Monotone B-Spline Smoothing for a Generalized Linear Model Response

Wei Wang & Dylan S. Small

General

Monotone B-Spline Smoothing for a Generalized Linear Model Response

Wei WANG and Dylan S. SMALL

Various methods have been proposed for smoothing under the monotonicity constraint. We review the literature and implement an approach of monotone smoothing with B-splines for a generalized linear model response. The approach is expressed as a quadratic programming problem and is easily solved using the statistical software R. In a simulation study, we find that the approach performs better than other approaches with much faster computation time. The approach can also be used for smoothing under other shape constraints or mixed constraints. Supplementary materials of the appendices and R code to implement the developed approach is available online.

KEY WORDS: Monotonicity constraint; Quadratic programming; Regression models.

1. INTRODUCTION

1.1 Motivation and Monotonic Regression Examples

In many fields of study, it is thought, based on subject matter knowledge, that a certain predictor has a monotonic effect on a response. For instance, the probability of death is assumed to be an increasing function of the severity of burn injury (Wolfe, Roi, and Margosches 1981). A brand's unit sales are a decreasing function of price (Brezger and Steiner 2008). A dose–response curve is often assumed to be nondecreasing (Müller and Schmitt 1988; Kelly and Rice 1990; Dette, Neumeyer, and Pilz 2005; Kong and Eubank 2006). Higher levels of a biomarker are often assumed to be associated with increasing disease risk (Ghosh 2007).

In this article, we implement an approach for estimating a monotone relationship between a continuous predictor and a

generalized linear model (GLM) response using B-splines. The response variable follows a distribution in the exponential family, which includes the normal, binomial, Poisson, gamma, and other distributions. An invertible (link) function transforms the mean of the response to a possibly nonlinear form of the predictor. We use a linear combination of B-spline basis functions with unknown coefficients to approximate the nonlinear form. Thus, the link function transforms the response mean to a linear combination of B-spline basis functions, where we estimate the unknown coefficients under the monotonicity constraint. We consider an example of estimating a thought-to-be monotonically increasing relationship, the survival probability of prematurely born infants given the birth weight, which will be presented in Section 5.

1.2 Review of Semi/Nonparametric Monotonic Regression Literature

To model the dependence of a response on a predictor in a non/semiparametric way, kernels, local polynomials, or splines are popular techniques. The resulting regression estimate is not necessarily monotone and extra conditions or steps are required to produce a monotone estimate. A classical method to produce a monotone estimate is isotonic regression by the "Pool-Adjacent-Violators-Algorithm" (PAVA) (Brunk 1958; Barlow et al. 1972). Usually isotonic regression is not smooth and behaves like a step function, which is a constant function in some region and then jumps to a higher level. A hybrid approach is then to combine smoothing methods with isotonic regression to obtain both a smooth and a monotone estimate (Friedman and Tibshirani 1984; Mukerjee 1988; Mammen 1991; Ghosh 2007).

In addition to the hybrid approach, various methods of monotone smoothing have been developed (Kappenman 1987; Müller and Schmitt 1988; Ramsay 1988; He and Shi 1998; Ramsay 1998; Hall and Huang 2001; Zhang 2004; Turlach 2005; Bollaerts, Eilers, and Mechelen 2006; Dette, Neumeyer, and Pilz 2006; Park and Park 2006; Bhattacharya and Kong 2007). Dette and Pilz (2006) and Dette and Scheder (2010) conducted detailed numerical comparisons of various estimators and found their performances are similar.

Most previous research on monotone smoothing focuses on studying a normal response. Approaches have been proposed to study a more general GLM response using splines (Wang 2000; Leitenstorfer and Tutz 2007; Tutz and Leitenstorfer 2007).

Wei Wang is Senior Statistician, Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104 (E-mail: *wangw@email.chop.edu*). Dylan S. Small is Professor, Department of Statistics, the Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *dsmall@wharton.upenn.edu*). The authors thank the referees and the editor for helpful comments, and thank Scott Lorch for the premature baby data and providing insights about the dataset.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/tas*.

These methods are flexible but involve complex computation. Wang (2000) defined a two-step iterative algorithm involving approximation of the "relative curvature curves" and solution of differential equations. The approaches in Leitenstorfer and Tutz (2007) and Tutz and Leitenstorfer (2007) are based on B-splines and integrated B-splines, respectively. The boosting technique originating from machine learning for classification problems is used. Implementation of the methods can be computationally intensive. In our data analysis of a binary response and a continuous predictor from 12,607 observations, execution of both implementations exceeded memory size limit on a Mac OS X (10.7.5, 2 × 2.66 GHz 6-Core Intel Xeon, 24GB 1333 MHz DDR3).

1.3 B-Splines for Monotonic Regression

In an effort to retain the good statistical properties of B-splines while reducing the computational intensity, we implement a monotone smoothing approach for a GLM response using B-splines. B-splines are piecewise polynomials and are more flexible than single polynomials. A smooth function being approximated by B-splines is usually written as a linear combination of B-spline basis functions (a brief definition is presented in Appendix A). In the expansion, if the coefficients are in an increasing order, the linear combination is then increasing (de Boor 1978). This constraint on the coefficients thus imposes a sufficient condition for the approximated smooth function to be monotonically increasing.

This property of B-splines has been used in Kelly and Rice (1990), Kong and Eubank (2006), and Leitenstorfer and Tutz (2007). Kelly and Rice (1990) and Kong and Eubank (2006) modeled a normal response and solved the constrained optimization problems using Fortran routines. Leitenstorfer and Tutz (2007) used boosting techniques to enforce the constraint. Our approach is expressed as a well-studied quadratic programming problem and so can be easily solved using standard optimization packages. The approach is easily implemented as we illustrate using the R function *solve.QP* in the package *quadprog*.

1.4 Outline

We present details of the model-fitting procedure in Section 2. In Section 3, we compare the performance of our estimator with the following estimators in simulations for a binary response: the conventional B-spline estimator, and the estimators in Leitenstorfer and Tutz (2007) and Tutz and Leitenstorfer (2007). In the settings we tried, our estimator generally outperforms others and also improves on computation time. In Section 4, we apply our approach to estimate the survival probability of infants as a function of birth weight assuming the survival probability is monotonically increasing. The approach we develop can also be used for estimation under other shape constraints besides monotonicity (e.g., nonnegativity) or multiple constraints simultaneously; this will be discussed in Section 5.

2. APPROACH

Let y be the response with a distribution in the exponential family, which includes distributions of Bernoulli/binomial, nor-

mal, Poisson, and others (McCullagh and Nelder 1989). Let μ be the mean of y. Let g be the link function. We assume $g(\mu)$ is connected with the predictor x through a smooth function f as $g(\mu) = f(x)$. In the following, however, we adopt the formulation in Leitenstorfer and Tutz (2007) and Tutz and Leitenstorfer (2007) that $g(\mu) = \beta_0 + f(x)$, where β_0 is the intercept term. In this way, f is confounded with β_0 and is only estimable under an identifiability constraint. Let x_i be the ith predictor value. We add the constraint $\sum_i f(x_i) = 0$ (Hastie and Tibshirani 1990; Wood 2006). In the case of a linear f, we have $g(\mu) = \beta_0 + \beta x$, where β is the slope and the constraint requires that x is centered by its sample mean.

For a Bernoulli y, its canonical link function has the logistic form, $g(\mu) = \ln[\mu/(1 - \mu)]$. It thus follows $\mu = 1/[1 + \exp(-\beta_0 - f(x))]$. Let y_i denote the ith response with mean μ_i, $i = 1, \ldots, n$. For the Bernoulli y_i, the log-likelihood of the ith individual is $\log[\mu_i^{y_i}(1 - \mu_i)^{1-y_i}] = y_i \log[\mu_i/(1 - \mu_i)] + \log(1 - \mu_i)$. The log-likelihood of all the individuals follows as $\sum_{i=1}^{n}\{y_i \log[\mu_i/(1 - \mu_i)] + \log(1 - \mu_i)\}$. In the following, we expand f by a linear combination of B-spline basis functions. We estimate the unknown coefficients via maximizing the log-likelihood function under the monotonicity constraint imposed on f. We focus on a Bernoulli response with logistic link function.

Let ϕ_p's be a series of B-spline basis functions with knots given. The expansion of f by ϕ_p's gives

$$f(x) = \sum_{p=1}^{P} \phi_p(x)\beta_p \equiv \boldsymbol{\phi}(x)'\boldsymbol{\beta}, \qquad (1)$$

where $\boldsymbol{\phi}(x) = (\phi_1(x), \ldots, \phi_P(x))'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_P)'$ is the unknown coefficient vector. Suppose f has an increasing effect on y. By de Boor (1978) (see also Appendix A), the monotonicity constraints on the coefficients is then $\beta_2 - \beta_1 \geq 0, \ldots, \beta_p - \beta_{p-1} \geq 0, \ldots, \beta_P - \beta_{P-1} \geq 0$. Expressed in a vector/matrix form, we have

$$\begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \end{pmatrix} \geq \mathbf{0},$$

where $\mathbf{0}$ is a zero vector and the inequality $\geq$ holds elementwise. Let $\mathbf{A}$ denote the $(P - 1) \times P$ matrix on the left-hand side of the inequality, which has elements -1 and 1 on the two main diagonals and elements 0 elsewhere. The constraint is thus

$$\mathbf{A}\boldsymbol{\beta} \geq \mathbf{0}. \qquad (2)$$

If f has a decreasing effect on the response, the constraint is then $(-\mathbf{A})\boldsymbol{\beta} \geq \mathbf{0}$.

Let $\Lambda(\boldsymbol{\beta})$ be the log-likelihood function. Our goal is to estimate $\boldsymbol{\beta}$ via maximizing $\Lambda(\boldsymbol{\beta})$ under the constraint (2). By definition, we have $g(\mu_i) = \beta_0 + \boldsymbol{\phi}(x_i)'\boldsymbol{\beta}$. For a Bernoulli y_i with a logistic link function, we get $\log[\mu_i/(1 - \mu_i)] = \beta_0 + \boldsymbol{\phi}(x_i)'\boldsymbol{\beta}$ and $\mu_i = 1/[1 + \exp(-\beta_0 - \boldsymbol{\phi}(x_i)'\boldsymbol{\beta})]$. Hence, $\Lambda(\boldsymbol{\beta}) = \sum_{i=1}^{n}\{y_i[\beta_0 + \boldsymbol{\phi}(x_i)'\boldsymbol{\beta}] - \log[1 + \exp(\beta_0 + \boldsymbol{\phi}(x_i)'\boldsymbol{\beta})]\}$.

Throughout the article, we use cubic B-splines. To control the smoothness of the fitted f, usually a penalty term on the roughness of the fit, $\int [f''(x)]^2 dx$, is added to the log-likelihood func-

tion for maximization over $\boldsymbol{\beta}$. From (1), the penalty term can be expressed as $\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}$, where $\mathbf{P} = \int (\partial^2 \boldsymbol{\phi}/\partial x^2)(\partial^2 \boldsymbol{\phi}/\partial x^2)' dx$. Let λ be a fixed smoothing parameter associated with f. Estimation of $\boldsymbol{\beta}$ is obtained via maximizing the penalized log-likelihood $\Lambda(\boldsymbol{\beta}) - (\lambda/2)\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}$ under the constraint $\mathbf{A}\boldsymbol{\beta} \geq 0$. For a fixed λ, we find the estimate by iteration. Given the previous estimate, we approximate the log-likelihood and get the current estimate by solving a quadratic programming problem. Detailed derivation is presented in Appendix B. To choose λ, we use common information criteria discussed at the end of this section.

Let v_i be the variance of y_i. Let η_i be the linear predictor $\beta_0 + \boldsymbol{\phi}(x_i)'\boldsymbol{\beta}$. By definition, η_i is a function of μ_i, that is, $g(\mu_i) = \eta_i$. Let $w_i = v_i^{-1}(\partial \mu_i / \partial \eta_i)^2$ and $d_i = w_i(\partial \eta_i / \partial \mu_i)$. For Bernoulli responses with a logistic link, we have $v_i = w_i = \mu_i(1 - \mu_i)$ and $d_i = 1$. Let $\phi_i \equiv \phi(x_i) = (\phi_1(x_i), \ldots, \phi_P(x_i))'$. Define a $P \times n$ matrix $\mathbf{B}' = (\phi_1, \ldots, \phi_n)$. That is, the (p, i)th element of $\mathbf{B}'$ is $\phi_p(x_i)$, $p = 1, \ldots, P$, $i = 1, \ldots, n$. Let $\tilde{\boldsymbol{\beta}}$ be the estimate computed from the previous iteration, based on which we define $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_n)'$, $\tilde{\mathbf{D}} = \text{diag}(\tilde{d}_1, \ldots, \tilde{d}_n)$, and $\tilde{\mathbf{W}} = \text{diag}(\tilde{w}_1, \ldots, \tilde{w}_n)$. Let $\mathbf{y} = (y_1, \ldots, y_n)'$. As calculated in Appendix B, maximizing the penalized log-likelihood $\Lambda(\boldsymbol{\beta}) - (\lambda/2)\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}$ over $\boldsymbol{\beta}$ can be approximated by the iterative optimization problem

$$\min_{\boldsymbol{\beta}} \left\{ -\left[(\mathbf{y} - \tilde{\boldsymbol{\mu}})\tilde{\mathbf{D}} + \tilde{\boldsymbol{\beta}}'\mathbf{B}'\tilde{\mathbf{W}} \right]' \mathbf{B}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}' \left(\mathbf{B}'\tilde{\mathbf{W}}\mathbf{B} + \lambda\mathbf{P} \right) \boldsymbol{\beta} \right\}. \tag{3}$$

Given the estimate from the previous iteration, solving (3) alone for $\boldsymbol{\beta}$ gives the conventional B-spline estimate at the current iteration. The whole procedure can be viewed as a modified iteratively weighted least-square procedure (see, e.g., O'Sullivan, Yandell, and Raynor 1986; Eilers and Marx 1996). Solving (3) under the constraint (2) for current estimate of $\boldsymbol{\beta}$ is a quadratic programming problem (Goldfarb and Idnani 1982, 1983). The optimal solution has to be found in the constrained region $\mathbf{A}\boldsymbol{\beta} \geq 0$.

Common criteria for choosing a smoothing parameter λ include Akaike information criterion (AIC), Bayesian information criterion (BIC), or generalized cross-validation (GCV). Calculation of all the criteria involves the "effective degrees of freedom" (df), which is obtained as the trace of the matrix $\mathbf{H} = \mathbf{B}(\mathbf{B}'\mathbf{WB} + \lambda\mathbf{P})^{-1}\mathbf{B}'\mathbf{W}$ (O'Sullivan, Yandell, and Raynor 1986). The AIC is defined as twice the df minus twice the log-likelihood, and BIC replaces twice the df as $\log(n)$ multiplying the df. The GCV is computed as $\sum_i (y_i - \mu_i)^2 / [v_i(n - \text{df})^2]$ using parameter estimates. Usually a sequence of the smoothing parameters is specified. At each point, the model is fitted and the criterion is calculated. The optimal smoothing parameter is chosen as the one where the criterion attains the minimum.

3. SIMULATION

We conduct simulations to compare the performance of the monotone B-spline estimator (MB) with other estimators: the conventional unrestricted B-spline estimator (CB), the estimator proposed by Leitenstorfer and Tutz (2007) (BB: boosting B-spline), and estimator proposed by Tutz and

Leitenstorfer (2007) (BI: boosting integrated B-spline). Estimators BB and BI are implemented through R packages from *http://www.statistik.lmu.de/institut/lehrstuhl/semsto/software/index.html*. In both approaches, one crucial tuning parameter plays the role of a smoothing parameter, is selected using the AIC or BIC criterion. The other parameters take the recommended values as in Leitenstorfer and Tutz (2007) and Tutz and Leitenstorfer (2007).

For the CB and MB estimators, we take 100 grid points of the smoothing parameter λ equally spaced between -8 and 6 on a log scale. We also use the information criterion AIC or BIC to choose the optimal λ. The models we will use are not of overly complicated forms, and so to save computation time, we use a small number of basis functions, 5, for all the estimators. The knots are equally spaced within the range of the data.

The predictor x is simulated from the Uniform $[0, 1]$ distribution and the response is simulated from a Bernoulli distribution. Components of the linear predictor are $\beta_0 = 0.5$ and f based on the following forms $f_1 = \sqrt{x}$, $f_2 = x + \sin(2\pi x)/(2\pi)$, $f_3 = x$, $f_4 = 1/(1 + \exp(5 - 10x))$, and $f_5 = 0.9x^2 + 0.1x$. The functions $f_1, \ldots, f_5$ are scaled by a factor of 4 and then shifted to range from -2 to 2. Shapes of the transformed functions are plotted in Figure 1.

We consider sample sizes 100 and 500 and run 1000 simulations in each scenario. Let x_j^m and x_j^M be the minimum and the maximum of the simulated x values in the jth simulation. Evaluation of the estimated mean functions is based on 500 equally spaced points within the range: $[\max\{x_j^m, \ j\}, \min\{x_j^M, \ j\}]$. At each point, we calculate the squared bias, variance, and mean squared error (MSE). The averaged values over the 500 points are summarized in Table 1 for each estimator. The associated computation time per simulation is shown in Table 2.

In Table 1, models 1–5 correspond to f_1–f_5, respectively. The smallest value is in bold font for each combination of model and n. The bias of the MB estimator is close to that of the CB estimator. The MB estimator has smaller variance with a reduction of the MSE except for model 4 when $n = 500$. Estimators BI and BB perform best for models 4 and 5, respectively. On the
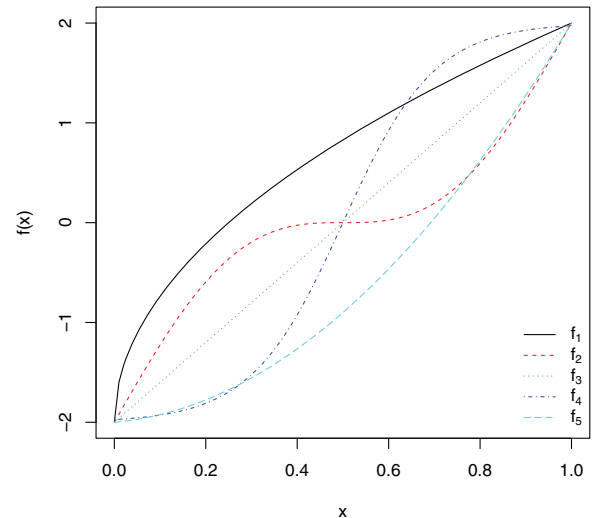

Figure 1. Functions f_1–f_5 used in the linear predictor for simulations.

Table 1. Averaged squared bias (Bias²), variance (Var.), and MSE of estimating the mean functions. Numbers are in a 10^{-3} scale

n	Estimator	Model 1			Model 2			Model 3			Model 4			Model 5		
		Bias²	Var.	MSE	Bias²	Var.	MSE	Bias²	Var.	MSE	Bias²	Var.	MSE	Bias²	Var.	MSE
100	BB (AIC)	0.26	3.71	3.97	2.63	4.93	7.56	0.08	4.71	4.79	2.22	4.24	6.46	**0.13**	4.60	**4.73**
	BB (BIC)	0.26	4.02	4.28	2.32	5.13	7.45	0.29	4.94	5.23	2.52	4.47	6.99	0.29	4.87	5.16
	BI (AIC)	0.27	4.66	4.93	**0.59**	5.79	6.38	0.57	5.77	6.34	**0.95**	4.33	**5.28**	0.53	6.08	6.61
	BI (BIC)	0.67	4.84	5.51	1.11	5.92	7.03	1.13	5.94	6.96	0.93	4.48	5.61	0.93	6.27	7.20
	CB (AIC)	**0.17**	4.28	4.45	0.76	6.68	7.44	< 0.01	4.75	4.75	1.08	5.03	6.11	0.16	5.46	5.62
	CB (BIC)	0.39	3.42	3.81	2.11	4.84	6.95	< 0.01	3.83	3.83	1.92	4.25	6.17	0.66	4.45	5.11
	MB (AIC)	0.22	3.61	3.83	1.45	4.82	**6.27**	< 0.01	4.16	4.16	1.84	4.20	6.04	0.31	4.67	4.98
	MB (BIC)	0.42	**3.24**	**3.66**	2.49	**4.13**	6.62	< 0.01	**3.74**	**3.74**	2.24	**3.91**	6.15	0.75	**4.21**	4.96
500	BB (AIC)	0.44	0.90	1.34	3.27	**1.11**	4.38	0.08	1.11	1.19	1.73	0.94	2.67	0.10	**0.95**	**1.05**
	BB (BIC)	0.48	0.92	1.40	3.40	1.20	4.60	0.13	1.14	1.27	1.84	0.96	2.80	0.15	0.96	1.11
	BI (AIC)	0.47	1.25	1.72	0.83	1.49	2.32	0.70	1.38	2.08	0.71	**0.82**	**1.53**	0.64	1.26	1.90
	BI (BIC)	0.52	1.27	1.79	0.87	1.51	2.38	0.72	1.39	2.11	0.71	0.83	1.54	0.66	1.28	1.94
	CB (AIC)	**0.11**	1.21	1.32	**0.22**	1.76	1.98	< 0.01	1.01	1.01	**0.51**	1.24	1.75	**0.09**	1.26	1.35
	CB (BIC)	0.55	0.87	1.42	1.73	1.47	3.20	< 0.01	0.75	0.75	1.82	1.09	2.91	0.57	1.04	1.61
	MB (AIC)	0.14	1.08	**1.22**	0.39	1.41	**1.80**	< 0.01	0.92	0.92	1.30	0.90	2.20	0.12	1.08	1.20
	MB (BIC)	0.56	**0.85**	1.41	1.92	1.28	3.20	< 0.01	**0.74**	**0.74**	2.32	0.83	3.15	0.58	1.02	1.60

NOTE: The smallest value is in bold font for each combination of model and n.

other hand, compared with CB, estimator BB has larger MSE in models 2–4 when $n = 100$ and in models 2 and 3 when $n = 500$. Estimator BI has larger MSE in models 1, 3, and 5 than CB. The MB estimator has the smallest MSE for models 1–3.

Table 2 shows the computation time per simulation associated with the estimators (AIC and BIC together) for each combination of model and n. When $n = 100$, the CB and MB estimators' time is about 1/3 or 1/4 of BI, and between 1/7 and 1/4 of BB. When $n = 500$, the CB and MB estimators are at least two times faster than BI and about nine times faster than BB.

In summary, estimator MB improves CB in nearly all of the settings in terms of the MSE and has the smallest MSE in most of the settings.

4. DATA ANALYSIS

We study data on premature infants born at a birth weight between 500 and 2000 g in Los Angeles and surrounding counties between 1997 and 1999. This is a subset of the data described in Lorch et al. (2012). We are interested in estimating the probability of survival given the birth weight. Survival refers to whether the infant is alive 60 months post gestation. In total, 11396 infants survived and 1211 died. For premature infants, survival is thought to be an increasing function of birth weight (Hack et al. 1991; Cooper et al. 1998).

Table 2. Computation time (minutes) per simulation associated with the estimators (AIC and BIC together) for each model

n	Estimator	Model 1	Model 2	Model 3	Model 4	Model 5
100	BB	2.77	2.70	3.05	3.05	2.56
	BI	1.56	1.60	1.61	2.46	1.74
	CB&MB	0.42	0.37	0.53	0.67	0.53
500	BB	12.02	12.04	14.61	14.31	15.91
	BI	3.40	3.29	3.42	3.99	3.99
	CB&MB	1.36	1.26	1.51	1.69	1.41

For a description of the data, we group the birth weight into 16 bins with bin width 100 g. The minimum number of infants in a group is 448. We then calculate the sample average of survivals within each bin and plot them against the mid-points of the bin in Figure 2. In the figure, we see the survival rate increases sharply at low birth weight, and then the increasing rate decreases and becomes flat as the birth weight approaches 2000 g. The estimated survival probabilities using estimators CB and MB are also shown in Figure 2. We use 12 equally spaced knots, and the optimal λ was chosen by AIC or BIC from a grid of 100 points equally spaced between −5 and 15 on the log scale. In
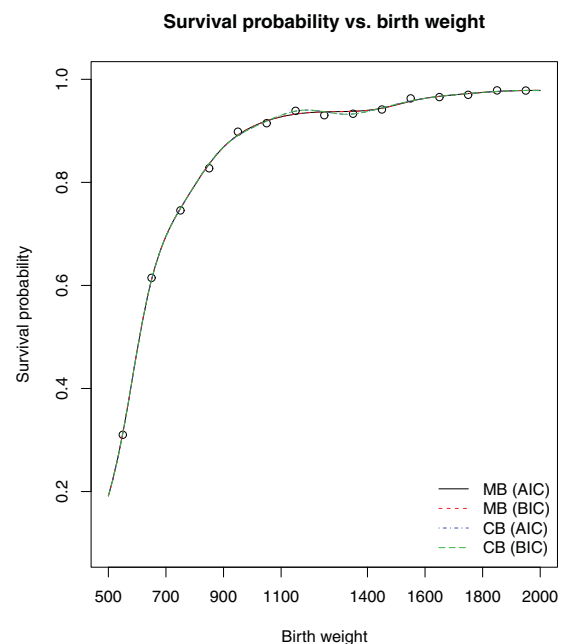

Figure 2. Survival rates and the estimated survival probabilities by estimators CB and MB where smoothing parameters are selected by the AIC or BIC criterion.
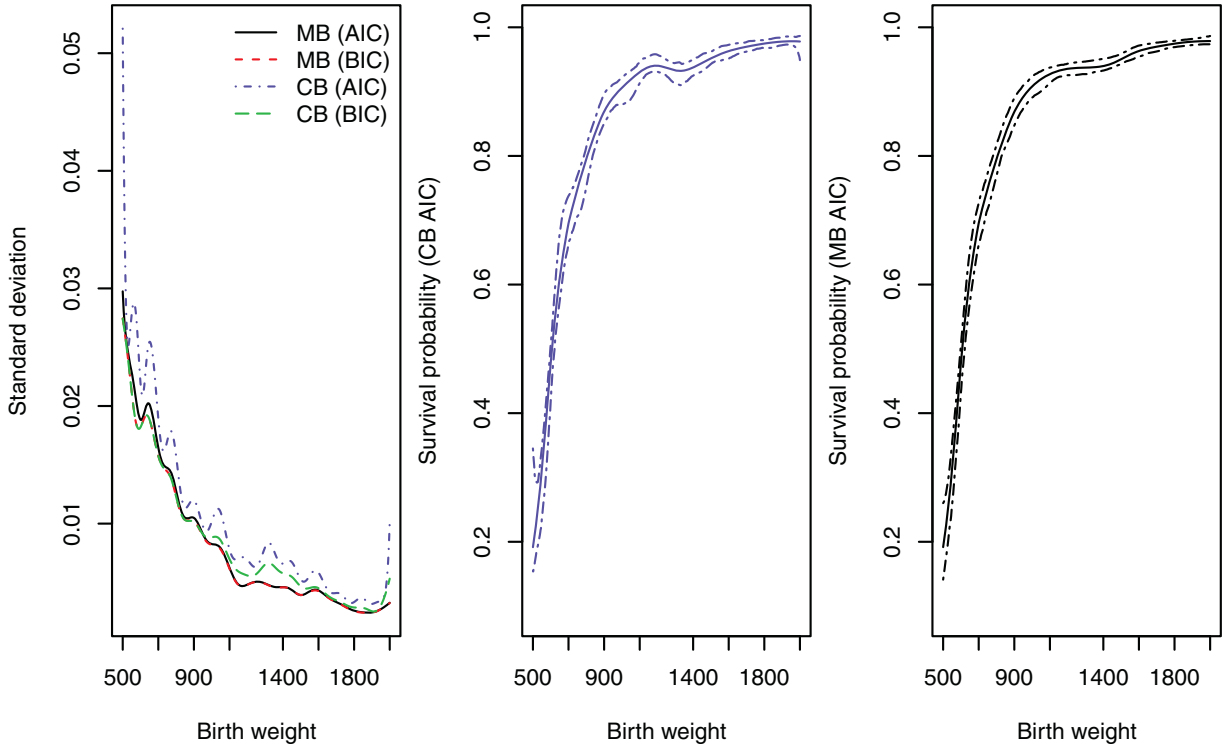
Figure 3. Standard deviations (left panel) of the CB and MB estimates, and confidence intervals of the CB AIC estimate (middle panel) and of the MB AIC estimate (right panel) obtained from the bootstrap.

the plot, we see that CB AIC and CB BIC estimates overlap, and so do the two MB estimates. The CB estimates do not produce a monotone curve in the region between 1100 and 1400.

In total, implementation of the approach took 50 min on the platform described at the end of Section 1.2. We also tried estimating the survival probability with estimators BB and BI. However, executions of the implementation halted with error message of exceeding memory size limit and so the analysis results using these approaches are not available.

To obtain standard deviations of the estimates and to build confidence intervals at each birth weight value, we use the bootstrap method as in Dilleen, Heimann, and Hirsch (2003). We resample the paired birth weight and survival status with 200 bootstrap replications. At each birth weight, we obtain the estimated standard deviation as the sample standard deviation of the 200 estimates. We use the 97.5% and 2.5% quantiles of the estimates as the upper and lower bound of a 95% confidence interval. Figure 3 shows the standard deviations for all estimators and confidence intervals for the CB (AIC) and MB (AIC) estimators. In the plots, we observe less variability in the MB estimators.

5. SUMMARY AND DISCUSSION

A smooth function being approximated by B-splines is usually written as a linear combination of B-spline basis functions. In the expansion, an increasing order of the coefficients imposes a sufficient condition for the approximated smooth function to be monotonically increasing (de Boor 1978). We implement a monotone smoothing approach based on B-splines for a GLM

response using the R function "solve.QP" of the package "quadprog." In the simulation studies, the approach generally outperforms the other estimators with much faster computation time.

Using B-spline properties (de Boor 1978), the approach can also be applied to estimate a smooth function under other shape constraints by changing the matrix $\mathbf{A}$ in (2). For example, nonnegativity of the coefficients, β_k's, guarantees the estimate to be nonnegative. These nonnegative constraints can then be imposed by a diagonal matrix $\mathbf{A}$. If the estimate is required to be convex, that is, the second-order derivatives being nonnegative, we can look at the second-order differences of the coefficients.

For instance, we assume the knots are equally spaced. Let $\mathbf{A}$ be a $(P - 2) \times P$ matrix with elements 1, -2, and 1 on the main diagonals and element 0 elsewhere as

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & & \ddots & & \ddots \\ & & 1 & -2 & 1 \end{pmatrix}.$$

Then (2) requires the nonnegativity of the second-order differences of β_p's, which in turn ensures convexity of the estimate. If the knots are not equally spaced, the structure of $\mathbf{A}$ remains the same but the values of nonzero elements will change according to formulas of B-spline derivatives (de Boor 1978). The approach can also handle more than one shape constraint simultaneously by binding multiple $\mathbf{A}$ matrices into one.

SUPPLEMENTARY MATERIALS

Appendix A and B: A brief description of the B-spline basis functions and derivation of Equation (3). (.pdf file)

R code for MB estimator: R code to implement the developed MB estimator in the article. (.R file)

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*, New York: Wiley. [28]

Bhattacharya, R., and Kong, M. (2007), "Consistency and Asymptotic Normality of the Estimated Effective Doses in Bioassay," *Journal of Statistical Planning and Inference*, 137, 643–658. [28]

Bollaerts, K., Eilers, P. H. C., and Mechelen, I. V. (2006), "Simple and Multiple *p*-Splines Regression With Shape Constraints," *British Journal of Mathematical and Statistical Psychology*, 59, 451–469. [28]

Brezger, A., and Steiner, W. J. (2008), "Monotonic Regression Based on Bayesian P-Splines: An Application to Estimating Price Response Functions From Store-Level Scanner Data," *Journal of Business & Economic Statistics*, 26, 90–104. [28]

Brunk, H. D. (1958), "On the Estimation of Parameters Restricted by Inequalities," *Annals of Mathematical Statistics*, 29, 437–454. [28]

Cooper, T. R., Berseth, C. L., Adams, J. M., and Weisman, L. E. (1998), "Actuarial Survival in the Premature Infant Less Than 30 Weeks' Gestation," *Pediatrics*, 101, 975–978. [31]

de Boor, C. (1978), *A Practical Guide to Splines*. New York: Springer. [29,32]

Dette, H., Neumeyer, N., and Pilz, K. F. (2005), "A Note on Nonparametric Estimation of the Effective Dose in Quantal Bioassay," *Journal of American Statistical Association*, 100, 503–510. [28]

Dette, H., Neumeyer, N., and Pilz, K. F. (2006), "A Simple Nonparametric Estimator of a Strictly Monotone Regression Function," *Bernoulli*, 12, 469–490. [28]

Dette, H., and Pilz, K. F. (2006), "A Comparative Study of Monotone Nonparametric Kernel Estimates," *Journal of Statistical Computation and Simulation*, 76, 41–56. [28]

Dette, H., and Scheder, R. (2010), "A Finite Sample Comparison of Nonparametric Estimates of the Effective Dose in Quantal Bioassay," *Journal of Statistical Computation and Simulation*, 80, 527–544. [28]

Dilleen, M., Heimann, G., and Hirsch, I. (2003), "Non-Parametric Estimators of a Monotonic Dose-Response Curve and Bootstrap Confidence Intervals," *Statistics in Medicine*, 22, 869–882. [32]

Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing With b-Splines and Penalties," *Statistical Science*, 11, 89–121. [30]

Friedman, J., and Tibshirani, R. (1984), "The Monotone Smoothing of Scatterplots," *Technometrics*, 26, 243–250. [28]

Ghosh, D. (2007), "Incorporating Monotonicity into the Evaluation of a Biomarker," *Biostatistics*, 8, 402–413. [28]

Goldfarb, D., and Idnani, A. (1982), "Dual and Primal-Dual Methods for Solving Strictly Convex Quadratic Programs," in *Numerical Analysis*, ed. J. P. Hennart, Springer-Verlag, pp 226–239. [30]

Goldfarb, D., and Idnani, A. (1983), "A Numerically Stable Dual Method for Solving Strictly Convex Quadratic Programs," *Mathematical Programming*, 27, 1–33. [30]

Hack, M., Horbar, J. D., Malloy, M. H., Wright, L., Tyson, J. E., and Wright, E. (1991), "Very Low Birth Weight Outcomes of the National Institute of Child Health and Human Development Neonatal Network," *Pediatrics*, 87, 587–597. [31]

Hall, P., and Huang, L. S. (2001), "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics*, 29, 624–647. [28]

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, Boca Raton, FL: Chapman & Hall/CRC. [29]

He, X., and Shi, P. (1998), "Monotone b-Spline Smoothing," *Journal of American Statistical Association*, 93, 643–650. [28]

Kappenman, R. (1987), "Nonparametric Estimation of Dose-Response Curves With Application to ed50 Estimation," *Journal of Statistical Computation and Simulation*, 28, 1–13. [28]

Kelly, C., and Rice, J. (1990), "Monotone Smoothing With Application to Dose-Response Curves and the Assessment of Synergism," *Biometrics*, 46, 1071–1085. [28,29]

Kong, M., and Eubank, R. L. (2006), "Monotone Smoothing With Application to Dose-Response Curve," *Communications in Statistics: Simulation and Computation*, 35, 991–1004. [28,29]

Leitenstorfer, F., and Tutz, G. (2007), "Generalized Monotonic Regression Based on b-Splines With an Application to Air Pollution Data," *Biostatistics*, 8, 654–673. [28,29,30]

Lorch, S. A., Baiocchi, M., Ahlberg, C. E., and Small, D. S. (2012), "The Differential Impact of Delivery Hospital on the Outcomes of Premature Infants," *Pediatrics*, 130, 270–278. [31]

Mammen, E. (1991), "Estimating a Smooth Monotone Regression Function," *The Annals of Statistics*, 19, 724–740. [28]

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*. (2nd ed.), New York: Chapman & Hall. [29]

Mukerjee, H. (1988), "Monotone Nonparametric Regression," *The Annals of Statistics*, 16, 741–750. [28]

Müller, H.-G., and Schmitt, T. (1988), "Kernel and Probit Estimates in Quantal Bioassay," *Journal of American Statistical Association*, 83, 750–759. [28]

O'Sullivan, F., Yandell, B. S., and Raynor, W. J. J. (1986), "Automatic Smoothing of Regression Functions in Generalized Linear Models," *Journal of American Statistical Association*, 81, 96–103. [30]

Park, D., and Park, S. (2006), "Parametric and Nonparametric Estimators of ed100α," *Journal of Statistical Computation and Simulation*, 76, 661–672. [28]

Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3, 425–461. [28]

——— (1998), "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society*, Series B, 60, 365–375. [28]

Turlach, B. A. (2005), "Shape Constrained Smoothing Using Smoothing Splines," *Computational Statistics*, 20, 81–103. [28]

Tutz, G., and Leitenstorfer, F. (2007), "Generalized Smooth Monotonic Regression in Additive Modeling," *Journal of Computational and Graphical Statistics*, 16, 165–188. [28,29,30]

Wang, Z. (2000), "An Algorithm for Generalized Monotonic Smoothing," *Journal of Applied Statistics*, 27, 495–507. [28]

Wolfe, R. A., Roi, L. D., and Margosches, E. (1981), "Monotonic Dichotomous Regression: A Burn Care Example," *Biometrics*, 37, 157–167. [28]

Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R,* Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science. [29]

Zhang, J. T. (2004), "A Simple and Efficient Monotone Smoother Using Smoothing Splines," *Nonparametric Statistics*, 16, 779–796. [28]