C. R. Acad. Sci. Paris, Ser. I ••• (••••) •••–•••

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

On estimation in a spatial functional linear regression model with derivatives

Estimation dans un modèle de régression fonctionnelle spatiale avec dérivées

Stéphane Bouka [b], Sophie Dabo-Niang [a,c], Guy Martial Nkiet [b]

[a] *Laboratoire LEM, CNRS 9221, University of Lille, France*
[b] *Laboratoire URMI, University of Masuku, Franceville, Gabon*
[c] *INRIA–MODAL, Lille, France*

A R T I C L E I N F O

Article history:
Received 5 December 2016
Accepted after revision 23 February 2018
Available online xxxx

Presented by Paul Deheuvels

A B S T R A C T

This paper deals with functional linear regression for spatial data. We study the asymptotic properties of an estimator of a linear model where a spatial scalar response variable is related to a spatial functional explanatory variable and to its derivative. Convergence results with rate of this estimator are derived.

R É S U M É

Cet article aborde l'estimation de la régression linéaire fonctionnelle dans un cadre spatial. Nous étudions les propriétés asymptotiques de l'estimateur d'un modèle où une variable réponse réelle est liée à une variable dépendante fonctionnelle et sa dérivée. Nous établissons des résultats de convergence pour cet estimateur, et des vitesses de convergence sont données.

Version française abrégée

Nous considérons le modèle défini par

$$Y_{\mathbf{i}} = \langle \beta, X_{\mathbf{i}} \rangle_E + \left\langle \gamma, X'_{\mathbf{i}} \right\rangle_F + \epsilon_{\mathbf{i}}\,, \quad \mathbf{i} \in D \subset \mathbb{Z}^d, \quad d \geq 2,$$

où E est l'espace de Sobolev d'ordre $(2,1)$ ($f \in E$ signifie que f et f' appartiennent à $F := L^2[0,1]$), β et γ sont des fonctions inconnues appartenant respectivement à E et F, $X_{\mathbf{i}}$ et $\epsilon_{\mathbf{i}}$ sont des variables aléatoires centrées et indépendantes à valeurs dans E et $\mathbb{R}$, respectivement, et $X'_{\mathbf{i}}$ désigne la dérivée première de $X_{\mathbf{i}}$. Nous supposons que $(Y_{\mathbf{i}}, X_{\mathbf{i}})$ a

E-mail addresses: anaboune26@yahoo.fr (S. Bouka), sophie.dabo@univ-lille3.fr (S. Dabo-Niang), gnkiet@hotmail.com (G.M. Nkiet).

https://doi.org/10.1016/j.crma.2018.02.013
1631-073X/

la même distribution qu'un vecteur aléatoire (Y, X), et que le processus est observé dans une région $\mathcal{I}_{\mathbf{n}} = \{1, 2, ..., n\}^d$ où $\mathbf{n} = (n, ..., n)$, $n \in \mathbb{N}^*$. Nous nous intéressons à l'estimation des paramètres β et γ, puis à la prédiction à un site non visité. L'estimateur de (β, γ) est $(\widehat{\beta}_{\mathbf{n}}, \widehat{\gamma}_{\mathbf{n}})$ défini, comme dans [9], sur la base d'estimateurs empiriques donnés de (2) à (5), par $\widehat{\beta}_{\mathbf{n}} = (S_{\mathbf{n},\beta} + \psi_{\mathbf{n}} I)^{-1} u_{\mathbf{n},\beta}$ et $\widehat{\gamma}_{\mathbf{n}} = (S_{\mathbf{n},\gamma} + \psi_{\mathbf{n}} I)^{-1} u_{\mathbf{n},\gamma}$. L'opérateur $T_{\mathbf{n}}$ (resp. T) étant l'un des opérateurs suivants : $\Gamma_{\mathbf{n}}$, $\Gamma'_{\mathbf{n}}$, $\Gamma'^{*}_{\mathbf{n}}$ et $\Gamma''_{\mathbf{n}}$ (resp. $\Gamma, \Gamma', \Gamma'^{*}, \Gamma''$), on a le théorème suivant.

Théorème 3.1. *Sous les Hypothèses 3.1–3.5 avec* $\alpha_{1,\infty}(t) = O(t^{-\theta})$, $\theta \geq d+1$, *on a :*

$$\mathbb{E}\left(\|T_{\mathbf{n}} - T\|_{\infty}^2\right) = O\left(n^{-d} \log n\right).$$

Notant $\mathcal{HS}$ l'espace des opérateurs de Hilbert–Schmidt et considérant les semi-normes $\|.\|_{\Gamma} := \|\Gamma^{1/2}(.)\|_E$ et $\|.\|_{\Gamma''} := \|\Gamma''^{1/2}(.)\|_F$, on en déduit le corollaire suivant.

Corollaire 3.1. *Soit* $(v_j)_{j\geq 1}$ *une base orthonormée de* E *constituée de vecteurs propres associés aux valeurs propres* $(\lambda_j)_{j\geq 1}$ *de* Γ *avec* $\lambda_j = O(r^j)$, $0 < r < 1$, $j \geq 1$. *Alors, sous les hypothèses du Théorème 3.1, on a :*
(i) $\|T_{\mathbf{n}} - T\|_{L^2(\mathcal{HS})} = O\left(n^{-d/2} \log n\right)$.
(ii)

$$\|\beta - \widehat{\beta}_{\mathbf{n}}\|_{\Gamma}^2 = O_p\left(\frac{\psi_{\mathbf{n}}^2}{\phi_{\mathbf{n}}^2}\right) + O_p\left(\frac{(\log n)^2}{\phi_{\mathbf{n}}^2 \psi_{\mathbf{n}}^2 n^d}\right) \text{ et } \|\gamma - \widehat{\gamma}_{\mathbf{n}}\|_{\Gamma''}^2 = O_p\left(\frac{\psi_{\mathbf{n}}^2}{\phi_{\mathbf{n}}^2}\right) + O_p\left(\frac{(\log n)^2}{\phi_{\mathbf{n}}^2 \psi_{\mathbf{n}}^2 n^d}\right).$$

Posant $\mathcal{I}_{\mathbf{n}+1_d} = \{1, 2, ..., n+1\}^d$, le prédicteur et sa version «théorique» à un site non visité $\mathbf{j} \in \mathcal{I}_{\mathbf{n}+1_d} \setminus \mathcal{I}_{\mathbf{n}}$ sont respectivement définis par $\widehat{Y}_{\mathbf{j}} = \langle \widehat{\beta}_{\mathbf{n}}, X_{\mathbf{j}} \rangle_E + \left\langle \widehat{\gamma}_{\mathbf{n}}, X'_{\mathbf{j}} \right\rangle_F$ et $Y^*_{\mathbf{j}} = \langle \beta, X_{\mathbf{j}} \rangle_E + \left\langle \gamma, X'_{\mathbf{j}} \right\rangle_F$. On a alors ce qui suit.

Corollaire 3.2. *Sous les hypothèses du Corollaire 3.1, on a, pour tout* $\mathbf{j} \in \mathcal{I}_{\mathbf{n}+1_d} \setminus \mathcal{I}_{\mathbf{n}}$ *:*

$$\mathbb{E}\left[\left(\widehat{Y}_{\mathbf{j}} - Y^*_{\mathbf{j}}\right)^2\right] = O\left(\frac{\psi_{\mathbf{n}}^2}{\phi_{\mathbf{n}}^2}\right) + O\left(\frac{(\log n)^2}{\phi_{\mathbf{n}}^2 \psi_{\mathbf{n}}^2 n^d}\right).$$

1. Introduction

Several types of functional linear models for independent data have been developed over the years, serving different purposes. The most studied is perhaps the functional linear model for scalar response, originally introduced by [6]. Estimation and prediction problems for this model and some of its generalizations have been tackled mainly for independent data (see, e.g., [1], [2], [8], [9]). Some works exist on functional spatial linear prediction using krigging methods (see, e.g., [3], [4], [7], [10]). They highlight the interest of considering spatial linear functional models. In this paper, the results obtained in [9] on estimation and prediction in the functional linear model with derivatives for independent data are extended to the spatial case. Namely, we consider the model given by:

$$Y_{\mathbf{i}} = \langle \beta, X_{\mathbf{i}} \rangle_E + \left\langle \gamma, X'_{\mathbf{i}} \right\rangle_F + \epsilon_{\mathbf{i}}, \quad \mathbf{i} \in D \subset \mathbb{Z}^d, \quad d \geq 2, \tag{1}$$

where E is the Sobolev space of order $(2, 1)$ ($f \in E$ means that f and f' belong to $F := L^2[0, 1]$), β and γ are unknown functions belonging to E and F, respectively, $X_{\mathbf{i}}$ and $\epsilon_{\mathbf{i}}$ are centered and independent random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ and valued into E and $\mathbb{R}$ respectively, and $X'_{\mathbf{i}}$ stands for the first derivative of $X_{\mathbf{i}}$. We assume that $(Y_{\mathbf{i}}, X_{\mathbf{i}})$ has the same distribution as a random vector (Y, X), and that the process is observed in a region $\mathcal{I}_{\mathbf{n}} = \{1, 2, ..., n\}^d$ where $\mathbf{n} = (n, ..., n)$, $n \in \mathbb{N}^*$. In Section 2, estimators of β and γ are introduced, as well as a predictor at a non-visited site. Section 3 gives assumptions and establishes asymptotic results, namely convergences with rate of the estimates. Finally, Section 4 is devoted to some indications for proving the theoretical results presented in this note.

2. Estimation and prediction

We propose estimators of β and γ by using the same approach than [9]. Denoting by $\otimes_E$ (resp. $\otimes_F$) the tensor product defined by $(u \otimes_E v)(h) =< u, h >_E v$ (resp. $(u \otimes_F v)(h) =< u, h >_F v$), we consider the covariance and cross-covariance operators defined by

$$\Gamma = \mathbb{E}(X \otimes_E X), \quad \Gamma'^{*} = \mathbb{E}(X \otimes_E X'), \quad \Gamma' = \mathbb{E}(X' \otimes_F X), \quad \Gamma'' = \mathbb{E}(X' \otimes_F X')$$

and we set: