Applied Survival Analysis

Hosmer & lemeshow

1999

CHAPTER 6

Assessment of Model Adequacy

6.1 INTRODUCTION

Model-based inferences depend completely on the fitted statistical model. For these inferences to be "valid" in any sense of the word, the fitted model must provide an adequate summary of the data upon which it is based. A complete and thorough examination of a model's fit and adherence to model assumptions is just as important as careful model development.

The goal of statistical model development is to obtain the model which best describes the "middle" of the data. The specific definition of "middle" depends on the particular type of statistical model, but the idea is basically the same for all statistical models. In the normal errors linear regression model setting, we can describe the relationship between the observed outcome variable and one of the covariates with a scatterplot. This plot of points for two or more covariates is often described as the "cloud" of data. In model development we find the regression line, plane or hyperplane that best fits/splits the cloud. The notion of "best" in this setting means that we have equal distances from observed points to fitted points above and below the surface. A "generic" main effects model with some nominal covariates, which treats continuous covariates as linear, may not have enough tilts, bends or turns to fit/split the cloud. Each step in the model development process is designed to tailor the regression surface to the observed cloud of data.

In most, if not all, applied settings the results of the fitted model will be summarized for publication using point and interval estimates of clinically interpretable measures. Examples of summary measures include the mean difference in linear regression, the odds ratio in logistic regression and the hazard ratio for the proportional hazards regression

model. Since any summary measure is only as good as the model it is based on, it is vital that one evaluate how well the fitted regression surface describes the data cloud. This process is generally referred to as *assessing the adequacy of the model*; like model development, it involves a number of steps. Performing these in a thorough and conscientious manner will assure that the inferential conclusions based on the fitted model are the best and most valid possible.

The methods for assessment of a fitted proportional hazards model are essentially the same as for other regression models, and we assume some experience with these, particularly with logistic regression [see Hosmer and Lemeshow (1989, Chapter 5)]. Requirements for model assessment are: (1) methods for testing the assumption of proportional hazards, (2) subject-specific diagnostic statistics that extend the notions of leverage and influence to the proportional hazards model and (3) overall summary measures of goodness-of-fit.

6.2 RESIDUALS

Central to the evaluation of model adequacy in any setting is an appropriate definition of a residual. As we discussed in Chapter 1, the fact that the outcome variable is time to some event and the observed values may be incomplete or censored is what sets a regression analysis of survival time apart from other regression models. In earlier chapters we suggested that the semiparametric proportional hazards model is a useful model for data of this type and we described why and how it may be fit using the partial likelihood. This combination of data, model and likelihood make definition of a residual much more difficult in modeling survival time than is the case with other statistical models.

Consider a logistic regression analysis of a binary outcome variable. In this setting, values of the outcome variable are "present" ($y = 1$) or "absent" ($y = 0$) for all subjects. The fitted model provides estimates of the probability that the outcome is present (i.e., the mean of Y). Thus, a natural definition of the residual is the difference between the observed value of the outcome variable and that predicted by the model. This form of the residual also follows as a natural consequence of characterizing the observed value of the outcome as the sum of a systematic component and an error component. The two key assumptions in this definition of a residual are: (1) the value of the outcome is known and (2) the fitted model provides an estimate of the "mean of the dependent variable" or systematic component of the model. Since assumption 2

and, more than likely, assumption 1 are not true when using the partial likelihood to fit the proportional hazards model to censored survival data, there is no obvious analog to the usual "observed minus predicted" residual used with other regression models.

The absence of an obvious residual has lead to the development of several different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazards model. Most software packages provide access to at least one of these residuals. Only two packages, SAS and S-PLUS, have full residual analysis capabilities at this time. This situation is likely to change as other packages update and modify their proportional hazards routines.

We assume, for the time being, that there are p covariates and that the n independent observations of time, covariates and censoring indicator are denoted by the triplet $(t_i, \mathbf{x}_i, c_i)$, $i = 1, 2, \ldots, n$, where $c_i = 1$ for uncensored observations and is zero otherwise. Schoenfeld (1982) proposed the first set of residuals for use with a fitted proportional hazards model and packages providing them refer to them as the "Schoenfeld residuals." These are based on the individual contributions to the derivative of the log partial likelihood. This derivative for the kth covariate is shown in (3.21) and is repeated here as

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{n} c_i \left\{ x_{ik} - \frac{\sum\limits_{j \in R(t_i)} x_{jk} e^{\mathbf{x}_j' \boldsymbol{\beta}}}{\sum\limits_{j \in R(t_i)} e^{\mathbf{x}_j' \boldsymbol{\beta}}} \right\}$$

$$= \sum_{i=1}^{n} c_i \left\{ x_{ik} - \bar{x}_{w_i k} \right\} , \qquad (6.1)$$

where

$$\bar{x}_{w_i k} = \frac{\sum\limits_{j \in R(t_i)} x_{jk} e^{\mathbf{x}_j' \boldsymbol{\beta}}}{\sum\limits_{j \in R(t_i)} e^{\mathbf{x}_j' \boldsymbol{\beta}}} . \qquad (6.2)$$

The estimator of the Schoenfeld residual for the ith subject on the kth covariate is obtained from (6.1) by substituting the partial likelihood estimator of the coefficient, $\hat{\boldsymbol{\beta}}$, and is

$$\hat{r}_{ik} = c_i \left(x_{ik} - \hat{\bar{x}}_{w_i k} \right), \qquad (6.3)$$

where

$$\hat{\bar{x}}_{w_i k} = \frac{\sum\limits_{j \in R(t_i)} x_{jk} e^{\mathbf{x}_j' \hat{\boldsymbol{\beta}}}}{\sum\limits_{j \in R(t_i)} e^{\mathbf{x}_j' \hat{\boldsymbol{\beta}}}}$$

is the estimator of the risk set conditional mean of the covariate. Since the partial likelihood estimator of the coefficient, $\hat{\boldsymbol{\beta}}$, is the solution to the equations obtained by setting (6.1) equal to zero, the sum of the Schoenfeld residuals is zero. Software packages set the value of the estimate of the Schoenfeld residual to missing for subjects whose observed survival time is censored.

Grambsch and Therneau (1994) suggest that scaling the Schoenfeld residuals by an estimator of its variance yields a residual with greater diagnostic power than the unscaled residuals. Let the vector of p Schoenfeld residuals for the ith subject be denoted as

$$\hat{\mathbf{r}}_i' = (\hat{r}_{i1}, \hat{r}_{i2}, \ldots, \hat{r}_{ip}) ,$$

where $\hat{r}_{ik}$ is the estimator in (6.3), with the convention that $\hat{r}_{ik} =$ missing if $c_i = 0$. Let the estimator of the $p \times p$ covariance matrix of the vector of residuals for the ith subject, as reported in Grambsch and Therneau (1994), be denoted by $\widehat{\mathrm{Var}}(\hat{\mathbf{r}}_i)$, and the estimator is missing if $c_i = 0$. The vector of scaled Schoenfeld residuals is the product of the inverse of the covariance matrix times the vector of residuals, namely

$$\hat{\mathbf{r}}_i^* = \left[\widehat{\mathrm{Var}}(\hat{\mathbf{r}}_i) \right]^{-1} \hat{\mathbf{r}}_i . \qquad (6.4)$$

The elements in the covariance matrix $\widehat{\mathrm{Var}}(\hat{\mathbf{r}}_i)$ are, in the current setting, a weighted version of the usual sum-of-squares matrix computed using the data in the risk set. For the ith subject, the diagonal elements in this matrix are

$$\widehat{\mathrm{Var}}(\hat{\mathbf{r}}_i)_{kk} = \sum_{j \in R(t_i)} \hat{w}_{ij} \left(x_{jk} - \hat{\bar{x}}_{w_i k} \right)^2 ,$$

and the off-diagonal elements are

$$\hat{\text{Var}}(\hat{\mathbf{r}}_i)_{kl} = \sum_{j \in R(t_i)} \hat{w}_{ij}\left(x_{jk} - \hat{\bar{x}}_{w_i k}\right)\left(x_{jl} - \hat{\bar{x}}_{w_i l}\right)$$

where

$$\hat{w}_{ij} = \frac{e^{\mathbf{x}_j'\hat{\boldsymbol{\beta}}}}{\sum_{l \in R(t_i)} e^{\mathbf{x}_l'\hat{\boldsymbol{\beta}}}} \;.$$

Grambsch and Therneau (1994) suggest use of an easily computed approximation for the scaled Schoenfeld residuals. This suggestion is based on their experience that the matrix, $\hat{\text{Var}}(\hat{\mathbf{r}}_i)$, tends to be fairly constant over time. If this matrix is constant, its inverse may be approximated by multiplying the estimator of the covariance matrix of the estimated coefficients by the number of events (i.e., the observed number of uncensored survival times m),

$$\left[\hat{\text{Var}}(\hat{\mathbf{r}}_i)\right]^{-1} = m\hat{\text{Var}}(\hat{\boldsymbol{\beta}}).$$

The approximate scaled Schoenfeld residuals are the ones computed by software packages, namely

$$\hat{\mathbf{r}}_i^* = m\hat{\text{Var}}(\hat{\boldsymbol{\beta}})\hat{\mathbf{r}}_i. \tag{6.5}$$

Subsequent references to the scaled Schoenfeld residuals, $\hat{\mathbf{r}}_i^*$, will mean the approximation in (6.5), not the true scaled residual in (6.4).

The counting process approach is an extremely useful and powerful tool for studying the proportional hazards model. Most descriptions of it, however, including those in statistical software manuals, are difficult to understand without knowledge of calculus. In this section and those that follow we try to present the counting process results in an intuitive and easily understood manner. An expanded introduction to the counting process approach is presented in Appendix 2. A complete development of the theory as well as applications to other settings may be found in Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993).

Assume that we follow a single subject with covariates denoted by $\mathbf{x}$ from time "zero" and that the event of interest is death. We could use as the outcome any other event that can occur only once or the first occurrence of an event such as drug use. The counting process represen-

tation of the proportional hazards model is a linear-like model that "counts" whether the event occurs (e.g., the subject dies) at time t. The basic model is

$$N(t) = \Lambda(t, \mathbf{x}, \boldsymbol{\beta}) + M(t), \tag{6.6}$$

where the function $N(t)$ is the "count" that represents the observed part of the model, the function $\Lambda(t, \mathbf{x}, \boldsymbol{\beta})$ is the "systematic component" of the model, and the function $M(t)$ is the "error component."

The function $N(t)$ is defined to be equal to zero until the exact time the event occurs and is equal to one thereafter. If the total length of follow-up is one year, and our subject dies on day 200, then

$$N(t) = \begin{cases} 0 \text{ for } t < 200, \\ 1 \text{ for } t \geq 200. \end{cases}$$

If the subject does not die during the one year of follow-up, then the count is always zero, $N(t) = 0$. Hence, the maximum value of the count function occurs at the end of follow-up of the subject and is equal to the value of the censoring indicator variable.

The systematic component of the model is, as we show in Appendix 2, equal to the cumulative hazard at time t under the proportional hazards model,

$$\Lambda(t, \mathbf{x}, \boldsymbol{\beta}) = H(t, \mathbf{x}, \boldsymbol{\beta}),$$

until follow-up ends on the subject and it is equal to zero thereafter. Thus, the value of the function for a subject who either dies or is censored on day 200 is

$$\Lambda(t, \mathbf{x}, \boldsymbol{\beta}) = \begin{cases} e^{\mathbf{x}'\boldsymbol{\beta}} H_0(t) & \text{for } t < 200, \\ e^{\mathbf{x}'\boldsymbol{\beta}} H_0(200) & \text{for } t \geq 200. \end{cases}$$

where $H_0(t)$ is the cumulative baseline hazard function. It follows that the maximum value for the systematic component also occurs at the end of follow-up, regardless of whether the event occurred. The function $M(t)$ in (6.6) is, under suitable mathematical assumptions, called a martingale and plays the role of the error component. It has many of the same properties that error components in other models have, in par-

ticular its mean is zero under the correct model. If we rearrange (6.6), $M(t)$ may be expressed in the form of a "residual" as

$$M(t) = N(t) - \Lambda(t, \mathbf{x}, \boldsymbol{\beta}).$$

(6.7)

The quantity in (6.7) is called the *martingale residual*. In theory it has a value at each time t, but the most useful choice of time at which to compute the residual is the end of follow-up, yielding a value for the ith subject of

$$
\begin{aligned}
M(t_i) &= N(t_i) - H(t_i, \mathbf{x}, \boldsymbol{\beta}) \\
&= c_i - H(t_i, \mathbf{x}, \boldsymbol{\beta}),
\end{aligned}
$$

(6.8)

since $\Lambda(t_i, \mathbf{x}, \boldsymbol{\beta}) = H(t_i, \mathbf{x}, \boldsymbol{\beta})$. For ease of notation, let $M_i = M(t_i)$. The estimator obtained by substituting the value of the partial likelihood estimator of the coefficients, $\hat{\boldsymbol{\beta}}$, is

$$\hat{M}_i = c_i - \hat{H}(t_i, \mathbf{x}, \hat{\boldsymbol{\beta}}).$$

(6.9)

The estimator $\hat{H}(t_i, \mathbf{x}, \hat{\boldsymbol{\beta}})$ is defined in (3.41). We used this residual, (6.9), in Chapter 5 for the graphical methods to assess the scale of a continuous covariate.

The residual in (6.9) has also been called the Cox–Snell or modified Cox–Snell residual, see Cox and Snell (1968) and Collett (1994). This terminology is due to the work of Cox and Snell, who showed that the values of $\hat{H}(t_i, \mathbf{x}, \hat{\boldsymbol{\beta}})$ may be thought of as observations from a censored sample with an exponential distribution and parameter equal to 1.0. Unfortunately, this distribution theory has not proven to be as useful for model evaluation as the theory derived from the counting process approach.

Using the counting process approach, the expressions in (6.7) and (6.8) are a completely natural way to define a residual. To see why it also makes sense to consider (6.8) as a residual in the proportional hazards regression model, assume for ease of notation that there are no ties and that the value of the baseline hazard at time t_i may be expressed as

$$h_0(t_i) = \frac{c_i}{\sum\limits_{j \in R(t)} e^{\mathbf{x}'_j \boldsymbol{\beta}}},$$

(6.10)

and the expression for the cumulative baseline hazard is

$$H_0(t_i) = \sum\limits_{t_j \le t_i} h_0(t_j).$$

(6.11)

The Breslow estimator of the cumulative baseline hazard is obtained from (6.11) by substituting the value of the partial likelihood estimator of the coefficients, $\hat{\boldsymbol{\beta}}$. Under these assumptions, the derivative in (6.1) may be expressed as

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum\limits_{i=1}^{n} x_{ik} \left[c_i - H(t_i, \mathbf{x}, \boldsymbol{\beta}) \right].$$

(6.12)

The expression in (6.12) is similar to the equations obtained for other models, such as linear and logistic regression, in that it expresses the partial derivative as a sum of the value of the covariate times an "observed minus expected" residual.

The p equations obtained by setting (6.1) equal to zero for each covariate are called the *score equations* and some authors, for example, Collett (1994), call the Schoenfeld residual in (6.3) the *score residual*. However, an entirely different residual of the same name is the one currently calculated by software packages. This residual is obtained by expressing the martingale residual representation shown in (6.12) in a slightly different form. The score equation for the kth covariate may be expressed as

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_k} = \sum\limits_{i=1}^{n} L_{ik}.$$

(6.13)

The expression for L_{ik} is somewhat complex. Readers who are willing to accept without further elaboration that the estimator of L_{ik} is the score *process* residual provided by software packages may skip the next paragraph where we describe L_{ik} in more detail.

The score process residual for the ith subject on the kth covariate in (6.13) may be expressed as

$$L_{ik} = \sum_{j=1}^{n} \left(x_{ik} - \bar{x}_{w_j k} \right) dM_i(t_j). \qquad (6.14)$$

The mean in the expression, $\bar{x}_{w_j k}$, is the value of (6.2) computed at t_j. The quantity $dM_i(t_j)$ is the change in the martingale residual for ith subject at time t_j and is

$$dM_i(t_j) = dN_i(t_j) - Y_i(t_j) e^{x_i' \beta} h_0(t_j). \qquad (6.15)$$

The first part of (6.15), $dN_i(t_j)$, is the change in the count function for the ith subject at time t_j. This will be always equal to zero for censored subjects. For noncensored subjects, it will be equal to zero except at the actual observed survival time, when it will be equal to one. That is, $dN_i(t_i) = 1$ for noncensored subjects. In the second part of (6.15), the function $Y_i(t_j)$ is called the *at risk process* and is defined as follows:

$$Y_i(t_j) = \begin{cases} 1 \text{ if } t_i \geq t_j \\ 0 \text{ if } t_i < t_j \end{cases}$$

and $h_0(t_j)$ is the value of (6.10) evaluated at t_j. An expanded computational formula yields the estimator

$$\hat{L}_{ik} = c_i \times \left(x_{ik} - \hat{\bar{x}}_{w_j k} \right) - x_{ik} \times \hat{H}\left(t_i, x, \hat{\beta}\right) + e^{x_i' \hat{\beta}} \sum_{t_j \leq t_i} \hat{\bar{x}}_{w_j k} \frac{c_j}{\sum_{l \in R_j} e^{x_l' \hat{\beta}}}. \qquad (6.16)$$

Let the vector of p score process residuals for the ith subject be denoted as

$$\hat{L}_i' = \left(\hat{L}_{i1}, \hat{L}_{i2}, ..., \hat{L}_{ip} \right). \qquad (6.17)$$

Before moving on, we provide a brief summary of residuals. The martingale residual, $\hat{M}_i$ in (6.9) has the form typically expected of a residual in that it resembles the difference between an observed outcome and a predicted outcome. The other three residuals (score process, Schoenfeld and scaled Schoenfeld) are covariate-specific. Every subject

has a value of the score process residual for the kth covariate, $\hat{L}_{ik}$ in (6.16), but the Schoenfeld residual in (6.3) and the scaled Schoenfeld residual in (6.5) are defined only at the observed survival times. Thus, there will be m subjects with values for these residuals. Each of these residuals provides a useful tool for examining one or more aspects of model adequacy. [Barlow and Prentice (1988) consider variations in the martingale residual obtained by including other functions of time in (6.8). As of this time, none of their generalized residuals have been added to software packages.]

We now consider methods for verifying the proportional hazards assumption.

6.3 METHODS FOR ASSESSING THE PROPORTIONAL HAZARDS ASSUMPTION

The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model, as discussed in detail in Chapter 4. Specifically, the proportional hazards model has a log-hazard function of the form

$$\ln[h(t, x, \beta)] = \ln[h_0(t)] + x' \beta. \qquad (6.18)$$

This function has two parts, the log of the baseline hazard function, $\ln[h_0(t)]$, and the linear predictor, $x' \beta$. Methods for building the linear predictor part of the model are discussed in detail in Chapter 5. The proportional hazards assumption characterizes the model as a function of time, not of the covariates per se. Assume for the moment that the model contains a single dichotomous covariate. A graph of the log-hazard, (6.18), over time would produce two curves, one for $x = 0$, $\ln[h_0(t)]$, and one for $x = 1$, $\ln[h_0(t)] + \beta$. Regardless of how simple or complicated the baseline hazard function is, the vertical distance between these two curves at any point in time is β. This fact is the reason that the hazard ratio, $\exp(\beta)$, has such a simple and useful interpretation.

As a second example, suppose age is the only covariate in the model and that it is scaled linearly. Consider the graphs of the log-hazard function for age a and age $a + 10$. If the coefficient, β, is positive, the upper to lower vertical distance between the two curves will be 10β at

every point in time. Assessing the proportional hazards assumption is an examination of the extent to which the two curves are equidistant over time.

There are, effectively, an infinite number of ways the model in (6.18) can be changed to yield non-proportional hazard functions or log-hazard functions that are not equidistant. As a result, a large number of tests and procedures have been proposed. However, recent developmental work by Grambsch and Therneau (1994) and simulation comparisons by Ng'andu (1997) have shown that one easily performed test and an associated graph yield a powerful and effective method for examining this critical assumption.

Grambsch and Therneau (1994) consider an alternative to the model in (6.18), originally proposed by Schoenfeld (1982), that has the following specific form of time-varying coefficient:

$$\beta_j(t) = \beta_j + \gamma_j g_j(t), \qquad (6.19)$$

where $g_j(t)$ is a specified function of time. The rationale behind this model is that the effect of a covariate may change over the period of follow-up. For example, the baseline value of a specific test may lose its relevance over time. The opposite could also occur, where a baseline measure is more predictive of survival later in follow-up. Under this model, Grambsch and Therneau show that the scaled Schoenfeld residuals in (6.4), and their approximation in (6.5), have, for the jth covariate, a mean at time t of approximately

$$E\left[r_j^*(t)\right] \cong \gamma_j g_j(t). \qquad (6.20)$$

The result in (6.20) suggests that a plot of the scaled Schoenfeld residuals over time may be used to visually assess whether the coefficient γ_j is equal to zero and, if not, what the nature of the time dependence, $g_j(t)$, may be. Grambsch and Therneau derive a generalized least squares estimator of the coefficients and a score test of the hypothesis that they are equal to zero, given specific choices for the functions $g_j(t)$. In addition, they show that specific choices for the function yield previously proposed tests. For example, use of $g(t) = \ln(t)$ yields a model first suggested by Cox (1972) and a test by Gill and Schumacher (1987) discussed by Chappell (1992). With this function, the model in (6.19) is

$$\beta_j(t) = \beta_j + \gamma_j \ln(t),$$

and the linear predictor portion of the model in (6.19) is

$$\beta_j x_j + \gamma_j x_j \ln(t). \qquad (6.21)$$

The form of the linear predictor in (6.21) suggests that another way to test the hypothesis that $\gamma_j = 0$ is via the partial likelihood ratio test, score test or Wald test obtained when the interaction $x_j \ln(t)$ is added to the proportional hazards model. The advantage of this approach over the generalized least squares score test proposed by Grambsch and Therneau is that it may be done using the model fitting software in many statistical software packages. One should note that when the interaction term, $x_j \ln(t)$, is included in the model the partial likelihood becomes much more complicated. Since the interaction is a function of time, its value must be recomputed for each term in the risk set at each observed survival time. The interaction term is not simply the product of the covariate and the subject's observed value of time.

Other functions of time have been suggested. Quantin, et al. (1996) propose using $g(t) = \ln[H_0(t)]$. Based on simulations reported in their paper, this test appears to have good power, but it is not as easy to compute as the test based on $g(t) = \ln(t)$. This is because the Breslow estimator of $H_0(t)$ must be computed and must be accessible at each observed survival time. Based on the simulations in Quantin et al. (1996) and Ng'andu (1997), the test with $g(t) = \ln(t)$ has power nearly as high as or higher than all other commonly used tests to detect reasonable alternatives to proportional hazards. For this reason, we consider only the model in (6.21). These same simulations show that the performance of the partial likelihood score test and the Grambsch and Therneau generalized least squares score test are essentially the same. Thus, we will use the more easily computed model-based forms of the test.

Before evaluating the fitted models from the UIS developed in Chapter 5 for proportional hazards, we consider the methods in some simpler models. In the case of models containing nominal scale covariates, a purely graphical bivariate assessment may be obtained from the plots of the "log-negative-log" of the within-group Kaplan-Meier estimator of the survivorship functions versus log-time. If the hazard functions are proportional, this plot should have parallel lines and the vertical distance between each line and that of the reference group

should be approximately equal to the coefficients from the fit of the proportional hazards model. One disadvantage of this graphical approach is that it is a univariate method that may be used only for nominal scale covariates or grouped continuous covariates. In addition it is difficult to visually assess whether the plotted lines deviate significantly from being parallel, particularly when sample sizes are small.

As an example of a multivariable model, we consider a situation in which the model contains one dichotomous covariate, denoted d, and one continuous covariate, denoted x. This data setting is used in several examples. In each example, two models were fit: the main effects model containing d and x and a second model obtained by including the interactions of each covariate with time, $d\ln(t)$ and $x\ln(t)$. For numerical reasons it is preferable to center log-time about its mean and use $\left[\ln(t) - \overline{\ln}(t)\right]$ in the interaction. As noted above, models containing interactions with log-time are easily fit in many statistical software packages.

The data for the first example illustrate a model in which the hazard is proportional in both covariates. The results of fitting the main effects model and the model with interactions with log-time are shown in Table 6.1. The main effects model in Table 6.1 shows that both covariates are highly significant. The p-values for the Wald statistics for both interaction coefficients are not significant, suggesting that the hazard function may be proportional in the two covariates. The value of the partial likelihood ratio test for the addition of the two interaction variables is $G = 0.134$ and, with 2 degrees-of-freedom, the p-value is 0.94. This is confirmed by the graphs in Figure 6.1. The scaled residuals scatter in a

Table 6.1 Estimated Coefficients, Standard Errors, z-Scores, Two-Tailed p-Values and 95% Confidence Intervals for Models with a Proportional Hazard Function in Both Covariates ($n = 100$ with 30% Censoring)

Variable	Coeff.	Std. Err.	z	P>|z|	95% CIE
d	0.579	0.249	2.33	0.020	0.092, 1.066
x	0.180	0.032	5.66	<0.001	0.118, 0.243
d	0.573	0.253	2.27	0.023	0.078, 1.068
x	0.186	0.035	5.26	<0.001	0.116, 0.255
$d\times\ln(t)$	−0.002	0.163	0.01	0.988	−0.322, 0.317
$x\times\ln(t)$	0.007	0.020	0.36	0.716	−0.032, 0.047

nonsystematic way about the zero line, and the polygon connecting the values of the smoothed residuals has approximately a zero slope and crosses the zero line several times. The initial upward trend in the smoothed residuals for the dichotomous covariate in Figure 6.1a is due to a few large negative residuals among the shortest survival times. One should also note that there are two bands of residuals for the dichotomous covariate. The upper band corresponds to subjects with $d = 1$ and the bottom one to those with $d = 0$. There are 70 points in the graph, as this is the observed number of survival times.

In a second example, the model is nonproportional in the continuous covariate. Table 6.2 presents the results of fitting the two models. The main effects model in Table 6.2 shows that both covariates are highly significant. The model with the interactions with log-time shows that the model may be non-proportional in the continuous covariate, as $p = 0.024$ for the Wald test for the $x \times \ln(t)$ coefficient. The value of the partial likelihood ratio test for the addition of the two interaction variables is $G = 5.424$ and, with 2 degrees-of-freedom, the p-value is 0.066. The interactions model shows some numeric instability, as the estimated standard error for the $d \times \ln(t)$ term is quite large. Based on this, two models (each containing only one of the log-time interactions) were fit, and the results supported the observation of non-proportionality in x but not d. Graphs of the scaled Schoenfeld residuals are shown in Figure 6.2.

Under the time-varying coefficient model in (6.21), if the covariate has a proportional hazard, the plot of the scaled Schoenfeld residuals and its smooth should show no trend over time. This is observed in Figure 6.2a, where it can be seen that the smoothed residuals have essentially a slope of zero. The apparent initial positive slope is due to one or two large negative residuals. On the other hand, the polygon based on the smoothed scaled residuals for the continuous covariate, shown in Figure 6.2b, displays a consistent positive slope, suggesting that the importance of the covariate increases over time and thus has a nonproportional hazard.

As a third example, data were created such that the hazard function was nonproportional for the dichotomous covariate and was proportional in the continuous covariate. Table 6.3 presents the results of fitting the main effects and log-time interactions models. Graphs of the scaled Schoenfeld residuals are presented in Figure 6.3.

The results in Table 6.3 show evidence of nonproportional hazards in the dichotomous covariate, as $p = 0.001$ for the Wald test of the coef-