Simulation Studies of Smoothing Parameter Estimates and Bayesian Confidence Intervals in Bernoulli SS ANOVA Models

Yuedong Wang
Department of Biostatistics, University of Michigan,
Ann Arbor, Michigan 48109

Grace Wahba, Rick Chappell
Department of Statistics, University of Wisconsin,
Madison, WI 53706

Chong Gu
Department of Statistics, Purdue University
West Lafayette, IN 47907

Key Words and Phrases: generalized cross validation; GRKPACK; Newton-Raphson iteration; RKPACK; smoothing parameter; smoothing spline ANOVA; unbiased risk estimate.

ABSTRACT

Wahba, Wang, Gu, Klein and Klein (1994b) introduced Smoothing Spline ANalysis of VAriance (SS ANOVA) method for data from exponential families. They used GCV and UBR methods to choose smoothing parameters. They also constructed Bayesian confidence intervals for SS ANOVA estimates. In this paper, we report simulation results concerning the performance of SS ANOVA estimates with smoothing parameters chosen by GCV or UBR method and their Bayesian confidence intervals. We also compare the performance between GCV and UBR.

1 SS ANOVA Method

Nelder and Wedderburn (1972) introduced a collection of statistical regression models known as generalized linear models (GLM's) for the analysis of data from exponential families (McCullagh and Nelder, 1989). GLM's have been used successfully in many areas. To achieve greater flexibility than a linear GLM regression surface, O'Sullivan, Yandell and Raynor (1986) and Gu (1990) used penalized likelihood method with smoothing splines and thin plate splines. Wahba et al. (1994b) introduced the SS ANOVA method for data from exponential families. In this section, we briefly introduce the SS ANOVA method. See also Wahba, Gu, Wang and Chappell (1994a), Wahba, Wang, Gu, Klein and Klein (1994c) and Wang (1994b) for details of the SS ANOVA method. The Gaussian case was considered in detail in Gu and Wahba (1993a) and Gu and Wahba (1993b).

Suppose data have the form $(y_i; \boldsymbol{t}_i),\ i = 1, 2, \cdots, n$, where $\boldsymbol{t}_i = (t_{1i}, \cdots, t_{di})$. y_i's are independent observations with density function

$$g(y_i; f_i, \phi) = \exp((y_i h(f_i) - b(f_i))/a(\phi) + c(y_i, \phi)), \tag{1}$$

where $f_i = f(\boldsymbol{x}_i)$ is the parameter of interest and $h(f_i)$ is a monotone transformation of f_i known as the *canonical parameter*. ϕ is an unknown scale parameter. The purpose is to investigate the global relationship between f and $\boldsymbol{x}$.

Assume $t_j \in \mathcal{T}^{(j)}$, where $\mathcal{T}^{(j)}$ is a measurable space. Let $\mathcal{T} = \mathcal{T}^{(1)} \otimes \cdots \otimes \mathcal{T}^{(d)}$. Let $d\mu_j$ be a probability measure on $\mathcal{T}^{(j)}$ and let $\mathcal{H}^{(j)}$ be a reproducing kernel Hilbert space (RKHS) (Aronszajn, 1950) of functions on $\mathcal{T}^{(j)}$ with $\int_{\mathcal{T}^{(j)}} f_j(t_j) d\mu_j = 0$ for $f_j(t_j) \in \mathcal{H}^{(j)}$. Let $\{1^{(j)}\}$ be the one dimensional space of constant functions on $\mathcal{T}^{(j)}$. Consider the RKHS

$$
\begin{aligned}
\mathcal{G} &= \prod_{j=1}^{d} (\{1^{(j)}\} \oplus \{\mathcal{H}^{(j)}\}) \\
&= \{1\} \oplus \sum_{j} \mathcal{H}^{(j)} \oplus \sum_{j<k} (\mathcal{H}^{(j)} \otimes \mathcal{H}^{(k)}) \oplus \cdots, \qquad (2)
\end{aligned}
$$

where $\{1\}$ denotes the constant functions on $\mathcal{T}$. An element f_j in $\mathcal{H}^{(j)}$ is called a *main effect*, an element f_{jk} in $\mathcal{H}^{(j)} \otimes \mathcal{H}^{(k)}$ is called a *two factor interaction*, and so on.

Similar to ANOVA, usually model space is a subspace $\mathcal{M}$ of $\mathcal{G}$. When a model is chosen, we can regroup and write the model space as

$$
\mathcal{M} = \mathcal{H}^0 \oplus \sum_{j=1}^{q} \mathcal{H}^j, \qquad (3)
$$

where $\mathcal{H}^0$ is a finite dimensional space containing functions which are not going to be penalized, usually lower order polynomials. See Wahba et al. (1994b) for details. A SS ANOVA estimate is the solution to the following variational problem:

$$
\min_{f \in \mathcal{M}} \left\{ -\sum_{i=1}^{n} l_i(f_i) + \frac{n}{2} \sum_{j=1}^{q} \lambda_j \|P_j f\|^2 \right\}, \qquad (4)
$$

where $l_i(f_i)$ is the log likelihood. The first part in (4) measures the goodness of fit. In the second part, P_j is the orthogonal projector in $\mathcal{M}$ onto $\mathcal{H}^j$ and $\|P_j f\|^2$ is a quadratic roughness penalty. λ_j's are a set of smoothing parameters. They control the trade-off between the goodness of fit and the roughness of the estimate. Writing $\lambda_j = \lambda/\theta_j$, (4) becomes

$$
\min_{f \in \mathcal{M}} \left\{ -\sum_{i=1}^{n} l_i(f_i) + \frac{n}{2} \lambda \|P_* f\|_\Theta^2 \right\}, \qquad (5)
$$

where $P_* = \sum_{j=1}^{q} P_j$ is the orthogonal projection in $\mathcal{M}$ onto $\mathcal{H}_* = \sum_{j=1}^{q} \mathcal{H}^j$.

$\|f\|_\Theta^2 = \|P_0 f\|^2 + \sum_{j=1}^q \theta_j^{-1} \|P_j f\|^2$ is a modified norm indexed by $\Theta = (\theta_1, \cdots, \theta_q)$. We denote R_j as the reproducing kernel (RK) (Aronszajn, 1950) for $\mathcal{H}^j$ under the original norm. The RK for $\sum_{j=1}^q \mathcal{H}^j$ under $\|\cdot\|_\Theta$ is $R_\Theta = \sum_{j=1}^q \theta_j R_j$.

The solution to (5) has the form (Wahba, 1990; O'Sullivan et al., 1986)

$$f_{\lambda,\Theta}(\boldsymbol{t}) = \boldsymbol{\phi}(\boldsymbol{t})^T \boldsymbol{d} + \boldsymbol{\xi}(\boldsymbol{t})^T \boldsymbol{c}, \tag{6}$$

where $\boldsymbol{\phi}^T(\boldsymbol{t}) = (\phi_1(\boldsymbol{t}), \cdots, \phi_M(\boldsymbol{t}))$, $\{\phi_v\}_{v=1}^M$ is a set of basis functions of $\mathcal{H}^0$, $M = \dim(\mathcal{H}^0)$, $\boldsymbol{\xi}^T(\boldsymbol{t}) = (R_\Theta(\boldsymbol{t}_1, \boldsymbol{t}), \cdots, R_\Theta(\boldsymbol{t}_n, \boldsymbol{t}))$. $\boldsymbol{c}_{n\times 1}$ and $\boldsymbol{d}_{M\times 1}$ are vectors of coefficients to be estimated. Substituting (6) into (5), we can estimate $\boldsymbol{c}$ and $\boldsymbol{d}$ by minimizing

$$I(\boldsymbol{c}, \boldsymbol{d}) = -\sum_{i=1}^n l_i(\boldsymbol{\phi}^T(\boldsymbol{t}_i)\boldsymbol{d} + \boldsymbol{\xi}^T(\boldsymbol{t}_i)\boldsymbol{c}) + \frac{n}{2}\lambda \boldsymbol{c}^T Q_\Theta \boldsymbol{c}, \tag{7}$$

where Q_Θ is a $n \times n$ matrix with $Q_\Theta(i,k) = R_\Theta(\boldsymbol{t}_i, \boldsymbol{t}_k)$. Since l_i's are not quadratic, (7) can not be solved directly. But if all $l_i(f_i)$'s are strictly concave, we can use Newton-Raphson procedure to compute $\boldsymbol{c}$ and $\boldsymbol{d}$ for fixed λ and Θ. Let $u_i = -dl_i/df_i$, $\boldsymbol{u}^T = (u_1, \cdots, u_n)$, $w_i = -d^2 l_i/df_i^2$, $W = \operatorname{diag}(w_1, \cdots, w_n)$, $S = (\boldsymbol{\phi}(\boldsymbol{t}_1), \cdots, \boldsymbol{\phi}(\boldsymbol{t}_n))^T$, $\boldsymbol{f} = S\boldsymbol{d} + Q_\Theta \boldsymbol{c}$. With the following transformations, $\tilde{Q}_\Theta = W_-^{1/2} Q_\Theta W_-^{1/2}$, $\tilde{\boldsymbol{c}} = W_-^{-1/2}\boldsymbol{c}$, $\tilde{S} = W_-^{1/2} S$, $\tilde{\boldsymbol{d}} = \boldsymbol{d}$, $\tilde{\boldsymbol{y}} = W_-^{-1/2}(W_- \boldsymbol{f}_- - \boldsymbol{u}_-)$, where the subscript minus indicates quantities evaluated at the previous Newton-Raphson iteration, the Newton-Raphson iteration equation can be simplified to

$$\begin{aligned} (\tilde{Q}_\Theta + n\lambda I)\tilde{\boldsymbol{c}} + \tilde{S}\tilde{\boldsymbol{d}} &= \tilde{\boldsymbol{y}}, \\ \tilde{S}^T \tilde{\boldsymbol{c}} = \boldsymbol{0}. \end{aligned} \tag{8}$$

Choosing appropriate smoothing parameters is crucial for effectively estimating the true function from data by fitting smoothing spline models. Gu

(1992) showed that we should choose smoothing parameters at each Newton-Raphson iteration. The generalized cross validation (GCV) method estimates smoothing parameters at each iteration by minimizing the GCV score

$$V(\lambda, \Theta) = \frac{1/n\|(I - A(\lambda, \Theta))\tilde{\boldsymbol{y}}\|^2}{[(1/n)tr(I - A(\lambda, \Theta))]^2}, \tag{9}$$

where $A(\lambda, \Theta)$ satisfies

$$(w_{1-}^{1/2} f_{\lambda,\Theta}(\boldsymbol{t}_1), \cdots, w_{n-}^{1/2} f_{\lambda,\Theta}(\boldsymbol{t}_n))^T = A(\lambda, \Theta)\tilde{\boldsymbol{y}}, \tag{10}$$

and $f_{\lambda,\Theta}(\boldsymbol{t}_i)$'s are computed from the solution of (8).

The unbiased risk (UBR) method estimates smoothing parameters at each iteration by minimizing the following unbiased risk estimate

$$\tilde{U}(\lambda, \Theta) = \frac{1}{n}\|(I - A(\lambda, \Theta))\tilde{\boldsymbol{y}}\|^2 + 2\frac{\hat{\sigma}^2}{n} trA(\lambda, \Theta), \tag{11}$$

where $\hat{\sigma}^2 = 1/n \sum_{i=1}^n u_{i-}^2/w_{i-}$, an estimate of dispersion parameter ϕ. If the dispersion parameter is known to be 1, such as in the case of binary data and Poisson data, a better estimate is

$$U(\lambda, \Theta) = \frac{1}{n}\|(I - A(\lambda, \Theta))\tilde{\boldsymbol{y}}\|^2 + \frac{2}{n} trA(\lambda, \Theta). \tag{12}$$

As in Gu (1992), Lemma 1 and Lemma 2 in Appendix A provide theoretical justification to the use of GCV and UBR method.

A generic code RKPACK (Gu, 1989; Gu and Wahba, 1991) is available to solve (8) and estimate λ and Θ via GCV (option V) or the UBR method at the same time. When using the UBR method, we can either specify $\sigma^2 = 1$ (option U) or estimate σ^2 (option U˜). Thus we can solve $\boldsymbol{c}$ and $\boldsymbol{d}$ in (8) by iteratively utilizing RKPACK until converge. Wang (1994a) developed GRKPACK for binary, binomial, Poisson and Gamma data based on RKPACK.

Based on a Bayes model, Wahba et al. (1994b) also proved that the posterior mean of this Bayes model approximately equals to the solution to (4). The posterior variance can be used to construct Bayesian confidence intervals. The Bayes model and formulae are given in Appendix B. See Wang (1994a) for discussions about how to calculate these Bayesian confidence intervals. See also Wang and Wahba (1995).

2 Simulations

In this section, we conduct 4 simulations to

1. Evaluate the performance of the SS ANOVA method;
2. Compare the performance of the GCV method and the UBR method for choosing smoothing parameters;
3. Assess the effectiveness of the GCV method and the UBR method. We compare the SKL's of estimates with smoothing parameters chosen by GCV or UBR with the SKL's of estimates with optimal smoothing parameters;
4. Evaluate the performance of Bayesian confidence intervals.

Binary data are generated based on an additive model

$$\text{logit}[\mathrm{P}(Y=1|t_1,t_2)] = f(t_1,t_2) = C + f_1(t_1) + f_2(t_2), \qquad 0 \le t_1,\ t_2 \le 1,$$

where C is a constant, $f_1(t_1)$, $f_2(t_2)$ are smooth functions which integrate to zero over $[0,1]$. Two different underlying functions f are used:

Case I	$C=0, \quad f_1(t_1) = 2t_1 - 1, \quad f_2(t_2) = t_2 - 0.5;$
Case II	$C=0,$

$$f_1(t_1) = -8(t_1 - 0.5)^2 + 8(t_1 - 0.5)^3 \mathrm{I}_{\{t_1 > 0.5\}} - A_1,$$
$$f_2(t_2) = (2.5t_2 - 1)^3 - A_2,$$

where $\mathrm{I}_{\{t_1>0.5\}} = 1$ if $t_1 > 0.5$ and $\mathrm{I}_{\{t_1>0.5\}} = 0$ otherwise. A_1 and A_2 are constants such that the integrals of f_1 and f_2 over $[0, 1]$ equal zero. For each case, the two main effects and the 3-dimensional probability surface are plotted in Figure 1. Two different sample sizes are used: $n = 200$ and $n = 400$. Hence we have $2 \times 2 = 4$ simulations. We generated the design points $\boldsymbol{t}_i$ from an uniform distribution on $[0, 1]^2$ and generated binary responses using the underlying function. Designs and responses are generated for 100 replications of each of these 4 simulations. Notice that we have a random design. That is, the designs are different for different replications. We use the Fortran routines `uni` of the Core Mathematics Library (Cmlib) from the National Bureau of Standards to generate random numbers in all the simulations. The symmetrized Kullback-Leibler discrepancy SKL $= \frac{1}{n}\sum_{i=1}^{n}(p_i - \hat{p}_i) \times (f_i - \hat{f}_i)$ measures the performance of an estimate, where $p = e^f/(1 + e^f)$ is the probability of $Y = 1$. In one replication of Case I with n=200, RKPACK failed to find a reasonable descent direction and the estimation was poor. We excluded this replication in the following plots.

First, we evaluate the performance of the SS ANOVA method. The UBR method (option U) is used to choose the smoothing parameters since it works better than option U˜ of UBR and GCV (see below). We select the 5th, 25th, 50th, 75th and 95th best estimates ordered by SKL. Their main effects estimates are plotted in Figure 2 for Case I and Figure 3 for Case II. Their estimates of the probability functions are plotted in Figure 4 for Case I and n=200, Figure 5 for Case I and n=400, Figure 6 for Case II and n=200, and Figure 7 for Case II and n=400. We conclude from these plots that the SS ANOVA method can capture the shape of a underlying model most of the

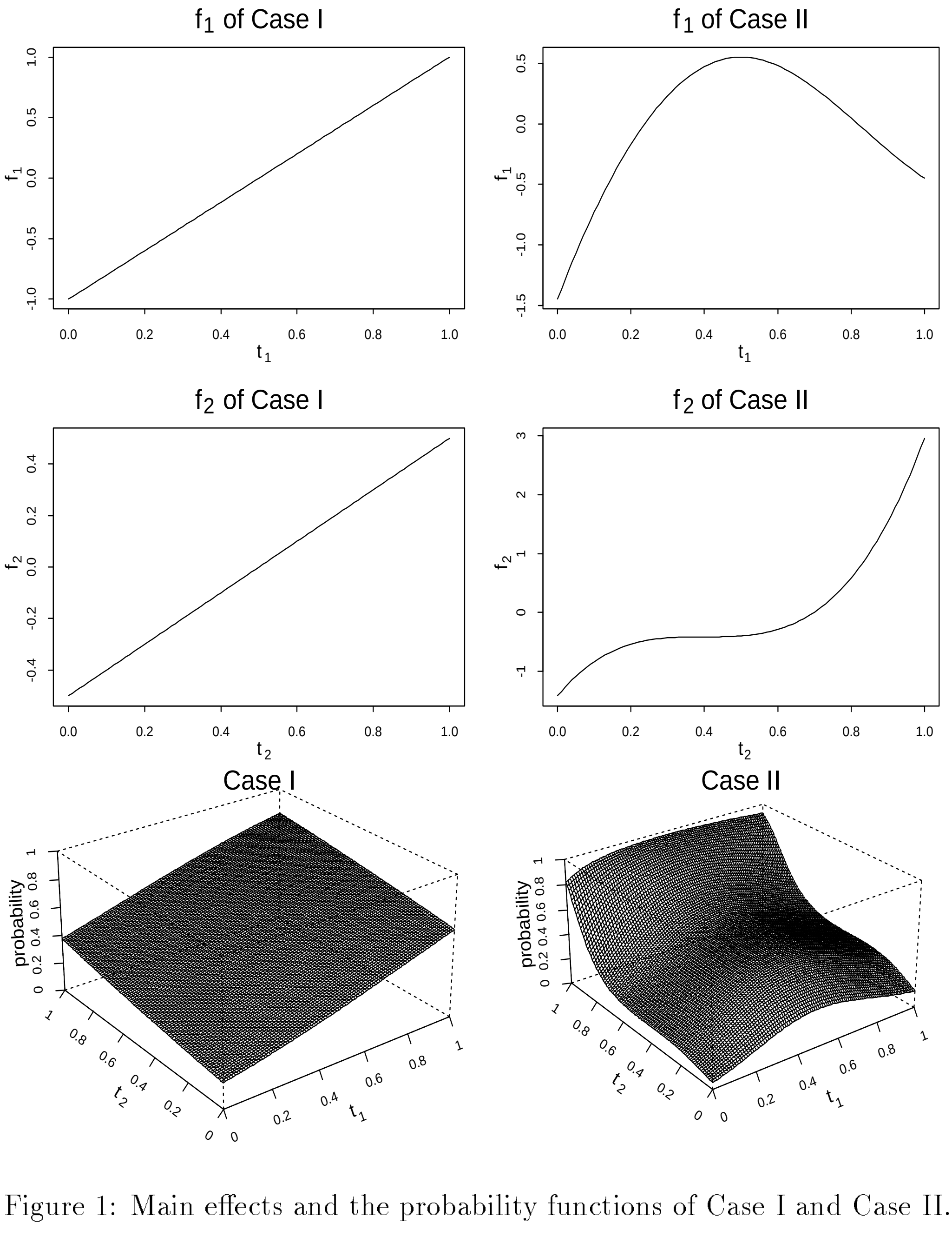

Figure 1: Main effects and the probability functions of Case I and Case II.

time even for small sample sizes. But it needs a moderate large sample size to capture the shapes of components.

To compare the GCV method and the option U˜ of UBR with the option U of UBR method for choosing smoothing parameters, we repeat the same simulations with option U˜ and V. As shown in Figure 8, the SKL's with option U are uniformly smaller than the SKL's with option U˜ , and the SKL's with option U are smaller than the SKL's with option V most of the time. The differences become smaller when the sample size increases. So the option U of UBR is recommended if there is no over- or under-dispersion.

These simulations examine estimation in the case of untied Bernoulli data, for which anomalous dispersion is not an issue. Though U˜ is not in practice used in such cases, in which ϕ is identically unity, its appearance in this simulation is still useful: it assesses the increase in SKL incurred by estimating ϕ. This increase is relevant to binomial situations in which overdispersion is possible or, equivalently, when the quantity to be minimized in equation (4) is considered a penalized log-quasilikelihood.

To assess the performance of the GCV method and the UBR method, for a fixed underlying function and a fixed sample, we use the simplex method to search for the minimum SKL as λ_1 and λ_2 vary. That is, for 3 fixed pairs of λ_1 and λ_2, we used RKPACK to calculate the estimates with fixed smoothing parameters. The SKL's for each pair of λ_1 and λ_2 can be calculated since we know the true function. A new pair of λ_1 and λ_2 is found by the simplex method. This procedure is iterated until convergence. The SKL reaches a minimum at the converged point. Comparing the SKL's using the GCV method and the UBR method with the optimal SKL's, Figure 9 shows that both GCV and UBR work well.

To evaluate the performance of Bayesian confidence intervals, we record the number of data points at which the confidence intervals cover the true

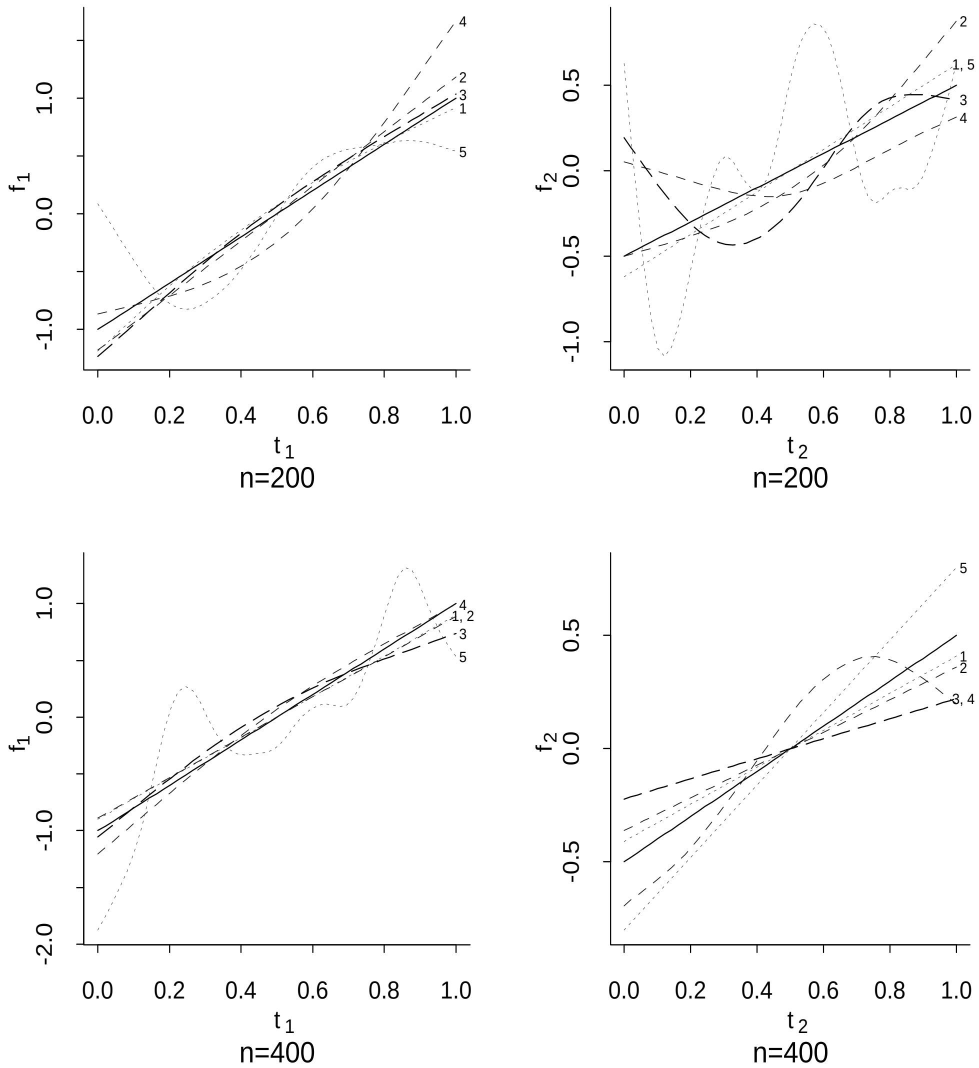

Figure 2: Estimates of the main effects of Case I. Solid lines are the true main effects. Five dashed lines in each graph are main effect estimates: 1, 2, 3, 4 and 5 are the 5th, 25th, 50th, 75th, 95th best estimates ordered by SKL.

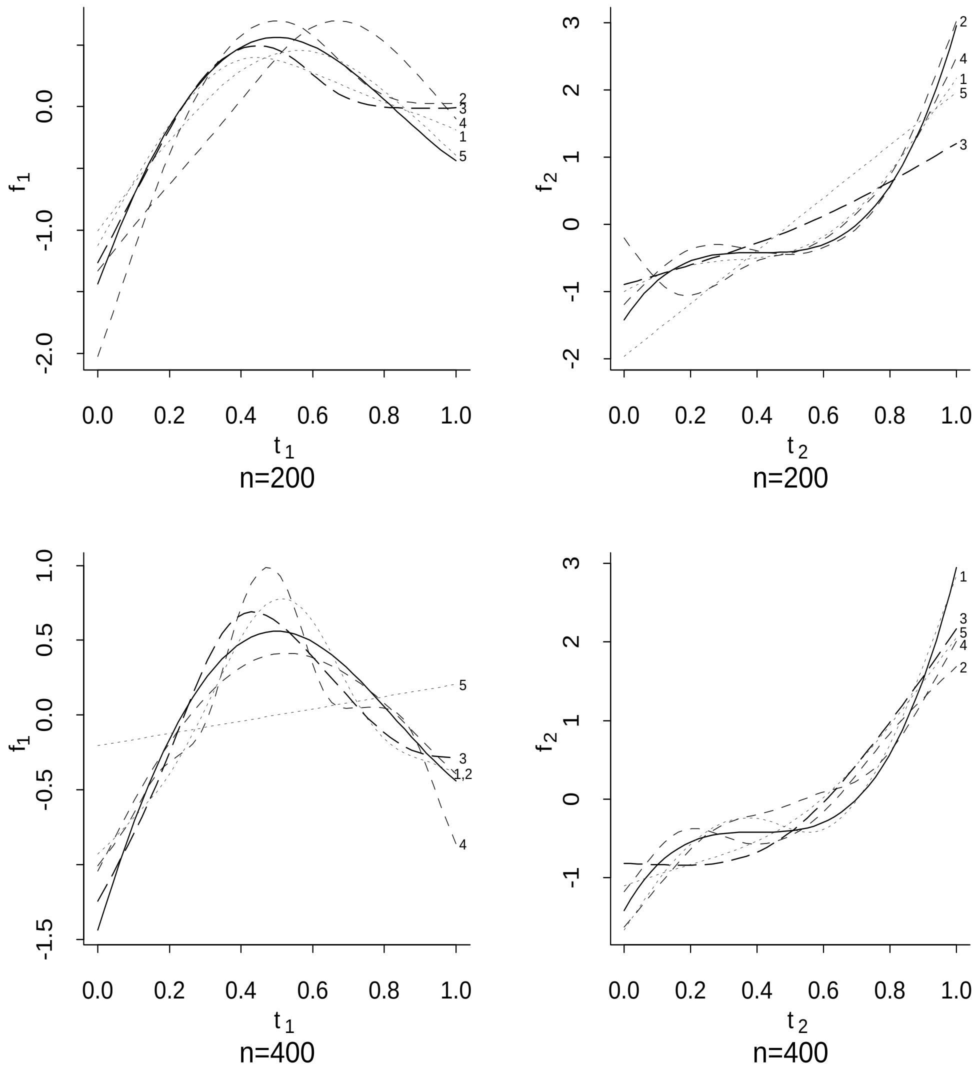

Figure 3: Estimates of the main effects of Case II. Solid lines are the true main effects. Five dashed lines in each graph are main effect estimates: 1, 2, 3, 4 and 5 are the 5th, 25th, 50th, 75th, 95th best estimates ordered by SKL.

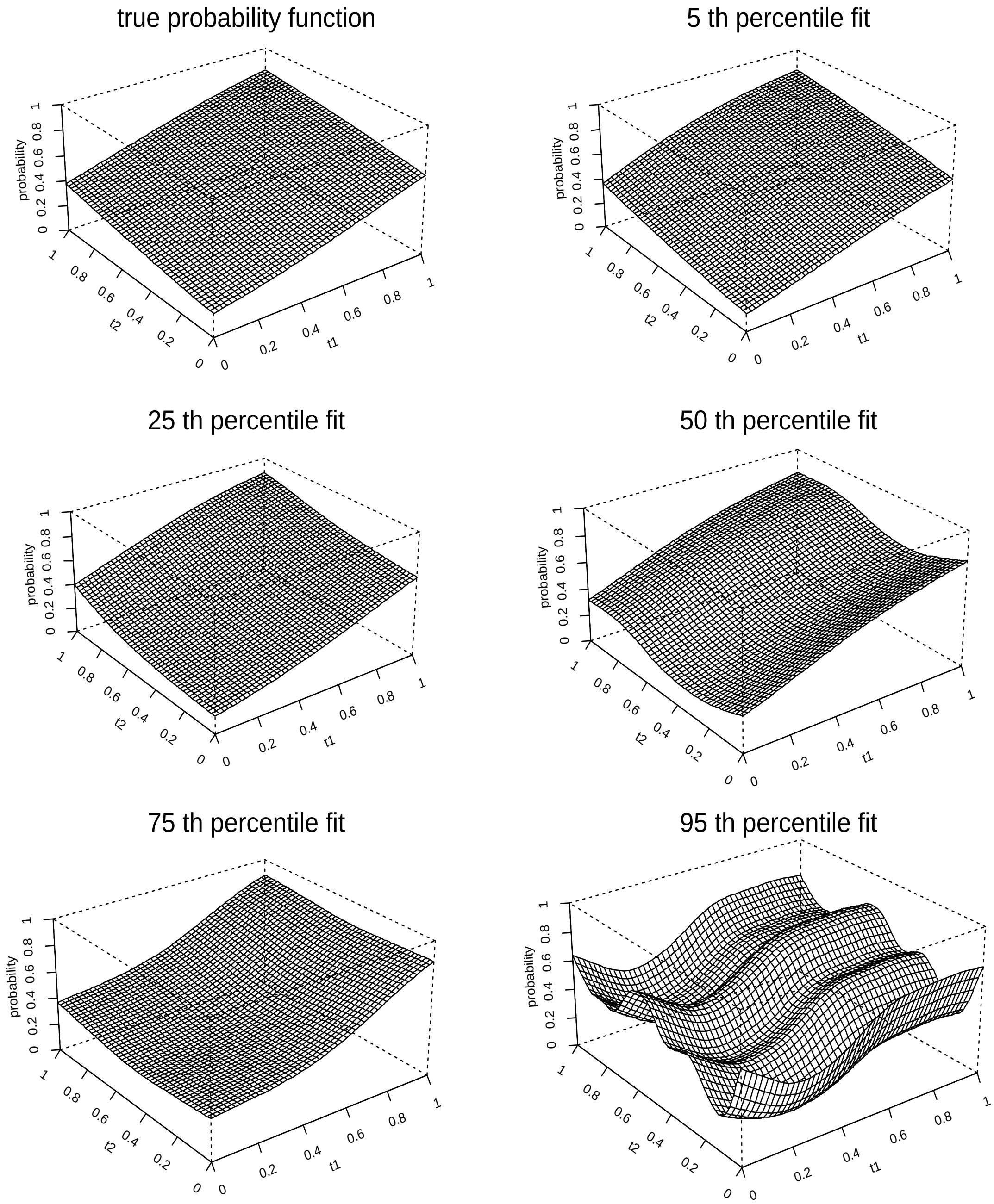

Figure 4: The true probability function and their estimates for Case I with n=200.

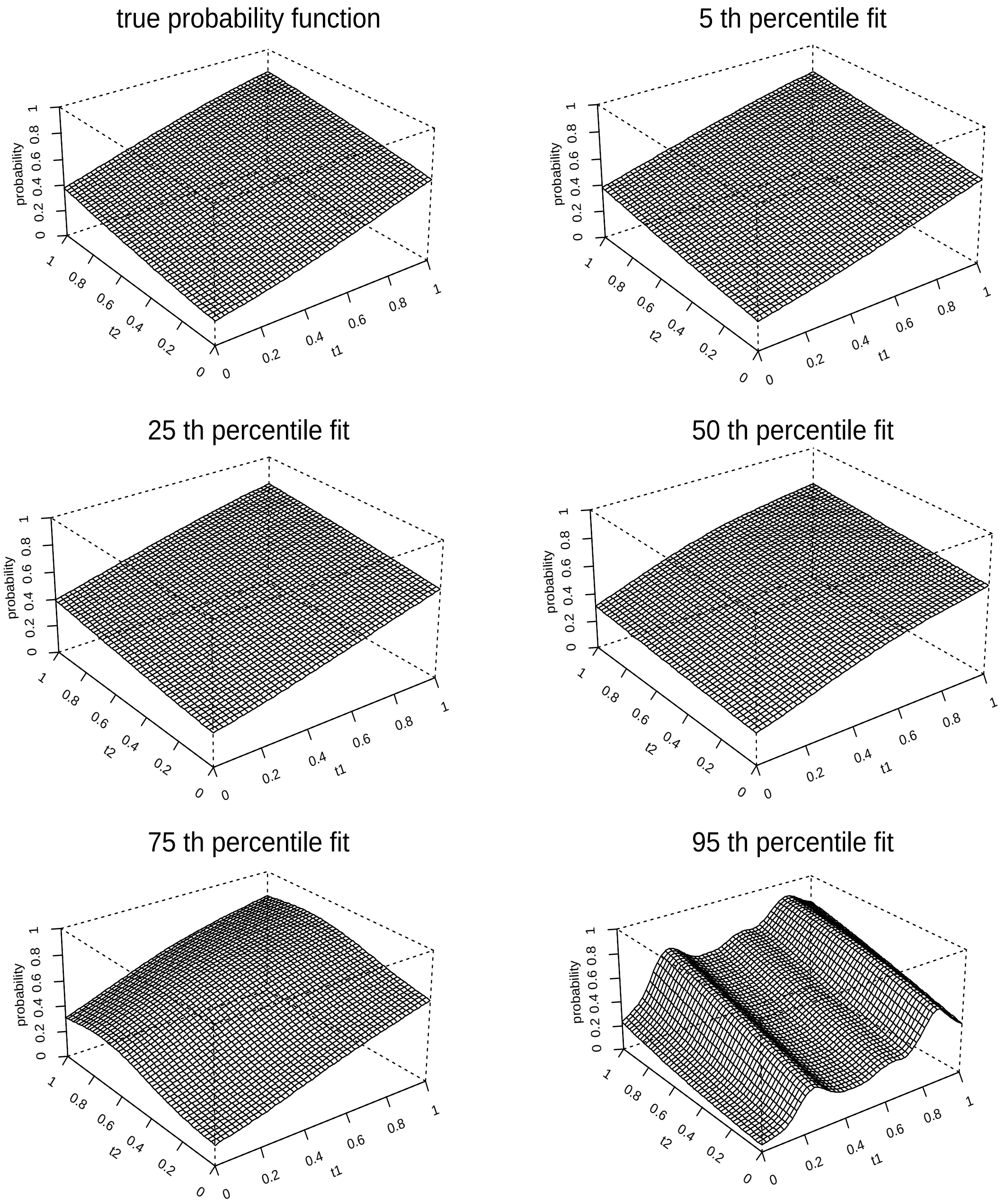

Figure 5: The true probability function and their estimates for Case I with n=400.

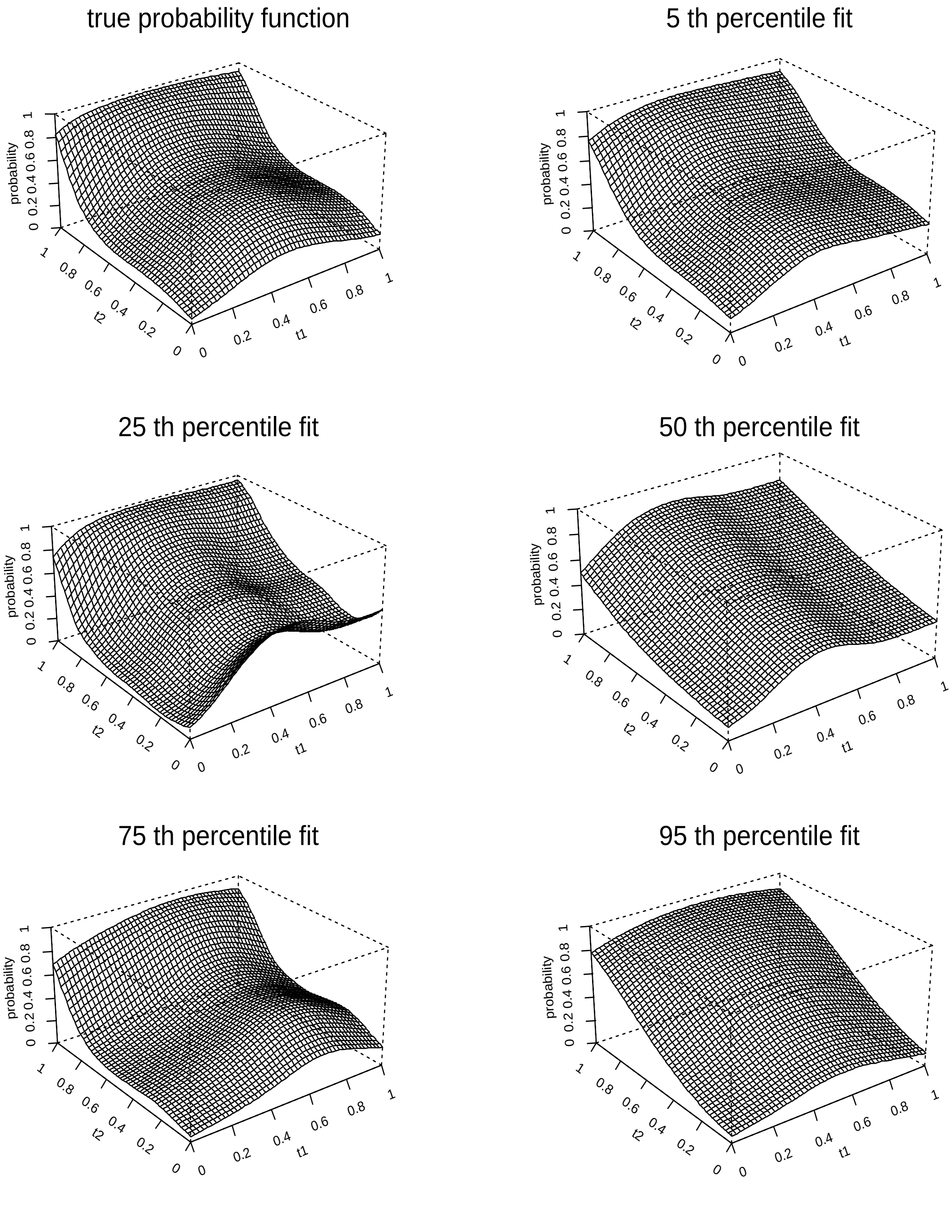

Figure 6: The true probability function and their estimates for Case II with n=200.

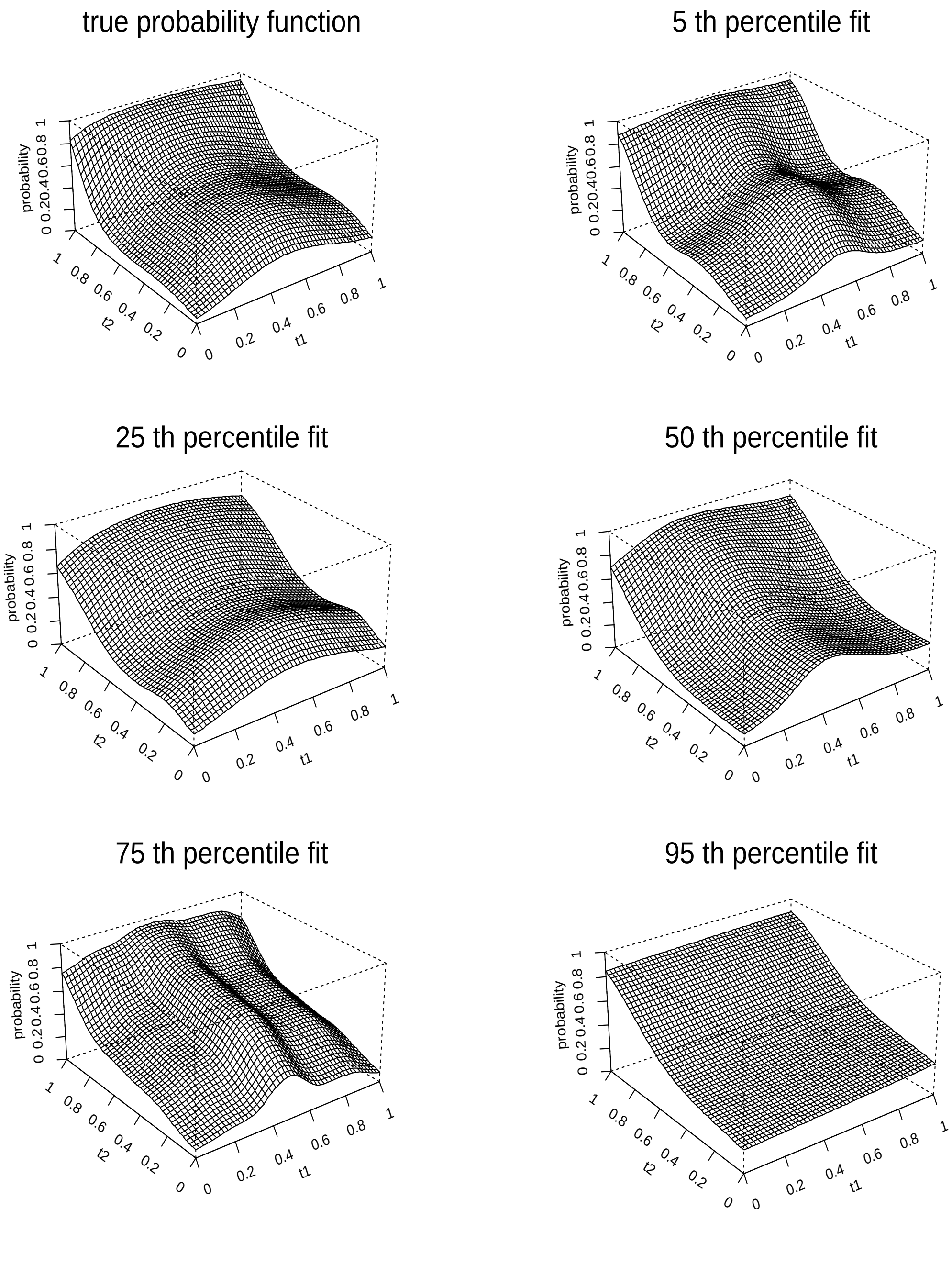

Figure 7: The true probability function and their estimates for Case II with n=400.

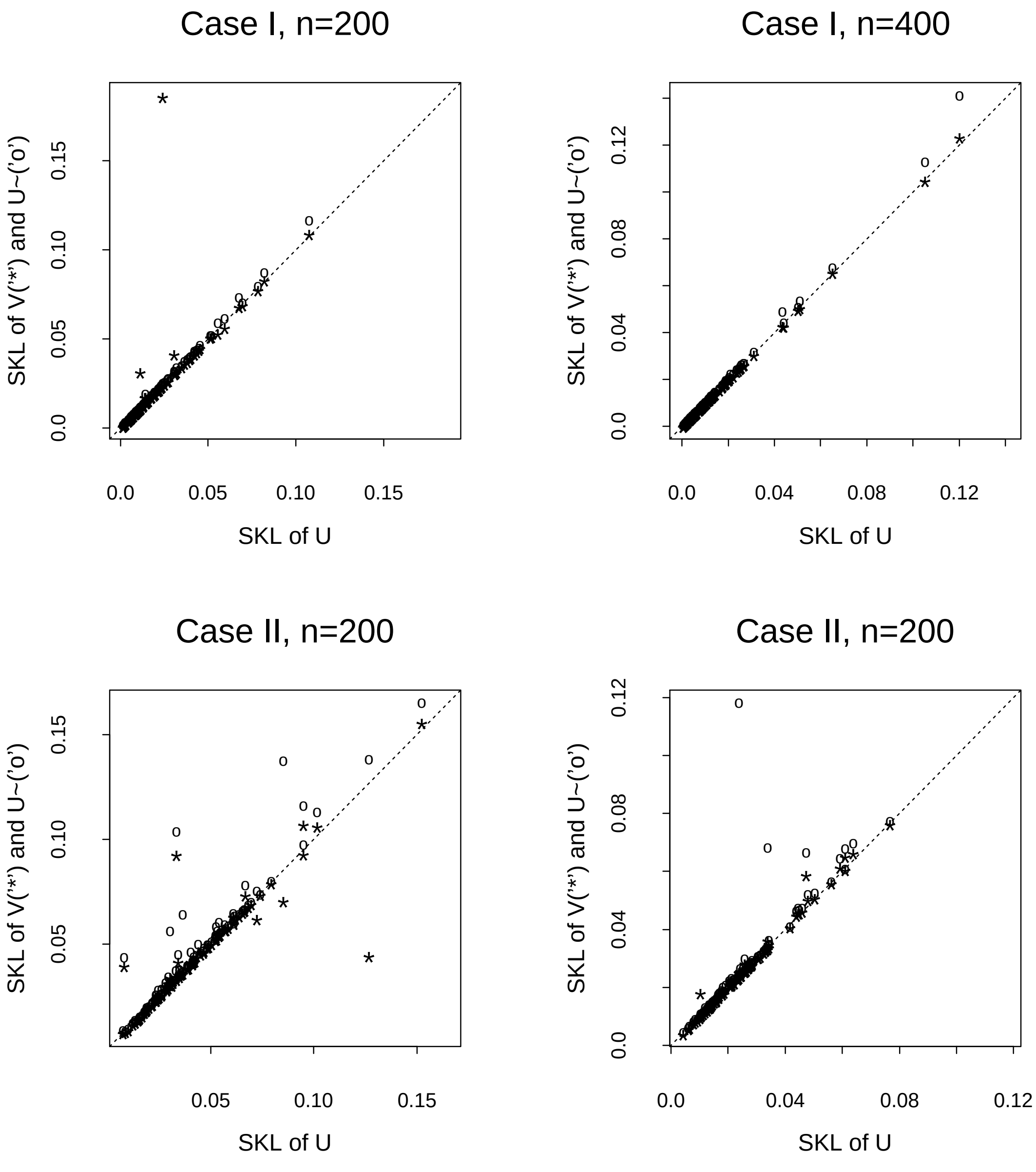

Figure 8: Comparison of option U with option V and option U˜ by SKL.

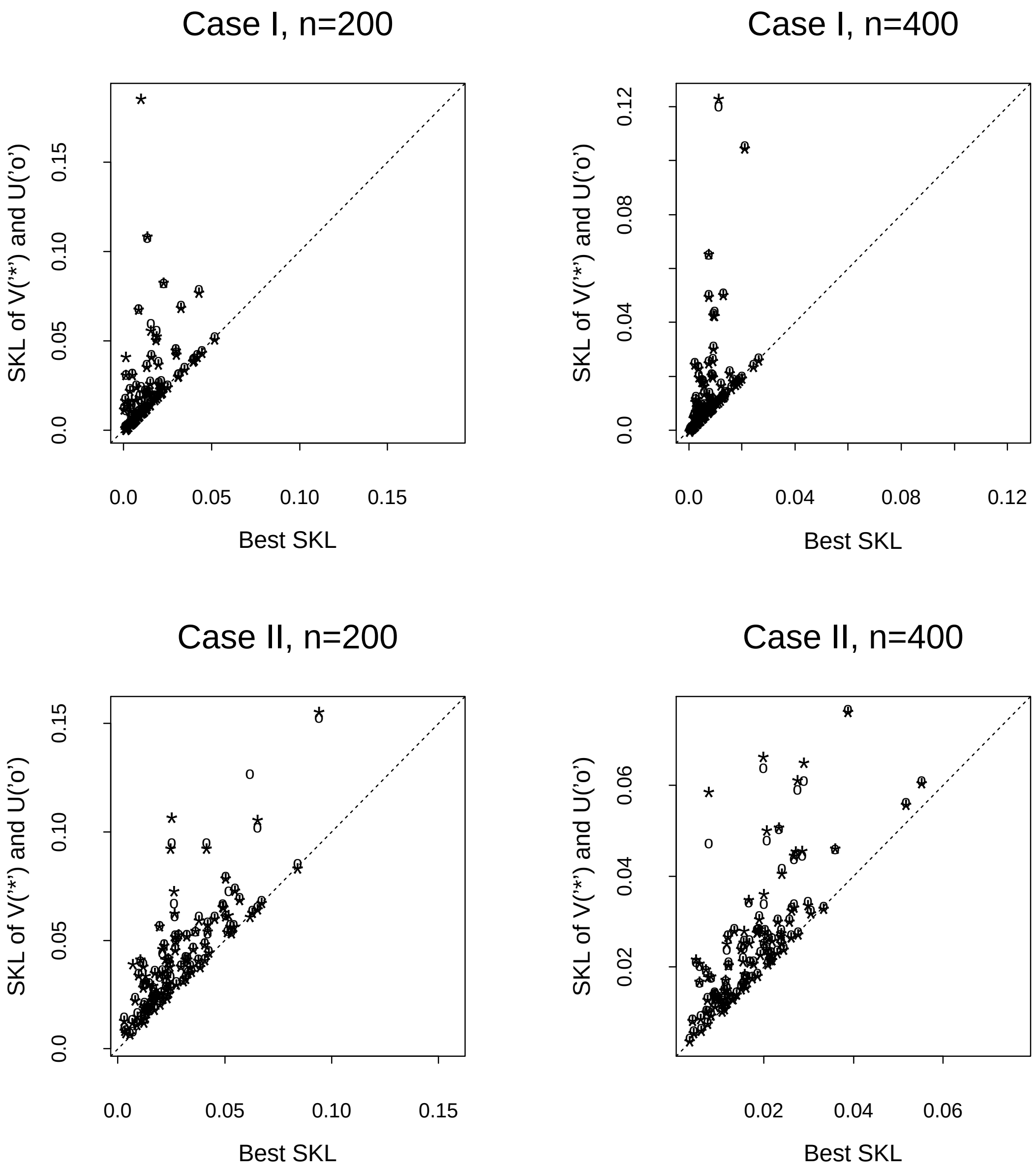

Figure 9: Comparison of option U and V with the optimum by SKL.

values of f, f_1 and f_2 (Smoothing parameters are chosen by option U). These numbers are then divided by the sample sizes to form the coverage percentage of the intervals on design points. Box-plots of the coverage of the overall and component-wise functions are shown in Figure 10 for Case I and Case II. Notice that for Case I, two components f_1 and f_2 are linear and go through the point zero, so the confidence intervals cover nearly all or none of the true function (see Gu and Wahba (1993b)). These box-plots indicate that usually the coverages of the overall functions are better than the coverage component-wise. The average coverages are not far from the nominal values but some components have big variances. "Typical" 90% Bayesian intervals for each case are shown in Figure 11 for Case I and Case II.

A Appendix

Lemma 1 *Suppose that $h(f) = f$, ϕ is known, $b \in C^2$ and b'' is uniformly bounded away from 0. If $|f_{i-} - f_i| = o(1)$ uniformly in i, then*

$$\tilde{y}_i = f_i + \epsilon_i + o_p(1),$$

where $\tilde{y}_i = f_{i-} - u_{i-}/w_{i-}$, ϵ_i has mean 0 and variance w_i^{-1}.

The *symmetrized Kullback-Leibler discrepancy* is defined as $SKL(f, \hat{f}) = \frac{1}{n}\sum_{i=1}^{n}[\mathrm{E}_{g_i}\ln(g_i/\hat{g}_i) + \mathrm{E}_{\hat{g}_i}\ln(\hat{g}_i/g_i)$, where $g_i = g(\cdot; f_i, \phi)$ and $\hat{g}_i = g(\cdot; \hat{f}_i, \phi)$.

Lemma 2 *Suppose that ϕ is known. $h, b \in C^2$. Then we have*

$$SKL(f, f_{\lambda,\Theta}) \approx \tilde{U}(\lambda, \Theta)$$

with σ^2 estimated by $\hat{\sigma}^2$.

See Wang (1994b) for proofs of these two lemmas.

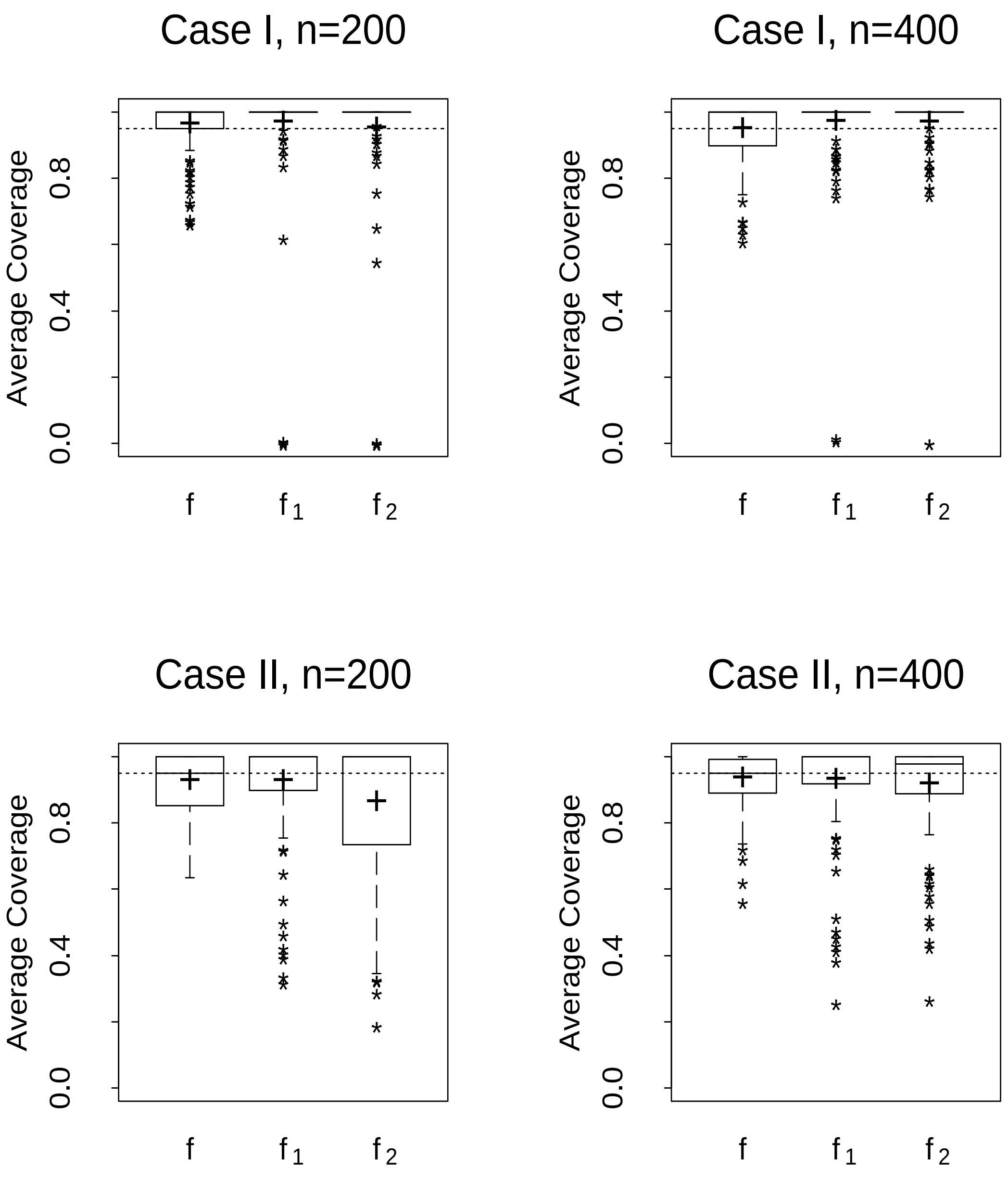

Figure 10: Coverage percentages of 95% Bayesian intervals. Pluses: sample means; dotted lines: nominal coverage.

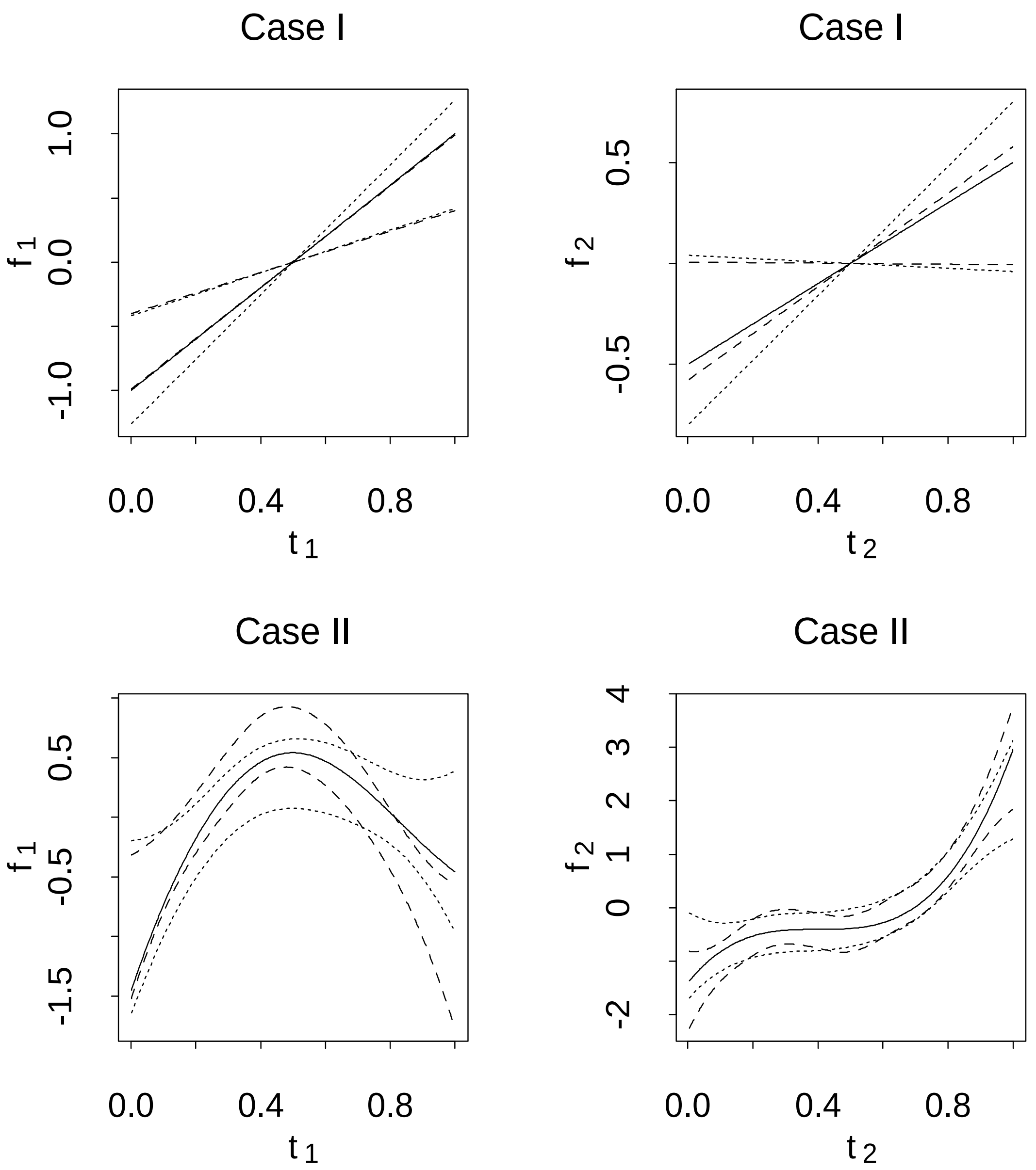

Figure 11: Display of the 90% intervals. Solid lines: true functions; dotted lines: n=200 intervals; dashed lines: n=400 intervals.

B Appendix

Let the prior for $f(\boldsymbol{t})$ be

$$F_\xi(\boldsymbol{t}) = \sum_{\nu=1}^{M} \tau_\nu \phi_\nu(\boldsymbol{t}) + b^{\frac{1}{2}} \sum_{\beta=1}^{q} \sqrt{\theta_\beta} Z_\beta(\boldsymbol{t}), \tag{13}$$

where $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_M)^T \sim N(\mathbf{o}, \xi I)$, Z_β are independent, zero mean Gaussian stochastic processes, independent of $\boldsymbol{\tau}$, with $\mathrm{E} Z_\beta(\boldsymbol{t}) Z_\beta(\boldsymbol{s}) = R_\beta(\boldsymbol{t}, \boldsymbol{s})$. With $\xi \to \infty$, Wahba et al (1994c) derived the approximate posterior mean and covariance. We list them in the following theorem.

Lemma 3 *Let $\boldsymbol{c}$ and $\boldsymbol{d}$ be a solution to (7). Let λ, Θ be smoothing parameters at convergence and W, Q_Θ be matrix based on converged values. Let $n\lambda = \hat{\sigma}^2/b$ and $M = Q_\Theta + n\lambda W^{-1}$. For $g_{0,\nu}(\boldsymbol{t}) = \tau_\nu \phi_\nu(\boldsymbol{t})$, $g_\beta(\boldsymbol{t}) = b^{\frac{1}{2}} \sqrt{\theta_\beta} Z_\beta(\boldsymbol{t})$, $\nu = 1, \cdots, M$, $\beta = 1, \cdots, q$, we have*

$$\begin{aligned}
E(g_{0,\nu}(\boldsymbol{t})|\boldsymbol{y}) &\approx d_\nu \phi_\nu(\boldsymbol{t}), \\
E(g_\beta(\boldsymbol{t})|\boldsymbol{y}) &\approx \sum_{i=1}^{n} c_i \theta_\beta R_\beta(\boldsymbol{t}, \boldsymbol{t}_i), \\
\frac{1}{b} Cov(g_{0,\nu}(\boldsymbol{s}), g_{0,\mu}(\boldsymbol{t})|\boldsymbol{y}) &\approx \phi_\nu(\boldsymbol{s}) \phi_\mu(\boldsymbol{t}) \boldsymbol{e}_\nu^T (S^T M^{-1} S)^{-1} \boldsymbol{e}_\mu, \\
\frac{1}{b} Cov(g_\beta(\boldsymbol{s}), g_{0,\nu}(\boldsymbol{t})|\boldsymbol{y}) &\approx -d_{\nu,\beta}(\boldsymbol{s}) \phi_\nu(\boldsymbol{t}), \\
\frac{1}{b} Cov(g_\beta(\boldsymbol{s}), g_\beta(\boldsymbol{t})|\boldsymbol{y}) &\approx \theta_\beta R_\beta(\boldsymbol{s}, \boldsymbol{t}) - \sum_{i=1}^{n} c_{i,\beta}(\boldsymbol{s}) \theta_\beta R_\beta(\boldsymbol{t}, \boldsymbol{t}_i), \\
\frac{1}{b} Cov(g_\gamma(\boldsymbol{s}), g_\beta(\boldsymbol{t})|\boldsymbol{y}) &\approx -\sum_{i=1}^{n} c_{i,\gamma}(\boldsymbol{s}) \theta_\beta R_\beta(\boldsymbol{t}, \boldsymbol{t}_i),
\end{aligned}$$

where $\boldsymbol{e}_\nu$ *is the* ν*th unit vector, and* $(d_{1,\beta}(\boldsymbol{s}), \cdots, d_{M,\beta}(\boldsymbol{s})) = d_\beta(\boldsymbol{s})^T$ *and* $(c_{1,\beta}(\boldsymbol{s}), \cdots, c_{n,\beta}(\boldsymbol{s})) = c_\beta(\boldsymbol{s})^T$ *are given by*

$$d_\beta(\boldsymbol{s}) = (S^T M^{-1} S)^{-1} S^T M^{-1} \begin{pmatrix} \theta_\beta R_\beta(\boldsymbol{s}, \boldsymbol{t}_1) \\ \vdots \\ \theta_\beta R_\beta(\boldsymbol{s}, \boldsymbol{t}_n) \end{pmatrix}, \tag{14}$$

$$c_\beta(\boldsymbol{s}) = [M^{-1} - M^{-1} S (S^T M^{-1} S)^{-1} S^T M^{-1}] \begin{pmatrix} \theta_\beta R_\beta(\boldsymbol{s}, \boldsymbol{t}_1) \\ \vdots \\ \theta_\beta R_\beta(\boldsymbol{s}, \boldsymbol{t}_n) \end{pmatrix}. \tag{15}$$

ACKNOWLEDGEMENTS

This research was supported by the NSF Grant DMS9121003, DMS9301511 and NIH Grants R01 EY09946, R29 EY10769, P60 DK20572, P30 HD18258.

References

Aronszajn, N. (1950). Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68**: 337–404.

Gu, C. (1989). RKPACK and its applications: Fitting smoothing spline models, *Proceedings of the Statistical Computing Section, ASA:* pp. 42–51. RKPACK is available by ftp in `lib.stat.cmu.edu/general/rkpk`.

Gu, C. (1990). Adaptive spline smoothing in non-Gaussian regression models, *Journal of the American Statistical Association* **85**: 801–807.

Gu, C. (1992). Cross-validating non Gaussian data, *Journal of Computational and Graphical Statistics* **2**: 169–179.

Gu, C. and Wahba, G. (1991). Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method, *SIAM J. Sci. Stat. Comput.* **12**: 383–398.

Gu, C. and Wahba, G. (1993a). Semiparametric ANOVA with tensor product thin plate spline, *Journal of the Royal Statistical Society B* **55**: 353–368.

Gu, C. and Wahba, G. (1993b). Smoothing spline ANOVA with componentwise Bayesian confidence intervals, *Journal of Computational and Graphical Statistics* **2**: 97–117.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall, London.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society, Serr. A* **135**: 370–384.

O'Sullivan, F., Yandell, B. and Raynor, W. (1986). Automatic smoothing of regression functions in generalized linear models, *Journal of the American Statistical Association* **81**: 96–103.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.

Wahba, G., Gu, C., Wang, Y. and Chappell, R. (1994a). Soft classification, a.k.a. risk estimation, via penalized log likelihood and smoothing spline analysis of variance, *The Mathematics of Generalization, Santa Fe Institute Studies in the Sciences of Complexity*, Vol. XX, 329-360, D. Wolpert eds,.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1994b). Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy, Technical Report 940, University of Wisconsin-Madison, Dept. of Statistics. Available by ftp in `ftp.stat.wisc.edu/pub/wahba/exptl.ssanova.ps.gz`.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1994c). Structured machine learning for 'soft' classification with smoothing spline ANOVA and stacked tuning, testing and evaluation, *Advances in Neural Information Processing 6* pp. 415–422. J. D. Cowan, G. Tesauro and J. Alspector, eds.

Wang, Y. (1994a). GRKPACK: fitting smoothing spline analysis of variance models to data from exponential families, Technical Report 942, University of Wisconsin-Madison, Dept. of Statistics. GRKPACK is available by ftp in `lib.stat.cmu.edu/general/grkpack`.

Wang, Y. (1994b). Smoothing spline analysis of variance of data from exponential families, Ph.D. Thesis, University of Wisconsin-Madison, Dept. of Statistics.

Wang, Y. and Wahba, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals, *J. Statist. Comput. Simul.* **51**: 263–279.