Kybernetika

Petr Novák

Goodness-of-fit test for the accelerated failure time model based on martingale residuals

Kybernetika, Vol. 49 (2013), No. 1, 40--59

Persistent URL: http://dml.cz/dmlcz/143239

KYBERNETIKA — VOLUME *49* (2013), NUMBER 1, PAGES 40–59

GOODNESS-OF-FIT TEST FOR THE ACCELERATED FAILURE TIME MODEL BASED ON MARTINGALE RESIDUALS

Petr Novák

The Accelerated Failure Time model presents a way to easily describe survival regression data. It is assumed that each observed unit ages internally faster or slower, depending on the covariate values. To use the model properly, we want to check if observed data fit the model assumptions. In present work we introduce a goodness-of-fit testing procedure based on modern martingale theory. On simulated data we study empirical properties of the test for various situations.

Keywords: accelerated failure time model, survival analysis, goodness-of-fit

Classification: 62N01, 62N03

1. INTRODUCTION

Let us observe survival data representing time which passes from beginning of an experiment until some pre-defined failure. We suppose that the data may be incomplete in a way that some objects may be removed from the observation prior to reaching the failure, which we call right censoring. We want to model the dependence of the time to failure on available covariates. The Accelerated Failure Time model (AFT, see [2]) presents an alternative to the most widely used and well described Cox proportional hazard model (see [3]). In the AFT model, we assume the log-linear dependence

$$\log T_i^* = -Z_i^T \beta_0 + \epsilon_i,$$

where T_i^*, $i = 1, \ldots, n$, are the real failure times, $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ covariates, β_0 the vector of real parameters and ϵ_i are *iid* random variables with an unknown distribution. Denote C_i the censoring times, $T_i = \min(T_i^*, C_i)$ the times of the end of observation and $\Delta_i = I(T_i^* \leq C_i)$ noncensoring indicators. Suppose that T_i^* and C_i are independent for all i given Z_i, and that the censoring distribution does not depend on the regression parameters. We observe independent data (T_i, Δ_i, Z_i), $i = 1, \ldots, n$.

We assume T_i^* to have a continuous distribution. Denote $F_i(t) = P(T_i^* \leq t)$ their distribution function, $f_i(t)$ density, $S_i(t) = 1 - F_i(t)$ the survival function, $\alpha_i(t) = \lim_{h\searrow 0} P(t \leq T_i^* < t + h | T_i^* \geq t)/h = f_i(t)/S_i(t)$ the hazard function and $A_i(t) =$

$\int_0^t \alpha_i(s)\,\mathrm{d}s$ the cumulative hazard. For the AFT model, we have

$$\alpha_i(t) = \alpha_0(\exp(Z_i^T\beta_0)t)\exp(Z_i^T\beta_0).$$

The baseline hazard $\alpha_0(t)$ is the hazard rate of $\exp(\epsilon_i)$, is completely unknown and will be estimated nonparametrically.

The data may be represented as counting processes, denote $N_i(t) = I(T_i \le t, \Delta_i = 1)$, $Y_i(t) = I(t \le T_i)$, intensities $\lambda_i(t) = Y_i(t)\alpha_i(t)$ and cumulative intensities $\Lambda_i(t) = \int_0^t \lambda_i(s)\,\mathrm{d}s$. All functions and processes are studied on an interval $t \in [0,\tau]$, where $\tau < \infty$ is some point beyond the last observed survival time. It can be shown, that under the model assumptions, $\Lambda_i(t)$ are the compensators of corresponding processes $N_i(t)$ with respect to $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), \boldsymbol{Z}_i, 0 \le s \le t, i = 1,\dots,n\}$ (see [5]). Therefore $M_i(t) := N_i(t) - \Lambda_i(t)$ are $\mathcal{F}_t$-martingales (Doob-Meyer decomposition). The log-likelihood for the data can be then rewritten with the help of the counting processes as

$$l(t) = \sum_{i=1}^n \int_0^t \left(\log(\alpha_i(s))\,\mathrm{d}N_i(s) - Y_i(s)\alpha_i(s)\,\mathrm{d}s\right),$$

and by taking the derivative with respect to model parameters we get the score process $U(t,\boldsymbol{\beta})$. For estimation of the parameters we solve the equations $U(\boldsymbol{\beta}) \equiv U(\tau,\boldsymbol{\beta}) = \mathbf{0}$.

To obtain reliable estimates, the model assumptions must be met. However, the data can deviate from the model, for example if the dependence is different than log-linear or if we neglected one or more covariates. To check if the model holds, one must consider some goodness-of-fit testing procedure. In section 2, we present a goodness-of-fit statistic for the AFT model based on martingale approach and resampling techniques.

The model can also be generalized to accommodate time-varying covariates. In section 3 we explore the approach proposed by Cox and Oakes [4] and further studied by Lin and Ying [9] and we present a generalization of the goodness-of-fit test statistic. On simulated examples we study the empirical properties of the test in various situations for both time-invariant and time-dependent covariates in section 4.

2. THE TEST STATISTIC – TIME-INVARIANT COVARIATES

In the case of fixed covariates, it is possible to employ methods based on classic linear regression. One can compute the residuals $r_i = \log T_i + Z_i^T\hat{\beta}$ (for uncensored observations, some adjustment is needed for censored data, see Buckley and James [2]), divide into subsets i. e. by the values of one of the covariates and check the equality of the means of these subsets with t-test or Wilcoxon test. This method is very straightforward and one can easily get an idea whether the dependence on each covariate is well described by the AFT model. The downside is that the residuals are neither independent nor identically distributed, and therefore the mentioned two-sample tests do not yield exact results. Also it cannot be adapted outright to accommodate time-varying covariates but one must take into account the type of dependence.

Bagdonavičius and Nikulin [1], p. 234, present a method for testing the model with repeated observations under each (possibly stepwise) covariate setting. This is useful in industrial testing, where the covariates often represent stress levels and can be set as desired.

Goodness-of-fit procedures based on sums of martingale residuals were proposed by Lin and Spiekerman [6] and Bagdonavičius and Nikulin [1], p. 252 for the AFT model with parametric baseline hazard and by Lin et al (1993) [7] for the Cox proportional hazards model. Here we propose a similar testing procedure also for the AFT model.

We use similar notation as Lin et al (1998) [8], using time-transformed counting processes. Let

$$N_i^*(t,\beta) = N_i(te^{-Z_i^T\beta}), \qquad Y_i^*(t,\beta) = Y_i(te^{-Z_i^T\beta}), \qquad i = 1,\dots,n.$$

$$S_0^*(t,\beta) = \sum_{i=1}^n Y_i^*(t,\beta), \qquad S_1^*(t,\beta) = \sum_{i=1}^n Y_i^*(t,\beta)Z_i,$$

$$E^*(t,\beta) = \frac{S_1^*(t,\beta)}{S_0^*(t,\beta)}, \qquad \hat{A}_0(t,\beta) = \int_0^t \frac{J(s)}{S_0^*(t,\beta)}\,\mathrm{d}N_\bullet^*(s,\beta),$$

for $J(s) = I(S_0^*(t,\beta) > 0)$. $\hat{A}_0(t,\beta)$ is the well-known Nelson–Aalen estimator of $A_0(t)$. With some algebra, the score process may be rewritten as

$$U(t,\beta) = \sum_{i=1}^n \int_0^t Q_0(s)(Z_i - E^*(s,\beta))\,\mathrm{d}N_i^*(s,\beta),$$

with $Q_0(s) = (\frac{s\alpha_0'(s)}{\alpha_0(s)} + 1)$. The estimated parameters $\hat{\beta}$ are taken as those minimizing $\|U(\beta)\|$, because the score process is not continuous in β. It can be shown, that also with other choices of $Q(s,\beta)$, such as $Q_1 \equiv 1$ or $Q_2(s,\beta) = \frac{1}{n}S_0^*(s,\beta)$, the estimated parameters are consistent and $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converge to a zero mean Gaussian process [8]. In further examples, we use simply $Q_1 \equiv 1$. Denote the martingale residuals

$$M_i^*(t,\beta) = N_i^*(t,\beta) - \int_0^t Y_i^*(s,\beta)\,\mathrm{d}A_0(s)$$

and their empirical counterparts

$$\hat{M}_i^*(t,\beta) = N_i^*(t,\beta) - \int_0^t Y_i^*(s,\beta)\,\mathrm{d}\hat{A}_0(s,\beta).$$

When the model holds, the martingale residuals should fluctuate around zero, otherwise they would deviate from zero systematically. The proposed test process is

$$W(t) = n^{-\frac{1}{2}} \sum_{i=1}^n w_i \hat{M}_i^*(t,\hat{\beta}),$$

where $w_i := f(Z_i)I(Z_i \leq z)$ are weights with a bounded function f and a vector of constants z. There are many possibilities how to choose the weights, most simple choice is to set $f(Z_i) = Z_i$ or $f(Z_i) \equiv 1$ and the elements z_k of the vector z as quantiles of corresponding covariates or no truncation ($z = \infty$). One can also try using the test with various weights and compare the results, see 4.1.

The idea of the test is to measure the distance of the process from zero, which can be done by computing

$$\sup_{t\in[0,\tau]} |W(t)| \qquad \text{or} \qquad \sup_{t\in[\delta,\tau]} \left| \frac{W(t)}{\sqrt{\widehat{\text{var}} W(t)}} \right|$$

with a suitable variance estimator and some small positive number δ to avoid possible problems at the edges.

As we show later, under the null hypothesis, the asymptotic distribution of $W(t)$ is a zero mean Gaussian process with a covariance function which is difficult to obtain. To assess whether the difference from zero is significant for given data, it is possible to devise a process $\hat{W}(t)$ which has the same limiting distribution under the null hypothesis and is easy to replicate. Denote

$$S_w(t,\beta) = \sum_i w_i Y_i^*(s,\beta), \qquad E_w^*(t,\beta) = \frac{S_w^*(t,\beta)}{S_0^*(t,\beta)}$$

$$f_N(t) = \frac{1}{n} \sum_i \Delta_i w_i f_0(t) t Z_i, \qquad f_Y(t) = \frac{1}{n} \sum_i w_i g_0(t) t Z_i, \tag{1}$$

where $f_0(t)$ and $g_0(t)$ are the baseline densities of e^{ϵ_i} and $T_i e^{Z_i^T \beta_0}$, respectively. Let $\hat{f}_N$ and $\hat{f}_Y$ be their empirical counterparts with kernel estimates $\hat{f}_0(t)$ and $\hat{g}_0(t)$. The quantities $\exp(\epsilon_i)$ can be consistently estimated by $\exp(r_i)$, with r_i being the modified regression residuals of [2]. Or we can estimate their distribution by resampling from $\hat{F}_0(t) = 1 - e^{-\hat{A}_0(t,\beta)}$. Estimates for $T_i e^{Z_i^T \beta_0}$ can be obtained by inserting the estimated parameters $\hat{\beta}$. As for the kernel estimate, it suffices to take Gaussian kernel with Silverman's commonly used bandwith $1.06\hat{\sigma} n^{-1/5}$ (Silverman, 1986 [12]).

With some algebra, it can be shown that

$$U(t,\beta_0) = \sum_{i=1}^n \int_0^t Q(s,\beta_0)(Z_i - E^*(s,\beta_0))\, \mathrm{d}M_i^*(s,\beta_0).$$

Take G_i, $i = 1,\dots,n$ as *iid* standard normals, let

$$U_w^G(t,\beta) = \sum_{i=1}^n \int_0^t Q(s,\beta)(w_i - E_w^*(s,\beta))\, \mathrm{d}\hat{M}_i^*(s,\beta) G_i,$$

$$U^G(t,\beta) = \sum_{i=1}^n \int_0^t Q(s,\beta)(Z_i - E^*(s,\beta))\, \mathrm{d}\hat{M}_i^*(s,\beta) G_i.$$

Take $\hat{\beta}^*$ as the solution of the equation

$$U(\beta) = U^G(\hat{\beta}).$$

It is of note, that $n^{\frac{1}{2}}(\hat{\beta} - \hat{\beta}^*)$ has the same limiting distribution as $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ (see [8]), which is useful for approximating the distribution of $\hat{\beta}$. Now we have all the components needed to introduce the main result:

Theorem 2.1. Under the assumptions (i) – (vi) from the Appendix, given the observed data $(N_i(t), Y_i(t), Z_i)$, $i = 1, \dots, n$, the process $W(t)$ from above has asymptotically the same distribution as

$$\begin{aligned}\hat{W}(t) =& \frac{1}{\sqrt{n}} U_w^G(t, \hat{\beta}) - \sqrt{n}\left(\hat{f}_N(t) + \int_0^t \hat{f}_Y(s)\,\mathrm{d}\hat{A}_0(s, \hat{\beta})\right)^T (\hat{\beta} - \hat{\beta}^*) \\ &- \frac{1}{\sqrt{n}} \int_0^t S_w(s, \hat{\beta})\,\mathrm{d}(\hat{A}_0(s, \hat{\beta}) - \hat{A}_0(s, \hat{\beta}^*)).\end{aligned}$$

P r o o f. The proof is deferred into the Appendix. □

We can now compute $W(t)$ for the studied data set and replicate $\hat{W}(t)$ many times. The desired p-value p of the test is the proportion of cases, in which the statistics computed from the replicated $\hat{W}(t)$ exceed the statistic computed from $W(t)$. If $p < \alpha$, we reject the hypothesis that the data follow the AFT model. The variance for the standardised variant can be computed directly from the resampled processes.

It is also possible to divide the interval $[0, \tau]$ into k subintervals, i. e. quartiles, and compute the statistic in each of the parts separately and obtain k p-values $p_1, \dots, p_k$. One possibility is then to reject the hypothesis whenever we would reject in one of the subintervals (if $\min(p_1, \dots, p_k) < \alpha$), which can lead to violating the general level of significance of the test. Another possibility is to use a Bonferroni approximation for multiple-testing and reject only if $\min(p_1, \dots, p_k) < \alpha/k$.

3. THE TEST STATISTIC – TIME-VARYING COVARIATES

We can also work with time-dependent covariates $Z_i(t)$. Cox and Oakes [4] and Lin and Ying [9] proposed representing the failure times via following time transformation:

$$e^{\epsilon_i} = h_i(T_i^*, \beta_0) = \int_0^{T_i^*} e^{Z_i^T(s)\beta_0}\,\mathrm{d}s,$$

where ϵ_i are (iid) and $Z_i(s) = (Z_{i1}(s), \dots, Z_{ip}(s))^T$ is a p-dimensional covariate process. Take the transformed counting processes as

$$N_i^{*+}(t, \beta) = \Delta_i I(h_i(T_i, \beta) \leq t), \qquad Y_i^{*+}(t, \beta) = I(h_i(T_i, \beta) \geq t).$$

We can then define the processes S_0^{*+}, S_1^{*+}, E^{*+}, $\hat{A}_0^+$, $\hat{M}_i^{*+}(t, \beta)$ and $U^+(t, \beta)$ in the same way as their equivalents in the case with the fixed covariates, using N_i^{*+} and Y_i^{*+}

instead of N_i^* and Y_i^*. Constructing the test is not entirely similar, because the weights $w_i = f(Z_i(t))I(Z_i(t) \leq z)$ would be time-dependent. For practical reasons, we work here with time-invariant weights, but it could be also shown that it is possible to use time-varying weights if they are predictable.

With appropriate weights and transformed counting processes we compute S_w^{*+}, E_w^{*+}, U_w^{G+} and U^{G+} in the same way as above. Because $e^{Z_i^T(s)\beta}$ is positive, $h_i(t,\beta)$ is increasing in t. Therefore for each fixed β an inverse function $h_i^{-1}(t,\beta)$ can be found, for which $h_i^{-1}(h_i(t,\beta),\beta) = h_i(h_i^{-1}(t,\beta),\beta) = t$. Let again f_0 and g_0 be the density functions of $h_i(T_i^*,\beta_0)$ and $h_i(T_i,\beta_0)$, respectively. Denote

$$f_N^+(t) = \frac{1}{n}\sum_i \Delta_i w_i f_0(t)\frac{\partial}{\partial\beta}\Big(- h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0}, \tag{2}$$

$$f_Y^+(t) = \frac{1}{n}\sum_i w_i g_0(t)\frac{\partial}{\partial\beta}\Big(- h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0} \tag{3}$$

and their empirical counterparts $\hat{f}_N^+(t)$ and $\hat{f}_Y^+(t)$ obtained by inserting $\hat{\beta}$ and kernel estimates $\hat{f}_0$ and $\hat{g}_0$. Again, $h_i(T_i,\beta_0)$ can be simply estimated by inserting $\hat{\beta}$. The estimation of $e^{\epsilon_i} = h_i(T_i^*,\beta_0)$ can be done similarly as in the case with constant covariates but is not as straightforward as one has to take into account the type of dependence, or one can estimate the distribution by using $\hat{F}_0^+(t) = 1 - e^{-\hat{A}_0^+(t,\beta)}$

Theorem 3.1. Suppose that (i) – (vi) rewritten for the modified variables and processes and also the assumptions (C1) – (C3) for $Z_i(t)$ (Lin and Ying [9], see Appendix) hold. Suppose that for fixed β the image of $h_i(t,\beta), t \in [0,\infty]$ does not depend on β. Let $w_i = f(Z_i(t_0))I(Z_i(t_0) \leq z)$ for a fixed time-point t_0.
Then given the data $(N_i(t), Y_i(t), Z_i(t))$, $i = 1,\ldots,n$, the resampled process $\hat{W}^+(t)$ constructed in the same way as $\hat{W}(t)$ with modified components from above has asymptotically the same distribution as $W^+(t) = \frac{1}{\sqrt{n}}\sum w_i \hat{M}_i^{*+}(t)$.

Proof. The proof is deferred into the Appendix. □

Now it is possible to perform the test in the same way as in the case with constant covariates by computing the observed process $W(t)$ and comparing with the replicated processes $\hat{W}(t)$.

The simplest case would be, if the covariate represents an additional influence which is added in given time s_i for each observed individual,

$$Z_i(t) = \begin{cases} 1 & t > s_i \\ 0 & t \leq s_i. \end{cases}$$

This means that at the time s_i the observed individual starts to age faster or slower. The covariate processes have clearly bounded variation. We get

$$h_i(t,\beta) = \min(t,s_i) + e^{\beta}(t-s_i)^+, \qquad h_i^{-1}(t,\beta) = \min(t,s_i) + e^{-\beta}(t-s_i)^+.$$

The weights for $W^+(t)$ can be chosen as $w_i = I(s_i \le z)$ for some z, i.e. $z = median(s_i)$ etc. Or we can simply sum all the residuals ($w_i \equiv 1$).

In this case, we have

$$\frac{\partial}{\partial\beta}\Big(- h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0} = -\frac{\partial}{\partial\beta}\Big(\min(t,s_i) + e^{\beta_0-\beta}(t-s_i)^+\Big)_{\beta=\beta_0} = (t-s_i)^+,$$

therefore $f_N^+(t)$ and $f_Y^+(t)$ are easy to compute.

Also the model with constant covariates can be viewed as a special case, with

$$h_i(t,\beta) = te^{Z_i^T\beta}, \qquad h_i^{-1}(t,\beta) = te^{-Z_i^T\beta},$$

therefore

$$\frac{\partial}{\partial\beta}\Big(- h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0} = -\frac{\partial}{\partial\beta}\Big(te^{Z_i^T(\beta_0-\beta)}\Big)_{\beta=\beta_0} = tZ_i^T,$$

and inserting into 2 and 3 we get $f_N(t)$ and $f_Y(t)$ as in 1.

4. SIMULATION STUDY

We shall use the proposed test in various situations. We want to study whether the test holds its level of significance and the empirical power of the test against certain alternatives for various sample sizes. Each time we consider noncensored data and data with about one quarter of the observations randomly and independently censored. As the test statistic, we took $\sup|W(t)|$ and $\sup\left|\frac{W(t)}{\sqrt{\widehat{\mathrm{var}} W(t)}}\right|$ with the variance estimated from the resampled processes. Both statistics were computed over the whole time interval and over four separated subintervals divided by the quartiles of $T_i e^{Z_i\hat\beta}$ or $h_i(T_i,\hat\beta)$ respectively. The p-value is taken as the proportion of samples in which the replicated statistics exceed the observed one. For the supremum over the whole interval, we reject the hypothesis if the p-value is lower than the significance level of $\alpha = 5\%$. If we compute the p-values over the quartiles separately, we reject firstly when $\min(p_1,p_2,p_3,p_4) < \alpha/4$, using the Bonferroni correction. Each time, 500 samples were generated and for each sample, $\hat W(t)$ was generated 200×. To examine the empirical power, we generate data from different models and observe the proportion of rightfully rejected samples. To see if the tests hold the significance level, we generate from the AFT model itself and observe the proportion of wrongfully rejected samples.

4.1. Constant covariates

First, we generated data from the AFT model itself, with lognormal baseline hazard $LN(\mu = 5, \sigma^2 = 1)$, $\beta = 1$ and one covariate Z_i generated as *(iid)* from $N(3,1)$. Censoring was generated independently in the same way with baseline LN(6.5,1). We compare weights with $f(Z_i)$ as either Z_i or equal to 1 and z as either the sample median of Z_i or infinity (no observations left out). Each time, 500 samples of 1000 observations were tested (Table 1).

Test		$[0,\tau]$				Quartiles – Bonferroni			
Statistic		$\sup\|W(t)\|$		$\sup\left\|\frac{W(t)}{\sqrt{\widehat{\text{var}}W(t)}}\right\|$		$\sup\|W(t)\|$		$\sup\left\|\frac{W(t)}{\sqrt{\widehat{\text{var}}W(t)}}\right\|$	
Censoring		NC	C	NC	C	NC	C	NC	C
The proportion of wrongfully rejected samples from the AFT model:									
Z_i	median	0	0	0.01	0.006	0	0	0.006	0.004
Z_i	∞	0	0	0	0	0	0	0	0
1	median	0	0	0.006	0.002	0	0.002	0.002	0.004
1	∞	0	0	0	0	0	0	0	0
The proportion of rightfully rejected samples from the Cox model:									
Z_i	median	0	0	0.784	0.636	0.654	0.464	0.746	0.574
Z_i	∞	0	0	0.872	0.65	0.044	0.016	0.674	0.456
1	median	0	0	0.918	0.738	0.746	0.574	0.858	0.688
1	∞	0	0	0	0	0	0	0	0

Tab. 1. The empirical level of significance of the test and empirical power against the alternative of the Cox model for various weights w_i. C – censoring, NC – without censoring.

Next, we generated data from the Cox model $\alpha_i(t) = e^{Z_i\beta}\alpha_0(t)$ with the same baseline hazard, parameter and covariates as above and the censoring in the same way with lognormal baseline distribution LN(5.7,1). To see which weights yield the highest power against this alternative, we compare the results for the four types of weights used above (Table 1).

We can see, that the empirical level of significance tends to be very low, in some cases even with no rejected samples whatsoever. This indicates that the test is overly conservative, leading possibly to a loss of power. In further examples, we will try to overcome this problem by removing the Bonferroni correction, thus rejecting the hypothesis, whenever we would reject in one of the quartiles. One has to be careful, because the removal of the correction may lead to exceeding the level of significance. Regarding the empirical power against the alternative of the Cox model, we may see that the weights $w_i = I(Z_i \leq median(Z_j))$ yield the best empirical power.

We now use these weights for testing samples of size ranging from 100 to 2000 (Table 2). The results below indicate that with increasing sample size the empirical power gets higher, however, for a reasonable power a large number of observations is still needed. Standardising with the deviation process and dividing into quartiles adds some power. With censoring, the power diminishes greatly. If we do not use the Bonferroni correction for the division in quartiles, the empirical level of significance stays below 0.05 by not much for the large samples, for the smaller, the test still seems to be too conservative.

On Figure 1 we see the test process $W(t)$ for one case generated from the Cox model with $n = 200$ and its 50 replications $\hat{W}(t)$ under the hypothesis of the AFT model.

Test	$[0,\tau]$				Quartiles			
Statistic	$\sup\lvert W(t)\rvert$		$\sup\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}}W(t)}}\right\rvert$		$\sup\lvert W(t)\rvert$		$\sup\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}}W(t)}}\right\rvert$	
Censoring	NC	C	NC	C	NC	C	NC	C
The proportion of wrongfully rejected samples from the AFT model:								
100	0	0	0	0.004	0	0.002	0.002	0.008
200	0	0	0.004	0.004	0.006	0.002	0.014	0.006
500	0	0	0	0	0		0.01	0.014
1000	0	0	0.006	0.002	0.004	0.004	0.026	0.018
2000	0	0	0.004	0.004	0.002	0.004	0.022	0.026
The proportion of rightfully rejected samples from the Cox model:								
100	0	0	0.016	0.004	0.026	0.018	0.036	0.014
200	0	0	0.112	0.042	0.124	0.056	0.202	0.100
500	0	0	0.476	0.302	0.558	0.376	0.688	0.444
1000	0	0	0.918	0.738	0.928	0.82	0.968	0.848
2000	0	0	1	0.998	1	0.988	1	1

Tab. 2. The empirical level of significance and the empirical power against the Cox model for various sample sizes.

Around the time index 100, the observed process tends to exceed the resampled values, which suggests that the model does not fit the data. The variance of the resampled processes increases with time, however in the displayed case the observed process is well between the resampled ones at the end of the time interval. Therefore the non-standardised statistic may not detect the deviation from the AFT model because the supremum of the resampled processes is near the end where the variance is larger. Variance standardising and division into quartiles helps to overcome this problem.

4.2. Time-varying covariates

Consider data with a single jump in one covariate, $Z_i(t) = I(t > s_i)$. First, we generated data from the AFT model itself, with lognormal baseline distribution $LN(5,1)$ and $\beta = 1$. The jump times s_i were generated as (*iid*) $LN(4,1)$. Censoring times were generated independently with the same distribution of jumps and baseline distribution $LN(6,1)$. We applied the test of the AFT model with weights $w_i = I(s_i \le median(s_j))$, with the plain supremum statistic, variance-adjusted version and the supremum computed over the quartiles using the Bonferroni correction. For results, see Table 3. As we see, the test holds its level of significance is in all cases below 5%, but sometimes, especially for the smaller samples, is again overly conservative.

Next, we generated from the Cox model $\alpha_i(t) = \exp(Z_i(s)\beta)\alpha_0(t)$ with the same setting and censoring generated the same way from $LN(6,1)$. For the obtained empirical power see Table 4. Without standardising or dividing into quartiles, the empirical power is surprisigly zero – this fact is discussed later. Observing the nonstandardised statistic in the quartiles separately yields better results, the power increases with the sample size. With the standardising, the power is even higher, and for each sample size stays approximately the same regardless of dividing into quartiles or censoring.

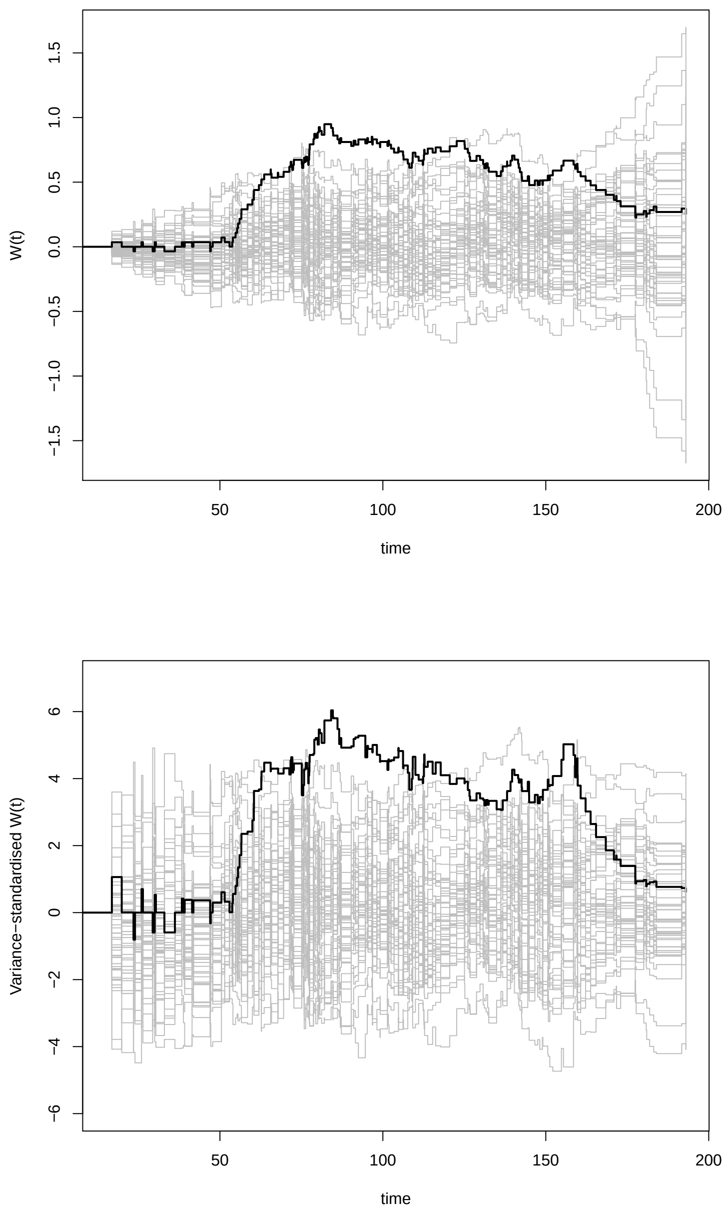

Fig. 1. The statistic W(t) (bold) with its 50 replications under the AFT model, non-standardised and variance-standardised version.

Test Statistic	$[0,\tau]$ sup $\lvert W(t)\rvert$		sup $\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}} W(t)}} \right\rvert$		Quartiles – Bonferroni sup $\lvert W(t)\rvert$		sup $\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}} W(t)}} \right\rvert$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0	0	0.01	0.004	0.004	0	0.004	0.004
200	0	0	0.014	0.02	0.006	0.01	0.004	0.01
500	0	0	0.022	0.02	0.002	0.004	0.014	0.016
1000	0	0	0.028	0.008	0	0.004	0.016	0.004
2000	0	0	0.048	0.028	0.012	0.012	0.036	0.028

Tab. 3. The empirical level of significance when generating from the AFT model with a time-varying covariate.

Test Statistic	$[0,\tau]$ sup $\lvert W(t)\rvert$		sup $\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}} W(t)}} \right\rvert$		Quartiles – Bonferroni sup $\lvert W(t)\rvert$		sup $\left\lvert \frac{W(t)}{\sqrt{\widehat{\mathrm{var}} W(t)}} \right\rvert$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0	0	0.238	0.218	0.132	0.128	0.178	0.16
200	0	0	0.754	0.704	0.38	0.298	0.662	0.584
500	0	0	1	0.998	0.846	0.808	0.996	0.996
1000	0	0	1	1	1	0.998	1	1
2000	0	0	1	1	1	1	1	1

Tab. 4. The empirical power against the Cox model with a time-varying covariate.

Finally, we generated data from the AFT model with one confounding covariate, with T_i^* satisfying $e^{\epsilon_i} = \int_0^{T_i^*} e^{Z_i(s)\beta_1 + X_i\beta_2}\,\mathrm{d}s$ with $Z_i(s) = I(s > s_i)$ same as above, X_i independent, generated from $N(3,1)$ and $\beta_1 = \beta_2 = 1$. We test whether the model holds if we try fitting it using just the covariate $Z_i(t)$. For results, see Table 5. For reasons discussed below, using the plain statistic without standardising or dividing into quartiles, the power is very low. However, if we standardise by the standard deviation process or observe the statistic in the quartiles separately, the empirical power is reasonably high. Also censoring does not reduce the power much.

For checking why the test does not reject the alternative with the plain statistic, we want to visualize the observed and resampled processes. For one generated data set of $n = 200$ with the confounding covariate we can see the test process $W(t)$ and its 50 replications under the hypothesis of the AFT model on Figure 2. The observed process lies between the replicated processes only for larger time values but exceeds them otherwise. We may therefore suppose that the model does not fit the data well.

Again, the variance of the processes increases with time. This is the reason for low

Test Statistic	$[0,\tau]$ $\sup\lvert W(t)\rvert$		$\sup\left\lvert\frac{W(t)}{\sqrt{\widehat{\text{var}}W(t)}}\right\rvert$		Quartiles – Bonferroni $\sup\lvert W(t)\rvert$		$\sup\left\lvert\frac{W(t)}{\sqrt{\widehat{\text{var}}W(t)}}\right\rvert$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0.012	0.062	0.25	0.21	0.256	0.2	0.266	0.2
200	0.002	0.012	0.404	0.376	0.422	0.394	0.426	0.398
500	0	0	0.848	0.754	0.852	0.776	0.868	0.776
1000	0	0	0.996	0.972	0.996	0.98	0.996	0.98
2000	0	0	1	1	1	1	1	1

Tab. 5. The empirical power against the AFT model with an ommited covariate.

rejection using the plain supremum statistic, because the supremum of each replicated process lies near the end of the observed interval. The observed process, however, deviates from the model notably from the beginning to the middle of the time interval, where the supremum is smaller. This can be overcome by using the variance-standardised processes or disregarding the last quartile of the data.

5. CONCLUSION

In present work we introduced a new goodness-of-fit test for the accelerated failure time model, based on martingale residuals and resampling techniques. Both the classic case with constant covariates and the generalization with time-dependent covariates were explored. Using simulated data we estimated the empirical properties of the test in various situations. In some cases, a large number of observations was needed to obtain a reasonable empirical power. As seen in the example situations, one has to consider more types of weights, work with the test process only on a part of the time interval or use the variance standardised statistic to detect possible deviation from the model.

The empirical power of the test against other important alternatives could be further studied, i. e. considering different baseline or covariate distributions or the alternative of an entirely different model. Also a more detailed analysis of which weights work best in each case as seen in the section 4.1 could be conducted. Moreover, the generalization for time-dependent covariates provides a broad range of applications which may yet be explored.

6. APPENDIX

We now prove the asymptotic equivalency of $W(t)$ and $\hat{W}(t)$. First we work with fixed covariates and then we generalize the proof also for time-dependent covariates.

6.1. Preliminaries

We will treat the covariates Z_i as random variables. Suppose, that:

(i) Z_i are bounded,

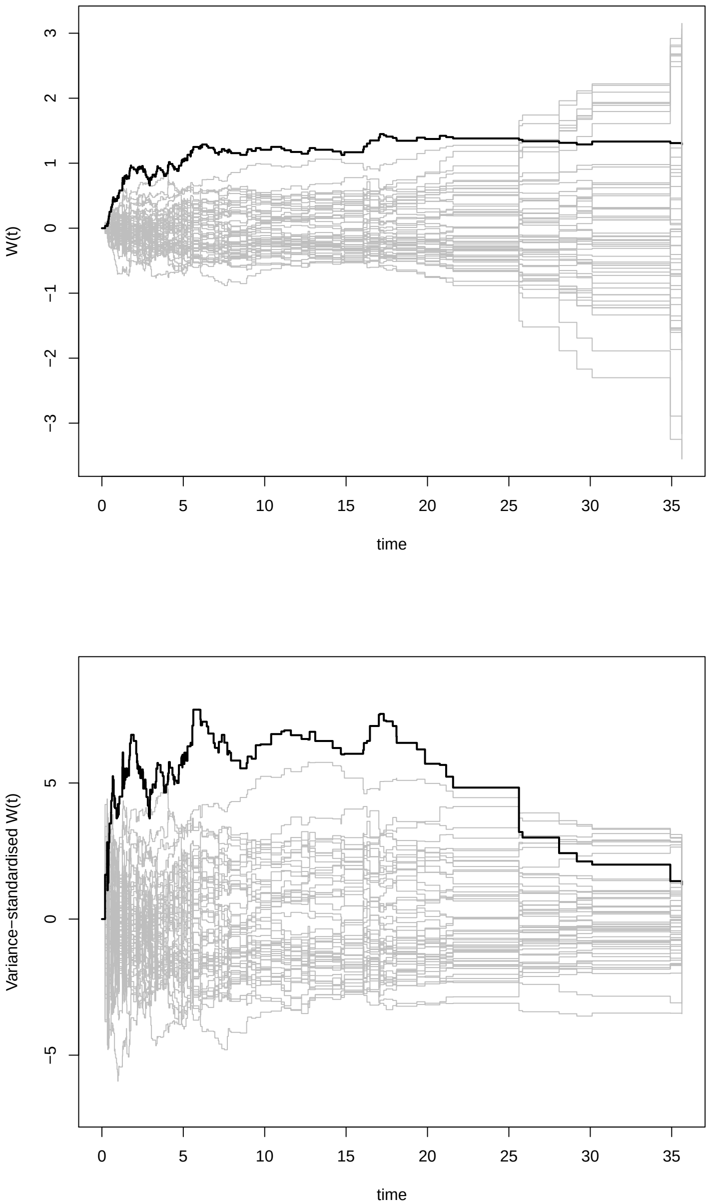

Fig. 2. The statistic W(t) (bold) with its 50 replications under the AFT model, non-standardised and variance-standardised version.

(ii) (N_i^*, C_i, Z_i) are (iid),

(iii) Q, E^*, E_w^* and $\frac{1}{n}S_w$ have bounded variation and converge almost surely to continuous functions q, e, e_w and s_w, respectively,

(iv) $C_i^* = C_i e^{Z_i^T\beta_0}$ have a uniformly bounded density and $A_0(t)$ has a bounded second derivative,

(v) $f_N(t) = \frac{1}{n}\sum_i \Delta_i w_i f_0(t) t Z_i$ and $f_Y(t) = \frac{1}{n}\sum_i w_i g_0(t) t Z_i$ have bounded variation and converge almost surely to $f_N^0(t)$ and $f_Y^0(t)$, respectively,

(vi) The kernel estimates $\hat{f}_0$ and $\hat{g}_0$ have a bounded variation and converge in probability, uniformly in $t \in [0,\tau]$, to f_0 and g_0, respectively.

Lin et al [8] shows, that under i-iv for $d_n \to 0$:

$$\sup_{\|\beta-\beta_0\|<d_n} \|U(\beta) - U(\beta_0) + nA(\beta-\beta_0)\|/(n^{\frac{1}{2}} + n\|\beta-\beta_0\|) = o_P(1), \tag{4}$$

$$\sup_{t\in[0,\tau],\|\beta-\beta_0\|<d_n} \left|n^{\frac{1}{2}}(\hat{A}_0(t,\beta) - \hat{A}_0(t,\beta_0)) - b^T(t) n^{\frac{1}{2}}(\beta-\beta_0)\right| = o_P(1), \tag{5}$$

where $A = \int_0^\tau q(t) E[Y_1^*(t,\beta_0)(Z_1 - e(t))^{\otimes 2}]\,\mathrm{d}(\alpha_0(t)t)$ and $b(t) = -\int_0^t e(s)\,\mathrm{d}(\alpha_0(s)s)$.

6.2. Convergence for sums of N_i^* and Y_i^*

First, we need to show the asymptotic properties of f_N and f_Y:

Lemma 6.1. Conditional on Z_i, under (i)–(vi) for $d_n \to 0$:

$$\sup_{t\in[0,\tau],\|\beta-\beta_0\|<d_n} \left|n^{-\frac{1}{2}} \sum w_i(N_i^*(t,\beta) - N_i^*(t,\beta_0)) + f_N^T(t) n^{\frac{1}{2}}(\beta-\beta_0)\right| = o_P(1), \tag{6}$$

$$\sup_{t\in[0,\tau],\|\beta-\beta_0\|<d_n} \left|n^{-\frac{1}{2}} \sum w_i(Y_i^*(t,\beta) - Y_i^*(t,\beta_0)) - f_Y^T(t) n^{\frac{1}{2}}(\beta-\beta_0)\right| = o_P(1), \tag{7}$$

with f_N and f_Y defined in 1.

P r o o f. In this proof we treat Z_i as fixed values. We have

$$\begin{aligned}
n^{-\frac{1}{2}} &\sum w_i(N_i^*(t,\beta) - N_i^*(t,\beta_0)) \\
&= n^{-\frac{1}{2}} \sum w_i \Delta_i [I(T_i^* e^{Z_i^T\beta} \le t) - I(T_i^* e^{Z_i^T\beta_0} \le t)] \\
&= n^{-\frac{1}{2}} \sum w_i \Delta_i [I(T_i^* \le t e^{-Z_i^T\beta}) - I(T_i^* \le t e^{-Z_i^T\beta_0})] \\
&= n^{-\frac{1}{2}} \sum w_i \Delta_i [I(t e^{-Z_i^T\beta_0} < T_i^* \le t e^{-Z_i^T\beta}) - I(t e^{-Z_i^T\beta} < T_i^* \le t e^{-Z_i^T\beta_0})].
\end{aligned}$$

From Lemma 1 of Lin and Ying [9] it follows that uniformly in $t \in [0,\tau]$:

$$\sup_{\|\beta-\beta_0\|<d_n} \left|n^{-\frac{1}{2}} \sum w_i(N_i^*(t,\beta) - N_i^*(t,\beta_0)) - n^{-\frac{1}{2}} E \sum w_i(N_i^*(t,\beta) - N_i^*(t,\beta_0))\right| = o_P(1)$$

and analogically for Y^*. Hence, it suffices to compute the expectation of the sum of indicators. For summand i we have $E[w_i\Delta_i[I(\cdot)-I(\cdot)]] = E[w_i\Delta_i E[I(\cdot)-I(\cdot)|\Delta_i]]$. The inner expectation equals to

$$\begin{aligned}&E[I(te^{-Z_i^T\beta_0} < T_i^* \le te^{-Z_i^T\beta}) - I(te^{-Z_i^T\beta} < T_i^* \le te^{-Z_i^T\beta_0})]\\ &= P(t < T_i^* e^{Z_i^T\beta_0} \le te^{Z_i^T(\beta_0-\beta)}) - P(te^{Z_i^T(\beta_0-\beta)} < T_i^* e^{Z_i^T\beta_0} \le t).\end{aligned}$$

Either the first or the second probability is zero, because the cases are mutually exclusive. Assume first, that $te^{Z_i^T(\beta_0-\beta)} > t$, which is equivalent with $Z_i^T\beta_0 > Z_i^T\beta$. Because $T_i^* e^{Z_i^T\beta_0}$ are (iid) with the distribution function F_0, we have

$$\begin{aligned}&P(t < T_i^* e^{Z_i^T\beta_0} \le te^{Z_i^T(\beta_0-\beta)}) = F_0(te^{Z_i^T(\beta_0-\beta)}) - F_0(t)\\ &\quad = f_0(t)t(e^{Z_i^T(\beta_0-\beta)} - 1) + o_P(1) = f_0(t)tZ_i^T(\beta_0-\beta) + o_P(1).\end{aligned}$$

We used the Taylor expansion for $\beta \to \beta_0$ twice. For $Z_i^T\beta_0 < Z_i^T\beta$ we get the same result, because

$$-P(te^{Z_i^T(\beta_0-\beta)} < T_i^* e^{Z_i^T\beta_0} \le t) = -(F_0(t) - F_0(te^{Z_i^T(\beta_0-\beta)})).$$

Therefore we get the desired result with an conditional expectation with Δ_i, we have

$$n^{-\frac{1}{2}}\sum w_i(N_i^*(t,\beta) - N_i^*(t,\beta_0)) = E[(\frac{1}{n}\sum w_i\Delta_i f_0(t)tZ_i)^T(\beta_0-\beta)n^{\frac{1}{2}}] + o_P(1).$$

Because the censoring is independent, due to SLNN we can replace the expectation with the observed quantity:

$$= (\frac{1}{n}\sum w_i\Delta_i f_0(t)tZ_i)^T(\beta_0-\beta)n^{\frac{1}{2}} + o_P(1) = -n^{\frac{1}{2}} f_N^T(t)(\beta-\beta_0) + o_P(1).$$

For the sums of Y_i^*, we have

$$\begin{aligned}n^{-\frac{1}{2}}\sum w_i(Y_i^*(t,\beta) - Y_i^*(t,\beta_0)) &= n^{-\frac{1}{2}}\sum w_i[I(T_i \ge te^{-Z_i^T\beta}) - I(T_i \ge te^{-Z_i^T\beta_0})]\\ &= n^{-\frac{1}{2}}\sum w_i[I(t > \min(T_i^* e^{Z_i^T\beta_0}, C_i e^{Z_i^T\beta_0}) \ge te^{Z_i^T(\beta_0-\beta)})\\ &\qquad - I(te^{Z_i^T(\beta_0-\beta)} > \min(T_i^* e^{Z_i^T\beta_0}, C_i e^{Z_i^T\beta_0}) \ge t)].\end{aligned}$$

We assumed that $C_i^* = C_i e^{Z_i^T\beta_0}$ have a bounded density and therefore $T_i e^{Z_i^T\beta_0} = \min(T_i^*, C_i)e^{Z_i^T\beta_0}$ can be assumed to have a density g_0. Computing again the expectation and using the Taylor expansion, we get

$$\begin{aligned}n^{-\frac{1}{2}}\sum w_i(Y_i^*(t,\beta) - Y_i^*(t,\beta_0)) &= (n^{-1}\sum w_i g_0(t)tZ_i)^T(\beta-\beta_0)n^{\frac{1}{2}} + o_P(1)\\ &= n^{\frac{1}{2}} f_Y^T(t)(\beta-\beta_0) + o_P(1).\end{aligned}$$

□

6.3. The convergence of the statistic $W(t)$ and $\hat{W}(t)$

Proof of Theorem 2.1. We show the asymptotic equivalence by proving the convergence of finite-dimensional distributions and tightness, with the help of multivariate functional central limit theorem given by Pollard (see [10]).

$$\begin{aligned}W(t) =&n^{-\frac{1}{2}}\sum_i w_i\hat{M}_i^*(t,\hat{\beta})\\ =&n^{-\frac{1}{2}}\sum_i w_i M_i^*(t,\beta_0)+n^{-\frac{1}{2}}\sum_i w_i(\hat{M}_i^*(t,\hat{\beta})-M_i^*(t,\beta_0))\\ =&n^{-\frac{1}{2}}\sum_i w_i M_i^*(t,\beta_0)+n^{-\frac{1}{2}}\sum_i w_i(N_i^*(t,\hat{\beta})-N_i^*(t,\beta_0))\\ &-n^{-\frac{1}{2}}\sum_i w_i\int_0^t\Big(Y_i^*(s,\hat{\beta})\,\mathrm{d}\hat{A}_0(s,\hat{\beta})-Y_i^*(s,\beta_0)\,\mathrm{d}A_0(s)\Big).\end{aligned}$$

Applying (6) and adding and subtracting $Y_i^*(s,\hat{\beta})dA_0(s)$ and $Y_i^*(s,\beta_0)d\hat{A}_0(s,\hat{\beta})$ we get

$$\begin{aligned}W(t) =&n^{-\frac{1}{2}}\sum_i w_i M_i^*(t,\beta_0)-n^{\frac{1}{2}}f_N^T(t)(\hat{\beta}-\beta_0)\\ &-n^{-\frac{1}{2}}\sum_i w_i\int_0^t Y_i^*(s,\beta_0)\,\mathrm{d}\left(\hat{A}_0(s,\hat{\beta})-A_0(s)\right)\\ &-n^{-\frac{1}{2}}\sum_i w_i\int_0^t(Y_i^*(s,\hat{\beta})-Y_i^*(s,\beta_0))\,\mathrm{d}A_0(s)+o_P(1).\end{aligned}$$

With the help of (4) and (5) we have

$$\begin{aligned}n^{\frac{1}{2}}(\hat{A}_0(s,\hat{\beta})-A_0(s)) &= n^{\frac{1}{2}}(\hat{A}_0(s,\beta_0)-A_0(s))+b^T(t)n^{\frac{1}{2}}(\hat{\beta}-\beta_0)+o_P(1)\\ &= n^{\frac{1}{2}}\sum_i\int_0^t\frac{\mathrm{d}M_i^*(s,\beta_0)}{S_0^*(s,\beta_0)}+b^T(t)n^{-\frac{1}{2}}A^{-1}U(\beta_0)+o_P(1).\end{aligned}$$

We apply (7) on the last term of $W(t)$ and then (4) for $n^{\frac{1}{2}}(\hat{\beta}-\beta_0)=n^{-\frac{1}{2}}A^{-1}U(\beta_0)+o_P(1)$:

$$\begin{aligned}W(t) &= n^{-\frac{1}{2}}\sum_i w_i M_i^*(t,\beta_0)-n^{\frac{1}{2}}\left(f_N(t)+\int_0^t f_Y(s)\,\mathrm{d}A_0(s)\right)^T(\hat{\beta}-\beta_0)\\ &-n^{-\frac{1}{2}}\sum_i\int_0^t\frac{S_w(s,\beta_0)}{S_0^*(s,\beta_0)}\,\mathrm{d}M_i^*(s,\beta_0)-n^{-\frac{1}{2}}\int_0^t S_w(s,\beta_0)\,\mathrm{d}b^T(s)A^{-1}U(\beta_0)+o_P(1)\\ &= n^{-\frac{1}{2}}\sum\int_0^t(w_i-E_w^*(s,\beta_0))\,\mathrm{d}M_i^*(s,\beta_0)\\ &-n^{-\frac{1}{2}}\left(f_N(t)+\int_0^t f_Y(s)\,\mathrm{d}A_0(s)+\int_0^t\frac{1}{n}S_w(s,\beta_0)\,\mathrm{d}b(s)\right)^T A^{-1}U(\beta_0)+o_P(1).\end{aligned}$$

The limiting process can be found similarly as in [8]. Write

$$U_M(t) = n^{-\frac{1}{2}} \sum M_i^*(t,\beta_0), \qquad U_{MZ}(t) = n^{-\frac{1}{2}} \sum Z_i M_i^*(t,\beta_0),$$

$$U_{MW}(t) = n^{-\frac{1}{2}} \sum w_i M_i^*(t,\beta_0).$$

For fixed t, each of the processes is a sum of iid zero-mean terms and therefore the finite-dimensional convergence of (U_M, U_{MZ}, U_{MW}) follows from multivariate central limit theorem. For each t, $M_i^*(t,\beta_0)$, $Z_i M_i^*(t,\beta_0)$ and $w_i M_i^*(t,\beta_0)$ can be written as sums and products of monotone functions, and therefore are manageable in sense of [10], p. 38. It then follows from the functional central limit theorem ([10], p. 53) that (U_M, U_{MZ}, U_{MW}) is tight and converges weakly to a zero-mean Gaussian process, say (W_M, W_{MZ}, W_{MW}). By the Skorokhod – Dudley – Wichura theorem ([11], p. 47), an equivalent process (U_M, U_{MZ}, U_{MW}) in an alternative probability space can be found, in which the convergence becomes almost sure. Because $Q(t,\beta_0)$, $E^*(t,\beta_0)$, $E_w^*(t,\beta_0)$, $\frac{1}{n}S_w(t,\beta_0)$, $f_N(t)$ and $f_Y(t)$ have bounded variation and converge almost surely to q, e, e_w, s_w, $f_N^0(t)$ and $f_Y^0(t)$, respectively, then $W(t)$ converges in $D[0,\tau]$ to

$$\int_0^t \mathrm{d}W_{MW}(s) - \int_0^t e_w(s,\beta_0)\,\mathrm{d}W_M(s) - c^T(t)\Big(\int_0^\tau q(s)\,\mathrm{d}W_{MZ} - \int_0^\tau q(s)e(s,\beta_0)\,\mathrm{d}W_M\Big),$$

where $c(t) = f_N^0(t) + \int_0^t f_Y^0(s)\,\mathrm{d}A_0(s) + \int_0^t s_w(s,\beta_0)\,\mathrm{d}b(s)$, which has zero mean and covariance function

$$\sigma(t_1,t_2) = E\Big(\Big[\int_0^{t_1} (w_1 - e_w(s,\beta_0))\,\mathrm{d}M_1^*(s,\beta_0) - c^T(t_1)A^{-1}\int_0^\tau q(s)[Z_1 - e(s,\beta_0)]\,\mathrm{d}M_1^*(s,\beta_0)\Big]$$
$$\times \Big[\int_0^{t_2} (w_1 - e_w(s,\beta_0))\,\mathrm{d}M_1^*(s,\beta_0) - c^T(t_2)A^{-1}\int_0^\tau q(s)[Z_1 - e(s,\beta_0)]\,\mathrm{d}M_1^*(s,\beta_0)\Big]\Big).$$

For $\hat{W}(t)$, we have

$$\begin{aligned}
\hat{W}(t) =& n^{-\frac{1}{2}} U_w^G(t,\hat\beta) - n^{\frac{1}{2}} \left(\hat f_N(t) + \int_0^t \hat f_Y(s)\,\mathrm{d}\hat A_0(s,\hat\beta)\right)^T (\hat\beta - \hat\beta^*) \\
&- n^{-\frac{1}{2}} \int_0^t S_w(s,\hat\beta)\,\mathrm{d}(\hat A_0(s,\hat\beta) - \hat A_0(s,\hat\beta^*)) \\
=& n^{-\frac{1}{2}} \sum \int_0^t (w_i - E_w^*(s,\hat\beta))\,\mathrm{d}\hat M_i^*(s,\hat\beta) G_i \\
&- n^{\frac{1}{2}} \left(\hat f_N(t) + \int_0^t \hat f_Y(s)\,\mathrm{d}\hat A_0(s,\hat\beta)\right)^T (\hat\beta - \hat\beta^*) \\
&- n^{\frac{1}{2}} \int_0^t \frac{1}{n} S_w(s,\hat\beta)\,\mathrm{d}b(s)(\hat\beta - \hat\beta^*) + o_P(1) \\
=& n^{-\frac{1}{2}} \sum \int_0^t (w_i - E_w^*(s,\hat\beta))\,\mathrm{d}\hat M_i^*(s,\hat\beta) G_i \\
&- n^{-\frac{1}{2}} \left(\hat f_N(t) + \int_0^t \hat f_Y(s)\,\mathrm{d}\hat A_0(s,\hat\beta) + \int_0^t \frac{1}{n} S_w(s,\hat\beta)\,\mathrm{d}b(s)\right)^T A^{-1} U(\hat\beta^*) + o_P(1).
\end{aligned}$$

We used (4) for

$$n^{\frac{1}{2}}(\hat\beta - \hat\beta^*) = n^{-\frac{1}{2}} A^{-1} U(\hat\beta^*) + o_P(1)$$

and (5) for

$$n^{\frac{1}{2}}(\hat{A}_0(t,\hat{\beta}) - \hat{A}_0(t,\hat{\beta}^*)) = b^T(t)n^{\frac{1}{2}}(\hat{\beta} - \hat{\beta}^*) + o_P(1).$$

The score process satisfies $U(\hat{\beta}^*) = U^G(\hat{\beta})$ and therefore we see that $\hat{W}(t)$ consists of the same parts as $W(t)$, with β_0, $M_i^*(t,\beta_0)$, $f_N(t)$ and $f_Y(t)$ replaced with $\hat{\beta}$, $G_i\hat{M}_i^*(t,\hat{\beta})$, $\hat{f}_N(t)$ and $\hat{f}_Y(t)$. The resampled martingale residuals $G_iM_i^*(t,\hat{\beta})$ have the same distribution as their theoretical counterparts, and the kernel estimates of f_0 and g_0 converge uniformly to the real densities. Therefore $\hat{W}(t)$ has the same limiting finite-dimensional distributions as $W(t)$. Tightness follows also by the same arguments as for $W(t)$. □

6.4. Time-varying covariates

We can modify the assumptions (i) – (vi) to accommodate N_i^{*+}, Y_i^{*+} and all derived generalized processes. Suppose also, that following assumptions for $Z_i(t)$ of Lin and Ying [9] hold:

(C1) $\forall\, i = 1,\dots,n, k = 1,\dots,p$: $Z_{ik}(t)$ have uniformly bounded total variation, i. e. $\exists\, D : Z_{ik}(0) + \int_0^\tau |\,\mathrm{d}Z_{ik}(s)| \le D$.

Because of (C1), Z_{ik} can be decomposed into $Z_{ik}(t) = Z_{ik}(0) + Z_{ik}^+(t) - Z_{ik}^-(t)$, where $Z_{ik}^{\pm}(\cdot)$ are increasing functions with $Z_{ik}^{\pm}(0) = 0$.

(C2) There exist $\eta_0 > 0$ and $\kappa_0 > 0$, such that $\forall\, k = 1,\dots,p$:

$$\sup_{|t-s|+\|\beta_1-\beta_2\|\le n^{-\kappa_0}} \frac{1}{n}\sum_{i=1}^n |Z_{ik}^{\pm}(h^{-1}(t,\beta_1)) - Z_{ik}^{\pm}(h^{-1}(s,\beta_2))| = O_P(n^{-\frac{1}{2}-\eta_0})$$

and for $d_n > 0$, $d_n \to 0$ exists ϵ_0, such that $\forall\, k = 1,\dots,p$:

$$\sup_{|t-s|+\|\beta_1-\beta_2\|\le d_n} \frac{1}{n}\sum_{i=1}^n |Z_{ik}^{\pm}(h^{-1}(t,\beta_1)) - Z_{ik}^{\pm}(h^{-1}(s,\beta_2))| = o_P(max(d_n^{\epsilon_0}, n^{-\epsilon_0})).$$

(C3) f_0 and f_0' are bounded, $\int_0^\tau \left(\frac{f_0'(s)}{f_0(s)}\right)^2 f(s)\,\mathrm{d}s < \infty$ and $\int_0^\tau x^{\epsilon_0} f(x)\,\mathrm{d}x < \infty$ for some $\epsilon_0 > 0$.

Then the results (4) and (5) (due to Lin et al [8]) hold for the case of time-dependent covariates, too. We can also extend (6) and (7):

Lemma 6.2. Suppose that for fixed β the image of $h_i(t,\beta), t \in [0,\infty]$ does not depend on β. Conditional on Z_i, under the assumptions (i) – (vi) rewritten for the the modified variables and processes and (C1) – (C3), for $d_n \to 0$:

$$\sup_{t\in[0,\tau],\|\beta-\beta_0\|<d_n} \left|n^{-\frac{1}{2}}\sum w_i(N_i^{*+}(t,\beta) - N_i^{*+}(t,\beta_0)) + (f_N^+(t))^T n^{\frac{1}{2}}(\beta-\beta_0)\right| = o_P(1),$$

$$\sup_{t\in[0,\tau],\|\beta-\beta_0\|<d_n} \left|n^{-\frac{1}{2}}\sum w_i(Y_i^{*+}(t,\beta) - Y_i^{*+}(t,\beta_0)) - (f_Y^+(t))^T n^{\frac{1}{2}}(\beta-\beta_0)\right| = o_P(1),$$

with f_N^+ and f_Y^+ defined in 2 and 3.

Proof of Theorem 2.1. We proceed similarly as in the proof of Lemma 6.1:

$$\begin{aligned}
&n^{-\frac{1}{2}}\sum w_i(N_i^{*+}(t,\beta)-N_i^{*+}(t,\beta_0))\\
&\quad= n^{-\frac{1}{2}}\sum w_i\Delta_i[I(h_i(T_i^*,\beta)\le t)-I(h_i(T_i^*,\beta_0)\le t)]\\
&\quad= n^{-\frac{1}{2}}\sum w_i\Delta_i[I(T_i^*\le h_i^{-1}(t,\beta))-I(T_i^*\le h_i^{-1}(t,\beta_0))]\\
&\quad= n^{-\frac{1}{2}}\sum w_i\Delta_i[I(h_i^{-1}(t,\beta_0)<T_i^*\le h_i^{-1}(t,\beta))-I(h_i^{-1}(t,\beta)<T_i^*\le h_i^{-1}(t,\beta_0))].
\end{aligned}$$

Again, it can be shown that it suffices to compute the expectation of the sum of indicators ([9], Lemma 1). For each part, the inner conditional expectation is

$$\begin{aligned}
&E[I(h_i^{-1}(t,\beta_0)<T_i^*\le h_i^{-1}(t,\beta))-I(h_i^{-1}(t,\beta)<T_i^*\le h_i^{-1}(t,\beta_0))]\\
&\quad= P(t<h_i(T_i^*,\beta_0)\le h_i(h_i^{-1}(t,\beta),\beta_0))-P(h_i(h_i^{-1}(t,\beta),\beta_0)<h_i(T_i^*,\beta_0)\le t).
\end{aligned}$$

Both cases are mutually exclusive, suppose first that $h_i(h_i^{-1}(t,\beta),\beta_0)>t$. Because $h_i(T_i^*,\beta_0)=e^{\epsilon_i}$ are (iid), we have

$$\begin{aligned}
P(t<h_i(T_i^*,\beta_0)\le h_i(h_i^{-1}(t,\beta),\beta_0)) &= F_0(h_i(h_i^{-1}(t,\beta),\beta_0))-F_0(t)\\
&= f_0(t)\Big(\frac{\partial}{\partial\beta}\Big(h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0}\Big)^T(\beta-\beta_0)+o_P(1)
\end{aligned}$$

using Taylor expansion for $\beta\to\beta_0$. For $h_i(h_i^{-1}(t,\beta),\beta_0)<t$ we get again the same result. Inserting into the sum and replacing the expectation with respect to Δ_i with the observed quantity we get

$$\begin{aligned}
&n^{-\frac{1}{2}}\sum w_i(N_i^{*+}(t,\beta)-N_i^{*+}(t,\beta_0))\\
&\qquad= \Big(\frac{1}{n}\sum w_i\Delta_i f_0(t)\frac{\partial}{\partial\beta}\Big(h_i(h_i^{-1}(t,\beta),\beta_0)\Big)_{\beta=\beta_0}\Big)^T(\beta-\beta_0)n^{\frac{1}{2}}+o_P(1)\\
&\qquad= -n^{\frac{1}{2}}(f_N^+(t))^T(\beta-\beta_0)+o_P(1).
\end{aligned}$$

In similar way we obtain also the result for sums of Y_i^{*+}. □

Proof of Theorem 3.1. The proof is analogous to the proof of Theorem 2.1, using Lemma 6.2. □

ACKNOWLEDGEMENT

This work was supported by grants of the Ministry of Education, Youth and Sports of the Czech Republic 1M06047 and SVV 261315/2011.

(Received December 12, 2011)

REFERENCES

[1] V. Bagdonavičius and M. Nikulin: Accelerated Life Models. Chapman and Hall / CRC, Boca Raton 2002.

[2] J. Buckley, and I. R. James: Linear regression with censored data. Biometrika *66* (1979), 429–436.

[3] D. R. Cox: Regression models and life tables. J. Roy. Statist. Soc. Ser. B *34* (1972), 187–220.

[4] D. R. Cox and D.Oakes: Analysis of Survival Data. Chapman and Hall / CRC, Boca Raton1984.

[5] T. R. Fleming and D. P. Harrington: Counting Processes and Survival Analysis. Wiley, New York 1991.

[6] D. Y. Lin and C. F. Spiekerman: Model checking techniques for parametric regression with censored data. Scand. J. Statist. *23* (1996), 157–177.

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

[8] D. Y. Lin, L. J. Wei, and Z. Ying: Accelerated failure time models for counting processes. Biometrika *85* (1998), 605–618.

[9] D. Y. Lin and Z. Ying: Semiparametric inference for the accelerated life model with time-dependent covariates. J. Statist. Plann. Inference *44* (1995), 47–63.

[10] D. Pollard: Empirical Processes: Theory and Applications. Hayward, California: IMS 1990.

[11] G. R. Shorack and J. A. Wellner: Empirical Processes with Applications to Statistics. Wiley, New York 1986.

[12] B. W. Silverman: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London 1986.

Petr Novák, Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8 and Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.

e-mail: novakp@karlin.mff.cuni.cz