University of Vienna
Department of Medical Computer Sciences
Section of Clinical Biometrics
Section Head: Prof. M. Schemper
A-1090 VIENNA, Spitalgasse 23

Phone: (+43)(1) 40400/6688
Fax: (+43)(1) 40400/6687

e-mail: imc@akh-wien.ac.at

Technical Report 10/1999
updated in January 2001

The application of Firth's procedure to Cox and logistic regression

Georg HEINZE

e-mail: georg.heinze@akh-wien.ac.at

Abstract

The phenomenon of separation or monotone likelihood is observed in the fitting process of a logistic or a Cox model if the likelihood converges while at least one parameter estimate diverges to $\pm$ infinity. Monotone likelihood primarily occurs in small samples with several unbalanced and highly predictive covariates, and, for Cox regression, with a high percentage of censoring. A procedure by Firth (1993) originally developed to reduce the bias of maximum likelihood estimates is shown to provide an ideal solution to monotone likelihood (cf. Heinze & Schemper, 2001, 2000). It produces finite parameter estimates by means of penalized maximum likelihood estimation. Corresponding Wald tests and confidence intervals are available but it is shown that penalized likelihood ratio tests and profile penalized likelihood confidence intervals are often preferable.

This Technical Report presents the complete results of an extensive simulation study exploring the properties of Firth's procedure in logistic and Cox regression. The empirical bias of parameter estimates obtained by Firth's procedure is compared to that resulting from ordinary maximum likelihood estimation in logistic and Cox regression. For logistic regression, bias comparisons also include a Bayesian procedure proposed by Clogg et al. (1991) and exact logistic regression (Hirji, Tsiatis & Mehta, 1989). Results show the clear advantage of Firth's procedure over previous options of analysis. Furthermore, empirical coverage probabilities by Wald and profile penalized likelihood confidence intervals for Firth-type parameter estimates are presented.

Two new SAS macro programs, FL and FC, were written to facilitate the application of Firth's procedure to logistic and Cox regression, respectively. The present report contains the complete User's Guide to these macro programs including syntax, computational methods and examples.

Contents

1 Introduction

In logistic regression it has been recognized that with small to medium-sized data sets situations may arise where, although the likelihood converges, at least one parameter estimate is infinite (Albert & Anderson, 1984; Santner & Duffy, 1986; Lesaffre & Albert, 1989; Hirji, Tsiatis & Mehta, 1989; Clarkson & Jennrich, 1991; Kolassa, 1997). These situations occur if the responses and nonresponses can be perfectly separated by a single covariate or by a non-trivial linear combination of covariates. Therefore Albert & Anderson (1984) denoted such situations by *separation*. In general one does not assume infinite parameter values in underlying populations. An infinite estimate therefore can be regarded as extremely inaccurate, the inaccuracy being properly expressed by resulting Wald confidence intervals of infinite width (Lesaffre & Albert, 1989).

The problem of separation is by no means negligible and may occur even if the underlying model parameters are low. In Table 1 we show how the probability of separation depends on the sample size and on the number, strength and degree of balance of dichotomous covariates.

A similar problem ('*monotone likelihood*', cf. Bryson & Johnson, 1981) is familiar to statisticians who often apply Cox's (1972) model in biomedicine or who investigate small sample properties of the model by simulation. In general, one does not assume infinite parameter values in underlying populations. The problem of monotone likelihood is rather caused by a breakdown of the standard maximum likelihood method under special conditions in a sample. For a single covariate this occurs (cf. Tsiatis, 1981) when at each failure time the covariate value 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.

In Table 2 we show how the probability for the occurrence of monotone likelihood depends on sample size, on the proportion of censoring of survival times and on the strength and degree of balance of dichotomous covariates. 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.

In this report the properties of a procedure are explored which arrives at finite estimates for β by a modification of the score function of logistic regression. This modification was originally developed by Firth (1993, 1992a, 1992b) to reduce the bias of maximum likelihood estimates in generalized linear models. These estimates are biased

Table 1: The probability of separation (%) in logistic regression. Each entry is based on 1000 samples. n, k, B_X, and B_Y denote sample size, number and degree of balance of dichotomous covariates and degree of balance of outcome, respectively.

			B_X=1:1				B_X=1:4			
			Odds ratio				Odds ratio			
n	k	B_Y	1	2	4	16	1	2	4	16
30	3	1:1	0	3	10	53	17	25	43	74
	5		2	7	24	75	30	41	58	85
	10		12	38	78	98	56	71	86	98
50	3		0	0	1	18	2	5	15	46
	5		0	0	2	32	6	9	22	53
	10		0	1	20	78	10	19	36	74
100	3		0	0	0	2	0	1	1	10
	5		0	0	0	1	0	0	1	9
	10		0	0	0	17	0	0	1	11
30	3	1:4	16	21	33	61	58	30	29	55
	5		27	32	45	75	76	52	48	76
	10		58	68	86	98	92	82	89	98
50	3		3	4	7	28	34	10	5	25
	5		4	6	10	39	44	18	9	41
	10		9	12	34	86	67	34	35	76
100	3		0	0	0	6	5	1	0	5
	5		0	0	0	7	6	1	0	4
	10		0	0	0	22	14	2	0	10

Table 2: The probability of monotone likelihood (%) in Cox regression. Each entry is based on 1000 samples. n, k, B_X, and $\%c$ denote sample size, number and degree of balance of dichotomous covariates, and percentage of censoring, respectively.

			$B_X = 1:1$					$B_X = 1:4$				
			Hazard ratio					Hazard ratio				
n	k	$\%c$	1	2	4	16	64	1	2	4	16	64
50	5	0	0	0	0	0	12	0	0	0	0	20
		50	0	0	0	1	27	3	13	30	57	74
		90	23	38	67	92	98	78	95	99	100	100
	15	0	0	0	0	0	17	0	0	0	0	16
		50	0	0	1	10	41	6	23	38	49	69
100	5	0	0	0	0	0	1	0	0	0	0	2
		50	0	0	0	0	3	0	1	2	9	20
		90	4	11	30	64	83	45	80	96	100	100
	15	0	0	0	0	0	2	0	0	0	0	1
		50	0	0	0	0	7	0	1	2	2	7
200	5	0	0	0	0	0	0	0	0	0	0	0
		50	0	0	0	0	0	0	0	0	0	1
		90	0	0	4	16	37	7	35	73	100	100
	15	0	0	0	0	0	0	0	0	0	0	0
		50	0	0	0	0	0	0	0	0	0	0
		90	0	0	2	9	38	17	68	96	100	100

away from zero (cf. e. g. Schaefer, 1983; Cordeiro & McCullagh, 1991) and the occurrence of infinite parameter estimates in situations of separation can be interpreted as an extreme consequence of this property. As will be shown in § 2, Firth's procedure can also be applied to Cox regression, supplying finite estimates in case of a breakdown of the maximum likelihood principle.

In the following section we first review some principle ideas of Firth (1993), then focus on their implementation in logistic regression (FL) and Cox regression (FC), and suggest confidence intervals based on the profile penalized likelihood. In § 3, the empirical performance of Firth's procedure is explored for Cox regression in comparison with maximum likelihood (Cox regression), and for logistic regression in comparison with maximum likelihood, exact logistic regression (Hirji et al., 1989) and a Bayesian logistic regression procedure suggested by Clogg, Rubin, Schenker, Schultz & Weidman (1991). In § 4, two SAS macro programs are introduced that facilitate application of Firth's procedure to Cox and logistic regression. Use of the macro programs is exemplified by some data sets.

2 Firth's modified score procedure

2.1 General considerations

Maximum likelihood estimates of regression parameters β_r ($r = 1, \ldots, k$) are obtained as solutions to the score equations $\partial \log L/\partial\beta_r \equiv U(\beta_r) = 0$ where L is the likelihood function. The small sample bias of these estimates results from the combination of curvature and unbiasedness of the score function as was pointed out by Firth (1993). He suggested to base estimation on modified score equations

$$U(\beta_r)^* \equiv U(\beta_r) + 0.5 \text{ trace}\{I(\beta)^{-1}[\partial I(\beta)/\partial\beta_r]\} = 0 \qquad (r = 1, \ldots, k) \tag{2·1}$$

where $I(\beta)^{-1}$ is the inverse of the information matrix evaluated at β. The modified score function $U(\beta)^*$ 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. By using this modification, Firth (1993) showed that the $O(n^{-1})$ bias of maximum likelihood estimates $\hat\beta$ is removed.

2.2 Application to logistic regression

If this idea is applied to a logistic model $\text{pr}(y_i = 1|x_i, \beta) = \pi_i = [1 + \exp(-\sum_{r=1}^{k} x_{ir}\beta_r)]^{-1}$, then the score function $U(\beta_r) = \sum_{i=1}^{n} (y_i - \pi_i)\, x_{ir}$ is replaced by the modified score function

$$U(\beta_r)^* = \sum_{i=1}^{n} [y_i - \pi_i + h_i\,(1/2 - \pi_i)]\, x_{ir} \qquad (r = 1, \ldots, k) \tag{2·2}$$

where the h_i's are the i-th diagonal elements of the 'hat' matrix $H = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}$, with $W = \text{diag}\{\pi_i(1 - \pi_i)\}$. The resulting Firth-type estimates (FL) can also be obtained by performing ordinary maximum likelihood estimation on iteratively modified data formed by splitting each original observation i into two new observations having response values y_i and $1 - y_i$ with weights $1 + h_i/2$ and $h_i/2$, respectively. The contribution of the new observations to the score function is $[(y_i - \pi_i)(1 + h_i/2) + (1 - y_i - \pi_i)h_i/2]\, x_{ir} = [y_i - \pi_i + h_i(1/2 - \pi_i)]\, x_{ir}$. Therefore standard logistic regression programs applied to such iteratively modified data will yield estimates defined by (2·1). The splitting of each original observation into a response and a nonresponse guarantees finite estimates. If data are summarized into g groups of distinct covariate vectors and represented in 'y_i out of m_i' notation where now $i = 1, \ldots, g$, then FL estimates and their variance matrix can be computed using standard software by iteratively adding $h_i(1/2 - \pi_i)$ to y_i (cf. Firth, 1992a; p. 95).

Estimation of standard errors can be based on the roots of the diagonal elements of $I(\hat\beta)^{-1}$, which is a first–order approximation to $[-\partial^2 \log L^*/(\partial\beta)^2]^{-1}$ (cf. Firth, 1993;

p. 36). Alternatively, one may base estimation of standard errors on minus the second derivative of the modified score function, treating the h_i's as fixed, resulting in standard errors that may be slightly too low. From our experience the difference between both methods is negligibly small.

As FL estimates typically will be lower in absolute value than maximum likelihood (ML) estimates, their variance will be reduced as well, so the bias reduction should not lead to any substantial loss in power.

Under the following condition FL and the procedure by Clogg et al. (1991) formally agree: for a saturated model where $g = k$, g being the number of distinct covariate patterns, and with $\bar{p} = \sum_{i=1}^{n} y_i/n = 0.5$. In that case $h_i = 1$ $(i = 1, \ldots, g)$ because $0 \leq h_i \leq 1$ and $\sum_i h_i = k$ (cf. e. g. Belsley, Kuh & Welsch, 1980; pp. 66–67), and therefore both adjustments, $h_i/2$ and $\bar{p}\ k/g$, assume the value of 1/2.

2.3 Application to Cox regression

For applying Firth's procedure to Cox regression we only need the derivatives of the information matrix used in the definition of the score modification term 0.5 trace$\{I(\beta)^{-1}[\partial I(\beta)/\partial \beta_r]\} \equiv a_r$.

Let m denote the number of distinct survival times $t_{(j]}$ $(j = 1, \ldots, m)$ among the n survival times t_i $(i = 1, \ldots, n)$, $x_i = (x_{i1}, \ldots, x_{ir}, \ldots, x_{ik})$ a covariate vector related to each individual, d_j the number of deaths at $t_{(j)}$, s_j the vector sum of the covariates of the d_j individuals (s_{ij} referring to the r–th component of s_j), and R_j the set of individuals alive and uncensored prior to $t_{(j)}$. Then the log likelihood and its first, second and third derivatives are defined as follows:

$$
\begin{aligned}
\log L(\beta) &= \sum_{j=1}^{m} \left[\beta s_j - d_j \log\left\{\sum\nolimits_{h\in R_j} \exp(x_h\beta)\right\}\right] \\
\frac{\partial \log L(\beta)}{\partial \beta_r} &= \sum_{j=1}^{m} \left\{ s_{jr} - \frac{d_j \sum_{h\in R_j} x_{hr} \exp(x_h\beta)}{\sum_{h\in R_j} \exp(x_h\beta)} \right\} \\
\frac{\partial^2 \log L(\beta)}{\partial \beta_r \partial \beta_s} &= -\sum_{j=1}^{m} d_j \left[\frac{\sum_{h\in R_j} x_{hr} x_{hs} \exp(x_h\beta)}{\sum_{h\in R_j} \exp(x_h\beta)} - \right. \\
&\quad \left. - \frac{\left\{\sum_{h\in R_j} x_{hr} \exp(x_h\beta)\right\} \left\{\sum_{h\in R_j} x_{hs} \exp(x_h\beta)\right\}}{\left\{\sum_{h\in R_j} \exp(x_h\beta)\right\}^2} \right] \\
\frac{\partial^3 \log L(\beta)}{\partial \beta_r \partial \beta_s \partial \beta_t} &= -\sum_{j=1}^{m} d_j \left\{ \left(\frac{S_{j,rst}}{S_{j,0}} - \frac{S_{j,rs} S_{j,t}}{S_{j,0}^2} \right) - \frac{S_{j,s}}{S_{j,0}} \left(\frac{S_{j,rt}}{S_{j,0}} - \frac{S_{j,r} S_{j,t}}{S_{j,0}^2} \right) - \right. \\
&\quad \left. - \frac{S_{j,r}}{S_{j,0}} \left(\frac{S_{j,st}}{S_{j,0}} - \frac{S_{j,s} S_{j,t}}{S_{j,0}^2} \right) \right\}
\end{aligned}
$$

where

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

When a model based on the modified score function is fitted by the Newton-Raphson algorithm (see e.g. Collett, 1994; pp. 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. There are two alternatives by which Wald tests can be obtained: First, by inserting the FC estimated $\hat{\beta}$ into the definition of the information matrix and then proceeding the usual way to get standard errors. This way was suggested by Firth (1993, p. 36). Second, by 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.

2.4 Profile likelihood confidence intervals

Independent of whether $\hat{\beta}$ is obtained by ML or by Firth's modified score procedure for Cox or for logistic regression its distribution may be distinctly nonnormal and then likelihood ratio tests are preferable. In our case the likelihood ratio statistic LR is defined by $LR = 2\left[\log L(\hat{\gamma}, \hat{\delta})^* - \log L(\gamma_0, \hat{\delta}_{\gamma_0})^*\right]$, 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 $\gamma = \gamma_0$. Note that for evaluating the penalized likelihood at $(\gamma_0, \hat{\delta}_{\gamma_0})$, the actual value of γ_0 has to be used for the computation of the penalization term $|I(\beta)|^{0.5}$, even if $\gamma_0 = 0$.

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.

3 Simulation study

3.1 Logistic regression

A simulation study was conducted to explore and compare the empirical performance of standard maximum likelihood fitting (ML), Firth-type fitting (FL), and fitting with augmented data according to Clogg et al. (1991) (CL). We also include results from exact logistic regression (XL) for simulated scenarios where it might be an option.

FL has been defined in the previous section, and parameter estimation was carried out using the SAS macro program `%FL` introduced in § 4. Fitting by ML follows implementation of procedure LOGISTIC of SAS/STAT (1999b), i. e. up to 25 iterations are carried out until parameter values converge. However, the algorithm is stopped if quasicomplete or complete separation are detected as follows: complete separation is declared if the probability of belonging to the correct response group is one for all observations, and quasicomplete separation is declared if an observation is identified to have an extremely large probability of correct allocation, and any of the diagonal elements of the dispersion matrix for the standardized observations vectors (all independent variables standardized to zero mean and unit variance) exceeds 5000.

XL models were obtained by first employing the efficient algorithm described by Hirji et al. (1987) to compute the exact null distribution of a sufficient statistic, and then by computing exact maximum likelihood or median unbiased estimates for non-separated and separated data sets, respectively.

The effect of the following factors on the 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 (30, 50, 100), number of binary covariates $k-1$ (3, 5, 10), identical odds ratio $\exp(\beta)$ associated with each covariate (1, 2, 4, 16) and identical degree of balance B_X of each covariate (1:1, 1:4). For each cell an intercept parameter β_1 was determined to obtain a balance of responses and nonresponses B_Y of 1:1 and of 1:4.

Covariate values x_{ir} ($i = 1, \ldots, n$; $r = 2, \ldots, k$) were sampled using random number generator RANBIN of SAS Institute (1999a), and x_{i1} was fixed to 1. Outcomes y_i were sampled from a binomial distribution with parameter $\pi_i = [1 + \exp(-\sum_{r=1}^{k} x_{ir}\beta)]^{-1}$ again using RANBIN.

The typical performance of the investigated procedures can be understood by means of results contained in Tables 3–4. We learn that the bias of $\hat{\beta}_2$ with FL is small, that it is larger with both CL and XL, and that it is largest with ML. In agreement with Hirji et al. (1989) there is a certain proportion of simulated data sets where XL parameter estimates are unavailable due to degenerate conditional distributions (see Table 5 for results).

Table 3: Average bias × 100 of parameter estimates in logistic regression. Each entry is based on 1000 samples. n, k, and B_X denote sample size, number and degree of balance of dichotomous covariates, respectively. The expected marginal balance of responses and nonresponses is fixed at 1:1. The β's of 0, 0.39, 1.69 and 2.77 correspond to odds ratios of 1, 2, 4, and 16, respectively.

			B_X=1:1				B_X=1:4			
			100β				100β			
n	k	Method	0	39	169	277	0	39	169	277
30	3	ML	-4	32	102	566	-7	88	186	424
		FL	-3	1	1	-6	-2	-1	-5	-19
		CL	-2	-1	-8	-48	-2	-7	-21	-63
		XL	-3	4	6	-35	-2	-2	-10	-42
	5	ML	4	60	274	1115	3	87	224	783
		FL	2	4	3	-24	-2	-3	-9	-36
		CL	2	-2	-19	-99	-2	-9	-28	-94
		XL	2	6	-5	-97	-2	-5	-19	-88
	10	ML	-27	574	1118	1168	-8	326	897	1292
		FL	0	3	-23	-130	2	8	-6	-89
		CL	-2	-15	-56	-172	2	-3	-34	-140
50	3	ML	0	7	24	190	-2	28	74	218
		FL	0	0	1	0	-2	0	-1	-5
		CL	0	-1	-4	-27	-2	-6	-16	-47
		XL	0	3	7	6	-2	4	5	-4
	5	ML	0	12	38	456	1	27	84	433
		FL	0	1	0	0	1	-1	-2	-5
		CL	0	-2	-13	-67	1	-7	-20	-74
		XL	0	3	8	-15	2	2	2	-21
	10	ML	-2	47	488	1746	7	76	267	1346
		FL	-1	1	2	-46	3	6	9	-19
		CL	-1	-8	-38	-142	2	-1	-19	-108
100	3	ML	0	4	10	34	1	5	9	58
		FL	0	1	2	2	1	-2	-3	-1
		CL	0	0	0	-11	1	-6	-13	-30
	5	ML	1	3	11	54	2	4	11	89
		FL	1	-2	-1	0	2	-2	-3	2
		CL	1	-3	-7	-37	2	-6	-16	-56
	10	ML	1	11	34	429	1	15	32	233
		FL	1	0	2	8	1	3	4	5
		CL	1	-3	-19	-97	1	1	-10	-71

Table 4: Average bias × 100 of parameter estimates in logistic regression *(cont)*. Each entry is based on 1000 samples. n, k, and B_X denote sample size, number and degree of balance of dichotomous covariates, respectively. The expected marginal balance of responses and nonresponses is fixed at 1:4. The β's of 0, 0.39, 1.69 and 2.77 correspond to odds ratios of 1, 2, 4, and 16, respectively.

			B_X=1:1				B_X=1:4			
			100β				100β			
n	k	Method	0	39	169	277	0	39	169	277
30	3	ML	-12	78	204	619	-247	-62	131	553
		FL	-5	-5	-2	-23	7	1	2	-22
		CL	-5	-1	-2	-54	-18	-16	-15	-64
		XL	-5	-5	-5	-60	22	3	-5	-61
	5	ML	19	204	441	1123	-232	1	297	936
		FL	4	7	0	-39	6	1	-2	-39
		CL	4	9	-12	-98	-15	-12	-17	-87
		XL	3	0	-19	-111	26	5	-17	-106
	10	ML	57	599	1108	1071	-174	356	907	1259
		FL	-1	-10	-31	-139	6	-4	-15	-100
		CL	-1	-16	-54	-173	-9	-13	-36	-140
50	3	ML	-1	19	50	307	-148	-40	18	265
		FL	-2	-3	0	-3	2	-3	0	-6
		CL	-2	-1	0	-29	-17	-17	-17	-55
		XL	-2	1	8	-5	1	-5	2	-11
	5	ML	2	34	103	573	-125	-23	87	507
		FL	2	4	8	1	5	4	6	-11
		CL	2	5	-3	-65	-11	-8	-11	-70
		XL	2	9	14	-22	6	4	6	-33
	10	ML	-3	131	695	1747	-95	119	539	1550
		FL	1	-4	-2	-49	5	5	9	-31
		CL	0	-9	-36	-140	-7	-4	-17	-109
100	3	ML	-1	3	12	90	-21	-4	9	74
		FL	-1	-2	1	5	3	0	3	4
		CL	-1	-1	1	-11	-5	-7	-8	-34
	5	ML	0	9	18	124	-20	5	16	77
		FL	0	2	2	7	2	4	3	-1
		CL	0	3	-2	-35	-5	-3	-9	-51
	10	ML	1	12	56	504	-25	6	38	320
		FL	0	-3	1	4	-1	-1	4	7
		CL	0	-4	-18	-97	-7	-4	-8	-69

Table 5: Probability (%) of degenerate conditional distributions in exact logistic regression. Each entry is based on 1000 samples. n, k, B_X, and B_Y denote sample size, number and degree of balance of dichotomous covariates, and expected marginal balance of responses and nonresponses, respectively.

			B_X=1:1				B_X=1:4			
			Odds ratio				Odds ratio			
n	k	B_Y	1	2	4	16	1	2	4	16
30	3	1:1	0	0	0	0	0	0	0	0
	5		0	0	0	2	0	0	0	2
50	3		0	0	0	0	0	0	0	0
	5		0	0	0	0	0	0	0	0
30	3	1:4	0	0	0	0	0	0	0	1
	5		1	1	1	3	1	1	2	4
50	3		0	0	0	0	0	0	0	0
	5		0	0	0	0	0	0	0	1

The empirical coverage under FL by one–sided 97.5% confidence intervals of Wald type and by those based on the profile penalized likelihood was equally satisfactory for low odds ratios and balanced covariates (Tables 6–9). For situations where separation occurs the profile of the penalized likelihood function becomes highly unsymmetric. Therefore Wald tests and confidence intervals become unsuitable. This is reflected in one–sided coverage probabilities substantially departing from 97.5% in situations with a high probability of separation. In these cases coverage by profile penalized likelihood confidence intervals is much more satisfactory.

Summarizing, the study confirmed the safe use of FL in general and its clear superiority over ML particularly in situations of high parameter values and/or unbalanced covariates. Especially for such situations inference should be based on penalized likelihood ratio tests and profile penalized likelihood ratio confidence intervals rather than on Wald type methods.

Table 6: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in logistic regression using Firth's modified score procedure. Each entry is based on 1000 samples. n and k denote sample size and number of dichotomous covariates, respectively. Balance of covariates and balance of outcome are both fixed at 1:1.

		Wald				Profile penalized likelihood			
		Odds ratio				Odds ratio			
n	k	1	2	4	16	1	2	4	16
30	3	99/99	98/99	99/99	95/100	98/98	98/97	98/98	96/100
	5	99/98	98/99	99/98	95/100	97/97	97/98	97/97	96/100
	10	99/99	99/100	99/99	92/100	98/97	98/99	98/97	97/100
50	3	98/98	99/99	98/98	96/100	97/98	98/97	97/98	97/99
	5	98/98	98/99	98/98	97/100	96/97	98/98	96/97	98/99
	10	98/99	99/99	98/99	94/100	97/98	98/98	97/98	98/100
100	3	98/97	97/97	98/97	96/100	98/97	97/97	98/97	97/98
	5	98/98	98/98	98/98	97/100	98/98	97/98	98/98	98/98
	10	99/98	98/99	99/98	96/100	98/97	98/98	98/97	97/99

Table 7: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in logistic regression using Firth's modified score procedure *(cont)*. Each entry is based on 1000 samples. n and k denote sample size and number of dichotomous covariates, respectively. Balance of covariates and balance of outcome are fixed at 1:4 and 1:1, respectively.

		Wald				Profile penalized likelihood			
		Odds ratio				Odds ratio			
n	k	1	2	4	16	1	2	4	16
30	3	100/100	99/100	100/100	97/100	97/97	97/99	97/97	97/100
	5	100/100	100/100	100/100	97/100	98/98	98/98	98/98	96/100
	10	100/99	99/100	100/99	96/100	97/98	98/98	97/98	97/100
50	3	99/99	98/99	99/99	97/100	98/98	97/98	98/98	97/100
	5	99/100	99/99	99/100	97/100	97/98	98/98	97/98	97/99
	10	99/99	99/99	99/99	96/100	98/98	98/97	98/98	98/100
100	3	99/98	98/99	99/98	97/100	98/97	97/98	98/97	97/98
	5	99/99	98/99	99/99	97/100	98/98	98/98	98/98	98/99
	10	98/98	98/98	98/98	97/99	97/97	98/97	97/97	98/98

Table 8: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in logistic regression using Firth's modified score procedure *(cont)*. Each entry is based on 1000 samples. n and k denote sample size and number of dichotomous covariates, respectively. Balance of covariates and balance of outcome are fixed at 1:1 and 1:4, respectively.

		Wald				Profile penalized likelihood			
		Odds ratio				Odds ratio			
n	k	1	2	4	16	1	2	4	16
30	3	99/99	99/100	99/99	96/100	98/98	97/98	98/98	97/100
	5	99/100	99/100	99/100	95/100	98/97	98/99	98/97	96/100
	10	100/100	100/100	100/100	92/100	99/99	99/100	99/99	98/100
50	3	99/99	98/99	99/99	96/100	98/98	98/98	98/98	97/100
	5	98/99	98/99	98/99	96/100	97/98	98/98	97/98	96/100
	10	99/99	97/100	99/99	92/100	97/98	97/99	97/98	97/100
100	3	98/98	97/98	98/98	96/100	97/98	97/97	97/98	97/99
	5	98/99	98/99	98/99	97/99	98/98	97/98	98/98	97/98
	10	98/98	98/99	98/98	95/100	98/98	97/99	98/98	97/99

Table 9: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in logistic regression using Firth's modified score procedure *(cont)*. Each entry is based on 1000 samples. n and k denote sample size and number of dichotomous covariates, respectively. Balance of covariates and balance of outcome are both fixed at 1:4.

		Wald				Profile penalized likelihood			
		Odds ratio				Odds ratio			
n	k	1	2	4	16	1	2	4	16
30	3	100/99	100/100	100/99	98/100	100/97	99/98	100/97	97/100
	5	100/99	100/100	100/99	96/100	99/98	98/97	99/98	96/100
	10	100/100	100/100	100/100	97/100	99/98	98/99	99/98	98/100
50	3	100/98	100/99	100/98	97/100	99/98	99/98	99/98	97/100
	5	100/99	100/99	100/99	97/100	99/98	98/97	99/98	97/100
	10	100/98	100/99	100/98	96/100	99/97	98/98	99/97	98/100
100	3	100/96	99/98	100/96	96/100	98/96	97/98	98/96	97/98
	5	100/98	99/98	100/98	96/100	98/98	98/97	98/98	97/98
	10	99/98	99/99	99/98	97/100	98/97	98/98	98/97	98/99

3.2 Cox regression

The performance of the standard fitting procedure for Cox's model (SC) and of the previously presented Firth-type fitting (FC) was explored and compared by a comprehensive Monte Carlo study. FC has been defined in the previous section, and paramter estimation was carried out using the program `FC` introduced in § 4. Fitting by SC follows the implementation by procedure PHREG of SAS/STAT (1999b), i. e. the iterative algorithm is stopped when the change in the log likelihood is less than 10^{-6}.

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 dichotomous covariate (1:1, 1:4).

Covariate values x_{ir} ($i = 1, \ldots, n; r = 1, \ldots, k$) were sampled using the uniform random number generator G05CAF of NAG (1998). Exponentially distributed survival times with hazards $\exp(-\sum_{r=1}^{k} x_i r \beta_r)$ 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 an expected 50% and 90% censoring of survival times.

From Tables 10–12 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.

Table 10: Average bias × 100 of parameter estimates in Cox regression. Each entry is based on 1000 samples. n, k, B_X, and $\%c$ denote sample size, number and degree of balance of dichotomous covariates, and percentage censoring, respectively. SC-FC and SC-conv denote standard Cox analysis with Firth's modification plugged in for samples subject to monotone likelihood and standard Cox analysis evaluated for samples not subject to monotone likelihood only, respectively. For SC-conv, 'NA' denotes situations where all simulated samples were subject to monotone likelihood. The β's of 0, 0.39, 1.69, 2.77, and 4.16 correspond to odds ratios of 1, 2, 4, 16, and 32, respectively.

				$B_X = 1:1$					$B_X = 1:4$				
				100β					100β				
n	k	$\%c$	Method	0	69	139	277	416	0	69	139	277	416
50	5	0	FC	0	4	12	13	11	-7	-3	7	13	9
			SC	0	5	15	25	209	-4	2	15	30	310
			SC-FC	0	5	15	23	36	-4	2	15	29	36
			SC-conv	0	5	15	23	19	-4	2	15	29	11
		50	FC	4	8	13	25	24	-4	0	-2	1	-24
			SC	4	11	20	60	440	10	52	135	308	770
			SC-FC	4	11	20	46	52	4	12	12	15	-11
			SC-conv	4	11	20	44	20	4	8	6	13	-24
		90	FC	-4	4	3	-19	-106	-16	-33	-68	-171	-295
			SC	-32	202	503	1359	1958	376	689	1008	1285	1221
			SC-FC	-4	12	12	-15	-105	-13	-32	-68	-171	-295
			SC-conv	1	2	-1	-26	-78	-32	-38	-90	NA	NA
	15	0	FC	0	18	28	54	65	-9	16	23	50	70
			SC	0	22	40	92	376	-6	24	37	84	364
			SC-FC	0	22	40	91	125	-6	24	37	84	128
			SC-conv	0	22	40	91	110	-6	24	37	84	113
		50	FC	0	24	64	222	353	-12	15	44	124	196
			SC	1	39	112	617	1496	3	60	150	458	1673
			SC-FC	1	39	110	376	478	-7	29	67	178	242
			SC-conv	1	39	107	299	372	-9	28	80	204	224

Table 11: Average bias × 100 of parameter estimates in Cox regression (*cont*). Each entry is based on 1000 samples. n, k, B_X, and $\%c$ denote sample size, number and degree of balance of dichotomous covariates, and percentage censoring, respectively. SC-FC and SC-conv denote standard Cox analysis with Firth's modification plugged in for samples subject to monotone likelihood and standard Cox analysis evaluated for samples not subject to monotone likelihood only, respectively. For SC-conv, 'NA' denotes situations where all simulated samples were subject to monotone likelihood. The β's of 0, 0.39, 1.69, 2.77, and 4.16 correspond to odds ratios of 1, 2, 4, 16, and 32, respectively.

				$B_X = 1:1$					$B_X = 1:4$				
				100β					100β				
n	k	$\%c$	Method	0	69	139	277	416	0	69	139	277	416
100	5	0	FC	1	4	5	6	11	-3	1	1	6	10
			SC	1	4	6	10	35	-1	3	4	12	54
			SC-FC	1	4	6	10	24	-1	3	4	12	26
			SC-conv	1	4	6	10	22	-1	3	4	12	22
		50	FC	0	3	5	10	14	-3	-1	-2	0	7
			SC	0	4	7	17	76	0	8	16	38	209
			SC-FC	0	4	7	17	34	0	5	8	16	35
			SC-conv	0	4	7	17	28	0	5	8	15	25
		90	FC	1	5	8	0	-20	-5	-10	-35	-123	-241
			SC	-2	50	146	550	1317	190	487	743	1180	1303
			SC-FC	1	13	20	12	-11	4	-5	-34	-123	-241
			SC-conv	1	9	12	3	-26	-2	-18	-54	NA	NA
	15	0	FC	0	7	15	23	31	-6	6	12	22	30
			SC	0	9	18	32	74	-4	9	16	31	67
			SC-FC	0	9	18	32	53	-4	9	16	31	52
			SC-conv	0	9	18	32	50	-4	9	16	31	49
		50	FC	-1	11	19	41	69	-1	7	13	31	56
			SC	-1	14	26	63	215	2	16	25	53	178
			SC-FC	-1	14	26	63	115	2	14	25	53	97
			SC-conv	-1	14	26	63	103	2	14	25	53	89

Table 12: Average bias × 100 of parameter estimates in Cox regression (*cont*). Each entry is based on 1000 samples. n, k, B_X, and $\%c$ denote sample size, number and degree of balance of dichotomous covariates, and percentage censoring, respectively. SC-FC and SC-conv denote standard Cox analysis with Firth's modification plugged in for samples subject to monotone likelihood and standard Cox analysis evaluated for samples not subject to monotone likelihood only, respectively. For SC-conv, 'NA' denotes situations where all simulated samples were subject to monotone likelihood. The β's of 0, 0.39, 1.69, 2.77, and 4.16 correspond to odds ratios of 1, 2, 4, 16, and 32, respectively.

				$B_X = 1:1$					$B_X = 1:4$				
				100β					100β				
n	k	$\%c$	Method	0	69	139	277	416	0	69	139	277	416
200	5	0	FC	0	1	3	5	5	-1	-1	2	4	4
			SC	0	1	3	6	10	-1	0	3	7	11
			SC-FC	0	1	3	6	10	-1	0	3	7	11
			SC-conv	0	1	3	6	10	-1	0	3	7	11
		50	FC	1	1	2	5	8	-1	-1	0	-2	-1
			SC	1	1	3	8	15	1	2	5	6	20
			SC-FC	1	1	3	8	15	1	2	5	6	14
			SC-conv	1	1	3	8	15	1	2	5	6	13
		90	FC	1	-1	3	6	3	1	-4	-12	-72	-178
			SC	2	4	22	90	408	37	151	420	986	1269
			SC-FC	2	3	12	24	27	11	8	-5	-71	-178
			SC-conv	2	3	11	21	5	9	-1	-24	-122	NA
	15	0	FC	0	4	5	11	16	-2	3	4	10	15
			SC	0	4	6	13	24	-1	4	6	13	22
			SC-FC	0	4	6	13	24	-1	4	6	13	22
			SC-conv	0	4	6	13	24	-1	4	6	13	22
		50	FC	1	6	10	23	29	0	2	6	18	26
			SC	1	7	12	29	45	2	6	11	25	41
			SC-FC	1	7	12	29	44	2	6	11	25	41
			SC-conv	1	7	12	29	44	2	6	11	25	41
		90	FC	1	2	29	149	344	0	-3	-1	-1	2
			SC	1	11	67	447	1277	30	137	309	652	1373
			SC-FC	1	10	60	264	465	8	3	0	-1	2
			SC-conv	1	10	58	229	300	7	2	10	NA	NA

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 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 much more satisfactory.

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

Table 13: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in Cox regression using Firth's modified score procedure. Each entry is based on 1000 samples. n, k and $\%c$ denote sample size, number of dichotomous covariates, and percentage censoring, respectively. Balance of covariates is fixed at 1:1.

			Wald					Profile penalized likelihood				
			Hazard ratio					Hazard ratio				
n	k	$\%c$	1	2	4	16	64	1	2	4	16	64
50	5	0	96/97	97/97	96/97	97/98	100/97	96/97	96/97	96/97	95/99	99/98
		50	98/97	96/97	98/97	97/98	100/96	98/96	95/97	98/96	94/98	99/97
		90	99/99	100/98	99/99	100/93	100/91	98/98	97/96	98/98	99/95	100/94
	15	0	94/95	92/98	94/95	90/99	96/99	94/94	91/97	94/94	87/99	91/99
		50	95/95	91/96	95/95	83/99	85/99	93/94	89/96	93/94	78/99	77/99
100	5	0	97/97	97/97	97/97	98/98	100/97	97/97	97/97	97/97	97/98	98/98
		50	98/98	97/97	98/98	97/98	100/97	98/97	97/97	98/97	96/98	97/97
		90	99/99	99/98	99/99	100/96	100/93	97/98	98/97	97/98	98/97	99/96
	15	0	97/95	96/98	97/95	93/99	97/98	97/95	95/97	97/95	92/99	94/99
		50	97/96	97/98	97/96	90/99	95/99	96/96	95/98	96/96	87/99	90/100
200	5	0	97/98	97/98	97/98	96/98	98/98	97/98	97/98	97/98	96/98	97/98
		50	98/99	97/99	98/99	97/98	98/97	97/98	97/99	97/98	96/98	97/98
		90	98/99	99/96	98/99	100/96	100/96	97/98	98/96	97/98	98/97	99/97
	15	0	97/97	96/98	97/97	94/99	96/98	97/97	96/98	97/97	93/99	95/99
		50	97/97	96/98	97/97	93/100	95/99	96/97	96/98	96/97	92/100	92/99
		90	98/98	97/97	98/98	87/98	87/98	97/98	96/97	97/98	82/99	80/99

Table 14: Coverage probability × 100 of one–sided left/right 97.5% confidence intervals in Cox regression using Firth's modified score procedure (*cont.*). Each entry is based on 1000 samples. n, k and $\%c$ denote sample size, number of dichotomous covariates, and percentage censoring, respectively. Balance of covariates is fixed at 1:4.

			Wald					Profile penalized likelihood				
			Hazard ratio					Hazard ratio				
n	k	$\%c$	1	2	4	16	64	1	2	4	16	64
50	5	0	98/94	98/94	98/94	97/97	100/96	97/95	96/96	97/95	96/97	99/97
		50	99/95	100/94	99/95	100/94	100/93	98/95	98/95	98/95	98/95	100/96
		90	100/97	100/94	100/97	100/86	100/67	99/96	100/94	99/96	100/91	100/95
	15	0	96/94	93/96	96/94	91/99	96/99	95/93	92/96	95/93	88/100	94/99
		50	96/93	94/96	96/93	88/98	92/98	95/92	92/96	95/92	84/99	86/99
100	5	0	98/95	98/96	98/95	98/97	100/96	98/96	97/97	98/96	97/98	98/97
		50	99/96	99/95	99/96	99/94	100/94	98/96	98/96	98/96	97/96	99/95
		90	100/97	100/94	100/97	100/85	100/82	99/96	100/95	99/96	100/91	100/93
	15	0	96/93	95/96	96/93	94/98	98/98	96/94	94/96	96/94	93/99	94/98
		50	98/94	97/95	98/94	92/98	95/98	97/94	95/96	97/94	89/99	91/99
200	5	0	98/95	98/96	98/95	99/98	99/97	98/96	98/97	98/96	98/98	98/97
		50	98/97	99/97	98/97	99/96	100/96	98/97	98/97	98/97	98/97	98/98
		90	100/97	100/95	100/97	100/89	100/88	99/97	99/97	99/97	100/94	100/89
	15	0	98/95	96/97	98/95	95/99	96/98	97/96	96/97	97/96	95/99	94/99
		50	98/96	97/97	98/96	94/98	96/99	98/96	96/98	98/96	93/98	94/99
		90	100/97	100/96	100/97	100/94	99/93	98/97	99/97	98/97	97/97	97/97

4 User's Guide to the SAS macros FL and FC

4.1 FL: Logistic regression

4.1.1 Overview

The SAS macro `%FL` was written to facilitate application of Firth's modified score procedure in logistic regression analysis. It is available at the WWW site `http://www.akh-wien.ac.at/imc/biometrie/fl`. Supplied with a SAS data set as input, its output contains FL–type logistic regression parameters, standard errors, confidence limits, p-values, the value of the maximized penalized log likelihood, and the number of iterations needed to arrive at the maximum. All output produced is stored in a SAS data set. Furthermore, parameter estimates, penalized log likelihood and covariance matrix are saved in a SAS data set that has the same structure as the output data set that can be obtained by SAS/PROC LOGISTIC (SAS Institute, 1999b) using the OUTEST option in the PROC LOGISTIC statement of that procedure. Efficient processing of multiple input data sets is possible through a BY variable, similarly to PROC LOGISTIC. Finally, the option of supplying offset values to the program enables computation of penalized likelihood ratio tests or exploration of penalized log likelihood profiles.

We exemplify use of the `%FL` SAS macro program by means of the analysis of an endometrial cancer study and thank Dr. E. Asseryanis from the Vienna General Hospital for providing the data set. Purpose of this study of 79 patients was to explain the state of the endometrium by the putative risk factors (=covariates) neovasculization (NV), pulsatility index of arteria uterina (PI), and endometrium height (EH). The state of the endometrium was histologically graded (HG) and classified as 0 (=0-II) and 1 (=III-IV) for 30 and 49 patients, respectively. NV is coded as 1 (present) for 13 patients and 0 (absent) for 66 patients. The two continuous covariates PI and EH range from 0 to 49 and from 0.27 to 3.61, with medians of 16 and 1.64, respectively. The data are given in Table 15.

Table 15: Endometrial cancer data set.

NV	PI	EH	HG	NV	PI	EH	HG	NV	PI	EH	HG
0	13	1.64	0	0	28	1.50	0	0	29	2.02	0
0	16	2.26	0	0	11	1.33	0	0	15	2.29	0
0	8	3.14	0	0	19	2.37	0	0	12	2.33	0
0	34	2.68	0	0	10	1.82	0	0	3	2.90	0
0	20	1.28	0	0	10	3.13	0	0	20	1.70	0
0	5	2.31	0	0	18	1.31	0	0	23	1.41	0
0	17	1.80	0	0	14	1.92	0	0	12	2.25	0
0	10	1.68	0	0	21	1.64	0	0	22	1.54	0
0	26	1.56	0	0	11	2.01	0	0	42	1.97	0
0	17	2.31	0	0	17	1.88	0	0	15	1.75	0
0	8	2.01	0	0	25	1.93	0	0	13	2.16	0
0	7	1.89	0	0	16	2.11	0	0	14	2.57	0
0	20	3.15	0	0	19	1.29	0	0	19	1.37	0
0	10	1.23	0	0	15	1.72	0	0	12	3.61	0
0	18	1.27	0	0	33	0.75	0	0	13	2.04	0
0	16	1.76	0	0	24	1.92	0	0	10	2.17	0
0	18	2.00	0	0	48	1.84	1	0	12	1.69	1
0	8	2.64	1	0	12	1.11	1	1	49	0.27	1
0	29	0.88	1	0	19	1.61	1	0	6	1.84	1
0	12	1.27	1	0	2	1.18	1	0	5	1.30	1
0	20	1.37	1	1	22	1.44	1	0	17	0.96	1
1	38	0.97	1	1	40	1.18	1	1	11	1.01	1
1	22	1.14	1	1	5	0.93	1	1	21	0.98	1
1	7	0.88	1	1	0	1.17	1	0	5	0.35	1
1	25	0.91	1	0	21	1.19	1	1	19	1.02	1
1	15	0.58	1	0	15	1.06	1	0	33	0.85	1
0	7	0.97	1								

Suppose the data has been stored in a SAS data set `bsp.endo`. To obtain an FL analysis with a table containing variable names, parameter estimates, standard errors, profile penalized likelihood confidence limits and p–values based on penalized likelihood ratio tests, the macro call

```
title "Analysis of endometrian cancer study";
%fl(data=bsp.endo, y=hg, varlist=nv pi eh);
```

will produce the following three output pages:

```
              Analysis of endometrian cancer study             1


  FFFFF L                Logistic regression
  F     L                with Firth's modified score function
  FFFF  L
  F     L
  F     LLLLL

  Author:                Georg Heinze
  Version:               2000.03

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

  Data set:              BSP.ENDO
  Dependent variable:    HG
  Independent variables: NV PI EH

  Table with parameter estimates saved as _OUTTAB.
  Estimates and covariance matrix saved as _OUTEST.
```

Page 2:

```
              Analysis of endometrian cancer study              2

                   Model fitting information

                    Penalized
                       log       Number of      Number of
    Iterations     likelihood    responses    nonresponses

         8          -24.0373         30             49
```

Page 3:

```
              Analysis of endometrian cancer study              3

  FL estimates, profile penalized likelihood confidence limits
              and penalized likelihood ratio tests

           Parameter  Standard    Lower     Upper      Pr >
 Variable   estimate    Error   95% c.l.  95% c.l.  Chi-Square


 INTERCEP    3.77456   1.48869   1.08254   7.20925    0.00419
 NV          2.92927   1.55076   0.60977   7.85456    0.00912
 PI         -0.03475   0.03958  -0.12446   0.04046    0.38748
 EH         -2.60416   0.77602  -4.36518  -1.23272    0.00003
```

4.1.2 Syntax

The following options are available in %FL:

%fl(<data=*SAS data set*,>
    <y=*variable*,>
    varlist=*variables*,
    <noint=*value*,>
    <odds=*value*,>
    <test=*variables*,>
    <profile=*variables*,>

```
<profsel=value,>
<profser=value,>
<outprof=SAS data set,>
<pl=value,>
<print=value,>
<outest=SAS data set,>
<outtab=SAS data set,>
<out=SAS data set,>
<pred=variable,>
<lower=variable,>
<upper=variable,>
<h=variable,>
<maxit=value,>
<maxhs=value,>
<epsilon=value,>
<maxstep=value>
<by=variables,>
<offset=SAS data set>);
```

These options can be categorized into basic options, output options, model fitting options, and options useful for simulation.

4.1.3 Basic options

- `data=`*SAS data set* names the input SAS data set. The default value is `_LAST_`.

- `y=`*variable* names the dependent variable containing values 1 and 0 only. The default value is Y.

- `varlist=`*variables* names a list of independent variables, separated by blanks. There is no default value. This option is required.

- `noint=`*value* suppresses estimation of an intercept if set to 1. The default value is 0.

- `odds=`*value* requests estimates and confidence limits for odds ratios if set to 1. The default value is 0.

- `test=`*variables* requests a likelihood ratio test for the null hypothesis that all parameters corresponding to variables specified in this option are zero.

- `pl=`*value* specifies the method of computing confidence limits and tests for parameters. If `pl`=1, then profile penalized likelihood confidence limits for parameters and penalized likelihood ratio tests will be computed. Otherwise, confidence limits and tests will be based on estimated variance (Wald method). Default value is 1.

- `profile=`*variables* requests a plot of the profile penalized log likelihood function for all variables specified in this options. Of course, these variables have to appear also in the `varlist` option. The x-axis ranges for this plot are automatically chosen by the macro, but can also be specified by the user in terms of standard errors to the left and right from the point estimate (options `profsel` and `profser`, respectively). Also the number of profile likelihood evaluations (`profn`, default value=100) can be chosen. If the `profile` options is used, then the data set specified in option `outprof` will consist of the variables

  - `_name_` containing the covariate's name
  - `_b_` containing the values on the x-axis (the values for β)
  - `_profli_` containing the values of the profile penalized log likelihood
  - `_normal_` containing the values $\ell_{\max} - 0.5(\beta - \hat{\beta})^2/\hat{\sigma}^2$ (where $\ell_{\max}$ is the maximized penalized log likelihood) which represent the Wald (normal) approximation to the profile penalized log likelihood
  - `_refer_` containing the reference line (the values of β where the profile penalized log likelihood function and normal approximation crosses the reference line are the profile penalized likelihood and Wald confidence limits, respectively).
  - any BY-variables specified

4.1.4 Model fitting options

- `maxit=`*value* specifies the maximum number of iterations. Default value is 50.

- `maxhs=`*value* specifies the maximum number of step–halvings allowed in one iteration. Default value is 5.

- `epsilon=`*value* specifies the maximum allowed sum of absolute changes in parameter values to declare convergence. Default value is 0.0001.

- `maxstep=`*value* specifies the maximum change of parameter values allowed in one iteration. Default value is 100.

4.1.5 Output options

- `print=`*value* suppresses printed output if set to 0. Default value is 1.

- `outest=`*SAS data set* names a SAS data set containing parameter estimates, penalized log likelihood and covariance matrix. There is no default value. The data set contains a variable for the intercept parameter and one variable for each explanatory variable in the `varlist` option. The `outest` data set contains one observation for each BY–group containing the FL–type estimates of the regression coefficients. Additionally, there are observations containing the rows of the estimated covariance matrix of the parameter estimators for each `by` group. The `outest` data set contains the following variables:

  – any BY variables specified

  – `_CODE_`, a variable containing the value -1 indicating a line with parameter estimates or the subsequent numbers of the covariates indicating lines containing corresponding rows of the estimated covariance matrix.

  – if `noint`=0, the variable `INTERCPT`

  – one variable for each explanatory variable in the `varlist` statement.

  – `_LINK_`, a character variable of length 8 with the value LOGIT

  – `_PENLIK_`, the maximized penalized log likelihood at the FL estimate (where `_CODE_`=-1) or, if `pl`=1, the maximized penalized log likelihood with the restriction that the corresponding parameter is 0 (where `_CODE_`> 0)

  – `_LNLIKE_`, the log likelihood at the FL estimate (where `_CODE_`=-1) or, if `pl`=1, the log likelihood at the FL estimate maximizing the penalized likelihood with the restriction that the corresponding parameter is 0 (where `_CODE_`> 0)

  – `_IT_`, the number of iterations needed to arrive at the maximum of the penalized likelihood

  – `_RESP_`, the number of responses

  – `_NORESP_`, the number of nonresponses

  – `_TYPE_`, a character variable of length 8 with two possible values: PARMS for parameter estimates or COV for covariance estimates

  – `_NAME_`, a character variable of length 8 containing the name ESTIMATE for parameter estimates or the name of each explanatory variable or INTERCPT for the covariance estimates

- outtab=*SAS data set* names a SAS data set containing parameter estimates, standard errors, Wald confidence limits and p–values. The default value is _OUTTAB. The data set contains one observation per explanatory variable or intercept parameter and BY–group. It contains the following variables:

  - any BY variables specified
  - _RBY_, the ascending rank of the corresponding BY-group
  - _VAR_, the subsequent number for each explanatory variable in the varlist option or 0 for the intercept parameter
  - _NAME_, a character variable of length 8 containing the name of each explanatory variable or INTERCPT
  - BETA, the FL parameter estimates
  - STDERR, the estimated standard error of the corresponding parameter estimate
  - CI_LO, the lower confidence limit for the parameter estimate
  - CI_UP, the upper confidence limit for the parameter estimate
  - P_VALUE, the p–value for $H_0 : \beta_r = 0$.
  - _ITER_, the number of iterations that the model fitting algorithm needed to arrive at the maximum of the penalized log likelihood.

  If the options odds is set to 1, then this data set will also contain the following variables:

  - ODDS, the estimated odds ratio
  - OR_LO, the lower confidence limit for the odds ratio
  - OR_UP, the upper confidence limit for the odds ratio

- out=*SAS data set* creates a new SAS data set that contains all the variables in the input data set and the predicted probability of an event reponse, the lower and upper confidence intervals for the predicted probability, and the diagonal element of the hat matrix.

  - pred=*variable* assigns a name for the variable in the out data set containing the predicted probabilities of an event response. Default value is _PRED_.
  - lower=*variable* assigns a name for the variable in the out data set containing the lower confidence limit of the predicted probability. Default value is _LOWER_.

– upper=*variable* assigns a name for the variable in the out data set containing the upper confidence limit of the predicted probability. Default value is _UPPER_.

– h=*variable* assigns a name for the variable in the out data set containing the diagonal element of the hat matrix. Default value is _H_.

4.1.6 Options useful for simulation

- by=*variables* requests separate analyses on observations in groups defined by the BY variable(s).

- offset=*SAS data set* names an input data set containing offset values of parameter estimates. This data set should contain the same variables as are specified in varlist, plus a variable named INTERCEP (unless noint=1) and, if the by–option is used, the variable(s) specified in this option. Therefore the offset data set should have as many observations as there are BY–groups in the input data set. If a particular parameter in a particular BY–group should be estimated, then its value should be missing in the offset data set, otherwise the parameter will be treated as fixed at the value found in the offset data set. If a variable contained in varlist ist not defined in the offset data set, its parameter value will be estimated in any BY–group.

4.1.7 Titles

Titles 1–3 are not used by the macro. These titles can be set by the user in a statement before the macro call. Titles 4 and 5 are used by the macro. These titles are deleted on exit.

4.1.8 Printed output

If print=0, %FL does not produce any printed output. Otherwise, printed output includes the following:

- if the by option is not used, the number of iterations needed to arrive at the maximum of the penalized log likelihood

- for each BY–group a table containing, for each parameter, the FL estimate, its estimated standard error, Wald lower and upper confidence limits, and Wald *p*–value

- if the `test` option is used, an additional page containing the penalized log likelihood χ^2, associated degrees of freedom and p-value for the test that all parameters corresponding to variables in the `test` option are zero

4.1.9 Computational details

Parameter values are estimated using a Newton–Rhaphson algorithm. With parameter vector $\beta = (\beta_1, \ldots, \beta_k)'$, the gradient vector g_β and the information matrix I_β are given, respectively, by (see also § 2)

$$\begin{aligned} g_\beta &= \sum_{i=1}^{n} \{y_i - \pi_i + h_i\,(1/2 - \pi_i)\}\, x_{ir} \\ I_\beta &= \sum_{i=1}^{n} -\frac{\partial^2 \log L_i}{\partial \beta^2} = \sum_{i=1}^{n} -\frac{\partial^2 \{y_i \pi_i + (1-y_i)(1-\pi_i)\}}{\partial \beta^2} \end{aligned} \tag{4·3}$$

With a starting value of $\beta^{(0)}$, the FL estimate $\hat\beta$ is obtained iteratively until convergence is obained:

$$\beta^{(s+1)} = \beta^{(s)} + I^{-1}_{\beta^{(s)}} g_{\beta^{(s)}} \tag{4·4}$$

If the penalized log likelihood evaluated at $\beta^{(s+1)}$ is less than that evaluated at $\beta^{(s)}$, then $\beta^{(s+1)}$ is recomputed by step–halving. The maximum number of step–halvings within one iteration can be specified using the `maxhs` option. If the step length for any parameter estimate in one iteration is longer than the value specified in the `maxstep` option, then its absolute value is truncated to this value.

From some experience with real and simulated data sets we learned that the choice of starting values is crucial for convergence of the `%FL` program if offset values are specified. Therefore, starting values can not be user–supplied. At beginning, all parameter values except the intercept and values specified in the `offset` data set are set to zero. The starting value of the intercept $\beta_1^{(0)}$ is calculated such that the average value of the linear predictor η equals the logit of the average outcome:

$$\beta_1^{(0)} = \log\left(\frac{\bar p}{1-\bar p}\right) - \bar\eta \tag{4·5}$$

where $\bar p = n^{-1}\sum_{i=1}^{n} y_i$, $\bar\eta = n^{-1}\sum_{i=1}^{n}\eta_i$, and $\eta_i = \sum_{r=1}^{k} x_{ir}\beta_r$.

Computation of profile penalized likelihood confidence intervals for parameter follows the algorithm of Venzon & Moolgavkar (1988), implemented also in PROC LOGISTIC of SAS Institute (1999b), p. 416f. The limits of the continuous set of values for which LR does not exceed the $(1-\alpha)$100th percentile of the χ^2_1-distribution (see also § 2.4), i. e. the profile penalized likelihood confidence limits, are found iteratively by approximating the penalized log likelihood function in a neighborhood of β by the quadratic function

$$\tilde{\ell}(\beta+\delta) = \ell(\beta) + \delta' g + \frac{1}{2}\delta' V \delta$$

where $g = g(\beta)$ is the gradient vector and $V = V(\beta)$ is the Hessian matrix. Suppose the confidence limits for parameter β_r are to be computed. The increment vector δ for the next iteration is obtained by solving the likelihood equations

$$\frac{\partial}{\partial\delta}\{\tilde{\ell}(\beta+\delta) + \lambda(e_r'\delta - \theta)\} = 0$$

where λ is a Lagrange multiplier, e_r is the rth unit vector, and θ is an unknown constant. The solution is

$$\delta = -V^{-1}(g + \lambda e_r) \qquad (4{\cdot}6)$$

Let $\ell_0 = \ell_{\max} - \chi^2_{1,1-\alpha}/2$, where $\ell_{\max}$ is the maximized value of the penalized log likelihood at the FL estimate. By substituting (4·6) into the equation $\tilde{\ell}(\beta+\delta) = \ell_0$, λ can be estimated as

$$\lambda = \pm\left(\frac{2\left(\ell_0 - \ell(\beta) + \frac{1}{2}g'V^{-1}g\right)}{e_r'V^{-1}e_r}\right)^{\frac{1}{2}}$$

The upper confidence limit for β_r is computed by starting at the FL estimate of β and iterating with positive values of λ until convergence is attained. The process is repeated for the lower confidence limit using negative values of λ.

Convergence is controlled by value ϵ specified with the `epsilon` option. Convergence is declared on the current iteration if the following two conditions are satisfied:

$$|\ell(\beta) - \ell_0| \le \epsilon$$

and

$$(g + \lambda e_r)' V^{-1}(g + \lambda e_r) \le \epsilon$$

4.1.10 Examples

Consider the data set introduced in § 4.1.1. To obtain estimated odds ratios and corresponding confidence intervals, the macro is called using the `odds` option

```
%fl(data=bsp.endo, y=y, varlist=nv eh pi, odds=1);
```

leading to the following output:

Page 1 omitted

```
          FL estimates, profile penalized likelihood confidence limits
                      and penalized likelihood ratio tests

                    Parameter    Standard      Lower        Upper
    Variable         estimate       Error    95% c.l.     95% c.l.

    INTERCEP          3.77456     1.48869     1.08254      7.20925
    NV                2.92927     1.55076     0.60977      7.85456
    EH               -2.60416     0.77602    -4.36518     -1.23272
    PI               -0.03475     0.03958    -0.12446      0.04046

       Pr >           Odds       Lower        Upper
    Chi-Square        ratio     95% c.l.     95% c.l.

      0.00101       43.5782     2.95216      1351.88
      0.00912       18.7140     1.84000      2577.46
      0.00003        0.0740     0.01271         0.29
      0.38748        0.9658     0.88298         1.04
```

The estimated probabilities of a high histological grading (III-IV) can be obtained by using the `out` option:

```
%fl(data=bsp.endo, y=hg, varlist=nv eh pi, out=prob, pred=prob,
    lower=p_lo, upper=p_up);
proc print data=prob;
var nv eh pi hg prob p_lo p_up;
run;
```

The output data set 'PROB' contains predicted probabilities as well as confidence limits:

Output omitted

OBS	NV	EH	PI	HG	PROB	P_LO	P_UP
1	0	1.64	13	0	0.27928	0.15998	0.44085
2	0	1.50	28	0	0.24885	0.10130	0.49335
3	0	2.02	29	0	0.07630	0.01839	0.26702
4	0	2.26	16	0	0.06496	0.01860	0.20297
5	0	1.33	11	0	0.48220	0.28478	0.68534
6	0	2.29	15	0	0.06237	0.01727	0.20115
7	0	3.14	8	0	0.00919	0.00074	0.10362
8	0	2.37	19	0	0.04489	0.01041	0.17354
9	0	2.33	12	0	0.06238	0.01637	0.21010
10	0	2.68	34	0	0.01230	0.00101	0.13286

Output omitted

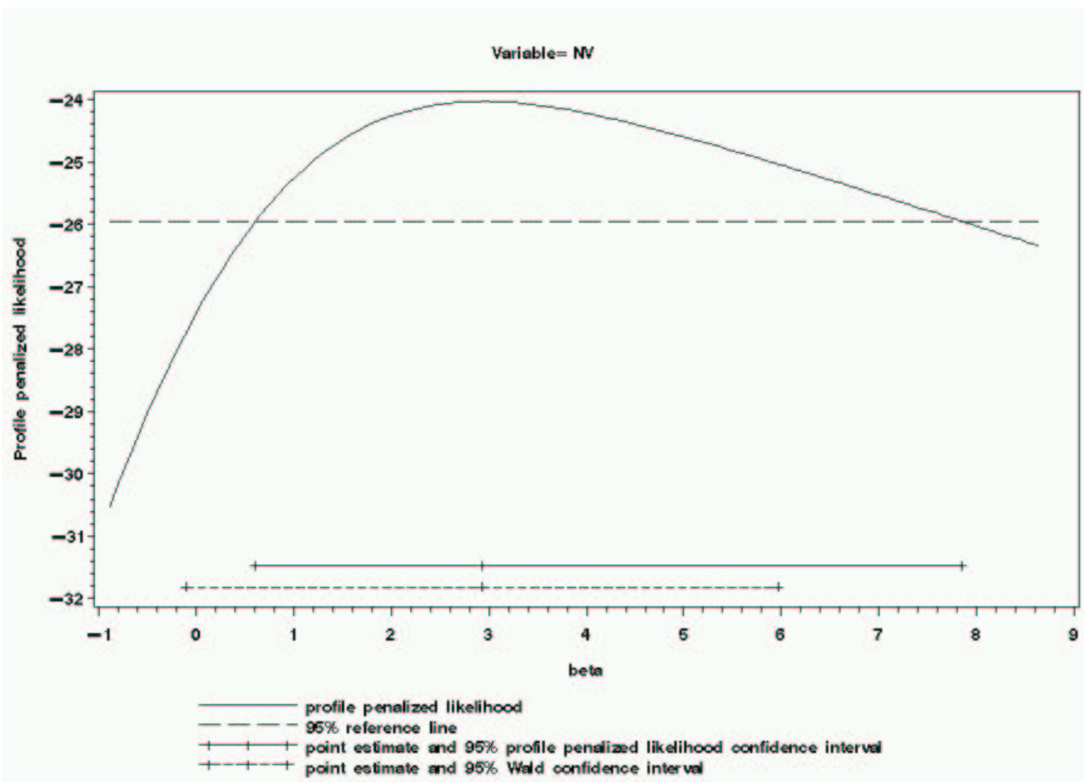
The `profile` option can be used to obtain a plot of the profile of the penalized log likelihood for a parmater.

```
%fl(data=endo2, y=hg, varlist=nv eh pi, profile=nv);
```

The values of the penalized log likelihood are plotted against the parameter values of *NV* by the macro (see Fig. 1, only available in the PS version of this document).

Fig. 1 shows the resulting graph.

Figure 1: Profile penalized likelihood function for parameter β_{NV} of the endometrian cancer study.



The `test` option can be used to test the simultanous effect of more than one independent variables on the outcome variable. In our example, to test the hypothesis $H_0 : \beta_{EH} = \beta_{PI} = 0$ the macro call

```
%fl(data=bsp.endo, y=hg, varlist=nv pi eh,
 test=eh pi);
```

is submitted. On the fourth page of output, the following table is displayed:

Tested variable (s)	Penalized log likelihood Chi-square	Degrees of freedom	Pr > Chi-square
EH PI	17.8667	2	.000131916

4.2 FC: Cox regression

4.2.1 Overview

The SAS–macro `%FC` was written to facilitate application of Firth's modified score procedure in Cox regression analysis. It is available at the WWW site `http://www.akh-wien.ac.at/imc/biometrie/fc`. Supplied with a SAS data set as input, its output contains FC-type Cox regression parameters, standard errors, confidence limits, p-values, the value of the maximized penalized log likelihood, and the number of iterations needed to arrive at the maximum. The parameter estimates, the maximized penalized log likelihood and the estimated covariance matrix are stored in a SAS data set that has the same structure as the output data set that can be obtained by SAS/PROC PHREG (SAS Institute, 1999b) using the OUTEST option in the PROC PHREG statement. Furthermore, parameter estimates, confidence limits and p-values are stored in a second data set. The structure of this data set resembles the output table written to the output window. The SAS–macro accesses the dynamic link library `fcdll.dll` which is written in FORTRAN and compiled for Windows-32 compatible PC. The functions of this library can also be accessed from programs other than SAS, provided the input data follow the structure outlined in § 4.2.10.

Multiple input data sets can be efficiently processed using grouping variables similarly to PROC PHREG. Finally, offset values can be used to compute profiles of penalized likelihood function or penalized likelihood ratio tests.

Use of `%FC` is exemplified using a breast cancer data set (Tables 16 and 17). Survival times of 100 patients were recorded (74 of them censored) and also values of four potential risk factors: tumour stage (T), nodal status (N), histological grading (G) and Cathepsin D immunoreactivity (CD). For analysis these factors were dichotomised to levels of 0 and 1 (unfavourable). The survival and censoring times are stored in variable *MONTHS*, with the censoring indicator *CENS*, 0 indicating a censored survival time.

Table 16: Breast cancer data set.

T	N	G	CD	MONTHS	CENS	T	N	G	CD	MONTHS	CENS
1	0	1	0	1.5	1	0	0	1	0	72.0	0
0	0	1	1	3.5	1	0	1	0	0	72.0	0
1	1	1	1	8.1	1	1	1	1	0	72.0	0
1	0	1	1	8.7	1	0	0	1	1	72.0	0
1	1	1	0	12.1	1	0	0	1	0	72.0	0
1	1	1	1	12.4	1	1	1	1	1	72.0	0
1	0	1	0	13.1	1	0	0	1	1	72.0	0
1	1	1	1	14.8	1	0	0	0	0	72.0	0
1	0	1	0	18.8	1	0	0	0	0	72.0	0
1	1	1	1	20.0	1	0	0	1	1	72.0	0
1	1	1	1	20.7	1	0	0	1	0	72.0	0
0	1	1	0	21.6	1	0	0	0	0	72.0	0
0	1	1	0	23.4	0	0	0	1	0	72.0	0
1	0	1	1	23.7	1	0	0	1	0	72.0	0
1	1	1	1	26.6	1	0	0	0	0	72.0	0
0	0	0	0	28.0	0	0	0	1	0	72.0	0
0	1	1	0	28.4	0	0	0	1	0	72.0	0
1	0	1	0	30.4	0	0	1	0	0	72.0	0
1	0	1	1	30.5	0	0	1	1	0	72.0	0
0	0	0	0	30.9	0	0	1	1	0	72.0	0
1	0	1	0	33.1	0	0	1	1	0	72.0	0
1	0	1	0	36.9	0	0	0	0	1	72.0	0
1	0	1	0	37.3	0	0	0	1	0	72.0	0
0	0	1	1	39.9	0	1	0	0	0	72.0	0

Table 17: Breast cancer data set (*cont*).

T	N	G	CD	MONTHS	CENS	T	N	G	CD	MONTHS	CENS
1	0	1	0	40.2	1	1	0	1	0	72.0	0
1	0	1	1	41.3	1	0	0	0	1	72.0	0
0	0	0	0	41.6	0	0	0	0	0	72.0	0
0	0	1	0	44.9	1	1	0	1	0	72.0	0
0	1	1	0	47.5	0	1	0	1	0	72.0	0
0	1	1	0	47.5	1	0	1	1	0	72.0	0
0	1	1	0	48.0	0	1	0	1	1	72.0	0
1	1	1	1	49.0	1	0	0	0	0	72.0	0
1	0	1	0	50.0	0	0	0	0	0	72.0	0
1	1	1	1	55.3	1	1	0	1	0	72.0	0
1	0	0	0	56.8	0	0	0	1	1	72.0	0
1	0	0	0	58.0	0	0	0	1	0	72.0	0
1	1	1	1	58.4	1	0	0	0	0	72.0	0
0	0	1	0	58.7	0	0	1	1	0	72.0	0
0	1	1	0	59.7	1	0	0	1	1	72.0	0
1	0	1	0	60.2	1	1	0	1	0	72.0	0
1	1	1	1	60.6	1	1	0	1	1	72.0	0
0	0	1	0	62.4	0	1	1	1	1	72.0	0
1	1	1	0	62.6	1	1	0	0	0	72.0	0
0	0	1	0	63.5	0	0	0	0	0	72.0	0
1	0	1	0	64.2	0	0	1	0	0	72.0	0
0	0	0	0	64.9	0	1	0	1	0	72.0	0
0	0	0	1	67.3	0	0	0	1	1	72.0	0
0	0	1	0	70.1	0	0	0	0	1	72.0	0
0	1	1	0	70.3	1	0	0	1	0	72.0	0
1	0	0	0	72.0	0	0	1	0	0	72.0	0

Suppose the data has been stored in SAS data set `bsp.breast`. To obtain an FC analysis with a table containing variable names, parameter estimates, standard errors, confidence limits and p–values, one submits the macro call

```
%fc(data=bsp.breast, varlist=t n g cd, time=months, cens=cens);
```

The following output is produced (*Page 1*):

```
                Analysis of breast cancer study                1


  FFFFF  CCC             Cox proportional hazards regression
  F     C                using Firth's modified score function
  FFFF  C
  F     C
  F      CCC


  Author:                Georg Heinze
  Version:               2000.12

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

  Data set:              BSP.BREAST
  Dependent variable:    TIME
  Censoring indicator:   CENS
  Censoring value:       0
  Independent variables: T N G CD

  Table with parameter estimates saved in _TAB.
  Estimates and covariance matrix saved in _EST.

  Covariance matrix is based on inverse Fisher information.
```

Page 2:

Model fitting information

Iterations	Penalized log likelihood	Null penalized log likelihood	Likelihood ratio Chi-square	Prob>Chi
6	-92.3524	-110.328	35.9514	<.0001

Number of events	Censored	Number of observations	% events	% censored
26	74	100	26	74

Page 3:

Analysis of breast cancer study 3

FC estimates, profile penalized likelihood confidence limits
and penalized likelihood ratio tests

Variable	Parameter estimate	Standard error	Standardized estimate	Lower 95% c.l.	Upper 95% c.l.	Pr > Chi-Square
T	1.22444	0.49160	0.60619	0.30950	2.24836	0.0082
N	0.91888	0.42257	0.42864	0.11374	1.76351	0.0253
G	2.42442	1.47354	1.06343	0.38231	7.28066	0.0136
CD	0.39711	0.44186	0.18198	-0.46702	1.25612	0.3645

4.2.2 Installation

Since the core routines reside in a dynamic link library that has to be accessed by SAS, some installation steps are necessary.

1. Copy the file ‘dforrt.dll’ into your `windows\system32` folder (e. g. `C:\Winnt\system32\`) if it’s not already there.

2. (a) Create a subfolder in your SAS working directory for DLLs (e. g. `s:\sasdll`). Copy the file ’fcdll.def’ into that subfolder. If you call the subfolder `s:\sasdll`, go to step 3.

   (b) If you call the subfolder `MYFOLDER\DLLs`, then in file ‘fc.sas’ change the first line which is now

   ```
   filename SASCBTBL 's:\sasdll\fcdll.def';
   ```

   to

   ```
   filename SASCBTBL 'MYFOLDER\DLLs\fcdll.def';
   ```

   Save ’fc.sas’ to the folder where you store SAS macros.

3. Copy the file ‘fcdll.dll’ into the folder you created at step 2.a.

4. In file SASV8.CFG, located in your sasroot folder (usually called something like `C:\Program Files\SAS Institute\SAS\V8\`) insert the line

   ```
   "s:\sasdll"
   ```

   or

   ```
   "MYFOLDER\DLLs"
   ```

   , respectively (see Step 2.a), after the two lines

   ```
   /* Setup the SAS System load image search paths definition */
   -PATH          (
   ```

   and save SASV8.CFG.

5. Restart SAS.

4.2.3 Syntax

The following options are available in %FC:

```
%fc(<data=SAS data set,>
<time=variable,>
<cens=variable,>
<censval=value,>
varlist=variables,
<pl=value,>
<risk=value,>
<profile=variables,
<profsel=value,>
<profser=value,>
<profn=value,>
<outprof=SAS data set,>
<outmod=SAS data set,>
<outest=SAS data set,>
<outtab=SAS data set,>
<print=value,>
<test=variables,>
<maxit=value,>
<maxhs=value,>
<epsilon=value,>
<maxstep=value>
<by=variables,>
<offset=SAS data set>);
```

These options can be categorized into basic options, output options, model fitting options, and options useful for simulation.

4.2.4 Basic options

- data=*SAS data set* names the input SAS data set. The default value is _LAST_.
- time=*variable* names a variable containing survival times. The default value is time.
- cens=*variable* names a variable containing the censoring indicator for each survival time. Default value is cens.

- `censval=`*value* names the censoring value. The default value is 0, meaning that if `censval`=0, then the corresponding survival time is treated as censored.

- `varlist=`*variables* names a list of independent variables, separated by blanks. There is no default value. This option is required.

- `pl=`*value* requests profile penalized likelihood confidence intervals and penalized likelihood ratio tests for parameters if set to 1 (fast algorithm of Venzon & Moolgavkar, 1988) or 2 (slow but robust simple binary search). With some data sets we noticed numerical problems when `pl` is set to 1. These problems usually disappear if you use `pl=2`. If `pl`=0, Wald confidence intervals and tests will be computed. Default value is 1.

- `profile=`*variables* requests a plot of the profile penalized log likelihood function for all variables specified in this options. Of course, these variables have to appear also in the `varlist` option. The x-axis ranges for this plot are automatically chosen by the macro, but can also be specified by the user in terms of standard errors to the left and right from the point estimate (options `profsel` and `profser`, respectively). Also the number of profile likelihood evaluations (`profn`, default value=100) can be chosen. If the `profile` options is used, then the data set specified in option `outprof` will consist of the variables

  - `_name_` containing the covariate's name
  - `_b_` containing the values on the x-axis (the values for β)
  - `_profli_` containing the values of the profile penalized log likelihood
  - `_normal_` containing the values $\ell_{\max} - 0.5(\beta - \hat{\beta})^2/\hat{\sigma}^2$ (where $\ell_{\max}$ is the maximized penalized log likelihood) which represent the Wald (normal) approximation to the profile penalized log likelihood
  - `_refer_` containing the reference line (the values of β where the profile penalized log likelihood function and normal approximation crosses the reference line are the profile penalized likelihood and Wald confidence limits, respectively).
  - any BY-variables specified

4.2.5 Output options

- `risk=`*value* requests estimated relative risks and confidence intervals to be included in the output table if set to 1. Default value is 0.

- `print=`*value* suppresses printed output if set to 0. Default value is 1.

- `outmod=`*SAS data set* names a SAS data set containing the number of iterations (`_it_`), the null penalized log likelihood (`_penli0_`), the maximized penalized log likelihood (`_penlik_`), the model chi^2 test statistic (`_modchi_`), the associated P-value (`_p_`), number and percentage of non-censored and censored observations (`_events`, `_cens_`, `_pev_` and `_pce_`, respectively), the total number of observations (`_nobs_`), and, if used, the variables specified in the `by` option.

- `outest=`*SAS data set* names a SAS data set containing parameter estimates, penalized log likelihood and covariance matrix. There is no default value. The data set contains one variable for each explanatory variable in the `varlist` option. The `outest` data set contains one observation for each `by` group containing the FC-type estimates of the regression coefficients. Additionally, there are observations containing the rows of the estimated covariance matrix of the parameter estimators for each BY–group. The `outest` data set contains the following variables:

  - any BY variables specified
  - one variable for each explanatory variable in the `varlist` option.
  - `_penlik_`, the maximized penalized log likelihood at the FC estimate
  - `_TYPE_`, a character variable of length 8 with two possible values: PARMS for parameter estimates or COV for covariance estimates
  - `_NAME_`, a character variable of length 8 containing the name of the `time` variable for parameter estimates or the name of each explanatory variable for the covariance estimates

- `outtab=`*SAS data set* names a SAS data set containing parameter estimates, standard errors, confidence limits and p–values. The default value is `_OUTTAB`. The data set contains one observation per explanatory variable and BY–group. It contains the following variables:

  - any BY variables specified
  - `_var_`, the subsequent number for each explanatory variable in the `varlist` option
  - `_name_`, the name of each explanatory variable in the model (as specified in the `varlist` option)
  - `_beta_`, the FC parameter estimates
  - `_stderr_`, the estimated standard error of the corresponding parameter estimate

– `_bstd_`, the standardized parameter estimate

– `_lo_`, the lower confidence limit for the parameter estimate

– `_up_`, the upper confidence limit for the parameter estimate

– `_p_`, the p–value for $H_0 : \beta_r = 0$.

The method of computation of confidence intervals and p-values (Wald or profile penalized likelihood) can be controlled using the `pl` option. The significance level of the intervals can be set by the `alpha` option.

4.2.6 Model fitting options

- `test=`*variables* requests a penalized likelihood ratio test of the null hypothesis that all parameters listed in the `test` option are zero.
- `maxit=`*value* specifies the maximum number of iterations. Default value is 25.
- `maxhs=`*value* specifies the maximum number of step-halvings allowed in one iteration. Default value is 2.
- `epsilon=`*value* specifies the maximum allowed change in penalized log likelihood to declare convergence. Default value is 10^{-6}.
- `maxstep=`*value* specifies the maximum change of (standardized) parameter values allowed in one iteration. Default value is 1.

4.2.7 Options useful for simulation

- `by=`*variables* requests separate analyses on observations in groups defined by the BY variable(s).
- `offset=`*SAS data set* names an input data set containing offset values of parameter estimates. This data set should contain the same variables as are specified in `varlist`, plus, if the `by`–option is used, the variable(s) specified in this option. Therefore the `offset` data set should have as many observations as there are BY–groups in the input data set. If a particular parameter in a particular BY–group should be estimated, then its value should be missing in the `offset` data set, otherwise the parameter will be treated as fixed at the value found in the `offset` data set. If a variable contained in `varlist` ist not defined in the `offset` data set, its parameter value will be estimated in any BY–group.

4.2.8 Titles

Titles 1–3 are not used by the macro. These titles can be set by the user in a statement before the macro call. Titles 4 and 5 are used by the macro. These titles are deleted on exit.

4.2.9 Printed output

Unless `print=0`, printed output usually consists of three pages. The first page includes

- the name of the input data set
- the name of the variable containing survival times
- the name of the variable containing the censoring indicator values
- the censoring value
- a message on where estimates, confidence limits and covariance matrix have been stored to

The second page includes

- the number of iterations needed to arrive at the maximum of the penalized log likelihood
- the value of the maximized penalized log likelihood
- the value of the null penalized log likelihood
- the χ^2 statistic for testing the hypothesis that all parameters in the model are zero
- the P-value associated with this test statistic
- a summary of the number and percentage of events and censored observations

The third page includes a table containing variable names, FC parameter estimates and associated estimated standard errors, confidence intervals for the parameters and p-values. If the `risk` option was set to 1, then this table also includes the estimated risk ratios and associated confidence intervals.

If a special penalized likelihood ratio test for testing more than one parameter at a time was requested by using the `test` option, an additional page gives information on the χ^2-statistic for testing the hypothesis that all parameters listed in the `test` option are 0, and the associated degrees and freedom and p-value.

Pages 2, 3 and 4 are repeated for all BY-groups if the `by`-option was used.

4.2.10 Use of 'fcdll.dll' with applications other than SAS

In a PROC IML, the macro `%FC` calls the dynamic link library `fcdll.dll`. This library supplies two subroutines: FIRTHCOX (estimation of parameters) and PLCOMP (estimation of profile penalized likelihood confidence intervals).

Subroutine FIRTHCOX

This subroutine takes as arguments the arrays CARDS (as*input*), PARMS (*updated* by the routine) and IOARRAY (*updated* by the routine). Let n and k denote sample size and number of covariates. Then the parameter arrays are of following dimension:

CARDS	$n \times (k+2)$
PARMS	11×1
IOARRAY	$(k+3) \times k$

Parameter CARDS is an array of REAL*8 and defined as follows:

$$\text{CARDS} = [\text{X}][\text{T}][\text{C}]$$

where X, T and C denote the $n \times k$ matrix of covariates, the $n \times 1$ vector of survival times and the $n \times 1$ vector of censoring indicators. Note that the data have to be sorted by ascending survival time. PARMS has the following entries:

Row	Description (recommended value)
1	Input: n
2	Input: k
3	Input: IFIRTH: 1 for penalized ml estimation, 0 for ml estimation (1)
4	Input: MAXIT: maximum number of iterations (25)
5	Input: MAXHS: maximum number of step-halvings per iteration (3)
6	Input: STEP: maximum step size per iteration (1.0)
7	Input: CRILI: convergence criterion for parameter estimates (0.000001)
8	Output: JCODE: returncode; 0=ok
9	Output: ISEP: monotone likelihood; 0=ok
10	Output: ITER: number of iterations carried out
11	Output: XL: (penalized) log likelihood
12	Output: XL0: null (penalized) log likelihood

IOARRAY has as many columns as parameters to be estimated. The lines are as follows (recommended values in brackets):

Line	description
1	Input: IFLAG: 1 for 'parameter value is to be estimated', 0 for 'parameter value is to be fixed' (1)
2	Input: OFFSET: offset-values for parameters (if IFLAG=0)
3	Output: B: parameter estimates
4 to 3+k	Output: covariance matrix of estimates

Subroutine PLCOMP

Subroutine PLCOMP again has three arguments: CARDS (*input*), PARMS (*update*) and IOARRAY (*update*). These arrays are of the following dimensions:

CARDS	$n \times (k+2)$
PARMS	9×1
IOARRAY	$6 \times k$

CARDS is defined the same way as in FIRTHCOX. PARMS takes the following arguments:

Row	Description (recommended value)
1	Input: n
2	Input: k
3	Input: IFIRTH: 1 for penalized ml estimation, 0 for ml estimation (1)
4	Input: MAXIT: maximum number of iterations (25)
5	Input: MAXHS: maximum number of step-halvings per iteration (3)
6	Input: STEP: maximum step size per iteration (1.0)
7	Input: CRILI: convergence criterion for parameter estimates (0.000001)
8	Input: ALPHA: significance level (the confidence level is therefore 1-ALPHA) (0.05)
9	Output: returncode; 0=ok

IOARRAY has as many columns as there are parameters to be estimated. The rows are defined as follows (recommended values in brackets):

Row	Description
1	Input: IFLAG: 0 for 'parameter is to be fixed' 1 for 'parameter is to be estimated', 2 for 'parameter is to be estimated, and OFFSET value is to be used as starting value' (1)
2	Input: OFFSET: offset values for parameters (if IFLAG=0)
3	Input: B: parameter estimates
4	Output: lower confidence limits
5	Output: upper confidence limits
6	Output: penalized likelihood ratio p-values for parameters
7	Output: number of iterations needed for lower confidence limit
8	Output: number of iterations needed for upper confidence limit

4.2.11 Computational details

Before calling the DLL, each covariate is standardized. Parameter estimates, covariance matrix and confidence limits are retransformed to the original scale after computations. Estimation is based on a Newton–Rhaphson algorithm. Starting values are taken as 0. The commonly used Breslow–approximation is used for computation of the likelihood and its derivatives with tied survival times. With parameter vector $\beta = (\beta_1, \ldots, \beta_k)'$, the gradient vector $g(\beta)$ and the information matrix $I(\beta)$ are given, respectively, by (see also § 2)

$$
\begin{aligned}
g_r(\beta) &= \sum_{j=1}^{m} \left\{ s_{jr} - \frac{d_j \sum_{h \in R_j} x_{hr} \exp(x_h \beta)}{\sum_{h \in R_j} \exp(x_h \beta)} \right\} + 0.5 \mathrm{trace} \left\{ I(\beta)^{-1} \frac{\partial I(\beta)}{\partial \beta_r} \right\} \\
I_{r,s}(\beta) &= -\frac{\partial^2 \log L(\beta)}{\partial \beta_r \partial \beta_s} \\
&= \sum_{j=1}^{m} d_j \left[\frac{\sum_{h \in R_j} x_{hr} x_{hs} \exp(x_h \beta)}{\sum_{h \in R_j} \exp(x_h \beta)} - \right. \\
&\quad \left. - \frac{\left\{ \sum_{h \in R_j} x_{hr} \exp(x_h \beta) \right\} \left\{ \sum_{h \in R_j} x_{hs} \exp(x_h \beta) \right\}}{\left\{ \sum_{h \in R_j} \exp(x_h \beta) \right\}^2} \right]
\end{aligned}
\qquad (4{\cdot}7)
$$

where $\partial I_{r,s}(\beta)/\partial\beta_t = -\partial^3 \log L(\beta)/(\partial\beta_r \partial\beta_s \partial\beta_t)$ (see § 2.2), $r, s, t = 1, \ldots, k$; m denoting the number of distinct survival times $t_{(j]}$ ($j = 1, \ldots, m$) among the n survival times t_i ($i = 1, \ldots, n$); $x_i = (x_{i1}, \ldots, x_{ir}, \ldots, x_{ik})$ the covariate vector related to each individual, d_j the number of deaths at $t_{(j)}$, s_j the vector sum of the covariates of the d_j individuals

(s_{ij} referring to the rth component of s_j), and R_j the set of individuals alive and uncensored prior to $t_{(j)}$. With a starting value of $\beta^{(0)}$, the FC estimate $\hat{\beta}$ is obtained iteratively until convergence is attained:

$$\beta^{(s+1)} = \beta^{(s)} + I(\beta^{(s)})^{-1} g(\beta^{(s)}) \tag{4·8}$$

If the penalized log likelihood evaluated at $\beta^{(s+1)}$ is less than that evaluated at $\beta^{(s)}$, then $\beta^{(s+1)}$ is recomputed by step–halving. The maximum number of step–halvings within one iteration can be specified using the `maxhs` option. If the step length for any parameter estimate in one iteration is longer than the value specified in the `maxstep` option, then its absolute value is truncated to this value. The iterative process is continued until the penalized log likelihood does not increase by more than the value specified in the `epsilon` option.
Computation of profile penalized likelihood confidence intervals follows the algorithm of Venzon & Moolgavkar (1988). This algorithm has already been described in § 4.1.9 on p. 34.

4.2.12 Examples

Consider the breast cancer data set introduced in § 4.2.1. To obtain Wald tests for all covariates based on the inverse fisher information matrix, one submits the call

```
%fc(data=bsp.breast, time=months, cens=cens, censval=0, varlist=t n g cd,
    pl=0);
```

leading to the following output (compare to the output of p. 44):

Output omitted

FC estimates and Wald confidence limits and tests

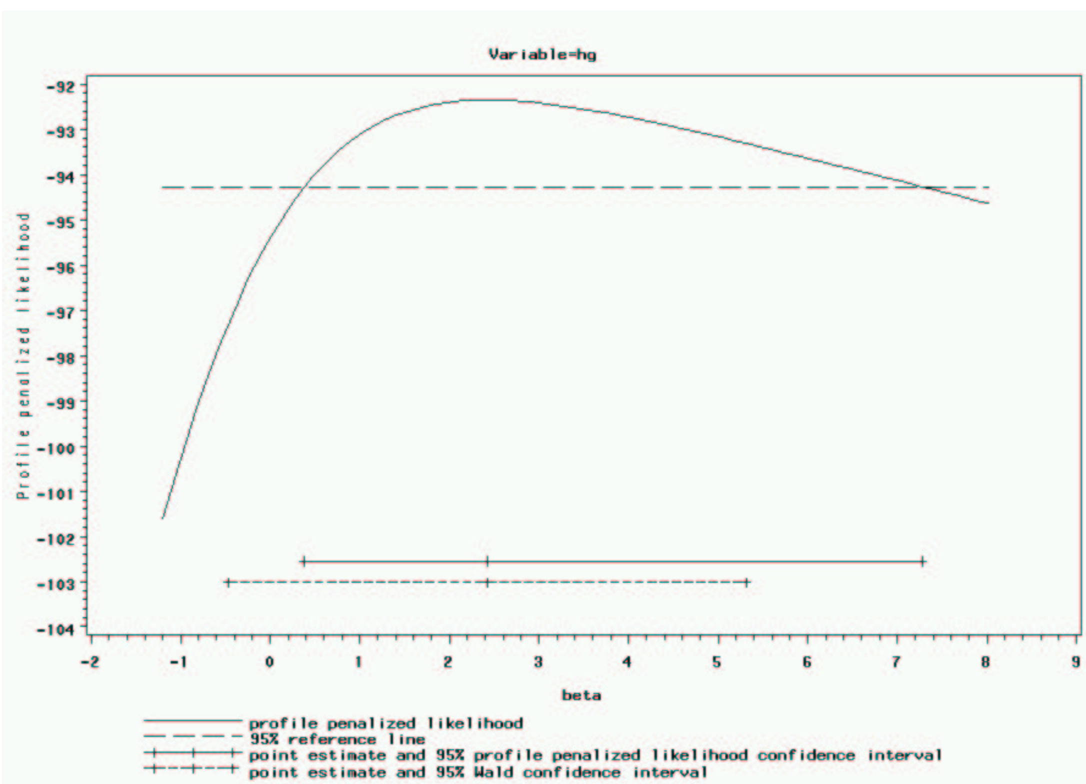
Variable	Parameter estimate	Standard error	Standardized estimate	Lower 95% c.l.	Upper 95% c.l.	Pr > Chi-Square
T	1.22444	0.49160	0.60619	0.26091	2.18796	0.0127
N	0.91888	0.42257	0.42864	0.09066	1.74711	0.0297
G	2.42442	1.47354	1.06343	-0.46368	5.31251	0.0999
CD	0.39711	0.44186	0.18198	-0.46891	1.26313	0.3688

To obtain the profile of the penalized log likelihood as a function of parameter β_G, the `profile` option of the macro can be used. We call the macro by submitting

```
%fc(data=breast2, time=months, cens=cens, varlist=t n g cd,
    profile=g);
```

Fig. 2 shows the graph of the profile penalized likelihood function for parameter β_G.

Figure 2: Profile penalized likelihood function for parameter β_G of breast cancer example.



The `test` option can be used to test the simultanous effect of more than one independent variables on survival. In our example, to test the hypothesis $H_0 : \beta_{CD} = \beta_G = 0$ the macro call

```
%fc(data=bsp.breast, time=months, cens=cens, varlist=t n g cd,
 test=g cd);
```

is submitted. On the fourth page of output, the following table is displayed:

```
            Penalized likelihood ratio tests for parameters


                                          Degrees
              Tested                         of          Pr >
            parameters     Chi-Square     freedom     Chi-Square


               G CD          7.38138          2          0.0250
```

References

ALBERT, A. & ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.

BELSLEY, D. A., KUH, E., & WELSCH, R. W. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

BRYSON, M. C. & JOHNSON, M. E. (1981). The incidence of monotone likelihood in the Cox model. *Technometrics* **23**, 381–383.

CLARKSON, D. B. & JENNRICH, R. I. (1991). Computing extended maximum likelihood estimates for linear parameter models. *Journal of the Royal Statistical Society B* **53**, 417–426.

CLOGG, C. C., RUBIN, D. B., SCHENKER, N., SCHULTZ, B., & WEIDMAN, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* **86**, 68–78.

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

CORDEIRO, G. M. & MCCULLAGH, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B* **53**, 629–643.

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

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

FIRTH, D. (1992b). Generalized linear models and Jeffreys priors: an iterative weighted least-squares approach. In Dodge, Y. & Whittaker, J., editors, *Computational Statistics*, volume 1, pages 553–557, Heidelberg. Physica-Verlag.

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

HEINZE, G. & SCHEMPER, M. (2000). A solution for the problem of separation in logistic regression. *Submitted* .

HEINZE, G. & SCHEMPER, M. (2001). A solution to the problem of monotone likelihood in Cox regression. *Biometrics* **57**.

HIRJI, K. F., MEHTA, C. R., & PATEL, N. R. (1987). Computing distributions for exact logistic regression. *Journal of the American Statistical Association* **82**, 1110–1117.

HIRJI, K. F., TSIATIS, A. A., & MEHTA, C. R. (1989). Median unbiased estimation for binary data. *The American Statistician* **43**, 7–11.

Kolassa, J. E. (1997). Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scandinavian Journal of Statistics* **24**, 523–530.

Lesaffre, E. & Albert, A. (1989). Partial separation in logistic discrimination. *Journal of the Royal Statistical Society B* **51**, 109–116.

NAG (1998). *NAG Fortran Library Manual - Mark 18.* Numerical Algorithms Group Ltd., Oxford.

Santner, T. J. & Duffy, D. E. (1986). A note on A. Albert and J. A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **73**, 755–758.

SAS Institute (1999a). *SAS Language Reference, Version 8.* SAS Institute Inc., Cary, NC.

SAS Institute (1999b). *SAS/STAT User's Guide, Version 8.* SAS Institute Inc., Cary, NC.

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

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics* **9**, 93–108.

Venzon, D. J. & Moolgavkar, S. H. (1988). A method for computing profile-likelihood based confidence intervals. *Applied Statistics* **37**, 87–94.