Aust. N. Z. J. Stat. **50**(2), 2008, 179–198 doi: 10.1111/j.1467-842X.2008.00507.x

ON SEMIPARAMETRIC REGRESSION WITH O'SULLIVAN PENALIZED SPLINES

M. P. WAND[1]* AND J. T. ORMEROD[2]

University of Wollongong and The University of New South Wales

Summary

An exposition on the use of O'Sullivan penalized splines in contemporary semiparametric regression, including mixed model and Bayesian formulations, is presented. O'Sullivan penalized splines are similar to P-splines, but have the advantage of being a direct generalization of smoothing splines. Exact expressions for the O'Sullivan penalty matrix are obtained. Comparisons between the two types of splines reveal that O'Sullivan penalized splines more closely mimic the natural boundary behaviour of smoothing splines. Implementation in modern computing environments such as MATLAB, R and BUGS is discussed.

Key words: additive models; Markov chain Monte Carlo; mixed models; P-splines; smoothing splines.

1. Introduction

Splines continue to play a central role in non-parametric and semiparametric regression modelling. Recent synopses include Eubank (1999), Gu (2002), Ruppert, Wand & Carroll (2003) and Denison *et al.* (2002). In all but the last reference, smooth functional relationships are fitted using a large basis of spline functions subject to penalization. Up until the mid-1990s, most literature on spline-based non-parametric regression was concerned with *smoothing splines*, and their multivariate extension, *thin-plate splines*, in which the penalty takes a particular form, and the number of basis functions roughly equals the sample size (e.g. Wahba, 1990; Green & Silverman, 1994). In recent years, however, there has been a great deal of research into more general spline/penalty strategies, most of which use considerably fewer basis functions. Driving forces include:

(i) more complicated models, often with several smooth functions;
(ii) larger data sets, in which smoothing and thin-plate splines become computationally intractable;
(iii) mixed model and Bayesian representations of smoothers that lend themselves to the use of established software, such as BUGS (Spiegelhalter, Thomas & Best, 2000), lme() in R (R Development Core Team, 2007), and PROC MIXED in SAS (SAS Institute, Inc., 2007), provided that the number of basis functions is relatively low.

*Author to whom correspondence should be addressed.
[1]School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia. e-mail: mwand@uow.edu.au
[2]School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW 2052, Australia.
Acknowledgments: The authors are grateful for comments from an associate editor and a referee, which led to considerable improvements in this article. Feedback from Iain Currie, Paul Eilers and Brian Marx is also gratefully acknowledged. The first author would like to acknowledge the hospitality of the Indian Statistical Institute, Calcutta, where some of this research took place.

Ruppert *et al.* (2003) summarize and provide access to many of these developments. The term *penalized splines* has emerged as a descriptor for general spline fitting subject to penalties.

O'Sullivan (1986, Section 3) introduced a class of penalized splines based on B-spline basis functions. O'Sullivan penalized splines are a direct generalization of smoothing splines, in that the latter arise when the maximal number of B-spline basis functions is included. Like smoothing splines, O'Sullivan penalized splines possess the attractive feature of natural boundary conditions (e.g. Green & Silverman, 1994, p. 12). They have also become the most widely used class of penalized splines in statistical analyses, as a result of their implementation in the popular R and S-PLUS (Insightful Corporation, 2007) function smooth.spline() and associated generalized additive model software (e.g. the gam library in R; Hastie, 2006).

Despite the omnipresence of O'Sullivan penalized splines, their use in semiparametric regression contexts, particularly those involving mixed model and Bayesian representations, is not very common. Recently, Welham *et al.* (2007) showed how most of the commonly used penalized splines can be treated within a single mixed model framework, although they did not work explicitly with the form given in O'Sullivan (1986).

In this paper we: (i) provide an exact matrix expression for the penalty of O'Sullivan splines that can be implemented in a few lines of a matrix-based computing language; (ii) compare O'Sullivan splines with their closest penalized spline relative, P-splines (Eilers & Marx, 1996), revealing some noticeable differences near the boundaries; (iii) demonstrate explicitly, including with R code, how O'Sullivan splines can be simply added to the mixed model-based regression armoury; and (iv) investigate their efficacy in Bayesian semiparametric regression using MCMC software such as BUGS and its variants. We conclude that the attractive features of O'Sullivan penalized splines – smoothness, numerical stability, natural boundary properties, direct generalization of smoothing splines – make them a very good choice of basis in semiparametric regression.

Section 2 provides a brief description of O'Sullivan penalized splines. A comparison with P-splines is made in Section 3. Section 4 describes a mixed model representation of O'Sullivan penalized splines, and how they can be used in models that benefit from this representation. Issues concerning Bayesian penalized spline smoothing and Markov chain Monte Carlo are described in Section 5. O'Sullivan splines of general degree are described in Section 6. A closing discussion is given in Section 7. An appendix contains relevant R code.

2. O'Sullivan penalized splines

O'Sullivan penalized splines have already been described several times in the literature. A recent reference is the Chapter 5 Appendix of Hastie, Tibshirani & Friedman (2001). A brief sketch is given here for convenience.

Consider the simplest non-parametric regression setting

$$y_i = f(x_i) + \varepsilon_i, \quad 1 \le i \le n, \tag{1}$$

where $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}$. Suppose that an estimate of f is required over $[a, b]$, an interval containing the x_is. For an integer $K \le n$, let $\kappa_1, \ldots, \kappa_{K+8}$ be a knot sequence such that

$$a = \kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 < \kappa_5 < \cdots < \kappa_{K+4} < \kappa_{K+5} = \kappa_{K+6} = \kappa_{K+7} = \kappa_{K+8} = b,$$

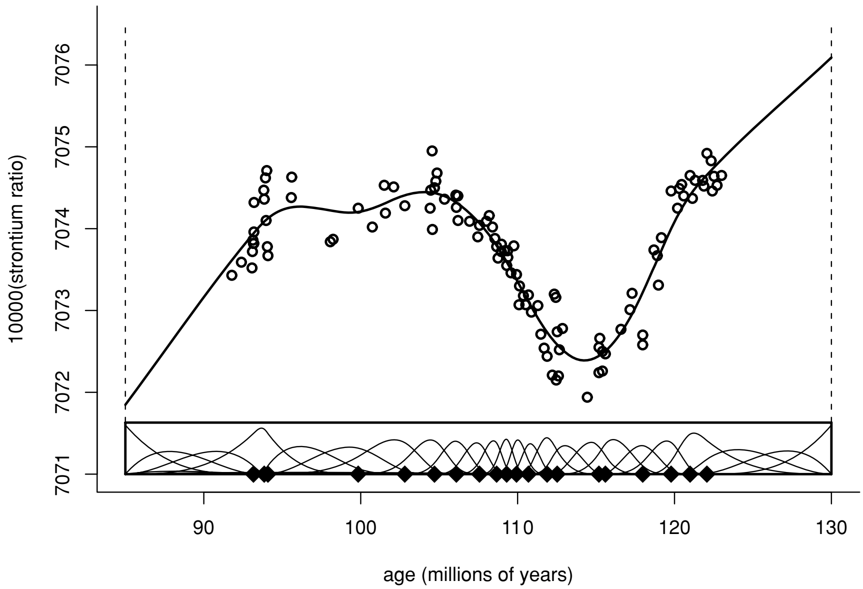

Figure 1. Illustration of natural boundary properties of a 20-interior knot O'Sullivan penalized spline fit to fossil data (source: Chaudhuri & Marron, 1999) over the interval [85, 130] millions of years. The interior knots are shown as solid diamonds. Inset: the 24 B-spline basis functions.

and let $B_1, \ldots, B_{K+4}$ be the cubic B-spline basis functions defined by these knots (see, for example, pp. 160–161 of Hastie *et al.*, 2001). Set up the $n \times (K+4)$ design matrix $\boldsymbol{B}$ with (i,k)th entry $B_{ik} = B_k(x_i)$, and the $(K+4) \times (K+4)$ penalty matrix $\boldsymbol{\Omega}$ with (k,k')th entry

$$\boldsymbol{\Omega}_{kk'} = \int_a^b B_k''(x) B_{k'}''(x)\, dx.$$

Then an estimate of f at location $x \in \mathbb{R}$ can be obtained as

$$\widehat{f}_O(x;\lambda) \equiv \boldsymbol{B}_x \widehat{\boldsymbol{\nu}}_O, \quad \text{where} \quad \widehat{\boldsymbol{\nu}}_O \equiv (\boldsymbol{B}^\top \boldsymbol{B} + \lambda \boldsymbol{\Omega})^{-1} \boldsymbol{B}^\top \boldsymbol{y}, \tag{2}$$

$\boldsymbol{B}_x \equiv [B_1(x), \ldots, B_{K+4}(x)]$, and $\lambda > 0$ is a smoothing parameter.

Note that the cubic smoothing spline arises in the special case $K = n$ and $\kappa_{k+4} = x_k$, $1 \leq k \leq n$, provided that the x_is are distinct (e.g. Green & Silverman, 1994, Section 3.6). Apart from giving a smooth (twice continuously differentiable) scatterplot smooth, $\widehat{f}_O(\cdot;\lambda)$ has good numerical properties. The basis functions are bounded and so not prone to overflow problems. Moreover, $\boldsymbol{B}^\top \boldsymbol{B}$ is four-banded, which leads to $O(n)$ algorithms when K is close to n (e.g. Hastie *et al.*, 2001). In addition, $\widehat{f}_O(\cdot;\lambda)$ satisfies so-called natural boundary conditions, meaning that

$$\widehat{f}_O''(a;\lambda) = \widehat{f}_O'''(a;\lambda) = \widehat{f}_O''(b;\lambda) = \widehat{f}_O'''(b;\lambda) = 0,$$

and implying that $\widehat{f}_O(\cdot;\lambda)$ is approximately linear over $[a, \kappa_5]$ and $[\kappa_{K+4}, b]$ (linearity is exact if $\kappa_5 = \min(x_i)$ and $\kappa_{K+4} = \max(x_i)$). Figure 1 illustrates these natural boundary properties of $\widehat{f}_O(\cdot;\lambda)$ for data on ratios of strontium isotopes found in fossil shells and their age; see Chaudhuri & Marron (1999) for details. Furthermore, $\widehat{f}_O(\cdot;\lambda)$ approximates the least squares line as $\lambda \to \infty$. The implication for mixed model smoothing is that the induced fixed effects component corresponds to straight line basis functions. Details are given in Section 4.

Computation of the design matrix $\boldsymbol{B}$ is usually quite easy. For example, B-splines are readily available in the MATLAB (The Mathworks, Inc., 2007), R and S-PLUS computing environments. Otherwise, recurrence formulae (e.g. de Boor, 1978; Eilers & Marx, 1996) can be called upon. However, computation of $\boldsymbol{\Omega}$ requires some additional effort. In Section 6, while treating general degree O'Sullivan penalized splines, we derive an exact matrix algebraic expression for the corresponding penalty matrices. In the cubic case, our theorem reduces to the expression

$$\boldsymbol{\Omega} = (\widetilde{\boldsymbol{B}}'')^\top \text{diag}(\boldsymbol{w})\widetilde{\boldsymbol{B}}'', \tag{3}$$

where $\widetilde{\boldsymbol{B}}''$ is the $3(K+7) \times (K+4)$ matrix with (i, j)th entry $B_j''(\widetilde{x}_i)$, $\widetilde{x}_i$ is the ith entry of the vector

$$\widetilde{\boldsymbol{x}} = \left(\kappa_1, \frac{\kappa_1 + \kappa_2}{2}, \kappa_2, \kappa_2, \frac{\kappa_2 + \kappa_3}{2}, \kappa_3, \ldots, \kappa_{K+7}, \frac{\kappa_{K+7} + \kappa_{K+8}}{2}, \kappa_{K+8}\right),$$

and $\boldsymbol{w}$ is the $3(K+7) \times 1$ vector given by

$$\boldsymbol{w} = \left(\frac{1}{6}(\Delta\boldsymbol{\kappa})_1, \frac{4}{6}(\Delta\boldsymbol{\kappa})_1, \frac{1}{6}(\Delta\boldsymbol{\kappa})_1, \frac{1}{6}(\Delta\boldsymbol{\kappa})_2, \frac{4}{6}(\Delta\boldsymbol{\kappa})_2, \right.$$
$$\left. \frac{1}{6}(\Delta\boldsymbol{\kappa})_2, \ldots, \frac{1}{6}(\Delta\boldsymbol{\kappa})_{K+7}, \frac{4}{6}(\Delta\boldsymbol{\kappa})_{K+7}, \frac{1}{6}(\Delta\boldsymbol{\kappa})_{K+7}\right),$$

where $(\Delta\boldsymbol{\kappa})_k \equiv \kappa_{k+1} - \kappa_k$, $1 \le k \le K+7$. Result (3) is none other than Simpson's rule applied over each of the inter-knot differences. This is because each $B_i''B_j''$ function is piecewise-quadratic. For commonly used values of K, (3) allows straightforward computation of $\boldsymbol{\Omega}$ in matrix-based languages such as MATLAB, R and S-PLUS. In the Appendix, we demonstrate the computation of $\boldsymbol{\Omega}$ in four lines of R code.

Finally, we mention knot choice. The R and S-PLUS function smooth.spline() uses

$$\kappa_k \simeq \left(\frac{k}{K+1}\right)\text{th sample quantile of the unique } x_i\text{s},$$

where

$$K = \begin{cases} n & n < 50 \\ 100 & n = 200 \\ 140 & n = 800 \\ 200 + (n - 3200)^{1/5} & n > 3200. \end{cases}$$

Other values of n between 50 and 3200 are handled by means of a logarithmic interpolation. For many functional relationships, fewer knots are sufficient. Figure 1 is one example, in which only $K = 20$ interior knots are used without compromising the quality of the fit. A common default in the penalized spline literature is $K = \min(n_U/4, 35)$, where n_U is the number of unique x_is (e.g. Ruppert *et al.*, 2003). Ruppert (2002) discusses 'hi-tech' choices of K. The distribution of the knots, for a given K, may have some effect on the results. As mentioned above, smooth.spline() uses quantile-based knots, whereas Eilers & Marx (1996), for example, recommend equally spaced knots. In most situations this effect will be minor.

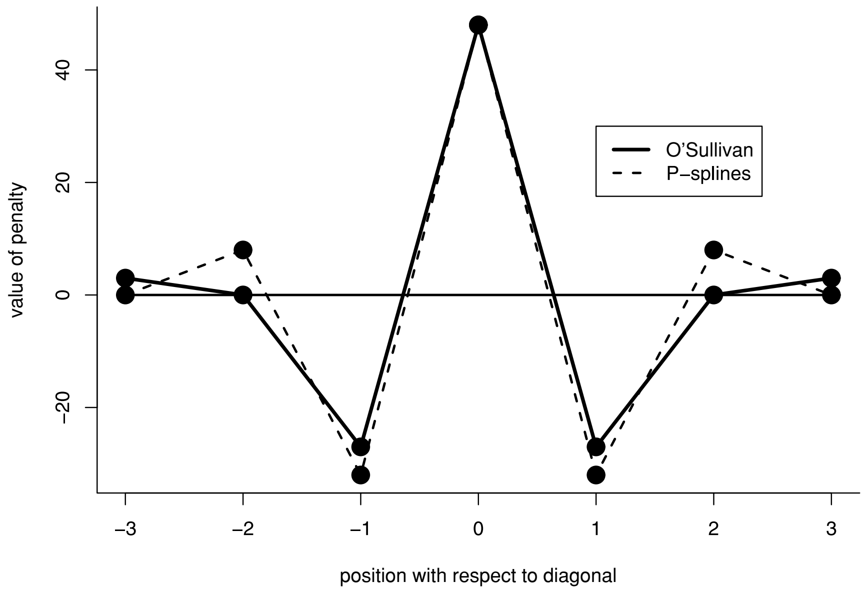

Figure 2. Comparison of near-diagonal entries of the penalty matrices for O'Sullivan penalized splines and cubic P-splines with $k = 2$ and equally spaced interior knots.

However, for either strategy it is possible to construct regression functions and predictor variable distributions for which problems arise. More sophisticated knot placement strategies may help. For example, Luo & Wahba (1997) propose more sophisticated basis function reduction methods that could be adapted to the current context.

3. Comparison with P-splines

The closest relatives of O'Sullivan penalized splines are the P-splines of Eilers & Marx (1996). If the interior knots $\kappa_5, \ldots, \kappa_{K+4}$ are taken to be equally spaced, the family of cubic P-splines is given by (2) with the $\boldsymbol{\Omega}$ replaced by $\boldsymbol{D}_k^\top \boldsymbol{D}_k$, where $\boldsymbol{D}_k$ is the kth-order differencing matrix. This differencing penalty corresponds to a discrete approximation to the integrated square of the kth derivative of the B-spline smoother. The choice $k = 2$ leads to the cubic P-spline estimate

$$\widehat{f}_P(x; \lambda) = \boldsymbol{B}_x \widehat{\boldsymbol{\nu}}_P, \quad \text{where} \quad \widehat{\boldsymbol{\nu}}_P \equiv \left(\boldsymbol{B}^\top \boldsymbol{B} + \lambda \boldsymbol{D}_2^\top \boldsymbol{D}_2\right)^{-1} \boldsymbol{B}^\top \boldsymbol{y}, \tag{4}$$

having the property that $\widehat{f}_P(\cdot; \lambda)$ approaches the least squares line as $\lambda \to \infty$. In this sense, (4) is the closest relative of $\widehat{f}_O(\cdot; \lambda)$. If the interior knots are equally spaced, the bands in the interior rows are, up to multiplicative factors, as follows:

O'Sullivan penalized splines (2):	3,	0,	−27,	48,	−27,	0,	3
Cubic P-splines; second-order difference (4):	0,	8,	−32,	48,	−32,	8,	0

Figure 2 facilitates visual comparison of the two. It is seen that the differences are relatively small, although not negligible.

What are the relative advantages of smoothers based on cubic P-splines and O'Sullivan penalized splines, or O-splines for short? A theoretical comparison between P-splines and O-splines in terms of estimation performance, perhaps in the spirit of Hall & Opsomer (2005), would be ideal – although this is beyond the scope of the current paper.

Eilers & Marx (1996) partially justify the use of P-splines rather than O-splines on the simplicity of the P-spline penalty matrix. However, as seen from (3), the penalty matrix needed for O-splines is straightforward to obtain. Furthermore, the discrete approximation of P-splines requires equally spaced knots which, depending on f, may not be desirable.

A possible advantage of P-splines is the option of higher-order penalties, although the resulting smoothers can have erratic extrapolation behaviour. A possible advantage of O-splines is their direct relationship with time-honoured smoothing splines, and their attractive theoretical properties (e.g. Nussbaum, 1985). From the results described in Section 2, it is clear that O-splines approach smoothing splines as $K \to n$. But how close are O-splines to smoothing splines for common (smaller) choices of K, and are they closer than P-splines with the same value of K and the same interior knots? To address these questions, we conducted an empirical study based on the 18 homoscedastic non-parametric regression settings in Wand (2000). For O-splines, we used $K = 100$ equally spaced interior knots with four repeated knots at each boundary, as described in Section 2. However, for P-splines, we used the knot sequence described in the Appendix of Eilers & Marx (1996), which involves extending the knots beyond the boundary, rather than repeating them. For each setting, 200 samples were generated, and smoothing spline estimates $\widehat{f}_S$, with the smoothing parameter chosen by means of the generalized cross-validation, were obtained. We then computed $\widehat{f}_O$ and $\widehat{f}_P$ to have the same effective degrees of freedom as $\widehat{f}_S$, and recorded closeness measures $d(\widehat{f}_O, \widehat{f}_S; A)$ and $d(\widehat{f}_P, \widehat{f}_S; A)$, where

$$d(f, g; A) \equiv \int_A (f - g)^2.$$

We took A corresponding to the intervals (a, κ_5) (left boundary), (κ_5, κ_{K+5}) (interior), (κ_{K+5}, b) (right boundary) and (a, b) (total region), where the κ_k denote the knots used for the O-spline fits. The Wand (2000) settings all involve predictor data within the unit interval. We took $(a, b) = (-0.1, 1.1)$ to assess behaviour beyond the range of the data. Wilcoxon tests on the 200 differences $d(\widehat{f}_O, \widehat{f}_S; A) - d(\widehat{f}_P, \widehat{f}_S; A)$ were carried out for each setting and choice of A. Apart from being distribution-free, Wilcoxon tests have the advantage of being invariant to normalization and to whether differences or ratios are used. In all 72 cases O-splines were closer to smoothing splines than P-splines, in the sense that the Wilcoxon P-value was less than 0.01.

To appreciate the practical significance of these results, we plotted the data and estimates at the 90th percentiles of each of the $d(\widehat{f}_O, \widehat{f}_S; A)$ and $d(\widehat{f}_P, \widehat{f}_S; A)$ samples, corresponding to relatively high discrepancies from $\widehat{f}_S$. Some examples are shown in Figure 3.

In the interior, all estimates based on the same data, and having the same degrees of freedom, are almost indistinguishable with the naked eye. However, big differences occur at the boundary. P-splines have a tendency to deviate from the natural boundary behaviour of smoothing splines. We also observed this phenomenon in the other 16 settings. Further study into this differing extrapolation behaviour would be worthwhile. We speculate that it comes

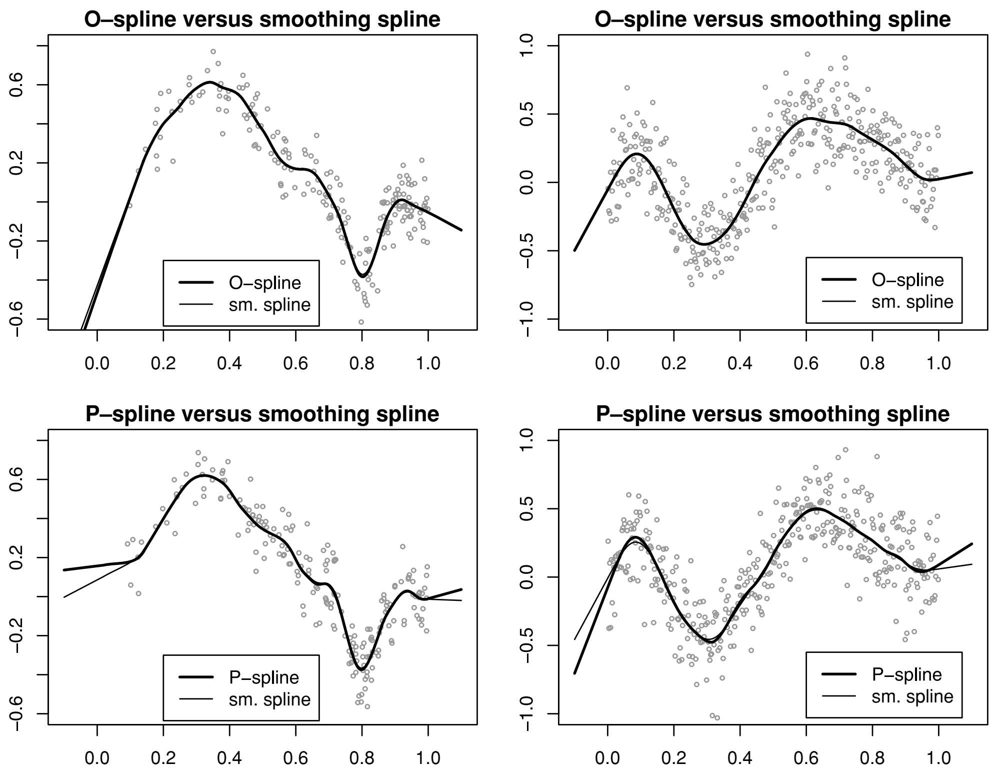

Figure 3. O-spline and P-spline fits compared with smoothing spline fits corresponding to the 90th percentiles of the $d(\widehat{f}_O, \widehat{f}_S; A)$ and $d(\widehat{f}_P, \widehat{f}_S; A)$ samples, for two of the homoscedastic settings of Wand (2000).

from differences between the exact integral penalty and its discrete approximation near the boundary.

4. Mixed model formulation

There are several ways by which $\widehat{\boldsymbol{\nu}}_O$ in (2) can be expressed as a best linear unbiased predictor (BLUP) in a mixed model (e.g. Speed, 1991; Verbyla, 1994). However, from a software standpoint, the most convenient form is $\widehat{\boldsymbol{\nu}}_O = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{u}})$ where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{u}}$ are (empirical) BLUPs of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ in the mixed model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \quad \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \boldsymbol{I} \end{bmatrix} \right) \tag{5}$$

for some design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$. An explicit expression for the BLUP in (5) (e.g. Ruppert *et al.*, 2003; Section 4.5.3) is

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \widehat{\boldsymbol{\nu}}_O = \left(\boldsymbol{C}^\top \boldsymbol{C} + \lambda \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \right)^{-1} \boldsymbol{C}^\top \boldsymbol{y}, \quad \lambda = \sigma_u^2 / \sigma_\varepsilon^2,$$

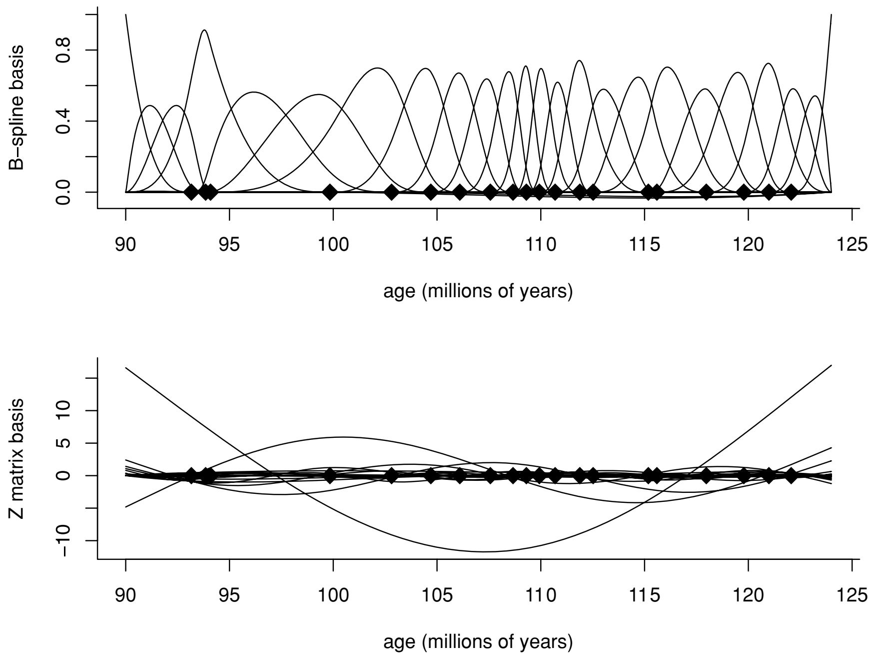

Figure 4. Comparison of the B-spline basis and the $\boldsymbol{Z}$ matrix basis for the fossil data example of Figure 2. The interior knots are shown as solid diamonds.

where $\boldsymbol{C} = [\boldsymbol{X} \mid \boldsymbol{Z}]$, and $\boldsymbol{I}$ is the identity matrix with the same number of columns as $\boldsymbol{Z}$. This 'canonical form' can be achieved if a $(K+4) \times (K+4)$ linear transformation matrix $\boldsymbol{L}$ can be found such that $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{L}$ and

$$\boldsymbol{L}^\top \boldsymbol{\Omega} \boldsymbol{L} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix}.$$

The usual method for obtaining $\boldsymbol{L}$ is spectral decomposition (e.g. Nychka & Cummins, 1996; Cantoni & Hastie, 2002; Welham *et al.*, 2007). It follows from results in the smoothing spline literature (e.g. Speed, 1991, Section 6) that

$$\text{rank}(\boldsymbol{\Omega}) = K + 2.$$

Hence, the spectral decomposition of $\boldsymbol{\Omega}$ is of the form $\boldsymbol{\Omega} = \boldsymbol{U}\ \text{diag}(\boldsymbol{d})\ \boldsymbol{U}^\top$, where $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{d}$ is a $(K+4) \times 1$ vector with exactly two zero entries and $K+2$ positive entries. Let $\boldsymbol{d}_Z$ be the $(K+2) \times 1$ sub-vector of $\boldsymbol{d}$ containing these positive entries, and let $\boldsymbol{U}_Z$ be the $(K+4) \times (K+2)$ sub-matrix of $\boldsymbol{U}$ with columns corresponding to the positive entries of $\boldsymbol{d}$. Then an appropriate linear transformation is $\boldsymbol{L} = [\boldsymbol{U}_X \mid \boldsymbol{U}_Z\ \text{diag}(\boldsymbol{d}_Z^{-1/2})]$. This leads to the fixed and random effects design matrices

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{U}_X \quad \text{and} \quad \boldsymbol{Z} = \boldsymbol{B}\boldsymbol{U}_Z \text{diag}\big(\boldsymbol{d}_Z^{-1/2}\big). \tag{6}$$

However, following again from the aforementioned smoothing spline literature (e.g. Speed, 1991, Section 6), $\boldsymbol{B}\boldsymbol{U}_X$ is a basis for the space of straight lines, so the simpler specification $\boldsymbol{X} = [1\ x_i]_{1 \le i \le n}$ may be used instead without affecting the fit. Figure 4 allows the comparison of

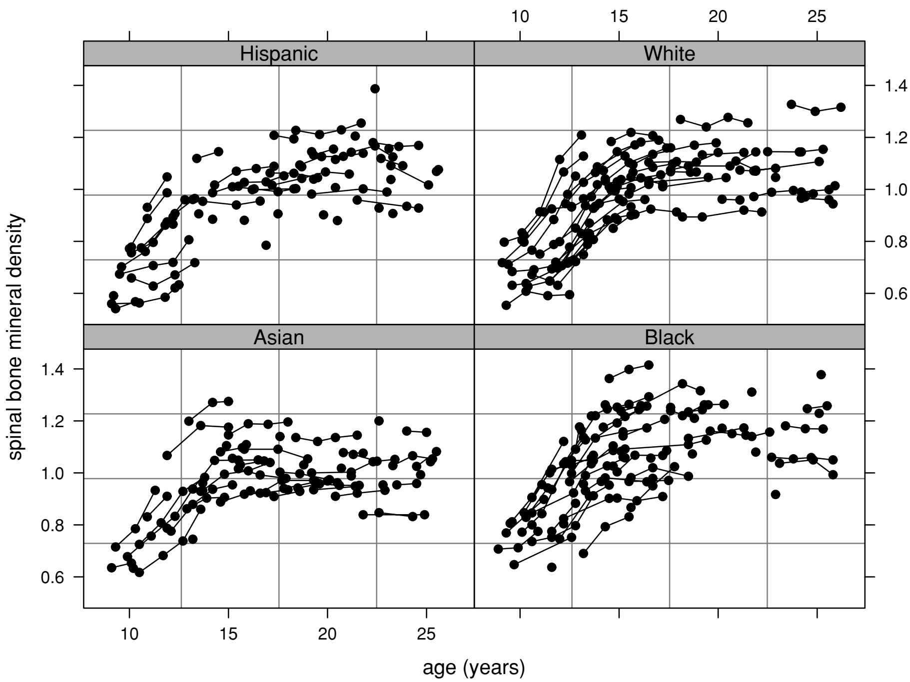

Figure 5. Spinal bone mineral data versus age for female subjects across four ethnicity groups (source: Bachrach *et al.*, 1999). Lines connect measurements taken from the same subject.

the original B-spline basis, corresponding to $\boldsymbol{B}$, and the basis corresponding to $\boldsymbol{Z}$. Notice the damping of the $\boldsymbol{Z}$ matrix basis functions with increasing oscillation. This compensates for the fact that the penalty is a multiple of the identity matrix.

In the Appendix, it is shown how the R linear mixed model function lme() can be used to obtain $\widehat{f_O}(\cdot;\lambda)$ based on (5), with $\boldsymbol{Z}$ given by (6). For simple scatterplot smoothing there is little difference between this approach and direct use of smooth.spline(), and the answers are equivalent if the knot sequence and λ values are equal. The default choice of λ differs: lme() uses restricted maximum likelihood (REML) to choose λ, whereas smooth.spline() uses generalized cross-validation (GCV).

The main advantage of the mixed model formulation of penalized splines is the incorporation into more complex models. Several examples are given in, for example, Ruppert *et al.* (2003). We will briefly describe one of them here. Figure 5 displays a longitudinal data set on bone mineral acquisition in young females (source: Bachrach *et al.*, 1999). The data consist of spinal bone mineral density (SBMD) measurements on each of 230 female subjects aged between 8 and 27. Each subject is measured between one and four times. Let n_i denote the number of measurements for subject i. The subjects have been divided into four ethnic groups: Asian, Black, Hispanic and White.

A useful additive mixed model for these data is

$$\begin{aligned}\texttt{SBMD}_{ij} = U_i + f(\texttt{age}_{ij}) + \beta_1\texttt{Black}_i + \beta_2\texttt{Hispanic}_i \\ + \beta_3\texttt{White}_i + \varepsilon_{ij}, \quad 1 \le i \le 230, \quad 1 \le j \le n_i, \qquad (7)\end{aligned}$$

where the U_i are independent and identically distributed (i.i.d.) N(0, σ_u^2) random intercepts for each subject, and $\texttt{Black}_i$, $\texttt{Hispanic}_i$ and $\texttt{White}_i$ are ethnicity indicators. The ε_{ij} are i.i.d. N (0, σ_ε^2) random errors. More sophisticated models that account for, say, serial correlation could be entertained.

O'Sullivan penalized splines can be used to fit (7) with the design matrices set up as follows. Based on the $\texttt{age}_i$ values and appropriate knots, set up

$$\boldsymbol{Z}_{\text{spline}} = \boldsymbol{B}\boldsymbol{U}_Z \text{diag}\big(\boldsymbol{d}_Z^{-1/2}\big)$$

analogous to the $\boldsymbol{Z}$ matrix of (6) for simple scatterplot smoothing. In the Appendix, when fitting data of this type, we use 15 interior knots corresponding to quantiles of the unique age values. Form

$$\boldsymbol{X} = \begin{bmatrix} 1 & \texttt{age}_{11} & \texttt{Black}_1 & \texttt{Hispanic}_1 & \texttt{White}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \texttt{age}_{1n_1} & \texttt{Black}_1 & \texttt{Hispanic}_1 & \texttt{White}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \texttt{age}_{230,1} & \texttt{Black}_{230} & \texttt{Hispanic}_{230} & \texttt{White}_{230} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \texttt{age}_{230,n_{230}} & \texttt{Black}_{230} & \texttt{Hispanic}_{230} & \texttt{White}_{230} \end{bmatrix} \text{ and}$$

$$\boldsymbol{Z}_{\text{subj}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Concatenate $\boldsymbol{Z}_{\text{subj}}$ and $\boldsymbol{Z}_{\text{spline}}$ to form

$$\boldsymbol{Z} = [\boldsymbol{Z}_{\text{subj}} | \boldsymbol{Z}_{\text{spline}}].$$

The appropriate mixed model is then

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \quad \text{cov}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \sigma_U^2 \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_u^2 \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \sigma_\varepsilon^2 \boldsymbol{I} \end{bmatrix}. \tag{8}$$

The Appendix contains R code for fitting this model. Note, in particular, that it circumvents explicit specification of $\boldsymbol{Z}_{\text{subj}}$. This is important for large longitudinal datasets.

5. Bayesian analysis and Markov chain Monte Carlo

A particularly attractive advantage of penalized splines, compared with smoothing splines, is the ease with which they can be fed into Markov chain Monte Carlo (MCMC) schemes for fitting Bayesian semiparametric regression models – as a result of the reduction in the number of basis functions. For simple scatterplot smoothing, this involves the Bayesian version of (5), namely

$$\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{u}, \sigma_\varepsilon^2 \sim \mathrm{N}\big(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \sigma_\varepsilon^2 \boldsymbol{I}\big), \quad \boldsymbol{u} \mid \sigma_u^2 \sim \mathrm{N}\big(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}\big),$$

and suitable (usually diffuse) prior distributions for $\boldsymbol{\beta}$, σ_u^2 and σ_ε^2. However, the big advantages of a Bayesian/MCMC approach are realized when handling complications such as measurement error (e.g. Carroll *et al.*, 2006) and generalized responses (e.g. Zhao *et al.*, 2006), which are hindered by intractable integrals in the likelihood.

Crainiceanu, Ruppert & Wand (2005) focussed on the use of the MCMC package WinBUGS (the Windows version of BUGS, Spiegelhalter *et al.*, 2000) for Bayesian penalized spline models. They reported that the choice of basis functions can have a substantial impact on the convergence of the chain. We decided to conduct some convergence checks for MCMC fitting of the regression model

$$\mathrm{logit}\{P(\texttt{union}_i = 1 \mid \texttt{wage}_i)\} = f(\texttt{wage}_i), \tag{9}$$

with f estimated by means of O'Sullivan penalized splines. Here $(\texttt{wage}_i, \texttt{union}_i)$, $1 \le i \le 534$, are pairs of wage amounts (dollars per hour) and trade union membership indicators for a sample of U.S. workers (source: Berndt, 1991). We expressed (9) as the Bayesian logistic mixed model:

$$\mathrm{logit}\{P(\texttt{union}_i = 1 \mid \texttt{wage}_i)\} = (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u})_i, \quad 1 \le i \le 534,$$

where $\boldsymbol{X} = [1 \ \texttt{wage}_i]_{1 \le i \le 534}$ and $\boldsymbol{Z} = \boldsymbol{B}\boldsymbol{U}_Z \ \mathrm{diag}\,(\boldsymbol{d}_Z^{-1/2})$, using the notation of Section 4. We used 15 interior knots with quantile spacing.

Following the advice of Zhao *et al.* (2006), we used WinBUGS to generate chains of length 5000 after a burn-in of 5000 and applied a thinning factor of 5, resulting in posterior samples of size 1000. Also in keeping with the recommendations of Zhao *et al.* (2006), we placed diffuse priors on the fixed effect parameters and variance component: β_0, β_1 independent N $(0, 10^8)$ and the prior density of σ_u^2 proportional to $(\sigma_u^2)^{-1.01} e^{-1/(100\sigma_u^2)}$, the inverse gamma distribution with shape and rate parameter both 0.01, after scaling the predictor to have unit variance. Zhao *et al.* (2006) found that the results can be sensitive to the choice of the inverse gamma hyperparameter.

The pointwise posterior mean effect of wage on the probability of trade union membership, together with 95% pointwise credible sets, is shown in Figure 6. Figure 7 allows assessment of the convergence of the MCMC at each quartile of the wage sample, and is seen to be excellent in each case. We also conducted convergence checks for larger logistic additive models involving up to six predictors and three smooth functions, and found the mixing to be very good when O-splines were used.

Several examples of semiparametric regression with WinBUGS, including code, are given in Crainiceanu *et al.* (2005) and Zhao *et al.* (2006).

6. General degree extension

Cubic O'Sullivan penalized splines have a natural extension to general odd-degree splines. Higher-degree splines have a role to play when smoother curve estimates are required.

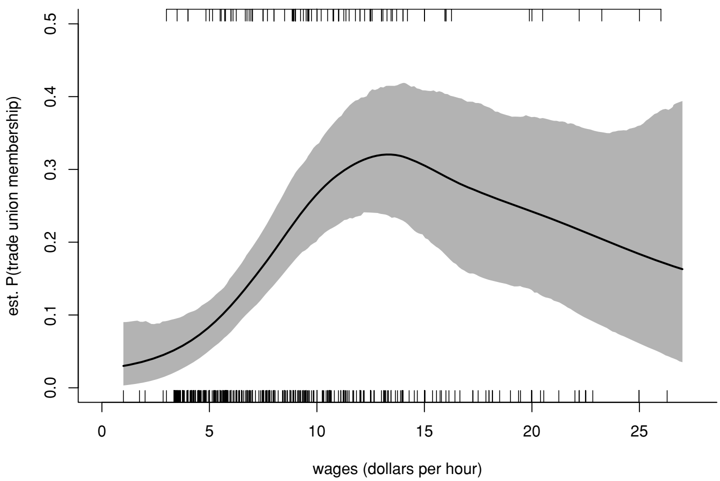

Figure 6. Fit of $\text{logit}\{P(\texttt{union}_i{=}1\,|\,\texttt{wage}_i)\} = f(\texttt{wage}_i)$ for trade union membership data (source: Berndt, 1991) using O'Sullivan penalized splines.

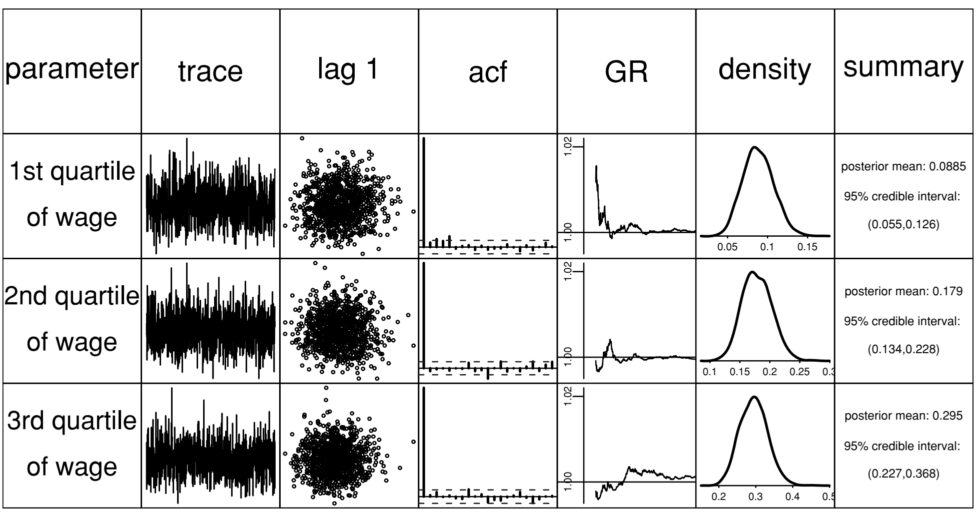

Figure 7. Assessment of MCMC convergence for the fit shown in Figure 6 at each quartile of wage. The columns are: quartile of wage, trace plot of sample of corresponding coefficient, plot of sample against 1-lagged sample, sample autocorrelation function, Gelman–Rubin $\sqrt{\widehat{R}}$ diagnostic, kernel estimates of posterior density and basic numerical summaries.

This arises, for example, in feature significance methodology (e.g. Chaudhuri & Marron, 1999; Hannig & Marron, 2006), in which first and second derivatives of the fit are required.

Return to the simple non-parametric regression setting (1) and let m be a general positive integer. Form the knot sequence

$$a = \kappa_1 = \cdots = \kappa_{2m} < \kappa_{2m+1} < \cdots < \kappa_{2m+K} < \kappa_{2m+K+1} = \cdots = \kappa_{4m+K} = b,$$

and let $B_{2m-1,1}, \ldots, B_{2m-1,K+2m}$ be the degree $(2m-1)$ B-spline basis defined by these knots. O'Sullivan penalized splines of order m then take the general form

$$\widehat{f}_O(x; m, \lambda) \equiv \boldsymbol{B}_{2m-1,x}\widehat{\boldsymbol{\nu}}_O, \quad \text{where} \quad \widehat{\boldsymbol{\nu}}_O \equiv \left(\boldsymbol{B}_{2m-1}^\top \boldsymbol{B}_{2m-1} + \lambda \boldsymbol{\Omega}^{(m)}\right)^{-1} \boldsymbol{B}_{2m-1}^\top \boldsymbol{y}.$$

Here $\boldsymbol{B}_{2m-1}$ is the $n \times (K+4m)$ design matrix with (i,k)th entry $B_{2m-1,k}(x_i)$, $\boldsymbol{B}_{2m-1,x} \equiv [B_{2m-1,1}(x), \ldots, B_{2m-1,K+2m}(x)]$, and $\boldsymbol{\Omega}^{(m)}$ is the $(K+2m) \times (K+2m)$ penalty matrix with (k, k')th entry

$$\boldsymbol{\Omega}_{kk'}^{(m)} = \int_a^b B_{2m-1,k}^{(m)}(x) B_{2m-1,k'}^{(m)}(x)\, dx.$$

In the special case in which the interior knots coincide with the x_is, assumed distinct, $\widehat{f}_O(\cdot; m, \lambda)$ corresponds to the order m smoothing spline; that is, the minimizer of

$$\sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b f^{(m)}(x)^2\, dx$$

(e.g. Schoenberg, 1964).

We are now ready to state our result for the exact computation of O'Sullivan spline penalty matrices.

Theorem. *The penalty matrix $\boldsymbol{\Omega}^{(m)}$ admits the exact explicit expression*

$$\boldsymbol{\Omega}^{(m)} = \left(\widetilde{\boldsymbol{B}}^{(m)}\right)^\top \operatorname{diag}(\boldsymbol{w}) \widetilde{\boldsymbol{B}}^{(m)},$$

where $\widetilde{\boldsymbol{B}}^{(m)}$ is the $(2m-1)(K+4m-1) \times (K+2m)$ matrix with (i,j)th entry $B_{2m-1,j}^{(m)}(\widetilde{x}_i)$, and $\boldsymbol{w}$ is a $(2m-1)(K+4m-1) \times 1$ vector with ith entry w_i. The $\widetilde{x}_i$ and w_i values are obtained according to

$$\widetilde{x}_{(2m-1)(\ell-1)+\ell'+1} = \kappa_\ell + \ell' h_{m,\ell}, \quad w_{(2m-1)(\ell-1)+\ell'+1} = h_{m,\ell}\,\omega_{m,\ell'}$$

for $1 \le \ell \le K+4m-1$, $0 \le \ell' \le 2m-2$. Here, for $1 \le k \le K+2m$, $h_{1,k} = \kappa_{k+1} - \kappa_k$, and, for $m \ge 2$, $h_{m,k} = (\kappa_{k+1} - \kappa_k)/(2m-2)$. Finally, for all $m \ge 1$,

$$\omega_{m,k} = \frac{(-1)^k}{k!(2m-2-k)!} \int_0^{2m-2} \frac{t(t-1)\cdots(t-2m+2)}{t-k}\, dt, \quad k = 0, \ldots, 2m-2.$$

Proof. The (k, k')th entry of $\boldsymbol{\Omega}^{(m)}$ is

$$\boldsymbol{\Omega}_{kk'}^{(m)} = \int_a^b B_{2m-1,k}^{(m)}(x) B_{2m-1,k'}^{(m)}(x)\, dx = \sum_{i=1}^{K+4m-1} \int_{\kappa_i}^{\kappa_{i+1}} B_{2m-1,k}^{(m)}(x) B_{2m-1,k'}^{(m)}(x)\, dx. \quad (10)$$

Because $B_{2m-1,k}^{(m)}(x)$, $B_{2m-1,k'}^{(m)}(x)$ are degree $m-1$ polynomials on each interval $x \in (\kappa_i, \kappa_{i+1})$ for $1 \le i \le K+4m-1$, the function $B_{2m-1,k}^{(m)}(x)\, B_{2m-1,k'}^{(m)}(x)$ is a degree $2(m-1)$ polynomial on the same interval. The result follows by applying to the right-hand side of (10) the Newton–Cotes integration $(2m-1)$-point rule (e.g. Whittaker & Robinson, 1967) which is exact for polynomials of degree $2(m-1)$ or lower.

TABLE 1
Table of $\omega_{m,k} \equiv \frac{(-1)^k}{k!(2m-2-k)!} \int_0^{2m-2} \frac{t(t-1)\cdots(t-2m+2)}{t-k} \, dt$ *values for* $m \le 4$

$m \backslash k$	0	1	2	3	4	5	6
1	1						
2	1/3	4/3	1/3				
3	14/45	64/45	8/15	64/45	14/45		
4	41/140	54/35	27/140	68/35	27/140	54/35	41/140

Table 1 provides values of $\omega_{m,k}$ for O'Sullivan polynomials up to degree 7. This, together with the Theorem, allows direct computation of penalty matrices of O'Sullivan splines for $m \le 4$. Higher values of m require one-off calculation of the $\omega_{k,m}$ through, say, a symbolic computation package such as MAPLE (Waterloo Maple Inc., 2007).

Recall from Section 2 that, in the case of cubic O'Sullivan splines, Newton–Cotes integration reduces to Simpson's rule, and a simpler, more revealing, expression results in the shape of (3).

7. Closing remarks

Smoothing splines have a special place in non-parametric and semiparametric regression. They are based on simple and intuitive principles, have an attractive theory (e.g. Nussbaum, 1985; Wahba, 1990; Eubank, 1994; Solo, 2000), and possess good practical properties such as natural boundary behaviour. Penalized splines, including P-splines, have gained popularity for reasons stated in the Introduction. However, proponents of penalized splines have been viewed by some, especially in the smoothing spline community, as ignoring the benefits that have been established for smoothing splines over the past few decades. O'Sullivan penalized splines, being a direct generalization and closer approximation of smoothing splines, provide an attractive link between the two streams of semiparametric regression research, and allow analysts to enjoy the best of both worlds.

Appendix: R implementation

In this Appendix we provide R code for the use of O'Sullivan penalized splines in the simplest semiparametric regression setting: scatterplot smoothing. The extension to more complex models, such as those described by Ngo & Wand (2004) and Crainiceanu *et al.* (2005), is straightforward. We illustrate one of these extensions: additive mixed models.

Direct scatterplot smoothing with user choice of smoothing parameter

Obtain scatterplot data corresponding to environmental data from the R package LATTICE. Set up plotting grid, knots and smoothing parameter:

```
library(lattice) ; attach(environmental)
x <- radiation ; y <- ozone^(1/3)
a <- 0 ; b <- 350 ; xg <- seq(a,b,length=101)
numIntKnots <- 20 ; lambda <- 1000
```

Set up the design matrix and related quantities:

```
library(splines)
```

```
intKnots <- quantile(unique(x),seq(0,1,length=
        (numIntKnots+2))[-c(1,(numIntKnots+2))])
names(intKnots) <- NULL
B <- bs(x,knots=intKnots,degree=3,
      Boundary.knots=c(a,b),intercept=TRUE)
BTB <- crossprod(B,B) ; BTy <- crossprod(B,y)
```

Create the $\mathbf{\Omega}$ matrix:

```
formOmega <- function(a,b,intKnots)
{
   allKnots <- c(rep(a,4),intKnots,rep(b,4))
   K <- length(intKnots) ; L <- 3*(K+8)
   xtilde <- (rep(allKnots,each=3)[-c(1,(L-1),L)]+
              rep(allKnots,each=3)[-c(1,2,L)])/2
   wts <- rep(diff(allKnots),each=3)*rep(c(1,4,1)/6,K+7)
   Bdd <- spline.des(allKnots,xtilde,derivs=rep(2,length(xtilde)),
                  outer.ok=TRUE)$design
   Omega <- t(Bdd*wts)
   return(Omega)
}
Omega <- formOmega(a,b,intKnots)
```

Obtain the coefficients:

```
nuHat <- solve(BTB+lambda*Omega,BTy)
```

For large K, the following alternative Cholesky-based approach can be considerably faster ($O(K)$, because $\boldsymbol{B}^\top \boldsymbol{B} + \lambda\mathbf{\Omega}$ is banded diagonal):

```
cholFac <- chol(BTB+lambda*Omega)
nuHat <- backsolve(cholFac,forwardsolve(t(cholFac),BTy))
```

Display the fit:

```
Bg <- bs(xg,knots=intKnots,degree=3,Boundary.knots=c(a,b),
intercept=TRUE)
 fhatg <- Bg
     plot(x,y,xlim=range(xg),bty="l",type="n",xlab="radiation",
     ylab="cuberoot of ozone",main="(a) direct fit; user
     choice of smooth. par.")
lines(xg,fhatg,lwd=2)
points(x,y,lwd=2)
```

Mixed model scatterplot smoothing with REML choice of smoothing parameter

Obtain the spectral decomposition of $\mathbf{\Omega}$:

```
eigOmega <- eigen(Omega)
```

Obtain the matrix for the linear transformation of $\boldsymbol{B}$ to $\boldsymbol{Z}$:

```
indsZ <- 1:(numIntKnots+2)
UZ <- eigOmega$vectors[,indsZ]
LZ <- t(t(UZ)/sqrt(eigOmega$values[indsZ]))
```

Perform a stability check:

```
indsX <- (numIntKnots+3):(numIntKnots+4)
UX <- eigOmega$vectors[,indsX]
L <- cbind(UX,LZ)
stabCheck <- t(crossprod(L,t(crossprod(L,Omega))))
if (sum(stabCheck^2) > 1.0001*(numIntKnots+2))
    print("WARNING: NUMERICAL INSTABILITY ARISING FROM SPECTRAL
          DECOMPOSITION")
```

Form the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices:

```
X <- cbind(rep(1,length(x)),x)
Z <- B%*%LZ
```

Fit using lme() with REML choice of smoothing parameter:

```
library(nlme)
group <- rep(1,length(x))
gpData <- groupedData(y x|group,data=data.frame(x,y))
fit <- lme(y -1+X,random=pdIdent(-1+Z),data=gpData)
```

Extract coefficients and plot scatterplot smooth over a grid:

```
betaHat <- fit$coef$fixed
uHat <- unlist(fit$coef$random)
Zg <- Bg%*%LZ
fhatgREML <- betaHat[1] + betaHat[2]*xg + Zg
plot(x,y,xlim=range(xg),bty="l",type="n",xlab="radiation",
      ylab="cuberoot of ozone",main="(b) mixed model fit;
      REML choice of smooth. par.")
lines(xg,fhatgREML,lwd=2)
points(x,y,lwd=2)
```

Execution of the above code leads to Figure 8.

Fitting an additive mixed model

The spinal bone mineral density data of Bachrach *et al.* (1999) are not publicly available. Therefore we will illustrate the fitting of additive mixed models using simulated data. For simplicity, we will use two ethnicity categories rather than four.

Generate data:

```
set.seed(394600) ; m <- 230 ; nVals <- sample(1:4,m,replace=TRUE)
betaVal <- 0.1 ; sigU <- 0.25 ; sigEps <- 0.05
f <- function(x)
   return(1 + pnorm((2*x-36)/5)/2)
```

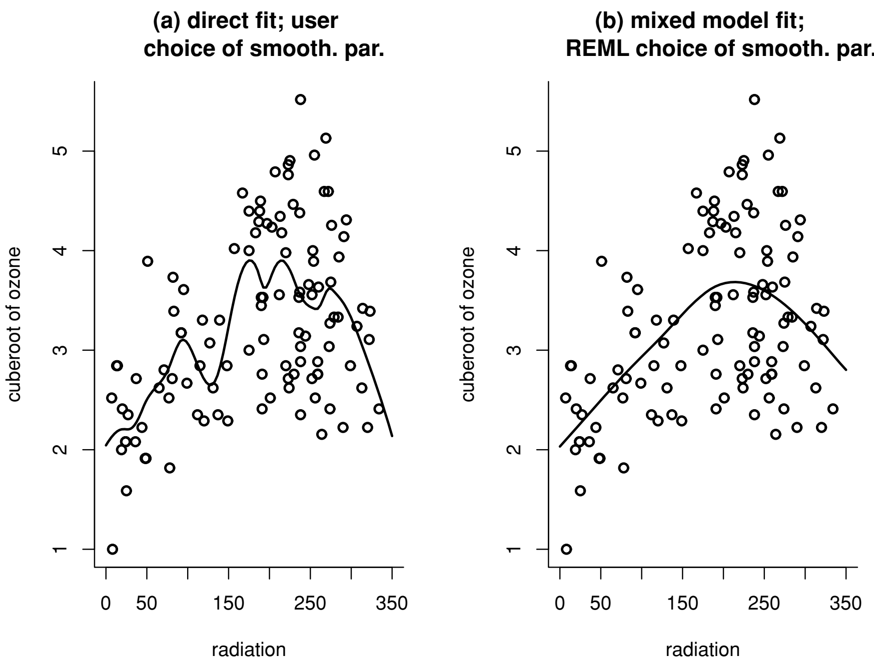

Figure 8. Plots obtained from execution of the first two chunks of code in this Appendix.

```
U <- rnorm(m,0,sigU)
age <- NULL ; ethnicity <- NULL
Uvals <- NULL ; idNum <- NULL
for (i in 1:m)
{
   idNum <- c(idNum,rep(i,nVals[i]))
   stt <- runif(1,8,28-(nVals[i]-1))
   age <- c(age,seq(stt,by=1,length=nVals[i]))
   xCurr <- sample(c(0,1),1)
   ethnicity <- c(ethnicity,rep(xCurr,nVals[i]))
   Uvals <- c(Uvals,rep(U[i],nVals[i]))
}
epsVals <- rnorm(sum(nVals),0,sigEps)
SBMD <- f(age) + betaVal*ethnicity + Uvals + epsVals
```

Set up basic variables for the spline component:

```
a <- 8 ; b <- 28; numIntKnots <- 15
intKnots <- quantile(unique(age),seq(0,1,length=
                  (numIntKnots+2))[-c(1,(numIntKnots+2))])
```

Obtain the spline component of the $\boldsymbol{Z}$ matrix:

```
B <- bs(age,knots=intKnots,degree=3,
        Boundary.knots=c(a,b),intercept=TRUE)
```

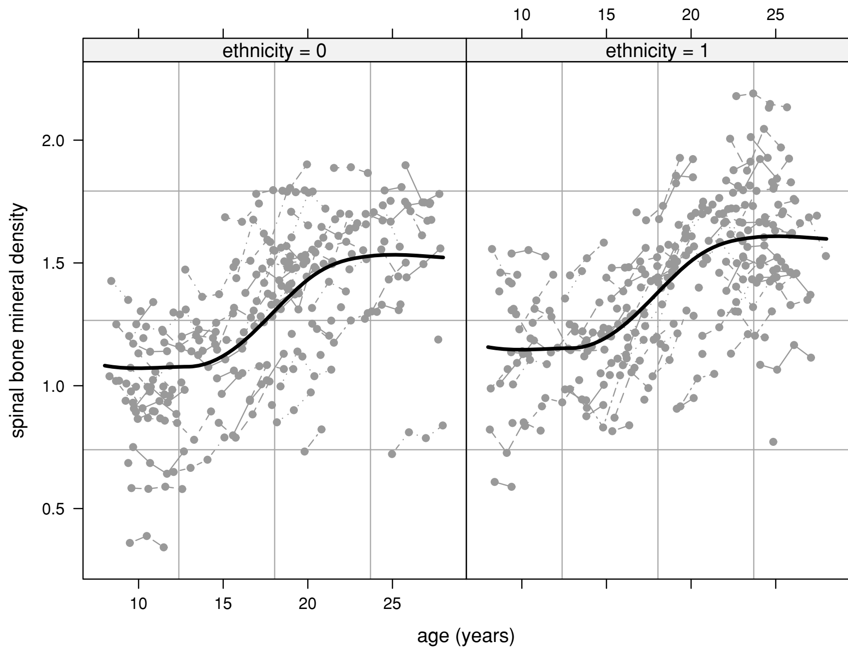

Figure 9. Plot obtained from execution of the last chunk of code in this Appendix.

```
Omega <- formOmega(a,b,intKnots)
eigOmega <- eigen(Omega)
indsZ <- 1:(numIntKnots+2)
UZ <- eigOmega$vectors[,indsZ]
LZ <- t(t(UZ)/sqrt(eigOmega$values[indsZ]))
ZSpline <- B
```

Obtain the $\boldsymbol{X}$ matrix:

```
X <- cbind(rep(1,length(SBMD)),age,ethnicity)
```

Set up the variables required for fitting via lme(). Note that the random intercept is taken care of via the tree identification numbers variable `idNum`, and that explicit formation of the random effect contribution to the $\boldsymbol{Z}$ matrix is not required.

```
groupVec <- factor(rep(1,length(SBMD)))
ZBlock <- list(list(groupVec=pdIdent(ZSpline-1)),list(idNum=
pdIdent(1)))
ZBlock <- unlist(ZBlock,recursive=FALSE)
dataFr <- groupedData(SBMD ethnicity|groupVec,
                      data=data.frame(SBMD,X,ZSpline,idNum))
fit <- lme(SBMD -1+X,data=dataFr,random=ZBlock)
betaHat <- fit$coef$fixed
uHat <- unlist(fit$coef$random)
uSplineHat <- uHat[1:ncol(ZSpline)]
```

Plot the data and fitted curve estimates together:

```
ng <- 101 ; ageg <- seq(a,b,length=ng)
Bg <- bs(ageg,knots=intKnots,degree=3,Boundary.knots=c(a,b),
intercept=TRUE)
ZgSpline <- Bg

plotMatrix0 <- cbind(rep(1,ng),ageg,rep(0,ng),ZgSpline)
fhatgREML <- plotMatrix0

xLabs <- paste("ethnicity =",as.character(ethnicity))
pobj <- xyplot(SBMD age|xLabs,groups=idNum,xlab="age (years)",
        ylab="spinal bone mineral density",subscripts=TRUE,
        panel=function(x,y,subscripts,groups)
        {
           panel.grid() ; panel.superpose(x,y,subscripts,groups,
                                type="b",col="grey60",pch=16)
           panelInd <- any(ethnicity[subscripts]==1)
           panel.xyplot(ageg,fhatgREML+panelInd*betaHat[3],
                       lwd=3,type="l",col="black")
        })
print(pobj)
```

Print approximate 95% confidence intervals for key parameters:

```
print(intervals(fit))
```

Execution of the above code should lead to an outcome similar to that in Figure 9 (the simulated data may differ between platforms).

References

BACHRACH, L.K., HASTIE, T., WANG, M.-C., NARASIMHAN, B. & MARCUS, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, Black and Caucasian youth. A longitudinal study. *J. Clin. Endocrin. Metab.* **84**, 4702–4712.

BERNDT, E.R. (1991). *The Practice of Econometrics: Classical and Contemporary*. Reading, MA: Addison-Wesley.

CANTONI, E. & HASTIE, T. (2002). Degrees of freedom tests for smoothing splines. *Biometrika* **89**, 251–265.

CARROLL, R.J., RUPPERT, D., STEFANSKI, L.A. & CRAINICEANU, C.M. (2006). *Measurement Error in Nonlinear Models*, 2nd edn. Boca Raton, FL: Chapman & Hall/CRC.

CHAUDHURI, P. & MARRON, J.S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94**, 807–823.

CRAINICEANU, C., RUPPERT, D. & WAND, M.P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *J. Statist. Software*, Volume 14, Issue 14.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Berlin: Springer-Verlag.

DENISON, D.G.T., HOLMES, C.C., MALLICK, B.K. & SMITH, A.F.M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. Chichester, UK: Wiley.

EILERS, P.H.C. & MARX, B.D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.* **11**, 89–121.

EUBANK, R.L. (1994). A simple smoothing spline. *Amer. Statist.* **48**, 103–106.

EUBANK, R.L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.

GREEN, P.J. & SILVERMAN, B.W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

GU, C. (2002). *Smoothing Spline ANOVA Models*. New York: Springer.

HALL, P. & OPSOMER, J.D. (2005). Theory for penalized spline regression. *Biometrika* **92**, 105–118.

HANNIG, J. & MARRON, J.S. (2006). Advanced distribution theory for SiZer. *J. Amer. Statist. Assoc.* **101**, 484–499.

HASTIE, T. (2006). GAM 0.97. R package. http://cran.r-project.org.

HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. (2001). *The Elements of Statistical Learning*. New York: Springer-Verlag.

INSIGHTFUL CORPORATION (2007). Seattle, WA.

LUO, Z. & WAHBA, G. (1997). Hybrid adaptive splines. *J. Amer. Statist. Assoc.* **92**, 107–116.

NGO, L. & WAND, M.P. (2004). Smoothing with mixed model software. *J. Statist. Software*, Volume 9, Issue 1.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13**, 984–997.

NYCHKA, D. & CUMMINS, D.J. (1996). Comment on paper by Eilers and Marx. *Statist. Sci.* **11**, 104–105.

O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.* **1**, 505–527.

R DEVELOPMENT CORE TEAM (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

RUPPERT, D. (2002). Selecting the number of knots for penalized splines. *J. Comput. Graph. Statist.* **11**, 735–757.

RUPPERT, D., WAND, M.P. & CARROLL, R.J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

SAS INSTITUTE, INC. (2007). Cary, NC.

SCHOENBERG, I.J. (1964). Spline functions and the problem of gradation. *Proc. Natl. Acad. Sci. USA* **52**, 947–950.

SOLO, V. (2000). A simple derivation of the smoothing spline. *Amer. Statist.* **54**, 40–43.

SPEED, T. (1991). Comment on paper by Robinson. *Statist. Sci.* **6**, 42–44.

SPIEGELHALTER, D., THOMAS, A. & BEST, N. (2000). WinBUGS Version 1.3 User Manual. www.hrc-bsu.cam.ac.uk/bugs.

THE MATHWORKS, INC. (2007). Natick, MA.

VERBYLA, A.P. (1994). Testing linearity in generalized linear models. In: *17th International Biometrics Conference.* Hamilton (Ontario), Canada., 177.

WAHBA, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

WAND, M.P. (2000). A comparison of regression spline smoothing procedures. *Comput. Statist.* **15**, 443–462.

WATERLOO MAPLE INC. (2007). Waterloo (Ontario), Canada.

WELHAM, S.J., CULLIS, B.R., KENWARD, M.G. & THOMPSON, R. (2007). A comparison of mixed model splines for curve fitting. *Aust. N. Z. J. Statist.* **49**, 1–23.

WHITTAKER, E.T. & ROBINSON, G. (1967). The Newton–Cotes formulae of integration. Section 76 in *The Calculus of Observations: A Treatise on Numerical Mathematics*, 4th edn. pp. 152–156. New York: Dover.

ZHAO, Y., STAUDENMAYER, J., COULL, B.A. & WAND, M.P. (2006). General design Bayesian generalized linear mixed models. *Statist. Sci.* **21**, 35–51.