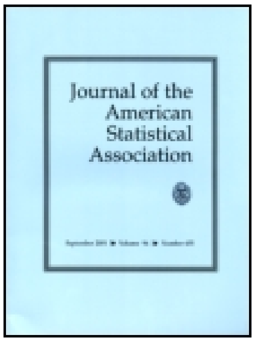

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Simultaneous Confidence Bands for Penalized Spline Estimators

Tatyana Krivobokova, Thomas Kneib & Gerda Claeskens

To cite this article: Tatyana Krivobokova, Thomas Kneib & Gerda Claeskens (2010) Simultaneous Confidence Bands for Penalized Spline Estimators, Journal of the American Statistical Association, 105:490, 852-863, DOI: 10.1198/jasa.2010.tm09165

To link to this article: https://doi.org/10.1198/jasa.2010.tm09165

Simultaneous Confidence Bands for Penalized Spline Estimators

Tatyana KRIVOBOKOVA, Thomas KNEIB, and Gerda CLAESKENS

In this article we construct simultaneous confidence bands for a smooth curve using penalized spline estimators. We consider three types of estimation methods: (a) as a standard (fixed effect) nonparametric model, (b) using the mixed-model framework with the spline coefficients as random effects, and (c) a full Bayesian approach. The volume-of-tube formula is applied for the first two methods and compared with Bayesian simultaneous confidence bands from a frequentist perspective. We show that the mixed-model formulation of penalized splines can help obtain, at least approximately, confidence bands with either Bayesian or frequentist properties. Simulations and data analysis support the proposed methods. The R package *ConfBands* accompanies the article.

KEY WORDS: B-spline; Bayesian penalized spline; Confidence band; Mixed model; Penalization.

1. INTRODUCTION

Penalized spline smoothing has received much attention over the last decade. Eilers and Marx (1996) coined the term "P-spline" estimator to denote a version of the O'Sullivan (1986) estimator with a simplified penalty matrix. The idea is to estimate the function of interest by some spline. Thereby a generous basis dimension is taken, and penalization with an integrated squared derivative of the spline function helps avoid overfitting. A small parameter dimension, a flexible choice of basis and penalties, and direct links to mixed and Bayesian models have made this smoothing technique popular (see Ruppert, Wand, and Carroll 2003 for examples and applications).

The theoretical properties of penalized splines have been less well explored. Some first results were presented by Hall and Opsomer (2005), Li and Ruppert (2008), and Kauermann, Krivobokova, and Fahrmeir (2009). Recently Claeskens, Krivobokova, and Opsomer (2009) showed that, depending on the number of knots, the asymptotic scenario of the penalized spline estimator is similar to that of either the regression spline or smoothing spline estimator. In this manner, the optimal asymptotic orders for the number of spline functions and the smoothing parameter were obtained. These new results can now be applied for inference, particularly for constructing simultaneous confidence bands.

In general, simultaneous confidence bands for a function f are constructed by studying the asymptotic distribution of $\sup_{a\le x\le b}|\hat{f}(x)-f(x)|$. The approach of Bickel and Rosenblatt (1973) relates this to a study of the distribution of $\sup_{a\le x\le b}|Z(x)|$, with $Z(x)$ a (standardized) Gaussian process satisfying certain conditions, which they showed to have an asymptotic extreme value distribution. This approach to constructing confidence bands has been used in the context of nonparametric estimation by, among others, Härdle (1989) for M-estimators and Claeskens and Van Keilegom (2003) for local polynomial likelihood estimators. Hall (1991) studied the convergence of normal extremes and found them to be slow, with the consequence that all those confidence bands do not perform satisfactorily for small samples, and bootstrap methods are often applied (see, e.g., Neumann and Polzehl 1998; Claeskens and Van Keilegom 2003).

Knafl, Sacks, and Ylvisaker (1985) and Hall and Titterington (1988) developed confidence bands based on large-sample upper bounds for the size of $\sup_{a\le x\le b}|\hat{f}(x)-f(x)|$. The main challenge with this approach is to take the bias of a nonparametric estimator into account. The choice of the smoothing parameter is a delicate matter as well. Eubank and Speckman (1993) applied a similar technique to obtain confidence bands for a periodic twice-differentiable function, using a kernel estimator. In this manner, the smoothing parameter was chosen data-driven and the bias was approximated using the estimator of the second derivative of the underlying mean function. Xia (1998) extended the approach of Eubank and Speckman (1993) using local polynomial estimators.

Another attractive approach is to construct confidence bands based on the volume of tube formula. Sun (1993) studied the tail probabilities of suprema of Gaussian random processes, which can be used for the construction of simultaneous confidence bands. It turns out that the leading coefficient in the approximation of the tail probability $P(\sup_{a\le x\le b}|Z(x)|>c)$ for $c\to\infty$ is connected through Weyl's (1939) formula for the volume of a tube of a manifold [also referred to as a Hotelling (1939) formula] to the volume of the manifold embedded in a unit sphere. The main attraction of this method is its straightforward extendability to more general and high-dimensional settings. However, the problems of choosing the smoothing parameter and handling the bias remain important. Sun and Loader (1994) suggested a bias correction for a particular class of functions, but left the smoothing parameter choice open. Zhou, Shen, and Wolfe (1998, theorem 4.2) used the volume-of-tube formula for estimation by regression splines (without using a penalty), but

Tatyana Krivobokova is Associate Professor, Courant Research Center PEG and Institute for Mathematical Stochastics, Georg August Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany (E-mail: *tkrivob@gwdg.de*). Thomas Kneib is Professor, Mathematics Department, Carl von Ossietzy Universität Oldenburg, 26111 Oldenburg, Germany. Gerda Claeskens is Professor, ORSTAT and Leuven Statistics Research Center, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium. The authors thank the editor, the referees, and Dr. Tatiyana Apanasovich for constructive remarks. They also thank Manuel Wiesenfarth for the substantial improvements to the R program used for the Bayesian confidence bands. The first author acknowledges the support of the German Research Foundation (Deutsche Forschungsgemeinschaft) as part of the Institutional Strategy of the University of Göttingen.

Journal of the American Statistical Association
June 2010, Vol. 105, No. 490, Theory and Methods
DOI: 10.1198/jasa.2010.tm09165

did not account for the bias, leading to undercoverage. We use this method to construct confidence bands for estimation by penalized spline estimators in the fixed and mixed-model framework.

In contrast to the frequentist setting, the Bayesian confidence bands are constructed based on the posterior distribution of the underlying process, given the data. Even though highest posterior density credible bands should be optimal from a theoretical perspective, they are generally hard to obtain, particularly when the estimation is based on Markov chain Monte Carlo (MCMC) simulation techniques (as is the case in this article and as is common practice in complex statistical models). In this case the posterior density is not available, and, consequently, confidence intervals are typically constructed based on sample quantiles obtained from the Monte Carlo output. The difficulty in constructing simultaneous confidence bands then lies in combining the sample quantiles such that a simultaneous coverage for a vector parameter is achieved. Besag et al. (1995) proposed combining appropriate order statistics of the univariate samples. Crainiceanu et al. (2007) considered simultaneous confidence bands when posterior normality for the parameter vector can be assumed. Held (2004) constructed simultaneous posterior probability statements about vector parameters based on a Rao–Blackwellized estimate of the posterior density. In principle, the posterior probabilities could be inverted to obtain a highest posterior density credible band, but the computational burden is high because additional simulations are required to obtain the posterior density estimate.

In this work we demonstrate the advantages of the mixed-model formulation, which combines both frequentist and Bayesian approaches. We develop a new approach for the mixed model-based confidence bands, as well as a new Bayesian simultaneous confidence band. The confidence bands obtained in the marginal mixed-model framework are identical to the Bayesian ones up to an unaccounted variability due to variance estimation. Because the Bayesian confidence bands (and thus the marginal mixed model-based ones) tend to be conservative in the nonparametric setting (see Cox 1993), we show how the confidence bands with the approximately frequentist properties can be obtained using the mixed-model representation of penalized splines. In this approach, no explicit bias estimation is necessary, and the smoothing parameter is estimated in the usual way from the corresponding (restricted) likelihood.

We introduce the curve estimators in Section 2. In Sections 3 and 4 we construct confidence bands for each setting, obtaining a new result for the mixed models as well as for the Bayesian method. We follow with a comparison and discussion in Section 5, and simulation results and a data example in Sections 6 and 7.

2. PENALIZED SPLINES IN THREE FRAMEWORKS

We wish to construct a simultaneous confidence band for an unknown smooth function $f \in C^q([a, b])$, which is a q times continuously differentiable function. We have observations (Y_i, x_i), with $x_i \in [a, b]$, $i = 1, \ldots, n$, from the model

$$Y_i = f(x_i) + \varepsilon_i. \tag{1}$$

The residuals ε_i are assumed to be independent and identically distributed as $\mathrm{N}(0, \sigma_\epsilon^2)$.

We first introduce some notation and explain the three frameworks for penalized splines.

2.1 Penalized Spline Estimator

We denote by $S(p+1; \underline{\tau})$ the set of spline functions of degree p with knots $\underline{\tau} = \{a = \tau_0 < \tau_1 < \cdots < \tau_K < \tau_{K+1} = b\}$. This set consists of all functions that are a polynomial of degree p on each interval $[\tau_j, \tau_{j+1}]$, and are $p-1$ times continuously differentiable. The set $S(1, \underline{\tau})$ consists of piecewise constant functions with jumps at the knots.

A penalized spline estimator of degree p based on the set of knots $\underline{\tau}$ is the solution to

$$\min_{s(x) \in S(p+1; \underline{\tau})} \left[\sum_{i=1}^{n} \{Y_i - s(x_i)\}^2 + \lambda \int_a^b \left\{ s^{(q)}(x) \right\}^2 dx \right], \tag{2}$$

with $q \le p$. Denote by a row vector $\mathbf{P}(x, \underline{\tau}) = \{P_1(x, \underline{\tau}), \ldots, P_{K+p+1}(x, \underline{\tau})\}$ a basis for $S(p+1, \underline{\tau})$. One example is the set of polynomial and piecewise polynomial functions $\{1, x, \ldots, x^p, (x-\tau_1)_+^p, \ldots, (x-\tau_K)_+^p\}$; another example is a basis of B-spline functions of degree p. With this notation, the spline function can be written as $s(x) = \mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}$, with an unknown parameter $\boldsymbol{\theta}$ of length $K + p + 1$ and (2) can be represented as minimization problem over $\boldsymbol{\theta}$.

The penalty in (2) is the integrated squared qth derivative of the spline function, which is assumed to be finite. Let $\mathbf{D}$ be the matrix such that $\int_a^b [\{\mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}\}^{(q)}]^2\, dx = \boldsymbol{\theta}^t \mathbf{D} \boldsymbol{\theta}$. Define the spline basis matrix $\mathbf{P} = \{\mathbf{P}(x_1, \underline{\tau})^t, \ldots, \mathbf{P}(x_n, \underline{\tau})^t\}^t$, and the response vector $\mathbf{Y} = (Y_1, \ldots, Y_n)^t$, then, for a given λ, the penalized spline estimator can be written as

$$\tilde{\mathbf{f}} = \mathbf{P}\tilde{\boldsymbol{\theta}} = \mathbf{P}(\mathbf{P}^t\mathbf{P} + \lambda \mathbf{D})^{-1}\mathbf{P}^t\mathbf{Y}, \tag{3}$$

where the estimator $\tilde{\mathbf{f}} = \{\tilde{f}(x_1), \ldots, \tilde{f}(x_n)\}^t$.

The penalty constant λ plays the role of a smoothing parameter. It can be estimated with any data-driven method that asymptotically minimizes the average mean squared error, such as (generalized) cross-validation or the Akaike information criterion (AIC). Replacing λ by its estimate $\hat{\lambda}$ leads to $\hat{\mathbf{f}} = \mathbf{P}(\mathbf{P}^t\mathbf{P} + \hat{\lambda}\mathbf{D})^{-1}\mathbf{P}^t\mathbf{Y}$.

2.2 Penalized Spline Estimators as Predictors in Mixed Models

A penalized spline estimator is equivalent to a best linear unbiased predictor (BLUP) in the corresponding mixed model (Brumback, Ruppert, and Wand 1999). To show this, we first decompose

$$\mathbf{P}\boldsymbol{\theta} = \mathbf{P}(\mathbf{F}_\beta \boldsymbol{\beta} + \mathbf{F}_u \mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \tag{4}$$

such that $(\mathbf{F}_\beta, \mathbf{F}_u)$ is of full rank, providing uniqueness of transformation, and $\mathbf{F}_\beta^t \mathbf{F}_u = \mathbf{F}_u^t \mathbf{F}_\beta = \mathbf{F}_\beta^t \mathbf{D} \mathbf{F}_\beta = 0$, $\mathbf{F}_u^t \mathbf{D} \mathbf{F}_u = \mathbf{I}_{K+p+1-q}$, ensuring that only coefficients $\mathbf{u}$ are penalized. In this manner, $\mathbf{P}$ is a $n \times (K+p+1)$, $\mathbf{X}$ is a $n \times q$, and $\mathbf{Z}$ is a $n \times \widetilde{K}$ matrix, with $\widetilde{K} = K + p + 1 - q$. There are several approaches to obtaining such a decomposition (for more details, see, e.g., Durban and Currie 2003 or Fahrmeir, Kneib, and Lang 2004). If we assume that $\mathbf{Y}|\mathbf{u} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma_\epsilon^2 \mathbf{I}_n)$ and $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_{\widetilde{K}})$, this leads to the standard linear mixed model with the BLUP

$$\tilde{\mathbf{f}}_m = \mathbf{P}_m \tilde{\boldsymbol{\theta}}_m = \mathbf{P}_m \left(\mathbf{P}_m^t \mathbf{P}_m + \frac{\sigma_\epsilon^2}{\sigma_u^2} \mathbf{D}_m \right)^{-1} \mathbf{P}_m^t \mathbf{Y},$$

where $\mathbf{P}_m = [\mathbf{X}, \mathbf{Z}]$, $\boldsymbol{\theta}_m = [\boldsymbol{\beta}, \mathbf{u}]$, $\mathbf{D}_m = \mathrm{diag}\{\mathbf{0}_q, \mathbf{1}_{\tilde{K}}\}$.

If we also replace σ_ϵ^2 and σ_u^2 with the corresponding (restricted) maximum likelihood estimators in the mixed model, this results in the estimated BLUP (EBLUP) $\hat{\mathbf{f}}_m = \mathbf{P}_m\hat{\boldsymbol{\theta}}_m = \mathbf{P}_m(\mathbf{P}_m^t\mathbf{P}_m + \hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2\mathbf{D}_m)^{-1}\mathbf{P}_m^t\mathbf{Y}$. Due to the construction of $\mathbf{P}_m$, there always exists a square-invertible matrix $\mathbf{L}$, such that $\mathbf{P} = \mathbf{P}_m\mathbf{L}$ and $\mathbf{D} = (\mathbf{L}^{-1})^t\mathbf{D}_m\mathbf{L}^{-1}$. Thus we do not further distinguish between the different forms for the model and penalty matrices. However, the notation with a subscript 'm,' as in $\hat{\mathbf{f}}_m$ and $\hat{\boldsymbol{\theta}}_m$, indicates that that the estimators are obtained in the mixed-model framework. Note that the smoothing parameter in this mixed-model formulation is the ratio of two variance components $\lambda = \sigma_\epsilon^2/\sigma_u^2$.

2.3 Bayesian Penalized Splines

In a Bayesian framework, the penalty on spline coefficients is related to a specific prior distribution for $\boldsymbol{\theta}$. For example, a quadratic penalty $\boldsymbol{\theta}^t\mathbf{D}\boldsymbol{\theta}/(2\sigma_\theta^2)$ is the special case of a Gaussian prior $\pi(\boldsymbol{\theta}) \propto \exp\{-\boldsymbol{\theta}^t\mathbf{D}\boldsymbol{\theta}/(2\sigma_\theta^2)\}$, where the scaled penalty $\mathbf{D}/\sigma_\theta^2$ equals the precision matrix of the prior. Assuming normality for the responses Y_i, the posterior $\pi(\boldsymbol{\theta}|\mathbf{Y})$ for the spline coefficients under this prior is given by

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{Y}) &\propto \pi(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\\ &\propto \prod_{i=1}^{n}\exp\left[-\frac{1}{2\sigma_\epsilon^2}\{Y_i - \mathbf{P}(x_i,\underline{\tau})\boldsymbol{\theta}\}^2\right]\\ &\quad\times \exp\left(-\frac{1}{2\sigma_\theta^2}\boldsymbol{\theta}^t\mathbf{D}\boldsymbol{\theta}\right), \qquad (5)\end{aligned}$$

where $\pi(\mathbf{Y}|\boldsymbol{\theta})$ corresponds to the likelihood of the observation model (1). By taking logarithms and multiplying with $-2\sigma_\epsilon^2$, maximizing the posterior distribution in (5) is equivalent to minimizing $\sum_{i=1}^n\{Y_i - \mathbf{P}(x_i,\underline{\tau})\boldsymbol{\theta}\}^2 + \sigma_\epsilon^2\boldsymbol{\theta}^t\mathbf{D}\boldsymbol{\theta}/\sigma_\theta^2$, so that the penalized spline estimator (3) and the posterior mode coincide for a fixed variance and smoothing parameter. Similar to the mixed-model interpretation of penalized splines, the smoothing parameter corresponds to the ratio of the error variance and the prior variance. The mixed-model representation is a simple reparameterization of the Bayesian formulation of penalized splines that avoids the partial impropriety in the Gaussian prior if $\mathbf{D}$ is rank-deficient. Fahrmeir, Kneib, and Lang (2004) used this connection to derive empirical Bayes estimators based on mixed-model methodology yielding posterior mode estimators.

In a fully Bayesian formulation, additional hyperpriors are assigned to the error variance σ_ϵ^2 and the prior variance σ_θ^2. The simplest and conjugate choices are inverse gamma distributions, and a standard choice is $\sigma_\epsilon^2 \sim \mathrm{IG}(0.001, 0.001)$ and $\sigma_\theta^2 \sim \mathrm{IG}(0.001, 0.001)$. Inferences in the fully Bayesian approach are then typically based on Markov chain Monte Carlo (MCMC) simulation techniques (see Brezger and Lang 2006 for details).

3. SIMULTANEOUS BAYESIAN CREDIBLE BANDS

In this section we focus on Bayesian credible bands derived from MCMC simulation output. In all approaches, we assume that we are interested in computing simultaneous credible bands for a collection of function evaluations $\hat{\mathbf{f}} = \mathbf{P}\hat{\boldsymbol{\theta}} = \{\hat{f}(x_1), \ldots, \hat{f}(x_n)\}^t$ based on simulation realizations $f^{(j)}(x_1), \ldots, f^{(j)}(x_n)$, $j = 1, \ldots, J$.

Note that the Bayesian confidence bands are conceptually different from frequentist confidence bands. The construction is based on the posterior distribution, and one seeks a confidence region I_α such that $P_{\mathbf{f}|\mathbf{Y}}(\mathbf{f} \in I_\alpha) = 1 - \alpha$,; that is, the coverage is defined in terms of the posterior distribution of $\mathbf{f} = \{f(x_1), \ldots, f(x_n)\}^t$ given the observed data $\mathbf{Y}$.

An obvious way to construct a simultaneous credible region for $\hat{\mathbf{f}}$ was outlined by Crainiceanu et al. (2007). Suppose that $\hat{\mathbf{f}}$ is the posterior mean estimator and that the posterior standard deviation $\sqrt{\widehat{\mathrm{var}}\{\hat{f}(x_i)\}}$ for each point contained in $\hat{\mathbf{f}}$ has been computed. By assuming approximate posterior normality and deriving the $(1-\alpha)$ sample quantile c_b of

$$\max_{i=1,\ldots,n}\left|\frac{f^{(j)}(x_i) - \hat{f}(x_i)}{\sqrt{\widehat{\mathrm{var}}\{\hat{f}(x_i)\}}}\right|, \qquad j = 1, \ldots, J, \qquad (6)$$

a simultaneous credible region is given by the hyperrectangular

$$\left[\hat{f}(x_i) - c_b\sqrt{\widehat{\mathrm{var}}\{\hat{f}(x_i)\}},\ \hat{f}(x_i) + c_b\sqrt{\widehat{\mathrm{var}}\{\hat{f}(x_i)\}}\right], \qquad i = 1, \ldots, n. \qquad (7)$$

These confidence bands implicitly rely on the approximate normality. In particular, the standard deviation is used as a measure of uncertainty (assuming symmetry of the posterior distribution), and the posterior mean is considered a center point. Thus the full posterior distribution information contained in the sample is not utilized.

Alternatively, we propose a new simultaneous credible band that avoids the assumption of posterior normality but is still based on pointwise measures of uncertainty. To be more specific, we base our considerations on the pointwise credible intervals derived from the $\alpha/2$ and $1-\alpha/2$ quantiles of the samples $f^{(j)}(x_1), \ldots, f^{(j)}(x_n)$, $j = 1, \ldots, J$. In a second step, these pointwise credible intervals are scaled with a constant factor until $(1-\alpha)100\%$ of all sampled curves are contained in the credible band. The rationale is as follows. The pointwise credible intervals indicate where information on the estimated curve is sparse, corresponding to wider intervals, or dense, corresponding to narrower intervals. In the approach of Crainiceanu et al. (2007), this information is obtained from posterior standard deviations. But this approach has the drawback of treating overestimation and underestimation of the penalized spline in a symmetric fashion, whereas the quantile-based approach allows for different uncertainties for overestimation and underestimation of the curve. This may be of particular relevance in local minima and maxima, where uncertainty might be attributed more strongly to one of the directions. Although such differences may be generally small in Gaussian smoothing situations, they typically will become more relevant in non-Gaussian observation models. For example, the likelihood, and thus also the posterior in binary regression models or for Poisson data, will be inherently asymmetric such that posterior normality will be questionable, at least for moderate sample sizes. In such situations, the proposed new band will be useful, because it involves possibly asymmetric local measures of uncertainty. Note that determining Bayesian simultaneous confidence bands in non-Gaussian regression models is easy, because all computations rely only on the sampled curves and do not involve the simulation model. A further advantage over the credible band approach of Crainiceanu et al. (2007) is that our proposal does not depend on a specific point estimator, because our credible band makes full use of the posterior sample information, considering a $1-\alpha$ sample of the *curves* to determine the required scaling factor.

4. SIMULTANEOUS CONFIDENCE BANDS WITH THE VOLUME–OF–TUBE FORMULA

4.1 The Use of the Volume-of-Tube Formula

The construction of simultaneous confidence bands using Weyl's (1939) volume-of-tube formula has been considered by, among others, Naiman (1986), Johansen and Johnstone (1990), and Sun and Loader (1994). Rigorous proofs have been given by Sun (1993). Here we sketch the basic ideas for completeness, which we use in subsequent sections.

We consider the regression model (1) and some unbiased estimator $\tilde{f}(x) = \mathbf{l}(x)^t\mathbf{Y}$ with $\operatorname{var}\{\tilde{f}(x)\} = \sigma_\epsilon^2\|\mathbf{l}(x)\|^2$. Because $\tilde{f}(x)$ is unbiased, $Z(x) = \{\tilde{f}(x) - f(x)\}\sigma_\epsilon^{-1}\|\mathbf{l}(x)\|^{-1}$ is a mean-0 Gaussian random field with $\operatorname{var}\{Z(x)\} = 1$ and

$$\operatorname{cov}\{Z(x_1), Z(x_2)\} = \left(\frac{\mathbf{l}(x_1)}{\|\mathbf{l}(x_1)\|}\right)^t \left(\frac{\mathbf{l}(x_2)}{\|\mathbf{l}(x_2)\|}\right) \equiv \sum_{i=1}^n v_i(x_1)v_i(x_2),$$

where $\sum_{i=1}^n v_i^2(x) = 1$. The set $V_n = \{\mathbf{v}(x) : x \in [a,b], \mathbf{v}(x) = (v_1(x), \ldots, v_n(x))\}$ is a one-dimensional manifold embedded in S^{n-1}, which is a unit sphere in $\mathbb{R}^n$. Let $\kappa_0 = \int_a^b \|\frac{d}{dx}\mathbf{v}(x)\|\,dx$ be the length of V_n, and define the vector $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{f}$. Then Sun and Loader (1994) obtained that

$$\begin{aligned} \alpha &= P\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon}|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c\right) \\ &= \frac{\kappa_0}{\pi}\exp(-c^2/2) + 2\{1-\Phi(c)\} + o\{\exp(-c^2/2)\}, \end{aligned} \tag{8}$$

with $\Phi(\cdot)$ denoting the distribution function of a standard normal distribution. If σ_ϵ is unknown and is estimated with some $\hat{\sigma}_\epsilon$ such that $\varsigma\hat{\sigma}_\epsilon^2/\sigma_\epsilon^2 \sim \chi_\varsigma^2$, then

$$\alpha \approx \frac{\kappa_0}{\pi}\left(1 + \frac{c^2}{\varsigma}\right)^{-\varsigma/2} + P(|t_\varsigma| > c), \tag{9}$$

with t_ς a t-distributed random variable with ς degrees of freedom. A value for c is obtained from (9), and the simultaneous $100(1-\alpha)\%$ confidence band for $f(x)$ for x in the interval $[a,b]$ is constructed as

$$\left[\tilde{f}(x) - c\hat{\sigma}_\epsilon\|\mathbf{l}(x)\|, \tilde{f}(x) + c\hat{\sigma}_\epsilon\|\mathbf{l}(x)\|\right]. \tag{10}$$

4.2 Simultaneous Confidence Bands for Penalized Spline Estimators

Now consider the penalized spline estimator with $\mathbf{l}(x) = \mathbf{P}(\mathbf{P}^t\mathbf{P} + \lambda\mathbf{D})^{-1}\mathbf{P}^t(x, \underline{\tau})$. In contrast to the setting of the previous section, $\mathbf{l}(x)$, as well as any other nonparametric estimator, is biased. A penalized spline estimator has two contributions to the bias. The approximation bias is due to the spline representation of the true function, whereas the shrinkage bias enters via the penalization. Theorem 1 of Claeskens, Krivobokova, and Opsomer (2009) stated that, depending on some assumptions on the number of knots K, the sample size n and the penalty λ, the theoretical properties of the penalized spline estimators are similar to either those of regression splines or those of smoothing splines, with a clear breakpoint between the two cases. In the latter case—that is, if the penalized splines asymptotics is close to that of smoothing splines—the shrinkage bias dominates the average mean squared error, whereas the approximation bias vanishes with the growing number of knots. Namely, the average squared approximation bias is of order $O(K^{-2q})$ with $K \sim \widetilde{C}n^{\nu/(2q+1)}$ for some constants $\widetilde{C}$ and $\nu > 1$, whereas the average squared shrinkage bias is of order $O\{n^{-2q/(2q+1)}\}$. Subsequently we assume that sufficiently many knots are taken so that we can replace $f(x)$ with $\mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}$ directly, with a negligible approximation bias [see also assumption (A3) in the Appendix].

Thus, to construct confidence bands, we instead deal with

$$P_{\mathbf{Y}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon} + m(x)|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c^*\right) = \alpha,$$

with $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{P}\boldsymbol{\theta}$, the shrinkage bias $m(x) = \mathbf{l}(x)^t\mathbf{P}\boldsymbol{\theta} - \mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}$ and a critical value c^* that accounts for the bias. The critical value c^* is typically difficult to find because of the unknown bias. Ignoring the shrinkage bias can lead to serious undercoverage, as demonstrated in the simulation study presented in Section 6. Sun and Loader (1994) found that a plug-in correction with $m(x)$ replaced by an estimator failed badly, in some cases being even worse than no correction. They suggested a bias-correction procedure for a class of functions with Lipschitz continuous $m(x)/\|\mathbf{l}(x)\|$, based on the estimator of $\max_{x\in[a,b]} |m(x)|/\|\mathbf{l}(x)\|$. In their simulation study with the local polynomial regression estimates, the resulting coverage of the confidence bands appeared to be conservative and highly dependent on the choice of smoothing parameter. Sun and Loader (1994) did not discuss a strategy for the best smoothing parameter choice in their setting. Clearly, choosing a smoothing parameter smaller than the optimal one in the mean squared error sense reduces the bias; however, no general guideline is available on how small the smoothing parameter should be chosen. Note also that so far, we have assumed the smoothing parameter (λ or $\sigma_\epsilon^2/\sigma_u^2$) to be known. Replacing smoothing parameter by its estimator introduces an extra source of variability, which one has to account for.

In general, in this framework for penalized splines, one faces the same problems as for any other nonparametric estimator: the need for bias correction and an appropriate choice of smoothing parameter. In the next section we consider simultaneous confidence bands resulting from the mixed-model representation of penalized splines and propose a simple bias correction for the standard nonparametric setting considered in this section.

4.3 Simultaneous Confidence Bands With the Mixed-Model Representation of Penalized Splines

4.3.1 Confidence Bands Based on the Marginal Mixed Model. Here we consider the mixed-model representation of penalized splines. We approximate $f(x)$ by $\mathbf{P}(x, \underline{\tau})\theta = \mathbf{X}(x)\boldsymbol{\beta} + \mathbf{Z}(x)\mathbf{u}$, with $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \sigma_u^2\mathbf{I}_{\widetilde{K}})$ as in (4). Here $\mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}$ is random because of the randomness of $\mathbf{u}$. Note that Sun, Raz, and Faraway (1999) worked with a similar mixed model, but they aimed to build confidence bands around the marginal mean of $\mathbf{Y}$, that is, around $\mathbf{X}(x)\boldsymbol{\beta}$ only. From the standard results on mixed models, it is known that

$$Z_m(x) = \frac{\mathbf{P}(x, \underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\mathrm{var}\{\mathbf{P}(x, \underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}}$$
$$= \frac{\mathbf{P}(x, \underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\sigma_\epsilon^2 \mathbf{P}(x, \underline{\tau})(\mathbf{P}^t\mathbf{P} + \sigma_\epsilon^2/\sigma_u^2\mathbf{D})^{-1}\mathbf{P}(x, \underline{\tau})^t}} \sim \mathrm{N}(0, 1).$$

Because $\mathrm{cov}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) = \sigma_\epsilon^2(\mathbf{P}^t\mathbf{P} + \sigma_\epsilon^2/\sigma_u^2\mathbf{D})^{-1}$, $Z_m(x)$ is a nonsingular Gaussian mean-0 random field with $\mathrm{var}\{Z_m(x)\} = 1$ and

$$\mathrm{cov}\{Z_m(x_1), Z_m(x_2)\} = \left(\frac{\mathbf{l}_m(x_1)}{\|\mathbf{l}_m(x_1)\|}\right)^t \left(\frac{\mathbf{l}_m(x_2)}{\|\mathbf{l}_m(x_2)\|}\right)$$
$$\equiv \sum_{i=1}^{\widetilde{K}} v_{m,i}(x_1)v_{m,i}(x_2),$$

where $\mathbf{l}_m(x) = (\mathbf{P}^t\mathbf{P} + \sigma_\epsilon^2/\sigma_u^2\mathbf{D})^{-1/2}\mathbf{P}(x, \underline{\tau})^t$ is the $\widetilde{K} \times 1$ vector and $\mathbf{V}_{m,\widetilde{K}} = \{\mathbf{v}_m(x) : x \in [a, b], \mathbf{v}_m(x) = (v_{m,1}(x), \ldots, v_{m,\widetilde{K}}(x))\}$ is a one-dimensional manifold embedded in $S^{\widetilde{K}-1}$. We replace κ_0 in (8) with the length of the mixed-model manifold, $\kappa_{m,0} = \int_a^b \|\frac{d}{dx}\mathbf{v}_m(x)\|\, dx$, to obtain that

$$\alpha = P_{\mathbf{Y},\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}_m(x)^t\boldsymbol{\epsilon}_m|}{\sigma_\epsilon\|\mathbf{l}_m(x)\|} \geq c_m\right)$$
$$= P_{\mathbf{Y},\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon} + m(x, \mathbf{u})|}{\sigma_\epsilon\|\mathbf{l}_m(x)\|} \geq c_m\right) \quad (11)$$
$$= \frac{\kappa_{m,0}}{\pi}\exp(-c_m^2/2) + 2\{1 - \Phi(c_m)\} + o\{\exp(-c_m^2/2)\}, \quad (12)$$

with $\boldsymbol{\epsilon}_m = (\mathbf{P}^t\mathbf{P} + \sigma_\epsilon^2/\sigma_u^2\mathbf{D})^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}_{\widetilde{K}})$. An unknown σ_ϵ can be replaced by any consistent estimator leading to an expression similar to (9).

Thus our confidence band obtained in the marginal mixed-model framework is

$$[\tilde{f}_m(x) - c_m\hat{\sigma}_\epsilon\|\mathbf{l}_m(x)\|, \tilde{f}_m(x) + c_m\hat{\sigma}_\epsilon\|\mathbf{l}_m(x)\|]. \quad (13)$$

In practice, the smoothing parameter $\sigma_\epsilon^2/\sigma_u^2$ must be replaced by its estimator. The following lemma shows that the variability due to smoothing parameter estimation can be ignored in the mixed-model framework for n sufficiently large.

Lemma 1. Under assumptions (A1)–(A3) in the Appendix, it holds

$$\frac{\hat{\mathbf{l}}(x)^t\mathbf{Y} - \mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}}{\|\hat{\mathbf{l}}_m(x)\|} = \frac{\mathbf{l}_m(x)^t\boldsymbol{\epsilon}_m}{\|\mathbf{l}_m(x)\|} + O_p\left(n^{-1/(4q+2)}\right), \quad (14)$$

$$\frac{\hat{\mathbf{l}}(x)^t\mathbf{Y} - \mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}}{\|\hat{\mathbf{l}}(x)\|} = \frac{\mathbf{l}(x)^t\mathbf{Y} - \mathbf{P}(x, \underline{\tau})\boldsymbol{\theta}}{\|\mathbf{l}(x)\|} + O_p\left(n^{-1/(4q+2)}\right), \quad (15)$$

with $\hat{\mathbf{l}}_m(x) = \mathbf{l}_m(x; \hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2)$ and $\hat{\mathbf{l}}(x) = \mathbf{l}(x; \hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2)$.

The proof is given in the Appendix. Note that using a smaller q implies a smaller variability due to smoothing parameter estimation.

Using the same marginal mixed-model framework for penalized splines, Ruppert, Wand, and Carroll (2003) suggested a Monte Carlo procedure for estimating c_m; that is, a sufficiently large number ($N = 10{,}000$, say) of realizations of the random variable $(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) \overset{\text{approx.}}{\sim} \mathrm{N}(0, \hat{\sigma}_\epsilon^2(\mathbf{P}^t\mathbf{P} + \hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2\mathbf{D})^{-1})$ are generated, and the corresponding values of

$$C = \max_{j=1,\ldots,M}\left[\frac{\mathbf{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\widehat{\mathrm{var}}\{\mathbf{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}}\right]$$

are calculated for a specified grid of x values $z_1, \ldots, z_M$. Their critical value, $\hat{c}_m$, is the empirical $(1 - \alpha)$ quantile of the thus-obtained values $C_1, \ldots, C_N$. A simultaneous confidence band is given by the hyperrectangular

$$\Big[\mathbf{P}(z_j, \underline{\tau})\hat{\boldsymbol{\theta}}_m - \hat{c}_m\sqrt{\widehat{\mathrm{var}}\{\mathbf{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}},$$
$$\mathbf{P}(z_j, \underline{\tau})\hat{\boldsymbol{\theta}}_m + \hat{c}_m\sqrt{\widehat{\mathrm{var}}\{\mathbf{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}\Big] \quad (16)$$

for $j = 1, \ldots, M$. Note that this approach also does not take into account the variability due to variance parameter estimation. Thus (13) and (16) can be expected to be approximately equal. Obviously, (7) is in fact the Bayesian version of (16), where the variability due to parameter estimation is taken into account. But because in our simulation study reported in Section 6, we found no significant differences between the results obtained immediately from the tube formula and from (6), we believe that the tube formula offers an attractive alternative to the computationally intensive simulation-based techniques.

4.3.2 Confidence Bands Based on the Conditional Mixed Model. Let us now treat $\mathbf{u}$ in (4) as fixed and consider the probability

$$\alpha = P_{\mathbf{Y}|\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{P}(x, \underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c^*\right)$$
$$= P_{\mathbf{Y}|\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon} + m(x, \mathbf{u})|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c^*\right), \quad (17)$$

where $\mathbf{l}(x) = \mathbf{l}(x, \sigma_\epsilon^2/\sigma_u^2)$. Up to a smoothing parameter, this is exactly the probability discussed in Section 4.2. As mentioned in Section 4.2, a plug-in correction of the form

$$\left[\tilde{f}(x) - \left(c + \frac{m(x, \mathbf{u})}{\sigma_\epsilon\|\mathbf{l}(x)\|}\right)\sigma_\epsilon\|\mathbf{l}(x)\|,\right.$$
$$\left.\tilde{f}(x) + \left(c - \frac{m(x, \mathbf{u})}{\sigma_\epsilon\|\mathbf{l}(x)\|}\right)\sigma_\epsilon\|\mathbf{l}(x)\|\right], \quad (18)$$

with the bias $m(x, \mathbf{u})$ replaced by its estimate and c obtained from (8), performs poorly. We suggest instead using the critical value c_m obtained from (12) in place of c^*. To justify this, we compare (11) with (17) averaged over $\mathbf{u}$, that is, with

$$\alpha = E_{\mathbf{u}}P_{\mathbf{Y}|\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon} + m(x, \mathbf{u})|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c^*\right)$$
$$= P_{\mathbf{Y},\mathbf{u}}\left(\max_{x\in[a,b]} \frac{|\mathbf{l}(x)^t\boldsymbol{\epsilon} + m(x, \mathbf{u})|}{\sigma_\epsilon\|\mathbf{l}(x)\|} \geq c^*\right).$$

Thus

$$P_{\mathbf{Y},\mathbf{u}}\left(\max_{x\in[a,b]}|Z_m(x)|\geq c_m\right) = P_{\mathbf{Y},\mathbf{u}}\left(\max_{x\in[a,b]}|Z_m(x)|\frac{\|\mathbf{l}_m(x)\|}{\|\mathbf{l}(x)\|}\geq c^*\right). \quad (19)$$

Formally, from (19), it follows that $c_m \leq c^* \leq r_m c_m$, with $r_m = \max_{x\in[a,b]}\|\mathbf{l}_m(x)\|/\|\mathbf{l}(x)\|$. To get an idea how large r_m can be, note that if K is small and $\lambda \to 0$, then the penalized spline estimator converges to a projection (regression spline) estimator with $\mathbf{l}_m(x)\to\mathbf{l}(x)$ for all $x\in[a,b]$, so that $r_m \to 1$. As $K\to n$, a penalized spline estimator converges to a smoothing spline estimator for which it is known that $\sqrt{\mathrm{ave}_x\|\mathbf{l}_m(x)\|^2/\mathrm{ave}_x\|\mathbf{l}(x)\|^2}\approx\sqrt{2q/(2q-1)}$ (see Wahba 1983). For example, for $q=2$, we have $\sqrt{2q/(2q-1)}\approx 1.15$. Another way to look at c_m involves the following theorem.

Theorem 1. For the critical values c and c_m, obtained from (8) and (12) respectively,

$$c_m^2 = c^2 + 2\frac{\kappa_{m,0}-\kappa_0}{\kappa_0} + o(1),$$

where $o(1)$ converges to 0 as $c\to\infty$. In addition, if

(A4) $\kappa_{m,0}\kappa_0{}^{-1} = \max_{x\in[a,b]}\|\mathbf{l}_m(x)\|^2\|\mathbf{l}(x)\|^{-2} + o(1)$

is fulfilled and the mixed model (4) with $\mathbf{u}\sim \mathrm{N}(\mathbf{0},\sigma_u^2\mathbf{I}_{\tilde{K}})$ holds, then

$$c_m^2 = c^2 + \max_{x\in[a,b]}\frac{2\,\mathrm{var}_{\mathbf{u}}\{m(x,\mathbf{u})\}}{\sigma_\epsilon^2\|\mathbf{l}(x)\|^2} + o(1). \quad (20)$$

The proof of the theorem is provided in the Appendix, and the assumption (A4) is easy to check in practice using the R-package *ConfBands* that accompanies the article. We can conclude that the critical value c_m automatically accounts for the bias; thus one can build a confidence band,

$$\left[\tilde{f}_m(x) - c_m\hat{\sigma}_\epsilon\|\mathbf{l}(x,\sigma_\epsilon^2/\sigma_u^2)\|, \tilde{f}_m(x) + c_m\hat{\sigma}_\epsilon\|\mathbf{l}(x,\sigma_\epsilon^2/\sigma_u^2)\|\right], \quad (21)$$

which will have approximate coverage probability $1-\alpha$, assuming that sufficient knots are taken so that the approximation bias is negligible. Lemma 1 justifies replacing the smoothing parameter with its estimate. In fact, the confidence band (21) is similar in spirit to the bias correction suggested by Sun and Loader (1994), but we avoid explicit estimation of $\max_{x\in[a,b]}|m(x,\mathbf{u})|/\|\mathbf{l}(x)\|$, instead using the critical value c_m obtained from the marginal mixed-model framework.

5. CONFIDENCE BANDS IN THREE FRAMEWORKS

The confidence bands discussed in Sections 3 and 4 are obtained in different frameworks. They rely on different assumptions about the function f, the corresponding estimators use different smoothing parameter estimates, and the interpretation of the confidence bands is different as well. In the standard nonparametric model with f as a fixed, sufficiently smooth function, the frequentist confidence bands are calculated with respect to the distribution of the data, given the function f. In other words, if one samples the data with the same mean function f many times, then one can expect that in $100(1-\alpha)\%$ cases, the true f will be inside the bands. In the Bayesian framework, f is considered a sample path of a stochastic process, and one is looking for the posterior probability that the true f is within the band, given the data. In the finite-dimensional parametric setting, both intervals—frequentist and Bayesian—are asymptotically equivalent. The well-known Bernstein–von Mises theorem states that the posterior distribution of the finite-dimensional parameter vector around its posterior mean is close to the distribution of the maximum likelihood estimate around the truth, and here the Bayesian confidence sets have good frequentist coverage properties. Unfortunately, this is not true in the nonparametric regression context. In particular, Cox (1993) showed that the Bayesian coverage probability for Bayesian smoothing splines with Gaussian priors tends to be larger than $(1-\alpha)100\%$ (see also Freedman 1999). Thus we would expect our Bayesian credible bands to be conservative for $f\in C^q[a,b]$.

The mixed-model-based bands can be considered intermediate between these two. On the one hand, they can be viewed as empirical versions of the Bayesian confidence bands (with σ_ϵ, σ_u, and $\boldsymbol{\beta}$ treated as fixed) having the same interpretation. On the other hand, the mixed-model-based band can be seen as a confidence band averaged over $\mathbf{u}$. Thus, as shown in the previous section, the mixed-model formulation of penalized splines can help obtain confidence bands that have asymptotically either Bayesian or frequentist properties. Namely, the confidence band (13) is approximately equivalent to the Bayesian one and the band as defined in (21) has approximately frequentist properties. Our simulation results presented in Section 6 confirm this.

The following theorem gives the asymptotic width of the intervals considered in the article.

Theorem 2. Under assumptions (A1)–(A3), the width of the confidence bands (10), (13), and (21), based on the volume-of-tube formula for a penalized spline estimator, has asymptotic order $O_p(\sqrt{\log K^2}n^{-q/(2q+1)}) = O_p(\sqrt{\log n^{2\nu/(2q+1)}}\times n^{-q/(2q+1)})$, $\nu > 1$.

The proof is provided in the Appendix. This theorem also holds if the smoothing parameter is replaced by its estimator $\hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2$, as follows immediately from Lemma 1. It follows that the interval becomes narrower with increasing n and wider with increasing K.

Up to a constant $\sqrt{2\nu/(2q+1)}$, this asymptotic order coincides with that given by Eubank and Speckman (1993) for a twice-differentiable function ($q=2$), namely $O_p(\sqrt{\log n}n^{-2/5})$, which is slightly slower than the optimal rate of $(\log n/n)^{q/(2q+1)}$ of Hall and Titterington (1988). Eubank and Speckman (1993) stressed that Hall and Titterington (1988) "chose a smoothing parameter designed to minimize the length of their intervals, rather than MSE" and conjectured that their rate of $\sqrt{\log n}n^{-2/5}$ is the best attainable with the smoothing parameter that minimizes the MSE. Using penalized splines, one can get narrow intervals not only by taking a larger smoothing parameter, but also by choosing a smaller K. However, K should not be taken too small, to avoid a growing approximation bias. More discussion on a practical choice of K is provided in Section 6.

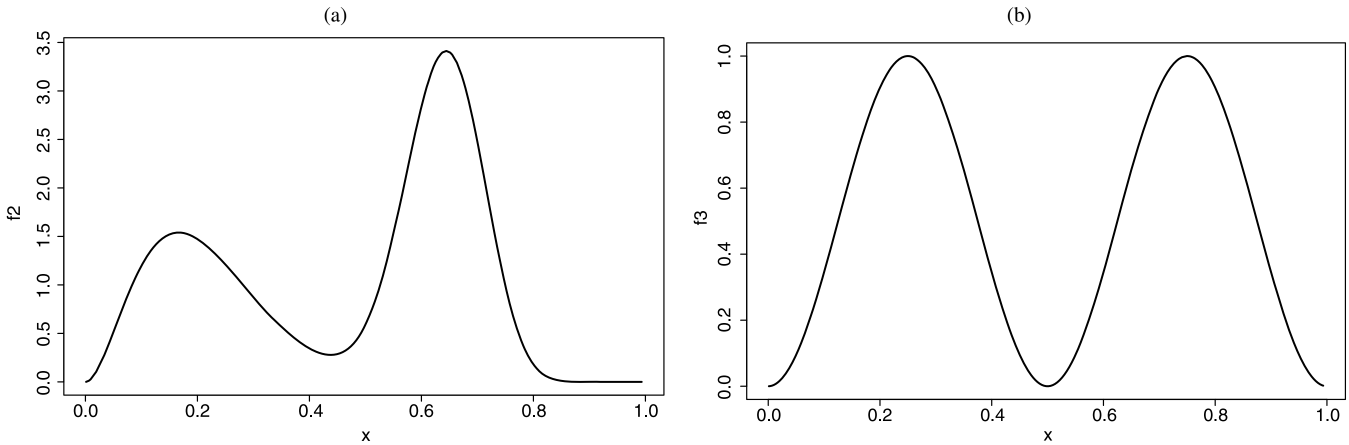

Figure 1. Functions used in the simulations: (a) $f_1(x)$, (b) $f_2(x)$.

6. SIMULATIONS

To assess the performance of the discussed approaches, we ran a simulation study. We considered two functions. The first function,

$$f_1(x) = \tfrac{6}{10}\beta_{30,17}(x) + \tfrac{4}{10}\beta_{3,11}(x),$$

with $\beta_{l,m}(x) = \Gamma(l+m)\{\Gamma(l)\Gamma(m)\}^{-1}x^{l-1}(1-x)^{m-1}$, was used by Wahba (1983), and the second,

$$f_2(x) = \sin^2\{2\pi(x-0.5)\}$$

was considered by Eubank and Speckman (1993) and Xia (1998). These functions are shown in Figure 1. The x values are taken to be uniformly distributed over [0, 1]. Three sample sizes were considered: small ($n = 50$), moderate ($n = 250$), and large ($n = 500$). The errors are taken to be independent $N(0, \sigma_\epsilon^2)$ distributed with $\sigma_\epsilon = 0.3$. Simulation results also are available for $\sigma_\epsilon = 0.1$ and $\sigma_\epsilon = 0.6$, but because no significant differences were found, we do not report these results here. We estimated the curves with different numbers of equidistant knots (i.e., $K = 15$, 40, 100 and $K = 200$), depending on the sample size. We used a B-spline basis of degree 3 and the integrated squared second derivative of the spline function as a penalty. The results for the 95% confidence bands reported in Table 1 are based on a Monte Carlo sample of size 1000.

The rows labeled **F** represent the coverage probabilities and corresponding areas for the confidence bands (10) built under the fixed-effects nonparametric model without any bias correction, as described in Section 4.2. Because the volume-of-tube formula assumes that the errors are normally distributed but does not require $n \to \infty$, we find no improvement in coverage with increasing n; this holds for both functions. As expected from the results of Theorem 2, the width (and thus the area) of the bands becomes smaller as n increases. Overall, the confidence bands in the standard nonparametric framework that ignore the bias have on average a 5%–10% smaller coverage for all combinations of n and K.

The rows labeled **C** show the coverage probability of the mixed-model-based bands, conditional on $\mathbf{u}$, as discussed in Section 4.3.2. They should result in a coverage probability close to the nominal value, which indeed is this case. In general, using a moderate number of knots ($K = 25$ to 50, depending on the sample size) is recommended in this framework, because the width of the interval is growing with increasing K. Overall, the bias correction resulting from the mixed-model representation of penalized splines is not only simple, but also very efficient.

The rows labeled **M** represent the coverage probabilities and corresponding areas for the confidence bands resulting from the marginal mixed-model framework, as discussed in Section 4.3.1. These bands appear to become increasingly conservative with growing K. This agrees with the findings of Cox (1993). Taking a small number of knots leads to a nearly parametric model in which the smoothing parameter is of little importance, which eliminates the differences between the mixed-model representation of penalized splines and its standard nonparametric formulation. Thus, in the marginal mixed-model framework, taking a small K will imply less conservative bands from the frequentist standpoint.

Finally, we consider the Bayesian confidence bands based on posterior normality (6) (denoted as **N**). These bands are conceptually close to the marginal mixed-model-based bands, as is also reflected in their similar behavior. This supports the asymptotic results of Lemma 1, which suggest that the variability due to smoothing parameter can be ignored. Our new proposed confidence bands, denoted by **U**, are typically somewhat wider than the **N** bands, but generally yield similar coverage, because the data were simulated from the normal distribution. Another method for constructing Bayesian simultaneous credible bands, based on order statitics of the samples, has been proposed by Besag et al. (1995). But our simulation study showed that the resulting credible bands suffered from serious undercoverage, and we omit further details here.

Overall, we found that the frequentist confidence bands without any bias correction led to undercoverage, whereas the Bayesian confidence bands typically became conservative with increasing K. The confidence bands based on the conditional mixed model result in confidence bands with coverage at best close to nominal.

7. EXAMPLES

Here we present two examples to illustrate our methods. In the first example, we consider the data on ratios of strontium isotopes found in fossil shells and their age, collected by

Table 1. Coverage probabilities and (areas) for $f_1(x)$ and $f_2(x)$, with nominal level 0.95 using **F** a fixed-effects model, **C** a mixed model conditional on **u**, and **M** a marginal mixed-effects model and Bayesian method based on normal posteriors **N** and univariate credible bands **U**. The range of the standard errors for the reported average areas is 2.6%–15.8% of the area for $f_1(x)$ and 2.9%–23.7% of the area for $f_2(x)$

	$n = 50$		$n = 250$			$n = 500$			
	$K = 15$	40	$K = 15$	40	100	$K = 15$	40	100	200
f_1									
F	0.91	0.86	0.89	0.88	0.90	0.87	0.90	0.90	0.89
	(0.91)	(0.91)	(0.43)	(0.44)	(0.45)	(0.31)	(0.33)	(0.33)	(0.33)
C	0.92	0.94	0.94	0.96	0.95	0.93	0.96	0.96	0.96
	(0.90)	(0.93)	(0.46)	(0.50)	(0.50)	(0.34)	(0.38)	(0.39)	(0.39)
M	0.96	0.97	0.96	0.99	0.99	0.94	0.99	0.99	0.99
	(0.98)	(1.04)	(0.48)	(0.56)	(0.58)	(0.35)	(0.43)	(0.44)	(0.44)
N	0.95	0.96	0.95	0.98	0.99	0.94	0.99	0.99	1.00
	(0.96)	(1.01)	(0.47)	(0.56)	(0.57)	(0.34)	(0.43)	(0.44)	(0.44)
U	0.95	0.96	0.95	0.98	0.99	0.95	0.98	0.99	1.00
	(0.97)	(1.02)	(0.48)	(0.57)	(0.59)	(0.35)	(0.44)	(0.45)	(0.45)
f_2									
F	0.85	0.86	0.78	0.73	0.81	0.86	0.88	0.85	0.88
	(0.70)	(0.71)	(0.32)	(0.32)	(0.33)	(0.25)	(0.25)	(0.25)	(0.25)
C	0.93	0.93	0.95	0.96	0.94	0.95	0.96	0.95	0.96
	(0.68)	(0.69)	(0.35)	(0.37)	(0.37)	(0.27)	(0.28)	(0.28)	(0.28)
M	0.97	0.97	0.98	0.98	0.99	0.98	1.00	0.99	1.00
	(0.76)	(0.78)	(0.39)	(0.42)	(0.42)	(0.29)	(0.32)	(0.32)	(0.32)
N	0.96	0.97	0.97	0.99	0.99	0.97	0.99	0.99	0.99
	(0.76)	(0.78)	(0.39)	(0.42)	(0.43)	(0.29)	(0.32)	(0.33)	(0.33)
U	0.96	0.97	0.97	0.98	0.99	0.97	0.99	0.99	0.99
	(0.77)	(0.79)	(0.40)	(0.43)	(0.44)	(0.30)	(0.33)	(0.33)	(0.34)

T. Bralower of the University of North Carolina. First analyzed by Chaudhuri and Marron (1999), this data set was used in section 6.5 on simultaneous confidence bands of Ruppert, Wand, and Carroll (2003). Figure 2 shows 106 observations. Ruppert, Wand, and Carroll (2003) found the critical value using the simulation procedure described in Section 4.3.1, building the band (16). From five independent simulations of $N = 10{,}000$ each, they obtained $\hat{c}_m \simeq 3.172, 3.198, 3.172, 3.201, 3.199$.

From Theorem 2, it follows that the critical values and thus the width of the band depends on the number of knots. We illustrate this result using the fossil shell data. For the estimation, we used penalized splines with a third-degree B-spline ba-

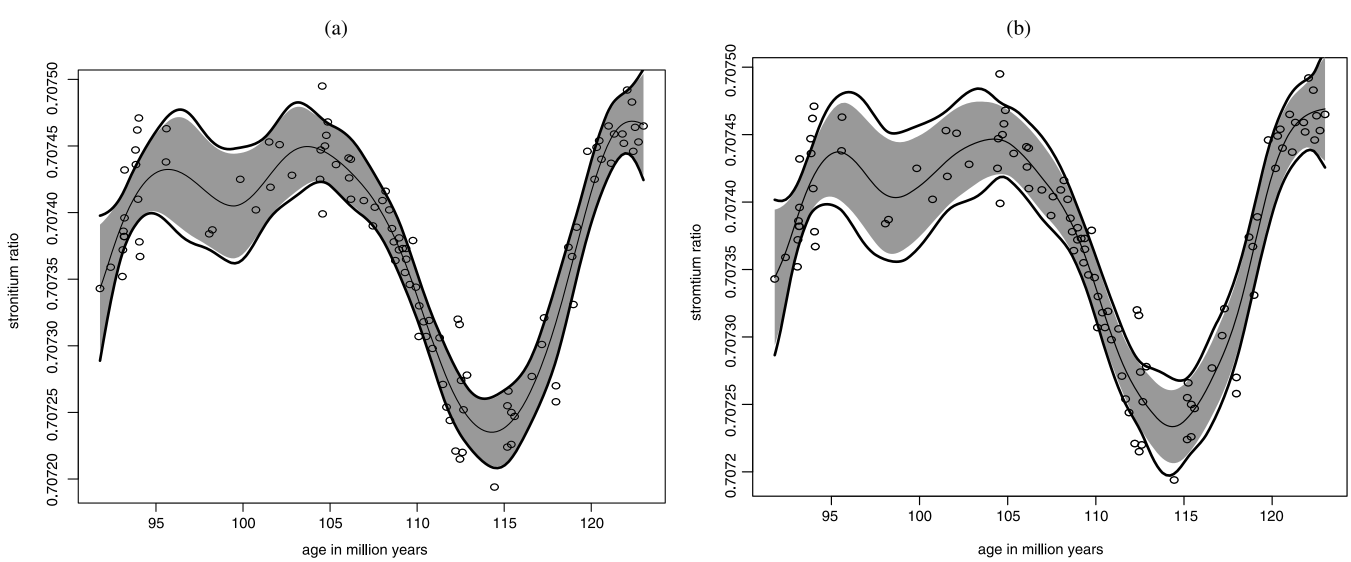

Figure 2. Fossil data with the confidence bands obtained from the conditional mixed model (shaded area) and the marginal mixed model (bold lines) (a) based on 10 knots, (b) based on 80 knots.

sis and second-order penalty. We found critical values from the marginal mixed model using 10 and 80 knots. The volume-of-tube formula (12) yielded critical values $c_m(K=10) = 3.229$ and $c_m(K=80) = 3.380$. The simulation-based critical values of Ruppert, Wand, and Carroll (2003) for the band (16) with $M = 150$ and $N = 10{,}000$ in five independent runs resulted in $\hat{c}_m(K=10) \simeq 3.107, 3.089, 3.080, 3.106, 3.092$, and to $\hat{c}_m(K=80) \simeq 3.251, 3.267, 3.255, 3.253, 3.272$. Obviously, the growing number of knots results in somewhat larger critical values. Note that Ruppert, Wand, and Carroll (2003) did not provide the number of knots used, but it can conjecture to be around 30–40.

The gray area in Figure 2 represents the simultaneous confidence bands (21) obtained from the conditional mixed model, and the bold lines indicate the bands (13) based on the marginal mixed model, using 10 knots (left) and 80 knots (right). The marginal mixed model-based bands (13) and (16), with the critical values obtained from the simulations and with the volume-of-tube formula, respectively, are indistinguishable on the plot. As mentioned earlier, the marginal mixed model-based bands become increasingly conservative as the number of knots grows, whereas the small number of knots leads to a nearly parametric model with a smaller difference between marginal and conditional mixed model-based bands, which is clearly shown in Figure 2.

The second example involves undernutrition in Kenyan children. The data come from the 2003 Kenya Demographic and Health Survey (2003 KDHS) carried out by the Kenya Central Bureau of Statistics, and are available free of charge from *www.measuredhs.com* for research purposes. We consider the so-called Z-score for stunting, depending on the child's age. The Z-score for stunting is defined as the standardized height for age, $Z = (H - m)/\sigma$, where H is child's height and m, σ are the median and the standard deviation of some reference population. Children with a Z-score < -2 are considered stunted. All 4686 observations are shown in the left plot of Figure 3. The data are cross-sectional (i.e., no same individuals) and available at 1-month intervals. Note the low signal-to-noise ratio for this data set. The gray area in the right plot shows the confidence band (21) using the conditional mixed-model representation of penalized splines based on 80 knots, and the bold lines indicate the corresponding marginal mixed model-based bands (13), which are indistinguishable from the Bayesian confidence bands. These confidence bands can be used to perform further analyses, such as a formal test for significance of bumps and dips between 1 and 4 years.

8. DISCUSSION

In this article we have considered the construction of simultaneous confidence bands in three frameworks for penalized splines. We used the volume-of-tube formula in the standard nonparametric setting and for the mixed-model representation of penalized splines. A full Bayesian analog of the mixed-model representation of penalized splines, as well as a new approach for Bayesian credible bands, were considered. We found that the volume-of-tube formula for the mixed-model formulation of penalized splines delivers results nearly identical to those for the full Bayesian framework, but at considerably less computational cost. Our main finding is that the mixed-model formulation of penalized splines also helps build the simultaneous bands with the frequentist coverage. Thus an explicit bias estimation is not needed, and the smoothing parameter is estimated from the corresponding (restricted) likelihood. Our approach appears to be effective in simulations, extremely fast, and easy to implement. The R package *ConfBands* that accompanies the article allows us to obtain all of the confidence bands discussed.

It is important to note that the volume-of-tube formula relies on the Gaussian distribution assumption for the errors. However, if the sample size is large and the central limit theorem applies, then the volume-of-tube formula is still valid for models with any non-Gaussian additive independent errors. Moreover, Loader and Sun (1997) showed that the volume-of-tube formula holds without modifications for spherically symmetric errors. In the linear regression context, some modifications of

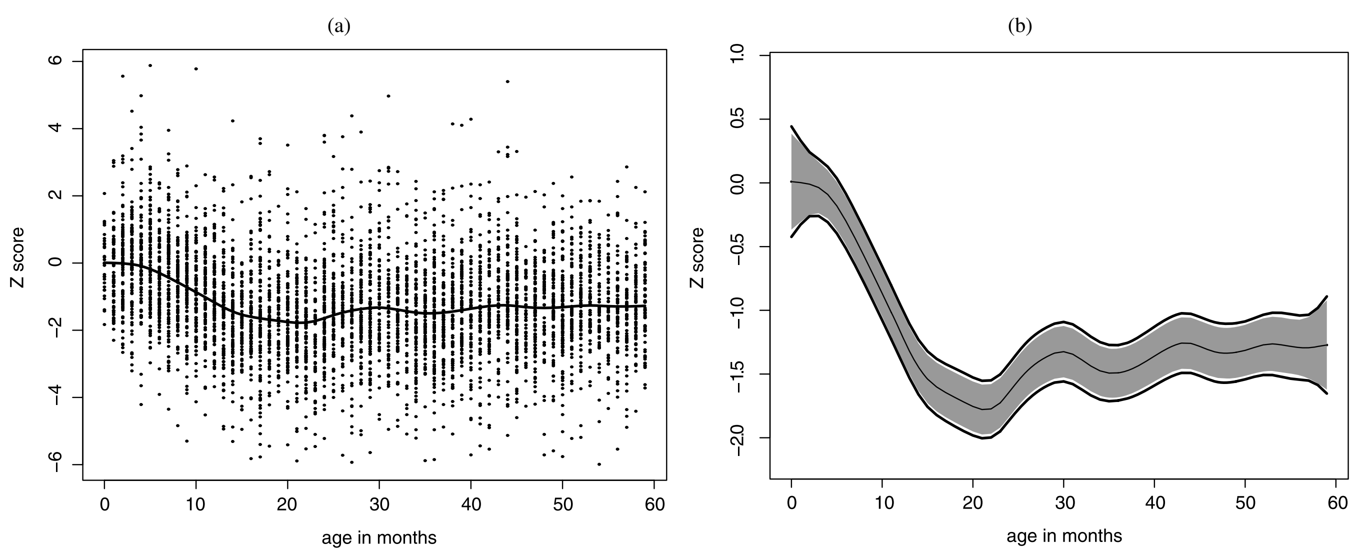

Figure 3. Z-scores for underweight of children in Kenya. (a) The data and a nonparametric fit based on 50 knots. (b) Simultaneous confidence bands based on the conditional mixed model (shaded area) and the marginal mixed model (bold lines).

the volume-of-tube formula have been developed to adjust for cases with heteroscedastic errors (Faraway and Sun 1995) and correlated errors (Sun, Raz, and Faraway 1999), whereas Sun, Loader, and McCormick (2000) considered generalized linear models. Extensions of our work to a generalized framework, as well as the handling of correlated and heteroscedastic data, offer interesting directions for further research.

APPENDIX: TECHNICAL DETAILS

A.1 Proofs

We adopt the framework of Claeskens, Krivobokova, and Opsomer (2009) and use the same assumptions.

(A1) Let $\delta = \max_{0\le j\le K}(\delta_j)$, $\delta_j = \tau_j - \tau_{j-1}$. There exists a constant $M > 0$, such that $\delta/\min_{1\le j\le K}(\delta_j) \le M$ and $\max_{0\le j\le K} |\delta_{j+1} - \delta_j| = o(K^{-1})$.

(A2) For deterministic design points $x_i \in [a, b]$, $i = 1, \ldots, n$, assume that there exists a distribution function Q with corresponding positive continuous design density ρ such that, with Q_n the empirical distribution of $x_1, \ldots, x_n$, $\sup_{x\in[a,b]} |Q_n(x) - Q(x)| = o(K^{-1})$.

(A3) $K_q = (K + p + 1 - q)(\lambda\check{c})^{1/2q} n^{-1/2q} > 1$ for some constant $\check{c}$ that depends only on q and the design density ρ and $K \sim \widetilde{C} n^{\nu/(2q+1)}$ for some constants $\widetilde{C}$ and $\nu > 1$.

Proof of Lemma 1. We denote $\sigma_\epsilon^2/\sigma_u^2 = \lambda_m$. Because $\hat{\lambda}_m$ is a maximum likelihood estimator, a routine calculation shows that

$$\hat{\lambda}_m \overset{\text{approx.}}{\sim} \mathrm{N}\left(\lambda_m, \frac{2\lambda_m^2}{\mathrm{tr}(\mathbf{S}^2) - p + o(1)}\right),$$

where $o(1) \to 0$ as $n \to \infty$, $\mathrm{tr}(\cdot)$ denotes the trace of the matrix and $\mathbf{S} = \mathbf{P}(\mathbf{P}^t\mathbf{P} + \lambda_m\mathbf{D})^{-1}\mathbf{P}^t$. We prove eq. (15) only; the proof of (14) is completely analogous. Using

$$\frac{\partial\|\mathbf{l}(x)\|^{-1}}{\partial\lambda} = \frac{\mathbf{l}^t(x)(\mathbf{I}_n - \mathbf{S})\mathbf{l}(x)}{\lambda\|\mathbf{l}(x)\|^3}, \qquad \frac{\partial\mathbf{l}(x)}{\partial\lambda} = \frac{(\mathbf{S} - \mathbf{I}_n)\mathbf{l}(x)}{\lambda},$$

and applying the delta method results in

$$\hat{\mathbf{l}}(x)^t\mathbf{f} \overset{\text{approx.}}{\sim} \mathrm{N}\left\{\mathbf{l}(x)^t\mathbf{f}, \mathrm{var}(\hat{\lambda}_m)\left(\frac{\partial\mathbf{l}(x)^t\mathbf{f}}{\partial\lambda_m}\right)^2\right\},$$

$$\|\hat{\mathbf{l}}(x)\|^{-1} \overset{\text{approx.}}{\sim} \mathrm{N}\left\{\|\mathbf{l}(x)\|^{-1}, \mathrm{var}(\hat{\lambda}_m)\left(\frac{\partial\|\mathbf{l}(x)\|^{-1}}{\partial\lambda_m}\right)^2\right\}.$$

With this, we find that

$$\mathrm{var}(\hat{\mathbf{l}}(x)^t\mathbf{f}) = \frac{[\{(\mathbf{I}_n - \mathbf{S})\mathbf{l}(x)\}^t\mathbf{f}]^2}{\mathrm{tr}(\mathbf{S}^2) - p + o(1)},$$

$$\mathrm{var}\left(\|\hat{\mathbf{l}}(x)\|^{-1}\right) = \frac{\{\mathbf{l}^t(x)(\mathbf{I}_n - \mathbf{S})\mathbf{l}(x)\}^2}{2\|\mathbf{l}(x)\|^6\{\mathrm{tr}(\mathbf{S}^2) - p + o(1)\}}.$$

To obtain the asymptotic orders, we use the results of Claeskens, Krivobokova, and Opsomer (2009). In particular, from their theorem 1, under assumptions (A1)–(A3), $n^{-1}\sum_{i=1}^n \mathrm{var}\{\tilde{f}(x_i)\} = n^{-1} \times \mathrm{tr}(\mathbf{S}^2) = O(n^{-2q/(2q+1)})$ for $K_q > 1$. Thus, $\mathrm{var}\{\tilde{f}(x)\} = \sigma_\epsilon^2\|\mathbf{l}(x)\|^2 = O(n^{-2q/(2q+1)})$ for any $x \in [a, b]$ and $\mathrm{tr}(\mathbf{S}^2) = O(n^{1/(2q+1)})$. With the arguments used in the proof of the asymptotic order of $\mathrm{tr}(\mathbf{S}^2)$ of Claeskens, Krivobokova, and Opsomer (2009), it is not difficult to see that $\mathrm{tr}(\mathbf{S})$ and $\mathrm{tr}(\mathbf{S}^3)$ have the same order $O(n^{1/(2q+1)})$, implying in particular that $\mathbf{l}(x)^t\mathbf{S}\mathbf{l}(x) = O(n^{-2q/(2q+1)})$. Noting that $\{(\mathbf{I}_n - \mathbf{S})\mathbf{l}(x)\}^t\mathbf{f} = \mathbf{l}^t(x)\{\mathbf{f} - E(\tilde{\mathbf{f}})\}$, we conclude that its asymptotic order is the same as that of the bias of $\tilde{\mathbf{f}}(x)$, that is, $O(n^{-q/(2q+1)})$. Thus we obtain $\hat{\mathbf{l}}(x)^t\mathbf{f} = \mathbf{l}(x)^t\mathbf{f} + O_p(n^{-1/2})$ and $\|\hat{\mathbf{l}}(x)\|^{-1} = \|\mathbf{l}(x)\|^{-1} + O_p(n^{(2q-1)/(4q+2)})$. Finally,

$$\begin{aligned}
&\frac{\hat{\mathbf{l}}(x)^t\mathbf{Y} - \mathbf{P}(x,\underline{\tau})^t\boldsymbol{\theta}}{\|\hat{\mathbf{l}}(x)\|}\\
&= \frac{\mathbf{l}(x)^t\mathbf{Y} - \mathbf{P}(x,\underline{\tau})^t\boldsymbol{\theta} + \{\hat{\mathbf{l}}(x) - \mathbf{l}(x)\}^t\{\mathbf{f} + O_p(1)\}}{\|\mathbf{l}(x)\|}\frac{\|\mathbf{l}(x)\|}{\|\hat{\mathbf{l}}(x)\|}\\
&= \left\{\frac{\mathbf{l}(x)^t\mathbf{Y} - \mathbf{P}(x,\underline{\tau})^t\boldsymbol{\theta}}{\|\mathbf{l}(x)\|} + O_p\left(n^{-1/(4q+2)}\right)\right\}\left\{1 + O_p\left(n^{-1/(4q+2)}\right)\right\},
\end{aligned}$$

proving the lemma.

Proof of Theorem 1. From (8) and (12), we conclude that

$$\begin{aligned}
&\frac{\kappa_{m,0}}{\pi}\exp(-c_m^2/2) - 2\Phi(c_m) + o\{\exp(-c_m^2/2)\}\\
&\qquad = \frac{\kappa_0}{\pi}\exp(-c^2/2) - 2\Phi(c) + o\{\exp(-c^2/2)\},
\end{aligned}$$

leading to $\exp(-c_m^2/2) = \exp(-c^2/2)\kappa_0\kappa_{m,0}{}^{-1}[1 + o\{\exp(-c^2/2)\}]$. Taking the logarithm from the both sides of the last equality and using the Taylor expansion of $\log(\kappa_{m,0})$ around $\log(\kappa_0)$, we find that

$$c_m^2 = c^2 + 2\frac{\kappa_{m,0} - \kappa_0}{\kappa_0} + o(1),$$

where $o(1)$ converges to 0 as $c \to \infty$. Note now that

$$\begin{aligned}
\mathrm{var}_{\mathbf{u}}\{m(x,\mathbf{u})\} &= \mathrm{var}_{\mathbf{u}}\{\mathbf{P}(x)(\mathbf{P}^t\mathbf{P} + \sigma_\epsilon^2/\sigma_u^2\mathbf{D})^{-1}\sigma_\epsilon^2/\sigma_u^2\mathbf{D}\boldsymbol{\theta}\}\\
&= \sigma_\epsilon^2\left(\|\mathbf{l}_m(x)\|^2 - \|\mathbf{l}(x)\|^2\right).
\end{aligned}$$

To obtain (20), it remains to apply (A4).

To prove Theorem 2, we need the following lemma.

Lemma 2. Under assumption (A1), it holds that

$$\begin{aligned}
\|\mathbf{l}_m'(x)\|^2\|\mathbf{l}_m(x)\|^{-2} &= c_1(x)K^2,\\
\{\mathbf{l}_m^t(x)\mathbf{l}_m'(x)\}^2\|\mathbf{l}_m(x)\|^{-4} &= \tilde{c}_1(x)K^2,\\
\|\mathbf{l}'(x)\|^2\|\mathbf{l}(x)\|^{-2} &= c_2(x)K^2,\\
\{\mathbf{l}^t(x)\mathbf{l}'(x)\}^2\|\mathbf{l}(x)\|^{-4} &= \tilde{c}_2(x)K^2
\end{aligned}$$

for some positive constants $c_1(x)$, $c_2(x)$, $\tilde{c}_1(x)$, $\tilde{c}_2(x)$ depending only on p and x.

Proof. Without loss of generality, we take p-degree (order $p+1$) B-splines as basis functions, so that for $x \in [\tau_i, \tau_{i+1})$, $i = 0, \ldots, K$, the basis vector takes the form $\mathbf{P}(x,\underline{\tau}) = \{\mathbf{0}_i, P_{i-p,p+1}(x), \ldots, P_{i,p+1}(x), \mathbf{0}_{K-i}\}$, with $P_{i,p+1}(x)$ denoting an ith B-spline of degree p evaluated at x and $\mathbf{0}_i$ as an i-dimensional vector of 0s. It is known (see, e.g., Zhou, Shen, and Wolfe 1998) that $\mathbf{P}'(x,\underline{\tau}) = p\mathbf{P}_p(x,\underline{\tau})\boldsymbol{\Delta}$, where $\mathbf{P}_p(x,\underline{\tau}) = \{\mathbf{0}_i, P_{i-p+1,p}(x), \ldots, P_{i,p}(x), \mathbf{0}_{K-i}\}$ and $\boldsymbol{\Delta}$ is a $(K+p) \times (K+p+1)$ matrix of weighted first-order differences, that is, a matrix with the rows $\{\mathbf{0}_j, -(\tau_{j+1+p} - \tau_{j+1})^{-1}, (\tau_{j+1+p} - \tau_{j+1})^{-1}, 0, \ldots\}$, $j = 0, \ldots, K+p-1$, each of length $K+p+1$. With this for $x \in [\tau_i, \tau_{i+1})$ we can rewrite $\|\mathbf{l}_m(x)\|^2 = \sum_{s,t=0}^p P_{i-s,p+1}(x)P_{i-t,p+1}(x)h_{st}$ and $\|\mathbf{l}(x)\|^2 = \sum_{s,t=0}^p P_{i-s,p+1}(x)P_{i-t,p+1}(x)\tilde{h}_{st}$, where $h_{st} = \{(\mathbf{P}^t\mathbf{P} + \lambda_m\mathbf{D})^{-1}\}_{i+1+s,i+1+t}$ and $\tilde{h}_{st} = \{(\mathbf{P}^t\mathbf{P} + \lambda_m\mathbf{D})^{-1}\mathbf{P}^t\mathbf{P}(\mathbf{P}^t\mathbf{P} + \lambda_m\mathbf{D})^{-1}\}_{i+1+s,i+1+t}$. Moreover,

$$\begin{aligned}
\|\mathbf{l}_m'(x)\|^2 = p^2\sum_{s,t=0}^p h_{st}&\left(\frac{P_{i-s,p}(x)}{\tau_{i+s+p} - \tau_{i+s}} - \frac{P_{i+1-s,p}(x)}{\tau_{i+1+s+p} - \tau_{i+1+s}}\right)\\
\times&\left(\frac{P_{i-t,p}(x)}{\tau_{i+t+p} - \tau_{i+t}} - \frac{P_{i+1+t,p}(x)}{\tau_{i+1+t+p} - \tau_{i+1+t}}\right),
\end{aligned}$$

$$\{\mathbf{l}_m^t(x)\mathbf{l}_m'(x)\}^2 = p^2\left\{\sum_{s,t=0}^{p} h_{st}\left(\frac{P_{i-s,p}(x)}{\tau_{i+s+p}-\tau_{i+s}} - \frac{P_{i+1-s,p}(x)}{\tau_{i+1+s+p}-\tau_{i+1+s}}\right) \times P_{i-t,p}(x)\right\}^2,$$

where $P_{i-p,p}(x) = P_{i+1,p}(x) = 0$. The analogous expressions hold for $\|\mathbf{l}'(x)\|$ and $\{\mathbf{l}^t(x)\mathbf{l}'(x)\}^2$, with h_{st} replaced by $\tilde{h}_{st}$. According to (A1), there exist positive constants $c_{j,k} < \infty$ independent of n and K such that $(\tau_{j+k+p} - \tau_{j+k}) = c_{j,k}^{-1}p/K$. Thus, because $0 \le P_{i,p+1}(x) \le 1$, and $0 \le P_{i,p}(x) \le 1$ for any i and only h_{st} depends on K and n, we find that

$$c_1(x) = \left(\sum_{s,t=0}^{p} h_{st}\{P_{i-s,p}(x)c_{i,s} - P_{i+1-s,p}(x)c_{i+1,s}\} \times \{P_{i-t,p}(x)c_{i,t} - P_{i+1-t,p}(x)c_{i+1,t}\}\right) \Big/ \sum_{s,t=0}^{p} h_{st}P_{i-s,p+1}(x)P_{i-t,p+1}(x)$$

is independent of K and n. Similar expressions can be obtained for $\tilde{c}_1(x)$, $c_2(x)$, and $\tilde{c}_2(x)$.

Proof of Theorem 2. The width of the confidence band based on the volume-of-tube formula for penalized splines at a fixed x is determined by the critical value c or c_m and the standard deviation $\sigma_\epsilon\|\mathbf{l}(x)\|$ or $\sigma_\epsilon\|\mathbf{l}_m(x)\|$. From (8) and (12), it follows that $c = \sqrt{\log[\kappa_0^2\{1+O(1)\}]}$ and $c_m = \sqrt{\log[\kappa_{m,0}^2\{1+O(1)\}]}$, where $O(1)$ is bounded for $c, c_m \to \infty$. As discussed in the proof of Lemma 1, the standard deviation $\sigma_\epsilon\|\mathbf{l}(x)\| = O(n^{-q/(2q+1)})$ and $\sigma_\epsilon\|\mathbf{l}_m(x)\| = O(n^{-q/(2q+1)})$. It remains to find the order of $\kappa_{m,0}$ and κ_0. By definition,

$$\kappa_{m,0} = \int_a^b \left\|\frac{d}{dx}\frac{\mathbf{l}_m(x)}{\|\mathbf{l}_m(x)\|}\right\| dx = \int_a^b \frac{\sqrt{\|\mathbf{l}_m(x)\|^2\|\mathbf{l}_m'(x)\|^2 - \{\mathbf{l}_m(x)^t\mathbf{l}_m'(x)\}^2}}{\|\mathbf{l}_m(x)\|^2}\,dx,$$

$$\kappa_0 = \int_a^b \left\|\frac{d}{dx}\frac{\mathbf{l}(x)}{\|\mathbf{l}(x)\|}\right\| dx = \int_a^b \frac{\sqrt{\|\mathbf{l}(x)\|^2\|\mathbf{l}'(x)\|^2 - \{\mathbf{l}(x)^t\mathbf{l}'(x)\}^2}}{\|\mathbf{l}(x)\|^2}\,dx.$$

Using Lemma 2, we find that $\kappa_{m,0} = O(K) = O(n^{\nu/(2q+1)})$ and $\kappa_0 = O(K) = O(n^{\nu/(2q+1)})$, and the width of the confidence band based on the volume-of-tube formula for penalized splines has asymptotic order $O_p(\sqrt{\log n^{2\nu/(2q+1)}}n^{-q/(2q+1)})$, $\nu > 1$.

[Received March 2009. Revised December 2009.]

REFERENCES

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66. [853,858]

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 1071–1095. [852]

Brezger, A., and Lang, S. (2006), "Generalized Additive Regression Based on Bayesian P-Splines," *Computational Statistics and Data Analysis*, 50, 967–991. [854]

Brumback, B. A., Ruppert, D., and Wand, M. P. (1999), Comment on "Variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based Prior," by T. S. Shively, R. Koher, and S. Wood, *Journal of the American Statistical Association*, 94, 794–797. [853]

Chaudhuri, P., and Marron, J. (1999), "SiZer for Exploration of Structures in Curves," *Journal of the American Statistical Association*, 94 (447), 807–823. [859]

Claeskens, G., and Van Keilegom, I. (2003), "Bootstrap Confidence Bands for Regression Curves and Their Derivatives," *The Annals of Statistics*, 31 (6), 1852–1884. [852]

Claeskens, G., Krivobokova, T., and Opsomer, J. (2009), "Asymptotic Properties of Penalized Spline Estimators," *Biometrika*, 96 (3), 529–544. [852, 855,861]

Cox, D. (1993), "An Analysis of Bayesian Inference for Nonparametric Regression," *The Annals of Statistics*, 21 (2), 903–923. [853,857,858]

Crainiceanu, C., Ruppert, D., Carroll, R., Adarsh, J., and Goodner, B. (2007), "Spatially Adaptive Penalized Splines With Heteroscedastic Errors," *Journal of Computational and Graphical Statistics*, 16, 265–288. [853-855]

Durban, M., and Currie, I. (2003), "A Note on P-Spline Smoothing With B-Splines and Penalties," *Computational Statistics*, 11 (2), 89–121. [853]

Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing With *B*-Splines and Penalties" (with discussion), *Statistical Science*, 11 (2), 89–121. [852]

Eubank, R. L., and Speckman, P. L. (1993), "Confidence Bands in Nonparametric Regression," *Journal of the American Statistical Association*, 88 (424), 1287–1301. [852,857,858]

Fahrmeir, L., Kneib, T., and Lang, S. (2004), "Penalized Structured Additive Regression for Space-Time Data: A Bayesian Perspective," *Statistica Sinica*, 14, 731–761. [853,854]

Faraway, J., and Sun, J. (1995), "Simultaneous Confidence Bands for Linear Regession With Heteroscedastic Errors," *Journal of the American Statistical Association*, 90, 1094–1098. [861]

Freedman, D. (1999), "On the Bernstein–Von Mises Theorem With Infinite-Dimensional Parameters," *The Annals of Statistics*, 27 (4), 1119–1140. [857]

Hall, P. (1991), "On Convergence Rates of Suprema," *Probability Theory and Related Fields*, 89 (4), 447–455. [852]

Hall, P., and Opsomer, J. (2005), "Theory for Penalized Spline Regression," *Biometrika*, 92, 105–118. [852]

Hall, P., and Titterington, D. (1988), "On Confidence Bands in Nonparametric Density Estimation and Regression," *Journal of Multivariate Analysis*, 27, 228–254. [852,857]

Härdle, W. (1989), "Asymptotic Maximal Deviation of *M*-Smoothers," *Journal of Multivariate Analysis*, 29 (2), 163–179. [852]

Held, L. (2004), "Simultaneous Posterior Probability Statements From Monte Carlo Output," *Journal of Computational and Graphical Statistics*, 13, 20–35. [853]

Hotelling, H. (1939), "Tubes and Spheres in *n*-Spaces, and a Class of Statistical Problems," *American Journal of Mathematics*, 61, 440–460. [852]

Johansen, S., and Johnstone, I. M. (1990), "Hotelling's Theorem on the Volume of Tubes: Some Illustrations in Simultaneous Inference and Data Analysis," *The Annals of Statistics*, 18 (2), 652–684. [855]

Kauermann, G., Krivobokova, T., and Fahrmeir, L. (2009), "Some Asymptotic Results on Generalized Penalized Spline Smoothing," *Journal of the Royal Statistical Society, Ser. B*, 71 (2), 487–503. [852]

Knafl, G., Sacks, J., and Ylvisaker, D. (1985), "Confidence Bands for Regression Functions," *Journal of the American Statistical Association*, 80 (391), 683–691. [852]

Li, Y., and Ruppert, D. (2008), "On the Asymptotics of Penalized Splines," *Biometrika*, 95 (2), 415–436. [852]

Loader, C., and Sun, J. (1997), "Robustness of Tube Formula Based Confidence Bands," *Journal of Computational and Graphical Statistics*, 6 (2), 242–250. [860]

Naiman, D. (1986), "Conservative Confidence Bands in Curvilinear Regression," *The Annals Statistics*, 14, 896–906. [855]

Neumann, M. H., and Polzehl, J. (1998), "Simultaneous Bootstrap Confidence Bands in Nonparametric Regression," *Journal of Nonparametric Statistics*, 9 (4), 307–333. [852]

O'Sullivan, F. (1986), "A Statistical Perspective on Ill-Posed Inverse Problems" (with discussion), *Statistical Science*, 1, 505–527. [852]

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge, U.K.: Cambridge University Press. [852,856,859,860]

Sun, J. (1993), "Tail Probabilities of the Maxima of Gaussian Random Fields," *The Annals of Probability*, 21 (1), 34–71. [852,855]

Sun, J., and Loader, C. R. (1994), "Simultaneous Confidence Bands for Linear Regression and Smoothing," *The Annals of Statistics*, 22 (3), 1328–1345. [852,855,857]

Sun, J., Loader, C., and McCormick, W. (2000), "Confidence Bands in Generalized Linear Models," *The Annals Statistics*, 28 (2), 429–460. [861]

Sun, J., Raz, J., and Faraway, J. J. (1999), "Confidence Bands for Growth and Response Curves," *Statistica Sinica*, 9 (3), 679–698. [855,861]

Wahba, G. (1983), "Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline," *Journal of the Royal Statistical Society, Ser. B*, 45 (1), 133–150. [857,858]

Weyl, H. (1939), "On the Volume of Tubes," *American Journal of Mathematics*, 61, 461–472. [852,855]

Xia, Y. (1998), "Bias-Corrected Confidence Bands in Nonparametric Regression," *Journal of the Royal Statistical Society, Ser. B*, 60 (4), 797–811. [852,858]

Zhou, S., Shen, X., and Wolfe, D. A. (1998), "Local Asymptotics for Regression Splines and Confidence Regions," *The Annals of Statistics*, 26 (5), 1760–1782. [852,861]