ARTICLE TEMPLATE

Goodness-of-fit test for the parametric proportional hazard regression model with interval-censored data

Rieko Sakurai[a,c] and Satoshi Hattori[b]

[a]Graduate School of Medicine, Kurume University, Japan; [b]Department of Integrated Medicine, Biomedical Statistics, Osaka University Graduate School of Medicine, Japan; [c]Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Japan

ARTICLE HISTORY
Compiled May 14, 2018

ABSTRACT
Interval-censored data are common in medical research. Fully parametric models provide simple and efficient inference for the estimation of survival functions using interval-censored observations. Inference based on a parametric regression model requires the complete specification of the probability density function, and therefore, correctly specifying the model is crucial, while the regression diagnostic is a very important step. However, regression diagnostic methods for use with the interval-censored data have not been completely developed. Here, we developed a model-checking procedure based on the cumulative martingale residuals for the interval-censored observations. We employed the conditional expectation of residuals for the diagnostics, because the data showing the exact failure time cannot be obtained for the interval-censoring analyses, and developed the formal resampling-based supremum-type test and graphical model-checking techniques. A simulation study demonstrated an excellent performance of the proposed method during the detection of misspecified functional form of covariates in the finite sample. Furthermore, we used this method for the analysis of the medical checkup data obtained in Japan.

KEYWORDS
Goodness-of-fit test; Covariate functional form; Interval-censoring; Cox-Snell residual; Martingale residual;

1. Introduction

Interval-censoring commonly occurs in practice, *e.g.*, in cancer clinical trials, where tumor recurrence is often of the primary interest and it is evaluated by using the periodic computed tomography measurements, and the time-to-recurrence can only be determined as the interval between two successive measurements. Survival analysis techniques that provide the statistical methods for the analysis of censored observations play important roles in medical research and drug development. Among many types of censored observations, statistical methods for the right-censored data have been developed the most.

Of the semiparametric models, Cox proportional hazards regression models [1] are extensively applied in practice for the analysis of the right-censored data. The prefer-

CONTACT R. Sakurai Email: sakurai.rieko@gmail.com

ence for this model in contrast to the completely parametric model [2] may be due to the success of the partial likelihood method [1] and the estimation using less model assumptions, since the parametric model relies on strong assumptions, and therefore, it requires careful model identification. However, model identification in the presence of censoring is difficult. Drawing histograms to assess the distributional assumptions for the censored observations is impossible, and a graphical diagnosis using residuals is not straightforward.

Inference procedures for various semiparametric models with interval censored data have been developed, including the Cox proportional hazards model [3,4], the accelerated failure time model [6], the additive hazards model [7], and the linear transformation model [8]. In general, these approaches are complicated, and the nonparametric portion of these semiparametric models cannot be estimated well [4,5]. However, the fully parametric likelihood-based inference is a very simple and attractive alternative, especially when the primary goal is to estimate the survival or hazards functions [9]. Unfortunately, few methods are available for the checking of model assumptions when using the interval-censored data. Therefore, we developed here a method to test the goodness-of-fit (GoF) of the parametric model with the interval-censored data, which may help overcome some of the difficulties associated with the model identification.

The GoF method has two aspects, the measurement of the overall fit of the model and the analysis of covariate function correctness [5]. For the overall fit, the GoF statistics represent the difference between the parametric maximum likelihood estimator and the nonparametric maximum likelihood estimator (NPMLE). Here, an issue for interval-censoring is the slower convergence of the NPMLE than that of the standard square-root-n estimators [5]. This is crucial for the stable estimation of the statistics [10]. The GoF of the overall fit is directly affected since it uses the NPMLE. Several researchers have focused on the potential solutions of this issue, *e.g.*, Pan [11] proposed a two-sample test via multiple imputation with the approximate Bayesian bootstrap method, while Ren [10] proposed the use of the Anderson-Darling type GoF test via the resampling method based on m out n bootstraps, which allows the resampling of smaller sample size m ($< n$). Furthermore, Li and Ma [12] extensively applied the repeated measurement model with random effects. Alternatively, Nysen *et al.* [13] proposed the modified Akaike's Information Criterion (AIC) for the comparison in a more general parametric model, instead of the NPMLEs, and the analysis of the model assumption, which was used for the diagnostics of the competing risk model as well [14,15].

However, the GoF aimed at measuring the correctness of covariate functions has been considerably less developed. Here, to obtain the information necessary for the improvement of the current model, residual analyses have been useful. In the survival analyses, the distribution of the residuals is naturally skewed, and therefore, the GoF cannot be diagnosed straightforwardly by simply plotting the residuals. Nevertheless, many residuals are available for the use in the right-censored regression models, including Cox-Snell [16], the adjusted Cox-Snell, and martingale residuals [17] (for further details on the right-censored data-associated residual analysis, see [18],[19], Collett [9]). Residual analysis methods have not been studied in the context of the interval-censored data, since the interval censored data model is not a straightforward extension of the right-censored data model. Additionally, interval-censoring increases the ambiguousness of the plots, since the data obtained are not time points but, instead, time intervals [20].

The residuals are considered random variables and censored observations when the original event-time data are censored [2]. Accordingly, we determined the residual

counterparts for the interval-censored data. Here, imputation approach can be used, but the counterpart from the NPMLE of the interval-censored residuals must be estimated again, which again leads to the convergence problem. To overcome this, we used test statistics with the conditional expectation of the residuals.

Farrington [20] defined the martingale residual for the interval-censored data which is obtained by using the conditional expectation of the Cox-Snell residuals through the relationship between the martingale and Cox-Snell residuals. Since a martingale residual, as defined by Farrington [20], has the asymmetry range and its distributional property is unclear, plotting these residuals may not be informative. For this, we considered cumulative residuals for the parametric survival model with the right-censored data[21]. Their cumulative residuals were shown to be a zero-mean Gaussian process if the model is true. A previously constructed GoF test can help alleviate the ambiguity of residual behavior diagnostics and it can be applied for a semi-parametric survival model with right-censoring as well [22].

In this study, we focus on the parametric proportional hazard (PH) model for interval-censored data, and developed regression diagnostic techniques based on the cumulative residuals [21,22] with Farrington's martingale residuals. Here, we present the notations and models, and afterward, propose a model-checking procedure. Following this, we introduce the results obtained in a simulation study and by applying the proposed method to a medical examination dataset obtained in the Kouseiren Health Care Center, Kagoshima, Japan. Finally, we demonstrated theoretically that the proposed cumulative residuals converge to the zero-mean Gaussian process in Appendix.

2. Methods

2.1. *Preliminary*

Considering T an non-negative random variable representing the failure time of interest, and Z a q-dimensional vector of baseline covariates, the pth element of Z can be denoted by $Z^{(p)}$. That is, $Z^{\top}=(Z^{(1)}, Z^{(2)}, ..., Z^{(q)})$. For n subjects, with the subscript i, which is a indicator of subjects, $\{(T_i, Z_i); i = 1, 2, ..., n\}$ are independent and identically distributed (i.i.d) copies of (T, Z). In the presence of interval censoring, T can not be observed. Instead, we can observe two time points L and R, and only know the event has occurred in the time interval, $(L, R]$. Thus, the observed data is denoted as (L, R, Z).

There exist several interval censoring schemes. The general one was introduced by [4,5,8], showing that if we assume that $0 \leq Y_{\{i,1\}} < Y_{\{i,2\}} < ... < Y_{\{i,K_i\}} \leq \infty$ are ordered examination times for the ith subject. K_i is the maximum examination times for each subject. If $K_i = 1$ or 2 for all subjects, then the observation scheme becomes case-1 or case-2, respectively. For $(L_i, R_i]$ as the smallest interval that brackets T_i, *i.e.*, $L_i = \max\{Y_{\{i,k\}} : Y_{\{i,k\}} < T_i\}$ and $R_i = \min\{Y_{\{i,k\}} : Y_{\{i,k\}} \geq T_i\}$. $L_i = 0$ indicates that the ith subject is left-censored, while $R_i = \infty$ indicates that the subject is right-censored. The sequence of examination times may not be completely observed; however, only the values of L_i and R_i have to be known, since the likelihood for $K_i > 2$ can be reduced to those in case-2.

2.2. *Parametric PH model for the interval-censored data*

Let the conditional survival function given Z be $S(t|Z;\theta)$, where θ is a finite-dimensional unknown parameter vector. The survival function of the parametric proportional hazard (PH) model can be written as $S(t,Z) = \{S_0(t)\}^{\exp(\beta^{\star\top} Z)}$, where $\beta^\star$ is a vector of regression coefficients and $S_0(t)$ is the baseline survival function. One of a representative parametric survival model is the accelerated failure time (AFT) model, which is defined as

$$\log T = \beta_0 + \beta^\top Z + \sigma\epsilon,$$

where the unknown parameters can be denoted by $\theta^\top = (\beta_0,\ \beta^\top, \sigma)$, where $\beta^\top = (\beta_1, \beta_2, ..., \beta_q)$ is an unknown regression coefficient vector and σ is an unknown scale parameter, and ϵ is a random variable with a known distribution. If ϵ follows the standard extreme value distribution, the T follows the Weibull distribution. The survival function can be expressed by $S(t|Z;\theta) = \exp(-(te^{-\beta^\top Z})^{1/\sigma})$. When σ is fixed 1, the Weibull model is reduced to the Exponential model. The Weibull AFT model is a specific form of the parametric PH model for $\beta^\star = -\beta/\sigma$ and $S_0(t) = \exp(-t^{1/\sigma})$.

Another example of parametric survival model is the piecewise exponential (PE) model [23]. It is assumued that the hazard rate function is constant for the time interval, $I_j = (\tau_{j-1}, \tau_j]$ where τ_j $(j = 1, 2..., J)$ is the change point of hazard rate such as $0 = \tau_0 < \tau_1 < ... < \tau_{J-1} < \tau_J = \infty$. The conditional hazard rate function given Z of the PE model can be defined as

$$\log \lambda(t|Z) = \log \lambda_j + \beta^\top Z$$

for $t \in I_j$, where λ_j represents a constant baseline hazard for the $t \in I_j$. The survival function can be expressed as

$$S(t|Z;\theta) = \prod_{j=1}^{J-1} \exp(-\lambda_j e^{\beta^\top Z}(\tau_j - \tau_{j-1}) - \lambda_j e^{\beta^\top Z}(t - \tau_{j-1}))$$

for $t \in I_j$, where $\theta^\top = (\lambda_1, \lambda_2, ..., \lambda_J, \beta^\top)$. The parametric PH model is also a specific form of the PE model for $\beta^\star = -\beta$ and $S_0(t) = \exp(-\sum_{j=1}^{J} \lambda_j t_j)$ where $t_j = \max(0, \min(\tau_j - \tau_{j-1}, t - \tau_{j-1}))$.

Suppose that (i) the examination times (L, R) are independent of T conditional on Z, and (ii) the joint distribution of (L, R, Z) are independent of the parameter interest θ. Then, the log-likelihood function can be written as

$$\ell(\theta) = \sum_{i=1}^{n} \log\{S(L_i|Z_i;\theta) - S(R_i|Z_i;\theta)\}.$$

Define,

$$U(\theta) = \sum_{i=1}^{n} U_i(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\ell_i(\theta), \tag{1}$$

and

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'}\ell_i(\theta). \tag{2}$$

Assuming (i) and (ii), the standard large-sample likelihood theory generally applies [5]. The true value of θ is represented as θ_0, while the maximum likelihood estimator was designated as $\hat{\theta}$, which can be easily obtained, *e.g.*, by using the standard Newton-Raphson method.

2.3. *Residuals for the interval-censored data*

Cox-Snell residuals [16] can be defined as $V_i = -\log S(T_i|Z_i;\theta_0)$, and V_i follows the unit exponential distribution if the fitted model is true. In the presence of censoring, V_i cannot be calculated. For right-censored observations, residuals for the censored observation are defined using the conditional expectation of V_i given the observed data, which is called the adjusted Cox-Snell residual. Farrington [20] argued for the extension of the adjusted Cox-Snell residuals to the interval-censored observations, and defined the Cox-Snell residuals for interval-censoring as follows:

$$\begin{aligned} r_i^C(\theta_0) &= r_i^C(L_i, R_i;\theta_0) = \mathsf{E}[V_i|T_i \in (L_i, R_i], Z_i] \\ &= \frac{S(L_i|Z_i;\theta_0)\{1-\log S(L_i|Z_i;\theta_0)\} - S(R_i|Z_i;\theta_0)\{1-\log S(R_i|Z_i;\theta_0)\}}{S(L_i|Z_i;\theta_0) - S(R_i|Z_i;\theta_0)}. \end{aligned}$$

Based on the relationship between the martingale residual and the Cox-Snell residual for right-censored observations, Farrington [20] defined the martingale residuals as $r_i^L(\theta_0) = 1 - r_i^C(\theta_0)$. Interestingly, the martingale residuals corresponds to the score function for β_0, in other words, $U_i(\beta_p) = Z_i^{(p)} r_i^L(\theta_0)$ holds [20]. Its characteristic is common to the right-censored data. Another common characteristic is $\mathsf{E}[r_i^L(\theta_0)] = 0$, which is valid for the correctly specified fitted model. Thus, we proposed a regression diagnostic method aggregating the Farrington's martingale residuals.

2.4. *The proposed GoF test statistics*

The GoF test statistics can be defined as follows

$$W(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i \le z)\hat{r}_i^L,$$

where $\hat{r}_i^L = r_i^L(\hat{\theta})$, $z^\top = (z^{(1)}, z^{(2)}, ..., z^{(q)})$ is a p-dimensional vector of covariates, $\mathsf{I}(\cdot)$ represents an indicator function, and $\mathsf{I}(Z_i \le z) = \prod_{p=1}^{q} \mathsf{I}(Z_i^{(p)} \le z^{(p)})$. If the fitted model holds, $W(z)$ converges weakly to the zero-mean Gaussian process. To perform a formal GoF test, $W_{omn} = \sup_z |W(z)|$ was defined, where the supremum is taken over $Z \in \mathsf{R}^q$. Deriving the null distribution of W_{omn} analytically is difficult, since the covariance function of $W(z)$ would be complicated. Therefore, we employed a

simulation-based technique according to the previously proposed concept [21]

$$\begin{aligned}\hat{W}(z) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i \le z) \hat{r}_i^L G_i \\ &+ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{I}(Z_j \le z) \frac{\partial \hat{r}_j^L}{\partial \theta}\Big|_{\theta=\hat{\theta}} I(\hat{\theta})^{-1} \sum_{i=1}^{n} U_i(\hat{\theta}) G_i,\end{aligned}$$

where the definition of $U(\theta)$ and $I(\theta)$ were (1) and (2) in Section 2.2, respectively, and G_i represents a random variable that follows the standard normal distribution, independent of data used. Conditional on data, $\hat{W}(z)$ converges weakly to a zero-mean Gaussian process with the same covariance function as the limiting zero-mean Gaussian process of $W(z)$ when the fitted model holds (Appendix). Therefore, the unconditional distribution of $W_{omn} = \sup_z |W(z)|$ is asymptotically equivalent to the conditional distribution of $\hat{W}_{omn} = \sup_z |\hat{W}(z)|$ of the given data. $\{G_i\}$ represent the only random quantities in $\hat{W}(z)$, and any number of realizations of $\hat{W}(z)$ can be generated by using computer generation of $\{G_i\}$. The p-value for W_{omn} can be defined as $M^{-1}\sharp\{\hat{W}_{omn} > W_{omn}\}$, where $\sharp\{\cdot\}$ represents the number for realizations satisfying the condition in the bracket and M is a sufficient large number for the realization of $\hat{W}_{omn}$. The test with W_{omn} is based on $W(z)$, which accumulates residuals along the direction of all covariates in Z_i. Therefore, the test with W_{omn} was named the omnibus test.

In the model-identification processes, any improvements of the current model are important, especially when a misspecification of the current model is suggested. Let $Z_i^{(p)}$ be the kth element of Z_i and $z^{(p)}$ be a scalar. Then, a special case of $W(z)$ is defined as

$$W^{(p)}(z^{(p)}) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i^{(p)} \le z^{(p)}) \hat{r}_i^L.$$

Afterward, it can be demonstrated that, if the fitted model holds, $W^{(p)}(z^{(p)})$ converges weakly to a zero-mean Gaussian process.

$$\begin{aligned}\hat{W}^{(p)}(z^{(p)}) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i^{(p)} \le z^{(p)}) \hat{r}_i^L G_i \\ &+ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{I}(Z_j^{(p)} \le z^{(p)}) \frac{\partial \hat{r}_j^L}{\partial \theta}\Big|_{\theta=\hat{\theta}} I(\hat{\theta})^{-1} \sum_{i=1}^{n} U_i(\hat{\theta}) G_i\end{aligned}$$

Depending on data, $\hat{W}^{(p)}(z^{(p)})$ converges weakly to a zero-mean Gaussian process with the same covariance function as the limiting zero-mean Gaussian process of $W^{(p)}(z^{(p)})$ when the fitted model holds, and a supremum type GoF test based on $W_k = \sup_{z^{(p)}} |W^{(p)}(z^{(p)})|$ can be performed, where the supremum is taken over $z^{(p)} \in (-\infty, \infty)$. This test is expected to be a powerful tool for the detection of the misspecification of the $Z^{(p)}$ functional form in the fitted model, and therefore, we named it the functional-form test. In addition to the formal evaluation using the p-value, the plotting of some realizations of the null process $\tilde{W}^{(p)}(z^{(p)})$ with the realization of $W^{(p)}(z^{(p)})$ is very useful, together with the graphical examination of the

differences between the realization of $W^{(p)}(z^{(p)})$ and the null processes.

3. Simulation study

We evaluated the size and power of the proposed method for the regression diagnostics. We considered two observation processes of time-to-events: fixed and regular interval, that is $t = 1, 2, ..., 10$ for every subject (OP1) and a random interval with the fixed time points, $t = 1, 2, ..., 10$, perturbed by the addition of a random number following the uniform distribution at $[-0.25, 0.25]$ (OP2). For each of these observation processes, we considered three datasets.

In the Dataset 1, the failure time was generated by using the Weibull AFT model,

$$Dataset\ 1 : \log T = Z_1 + Z_1^2 + Z_2 + Z_2^2 + \epsilon,$$

where $Z_1 \sim U[-1, 1]$, $Z_2 \sim N(0, 1)$ and ϵ follows the standard extreme value distribution. We fitted the following Weibull AFT models to this dataset:

$$\begin{aligned} Model\ 1 : \log T &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_1^2 + \beta_3 Z_2 + \beta_4 Z_2^2 + \sigma\epsilon, \\ Model\ 2 : \log T &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_2^2 + \sigma\epsilon, \end{aligned}$$

and

$$Model\ 3 : \log T = \beta_0 + \beta_1 Z_1 + \beta_2 Z_1^2 + \beta_3 Z_2 + \sigma\epsilon.$$

In Dataset 2, the failure time was generated from the PE model,

$$Dataset\ 2 : \log \lambda(t|Z) = \log \lambda_j + Z_1 + Z_1^2 + Z_2 + Z_2^2$$

for $t \in I_j$, with $\tau = \{1, 4, 9, 16\}$ and $\lambda = \{0.05, 0.1, 3, 2\}$, where $Z_1 \sim U[-1, 1]$ and $Z_2 \sim N(0, 1)$. We fitted the following PE models:

$$\begin{aligned} Model\ 1 : \log \lambda(t|Z) &= \log \lambda_j + \beta_1 Z_1 + \beta_2 Z_1^2 + \beta_3 Z_2 + \beta_4 Z_2^2, \\ Model\ 2 : \log \lambda(t|Z) &= \log \lambda_j + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_2^2, \end{aligned}$$

and

$$Model\ 3 : \log \lambda(t|Z) = \log \lambda_j + \beta_1 Z_1 + \beta_2 Z_1^2 + \beta_3 Z_2.$$

In Dataset 3, the failure time was generated from the Weibull AFT model, where ϵ follows the normal distribution. The Weibull AFT model was applied to Dataset 3, where ϵ was assumed to follow the standard extreme value distribution, and therefore, the error distribution in the Dataset 3 was misspecified, while Dataset 1 and 2 were correctly specified in terms of the error distribution. Considering the regression model, for each dataset, Model 1 was correctly specified, and Models 2 and 3 were misspecified in terms of Z_1 and Z_2, respectively. Using each model, we evaluated the size or power of the omnibus test and performed the functional form tests for Z_1 and Z_2 based on 1,000 simulated null processes. The number of subjects was set to 100, 200, or 500. The parameters were estimated by using the SAS procedure, PROC LIFEREG, while the proposed GoF test statistics was implemented with R.

Table 1. Summary of empirical sizes and powers for the omnibus and functional form tests with the nominal level of 0.05; OP denotes the observation process.

					Functional-form test	
OP	Dataset	N	Model	Omnibus test	Z_1	Z_2
OP1	Dataset 1	100	Model 1	0.079	0.062	0.073
			Model 2	0.306	0.576	0.075
			Model 3	0.975	0.077	1
		200	Model 1	0.058	0.052	0.058
			Model 2	0.493	0.870	0.064
			Model 3	1	0.068	1
		500	Model 1	0.054	0.057	0.051
			Model 2	0.880	0.998	0.053
			Model 3	1	0.050	1
	Dataset 2	100	Model 1	0.096	0.073	0.083
			Model 2	0.212	0.495	0.071
			Model 3	0.761	0.075	0.985
		200	Model 1	0.070	0.070	0.060
			Model 2	0.388	0.776	0.065
			Model 3	0.990	0.050	1
		500	Model 1	0.067	0.069	0.058
			Model 2	0.810	1	0.060
			Model 3	1	0.060	1
	Dataset 3	100	Model 1	0.136	0.060	0.110
			Model 2	0.273	0.504	0.062
			Model 3	0.962	0.066	1
		200	Model 1	0.060	0.046	0.059
			Model 2	0.426	0.793	0.059
			Model 3	1	0.067	1
		500	Model 1	0.053	0.044	0.062
			Model 2	0.771	0.989	0.068
			Model 3	1	0.064	1
OP2	Dataset 1	100	Model 1	0.082	0.052	0.076
			Model 2	0.304	0.567	0.073
			Model 3	0.967	1	0.092
		200	Model 1	0.071	0.052	0.071
			Model 2	0.492	0.864	0.068
			Model 3	0.999	0.068	1
		500	Model 1	0.066	0.048	0.056
			Model 2	0.886	0.999	0.054
			Model 3	1	0.050	1
	Dataset 2	100	Model 1	0.102	0.075	0.088
			Model 2	0.229	0.557	0.082
			Model 3	0.792	0.063	0.99
		200	Model 1	0.063	0.062	0.071
			Model 2	0.389	0.793	0.060
			Model 3	0.983	0.047	1
		500	Model 1	0.058	0.054	0.065
			Model 2	0.811	0.997	0.057
			Model 3	1	0.054	1
	Dataset 3	100	Model 1	0.080	0.072	0.080
			Model 2	0.730	0.885	0.838
			Model 3	0.743	0.839	0.894
		200	Model 1	0.060	0.053	0.061
			Model 2	0.772	0.956	0.875
			Model 3	0.963	0.923	0.999
		500	Model 1	0.065	0.049	0.074
			Model 2	0.846	0.993	0.872
			Model 3	0.999	0.918	1

3.1. *Results*

In Table 1, empirical sizes and powers obtained for the nominal 5% level tests are presented. When using Model 1, which correctly specified the data-generating models, all omnibus and functional form tests were shown to have the empirical sizes close to the nominal level of 5%, regardless of the sample size. As Model 2 had a misspecified functional form for Z_1, using this model, the empirical powers of the functional form tests for Z_1 were shown to be larger than those obtained in the other tests. Similarly, with Model 3, which had a misspecified functional form for Z_2, the functional form tests for Z_2 had larger empirical powers than those in the other tests. These results indicate that the functional form tests are sensitive to the misspecification of the functional forms of covariates, and that they are useful for determining the nature of a misspecification. This observation can be used to improve the current model. Using Dataset 3, where both the error distribution and functional form were misspecified, the performance of functional form tests was shown to depend on the observed time point. For the regular time points, the functional form tests work well. However, when the observed time points are perturbed, the functional form tests cannot distinguish the misspecified covariate functions.

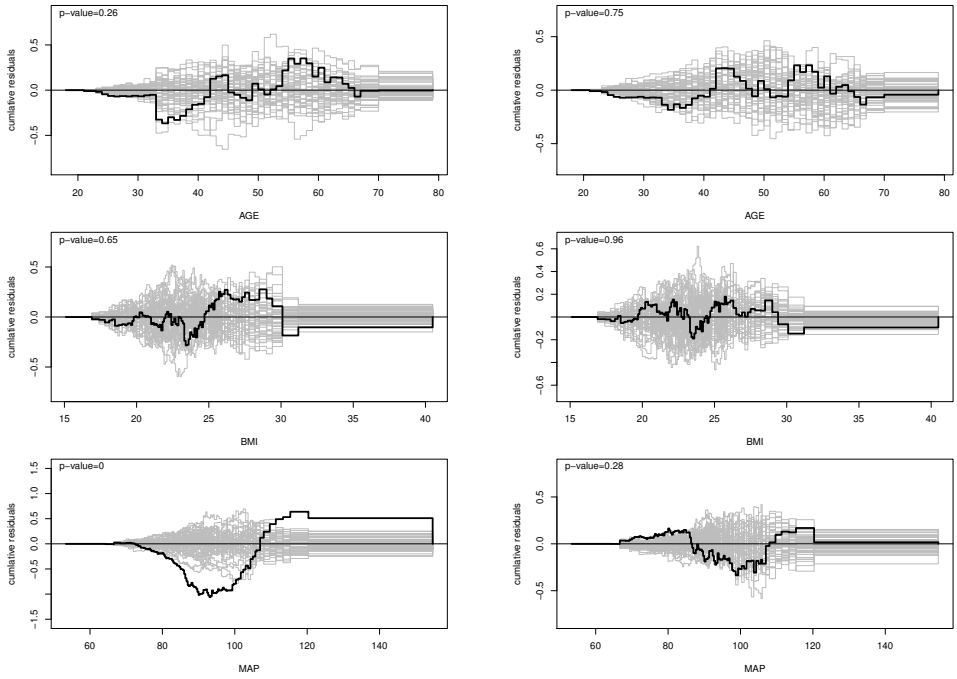
4. Application example

We illustrated our proposed method using the medical examination data collected at the Kouseiren Health Care Center in Kagoshima, Japan. This dataset included data obtained from 2,656 subjects enrolled in 2001, who received the annual medical checkups, and were followed until 2011. The objective of this analysis was to predict the time-to-initiation of the hypertension therapy. Each subject was asked whether he/she had received hypertension therapy at each annual medical checkup: the exact date of the initiation of hypertension therapy was not recorded, showing that the time-to-initiation of a hypertension therapy was interval-censored. Assuming that we were interested in estimating the survival function of a subject with a given covariate in 2001 accurately, we applied the Weibull AFT model with age, BMI (body mass index), DBP (diastolic blood pressure), and SBP (systolic blood pressure). DBP and SBP were shown to be highly correlated, and therefore, we transformed these parameters into a single parameter, the mean arterial pressure (MAP), defined as MAP:= $\frac{2}{3}$DBP+$\frac{1}{3}$SBP. No hypertension therapies have been initiated for 75 % of the analyzed subjects, and they were right-censored in 2011, while the remainder of the patients was interval-censored.

We began by applying the model:

$$Model\ 1:\ \log T \quad = \quad \beta_0 + \beta_1 AGE + \beta_2 BMI + \beta_3 MAP + \sigma\epsilon,$$

where ϵ followed the standard extreme value distribution. All regression coefficients significantly differed from 0 ($p<0.0001$). When applying the proposed method, p-values were evaluated with 1,000 simulated realizations of the null processes. The p-value of the omnibus test was 0.032, indicating that the Model 1 was misspecified. In Figure 1 (a), graphical plots of 50 randomly selected realizations of the null process $\hat{W}^{(p)}(z^{(p)})$ are shown with the realization of $W^{(p)}(z^{(p)})$ for AGE, BMI, and MAP, respectively. The realization of $W^{(p)}(z^{(p)})$ for MAP appears very different from the realizations of the simulated null process. Correspondingly, the p-values of the functional form tests



(a) Model 1 (b) Model 2

Figure 1. Plots of the cumulative martingale residuals for the functional form tests with 50 randomly selected simulated null processes for the Model 1 applied to the health checkup data.

were p=0.259 for age, p=0.650 for BMI, and p=0.001 for MAP. Therefore, the linearity assumption of MAP in Model 1 was considered suspicious. We attempted to improve the GoF by modifying the functional form of MAP, and applied the model with the quadratic term of MAP, as follows:

$$Model\ 2: \ \log T \quad = \quad \beta_0 + \beta_1 AGE + \beta_2 BMI + \beta_3 MAP + \beta_4 MAP^2 + \sigma\epsilon.$$

Once again, all regression coefficients, including the quadratic term for MAP, significantly differed from 0 (p<0.0001). The p-value of the omnibus test was 0.430. In Figure 1 (b), graphical plots for the assessment of Model 2 are presented. The corresponding p-values of the functional form tests were p=0.747 for AGE, p=0.963 for BMI, and p=0.276 for MAP. GoF for the functional form of MAP was shown to be improved and substantial misspecification of Model 2 was not suggested.

5. Discussion

Although fully parametric survival regression models are very important for the analyses of the interval-censored observations, regression diagnostic techniques are less developed. Therefore, we developed a GoF test and graphical diagnostic method, and the results of our simulation study and the application of our method for the analysis of medical examination data showed that it can efficiently detect the misspecification of the functional forms of covariates. In addition, we confirmed theoretically that the proposed method operate normally for the interval censored data under the suitable conditions (see Appendix).

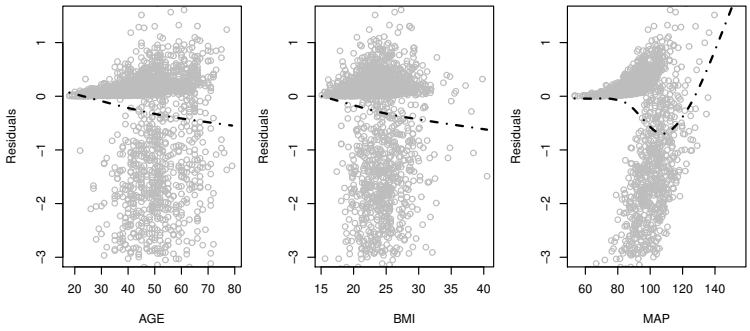


Figure 2. Plots of the adjusted Cox-Snell residual log with a smoothing curve (dotdash curve) for Model 1 applied to the medical checkup dataset. Smoothing curves were included using the function *smooth.spline* in R, according to a previous report by Hastie and Tibshirani [26].

Additionally, we attempted to precisely define further steps necessary for the improvement of the current model. As demonstrated in the example presented in this study, the proposed method can indicate possible areas of improvement and it is useful for the efficient model-building. In our analysis of the medical examination data, the graphical plots obtained for the cumulative residuals suggested the non-linearity of MAP. Sets of graphical plots for the functional form tests, such as those presented in Figure 1 are very informative for the determination of the covariates for which the functional form should be modified. However, since the residuals were aggregated, it is difficult to estimate the type of the non-linear function that should be further applied based on the plots of cumulative residuals. We added a quadratic term for MAP to capture the non-linearity in an ad hoc manner. For the right-censored Weibull AFT model, Lindqvist *et al.* [24] proposed a method for the determination of the functional form of a covariate by smoothing the log of the adjusted Cox-Snell residuals over the covariate values. A similar approach can be employed for the interval-censored observations. In Figure 2, we present the plots of $(Z_i^{(p)}, \log(1-\hat{r}_i^C))$ for Model 1 with a smoothing curve for AGE, BMI, and MAP. The results presented in Figure 2 suggest that a quadratic pattern remains for the MAP residuals. However, the possible insights obtained by using the results presented in Figure 2 may be highly subjective, but coupled with the formal evaluation by using our proposed method, these plots may be useful for the evaluation of model improvements. As another candidate approach, Hashimoto *et al.* [25] proposed the modified deviance residuals for the normal linear regression model.

The proposed method is limited such that the deviation from the true model in terms of parameters, except for the covariate effects, must be negligible, and covariates need to be bounded, which is assumed for the theoretical consideration. However, the condition of the boundedness of covariates may be relaxed, and the influence for real data analysis is relatively small since we deal with finite samples. Actually, the proposed method operate normally in the simulation using covariates sampled from the normal distribution. We will deepen more considerations in next research. Taken together, further research is required for the development and integration of the methods to obtain the clue of model improvements (*e.g.*, [25]), which can then allow a more efficient construction of models.

Appendix

We consider the parametric PH model such as $S(t,Z) = \{S_0(t)\}^{\exp(\beta^{\star\top} Z)}$ where $\beta^\star$ is a vector of regression coefficients and $S_0(t)$ is the baseline survival function. The Weibull AFT and the PE model are included in the parametric PH model. For this model, we confirm that the proposed GoF test statistics operate normally under the suitable conditions. First, we organize the sufficient conditions of observed data L, R and Z for the boundedness of the martingale residuals $r^L(\theta_0)$ and its derivative for $\beta^\star$. By using these boundedness, secondly, we show that $W(z)$ and $\hat{W}(z)$ convergent weakly to the same zero-mean Gaussian distribution.

Conditions for boundedness of $\boldsymbol{r^L(\theta_0)}$ *and* $\frac{\partial}{\partial \beta^\star} \boldsymbol{r^L(\theta_0)}$

Conditions are the followings:

(1) A positive number, η, exists, which satisfies with $P(L - R \geq \eta) = 1$.
(2) The union of the support of L and R is contained in an interval $[\tau_0, \tau_1]$, where $0 < \tau_0 < \tau_1 < \infty$.
(3) (a) The distribution of Z is not concentrated on any proper affine subspace of R^q (i.e. of dimension $q-1$ or smaller). (b) Z is bounded; there exists some positive constant $\mathfrak{K}$ such as $|Z| \leq \mathfrak{K}$.

Theorem 5.1. *Suppose that $S_0(t)$ is a monotone decreasing function such as $S_0(t) = \exp(-a(t))$ where $a(t)$ is a monotone increasing function such as $a(0) = 0$ and $a(\infty) = \infty$. Then, under the conditions (3), $S(t|Z;\theta_0) = \{S_0(t)\}^{\exp(\beta^{\star\top} Z)}$ is strictly positive and bounded continuous function on $t \in [\tau_0, \tau_1]$ such as $0 < \tau_0 < \tau_1 < \infty$.*

Proof. For fixed Z, the following inequality,

$$0 = S_0(\infty) < S_0(\tau_1) < S_0(\tau_0) < S_0(0) = 1$$

holds. While, by the condition (3)

$$|\beta^{\star\top} Z| \leq |\beta^\star||Z| < \infty$$
$$\therefore 0 < \exp(\beta^{\star\top} Z) < \infty.$$

Thus, for all t,

$$S(t|Z;\theta_0) \to 0 \text{ as } \exp(\beta^{\star\top} Z) \to \infty$$

and

$$S(t|Z;\theta_0) \to 1 \text{ as } \exp(\beta^{\star\top} Z) \to 0.$$

Therefore, $S(t|Z,\theta_0) \in (0,1)$ holds, and then, $|\log S(t|Z,\theta_0)| < \infty$ holds. □

Corollary 5.2. *Suppose $S_0(t)$ and Z are same as that of Theorem 5.1. Then, for any two time points t_1 and t_2 such as $t_1 < t_2 \in [\tau_0, \tau_1]$, which implies the condition (1)*

and (2), and $0 < \epsilon < \infty$,

$$\begin{aligned}&\mathsf{P}_{t,Z}\{(S(t_1|Z;\theta_0) - S(t_2|Z;\theta_0)) > \epsilon\} = 1,\\ &\mathsf{P}_{t,Z}\{|S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0) - S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)| < \infty\} = 1\end{aligned}$$

and

$$\mathsf{P}_{t,Z}\{|S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0)^2 - S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)^2| < \infty\} = 1$$

hold.

Proof. The following inequality holds for $t_1 < t_2$ and fixed Z:

$$\begin{aligned}&S_0(t_1) > S_0(t_2)\\ &\Rightarrow \{S_0(t_1)\}^{\exp(\beta^{\star\top}Z)} > \{S_0(t_2)\}^{\exp(\beta^{\star\top}Z)}\\ &\Rightarrow S(t_1|Z;\theta_0) > S(t_2|Z;\theta_0).\end{aligned}$$

Thus, $S(t_1|Z;\theta_0) - S(t_2|Z;\theta_0) > 0$ holds, and $S(t_1|Z;\theta_0) - S(t_2|Z;\theta_0) < \infty$ holds since we obtain the result of boundedness of $S(t_1|Z;\theta_0)$ and $\log S(t_1|Z;\theta_0)$ by Theorem 5.1. Moreover, the following inequality holds:

$$\begin{aligned}&|S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0) - S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)|\\ &< |S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0)| + |S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)|\\ &\le |S(t_1|Z;\theta_0)||\log S(t_1|Z;\theta_0)| + |S(t_2|Z;\theta_0)||\log S(t_2|Z;\theta_0)|\\ &< \infty.\end{aligned}$$

Similarly, the following inequalities holds:

$$\begin{aligned}&|S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0)^2 - S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)^2|\\ &< |S(t_1|Z;\theta_0)\log S(t_1|Z;\theta_0)^2| + |S(t_2|Z;\theta_0)\log S(t_2|Z;\theta_0)^2|\\ &\le |S(t_1|Z;\theta_0)|\log S(t_1|Z;\theta_0)^2 + |S(t_2|Z;\theta_0)|\log S(t_2|Z;\theta_0)^2\\ &< \infty\end{aligned}$$

by Theorem 5.1. □

Now, we can prove the boundedness of the martingale residuals $r^L(\theta_0)$ and its derivative for $\beta^\star$, $\frac{\partial r^L(\theta_0)}{\partial \beta^\star}$. $r^L(\theta_0)$ and $\frac{\partial r^L(\theta_0)}{\partial \beta^\star}$ are provided as

$$\begin{aligned}r^L(\theta_0) =& \frac{S(L|Z;\theta_0)\log S(L|Z;\theta_0) - S(R|Z;\theta_0)\log S(R|Z;\theta_0)}{S(L|Z;\theta_0) - S(R|Z;\theta_0)},\\ \frac{\partial}{\partial\beta^\star}r^L(\theta_0) =& -Z\Big[\frac{\log S(L|Z;\theta_0)S(L|Z;\theta_0)(1+\log S(L|Z;\theta_0))}{S(L|Z;\theta_0) - S(R|Z;\theta_0)}\\ &- \frac{\log S(R|Z;\theta_0)S(R|Z;\theta_0)(1+\log S(R|Z;\theta_0))}{S(L|Z;\theta_0) - S(R|Z;\theta_0)} - r^L(\theta_0)^2\Big].\end{aligned}$$

The $S_0(t)$ of the Weibull AFT and PE model satisfy with the conditions for that of Theorem 5.1. Therefore, under the conditions (1)–(3), the boundedness of martingale

residuals $r^L(\theta_0)$ and its derivative $\frac{\partial}{\partial \beta^\star} r^L(\theta_0)$ are hold by Theorem 5.1 and Corollary 5.2.

Weak convergence of $W(z)$ and $\hat{W}(z)$

We shall demonstrate that $W(z)$ and $\hat{W}(z)$ have the same zero-mean Gaussian distribution in the same line of Lin and Spiekerman [21]. By Taylor series expansion, $W(z) = \tilde{W}(z) + o_p(1)$ uniformly, where

$$\begin{aligned}
\tilde{W}(z) =& \frac{1}{\sqrt{n}}\Big[\sum_{i=1}^{n} \mathsf{I}(Z_i \le z) r_i^L(\theta_0) + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{I}(Z_j \le z) \frac{\partial r_j^L(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} \sqrt{n}(\hat{\theta} - \theta_0)\Big] \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i \le z) r_i^L(\theta_0) \\
&+ \frac{1}{n} \sum_{j=1}^{n} \mathsf{I}(Z_j \le z) \frac{\partial r_j^L(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} \Big(\frac{1}{n} I(\theta_0)\Big)^{-1} \frac{1}{\sqrt{n}} \sum_{l=1}^{n} U_l(\theta_0) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i \le z) r_i^L(\theta_0) + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathsf{I}(Z_j \le z) \frac{\partial r_j^L(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} I(\theta_0)^{-1} \sum_{l=1}^{n} U_l(\theta_0),
\end{aligned}$$

where we use the standard maximum likelihood theory, *i.e.*,

$$\sqrt{n}(\hat{\theta} - \theta_0) = \Big(\frac{1}{n} I(\theta_0)\Big)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(\theta_0) + o_p(1).$$

$\hat{W}(z)$ is obtained from $\tilde{W}(z)$ by multiplying G_i $(i = 1, ..., n)$ and replacing θ_0 with $\hat{\theta}$. Let

$$H_i(z; \theta) = \mathsf{I}(Z_i \le z) r_i^L(\theta) + J(\theta) \mathcal{J}(\theta)^{-1} U_i(\theta),$$

where $J(\theta) = n^{-1} \sum_{i=1}^{n} \mathsf{I}(Z_i \le z) \frac{\partial r_i^L(\theta)}{\partial \theta}$ and $\mathcal{J}(\theta) = n^{-1} I(\theta)$. Then, the covariance function for $\hat{W}(\cdot)$ between z_1 and z_2 is

$$\begin{aligned}
\mathsf{C}[\hat{W}(z_1), \hat{W}(z_2)] &= \mathsf{E}_G\Big[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_i(z_1; \hat{\theta}) G_i \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_i^\top(z_2; \hat{\theta}) G_i\Big] \\
&= \frac{1}{n} \sum_{i=1}^{n} H_i(z_1; \hat{\theta}) H_i^\top(z_2; \hat{\theta}) \mathsf{E}_G[G_i^2] \\
&= \frac{1}{n} \sum_{i=1}^{n} H_i(z_1; \hat{\theta}) H_i^\top(z_2; \hat{\theta}).
\end{aligned}$$

This equation converges almost surely to $\mathsf{E}\big[H_1(z_1; \theta_0) H_1^\top(z_2; \theta_0)\big]$, which is the asymptotic covariance function for $\tilde{W}(z)$, due to the strong consistency of $\hat{\theta}$ and the strong law of large numbers. Following this, the finite-dimensional distributions of $\hat{W}(z)$ and $\tilde{W}(z)$ are both asymptotically normal, according to the Lindeberg-Feller theorem and the Cramér–Wold device.

To show the tightness of $\tilde{W}(z)$, we utilized the application of Bickel and Wichura [27] who developed the convergence criteria of stochastic processes for the function in D_q, which is the uniform closure in the space of all bounded functions of the q-dimensional vector subspace of functions. Define

$$\tilde{W}(z) = \tilde{W}_1(z) + \tilde{W}_2(z),$$

where

$$\tilde{W}_1(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{I}(Z_i \leq z) r_i^L(\theta_0)$$

and

$$\tilde{W}_2(z) = \frac{1}{n} \sum_{i=1}^{n} \mathsf{I}(Z_i \leq z) \frac{\partial r_i^L(\theta)}{\partial \theta}|_{\theta=\theta_0} \mathcal{J}(\theta_0)^{-1} \mathcal{S}(\theta_0),$$

where $\mathcal{S}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(\theta_0)$. We apply Theorem 1 and 3 of [27] to show the tightness of $\tilde{W}_1(z)$ and $\tilde{W}_2(z)$, respectively.

For q-dimensional vector $z_i = \{z_i^{(p)}\}_{p=1,\ldots,q}$ $(i = 1, 2, 3)$ with the order such as $z_1^{(p)} < z_2^{(p)} < z_3^{(p)}$ for each p, the following inequality holds:

$$\begin{aligned}
&\mathsf{E}_z[|\tilde{W}_1(z_2) - \tilde{W}_1(z_1)|^2 |\tilde{W}_1(z_3) - \tilde{W}_1(z_2)|^2] \\
&\quad = \mathsf{E}_z[|n^{-1/2} \sum_{i=1}^{n} \prod_{p=1}^{q} \mathsf{I}(z_1^{(p)} < Z_i^{(p)} < z_2^{(p)}) r_i^L(\theta_0)|^2 \\
&\qquad |n^{-1/2} \sum_{i=1}^{n} \prod_{p=1}^{q} \mathsf{I}(z_2^{(p)} < Z_i^{(p)} < z_3^{(p)}) r_i^L(\theta_0)|^2] \\
&\quad \leq \mathsf{E}_z(\sum_{i=1}^{n} r_i^L(\theta_0)^2)^2 \prod_{p=1}^{q} \mathsf{P}(z_1^{(p)} < Z^{(p)} < z_3^{(p)})^2. \qquad (3)
\end{aligned}$$

The boundedness of $r_i^L(\theta_0)$ holds under the conditions (1)–(3) (See previous section), and $\mathsf{E} r_i^L(\theta_0) = 0$ when the model is specified correctly. Therefore, the left side of (3) is bounded, and its mean is zero. According to Bickel and Wichura [27], define,

$$M_p'' = \sup\{|\tilde{W}_1(z_2^{(p)}) - \tilde{W}_1(z_1^{(p)})| \wedge |\tilde{W}_1(z_3^{(p)}) - \tilde{W}_1(z_2^{(p)})| : z_1^{(p)} < z_2^{(p)} < z_3^{(p)} \in Z^{(p)}\}$$

and $M'' = \max_p M_p''$. By Chebyshev's inequality, (3) implies that

$$\mathsf{P}(M'' \geq \lambda) \leq \mathfrak{k} \lambda^{-4} \mu(Z)^2$$

where $\mu(Z)$ is a finite nonnegative measure on Z which corresponds to $\prod_{p=1}^{q} \mathsf{P}(z_1^{(p)} < Z^{(p)} < z_3^{(p)})$ in (3), and $\mathfrak{k}$ is the bound of $(\sum_{i=1}^{n} r_i^L(\theta_0)^2)^2$ in (3). Therefore, Theorem 1 of [27] holds for $(\gamma, \beta) = (4, 2)$. Moreover, suppose that each $\tilde{W}_1^{(p)}(z)$ vanishes along

the lower boundary of $Z^{(p)}$. Consequently, the tightness of $\tilde{W}_1(z)$ holds from Theorem 3 of [27].

Similarly, for the tightness of $\tilde{W}_2(z)$, the following equation holds:

$$
\begin{aligned}
&\mathsf{E}_z[|\tilde{W}_2(z_2)-\tilde{W}_2(z_1)|^2|\tilde{W}_2(z_3)-\tilde{W}_2(z_2)|^2] \\
&= \mathsf{E}_z[|n^{-1}\sum_{i=1}^{n}\mathsf{I}(z_1^{(p)}<Z_i^{(p)}<z_2^{(p)})\frac{\partial r_i^L(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\mathcal{J}(\theta_0)^{-1}\mathcal{S}(\theta_0)|^2 \\
&\qquad |n^{-1}\sum_{i=1}^{n}\mathsf{I}(z_2^{(p)}<Z_i^{(p)}<z_3^{(p)})\frac{\partial r_i^L(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\mathcal{J}(\theta_0)^{-1}\mathcal{S}(\theta_0)|^2] \\
&\le \mathsf{E}_z\big(n^{-1}\sum_{i=1}^{n}\frac{\partial r_i^L(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}^2\big)^2\big(\mathcal{J}(\theta_0)^{-1}\mathcal{S}(\theta_0)\big)^4\prod_{p=1}^{q}\mathsf{P}(z_1^{(p)}<Z^{(p)}<z_3^{(p)})^2. \qquad (4)
\end{aligned}
$$

With maximum likelihood theory, $\mathsf{E}[\mathcal{J}(\theta_0)^{-1}\mathcal{S}(\theta_0)]\to 0$ as $\hat{\theta}\to\theta_0$, and the boundedness of $\mathcal{J}(\theta_0)^{-1}\mathcal{S}(\theta_0)$ holds. The boundedness of $\frac{\partial r_i^L(\theta)}{\partial\theta}$ holds under the conditions (1)–(3) (See previous section). Therefore, the left side of (4) is bounded, and its mean is zero. Then, the tightness of $\tilde{W}_2(z)$ holds in the same line for $\tilde{W}_1$.

The tightness of $\hat{W}(z)$ has the same argument as that of $\tilde{W}(z)$, since the mean and covariance of $\hat{W}(z)$ is invariant against the operator G_i. For $\hat{W}(z)=\hat{W}_1(z)+\hat{W}_2(z)$, the tightness holds for $\hat{W}_1(z)$ and $\hat{W}_2(z)$, respectively, the same as for $\tilde{W}_1(z)$ and $\tilde{W}_2(z)$. Since the boundedness of $r_i^L(\theta)$, $\frac{\partial r_i^L(\theta)}{\partial\theta}$ and $\mathcal{J}(\hat{\theta})^{-1}\mathcal{S}(\hat{\theta})$ hold in the same line for $\tilde{W}(z)$, and $\mathsf{E}r_i^L(\hat{\theta})\to 0$ as $\hat{\theta}\to\theta_0$ and $\mathsf{E}[\mathcal{J}(\hat{\theta})^{-1}\mathcal{S}(\hat{\theta})]=0$ for $\hat{W}_1(z)$ and $\hat{W}_2(z)$, respectively. Therefore, the conditionals on data, $\hat{W}(z)$ have the same zero-mean Gaussian distribution as $\tilde{W}(z)$ asymptotically.

References

[1] Cox, D. R., 1972. Regression models and life tables. *Journal of the Royal Statistical Society*, Series B, 34, 187-220.

[2] Lawless, J. F., 2011. *Statistical Models and Methods for Lifetime Data (Second Edition).* John Wiley & Sons.

[3] Zhang, Y., Hua, L., & Huang, J., 2010. A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37, 338-354.

[4] Huang, J., & Wellner, J. A., 1997. Interval censored survival data: a review of recent progress. *In Proceedings of the First Seattle Symposium in Biostatistics*. Springer, New York, NY. pp. 123-169

[5] Sun, J., 2007. *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer Science and Business Media.

[6] Tian, L. and Cai, T. 2006. On the accelerated failure time model for current status and interval censored data. *Biometrika*, 93, 329-342.

[7] Zeng, D., Cai, J. & Shen, Y. 2006. Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, 16, 287-302.

[8] Zeng, D., Mao, L., & Lin, D. Y., 2016. Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2), 253-271.

[9] Collett, D., 2014. *Modelling Survival Data in Medical Research (3rd edition).* Chapman and Hall, London.

[10] Ren, J. J., 2003. Goodness of fit tests with interval censored data. *Scandinavian Journal of Statistics*, 30, 211-226.
[11] Pan, W., 2000. A two-sample test with interval censored data via multiple imputation. *Statistics in Medicine*, 19(1), 1-11.
[12] Li, J., & Ma, S., 2010. Interval-censored data with repeated measurements and a cured subgroup. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(4), 693-705.
[13] Nysen, R., Aerts, M., & Faes, C., 2012. Testing goodness of fit of parametric models for censored data. *Statistics in Medicine*, 31(21), 2374-2385.
[14] Hudgens, M. G., Chenxi Li, C., & Fine, J. P., 2014. Parametric likelihood inference for interval censored competing risks data. *Biometrics*, 70(1), 1-9.
[15] Sparling, Y. H., Younes, N., Lachin, J. M., & Bautista, O.M., 2006. Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7, 599-614.
[16] Cox, D. R., & Snell, E. J., 1968. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248-275.
[17] Barlow, W. E., & Prentice, R. L., 1988. Residuals for relative risk regression. *Biometrika*, 75, 65-74.
[18] Fleming, T.R., & Harrington, D., 1991. *Counting Processes and Survival Analysis.* John Wiley & Son Inc., New York.
[19] Therneau, T. M., Grambsch, P. M., & Fleming, T. R., 1990. Martingale-based residuals for survival models. *Biometrika*, 77, 147-160.
[20] Farrington, C. P., 2000. Residuals for proportional hazards models with interval-censored survival data. *Biometrics*, 56(2), 473-482.
[21] Lin, D. Y., & Spiekerman, C. F., 1996. Model checking techniques for parametric regression with censored data. *Scandinavian Journal of Statistics*, 23, 157-177.
[22] Lin, D. Y., Wei, L. J., & Ying, Z., 1993. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80, 557-572.
[23] Lindsey, J. C., & Ryan, L. M., 1998. Methods for interval-censored data. *Statistics in Medicine*, 17, 219-238.
[24] Lindqvist, B. H., Kvaløy, J. T., & Aaserud, S., 2015. Residual plots to reveal the functional form for covariates in parametric accelerated failure time models. *Lifetime Data Analysis*, 21(3), 353-378.
[25] Hashimoto, E. M., Ortega, E. M., Cancho, V. G., & Cordeiro, G. M., 2010. The log-exponentiated Weibull regression model for interval-censored data. *Computational Statistics and Data Analysis*, 54(4), 1017-1035.
[26] Hastie, T. J., & Tibshirani, R. J., 1990. *Generalized Additive Models (Vol. 43)*. CRC Press.
[27] Bickel, P. J., & Wichura, M. J., 1971. Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, 42(5), 1656-1670.