Nonparametric tests for semiparametric regression models

Federico Ferraccioli[1, 2], Laura M. Sangalli[2], and Livio Finos[3*]

[1]*Dept. of Statistical Science, University of Padova, Italy*
[2]*MOX - Dept. of Mathematics, Politecnico di Milano, Italy*
[3*]*Corresponding author, DPSS, University of Padova, Italy*

Abstract

Semiparametric regression models have received considerable attention over the last decades, because of their flexibility and their good finite sample performances. Here we propose an innovative nonparametric test for the linear part of the models, based on random sign-flipping of an appropriate transformation of the residuals, that exploits a spectral decomposition of the residualizing matrix associated with the nonparametric part of the model. The test can be applied to a vast class of extensively used semiparametric regression models with roughness penalties, with nonparametric components defined over one-dimensional, as well as over multi-dimensional domains, including for instance models based on univariate or multivariate splines. We prove the good asymptotic properties of the proposed test. Moreover, by means of extensive simulation studies, we show the superiority of the proposed test with respect to current parametric alternatives, demonstrating its excellent control of the Type I

error, accompanied by a good power, even in challenging data scenarios, where instead current parametric alternatives fail.

Keywords: functional data analysis, smoothing, roughness penalty, sign-flip

1   Introduction

Semiparametric regression models have a long history in statistics [see, e.g., the textbooks Green and Silverman, 1994, Bickel et al., 1998, Ruppert et al., 2003, and references therein]. Because of their flexibility and versatility, they have been the object of an extensive and still very active literature. In this work, we propose an efficient (conditional) resampling-based test [Pesarin, 2001, Hemerik and Goeman, 2018b, Chung and Romano, 2013] for the linear component in partially linear and semiparametric regression models with roughness penalties. The test can be applied to a vast class of extensively used models, with nonparametric components defined over one-dimensional, as well as over multi-dimensional domains, including manifold domains. This embraces for instance the highly popular semiparametric regression models based on splines [see e.g. Heckman, 1986, Yu and Ruppert, 2002, Wand and Ormerod, 2008, Wang, 2019, and references therein], on thin-plate splines [see, e.g., Wood, 2003], and on spherical splines [Wahba, 1981], as well as semiparametric models based on recent smoothing techniques over two-dimensional (possibly irregularly-shaped or curved) domains, such as soap film smoothing [Wood et al., 2008], bivariate-splines over triangulations [Lai and Schumaker, 2007, Baramidze et al., 2006, Lai et al., 2009, Guillas and Lai, 2010, Lai and Wang, 2013, Wang et al., 2020], and Spatial Regression with Partial Differential Equation regularization (SR-PDE) [see e.g., Sangalli et al., 2013, Azzimonti et al., 2015, Ettinger et al., 2016, Wilhelm et al., 2016, Sangalli, 2021].

Various classical approaches are available to make inference in the context of semiparametric regressions, and different strategies have been proposed to cope with the bias induced

by the roughness penalty. Some possibilities include undersmoothing approaches developed for nonparametric models [see e.g. the review in Hall and Horowitz, 2013], Bayesian approaches [Wahba, 1983, Nychka, 1988, Marra and Wood, 2012] and various corrections of Wald type test statistics, such as the sandwich estimators in Gray [1994] and Yu and Ruppert [2002], and the Speckman's version in Speckman [1988] and Holland [2017]. These approaches might nonetheless have poor performances in the finite sample scenario, due to the effects of the roughness penalty [see, e.g. Maas and Hox, 2004, Freedman, 2006]. In particular, as also evidenced by the simulation studies reported in this work, such tests have a poor control of Type I error.

Here we propose an innovative test for the linear part of semiparametric regression models, based on conditional resampling of a transformation of the residuals. This test, unlike other proposals, allows to overcome the problem of dependence in the residuals that is particularly strong in semiparametric models. Some approaches proposed in the context of classical regression models, such as those in Huh and Jhun [2001] and Kherad-Pajouh and Renaud [2010], derive transformed residuals from spectral decomposition of the residualizing matrix that projects into the residual space. In the setting considered by these authors, the conditional distribution of the test statistic can be defined on the basis of permutations [see, e.g., Pesarin, 2001, Chung and Romano, 2013, Pauly et al., 2015, Winkler et al., 2014], rotations [Solari et al., 2014] or sign-flips [Hemerik et al., 2020] of such transformed residuals. These approaches are nonetheless not valid in this context, since the residualizing matrix is not idempotent in the case of penalized regression models. Because of this, the transformed residuals are not spherical (i.e., they are not homoscedastic and independent) and the standard permutation, rotation or sign-flip procedures become invalid in our context. To overcome this problem we here study a conditional sign-flip procedure, named eigen sign-flip test, that preserves the finite sample covariance structure of the residuals, hence ensuring asymptotically exactness of the derived test. This idea has

3

been explored in Ferraccioli et al. [2021], restricted to a specific case of SR-PDE model. The current work addresses instead the broad spectrum of highly popular semiparametric regression models mentioned above. Moreover, we study in detail the asymptotic properties of the test. In particular, we prove the asymptotic exactness of the test and derive similar results for interval hypothesis and confidence intervals. Some of the obtained results leverage on the asymptotic properties of the estimator of the nonparametric part of the model. Such properties in turn depend on conditions that are model-specific, since they depend for instance on the dimension and geometry of the domain over which the nonparametric term is defined, on the roughness term being considered, on the type of basis, etc. In the present work, we hence define assumptions that are general enough to cover a variety of semiparametric regression models, and refer the reader to other works for the appropriate specifications of such assumptions for the specific model being considered (e.g., to Claeskens et al. [2009] for univariate penalized splines estimators, to Holland [2017] for multivariate penalized splines estimators, to Xiao [2019] for general penalized splines and to Arnone et al. [2021] for SR-PDE.)

The paper is organized as follows. In Section 2 we briefly review the semiparametric penalized regression framework, outlining the forms of the associated discrete estimators. In Section 3 we recall some classical parametric approaches for inference on the linear part of a semiparametric regression model and summarize the properties of the score test statistic in this context. In Section 4 we present the Eigen-sign flip test and describe its theoretical and asymptotic properties. In Section 5 we compare our proposal to more classical parametric approaches in extensive simulation studies. In Section 6 we present an application to the study of human development in Nigeria. Finally, some discussions and possible directions for future research are outlined in Section 7.

4

2 Semiparametric regression

Let $y_i \in \mathbb{R}$ be the value of the variable of interest observed in correspondence of covariates $\mathbf{x}_i \in \mathbb{R}^q$ and of $\mathbf{p}_i \in \Omega \subseteq \mathbb{R}^d$, $d \geq 1$. We consider the semiparametric model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + f(\mathbf{p}_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^q$ is the vector of regression parameters, f is a real-valued smooth function on Ω, and ϵ_i are i.i.d. random errors with $\mathbb{E}(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma^2$.

The interest is to estimate both the linear coefficients $\boldsymbol{\beta}$ and the nonparametric component f. However, the estimation of $(\boldsymbol{\beta}, f)$ in model (1) via maximum likelihood is usually inappropriate or infeasible, due to the infinite-dimensionality of the nonparametric component f. To avoid this problem, some type of roughness penalty can be imposed, in order to reduce the space of possible solutions. In general, the resulting penalized likelihood estimators for $\boldsymbol{\beta}$ and f are the solution of the minimization problem

$$\underset{\boldsymbol{\beta}, f}{\mathrm{argmin}} \sum_{i=1}^{n} (y_i - f(\mathbf{p}_i) - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \lambda_n \mathcal{P}(f), \tag{2}$$

where $\mathcal{P}(\cdot)$ is some type of roughness penalty. Depending on the assumptions on the domain $\Omega \subseteq \mathbb{R}^d$, on the dimension d, and on the required smoothness of the function f, various proposals for $\mathcal{P}(f)$ have been considered in the literature, and different discretization procedures have been adopted to reduce the infinite-dimensional estimation problem (2) to a finite dimensional one. For instance, for $d = 1$ and Ω an interval of the real line, model (1)-(2) can involve the classical and extensively used O'Sullivan splines [O'Sullivan, 1986, Heckman, 1986, Yu and Ruppert, 2002, Wand and Ormerod, 2008], whose penalty is the integrated squared derivative of some order, and can for instance rely on B-spline bases. When Ω is the real plane, it is possible to use thin-plate splines [see, e.g., Duchon, 1977, Wahba, 1990, Wood, 2003], which involve as penalty the so-called thin-plate energy. Moreover, various recent techniques target two-dimensional bounded planar domains $\Omega \subset \mathbb{R}^2$,

including: soap-film smoothing [Wood et al., 2008], that considers a penalty involving the Laplacian of f; bivariate-splines over triangulations [Lai and Schumaker, 2007, Guillas and Lai, 2010, Lai and Wang, 2013], whose regularizing term may include high-order derivatives; SR-PDE [Sangalli et al., 2013, Azzimonti et al., 2015], where the regularizing term can involve general second-order partial differential equations, and the estimation problem is discretized via finite element bases [Sangalli et al., 2013, Azzimonti et al., 2015] or advanced spline bases [Wilhelm et al., 2016]. Some of these techniques also permit the constructions of semiparametric models over spherical domains [Wahba, 1981, Baramidze et al., 2006, Lai et al., 2009] and general surface domains [Ettinger et al., 2016, Wilhelm et al., 2016].

2.1 Discrete estimators

The estimation of model (1) usually involves the representation of the nonparametric component f through some type of basis expansion, depending on the penalization being considered. Let $\Psi \in \mathbb{R}^n \times \mathbb{R}^K$ be the matrix of the evaluations of the K basis functions $\psi_1, \ldots, \psi_K$ at the n data locations $\mathbf{p}_1, \ldots, \mathbf{p}_n$, that is,

$$\Psi = \begin{bmatrix} \psi_1(\mathbf{p}_1) & \ldots & \psi_K(\mathbf{p}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{p}_n) & \ldots & \psi_K(\mathbf{p}_n) \end{bmatrix}.$$

Then, we write $(f(\mathbf{p}_1), \ldots, f(\mathbf{p}_n))^\top = \Psi \boldsymbol{\gamma}$ for some vector of coefficients $\boldsymbol{\gamma} \in \mathbb{R}^K$. Moreover, let P denote the $K \times K$ positive semidefinite matrix representing the discretization of the penalty $\mathcal{P}(\cdot)$. Finally, set $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and denote by $X \in \mathbb{R}^n \times \mathbb{R}^q$ the design matrix, whose i-th row is given by $\mathbf{x}_i$. The estimation problem (2) is therefore discretized as

$$\underset{\boldsymbol{\beta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \big\{ (\mathbf{y} - X\boldsymbol{\beta} - \Psi\boldsymbol{\gamma})^\top (\mathbf{y} - X\boldsymbol{\beta} - \Psi\boldsymbol{\gamma}) + \lambda_n \boldsymbol{\gamma}^\top P \boldsymbol{\gamma} \big\}. \tag{3}$$

The solution to (3) is uniquely determined by the normal equations

$$X^\top(\mathbf{y} - X\boldsymbol{\beta} - \Psi\boldsymbol{\gamma}) = 0, \tag{4}$$

$$(\Psi^\top\Psi + \lambda P)\boldsymbol{\gamma} - \Psi^\top(\mathbf{y} - X\boldsymbol{\beta}) = 0. \tag{5}$$

Setting

$$\begin{aligned} Q &= I_n - X(X^\top X)^{-1}X^\top, \\ \Lambda &= I_n - \Psi(\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top, \end{aligned} \tag{6}$$

the explicit form of the estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are, respectively,

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X^\top\Lambda X)^{-1}X^\top\Lambda\mathbf{y}, \\ \hat{\boldsymbol{\gamma}} &= (\Psi^\top Q\Psi + \lambda P)^{-1}\Psi^\top Q\mathbf{y}, \end{aligned} \tag{7}$$

or equivalently

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X^\top X)^{-1}X^\top(\mathbf{y} - \Psi\hat{\boldsymbol{\gamma}}), \\ \hat{\boldsymbol{\gamma}} &= (\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top(\mathbf{y} - X\hat{\boldsymbol{\beta}}). \end{aligned} \tag{8}$$

3   Inference on $\boldsymbol{\beta}$

In semiparametric regression, a natural question is whether the covariates X have an effect on the variable of interest. We are thus interested in the system of hypotheses

$$\text{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{versus} \qquad \text{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0. \tag{9}$$

A standard approach to verify (9) is to use a Wald type test [see, e.g., Schervish, 2012], based on the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. The study of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, in semiparametric regression models, has been tackled by a number of works. See for instance

Heckman [1986], Yu and Ruppert [2002], Li and Ruppert [2008], Holland [2017], Xiao [2019], Yu et al. [2019], Wang et al. [2020] for semiparametric models based on univariate and bivariate splines.

The parametric Wald type test may nonetheless have poor performances in small sample scenarios, due to the overestimation of the variance of the test statistic, induced by the penalization. A number of corrections to Wald type test have been proposed to avoid this issue, such as the sandwich estimators in Gray [1994] and Yu and Ruppert [2002] and the Speckman's version in Speckman [1988] and Holland [2017]. Nonetheless, these approaches can only partially solve the problem, and may lead to a poor control of the Type I error, especially when a strong temporal/spatial structure in the covariates is present, as indicated by the simulations carried on in Section 5.

In the Section 4 we introduce an innovative nonparametric alternative for testing on $\boldsymbol{\beta}$. Such proposal is based on the score statistic. For this reason, in the remainder of this section we review the properties of the score statistic in the context of penalized semiparametric regression. The proposed method does not rely on the estimation of the Fisher information matrix to define the null distribution, which is implicitly recovered by an appropriate nonparametric resampling procedure, as described in Section 4.

3.1 Properties of the score statistic in penalized semiparametric regression

We first study the distributional properties of the score statistic, which constitute the base of the nonparametric test defined in Section 4. Using the normal equation (4), we can define the classical score test statistic

$$X^\top(\mathbf{y} - X\boldsymbol{\beta}_0 - \Psi\boldsymbol{\gamma}). \tag{10}$$

Since $\boldsymbol{\gamma}$ is unknown, we can use the plug-in $\hat{\boldsymbol{\gamma}}$. Substituting $\hat{\boldsymbol{\gamma}}$ in expression (10), we define the test statistic T as

$$
\begin{aligned}
T &= n^{-1/2} X^\top (\mathbf{y} - \Psi\hat{\boldsymbol{\gamma}} - X\boldsymbol{\beta}_0) \\
&= n^{-1/2} X^\top (\mathbf{y} - \Psi(\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top(\mathbf{y} - X\boldsymbol{\beta}_0) - X\boldsymbol{\beta}_0) \\
&= n^{-1/2} X^\top ((I - \Psi(\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top)\mathbf{y} - (I - \Psi(\Psi^\top\Psi + \lambda P)^{-1}\Psi^\top)X\boldsymbol{\beta}_0) \\
&= n^{-1/2} X^\top \Lambda (\mathbf{y} - X\boldsymbol{\beta}_0) = n^{-1/2} X^\top \Lambda \mathbf{r}.
\end{aligned}
\tag{11}
$$

with $\mathbf{r} = \mathbf{y} - X\boldsymbol{\beta}_0$. We make the following assumption:

(A1) For n large enough, the matrix $\Psi^\top\Psi$ is positive definite.

Assumption (A1) is quite general; its specification depends on the basis considered. In particular, this specification usually involves conditions on the nodes of the basis and their position with respect to the design points $\mathbf{p}_1, \dots, \mathbf{p}_n$. More specifically, it involves the type of basis, the rate at which the number of bases K grows with n, the minimum distance between the nodes, and the density of the design points inside the domain. For instance, in the case of univariate penalized splines estimators, (A1) follows from Assumptions 1–3 in Claeskens et al. [2009]. In the case of multivariate penalized splines estimators, it follows from Assumptions 1–2 in Holland [2017]. In the case of SR-PDE, it follows from Assumptions 3–5 in Arnone et al. [2021].

Here we consider the case of fixed designs, thus implicitly conditioning on the sample points and the covariates. Similar results can be obtained in the random design scenario, by introducing further assumptions on the distribution of the design points and covariates (e.g. that the covariates are realizations of continuous processes on Ω).

Under (A1), we can consider the Demmler and Reinsch [1975] decomposition

$$
(\Psi^\top\Psi)^{-1/2} P (\Psi^\top\Psi)^{-1/2} = U \operatorname{diag}(\rho) U^\top,
$$

where U is the matrix of eigenvectors, and ρ is the corresponding vector of eigenvalues $\{\rho_k\}_{k=1}^K$ [see Eubank, 1999, for details]. Let us also denote $A = \Psi(\Psi^\top\Psi)^{-1/2}U$. Note that this matrix is semi-orthogonal, i.e., $A^\top A = I_K$ and $AA^\top = \Psi(\Psi^\top\Psi)^{-1}\Psi^\top$. Following Demmler and Reinsch [1975], we can rewrite the matrix Λ in (6) as

$$\Lambda = I_n - A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top.$$

Using this decomposition, we can now study the behaviour of the bias of the test statistic T, in terms of the eigenvalues ρ_k.

Lemma 3.1. *Assume (A1) and let $\tilde{\tilde{X}} = A^\top X$ and $\tilde{\tilde{\gamma}} = A^\top\Psi\gamma$. Let also $\tilde{\tilde{\mathbf{x}}}_i$ be the q-dimensional vectors corresponding to the rows of $\tilde{\tilde{X}}$, and $\tilde{\tilde{\gamma}}_i$ be the elements of the vector $\tilde{\tilde{\gamma}}$. Under the null hypothesis (9), the bias $\mathbf{b}_\lambda$ of T is*

$$\mathbf{b}_\lambda \leq n^{-1/2}\lambda\max_i(|\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i|)\sum_{i=1}^K\rho_i, \tag{12}$$

where the inequality is considered element-wise.

Proof. Denote by $\boldsymbol{\epsilon}$ the n-dimensional vector of i.i.d. residuals. Under the null hypothesis, we have

$$\begin{aligned}
\mathbb{E}(T) &= n^{-1/2}\mathbb{E}(X^\top\Lambda\mathbf{r}) \\
&= n^{-1/2}\mathbb{E}(X^\top\Lambda(\mathbf{y} - X\boldsymbol{\beta}_0)) \\
&= n^{-1/2}X^\top\Lambda\mathbb{E}(\Psi\gamma + X\boldsymbol{\beta}_0 - X\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}) \\
&= n^{-1/2}X^\top\Lambda\Psi\gamma + n^{-1/2}X^\top\Lambda\mathbb{E}(\boldsymbol{\epsilon}) \\
&= n^{-1/2}X^\top\Lambda\Psi\gamma = \mathbf{b}_\lambda,
\end{aligned}$$

since the term $\mathbb{E}(\boldsymbol{\epsilon})$ is zero by assumption. Using the decomposition in (3.1), it follows that

$$\begin{aligned}
\Lambda &= I_n - A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top \\
&= AA^\top - AA^\top + I_n - A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top \\
&= A(I_K - (I_K + \lambda\mathrm{diag}(\rho))^{-1})A^\top - AA^\top + I_n. \tag{13}
\end{aligned}$$

10

Substituting (13) in $\mathbf{b}_\lambda$, we obtain

$$\mathbf{b}_\lambda = n^{-1/2}[X^\top(A(I_K - (I_K + \lambda\mathrm{diag}(\rho))^{-1})A^\top - AA^\top + I_n)\Psi\boldsymbol{\gamma}]$$
$$= n^{-1/2}[X^\top A(I_K - (I_K + \lambda\mathrm{diag}(\rho))^{-1})A^\top\Psi\boldsymbol{\gamma} - X^\top AA^\top\Psi\boldsymbol{\gamma} + X^\top\Psi\boldsymbol{\gamma}]$$
$$= n^{-1/2}[X^\top A(I_K - (I_K + \lambda\mathrm{diag}(\rho))^{-1})A^\top\Psi\boldsymbol{\gamma} - X^\top\Psi(\Psi^\top\Psi)^{-1}\Psi^\top\Psi\boldsymbol{\gamma} + X^\top\Psi\boldsymbol{\gamma}]$$
$$= n^{-1/2}X^\top A(I_K - (I_K + \lambda\mathrm{diag}(\rho))^{-1})A^\top\Psi\boldsymbol{\gamma}.$$

Using the notation $\tilde{\tilde{X}} = A^\top X$ and $\tilde{\tilde{\boldsymbol{\gamma}}} = A^\top\Psi\boldsymbol{\gamma}$, the bias can therefore be rewritten as

$$\mathbf{b}_\lambda = n^{-1/2}\sum_{i=1}^{K}\left(1 - \frac{1}{1 + \lambda\rho_i}\right)\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i = n^{-1/2}\sum_{i=1}^{K}\left(\frac{\lambda\rho_i}{1 + \lambda\rho_i}\right)\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i, \tag{14}$$

where $\tilde{\tilde{\mathbf{x}}}_i$ are the q-dimensional vectors corresponding to the rows of $\tilde{\tilde{X}}$, and $\tilde{\tilde{\gamma}}_i$ the elements of the vector $\tilde{\tilde{\boldsymbol{\gamma}}}$. Equation (14) highlights that the bias is a sum of K contributions, weighted by the eigenvalues ρ_k, and moderated by λ. Since the function $x/(1 + x) < x$, for $x > 0$, we can bound the bias as follows

$$\mathbf{b}_\lambda = n^{-1/2}\sum_{i=1}^{K}\left(\frac{\lambda\rho_i}{1 + \lambda\rho_i}\right)\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i \leq n^{-1/2}\lambda\sum_{i=1}^{K}\rho_i|\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i| \leq \max_i(|\tilde{\tilde{\mathbf{x}}}_i\tilde{\tilde{\gamma}}_i|)n^{-1/2}\lambda\sum_{i=1}^{K}\rho_i.$$

$\square$

The expression (14) highlights how the bias depends on the chosen penalization through the eigenvalues ρ_k. We finally make the following assumption.

(A2) The smoothing parameter $\lambda = \lambda_n$ is chosen so that $\lambda\sum_{i=1}^{K}\rho_i = o(1)$.

Thanks to Lemma 3.1, assumption (A2) implies the asymptotic unbiasedness of score statistic T, since f is a continuous function on the bounded domain Ω and the covariates are realizations of a continuous process on Ω. This is a standard assumption when studying the asymptotic properties of semiparametric and nonparametric penalized regression models. Likewise for Assumption (A1), also Assumption (A2) needs to be specified depending on

11

the penalty and basis considered. Indeed, Assumptions (A1)–(A2) are intentionally left quite general to embrace various semiparametric models; moreover, the precise rates of convergence are not of direct interest in this work. Theorem 1 in Claeskens et al. [2009] gives for instance the appropriate rates for λ in the case of univariate penalized splines estimators, Theorem 3 in Holland [2017] gives it for multivariate penalized spline estimators, while Lemma 3 in Arnone et al. [2021] gives it for SR-PDE estimators.

We can now state the main result for the asymptotic distribution of the test statistic T.

Theorem 3.2. *Let $\nu = \sigma^2 X^\top \Lambda^2 X$. Under the assumptions (A1)–(A2), the test statistic T in (11) is asymptotically normal under the null hypothesis (9), with*

$$\sqrt{n}\nu^{-1/2} T \stackrel{.}{\sim} \mathcal{N}(\mathbf{0}, \mathbb{I}_q).$$

Proof. We know that

$$\mathbb{E}(T) = n^{-1/2}\mathbb{E}(X^\top \Lambda \mathbf{r}) = n^{-1/2}X^\top \Lambda \Psi \boldsymbol{\gamma} + n^{-1/2}X^\top \Lambda \mathbb{E}(\boldsymbol{\epsilon})$$

$$= n^{-1/2}\mathbf{b}_\lambda + n^{-1/2}\sum_{i=1}^{n}[X^\top \Lambda]_i \mathbb{E}(\epsilon_i)$$

where the notation $[X^\top \Lambda]_i$ is used to indicate the i-th column of the $q \times n$ matrix $X^\top \Lambda$. Under assumption (A2), it follows from (12) that the bias $\mathbf{b}_\lambda$ is asymptotically zero. The expected value $\mathbb{E}(T)$ is therefore asymptotically zero. For the variance, under the null hypothesis we have

$$\mathrm{Var}(T) = \mathrm{Var}(n^{-1/2}X^\top \Lambda \mathbf{r}) = n^{-1}X^\top \Lambda \mathrm{Var}(\Psi \boldsymbol{\gamma} + X\boldsymbol{\beta}_0 + \boldsymbol{\epsilon} - X\boldsymbol{\beta}_0)\Lambda X$$

$$= n^{-1}X^\top \Lambda \mathrm{Var}(\boldsymbol{\epsilon})\Lambda X = n^{-1}\sigma^2 X^\top \Lambda^2 X.$$

Substituting the expression of Λ from equation (13) in the previous expression, we obtain

$$
\begin{aligned}
\mathrm{Var}(T) &= n^{-1}\sigma^2 X^\top (I_n - A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top)^2 X \\
&= n^{-1}\sigma^2 X^\top (I_n + A(I_K + \lambda\mathrm{diag}(\rho))^{-2}A^\top - 2A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top)X \\
&= n^{-1}\sigma^2 (X^\top X + X^\top A(I_K + \lambda\mathrm{diag}(\rho))^{-2}A^\top X - 2X^\top A(I_K + \lambda\mathrm{diag}(\rho))^{-1}A^\top X) \\
&= n^{-1}\sigma^2 X^\top X + n^{-1}\sigma^2 (\tilde{\tilde{X}}^\top (I_K + \lambda\mathrm{diag}(\rho))^{-2}\tilde{\tilde{X}} - 2\tilde{\tilde{X}}^\top (I_K + \lambda\mathrm{diag}(\rho))^{-1}\tilde{\tilde{X}}) \\
&= n^{-1}\sigma^2 X^\top X + n^{-1}\sigma^2 \tilde{\tilde{X}}^\top ((I_K + \lambda\mathrm{diag}(\rho))^{-2} - 2(I_K + \lambda\mathrm{diag}(\rho))^{-1})\tilde{\tilde{X}}
\end{aligned}
$$

Using the notation $\tilde{\tilde{X}} = A^\top X$ and completing the square in the second term, we hence get

$$
\begin{aligned}
\mathrm{Var}(T) &= n^{-1}\sigma^2 X^\top (I_n - AA^\top)X + n^{-1}\sigma^2 \tilde{\tilde{X}}^\top ((I_K + \lambda\mathrm{diag}(\rho))^{-1} - I_K)^2 \tilde{\tilde{X}} \\
&= n^{-1}\sigma^2 X^\top X - n^{-1}\sigma^2 X^\top AA^\top X + n^{-1}\sigma^2 \tilde{\tilde{X}}^\top ((I_K + \lambda\mathrm{diag}(\rho))^{-1} - I_K)^2 \tilde{\tilde{X}} \\
&= n^{-1}\sigma^2 X^\top (I_n - AA^\top)X + n^{-1}\sigma^2 \sum_{i=1}^{K} \left(\frac{\lambda\rho_i}{1 + \lambda\rho_i} \right)^2 \tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top,
\end{aligned}
\tag{15}
$$

where $\tilde{\tilde{\mathbf{x}}}_i$ are the q-dimensional vectors corresponding to the rows of $\tilde{\tilde{X}}$. Note that the first term does not depend on λ. As for the second term, since $x^2/(1+x)^2 < x^2$, for $x > 0$, we have

$$
\begin{aligned}
\sigma^2 n^{-1} \sum_{i=1}^{K} \left(\frac{\lambda\rho_i}{1 + \lambda\rho_i} \right)^2 \tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top &\leq \sigma^2 n^{-1}\lambda^2 \sum_{i=1}^{K} \rho_i^2 |\tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top| \\
&\leq \sigma^2 \max(|\tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top|) n^{-1}\lambda^2 \sum_{i=1}^{K} \rho_i^2 \\
&\leq \sigma^2 \max(|\tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top|) n^{-1}\lambda^2 \left(\sum_{i=1}^{K} \rho_i \right)^2 \\
&= \sigma^2 \max(|\tilde{\tilde{\mathbf{x}}}_i \tilde{\tilde{\mathbf{x}}}_i^\top|) \left(n^{-1/2}\lambda \sum_{i=1}^{K} \rho_i \right)^2,
\end{aligned}
$$

where the maximum is taken element-wise. Therefore, for n large enough (since the co-variates are realizations of a continuous process on Ω), assumption (A2) implies that the

13

second term in (15) vanishes faster than the first term. Concerning the first term in (15), with a similar argument it is easy to check that the matrix $AA^\top$ is idempotent with rank K. Thus, it admits the spectral decomposition $AA^\top = U\mathrm{diag}(1,\ldots,1,0,\ldots,0)U^\top$, with the first K non-null eigenvalues equal to 1. The term $X^\top(I_n - AA^\top)X$ is therefore the sum of $n-K$ components with bounded variance, since the covariates are realizations of a continuous process on Ω, thus the Feller condition is satisfied. It follows from the central limit theorem [see, e.g., Van der Vaart, 2000] that the test statistic T is also asymptotically normal. $\qquad\square$

4 Eigen sign-flip test for the linear component in penalized semiparametric regression models

In the classical linear regression case, under the standard assumption of i.i.d. random noise, the score statistics can also be viewed as a sum of n contributions that have asymptotically zero mean, under the null hypothesis H_0 (9). This information can be used to derive the null distribution of the test statistic, without the need of a direct estimation of the Fisher information. In the context of semiparametric regression, instead, a first naive attempt to derive the distribution of the test statistic can be made by random permutations (or sign-flips) of the contributions of the score [Winkler et al., 2014, Hemerik et al., 2020]. This approach, attempted in Ferraccioli [2020] for a simple type of SR-PDE model [Sangalli et al., 2013], might nonetheless be not optimal in the semiparametric regression setting. The reason for this lies in the fact that naive permutation does not account for the correlation between residuals, nor for the bias of the estimates, which is inherent to semiparametric models. To solve this issue, always considering a special case of SR-PDE model, Ferraccioli et al. [2021] defines a new test statistic, that leverages on the spectral decomposition of the matrix Λ, leading to the definition of the eigen sign-flip test.

14

We here defined the eigen sign-flip test on $\boldsymbol{\beta}$ for a general forms of penalized semi-parametric regression models. We study the properties of the test, proving its asymptotic distribution. A thorough discussion on the nature of the proposed test in given in Section 4.2.

Definition 1 (Eigen sign-flip test). *Let us consider the singular value decomposition $\Lambda = VDV^\top$. Set $\Pi = \text{diag}(\pi_1, \ldots, \pi_n)$, where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ is a random vector uniformly distributed in $\{-1, 1\}^n$. Let us also define the n-dimensional vectors $\tilde{X} = D^{1/2}V^\top X$ and $\tilde{\mathbf{r}} = D^{1/2}V^\top \mathbf{r} = D^{1/2}V^\top(\mathbf{y} - X\boldsymbol{\beta}_0)$. The eigen sign-flip statistics is defined as*

$$T_\Pi = n^{-1/2}X^\top V D^{1/2}\Pi D^{1/2}V^\top \Lambda(\mathbf{y} - X\boldsymbol{\beta}_0) = n^{-1/2}\tilde{X}^\top \Pi \tilde{\mathbf{r}}.$$

Note that the observed statistic $T = T_I$ corresponds to the case where $\pi_i = 1$, $i = 1, \ldots, n$. As standard in permutational approaches, the component-wise p-values are thus computed as the rank of T_I with respect to a sample of M sign-flips $\boldsymbol{\pi}$, divided by M [see, e.g., Pesarin, 2001].

4.1 Asymptotic properties of the eigen sign-flip test

We now study the asymptotic properties of the test statistic T_Π in Definition 1. We first show that the asymptotic distribution of the test statistic T_Π is the same as T_I. We then show that the eigen sign-flip test is asymptotically exact.

Theorem 4.1. *Let $\nu = \sigma^2 X^\top \Lambda^2 X$. Under the assumptions (A1)–(A2), for any given Π, the distribution of T_Π is asymptotically normal, with*

$$\sqrt{n}\nu^{-1/2}T_\Pi \overset{\cdot}{\sim} \mathcal{N}(\mathbf{0}, \mathbb{I}_q).$$

15

Proof. For the expected value, under the null hypothesis we have

$$
\begin{aligned}
\mathbb{E}(T_\Pi) &= n^{-1/2}\mathbb{E}(\tilde{X}^\top \Pi \tilde{\mathbf{r}}) \\
&= n^{-1/2}\mathbb{E}(X^\top V \Pi V^\top \Lambda(\mathbf{y} - X\boldsymbol{\beta}_0)) \\
&= n^{-1/2}X^\top V \Pi V^\top \Lambda\mathbb{E}(\mathbf{y} - X\boldsymbol{\beta}_0) \\
&= n^{-1/2}X^\top V \Pi V^\top \Lambda\Psi\boldsymbol{\gamma}.
\end{aligned}
$$

Following the same reasoning of the proof of Theorem 3.2, but with the quantity $V\Pi V^\top X$ in place of X, we can show that the expected value of T_Π is asymptotically zero.

As for the variance, under the null hypothesis we have

$$
\begin{aligned}
\mathrm{Var}(T_\Pi) &= \mathrm{Var}(n^{-1/2}\tilde{X}^\top \Pi \tilde{\mathbf{r}}) \\
&= \mathrm{Var}(n^{-1/2}X^\top V \Pi D V^\top(\mathbf{y} - X\boldsymbol{\beta}_0)) \\
&= n^{-1}X^\top V D \Pi V^\top \mathrm{Var}(\boldsymbol{\epsilon}) V \Pi D V^\top X \\
&= n^{-1}\sigma^2 X^\top \Lambda^2 X = n^{-1}\nu.
\end{aligned}
$$

It follows from the central limit theorem [Van der Vaart, 2000], that the test statistic T_Π is also asymptotically normal. $\qquad\square$

Remark 1. *Note that the bias in the mean of the test statistic is intrinsic in the regularization approach, and cannot be avoided in the finite sample scenario. Because of this bias, we are only able to reach asymptotically exact results.*

Remark 2. *Note also that the matrix Π is defined so that it commutes with D. This is necessary to ensure that the variance of the test statistic is invariant under the action of Π.*

We now introduce some notation before establishing the main result, that constitutes the pivot point to prove the asymptotic control of the probability of Type I error. For

16

the sake of simplicity of exposition, we consider the results for a single covariate case in the remainder of this section and in Section 4.3. In Section 4.4, we outline the procedure for the general multivariate case. Let $\alpha \in [0, 1)$. For any $a \in \mathbb{R}$, let $\lceil a \rceil$ be the smallest integer which is larger than or equal to a and let $\lfloor a \rfloor$ be the largest integer which is at most a. We consider all the possible $w = 2^n$ sign-flips $\Pi_1, \ldots, \Pi_w$, where $\Pi_1 = I$. For a given value of the test statistic T_I^n, we hence consider all the associated sign-flipped values $T_I^n, T_{\Pi_2}^n, \ldots, T_{\Pi_w}^n$, where we use the superscript n to highlight the sample size. We denote by $T_{(1)}^n \leq \ldots \leq T_{(w)}^n$ the corresponding sorted value. Finally, we write $T_{[1-\alpha]}^n = T_{(\lceil 1-\alpha \rceil w)}^n$.

Theorem 4.2. *Consider the test that rejects H_0 if and only if $T_I^n > T_{[1-\alpha]}^n$. Then, under the null hypothesis, the test is asymptotically exact and the rejection probability $\mathbb{P}(T_I^n > T_{[1-\alpha]}^n)$ is at most α.*

Proof. We need to show that the asymptotic distribution of the 2^n-dimensional vector of test statistics $\mathbf{T} = (T_I^n, \ldots, T_{\Pi_w}^n)^\top$ is invariant under sign-flip transformations Π, that is $\mathbf{T} \stackrel{d}{=} \Pi \circ \mathbf{T}$, where $\stackrel{d}{=}$ represent the equality in distribution and the composition stands for $\Pi \circ \mathbf{T} = \Pi \circ (T_I^n, T_{\Pi_2}^n, \ldots, T_{\Pi_w}^n)^\top = (T_{\Pi I}^n, T_{\Pi \Pi_2}^n, \ldots, T_{\Pi \Pi_w}^n)^\top$. This will prove the asymptotic control of the Type I error through Theorem 15.2.1 in Lehmann and Romano [2008] and Theorem 1 in Hemerik and Goeman [2018a].

Let $\tilde{\mathbf{X}}$ be the diagonal $n \times n$ matrix with elements $(\tilde{X}_1, \ldots, \tilde{X}_n)$. The test statistic in Definition 1 can be rewritten as

$$T_\Pi = \tilde{X}^\top \Pi \tilde{\mathbf{r}} = \mathbb{1}_n^\top \Pi \tilde{\mathbf{X}} \tilde{\mathbf{r}},$$

where $\mathbb{1}_n$ is the n-dimensional unit vector. The test statistic T_Π can hence be viewed as sum of n contributions, where each element of $\tilde{\mathbf{X}} \tilde{\mathbf{r}}$ is sign-flipped through Π. Similarly, the variance of T_Π can be written as

$$\mathrm{Var}(T_\Pi) = n^{-1} \sigma^2 \mathbb{1}_n^\top \tilde{\mathbf{X}} D \tilde{\mathbf{X}} \mathbb{1}_n;$$

17

see also Theorem 4.1.

To evaluate the joint distribution of the test statistics $\mathbf{T}$, let us now define $\mathbf{\Pi}$ as the $2^n \times n$ matrix collecting all the $w = 2^n$ vectors of sign-flip row-wise. Therefore, we can write $\mathbf{T} = \mathbf{\Pi}\tilde{\mathbf{X}}\tilde{\mathbf{r}}$, and $\Pi \circ \mathbf{T} = \mathbf{\Pi}\mathbf{\Pi}\tilde{\mathbf{X}}\tilde{\mathbf{r}}$. The joint distribution of $\mathbf{T}$ is multivariate normal with variance $\mathrm{Var}(\mathbf{T}) = \sigma^2 n^{-1} \mathbf{\Pi}\tilde{\mathbf{X}}D\tilde{\mathbf{X}}\mathbf{\Pi}^\top$ and asymptotically zero mean. We now have to show that $\Pi \circ \mathbf{T}$ follows the same asymptotic multivariate normal distribution. First note that the transformation Π does not affect the expected value, that remains asymptotically zero. Furthermore, for the variance we have

$$
\begin{aligned}
\mathrm{Var}(\Pi \circ \mathbf{T}) &= n^{-1}\sigma^2 \mathbf{\Pi}\mathbf{\Pi}\tilde{\mathbf{X}}D\tilde{\mathbf{X}}\mathbf{\Pi}\mathbf{\Pi}^\top \\
&= n^{-1}\sigma^2 \mathbf{\Pi}\mathbf{\Pi}\mathbf{\Pi}\tilde{\mathbf{X}}D\tilde{\mathbf{X}}\mathbf{\Pi}^\top \\
&= n^{-1}\sigma^2 \mathbf{\Pi}\tilde{\mathbf{X}}D\tilde{\mathbf{X}}\mathbf{\Pi}^\top.
\end{aligned}
$$

Thanks to Theorem 15.2.1 in Lehmann and Romano [2008] and Theorem 1 in Hemerik and Goeman [2018a], this yields the null invariance $\mathbf{T} \overset{d}{=} \Pi \circ \mathbf{T}$. It follows that under H_0, $\mathbb{P}(T_I^n > T_{[1-\alpha]}^n) \leq \alpha$.

$\square$

Remark 3. *The previous result is still valid in the case when $w \neq 2^n$, i.e., when not every element of $\{1, -1\}^n$ is used once [see, e.g. Hemerik and Goeman, 2018a]. For computational reasons, it is in fact common practice to sample uniformly from $\{-1, 1\}^n$, with or without replacement. The same holds also for the results in the next section.*

4.2 On the nature of the eigen sign-flip test

A few comments on this approach may be useful to understand its nature.

In the simpler context of linear regression models, it is possible to define a residualizing matrix that projects into the residual space. This is an orthogonal projection matrix; as

such, it is idempotent, and all its eigenvalues are zero or one. Thanks to this, for classical linear regression models, Kherad-Pajouh and Renaud [2010] propose to pre-multiply the residuals by the semiorthogonal matrix defined by the eigenvectors corresponding to the non-null eigenvalues. This pre-multiplication transforms the n residuals into pseudo-residuals, reducing their cardinality. The number of these new pseudo-residuals is equal to the rank of the residualizing projection matrix (i.e., the number of non zeros eigenvalues, usually equal to the number of covariates). Being the remaining eigenvalues of residualizing matrix equal to one, the resulting pseudo-residuals are now independent and homoscedastic (i.e., spherical). In particular, Kherad-Pajouh and Renaud [2010] suggest the use of a permutation approach, while Solari et al. [2014] extend it to the more general framework of rotations matrices.

Unfortunately, within the semiparametric regression framework, the residualizing matrix Λ is not a projection matrix and is not idempotent; therefore, its eigenvalues do not take values in $\{0, 1\}$. A multiplication by these eigenvalues would act as a scaling factor for the residuals, making them independent, but not homoscedastic. For this reason, defining Π as a permutation or a rotation matrix would not be a valid solution. Defining instead Π as sign-flipping matrix ensures the commutative property $\Pi D = D \Pi$, as highlighted in Remark 2. This property is indeed crucial, as it guarantees that the variance of the test statistics is constant over Π, as proved in Theorem 4.1. It is also worth to emphasize that the test guarantees only asymptotic exactness since the penalization induces a bias in the estimation of the mean – which vanishes with increasing n – while, for fixed n, the variance remains constant for all the test statistics defining the null distribution. On the contrary, the standard parametric Wald test needs asymptotic results for both the mean and the variance, such as those obtained in Section 3.1. Similar considerations could be drawn for a naive sign-flip score test that does not decomposes the matrix Λ. In this case, for finite samples, the sign-flipped test statistic would not have variance equal to observed test

statistic. This would lead to performances that are comparable to the parametric Wald test, as shown in Ferraccioli [2020]. This difference between the eigen sign-flip test and the other competitors is crucial in providing an adequate control of Type I error, as shown by the simulations in Section 5.

4.3 Interval hypotheses, two-sided tests and confidence intervals

So far we have defined the eigen sign-flip test for point-wise null hypothesis. The most common situation in practice is nonetheless to define interval null hypotheses. As for standard approaches, we need to prove that the p-value computed under any point-wise null hypothesis $H_0 : \beta = \beta_0 - \epsilon$ ($\forall\ \epsilon > 0$) has also Type I error probability bounded by α. For convenience, let us define the test statistic as a function of the tested coefficient, that is $T_\Pi(\beta_0) = n^{-1/2} X^\top V \Pi V^\top \Lambda (\mathbf{y} - X\beta_0) = n^{-1/2} X^\top V \Pi D V (\mathbf{y} - X\beta_0)$. We now give two results for interval hypothesis and two sided tests, and a third result on confidence intervals.

Corollary 4.2.1 (Interval hypotheses)**.** *Consider the hypotheses* $H_0^{'} :\ \beta = \beta_0 - \epsilon$ *and* $H_1^{'} :\ \beta > \beta_0 - \epsilon$, *with* $\epsilon > 0$. *Then for every* $\epsilon > 0$, $\mathbb{P}(T_\Pi(\beta_0 - \epsilon) \geq T_I(\beta_0 - \epsilon)) \leq \mathbb{P}(T_\Pi(\beta_0) \geq T_I(\beta_0))$. *The same is true for the opposite hypothesis: that is,* $\epsilon < 0$ *and* $H_1' :\ \beta < \beta_0 + \epsilon$.

Proof. We have

$$\mathbb{P}(T_\Pi(\beta_0 - \epsilon) \geq T_I(\beta_0 - \epsilon)) =$$
$$= \mathbb{P}(T_\Pi(\beta_0) - n^{-1/2} X^\top V \Pi D V^\top X \epsilon \geq T_I(\beta_0) - n^{-1/2} X^\top D V^\top X \epsilon) =$$
$$= \mathbb{P}(T_\Pi(\beta_0) + n^{-1/2} X^\top V (I - \Pi) D V^\top X \epsilon \geq T_I(\beta_0)) =$$
$$\leq \mathbb{P}(T_\Pi(\beta_0) \geq T_I(\beta_0)).$$

Note that last inequality holds since

$$n^{-1/2} X^\top V (I - \Pi) DVX\epsilon = \mathbb{1}_n^\top \tilde{\mathbf{X}} (I - \Pi) \tilde{\mathbf{X}} \mathbb{1}_n \epsilon \geq 0 \quad \forall \Pi.$$

Note that $(I - \Pi)$ is a diagonal matrix with non-negative diagonal entries, thus it is positive semi-definite for all Π. Therefore any quadratic form of it is non-negative. $\qquad \square$

Corollary 4.2.2 (Two-sided test). *Consider $\alpha_1, \alpha_2 \in (0, 1]$ and such that $\alpha_1 + \alpha_2 < 1$. Then, under $H_0 : \beta = \beta_0$, as $n \to \infty$,*

$$\mathbb{P}\left[(T_I^n < T_{[\alpha_1]}^n) \cup (T_I^n > T_{[1-\alpha_2]}^n)\right] \to \alpha_1 + \alpha_2.$$

That is, the eigen sign-flip test controls the Type I error asymptotically when testing $H_0 : \beta = \beta_0$ against the two sided alternative $H_1 : \beta \neq \beta_0$.

Proof. Theorem 4.2 proves that $\mathbb{P}\left[T_I^n < T_{[\alpha_1]}^n\right] \to \alpha_1$ and $\mathbb{P}\left[T_I^n > T_{[1-\alpha_2]}^n\right] \to \alpha_2$. This, together with the fact that $\mathbb{P}\left[(T_I^n < T_{[\alpha_1]}^n) \cap (T_I^n > T_{[1-\alpha_2]}^n)\right] \to 0$ proves the corollary. $\quad \square$

As consequence of the two lemmas above, we can also derive confidence intervals for the parameter β.

Corollary 4.2.3 (Confidence Interval). *Let $\alpha \in (0, 1]$. Then the set*

$$\Theta(\beta) = \{\beta : \mathbb{P}(T_\Pi(\beta) \geq (\leq) T_I(\beta)) \geq (\leq)\alpha\}$$

is a one-sided confidence interval for parameter β with asymptotic coverage $1 - \alpha$. Let also $\alpha_1, \alpha_2 \in (0, 1]$ such that $\alpha_1 + \alpha_2 < 1$. Similarly, the set

$$\Theta(\beta) = \{\beta : (\mathbb{P}(T_\Pi(\beta) \geq T_I(\beta)) \geq \alpha_1) \cap (\mathbb{P}(T_\Pi(\beta) \leq T_I(\beta)) \leq \alpha_2)\}$$

is a two-sided confidence interval with asymptotic coverage $1 - (\alpha_1 + \alpha_2)$.

Proof. The proof follows directly from Corollaries 4.2.1 and 4.2.2. $\qquad \square$

21

4.4 Testing a subset of the covariates

We now deal with the case where we have multiple covariates and we are interested in testing a subset of the covariates. Specifically, assume $X \in \mathbb{R}^n \times \mathbb{R}^q$ represents the set of covariates of interest, with associated vector of coefficients $\boldsymbol{\beta}$, and $Z \in \mathbb{R}^n \times \mathbb{R}^p$ the set of covariates associated with the vector of nuisance coefficients $\boldsymbol{\zeta}$. The minimization problem in (2) then becomes

$$\underset{\boldsymbol{\beta},f}{\operatorname{argmin}} \sum_{i=1}^n (y_i - f(\mathbf{p}_i) - \mathbf{x}_i^\top \boldsymbol{\beta} - \mathbf{z}_i^\top \boldsymbol{\zeta})^2 + \lambda_n \mathcal{P}(f). \tag{16}$$

We might be interested in testing

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{versus} \qquad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

for any value of $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$. Let us define $\Psi^* = [Z|\Psi]$ the $n \times (p+K)$ matrix composed by the covariates associated with the nuisance parameters and the bases for the nonparametric part of the model, with coefficients $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \boldsymbol{\gamma})$. We can then rewrite equation (5) as

$$(\Psi^{*\top}\Psi^* + \lambda P^*)\boldsymbol{\theta} - \Psi^\top(\mathbf{y} - X\boldsymbol{\beta}) = 0, \quad P^* = \begin{bmatrix} \mathbb{O}_{p \times p} & \mathbb{O}_{p \times K} \\ \mathbb{O}_{K \times p} & P_{K \times K} \end{bmatrix},$$

where $\mathbb{O}$ is a matrix of zeros.

Definition 1 of the eigen sign-flip test remains valid also in this case, with the only modification of the matrix Λ in (6), where Ψ is replaced by Ψ^*. Moreover, the following corollary provides the extension of Theorem 4.2 to the case where $\boldsymbol{\beta}$ is a vector.

Corollary 4.2.4. *Consider the test that rejects $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ if and only if $\varphi(T_I^n) > \varphi(T_{[1-\alpha]}^n)$, where $\varphi(\cdot)$ is any Nonparametric Combining function [Pesarin, 2001, Section 6.2]. Then, under the null hypothesis, the test is asymptotically exact and the rejection probability $\mathbb{P}(\varphi(T_I^n) > \varphi(T_{[1-\alpha]}^n))$ is at most α.*

Proof. In order to extend the proof of Theorem 4.2 to the multivariate framework, we need to rely on the Nonparametric Combination of dependent test statistics, as defined, e.g., in Section 6.2 of Pesarin [2001]. First of all, recall that the test statistic T is a vector itself. Moreover, Theorem 3.2 proves the asymptotic multivariate normality of T and Theorem 4.1 shows that the sign-flipped vectors of test statistics $T_I^n, \ldots, T_{\Pi_w}^n$ share the same distribution. Therefore, the matrix $\mathbf{T} = (T_I^n, \ldots, T_{\Pi_w}^n)^\top$ is equal in distribution to $\Pi \circ \mathbf{T}$. More precisely, $\mathbf{T}$ is the 2^n-dimensional vector of test statistics $T_{(\cdot)}^n$, i.e., each row of $\mathbf{T}$ is a sampling from the multivariate test statistics T. We can therefore use any Nonparametric Combination function Pesarin [2001] to obtain a p-value. $\square$

Among the most commonly used Nonparametric Combining functions, defined, e.g., in Pesarin [2001], are the max-T, sum-T or Mahalanobis distance. As an illustrative example, a p-value based on the min-p combining function [Westfall and Young, 1993] rejects the multivariate null hypothesis if the maximum value of T is larger than the $1 - \alpha$ quantile of the distribution of the maxima computed over the w elements of $(T_I^n, \ldots, T_{\Pi_w}^n)$.

5 Simulation studies

In this section we present two simulation studies, to investigate the finite sample performances of the proposed test. Simulation 1, in Section 5.1, considers a semiparametric model based on classical univariate splines [as for instance in Heckman, 1986, Wand and Ormerod, 2008]. Simulation 2, in Section 5.2, considers instead a semiparametric model based on SR-PDE [Sangalli et al., 2013]. In these different settings, we compare the performances of three different tests:

- Wald: a classical Wald type test based on the asymptotic distribution of $\hat{\boldsymbol{\beta}}$;

- Speck: a similar Wald type test based on the asymptotic distribution of the Speckman

version of the estimator [Speckman, 1988], as derived in Holland [2017];

- ESF: the Eigen sign-flip score test introduced in Definition 1.

The results show the performances of the tests over 1000 simulation repetitions.

5.1 Simulation 1

In Simulation 1, we simulate from model (1), with $\Omega = [0, 1]$ and $p_1, \ldots, p_n$ randomly sampled from a uniform distribution on Ω, with $n = 200$. For the nonparametric component of the model, we consider the test function 1 from the function `gamSim` in the `R` package `mgcv` [Wood, 2015, 2017], defined as $0.2p^{11}(10(1 - p))^6 + 10(10p)^3(1 - p)^{10}$. We consider $q = 1$ covariate, and we generate $x_1, \ldots, x_n$ according to four different stochastic processes:

(a) an i.i.d. random sample from $\mathcal{N}(0, 0.1^2)$;

(b) a Gaussian random field on $[0, 1]$ with mean zero and scale 0.01;

(c) the function $\sqrt{p + 2}$ on $[0, 1]$, with added an i.i.d. random sample from $\mathcal{N}(0, 0.1^2)$;

(d) the function $\sqrt{p + 2}$ on $[0, 1]$, with added a Gaussian random field with mean zero and scale 0.01.

The covariates and the true f are standardized, before computing the response variable y, so that their relative contributions to the response are comparable. We consider both $\beta_0 = 0$ and other 10 different values of β_0, from 0.01 to 0.1, to check both the Type I error and the power of the test. Finally, we add i.i.d. normal random errors $\epsilon_1, \ldots, \epsilon_n$, with zero mean and standard deviation 0.1. For each test case, the generation of the covariates and noise is repeated 1000 times.

The model is estimated using cubic B-spline bases, with 200 equispaced internal nodes on Ω, using the implementation in Wand and Ormerod [2008]. The smoothing parameter

is chosen via cross-validation. The tests are performed with nominal value 0.05. For the proposed eigen sign-flip test, we consider 1000 random sign-flips.

Table 1: Simulation 1. Proportion of Type I error in 1000 replicates (nominal value of the test: 0.05)

	(a)	(b)	(c)	(d)
Wald	0.057	0.260	0.054	0.262
Speck	0.023	0.089	0.023	0.089
ESF	0.051	0.038	0.045	0.042

Table 1 shows the control of Type I error and Figure 1 shows the power functions for the three competing tests. The table and figure immediately highlight that the most challenging scenarios are cases (b) and (d), where the covariates have been generated with a dependence structure, sampling from a Gaussian process. The classic parametric test (Wald) shows an extremely poor control of the Type I error in these two cases, with an observed proportion of Type I error of over 26%, when the nominal value of the test is 5%. This behavior is possibly due to the poor estimation of the variance induced by the regularized estimates. The Speckman variant appears more robust, partly correcting for the misspecified variance. Nonetheless, this test is significantly under-conservative in cases (b) and (d), with a proportion of Type I error of almost 9%, while it is over-conservative in cases (a) and (c), where it returns a proportion of Type I error of about 2.3%. The proposed eigen sign-flip score test, on the contrary, maintains an extremely good control of the Type I error, under all scenarios, and it is never under-conservative. Also in the challenging cases (b) and (d), at the cost of a slightly loss of power, it manages to keep a proportion of Type I error very close (and just slightly inferior) to the nominal value of the test.

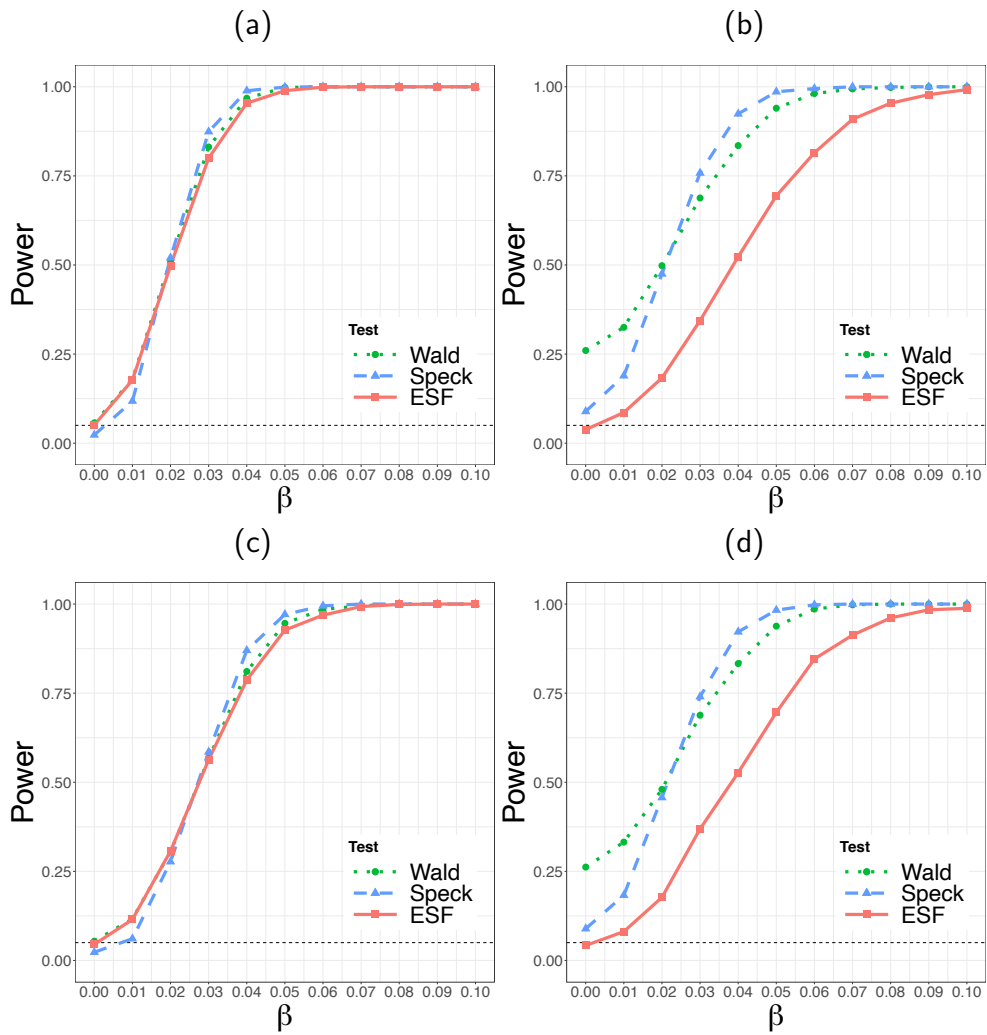We also considered the case of multiple covariates, following the simulation scheme

25

Figure 1: Simulation 1. Power of the Wald test (green dotted), of its Speckman variant (cyan dashed) and of the proposed eigen sign-flip (red solid).

detailed above, but including simultaneously all four covariates (a), (b), (c) and (d) in the data generation, and testing one parameter at a time, considering the other parameters as nuisance, as detailed in Section 4.4. The same considerations as those detailed for the simulation in Figure 1 can be drown (results non included for sake of space).

5.2 Simulation 2

In Simulation 2, we simulate from model (1), with $\Omega = [0,1] \times [0,1]$ and $\mathbf{p}_1, \ldots, \mathbf{p}_n$ randomly sampled from a uniform distribution on Ω, with $n = 225$. For the nonparametric component of the model, we consider the test function 2 from the function `gamSim` in the `R` package `mgcv` [Wood, 2015, 2017], defined as

$$0.4\pi^{0.3}\left(1.2\exp\left(-\frac{(p_1 - 0.2)^2}{0.3^2} - \frac{(p_2 - 0.3)^2}{0.4^2}\right) + 0.8\exp\left(-\frac{(p_1 - 0.7)^2}{0.3^2} - \frac{(p_2 - 0.8)^2}{0.4^2}\right)\right).$$

We consider $q = 1$ covariate, and we generate $x_1, \ldots, x_n$ according to four different stochastic processes:

(a) a Gaussian random field with zero mean and scale 0.05;

(b) a Matern random field with $\nu = 1$, $\sigma = 2$ and scale 0.1;

(c) the function $\cos(5(p_1 + p_2)) + (2p_1 - p_1 p_2^2)^2$ with added a Gaussian random field with scale 0.05;

(d) the function $\cos(5(p_1 + p_2)) + (2p_1 - p_1 p_2^2)^2$ with added a Matern random field with $\nu = 1$, $\sigma = 2$ and scale 0.1.

The covariates and the true f are standardized, before computing the response variable y, so that their relative contributions to the response are comparable. We consider both $\beta_0 = 0$ and other 10 different values of β_0, from 0.01 to 0.1, to check both the Type I error and the power of the test. Finally, we add i.i.d. normal random errors $\epsilon_1, \ldots, \epsilon_n$, with zero

27

mean and standard deviation 0.1. For each test case, the generation of the covariates and noise is repeated 1000 times.

The model is estimated using SR-PDE, with linear finite elements on a mesh having 225 nodes on a regular lattice over Ω, implemented using the package `fdaPDE`. The smoothing parameter is chosen via cross-validation. The tests are performed with nominal value 0.05. For the proposed eigen sign-flip test, we consider 1000 random sign-flips.

Table 2: Simulation 2. Proportion of Type I error in 1000 replicates (nominal value of the test: 0.05)

	(a)	(b)	(c)	(d)
Wald	0.146	0.176	0.270	0.319
Speck	0.084	0.094	0.102	0.103
ESF	0.041	0.040	0.054	0.046

The results are presented in Table 2 and Figure 2. The classic parametric test (Wald) has poor performances and very low control of the Type I error in all the scenarios, with proportion of Type I error of about 15% and higher. The Speckam variant is always more robust than the Wald, but it is often severely under-conservative, with observed proportion of Type I error of about 10%. The proposed eigen sign-flip, on the contrary, at a loss of some power, permits an extremely good control of the Type I error, even in the more challenging scenarios, where the covariate has a strong spatial structure.

6 Study of human development in Nigeria

In this section we apply the proposed methodology to the analysis of human development in Nigeria. In particular, we are interested in better understanding the difference in socio-economic and health conditions in the various states of the country. Unfortunately, data
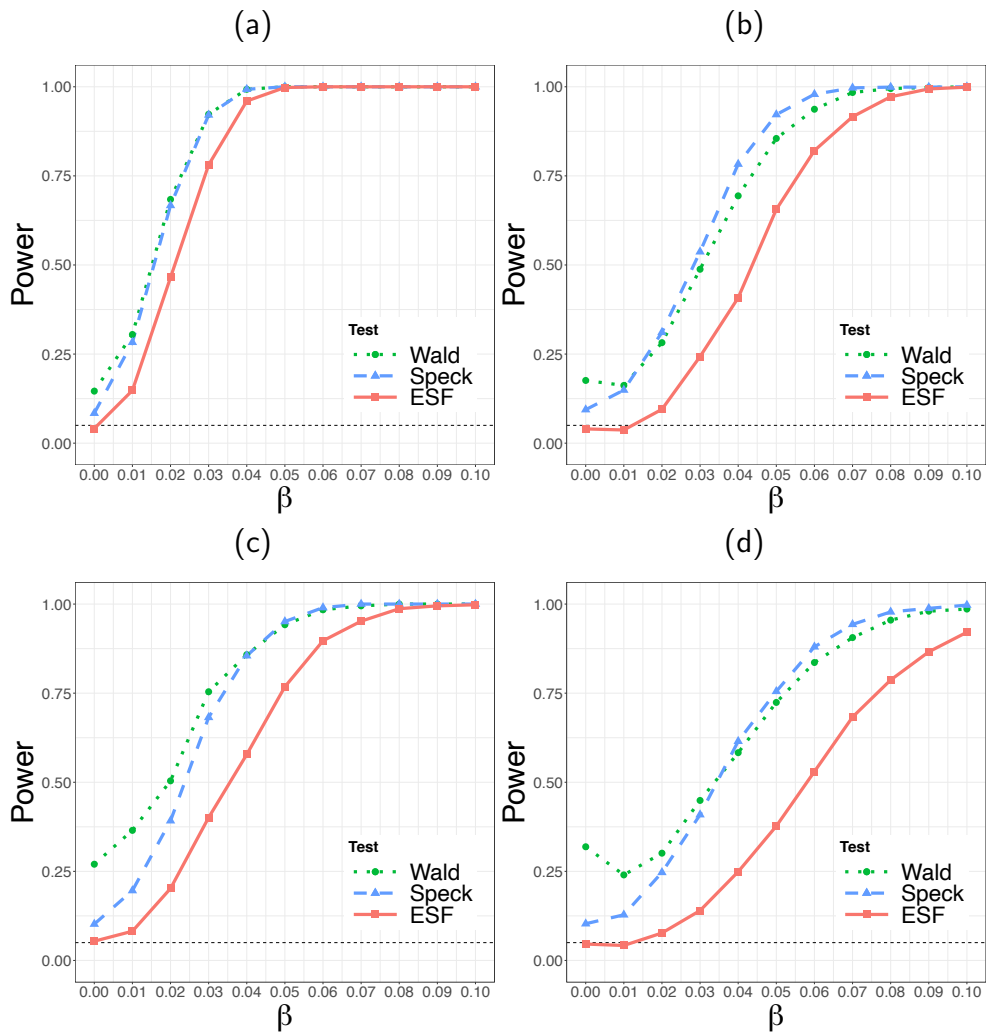
Figure 2: Simulation 2. Power of the Wald test (green dotted), of its Speckman variant (cyan dashed) and of the proposed eigen sign-flip (red solid).

at national and subnational level are often poor or not publicly available. This lack of information and of public domain surveys hamper the efforts to identify and develop targeted interventions in troubled areas [Jerven, 2013]. An alternative to traditional data consists in using other sources of openly accessible data, such as data from social media, mobile phone networks, or satellites. In particular, a popular recent approach leverages on satellite images of luminosity at night to estimate economic activity [Chen and Nordhaus, 2011, Jean et al., 2016]. These images highlight urban areas, which typically offer better provisions of basic services such as electricity, water and public health, as well as more job opportunities, with respect to rural areas.

Here we use open satellite data (NASA Worldview Snapshots), together with demographic data, to predict human development. Specifically, as a response variable, we consider the Human Development Index (HDI) (available at https://globaldatalab.org/shdi), an aggreated index that takes into account multiple dimensions at the household and individual level in health, education and standard of living. This index is available at states level, for the 36 states of Nigeria, and for the Federal Capital Territory. The values of this index are shown in panel d of Figure 3. As covariates, in the parametric part of the model, we use the population density, $\mathbf{x}_{Pop}$, of each state (data from the National Bureau of Statistics, Nigeria), shown in panel e of Figure 3, and the three satellite images shown in the top panels of the same figure, that are

- Nightlight luminosity, $\mathbf{x}_{Night}$, obtained via the VIIRS Nighttime Imagery, that captures low-light emission sources, under varying illumination conditions (panel a);

- Short-Wave Infrared, $\mathbf{x}_{SWIR}$, that highlights bare soils, such as deserts (panel b);

- Near Infrared, $\mathbf{x}_{NIR}$, that highlights vegetation (panel c).

We are interested in identifying significant effects of these covariates on human development,

30

considering the model

$$\mathbf{y} = \mathbf{f} + \mathbf{x}_{Pop}\beta_{Pop} + \mathbf{x}_{Night}\beta_{Night} + \mathbf{x}_{SWIR}\beta_{SWIR} + \mathbf{x}_{NIR}\beta_{NIR} + \epsilon.$$

Since the HDI, the response variable, and one of the covariate, the population density, are available at state level, we also aggregate the other three covariates at state level, considering their areal means. We then apply SR-PDE, considering the data located at the capitals of each state. We use a mesh with 320 nodes and select the smoothing parameter through generalized cross-validation ($\lambda_n = 0.1$). We hence perform significance tests on each covariate, one at a time, considering the other parameters as nuisance, as described in Section 4.4, using the eigen sign-flip procedure with 5000 random sign-flips.

Nightlight results significant ($p < 0.005$), with a positive impact on human development (the estimated coefficients is 0.29). The finding on nightlight is in line with other recent research studies [Chen and Nordhaus, 2011, Jean et al., 2016]. The presence of urban areas, in fact, plays a huge role in the overall wealth of the population. This of course does not imply a causality effect, since increased wealth has itself an impact on the development of urban areas. Nightlight is nonetheless a good indicator of the socio-economic status at local level, that does not require any official statistics, as previously discussed. Short-wave infrared seems to be slightly significant ($0.05 < p < 0.1$), with a negative impact on human development (the estimated coefficients is -0.016). The result might suggest that the presence of deserted areas with large amounts of bare soil lead to a decrease in human development. The more advanced states are indeed close to the ocean, in the southern part of the country. The northern part instead, that is mostly deserted, is not very populated. It is also worth noting that the aggregation at state level averages localized features, such as the presence of rivers, lakes or small vegetation, possibly reducing important informations. The third satellite covariate, near infrared, does not appear significant ($p > 0.1$). The same apply for population density ($p > 0.1$). This is possibly due to the fact that the distribution is highly skewed, with most of the population residing in the state of Lagos,
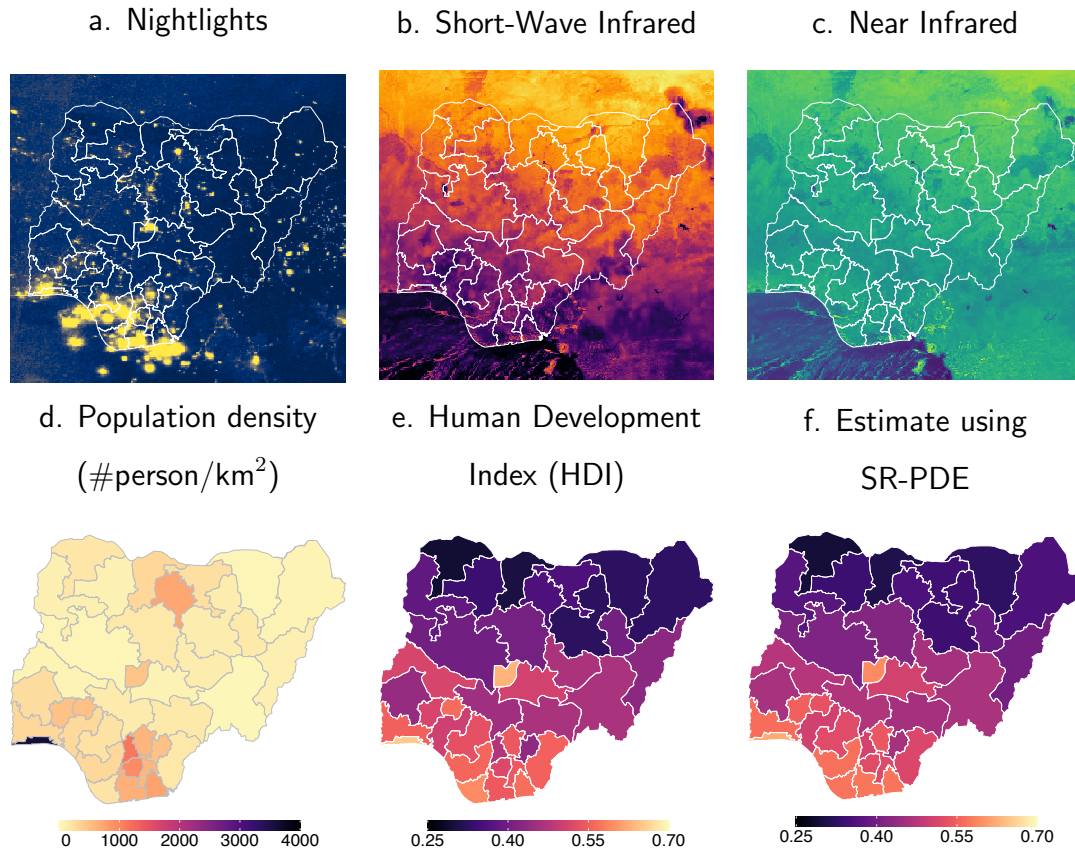
31

Figure 3: Panels a. – d. show the covariates used: nightlight luminosity (a), Short-wave infrared (b), Near infrared (c) satellite images, and population density (d). Panel e shows the observed HDI. Panel f shows the HDI predicted by the SR-PDE model. Imagery from the Worldview Snapshots application (https://wvs.earthdata.nasa.gov)

in the south-west of the country (see panel d in Figure 3). Panel f in Figure 3 shows the predicted MPI values, highlighting the very high explicative power of the model.

7 Discussion

This paper describes a strongly innovative and highly promising inferential approach in the context of semiparametric regression with roughness penalties. The paper focuses on tests for the linear part of the models. On the other hand, similar ideas can be used to develop tests and confidence bands on the nonlinear part of the models. Moreover, the described approach could be extended to deal with semiparametric regression with spatio-temporal components [see, e.g. Ugarte et al., 2009, 2010, Aguilera-Morillo et al., 2017, Marra et al., 2012, Augustin et al., 2013, Bernardi et al., 2018], further broadening the spectrum of potential models that could benefit from our proposal. These developments will be objects of dedicated future studies. We are confident this inferential approach will become popular and will prove to be highly valuable in the varied contexts where semiparametric regression is used.

References

M Carmen Aguilera-Morillo, María Durbán, and Ana M Aguilera. Prediction of functional data with spatial dependence: a penalized approach. *Stochastic environmental research and risk assessment*, 31(1):07–22, 2017.

Eleonora Arnone, Alois Kneip, Fabio Nobile, and Laura M. Sangalli. Some first results

on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*, page DOI: 10.5705/ss.202019.0346, 2021.

Nicole H Augustin, Verena M Trenkel, Simon N Wood, and Pascal Lorance. Space-time modelling of blue ling for fisheries stock management. *Environmetrics*, 24(2):109–119, 2013.

Laura Azzimonti, Laura M Sangalli, Piercesare Secchi, Maurizio Domanin, and Fabio Nobile. Blood flow velocity field estimation via spatial regression with pde penalization. *Journal of the American Statistical Association*, 110(511):1057–1071, 2015.

Victoria Baramidze, Ming-Jun Lai, and CK Shum. Spherical splines for data interpolation and fitting. *SIAM Journal on Scientific Computing*, 28(1):241–259, 2006. ISSN 1064-8275. doi: 10.1137/040620722.

Mara S. Bernardi, Michelle Carey, James O. Ramsay, and Laura M. Sangalli. Modeling spatial anisotropy via regression with partial differential regularization. *Journal of Multivariate Analysis*, 167:15–30, 2018. ISSN 0047-259X. doi: 10.1016/j.jmva.2018.03.014.

Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and John A. Wellner. *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag, New York, 1998. ISBN 0-387-98473-9. Reprint of the 1993 original.

Xi Chen and William D Nordhaus. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594, 2011.

EunYi Chung and Joseph P. Romano. Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484 – 507, 2013. doi: 10.1214/13-AOS1090.

Gerda Claeskens, Tatyana Krivobokova, and Jean D Opsomer. Asymptotic properties of penalized spline estimators. *Biometrika*, 96(3):529–544, 2009.

A Demmler and C Reinsch. Oscillation matrices with spline smoothing. *Numerische Mathematik*, 24(5):375–382, 1975.

Jean Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. pages 85–100. Lecture Notes in Math., Vol. 571, 1977.

Bree Ettinger, Simona Perotto, and Laura M. Sangalli. Spatial regression models over two-dimensional manifolds. *Biometrika*, 103(1):71–88, 2016. ISSN 0006-3444. doi: 10.1093/biomet/asv069.

Randall L. Eubank. *Nonparametric regression and spline smoothing*, volume 157 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, second edition, 1999. ISBN 0-8247-9337-4.

Federico Ferraccioli. *Nonparametric methods for complex spatial domains: density estimation and hypothesis testing*. PhD thesis, Università degli Studi di Padova, 2020.

Federico Ferraccioli, Laura M Sangalli, and Livio Finos. Some first inferential tools for spatial regression with differential regularization. *Journal of Multivariate Analysis*, page doi:10.1016/j.jmva.2021.104866, 2021.

David A Freedman. On the so-called "huber sandwich estimator" and "robust standard errors". *The American Statistician*, 60(4):299–302, 2006.

Robert J Gray. Spline-based tests in survival analysis. *Biometrics*, pages 640–652, 1994.

Peter J Green and Bernard W Silverman. *Nonparametric regression and generalized linear models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1994. ISBN 0-412-30040-0. doi: 10.1007/978-1-4899-4473-3.

Serge Guillas and Ming-Jun Lai. Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, 22(3-4):477–497, 2010. ISSN 1048-5252. doi: 10.1080/10485250903323180.

Peter Hall and Joel Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921, 2013. ISSN 0090-5364. doi: 10.1214/13-AOS1137.

Nancy E Heckman. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(2):244–248, 1986.

Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *TEST*, 27(4): 811–825, 2018a.

Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 27(4):811–825, December 2018b.

Jesse Hemerik, Jelle J Goeman, and Livio Finos. Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):841–864, 2020.

Ashley D Holland. Penalized spline estimation in the partially linear model. *Journal of Multivariate Analysis*, 153:211–235, 2017.

Myung-Hoe Huh and Myoungshic Jhun. Random permutation testing in multiple linear regression. *Communications in statistics-Theory and Methods*, 30(10):2023–2032, 2001.

Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

Morten Jerven. *Poor numbers: how we are misled by African development statistics and what to do about it.* Cornell University Press, 2013.

Sara Kherad-Pajouh and Olivier Renaud. An exact permutation method for testing any effect in balanced and unbalanced fixed effect anova. *Computational Statistics & Data Analysis*, 54(7):1881–1893, 2010.

Ming-Jun Lai and Larry L. Schumaker. *Spline functions on triangulations*, volume 110 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2007. ISBN 978-0-521-87592-9; 0-521-87592-7. doi: 10.1017/CBO9780511721588.

Ming-Jun Lai and Li Wang. Bivariate penalized splines for regression. *Statistica Sinica*, 23 (3):1399–1417, 2013. ISSN 1017-0405.

Ming-Jun Lai, C. K. Shum, V. Baramidze, and P. Wenston. Triangulated spherical splines for geopotential reconstruction. *Journal of Geodesy*, 83(4):695–708, 2009.

Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses.* Springer Science & Business Media, 2008.

Yingxing Li and David Ruppert. On the asymptotics of penalized splines. *Biometrika*, 95 (2):415–436, 2008.

Cora JM Maas and Joop J Hox. Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, 58(2):127–137, 2004.

Giampiero Marra and Simon N Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74, 2012.

Giampiero Marra, David L Miller, and Luca Zanin. Modelling the spatiotemporal distribution of the incidence of resident foreign population. *Statistica Neerlandica*, 66(2): 133–160, 2012.

Douglas Nychka. Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83(404):1134–1143, 1988. ISSN 0162-1459.

Finbarr O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical science*, pages 502–518, 1986.

Markus Pauly, Edgar Brunner, and Frank Konietschke. Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):461–473, 2015. doi: https://doi.org/10.1111/rssb.12073. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12073`.

Fortunato Pesarin. *Multivariate permutation tests: with applications in biostatistics*, volume 240. Wiley Chichester, 2001.

David Ruppert, Matt P Wand, and Raymond J Carroll. *Semiparametric regression*. Number 12. Cambridge university press, 2003.

Laura M Sangalli. Spatial regression with partial differential equation regularisation. *International Statistical Review*, 89(3):505—-531, 2021. doi: DOI: 10.1111/insr.12444.

Laura M Sangalli, James O Ramsay, and Timothy O Ramsay. Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75 (4):681–703, 2013.

Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

Aldo Solari, Livio Finos, and Jelle J Goeman. Rotation-based multiple testing in the multivariate linear model. *Biometrics*, 70(4):954–961, 2014.

Paul Speckman. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(3):413–436, 1988.

María Dolores Ugarte, Tomás Goicoa, Ana F Militino, and María Durbán. Spline smoothing in small area trend estimation and forecasting. *Computational Statistics & Data Analysis*, 53(10):3616–3629, 2009.

MD Ugarte, T Goicoa, and AF Militino. Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics: The official journal of the International Environmetrics Society*, 21(3-4):270–289, 2010.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Grace Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981. ISSN 0196-5204. doi: 10.1137/0902002.

Grace Wahba. Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B. Methodological*, 45(1):133–150, 1983. ISSN 0035-9246.

Grace Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990. ISBN 0-89871-244-0. doi: 10.1137/1.9781611970128.

Matt P Wand and JT Ormerod. On semiparametric regression with o'sullivan penalized splines. *Australian & New Zealand Journal of Statistics*, 50(2):179–198, 2008.

Li Wang, Guannan Wang, Min-Jun Lai, and Lei Gao. Efficient estimation of partially linear models for spatial data over complex domains. *Statistica Sinica*, 30:347–369, 2020.

Yuedong Wang. *Smoothing splines: methods and applications*. Chapman and Hall/CRC, 2019.

Peter H Westfall and S Stanley Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons, 1993.

Matthieu Wilhelm, Luca Dedè, Laura M. Sangalli, and Pierre Wilhelm. IGS: an IsoGeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering*, 302:70–89, 2016. ISSN 0045-7825. doi: 10.1016/j.cma.2015.12.028.

Anderson M Winkler, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols. Permutation inference for the general linear model. *Neuroimage*, 92: 381–397, 2014.

Simon Wood. Package 'mgcv'. *R package version*, 1:29, 2015.

Simon N Wood. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114, 2003.

Simon N Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

Simon N Wood, Mark V Bravington, and Sharon L Hedley. Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):931–955, 2008.

Luo Xiao. Asymptotic theory of penalized splines. *Electronic Journal of Statistics*, 13(1): 747–794, 2019.

Shan Yu, Guannan Wang, Li Wang, Chenhui Liu, and Lijian Yang. Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*, 2019.

Yan Yu and David Ruppert. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460):1042–1054, 2002.

Supplementary material - Nonparametric tests for semiparametric regression models

Federico Ferraccioli[1, 2], Laura M. Sangalli[2], and Livio Finos[3*]

[1]*Dept. of Statistical Science, University of Padova, Italy*

[2]*MOX - Dept. of Mathematics, Politecnico di Milano, Italy*

[3*]*Corresponding author, DPSS, University of Padova, Italy*

8   Additional results

In this supplementary material we include some additional results. In particular, in Section 8.1 we explore the robustness/sensitivity of the considered tests to different specifications of the smoothing parameter. In Section 8.2 we illustrate the behavior of the tests for different sample sizes and different number of basis functions. In Section 8.3 we explore the case of multiple covariates. Finally, in Section 8.4 we investigates the performances of the considered tests when data are generated from a fully parametric model.

8.1   Sensitivity to the smoothing parameter

In this section we show the sensitivity of the considered tests to different values of the smoothing parameter λ. We generate the data as in Simulation 1, case (c), in Section 5.1
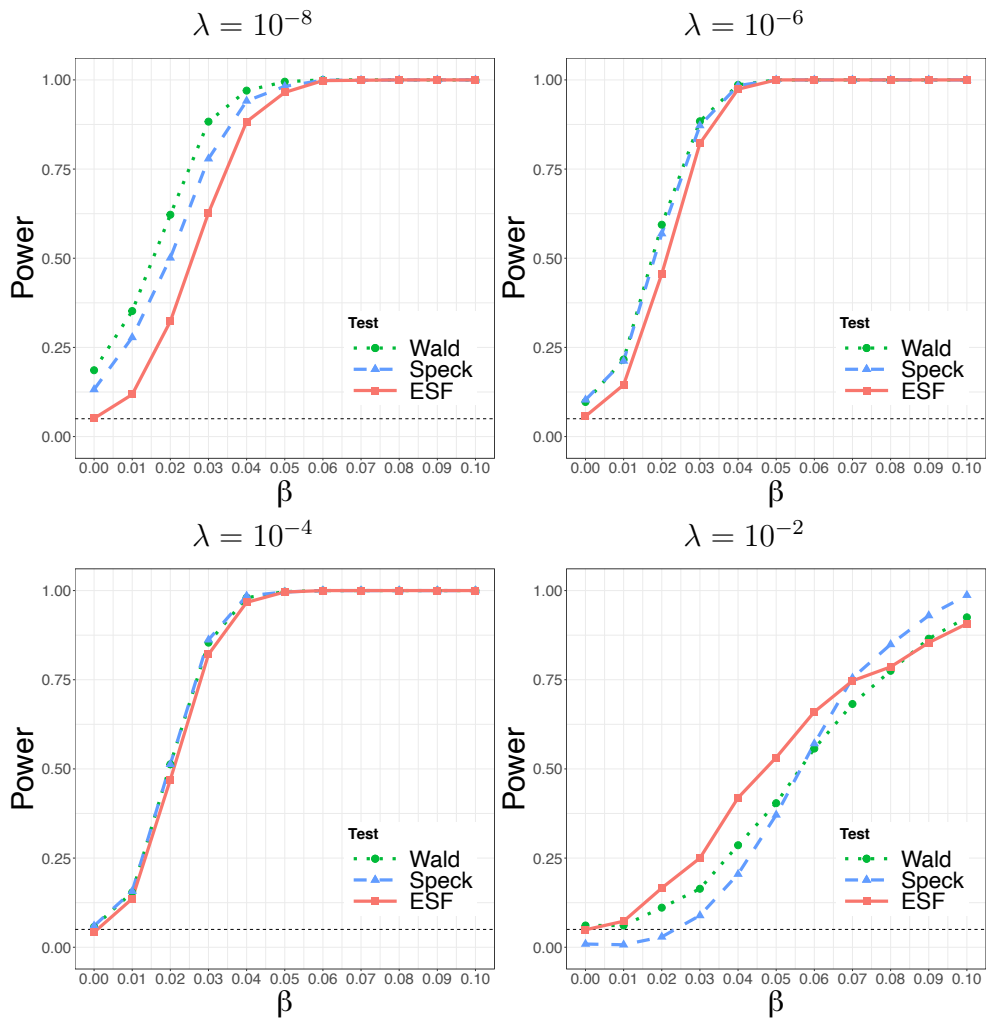
Figure 4: Simulation 1, case (c), Section 5.1 of the paper; different smoothing parameters $\lambda = 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}$. Power of Wald test (green dotted), Speckman (cyan dashed) and the proposed eigen sign-flip (red solid).

of the paper. We then obtain the estimates setting four different values of the smoothing parameter ($\lambda = 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}$) and perform the tests. The corresponding results, shown in Figure 4, demonstrate a good robustness of the proposed eigen sign-flip test to the choice of the smoothing parameter, differently from the parametric alternatives, Wald and Speckman, that are instead more sensitive, possibly due to the incorrect estimation of the variance. In particular, when λ is too small (top left panel of Figure 4), the Wald and Speckman tests show under-conservative behavior, with poor control of the Type-I error, while instead the proposed eigen sign-flip test maintains very good control of the Type-I error. When λ is too large (bottom right panel of the same figure), all the tests tend to lose power, with the drop in power being more severe for Wald and Speckman tests.

8.2 Different sample sizes and number of basis functions

In this section we show the performances of the considered tests for different sample sizes and number of basis functions. We generate the data as in Simulation 1, case (c), in Section 5.1 of the manuscript, with different sample sizes, $n = 200, 500$. We then obtain the estimates using different numbers of basis functions, $K = 20, 200$, and finally perform the considered tests. The corresponding results are shown in Figure 5. The three considered tests behave similarly. In particular, comparing the top panels in Figure 5, corresponding to $n = 200$, to the bottom panels of the same figure, corresponding to $n = 500$, we see a clear increase in power for all tests, as n increases, both for $K = 20$ and for $K = 200$. Comparing the left panels in Figure 5, corresponding to $K = 20$ bases, to the right panels of the same figure, corresponding to $K = 200$ bases, we observed a loss in power for all methods for smaller sample sizes ($n = 200$ in the top panels), whilst we do not observe any appreciable loss in power for larger sample sizes ($n = 500$ in the bottom panels). There is no particular difference in terms of power and Type-I error control with respect to K.
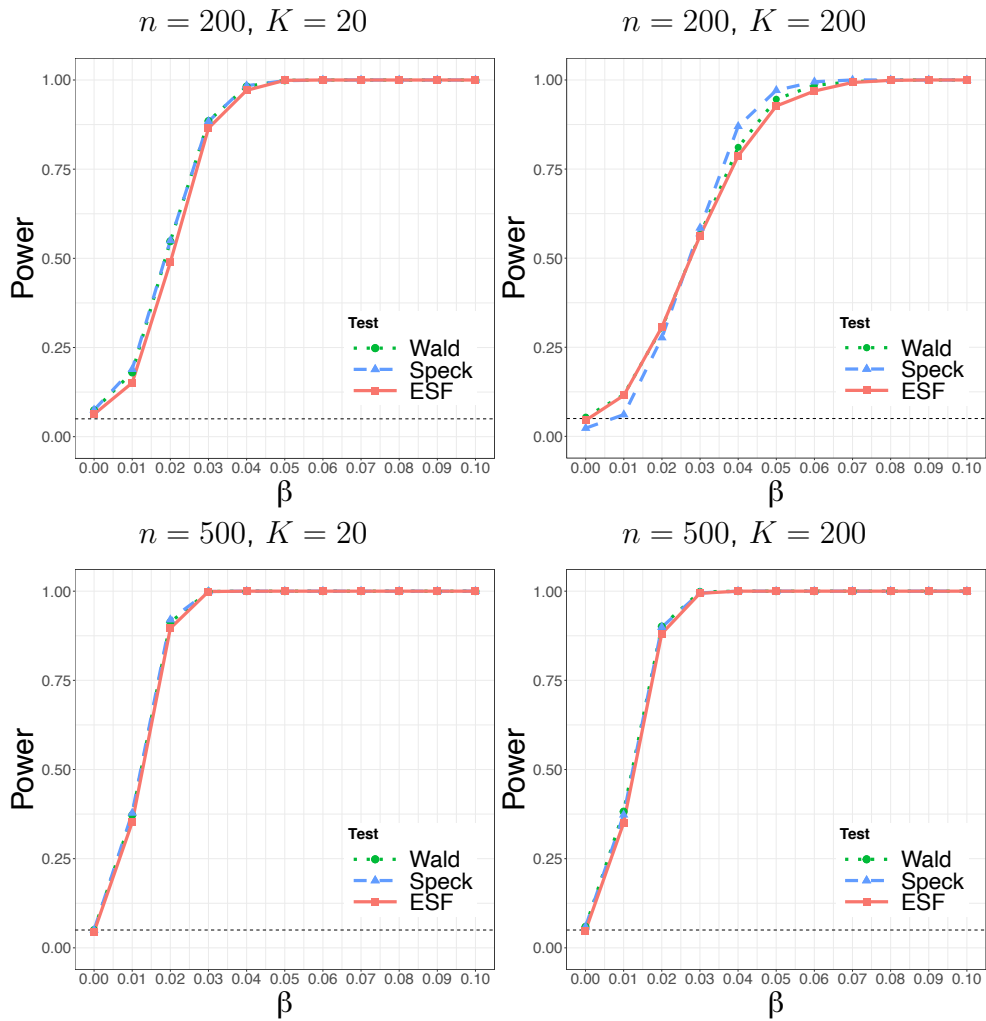
Figure 5: Simulation 1, case (c); different sample sizes, $n = 200$ (top panels) and $n = 500$ (bottom panels), and different numbers of bases $K = 20$ (left panels) and $K = 200$ (right panels). Power of Wald test (green dotted), Speckman (cyan dashed) and the proposed eigen sign-flip (red solid).

8.3 Multiple covariates

In this section we show the performances of the considered tests in the case of multiple covariates. Specifically, we include in the model two covariates, we hence test one covariate (corresponding to x_i in Section 4.4), considering the other covariate as nuisance variable (corresponding to z_i in Section 4.4). In the top left panel of Figure 6 of this response (corresponding to Figure 7 of the new Supplementary Material) we show the results obtained when generating data using (a) and (b) as covariates, and hence testing (a) as covariate of interest (playing the role of x_i in Section 4.4), while considering (b) as nuisance (playing the role of z_i in Section 4.4). In the top right panel of Figure 6, we show the results obtained when generating data using (a) and (b) as covariates, and hence testing (b) as covariate of interest, while considering (a) as nuisance. In the bottom left panel of Figure 6 we show the results obtained when generating data using (c) and (d) as covariates, and hence testing (c) as covariate of interest, while considering (d) as nuisance. In the bottom right panel of Figure 6 we show the results obtained when generating data using (c) and (d) as covariates, and hence testing (d) as covariate of interest, while considering (c) as nuisance. The results are analogous to those shown in the paper. In particular, the estimation of the nuisance parameter does not display an appreciable bias, thus not impairing the performances of the tests.

8.4 Data generation from a fully parametric model

In this section we explore the performances of the considered tests, under the same specifications of Simulation 1, cases (a)–(d), but setting $f = 0$ in the data generation. Specifically, we generate the data from a fully parametric model, but then estimate β and f, considering a semiparametric model. The results are shown in Figure 7. All three tests display the same power curves. In particular, the Wald and Speckman tests perform very well in all
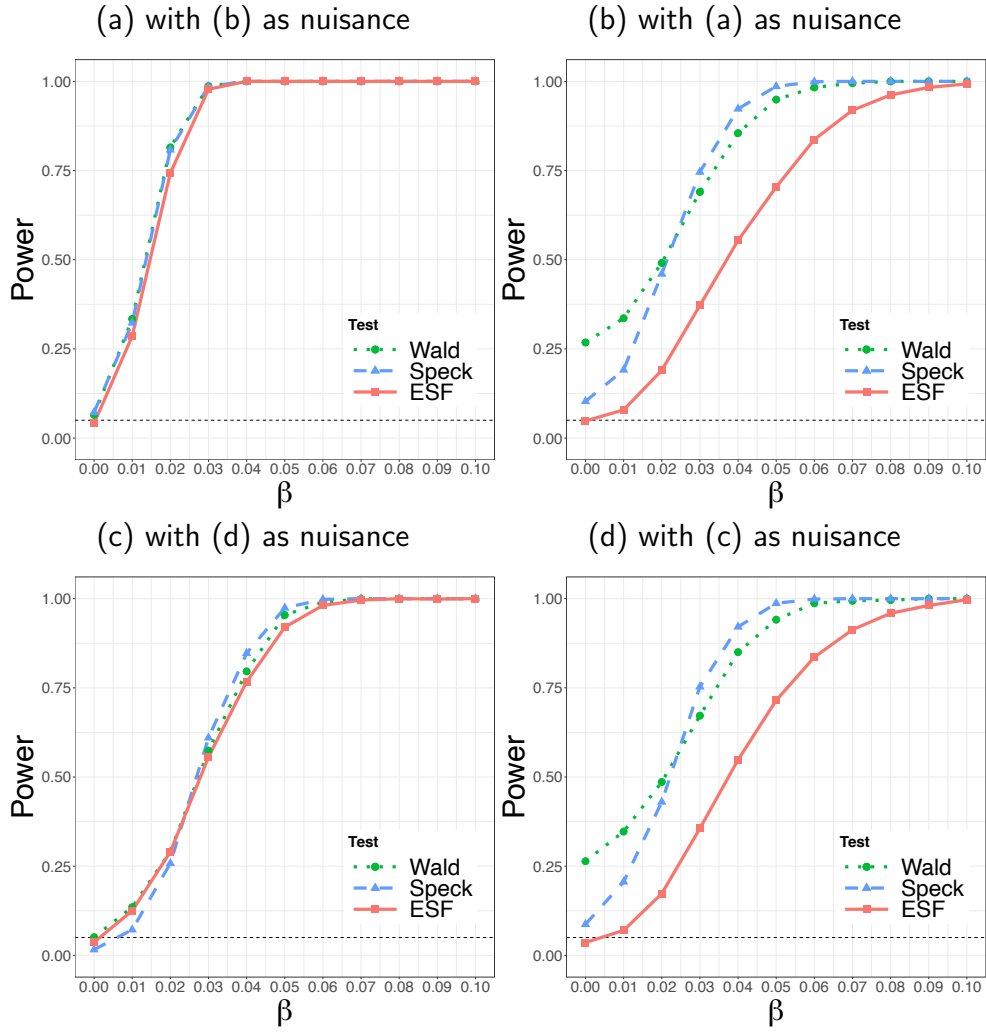
Figure 6: Simulation 1, multiple covariates. Top left: we include (a) and (b) as covariates in the data generation, and then test (a) as covariate of interest (corresponding to x_i in Section 4.4), while considering (b) as nuisance (corresponding z_i in Section 4.4). Top right: we include (a) and (b) as covariates in the data generation and the test (b) as covariate of interest, while considering (a) as nuisance. Bottom left: we include (c) and (d) as covariates in the data generation, and then test (c) is as covariate of interest, while considering (d) as nuisance. Bottom right: we include (c) and (d) as covariates in the data generation, and then test (d) as covariate of interest, while considering (c) as nuisance.
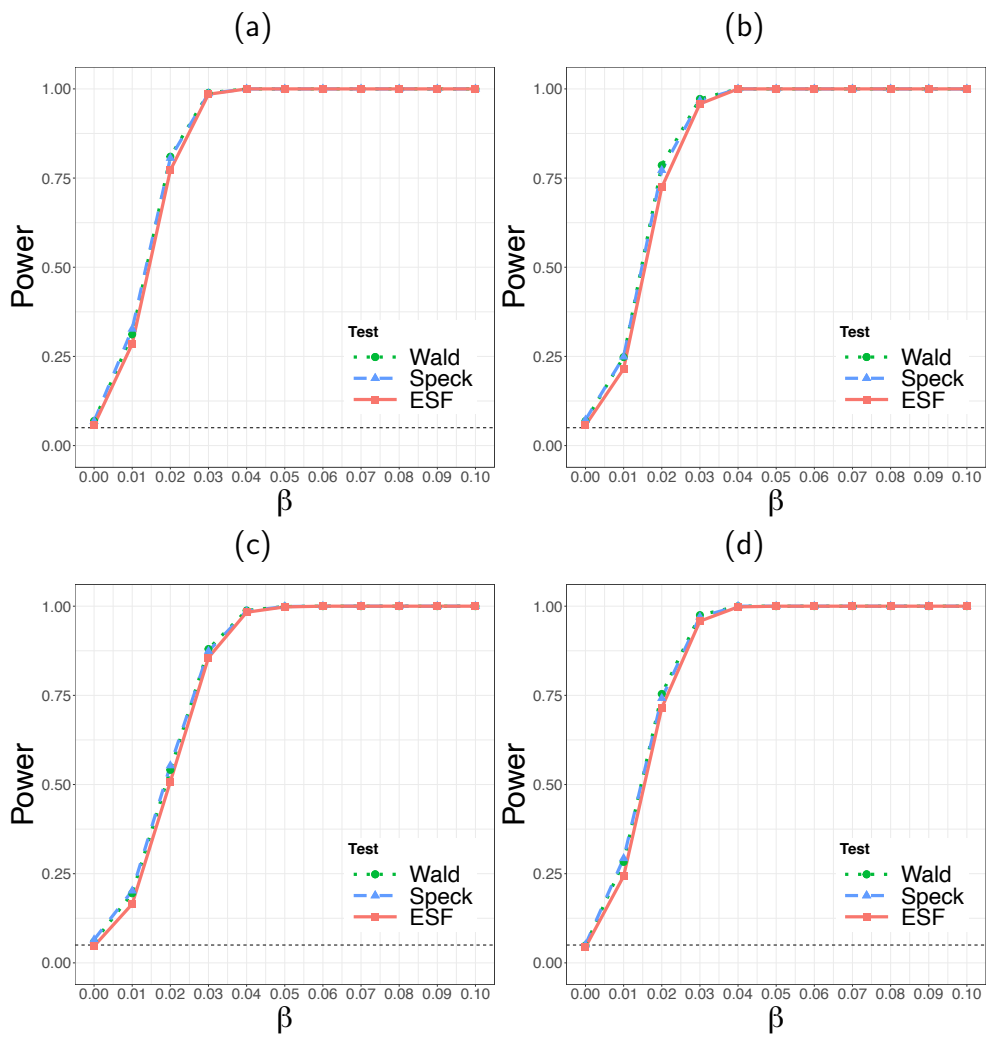
6

Figure 7: Simulation 1, cases (a)–(d), setting $f = 0$ in the data generation (fully parametric model) and obtaining the estimates using the semiparametric model. Power of Wald test (green dotted), Speckman (cyan dashed) and the proposed eigen sign-flip (red solid).

7

cases, with a good control of type-I error, since they are not affected by the nonparametric component; this is due to the good estimation of f, close to the null function. Noticeably, the proposed eigen sign-flip test displays the same performances of the parametric alternatives, without any appreciable loss in power.

MOX Technical Reports, last issues

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

85/2023 Arnone, E.; De Falco, C.; Formaggia, L.; Meretti, G.; Sangalli, L.M.
Computationally efficient techniques for Spatial Regression with Differential Regularization

83/2023 Cavinato, L.; Massi, M.C.; Sollini, M.; Kirienko , M.; Ieva, F.
Dual adversarial deconfounding autoencoder for joint batch-effects removal from multi-center and multi-scanner radiomics data

81/2023 Buchwald, S.; Ciaramella, G; Salomon, J.; Sugny, D.
A SPIRED code for the reconstruction of spin distribution

82/2023 Pozzi, G.; Ciarletta, P.
Geometric control by active mechanics of epithelial gap closure

79/2023 Agosti, A.; Bardin, R.; Ciarletta, P.; Grasselli, M.
A diffuse interface model of tumour evolution under a finite elastic confinement

78/2023 Antonietti, P.F.; Bonizzoni, F.; Corti, M.; Dall'Olio, A.
Discontinuous Galerkin for the heterodimer model of prion dynamics in Parkinson's disease

77/2023 Fumagalli, I.; Corti, M.; Parolini, N.; Antonietti, P. F.
Polytopal discontinuous Galerkin discretization of brain multiphysics flow dynamics

76/2023 Ieva, F.; Galliani, G.; Secchi, P.
The impact of public transport on the diffusion of COVID-19 pandemie in Lombardy during 2020

74/2023 Pidò, S.; Pinoli, P.; Crovari, P.; Ieva, F.; Garzotto, F.; Ceri, S.
Ask Your Data—Supporting Data Science Processes by Combining AutoML and Conversational Interfaces

75/2023 Archetti, A.; Ieva, F.; Matteucci, M.
Scaling survival analysis in healthcare with federated survival forests: A comparative study on heart failure and breast cancer genomics