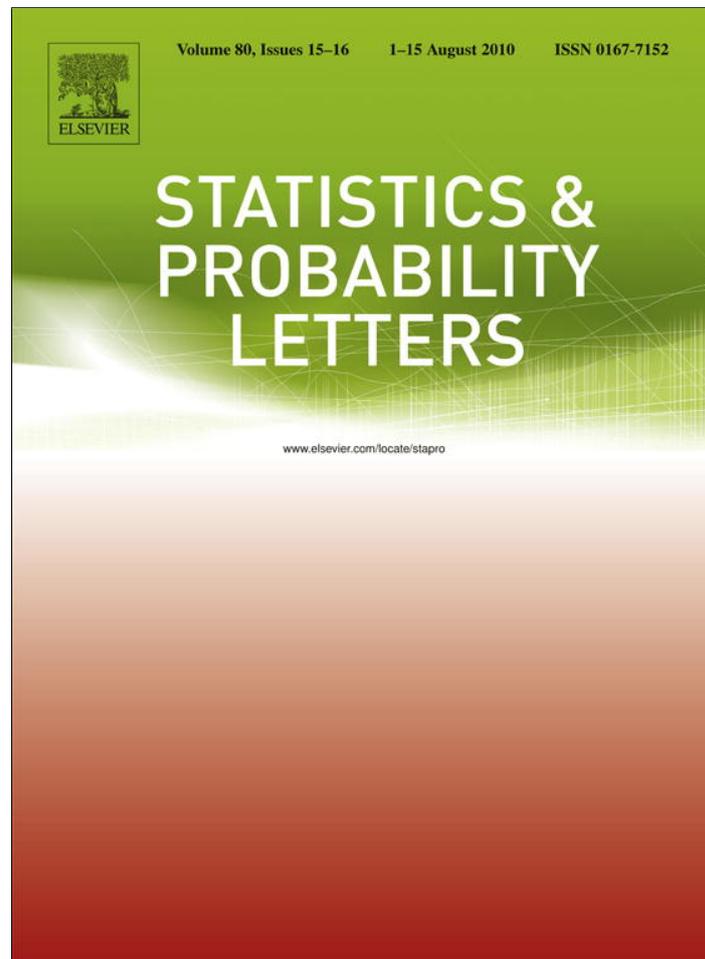


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Marginal longitudinal semiparametric regression via penalized splines

M. Al Kadiri^a, R.J. Carroll^b, M.P. Wand^{a,*}^a Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, New South Wales, Australia^b Department of Statistics, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Article history:

Received 16 July 2009

Received in revised form 10 December 2009

Accepted 11 April 2010

Available online 22 April 2010

Keywords:

Additive models

Best prediction

Maximum likelihood

Gibbs sampling

Nonparametric regression

Restricted maximum likelihood

Varying coefficient models

ABSTRACT

We study the marginal longitudinal nonparametric regression problem and some of its semiparametric extensions. We point out that, while several elaborate proposals for efficient estimation have been proposed, a relative simple and straightforward one, based on penalized splines, has not. After describing our approach, we then explain how Gibbs sampling and the BUGS software can be used to achieve quick and effective implementation. Illustrations are provided for nonparametric regression and additive models.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The past decade has seen a great deal of interest and activity in nonparametric regression for longitudinal data. A prominent component of this research is the *marginal longitudinal nonparametric regression* problem in which the covariance matrix of the responses for each subject is not modelled conditionally, and instead is an unspecified parameter to be estimated.

Ruppert et al. (2009, Section 3.9) provide a summary of research on this problem up until about 2008. Whilst Zeger and Diggle (1994) is an early reference for marginal longitudinal nonparametric regression, the area started to heat up in response to Lin and Carroll (2001), where it was shown that ordinary kernel smoothers are more efficient if so-called working independence is assumed. This spawned a flurry of activity on the problem. Relevant references include: Lin and Carroll (2000, 2006), Welsh et al. (2002), Wang (2003), Linton et al. (2003), Lin et al. (2004), Carroll et al. (2004), Hu et al. (2004), Chen and Jin (2005), Wang et al. (2005) and Fan et al. (2007), Sun et al. (2007), Fan and Wu (2008) and Carroll et al. (2009a,b).

In this article, we describe a relatively simple approach to the marginal longitudinal regression problem and its semiparametric extensions. Our approach is the natural one arising from the mixed model representation of penalized splines (e.g. Brumback et al., 1999; Ruppert et al., 2003) with estimation and inference done using maximum likelihood and best prediction. There is also the option of adopting a Bayesian standpoint and calling upon Markov chain Monte Carlo to achieve approximate inference. An interesting aspect of our marginal longitudinal semiparametric regression models is that Gibbs sampling applies with draws from standard distributions. For the Bayesian version of our model, the BUGS inference engine (Lunn et al., 2000) can be used for fitting, and we provide some illustrative code.

The penalized spline/mixed model approach means that semiparametric extensions of the marginal longitudinal regression problem can be handled straightforwardly. We describe extensions to additive and varying coefficient models, although other extensions can be handled similarly.

* Corresponding author.

E-mail address: mwand@uow.edu.au (M.P. Wand).

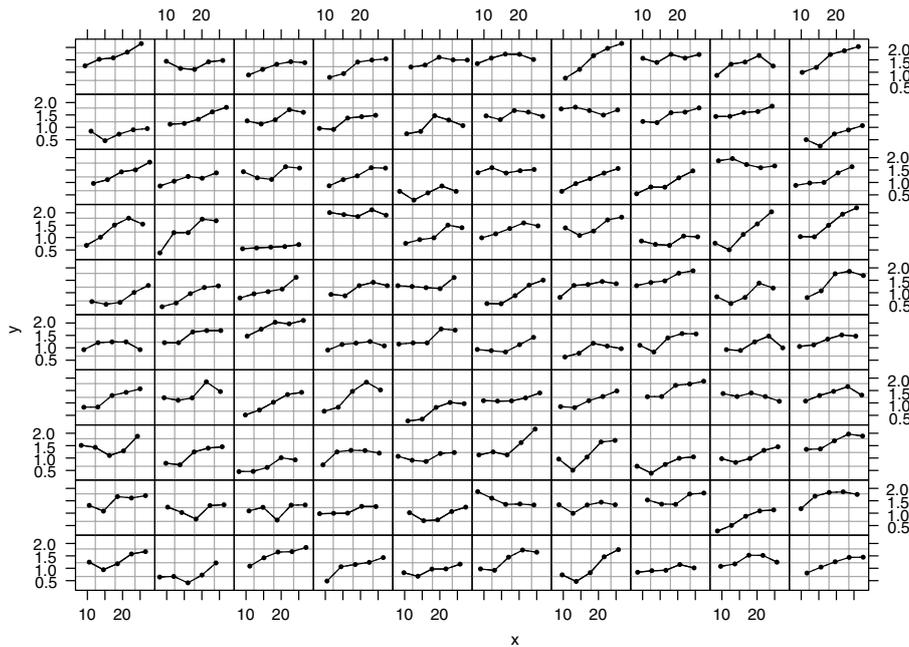


Fig. 1. A data set simulated from a version of the marginal longitudinal nonparametric regression model (1) with $m = 100, n = 5$ and f and Σ as described in the text.

Section 2 describes the penalized spline approach and identifies the mixed model structures required to handle marginal longitudinal semiparametric regression problems. In Section 3, we discuss fitting via maximum likelihood and best prediction. Section 4 describes Bayesian inference via Gibbs sampling and BUGS. Illustrations are provided in Section 5 and closing discussion is given in Section 6.

2. Marginal longitudinal nonparametric regression and extensions

For $1 \leq i \leq m$ subjects, we observe $1 \leq j \leq n$ ($n \ll m$) scalar responses y_{ij} and predictors x_{ij} . Let \mathbf{y}_i be the vector of responses for the i th subject and \mathbf{x}_i be defined similarly. The covariance matrix of a random vector \mathbf{v} is denoted by $\text{Cov}(\mathbf{v})$. The marginal longitudinal nonparametric regression model is then

$$E(y_{ij}) = f(x_{ij}), \quad \text{Cov}\{y_i | f(\mathbf{x}_i)\} = \Sigma, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \tag{1}$$

for some real-valued smooth function f and $n \times n$ covariance matrix Σ . The notation $f(\mathbf{x}_i)$ means that the function f is applied element-wise to each of the entries of \mathbf{x}_i . We use $\text{Cov}\{y_i | f(\mathbf{x}_i)\}$ rather than $\text{Cov}(y_i)$ to allow for the possibility that $f(\mathbf{x}_i)$ is random according to the model, although this is not a requirement.

Fig. 1 shows a simulated data set for model (1), with $m = 100, n = 5$,

$$f(x) = 1 + \frac{1}{2} \Phi((2x - 36)/5) \quad \text{and} \quad \Sigma = \begin{bmatrix} 0.122 & 0.098 & 0.078 & 0.063 & 0.050 \\ 0.098 & 0.122 & 0.098 & 0.078 & 0.063 \\ 0.078 & 0.098 & 0.122 & 0.098 & 0.078 \\ 0.063 & 0.078 & 0.098 & 0.122 & 0.098 \\ 0.050 & 0.063 & 0.078 & 0.098 & 0.122 \end{bmatrix}, \tag{2}$$

where Φ is the standard normal distribution function. The main problem is efficient estimation of f from data such as that shown in Fig. 1. Estimation of Σ may also be of interest.

Our approach to function estimation involves spline models for f of the form

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^K u_k z_k(x) \tag{3}$$

where z_1, \dots, z_K is a rich set of spline basis functions. A simple basis arises from setting $z_k(x) = (x - \kappa_k)_+$ where $\kappa_1, \dots, \kappa_K$ is a dense set of knots placed over the range of the x_i s and $r_+ = \max(0, r)$ for any real number r . However, we recommend a smoother and more numerically stable choice for z_k , such as those described in Welham et al. (2007) and Wand and Ormerod (2008). The number of basis functions K has a minor effect on the efficacy of (3) and, for most signals arising in practice, $K = 25$ is sufficient. Li and Ruppert (2008) give some interesting asymptotics that provide support for this maxim.

To avoid over-fitting the spline coefficients $u_k, 1 \leq k \leq K$, need to be penalized in some way. A convenient penalization mechanism is to treat the u_k as a random sample from a distribution with mean zero and variance σ^2 . This permits the

following linear mixed model representation of (1) and (3):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{4}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_1 \\ \vdots & \vdots \\ \mathbf{1} & \mathbf{x}_m \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} z_1(\mathbf{x}_1) & \cdots & z_K(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ z_1(\mathbf{x}_m) & \cdots & z_K(\mathbf{x}_m) \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix}.$$

The random vectors on the right-hand side of (4) have mean zero and covariance matrix:

$$\text{Cov} \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \otimes \boldsymbol{\Sigma} \end{bmatrix},$$

where \otimes denotes Kronecker product. For fixed values of σ^2 and $\boldsymbol{\Sigma}$, we can call upon best linear unbiased prediction (e.g. Robinson, 1991) to estimate $\boldsymbol{\beta}$ and \mathbf{u} and, hence, the regression function f . In practice, though, both σ^2 and $\boldsymbol{\Sigma}$ need to be estimated and a convenient assumption for achieving this aim is

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \otimes \boldsymbol{\Sigma} \end{bmatrix} \right). \tag{5}$$

From now on, we will assume that the Gaussian linear mixed model (4) and (5) is reasonably assumed. Sections 3 and 4 describe two approaches to fitting and inference. Before getting to that, we describe some semiparametric extensions of (1).

2.1. Additive models extension

Suppose now that, corresponding to each y_{ij} , several predictor variables are available. There are a number of semiparametric regression extensions of (1) that could be considered. In this section, we focus on the additive model extension. To keep the notation simple, we restrict discussion to the situation where there are two continuous predictors with the j th measurement on subject i denoted by x_{1ij} and x_{2ij} . The marginal longitudinal additive model for such data is

$$E(y_{ij}) = \beta_0 + f_1(x_{1ij}) + f_2(x_{2ij}), \quad \text{Cov}\{y_i | f_1(\mathbf{x}_{1i}), f_2(\mathbf{x}_{2i})\} = \boldsymbol{\Sigma}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n \tag{6}$$

where f_1 and f_2 are smooth functions. If each of these is modelled as a penalized spline:

$$f_1(x_1) = \beta_{11}x_1 + \sum_{k=1}^{K_1} u_{1k}z_{1k}(x_1) \quad \text{and} \quad f_2(x_2) = \beta_{21}x_2 + \sum_{k=1}^{K_2} u_{2k}z_{2k}(x_2) \tag{7}$$

with coefficients independently distributed subject to

$$u_{1k} \text{ i.i.d. } N(0, \sigma_1^2) \quad \text{and} \quad u_{2k} \text{ i.i.d. } N(0, \sigma_2^2)$$

then a Gaussian linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

arises. The differences between this model and that of Section 2 are that the design matrices are now

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_{11} & \mathbf{x}_{21} \\ \vdots & \vdots & \vdots \\ \mathbf{1} & \mathbf{x}_{1m} & \mathbf{x}_{2m} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} z_{11}(\mathbf{x}_{11}) & \cdots & z_{1K_1}(\mathbf{x}_{11}) & z_{21}(\mathbf{x}_{21}) & \cdots & z_{2K_2}(\mathbf{x}_{21}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_{11}(\mathbf{x}_{1m}) & \cdots & z_{1K_1}(\mathbf{x}_{1m}) & z_{21}(\mathbf{x}_{2m}) & \cdots & z_{2K_2}(\mathbf{x}_{2m}) \end{bmatrix}$$

where \mathbf{x}_{1i} is the $n \times 1$ vector containing the x_{1ij} measurements and \mathbf{x}_{2i} is defined similarly. The coefficient vectors are

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$

where \mathbf{u}_1 is the $K_1 \times 1$ vector containing the u_{1k} and \mathbf{u}_2 is defined similarly. The covariance matrix of the spline coefficients and errors is now

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_m \otimes \boldsymbol{\Sigma} \end{bmatrix} \right). \tag{8}$$

Fitting via maximum likelihood and best prediction is analogous to that described in Section 3. The main difference is that there are two variance parameters σ_1^2 and σ_2^2 (and in extensions to additive models with d smooth functions there will be d such variance components) as well as the error covariance matrix Σ . Maximum likelihood fitting, described in Section 3, requires an expression for $\mathbf{V} \equiv \text{Cov}(\mathbf{y})$. For the current model, this matrix takes the form

$$\mathbf{V} = \mathbf{V}(\sigma_1^2, \sigma_2^2, \Sigma) = \sigma_1^2 \mathbf{Z}_{[1]} \mathbf{Z}_{[1]}^T + \sigma_2^2 \mathbf{Z}_{[2]} \mathbf{Z}_{[2]}^T + \mathbf{I}_m \otimes \Sigma$$

where $\mathbf{Z}_{[1]}$ and $\mathbf{Z}_{[2]}$ correspond to the column-wise partitioning of \mathbf{Z} according to the basis functions for f_1 and f_2 (i.e. $\mathbf{Z} = [\mathbf{Z}_{[1]} \mathbf{Z}_{[2]}]$).

Before closing this section, we briefly mention that the model

$$E(y_{ij}) = \beta_0 + \beta_1 x_{1ij} + f_2(x_{2ij}), \quad \text{Cov}\{\mathbf{y}_i | \mathbf{x}_{1i}, f_2(\mathbf{x}_{2i})\} = \Sigma, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{9}$$

is a simpler type of additive model than (6), since it only has one smooth function component. This is a *bona fide* semiparametric regression model since the right-hand side has the effect of the x_{1ij} s modelled parametrically and the effect of the x_{2ij} s modelled nonparametrically. However, the linear mixed model attached with this model is on par with that treated in Section 2. In particular, the random component structure (5) applies to (9).

2.2. Incorporation of interactions

An important extension of (6) is

$$E(y_{ij}) = \beta_0 + f_{x_{3ij}}(x_{1ij}) + f_2(x_{2ij}), \quad \text{Cov}\{\mathbf{y}_i | f_{x_{3ij}}(\mathbf{x}_{1i}), f_2(\mathbf{x}_{2i})\} = \Sigma, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{10}$$

where the x_{3ij} correspond to measurements on a categorical variable x_3 . The cross-sectional data version of such models are treated, using penalized splines, by Coull et al. (2001) and Ruppert et al. (2003, Chapter 12).

Interestingly, the Gaussian linear mixed model for fitting (10) takes the same form as that for fitting the additive model (6). In particular, the covariance matrix of the random effects and error vectors takes the same as that given in (8), but with two additional variance parameters. The only other difference is the form of the design matrices \mathbf{X} and \mathbf{Z} . The aforementioned references provide details on their construction.

2.3. Varying coefficient models

Another type of multiple-predictor semiparametric regression model is that involving varying coefficients. The simplest marginal longitudinal varying coefficient model is

$$E(y_{ij}) = f_0(s_{ij}) + f_1(s_{ij})x_{ij}, \quad \text{Cov}\{\mathbf{y}_i | f_0(\mathbf{s}_i), f_1(\mathbf{s}_i)\} = \Sigma, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{11}$$

where s_{ij} are the longitudinal measurements on a continuous predictor variable s and x_{ij} are the measurements on a second predictor x . The modifying effect of s on the linear relationship between $E(y)$ and x is modelled flexibly through the varying coefficients $f_0(s)$ and $f_1(s)$. Sun et al. (2007) paid particular attention to models of this type.

As with the interaction extension, the Gaussian linear mixed model for fitting the varying coefficient model (11) takes the same form as that for fitting the additive model (6). In particular, the covariance matrix of the random effects and error vectors is exactly the same as that given in (8). Only the design matrices require modification. Section 12.4 of Ruppert et al. (2003) provides the relevant details.

3. Maximum likelihood estimation and best prediction

Each of the marginal longitudinal semiparametric regression models in the previous section, and their extensions to d smooth functions, can be handled using the Gaussian linear mixed model

$$\mathbf{y} | \mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{I}_m \otimes \Sigma), \quad \mathbf{u} \sim N(\mathbf{0}, \text{blockdiag}(\sigma_\ell^2 \mathbf{I}_{K_\ell}))_{1 \leq \ell \leq d} \tag{12}$$

Here, K_ℓ corresponds to the number of spline basis functions used in the ℓ th smooth function estimate. Let $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_d^2)$ be the vector of variance parameters. Then, the log-likelihood of \mathbf{y} under (12) is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \Sigma) = -\frac{1}{2} \{n \log(2\pi) + \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \tag{13}$$

where

$$\mathbf{V} = \mathbf{V}(\boldsymbol{\sigma}^2, \Sigma) \equiv \text{Cov}(\mathbf{y}) = \sum_{\ell=1}^d \sigma_\ell^2 \mathbf{Z}_{[\ell]} \mathbf{Z}_{[\ell]}^T + \mathbf{I}_m \otimes \Sigma$$

and $[\mathbf{Z}_{[1]} \dots \mathbf{Z}_{[d]}]$ is the partition of \mathbf{Z} corresponding to the basis functions for each smooth function estimate.

For any fixed values of $\boldsymbol{\sigma}^2$ and Σ , the fixed effects solution is

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\sigma}^2, \Sigma) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}. \tag{14}$$

On substitution into (13), we obtain the *profile log-likelihood* for (σ^2, Σ) as:

$$\ell_p(\sigma^2, \Sigma) = -\frac{1}{2} [\log |\mathbf{V}| + \mathbf{y}^T \mathbf{V}^{-1} \{\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}\} \mathbf{y}] - \frac{n}{2} \log(2\pi). \tag{15}$$

However, the *restricted log-likelihood* (Patterson and Thompson, 1971)

$$\ell_R(\sigma^2, \Sigma) = \ell_p(\sigma^2, \Sigma) - \frac{1}{2} \log |\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}| \tag{16}$$

is usually preferred, since it accounts for estimation of the fixed effects vector β . The maximizers of $\ell_R(\sigma^2, \Sigma)$ are often labelled the *restricted maximum likelihood* or *REML* estimates of σ^2 and Σ .

Likelihood-based estimation of the model parameters β, σ^2 and Σ thus involves:

1. Obtain the REML estimates $\hat{\sigma}^2$ and $\hat{\Sigma}$ by maximizing $\ell_R(\sigma^2, \Sigma)$.
2. Obtain the maximum likelihood estimate of $\hat{\beta} = \hat{\beta}(\hat{\sigma}^2, \hat{\Sigma})$ according to (14).

Step 1 is by far the more challenging, since it involves multivariate numerical optimization.

Lastly, there is the problem of estimating spline coefficients \mathbf{u} . Since \mathbf{u} is random, we cannot appeal to maximum likelihood and instead have to rely on best prediction:

$$\tilde{\mathbf{u}}(\sigma^2, \Sigma) \equiv E(\mathbf{u}|\mathbf{y}) = \mathbf{G}_{\sigma^2} \mathbf{Z}^T \mathbf{V}(\sigma^2, \Sigma)^{-1} \{\mathbf{y} - \mathbf{X}\tilde{\beta}(\sigma^2, \Sigma)\}$$

where $\mathbf{G}_{\sigma^2} = \text{blockdiag}_{1 \leq \ell \leq d}(\sigma_\ell^2 \mathbf{I}_{K_\ell})$. An appropriate estimator for \mathbf{u} in this context is the empirical best predictor

$$\hat{\mathbf{u}} = \mathbf{G}_{\hat{\sigma}^2} \mathbf{Z}^T \mathbf{V}(\hat{\sigma}^2, \hat{\Sigma})^{-1} \{\mathbf{y} - \mathbf{X}\tilde{\beta}(\hat{\sigma}^2, \hat{\Sigma})\}.$$

It is straightforward to construct estimates of the regression function f at arbitrary locations $x \in \mathbb{R}$ using $\hat{\beta}$ and $\hat{\mathbf{u}}$.

Despite (12) being a relatively simple linear mixed model, we have not yet been successful in fitting it with standard mixed model software such as `lme()` (Pinheiro et al., 2009) in the R computing language (R Development Core Team, 2009). This led us to also consider the Bayesian inference version and implementation via Gibbs sampling, as the next section describes. Note, however, that the approach described in the current section can be implemented via direct programming using either Newton–Raphson or Expectation–Maximization iteration.

4. Bayesian inference

An alternative inference strategy, which permits more direct implementation in standard software, involves working with a hierarchical Bayesian version of the Gaussian linear mixed model (12). This entails treating β, σ^2 and Σ as random and setting prior distributions for each of them. The most convenient choice, because of conjugacy properties, are priors of the form:

$$\beta \sim N(\mathbf{0}, \mathbf{F}), \quad \sigma_\ell^2 \sim \text{Inverse-Gamma}(A_\ell, B_\ell) \quad \text{and} \quad \Sigma \sim \text{Inverse-Wishart}(a, \mathbf{B}) \tag{17}$$

where $A_\ell, B_\ell, 1 \leq \ell \leq d$, are positive constants and \mathbf{F} and \mathbf{B} both positive definite matrices. Throughout this section, let $[x]$ denote the density function of x . Then, the notation $\sigma^2 \sim \text{Inverse-Gamma}(A, B)$ means that

$$[\sigma^2] = \frac{B^A}{\Gamma(A)} (\sigma^2)^{-A-1} e^{-B/\sigma^2}, \quad \sigma^2, A, B > 0.$$

The notation $\Sigma \sim \text{Inverse-Wishart}(a, \mathbf{B})$, where Σ is $n \times n$, means that

$$[\Sigma] = C_{n,a}^{-1} |\mathbf{B}|^{a/2} |\Sigma|^{-(a+n+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{B}\Sigma^{-1}) \right\}, \quad a > 0, \quad \Sigma, \mathbf{B} \text{ both positive definite}$$

where $C_{n,a} \equiv 2^{an/2} \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma\left(\frac{a+1-i}{2}\right)$.

Bayesian inference is based on the posterior density functions:

$$[\beta|\mathbf{y}], \quad [\mathbf{u}|\mathbf{y}], \quad [\sigma^2|\mathbf{y}] \quad \text{and} \quad [\Sigma|\mathbf{y}]. \tag{18}$$

The probability calculus required to obtain each of these is unwieldy and, in practice, either analytic or Monte Carlo approximations need to be called upon. As shown in Section 4.1, the Markov chain Monte Carlo method *Gibbs sampling* is straightforward to implement for the Bayesian version of (12) and the priors (17) and, upon convergence, yields samples of arbitrary size from the posterior densities (18). The software package BUGS (Lunn et al., 2000) facilitates this approach to approximate Bayesian inference and illustrative code is given in Section 4.2.

A final, albeit important, aspect of this approach to fitting and inference is choice of the hyperparameters $\mathbf{F}, A_\ell, B_\ell, a$ and \mathbf{B} . If the analyst has specific prior beliefs about the model parameters, then there is the opportunity to choose the hyperparameters so that the prior densities reflect those beliefs. More often than not such prior beliefs are absent and vague priors should be used. Reasonable choices for the fixed effects and variance hyperparameters, assuming that the data have been suitably standardized, are:

$$\mathbf{F} = 10^8 \mathbf{I} \quad \text{and} \quad A_\ell = B_\ell = 0.01. \tag{19}$$

Reasonable choices for the hyperparameters associated with Σ are

$$a = n \quad \text{and} \quad \mathbf{B} = 0.01\mathbf{I}_n. \tag{20}$$

We recommend that the sensitivity checks be done on these hyperparameters. An example of such a check, for a different Bayesian penalized spline model, is Figure 3 of Zhao et al. (2006). Section 3.2 of Crainiceanu et al. (2007) has detailed discussion on this issue, including the consequences of failing to pre-standardize the data.

4.1. Gibbs sampling scheme

The hierarchical Bayesian model specified by (12) and (17) can be fitted using a Gibbs sampling scheme with draws from standard distributions. We give the details here.

First, we note (e.g. Robert and Casella, 2004, p. 371) that Gibbs sampling requires successive draws from the full conditional distributions for each member of a particular partition of the parameters in the model. For the present model, we use the partition:

$$\begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix}, \quad \sigma_1^2, \dots, \sigma_d^2, \Sigma.$$

As an example, the full conditional distribution for σ_1^2 is

$$\sigma_1^2 | \mathbf{y}, \begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix}, \quad \sigma_2^2, \dots, \sigma_d^2, \Sigma.$$

We denote this by ' $\sigma_1^2 | \text{rest}$ ' for short. Let $\mathbf{C} \equiv [\mathbf{X} \ \mathbf{Z}]$ and, as before, let $\mathbf{G}_{\sigma^2} \equiv \text{blockdiag}_{1 \leq \ell \leq d}(\sigma_\ell^2 \mathbf{I}_{K_\ell})$. Then, the required full conditionals for Gibbs sampling are:

$$\begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix} | \text{rest} \sim N((\mathbf{C}^T (\mathbf{I}_m \otimes \Sigma^{-1}) \mathbf{C} + \mathbf{G}_{\sigma^2})^{-1} \mathbf{C}^T \Sigma^{-1} \mathbf{y}, (\mathbf{C}^T (\mathbf{I} \otimes \Sigma^{-1}) \mathbf{C} + \mathbf{G}_{\sigma^2})^{-1}),$$

$$\sigma_\ell^2 | \text{rest} \sim \text{Inverse-Gamma} \left(A_\ell + \frac{1}{2} K_\ell, B_\ell + \frac{1}{2} \|\mathbf{u}_\ell\|^2 \right), \quad 1 \leq \ell \leq d$$

and

$$\Sigma | \text{rest} \sim \text{Inverse-Wishart}(a + m, \mathbf{B} + (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})^T).$$

Provided that a is an integer then the Inverse-Wishart draws for Σ can be achieved by setting

$$\Sigma = \left(\sum_{i=1}^{a+m} \mathbf{v}_i \mathbf{v}_i^T \right)^{-1}$$

where the \mathbf{v}_i are independent $N(\mathbf{0}, \{\mathbf{B} + (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})^T\}^{-1})$ random vectors.

Interestingly, the fact that Σ is unstructured means that Gibbs sampling is exact. This is not the case if Σ is structured (e.g. autoregressive) and more complicated Markov chain Monte Carlo schemes are then required.

4.2. Implementation in BUGS

The BUGS language supports implementation of our Bayesian marginal longitudinal semiparametric regression models. It is recommended that the spline basis functions be set up outside of BUGS. We do this in R and then call BUGS using the BRugs package (Ligges et al., 2009). We pass the regression data to BUGS using matrices. For example, the variable yMat is an $m \times n$ matrix with (i, j) entry containing y_{ij} . Our BUGS code for fitting the marginal longitudinal nonparametric regression model is:

```
model
{
  for (i in 1:m)
  {
    for (j in 1:n)
    {
      mu[i,j] <- beta0 + beta1*xMat[i,j] + inprod(u[],Z[(i-1)*n+j,])
    }
    yMat[i,1:n] ~ dmnorm(mu[i,],Omega[1:n,1:n])
  }
  for (k in 1:K)
```

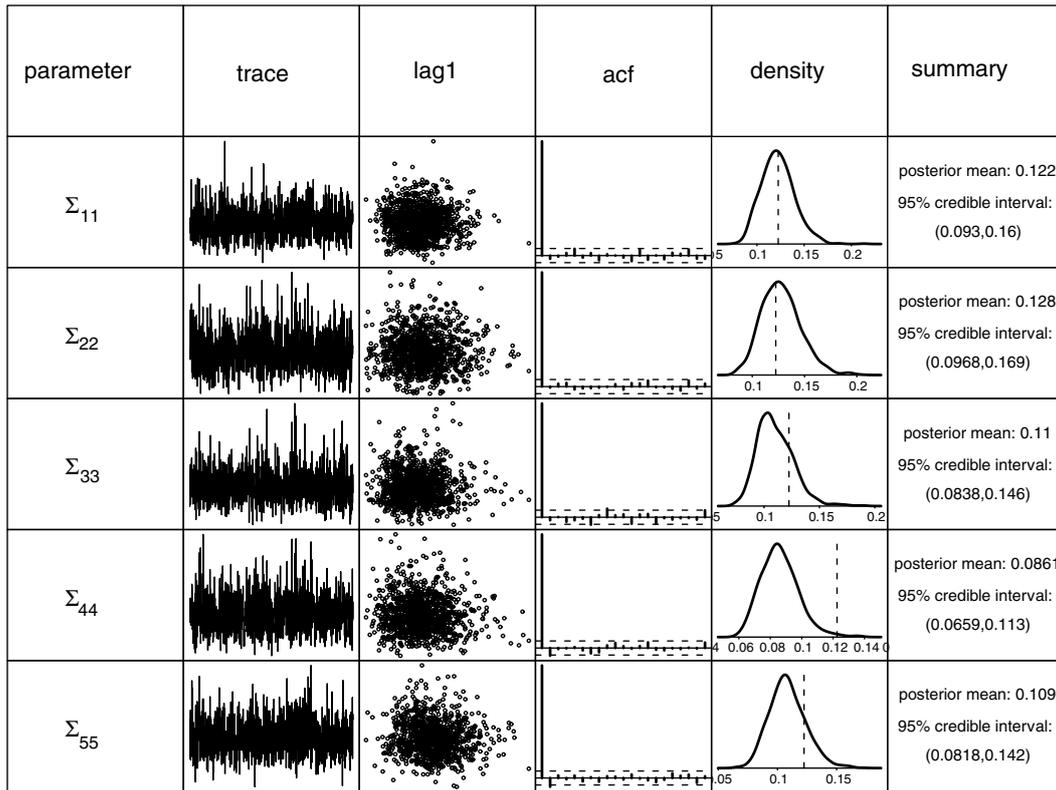


Fig. 2. Summary of MCMC-based inference for the diagonal entries of Σ in the fitted marginal longitudinal nonparametric regression model. The columns are: parameter, trace plot of MCMC sample, plot of sample against 1-lagged sample, sample autocorrelation function, kernel estimates of posterior density and basic numerical summaries. True values of the parameters are shown as vertical dashed lines in the posterior density estimate.

```

{
  u[k] ~ dnorm(0,tau)
}
beta0 ~ dnorm(0,1.0E-8); beta1 ~ dnorm(0,1.0E-8)
tau ~ dgamma(0.01,0.01); Omega[1:n,1:n] ~ dwish(R[,],n)
for (i in 1:n)
{
  for (j in 1:n)
  {
    R[i,j] <- 0.01*equals(i,j)
  }
}
sigma <- 1/sqrt(tau); Sigma[1:n,1:n] <- inverse(Omega[,])
}

```

Note that BUGS uses precision matrices rather than covariance matrices in its multivariate normal distribution specification. Hence, the above code uses the variable Ω , corresponding to $\Omega = \Sigma^{-1}$. Similarly, the precision parameter τ corresponds to $\tau = 1/\sigma^2$ where σ^2 is the spline penalization variance component.

5. Illustrations

We tested out BUGS fitting of the four types of models presented in Section 2 on several sets of simulated data, as well as some empirical data. We now present some of these results.

5.1. Nonparametric regression and simulated data

We fitted the penalized spline model (12) to the data of Fig. 1. The y_{ij} were generated according to (2). The x_{ij} are equally spaced but with the starting positions x_{i1} were generated uniformly from the interval (8, 12). We used the diffuse priors given by (19) and (20). A burn-in period of 5000 was used, followed by 5000 iterations with a thinning factor of 5—resulting in samples of size 1000 being retained for inference. The BUGS program, took 5.5 min to run on the third author’s laptop computer (Mac OS X; 2.33 GHz processor, 3 GBytes of RAM).

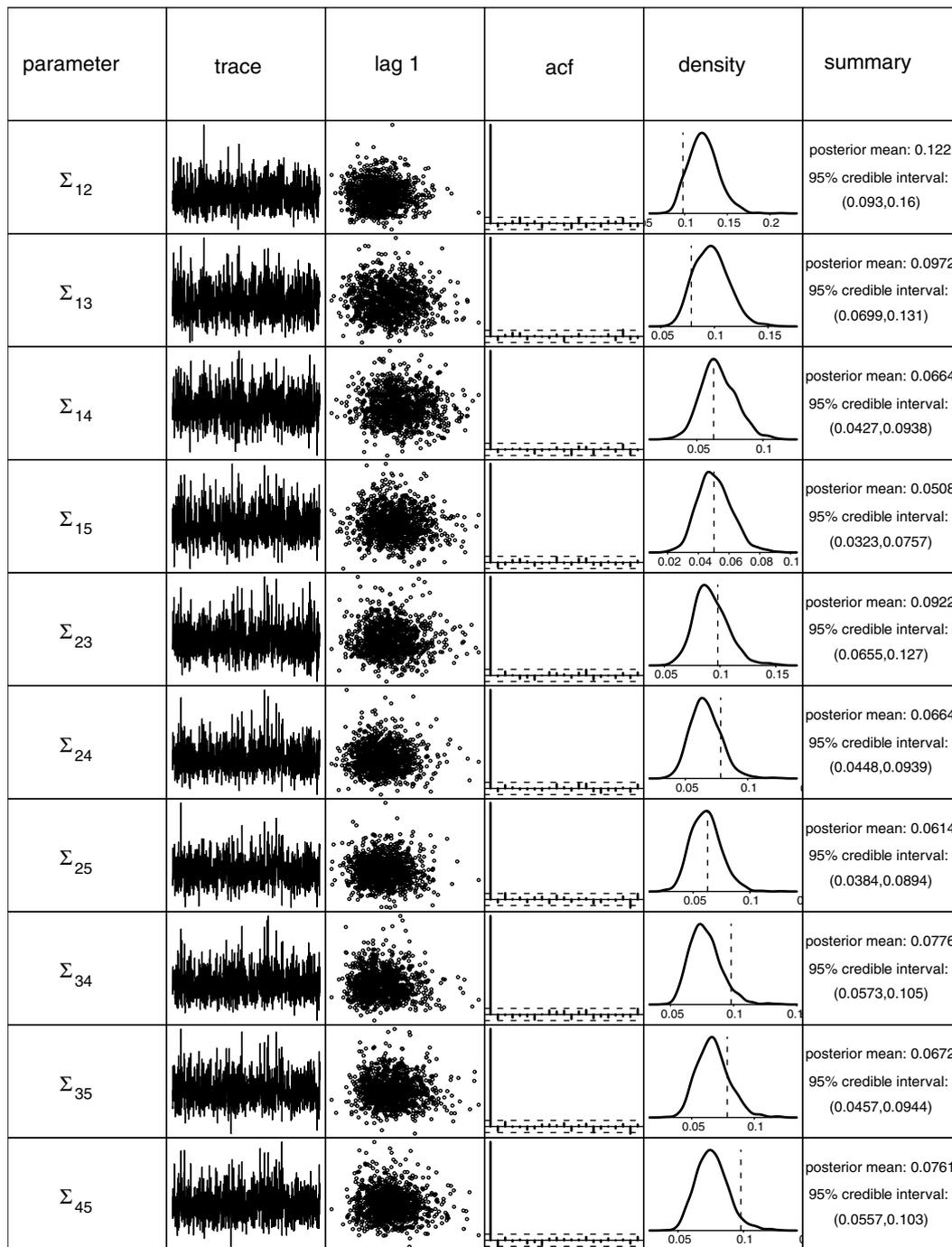


Fig. 3. Summary of MCMC-based inference for the off-diagonal entries of Σ in the fitted marginal longitudinal nonparametric regression model. The columns are: parameter, trace plot of MCMC sample, plot of sample against 1-lagged sample, sample autocorrelation function, kernel estimates of posterior density and basic numerical summaries. True values of the parameters are shown as vertical dashed lines in the posterior density estimate.

Fig. 2 summarizes the BUGS output for the diagonal entries of Σ . The chains mix quite well with no significant autocorrelation. In addition, the true values of Σ_{ii} are captured by the 95% credible intervals in four out of the five cases.

The results for the off-diagonal entries of Σ are summarized in Fig. 3. All ten of the true values of Σ_{ij} are captured by the 95% credible intervals.

The Bayesian penalized spline estimate of f is shown in Fig. 4, with and without the data. The thick solid curves correspond to the posterior means of (3) over a grid of x_s . The dashed curves are corresponding 95% credible sets. The regression function from which the data were generated is shown for comparison. The pointwise posteriors are seen to cover the true f quite well.

The good results presented in this section are typical of the performances we observed over several runs, as well as different choices for f and Σ . An interesting future project would be a large scale simulation study that compares this approach with existing methods.

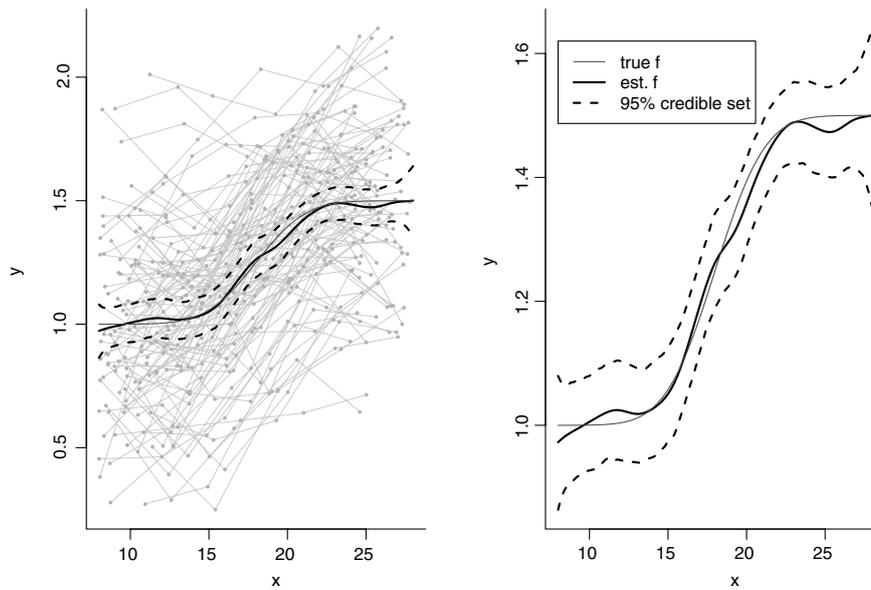


Fig. 4. Left panel: estimated regression function f , together with the longitudinal data on which it is based. Dashed curves correspond to pointwise 95% credible sets. The true f is shown as a thin grey curve. Right panel: estimated regression function f , with data omitted to allow better visualization.

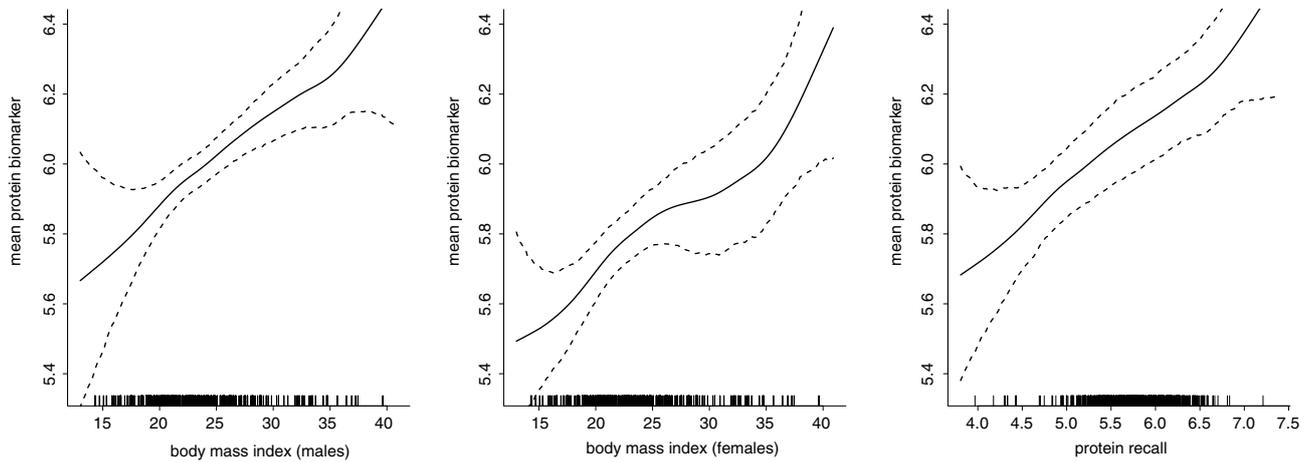


Fig. 5. Estimated functions for the additive/interaction model (10) for the nutritional epidemiology data described in the text. The dashed curves correspond to 95% pointwise credible sets.

5.2. Additive/interaction model and nutritional epidemiology data

We fitted the Bayesian additive/interaction model (10) to data from a nutritional epidemiology study (source: Kipnis et al., 2003). The sample sizes are $m = 294$ and $n = 2$ and each penalized spline fit involved $K = 25$ basis functions. The BUGS program, with 5000 burn-in samples and 5000 retained samples (thinned by a factor of 5), took 52 min to run on the third author's laptop computer. In the notation of (10), the variables are

- y = logarithm of intake of protein as measured by the biomarker urinary nitrogen,
- x_1 = body mass index,
- x_2 = logarithm of intake of protein as measured by a 24-hour recall instrument,
- x_3 = gender.

Fig. 5 shows the estimates of f_{male} , f_{female} and f_2 and corresponding 95% pointwise credible sets.

Fig. 6 summarizes the MCMC-based inference for the entries Σ . Excellent chain mixing is observed. The within-subject covariance is significantly positive, as indicated by the 95% credible set for Σ_{12} .

6. Discussion

It is somewhat of a quirk that the mixed model-based penalized spline approach to marginal longitudinal nonparametric regression has not been explored in depth until now. Nevertheless, as we have illustrated in the previous section, it is a viable approach that is readily implemented in standard software. Another advantage of this approach is that complications

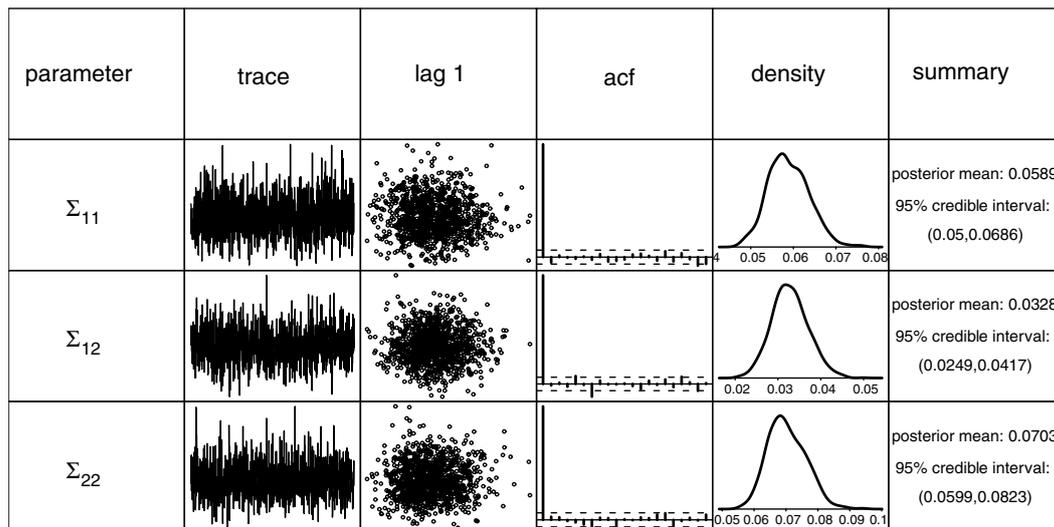


Fig. 6. Summary of MCMC-based inference for the entries of Σ in the fitted additive/interaction model (10). The columns are: parameter, trace plot of MCMC sample, plot of sample against 1-lagged sample, sample autocorrelation function, kernel estimates of posterior density and basic numerical summaries.

such as missingness can be handled within the same likelihood-based or Bayesian frameworks. It would be interesting to see if the asymptotic efficiency results established for other approaches (e.g. Wang et al., 2005) also apply here. Another interesting line of inquiry is the use of penalized splines in flexible marginal modelling of longitudinal data with binary and count responses. Heagerty (1999), for example, describes the parametric models of this type. We believe that such models are amenable to a penalized spline approach, with implementation in BUGS, but we have not explored this deeply.

Acknowledgements

Wand's research was partially supported by Australian Research Council Discovery Project DP0877055. Carroll's research was supported by a grant from the US National Cancer Institute (CA57030) and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology, Saudi Arabia.

References

- Brumback, B.A., Ruppert, D., Wand, M.P., 1999. Comment on paper by Shively, Kohn & Wood. *Journal of the American Statistical Association* 94, 794–797.
- Carroll, R.J., Hall, P., Apanasovich, T.V., Lin, X., 2004. Histopline method in nonparametric regression models with application to clustered/longitudinal data. *Statistica Sinica* 14, 649–674.
- Carroll, R.J., Maity, A., Mammen, E., Yu, K., 2009a. Nonparametric additive regression for repeatedly measured data. *Biometrika* 96, 383–398.
- Carroll, R.J., Maity, A., Mammen, E., Yu, K., 2009b. Efficient semiparametric marginal estimation for the partially linear additive model for longitudinal/clustered data. *Statistics in Biosciences* 1, 10–31.
- Chen, K., Jin, Z., 2005. Local polynomial regression analysis of clustered data. *Biometrika* 92, 59–74.
- Coull, B.A., Ruppert, D., Wand, M.P., 2001. Simple incorporation of interactions into additive models. *Biometrics* 57, 539–545.
- Crainiceanu, C.M., Ruppert, D., Carroll, R.J., Joshi, A., Goodner, B., 2007. Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* 16, 265–288.
- Fan, J., Huang, T., Li, R., 2007. Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* 102, 632–641.
- Fan, J., Wu, Y., 2008. Semiparametric estimation of covariance matrixes for longitudinal data. *Journal of the American Statistical Association* 103, 1520–1533.
- Heagerty, P.J., 1999. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* 55, 688–698.
- Hu, Z.H., Wang, N., Carroll, R.J., 2004. Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika* 91, 251–262.
- Kipnis, V., Subar, A.F., Midthune, D., Freedman, L.S., Ballard-Barbash, R., Troiano, R., Bingham, S., Schoeller, D.A., Schatzkin, A., Carroll, R.J., 2003. The structure of dietary measurement error: results of the OPEN biomarker study. *American Journal of Epidemiology* 158, 14–21.
- Li, Y., Ruppert, D., 2008. On the asymptotics of penalized splines. *Biometrika* 95, 415–436.
- Lin, X., Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* 95, 520–534.
- Lin, X., Carroll, R.J., 2001. Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association* 96, 1045–1056.
- Lin, X., Carroll, R.J., 2006. Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society, Series B* 68, 68–88.
- Lin, X., Wang, N., Welsh, A.H., Carroll, R.J., 2004. Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika* 91, 177–193.
- Linton, O.B., Mammen, E., Lin, X., Carroll, R.J., 2003. Accounting for correlation in marginal longitudinal nonparametric regression. In: Lin, D.Y., Heagerty, P.J. (Eds.), *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*. Springer, New York, pp. 23–33.
- Ligges, U., Thomas, A., Spiegelhalter, D., Best, N., Lunn, D., Rice, K., Sturtz, S., 2009. BRugs 0.5-3: Analysis of graphical models using MCMC techniques. R package. <http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/2.10>.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10, 325–337.
- Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., The R Core Team, 2009. nlme 3.1: linear and nonlinear mixed effects models. R package. www.R-project.org.

- R Development Core Team, 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0. www.R-project.org.
- Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods, second ed.. Springer-Verlag, New York.
- Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. *Statistical Science* 6, 15–51.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression*. Cambridge University Press, New York.
- Ruppert, D., Wand, M.P., Carroll, R.J., 2009. Semiparametric regression during 2003–2007. *Electronic Journal of Statistics* 3, 1193–1256.
- Sun, Y., Zhang, W., Tong, H., 2007. Estimation of the covariance matrix of random effects in longitudinal studies. *The Annals of Statistics* 35, 2795–2814.
- Wand, M.P., Ormerod, J.T., 2008. On semiparametric regression with O'Sullivan penalized splines. *Australian and New Zealand Journal of Statistics* 50, 179–198.
- Wang, N., 2003. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* 90, 43–52.
- Wang, N., Carroll, R.J., Lin, X., 2005. Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association* 100, 147–157.
- Welham, S.J., Cullis, B.R., Kenward, M.G., Thompson, R., 2007. A comparison of mixed model splines for curve fitting. *Australian and New Zealand Journal of Statistics* 49, 1–23.
- Welsh, A.H., Lin, X., Carroll, R.J., 2002. Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association* 97, 482–493.
- Zeger, S., Diggle, P.J., 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 50, 689–699.
- Zhao, Y., Staudenmayer, J., Coull, B.A., Wand, M.P., 2006. General design Bayesian generalized linear mixed models. *Statistical Science* 21, 35–51.