STATISTICS IN TRANSITION new series, March 2026
Vol. XX, No. X, pp. xx–XX, https://doi.org/10.59139/stattrans-2026-xxx
Received: xx.xx.xxxx; accepted: xx.xx.xxxx

Jackknife-based diagnostics for non-monotonic hazard survival model with interval-censored data

Jayanthi Arasan[1]

Abstract

This study focuses on jackknife-based model diagnostics for a non-monotonic two-parameter hazard survival regression model (TBPR) when data is interval and right-censored. This distribution is very flexible, because it accommodates both monotonic and bathtub-shaped hazard rates. This research proposes a bias-corrected jackknife harmonic mean and a random imputation technique to obtain the altered Cox-Snell ($r^*_{C_i}$), adjusted Martingale ($r^*_{M_i}$) and Schoenfeld ($r^*_{S_i}$) residuals. Two simulation studies were conducted to assess the performances of the altered residuals and their ability to detect extreme observations and outliers at various censoring proportions (cp) and sample sizes (n) for this model. The results indicated that the altered residuals based on jackknife outperformed other residuals at cp and n levels. The proposed methods are then illustrated using a real dataset on Hodgkin's Disease with the prior treatment group as the covariate. The results showed that the altered residuals work well to address model adequacy and identify potential outliers in the dataset.

Keywords: Jackknife, interval-censored, outliers, covariate.

1. Introduction

Survival data with non-monotonic or bathtub hazard rates is commonly encountered in medical research. Some examples include lifetimes of kidney or heart transplant patients, lifetime of curability of breast cancer and lung cancer patients. The two-parameter distribution with bathtub shape (TPB) model was proposed by (Chen 2000) and extended by (Ismail, Arasan, Safie & Mohd Safari 2022) to incorporate covariates, resulting in what is known as the TPB regression (TPBR) model. This model is very flexible compared to other survival models as it accommodates both monotonic and non-monotonic, namely bathtub shaped hazard rates, see (Chen 2000).

This research focuses on the model diagnostics for the TBPR model when data is both right and interval-censored. Although residual analysis plays a central role in model diagnostics, traditional approaches such as the Cox-Snell residual often fail to perform well under right or interval censoring, particularly when the underlying hazard is non-monotonic. Existing adjustments like midpoint imputation or bootstrap methods have been explored primarily for simpler monotonic hazard models. However, real-world survival data often involve more complex hazard shapes (e.g. bathtub) and censoring types. This study addresses this gap by proposing a jackknife-based adjustment to residuals specifically for the

[1]Department of Mathematics and Statistics, Universiti Putra Malaysia, Malaysia.
E-mail: jayanthi@upm.edu.my. ORCID: https://orcid.org/0000-0003-1805-9601.

TBPR model, a flexible model that accommodates both monotonic and non-monotonic hazard functions.

Interval-censored data is prevalent in many clinical and longitudinal studies, primarily due to constraints such as time, cost, and the necessity for periodic inspections conducted at varying intervals. Data is interval-censored when the lifetime of the i^{th} patient lies within an interval, $t_{L_i} < t_i < t_{R_i}$, where t_{L_i} and t_{R_i} denote the left and right endpoints of the observed interval, respectively.

A special case of the interval-censored data, where $t_{L_i} < t_i < \infty$, gives us the right-censored data, see (Sun 2006), who provided a detailed overview of statistical methods for analyzing interval-censored failure time data, covering techniques like maximum likelihood, nonparametric, semiparametric, and Bayesian methods. (Lawless 1982) discussed statistical methods for analyzing interval-censored data, including current status data as a special case.

The use of computer intensive techniques such as the jackknife and bootstrap can be found in (Arasan & Lunn 2008), who compared alternative confidence interval estimation methods, including bootstrap and jackknife techniques, for the parameters of a parallel two-component system model with dependent failure and time-varying covariates, showing that the jackknife method outperforms bootstrap techniques for censored data. (Arasan & Lunn 2009), extended a parallel system survival model based on the bivariate exponential to include a time-varying covariate, evaluating parameter estimates at various censoring levels, comparing fixed vs. time-varying covariate models, and studying Wald, likelihood ratio, and jackknife methods for constructing confidence intervals, with applications to diabetic retinopathy data.

Following that, (Manoharan, Arasan, Midi & Adam 2015) compared the performance of Wald, likelihood ratio, and jackknife confidence intervals for the parameters of the log-normal distribution in the presence of left-truncated and right-censored survival data, finding that the jackknife method outperformed the others, particularly for small sample sizes with left-truncated data and low censoring. (Kiani, Arasan, Midi et al. 2012) examined the Gompertz model with time-dependent covariates and right-censored data, comparing its performance at different censoring levels and sample sizes, and evaluating Wald and jackknife methods for confidence intervals.

Survival models with interval-censored data have been explored by authors such as (Kiani & Arasan 2013), who extended the Gompertz model with time-dependent covariates for interval-censored data, comparing the performance of Wald and likelihood ratio methods for confidence interval estimation. The study highlighted the effectiveness of these methods in handling interval-censored data. (Fang, Arasan, Midi & Bakar 2015) compared jackknife and bootstrap confidence interval estimates for the parameters of a log-logistic model with censored data and covariates, evaluating their performance through coverage probability studies at various error probability levels and censoring proportions.

(Alharbi, Jayanthi, Haizum & Ling 2022) extended the generalized exponential model to include covariates for interval-censored data, evaluating the maximum likelihood estimator and Wald confidence intervals, with better performance observed at larger sample sizes and lower censoring proportions. Then, (Al-Hakeem, Arasan, Mustafa & Peng 2023) extended the generalized exponential distribution to incorporate time-dependent covariates for

interval-censored data, comparing maximum likelihood estimations and finding better performance with larger sample sizes and lower attendance probabilities. (Manoharan, Arasan, Midi & Adam 2020) assessed the performance of local influential diagnostics for the extended log-normal model with time-dependent covariates, left-truncation, and case-k interval censoring, comparing it with global diagnostics through a simulation study.

More recently, several models and inference methods have been developed for interval-censored survival data. For instance, (Zhou, Sun & Ibrahim 2021, Zhou & Sun 2021) explored transformation models and estimation techniques. (García Meixide, Lema & Vilar 2024) proposed a sparse neural network AFT model for interval-censored outcomes, demonstrating improved prediction performance over classical methods using real-world biomedical data. (Lou, Li & Sun 2024) developed a two-step semiparametric transformation approach to handle missing covariate issues, supported by simulations and an Alzheimer's disease dataset. (Zhang, Li & Weng 2023) introduced a valid inference procedure post-variable selection for the Cox model with interval-censored data, using lasso and asymptotic techniques. Lastly, (Pal, Peng & Aselisewine 2023) discussed a support vector–based semiparametric cure model that accommodates interval-censored survival times.

Other research related to survival models with covariates include (Arasan & Ehsani 2011), who applied a repairable system model for interval failure data with a time-dependent covariate, evaluating several NHPP-based models on ball bearing failure data and using bootstrapping for variance estimation. They found that the proposed model was effective and easy to implement. (Manoharan, Arasan, Midi & Adam 2017) extended the three-parameter log-normal survival model to incorporate left-truncated and right-censored data with covariates. They applied bootstrap inferential procedures to estimate the parameters and assessed the model's performance through a simulation study.

The Cox-Snell residuals (r_{C_i}) are commonly used for checking the fit of a model in survival analysis. When the data is positively skewed because of censoring, the Cox-Snell residuals tend to be smaller or less informative because they are based on the assumption that all observations are fully observed, which is not the case with censored data, as pointed out by (Cox & Snell 1968). To correct this, the Cox-Snell residuals can be modified by adding a positive surplus to make it more reliable. Two conventional modifications of r_{C_i} take the surplus as the mean ($r^{'}_{C_i}$) and median ($r^{''}_{C_i}$) of the standard exponential distribution, see (Cox & Snell 1968). The use of the median of the standard exponential distribution for the surplus was proposed by (Crowley & Hu 1977) as they found the mean tends to inflate the residual far too much. Normally, the arithmetic mean works well when the data is simple and does not have extreme values or outliers, as discussed by (Huber 1981). If the data contains extreme values, the arithmetic mean may not be ideal, as it can overly increase the residuals.

For survival data, which can often follow an exponential or skewed pattern, the geometric mean is a better option because it handles this kind of data more effectively. However, when the data contains extreme values or outliers, the harmonic mean is preferred because it is less affected by these extremes. (Naslina, Jayanthi, Syahida & Bakri 2020) and (Lai & Arasan 2020) deduced that the modified Cox-Snell residuals for the Gompertz model based on the empirical harmonic mean perform better than both standard and other modified Cox-Snell residuals. (Arasan & Midi 2021) concluded that harmonic mean and jackknife har-

monic mean residuals perform significantly better, especially when censoring proportions are high.

In the case of interval-censored data, where the exact timing of an event is unknown but falls within a specified range, traditional methods may not yield accurate results. When data is interval-censored, the Cox-Snell residuals themselves are also interval-censored. (Farrington 2000) recommends replacing the interval residuals with expected values under $\exp(1)$. However, this approach may be impractical for more complex models or when the data exhibits mixed-case censoring. To address this, this study proposes a change to the Cox-Snell residuals by using the jackknife bias-corrected harmonic mean and random imputation, which is better at dealing with heavy censoring. This adjustment is expected to give more reliable results, as shown in (Arasan & Midi 2023), especially when the data is censored in different ways and contains outliers, which can improve model assessment accuracy.

(Arasan & Midi 2023) introduced a method using the bias-corrected bootstrap harmonic mean and random imputation to adjust residuals for the extreme minimum value regression with right- and interval-censored data. The extreme minimum value regression model only accommodates monotonic hazards with a simpler data structure. Their study demonstrated that these adjusted residuals were effective for assessing model adequacy and identifying influential observations. In contrast, the current study focuses on modifying the Cox-Snell residuals using the jackknife bias-corrected harmonic mean and random imputation for a two-parameter distribution with a bathtub-shaped hazard, which has a more complex data structure. While both the study by (Arasan & Midi 2023) and the present work aim to improve residuals in the presence of censoring, our approach emphasizes the jackknife technique, particularly in cases of mixed censoring with a non-monotonic hazard rate. Although these two studies explore similar goals, they propose different models and computational techniques for addressing residual issues in survival analysis.

2. Methodology

2.1. The model

Let T be a non-negative random variable representing the survival time of an event. The density and survivor functions for the TBP model by (Chen 2000) are given by Eqs. (1) and (2).

$$f(t,\lambda,\gamma) = \lambda\gamma t^{\gamma-1}\exp\left(t^{\gamma}+\lambda(1-e^{t^{\gamma}})\right), \quad (1)$$

$$S(t,\lambda,\gamma) = \exp\left(\lambda\left(1-e^{t^{\gamma}}\right)\right), \quad t>0. \quad (2)$$

The effect of the covariates can be incorporated into the model by allowing the parameter λ to be a function of the covariates. If the vector of covariate values is $x' = (x_0, x_1, \cdots, x_{p-1})$, and the vector of regression coefficients is $\beta' = (\beta_0, \beta_1, \cdots, \beta_{p-1})$, then $\lambda = e^{-\beta' x}$, where $\gamma > 0$ represents an unknown parameter. The density and survivor functions for the TBPR

model are given by Eqs. (3) and (4).

$$f(t,\beta,\gamma) = \gamma t^{\gamma-1}\exp\left(-\beta'\mathbf{x}+t^{\gamma}+e^{-\beta'\mathbf{x}}(1-e^{t^{\gamma}})\right), \tag{3}$$

$$S(t,\beta,\gamma) = \exp\left(e^{-\beta'\mathbf{x}}(1-e^{t^{\gamma}})\right), \quad t>0. \tag{4}$$

The distribution has a monotonically increasing hazard function when $\gamma \geq 1$ and may have a bathtub-shaped hazard function when $\gamma < 1$. Consider the case where there are lifetimes for $i = 1,2,\ldots,n$ observations. Let the left and right endpoints for the i^{th} subject be t_{L_i} and t_{R_i}, respectively. To distinguish between censoring types for each observation, we define an indicator variable δ_i as follows:

$$\delta_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ observation is interval-censored,} \\ 0 & \text{if the } i^{\text{th}} \text{ observation is right-censored.} \end{cases} \tag{5}$$

The likelihood function for the full sample with interval and right-censored data is shown by Eq. (6).

$$l(\beta,\gamma) = \prod_{i=1}^{n}\left[S(t_{L_i})-S(t_{R_i})\right]^{\delta_i}\left[S(t_{L_i})\right]^{(1-\delta_i)}. \tag{6}$$

So, for the TBPR model the likelihood and log-likelihood functions for the full sample are shown by Eqs. (7) and (8).

$$l(\beta,\gamma) = \prod_{i=1}^{n}\left[e^{-e^{\beta'x_i}(1-exp(t_{L_i}^{\gamma}))}-e^{-e^{\beta'x_i}(1-exp(t_{R_i}^{\gamma}))}\right]^{\delta_i}\left[e^{-e^{\beta'x_i}(1-e^{t_{L_i}^{\gamma}})}\right]^{(1-\delta_i)} \tag{7}$$

$$\begin{aligned} L(\beta,\gamma) = {} & \sum_{i=1}^{n}\delta_i\left\{\log\left[e^{-e^{\beta'x_i}(1-exp(t_{L_i}^{\gamma}))}-e^{-e^{\beta'x_i}(1-exp(t_{R_i}^{\gamma}))}\right]\right\} \\ & +(1-\delta_i)\left\{e^{-e^{\beta'x_i}(1-e^{t_{L_i}^{\gamma}})}\right\}. \end{aligned} \tag{8}$$

The estimates for β and γ are obtained by solving the likelihood equations using any iterative technique suited for nonlinear equations. The inverse of the observed information matrix, denoted as $i(\widehat{\beta},\widehat{\gamma})$, can be computed from the second partial derivatives of the log-likelihood function, evaluated at $\widehat{\beta}$ and $\widehat{\gamma}$, providing estimates for the variance and covariance, as shown in Eq. (9).

$$\widehat{Var}(\widehat{\beta},\widehat{\gamma}) = [i(\widehat{\beta},\widehat{\gamma})]^{-1}. \tag{9}$$

2.2. The residuals

The Cox-Snell residual for the i^{th} subject is given by $r_{C_i} = \hat{H}(t_i) = -\log(\hat{S}(t_i))$, where $\hat{H}(t_i)$ and $\hat{S}(t_i)$ are the estimated cumulative hazard and survivor functions, respectively. As discussed by (Cox & Snell 1968), a challenge arises when dealing with censored data, particularly right-censored observations, as these residuals tend to underestimate the true values. To address this, we propose modified Cox-Snell residuals using bias-corrected harmonic means via the jackknife method and random imputation, depending on the type of censoring.

Right-Censored Data

To adjust Cox-Snell residuals under right-censoring, the jackknife bias-corrected harmonic mean is applied. The i^{th} jackknife sample is constructed by removing the i^{th} observation from the original dataset of n observations, as described by (Efron & Tibshirani 1994) and defined in Eq. (10).

$$t_{(i)} = (t_1, t_2, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n) \tag{10}$$

Let $\hat{\theta}_{h(i)}$ be the harmonic mean from the i^{th} jackknife sample. The average of these harmonic means is:

$$\hat{\theta}_{h(.)} = \sum_{i=1}^{n} \frac{\hat{\theta}_{h(i)}}{n}$$

The jackknife estimate of bias is given by $(n-1)(\hat{\theta}_{h(.)} - \hat{\theta}_h)$, where $\hat{\theta}_h$ is the harmonic mean of the full dataset. The jackknife bias-corrected estimate can then be obtained as shown in Eq. (11).

$$\hat{\theta}_{h_{jack}} = n\hat{\theta}_h - (n-1)(\hat{\theta}_{h(.)}). \tag{11}$$

The altered Cox-Snell residual when data is right-censored using the jackknife bias-corrected estimate is given by Eq. (12).

$$r_{C_i}^{\text{jack}} = r_{C_i} + \hat{\theta}_{h_{jack}} \quad \text{for the } i^{\text{th}} \text{ subject} \tag{12}$$

Interval-Censored Data

Residual analysis under interval-censoring is more complex (Farrington 2000). A practical method introduced by (Arasan & Midi 2023) uses random imputation. Let $S(\cdot)$ denote the model-based survivor function. For the i^{th} subject, generate R values from the uniform distribution $U(S(t_{R_i}), S(t_{L_i}))$, then transform these to obtain pseudo-lifetimes ${t_i}^r$, for $r = 1, 2, \ldots, R$. The imputed lifetime is estimated as:

$$t_i^{'} = \sum_{r=1}^{R} \frac{{t_i}^r}{R}$$

The adjusted Cox-Snell residual for interval-censored data is then:

$$r_{C_i}^{\text{int}} = \hat{H}(t_i^{'}) = -\log(\hat{S}(t_i^{'})) \tag{13}$$

General Formulation

Combining both scenarios, the modified Cox-Snell residual is defined as:

$$r_{C_i}^{*} = \begin{cases} r_{C_i}^{\text{jack}} & \text{if data is right-censored} \\ r_{C_i}^{\text{int}} & \text{if data is interval-censored} \end{cases} \tag{14}$$

Following that, the adjusted martingale and deviance residuals using the jackknife bias-corrected estimate are given by Eqs. (15) and (16).

$$r_{M_i}^{*} = \delta_i - r_{C_i}^{*}. \tag{15}$$

$$r_{D_i}^{*} = \text{Sgn}(r_{M_i}^{*})[-2(r_{M_i}^{*} + \delta_i \ln(\delta_i - r_{M_i}^{*}))]^{1/2}. \tag{16}$$

The score or Schoenfeld residual (r_{S_i}) was proposed by (Schoenfeld 1982), and is derived from the first derivatives of the log-likelihood function with respect to its parameters. Consequently, these residuals exhibit varying values for each parameter in the model. Since the data is both interval- and right-censored, the adjusted score residuals can be obtained using the imputed lifetimes discussed in Section 2.2. Let,

$$\tilde{t} = \begin{cases} t_i^{'} & \text{for } t_i \text{ interval-censored,} \\ t_{L_i} & \text{for } t_i \text{ right-censored.} \end{cases} \tag{17}$$

The log-likelihood for the full sample is given by Eq. (18).

$$\begin{aligned} \ell(\beta,\gamma) &= \sum_{i=1}^{n} \delta_i \log f(\tilde{t}_i,\beta,\gamma) + (1-\delta_i)\log S(\tilde{t}_i,\beta,\gamma) \\ &= \sum_{i=1}^{n} \delta_i \left[\log\gamma + (\gamma-1)\log\tilde{t}_i - \beta'\mathbf{x}_i + \tilde{t}_i^{\gamma} + e^{-\beta'\mathbf{x}_i}(1-e^{\tilde{t}_i^{\gamma}})\right] \\ &\quad + (1-\delta_i)\left[e^{-\beta'\mathbf{x}_i}(1-e^{\tilde{t}_i^{\gamma}})\right] \end{aligned} \tag{18}$$

The adjusted score residuals ($r_{S_i}^{*}$) can now be calculated from the components of the first derivatives of the log-likelihood function with respect to its parameters, β and γ, evaluated at their respective MLEs, see Eqs. (19) and (20).

$$\frac{\partial L(\beta,\gamma)}{\partial \beta_j} = \sum_{i=1}^{n} -x_{ij}\left[\delta_i + e^{-\beta'\mathbf{x}_i}(1-e^{\tilde{t}_i^{\gamma}})\right], j = 0, 1\cdots, p-1, \tag{19}$$

$$\frac{\partial L(\beta,\gamma)}{\partial \gamma} = \sum_{i=1}^{n}\left[\delta_i\left(\frac{1}{\gamma}+\ln\tilde{t}_i\right)+\tilde{t}_i^{\gamma}\ln\tilde{t}_i\left(\delta_i-e^{-\beta'\mathbf{x}_i}e^{\tilde{t}_i^{\gamma}}\right)\right]. \quad (20)$$

The plot of $r^*_{S_i}$ versus the observation number should be randomly distributed around zero for a good fit. Index plots of the score residuals for each covariate in the fitted model are useful at indicating extreme observations and outliers.

3. Simulation study

Two simulation studies were designed to assess different aspects of the proposed residual diagnostics as follows.

- **Simulation Study I (Sim I)** was designed to identify the most suitable modified residuals by evaluating their ability to assess model adequacy across different levels of censoring and sample sizes.
- **Simulation Study II (Sim II)** builds on the findings of Sim I and investigates the effectiveness of the best-performing residuals from Sim I in detecting extreme or influential observations, which is vital for model diagnostics in clinical survival data.

Sim I was conducted using 1000 replications, at $n = 50$, 80 and $n = 120$, with approximate right censoring proportions (cp) of 0.30, 0.40, 0.50, 0.55 and 0.60 for the TBPR model. The objective is to compare the effectiveness of altered Cox-Snell residuals, utilizing bias-corrected jackknife harmonic mean and multiple imputation, $r^*_{C_i}$ against $r^{'}_{C_i}$ and $r^{''}_{C_i}$, using mid-point imputation. Mid-point imputation estimates lifetimes by using the average of the t_{L_i} and t_{R_i}. It assumes the true value is near the middle of the range. The simulation study only examines the effectiveness of the altered Cox-Snell residuals, as the values of the martingale and deviance residuals are based on these altered Cox-Snell residuals.

The values of β_0, β_1, and γ were set to 3.3, 0.95, and 0.42, respectively, to mimic the lifetime of cancer data, measured in months. Survival times were derived using the inverse transformation method. For the i^{th} observation, the censoring time c_i follows an exponential distribution with parameter μ, where the value of μ is adjusted to achieve the desired approximate right censoring proportion in our dataset. The covariate was simulated as a categorical variable with proportions set to $P = 0.5$ to mimic the distribution of treatment types among patients. The parameter estimates can be obtained by solving the likelihood equations using an iterative procedure designed for nonlinear equations. In this study, the maximum likelihood estimators for all parameters were computed employing the Newton-Raphson iterative method.

To generate interval-censored data, we utilized a sequence of 24 check-up times, $\tau_1, \tau_2, \cdots, \tau_\kappa$, spaced at two-month intervals, assuming all subjects attended these check-ups. Subsequently, we determined whether the uncensored lifetimes, t_i, fell within any of these intervals. If t_i fell within the interval (τ_m, τ_{m+1}) where $m \leq \kappa$, then the corresponding left and right bounds, t_{L_i} and t_{R_i}, for the i^{th} observation, were set to τ_m and τ_{m+1}, respectively. Otherwise, if $t_i > \tau_\kappa$, t_i would be right-censored at τ_κ.

To assess the efficacy of various modifications of the Cox-Snell residuals, it is necessary to derive the estimated Kaplan-Meier survivor function based on the values of these altered residuals. Let $\hat{S}(r^*_{C_i})$ represent the estimated Kaplan-Meier survivor function derived from the adjusted Cox-Snell residuals. The plot of $\log[-\log(\hat{S}(r^*_{C_i}))]$ against $\log(r^*_{C_i})$ should ideally manifest as a linear function with unit slope and intercept zero, as expected when the residuals follow an unit exponential distribution under a correctly specified model. Consequently, by applying the same methodology to the other residuals, their performances can be compared based on the mean absolute deviation (MAD) of three key metrics: the intercept, slope, and correlation coefficient R, from their ideal values of 0, 1, and 1, respectively, indicating a well-fitting model.

Sim II, with 1000 replications, was also carried out using sample sizes of $50, 80$, 200 and 360, along with approximate right censoring proportions (cp) set at 0.10 and 0.30 for the TBPR model. The covariate was simulated as categorical variable with a proportion set to $P = 0.5$ to mimic the distribution of two different treatment types among patients. The purpose of this simulation study is to assess and compare the effectiveness of the best adjusted residuals in detecting extreme observations and outliers. Two data points were randomly chosen and perturbed to yield extreme observations compared to the others. This was achieved by altering the m^{th} lifetime, t_m, by an amount $\omega = 3.5$ scaled by the standard deviation of the lifetimes, s_t, and the largest censored observation, t_{max}, resulting in $t^{'}_m = t_m + \omega s_t + t_{max}$.

The detection percentage was determined based on whether the randomly selected outliers produced the two largest absolute values of the adjusted residuals. For the score residuals, the residual corresponding to the covariate parameter was used to detect outliers. The detection rate was further categorized into two cases: the percentage of datasets where both outliers were detected and the percentage of datasets where only one outlier was detected. In some cases, the methods did not detect any outliers. The overall detection rate was calculated by considering both full and partial detections. Specifically, full weight was given to cases where both outliers were detected, while half weight was assigned to cases where only one outlier was detected.

3.1. Simulation results

The results of Sim I are given in Figures 1-3. The plots demonstrate that the newly proposed adjusted residual, $r^*_{C_i}$ consistently exhibits significantly lower MAD values for intercept, slope, and correlation coefficient (R) across all levels of censoring proportions and sample sizes. This indicates superior performance by $r^*_{C_i}$ in assessing model adequacy. Although performance of $r^{''}_{C_i}$ is marginally superior to that of $r^{'}_{C_i}$, $r^*_{C_i}$ notably surpasses both in indicating a well-fitted model. As n increases, the gap narrows, but $r^*_{C_i}$ still maintains a clear advantage, supporting its robustness across different data conditions. These results confirm the effectiveness of the jackknife bias correction and random imputation method in improving residual-based diagnostics for right- and interval-censored data.

Sim II results, shown in Table 1, evaluate the ability of four adjusted residuals, $r^*_{C_i}$, $r^*_{M_i}$, $r^*_{D_i}$, and $r^*_{S_i}$, to detect outliers under two censoring scenarios: $cp = 0.10$ and $cp = 0.30$. The values outside parentheses correspond to $cp = 0.10$, while those in parentheses refer to

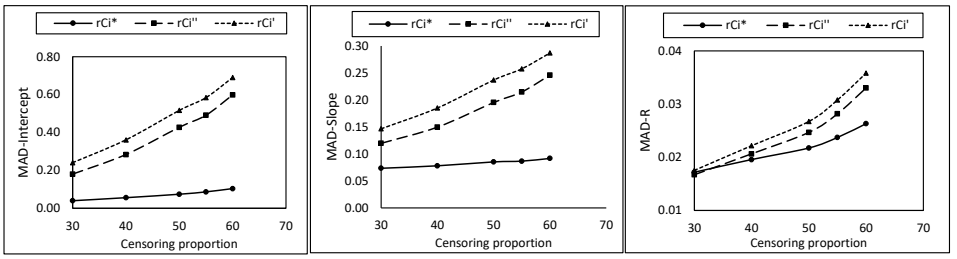

Figure 1. MAD for TBPR model at $n = 50$

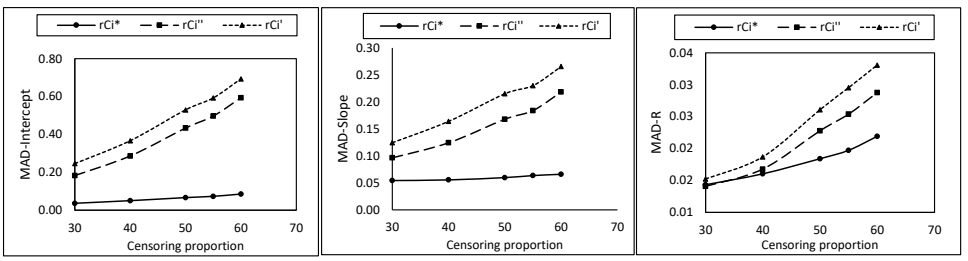

Figure 2. MAD for TBPR model at $n = 80$

$cp = 0.30$. The goal is to measure how often the two deliberately perturbed observations are correctly identified as the most extreme. The results indicate that the newly proposed adjusted Cox-Snell residual, $r^*_{C_i}$, performs best in detecting both outliers in over 99% of simulations even at small sample sizes, and maintaining perfect detection rates at $n = 200$ and above. It is followed by $r^*_{M_i}$, then $r^*_{D_i}$ and $r^*_{S_i}$, in terms of detection accuracy.

The performance of all methods improve as n increases and when $n = 360$, where all residuals except the adjusted score residuals achieved 100% detection for both outliers. When $cp = 0.3$, $r^*_{C_i}$ and $r^*_{M_i}$ remain robust, maintaining overall detection above 95%, even with a sample size as low as $n = 50$. However, the performance of $r^*_{D_i}$ declines rapidly, achieving only 65.9% overall detection compared to $r^*_{S_i}$, which achieves 86.3%. Once again, all performances improve as n increases, particularly $r^*_{D_i}$, which begins to outperform $r^*_{S_i}$ when $n \geq 200$. However, only $r^*_{C_i}$ and $r^*_{M_i}$ achieve 100% detection for both outliers at all censoring levels when $n = 360$.

These findings demonstrate the effectiveness of $r^*_{C_i}$ for assessing model adequacy and detecting outliers, especially in complex censoring settings. Comparing the performance of different residuals across sample sizes and censoring levels also offers practical guidance for researchers and clinicians in selecting suitable diagnostics for survival analysis. Together, the results from both simulation studies confirm the reliability and robustness of the proposed residual adjustments.

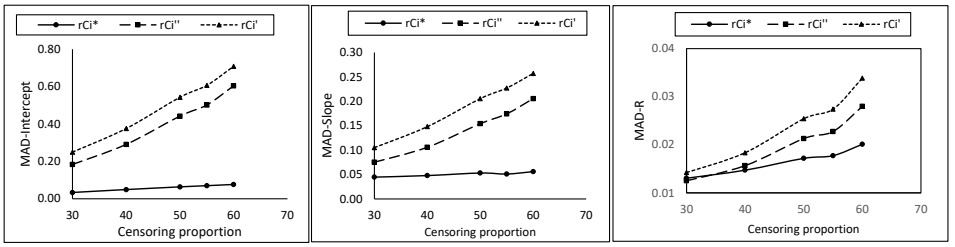

Figure 3. MAD for TBPR model at $n = 120$

Table 1. Percentage detection for different sample sizes and censoring proportions

n	Type	2 obs	1 obs	Overall
50	$r^*_{C_i}$	99.0 (97.5)	1.0 (2.5)	99.5 (98.8)
	$r^*_{M_i}$	97.3 (90.2)	2.7 (9.8)	98.7 (95.1)
	$r^*_{D_i}$	71.7 (39.6)	26.9 (52.5)	85.2 (65.9)
	$r^*_{S_i}$	76.1 (76.3)	20.8 (20.0)	86.5 (86.3)
80	$r^*_{C_i}$	100.0 (99.1)	0.0 (0.9)	100.0 (99.6)
	$r^*_{M_i}$	99.7 (97.1)	0.3 (2.9)	99.9 (98.6)
	$r^*_{D_i}$	86.5 (55.3)	13.2 (40.3)	93.1 (75.5)
	$r^*_{S_i}$	84.1 (80.3)	14.5 (16.9)	91.4 (88.8)
200	$r^*_{C_i}$	100.0 (100.0)	0.0 (0.0)	100.0 (100.0)
	$r^*_{M_i}$	100.0 (100.0)	0.0 (0.0)	100.0 (100.0)
	$r^*_{D_i}$	99.0 (89.1)	1.0 (10.9)	99.5 (94.6)
	$r^*_{S_i}$	94.3 (88.5)	5.7 (10.7)	97.2 (93.9)
360	$r^*_{C_i}$	100.0 (100.0)	0.0 (0.0)	100.0 (100.0)
	$r^*_{M_i}$	100.0 (100.0)	0.0 (0.0)	100.0 (100.0)
	$r^*_{D_i}$	100.0 (98.5)	0.0 (1.5)	100.0 (99.3)
	$r^*_{S_i}$	97.1 (93.8)	2.9(6.1)	98.6 (96.9)

Values outside parentheses correspond to $cp = 0.10$, while values in parentheses correspond to $cp = 0.30$.

4. Real example on Hodgkin's Disease

In this section, we apply the proposed methods to a real dataset to demonstrate the practical applicability of the modified residuals. The dataset comprises the survival times (in months) of 35 patients diagnosed with Hodgkin's Disease and treated with nitrogen mustards, as originally analyzed by (Bartolucci & Dickey 1977). The survival time represents the duration from treatment initiation to either death or censoring. Patients were classified into two groups: Group 1 received minimal or no prior therapy, while Group 2 underwent heavy prior therapy. Among these patients, 9 were right-censored, resulting in a censoring proportion of $cp = 0.257$, which falls within the range considered in our simulation studies.

This dataset was selected to evaluate the diagnostic performance of the adjusted residuals in detecting model fit and influential observations in a real-world clinical scenario. We focus on checking the fit of the TBPR model and testing the modified residuals with both right- and interval-censored data. By comparing the results from our simulations with the

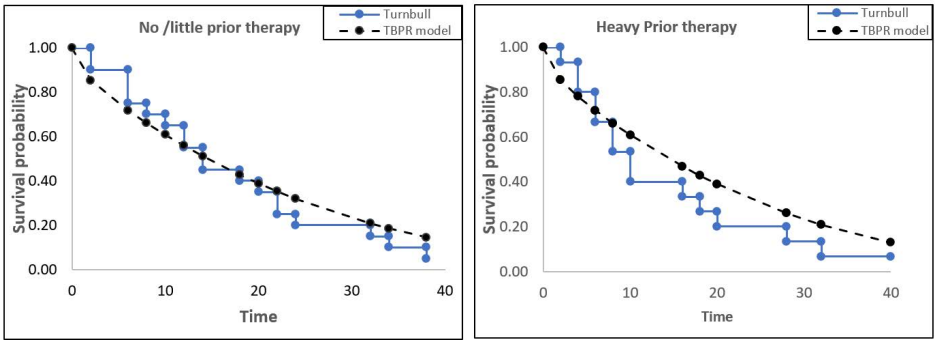

Figure 4. Turnbull and TBPR Survivor Function Estimates for Hodgkin's Disease data by Group

real data, this section will help confirm the findings from the simulation study and shows how useful the proposed methods are for model diagnostics in a clinical context.

The TBPR model was fitted using the treatment group as a categorical covariate. To align with the objectives of this study, the data was modified to create interval-censored data with a 2-month width. To assess the model fit, we obtain the Turnbull estimate of the survivor function (TB) and compare it with the parametric survivor function obtained using the TBPR model for each patient group. Figure 4 presents the plots, indicating that employing the TBPR model would be rather appropriate for the dataset. The survival functions for the two groups indicate that the patients who received little or no prior therapy have slightly better chances of survival than the patients who received heavy prior therapy.

Table 2 displays the parameter estimates obtained from fitting the TBPR model to the Hodgkin's Disease dataset, with group as the covariate. The p-value associated with β_1 indicates a lack of statistically significant difference between patients who received minimal or no prior treatment and those who underwent heavy prior treatments. According to the estimated parameters, the median survival time for patients in the first and second groups is 14.5 and 14.2 months, respectively, indicating a relatively small difference. Figures 5 and 6 show the index plots for $r^*_{C_i}$, $r^*_{M_i}$,$r^*_{D_i}$, and $r^*_{S_i}$ for the Hodgkin's Disease data.

All plots except the $r^*_{D_i}$ plot indicate that observation 30 exceeds the two standard deviations from the mean limit, respectively. Thus, it is important that we investigate this observation thoroughly. Patient 30 had the longest censored lifetime of approximately 40 months among those who underwent extensive prior treatment. All other observations, whether they experienced failure (uncensored) or were censored, had survival times shorter than observation 18. The $r^*_{S_i}$ plot was the only one that singled out observation 18 in addition to observation 30. This was the second largest censored lifetime of approximately 30 months among those who underwent extensive prior treatment. All other patients, whether they experienced an event (failure) or were censored, had survival times shorter than that of observation 18. However, since observation 18 was not flagged as extreme in either the $r^*_{C_i}$ and $r^*_{M_i}$ plots, both of which exhibited superior performance in the simulation study, it is unlikely to be a true outlier.

Table 2. Estimates and 95% Wald interval for the parameters of TBPR Model

Parameter	Estimates	Std.Err	Z	P Val	lower	upper
β_0	2.794	0.693	4.030	0.000	1.435	4.153
β_1	-0.0199	0.404	-0.049	0.961	-0.811	0.772
γ	0.341	0.035	9.830	0.000	0.273	0.410

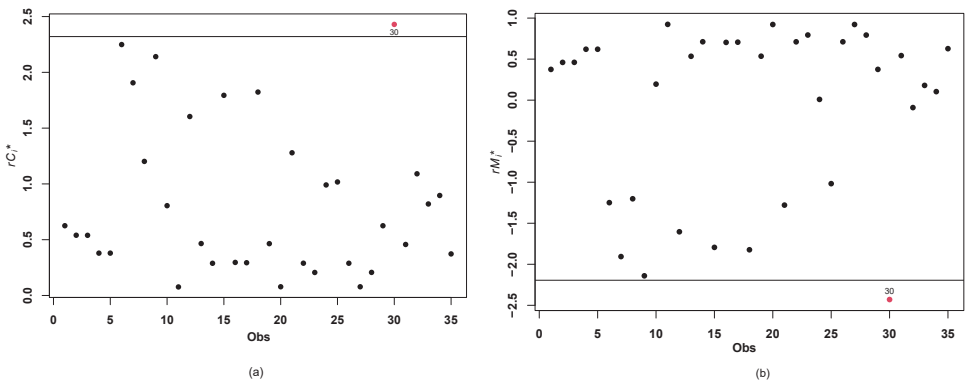

Figure 5. Index plot of adjusted Cox-Snell (a) and adjusted martingale (b) residuals for Hodgkin's Disease data.

The analysis of the Hodgkin's Disease dataset confirmed the findings from the simulation studies. The adjusted residuals, particularly $r^*_{C_i}$ and $r^*_{M_i}$, effectively identified observation 30 as an influential outlier, demonstrating their reliability across both simulated and practical scenarios. This is clearly illustrated in Figure 5, where observation 30 appears as a distinct outlier with substantially higher residual values in both the $r^*_{C_i}$ and $r^*_{M_i}$ plots, reinforcing the simulation findings. These results support the practical effectiveness of the proposed methods in real-world survival analysis, even under moderate censoring. The TBPR model showed a good fit to the data, and patients with minimal or no prior therapy exhibited slightly better survival outcomes.

These real data findings align with those observed in Simulation Study II. The adjusted residuals $r^*_{C_i}$ and $r^*_{M_i}$ consistently identified the most extreme observations, confirming their robustness and diagnostic value. Figure 5 highlights observation 30 as a clear outlier, further validating the ability of these residuals to detect influential cases. In addition, the comparison of model-based and Turnbull survivor curves supports the adequacy of the TBPR model. Together, the results demonstrate that the proposed residual adjustments serve as reliable tools for evaluating model fit and identifying outliers, making them valuable for practical use in survival analysis with interval- or right-censored data.

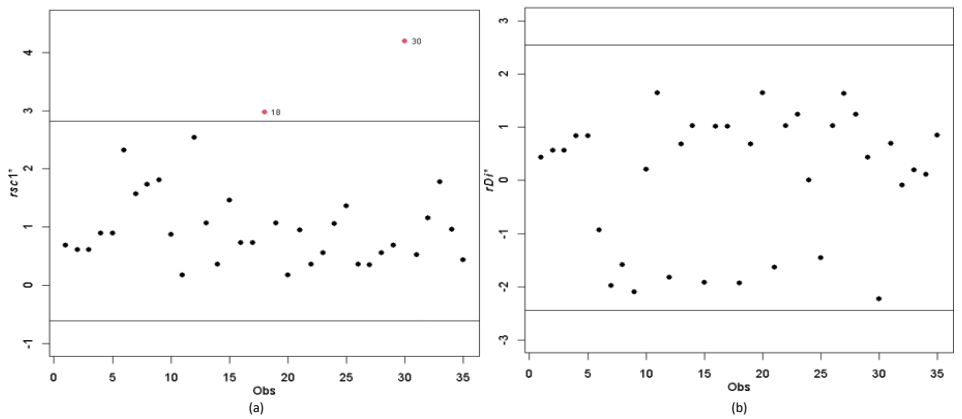

Figure 6. Index plot of adjusted score (a) and adjusted deviance (b) residuals for Hodgkin's Disease data.

5. Conclusion

In this study, we aimed to develop and evaluate modified residuals for model diagnostics in survival analysis, particularly for the TBPR model with right- and interval-censored data. Specifically, we explored the performance of various residuals modified using bias-corrected jackknife harmonic means with random imputation. The results of the first simulation study showed that the newly proposed adjusted residual, $r^*_{C_i}$, consistently outperformed other variations of the Cox-Snell residuals in detecting model fit, exhibiting significantly lower mean absolute deviation (MAD) values across all levels of censoring proportions and sample sizes. While r''_{C_i} performed slightly better than r'_{C_i}, $r^*_{C_i}$ still proved superior in identifying well-fitted models.

The second simulation study showed that $r^*_{C_i}$ and $r^*_{M_i}$ residuals were particularly effective in detecting influential observations, while $r^*_{S_i}$ and $r^*_{D_i}$ residuals performed poorly, although their performance improved as sample sizes increased. These findings highlight the importance of selecting the appropriate residuals for specific types of censoring and sample sizes. Thus, the newly proposed $r^*_{C_i}$ residual significantly outperformed other variations of the Cox-Snell residuals in terms of model diagnostics and detecting extreme observations. Our objective was to improve upon traditional residuals and test their effectiveness through both simulation and real data.

We also applied the proposed methods to a modified real dataset on Hodgkin's Disease patients. The goal was to demonstrate how the proposed residual adjustments perform in a real-world survival analysis. We focused on assessing the fit of the TBPR model and testing the modified residuals with both right- and interval-censored data, which are common in survival analysis. By comparing the results from our simulations with the real data, we were able to confirm the findings from the simulation study and show how useful the proposed methods are for model diagnostics in a clinical context. The results showed that the modified residuals were effective in detecting extreme and influential observations, for

the TBPR model, which aligns with the objectives of this study. For instance, the $r^*_{C_i}$ and $r^*_{M_i}$ residuals successfully identified outliers in the Hodgkin's Disease data, confirming their utility in practical, real-world survival analysis. Therefore, the study's objectives were successfully achieved: the proposed modifications to the Cox-Snell residuals improved model diagnostics and provided meaningful results in both simulated and real-world data contexts.

The methods presented in this research, being computationally intensive and empirically driven, can easily be extended to other models, such as bivariate or parallel-system models, and can handle different types of data, including truncated, left-censored, and mixed-case censored data. Further exploration could involve using double bootstrap techniques to refine these diagnostics. Finally, the analysis of the Hodgkin's Disease dataset illustrates that the TBPR model is suitable for the data, with patients who received minimal or no prior therapy having slightly better survival outcomes compared to those who received heavy prior therapy.

References

Al-Hakeem, H. A., Arasan, J., Mustafa, M. S. B. and Peng, L. F., (2023). Generalized exponential distribution with interval-censored data and time dependent covariate. *Communications in Statistics-Simulation and Computation,* 52(12), pp. 6149–6159.

Alharbi, N., Jayanthi, A., Haizum, A. and Ling, W., (2022). Assessing performance of the generalized exponential model in the presence of the interval censored data with covariate. *Austrian Journal of Statistics,* 51(1), pp. 52–69.

Arasan, J., Ehsani, S., (2011). Modeling repairable system failures with interval failure data and time dependent covariate. *Journal of Modern Applied Statistical Methods,* 10, pp. 618–624.

Arasan, J., Lunn, M., (2008). Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics,* 23, pp. 605–622.

Arasan, J., Lunn, M., (2009). Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference,* 139(3), pp 944–951.

Arasan, J., Midi, H., (2021). Jackknife and bootstrap estimates for modified residuals of the log-logistic model. *in AIP Conference Proceedings*, Vol. 2423(1), AIP Publishing LLC, p. 070009.

Arasan, J., Midi, H., (2023). Bootstrap based diagnostics for survival regression model with interval and right-censored data. *Austrian Journal of Statistics,* 52(2), pp. 66–85.

Bartolucci, A. A., Dickey, J. M., (1977). Comparative bayesian and traditional inference for gamma-modeled survival data. *Biometrics* pp. 343–354.

Chen, Z., (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters,* 49(2), pp. 155–161.

Cox, D. R., Snell, E. J., (1968). 'A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological),* 30(2), pp. 248–265.

Crowley, J., Hu, M., (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association,* 72(357), pp. 27–36.

Efron, B., Tibshirani, R. J., (1994). *An Introduction to the Bootstrap*, Chapman & Hall/CRC, New York.

Fang, L. Y., Arasan, J., Midi, H. and Bakar, M. R. A., (2015). Jackknife and bootstrap inferential procedures for censored survival data. *in* AIP Conference Proceedings, Vol. 1682(1), AIP Publishing.

Farrington, C. P., (2000). Residuals for proportional hazards models with interval-censored survival data. *Biometrics,* 56(2), pp. 473–482.

García Meixide, A., Lema, M. and Vilar, J. M., (2024). A bayesian semiparametric mixture cure model for doubly interval-censored data. *Computational Statistics & Data Analysis,* 190, p. 107925.

Huber, P. J., (1981). *Robust Statistics*, Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Ismail, I., Arasan, J., Safie, M. and Mohd Safari, M. A., (2022). Bathtub hazard model with covariate and right censored data. *Journal of Quality Measurement and Analysis JQMA,* 18(3), pp. 1–15.

Kiani, K., Arasan, J., (2013). Gompertz model with time-dependent covariate in the presence of interval-, right-and left-censored data. *Journal of Statistical Computation and Simulation,* 83(8), pp. 1472–1490.

Kiani, K., Arasan, J., Midi, H. et al., (2012). Interval estimations for parameters of gompertz model with time-dependent covariate and right censored data. *Sains Malaysiana,* 41(4), pp. 471–480.

Lai, M. C., Arasan, J., (2020). Single covariate log-logistic model adequacy with right and interval censored data. *Journal of Quality Measurement and Analysis,* 16(2), pp. 131–140.

Lawless, J. F., (1982). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

Lou, Y., Li, G. and Sun, J., (2024). Interval-censored quantile regression based on fractional counting process. *Statistical Analysis and Modeling,* XX(XX), pp. 1–20.

Manoharan, T., Arasan, J., Midi, H. and Adam, M. B., (2015). A coverage probability on the parameters of the log-normal distribution in the presence of left-truncated and right-censored survival data. *Malaysian Journal of Mathematical Sciences,* 9(1).

Manoharan, T., Arasan, J., Midi, H. and Adam, M. B., (2017). Bootstrap intervals in the presence of left-truncation, censoring and covariates with a parametric distribution. *Sains Malaysiana,* 46(12), pp. 2529–2539.

Manoharan, T., Arasan, J., Midi, H. and Adam, M. B., (2020). Influential measures on log-normal model for left-truncated and case-k interval censored data with time-dependent covariate. *Communications in Statistics-Simulation and Computation,* 49(6), pp. 1445–1466.

Naslina, A. M. N. N., Jayanthi, A., Syahida, Z. H. and Bakri, A. M., (2020). Assessing the goodness of fit of the gompertz model in the presence of right and interval censored data with covariate. *Austrian Journal of Statistics,* 49(3), pp. 57–71.

Pal, N., Peng, Y. and Aselisewine, S., (2023). Bayesian cure rate modeling of interval censored data based on negative binomial distribution. *Statistics & Probability Letters,* 204, p. 109984.

Schoenfeld, D., (1982). Partial residuals for the proportional hazards regression model. *Biometrika,* 69(1), pp. 239–241.

Sun, J., (2006). *The statistical analysis of interval-censored failure time data*, Vol. 3, No. 1, Springer.

Zhang, J., Li, G. and Weng, C., (2023). A semiparametric transformation model for multivariate interval-censored failure time data. *Lifetime Data Analysis,* 30, pp. 41–65.

Zhou, H., Sun, J., (2021). Semiparametric transformation model for interval-censored survival data with covariate measurement error. *Biometrics,* 77(2), pp. 523–535.

Zhou, H., Sun, J. and Ibrahim, J. G., (2021). A review on interval-censored survival data: models, inference methods, and applications. *Statistical Methods in Medical Research,* 30(5), pp. 1312–1336.