International Journal of Scientific & Engineering Research Volume 9, Issue 4, April-2018
ISSN 2229-5518

Inference on quantile residual life with left-truncated and right-censored data

FENG Hailin · GAO Wei · MA jiulong ·
School of mathematics and statistics, Xidian University, Xi'an, Shaan xi, China

Inference on quantile residual life with left-truncated and right-censored data

FENG Hailin · GAO Wei · MA jiulong ·
School of mathematics and statistics, Xidian University, Xi'an, Shaan xi, China

Abstract The left-truncated and right-censored data is a kind of common important test data in the field of biomedical and engineering reliability research. A quantile residual life prediction model is established for the left-truncated and right-censored in this paper. In the prediction model, with the aid of Cox model as an auxiliary mode to estimate survival function, meanwhile the unknown parameter estimation method is proposed by combining with the characteristics of the left truncated data. Furthermore, the asymptotic properties for the proposed estimator are derived. Simulation studies are performed to demonstrate that the proposed estimator work well in finite-sample situations. Finally, an example analysis of Channing House data sets is given.

Keywords left-truncated , right-censored , residual life , quantile , Cox model

1 Introduction

The concept of quantile residual life proposed by Since Haines and Singpurwalla in 1974 [1], as an alternative to the mean residual life, has been extensively studied. The quantile residual life can provide a more comprehensive description on the residual life than the mean residual life. Especially, the mean residual life is sensitive and even incalculable for life distribution with skewed or asymmetric, however the quantile residual life is well in the case [1]. So the quantile residual life research gets more and more attention. As far as the authors know, The representative studies of quantile residual life prediction just for the right-censored data include Jeong et al(2008) [2], Jung S H et al(2009) [3], Wang H J et al(2012) [4] and Lin C et al(2015) [5].

However, in many practical trials, such as the HIV/AIDS history study [6], the Alzheimer disease study [7], the collected data is not only the right-censored, but also the left-truncated. In addition, the Channing House data, the Canadian Study of Healthy and Aging dementia data, and the Spain unemployment data are also the right-censored and left-truncated. Right-censored data is a common type of data that cannot be observed due to human or other reasons during the observation and experimentation. Left truncated data refers to the period from the occurrence of the event to the time when the event is recorded. For example, from patients getting sick, only those who have lived for a period of time and received treatment are recorded. If the patient has died before treatment, one will not be recorded, and the data is the left-truncated data. Left-truncated and right-censored data often appear in real situations such as prevalent cohort studies, cancer screening, and labor economics. However, there are little literature that is the study of the problem of the quantile residual life under the left-truncated and right-censored data, so the quantile residual life study for this kind of data is necessary and very urgent.

In this paper, we build a quantile residual life prediction model for the left-truncated and right-censored, and propose a parameter estimation method. The method uses Cox model as an auxiliary model to estimate the survival function, using the characteristics of the left-truncated data to estimated unknown parameters. We derive the asymptotic properties for the proposed estimator. Numerical simulations and example analysis verify the accuracy and effectiveness of the proposed method.

The paper is arranged as follows. The second section introduces the symbols to be used. In the third part, the quantile residual life prediction model based on left-truncated and right-censored data is proposed, and the correlation analysis is given. In the fourth section, the asymptotic of the proposed estimators is derived. In the fifth section, the finite sample properties of the proposed estimators are verified by simulation studies. The sixth section gives an example analysis. The seventh section summarizes the article.

2 Notations

As shown in Fig.1, $\widetilde{T}$ is all the survival time variables with distribution function of F on $[0,\tau)$. $\widetilde{A}$ is the truncation variable that is subject to uniform distribution on$[0,\kappa]$. $\widetilde{A}$ and $\widetilde{T}$ are independent of each other.

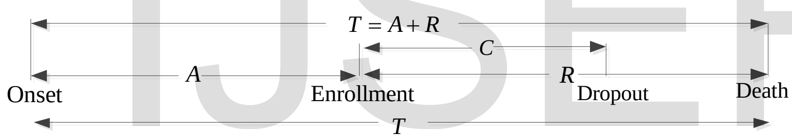

Figure 1 truncation and censoring of time

$S(\bullet) = 1 - F(\bullet)$ is the survival function of the unbiased data. In the presence of the left-truncation sample, only if the $\widetilde{T} \geq \widetilde{A}$, the individual can be observed. T is the failure time observed and A the truncation time observed. So the joint distribution (T, A) is similar to that $\left(\widetilde{T}, \widetilde{A}\right)\Big|\widetilde{T} \geq \widetilde{A}$. Let $R = T - A$, where R is the residual life. $A + C$ is the censored time, where C is the period from the recorded to the censored. $Z = \left(1, \widetilde{Z}^T\right)$ is a $1 \times p$ vector indicating the covariates. Assume that C and (A, R) are independent of each other at the given Z . In the case, we can observe $(Y_i, A_i, \delta_i, Z_i), i = 1, ..., n$, where $Y_i = \min(A_i + R_i, A_i + C_i)$, $\delta_i = I(R_i \leq C_i)$.

Let f be the unbiased density function of T given covariate $Z = z$, so the biased density function of the left-truncation data T is

$g(t|z) = tf(t|z)/\mu(z)$, $\mu(z) = \int_0^\infty uf(u|z)du$

where $f(t|z)$ is unbiased density, $g(t|z)$ is biased density, $\mu(z) = \int_0^\infty S(t|Z=z)dz$, and $\mu(z) < \infty$ when covariate z is given.

3 Prediction model on quantile residual life with left-truncated and censored data

Usually, at the time t_0 ,the $\tau-$ quantile residual life is defined as

$$\theta_\tau(t_0|Z) = \tau - quantile(T - t_0|T \geq t_0, Z). \tag{1}$$

We can obtain formula (2) from the above formula.

$$P\{T - t_0 \geq \theta_\tau(t_0|Z)|T \geq t_0, Z\} = \tau. \tag{2}$$

Formula (2) means

$$P\{T - t_0 \geq \theta_\tau(t_0|Z)|Z\} = \tau P\{T \geq t_0|Z\}. \tag{3}$$

Let the conditional survival function of T be $S(t|Z) = P(T \geq t|Z)$, Note that $\theta_\tau(t_0|Z)$ is not uniquely determined by $S(t|Z) = P(T \geq t|Z)$. In practice, it is first necessary to build the model $S(t|Z)$. and then infer the value of θ_τ at a fixed time point t_0 under a given covariates Z . For the sample of the survival time over t_0 (namely $(T - t_0|T \geq t_0, Z)$), the condition residual life for survival function is

$$S(t|t_0, Z) = S(t + t_0|Z)/S(t_0|Z). \tag{4}$$

Further, there is

$$S(t + \theta_\tau(t|Z)|Z) = \tau S(t|Z). \tag{5}$$

Suppose $\theta_\tau(t, Z)$ is the only solution of equation (5). We estimate the conditional quantile residual life function by solving equation (6) at time t_0.

$$\widehat{U}(\theta_\tau(t_0, Z)) = \widehat{S}(t_0 + \theta_\tau(t_0, Z)|Z) - \tau\widehat{S}(t_0|Z) = 0. \tag{6}$$

$\widehat{S}(|Z)$ is the consistent estimation of $S(|Z)$. Note that if equation (5) has the solution, we consider the definition $\theta_\tau(t, Z) = \inf\{\theta_\tau(t_0, Z) : S(t_0 + \theta_\tau(t_0|Z)|Z) \leq \tau S(t_0|Z)\}$, this is, $S^{-1}(\tau S(t_0|Z)|Z) - t_0$, where $S^{-1}(\theta_\tau|Z) = \inf\{t : S(t|Z) \leq \tau\}$. Further, let $\widehat{\theta}_\tau(t_0, Z) = \widehat{S}^{-1}\left(\tau\widehat{S}(t_0|Z)|Z\right) - t_0$ be the solution of equation (6).

Now an important question is the estimator of $S(t|Z)$ .Because one of the main focuses of the application of the Cox proportional hazards model is to analysis left-truncated data. So we consider the Cox proportional hazards model as an auxiliary model

$$\Lambda(t|Z) = \Lambda_0(t)\exp\left(\beta^T Z\right). \tag{7}$$

Where $\Lambda_0(t)$ is the unknown baseline risk function, and β is the unknown regression coefficient vector of Z .The conditional survival function $S(t|Z)$ can be estimated by formula (8),

$$\widehat{S}(t|Z) = \exp\left\{-\widehat{\Lambda}(t|Z)\right\} = \exp\left\{-\exp\left\{\widehat{\beta}^T Z\right\}\widehat{\Lambda}_0(t)\right\}. \tag{8}$$

$\widehat{\beta}$ is the estimation of β ,and $\widehat{\Lambda}_0(t)$ is estimation of $\Lambda_0(t)$ by Breslow (1972) [8].

Now, we give an estimation method for parameter β .Considering the data type of the left truncation, we need to combine the characteristics of the left truncated data[9, 10],then the joint density function (A, T) is expressed as

$$f_{A,T}(a, t|z) = f_A(a|z) f_{T|A}(t|a, z) = [S(a|z)/\mu(z)] [f(t|z\, I(t > a))/S(a|z)]. \tag{9}$$

Given the truncation time $A = a$,the conditional likelihood function of Y is

$$L_C(\beta) = \prod_{i=1}^n \left[f(Y_i|Z_i, \beta)^{\delta_i} S(Y_i|Z_i, \beta)^{1-\delta_i} \Big/ S(A_i|Z_i, \beta) \right]. \tag{10}$$

L_C can be further expressed as

$$L_C(\beta, \Lambda_0) = L_P(\beta) L_M(\beta, \Lambda_0). \tag{11}$$

where

$$L_P(\beta) = \prod_i \left[\exp\left(\beta^T Z_i\right) \Big/ \sum_{j \in R(Y_i)} \exp\left(\beta^T Z_j\right) \right]^{\delta_i}. \tag{12}$$

L_M is the marginal likelihood of the truncation time A ,denoted as

$$\begin{aligned} L_M\left(\beta, \widehat{\Lambda}_\beta\right) &= \prod_{i=1}^n \left[S(A_i|Z_i)/\mu_{\beta,\Lambda}(Z_i)\right] \\ &= \prod_{i=1}^n \left[\exp\left\{-\Lambda(A_i)\exp\left(\beta^T Z_i\right)\right\} \Big/ \int_0^\infty \exp\left\{-\Lambda(u)\exp\left(\beta^T Z_i\right)\right\} du\right]. \end{aligned} \tag{13}$$

Further, the estimation equation of $\widehat{\beta}$ can be obtained

$$\begin{aligned} &M_P(\beta) = \sum_{i=1}^n \delta_i \{Z_i - \\ &\left[\sum_{j=1}^n Z_j \exp\left(\beta^T Z_j\right) I(Y_j \geqslant Y_i \geqslant A_j) \Big/ \sum_{j=1}^n \exp\left(\beta^T Z_j\right) I(Y_j \geqslant Y_i \geqslant A_j)\right]\Big\} = 0. \end{aligned} \tag{14}$$

Wang et al (1993)[1]proved that the partial likelihood estimator obtained from the solution of M_P is approximately as efficient as the maximize of L_C . $\widehat{\Lambda}_0$ can be obtained from the following formula

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^n \left[I(Y_i \le t)\,\delta_i \Big/ \sum_{j=1}^n Q_j(Y_i)\, e^{\widehat{\beta}^T Z_j(Y_i)} \right]. \tag{15}$$

where $Q_i(t) = I(Y_i \ge t)$. Therefore, the estimator of $\widehat{\theta}_\tau(t_0, Z)$ can be obtained by solving the equation

$$\widehat{U}(\theta_\tau(t_0, Z)) = \widehat{S}(t_0 + \theta_\tau(t_0, Z)|Z) - \tau \widehat{S}(t_0|Z) = 0. \tag{16}$$

4 Asymptotic properties

In order to prove the consistency of $\widehat{\theta}_\tau(t_0, Z)$,that is, $\widehat{\theta}_\tau(t_0, Z) \to \theta_\tau(t_0, Z)$, we need to use the strong consistency of $\widehat{\beta}$ and $\widehat{\Lambda}_0$ .For the proof for the asymptotic normality of $\widehat{\theta}_\tau(t_0, Z)$, we first need to apply the asymptotic theorem proposed by Andersen and Gill(1982)[11]to obtain $\sqrt{n}\{\widehat{\Lambda}(\cdot)-\Lambda(\cdot)\}$ convergence weakly to a zero-mean Gaussian process. Then, through the functional delta-method and some standard counting process techniques, we can get the results.

In order to prove the results, we need the following assumptions and regularity conditions:

(A1) The covariate $Z(\cdot)$ has uniformly bounded total variation.

(A2) $\beta_0 \in \mathrm{B} \subset R^P$,and B is open, convex and bounded.

(A3) $\alpha = \sup\{t : Y(t) > 0\}$, $\Lambda_0(t)$ is continuous and $\Lambda_0(\alpha) < \infty$.

(A4) Ω is positive definite, which is the asymptotic matrix of $\sqrt{n}\left(\widehat{\beta}-\beta_0\right)$.

We have the following consistency and asymptotic normality.

Theorem 4.1 *When the above condition holds and when $n \to \infty$, $\theta_\tau(t_0, Z)$ is the only solution of equation (5) for given covariate Z and a specific time point t_0 ,that is,*

$$\widehat{\theta}_\tau(t_0, Z) \xrightarrow{P} \theta_\tau(t_0, Z).$$

Theorem 4.2 *When the above conditions are satisfied, given the covariate Z and the specific time point t_0 ,when $n \to \infty$,that is,*

$$\sqrt{n}\left\{\widehat{\theta}_\tau(t_0, Z) - \theta_\tau(t_0, Z)\right\} \xrightarrow{\ell} N\left(0, \sigma^2\right).$$

where $\xrightarrow{\ell}$ means convergence in distribution. And $\sigma^2 = \left[\alpha^2 S(t_0 \,|Z)^2 \big/ g(t_0+\theta_\tau \,|Z)^2\right]\sum$

where

$\sum = \int_{t_0}^{t_0+\theta_\tau}\left[\exp\left(2\beta_0^T Z(u)\right)/s^{(0)}(\beta_0, u)\right]d\Lambda_0(u) + \left(h(t_0+\theta_\tau) - h(t_0)\right)^T \Omega^{-1}\left(h(t_0+\theta_\tau) - h(t_0)\right)$

$\Omega = \int_0^\infty \left\{\left[s^{(2)}(\beta_0, t)/s^{(0)}(\beta_0, t)\right] - \overline{z}(\beta_0, t)^{\otimes 2}\right\} s^{(0)}(\beta_0, t)\, d\Lambda_0(t)$

The proof of theorem 1 and theorem 2 can be found in Lin C (2015)[12].

5 Numerical studies

Before we do numerical simulation, we produce left-truncated and right-censored data. First, we generate independent pairs $\left(\widetilde{A}, \widetilde{T}\right)$,where $\widetilde{T}$ is the survival time obeying the model $\Lambda(t) = \Lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2)$, where $\widetilde{A}$ obeys the uniform distribution $U(0, a)$, and we choose different a to get different truncation probabilities. Then, we select n pairs that match the condition $\widetilde{T} \geq \widetilde{A}$ .Censored variable C obeys $U(0, c)$,where c is used to control the censoring rate. The censoring indicator function is $I(A_i + R_i \leq A_i + C_i)$,where $A_i + C_i$ is the total censored time.

In the simulation, we set $\beta_1 = 0.5$, $\beta_2 = 1$, and consider two baseline risk functions $\Lambda_0(t) = t$, $\Lambda_0(t) = 0.5t^2$,and covariate Z_1 obeys $N(1, 1)$,Z_2 obeys $B(0.5)$ .Further, when the truncation rate (a%) is 5%,10% and the censored rate (c%)is 15%,30% and the quantile is 0.3,0.5,0.7 with sample size $n = 200$, repeating 500 times to get independent estimators,we calculate the estimated value of $\widehat{\theta}_\tau(t_0, Z)$ under the different baseline risk functions at the time $t_0 = 0.5, 1$ respectively.

In order to verify that ignoring the left-truncation will cause problems, we have carried

Table 1 $\Lambda_0(t) = t$

a%	c%	t_0	τ	true	method	estimator	Bias	SE	SD	CP	p-value
5%	15%	0.5	0.3	0.6469	proposed	0.6465	0.0004	0.0249	0.0248	96.68	0.0048
					K-M	0.6821	-0.0352	0.0459	0.0436	94.40	0.0805
			0.5	0.4908	proposed	0.4983	-0.0075	0.0302	0.0184	97.17	0.0125
					K-M	0.4921	-0.0013	0.0459	0.0436	94.40	0.0815
			0.7	0.3521	proposed	0.3507	0.0014	0.0607	0.0473	94.04	0.0004
					K-M	0.3595	-0.0074	0.0778	0.0771	92.89	0.0375
		1	0.3	0.3176	proposed	0.3165	0.0011	0.0353	0.0335	98.22	0.0001
					K-M	0.3195	-0.0019	0.0495	0.0482	97.49	0.0549
			0.5	0.2006	proposed	0.2048	-0.0042	0.0874	0.0829	98.58	0.0053
					K-M	0.2095	-0.0089	0.0941	0.0947	96.58	0.0709
			0.7	0.1115	proposed	0.1045	0.0070	0.0162	0.0102	94.79	0.0017
					K-M	0.1049	0.0066	0.0772	0.0733	93.95	0.0112
	30%	0.5	0.3	0.6469	proposed	0.6432	0.0037	0.0710	0.0622	92.14	0.0044
					K-M	0.6902	-0.0433	0.0998	0.0844	91.06	0.0165
			0.5	0.4908	proposed	0.5100	-0.0192	0.0156	0.0145	97.76	0.0033
					K-M	0.5692	-0.0784	0.0309	0.0242	92.23	0.0661
			0.7	0.3521	proposed	0.3496	0.0025	0.0487	0.0485	96.83	0.0012
					K-M	0.3868	-0.0347	0.0681	0.0669	93.62	0.0208
		1	0.3	0.3176	proposed	0.3087	0.0089	0.0352	0.0231	97.96	0.0058
					K-M	0.3213	-0.0037	0.0580	0.0499	96.98	0.0235
			0.5	0.2006	proposed	0.2065	-0.0059	0.0326	0.0258	96.15	0.0015
					K-M	0.2104	-0.0098	0.0615	0.0533	95.26	0.0045
			0.7	0.1115	proposed	0.1234	-0.0119	0.1089	0.0033	92..60	0.0042
					K-M	0.1720	-0.0605	0.0166	0.0158	92.25	0.0222
10%	15%	0.5	0.3	0.6469	proposed	0.6468	0.0001	0.0197	0.0110	90.41	0.0014
					K-M	0.6529	-0.0060	0.0385	0.0211	89.79	0.0079
			0.5	0.4908	proposed	0.4906	0.0002	0.0630	0.0585	95.27	0.0092
					K-M	0.5329	-0.0421	0.0853	0.0617	93.36	0.0493
			0.7	0.3521	proposed	0.3301	0.0220	0.0961	0.0912	98.41	0.0001
					K-M	0.3942	-0.0421	0.1228	0.1165	95.64	0.0627
		1	0.3	0.3176	proposed	0.2985	0.0191	0.0288	0.0222	94.55	0.0046
					K-M	0.3425	-0.0249	0.0422	0.0984	90.30	0.0060
			0.5	0.2006	proposed	0.1975	0.0031	0.0073	0.0077	97.43	0.0025
					K-M	0.2286	-0.0280	0.0090	0.0088	92.60	0.0051
			0.7	0.1115	proposed	0.1090	0.0025	0.0075	0.0060	95.06	0.0081
					K-M	0.1239	-0.0124	0.0082	0.0076	93.74	0.0093
	30%	0.5	0.3	0.6469	proposed	0.6429	0.0040	0.0554	0.0423	98.56	0.0060
					K-M	0.6616	-0.0147	0.0621	0.0517	95.12	0.0138
			0.5	0.4908	proposed	0.5053	-0.0145	0.0394	0.0256	95.64	0.0062
					K-M	0.5057	-0.0149	0.0496	0.0399	95.43	0.0094
			0.7	0.3521	proposed	0.3564	-0.0043	0.0447	0.0409	95.37	0.0053
					K-M	0.3578	-0.0057	0.0772	0.0691	95.32	0.0086
		1	0.3	0.3176	proposed	0.3282	-0.0106	0.0882	0.0785	97.90	0.0099
					K-M	0.3394	-0.0218	0.0970	0.0869	95.02	0.0186
			0.5	0.2006	proposed	0.1900	0.0106	0.0760	0.0659	97.45	0.0058
					K-M	0.2500	-0.0494	0.0822	0.0764	92.96	0.0056
			0.7	0.1115	proposed	0.1249	-0.0134	0.0887	0.0836	91.74	0.0053
					K-M	0.1773	-0.0658	0.0967	0.0917	91.86	0.0081

Table 2 $\Lambda_0(t) = 0.5t^2$

a%	c%	t_0	τ	true	method	estimator	Bias	SE	SD	CP	p-value
5%	15%	0.5	0.3	0.6134	proposed	0.6054	0.0080	0.0048	0.0017	93.05	0.0023
					K-M	0.6379	-0.0245	0.0064	0.0054	92.52	0.0054
			0.5	0.4464	proposed	0.4455	0.0009	0.0197	0.0182	92.45	0.0027
					K-M	0.4697	-0.0233	0.0284	0.0259	88.20	0.0211
			0.7	0.2920	proposed	0.2991	-0.0071	0.0506	0.0429	99.79	0.0014
					K-M	0.3245	-0.0325	0.0799	0.0706	97.63	0.0023
		1	0.3	0.3114	proposed	0.3118	-0.0004	0.0236	0.0147	90.60	0.0013
					K-M	0.3154	-0.0040	0.0736	0.0647	86.99	0.0145
			0.5	0.1988	proposed	0.2128	-0.0140	0.0184	0.0158	92.59	0.0026
					K-M	0.2621	-0.0633	0.0481	0.0405	86.27	0.0041
			0.7	0.1111	proposed	0.1009	0.0102	0.0822	0.0780	91.93	0.0012
					K-M	0.1283	-0.0172	0.1020	0.0968	85.61	0.0031
	30%	0.5	0.3	0.6134	proposed	0.6359	-0.0225	0.0179	0.0116	94.00	0.2971
					K-M	0.6517	-0.0383	0.0364	0.0354	97.34	0.0002
			0.5	0.4464	proposed	0.4368	0.0096	0.0614	0.0577	94.63	0.0048
					K-M	0.4914	-0.0450	0.0715	0.0657	90.59	0.0140
			0.7	0.2920	proposed	0.3060	-0.0140	0.0165	0.0154	91.28	0.0045
					K-M	0.3302	-0.0382	0.0467	0.0392	88.73	0.0585
		1	0.3	0.3114	proposed	0.2993	0.0121	0.0875	0.0830	87.81	0.0059
					K-M	0.3343	-0.0229	0.0499	0.0496	98.20	0.0016
			0.5	0.1988	proposed	0.1681	0.0307	0.0719	0.0682	92.15	0.0045
					K-M	0.2509	-0.0521	0.0480	0.0435	94.47	0.0054
			0.7	0.1111	proposed	0.1114	-0.0003	0.0211	0.0185	92.77	0.0024
					K-M	0.1329	-0.0218	0.0363	0.0243	88.66	0.0035
10%	15%	0.5	0.3	0.6134	proposed	0.5922	0.0212	0.0032	0.0027	95.27	0.0401
					K-M	0.6412	-0.0278	0.0085	0.0056	89.24	0.0987
			0.5	0.4464	proposed	0.4423	0.0041	0.0029	0.0010	94.99	0.0008
					K-M	0.4684	-0.0220	0.0062	0.0028	92.98	0.0112
			0.7	0.2920	proposed	0.2822	0.0098	0.0106	0.0103	95.53	0.0027
					K-M	0.3292	-0.0372	0.0360	0.0291	93.00	0.0600
		1	0.3	0.3114	proposed	0.3242	-0.0128	0.0294	0.0207	93.50	0.0030
					K-M	0.3259	-0.0145	0.0338	0.0334	92.19	0.0055
			0.5	0.1988	proposed	0.2018	-0.0030	0.0483	0.0407	95.95	0.0047
					K-M	0.2075	-0.0087	0.0769	0.0678	93.67	0.0081
			0.7	0.1111	proposed	0.1115	-0.0004	0.0204	0.0142	95.90	0.0061
					K-M	0.1174	-0.0063	0.0242	0.0183	89.80	0.0130
	30%	0.5	0.3	0.6134	proposed	0.6169	-0.0035	0.0247	0.0193	96.50	0.0015
					K-M	0.6183	-0.0049	0.0349	0.0331	93.92	0.0060
			0.5	0.4464	proposed	0.4843	-0.0379	0.0185	0.0073	95.61	0.0088
					K-M	0.4852	-0.0388	0.0555	0.0424	88.18	0.0689
			0.7	0.2920	proposed	0.2853	0.0067	0.0265	0.0200	88.90	0.0052
					K-M	0.3219	-0.0299	0.0608	0.0576	85.44	0.0094
		1	0.3	0.3114	proposed	0.2933	0.0181	0.0219	0.6194	95.06	0.0090
					K-M	0.3390	-0.0276	0.0652	0.0657	88.67	0.0116
			0.5	0.1988	proposed	0.1841	0.0147	0.0212	0.0139	98.83	0.0092
					K-M	0.3285	-0.1297	0.0450	0.0366	92.73	0.0132
			0.7	0.1111	proposed	0.1011	0.0100	0.0138	0.0123	96.72	0.0032
					K-M	0.1768	-0.0657	0.0239	0.0176	90.30	0.0187

out a comparison simulation experiment. In the simulation, we neglect the left-truncation case, and regard the data as the standard right-censored data. We apply popular of Kaplan-Meier method to estimate the survival function, i.e. let $\widetilde{S}(t|Z)$ be the Kaplan-Meier estimation, and $\widetilde{\theta}_\tau(t_0, Z)$ be the corresponding estimations result.

The simulation results show the parameter estimation in terms of empirical bias (Bias), empirical standard deviation (SD), standard errors (SE), the coverage probabilities of the 95% Wald confidence intervals (CP) and p-value. Table 1 and table 2 show:

(1) The empirical bias obtained by the proposed method is smaller;

(2) The empirical standard deviation and standard errors are close to the proposed method;

(3) The confidence intervals have a more reasonable coverage for estimator by the proposed method;

(4) The p-value obtained by the proposed method is smaller.

These conclusions show that the estimators obtained by the proposed method are more accurate and effective than the Kaplan-Meier estimation method.

6 Real data example

In this section, the Channing House data is used to evaluate the proposed method. Channing House is an American retired Center located in California City, Alto, and Palo. Channing House data is collected and recorded from 1964 to July 1, 1975. During this period, a total of 97 men and 365 women lived in the center. In addition, the age at which all members enter and leave the retirement center is also recorded. According to the record, it is found that the data set belongs to the right-censored data type, because many members were still alive when the recording was finished. Only 46 men and 130 women died during the study at the Channing House retirement center, resulting in a censored rate of about 61.9%. When the members died before 1965, they were not recorded, so the data was left truncated. Further studies have shown that if we choose to enter the center older than 786 months (a total of 448 individuals, including 94males and 354 females), the sub-sample data type is left truncated [13].

This article is interested in the effect of gender differences on the survival time of members. With the representation $w = 1$ of male individuals, $w = 0$ indicating the individual females, the prediction value of the quantile residual life is estimated at $t_0 = 800, 900, 1000$ with $\tau = 0.3, 0.5, 0.7$.

Table 3, table 4 and table 5 show the male data, the female data ,the overall data, corresponding parameter estimation in terms of empirical standard deviation(SD), standard errors(SE), the coverage probabilities of the 95% Wald confidence intervals(CP)and p-value, respectively. Comparison of simulation results of table 3 and table 4 show that women have a longer survival time than men. In the case of neglecting the left-truncation, the prediction results are larger. In addition, table 5 shows:

(1) The empirical standard deviation (SD) and standard errors (SE) are smaller and close of the estimations obtained by the proposed method;

(2) The confidence intervals have a more reasonable coverage for estimator by the proposed

method;

(3) The p-value obtained by the proposed method is smaller.

These conclusions show that the estimators obtained by the proposed method are more accurate and effective than the Kaplan-Meier estimation method. The left-truncated data type has some influence on the prediction of quantile residual life, which is consistent with our expected results.

Table 3 Male data analysis

t_0	τ	method	estimator	SE	SD	CP	p-value
800	0.3	proposed	260.7788	8.180	8.085	87.78	0.0020
		K-M	275.5319	9.705	9.615	85.31	0.0180
	0.5	proposed	209.9760	9.520	9.409	97.60	0.0957
		K-M	217.4936	10.78	10.30	93.68	0.2430
	0.7	proposed	99.7198	10.75	10.62	97.19	0.0137
		K-M	106.1145	12.39	12.25	91.14	0.0821
900	0.3	proposed	123.2782	8.326	8.294	98.23	0.0817
		K-M	139.5018	11.29	11.04	95.01	0.1089
	0.5	proposed	119.0039	9.663	9.644	90.03	0.0901
		K-M	149.0451	10.96	10.72	89.04	0.1100
	0.7	proposed	66.1744	10.36	10.24	94.47	0.0567
		K-M	84.4169	17.68	17.47	89.41	0.0887
1000	0.3	proposed	141.1365	7.321	7.300	91.13	0.0065
		K-M	159.2402	9.364	9.354	82.40	0.0308
	0.5	proposed	66.0061	8.944	8.918	91.36	0.0070
		K-M	78.3301	9.020	8.998	83.30	0.0115
	0.7	proposed	34.0132	10.40	10.37	84.01	0.0145
		K-M	42.5072	15.71	15.64	72.25	0.0337

Table 4 Female data analysis

t_0	τ	method	estimator	SE	SD	CP	p-value
800	0.3	proposed	287.7880	8.180	80.85	98.86	0.0100
		K-M	300.5319	9.705	96.15	91.13	0.0310
	0.5	proposed	210.9760	9.520	94.09	89.76	0.0137
		K-M	221.4936	10.78	10.03	93.68	0.0424
	0.7	proposed	169.7198	10.75	10.62	97.19	0.0045
		K-M	176.1145	12.39	12.25	86.11	0.0173
900	0.3	proposed	172.2782	13.26	12.99	98.23	0.0071
		K-M	179.5018	14.29	14.04	95.01	0.0580
	0.5	proposed	149.0039	16.63	16.44	90.03	0.0084
		K-M	159.0450	20.96	20.72	89.04	0.0158
	0.7	proposed	96.1744	10.36	10.24	94.47	0.0090
		K-M	104.4416	17.68	17.43	89.44	0.0239
1000	0.3	proposed	168.8850	11.31	11.11	88.85	0.0120
		K-M	188.6832	13.68	12.54	86.83	0.0950
	0.5	proposed	85.9315	11.77	11.51	93.15	0.0100
		K-M	98.7777	17.30	17.10	77.77	0.0321
	0.7	proposed	59.5464	12.35	12.15	95.46	0.0063
		K-M	64.6024	16.32	16.13	80.24	0.0579

Table 5 Overall data analysis

t_0	τ	method	estimator	SE	SD	CP	p-value
800	0.3	proposed	285.9151	8.625	8.620	91.51	0.0010
		K-M	293.2713	9.187	9.178	82.71	0.0430
	0.5	proposed	208.5374	7.225	7.122	94.50	0.0017
		K-M	216.3750	9.375	9.366	83.75	0.0239
	0.7	proposed	153.4204	7.520	7.349	94.20	0.0038
		K-M	167.7856	10.57	10.09	88.56	0.0137
900	0.3	proposed	169.7023	6.847	6.327	97.02	0.0008
		K-M	183.6637	11.79	11.46	86.37	0.0066
	0.5	proposed	132.2268	6.353	6.241	90.00	0.0006
		K-M	156.2966	8.620	8.460	86.64	0.0307
	0.7	proposed	85.8328	7.327	7.214	93.28	0.0024
		K-M	94.0642	8.205	8.192	84.27	0.0164
1000	0.3	proposed	151.1365	7.321	7.300	91.13	0.0082
		K-M	179.2402	9.364	9.154	82.40	0.0137
	0.5	proposed	76.6006	8.944	8.918	90.61	0.0010
		K-M	88.3301	9.020	8.993	83.33	0.0767
	0.7	proposed	44.0132	10.40	10.37	84.01	0.0078
		K-M	62.5072	11.72	11.64	72.50	0.0253

7 Concluding remarks

In this paper, the quantile residual life prediction model is established under the left-truncated and right-censored data. Estimator of the model parameters is obtained by the estimation algorithm with utilizing characteristics of the left-truncated data type. The asymptotic of the proposed estimators are derived. Finally, we conduct simulation studies to evaluate its finite sample performance, and the simulation of the real data also verifies the accuracy of the proposed method. In practice, the covariate may be changing with the time. It would lead to lower efficiency of the estimator if we neglect the influence of the time-varying covariate. Thus we will study these situations in the future.

Funding

Feng's work was funded by the National Natural Science Foundation of China(NSFC), NO.71271165.

References

[1] Jeong J H. Statistical Inference on Residual Life[M]. Springer New York, 2014.

[2] Jeong J H, Jung S H, Costantino J P. Nonparametric inference on median residual life function [J]. Biometrics, 2008, 64(1): 157-163.

[3] Jung S H, Jeong J H, Bandos H. Regression on quantile residual life[J]. Biometrics, 2009, 65(4): 1203-1212.

[4] Wang H J, Wang L. Locally weighted censored quantile regression [J]. Journal of the American Statistical Association, 2012.

[5] Lin C, Zhang L, Zhou Y. Conditional quantile residual lifetime models for right censored data[J]. Lifetime data analysis, 2015, 21(1): 75-96.

[6] Lagakos S W, Barraj L M, Gruttola V D. Nonparametric Analysis of Truncated Survival Data, with Application to AIDS [J]. Biometrika, 1988, 75(3):515-523.

[7] Stern Y, Tang M X, Albert M S, et al. Predicting time to nursing home care and death in individuals with Alzheimer disease[J]. Jama, 1997, 277(10): 806-812.

[8] Stern Y, Tang M X, Albert M S, et al. Predicting time to nursing home care and death in individuals with Alzheimer disease[J]. Jama, 1997, 277(10): 806-812.

[9] Andersen P K, rnulf Borgan, Gill R D, et al. Statistical Models Based on Counting Processes[J]. Technometrics, 1994, 52(1):1046.

[10] Ripley B D, Solomon P J. Statistical models for prevalent cohort data[J]. Biometrics, 1995, 51(1):373.

[11] Andersen P K, Gill R D. Cox's Regression Model for Counting Processes: A Large Sample Study [J]. Annals of Statistics, 1982, 10(4):1100-1120.

[12] Lin C, Li Z, Yong Z. Conditional quantile residual lifetime models for right censored data[J]. Lifetime Data Analysis, 2015, 21(1):75.

[13] Chen X R, Zhou Y. Quantile Regression for Right-Censored and Length-Biased Data[J]. Acta Mathematicae Applicatae Sinica, English Series, 2012, 28(3):443-462.