MOX-Report No. 33/2020

Smooth LASSO Estimator for the Function-on-Function Linear Regression Model

Centofanti, F.; Fontana, M.; Lepore, A.; Vantini, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Smooth Lasso Estimator for the Function-on-Function Linear Regression Model

Fabio Centofanti[1], Matteo Fontana[2], Antonio Lepore[1], and Simone Vantini*[2]

[1]Department of Industrial Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125, Naples, Italy
[2]MOX - Modelling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milan, Italy

Abstract

A new estimator, named as S-LASSO, is proposed for the coefficient function of a functional linear regression model where values of the response function, at a given domain point, depends on the full trajectory of the covariate function. The S-LASSO estimator is shown to be able to increase the interpretability of the model, by better locating regions where the coefficient function is zero, and to smoothly estimate non-zero values of the coefficient function. The sparsity of the estimator is ensured by a functional LASSO penalty whereas the smoothness is provided by two roughness penalties. The resulting estimator is proved to be estimation and pointwise sign consistent. Via an extensive Monte Carlo simulation study, the estimation and predictive performance of the S-LASSO estimator are shown to be better than (or at worst comparable with) competing estimators already presented in the literature before. Practical advantages of the S-LASSO estimator are illustrated through the analysis of the well known *Canadian weather* and *Swedish mortality data* .

Keywords: functional data analysis, functional regression, LASSO, B-splines, roughness penalties.

*Corresponding author. e-mail: simone.vantini@polimi.it

1 Introduction

Functional linear regression (FLR) is the generalization of the classical multivariate regression to the context of the functional data analysis (FDA) (e.g. Ramsay and Silverman (2005); Horváth and Kokoszka (2012); Hsing and Eubank (2015); Kokoszka and Reimherr (2017)), where either the predictor or the response or both have a functional form. In particular, we study the Function-on-Function (FoF) linear regression model, where both the predictor and the response variable are functions and each value of the latter, for any domain point, depends on the full trajectory of the former. The model is as follows

$$Y_i(t) = \int_{\mathcal{S}} X_i(s)\,\beta(s,t)\,ds + \varepsilon_i(t) \quad t \in \mathcal{T}, \tag{1}$$

for $i = 1, \ldots, n$. The pairs (X_i, Y_i) are independent realizations of the predictor X and the response Y, which are assumed to be smooth random process with realizations in $L^2(\mathcal{S})$ and $L^2(\mathcal{T})$, i.e., the Hilbert spaces of square integrable functions defined on the compact sets $\mathcal{S}$ and $\mathcal{T}$, respectively. Without loss of generality, the latter are also assumed with functional mean equal to zero. The functions ε_i are zero-mean random errors, independent of X_i, and have autocovariance function $K(t_1, t_2)$, t_1 and $t_2 \in \mathcal{T}$. The function β is smooth in $L^2(\mathcal{S} \times \mathcal{T})$, i.e., the Hilbert space of bivariate square integrable functions defined on the compact set $\mathcal{S} \times \mathcal{T}$, and is hereinafter referred to as *coefficient function*. For each $t \in \mathcal{T}$, the contribution of X_i to the conditional value of $Y_i(t)$ is generated by $\beta(\cdot, t)$, which works as continuous set of weights of the predictor evaluations.

The interpretability of the model in Equation (1) is of great practical interest and is based on the knowledge of the parts of the domain $\mathcal{S} \times \mathcal{T}$ where the predictor X influences the response Y (*non-null region*) or not (*null region*) i.e., β is different from or equal to zero, respectively.

FLR analysis is a hot topic in the FDA literature, a comprehensive review of the main results is provided by Morris (2015) as well as Ramsay and Silverman (2005); Horváth and Kokoszka (2012) and Cuevas (2014) give worthwhile modern perspectives. Although the research efforts have been focused mainly on the case where either the predictor or the response have functional form (Cardot et al., 2003; Li et al., 2007; Hall et al., 2007), the interest in the FoF linear regression has increased in the very last years. Besse and Cardot

(1996) developed a spline based approach to estimate the coefficient function β, Ramsay and Silverman (2005) proposed an estimator assumed to be in a finite dimension tensor space spanned by two basis sets and where regularization is achieved by either truncation or roughness penalties. Yao et al. (2005b) built up an estimation method based on the principal component decomposition of the autovariance function of both the predictor and the response based on the *principal analysis by conditional expectation* (PACE) method (Yao et al., 2005a). This estimator was extended by Chiou et al. (2014) to the case of multivariate functional responses and predictors. A general framework for the estimation of the coefficient function was proposed by Ivanescu et al. (2015) by means of the mixed model representation of the penalized regression. An extension of the ridge regression (Hastie et al., 2009) to the FoF linear regression with an application to the Italian gas market was presented in Canale and Vantini (2016). To take into account the case when the errors ε_i are correlated, in Scheipl et al. (2015) the authors developed a general framework for additive mixed models by extending the work of Ivanescu et al. (2015).

Few works address the issue of the interpretability in FLR. In the function-on-scalar setting, James et al. (2009) proposed the FLiRTI (*Functional Linear Regression That's Interpretable*) estimator that is able to recover the sparseness of the coefficient function, by imposing L_1-penalties on the coefficient function itself and its first two derivatives. An estimator obtained in two stages was proposed by Zhou et al. (2013), where an initial estimate is obtained by means of a Dantzig selector (Candes et al., 2007) refined via the group *Smoothly Clipped Absolute Deviation* (SCAD) penalty (Fan and Li, 2001). The most recent work that addresses the issue of interpretability is that of Lin et al. (2017), who proposed a *Smooth and Locally Sparse* (SLoS) estimator of the coefficient function based on the combination of the smoothing spline method with the functional SCAD penalty. However, to the best of the author knowledge, no effort has been made in the literature to obtain an interpretable estimator for the FoF linear regression model.

An interpretable estimator of the coefficient function β, named S-LASSO (Smooth plus LASSO), is proposed in this work, that is locally sparse (i.e., is zero on the null region) and, at the same time smooth on the non-null region. The property of sparseness of the S-LASSO estimator is provided by a functional L_1 penalty, which is the functional generalization of

the classical Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996). Whereas, two roughness penalties, introduced in the objective function, ensure smoothness of the estimator on the non-null region. From a computational point of view, the S-LASSO estimator is obtained as the solution of a single optimization problem, by means of a new version of the *orthant-wise limited-memory quasi-Newton* (OWL-QN) algorithm (Andrew and Gao, 2007), specifically designed to solve optimization problems involving L_1 penalties.

The paper is organized as follows. In Section 2, the S-LASSO estimator is presented. In Section 3, asymptotic properties of the S-LASSO estimator are discussed in terms of consistency and pointwise sign consistency. In Section 4, by means of a Monte Carlo simulation study the S-LASSO estimator is compared, in terms of estimation error and prediction accuracy, with competing estimators already proposed in the literature before. In Section 5, the potential of the S-LASSO estimator are demonstrated with respect of two benchmark datasets: the *Canadian weather* and *Swedish mortality data*. Proofs and algorithm description are given in the Supplementary Material.

2 Methodology

In Section 2.1, we briefly describe the smoothing spline estimator. Readers who are already familiar with this approach may skip to the next subsection. In Section 2.2, we briefly review the LASSO method, which is among the most famous methods to obtain sparse coefficient estimator for the multivariate linear regression model. Then, the LASSO penalty is extended to the FoF linear regression model. In Section 2.3, the S-LASSO estimator is defined.

2.1 The smoothing spline estimator

The smoothing spline estimator of the FoF linear regression model (Ramsay and Silverman, 2005) is the first key component of the S-LASSO estimator. It is based on the assumption that the coefficient function β may be well approximated by an element in the tensor product space generated by two spline function spaces, where a spline is a function defined piecewise by polynomials. Well-known basis functions for the spline space are the B-

splines. A B-spline basis is a set of spline functions uniquely defined by an order k and a non-decreasing sequence of $M+2$ knots, that we hereby assume to be equally spaced in a general domain $\mathcal{D}$. Cubic B-splines are B-splines of order $k=4$. Each B-spline function is a positive polynomial of degree $k-1$ over each subinterval defined by the knot sequence and is non-zero over no more than k of these subintervals (i.e., the compact support property). Note that the tensor products between the elements of the basis of two vector spaces is a basis for the tensor product space between the two spaces. Therefore, the set of the tensor products between the elements of two B-splines sets is a basis for the tensor product space of the corresponding spline function spaces, with properties that readily follow from those of B-splines defined on one-dimensional domains. In our setting, besides the computational advantage (Hastie et al., 2009), the compact support property is fundamental because it allows one to link the values of β over a given region to the B-splines with support in the same region and to discard all the B-splines that are outside that region. Thorough descriptions of splines and B-splines are in De Boor et al. (1978) and Schumaker (2007).

The smoothing spline estimator (Ramsay and Silverman, 2005) is defined as

$$\hat{\beta}_S \;=\; \underset{\alpha\in\mathbb{S}_{k_1,k_2,M_1,M_2}}{\operatorname{argmin}} \Big\{ \sum_{i=1}^{n} ||Y_i - \int_{\mathcal{S}} X_i(s)\,\alpha(s,\cdot)\,ds||^2 + \lambda_s||\mathcal{L}_s^{m_s}\alpha||^2 + \lambda_t||\mathcal{L}_t^{m_t}\alpha||^2 \Big\}, \quad (2)$$

where $\mathbb{S}_{k_1,k_2,M_1,M_2}$ is the tensor product space generated by the sets of B-splines of orders k_1 and k_2 associated with the non-decreasing sequences of M_1+2 and M_2+2 knots defined on $\mathcal{S}$ and $\mathcal{T}$, respectively. $\mathcal{L}_s^{m_s}$ and $\mathcal{L}_t^{m_t}$, with $m_s \leq k_1-1$ and $m_t \leq k_2-1$, are the m_s-th and m_t-th order linear differential operators applied to α with respect to the variables s and t, respectively. The symbol $||\cdot||$ denotes the L^2-norm corresponding to the inner product $< f, g >= \int fg$. λ_s and λ_t are parameters generally referred to as *roughness parameters*. The aim of the second and third terms on the right-hand side of Equation (2) is that of penalizing features along s and t directions. A common practice, when dealing with cubic splines, is to choose $m_s = 2$ and $m_t = 2$, which results in the penalization of the curvature of the final estimator. When $\lambda_s = \lambda_t = 0$, the wiggliness of the estimator is not penalized and the resulting estimator is the one that minimizes the sum of squared errors. On the contrary, as $\lambda_s \to \infty$ and $\lambda_t \to \infty$, $\hat{\beta}_S$ converges to a bivariate polynomial with degree equal to $|\max(m_s, m_t) - 1|$. However, there is no guarantee that $\hat{\beta}_S$ is a sparse estimator, i.e., it is exactly equal to zero in some part of the domain $\mathcal{S}\times\mathcal{T}$.

2.2 LASSO and the Functional LASSO Penalty

Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be the realizations of q scalar response variables $Y_1^s, \ldots, Y_q^s$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ the realizations of p scalar covariates $X_1^s, \ldots, X_p^s$ and $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$, with $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip})^T$, the coefficient vectors. Then, by assuming $Y_1^s, \ldots, Y_q^s, X_1^s, \ldots, X_p^s$ having zero mean, the multivariate linear regression model can be rewritten as follows

$$\boldsymbol{y}_i = \sum_{j=1}^{q} \boldsymbol{x}_i^T \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_i, \tag{3}$$

with $i = 1, \ldots, n$, where the errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are uncorrelated with $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$, independent of $X_1^s, \ldots, X_p^s$. Then, the multivariate LASSO estimator of the coefficient vector $\boldsymbol{\beta}$ is (Tibshirani, 1996)

$$\hat{\boldsymbol{\beta}}_{L1}, \ldots, \hat{\boldsymbol{\beta}}_{Lq} = \underset{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{q} \left(y_{ij} - \boldsymbol{x}_i^T \boldsymbol{\alpha}_j \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{q} |\boldsymbol{\alpha}_j|^T \mathbf{1} \leq \lambda^*, \tag{4}$$

where $|\boldsymbol{\alpha}_j| = (|\alpha_{j1}|, \ldots, |\alpha_{jp}|)^T$, with $|\cdot|$ denoting the absolute value, $\mathbf{1}$ the unit vector of length p and λ^* a positive constant. Equivalently,

$$\hat{\boldsymbol{\beta}}_{L1}, \ldots, \hat{\boldsymbol{\beta}}_{Lq} = \underset{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{q} \left(y_{ij} - \boldsymbol{x}_i^T \boldsymbol{\alpha}_j \right)^2 + \lambda_L \sum_{j=1}^{q} |\boldsymbol{\alpha}_j|^T \mathbf{1} \right\}, \tag{5}$$

where λ_L is a positive constant in one-to-one inverse correspondence with λ^*. The constant λ_L, usually called *regularization parameter*, controls the degree of shrinkage towards zero applied to the resulting estimator. The larger this value, the larger the probability that some elements of the coefficient estimates are exactly zero. As stated in Section 1, the LASSO was introduced to improve the prediction accuracy (in terms of expected mean square error) by making the final estimator biased, as well as its interpretability, by performing automatic variable selection (Tibshirani, 1996). The LASSO penalty constrains the resulting estimator to belong to a region that is a cross-polytope with corners on the axis and, thus, it maximizes the probability of obtaining a coefficient vector with zero elements.

Similar properties can be inherited by the S-LASSO estimator, by observing that the multivariate linear model matches the FoF linear regression model if $\sum_{j=1}^{q} \boldsymbol{x}_i^T \boldsymbol{\beta}_j$ is replaced by $\int_{\mathcal{S}} X_i(s)\, \beta(s,t)\, ds$, and, $\boldsymbol{y}_i$ and $\boldsymbol{\varepsilon}_i$ are substituted by $Y_i(t)$ and $\varepsilon_i(t)$, with $t \in \mathcal{T}$, respectively. Therefore, a straightforward generalization of the LASSO penalty $\lambda_L \sum_{j=1}^{q} |\boldsymbol{\alpha}_j|^T \mathbf{1}$ in

Equation (5) to the FoF linear regression model, referred to as *functional LASSO penalty*, is

$$\mathcal{P}_L(\alpha) = \lambda_L \int_{\mathcal{S}} \int_{\mathcal{T}} |\alpha(s,t)| ds dt, \tag{6}$$

for $\alpha \in L^2(\mathcal{S} \times \mathcal{T})$.

2.3 The S-LASSO Estimator

In view of the functional LASSO penalty of Equation (6) and the smoothing spline estimator of Equation (2), the S-LASSO estimator is defined as follows

$$\hat{\beta}_{SL} = \underset{\alpha \in \mathbb{S}_{k_1,k_2,M_1,M_2}}{\operatorname{argmin}} \Big\{ \sum_{i=1}^{n} ||Y_i - \int_{\mathcal{S}} X_i(s)\, \alpha(s,\cdot)\, ds||^2 + \lambda_s ||\mathcal{L}_s^{m_s} \alpha||^2 + \lambda_t ||\mathcal{L}_t^{m_t} \alpha||^2 + \mathcal{P}_L(\alpha) \Big\}, \tag{7}$$

where the last two terms represent the two roughness penalties introduced in Equation (2) to control the smoothness of the coefficient function estimator. It is worth noting that, in general, the estimator smoothness may be also controlled by opportunely choosing the dimension of the space $\mathbb{S}_{k_1,k_2,M_1,M_2}$, that is, by fixing k_1 and k_2, and choosing M_1 and M_2 (Ramsay and Silverman, 2005). However, this strategy is not suitable in this case. To obtain a sparse estimator, the dimension of the space $\mathbb{S}_{k_1,k_2,M_1,M_2}$ must be in fact as large as possible. In this way, the value of β in a given region is strictly related to the coefficients of the B-spline functions defined on the same part of the domain and, thus, they tend to be zero in the null region. On the contrary, when the dimension of $\mathbb{S}_{k_1,k_2,M_1,M_2}$ is small, there is a larger probability that some B-spline functions have support both in the null and non-null regions and, thus the corresponding B-spline coefficients result different from zero. Therefore, we find suitable the use of the two roughness penalty terms in Equation (7).

To compute the S-LASSO estimator, let us consider the space $\mathbb{S}_{k_1,k_2,M_1,M_2}$ generated by the two sets of B-splines $\boldsymbol{\psi}^s = \left(\psi_1^s, \ldots, \psi_{M_1+k_1}^s\right)^T$ and $\boldsymbol{\psi}^t = \left(\psi_1^t, \ldots, \psi_{M_2+k_2}^t\right)^T$, of order k_1 and k_2 and non-decreasing knots sequences $\Delta^s = \{s_0, s_1, \ldots, s_{M_1}, s_{M_1+1}\}$ and $\Delta^t = \{t_0, t_1, \ldots, t_{M_2}, t_{M_2+1}\}$, defined on $\mathcal{S} = [s_0, s_{M_1+1}]$ and $\mathcal{T} = [t_0, t_{M_2+1}]$, respectively. Similarly to the standard smoothing spline estimator, by performing the minimization in

Equation (7) over $\alpha \in \mathbb{S}_{k_1,k_2,M_1,M_2}$, we implicitly assume that β can be suitably approximated by an element in $\mathbb{S}_{k_1,k_2,M_1,M_2}$, that is

$$\beta(s,t) \approx \tilde{\beta}(s,t) \doteq \sum_{i=1}^{M_1+k_1} \sum_{j=1}^{M_2+k_2} b_{ij}\psi_i^s(s)\psi_j^t(t) = \boldsymbol{\psi}^s(s)^T \boldsymbol{B}\boldsymbol{\psi}^t(t) \quad s \in \mathcal{S}, t \in \mathcal{T}, \tag{8}$$

with $\boldsymbol{B} = \{b_{ij}\} \in \mathbb{R}^{M_1+k_1 \times M_2+k_2}$ and b_{ij} are scalar coefficients. So stated, the problem of estimating β has been reduced to the estimation of the unknown coefficient matrix $\boldsymbol{B}$. Let $\alpha(s,t) = \boldsymbol{\psi}^s(s)^T \boldsymbol{B}_\alpha \boldsymbol{\psi}^t(t)$, $s \in \mathcal{S}, t \in \mathcal{T}$, in $\mathbb{S}_{k_1,k_2,M_1,M_2}$, where $\boldsymbol{B}_\alpha = \{b_{\alpha,ij}\} \in \mathbb{R}^{M_1+k_1 \times M_2+k_2}$. Then, the first term of the right-hand side of Equation (7) may be rewritten as

$$\sum_{i=1}^{n} ||Y_i - \int_{\mathcal{S}} X_i(s)\alpha(s,\cdot)\,ds||^2 = \sum_{i=1}^{n} \int_{\mathcal{T}} Y_i(t)^2\,dt - 2\,\mathrm{Tr}\left[\boldsymbol{X}\boldsymbol{B}_\alpha\boldsymbol{Y}^T\right] + \mathrm{Tr}\left[\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_\alpha\boldsymbol{W}_t\boldsymbol{B}_\alpha^T\right], \tag{9}$$

whereas, the roughness penalties of the left side of Equation (7) become

$$\lambda_s||\mathcal{L}_s^{m_s}\alpha||^2 = \lambda_s\,\mathrm{Tr}\left[\boldsymbol{B}_\alpha^T\boldsymbol{R}_s\boldsymbol{B}_\alpha\boldsymbol{W}_t\right] \quad \lambda_t||\mathcal{L}_t^{m_t}\alpha||^2 = \lambda_t\,\mathrm{Tr}\left[\boldsymbol{B}_\alpha^T\boldsymbol{W}_s\boldsymbol{B}_\alpha\boldsymbol{R}_t\right], \tag{10}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)^T$, with $\boldsymbol{X}_i = \int_{\mathcal{S}} X_i(s)\boldsymbol{\psi}^s(s)\,ds$, $\boldsymbol{Y} = (\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n)^T$ with $\boldsymbol{Y}_i = \int_{\mathcal{T}} Y_i(t)\boldsymbol{\psi}^t(t)\,dt$ $\boldsymbol{W}_s = \int_{\mathcal{S}} \boldsymbol{\psi}^s(s)\boldsymbol{\psi}^s(s)^T\,ds$, $\boldsymbol{W}_t = \int_{\mathcal{T}} \boldsymbol{\psi}^t(t)\boldsymbol{\psi}^t(t)^T\,dt$, $\boldsymbol{R}_s = \int_{\mathcal{S}} \mathcal{L}_s^{m_s}[\boldsymbol{\psi}^s(s)]\,\mathcal{L}_s^{m_s}[\boldsymbol{\psi}^s(s)]^T\,ds$ and $\boldsymbol{R}_t = \int_{\mathcal{T}} \mathcal{L}_t^{m_t}[\boldsymbol{\psi}^t(t)]\,\mathcal{L}_t^{m_t}[\boldsymbol{\psi}^t(t)]^T\,dt$. The term $\mathrm{Tr}[\boldsymbol{A}]$ denotes the trace of a square matrix $\boldsymbol{A}$. Note that, if the functional LASSO penalty would be expressed as a function of $|\boldsymbol{B}_\alpha| \doteq \{|b_{\alpha,ij}|\}$, then, standard optimization algorithms for L_1-regularized loss would be easily implemented to solve the problem in Equation (7). Unfortunately, the most simple form we obtain is as follows

$$\lambda_L \int_{\mathcal{S}}\int_{\mathcal{T}} |\alpha(s,t)|dsdt = \lambda_L \int_{\mathcal{S}}\int_{\mathcal{T}} |\boldsymbol{\psi}^s(s)^T\boldsymbol{B}_\alpha\boldsymbol{\psi}^t(t)|dsdt. \tag{11}$$

By the following theorem, we shall provide a practical way to face the issue by finding a reasonable approximation of the functional LASSO penalty.

Theorem 1. *Let* $\mathbb{S}_{k_1,k_2,\Delta_{1,e}\Delta_{2,e}} = \mathrm{span}\{B_{i_1}B_{i_2}\}_{i_1=1,i_2=1}^{M_1+k_1,M_2+k_2}$, *with* $\{B_{i_j}\}$ *the set of B-splines of orders* k_j *and non-decreasing knots sequences* $\Delta_j = \{x_{j,0}, x_{j,1}, \ldots, x_{j,M_j}, x_{j,M_j+1}\}$ *defined on the compact set* $\mathcal{D}_j = \left[x_{j,0}, x_{j,M_j+1}\right]$ *and* $\Delta_{j,e}$ *the extended partitions corresponding to* Δ_j *defined as* $\Delta_{j,e} = \{y_{j,l}\}_{l=1}^{M_j+2k_j}$ *where* $y_{j,1},\ldots,y_{j,k_j} = x_{j,0}$, $y_{j,1+k_j},\ldots,y_{j,M_j+k_j} =$

$x_{j,1},\ldots,x_{j,M_j}$ *and* $y_{j,M_j+1+k_j},\ldots,y_{j,M_j+2k_j}=x_{j,M_j+1}$*, for* $j=1,2$*. Then, for* $f(z_1,z_2)=\sum_{i_1=1}^{M_1+k_1}\sum_{i_2=1}^{M_2+k_2}c_{i_1i_2}B_{i_1}(z_1)B_{i_2}(z_2)\in\mathbb{S}_{k_1,k_2,\Delta_{1,e}\Delta_{2,e}}$*, with* $z_1\in\mathcal{D}_1$ *and* $z_2\in\mathcal{D}_2$*,*

$$0\leq||f||_{\ell^1,\Delta_{1,e},\Delta_{2,e}}-||f||_{L^1}=O\left(\frac{1}{M_1}\right)+O\left(\frac{1}{M_2}\right),\tag{12}$$

where

$$||f||_{\ell^1,\Delta_{2,e},\Delta_{1,e}}=\sum_{i_1=1}^{M_1+k_1}\sum_{i_2=1}^{M_2+k_2}|c_{i_1i_2}|\frac{(y_{1,i_1+k_1}-y_{1,i_1})(y_{2,i_2+k_2}-y_{2,i_2})}{k_1k_2},\tag{13}$$

and

$$||f||_{L^1}=\int_{\mathcal{D}_1}\int_{\mathcal{D}_2}|f(z_1,z_2)|dz_1dz_2.\tag{14}$$

The interpretation of the above theorem is quite simple. It basically says that for large values of M_1 and M_2, $||f||_{L^1}$ is well approximated from the top by $||f||_{\ell^1,\Delta_{2,e},\Delta_{1,e}}$ and the approximation error tends to zero as $M_1,M_2\to\infty$. By using this result, the functional LASSO penalty $\mathcal{P}_L(\alpha)$ can be approximated by

$$\mathcal{P}_L(\alpha)\approx\lambda_L\sum_{i=1}^{M_1+k_1}\sum_{j=1}^{M_2+k_2}|b_{\alpha,ij}|\frac{\left(s^e_{i+k_1}-s^e_i\right)\left(t^e_{j+k_2}-t^e_j\right)}{k_1k_2}=\lambda_L\boldsymbol{w}_s^T|\boldsymbol{B}_\alpha|\boldsymbol{w}_t,\tag{15}$$

where $\{s^e_i\}$ and $\{t^e_i\}$ are the extended partitions associated with Δ^s and Δ^t, respectively, $\boldsymbol{w}_s=\left[\frac{\left(s^e_{1+k_1}-s^e_1\right)}{k_1},\ldots,\frac{\left(s^e_{M_1+2k_1}-s^e_{M_1+k_1}\right)}{k_1}\right]^T$ and $\boldsymbol{w}_t=\left[\frac{\left(t^e_{1+k_2}-t^e_1\right)}{k_2},\ldots,\frac{\left(t^e_{M_2+2k_2}-t^e_{M_2+k_2}\right)}{k_2}\right]^T$. Therefore, upon using the approximation in in Equation (15), Equation (9) and Equation (10), the optimization problem in Equation (7) becomes

$$\begin{aligned}\hat{\boldsymbol{B}}_{SL}\approx\underset{\boldsymbol{B}_\alpha\in\mathbb{R}^{(M_1+k_1)\times(M_2+k_2)}}{\operatorname{argmin}}\Big\{&\sum_{i=1}^n\int_{\mathcal{T}}Y_i(t)^2\,dt-2\operatorname{Tr}\left[\boldsymbol{X}\boldsymbol{B}_\alpha\boldsymbol{Y}^T\right]+\operatorname{Tr}\left[\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}_\alpha\boldsymbol{W}_t\boldsymbol{B}_\alpha^T\right]\\&+\lambda_s\operatorname{Tr}\left[\boldsymbol{B}_\alpha^T\boldsymbol{R}_s\boldsymbol{B}_\alpha\boldsymbol{W}_t\right]+\lambda_t\operatorname{Tr}\left[\boldsymbol{B}_\alpha^T\boldsymbol{W}_s\boldsymbol{B}_\alpha\boldsymbol{R}_t\right]+\lambda_L\boldsymbol{w}_s^T|\boldsymbol{B}_\alpha|\boldsymbol{w}_t\Big\}.\end{aligned}\tag{16}$$

Then, the coefficient β is estimated by $\hat{\beta}_{SL}(s,t)=\boldsymbol{\psi}^s(s)^T\hat{\boldsymbol{B}}_{SL}\boldsymbol{\psi}^t(t)$ for $s\in\mathcal{S}$ and $t\in\mathcal{T}$. The optimization problem with L_1-regularized loss in Equation (19) is (i) convex, being sum or integral of convex function; and (ii) has a unique solution given some general conditions on the matrix $\boldsymbol{W}_t\otimes\boldsymbol{X}^T\boldsymbol{X}$ (with $\otimes$ the Kronecker product). Unfortunately, the objective function is not differentiable at zero, and thus it has not a closed-form solution. In view of this, general purpose gradient-based optimization algorithms – as for instance the *L-BFGS*

quasi-Newton method (Nocedal and Wright, 2006) – and classical optimization algorithms for solving LASSO problems – such as coordinate descent (Friedman et al., 2010) and least-angle regression (LARS) (Efron et al., 2004) – are not suitable. In contrast, we found very promising a modified version of the *orthant-wise limited-memory quasi-Newton* (OWL-QN) algorithm proposed by Andrew and Gao (2007).

The OWL-QN algorithm is based on the fact that the L_1 norm is differentiable for the set of points in which each coordinate never changes sign (i.e., *orthant*), being a linear function of its argument. In each orthant, the second-order behaviour of an objective function of the form $f(\boldsymbol{x}) = l(\boldsymbol{x}) + C||\boldsymbol{x}||_1$, to be minimized, is determined by l alone. The function $l : \mathbb{R}^r \to \mathbb{R}$ is a convex, bounded below, continuously differentiable with continuously differentiable gradient ∇l, $\boldsymbol{x} = (x_1, \ldots, x_r)^T$, C is a given positive constant, and $||\cdot||_1$ is the usual ℓ_1 norm. Therefore, Andrew and Gao (2007) propose to derive a quadratic approximation of the function l that is valid for some orthant containing the current point and then to search for the minimum of the approximation, by constraining the solution in the orthant where the approximation is valid. There may be several orthants containing or adjacent to a given point. The choice of the orthant to explore is based on the *pseudo-gradiant* $\diamond f(\boldsymbol{x})$ of f at $\boldsymbol{x}$, whose components are defined as

$$\diamond_i f(\boldsymbol{x}) = \begin{cases} \frac{\partial l(x)}{\partial x_i} + C\,\mathrm{sign}(x_i) & \text{if } |x_i| > 0 \\ \frac{\partial l(x)}{\partial x_i} + C & \text{if } x_i = 0, \frac{\partial l(x)}{\partial x_i} < -C \\ \frac{\partial l(x)}{\partial x_i} - C & \text{if } x_i = 0, \frac{\partial l(x)}{\partial x_i} > C \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

where $\mathrm{sign}(\cdot)$ is the sign function.

However, the objective function of the optimization problem in Equation (19) is of the form $f^*(\boldsymbol{x}) = l(\boldsymbol{x}) + C||\boldsymbol{D}\boldsymbol{x}||_1$, where $\boldsymbol{D} = \{d_i\} \in \mathbb{R}^{r\times r}$ is a diagonal matrix of positive weights. To take into account the weights, the OWL-QN algorithm must be implemented with a different *pseudo-gradiant* $\diamond f^*(\boldsymbol{x})$ of f^* at $\boldsymbol{x}$, whose components are defined as

$$\diamond_i f^*(\boldsymbol{x}) = \begin{cases} \frac{\partial l(x)}{\partial x_i} + d_i C\,\mathrm{sign}(x_i) & \text{if } |x_i| > 0 \\ \frac{\partial l(x)}{\partial x_i} + d_i C & \text{if } x_i = 0, \frac{\partial l(x)}{\partial x_i} < -C \\ \frac{\partial l(x)}{\partial x_i} - d_i C & \text{if } x_i = 0, \frac{\partial l(x)}{\partial x_i} > C \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

A more detailed description of the OWL-QN algorithm for objective functions of the form $l(\boldsymbol{x})+C||\boldsymbol{D}\boldsymbol{x}||_1$ is given found in the Supplementary Material. Note that, the optimization problem in Equation (19) can be rewritten by vectorization as

$$\hat{\boldsymbol{b}}_{SL} \approx \hat{\boldsymbol{b}}_{app} = \underset{\boldsymbol{b}_\alpha \in \mathbb{R}^{(M_1+k_1)(M_2+k_2)}}{\operatorname{argmin}} \Big\{ -2\,\operatorname{vec}\left(\boldsymbol{X}^T\boldsymbol{Y}\right)^T \boldsymbol{b}_\alpha + \boldsymbol{b}_\alpha^T \left(\boldsymbol{W}_t \otimes \boldsymbol{X}^T\boldsymbol{X}\right) \boldsymbol{b}_\alpha + \lambda_s \boldsymbol{b}_\alpha^T \boldsymbol{L}_{wr} \boldsymbol{b}_\alpha + \lambda_t \boldsymbol{b}_\alpha^T \boldsymbol{L}_{rw} \boldsymbol{b}_\alpha + \lambda_L ||\boldsymbol{W}_{st}\boldsymbol{b}_\alpha||_1 \Big\}, \tag{19}$$

where $\hat{\boldsymbol{b}}_{SL} = \operatorname{vec}\left(\hat{\boldsymbol{B}}_{SL}\right)$, $\boldsymbol{L}_{rw} \doteq (\boldsymbol{R}_t \otimes \boldsymbol{W}_s)$ and $\boldsymbol{L}_{wr} \doteq (\boldsymbol{W}_t \otimes \boldsymbol{R}_s)$, and $\boldsymbol{W}_{st}$ is the diagonal matrix whose diagonal elements are $\boldsymbol{w}_s^T \otimes \boldsymbol{w}_t^T$. Moreover, for generic a matrix $\boldsymbol{A} \in \mathbb{R}^{j\times k}$, $\operatorname{vec}(\boldsymbol{A})$ indicates the vector of length jk obtained by writing the matrix $\boldsymbol{A}$ as a vector column-wise. Therefore, the OWL-QN with pseudo-gradient as in Equation (18) can be straightforwardly applied.

In the following, we summarize all the parameters that need to be set to obtain the S-LASSO estimator. The orders k_1 and k_2 should be chosen with respect to the degree of smoothness we want to achieve, and the computational efforts. The larger k_1 and k_2 the smoother the resulting estimator. As stated before, M_1 and M_2 should be as large as possible to ensure that the null region is correctly captured and the approximation in Equation (15) is valid, with respect to the maximum computational efforts. A standard choice is $k_1 = k_2 = 4$, i.e., cubic B-splines, with equally spaced knot sequences. In the smoothing spline estimator, the choice of M_1 and M_2 is not crucial (Cardot et al., 2003), because the smoothness of the estimator is controlled by the two smoothing parameters. Finally, at given k_1, k_2, M_1, and M_2, the optimal values of λ_s, λ_t and λ_L can be selected as those that minimize the the estimated prediction error function $CV(\lambda_s, \lambda_t, \lambda_L)$, i.e., $CV(\lambda_s, \lambda_t, \lambda_L)$, over a grid of candidate values (Hastie et al., 2009). However, although this choice could be optimal for the prediction performance, it may affect the interpretability of the model. Much more interpretable models, with a slight decrease in predictive performance, may in fact exist. The *k-standard error* rule, which is a generalization of the *one-standard error* rule (Hastie et al., 2009), may be a more reasonable choice. That is, to choose the most sparse model whose error is no more than k standard errors above the error of the best model. In practice, as spareness is controlled by the parameter λ_L, we first find the best model in terms of estimated prediction error at given λ_L and then, among the selected

models, we apply the k-standard error rule. This rule may be particularly useful when $CV(\lambda_s, \lambda_t, \lambda_L)$ is flat with respect to λ_L, in this case, it chooses the simplest model among those achieving similar estimated prediction error.

3 Theoretical Properties of the S-LASSO Estimator

In this section, we provide some theoretical results on the S-LASSO estimator, under some regularity assumptions, i.e., the estimation consistency (Theorem 2) and the pointwise sign consistency (Theorem 3) of $\hat{\beta}_{SL}$. All proofs are in the Supplementary Material. The following regularity conditions are assumed.

C 1. $||X||_2$ *is almost surely bounded, i.e.,* $||X||_2 \leq c < \infty$.

C 2. β *is in the Hölder space* $C^{p',\nu}(\mathcal{S} \times \mathcal{T})$ *defined as the set of functions* f *on* $\mathcal{S} \times \mathcal{T}$ *having continuous partial and mixed derivatives up to order* p' *and such that the partial and mixed derivatives of order* p' *are Hölder continuous, that is,* $|f^{(p')}(\boldsymbol{x}_1) - f^{(p')}(\boldsymbol{x}_2)| \leq c||\boldsymbol{x_1} - \boldsymbol{x}_2||^{\nu}$, *for some constant* c, *integer* p' *and* $\nu \in [0, 1]$, *and for all* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S} \times \mathcal{T}$, *where* $f^{(p')}$ *is the partial and mixed derivatives of order* p'. *Moreover, let* $p \doteq p' + \nu$ *such that* $3/2 < p \leq k_1 - 1$ *and* $3/2 < p \leq k_2 - 1$.

C 3. $M_1 = o\left(n^{1/4}\right)$, $M_2 = o\left(n^{1/4}\right)$, $M_1 = \omega\left(n^{\frac{1}{2p+1}}\right)$, $M_2 = \omega\left(n^{\frac{1}{2p+1}}\right)$, *where* $a_n = \omega(b_n)$ *means* $\frac{a_n}{b_n} \to \infty$ *for* $n \to \infty$,

C 4. *There exist two positive constants* b *and* B *such that*

$$b \leq \Lambda_{min}\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right) \leq \Lambda_{max}\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right) \leq B, \tag{20}$$

where $\Lambda_{min}(\boldsymbol{M})$ *and* $\Lambda_{max}(\boldsymbol{M})$ *denote the minimum and maximum eigenvalues of the matrix* $\boldsymbol{M}$.

C 5. $\lambda_s = o\left(M_1^{-2m_s+1}\right)$, $\lambda_t = o\left(M_2^{-2m_t+1}\right)$.

C.1 and C.2 are the anoulogus of (H1) and (H2) in Cardot et al. (2003) for a bivariate regression function. C.2 ensures that β is sufficiently smooth. C.3 provides information on the growth rate of the number of knots M_1 and M_2. C.4 is the anolugus of condition (F)

of Fan et al. (2004) and assumes that the matrix $\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right)$ has reasonably good behaviour, whereas, C.5 provides guidance on the choice of the parameters λ_s and λ_t.

Theorem 2 shows that with probability tending to one there exists a solution of the optimization problem in Equation (7) that converges to $\tilde{\beta}$, chosen such that $||\beta - \tilde{\beta}||_\infty = O(M_1^{-p}) + O(M_2^{-p})$. To prove Theorem 2, in addition to C.1-C.5, the following condition is considered.

C 6. $\lambda_L = o\left(M_1^{-1}M_2^{-1}\right)$.

The first result is about the convergence rate of $\hat{\beta}_{SL}$ to β in terms of L_∞-norm.

Theorem 2. *Under assumptions C.1-C.6, there exists a unique solution $\hat{\beta}_{SL}$ of the optimization problem in Equation* (7)*, such that*

$$||\hat{\beta}_{SL} - \beta||_\infty = O_p\left(M_1^{1/2}M_2^{1/2}n^{-1/2}\right). \tag{21}$$

According to the above theorem, there exists an estimator $\hat{\beta}_{SL}$ of β that is root-n/M_1M_2 consistent.

Before stating Theorem 3, let us define with $\boldsymbol{b}_{(1)}$ the vector whose entries are the q non-zero elements of $\boldsymbol{b}$ that are and with $\boldsymbol{b}_{(2)}$ the vector whose entries are the $(M_1 + k_1)(M_2 + k_2) - q$ elements of $\boldsymbol{b}$ that are equal to zero. In what follows, we assume, without loss of generality, that $\boldsymbol{b} = \begin{bmatrix}\boldsymbol{b}_{(1)}^T & \boldsymbol{b}_{(2)}^T\end{bmatrix}^T$ and that a matrix $\boldsymbol{A}_l \in \mathbb{R}^{(M_1+k_1)(M_2+k_2)\times(M_1+k_1)(M_2+k_2)}$ can be expressed in block-wise form as

$$\boldsymbol{A}_l = \begin{bmatrix} \boldsymbol{A}_{l,11} \in \mathbb{R}^{q\times q} & \boldsymbol{A}_{l,12} \in \mathbb{R}^{q\times(M_1+k_1)(M_2+k_2)-q} \\ \boldsymbol{A}_{l,21} \in \mathbb{R}^{(M_1+k_1)(M_2+k_2)-q\times q} & \boldsymbol{A}_{l,22} \in \mathbb{R}^{(M_1+k_1)(M_2+k_2)-q\times(M_1+k_1)(M_2+k_2)-q} \end{bmatrix}.$$

To prove Theorem 3, in addition to C.1-C.5, the following conditions are considered.

C 7. *(S-LASSO irrepresentable condition (SL-IC)) There exists λ_s, λ_t, λ_L, and a constant $\eta > 0$ such that, element-wise,*

$$\begin{aligned}\Big|\boldsymbol{W}_{st,21}^{-1}\Big\{ &\left[\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right)_{21} + n^{-1}\lambda_s\boldsymbol{L}_{wr,21} + n^{-1}\lambda_t\boldsymbol{L}_{rw,21}\right] \\ &\left[\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right)_{11} + n^{-1}\lambda_s\boldsymbol{L}_{wr,11} + n^{-1}\lambda_t\boldsymbol{L}_{rw,11}\right]^{-1} \\ &\left[\boldsymbol{W}_{st,11}\,\mathrm{sign}\left(\boldsymbol{b}_{\alpha(1)}\right) + 2\lambda_L^{-1}\lambda_s\boldsymbol{L}_{wr,11}\boldsymbol{b}_{(1)} + 2\lambda_L^{-1}\lambda_t\boldsymbol{L}_{rw,11}\boldsymbol{b}_{(1)}\right] \\ &-2\lambda_L^{-1}\lambda_s\boldsymbol{L}_{wr,21}\boldsymbol{b}_{(1)} - 2\lambda_L^{-1}\lambda_t\boldsymbol{L}_{rw,21}\boldsymbol{b}_{(1)}\Big\}\Big| \leq 1 - \eta.\end{aligned}$$

C 8. *The functions $\varepsilon_i(t)$ in Equation (1) are zero mean Gaussian random processes with autocovariance function $K(t_1, t_2)$, t_1 and $t_2 \in \mathcal{T}$, independent of X_i.*

C 9. *Given $\rho \doteq \min |\left[\left(\boldsymbol{W}_t \otimes \boldsymbol{X}^T\boldsymbol{X}\right)_{11} + \lambda_s \boldsymbol{L}_{wr,11} + \lambda_t \boldsymbol{L}_{rw,11}\right]^{-1} \left[\left(\boldsymbol{W}_t \otimes \boldsymbol{X}^T\boldsymbol{X}\right)_{11} \boldsymbol{b}_{(1)}\right]|$ and $C_{min} \doteq \Lambda_{min}\left[\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right)_{11}\right]$, $\Lambda_{min}(\boldsymbol{W}_t) M_2 \to c_w$ as $n \to \infty$, with $0 < c_w < \infty$, and the parameters λ_s, λ_t and λ_L are chosen such that*

(a)

$$\frac{M_1^2 M_2^2 \log\left[(M_1 + k_1)(M_2 + k_2) - q\right]}{\lambda_L^2}\left[nc^2 + \frac{\lambda_s^2 \Lambda_{max}^2(\boldsymbol{L}_{wr})}{nC_{min}} + \frac{\lambda_t^2 \Lambda_{max}^2(\boldsymbol{L}_{rw})}{nC_{min}}\right] = o(1),$$

(b)

$$\frac{1}{\rho}\Bigg\{\sqrt{\frac{M_1 M_2 \log(q)}{nC_{min}}} + \frac{\lambda_L}{nM_1M_2}\Lambda_{min}^{-1}\left[\left(\boldsymbol{W}_t \otimes n^{-1}\boldsymbol{X}^T\boldsymbol{X}\right)_{11} + \lambda_s n^{-1}\boldsymbol{L}_{wr,11} + \lambda_t n^{-1}\boldsymbol{L}_{rw,11}\right] ||\operatorname{sign}\left(\boldsymbol{b}_{(1)}\right)||_2\Bigg\} = o(1).$$

The SL-IC in C.7 is the straightforward generalization to the problem in Equation (7) of the elastic irrepresentable condition described in Jia and Yu (2010). It is a consequence of the standard Karush−Kuhn−Tucker (KKT) conditions applied to the optimization problem in Equation (19). C.8 gives some conditions on the relationship of λ_s, λ_t, and λ_L with M_1, M_2 and n. In the classical setting, an estimator is sign selection consistent if it has the same sign of the true parameter with probability tending to one. Analogously, we say that an estimator of the coefficient function β is pointwise sign consistent if, in each point of the domain, it has the same sign of β with probability tending to one. The following theorem states that, under opportune assumptions, the S-LASSO estimator is pointwise sign consistent.

Theorem 3. *Under assumptions C.1-C.5 and C.7-C.9, $\hat{\beta}_{SL}$ is pointwise sign consistent, that is, for all $s \in \mathcal{S}$ and $t \in \mathcal{T}$,*

$$\Pr\left\{\operatorname{sign}\left[\hat{\beta}_{SL}(s,t)\right] = \operatorname{sign}\left[\beta(s,t)\right]\right\} \to 1, \tag{22}$$

as $n \to \infty$.

4 Simulation Study

In this section, we conduct a Monte Carlo simulation study to explore the performance of the S-LASSO estimator. We consider the four following different scenarios

- **Scenario I** The coefficient function is zero all over the domain, i.e., $\beta(s,t)=0$, if $(s,t) \in [0,1] \times [0,1]$.

- **Scenario II** β is different from zero in the central part of the domain (Figure 1(a)), i.e.,

$$\beta(s,t) = \begin{cases} -\left(\frac{s-0.5}{0.25}\right)^2 - \left(\frac{t-0.5}{0.25}\right)^2 + 1 & \text{if } 0.5 - 0.25\sqrt{1-(t-0.5)^2} \leq s \leq 0.5 + 0.25\sqrt{1-(t-0.5)^2} \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

- **Scenario III** β is different from zero on the edge of the domain (Figure 1(b)), i.e.,

$$\beta(s,t) = \begin{cases} 0.5(1-t)\sin\left[10\pi\left(t - 1.05 + \sqrt{1-(s-0.5)^2}\right)\right] & \text{if } t \leq 1.05 - \sqrt{1-(s-0.5)^2} \\ 0.5\sin\left(10\pi\left(s + 1.05 + \sqrt{1-(t-0.5)^2}\right)\right) & \text{if } s \leq -0.05 - \sqrt{1-(t-0.5)^2} \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

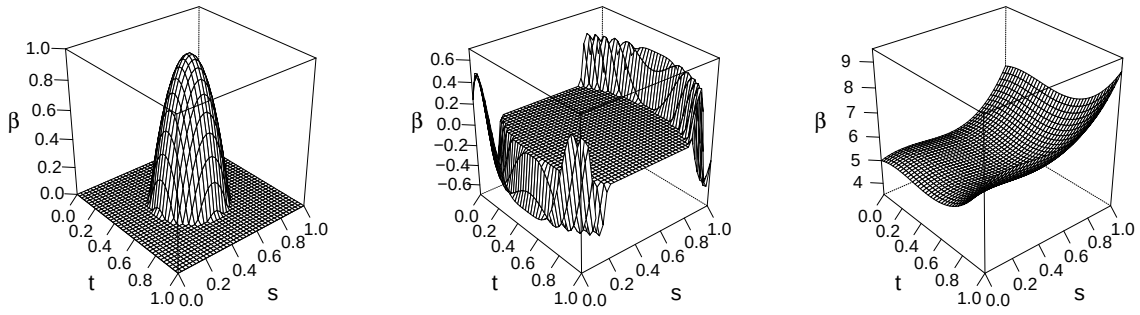
- **Scenario IV** β is non-null everywhere (Figure 1(c)).

$$\beta(s,t) = \left(\frac{t-0.5}{0.5}\right)^3 + \left(\frac{s-0.5}{0.5}\right)^3 + \left(\frac{t-0.5}{0.5}\right)^2 + \left(\frac{s-0.5}{0.5}\right)^2 + 5. \tag{25}$$

  This scenario is not expected to be favourable to the S-LASSO estimator.

The independent observations of the covariates X_i and errors ε_i are generated as $X_i = \sum_{j=1}^{32} x_{ij}\psi_i^x$ and $\varepsilon_i = k\sum_{j=1}^{20} e_{ij}\psi_i^\varepsilon$, where the coefficients x_{ij} and e_{ij} are independent realizations of standard normal random variable, and $\psi_i^x(s)$ and $\psi_i^\varepsilon(s)$ are cubic B-splines with evenly spaced knot sequence (the numbers of basis has been randomly chosen between 10 and 50). In Scenario I, the constant k is chosen equal to 1; whereas, in Scenario II, Scenario III and Scenario IV, it is chosen such that the modified signal-to-noise ratio function $MSN \doteq \mathrm{Var}[\mathrm{E}(Y_i|X_i)] + \max \mathrm{Var}[\mathrm{E}(Y_i|X_i)]/\mathrm{Var}(\varepsilon_i)$ is equal to 4.

For each scenario, we generate 100 datasets composed of a training set with sample size n and a test set T with size N equal to 4000 that are used to estimate the coefficient



(a) (b) (c)

Figure 1: True coefficient function β for Scenario II (a), Scenario III (b) and Scenario IV (c) in the simulation study.

function and to test its predictive performance. This is repeated for three different sample sizes $n = 150, 500, 1000$. As in Lin et al. (2017), we consider the integrated squared error (ISE) to asses the quality of the estimator $\hat{\beta}$ of the coefficient function β. In particular, the ISE over the null region (ISE_0) and the non-null region (ISE_1) are defined as

$$\text{ISE}_0 = \frac{1}{A_0} \int \int_{N(\beta)} \left(\hat{\beta}(s,t) - \beta(s,t) \right)^2 dsdt, \tag{26}$$

and

$$\text{ISE}_1 = \frac{1}{A_1} \int \int_{NN(\beta)} \left(\hat{\beta}(s,t) - \beta(s,t) \right)^2 dsdt, \tag{27}$$

where A_0 and A_1 are the measures of the null ($N(\beta)$) and non-null ($NN(\beta)$) regions, respectively. The ISE_0 and the ISE_1 are indicators of the estimation error of $\hat{\beta}$ over both the null and the non-null regions. Moreover, predictive performance is measured to the prediction mean squared error (PMSE), defined as

$$\text{PMSE} = \frac{1}{N} \sum_{(X,Y) \in T} \int_0^1 \left(Y(t) - \int_0^1 X(s) \hat{\beta}(s,t) \, ds \right)^2 dt, \tag{28}$$

where $\hat{\beta}$ is based on the observations in the training set. The observations in the test set centred by means of the sample mean functions estimated through the training set observations.

The S-LASSO estimator is compared with four different estimators of β that are representative of the state of the art of the FoF linear regression model estimation. The first

two are those proposed by Ramsay and Silverman (2005), where the coefficient function estimator is assumed to be in a finite dimension tensor space with regularization achieved either by choosing the dimension of the tensor space or by introducing roughness penalties. They will be referred to as TRU and SMOOTH estimators, respectively. The third and fourth ones are those proposed by Yao et al. (2005b), based on the functional principal components analysis (referred to as PCA), and by Canale and Vantini (2016), based on a ridge-type penalization (referred to as RIDGE). The TRU, SMOOTH and S-LASSO are computed by using cubic B-splines with evenly space knot sequences. The dimensions of the B-spline sets that generate the tensor product space for the SMOOTH and S-LASSO estimator are both set equal to 60. All the tuning parameters of the five considered estimators are chosen by means of 10-fold cross-validation, viz., the dimension of the tensor basis space for the TRU, the roughness penalties for the SMOOTH, the numbers of retained principal components for the PCA, the penalization parameter for the RIDGE and λ_s, λ_t and λ_L for the S-LASSO. In particular for Scenario I, where the CV is predominating flat, we use the 10-fold cross-validation with the 0.5-standard deviation rule, whereas for the other scenarios the selected parameters are those corresponding to the minimum estimated prediction errors.

The performance of the estimators in terms of ISE_0 is displayed in Figure 2. It is not surprising that the estimation error of β over $N(\beta)$ of the S-LASSO estimator is significantly smaller than those of the other estimators, being the capability of recovering sparseness of β its main feature. In Scenario I, the RIDGE estimator is the only one that performs comparably to the S-LASSO estimator. This is in accordance with the multivariate setting where it is well known that, when the response is independent of the covariates, the ridge estimator is able to shrink all the coefficients towards zero. The TRU, SMOOTH, and PCA estimators have difficulties to correctly identify $N(\beta)$ for all sample sizes. Nevertheless, their performance is very poor at $n = 150$. In Scenario II, the S-LASSO estimator is still the best one to estimate β over $N(\beta)$. However, in this case, the RIDGE estimator performance is unsatisfactory and is mainly caused by the lack of smoothness control that makes the estimator over-rough, especially at small n. Among the competitor estimators, the SMOOTH one has the best performance. In Scenario III, results are similar

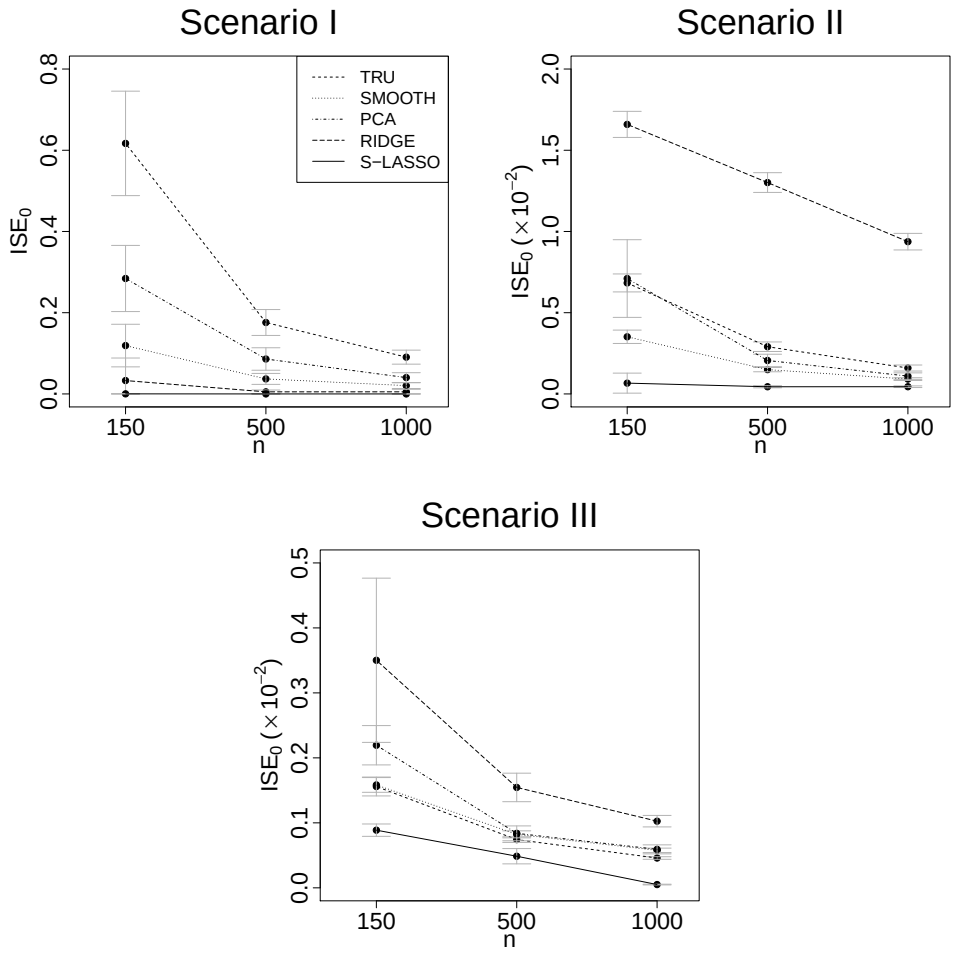


Figure 2: The integrated squared error on the null region (ISE_0) along with $\pm 0.5(\textit{standard error})$ for the TRU(·····), SMOOTH(·····), PCA(-·-·), RIDGE(- - -), and S-LASSO (——) estimators.

to those of Scenario II, even if the TRU estimator appears as the best alternative. Both PCA and RIDGE estimators are not able to successfully recover sparseness of β for $n = 150$. For the former, the cause is the number of observations not sufficient to capture the covariance structure of the data, whereas for the latter, it is due to the excessive roughness of the estimator.

Results in terms of ISE_1 are summarized in Figure 3. It is worth noting that, in this case, as expected the performance of the S-LASSO estimator is generally worse than that of the SMOOTH estimator. In some cases, it is worse than that of the TRU estimator

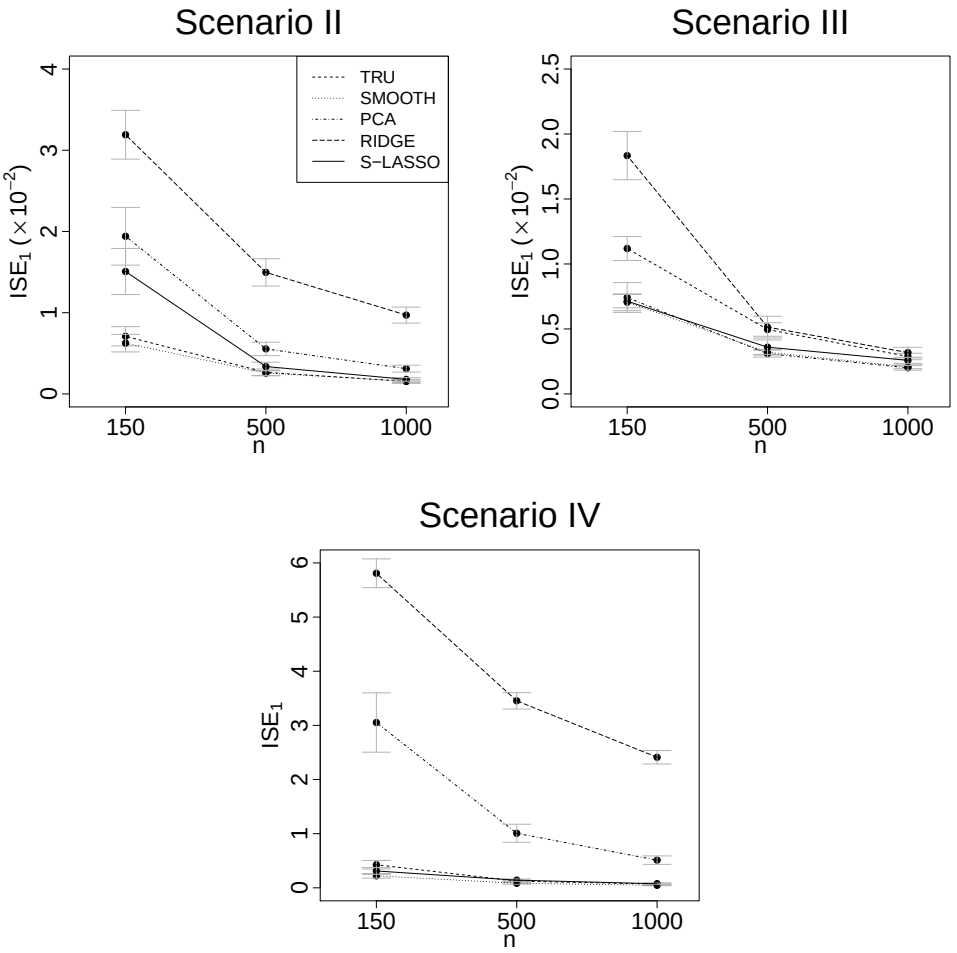


Figure 3: The integrated squared error on the non-null region (ISE_1) along with ± 0.5(*standard error*) for the TRU(·····), SMOOTH(·····), PCA(-·-·), RIDGE(- - -), and S-LASSO (——) estimators.

as well. However, in Scenario II performance differences between the S-LASSO estimator and TRU or SMOOTH estimators become negligible as sample size increases. The PCA and RIDGE estimators are always less efficient. The results are similar for Scenario III, where the performance of the S-LASSO estimator is comparable with that of the SMOOTH estimator. By comparing to the classical LASSO method, the behaviour of the S-LASSO estimator — in terms of ISE_1 — is not surprising. Indeed, it is well known that LASSO method does nice variable selection, even if it tends to overshrink the estimators of the non-null coefficients (Fan et al., 2004; James and Radchenko, 2009). By looking at the

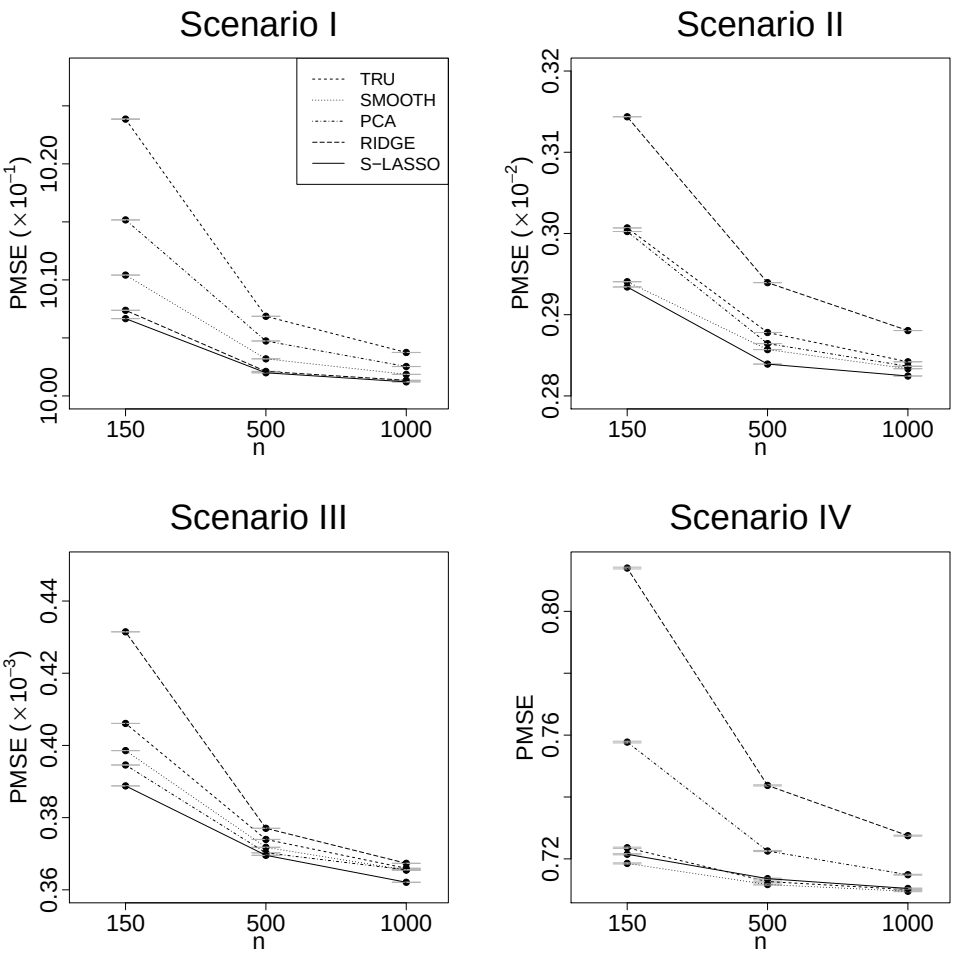


Figure 4: The prediction mean squared error (PMSE) along with ± 0.5(*standard error*) for the TRU(·····), SMOOTH(·····), PCA(-·-·), RIDGE(- - -), and S-LASSO (——) estimators.

result for Scenario II and III, we surmise that this phenomenon arises in the FoF linear regression model as well. Finally, in Scenario IV, where β is always different from zero, the S-LASSO estimator,performs comparably to the SMOOTH (i.e., the S-LASSO estimator with $\lambda_L = 0$). In this case β is not sparse and, thus, the functional LASSO penalty does not help.

Figure 4 shows PMSE averages and corresponding standard errors for all the considered estimators. Since PMSE is strictly related to the ISE_0 and the ISE_1, results are totally consistent with those of Figure 2 and Figure 3. In particular, the S-LASSO estimator outperforms all the competitor ones in favorable scenarios (viz., Scenario I, II, and III), being

the corresponding PMSE lower than that achieved by the other competing estimators. In these scenarios, although the performance of the S-LASSO estimator in terms of ISE_1 is not excellent, the clear superiority in terms of ISE_0 compensates and gives rise to smaller PMSE. Otherwise, for Scenario IV, where the coefficient function is not sparse, the performance of the S-LASSO estimator is very similar to that of the SMOOTH estimator, which is the best one in this case. This is encouraging, because, it proves that the performance of the S-LASSO estimator does not dramatically decline in less favourable scenarios.

In summary, the S-LASSO estimator outperforms the competitors both in terms of estimation error on the null region and prediction accuracy on a new dataset, as well as that it is able to estimate competitively the coefficient function on the non-null region. On the other hand, in order to achieve sparseness, the S-LASSO tends to overshrink the estimator of the coefficient function on the non-null region. This means that, as in the classical setting (James and Radchenko, 2009), there is a trade-off between the ability of recovering sparseness and the estimation accuracy on the non-null region of the final estimator. Moreover, even when the coefficient function is not sparse (Scenario IV), the proposed estimator demonstrates to have both good prediction and estimation performance. This is another key property of the proposed estimator that, encourages practitioners to use the S-LASSO estimator even when there is not prior knowledge about the shape of the coefficient function. Finally, it should be noticed that, in scenarios similar to those analysed, the PCA and RIDGE estimators should not be preferred with respect to the TRU, SMOOTH and S-LASSO ones.

5 Real-Data Examples

In this section, we analyse two real-data examples. We aim to confirm that the S-LASSO estimator has advantages in terms of both prediction accuracy and interpretability, over the SMOOTH estimator, which has been demonstrated in Section 4 to be the best alternative among the competitors. The datasets used in the examples are the *Canadian weather* and *Swedish mortality*. Both are classical benchmark functional data sets thoroughly studied in the literature.

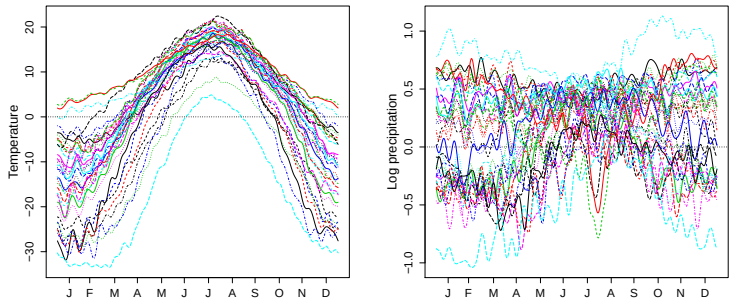


Figure 5: Daily mean temperature and log-daily rainfall profiles at 35 cities in Canada over the year.

5.1 Canadian Weather Data

The Canadian weather data have been studied by Ramsay and Silverman (2005) and Sun et al. (2018). The data set contains the daily mean temperature curves, measured in Celsius degree, and the log-scale of the daily rainfall profiles, measured in millimeter, recorded at 35 cities in Canada. Both temperature and rainfall profiles are obtained by averaging over the years 1960 trough 1994. Figure 5 shows the profiles. The aim is to predict the log-daily rainfall based on the daily temperature using the model reported in Equation (1). Figure 6 shows the S-LASSO and SMOOTH estimates of the coefficient function β. The SMOOTH estimate is obtained using a Fourier basis—to take into account the periodicity of the data—and roughness penalties were chosen by using 10-fold cross-validation over an opportune grid of values. 10-fold cross-validation is used to set the parameters λ_s, λ_t and λ_L as well.

The S-LASSO estimates is roughly zero over large domain portions. In particular, except for values from July through August, it is always zero in summer months (i.e., late June, July, August and September) and in January and February. This suggests in those months rainfalls are not significantly influenced by daily temperature throughout the year. Otherwise, temperature in fall months (i.e., October, November and December) gives strong positive contribution on the daily rainfalls. In other words, the higher (the lower) the temperature in October, November and December, the heavier (the lighter) the precipitations throughout the year. It is interesting that the S-LASSO estimate in spring

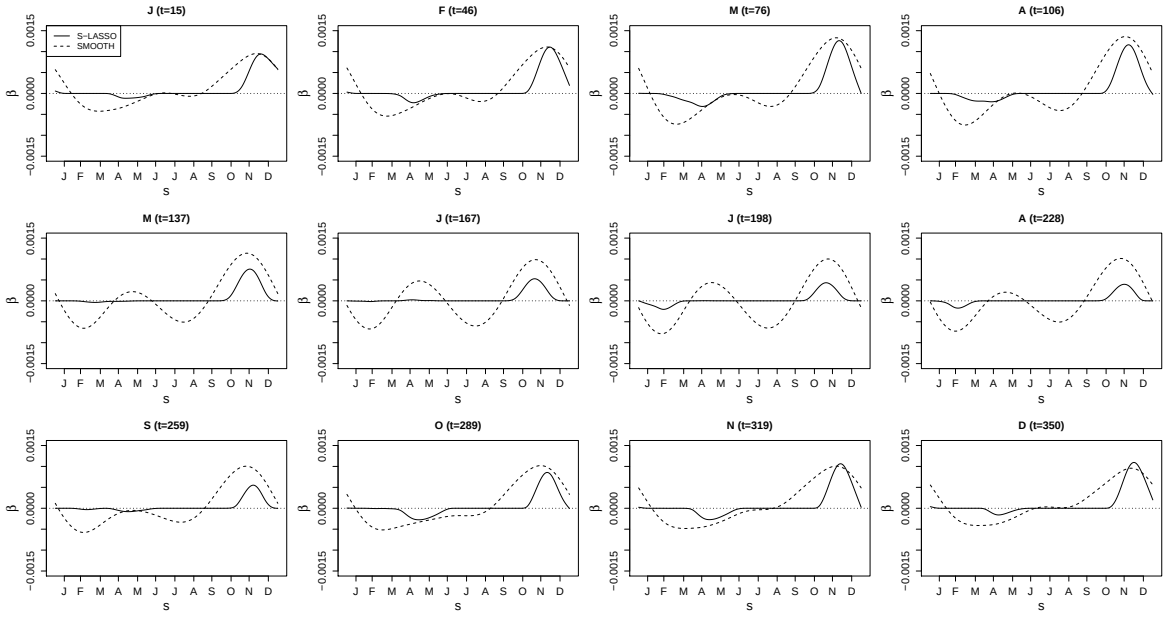
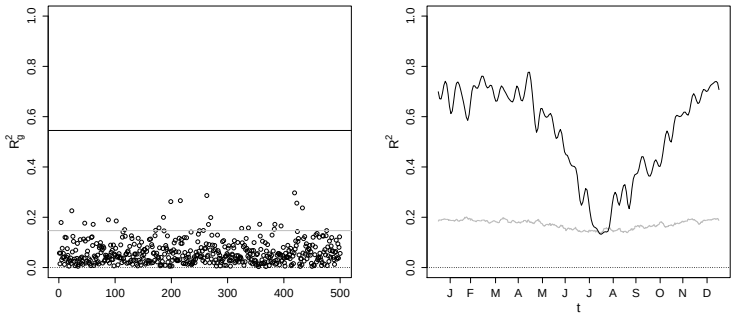


Figure 6: S-LASSO (solid line) and SMOOTH (dashed line) estimates of the coefficient functions β at different months for the Canadian weather data.

months (i.e., March, April and May) is negative for values of t form January through April, and from October through December. This suggests that the higher (the lower) the temperature in the spring the lighter (the heavier) the daily rainfalls from October through April. Finally, it is evidenced a small influence of the temperature in February on precipitations in July and August. It is worth noting that the S-LASSO estimate is more interpretable than the SMOOTH estimates, which does not allow for a straightforward interpretation. Moreover, the S-LASSO estimate appears to have, even if slightly, better prediction performance than the SMOOTH one. Indeed, 10-fold cross-validation mean squared errors are 22.314 and 22.365, respectively.

Finally, we perform two permutation tests to asses the statistical significance of the S-LASSO estimator. The first test is based on the *global functional coefficient of determination* defined as $R^2_g \doteq \int_{\mathcal{T}} \frac{\mathrm{Var}[\mathrm{E}(Y(t)|X)]}{\mathrm{Var}[Y(t)]} dt$ (Horváth and Kokoszka, 2012), with $\mathcal{T} = [0, 365]$. In Figure 7(a) the solid black line indicates the observed R^2_g that is equal to 0.55. The bold points represent 500 R^2_g values obtained by means of random permutations of the response variable. Whereas, the grey line correspond to the 95th sample percentile. All 100 values of R^2 as well as the value of the 95th sample percentile is far below 0.55, which gives a



(a) (b)

Figure 7: For the Canadian weather data, (a) R_g^2 from permuting the response 500 times, where the black line corresponds to the observed R_g^2 and the grey line to the 95th sample percentile; (b) the black line is the observed R^2 and the grey line is the pointwise 95th sample percentile curve.

strong evidence of a significant relationship between rainfalls and temperature, globally. By a second test, we aim to analyse the statistical significance pointwise, i.e., for each $t \in \mathcal{T}$. It is based on the *pointwise functional coefficient of determination* defined as $R^2(t) \doteq \frac{\mathrm{Var}[\mathrm{E}(Y(t)|X)]}{\mathrm{Var}[Y(t)]}$ for $t \in \mathcal{T}$ (Horváth and Kokoszka, 2012). Figure 7(b) shows the observed R^2 (solid black line) along with the pointwise 95th sample percentile curve. The latter has been obtained by means of 500 R^2 values obtained by randomly permuting the response variable. The observed R^2 is far above the 95th sample percentile curve, except for some summer months (viz., July and August). As global conclusion, we can state that the temperature has a large influence on the rainfalls in autumn, winter and spring.

5.2 Swedish Mortality Data

The Swedish mortality data, available from the Human Mortality Database (`http://mortality.org`), are regarded as a very reliable dataset on long-term longitudinal mortalities. In particular, we focus on the log-hazard rate functions of the Swedish females mortality data for year-of-birth cohorts that refer to females born in the years 1751-1894 with ages 0-80. The value of a log-hazard rate function at a given age is the natural logarithm of the ratio of the number of females who died at that age and the number of females alive with the same age. Note that, those data have been analysed also by Chiou and

Müller (2009) and Ramsay et al. (2009). Figure 8 shows the 144 log-hazard functions.

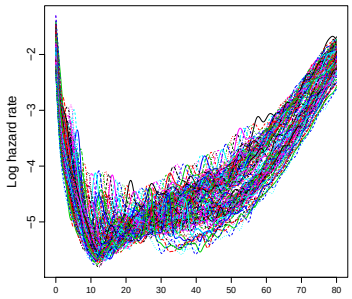


Figure 8: Log-hazard rates as a function of age for Swedish female cohorts born in the years 1751-1894.

The aim of the analysis is to explore the relationship of the log-hazard rate function for a given year with the log-hazard rate curve of the previous year by means of the model reported in Equation (1). Our interest is to identify what features of the log-hazard rate functions for a given year influence the log-hazard rate of the following year.

Figure 9 shows the S-LASSO and SMOOTH estimates of coefficient function β. The unknown parameters to obtain the SMOOTH and S-LASSO estimates are chosen as in the Canadian weather example, but in this case B-splines are used for both estimators. The S-LASSO estimate is zero almost over all the domain except for few regions. In particular, at given t, the S-LASSO estimate is different from zero in an interval located right after that age. This can likely support the conjecture that if an event influences the mortality of the Swedish female at a given age, it impacts on the the death rate below that age born in the following years. Nevertheless, this expected dependence is poorly pointed out by the SMOOTH estimator, where this behaviour is confounded by less meaningful periodic components. It is interesting to note that the S-LASSO estimate at high values of t is slightly different from zero for ages ranging from 40 to 60. This shows that if an event affecting the death rate occurs in that range, the log-hazard functions of the following cohorts will be influenced at high ages (i.e., corresponding to high values of t). On the contrary, the wiggle of the SMOOTH estimate does not allow drawing such conclusions.

Finally, we perform the two permutation test already described in the Canadian weather

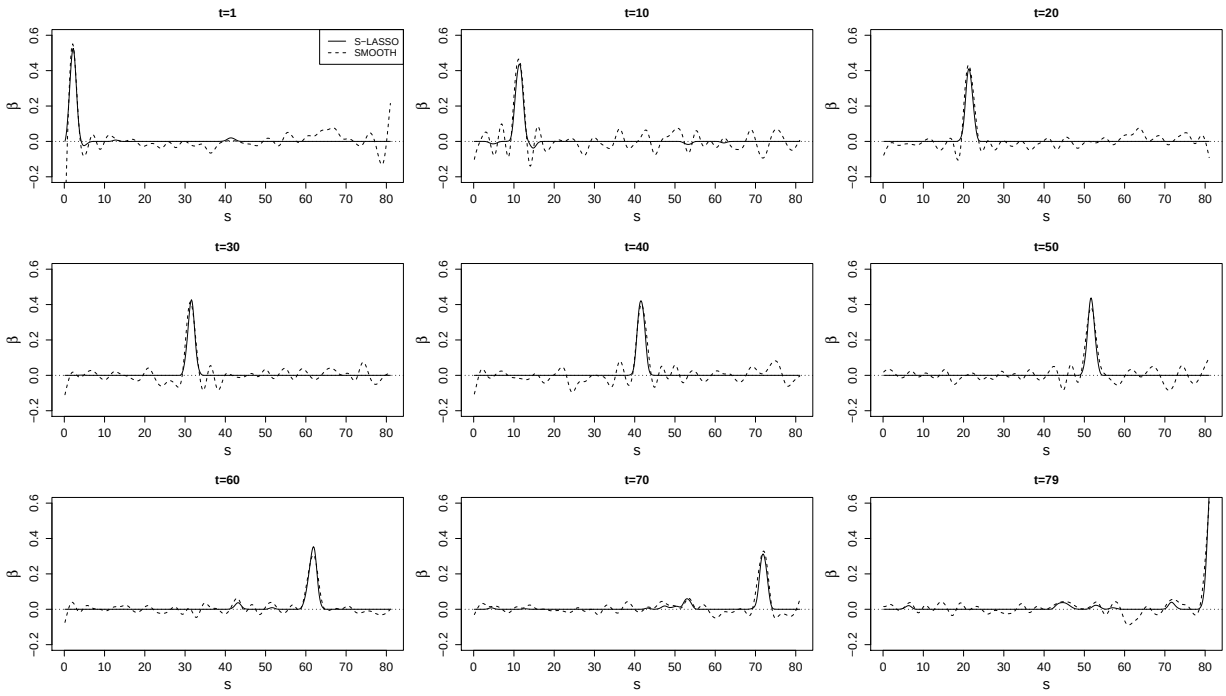
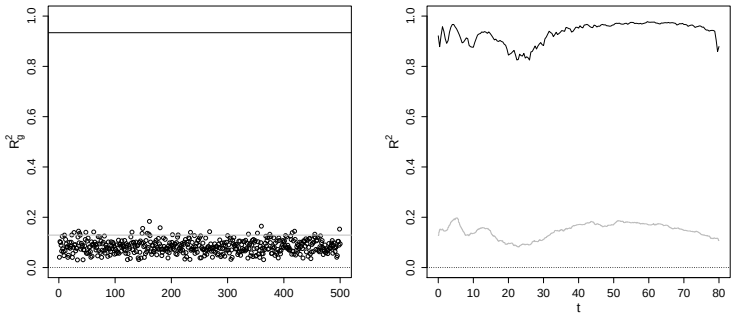


Figure 9: S-LASSO (solid line) and SMOOTH (dashed line) estimates of the coefficient functions β at different values of t for the Swedish mortality data.

data example. Figure 10 shows the results. Both the observed R_g^2 and R^2 are far above the 95th sample percentile (Figure 10(a)) and the pointwise 95th sample percentile curve (Figure 10(b)) respectively. This significantly evidences a relation between two consecutive log-hazard rate functions for all ages.



(a) (b)

Figure 10: For the Swedish mortality data, (a) R^2_g from permuting the response 500 times, where the black line corresponds to the observed R^2_g and the grey line to the 95th sample percentile; (b) the black line is the observed R^2 and the grey line is the pointwise 95th sample percentile curve.

6 Conclusion

The LASSO is one of the most used and popular method to estimate coefficients in classical linear regression models as it ensures both prediction accuracy and interpretability of the phenomenon under study (by simultaneously performing variable selection). The S-LASSO estimator, proposed in this paper, for the coefficient function of a Function-on-Function (FoF) linear regression model. Specifically, it is an extension to the functional setting of the multivariate LASSO estimator. As the latter, the S-LASSO estimator is able both to increase both the prediction accuracy of the estimated model, via continuous shrinking, and the interpretability of the model, by identifying the null region of the regression coefficient, that is the region where the coefficient function is exactly zero.

The S-LASSO estimator is obtained by combining several elements: the functional LASSO penalization, which has the task of shrinking towards zero the estimator on the null region; the B-splines, which are essential to ensure sparsity of the estimator because of the compact support property; and two roughness penalties, which are needed to ensure smoothness of the estimator, also when the number of B-splines escalates to ensure sparsity. We proved that the S-LASSO estimator is both estimation and pointwise sign consistent, i.e., the estimation error in terms of L^2-norm goes to zero in probability and the S-LASSO

estimator has the same sign of the coefficient function with probability one. Moreover, we have shown via an extensive Monte Carlo simulation study that, with regard to other methods that have already appeared in the literature before, the S-LASSO estimator is much more interpretable, on the one hand, and has still good estimation and appealing predictive performance, on the other. Consistently with the behaviour of the classical LASSO estimator (Fan et al., 2004), the S-LASSO estimator is shown to over-shrink the coefficient function over the non-null region.

To the best of the authors knowledge, this is the first work that addresses the issue of interpretability in the FoF linear regression. However, although the functional LASSO penalty produces an estimator with good properties, other penalties, e.g. the *SCAD* (Fan and Li, 2001) and *adaptive LASSO* (Zou, 2006), properly adapted to the functional setting, may guarantee even better performance both in terms of interpretabilty and prediction accuracy.

References

Andrew, G. and J. Gao (2007). Scalable training of l 1-regularized log-linear models. In *Proceedings of the 24th international conference on Machine learning*, pp. 33–40. ACM.

Besse, P. C. and H. Cardot (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics 24*(4), 467–487.

Canale, A. and S. Vantini (2016). Constrained functional time series: Applications to the italian gas market. *International Journal of Forecasting 32*(4), 1340–1351.

Candes, E., T. Tao, et al. (2007). The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics 35*(6), 2313–2351.

Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statistica Sinica*, 571–591.

Chiou, J.-M., Y.-T. Chen, and Y.-F. Yang (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 1571–1596.

Chiou, J.-M. and H.-G. Müller (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association 104*(486), 572–585.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference 147*, 1–23.

De Boor, C., C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor (1978). *A practical guide to splines*, Volume 27. springer-verlag New York.

Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. *The Annals of statistics 32*(2), 407–499.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association 96*(456), 1348–1360.

Fan, J., H. Peng, et al. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics 32*(3), 928–961.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software 33*(1), 1.

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics New York, NY, USA:.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Ivanescu, A. E., A.-M. Staicu, F. Scheipl, and S. Greven (2015). Penalized function-on-function regression. *Computational Statistics 30*(2), 539–568.

James, G. M. and P. Radchenko (2009). A generalized dantzig selector with shrinkage tuning. *Biometrika 96*(2), 323–337.

James, G. M., J. Wang, J. Zhu, et al. (2009). Functional linear regression thats interpretable. *The Annals of Statistics 37*(5A), 2083–2108.

Jia, J. and B. Yu (2010). On model selection consistency of the elastic net when p n. *Statistica Sinica*, 595–611.

Kokoszka, P. and M. Reimherr (2017). *Introduction to functional data analysis*. CRC Press.

Ledoux, M. and M. Talagrand (2013). *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.

Li, Y., T. Hsing, et al. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis 98*(9), 1782–1804.

Lin, Z., J. Cao, L. Wang, and H. Wang (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics 26*(2), 306–318.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application 2*, 321–359.

Nocedal, J. and S. Wright (2006). *Numerical optimization*. Springer Science & Business Media.

Ramsay, J. and B. Silverman (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.

Ramsay, J. O., G. Hooker, and S. Graves (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.

Scheipl, F., A.-M. Staicu, and S. Greven (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics 24*(2), 477–501.

Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge University Press.

Sun, X., P. Du, X. Wang, and P. Ma (2018). Optimal penalized function-on-function regression under a reproducing kernel hilbert space framework. *Journal of the American Statistical Association*, 1–11.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 2873–2903.

Zhou, J., N.-Y. Wang, and N. Wang (2013). Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica 23*(1), 25.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101*(476), 1418–1429.

Zou, H. and H. H. Zhang (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics 37*(4), 1733.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

32/2020 Menafoglio, A.; Sgobba, S.; Lanzano, G.; Pacor, F.
Simulation of seismic ground motion fields via object-oriented spatial statistics: a case study in Northern Italy

31/2020 Bernardi, M.S.; Africa, P.C.; de Falco, C.; Formaggia, L.; Menafoglio, A.; Vantini, S.
On the Use of Interfeometric Synthetic Aperture Radar Data for Monitoring and Forecasting Natural Hazards

30/2020 Massi, M.C., Gasperoni, F., Ieva, F., Paganoni, A.M., Zunino, P., Manzoni, A., Franco, N.R., e
A deep learning approach validates genetic risk factors for late toxicity after prostate cancer radiotherapy in a REQUITE multinational cohort

29/2020 Piersanti, R.; Africa, P.C.; Fedele, M.; Vergara, C.; Dede', L.; Corno, A.F.; Quarteroni, A.
Modeling cardiac muscle fibers in ventricular and atrial electrophysiology simulations

26/2020 Zonca, S.; Antonietti, P.F.; Vergara, C.
A Polygonal Discontinuous Galerkin formulation for contact mechanics in fluid-structure interaction problems

28/2020 Regazzoni, F.; Dedè, L.; Quarteroni, A.
Biophysically detailed mathematical models of multiscale cardiac active mechanics

27/2020 Spreafico, M.; Ieva, F.; Fiocco, M.
Modelling dynamic covariates effect on survival via Functional Data Analysis: application to the MRC BO06 trial in osteosarcoma

Regazzoni, F.; Dedè, L.; Quarteroni, A.
Biophysically detailed mathematical models of multiscale cardiac active mechanics

23/2020 Spreafico, M.; Ieva, F.
Functional modelling of recurrent events on time-to-event processes

20/2020 Almi, S.; Belz, S.; Micheletti, S.; Perotto, S.
A DIMENSION-REDUCTION MODEL FOR BRITTLE FRACTURES ON THIN SHELLS WITH MESH ADAPTIVITY