2009

Marginal longitudinal semiparametric regression via penalized splines

Mohammad Alkadiri  
ma981@uow.edu.au

R. J. Carroll  
*Texas A&M University*

M. P. Wand  
*University of Wollongong, mwand@uow.edu.au*

---

**Recommended Citation**  
Alkadiri, Mohammad; Carroll, R. J.; and Wand, M. P., Marginal longitudinal semiparametric regression via penalized splines, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 05-09, 2009, 13p.  
Marginal Longitudinal Semiparametric Regression via Penalized Splines.

Alkadiri, M., Carroll, R.J. and Wand, M.P.
Marginal longitudinal semiparametric regression via penalized splines

BY M. ALKADIRI¹, R.J. CARROLL² AND M. P. WAND¹

¹ Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, New South Wales, Australia

² Department of Statistics, Texas A&M University, College Station, Texas, USA

16th July, 2009

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.

Keywords: Additive models; Best prediction; Maximum likelihood; Gibbs sampling; Nonparametric regression; Restricted maximum likelihood; Varying coefficient models.

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, Wand & Carroll (2009; Section 3.9) provide a summary of research on this problem up until about 2008. Whilst Zeger & Diggle (1994) is an early reference for marginal longitudinal nonparametric regression, the area started to heat up in response to Lin & 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: Welsh, Lin & Carroll (2002), Wang (2003), Linton, Mammen, Lin & Carroll (2003), Lin, Wang, Welsh & Carroll (2004), Carroll, Hall, Apanasovich & Lin (2004), Hu, Wang & Carroll (2004), Chen & Jin (2005), Wang, Carroll & Lin (2005), Lin & Carroll (2006) and Fan, Huang & Li (2007), Sun, Zhang & Tong (2007) and Fan & Wu (2008).

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, Ruppert & Wand, 1999; Ruppert, Wand & Carroll, 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. The Bayesian version of our models means that 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.

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 describe 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

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

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

Figure 1 shows a simulated data set for model (1), with $m = 100$, $n = 10$,

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

where $\Phi$ is the standard normal distribution function. The main problem is efficient estimation of $f$ from data such as that shown in Figure 1. Estimation of $\Sigma$ 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 $z_1, \ldots, z_K$ is a rich set of spline basis functions. A simple basis arises from setting $z_k(x) = (x - \kappa_k)_+$, where $\kappa_1, \ldots, \kappa_K$ is a dense set of knots placed over the range of the $x_i$s. However, we recommend a smoother and more numerically stable choice for $z_k$, such as those described in Welham, Cullis, Kenward & Thompson (2007) and Wand & 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 & 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 penalisation mechanism is to treat the $u_k$ as a random sample from a distribution with mean zero and variance $\sigma^2$. This permits the following linear mixed model representation of (1) and (3):

$$\begin{align*} \mathbf{y} &= \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{\varepsilon} \tag{4} \\
\beta &= \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix} \end{align*}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} z_1(x_1) & \cdots & z_K(x_1) \\ \vdots & \ddots & \vdots \\ z_1(x_m) & \cdots & z_K(x_1) \end{bmatrix}, \quad \mathbf{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$
From now on we will assume that the random vectors on the right-hand side of (4) have mean zero and covariance matrix:

\[
\text{Cov} \begin{bmatrix}
  \mathbf{u} \\
  \varepsilon_1 \\
  \varepsilon_2 \\
  \vdots \\
  \varepsilon_m
\end{bmatrix} = \begin{bmatrix}
  \sigma^2 I & 0 & 0 & \cdots & 0 \\
  0 & \Sigma & 0 & \cdots & 0 \\
  0 & 0 & \Sigma & \cdots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \cdots & \Sigma
\end{bmatrix} = \begin{bmatrix}
  \sigma^2 I & 0 \\
  0 & I_m \otimes \Sigma
\end{bmatrix}.
\]

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

\[
\begin{bmatrix}
  \mathbf{u} \\
  \varepsilon
\end{bmatrix} \sim N \left( \begin{bmatrix}
  \mathbf{0} \\
  \mathbf{0}
\end{bmatrix}, \begin{bmatrix}
  \sigma^2 I & 0 \\
  0 & I_m \otimes \Sigma
\end{bmatrix} \right).
\]

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
This is a bona fide simpler type of additive model than (6) since it only has one smooth function component.

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

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

with coefficients independently 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

\[y = X \beta + Z u + \varepsilon\]

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

\[
X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1m} & x_{2m} \end{bmatrix}, \quad Z = \begin{bmatrix} z_{11}(x_{11}) & \cdots & z_{1K_1}(x_{11}) & \cdots & z_{2K_2}(x_{21}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ z_{11}(x_{1m}) & \cdots & z_{1K_1}(x_{1m}) & \cdots & z_{2K_2}(x_{2m}) \end{bmatrix}
\]

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

\[
\beta = \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \end{bmatrix} \quad \text{and} \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}
\]

where \( u_1 \) is the \( K_1 \times 1 \) vector containing the \( u_{1k} \) and \( u_2 \) is defined similarly. The covariance matrix of the spline coefficients and errors is now

\[
\begin{bmatrix} u_1 \\ u_2 \\ \varepsilon \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 I & 0 & 0 \\ 0 & \sigma_2^2 I & 0 \\ 0 & 0 & I_m \otimes \Sigma \end{bmatrix} \right). \tag{8}
\]

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 \( \sigma_1^2 \) and \( \sigma_2^2 \) (and in extensions to additive models with \( d \) smooth functions there will be \( d \) such variance components) as well as the error covariance matrix \( \Sigma \). Maximum likelihood fitting, described in Section 3, requires an expression for \( V \equiv \text{Cov}(y) \). For the current model, this matrix takes the form

\[
V = V(\sigma_1^2, \sigma_2^2, \Sigma) = \sigma_1^2 Z_{[1]} Z_{[1]}^T + \sigma_2^2 Z_{[2]} Z_{[2]}^T + I_m \otimes \Sigma
\]

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

Before closing this section we briefly mention that the model

\[
E(y_{ij}) = \beta_0 + \beta_1 x_{1ij} + f_2(x_{2ij}), \quad \text{Cov}(y_i) = \Sigma, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{9}
\]

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 Varying Coefficient Models Extension

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

\[ E(y_{ij}) = f_0(s_{ij}) + f_1(s_{ij}) x_{ij}, \quad \text{Cov}(y_i|f_0(s_i), f_1(s_i)) = \Sigma, \quad 1 \leq i \leq m, \ 1 \leq j \leq n \quad (10) \]

where the \( s_{ij} \) are longitudinal measurements on a continuous predictor variable \( s \) and the \( x_{ij} \) are 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, Zhang & Tong (2007) paid particular attention to models of this type.

Interestingly, the Gaussian linear mixed model for fitting the varying coefficient model (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 is exactly the same as that given at (8). The only difference is that the fixed effect matrices are now

\[
X = \begin{bmatrix}
1 & s_1 & x_1 & s_1 \odot x_1 \\
\vdots & \vdots & \vdots & \vdots \\
1 & s_m & x_m & s_m \odot x_m
\end{bmatrix}
\quad \text{and} \quad 
\begin{bmatrix}
\beta_0 \\
\beta_{10} \\
\beta_{11}
\end{bmatrix}
\]

whilst the design matrix for the random effects component is

\[
Z = \begin{bmatrix}
z_1(s_1) & \cdots & z_K(s_1) & x_1 \odot z_1(s_1) & \cdots & x_1 \odot z_K(s_1) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
z_1(s_m) & \cdots & z_K(s_m) & x_m \odot z_1(s_m) & \cdots & x_m \odot z_K(s_m)
\end{bmatrix},
\]

with \( a \odot b \) denoting the element-wise product of vectors \( a \) and \( b \).

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

\[ y|u \sim N(X\beta + Zu, I_m \otimes \Sigma), \quad u \sim N(0, \text{blockdiag}(\sigma^2_{1\ell} I_{K_\ell})). \quad (11) \]

Here \( K_\ell \) corresponds to the number of spline basis functions used in the \( \ell \)th smooth function estimate. Let \( \sigma^2 = (\sigma_1^2, \ldots, \sigma_d^2) \) be the vector of variance parameters. Then the log-likelihood of \( y \) under (11) is

\[
\ell(\beta, \sigma^2, \Sigma) = -\frac{1}{2} \left\{ n \log(2\pi) + \log |V| + (y - X\beta)^T V^{-1} (y - X\beta) \right\} \quad (12)
\]

where

\[
V = V(\sigma^2, \Sigma) \equiv \text{Cov}(y) = \sum_{\ell=1}^d \sigma^2_{\ell\ell} Z_{[\ell]} Z_{[\ell]}^T + I_m \otimes \Sigma
\]

and \([Z_{[1]} \cdots Z_{[d]}]\) is the partition of \( Z \) corresponding to the basis functions for each smooth function estimate.

For any fixed values of \( \sigma^2 \) and \( \Sigma \) the fixed effects solution is

\[
\hat{\beta}(\sigma^2, \Sigma) = (X^T V^{-1} X)^{-1} X^T V^{-1} y \quad (13)
\]

On substitution into (12) we obtain the profile log-likelihood for \((\sigma^2, \Sigma)\) as:

\[
\ell_P(\sigma^2, \Sigma) = -\frac{1}{2} \left[ \log |V| + y^T V^{-1} \{ I - X(X^T V^{-1} X)^{-1} X^T V^{-1} \} y \right] - \frac{n}{2} \log(2\pi). \quad (14)
\]
However, the restricted log-likelihood (Patterson & Thompson, 1971)

$$
\ell_R(\sigma^2, \Sigma) = \ell_P(\sigma^2, \Sigma) - \frac{1}{2} \log |X^TV^{-1}X|
$$

(15)
is usually preferred since it accounts for estimation of the fixed effects vector \( \beta \). The maximizers of \( \ell_R(\sigma^2, \Sigma) \) are often labelled the restricted maximum likelihood or REML estimates of \( \sigma^2 \) and \( \Sigma \).

Likelihood-based estimation of the model parameters \( \beta, \sigma^2 \) and \( \Sigma \) thus involves:

1. Obtain the REML estimates \( \hat{\sigma}^2 \) and \( \hat{\Sigma} \) by maximising \( \ell_R(\sigma^2, \Sigma) \).

2. Obtain the maximum likelihood estimate of \( \hat{\beta} = \tilde{\beta}(\hat{\sigma}^2, \hat{\Sigma}) \) according to (13).

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

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

$$
\tilde{u}(\sigma^2, \Sigma) \equiv E(y|u) = G_{\sigma^2}Z^TV(\sigma^2, \Sigma)^{-1}\{y - X\tilde{\beta}(\sigma^2, \Sigma)\}
$$

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

$$
\hat{u} = G_{\hat{\sigma}^2}Z^TV(\hat{\sigma}^2, \hat{\Sigma})^{-1}\{y - X\hat{\beta}(\hat{\sigma}^2, \hat{\Sigma})\}.
$$

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

Despite (11) being a relatively simple linear mixed model, we have not yet been successful in fitting it with standard mixed model software such as \( \text{nlme}() \) (Pinheiro et al. 2008) in the \( \mathbb{R} \) computing language (R Core Development Team, 2009). This led us to also consider the Bayesian inference version and implementation via Gibbs sampling, as the next section describes.

### 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 (11). This entails treating \( \beta, \sigma^2 \) and \( \Sigma \) as random and setting prior distributions for each of them. The most convenient choice, because of conjugacy properties, are priors of the form:

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

(16)

where \( A_\ell, B_\ell, 1 \leq \ell \leq d \), are positive constants and \( F \) and \( B \) 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 \( \Sigma \sim \text{Inverse-Wishart}(a, B) \), where \( \Sigma \) is \( n \times n \), means that

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

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

Bayesian inference is based on the posterior density functions:

$$
|\beta|, \quad |u| \quad \text{and} \quad |\Sigma|.
$$

(17)
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 (11) and the priors (16) and, upon convergence, yields samples of arbitrary size from the posterior densities (17). 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 \( F, A_\ell, B_\ell, a \) and \( B \). If the analyst has specific prior beliefs about the model parameters then there is the opportunity to choose the prior densities to 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:

\[
F = 10^8 I \quad \text{and} \quad A_\ell = B_\ell = 0.01. 
\]  

(18)

Reasonable choices for the hyperparameters associated with \( \Sigma \) are

\[
a = n \quad \text{and} \quad B = 0.01 I_n. 
\]  

(19)

### 4.1 Gibbs Sampling Scheme

The hierarchical Bayesian model specified by (11) and (16) can be fitted using a Gibbs sampling scheme with draws from standard distributions. We give the details here.

First, we note (e.g. Robert & 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:

\[
[ \beta_u^\intercal, \sigma^2_1, \ldots, \sigma^2_d, \Sigma]. 
\]

As an example, the full conditional distribution for \( \sigma^2_1 \) is

\[
\sigma^2_1 \mid \text{rest} \sim \text{Inverse-Gamma}(A_1 + \frac{1}{2} K_1, B_1 + \frac{1}{2} \| u_1 \|^2), \quad 1 \leq \ell \leq d 
\]

and

\[
\Sigma \mid \text{rest} \sim \text{Inverse-Wishart}(a + m, B + (y - X\beta - Zu)(y - X\beta - Zu)^T). 
\]

Provided that \( a \) is an integer then the Inverse-Wishart draws for \( \Sigma \) can be achieved by setting

\[
\Sigma = \left( \sum_{i=1}^{a+m} v_i v_i^T \right)^{-1} 
\]

where the \( v_i \) are independent \( N(0, \{ B + (y - X\beta - Zu)(y - X\beta - Zu)^T \}^{-1} \) random vectors.

Interestingly, the fact that \( \Sigma \) is unstructured means that Gibbs sampling is exact. This is not the case if \( \Sigma \) 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. 2007). We pass the regression data to BUGS using matrices. For example, the variable $y_{ij}$ 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:

```plaintext
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)
  {
    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 Omega, corresponding to $\Omega = \Sigma^{-1}$. Similarly, the precision parameter tau corresponds to $\tau = 1/\sigma^2$ where $\sigma^2$ is the spline penalisation variance component.

5 Illustrations

We tested out BUGS fitting of the three types of models presented in Section 2 on several sets of simulated data. The simulation aspect also allows for comparisons were done with the true functions and marginal covariance matrix that generated the data. We now present some of these results as illustration of the methodology and its good performance.

5.1 Illustration for Nonparametric Regression

We fitted the penalized spline model (11) to the data of Figure 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 (18) and (19). 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.
Figure 2: *Summary of MCMC-based inference for the diagonal entries of $\Sigma$ 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.

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

The results for the off-diagonal entries of $\Sigma$ are summarized in Figure 3. All ten of the true values of $\Sigma_{ij}$ are captured by the 95% credible intervals in four out of the five cases.

The Bayesian penalized spline estimate of $f$ is shown in Figure 4, with and without the data. The thick solid curves correspond to the posterior means of $f$ 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 gridwise 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 $\Sigma$. An interesting future project would be a large scale simulation study that compares this approach with existing methods.

### 5.2 Illustration for Additive Models

We simulated data according to the model

$$E(y_{ij}) = \sin(2\pi(x_{1ij}^2 - 0.1)) + \sin(3\pi(0.05 - x_{2ij})),$$

$$\text{Cov}(y_i) = 0.361_5 I_5^T + 0.25 I_5$$
Figure 3: Summary of MCMC-based inference for the off-diagonal entries of $\Sigma$ 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.

for $1 \leq i \leq 200$ and $1 \leq j \leq 5$. Here $1_d$ denotes the $d \times 1$ vector of ones. For each $i$ we generated the $x_{1ij}$ to be

\[ x_{1ij} = r_i + 1/n, \quad j = 1, \ldots, n \]
where \( r_i \) is uniformly distributed on \((0, 1/n)\). An identical strategy was used for the \( x_{2ij} \). Even though the \( x_{1ij} \) and \( x_{2ij} \) were generated from a random process, they are considered fixed in the present analysis.

Figure 5 shows the posterior means of \( f_1 \) and \( f_2 \) and accompanying pointwise 95% credible sets. These answers were obtained via BUGS. Agreement with the true \( f_1 \) and \( f_2 \) is seen to be very good. The numerical summaries for the posterior of \( \Sigma \) are consistent with the truth from which the data were generated, although these are not shown because of space considerations.

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 such as missingness can be handled within the same likelihood-based or Bayesian framworks. It would be interesting to see if the asymptotic efficiency results established for other approaches (e.g. Wang, Carroll & Lin, 2005) also apply here.

Acknowledgments

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 Insti-
Figure 5: Posterior mean estimates of $f_1$ and $f_2$. Dashed curves correspond to pointwise 95% credible sets. The true $f_1$ and $f_2$ are shown as thin grey curves. Vertical alignment is achieved by plotting $f_1(x_1) + f_2(x_2)$ versus $x_1$ in the left panel, where $x_2$ is the average of the $x_{2ij}$s. The right panel is $f_1(x_1) + f_2(x_2)$ versus $x_2$.

References


Institute (CA57030) and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology, Saudi Arabia.


