

## INFLUENCE DIAGNOSTICS FOR THE PROPORTIONAL HAZARDS MODEL

L.A. WEISSFELD

*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA*

Received April 1989

Revised July 1989

*Abstract:* Influence diagnostics based on Cook's curvature diagnostic (1986) are developed for the proportional hazards model. Three perturbation schemes are considered: perturbation of the likelihood, perturbation of the censoring information and perturbation of covariate values.

*Keywords:* Censored data, curvature, influence, perturbation diagnostic.

### 1. Introduction

The Cox proportional hazards model is often the model of choice when analysing censored data and many methods have been proposed for residual analysis and assessment of influential observations for this model. Problems existed with the earlier methods of residual analysis (Crowley and Storer, 1983); however Barlow and Prentice (1988) have defined residuals that are closely related to the empirical influence function and other standard regression diagnostics circumventing these problems. Other commonly used regression diagnostics for the linear regression model have also been extended to the Cox proportional hazards model. These include the influence function (Reid and Crépeau, 1985; Cain and Lange, 1984) and estimates of the deleted values of the regression coefficients based on an augmented design matrix (Storer and Crowley, 1985).

The focus of this paper is on the application of Cook's (1986) influence diagnostic to the proportional hazards model. This approach is based on perturbation of the likelihood function and perturbation of covariates included in the model. The proposed diagnostics differ from the diagnostic presented in Pettit and Bin Daud (1989) which is based on an approximation to the likelihood displacement. These methods are illustrated using a lung cancer data set (Kalbfleisch and Prentice, 1980) and the Stanford heart transplant data (Miller and Halpern, 1982).

### 2. The Cox proportional hazards model

Let  $T_1, \dots, T_n$  be the observed time to event and let  $\delta_i = I\{T_i \leq U_i\}$ , where  $U_1, \dots, U_n$  are censoring times and  $U_i$  is assumed to be independent of  $T_i$ . Then, under the assumption of the proportional hazards model, the hazard function is given by

$$\lambda(t|x) = \lambda_0(t) \exp(x^T \beta),$$

where  $\lambda_0(t)$  is an arbitrary baseline hazard function,  $(\beta_1, \dots, \beta_p)^T$  is a parameter vector, and  $(x_1, \dots, x_p)^T$

is a covariate vector. Estimates of  $\beta$  are obtained from the partial likelihood function (Cox, 1972) which is given by

$$L(\beta) = \prod_{i=1}^n \left\{ \frac{\exp(x_i^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)} \right\} \delta_i, \tag{1}$$

where  $x_i$  denotes the covariate vector associated with the  $i$ th time and  $R(t_i)$  denotes the risk set at  $t_i$ . Estimates of  $\beta$  are obtained by maximizing the log of the likelihood function and covariance estimates are obtained from the inverse of the information matrix. In particular, the likelihood equations obtained from (1) are give by

$$0 = \frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^n \delta_i \left\{ x_{ij} - \frac{\sum_{l \in R(t_i)} x_{lj} \exp(x_l^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)} \right\}, \quad j = 1, \dots, p, \tag{2}$$

and estimates are obtained by solving (2) for  $\beta_1, \dots, \beta_p$ .

**3. Curvature diagnostics**

To measure the influence that observations have on the parameter estimates, the effect of perturbations on either elements in the score vector or the covariate vector can be measured by computing the curvature of the likelihood function (Cook, 1986). Let  $\omega = (\omega_1, \dots, \omega_n)$  denote the vector of perturbations, then the curvature is defined by

$$C_i = 2 |I^T \ddot{F} I|,$$

where  $|I| = 1$ , the  $(i, j)$ th element of  $\ddot{F}$  is given by  $\partial^2 L(\hat{\beta} | \omega) / \partial \omega_i \partial \omega_j$  and  $L(\hat{\beta} | \omega)$  denotes the estimated perturbed likelihood. The computation of this measure is simplified by noting that

$$\ddot{F} = \Delta^T I^{-1} \Delta, \tag{3}$$

where  $I$  is the information matrix and the  $ij$ th element of  $\Delta$  is given by  $\partial^2 \log L(\beta | \omega) / \partial \beta_i \partial \omega_j$  evaluated at  $\beta = \hat{\beta}$  and  $\omega = \omega_0$ . The diagnostic of interest is based on the maximum curvature,  $C_{\max}$ , which is the maximum eigenvalue of  $\ddot{F}$  defined in (3). The corresponding eigenvector,  $l_{\max}$ , is then examined with large or small values pointing to observations which are possibly influential.

The application of this diagnostic depends upon the choice of a perturbation scheme. For the Cox proportional hazards model there are many ways to perturb the data: the individual elements of the sum in the likelihood function can be perturbed, the vector indicating whether or not an individual is censored,  $\delta_i$ , can be perturbed, or the covariate values can be perturbed. Perturbation of the likelihood provides information about the stability of the likelihood function as well as information regarding the likelihood ratio test. To implement the first perturbation scheme, that is, perturbation of the likelihood function, the elements of the likelihood given in (1) are perturbed as follows:

$$\log L(\beta | \omega) = \sum_{i=1}^n \delta_i \omega_i \left\{ x_i^T \beta - \log \left[ \sum_{l \in R(t_i)} e^{x_l^T \beta} \right] \right\}.$$

Based on this perturbation scheme,  $\Delta$  takes on the form

$$\Delta_{ji} = \frac{\partial^2 \log L(\beta | \omega)}{\partial \beta_j \partial \omega_i} = \delta_i x_{ij} - \delta_i \frac{\sum_{l \in R(t_i)} x_{lj} \exp(x_l^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)}, \quad i = 1, \dots, n, \quad j = 1, \dots, p, \tag{4}$$

where the derivative is evaluated at  $\omega_i = 0$ . It is of interest to note that the  $\Delta_{ji}$ 's are Schoenfeld's (1982)

partial residuals. The appropriate curvature diagnostic will then be based on the eigenvector associated with the maximum eigenvalue of  $\ddot{F} = \Delta^T I^{-1} \Delta$  as defined in (3) and (4). The disadvantage of this perturbation scheme is that all censored observations, that is, those observations with  $\delta_i = 0$ , will have a curvature of 0 under this definition. Thus, censored observations which are influential, will not be flagged by this diagnostic.

Another type of perturbation scheme can be used to circumvent this problem. This involves perturbing  $\delta = (\delta_1, \dots, \delta_n)$  by an amount  $\omega = (\omega_1, \dots, \omega_n)$ . This is the same as perturbing the counting process  $N_i(t) = I\{T_i \leq t, \delta_i = 1\}$  which counts the number of events for each observation. Under this scheme the perturbed likelihood takes on the form

$$\log L(\beta | \omega) = \sum_{i=1}^n \left\{ (\delta_i + \omega_i) x_i^T \beta + (\delta_i + \omega_i) \log \left[ \sum_{l \in R(t_i)} \exp(x_l^T \beta) \right] \right\}$$

and the matrix  $\Delta$  needed for computation of  $\ddot{F}$  in (3) is given by

$$\Delta_{ji} = \frac{\partial^2 \log L(\beta | \omega)}{\partial \beta_j \partial \omega_i} = x_{ij} - \frac{\sum_{l \in R(t_i)} x_{lj} \exp(x_l^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)}, \quad j = 1, \dots, p, \quad i = 1, \dots, n, \tag{5}$$

where the derivative is evaluated at  $\omega_i = 0$ . The curvature diagnostic based on this scheme has the advantage of being able to detect censored observations which may be influential as well as uncensored observations. This diagnostic has the same form as that given in (4); however, values are defined for both censored and uncensored observations, while (4) takes on nonzero values only for uncensored observations.

To determine the effect of perturbations in the covariate matrix,  $X$ , the  $k$ th covariate is perturbed as follows:

$$x'_{ik} = x_{ik} + \omega_{ik} s_k, \quad i = 1, \dots, n,$$

where  $s_k$  is a scale factor, typically, the standard deviation of the  $k$ th coefficient. The maximum curvature is computed as before with  $\Delta$  being of dimension  $p \times np$ . If  $\Delta$  is partitioned into  $p$  submatrices, that is,  $\Delta = (\Delta_1, \dots, \Delta_p)$ , then the  $j$ th submatrix is given by

$$\begin{aligned} \Delta_{mkj} &= \frac{\partial^2 \log L(\beta | \omega)}{\partial \beta_j \partial \omega_{mk}} \\ &= \sum_{i=1}^n \left\{ \frac{\beta_k s_k \exp(x_m^T \beta) \sum_{l \in R(t_i)} x_{lj} \exp(x_l^T \beta)}{\left\{ \sum_{l \in R(t_i)} \exp(x_l^T \beta) \right\}^2} - \frac{x_{mj} \beta_k s_k \exp(x_m^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)} \right\} \delta_i \delta_{m \in R(t_i)} \\ &\quad (j \neq k, \quad j, k = 1, \dots, p, \quad m = 1, \dots, n) \\ &= \delta_m s_k - \sum_{i=1}^n \left\{ \frac{s_k \exp(x_m^T \beta) + x_{mj} \beta_k s_k \exp(x_m^T \beta)}{\sum_{l \in R(t_i)} \exp(x_l^T \beta)} \right. \\ &\quad \left. - \frac{\beta_k s_k \exp(x_m^T \beta) \sum_{l \in R(t_i)} x_{lk} \exp(x_l^T \beta)}{\left[ \sum_{l \in R(t_i)} \exp(x_l^T \beta) \right]^2} \right\} \delta_i \delta_{m \in R(t_i)} \\ &\quad (j = k = 1, \dots, p, \quad m = 1, \dots, n), \tag{6} \end{aligned}$$

where  $t_1 < \dots < t_n$  are the observed times,  $\delta_{m \in R(t_i)} = 1$  if  $m \in R(t_i)$  and 0 otherwise, and the derivative is

evaluated at  $\omega_{jk} = 0$ . Thus, for the  $j$ th covariate, the curvature diagnostic is based on the eigenvector associated with the maximum eigenvalue of

$$\Delta_j^T I^{-1} \Delta_j,$$

where  $\Delta_j$  is as defined in (6). This diagnostic has an advantage since it measures influence in terms of the location of observations in the covariate space.

#### 4. Examples

The use of the perturbation diagnostics presented in the previous section is illustrated in several examples. The first of these examples is the Stanford heart transplant data presented in Miller and Halpern (1982). Table 1 presents the curvature diagnostics for selected observations, obtained from fitting a proportional hazards model with age and mismatch score included as covariates. Examining the results obtained from perturbation of the covariates, the observations with failure times of 538 and 2474 days are influential for the estimation of the mismatch score coefficient, while observations with survival times of 1612 and 3410 days are flagged as influential for the estimation of the age coefficient. When considering the observations flagged by the diagnostics based on perturbation of the failure times or censoring vector, given by (4) and (5), observations with failure times at 10, 42 and 86 are flagged as influential, with the influential observation at time 10 being a subject aged 13.

To assess the importance of each of these observations, they were deleted from the data set one at a time. Based on these deletion results, deletion of the failure time at 538 days results in a large change in the standard error of the coefficient for mismatch score as well as resulting in a change of 0.4 standard errors in the age coefficient. Similarly, deletion of the failure at 2474 days results in a large change in the standard error of the age coefficient and a change of 0.11 standard errors in the coefficient for mismatch score. The observation flagged as most influential for age, the survival time of 1612 days, had little effect on either age or mismatch score when it was deleted from the data set. Deletion of the failures at 10, 42 and 86 days all result in large changes in the coefficients for both age and mismatch score. The magnitude of change observed when either of the failures at 538 and 2474 days, or the survival time of 1612 days are deleted is not surprising since the eigenvalue for the age coefficient is 0.0098, while the maximum eigenvalue for the mismatch score is 0.0161. Neither of these eigenvalues indicate that perturbation of the covariate values will cause substantial changes in the results obtained from this model. On the other hand, the maximum eigenvalues for the perturbation of the times and  $\delta_i$  are 1.6 and 2.3, respectively indicating that observations flagged by these diagnostics will have an impact on the results obtained from this model.

Table 1  
Curvature diagnostics obtained from the Stanford heart transplant data example

Time	$\delta_i$	Age	Mismatch score	Curvature diagnostics			
				Age	Mismatch score	Likelihood	$\delta_i$
538	1	49	276	0.006	0.345	-0.151	-0.082
2474	1	52	170	-0.075	0.197	-0.105	-0.074
2805	0	48	120	-0.140	0.019	0.000	-0.045
3410	0	45	98	-0.143	-0.021	0.000	-0.023
1612	0	51	125	-0.153	0.032	0.000	-0.072
10	1	13	149	0.114	0.001	0.237	0.225
86	1	12	126	0.114	0.002	0.244	0.225
42	1	19	63	0.113	-0.008	0.243	0.201

Table 2  
Curvature diagnostics for lung cancer data

Obs.	Time	$\delta_i$	Curvature diagnostics							Likelihood	$\delta_i$										
			$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$												
36	287	1	60	66	25	0	1	0	1	0	1	0.314	-0.497	-0.406	0.542	-0.746	-0.254	-0.482	0.058	0.058	
30	177	1	50	66	16	0	0	1	0	1	0	1	0.326	-0.402	-0.138	0.357	0.053	-0.126	-0.363	0.068	0.075
34	231	0	50	52	9	1	0	0	0	0	0	0	0.459	-0.051	-0.130	-0.216	-0.040	-0.281	0.134	0.000	-0.034
22	90	1	60	50	22	0	0	1	0	0	0	0	-0.029	0.093	-0.092	0.174	0.087	-0.557	0.182	0.088	0.079
18	51	1	30	59	87	0	1	0	0	0	0	0	0.302	-0.106	-0.763	0.145	-0.305	0.034	0.157	-0.086	-0.041
3	8	1	40	63	58	1	0	0	1	0	0	1	-0.135	0.104	0.083	-0.129	0.093	0.166	0.086	-0.496	-0.447
6	11	1	70	48	11	1	1	0	0	1	0	1	-0.146	0.117	0.109	-0.122	0.094	0.164	0.093	-0.404	-0.405

$x_1$  = performance status,  $x_2$  = age,  $x_3$  = months since diagnosis,  $x_4$  = squamous cell,  $x_5$  = small cell,  $x_6$  = adeno,  $x_7$  = standard/test.

In fact, deletion of any of these three observations, the failures at 10, 42 and 86 days, has a large impact on parameter estimates.

This example was also examined in Reid and Crépeau (1985) and Pettit and Bin Daud (1989). The results obtained with the curvature diagnostics presented here are similar to those obtained from the influence function diagnostic in Reid and Crépeau (1985). The observation with a failure time of 10 days and age of 56 was not flagged as influential by the curvature diagnostic while it was flagged by the influence function. Deletion of this observation results in the largest change in the coefficient of mismatch score while deletion of the most influential observation flagged by the mismatch curvature diagnostic, the failure at 538 days, resulted in the largest change in the standard error of the mismatch coefficient.

For comparison with the analysis in Pettit and Bin Daud (1989), curvature diagnostics were also computed for the Stanford heart transplant data with the model including individuals who survived more than 10 days and the covariates  $(\text{age} - 41.9)/10$ ,  $(\text{age} - 41.9)^2/100$  and mismatch score. The results obtained did not agree well with those of Pettit and Bin Daud. As in Pettit and Bin Daud (1989), several observations with ages less than 20 were flagged as influential; however, observations of other types were also flagged by the diagnostics presented here. In particular, observations with ages of 58, 52, 56 and 53 were also flagged. These observations all had an impact on parameter estimates when deleted from the data set.

Table 2 presents results from the lung cancer data set given in Kalbfleisch and Prentice (1980). This data set includes information on 40 patients with lung cancer and examines the effectiveness of a treatment while controlling for age, months treated, performance status, tumor type and treatment. The results indicate that observations 30 and 36 are the two most influential observations in the data set. There is nothing to distinguish these two observations as being influential; however, deletion of either of these observations results in substantial changes in most of the coefficients. In particular, deletion of observation 30 causes the effect of the variable  $x_6$ , the indicator variable for adeno tumor cell type, to become statistically significant, while deletion of observation 36 causes the significance level of the coefficient associated with this variable to become over twice as large. Observations 33 and 18 were also flagged as being influential by the curvature diagnostics based on perturbation of the covariate values, but deletion of these observations resulted in little change in the parameter estimates. When the diagnostics for the failure times and perturbation of  $\delta_i$  were examined, observations 3 and 6 were flagged as the most influential. Deletion of either of these observations from the data set resulted in changes in the parameter estimates; but these changes were not large enough to affect the inferences drawn from this analysis. These results agree well with those obtained by Reid and Crépeau (1985).

## 5. Conclusions

The curvature diagnostics presented in this paper are useful for detecting influential observations when the Cox proportional hazards model is fit to censored data. In particular, the diagnostic based on perturbation of covariate values (6) is useful for locating observations which have an impact on estimation of individual coefficients. The value of  $\omega_i$ , the perturbation vector, can also be varied so that a more detailed influence analysis can be done. The diagnostics based on perturbing the likelihood function (4) and the censoring vector (5),  $\delta_i$ , are also useful for this type of analysis; however, the diagnostic based on the censoring information is more useful since it treats both censored and uncensored observations in the same way. These diagnostics are useful for flagging observations which may influence the results obtained from the likelihood ratio test.

The results obtained in the examples also indicate that censored observations can be influential even though they do not play the same role in the estimation of parameters as the uncensored observations. This is not the case for the diagnostic based on perturbation of the likelihood (4), where censored observations will always have a value of 0. Those censored observations which are influential will be flagged by the

other two diagnostics as well as the diagnostics presented by Cain and Lange (1984), Pettit and Bin Daud (1989) and Reid and Crépeau (1985). The curvature diagnostics presented here have the added advantage of flagging observations which are influential due to masking.

## References

- Barlow, W.E. and R.L. Prentice (1988), Residuals for relative risk regression, *Biometrika* **75**, 65–74.
- Cain, K.C. and N.T. Lange (1984), Approximate case influence for the proportional hazards regression model with censored data, *Biometrics* **40**, 493–499.
- Cook, R.D. (1986), Assessment of local influence, *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.
- Crowley, J. and B.E. Storer (1983), Comment on a reanalysis of the Stanford heart transplant data (by Aitkin et al.), *J. Amer. Statist. Assoc.* **72**, 27–36.
- Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data* (Wiley, New York).
- Miller, R. and J. Halpern (1982), Regression with censored data, *Biometrika* **69**, 521–531.
- Pettit, A.N. and I. Bin Daud (1989), Case-weighted measures of influence for proportional hazards regression, *Appl. Statist.* **38**(1), 51–67.
- Reid, N. and H. Crépeau (1985), Influence functions for proportional hazards regression, *Biometrika* **72**, 1–9.
- Schoenfeld, D. (1982), Partial residuals for the proportional hazards regression model, *Biometrika* **69**, 239–241.
- Storer, B.E. and J. Crowley (1985), A diagnostic for Cox regression and general conditional likelihoods, *J. Amer. Statist. Assoc.* **80**, 139–147.