CROSS-VALIDATION IN SURVIVAL ANALYSIS

PIERRE J. M. VERWEIJ AND HANS C. VAN HOUWELINGEN

Department of Medical Statistics, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands

SUMMARY

predictive value of a statistical model is conceptually different from the explained variation. In this er we construct a measure of the predictive value of the Cox proportional hazards model, computed from leave-one-out regression coefficients. These coefficients can also be used to calculate a shrinkage factor ch can be applied to improve the predictions and that can be used in R^2-type measures of the proportion xplained variation. Our methods are illustrated by a study of chemotherapy for advanced ovarian cer.

1. INTRODUCTION

building a regression model, it is usual to search for the one that best describes the available a. In this procedure, the explained variation, as measured by the log-likelihood, is maximized. wever, to assess how well the model predicts future data, its predictive value has to be asured. In linear and logistic models this can be done by measures like Allen's[1] *PRESS*, llows's[2] C_p and Akaike's[3] information criterion *AIC*.

In survival analysis based on the Cox proportional hazards model,[4] *PRESS* and Mallows's C_p not available. Furthermore, *AIC* is not easy to interpret, because the components of the rtial likelihood are dependent. In this paper, the methods of Van Houwelingen and Le Cessie[5] applied to develop a measure of the predictive value of a proportional hazards model.

In the Cox model the hazard function for an individual i is given by

$$h_i(t) = h_0(t)w_i,$$

ere h_0 is a baseline hazard function and w_i is the relative risk defined by

$$w_i = \exp(X_i\beta).$$

is a row vector of covariates and β is a parameter vector. A maximum likelihood estimate of an be found by maximizing the partial likelihood

$$L(\beta) = \prod_{i=1}^{n}\left(\frac{w_i}{\sum_{R_i} w_j}\right)^{d_i}.$$

ere d_i is the censoring indicator and R_i is the risk set of all individuals still alive and uncensored st before time t_i.

In the following section a cross-validated likelihood is introduced, which serves as a measure of e predictive value of a statistical model. It is applied to a set of survival data in Section 3. In

277–6715/93/242305–10$10.00

Section 4 we illustrate how predictions can be improved by shrinkage. The application of th cross-validated likelihood in measures of explained variation is demonstrated in Section 5.

2. CROSS-VALIDATED LIKELIHOOD IN THE COX MODEL

We suppose there are n observations and a regression model is used to describe the data. The log-likelihood is denoted by $l(\beta)$, where β is the regression coefficient. We define the contribution of observation i to the log-likelihood as

$$l_i(\beta) = l(\beta) - l_{(-i)}(\beta),$$

where $l_{(-i)}(\beta)$ is the log-likelihood when observation i is left out. The value of β that maximizes $l_{(-i)}(\beta)$ is denoted by $\hat{\beta}_{(-i)}$.

If the components of the likelihood are independent, as in linear and logistic regression models, $l_i(\beta)$ simply equals the contribution of the ith component and

$$\sum_{i=1}^{n} l_i(\beta) = l(\beta).$$

We define the cross-validated log-likelihood cvl by

$$cvl = \sum_{i=1}^{n} l_i(\hat{\beta}_{(-i)}). \tag{1}$$

For a given model the cvl measures how well every observation i can be predicted using the other observations and thus serves as a measure of predictive value.

The difference between the cvl and log-likelihood is best demonstrated in the linear model with known variance. Apart from a constant, $-2 \times$ log-likelihood in this case is equal to the residual sum of squares $\Sigma(y_i - X_i\hat{\beta})^2$, while the cvl equals the predicted sum of squares $PRESS = \Sigma(y_i - X_i\hat{\beta}_{(-i)})^2$ (which in turn is approximately equal to Mallows's C_p).

In the Cox model, $l_i(\beta)$ is derived as follows. The partial likelihood when no ties are present is given by

$$L(\beta) = \prod_{j=1}^{n} \left(\frac{w_j}{\sum_{k \in R_j} w_k} \right)^{d_j}.$$

When individual i is left out, the ith factor drops out and individual i is removed from all risk sets before time t_i. If the t_is are sorted, so that $t_j < t_i$ for $j < i$, this leads to

$$L_{(-i)}(\beta) = \prod_{j<i} \left(\frac{w_j}{\sum_{k \in R_j} w_k - w_i} \right)^{d_j} \prod_{j>i} \left(\frac{w_j}{\sum_{k \in R_j} w_k} \right)^{d_j}.$$

The contribution $L_i(\beta)$ of individual i to the partial likelihood is equal to $L(\beta)/L_{(-i)}(\beta)$, which leads to

$$L_i(\beta) = \prod_{j<i} (1 - p_{ij})^{d_j} p_{ii}^{d_i}$$

with

$$p_{ij} = \frac{w_i}{\sum_{k \in R_j} w_k},$$

the probability that individual i dies at time t_j, given the risk sets and survival times. Thus, $L_i(\beta)$ is the conditional probability that individual i survives to t_{i-1} and, if $d_i = 1$, dies at time t_i. The contribution $l_i(\beta)$ to the log-likelihood is

$$l_i(\beta) = \sum_{j<i} d_j \ln(1 - p_{ij}) + d_i \ln(p_{ii}). \tag{2}$$

In the null model without covariates, $p_{ij} = 1/(n - j + 1)$ for all i, and if there is no censoring, $l_i(0) = -\ln(n)$ for all i.

For the computation of the cross-validated log-likelihood *cvl*, the leave-one-out regression coefficients $\hat{\beta}_{(-i)}$ are needed. The determination of these coefficients involves the fitting of n Cox models, each with $n - 1$ observations. Approximations for the $\hat{\beta}_{(-i)}$s are given in the Appendix.

3. CHEMOTHERAPY FOR ADVANCED OVARIAN CANCER

We consider 358 patients with advanced ovarian cancer, who were treated with chemotherapy. The survival times from 268 of these patients were analysed by Van Houwelingen *et al.*[6] Here we include a further 90 patients who received chemotherapy that differed slightly from the treatments considered in the original paper. Five prognostic variables were measured at diagnosis: FIGO stage, Karnofsky performance status, diameter of the residual tumour after surgery, Broders's grade and ascites.

FIGO stage expresses the site of the metastases as III (tumour extension beyond the pelvis or retroperitoneal nodes) or IV (distant metastases). The Karnofsky performance status measures the ability of the patient to lead her daily life on a scale from 0 (very bad) to 100 (very good): categories are $\leqslant 60$, 70, 80, 90 and 100. The diameter is classified as: microscopic, <1, 1–2, 2–5 and >5 cm. Broders's grade, a histological grade based on the percentage of undifferentiated cells present and the degree of anaplasia, has the following categories: 1, 2, 3, 4 and unknown. The last is considered as a separate category and not as a missing value. Ascites are coded as absent, present and unknown, the last again representing a separate category.

Karnofsky status and tumour diameter could have been coded as continuous covariates, but we chose a categorical coding to emphasize the difference between log-likelihood and cross-validated log-likelihood. Both these measures were computed for a series of Cox models built using a forward stepwise procedure. (This is used only for illustration and is not recommended routinely.) Furthermore, after each step Akaike's information criterion *AIC*, defined as the log-likelihood minus the degrees of freedom of the model, was computed. However, it should be repeated that the interpretation of *AIC* in the Cox model is not clear, because the components of the partial likelihood are dependent. Log-likelihood, *AIC* and *cvl* are given in Table I.

The log-likelihood increases with every extension of the model, the significance of the increment being measured by the likelihood ratio test. The contributions of Broders's grade and ascites are not significant ($p = 0{\cdot}05$).

With *AIC*, a covariate with dimension d (that is $d + 1$ categories) is entered into the model if *AIC* increases, that is if the increment of the log-likelihood is greater than d. According to this criterion, Broders's grade can be entered into the model, but ascites cannot.

The cross-validated log-likelihood *cvl* measures the predictive value of the model. For each step, the increment in *cvl* is approximately equal to that of *AIC*, but *cvl* decreases when Broders's grade is entered into the model. Thus, adding this covariate to the model will not improve the predictions derived from the model. This is also the case for ascites.

Table I. Cox models for survival with five covariates in 358 patients with advanced ovarian cancer treated by chemotherapy. Degrees of freedom (d.f.), log-likelihood (l), Akaike's information criterion (AIC) and cross-validated log-likelihood (cvl) for five Cox models and differences Δl (including p-value), ΔAIC and Δcvl from the preceding model

Step	Covariate entered	d.f.	l	Δl	AIC	ΔAIC	cvl	Δcvl
0		0	−1414·3		−1414·3		−1679·5	
1	FIGO	1	−1401·3	13·0 $p = 0·00$	−1402·3	12·0	−1667·6	12·0
2	Karnofsky status	5	−1389·3	12·0 $p = 0·00$	−1394·3	8·0	−1660·3	7·3
3	Diameter	9	−1379·0	10·3 $p = 0·00$	−1388·0	6·3	−1654·6	5·7
4	Broders's grade	13	−1374·6	4·4 $p = 0·07$	−1387·6	0·4	−1655·1	−0·5
5	Ascites	15	−1373·1	1·5 $p = 0·22$	−1388·1	−0·5	−1656·6	−1·5

We consider the model after step 4 as final, since according to all three criteria, ascites is not needed in the model.

4. MODEL CALIBRATION

The predictive value of a statistical model measures the fit to future data, while the explained variation of a model measures the fit to the data from which the model was derived. If the model is extended, the explained variation increases, but the predictive value increases less (or may even decrease). This discrepancy is described by Copas[7] as 'the amount by which the fit to new data falls short of the fit to the old data'.

In the Cox model, predictions of individual hazard or survivor functions are made by using the estimated relative risk $\hat{w}_i = \exp(X_i\hat{\beta})$ or, equivalently, the prognostic index $X_i\hat{\beta}$. In the model calibration procedure of Van Houwelingen and Le Cessie[5] the prediction process is mimicked as follows. First obtain a prognostic index $X_i\hat{\beta}_{(-i)}$ for each individual i, based on the other individuals. Then perform Cox regression on the complete data set with $X_i\hat{\beta}_{(-i)}$ as the only covariate. The log-likelihood of this model is denoted by $l^*(c)$, where c is the regression coefficient of the prognostic index. It is interesting to compare $l^*(c)$ with $l(c)$, the partial log-likelihood in the model with $X_i\hat{\beta}$ as the only covariate. These functions are plotted in Figure 1 for the final model in the example of Section 3.

First we note that $l^*(c) < l(c)$ for all c, reflecting the fact that the fit to new data is not as good as the fit to the old data. By definition of $\hat{\beta}$, $l(c)$ is maximized by $c = 1$ and $l(1)$ is a measure of the fit to the data from which the model was derived, while $l^*(1)$ measures the fit to future data. Thus $l^*(1)$ is a measure of the predictive value of the model, and as such it is an alternative to the cvl, introduced in Section 2. Furthermore, $l^*(c)$ is maximized by $\hat{c}$, which in most cases is less than 1 and is therefore called a shrinkage factor. In the final model of Section 3, $\hat{c} = 0·78$. Predictions of individual hazard and survivor functions can be improved by using the adjusted prognostic index

$$API_i = \bar{X}\hat{\beta} + \hat{c}(X_i - \bar{X})\hat{\beta},$$

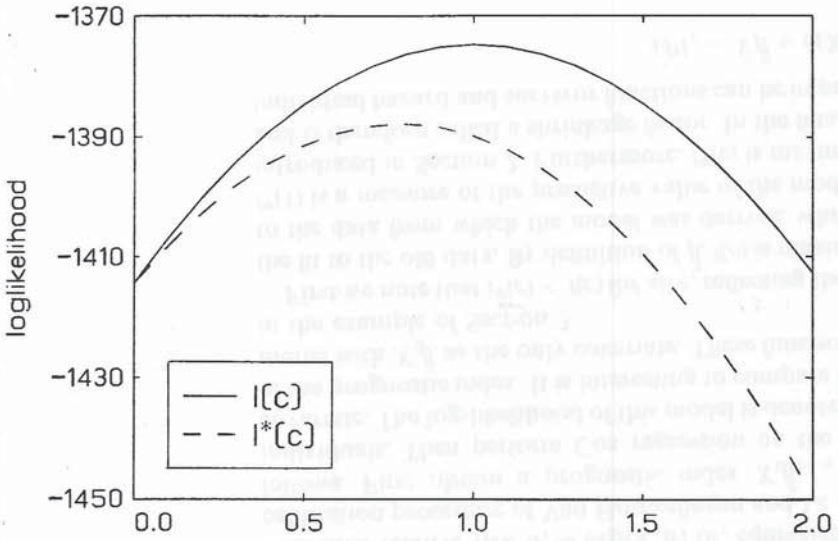

Figure 1. Partial log-likelihood $l(c)$ for the model with $X_i\hat{\beta}$ as the only covariate and $l^*(c)$ for the model with $X_i\hat{\beta}_{(-i)}$ as the only covariate

Table II. Advanced ovarian cancer treated by chemotherapy: log-likelihood $l^*(c)$ of the model with $X_i\hat{\beta}_{(-i)}$ as the only covariate at $c = 1$ and at $c = \hat{c}$ and their differences from the preceding model

Step	Covariate entered	d.f.	$l^*(1)$	$\Delta l^*(1)$	$\hat{c}$	$l^*(\hat{c})$	$\Delta l^*(\hat{c})$
0		0	− 1414·3			− 1414·3	
1	FIGO	1	− 1402·4	11·9	0·95	− 1402·4	11·9
2	Karnofsky status	5	− 1395·3	7·1	0·86	− 1394·8	7·6
3	Diameter	9	− 1389·8	5·5	0·83	− 1388·6	6·2
4	Broders's grade	13	− 1390·1	− 0·3	0·78	− 1387·9	0·7
5	Ascites	15	− 1391·6	− 1·5	0·74	− 1388·6	− 0·7

where deviations from the mean are multiplied by the shrinkage factor. The results of shrinkage in the example of Section 3 are shown in Table II.

The log-likelihood $l^*(1)$ increases until Broders's grade is entered into the model, which agrees with the behaviour of the *cvl* in Section 3. Furthermore, the differences of $l^*(1)$ from the null model, which can be computed from Table II, are approximately equal to those for *cvl* from the null model, which can be computed from Table I. In turn, these values appear approximately equal to the difference of the log-likelihood from the null model, which can be computed from those in Table I, multiplied by $\hat{c}^2$.

If the model is extended, the amount of shrinkage to the mean increases, which is demonstrated by the values of the shrinkage factor $\hat{c}$. The maximal predictive value $l^*(\hat{c})$, which is achieved if the shrinkage factor is applied in the predictions, increases until ascites is entered into the model. Thus, $l^*(\hat{c})$ increases for more steps than $l^*(1)$. If ascites is entered into the model, however, even the application of the shrinkage factor cannot prevent the predictions from becoming poorer.

In the final model of Section 3 (including FIGO stage, Karnofsky status, diameter and Broders's grade), survival curves corresponding to the 25th, 50th and 75th percentiles of the

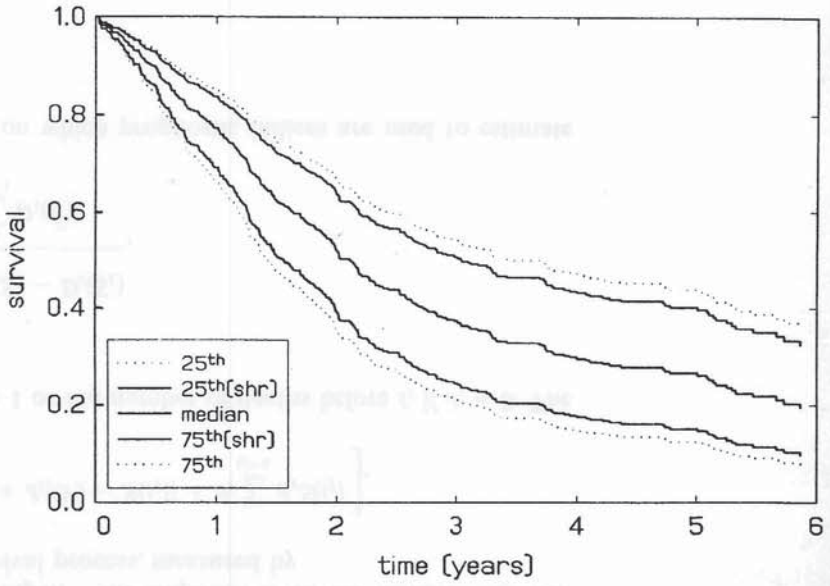

Figure 2. Estimated survivor functions corresponding to the 25th, 50th and 75th percentiles of the prognostic index $X_i\hat{\beta}$ without and with shrinkage

prognostic index $X_i\hat{\beta}$ were estimated. Figure 2 shows that after applying the shrinkage factor the good predictions become poorer and the bad predictions improve. Shrinkage has no effect on the predicted survival for the median of the prognostic indices.

5. APPLICATION OF *cvl* IN MEASURES OF EXPLAINED VARIATION

So far, we have measured explained variation by the log-likelihood. An equivalent measure is the difference between the log-likelihood $l(\hat{\beta})$ of the model and the log-likelihood $l(0)$ of the null model with no covariates. This difference can be transformed into a measure of the proportion of explained variation (Reference 8, pp. 208–209; Reference 9):

$$R^2 = 1 - \exp\left(-\frac{2}{n}(l(\beta) - l(0))\right).$$

However, R^2 measures the proportion of variation explained by the model for the data from which the model was derived. To measure the proportion of explained variation in future data, the log-likelihood can be replaced by the cross-validated log-likelihood *cvl*, which results in a cross-validated coefficient

$$R^2_{cv} = 1 - \exp\left(-\frac{2}{n}(cvl - cvl_{null})\right)$$

where cvl_{null} is the cross-validated log-likelihood for the model without covariates, as defined in Section 2. In the final model of Section 3, $R^2 = 0{\cdot}199$, while $R^2_{cv} = 0{\cdot}126$. Observe that R^2_{cv} is approximately equal to $\hat{c}^2R^2$, with $\hat{c} = 0{\cdot}78$ the shrinkage factor of the model.

The effect of censoring on a measure of explained variation can be made more explicit if the individual censoring times C_i are known. In practice this is not the case, but the C_is can all be

chosen to be equal to T_0, say, which must be less than the maximal follow-up. In that case the explained variation is censored at T_0. For the expectation of $\Sigma l_i(\beta) - l_i(0)$, with $l_i(\beta)$ and $l_i(0)$ as in equation (2), it can be shown that

$$E\left[\sum_{i=1}^{n} l_i(\beta) - l_i(0)\right] = \sum_{i=1}^{n}\left[\int_0^{T_0} \ln\left(\frac{f_i(t)}{\bar{f}(t)}\right) f_i(t)\,dt + \ln\left(\frac{S_i(T_0)}{\bar{S}(T_0)}\right) S_i(T_0)\right], \tag{3}$$

in which the S_is are the individual survivor functions, $\bar{S}$ is their mean and the f_is and $\bar{f}$ are the corresponding probability densities. This formula is valid for continuous survivor functions and needs slight adaptation in case of discontinuous S_is and $\bar{S}$. The right-hand side of (3) can be seen as a measure of the dispersion of the S_is and hence of explained variation.

We now calculate (3) for the final model of Section 3. T_0 is chosen to be 4 years, which is the minimal follow-up for all patients. Furthermore, $\bar{S}$ is estimated by the Kaplan–Meier method. If the S_is are estimated based on the prognostic indices $X_i\hat{\beta}$, the right-hand side is equal to 37·2, with corresponding $R^2 = 0{\cdot}189$. If the adjusted prognostic indices API_i of Section 4 are used instead, the result is 24·0, with corresponding $R^2 = 0{\cdot}125$. The R^2 measures are approximately equal to R^2 and R^2_{cv}, calculated above. The difference can be partly explained by the censoring of the explained variation.

Finally, we compare our results with two other measures of the proportion of explained variation for survival data. The first measure, which was proposed by Korn and Simon,[10] is of the form

$$\frac{R(\bar{S}) - \frac{1}{n}\sum_{i=1}^{n} R(S_i)}{R(\bar{S})},$$

where $R(S)$ is the mean square error for a prediction of the survival time T, based on the survivor function S. As before, the prediction error is censored at $T_0 = 4$ years and $\bar{S}$ is estimated by the Kaplan–Meier method. If the S_is are estimated using the $X_i\hat{\beta}$s, the proportion of explained variation is 0·175. If the API_is are used, the proportion is only 0·116. These values are approximately equal to R^2 and R^2_{cv}, based on (3) with $T_0 = 4$ years.

In the second measure, proposed by Schemper,[11] the emphasis is on the distance between a survivor function S and the individual survival process, measured by

$$D_i(S) = \frac{1}{k_i}\left[\sum_{t_j<t_i} d_j(1 - S(t_j)) + d_i|0{\cdot}5 - S(t_i)| + d_i\sum_{t_j>t_i} d_j S(t_j)\right],$$

where k_i is the total number of deaths if $d_i = 1$ or the number of deaths before t_i if $d_i = 0$. The related proportion of explained variation is

$$\frac{\sum_{i=1}^{n} D_i(\bar{S}) - D_i(S_i)}{\sum_{i=1}^{n} D_i(\bar{S})},$$

which is equal to 0·123 or 0·105 depending on which prognostic indices are used to estimate the S_is.

6. DISCUSSION

In this paper we have considered predictive value as an alternative for explained variation. Each concept gives a certain perspective on a statistical model. The explained variation, measured by the log-likelihood, refers to the data from which the model is derived. The predictive value refers to the fit to future data. It can be measured by *AIC* (if the components of the likelihood are independent) or *cvl* (in the Cox model).

The use of the log-likelihood (and associated *P*-value) or, on the other hand, of *AIC* (or *cvl*) can lead to different selections of covariates. This can be explained by the significance level α of *AIC*. For a covariate with dimension d (that is, $d+1$ categories), $\alpha = \Pr(AIC > 0) = \Pr(\Delta l > d)$, in which Δl denotes the difference in log-likelihood and the probabilities are conditional under the null hypothesis that the covariate has no effect. This leads to $\alpha = \Pr(\chi^2_{[d]} > 2d)$. If $d < 7$, then $\alpha > 0{\cdot}05$ and *AIC* is less conservative than the likelihood ratio test, which was demonstrated in Section 3 for Broders's grade. On the other hand, *AIC* is more conservative than the likelihood ratio test if $d \geqslant 7$.

A similar phenomenon may be observed when the *cvl* is used. It is possible that the inclusion of a covariate does not increase the log-likelihood significantly, while the same covariate may increase the *cvl*. The opposite can also occur when, for instance, a high-dimensional interaction increases the log-likelihood significantly, but decreases the *cvl*.

Another way to assess predictive value is by performing Cox regression in a model with the prognostic index $X_i\hat{\beta}_{(-i)}$ as the only covariate. The log-likelihood $l^*(c)$ of this model is maximized by the shrinkage factor $\hat{c}$. Application of $\hat{c}$ in the prognostic index leads to an adjusted index, which is more realistic.

A third method to assess predictive value is to use a R^2-type measure of the proportion of explained variation, in which the adjusted prognostic indices are used. This leads to a smaller, but again more realistic, proportion of explained variation.

APPENDIX: APPROXIMATE COEFFICIENTS $\hat{\beta}_{(-i)}$

We consider four ways of approximating the coefficients $\hat{\beta}_{(-i)}$, which are not restricted to survival analysis but can be applied to other regression models as well.

The first utilizes that $\hat{\beta}_{(-i)}$ by definition maximizes $l_{(-i)}(\beta)$, so the first derivative must equal zero at $\beta = \hat{\beta}_{(-i)}$. A first-order Taylor approximation at $\beta = \hat{\beta}$ leads to

$$\hat{\beta}_{(-i)} = \hat{\beta} + \left(\frac{\partial^2 l}{\partial \beta^2} - \frac{\partial^2 l_i}{\partial \beta^2}\right)^{-1} \frac{\partial l_i}{\partial \beta}, \tag{4}$$

where all derivatives are evaluated at $\hat{\beta}$. In linear models, (4) gives exact values of $\hat{\beta}_{(-i)}$, whereas in logistic models the results are equivalent to the one-step approximations of Pregibon.[12]

In the Cox model, the first derivative of $l_i(\beta)$ is given by

$$\frac{\partial l_i}{\partial \beta}(\beta) = -\sum_{j<i} d_j \frac{p_{ij}}{1-p_{ij}}(X_i - \bar{X}_j) + d_i(X_i - \bar{X}_i),$$

with p_{ij} defined as in Section 2 and

$$\bar{X}_j = \frac{\sum_{k \in R_j} w_k X_k}{\sum_{k \in R_j} w_k} = \sum_{k \in R_j} p_{kj} X_k,$$

the weighted mean of the covariates of all individuals in R_j. The second derivative of $l_i(\beta)$ is

$$\frac{\partial^2 l_i}{\partial \beta^2}(\beta) = -\sum_{j<i} d_j \left[\frac{p_{ij}}{(1-p_{ij})^2}(X_i - \bar{X}_j)(X_i - \bar{X}_j)' - \frac{p_{ij}}{1-p_{ij}} \operatorname{var}(X_j) \right] - d_i \operatorname{var}(X_i),$$

with

$$\operatorname{var}(X_j) = \sum_{k \in R_j} p_{kj}(X_k - \bar{X}_j)(X_k - \bar{X}_j)',$$

the weighted variance of the covariates of all individuals in R_j.

The computation of the second derivatives can consume a lot of computer time and/or storage. Omitting this term leads to

$$\hat{\beta}_{(-i)} = \hat{\beta} + \left(\frac{\partial^2 l}{\partial \beta^2} \right)^{-1} \frac{\partial l_i}{\partial \beta}, \tag{5}$$

which serves as a second approximation to the coefficients.

The third method is to add an indicator variable Z_i for individual i to the model, that is $X\beta$ is replaced by $X\beta^* + Z_i\gamma$. In models with independent components, this method has the following features: $\hat{\beta}_{(-i)}$ is the maximum likelihood estimate of β^*, $\hat{\beta}_{(-i)}$ is already well approximated after one Newton–Raphson iteration (with initial values $\hat{\beta}$ and 0 for β^* and γ respectively) and the results are the same as in the first method. In the Cox model these features do not hold, but the augmented model can still give a good approximation for $\hat{\beta}_{(-i)}$.

Finally we consider the method of Storer and Crowley,[13] who proposed to add two time-dependent indicator variables $Z1_i(t)$ and $Z2_i(t)$ for individual i to the model. For individual i, $Z1_i(t) = 1$ if $t < t_i$ and $Z2_i(t) = 1$ if $t = t_i$. Furthermore $Z1_i(t)$ and $Z2_i(t)$ are identically zero for all other individuals. The augmented Cox model for the hazard function is

$$h(t) = h_0(t) \exp(X\beta^* + Z1_i(t)\gamma_1 + Z2_i(t)\gamma_2).$$

The values of β^*, γ_1 and γ_2 that maximize the partial likelihood of this model are $\hat{\beta}_{(-i)}$, $-\infty$ and $+\infty$ respectively. Because of the infinite values Newton–Raphson iterations will not converge, but again only one iteration (with initial values $\beta^* = \hat{\beta}$, $\gamma_1 = 0$ and $\gamma_2 = 0$) can give a good approximation of $\hat{\beta}_{(-i)}$.

All four methods lead to nearly the same approximations of the leave-one-out regression coefficients and hence to approximately the same values of *cvl*.

REFERENCES

1. Allen, D. M. 'The relation between variable selection and data augmentation and a method for prediction', *Technometrics*, **16**, 125–127 (1974).
2. Mallows, C. L. 'Some comments on C_p', *Technometrics*, **15**, 661–675 (1973).
3. Akaike, H. 'Information theory and an extension of the entropy maximization principle', in Petrov, B. N. and Csak, F. (eds.), *Proceedings of the Second International Symposium on Information Theory*, Akademia, Kiado, 1973, pp. 267–281.
4. Cox, D. R. 'Regression models and life tables' (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
5. Van Houwelingen, J. C. and Le Cessie, S. 'Predictive value of statistical models', *Statistics in Medicine*, **9**, 1303–1325 (1990).
6. Van Houwelingen, J. C., Ten Bokkel Huinink, W. W., Van der Burg, M. E. L., Van Oosterom, A. T. and Neijt, J. P. 'Predictability of the survival of patients with advanced ovarian cancer', *Journal of Clinical Oncology*, **7**, 769–773 (1989).

7. Copas, J. B. 'Regression, prediction and shrinkage' (with discussion), *Journal of the Royal Statistical Society, Series B*, **45**, 311–354 (1983).
8. Cox, D. R. and Snell, E. J. *The Analysis of Binary Data*, second edition, Chapman and Hall, London, 1989.
9. Nagelkerke, N. J. D. 'A note on the general definition of the coefficient of determination', *Biometrika*, **78**, 691–692 (1991).
10. Korn, E. L. and Simon, R. 'Measures of explained variation for survival data', *Statistics in Medicine*, **9**, 487–503 (1990).
11. Schemper, M. 'The explained variation in proportional hazards regression', *Biometrika*, **77**, 216–218 (1990).
12. Pregibon, D. 'Logistic regression diagnostics', *Annals of Statistics*, **9**, 705–724 (1981).
13. Storer, B. E. and Crowley, J. 'A diagnostic for Cox regression and general conditional likelihoods', *Journal of the American Statistical Association*, **80**, 139–147 (1985).