in Table 6.6 and indicate that SDM is significantly more appropriate than SEM in all almost cases, SDM is rejected in favor to SAR model in case of CO_2, CO with the two structures of the spatial weight matrix. Tables 6.8 and 6.9 report both SAR and SEM estimates for all cases. As expected the spatial autoregressive parameters λ and ρ remain highly significant for all pollutants and the SAR model coefficients are significant in the case of CO_2 and CO. The SDM model was rejected in favor of the SAR when using our methodology while when estimating the usual model (6.2) of Marbuah & Amuakwa-Mensah (2017), the coefficients of $lnIncomepc$, $lnIncomepc^2$, $Lag\ lnInocmepc$ and $Lag\ lnIncomepc^2$ are non-significant when the dependent variable is CO_2 or CO. In this case, SDM model was rejected in favor of the SAR based on the results of the Wald test in Table 6.7. When comparing SEM, SAR and SDM, the coefficient of $Factor_2$ is significant for all air emission pollutants compare to that of $Factor_1$ particularly in cases where SAR or SEM is rejected in favor of SDM.

In basis OLS specification, the effect of a change in one of the covariates on the response variable is equal to the coefficient estimate. This is not the case in spatial regression model when testing spatial spillovers effects (LeSage & Pace, 2009). The correct interpretation of the spatial autoregressive coefficient in a SDM is based on the partial derivatives or the impact perspective. The advantages of this technique is that the spillover effects is break down into direct and indirect effects (see LeSage & Pace (2009)). Table 6.12 displays the direct, indirect and total effects through the SDM (6.7) for all air emissions. Significant direct impact of the covariates are slightly different from the coefficient estimates because of the feedback effects that reflect the impact through surrounding municipalities. We note also that spillover effect of $Factor_2$ is negative in all cases meaning if it is significant, increasing in $Factor_2$ (population density) in a specific municipality will decrease the emission in surrounding municipality, vice versa. A deep look at Table 6.12 argues that spillover effects exist for all covariates, namely population density, income and square of per capita, at nearly 70%. These results confirm a large part of result founded by Marbuah & Amuakwa-Mensah (2017), where direct effect was at most cases participate in more than 50% of the total effect. Even though that income and population have a greater impact on emission in the specific municipality, consequences on other municipality cannot be neglected.

6.5 Conclusion

This chapter considers impact of economic growth on environment using spatial econometric tools to take into account the spatial dependence and spatial spillover effect induced by neighboring. For that, we used panel data on seven different air emission per capita in the 290 municipalities of Sweden over the period of $2005 - 2013$. In order to detect spatial regimes in emission, we undertake some preliminary investigations by deleting spatial correlation between covariates and doing exploratory spatial analyses; global Moran's I, Geary's C indexes and tests and local indicator of spatial association. The obtained results show clear evidence of spatial dependence and spatial spillover effect, so to empirically validate these results we re-examine the relationship between emission, income and population density using spatial models instead of the classical regression model in basic EKC methodology. We start by testing weather an OLS model is rejected on favor to the spatial lag model or spatial error model using the classic and robust LM test based on the residuals of the OLS model. Both tests accept spatial models (SAR and SEM). A more general model (SDM) was estimated first to consider the spillover effect with spatially lag covariates. The results in combination with a Wald test reinforce the choice of SDM model regarding the SEM model for the seven air emissions and the SAR model for two emissions (CO_2 and CO) independently of the used spatial weight matrix. Finally to interpret correctly the estimates we analyze the impact of the covariates on the dependent variable using the derivative form. The direct and indirect impact was highly significant indicating the presence of the spillover effect between municipalities. The obtained results are in line with that of Marbuah & Amuakwa-Mensah (2017) with some improvement in reducing the colinearity of the coefficients thank to the factors variables based on principal components loading that avoid the multicollinearity of the covariates.

6.6 Appendix

Table 6.5: Correlation matrix of $lnIncomepc$, $lnIncomepc^2$ and $lnPopdens$

	$lnIncomerpc$	$lnIncomepc^2$	$logPopdens$
$lnIncomepc$	1	0.997	−0.609
$logIncomepc^2$	0.997	1	−0.614
$logPopdens$	−0.609	−0.614	1

Table 6.6: Parameters estimation of SDM and Wald test

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
5-knn							
λ	0.486***	0.518***	0.503***	0.227***	0.404***	0.398***	0.403***
	(0.021)	(0.022)	(0.021)	(0.016)	(0.020)	(0.020)	(0.020)
$Factor_1$	−0.001	−0.008***	0.001	0.003***	0.000	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
$Factor_2$	−0.218***	−0.125***	−0.200***	−0.335***	−0.333***	−0.331	−0.331***
	(0.013)	(0.013)	(0.014)	(0.006)	(0.012)	(0.001)	(0.012)
$Lag\ Factor_1$	−0.002	−0.002	−0.004***	−0.004	−0.002**	−0.002	−0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$Lag\ Factor_2$	−0.031**	−0.061***	−0.070***	−0.013	−0.063***	−0.054***	−0.033**
	(0.017)	(0.017)	(0.018)	(0.008)	(0.015)	(0.014)	(0.015)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test							
SAR(H_0) vs SDM	3.284	15.6***	15.348***	3.5102	20.883***	19.221***	5.224*
SEM(H_0) vs SDM	515.51***	587.15***	592.27***	195.16***	417.56***	428.44***	395.92***
ID							
λ	0.655***	0.704***	0.667**	0.253***	0.543***	0.535***	0.553***
	(0.025)	(0.024)	(0.024)	(0.022)	(0.024)	(0.024)	(0.024)
$Factor_1$	−0.001*	−0.008***	0.001	0.003***	0.000	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
$Factor_2$	−0.224***	−0.125***	−0.205***	−0.342	−0.337***	−0.337***	−0.334***
	(0.133)	(0.013)	(0.014)	(0.006)	(0.012)	(0.011)	(0.012)
$Lag\ Factor_1$	0.005*	−0.006**	−0.008***	0.000	−0.008***	−0.007***	−0.008***
	(0.003)	(0.003)	(−0.008)	(0.001)	(0.002)	(0.002)	(0.002)
$Lag\ Factor_2$	−0.053**	−0.085***	−0.104***	−0.009	−0.107***	−0.098***	−0.082***
	(0.026)	(0.026)	(0.027)	(0.012)	(0.023)	(0.021)	(0.022)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test							
SAR(H_0) vs SDM	4.136	12.478***	15.432***	0.9883	24.593***	24.682***	13.296***
SEM(H_0) vs SDM	709.8***	875.46***	791.29***	129.53***	513.85***	533.73***	529.85***

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

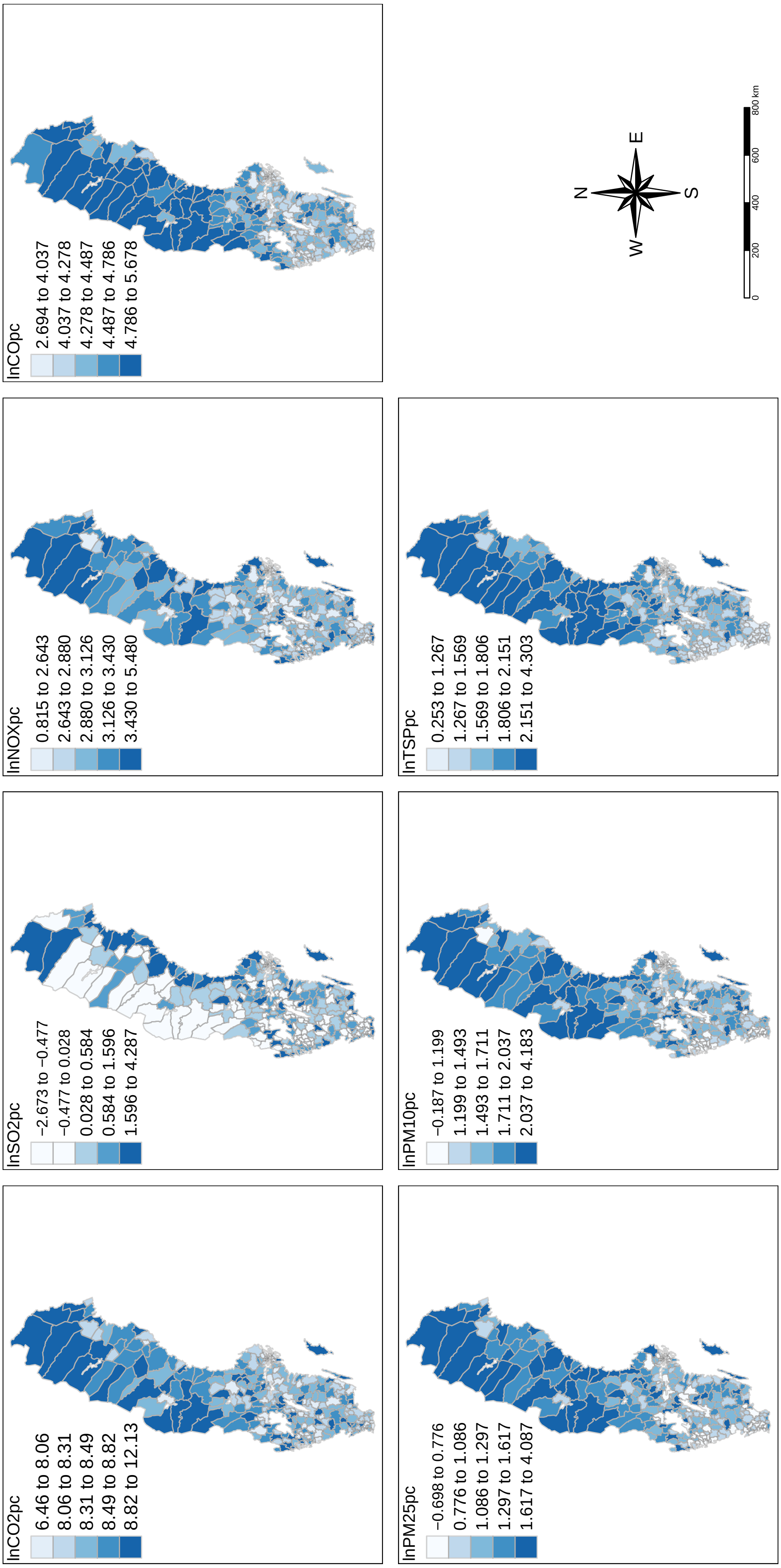

Figure 6.1: Pollutant emission degrees in the 290 Swedish municipalities

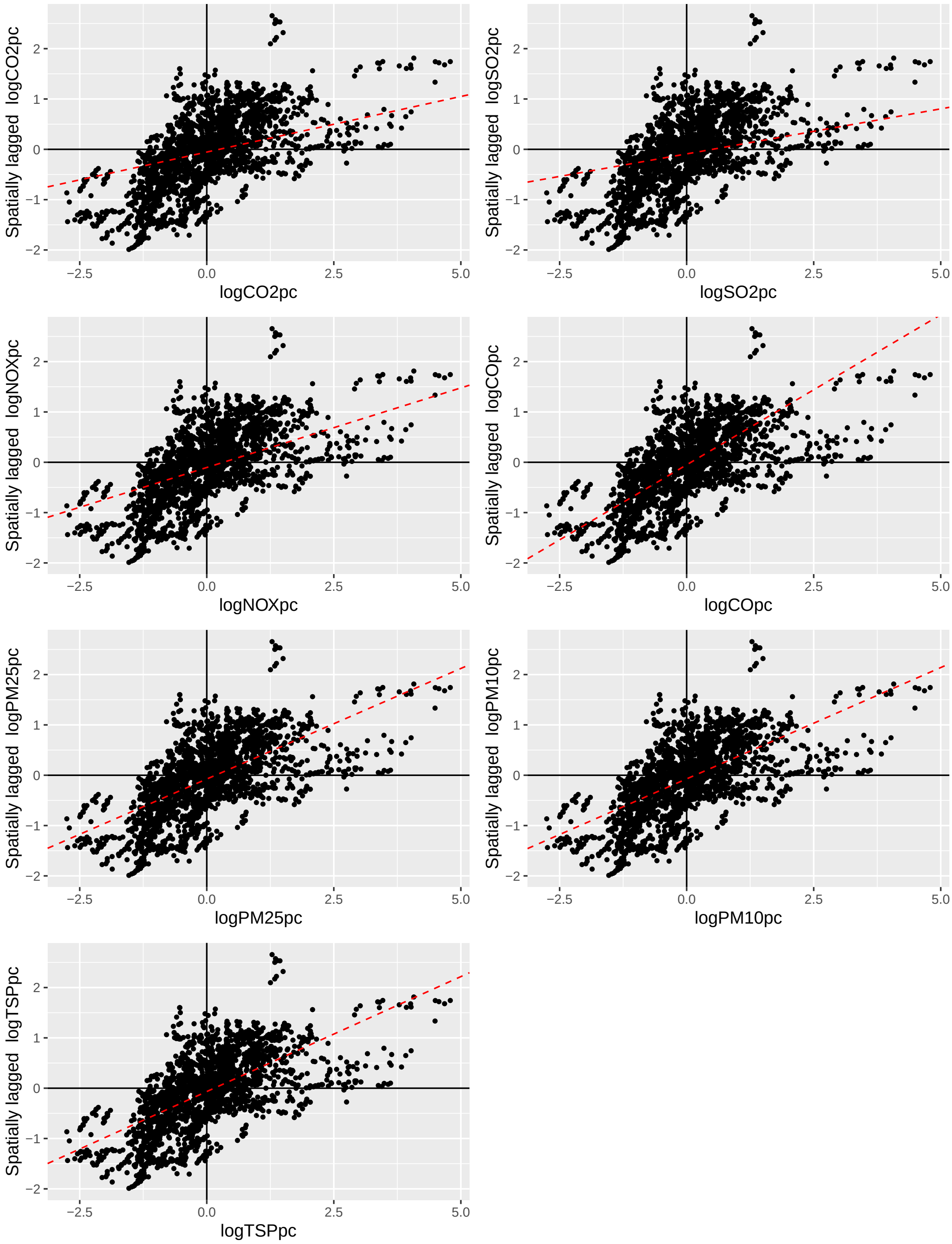

Figure 6.2: Moran's I scatter plot for air pollution emission

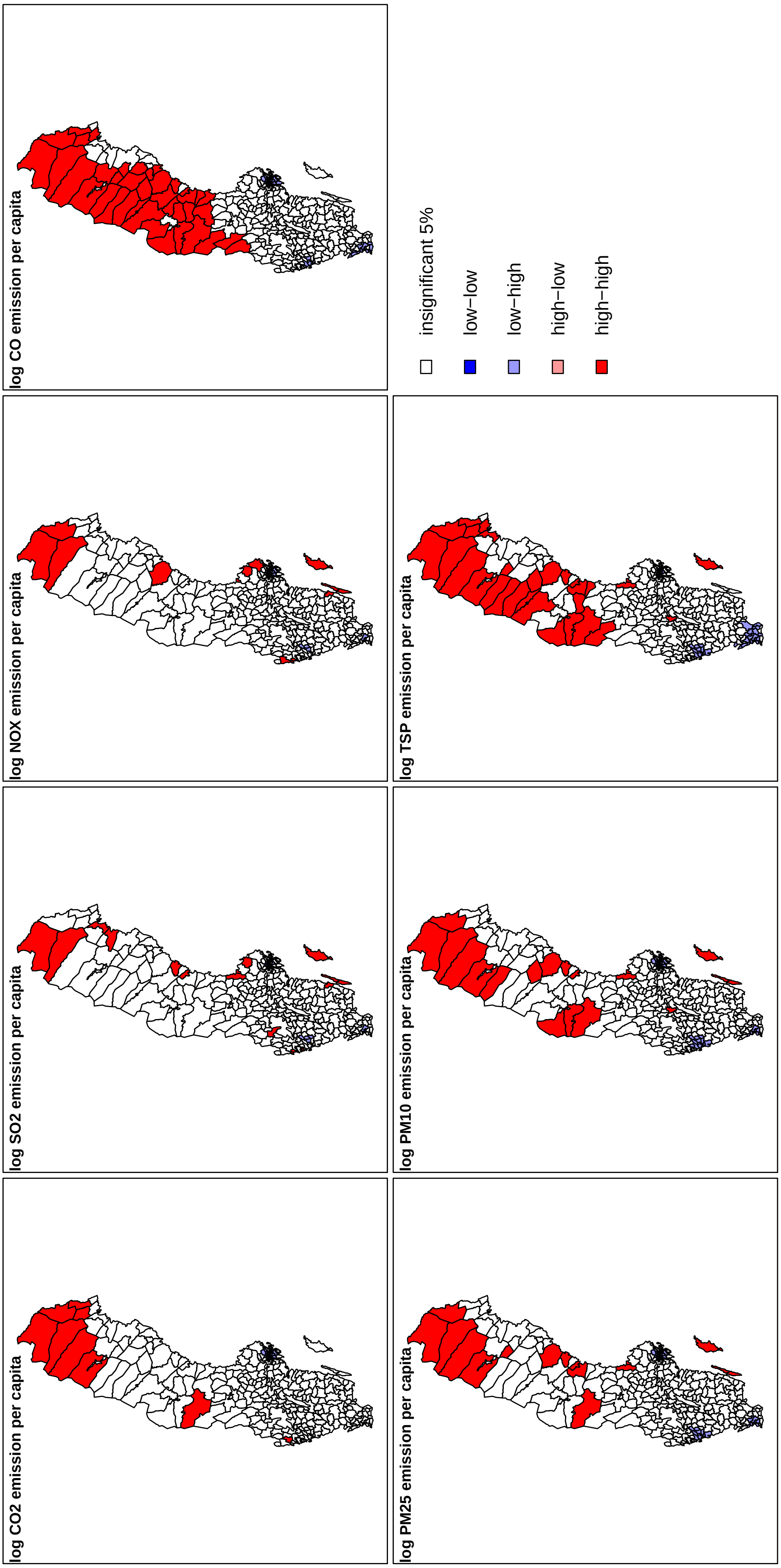

Figure 6.3: Spatial clustering, Anselin's Local Moran's I

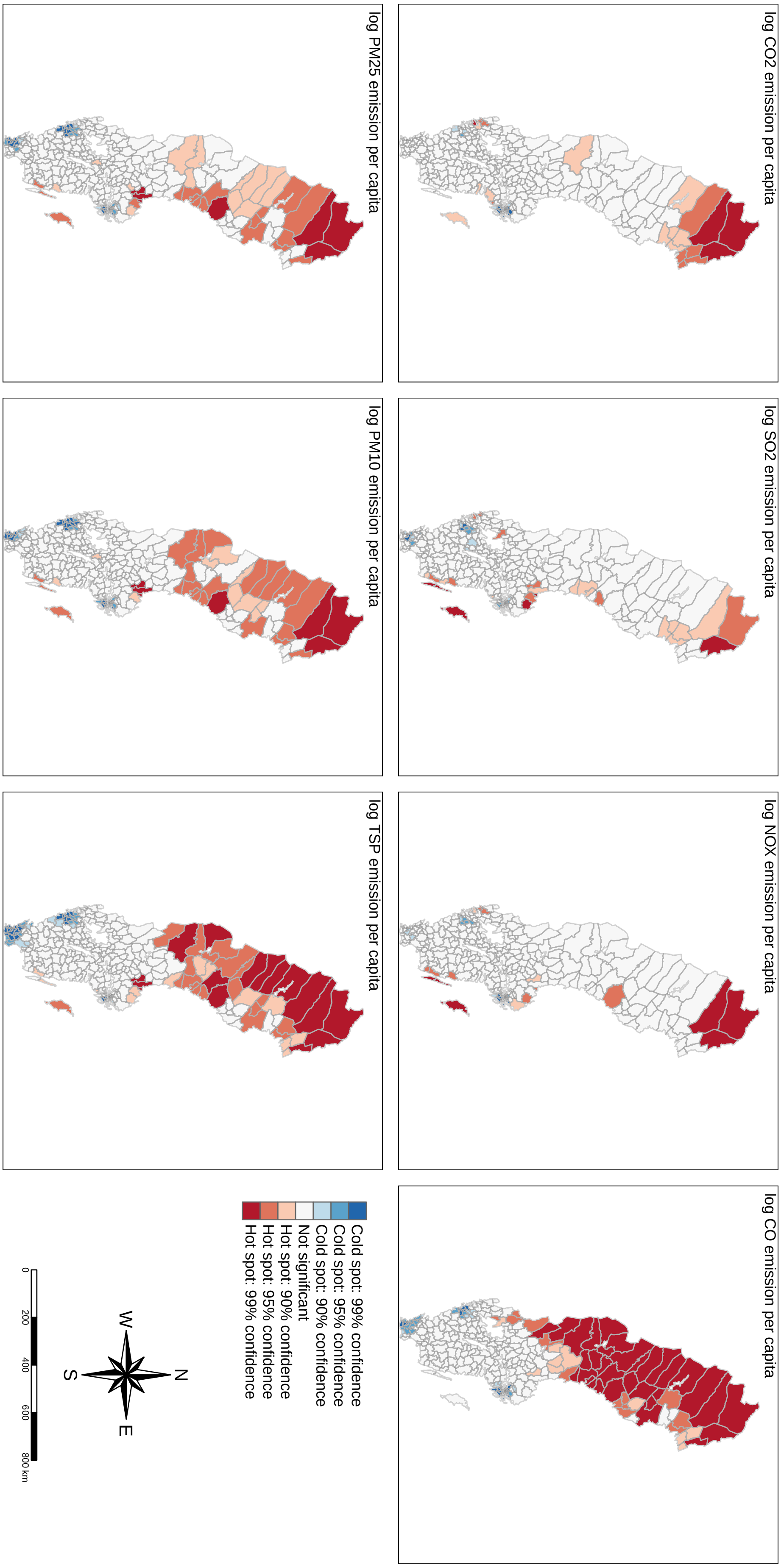

Figure 6.4: Hot Spot (Getis-Ord Gi*) Analyses

Table 6.7: Parameters estimation of SDM and Wald test with Marbuah & Amuakwa-Mensah (2017) methodology

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
5-knn							
λ	0.486***	0.515***	0.506***	0.228***	0.407***	0.402***	0.405***
	(0.021)	(0.022)	(0.021)	(0.016)	(0.020)	(0.020)	(0.020)
$lnIncomepc$	0.035	0.316**	0.813***	0.096	0.700***	0.625***	0.322***
	(0.144)	(0.144)	(0.008)	(0.067)	(0.126)	(0.116)	(0.123)
$lnIncomepc^2$	−0.003	−0.026***	−0.044***	−0.002	−0.038***	−0.034***	−0.019***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
$lnPopdens$	−0.131***	−0.078***	−0.124	−0.201***	−0.203***	−0.201***	−0.200***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.007)	(0.007)
$Lag\ lnIncomepc$	0.111	0.694**	0.403	0.055	0.403*	0.250	0.284
	(0.274)	(0.273)	(0.281)	(0.127)	(0.239)	(0.220)	(0.233)
$Lag\ lnIncomepc^2$	−0.008	−0.040***	−0.026*	−0.003	−0.024*	−0.015	−0.017
	(0.015)	(0.015)	(0.015)	(0.007)	(0.013)	(0.012)	(0.013)
$Lag\ lnPopdens$	−0.019*	−0.036***	−0.040***	−0.007	−0.035***	−0.030***	−0.019**
	(0.010)	(0.010)	(0.011)	(0.005)	(0.009)	(0.008)	(0.009)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test							
SAR(H_0) vs SDM	3.412	21.467***	15.793***	3.494	21.979***	18.866***	6.228
SEM(H_0) vs SDM	516.05***	587.23***	606.96***	196.49***	431.18***	441.12***	402.62***
ID							
λ	0.660***	0.700***	0.677***	0.256***	0.553***	0.546***	0.558***
	(0.024)	(0.024)	(0.149)	(0.022)	(0.024)	(0.023)	(0.024)
$lnIncomepc$	0.241*	0.343**	−0.053***	0.116	0.766***	0.699***	0.378***
	(0.145)	(0.114)	(0.149)	(0.069)	(0.127)	(0.117)	(0.124)
$lnIncomepc^2$	−0.015*	−0.028***	−0.053***	−0.003	−0.042***	−0.038***	−0.022***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
$lnPopdens$	−0.137***	−0.081***	−0.130***	−0.205***	−0.207***	−0.206***	−0.203***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.007)	(0.007)
$Lag\ lnIncomepc$	1.042**	2.257***	1.441***	−0.153	1.184***	0.804**	1.091**
	(0.507)	(0.503)	(0.520)	(0.239)	(0.444)	(0.409)	(0.432)
$Lag\ lnIncomepc^2$	−0.059**	−0.124***	−0.083***	0.008	−0.069***	−0.050**	−0.065***
	(0.027)	(0.026)	(0.027)	(0.013)	(0.023)	(0.022)	(0.023)
$Lag\ lnPopdens$	−0.026	−0.039**	−0.049***	−0.005	−0.053***	−0.050***	−0.042***
	(0.016)	(0.016)	(0.016)	(0.007)	(0.014)	(0.013)	(0.013)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald test							
SAR(H_0) vs SDM	7.944**	31.686***	19.866***	1.179	27.851***	24.585***	18.317***
SEM(H_0) vs SDM	735.04***	880.72***	859.04***	132.62***	545.67***	570.49***	349.01***

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

Table 6.8: Parameters estimation of SAR and SEM model with spatial weight matrix based on KNN ($k = 5$)

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
Model: SAR							
λ	0.486^{***}	0.517^{***}	0.505^{***}	0.228^{***}	0.408^{***}	0.403^{***}	0.405^{***}
	(0.021)	(0.024)	(0.021)	(0.016)	(0.020)	(0.020)	(0.020)
$Factor_1$	-0.002^{***}	-0.008^{***}	0.001	0.003^{***}	0.000	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
$Factor_2$	-0.220^{***}	-0.128^{***}	-0.202^{***}	-0.336^{***}	-0.335^{***}	-0.333^{***}	-0.332^{***}
	(0.013)	(0.013)	(0.0134)	(0.006)	(0.012)	(0.011)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model: SEM							
ρ	0.518^{***}	0.717^{***}	0.545^{***}	0.524^{***}	0.520^{***}	0.537^{***}	0.515^{***}
	(0.022)	(0.024)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
$Factor_1$	-0.002^{**}	-0.010^{***}	0.002^{**}	0.003^{***}	0.001	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
$Factor_2$	-0.255^{***}	-0.144^{***}	-0.230^{***}	-0.364^{***}	-0.377^{***}	-0.377^{***}	-0.375^{***}
	(0.014)	(0.014)	(0.015)	(0.006)	(0.012)	(0.011)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

Table 6.9: Parameters estimation of SAR and SEM model with spatial weight matrix based on inverse distance

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
Model: SAR							
λ	0.656^{***}	0.704^{***}	0.670^{***}	0.254^{***}	0.550^{***}	0.542^{***}	0.555^{***}
	(0.245)	(0.024)	(0.024)	(0.022)	(0.024)	(0.024)	(0.024)
$Factor_1$	-0.001^{*}	-0.008^{***}	0.001	0.003^{***}	0.000	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
$Factor_2$	-0.226^{***}	-0.128^{***}	-0.208^{***}	-0.343^{***}	-0.340^{***}	-0.339^{***}	-0.336^{***}
	(0.013)	(0.013)	(0.014)	(0.006)	(0.020)	(0.011)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model: SEM							
ρ	0.695^{***}	0.717^{***}	0.711^{***}	0.680^{***}	0.700^{***}	0.717^{***}	0.702^{***}
	(0.026)	(0.024)	(0.025)	(0.027)	(0.025)	(0.024)	(0.025)
$Factor_1$	-0.002^{**}	-0.010^{***}	0.002^{***}	0.003^{***}	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
$Factor_2$	-0.259^{***}	-0.144^{***}	-0.235^{***}	-0.367^{***}	-0.379^{***}	-0.381^{***}	-0.376^{***}
	(0.014)	(0.014)	(0.015)	(0.006)	(0.012)	(0.011)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

Table 6.10: Parameters estimation of SAR and SEM model with spatial weight matrix based on KNN ($k = 5$) (Marbuah & Amuakwa-Mensah (2017) Model)

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
Model: SAR							
λ	0.487***	0.517***	0.508***	0.229***	0.412***	0.406***	0.407***
	(0.021)	(0.022)	(0.021)	(0.016)	(0.020)	(0.020)	(0.020)
$lnIncomepc$	0.044	0.322**	0.828***	0.099	0.714***	0.639***	0.328***
	(0.144)	(0.144)	(0.148)	(0.067)	(0.126)	(0.116)	(0.123)
$lnIncomepc^2$	−0.004	−0.026***	−0.044***	−0.003	−0.039***	−0.034***	−0.019***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
$lnPopdens$	−0.131***	−0.078***	−0.125***	−0.201***	−0.203***	−0.202***	−0.199***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model: SEM							
ρ	0.521***	0.527***	0.551***	0.523***	0.524***	0.540***	0.517***
	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)
$lnIncomepc$	0.211	0.464***	1.041***	0.029	0.841***	0.723***	0.359***
	(0.163)	(0.164)	(0.167)	(0.071)	(0.139)	(0.127)	(0.135)
$lnIncomepc^2$	−0.014	−0.035***	−0.055***	0.001	−0.046***	−0.039***	−0.020***
	(0.009)	(0.009)	(0.009)	(0.004)	(0.008)	(0.007)	(0.008)
$lnPopdens$	−0.153***	−0.087***	−0.141***	−0.218***	−0.228	−0.228***	−0.225
	(0.009)	(0.009)	(0.009)	(0.004)	(0.007)	(0.007)	(0.007)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

Table 6.11: Parameters estimation of SAR and SEM model with spatial weight matrix based on inverse distance (Marbuah & Amuakwa-Mensah (2017) Model)

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
Model: SAR							
λ	0.663***	0.704***	0.682***	0.258***	0.561***	0.554***	0.561***
	(0.024)	(0.024)	(0.023)	(0.022)	(0.024)	(0.023)	(0.024)
$lnIncomepc$	0.243*	0.336***	0.986***	0.123*	0.787***	0.724***	0.387***
	(0.144)	(0.144)	(0.148)	(0.068)	(0.127)	(0.117)	(0.123)
$lnIncomepc^2$	−0.015*	−0.027***	−0.053***	−0.004	−0.043***	−0.039***	−0.022***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
$lnPopdens$	−0.136***	−0.078***	−0.128***	−0.205*	−0.206***	−0.206***	−0.202***
	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)	(0.007)	(0.007)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model: SEM							
ρ	0.702***	0.719***	0.721$***$	0.680***	0.707***	0.723***	0.705***
	(0.025)	(0.161)	(0.024)	(0.027)	(0.025)	(0.024)	(0.025)
$lnIncomepc$	0.389**	0.496***	1.181***	0.086	0.890***	0.808***	0.400***
	(0.161)	(0.161)	(0.165)	(0.071)	(0.138)	(0.126)	(0.134)
$lnIncomepc^2$	−0.023***	−0.037***	−0.063***	−0.002	−0.049***	−0.044***	−0.023***
	(0.009)	(0.009)	(0.009)	(0.004)	(0.008)	(0.007)	(0.007)
$lnPopdens$	−0.156***	−0.088***	−0.146***	−0.220***	−0.230***	−0.231***	−0.226***
	(0.009)	(0.009)	(0.009)	(0.004)	(0.007)	(0.007)	(0.007)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standards errors are given between parentheses. $*p < 0.1$, $**\ p < 0.05$ and $***p < 0.01$

Table 6.12: Impact analysis for SDM model.

Variable	Dependent variable						
	$lnCO_2pc$	$lnSO_2pc$	$lnNO_xpc$	$lnCOpc$	$lnPM_{2.5}pc$	$lnPM_{1.0}pc$	$lnTSPpc$
5-knn							
Direct							
$Factor_1$	−0.001	−0.009***	0.001	0.003***	0.000	0.001	−0.001
$Factor_2$	−0.230***	−0.132***	−0.211***	−0.339***	−0.344***	−0.342***	−0.341***
$Lag\ Factor_1$	−0.002	−0.002	−0.005***	−0.000*	−0.003*	−0.002	−0.002
$Lag\ Factor_2$	−0.033*	−0.064***	−0.074***	−0.013*	−0.065***	−0.056***	−0.035**
Indirect							
$Factor_1$	−0.001	−0.008***	0.001	0.001***	0.000	0.000	−0.000
$Factor_2$	−0.195***	−0.126***	−0.191***	−0.095***	−0.215***	−0.209***	−0.213***
$Lag\ Factor_1$	−0.002	−0.002	−0.004***	−0.000*	−0.002*	0.001	−0.001
$Lag\ Factor_2$	−0.028*	−0.061***	−0.067***	−0.004*	−0.040***	−0.034***	−0.022**
Total							
$Factor_1$	−0.002	−0.017***	0.002	0.004***	0.000	0.001	−0.001
$Factor_2$	−0.425***	−0.259***	−0.403***	−0.434***	−0.558***	−0.550***	−0.554***
$Lag\ Factor_1$	−0.004	−0.006	−0.009***	−0.001*	−0.004*	0.003	−0.003
$Lag\ Factor_2$	−0.061∗	−0.125***	−0.141***	−0.017*	−0.105***	−0.090***	−0.056**
ID							
Direct							
$Factor_1$	−0.001	−0.009***	0.001	0.003***	0.000	0.001	−0.001
$Factor_2$	−0.235***	−0.132***	−0.215***	−0.344***	−0.346***	−0.345***	−0.344***
$Lag\ Factor_1$	−0.005*	−0.006**	−0.009***	−0.000	−0.008***	−0.007***	−0.008***
$Lag\ Factor_2$	−0.056**	−0.090***	−0.109***	−0.009	−0.110***	−0.100***	−0.084***
Indirect							
$Factor_1$	−0.002	−0.019***	0.002	0.001***	0.000	0.001	−0.001
$Factor_2$	−0.416***	−0.289***	−0.400***	−0.114***	−0.392***	−0.378***	−0.404***
$Lag\ Factor_1$	−0.009*	−0.014**	−0.016***	−0.000	−0.009***	−0.008***	−0.009***
$Lag\ Factor_2$	−0.099**	−0.196***	−0.202***	−0.003	−0.125***	−0.110***	−0.099***
Total							
$Factor_1$	−0.004	−0.028***	0.003	0.004***	0.000	0.001	−0.002
$Factor_2$	−0.651***	−0.421***	−0.614***	−0.458***	−0.738***	−0.723***	−0.748***
$Lag\ Factor_1$	−0.014*	−0.020**	−0.025***	−0.001	−0.017***	−0.015***	−0.017***
$Lag\ Factor_2$	−0.154**	−0.286***	−0.311***	−0.013	−0.235***	−0.2102***	−0.183***

$* p < 0.1$, $** \ p < 0.05$ and $*** p < 0.01$

Chapter 7

General conclusion and perspectives

This thesis aims to contribute to spatial econometrics by proposing dedicated tools dealing with real-valued and functional spatial data in particular two types of models: the first is a spatial lag model with real-valued and functional covariate assuming both exogenous and endogenous spatial weight matrices and the second is a partially linear Probit model with spatial heteroskedasticity.

First, we propose a functional linear autoregressive spatial model able to describe and/or predict some spatially inter-connected real-valued response variable according to a functional covariate. We define the structure of the inter-dependence of the response variable in two different ways. In the first contribution, we assume the classical form by assuming an exogenous spatial weight matrix based on the notion of how geographic distance affects dependency while in the second contribution we propose a composite structure combining geographic distance and a linear process leading to an endogenous structure of the matrix. In both models, estimation procedure of the parameters of interest is done with a two step approach consisting of reducing the infinite dimension of the functional co-variate with a truncation technique detailed in Chapter 2 and applying a maximum likelihood algorithm. The consistency as well the asymptotic normality of the estimates are established in the first contribution in addition to a Monte Carlo study and real-data application that demonstrate the performance of the proposed models compared to the generalized functional linear model. For the second contribution, Monte Carlo study was conducted, the results highlight the performance of the proposed method compared to the first proposed model which miss-specifies the endogenous weight matrix used in the simulations.

The third contribution in Chapter 5, is a generalization of the heteroskedasticity Probit model proposed by Alvarez & Brehm (1995), in situations where assuming a constant variance error model is not realistic. Then we propose a general Probit model with a multiplicative function of the disturbance variances depending on spatial covariates. For more flexibility, a semi-parametric model is considered. The estimation methodology combines the weighted likelihood method to estimate the nonparametric component and a profile likelihood for the parametric component. Finite sample behavior of the estimates is given in addition to indications on how to extend the asymptotic results of Severini & Wong (1992) to the proposed context.

Chapter 6 treats the problem of environment degradation and economic growth. In particular, we attempt to explain the behavior of some pollutants emissions according to income per capita and population density in Sweden. This study is based on spatial econometric tools given in the previous chapters. Firstly, we provide arguments about the presence of spatial dependence on the pollution data by exploratory spatial data analysis (ESDA) and the local indicator spatial association (LISA), then we fit three spatial models to the data; (i) the general Durbin model, (ii) the spatial lag model (iii) and the spatial error model.

In the different contributions, some weak points may be the topics of future investigations. In a theoretical point of view, regarding the functional SAR model of Chapter 3, one may add some spatial lagged functional variable in order to explain some potential spillover effect often present in spatial data. We can also investigate other functional covariates including more complex features (e.g. correlated random functions). We could explore more applications to other fields such as continuous tracking, monitoring of movements, health data, continuously recorded climate data or financial space-time series data. Generalized spatial functional linear models may also be source of future investigations.

It is well known in spatial analyses that the spatial weight matrix is a key element and has been always imposed rather than estimated. Miss-specifying this matrix is a critical step and may produce inconsistent estimates. In some situations, practitioners have enough information to specify weighting matrices but this is not always possible. Then choosing an adequate matrix that reflects the spatial relationship in the concerned dataset is crucial. Alternative methods are the baseline of dynamic researches. Chapter 4 aims to contribute to this task. In fact, we relax the assumption of exogenous matrix by using an endogenous alternative. However, we use a parametric spatial matrix structure, this assumption may be relaxed by using a nonparametric weight matrix. Few authors (Beenstock & Felsenstein, 2012; Sun, 2016) treated this topic and proposed non parametric spatial weight matrix. Extending these works to the functional spatial lag model may be interesting. We have also to apply some of our contributions to real data, namely many potential applications may be considered within the FSAR model with endogenous spatial weight matrix and the spatial Probit model in Chapter 5.

The empirical study of the Chapter 5 shows that the estimations performs for the parametric and non parametric components but we have weak estimation precision of the spatial parameter. This requires further investigations to improve the estimates. Using local polynomial for the non-parametric estimate or adequate instrumental variables may be helpful.

List of Figures

List of Tables

Bibliography

Aguilera, A., Gutiérrez, R., & Valderrama, M. (1996). Approximation of estimators in the pca of a stochastic process using b-splines. *Communications in Statistics - Simulation and Computation*, *25*, 671–690. doi:10.1080/03610919608813336.

Aguilera, A. M., Gutiérrez, R., Ocaña, F. A., & Valderrama, M. J. (1995). Computational approaches to estimation in the principal component analysis of a stochastic process. *Applied Stochastic Models and Data Analysis*, *11*, 279–299. doi:10.1002/asm.3150110402.

Ahmed, M. S. (2017). *Contribution à la statistique spatiale et l'analyse de données fonctionnelles*. Ph.D. thesis. URL: http://www.theses.fr/2017LIL30047 thèse de doctorat dirigée par Dabo, Sophie et Attouch, Mohammed Kadi Mathématiques appliquées et applications des mathématiques Lille 3 2017.

Ahrens, A., & Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, *3*, 128–155. doi:10.3390/econometrics3010128.

Aklin, M. (2016). Re-exploring the trade and environment nexus through the diffusion of pollution. *Environmental & Resource Economics*, *64*, 663–682. doi:10.1007/s10640-015-9893-1.

Alvarez, R. M., & Brehm, J. (1995). American ambivalence towards abortion policy: Development of a heteroskedastic probit model of competing values. *American Journal of Political Science*, *39*, 1055–1082. URL: http://www.jstor.org/stable/2111669.

Amato, U., Antoniadis, A., & Feis, I. D. (2006). Dimension reduction in functional regression with applications. *Computational Statistics and Data Analysis*, *50*, 2422 – 2446. Statistical signal extraction and filtering.

Ancot, J. P. L. K., L, & Molle, W. (1982). opics in regional development modelling. In A. A. M. Albegov, & F. Snickars (Eds.), *Regional Development Modelling in Theory and Practice* (pp. 165–182). Amsterdam: North Holland: Oxford University Press.

Anselin, L. (1980). *Estimation Methods for Spatial Autoregressive Structures: A Study in Spatial Econometrics*. Cornell University Program in Urban and Regional Studies Regional Science Dissertation and Monograph Series. Program in Urban and Regional Studies, Cornell University.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models* volume 4. Springer Science & Business Media.

Anselin, L. (1990). Spatial dependence and spatial structural instability in applied regression analysis*. *Journal of Regional Science*, *30*, 185–207. doi:10.1111/j.1467-9787.1990.tb00092.x.

Anselin, L. (1996). The moran scatterplot as an esda tool to assess local instability in spatial association. In M. Fischer, H. J. Scholten, & D. Unwin (Eds.), *Spatial Analytical Perspectives on GIS*. London: Taylor & Francis.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, *27*, 247–267.

Anselin, L., & Bera, A. (1998). patial dependence in linear regression models with an introduction to spatial econometrics. In . Giles, David E. A., & A. Ullah (Eds.), *Handbook of applied economic statistics* chapter 7. (pp. 237–289). New York : Marcel Dekker.

Anselin, L., & Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, *23*, 112–131. doi:10.1111/j.1538-4632.1991.tb00228.x.

Arbia, G., & Baltagi, B. H. B. H. (2010). *Spatial econometrics : methods and applications*. Heidelberg : Physica-Verlag. Includes bibliographical references.

Baicker, K. (2005). The spillover effects of state spending. *Journal of Public Economics*, *89*, 529–544.

Baltagi, B. H., Song, S. H., & Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, *117*, 123 – 150. doi:https://doi.org/10.1016/S0304-4076(03)00120-9.

Beenstock, M., & Felsenstein, D. (2012). Nonparametric estimation of the spatial connectivity matrix using spatial panel data. *Geographical Analysis*, *44*, 386–397. doi:10.1111/j.1538-4632.2012.00851.x.

Berke, O. (2004). Exploratory disease mapping: kriging the spatial risk function from regional count data. *International Journal of Health Geographics*, *3*, 18. doi:10.1186/1476-072X-3-18.

Beron, K. J., & Vijverberg, W. P. M. (2004). Probit in a spatial context: A monte carlo analysis. In L. Anselin, R. J. G. M. Florax, & S. J. Rey (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications* (pp. 169–195). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-05617-2_8.

Bhat, C. R., & Sener, I. N. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems*, *11*, 243–272. doi:10.1007/s10109-009-0077-9.

Bockstael, N. E. (1996). Modeling economics and ecology: The importance of a spatial perspective. *American Journal of Agricultural Economics*, *78*, 1168–1180.

Bohorquez, M., Giraldo, R., & Mateu, J. (2016). Optimal sampling for spatial prediction of functional data. *Statistical Methods & Applications*, *25*, 39–54. doi:10.1007/s10260-015-0340-9.

Bohorquez, M., Giraldo, R., & Mateu, J. (2017). Multivariate functional random fields: prediction and optimal sampling. *Stochastic Environmental Research and Risk Assessment*, *31*, 53–70.

Bohorquez, R., Marthaand Giraldo, & Mateu, J. (2016). Optimal sampling for spatial prediction of functional data. *Statistical Methods & Applications*, *25*, 39–54. doi:10.1007/s10260-015-0340-9.

Bosq, D. (2000). *Linear processes in function spaces* volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. doi:10.1007/978-1-4612-1154-9.

Bradford, D., A, F. R., H., S. S., & Wagner, M. (2005). The environmental kuznets curve: Exploring a fresh specification. *Journal of Economic Analysis & Policy*, *4*, 1–30.

Brumback, B. A., & Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, *93*, 961–976. doi:10.1080/01621459.1998.10473755.

Burnett, J. W., Bergstrom, J. C., & Dorfman, J. H. (2013a). A spatial panel data approach to estimating u.s. state-level energy emissions. *Energy Economics*, *40*, 396–404. doi:10.1016/j.eneco.2013.07.0.

Burnett, J. W., Bergstrom, J. C., & Wetzstein, M. E. (2013b). Carbon dioxide emissions and economic growth in the u.s. *Journal of Policy Modeling*, *35*, 1014–1028. doi:10.1016/j.jpolmod.2013.08.

Cai, T. T., & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, *107*, 1201–1216. doi:10.1080/01621459.2012.716337.

Cardot, H., Faivre, R., & Goulard, M. (2003). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics*, *30*, 1185–1199.

Cardot, H., Ferraty, F., & Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*, *45*, 11 – 22.

Cardot, H., & Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, *92*, 24 – 41.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489. doi:10.1080/01621459.1997.10474001.

Case, A. (1991). Spatial patterns in household demand. *Econometrica*, *59*, 953–65.

Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics*, *22*, 491–508.

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

Case, A. C., Rosen, H. S., & Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, *52*, 285 – 307. URL: http://www.sciencedirect.com/science/article/pii/004727279390036S. doi:https://doi.org/10.1016/0047-2727(93)90036-S.

Castro, P. E., Lawton, W. H., & Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, *28*, 329–337.

Chakir, R., & Parent, O. (2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, *88*, 327–344. doi:10.1111/j.1435-5957.2009.00239.x.

Cliff, A., & Ord, J. (1981). *Spatial Processes: Models & Applications*. Pion.

Cliff, A. D. A. D., & Ord, .-j. a., J. K. (John Keith) (1973). *Spatial autocorrelation*. London : Pion. Includes index.

Cliff, A. D. A. D., & Ord, J. K. J. K. (1969). The problem of spatial autocorrelation. london: Pion. *Regional science*, (pp. 25 –55).

Cohen, J., & Morrison Paul, C. (2004). Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *The Review of Economics and Statistics*, *86*, 551–560.

Cole, M. (2004). Trade, the pollution haven hypothesis and the environmental kuznets curve: examining the linkages. *Ecological Economics*, *48*, 71–81.

Comte, F., & Johannes, J. (2012). Adaptive functional linear regression. *The Annals of Statistics*, *40*, 2765–2797. doi:10.1214/12-AOS1050.

Conley, T., & Ligon, E. (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth*, *7*, 157–87.

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

Cortes-D, D. L., Camacho-Tamayo, J. H., & Giraldo, R. (2016). Spatial prediction of soil penetration resistance using functional geostatistics. *Scientia Agricola*, *73*, 455 – 461.

Crabbé, K., & Vandenbussche, H. (2008). Spatial tax competition in the eu 15.

Crambes, C., Kneip, A., Sarda, P. et al. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, *37*, 35–72.

Cressie, N. A. (1993). Wiley.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, *147*, 1 – 23. doi:https://doi.org/10.1016/j.jspi.2013.04.002.

Dabo-Niang, S., Yao, A.-F., Pischedda, L., Cuny, P., & Gilbert, F. (2010). Spatial mode estimation for functional random fields with application to bioturbation problem. *Stochastic Environmental Research and Risk Assessment*, *24*, 487–497. doi:10.1007/s00477-009-0339-6.

Deng, Q., Yang, K., & Luo, Y. (2017). Spatiotemporal patterns of pm2.5 in the beijing–tianjin–hebei region during 2013–2016. *Geology, Ecology, and Landscapes*, *1*, 95–103. doi:10.1080/24749508.2017.1332851.

Dinda, S. (2004). Environmental kuznets curve hypothesis: A survey. *Ecological Economics*, *49*, 431 – 455. doi:https://doi.org/10.1016/j.ecolecon.2004.02.011.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, *5*, 9–28. doi:10.1080/17421770903541772.

Elsom, D. (1978). Spatial correlation analysis of air pollution data in an urban area. *Atmospheric Environment (1967)*, *12*, 1103 – 1107. doi:https://doi.org/10.1016/0004-6981(78)90356-6.

Ertur, C., & Koch, W. (2011). A contribution to the theory and empirics of schumpeterian growth with worldwide interactions. *Journal of Economic Growth*, *16*, 215–255.

Everett, T., Ishwaran, M., Ansaloni, G. P., & Rubin, A. (2010). *Economic growth and the environment*. MPRA Paper University Library of Munich, Germany.

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

Fan, J., & Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *62*, 303–322.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, *39*, 254–261.

Farber, S., Páez, A., & Volz, E. (2009). Topology and dependency tests in spatial and network autoregressive models. *Geographical Analysis*, *41*, 158–180. doi:10.1111/j.1538-4632.2009.00751.x.

Fedorov, V. (1989). Kriging and other estimators of spatial field characteristics (with special reference to environmental studies). *Atmospheric Environment (1967)*, *23*, 175 – 184. doi:https://doi.org/10.1016/0004-6981(89)90110-8.

Ferraty, F., & Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, *330*, 139 – 142. doi:https://doi.org/10.1016/S0764-4442(00)00140-3.

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

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In L. Anselin, R. J. G. M. Florax, & S. J. Rey (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications* (pp. 145–168). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-05617-2_7.

Frank, I. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*, 109–135.

Gao, J., Lu, Z., & Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *The Annals of Statistics*, *34*, 1395–1435. doi:10.1214/009053606000000317.

Garthoff, R., & Otto, P. (2017). Control charts for multivariate spatial autoregressive models. *A Journal of the German Statistical Society*, *101*, 67–94. doi:10.1007/s10182-016-0276-x.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*, 115–146.

Giacomini, R., & Granger, C. (2004). Aggregation of space-time processes. *Journal of Econometrics*, *118*, 7–26.

Giraldo, R. (2014). Cokriging based on curves, prediction and estimation of the prediction variance. *InterStat*, *2*, 1–30.

Giraldo, R., Delicado, P., & Mateu, J. (2010a). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*, 66–82. doi:10.1007/s13253-009-0012-z.

Giraldo, R., Delicado, P., & Mateu, J. (2010b). Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, *15*, 66–82. doi:10.1007/s13253-009-0012-z.

Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Journal of Agricultural, Biological, and Environmental Statistics*, *18*, 411–426. doi:10.1007/s10651-010-0143-y.

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review*, *23*, 139–159. doi:10.1177/016001760002300201.

Green, P. J., & Yandell, B. S. (1985). Semi-parametric generalized linear models. In R. Gilchrist, B. Francis, & J. Whittaker (Eds.), *Generalized Linear Models* (pp. 44–55). New York, NY: Springer US.

Grenander, U. (1950). Stochastic processes and statistical inference. *Arkiv för Matematik*, *1*, 195–277. doi:10.1007/BF02590638.

Grossman, G., Krueger, A., & of Economic Research, N. B. (1991). *Environmental Impacts of a North American Free Trade Agreement*. Number 3914 in Working paper series (National Bureau of Economic Research). National Bureau of Economic Research.

Grossman, G. M., & Krueger, A. (1995). Economic growth and the environment. *The Quarterly Journal of Economics*, *110*, 353–377.

Grossman, G. M., Krueger, A. B., Tang, K., Research, G. F. F., Brown, D., Evans, G., Sharing, G. S. F., Grossman, G. H., Krueger, A. B., School, W. W., & School, W. W. (1993). Environmental impacts of a north american free trade agreement," in mexican-u.s. free trade agreement, edited by.

Halkos, G., & Tzeremes, N. (2011). *Regional environmental efficiency and economic growth: NUTS2 evidence from Germany, France and the UK*. MPRA Paper 33698 University Library of Munich, Germany.

Han, X., & Lee, L. (2016). Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. *Journal of Business & Economic Statistics*, *34*, 642–660. doi:10.1080/07350015.2016.1167058.

Harvey, A. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*, 461–65.

Hastie, T., Buja, A., & Tibshirani, R. (1995). Penalized discriminant analysis. *The Annals of Statistics*, *23*, 73–102.

Hastie, T., & Mallows, C. (1993). A statistical view of some chemometrics regression tools: Discussion. *Technometrics*, *35*, 140–143.

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

Ho, C.-Y., Wang, W., & Yu, J. (2013). Growth spillover through trade: A spatial dynamic panel data approach. *Economics Letters*, *120*, 450–453.

Holloway, G., & Lapar, M. L. A. (2007). How big is your neighbourhood? spatial implications of market participation among filipino smallholders. *Journal of Agricultural Economics*, *58*, 37–60. doi:10.1111/j.1477-9552.2007.00077.x.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York. doi:10.1007/978-1-4614-3655-3.

Hoshino, T. (2018). Semiparametric spatial autoregressive models with endogenous regressors: With an application to crime data. *Journal of Business & Economic Statistics*, *36*, 160–172. doi:10.1080/07350015.2016.1146145.

Hsing, T., & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley.

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

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *64*, 411–432.

James, G. M., & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *63*, 533–550.

Jenish, N., & Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, *150*, 86 – 98.

Jenish, N., & Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, *170*, 178 – 190.

Kaika, D., & Zervas, E. (2013). The environmental kuznets curve (ekc) theory—part a: Concept, causes and the co2 emissions case. *Energy Policy*, *62*, 1392–1402.

Keene, A., & Deller, S. C. (2015). Evidence of the environmental kuznets' curve among us counties and the impact of social capital. *International Regional Science Review*, *38*, 358–387. doi:10.1177/0160017613496633.

Kelejian, H. H., & Piras, G. (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, *46*, 140–149.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*, 99–121.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509–533. doi:10.1111/1468-2354.00027.

Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, *104*, 219–257.

Kelejian, H. H., & Robinson, D. P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science*, *72*, 297–312. doi:10.1007/BF01434278.

Kijima, M., Nishide, K., & Ohyama, A. (2010). Economic models for the environmental kuznets curve: A survey. *Journal of Economic Dynamics and Control*, *34*, 1187–1201.

Kooijman, S. A. L. M. (1976). Some remarks on the statistical analysis of grids especially with respect to ecology. In B. Van Rootselaar (Ed.), *Annals of Systems Research: Publikatie van de Systeemgroep Nederland Publication of the Netherlands Society for Systems Research* (pp. 113–132). Boston, MA: Springer US. doi:10.1007/978-1-4613-4243-4_6.

Kostov, P. (2010). Model boosting for spatial weighting matrix selection in spatial lag models. *Environment and Planning B: Planning and Design*, *37*, 533–549. doi:10.1068/b35137.

Kostov, P. (2013). Spatial weighting matrix selection in spatial lag econometric model. *Econometrics*, *1*, 20–30. doi:10.12966/e.05.01.2013.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, *45*, 1–28.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica Journal of the Econometric Society*, *72*, 1899–1925. doi:10.1111/j.1468-0262.2004.00558.x.

Lee, L. (2007). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *137*, 489–514. doi:10.1016/j.jeconom.2005.10.004.

Lee, L., & Yu, J. (2017). Qml estimation of spatial dynamic panel data models with endogenous time varying spatial weight matrices. *Journal of Econometrics*, (pp. 173–201).

LeSage, J., & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2*, 1–33.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL : CRC Press. Formerly CIP.

Lin, X., & Lee, L. (2010). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, *157*, 34–52. doi:10.1016/j.jeconom.2009.10.035.

List, J., & Gallet, C. (1999). The environmental kuznets curve: does one size fit all? *Ecological Economics*, *31*, 409–423.

M. Escabias, A. A., & Valderrama, M. (2007). Functional pls logit regression model. *Computational Statistics and Data Analysis*, *51*, 4891 – 4902.

Maddison, D. (2006). Environmental kuznets curves: A spatial econometric approach. *Journal of Environmental Economics and Management*, *51*, 218–230.

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

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.

Marbuah, G., & Amuakwa-Mensah, F. (2017). Spatial analysis of emissions in sweden. *Energy Economics*, *68*, 383 – 394. doi:https://doi.org/10.1016/j.eneco.2017.10.003.

Martinetti, D., & Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics*, *64*, 30 – 45. doi:https://doi.org/10.1016/j.regsciurbeco.2017.02.002.

Marx, B. D., & Eilers, P. H. C. (1999). Generalized linear regression on sampled signals and curves: A p-spline approach. *Technometrics*, *41*, 1–13.

McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, *32*, 335–348. doi:10.1111/j.1467-9787.1992.tb00190.x.

Messner, J. W., Mayr, G. J., Zeileis, A., & Wilks, D. S. (2014). Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Monthly Weather Review*, *142*, 448–456.

Mizruchi, M. S., & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, *30*, 190–200.

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

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

Muller, H.-G., & Stadtmuller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, *33*, 774–805. doi:10.1214/009053604000001156.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A (General)*, *135*, 370–384.

Nerini, D., Monestiez, P., & Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, *101*, 409–418. doi:10.1016/j.jmva.2009.03.005.

Oliveros, M. E. G., Halliday, S. V., Posada, M. M. B., & Bachmann, R. (2010). Contradictions and power play in service encounters: an activity theory approach. *Cadernos EBAPE BR*, *8*, 353 – 369.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120–126.

Panayotou, T. (2003). Economic growth and the environment. *Economic Survey of Europe*, (pp. 45–72).

Pineda-Ríos, W., & Giraldo, R. (2016). Functional sar model. URL: https://arxiv.org/pdf/1609.03680.pdf.

Pinkse, J., & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, *85*, 125–154.

Pinkse, J., & Slade, M. E. (2010). The future of spatial econometrics. *Journal of Regional Science*, *50*, 103–117. doi:10.1111/j.1467-9787.2009.00645.x.

Qu, X., & Lee, L. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, *184*, 209 – 232.

Ramsay, J., & Silverman, B. (1997). Functional data analysis. *New York*, .

Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, *47*, 379–396. doi:10.1007/BF02293704.

Ramsay, J. O., & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B (Methodological)*, *53*, 539–572.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics (2nd ed.). Springer, New York.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, *14*, 1–17.

Robinson, P. (2010). Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics*, *157*, 6 – 17. doi:https://doi.org/10.1016/j.jeconom.2009.10.031. Nonlinear and Nonparametric Methods in Econometrics.

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

Ruiz-Medina, M. (2011). Spatial autoregressive and moving average hilbertian processes. *Journal of Multivariate Analysis*, *102*, 292–305. doi:https://doi.org/10.1016/j.jmva.2010.09.005.

Ruiz-Medina, M. D. (2012). Spatial functional prediction from spatial autoregressive hilbertian processes. *Environmetrics*, *23*, 119–128. doi:10.1002/env.1143.

Rupasingha, A., Goetz, S. J., Debertin, D. L., & Pagoulatos, A. (2004). The environmental kuznets curve for us counties: A spatial econometric analysis with extensions. *Papers in Regional Science*, *83*, 407–424. doi:10.1007/s10110-004-0199-x.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Severini, T. A., & Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association*, *89*, 501–511.

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

Shafik, N. (1994). Economic development and environmental quality: An econometric analysis. *Oxford Economic Papers*, *46*, 757–73.

Shi, W., & Lee, L. (2018). A spatial panel data model with time varying endogenous weights matrices and common factors. *Regional Science and Urban Economics*, *72*, 6 – 34. doi:https://doi.org/10.1016/j.regsciurbeco.2017.03.007. New Advances in Spatial Econometrics: Interactions Matter.

Smith, C. K., & C., R. J. (2004). Diagnosis murder: The death of state death taxes. *Economic Inquiry*, *42*, 537–559. doi:10.1093/ei/cbh080.

Stern, D. (2002). Explaining changes in global sulfur emissions: an econometric decomposition approach. *Ecological Economics*, *42*, 201–220.

Stern, D. (2017). The environmental kuznets curve after 25 years. *Journal of Bioeconomics*, *19*, 7–28.

Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World Development*, *32*, 1419 – 1439. doi:https://doi.org/10.1016/j.worlddev.2004.03.004.

Su, L. (2004). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, *167*, 1899–1925.

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

Sulemana, I., James, H. S., & Rikoon, J. S. (2017). Environmental kuznets curves for air pollution in african and developed countries: exploring turning point incomes and the role of democracy. *Journal of Environmental Economics and Policy*, *6*, 134–152. doi:10.1080/21606544.2016.1231635.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, *195*, 134 – 153. doi:https://doi.org/10.1016/j.jeconom.2016.07.005.

Ternynck, C. (2014). Spatial regression estimation for functional data with spatial dependency. *Journal de la Societe Française de Statistique*, *155*, 138–160.

Tingting Huang, H. W.-S. W., Gilbert Saporta (2019). Spatial functional linear model and its estimation method. URL: https://arxiv.org/abs/1811.00314.

Tjøstheim, D. (1987). Spatial series and time series: similarities and differences. In *Spatial processes and spatial time series analysis, Proceedings of the 6th Franco-Belgian Meeting of Statisticians. Publications des Facultés Universitaires Saint-Louis, Brussels* (pp. 217–228).

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, *46*, 234–240.

Torras, M., & Boyce, J. K. (1998). Income, inequality, and pollution: a reassessment of the environmental kuznets curve. *Ecological Economics*, *25*, 147–160.

Vincent, J. (1997). Testing for environmental kuznets curves within a developing country. *Environment and Development Economics*, *2*, 417–431.

Wang, F., Gao, M., Liu, J., & Fan, W. (2018). The spatial network structure of china's regional carbon emissions and its network effect. *Energies*, *11*. doi:10.3390/en11102706.

Wang, H., Iglesias, E. M., & Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, *172*, 77 – 89. doi:https://doi.org/10.1016/j.jeconom.2012.08.005.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.

Wong, R. K. W., Yao, F., & Lee, T. C. M. (2014). Robust estimation for generalized additive models. *Journal of Computational and Graphical Statistics*, *23*, 270–289. doi:10.1080/10618600.2012.756816.

Yang, K., & Lee, L. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics*, *196*, 196–214. doi:10.1016/j.jeconom.2016.04.019.

Zheng, Y., & Zhu, J. (2012). On the asymptotics of maximum likelihood estimation for spatial linear models on a lattice. *Sankhya Applied Mathematical Statistics and Probability*, *74*, 29–56. doi:10.1007/s13171-012-0009-5.

