UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPRDC TR 85-10	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Navy Personnel Research and Development Center 6b. OFFICE SYMBOL (if applicable) 41	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) San Diego, California 92152-6800	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (if applicable) Code 200	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code) Box 8e Arlington, Virginia 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: PE62763N PROJECT NO: RF63-521 TASK NO.: 806 WORK UNIT NO:

11. TITLE (include Security Classification)
A NEW HAZARD FUNCTION ESTIMATOR OF PERFORMANCE TIME

12. PERSONAL AUTHOR(S)
Bloxom, Bruce

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM Oct 83 TO Jul 84	1984 November	34

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	human performance, performance time, hazard function
05	09		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The analysis of performance time is used to study a wide variety of manpower, training, and human factors problems in the Navy. The cumulative distribution of performance time provides some indication of the rate of performance, but it can lead to false conclusions as well. One method of measuring the rate of performance with a shifting residual basis of analysis, called hazard analysis, consists of estimating a hazard function, which shows the rate of performance as a function of time. Although various procedures have been developed for estimating a hazard function, none of the procedures has been shown to produce plausibly smooth and precise estimates under a variety of conditions. This research proposed a constrained quadratic spline as an estimator of the hazard function of performance time. A maximum penalized likelihood procedure was used to fit the estimator to a sample of psychological response times. The procedure was also used in a simulation study of the precision of the estimator.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Bruce Bloxom	22b. TELEPHONE (include Area Code) (619) 225-2231 22c. OFFICE SYMBOL 41

DD FORM 1473, 84 JAN

83 APR EDITION MAY BE USED UNTIL EXHAUSTED
ALL OTHER EDITIONS ARE OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

NPRDC TR 85-10

NOVEMBER 1984

A NEW HAZARD FUNCTION ESTIMATOR OF PERFORMANCE TIME

NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

NPRDC TR 85-10 November 1984

A NEW HAZARD FUNCTION ESTIMATOR OF PERFORMANCE TIME

Bruce Bloxom

Reviewed by
Richard C. Sorenson

Approved by
Robert Penn

Released by
J. W. Renard
Captain, U.S. Navy
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152-6800

FOREWORD

This effort was conducted within exploratory development task PE62763N RF63-521-806 (Future Technologies for Manpower and Personnel). Part of this task is devoted to Techwatch, a systematic effort to monitor technological developments in a broad range of disciplines across both western and Soviet block scientific communities. The goal of Techwatch is to identify technological opportunities of potential impact on manpower, personnel, training, and human factors. Summaries of the state of technology will be prepared and published, from time to time, and new projects will be proposed to exploit particularly promising technological areas. This report deals with hazard analysis, a statistical approach to analyzing performance times. Results are intended for researchers involved in studying human performance.

J. W. RENARD
Captain, U.S. Navy
Commanding Officer

J. W. TWEEDDALE
Technical Director

SUMMARY

Problem

The analysis of performance time is used to study a wide variety of manpower, training, and human factors problems in the Navy. For example, attrition can be measured by the number of months of service before separation, trainee success can be measured by the time required to satisfactorily complete a series of training modules, and the vigilance of a radar observer can be measured by the length of time he or she can work without making response errors. In each of these contexts, the cumulative distribution of performance time provides some indication of the rate of performance, but it can lead to false conclusions as well. In the attrition example, a comparison of the percentages of high-school graduates and nongraduates at various points in their first tour of duty shows a smaller percentage of nongraduates at each point. This comparison might lead one to believe that the attrition from nongraduates is greater across the entire term of service. However, if the percentage still in the service at each point is based on the number remaining just previously, a very different result can be obtained.

One method of measuring the rate of performance with a shifting residual basis of analysis, called hazard analysis, consists of estimating a hazard function, which shows the rate of performance as a function of time. Although various procedures have been developed for estimating a hazard function, none of the procedures has been shown to produce plausibly smooth and precise estimates under a variety of conditions. Precision has been especially a problem in the upper quantiles of the distribution.

Purpose

The purpose of this research was to (a) develop a hazard function estimator that is visually smooth and (b) study its precision with various sample sizes and under various shapes of the hazard function being estimated.

Approach

The hazard function estimator proposed in this research is a quadratic spline, which is a piecewise quadratic polynomial constrained to be continuous and to have continuous first-order derivatives where the pieces are joined, points that are commonly termed "knots." The estimator function may be constrained to be either increasing or decreasing at the knots and either convex or inflected. A procedure for fitting the spline to data was proposed and its use illustrated in analysis of a set of empirical data. To examine the precision of the estimator as a function of sample size and variation in the form of the hazard function, a simulation study was conducted.

Results

When the hazard function was constrained to be isotonic or bidirectional and to have a small number of inflections, the empirical example demonstrated that a maximum penalized likelihood procedure can provide rapid computation of this kind of hazard estimator and can produce a good fit to the data. The shape of the hazard estimate in the empirical example resembled the shape of hazard estimates frequently obtained under two models in the simulation study when moderate sample sizes (75 and 150) were used. In one of those models, the true hazard function increased and then leveled off at the 75th percentile; in the other model, the true hazard function increased and then decreased to about three fourths of its maximum before leveling off.

Conclusions

The simulation study indicated that more substantial decreases (by half or more) in the tail of the true hazard function could be detected reliably by the proposed method when moderate sample sizes were used. The method, therefore, has practicality, because such decreases have been found in empirical studies as well as in analyses of parametric distributions (e.g., the inverse Gaussian distribution). The study also indicated that the proposed method could reliably detect smaller decreases in the true hazard function and could provide rather precise pointwise estimates of the tail of that function when a large sample size (N = 500) was used. The large sample size also resulted in much more accurate estimation of the hazard function across the middle range of the distribution, but bias was still perceptible. Finally, the study indicated that transforming the data to increase the dispersion may be an effective means of improving the method's precision with moderate sample sizes. Which transformation to use with which set of constraints is a subject for further study.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Background and Problem

The analysis of performance time is used to study a wide variety of manpower, training, and human factors problems in the Navy. For example, attrition can be measured by the number of months of service before separation, the success of a trainee can be measured by the time required to complete a series of training modules satisfactorily, and the vigilance of a radar observer can be measured by the length of time he or she can work without making response errors.

In each of these contexts, the cumulative distribution of performance time provides some indication of the rate of performance, but it can lead to false conclusions as well. In the attrition example, a comparison of the percentages of high-school graduates and nongraduates at various points in their first tour of duty shows a smaller percentage of nongraduates at each point. This comparison might lead one to believe that the attrition from nongraduates is greater across the entire term of service. However, if the percentage still in the service at each point is based on the number remaining just previously, a very different result can be obtained. For some groups of enlistees late in their term of service, the percentage remaining is greater for the nongraduates than for the graduates. The less educated enlistees at this point may be making more of an effort to remain, perhaps to avoid what for them is a return to an uncertain civilian job market. This finding would be missed by an inspection only of the cumulative distribution functions or other performance time summary statistics.

Measuring the rate of performance with a shifting residual basis of analysis, as just described, is called hazard analysis. The analysis consists of estimating a hazard function, which shows the rate of performance as a function of time. Although various procedures have been developed for estimating a hazard function, none of the procedures has been shown to produce plausibly smooth and precise estimates under a variety of conditions. Precision, particularly variability, has been especially a problem in the upper quantiles of the distribution.

The problem addressed by this research may be stated as follows: Assume a process that is initiated at time $x = 0$ and is completed at randomly distributed time $x = t$. Given N independently and identically distributed observations, T_i, we need to estimate the rate at which the process is completed at each point on $x > 0$. Specifically, we need to estimate the conditional density function, $r(x) = f(x)/\{1-F(x)\}$, where $f(x)$ is the unknown probability density function and $F(x)$ is the unknown cumulative distribution function. Because of its use in reliability theory (e.g., Barlow & Proschan, 1975), $r(x)$ is commonly called the hazard function.

Motivation for estimating a distribution's hazard function can be found in problems posed in psychology as well as in biostatistics, actuarial science, and engineering reliability (Lawless, 1983). In psychology, examining the hazard function of a response time distribution has been proposed as an alternative to examining the distribution's moments when one is trying to choose among alternative parametric models of decision processes (e.g., Burbeck & Luce, 1982). The hazard function has also been suggested as part of a nonparametric measure of cognitive capacity (Townsend & Ashby, 1978). In addition, comparing the hazard functions of two response time distributions has been suggested as a nonparametric device for assessing simply whether one process's response times are more rapid than another's. It has been shown that, if one distribution's hazard

function dominates another's overall x, then all quantiles of the first distribution are less than the corresponding quantiles of the second distribution (Barlow & Proschan, 1975).

Purpose

The purpose of this research was to (a) develop a hazard function estimator that is visually smooth, and (b) study its precision with various sample sizes and under various shapes of the hazard function being estimated.

APPROACH

The hazard function estimator proposed in this research is a quadratic spline, which is a piecewise quadratic polynomial constrained to be continuous and to have continuous first-order derivatives where the pieces are joined, points on x that are commonly termed "knots." Because the pieces are quadratic, inflections, such as departures from a convex form, can occur only at the knots. Also, linear constraints on the coefficients of the polynomials can be used to constrain the function to be convex or inflected and either increasing or decreasing at the knots (Wright & Wegman, 1980).

Two features of a quadratic spline suggest the ways it can provide reasonably precise hazard function estimates. The first feature is the smoothness that can be produced by linear constraints on the function. This smoothness is characteristic of hazard functions of many commonly considered parametric families of distributions: Gamma, generalized gamma, Weibull, inverse Gaussian, Rayleigh, and numerous other distributions have hazard functions that are isotonic (monotone increasing or decreasing) or bidirectional (increasing-then-decreasing or decreasing-then-increasing) functions of x and that have a small number of inflections (Glaser, 1980; Burbeck & Luce, 1982). The second feature that makes a quadratic spline a promising hazard estimator is its flexibility. Second and higher order derivatives of the spline are not defined at the knots, thus allowing the function to have rather pronounced changes in curvature at those points. With knots no closer than at every decile, a quadratic spline can rather closely approximate even the very peaked hazard function of an inverse Gaussian (Wald) distribution. Furthermore, the spline can become linear in the upper tail of the distribution as is necessary when the random variable has an exponential component (Ashby, 1982).

Numerous alternative estimators of a hazard function have been and could be developed (e.g., Lawless, 1983; Singpurwalla & Wong, 1983). A higher order polynomial spline could be employed with a penalty placed on the spline's roughness; for example, on its second and higher order derivatives (e.g., Rice & Rosenblatt, 1983). This method would clearly be appropriate when the number of peaks and inflections in the hazard function is not known or is reasonably believed to be small. Alternatively, a spline could be employed as a density estimator (e.g., Mendelsohn & Rice, 1982) that is then used to compute a hazard estimator; however, it is not clear how that hazard estimator would be constrained to be convex or isotonic (Bloxom, 1983). Finally, there have been proposals of other kinds of hazard estimators that are isotonic or bidirectional (e.g., Bartoszynski, Brown, McBride, & Thompson, 1981); however, those estimators are not constrained to have a small number of inflections.

The quadratic spline expression proposed in this research as a hazard function estimator was provided with constraints that are sufficient for the spline to be isotonic or bidirectional and convex or inflected. A procedure for fitting the spline to data was proposed and its use illustrated in analysis of a set of empirical data. To examine the

precision of the estimator as a function of sample size and variation in the form of the hazard function, a small simulation study was conducted.

RESULTS AND DISCUSSION

Hazard Estimators

A hazard function, as defined earlier, completely characterizes a distribution in the sense that the probability density function and the cumulative distribution function can be expressed in terms of the hazard function:

$$f(x) = r(x) \exp\{-\int_0^x r(u)\ du\} \tag{1}$$

and

$$F(x) = 1 - \exp\{-\int_0^x r(u)\ du\}. \tag{2}$$

If f(x) and F(x) satisfy their definitions, that is, if $f(x) \geq 0$ and $F(x) \to 1$ as $x \to \infty$, then it can be shown that $r(x) \geq 0$ for all x and $\int_0^x r(u)\ du \to \infty$ as $x \to \infty$ (Barlow & Proschan, 1975). In this research, a completely defined hazard function was defined as one with these last two properties. Substitution of this estimator into Equations 1 and 2 produces estimators of f(x) and F(x) such that $\hat{f}(x) \geq 0$ and $\hat{F}(x) \to 1$ as $x \to \infty$.

Quadratic Spline

A formulation of a quadratic spline that can provide a completely defined hazard estimator is:

$$\hat{r}(x) = \underline{b}'\ \underline{S}(x) \tag{3}$$

where $\underline{b}$ is a vector of K+3 coefficients, $\underline{S}'(x)$ is the vector $\{1, x, x^2, (x-a_1)_+^2, (x-a_2)_+^2, \text{---}, (x-a_K)_+^2\}$, $0<a_1<a_2<\text{---}<a_K$ are K knots joining K+1 quadratic pieces, $a_0 = 0$ is the lower bound of x, and the function $(x-a_j)_+ = (x-a_j)$ if $x \geq a_j$ and = 0 otherwise.

This formulation has a number of useful features. First, like r(x), the function is defined over the interval $(0,\infty)$ on x. Second, because of the one-to-one correspondence between coefficients and quadratic terms, $\hat{r}(x)$ can be smoothed to be quadratic at any knot simply by setting the corresponding coefficient equal to zero. Third, obtaining estimates of f(x) and F(x) requires integrating nothing more complicated than quadratic terms $(x-a_j)_+^2$. Fourth, obtaining hazard estimates that are isotonic, bidirectional, or convex requires nothing more complicated than placing linear constraints on the vector of coefficients, $\underline{b}$, during the estimation of that vector.

In the history of the study of splines, two concerns have been noted regarding the use of the formulation in Equation 3. One concern is that computation of Equation 3 is somewhat slower than other formulations, most notably B-splines (Shumaker, 1981),

because the number of terms being computed increases as x increases. However, in spite of this relative slowness, estimating $\hat{r}(x)$ is quite rapid, even in a maximum likelihood estimation where $\hat{r}(x)$ is computed for each data point in each iteration. A second concern is that $\underline{b}$ becomes unstable if one knot is placed very close to another knot because the corresponding terms of $\underline{S}(x)$ become collinear. However, when one is estimating a hazard function, knots do not generally need to be very close to give the spline sufficient flexibility to approximate a wide variety of hazard functions. Furthermore, even if a pair of knots should need to be placed in close proximity, instability of $\underline{b}$ does not imply instability of $\hat{r}(x)$, which is the estimate of interest.

An argument in favor of using B-splines instead of Equation 3 is that a wide variety of shapes of functions can be approximated by B-splines that (1) have adjacent knots equated, (2) have been integrated, or (3) have their coefficients constrained to be nonnegative. However, the formulation in Equation 3, as developed in this study, is sufficient for approximating hazard functions that are plausible in a wide variety of applications. Furthermore, using B-splines with nonnegative coefficients to approximate some plausible hazard functions, such as those that are bidirectional with a nonzero asymptote, requires a mixture of B-splines and integrated B-splines; that mixture may require estimating more coefficients than are needed when Equation 3 is used.

Constraints

If the quadratic spline in Equation 3 is to be a completely defined hazard estimator, linear constraints need to be placed on $\underline{b}$. The approach proposed here places a set of constraints on the spline, $\hat{r}(a_j)$, on its first derivative, $\hat{r}'(a_j)$, and on the first backward difference of its derivative, $\Delta\hat{r}'(a_j) = \{\hat{r}'(a_j) - \hat{r}'(a_{j-1})\}$, at the knots, a_j. Each constraint can be written as a linear function of the spline's coefficients; from Equation 3, these are:

$$\hat{r}(a_j) = \underline{b}'\,\underline{S}(a_j),$$

$$\hat{r}'(a_j) = \underline{b}'\,\underline{S}^*(a_j),$$

and

$$\Delta\,\hat{r}'(a_j) = \underline{b}'\{\underline{S}^*(a_j) - \underline{S}^*(a_{j-1})\},$$

where each element of $\underline{S}^*(x)$ is the first derivative of the corresponding element of $\underline{S}(x)$. Because the pieces of $\hat{r}(x)$ are quadratic between knots, the constraints also place limitations on $\hat{r}(x)$, $\hat{r}'(x)$, and $\hat{r}''(x)$ between knots.

Table 1 lists classes of constraints that form sufficient conditions for $\hat{r}(x)$ to be isotonic increasing, isotonic decreasing, bidirectional increasing-then-decreasing, or bidirectional decreasing-then-increasing. These four functional shapes are of interest here because they characterize the hazard functions of a wide variety of distributions of x (see Burbeck & Luce, 1982; or Glaser, 1980, for specific examples). For isotonic $\hat{r}(x)$, Table 1 also lists subclasses of constraints sufficient for $\hat{r}(x)$ to be positive convex (i.e., to have nonnegative second derivatives), to be negative convex (i.e., to have nonpositive second derivatives), or to have a single inflection (i.e., to have exactly one positive convex segment and one negative convex segment). For bidirectional $\hat{r}(x)$, Table 1 also lists subclasses of constraints sufficient for $\hat{r}(x)$ to be convex, have one inflection, or have two inflections.

Table 1

Constraints on Hazard Estimator by Class of Function

Class	Subclass	Constraints
Isotonic increasing	---	$\hat{r}(a_0) = 0$
		$\hat{r}'(a_j) \geq 0$, $j \leq K$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$
	Positive convex	$\hat{r}(a_0) = 0$
		$\hat{r}'(a_0) \geq 0$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq K$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
	Negative convex	$\hat{r}(a_0) = 0$
		$\Delta\ \hat{r}'(a_j) \leq 0$, $j \leq K$
		$\hat{r}'(a_K) \geq 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
	One inflection	$\hat{r}(a_0) = 0$
		$\hat{r}'(a_0) \geq 0$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq i$
		$\Delta\ \hat{r}'(a_j) \leq 0$, $i+1 \leq j \leq K$
		$\hat{r}'(a_K) \geq 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$

Table 1 (Continued)

Class	Subclass	Constraints
Isotonic decreasing	---	$\hat{r}'(a_j) \leq 0$, $j \leq K-1$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$
	Positive convex	$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq K$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$
	One inflection	$\hat{r}'(a_0) \leq 0$
		$\Delta\ \hat{r}'(a_j) \leq 0$, $j \leq i$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $i+1 \leq j \leq K$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$

Table 1 (Continued)

Class	Subclass	Constraints
Bidirectional increasing-then-decreasing	---	$\hat{r}(a_0) = 0$
		$\hat{r}'(a_j) \geq 0$, $j \leq i$
		$\hat{r}'(a_j) \leq 0$, $i+1 \leq j \leq K+1$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$
	One inflection	$\hat{r}(a_0) = 0$
		$\Delta\ \hat{r}'(a_j) \leq 0$, $j \leq i$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $i+1 \leq j \leq K$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$
	Two inflections	$\hat{r}(a_0) = 0$
		$\hat{r}'(a_0) \geq 0$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq i$
		$\Delta\ \hat{r}'(a_j) \leq 0$, $i+1 \leq j \leq m$
		$\Delta\ \hat{r}'(a_j) \geq 0$, $m+1 \leq j \leq K$
		$\hat{r}'(a_K) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
		$\hat{r}(a_K) > 0$

Table 1 (Continued)

Class	Subclass	Constraints
Bidirectional decreasing-then-increasing	---	$\hat{r}'(a_j) \leq 0$, $j \leq i-1$
		$\hat{r}'(a_i) = 0$
		$\hat{r}(a_i) \geq 0$
		$\hat{r}'(a_j) \geq 0$, $i+1 \leq j \leq K$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
	Positive convex	$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq K$
		$\hat{r}(a_i) \geq 0$, $0 < i < K$
		$\hat{r}'(a_i) = 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$
	One inflection	$\Delta\ \hat{r}'(a_j) \geq 0$, $j \leq m$
		$\hat{r}(a_i) \geq 0$, $0 < i < m$
		$\hat{r}'(a_i) = 0$
		$\Delta\ \hat{r}'(a_j) \leq$, $m+1 \leq j \leq K$
		$\hat{r}'(a_K) \geq 0$
		$\Delta\ \hat{r}'(a_{K+1}) = 0$, any $a_{K+1} > a_K$

Each set of constraints in Table 1 is sufficient for a completely defined $\hat{r}(x)$. If $\hat{r}(x)$ is isotonic and increasing, the sufficient conditions require that $\hat{r}(x)$ be nonnegative at the lower bound, a_0, and have a nonnegative slope at the lower bound and at all knots, $a_j > a_0$. Because the function is quadratic between knots, these constraints are sufficient for its slope to be nonnegative between knots as well as at them. If $\hat{r}(x)$ is isotonic and decreasing, the sufficient conditions for a completely defined estimator require that $\hat{r}(x)$ be greater than zero at a_K and have a nonpositive slope at the lower bound and at all knots, $a_j > a_0$. These constraints make the slope of $\hat{r}(x)$ nonpositive between knots as well as at them. The sufficient conditions for a completely defined $\hat{r}(x)$ also require that $\hat{r}(x)$ be linear with a nonnegative slope, for increasing $\hat{r}(x)$, or with a zero slope, for decreasing $\hat{r}(x)$, where $x > a_K$, that is, in the upper tail of the distribution.

If $\hat{r}(x)$ is bidirectional and increasing-then-decreasing, the sufficient conditions in Table 1 for a completely defined $\hat{r}(x)$ require that $\hat{r}(x)$ be nonnegative at the lower bound, have a nonnegative slope at the lower bound and at all knots a_j, $j = 1,---, i$ for some $i < (K-2)$, have a nonpositive slope at all knots a_j, $j = i+1,---,K-1$, have a flat slope and be positive at a_K, and be linear with a zero slope where $x > a_K$. If $\hat{r}(x)$ is bidirectional and decreasing-then-increasing, the conditions in Table 1 require that $\hat{r}(x)$ have a nonpositive slope at the lower bound and at all knots a_j, $j = 1,---,i-1$ for some $i < K$, be nonnegative with a flat slope at a_i, have a nonnegative slope at all knots a_j, $j = i+1,---,K$, and be linear where $x > a_K$.

Besides being sufficient conditions for a completely defined $\hat{r}(x)$, subclasses of constraints in Table 1 limit the number of inflections in $\hat{r}(x)$ by placing restrictions on the first backward difference in the slope at the knots. For example, one set of sufficient conditions for an $\hat{r}(x)$ with no inflections is the subclass of constraints for an increasing positive convex function. This subclass constrains all of the quadratic segments between knots to be positive convex by requiring the first backward difference in slope at all knots a_j, $j = 1,---,K$, to be nonnegative. Another set of sufficient conditions for an $\hat{r}(x)$ with no inflections is the subclass of constraints for an increasing negative convex function; this subclass requires the first backward difference in slope to be nonpositive at all knots, a_j, $j = 1,---,K$, and, thus, constrains the quadratic segments between knots to be negative convex.

Sufficient conditions for $\hat{r}(x)$ to have one inflection are subclasses of constraints in which the first backward difference in slope is nonnegative (or nonpositive) for all knots, a_j, $j = 1,---,i$, and is nonpositive (or nonnegative) for all knots a_j, $j = i+1,---,K$. These constraints form an $\hat{r}(x)$ that is positive convex below a_i and negative convex above a_i. Similarly, sufficient conditions for $\hat{r}(x)$ to have two inflections are subclasses of constraints that make the function positive (or negative) convex up to some knot, a_i, negative (or positive) convex between knots a_i and a_m, $i < m < K$, and positive (or negative convex above a_m.

Parameter Estimation

Two problems were encountered in an application of the hazard estimator: (a) selecting the location of the knots, a_j, and (b) selecting a function to optimize when estimating the coefficients, $\underline{b}$.

Knots

One consideration in selecting the location of the knots is that they be placed so that the true r(x) can be closely approximated by $\hat{r}(x)$. In practice, an optimal approach to this problem (e.g., de Boor, 1978) is difficult because the true r(x) is unknown. However, one can distribute a moderate number of knots, either at equal intervals (Shumaker, 1981) or at equally spaced quantiles on x (Wahba, 1976), so that knots are likely to be placed in regions where r(x) departs markedly from a quadratic polynomial form.

A second consideration in selecting the location of the knots is that there be enough data between knots to avoid severe sampling error in the pieces of the spline. Based on practical experience in fitting splines to data, Wold (1974) suggested that there be at least four or five data points between each pair of knots. Lii and Rosenblatt (1975) illustrated how extremely unstable a spline (density) estimator can become in segments where the data are sparse.

In the empirical example and in the simulations that will be described later, a knot is placed at each of the nine deciles of the sample's distribution. This placement distributes the knots widely on the distribution of x as is necessary for close approximation of a variety of forms of r(x). Also, it is an attempt to avoid severe sampling error across $\hat{r}(x)$: When the sample size is at least 50, Wold's recommended minimum number of data points is between each pair of knots. Further control of sampling error is sought by constraining $\hat{r}(x)$ to be linear above the ninth decile, a_K; this constraint is included in each set of constraints in Table 1.

Coefficients

A natural function to optimize when estimating the spline's coefficients, $\underline{b}$, is the likelihood function

$$\prod_i \hat{f}(T_i)$$

or, more practically, the log-likelihood function

$$\sum_i \ln \hat{f}(T_i).$$

One attractive feature of such an approach is the close connection, through substitution in Equation 1, between the estimator, $\hat{r}(x)$, and the function being optimized. Another attractive feature is that, if each constraint on $\hat{r}(x)$ is incorporated into a penalty function, $\Psi_p(\hat{\underline{b}})$, is added to the log likelihood. For example, in

$$g = \sum_i \ln \hat{f}(T_i) + \sum_p \Psi_p(\hat{\underline{b}}), \tag{4}$$

then optimizing g has a Bayesian interpretation. This interpretation assumes that

$$\exp\{\sum_{p} \Psi_p(\underline{\hat{b}})\}$$

is proportional to the prior joint distribution of $\underline{\hat{b}}$. Then the "penalized" likelihood,

$$G = \{\prod_{i} \hat{f}(T_i)\} \cdot \exp\{\sum_{p} \Psi_p(\underline{\hat{b}})\}$$

is proportional to the posterior joint distribution of $\underline{\hat{b}}$. From this perspective, maximizing g results in the same $\underline{b}$ as maximizing the posterior likelihood.

Using Equation 4 requires the selection of penalty functions, $\Psi_p(\underline{\hat{b}})$, but finding practical functions does not seem to be difficult. Fiacco and McCormick (1968) and Mylander, Holmes, and McCormick (1973) developed an algorithm that uses $\Psi_p(\underline{\hat{b}}) = -\Phi(\underline{\hat{b}})^2/u$ where $\Phi(\underline{\hat{b}})$ is an equality constraint; for example, $\Phi(\underline{\hat{b}}) = \hat{r}'(a_K) = 0$, and $\Psi_p(\underline{\hat{b}}) = u \ln \theta(\underline{\hat{b}})$ where $\theta(\underline{\hat{b}})$ is a nonnegativity constraint (e.g., $\theta(\underline{\hat{b}}) = \hat{r}'(a_0) \geq 0$). The constant u is chosen to be small (e.g., 2^{-16}), so that its effect on Equation 4 is negligible when the constraints are satisfied). Using these penalty functions results in estimates of $\underline{b}$ that satisfy the constraints and yield close fits to the data, as described in the numerical example (see Empirical Example).

With the penalty function approach shown in Equation 4 and linear constraints of the kind given in Table 1, maximizing Equation 4 can be accomplished by any one of a number of optimization algorithms. The algorithm by Mylander et al. (1973) was used for the numerical example and the simualtion study that follows; it employs the Gauss-Newton method of optimization that requires first- and second-order partial derivatives of the log-likelihood; from Equations 1 and 3 these are

$$\frac{\partial}{\partial \underline{b}} \{\sum_{i} \hat{f}(T_i)\} = \sum_{i} \underline{b}' \underline{S}(T_i)\}^{-1} \underline{S}(T_i) - \sum_{i} \underline{V}(T_i)$$

and

$$\frac{\partial^2}{\partial \underline{b}\underline{b}'} \{\sum_{i} \hat{f}(T_i)\} = -\sum_{i} \underline{b}' \underline{S}(T_i)\}^{-2} \underline{S}(T_i) \underline{S}'(T_i),$$

respectively, where $\underline{v}(T_i)$ is a vector of elements of $\underline{S}(T_i)$ that have been integrated over $\{0, T_i\}$. With double precision of an IBM 3033, this algorithm converges on an optimum in about 10 seconds with N = 75 and in about 35-40 seconds with N = 500.

Choosing Among Alternative $\hat{b}$

In most applications, using more than one class or subclass of constraints of the kind shown in Table 1 is plausible a priori. It then becomes important to decide empirically which constraints are most plausible. For example, in Burbeck and Luce (1982), it was important to decide empirically whether a distribution's hazard function was isotonic increasing or peaked; that is, bidirectional increasing-then-decreasing.

When alternative sets of constraints are plausible, g in Equation 4 can be optimized for each of the sets, resulting in alternative estimates of $\underline{b}$. A natural extension of this approach is to select as the most plausible estimate that $\hat{\underline{b}}$ for which g is largest. Although this approach is used in the following numerical example, the approach is problematic because the dimensionality of $\hat{\underline{b}}$ is not well defined and can vary across sets of constraints. When some of the inequality constraints placed on $\hat{\underline{b}}$ are on their boundaries, it is unclear what the effecive number of parameters is. In such a case, it is unclear how to adjust comparisons of g for variations in the dimensionality of $\hat{\underline{b}}$. Some alternatives were considered in the simulation study.

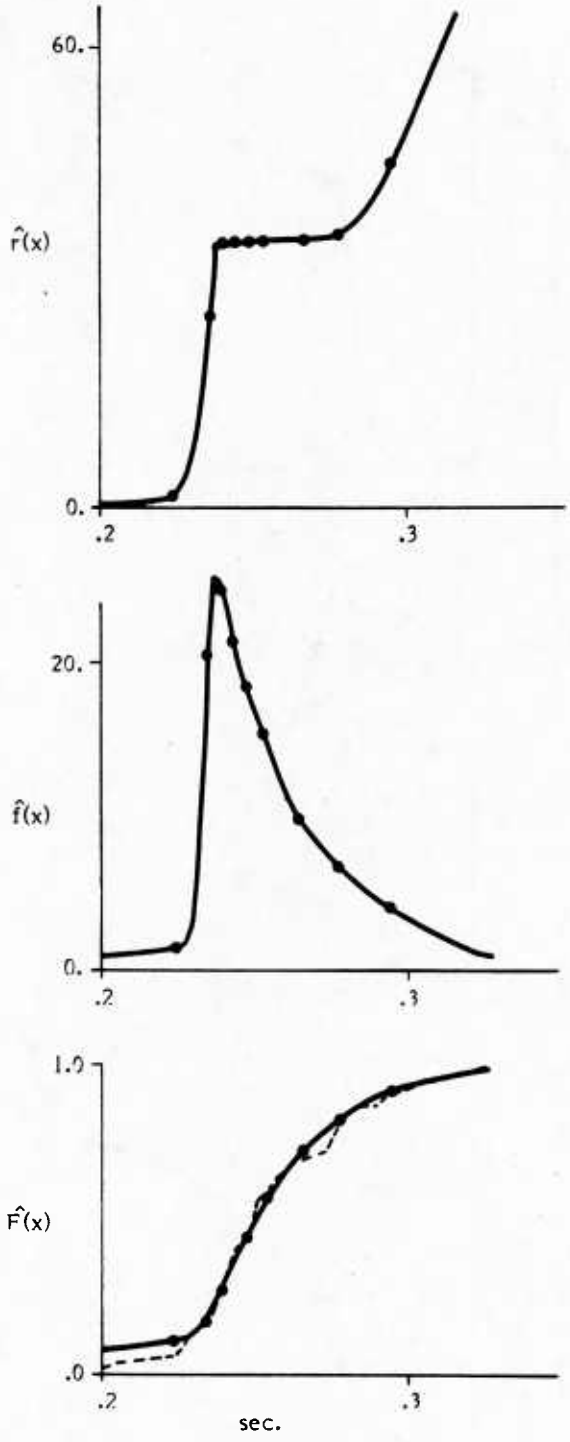
Empirical Example

To illustrate the procedure just described, response times collected by Kohfeld, Santee, and Wallace (1981) were used to obtain $\hat{r}(x)$. The data were response times collected from a single subject in a single session with a Donders multistage decision task on which the subject was highly practiced. Knots $a_1,---,a_9$ were placed at the nine sample deciles, .226, 236, .239,---, .295 seconds, respectively. The lower bound, a_0, was set equal to 0.

In analyzing these data, one question is whether the hazard function is an increasing or a peaked function of x; that is, whether the likelihood of a response increases or increases-then-decreases as the subject goes longer without responding (Burbeck & Luce, 1982). Under the constraints for an isotonic increasing $\hat{r}(x)$ in Table 1, optimizing Equation 4 yielded g = 212.40. Under the constraints for a bidirectional increasing-then-decreasing $\hat{r}(x)$, the estimate was constrained to be nondecreasing from knots a_0 to a_i, nonincreasing from knots a_{i+1} to a_8, and flat at and above a_9, the ninth decile; optimizing Equation 4 with i = 2,---,7 yielded values of g = 211.23, 211.23, 211.30, 211.27, 211.18, and 211.44 respectively. Because all of these latter values of g were less than the g obtained with the constraints for an isotonic increasing $\hat{r}(x)$, the increasing hazard function was accepted as more plausible than the increasing-then-decreasing hazard function. Figure 1 shows the isotonic increasing hazard estimate and, from substitution in Equation 1, its density and cumulative distribution function estimates (plotted with the sample cumulative step function). As can be seen from the plot of the cumulative distribution function, the fit to the data is quite close.

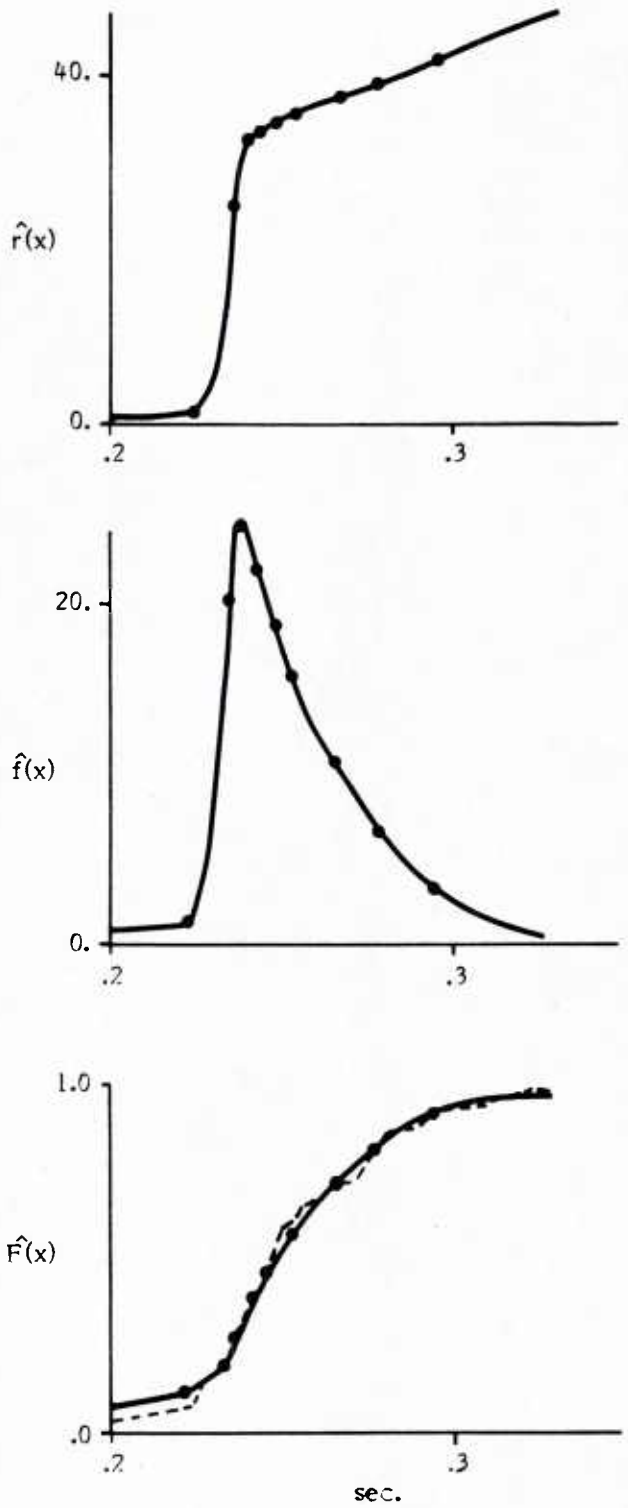
Given that the hazard function appears to be isotonic increasing, we ask where an inflection, if any, occurs in the function. The two inflections in $\hat{r}(x)$ in Figure 1 may be viewed with some scepticism because most commonly used parametric distributions with increasing hazard functions have no more than one inflection in those functions.

Three subclasses of constraints in Table 1 limit an isotonic increasing hazard estimate to no more than one inflection. Using the constraints for the positive convex or the negative convex function results in a hazard estimate with no inflection. The constraints for the one-inflection functions specify a knot, a_i, below which $\hat{r}(x)$ is positive convex and above which $\hat{r}(x)$ is negative convex. To select the most plausible estimate here, Equation 4 was optimized under the positive convex set of constraints, the negative convex set of constraints, and eight (i = 1,---,8) sets of positive-then-negative convex constraints. The largest value of g was 211.55, which was obtained under the positive-then-negative convex constraints with i = 2. Figure 2 shows this hazard estimate and its density and cumulative distribution function estimates (plotted with the sample cumulative step function). It is interesting to note that, even though the upper tail of $\hat{r}(x)$ in



- - - - Sample cumulative distribution function for 100 response times of a subject in a multistage decision task (Kohfeld et al., 1981).

Figure 1. Estimates of isotonic hazard ($\hat{r}$ (x)), density ($\hat{f}$ (x)), and cumulative ($\hat{F}$ (x)) distribution.



- - - - Sample cumulative distribution function for 100 response times of a subject in a multistage decision task (Kohfeld et al., 1981).

Figure 2. Estimates of one-inflection hazard ($\hat{r}$ (x)), density ($\hat{f}$ (x)), and cumulative ($\hat{F}$ (x)) distribution functions.

Figure 2 is much less steep than the upper tail of $\hat{r}(x)$ in Figure 1 and has no inflection, the fit of $\hat{F}(x)$ to the data in the cumulative step function is nearly as close as in Figure 1. Thus, if it is assumed that the true hazard function has only one inflection, the inflection in the upper tail of $\hat{r}(x)$ in Figure 1 may be attributable to error of estimation. This conclusion makes sense when one examines $\hat{r}(x)$ as a function of $\hat{f}(x)$ and $\hat{F}(x)$ in $\hat{r}(x) = \hat{f}(x)/\{1-\hat{F}(x)\}$, where it is seen that even a small amount of error in $\hat{F}(x)$ can produce a large error in $\hat{r}(x)$ as $\hat{F}(x) \to 1$.

Simulation Study

As shown in the preceding numerical example, the proposed method of hazard estimation can provide a close fit to a set of empirical data. However, the example does not indicate how precisely the method recovers the true hazard function. The simulation study was designed to provide a partial assessment of the precision of the method. It examines how accurately a variety of increasing and increasing-then-decreasing hazard functions are estimated with samples of the size often employed in psychological experiments.

Method

Four hazard functions (see Figure 3) define the distributions sampled in this study. The first distribution, Model A, has the hazard function

$$r_A(x) = 17.9\,x - 22.4\,x^2 - 17.5\,(x - .4)_+^2 + 79.8\,(x - .55)_+^2 - 39.9\,(x - .7)_+^2.$$

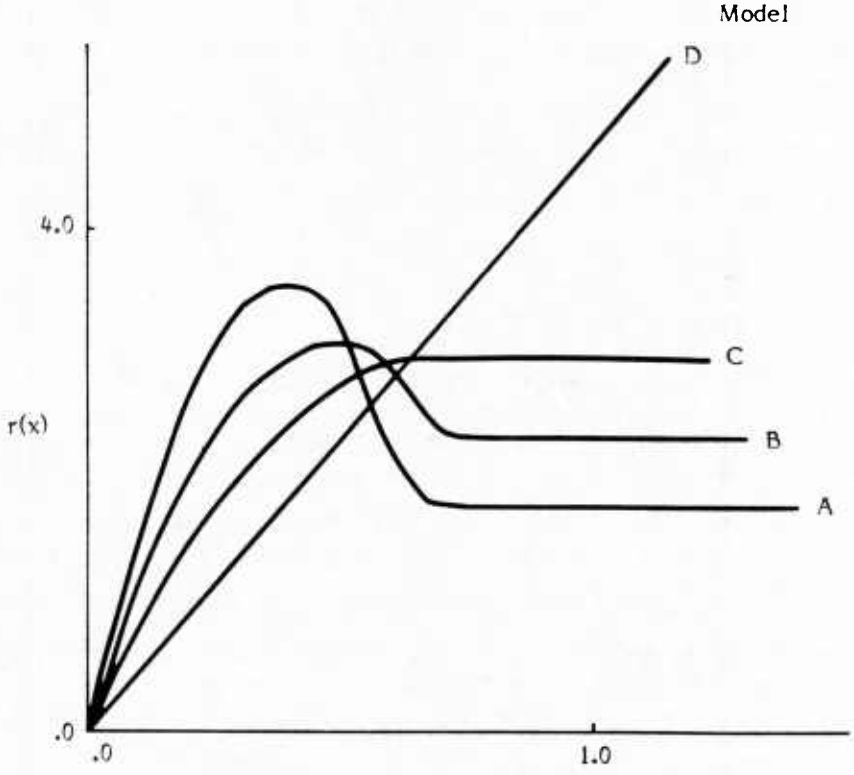


Figure 3. Hazard functions for four simulation models. Each function is shown up to the 95th percentile of the model's response time distribution.

This function has important features in common with some large-sample estimates of hazard functions obtained in Burbeck and Luce (1982): The function increases up to the sixth decile and then decreases to the eighth decile, where it approaches an asymptote that is half the maximum of the function. The second distribution, Model B, has the hazard function

$$r_B(x) = 12.5\ x - 12.5\ x_2 - 12.5\ (x - .5)_+^2 + 50.0\ (x - .625)_+^2 - 25.0\ (x - .75)_+^2.$$

This function resembles $r_A(x)$ with the exception of its asymptote, which is three fourths the maximum of the function. The third distribution, Model C, has the hazard function

$$r_C(x) = 8.9\ x - 6.8\ x^2 + 6.8\ (x - .66)_+^2.$$

This function increases to an asymptote at the distribution's 75th percentile. Its negative convex form and flat upper tail mimic the hazard functions of some parameterizations of the gamma and Weibull distributions. The fourth distribution, Model D, has the hazard function

$$r_D(x) = 4.6\ x.$$

This function is the linear hazard function of the Rayleigh distribution, which has enjoyed some popularity in reliability engineering studies. It is included here to provide a model in which r(x) is increasing in the upper tail of the distribution.

Two sample sizes, 75 and 150, were used in this study. Although sample sizes larger than these are sometimes available in psychological experiments, these represent the approximate numbers of observations often obtained in a single experimental condition.

For each of the four models and each of the two sample sizes, 20 samples were drawn using random seeds and a uniform random number generator LLRANDOMII (Lewis & Uribe, 1981). Sample values, T_i, were obtained by inverting $F(T_i) = 1 - \exp\{\int_0^{T_i} r(u)\ du\}$. Twenty replications do not provide a very precise estimate of the distribution of $\hat{r}(x)$ at each x. However, they can provide results that indicate regions of x where bias or sampling variation in $\hat{r}(x)$ is quite large.

The procedure used to estimate r(x) in each replication was identical to the procedure used in the first phase of the numerical example. Knots were placed at the nine sample deciles. The function g in Equation 4 was maximized under the constraints for an isotonic increasing $\hat{r}(x)$, using the notation g_I for this maximum of Equation 4. The function g in Equation 4 was also maximized under the constraints for an increasing-then-decreasing $\hat{r}(x)$ with a_i at knots 7, 6, 5,---. Whichever of these placements of a_i produced the largest value of Equation 4, g_{ID}, provided the $\hat{r}(x)$ that was accepted as the most plausible increasing-then-decreasing estimate. The choice between the increasing $\hat{r}(x)$ and the increasing-then-decreasing $\hat{r}(x)$ was based on the contrast, $c = g_I - g_{ID}$. As in the numerical example, the increasing $\hat{r}(x)$ was chosen if $c \geq 0$; the increasing-then-decreasing $\hat{r}(x)$ was chosen otherwise. Two other criteria for c were tried and were found to select the correct set of constraints less accurately.

Three methods were used to assess the precision of the hazard estimates for each model and sample size. One method was to examine the proportion of replications in which $\hat{r}(x)$ was correctly chosen as increasing (IHF) or increasing-then-decreasing (IDHF). The true hazard functions in Models A and B are IDHF; in Models C and D, they are IHF. The second method was to examine the integrated square discrepancy between $\hat{r}(x)$ and $r(x)$ over a finite range on x. When averaged across replications, this measure is the sample analogue of the mean integrated square error (MISE) sometimes used in studies of asymptotic properties of function estimators (e.g., Kronmal & Tarter, 1968; Hall, 1982). The third method of assessing the precision of $\hat{r}(x)$ was to plot the mean of the 20 $\hat{r}(x)$'s as a function of x. Also, a high- and low-order statistic of the $\hat{r}(x)$'s was plotted at each of 10 points on x. In contrast to the first two methods, this method was intended to indicate approximately how bias and sampling variation in $\hat{r}(x)$ change across the range of x.

Results

The first section of Table 2 shows the proportion of replications in which $\hat{r}(x)$ was correctly classified as IHF or IDHF when $c = g_I - g_{ID}$ was compared with zero. For Models A and B, this is the proportion of replications in which $c < 0$; for Models C and D, this is the proportion of replications in which $c \geq 0$. The proportion was quite high for all models except Model B, where almost half of the $\hat{r}(x)$ were incorrectly chosen to be IHF. It is not surprising that the error rate for Model B is greater than for Model A, because $r_B(x)$ decreases much less than $r_A(x)$ and thus more resembles an IHF function. It is somewhat surprising that the error rate for Model C is so low, considering that $r_C(x)$ is level above the 75th percentile. If $\hat{r}(x)$ were unbiased, one would expect that sampling error would result in some of the $\hat{r}_C(x)$ being IDHF. However, other results, to be reported below, suggest that $\hat{r}(x)$ has a bias that tends to make it IHF.

Table 2

Results of the Use of Three Log-Likelihood Criteria: Proportions of Replications That Correctly Classify the Hazard Estimate

Criterion for $c = g_I - g_{ID}$	Sample Size	Proportions by Simulation Model			
		Increasing-Then-Decreasing		Increasing	
		Model A	Model B	Model C	Model D
Zero	75	1.00	.60	.95	1.00
	150	1.00	.55	1.00	1.00
Akaike (1974)	75	--	1.00	.15	--
	150	--	.95	1.00	--
Schwartz (1978)	75	--	1.00	.00	--
	150	--	1.00	.10	--

Note. N = 20 replications for each model under each sample size.

Table 2 also shows the proportion of replications in which $\hat{r}(x)$ was correctly classified when c was compared with two criteria other than zero (Akaike, 1974; Schwartz, 1978). Both criteria take into account that g_I and g_{ID} are optimized under different effective numbers of parameters. When g_I is optimized, the slope of $\hat{r}(x)$ is not constrained to equal zero in the upper tail. Thus, obtaining g_I requires estimating one more parameter than does obtaining g_{ID}, under which $\hat{r}(x)$ is constrained to have a zero slope in the upper tail. Because of this difference between g_I and g_{ID}, a criterion proposed by Akaike (1974) indicates that $\hat{r}(x)$ should be chosen as IHF if $c > 1.00$ instead of zero. A Bayesian criterion proposed by Schwartz (1978) indicates that $\hat{r}(x)$ should be chosen as IHF if $c \geq .5 \log N$. The proportion of replications in which these criteria resulted in the correct classification of $\hat{r}(x)$ was quite high for Model B using each of these criteria. However, both criteria quite frequently misclassified $\hat{r}_C(x)$ as IDHR. Even the high proportion of correct classifications obtained using Akaike's criterion with 150 observations was not sustained with a sample size of 500 (see Additional Analyses). Thus, it appears that these two criteria are not necessarily improvements over the zero criterion for c in the present context, although they have been found to be quite useful in contexts where inequality constraints are not used (e.g., Neftci, 1982).

In addition to examining the proportion of $\hat{r}(x)$ that were correctly classified, the precision of each $\hat{r}(x)$ was measured by the integrated square discrepancy (ISD),

$$ISD = (sy)^{-1} \int_0^y \{\hat{r}(x) - r(x)\}^2 \, dx$$

where y is the 95th percentile on the true F(x). Although both $\hat{r}(x)$ and r(x) are defined over $\{0,\infty\}$, that range of integration is not used here because the linear tails of those functions would result in $ISD \to \infty$ as $y \to \infty$ whenever $\hat{r}(x) \neq r(x)$ in the upper tail. The scale factor,

$$s = y^{-1} \int_0^y \{r(x)^2\} \, dx,$$

was chosen to adjust ISD for differences in the magnitude of r(x) across models.

Table 3 gives the mean ISD for the 20 replications under each model and sample size. Not surprisingly, the larger sample size resulted in a more accurate estimate (i.e., a smaller mean ISD) under each model, although it is not much more accurate in the case of the IHF models, C and D. Also, accuracy was greater when r(x) was lower in the upper tail: Model A has the lowest tail; Model D has the highest tail.

To further examine the precision of the hazard estimates, the mean $\hat{r}(x)$ for each model and sample size was plotted as a function of x. Figure 4 shows these results as broken lines, plotted along with the corresponding true r(x) as solid lines up to the 95th percentile of x. Included on each graph are vertical lines connecting the 3rd and 18th order statistics (14th and 86th prcentiles) of the 20 $\hat{r}(x)$; these are given at the 5th, 15th, 25th,---, 95th percentiles on x. These results suggest, not surprisingly, that positive bias and sampling variability are increasing functions of r(x) and decreasing functions of N. The results also show a pronounced positive bias occurring in the upper tail of Models B, C, and D, consistent with the finding that these models have the highest mean ISD. This bias is also consistent with the frequency of $g_I > g_{ID}$ for these models. Finally, there is a

Table 3

Mean Integrated Square Discrepancy (ISD) of the Hazard Estimate by Simulation Model

Sample Size	Mean ISD			
	Model A	Model B	Model C	Model D
75	.085	.177	.199	.267
150	.053	.081	.120	.199

Note. N = 20 replications for each model under each sample size.

visually small but reliable negative bias in the lower tail for each model and sample size. This result, considered along with the positive bias in the upper tail, suggests that the sample quantiles may be less dispersed on x than are the corresponding population quantiles for the rather broad class of distributions sampled here.

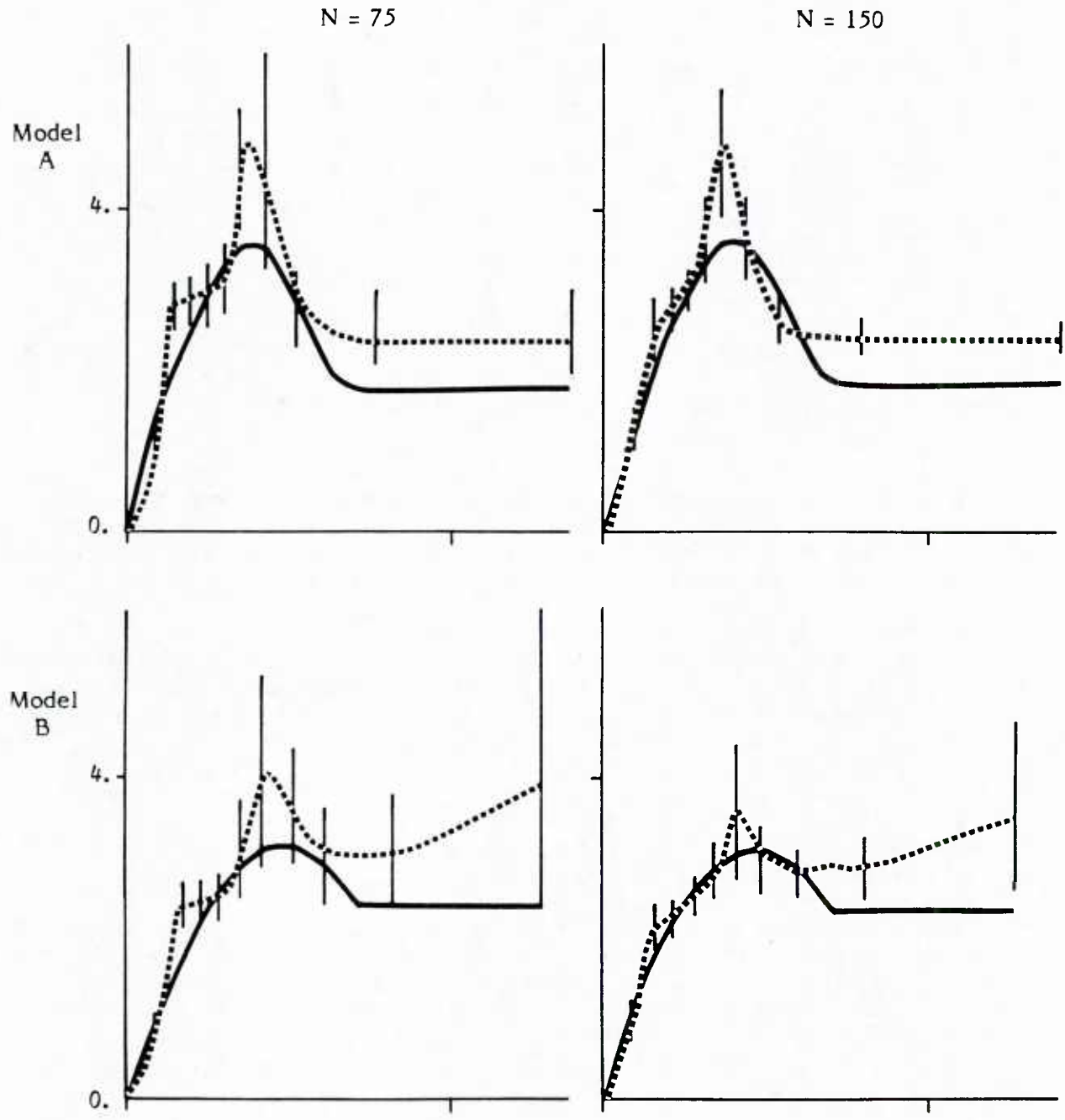
An important question in this analysis is whether the proposed $\hat{r}(x)$ is sufficiently precise to provide useful estimates, given the sample sizes employed here. The results in Figure 4 do not make the procedure look generally promising, given a sample size of 75. However, fairly accurate detection of a decreasing r(x) in the upper tail was obtained for one model, A, with this sample size. That model has an r(x) that resembles many large-sample empirical estimates obtained by Burbeck and Luce (1982) and that decreases as much as the hazard function of some parameterizations of the often-proposed Wald, or inverse Gaussian, distribution. Thus, with a sample size as small as 75, the proposed procedure may be useful for detecting a substantially decreasing r(x) even if it is not useful for precise estimation of that function.

Given a sample size of 150, Figure 4 suggests that the proposed $\hat{r}(x)$ can provide fairly unbiased estimation up through the eighth decile when r(x) is IHF. However, when r(x) is IDHF, the estimates tend to be incorrectly classified as IHF if r(x) decreases only slightly (Model B), and are markedly biased and variable at the peak of r(x) if r(x) decreases more substantially (Model A). When r(x) decreases substantially, $\hat{r}(x)$ provides a fairly good estimate of where the peak of r(x) occurs on x, but it does not provide a reliable estimate of the height of that peak.

Additional Analyses

Although the preceding analyses provide useful results, three additional analyses were conducted to probe the effects of (a) using a substantially larger sample size, (b) transforming the data to increase the dispersion on x, and (c) constraining r(x) to have no more than one inflection (as in the empirical example).

To probe the effect of using a larger sample size, 20 replications with N = 500 were obtained for Models B and C. The sets of constraints on $\hat{r}(x)$ were the same as those used in the simulation study. For Model B, $\hat{r}(x)$ was correctly classified as IDHF in 90 percent



———— Hazard function.

- - - - Mean hazard estimate.

| Variation of hazard estimate.

Figure 4. Hazard function, mean hazard estimate, and variation of hazard estimate for four simulation models (N = 75, 150).

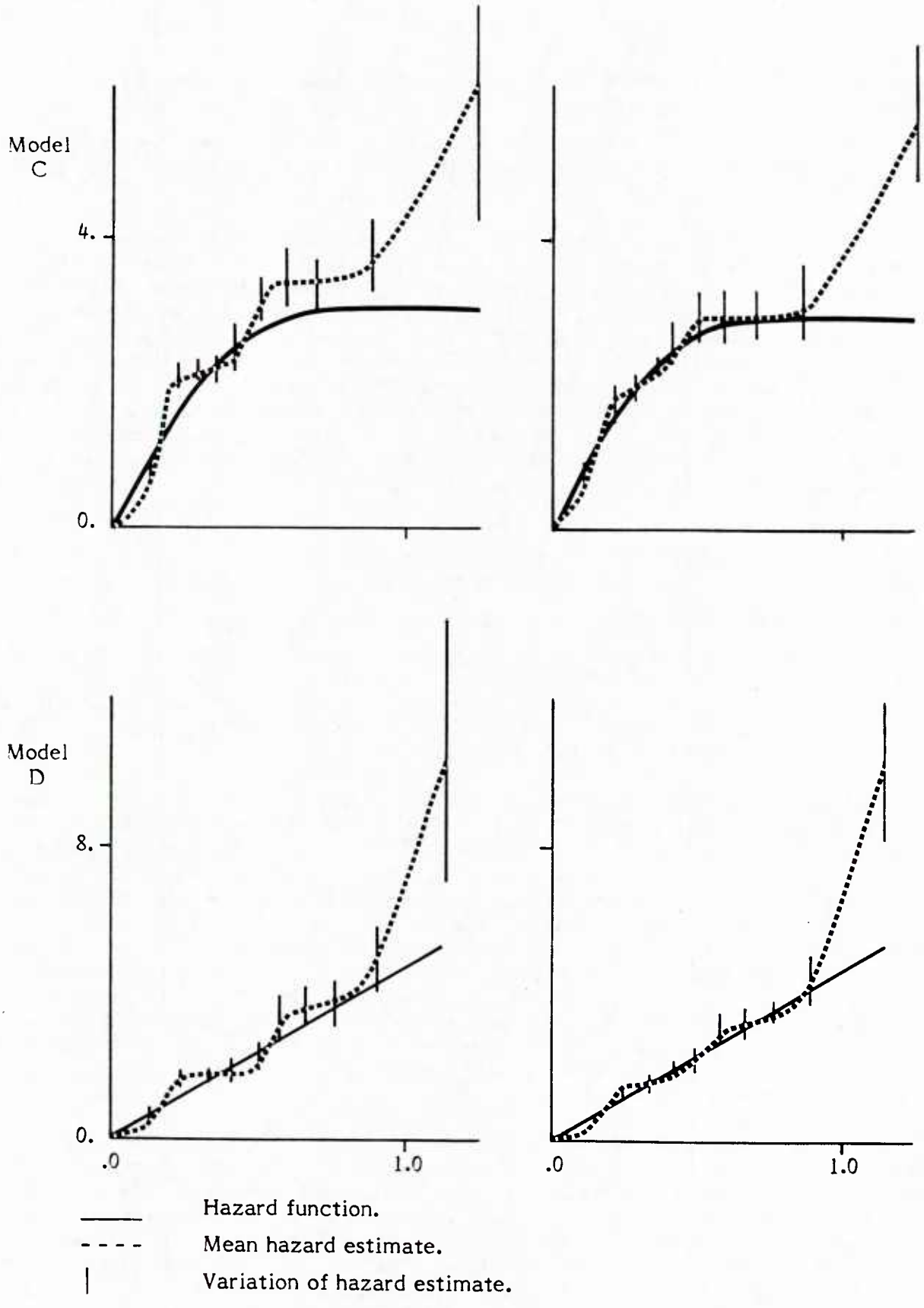


Figure 4. Continued.

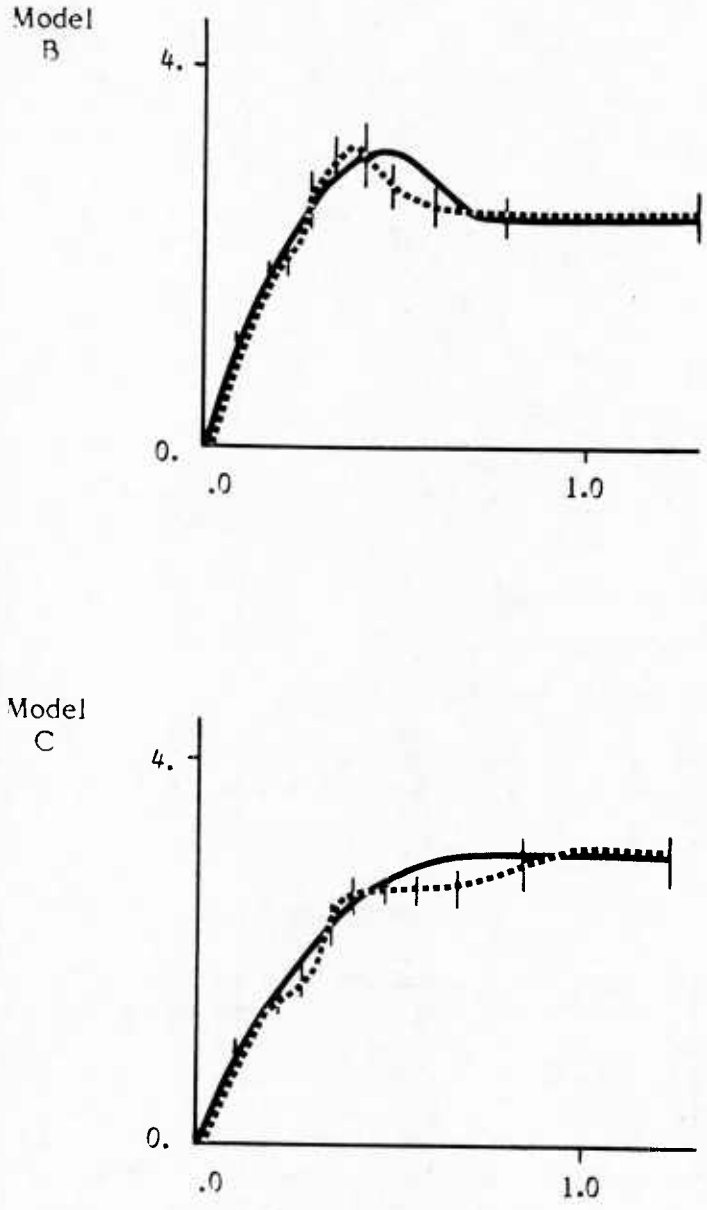
of the replications; for Model C, $\hat{r}(x)$ was correctly classified as IHF in 95 percent of the replications. The use of Akaike's (1974) and Schwartz's (1978) criteria for $c = g_I - g_{ID}$ resulted in all $\hat{r}(x)$ being incorrectly classified for Model C. These results lend further support to the use of the zero criterion for c when choosing between IHF and IDHF constraints.

Both the global and pointwise measures of precision indicate that $\hat{r}(x)$ was much improved when the substantially larger sample size was used. The mean ISD indices were .009 and .005 for Models B and C respectively, compared with values of .081 and .120 obtained with samples of 150. Figure 5 shows r(x), the mean $\hat{r}(x)$ and the 3rd and 18th order statistics of $\hat{r}(x)$ for the 5th, 15th,---, 95th percentiles on x. The larger sample size substantially improved the precision of $\hat{r}(x)$ in the upper tail, resulting in the smaller mean ISD as well as the better overall visual proximity to r(x). However, using the larger sample size did not completely remove what appears to be a positive bias in $\hat{r}(x)$ in the middle of the distribution and a negative bias between the middle and each of the tails. Further research is needed to ascertain whether the bias is due to the nature of the constraints or due to the bias that is nearly always present in the order statistics (Johnson & Kotz, 1970).

To probe the effect of increasing the dispersion on x, the 40 replications with N = 75 from Models B and C were transformed by replacing each order statistic, $T_{(i)}$, with a point halfway between the adjacent order statistics, $T_{(i-1)}$ and $T_{(i+1)}$. This transformation tends to shift each data point in the direction of lower local density. In a unimodal distribution, it tends to shift data toward the tails. In estimating $\hat{r}(x)$ from the transformed data, the sets of constraints were the same as those used in the simulation study reported above. Also, for each replication, the zero criterion for c was used in choosing between the IHF and IDHF constraints.

All measures of precision indicated that $\hat{r}(x)$ was at least somewhat improved by transforming the data. For Model B, $\hat{r}(x)$ was correctly classified as IHF in 85 percent of the replications; for Model C, $\hat{r}(x)$ was correctly classified as IDHF in 90 percent of the replications. The first of these percentages is an improvement over the 60 percent accuracy obtained with the untransformed data (see Table 1). The mean ISD indices were .096 and .168 for Models B and C respectively; each of these is better than the corresponding indices, .177 and .199, obtained with the untransformed data. An inspection of the pointwise distribution of $\hat{r}(x)$ indicated that, for both models, transforming the data reduced both the bias and variability of the estimator in the upper tail of x but had little effect on precision elsewhere. Even so, the improvement in the upper tail is encouraging, because that is where $\hat{r}(x)$ was least precise before the transformation. A modification of the transformation, say, by differentially weighting $T_{(i-1)}$ and $T_{(i+1)}$, may improve the precision still further.

To assess the effect of constraining $\hat{r}(x)$ to have no more than one inflection, the 20 samples of 150 from Model C were used to obtain $\hat{r}(x)$ under limited-inflection constraints, as in the second part of the empirical example described previously. The mean ISD for these estimates was .060, compared with the mean ISD of .120 obtained with the same data when $\hat{r}(x)$ was constrained only to be isotonic increasing. Figure 6 shows r(x), the mean $\hat{r}(x)$ and the order statistics of $\hat{r}(x)$ in the manner depicted in Figures 4 and 5. When Figure 6 is compared with the results in Figure 4 for Model C with N = 150, it is clear that constraining $\hat{r}(x)$ to have no more than one inflection can increase the precision in the upper tail. However, this increase in precision is at the expense of additional positive bias in $\hat{r}(x)$ between the seventh and eighth deciles. Thus, although a less inflected $\hat{r}(x)$ has a more plausible visual smoothness, it is not necessarily more precise pointwise than an estimate constrained only to be isotonic.

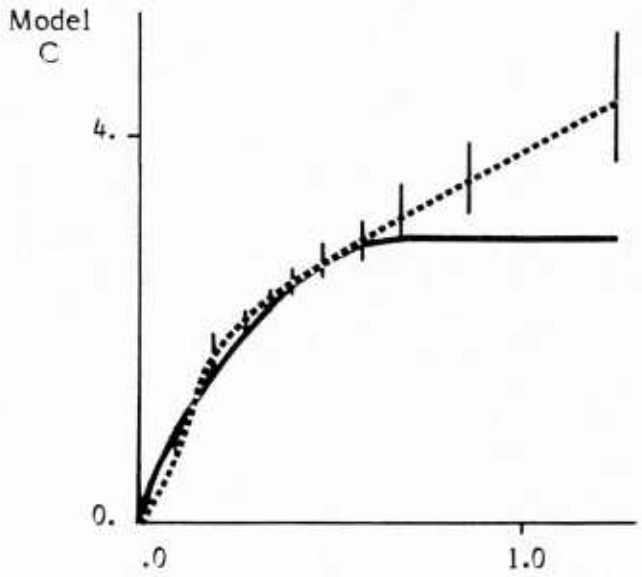


——— Hazard function.

- - - - Mean hazard estimate.

| Variation of hazard estimate.

Figure 5. Hazard function, mean hazard estimate, and variation of hazard estimate for simulation Models B and C (N = 500).



_____ Hazard function.

- - - - Mean hazard estimate.

| Variation of hazard estimate.

Figure 6. Hazard function, mean hazard estimate, and variation of hazard estimate for simulation Model C (N = 150) with limited-inflection constraints.

CONCLUSIONS

When the hazard function was constrained to be isotonic or bidirectional and to have a small number of inflections, the empirical example demonstrated that a maximum penalized likelihood procedure can provide rapid computation of this kind of hazard estimator and can produce a good fit to the data. The shape of the hazard estimate in the empirical example resembled the shape of hazard estimates frequently obtained under two models in the simulation study when moderate sample sizes (75 and 150) were used. In one of those models, the true hazard function increased and then leveled off at the 75th percentile; in the other model, the true hazard function increased and then decreased to about three fourths of its maximum before leveling off. Thus, in light of the simulation results, it is difficult to decide whether the empirical example's true hazard function is an increasing or an increasing-then-slightly-decreasing function.

The simulation study indicated that more substantial decreases (by half or more) in the tail of the true hazard function could be detected reliably by the proposed method when moderate sample sizes were used. The method, therefore, has practicality, because such decreases have been found in empirical studies as well as in analyses of parametric distributions (e.g., the inverse Gaussian distribution). The study also indicated that the proposed method could reliably detect smaller decreases in the true hazard function and could provide rather precise pointwise estimates of the tail of that function when a large sample size (N = 500) was used. The large sample size also resulted in much more accurate estimation of the hazard function across the middle range of the distribution, but bias was still perceptible. Finally, the study indicated that transforming the data to increase the dispersion on x may be an effective means of improving the method's precision with moderate sample sizes. Which transformation to use with which set of constraints is a subject for further study.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19, 716-723.

Ashby, F. G. (1982). Testing assumptions of exponential additive reaction time models. Memory and Cognition, 10, 125-134.

Barlow, R. E., & Proschan, F. (1975). Statistical theory of reliability and life testing: Probability models. New York: Holt, Rinehart & Winston.

Bartoszynski, R., Brown, B. W., McBride, C. M., & Thompson, J. R. (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. Annals of Statistics, 9, 1050-1060.

Bloxom, B. (1983). Some problems in estimating response time distributions. In H. Wainer & S. Messick (Eds.). Principals of modern psychological measurement: A festschrift in honor of Frederick M. Lord (pp. 303-328). Hillsdale, NJ: L. Erlbaum Assoc.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. Perception and Psychophysics, 32, 117-133.

de Boor, C. (1978). A practical guide to splines. New York: Springer-Verlag, 1978.

Fiacco, A. V., & McCormick, G. P. (1968). Nonlinear programming: Sequential unconstrained minimization techniques. New York: John Wiley & Sons.

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. Journal of the American Statistical Association, 75, 667-672.

Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. Stochastic Processes and Their Applications, 13, 11-25.

Johnson, N. I., & Kotz, S. (1970). Continuous univariate distributions-1. Boston: Houghton Mifflin.

Kohfeld, D. L., Santee, J. L., & Wallace, N. D. (1981). Loudness and reaction time: II. Identification of detection components at different intensities and frequencies. Perception and Psychophysics, 29, 550-562.

Kronmal, R. A., & Tarter, M. E. (1968). The estimation of probability densities and cumulatives by Fourier series methods. Journal of the American Statistical Association, 63, 925-952.

Lawless, J. F. (1983). Statistical methods in reliability. Technometrics, 25, 305-316.

Lewis, P. A. W., & Urbine, L. (1981). The new Naval Postgraduate School random number package LLRANDOMII (computer program) (Naval Postgraduate School Report NPS55-81-005). Monterey, CA: Naval Postgraduate School.

Lii, K. S., & Rosenblatt, M. (1975). Asymptotic behavior of a spline estimate of a density function. Computation and Mathematics with Applications, 1, 223-235.

Mendelsohn, J., & Rice, J. (1982). Deconvolution of microfluorometric histograms with B-splines. Journal of the American Statistical Association, 77, 748-753.

Mylander, W. C., Holmes, R. L., & McCormick, G. P. (1973). A guide to SUMT-Version 4: The computer program implementing the sequential unconstrained minimization technique for nonlinear programming (computer program manual) (Res. Rep. RAC-P-63). McLean, VA: Research Analysis Corporation.

Neftci, S. N. (1982). Specification of economic time series models using Akaike's criterion. Journal of American Statistical Association, 77, 537-540.

Rice, J., & Rosenblatt, M. (1983). Smoothing splines: Regression, derivatives, and deconvolution. Annals of Statistics, 11, 141-156.

Schwartz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Shumaker, L. L. (1981). Spline functions: Basic theory. New York: John Wiley & Sons.

Singpurwalla, N. D., & Wong, M.-Y. (1983). Estimation of the failure rate--A survey of nonparametric methods. Part I: Non-Bayesian methods. Communications in Statistics--Theory and Methods, A12, 559-588.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modelling capacity in simple processing systems. In N. J. Castellan and F. Restle (Eds.). Cognitive theory: Volume 3 (pp 199-239). Hillsdale, NJ: L. Erlbaum Assoc.

Wahba, G. (1976). Histosplines with knots which are order statistics. Journal of the Royal Statistical Society (Series B), 38, 140-151.

Wold, S. (1974). Spline functions in data analysis. Technometrics, 16, 1-11.

Wright, J. W., & Wegman, E. J. (1980). Isotonic, convex, and related splines. Annals of Statistics, 8, 1023-1035.

DISTRIBUTION LIST

Military Assistant for Training and Personnel Technology (OSUDS) (R&AT)
Chief of Naval Operations (OP-01B7) (2)
Chief of Naval Material (NMAT 0722)
Commander, Naval Military Personnel Command (NMPC-013C)
Chief of Naval Research (Code 200), (Code 270), (Code 400), (Code 440), (Code 442), (Code 442PT)
Office of Naval Research, Detachment Pasadena
Chief of Naval Education and Training (Code N-21)
Commanding Officer, Naval Training Equipment Center (Technical Library) (5)
Program Manager, Life Sciences Directorate, Bolling Air Force Base (AFOSR/NL)
Commanding Officer, U.S. Coast Guard Research and Development Center, Avery Point
Superintendent, Naval Postgraduate School
Director of Research, U.S. Naval Academy
Defense Technical Information Center (DDA) (12)

U21497Z

DEPARTMENT OF THE NAVY

NAVY PERSONNEL RESEARCH AND
DEVELOPMENT CENTER
SAN DIEGO, CALIFORNIA 92152

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

POSTAGE AND FEES PAID
DEPARTMENT OF THE NAVY
DOD-316

U.S.MAIL