Statistica Sinica **9**(1999), 679-698

CONFIDENCE BANDS FOR GROWTH AND RESPONSE CURVES

Jiayang Sun, Jonathan Raz* and Julian J. Faraway*

*Case Western Reserve University and * The University of Michigan*

Abstract: Linear models with structured covariance matrices are applicable to a wide variety of longitudinal data sets. This paper presents a method for constructing simultaneous confidence bands for the growth or response curve. The bands can be defined on finite intervals and various other subsets of the real line. Previous results on simultaneous confidence bands in linear regression with independent errors are generalized to include correlated errors when the covariance matrix depends on a finite number of unknown parameters. A correction for the additional variability arising from estimation of unknown covariance parameters is presented. The method is evaluated using data simulated from several different linear models, and it is applied to the analysis of pig metabolite concentrations after brief myocardial ischemia.

Key words and phrases: Longitudinal data, mixed effects, nuclear magnetic resonance, repeated measurements, structured covariance, tube method.

1. Introduction

Longitudinal data arise in a wide variety of applications, including pharmacokinetics (Vonesh and Carter (1987)), health research (Jennrich and Schluchter (1986), Gelfand, Hills, Racine-Poon and Smith (1990), Carter, Resnick, Ariet, Shieh and Vonesh (1992)) and education (Raudenbush (1988)). Longitudinal studies allow investigation of the properties of the growth or response curve, such as the shape of the curve, the intervals in which it exceeds a threshold, or the presence and amplitude of local extrema.

For example, Schwartz, Schaefer, Gober, Meyerhoff, Smekal, Massie and Weiner (1990) studied myocardial reactive hyperemia by measuring phosphocreatine (PCr) concentration in the pig heart before, during, and after occlusion of the coronary artery. The investigators were interested in the shape of the mean response curve, especially during and immediately following occlusion. Figure 1 shows the PCR measurements for each of 12 pigs, and Figure 2 shows the means at each time. Figure 2 also shows a regression spline estimate of the response curve based on a fit of a linear model with autoregressive errors to the data in Figure 1. The PCr measurements vary considerably between and within pigs,

suggesting a large amount of uncertainty in the estimate of the true response curve.

Although existing methods provide estimates of the response curve, and inference concerning the parameters of the function used to describe the curve, they provide no way to quantify the uncertainty in the estimate of the curve itself. This article presents a new method for constructing simultaneous confidence bands (SCBs) for a portion of a growth or response curve when the estimated curve is constructed from maximum likelihood (ML) or restricted maximum likelihood (REML) estimates of the parameters of a linear model with a structured covariance matrix (Jennrich and Schluchter (1986)). Figure 3 shows an SCB for the PCr response curve that is restricted to the interval that was of primary interest to the investigators.

Related work on SCBs for a linear regression function goes back to Scheffé (1959, p.68ff) where errors are independent and homoscedastic, and the predictor space is unconstrained, i.e. the domain of interest (denoted by $\mathcal{T}$ throughout this paper) is $\mathbb{R}^{q+1}$, where $q+1$ is the dimension of the predictor space. Therefore any bands using constraints on the domain should be narrower than Scheffé's bands. Previous publications that discussed bands defined on a constrained domain include Bohrer and Francis (1972), Casella and Strawderman (1980), Halperin, Rastogi, Ho and Yang (1967), Wynn (1984), Uusipaikka (1983), Naiman (1987), Hall and Titterington (1988), Härdle and Marron (1991) and Sun and Loader (1994). Other motivations for informative or narrow SCBs can be found in Cox (1965). Also see Seber (1977) for a survey of earlier methods and Gafarian (1978) for fixed-width and hyperbolic bands. Faraway and Sun (1994) considered the case in which errors are independent and heteroscedastic, and they developed an adjustment to account for the extra variation introduced by estimation of unknown weights in a weighted least squares regression.

In longitudinal data, the errors are correlated, and ML and REML are often used, so a different formula for the SCB is needed. Elston and Grizzle (1962) modified Scheffé's bands to account for the correlation structure, and hence their bands will be too wide when we are only interested in a nontrivial subset of $\mathbb{R}^{q+1}$. Steward (1987) considered line-segment bands so that he could reduce the calculation of the coverage probability to a calculation at the two end points. Here, we apply the tube method (Naiman (1987) and Sun and Loader (1994)) to constructing SCBs when the covariances are known, and then modify the bands for the case when the covariance parameters are unknown (Section 3.1). We develop a derivative method and a perturbation method for adjusting the bands to account for the variation in ML and REML estimates of the covariance parameters (Section 3.2). We evaluate the SCBs using simulated data from a variety of linear models with structured covariance matrices (Section 4). We

apply the SCB to magnetic resonance spectroscopy measurements of metabolites in pig hearts (Section 5), and we compare our SCB to Scheffé-type bands and to Bonferroni t intervals. The linear model is described in Section 2.

Our bands have width proportional to the estimated standard deviation of the estimator of the response curve. Typically, this will mean that the area bounded by our bands is smaller than the area bounded by fixed width confidence bands. Our formula is applicable in principle to any linear smoothers (Sun and Loader (1994)) and to a large class of domains $\mathcal{T}$, for example, the domain can be a union of finite intervals and/or points. Advantages of tube methods over alternative methods in general can be found in Loader (1993).

2. The Linear Model for Growth and Response Curves

Based on earlier work of Harville (1977), Laird and Ware (1982) proposed a linear mixed effects model for longitudinal data, and they discussed ML and REML estimation using the EM algorithm. Jennrich and Schluchter (1986) discussed linear mixed effects models and other linear models with structured covariance matrices, and they considered estimation using the Newton-Raphson algorithm. We consider a special case of these linear models in which the regression function depends only on time.

Suppose that longitudinal data are acquired from n individuals, and that p_i measurements are made on individual i. Let the response of individual i at time t_{ij} be denoted by Y_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, p_i$, and let $x_0(t), \ldots, x_q(t)$ be specified smooth functions (such as powers of t), where x_0 typically is set equal to one. Then a reasonable linear model is

$$Y_{ij} = f(t_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n; \quad j = 1, \ldots, p_i, \tag{1}$$

where $f(t) = \sum_{k=0}^{q} \beta_k \, x_k(t)$. Our goal is to find SCBs for $f(t)$, for all $t \in \mathcal{T}$, a one-dimensional domain, for example, $\mathcal{T} = [a, b]$ or $\mathcal{T} = [a, b] \cup [c, d]$. We consider only the one-dimensional case, but an extension to higher dimensional $\mathcal{T}$ is possible.

Let $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip_i})^T$. Assume that $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are independent and that $\boldsymbol{\epsilon}_i$ is multivariate normal with mean equal to the zero vector and covariance matrix $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{R}_i(\boldsymbol{\theta})$, as in Jennrich and Schluchter (1986), where σ^2 is an unknown parameter, and the $p_i \times p_i$ matrix $\mathbf{R}_i(\boldsymbol{\theta})$ may depend on an unknown $r \times 1$ parameter vector $\boldsymbol{\theta}$. In mixed effects models, $\mathbf{R}_i(\boldsymbol{\theta})$ has the form $\mathbf{Z}_i \mathbf{D}_i(\boldsymbol{\theta}_1) \mathbf{Z}_i^T + \mathbf{W}_i(\boldsymbol{\theta}_2)$, where $\mathbf{Z}_i$ is a specified matrix, $\mathbf{D}_i$ is a specified matrix function of an unknown vector $\boldsymbol{\theta}_1$, $\mathbf{W}_i$ is a specified matrix function of an unknown vector $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$.

If we define $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip_i})^T$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_q)^T$, and the $p_i \times (q + 1)$ matrix $\mathbf{X}_i$ with elements $x_k(t_{ij})$, then the model has the form of a linear model

with a structured covariance matrix: $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, $i = 1, \ldots, n$. Further, let $N = \sum_{i=1}^{n} p_i$, $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^T$ be an N-vector, $\mathbf{X}^T = (\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T)$ be an $N \times (q+1)$ matrix, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \ldots, \boldsymbol{\epsilon}_n^T)^T$. Then the model can be written in the usual matrix form: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ has an N-dimensional normal distribution with zero mean and block diagonal covariance function

$$\boldsymbol{\Sigma} \;=\; \sigma^2\mathbf{R}(\boldsymbol{\theta}) \;\equiv\; \sigma^2\mathrm{diag}\{\mathbf{R}_1(\boldsymbol{\theta}), \ldots, \mathbf{R}_n(\boldsymbol{\theta})\}.$$

Note that we consider only linear models of growth and response curves, but many models of longitudinal data are nonlinear in the parameters (Lindstrom and Bates (1990) and Vonesh and Carter (1992)). In the latter case, our confidence bands might be generalized by constructing a linear approximation to the fitted nonlinear model, with an additional correction for this approximation.

Our method requires that the estimators of $\boldsymbol{\beta}$ and σ^2 be independent. Under mild regularity conditions, this assumption is asymptotically satisfied by the ML (Jennrich and Schluchter (1986), Miller (1977) and Harville (1977)) and REML estimators (Jiang (1998)). These theoretical results are illustrated by our simulation study.

3. Simultaneous Confidence Bands

3.1. Approximation formula

If $\boldsymbol{\theta}$ were known, we could estimate $\boldsymbol{\beta}$ by a generalized least squares estimator:

$$\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{Y},$$

and hence $f(t)$ by

$$\hat{f}(t) = \sum_{k=0}^{q} \hat{\beta}_k\, x_k(t) = \mathbf{x}(t)^T\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \equiv \mathbf{l}(t,\boldsymbol{\theta})^T\mathbf{Y} = \sum_{i=1}^{n} \mathbf{l}_i(t,\boldsymbol{\theta})^T\mathbf{Y}_i,$$

where $\mathbf{x}(t)$ is a $q+1$-vector with elements $x_k(t)$, $\mathbf{l}(t,\boldsymbol{\theta})$ is an N-vector, and $\mathbf{l}_i(t,\boldsymbol{\theta})$ is a p_i-vector with $\mathbf{l}(t,\boldsymbol{\theta})^T = (\mathbf{l}_1(t,\boldsymbol{\theta})^T, \ldots, \mathbf{l}_n(t,\boldsymbol{\theta})^T) = \mathbf{x}(t)^T\{\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{X}\}^{-1}$ $\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}$. Thus, $\mathrm{Var}\,\{\hat{f}(t)\} = \mathbf{l}(t,\boldsymbol{\theta})^T\boldsymbol{\Sigma}\,\mathbf{l}(t,\boldsymbol{\theta}) = \sigma^2||\mathbf{l}^R(t,\boldsymbol{\theta})||^2$, where $\mathbf{l}^R(t,\boldsymbol{\theta}) = \mathbf{V}^{\mathbf{T}}(\boldsymbol{\theta})\mathbf{l}(\mathbf{t},\boldsymbol{\theta})$, $\mathbf{V}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})^T = \mathbf{R}(\boldsymbol{\theta})$ is the Cholesky decomposition of $\mathbf{R}$, and

$$||\mathbf{l}^R(t,\boldsymbol{\theta})||^2 = \mathbf{l}(t,\boldsymbol{\theta})^T\mathbf{R}(\boldsymbol{\theta})\mathbf{l}(t,\boldsymbol{\theta}) = \mathbf{x}(t)^T\{\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{X}\}^{-1}\mathbf{x}(t).$$

A reasonable simultaneous confidence interval is of the form

$$\left[\;\hat{f}(t) - c\sigma\|\mathbf{l}^R(t,\boldsymbol{\theta})\|, \qquad \hat{f}(t) + c\sigma\|\mathbf{l}^R(t,\boldsymbol{\theta})\|\;\right], \tag{2}$$

where c is such that the coverage probability of $f(t)$ in the interval for all $t \in \mathcal{T}$ equals a prescribed level, $1 - \alpha$.

When σ^2 and $\boldsymbol{\theta}$ are *unknown*, it is natural to replace them by some good estimators $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}$, for example ML or REML estimators. Then we need to study SCBs of the form

$$\left[\ \hat{f}(t) - c\hat{\sigma}\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|, \qquad \hat{f}(t) + c\hat{\sigma}\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\| \ \right] \quad \text{for some } c. \tag{3}$$

We adjust the usual ML estimator, $\tilde{\sigma}^2$, by setting $\hat{\sigma}^2 = N\tilde{\sigma}^2/\nu$ where $\nu = N-q-1$ because this reduces the bias and produces superior results in our simulations. REML estimators also were introduced to reduce the bias of ML estimators. The $\hat{\sigma}^2$ below refers to either the adjusted ML estimator or the REML estimator.

The problem of finding a $1-\alpha$ SCB for $f(t), \forall t \in \mathcal{T} \subseteq \mathbb{R}$, in the form (3) is equivalent to choosing c such that

$$\begin{aligned}\alpha &= \mathrm{pr}\Big\{|\hat{f}(t) - f(t)| > c\hat{\sigma}\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|, \ \text{ for some } t \in \mathcal{T}\Big\} \\ &= \mathrm{pr}\Big\{\sup_{\mathbf{t}\in\mathcal{T}} \frac{|\hat{f}(t) - f(t)|}{\sigma\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}\frac{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|} > c\frac{\hat{\sigma}}{\sigma}\Big\}. \end{aligned} \tag{4}$$

Note that $Z_n(t) = |\hat{f}(t) - f(t)|/\{\sigma\|\mathbf{l}^R(t,\boldsymbol{\theta})\|\}$ is a different process from a corresponding process in Faraway and Sun (1994). When the data are correlated, the SCB must be corrected for the extra variation introduced by estimating $\boldsymbol{\theta}$.

Proposition 1. *Suppose that σ^2 and $\boldsymbol{\theta}$ are estimated by some consistent estimators, e.g. adjusted MLE or REML estimators. Define*

$$\mathbf{S}(t) = \frac{\mathbf{l}^R(t,\boldsymbol{\theta})}{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|} = \frac{\mathbf{V}(\boldsymbol{\theta})^T\mathbf{l}(t,\boldsymbol{\theta})}{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}, \qquad a = \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|}{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}, \tag{5}$$

where $\mathbf{R}(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})^T$ is the Cholesky decomposition of $\mathbf{R}$. Suppose that a' is a positive constant such that $P(a < a') = o(\alpha)$ as $n \to \infty$ and $\alpha \to 0$. Then an approximation to (4) *is*

$$\alpha \approx E \cdot P(|t_\nu| > ca') + \frac{\kappa_0}{\pi}\left(1 + \frac{c^2a'^2}{\nu}\right)^{-\nu/2}, \tag{6}$$

where t_ν denotes a t-distributed random variable with ν degrees of freedom, $\kappa_0 = \int_{t\in\mathcal{T}} \|\mathbf{S}'(t)\|dt$, the volume of $\{\mathbf{S}(t),\, t \in \mathcal{T}\}$, and E is the Euler-Poincaré characteristic of $\{\mathbf{S}(t),\, t \in \mathcal{T}\}$.

A derivation of (6) is in the Appendix.

The approximation (6) may be slightly conservative in the sense that the resulting coverage probability is slightly higher than $1-\alpha$, but the coverage probability gets very close to $1-\alpha$ as $N \to \infty$ and $\alpha \to 0$. For more discussion on this issue, see the end of the Appendix.

The resulting SCBs depend on constants E, κ_0, and a', which must be computed or estimated. The Euler-Poincaré characteristic E is equal to zero if $\{S(t), t \in \mathcal{T}\}$ has no boundary, one if $\{S(t), t \in \mathcal{T}\}$ is a continuous curve with two disjoint ends (as in most cases when $\mathcal{T}$ is an interval), and each extra disjoint segment or point adds one to E if the disjoint points/ends are not too close to each other. (Consult Kreyszig (1968) for details.) Two methods for estimating a' are presented in the next section. A simple approximation to κ_0 when $\mathcal{T} = [a, b]$ is obtained by partitioning $[a, b]$ into $a = z_0 < \cdots < z_m = b$, and computing

$$\kappa_0 = \sum_{i=1}^{m} \int_{z_{i-1}}^{z_i} \|\mathbf{S}'(x)\| dx \approx \sum_{i=1}^{m} \|\mathbf{S}(z_i) - \mathbf{S}(z_{i-1})\|.$$

When $\boldsymbol{\theta}$ is unknown, κ_0 is estimated by substituting $\hat{\boldsymbol{\theta}}$ in the definition of $\mathbf{S}$.

Of course, the resulting coverage probability, based on c from (6) and the estimated constants, is only approximately $1-\alpha$. However, it should get closer to the nominal level as the sample size increases. Indeed, as shown in our simulation, using estimated constants in our formula is quite satisfactory even for moderate sample sizes.

3.2. Correction constant a'

In Faraway and Sun (1994), the correction constants were estimated by replacing the true unknown weights and regression function with non-parametric estimators derived from a kernel regression with a small window width. In the models with structured covariance matrices considered here, it is not clear how to develop an analogous procedure for estimating the correction constant a'. As alternatives, we propose two methods: the derivative method and the perturbation method.

The *derivative method* is similar to the delta method. Recall that if $\hat{\boldsymbol{\theta}}$ is the ML estimator,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(N^{-1/2}), \qquad \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \ \sim \ \mathcal{N}(0, \ \boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

where $\sim$ denotes "has approximately same distribution as, for large N", and $\boldsymbol{I}^{-1}(\boldsymbol{\theta}) = O_p(N^{-1})$ is the inverse of the Fisher information of $\boldsymbol{\theta}$ from $\mathbf{Y}$. Let $\boldsymbol{I}(\boldsymbol{\theta})^{-1} = \mathbf{U}(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})^T$ be the Cholesky decomposition of the inverse of the Fisher information, so $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim \mathbf{U}(\hat{\boldsymbol{\theta}})\mathbf{e}$ where $\mathbf{e}$ is an r dimensional standard normal random vector and $\mathbf{U} = O_p(N^{-1/2})$. Hence, with ∇ denoting the gradient with respect to $\boldsymbol{\theta}$, we have

$$a^2 = \inf_{t \in \mathcal{T}} \frac{\|\mathbf{l}^R(t, \hat{\boldsymbol{\theta}})\|^2}{\|\mathbf{l}^R(t, \boldsymbol{\theta})\|^2}$$

$$\geq \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|^2}{||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2 + |(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\nabla||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2|}$$
$$\sim \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|^2}{||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2 + |\mathbf{e}^T\mathbf{U}(\hat{\boldsymbol{\theta}})^T\nabla||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2|},$$

where the inequality is an approximate inequality, for large N. Note that for a nonnegative constant d and vector $\mathbf{A}$, $E(d/(d+|\mathbf{e}^T\mathbf{A}|) \geq d/(d+E|\mathbf{e}^T\mathbf{A}|)$ by Jensen's inequality. So, if N is large enough and dimension of $\boldsymbol{\theta}$ is one, i.e. $r=1$, a reasonable estimate of a' is

$$\hat{a}' = \left\{ \inf_{t\in\mathcal{T}} \frac{||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2}{||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2 + (2/\pi^{1/2})|\mathbf{U}(\hat{\boldsymbol{\theta}})^T\nabla||\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})||^2|} \right\}^{1/2}. \tag{7}$$

For example, in the compound symmetry model discussed in Section 4, (7) reduces to

$$\hat{a}' = \left\{1 + 2/(\pi n)^{1/2}\right\}^{-1/2}. \tag{8}$$

For some covariance structures, it is difficult to compute the derivatives of $||\mathbf{l}^R(t,\boldsymbol{\theta})||^2$. For this reason, we propose the *perturbation* method, which does not use derivatives. As in the derivative method, we have, for some standard normal vector $\mathbf{e}$,

$$a^2 = \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\ \hat{\boldsymbol{\theta}})\|^2}{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}}+\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\|^2}$$
$$\sim \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|^2}{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}}+\mathbf{U}(\hat{\boldsymbol{\theta}})\mathbf{e})\|^2}$$
$$= \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|^2}{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}}+||\mathbf{e}||\mathbf{U}(\hat{\boldsymbol{\theta}})\mathbf{e}/||\mathbf{e}||)\|^2},$$

where $||\mathbf{e}||^2$ has a χ^2 distribution with degrees of freedom q and $||\mathbf{e}||$ has mean

$$\mu_q \equiv \Gamma(\frac{q}{2}+\frac{1}{2})2^{1/2}\{\Gamma(\frac{q}{2})\}^{-1}.$$

In a small neighborhood of $\hat{\theta}$, $\|\mathbf{l}^R(t,\theta)\|$ is approximately linear in $\boldsymbol{\theta}$, and hence its extreme values are approximately at vertices of the neighborhood. Hence, as

$$\max_{||\mathbf{d}||=\mathbf{1}} \mathbf{b}'\mathbf{d} = ||\mathbf{b}||, \qquad \text{for any vector } \mathbf{b},$$

a reasonable estimate for a' is

$$\hat{a}' = \left\{ \text{ave}_{\mathbf{u}\in\mathcal{U}} \inf_{t\in\mathcal{T}} \frac{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|^2}{\|\mathbf{l}^R(t,\ \hat{\boldsymbol{\theta}}+\mu_q\mathbf{u})\|^2} \right\}^{1/2}. \tag{9}$$

Here $||\mathbf{b}||$ is the l_2 norm of $\mathbf{b}$, and

$$\mathcal{U} = \{\mathbf{u} : \ \mathbf{u} = (s_1||\mathbf{v}_1||, \ldots, s_q||\mathbf{v}_q||)^T, \ s_i = 1 \text{ or } -1, \ i = 1, \ldots, q\}$$

with $\mathbf{v}_i$ being the ith row vector of $\mathbf{U}$, and the average is over $\mathbf{u} \in \mathcal{U}$ such that $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is positive definite.

4. Simulation

4.1. Compound symmetry

In this and the following subsection we take $p_i \equiv p$ and space the t_{ij} evenly in the interval $\mathcal{T}$. Let $f(t_{ij}) = \beta_0 + \beta_1 t_{ij}$ and set $\mathbf{Z}_i = \mathbf{1}_p$ in a mixed effects model so that $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_p + \sigma_a^2 \mathbf{1}_p \mathbf{1}_p^T$ where σ_a^2 is the variance of the random intercept, $\mathbf{1}_p$ is a vector of 1's, $\boldsymbol{I}_p$ is a $p \times p$ identity matrix and set $\theta = \sigma_a^2/\sigma^2$. In this special case, we may estimate the a' directly by (8).

One hundred thousand replications were used and standard errors for the empirical coverage rates are no more than 1 unit in the 3rd decimal place. The standard case is $n = 15$, $p = 5$, $\sigma = 1$, $\theta = 1$ and $\mathcal{T} = [-0.5, 0.5]$. We vary only one setting at a time from the standard setting in Table 1, which gives the empirical coverage rates. We see that the actual levels at 95% are somewhat conservative but then this is expected due to the nature of the approximating formula. The 99% SCBs are all close to their nominal level.

Table 1. Empirical coverage rates for compound symmetry structure.

Nominal	Standard	Wider Range $\mathcal{T} = [-1.5, 1.5]$	Larger θ $\theta = 2$	Smaller θ $\theta = 0.5$	Larger n $n = 50$	Larger p $p = 10$
0.95	0.961	0.967	0.960	0.962	0.962	0.962
0.99	0.991	0.992	0.990	0.991	0.992	0.991

4.2. Random slope and intercept

In this model, $\mathbf{R}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \mathbf{D}_i(\boldsymbol{\theta}_1) \mathbf{Z}_i^T + \sigma^2 \boldsymbol{I}$ where $\mathbf{Z}_i = (\mathbf{1}, t_i)$ and

$$\mathbf{D}_i = \begin{pmatrix} \theta_1 & \theta_3 \\ \theta_3 & \theta_2 \end{pmatrix}. \tag{10}$$

In this case, we could not compute the estimated a' by (8), so the perturbation method (9) was used. The standard model to which other cases will be compared has $(\theta_1, \theta_2, \theta_3) = (1, 1, 0.5)$ and $\sigma = 1$, $n = 15$, $p = 5$, and $\mathcal{T} = [-0.5, 0.5]$. As before, 100,000 replications were used and standard errors are no more than 1 unit in the 3rd decimal place. We vary only one setting at a time from the standard setting in Table 2.

We performed additional simulations using REML estimation as implemented in SAS PROC MIXED (SAS Institute Inc. (1992)). Only 1,000 replications were used for the REML simulations due to the much less efficient calculations in the SAS macro as compared with the C program used to compute the ML estimates. Thus, the standard errors for the REML results range between 0.0014 and 0.008. In this model, the correlation between the REML estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ is identical to that between the ML estimators.

Table 2. Empirical coverage rates for random slope and intercept structure.

Method	Nominal Level	Standard	Wider Range $\mathcal{T} = [-1.5, 1.5]$	Larger n $n = 50$	Even larger n $n = 200$	Larger p $p = 10$
ML	0.95	0.970	0.984	0.963	0.960	0.954
	0.99	0.994	0.997	0.997	0.993	0.989
REML	0.95	0.982	0.990	0.969	0.935	0.963
	0.99	0.997	0.998	0.998	0.989	0.992

The ML results are somewhat conservative, although if there is to be an inaccuracy, better that it be this kind. Accuracy improves with increasing n and p. The REML estimates of the random effects variance were consistently larger than the ML estimates, and thus the REML confidence bands tended to be wider, leading to the somewhat more conservative results. (In general, ML estimates of variance components are biased downwards, and REML estimates tend to be less biased or unbiased.)

5. Data Example and Further Simulation Results

Schwartz et al. (1990) measured PCr concentration in the pig heart using phosphorus (^{31}P) nuclear magnetic resonance spectroscopy. PCr concentration was measured every 4.8 seconds for 24 seconds of no intervention (5 measurements), 24 seconds during which the anterior descending coronary artery was occluded (5 measurements), and 62 seconds of reactive hyperemia after release of the occlusion (13 measurements). The design was unbalanced, since the investigators decided to acquire a few extra measurements per pig after all the data had been collected from the first few pigs.

The pigs vary considerably in their baseline level of PCr, which has no scientific importance. For this reason, the researchers normalized the data by averaging the five control responses and dividing each response by the control average for that pig. We found that this transformation also reduced the skewness and stabilized the variance, so we analyzed the data in the same form as the researchers did. Figures 1 and 2 show the data after normalization.

One approach to analyzing these data would be to construct a parametric model that explicitly includes parameters representing various features of the response curve and to construct simultaneous confidence intervals for the model parameters. However, this alternative approach would not provide inference about the uncertainty in the shape of the curve except through the uncertainty in the estimated parameters and would not be much more efficient than a confidence band if the number of parameters were large. In contrast, SCBs applied to the normalized data provide inference directly related to the shape of the response curve.

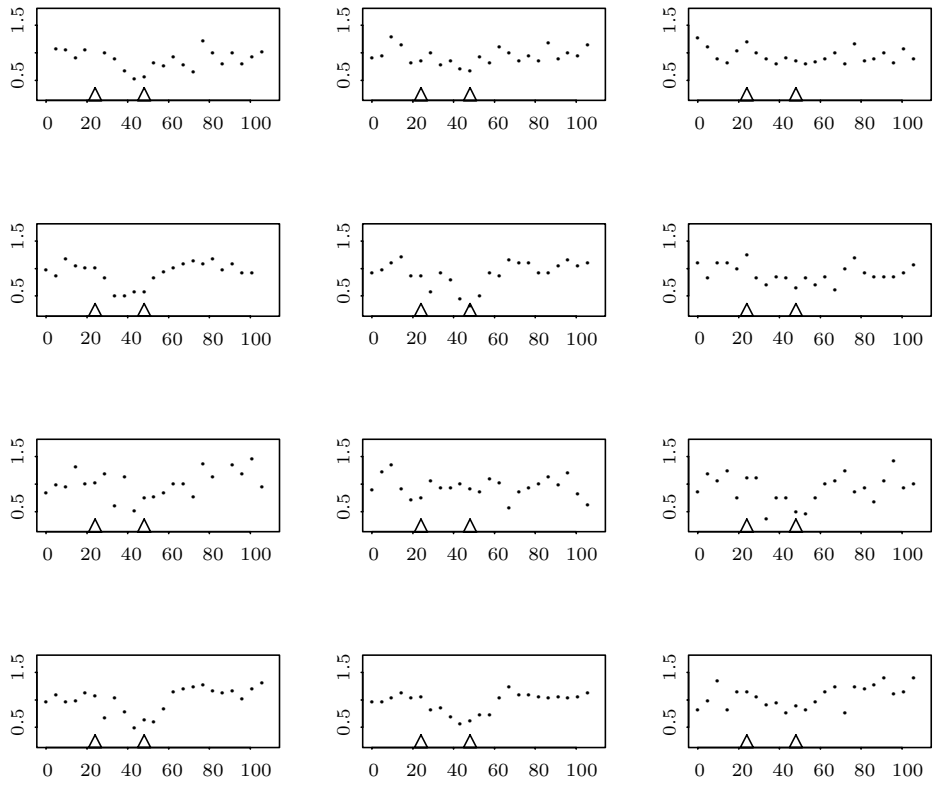

Figure 1. Plots of relative PCr concentration against time in seconds for each of 12 pigs. The triangles indicate the start and end of coronary artery occlusion.

We modeled the response curve using fixed knot quadratic B-splines with a knot at the start of occlusion (24.0 seconds from the start of data acquisition), two during occlusion (33.6 and 43.2 seconds), one at the start of reactive hyperemia

(52.8 seconds), and two during hyperemia (62.4 and 81.6 seconds). This model assumes a quadratic function before occlusion. The knot at 24 seconds reflects possible sudden changes at the time of occlusion. The two knots during occlusion are equally spaced between the end of occlusion and the start of hyperemia. The two knots during hyperemia were chosen under the assumption that the response function is less smooth near the beginning of this period and more smooth later in the period.

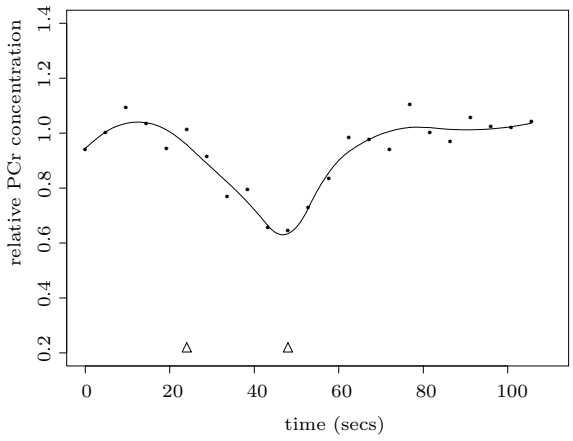

Figure 2. Mean relative PCr concentration against time with fitted regression spline. The plot gives the mean of non-missing data points at each of the 23 times. The regression spline estimate was computed by fitting a linear model with autoregressive errors to the data from 12 pigs. The triangles indicate the start and end of coronary artery occlusion.

The investigators did not expect systematic changes before occlusion; we chose to fit a general model that allowed for possible unanticipated changes in the mean. We could have modified the B-spline model to enforce a constant mean before occlusion. The procedure for constructing confidence bands would be unchanged in this case.

Our quadratic spline model assumes that the response function is continuous and has a continuous first derivative. At the suggestion of a referee, we compared the fit of our model with that of a model that does not enforce the constraint of a continuous first derivative at time 24 seconds. The two models gave very similar estimated response curves.

Using SAS PROC MIXED (SAS Institute Inc. (1992)), we computed REML estimates for models with several different covariance structures, including a structure defined by assuming first order autoregressive (AR(1)) errors, several

structures defined by including various spline coefficients as random effects, and structures defined by combining random effects with AR(1) errors. Based on Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) (Schwartz, *et al.* (1990)), we selected the AR(1) model. This model implies stationary errors, possibly incorrect because of the normalization.

Since the investigators were primarily interested in the response curve during and immediately following occlusion, we defined the interval $\mathcal{T}$ to start at the beginning of occlusion (24 seconds) and end at the measurement at which the mean relative PCr concentration (Figure 2) appears to return to the baseline level (76.8 seconds, which is the sixth measurement after release of occlusion). Then we constructed an approximate 95 per cent SCB (Figure 3) with a correction factor a' that was computed using the perturbation method. This band indicates strong evidence for a decrease in PCr concentration due to the occlusion, but suggests considerable uncertainty about the shape of the response curve, time to minimum, time to return to baseline, rate of decline, and other features.

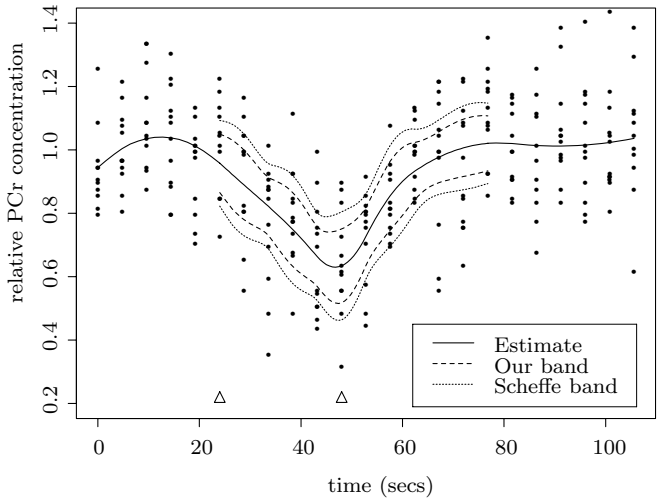

Figure 3. Scatterplot of PCr data from 12 pigs with fitted regression spline (solid line), our 95 percent simultaneous confidence band (dashed lines) for a finite interval, 95 percent Scheffé-type band (dotted lines). The analysis assumed autoregressive errors within pigs and independence among pigs. The decrease in PCr concentration during occlusion is significant based on the simultaneous confidence band for a finite interval, but the band indicates considerable uncertainty concerning the shape of the curve, time to minimum, time to return to baseline, rate of decline, and other features. The triangles indicate the start and end of coronary artery occlusion.

Figure 3 also shows the wider Scheffé-type band that is not restricted to the interval. This band was constructed using a generalization of a method proposed by Elston and Grizzle (1962). Details are given in the Appendix. For comparison with our bands defined on the interval $\mathcal{T}$, Figure 3 shows the Scheffé-type bands for the regression curve defined on this interval.

Simultaneous confidence intervals at the design points can be constructed using Bonferroni's inequality. Figure 4 shows the Bonferroni t intervals at the 12 time points between 24 and 76.8 seconds. These intervals were based on the mean of non-missing values at each of the 12 times, and they do not depend on the spline model. At most time points, the t intervals are noticeably wider than the SCB, even though our SCB applies to the regression function evaluated at every point on the real line between 24 and 76.8 seconds, while the t intervals only apply to 12 discrete points.

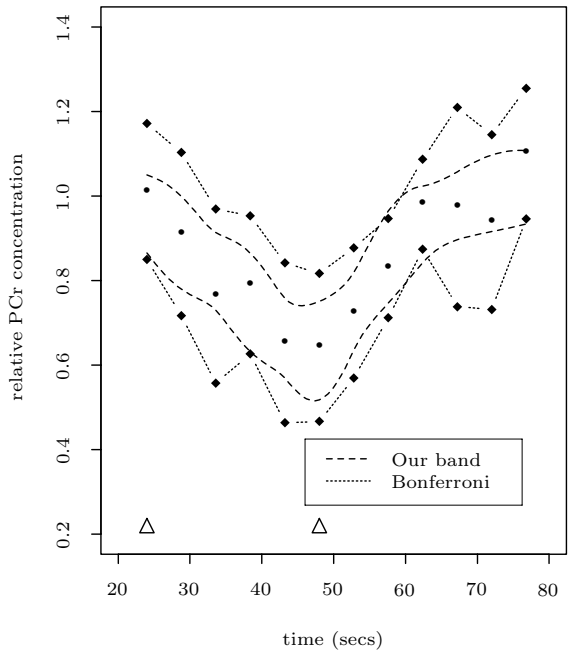

Figure 4. Mean relative PCr concentration (black dots) at 12 times with our 95 percent simultaneous confidence band for a finite interval (dashed lines), and 95 percent Bonferroni t confidence intervals. The endpoints of the Bonferroni intervals are indicated by black diamonds. The open triangles indicate the start and end of coronary artery occlusion.

We also constructed an SCB based on an analysis of the unnormalized data using a model that included both a random intercept and AR(1) errors. This second band was much wider than the one shown in Figure 3, reflecting the variability in the baseline PCr concentration and the consequent uncertainty in the estimate of the fixed intercept.

Since the data and model in this application were more complicated than those used in the simulation studies (Section 4), we performed additional simulations using artificial data designed to resemble the real PCr data. The simulated data were generated from the spline model with AR(1) errors, both with and without a random intercept. The true parameter values in the simulation were set equal to the estimates from the analysis of the real data after normalization (for the simulations without a random intercept in the model) and the estimates before normalization (for the simulations with a random intercept). Each simulated data value was defined to be missing with probability 10%. The value of p_i was 23 minus the number of missing responses for individual i, and n was taken to be 12 (as in the real data), 50, and 100. For each of 1,000 simulated data sets, both ML and REML estimates were computed and both 95% and 99% bands were constructed. Tables 3 and 4 give the empirical coverage rates, which are close to the nominal rates. The bands tend to be slightly too narrow when $n = 12$, and somewhat conservative when $n = 50$.

In the $n = 50$ and $n = 100$ simulations, we saved the ML and REML estimators in each replication, and computed the correlation between the estimators of β and σ^2. This correlation was small and decreased with increasing n.

Table 3. Empirical coverage rates for spline models with AR(1) errors.

Method	Nominal Level	$n = 12$	$n = 50$	$n = 100$
ML	0.95	0.942	0.974	0.947
	0.99	0.987	0.995	0.992
REML	0.95	0.950	0.974	0.947
	0.99	0.990	0.995	0.992

Table 4. Empirical coverage rates for spline models with random intercept and AR(1) errors.

Method	Nominal Level	$n = 12$	$n = 50$	$n = 100$
ML	0.95	0.944	0.974	0.972
	0.99	0.984	0.998	0.994
REML	0.95	0.946	0.976	0.973
	0.99	0.985	0.998	0.994

6. Discussion

We have derived SCBs for linear models with structured covariance matrices. Our bands cover an infinite number of points in any compact interval. In the application to pig metabolite data, the SCB captured the dip in the response curve following occlusion, despite the small number of pigs. The simulation results indicated that the bands perform reasonably well even in such difficult applications.

SCBs might be useful in applications to many other types of longitudinal data sets. For example, Guthrie and Buchwald (1991) suggested analyzing differences between two electrophysiological time series using multiple t-tests at each time point with a correction for multiple comparison under the assumption of AR(1) errors. An alternative and possibly more informative approach would be to construct an SCB based on the difference of the two series.

Acknowledgements

We thank Gregory Schwartz, M.D., for supplying the PCr data. The work of Jonathan Raz was supported by National Science Foundation award DMS-91-16730. The work of Jiayang Sun was supported in part by National Science Foundation awards DMS-92-03357 and DMS-9504515.

Appendix

Derivation of (6). Denote the right hand side of (4) by $P(c)$. Let $\mathbf{l} = \mathbf{l}(t, \boldsymbol{\theta})$, $\hat{\mathbf{l}} = \mathbf{l}(t, \hat{\boldsymbol{\theta}})$, $\mathbf{l}^R = \mathbf{l}^R(t, \boldsymbol{\theta})$, and $\delta(t) = |\left\langle \mathbf{l} - \hat{\mathbf{l}}, \mathbf{Y} \right\rangle|$. Further, let $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, so that $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$. Under the model (1), $f(t) = \langle \mathbf{l}(t, \boldsymbol{\theta}), \boldsymbol{\mu} \rangle$. So, the difference between $f(t)$ and its estimator can be bounded by two terms:

$$|\hat{f}(t) - f(t)| = |\langle \mathbf{l}, \boldsymbol{\epsilon} \rangle + \langle \mathbf{l}, \boldsymbol{\mu} \rangle + \left\langle \hat{\mathbf{l}} - \mathbf{l}, \mathbf{Y} \right\rangle - f| \ \leq \ |\langle \mathbf{l}(t, \boldsymbol{\theta}), \boldsymbol{\epsilon} \rangle| + \delta(t), \quad \text{(A.1)}$$

The normalized first term in (A.1) is

$$Z(t) = \frac{\langle \mathbf{l}(t, \boldsymbol{\theta}), \boldsymbol{\epsilon} \rangle}{\sigma \|\mathbf{l}^R(t, \boldsymbol{\theta})\|} = \left\langle \frac{\mathbf{l}}{\|\mathbf{l}^R\|}, \frac{\boldsymbol{\epsilon}}{\sigma} \right\rangle,$$

which has the same distribution as $\left\langle \mathbf{l}/\|\mathbf{l}^R\|, \mathbf{V}(\boldsymbol{\theta})\mathbf{e} \right\rangle$ and as $\langle \mathbf{S}(t), \mathbf{e} \rangle$, where $\mathbf{S}(t)$ is defined by (5) and $\mathbf{e}$ is an N dimensional standard normal random vector independent of $\boldsymbol{\theta}$ and t. Note that $||\mathbf{S}(t)|| = 1$ for all $t \in \mathcal{T}$. So, $Z(t)$ is a Gaussian random field on $\mathcal{T}$ with zero mean, unit variance, and covariance function $\mathbf{S}(t)^T\mathbf{S}(t')$. Denote by a the minimum of the ratio of $\mathbf{l}^R$ and $\hat{\mathbf{l}}^R$ (see (5)), and let $\delta = \sup_{t \in \mathcal{T}}[\delta(t)/\{\sigma\|\mathbf{l}^R(t, \boldsymbol{\theta})\|\}]$, the difference in the regression estimator

due to the variation in estimating $\boldsymbol{\theta}$. Then it is straightforward to see that

$$P(c) \le \text{pr}\left(\sup_{t\in\mathcal{T}}\left[\left\{|Z(t)| + \frac{\delta(t)/\sigma}{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}\right\}\frac{\|\mathbf{l}^R(t,\boldsymbol{\theta})\|}{\|\mathbf{l}^R(t,\hat{\boldsymbol{\theta}})\|}\right] > c\frac{\hat\sigma}{\sigma}\right)$$
$$\le \text{pr}\left\{\sup_{t\in\mathcal{T}}|Z(t)| > c\frac{\hat\sigma}{\sigma}a - \delta\right\} \;\le\; 2\text{pr}\left\{\sup_{t\in\mathcal{T}} Z(t) > c\frac{\hat\sigma}{\sigma}a - \delta\right\}.$$

Since both ML and REML estimators are consistent under some regularity conditions, we have $a = 1 + o_p(1), \delta = o_p(1)$, as $n \to \infty$. So we can bound a by a positive constant a' such that $a \ge a'$ in probability as $n \to \infty$ and

$$P(c) \le 2\text{pr}\left\{\sup_{t\in\mathcal{T}} Z(t) > \frac{c}{\nu^{1/2}}\frac{\nu^{1/2}\hat\sigma}{\sigma}a'\right\} + o(\alpha) \tag{A.2}$$

Faraway and Sun (1994) set a' identically equal to one and used corrections γ' (due to estimating σ^2 when the weights are unknown) and δ' (a positive constant such that $\delta \le \delta'$ in probability as $n \to \infty$). Our simulation results (Section 4 and Section 5) confirm that the bias δ and the small finite-sample dependence between $\hat\sigma^2$ and $\hat\beta$ (which is related to γ') are negligible, at least for the estimators and models considered in this paper and for our estimators of a'. Other applications may require an additional adjustment for estimating γ' and δ', which may be carried out by combining results of Faraway and Sun (1994) with the estimators of a' given in Section 3.2.

When either adjusted ML or REML estimators are used, $\nu^{1/2}\hat\sigma/\sigma$ has *approximately* a χ distribution with degrees of freedom ν, that is, it has a probability density function $f(y,\nu) = y^{\nu-1}\exp(-y^2/2)/\{2^{\nu/2-1}\Gamma(\nu/2)\}$. It is also easy to see from Equation (3.7) in Jiang (1998) that $\hat\sigma$ is asymptotically independent of $Z(t)$. Thus, if the manifold $\{\mathbf{S}(t),\, t \in \mathcal{T}\}$ has no boundary (e.g. a closed circle in the right panel of Figure 5), an application of the tail approximation to the maximum of a Gaussian random field gives

$$P(c) \dot\le 2\int_0^\infty \text{pr}\left\{\sup_{t\in\mathcal{T}} Z(t) \ge \frac{c'}{\nu^{1/2}}y\right\} f(y,\nu)dy + o(\alpha)$$
$$\approx \int_0^\infty \kappa_0\frac{1}{\pi}\exp\{-\frac{1}{2}(\frac{c'}{\nu^{1/2}}y)^2\}f(y,\nu)\,dy \tag{A.3}$$
$$\approx \frac{\kappa_0}{\pi}\left(1 + \frac{c'^2}{\nu}\right)^{-\nu/2} \tag{A.4}$$

where $c' = ca'$, κ_0 is defined in the proposition, and $\dot\le$ is an approximate inequality as $\alpha \to 0$. In the case that the manifold $\{\mathbf{S}(t),\, t \in \mathcal{T}\}$ has a boundary

there is an additional term in (A.3) and hence also in (A.4), described below. Note the fact that the tail probability $P\{\max_{t\in\mathcal{T}} Z(t) > z\}$ corresponds to a probability computed on a tubular area about the manifold $\{\mathbf{S}(t), t \in \mathcal{T}\}$ with a radius depending on z (cf. Sun (1993)). So, in the case that $\mathcal{T}$ is a finite interval and the two end points of $S(t)$ are not equal, the tubular area has two half disks (see the left panel of Figure 5 and compare this with the tubular area without caps shown in the right panel). By (A.2) this results in the additional term

$$2P\{Z(a) \geq c'\hat{\sigma}/\sigma\} = P\{|t_\nu| > c'\}, \tag{A.5}$$

accounting for the coverage probability in the two half disks. Here a is one of the two end points of $\mathcal{T}$, different from the a considered earlier. In more general cases, the additional term is $E \cdot P\{|t_\nu| > c'\}$, where E is Euler-Poincaré characteristic of $\{\mathbf{S}(t), t \in \mathcal{T}\}$ (cf. Sun and Loader (1994)). The constant E is one if $\mathcal{T}$ is a finite interval and the end values of $\mathbf{S}(t)$ are not equal. (See Figure 5 and the discussion in Section 3.1.) This derivation is similar to the derivation of the boundary correction term in Sun and Loader (1994).

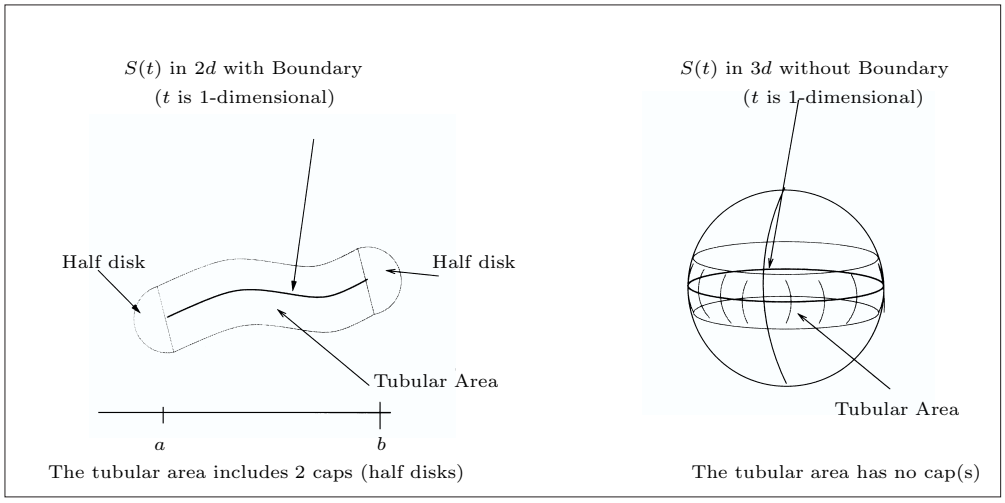

Figure 5. Tubular Area of $\{S(t), t \in \mathcal{T}\}$. If $\{S(t), t \in \mathcal{T}\}$ has two boundary points a and b, then the tubular area includes two half disks. If $\{S(t), t \in \mathcal{T}\}$ has no boundary, then the tubular area is the region without the two caps.

Although setting the right hand side of formula (A.4) (plus $E \cdot P\{|t_\nu| > c'\}$) to α gives a conservative confidence band, the nominal level will be closely approximated if a' is close to a. Thus, the final (conservative) approximation

formula is

$$\alpha \approx E \cdot P(|t_\nu| > ca') + \frac{\kappa_0}{\pi}\left(1 + \frac{c^2 a'^2}{\nu}\right)^{-\nu/2}.$$

When $\mathcal{T}$ is reduced to one point, the second term of (6) is zero, $E = 1$, and (6) reduces to the familiar t interval, approximately.

In other applications of the tube formula, the approximation "$\approx$" in (A.4) should be replaced by "$\leq$" when the tube self-overlaps. So, this would definitely produce a conservative formula. Fortunately, this cannot occur within the class of regression functions we have chosen when α is not too large. See Sun and Loader (1994) for more discussion on this point.

Derivation of the Scheffé-type bands shown in Figure 3. Define $\mathbf{l}$, $\hat{\mathbf{l}}$, $\mathbf{l}^R$ as in the derivation of (6) and let $\boldsymbol{\epsilon} = \mathbf{V}\mathbf{e}$ where $\mathbf{V}$ is the Cholesky triangle of $\mathbf{R}$. Then we wish to determine c such that

$$\alpha \geq \text{pr}\left\{|\hat{f}(t) - f(t)| > c\hat{\sigma}\|\mathbf{l}^R(t, \hat{\boldsymbol{\theta}})\|, \text{ for some } t \in \mathcal{T}\right\}.$$

Let $P(c)$ equal the right hand side. Suppose c satisfies

$$\begin{aligned} P(c) &= \text{pr}\left\{\sup_{t\in\mathcal{T}} \frac{|\hat{\mathbf{l}}^T\mathbf{Y} - \hat{\mathbf{l}}^T\mathbf{X}\boldsymbol{\beta}|}{\sigma\|\hat{\mathbf{l}}^R\|} > c\frac{\hat{\sigma}}{\sigma}\right\} \\ &= \text{pr}\left\{\sup_{t\in\mathcal{T}} \frac{|\hat{\mathbf{l}}^T\boldsymbol{\epsilon}|}{\sigma\|\hat{\mathbf{l}}^R\|} > c\frac{\hat{\sigma}}{\sigma}\right\} \\ &= \text{pr}\left\{\sup_{\mathbf{t}\in\mathcal{T}} \frac{|\left\langle\hat{\mathbf{l}}^R, \mathbf{e}\right\rangle|}{\sigma\|\hat{\mathbf{l}}^R\|} > c\frac{\hat{\sigma}}{\sigma}\right\}. \end{aligned}$$

Let $\mathcal{L}$ be the linear space spanned by the columns of $\mathbf{V}^{-1}\mathbf{X}$ and let $\mathbf{L} = (\mathbf{V}^T)^{-1}\{\mathbf{X}^T\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{X}\}^{-1}\mathbf{V}^{-1}$ be the projection matrix to $\mathcal{L}$. Since $\mathbf{l}^R/\|\mathbf{l}^\mathbf{R}\|$ is a unit vector in $\mathcal{L}$, we have

$$\begin{aligned} P(c) &\leq \text{pr}\left\{\sup_{\mathbf{u}\in\mathcal{L},\|\mathbf{u}\|=1} |\langle\mathbf{u}, \mathbf{e}\rangle|/\sigma > c\frac{\hat{\sigma}}{\sigma}\right\} \\ &= \text{pr}\left\{\frac{||\mathbf{L}\mathbf{e}||^2/\sigma^2}{\hat{\sigma}^2/\sigma^2} > c^2\right\} \end{aligned}$$

If $\boldsymbol{\theta}$ is finite dimensional and $E(\mathbf{e}^3) = 0$, the denominator and numerator of the last expression are asymptotically independent (Jiang (1998)), so $||\mathbf{L}\mathbf{e}||^2/[\hat{\sigma}^2\,(1+q)]$ has an approximate F distribution with $1+q$ and $n-q-1$ degrees of freedom. Thus $c \approx \sqrt{(1+q)F^\alpha_{1+q,\nu}}$, where F^α is the upper α quantile of the F distribution.

To account for the additional variability due to estimation of $\boldsymbol{\theta}$, we use the correction constant $\hat{a}'$ described in Section 3.

References

Bohrer, R. and Francis, G. K. (1972). Sharp one-sided confidence bounds over positive regions. *Ann. Math. Statist.* **43**, 1541-1548.

Carter, R. L., Resnick, M. B., Ariet, M., Shieh, G. and Vonesh, E. F. (1992). A random coefficient growth curve analysis of mental development in low-birth-weight infants. *Statist. in Medicine* **11**, 243-256.

Casella, G. and Strawderman, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *J. Amer. Statist. Assoc.* **75**, 862-868.

Cox, D. R. (1965). A remark on multiple comparison methods. *Technometrics* **7**, 223-224.

Elston, R. C. and Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bounds. *Biometrics* **18**, 148-159.

Faraway, J. and Sun, J. (1995). Simultaneous confidence bands for linear regression with heteroscedastic errors. *J. Amer. Statist. Assoc.* **90**, 1094-1098.

Gafarian, A. V. (1964). Confidence bands in straight line regression. *J. Amer. Statist. Assoc.* **59**, 182-213.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972-985.

Guthrie, D. and Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology* **28**, 240-244.

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation of variance in non-parametric regression. *J. Multivariate Anal.* **27**, 228-254.

Halperin, M., Rastogi, S. C., Ho, I. and Yang, Y. Y. (1967). Shorter confidence bands in linear regression. *J. Amer. Statist. Assoc.* **62**, 1050-1068.

Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778-796.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problem. *J. Amer. Statist. Assoc.* **72**, 320-340.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-820.

Jiang, J. (1998). Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statist. Sinica* **8**, 861-885.

Kreyszig, E. (1968). *Introduction to Differential Geometry and Riemannian Geometry.* University of Toronto Press.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lindstrom, M. J. and Bates, J. H. (1990). Random-effects models for longitudinal data. *Biometrics* **46**, 673-687.

Loader, C. (1993). Nonparametric regression, confidence bands and bias correction. *Proceedings of the* 25*th Symposium on the Interface Between Computer Science and Statistics*, 131-136.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5**, 746-762.

Naiman, D. (1987). Simultaneous confidence bounds in multiple regression using predictor variable constraints. *J. Amer. Statist. Assoc.* **82**, 214-219.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: a review. *J. Educational Statist.* **13**, 85-116.

SAS Institute Inc. (1992). SAS Technical Report P-229, *SAS/STAT Software: Changes and Enhancements, Release* 6.07, SAS Institute Inc., Cary, NC, Chapter 16, The MIXED Procedure, 287-366.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley, New York.

Schwartz, G. G., Schaefer, S., Gober, J., Meyerhoff, D. J., Smekal, A., Massie, B. and Weiner, M. W. (1990). Myocardial high-energy phosphates in reactive hyperemia. *Amer. J. Phys.* **259**, H1190-H1196.

Seber, G. (1977). *Linear Regression Analysis*. John Wiley, New York.

Stewart, P. W. (1987). Line-segment confidence bands for repeated measures. *Biometrics* **43**, 629-640.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* **21**, 34-71.

Sun, J. and Loader, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22**, 1328-1345.

Uusipaikka, Esa (1983). Exact confidence bands for linear regression over intervals. *J. Amer. Statist. Assoc.* **78**, 638-644.

Vonesh, E. F. and Carter, R. L. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics* **43**, 617-628.

Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* **48**, 1-17.

Wynn, H. P. (1984). An exact confidence band for one-dimensional polynomial regression. *Biometrika* **71**, 375-380.

Department of Statistics, Case Western Reserve University, Cleveland, Ohio 44106-7054, U.S.A
E-mail: jiayang@sun.cwru.edu

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.
E-mail: jonraz@umich.edu

Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109-1027, U.S.A.
E-mail: faraway@umich.edu

(Received August 1997; accepted September 1998)