N.C.B. Jespersen

# Dichotomizing a Continuous Covariate in the Cox Regression Model

Preprint
April
1986
3

Institute of Mathematical Statistics
University of Copenhagen

N.C.B. Jespersen

# DICHOTOMIZING A CONTINUOUS COVARIATE IN THE COX REGRESSION MODEL

Preprint 1986 No. 3

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

April 1986

Abstract:

We consider the Cox regression model with one continuous covariate. Sometimes it is convenient to dichotomize the covariate. In this note we propose a likelihood based method of choosing a threshold for the continuous covariate. Furthermore we propose a natural test statistic for the hypothesis of no effect of the covariate in this framework and derive the large sample properties of this test statistic. The method is illustrated by an application to the influence of estrogen receptor level on breast cancer recurrence and mortality.

## 1. Introduction.

In this note we will consider the proportional hazards model (see e.g. Cox (1972)) for survival data. We thus specify the mortality $\lambda_i(\cdot)$ for individual $i$ in the following way

$$(1.1) \qquad \lambda_i(t) = \lambda_0(t)\exp\{\underline{\beta}'\underline{x}^i\}$$

where $\underline{x}^i$ is a covariate vector describing the characteristics of individual $i$ and $\lambda_0(\cdot)$ is a common underlying intensity function. In general the components of $\underline{x}^i$ can be both continuous and discrete, and sometimes it is convenient to discretize a covariate originally being continuous. This can be the case when the aim of the analysis is to use the covariate as a predictor i.e. to divide the individuals into two (or more) "risk" groups according to the value of the covariate or if the exact measurement of the covariate is unreliable Finally, in some situations a "linear" dependence on the covariate is inappropriate and a threshold dependence in fact more natural. Sometimes a theoretically based threshold value can be given and applied but one is often forced to choose it after inspection of data. In this note we will consider the case with one covariate only and propose a likelihood based method of choosing a threshold $\gamma$. We will thus specify the mortality as follows

$$\lambda_i(t) = \lambda_0(t)\exp\{\beta\cdot I(x_i\leq\gamma)\} \;;\; \beta,\gamma \text{ varying}$$

We will consider the problem of testing the hypothesis H: $\beta=0$. Note that the parameter $\gamma$ is present only under the alternative hypothesis. Davies (1977) discusses hypothesis testing in statistical models with this special feature. One of his proposals is to base the test on the normalized score test process ($\gamma$ being the 'time-parameter'). We will show that the score test process converges weakly to the tied-down Ornstein-Uhlenbeck process ( a Brownian bridge normalized to have a constant variance) as the number of individuals tends to infinity subject to some mild regularity conditions. The theory was developed in connection with a study of the effect of hormone receptor level on the recurrence/-death intensity for primary breast cancer patients, and we include a numerical example from that study. For related problems and similar approaches see Miller & Siegmund (1982) and Matthews et. al (1985).

## 2. Mathematical framework and notation.

In this note we will use the counting process set-up for survival data (see e.g. Andersen & Gill (1982)). Let $T_1,\ldots,T_n$ be independent lifetimes and $C_1,\ldots,C_n$ the corresponding censoring times. We thus observe $n$ independent univariate counting processes $N_1,\ldots,N_n$, where $N_i(t) = I(T_i\leq t\wedge C_i)$ i.e. $N_i$ jumps when individual $i$ is observed to die, over the time interval $[0,1]$. For mathematical convenience in the asymptotic theory we will model the (one-dimensional) covariates as i.i.d. random variables $X_1,\dots,X_n$. Let $F^n$ be the $\sigma$-algebra spanned by $X_1,\dots,X_n$ and $(F_t^n)_{t\in[0,1]}$ a right-continuous non-decreasing family of $\sigma$-algebras; $F_t^n$ represent everything that happens up to time $t$ for the $n$ individuals. Introduce $G_t^n = F^n \vee F_t^n$. In our set-up, properties of stochastic processes such as being a martingale or a predictable process are relative to the right continuous family of $\sigma$-algebras $(G_t^n)_{t\in[0,1]}$. A generalization accomodating more complicated censoring patterns than right-censoring is obtained by introducing $Y_i(\cdot)$, being predictable processes taking values in $\{0,1\}$ and indicating when individual $i$ is under observation. In the case of right-censoring $Y_i(t) = I(t\leq C_i ,t\leq T_i)$.

Let $N_\cdot^n(t) = \sum_{i=1}^n N_i(t)$ , $Y_\cdot^n(t) = \sum_{i=1}^n Y_i(t)$.

For convenience we will drop the superscript $n$ in the sequel.

## 3. Statistical model.

We will now specify a statistical model for the observations by modelling the intensity process for $(N_i(s))_{\substack{s\in[0,1] \\ i=1\dots n}}$

in the following way

$$(3.1) \qquad \lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\beta\cdot I(X_i\leq\gamma)\}$$

where $\beta$ and $\gamma$ are unknown real parameters to be estimated and $\lambda_0(\cdot)$ is an unknown intensity function to be estimated. The parameter $\gamma$

is thus the threshold, $\beta$ the effect of having an X-value less than $\gamma$ and $\lambda_0(\cdot)$ is the intensity function for the underlying life time distribution. Note that the model specification (3.1) implies that conditionally on $X_1 = x_1$, $X_2 = x_2, \dots, X_n = x_n$ and keeping $\gamma$ fixed we have a standard proportional hazards model with covariates $z_i = I\{x_i \le \gamma\}$. The statistical inference for $\beta$ and $\gamma$ will be based on the partial likelihood function L (see Andersen & Gill (1982)).

$$(3.2) \quad \log L(\beta,\gamma) = \beta \sum_{i=1}^{n} N_i(1) I(X_i \le \gamma) - \int_0^1 \log\Big(\sum_{i=1}^{n} Y_i(u) \exp\{\beta I(X_i \le \gamma)\}\Big) N.(du)$$

In section 4 we will briefly discuss the estimation of $\beta, \gamma$ and $\lambda_0(\cdot)$ and devote the rest of the paper to testing the hypothesis $H_0 : \beta = 0$.

## 4. Estimation.

For *fixed* $\gamma$ the log likelihood (3.2) is known to have nice properties (see Andersen & Gill (1982)). We can find the maximum likelihood estimate $\hat{\beta}_\gamma$ for $\beta$ and the asymptotic distribution of $\hat{\beta}_\gamma$ is well known, subject to some regularity assumptions. The partially maximized likelihood - which is also denoted L -

$$(4.1) \qquad L(\gamma) = \sup_\beta L(\beta,\gamma) = L(\hat{\beta}_\gamma, \gamma)$$

is piecewise constant between the values of $X_1, \dots X_n$. Therefore a maximum likelihood estimator $\hat{\gamma}$ for $\gamma$ exists but it is not unique. When $\beta$ and $\gamma$ have been estimated we can estimate the integrated intensity

$$\Lambda_0(t) = \int_0^t \lambda_0(u)\,du$$

$$\hat{\Lambda}_0(t) = \int_{[0,t]} \left[\sum_{i=1}^{n} Y_i(u)\exp\{\hat{\beta} I(X_i \leq \hat{\gamma})\}\right]^{-1} N.(du).$$

(See e.g. Breslow (1972)).

## 5. Testing $H_0$: $\beta = 0$.

We will consider the hypothesis $H_0$: $\beta = 0$. Standard maximum likelihood theory - though based on the partial likelihood function - suggests that the test of the null-hypothesis should be based on the likelihood ratio test statistic

$$(5.1) \qquad -2\log Q = -2\log\left(\frac{L(0)}{L(\hat{\beta},\hat{\gamma})}\right) = \sup_{\gamma} \; -2\log\left(\frac{L(0)}{L(\hat{\beta}_\gamma,\gamma)}\right)$$

$$= \sup_{\gamma}\left[-2\log Q_\gamma\right]$$

where $Q_\gamma = \dfrac{L(0)}{L(\hat{\beta}_\gamma,\gamma)}$ which is the usual likelihood ratio when $\gamma$ is fixed. We can interpret $-2\log Q_\gamma$ as the distance between the hypothesis $H_0$: $\beta = 0$ and the hypothesis $H_\gamma$: $\beta \in R, \gamma$ fixed. So $-2\log Q$ is the supremum of such distances. Instead of considering $-2\log Q_\gamma$ one could use as a distance $|S_\gamma|$ normalized by its variance, where $S_\gamma$ is the score test statistic for fixed $\gamma$:

$$(5.2) \qquad S_\gamma = \frac{\partial}{\partial\beta}\log L(\beta,\gamma)\Big|_{\beta=0} = \sum_{i=1}^{n} N_i(1)I(X_i \leq \gamma) - \int_0^1 \left[\frac{\sum_{i=1}^{n} Y_i(u)I(X_i \leq \gamma)}{\sum_{i=1}^{n} Y_i(u)}\right] N.(du)$$

A natural test statistic for $H_0$: $\beta = 0$ based on the score tests is then $\sup_{\gamma}|S_\gamma|$ (see also Davies (1977)). In order to derive large sample properties of $\sup_{\gamma}|S_\gamma|$ we have to make some regularity assumptions to be further interpreted below. First we introduce some notation. If $I \subseteq [0,1]$ is an interval, $t \in I$, we define

$$B_i(\varepsilon,t;I) = \{\exists s\in I\cap]t-\varepsilon,t] : Y_i(s) \neq Y_i(t)\}$$

$$\rho(\varepsilon,I) = \sup_{t\in I} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} PB_i(\varepsilon,t;I)$$

Condition A. There exists a function $\psi$ defined on $[0,1]$ so that

(A.1) $\forall t\in[0,1]: \frac{1}{n}\sum_{i=1}^{n} Y_i(t) \xrightarrow{P} \psi(t)$.

and

(A.2) $\int_0^1 \lambda_0(t)/\psi(t)\ dt < +\infty$. □

Condition B. There exists a partition $0=t_0<t_1<..<t_K=1$ so that

(B.1) $\forall i=1,..,K: \rho(\varepsilon,]t_{i-1},t_i]) \to 0$ as $\varepsilon \to 0$. □

Condition C. The $X_i$'s are i.i.d. with a common continuous distribution function $F$. □

Theorem 5.1 Under conditions A, B and C the process

$$\left(\frac{1}{\sqrt{N.(1)}} S_{F^{-1}(s)}\right)_{s\in[0,1]}$$ converges weakly to the Brownian bridge,

$W^0$, in $D[0,1]$.

Proof $F(X_1),\ldots,F(X_n)$ are i.i.d. and uniformly distributed. Having observed this, the theorem is a consequence of Theorem 7.1 □

Now, $\mathrm{Var}(S_\gamma)=F(\gamma)\ (1-F(\gamma))\ (EN.(1) - \int_0^1\lambda_0(u)du)$ (see Lemma 7.7), therefore a natural test statistic for the hypothesis $H:\beta=0$ would be based on the process $T_\gamma = (N.(1)F(\gamma)(1-F(\gamma)))^{-\frac{1}{2}}\ S_\gamma$

We will discuss two test statistic. The likelihood considerations in the beginning of this section lead to $\sup_\gamma |T_\gamma|$. However, the supremum of the tied-down Ornstein-Uhlenbeck process is $+\infty$ with probability 1. Therefore we must confine ourselves to compact subintervals when taking supremums. For fixed $\varepsilon>0$ we thus introduce

(5.3) $$\tilde{T}_\varepsilon = \sup_{F^{-1}(\varepsilon)\leq\gamma\leq F^{-1}(1-\varepsilon)} |T_\gamma|$$

(5.4) $$\tilde{\tilde{T}}_\varepsilon = \sup_{\hat{F}_n^{-1}(\varepsilon)\leq\gamma\leq \hat{F}_n^{-1}(1-\varepsilon)} (N.(1)\hat{F}_n(\gamma)\ (1-\hat{F}_n(\gamma)))^{-\frac{1}{2}}|S_\gamma|$$

($\hat{F}_n(\cdot)$ denotes the empirical distribution function for the $X_i$'s). We thus only take supremum over some prespecified fraction of F's support.

<u>Theorem 5.2</u> Under conditions A, B and C $\tilde{T}_\varepsilon$ will converge in distribution to

$$T = \sup_{\varepsilon\leq s\leq 1-\varepsilon} \frac{|W^0(s)|}{\sqrt{s(1-s)}} .$$

If, in addition, the support of F is an interval $\tilde{\tilde{T}}_\varepsilon$ will have the same limiting distribution. Finally, an approximation to this distribution is given by

(5.5) $$P(T^2\geq t) \simeq {}^{4\varphi(t)}/t + 2\varphi(t)(t-1/t)\log((1-\varepsilon/\varepsilon))$$

where $\varphi(t)$ is the standard normal density $(2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}t^2)$.

<u>Proof</u>

Rewrite $\tilde{T}_\varepsilon = \sup_{\varepsilon\leq s\leq 1-\varepsilon} \frac{|S_{F^{-1}(s)}|}{\sqrt{s(1-s)}}$. Then the first statement is a consequence of Theorem 5.1. The second statement follows from standard probabilistic arguments. Finally, the approximation (5.5) is derived in Miller & Siegmund (1982). □

For practical purposes $\tilde{\tilde{T}}_\varepsilon$ is preferable to $\tilde{T}_\varepsilon$ because the common distribution function F is usually unknown. One has to decide what ε to choose. Should it be 0.05, 0.025? For a discussion on this matter see Miller & Siegmund (1982). A simple way of circumventing this problem is to use $\frac{1}{\sqrt{N.(1)}} \sup_\gamma |S_\gamma|$ as a test statistic. Obviously, this test statistic is less sensitive to significance of γ's in the tail of F. Finally, the asymptotic distribution of the test statistic is the same as that of the Kolmogorov-Smirnov test statistic (a consequence of Theorem 5.1).

## 6. Regularity conditions and right censoring.

In this section we will comment on the regularity conditions A and B. Notice that $\forall t$: $\psi(t) \leqq 1$ so (A.2) implies $\int_0^1 \lambda_0(t)dt < +\infty$. Conversely if $\int_0^1 \lambda_0(t)dt < \infty$ and ψ is bounded away from zero then (A.2) holds. Interpreting ψ(t) as the fraction of individuals under study at time t (A.2) roughly says that if the fraction of individuals under study is small then the mortality should not be to large - in other words if the mortality is high at a specific timepoint we have to be sure that we have many individuals under risk at that time.

Roughly speaking Conditon B ensures that the censoring does not cluster to heavily such as would be the case if clustering could only take place at fixed timepoints. In the rest of this section we will discuss the conditions A and B in the case of right-censoring. We will consider two types of censoring, namely fixed censoring and type I censoring i.e. the censoring times are i.i.d. and independent of the survival times.

Proposition 6.1. (Fixed censoring)

Let $C_i = c_i$ be a fixed censoring time for individual i. If there exists a strictly positive function $\varphi : [0,1] \to R_+$ which is continuous from the left with right limits and a finite number of discontinuities so that

$$(6.1) \qquad \forall t\in[0,1] : \frac{1}{n}\sum_{i=1}^{n} I(c_i \geq t) \to \varphi(t)$$

and if

$$(6.2) \qquad \int_0^1 \lambda_0(t)dt < +\infty$$

then conditions A and B hold true.

Proof. Note that

$$Y_i(t) = I(T_i \geq t, C_i \geq t) = I(T_i \geq t)\, I(c_i \geq t) \text{ so}$$

$\frac{1}{n}\sum_{i=1}^{n} Y_i(t) \xrightarrow{P} \varphi(t)\exp\{-\int_0^t \lambda_0(s)ds\}$ showing that (A.1) is fulfilled. (A.2) follows from (6.2) and the fact that $\varphi$ and thus $\psi(t) = \varphi(t)\exp\{-\int_0^t \lambda_0(u)du\}$ is bounded away from zero. Let $t_1,\dots,t_{k-1}$ be the jump times for $\varphi$ and set $t_0 = 0, t_k = 1$. Now

$$B_i(\varepsilon,t;]t_{j-1},t_j]) = \{s\in]t_{j-1},t_j]\cap]t-\varepsilon,t]:Y_i(s) \neq Y_i(t)\}$$
$$\subseteq \{t-\varepsilon<T_i\leqq t\}\cup\{t-\varepsilon<c_i\leqq t\}$$

so

$$\rho(\varepsilon,]t_{j-1},t_j])\leqq \sup_{t\in]t_{j-1},t_j]} P(t-\varepsilon<T_1\leqq t) + \sup_{t\in]t_{j-1},t_j]} \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I(t-\varepsilon\leqq c_i<t)$$

$$= \sup_{t\in]t_{j-1},t_j]} P(t-\varepsilon<T_1\leqq t) + \sup_{t\in]t_{j-1},t_j]} (\varphi(t-\varepsilon) - \varphi(t))$$

Since $\varphi$ is uniformly continuous on $]t_{j-1},t_j]$ and $T_1$ has a continuous distribution function we see that $\rho(\varepsilon,]t_{j-1},t_j]) \to 0$ as $\varepsilon\to 0$ and conditon B holds true □

This proposition has its counterpart for type I censoring which can be formulated as follows.

Proposition 6.2. (Type I censoring)

Let $C_i$ be i.i.d. and independent of the $T_i$'s. Assume that the distribution function F for $C_i$ has at most a finite number of discontinuities. If

(6.3) $\int_0^1 \lambda_0(t)dt < +\infty$

and

(6.4) $F(1) < 1$

then conditions A and B hold true.

Proof. Translate the proof of Proposition 6.1 □

Finally we will give an example where one actually needs a partition in Condition B.

Example 6.1.

We will consider the following kind of censoring mechanism: at a given time $t_0$ one tosses a coin to decide whether or not an individual should be excluded from the study. Formally, let $R_i$ be i.i.d $P(R_i=1) = 1 - P(R_i=0) = \alpha \in ]0,1[$ and define the rightcensoring as follows

$$(6.5) \qquad C_i = \begin{cases} 2 \text{ if } R_i = 0 \\ t_0 \text{ if } R_i = 1 \end{cases}$$

Let $F$ be the distribution function for $C_i$ then

$$F(t) = \begin{cases} 0 & t < t_0 \\ \alpha & t_0 \leqq t < 2 \\ 1 & t \geqq 2 \end{cases}$$

so $F$ as has only one discontinuity and $F(1) = \alpha < 1$ so the according to Proposition 6.2 conditions A and B hold true, but it is not true that $\rho(\varepsilon,[0,1]) \to 0$ as $\varepsilon \to 0$.

## 7. Large sample properties of the score test process.

Let F be the common, continous distribution function for the $X_i$'s. Unless otherwise stated we will throughout this section assume that F is uniform. Let $\tilde{N}_i(s) = I(X_i<s)$. Then the score test S can be rewritten

$$(7.1) \qquad S_\gamma = \sum_{i=1}^{n} N_i(1)\tilde{N}_i(\gamma) - \int_0^1 \frac{\sum_{i=1}^{n} Y_i(u)\tilde{N}_i(\gamma)}{\sum_{i=1}^{n} Y_i(u)} N.(du)$$

Introduce

$$(7.2)\qquad U_n(s,t) = \sum_{i=1}^{n} N_i(t)\tilde{N}_i(s) - \int_0^t \frac{\sum_{i=1}^{n} Y_i(u)\tilde{N}_i(s)}{\sum_{i=1}^{n} Y_i(u)} N.(du)$$

Then $S_\gamma = U_n(\gamma,1)$. Let $\Lambda_i(\cdot)$ denote the compensator for $N_i(\cdot)$ i.e. $\Lambda_i(\cdot)$ is the unique adapted increasing process so that $M_i(\cdot) := N_i(\cdot) - \Lambda_i(\cdot)$ is a martingale. $\Lambda_i(t) = \int_0^t Y_i(u)\lambda_0(u)du$. Let furthermore $M. = M_1 + \ldots + M_n$, $\Lambda. = \Lambda_1 + \ldots + \Lambda_n$. Define $\xi_{i,n}$ in the following way

$$(7.3)\qquad \xi_{i,n}(t) = N_i(t) - \int_0^t \frac{Y_i(u)}{\sum_{i=1}^{n} Y_i(u)} N.(du) = M_i(t) - \int_0^t \frac{Y_i(u)}{\sum_{i=1}^{n} Y_i(u)} M.(du)$$

Rewrite (7.2) as follows

$$(7.4)\qquad U_n(s,t) = \sum_{i=1}^{n} N_i(t)\tilde{N}_i(s) - \int_0^t \frac{\sum_{i=1}^{n} Y_i(u)\tilde{N}_i(s)}{\sum_{i=1}^{n} Y_i(u)} N.(du)$$

$$= \sum_{i=1}^{n} \left[ N_i(t) - \int_0^t \frac{Y_i(u)}{\sum_{i=1}^{n} Y_i(u)} N.(du) \right] \tilde{N}_i(s)$$

$$= \sum_{i=1}^{n} \xi_{i,n}(t)\tilde{N}_i(s)$$

Note that $\sum_{i=1}^{n} \xi_{i,n}(t) \equiv 0$. In Lemma 7.2 and Lemma 7.3 below we will deduce some consequences of conditions A and B. But first we will state a lemma for easy reference.

<u>Lemma 7.1</u> Let $Z_1, Z_2, \ldots$ be independent random variables with uniformly bounded variances i.e. $\sup_n \mathrm{Var}(Z_n) < +\infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i - EZ_i) \overset{a.s.}{\rightarrow} 0$$

<u>Proof</u>. See Chung [4] $\square$

<u>Lemma 7.2</u> (A.1) and (B.1) imply that $\psi$ is uniformly continuous on the intervals $]t_{i-1}, t_i]$ $i = 1, \ldots, K$

<u>Proof</u>. Set $I = ]t_{i-1}, t_i]$- Let $t \in I, \delta > 0$ with $t - \delta \in I$ then

$$|\psi(t-\delta) - \psi(t)| \leqq |\psi(t) - \frac{1}{n}\sum_{j=1}^{n} Y_j(t)| + |\psi(t-\delta) - \frac{1}{n}\sum_{j=1}^{n} Y_j(t-\delta)| + \frac{1}{n}\sum_{j=1}^{n} I(Y_j(t) \neq Y_j(t-\delta))$$

$$\leqq |\psi(t) - \frac{1}{n}\sum_{j=1}^{n} Y_j(t)| + |\psi(t-\delta) - \frac{1}{n}\sum_{j=1}^{n} Y_j(t-\delta)| + |\frac{1}{n}\sum_{j=1}^{n} I_{B_j(2\delta, t_j I)} - PB_j(2\delta, t; I)|$$

$$+ \frac{1}{n}\sum_{j=1}^{n} PB_j(2\delta, t; I)$$

Taking $\limsup_{n\to\infty}$ on both sides we see that almost surely - using Lemma 7.1 - $|\psi(t-\delta) - \psi(t)| \leqq \rho(2\delta, I)$ showing that $\psi$ is uniformly continuous on I $\square$

<u>Lemma 7.3</u> Let $Z_1, Z_2, \ldots$ be i.i.d. random variables which are independent of $Y_1(\cdot), Y_2(\cdot), \ldots$ . Assume that $EZ_1^2 < +\infty$ and let $\alpha = EZ_1$ . If Conditions A and B hold true then

$$(7.5) \qquad \sup_{t\in[0,1]} |\frac{1}{n}\sum_{j=1}^{n} Y_j(t) Z_j - \psi(t)\,\alpha| \rightarrow 0 \quad \text{almost surely}$$

Proof Let $\varepsilon > 0$ and $I = ]t_{i-1}, t_i]$. Since $\psi$ is uniformly continuous on I (see Lemma 7.2) we can choose $t_{i-1}=u_0<u_1<\ldots<u_N = t_i$ so that $\sup_{s,t\in ]u_j,u_{j+1}]} |\psi(s)-\psi(t)|<\varepsilon; j=0,\ldots,n-1$ and, letting $\delta:=\max_j |u_{j+1}-u_j|$, $\rho(\delta,I)<\varepsilon$. Since the $Y_j(t)Z_j$'s have uniformly bounded variances $\frac{1}{n}\sum_{j=1}^{n} Y_j(t)Z_j \xrightarrow{a.s.} \psi(t)\alpha$ according to Lemma 7.1. We can now choose $n_0$ (which is stochastic) so that for $n \geq n_0$

$$(7.6)\qquad \left|\frac{1}{n}\sum_{j=1}^{n} Y_j(u_k)Z_j-\psi(u_k)\alpha\right|<\varepsilon;\ k = 0,\ldots,N$$

$$(7.7)\qquad \left|\frac{1}{n}\sum_{j=1}^{n} (|Z_j|\cdot I_{B_j(\delta,u_k;I)}-E|Z_j|PB_j(\delta,u_k;I))\right|<\varepsilon;\ k = 0,\ldots,N$$

For $t\in ]u_k,u_{k+1}]$ and $n \geqq n_0$ we then have

$$\left|\frac{1}{n}\sum_{j=1}^{n} Y_j(t)Z_j-\alpha\psi(t)\right| \leqq \left|\frac{1}{n}\sum_{j=1}^{n} (Y_j(t)-Y_j(u_{k+1}))Z_j\right|+\left|\frac{1}{n}\sum_{j=1}^{n} Y_j(u_{k+1})Z_j-\alpha\psi(u_{k+1})\right|$$

$$+ |\alpha\psi(u_{k+1}) - \alpha\psi(t)|$$

$$\leqq\left|\frac{1}{n}\sum_{j=1}^{n} (|Z_j|I_{B_j(\delta,u_{k+1};I)}-E|Z_j|PB_j(\delta,u_{k+1};I))\right|$$

$$+ \frac{1}{n}\sum_{j=1}^{n} PB_j(\delta,u_{k+1};I)E|Z_1|$$

$$+\left|\frac{1}{n}\sum_{j=1}^{n} Y_j(u_{k+1})Z_j-\alpha\psi(u_{k+1})\right|+|\alpha|\cdot|\psi(t)-\psi(u_{k+1})|$$

$$\leqq\varepsilon+E|Z_1|\frac{1}{n}\sum_{j=1}^{n} PB_j(\delta,u_{k+1};I) + \varepsilon + |\alpha|\varepsilon$$

This implies that

$$\limsup_{n\to\infty} \sup_{t\in I} \left|\frac{1}{n}\sum_{j=1}^{n} Y_j(t)Z_j-\alpha\psi(t)\right|\leqq(2+|\alpha|+E|Z_n|)\varepsilon$$

almost surely.

Letting $\varepsilon \to 0$ through rationals we obtain (7.5). □

Remark. Simple examples show that it is not possible to replace Condition C with, say continuity of $\psi$.

Now we are able to show the following fundamental result.

Lemma 7.4. Assume that Conditions A and B hold true. For fixed $t \in [0,1]$, the finite-dimensional distributions of $\frac{1}{\sqrt{n}} U_n(\cdot,t)$ converge weakly to those of the process $(\int_0^t \psi(u)\lambda_0(u)du)^{\frac{1}{2}} W^0$, where $W^0$ is the Brownian bridge on $D[0,1]$.

Proof. The basic idea is to apply the standard limit theorems for the proportional hazard model (see Andersen & Gill (1982)). If $s_1,..,s_m \in [0,1]$ we define $Z_{i,k} = I(X_i \leq s_k)$, $i = 1,..n$; $k = 1,..,m$ and $\underline{Z}_i = (Z_{i,1},...,Z_{i,m})$. Applying Lemma 7.3 to appropriate sequences of random variables yields

$$(7.8) \quad \sup_t \left|\frac{1}{n}\sum_{i=1}^{n} Y_i(t) - \psi(t)\right| \xrightarrow{P} 0$$

$$(7.9) \quad \sup_t \left|\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\underline{Z}_i - \psi(t)E\underline{Z}_1\right| \xrightarrow{P} 0$$

$$(7.10) \quad \sup_t \left|\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\underline{Z}_i^{\otimes 2} - \psi(t)E\underline{Z}_1^{\otimes 2}\right| \xrightarrow{P} 0$$

(here $\underline{Z}_i^{\otimes 2}$ denotes the m x m matrix with (j,k)'th element $Z_{i,j}Z_{i,k}$). Arguing as in Andersen & Gill (1982), Theorem 3.2 we get that

$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_{i,n}(\cdot)\underline{Z}_i = \frac{1}{\sqrt{n}} (U_n(s_1,\cdot),..,U_n(s_m,\cdot))$ converges in distribution

to a Gaussian process with independent increments and variance function $t \to \int_0^t \psi(u)\lambda_0(u)du\ (E\underline{Z}_1^{\otimes 2} - (E\underline{Z}_1)^{\otimes 2})$.

Now, $E\underline{Z}_1^{\otimes 2}-(E\underline{Z}_1)^{\otimes 2}= (s_j \wedge s_k - s_j s_k)_{j,k}$ which is the covariance matrix for $(W^0(s_1)),\ldots,W^0(s_k))$, $W^0$ denoting standard Brownian bridge. □

In order to prove the weak convergence of the process $\frac{1}{\sqrt{n}} U_n(\cdot,t)$ we need two more lemmas.

Lemma 7.5 Assume that Conditions A and B hold true. Then

$$(7.11) \qquad P\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{i,n}(t)^2>M\right)\to 0 \text{ as } n \to \infty$$

where $M = \int_0^t \psi(u)\lambda_0(u)du + \left(\int_0^t \lambda_0(u)du\right)^2 + 1$

Proof.

$$\frac{1}{n}\sum_{i=1}^{n}\xi_{i,n}(t)^2=\frac{1}{n}\sum_{i=1}^{n}\{N_i(t)+\left(\int_0^t \frac{Y_i}{\Sigma Y_i}dN.\right)^2- 2N_i(t)\int_0^t \frac{Y_i}{\Sigma Y_i}dN.\}$$

$$\leqq \frac{1}{n} N.(t) +\left(\int_0^t \frac{1\{Y.>0\}}{Y.} dN.\right)^2$$

(7.11) will follow if we can show

$$(7.12) \qquad \frac{1}{n}N.(t) \xrightarrow{P} \int_0^t \psi(u)\lambda_0 du$$

$$(7.13) \qquad \int_0^t \frac{1\{Y.>0\}}{Y.} dN. \xrightarrow{P} \int_0^t \lambda_0(u)du$$

In order to show (7.12) we note that

$$E\frac{1}{n}N.(t) = \frac{1}{n}E\Lambda.(t) = \frac{1}{n}\int_0^t EY.(u)\lambda_0(u)du = \int_0^t E\left(\frac{1}{n}Y.(u)\right)\lambda_0(u)du \to \int_0^t \psi(u)\lambda_0(u)du$$

by dominated convergence and Var $\left(\frac{1}{n}N.(t)\right) =$

$$\frac{1}{n^2}\sum_{i=1}^{n} \text{Var } N_i(t) \leqq \frac{1}{n} \cdot \frac{1}{n} E\, N.(t) \to 0.$$

So (7.12) follows by Chebycheff's inequality.

Write $\int_0^t \frac{1\{Y.>0\}}{Y.}\, dN. = A_n + B_n$ where $A_n = \int_0^t \frac{1\{Y.>0\}}{Y.} dM.$ and

$B_n = \int_0^t \frac{1\{Y.>0\}}{Y.}\, d\Lambda. = \int_0^t 1\{Y.(u)>0\}\lambda_0(u)du.$

Again, using dominated convergence $EB_n \to \int_0^t \lambda_0(u)du$ and Var $B_n \to 0$ so $B_n \overset{P}{\to} 0$. Since M. is a mean zero martingale $EA_n = 0$. Now

$$\mathrm{Var}(A_n) = EA_n^2 = E\int_0^t \frac{1\{Y.>0\}}{Y.^2}\, d\Lambda. = E\int_0^t \frac{1\{Y.(u)>0\}}{Y.(u)}\, \lambda_0(u)du$$

$$= \int_0^t E\left\{\frac{1\{Y.(u)>0\}}{Y.(u)}\right\} \lambda_0(u)du.$$ By dominated convergence

$$E\left\{\frac{1\{Y.(u)>0\}}{Y.(u)}\right\} = \frac{1}{n} E\left\{\frac{1\{\frac{1}{n}Y.(u)>0\}}{\frac{1}{n}Y.(u)}\right\} \to 0.$$

Again, by dominated convergence, $\mathrm{Var}(A_n) \to 0$. □

<u>Lemma 7.6</u> Let $\xi_i, i=1,\ldots,n$, be random variables with $\sum_{i=1}\xi_i \equiv 0$. Let $X_1,\ldots,X_n$ be i.i.d. uniformly distributed and independent of the $\xi_i$'s. Let $\tilde N_i(s) = I(X_i \leq s)$, $\tilde N_i(s,t) = \tilde N_i(t) - \tilde N_i(s)$. Then

$$(7.14) \qquad \forall s_1 \leq s \leq s_2:\ E\left|\sum_{i=1}^n \xi_i \tilde N_i(s_1,s)\right|^2 \left|\sum_{i=1}^n \xi_i \tilde N_i(s,s_2)\right|^2 \leq 7\cdot E\left(\sum_{i=1}^n \xi_i^2\right)^2 (s_2-s_1)^2$$

<u>Proof</u>. The left hand side of (7.14) equals.

$$(7.15) \qquad \sum_i \sum_j \sum_k \sum_l E\, \xi_i \xi_j \xi_k \xi_l\ \tilde N_i(s_1,s)\tilde N_j(s_1,s)\tilde N_k(s,s_2)\tilde N_l(s,s_2)$$

Since $\tilde N_i(s_1,s)\tilde N_i(s,s_2) \equiv 0$ we only have to sum over terms where $k \neq i \neq l$ and $k \neq j \neq l$. Introduce the convention that $\ddagger$ means that the summation only includes terms with all indices different for instance $\sum^{\ddagger}_{i,j,k} = \sum_{i\neq j, j\neq k, i\neq k}$. With this convention we can rewrite (7.15) in the following way

$$(7.16) \qquad \mathop{\ddagger}_{i,j,k,l} + \mathop{\ddagger}_{i,i,k,l} + \mathop{\ddagger}_{i,j,k,k} + \mathop{\ddagger}_{i,i,j,j}$$

where the term corresponding to $(i,j,k,l)$ is

$E(\xi_i\xi_j\xi_k\xi_l\tilde{N}_i(s_1,s)\tilde{N}_j(s_1,s)\tilde{N}_k(s,s_2)\tilde{N}_l(s,s_2))$. Now

$$(7.17) \qquad \mathop{\ddagger}_{i,i,j,j} =(s-s_1)(s_2-s) \sum_{i\neq j} E\xi_i^2\xi_j^2=(s-s_1)(s_2-s)\left[E(\sum_{i=1}^{n}\xi_i^2)^2-\sum_1^n E\xi_i^4\right]$$

$$(7.18) \qquad \mathop{\ddagger}_{i,j,k,k} =(s-s_1)^2(s_2-s) \mathop{\ddagger}_{i,j,k} E\xi_i^2\xi_j\xi_k$$

$$(7.19) \qquad \mathop{\ddagger}_{i,i,k,l} =(s-s_1)(s_2-s)^2 \mathop{\ddagger}_{i,j,k} E\xi_i^2\xi_j\xi_k$$

and

$$(7.20) \qquad \mathop{\ddagger}_{i,j,k} E\xi_i^2\xi_j\xi_k= \sum_{i,j,k} - \mathop{\ddagger}_{i,i,j} - \mathop{\ddagger}_{i,j,i} - \mathop{\ddagger}_{i,j,j} - \sum_{i,i,i}$$

$$= 0 - 2\sum_{i\neq j} E\xi_i^3\xi_j - \sum_{i\neq j} E\xi_i^2\xi_j{}^2 - \sum_1^n E\xi_i^4 = -\left[E(\sum_1^n\xi_i^2)^2 - 2\sum_1^n E\xi_i^4\right]$$

Finally

$$(7.21) \qquad \mathop{\ddagger}_{i,j,k,l} = (s-s_1)^2(s_2-s)^2 \mathop{\ddagger}_{i,j,k,l} E\xi_i\xi_j\xi_k\xi_l$$

$$= (s-s_1)^2(s_2-s)^2\left[\sum_{i,j,k,l} E\xi_i\xi_j\xi_k\xi_l - 6\mathop{\ddagger}_{i,j,k} E\xi_i^2\xi_j\xi_k\right.$$

$$\left.- 2\mathop{\ddagger}_{i,j} E\xi_i^2\xi_j^2 - 4\mathop{\ddagger}_{i,j} E\xi_i^2\xi_j - \sum_i E\xi_i^4\right]= (s-s_1)^2(s_2-s)^2\left[4E(\sum_i\xi_i^2) - 7\sum_i E\xi_i^4\right]$$

Collecting terms (7.17) - (7.21) gives

$$(7.15)=(s-s_1)^2(s_2-s)^2\left[4E(\sum_i\xi_i^2)^2-7\sum_i E\xi_i^4\right]-(s-s_1)(s_2-s)(s_2-s_1)\left[E(\sum_i\xi_i^2)^2-2\sum_i E\xi_i^4\right]$$

$$+ (s-s_1)(s_2-s)\left[E(\sum_i\xi_i^2)^2 - \sum_i E\xi_i^4\right] \leqq (s_2-s_1)^2\cdot 7\, E(\sum_i\xi_i^2)^2 \qquad \square$$

Theorem 7.1

Assume that Conditions A,B and C hold true. Then for each fixed $t \in [0,1]$, $\frac{1}{\sqrt{n}} U_n(\cdot,t)$ converges weakly to $(\int_0^t \psi(u)\lambda_0(u)du)^{\frac{1}{2}} W^0$ where $W^0$ is a Brownian bridge on $D[0,1]$.

Proof According to Lemma 7.5 we can find $M<\infty$ so that

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\xi_{i,n}(t)^2>M\right) \to 0 \text{ as } n \to \infty. \text{ Define}$$

$$\tilde{\xi}_{i,n}(t)=\xi_{i,n}(t)\cdot 1\{\frac{1}{n}\sum_{i=1}^{n}\xi_{i,n}(t)^2 \leqq M\} \quad \text{and}$$

$$\tilde{U}_n(s,t) = \sum_{i=1}^{n}\tilde{\xi}_{i,n}(t)\tilde{N}_i(s). \text{ Notice that}$$

$$(7.22) \qquad P(U_n(\cdot,t)\neq\tilde{U}_n(\cdot,t))\leqq P(\frac{1}{n}\sum_{i=1}^{n}\xi_{i,n}(t)^2 > M) \to 0$$

so Lemma 7.4 implies that the finite-dimensional distributions of $\frac{1}{\sqrt{n}}\tilde{U}_n(\cdot,t)$ converge weakly to those of the limiting distribution indicated in the theorem. Now, Lemma 7.6 shows that

$$(7.23) \quad E\left|\frac{1}{\sqrt{n}}\tilde{U}_n(s,t)-\frac{1}{\sqrt{n}}\tilde{U}_n(s_1,t)\right|^2\left|\frac{1}{\sqrt{n}}\tilde{U}_n(s_2,t)-\frac{1}{\sqrt{n}}\tilde{U}_n(s,t)\right|^2\leqq 7\cdot M\cdot(s_2-s_1)^2$$

so by Theorem 15.6 in Billingsley (1968) we obtain weak convergence of $\frac{1}{\sqrt{n}}\tilde{U}_n(\cdot,t)$ to the limit stated. But by (7.22) this implies that $\frac{1}{\sqrt{n}}U_n(\cdot,t)$ converges weakly to the same limit. □

Remark 1. Let $\varphi(t) = \int_0^t \psi(u)\lambda_0(u)du$. As a matter of fact we have shown that for each fixed $t\in[0,1]$ the process $\frac{1}{\sqrt{n}}U_n(\cdot,t)$ converges weakly to a Gaussian process with covariance function $(s_1,s_2) \to (s_1\wedge s_2-s_1s_2)\varphi(t)$ and for each fixed $s\in[0,1]$ the process $\frac{1}{\sqrt{n}}U_n(s,\cdot)$ converges weakly to a Gaussian process with covariance function $(t_1,t_2)\to s(1-s)\varphi(t_1\wedge t_2)$.

Furthermore, there exists a Gaussian two-parameter process with covariance function $((s_1,t_1),(s_2,t_2)) \to (s_1 \wedge s_2 - s_1 s_2)\varphi(t_1 \wedge t_2)$, namely a <u>Kiefer-process</u> where the t-parameter is transformed with the continuous function $\varphi$ (for the notion of Kiefer-processes see e.g. Csörgö & Revesz (1968)). Therefore, it seems plausible that, with a little bit more effort and perhaps further regularity conditions, it is possible to show weak convergence of $\frac{1}{\sqrt{n}}U_n(\cdot,\cdot)$ in a suitable space $D[0,1]^2$ (see e.g. Bickel & Wichura (1971) or Straf (1972)) to the modified Kiefer-process.$\square$

<u>Remark 2</u>. Since $\frac{1}{n}$ N.(t) $\xrightarrow{P} \int_0^t \psi(u)\lambda_0(u)du$ (see the proof of Lemma 7.5) we can conclude that for <u>fixed</u> $t \in [0,1]$ the process $\frac{1}{\sqrt{N.(t)}}U_n(\cdot,t)$ converges weakly to a Brownian bridge.

In Lemma 7.7 below we will drop the assumption that the $X_i$'s are uniformly distributed but will allow for a general common distribution function F.

<u>Lemma 7.7</u> The variance of the $U_n(.,.)$ is given by

$$\text{(7.24)} \qquad \operatorname{Var} U_n(s,t) = F(s)(1-F(s))\,(E\,N.(t) - \int_0^t \lambda_0(u)du)$$

<u>Proof</u> $\quad \operatorname{Var} U_n(s,t) = \operatorname{Var} \sum_{i=1}^n \xi_{i,n}(t)\tilde{N}_i(s)$

$$= \operatorname{Var}(E(\sum_{i=1}^n \xi_{i,n}(t)\tilde{N}_i(s) \,|\, \xi_{i,n}(t)\ ,\ i=1,\dots,n))$$

$$+ E(\operatorname{Var}(\sum_{i=1}^n \xi_{i,n}(t)\tilde{N}_i(s) \,|\, \xi_{i,n}(t)\ ,\ i=1,\dots,n))$$

$$= 0 + E\sum_{i=1}^n \xi_{i,n}(t)^2 \operatorname{Var}\tilde{N}_i(s) = F(s)(1-F(s)\sum_{i=1}^n E\xi_{i,n}(t)^2.$$

Now, $\xi_{i,n}(t)^2 = M_i(t)^2 + (\int_0^t \frac{Y_i}{Y.} dM.)^2 - 2M_i(t)\int_0^t \frac{Y_i}{Y.} dM.$

so

$$E\xi_{i,n}(t)^2 = EN_i(t) + E\int_0^t \left[\frac{Y_i(u)}{Y.(u)}\right]^2 Y.(u)\lambda_0(u)du$$

$$- 2E\int_0^t \frac{Y_i(u)^2}{Y.(u)} \lambda_0(u)du = E\ N_i(t) - E\int_0^t \frac{Y_i(u)}{Y.(u)} \lambda_0(u)du.$$

Adding these terms yields (7.24). □

## 8. An example.

The present investigation was motivated by a study of the role of estrogen receptors as a possible risk factor for low-risk primary breast cancer patients; full medical details were discussed by Thorpe et al. (1986). Usually the estrogen receptor measurements (ER for short) are dichotomized using a threshold of $9.10^{-15}$mol/mg protein i.e. patients with ER $\leq$ 9 are said to be ER÷ and patients with ER > 9 are said to be ER+ (though thresholds of 1,2,6 and 19 have also been proposed in the literature). ER+ patients are expected to have a better survival than ER÷ patients. Since the estrogen content in the tumor is strongly related to the menopausal state we have performed a Cox regression analysis for both pre- and postmenopausal patients. For premenopausal patients (n = 193) it turned out that the only significant covariate was the ER-status ('standard' covariates such as tumor size, age, malignancy grading etc. being non-significant). To shed some light on the relevance in the present data of the conventional threshold of 9, we derived the profile log likelihood (for each $\gamma$, we have maximized over $\beta$), c.f. fig. a.

It is seen that there is a peak at $\gamma=7$ corresponding to $7\cdot 10^{-15}$ mol/mg as the 'best' threshold. We are now interested in the hypothesis $H_0:\beta=0$. We find $\tilde{\tilde{T}}_\varepsilon$ 3.4 ($\varepsilon=1/20$) and $\sup_\gamma |S\gamma|/\sqrt{N.(1)}=1.6$ corresponding to p-values of 2.3% and 1.6% respectively, indicating the significance of ER. For postmenopausal patients (n=498), using again the conventional threshold $\gamma=9$, there were no significant prognostic factors. Performing the same analysis as above, fig. b shows the profile log likelihood. There is no peak corresponding to the insignificance of ER. We find $\tilde{\tilde{T}}_\varepsilon=0.7$ ($\varepsilon=1/20$) and $\sup_\gamma|S\gamma|/\sqrt{N.(1)}=0.3$ corresponding to p-values of 55% and 99% respectively both being insignificant.

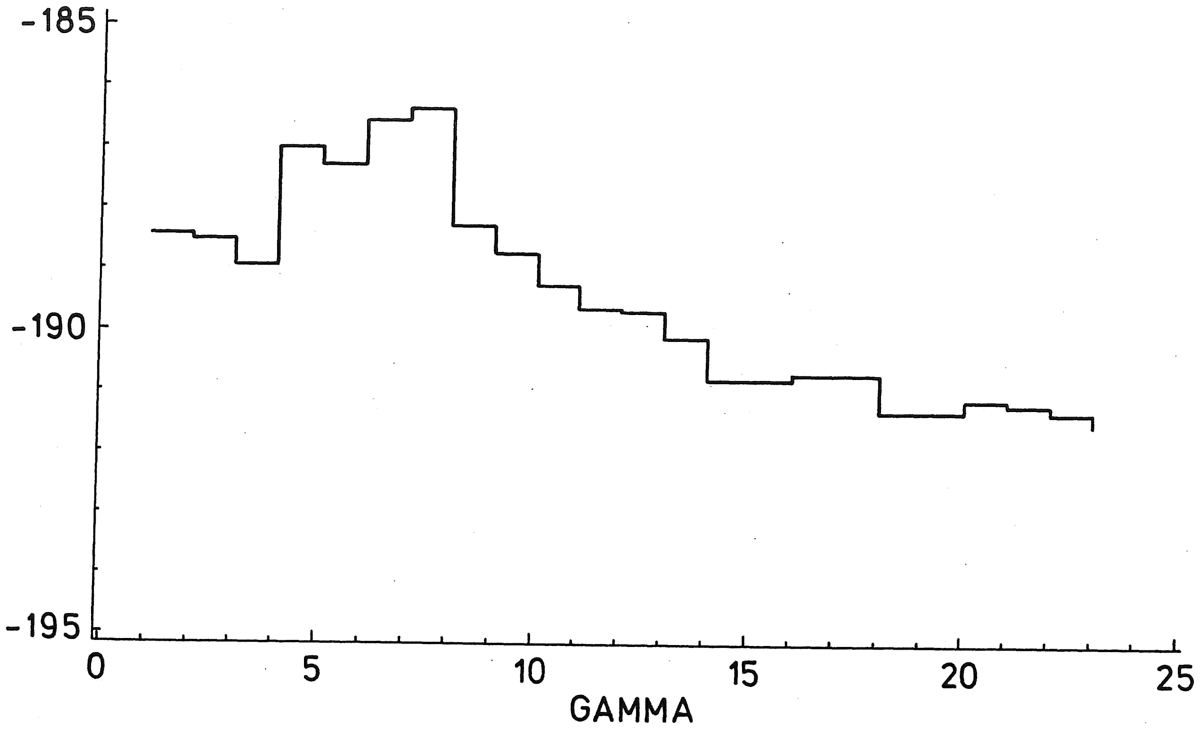


fig. a. Profile log partial likelihood for premenopausal women (n=193)

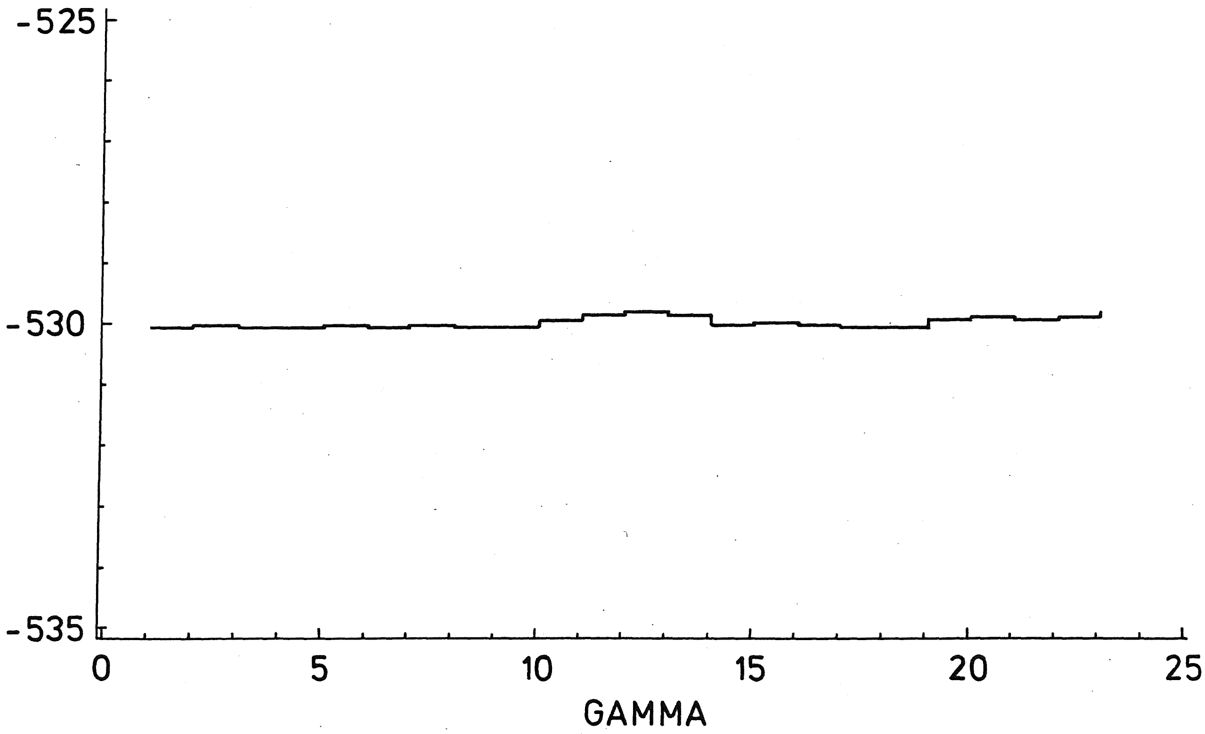


fig. b. Profile log partial likelihood for postmenopausal women (n=498)

## 9. Miscellaneous.

In this section we will briefly relate the results in this paper to two other recent papers.

Miller & Siegmund (1982) consider the following situation. Let X be a quantitative variable being potentially predictive of an event E. Values of this variable are available for all individuals. For each x we form the 2x2 table

| | X<x | X>x |
|---|---|---|
| E | a | b |
| $E^c$ | c | d |

and use $\max_x \chi_x^2$, $\chi_x^2=N(ad-bc)[(a+b)(a+c)(b+d)(c+d)]^{-1}$, $N=a+b+c+d$, as a test statistic for the hypothesis of no predictive effect of X. Miller & Siegmund show, roughly speaking, that $(\chi_x^2)_x$, as a stochastic process, converges in distribution to

$$\left(\frac{(W_s^o)^2}{s(1-s)}\right)_{s\in]0,1[}$$ as N tends to infinity. Noting that $\chi_x^2$ in fact is a score test statistic normalised with its (conditional) variance we see that this result is similar to ours.

Wei (1984) considers the proportional hazards model with one covariate X i.e. the hazard for a patient with covariate X=x is of the form $\lambda(t;x)=\lambda_0(t)\exp\{\alpha x\},\alpha\in R$, $\lambda_0(\cdot)$ is an (arbitrary) intensity function. He introduces a goodness-of-fit test statistic for this proportional hazards assumption as follows. At each timepoint t we calculate the derivative of the log partial likelihood function and insert the proportional hazards estimate of $\alpha$ i.e. $S_t=\frac{\partial}{\partial\alpha}\log L(\alpha;t)\big|_{\alpha=\hat{\alpha}}$. Wei suggests that $T=\sup_t S_t^2$ could be used as goodnes-of-fit test statistic. Wei's argument for using this statistic is that $\frac{\partial}{\partial\alpha}\log L(\alpha;t)$ has a nice interpretation as a difference between the observed and expected number of deaths (if X is binary at least) and since $\alpha$ is unknown we replace it with the proportional hazards estimate $\hat{\alpha}$. He shows that $(S_t)_t$, properly normalised, converges in distribution to a time transformed Brownian bridge. Now, consider the extension of the proportional hazards model given by $\lambda(t;x)=\lambda_0(t)\exp\{(\alpha+\beta 1\{t\leq\tau\})X\}$ $\alpha,\beta,\tau\in R$, $\lambda_0(\cdot)$ on (arbitraty) intensity function. The porportional hazards model corresponds to $H_0:\beta=0$. Note that $\tau$ vanishes under $H_0$ so following Davies (1977) we should base a test for $H_0$ on the process $(\tilde{S}_\tau)_\tau$ where $\tilde{S}_\tau=\frac{\partial}{\partial\beta}\log L(\alpha,\beta,\tau)\big|_{\substack{\alpha=\hat{\alpha}\\ \beta=0}}$, $L(\alpha,\beta,\tau)$ is the partial likelihood function and $\hat{\alpha}$ is the proportional hazards estimate. A little calculus shows that $\tilde{S}_\tau=S_\tau$ so Wei's goodnes-of-fit test is a natural test for $H_0:\beta=0$ in this model. This suggests that the goodnes-of-fit test is good against alternatives with a change in the regression coefficient at an unknown timepoint as indicated above.

Acknowledgements.

I am indebted to Niels Keiding for several detailed discussions of the manuscript providing a number of improvements in the presentation and a stimulating interest for the problem. I owe thanks to Søren Asmussen for introducing Condition B and to Per Kragh Andersen for stimulating discussions. Finally I will thank Susan Thorpe and DBCG for supplying the data used in section 8.

References.

Andersen, P.K. and Gill, R. (1982). Cox's regression Model for Counting Processes: A Large Sample Study. Ann. of Statist., 4, 1100-1120.

Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist., 42, 1656-1670.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Breslow, N. (1972). Contribution to the discussion of the paper by D.R. Cox. J. Roy. Statist. Soc. B., 34, 187-220.

Chung, K.L. (1968). A Course in Probability Theory. Harcourt, Brace & World, Inc., New York.

Cox, D.R. (1972). Regression models and life tables(with discussion). J. roy. Statist. Soc. B., 34, 187-220.

Csörgö, M. and Révész, P. (1968). Strong Approximations in Probability and Statistics. Akadémiai Kiadó, Budapest.

Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrica, 64, 247-254.

Matthews, D.E., Farewell, V.T. and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. Ann. of Stat., 13, 583-591.

Miller, R. and Siegmund, D. (1982). Maximally Selected Chi Square Statistics. Biometrics, 38, 1011-1016.

Smirnov, N. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. Ann. Math. Statist., 19, 279-281.

Straf, M.L. (1972). Weak convergence of Stochastic Processes with several Parameters. In sixth Berkeley Symposium on Mathematics, Statistics and Probability (Eds Le Cam et al.), Vol. II, 187-221. University of California Press.

Thorpe, S., Jespersen, N.C.B., Rose, C., Rasmussen, B.B. and Mouridsen, H.T. (1986). Prognostic Value of Steroid Hormone Receptors: I Multivariate Analysis of Low Risk, Untreated Primary Breast Cancer. Unpublished manuscript.

Wei, L.J. (1984). Testing Goodness of Fit for Proportional Hazards Model with censored Observations. JASA, vol. 79, 387, 649-652.

PREPRINTS 1985

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Cohn, Harry: Almost Sure Convergence for Stochastically Monotone Temporally Homogeneous Markov Processes and Applications.

No. 2 Rootzén, Holger: Maxima and Exceedances of Stationary Markov Chains.

No. 3 Asmussen, Søren und Johansen, Helle: Über eine Stetigkeitsfrage betreffend das Bedienungssystem GI/GI/s.

No. 4 Asmussen, Søren and Thorisson, Hermann: A Markov Chain Approach to Periodic Queues.

No. 5 Hald, Anders: Galileo as Statistician.

No. 6 Johansen, Søren: The Mathematical Structure of Error Correction Models.

No. 7 Johansen, Søren and Johnstone, Iain: Some Uses of Spherical Geometry in Simultaneous Inference and Data Analysis.

PREPRINTS 1986

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jespersen, N.C.B.: On the Structure of Simple Transformation Models.

No. 2 Dalgaard, Peter and Johansen, Søren: The Asymptotic Properties of the Cornish-Bowden-Eisenthal Median Estimator.

No. 3 Jespersen, N.C.B.: Dichotomizing a Continuous Covariate in the Cox Regression Model.