

# Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves

Babette A. Brumback and John A. Rice

## Abstract

We introduce a class of models for an additive decomposition of groups of curves stratified by crossed and nested factors, generalizing smoothing splines to such samples by associating them with a corresponding mixed effects model. The models are also useful for imputation of missing data and exploratory analysis of variance. We prove that the best linear unbiased predictors (BLUP) from the extended mixed effects model correspond to solutions of a generalized penalized regression where smoothing parameters are directly related to variance components, and we show that these solutions are natural cubic splines. The model parameters are estimated using a highly efficient implementation of the EM algorithm for restricted maximum likelihood (REML) estimation based on a preliminary eigenvector decomposition. Variability of computed estimates can be assessed with asymptotic techniques or with a novel hierarchical bootstrap resampling scheme for nested mixed effects models. Our methods are applied to menstrual cycle data from studies of reproductive function that measure daily urinary progesterone; the sample of progesterone curves is stratified by cycles nested within subjects nested within conceptive and non-conceptive groups.

KEY WORDS: Mixed effects model; Penalized regression; Variance component; Smoothing parameter; Hierarchical Bootstrap; Menstrual data

## 1. INTRODUCTION

Curve data arise frequently in scientific studies and form an active topic of current statistical research. Familiar examples of such data include growth curves, biomarkers measured over time, and reaction curves in chemical experiments. Our own work is motivated by studies of reproductive function where the observational unit is the profile of a particular hormone

---

Babette A. Brumback is Postdoctoral Fellow, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. John A. Rice is Professor, Statistics Department, University of California, Berkeley, CA 94720. This research was partly supported by National Institute for Environmental Health Sciences grants NIEHS 6198 and NIEHS 5T32 ESO7142-14 (Brumback) and National Science Foundation grant NSF DMS-9213353 (Rice), and the paper was partly written as part of the first author's thesis at the Statistics Department, University of California, Berkeley. We thank Terry Speed, Shanna Swan, and Bill Lasley for numerous discussions and Trevor Hastie, Linda Kaufman, and Louise Ryan for helpful comments.

measured daily in urine over the course of a menstrual cycle; see, for example, Munro et al. (1991). One of the interesting features of curve data from menstrual cycles is that cycles are nested in women, who in turn may be nested in groups. Furthermore, the nesting is typically unbalanced, with women contributing differing numbers of cycles and categorized by groups of unequal sizes. Imbalance can also occur within the observational units, as it is not uncommon for cycles to be measured with missing data. In this paper, we extend the smoothing spline model for individual curves (Silverman 1985; Wahba 1978) to generalized smoothing spline models for samples of curves stratified by nested and crossed factors. In so doing, we create for the menstrual data a smoothing spline estimation of group means as well as of subject and cycle departures that handles appropriately the unbalanced nested sample structure together with the missing data. While our models were motivated by data measured over menstrual cycles (e.g. hormone profiles, other biomarkers, and drug reactivity), other examples of nested and crossed samples of curves come to mind. Consider samples of growth curves stratified by completely crossed block and treatment factors, for instance, or longitudinally sampled biomarkers from patients stratified by partially crossed age and ethnicity factors nested within sex. We anticipate future application of these methods to data arising from a broad range of scientific inquiry.

For illustration we use a sample of urinary metabolite progesterone curves measured over 21 conceptive and 70 non-conceptive menstrual cycles as part of continuing studies of early pregnancy loss conducted by the Institute for Toxicology and Environmental Health at the University of California, Davis in collaboration with the Reproductive Epidemiology Section of the California Department of Health Services in Berkeley. Our sample comes from patients with healthy reproductive function enrolled in an artificial insemination clinic where insemination attempts are well-timed for each menstrual cycle. As is standard practice in endocrinological research (Yen and Jaffe 1991), progesterone profiles are aligned by the day of ovulation, here determined by serum luteinizing hormone, then truncated at each end to present curves of equal length. Figure 1 presents superpositions of nonconceptive profiles stratified by subject, and Figure 2 shows conceptive profiles.

Two original aims of our analysis were to characterize differences in conceptive and non-conceptive cycles prior to implantation (which occurs approximately day 7 or 8 following ovulation) and to explore the between versus within woman variation in non-conceptive cycle profiles. We were thus interested in analysis of variance type questions involving

both factors, subject and group. To accommodate the missing data and approximate the underlying physiology, where daily changes in integrated hormone production can be taken to vary smoothly, we developed a class of models where the factor effects and individual curves are presumed smooth. A particularly useful byproduct of our work turned out to be the capacity of the models to impute missing data in a given cycle by combining a smoothing spline interpolation from neighboring days with additional information from related cycles.

The models and computational methods developed in this paper represent a synthesis of ideas from the literature on mixed effects models, smoothing splines, analysis of variance, and samples of curves. Most contemporary research in the area of samples of curves assumes, either tacitly or explicitly, that the individual curves are pairwise uncorrelated (Anderson and Jones 1995; Ramsay and Dalzell 1991; Rice and Silverman 1991). An exception can be found in Barry (1996), where  $n_i$  individuals indexed by  $j$  and assigned to treatment  $i$  contribute curves that are modeled as the sum of a random mean for treatment  $i$  plus a random intercept term for subject  $j$  plus white noise. The treatment means are assigned a Gaussian prior with an integrated Wiener process component motivated by the correspondence to linear smoothing splines described in Wahba (1978). Much earlier, analysis of variance techniques involving random polynomial coefficients were applied to growth curve problems (Potthoff and Roy 1964; Rao 1965). Our concern is with nonparametric procedures.

In the next section, we outline the nexus between smoothing splines phrased in terms of penalized regression models and alternatively in terms of mixed effects models with certain assumptions on the design matrices and on the covariance of random effects. We then introduce smoothing spline models for the analysis of samples of curves stratified by nested and crossed factors, and subsequently derive dual representations of the models as both penalized regressions and mixed effects linear models. Section 3 describes the computational aspects of estimation within the smoothing spline models, focusing on an efficient approach based on a preliminary eigenvector decomposition. Asymptotic and bootstrap approaches to assessing variability of the smooth curve estimates are also discussed. The methodology is illustrated throughout with the progesterone data.

## 2. SMOOTHING SPLINES AND MIXED EFFECTS MODELS

A review of smoothing splines will first be given, with particular emphasis on the correspondence between smoothing splines and mixed effects models, which arises from the

derivation of a smoothing spline as a Bayes estimate (Kimeldorf and Wahba 1970; Wahba 1978). While the setting of these papers was in continuous time, Silverman (1985) developed the discrete time analog and Speed (1991) pointed out the connection with BLUP and REML in a mixed effects model. We use the mixed effects model as a basis for extending the smoothing spline procedures from one curve to a sample of curves. In 2.1 we present a model for nested samples of curves, giving equivalent formulations in both mixed model and penalized regression terms; in 2.2 we repeat the exercise for a crossed sample of curves with two factors. While for the illustrative purposes of this paper our models are detailed with cubic smoothing splines, the methods pertain to smoothing splines of other orders and more generally to penalized regressions convertible into mixed effects models (all that is required is a quadratic form for the penalty.)

Consider the regression problem where we have observations  $y_i$  at design points  $t_i$ ,  $i = 1, \dots, n$  and the observations are assumed to satisfy

$$y_i = s(t_i) + \varepsilon_i, \tag{1}$$

with  $s(t)$  a smooth curve and  $\varepsilon_i$ ,  $i = 1, \dots, n$  i.i.d.  $N(0, \sigma^2)$ . Let  $\mathcal{W}_2^2[a, b]$  be the Sobolev space of functions whose second derivatives are square integrable, where  $[a, b]$  is any interval containing the design points. The  $s(t) \in \mathcal{W}_2^2[a, b]$  that minimizes the residual sum of squares plus a roughness penalty,

$$\sum_{i=1}^n (y(t_i) - s(t_i))^2 + \lambda \int (s''(t))^2 dt, \tag{2}$$

is called the cubic smoothing spline fit to the data, since the minimizing function  $\hat{s}(t)$  is a natural cubic spline. The smoothing parameter  $\lambda$  governs the tradeoff between smoothness and goodness-of-fit.

Let  $B_i$ ,  $i = 1, \dots, n$  be a basis of functions for the  $n$ -dimensional space  $\mathcal{S}(t_2, \dots, t_{n-1})$  of natural cubic splines with knots at the interior design points,  $t_2, \dots, t_{n-1}$ , such that  $B_1$  and  $B_2$  span the linear functions and  $B_i$ ,  $i = 3, \dots, n$  span those functions orthogonal to  $B_1$  and  $B_2$ . Also define

$$\mathbf{Q} = \left( \int B_i''(t) B_j''(t) dt \right)_{ij} \quad i = 3, \dots, n, \quad j = 3, \dots, n \tag{3}$$

and let the singular value decomposition of  $\mathbf{Q}$  be represented by  $\mathbf{U}\mathbf{D}\mathbf{U}^T$ . Letting  $\mathbf{s}$  denote  $s(t)$  evaluated at  $t_1, \dots, t_n$ , it can be shown that the fitted smoothing spline evaluated at

the design points,  $\hat{\mathbf{s}} = (\hat{s}(t_1), \dots, \hat{s}(t_n))^T$ , equals the BLUP solution  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$  to a mixed effects model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (4)$$

where  $\boldsymbol{\beta}$  and  $\mathbf{u}$  are respectively vectors of fixed and random effects corresponding to known design matrices  $\mathbf{X}$  and  $\mathbf{Z}$ , while  $\boldsymbol{\varepsilon}$  is the vector of random errors. The mixed effects model is specified by  $\mathbf{X} = (B_j(t_i))_{ij}$   $i = 1, \dots, n$ ,  $j = 1, 2$ ;  $\mathbf{B} = (B_j(t_i))_{ij}$   $i = 1, \dots, n$ ,  $j = 3, \dots, n$ ;  $\mathbf{Z} = \mathbf{B}\mathbf{U}\mathbf{D}^{-1/2}$ ;  $\mathbf{u} \sim MVN(\mathbf{0}, (\sigma^2/\lambda)\mathbf{I}_{n-2})$ ; and  $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}_n)$  independent of  $\mathbf{u}$ .

The BLUP solution is given by the equations

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (5)$$

$$\hat{\mathbf{u}} = (\sigma^2/\lambda)\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad (6)$$

where the parameters of  $V = \text{var}(\mathbf{y}) = (\sigma^2/\lambda)\mathbf{Z}\mathbf{Z}' + \sigma^2\mathbf{I}_n$  are presumed known,  $\hat{\boldsymbol{\beta}}$  is the GLS estimator for  $\boldsymbol{\beta}$ , and  $\hat{\mathbf{u}}$  is the regression of  $\mathbf{u}$  on  $\mathbf{y}$ , with the mean  $\mathbf{X}\boldsymbol{\beta}$  of  $\mathbf{y}$  replaced by its obvious linear estimator. Given only the first and second moments of  $\mathbf{u}$  and  $\boldsymbol{\varepsilon}$ , the BLUP solutions for  $\boldsymbol{\beta}$  and  $\mathbf{u}$  are the best linear unbiased predictors in the sense defined by Robinson (1991). With the additional assumption that  $\mathbf{u}$  and  $\boldsymbol{\varepsilon}$  follow the multivariate normal distribution,  $\hat{\boldsymbol{\beta}}$  and  $\hat{\mathbf{u}}$  are respectively the MLE of  $\boldsymbol{\beta}$  and conditional mean of  $\mathbf{u}$  given  $\mathbf{y}$ . It is easy to show that if  $\mathbf{X}'\mathbf{Z} = \mathbf{0}$ , then  $\mathbf{X}'\mathbf{V} = \sigma^2\mathbf{X}'$ ,  $\mathbf{X}'\mathbf{V}^{-1} = (1/\sigma^2)\mathbf{X}'$ , and therefore (5 – 6) reduce to

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (7)$$

$$\hat{\mathbf{u}} = (\sigma^2/\lambda)\mathbf{Z}'\mathbf{V}^{-1}\mathbf{y}, \quad (8)$$

the ordinary least squares estimator for  $\boldsymbol{\beta}$  and the regression of  $\mathbf{u}$  on  $\mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ : in this case,  $\boldsymbol{\beta}$  is estimated as if  $\mathbf{Z}\mathbf{u}$  were not present, and  $\mathbf{u}$  is estimated as if  $\mathbf{X}\boldsymbol{\beta}$  were not present.

For computational convenience as in Hastie and Tibshirani (1990), we let the  $B_i$  above be a basis for the  $(n + 2)$ -dimensional space of cubic splines with knots  $(t_2, \dots, t_{n-1})$  and calculate  $\mathbf{Q}$  using B-spline derivatives, readily available in Splus. The solutions to (2) and (4) are unchanged since the natural cubic splines are a subset of the cubic splines with the same knots. The resulting columns of  $\mathbf{Z}$  are presented graphically in Figure 4. Notice that columns representing smoother functions have larger norms, and recall that the components of  $\mathbf{u}$  are identically distributed. An alternative parametrization could use the same columns

of  $\mathbf{Z}$  but standardized columns, absorbing the norms into the distribution of  $\mathbf{u}$ . The basis for the natural cubic splines given by the orthogonalized columns of  $\mathbf{X}$  combined with the standardized columns of  $\mathbf{Z}$  is known as the Demmler-Reinsch basis (Demmler and Reinsch 1975; Eubank 1988). We see that (4) gives more weight *a priori* to smooth curves: the data is modeled as the sum of a fixed line plus a random wiggly curve distributed as a randomly weighted sum of progressively rougher curves, where *a priori* the weights for smoother curves are bigger. We note that the choice of  $\mathbf{Z}$  and  $\mathbf{u}$  is not uniquely determined; alternative parametrizations may be used provided  $\text{var}(\mathbf{Z}\mathbf{u}) = (\sigma^2/\lambda)\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T$ . When the  $B_i$ ,  $i = 1, \dots, n$  can be chosen so that  $\mathbf{X}^T\mathbf{Z} = \mathbf{0}$ , then by (7) and (8), the fitted curve is the sum of a line fit by OLS plus a regression of the randomly weighted sum of curves on data  $\mathbf{y}$  assumed to have mean zero.

The variance components  $\sigma^2/\lambda$  and  $\sigma^2$  derive from the smoothing parameter  $\lambda$  and error variance  $\sigma^2$  of (2). In the preceding discussion,  $\lambda$  and  $\sigma^2$  are presumed known; whereas in practice,  $\lambda$  and  $\sigma^2$  must be estimated from the data. Theoretical and practical discussions concerning selection of the smoothing parameter  $\lambda$  within the penalized regression framework are abundant in the literature. See Eubank (1988) for a good review. In the penalized regression setting, cross validation and generalized cross validation are the most popular techniques for choosing  $\lambda$ . Note that  $\sigma^2$  is not needed by these methods for estimating  $\lambda$ , and also that it cancels out of the expressions (5) and (6) for  $\hat{\boldsymbol{\beta}}$  and  $\hat{\mathbf{u}}$ . In the context of mixed effect models, it is natural to use maximum likelihood or restricted maximum likelihood (REML) estimation for the variance components  $\sigma^2/\lambda$  and  $\sigma^2$ . Using likelihood based methods for selecting  $\lambda$  is hardly a new idea; according to Speed (1991), REML estimation of  $\lambda$  coincides with the generalized maximum likelihood (GML) procedure discussed in Wahba (1985), which also compares GML to GCV.

We conclude this section with a discussion of the relationships between models (1) and (4). Model (1) is infinite dimensional and model (4) can be viewed as a finite dimensional approximation to it which specifies that  $s(t)$  belongs to the span of a spline basis and also that the coefficients in this basis arise from a particular stochastic model. This finite dimensional Bayesian model was proposed by Silverman (1985) as an alternative to the infinite dimensional Bayesian model of Wahba (1978), in which the smoothing spline (2) is the posterior mean. Eubank (1988) contains extensive discussion and literature references on the spline approximation (2) to  $s(t)$  of (1); generally, smoothing splines are effective

nonparametric estimates and are well understood from practical and theoretical points of view within the context of model (1). The applicability of inferences based on the Bayesian model to the deterministic model (1) is more murky. For example, the coverage properties of Wahba's (1983) Bayesian confidence intervals are not entirely satisfactory; again, we refer to Eubank (1988, Section 5.5) for extensive discussion. In our context, it seems plausible to us to view the multiple curves from multiple women as realizations of stochastic processes, and consequently to use elaborations of model (4) for estimation and inference. This is essentially the point of view taken by Rice and Silverman (1991), who, however, allow greater flexibility in modeling the covariance structure. In any case, we find it useful to note and exploit the algebraic equivalence of the deterministic and stochastic models for computation and we are hopeful that, as has repeatedly been the case in signal processing, procedures derived from stochastic models are effective in a broader realm.

## 2.1 Modeling a Nested Sample of Curves

We now consider a generalization of the regression problem in (2) encompassing observations from a nested sample of curves. The basic idea is to set forth an additive model in which the observations from one curve are of the form

$$y_{il} = s_{g(i)}(t_{il}) + s_{w(i)}(t_{il}) + s_{c(i)}(t_{il}) + \varepsilon_{il}, \quad (9)$$

$l = 1, \dots, n_i$ ,  $i = 1, \dots, N$  where  $s_{g(i)}(t)$ ,  $s_{w(i)}(t)$ , and  $s_{c(i)}(t)$  represent a smooth group mean, smooth subject departure from group mean, and smooth cycle departure from subject mean;  $g$ ,  $w$ ,  $c$  index the three layers of nesting ( $g$  standing for group,  $w$  for woman or subject, and  $c$  for cycle);  $g(i)$ ,  $w(i)$ , and  $c(i)$  represent the group, subject, and cycle comprising the  $i^{\text{th}}$  sample curve observed at  $n_i$  design points; and  $\varepsilon_{il}$ ,  $l = 1, \dots, n_i$ ,  $i = 1, \dots, N$  are i.i.d.  $N(0, \sigma^2)$ . Figure 3 presents results of fitting the generalized model to the progesterone data, depicted for one subject who contributed five nonconceptive cycles. Observe the pleasing graphical summary of the data in terms of a smooth curve analysis of the variance due to group, subject, and cycle effects. Also regard the proficient imputation of missing data for Cycle 2 and for the deleted mid-cycle data of Cycle 5.

Since the appropriate analog of (2) is not *a priori* clear in this context, we borrow from the long history of ANOVA with mixed effects models to suggest one. We then prove that the solutions to the resulting optimization problems are natural cubic splines.

To build a mixed effects model from (9), we work initially with one curve and specify

$$\begin{aligned}
\mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta}_{g(i)} + \mathbf{Z}_i \mathbf{u}_{g(i)} & + & \mathbf{X}_i \boldsymbol{\beta}_{w(i)} + \mathbf{Z}_i \mathbf{u}_{w(i)} \\
\text{curve } i &= \text{smooth group } g(i) \text{ mean} & + & \text{smooth woman } w(i) \text{ deviation} \\
& & + & \mathbf{X}_i \boldsymbol{\beta}_{c(i)} + \mathbf{Z}_i \mathbf{u}_{c(i)} & + & \boldsymbol{\varepsilon}_i \\
& & + & \text{smooth cycle } c(i) \text{ deviation} & + & \text{noise.}
\end{aligned} \tag{10}$$

The design matrices, which unlike those in (4) are subscripted to indicate the time points specific to curve  $i$ , are as follows. Let  $\{\nu_i\}_{i=1}^T$  be the ordered set of  $T$  unique observation times from all the curves taken together, and let  $B_j$ ,  $j = 1, 2, \dots, T$  be a basis of  $S(\nu_2, \dots, \nu_{T-1})$ . As before, let  $B_1$  and  $B_2$  span the linear functions, let the remaining  $B_j$  be orthogonal to them, and let the resulting matrix  $\mathbf{Q}$  have the singular value decomposition  $\mathbf{U}\mathbf{D}\mathbf{U}^T$ . Then take  $\mathbf{X}_i = (B_j(t_{il}))_{jl}$ ,  $j = 1, 2$ ,  $l = 1, \dots, n_i$ ;  $\mathbf{B}_i = (B_j(t_{il}))_{jl}$ ,  $j = 3, \dots, T + 2$ ,  $l = 1, \dots, n_i$ ;  $\mathbf{Z}_i = \mathbf{B}_i \mathbf{U} \mathbf{D}^{-1/2}$ ;  $\mathbf{u}_{g(i)} \sim MVN(\mathbf{0}, (\sigma^2/\lambda) \mathbf{I}_{T-2})$ ;  $\mathbf{u}_{w(i)} \sim MVN(\mathbf{0}, (\sigma^2/\lambda_w) \mathbf{I}_{T-2})$ ;  $\mathbf{u}_{c(i)} \sim MVN(\mathbf{0}, (\sigma^2/\lambda_c) \mathbf{I}_{T-2})$ ;  $\boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ ; and  $\mathbf{u}_{g(i)}$  independent of  $\mathbf{u}_{w(i)}$  independent of  $\mathbf{u}_{c(i)}$  independent of  $\boldsymbol{\varepsilon}_i$ . It is again true that the columns of  $\mathbf{X}_i$  represent the linear and constant functions and are orthogonal to the columns of  $\mathbf{Z}_i$ , and that the columns of  $\mathbf{Z}_i$  corresponding to smoother functions have larger norms.

To complete the specification of the nested smoothing spline model, we stack the curves and specify independence amongst the collection of random vectors:  $\mathbf{u}_{g(i)}$  is independent of  $\mathbf{u}_{g(j)}$  if  $g(i) \neq g(j)$ ;  $\mathbf{u}_{w(i)}$  is independent of  $\mathbf{u}_{w(j)}$  if  $w(i) \neq w(j)$ ;  $\mathbf{u}_{c(i)}$  is independent of  $\mathbf{u}_{c(j)}$  if  $c(i) \neq c(j)$ ; and  $\boldsymbol{\varepsilon}_i$  is independent of  $\boldsymbol{\varepsilon}_j$  for  $i \neq j$ . Additionally,  $\mathbf{u}_{g(i)}$ ,  $\mathbf{u}_{w(j)}$ ,  $\mathbf{u}_{c(k)}$ , and  $\boldsymbol{\varepsilon}_l$  are mutually independent for all  $(i, j, k, l)$ . The resulting model is

$$\mathbf{y} = \mathbf{X}_g \boldsymbol{\beta}_g + \mathbf{Z}_g \mathbf{u}_g + \mathbf{X}_w \boldsymbol{\beta}_w + \mathbf{Z}_w \mathbf{u}_w + \mathbf{X}_c \boldsymbol{\beta}_c + \mathbf{Z}_c \mathbf{u}_c + \boldsymbol{\varepsilon}, \tag{11}$$

where  $\boldsymbol{\beta}_g$ ,  $\boldsymbol{\beta}_w$ , and  $\boldsymbol{\beta}_c$  are column vectors comprising stacked fixed effects vectors:  $\boldsymbol{\beta}_g$  stacks the unique  $\boldsymbol{\beta}_{g(i)}$ ,  $\boldsymbol{\beta}_w$  the  $\boldsymbol{\beta}_{w(i)}$ , and  $\boldsymbol{\beta}_c$  the  $\boldsymbol{\beta}_{c(i)}$ ; similarly, for the random effects vectors the  $\mathbf{u}_{g(i)}$  are stacked in  $\mathbf{u}_g$ , the  $\mathbf{u}_{w(i)}$  in  $\mathbf{u}_w$ , and the  $\mathbf{u}_{c(i)}$  in  $\mathbf{u}_c$ . Correspondingly,  $\mathbf{X}_g$ ,  $\mathbf{X}_w$ , and  $\mathbf{X}_c$  are block diagonal fixed effects design matrices composed of blocks of stacked  $\mathbf{X}_i$ ; similarly,  $\mathbf{Z}_g$ ,  $\mathbf{Z}_w$ , and  $\mathbf{Z}_c$  are block diagonal random effects design matrices comprising blocks of stacked  $\mathbf{Z}_i$ .

As is typical in nested ANOVA models, the fixed effects are not identifiable for each factor level at every layer of the nest. The usual solution is to impose constraints that

the fixed effects sum to zero at each layer save the top-most (group layer); accordingly, we constrain the cycle fixed effects to sum to zero within subjects and the subject fixed effects to sum to zero within groups, and rewrite (10) as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{Z}_i (\mathbf{u}_{g(i)} + \mathbf{u}_{w(i)} + \mathbf{u}_{c(i)}) + \boldsymbol{\varepsilon}_i \quad (12)$$

where  $\boldsymbol{\beta}_i$  is defined to equal  $\boldsymbol{\beta}_{g(i)} + \boldsymbol{\beta}_{w(i)} + \boldsymbol{\beta}_{c(i)}$ . While identifiability issues do not arise for the random effects portion of the model, there is nevertheless a linear dependency among the BLUP estimates for the  $\mathbf{u}_{g(i)}$ ,  $\mathbf{u}_{w(i)}$ , and  $\mathbf{u}_{c(i)}$ . Specifically, the estimates for the  $\mathbf{u}_{g(i)}$  and  $\mathbf{u}_{w(i)}$  are linear functions of the estimates for the  $\mathbf{u}_{c(i)}$ .

Letting  $\mathbf{Z} = (\mathbf{Z}_g \mid \mathbf{Z}_w \mid \mathbf{Z}_c)$ ;  $\mathbf{X}$  denote the block diagonal matrix with blocks  $\mathbf{X}_i$ ;  $\mathbf{u} = (\mathbf{u}_g^T, \mathbf{u}_w^T, \mathbf{u}_c^T)^T$ ; and  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_N^T)^T$ ; the model is seen to belong to the class of mixed effects models addressed by Robinson, and furthermore to the class of variance components models described in detail by Searle, Casella, and McCulloch (1992). The pictorial representation shown below may prove helpful.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \end{pmatrix} = \begin{pmatrix} x_1 & & & & \\ & x_2 & & & \\ & & x_3 & & \\ & & & x_4 & \\ & & & & \ddots \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \vdots \end{pmatrix} +$$

$$\begin{pmatrix} z_1 & | & z_1 & | & z_1 & & & & \\ z_2 & | & z_2 & | & & z_2 & & & \\ & z_3 & | & & z_3 & & & & \\ & & z_4 & | & & & z_4 & & \\ \vdots & | & \vdots & | & \vdots & & \vdots & & \\ \text{groups} & & \text{women} & & \text{cycles} & & & & \end{pmatrix} \begin{pmatrix} u_{g1} \\ u_{g2} \\ u_{w1} \\ u_{w2} \\ u_{w3} \\ u_{c1} \\ u_{c2} \\ u_{c3} \\ u_{c4} \\ \vdots \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \vdots \end{pmatrix}$$

In the pictorial example, the first and second curves  $\mathbf{y}_1$  and  $\mathbf{y}_2$  in the sample belong to group 1, woman 1, and cycles 1 and 2;  $\mathbf{y}_3$  belongs to group 2, woman 2 and cycle 3; and  $\mathbf{y}_4$  to group 2, woman 3 and cycle 4.

Notice that implicit in the nested smoothing spline model is the assumption that curves arising from the same woman or group are correlated, with curves  $i$  and  $j$  having covariance

$$\text{cov}(\mathbf{y}_i, \mathbf{y}_j) = \begin{cases} \mathbf{0} & g(i) \neq g(j) \\ (\sigma^2/\lambda_g) \mathbf{Z}_i \mathbf{Z}_j' & g(i) = g(j), w(i) \neq w(j) \\ (\sigma^2/\lambda_g + \sigma^2/\lambda_w) \mathbf{Z}_i \mathbf{Z}_j' & w(i) = w(j), c(i) \neq c(j) \\ (\sigma^2/\lambda_g + \sigma^2/\lambda_w + \sigma^2/\lambda_c) \mathbf{Z}_i \mathbf{Z}_j' & c(i) = c(j) \end{cases} \quad (13)$$

Hence, related curves are correlated, and the correlation increases with relatedness. Since the affine components are modeled with fixed effects, only the non-affine components of the curves are correlated. This could be remedied, however, by converting the fixed effects into random effects with a suitably large variance, possibly to be estimated from the data as an additional variance component. We choose to maintain the fixed effects for two reasons: 1) estimation is computationally faster and easier with fewer variance components, and 2) our interest lies in a straightforward generalization of the single curve smoothing spline model, whose equivalent mixed effects model requires fixed effects. We should point out that in the limiting case where the variance of a random effect  $u_0$  goes to infinity, the BLUP predictor for  $u_0$  is identical to the BLUP estimator for its fixed effect alternative  $\beta_0$ .

It turns out that the nested mixed effects model (11) has an equivalent representation as a penalized regression, where the fitted nested curves incur penalties of the same variety as the cubic smoothing spline. This is formalized in Theorem 1.

**Theorem 1** The BLUP solutions summarized by  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$  to the nested mixed effects model (11) are equivalent to the fitted curves given by a corresponding penalized regression. The penalized regression fits the collection of smooth curves  $s_{g(i)}(t)$ ,  $s_{w(i)}(t)$ , and  $s_{c(i)}(t)$  in (9) by minimizing the residual sum of squares plus a roughness penalty generalized for nested samples of curves:

$$\begin{aligned} \arg \min & \sum_i \sum_l (y_{il} - (s_{g(i)}(t_{il}) + s_{w(i)}(t_{il}) + s_{c(i)}(t_{il})))^2 \\ & + \lambda_g \sum_k \int (s_{g_k}''(t))^2 dt + \lambda_w \sum_j \int (s_{w_j}''(t))^2 dt + \lambda_c \sum_i \int (s_{c_i}''(t))^2 dt \end{aligned} \quad (14)$$

over the space of smooth curves  $s_{g_k}(t)$ ,  $s_{w_j}(t)$ , and  $s_{c_i}(t) \in \mathcal{W}_2^2[a, b]$ , where  $g_k$  indexes groups or the unique  $g(i)$ ,  $w_j$  indexes women, and  $c_i$  indexes cycles. Moreover, the unique set of curves minimizing (14) is a subset of the natural cubic splines with knots at the collective interior design points,  $(\nu_2, \dots, \nu_{T-1})$ .

For proof see the Appendix.

The parameters  $\lambda_g$ ,  $\lambda_w$ , and  $\lambda_c$  in (14) stem from the variance components  $\sigma^2/\lambda_g$ ,  $\sigma^2/\lambda_w$ , and  $\sigma^2/\lambda_c$  of (10), and can be viewed as hierarchical smoothing parameters. The smoothing parameters  $\lambda_g$ ,  $\lambda_w$ , and  $\lambda_c$  control the amplitude of the non-affine components of the hierarchical smooth fits; the variance components  $\sigma^2/\lambda_g$ ,  $\sigma^2/\lambda_w$ , and  $\sigma^2/\lambda_c$  represent the relative importance of group, woman, and cycle effects in the wiggly portion of the

overall fit. Estimates of the variance components can be viewed as measures of the relative contribution to the overall variation of group, woman, and cycle effects. For instance, if the estimates suggest  $\sigma^2/\lambda_w \ll \sigma^2/\lambda_c$ , they offer evidence that cycles from different women are as similar as cycles from the same woman. If the estimates indicate  $\sigma^2/\lambda_g \ll \sigma^2/\lambda_c$  and  $\sigma^2/\lambda_w \ll \sigma^2/\lambda_c$ , we can simplify our model, dropping the group and subject factors without affecting the smooth curve decomposition of the data. This is equivalent to fitting separate smoothing splines to each curve in the sample, then averaging across cycles within women and women within groups to calculate smooth woman departures from smooth group means.

## 2.2 Modeling a Crossed Sample of Curves

Turning next to a crossed sample of curves, we set forth the model

$$y_{il} = s_\mu(t_{il}) + s_{b(i)}(t_{il}) + s_{\tau(i)}(t_{il}) + \varepsilon_{il}, \quad (15)$$

$l = 1, \dots, n_i$ ,  $i = 1, \dots, N$  where  $s_\mu(t)$ ,  $s_{b(i)}(t)$ , and  $s_{\tau(i)}(t)$  represent a smooth overall mean, smooth block main effect, and smooth treatment main effect;  $\mu$ ,  $b$ , and  $\tau$  index the overall mean plus two crossed factors that stratify the sample ( $b$  standing for block and  $\tau$  for treatment);  $b(i)$  and  $\tau(i)$  represent the block and treatment assigned to the  $i^{\text{th}}$  sample curve observed at  $n_i$  design points; and  $\varepsilon_{il}$ ,  $l = 1, \dots, n_i$ ,  $i = 1, \dots, N$  are i.i.d.  $N(0, \sigma^2)$ .

The model specified by (15) differs from (9) in that the factors are crossed rather than nested: whereas in (9) curves from one woman  $w_j$  can belong only to one group  $g_k$ , in (15), curves assigned to treatment  $\tau_j$  will span all the blocks  $b_k$ . Consequently, the mixed effects model corresponding to (15) will differ from (11) in the structure of the fixed and random effects design matrices, whereas, at the level of a single curve, the mixed effects model will closely resemble (10). For one curve, the model is specified as

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta}_\mu + \mathbf{Z}_i \mathbf{u}_\mu & + & \mathbf{X}_i \boldsymbol{\beta}_{b(i)} + \mathbf{Z}_i \mathbf{u}_{b(i)} \\ \text{curve } i &= \text{smooth overall mean} & + & \text{smooth main effect of block } b(i) \\ & + \mathbf{X}_i \boldsymbol{\beta}_{\tau(i)} + \mathbf{Z}_i \mathbf{u}_{\tau(i)} & + & \boldsymbol{\varepsilon}_i \\ & + \text{smooth main effect of treatment } \tau(i) & + & \text{interactions and noise,} \end{aligned} \quad (16)$$

where the design matrices, the fixed and random effects, and the error term are constructed as in subsection 2.1, except that we label the smoothing parameters as  $\lambda_\mu$ ,  $\lambda_b$ , and  $\lambda_\tau$  to coincide with the labeling of the random effects vectors. We borrow further from that

section and specify likewise independence amongst the collection of random vectors. The resulting model is

$$\mathbf{y} = \mathbf{X}_\mu \boldsymbol{\beta}_\mu + \mathbf{Z}_\mu \mathbf{u}_\mu + \mathbf{X}_b \boldsymbol{\beta}_b + \mathbf{Z}_b \mathbf{u}_b + \mathbf{X}_\tau \boldsymbol{\beta}_\tau + \mathbf{Z}_\tau \mathbf{u}_\tau + \boldsymbol{\varepsilon}, \quad (17)$$

where the fixed and random effects vectors and associated design matrices are again stacked versions of those in (16). As crossed ANOVA models come upon identifiability issues similar to those of their nested counterparts, care must again be taken in estimating the fixed effects. The usual solution is to set the effects belonging to the overall mean and to one level of either the block or treatment factors equal to zero; see Scheffé(1959). Estimates can then be adjusted to reflect constraints such as 1) block main effects summing to zero and 2) treatment main effects summing to zero. It is again true that the BLUP predictors for the random effects satisfy linear constraints.

The covariance of curves in the crossed smoothing spline model is given by

$$\text{cov}(\mathbf{y}_i, \mathbf{y}_j) = \begin{cases} (\sigma^2/\lambda_\mu) \mathbf{Z}_i \mathbf{Z}_j' & b(i) \neq b(j), \tau(i) \neq \tau(j) \\ (\sigma^2/\lambda_\mu + \sigma^2/\lambda_b) \mathbf{Z}_i \mathbf{Z}_j' & b(i) = b(j), \tau(i) \neq \tau(j) \\ (\sigma^2/\lambda_\mu + \sigma^2/\lambda_\tau) \mathbf{Z}_i \mathbf{Z}_j' & \tau(i) = \tau(j), b(i) \neq b(j) \\ (\sigma^2/\lambda_\mu + \sigma^2/\lambda_b + \sigma^2/\lambda_\tau) \mathbf{Z}_i \mathbf{Z}_j' & b(i) = b(j), \tau(i) = \tau(j) \end{cases}$$

Notice that here the covariance has a crossed block matrix structure in contrast to the nested structure of (13). While the formulation in (16), which models the variance of the error term  $\boldsymbol{\varepsilon}_i$  as white noise, is convenient for pedagogic purposes, for practical purposes  $\boldsymbol{\varepsilon}_i$  should probably be modeled as the sum of interaction splines plus white noise. When one curve is measured at each combination  $(b_j, \tau_k)$  of block and treatment levels, we could introduce an interaction spline to model the smooth contribution of block  $b_j$  and treatment  $\tau_k$  beyond that of the sum of two smooth main effects. Our terminology should not be confused with the interaction splines of Wahba (1988). Extension of the ideas and results predicated on (16) is straightforward.

The crossed mixed effects model (17) has its own equivalent representation as a penalized regression, formalized in Theorem 2.

**Theorem 2** Let  $\boldsymbol{\beta} = (\boldsymbol{\beta}_\mu^T, \boldsymbol{\beta}_b^T, \boldsymbol{\beta}_\tau^T)^T$ ;  $\mathbf{u} = (\mathbf{u}_\mu^T, \mathbf{u}_b^T, \mathbf{u}_\tau^T)^T$ ; and  $\mathbf{X}$  and  $\mathbf{Z}$  be the corresponding design matrices. The BLUP solutions summarized by  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$  to the crossed mixed effects model in (17) are equivalent to the fitted curves given by a corresponding penalized regression. The penalized regression fits the collection of smooth curves  $s_\mu(t)$ ,

$s_{b(i)}(t)$ , and  $s_{\tau(i)}(t)$  in (15) by minimizing the residual sum of squares plus a roughness penalty generalized for crossed samples of curves:

$$\begin{aligned} \arg \min \sum_i \sum_l (y_{il} - (s_\mu(t_{il}) + s_{b(i)}(t_{il}) + s_{\tau(i)}(t_{il})))^2 \\ + \lambda_\mu \int (s_\mu''(t))^2 dt + \lambda_b \sum_j \int (s_{b_j}''(t))^2 dt + \lambda_\tau \sum_k \int (s_{\tau_k}''(t))^2 dt \end{aligned} \quad (18)$$

over the space of smooth curves  $s_\mu(t)$ ,  $s_{b_j}(t)$ , and  $s_{\tau_k}(t) \in \mathcal{W}_2^2[a, b]$ , where  $b_j$  indexes blocks or the unique  $b(i)$ , and  $\tau_k$  indexes treatments. Moreover, the unique set of curves minimizing (18) is a subset of the natural cubic splines with knots at the collective interior design points,  $(\nu_2, \dots, \nu_{T-1})$ .

For proof see the Appendix.

It is interesting to find that the form of the penalty in the nested case goes through unchanged to the crossed case; each in the collection of fitted smooth curves is penalized by the normed second derivative squared multiplied by the appropriate smoothing parameter, and then the penalties are summed together with equal weight. The difference between (18) and (14) is somewhat hidden, residing primarily in the nature of the residual sum of squares (RSS) but also in the interplay between the RSS and the penalty. In hindsight, the structure of the two penalized regressions appears sensible: penalties are treated equally at each layer but differ between layers by a multiplicative factor; prospectively, however, the basis for this structure was not apparent. The identification of the mixed effects ANOVA model with the penalized regression provides us with a convincing rationale for both procedures.

### 3. ESTIMATION

Estimation in the smoothing spline models of the previous section can proceed with standard techniques for estimation within the mixed model and variance component framework. The variance parameters such as  $\lambda_g$ ,  $\lambda_w$ ,  $\lambda_c$ , and  $\sigma^2$  are first estimated with ML or REML methodology, then the fixed and random effects are estimated with the BLUP equations by substituting estimates for the variance components. In subsection 3.1 we discuss computational aspects of the estimating procedure, while subsection 3.2 addresses variability of the estimates.

#### 3.1 Computation

The models of Section 2 each have the form of a variance components model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^r \mathbf{W}_j \mathbf{u}_j + \boldsymbol{\varepsilon} \quad (19)$$

with  $\text{var}(\mathbf{u}_j) = (\sigma^2/\lambda_j)\mathbf{I}$ ;  $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ ; and therefore  $\text{var}(\mathbf{y}) = \mathbf{V}(\boldsymbol{\theta}) = \sum(\sigma^2/\lambda_j)\mathbf{W}_j\mathbf{W}_j^T + \sigma^2\mathbf{I}$ , where  $\boldsymbol{\theta} = (\sigma^2, \sigma^2/\lambda_j, j = 1, \dots, r)$  is the vector of variance components. For example, in the nested model  $r = 3$ ; the  $\mathbf{u}_j$  are  $\mathbf{u}_g, \mathbf{u}_w$ , and  $\mathbf{u}_c$ ; the  $\mathbf{W}_j$  are  $\mathbf{Z}_g, \mathbf{Z}_w$ , and  $\mathbf{Z}_c$ ; and the  $\lambda_j$  are  $\lambda_g, \lambda_w$ , and  $\lambda_c$ . We use REML estimation for the variance components, first because it possesses desirable properties as reviewed in Speed (1995), but also because it agrees with Wahba's generalized maximum likelihood procedure.

The computational methods outlined below for REML estimation can easily be transferred to methods for ML estimation, since  $\mathcal{R}(\mathbf{X}) \perp \mathcal{R}(\mathbf{W}_j)$  implies that the likelihood  $p(\mathbf{y})$  can be factored, keeping the variance components separate from the fixed effects. Restricted maximum likelihood is so named because it maximizes the likelihood in  $\boldsymbol{\theta}$  of the restricted data  $\mathbf{x} = P_{\mathcal{R}(\mathbf{X})^\perp} \mathbf{y}$ , viz., the original data projected onto the so-called space of error contrasts,  $\mathcal{R}(\mathbf{X})^\perp$ . For computation of REML estimates, it is helpful to switch to coordinates based on a full rank basis for the subspace  $\mathcal{R}(\mathbf{X})^\perp$ , in order to work with a non-singular normal distribution. This is accomplished by first selecting a matrix  $\mathbf{K}^T$  with the highest rank possible such that  $\mathbf{K}^T \mathbf{X} = \mathbf{0}$  and  $\mathbf{K}^T$  has full row rank, then switching coordinates from  $\mathbf{y}$  to  $\mathbf{K}^T \mathbf{y}$ . A convenient choice for  $\mathbf{K}^T$  takes as its rows the orthonormal eigenvectors of  $\mathbf{X}\mathbf{X}^T$  corresponding to the zero eigenvalues. With this choice,  $\mathbf{K}^T \mathbf{K} = \mathbf{I}_{n-q}$  where  $q$  is the rank of  $\mathbf{X}\mathbf{X}^T$ . For the smoothing spline models, where  $\mathbf{X}$  is block diagonal with blocks  $\mathbf{X}_i$  ( $i$  indexing curves), we can construct  $\mathbf{K}^T$  as block diagonal with blocks  $\mathbf{K}_i^T$  having rows that are the orthonormal eigenvectors of  $\mathbf{X}_i \mathbf{X}_i^T$ .

We use the EM algorithm of Dempster, Laird and Rubin (1977) to maximize the restricted likelihood in the variance components. The complete data is specified as  $\mathbf{K}^T \mathbf{y}$ ;  $\mathbf{u}_j$ ,  $j = 1, \dots, r$ ; and  $\boldsymbol{\varepsilon}$  (using  $\mathbf{K}^T \boldsymbol{\varepsilon}$  gives the same results); the complete data sufficient statistics for the variance components are thus  $\mathbf{t}_0 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$  (for  $\sigma^2$ ) and  $\mathbf{t}_j = \mathbf{u}_j^T \mathbf{u}_j$  (for  $\sigma^2/\lambda_j$ ),  $j = 1, \dots, r$ . Letting  $\mathbf{V}_K$  denote  $\text{var}(\mathbf{K}^T \mathbf{y})$ , E-step expectations of the  $\mathbf{t}_j$  conditional on the actual data  $\mathbf{K}^T \mathbf{y}$  are computed as

$$\begin{aligned} E(\mathbf{t}_0 | \mathbf{K}^T \mathbf{y}) &= \text{tr}(\text{var}(\boldsymbol{\varepsilon} | \mathbf{K}^T \mathbf{y})) + E(\boldsymbol{\varepsilon} | \mathbf{K}^T \mathbf{y})^T E(\boldsymbol{\varepsilon} | \mathbf{K}^T \mathbf{y}) \\ &= \text{tr}(\sigma^2 \mathbf{I} - \sigma^4 \mathbf{K} \mathbf{V}_K^{-1} \mathbf{K}^T) + \sigma^4 \mathbf{y}^T \mathbf{K} \mathbf{V}_K^{-2} \mathbf{K}^T \mathbf{y} \end{aligned}$$

$$\begin{aligned}
E(t_j | \mathbf{K}^T \mathbf{y}) &= \text{tr}((\sigma^2/\lambda_g) \mathbf{I} - (\sigma^2/\lambda_g)^2 \mathbf{W}_j^T \mathbf{K} \mathbf{V}_K^{-1} \mathbf{K}^T \mathbf{W}_j) \\
&\quad + (\sigma^2/\lambda_g)^2 \mathbf{y}^T \mathbf{K} \mathbf{V}_K^{-1} \mathbf{K}^T \mathbf{W}_j \mathbf{W}_j^T \mathbf{K} \mathbf{V}_K^{-1} \mathbf{K}^T \mathbf{y}, \quad j = 1, \dots, r
\end{aligned}$$

Letting  $m_j$  denote the dimension of  $\mathbf{u}_j$ , and recalling that  $n$  is the dimension of  $\boldsymbol{\varepsilon}$ , the M-step maximum likelihood solutions are  $\hat{\sigma}^2 = t_0/n$  and  $(\hat{\sigma}^2/\lambda_j) = t_j/m_j$ . The EM algorithm alternates the E- and M-steps in an iterative fashion. We see that the computational requirements of the E-step depend on the orders of 1) the multiplications of  $\mathbf{V}_K^{-1}$  and  $\mathbf{K}^T \mathbf{W}_j$  and 2) the inversions  $\mathbf{V}_K^{-1}$ . In our models, these matrices are typically very large and do not possess one of the patterned structures specially accommodated by existing routines such as SAS Proc Mixed, which therefore spend at each iteration an inordinate amount of time blindly inverting the covariance matrix. For the sample of 91 progesterone curves,  $\mathbf{V}_K^{-1}$  has approximate size  $2000 \times 2000$ , requiring approximately 32 megabytes (M) of storage. The amount of time required for one inversion depends heavily upon the random access memory (RAM) capacity of the computer relative to the size of the matrix; even the fastest computers will be overwhelmed by memory swapping if the supply of RAM is inadequate. For a computer equipped with plenty of RAM, such as our Sparc 2000 Data Center with one gigabyte RAM, just one inversion of a  $2000 \times 2000$  matrix will still take 15 minutes. Small increases in the number of progesterone curves result in dramatic increases in storage and time requirements. For instance, a 50% increase in the number of progesterone curves produces a  $3000 \times 3000$  REML covariance matrix requiring 72M of storage and two hours to invert. The computational burden for each E-step iteration is therefore very heavy, motivating our search for a more efficient algorithm. We now describe two simplifying approaches. Our computer programs were written in Matlab and executed on the Sparc 2000 Data Center.

Requiring that the smoothing parameters for different factors are all equal leads to substantial computational savings. For the nested model in (11), the requirement implies that the *a priori* variance of the smooth curves corresponding to group mean, subject departure from group mean, and cycle departure from subject mean are all equal. For the crossed model in (17), it implies that the variances associated with the overall mean, block main effects, and treatment main effects are all equal. A preliminary check of these assumptions can be made based on subsets of the data small enough for estimation within the unsimplified models and on asymptotic or bootstrap estimates of variability; refer to Section 4 for an example. Our first computational method requires equal smoothing pa-

rameters:  $\lambda_j = \lambda$  for all  $j$ . When this assumption holds,  $\mathbf{V}_K = (\sigma^2/\lambda)\mathbf{K}^T\mathbf{Z}\mathbf{Z}^T\mathbf{K} + \sigma^2\mathbf{I}$ , where  $\mathbf{Z} = (\mathbf{W}_1 | \dots | \mathbf{W}_r)$ . Therefore, the eigenvectors  $\boldsymbol{\xi}_k$  of  $\mathbf{V}_K$  are the same as those of  $\mathbf{K}^T\mathbf{Z}\mathbf{Z}^T\mathbf{K}$ , and the eigenvalues of  $\mathbf{V}_K$  are  $(\sigma^2/\lambda)d_k + \sigma^2$ , where  $d_k$  is the eigenvalue of  $\mathbf{K}^T\mathbf{Z}\mathbf{Z}^T\mathbf{K}$  corresponding to  $\boldsymbol{\xi}_k$ . Furthermore, since there are only two variance components ( $\sigma^2$  and  $\sigma^2/\lambda$ ), only two complete data sufficient statistics are necessary,  $t_0 = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$  and  $t_1 = \mathbf{u}^T\mathbf{u}$ , where  $\mathbf{u}^T = (\mathbf{u}_1^T, \dots, \mathbf{u}_r^T)$ . The E-step becomes

$$\begin{aligned} E(t_0|\mathbf{K}^T\mathbf{y}) &= \sigma^2n - \sigma^2\lambda \sum_k 1/(d_k + \lambda) + \lambda^2 \sum_k (\mathbf{y}^T\mathbf{K}\boldsymbol{\xi}_k)^2/(d_k + \lambda)^2 \\ E(t_1|\mathbf{K}^T\mathbf{y}) &= (\sigma^2/\lambda)(\sum m_j) - (\sigma^2/\lambda) \sum_k d_k/(d_k + \lambda) + \sum_k (\mathbf{y}^T\mathbf{K}\boldsymbol{\xi}_k)^2 d_k/(d_k + \lambda)^2. \end{aligned}$$

In words, after a one time eigenvector decomposition of  $\mathbf{V}_K$  and multiplication of  $\mathbf{y}^T\mathbf{K}\boldsymbol{\xi}_k$ , successive iterations of the EM algorithm require nothing but a series of quick dot products and scalar multiplications. The time to compute each iteration with a REML covariance matrix of approximate size  $2000 \times 2000$  is reduced to a split second even for computers with small capacity RAM; the price is a preliminary eigenvector decomposition taking about 2 hours. Recalling the size of the equations for the full-blown approach and the time required for just one inversion, it is clear that the computational savings are immense in both the space and time dimensions. It is generally true that for REML estimation, models with only two variance components yield to this kind of reduction. However, it is not usually true for ML estimation.

The results of our progesterone data analysis, partially displayed in Figure 3, were computed using equal smoothing parameters and the above algorithm. The total computational time required is about 1.5 hours. Since the covariance matrix of the progesterone data is block diagonal comprising two blocks of size  $1500 \times 1500$  and  $500 \times 500$  corresponding respectively to the nonconceptive and conceptive groups, we save much time by performing separate eigenvector decompositions for each block and combining the results; this takes approximately one hour. The EM algorithm then computes REML estimates in about 30 seconds using hundreds of iterations, more than are adequate for convergence. Finally, BLUP estimates of the fitted curves are computed with a sparse backsolve algorithm in approximately 20 minutes.

The fitted curves in Figure 3 were computed using the REML estimate for  $\lambda$ . For comparison, we also computed fitted curves for other  $\lambda$ . The effects on the fitted curves associated with Subject 11 Cycle 2 are shown in Figure 5. Observe in particular the effects

on the fit for the missing data after day 3. For smaller  $\lambda$ , there is less pressure on the cycle fit to track the fitted woman and group means, and thus the optimal fit has more flexibility to minimize the penalty on its second derivative by tending towards a straight line after day 3 with slope equal to that at day 3. At the same time, reducing  $\lambda$  results in cycle fits that more closely track non-missing cycle data. Increasing  $\lambda$  causes group, woman, and cycle fits to more closely resemble ordinary least squares straight line fits.

If the requirement of equal smoothing parameters is not reasonable, but the design points are the same for all curves, then computation at the E-step can still be reduced to some extent. In this case, the  $\mathbf{Z}_i$  and  $\mathbf{X}_i$  from Section 2 are the same for all curves  $i$ , and thus so are the  $\mathbf{K}_i$  defined in this section. Defining  $\mathbf{Z}_0 = \mathbf{Z}_i$ ,  $\mathbf{X}_0 = \mathbf{X}_i$ , and  $\mathbf{K}_0 = \mathbf{K}_i$ , it follows that the  $\mathbf{W}_j$  in (19) are block matrices comprising equal size blocks of repeated  $\mathbf{Z}_0$  and zero matrices; hence  $\mathbf{V}_K$  can be written

$$\mathbf{V}_K = \sum_j (\sigma^2/\lambda_j) \mathbf{A}_j \otimes \mathbf{D}, \quad (20)$$

where  $\otimes$  is the Kronecker product,  $\mathbf{A}_j$  is a matrix of ones and zeros representing indicators of factor levels ( $\mathbf{A}_j$  is a direct sum of matrices of ones), and  $\mathbf{D} = \mathbf{K}_0^T \mathbf{Z}_0 \mathbf{Z}_0^T \mathbf{K}_0$ . Therefore, the eigenvectors of  $\mathbf{V}_K$  are given by  $\boldsymbol{\xi}_{ik} = \boldsymbol{\eta}_i(\sigma^2, \lambda_1, \dots, \lambda_r) \otimes \boldsymbol{\gamma}_k$ , with  $\boldsymbol{\eta}_i(\sigma^2, \lambda_1, \dots, \lambda_r)$  representing the eigenvectors of  $\sum_j (\sigma^2/\lambda_j) \mathbf{A}_j$ , and  $\boldsymbol{\gamma}_k$  those of  $\mathbf{D}$ . The  $\mathbf{K}^T \mathbf{W}_j$  can also be written in a Kronecker product form compatible for multiplication with (20). Consequently, only the  $\boldsymbol{\eta}_i(\sigma^2, \lambda_1, \dots, \lambda_r)$  and the eigenvalues of  $\mathbf{V}_K$  need updating at each iteration of the E-step, and the order of the multiplications and inversions is reduced. For instance, for a sample of 100 curves each measured at the same 25 design points, the matrix  $\sum_j (\sigma^2/\lambda_j) \mathbf{A}_j$  would be  $100 \times 100$  (one eigenvector decomposition taking two seconds), and  $\mathbf{D}$   $25 \times 25$ . For designs that are balanced at the level of curves (in addition to that of design points), the computational savings can be stretched even further due to the associated simple structure of  $\sum_j (\sigma^2/\lambda_j) \mathbf{A}_j$ .

When the design points are not the same for all curves, as for instance when there is missing data,  $\mathbf{V}_K$  does not take a simplified Kronecker product form. A possibility for reducing the burden at each E-step is to place the simplified EM algorithm for no data missing within an “outer loop” EM algorithm that updates estimates for the missing data at each iteration. The resultant nested algorithm would give only approximate REML estimates; see Tanner (1991) for a helpful discussion of such approximate EM algorithms.

As a last resort, the variance components might be estimated from a subset of the data; Searle et al. (1992) mention a possibility for combining estimates from different subsets of the data.

### 3.2 Variability

Asymptotic variances of the REML estimates for the variance components of (19) can be obtained analytically with the equations derived in Searle (1970), where  $\text{var}(\hat{\boldsymbol{\theta}}_{\text{REML}}) = 2\mathbf{M}^{-1}$ ,  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_r) = (\sigma^2, \sigma^2/\lambda_1, \dots, \sigma^2/\lambda_r)$ ,  $\mathbf{M}$  is an  $r \times r$  matrix having elements  $m_{jk} = \text{tr}(\mathbf{P}\mathbf{W}_j\mathbf{W}_j^T\mathbf{P}\mathbf{W}_k\mathbf{W}_k^T)$ ,  $\mathbf{P} = \mathbf{K}\mathbf{V}_K^{-1}\mathbf{K}^T$ , and  $\mathbf{W}_0 = \mathbf{I}_n$ . The relevance of these asymptotic formulae is perhaps questionable and can be determined by the adherence to certain assumptions as mentioned in Miller (1977). We use them nonetheless to provide quick but crude measures of the reliability of findings suggested by variance component estimates from the nested model with unconstrained smoothing parameters. These estimates were obtained from two small subsets of the progesterone data and are presented along with their asymptotic standard deviations in Table 1. The table suggests that constraining the smoothing parameters to be equal is a reasonable simplification, as the three estimates are close relative to their estimated standard deviations. It also offers evidence that cycles vary more between women than they do within, as the estimate of  $\sigma^2/\lambda_w$  is further than two standard deviations from zero. Similarly, regarding the estimate of  $\sigma^2/\lambda_g$  as a univariate summary measure of the degree to which nonconceptive and conceptive cycles can be distinguished from each other, we find evidence that the two groups differ. However, this measure incorporates post-implantation observations of progesterone that are well-known to be higher for conceptive cycles. A more interesting assessment is based on the bias and variability of the fitted curves themselves. Were the variance components known, we could obtain analytically the mean squared errors for the BLUP estimates, conditional on realized values of the random effects. In practice, the variance components are unknown. One might approximate the variance for known variance components by substituting the variance component estimates. It is well established that this tends to downwardly bias the resultant estimates for the variance of the fit. Freedman and Peters (1984) give details in the context of approximate generalized least squares.

For this reason, we employ bootstrap techniques — see Efron and Tibshirani (1993) for a broad review — to study the reliability of our estimated smooth group means. Many

versions of the bootstrap are possible for use with the smoothing spline models; these vary by 1) degree of belief in the model (see Laird and Louis (1987) for an interesting discussion), and 2) estimand under study. However, bootstrap algorithms using hierarchical resampling schemes such as the two we describe have not been addressed in the literature and remain an interesting topic for further study.

First we outline a nonparametric version to study variability of fitted group means in the nested model. To make one bootstrap sample, construct a new sample of cycle curves stratified by subjects and groups in the following manner.

1. Draw a sample of size  $G_k$  with replacement from the  $G_k$  women in group  $k$ .
2. For the  $j^{th}$  woman selected into the above sample, draw a sample of size  $W_j$  with replacement from the  $W_j$  cycle curves she contributed to the original sample.

Since this method of resampling creates samples with differing patterns of missing data, computation of REML and BLUP estimates requires multiple eigenvector decompositions.

We therefore implemented a partially parametric version of the bootstrap to maintain the pattern of missing data present in the original sample. The computational time required to construct 35 bootstrap samples and estimate fitted group means using our efficient algorithm with a shared eigenvector decomposition and sparse backsolve is about 45 minutes. From the original fit we have the following objects:

1. The fitted group means  $\hat{s}_{g_k}$  for all groups in the original sample, indexed by  $k$ .
2. The fitted woman departures  $\hat{s}_{w_j}$  for all women in the original sample, indexed by  $j$ .
3. The fitted cycle departures  $\hat{s}_{c_i}$  for all cycles in the original sample, indexed by  $i$ .
4. A collection of residuals  $\hat{\varepsilon}_{il}$  indexed by the  $l^{th}$  design point for cycle  $i$ .

For each group, we build a sample in the following way. Consider a woman in that group. Construct for her a bootstrap mean by adding to the fitted group mean a random selection from (2). For each cycle she contributed to the original sample, build a bootstrap analogue by adding to her bootstrap mean a random selection from (3) and a random set of residuals drawn from (4). We evaluate the result at the same observation times as occurred in the original cycle, in order to maintain the same pattern of missing data. For each bootstrap sample so obtained we repeat the REML and BLUP procedures.

We remark that there are many unresolved issues concerning the relationship of our bootstrap probability mechanism to the real world mechanism that generated the original data. First, the procedure is partially parametric in that the fitted group means are taken as “truth”. Second, the objects in (2) and (3) are posterior means and as such are smoother than samples from the prior distribution of the model. Third, the residuals in (4) contain bias, unlike the random errors of the model. Related to the third issue but somewhat counterintuitive to the second is a phenomenon we noticed in which estimates of  $\lambda$  from the bootstrap samples are higher than the REML estimate from the original sample. Despite possible problems intrinsic to our procedure, we found that bootstrap estimates of fitted group means appear similar to the original estimates.

Fitted group means computed from the progesterone data are presented in Figure 6 and overlaid by a rough comparison based on daily means. It is reassuring that the qualitative differences revealed by the smoothing spline model agree with those summarized in the rough comparison. Particularly interesting for our colleagues in biomedical research is the comparison over the time between ovulation day and day 8. Two previous studies presented findings showing progesterone production to be higher in conceptive than in non-conceptive cycles during this time. The discrepancy created by our results may possibly be explained by our use of urinary rather than serum samples in the analysis (Stewart, Nakajima, Overstreet and Lasley (1993) used serum), and also by the nature of our sample — we do not have adequate data to make a paired comparison using paired nonconceptive and conceptive cycles from the same women (Baird et al. (1996) used paired data). However, the previous studies did not address the potential problem of multiple comparisons; several daily comparisons were made at the .05 significance level.

We used the bootstrap simulations to investigate the reliability of our finding that on average, progesterone production between ovulation and implantation is lower in conceptive than in non-conceptive cycles. See Figure 7 for an example of our results. Tabulation of results from 35 bootstrap samples found only 16 graphs (46%) with the nonconceptive (solid line) estimate lying above the conceptive estimate during days +2 to +8 as in Figure 6, strongly suggesting (presuming that the bootstrap procedure gives reasonable results) that this feature is not significant.

#### 4. SUMMARY

We have used the correspondence between penalized regression and mixed effects mod-

els to extend the smoothing spline model for individual curves to encompass samples of curves stratified by nested and crossed factors. We presented equivalent formulations of the generalized procedure in terms of both penalized regression and mixed effects models, and highlighted the relationship between the associated generalized smoothing parameters and variance components. Particular attention was given to the correlation structure inherent in the generalized models and its impact on smooth curve estimates, especially with respect to data imputation. An efficient method for computation of parameter estimates using a preliminary eigenvector decomposition was described along with asymptotic and bootstrap approaches to determining variability of the estimates and resultant fitted smooth curves. The exposition was illustrated with analyses of the urinary progesterone data, a sample of curves implicitly stratified by subjects nested within conceptive and nonconceptive groups.

Our research has generated interesting statistical questions that remain unanswered. Primary among these is our conjecture that the bootstrap procedure employing a hierarchical resampling scheme leads to an assessment of variability that is representative of the actual variability. We are also curious about how our REML estimates of the smoothing parameters correspond to those that a cross-validated determination would give. The development of computer programs for implementing our approximate nested EM is a task left open. There is also the possibility to extend our models in various directions. One such direction takes the smoothing parameters to be different for each estimated curve. In another direction one might incorporate in the style of linear regression an explicit modeling of categorical and continuous covariates in place of our analysis of variance framework for categorical covariates. A model with three continuous covariates  $x_1$ ,  $x_2$ , and  $x_3$  observed for each curve  $i$  might be written as

$$y_{il} = x_1(i)s_1(t_{il}) + x_2(i)s_2(t_{il}) + x_3(i)s_3(t_{il}) + \varepsilon_{il},$$

with  $s_1(t)$ ,  $s_2(t)$ , and  $s_3(t)$  representing three underlying smooth curves analogous to the parameters of a linear regression. Categorical covariates could be included using dummy variables to produce a broad class of models encompassing those of this article.

## APPENDIX: PROOFS

Proof of Theorem 1

We need the following lemmas.

*Lemma 1*

Suppose  $n \geq 2$  and that  $t_1 < \dots < t_n$ . Given any values  $y_1, \dots, y_n$ , there is a unique natural cubic spline  $\check{s}$  with knots at the points  $t_i$  satisfying

$$\check{s}(t_i) = y_i \text{ for } i = 1, \dots, n.$$

For proof see Theorems 2.1 and 2.2 in Green and Silverman (1994) chapter 2.

*Lemma 2*

Suppose  $n \geq 2$  and that  $a < t_1, \dots, t_n < b$ . There is a unique function  $\check{s}$  in  $\mathcal{W}_2^2[a, b]$  that minimizes  $\int (s''(t))^2 dt$  subject to  $s(t_i) = y_i$ ,  $i = 1, \dots, n$ ; moreover,  $\check{s}$  is a natural cubic spline with knots at the points  $t_i$ .

For proof see Theorem 2.3 in Green and Silverman (1994) chapter 2.

*Lemma 3*

Adding to the notation of Section 2.1, let  $\mathbf{X}(t) = (B_1(t), B_2(t))$ ,  $\mathbf{B}(t) = (B_3(t), \dots, B_T(t))$ ,  $\mathbf{B}''(t) = (B_3''(t), \dots, B_T''(t))$ ,  $\mathbf{Z}(t) = \mathbf{B}(t)\mathbf{U}\mathbf{D}^{-1/2}$ , and  $\mathbf{Z}''(t) = \mathbf{B}''(t)\mathbf{U}\mathbf{D}^{-1/2}$ . Furthermore, let  $(\alpha_1, \alpha_2)$  be defined through the one-to-one correspondence given by  $s(t) = \mathbf{X}(t)\alpha_1 + \mathbf{Z}(t)\alpha_2$  between functions  $s(t)$  in  $\mathcal{S}(\nu_2, \dots, \nu_{T-1})$  and pairs of coefficient vectors in  $\mathfrak{R}^2 \times \mathfrak{R}^{T-2}$ . Then

$$\int (s''(t))^2 dt = \alpha_2^T \alpha_2.$$

Proof:

$$\begin{aligned} s''(t) &= \mathbf{Z}''(t)\alpha_2 \text{ since } B_1''(t) = B_2''(t) = 0. \text{ Therefore } \int (s''(t))^2 dt = \int (\mathbf{Z}''(t)\alpha_2)^2 dt = \\ &= \int (\alpha_2^T \mathbf{Z}''(t)^T \mathbf{Z}''(t) \alpha_2) dt = \alpha_2^T \left( \int \mathbf{D}^{-1/2} \mathbf{U}^T \mathbf{B}''(t)^T \mathbf{B}''(t) \mathbf{U} \mathbf{D}^{-1/2} dt \right) \alpha_2 = \\ &= \alpha_2^T \left( \mathbf{D}^{-1/2} \mathbf{U}^T \mathbf{Q} \mathbf{U} \mathbf{D}^{-1/2} \right) \alpha_2 = \alpha_2^T \left( \mathbf{D}^{-1/2} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{U}^T \mathbf{U} \mathbf{D}^{-1/2} \right) \alpha_2 = \alpha_2^T \alpha_2. \quad \square \end{aligned}$$

Denote the functions minimizing (14) by  $\check{s}_{g_k}(t)$ ,  $k = 1, \dots, N_G$ ,  $\check{s}_{w_j}(t)$ ,  $j = 1, \dots, N_W$ , and  $\check{s}_{c_i}$ ,  $i = 1, \dots, N$ , where  $N_G$ ,  $N_W$ , and  $N$  are respectively the number of different groups, women, and cycles in the sample. By Lemma 1, there exist a unique collection of natural cubic splines  $\check{s}_{g_k}(t)$ ,  $k = 1, \dots, N_G$ ,  $\check{s}_{w_j}(t)$ ,  $j = 1, \dots, N_W$ , and  $\check{s}_{c_i}(t)$ ,  $i = 1, \dots, N$  with knots at  $(\nu_2, \dots, \nu_{T-1})$  such that  $\check{s}_{g_k}(\nu_l) = \check{s}_{g_k}(\nu_l)$ ,  $\check{s}_{w_j}(\nu_l) = \check{s}_{w_j}(\nu_l)$ , and  $\check{s}_{c_i}(\nu_l) = \check{s}_{c_i}(\nu_l)$  for  $l = 2, \dots, T-1$ . By Lemma 2,  $\lambda_g \int (\check{s}_{g_k}''(t))^2 dt \leq \lambda_g \int (\check{s}_{g_k}''(t))^2 dt$  for each  $k$ ,  $\lambda_w \int (\check{s}_{w_j}''(t))^2 dt \leq \lambda_w \int (\check{s}_{w_j}''(t))^2 dt$  for each  $j$ , and  $\lambda_c \int (\check{s}_{c_i}''(t))^2 dt \leq \lambda_c \int (\check{s}_{c_i}''(t))^2 dt$  for each  $i$ , with equality in any case only if  $\check{s} = \tilde{s}$ . Therefore, the solutions to (14) are unique and belong to the class of natural cubic spline functions having knots at the collective interior design points,  $(\nu_2, \dots, \nu_{T-1})$ .

In fact, the solutions to (14) belong to a smaller class of natural cubic splines. Let  $\phi$

index the strata corresponding to the set of factor levels  $\{g_k, k = 1, \dots, N_G, w_j, j = 1, \dots, N_W, \text{ and } c_i, i = 1, \dots, N\}$ ,  $\lambda(\phi)$  denote the corresponding smoothing parameter, and  $(\nu_2, \dots, \nu_{T(\phi)-1})$  denote the interior design points for stratum  $\phi$ . By Lemma 1, for each  $\phi$  there exists a unique natural cubic spline  $\check{s}_\phi(t)$  with knots at  $(\nu_2, \dots, \nu_{T(\phi)-1})$  such that  $\check{s}_\phi(t) = \tilde{s}_\phi(t)$  for  $t \in \{\nu_i\}_{i=2}^{T(\phi)-1}$ . By Lemma 2,  $\lambda(\phi) \int (\check{s}_\phi''(t))^2 dt \leq \lambda(\phi) \int (\tilde{s}_\phi''(t))^2 dt$ , with equality only if  $\check{s}_\phi = \tilde{s}_\phi$ . Therefore, the solutions to (14), indexed by  $\phi$ , belong to the class of natural cubic spline functions having knots at the relevant interior design points,  $(\nu_2, \dots, \nu_{T(\phi)-1})$ .

We now show that the solutions to (14) sampled at the appropriate design points are the BLUP solutions summarized by  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$ . By Lemma 3, functions  $s(t) \in \mathcal{S}(\nu_2, \dots, \nu_{T-1})$  can be written as  $\mathbf{X}(t)\boldsymbol{\alpha}_1 + \mathbf{Z}(t)\boldsymbol{\alpha}_2$  with  $\int (s''(t))^2 dt = \boldsymbol{\alpha}_2^T \boldsymbol{\alpha}_2$ . Therefore,

1.  $\sum_i \sum_l (y_{il} - (s_{g(i)}(t_{il}) + s_{w(i)}(t_{il}) + s_{c(i)}(t_{il})))^2$  can be written as  $\sum_i \|\mathbf{y}_i - (\mathbf{X}_i(\boldsymbol{\beta}_{g(i)} + \boldsymbol{\beta}_{w(i)} + \boldsymbol{\beta}_{c(i)}) + \mathbf{Z}_i(\mathbf{u}_{g(i)} + \mathbf{u}_{w(i)} + \mathbf{u}_{c(i)}))\|^2$ , letting  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  represent  $\mathbf{X}(t)$  and  $\mathbf{Z}(t)$  sampled at the design points specific to curve  $i$  as in Section 2.1, while
2.  $\lambda_g \int (s_{g_k}''(t))^2 dt$ ,  $\lambda_w \int (s_{w_j}''(t))^2 dt$ , and  $\lambda_c \int (s_{c_i}''(t))^2 dt$  can be written respectively as  $\lambda_g \sum_k \mathbf{u}_{g_k}^T \mathbf{u}_{g_k}$ ,  $\lambda_w \sum_j \mathbf{u}_{w_j}^T \mathbf{u}_{w_j}$ , and  $\lambda_c \sum_i \mathbf{u}_{c_i}^T \mathbf{u}_{c_i}$ .

Collecting terms into  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\boldsymbol{\beta}$ ,  $\mathbf{u}_g$ ,  $\mathbf{u}_w$ ,  $\mathbf{u}_c$ , and  $\mathbf{u}$  as in Section 2.1, we have shown so far that the solutions to (14) can be summarized by  $\mathbf{X}\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{u}^*$  where  $\boldsymbol{\beta}^*$  and  $\mathbf{u}^*$  are the solutions to

$$\arg \min \|\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})\|^2 + \lambda_g \mathbf{u}_g^T \mathbf{u}_g + \lambda_w \mathbf{u}_w^T \mathbf{u}_w + \lambda_c \mathbf{u}_c^T \mathbf{u}_c. \quad (21)$$

Defining  $p(\mathbf{y}|\mathbf{u})$  as a MVN density with mean  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$  and variance  $\sigma^2 \mathbf{I}$ , and  $p(\mathbf{u})$  as a MVN density with mean  $\mathbf{0}$  and variance  $\sigma^2/\lambda_g \mathbf{I}_{N_G} \oplus \sigma^2/\lambda_w \mathbf{I}_{N_W} \oplus \sigma^2/\lambda_c \mathbf{I}_N$  (where  $\oplus$  is the direct sum operation for matrices), the solutions to (21) are those to  $\arg \min -2 \log p(\mathbf{y}|\mathbf{u})p(\mathbf{u})$  and hence BLUP (Lindley and Smith 1972; Robinson 1991) for the model in (11). Therefore the BLUP fitted curves for (11) summarized by  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$  are equivalent to the fitted curves given by the penalized regression in (14).  $\square$

Proof of Theorem 2

The proof of Theorem 2 is very similar to the proof of Theorem 1 and thus is omitted.

## REFERENCES

- Anderson, S., and Jones, R. (1995), "Smoothing Splines for Longitudinal Data," *Statistics in Medicine*, 14, 1235-1248.
- Baird, D., Wilcox, A., Weinberg, C., Kamel, F., Musey, P., and Collins, C. (1996), "Preimplantation Estrogen and Progesterone Profiles in Conception and Nonconception Menstrual Cycles," *Human Reproduction*, to appear.
- Barry, D. (1996), "An Empirical Bayes Approach to Growth Curve Analysis," *The Statistician*, 45, 3-19.
- Demmler, A., and Reinsch, C. (1975), "Oscillation Matrices with Spline Smoothing," *Numerische Mathematik*, 24, 375-382.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-22.
- Dempster, A.P., Selwyn, M.R., Patel, C.M., and Roth, A.J. (1984), "Statistical and Computational Aspects of Mixed Model Analysis," *Applied Statistics*, 33, 203-214.
- Efron, B., and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, London, U.K.: Chapman & Hall.
- Eubank, R.L. (1988), *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.
- Freedman, D.A., and Peters, S.S. (1984), "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association*, 79, 97-106.
- Green, P.J., and Silverman, B.W. (1994), *Nonparametric Regression and Generalized Linear Models*, London, U.K.: Chapman & Hall.
- Hastie, T.J., and Tibshirani, R.J. (1990), *Generalized Additive Models*, London, U.K.: Chapman & Hall.
- Kimeldorf, G.S., and Wahba, G. (1970), "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *The Annals of Mathematical Statistics*, 41, 495-502.
- Laird, N.M., and Louis, T.A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," *Journal of the American Statistical Association*, 82, 739-750.
- Lindley, D.V., and Smith, A.F.M. (1972), "Bayes estimates for the linear model (with discussion)," *Journal of the Royal Statistical Society, Ser. B*, 34, 1-41.
- Miller, J.J. (1977), "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," *The Annals of Statistics*, 5, 746-762.
- Munro, C., Stabenfeldt, G., Cragun, J., Addiego, L., Overstreet, J., and Lasley, B. (1991), "Relationship of Serum Estradiol and Progesterone Concentrations to the Excretion Profiles of Their Major Urinary Metabolites as Measured by Enzyme Immunoassay and Radioimmunoassay," *Clinical Chemistry*, 37, 838-844.

- Potthoff, R.F., and Roy, S.N. (1964), "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems," *Biometrika*, 51, 313-326.
- Ramsay, J.O., and Dalzell, C.J. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Ser. B, 53, 539-561.
- Rao, C.R. (1965), "The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves," *Biometrika*, 52, 447-458.
- Rice, J.A., and Silverman, B.W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233-243.
- Robinson, G.K. (1991), "That BLUP is a Good Thing: The Estimation of Random Effects (with discussion)," *Statistical Science*, 6, 15-32.
- Scheffé, H. (1959), *The Analysis of Variance*, New York: John Wiley.
- Searle, S.R. (1970), "Large Sample Variances of Maximum Likelihood Estimators of Variance Components Using Unbalanced Data," *Biometrics*, 26, 505-524.
- Searle, S.R., Casella, G., and McCulloch, C.E. (1992), *Variance Components*, New York: John Wiley.
- Silverman, B.W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," *Journal of the Royal Statistical Society*, Ser. B, 47, 1-21.
- Speed, T.P. (1991), "Comment on 'That BLUP is a Good Thing: The Estimation of Random Effects'," *Statistical Science*, 6, 42-44.
- Speed, T.P. (1995), "REML: A Brief Review," Statistics Research Report No. SRR 004-95, The Australian National University, Centre for Mathematics and its Applications.
- Stewart, D., Nakajima, S., Overstreet, J., and Lasley, B. (1993), "Enhanced Ovarian Steroid Secretion Prior to Implantation in Early Human Pregnancy," *Journal of Clinical Endocrinology*, 76, 1470-1476.
- Tanner, M.A. (1991), *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, New York: Springer-Verlag.
- Thompson, E.A., and Shaw, R.G. (1990), "Pedigree Analysis for Quantitative Traits: Variance Components without Matrix Inversion," *Biometrics*, 46, 399-413.
- Wahba, G. (1978), "Improper Priors, Spline Smoothing, and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 364-372.
- Wahba, G. (1985), "A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem," *The Annals of Statistics*, 13, 1378-1402.
- Wahba, G. (1988), "Partial and interaction spline models," *Bayesian Statistics 3*, 479-491.
- Yen, S., and Jaffe, R.B. (1991), *Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management*, Philadelphia: W.B. Saunders Company.

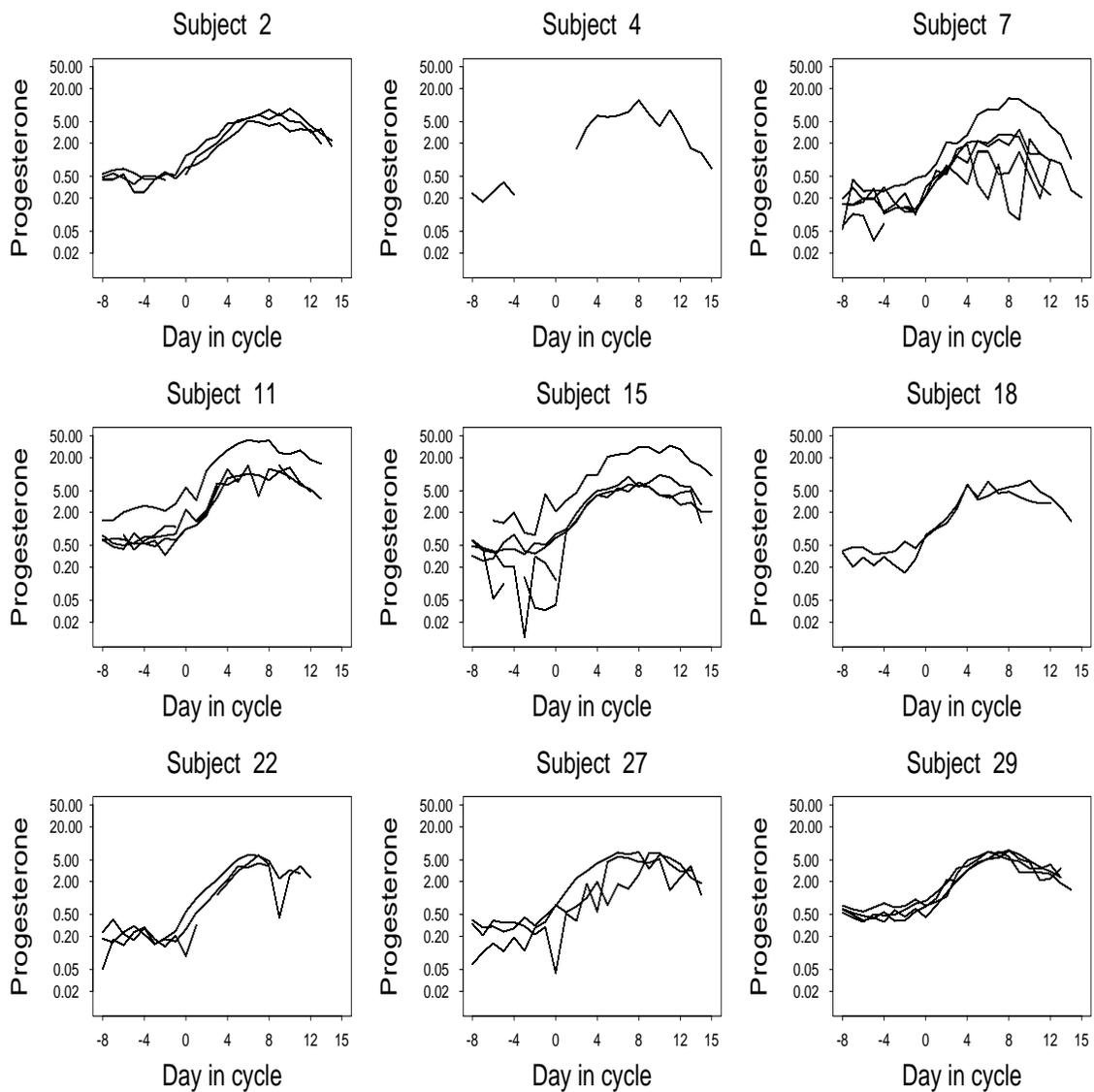


Figure 1: Observed records of progesterone metabolite concentrations (measured as ng PDG/mg Cr) from nonconceptive menstrual cycles, shown stratified by subject and graphed versus day in cycle relative to ovulation. Log concentrations were used in all analyses, and observations outside [day -8, day +15] excluded.

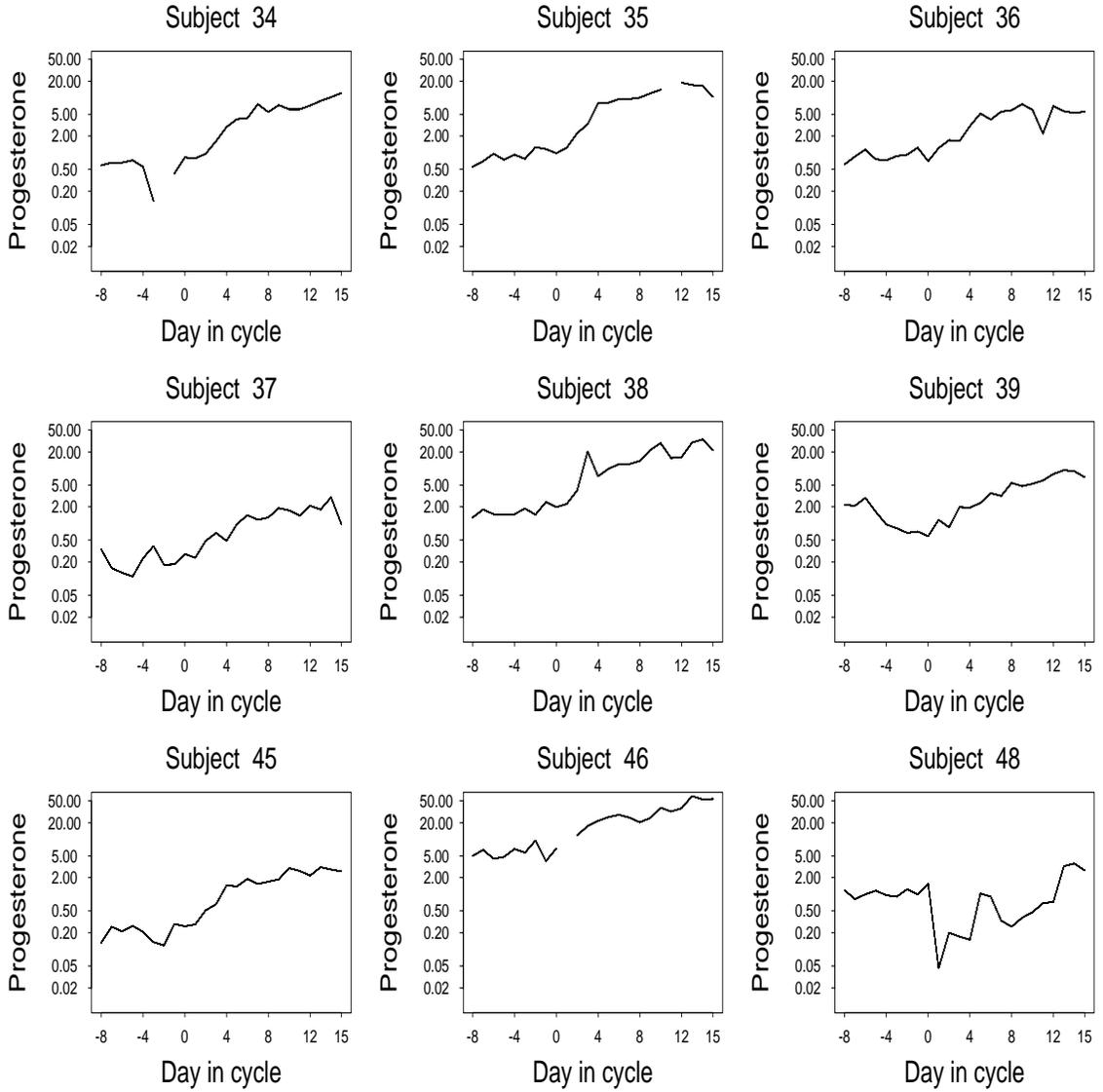


Figure 2: Progesterone data from nine conceptive cycles.

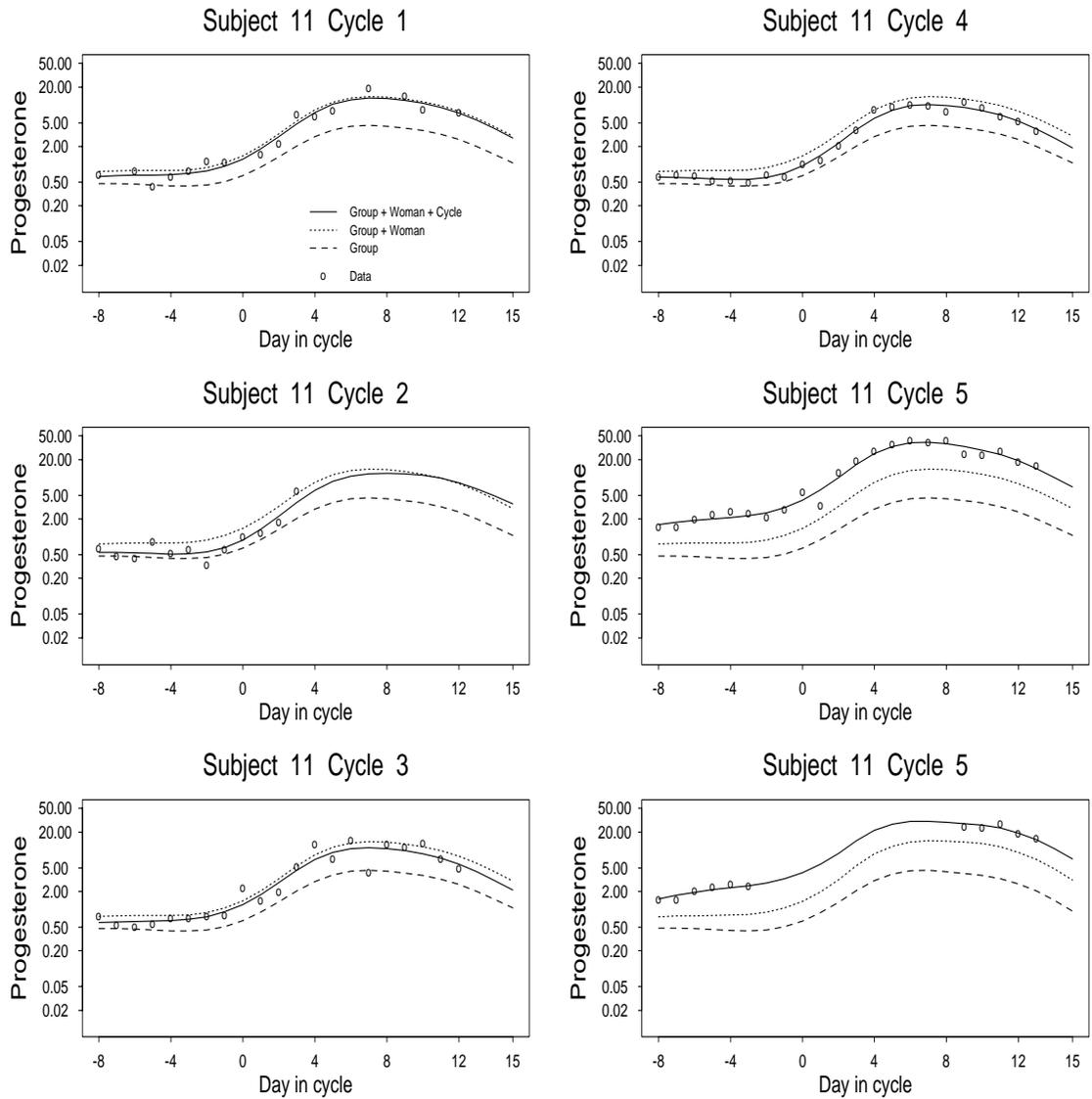


Figure 3: Results of applying our methods to the progesterone data, illustrating the smooth analysis of variance and proficient data imputation. Data was deleted from Cycle 5 in a subsequent analysis for comparison.

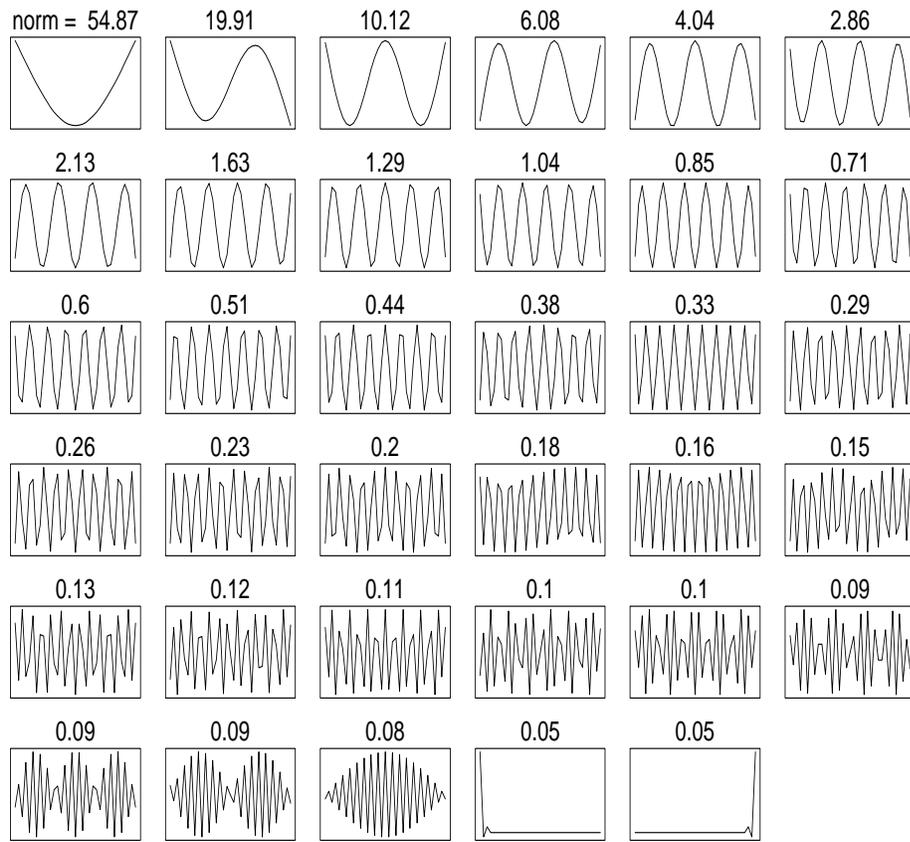


Figure 4: Columns of  $Z$  and their norms.

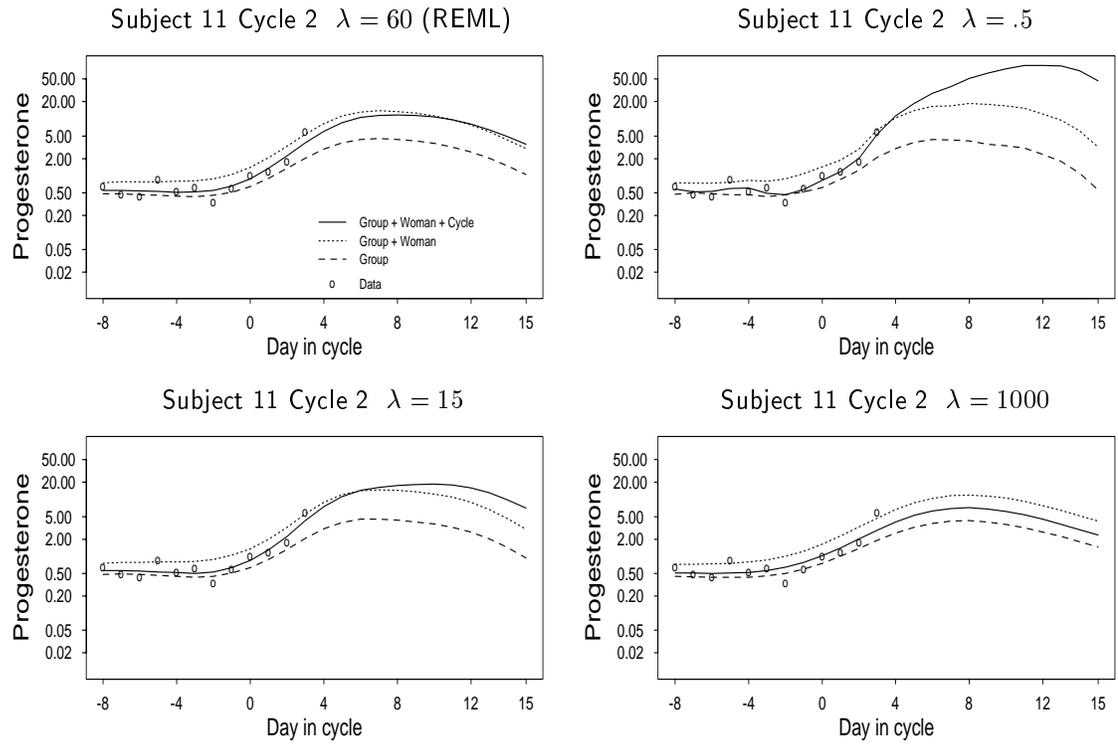


Figure 5: Comparison of fitted curves for Subject 11 Cycle 2 for varying values of the smoothing parameter  $\lambda$ .

	Smaller subset		Larger subset	
	Estimate	SD	Estimate	SD
$\sigma^2$	.080	.0065	.1655	.0088
$(\sigma^2/\lambda_g)$	.0107	.0048	.010	.0040
$(\sigma^2/\lambda_w)$	.0059	.0027	.0063	.0021
$(\sigma^2/\lambda_c)$	.0041	.0014	.0052	.0014

Table 1: Estimates of the variance components and their asymptotic standard deviations using two subsets of the data with the nested model and unconstrained smoothing parameters.

### Conceptive and nonconceptive group means

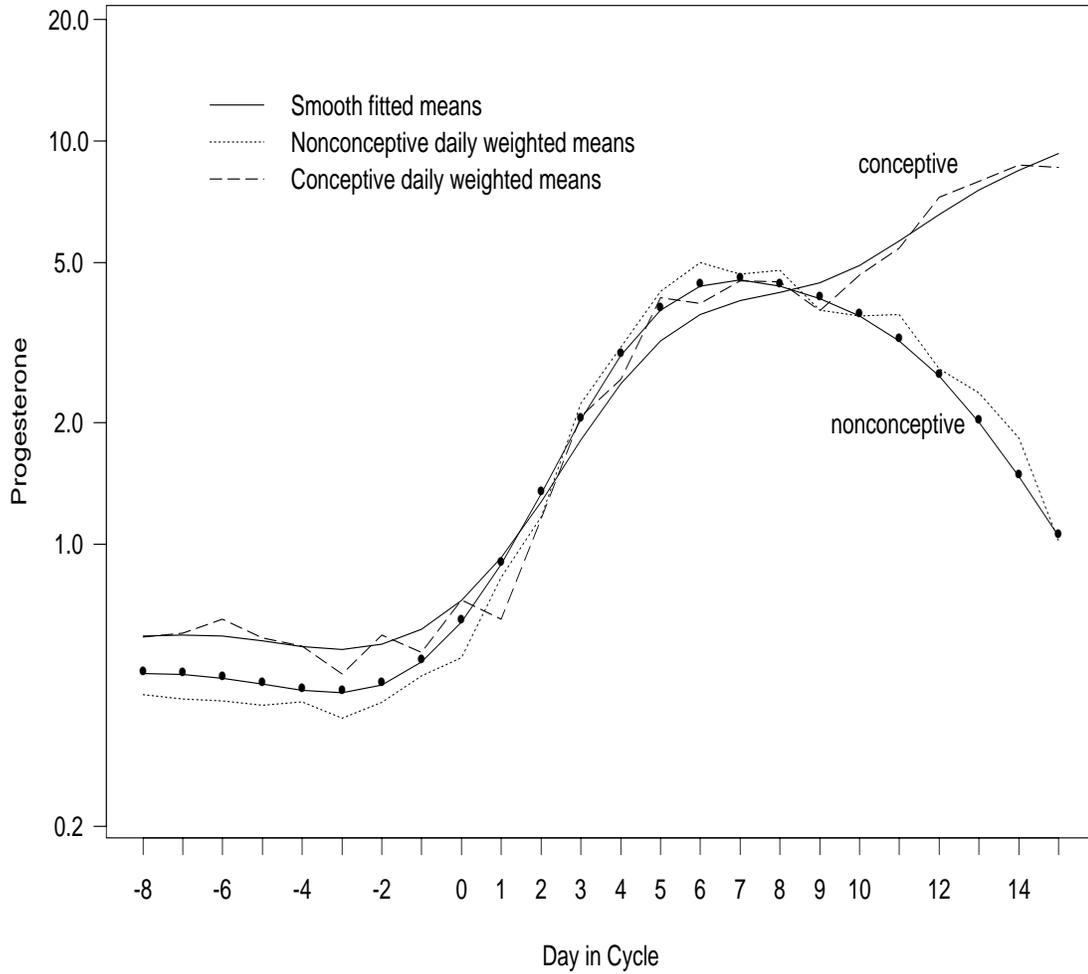


Figure 6: Smooth estimates for the nonconceptive and conceptive group means, overlaid by a rough comparison based on daily means inversely weighted by number of observations per subject.

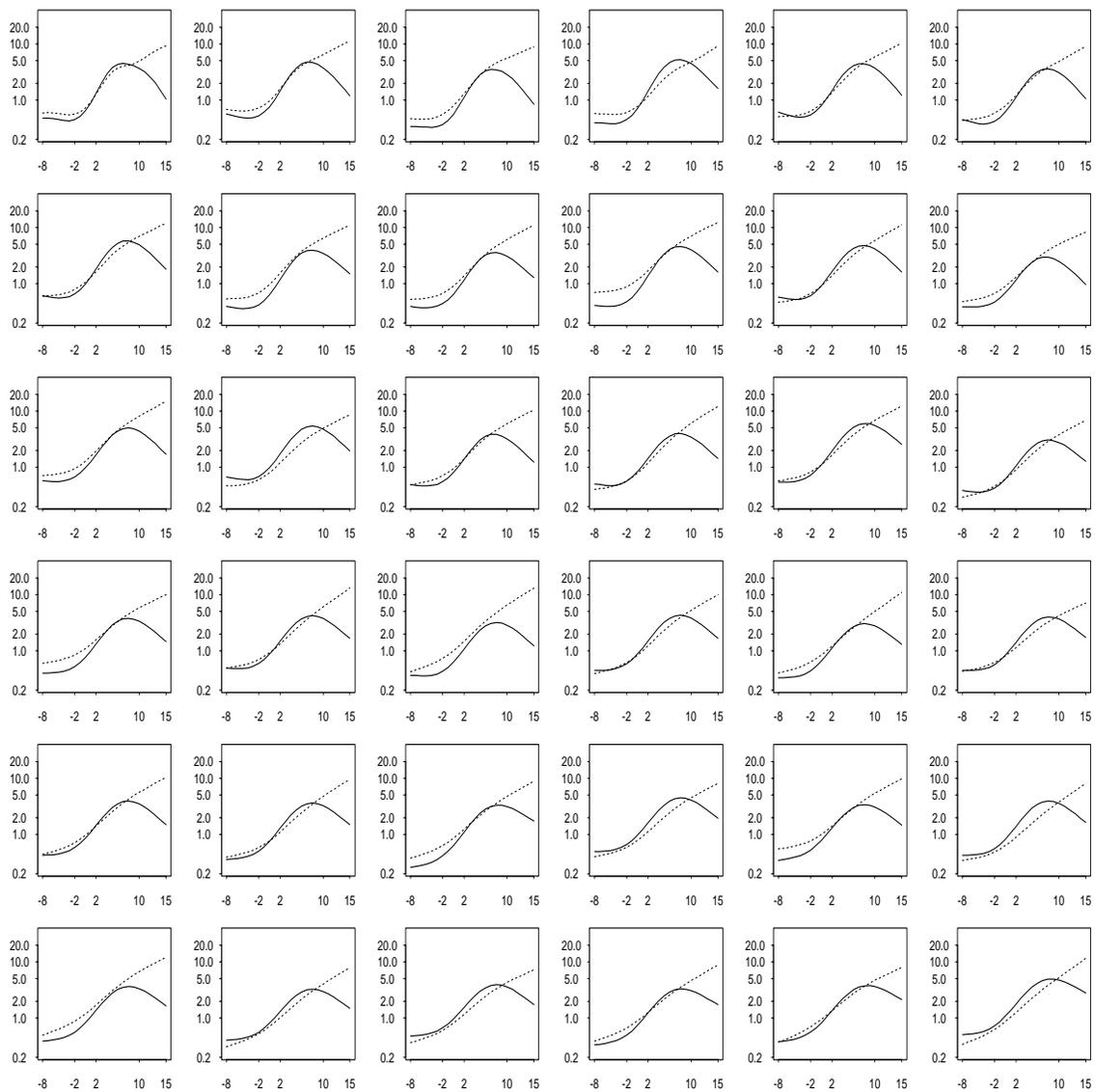


Figure 7: Example of 35 bootstrap simulations to compare fitted group means. Solid lines represent nonconceptive and dotted lines conceptive progesterone means, plotted versus day in cycle. The original fit is displayed in the first panel for comparison.