Comparison of Parametric and Bootstrap Approaches to Obtaining Confidence Intervals for Logspline Density Estimation

Charles KOOPERBERG and Charles J. STONE

In earlier articles, we developed an automated methodology for using cubic splines with tail linear constraints to model the logarithm of a univariate density function. This methodology was subsequently modified so that the knots were determined by stepwise addition-deletion and the remaining coefficients were determined by maximum likelihood estimation. An alternative approach, referred to as the free knot spline procedure, is to use the maximum likelihood method to estimate the knot locations as well as the remaining coefficients. This article compares various approaches to constructing confidence intervals for logspline density estimates, for both the stepwise procedure and the free knot procedure. It is concluded that a variation of the bootstrap, in which only a limited number of bootstrap simulations are used to estimate standard errors that are combined with standard normal quantiles, seems to perform the best, especially when coverages and computing time are both taken into account.

Key Words: BIC; Cubic splines; Free knots; Maximum likelihood estimation; Stepwise knot selection; Tail linear constraints.

1. INTRODUCTION

Polynomial splines have successfully been used to model unknown functions. The typical polynomial spline methodology employs a stepwise addition-deletion procedure. That is, initially a minimal model is fit to the data, after which additional basis functions are added to the model. In one-dimensional problems these additional basis functions typically involve new knots for a spline function, while in higher dimensional problems they may involve knots in one of the components or tensor products of lower dimensional splines. Which basis function is added next is usually decided by maximizing the Rao statistic,

Charles Kooperberg is Member, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024 (E-mail: clk@fhcrc.org). Charles J. Stone is Professor, Department of Statistics, University of California, Berkeley, CA 94720-3860 (E-mail: stone@stat.berkeley.edu).

Journal of Computational and Graphical Statistics, Volume 13, Number 1, Pages 106–122
DOI: 10.1198/1061860043047

which is based on a quadratic approximation to the log-likelihood. After a maximal model is reached, basis functions are deleted one at a time. Here typically the Wald statistic, also based on a quadratic approximation to the log-likelihood, is used to decide which basis function to remove next. Polynomial spline methodologies include MARS (Friedman 1991), Logspline (Kooperberg and Stone 1991, 1992), Hare (Kooperberg, Stone, and Truong 1995), and Polyclass (Kooperberg, Bose, and Stone 1997). Stone, Hansen, Kooperberg, and Truong (1997) contains an overview of polynomial spline methods.

Getting reliable confidence intervals corresponding to function estimates that are obtained using such stepwise procedures is challenging. After model selection has been carried out, the estimated function has a simple parametric form. However, treating the final model as a fixed parametric model, ignoring the large amount of model selection that may have occurred, yields confidence intervals with too low coverage.

At least for one-dimensional problems, such as logspline density estimation, this problem can potentially be circumvented by using free knot splines: polynomial splines in which the knot locations are treated as additional parameters to be estimated along with the other parameters. Although intuitively appealing, free knot splines have gotten relatively little attention in statistics because of both the numerical problems associated with implementing them and the mathematical difficulty in investigating their theoretical properties.

Recently, Stone and Huang (2002) successfully investigated the theoretical properties (rates of convergence) of statistical modeling with free knot splines and their tensor products in the general context of extended linear modeling, which includes not only density estimation but also regression, logistic regression, and hazard regression as special cases. These theoretical results add motivation for further investigations of the methodological aspects of statistical modeling with free knot splines.

Lindstrom (1999) developed a computationally feasible algorithm for fitting free knot splines in the context of univariate regression. In this algorithm she added a small penalty term to prevent knots from coalescing. She did not investigate the effect of using free knot splines for obtaining confidence intervals.

This article compares confidence intervals for logspline density estimates using a free knot spline procedure to that using a stepwise addition-deletion procedure. Computationally, the free knot spline procedure is considerably more time-consuming than the stepwise procedure for logspline density estimation. We believe that it is conceptually useful to think of the stepwise procedure as being a computational shortcut for approximating the free knot spline estimate. A proper understanding of properties of estimates and related confidence intervals based on free knot splines should provide us with insight about the stepwise procedure.

An alternative approach to accessing uncertainty caused by adaptive model selection is the Bayesian approach (Smith and Kohn 1996; Denison, Smith, and Mallick 1998; DiMatteo, Genovese, and Kass 2001; Hansen and Kooperberg 2002), in which Bayesian inference is carried out by putting priors on knot positions, number of knots, and perhaps parameters. Interestingly, while the Bayesian procedure supposedly gives an automatic mechanism for constructing credible (Bayesian confidence) intervals, most of these articles do not provide such intervals, and none study their properties. Because this article is primarily a comparison of free knot and stepwise procedures, we will discuss Bayesian methods only briefly in Sections 3.3 and 5.

Section 2 briefly reviews logspline density estimation in general and the free knot spline procedure in particular. Section 3 discusses the various approaches to obtaining confidence intervals. These approaches are compared by means of a simulation study in Section 4. Section 5 applies the various approaches to a real example. We end with a brief discussion.

2. BASIC METHODOLOGY

Consider an unknown density function f on a compact interval $[L, U]$, where $-\infty < L < U < \infty$. We will estimate $\eta = \log f$ by a cubic spline $\widehat{\eta}$ on $[L, U]$ having $J \geq 2$ knots $\gamma_1, \ldots, \gamma_J$ with $L < \gamma_1 < \cdots < \gamma_J < U$. To avoid unacceptably high variance of $\widehat{\eta}$ near L and U, we impose the tail linear constraints $\widehat{\eta}''(L) = 0$ and $\widehat{\eta}''(U) = 0$. We also require that $\widehat{f} = \exp(\widehat{\eta})$ be a density function; that is, that $\int_L^U \widehat{f}(y)\, dy = 1$. To this end, we write $\widehat{\eta} = \widehat{g} - \widehat{C}$, where $\widehat{g}$ satisfies the tail linear constraints and $\widehat{C}$ is the normalizing constant.

Given the knot sequence $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)$ let $\mathbb{G}_{\boldsymbol{\gamma}}$ denote the space of cubic splines on $[L, U]$ corresponding to $\boldsymbol{\gamma}$ and satisfying the tail linear constraints. Thus, a function g on $[L, U]$ is a member of $\mathbb{G}$ if and only if g is twice continuously differentiable on $[L, U]$, the restriction of g to each of the intervals $[L, \gamma_1]$, $[\gamma_1, \gamma_2]$, $\ldots$, $[\gamma_{J-1}, \gamma_J]$, $[\gamma_J, U]$ is a cubic polynomial, $g''(L) = 0$, and $g''(U) = 0$. Observe that $\mathbb{G}_{\boldsymbol{\gamma}}$ is a $(J+2)$-dimensional linear space. Set $p = J+1$, and let $1, B_{\boldsymbol{\gamma}1}, \ldots, B_{\boldsymbol{\gamma}p}$ be a basis of $\mathbb{G}_{\boldsymbol{\gamma}}$. Given $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, consider the corresponding candidate $\eta_{\boldsymbol{\gamma}}(\theta)$ for $\widehat{\eta}$ given by

$$\eta_{\boldsymbol{\gamma}}(y; \boldsymbol{\theta}) = \theta_1 B_{\boldsymbol{\gamma}1}(y) + \cdots + \theta_p B_{\boldsymbol{\gamma}p}(y) - C_{\boldsymbol{\gamma}}(\theta), \qquad L \leq y \leq U,$$

where

$$C_{\boldsymbol{\gamma}}(\theta) = \log\left(\int_L^U \exp(\theta_1 B_{\boldsymbol{\gamma}1}(y) + \cdots + \theta_p B_{\boldsymbol{\gamma}p}(y))\, dy\right).$$

Note that since $[L, U]$ is a compact interval $\exp \eta_{\boldsymbol{\gamma}}(y; \boldsymbol{\theta})$ is a positive density function on $[L, U]$ for every $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$.

Let $Y_1, \ldots, Y_n$ be a random sample of size n from a distribution having density f and log-density $\eta = \log f$. The corresponding log-likelihood is given by

$$\ell_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = \sum_{i=1}^n \eta_{\boldsymbol{\gamma}}(Y_i; \boldsymbol{\theta}) = \sum_{j=1}^p \theta_p \sum_{i=1}^n B_{\boldsymbol{\gamma}j}(Y_i) - nC_{\boldsymbol{\gamma}}(\boldsymbol{\theta}).$$

2.1 LOGSPLINE DENSITY ESTIMATION WITH STEPWISE KNOT SELECTION

For the stepwise addition-deletion procedure (Stone et al. 1997), let $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}^s$ denote the maximum likelihood estimate of $\boldsymbol{\theta}$ and set

$$\widehat{\ell}^s = \ell_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}^s\right) = \arg\max_{\boldsymbol{\theta}} \ell_{\boldsymbol{\gamma}}(\boldsymbol{\theta}).$$

(The superscript s is used to refer to estimates obtained by using stepwise addition-deletion of knots; the superscript f is used below to refer to estimates using free knots.) In this context we need to select both the number J of knots and the vector $\boldsymbol{\gamma}$ of knot locations. This selection is carried out by means of stepwise addition-deletion: after positioning a few initial knots at selected order statistics we add knots one at a time to increase the log-likelihood as much as possible until a maximum number of knots is reached; then we carry out a stepwise deletion procedure. Let $\widehat{\boldsymbol{\theta}}^s_J$ and $\widehat{\ell}^s_J$ denote the parameter estimates and the log-likelihood, using the stepwise addition-deletion procedure with J knots. To select J, we will employ the (generalized) Akaike information criterion $\text{AIC}_{J,a} = -2\widehat{\ell}^f_J + d_J a$, where d_J is the number of free parameters in a model with J basis functions and a is a complexity parameter. For the stepwise addition-deletion procedure we ignore the knot selection, so $\widehat{\boldsymbol{\theta}}^s_J$ has $p = d_J = J + 1$ free parameters. We use the AIC criterion with $a = \log n$ to select the number J of knots. Let $\widehat{\boldsymbol{\theta}}^s$, $\widehat{\eta}^s$, and $\widehat{f}^s$ denote the parameter estimates and the logspline estimates of the log-density and density function, respectively, using the stepwise addition-deletion procedure.

Kooperberg and Stone (1992) described an alternative stepwise procedure using only stepwise deletion of knots. In this procedure more initial knots are positioned at selected order statistics of the data and the "least significant" knots are deleted one at a time. In this article we will only show results using the stepwise addition-deletion procedure. We did, however, carry out all simulations for the stepwise deletion procedure as well, and the corresponding results are briefly summarized in Section 4.1.

2.2 LOGSPLINE DENSITY ESTIMATION WITH FREE KNOTS

Let $\left(\widehat{\boldsymbol{\gamma}}^f, \widehat{\boldsymbol{\theta}}^f\right)$ denote the maximum likelihood estimate of $(\boldsymbol{\gamma}, \boldsymbol{\theta})$, so that

$$\widehat{\ell}^f = \ell_{\widehat{\boldsymbol{\gamma}}^f}\left(\widehat{\boldsymbol{\theta}}^f\right) = \arg\max_{\boldsymbol{\gamma},\boldsymbol{\theta}} \ell_{\boldsymbol{\gamma}}(\boldsymbol{\theta}).$$

Computing maximum likelihood estimates with free knots is a highly nontrivial numerical problem, as the likelihood function $\ell_{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ is severely multimodal, and degenerate solutions exist when too many of the knots γ_j get close together.

Observe that the positive integer parameter J must also be chosen. Let $\widehat{\boldsymbol{\gamma}}^f_J$, $\widehat{\boldsymbol{\theta}}^f_J$, and $\widehat{\ell}^f_J$ now indicate the dependence of $\widehat{\boldsymbol{\gamma}}^f$, $\widehat{\boldsymbol{\theta}}^f$, and $\widehat{\ell}^f$, respectively, on J. For the free knot procedure $\widehat{\boldsymbol{\gamma}}^f_J$ has J free parameters and $\widehat{\boldsymbol{\theta}}^f_J$ has $p = J+1$ free parameters, so $d_J = 2J+1$. We select the value $\widehat{J}^f$ of J that minimizes $\text{AIC}_{J,2}$. Set $\widehat{\boldsymbol{\gamma}}^f = \widehat{\boldsymbol{\gamma}}^f_{\widehat{J}^f}$, $\widehat{\boldsymbol{\theta}}^f = \widehat{\boldsymbol{\theta}}^f_{\widehat{J}^f}$, and $\widehat{\eta}^f(y) = \eta_{\widehat{\boldsymbol{\gamma}}^f}\left(y; \widehat{\boldsymbol{\theta}}^f\right)$. We refer to $\widehat{f}^f(y) = \exp \widehat{\eta}^f(y)$ as the logspline estimate with free knots of the density f at y.

For logspline density estimation with free knots we choose $a = 2$ (AIC) as the complexity parameter. When free knots are used we have found $a = \log n$, which is used for the stepwise addition-deletion procedure, to be too large for two reasons. First, a large parameter promotes smaller models, with lower variance and somewhat larger bias. For exploratory data analysis this is quite desirable; however, for confidence intervals the coverages will be too low if estimates are overly biased. A second reason is that in the free knot spline approach all knots are parameters. Thus, a model with J knots now has $d_J = 2J + 1$ parameters and gets a penalty of $(2J + 1)a$, while for the stepwise approach such a model had $d_J = J + 1$ parameters and got a penalty of $(J + 1)a$. In our experience, however, the increase in log-likelihood that is achieved by adding a basis function with a fixed knot is considerably larger than the increase in log-likelihood that is achieved by making a fixed knot a free knot.

3. CONFIDENCE INTERVALS

3.1 Free Knot Splines

Let $\widehat{\nabla}^f_J \eta(y)$ denote the $(2J + 1)$-dimensional gradient of $\eta_{\boldsymbol{\gamma}}(y;\boldsymbol{\theta})$ at the maximum likelihood estimate $\left(\widehat{\boldsymbol{\gamma}}^f, \widehat{\boldsymbol{\theta}}^f\right)$, and let $\widehat{H}^f_J$ denote the corresponding Hessian when there are J free knots. Set $\widehat{\nabla}^f \eta(y) = \widehat{\nabla}^f_{\widehat{J}^f} \eta(y)$ and $\widehat{H}^f = \widehat{H}^f_{\widehat{J}^f}$. The standard error in the estimate $\widehat{\eta}(y)$ is given by

$$\mathrm{SE}^f(\widehat{\eta}^f(y)) = \sqrt{\left[\widehat{\nabla}^f \eta(y)\right]^T \left(-\widehat{H}^f\right)^{-1} \widehat{\nabla}^f \eta(y)}. \tag{3.1}$$

(To compute the various derivatives in this formula, Theorems 2.51, 2.55, and 4.27 of Schumaker (1991) are employed.) This leads to the nominal 95% confidence interval

$$\left(\widehat{\eta}^f(y) - 1.96\,\mathrm{SE}^f(\widehat{\eta}^f(y)),\ \widehat{\eta}^f(y) + 1.96\,\mathrm{SE}^f(\widehat{\eta}^f(y))\right)$$

for $\eta(y)$ and the corresponding nominal 95% confidence interval

$$\left(\exp\left(\widehat{\eta}^f(y) - 1.96\,\mathrm{SE}^f(\widehat{\eta}^f(y))\right),\ \exp\left(\widehat{\eta}^f(y) + 1.96\,\mathrm{SE}^f(\widehat{\eta}^f(y))\right)\right) \tag{3.2}$$

for $f(y)$.

3.2 Stepwise Addition-Deletion

For the stepwise addition-deletion procedures we can construct confidence intervals similar to those for the free knot procedure, except that the knots are now considered fixed. Thus, $\widehat{\nabla}^s_J \eta(y)$ is the $(J + 1)$-dimensional gradient of $\eta_{\boldsymbol{\gamma}}(y;\boldsymbol{\theta})$ at the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}^s$ and $\widehat{H}^s_J$ is the corresponding Hessian when there are J knots selected using the stepwise procedure. These quantities can now be used to construct confidence intervals using (3.1) and (3.2).

Alternatively, we can employ the usual bootstrap procedure to obtain confidence intervals corresponding to logspline density estimates. In this article we examine the coverage of bootstrap percentile intervals (Efron and Tibshirani 1993) for the log-density function.

That is, we take B (we used $B = 1{,}000$) samples $\mathbf{Y^i}$ with replacement of size n from the data $Y_1, \ldots, Y_n$, and for each sample $\mathbf{Y^i}$ we obtain the logspline density estimate. The 95% pointwise confidence interval for $\widehat{\eta}(y)$ is then from the 2.5th to the 97.5th percentile of the B bootstrap estimates for the log-density.

Clearly, the bootstrap is a computationally time-consuming procedure for getting confidence intervals, as we need to fit B logspline densities. However, it is still slightly faster than a procedure for fitting logspline densities with free knots.

A considerably cheaper approach is to hope that the logspline estimates of the log-density function have approximately a normal distribution, but that the estimates of the standard errors that are obtained using standard techniques are too small. If so, we can get by with a much smaller number B of bootstrap estimates (say $B = 25$) by using these estimates to obtain bootstrap estimates of $\mathrm{SE}(\widehat{\eta}(y))$ and then using Equation (3.2) or the equivalent to obtain confidence intervals for η.

3.3 A BAYESIAN APPROACH

Hansen and Kooperberg (2002) described a Bayesian approach to logspline density estimation, which involves a prior $p(J)$ on the dimension of the model, a prior $p(\beta \mid J)$ on the location of the knots, and a prior $P(\boldsymbol{\theta} \mid \beta, J)$ on the coefficients. They discussed several Bayesian versions for logspline density estimation, differing somewhat in the choice of priors and hyperparameters. Hansen and Kooperberg (2002) pointed out that, depending on how priors are selected, a Bayesian procedure can be similar in performance to a greedy stepwise procedure using AIC to select the number of knots when a geometric prior on the number of knots is used, or it can be similar to a smoothing spline approach when a uniform prior on the number of knots and a particular multivariate normal prior on the coefficients are used.

Given the data $Y_1, \ldots, Y_n$, the posterior distribution of $(J, \beta, \boldsymbol{\theta})$ is explored using a reversible jump Markov chain Monte Carlo (Green 1995) algorithm. To make (pointwise) 95% credible (Bayesian confidence) intervals about the logspline density estimate obtained from this Bayesian procedure, the 2.5th and 97.5th percentiles of all Markov chain Monte Carlo simulations are used. The algorithm of Hansen and Kooperberg (2002) for logspline density estimation is similar to algorithms for univariate regression using polynomial splines proposed by Denison, Mallick and Smith (1998) and Smith and Kohn (1996).

Because there are many more options and parameters that need to be selected for the Bayesian procedure, we opted not to include this procedure in the comparison in the next section, but we do include it in Section 5.

4. SIMULATED EXAMPLES

This section applies the various approaches to obtaining confidence intervals for logspline density estimation to data that were simulated from four distributions:

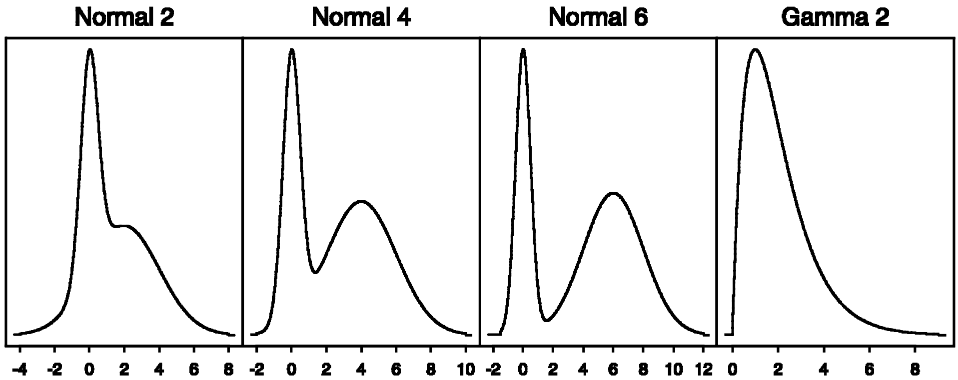

Figure 1. *The four densities used in the simulation study.*

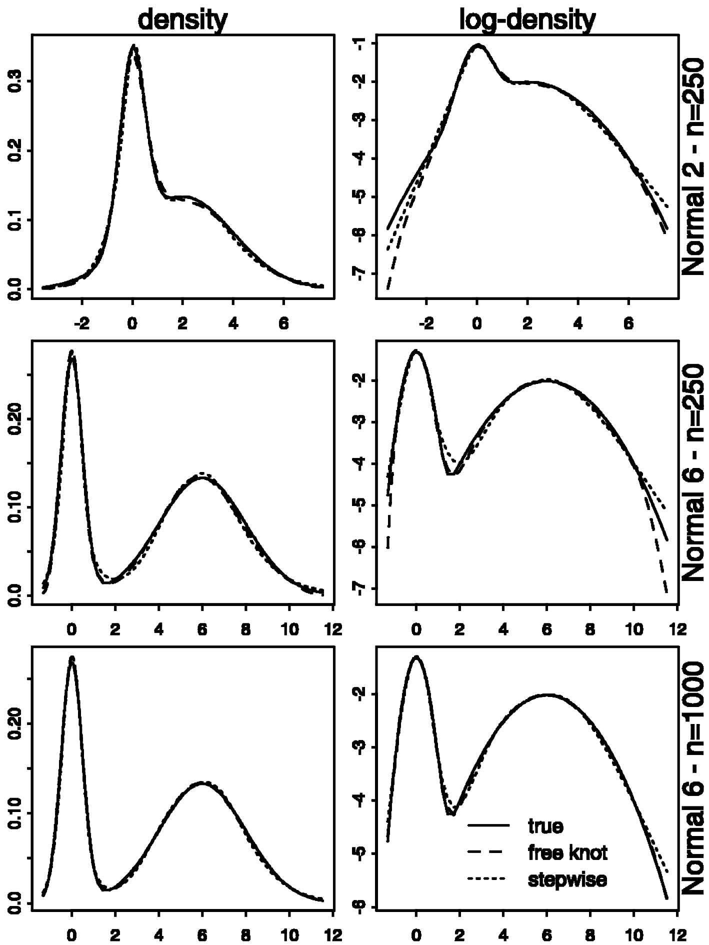

Figure 2. *Bias for three of the situations in the simulation study.*

Normal 2. A mixture of two normal distributions, so that the true density of Y is given by

$$f(y) = c\left(\frac{1}{3}f_{Z_1}(y) + \frac{2}{3}f_{Z_2}(y)\right)\text{ind}(-4, 8),$$

where Z_1 has a normal distribution with mean 0 and standard deviation 0.5, Z_2 has a normal distribution with mean 2 and standard deviation 2, ind$(\cdot)$ is the usual indicator function, and c is the normalizer to correct for the truncation to $(-4, 8)$.

Normal 4. As in example 1, but the mean of Z_2 is 4 and Y is truncated to $(-2, 10)$.

Normal 6. As in example 1, but the mean of Z_2 is 6 and Y is truncated to $(-1.5, 12)$.

Gamma 2. A gamma distribution with shape parameter 2 and mean 1, with Y truncated to the interval $(0, 9)$.

Figure 1 shows the four true densities. The Normal 2 density has one mode, but a clear second hump; Normal 4 has two, not very well separated, modes; Normal 6 has two well separated modes; and the Gamma 2 density is unimodal. For each of the four densities we generated $N = 250$ samples of size 250 and also N samples of size 1,000. For a particular distribution, let $\widehat{\eta}_i^u$ and $\widehat{f}_i^u$, $u \in \{f, s\}$ be the logspline density estimates for the log-density and density function, using the two different knot selection schemes described in Sections 2.1 and 2.2; set $\overline{\eta}^u(y) = N^{-1}\sum_i \widehat{\eta}_i^u(y)$, which we think of as a Monte Carlo estimate of $E(\widehat{\eta}^u(y))$; and set $\overline{f}^u(y) = \exp\overline{\eta}^u(y)$.

Figure 2 compares $f(y)$ (solid), $\overline{f}^f(y)$ (dashed), and $\overline{f}^s(y)$ (dotted), as well as $\eta(y)$, $\overline{\eta}^f(y)$, and $\overline{\eta}^s(y)$. Especially from the dip between the modes for the two Normal 6 examples we note that the free knot spline procedure appears less biased than the stepwise procedure. The Normal 6 density yields (for both approaches) the most biased results, while the results for the other three densities are fairly comparable. The bias in the extreme tails is partly due to the log-density estimates in the tails being restricted to be linear, while the true densities are approximately quadratic. Removing the tail-linear constraints reduces the bias but at the expense of a substantially increased variance. Since the confidence intervals, which except for the bootstrap percentile approach are all centered on the estimate, do not correct for this bias, we expect that the tail-linear constraints may be responsible for the coverages being somewhat too low in the extreme tails and, conceivably, also in the region between the modes for the two Normal 6 examples.

Table 1 shows the mean integrated squared bias for the log-density for both approaches, as computed over the 249 quantiles corresponding to the probabilities $i/250$, $i = 1, \ldots 249$ of the true density. From this table we see that the free knot procedure performs considerably better than the stepwise procedure for $n = 1{,}000$, but the results are similar for $n = 250$: for the Gamma 2 density, the least complicated density in our simulation, with $n = 250$ the stepwise procedure actually had less bias.

Let $\text{SD}^u(\widehat{\eta}^u(y)) = \sqrt{\text{var}\{\widehat{\eta}^u(y)\}}$ denote the pointwise sample standard deviation of the $N = 250$ estimates of $\widehat{\eta}^u$. Let $\text{SE}^u(\widehat{\eta}^u(y))$ be the standard errors as defined in Sections 3.1 and 3.2. In addition, let $\text{SEFX}^f(\widehat{\eta}^f(y))$ be the standard errors assuming the knots of

Table 1. Mean Integrated Squared Bias for Two Logspline Procedures

Density	*Free knot*	*Stepwise*
	n = 250	
Normal 2	0.0054	0.0052
Normal 4	0.0019	0.0052
Normal 6	0.0048	0.0096
Gamma 2	0.0099	0.0052
	n = 1,000	
Normal 2	0.0006	0.0029
Normal 4	0.0003	0.0019
Normal 6	0.0008	0.0031
Gamma 2	0.0002	0.0020

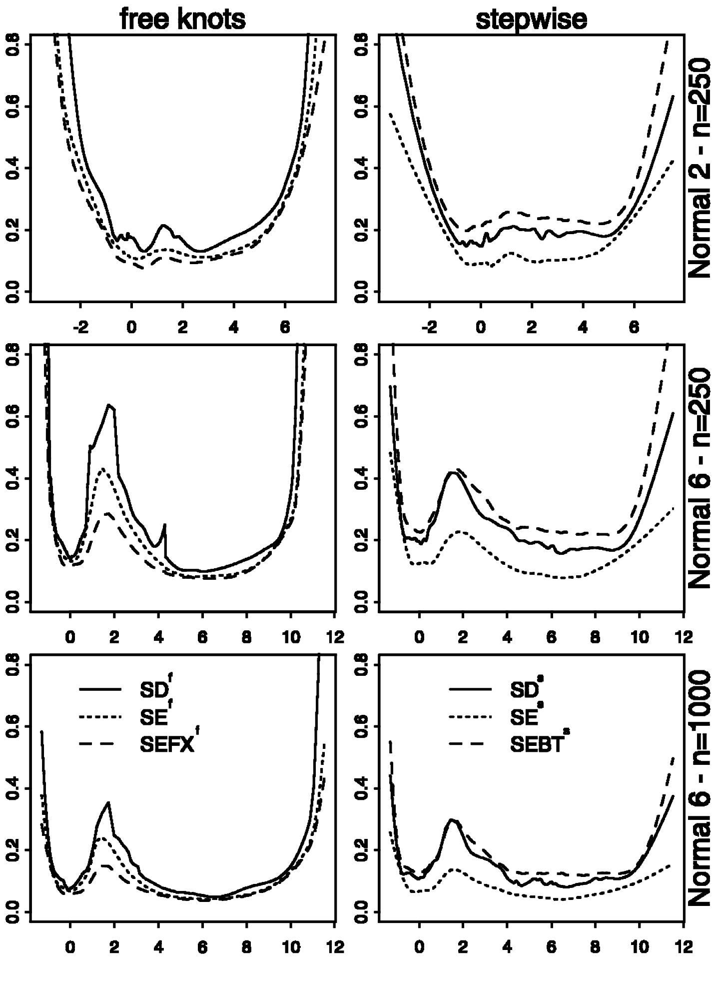

Figure 3. Estimates of the standard deviation and standard errors of the log-density for three of the situations in the simulation study.

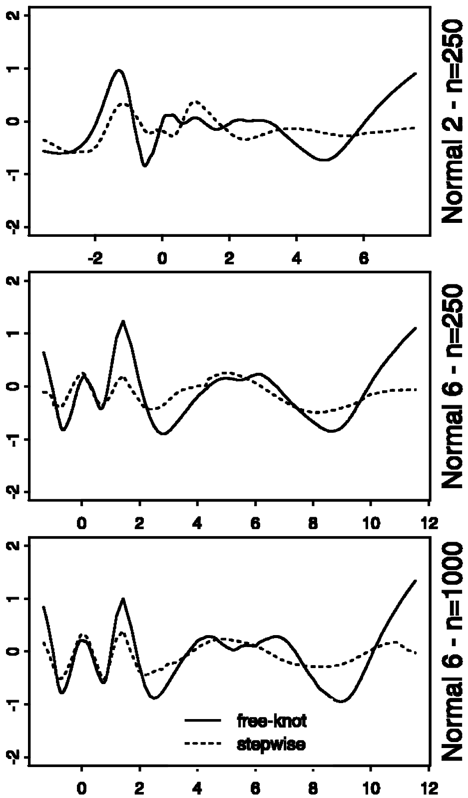

Figure 4. Estimates of the normalized bias of the log-density for three of the situations in the simulation study.

the free knot procedure are fixed (so that they make use only of the $(\widehat{J}^f + 1) \times (\widehat{J}^f + 1)$ Hessian matrix for the coefficients). Note that SE^s also assumes that the knots are fixed. Let $\mathrm{SEBT}^s(\widehat{\eta}^s(y))$ be the standard error of $\widehat{\eta}^s(y)$, estimated as the standard deviation from the log-density estimates of 25 bootstrap samples. Let $\overline{\mathrm{SE}}^u(\widehat{\eta}^u(y))$, $\overline{\mathrm{SEFX}}^f(\widehat{\eta}^f(y))$, and $\overline{\mathrm{SEBT}}^s(\widehat{\eta}^s(y))$ be the average of $\mathrm{SE}^u(\widehat{\eta}^u(y))$, $\mathrm{SEFX}^f(\widehat{\eta}^f(y))$, and $\mathrm{SEBT}^s(\widehat{\eta}^s(y))$, respectively, over the $N = 250$ simulations.

Figure 3 examines the various estimates of the standard error for three of the situations in the simulation study, for both the free knot spline and the stepwise addition-deletion procedure. From the plots on the left side of this figure we note that the $\overline{\mathrm{SEFX}}^f$ and even $\overline{\mathrm{SE}}^f$ consistently underestimate SD^f. This suggests that using free knot splines may improve the coverages of confidence intervals, but the intervals using the free knot spline procedure will probably still have too low coverages. We will see below that this is indeed the case. For the stepwise procedure we note that $\overline{\mathrm{SE}}^s$ is much smaller than $\overline{\mathrm{SD}}^s$, but $\overline{\mathrm{SEBT}}^s$ is actually slightly larger than $\overline{\mathrm{SD}}^s$, suggesting that the bootstrapping introduced some extra variability.

Set

$$\rho^u(y) = \frac{\overline{\eta}^u(y) - \eta(y)}{\mathrm{SD}^u(\widehat{\eta}^u(y))}, \qquad u \in \{f, s\}.$$

Table 2. Percent Coverages for Five Different Approaches to Obtaining Confidence Intervals for a Log-Density, Estimated Using Logspline

Procedure			*Stepwise addition-deletion*		
method	*Free knot*			*bootstrap*	
density	*SE^f*	*SEFX^f*	*SE^s*	*percentiles*	*SEBT^s*
			n = 250		
Normal 2	84.0	77.4	72.2	98.1	96.5
Normal 4	88.8	82.5	73.0	98.0	96.4
Normal 6	89.0	84.0	70.8	97.6	95.8
Gamma 2	86.2	81.2	58.6	98.2	97.4
			n = 1,000		
Normal 2	89.2	79.6	67.6	97.7	94.9
Normal 4	89.3	82.7	68.8	98.8	95.8
Normal 6	86.2	81.4	69.8	98.2	95.4
Gamma 2	84.0	77.3	57.9	98.2	95.6
Average	87.1	80.7	67.3	98.1	96.0

We interpret ρ^u as a *normalized bias*. Observe that ρ^u provides a rough indication of how much the actual coverage probabilities of corresponding nominal 95% confidence intervals would fall below 0.95. To see this, note that if W is normally distributed with standard deviation σ and nominal mean μ but actual mean μ^*, then $P(W - 1.96\sigma < \mu < W + 1.96\sigma)$ equals 0.943, 0.921, 0.830 according as $(\mu - \mu^*)/\sigma$ equals ± 0.25, ± 0.5, ± 1.0. Figure 4 shows the normalized bias of both logspline procedures for three of the simulation set-ups. Interestingly, this plot suggests that the normalized bias for the stepwise procedure is slightly smaller than that for the free knot procedure. Not surprisingly, the dip between the two peaks for the Normal 6 distribution is the region in which all procedures are most biased.

Table 2 shows the percent coverages of nominal pointwise 95% confidence intervals for the density estimates, averaged over the 249 quantiles corresponding to the probabilities $i/250$, $i = 1, \ldots 249$ of the true density. The numbers in the "coverage" columns are the averages of these numbers for the N realizations. (The standard errors of these averages are around 1%.) We show results for all the standard errors discussed above, as well as the bootstrap percentile approach, discussed in Section 3.2. Observe from this table that while the free knot standard errors (SE^f) yield much better coverages than the standard errors that keep the knots fixed (SEFX^f and SE^s), they still give coverages that are considerably too low. The differences in coverages between the intervals that keep the knots fixed and those using free knot standard errors suggest that an important reason for the low coverages of the former is that they assume that the knots are fixed.

Both bootstrap intervals provide coverages that are either accurate or slightly too conservative. Surprisingly, the coverages for the bootstrap percentile interval procedure are consistently too high. It is our impression that this is due to some instability in the stepwise logspline algorithm when there are many repeat observations, causing the intervals to be

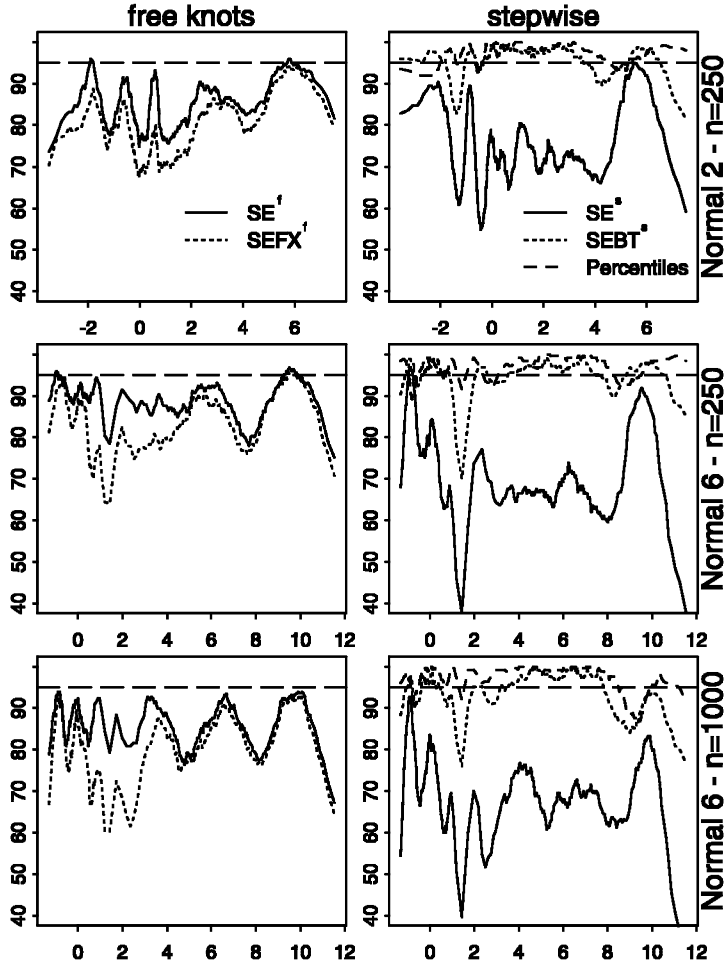

Figure 5. Coverages of various approaches to obtaining confidence intervals for logspline density estimates. The long dashes are the nominal 95% coverages.

occasionally too large. Similarly, the values of the SEBT^s were slightly too large in Figure 3. We should keep in mind that, after SE^s, SEBT^s is by far the cheapest to compute: it require only 25 bootstrap samples, as opposed to the 1,000 needed for the bootstrap percentiles, while the stepwise procedure is orders of magnitude faster than the free knot procedure.

Resampling from the fitted logspline density, rather than from the data, would prevent the problem of having repeat observations. However, with that approach the density from which we sample is no longer the true density, but rather a logspline density. A bootstrap procedure using resampling from the logspline density does therefore not address the uncertainty caused by approximating a true density (which is not a logspline density) by a logspline density.

For three of the simulation set-ups we show in Figure 5 the percent coverages of a variety of the confidence intervals. On the left side we observe that, for the intervals based on free knot splines, the two coverages follow each other fairly closely, having good and

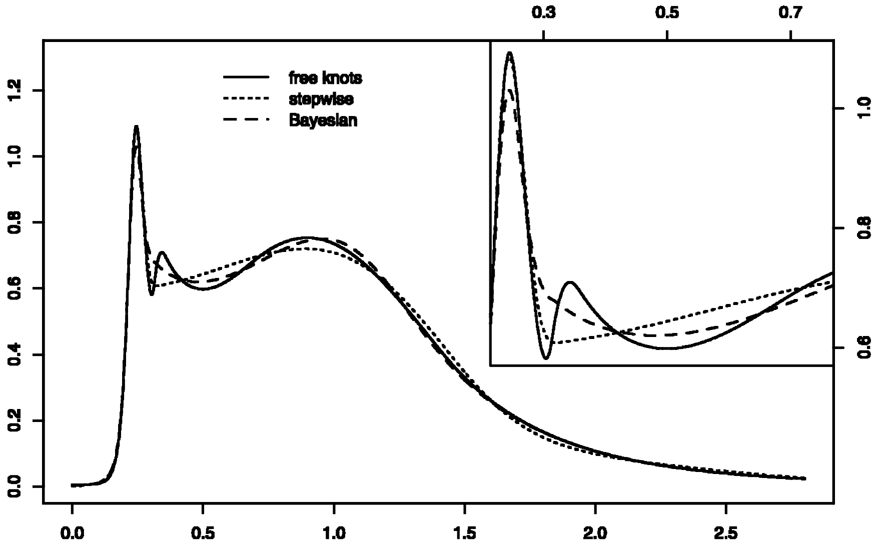

Figure 6. Logspline density estimation for the income data using a number of different procedures.

bad coverages at the same locations. For the stepwise procedures, the interval based on SE^s (solid) clearly has unacceptably low coverages. In fact, there are virtually no values of y for which the coverage is above 95% for any of the three targets that we show. The interval based on SEBT^s (dotted) has overall good coverage, but too low coverage at some spots, for example near the valley between the two peaks for the Normal 6 examples. The confidence intervals using bootstrap percentiles (short dashes) are almost everywhere close to the nominal values (long dashes).

4.1 Comparison With the Stepwise Deletion Procedure

We now briefly describe our experience in applying the stepwise deletion procedure of Kooperberg and Stone (1992). For the most part this procedure behaves comparably to the stepwise addition-deletion procedure used in the previous section. In particular:

- The stepwise deletion procedure yields slightly less biased results for the easy targets (Normal 2 and Gamma 2), but more biased results for the harder target (Normal 6). This is understandable, since the stepwise deletion procedure is less adaptive.
- SDs and SEs for both stepwise procedures are comparable. The normalized bias for the stepwise deletion procedure is, in places such as the dip for the Normal 6 distribution, much worse than for either the stepwise addition-deletion procedure or the free knot procedure.
- The overall coverage of the stepwise deletion procedure is very similar to that of the stepwise addition-deletion procedure; however, at the hard locations the coverages of the stepwise deletion procedures are much worse.

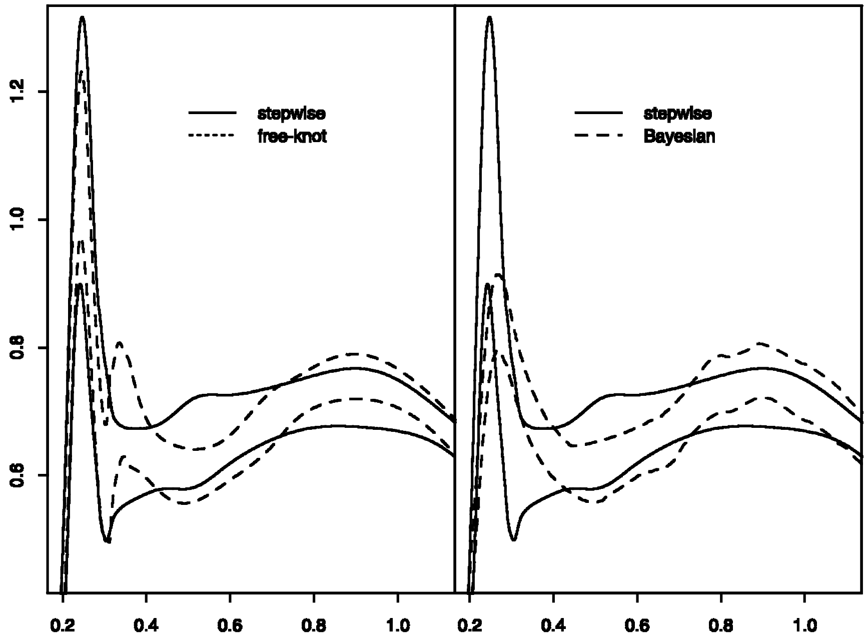

Figure 7. Comparison of the logspline pointwise confidence interval and Bayesian pointwise credible intervals for the income data.

5. INCOME DATA

We applied logspline density estimation with free knots to the income data, which we also discussed in Kooperberg and Stone (1991, 1992) and Stone, Hansen, Kooperberg, and Truong (1997). Figure 6 compares the logspline density estimation procedures with free knots and stepwise addition-deletion. The free knot spline fit has seven knots, and the stepwise addition-deletion fit has 10 knots. The "Bayesian" estimate, version (v) from Hansen and Kooperberg (2002), also has uniform priors on knot location, a uniform prior on the model size, and a multivariate normal prior distribution on the coefficients $\boldsymbol{\theta}$. Hansen and Kooperberg argued that this is a good choice of priors both with respect to the amount of smoothing and the MCMC exploration of possible logspline densities. We note that all estimates are very similar, except near the dip to the right of the peak. The free knot spline fit has a conspicuous small bump, which does not appear in the stepwise fit.

The left side of Figure 7 shows the (pointwise) 95% confidence intervals, using SE^f and SEBT^s. Observe that the confidence intervals using SE^f are considerably narrower than those using SEBT^s. The evidence for a small bump seems slim. Although this bump is present when we use free knots, the confidence interval allows for a curve without this bump. Because the confidence intervals are pointwise, this does not allow us to draw firm conclusions about the presence of the bump; however, out of 1,000 additional bootstrap estimates of the logspline density using stepwise addition-deletion, the extra bump was present in only 58 estimates. We also computed the bootstrap percentile intervals for the stepwise procedure (not shown in the figure). Overall, these intervals agree with the conclusion from the previous section: the bootstrap SE approach yields reasonable confidence intervals at a computing price that is much smaller than that for free knot splines or a full bootstrap approach.

The right side of Figure 7 shows the same pointwise confidence intervals based on SEBT^f as on the left side, but this time we added 95% credible intervals corresponding to the Bayesian estimate shown in Figure 6. The results shown in this figure are based on a run of 100,000 MCMC iterations, which takes a CPU time that is comparable to the bootstrap percentile approach and which is considerably larger than what is needed to obtain good point estimates. The 95% credible intervals are considerably smaller than the bootstrap intervals, suggesting that the coverages of the former intervals may be significantly under 95%. Several other versions that we explored of the Bayesian procedure proposed by Hansen and Kooperberg (2002) had too small intervals as well.

6. DISCUSSION

In this research we have carried out an investigation into various approaches for obtaining confidence intervals for logspline density estimates. To this end, we compared three methods for obtaining density estimates: the stepwise deletion procedure of Kooperberg and Stone (1992), the stepwise addition-deletion procedure of Stone et al. (1997), and a novel estimate employing free knot splines. Both stepwise estimates have similar bias behavior, except when the underlying curve has substantial spatial variation, in which case the stepwise addition-deletion procedure yields less biased estimates. The more adaptive free knot procedure evidently works well at yielding a nearly unbiased estimate, but it is very CPU intensive. It would be worthwhile to investigate other numerical approaches involving, for example, simulated annealing to get good starting values for the free knot spline procedure.

In order to obtain parametric pointwise confidence intervals for the various estimates, we first need to determine the corresponding standard errors. To this end, we have mainly employed the classical approach involving the negative of the Hessian matrix of the log-likelihood at the maximum likelihood estimates of the unknown parameters as an estimate of the Fisher information matrix. For the free knot procedure here we have properly taken into account the fact that the free knots are additional estimated parameters, something that is ignored for the stepwise procedures. The standard errors that we have obtained in this manner are evidently significantly too small for the estimates of the log-density and density functions, even for the free knot spline procedure, presumably because they ignore the adaptivity in the choice of the number of free knots. The resulting estimates of confidence intervals have, not surprising, substantially too low coverages. Bayesian credible intervals for density estimates that look reasonable appear too small. Bootstrap percentile intervals appear slightly ragged, suggesting that very large numbers of bootstrap samples are needed, and their coverages are too high. The bootstrap SE approach—estimating the standard error based on a limited number of bootstrap estimates and using “1.96” to obtain

95% confidence intervals—seems to have the best performance. The coverage is about right, the computational expense is low, and the pointwise confidence intervals are fairly smooth. This performance came as a pleasant surprise to us and suggests that the bootstrap SE approach deserves a more thorough investigation. We suspect that for the free-knot approach a bootstrap SE approach would also yield better confidence intervals. However, the large amount of computing needed for free-knot splines would make such an approach unattractive.

There is an interesting difference between the two stepwise procedures: the stepwise addition-deletion procedure is more adaptive than the stepwise deletion procedure and yields less biased results when there are sharp peaks or valleys in the density. As a result, for the bootstrap SE intervals, the coverages for the stepwise addition-deletion procedure are better than those for the stepwise deletion procedure.

We believe that the improvement of the free knot spline procedure over the stepwise addition-deletion procedure for logspline density estimation from Stone et al. (1997) is minimal and may not justify using the free knot spline procedure in practice (a similar conclusion was reached about the Bayesian approach by Hansen and Kooperberg (2002)). To some extent this was to be expected: for cubic splines, as used in logspline, the influence of a knot is global rather than local (after all, it is virtually impossible to detect the location of knots for cubic splines by eye); hence the exact location is only of secondary importance. This is quite different for linear splines, as used in Hare (Kooperberg, Stone, and Truong 1995), Polyclass (Kooperberg, Bose, and Stone 1997) and Triogram (Hansen, Kooperberg and Sardy 1998). Again, a similar conclusion was reached about the Bayesian approach by Hansen and Kooperberg (2002). Nevertheless, we believe that it is conceptually useful to think of the stepwise addition-deletion procedure for logspline density estimation as being a computational shortcut for approximating the free knot spline estimate.

ACKNOWLEDGMENTS

Charles Kooperberg was supported in part by National Institutes of Health grant CA74841. Charles J. Stone was supported in part by National Science Foundation grant DMS-9802071.

[Received January 2001. Revised August 2002.]

REFERENCES

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), "Automatic Bayesian Curve Fitting," *Journal of the Royal Statistical Society*, Ser. B., 60, 333–350.

DiMatteo, L., Genovese, C. R., and Kass, R. E. (2001), "Bayesian Curve-Fitting with Free-Knot Splines," *Biometrika*, 88, 1055–1071.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines" (with discussion), *The Annals of Statistics*, 19, 1–141.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Hansen, M. H., and Kooperberg, C. (2002). "Spline Adaptation in Extended Linear Models" (with discussion), *Statistical Science*, 17, 2–51.

Hansen, M., Kooperberg, C., and Sardy, S. (1998), "Triogram Models," *Journal of the American Statistical Association*, 93, 101–119.

Kooperberg, C., Bose, S., and Stone, C. J. (1997), "Polychotomous Regression," *Journal of the American Statistical Association*, 92, 117–127.

Kooperberg, C., and Stone, C. J. (1991), "A Study of Logspline Density Estimation," *Computational Statistics and Data Analysis*, 12, 327–347.

——— (1992), "Logspline Density Estimation for Censored Data," *Journal of Computational and Graphical Statistics*, 1, 301–328.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Hazard Regression," *Journal of the American Statistical Association*, 90, 78–94.

Lindstrom, M. J. (1999), "Penalized Estimation of Free-knot Splines," *Journal of Computational and Graphical Statistics*, 8, 333–352.

Schumaker, L. L. (1981), *Spline Functions: Basic Theory*, New York: Wiley.

Smith, M., and Kohn, R. (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317–343.

Stone, C. J., and Huang, J. Z. (2002), "Free Knot Splines in Concave Extended Linear Modeling," *Journal of Statistical Planning and Inference*, 108, 219–253.

Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997), "Polynomial Splines and Their Tensor Products in Extended Linear Modeling" (with discussion), *The Annals of Statistics*, 25, 1371–1470.