Computational Statistics (2003) 18:251–262

A note on P-spline additive models with correlated errors

Maria Durbán[1] and Iain D. Currie[2]

[1]Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Leganés, Madrid, 28911, Spain
[2] Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK

Summary

We consider additive models with k smooth terms and correlated errors, and use the penalised spline approach of Eilers & Marx (1996) to estimate the smooth functions. We obtain explicit expressions for the hat-matrix of the model and each individual curve. P-splines are represented as mixed models and REML is used to select the smoothing and correlation parameters. The method is applied to the analysis of some time series data.

Keywords: Additive model, P-spline, REML, back-fitting, serial correlation, mixed models

1 Introduction

Additive models (Buja, Hastie & Tibshirani 1989, Hastie & Tibshirani 1990) represent a response variable y as the sum of k smooth terms which act on explanatory variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$,

$$\mathrm{E}(y|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \alpha + f_1(\boldsymbol{x}_1) + \ldots + f_k(\boldsymbol{x}_k), \quad E(f_i(\boldsymbol{x}_i)) = 0.$$

Estimation of each $f_i(\boldsymbol{x}_i)$ is often achieved through a *scatterplot smoother.* Various authors have used different smoothers based on the objectives of their analysis: Opsomer & Ruppert (1999) used local linear regression as the smoothing method, in Durban, Hackett & Currie (1999) smoothing was done by loess and cubic smoothing splines, and Marx & Eilers (1998) and Coull, Ruppert & Wand (2001) proposed the use of penalised spline models. For many years, the back-fitting algorithm was the standard method of obtaining estimates in an additive model. Back-fitting is computationally efficient but it has two main disadvantages: 1) as an iterative procedure, it does not give explicit expressions for the estimated smooth terms, and 2) there is no back-fitting algorithm for an additive model with autocorrelated errors. Durban et al. (1999) and Opsomer (2000) gave explicit expressions for the backfitting estimators and studied their theoretical properties. The P-spline approach of Marx & Eilers (1998) has the advantage that no backfitting is needed; this approach has been used in several recent papers which use the fact that a penalised spline may be expressed as a mixed model (Coull, Ruppert & Wand 2001, Coull, Schwartz & Wand 2001). Recently Aerts, Claeskens & Wand (2002) used P-splines with a truncated lines basis and gave explicit expressions for the individual estimated curves; they did not consider a mixed model approach.

Smoothing in the presence of correlated errors has become an area of recent interest. For many years the main difficulty has been the joint estimation of the smoothing and correlation parameters, since the usual criteria for choosing the smoothing parameter tend to undersmooth (positive correlation) or oversmooth (negative correlation) the data when the assumption of independent data is relaxed (see for example Altman 1990, Hart 1991). More recently, Wang (1998) related a smoothing spline model to a mixed-effect model and showed that generalized maximum likelihood (GML) estimates of the smoothing parameter and correlation parameter are restricted maximum likelihood (REML) estimates. Currie & Durban (2002) used the mixed model representation of P-splines and showed that the number of knots selected to fit the P-spline has almost no influence on the estimation of the correlation parameter; they used likelihood ratio tests to select the order of the autoregressive process. Coull, Schwartz & Wand (2001) used additive models and allowed for correlation among repeated measurements, while Smith, Wong & Kohn (1998) presented a Bayesian approach for nonparametric estimation of an additive model with autocorrelated errors.

The plan of the paper is as follows. In section 2 we follow the P-spline approach of Eilers & Marx (1996) and give simple expressions for the estimated curves in an additive model with correlated errors; this allows us to give confidence intervals for the fitted functions and to use likelihood ratio tests for model selection. In section 3 we give a mixed model representation of

P-splines with correlated errors and in section 4 we apply our results to the analysis of some electricity consumption data.

2 A P-spline approach to nonparametric regression

We suppose that the response $\boldsymbol{y}$ depends on the predictor $\boldsymbol{x}$ in a smooth fashion, and so the model can be written

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon} \tag{1}$$

where, in this section, we assume for simplicity that $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.
Eilers & Marx (1996) introduced penalised B-splines (which they termed P-splines) as an alternative approach to nonparametric regression. B-splines are constructed from polynomial pieces of degree $bdeg$ which are joined at certain values of x, called knots; we suppose that the knots divide the range of x into ndx intervals ('n' intervals of length 'dx'). An important point is that typically the number of knots is smaller than the number of data points n. Eilers and Marx make two key assumptions: (a) $\mathrm{E}(y_i) = \sum_{j=1}^{k} a_j B_j(x_i)$, where $\boldsymbol{B} = (B_1(\boldsymbol{x}), \ldots, B_k(\boldsymbol{x}))$ is an $n \times k$ regression matrix with $k = bdeg + ndx$; (b) the regression coefficients a_j satisfy certain smoothness conditions based on finite differences of adjacent coefficients. Eilers and Marx then estimate the coefficients by minimising the following penalised least squares function

$$S = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}) + \lambda \boldsymbol{a}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{a}. \tag{2}$$

The parameter λ controls the smoothness of the fit with larger values of λ corresponding to smoother fitted curves; the matrix $\boldsymbol{D}$ is a difference matrix of dimension $(ndx + bdeg - pord) \times (ndx + bdeg)$, ($pord$ is the order of the penalty). The solution to the minimisation of (2) (conditional on the value of λ) is called a P-spline:

$$\hat{\boldsymbol{a}} = (\boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{B}'\boldsymbol{y}; \tag{3}$$

the fitted value of $\boldsymbol{y}$ is $\hat{\boldsymbol{y}} = \boldsymbol{B}\hat{\boldsymbol{a}} = \boldsymbol{H}\boldsymbol{y}$ where $\boldsymbol{H}$ is the hat-matrix

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{B}'. \tag{4}$$

2.1 P-splines with a general error structure

Much past work on smoothing is based on the assumptions of independent errors with constant variance. However, in many applications, these assumptions do not hold: time series, spatially correlated data or data with heteroscedastic errors are all examples where the simple assumptions in (1) do not hold. We extend (1) to the more general model $\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{V}$; the matrix $\boldsymbol{V}$ will be a matrix of weights in the case of heteroscedastic errors, or it will depend on one or more parameters in the case of correlated errors. Expression (2) becomes

$$S = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}) + \lambda \boldsymbol{a}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{a}, \tag{5}$$

a weighted version of the previous penalised least squares. For a given λ, the hat-matrix for this model is

$$\boldsymbol{H}_V = \boldsymbol{B}\left(\boldsymbol{B}'\boldsymbol{V}^{-1}\boldsymbol{B} + \lambda \boldsymbol{D}'\boldsymbol{D}\right)^{-1}\boldsymbol{B}'\boldsymbol{V}^{-1}. \tag{6}$$

2.2 Additive models

We suppose that the response $\boldsymbol{y}$ is explained by the additive model

$$\boldsymbol{y} = \alpha\boldsymbol{1} + f_1(\boldsymbol{x}_1) + \ldots + f_k(\boldsymbol{x}_k) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}). \tag{7}$$

One common method of estimation for an additive model is backfitting (Buja et al. 1989, Hastie & Tibshirani 1987). This method is computationally very efficient and can be used with any linear smoother (which includes P-splines), but, as an iterative procedure, it does not give explicit expressions for the estimated smooth curves. Durban et al. (1999) and more extensively (Durban 1999) gave explicit solutions to the backfitting algorithm; (Opsomer 2000) gave asymptotic properties of the backfitting estimators in the case of local polynomials. Here we give equivalent results when the smooth functions are estimated by P-splines with a B-spline basis.
We introduce an intercept term α to avoid identifiability problems, since otherwise the fitted curves would be unique only up to a constant. Thus we assume $\mathrm{E}(\boldsymbol{y}) = \alpha\boldsymbol{1}$ and define the centred regression matrices

$$\boldsymbol{B}_i^* = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}'/n)\boldsymbol{B}_i \tag{8}$$

with the properties that $\boldsymbol{B}_i^*\boldsymbol{1} = \boldsymbol{0}$ and $\boldsymbol{B}_i^{*\prime}\boldsymbol{1} = \boldsymbol{0}$, as in (Durban et al. 1999). We now write (7) in the form

$$\boldsymbol{y} = \alpha\boldsymbol{1} + \boldsymbol{B}_1^*\boldsymbol{a}_1 + \ldots + \boldsymbol{B}_k^*\boldsymbol{a}_k + \boldsymbol{\epsilon} = \boldsymbol{B}\boldsymbol{a} + \boldsymbol{\epsilon} \tag{9}$$

where $\boldsymbol{B} = (\boldsymbol{1} : \boldsymbol{B}_1^* : \ldots : \boldsymbol{B}_k^*)$ and $\boldsymbol{a} = (\alpha, \boldsymbol{a}_1', \ldots, \boldsymbol{a}_k')'$. We choose $\boldsymbol{a}$ by minimising

$$S = (\boldsymbol{y} - \alpha\boldsymbol{1} - \boldsymbol{B}_1^*\boldsymbol{a}_1 - \ldots - \boldsymbol{B}_k^*\boldsymbol{a}_k)'(\boldsymbol{y} - \alpha\boldsymbol{1} - \boldsymbol{B}_1^*\boldsymbol{a}_1 - \ldots - \boldsymbol{B}_k^*\boldsymbol{a}_k) + \boldsymbol{a}'\boldsymbol{P}\boldsymbol{a} \tag{10}$$

where $\boldsymbol{P} = \mathrm{blockdiag}(0, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_k)$ with $\boldsymbol{P}_i = \lambda_i \boldsymbol{D}_i'\boldsymbol{D}_i$. Taking derivatives with respect to each component of $\boldsymbol{a}$ we obtain the following system of equations:

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{B}_1^{*\prime}\boldsymbol{B}_1^* + \lambda_1 \boldsymbol{D}_1'\boldsymbol{D}_1 & \dots & \boldsymbol{B}_1^{*\prime}\boldsymbol{B}_k^* \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \boldsymbol{B}_k^{*\prime}\boldsymbol{B}_1^* & \dots & \boldsymbol{B}_k^{*\prime}\boldsymbol{B}_k^* + \lambda_k \boldsymbol{D}_k'\boldsymbol{D}_k \end{bmatrix} \begin{bmatrix} \alpha \\ \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_k \end{bmatrix} = \begin{bmatrix} \mathbf{1}' \\ \boldsymbol{B}_1^{*\prime} \\ \vdots \\ \boldsymbol{B}_k^{*\prime} \end{bmatrix} \boldsymbol{y}. \tag{11}$$

This is equivalent to the set of normal equations in (Hastie & Tibshirani 1987). In Appendix A we show that the hat-matrix $\boldsymbol{H}_k$ for model (9) can be written

$$\boldsymbol{H}_k = \mathbf{1}\mathbf{1}'/n + \boldsymbol{S}_i^* + (\boldsymbol{I} - \boldsymbol{S}_i^*)\boldsymbol{H}_{-i}^*, \quad i = 1, \dots, k, \tag{12}$$

where $\boldsymbol{S}_i^*$ is the centred smoother matrix of a model with a single smooth term and $\boldsymbol{H}_{-i}^*$ is the centered hat-matrix of a weighted additive model with $k-1$ smooth terms (see Appendix A in (21) and (22) for precise definitions). Any fitted curve is estimated by

$$\hat{f}_i(\boldsymbol{x}_i) = \boldsymbol{S}_i^*(\boldsymbol{I} - \boldsymbol{H}_{-i}^*)\boldsymbol{y}. \tag{13}$$

In Appendix A we minimise the weighted version of (10) and hence obtain the corresponding expressions to (12) and (13) when $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{V})$. Semiparametric additive models are a particular case of additive models where, for example, $f_1(\boldsymbol{x}_1) = \boldsymbol{X}\boldsymbol{\beta}$, $\boldsymbol{B}_1 = \boldsymbol{X}$ and $\boldsymbol{P}_1 = \mathbf{0}$.

3 Mixed model representation of P-splines

Verbyla, Cullis, Kenward & Welham (1999) showed that a fitted cubic smoothing spline could be written as an estimated straight line plus a predicted random effect. We follow the same methodology, but applied to P-splines with a B-spline basis. It is worth noticing that a cubic smoothing spline is a P-spline with $ndx = n$, $bdeg = 3$ and $pord = 2$ and so the methodology we present here extends previous work. Brumback, Ruppert & Wand (1999) and Coull, Ruppert & Wand (2001) among others have used other low-rank spline smoothers (based on truncated lines basis) in the mixed model context; however, the B-spline basis used by Marx & Eilers (1998) has better numerical properties (Aerts et al. 2002).
In the additive model (9) with general error structure, we assume that $\boldsymbol{a}^* = (\boldsymbol{a}_1', \dots, \boldsymbol{a}_k')' = \boldsymbol{G}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{u}$, and consider the mixed model

$$\boldsymbol{y} = \boldsymbol{B}\tilde{\boldsymbol{G}}\boldsymbol{\beta} + \boldsymbol{B}^*\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\epsilon}, \quad \boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_u^2), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{V}) \tag{14}$$

with $\tilde{\boldsymbol{G}}$=blockdiag$(1, \boldsymbol{G})$, $\boldsymbol{\beta} = (\alpha, \boldsymbol{b}')'$, $\boldsymbol{B}^* = (\boldsymbol{B}_1^* : \dots : \boldsymbol{B}_k^*)$, $\boldsymbol{u} = (\boldsymbol{u}_1', \dots, \boldsymbol{u}_k')'$, $\boldsymbol{\sigma}_u^2 = \text{blockdiag}(\sigma_{u_1}^2 \boldsymbol{I}_{r_1}, \dots, \sigma_{u_k}^2 \boldsymbol{I}_{r_k})$, and $\boldsymbol{u}$ and $\boldsymbol{\epsilon}$ are independent.

We need to choose $\boldsymbol{G}$ and $\boldsymbol{Z}$ such that (a) $[\boldsymbol{G} : \boldsymbol{Z}]$ is square and of full rank, and (b) (14) results in a set of mixed model equations (see for example Searle, Casella & McCulloch 1992) with solution (11). We give in Appendix B the definition of r_i and show how to construct $\boldsymbol{G}$ and $\boldsymbol{Z}$ and thus decompose $\boldsymbol{Ba}$ into fixed polynomial trends of degree $q_i - 1$, $i = 1, \ldots, k$ and k random components, and prove that the estimate of $\boldsymbol{a}$ by penalised least squares is equivalent to the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ from the mixed model (14).
Each smoothing parameter λ_i is given by the variance ratio $\sigma^2/\sigma^2_{u_i}$. Thus the estimation of the smoothing parameters is equivalent to variance component estimation in a mixed model, and the smoothing parameters can be chosen by maximising the residual log-likelihood

$$\begin{aligned}\ell(\sigma^2, \boldsymbol{\lambda}) &= -\tfrac{1}{2}\log|\boldsymbol{\Sigma}| - \tfrac{1}{2}\log|\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}| \\ &\quad -\tfrac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{\Sigma}^{-1})\boldsymbol{y}, \qquad (15)\end{aligned}$$

with $\tilde{\boldsymbol{X}} = \boldsymbol{B}\tilde{\boldsymbol{G}}$ and $\boldsymbol{\Sigma} = \sigma^2(\boldsymbol{V} + \boldsymbol{B}^*\boldsymbol{Z}\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}'\boldsymbol{B}^{*\prime})$, where $\boldsymbol{\Lambda}^{-1} = \boldsymbol{\sigma}^2_u/\sigma^2$.
In the case of correlated errors we take $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\rho})$ where $\boldsymbol{\rho}$ is a vector of unknown parameters and then maximise $\ell(\sigma^2, \boldsymbol{\lambda}, \boldsymbol{\rho})$.

4 Applications

In this section we give an illustrative example of the results presented in sections 2 and 3. Examples of P-splines with correlated or heteroscedastic errors with a single smooth term can be found in Currie & Durban (2002), and Coull, Schwartz & Wand (2001) fitted an additive model with an AR(1) error structure. In that paper, the authors propose the use of standard statistical software (the SAS procedure PROC MIXED or the S-PLUS function `lme()`) to fit additive models with correlated errors. However, the implementation of additive mixed models is not yet straightforward, and the calculation of standard errors and confidence intervals is not clear (as in the case of the `gam()` function in S-PLUS). Our results are based on the results of sections 2 and 3 and on computational methods for linear mixed-effects models presented by Pinheiro & Bates (2000).
We analysed the residential electricity data of Smith et al. (1998). The data consist of 264 observations of monthly electricity consumption ($\boldsymbol{y}$) and four independent variables: the number of heating degree days ($\boldsymbol{x}_1$), the number of cooling degree days ($\boldsymbol{x}_2$), average real electricity price ($\boldsymbol{x}_3$) and real disposable income ($\boldsymbol{x}_4$) (detailed description of the data can be found in Harris & Liu 1993). We also include month ($\boldsymbol{x}_5$) as an additional covariate. A `log` transformation of the data is used to obtain normality of residuals and the response was modelled as

$$\log(\boldsymbol{y}) = \boldsymbol{1}\alpha + f_1(\boldsymbol{x}_1) + f_2(\boldsymbol{x}_2) + f_3(\boldsymbol{x}_3) + f_4(\boldsymbol{x}_4) + f_5(\boldsymbol{x}_5) + \boldsymbol{\epsilon}, \qquad (16)$$

and $\epsilon \sim \mathcal{N}(0, \sigma^2 V)$.
We used the mixed model representation of P-splines to model the curves, and assumed that the errors follow an auto-regressive process. The smoothing

Error Variance Model	Number of Variance parameters	-2(log-lik)
I	6	-1143.93
AR(1)	7	-1240.46
AR(2)	8	-1248.28
AR(12)	18	-1375.89
AR(13)	19	-1379.77

Table 1: Log-likelihood values for models with independent, AR(1), AR(2), AR(12) and AR(13) errors

parameters and correlation parameters, together with the order of the autoregressive process, were estimated by minimisation of (15). Table 1 gives log-likelihood values for five of the fitted models; the model selected was an AR(13) compared with the order 12 autoregressive model fitted by Smith et al. (1998). Figure 2(e) shows that the autocorrelation function of the whitened residuals (defined as: $\tilde{e} = \hat{V}^{-1/2}\hat{e}$) seems consistent with white noise and Figure 2(f) shows that the assumption of normality is satisfactory. The estimated intercept $\hat{\alpha} = 0.38$. Figure 1 and Figure 2(a)-(d) show fitted curves, partial residuals and pointwise confidence intervals for all covariates;

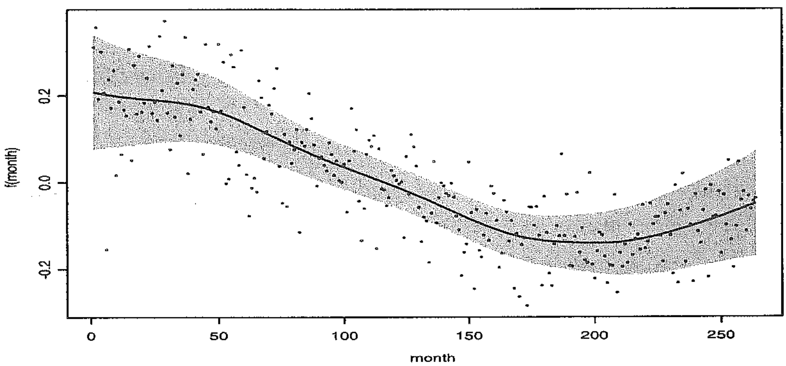

Figure 1: Estimated trend for month number together with partial residuals and 95% confidence intervals.

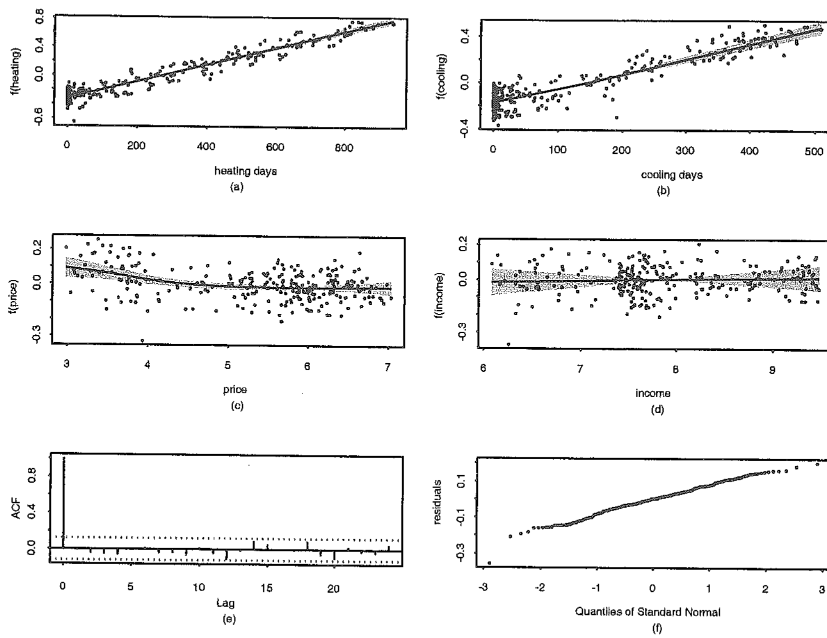

Figure 2: (a)-(d) Plot of fitted curves for heating days, cooling days, price and income, together with partial residuals and 95% pointwise confidence intervals, (e) autocorrelations of $\tilde{e}$, (f) normal probability plot of $\tilde{e}$

the fitted functions and confidence intervals were estimated using (13). Our plots are quite similar to those in Smith et al. (1998) although the function estimates for electricity price and disposable income in their paper are less smooth than our results; we prefer the smoother plots in our Figure 2. Our plots suggest that the log of electricity consumption is linear in heating days, cooling days and real disposable income, and nonlinear in time and price. A further comparison between the two approaches concerns identifiability: in Smith et al. (1998) the estimated functions are identifiable up to a constant (as is shown in their figure 6), and so the scale of their plots is misleading; with our approach we estimate centred fitted curves and so avoid identifiability problems.

5 Closing remarks

In this paper we have given closed form expressions for the components of a P-spline additive model with a general error structure. These closed forms enabled us to calculate confidence intervals for each individual curve, and the use of the mixed model representation of a P-spline gave an efficient method for both the estimation of the smoothing and correlation parameters, and the selection of the order of the autoregression process.

Appendix A

We derive expression (12) for the hat-matrix in the additive model (9) with k terms. First we define $\boldsymbol{B}^*_{-i} = (\boldsymbol{B}^*_1 : \ldots : \boldsymbol{B}^*_{i-1} : \boldsymbol{B}^*_{i+1} : \ldots : \boldsymbol{B}^*_k)$ and $\boldsymbol{a}_{-i} = (\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_{i-1}, \boldsymbol{a}'_{i+1}, \ldots, \boldsymbol{a}'_k)'$. We rewrite (9) as

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{B}\boldsymbol{a} \text{ where } \boldsymbol{B} = (\boldsymbol{1} : \boldsymbol{B}^*_i : \boldsymbol{B}^*_{-i}), \quad \boldsymbol{a} = (\alpha, \boldsymbol{a}'_i, \boldsymbol{a}'_{-i})',$$

and obtain $\hat{\boldsymbol{a}}$ by minimising

$$S = (\boldsymbol{y}-\alpha\boldsymbol{1}-\boldsymbol{B}^*_i\boldsymbol{a}_i-\boldsymbol{B}^*_{-i}\boldsymbol{a}_{-i})'(\boldsymbol{y}-\alpha\boldsymbol{1}-\boldsymbol{B}^*_i\boldsymbol{a}_i-\boldsymbol{B}^*_{-i}\boldsymbol{a}_{-i})+\boldsymbol{a}'_i\boldsymbol{P}_i\boldsymbol{a}_i+\boldsymbol{a}'_{-i}\boldsymbol{P}_{-i}\boldsymbol{a}_{-i}.$$

Taking derivatives with respect to the components of $\boldsymbol{a}$ we obtain

$$\frac{\partial S}{\partial \alpha} = \boldsymbol{1}'(\boldsymbol{y} - \hat{\alpha}\boldsymbol{1} - \boldsymbol{B}^*_i\hat{\boldsymbol{a}}_i - \boldsymbol{B}^*_{-i}\hat{\boldsymbol{a}}_{-i}) = \boldsymbol{1}'\boldsymbol{y} - \boldsymbol{1}'\boldsymbol{1}\hat{\alpha} = 0 \tag{17}$$

$$\frac{\partial S}{\partial \boldsymbol{a}_i} = -\boldsymbol{B}^{*\prime}_i(\boldsymbol{y} - \hat{\alpha}\boldsymbol{1} - \boldsymbol{B}^*_i\hat{\boldsymbol{a}}_i - \boldsymbol{B}^*_{-i}\hat{\boldsymbol{a}}_{-i}) + \boldsymbol{P}_i\hat{\boldsymbol{a}}_i = \boldsymbol{0} \tag{18}$$

$$\frac{\partial S}{\partial \boldsymbol{a}_{-i}} = -\boldsymbol{B}^{*\prime}_{-i}(\boldsymbol{y} - \hat{\alpha}\boldsymbol{1} - \boldsymbol{B}^*_i\hat{\boldsymbol{a}}_i - \boldsymbol{B}^*_{-i}\hat{\boldsymbol{a}}_{-i}) + \boldsymbol{P}_{-i}\hat{\boldsymbol{a}}_{-i} = \boldsymbol{0}. \tag{19}$$

Equation (17) yields $\hat{\alpha} = (\boldsymbol{1}'\boldsymbol{1})^{-1}\boldsymbol{1}'\boldsymbol{y} = \overline{\boldsymbol{y}}$. From (18) we obtain

$$(\boldsymbol{B}^{*\prime}_i\boldsymbol{B}^*_i + \boldsymbol{P}_i)\hat{\boldsymbol{a}}_i = \boldsymbol{B}^{*\prime}_i(\boldsymbol{y} - \boldsymbol{B}^*_{-i}\hat{\boldsymbol{a}}_{-i})$$

and so

$$\boldsymbol{B}^*_i\hat{\boldsymbol{a}}_i = \boldsymbol{S}^*_i(\boldsymbol{y} - \boldsymbol{B}^*_{-i}\hat{\boldsymbol{a}}_{-i}) \tag{20}$$

where

$$\boldsymbol{S}^*_i = \boldsymbol{B}^*_i(\boldsymbol{B}^{*\prime}_i\boldsymbol{B}^*_i + \boldsymbol{P}_i)^-\boldsymbol{B}^{*\prime}_i \tag{21}$$

is the centred smoother matrix from the model with the single smooth term $\boldsymbol{B}^*_i$. Here $\boldsymbol{A}^-$ represents a generalised inverse of $\boldsymbol{A}$ (although the generalised inverse is not unique, $\boldsymbol{S}^*_i$ is invariant to the choice of the generalised inverse; (see for example Harville 1999, chap. 9). Substitution of (20) in (19) yields

$$(\boldsymbol{B}^{*\prime}_{-i}(\boldsymbol{I} - \boldsymbol{S}^*_i)\boldsymbol{B}^*_{-i} + \boldsymbol{P}_{-i})\hat{\boldsymbol{a}}_{-i} = \boldsymbol{B}^{*\prime}_{-i}(\boldsymbol{I} - \boldsymbol{S}^*_i)\boldsymbol{y}$$

from which we find

$$B^*_{-i}\hat{a}_{-i} = H^*_{-i}y$$

where

$$H^*_{-i} = B^*_{-i}(B^{*\prime}_{-i}(I - S^*_i)B^*_{-i} + P_{-i})^{-}B^{*\prime}_{-i}(I - S^*_i) \qquad (22)$$

and H^*_{-i} is the centred hat-matrix of a weighted additive model (with weights $(I - S^*_i)$). Thus, $\hat{y} = \mathbf{1}\hat{\alpha} + B^*_i\hat{a}_i + B^*_{-i}\hat{a}_{-i} = H_k y$ where

$$H_k = \mathbf{1}\mathbf{1}'/n + S^*_i + (I - S^*_i)H^*_{-i} \qquad (23)$$

is the hat-matrix for an additive model with k terms, as required.

When $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 V)$, it is straightforward to show that (13) becomes

$$\hat{f}_i(x_i) = S^*_{iV}(I - H^*_{-iV})y$$

where

$$\begin{aligned} S^*_{iV} &= B^*_i\left(B^{*\prime}_i V^{-1} B^*_i + P_i\right)^{-} B^{*\prime}_i V^{-1} \quad \text{as in (21)} \\ H^*_{-iV} &= B^*_{-i}(B^{*\prime}_{-i}V^{-1}(I - S^*_{iV})B^*_{-i} + P_{-i})^{-}B^{*\prime}_{-i}V^{-1}(I - S^*_{iV}). \end{aligned}$$

Appendix B

The additive P-spline setup chooses a for given λ to minimise

$$S(a) = (y - Ba)'V^{-1}(y - Ba) + a'Pa \qquad (24)$$

with $B = (\mathbf{1} : B^*_1 : \ldots : B^*_k)$ and $P = \text{blockdiag}(0, P_1, \ldots, P_k)$ where $P_j = \lambda_j D'_j D_j$. The value of a that minimises (24) satisfies

$$(B'V^{-1}B + P)a = B'V^{-1}y. \qquad (25)$$

We show that (25) is the result of estimation in a mixed model. We write $Ba = B\tilde{G}\beta + B^*Zu$ with $\tilde{G}$ =blockdiag$(1, G)$, $B^* = (B^*_1 : \ldots : B^*_k)$ and G and Z defined as follows: $G = \text{blockdiag}(1, G_1, \ldots, G_k)$ where $G_i = (g_i, g_i^2, \ldots, g_i^{q_i-1})$, q_i is the order of penalty for the ith regressor x_i and $g'_i = (1, 2, \ldots, p_i)$ where p_i is the rank of B_i; $Z = \text{blockdiag}(Z_1, \ldots, Z_k)$ with $Z_i = D'_i(D_i D'_i)^{-1}$. Substituting $Ba = B\tilde{G}\beta + B^*Zu$ in (25) and using $D_i G_i = \mathbf{0}$ we find

$$B'V^{-1}B\tilde{G}\beta + B'V^{-1}B^*Zu + P^*Zu = B'V^{-1}y \qquad (26)$$

with P^* =blockdiag$(P_1, \ldots, P_k)$. Multiplying (26) by $\tilde{G}'$ gives

$$\tilde{G}'B'V^{-1}B\tilde{G}\beta + \tilde{G}'B'V^{-1}B^*Zu = \tilde{G}'B'V^{-1}y \qquad (27)$$

(again using $D_i G_i = 0$) while multiplying (26) by blockdiag$(0, Z')$ gives

$$Z'B^{*\prime}V^{-1}B\tilde{G}\beta + Z'B^{*\prime}V^{-1}B^*Zu + Z'P^*Zu = Z'B^{*\prime}V^{-1}y. \quad (28)$$

Let $r_i = ndx_i + bdeg_i - pord_i$ be the number of columns of D_i, the difference matrix for the ith regressor x_i. Then $Z'P^*Z = \text{blockdiag}(\lambda_1 I_{r_1}, \ldots, \lambda_k I_{r_k}) = \Lambda$. Then (27) and (28) can be written

$$\begin{bmatrix} \tilde{G}'B'V^{-1}B\tilde{G} & \tilde{G}'B'V^{-1}B^*Z \\ Z'B^{*\prime}V^{-1}B\tilde{G} & Z'B^{*\prime}V^{-1}B^*Z + \Lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \tilde{G}'B' \\ Z'B^{*\prime} \end{bmatrix} V^{-1}y.$$

Thus $\hat{\beta}$ and $\hat{u}$ are estimates that arise from the mixed model

$$y = B\tilde{G}b + B^*Zu + \epsilon$$

where $u \sim \mathcal{N}(0, \sigma_u^2)$, $\epsilon \sim \mathcal{N}(0, \sigma^2 V)$ and $\sigma_u^2 = \sigma^2\Lambda^{-1}$.

Acknowledgements

The authors would like to thank Mike Smith for supplying the electicity data. The work of Maria Durbán was supported by European Commission project HPCF CT - 2000 - 00041 and by DGES project BEC 2001-1270.

References

Aerts, M., Claeskens, G. & Wand, M. P. (2002), 'Some theory for penalized spline generalized linear models', *J. Stat. Plan. Infer.* **103**, 455–470.

Altman, N. (1990), 'Kernel smoothing of data with correlated errors', *J. Am. Stat. Assoc.* **85**, 749–759.

Brumback, B., Ruppert, D. & Wand, M. (1999), 'Comment on "Variable selection and function estimation in additive nonparametric regression using a data-based prior"', *J. Am. Stat. Assoc.* **94**, 794–797.

Buja, A., Hastie, T. & Tibshirani, R. (1989), 'Linear smoothers and additive models (with discussion)', *Ann. Stat.* **17**, 453–555.

Coull, A., Ruppert, D. & Wand, M. (2001), 'Simple incorporation of interactions into additive models', *Biometrics* **57**, 539–545.

Coull, B., Schwartz, J. & Wand, M. (2001), 'Respiratory health and air pollution: Additive mixed model analyses', *Biostatistics* **2**, 337–349.

Currie, I. D. & Durban, M. (2002), 'Flexible smoothing with P-splines: A unified approach', *Statistical Modelling* **2**, 333–349.

Durban, M. (1999), Modelling Spatial Trend and Local Competition Effects Using Semiparametric Additive Models, PhD thesis, Dept. Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, Scotland, U.K.

Durban, M., Hackett, C. & Currie, I. (1999), 'Approximate standard errors in semi-parametric additive models', *Biometrics* **55**, 699–703.

Eilers, P. & Marx, B. (1996), 'Flexible smoothing with B-splines and penalties', *Stat. Sci.* **11**, 89–121.

Harris, J. & Liu, L. (1993), 'Dynamic structural analysis and forecasting of residential electricity consumption', *Int. J. Forec.* **2**, 437–455.

Hart, J. (1991), 'Kernel regression estimation with time series errors', *J. Roy. Stat. Soc. B* **53**, 173–187.

Harville, D. A. (1999), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag.

Hastie, T. & Tibshirani, R. (1987), 'Generalized additive models: Some applications', *J. Am. Stat. Assoc.* **82**, 371–386.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

Marx, B. & Eilers, P. (1998), 'Direct generalized additive modeling with penalized likelihood', *Comput. Stat. Data An.* **28**, 193–209.

Opsomer, J. (2000), 'Asymptotic properties of backfitting estimators', *J. Multivariate Anal.* **73**, 166–179.

Opsomer, J. & Ruppert, D. (1999), 'A root-n consistent backfitting estimator for semiparametric additive modelling', *J. Comput. Graph. Stat.* **8**, 715–732.

Pinheiro, J. & Bates, D. (2000), *Mixed-Effects Models in S and S-PLUS*, Statistics and Computing, Springer.

Searle, S., Casella, G. & McCulloch, C. (1992), *Variance components*, Wiley Series in Probability and Mathematical Statistics.

Smith, M., Wong, C. & Kohn, R. (1998), 'Additive nonparametric regression with autocorrelated errors', *J. Roy. Stat. Soc. B* **60**, 311–331.

Verbyla, A., Cullis, B., Kenward, M. & Welham, S. (1999), 'The analysis of designed experiments and longitudinal data using smoothing splines', *J. Roy. Stat. Soc. C* **48**, 269–312.

Wang, Y. (1998), 'Smoothing spline models with correlated errors', *J. Am. Stat. Assoc.* **93**, 341–348.