

# On constrained smoothing and out-of-range prediction using $P$ -splines: A conic optimization approach



Manuel Navarro-García<sup>a,b</sup>, Vanesa Guerrero<sup>a,\*</sup>, María Durban<sup>a</sup>

<sup>a</sup> Department of Statistics, Universidad Carlos III de Madrid, Calle Madrid 126, Getafe 28901, Spain

<sup>b</sup> Komorebi AI Technologies, Avenida General Perón 26, Planta 4, Madrid 28020, Spain

## ARTICLE INFO

### Article history:

Received 5 June 2022

Revised 9 September 2022

Accepted 31 October 2022

### Keywords:

Data science

Penalized splines

Conic optimization

Smoothing

Prediction

## ABSTRACT

Decision-making is often based on the analysis of complex and evolving data. Thus, having systems which allow to incorporate human knowledge and provide valuable support to the decider becomes crucial. In this work, statistical modelling and mathematical optimization paradigms merge to address the problem of estimating smooth curves which verify structural properties, both in the observed domain in which data have been gathered and outwards. We assume that the curve to be estimated is defined through a reduced-rank basis ( $B$ -splines) and fitted via a penalized splines approach ( $P$ -splines). To incorporate requirements about the sign, monotonicity and curvature in the fitting procedure, a conic programming approach is developed which, for the first time, successfully conveys out-of-range constrained prediction. In summary, the contributions of this paper are fourfold: first, a mathematical optimization formulation for the estimation of non-negative  $P$ -splines is proposed; second, previous results are generalized to the out-of-range prediction framework; third, these approaches are extended to other shape constraints and to multiple curves fitting; and fourth, an open source Python library is developed: `cpsplines`. We use simulated instances, data of the evolution of the COVID-19 pandemic and of mortality rates for different age groups to test our approaches.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

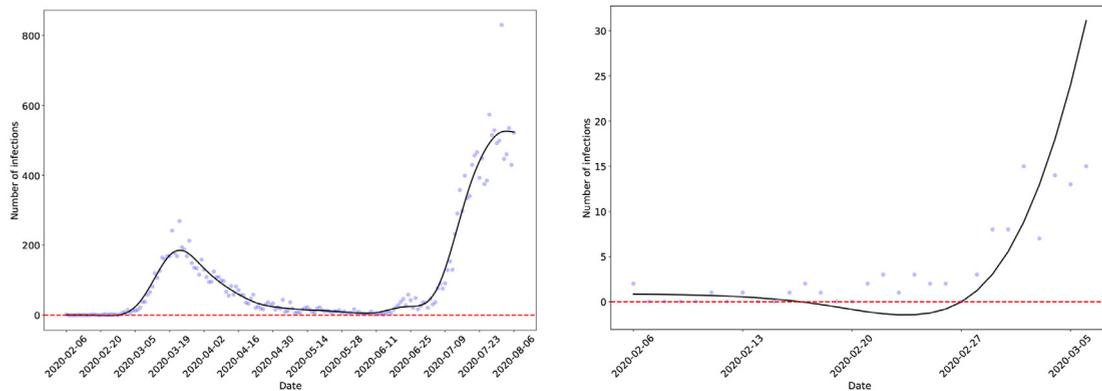
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

In many research areas, studying the relationship between a set of regressors and a response variable requires to estimate curves which satisfy certain properties. Examples include, for instance, the estimation of the number COVID-19 cases for a given date (non-negative values), probability densities (non-negative and integrate to one), economies of scale (convex and non-increasing because production at a larger scale are achieved at a lower cost), or mortality rates (higher values for older age groups). To carry out these kind of analyses, constrained estimation methods have to be developed. These constrained approaches must be flexible enough to incorporate additional requirements, such as non-negativity or other shape conditions. Furthermore, these requirements may need to be preserved when carrying out out-of-sample prediction. This work addresses the problem of constrained estimation and out-of-range prediction in the context of univariate regression combining the statistical framework of additive models [25] with conic optimization [8]. Thus, this work establishes

\* Corresponding author.

E-mail address: [vanesa.guerrero@uc3m.es](mailto:vanesa.guerrero@uc3m.es) (V. Guerrero).



(a) Number of daily COVID-19 cases in the Spanish region of Aragón from February 6th to August 6th 2020 (blue dots) and a smooth curve fitting the data (black line). The red dashed line corresponds to  $Y = 0$ .

(b) Zoom in of Figure 1a from February 6th to March 5th 2020 to show the negative number of cases predicted by the fitted curve.

**Fig. 1.** Evolution of COVID-19 pandemic in terms of the daily reported cases in the Spanish region of Aragón.

a new bond between statistics and mathematical optimization, as other recent works in the operations research literature [3,5,9,13,22,26].

One of the initial problems that motivated our work was the study of the evolution of the COVID-19 pandemic. The media has shared many graphs showing the number of daily positive cases, number of people in ICUs or number of fatalities together with a smooth curve capturing the general trend and diminishing the noise in the data. Figure 1a shows a scatter plot representing the daily number of COVID-19 cases in the Spanish region of Aragón from February 6th to August 6th 2020, together with a smooth curve to denoise the data. However, if this curve is fitted using an unconstrained approach, misleading results might be depicted, such as the negative number of COVID-19 cases from February 16th to February 27th 2020 (see Fig. 1b for a zoom in to show the negative behaviour of the curve in that range). Thus, addressing the problem of estimating smooth curves which satisfy, for instance, non-negativity conditions, becomes a must to avoid situations like the one shown in Fig. 1, which actually appeared on Spanish media in October 2020. Furthermore, being able to simulate different constrained prediction scenarios by incorporating expert knowledge is a challenge which has not been fully solved by existing short-term prediction approaches [1,2,23]. The methodology proposed in this paper allows the use to constrain the out-of-range predictions to emulate, for example, the evolution of the pandemic under different conditions such as, the growth rate during the second wave doubles or halves the one of the first wave.

A second motivating example for addressing constrained smoothing and prediction is related to demography. Most of the existing models for the graduation of mortality have to be biologically reasonable in the sense that, for example, mortality rates increase with age [10]. However, existing approaches [7] do not guarantee reasonable future predictions which are biologically consistent. Several attempts have been taken to remedy this problem. For example, [15] and [19] proposed smooth variants of previous models, and more recently [11] incorporated observed demographic information from the past years into the model. Nevertheless, none of these models impose directly the biological constraints needed, and, therefore, they cannot guarantee their fulfillment.

Estimating smooth curves which verify structural properties (both in the observed domain in which data have been gathered and outwards) without altering the fitting of the curve to the data too much becomes a very complex problem from the statistical point of view. This is thus the reason why we turn to mathematical optimization. The need to model complex data using a challenging statistical model is fulfilled by a conic optimization approach in this work. In particular, we focus on the penalized regression spline approach [20], frequently called  $P$ -splines, to estimate one (or more) unknown functions which relate a single predictor to a response variable.  $P$ -splines are based on a reduced-rank basis representation of the function to be estimated, and a modification of the likelihood function by adding a penalty term to control the smoothness of the fit. There are many choices of basis and penalties [21]. With respect to the basis choice, we focus on the use of  $B$ -splines [18], which consist of polynomial pieces connected by a set of knots in a continuous and differentiable way. Regarding the penalty, a discrete one is chosen, which is based on the adjacent coefficients of the  $B$ -spline basis. There are two main reasons for our choices: i) the good numerical properties of  $B$ -splines, and ii) the computational advantages of using local basis functions and discrete penalties without compromising the quality of fit of the estimated curve.

In the context of  $P$ -splines which aim to satisfy structural properties, most approaches have been focused on shape constraints. For instance, [6] proposed the use of additional asymmetric discrete penalties trying to enforce the constraints, and

[34] uses ad-hoc reparametrization of the  $B$ -spline basis and different penalties according to the requirements needed to be met. The aforementioned approaches yield sufficient but not necessary conditions about monotonicity and curvature of the estimated curves and are unable to address the non-negativity requirement. In [27] the constrained  $P$ -splines are computed via a weighted projection onto a polyhedral convex cone, which is a subset of the model space for the unconstrained penalized spline, and necessary and sufficient conditions for monotonicity and curvature requirements are provided. However, this approach cannot handle the non-negativity requirement and nor be generalized to a spline basis of arbitrary degree. Finally, the non-negativity requirement using  $P$ -splines is stated in Xia and Alizadeh [37] and extended to impose monotonicity and curvature conditions in Papp [32], Papp and Alizadeh [33], which use Bernstein polynomials instead, by means of a conic optimization approach. However, these works consider a definition of the knot sequence, which is needed to define the basis functions, that may provide undesirable effects in the boundaries of the estimated function. Furthermore, their approach cannot be used for carrying out (constrained) out-of-range prediction.

In summary, the contributions of this work are fourfold. First, the relationship between the coefficients of the polynomials which form the fitted function with the coefficients in its basis representation using  $B$ -splines is formally stated. An appropriate knot sequence is used, which avoids getting an undesirable behaviour of the fitted function in the boundaries of the domain and allows us to appropriately address the (constrained) out-of-range prediction problem. This result allows us to model the problem of fitting a non-negative function to a set of observations and beyond using a conic programming approach thanks to a characterization of non-negative polynomials in Bertsimas and Popescu [4]. Contrary to [6,34], our approach provides necessary and sufficient conditions for the estimation of a non-negative curve. Moreover, an arbitrary basis degree can be employed to guarantee these conditions, unlike [27]. Although the result in Bertsimas and Popescu [4] was first used in Monteiro et al. [28], the use of reduced-rank basis functions allows us to significantly reduce the number of decision variables in their optimization model as well as avoiding the knot selection problem thanks to our penalization term. Second, the problem of constrained out-of-range prediction, either forward or backward, using  $P$ -splines is successfully addressed for the first time, as far as the authors are aware. Following the approaches in Carballo et al. [12], Currie et al. [16], the values to be predicted are treated as missing data in the fitting procedure previously outlined, thus, yielding a general framework to ensure non-negativity for the out-of-range prediction. The third contribution consists of extending the non-negativity constraint in the constrained smoothing and prediction approaches to other requirements, such as monotonicity and curvature, and to the case in which multiple constrained functions of the same predictor and response need to be estimated simultaneously for different group of observations. These extensions allow us to simulate different prediction scenarios based, for instance, on past information. The fourth and last contribution is the development of an open source Python library, `cpsplines`<sup>1</sup>, which contains the implementations of all the methodologies developed in this paper.

The rest of this paper is organized as follows. Section 2 outlines the basic definitions and properties of  $P$ -splines for smoothing and out-of-range prediction, which are needed to develop our approach. Our conic programming approach for non-negative constrained  $P$ -splines is described in Section 3, whereas some extensions are presented in Section 4. Finally, in Section 5 the approaches presented in this paper are illustrated with simulated data as well as in real applications in the context of the COVID-19 pandemic evolution and in a demographic example. The paper concludes with some final remarks and future lines of research in Section 6.

## 2. Preliminaries

This section is devoted to introduce basic definitions and properties of  $P$ -splines using a basis of  $B$ -splines [18]. First, we focus on describing the estimation of a smooth curve which fits observed data, and then we describe how to perform out-of-range forward prediction using  $P$ -splines. We point out that the backward case would be analogous.

### 2.1. Smoothing $P$ -splines

Let us assume that a set of observations  $\{(x_i, y_i)\}$ ,  $i = 1, \dots, n$  is available, where  $x_i$  and  $y_i$  refer to the  $i$ th observation of the continuous covariate  $X$  and the response variable  $Y$  to be predicted, respectively. Without loss of generality, we assume that it is ordered in an increasing way with respect to the observations of  $X$ , namely  $x_1 \leq x_2 \leq \dots < x_n$ . Our aim is to estimate a function  $f : [x_1, x_n] \subset \mathbb{R} \rightarrow \mathbb{R}$  such that

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where  $\varepsilon_i \in \mathbb{R}$  is an error term.

Estimating  $f$  is, in general, challenging and many possibilities exist. In this work, the penalized regression smoothing spline approach, also known as  $P$ -splines [20], is used to estimate  $f$  in (1).  $P$ -splines consist of a basis function approach using splines for regression, together with a penalization term. Although many different spline bases can be used, this work focuses on the so-called  $B$ -splines [18]. The main reason behind this choice is their good numerical properties and their extended use by the scientific community.

<sup>1</sup> <https://github.com/ManuelNavarroGarcia/cpsplines>

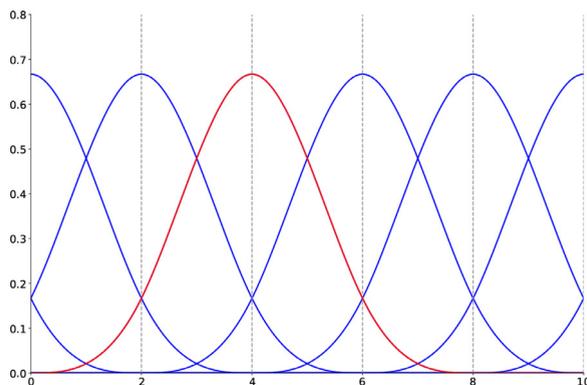


Fig. 2. Example of a basis of cubic ( $d = 3$ )  $B$ -splines with equally spaced knots. The dashed vertical lines show the position of the knots.

$B$ -splines are unidimensional non-negative basis functions constructed from polynomial pieces which join in a differentiable way at certain values: the knots. Given a knot sequence  $\mathbf{t} = \{t_j\}_{j=1}^m$ , i.e. an increasing and equally-spaced sequence of  $m \geq 1$  real numbers, a  $B$ -spline basis function of degree  $d$  is defined recursively for each  $j = 1, \dots, m$  as

$$B_{j,d,\mathbf{t}}(x) = \frac{x - t_j}{t_{j+d} - t_j} B_{j,d-1,\mathbf{t}}(x) + \frac{t_{j+d+1} - x}{t_{j+d+1} - t_{j+1}} B_{j+1,d-1,\mathbf{t}}(x), \tag{2}$$

where

$$B_{j,0,\mathbf{t}}(x) = \begin{cases} 1 & \text{if } x \in [t_j, t_{j+1}), \\ 0 & \text{otherwise.} \end{cases}$$

Although the knot sequence  $\mathbf{t}$  is assumed to be increasing and equally spaced for simplicity, all the results presented in this work also hold for non-decreasing and unevenly spaced knots. In particular, for a knot sequence consisting of  $m$  different elements,  $m - d - 1$   $B$ -splines of degree  $d$  can be generated. Figure 2 illustrates an example of a basis of cubic ( $d = 3$ )  $B$ -splines with equally spaced knots located at the vertical dashed lines. The  $B$ -spline drawn in red shows the overlapping nature of the basis in the interval between knots, i.e. there are four basis functions defined in each interval defined by two contiguous knots which are non-zero.

In this paper we work with cubic  $B$ -splines, and thus, we assume  $d = 3$  hereinafter. This choice is not arbitrary: cubic splines guarantee a notable level of smoothness, and even though higher degrees would provide smoother curves, the increment in the computational effort to compute them is usually a too high price to pay to end up with indistinguishable results. Nevertheless, we emphasize that the methodology we propose is also valid for  $d > 0$ . Therefore, we assume that the function  $f$  in (1) to be estimated in a domain  $[x_1, x_n]$  can be approximated by a function defined as a linear combination of cubic  $B$ -splines. Let the domain  $[x_1, x_n]$  be divided into  $k$  equal intervals by  $k + 1$  knots, usually called internal knots. Since, by definition, each interval will be covered by  $d + 1$  non-zero  $B$ -splines of degree  $d$  and, in our case  $d = 3$ , then  $k + 3$   $B$ -splines are needed to ensure that each interval in which the domain is split is covered by four non-zero cubic  $B$ -splines as in Fig. 2, and thus, avoid undesired effects in the boundaries because of a deficient number of basis functions covering the extreme intervals (the boundaries) as in Papp [32], Papp and Alizadeh [33], Xia and Alizadeh [37]. Therefore, a knot sequence consisting of  $k + 7$  knots is needed, i.e.  $\mathbf{t} = \{t_q\}_{q=1}^{k+7}$  such that  $t_4 = x_1$  and  $t_{k+4} = x_n$ . Then,  $f$  in (1) is to be estimated in  $[x_1, x_n]$  by means of a linear combination  $k + 3$  cubic  $B$ -splines defined over the knot sequence  $\mathbf{t}$ , yielding a curve  $S$  as

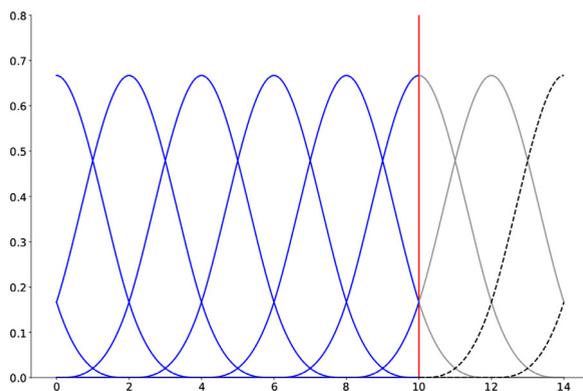
$$S(x) = \sum_{j=1}^{k+3} \theta_j B_{j,3,\mathbf{t}}(x). \tag{3}$$

Then, the values of the coefficients  $\theta_j$ ,  $j = 1, \dots, k + 3$  in (3) are sought so that the least squares criterion is optimized as follows:

$$\min_{\theta_j \in \mathbb{R}} \sum_{i=1}^n \left( y_i - \sum_{j=1}^{k+3} \theta_j B_{j,3,\mathbf{t}}(x_i) \right)^2. \tag{4}$$

$j=1, \dots, k+3$

The problem in (4) is a convex quadratic unconstrained program for which a closed formula for its optimal solution exists. The smoothness of the estimated fitted curve obtained from (4) depends on the number and position of the knots. Thus, this choice becomes crucial to get a good and smooth enough fit. In order to relax the role played by the knots in the shape of the estimated curve,  $P$ -splines are considered instead. The idea is to use a large number of knots and control the smoothness of the curve by adding a penalty on the coefficients  $\theta_j$ ,  $j = 1, \dots, k + 3$  to the objective function in (4). There



**Fig. 3.** Extended basis of cubic  $B$ -splines with evenly spaced knots. The red vertical line indicates the point in which the extension starts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are different choices for this penalty, being the one based on penalizing the integral of the squared second derivative of the fitted function a common choice [30,31]. In this work, a discrete approximation for that integral proposed in Eilers and Marx [20] is used, which relies on finite differences on adjacent coefficients representing the second derivative of  $S$  in (3). Then, the  $P$ -splines approach to estimate function  $f$  in (1) using a basis of  $B$ -splines is stated as the following convex quadratic unconstrained optimization problem, which depends on the smoothing parameter  $\lambda \geq 0$  :

$$\min_{\substack{\theta_j \in \mathbb{R} \\ j=1, \dots, k+3}} \sum_{i=1}^n \left( y_i - \sum_{j=1}^{k+3} \theta_j B_{j,3,t}(x_i) \right)^2 + \lambda \sum_{j=3}^{k+3} (\theta_j - 2\theta_{j-1} + \theta_{j-2})^2. \tag{5}$$

The penalty solves the problem of choosing the number and position of the knots, and it is the smoothing parameter  $\lambda$  the one controlling the smoothness of the fitted curve: imposing  $\lambda = 0$  reduces (5) to (4), and thus the optimal curve is that obtained by the least squares criterion, which could lead to very wiggly curves. On the other hand, if  $\lambda \rightarrow +\infty$ , the second derivative of  $S$  in (3) must be zero in order to reduce the second term in the objective, thus yielding a linear approximation to  $f$ , which is the least possible wiggly curve. Therefore, the choice of the position of the knots, provided that the number of knots is large enough, becomes much less relevant when the stated penalty is imposed.

### 2.2. Out-of-range prediction using $P$ -splines

As we mentioned earlier, one of the reasons that motivated our work was to solve the problem of predicting values out of the initial range of the observed sample of the covariate  $X$ , subject to certain constraints. In particular, we will focus on forward prediction, though the results presented in what follows can be easily derived for backward prediction. Hence, we suppose that  $\{x_i\}, i = n + 1, \dots, n + n_p$  is a set of values arranged in increasing order outside the convex hull of the interval  $[x_1, x_n]$  and to the right of  $x_n$ , this is  $x_i > x_n, i = n + 1, \dots, n + n_p$ .

Observe that the interval  $[x_1, x_{n+n_p}]$  is not contained in the interval  $[t_4, t_{k+4}]$ , defined by the knot sequence  $\mathbf{t} = \{t_q\}_{q=1}^{k+7}$  used to cover the range of the observations  $[x_1, x_n]$ . Therefore,  $\mathbf{t}$  has to be extended to cover the range defined by  $[x_1, x_{n+n_p}]$ , and thus, to construct a new extended  $B$ -spline basis. To this end, we define the extended knot sequence  $\mathbf{t}_+ = \{t_q\}_{q=1}^{k+k_p+7}$  by adding  $k_p$  more knots at the end of  $\mathbf{t}$ , while preserving the same step length as the knots in  $\mathbf{t}$ , until the inequality  $x_{n+n_p} \leq t_{k+k_p+4}$  holds. Figure 3 illustrates an extended  $B$ -splines basis of the one depicted in Fig. 2, where the extended domain starts at the red vertical line. We point out that some of the added  $B$ -splines are a continuation of the existing ones (shown in solid grey curves) while others take their first non-zero values on the extended domain (shown in black dashed lines).

Thus, thanks to the aforementioned extended basis, an extended version of  $S$  in (3) can be defined. This extension, which allows the out-of-range prediction,  $S_+$ , is stated as a linear combination of  $k + k_p + 3$   $B$ -splines defined through the coefficients  $\theta_j, j = 1, \dots, k + k_p + 3$  as

$$S_+(x) = \sum_{j=1}^{k+k_p+3} \theta_j B_{j,3,t_+}(x). \tag{6}$$

In order to estimate the unknown coefficients  $\theta_j$  we follow the approach in Carballo et al. [12] and [16], in which the values to be predicted are treated as missing data in the fitting procedure. This yields a weighted penalized least squares problem where the missing values, i.e.  $y_i$  for  $i = n + 1, \dots, n + n_p$ , are given zero weights. Then, the extension of  $P$ -splines

for forward prediction, this is the prediction version of problem (5), is stated as

$$\min_{\theta_j \in \mathbb{R}} \sum_{i=1}^{n+n_p} v_i \left( y_i - \sum_{j=1}^{k+k_p+3} \theta_j B_{j,3,t_+}(x_i) \right)^2 + \lambda \sum_{j=3}^{k+k_p+3} (\theta_j - 2\theta_{j-1} + \theta_{j-2})^2, \tag{7}$$

$j=1, \dots, k+k_p+3$

where the weights  $v_i = 1$  for  $i = 1, \dots, n$ , and  $v_i = 0$  for  $i = n + 1, \dots, n + n_p$ . Observe that, due to the definition of these weights, all the work done for the out-of-range prediction is performed by the penalty term since the first term in (7) equals to the first term in (5). In fact, given a penalty of order  $\ell$ , the new coefficients are combinations of order  $\ell - 1$  of the last  $\ell$  fitted coefficients (see [12] for details). In our case,  $\theta_j, j = k + 4, \dots, k + k_p + 3$  are linear combinations of  $\theta_{k+2}$  and  $\theta_{k+3}$ .

### 3. Non-negative constrained P-splines: A conic programming approach

Once the unconstrained P-spline approach for estimating function  $f$  in (1) have been introduced, we turn to the constrained modelling problem. In Section 3.1, we present some theoretical results needed to impose non-negativity constraints on the estimation models (5) and (7), i.e. that the estimated curves  $S$  in (3) or  $S_+$  in (6) are non-negative. The constrained versions of these models are stated in Sections 3.2 and 3.3, respectively. Furthermore, as a byproduct, these results will allow us to impose also shape constraints based on the derivatives of the smooth function and the simultaneous estimation of constrained smooth curves for grouped observations (Section 4).

#### 3.1. Characterization of non-negative polynomials

Estimating the function  $f$  in model (1) using the approach in (5) (or (7) in the forward prediction framework) does not ensure that the estimated curve satisfies any requirements about, for instance, its sign or shape as shown in Fig. 1. In Section 1 we discussed about the existence of many situations in which these requirements are needed to provide the estimation with meaningful and coherent interpretations. In particular, one of those requirements that naturally arises in many real applications concerns the non-negativity of the estimated curve on its whole domain.

A straightforward way of fitting a non-negative curve  $S$  as stated in (3) (or  $S_+$  as in (6)) using the approach in (5) (resp. (7)) consists of constraining all the coefficients  $\theta_j$  defining the linear combination of B-splines to be non-negative. However, this approach might yield very poor fits and it is not a necessary condition for  $S$  (resp.  $S_+$ ) to be non-negative. As an alternative, and taking into account that  $S$  and  $S_+$  are cubic piecewise polynomials, we propose the use of a characterization of non-negative univariate polynomials to impose non-negativity constraints in (5) (resp. (7)) with the least possible impact on the fitted curve.

To do that, recall that each B-spline  $B_{j,3,t}(x)$  for  $j = 1, \dots, k + 3$  is non-zero on the interval  $[t_j, t_{j+4})$ , and it is composed of four cubic polynomials  $b_{j,q}(x)$ , each of them defined in the interval  $[t_q, t_{q+1})$ , with  $q = j, \dots, j + 3$ . Therefore, for  $x \in [t_q, t_{q+1})$  each B-spline is a cubic polynomial of the form

$$b_{j,q}(x) = \sum_{r=0}^3 g_{j,q,r} x^r,$$

where  $g_{j,q,r}$  are the coefficients of cubic ( $r = 3$ ), quadratic ( $r = 2$ ), linear ( $r = 1$ ) and the constant term ( $r = 0$ ) of  $b_{j,q}(x)$ . Then, the value of the function  $S$  in (3) for every  $x \in [t_q, t_{q+1})$  is

$$S(x) = \sum_{j=q-3}^q \theta_j b_{j,q}(x) = \sum_{j=q-3}^q \theta_j \sum_{r=0}^3 g_{j,q,r} x^r = \sum_{r=0}^3 \left( \sum_{j=q-3}^q \theta_j g_{j,q,r} \right) x^r.$$

Hence, in each interval  $[t_q, t_{q+1})$ ,  $S$  in (3) can be written as  $S(x) = \alpha_q x^3 + \beta_q x^2 + \gamma_q x + \delta_q, q = 4, \dots, k + 3$ , where

$$\alpha_q = \sum_{j=q-3}^q \theta_j g_{j,q,3}, \quad \beta_q = \sum_{j=q-3}^q \theta_j g_{j,q,2}, \quad \gamma_q = \sum_{j=q-3}^q \theta_j g_{j,q,1}, \quad \delta_q = \sum_{j=q-3}^q \theta_j g_{j,q,0}. \tag{8}$$

It is worth pointing out that the same result holds for a general degree  $d$ , and it will be used in Section 4 for quadratic and linear B-splines and in Supplementary Material S.1 for quartic basis functions. Furthermore, the same result holds for  $S_+$  in (6) using the extended knot sequence  $\mathbf{t}_+ = \{t_q\}_{q=1}^{k+k_p+7}$ .

The explicit representation of the coefficients in (8) allows us to use the characterization of non-negative polynomials in Bertsimas and Popescu [4]. This result, which is reproduced in Theorem 1, gives a necessary and sufficient condition for non-negative univariate polynomials of any degree in an interval of the form  $[a, b] \subset \mathbb{R}$ .

**Theorem 1.** (Proposition 1(d) in Bertsimas and Popescu [4]) *The polynomial  $g(x) = \sum_{r=0}^s c_r x^r$  satisfies  $g(x) \geq 0$  for all  $x \in [a, b]$  if and only if there exists a  $(s + 1) \times (s + 1)$  matrix  $\mathbf{Z} = [z_{ij}]_{i,j=0,\dots,s}$  such that*

$$\sum_{\substack{i,j=0,\dots,s \\ i+j=2q-1}} z_{ij} = 0, \quad q = 1, \dots, s,$$

$$\sum_{\substack{i,j=0,\dots,s \\ i+j=2q}} z_{ij} = \sum_{m=0}^q \sum_{r=m}^{s+m-q} c_r \binom{r}{m} \binom{s-r}{q-m} a^{r-m} b^m, \quad q = 0, \dots, s,$$

$$\mathbf{Z} \in \mathbb{S}^{s+1},$$

where  $\mathbb{S}^{s+1}$  is the cone of positive semidefinite matrices of dimension  $(s + 1) \times (s + 1)$ .

Thanks to the explicit relations in (8), the conditions stated in Bertsimas and Popescu [4] and reproduced in Theorem 1 can be incorporated into the optimization problems (5) and (7) to impose non-negativity of the estimated curves, yielding conic optimization models which are described hereinafter.

### 3.2. Non-negative smoothing P-splines

This section is devoted to reformulate the estimation problem in (5) to ensure that the curve  $S$ , given by (3) and approximating  $f$  in the prediction model (1), is non-negative.

In what follows some notation is introduced which is needed to state the constrained version of the problem in (5) as a conic optimization model. Let  $\mathbf{y}^\top = (y_1, \dots, y_n)$  be the  $n$ -dimensional vector containing the  $n$  observations of the response variable  $Y$  and let  $\mathbf{B} = [B_{j,3,t}(x_i)]_{i=1,\dots,n}^{j=1,\dots,k+3}$  be an  $n \times (k + 3)$  matrix containing the evaluations of the  $k + 3$  B-splines on the

$n$  observations of the covariate  $X$ . Let  $\mathbf{D}$  be a  $(k + 1) \times (k + 3)$  matrix defined as

$$\mathbf{D} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ & & \vdots & & & \ddots & & \vdots & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}.$$

The  $(k + 3)$ -dimensional vector  $\boldsymbol{\theta}^\top = (\theta_1, \dots, \theta_{k+3})$  contains the coefficients defining the curve  $S$  in (3), and thus the decision variables in (5). Observe that, for  $\lambda \geq 0$ , the previous definitions allow us to write the problem (5) in matrix form as

$$\min_{\boldsymbol{\theta}} -2\mathbf{y}^\top \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\theta}^\top (\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D}) \boldsymbol{\theta}$$

We point out that the constant term  $\mathbf{y}^\top \mathbf{y}$  has been left out in the objective function in

s.t.  $\boldsymbol{\theta} \in \mathbb{R}^{k+3}$ .

(9). This matrix notation is used hereinafter to state the constrained models. Furthermore, the convex quadratic optimization model in (9) is stated as a second-order cone program in Appendix A. Such reformulation is used in what follows to state the constrained approach in the form of a standard conic optimization model.

In order to impose that the curve  $S$  obtained from the solution of (9) is non-negative, the conditions of Theorem 1 for cubic polynomials are imposed over the coefficients  $\boldsymbol{\theta}$ . To do that, the characterization in (8) is used yielding the following conic optimization problem:

$$\begin{aligned} \min_{\mathbf{u}, \boldsymbol{\theta}, \mathbf{Z}_q} \quad & u - 2\mathbf{y}^\top \mathbf{B}\boldsymbol{\theta} \\ \text{s.t.} \quad & \langle \mathbf{H}_\ell, \mathbf{Z}_q \rangle_F = 0, \quad q = 4, 5, \dots, k + 3, \quad \ell = 1, 2, 3 \\ & \begin{pmatrix} \langle \mathbf{H}_4, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_5, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_6, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_7, \mathbf{Z}_q \rangle_F \end{pmatrix} = \mathbf{W}_q \mathbf{G}_q \boldsymbol{\theta}, \quad q = 4, 5, \dots, k + 3 \\ & ((\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D})^{1/2} \boldsymbol{\theta}, u, \frac{1}{2}) \in \mathcal{Q}_r^{k+5} \\ & \mathbf{Z}_q \in \mathbb{S}^4, \quad q = 4, 5, \dots, k + 3 \\ & \boldsymbol{\theta} \in \mathbb{R}^{k+3}, \end{aligned} \tag{10}$$

where  $\langle \mathbf{U}, \mathbf{V} \rangle_F = \text{Trace}(\mathbf{U}^\top \mathbf{V})$  denotes the Frobenius inner product, and matrices  $\mathbf{H}_\ell$ ,  $\mathbf{W}_q$  and  $\mathbf{G}_q$ ,  $\ell = 1, \dots, 7$  and  $q = 4, 5, \dots, k + 3$  arise from the conditions in Theorem 1, as well as the matrices  $\mathbf{Z}_q$ , which are decision variables in the optimization model. Observe that the feasible region of (10) involves the cone of positive semidefinite matrices of dimension

$4 \times 4, \mathbb{S}^4$ , and the rotated second-order cone,  $\mathcal{Q}_r^{k+5} = \{(\mathbf{x}, y, z) \in \mathbb{R}^{k+5} : \|\mathbf{x}\|_2^2 \leq 2yz, y \geq 0, z \geq 0\}$ . We refer the reader to Appendix B to browse through the expressions of the mentioned matrices. We point out that model (10) also can be adapted for the case in which  $S$  is required to be above a certain positive threshold. We refer the reader to Appendix B to see the details.

### 3.3. Non-negative out-of-sample prediction with P-splines

In order to address the non-negative forward prediction using P-splines we consider  $\{(x_i, y_i)\}, i = 1, \dots, n$  and  $\{x_i\}, i = n + 1, \dots, n + n_p$  as stated in Section 2.2. Let  $\mathbf{y} = (y_1, \dots, y_n)^\top$  be the observed values from the response variable  $Y$ , and  $\mathbf{y}_p = (y_{n+1}, \dots, y_{n+n_p})^\top$  the unknown values to be predicted. Analogously,  $\mathbf{x}$  and  $\mathbf{x}_p$  are defined for the predictor but with the difference that both  $\mathbf{x}$  and  $\mathbf{x}_p$  contain known values: on the one hand, the ones for which  $\mathbf{y}$  is known, and on the other hand, the values for which the out-of-range prediction is to be made. Therefore, we define

$$\mathbf{x}_+ = \begin{pmatrix} \mathbf{x} \\ \mathbf{x}_p \end{pmatrix}, \quad \mathbf{y}_+ = \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_p \end{pmatrix}.$$

Considering the extended knot sequence  $\mathbf{t}_+$  defined in Section 2.2,  $\lambda \geq 0$  and the vector  $\boldsymbol{\theta}_+ = (\theta_1, \dots, \theta_{k+k_p+3})^\top$ , the optimization problem in (7) can be rewritten in matrix form as

$$\begin{aligned} \min_{\boldsymbol{\theta}_+} \quad & -2\mathbf{y}_+^\top \mathbf{V} \mathbf{B}_+ \boldsymbol{\theta}_+ + \boldsymbol{\theta}_+^\top (\mathbf{B}_+^\top \mathbf{V} \mathbf{B}_+ + \lambda \mathbf{D}_+^\top \mathbf{D}_+) \boldsymbol{\theta}_+ \\ \text{s.t.} \quad & \boldsymbol{\theta}_+ \in \mathbb{R}^{k+k_p+3}, \end{aligned} \tag{11}$$

where  $\mathbf{B}_+$  and  $\mathbf{D}_+$  are the basis and the difference matrices as defined in Section 3.2, respectively, but extended to include the additional coefficients needed for forward prediction. Furthermore, matrix  $\mathbf{V}$  is a  $(n + n_p) \times (n + n_p)$  weight diagonal matrix with 0 entries if the data is missing, i.e. observed values of the regressor for which we are making the forward prediction, and 1 if the data is observed, i.e. observed values of the regressor and the response variable. The constant term  $\mathbf{y}_+^\top \mathbf{y}_+$  has been omitted in the objective function of problem (11).

The non-negativity in the fitting and forward prediction region, i.e. in the domain  $[x_1, x_{n+n_p}]$ , can be stated similarly to the approach in (10) and thanks again to Theorem 1 and the characterization in (8). The same matrices  $\mathbf{H}_\ell, \ell = 1, \dots, 7$  and, for  $q = 4, \dots, k + k_p + 3$ , the matrices  $\mathbf{W}_q$  and  $\mathbf{G}_{+,q}$  are also considered, where  $\mathbf{G}_{+,q}$  is an extended version of  $\mathbf{G}_q$  as described in Appendix B, and the reformulation in Appendix A. Then, the problem of estimating the function  $f$  in the regression Eq. (1) in the domain  $[x_1, x_{n+n_p}]$  is stated as the following conic optimization problem:

$$\begin{aligned} \min_{u, \boldsymbol{\theta}_+, \mathbf{Z}_q} \quad & u - 2\mathbf{y}_+^\top \mathbf{V} \mathbf{B}_+ \boldsymbol{\theta}_+ \\ \text{s.t.} \quad & \langle \mathbf{H}_\ell, \mathbf{Z}_q \rangle_F = 0, \quad q = 4, 5, \dots, k + k_p + 3, \quad \ell = 1, 2, 3 \\ & \begin{pmatrix} \langle \mathbf{H}_4, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_5, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_6, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_7, \mathbf{Z}_q \rangle_F \end{pmatrix} = \mathbf{W}_q \mathbf{G}_{+,q} \boldsymbol{\theta}_+, \quad q = 4, 5, \dots, k + k_p + 3 \\ & ((\mathbf{B}_+^\top \mathbf{V} \mathbf{B}_+ + \lambda \mathbf{D}_+^\top \mathbf{D}_+)^{1/2} \boldsymbol{\theta}_+, u, \frac{1}{2}) \in \mathcal{Q}_r^{k+k_p+5} \\ & \mathbf{Z}_q \in \mathbb{S}^4, \quad q = 4, 5, \dots, k + k_p + 3 \\ & \boldsymbol{\theta}_+ \in \mathbb{R}^{k+k_p+3}. \end{aligned} \tag{12}$$

## 4. Extensions to shape-constrained P-splines and multiple curves estimation

Thanks to the results obtained in Section 3, imposing conditions on the shape in the estimated curves, i.e. monotonicity and curvature, becomes straightforward. The details on how these requirements can be considered in the fitting procedure are given in Section 4.1. We point out that the extensions to shape-constrained P-splines to be described in this section are separate models and they can be combined in any way to obtain, for instance, non-negative and convex curve (which might indeed yield a non-monotone curve) or a monotone curve with no additional conditions about its sign or curvature. In addition, our methodology is extended in Section 4.2 to the case in which multiple curves have to be simultaneously constrained for groups of observed data. The challenge in this multiple setting consists of imposing a relative position in the curves, in the sense that these curves do not overlap. We point out that the following results are described in the case of within sample prediction, but their extensions to the out-of-sample prediction setting is straightforward once an extended basis of B-splines is considered as described in Sections 2.2 and 3.3. Furthermore, the methodology proposed in Section 3 and the extensions described hereinafter allow us to construct confidence bands using the standard bootstrap techniques [17]. This is illustrated with a synthetic dataset in Supplementary Material S.2.

### 4.1. Shape-constrained smoothing P-splines

In this section we describe how to impose other constraints on the curve  $S$  in (5), such as monotonicity or curvature. Thanks to the conditions on the smoothness of  $S$ , its derivative exists and it is again a piecewise polynomial, thus allowing us to apply Theorem 1 in higher order derivatives of  $S$  to enforce sign constraints on them. We will focus on the cases of non-decreasing curves and convex curves, but the extensions to non-increasing or concavity are straightforward.

#### 4.1.1. Monotonicity

Imposing that the curve  $S$  in (3) is non-decreasing follows from applying the results obtained in Section 3.2 to its first derivative. Since  $S$  has the form of a continuous and twice continuously differentiable piecewise cubic polynomial, its derivative is a continuous and once differentiable piecewise quadratic polynomial. Thus, Theorem 1 and an adapted version of the characterization in (8) for quadratic polynomials can be used to impose the non-negativity of the derivative of  $S$ , which equates to the non-decreasing condition under study.

Hence, given a set of observations and a knot sequence in the conditions described in Section 2.1, the problem of estimating a function  $f$  in the regression Eq. (1) by means of non-decreasing P-splines, using B-splines as basis functions as expressed in (3), is stated as the following conic optimization problem:

$$\begin{aligned}
 \min_{u, \boldsymbol{\theta}, \mathbf{Z}_q} \quad & u - 2\mathbf{y}^\top \mathbf{B}\boldsymbol{\theta} \\
 \text{s.t.} \quad & \langle \widehat{\mathbf{H}}_\ell, \mathbf{Z}_q \rangle_F = 0, \quad q = 4, 5, \dots, k+3, \quad \ell = 1, 2 \\
 & \begin{pmatrix} \langle \widehat{\mathbf{H}}_3, \mathbf{Z}_q \rangle_F \\ \langle \widehat{\mathbf{H}}_4, \mathbf{Z}_q \rangle_F \\ \langle \widehat{\mathbf{H}}_5, \mathbf{Z}_q \rangle_F \end{pmatrix} = \widehat{\mathbf{W}}_q \widehat{\mathbf{G}}_q \boldsymbol{\theta}, \quad q = 4, 5, \dots, k+3 \\
 & ((\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D})^{1/2} \boldsymbol{\theta}, u, \frac{1}{2}) \in \mathcal{Q}_r^{k+5} \\
 & \mathbf{Z}_q \in \mathbb{S}^3, \quad q = 4, 5, \dots, k+3 \\
 & \boldsymbol{\theta} \in \mathbb{R}^{k+3},
 \end{aligned} \tag{13}$$

where the matrices  $\widehat{\mathbf{H}}_\ell$ ,  $\widehat{\mathbf{G}}_q$  and  $\widehat{\mathbf{W}}_q$ , for  $\ell = 1, \dots, 5$  and  $q = 4, \dots, k+3$  are adapted versions of their counterparts stated in Section 3.2 (see Appendix C). We point out that the decision variables  $\mathbf{Z}_q \in \mathbb{S}^3$  arise from imposing non-negativity on a quadratic polynomial by means of Theorem 1.

The proposed methodology allows us to incorporate conditions about the growth rate of the estimated curve, both in the domain in which values of the covariate and the response are observed or in the out-of-range interval in which the prediction is performed. Thus, this approach enables us to simulate, for instance, different forward prediction scenarios setting different thresholds for the derivative of  $S_+$  in (6).

Solving (13) yields a curve  $S$ , as stated in (3), which approximates  $f$  in the regression model (1), that is non-decreasing. If, for instance, we want to fit a curve  $S$  which is non-negative and non-decreasing, the models (10) and (13) have to be merged provided that two variants of decision variables  $\mathbf{Z}_q$  appear in the new model: the ones involved in the sign conditions ( $\mathbf{Z}_q^1 \in \mathbb{S}^4$ ) and those involved in the monotonicity ones ( $\mathbf{Z}_q^2 \in \mathbb{S}^3$ ).

#### 4.1.2. Curvature

Imposing that the curve  $S$  in (3) is convex is equivalent to enforce that its second derivative is non-negative. Thus, the reasoning applied in Section 3.2 can also be used to model this shape constraint. In this case, as the second derivative of  $S$  is a continuous piecewise polynomial of degree one, an adapted version of the characterization in (8) for degree one is required.

Hence, the problem of estimating a function  $f$  in the regression Eq. (1) by means of convex P-splines, using B-splines as basis functions as expressed in (3), is stated as the following conic optimization problem:

$$\begin{aligned}
 \min_{u, \boldsymbol{\theta}, \mathbf{Z}_q} \quad & u - 2\mathbf{y}^\top \mathbf{B}\boldsymbol{\theta} \\
 \text{s.t.} \quad & \langle \overline{\mathbf{H}}_1, \mathbf{Z}_q \rangle_F = 0, \quad q = 4, 5, \dots, k+3 \\
 & \begin{pmatrix} \langle \overline{\mathbf{H}}_2, \mathbf{Z}_q \rangle_F \\ \langle \overline{\mathbf{H}}_3, \mathbf{Z}_q \rangle_F \end{pmatrix} = \overline{\mathbf{W}}_q \overline{\mathbf{G}}_q \boldsymbol{\theta}, \quad q = 4, 5, \dots, k+3 \\
 & ((\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D})^{1/2} \boldsymbol{\theta}, u, \frac{1}{2}) \in \mathcal{Q}_r^{k+5} \\
 & \mathbf{Z}_q \in \mathbb{S}^2, \quad q = 4, 5, \dots, k+3 \\
 & \boldsymbol{\theta} \in \mathbb{R}^{k+3},
 \end{aligned} \tag{14}$$

where the matrices  $\bar{\mathbf{H}}_1, \bar{\mathbf{H}}_2, \bar{\mathbf{H}}_3, \bar{\mathbf{G}}_q$  and  $\bar{\mathbf{W}}_q$ , for  $q = 4, \dots, k + 3$ , are adapted versions of their counterparts stated in Section 3.2 (see Appendix D) and the decision variables  $\mathbf{Z}_q \in \mathbb{S}^2$  arise from the conditions in Theorem 1. We point out that (14) has constraints involving  $2 \times 2$  positive semidefinite matrices, and thus it can be equivalently rewritten as a second order cone problem. However, this only happens when spline basis of degree 3 are being considered, as it is our case, but this is not true if higher degrees are used instead. Therefore, we use the approach that can cope with a general setting.

As detailed at the end of Section 4.1.1, fitting a curve  $S$  which satisfies sign, monotonicity and/or curvature requirements simultaneously is achieved by merging the models (10), (13) and/or (14) and accommodating the necessary amount and type of decision variables  $\mathbf{Z}_q$ .

#### 4.2. Multiple curves

One of the main advantages of spline modelling is that it can be embedded into an almost endless variety of statistical models, and so are the methodology proposed in this paper to constrain the estimated curves in the  $P$ -splines framework. In this section we address a very common scenario in which the functional form of the predictor variable varies across groups defined by levels of a categorical variable. Furthermore, under this scenario, our aim will be to jointly estimate non-overlapping curves using the results developed in previous sections.

Consider the set of observations  $\{(x_i, y_{i,g})\}$ , where  $x_i$  denotes the  $i$ th observation of a continuous predictor  $X$  and  $y_{i,g}$  refers to the  $i$ th observation of  $g$ th group of the response variable  $Y$ ,  $i = 1, \dots, n$  and  $g = 1, \dots, G$ . Then, the model we wish to fit is

$$y_{i,g} = f_g(x_i) + \varepsilon_{i,g}, \quad i = 1, \dots, n, \quad g = 1, \dots, G, \tag{15}$$

where  $f_1, f_2, \dots, f_G$  are  $G$  different functions depending on the group and  $\varepsilon_{i,g} \in \mathbb{R}$  are error terms.

As stated in Section 2, each function  $f_g, g = 1, \dots, G$ , is to be approximated by a linear combination of  $B$ -splines defined over a knot sequence  $\mathbf{t} = \{t_q\}_{q=1}^{k+7}$  which covers the domain of observed values of the covariate, thus yielding the curves

$$S_1(x) = \sum_{j=1}^{k+3} \theta_{j,1} B_{j,3,\mathbf{t}}(x), \quad S_2(x) = \sum_{j=1}^{k+3} \theta_{j,2} B_{j,3,\mathbf{t}}(x), \quad \dots, \quad S_G(x) = \sum_{j=1}^{k+3} \theta_{j,G} B_{j,3,\mathbf{t}}(x), \tag{16}$$

where  $\theta_{j,g}, j = 1, \dots, k + 3, g = 1, \dots, G$ , denote the coefficients of the  $B$ -spline bases expansions in each group. Observe that the set of  $B$ -splines for estimating each function is the same since the values of the predictor variable are the same in each group. However, this is not necessary, and results would be easily extended to the case in which the values of the covariate were not the same for all the groups.

The problem of simultaneously estimating the  $G$  smooth curves in (15), following the  $P$ -splines approach using a basis of  $B$ -splines, is stated as the following convex quadratic unconstrained optimization problem:

$$\min_{\substack{\theta_{j,g} \in \mathbb{R} \\ j=1, \dots, k+3 \\ g=1, \dots, G}} \sum_{g=1}^G \left\{ \sum_{i=1}^n \left( y_{i,g} - \sum_{j=1}^{k+3} \theta_{j,g} B_{j,3,\mathbf{t}}(x_i) \right)^2 + \lambda_g \sum_{j=3}^{k+3} (\theta_{j,g} - 2\theta_{j-1,g} + \theta_{j-2,g})^2 \right\}. \tag{17}$$

Observe that  $G$  smoothing parameters have to be selected, thus augmenting the computational effort required to determine them via, for instance, a cross-validation approach. Furthermore, (17) is a separable problem, namely the curve in each group could have been fitted solving  $G$  optimization problems separately. However, the goal of stating the multiple curves fitting in this way is to present its constrained version, which is proposed in this paper for the first time as far as the authors are aware. Furthermore, estimating all the curves simultaneously in the same optimization problem is crucial when the aim is to carry out inference on them, or the sample size is small.

In what follows, a constrained version of (17) is stated in which the non-overlapping of the curves is required. This would be equivalent to establish an order among the curves that has to be satisfied in the domain in which values of the covariate and the response are observed and in the out-of-range interval in which the prediction is performed. Without loss of generality, we assume that the fitted curve for observations in group  $g$  has to be above the curve for observations in group  $g + 1$ , for  $g = 1, \dots, G - 1$ . This yields a constrained version of (17) based on constraining to non-negativity the differences between adjacent curves  $h_g$  defined as:

$$h_g(x) := S_g(x) - S_{g+1}(x), \quad g = 1, 2, \dots, G - 1. \tag{18}$$

Therefore, the problem of estimating simultaneous curves in model (15) by means of non-overlapping  $P$ -splines, using  $B$ -splines as basis functions as expressed in (16), is stated as the following conic optimization problem:

$$\begin{aligned}
 & \min_{u, \tilde{\boldsymbol{\theta}}, \mathbf{Z}_{q,g}} \quad u - 2\tilde{\mathbf{y}}^\top \tilde{\mathbf{B}}\tilde{\boldsymbol{\theta}} \\
 & \text{s.t.} \quad \boldsymbol{\tau}_{q,g} = \mathbf{0}_3, \quad q = 4, 5, \dots, k + 3, \quad g = 1, 2, \dots, G - 1 \\
 & \quad \begin{pmatrix} \rho_{q,1} \\ \rho_{q,2} \\ \vdots \\ \rho_{q,G-1} \end{pmatrix} = \mathbf{F}_q \tilde{\boldsymbol{\theta}}, \quad q = 4, 5, \dots, k + 3 \\
 & \quad \left( \tilde{\mathbf{B}}^\top \tilde{\mathbf{B}} + \lambda \tilde{\mathbf{D}}^\top \tilde{\mathbf{D}} \right)^{1/2} \tilde{\boldsymbol{\theta}}, u, \frac{1}{2} \in \mathcal{Q}_r^{(k+3)G+2} \\
 & \quad \mathbf{Z}_{q,g} \in \mathbb{S}^4, \quad q = 4, 5, \dots, k + 3, \quad g = 1, 2, \dots, G - 1 \\
 & \quad \tilde{\boldsymbol{\theta}} \in \mathbb{R}^{G(k+3)},
 \end{aligned} \tag{19}$$

where

$$\tilde{\boldsymbol{\theta}} = \text{vec}(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2 \mid \dots \mid \boldsymbol{\theta}_G), \quad \tilde{\mathbf{y}} = \text{vec}(\mathbf{y}_1 \mid \mathbf{y}_2 \mid \dots \mid \mathbf{y}_G),$$

$$\tilde{\mathbf{B}} = \mathbf{I}_G \otimes \mathbf{B}, \quad \tilde{\mathbf{P}} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_G) \otimes \mathbf{D}^\top \mathbf{D},$$

$$\boldsymbol{\tau}_{q,g} = \begin{pmatrix} \langle \mathbf{H}_1, \mathbf{Z}_{q,g} \rangle_F \\ \langle \mathbf{H}_2, \mathbf{Z}_{q,g} \rangle_F \\ \langle \mathbf{H}_3, \mathbf{Z}_{q,g} \rangle_F \end{pmatrix}, \quad \boldsymbol{\rho}_{q,g} = \begin{pmatrix} \langle \mathbf{H}_4, \mathbf{Z}_{q,g} \rangle_F \\ \langle \mathbf{H}_5, \mathbf{Z}_{q,g} \rangle_F \\ \langle \mathbf{H}_6, \mathbf{Z}_{q,g} \rangle_F \\ \langle \mathbf{H}_7, \mathbf{Z}_{q,g} \rangle_F \end{pmatrix}, \quad q = 4, 5, \dots, k + 3, \quad g = 1, 2, \dots, G - 1,$$

$$\mathbf{F}_q = \begin{pmatrix} \mathbf{W}_q \mathbf{G}_q & -\mathbf{W}_q \mathbf{G}_q & \mathbf{0}_{4 \times (k+3)} & \cdots & \mathbf{0}_{4 \times (k+3)} & \mathbf{0}_{4 \times (k+3)} \\ \mathbf{0}_{4 \times (k+3)} & \mathbf{W}_q \mathbf{G}_q & -\mathbf{W}_q \mathbf{G}_q & \cdots & \mathbf{0}_{4 \times (k+3)} & \mathbf{0}_{4 \times (k+3)} \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{0}_{4 \times (k+3)} & \mathbf{0}_{4 \times (k+3)} & \mathbf{0}_{4 \times (k+3)} & \cdots & \mathbf{W}_q \mathbf{G}_q & -\mathbf{W}_q \mathbf{G}_q \end{pmatrix}, \quad q = 4, 5, \dots, k + 3,$$

and  $\boldsymbol{\theta}_g^\top = (\theta_{1,g}, \theta_{2,g}, \dots, \theta_{k+3,g})$  and  $\mathbf{y}_g^\top = (y_{1,g}, y_{2,g}, \dots, y_{n,g})$  are the coefficients of the  $B$ -splines and the  $n$  observations of the response variable  $Y$  corresponding to the  $g$ th group,  $g = 1, \dots, G$ , respectively. The operator  $\text{vec}$  denotes the vectorization operation, that converts a matrix to a vector by stacking its columns,  $\text{diag}$  refers to the operator that maps a tuple to the corresponding diagonal matrix,  $\otimes$  denotes the Kronecker product,  $\mathbf{I}_u$  is the identity matrix of order  $u$ ,  $\mathbf{0}_u$  refers to a column vector of zeroes of dimension  $u$ ,  $\mathbf{B}$  and  $\mathbf{D}$  are the basis and the difference matrices as defined in Section 3.2, respectively, and the matrices  $\mathbf{H}_\ell$ ,  $\mathbf{G}_q$  and  $\mathbf{W}_q$  are the ones defined for problem (10) and stated in Appendix B. Furthermore, the banded matrices  $\mathbf{F}_q$  contain the coefficients required to enforce the non-negativity in each interval  $[t_q, t_{q+1})$  by means of Theorem 1 to the curves  $h_g$  in (18).

The extension of (19) to the forward prediction setting is straightforward once an extended basis of  $B$ -splines is considered. Similar to the non-negativity condition imposed, which translates into the non-overlapping of the set of curves, monotonicity and curvature conditions can be addressed as in Sections 4.1.1 and 4.1.2. Finally, the optimization model in (19) easily extends to the case in which a threshold on the separation of the multiple curves has to be imposed. To do so, this threshold is added in the constant terms for the underneath curve as described in Appendix B. Moreover, this methodology enables to consider different separating thresholds at each pair of curves.

### 5. Numerical experiments

We apply our methodology to a series of data sets, two of them from a real data modelling situation, and a collection of simulated data sets used to illustrate the broad range of circumstances of applicability and the good performance against other state-of-the-art techniques. In Section 5.1, we compare our methodology with the one stated in Papp [32], Papp and Alizadeh [33] using one of their simulation scenarios. Then, Section 5.2 is devoted to a real application in the context of demography, in which the coherent out-of-range forward prediction of mortality rates for different age groups is carried

**Table 1**

Mean ( $\mu_p$ ) and standard deviation ( $\sigma_p$ ) over 100 experiments of the  $L_p$  distances,  $p \in \{1, 2, \infty\}$ , between the observed and the estimated values using the function (20) and  $\sigma = 0.15$ .

	Model	$10^2\mu_2$	$10^2\sigma_2$	$\mu_\infty$	$\sigma_\infty$	$10\mu_1$	$10\sigma_1$
Unconstrained	Problem (9)	0.800	0.096	0.256	<b>0.058</b>	0.619	0.084
Non-decreasing	PA1	1.344	0.412	0.481	0.114	0.749	0.118
	PA2	0.807	0.477	0.339	0.127	0.610	0.113
	Problem (13)	<b>0.704</b>	<b>0.092</b>	<b>0.251</b>	0.062	<b>0.522</b>	<b>0.082</b>

**Table 2**

Mean ( $\mu_p$ ) and standard deviation ( $\sigma_p$ ) over 100 experiments of the  $L_p$  distances,  $p \in \{1, 2, \infty\}$  between the observed and the estimated values using the function (20) and  $\sigma = 0.30$ .

	Model	$10^2\mu_2$	$10^2\sigma_2$	$\mu_\infty$	$\sigma_\infty$	$10\mu_1$	$10\sigma_1$
Unconstrained	Problem (9)	1.403	<b>0.178</b>	0.429	<b>0.104</b>	1.094	<b>0.157</b>
Non-decreasing	PA1	4.426	0.911	0.715	0.110	1.478	0.177
	PA2	4.052	7.799	0.581	0.307	1.290	0.679
	Problem (13)	<b>1.265</b>	0.191	<b>0.405</b>	0.108	<b>0.969</b>	0.170

out. After that, Section 5.3 shows an application arisen in the context of COVID-19 pandemic. Finally, we refer the reader to Supplementary Material S.1 for another simulation study to compare the work in Papp [32], Papp and Alizadeh [33] and ours and to Supplementary Material S.2 for an illustration about the calculation of confidence bands within our framework.

All the optimization models involved in the experiments have been solved using MOSEK [29] under Python in a PC Intel Core i7-8550, 8GB of RAM. An open source Python library, cpsplines, which contains all the implementations of the methodologies proposed in this paper is publicly available at <https://github.com/ManuelNavarroGarcia/cpsplines>. Furthermore, all the simulations and results in this section can be reproduced using the code in [https://github.com/ManuelNavarroGarcia/paper\\_cpsplines\\_1d](https://github.com/ManuelNavarroGarcia/paper_cpsplines_1d).

The smoothing parameter  $\lambda$  is chosen in all our experiments using the Generalized Cross-Validation (GCV) criterion [24], using the Sequential Least Squares Programming (SLSQP) method implemented in SciPy [36] for its optimization and using the default settings.

### 5.1. Non-decreasing smoothing P-splines

In this section, we illustrate the performance of the non-decreasing P-splines model in (13) using a simulation scenario proposed in Papp [32], Papp and Alizadeh [33]. We simulated data using the model (1), where  $f$  is the smooth non-decreasing function given by

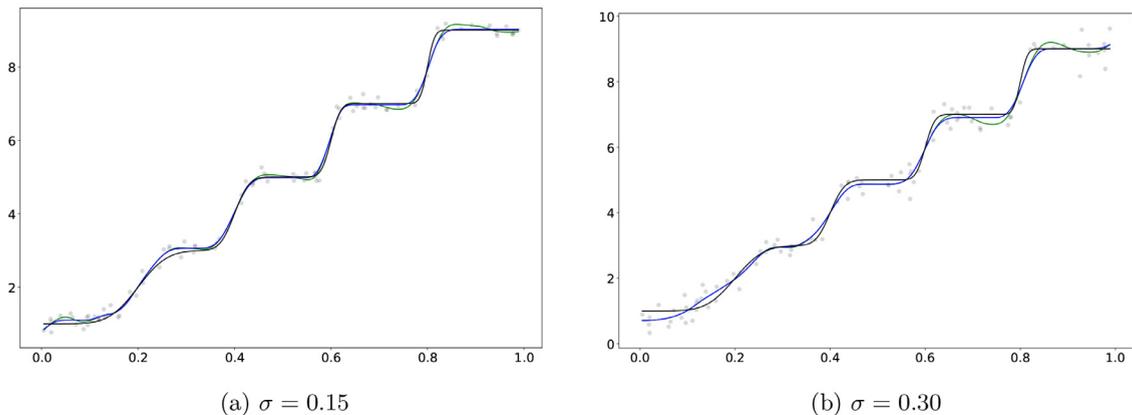
$$f(x) = 5 + \sum_{i=1}^4 \operatorname{erf} \left[ 15i \left( x - \frac{i}{5} \right) \right], \quad x \in [0, 1], \tag{20}$$

where  $\operatorname{erf}(x)$  denotes the error function

$$\operatorname{erf}(x) = \int_0^x e^{-t^2} dt,$$

and  $\varepsilon$  is normally distributed with zero mean and standard deviation  $\sigma$ . To carry out the simulation study, samples  $\{(x_i, y_i)\}_{i=1}^n$  of  $n = 100$  observations are drawn, where  $\{x_i\}$  are the values of the regressor obtained from a uniform distribution  $U(0, 1)$  and  $\{y_i\}$  are the observed values of the response variable, obtained from (1) using as  $f$  the function in (20) and  $\varepsilon$  as described above. As in Papp [32], Papp and Alizadeh [33], two different situations are considered:  $\sigma = 0.15$  and  $\sigma = 0.30$  and, for both settings, 100 samples are generated. In order to define the B-spline basis functions, the fitting region is split into  $k = 40$  equal length subintervals.

The results obtained from solving (13) are compared with the ones obtained by the two approaches stated in Papp [32], Papp and Alizadeh [33] to get a smooth monotone curve in the aforementioned simulation setting, as well as with the unconstrained counterpart in (9). Tables 1 and 2 show, for the two simulation scenarios ( $\sigma = 0.15$  and  $\sigma = 0.3$ ), respectively, the means  $\mu_p$  and standard deviations  $\sigma_p$  in 100 instances of the  $L_p$ ,  $p \in \{1, 2\}$ , divided by the number of observations, and the  $L_\infty$  distances between the predicted values for each approach and the theoretical curve, i.e.  $f$  in (20). We refer to the methodology in Papp [32], Papp and Alizadeh [33] with and without penalty as PA1 and PA2, respectively, in Tables 1 and 2. The values of  $\mu_p$  and  $\sigma_p$ ,  $p \in \{1, 2, \infty\}$  for PA1 and PA2 reported are the ones shown in Tables 3.1 and 3.2 in Papp [32]. The best result among the four approaches considered is highlighted in bold. Observe that, for all the distances, (13) outperforms the constrained approaches in PA1 and PA2 and also the unconstrained case. Thus, we can conclude that, in this case, imposing the non-decreasing requirement yields a better approximation of the theoretical curve in (20). Furthermore, both (9) and (13) show better results according to the metrics considered than PA1 and PA2. We point out that the curves obtained by the unconstrained model (9) do not satisfy, in general, the non-decreasing condition, which make them not



**Fig. 4.** Graphs of the theoretical curve (20) (black) and the curves obtained from the unconstrained (9) (green) and the non-decreasing constrained (13) (blue) approaches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

directly comparable with the ones obtained by the constrained models under study, since a lower error might be obtained but the conditions required could not be met.

To get some insights about the simulation study presented in this section, two of its instances are visualized in Fig. 4. In particular, Fig. 4a depicts a  $\sigma = 0.15$  case and Fig. 4b a  $\sigma = 0.30$  one. These graphs show the theoretical curve (20) (in black), which we aim to estimate from sample data, the unconstrained fitted curve using the model (9) (in green) and the non-decreasing fitted curve using the model (13) (in blue), together with the sample points (grey dots) used to estimate these curves. Observe that the unconstrained and the monotonic curves are very close to each other at the regions where the theoretical curve increases, but clearly differ where it flattens. In these areas, the unconstrained fitted curve displays up and downs, which goes against the non-decreasing behaviour of the theoretical function (20). This wiggly pattern disappears when the monotonicity constraints are enforced, unraveling the same trend that the theoretical curve has.

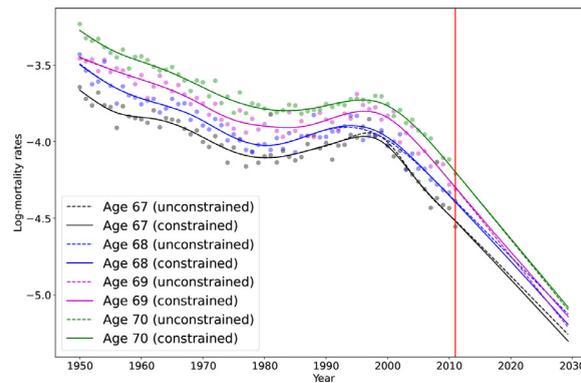
## 5.2. Coherent prediction of mortality rates

Prediction of future mortality rates is of vital importance in many areas, for example, the insurance industry needs them for annuity pricing, and governments rely on them to ensure the sustainability of public pensions systems. When a model is used to predict future mortality rates, the predicted values should be biologically reasonable in the sense that, for example, the rates change smoothly over time and increase with age. Then, if the prediction model is not subject to any constraints, or individual ages are predicted separately, these basic conditions might not be satisfied (even if they are in the domain of the observed data). To illustrate this situation we use mortality data of female Danish population. The data set is freely available from the Human Mortality Database [35]. For each year between 1950 and 2011, we have the number of deaths scaled by the size of the population (the mortality rate) for ages 67, 68, 69 and 70, and our aim is to model the logarithm of the mortality rate, as it is common in actuarial sciences, and also to carry out a forward prediction up to 2025. Thus,  $n = 62$  observations are available for each of the  $G = 4$  groups. To do so, we must take into account that a coherent forecasting in this context must ensure that mortality rates for older ages are higher than those for younger ones. Thanks to the methodology described in Section 4.2 such biologically coherent predictions are to be carried out solving the forward prediction extension to problem (19).

The domain of the data [1950, 2011] is divided into  $k = 10$  equal length subintervals. Since we are interested in making predictions until 2025 and the step length is approximately 6 years,  $k_p = 3$  extra knots are needed to reach the desired value (year 2025). Finally, a separating threshold of 0.05 between the four curves is imposed into the forward prediction extension to problem (19) to illustrate that our methodology holds also for arbitrary thresholds different from zero (non-negativity).

Figure 5 shows the fit and forward prediction of the log-mortality rates for ages 67 to 70 of Danish women, using the out-of-sample prediction extensions of problems (17) and (19). We can see that the estimated curves fit very well the pattern of mortality rates in the region where data are observed and both approaches, constrained and unconstrained, yield almost identical results. However, in the prediction region, the unconstrained curves cross-over, violating the biological premises mentioned above. This issue is solved using the methodology proposed in Section 4.2 which allows us to enforce the required conditions.

As final remarks, we point out that the solid and dashed curves for 70-year-old women coincide. The reason for this is that such age (group) is the one chosen as the first in the sequential order imposed to avoid overlapping, and thus the curve corresponding to 69-year-old women log-mortality is the one that has to change its shape to ensure that the separation between both ages is met. This reasoning extends to the other age groups. To conclude, observe that curves in the forward prediction region would also depend on the number of additional knots added to carry out the prediction.



**Fig. 5.** Curves of the log-mortality rates for four age groups of Danish women. Dashed lines correspond to the unconstrained forward prediction version of the model in (17) and solid lines to its constrained counterpart in (19). The red vertical line separates the smoothing and the prediction regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 5.3. COVID-19: Incorporating premises in the prediction of the number of infected people

The onset of the COVID-19 pandemic has favoured the appearance of a large variety of short-term prediction approaches to predict the daily number of infected people or fatalities in a certain zone and period. Being able to accurately predict the behaviour of the curve in future stages of the pandemic is crucial for the governments, whose decisions are mainly data-driven and have the aim of controlling the spread of the virus. The future shape of the curve will vary based on their decisions and, thanks to the methodology proposed in this paper, the effect that past decisions have had to *flatten the curve*, for instance to reduce the growth rate by a certain factor, could be incorporated in the forecasting model to simulate different prediction scenarios based on the observed effect different measures in the past. To illustrate this idea, we use the number of infected people per day in the Spanish region of Aragón between February 6th and August 6th 2020 (the first 183 days, i.e. the first half a year since the first detected infection). This data set, together with information related to other Spanish regions, provinces and further statistical information, is periodically published by the Spanish Epidemiological Center (Centro Nacional de Epidemiología) [14].

We consider that the domain of the data, this is, the values for the covariate “day of the year” is  $[0,182]$ , where value 0 represents February 6th 2020 and 182 corresponds to August 6th 2020. Then,  $x_i = i$  and  $y_i$  represents the number of infected people in  $i$ th day  $i = 0, \dots, 182$ , yielding the sample  $\{(x_i, y_i)\}_{i=1}^n$ , where  $n = 183$  observations. The domain of the data  $[0, 182]$  is divided into  $k = 30$  subintervals of the same length to construct the  $B$ -spline basis.

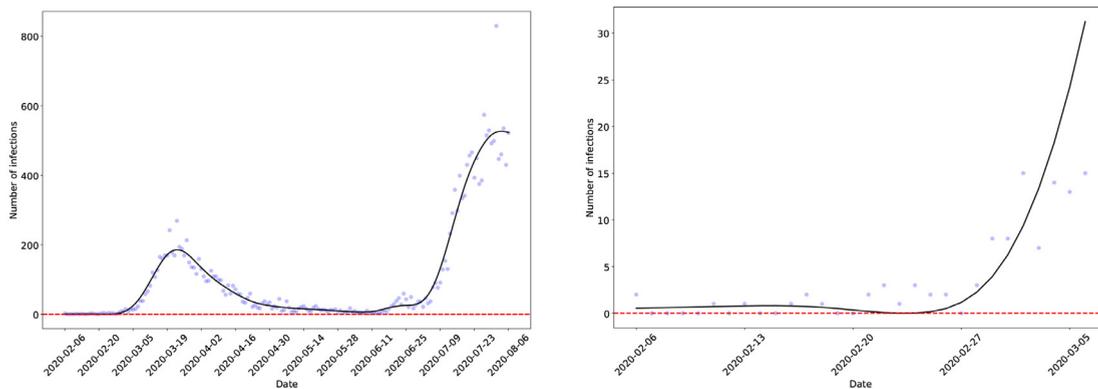
Figure 1 shows the problem that arises when the unconstrained  $P$ -splines approach, i.e. the optimization model in (9), is used to obtain a smooth curve fitting the observed data: the curve became negative from February 16th to February 27th 2020. If the non-negative constrained model is used instead, i.e. the optimization model in (10), the obtained curve is able to capture the trend without trying to interpolate, as well as being coherent with the nature of the data (see Fig. 6).

The range of dates selected corresponds to the period in which the *first wave* in Spain was over and the *second wave* was starting to diminish, as it can be observed in Fig. 6a. Then, in order to predict the evolution of the pandemic during the second wave the information enclosed in the first one can be used to simulate different scenarios. Therefore, we suppose that we face a situation in which the number of reported COVID-19 cases have been rising for some days and we aim to predict how this number evolves for some days ahead assuming different premises such as the growth rate. In the following study, we suppose we have gathered data until July 15th 2020 and we aim to predict the number of infected people 14 days ahead (i.e., the forecasting horizon is set to July 29th 2020), constraining the curve in the observed domain to be non-negative and the out-of-range prediction to satisfy certain requirements described in what follows.

From now on, we consider that the domain of the data, this is, the values for the covariate “days since first detected infection”, is  $[0,161]$ , where 0 represents February 6th 2020 and 161 corresponds to July 15th 2020, and this is split into  $k = 27$  equal length subintervals. Based on the aforementioned step length of the knot sequence, three extra knots are needed to reach the forecasting horizon, thus  $k_p = 3$  is taken.

Figure 7 shows the results of four possible out-of-range prediction scenarios obtained using the model (12) and some of its variants as presented in Section 4. The details about the conditions imposed in each case are described below:

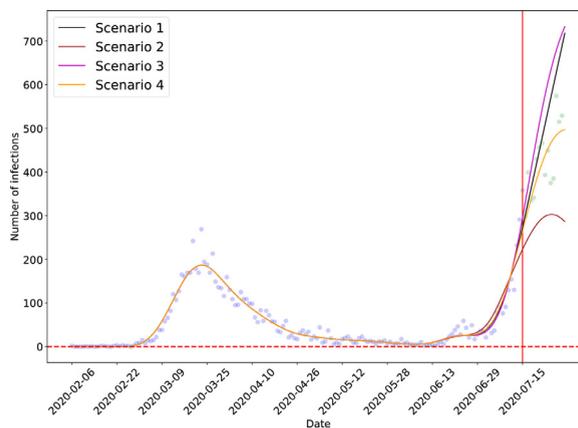
- **Scenario 1:** Non-negativity is imposed both in the observed and the forward prediction domains.
- **Scenario 2:** Besides the non-negativity requirement, equal growth rates in the first and second waves are enforced. To achieve this result, we computed the derivatives in an interval within the domain region of the curve depicted in Fig. 6a with the same length as the prediction interval, and the derivatives of the curve at the forward prediction regions are constrained to be the same.



(a) Number of daily COVID-19 cases in the Spanish region of Aragón from February 6th to August 6th 2020 (blue dots) together the non-negative fitted curve smoothing the data (black line). The red dashed line corresponds to number of cases equal to 0.

(b) Zoom in of Figure 6a from February 6th to March 5th 2020 to show that the negative number of cases predicted by the unconstrained fitted curve in Figure 1b disappears when our constrained approach is used.

**Fig. 6.** Evolution of COVID-19 pandemic in terms of the daily reported cases in the Spanish region of Aragón imposing non-negativity on the fitted curve by means of the optimization model in (10).



**Fig. 7.** Smoothing and forecasting curves obtained for the number of daily COVID-19 cases in the Spanish region of Aragón data. The solid red vertical line depicts the day in which forecasting begins and the red dashed horizontal line corresponds to the horizontal axis (number of cases equal to 0). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- **Scenarios 3 and 4:** To add more flexibility in capturing the second wave, we introduce time lags in the conditions imposed on the derivatives of the curve in the out-of-range prediction region. In Scenario 3 (resp. 4), the growth rate is imposed to be twice (resp. 1.5 times larger) the one of the first wave but with a two-day (resp. one-day) time lag.

To obtain the curves in Fig. 7, only the observations in the fitting region (blue dots) are used in the estimation procedures, and the observations in the forward prediction region (green dots) are just plotted to compare the estimated curves with the reported number of cases those days. We observe that the black curve (Scenario 1) shows an increasing trend (as expected), but the growth rate remains constant since there are not other active constraints in the prediction domain. In other words, this curve will never reach a stationary point in which it starts decreasing and thus, it depicts an unrealistic situation. On the contrary, the brown curve (Scenario 2) is able to predict where the peak would happen. However, the assumption about the same duration of both waves may be considered as a very strong one. To overcome this eventuality, we illustrate the Scenario 3 (magenta line), which is also able to capture the peak and allows the analyst to incorporate a lag parameter to extend or reduce the duration of the wave. Finally, for illustrative purposes, we show a fourth scenario (orange line) in which the growth rates in the prediction region is 1.5 times larger than the one in the observed region and 1 additional day (lag) is added for each day that goes by.

The results shown in Fig. 7 highlight that the methodology jointly with its variants fed by these two parameters (the one scaling the derivative and the one lagging the curve) are a simple but very flexible tool to simulate different plausible scenarios in an epidemiological context.

### 6. Conclusions

In this paper, we addressed the problem of estimating smooth curves in the univariate regression framework which satisfy requirements about their sign, monotonicity and curvature, both in the observed domain in which data have been gathered and outwards. These curves have been fitted using a cubic  $P$ -splines approach under constraints which ensure the fulfillment of the required conditions. To do so, a theoretical result in Bertsimas and Popescu [4] about non-negative polynomials has been used, yielding a conic optimization model for such aim. Furthermore, the proposed methodology is extended to the out-of-range prediction using  $P$ -splines following the approach in Carballo et al. [12] and Currie et al. [16] of treating the values to be predicted as missing values. Our approach successfully conveys out-of-range constrained forecasting using  $P$ -splines for the first time, as far as the authors are aware. As a final extension of the proposed methodology, the case of multiple curves estimation is addressed to deal with the problem of simultaneously obtain a set of curves which represent different groups and that must satisfy, for instance, non-overlapping requirements. The stated conic programming models which give solution to these different situations have been tested on simulated and real data. The first real application arises in the context of the COVID-19 pandemic, to ensure a non-negative smoothing of number of daily reported cases in the Spanish region of Aragón and also simulate different scenarios in the forecast of the evolution of the pandemic, while the second comes from a demographic framework, in which the curves representing different groups of age cannot overlap because of biological reasons.

Several extensions of the proposed methodology remain unexplored and entail a challenge for the near future. Particularly, two research lines are the natural continuation of this work. First, studying the general case in which a  $m$ -dimensional smooth hypersurface has to be estimated instead of a curve requires a deeper study to understand how shape conditions could be imposed in that case in terms of easy-to-model constraints in the fitting optimization model. Second, integrating a variable selection approach which allows us to find a good subset of covariates and interaction terms in the smoothing regression setting and within the constrained framework proposed in this paper.

### Funding

This research is supported by projects PID2019-104901RB-I00, PID2019-110886RB-I00 and IJC2020-045220-I (funded by MCIN/AEI/10.13039/501100011033 and the last also supported by European Union “NextGenerationEU/PRTR” funds), FQM-329, P18-FR-2369 and US-13811 (Junta de Andalucía), IND2020/TIC-17526 (Comunidad de Madrid) and Universidad Carlos III de Madrid (Read & Publish Agreement CRUE-CSIC 2022). This support is gratefully acknowledged.

### Data availability

We have shared the links to our data and code within the manuscript.

### Acknowledgments

The authors would like to thank the reviewers for their comments on prior versions of this manuscript.

### Appendix A

The optimization model in (9) is a convex quadratic problem. Quadratic programming is a special case of second-order cone programming, and therefore (9) can be reformulated as that. To do so, let us consider a non-negative auxiliary variable  $u$  to make the objective function linear:

$$\begin{aligned}
 & \min_{u, \theta} \quad u - 2\mathbf{y}^T \mathbf{B}\theta \\
 & \text{s.t.} \quad \theta^T (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{D}) \theta \leq u \\
 & \quad \quad \theta \in \mathbb{R}^{k+3}, \quad u \geq 0.
 \end{aligned} \tag{A.1}$$

Then, the new constraint is expressed using a rotated second-order cone, yielding the following reformulation for (9):

$$\begin{aligned}
 & \min_{u, \theta} \quad u - 2\mathbf{y}^T \mathbf{B}\theta \\
 & \text{s.t.} \quad ((\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{D})^{1/2} \theta, u, \frac{1}{2}) \in \mathcal{Q}_r^{k+5}
 \end{aligned} \tag{A.2}$$

where  $\mathcal{Q}_r^{n+2} = \{(\mathbf{x}, y, z) \in \mathbb{R}^{n+2} : \|\mathbf{x}\|_2^2 \leq 2yz, y \geq 0, z \geq 0\}$ .

**Appendix B**

In what follows, the expressions of matrices  $\mathbf{H}_\ell$ ,  $\ell = 1, \dots, 7$  in (10) and (12),  $\mathbf{W}_q$  and  $\mathbf{G}_q$ ,  $q = 4, 5, \dots, k + 3$ , in (10) and  $\mathbf{W}_q$  and  $\mathbf{G}_{+,q}$ ,  $q = 4, 5, \dots, k + k_p + 3$ , in (12) are provided.

On one hand, matrices  $\mathbf{H}_1, \dots, \mathbf{H}_7$  allow us to define the left-hand sides of the conditions in Theorem 1:

$$\mathbf{H}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\mathbf{H}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_5 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_6 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_7 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

On the other hand,  $\mathbf{G}_q$  and  $\mathbf{W}_q$  allow to define the right-hand sides of such conditions,  $q = 4, \dots, k + 3$ . Let  $\mathbf{G}_q$  be the  $4 \times (k + 3)$  block matrix containing the coefficients of cubic, quadratic, linear and constant terms of the  $B$ -spline basis in the interval  $[t_q, t_{q+1})$  which is defined as

$$\mathbf{G}_q = \left( \begin{array}{c|cccc} & \mathcal{G}_{q-3,q,0} & \mathcal{G}_{q-2,q,0} & \mathcal{G}_{q-1,q,0} & \mathcal{G}_{q,q,0} \\ \mathbf{0}_{4 \times (q-4)} & \mathcal{G}_{q-3,q,1} & \mathcal{G}_{q-2,q,1} & \mathcal{G}_{q-1,q,1} & \mathcal{G}_{q,q,1} \\ & \mathcal{G}_{q-3,q,2} & \mathcal{G}_{q-2,q,2} & \mathcal{G}_{q-1,q,2} & \mathcal{G}_{q,q,2} \\ & \mathcal{G}_{q-3,q,3} & \mathcal{G}_{q-2,q,3} & \mathcal{G}_{q-1,q,3} & \mathcal{G}_{q,q,3} \end{array} \right) \mathbf{0}_{4 \times (k+3-q)},$$

where  $\mathbf{0}_{u \times v}$  represents a null matrix with  $u$  rows and  $v$  columns. Furthermore, let  $\mathbf{W}_q$  be a  $4 \times 4$  matrix which allows us to retain some of the right-hand-sides in the conditions in Theorem 1, which is defined as

$$\mathbf{W}_q = \begin{pmatrix} 1 & t_q & t_q^2 & t_q^3 \\ [6pt]3 & 2t_q + t_{q+1} & t_q^2 + 2t_q t_{q+1} & 3t_q^2 t_{q+1} \\ [6pt]3 & 2t_{q+1} + t_q & t_{q+1}^2 + 2t_q t_{q+1} & 3t_q t_{q+1}^2 \\ [6pt]1 & t_{q+1} & t_{q+1}^2 & t_{q+1}^3 \end{pmatrix}. \tag{B.1}$$

For the forecasting approach in (12), extended versions of the previous matrices are needed. Then, for  $q = 4, 5, \dots, k + k_p + 3$  one has

$$\mathbf{G}_{+,q} = \left( \begin{array}{c|cccc} & \mathcal{G}_{q-3,q,0} & \mathcal{G}_{q-2,q,0} & \mathcal{G}_{q-1,q,0} & \mathcal{G}_{q,q,0} \\ \mathbf{0}_{4 \times (q-4)} & \mathcal{G}_{q-3,q,1} & \mathcal{G}_{q-2,q,1} & \mathcal{G}_{q-1,q,1} & \mathcal{G}_{q,q,1} \\ & \mathcal{G}_{q-3,q,2} & \mathcal{G}_{q-2,q,2} & \mathcal{G}_{q-1,q,2} & \mathcal{G}_{q,q,2} \\ & \mathcal{G}_{q-3,q,3} & \mathcal{G}_{q-2,q,3} & \mathcal{G}_{q-1,q,3} & \mathcal{G}_{q,q,3} \end{array} \right) \mathbf{0}_{4 \times (k+k_p+3-q)},$$

and  $\mathbf{W}_q$  are the ones defined in (B.1) but for a larger sequence  $q = 4, 5, \dots, k + k_p + 3$ .

For the sake of completeness, we describe in what follows how model (10) changes when the unknown function  $f$  to be estimated in (1) is required to be greater or equal than a given positive threshold  $\alpha$ . Observe that this condition is equivalent to impose the non-negativity requirement to the curve  $f_\alpha(x) := f(x) - \alpha$ , so we need to determine the form of the coefficients of  $f_\alpha$  in order to apply Theorem 1. Since the coefficients of  $f$  at the interval  $[t_q, t_{q+1})$  are given by  $\mathbf{G}_q \boldsymbol{\theta}$ , the corresponding coefficients for  $f_\alpha$  are given by  $\mathbf{G}_q \boldsymbol{\theta} - (\alpha, 0, 0, 0)^\top$ . Therefore, the conic optimization model yielding a curve

which is greater or equal than  $\alpha > 0$  is given by:

$$\begin{aligned}
 & \min_{u, \boldsymbol{\theta}, \mathbf{Z}_q} \quad u - 2\mathbf{y}^\top \mathbf{B}\boldsymbol{\theta} \\
 & \text{s.t.} \quad \langle \mathbf{H}_\ell, \mathbf{Z}_q \rangle_F = 0, \quad q = 4, 5, \dots, k+3, \quad \ell = 1, 2, 3 \\
 & \quad \begin{pmatrix} \langle \mathbf{H}_4, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_5, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_6, \mathbf{Z}_q \rangle_F \\ \langle \mathbf{H}_7, \mathbf{Z}_q \rangle_F \end{pmatrix} = \mathbf{W}_q \mathbf{G}_q \boldsymbol{\theta} - \mathbf{W}_q \begin{pmatrix} \alpha \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad q = 4, 5, \dots, k+3 \\
 & \quad \left( (\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D})^{1/2} \boldsymbol{\theta}, u, \frac{1}{2} \right) \in \mathcal{Q}_r^{k+5} \\
 & \quad \mathbf{Z}_q \in \mathbb{S}^4, \quad q = 4, 5, \dots, k+3 \\
 & \quad \boldsymbol{\theta} \in \mathbb{R}^{k+3},
 \end{aligned} \tag{B.2}$$

**Appendix C**

This appendix is devoted to present the matrices  $\widehat{\mathbf{H}}_\ell$ ,  $\widehat{\mathbf{W}}_q$  and  $\widehat{\mathbf{G}}_q$ , for  $\ell = 1, \dots, 5$ ,  $q = 4, 5, \dots, k+3$ , in (13). First,  $\widehat{\mathbf{H}}_1, \dots, \widehat{\mathbf{H}}_5$  are auxiliary matrices used to state the left-hand sides of the conditions in Theorem 1:

$$\begin{aligned}
 \widehat{\mathbf{H}}_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \widehat{\mathbf{H}}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \widehat{\mathbf{H}}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
 \widehat{\mathbf{H}}_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \widehat{\mathbf{H}}_5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
 \end{aligned}$$

Then,  $\widehat{\mathbf{G}}_q$  and  $\widehat{\mathbf{W}}_q$  are used to formulate the right-hand sides of these conditions,  $q = 4, \dots, k+3$ . In this paradigm,  $\widehat{\mathbf{G}}_q$  are  $3 \times (k+3)$  matrices obtained by deleting the first row of matrices  $\mathbf{G}_q$  as defined in Appendix B:

$$\widehat{\mathbf{G}}_q = \left( \begin{array}{c|cccc} \mathbf{0}_{3 \times (q-4)} & \mathcal{G}_{q-3,q,1} & \mathcal{G}_{q-2,q,1} & \mathcal{G}_{q-1,q,1} & \mathcal{G}_{q,q,1} \\ & \mathcal{G}_{q-3,q,2} & \mathcal{G}_{q-2,q,2} & \mathcal{G}_{q-1,q,2} & \mathcal{G}_{q,q,2} \\ & \mathcal{G}_{q-3,q,3} & \mathcal{G}_{q-2,q,3} & \mathcal{G}_{q-1,q,3} & \mathcal{G}_{q,q,3} \\ \hline & \mathbf{0}_{3 \times (k+3-q)} & & & \end{array} \right).$$

Furthermore, let  $\widehat{\mathbf{W}}_q$  be a  $3 \times 3$  matrix which contain the contribution of the interval bounds  $t_q$  and  $t_{q+1}$  on the right-hand-side conditions in Theorem 1 defined as

$$\widehat{\mathbf{W}}_q = \begin{pmatrix} 1 & 2t_q & 3t_q^2 \\ 2 & 2t_q + 2t_{q+1} & 6t_q t_{q+1} \\ 1 & 2t_{q+1} & 3t_{q+1}^2 \end{pmatrix}.$$

**Appendix D**

In this appendix, we present the expressions of matrices  $\overline{\mathbf{H}}_\ell$ ,  $\overline{\mathbf{W}}_q$  and  $\overline{\mathbf{G}}_q$ , for  $\ell = 1, 2, 3$ ,  $q = 4, 5, \dots, k+3$ , in (14).

First,  $\overline{\mathbf{H}}_1, \overline{\mathbf{H}}_2, \overline{\mathbf{H}}_3$  are used to model the left-hand sides of the conditions in Theorem 1 applied to degree one polynomials:

$$\overline{\mathbf{H}}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \overline{\mathbf{H}}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \overline{\mathbf{H}}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Second,  $\overline{\mathbf{G}}_q$  contain the coefficients of cubic and quadratic terms of the B-spline, i.e., they are obtained by retaining only the last two rows of matrices  $\mathbf{G}_q$ ,  $q = 4, \dots, k+3$ :

$$\overline{\mathbf{G}}_q = \left( \begin{array}{c|cccc} \mathbf{0}_{2 \times (q-4)} & \mathcal{G}_{q-3,q,2} & \mathcal{G}_{q-2,q,2} & \mathcal{G}_{q-1,q,2} & \mathcal{G}_{q,q,2} \\ & \mathcal{G}_{q-3,q,3} & \mathcal{G}_{q-2,q,3} & \mathcal{G}_{q-1,q,3} & \mathcal{G}_{q,q,3} \\ \hline & \mathbf{0}_{2 \times (k+3-q)} & & & \end{array} \right),$$

Finally, let  $\bar{W}_q$  be a  $2 \times 2$  matrix which allows us to retain some of the right-hand-side conditions in [Theorem 1](#) defined as

$$\bar{W}_q = \begin{pmatrix} 2 & 6t_q \\ 2 & 6t_{q+1} \end{pmatrix}.$$

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.amc.2022.127679](https://doi.org/10.1016/j.amc.2022.127679).

## References

- [1] S. Benítez-Peña, E. Carrizosa, V. Guerrero, M.D. Jiménez-Gamero, B. Martín-Barragán, C. Molero-Río, P. Ramírez-Cobo, D. Romero Morales, M.R. Sillero-Denamiel, On sparse ensemble methods: an application to short-term predictions of the evolution of COVID-19, *Eur. J. Oper. Res.* 295 (2021) 648–663, doi:[10.1016/j.ejor.2021.04.016](https://doi.org/10.1016/j.ejor.2021.04.016).
- [2] D. Bertsimas, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, A. Jacquillat, D. Lahlou Kitane, G. Lukin, M. Li, M. Luca, O. Nohadani, A. Orfanoudaki, T. Papalexopoulos, I. Paskov, J. Pauphilet, O.L. Skali, B. Stellato, B.H. Tazi, K. Villalobos Carballo, H. Wiberg, C. Zeng, From predictions to prescriptions: a data-driven response to COVID-19, *Health Care Manage. Sci.* 24 (2021) 253–272, doi:[10.1007/s10729-020-09542-0](https://doi.org/10.1007/s10729-020-09542-0).
- [3] D. Bertsimas, A. King, OR forum—an algorithmic approach to linear regression, *Oper. Res.* 64 (1) (2016) 2–16, doi:[10.1287/opre.2015.1436](https://doi.org/10.1287/opre.2015.1436).
- [4] D. Bertsimas, I. Popescu, On the relation between option and stock prices: a convex optimization approach, *Oper. Res.* 50 (2) (2002) 358–374, doi:[10.1287/opre.50.2.358.424](https://doi.org/10.1287/opre.50.2.358.424).
- [5] R. Blanquero, E. Carrizosa, C. Molero-Río, D. Romero Morales, Sparsity in optimal randomized classification trees, *Eur. J. Oper. Res.* 284 (1) (2020) 255–272, doi:[10.1016/j.ejor.2019.12.002](https://doi.org/10.1016/j.ejor.2019.12.002).
- [6] K. Bollaerts, P.H. Eilers, I. Van Mechelen, Simple and multiple P-splines regression with shape constraints, *Br. J. Math. Stat. Psychol.* 59 (2) (2006) 451–469, doi:[10.1348/000711005X84293](https://doi.org/10.1348/000711005X84293).
- [7] H. Booth, L. Tickle, Mortality modelling and forecasting: a review of methods, *Ann. Actuarial Sci.* 3 (1–2) (2008) 3–43, doi:[10.1017/s1748499500000440](https://doi.org/10.1017/s1748499500000440).
- [8] S. Boyd, S.P. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [9] V. Bucarey, M. Labbé, J.M. Morales, S. Pineda, An exact dynamic programming approach to segmented isotonic regression, *Omega* 105 (2021) 102–516, doi:[10.1016/j.omega.2021.102516](https://doi.org/10.1016/j.omega.2021.102516).
- [10] A.J.G. Cairns, D. Blake, K. Dowd, Pricing death: frameworks for the valuation and securitization of mortality risk, *ASTIN Bull.* 36 (1) (2006) 79–120, doi:[10.2143/ast.36.1.2014145](https://doi.org/10.2143/ast.36.1.2014145).
- [11] C.G. Camarda, Smooth constrained mortality forecasting, *Demogr. Res.* 41 (38) (2019) 1091–1130, doi:[10.4054/DemRes.2019.41.38](https://doi.org/10.4054/DemRes.2019.41.38).
- [12] A. Carballo, M. Durban, G. Kauermann, D.J. Lee, A general framework for prediction in penalized regression, *Stat. Modell.* 21 (4) (2021) 293–312, doi:[10.1177/1471082X19896867](https://doi.org/10.1177/1471082X19896867).
- [13] E. Carrizosa, V. Guerrero, rs-Sparse principal component analysis: a mixed integer nonlinear programming approach with VNS, *Comput. Oper. Res.* 52 (2014) 349–354, doi:[10.1016/j.cor.2013.04.012](https://doi.org/10.1016/j.cor.2013.04.012).
- [14] Centro Nacional de Epidemiología, Instituto de Salud Carlos III (Sapin), Accessed: September 2020, <https://cneccovid.isciii.es/covid19/>.
- [15] I.D. Currie, Smoothing constrained generalized linear models with an application to the Lee-Carter model, *Stat. Modell.* 13 (1) (2013) 69–93, doi:[10.1177/1471082X12471373](https://doi.org/10.1177/1471082X12471373).
- [16] I.D. Currie, M. Durban, P.H.C. Eilers, Smoothing and forecasting mortality rates, *Stat. Modell.* 4 (2004) 279–298.
- [17] A.C. Davison, D.V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, 1990.
- [18] C. De Boor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
- [19] A. Delwarde, M. Denuit, P.H.C. Eilers, Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach, *Stat. Modell.* 7 (1) (2007) 29–48, doi:[10.1177/1471082X0600700103](https://doi.org/10.1177/1471082X0600700103).
- [20] P.H.C. Eilers, B.D. Marx, Flexible smoothing with B-splines and penalties, *Stat. Sci.* 11 (2) (1996) 89–121.
- [21] P.H.C. Eilers, B.D. Marx, Splines, knots, and penalties, *WIREs Comput. Stat.* 2 (6) (2010) 637–653, doi:[10.1002/wics.125](https://doi.org/10.1002/wics.125).
- [22] C. Gambella, B. Ghaddar, J. Naoum-Sawaya, Optimization problems for machine learning: a survey, *Eur. J. Oper. Res.* 290 (3) (2021) 807–828, doi:[10.1016/j.ejor.2020.08.045](https://doi.org/10.1016/j.ejor.2020.08.045).
- [23] M. Goic, M.S. Bozanic-Leal, M. Badal, L.J. Basso, COVID-19: short-term forecast of ICU beds in times of crisis, *PLoS ONE* 16 (1) (2021), doi:[10.1371/journal.pone.0245272](https://doi.org/10.1371/journal.pone.0245272).
- [24] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics* 21 (2) (1979) 215–223.
- [25] T.J. Hastie, R.J. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 1990.
- [26] X. Ju, J.M. Rosenberger, V.C.P. Chen, F. Liu, Global optimization on non-convex two-way interaction truncated linear multivariate adaptive regression splines using mixed integer quadratic programming, *Inf. Sci.* 597 (2022) 38–52, doi:[10.1016/j.ins.2022.03.041](https://doi.org/10.1016/j.ins.2022.03.041).
- [27] M.C. Meyer, Constrained penalized splines, *Can. J. Stat.* 40 (1) (2012) 190–206, doi:[10.1002/cjs](https://doi.org/10.1002/cjs).
- [28] A.M. Monteiro, R.H. Tütüncü, L.N. Vicente, Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity, *Eur. J. Oper. Res.* 187 (2) (2008) 525–542, doi:[10.1016/j.ejor.2007.02.041](https://doi.org/10.1016/j.ejor.2007.02.041).
- [29] I. MOSEK ApS, MOSEK optimizer API for Python 9.2.14, 2020, <https://docs.mosek.com/9.2/pythonapi/index.html>.
- [30] F. O’Sullivan, A Statistical perspective on ill-posed inverse problems, *Stat. Sci.* 1 (4) (1986) 502–527.
- [31] F. O’Sullivan, Fast computation of fully automated logdensity and log-hazard estimators, *SIAM J. Sci. Stat. Comput.* 9 (2) (1988) 363–379. <https://epubs.siam.org/page/terms>
- [32] D. Papp, Optimization models for shape-constrained function estimation problems involving nonnegative polynomials and their restrictions, *Rutgers The State University of New Jersey-New Brunswick*, 2011 Ph.D. thesis.
- [33] D. Papp, F. Alizadeh, Shape-constrained estimation using nonnegative splines, *J. Comput. Graph. Stat.* 23 (1) (2014) 211–231, doi:[10.1080/10618600.2012.707343](https://doi.org/10.1080/10618600.2012.707343).
- [34] N. Pya, S.N. Wood, Shape constrained additive models, *Stat. Comput.* 25 (3) (2015) 543–559, doi:[10.1007/s11222-013-9448-7](https://doi.org/10.1007/s11222-013-9448-7).
- [35] University of California (USA), & Max Planck Institute for Demographic Research (Germany) (2020). Human Mortality Database. URL: [www.mortality.org](http://www.mortality.org).
- [36] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, S. Contributors, {SciPy} 1.0: fundamental algorithms for scientific computing in python, *Nat. Methods* 17 (2020) 261–272, doi:[10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [37] Y. Xia, F. Alizadeh, Second-Order Cone Programming for P-Spline Simulation Metamodeling, Technical Report, 2015. <https://arxiv.org/abs/1506.05536>