Published in final edited form as:

J Am Stat Assoc. 2016 ; 111(515): 1168–1181. doi:10.1080/01621459.2015.1076725.

Bayesian Nonparametric Longitudinal Data Analysis

Fernando A. Quintana,
Pontificia Universidad Católica de Chile, Santiago, Chile

Wesley O. Johnson,
University of California Irvine, USA

Elaine Waetjen, and
University of California Davis, USA

Ellen Gold
University of California Davis, USA

Abstract

Practical Bayesian nonparametric methods have been developed across a wide variety of contexts. Here, we develop a novel statistical model that generalizes standard mixed models for longitudinal data that include flexible mean functions as well as combined compound symmetry (CS) and autoregressive (AR) covariance structures. AR structure is often specified through the use of a Gaussian process (GP) with covariance functions that allow longitudinal data to be more correlated if they are observed closer in time than if they are observed farther apart. We allow for AR structure by considering a broader class of models that incorporates a Dirichlet Process Mixture (DPM) over the covariance parameters of the GP. We are able to take advantage of modern Bayesian statistical methods in making full predictive inferences and about characteristics of longitudinal profiles and their differences across covariate combinations. We also take advantage of the generality of our model, which provides for estimation of a variety of covariance structures. We observe that models that fail to incorporate CS or AR structure can result in very poor estimation of a covariance or correlation matrix. In our illustration using hormone data observed on women through the menopausal transition, biology dictates the use of a generalized family of sigmoid functions as a model for time trends across subpopulation categories.

Keywords

Bayesian Nonparametric; Covariance Estimation; Dirichlet Process Mixture; Gaussian process; Mixed Model; Ornstein-Uhlenbeck Process; Study of Women Across the Nation (SWAN)

1 Introduction

Longitudinal data are ubiquitous as evidenced by a plethora of recent books covering the topic (Hand and Crowder; 1996; Davidian and Giltinan; 1998; Verbeke and Molenberghs; 2000; Diggle, Heagerty, Liang and Zeger; 2002; Hardin and Hilbe; 2003; Hedeker and Gibbons; 2006; Fitzmaurice, Laird and Ware; 2011). Historically, continuous repeated measures data on the same individual were handled under the assumption of, say d, regularly spaced observation times and under an assumption of multivariate normality (see for

example Johnson and Wichern; 2007). Standard models included the assumption of common and either unstructured or structured covariance matrices for repeated observations on each individual. Traditional structured covariances arise by incorporating simple random effects.

Modern longitudinal data analysis makes no assumption about regular times of observation, and unstructured covariance matrices can result in difficulties due to having a large number of parameters relative to the sample size. Many statistical models and methods have thus been developed for handling irregular correlated data, e.g. Pullenayegum and Lim (2015).

Alternatively, models for dependent data have been developed (i) by including random and/or fixed functions in time, (ii) through the use of various forms of random effects, (iii) by using latent stochastic processes, or (iv) through a combination of functions and robust methods that accommodate without modeling covariance structure, among others (e.g. Laird and Ware; 1982; Zeger and Diggle; 1994; Shi, Weiss and Taylor; 1996; Zhang and Davidian; 2001; Li, Lin and Müller; 2010; He, Zhu and Fung; 2002). Here, we consider a combination of (i – iii), and then incorporate a nonparametric component that will add some degree of robustness to our ultimate model.

The first component involves the use of parametric or flexible functions in time, for example generalized sigmoid functions that are dictated by the biology in our illustration in Section 5, while other applications may require more flexible functions in time; data are allowed to have differing trends in distinct subpopulations. Random effects for individuals can be modeled with standard parametric distributions, or more flexibly with parametric or nonparametric mixtures (Li et al.; 2010). Finally, we incorporate a stochastic process component and its generalization to allow for more general covariance structure in time. Each of these model components induces its own dependence structure and differing specific choices that are made for inclusion in the final model result in distinct covariance structures for the observed data.

Another approach offers the possibility of directly modeling the covariance structure. Rather than incorporating a stochastic process, for example, Pourahmadi (1999), Daniels and Kass (2001), Smith and Kohn (2002) and Daniels and Pourahmadi (2002) directly model the covariance matrix for individual responses by reparametrizing the matrix using decompositions. Some of these models allow for the incorporation of (time dependent) covariates, including time itself, in the covariance structure and thus allow for non-homogenous structure in time. Some of them allow for shrinking towards some form of known structure, like AR structure or independence.

The main goal of this paper is to extend the general class of mixed models for longitudinal data by generalizing the GP part to be nonparametric, which to our knowledge has not been done. A standard GP model has exponential covariance structure, depends on only two parameters, and results in AR correlation of order one. Our premise is that this structure is overly simple for many data sets and that generalizing it will result in improvements under many circumstances. To this end, we implement a Dirichlet Process mixture model by mixing on the parameters of the covariance function of the process, resulting in a DPM of processes.

We present background material in section 2 and we develop our model and discuss inferences in section 3. In section 4 we carry out a simulation study that suggests that the proposed semi-parametric model can be naturally used as a method to estimate covariance and correlation matrices, which outperforms some traditional competing alternatives. Section 5 considers longitudinal follicle stimulating hormone (FSH) data as women pass through the menopausal transition (Waetjen et al.; 2011). Here, it is believed that the data will generally follow sigmoidal curves rising from lower values to higher values according to this shape. We build on this and analyze the data using a model for longitudinal FSH profiles with a 5-parameter generalization that allows for some departures from sigmoidal shape, in addition to the flexible proposed correlation structure. Comparison with alternative models are also provided. Conclusions are presented in section 6. Some technical and Monte Carlo simulation details are presented in Appendix A.

2 Background Material

In this section, we concentrate on providing an appropriate context for our proposed model, which will include background material for a number of other models and methods from the literature. We also discuss the Dirichlet Process and the Dirichlet Process Mixture.

2.1 Structured Covariances: Unequally Spaced Times of Observation

Assume that observations are made on individual i at times $\{t_{i1}, \ldots, t_{in_i}\}$, namely $\mathbf{y}_i = \{y_{ij} : j = 1, \ldots, n_i\}'$. Thus $\mathbf{y}_i$ is $n_i \times 1$. At time t_{ij} we allow for a vector of possibly time dependent covariates $\mathbf{x}'_{ij} = (1, x_{i1}(t_{ij}), \ldots, x_{ip}(t_{ij}))$, and assume that $E(y_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta}$. Define the $n_i \times (p + 1)$ design matrix $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})'$, leading to an assumed mean vector $E(\mathbf{y}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$. Then allow for a corresponding $n_i \times q$ design matrix $\mathbf{Z}_i$, with $q \leq p$ and with the column space of $\mathbf{Z}_i$ restricted to be contained in the column space of $\mathbf{X}_i$. A traditional model for longitudinal data formulated with these ingredients is the general linear mixed model of Laird and Ware (1982).

Since we model trends in time, let $f_i(t)$ be a particular function of time for individual i. This could be a parametric form that is allowed to differ according to particular characteristics of individuals, as in the illustration presented in Section 5 where we analyze hormone data collected on women who are going through menopause. We consider generalized sigmoidal functions of time that are allowed to be different for eight distinct types of women (varying according to age and race/ethnicity; see (5) and Figure 3). Other circumstances will require alternative parametric non-linear functions, for example with growth curve data (Davidian and Giltinan; 1998). In general, however, we would allow for the possibility of flexible (smooth) functions that can be approximated by various types of basis functions with unknown (possibly random) coefficients. See for example Shi et al. (1996), Zhang and Davidian (2001), He et al. (2002), and Li et al. (2010).

Our model will be based on a well known generalization of the linear mixed model (Diggle; 1988) that also allows for AR structure, namely

$$\boldsymbol{y}_i=\boldsymbol{\mu}_i+\boldsymbol{Z}_i\boldsymbol{b}_i+\boldsymbol{w}_i+\boldsymbol{\varepsilon}_i,\quad \boldsymbol{b}_i|\boldsymbol{\xi}\sim N_r\left(0,\boldsymbol{D}\left(\boldsymbol{\xi}\right)\right),\quad \boldsymbol{w}_i|\phi\sim N_{n_i}\left(0,\boldsymbol{H}_i\left(\phi\right)\right), \quad (1)$$

where $\boldsymbol{H}_i(\boldsymbol{\phi})$ is $n_i \times n_i$ and has a structural form, and where $\varepsilon_i \sim N\left(0,\ \sigma^2\boldsymbol{I}_{n_i}\right)$. The vectors $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ contain variance-covariance parameters for $\boldsymbol{b}_i$ and $\boldsymbol{w}_i$, respectively. In situations where a specific functional specification does not fit the data well, the fixed and/or random effects parts of the model may be augmented to include basis functions such as splines.

The vectors $\boldsymbol{w}_i$ are generated by mean zero Gaussian stochastic processes, $\{w_i(t) : t > 0\}$. If $\mathrm{Cov}(w_i(t+s),\ w_i(t))=\sigma^2_w\rho(s)$, with $\rho(s) = \rho^s$, the resulting stationary process is an Ornstein-Uhlenbeck process (OUP), (Rasmussen and Williams; 2006). The OUP gives an exponential covariance function and induces AR structure. Zeger and Diggle (1994) used $\rho(s) = \alpha + (1 - \alpha)\rho^s$. There are additional choices, including the possibility that σ^2_w could depend on t, resulting in a nonhomogeneous OUP (Zhang et al.; 1998). Taylor et al. (1994) used an integrated OUP (integrating over an OUP with exponential covariance function) that results in a covariance function that depends on both t and s. With structured covariance functions, the marginal covariance matrix for $\boldsymbol{y}_i$ is $\mathrm{Cov}(\boldsymbol{y}_i)=\sum_i\left(\boldsymbol{\xi},\phi,\sigma^2\right)=\boldsymbol{Z}_i\boldsymbol{D}(\boldsymbol{\xi})\boldsymbol{Z}_i^{'}+\boldsymbol{H}_i(\phi)+\sigma^2 I_{n_i}$. Diggle and Verbyla (1998) developed an alternative nonparametric estimator of the covariance function, based on sample variogram ordinates and squared residuals.

Model (1) allows for functional components and thus is quite general. But (1) can be even more general, namely, the model for the b_i's can be flexible as well. In the context of mixed models, Verbeke and Lesaffre (1996); Zhang and Davidian (2001); Kleinman and Ibrahim (1998a,b); Mukhopadhyay and Gelfand (1997); Müller and Rosner (1997) all developed flexible models for the random effects. More recently, Li, Lin and Müller (2010) developed a Bayesian semiparametric approach to model (1), without the GP, that allowed for flexibility in both the functional part, using smoothing splines, and using a DP model for the random effects.

The combination of choosing which terms to include in (1), and making particular choices for $\boldsymbol{H}(\boldsymbol{\phi})$ and $\boldsymbol{D}(\boldsymbol{\xi})$, when the corresponding effects are included in the model, determines the covariance structure for the data. There are obviously many choices. Using the Bayesian approach, there is also the issue of selecting prior distributions for $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$, as well as other parameters, resulting in yet more choices.

2.2 Reparametrizing Covariances by Decomposition

An alternative approach to modeling longitudinal data involves modeling the covariance matrices for the $\boldsymbol{y}_i$'s (Pourahmadi; 1999; Pourahmadi and Daniels; 2002; Daniels and Pourahmadi; 2002; Smith and Kohn; 2002). In all of these approaches the authors take advantage of the Cholesky factorization of a covariance matrix, namely that, for a given covariance matrix Σ, there exists a lower triangular $\boldsymbol{T}$ with ones on the diagonal, and diagonal $\boldsymbol{\lambda}$ such that $\boldsymbol{T}\Sigma\boldsymbol{T}^{'} = \boldsymbol{\lambda}$. The components of $\boldsymbol{\lambda}$ are only restricted to be positive, and

the off diagonal elements of $\boldsymbol{T}$ are unrestricted. Smith and Kohn (2002) simplify this structure resulting in greater parsimony. Pourahmadi (1999) seeks parsimony by modeling all parameters as functions of covariates (including time), thus reducing the number of parameters.

A major issue for Bayesian analysis in this setting is the choice of prior distribution for the covariance matrix. In the unstructured case, the inverse Wishart has been a standard choice. The historically standard so-called noninformative prior is the Jeffreys (1961) prior, while Yang and Berger (1994) developed a reference prior that results in considerably improved frequentist properties at the expense of increased computational effort. Leonard and Hsu (1992), Daniels and Kass (2001), Daniels and Pourahmadi (2002) and Pourahmadi and Daniels (2002) provide further Bayesian developments. For an approach that focuses on the correlation matrix, see Liechty et al. (2004).

2.3 The Dirichlet Process and Dirichlet Process Mixture

The DP is most easily characterized by the Sethuraman (1994) construction of it. Let G_0 be a known distribution and let $M > 0$ be a positive constant. Then we say $G \sim DP(G_0, M)$ provided

$$G(\cdot) = \sum_{k=1}^{\infty} \omega_k \delta_{v_k}(\cdot), \quad v_k \overset{\text{iid}}{\sim} G_0,$$

where $\delta_v(\cdot)$ defines point mass at v and where $\omega_k = V_k \prod_{r=1}^{k-1}(1 - V_r)$ with $V_r \overset{\text{iid}}{\sim} \text{Beta}(1, M)$. Thus G is a random distribution that is discrete with probability one. G_0 is called the base or centering distribution since $E(G) = G_0$. For any appropriate set A, $G(A) \sim \text{Beta}(M G_0(A), M G(A^c))$, and thus $G(A) \xrightarrow{p} G_0(A)$ as $M \to \infty$. For small M, there is much more flexibility than for large M.

The DPM takes advantage of the discreteness of the DP. Consider a parametric density function that depends on parameters $\boldsymbol{\nu}$, $f(\cdot|\boldsymbol{\nu})$, and let $\boldsymbol{\nu}|G \sim G$, $G \sim DP(G_0, M)$. We obtain the DP mixture $f(\cdot|G) = \int f(\cdot|\boldsymbol{\nu}) G(d\boldsymbol{\nu}) = \sum_{k=1}^{\infty} \omega_k f(\cdot/v_k)$, using the Sethuraman (1994) construction.

3 General Model and Inference

In this section, we build on the generic model (1) by generalizing the GP part using DP mixtures of Gaussian processes. The $\boldsymbol{b}_i$'s are modeled with parametric distributions to avoid identifiability problems. We first develop the DPM of GPs, then we discuss the full model and finally, posterior inference, which is based on Markov chain Monte Carlo (MCMC) approximation to the joint posterior. Details of its implementation are given in the Appendix A.

3.1 DPM of GPs

Our extension of model (1) aims at introducing flexibility beyond the exponential covariance structure implied by the OUP, which was discussed in the next paragraph after (1). Consider first the GP, $\boldsymbol{w}_i$, for the ith subject, with covariance matrix of the form $\boldsymbol{H}_i(\phi) = \sigma_\omega^2 \tilde{\boldsymbol{H}}_i(\rho)$, where $\phi = \left(\sigma_\omega^2, \rho\right)$ and $\left\{\tilde{\boldsymbol{H}}_i(\rho)\right\}_{k,\ell} = \rho^{|t_{i\ell} - t_{ik}|}$. Recall that $(t_{i1}, \ldots, t_{in_i})$ are the times at which observations $\boldsymbol{y}_i = \{y_{ij} : j = 1, \ldots, n_i\}'$ for the ith subject are made. We model $\boldsymbol{\phi}_i | G \sim G$ with $G \sim DP(G_0, M)$ so that

$$f(\boldsymbol{\omega}_i / G) = \int N(\boldsymbol{\omega}_i | 0, \sigma_\omega^2 \tilde{\boldsymbol{H}}_i(\rho)) dG(\phi) = \sum_{k=1}^{\infty} \omega_k N_{n_i}(\boldsymbol{\omega}_i | 0, \tilde{\sigma}_{\omega k}^2 \tilde{\boldsymbol{H}}_i(\tilde{\rho}_k)), \tag{2}$$

an infinite mixture of multivariate normal densities, where $\left(\tilde{\sigma}_{\omega k}^2, \tilde{\rho} k\right) \overset{\text{iid}}{\sim} G_0$, and the ω_k's were defined in subsection 2.3. The base distribution G_0 will be discussed below. A related spatial DP with exponential covariance function in the base distribution was developed in Gelfand et al. (2005).

Model (2) implies clustering on autocorrelation structure across subjects. Moreover, using the Sethuraman construction above, we have $\operatorname{Cov}(w_i(t), w_i(t+s) | G) = \sum_{k=1}^{\infty} \omega_k \tilde{\sigma}_{wk}^2 \tilde{\rho}_k^s$. Thus, if the ith subject has equally spaced times between observations, the corresponding covariance matrix has homogeneous diagonals with decreasing values for $s = 1, 2, \ldots$, resulting in a clear generalization of simple AR(1). This hypothetical individual would have a covariance matrix for their $\boldsymbol{w}_i$ that is an example of a Toeplitz matrix (Ray; 1970). However, individuals in data with irregular times of observation would not.

As usual in DPM models, it is useful to break the mixture (2) by introducing latent parameters $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n$ so that the model can be stated hierarchically. We now extend (1) using (2) and propose the following hierarchical model for the entire data set $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$:

$$\begin{aligned}
\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{w}_i, \sigma^2 &\overset{\text{ind}}{\sim} N_{n_i}(\boldsymbol{\mu}_i + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{w}_i, \sigma^2 \boldsymbol{I}) \\
\boldsymbol{w}_i | \phi_i = (\sigma_{wi}^2, \rho_i) &\overset{\text{ind}}{\sim} N_{n_i}\left(0, \sigma_{wi}^2 \tilde{\boldsymbol{H}}_i(\rho_i)\right) \\
\phi_1, \ldots, \phi_n | G &\overset{\text{iid}}{\sim} G \\
G &\sim DP(G_0, M) \\
\boldsymbol{b}_i &\overset{\text{iid}}{\sim} N(0, \boldsymbol{D}(\boldsymbol{\xi})) \\
\sigma, \boldsymbol{\beta}, \boldsymbol{\xi} &\sim U(0, A) \times N(\boldsymbol{\beta}_0, \boldsymbol{B}) \times p(\boldsymbol{\xi}),
\end{aligned} \tag{3}$$

where $\boldsymbol{\mu}_i = \boldsymbol{X}_i \boldsymbol{\beta}$ and $\boldsymbol{w}_i$ and $\boldsymbol{b}_i$ are assumed independent for all $1 \leq i \leq n$. This model thus extends the exponential covariance function of the OUP as noted earlier by allowing individual structures that can be clustered across subjects through the $\boldsymbol{\phi}_i$ parameters. From here on, we will often refer to model (3) as the nonparametric autoregressive model (NPAR).

One advantage of the OUP structure built in model (3) is that $\tilde{\boldsymbol{H}}_i(\rho_i)$ can be analytically inverted and is a tridiagonal matrix, which follows by considering results available from the theory of antedependent random variables (Gabriel; 1962; Zimmerman and Núñez-Antón; 2010). Details are given in the Appendix B. In particular, the problem of evaluating $f(\boldsymbol{w}_i|\boldsymbol{\phi}_i)$ is further simplified by using this form for the inverse covariance matrix to show that if $\boldsymbol{w}_i \sim N_{n_i}(0, \sigma^2_{wi}\tilde{\boldsymbol{H}}(\rho_i))$, then letting $r_{i,k}=\rho_i^{|t_{i,k+1}-t_{ik}}$ for $k = 1, \ldots, n_i - 1$

$$w_{i1}\sim N_1(0,\sigma_i^2),\quad w_{i1}|w_{i1}=\tilde{w}_1,\ldots,w_{ik-1}=\tilde{w}_{k-1}\sim N_1(\tilde{w}_{k-1}r_{ik-1},\sigma_i^2(1-r_{ik-1}^2)). \quad (4)$$

Thus we obtain $f(\boldsymbol{w}_i|\boldsymbol{\phi}_i)$ as the product of n_i univariate normal probability densities, which makes it very simple to obtain the full conditional distribution of $\boldsymbol{w}_i$ in the Gibbs sampling algorithm to be discussed later. Without these simplifications, the MCMC method would involve the inversion of potentially n distinct covariance matrices at each iteration. This is expensive computation that is saved by having explicit forms for the inverses, resulting in the simplification of model specification (3).

3.2 Marginal Augmented Data Likelihood and Prior Specification

Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)$, $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)$ and define $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\xi}, \boldsymbol{\phi})$. As usual in DP-related models, inference involves marginalizing the random G to obtain the marginal prior distribution for $\boldsymbol{\phi}$, which has the well known Polya' Urn representation (Blackwell and MacQueen; 1973; Escobar; 1994)

$$\phi_1\sim G_0,\quad \phi_k|(\phi_1,\ldots,\phi_{k-1})\sim\frac{MG_0+\sum_{\ell=1}^{k-1}\delta_{\phi_\ell}}{M+k-1}.$$

Consider now the augmented data likelihood arising from model (3)

$$L(\boldsymbol{\beta},\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{b},\boldsymbol{w})=f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{w},\boldsymbol{b},\sigma^2)f(\boldsymbol{b}|\boldsymbol{\xi})f(\boldsymbol{w}|\phi),$$

which we combine with $p(\sigma^2, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\phi}) = p(\sigma^2)p(\boldsymbol{\beta})p(\boldsymbol{\xi})p(\boldsymbol{\phi})$, the marginal prior distribution, where we assume a priori independence of the four components. The corresponding joint posterior is then

$$\begin{aligned} p(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{w}|\boldsymbol{y}) &\propto \left\{\prod_{i=1}^n f(\boldsymbol{y}_i,\boldsymbol{b}_i,\boldsymbol{w}_i|\boldsymbol{\beta},\boldsymbol{\theta})\right\} p(\sigma^2)p(\boldsymbol{\beta})p(\boldsymbol{\xi})p(\phi) \\ &= \left\{\prod_{i=1}^n f(\boldsymbol{y}_i|\boldsymbol{\beta},\boldsymbol{w}_i,\boldsymbol{b}_i,\sigma^2)f(\boldsymbol{b}_i|\boldsymbol{\xi})f(\boldsymbol{w}_i|\phi_i)\right\} p(\sigma^2)p(\boldsymbol{\beta})p(\boldsymbol{\xi})p(\phi). \end{aligned}$$

The first term in the product of the augmented data likelihood is the $N_{ni}(\boldsymbol{\mu}_i + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{w}_i, \sigma^2\boldsymbol{I})$ pdf evaluated at $\boldsymbol{y}_i$. The second term is the $N_l(0, \boldsymbol{D}(\boldsymbol{\xi}))$ pdf evaluated at $\boldsymbol{b}_i$. The third term is obtained as the product of the normal densities given in (4). We assume a proper prior for $\boldsymbol{\xi}$ and we introduce a precise form when we get to the examples. For example, if the b_i's are

scalars corresponding to random intercepts, then $\xi \equiv \sigma_b^2$, and we would assign an inverse gamma distribution to it. Our choice for G_0 is the cross product of a $U(0, 1)$ distribution (corresponding to the ρ_i's) and a $U(0, C)$ distribution (corresponding to the standard deviations, σ_i's). We consider specific cases of the NPAR model (3) in the sequel.

3.3 Posterior Inferences

We approximate the joint posterior using Gibbs sampling. We require the full conditional distributions

$$p(\boldsymbol{\beta}|\text{else})\ \ p(\sigma|\text{else})\ \ p(\boldsymbol{\xi}|\text{else})\ \ p(\boldsymbol{b}_i|\text{else})\ \ p(\boldsymbol{w}_i|\text{else})\ \ p(\phi|\text{else}).$$

Most of these are either standard distributions or can be handled with standard MCMC methods. The full conditional for $\boldsymbol{\phi}$ is also non standard, but many algorithms have been developed for handling this part. Neal (2000) summarizes them, and also develops some new algorithms, in conjugate and non-conjugate cases. In our non-conjugate case, we apply Neal's algorithm number 8. Details are given in Appendix A.

4 Estimating Covariance Structure

Since our paper emphasizes the importance of modeling/accommodating dependence in longitudinal data, we now explore how well our model works for estimating different covariance structures. For simplicity, we consider a particular version of our model (3). Covariance estimation is briefly discussed in subsection 4.1, and simulation based comparisons are given is subsection 4.2.

4.1 Methods

Consider the simplified version of the NPAR model (3)

$$\boldsymbol{y}_i = (\mu + b_i)e + \boldsymbol{w}_i + \boldsymbol{\varepsilon}_i,$$

with no covariates, where e is a vector of ones, and $b_i \sim N(0, \sigma_b^2)$. This model features a fixed overall intercept, and random intercepts and DPMs of OUPs for sampling units. We let $n_i = k$ for all i, and assume $t_{ij} = t_j$, namely that all individuals are observed at the same times. Under these assumptions, there is a common covariance matrix, say Σ, to estimate. There are of course other possible choices of NPAR depending on the circumstances. We are mainly interested in exhibiting the robustness of this choice to a number of possible covariance choices that would not be in the same form as the parametric version of this model. The posterior mean,

$$\hat{\sum}_{NPAR} = E\left\{\sigma_b^2 \boldsymbol{J} + \sigma_w^2 \boldsymbol{H}(\rho) + \sigma^2 \boldsymbol{I} | \text{data}\right\},$$

where $\boldsymbol{J}$ is a matrix of ones, is a natural estimate of Σ.

For comparison, we consider two Bayesian approaches to the simple unstructured model $\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. The straw man approach involves placing an inverse Wishart prior distribution on Σ (Inv-Wishart(13, diag{100, 11})). We also consider the computationally and conceptually nice prior of Daniels and Pourahmadi (2002), which was discussed in subsection 2.2.

We now introduce three common loss functions, which result in three distinct estimators that have been used for the estimation of general covariances:

$$\hat{\sum}_1 = E(\sum | \text{data}), \quad \hat{\sum}_2 = E(\sum^{-1} | \text{data})^{-1},$$
$$\text{vec}(\hat{\sum}_3) = \left\{ E(\sum^{-1} \otimes \sum^{-1} | \text{data} \right\}^{-1} \text{vec} \left\{ E(\sum^{-1} | \text{data}) \right\},$$

where the first estimator minimizes mean squared error loss, the second estimator minimizes the loss function

$$L_1(\hat{\sum}, \sum) = tr(\hat{\sum} \sum^{-1}) - log(|\hat{\sum} \sum^{-1}|) - dim(\sum)$$

and the third minimizes

$$L_2(\hat{\sum}, \sum) = tr((\hat{\sum} \sum^{-1} - \boldsymbol{I})^2)$$

(Yang and Berger; 1994).

In the next subsection, we make comparisons based on simulated data. While we only use $\hat{\sum}_{NPAR}$, regardless of the choice of loss function, when we use the other two models, we apply the estimator that is optimal for the choice of loss function.

4.2 Simulations

We simulated data from compound symmetry (CS), AR, mixture of CS and AR and from a non-standard structure. We generated 100 data sets each with 100 sequences of length 11, with mean 0 and covariance matrices as follows:

$$C_1(i,j) = 10I(i=j) + 7I(i \neq j) \quad \text{(CS)}$$
$$C_2(i,j) = 10 \times 0.4^{|i-j|} \quad \text{(AR(1))}$$
$$C_3(i,j) = 0.3C_1(i,j) + 0.7C_2(i,j) \quad \text{(mixture of CS and AR(1))}$$
$$C_4(i,j) = 10/\sqrt{1+|i-j|}.$$

For each estimator and each of the four types of data, we obtained the root mean squared error (RMSE); results are given in Tables 1 and 2 for covariance and correlation matrices, respectively. The RMSE is the square root of the average squared error. Namely, let Λ be a covariance or correlation matrix, and define λ to be the corresponding vector of unique parameter values in Λ. Let $\hat{\lambda}_i$ be the vector of corresponding estimates based on simulated data set *i*. Then

$$\text{RMSE}=\sqrt{\frac{1}{100}\sum_{i=1}^{100}\|\hat{\lambda}_i-\lambda\|^2/dim(\lambda)},$$

where $\|\cdot\|$ denotes Euclidean distance.

Comparisons among models should be made for each loss function and covariance model combination. When estimating these particular covariance structures, the NPAR model appears to dominate the others across the board. The Daniels and Pourahmadi (2002) model dominates the inverse Wishart for loss functions L_1 and L_2. When estimating correlation matrices, all models seem roughly equivalent on RMSE except for the relatively poor performance of NPAR under strictly AR structure. With the other two loss functions, the NPAR model appears to dominate while the Daniels and Pourahmadi (2002) method dominates the inverse Wishart.

We are not surprised that our estimator, which was designed to specifically model longitudinal correlation structure, could be superior to estimators that were designed to be optimal according to various loss functions that were not specifically intended for use with longitudinal data. It is worth pointing out that this is obviously a very small simulation study.

5 Hormone Data

In this section, we analyze hormone data that were collected on women during the menopausal transition. The data are described in subsection 5.1. We consider comparisons with several other models that are described in subsection 5.2. The data are analyzed in subsection 5.3.

5.1 The Data

We consider a small subset of data that were obtained from the Study of Women Across the Nation (SWAN), which is a multicenter prospective cohort study of women from five racial/ethnic groups at seven sites who have been followed to characterize the menopausal transition. The eligibility criteria for the SWAN cohort at baseline were: age of 42 to 52 years; self-identification as one of the five racial/ethnic groups (African American, Caucasian, Chinese, Japanese, Hispanic); and pre- or early peri-menopausal. The exclusion criteria included: (i) no menstrual period in more than 3 months before enrollment; (ii) having had a hysterectomy and/or bilateral oophorectomy before enrollment; and (iii) current lactation, or exogenous hormone use. Moreover, the women in this subsample must have completed transition to post-menopause, must not have used hormones during this 9-year period (one year of "baseline" data and 8 years of followup), must have complete serum hormone data for all follow-up years baseline through visit 9, and all serum hormone values during pre- and early peri-menopause must have been in a time window of days two through five of the menstrual cycle. The reason for these restrictions is to have the purest and most complete sample in terms of follow-up hormone information. An expanded description of

the study and of data related to ours is given in Waetjen et al. (2011). Additional details can be found at the SWAN web site (www.swanstudy.org/).

Our main interest here is to model the annual follicle stimulating hormone (FSH) concentrations through the menopausal transition with the goal of finding a model that fits the data well. Concentrations of FSH and other hormones have been modeled to increase according to a (four parameter) sigmoidal shape (Dennerstein et al.; 2007). SWAN measured FSH and other hormone concentrations from serum samples annually in days two through five of the menstrual cycle for women who were still menstruating or on any day that women came in for their annual visit if they were postmenopausal. Times of observation were centered on the year of final menstrual period (FMP), namely $t_i = 0$ corresponds to the year in which the final menses occurred, which is defined to be the actual time of last menses before a 12 month period in which there were none. Thus year -3 is 3 years prior to the FMP, and year +3 is three years after. The data included women who started at year −8 continuing through year 0, and women starting at year −2 and continuing through year 6 (after FMP). Covariates considered to affect the overall levels of concentration were: (i) race/ethnicity (Hispanics were excluded due to retention problems at the one site that recruited Hispanic participants), and (ii) age at the beginning of the study, dichotomized as (≤46/> 46) years. Our subset of the data included 9 observations for each of 162 women, and contained no missing observations. Methods for handling missing data in longitudinal studies are discussed at great length in Daniels and Hogan (2008).

In Appendix C, we show plots of the actual data and their averages corresponding to the age by race-ethnicity categories. From these plots it is clear that a simple sigmoid (“S”) shape could be inadequate for modeling curve shapes. We thus consider more general shapes in the next section.

Our data are similar to a subset of the data in Waetjen et al. (2011) in which the goal was to ascertain any relationship between serum estradiol (E2) and incidence of urinary incontinence symptoms during the menopausal transition. Our data included an extra year of data beyond that analysis, but do not include other covariates considered there, nor do they include information about incontinence. As a followup to this study, it will be possible to consider joint modeling of the hormone profile and the time to event, such as first onset of incontinence. Many instances of joint modeling occur in the literature, for example see Tsiatis and Davidian (2004) for an overview of frequentist approaches or Hanson, Branscum and Johnson (2011) for semiparametric Bayesian approaches.

5.2 Models for Comparison

We modeled the FSH concentrations, $\{\boldsymbol{y}_i : i = 1, \ldots, 162, n_i = 9\}$, in several ways. For all models, we allowed for distinct fixed functional effects, in time, that were either linear or “generalized sigmoidal” for each of the eight combinations of age and race/ethnicity. In addition we considered the possibility of adding (i) a DPM of OU processes, (ii) simple random effects, and/or (iii) a particular DPM of parametric random effects distributions, which is a Bayesian nonparametric generalization of random effects. The way we modeled functional fixed effects for age and race/ethnicity was to allow all coefficients for the linear, or sigmoidal, forms to be different depending on the particular combination of covariates.

Let $c(i) \in \{1, 2, \ldots, 8\}$ be an indicator variable describing the particular combination of four races and two ages corresponding to subject i. Here, we set $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_8')$, where $\boldsymbol{\beta}_\ell$ is the vector of fixed parameters associated with combination ℓ. We also make use of the five parameter generalized sigmoid curve that was discussed in Ricketts and Head (1999):

$$S(t|\boldsymbol{\beta}) = \beta_1 + \frac{\beta_2}{1 + f_t exp\{\beta_3(\beta_4 - t)\} + (1 - f_t) exp\{\beta_5(\beta_4 - t)\}}, \quad (5)$$

where

$$f_t = \frac{1}{1 + exp\{-C(\beta_4 - t)\}} \quad \text{and} \quad C \frac{2\beta_3\beta_5}{|\beta_3 + \beta_5|},$$

in which case the fixed effects become $\mu_{ij} = S(t_{ij}|\boldsymbol{\beta}_{c(i)})$. The parameter vectors are now five-dimensional and the curves defined by (5) are not restricted to be monotone, as would be the case of a pure sigmoidal curve. However, if β_3 and β_5 are both positive, then (5) is monotone and increasing, and if both are negative, then it is decreasing. For the moment we focus on the both positive case. We have that β_1 is the lower asymptote, and $\beta_1 + \beta_2$ is the upper asymptote. Setting $t = \beta_4$, we see that this is the point at which the curve is half way between asymptotes. In the special case where $\beta_3 = \beta_5 > 0$, it follows that the slope of the sigmoid at the midpoint is higher for larger $\beta_3 = \beta_5$. In the general case we can see that as t becomes small, f_t tends to zero, and as t becomes large, f_t tends to one. So for small values of t, relative to β_4, the rate of increase of the 5 parameter sigmoid tends to be governed by the part of the function with β_5, and for large values, relative to β_4, it tends to be governed by the part with β_3. Thus the function is allowed to increase at different rates to the left and to the right of β_4. Similar statements can be made if β_3 and β_5 are both negative. If one of these is negative and the other positive, (4) need not be monotone as will be seen in our data analysis where this is the case.

Using a model with fixed effects specified through (5) we can compare the estimated mean profiles for the eight groups when these curves are fitted, and we can also use the model for prediction of future or missing values. We have selected (5) for biological reasons, but could have used a variety of flexible forms to capture time trends.

We consider four specific models and two simplified versions:

I. Sigmoid mean with random intercept plus OUP:

$$\boldsymbol{y}_i = S(\boldsymbol{t}_i|\boldsymbol{\beta}_{c(i)}) + b_i\boldsymbol{e} + \boldsymbol{w}_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{t}_i' = (t_{i1}, \ldots, t_{i9})$, $b_i \overset{\text{ind}}{\sim} N(0, \sigma_b^2)$ are individual-specific random effects, $\boldsymbol{e}$ is a vector of ones, $\boldsymbol{w}_i$ is distributed as a DPM of OUPs as specified in (3), and where $S(\boldsymbol{t}_i|\boldsymbol{\beta}_{c(i)})$ is a vector with entries $S(t_{ij}|\boldsymbol{\beta}_{c(i)})$ for $j \in \{1, \ldots, 9\}$.

II. Linear mean and linear random effects plus OUP:

$$\boldsymbol{y}_i = \boldsymbol{\beta}_{c(i),0}\boldsymbol{e} + \boldsymbol{\beta}_{c(i),1}\boldsymbol{t}_i + b_{i0}\boldsymbol{e} + b_{i1}\boldsymbol{t}_i + \boldsymbol{w}_i + \boldsymbol{\varepsilon}_i,$$

with $b_{ir} \overset{\text{ind}}{\sim} N(0, \sigma^2_{b,r})$, $r = 0, 1$.

III. Sigmoid mean, simple random effects and no OUP:

$$\boldsymbol{y}_i = S(\boldsymbol{t}_i|\boldsymbol{\beta}_{c(i)}) + b_i\boldsymbol{e} + \boldsymbol{\varepsilon}_i,$$

where $b_i \sim N(0, \sigma^2_b)$ are individual-specific random effects. This is a *parametric* model.

IV. Sigmoid mean, DPM for random effects and no OUP: This model is a generalization of Model III that requires some explanation. It is more general due to use of a nonparametric structure for the random effects distributions, which are modeled exchangeably for the eight categories through the use of a particular choice of DPM. The model correlates these eight distributions in the sense of MacEachern (1999, 2000) and De Iorio et al. (2004, 2009), who have defined and developed various Dependent Dirichlet Processe (DDP) models. Our model for the random effects will be a particular DDP. We emphasize that no longitudinal correlation structure is built into this model. The model is described briefly below, and then discussed more carefully in Appendix D.

$$\boldsymbol{y}_i = S(\boldsymbol{t}_i|\boldsymbol{\beta}_{c(i)}) + b_i\boldsymbol{e} + \boldsymbol{\varepsilon}_i,$$

where $b_i = z'_{c(i)}\gamma_i + \eta_i$, where $z_{c(i)}$ is a vector with seven zeros and one one in the slot that corresponds to the age by race-ethnicity group $c(i)$, and $\boldsymbol{\gamma}_i$ is 8×1. Then let $\gamma_i|G \overset{\text{iid}}{\sim} G$ with $G \sim DP(M, G_0)$, and $\eta_i \sim N(0, s^2)$, so we have a DPM of ANOVA models for b_i. Specify the centering measure to be $G_0 = N(\boldsymbol{\lambda}, \Psi)$, with $\boldsymbol{\lambda}|\Psi \sim N(\boldsymbol{\lambda}_0, \Psi)$, and $\Psi \sim IW(\nu_\ell \Psi_0)$. By analogy with (2), the induced random-effects distribution can be expressed as $f(b_i|c(i), G) = \int N(b_i; z'_{c(i)}\gamma, s^2)dG(\gamma) = \sum_\ell^\infty \omega_\ell N(b_i; z'_{c(i)}\tilde{\gamma}_\ell, s^2)$, where $\{\tilde{\gamma}_\ell\} \overset{\text{iid}}{\sim} N(\boldsymbol{\gamma}, \Psi)$ and the weights $\{\omega_\ell\}$ are as in subsection 2.3. A standard choice for M is one, which we adopt. The parameters $\boldsymbol{\lambda}$ and Ψ add flexibility to the model, but are not explicit objects of interest. They are easily handled in the MCMC approximation to the joint posterior.

V. Same as I but with $\rho_i = 0$ for all i: The DP mixture of OUPs is based on the σ_i^2 parameters only. All the r_{ik} quantities from (4) are zero and the w_{ik} Gaussian variables are iid zero mean in this case.

VI. Same as II but with $\rho_i = 0$ for all i: The DP mixture of OUPs is based on the σ_i^2 parameters only. As in model V, all the r_{ik} quantities are zero and all the w_{ik} Gaussian variables are iid zero mean.

Models I and II are NPAR models, while III and IV combine some of the model features we have discussed so far and thus constitute interesting alternatives for the purpose of comparing results. Models V and VI are simplifications of I and II, respectively.

5.3 Data Analysis

We first fitted models (I–VI) from subsection 5.2 and calculated log pseudo-marginal likelihood (LPML) statistics for each model; see Christensen et al. (2010), Section 4.9.2, or Gelfand and Dey (1994). This criterion for model selection was first introduced by Geisser and Eddy (1979) and has been used extensively for model selection in recent years; see for example Hanson, Branscum and Johnson (2011).

Using generic notation, let $\mathcal{M}$ denote a model, and for an independent sample $\{y_1, \ldots, y_n\}$ with $y_i|\theta, \mathcal{M} \sim f_i(\cdot|\theta, \mathcal{M})$, $f(y_i|y_{(i)}, \mathcal{M})$ is defined as the predictive density of observation i in the data based on all the data except y_i; y_i might be a scalar or a vector. This has also been termed as the conditional predictive ordinate (CPO_i) (eg. Gelfand and Dey; 1994). The pseudo-marginal likelihood is defined to be $\prod_{i=1}^{n} f_i(y_i|y_{(i)}, \mathcal{M})$. Models with large values of the LPML are preferable to those with small values. In standard linear models, there are well-known formulas for CPOs. The standard generic MCMC based computational formula for CPO is

$$\mathrm{CPO}_i^{-1} \doteq (1/s)\sum_{k=1}^{s} 1/f_i(y_i|\theta^k, \mathcal{M}),$$

where $\{\theta^1, \ldots, \theta^s\}$ is an MCMC sample from the joint posterior of all unknown parameters and latents in the generic model. Our LPML and CPO are slightly more complicated, but straight forward to calculate.

The pseudo-marginal likelihood for our problem is defined to be $\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{y}_{(ij)}, \boldsymbol{X}_i, \mathcal{M})$, where $f(y_{ij}|\boldsymbol{y}_{(ij)}, \boldsymbol{X}_i, \mathcal{M})$ is the predictive density, under model $\mathcal{M}$, corresponding to individual i at time j based on the data minus y_{ij}. LPML values are given in Table 3. The clear winners are the sigmoid and linear NPAR models, with the sigmoid-OUP model being the overall winner. Models III, IV and VI are clearly the worst; recall that model III is a sigmoid-parametric model, model IV uses the DDP to model random effects in the absence of OUP structure and model VI uses linear rather than sigmoid structure, and leaves AR structure out of the OUP. The in between model, model V, includes sigmoid structure but leaves AR structure out of the OUP. *It thus seems clear based on this criterion that we prefer a model with both sigmoid and full OUP structure.*

Figure 1 shows plots of fitted values under model I (posterior mean of $S(t_i|\boldsymbol{\beta}_{(c(i))})+b_i+\boldsymbol{w}_i$) for eight randomly selected women, one from each age by race/ethnicity combination. The fitted values are quite close to the observed values. Observe the different time scales for each woman, which depend on when they entered the study. The corresponding figure for model II (not shown) is virtually identical to the eye. 95% probability intervals, calculated separately at each of the observed time points. Figure 2 shows dramatic differences in fitted

values for participant 41, depending on model. The plots for participant 41 based on models II (linear OUP) and model III (sigmoid parametric) are not shown because they were virtually identical to those for models I (sigmoid OUP) and IV (sigmoid DDP), respectively. The parametric and DDP sigmoid models fit a strictly increasing sigmoid shaped curve for participant 41 with no downturn at the end. But, it has to also be noted that patient 41 left the study at the time of FMP, so it may be expected that a down turn would have come sometime later; see Figure 3.

In response to a referee's query, based on their concern that intervals in Figure 1 were narrow and that the fits might involve over-fitting, we plotted predicted values and corresponding intervals for patient 1 based on using predictive distributions that use all data, and also based on predictive distributions that use all the data except for the case being predicted. This plot is given in Appendix E, and shows that case deleted predictions track the data well, but not as well as the full data based predictions. Moreover, prediction intervals based on case deletion are considerably wider, as would be expected.

Figure 3 shows model based future predictions (posterior mean curves) for the eight different types of patient, all on the same time scale. It thus makes sense to compare shapes and levels across race/ethnicity for the same age group, and between age groups for the same race/ethnicity. Generally speaking, all models that include sigmoid mean functions predict that women's FSH hormones will go up sigmoidally, and then curve downwards toward the end of the time frame, regardless of age-race/ethnicity category. Model II on the other hand predicts a simple linear increase in FSH hormone values in contrast to the others. We would have little to say about the distinctions in the eight Model II plots in Figure 3. However, the sigmoid based models tended to achieve their maximum values of FSH levels at or around the time of FMP. The most notable departure from that trend is the group of younger Chinese women, who were estimated to achieve their peak between one and two years after FMP. Also the estimated level of the peak was noticeably lower for the younger Chinese than it was for younger Japanese and Caucasian women, and the estimated peak FSH levels for older African Americans were noticeably lower than the corresponding estimates for Japanese and Caucasians. Our final inferences relate to the curves from model I, but the general trends discussed here are the same for all three sigmoid based models.

Statistical inferences are straight forward. First, consider the timing of the modeled FSH maximum. At each MCMC iteration, we numerically calculate the time that corresponds to the maximum level achieved in the modeled sigmoid curve for each covariate combination, resulting in a numerical approximation to the posterior distribution of the time it takes to achieve the maximum, relative to FMP. Figure 4 gives posterior estimates and 95% probability intervals for all eight combinations. The most dramatic inference is that Chinese women who are 46 years old and under at baseline achieve their maximum approximately between one and three years after FMP with 95% posterior probability, while corresponding intervals for younger women in the other race/ethnic groups are below this interval. The posterior probabilities that the timing of the maximum for younger Chinese women would be greater than that for African Americans, Caucasians and Japanese are 0.987, 0.9998 and 0.9990, respectively.

Intervals for African Americans, Caucasians and Chinese (see Figure 4) are above zero, so these maxima are evidently achieved after FMP. Among older women at baseline, there is a 0.95 posterior probability that timing for African Americans is greater than for Caucasians. The posterior probability that the difference in timing comparing younger to older Chinese women is positive, is one to four decimal places. There is a clear statistical difference in timing comparing age groups for Chinese women but not for the other groups.

We next compared slopes among women in the eight groups. The slope was taken to be the slope of a straight line that connects the values of the modeled FMP curve at the time of the maximum value minus 4 years and the time of the maximum value. These lines can be imagined by looking at the curves in Figure 4. Estimated slopes and 95% PIs for younger African Americans, Caucasians, Chinese and Japanese are 13.5 (9.9, 16.5), 15.3 (12.7, 17.4), 5.3 (1.8,7.6) and 18.1 (13.2, 22.6), respectively. The interval for younger Chinese women is lower and does not overlap any of the other intervals, so there is clearly a statistically important difference in their slope over this range than for all other race/ethnicities. The other slopes do not appear to be practically different from one another.

Finally, we estimated correlations among repeated responses on a new patient with *equally spaced times* of observation based on the joint predictive distribution under Model I. The estimated Toeplitz correlations for these times that were {1, 2, …, 8} years apart were: {0.43, 0.27, 0.21, 0.17, 0.15, 0.14, 0.14, 0.13}, respectively, which is quite distinct from AR structure. We observe that, after about four years, the correlations flatten out around 0.14. With a typical AR structure, the estimated correlations would continue to decrease across time.

6 Conclusions

Longitudinal data have often been modeled using functional models in time in conjunction with random effects in order to cope with longitudinal correlation. While it is clear that the actual dependence among repeated observations on the same individual will be somewhat accommodated by such models, it is not clear to what extent. Consequently, other authors have used Gaussian processes that incorporate AR structure that allows for higher correlation between observations that are observed closer in time than those observed farther apart. However, such correlation is constrained to be a power of the time difference. GP models with more complex covariance functions have been employed to lessen this issue, and to cope with possible lack of stationarity. In this paper, we developed a generalization of many popular existing models for longitudinal data that results in a novel and somewhat general Toeplitz structure for the GP part of the model. Moreover, at least in our example, we noticed that models without the GP part were noticeably less adequate than models that included EOP structure, according to the LPML criterion.

We showed, for our data, that the NPAR model with local linear structure is able to capture shapes in the actual hormone profiles. However, we also found that, for interpretive purposes, it was necessary to replace linear structure with sigmoid structure, which resulted in a larger LPML and where it was possible to make inferences about the different shapes of curves associated with distinct covariate combinations. For the data analyzed, it would not

have been sufficient to only incorporate sigmoid structure with either parametric or nonparametric random effects. It was necessary to also incorporate longitudinal correlated random effects, which was accomplished through an NPAR model.

As previously mentioned, a number of authors have developed models for handling non-stationary data. The NPAR model (3) is general enough to adopt non-stationary forms. One such example is model II from subsection 5.2, since the variance of $\boldsymbol{y}_i$ depends on the vector of times $\boldsymbol{t}_i$. Another example would follow by adding a sigmoid function with random effects for coefficients to model I in subsection 5.2. We could also model the log of σ^2 to be either a known functional form of time, like a polynomial, or to specify some form of nonparametric model for it. Finally, we could generalize our NPAR model by extending the DP to a DDP. Here, the model would include separate DP mixtures of OU processes for each age by ethnicity category, and they would be made to be dependent in the same way that was described in Appendix D for the DDP part of Model IV. While quite feasible, we did not pursue this particular modeling line in the FSH application because the proposed model already fits the sample profiles quite well, and because the simulation study suggests that the cluster-specific covariance structures are reasonably well estimated.

Finally, a referee asked that we compare our DP mixture model, Model I, with with comparable models that involved parametric finite mixtures. We did this and results are given in Appendix F. The bottom line is that our Model I was a substantial improvement on all models considered there.

Acknowledgments

We thank all the referees for comments and suggestions that greatly improved the paper. Fernando A. Quintana was partially funded by grant FONDECYT 1100010. The Study of Women's Health Across the Nation (SWAN) has grant support from the National Institutes of Health (NIH), DHHS, through the National Institute on Aging (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women's Health (ORWH) (Grants NR004061; AG012505, AG012535, AG012531, AG012539, AG012546, AG012553, AG012554, AG012495). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIA, NINR, ORWH or the NIH.

Clinical Centers: University of Michigan, Ann Arbor - Siobán Harlow, PI 2011 – present, MaryFran Sowers, PI 1994 – 2011; Massachusetts General Hospital, Boston, MA - Joel Finkelstein, PI 1999 – present; Robert Neer, PI 1994 – 1999; Rush University, Rush University Medical Center, Chicago, IL - Howard Kravitz, PI 2009 – present; Lynda Powell, PI 1994 – 2009; University of California, Davis/Kaiser - Ellen Gold, PI; University of California, Los Angeles - Gail Greendale, PI; Albert Einstein College of Medicine, Bronx, NY - Carol Derby, PI 2011 – present, Rachel Wildman, PI 2010 – 2011; Nanette Santoro, PI 2004 – 2010; University of Medicine and Dentistry - New Jersey Medical School, Newark - Gerson Weiss, PI 1994 – 2004; and the University of Pittsburgh, Pittsburgh, PA - Karen Matthews, PI.

NIH Program Office: National Institute on Aging, Bethesda, MD - Winifred Rossi 2012; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD - Program Officers.

Central Laboratory: University of Michigan, Ann Arbor - Daniel McConnell (Central Ligand Assay Satellite Services).

Coordinating Center: University of Pittsburgh, Pittsburgh, PA - Kim Sutton-Tyrrell, Co-PI 2001 – present; Maria Mori Brooks Co-PI 2012; New England Research Institutes, Watertown, MA - Sonja McKinlay, PI 1995 – 2001.

Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair.

We thank the study staff at each site and all the women who participated in SWAN.

Appendix A: Details for Posterior Simulation

Under the assumed prior distributions, and in the linear mean case, the full conditionals for $\boldsymbol{\beta}$ and for b_i are multivariate normal and are easily determined and sampled. When sigmoid curves (5) are used, updating $\boldsymbol{\beta}$ is done through random walk-type Metropolis within Gibbs steps. The full conditional for σ is easily handled by another random walk Metropolis within Gibbs step. In the simple case of the b_i's being only random intercepts and with an inverse gamma prior on the corresponding variance component, the full conditional for σ_b^2 is also inverse gamma. In general, if $\boldsymbol{D}(\boldsymbol{\xi})$ is diagonal, and if independent inverse gamma priors are assigned to the r components, then the full conditional for $\boldsymbol{\xi}$ will consist of r independent inverse gamma distributions. These comments apply whether or not there are functional components (in both the fixed effects and in the random effects parts). However more complex priors may be warranted in this case. The full conditional for component $\boldsymbol{w}_i$ is obtained using (4) by first obtaining the full conditional for w_{i1} given everything else except the remaining components of $\boldsymbol{w}_i$. Then the full conditional for w_{i2} is obtained, given everything else and the new iterate for w_{i1}, and so on for the rest of the components of $\boldsymbol{w}_i$. All of these distributions are easily determined, computed and sampled.

What remains are the full conditionals for the components of $\boldsymbol{\phi}$. MCMC algorithms for DP mixtures have been developed by several authors, including Escobar (1994), MacEachern and Müller (1998) and Neal (2000), among others. Conditional conjugacy means that a particular full conditional is in the same form as the prior for the particular factor. Thus if it is well known how to sample from the prior, it is possible to sample from the corresponding full conditional. For our DPM, we have $f(\boldsymbol{w} \mid \boldsymbol{\phi})p(\boldsymbol{\phi})$, where after marginalization over G, $p(\boldsymbol{\phi})$ is determined by the Polya urn scheme (Blackwell and MacQueen; 1973). From here, it is standard to obtain the full conditionals

$$p(\phi_i|\phi_{(i)}, \boldsymbol{w}, \text{else}) = p(\phi_i|\phi_{(i)}, \boldsymbol{w}_i) \propto p(\phi_i|\phi_{(i)}) f(\boldsymbol{w}_i|\phi_i),$$

where the (i) notation means all other components except i. Due to exchangeability,

$$\phi_i|\phi_{(i)} \sim \frac{MG_0 + \sum_{r \neq i} \delta_{\phi_r}}{M + n - 1}.$$

The final result is simple to obtain directly, or can be obtained from formulas (3.2–4) in Neal (2000), as well as from the aforementioned references. We have

$$\phi_i|\phi_{(i)}, \boldsymbol{w}_i \sim \tilde{q}_i K_i + \sum_{r \neq i} q_{ir} \delta_{\phi_r}$$

$$K_i \sim \frac{f(\boldsymbol{w}_i|\phi) g_0(\phi)}{\int f(\boldsymbol{w}_i|\phi) g_0(\phi) d\phi}, \quad \tilde{q}_i = \kappa M \int f(\boldsymbol{w}_i|\phi) g_0(\phi) d\phi, \quad q_{ir} = \kappa f(\boldsymbol{w}_i|\phi_r),$$

where g_0 is the pdf corresponding to G_0 and κ is the normalizing constant. Notice the integrals involved here. In the conjugate case, the integrals have analytic expressions due to the choice of G_0. In our non-conjugate case, however, we use Algorithm 8 in Neal (2000),

which allows for simulating some auxiliary variables that avoid these integrations altogether. We now describe this algorithm as it applies to our problem.

Let k denote the number of unique values in $\boldsymbol{\phi}$ (i.e., the number of clusters), and let $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_k$ denote such values. Define s_i to be an indicator of cluster membership, so that $\phi_i = \boldsymbol{\eta}_{s_i}$. The values of the s_i variables are themselves immaterial, but are treated only as class indicators. The idea of Neal (2000) consists of introducing m temporary additional auxiliary parameters $\boldsymbol{\eta}_{k+1}, \ldots, \boldsymbol{\eta}_{k+m}$, drawn from G_0, that are associated with "empty clusters", i.e., related to no observation, and that are discarded if unused. The Gibbs sampler acts first by updating the configuration indicators $\boldsymbol{s} = (s_1, \ldots, s_n)$. Let k^- the number of distinct s_j for $j \neq i$, and denote by n_ℓ^- the number of s_j for $j \neq i$ that are equal to ℓ. By exchangeability and the fact that the s_j values are arbitrary, when updating s_i we can assume that we are updating s_i for the last observation and that the s_j for other observations have values in $\{1, \ldots, k^-\}$. It then follows that the prior probability $P(s_i = \ell \mid \boldsymbol{s}_{(i)})$ is given by $n_\ell^-/(M+n-1)$ for $1 \leq \ell \leq k^-$ and by $(M/m)/(M+n-1)$ for $k^- < \ell \leq k^- + m$.

Combining the previous representation with the likelihood factor $f(\boldsymbol{w}_i \mid \boldsymbol{\phi})$, the relevant step in Algorithm 8 in Neal (2000) works as follows for $i = 1, \ldots, n$. Let $h = k^- + m$, and label the s_j in $\{1, \ldots, k^-\}$. If $s_i = s_j$ for some $j \neq i$, draw independent values from G_0 for $\boldsymbol{\eta}_{k^-+1}, \ldots, \boldsymbol{\eta}_h$. If $s_j \neq s_i$ for all j (i.e., observation i was in a singleton), assign s_i the label $k^- + 1$ and draw independent G_0 values for $\boldsymbol{\eta}_{k^-+2}, \ldots, \boldsymbol{\eta}_h$. Then, resample s_i from

$$P(s_i = \ell \mid \boldsymbol{s}_{(i)}, \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_h) = \begin{cases} \kappa \frac{n_\ell^-}{M+n-1} f(\boldsymbol{w}_i \mid \boldsymbol{\eta}_\ell) & \text{for } 1 \leq \ell \leq k^- \\ \kappa \frac{M/m}{M+n-1} f(\boldsymbol{w}_i \mid \boldsymbol{\eta}_\ell) & \text{for } k^- < \ell \leq h, \end{cases}$$

where κ is an appropriate normalizing constant.

To conclude the resampling step, drop all the $\boldsymbol{\eta}$ parameters that are not associated with observations, and draw $\boldsymbol{\eta}_\ell$ from the corresponding full conditional, which amounts to the posterior distribution of $\boldsymbol{\eta}_\ell \mid \{\boldsymbol{w}_i : s_i = \ell\}$.

Appendix B: Inverse of H~(ρ)

Let $t_1 < \cdots < t_n$ be real constants and write $r_k = \rho^{|t_{k+1} - t_k|}$ for $k = 1, \ldots, n-1$. Consider the generic symmetric $n \times n$ matrix $\tilde{\boldsymbol{H}}(\rho)$ with entries $\left\{\tilde{\boldsymbol{H}}(\rho)\right\}_{k,k} = 1$ for $k = 1, \ldots, n$, and for $\ell < k$

$$\left\{\tilde{\boldsymbol{H}}(\rho)\right\}_{\ell k} = r_\ell r_{\ell+1} \cdots r_{k-1}.$$

For instance, when $n = 4$,

$$\tilde{\boldsymbol{H}}(\rho)=\begin{bmatrix} 1 & r_1 & r_1r_2 & r_1r_2r_3 \\ r_1 & 1 & r_2 & r_2r_3 \\ r_1r_2 & r_2 & 1 & r_3 \\ r_1r_2r_3 & r_2r_3 & r_3 & 1 \end{bmatrix}.$$

This form is identical to that of antedependence. It then follows that (Zimmerman and Núñez-Antón; 2010)

$$\begin{gathered} \left\{\tilde{\boldsymbol{H}}(\rho)^{-1}\right\}_{k,k+1}=-\frac{r_k}{1-r_k^2},\ \ k=1\ldots,n-1 \\ \left\{\tilde{\boldsymbol{H}}(\rho)^{-1}\right\}_{1,1}=\frac{1}{1-r_1^2} \\ \left\{\tilde{\boldsymbol{H}}(\rho)^{-1}\right\}_{k,k}=\frac{1-r_{k-1}^2r_k^2}{1-r_{k-1}^2-r_k^2+r_{k-1}^2r_k^2},\ \ k=2,\ldots,n-1 \\ \left\{\tilde{\boldsymbol{H}}(\rho)^{-1}\right\}_{n,n}=\frac{1}{1-r_{n-1}^2}. \end{gathered}$$

and $\left\{\tilde{\boldsymbol{H}}(\rho)^{-1}\right\}_{k,\ell}=0$ otherwise. In addition, we have that

$$\det(\tilde{\boldsymbol{H}}(\rho))=1+\sum_{i=1}^{n-1}(-1)^i\prod_{1\le s_1<s_2<\cdots<s_i\le n-1}r_{s_i}^2.$$

For instance, again in the $n = 4$ case,

$$\tilde{\boldsymbol{H}}(\rho)^{-1}=\begin{bmatrix} \frac{1}{1-r_1^2} & \frac{-r_1}{1-r_1^2} & 0 & 0 \\ \frac{-r_1}{1-r_1^2} & \frac{1-r_1^2r_2^2}{1-r_1^2-r_2^2+r_1^2r_2^2} & \frac{-r_2}{1-r_2^2} & 0 \\ 0 & \frac{-r_2}{1-r_2^2} & \frac{1-r_2^2r_3^2}{1-r_2^2-r_3^2+r_2^2r_3^2} & \frac{-r_3}{1-r_3^2} \\ 0 & 0 & \frac{r_3}{1-r_3^2} & \frac{1}{1-r_3^2} \end{bmatrix};$$

and

$$\det(\tilde{\boldsymbol{H}}(\phi))=1-r_1^2-r_2^2-r_3^2+r_1^2r_2^2+r_1^2r_3^2+r_2^2r_3^2-r_1^2r_2^2r_3^2.$$

Appendix C: Plot of All Data According to Age and Race-Ethnicity Status

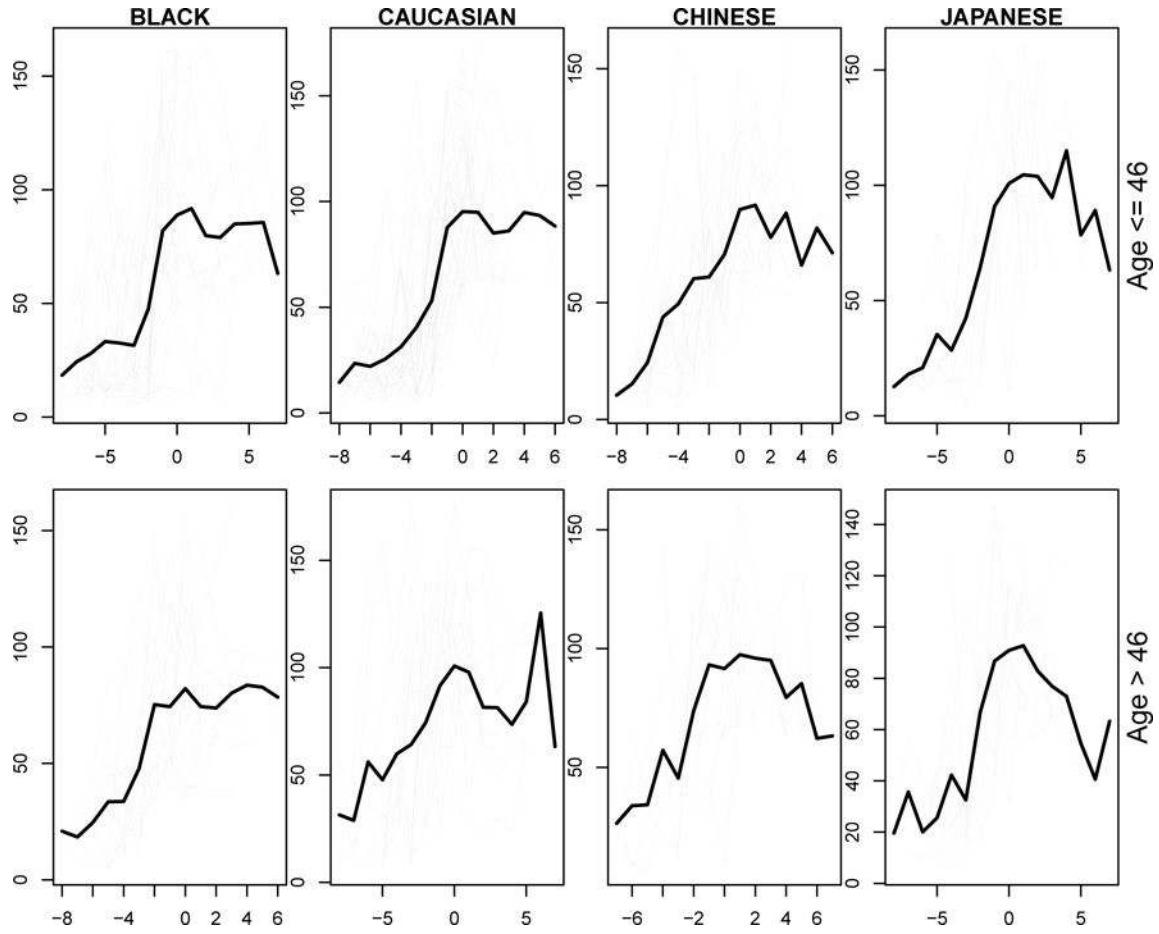

Figure 5.
Raw data plots. Thick black line is the average by group.

Appendix D: Details for Dependent Dirichlet Process

The random effects for the eight age by race-ethnicity categories in Model IV were modeled using a DPM of normal distributions. The mixing is on an eight dimensional vector with each component corresponding to one of the eight categories. Here, we discuss the underlying DDP model that is induced by this model.

Let $\Xi = \{1, 2, \ldots, 8\}$ denote the eight categories. Our model corresponds to specifying marginal DPM models for the random effects distribution for each of these categories, and a bit more. Namely, if we let $b_i | c(i) = k$, $\boldsymbol{\gamma} \sim N(\gamma_k, s^2) = N(z'_{c(i)}\gamma, s)$, and let $\gamma_k | G_k \sim G_k$, with $G_k \sim DP(M, G_{0k})$, $k = 1, \ldots, 8$, where G_{0k} is the kth marginal of G_0, then these marginal distributions are identical to the corresponding induced marginal distributions based on Model IV. By the Sethuraman construction, $G_k(\cdot) = \sum_{j=1}^{\infty} w_j \delta_{\gamma_k^j}(\cdot)$ where $\gamma_k^j \overset{\text{iid}}{\sim} G_{0k}$. Thus each of the eight random effects distributions is a DP mixture of normal distributions where the mixing is on the mean for that group. In the extreme case with very

large M, this is just a parametric normal random effects distribution. At the other end of the spectrum, this specification allows for mixtures of normals, that could be multimodal, for the random effects distributions. Then the "a bit more" part involves the fact that the collection of distributions $\{G_k : k \in \Xi\}$ are allowed to be correlated. The collection is obviously expressed as $\left\{G_k(\cdot)=\sum_{j=1}^{\infty} w_j \delta_{\gamma_k^j}(\cdot) : k \in \Xi\right\}$, where we now recognize that the vectors $(\gamma_k^j : k=1,\ldots,8)$ are iid in j, and that the components of these vectors are iid G_0, (conditional on (λ, Ψ)) which are assumed to be multivariate normal with non-diagonal covariance matrix in our specification. Note that the set of random probability measures $\{G_k : k \in \Xi\}$ is thus constructed from a single set of stick-breaking weights $\{w_j : j \geq 1\}$ and a single set of atoms $\{\boldsymbol{\gamma}_j : j \geq 1\}$. These common elements are what in the end induce dependence among the probability measures G_k. Thus we have described a collection of Dirichlet Processes that are dependent, which is termed as a DDP by its inventor (MacEachern; 1999, 2000). The purpose of the dependence is so that strength can be borrowed across the eight populations, which is the standard purpose of hierarchical model specifications such as the one here. MacEachern's definition of the DDP was more general in the sense that Ξ is allowed to be more general. There have now been many applications of the DDP to various problems, see for example De Iorio et al. (2004, 2009), and Dunson et al. (2007), among many others.

Appendix E: Prediction Plots for Patient 1 with and without the Response Being Predicted

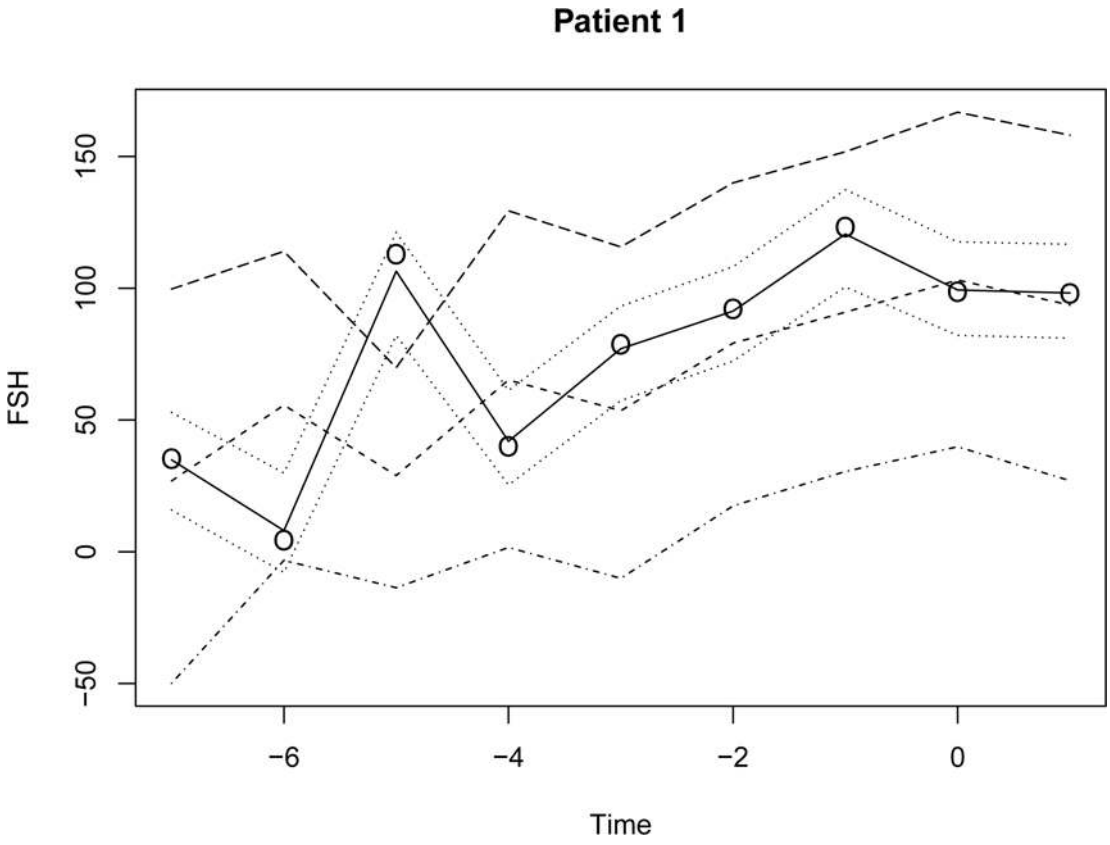

Figure 6.

Predictions and prediction bands based on all data and case deleted data: Solid and dotted gives predictions and intervals based on all data; dashed give predictions and intervals based on case deleted data.

Appendix F: Comparison Between Model I and a Parametric Finite Mixture Model

We make a comparison between our selected model, model I, and the more traditional finite K-term mixture model for a variety of values of K. These models are meant to be simplifications of our model I in the sense that the mixtures of the OU process parameters, (ρ, σ^2_w), are now classical fixed finite dimensional mixtures, rather than DPMs. Specifically, we consider $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ to be generated from the mixture distribution

$$\pi_1 \delta_{\phi_1} + \cdots + \pi_K \delta_{\phi_K},$$

where $\phi_1, \ldots, \phi_K \overset{\text{iid}}{\sim} G_0$ and $(\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K) \sim$ Dirichlet(1, …, 1). Here G_0 is the same baseline distribution considered earlier for model I. These models are indexed by the number of terms in the mixture, K, but otherwise are the same. They include a sigmoid function part that is identical to the one considered in the paper and where a separate function is allowed for the 8 age by ethnicity categories. They also include random effects terms, b, Gaussian process terms, w, and error terms, ε. These models were fit in WinBUGS for $K = 2, \ldots, 10$, and LPML values were obtained in each instance. The range of LPMLs was −7700.2, for $K = 3$, to −7665.4 for the model with $K = 5$. The model with $K = 5$ wins. However, since the LPML for our model I is −5966, it is a clear winner over all of the finite mixture models. In fact, all of our models I–VI have much larger LPML values than these.

References

Blackwell D, MacQueen J. Ferguson distributions via Pólya urn schemes. The Annals of Statistics. 1973; 1:353–355.

Christensen, R., Johnson, W., Branscum, A., Hanson, TE. Bayesian ideas and data analysis. CRC Press; Boca Raton, FL: 2010. Texts in Statistical Science SeriesAn introduction for scientists and statisticians

Daniels, MJ., Hogan, JW. Missing data in longitudinal studies. Chapman & Hall/CRC; Boca Raton, FL: 2008. Vol. 109 of Monographs on Statistics and Applied ProbabilityStrategies for Bayesian modeling and sensitivity analysis

Daniels MJ, Kass RE. Shrinkage estimators for covariance matrices. Biometrics. 2001; 57(4):1173–1184. [PubMed: 11764258]

Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika. 2002; 89(3):553–566.

Davidian, M., Giltinan, DM. Nonlinear Models for Repeated Measurement Data. Chapman & Hall Ltd; 1998.

De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. Biometrics. 2009; 65(3):762–771. [PubMed: 19210742]

De Iorio M, Müller P, Rosner G, MacEachern S. An ANOVA model for dependent random measures. Journal of the American Statistical Association. 2004; 99:205–215.

Dennerstein L, Lehert P, Burger H, Guthrie1 J. New findings from nonlinear longitudinal modelling of menopausal hormone changes. Human Reproduction Update. 2007; 13:551–557. [PubMed: 17616552]

Diggle PJ. An approach to the analysis of repeated measurements. Biometrics. 1988; 44(4):959–971. [PubMed: 3233259]

Diggle, PJ., Heagerty, PJ., Liang, KY., Zeger, SL. Analysis of longitudinal data. second. Vol. 25. Oxford University Press; Oxford: 2002. Vol. 25 of Oxford Statistical Science Series

Diggle PJ, Verbyla AP. Nonparametric estimation of covariance structure in longitudinal data. Biometrics. 1998; 54:401–415. [PubMed: 9629635]

Dunson DB, Pillai N, Park JH. Bayesian density regression. Journal of the Royal Statistical Society. Series B. Statistical Methodology. 2007; 69(2):163–183.

Escobar MD. Estimating normal means with a Dirichlet process prior. Journal of the American Statistical Association. 1994; 89(425):268–277.

Fitzmaurice, GM., Laird, NM., Ware, JH. Applied longitudinal analysis. second. John Wiley & Sons Inc; Hoboken, NJ: 2011. Wiley Series in Probability and Statistics

Gabriel KR. Ante-dependence analysis of an ordered set of variables. Ann Math Statist. 1962; 33:201–212.

Geisser S, Eddy WF. A predictive approach to model selection. J Amer Statist Assoc. 1979; 74(365): 153–160.

Gelfand AE, Dey DK. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society Series B (Methodological). 1994:501–514.

Gelfand AE, Kottas A, MacEachern SN. Bayesian nonparametric spatial modeling with Dirichlet process mixing. Journal of the American Statistical Association. 2005; 100:1021–1035.

Hand, DJ., Crowder, MJ. Practical Longitudinal Data Analysis. Chapman and Hall Ltd.; London: 1996. Texts in Statistical Science

Hanson TE, Branscum AJ, Johnson WO. Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. Lifetime Data Anal. 2011; 17(1):3–28. Supplementary material available online. [PubMed: 20369294]

Hardin, JW., Hilbe, JM. Generalized estimating equations. Chapman & Hall/CRC; Boca Raton, FL: 2003.

He X, Zhu ZY, Fung WK. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika. 2002; 89(3):579–590.

Hedeker, D., Gibbons, RD. Longitudinal data analysis. John Wiley & Sons; Hoboken, NJ: 2006. Wiley Series in Probability and Statistics, Wiley-Interscience

Jeffreys, H. Theory of Probability. Oxford University Press; 1961.

Johnson, RA., Wichern, DW. Applied multivariate statistical analysis. sixth. Pearson Prentice Hall; Upper Saddle River, NJ: 2007.

Kleinman KP, Ibrahim JG. A semi-parametric Bayesian approach to generalized linear mixed models. Statistics in Medicine. 1998a; 17:2579–2596. [PubMed: 9839349]

Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. Biometrics. 1998b; 54:921–938. [PubMed: 9750242]

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Leonard T, Hsu JSJ. Bayesian inference for a covariance matrix. Ann Statist. 1992; 20(4):1669–1696.

Li Y, Lin X, Müller P. Bayesian inference in semiparametric mixed models for longitudional data. Biometrics. 2010; 66(1):70–78. [PubMed: 19432777]

Liechty JC, Liechty MW, Müller P. Bayesian correlation estimation. Biometrika. 2004; 91(1):1–14.

MacEachern, SN. ASA Proceedings of the Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association; 1999. Dependent Nonparametric Processes.

MacEachern, SN. Dependent Dirichlet processes. Department of Statistics, The Ohio State University; 2000. Technical report

MacEachern SN, Müller P. Estimating Mixture of Dirichlet Process Models. Journal of Computational and Graphical Statistics. 1998; 7(2):223–338.

Mukhopadhyay S, Gelfand AE. Dirichlet process mixed generalized linear models. Journal of the American Statistical Association. 1997; 92(438):633–639.

Müller P, Rosner GL. A Bayesian population model with hierarchical mixture priors applied to blood count data. Journal of the American Statistical Association. 1997; 92:1279–1292.

Neal R. Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics. 2000; 9:249–265.

Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika. 1999; 86(3):677–690.

Pourahmadi M, Daniels MJ. Dynamic conditionally linear mixed models for longitudinal data. Biometrics. 2002; 58(1):225–231. [PubMed: 11890319]

Pullenayegum E, Lim L. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. Stat Methods Med Res. 2015; 42:121–130.

Rasmussen, CE., Williams, CKI. Gaussian processes for machine learning. MIT Press; Cambridge, MA: 2006. Adaptive Computation and Machine Learning

Ray WD. The inverse of a finite toeplitz matrix. Technometrics. 1970; 12:153–156.

Ricketts J, Head G. A five-parameter logistic equation for investigating asymmetry of curvature in baroreflex studies. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology. 1999; 277:R441–R454.

Sethuraman J. A constructive definition of Dirichlet priors. Statist Sinica. 1994; 4(2):639–650.

Shi M, Weiss RE, Taylor JMG. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. Applied Statistics. 1996; 45:151–163.

Smith M, Kohn R. Parsimonious covariance matrix estimation for longitudinal data. Journal of the American Statistical Association. 2002; 97(460):1141–1153.

Taylor J, Cumberland W, Sy J. A stochastic model for analysis of longitudinal AIDS data. Journal of the American Statistical Association. 1994; 89:727–736.

Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica. 2004; 14(3):809–834.

Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association. 1996; 91:217–221.

Verbeke, G., Molenberghs, G. Linear mixed models for longitudinal data. Springer-Verlag; New York: 2000. Springer Series in Statistics

Waetjen L, Johnson W, Xing G, Feng W, Greendale G, Gold E. Serum estradiol levels are not associated with urinary incontinence in midlife women transitioning through menopause. Menopause: The Journal of The North American Menopause Society. 2011; 18:1283–90.

Yang RY, Berger JO. Estimation of a covariance matrix using the reference prior. Ann Statist. 1994; 22(3):1195–1211.

Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics. 1994; 50:689–699. [PubMed: 7981395]

Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics. 2001; 57(3):795–802. [PubMed: 11550930]

Zhang D, Lin X, Raz J, Sowers M. Semiparametric stochastic mixed models for longitudinal data. J Amer Statist Assoc. 1998; 93(442):710–719.

Zimmerman, DL., Núñez-Antón, VA. Antedependence models for longitudinal data. CRC Press; Boca Raton, FL: 2010. Vol. 112 of Monographs on Statistics and Applied Probability

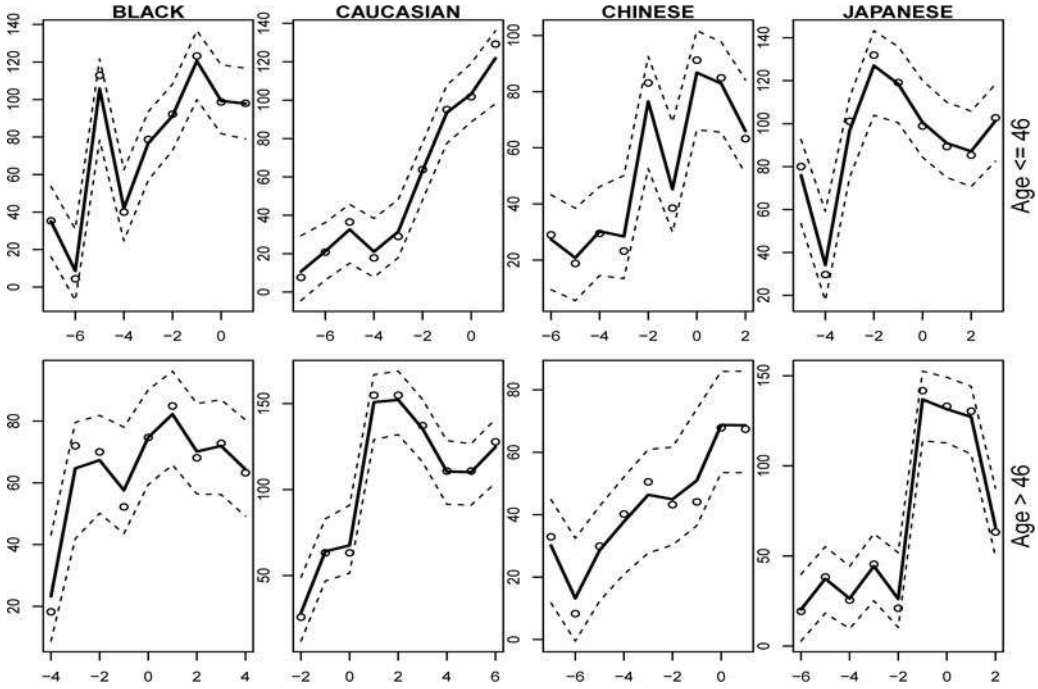

Figure 1.

Fitted FSH values and 95% intervals for eight randomly selected individuals using model I. Note the different scales.

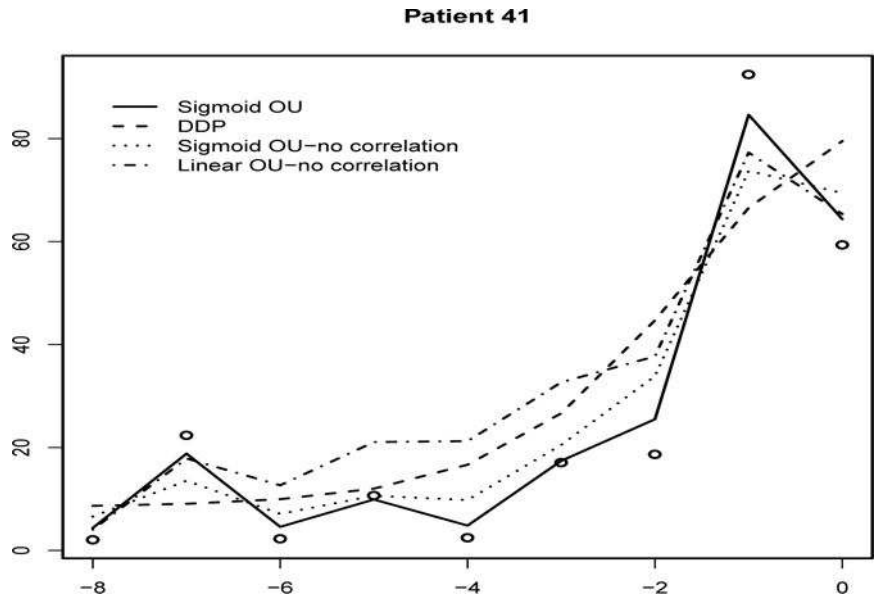

Figure 2.
Fits for patient 41

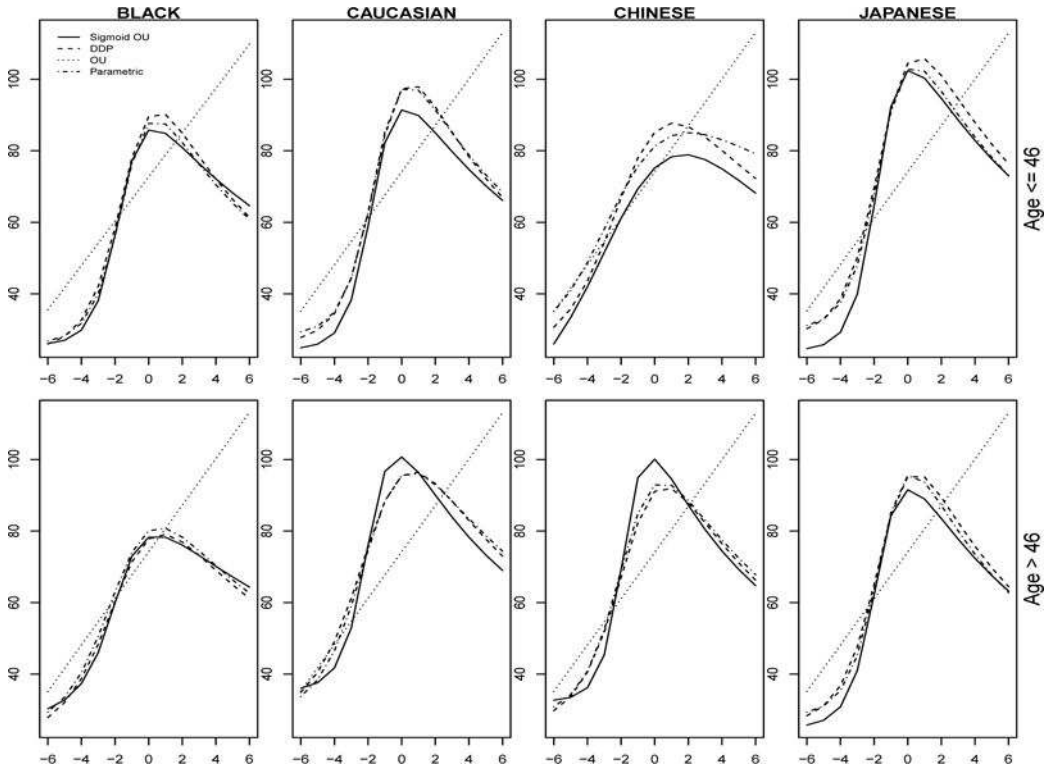

Figure 3.
Predictions of future hormone concentrations for eight types of women, using Model I

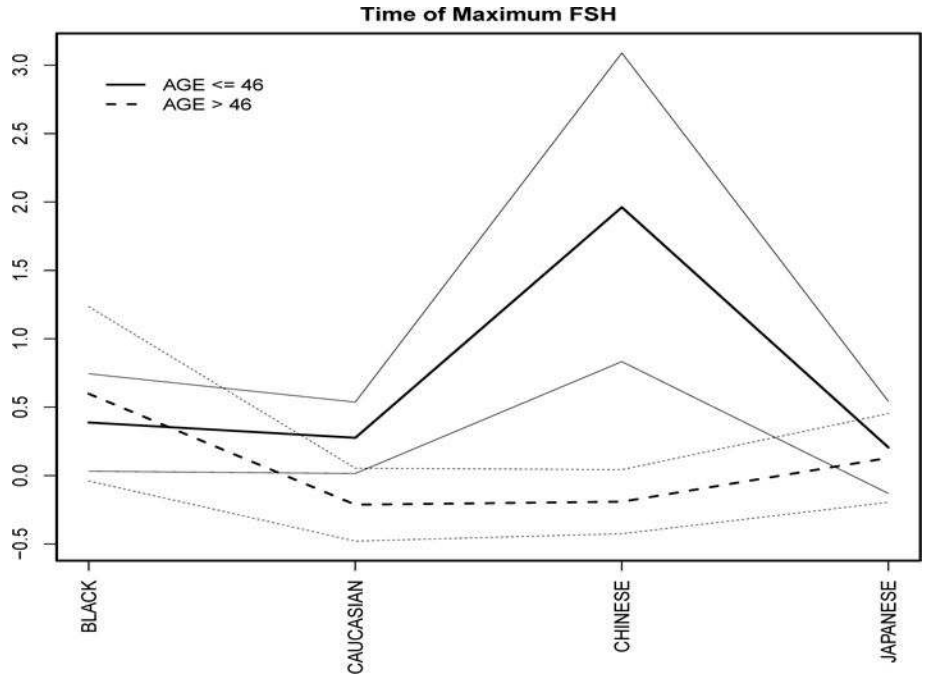

Figure 4.

Comparisons of estimated time to maximum (and 95% posterior probability intervals) over the eight covariate combinations, using Model I

Table 1

Simulation Results for Σ

Loss	Model	C_1	C_2	C_3	C_4
RMSE	NPAR	0.96	0.88	0.51	0.68
	DP (2002)	1.14	1.19	1.12	1.37
	Inv Wish	1.09	1.02	1.04	1.07
L_1	NPAR	0.04	0.07	0.03	0.05
	DP (2002)	0.73	0.80	0.76	0.77
	Inv Wish	1.13	1.09	1.14	1.11
L_2	NPAR	0.08	0.19	0.06	0.09
	DP (2002)	1.16	1.14	1.11	1.19
	Inv Wish	1.30	1.28	1.30	1.29

Table 2

Simulation Results for $\boldsymbol{R}$

Loss	Model	C_1	C_2	C_3	C_4
RMSE	NPAR	0.08	0.19	0.06	0.09
	DP (2002)	0.06	0.09	0.08	0.07
	Inv Wish	0.05	0.09	0.09	0.07
L_1	NPAR	0.09	0.08	0.03	0.06
	DP (2002)	0.70	0.83	0.78	0.85
	Inv Wish	0.96	0.87	0.91	0.88
L_2	NPAR	0.18	0.19	0.07	0.12
	DP (2002)	1.07	1.11	1.08	1.14
	Inv Wish	1.19	1.13	1.15	1.15

Table 3

LPML Values for Models I–VI

Model	I	II	III	IV	V	VI
	−5966	−6186	−6936	−6936	−6673	−6912