GOODNESS-OF-FIT INFERENCE FOR THE COX–AALEN ADDITIVE-MULTIPLICATIVE REGRESSION MODEL

DAVID KRAUS

Department of Probability and Statistics, Charles University
Sokolovská 83, CZ-186 75 Prague, Czech Republic
kraus@karlin.mff.cuni.cz

Abstract. The Cox–Aalen additive-multiplicative intensity model of Scheike and Zhang (2002) is considered. We study goodness-of-fit tests based on the stratified martingale residual process. Asymptotic distribution of the process is derived and the Kolmogorov–Smirnov type test is constructed. Several ways of overcoming the problem of complexity of the limiting distribution are discussed. The results are accompanied by a small Monte Carlo study.

1. Introduction

In survival analysis, regression models are used to explain occurence of events (failures) in time by the influence of explanatory variables (covariates). Let $Z_i = \{(Z_{i1}(t), \dots, Z_{ip}(t))^\mathsf{T}, t \in [0, \tau]\}$ be a vector of covariates (possibly time-dependent, i.e. predictable stochastic processes) for the i-th observed individual, Y_i be an indicator process (indicating by its value at time t whether the i-th individual is at risk of the event) and λ_i be an intensity process of the corresponding counting process N_i. Then the most popular model for the intensity process is the Cox proportional hazards model of the form

$$\lambda_i(t) = Y_i(t) \exp\{\beta_0^\mathsf{T} Z_i(t)\}\lambda_0(t),$$

where λ_0 is an unknown baseline hazard function and β_0 is a p-vector of unknown regression parameters. Another frequently used model is the Aalen additive model

$$\lambda_i(t) = Y_i(t) X_i(t)^\mathsf{T} \alpha(t)$$

with a vector of unknown functions $\alpha = \{(\alpha_1(t), \dots, \alpha_q(t))^\mathsf{T}, t \in [0, \tau]\}$ and a vector of covariates X_i.

There are several ways of combining these two models. Lin and Ying (1995) suggested the intensity process of the form

$$\lambda_i(t) = Y_i(t)\big\{g(\gamma_0^\mathsf{T} X_i(t)) + h(\beta_0^\mathsf{T} Z_i(t))\lambda_0(t)\big\}$$

Date: September 23, 2004.
2000 *Mathematics Subject Classification.* 62N03.
Key words and phrases. Counting process, Cox–Aalen model, goodness-of-fit, hazard regression, martingale, residual, survival analysis.

with known functions g and h. Later Scheike and Martinussen (2002) followed this idea and studied a model with

$$\lambda_i(t) = Y_i(t)\big\{X_i(t)^\mathsf{T}\alpha(t) + \exp\{\beta_0^\mathsf{T} Z_i(t)\}\lambda_0(t)\big\},$$

which allows for time-varying effects. A different approach was used by Scheike and Zhang (2002). Their model, which is here called the Cox–Aalen model, follows the form

$$\lambda_i(t) = Y_i(t)\exp\{\beta_0^\mathsf{T} Z_i(t)\}X_i(t)^\mathsf{T}\alpha(t), \qquad 0 \le t \le \tau, \tag{1}$$

where $(X_i^\mathsf{T}, Z_i^\mathsf{T})^\mathsf{T}$ is a $(q+p)$-vector of predictable covariates (usually $X_{i1} \equiv 1$). Some components of $\alpha(t)$ or $X_i(t)$ can even be negative, provided the whole term $X_i(t)^\mathsf{T}\alpha(t)$ is nonnegative (an intensity always has to be nonnegative). Scheike and Zhang (2002) argue that this model is the first step Taylor approximation of the model

$$\lambda_i(t) = Y_i(t)\exp\{\beta_0^\mathsf{T} Z_i(t)\}\lambda(t, X_i(t)) \tag{2}$$

of Dabrowska (1997). The model (2) can be used when one wants to assess the influence of several covariates (treated proportionally) and wants to avoid possible incorrect conclusions caused by omission or inadequate specification of effects of the other covariates which are not in focus. However, the model (2) is too general and therefore its estimation is somewhat complicated. On the contrary, the model of Scheike and Zhang (2002) with the Aalen additive regression as the baseline is easier to estimate although it still preserves the flexible nonparametric form with time-varying effects.

Scheike and Zhang (2002) suggested an estimation procedure and derived asymptotic properties of the estimators. Here we propose a test of goodness of fit.

For the Cox model many goodness-of-fit tests were developed. Among others we can mention the tests on parameters, the tests of the proportional hazards assumption or the tests on the functional form of a covariate (a survey of these tests can be found in Therneau and Grambsch, 2000). Besides these tests focused on specific departures from the Cox model, there are some global tests. One is based on the doubly cumulative hazard function (McKeague and Utikal, 1991). Another approach is based on the martingale residual process. This approach is here adapted for the situation of the Cox–Aalen model.

The idea was originated by Arjas (1988) who suggested a graphical procedure for assessing goodness of fit of the Cox model. The method is based on comparison of observed and expected number of failures within a given stratum. For each observed individual $i \in \{1, \dots, n\}$ this difference is expressed by the process $\hat{M}_i = N_i - \hat{\Lambda}_i$. Hence, if a stratum $I \subset \{1, \dots, n\}$ is chosen, the process $\Xi_I = \sum_{i \in I} \hat{M}_i$ should fluctuate around zero. Let $0 \le T_{(I,1)} \le T_{(I,2)} \le \cdots \le T_{(I,K)} \le \tau$ be ordered times of the actual failures in I. Then Arjas's plots are plots of the values $\sum_{i \in I} \hat{\Lambda}_i(T_{(I,k)})$ against $k = \sum_{i \in I} N_i(T_{(I,k)})$. The graph should be close to the line with slope 1 when the fit is good, and should differ otherwise.

Asymptotic behaviour of the residual process was studied by Marzec and Marzec (1993). Under certain conditions they showed that the limiting distribution of $n^{-1/2}\Xi_I$ is that of a continuous zero-mean Gaussian process. Later, Marzec and Marzec (1997) presented some generalisations and used a transformation of the limiting process to a martingale which enables construction of the Kolmogorov–Smirnov type test. The idea of Arjas's plots and stratified residual processes was successfully used by Volf (1996) also for the Aalen additive model.

In the situation of the Cox–Aalen model of (1), the stratified martingale residual process has the form

$$\text{(3)}\quad \Xi_I(t) = \sum_{i\in I}(N_i(t) - \hat{\Lambda}_i(t)) = \sum_{i\in I}\int_0^t \big[dN_i(s) - Y_i(s)\exp\{\hat{\beta}^\mathsf{T} Z_i(s)\}X_i(s)^\mathsf{T} Y^-(\hat{\beta}, s)dN(s)\big]$$

(the notation is explained in Section 2).

The paper is organised as follows. In Section 2 we recapitulate basic facts about the model as stated in Scheike and Zhang (2002) and introduce the notation. Section 3 establishes asymptotic properties of the residual process. Section 4 is devoted to the study of several testing procedures. In Section 5 a simulation study is presented.

2. Notation, previous results and assumptions

First, we recall the notation and the results of Scheike and Zhang (2002). We work in the standard counting process framework: $N = (N_1, \ldots, N_n)^\mathsf{T}$ is a multivariate counting process with components having intensities λ_i of the form (1), $\Lambda = (\Lambda_1, \ldots, \Lambda_n)^\mathsf{T}$ (with $\Lambda_i = \int \lambda_i$) is its compensator and $M = (M_1, \ldots, M_n)^\mathsf{T}$ (with $M_i = N_i - \Lambda_i$) is a martingale. The counting process background is general enough to allow e.g. for recurrent events on the same individual (i.e. possibly more than one jump of the corresponding process N_i) and repeated entering and leaving the riskset (i.e. the indicator process Y_i can switch between the values 1 and 0 repeatedly). For more details on counting processes see Andersen, Borgan, Gill and Keiding (1993).

The cumulative hazard function $A(t) = \int_0^t \alpha(s)ds$ and the parameter β_0 need to be estimated. As suggested by Scheike and Zhang (2002), β_0 is estimated by $\hat{\beta}$ which is given as the solution to the score equation $U(\beta, \tau) = 0$, where $U(\beta, \cdot)$ is the score process of the form

$$\text{(4)}\qquad U(\beta, t) = \int_0^t \big[Z(s)^\mathsf{T} - Z(s)^\mathsf{T} Y(\beta, s)Y^-(\beta, s)\big]dN(s).$$

Here

$$Y(\beta, t) = \big(Y_1(t)\exp\{\beta^\mathsf{T} Z_1(t)\}X_1(t), \ldots, Y_n(t)\exp\{\beta^\mathsf{T} Z_n(t)\}X_n(t)\big)^\mathsf{T},$$

$$Z(t) = \big(Z_1(t), \ldots, Z_n(t)\big)^\mathsf{T}, \qquad Y^-(\beta, t) = [Y(\beta, t)^\mathsf{T} W(\beta, t)Y(\beta, t)]^{-1}Y(\beta, t)^\mathsf{T} W(\beta, t),$$

and the weight matrix $W(\beta, t) = \mathrm{diag}[w_1(\beta, t), \ldots, w_n(\beta, t)]$ is diagonal with elements of the form $w_i(\beta, t) = \exp\{-\beta^\mathsf{T} Z_i(t)\}/h_i(t)$, where h_i are some functions not depending on β. Note that for $\beta = \beta_0$ we have

$$U(\beta_0, t) = \int_0^t \big[Z(s)^\mathsf{T} - Z(s)^\mathsf{T} Y(\beta_0, s)Y^-(\beta_0, s)\big]dM(s).$$

The estimation procedure of Scheike and Zhang (2002) consists of two steps:

(1) Solve the score equations to obtain $\hat{\beta}$.

(2) Estimate $dA(t)$ by the weighted least squares principle by $d\hat{A}(t) = Y^{-}(\hat{\beta}, t)dN(t)$.

Initial estimates $\hat{\beta}$ and $d\hat{A}(t)$ are computed with the choice $h_i \equiv 1$. The maximum likelihood weights have $h_i(t) = X_i(t)^{\mathsf{T}}\alpha(t)$. They can be estimated by smoothing $d\hat{A}(t)$ to obtain $\hat{\alpha}(t)$ and setting $h_i(t) = X_i(t)^{\mathsf{T}}\hat{\alpha}(t)$. Now the steps (1) and (2) can be repeated to find final estimates of β_0 and $dA(t)$.

Let us use some further notation which comes from Scheike and Zhang (2002). They introduced

$$S^{(j)(k)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) Z_i(t)^{\otimes j} X_i(t)^{\otimes k} \exp\{\beta^{\mathsf{T}} Z_i(t)\}^k w_i^k(\beta, t) \exp\{\beta^{\mathsf{T}} Z_i(t)\} X_i(t)^{\mathsf{T}}, \quad j+k \le 2,$$

which is understood as a linear operator on q-vectors when the dimensions do not match for matrix multiplication (i.e. when $j + k = 2$). For $j = k = 1$ an additional transposition is needed, so let

$$S^{(1)(1)}(\beta, t) = \sum_{i=1}^{n} Y_i(t) w_i(\beta, t) \exp\{\beta^{\mathsf{T}} Z_i(t)\} X_i(t) Z_i(t)^{\mathsf{T}} \exp\{\beta^{\mathsf{T}} Z_i(t)\} X_i(t)^{\mathsf{T}}.$$

For $k = 0$, $S^{(j)}$ denotes $S^{(j)(0)}$. Let

$$(5) \quad \Sigma_U(\beta, t) = \int_0^t \big[Z(s)^{\mathsf{T}} - S^{(1)}(\beta, s) Y^{-}(\beta, s)\big] \operatorname{diag}[dN(s)] \big[Z(s)^{\mathsf{T}} - S^{(1)}(\beta, s) Y^{-}(\beta, s)\big]^{\mathsf{T}}.$$

Next denote

$$(6) \quad \Sigma_J(\beta, t) = -\frac{\partial}{\partial \beta} U(\beta, t) = \int_0^t S^{(2)}(\beta, s) Y^{-}(\beta, s) dN(s) - \int_0^t S^{(1)}(\beta, s)[S^{(0)(1)}(\beta, s)]^{-1} S^{(1)(1)}(\beta, s) Y^{-}(\beta, s) dN(s),$$

where the last equality is thanks to the fact

$$(7) \qquad \frac{\partial}{\partial \beta} Y^{-}(\beta, t) v = -[S^{(0)(1)}(\beta, t)]^{-1} S^{(1)(1)}(\beta, t) Y^{-}(\beta, t) v$$

for any n-vector v (which is shown in Scheike and Zhang, 2002). We can use Taylor's expansion of $U(\cdot, \tau)$ around β_0 to get the equality

$$(8) \qquad n^{-1/2} U(\beta_0, \tau) = n^{-1} \Sigma_J(\beta^*, \tau)\{n^{1/2}(\hat{\beta} - \beta_0)\},$$

which will be used later.

Scheike and Zhang (2002) found certain conditions under which they proved large sample properties of $\hat{\beta}$ and $\hat{A}$. These conditions are similar to those of Andersen and Gill (1982) for the Cox model. Particularly, the conditions guarantee positive definiteness of the matrix $\sigma_J(\beta_0, \tau)$ (defined later on) and uniform convergence in probability of $n^{-1} S^{(j)(k)}$ to some uniformly continuous bounded functions $s^{(j)(k)}$ on $\mathcal{B} \times [0, \tau]$ for some neighbourhood $\mathcal{B}$ of

β_0. Moreover, for simplicity, the covariates are assumed to be bounded (which yields the Lindeberg condition needed in the proofs).

Let us now consider a stratum I and define

$$S_I^{(j)(k)}(\beta,s)=\sum_{i\in I}Y_i(t)Z_i(t)^{\otimes j}X_i(t)^{\otimes k}\exp\{\beta^\mathsf{T}Z_i(t)\}^k w_i^k(t)\exp\{\beta^\mathsf{T}Z_i(t)\}X_i(t)^\mathsf{T}$$

(with the same convention for $S_I^{(1)(1)}$ as for $S^{(1)(1)}$). By $|I|$ we denote the number of elements of I. In addition to the conditions C1–C4 of Scheike and Zhang (2002) we assume

C5. $|I|/n\to r$ as $n\to\infty$ (for some $r\in(0,1)$),

C6. $|I|^{-1}S_I^{(j)(k)}$ and $s_I^{(j)(k)}$ satisfy the same convergence and regularity conditions C2, C3 as $n^{-1}S^{(j)(k)}$ and $s^{(j)(k)}$.

Let 1_I be an n-vector with the i-th element indicating whether the i-th observation belongs to the stratum I. Further, we need to denote

$$\begin{aligned}
B(\beta,t)&=\int_0^t\big[1_I^\mathsf{T}-S_I^{(0)}(\beta,s)Y^-(\beta,s)\big]\operatorname{diag}[dN(s)]\big[1_I^\mathsf{T}-S_I^{(0)}(\beta,s)Y^-(\beta,s)\big]^\mathsf{T},\\
C(\beta,t)&=\int_0^t\big[Z(s)^\mathsf{T}-S^{(1)}(\beta,s)Y^-(\beta,s)\big]\operatorname{diag}[dN(s)]\big[1_I^\mathsf{T}-S_I^{(0)}(\beta,s)Y^-(\beta,s)\big]^\mathsf{T},\\
D(\beta,t)&=\int_0^t S_I^{(1)}(\beta,s)Y^-(\beta,s)dN(s)\\
&\quad-\left\{\int_0^t S_I^{(0)}(\beta,s)[S^{(0)(1)}(\beta,s)]^{-1}S^{(1)(1)}(\beta,s)Y^-(\beta,s)dN(s)\right\}^\mathsf{T},\\
b(\beta,t)&=r\int_0^t s_I^{(0)}(\beta,s)\alpha(s)ds-2r^2\int_0^t s_I^{(0)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}s_I^{(0)(1)}(\beta,s)\alpha(s)ds\\
&\quad+r^2\int_0^t s_I^{(0)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}[s^{(0)(2)}(\beta,s)\alpha(s)][s^{(0)(1)}(\beta,s)]^{-1}s_I^{(0)}(\beta,s)^\mathsf{T}ds,\\
c(\beta,t)&=r\int_0^t s_I^{(1)}(\beta,s)\alpha(s)ds-r\int_0^t s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}s_I^{(0)(1)}(\beta,s)\alpha(s)ds\\
&\quad-r\int_0^t[s^{(1)(1)}(\beta,s)\alpha(s)]^\mathsf{T}[s^{(0)(1)}(\beta,s)]^{-1}s_I^{(0)}(\beta,s)^\mathsf{T}ds\\
&\quad+r\int_0^t s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}[s^{(0)(2)}(\beta,s)\alpha(s)][s^{(0)(1)}(\beta,s)]^{-1}s_I^{(0)}(\beta,s)^\mathsf{T}ds,\\
d(\beta,t)&=r\int_0^t s_I^{(1)}(\beta,s)\alpha(s)ds-r\left\{\int_0^t s_I^{(0)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}s^{(1)(1)}(\beta,s)\alpha(s)ds\right\}^\mathsf{T},\\
\sigma_U(\beta,t)&=\int_0^t s^{(2)}(\beta,s)\alpha(s)ds-\int_0^t s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}[s^{(1)(1)}(\beta,s)\alpha(s)]ds\\
&\quad-\int_0^t\big\{s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}[s^{(1)(1)}(\beta,s)\alpha(s)]\big\}^\mathsf{T}ds\\
&\quad+\int_0^t s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}[s^{(0)(2)}(\beta,s)\alpha(s)][s^{(0)(1)}(\beta,s)]^{-1}s^{(1)}(\beta,s)^\mathsf{T}ds,
\end{aligned}$$

$$\sigma_J(\beta,t) = \int_0^t s^{(2)}(\beta,s)\alpha(s)ds - \int_0^t s^{(1)}(\beta,s)[s^{(0)(1)}(\beta,s)]^{-1}s^{(1)(1)}(\beta,s)\alpha(s)ds.$$

3. Asymptotic properties of the martingale residual process

Theorem 1. *The residual process $n^{-1/2}\Xi_I$ of (3) converges weakly in $D([0,\tau])$ to a zero-mean continuous Gaussian process $\xi_I(\cdot) = \gamma_1(\cdot) - d(\beta_0,\cdot)^{\mathsf{T}}\sigma_J(\beta_0,\tau)^{-1}\gamma_2(\tau)$, where $\gamma = (\gamma_1,\gamma_2^{\mathsf{T}})^{\mathsf{T}}$ is a $(1+p)$-dimensional zero-mean continuous Gaussian martingale with covariances given by*

$$\operatorname{cov}(\gamma_1(u),\gamma_1(t)) = b(\beta_0,u\wedge t), \qquad \operatorname{cov}(\gamma_2(u),\gamma_2(t)) = \sigma_U(\beta_0,u\wedge t),$$
$$\operatorname{cov}(\gamma_2(u),\gamma_1(t)) = c(\beta_0,u\wedge t).$$

The functions $\sigma_J(\beta_0,t)$ and $\sigma_U(\beta_0,t)$ can be estimated uniformly consistently by $n^{-1}\Sigma_J(\hat\beta,t)$ and $n^{-1}\Sigma_U(\hat\beta,t)$ given by (6) and (5), respectively. Uniformly consistent estimates of $b(\beta_0,t)$, $c(\beta_0,t)$ and $d(\beta_0,t)$ are $n^{-1}B(\hat\beta,t)$, $n^{-1}C(\hat\beta,t)$ and $n^{-1}D(\hat\beta,t)$, respectively.

Proof. To prove the theorem, we first find a martingale representation of the residual process Ξ_I and then apply Rebolledo's central limit theorem (Andersen and Gill, 1982, Theorem I.2).

Rewrite

$$\Xi_I(t) = \sum_{i\in I} M_i(t) - \int_0^t S_I^{(0)}(\hat\beta,s)Y^-(\hat\beta,s)dN(s) + \int_0^t S_I^{(0)}(\beta_0,s)\alpha(s)ds. \tag{9}$$

By Taylor's expansion in the first integral on the right hand side around β_0 we obtain

$$\int_0^t S_I^{(0)}(\hat\beta,s)Y^-(\hat\beta,s)dN(s) = \int_0^t S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)dN(s) + \left[\int_0^t \frac{\partial}{\partial\beta}\{S_I^{(0)}(\beta,s)Y^-(\beta,s)dN(s)\}\big|_{\beta=\tilde\beta}\right]^{\mathsf{T}}(\hat\beta-\beta_0), \tag{10}$$

where $\tilde\beta$ is on the line segment between β_0 and $\hat\beta$. Using (7), we compute

$$\int_0^t \frac{\partial}{\partial\beta}\{S_I^{(0)}(\beta,s)Y^-(\beta,s)dN(s)\}\big|_{\beta=\tilde\beta} = \int_0^t S_I^{(1)}(\tilde\beta,s)Y^-(\tilde\beta,s)dN(s) - \left\{\int_0^t S_I^{(0)}(\tilde\beta,s)[S^{(0)(1)}(\tilde\beta,s)]^{-1}S^{(1)(1)}(\tilde\beta,s)Y^-(\tilde\beta,s)dN(s)\right\}^{\mathsf{T}} = D(\tilde\beta,t).$$

Further note that $\alpha(s) = Y^-(\beta_0,s)\lambda(s)$. Therefore the first term in (10) and the last term in (9) collect together into an integral w.r.t. M. Put

$$\Gamma_1(t) = \int_0^t \left[1_I^{\mathsf{T}} - S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)\right]dM(s).$$

We can finally write

$$n^{-1/2}\Xi_I(t) = n^{-1/2}\Gamma_1(t) - \{n^{-1}D(\tilde\beta,t)\}^{\mathsf{T}}\{n^{1/2}(\hat\beta-\beta_0)\}. \tag{11}$$

Moreover, if the matrix $\Sigma_J(\beta^*,\tau)$ is invertible (which is, however, not necessary for the proof), then, denoting $\Gamma_2 = U(\beta_0,\cdot)$ and using (8), we get

$$n^{-1/2}\Xi_I(t) = n^{-1/2}\Gamma_1(t) - \{n^{-1}D(\tilde{\beta},t)\}^\mathsf{T}\{n\Sigma_J(\beta^*,\tau)^{-1}\}\{n^{-1/2}\Gamma_2(\tau)\}. \tag{12}$$

We need to establish the joint weak convergence of the martingale $n^{-1/2}\Gamma = n^{-1/2}(\Gamma_1,\Gamma_2^\mathsf{T})^\mathsf{T}$ to the zero-mean continuous Gaussian martingale γ of the theorem.

One can compute

$$\begin{aligned}
\langle n^{-1/2}\Gamma_1\rangle(t) &= n^{-1}\int_0^t \left[1_I^\mathsf{T} - S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)\right]\operatorname{diag}[\lambda(s)]\left[1_I^\mathsf{T} - S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)\right]^\mathsf{T} ds \\
&= n^{-1}\int_0^t S_I^{(0)}(\beta_0,s)\alpha(s)ds - 2n^{-1}\int_0^t S_I^{(0)}(\beta_0,s)[S^{(0)(1)}(\beta_0,s)]^{-1}S_I^{(0)(1)}(\beta_0,s)\alpha(s)ds \\
&+ n^{-1}\int_0^t S_I^{(0)}(\beta_0,s)[S^{(0)(1)}(\beta_0,s)]^{-1}[S^{(0)(2)}(\beta_0,s)\alpha(s)][S^{(0)(1)}(\beta_0,s)]^{-1}S_I^{(0)}(\beta_0,s)^\mathsf{T} ds \\
&\xrightarrow[n\to\infty]{\Pr} b(\beta_0,t).
\end{aligned}$$

Further, due to Scheike and Zhang (2002) we have

$$\begin{aligned}
\langle n^{-1/2}\Gamma_2\rangle(t) &= n^{-1}\int_0^t \left[Z(s)^\mathsf{T} - S^{(1)}(\beta_0,s)Y^-(\beta_0,s)\right]\operatorname{diag}[\lambda(s)] \\
&\times \left[Z(s)^\mathsf{T} - S^{(1)}(\beta_0,s)Y^-(\beta_0,s)\right]^\mathsf{T} ds \xrightarrow[n\to\infty]{\Pr} \sigma_U(\beta_0,t).
\end{aligned}$$

Similarly

$$\begin{aligned}
\langle n^{-1/2}\Gamma_2, n^{-1/2}\Gamma_1\rangle(t) &= n^{-1}\int_0^t \left[Z(s)^\mathsf{T} - S^{(1)}(\beta_0,s)Y^-(\beta_0,s)\right]\operatorname{diag}[\lambda(s)] \\
&\times \left[1_I^\mathsf{T} - S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)\right]^\mathsf{T} ds = n^{-1}\int_0^t S_I^{(1)}(\beta_0,s)\alpha(s)ds \\
&- n^{-1}\int_0^t S^{(1)}(\beta_0,s)[S^{(0)(1)}(\beta_0,s)]^{-1}S_I^{(0)(1)}(\beta_0,s)\alpha(s)ds \\
&- n^{-1}\int_0^t [S^{(1)(1)}(\beta_0,s)\alpha(s)]^\mathsf{T}[S^{(0)(1)}(\beta_0,s)]^{-1}S_I^{(0)}(\beta_0,s)^\mathsf{T} ds \\
&+ n^{-1}\int_0^t S^{(1)}(\beta_0,s)[S^{(0)(1)}(\beta_0,s)]^{-1}[S^{(0)(2)}(\beta_0,s)\alpha(s)] \\
&\times [S^{(0)(1)}(\beta_0,s)]^{-1}S_I^{(0)}(\beta_0,s)^\mathsf{T} ds \xrightarrow[n\to\infty]{\Pr} c(\beta_0,t).
\end{aligned}$$

Therefore the only thing that remains to be verified to establish the convergence $n^{-1/2}\Gamma \xrightarrow{\mathcal{D}} \gamma$ is the Lindeberg condition. This condition is satisfied thanks to the assumption of boundedness of the covariates.

Eventually, for $\bar{\beta} \xrightarrow{\Pr} \beta_0$ it can be shown by using Lenglart's inequality that $n^{-1}\Sigma_J(\bar{\beta},t) \to \sigma_J(\beta_0,t)$ and $n^{-1}D(\bar{\beta},t) \to d(\beta_0,t)$ uniformly (in $t \in [0,\tau]$) in probability. Hence the proposed weak convergence $n^{-1/2}\Xi_I \to \xi_I$ follows by the continuous mapping theorem. The consistency results for b, c, σ_U can be derived in a similar fashion as for d and σ_J. $\square$

Note that $S^{(j)(k)}(\beta,t)\alpha(t) = S^{(j)(k-1)}(\beta,t)$ and $S_I^{(j)(k)}(\beta,t)\alpha(t) = S_I^{(j)(k-1)}(\beta,t)$ for the maximum likelihood weights. If the ML weights are estimated, then $s^{(j)(k)}(\beta,t)\alpha(t) = s^{(j)(k-1)}(\beta,t)$ and $s_I^{(j)(k)}(\beta,t)\alpha(t) = s_I^{(j)(k-1)}(\beta,t)$. Hence the variances and covariances in Theorem 1 simplify, since $c = d$ and $\sigma_J = \sigma_U$.

4. Testing procedures

A graphical technique, an analogue of the residual plots of Arjas (1988), has already been explained in Section 1. In this section we describe some visual and mainly numerical methods of investigation of the martingale residual process. A direct use of the asymptotic distribution is not possible. This is caused by its complexity, since Theorem 1 states that ξ_I, the weak limit of $n^{-1/2}\Xi_I$, is a continuous zero-mean Gaussian process with the covariance function

$$\begin{aligned}\operatorname{cov}(\xi_I(t),\xi_I(u)) = b(\beta_0,t) - d(\beta_0,t)^{\mathsf{T}}\sigma_J(\beta_0,\tau)^{-1}c(\beta_0,u) - d(\beta_0,u)^{\mathsf{T}}\sigma_J(\beta_0,\tau)^{-1}c(\beta_0,t)\\ + d(\beta_0,t)^{\mathsf{T}}\sigma_J(\beta_0,\tau)^{-1}\sigma_U(\beta_0,\tau)\sigma_J(\beta_0,\tau)^{-1}c(\beta_0,u)\end{aligned}$$

for $0 \le t \le u \le \tau$. Thus it is seen that ξ_I is distributed neither as a martingale nor as a process with a known distribution. Therefore we cannot straightforwardly use for instance the Kolmogorov–Smirnov type test.

We show how the distribution of ξ_I reduces in a special case and then we describe how some approximations and transformations can be applied in the general situation.

4.1. **The case of one dichotomous covariate in the Cox part.** Let us consider the case of a one-dimensional covariate in the Cox part of the model (i.e. $p = 1$). Suppose that this covariate is time independent with only two possible values (z_0, z_1) and put $I = \{i : Z_i = z_1\}$. In the framework of the Cox model this special situation was studied by Marzec and Marzec (1993) (see also Wei, 1984). They showed that the residual process weakly converges to a time-transformed Brownian bridge. We derive a similar result for the Cox–Aalen model here.

In this special setup we observe that

$$s_I^{(j)(k)} = z_1^j s_I^{(0)(k)}, \quad s^{(1)(k)} = r(z_1 - z_0)s_I^{(0)(k)} + z_0 s^{(0)(k)}, \quad s^{(2)} = r(z_1^2 - z_0^2)s_I^{(0)} + z_0^2 s^{(0)}. \tag{13}$$

In this subsection, let us assume the use of the maximum likelihood weights (or their estimates). Recall that with these weights we have $s^{(j)(k)}\alpha = s^{(j)(k-1)}$ as well as $s_I^{(j)(k)}\alpha = s_I^{(j)(k-1)}$ and $\sigma_J = \sigma_U$. Using this fact and inserting from (13) into the expressions of the functions c, d, σ_J, we conclude after some computations that

$$c(\beta,t) = (z_1 - z_0)b(\beta,t), \quad d(\beta,t) = (z_1 - z_0)b(\beta,t), \quad \sigma_J(\beta,t) = (z_1 - z_0)^2 b(\beta,t).$$

Hence it is immediately seen that the covariance function of the limiting process ξ_I is

$$\operatorname{cov}(\xi_I(t),\xi_I(u)) = b(\beta_0,t)[1 - b(\beta_0,u)/b(\beta_0,\tau)], \qquad 0 \le t \le u \le \tau.$$

Thus

$$\xi_I(\cdot) \stackrel{\mathcal{D}}{=} [b(\beta_0,\tau)]^{1/2} W^0(b(\beta_0,\cdot)/b(\beta_0,\tau)),$$

where $W^0 = \{W^0(t) : 0 \le t \le 1\}$ is the Brownian bridge.

We can construct the Kolmogorov–Smirnov type test based on the test statistic $\sup_{0\le t\le\tau} |\Xi_I(t)|/[B(\hat\beta,\tau)]^{1/2}$, which is asymptotically distributed as the well-known random variable $\sup_{0\le t\le 1} |W^0(t)|$. A significantly large value of the statistic leads to the rejection of the hypothesis that the data come from the model (1).

4.2. **The simulation approximation.** Return now to the general situation. We shall describe how to approximate the asymptotic distribution of the martingale residual process through simulations. The simulations can be performed to obtain a sample from the limiting distribution, and hence to assess both graphically and numerically how unusual the observed residual process is. Our approach is based on the idea of Lin, Wei and Ying (1993).

In (12) we have found the martingale representation of the residual process of the form

$$\Xi_I(t) = \int_0^t \big[1_I^\mathsf{T} - S_I^{(0)}(\beta_0,s)Y^-(\beta_0,s)\big]\,dM(s) \\ - D(\tilde\beta,t)\Sigma_J(\beta^*,\tau)^{-1}\int_0^\tau \big[Z(s)^\mathsf{T} - Z(s)^\mathsf{T} Y(\beta_0,s)Y^-(\beta_0,s)\big]\,dM(s).$$

The limiting distribution can be approximated by plugging in the consistent estimate $\hat\beta$ in place of $\beta_0,\tilde\beta,\beta^*$, and by replacing the martingale increments $dM_i(t)$ by their simulated values. For the martingales M_i, $i = 1,\ldots,n$ it holds that $\mathsf{E}\,M_i(t) = 0$ and $\operatorname{var} M_i(t) = \mathsf{E}[M_i(t)^2] = \mathsf{E}\,\Lambda_i(t) = \mathsf{E}\,N_i(t)$. Therefore Lin et al. (1993) suggested to approximate M_i by $\tilde M_i = G_i N_i$ (i.e. $\tilde M = \operatorname{diag}[G]N$), where $G = (G_1,\ldots,G_n)$ is a random sample of standard normal variables independent of the data. Finally, we obtain the approximation

$$\tilde\Xi_I(t) = \int_0^t \big[1_I^\mathsf{T} - S_I^{(0)}(\hat\beta,s)Y^-(\hat\beta,s)\big]\operatorname{diag}[G]dN(s) \\ - D(\hat\beta,t)\Sigma_J(\hat\beta,\tau)^{-1}\int_0^\tau \big[Z(s)^\mathsf{T} - Z(s)^\mathsf{T} Y(\hat\beta,s)Y^-(\hat\beta,s)\big]\operatorname{diag}[G]dN(s).$$

It can be shown (by means similar to that of Lin et al., 1993) that the asymptotic distribution of $n^{-1/2}\tilde\Xi_I$ is equal to that of $n^{-1/2}\Xi_I$, i.e. the distribution of ξ_I. Thus, generating repeatedly a standard normal sample G and computing $\tilde\Xi_I$, we obtain a sample from the desired limiting distribution. We can visually assess goodness-of-fit by plotting Ξ_I together with an appropriate number of simulated $\tilde\Xi_I$. To test the hypothesis numerically we generate an adequately large number of realisations of $\tilde\Xi_I$ and estimate critical values or the p-value.

Note that the simulation of the asymptotic distribution is conditional on the data and therefore we do not obtain universal critical values of this distribution. In other words, the test is not distribution-free and we have to carry out the simulations for each particular data set separately. This can result in higher computational demands.

4.3. **The transformation method.** Another way of overcoming the problem with complexity of the asymptotic distribution of the residual process is based on the transformation idea of Khmaladze (1981). In the framework of testing whether a random variable follows a parametric form of the distribution, he suggested a transformation of empirical processes with plugged-in estimated parameters in order to obtain a well-known asymptotic distribution which does not depend on the distribution of the data. The idea was then used by Andersen, Borgan, Gill and Keiding (1993, VI.3.3.4) for testing goodness of fit of parametric models for intensities and by Marzec and Marzec (1997) for assessment of the Cox model.

Here we will find the compensator of ξ_I, say $\bar{\xi}_I$, and use the empirical counterpart Ψ_I of the martingale $\psi_I = \xi_I - \bar{\xi}_I$ as a basis for testing. The process ψ_I should be a martingale with respect to the filtration generated by ξ_I, i.e. with respect to $\mathcal{G}_t = \sigma\{\gamma(s), s \le t; \gamma_2(\tau)\}$, $t \in [0,\tau]$. As $d(\beta_0,\cdot)^\mathsf{T}\sigma_J(\beta_0,\tau)^{-1}\gamma_2(\tau)$ is measurable w.r.t. $\mathcal{G}_0$, we only need to compensate γ_1. The compensator of γ_1, say $\bar{\gamma}_1$, can be derived rather heuristically as follows (cf. Andersen et al., 1993, VI.3.3.4, pp. 464–466). Since the process γ is a Gaussian martingale, it has independent increments and $(d\gamma_1(t), \gamma_2(\tau)-\gamma_2(t))^\mathsf{T}$ is jointly normally distributed. Therefore

$$\begin{aligned} \mathsf{E}[d\gamma_1(t)|\gamma(s), s \le t; \gamma_2(\tau)] &= \mathsf{E}[d\gamma_1(t)|\gamma_2(\tau)-\gamma_2(t)] \\ &= \operatorname{cov}\{d\gamma_1(t), \gamma_2(\tau)-\gamma_2(t)\}[\operatorname{var}\{\gamma_2(\tau)-\gamma_2(t)\}]^{-1}[\gamma_2(\tau)-\gamma_2(t)] \\ &= c(\beta_0,dt)^\mathsf{T}\theta_U(\beta_0,t)^{-1}[\gamma_2(\tau)-\gamma_2(t)], \end{aligned}$$

where $\theta_U(\beta,t) = \sigma_U(\beta,\tau) - \sigma_U(\beta,t)$. Hence a natural candidate for the compensator of γ_1 is

$$\bar{\gamma}_1(t) = \int_0^t \mathsf{E}[d\gamma_1(s)|\gamma(u, u \le s; \gamma_2(\tau)] = \int_0^t [\gamma_2(\tau)-\gamma_2(s)]^\mathsf{T}\theta_U(\beta_0,s)^{-1}c(\beta_0,ds), \quad t \in [0,\tau].$$

The compensated residual process is then

$$\psi_I(t) = \gamma_1(t) - \int_0^t [\gamma_2(\tau)-\gamma_2(s)]^\mathsf{T}\theta_U(\beta_0,s)^{-1}c(\beta_0,ds).$$

Formal verification of the fact that ψ_I is actually a Gaussian martingale (with variance function $b(\beta_0,\cdot)$) is analogous to the proof of Lemma 3.2 of Marzec and Marzec (1997). Finally, the empirical counterpart of ψ_I can be

$$\Psi_I(t) = \Xi_I(t) - \int_0^t [U(\hat{\beta},\tau) - U(\hat{\beta},s)]^\mathsf{T}\Theta_U(\hat{\beta},s)^{-1}C(\hat{\beta},ds) \tag{14}$$

with $\Theta_U(\beta,t) = \Sigma_U(\beta,\tau) - \Sigma_U(\beta,t)$.

Now we shall investigate when $n^{-1/2}\Psi_I$ really converges to ψ_I. Let us confine ourselves to the interval $[0,\tau-\delta]$ (for a small $\delta > 0$). On this interval the weak convergence can be studied easily, whereas on the whole interval $[0,\tau]$ it becomes difficult because of the matrix inversion (since $\Theta_U(\hat{\beta},s)$ is close to zero for s close to τ). Some additional mild conditions to guarantee the convergence on $[0,\tau]$ were found by Marzec and Marzec (1997) in the case of the Cox model.

By Taylor's expansion of (14) around β_0, we obtain

$$(15)\quad n^{-1/2}\Psi_I(t) = n^{-1/2}\Gamma_1(t) - n^{-1/2}\int_0^t [U(\beta_0,\tau) - U(\beta_0,s)]^\mathsf{T}\Theta_U(\hat\beta,s)^{-1}C(\hat\beta,ds)$$
$$- n^{1/2}(\hat\beta-\beta_0)^\mathsf{T}\left[n^{-1}D(\tilde\beta,t) - n^{-1}\int_0^t \Theta_J(\beta^*,s)\Theta_U(\hat\beta,s)^{-1}C(\hat\beta,ds)\right],$$

where $\Theta_J(\beta,t) = \Sigma_J(\beta,\tau) - \Sigma_J(\beta,t)$. Now the limiting process is equal to ψ_I, if the last term in (15) vanishes asymptotically. This happens when $c = d$ and $\sigma_U = \sigma_I$, which is satisfied when we use the ML weights, as mentioned in Section 3. However, the efficient iterative estimation procedure requires solving the score equations repeatedly. Since this is the most time consuming part of the computation, it would be preferable to avoid this iteration. It is not necessary to perform this iterative estimation in order to achieve $c = d$ and $\sigma_U = \sigma_I$. It suffices to use consistent estimates of the ML weights both in the score equation and in the LS estimator. Therefore we suggest another estimation procedure:

(1) Estimate dA by the nonweighted least squares as if the model were Aalen's model (without the Cox part covariates). Smooth this estimate $d\tilde{A}$ to obtain $\tilde\alpha$.
(2) Set $w_i(\beta,t) = \exp\{-\beta^\mathsf{T}Z_i(t)\}/(X_i(t)^\mathsf{T}\tilde\alpha(t))$ and find $\hat\beta$ by solving the score equations.
(3) Obtain final estimates $d\hat{A}$ of dA by the weighted least squares principle with $w_i(\hat\beta,t) = \exp\{-\hat\beta^\mathsf{T}Z_i(t)\}/(X_i(t)^\mathsf{T}\tilde\alpha(t))$.

Nevertheless, our computational experience shows that even if we find only the initial estimates $\hat\beta$ and $d\hat{A}$ as described in Section 2 (i.e. we do not perform either the efficient estimation or the above suggested procedure), the results are still quite good (almost identical). This corresponds to the fact that the initial estimates and the efficiently weighted ones do not differ notably (see Scheike and Zhang (2002) and references therein).

Finally, the Kolmogorov–Smirnov type test statistic $\sup|\Psi_I(t)|/\{B(\hat\beta,\tau)\}^{1/2}$ is asymptotically distributed as the variable $\sup|W(t)|$ (where W denotes the Brownian motion).

5. Simulations

We performed a small simulation study in order to investigate performance of the proposed tests. We generated survival data following various models with various censoring patterns and estimated the Cox–Aalen model of the form

$$\lambda_i(t) = \{\alpha_1(t) + \alpha_2(t)X_i\}\exp\{\beta_1 Z_i\}.$$

Under the null hypothesis H_0 the samples came from the distribution with the intensity

$$\lambda_i(t) = \{0.5 + 0.2tX_i\}\exp\{0.5Z_i\},$$

where X_i was uniformly distributed on $[0,1]$ and Z_i had the standard normal distribution. Then two alternatives with additional covariates were considered: H_1 with

$$\lambda_i(t) = \{0.5 + 0.2tX_i + 0.7tX_i^*\}\exp\{0.5Z_i\},$$

where X_i^* had the alternative distribution on $\{0,1\}$ with probability 0.5, and H_2 having

$$\lambda_i(t) = \{0.5 + 0.2tX_i\}\exp\{0.5Z_i + 0.7Z_i^*\}$$

TABLE 1. Empirical sizes of the two tests on the nominal level of 0.05

		$I = \{i : X_i > 0.5\}$		$I = \{i : Z_i > 0\}$	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$
Without	Simulation	0.043	0.043	0.051	0.060
censoring	Transformation	0.067	0.054	0.073	0.047
Censoring	Simulation	0.044	0.049	0.052	0.047
U$[0, 5]$ (31 %)	Transformation	0.054	0.064	0.053	0.058
Censoring	Simulation	0.056	0.071	0.045	0.044
U$[0, 2.5]$ (51 %)	Transformation	0.041	0.056	0.013	0.009

TABLE 2. Empirical powers of the two tests on the nominal level of 0.05

		H_1		H_2	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$
Without	Simulation	0.726	0.979	0.889	0.995
censoring	Transformation	0.739	0.972	0.887	0.994
Censoring U$[0, 5]$	Simulation	0.553	0.879	0.794	0.980
(H_1 25 %, H_2 24 %)	Transformation	0.553	0.871	0.777	0.971
Censoring U$[0, 2.5]$	Simulation	0.313	0.633	0.664	0.955
(H_1 45 %, H_2 42 %)	Transformation	0.304	0.621	0.661	0.945

with Z_i^* having the same distribution as X_i^* in the previous situation. The covariates were generated independently. Three censoring schemes were considered: no censoring, moderate censoring (with censoring times having the uniform distribution on $[0, 5]$) and heavy censoring (with uniform censoring times on $[0, 2.5]$). The censoring times were mutually independent and independent of the survival times and of the covariates. The corresponding censoring rates are indicated in the tables. The sample sizes were $n = 100$ and 200. For the null hypothesis, stratification with respect to both covariates was studied, i.e. the stratum was first $I = \{i : X_i > 0.5\}$ and then $I = \{i : Z_i > 0\}$. Under the alternatives, the data were stratified with respect to the missing covariate: $I = \{i : X_i^* = 1\}$ under H_1, and $I = \{i : Z_i^* = 1\}$ under H_2.

Under the null hypothesis as well as under the alternatives, we generated the sample, estimated the model and tested goodness of fit. Two tests were performed: the test of Subsection 4.2 based on the simulation approximation (with 2000 realisations of the residual process) and the test of Subsection 4.3 based on the transformation. This was repeated 1000 times and empirical levels and powers of the tests on the nominal level of 0.05 were computed. Since the estimation of the model is highly time-consuming, we were able to carry out only 1000 repetitions in each situation, so our results give only a broad image of behaviour of the tests.

Table 1 reports the sizes of the tests in the above mentioned situations under H_0. It is seen that the tests maintain approximately their nominal significance levels which is not seriously affected by censoring. Table 2 confirms that the tests have good power against the alternatives H_1 and H_2 of missing covariates. When censoring is present, the power decreases. There is no important difference between the two versions (simulation and transformation) of the test.

Acknowledgement

The author is grateful to Petr Volf (Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic) for his helpful comments and suggestions.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 10, 1100–1120.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *J. Amer. Statist. Assoc.*, 83, 204–212.

Dabrowska, D. M. (1997). Smoothed Cox regression. *Ann. Statist.*, 25, 1510–1540.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Teor. Veroyatnost. i Primenen.*, 26, 246–265. In Russian. English translation in *Theory Probab. Appl.*, 26, 240–257.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80, 557–572.

Lin, D. Y. and Ying, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Ann. Statist.*, 23, 1712–1734.

Marzec, L. and Marzec, P. (1993). Goodness of fit inference based on stratification in Cox's regression model. *Scand. J. Statist.*, 20, 227–238.

Marzec, L. and Marzec, P. (1997). Generalized martingale-residual processes for goodness-of-fit inference in Cox's type regression models. *Ann. Statist.*, 25, 683–714.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Statist.*, 18, 177–195.

Scheike, T. H. and Martinussen, T. (2002). A flexible additive multiplicative hazard model. *Biometrika*, 89, 283–298.

Scheike, T. H. and Zhang, M.-J. (2002). An additive-multiplicative Cox–Aalen regression model. *Scand. J. Statist.*, 29, 75–88.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer, New York.

Volf, P. (1996). Analysis of generalized residuals in hazard regression models. *Kybernetika*, 32, 501–510.

Wei, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc.*, 79, 649–652.