Relative Risk Estimation

By

Finbarr O'Sullivan[1]

[1]Research supported by
National Science Foundation under Grant No. MCS-840-3239.

Technical Report No. 76
September 1986

Department of Statistics
University of California
Berkeley, California

Relative Risk Estimation

Finbarr O'Sullivan[1]

Department of Statistics
University of California
Berkeley. CA 94720.

ABSTRACT

An alternative to the local scoring method of Hastie and Tibshirani[15] is provided for nonparametric estimation of the relative risk in the Cox model. The method involves penalized partial likelihood. Computations are carried out using a damped Newton-Raphson iteration. Each iterate is evaluated using an appropriately preconditioned conjugate gradient algorithm. The algorithm is globally convergent under mild conditions. One-step diagnostics are developed and cross validation criteria are provided to guide the evaluation of the degree of smoothness of the estimator. These cross-validation scores have potential application to model selection in standard Cox regression contexts also. Bayesian confidence intervals akin to those of Wahba[32] are defined. The preformance of the methodology is illustrated on real and simulated data.

AMS 1980 subject classifications. Primary, 62-G05, Secondary, 62-J05, 41-A35, 41-A25, 47-A53, 45-L10, 45-M05.

Key words and phrases. B-splines, Cox Proportional Hazards Model, Cross Validation, Conjugate Gradient, Diagnostics, Penalized Partial Likelihood.

Running Head: Relative Risk Estimation

August 22, 1986

[1] Research supported by the National Science Foundation under Grant No. MCS-840-3239.

Relative Risk Estimation

Finbarr O'Sullivan[1]

Department of Statistics

University of California

Berkeley. CA 94720.

1. Introduction

The proportional hazards model of Cox[6] specifies that the hazard function for the survival time of an individual with covariate vector **x** is given by

$$\lambda(t \mid x) = \lambda_0(t)\, e^{\beta' \mathbf{x}} . \tag{1.1}$$

λ_0 is the baseline hazard and the relative risk $\beta'\mathbf{x}$ is a linear model in **x** with β the vector of parameters. Since its introduction this model has become a standard tool in survival analysis, see Carter, Wampler and Stablein[5], Cox and Oakes[7] and Kalbfleisch and Prentice[17] for example. In modern data analysis environments the detection and modeling of non-linear covariate effects has become an important practicable issue. Some approaches to this problem have been proposed by Lagakos[18], Leurgans[20], Stone[29] and Tibshirani[31]. A proposal based on a penalized likelihood ideas of Good and Gaskins[13] is described here. Although I will confine myself to a single one-dimensional covariate, the methods and algorithms naturally extend to the multivariate situation.

To begin, let the conditional hazard be expressed as:

$$\lambda(t \mid x) = \lambda_0(t)\, w(x)\, e^{\theta(x)} , \tag{1.2}$$

[1] Research supported by the National Science Foundation under Grant No. MCS-840-3239.

where w is a given nonnegative weight function. The function θ is to be estimated. Including the weight function makes it possible to use the technique as a component in a more elaborate algorithm. Thus just as one-dimensional smoothers are used as components in the ACE algorithm of Brieman and Friedman[2], the back-fitting algorithm of Hastie and Tibshirani[15] and the forward stepwise algorithm of Tibshirani[31], the present technique might be used iteratively to fit additive approximations to the relative risk when there are several covariates.

Suppose there is a sample of censored survival times and corresponding covariates. Let the distinct failures occur at times $t_1, t_2, \cdots, t_n$. The number of individuals with survival times equal to t_i is m_i. Of these m_i individuals d_i fail, the remaining survival times are censored. The covariates associated with the d_i individuals are denoted $x_{i(k)}$ for $k = 1, 2, \cdots, d_i$. Treating ties in the manner suggested by Peto[27] and Breslow[1], the penalized partial likelihood estimator is defined to be the minimizer of

$$l_{n\mu}(\theta) = \sum_{i=1}^{n} d_i \log[\sum_{j \in R_{(i)}} w(x_j) e^{\theta(x_j)}] - \sum_{k=1}^{d_i} \theta(x_{i(k)}) + \mu \int [\ddot{\theta}(s)]^2 ds \quad , \quad \mu > 0 \, , \tag{1.3}$$

where $R_{(i)}$ is a *risk* set of individuals with survival times greater than or equal to t_i. The first two terms make up the negative logarithm of a partial likelihood functional, the last term is a familiar roughness penalty used in nonparametric regression, see Wahba[33]. The parameter μ controls the smoothness of the estimator. As μ is increased the estimated relative risk function is forced to be more and more linear. The estimation problem is to find θ, in an appropriate function space, to minimize the penalized partial likelihood. As $l_{n\mu}(\theta) = l_{n\mu}(\theta + c)$ for any constant c, it will only be possible to identify θ up to a constant shift. Estimates of θ will be arbitrarily adjusted so that they integrate to zero.

The negative logarithm of the partial likelihood is convex so results in O'Sullivan Yandell and Raynor[24] can be easily adapted to show that the existence of a unique (up to a constant shift) minimizer of $l_{n\mu}$ is guaranteed provided there is a unique (up to a constant shift) minimizer of the negative log-partial likelihood term in (1.3) over the set of linear functions. Asymptotic convergence properties of the estimator are worked out in O'Sullivan[25]. Rates of convergence similar to those of familiar nonparametric smoothing spline regression estimators are obtained. These results extend to bounded multivariate time dependent covariates.

An algorithm for minimizing the penalized partial likelihood functional is described in § 2. The algorithm is a damped Newton-Raphson procedure. This has the attractive property of being globally convergent under mild conditions. Significant gains in computational efficiency result from using an appropriately preconditioned conjugate gradient method, see Golub and Van Loan[12], to evaluate iterates. One step diagnostics are defined and a cross-validation criterion for assessing the degree of smoothing is provided. Approximate pointwise Bayesian confidence intervals are also developed. The techniques are illustrated on real and simulated data.

Acknowledgements

I am indebted to Professors John Crowley and Grace Wahba for stimulating my interest in this topic. Conversations with Trevor Hastie and Robert Tibshirani clarified my understanding of the Local Scoring Algorithm.

2. Fitting Procedure

2.1. B-spline Representation

The minimization of $l_{n\mu}$ is carried out over the Sobolev space of continuous functions whose second derivative is square integrable. It is shown in O'Sullivan[22] that any solution to this minimization problem is a cubic spline with knots at x_j's for which $j \in R_{(1)}$. It follows that the minimization can be carried out over the set of functions of the form

$$\theta = \sum_{r=1}^{p} \theta_r B_r \tag{2.1}$$

where $p = \#R_{(1)}+2$, B_r are cubic B-splines with interior knots at x_j for $j \in R_{(1)}$, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)$ are B-spline coefficients. A standard reference on B-splines is the book by de Boor[8], for a statistical perspective see Silverman[28]. As one might expect, the local support property of B-splines leads to great gains in computational efficiency.

Letting $\mathbf{z}_i = (B_1(x_i), B_2(x_i), \cdots, B_p(x_i))'$ and $\mathbf{z}_{i(k)} = (B_1(x_{i(k)}), B_2(x_{i(k)}), \cdots, B_p(x_{i(k)}))'$, the objective function in (1.3) becomes

$$l_{n\mu}(\boldsymbol{\theta}) = \sum_{i=1}^{n} d_i \log[\sum_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}}] - \sum_{k=1}^{d_i} \mathbf{z}_{i(k)}'\boldsymbol{\theta} + \mu\boldsymbol{\theta}'\Sigma\boldsymbol{\theta} \tag{2.2}$$

where $\Sigma_{rs} = \int \ddot{B}_r(x)\ddot{B}_s(x)dx$ for $r,s = 1,2, \cdots p$. The goal is to find $\boldsymbol{\theta} \in R^p$ to minimize this function.

2.2. Damped Newton-Raphson Algorithm

The gradient or score vector, s($\boldsymbol{\theta}$), and the Hessian, G($\boldsymbol{\theta}$), of $l_{n\mu}$ are given by:

$$\begin{aligned} s(\boldsymbol{\theta}) &= \nabla_\theta l_{n\mu}(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{n} d_i \frac{\sum_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}} \mathbf{z}_j}{\sum_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}}} - \sum_{k=1}^{d_i} \mathbf{z}_{i(k)} + 2\mu\Sigma\boldsymbol{\theta} \end{aligned} \tag{2.3}$$

$$G(\boldsymbol{\theta}) = \sum_{i=1}^{n} d_i \frac{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}} \mathbf{z}_j \mathbf{z}_j'}{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}}} - \sum_{i=1}^{n} d_i \left\{ \frac{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}} \mathbf{z}_j}{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}}} \right\} \left\{ \frac{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}} \mathbf{z}_j}{\sum\limits_{j \in R_{(i)}} w(x_j) e^{\mathbf{z}_j'\boldsymbol{\theta}}} \right\}' + 2\mu\boldsymbol{\Sigma}$$

$$= \mathbf{H}^{(1)}(\boldsymbol{\theta}) - \mathbf{H}^{(2)}(\boldsymbol{\theta}) + 2\mu\boldsymbol{\Sigma}\,. \tag{2.4}$$

Due to the local support property of B-splines, $\mathbf{H}^{(1)}(\boldsymbol{\theta}) + \mu\boldsymbol{\Sigma}$ is 7-banded. $\mathbf{s}(\boldsymbol{\theta})$ and $\mathbf{H}^{(1)}(\boldsymbol{\theta}) + \mu\boldsymbol{\Sigma}$ can be evaluated in $O(n)$ Floating Point Operations (Flops), however, since $\mathbf{H}^{(2)}$ is full the computation of the entire hessian requires $O(np^2)$ Flops.

Basic Iteration

The damped Newton-Raphson iteration, see page 501 of Ortega and Rheinbold[26], has the form

$$\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\theta}^{(l)} - \alpha_l \mathbf{G}(\boldsymbol{\theta}^{(l)})^{-1}\mathbf{s}(\boldsymbol{\theta}^{(l)}) \qquad l = 0,1,2,\cdots \tag{2.5}$$

where the l'th step size α_l is chosen to minimize the objective function in the direction $-\mathbf{G}(\boldsymbol{\theta}^{(l)})^{-1}\mathbf{s}(\boldsymbol{\theta}^{(l)})$ from $\boldsymbol{\theta}^{(l)}$ at each iteration. If $\mathbf{G}(\boldsymbol{\theta})$ is strictly positive definite for all $\boldsymbol{\theta}$ then from Theorem 14.4.3 of Ortega and Rheinbold[26] the damped Newton-Raphson iteration is globally convergent. By convexity of the penalized partial likelihood, $\mathbf{G}(\boldsymbol{\theta})$ is always positive semi-definite - it is strictly positive definite whenever there is a unique minimizer of the objective function. Thus it is easy to show that the algorithm is globally convergent under a mild condition. The theorem in Ortega and Rheinbold[26] also shows that after a finite number of steps the iteration is quadratically convergent. Exact minimization with respect to the step size does not seem to be necessary - my implementation uses step size halving similar to that described by Hopkins in BMDP[9].

Preconditioned Conjugate Gradient Technique

The $\mathbf{H}^{(2)}(\boldsymbol{\theta})$ term makes the Hessian a full matrix. $\mathbf{H}^{(1)} + 2\mu\boldsymbol{\Sigma}$ is banded and its inverse can be computed very quickly ($O(p)$ computations). Since $[\mathbf{H}^{(1)}(\boldsymbol{\theta}) + 2\mu\boldsymbol{\Sigma}]^{-1}\mathbf{H}^{(2)}(\boldsymbol{\theta})$ tends to have low rank, from Golub and Van Loan[12] a conjugate gradient procedure with $[\mathbf{H}^{(1)}(\boldsymbol{\theta}) + 2\mu\boldsymbol{\Sigma}]$ as a preconditioner is well suited for computing the direction $\mathbf{G}(\boldsymbol{\theta}^{(l)})^{-1}\mathbf{s}(\boldsymbol{\theta}^{(l)})$. The method typically converges in just two or three iterations giving an effective $O(p^2)$ procedure for computation of the Newton-Raphson direction.

2.3. Some Comments

Alternative techniques for nonparametric estimation in the Cox model have been proposed by Hastie and Tibshirani[15]. These techniques are known as *Local Scoring* and *Local Likelihood*. For exponential family models the local scoring algorithm and the standard Newton-Raphson algorithm are closely related. For the Cox model this is no longer true. Here the local scoring procedure corresponds to some form of quasi-Newton iteration. Convergence properties of this algorithm have yet to be studied.

The damping factor α_l is very important in practice. Only with the damping factor does the algorithm lose sensitivity to starting guesses. Concerns about sensitivity to starting guesses have been raised by Brillinger[3]. If damping factors were incorporated into the algorithms of Hastie and Tibshirani[14] and O'Sullivan Yandell and Raynor[24], then these methods would also be globally convergent in many cases.

3. One-Step Diagnostics and Cross-Validation Scores

Since the number of parameters is large, exact one-step diagnostics such as those of Storer and Crowley[30] are prohibitively expensive to compute. For the same reason exact one-step Newton-Raphson cross validation, see Burman[4], is not computationally feasible either. If the baseline cumulative hazard were known then one-step diagnostics and model selection cross-validation scores would be easy to define. Using these diagnostics with the baseline cumulative hazard replaced by a nonparametric estimate seems to work very well.

3.1. Baseline Cumulative Hazard Known

Let Λ_0 be the baseline cumulative hazard. If Λ_0 is known then $\boldsymbol{\theta}$ could be chosen to minimize the penalized likelihood

$$\sum_{i=1}^{n}\left[\sum_{k=1}^{m_i}\Lambda_o(t_i)w(x_{i(k)})e^{\theta(x_{i(k)})}-\sum_{k=1}^{d_i}\theta(x_{i(k)})\right]+\mu\int[\ddot{\theta}(x)]^2dx \quad , \quad \mu>0\,. \tag{3.1}$$

Here $x_{i(k)}$ for $d_i < k \le m_i$ corresponds to censored observations at t_i. Letting $c_{i(k)}=0$ if $x_{i(k)}$ corresponds to a censored observation and $c_{i(k)}=1$ otherwise, (3.1) becomes

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\Lambda_o(t_i)w(x_{i(k)})e^{\theta(x_{i(k)})}-c_{i(k)}\theta(x_{i(k)})\right]+\mu\int[\ddot{\theta}(x)]^2dx \quad . \tag{3.2}$$

The weights and adjusted dependent variates for an Iterative Reweighted Least Squares algorithm would be:

$$w_{i(k)}=w(x_{i(k)})\Lambda_0(t_i)e^{\theta^{(l)}(x_{i(k)})} \quad ,$$

$$y_{i(k)}=\theta^{(l)}(x_{i(k)})+c_{i(k)}/w(x_{i(k)})-1 \quad , \tag{3.3}$$

for $l=1,2,\cdots$. Thus the next iterate, $\theta^{(l+1)}$, would be chosen to minimize

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}w_{i(k)}[y_{i(k)}-\theta(x_{(i(k))})]^2+\mu\int[\ddot{\theta}(x)]^2dx \quad . \tag{3.4}$$

At the final iteration the predicted values $\hat{\theta}_\mu = (\hat{\theta}_\mu(x_{1(1)}), \hat{\theta}_\mu(x_{1(2)}), \cdots, \hat{\theta}_\mu(x_{n(m_n)}))'$ are

$$\hat{\theta}_\mu = \mathbf{X}'[\mathbf{X}'\hat{\mathbf{W}}\mathbf{X} + 2\mu\Sigma]^{-1}\mathbf{X}\hat{\mathbf{W}}\hat{\mathbf{y}} \tag{3.5}$$

where $\hat{\mathbf{y}}$ is the vector of final adjusted dependent variates, $\hat{\mathbf{W}}$ the diagonal matrix of final weights, and $\mathbf{X}$ is a matrix with $(i(k),j)$'th entry equal to $B_j(x_{(i(k)})$. With this

$$\hat{\theta}_\mu = \hat{\mathbf{H}}\hat{\mathbf{y}} \;, \tag{3.6}$$

where $\hat{\mathbf{H}}$ is the linearized hat matrix. As in Eubank[11] the i'th diagonal element of $\hat{\mathbf{H}}$, call it $\hat{h}_i$, is referred to as the i'th leverage value. A fast $O(n)$ algorithm for computing the vector of leverages is given in O'Sullivan[23], see also Hutchinson and de Hoog[16].

Cross Validation Scores

The ordinary cross-validation score for the final weighted least squares problem is

$$V_0(\mu) = \sum_{i=1}^{n}\sum_{k=1}^{m_i} w_{i(k)} \left\{ \frac{[y_{i(k)}^{(l)} - \theta(x_{(i(k)})]}{[\ 1 - h_i\ ]} \right\}^2 \tag{3.7}$$

and the Generalized cross-validation score is

$$V(\mu) = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m_i} w_{i(k)}[y_{i(k)}^{(l)} - \hat{\theta}(x_{(i(k)})]^2}{[\ 1 - \bar{h}.\ ]^2} \;, \tag{3.8}$$

where $\bar{h}.$ is the mean leverage. From the discussion in O'Sullivan Yandell and Raynor[24], these cross validation scores should perform well from the point of view of the weighted mean square error

$$R(\mu) = \sum_{i=1}^{n}\sum_{k=1}^{m_i} \bar{w}_{i(k)}[\hat{\theta}_\mu(x_{i(k)}) - \theta_0(x_{(i(k)})]^2 \tag{3.9}$$

where θ_0 is the true value of θ and $\bar{w}_{i(k)} = \Lambda_0(t_i)e^{\theta_0(x_{i(k)})}$.

Approximate Bayesian Confidence Intervals

Bayesian confidence intervals similar to those of Leonard[19], Silverman[28], and Wahba[32] can also be defined. The approximate 95% confidence interval for $\theta(x)$ takes the form

$$\hat{\theta}(x) + 2 \cdot \hat{\sigma}(x) \quad , \tag{3.10}$$

where $\hat{\sigma}(x) = \sqrt{\mathbf{B}(x)'[\mathbf{X}\hat{\mathbf{W}}\mathbf{X} + 2\mu\Sigma]^{-1}\mathbf{B}(x)}$ and $\mathbf{B}(x) = (B_1(x), B_2(x), \cdots, B_p(x))'$.

3.2. Baseline Cumulative Hazard Unknown

Having minimized the penalized partial likelihood a nonparametric estimator of Λ_0, such as that of Breslow[1] (a variety of others are indicated in Cox and Oakes[7]), can be computed. Substituting this into the definition for the weights in (3.3) allows one to evaluate leverages, cross validation scores, and confidence intervals. These cross validation scores might also be used to select the number of variables entering a finite dimensional Cox regression model. It is interesting that the score vector for the penalized partial likelihood is exactly the same as the score vector for the full penalized likelihood with baseline cumulative hazard replaced by the Breslow estimate corresponding to the current value of θ.

4. Two Examples

4.1. Stanford Heart Transplant Data

The technique is used to study the effect of age in the well known Stanford Heart Transplant data reported by Miller and Halpern[21]. In these data, survival times are number of months until death following a heart transplant. More than 30% of the cases are censored. There are 43 distinct age values.

The degree of smoothness is chosen to minimize the ordinary cross-validation function (formula (3.7) with the baseline cumulative hazad replaced by the Breslow estimate). The estimate and 70% confidence bands are given in Figure 4.1. The relative risk is low and fairly flat from 20 to 40 years, it climbs quickly after 40 years. A Bootstrap analysis employed by Efron and Tibshirani[10] found no significant difference between a quadratic model and a non-parametric fit (based on the local scoring algorithm) for ages less than 40 years. The Bayesian confidence intervals agree with this finding. The age distribution is markedly skewed to the left and leverage values and the width of the confidence bands are larger at the lower end of the age distribution, see Figure 4.2. The explains the poorer resolution at lower ages.

4.2. Simulated Data

A set of data, (X_i, T_i, C_i) $i = 1,2, \cdots n$, was generated as follows: The X_i are independent and identically distributed uniform [0,1] random variables. Conditional on X_i

$$T_i = \min(Y_i, Z_i) \qquad \text{and} \qquad C_i = I_{[Y_i < Z_i]} \tag{4.1}$$

where Y_i is a failure time and Z_i is a censoring time. C_i is a censoring indicator. The hazard function for Y_i is

$$\lambda(t \mid X_i) = \Lambda_0(t)\, e^{\sin(2\pi X_i)} \tag{4.2}$$

with a Weibull baseline hazard $\Lambda_0(t) = t^{.58} e^{-4}$. The censoring time Z_i is exponential, generated independently of X_i and Y_i.

A sample of $n = 157$ observations was generated with the level of censoring being such that around 33% of the observations were censored. The simulated data has been set up to have some general features in common with the heart transplant data. A comparison between the baseline survival functions is given in Figure 4.3; $\exp(-\Lambda_0(T_{(i)})$ is plotted against $\exp(-\hat{\Lambda}_0(t_{(i)})$ where $\hat{\Lambda}_0$ is the estimated baseline cumulative hazard for the heart transplant data. $T_{(i)}$ and $t_{(i)}$ are the i'th smallest survival times in the simulated and real data sets. (Since partial likelihoods only depend on the ranks of survival times, an empirical qq-plot of simulated and actual survival times would not be that interesting).

The estimated relative risk with smoothing parameter chosen to minimize the ordinary cross-validation function is graphed in Figure 4.4. 70% confidence bands are also included. Leverage values are plotted against the covariate in Figure 4.5. Not surprisingly, leverages are larger when the x-values are near the edge of the interval [0,1]. There is also a tendency for leverages to increase with survival time. The ordinary and generalized cross-validation functions and the associated loss are graphed in Figure 4.6. The scale on the horizontal axis is in units of the logarithm of the trace of the hat-matrix ($\bar{h}.$). From (3.9) the loss is

$$R(\mu) = \sum_{i=1}^{n} \Lambda_0(t_i) e^{\sin(2\pi x_i)} [\hat{\theta}_\mu(x_i) - \sin(2\pi x_i)]^2 \ . \tag{4.3}$$

The minima of the cross-validation functions are very close to the minimizer of the conjectured risk. In this particular instance, the minimizer of ordinary cross validation function is a little closer to the minimizer of the true risk.

5. Discussion

Penalized partial likelihood combined with cross-validation is a practicable method for non-parametric estimation of relative risk in the Cox model. The example shows that the procedure will identify non-linear effects when they are genuinely present.

References

1. Breslow, N.E., "Contribution to discussion of paper by D.R. Cox," *J. R. Statist. Soc.*, vol. B 34, pp. 216-217, 1972.

2. Brieman, L. and Friedman, J. H., "Estimating optimal transformations for multiple regression and correlation (with discussion)," *Journal Amer. Statist. Assoc.*, vol. 80, no. 391, pp. 580-619, 1985.

3. Brillinger, D.R., "Contribution to discussion of paper by T. Hastie and R. Tibshirani (to appear)," *J. Statist. Sci.*, 1986.

4. Burman, P., "Estimation of generalized additive models," *Ann. Statist. (to appear)*, 1986.

5. Carter, W. H., Wampler, G. L., and Stablein, D. M., *Regression Analysis of Survival Data in Cancer Chemotherapy*, Marcel Dekker, Inc., New York, 1983.

6. Cox, D. R., "Regression models and life tables (with discussion).," *J. Roy. Statist. Soc. B.*, vol. 34, pp. 187-220, 1972.

7. Cox, D. R. and Oakes, D., *Analysis of Survival Data*, Chapman and Hall, London, 1984.

8. DeBoor, C., *A Practical Guide to B-Splines*, Springer-Verlag, New York, 1978.

9. Dixon, W.J., *BMDP Statistical Software*, University of California Press, Los Angeles, 1985.

10. Efron, B. and Tibshirani, R., "Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy," *J. Statist. Soc.*, vol. 1, pp. 54-77, 1986.

11. Eubank, R.L., "The hat matrix for smoothing splines," *Statistics and Probability Letters*, vol. 2, pp. 9-14, North Holland, 1984.

12. Golub, G.H. and VanLoan, Ch.F., *Matrix Computations*, North Oxford Academic, New York, 1983.

13. Good, I. J. and Gaskins, R. A., ''Non-parametric roughness penalties for probability densities,'' *Biometrika*, vol. 58, pp. 255-277, 1971.

14. Hastie, T. and Tibshirani, R., *Generalized additive models, cubic splines and penalized likelihood*, 1986. manuscript

15. Hastie, T. J. and Tibshirani, R. J., ''Generalized additive models (with discussion),'' *J. Statist. Sci. (to appear)*, 1986.

16. Hutchinson, M. F. and deHoog, F. R., ''Smoothing Noisy Data with Spline Functions,'' *Numerische Mathematik*, vol. 47, no. 1, pp. 99-107, 1985.

17. Kalbfleisch, J. D. and Prentice, R. L., *The Analysis of Survival Time Data*, Wiley, New York, 1980.

18. Lagakos, S.W., ''The graphical evaluation of explanatory variables in proportional hazard regression models,'' *Biometrika*, vol. 68, pp. 93-98, 1981.

19. Leonard, T., ''An empirical Bayesian approach to the smooth estimation of unknown functions,'' Tech. Rep. No. 2339, Math. Research Center, University of Wisconsin-Madison, 1982.

20. Leurgans, S., ''Graphical methods for proportional hazards models,'' Proceedings of the 13'th International Biometrics Conference, 1986.

21. Miller, R. and Halpern, J., ''Regression with censored data,'' *Biometrika*, vol. 69, pp. 521-531, 1982.

22. O'Sullivan, F., ''On the analysis of some penalised likelihood estimation schemes,'' Tech. Rep. No. 726, Statistics Dept., University of Wisconsin-Madison, 1983.

23. O'Sullivan, F., ''Contribution to the discussion of paper by B.W. Silverman,'' *J. R. Statist. Soc.*, vol. B 47, no. 1, pp. 39-40, 1985.

24. O'Sullivan, F., Yandell, B., and Raynor, W. J., ''Automatic smoothing of regression functions in generalized linear models,'' *J. Amer. Statist. Assoc.*, vol. 81, no. 393, pp. 96-104, 1986.

25. O'Sullivan, F., ''Nonparametric estimation in the Cox proportional hazards model,'' Tech. Rep. No. 64, Statistics Dept., University of California-Berkeley, 1986.

26. Ortega, J. M. and Rheinbold, W. C., *Iterative Solutions of Non-linear Equations in Several Variables*, Academic Press, 1970.

27. Peto, R., ''Contribution to discussion of paper by D.R. Cox,'' *J. R. Statist. Soc.*, vol. B 34, pp. 205-207, 1972.

28. Silverman, B. W., ''Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion),'' *J. R. Statist. Soc. Series B*, vol. 47, no. 1, pp. 1-50, 1985.

29. Stone, C. J., ''Contribution to the discussion of the paper by T. Hastie and R. Tibshirani (to appear),'' *J. Statist. Sci.*, 1986.

30. Storer, B.E. and Crowley, J., ''A diagnostic for Cox regression and general conditional likelihoods,'' *J. Amer. Statist. Ass.*, vol. 80, pp. 139-147, 1985.

31. Tibshirani, R., ''Non-parametric estimation of relative risk,'' Tech. Rep. No. Orion 22, Department of Statistics, Stanford University., 1983.

32. Wahba, G., ''Bayesian confidence intervals for the cross-validated smoothing spline,'' *J. R. Statist. Soc.*, vol. 45, pp. 133-150, 1983.

33. Wahba, G., ''Cross-validated spline methods for the estimation of multivariate functions from data on functionals,'' in *Statistics: An Appraisal*, ed. H. A. David and H. T. David, pp. 205-233, The Iowa State University Press, 1984.

Figure Legends

Figure 4.1 : Non-parametric estimate of the relative risk (solid line) along with 70% confidence bands (dotted line).

Figure 4.2 : Leverage values versus age.(- denotes an uncensored value, + a censored value.)

Figure 4.3 : Estimated baseline survival probability for the Heart Transplant data versus the baseline probabilities in the simulated data.

Figure 4.4 : Non-parametric estimate of the relative risk (solid line), and 70% confidence bands (dotted lines). The true risk is also included (dashed line).

Figure 4.5 : Scaled ordinary (dashed) and generalized (dotted) cross-validation scores. The true risk (see (3.9)) is the solid line.

Figure 4.6 : Leverage values for the simulated data. (- denotes an uncensored value, + a censored value.)

Figure 4.1

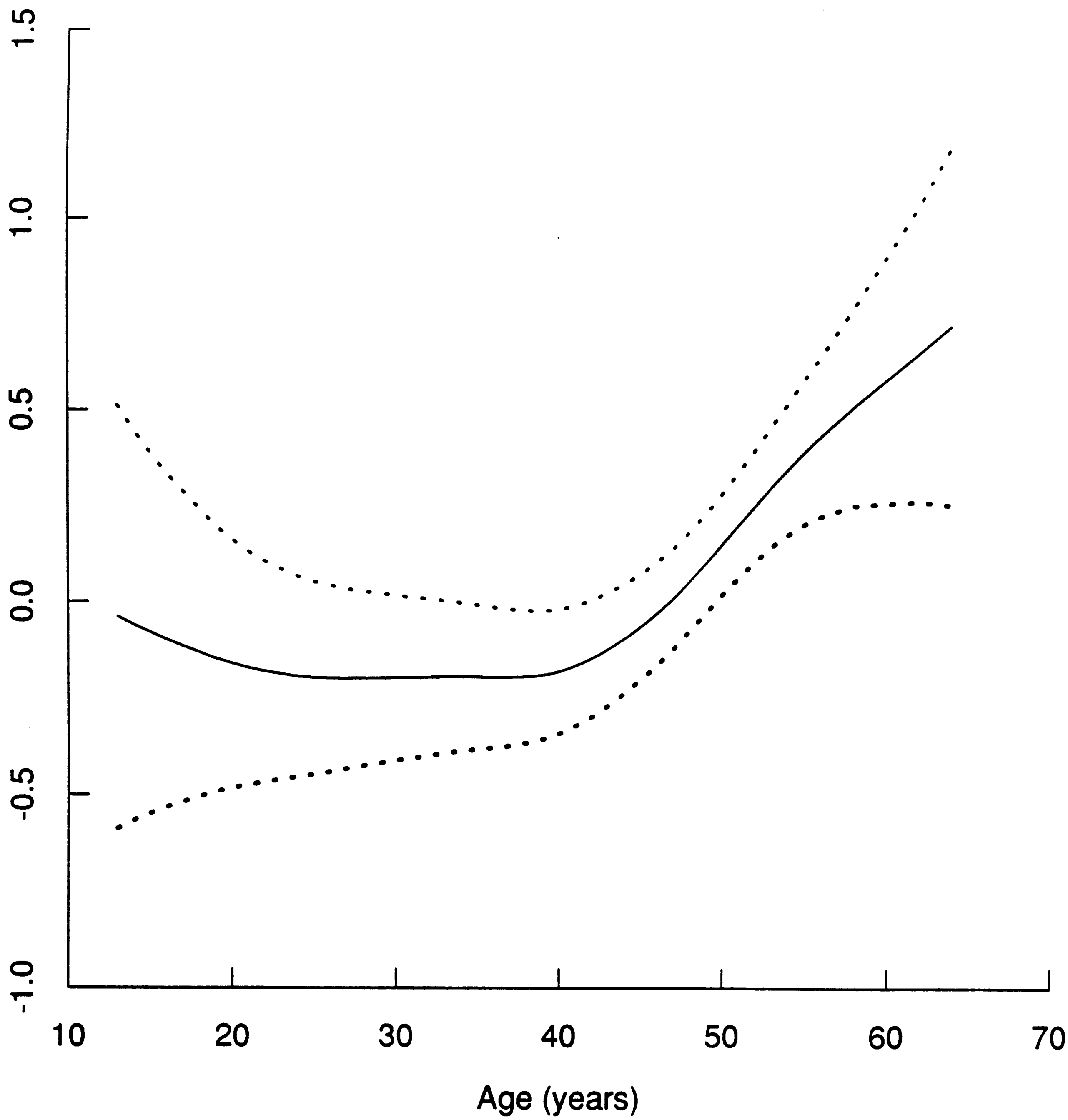


Figure 4.2

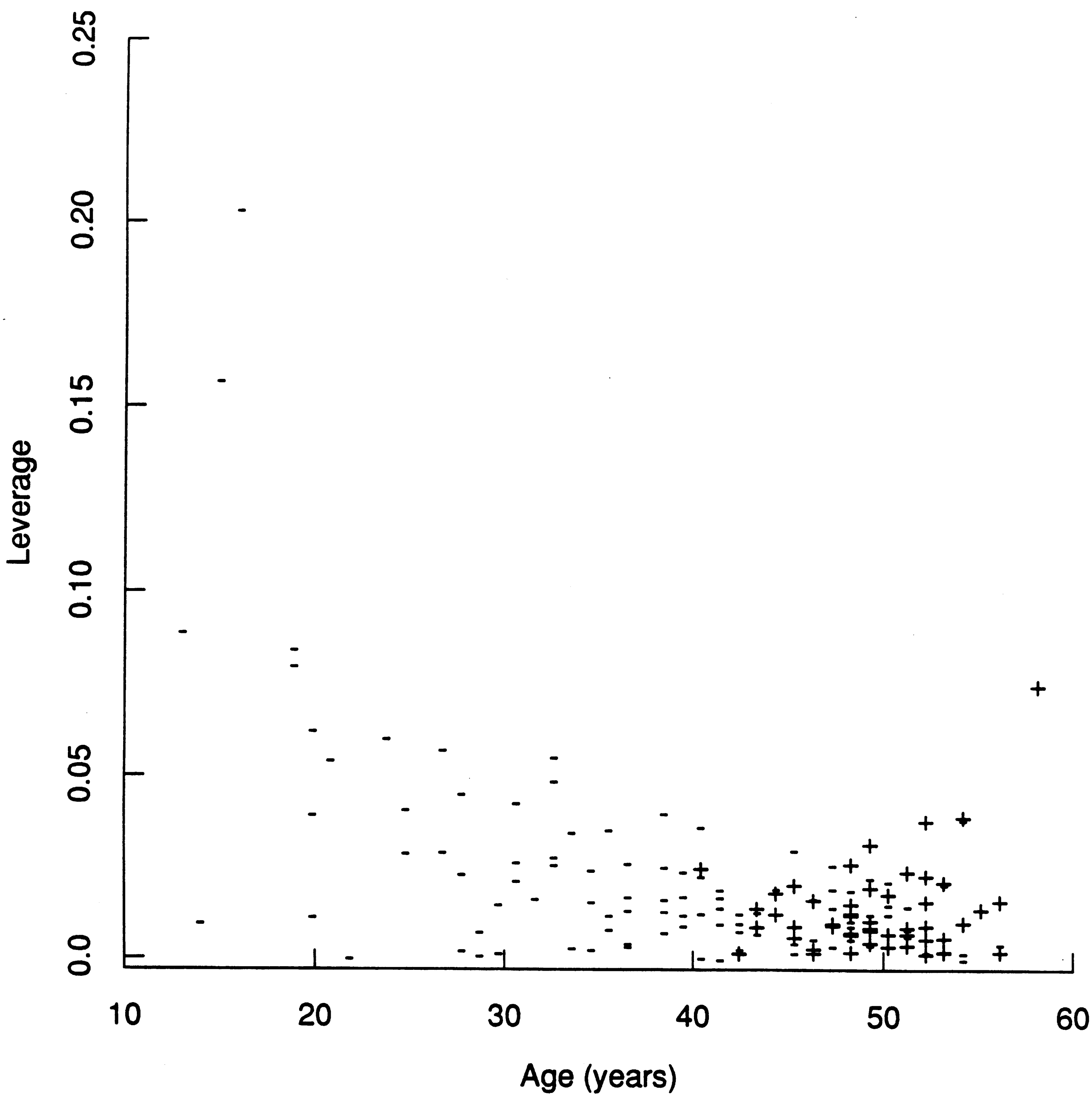


Figure 4.3

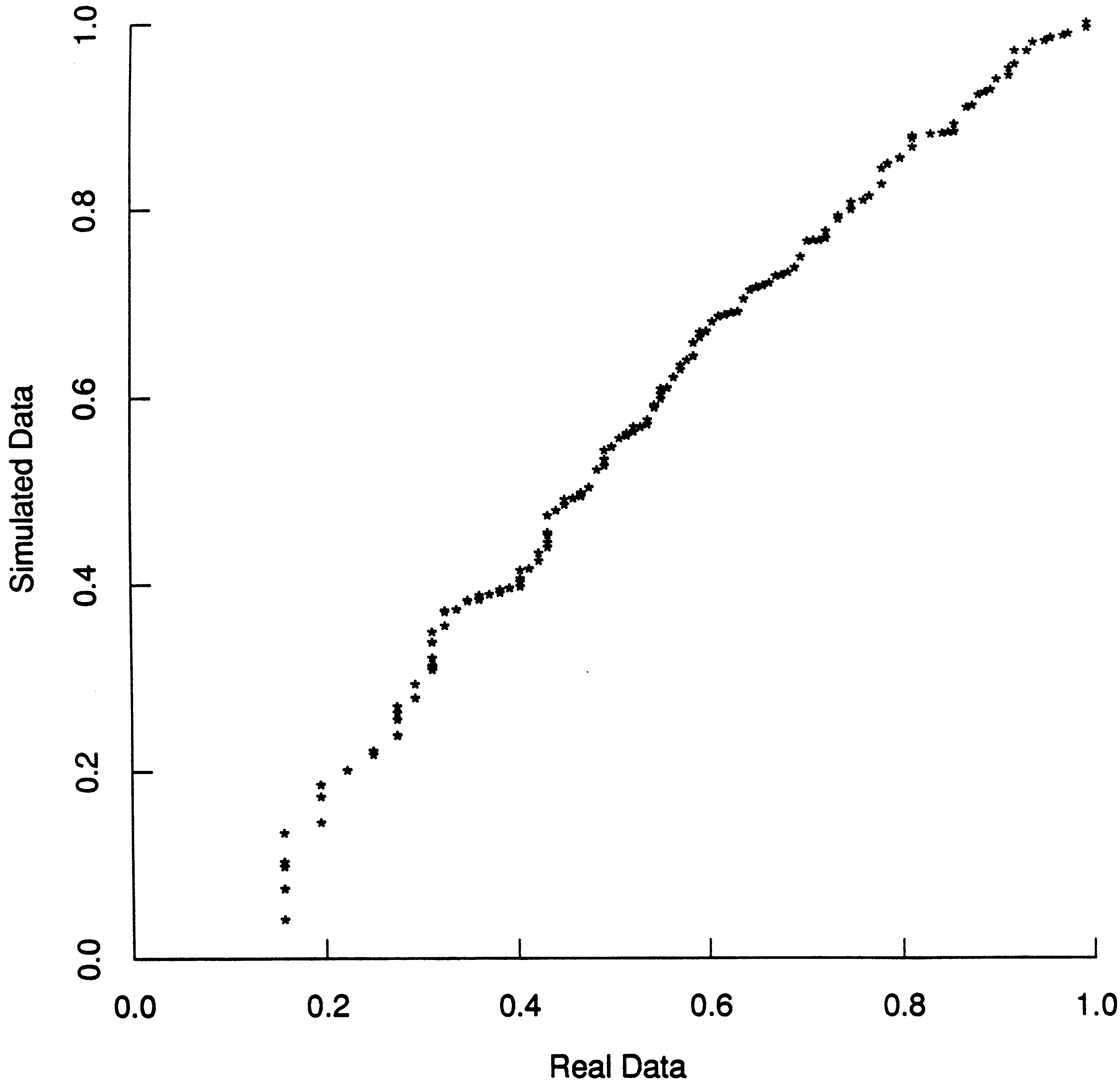


Figure 4.4

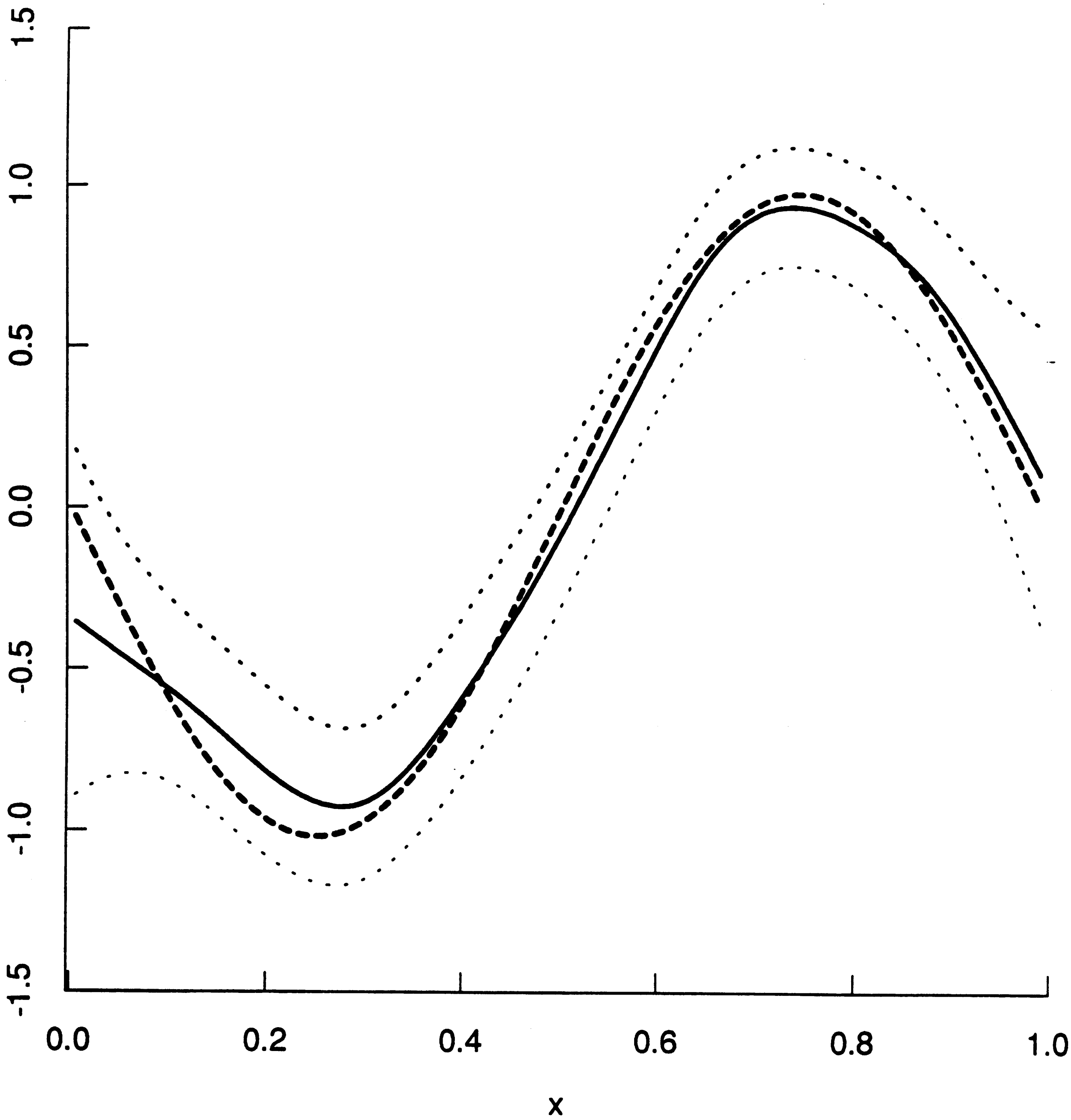
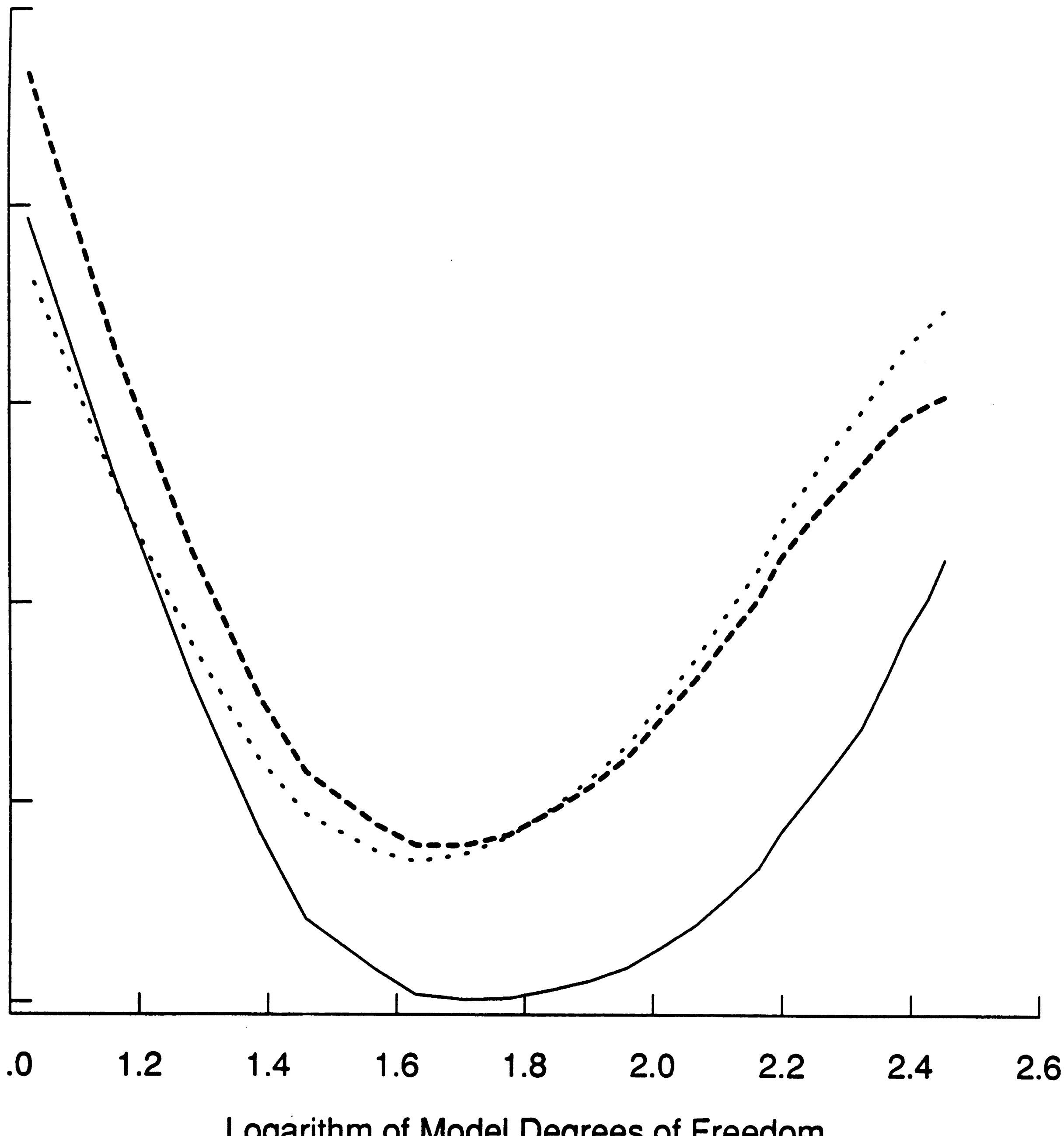


Figure 4.5



1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
Logarithm of Model Degrees of Freedom

Figure 4.6