Semiparametric Stochastic Mixed Models for Longitudinal Data

Daowen Zhang; Xihong Lin; Jonathan Raz; MaryFran Sowers


Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199806%2993%3A442%3C710%3ASSMMFL%3E2.0.CO%3B2-C


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/astata.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Semiparametric Stochastic Mixed Models for Longitudinal Data

Daowen ZHANG, Xihong LIN, Jonathan RAZ, and MaryFran SOWERS

We consider inference for a semiparametric stochastic mixed model for longitudinal data. This model uses parametric fixed effects to represent the covariate effects and an arbitrary smooth function to model the time effect and accounts for the within-subject correlation using random effects and a stationary or nonstationary stochastic process. We derive maximum penalized likelihood estimators of the regression coefficients and the nonparametric function. The resulting estimator of the nonparametric function is a smoothing spline. We propose and compare frequentist inference and Bayesian inference on these model components. We use restricted maximum likelihood to estimate the smoothing parameter and the variance components simultaneously. We show that estimation of all model components of interest can proceed by fitting a modified linear mixed model. We illustrate the proposed method by analyzing a hormone dataset and evaluate its performance through simulations.

KEY WORDS: Correlated data; Nonparametric regression; Penalized likelihood; Restricted maximum likelihood; Smoothing parameter; Smoothing spline; Variance components.

1. INTRODUCTION

Longitudinal studies are used in many fields of research, including epidemiology, clinical trials, and survey sampling. They are often characterized by the dependence of repeated observations over time within the same subject. Linear mixed models provide a flexible likelihood framework to model longitudinal data parametrically (Diggle, Liang, and Zeger 1994; Laird and Ware 1982). The recently developed SAS procedure PROC MIXED greatly increases the popularity of linear mixed models.

In the analysis of longitudinal data, the parametric assumption in linear mixed models may not always be appropriate. For example, in a longitudinal hormone study on progesterone (Sowers et al. 1995), urine samples were collected from 34 healthy women in a menstrual cycle and urinary progesterone was assayed on alternative days. Progesterone is a reproductive hormone responsible for normal fertility and menstrual cycling. The investigators were interested in the time course of the progesterone level in a menstrual cycle as well as the effects of age and body size on this hormone. Figure 1 plots the log-transformed progesterone level as a function of time measured in units in a standardized menstrual cycle. (For the definition of the standardized menstrual cycle, see Sec. 5.) Figure 1 suggests that the progesterone level varies over time in a complicated manner, and it is difficult to model its time trend using a simple parametric function. The same is true in other hormone studies (Sowers et al. 1998) and other areas of biomedical research, such as studies of growth (Donnelly, Laird, and Ware 1995) and HIV research on the time course of CD4 counts (Zeger and Diggle 1994). Hence it is of potential interest in these studies to model the time effect nonparametrically while accounting for the correlation of observations within the same subject.

For independent data, there is a rich literature on kernel and spline methods for nonparametric and semiparametric regression (Green and Silverman 1994; Speckman). Data-driven methods such as cross-validation (CV) and generalized cross-validation (GCV) are commonly used to select the smoothing parameter. Wahba (1978) proposed a Bayesian model for spline smoothing. Only limited work has been done on nonparametric and semiparametric regression for correlated data. Several authors considered kernel smoothing and spline smoothing for models with a single nonparametric function with correlated errors (Altman 1990; Diggle and Hutchinson 1989; Hart 1991; Rice and Silverman 1991). Zeger and Diggle (1994) proposed a semiparametric mixed model for longitudinal data similar to the one we consider in this article. They estimated the nonparametric function of time using a kernel smoother with the bandwidth parameter chosen via cross-validation and estimated the variance components by extrapolating the autocovariance function constructed under a parametric model for the time effect. For inference on the nonparametric function and the correlation parameters (especially the latter), very little has been done by these authors. In terms of modeling within-subject serial correlation, Zeger and Diggle (1994) assumed a stationary Gaussian process. This assumption may not be appropriate in some applications (Donnelly et al. 1995). For example, in the progesterone data example mentioned earlier, Figure 2 shows the empirical variance of the log progesterone level as a function of time. One can easily see that the variance changes over time.

In this article we extend the Zeger and Diggle (1994) model to a more general class of models termed semiparametric stochastic mixed models. This class of models assumes that the mean of the outcome variable is an arbitrary smooth function of time and parametric functions of the covariates. To allow for more flexible within-subject
variance–covariance structures, we propose various
stationary and nonstationary stochastic processes to model
serial correlation. Although our model consists of only a
single nonparametric function, it can be easily generalized
to accommodate interactions between the covariates and the
nonparametric time function; see Section 7 for details.

A key feature of this article is that, unlike Zeger and
Diggle (1994), we make inference on all model compo-
nents within a unified linear mixed model framework. In
contrast to the kernel smoothing approach of Zeger and
Diggle (1994), we estimate the nonparametric function of
time using a smoothing spline, which allows us to make
inference on this nonparametric function as part of the
overall inference procedure. The estimators of the regression co-
efficients and the natural cubic spline estimator of the non-
parametric function are obtained using maximum penalized
likelihood. The random effects and the stochastic process
are estimated using their conditional means given the data.
Frequentist and Bayesian inference on these model com-
nonents are proposed and compared. We treat the smoothing
parameter as an extra variance component and estimate it
jointly with the other variance components using restricted
maximum likelihood (REML). We show that all model pa-
rameters can be estimated from a modified linear mixed
model.

The article is organized as follows. Section 2 states the
model. Section 3 contains the estimation and inference pro-
cedures. Sections 3.1–3.3 present the maximum penalized
likelihood estimators of the regression coefficients and the
nonparametric time function, as well as the empirical Bayes
estimators of the random effects and the stochastic process.
Section 3.4 considers the biases and the frequentist co-
variances of these estimators under the assumption that the
nonparametric time function is a fixed parameter. Section
3.5 discusses the estimation procedures in Sections 3.1–3.3
from a Bayesian perspective and derives the Bayesian co-
variances of the model parameters by assuming the non-
parametric time function to be a random function. Our
Bayesian covariance calculation is motivated by the earlier
work of Wahba (1983), who showed that for independent
data, the Bayesian confidence intervals of the nonparamet-

ric function have good coverage probabilities when the true
nonparametric function is a fixed smooth function. In this
article we are interested in comparing the coverage prob-
babilities of the frequentist and Bayesian confidence inter-
vals of the nonparametric time function in the situation
where the true time function is a fixed smooth function,
which is the situation of our primary interest. Section 4 pro-
poses a joint estimation procedure for the smoothing parameter
and the variance components. Section 5 illustrates appli-
cation of the proposed methods to the progesterone data, and
Section 6 reports the results from a simulation study de-
signed to evaluate the performance of the proposed estima-
tion procedure. Section 7 ends the article with a discussion.

2. THE SEMIPARAMETRIC STOCHASTIC
MIXED MODEL

2.1 The Model Specification

Let the sample consist of \( m \) subjects, with the \( i \)th sub-
ject having \( n_i \) observations over time. Suppose that
\( Y_{ij} (i = 1, \ldots, m, j = 1, \ldots, n_i) \) is the response for the \( i \)th subject
at time point \( t_{ij} \) and satisfies

\[
Y_{ij} = X_{ij}' \beta + f(t_{ij}) + Z_{ij}' b_i + U_i(t_{ij}) + \epsilon_{ij},
\]

where \( \beta \) is a \( p \times 1 \) vector of regression coefficients
associated with covariates \( X_{ij} \), \( f(t) \) is a twice-differentiable
smooth function of time, the \( b_i \) are independent \( q \times 1 \) vectors of
random effects associated with covariates \( Z_{ij} \), the \( U_i(t) \)
are independent random processes used to model serial


correlation, and the \( \epsilon_{ij} \) are independent measurement errors.

We assume that \( b_i \) is distributed as normal \((0, \mathbf{D}(\phi))\), where
\( \mathbf{D} \) is a positive definite matrix depending on a parameter
vector \( \phi \); \( U_i(t) \) is a mean zero Gaussian process with cov-

ariance function \( \text{cov}(U_i(t), U_i(s)) = \gamma(\xi, \alpha; t, s) \) for a
specific parametric function \( \gamma(\cdot) \) that depends on a parameter

vector \( \xi \) and a scalar \( \alpha \) used to characterize the variance and
correlation of the process \( U_i(t) \); and \( \epsilon_{ij} \) is distributed as

normal\((0, \sigma^2)\). We further assume \( b_i, U_i(t) \), and \( \epsilon_{ij} \) to be
mutually independent.

Diggle et al. (1994) and Zeger and Diggle (1994) consid-
ered a similar model to that given in (1). But their model had
only a random intercept and assumed the stochastic process \( U_i(t) \) to be a stationary Gaussian process. If a parametric function is assumed for \( f(t) \), then model (1) reduces to the linear mixed model of Diggle et al. (1994, eq. 5.2.3).

### 2.2 The Gaussian Process Specification

Because the within-subject covariance often varies over time in longitudinal data, we propose to model serial correlation using various stationary and nonstationary Gaussian stochastic processes \( U_i(t) \). A specific choice of \( U_i(t) \) depends on the underlying subject-specific biological process in a particular application. A useful stochastic process \( U_i(t) \) to model nonstationary within-subject covariance is the nonhomogeneous Ornstein–Uhlenbeck (NOU) process (Rosenblatt 1974), which assumes the variance function \( \xi(t) \) to vary over time and the correlation function to decay exponentially over time \( \text{corr}(U_i(t), U_i(s)) = \exp(-\alpha|t-s|) \).

For example, we may assume a log-linear model for \( \xi(t) \) as \( \log(\xi(t)) = \xi_0 + \xi_1 \cdot t \). When \( \xi(t) \) is constant, the NOU process reduces to a stationary Ornstein–Uhlenbeck (OU) process (Diggle et al. 1994, sec 5.2).

An extension of the OU/NOU process is to assume \( \text{var}(U_i(t)) = \xi(t) \) and \( \text{corr}(U_i(t), U_i(s)) = \rho(\alpha; |t-s|) \) for some parametric functions \( \xi(\cdot) \) and \( \rho(\cdot) \). This includes the Gaussian correlation model \( \rho(\alpha; |t-s|) = \exp[-\alpha|t-s|^2] \) (Diggle et al. 1994, sec 5.2). Other useful choices for \( U_i(t) \) include (a) a Wiener process (Taylor, Cumberland, and Sy 1994); (b) an integrated Wiener process (Wahba 1978), which may lead to each predicted subject-specific curve similar to a natural cubic spline; (c) an integrated OU (IOU) process (Sandland and McGilchrist 1979), which is useful in growth curve analysis and assumes the relative growth rate to be a stationary OU process; and (d) an antependence process for equally spaced time points (Diggle et al. 1994, sec 5.2).

### 3. ESTIMATION AND INFERENCE FOR THE MEAN COMPONENTS

#### 3.1 Matrix Notation

Denote \( Y_i = (Y_{i1}, \ldots, Y_{in_i})^T \), and \( X_i, Z_i, U_i, \varepsilon_i \) similarly (\( i = 1, \ldots, m \)). Let \( t^0 = (t_{i1}^0, \ldots, t_{in_i}^0)^T \) be a vector of ordered distinct values of the time points \( t_{ij} = (1, \ldots, m, j = 1, \ldots, n_i) \) and let \( N_i \) be the incidence matrix for the \( i \)th subject connecting \( t_i = (t_{i1}, \ldots, t_{in_i})^T \) and \( t^0 \) such that the \( (j, l) \)th element of \( N_i \) is 1 if \( t_{ij} = t_{jl}^0 \) and 0 otherwise (\( j = 1, \ldots, n_i, l = 1, \ldots, r \)). Then model (1) can be written as

\[
Y_i = X_i\beta + N_i f + Z_i b_i + U_i + \varepsilon_i, \tag{2}
\]

where \( f = (f(t_{i1}^0), \ldots, f(t_{in_i}^0))^T \).

Further denoting \( Y = (Y_1^T, \ldots, Y_m^T)^T \) and \( X, N, \varepsilon \) similarly, and \( Z = \text{diag}(Z_1, \ldots, Z_m) \), we have

\[
Y = X\beta + Nf + Zb + U + \varepsilon, \tag{3}
\]

where \( b = (b_1^T, \ldots, b_m^T)^T \) is normal \((0, \mathcal{D}(\phi)) \), with \( \mathcal{D}(\phi) = \text{diag}(D(D), \ldots, D(D)) \); \( U = (U_1^T, \ldots, U_m^T)^T \) is normal \((0, \Gamma(\xi, \alpha)) \), with \( \Gamma(\xi, \alpha) = \text{diag}(\Gamma_1(t_1, t_1), \ldots, \Gamma_m(t_m, t_m)) \) and the \((j, j')\)th element \((j, j' = 1, \ldots, n)\) of \( \Gamma(t_i, t_j) \) being \( \xi_0 + \xi_1 \cdot t \); and \( \varepsilon = (\varepsilon_1^T, \ldots, \varepsilon_m^T)^T \) is normal \((0, \sigma^2I) \), with \( I \) denoting an identity matrix of dimension \( n = \sum_{i=1}^m n_i \).

#### 3.2 Estimation of the Regression Coefficients, Nonparametric Function, Random Effects, and Stochastic Process

For given variance components \( \theta = (\phi^T, \xi^T, \alpha, \sigma^2)^T \), the log-likelihood function of \((\beta, f)\), is, apart from a constant,

\[
l(\beta, f; Y) = -\frac{1}{2} \log |V| - \frac{1}{2} (Y - X\beta - Nf)^T V^{-1} (Y - X\beta - Nf),
\]

where \( V = \text{diag}(V_1, \ldots, V_m) \) and \( V_i = Z_i D Z_i^T + \Gamma_i + \sigma^2I \). Because \( f(t) \) is an infinite-dimensional parameter, we consider the maximum penalized likelihood estimator (MPLE) of \( \beta \) and \( f(t) \), which leads to a natural cubic spline estimator of \( f(t) \) (O’Sullivan, Yandell, and Raynor 1986).

Specifically, the MPLE of \((\beta, f(t))\) maximizes

\[
l(\beta, f; Y) - \frac{\lambda}{2} \int_{T_1}^{T_2} [f''(t)]^2 dt = l(\beta, f; Y) - \frac{\lambda}{2} f^T K f, \tag{4}
\]

where \( \lambda \geq 0 \) is a smoothing parameter controlling the balance between the goodness of fit and the roughness of the estimated \( f(t) \). \( T_1 \) and \( T_2 \) specify the range of \( t \), and \( K \) is the nonnegative definite smoothing matrix defined in equation (2.3) of Green and Silverman (1994).

For fixed smoothing parameter \( \lambda \) and variance components \( \theta \), differentiation of (4) with respect to \( \beta \) and \( f \) shows that the MPLE \((\hat{\beta}, \hat{f})\) solves

\[
\begin{bmatrix}
X^T W X & X^T W N \\
N^T W X & N^T W N + \lambda K
\end{bmatrix}
\begin{bmatrix}
\beta \\
\hat{f}
\end{bmatrix}
= \begin{bmatrix}
X^T W Y \\
N^T W Y
\end{bmatrix},
\tag{5}
\]

where \( W = V^{-1} = \text{diag}(W_1, \ldots, W_m) \) and \( W_i = V_i^{-1} \). Equation (5) has a unique solution if and only if the matrix \([X, NT]\) is of full rank, where \( T = [1, t^0] \) and \( I \) is an \( r \times 1 \) vector of 1’s. Note that this condition is also a necessary and sufficient condition for the coefficient matrix of equation (5), which we denote by \( C \), to be positive definite.

To study the theoretical properties of the MPLEs (Sects. 3.4 and 3.5), it is useful to derive a closed-form solution of (5). Denoting two nonnegative definite weight matrices by \( W_x = W - W N (N^T W N + \lambda K)^{-1} N^T W \) and \( W_f = W - W X (X^T W X)^{-1} X^T W \), some calculations give the MPLEs of \( \beta \) and \( f \) as

\[
\hat{\beta} = (X^T W_x X)^{-1} X^T W_x Y
\]

and

\[
\hat{f} = (N^T W_f N + \lambda K)^{-1} N^T W_f Y. \tag{7}
\]

Note that both \((X^T W_x X)^{-1}\) and \((N^T W_f N + \lambda K)^{-1}\) are positive definite given that \([X, NT]\) is of full rank. These two matrices are in fact the corresponding block-diagonal elements of \( C^{-1} \).
Let \( s_i = (s_{i1}, \ldots, s_{ic_i})^T \) be an arbitrary \( c_i \times 1 \) vector of time points. Estimation of the subject-specific random effects \( b_i \) and the subject-specific stochastic process \( U_i(s_i)(i = 1, \ldots, m) \) may proceed by calculating their conditional expectations given the data \( Y_i \), while estimating \( \beta \) and \( f \) by their MPLEs. This gives

\[
\hat{b}_i = DZ_i^TW_i(Y_i - X_i\hat{\beta} - \hat{f}_i)
\]

(8)

and

\[
\hat{U}_i(s_i) = \Gamma_i(s_i, t_i)W_i(Y_i - X_i\hat{\beta} - \hat{f}_i),
\]

(9)

where the \((j, j')\)th element of \( \Gamma_i(s_i, t_i) \) is \( \gamma(\xi, \alpha; s_{ij}, t_{ij'}) \) and \( \hat{f}_i = N\hat{f} \).

3.3 The Linear Mixed Model Representation

In view of the quadratic form of the penalty term in the penalized log-likelihood (4), we show in this section how to derive the MPLEs \( \hat{\beta} \) and \( \hat{f} \) by writing the semiparametric mixed model (3) as a modified linear mixed model. This linear mixed model representation provides a foundation for the estimation procedure of the smoothing parameter and the variance components (Sec. 4).

Following Green (1987), we can write \( f \) via a one-to-one linear transformation as

\[
f = T\delta + Ba,
\]

(10)

where \( \delta \) and \( a \) are vectors with dimensions 2 and \( r - 2 \), \( B = L(L^TL)^{-1}L \) is \( r \times (r - 2) \) full-rank matrix satisfying \( K = LL^T \) and \( L^TT = 0 \). Using the equality \( f^TWf = a^Ta \), we have that the penalized log-likelihood (4) differs by no more than an additive constant from

\[
\frac{1}{2} \log |V| - \frac{1}{2} (Y - X\beta - NT\delta - NBa)^T
\times V^{-1}(Y - X\beta - NT\delta - NBa) - \frac{1}{2\tau} a^Ta,
\]

(11)

where \( \tau = 1/\lambda \). Consequently, when \( \hat{f}(t) \) is a natural cubic spline, we can write the semiparametric mixed model (3) as a modified linear mixed model,

\[
Y = X\beta + NT\delta + NBa + Zb + U + \varepsilon,
\]

(12)

where \( \beta = (\beta^T, \delta^T)^T \) are the regression coefficients and \( \beta_a = (\alpha^T, \beta^T, U^T)^T \) are mutually independent random effects with a distributed as normal(0, \( \tau I \)), and \( (b, U) \) having the same distributions as those given in Section 3.1. Because the design matrix \( NB \) does not have a block-diagonal structure, the random effects \( a \) and \( (b, U) \) hence may be viewed as random effects generated from a crossed design.

It can be easily shown that the best linear unbiased predictors (BLUPs) \( \hat{\beta} \) and \( \hat{f} = T\hat{\delta} + Ba \) from linear mixed model (12) are identical to the MPLEs from (5)-(7). The same is true for the estimators of the random effects \( \hat{b}_i \) in (8) and the estimator of the stochastic process \( \hat{U}_i(s_i) \) in (9). Note that we here represent the MPLE \( \hat{f} \) as a linear combination of the estimated fixed effects \( \hat{\delta} \) and random effects \( \hat{a} \).

It is often computationally expensive to solve \( \hat{\beta} \) and \( \hat{f} \) from the standard normal equation of model (12) (see, e.g., Harville 1977, eqs. 3.1 and 3.3), because of the potentially high dimension of the random effects \( b_i \) and the lack of block-diagonal structure of \( V_* = \tau B_*B_*^T + V \), the marginal variance of \( Y \) under the mixed-model representation (12). However, one can show that the BLUPs \( (\hat{\beta}, \hat{\delta}, \hat{a}) \) of model (12) are identical to those solving

\[
\begin{bmatrix}
X^TWX_* & X^TWB_* \\
B_*^TWX_* & B_*^TWB_* + \lambda I
\end{bmatrix}
\begin{bmatrix}
\beta_* \\
\delta_*
\end{bmatrix}
= \begin{bmatrix}
X^TWY \\
B_*^TWY
\end{bmatrix},
\]

(13)

where \( X_* = [X, NT] \) and \( B_* = NB \). Note that (13) has a structure parallel to that of (5). An advantage of estimating \( \hat{\beta} \) and \( \hat{f} \) from (13), or equivalently from (5), is that it requires much less computation. This is because the dimension of \( a \) is much smaller and is of the same magnitude as the number of distinct time points, which is often much less than the total number of observations in many longitudinal studies. Furthermore, calculation of the weight matrix \( W \) in these equations requires only inverting the block-diagonal covariance matrix \( V \).

3.4 Biases and Covariances of the Regression Coefficients, Nonparametric Function, Random Effects, and Stochastic Process

In this section we study the biases and frequentist covariances of the parameter estimators obtained in Sections 3.2 and 3.3, assuming that \( f(t) \) is a fixed parameter. Simple calculations using (6) and (7) show that the biases of the MPLEs \( \hat{\beta} \) and \( \hat{f} \) are

\[
E(\hat{\beta}) - \beta = (X^TW_xX)^{-1}X^TW_xNf
\]

and

\[
E(\hat{f}) - f = -\lambda(N^TW_fN + \lambda K)^{-1}Kf.
\]

These biases go to 0 as \( \lambda \downarrow 0 \). Note that \( \hat{f} \) is a shrinkage estimator of \( f \), because the eigenvalues of \( \lambda(N^TW_fN + \lambda K)^{-1}K \) are between 0 and 1. Similarly, the biases in the estimators of the random effects \( b_i \) and the stochastic process \( \hat{U}_i(s_i) \) are

\[
E(\hat{b}_i) = D_iZ_i^TW_i[\lambda N_i(N^TW_fN + \lambda K)^{-1}K
- X_i(X^TW_xX)^{-1}XW_xN]f
\]

and

\[
E(\hat{U}_i(s_i)) = \Gamma_i(s_i, t_i)W_i[\lambda N_i(N^TW_fN + \lambda K)^{-1}K
- X_i(X^TW_xX)^{-1}XW_xN]f.
\]

The covariances of the MPLEs \( (\hat{\beta}, \hat{f}) \) can be easily calculated from (6) and (7),

\[
\text{cov}(\hat{\beta}) = (X^TW_xX)^{-1}
\times X^TW_xVW_xX(X^TW_xX)^{-1},
\]

(14)

and

\[
\text{cov}(\hat{f}) = (N^TW_fN + \lambda K)^{-1}
\times N^TW_fVW_fN(N^TW_fN + \lambda K)^{-1}.
\]

(15)
For independent data, (14) reduces to (5.4) of Green (1987). Under the classical nonparametric model,

\[ y_i = f(t_i) + \epsilon_i, \]

(16)

where the \( \epsilon_i \) are independent and follow normal(0, \( \sigma^2 \)), the covariance of \( \hat{f} \) given in (15) reduces to the familiar form \( \sigma^2 A^2 \), where \( A = (I + \lambda K)^{-1} \) and \( \lambda = \sigma^2 \lambda \) (Hastie and Tibshirani 1991).

The covariances of the estimators of the random effects \( b_i \) and the stochastic process \( U_i(s_i) \) are

\[
\text{cov}(\hat{b}_i - b_i) = D - DZ_i^T W_i Z_i D + DZ_i^T W_i \lambda_i C^{-1} \lambda_i^T W_i \lambda_i C^{-1} \lambda_i^T W_i Z_i D
\]

and

\[
\text{cov}(\hat{U}_i(s_i) - U_i(s_i))
= \Gamma(s_i, s_i) - \Gamma(s_i, t_i) W_i \lambda_i \Gamma(s_i, t_i)^T + \Gamma(s_i, t_i) W_i \lambda_i C^{-1} \lambda_i^T W_i \lambda_i C^{-1} \lambda_i^T W_i \Gamma(s_i, t_i)^T,
\]

where \( \lambda_i = [X_i, N_i] \).

3.5 Bayesian Formulation and Inference

In this section we first show that the MPLEs \( (\hat{\beta}, \hat{f}) \) and the random-effects estimators \( (\hat{b}, \hat{U}) \) obtained in Sections 3.2 and 3.3 can be derived within a Bayesian framework, and then derive the Bayesian standard errors of \( (\hat{\beta}, \hat{f}) \) in the spirit of Wahba (1983) as an alternative to the frequentist standard errors given in (14) and (15).

We first derive the MPLEs \( (\hat{\beta}, \hat{f}) \) from a Bayesian prospective. Assume that \( \beta \) has no informative prior on \( R^p \), the p-dimensional Euclidean space, and \( f(t) \) has a partially improper integrated Wiener prior (Wahba 1978),

\[
f(t) = \delta_0 + \delta_1 t + \lambda^{-1/2} \int_0^t \mathcal{W}(u) du,
\]

(17)

where \( \delta_0 \) and \( \delta_1 \) have noninformative priors on \((-\infty, +\infty)\) and \( \mathcal{W}(t) \) is the standard Wiener process. Note that the prior specification of \( f \) using (17) is equivalent to assuming \( f \) takes the form in (10) with a flat prior for \( \delta \), a normal prior \( \mathcal{N}(0, \tau I) \) for \( a \), and \( B = \Sigma^{1/2} \), where \( \Sigma \) is the integrated Wiener covariance matrix evaluated at \( t_0 \). Following Wahba (1978), one can easily show that the posterior modes and means of \( (\beta, f) \) under this Bayesian model are identical to the MPLEs \( (\hat{\beta}, \hat{f}) \), as are the estimators of the random effects and the stochastic process \( \hat{b} \) and \( \hat{U} \).

An alternative Bayesian formulation to obtain the MPLEs \( (\hat{\beta}, \hat{f}) \) using the posterior modes is based on the finite-dimensional Bayesian model of Green and Silverman (1994). Specifically, one can easily see that equation (4) is the posterior log-likelihood of \( (\beta, f) \) if one assumes a flat prior for \( \beta \) and a partially improper Gaussian prior for \( f \) whose log density has kernel \(-\lambda f^2 K f/2\).

Under the classical nonparametric model (16), Wahba (1985) suggested estimating the covariance of \( \hat{f} \) using the posterior covariance of \( \hat{f} \) under the Bayesian model (17). She showed that the resulting Bayesian confidence intervals of \( f \) have good coverage probabilities when \( f(t) \) is a fixed smooth function.

We suspect that the same may be true under the semiparametric stochastic mixed model (1). We hence propose estimating the covariances of \( \hat{\beta} \) and \( \hat{f} \) by their posterior covariances under the Bayesian models discussed earlier as an alternative to the covariance estimators given in Section 3.4. Some calculations show that the Bayesian covariances of \( \hat{\beta} \) and \( \hat{f} \) have much simpler forms:

\[
\text{cov}_B(\hat{\beta}) = (X^T W_x X)^{-1}
\]

and

\[
\text{cov}_B(\hat{f}) = (N^T W_f N + \lambda K)^{-1}.
\]

(18)

(19)

It is easy to show that these Bayesian covariances are identical to the covariances \( \text{cov}(\hat{\beta}) \) and \( \text{cov}(\hat{f}) = \text{cov}(T \hat{\delta} + B(\hat{a} - \hat{a})) \) calculated from the modified linear mixed model (12).

Under the classical nonparametric model (16), the Bayesian covariance of \( \hat{f} \) given in equation (19) reduces to the familiar form \( \sigma^2 A \) (Wahba 1983), where \( A \) is defined in Section 3.4.

Because the differences of the Bayesian covariance matrices in (18)–(19) and the frequentist covariance matrices in (14)–(15), \( \text{cov}_B(\beta) - \text{cov}(\beta) \) and \( \text{cov}_B(\hat{f}) - \text{cov}(\hat{f}) \), are always nonnegative definite, the Bayesian standard errors of \( \beta \) and \( f \) hence are often greater than their frequentist counterparts. As shown by Wahba (1983, p134), the average of the Bayesian variances of \( f \) is approximately equal to the average of the mean square errors of \( \hat{f} \), which is the average of the squared biases plus the average of the frequentist variances. Therefore, an interpretation for larger Bayesian standard errors of \( \hat{f} \) and \( \hat{\beta} \) could be that they account for the biases in \( \hat{f} \) and \( \hat{\beta} \).

The frequentist and Bayesian covariances become identical when \( \lambda \downarrow 0 \) or \( \lambda \uparrow \infty \). In Section 5 we compare through simulations the coverage probabilities of the resulting frequentist and Bayesian confidence interval estimators of \( f \) when the true \( f(t) \) is a fixed smooth function, which is the situation of our primary interest. Our view of the Bayesian standard error of \( f \) follows that of Berger (1980) and Wahba (1983), which is to construct the confidence interval of \( f \) under a prior and to examine how it performs in the case of interest; that is, when \( f(t) \) is a fixed parameter.

Compared to the frequentist covariances, the Bayesian covariances of \( \hat{b}_i \) and \( \hat{U}_i(s_i) \) also have simpler forms and are calculated by assuming that \( f \) has a partially improper prior given earlier in this section:

\[
\text{cov}_B(\hat{b}_i - b_i) = D - DZ_i^T W_i Z_i D + DZ_i^T W_i \lambda_i C^{-1} \lambda_i^T W_i Z_i D
\]

and

\[
\text{cov}_B(\hat{U}_i(s_i) - U_i(s_i)) = \Gamma(s_i, s_i) - \Gamma(s_i, t_i) W_i \lambda_i \Gamma(s_i, t_i)^T + \Gamma(s_i, t_i) W_i \lambda_i C^{-1} \lambda_i^T W_i \Gamma(s_i, t_i)^T.
\]

Note that these Bayesian covariances of \( \hat{b}_i \) and \( \hat{U}_i(s_i) \) are identical to those obtained from the modified linear mixed model (12).
4. ESTIMATION OF THE SMOOTHING PARAMETER AND VARIANCE COMPONENTS

We assume in Section 3 that the smoothing parameter \( \lambda \) and the variance components \( \theta \) are known when we make inference about the mean parameters. But they are often unknown in practice and need to be estimated from the data. Many authors have pointed out that it is crucial to have a good estimator of the smoothing parameter \( \lambda \) to ensure a good performance of the estimator \( \hat{f}(t) \) (Green and Silverman 1994; Wahba 1978a). A classical data-driven approach to selecting the smoothing parameter is cross-validation, which leaves out one subject’s entire data at a time (Rice and Silverman 1991; Zeger and Diggle 1994). But this approach is often expensive, and the subsequent inference on the variance components is difficult (Zeger and Diggle 1994).

Under the simple nonparametric model (16) for independent data, Wahba (1985) proposed estimating the smoothing parameter \( \lambda \) using generalized maximum likelihood (GML) by assuming \( f(t) \) has a partially improper integrated Wiener prior (17). Speed (1991) pointed out that the GML estimator of \( \tau = 1/\lambda \) is identical to the restricted maximum likelihood (REML) (Harville 1977) estimator of the variance component \( \tau \) under the linear mixed model \( \mathbf{Y} = \mathbf{T}\delta + \mathbf{B}a + \mathbf{e} \), where \( a \sim N(0, \tau I) \) and \( \mathbf{e} \sim N(0, \sigma^2 I) \), and \( \mathbf{T} \) and \( \mathbf{B} \) are defined in Section 3.3.

REML has also been widely used for selecting the smoothing parameter in the state-space model literature, where a smoothing spline is formulated using a state-space model and REML or ML is often used to estimate the smoothing parameter. For example, Ansley, Kohn, and Wong (1993) and Wecker and Ansley (1983) estimated the smoothing parameter using ML and REML in the classical nonparametric regression model (16) with a fixed unknown smooth function \( f(t) \) using the state-space model formulation.

Kohn, Ansley, and Tharm (1991) showed in an extensive simulation study that the REML estimator of \( \tau \) performs well and often better compared to the CV and GCV estimators in estimating the nonparametric function under (16). Further simulation studies by Ansley et al. (1993) gave similar conclusions on the performance of REML and GCV.

Motivated by these authors’ results and using the connection between the smoothing spline and the linear mixed models established in Section 3.3, we propose estimating \( \lambda \) and \( \theta \) simultaneously using REML by treating \( \tau \) as an extra variance component in the linear mixed model (12). The REML log-likelihood of \((\tau, \theta)\) is

\[
l_R(\tau, \theta; \mathbf{Y}) = -\frac{1}{2} \log |\mathbf{V}_\ast| - \frac{1}{2} \log |\mathbf{X}_\ast^T \mathbf{V}_\ast^{-1} \mathbf{X}_\ast|
- \frac{1}{2} (\mathbf{Y} - \mathbf{X}_\ast \hat{\beta}_\ast)^T \mathbf{V}_\ast^{-1} (\mathbf{Y} - \mathbf{X}_\ast \hat{\beta}_\ast),
\]

where \( \mathbf{X}_\ast \) and \( \mathbf{V}_\ast \) were defined in Section 3.3. Note that the REML log-likelihood \( l_R(\tau, \theta; \mathbf{Y}) \) in (20) can also be motivated from a finite-dimensional Bayesian model discussed in Section 3.5. Specifically, \( l_R(\tau, \theta; \mathbf{Y}) \) is the marginal log-likelihood of \( \mathbf{Y} \) obtained by assuming a flat prior for \((\beta, \delta)\) and a Gaussian prior for \( a \) as normal\((0, \tau I)\) in \( l(\beta, \delta, a; \mathbf{Y}) = l(\beta, \delta, a; \mathbf{Y}) \), and then integrating out \((\beta, \delta)\) and \( a \) (Harville 1977; Wahba 1985).

Differentiating \( l_R(\tau, \theta; \mathbf{Y}) \) with respect to \( \tau \) and \( \theta \) and using the identity \( \mathbf{V}_\ast^{-1}(\mathbf{Y} - \mathbf{X}_\ast \hat{\beta}_\ast) = \mathbf{V}_\ast^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{\xi}) \), the REML estimating equations for \( \tau \) and \( \theta \) are

\[
-\frac{1}{2} \text{Tr}(\mathbf{P}_* \mathbf{B}_* \mathbf{B}^T) + \frac{1}{2} (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{\xi})^T \mathbf{V}_\ast^{-1} \mathbf{B}_* \mathbf{B}^T \mathbf{V}_\ast^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{\xi}) = 0 \tag{21}
\]

and

\[
-\frac{1}{2} \text{Tr} \left( \frac{\partial \mathbf{V}}{\partial \theta_j} \right) + \frac{1}{2} (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{\xi})^T \mathbf{V}_\ast^{-1} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{V}_\ast^{-1} (\mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{N}\hat{\xi}) = 0, \tag{22}
\]

where

\[
\mathbf{P}_* = \mathbf{V}_\ast^{-1} - \mathbf{V}_\ast^{-1} \mathbf{X}_\ast (\mathbf{X}_\ast^T \mathbf{V}_\ast^{-1} \mathbf{X}_\ast)^{-1} \mathbf{X}_\ast^T \mathbf{V}_\ast^{-1},
\]

\[
\mathbf{V} \text{ and } \mathbf{C} \text{ were defined in Section 3.2, and } \mathbf{X} = [\mathbf{X}, \mathbf{N}] \text{.}
\]

Calculation of \( \mathbf{P}_* \) from (23) is often computationally prohibitive, because the \( n \times n \) covariance matrix \( \mathbf{V}_\ast \) does not have a block-diagonal structure. But it is often computationally feasible to calculate \( \mathbf{P}_* \) from (24), because \( \mathbf{V} \) is a block-diagonal matrix and the dimension of \( \mathbf{C} \) is about the same as the number of distinct time points, which is often much lower than the number of total observations. The Fisher scoring algorithm then can be used to solve (21) and (22) for \( \tau \) and \( \theta \).

For the classical nonparametric model (16), it can easily be shown that the REML estimator of the smoothing parameter \( \lambda \) from (21) reduces to Wahba’s (1985) GML estimator. It is interesting to note that the REML estimator of the residual variance \( \sigma^2 \) from (22) is identical to the conventional estimator \( \hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \hat{f}(t_i))^2 / \text{Tr}(\mathbf{I} - \mathbf{A}) \) (Green and Silverman 1994, eq. 3.19), where \( \mathbf{A} \) was given in Section 3.4. This REML property of \( \hat{\sigma}^2 \) provides a more systematic justification for the use of \( \hat{\sigma}^2 \) in the classical nonparametric model (16), and has not been recognized in the literature.

The covariance of \( \hat{\theta} \) can be estimated using the corresponding block of the inverse of the Fisher information matrix obtained from (21) and (22),

\[
\mathbf{I} = \begin{pmatrix}
\mathbf{I}_{\tau \tau} & \mathbf{I}_{\tau \theta} \\
\mathbf{I}_{\tau \theta} & \mathbf{I}_{\theta \theta}
\end{pmatrix}, \tag{25}
\]

where

\[
\begin{align*}
\mathbf{I}_{\tau \tau} &= \frac{1}{2} \text{Tr} (\mathbf{P}_* \mathbf{B}_* \mathbf{B}^T) \\
\mathbf{I}_{\tau \theta} &= \frac{1}{2} \text{Tr} \left( \mathbf{P}_* \mathbf{B}_* \mathbf{B}^T \mathbf{P}_* \frac{\partial \mathbf{V}}{\partial \theta_j} \right), \\
\mathbf{I}_{\theta \theta} &= \frac{1}{2} \text{Tr} \left( \mathbf{P}_* \frac{\partial \mathbf{V}}{\partial \theta_k} \mathbf{P}_* \frac{\partial \mathbf{V}}{\partial \theta_k} \right),
\end{align*}
\]

and

\[
\mathbf{I}_{\tau \tau} = \frac{1}{2} \text{Tr} (\mathbf{P}_* \mathbf{B}_* \mathbf{B}^T) \mathbf{P}_* \mathbf{B}_* \mathbf{B}^T,
\]
Table 1. Estimates of the Regression Coefficients, Variance Components and Smoothing Parameter in the Progesterone Data

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Parameter estimate</th>
<th>Bayesian SE</th>
<th>Frequentist SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>.9247</td>
<td>1.9236</td>
<td>1.9236</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-2.9127</td>
<td>2.3762</td>
<td>2.3761</td>
</tr>
<tr>
<td>$\phi$</td>
<td>.2617</td>
<td>.0719</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>.9693</td>
<td>.0617</td>
<td></td>
</tr>
<tr>
<td>$\xi_0$</td>
<td>-2.1563</td>
<td>.4469</td>
<td></td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>3.6710</td>
<td>1.2214</td>
<td></td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>-2.1598</td>
<td>.8071</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>.1366</td>
<td>.0163</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>7.2976</td>
<td>4.4997</td>
<td></td>
</tr>
</tbody>
</table>

Note that we here are interested mainly in using (25) to draw inference for the variance components $\theta$ but not for the smoothing parameter $\tau$. Our rationale for using (25) to construct the variance of $\hat{\theta}$ is to account for the loss of information from estimating $\tau$. We examine the performance of this variance estimate of $\hat{\theta}$ in Section 6 through simulations.

5. APPLICATION TO THE PROGESTERONE DATA

We applied the proposed semiparametric stochastic mixed model to analyzing the longitudinal progesterone data discussed in Section 1. Among the 34 study participants, a total of 492 observations were obtained, with each woman contributing from 11 to 28 observations over time. This gave an average of 14.5 observations per woman. The menstrual cycle lengths of these women ranged from 23 to 56 days, with an average of 29.6 days. To overcome the problem of unequal cycle lengths among the study participants, each woman’s menstrual cycle length was standardized uniformly to a reference 28-day cycle (Sowers et al. 1998). The reason of standardization is that it is biologically meaningful to assume that the change of the progesterone level for each woman depends on the time during a menstrual cycle relative to her cycle length. The standardization resulted in 98 distinct time points. A log transformation was applied to the progesterone level to make the normality assumption more plausible.

Figure 1 displays the log-transformed progesterone values during a standardized menstrual cycle. Figure 2 shows their empirical sample variances, which were calculated by grouping the data into 28 1-day intervals. Note that because urine samples were analyzed on alternative days, the observations within every 1-day interval came from different subjects and hence were independent. Examination of Figures 1 and 2 suggests that both the mean and variance of the log progesterone level change over time nonlinearly.

Denote by $Y_{ij}$ the $j$th log-transformed progesterone value measured at standardized day $t_{ij}$ since menstruation for the $i$th woman, and by $AGE_i$ and $BMI_i$, her age and body mass index. We considered the following semiparametric stochastic mixed model:

$$Y_{ij} = \beta_1 AGE_i + \beta_2 BMI_i + f(t_{ij}) + b_i + U_i(t_{ij}) + \epsilon_{ij},$$

where the $b_i$ are independent random intercepts following a normal$(0, \phi)$ distribution, the $U_i(t)$ are mean 0 Gaussian processes modeling serial correlation, and the $\epsilon_{ij}$ are independent measurement errors following a normal$(0, \sigma^2)$ distribution. To allow the variance of $Y_{ij}$ to vary over time, we assumed $U_i(t)$ to be a NOU process, which satisfies $\text{var}(U_i(t)) = \xi(t)$ with $\log(\xi(t)) = \xi_0 + \xi_1 t + \xi_2 t^2$ and $\text{corr}(U_i(t), U_i(s)) = \rho^{t-s}$. Note that here we assume that each woman’s serial correlation is the same on the transformed time scale. Other correlation structures are also possible; see Section 7 for more discussion. For the sake of computational stability in fitting model (26), standardized day since menstruation was centered at the median 14 days and divided by 10, whereas age and BMI were centered at the medians 36 years and 26 kg/m$^2$ and divided by 100. Hence $f(t)$ represents the progesterone profile for the population of women with age 36 years and BMI 26 kg/m$^2$.

Figure 1 superimposes the MLE of $f(t)$ and its pointwise 95% frequentist and Bayesian confidence intervals. The levels of progesterone remain relatively low and stable in the first half of a menstrual cycle and increase markedly after ovulation. They reach a peak around reference day 23.

Table 2. Simulation Results on Estimates of Model Parameters Based on 500 Replications

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>Model parameter</th>
<th>Relative bias</th>
<th>Empirical SE</th>
<th>Model-based Bayesian SE</th>
<th>Model-based frequentist SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 34$</td>
<td>$\beta_1$</td>
<td>-.0739</td>
<td>2.0142</td>
<td>1.9108</td>
<td>1.9107</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>.0077</td>
<td>2.2726</td>
<td>2.3604</td>
<td>2.3604</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>.0091</td>
<td>.0793</td>
<td>.0729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>.1755</td>
<td>.0698</td>
<td>.0687</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_0$</td>
<td>.0883</td>
<td>.5404</td>
<td>.4871</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_1$</td>
<td>.1149</td>
<td>1.4560</td>
<td>1.2940</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_2$</td>
<td>.1142</td>
<td>.9261</td>
<td>.8433</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>.0232</td>
<td>.0171</td>
<td>.0160</td>
<td></td>
</tr>
<tr>
<td>$m = 68$</td>
<td>$\beta_1$</td>
<td>.0583</td>
<td>1.3777</td>
<td>1.3614</td>
<td>1.3613</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>.0256</td>
<td>1.6377</td>
<td>1.1817</td>
<td>1.1816</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>.0082</td>
<td>.0450</td>
<td>.0505</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>.0147</td>
<td>.0442</td>
<td>.0421</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_0$</td>
<td>.0002</td>
<td>.3187</td>
<td>.3079</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_1$</td>
<td>-.0382</td>
<td>.8852</td>
<td>.8261</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\xi_2$</td>
<td>-.0612</td>
<td>.5789</td>
<td>.5465</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>-.0091</td>
<td>.0123</td>
<td>.0118</td>
<td></td>
</tr>
</tbody>
</table>
and then decrease. The frequentist and Bayesian confidence intervals of the natural cubic spline estimator $f(t)$ are very close. This finding reflects our theoretical results in Section 3.4, because the estimate of smoothing parameter $\tau$ is much larger than the other estimated variance components (Table 1).

Table 1 presents estimates of the regression coefficients, the variance components, and the smoothing parameter. Age and BMI were found to have no significant effect on progesterone level. Note that the frequentist and Bayesian standard errors of the estimates of the regression coefficients are almost identical. This again reflects our theoretical results in Section 3.4 and may be due to the same reason stated in the previous paragraph. The estimates of $(\xi_0, \xi_1, \xi_2)$ indicate that the variance of the outcome progesterone level varies strongly over time. The estimated variance curve superimposed in Figure 2 agrees quite well with the empirical variances.

6. A SIMULATION STUDY

We carried out a simulation study to evaluate the performance of the MPLS estimators of the regression coefficients and the nonparametric function with the smoothing parameter estimated using REML and the REML variance components. The design of the simulation study was identical to that of the original progesterone data. We generated data $(m = 34)$ from model (26) and set the true regression coefficients $\beta$, the fixed nonparametric function $f(t)$, and the true variance components $\theta$ equal to the estimates from the analysis of the real data. To study the influence of the sample size on the performance of the estimation procedure, we repeated this experiment with the entire design duplicated and the number of subjects doubled so that $m = 68$. We carried out 500 simulations under each experiment.

Table 2 presents the relative biases, empirical standard errors, and model-based Bayesian and frequentist standard errors of the parameter estimates. The relative bias is defined as the bias of a parameter estimate divided by its true value. The MPLSs of the regression coefficients are nearly unbiased, whereas the REML estimates of the variance components are slightly biased when $m = 34$ and become virtually unbiased when $m = 68$. The empirical and model-based standard errors of the estimates of the regression coefficients and the variance components closely agree with each other. The Bayesian and frequentist standard errors of the regression coefficient estimates are almost identical. This may be due to the fact that the bias in $\beta$ is negligible, and the true fixed nonparametric function $f(t)$ has a large curvature, which results in the values of the estimated smoothing parameter $\tilde{f}$ in all simulated datasets being much larger than the other variance components. Note that only Bayesian SEs of the variance component estimators are given, because their frequentist counterparts are difficult to calculate from (22) and the Bayesian SEs perform very well.

The bias in the estimated nonparametric function $f(t)$ is minimal and becomes smaller as the sample size doubles (Fig. 3). The bias is relatively higher at values of $t$ where the variance of the outcome variable is larger. Figure 4 shows that the pointwise model-based Bayesian and frequentist SEs both agree quite well with the empirical SEs. As expected, the frequentist SEs are smaller than the Bayesian SEs. All SEs decrease as the sample size increases.

Figure 5 compares the estimated 95% pointwise coverage probabilities of the frequentist and Bayesian confidence intervals of $f(t)$. Overall, the coverage probabilities using the Bayesian confidence intervals are slightly closer to the nominal value than those obtained from the frequentist confidence intervals for both $m = 34$ and $m = 68$. The means over time of the estimated Bayesian and frequentist coverage probabilities are 94% and 93% when $m = 34$ and 94.8% and 93.8% when $m = 68$. These results suggest that the Bayesian confidence intervals perform slightly better than their frequentist counterparts when the true $f(t)$ is a fixed smooth function. This may be due to the fact that the Bayesian SEs account for the bias in the MPLS $f$. Our simulation results are consistent with those of Wahba (1985) for independent data and suggest that the Bayesian confidence intervals have good sampling properties for a fixed smooth function $f(t)$.

7. DISCUSSION

In this article we have considered the inference for semiparametric stochastic mixed models for longitudinal data.
We used MPLLE to estimate the regression coefficients of the parametric fixed effects and to estimate the nonparametric function of time as a natural cubic spline. We used REML to estimate the smoothing parameter and the variance components simultaneously. A key feature of our approach is that it allows us to make systematic inference on all model parameters by representing a semiparametric model as a modified parametric linear mixed model.

Our simulation results show that the proposed method performs well in estimating the regression coefficients, the nonparametric function, and the variance components. Bayesian confidence intervals of the nonparametric function $f(t)$ appear to have slightly better coverage probabilities compared to their frequentist counterparts in the situation where the true $f(t)$ is a fixed smooth function, which is the situation of our primary interest. No practical difference is found in our simulation between the Bayesian and frequentist approaches in calculating the standard errors for the estimates of the regression coefficients.

Although our simulation study reveals the attractive performance of REML in estimating $f(t)$ and $\beta$, the asymptotic properties of REMLE still need to be investigated. This research requires studying consistency and asymptotic normality of $(f(t), \beta, \theta)$, which could be challenging and could require developing innovative statistical theory, as standard asymptotic by simply letting the number of subjects $m \to \infty$ might not be sufficient. This is because $m \to \infty$ does not guarantee that the number of distinct time points $r$ would also go to infinity. The latter is often required by consistency and asymptotic normality of $\hat{f}(t)$. To have $r \to \infty$, one might need to have the number of time points for each subject $n_i \to \infty$ if different subjects are observed at the same time points. It follows that consistency and asymptotic normality of $\hat{f}(t), \hat{\beta},$ and $\hat{\theta}$ require that both $m$ and $r$ (or $n_i$) go to infinity, and the normalizing constants for $\hat{f}(t), \hat{\beta},$ and $\hat{\theta}$ could well be different. Specifically, it is of potential interest to investigate whether the MPLLEs of the regression coefficients $\hat{\beta}$ and the REML estimators of the variance components $\hat{\theta}$ are $\sqrt{m}$ consistent and whether the conventional convergence rate $r^{-4/5}$ of the mean squared error of the MPLLE $\hat{f}$ still applies. The asymptotic could also depend on which of $m$ and $r$ (or $n_i$) goes to infinity first and which one goes there faster.

We have discussed using REML to estimate the smoothing parameter. Alternative approaches to selecting the smoothing parameter include cross-validation (CV) (Rice and Silverman 1991) and generalized cross-validation (GCV) (Wahba 1985). Note that GCV so far has not been well-defined for correlated data. Unlike REML, GCV poses the challenge of suffering the same problem as CV—that is, subsequent inference on the variance components could be difficult (Zeger and Diggle 1994). Further research is needed to compare the performance of the MPLLEs $(\hat{\beta}, \hat{f})$ using CV and GCV (to be defined) with that using REMLE.

In view of the fact that in longitudinal studies the variance–covariance of the outcome variable often varies over time (Donnelly et al. 1995; Taylor et al. 1994), we propose various stationary and nonstationary Gaussian stochastic processes to model the within-subject variability. Although in some cases the choice of the within-subject stochastic process does not have fundamental impact on the major inferences of interest, it could be crucial to the validity of statistical inference in other situations. Furthermore, a reasonable model for the subject-specific process would help investigators better understand the underlying biological process. In the presence of a nonstationary process, as in the progesterone data, the existing method for model diagnostics using a variogram (Diggle et al. 1994) is not applicable. More research is needed to develop new diagnostic tools to examine the appropriateness of the assumed covariance structure in such data.

When the variance changes over time in a longitudinal study, a question of interest is whether one should specify the time-varying variance in the random effects part or in the stochastic process part in model (1). Because the random effects model the between-subject variation, whereas the stochastic process models the within-subject variation, our first suggestion is that the decision could be made based on one's understanding of the underlying biological process, which could provide a guide on the causes of the change of the variance over time; that is, between or within-subject variation. Our second suggestion is to examine the data for each individual as a single time series to look for nonstationarity. For example, in the progesterone data, it is biologically plausible to assume that the variance changes with the mean progesterone level over time. This suggests that the time-varying variance may be due to the underlying biological process. Hence we modeled it using a NOU process in the stochastic process part of the model. This view is supported by examining each woman's data. Specifically, we found that for most women, the progesterone level is often stable during the first half of the cycle when the progesterone level is low, and becomes variable when the progesterone level increases during the second half of the cycle.

In the analysis of the progesterone data, we modeled the mean progesterone level as well as the serial correlation as a function of time measured in a standardized menstrual cycle. This implies that the serial correlation between consecutive days on the transformed time scale is the same for different women. In other words, we assume that, for example,
the correlation of two observations measured 1 day apart for a woman with a 25-day cycle is the same as the correlation of two observations measured 2 days apart for a woman with a 50-day cycle. This assumption may be reasonable if one believes the degree of similarity of two consecutive progesterone measures for a woman is driven by the time within a menstrual cycle relative to cycle length. An alternative approach suggested by one referee is to model the mean progesterone level on the transformed time scale and to model serial correlation on the original time scale. Because 30 out of 34 women had cycle lengths very close to the reference 28-day cycle (23–32 days) and only 4 women had long cycles (39–56 days), we expect that the standardization of the lengths of different women’s menstrual cycles has minimal impact on the inference. Because only very few women had long cycles in our data, we found it was difficult to determine which of the two scales was better for modeling serial correlation using the available data.

For simplicity, our model assumes a single nonparametric function and parametric covariate effects. One can easily extend our model and inference procedure to accommodate a more complicated mean structure, such as a model with an interaction between a covariate and a nonparametric function. For example, if we have a single covariate \( x \), then the interaction model can be written as

\[
y_{ij} = x_{ij} \beta + f(t_{ij}) + x_{ij} g(t_{ij}) + Z_{ij}^T \beta + U_i(t_{ij}) + \epsilon_{ij},
\]

(27)

where \( f(t) \) and \( g(t) \) are smooth nonparametric functions of times and \( g(t) \) is centered for the sake of identifiability. We can use a similar penalized likelihood approach to obtain natural cubic smoothing spline estimates of both \( f(t) \) and \( g(t) \). Specifically, from (10) we can write \( f \) and \( g \) as

\[
f = T \delta_1 + B \alpha_1 \quad \text{and} \quad g = t^0 \delta_2 + B \alpha_2,
\]

where \( t^0 \) centered about its mean and \( \alpha_1 \) and \( \alpha_2 \) are independent and follow normal(0, \( \tau_1 \)I) and normal(0, \( \tau_2 \)I). Plugging these expressions of \( f \) and \( g \) into (27), we can write (27) as a linear mixed model similar to (12). Estimation and inference on the model parameters can then proceed by applying the approach described in this article.

[Received December 1996. Revised October 1997.]

REFERENCES


