Akademie věd České republiky
Ústav teorie informace a automatizace

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

Petr Novák:

Goodness–of–fit test for the AFT model based on martingale residuals

No. 2309 October 2011

ÚTIA AV ČR, P. O. Box 18, 182 08 Prague, Czech Republic
Telex: 122018 atom c, Fax: (+42) (2) 688 4903
E-mail: utia@utia.cas.cz

Goodness-of-fit test for the AFT model based on martingale residuals

Petr Novák, KMPS MFF UK, ÚTIA AV ČR

24. května 2011

Abstract

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 statistics based on modern martingale theory. On simulated data we try to estimate empirical properties of the test for various situations.

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, Buckley&James 1979) presents an alternative to the most widely used and well described Cox proportional hazard model (Cox, 1972). In the AFT model, we assume the log-linear dependence

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

where T_i^*, $i = 1, ..., n$, are the real failure times, $Z_i = (Z_{i1}, ..., Z_{ip})^T$ covariates, β_0 the vector of real parameters and ϵ_i (iid). 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 T_i^* and C_i independent for all i. We observe independent data (T_i, Δ_i, Z_i), $i = 1, ..., n$.

We assume T_i^* to be continuous. 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 $A_i(t) = \int_0^t \alpha_i(s) ds$ the cumulative hazard. For the AFT model, we have

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

We assume that the baseline hazard $\alpha_0(t)$ is completely unknown and is estimated nonparametrically.

If we want to work with time-dependent covariates, the generalization of Lin&Ying (1995) may be used, in which the failure times are taken as

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

The data may be represented as counting processes, denote $N_i(t) = I(T_i \leq t, \Delta_i = 1)$, $Y_i(t) = I(t \leq T_i)$, intensities $\lambda_i(t) = Y_i(t)\alpha_i(t)$ and cumulative intensities $\Lambda_i(t) = \int_0^t \lambda_i(s) ds$. All functions and processes are 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 $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), \boldsymbol{X}_i, 0 \le s \le t, i = 1, ..., n\}$ (Fleming & Harrington, 1991). Therefore $M_i(t) := N_i(t) - \Lambda_i(t)$ are F_t-martingales (Doob-Meier decomposition). The log-likelihood for the data can be then rewritten with the help of the counting processes as

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

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 $U(\boldsymbol{\beta}) \equiv U(\tau, \boldsymbol{\beta}) = \mathbf{0}$.

We present a goodness-of-fit statistics for the AFT model based on martingale approach and resampling techniques for time-invariant covariates and for an important subclass of the time-varying covariates. On simulated examples we study the empirical properties of the test in various settings.

2 The test statistics - time-invariant covariates

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

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

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

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

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

$$U(t, \beta) = \sum_{i=1}^{n} \int_0^t Q(s, \beta)(Z_i - E^*(s, \beta)) dN_i^*(s, \beta),$$

with $Q(s, \beta) = (\frac{s\alpha_0'(s)}{\alpha_0(s)} + 1)$. 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 with $Q_1(s, \beta) \equiv 1$ or $Q_2(s, \beta) = \frac{1}{n} S_0^*(s, \beta)$ instead of $Q(s, \beta)$, the estimated parameters are consistent and $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ converge to a zero mean Gaussian process. In further examples, we use simply $Q(s, \beta) \equiv 1$. Denote the martingale residuals

$$M_i^*(t, \beta) = N_i^*(t, \beta) - \int_0^t Y_i^*(s, \beta) dA_0(s, \beta)$$

and their empirical counterparts

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

With some algebra, it can be shown that

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

The proposed test statistics is

$$W(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} f(Z_i) I(Z_i \le z) \hat{M}_i^*(t, \hat{\beta}),$$

where f is a bounded function and z is a vector of constants. Denote $f_i := f(Z_i)I(Z_i \le z)$. We want to find a statistics that is easy to replicate and has the same limiting distribution. Denote

$$S_f(t,\beta) = \sum_i f_i Y_i^*(s,\beta), \qquad E_f^*(t,\beta) = \frac{S_f^*(t,\beta)}{S_0^*(t,\beta)}$$

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

where $f_0(t)$ and $g_0(t)$ are the baseline densities of e_i^{ϵ} 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 $\hat{g}_0(t)$. Take G_i, $i = 1, ..., n$ as iid standard normals, let

$$U_f^G(t,\beta) = \sum_{i=1}^{n} \int_0^t Q(s,\beta)(f_i - E_f^*(s,\beta)) d\hat{M}_i^*(s,\beta) G_i,$$

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

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

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

We prove in the appendix, that $W(t)$ has asymptotically the same distribution as

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

We can compute $W(t)$ for the studied data set and replicate $\hat{W}(t)$ many times. For testing we may use

$$sup_{t\in[0,\tau]}|W(t)| \quad or \quad sup_{t\in[0,\tau]}|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}|$$

with a suitable variance estimator. If the statistics computed from $W(t)$ exceeds $(1-\alpha)\%$ of the statistics from the replicated $\hat{W}(t)$, we reject the hypothesis that the data follow the AFT model.

3 The test statistics - one jump in covariates

We can also work with time-dependent covariates $Z_i(t)$. Lin & Ying (1995) proposed representing the failure times via following time transformation:

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

where ϵ_i are (iid). Take the transformed counting processes as

$$N_i^*(t,\beta) = \Delta_i I(h_i(T_i,\beta) \le t), \qquad Y_i^*(t,\beta) = I(h_i(T_i,\beta) \ge t),$$

the processes S_0^*, S_1^*, E^*, $\hat{A}_0$, $\hat{M}_i(t,\beta)$ and $U(\beta,t)$ are then computed similarly as with fixed covariates. Constructing the test is not entirely similar, because the weights $f_i = f(X_i)I(X_i \leq x)$ cannot be used.

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. We easily show, that

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

For the statistics $W(t) = \frac{1}{\sqrt{n}} \sum f_i \hat{M}_i(t)$, the weights can be chosen as $f_i = I(s_i \leq z)$ for some z, i.e. $z = median(s_i)$ etc. Or we can simply sum all the residuals ($f_i \equiv 1$).

With this weigths and transformed counting processes we compute S_f, U_f^G and U^G in the same way as before. The resampled process is constructed as before, only with

$$\hat{f}_N(t) = \frac{1}{n} \sum_i \Delta_i f_i \hat{f}_0(t)(t - s_i)^+, \qquad \hat{f}_Y(t) = \frac{1}{n} \sum_i f_i \hat{g}_0(t)(t - s_i)^+$$

and it has the same limiting distribution as $W(t)$.

4 Simulation study

We try using the proposed test in various situations. We want to study the empirical power of the test against certain alternatives for various sample sizes. Each time we consider noncensored data and a data with about one quarter of the observations randomly and independently censored. As test statstics, we took $sup|W(t)|$ and $sup|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}|$ with the variance estimated form the resampled processes. Both statistics were computed over $[0,\tau]$ and over four separated subintervals divided by quartiles of $T_i e^{X_i\hat{\beta}}$ or $h_i(T_i,\hat{\beta})$. The hypothesis that the AFT model holds is rejected on the significancy level of 5%, if the observed statistics exceed 95% of the replicated statistics. In the case with the division into quartiles, we reject the model whenever we would reject in one of the quartiles. 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 Cox model $\alpha_i(t) = e^{Z_i\beta}\alpha_0(t)$ with lognormal baseline hazard LN(5,1). The covariates Z_i were taken as values from $N(3,1)$ with $\beta = 1$. For the weights $f_i = f(Z_i)I(Z_i \leq z)$ we took $f(Z_i) = Z_i$ and $f \equiv 1$ and $z = median(Z_i)$ and $z = 10\%quantile(Z_i)$. Samples of 1000 observations were tested to determine which weights suit this alternative best (Tab.1). It is clear

Test Statistics		$[0,\tau]$ $sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$		quartiles $\sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$	
Censoring		NC	C	NC	C	NC	C	NC	C
x	median	0	0	0	0.042	0	0.014	0	0.048
x	10%q	0.102	0.040	0.132	0.148	0.292	0.270	0.316	0.252
1	median	0.002	0	0.140	0.126	0.024	0.016	0.144	0.140
1	10%q	0.364	0.090	0.406	0.120	0.582	0.264	0.614	0.226

Tab. 1: The proportion of rightfully rejected samples from the Cox model for various weights f_i

that the weights $f_i = I(Z_i \leq 10\%q.Z)$ yield the best empirical power against the alternative of the Cox model. We now use these weights for testing samples of various sizes (Tab.2). The results below indicate, that with increasing sample size the empirical power gets higher, however, for a reasonable power a great number of observations is still needed. Standardising with the deviation process and dividing into quartiles adds some power. With censoring, the power diminishes greatly.

Test Statistics	$[0,\tau]$ $sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$		quartiles $\sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0.010	0.006	0.022	0.012	0.048	0.054	0.052	0.048
200	0.018	0.032	0.024	0.024	0.062	0.076	0.076	0.074
500	0.174	0.022	0.146	0.066	0.308	0.106	0.340	0.124
1000	0.364	0.090	0.406	0.120	0.582	0.264	0.614	0.226
2000	0.816	0.160	0.874	0.476	0.982	0.614	0.982	0.602

Tab. 2: The empirical power against the Cox model for various sample sizes

When we generated data from the AFT model with the same baseline distribution, covariates and β, the empirical level of significance was in almost every case below 5%, sometimes even smaller than 1%.

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 Cox model $\alpha_i(t) = exp(Z_i(s)\beta)\alpha_0(t)$ with lognormal baseline distribution $LN(5,1)$ and $\beta = 1$. The times to jump s_i were generated as (iid) $e^{-1} + LN(4,1)$. We applied the test of the AFT model with weights $f_i = I(s_i \leq median(s_j))$. For the results see Tab.3. Without standardising with the estimated standard deviation process $\sqrt{\widehat{var}W(t)}$ or dividing into quartiles, the empirical power is surprisigly zero. Observing the nonstandardised statistics 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.

Test Statistics	$[0,\tau]$ $sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$		quartiles $\sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0	0	0.082	0.068	0.012	0.020	0.102	0.084
200	0	0	0.442	0.444	0.026	0.046	0.472	0.492
500	0	0	0.984	0.992	0.168	0.216	0.988	0.992
1000	0	0	1	1	0.680	0.682	1	1
2000	0	0	1	1	1	1	1	1

Tab. 3: The empirical power against the Cox model with a time-varying covariate

When we generated data from the AFT model with the same setting instead, the empirical level of significance was almost every time below 5%.

Next, we generated data from the AFT model with one confounding covariate, with T_i^* satisfying $e^{\epsilon_i} = \int_0^{T_i^*} e^{Z_i(t)\beta_1 + X_i\beta_2} dt$ with $Z_i(t)$ 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. For results, see Tab.4. For some reasons, using the statistics plain without standardising or dividing into quartiles, the power gets lower with more observations. However, if we standardise by the standard deviation process or observe the statistics in the quartiles separately, the empirical power is in some cases even better than in the previous setting. Also censoring does not diminish the power much.

Test Statistics	$[0,\tau]$ $sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$		quartiles $\sup\|W(t)\|$		$sup\|\frac{W(t)}{\sqrt{\widehat{var}W(t)}}\|$	
Censoring	NC	C	NC	C	NC	C	NC	C
100	0.210	0.238	0.270	0.224	0.438	0.374	0.452	0.366
200	0.110	0.386	0.458	0.460	0.682	0.630	0.678	0.626
500	0	0.100	0.852	0.776	0.928	0.904	0.932	0.920
1000	0	0	0.992	0.984	0.996	0.992	0.996	0.996
2000	0	0	1	1	1	1	1	1

Tab. 4: The empirical power against the AFT model with an ommited covariate

5 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.

5.1 Preliminaries

We will treat the covariates Z_i as random variables. Suppose, that:
(i) Z_i are bounded.
(ii) (N_i^*, C_i^*, Z_i) are (iid), with C_i^* being time-transformed censoring times.
(iii) Q, E^*, E_f^* and $\frac{1}{n}S_f$ have bounded variation and converge almost surely to continuous functions q, e, e_f and s_f, respectively.
(iv) C_i^* have a uniformly bounded density.
(v) $f_N(t)$ and $f_Y(t)$ have bounded variation and converge almost surely to $f_N^0(t)$ and $f_Y^0(t)$, respectively.

Lin et al (1998) shows, that under i-iv for $d_n \to 0$:

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

$$sup_{|\beta-\beta_0|<d_n} \|n^{\frac{1}{2}}(\hat{A}_0(\beta) - \hat{A}_0(\beta_0)) - b^T(t)n^{\frac{1}{2}}(\beta - \beta_0)\| = o_P(1), \tag{2}$$

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

5.2 Convergence for sums of N_i^* and Y_i^*

First, we show in a similar way, that under (i)-(v):

$$sup_{|\beta-\beta_0|<d_n} \|n^{-\frac{1}{2}} \sum f_i(N_i^*(t,\beta) - N_i^*(t,\beta_0)) - f_N^T(t)n^{\frac{1}{2}}(\beta - \beta_0)\| = o_P(1), \tag{3}$$

$$sup_{|\beta-\beta_0|<d_n} \|n^{-\frac{1}{2}} \sum f_i(Y_i^*(t,\beta) - Y_i^*(t,\beta_0)) + f_Y^T(t)n^{\frac{1}{2}}(\beta - \beta_0)\| = o_P(1), \tag{4}$$

where f_N and f_Y are from above.

Assume first, that $\beta < \beta_0$ We have

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

From Lemma 1 of Lin&Ying (1993) follows, that

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

and analogically for Y^*. It suffices to compute the expectation of the sum of indicators. Because $T_i^* e^{Z_i^T\beta_0}$ are (iid), we have

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

We used the Taylor expansion for $\beta \to \beta_0$ twice. For $\beta > \beta_0$ we get the same result. Conditioning on Δ_i, we have

$$\begin{aligned} n^{-\frac{1}{2}} \sum f_i (N_i^*(t,\beta) - N_i^*(t,\beta_0)) &= (\frac{1}{n} \sum f_i \Delta_i f_0(t) t Z_i)^T (\beta - \beta_0)\sqrt{n} + o_P(1) \\ &= n^{\frac{1}{2}} f_N^T(t)(\beta - \beta_0) + o_P(1). \end{aligned}$$

Similarly for Y_i^*, we have

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

if $\beta > \beta_0$. We assumed, that also $C_i e^{Z_i^T \beta_0}$ are (iid) and therefore $min(T_i^*, C_i) e^{Z_i^T \beta_0}$ are also (iid) with a density g_0. Computing the expectation and using the Taylor expansion, we get

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

5.3 The convergence of the statistics $W(t)$ and $\hat{W}(t)$

We show the asymptotical equivalence by proving the convergence of finite-dimensional distributions and tightness, with the help of multivariate functional central limit theorem given by Pollard (1990).

$$\begin{aligned} W(t) =& \frac{1}{\sqrt{n}} \sum_i f_i \hat{M}_i^*(t,\hat{\beta}) \\ =& \frac{1}{\sqrt{n}} \sum_i f_i M_i^*(t,\beta_0) + \frac{1}{\sqrt{n}} \sum_i f_i (\hat{M}_i^*(t,\hat{\beta}) - M_i^*(t,\beta_0)) \\ =& \frac{1}{\sqrt{n}} \sum_i f_i M_i^*(t,\beta_0) + \frac{1}{\sqrt{n}} \sum_i f_i (N_i^*(t,\hat{\beta}) - N_i^*(t,\beta_0)) \\ & - \frac{1}{\sqrt{n}} \sum_i f_i \int_0^t \left(Y_i^*(s,\hat{\beta}) d\hat{A}_0(s,\hat{\beta}) - Y_i^*(s,\beta_0) dA_0(s) \right) \end{aligned}$$

Applying (3) and adding and substracting $Y_i^*(s,\hat{\beta})dA_0(s)$ and $Y_i^*(s,\beta_0)dA_0(s,\hat{\beta})$ we get

$$\begin{aligned} W(t) =& \frac{1}{\sqrt{n}} \sum_i f_i M_i^*(t,\beta_0) + n^{\frac{1}{2}} f_N^T(t)(\hat{\beta} - \beta_0) \\ & - \frac{1}{\sqrt{n}} \sum_i f_i \int_0^t Y_i^*(s,\beta_0) d\left(\hat{A}_0(s,\hat{\beta}) - A_0(s)\right) \\ & - \frac{1}{\sqrt{n}} \sum_i f_i \int_0^t (Y_i^*(s,\hat{\beta}) - Y_i^*(s,\beta_0)) dA_0(s) + o_P(1). \end{aligned}$$

With the help of (1) and (2) we have

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

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

$$\begin{aligned}
W(t) &= \frac{1}{\sqrt{n}}\sum_i f_i M_i^*(t,\beta_0) + n^{\frac{1}{2}}\left(f_N(t) + \int_0^t f_Y(s)dA_0(s)\right)^T(\hat{\beta}-\beta_0) \\
&- \frac{1}{\sqrt{n}}\sum_i \int_0^t \frac{S_f(s,\beta_0)}{S_0^*(s,\beta_0)}dM_i^*(s,\beta_0) - n^{-\frac{1}{2}}\int_0^t S_f(s,\beta_0)db^T(s)A^{-1}U(\beta_0) + o_P(1) \\
&= \frac{1}{\sqrt{n}}\sum \int_0^t (f_i - E_f^*(s,\beta_0))dM_i^*(s,\beta_0) \\
&+ \frac{1}{\sqrt{n}}\left(f_N(t) + \int_0^t f_Y(s)dA_0(s) - \int_0^t \frac{1}{n}S_f(s,\beta_0)db(s)\right)^T A^{-1}U(\beta_0) + o_P(1).
\end{aligned}$$

The limiting process can be found similarly as in Lin et al (1998). Write

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

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_{MF}) follows from multivariate central limit theorem. For each t, $M_i^*(t,\beta_0)$, $Z_i M_i^*(t,\beta_0)$ and $f_i M_i^*(t,\beta_0)$ can be written as sums and products of monotone functions, and therefore are manageable in sense of Pollard (1990), p.38. It then follows from the functional central limit theorem (Pollard, 1990, p.53) that (U_M, U_{MZ}, U_{MF}) is tight and converges weakly to a zero-mean Gaussian process, say (W_M, W_{ZM}, W_{MF}). By the Skorokhod-Dudley-Wichura theorem (Shorack & Wellner, 1986, p.47), an equivalent process (U_M, U_{MZ}, U_{MF}) in an alternative probability space can be found, in which the convergence becomes almost sure. Becausee $Q(t,\beta_0)$, $E^*(t,\beta_0)$, $E_f^*(t,\beta_0)$, $\frac{1}{n}S_f(t,\beta_0)$, $f_N(t)$ and $f_Y(t)$ have bounded variation and converge almost surely to q, e, e_f, s_f, $f_N^0(t)$ and $f_Y^0(t)$, respectively, then $W(t)$ converges in $D[0,\tau]$ to

$$\int_0^t dW_{MF}(s) - \int_0^t e_f(s,\beta_0)dW_M(s) + c^T(t)\int_0^\tau q(s)dW_{MZ} - c^T(t)\int_0^\tau q(s)e(s,\beta_0)dW_M,$$

where $c(t) = f_N^0(t) + \int_0^t f_Y^0(s)dA_0(s) - \int_0^t s_f(s,\beta_0)db(s)$, which has zero mean and covariance function

$$\begin{aligned}
\sigma(t_1,t_2) =& E\Big(\left[\int_0^{t_1}(f_1 - e_f(s,\beta_0))dM_1^*(s,\beta_0) + c^T(t_1)A^{-1}\int_0^\tau q(s)[Z_1 - e(s,\beta_0)]dM_1^*(s,\beta_0)\right] \\
&\times \left[\int_0^{t_2}(f_1 - e_f(s,\beta_0))dM_1^*(s,\beta_0) + c^T(t_2)A^{-1}\int_0^\tau q(s)[Z_1 - e(s,\beta_0)]dM_1^*(s,\beta_0)\right]\Big).
\end{aligned}$$

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

$$\begin{aligned}
\hat{W}(t) =& \frac{1}{\sqrt{n}}U_f^G(t,\hat{\beta}) + \sqrt{n}\left(\hat{f}_N(t) + \int_0^t \hat{f}_Y(s)d\hat{A}_0(s,\hat{\beta})\right)^T(\hat{\beta}-\hat{\beta}^*) \\
&- \frac{1}{\sqrt{n}}\int_0^t S_f(s,\hat{\beta})d(\hat{A}_0(s,\hat{\beta}) - \hat{A}_0(s,\hat{\beta}^*)) \\
=& \frac{1}{\sqrt{n}}\sum \int_0^t (f_i - E_f^*(s,\hat{\beta}))d\hat{M}_i^*(s,\hat{\beta})G_i + \frac{1}{\sqrt{n}}\left(\hat{f}_N(t) + \int_0^t \hat{f}_Y(s)d\hat{A}_0(s,\hat{\beta})\right)^T(\hat{\beta}-\hat{\beta}^*) \\
&- \sqrt{n}\int_0^t \frac{1}{n}S_f(s,\hat{\beta})db(s)(\hat{\beta}-\hat{\beta}^*) + o_P(1) \\
=& \frac{1}{\sqrt{n}}\sum \int_0^t (f_i - E_f^*(s,\hat{\beta}))d\hat{M}_i^*(s,\hat{\beta})G_i \\
&+ \frac{1}{\sqrt{n}}\left(\hat{f}_N(t) + \int_0^t \hat{f}_Y(s)d\hat{A}_0(s,\hat{\beta}) - \int_0^t \frac{1}{n}S_f(s,\hat{\beta})db(s)\right)^T A^{-1}U(\hat{\beta}^*) + o_P(1).
\end{aligned}$$

We used (1) for

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

and (2) 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 $G_i M_i^*(t,\hat\beta)$ have the same distribution as their theoretical counterparts, and the kernel estimates of f_0 and g_0 converge 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 $W(t)$.

5.4 Time-varying covariates

For the case with the one-jump covariate, the results (1) and (2) of Lin et al (1998) hold with assumptions slightly changed to accomodate $Z_i(t)$ as processes. The results (3) and (4) for the sums of the processes N_i^* and Y_i^* hold also, with

$$f_N(t) = \frac{1}{n}\sum_i \Delta_i f_i f_0(t)(t-s_i)^+, \qquad f_Y(t) = \frac{1}{n}\sum_i f_i g_0(t)(t-s_i)^+.$$

We take

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

Assume first, that $\beta < \beta_0$ We have

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

Again, it suffices to compute the expectation of the sum of indicators. Because $h_i(T_i^*,\beta_0)$ are (iid) and $h_i(h_i^{-1}(t,\beta),\beta_0) = min(t,s_i) + e^{\beta_0-\beta}(t-s_i)^+$, we have

$$\begin{aligned}
&EI[t < h_i(T_i^*,\beta_0) \le h_i(h_i^{-1}(t,\hat\beta),\beta_0)] = F_0(t < h_i(T_i^*,\beta_0) \le h_i(h_i^{-1}(t,\hat\beta),\beta_0)) \\
&\quad = F_0(min(t,s_i) + e^{\beta_0-\beta}(t-s_i)^+) - F_0(t).
\end{aligned}$$

For $t \le s_i$ the term is zero, for $t > s_i$ using Taylor expansion for $\beta \to \beta_0$ we get

$$\begin{aligned}
&= F_0(s_i + e^{\beta_0-\beta}(t-s_i)) - F_0(t) = f_0(t)(s_i + e^{\beta_0-\beta}(t-s_i) - t) + o_P(1) \\
&= f_0(t)(t-s_i)(e^{\beta_0-\beta} - 1) + o_P(1) = f_0(t)(t-s_i)(\beta-\beta_0) + o_P(1).
\end{aligned}$$

For $\beta > \beta_0$ we get the same result. Merging the two cases and adding into the sum we get $f_N(t)$ as shown above. In similar way we obtain also $f_Y(t)$.

Reference

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

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

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

[4] Lin D.Y., Wei L.J., Ying Z.: *Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals*, Biometrika 80, 557-572, 1993.

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

[6] Lin D.Y., Ying Z.: *Semiparametric inference for the accelerated life model with time-dependent covariates*, Journal od Statistical Planning and Inference 44, 47-63, 1995.

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

[8] Shorack G.R.,Wellner J.A.: *Empirical processes with applications to statistics*, Wiley, 1986.