BIOMETRICS 57, 114–119
March 2001

A Solution to the Problem of Monotone Likelihood in Cox Regression

Georg Heinze* **and Michael Schemper**

Section of Clinical Biometrics, Department of Medical Computer Sciences, Vienna University, Spitalgasse 23, A-1090 Vienna, Austria
* *email:* Georg.Heinze@akh-wien.ac.at

SUMMARY. The phenomenon of monotone likelihood is observed in the fitting process of a Cox model if the likelihood converges to a finite value while at least one parameter estimate diverges to $\pm\infty$. Monotone likelihood primarily occurs in small samples with substantial censoring of survival times and several highly predictive covariates. Previous options to deal with monotone likelihood have been unsatisfactory. The solution we suggest is an adaptation of a procedure by Firth (1993, *Biometrika* **80,** 27–38) originally developed to reduce the bias of maximum likelihood estimates. This procedure produces finite parameter estimates by means of penalized maximum likelihood estimation. Corresponding Wald-type tests and confidence intervals are available, but it is shown that penalized likelihood ratio tests and profile penalized likelihood confidence intervals are often preferable. An empirical study of the suggested procedures confirms satisfactory performance of both estimation and inference. The advantage of the procedure over previous options of analysis is finally exemplified in the analysis of a breast cancer study.

KEY WORDS: Bias reduction; Infinite estimates; Modified score; Penalized likelihood; Profile likelihood; Proportional hazards model; Separation; Survival analysis.

1. Introduction

Statisticians who often apply Cox's (1972) model in biomedicine or who investigate small-sample properties of the model by simulation are familiar with the problem of monotone likelihood, i.e., during the iterative fitting process, the likelihood converges to a finite value while at least one parameter estimate diverges to $\pm\infty$. In general, one does not assume infinite parameter values in underlying populations. The problem of monotone likelihood is rather one of nonexistence of the maximum likelihood estimate under special conditions in a sample. For a single covariate, this occurs when, at each failure time, the covariate value of the failed individual is the largest of all covariate values in the risk set at that time or when it is always the smallest. It also happens when the same is true for a linear combination of covariates (cf., Jacobsen, 1989).

In Table 1, we show how the probability for the occurrence of monotone likelihood depends on sample size, on the proportion of censoring of survival times, on the magnitude of the relative risk associated with dichotomous covariates, and on the degree of balance in their distribution. Furthermore, this probability increases with an increasing number of dichotomous covariates. Monotone likelihood rarely occurs with continuous covariates or uncensored samples. But highly censored samples with several strong covariates do have a good chance of causing monotone likelihood. Though, as we recognize, the phenomenon of monotone likelihood is by no means unusual under many conditions likely to occur in practice, only few authors have addressed this issue.

Monotone likelihood occurred in a breast cancer study (Lösch et al., 1998) that was analyzed in our department and motivated this work. In this clinical study, survival times of 100 patients were recorded (74 of them censored) and also values of four potential risk factors (i.e., covariates): tumor stage (pT), nodal status (N), histological grading (G), and cathepsin D immunoreactivity (CD). For analysis, these factors were dichotomized to levels of zero and one (unfavorable). Definitions of levels and the frequencies of their occurrence are given in Table 2. Median follow-up time of the study was 72 months. If a Cox model (SC) is fitted to this data set, risk factor G causes monotone likelihood, which is easily recognized by the extreme estimated relative risk of 13543327 (corresponding to a parameter estimate of $\hat{\beta}_G = 16.42$) and an insignificant Wald test.

Currently, if monotone likelihood caused by a covariate (G in our example) is detected in an analysis by Cox's model, the following options are available:

(a) Changing to a different type of model (cf., Johnson et al., 1982).
(b) Omission of G from the model.
(c) Stratification of the analysis on G (cf., Bryson and Johnson, 1981).
(d) Standard Cox regression analysis (SC) with the parameter estimate $\hat{\beta}_G$ set to a high value (e.g., in procedure PHREG of SAS/STAT (SAS, 1999), the $\hat{\beta}_G$ of that iteration is chosen at which the log likelihood had changed by less than 10^{-6}).

Table 1

Probabilities for the occurrence of monotone likelihood in Cox regression in samples of five dichotomous covariates

		$R = 1$		$R = 4$		$R = 16$	
	B:	1:1	1:4	1:1	1:4	1:1	1:4
n	%c:	50/90	50/90	50/90	50/90	50/90	50/90
50		0/23	3/78	0/67	30/99	1/92	59/100
100		0/4	0/45	0/30	2/96	0/64	9/100
200		0/0	0/7	0/4	0/73	0/16	0/100

n denotes sample size, %c expected percentage of censored survival times, R identical relative risk produced by covariates, and B gives the degree of balance identical for all dichotomous covariates; estimated probabilities ($\times 100$) are based on 1000 simulated samples each.

Expressing effects of covariates in terms of the relative risk they produce is not only common in the analysis of failure times but also useful. Models whose parameters have different interpretations (option (a)) may be less appealing. Furthermore, Cox's model may still be appropriate for all other covariates and also with respect to the multiplicativity of risk.

Neither (b) nor (c) provides any information about the effect of an unusually strong and therefore important parameter, option (b) being particularly inappropriate because the effects of the other three factors cannot be adjusted for G. In Table 2, the effects of pT, N, and CD under omission of G are overestimated compared with the results under an SC analysis (as defined by option (d)), while results for pT, N, and CD under stratification by G (not in Table 2) are very similar to the results by an SC analysis.

Option (d) is perhaps the most sensible one in practice. Though Wald tests and related confidence intervals cannot be used due to the extreme inflation of $\text{var}(\hat{\beta}_G)$ when $\hat{\beta}_G$ is set to a value of 16.42, likelihood ratio tests and profile likelihood confidence intervals could be used. What still remains unsatisfactory, however, is the arbitrary choice for $\hat{\beta}_G$ and thus the uncertainty of related estimates such as survival functions.

In this article, we present and suggest a procedure that avoids the arbitrary choice for $\hat{\beta}_G$ and arrives at a finite estimate for β_G by a modification of the score function of Cox's model. The modification needed to arrive at these results was originally derived by Firth (1992a,b, 1993) to reduce the bias of maximum likelihood estimates in generalized linear models. These estimates are biased away from zero and the occurrence of infinite parameter estimates in situations of monotone likelihood can be interpreted as an extreme consequence of this property. Schaefer (1983), Cordeiro and McCullagh (1991), Cordeiro and Cribari-Neto (1998), Leung and Wang (1998), and other authors have discussed the bias of maximum likelihood estimates and have suggested corrections that, however, are only applicable to finite estimates.

In the following section, the approach by Firth (1993) is formulated for Cox regression (FC). In Section 3, the empirical performance of the procedure is explored by simulation. Section 4 revisits the breast cancer study and presents and discusses results by an FC analysis.

2. Definition of Firth's Procedure for Cox Regression

We first review Cox's (1972) regression model and then some principle ideas of Firth (1993), and, finally, we deal with their implementation in Cox regression (FC).

In a sample of n individuals, we observe m distinct and uncensored survival times $t_{(j)}$ $(1 \leq j \leq m)$ among the n possibly censored survival times t_i $(1 \leq i \leq n)$. A covariate row vector $x_i = (x_{i1}, \ldots, x_{ir}, \ldots, x_{ik})$ is related to each individual. Let d_j denote the number of deaths at $t_{(j)}$ and s_j the vector sum of the covariates of the d_j individuals (s_{jr} referring to the rth component of s_j). The set of individuals alive and uncensored prior to $t_{(j)}$, the risk set, is denoted by R_j. A vector β of k regression parameters is to be estimated. Then the log likelihood is defined as

$$\log L(\beta) = \sum_{j=1}^{m} \left[\beta s_j - d_j \log \left\{ \sum\nolimits_{h \in R_j} \exp(x_h \beta) \right\} \right].$$

Maximum likelihood estimates $\hat{\beta}_r$ of regression parameters β_r, $1 \leq r \leq k$, are derived as solutions to the score equations $\partial \log L(\beta_r)/\partial \beta_r \equiv U(\beta_r) = 0$. Cox's model is discussed in more detail by Collett (1994), Marubini and Valsecchi (1995), and Cox and Oakes (1984).

For a wide class of regression models, Firth (1993) demonstrated that the bias of maximum likelihood parameter estimates $\hat{\theta}$ arises from the combination of the unbiasedness of the score function at the true value of θ and the curvature of the score function at θ. By introducing a suitable bias into the score function, the bias in $\hat{\theta}$ can be reduced. The required modification of the score function $U(\theta_r)^* = 0$, $1 \leq r \leq k$, was derived by Firth (1993) in detail, and we recommend this reference to the interested and mathematically inclined reader. Here we give only the main result applicable to the estimation

Table 2

Breast cancer study: risk factors and associated estimates of relative risk

		Estimates of relative risk (p-values)		
Factor	Frequencies	SC (G omitted)	SC	FC (Wald/LR)
pT (2-4) vs. pT (1)	43, 57	4.8 (0.002)	3.6 (0.01)	3.4 (0.01/0.01)
N (1-2) vs. N (0)	32, 68	3.1 (0.01)	2.6 (0.03)	2.5 (0.03/0.03)
G (2-3) vs. G (1)	74, 26	—	13543327 (1.0)	11.3 (0.1/0.01)
CD (pos.) vs. CD (neg.)	30, 70	1.7 (0.25)	1.5 (0.37)	1.5 (0.37/0.36)

Note: p-values of FC (LR) refer to penalized likelihood ratio tests; all other p-values refer to Wald tests.

of parameters β in Cox regression,

$$U(\beta_r)^* \equiv U(\beta_r) + a_r = 0, \qquad 1 \le r \le k,$$

with

$$a_r = 0.5 \operatorname{trace}\left[I(\beta)^{-1}\{\partial I(\beta)/\partial\beta_r\}\right],$$

where $I(\beta)^{-1}$ is the inverse of the information matrix evaluated at β, also known as the estimated covariance matrix of $\hat{\beta}$. The term in brackets, $\{\cdot\}$, is the derivative of the information matrix with respect to parameter β_r. For Cox regression, its elements are given by

$$\begin{aligned}\frac{\partial I_{rs}(\beta)}{\partial\beta_t} &= -\frac{\partial^3 \log L(\beta)}{\partial\beta_r\partial\beta_s\partial\beta_t} \\ &= \sum_{j=1}^{m} d_j \left\{ \left(\frac{S_{j,rst}}{S_{j,0}} - \frac{S_{j,rs}S_{j,t}}{S_{j,0}^2} \right) \right. \\ &\quad - \frac{S_{j,s}}{S_{j,0}} \left(\frac{S_{j,rt}}{S_{j,0}} - \frac{S_{j,r}S_{j,t}}{S_{j,0}^2} \right) \\ &\quad \left. - \frac{S_{j,r}}{S_{j,0}} \left(\frac{S_{j,st}}{S_{j,0}} - \frac{S_{j,s}S_{j,t}}{S_{j,0}^2} \right) \right\}, \\ &\qquad\qquad 1 \le r, s, t \le k,\end{aligned}$$

where

$$\begin{aligned} S_{j,0} &= \sum_{h\in R_j} \exp(x_h\beta), \\ S_{j,r} &= \sum_{h\in R_j} x_{hr} \exp(x_h\,\beta), \\ S_{j,s} &= \sum_{h\in R_j} x_{hs} \exp(x_h\,\beta), \\ S_{j,t} &= \sum_{h\in R_j} x_{ht} \exp(x_h\,\beta), \\ S_{j,rs} &= \sum_{h\in R_j} x_{hr}x_{hs} \exp(x_h\beta), \\ S_{j,rt} &= \sum_{h\in R_j} x_{hr}x_{ht} \exp(x_h\beta), \\ S_{j,st} &= \sum_{h\in R_j} x_{hs}x_{ht} \exp(x_h\beta), \end{aligned}$$

and

$$S_{j,rst} = \sum_{h\in R_j} x_{hr}x_{hs}x_{ht} \exp(x_h\,\beta).$$

When a model based on the modified score function is fitted by the Newton–Raphson algorithm (cf., Collett, 1994, p. 66–67), the term a_r is evaluated at each step of the iteration based on the current value of $\hat{\beta}$. No further adaptations are necessary when the Firth correction is used with Cox regression.

The modified score function is related to the penalized log likelihood and likelihood functions, $\log L(\beta)^* = \log L(\beta) + 0.5 \log|I(\beta)|$ and $L(\beta)^* = L(\beta)|I(\beta)|^{0.5}$, respectively. The penalty function $|I(\beta)|^{0.5}$ is known as Jeffreys invariant prior for this problem. Its influence is asymptotically negligible. For exponential family models in canonical parameterization, it removes the $O(n^{-1})$ bias from the parameter estimates (cf., Firth, 1993). This property can also be assumed to hold for Cox regression, which can be reformulated and interpreted as an exponential family model with canonical parameter β (cf., McCullagh and Nelder, 1989, p. 429).

Do finite FC parameter estimates always exist? At first, we consider a sample of two individuals for which survival time and a single covariate are recorded. Of course, the survival times have to be different (the larger one may be censored) and the covariate values are assumed to differ by, say, one. While an SC estimate does not exist, the modified score function then reduces to $U(\beta)^* = \{1-\exp(\beta)\}/\{1+\exp(\beta)\}+0.5$ having its root at $\exp(\hat{\beta}) = 3$. More generally, it can be shown that FC permits finite estimates of k parameters as long as the risk sets of at least k distinct failure times each have nonzero variance in at least one covariate and each covariate has nonzero variance in at least one of the risk sets. Consider the information matrix $I(\beta)$, the entries of which are

$$\begin{aligned} I_{rs}(\beta) = \sum_{j=1}^{m} d_j \left\{ \sum_{h\in R_j} x_{hr}\, x_{hs}\, w_h \right. \\ \left. - \left(\sum_{h\in R_j} x_{hr} w_h \right) \left(\sum_{h\in R_j} x_{hs} w_h \right) \right\}, \end{aligned}$$

with $1 \le r, s \le k$, and $w_h = \exp(x_h\beta)/\Sigma_{l\in R_j} \exp(x_l\beta)$. Each of the m summands corresponds to one risk set R_j and can be interpreted as a covariance matrix of the covariates of all individuals $h \in R_j$, weighted by w_h. As $\beta_r \to \pm\infty$, the highest (lowest) observed value of a covariate x_r in each risk set gains more and more weight compared with the others. Therefore, the variance of x_r in each risk set and consequently the determinant of $I(\beta)$ approach zero so that, even if the likelihood L is monotone in β, the penalized likelihood L^* is guaranteed to attain its maximum at some finite value of $\hat{\beta}$.

Hence, the FC method completely eliminates the occurrence of monotone likelihood. Only those problems of estimation remain that can also occur with the general linear model, e.g., problems due to multicollinearity or nearly degenerate covariate distributions.

There are two alternatives by which Wald tests can be obtained. The first alternative is inserting the FC estimated $\hat{\beta}$ into the definition of the information matrix and then proceeding the usual way to get standard errors. This was suggested by Firth (1993, p. 36). The second alternative is evaluating the second derivative of the penalized log-likelihood function using numerical differentiation (e.g., by subroutine D04AAF of NAG (1998)) and then again proceeding the usual way to evaluate the information matrix and standard errors of $\hat{\beta}$. From our experience, differences in estimated standard errors according to both alternatives are negligible. Compared to SC, standard errors of $\hat{\beta}$ under FC are generally smaller and always finite. Penalization by Jeffreys prior tends to shift a parameter estimate toward a value where the determinant of the inverse covariance matrix is maximized or, loosely speaking, where variances of parameter estimates are minimized.

Independent of whether $\hat{\beta}$ is obtained by SC or FC, its distribution may be distinctly nonnormal and then likelihood ratio tests are preferable (cf., Cox and Oakes, 1984, p. 34–37). In our case, the likelihood ratio statistic LR is defined

by $LR = 2\{\log L(\hat{\gamma}, \hat{\delta})^* - \log L(\gamma_0, \hat{\delta}_{\gamma_0})^*\}$, where $(\hat{\gamma}, \hat{\delta})$ is the joint penalized maximum likelihood estimate of $\beta = (\gamma, \delta)$, the hypothesis of $\gamma = \gamma_0$ being tested, and $\hat{\delta}_{\gamma_0}$ is the penalized maximum likelihood estimate of δ when $\gamma = \gamma_0$. The values of the profile of the penalized log-likelihood function for γ, $\log L(\gamma, \hat{\delta}_\gamma)^*$, are obtained by fixing γ at predefined values around $\hat{\gamma}$, $\hat{\delta}_\gamma$ denoting penalized maximum likelihood estimates of δ for γ fixed at the predefined values. A profile likelihood $(1-\alpha)100\%$ confidence interval for a scalar parameter γ is the continuous set of values γ_0 for which LR does not exceed the $(1-\alpha)100$th percentile of the χ^2_1-distribution. In Section 4, the profile of the penalized likelihood will be used to judge the adequacy of Wald tests.

3. An Empirical Study

The empirical performance of the standard fitting procedure from Cox's model (SC) and of the previously presented Firth-type fitting (FC) was explored by a comprehensive Monte Carlo study.

While FC has been defined in the previous section, fitting by SC follows option (d) of Section 1 as implemented by procedure PHREG of SAS/STAT (SAS, 1999).

The effect of the following factors on bias of parameter estimates and on the coverage probability of one-sided lower (extending to $-\infty$) and upper (extending to $+\infty$) 97.5% confidence intervals was investigated in a factorial design, generating 1000 samples for each cell: sample size n (50, 100, 200), number of independent dichotomous covariates k (5, 15), expected percentage of censored survival times %c (0, 50, 90), identical relative risk R associated with each covariate (1, 2, 4, 16, 64), and identical degree of balance B of each covariate (1:1, 1:4).

Covariate values X_r $(1 \le r \le k)$ were sampled using the uniform random number generator G05CAF of NAG (1998) and exponentially distributed survival times with hazards $\exp(-\Sigma X_r\beta_r)$ and β_r set to 0, log 2, log 4, log 16, and log 64, respectively, were obtained using G05DBF of NAG (1998). In some experiments, the generated survival times were subjected to administrative censoring using the model of a medical study. Individuals were assumed to enter the study at a constant rate in an interval $(0, \tau)$ and then to die according to the prescribed survival distribution. For each combination of k, %c, R, and B, a value of τ, the time of analysis, was determined to achieve expected 50 and 90% censoring of survival times.

While the complete numerical results of the Monte Carlo study are contained in a technical report (Heinze, 1999), the typical performance of SC and FC can already be understood by means of the results selected for Table 3. We learn that the bias of both FC and SC is relatively small unless R is high in the presence of high censoring. The bias generally gets smaller with increasing n.

There is a small but clear advantage of FC over SC in situations of high R and high censoring but rare occurrence of monotone likelihood. This bias-reducing property had been the original target of Firth's (1993) adaptation and is now empirically confirmed also for Cox regression. If monotone likelihood occurs, the estimates by FC are quite satisfactory in an absolute sense and even more so if compared with estimates by SC. The latter arrive at far too high parameter estimates.

The empirical coverage (under FC) by one-sided 97.5% confidence intervals of Wald type and by those based on the profile penalized likelihood was equally satisfactory for low values of R and low censoring. For situations where monotone likelihood occurs, the profile of the penalized likelihood function becomes highly unsymmetric (as will be illustrated by Figure 1 in the next section) and therefore Wald tests and confidence intervals become unsuitable. This is reflected in one-sided coverage probabilities substantially departing from 97.5% (e.g., 90 or 100%) in situations where monotone likelihood is likely to appear. However, in these cases, the corresponding coverage by profile penalized likelihood confidence intervals is more satisfactory. Empirical results given by Table 4 indicate the general appropriateness of profile penalized likelihood confidence intervals.

Summarizing, the study confirmed the safe use of FC generally and its clear superiority over SC particularly in situations of high censoring and high parameter values. Particularly for such situations, inference should be based on penalized likelihood ratio tests and profile penalized likelihood confidence intervals rather than on Wald-type methods.

Table 3

Average bias (×100) *of estimated parameter values in Cox regression using SC/FC*

			$100\beta = 0$ $(R = 1)$			$100\beta = 69$ $(R = 2)$			$100\beta = 139$ $(R = 4)$		
n	k	B	%c = 0	%c = 50	%c = 90	%c = 0	%c = 50	%c = 90	%c = 0	%c = 50	%c = 90
50	5	1:1	0/0	4/4	32/4	5/4	11/8	$202^+/4$	15/12	20/13	$503^+/3$
50	5	1:4	4/7	10/4	$376^+/16$	1/3	52/0	$689^+/33$	15/7	135/2	$1008^+/68$
100	5	1:1	1/1	0/0	2/1	4/4	4/3	50/5	6/5	7/5	$146^+/8$
100	5	1:4	1/3	0/3	$190^+/5$	3/1	8/1	$487^+/10$	4/1	16/2	$743^+/35$
200	5	1:1	0/0	1/1	2/1	1/1	1/1	4/1	3/3	3/2	22/3
200	5	1:4	1/1	1/1	37/1	0/1	2/1	$151^+/4$	3/2	5/0	$420^+/12$
200	15	1:1	0/0	1/1	1/1	4/4	7/6	11/2	6/5	12/10	67/29
200	15	1:4	1/2	2/0	30/0	4/3	6/2	137/3	6/4	11/6	$309^+/1$

Each entry is based on 1000 samples. $^+$ denotes experimental conditions where >50% of the samples produced monotone likelihood. n, k, B, %c, and R denote sample size, number and degree of balance of dichotomous covariates, expected percentage of censored survival times, and relative risk, respectively.

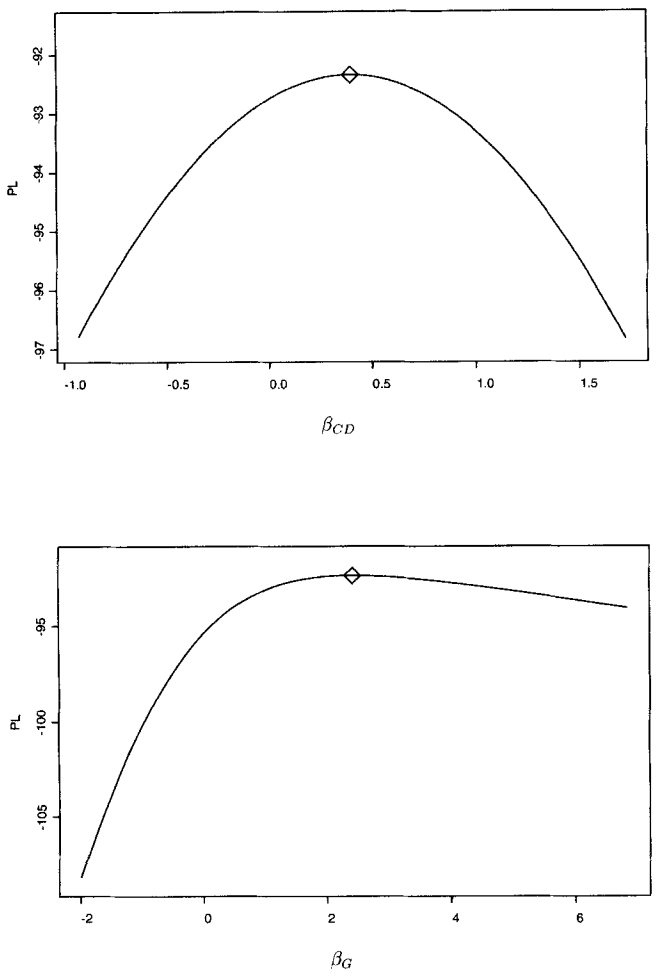

Figure 1. Profiles of the penalized log-likelihood function (PL) for factors CD (top) and G (bottom). The functions were obtained by fixing the investigated parameters, β_{CD} and β_G, at 100 predefined values evenly spread within ± 3 standard errors ($\hat{\sigma}(\beta_{CD}) = 0.44$, $\hat{\sigma}(\beta_G) = 1.47$) of the point estimates ($\hat{\beta}_{CD} = 0.40$, $\hat{\beta}_G = 2.43$), denoted by $\diamondsuit$.

4. Example and Further Aspects of Application

We now return to the breast cancer study introduced in Section 1. By means of the results given in Table 2, we have seen that current options of analysis in the presence of monotone likelihood (changing to a different type of model, omission of G from the model, stratification of the analysis on G, standard analysis with parameter estimate $\hat{\beta}_G$ set to a high value) are unsatisfactory. Reanalysis of the data set by FC leads to point estimates for pT, N, and CD that are slightly smaller than those by SC, as anticipated. The relative risk of 11.3 for G—contrasted with 13543327 by SC—is a plausible and well-communicable result.

The p-values from Wald and likelihood ratio tests agree well for factors pT, N, and CD but differ substantially for the large effect of G. Exploration of the profiles of the penalized log-likelihood function reveals approximately normal shapes for pT, N, and CD, while the shape for factor G is distinctly nonnormal (see Figure 1). This explains the failure of the Wald test to declare the strongest effect of this study as significant while the weaker effects of pT and N are significant at a level of $\alpha = 0.05$. With increasing parameter values, distributions of parameter estimates tend to become nonnormal and then likelihood ratio tests become preferable. Similarly, two-sided 95% confidence intervals according to Wald and according to the profile penalized likelihood are close for the effect of CD (0.63–3.54 and 0.63–3.51, respectively) but differ substantially for the effect of G (0.63–203 versus 1.47–1451).

Thus, application of FC and use of penalized likelihood ratio tests have provided better information about the risk factors of this study than the previously available procedures.

By means of the example, we have demonstrated how analysis should proceed if monotone likelihood is encountered. But should the FC procedure replace SC generally? Our empirical results indicated that the bias of parameter estimates in Cox models tends to be small unless unusually small samples with substantial censoring and several risk factors are to be analyzed. Therefore, we recommend the FC procedure in practice only for such samples and, of course, if monotone likelihood is occurring.

Table 4

Coverage probability ($\times 100$) *of one-sided left/right* 97.5% *profile penalized likelihood confidence intervals for Cox regression parameters using FC*

			$100\beta = 0$ ($R = 1$)			$100\beta = 69$ ($R = 2$)			$100\beta = 139$ ($R = 4$)		
n	k	B	$\%c = 0$	$\%c = 50$	$\%c = 90$	$\%c = 0$	$\%c = 50$	$\%c = 90$	$\%c = 0$	$\%c = 50$	$\%c = 90$
50	5	1:1	96/97	98/96	98/98	96/97	95/97	97/96	96/97	98/96	98/98
50	5	1:4	97/95	98/95	99/96	96/96	98/95	100/94	97/95	98/95	99/96
100	5	1:1	97/97	98/97	97/98	97/97	97/97	98/97	97/97	98/97	97/98
100	5	1:4	98/96	98/96	99/96	97/97	98/96	100/95	98/96	98/96	99/96
200	5	1:1	97/98	97/98	97/98	97/98	97/99	98/96	97/98	97/98	97/98
200	5	1:4	98/96	98/97	99/97	98/97	98/97	99/97	98/96	98/97	99/97
200	15	1:1	97/97	96/97	97/98	96/98	96/98	96/97	97/97	96/97	97/98
200	15	1:4	97/96	98/96	98/97	96/97	96/98	99/97	97/96	98/96	98/97

Each entry is based on 1000 samples. n, k, B, $\%c$, and R denote sample size, number and degree of balance of dichotomous covariates, expected percentage of censored survival times, and relative risk, respectively.

Monotone likelihood and the unavailability of suitable methods to cope with it previously severely limited the design and interpretation of small-sample simulation experiments for Cox regression (Loughin, 1998). Therefore, the suggested procedure may have strong impact on the design of future Monte Carlo studies of Cox's model. For similar reasons, it may also increase the applicability of bootstrap techniques to Cox regression for small samples.

Application of the FC approach including plots of profiles of the penalized log-likelihood function and corresponding likelihood ratio tests and confidence intervals are facilitated by a program *FC* available on request.

RÉSUMÉ

On est en présence d'un phénomène de vraisemblance monotone, lors de l'ajustement par un modèle de Cox, lorsque la vraisemblance tend vers une limite finie alors que l'un au moins des paramètres tend vers plus ou moins l'infini. Ce phénomène de vraisemblance monotone survient principalement dans des échantillons de petite taille où le processus de censure est important et plusieurs covariables sont fortement prédictives. Les possibilités existantes jusqu'à ce jour pour traiter les cas de vraisemblance monotone n'ont pas donné satisfaction. Nous suggérons une solution qui est une adaptation de la procédure que Firth (1993, *Biometrics* **80,** 27–38) a initialement développé dans le but de réduire le biais des estimateurs du maximum de vraisemblance. Cette procédure fournit des estimateurs bornés par les moyens d'une estimation pénalisée du maximum de vraisemblance. Les test de Wald et intervalle de confiance correspondant à ces estimateurs existent, mais on montre que le test du rapport des vraisemblances pénalisées et les intervalles de confiance calibrés pour la vraisemblance pénalisée sont souvent préférables. Une étude empirique de la procédure que nous suggérons, confirme des performances satisfaisantes tant pour l'estimation que pour l'inférence. Finalement, l'avantage de cette procédure par rapport aux précédentes possibilités est démontré par l'exemple de l'analyse d'une étude sur le cancer du sein.

REFERENCES

Bryson, M. C. and Johnson, M. E. (1981). The incidence of monotone likelihood in the Cox model. *Technometrics* **23,** 381–383.

Collett, D. (1994). *Modelling Survival Data in Medical Research.* London: Chapman and Hall.

Cordeiro, G. M. and Cribari-Neto, F. (1998). On bias reduction in exponential and non-exponential family regression models. *Communications in Statistics—Simulation and Computation* **27,** 485–500.

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B* **53,** 629–643.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

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

Firth, D. (1992a). Bias reduction, the Jeffreys prior and GLIM. In *Advances in GLIM and Statistical Modelling,* L. Fahrmeir, B. Francis, R. Gilchrist, and G. Tutz (eds), 91–100. New York: Springer-Verlag.

Firth, D. (1992b). Generalized linear models and Jeffreys priors: An iterative weighted least-squares approach. In *Computational Statistics,* Volume 1, Y. Dodge and J. Whittaker (eds), 553–557. Heidelberg: Physica-Verlag.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80,** 27–38.

Heinze, G. (1999). *The application of Firth's procedure to Cox and logistic regression.* Technical Report 10, Department of Medical Computer Sciences, Section of Clinical Biometrics, Vienna University, Vienna.

Jacobsen, M. (1989). Existence and unicity of MLEs in discrete exponential family distributions. *Scandinavian Journal of Statistics* **16,** 335–349.

Johnson, M. E., Tolley, H. D., Bryson, M. C., and Goldman, A. S. (1982). Covariate analysis of survival data: A small-sample study of Cox's model. *Biometrics* **38,** 685–698.

Leung, D. H.-Y. and Wang, Y.-G. (1998). Bias reduction using stochastic approximation. *Australian and New Zealand Journal of Statistics* **40,** 43–52.

Lösch, A., Tempfer, C., Kohlberger, P., Joura, E. A., Denk, M., Zajic, B., Breitenecker, G., and Kainz, C. (1998). Prognostic value of cathepsin D expression and association with histomorphological subtypes in breast cancer. *British Journal of Cancer* **78,** 205–209.

Loughin, T. M. (1998). On the bootstrap and monotone likelihood in the Cox proportional hazards regression model. *Lifetime Data Analysis* **4,** 393–403.

Marubini, E. and Valsecchi, M. G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies.* New York: John Wiley.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

NAG. (1998). *NAG Fortran Library Manual—Mark 18.* Oxford: Numerical Algorithms Group.

SAS. (1999). *SAS/STAT User's Guide,* Version 8. Cary, North Carolina: SAS Institute.

Schaefer, R. L. (1983). Bias correction in maximum likelihood logistic regression. *Statistics in Medicine* **2,** 71–78.

Received December 1999. *Revised July* 2000.
Accepted July 2000.