

Kullback-Based Selection Criteria for Accelerated Failure Time Models: A Comparative Study

Clarisse Houénafa Dete ¹, Bruno Enagnon Lokonon ¹, Kossi Essona Gneyou ²,
Marcel SENOU ³, and Romain Glèlè Kakai ¹

¹ Laboratoire de Biomathématiques et d'Estimations Forestières,
Faculté des Sciences Agronomiques, Université d'Abomey-Calavi,
Abomey-Calavi, Benin.
clarissedete@gmail.com
brunolokonon@gmail.com
glele.romain@gmail.com

² Department of Mathematics, Université de Lomé, Lomé, Togo.
kossi.gneyou@yahoo.fr

³ Faculté des Sciences Agronomiques, Université d'Abomey-Calavi,
Abomey-Calavi, Benin.
m_senou2002@yahoo.fr

ABSTRACT

Survival analysis involves modeling time-to-event data using multiple models, such as the Accelerated Failure Time (AFT) model. Choosing the most appropriate model is crucial. Various model selection criteria are available in the literature. This study examines the performance of new model selection criteria based on Kullback's Symmetric Divergence within the Parametric AFT model framework. Simulations and real data were used to compare the Kullback Information Criterion (KIC) and its corrected version (KIC_c) with traditional information criteria, such as the Akaike Information Criterion (AIC), corrected AIC (AIC_c), improved AIC (AIC_{sur}), and Bayesian Information Criterion. Results showed that KIC_c outperforms other criteria in selecting the true model order in small samples and maintains competitive predictive accuracy across conditions. KIC and KIC_c outperform the AIC family criteria for moderate and large sample sizes. The real data shows that AIC and AIC_{sur} tend to select overfitted models, while KIC_c and AIC_c favor underfitted models.

Keywords: Survival analysis, model selection criteria, performance, Accelerated Failure Time model.

2020 Mathematics Subject Classification: 62Nxx, 62F07, 62H12.

1 Introduction

Survival analysis is a field within biostatistics that analyzes data representing the duration until a specific event occurs. This duration is usually called failure time, survival time, or lifespan

(Jiang and Guterman, 2024). The statistical literature has created many models for analyzing survival data (Saikia and Barman, 2017). However, the semi-parametric Cox Proportional Hazard (CPH) model is the most used in the literature in recent years to analyze survival data (Mastor et al., 2023). Consistent with traditional statistical models, the CPH model depends on several assumptions. It assumes: i) proportional hazard, ii) independence of the survival times per individual in population-based research, and iii) the censoring is noninformative regarding the interested outcome (Deo et al., 2021). When the first assumption (proportional hazard) fails, the Accelerated Failure Time (AFT) model serves as an alternative to the CPH model (Ouerfelli et al., 2022). The name "accelerated failure time model" (Saikia and Barman, 2017) is derived from the combination of "failure" (such as death or disease) and "accelerated" (which) denotes the specific cause responsible for the increased rate of failure. We refer to the given factor as the "acceleration factor" (Saikia and Barman, 2017). In contrast to other well-known models that operate in the hazard function, such as the proportional hazards (PH) model (Choi et al., 2024), the AFT model considers factors that directly influence the failure time, facilitating impact assessment. This model is widely used in clinical trials and technical applications to measure variations in device usage (Lin, 2018; Mukin et al., 2023). Accelerated Failure Time models are particularly intriguing in clinical trials because they directly analyze the correlation between patient variables and failure times (Henderson et al., 2020).

AFT models can be classified into semi-parametric and parametric types. In the semi-parametric AFT model, the error distribution (or equivalently, the baseline hazard function) is unspecified, which offers flexibility but requires more complex estimation procedures, particularly for regression coefficients (Bhaskaran et al., 2024). In contrast, parametric AFT models assume a specific error distribution, such as Weibull, log-normal, or log-logistic, resulting in simpler and more efficient estimation through maximum likelihood methods. However, selecting the appropriate distribution is not straightforward, and model misspecification can lead to biased inferences (Majeed, 2020). Therefore, employing model selection criteria becomes essential in guiding the choice of the most appropriate parametric form.

An appropriate model selection criterion is a valuable tool for evaluating the suitability of a fitted model by determining if it provides an ideal equilibrium between the quality of fit and the simplicity of the model (Keerativibool and Siripanich, 2017). Several studies have been undertaken to evaluate the performance of model selection criteria (Liang et al., 2008; Dirick et al., 2015; Farooq and Karami, 2019). The initial model selection criterion that achieved broad acceptance was the Akaike Information Criterion (AIC) proposed by Akaike (1974). The AIC is a measure that provides an asymptotically unbiased estimate of the Kullback-Leibler distance between the true and candidate models. However, the AIC tends to favor overfitted models (Netzer et al., 2017). In cases with small to moderate sample sizes and a collection of high-dimensional models as candidates, the Akaike Information Criterion may also favor the larger models (Cavanaugh and Neath, 2019). This is why Hurvich and Tsai (1989) suggested using the corrected AIC (AIC_c) when the sample size is small. Within the same framework, Liang and Zou (2008) has introduced the improved AIC (AIC_{sur}). Moreover, it is advisable to optimize the performance of these criteria (AIC_c or AIC_{sur}) by using the number of events instead of the number of observations in the penalty term (Therneau et al., 2003; Liang and Zou, 2008). The AIC_c

and AIC_{sur} criteria in the AIC family have been derived using Kullback's directed divergence. Furthermore, literature frequently employs the Bayesian Information Criterion (BIC). Practitioners generally favor BIC over AIC due to its Bayesian rationale or inclination to select more concise models instead of AIC (Riansut, 2023). The Akaike Information Criterion (AIC) aims to minimize prediction error. It is known to be asymptotically efficient, meaning it tends to select models that predict future data well as the sample size increases. However, it is inconsistent, so it does not necessarily select the true model, even with large samples. In contrast, the Bayesian Information Criterion (BIC) is consistent (Cavanaugh and Neath, 2019), meaning it tends to select the true model as the sample size increases. Still, it is not asymptotically efficient, so its predictive performance may be lower compared to AIC in some settings. Next, Cavanaugh (1999) developed and examined the Kullback Information Criterion (KIC), another well-known criterion. Cavanaugh (1999) introduced the KIC estimator as an asymptotically unbiased measure of Kullback's symmetric divergence. In small-sample applications, the KIC appears to underestimate Kullback's symmetric divergence, as demonstrated by Cavanaugh (2004). Thus, Cavanaugh (2004) introduced KIC_c to address this issue. As a result, Cavanaugh (2004) proposed a revised version of KIC (KIC_c). Studies have already found that for linear models, generalized linear models, and nonlinear models (Cavanaugh, 2004; Acion, 2011; Azzaoui and Hafidi, 2012; Kim et al., 2014), the criteria based on Kullback's symmetric divergence are better than those based on Kullback's directed divergence. Regrettably, the research still lacks evidence of the performance of these criteria in parametric AFT models. The goal of this study is to compare the different model selection criteria, namely KIC , and KIC_c , in comparison to AIC , AIC_c , AIC_{sur} , and BIC , in parametric AFT models. After the introduction, we structure the study as follows: Section 1 outlines the methodology, section 2 presents our findings and discussions, and section 3 discusses the concise conclusion and potential directions for further research.

2 Methods

2.1 Selection criteria based on Kullback's Divergence

Let Y , θ_0 , and θ_k denote the observed data, the vector of the unknown parameters, and the vector of the parameters of the candidate model, respectively. The subscript k represents the dimension of the vector. Let $\hat{\theta}_k$ be a vector of the maximum likelihood estimator of θ_k , and let $g(Y|\hat{\theta}_k)$ represent the empirical likelihood corresponding to $g(Y|\theta_k)$. The true density of the data is denoted by $g(Y|\theta_0)$, while the candidate density is represented by $g(Y|\theta_k)$. Let us consider the objective of identifying the adjusted model $g(Y|\hat{\theta}_k)$; $k \in \{k_1, k_2, \dots, k_L\}$, which serves as the optimal approximation to $g(Y|\theta_0)$ within the set of families $\{\mathcal{G}(k_1), \mathcal{G}(k_2), \dots, \mathcal{G}(k_L)\}$. If $g(Y|\theta_0)$ includes $\mathcal{G}(k)$ and $\mathcal{G}(k)$ is structured such that no smaller family can include $g(Y|\theta_0)$, then $g(Y|\hat{\theta}_k)$ is correctly specified. $g(Y|\hat{\theta}_k)$ is considered overfit if $g(Y|\theta_0)$ pertains to $\mathcal{G}(k)$ and $\mathcal{G}(k)$ is such that families smaller than $\mathcal{G}(k)$ also contain $g(Y|\theta_0)$. $g(Y|\hat{\theta}_k)$ is classified as underfit if $g(Y|\theta_0)$ is not included of $\mathcal{G}(k)$. The appropriate model can be selected using the Kullback Directed Divergence (Kim et al., 2014; Cavanaugh and Neath, 2019). The Kullback's

Directed Divergence between $g(Y|\theta_0)$ and $g(Y|\theta_k)$ with regard to $g(Y|\theta_0)$ is defined as

$$I(\theta_0, \theta_k) = E_{\theta_0} \left\{ \ln \frac{g(Y|\theta_0)}{g(Y|\theta_k)} \right\}, \quad (2.1)$$

E_{θ} represents the expected value under $g(Y|\theta_0)$. Note that $I(\theta_0, \theta_k) \neq I(\theta_k, \theta_0)$ unless $\theta_k = \theta_0$. The *AIC* family criteria are formulated based on Kullback's directed divergence. Let us consider two parametric density functions, denoted as $g(Y|\theta)$ and $g(Y|\theta_*)$. We define:

$$d(\theta, \theta_*) = E_{\theta} \{-2 \ln g(Y|\theta_*)\}. \quad (2.2)$$

From 2.1 and 2.2, $2I(\theta_0, \theta_k)$ equals $d(\theta_0, \theta_k) - d(\theta_0, \theta_0)$. The independence of $d(\theta_0, \theta_0)$ from θ_k implies that the ranking of candidate models derived from $I(\theta_0, \theta_k)$ will be the same as the ranking based on $d(\theta_0, \theta_k)$. Consequently, for the purpose of this study, $d(\theta_0, \theta_k)$ can serve as a suitable substitute for $I(\theta_0, \theta_k)$ (Keerativibool and Siripanich, 2017). So, $d(\theta_0, \theta_k)$ can be write:

$$d(\theta_0, \hat{\theta}_k) = E_{\theta_0} \{-2 \ln g(Y|\hat{\theta}_k)\}|_{\theta_k=\hat{\theta}_k}. \quad (2.3)$$

However, evaluating $d(\theta_0, \hat{\theta}_k)$ is not feasible without prior knowledge of θ_0 . Akaike (1973, 1974) argue that $-2 \ln g(Y|\hat{\theta}_k)$ serves as a biased estimator of $d(\theta_0, \hat{\theta}_k)$. He proposes an asymptotically unbiased estimator, defined as follows:

$$AIC = -2 \ln g(Y|\hat{\theta}_k) + 2k. \quad (2.4)$$

Let us denote:

$$\Delta(\theta_0, k) = E_{\theta_0} \{d(\theta_0, \hat{\theta}_k)\}. \quad (2.5)$$

It may be deduced that:

$$\Delta(\theta_0, k) = E_{\theta_0} \{AIC\} + o(1). \quad (2.6)$$

However, when the sample size is small or k is significantly large relative to the sample size, *AIC* can favor overfitted models, thus restricting its efficacy (Acion, 2011). Hurvich and Tsai (1989) proposed *AIC_c* as an unbiased estimator of Kullback's directed divergence for small sample sizes. *AIC_c* is represented as follows:

$$AIC_c = -2 \ln g(Y|\hat{\theta}_k) + \frac{2nk}{n - k - 1}. \quad (2.7)$$

Cavanaugh (1999) noted that Kullback's directed divergence resulted in excessively high underfitting values for model selection, particularly in the context of over-parameterized models. An alternative approach to mitigate overfitting and underfitting problems involves reversing the roles of two models in defining the measure known as Kullback's symmetric divergence. Kullback's Symmetric Divergence is defined as follows:

$$\begin{aligned} 2J(\theta_0, \theta_k) &= 2I(\theta_0, \theta_k) + 2I(\theta_k, \theta_0) \\ &= \{d(\theta_0, \theta_k) - d(\theta_0, \theta_0)\} + \{d(\theta_k, \theta_0) - d(\theta_k, \theta_k)\}. \end{aligned} \quad (2.8)$$

Note that $J(\theta_0, \theta_k) = J(\theta_k, \theta_0)$; therefore, $J(\theta_0, \theta_k)$ is symmetric. In this context, $d(\theta_k, \theta_0)$ is defined as $E_{\theta_k} \{-2 \ln g(Y|\theta_0)\}$, while $d(\theta_k, \theta_k)$ is represented as $E_{\theta_k} \{-2 \ln g(Y|\theta_k)\}$. Eliminating the constant term $d(\theta_0, \theta_0)$ from the preceding equation, we have:

$$K(\theta_0, \theta_k) = d(\theta_0, \theta_k) + d(\theta_k, \theta_0) - d(\theta_k, \theta_k),$$

serving as a substitute for the metric $J(\theta_0, \theta_k)$. We will now consider the estimation of

$$K(\theta_0, \hat{\theta}_k) = K(\theta_0, \theta_k)|_{\theta_k = \hat{\theta}_k}.$$

Cavanaugh (1999) proposed an asymptotically unbiased estimator for symmetric divergence, defined as follows:

$$KIC = -2 \ln g(Y|\hat{\theta}_k) + 3k. \quad (2.9)$$

i.e., $E_{\theta_0}\{KIC\} + o(1) = \Omega(\theta_0, k)$.

Seghouane and Bekara (2004) introduced an unbiased estimator $\Omega(\theta_0, k)$ to address the significant negative bias present in KIC as follows:

$$KIC_c = -2 \ln g(Y|\hat{\theta}_k) + \frac{2nk}{n-k-1} - n\psi\left(\frac{n-k+1}{2}\right) + n \ln\left(\frac{n}{2}\right), \quad (2.10)$$

where $\psi(\cdot)$ represents the digamma function. Then KIC_c is an exactly unbiased estimation of $K(\theta_0, \hat{\theta}_k)$

$$E_{\theta_0}\{KIC_c\} = \Omega(\theta_0, k).$$

Due to the absence of a closed-form solution for the $\psi(\cdot)$ in KIC_c , Cavanaugh (2004) proposed an approximation of KIC_c define as:

$$KIC_c = -2 \ln g(Y|\hat{\theta}_k) + n \ln\left(\frac{n}{n-k+1}\right) + \frac{n((n-k+1)(2k+1)-2)}{(n-k-1)(n-k+1)}. \quad (2.11)$$

2.2 Parametric Accelerated Failure Time models

Let T represent the survival time, C denotes the censoring time, and x indicates the associated p covariates. Define $Z = \min\{T, C\}$ as the observed survival time and δ the censoring indicator time ($\delta = 1$ indicating an event and $\delta = 0$ indicating censoring). Let $f(t|x)$ and $S(t|x)$ denote the density and survival functions of T given x , respectively. The AFT model specifies the relationship between the survival function at each time t in T and the covariates as follows:

$$S(t|x) = S_0[t \exp(x' \beta)],$$

where S_0 represents the baseline survival function and $\beta^t = (\beta_1, \beta_2, \dots, \beta_p)$ denotes a vector of regression coefficients. The term $\exp(\beta^t x)$ is called the accelerated factor, which enhances the survival function when the covariate $x = 0$. The relationship between the survival time and a set of covariates can also be illustrated as a linear relation between the natural logarithm of survival time and the covariate x , that is

$$Y = \ln T = \alpha_0 + x' \beta + \sigma \epsilon,$$

where α_0 is the true intercept, $\sigma > 0$ is an unknown scale parameter, and ϵ stands for the error term, an independently and identically distributed (IID) term with an unspecified distribution function F , with a mean of 0, and a variance of 1 (Mustefa and Chen, 2021). The parametric AFT model assumes a specific distribution for the error term. Each distribution of ϵ corresponds to a distribution for T , and AFT models are designated according to the distribution of T rather

than that of ϵ or $\log T$ (Wu, 2018; Collett, 2023). When the error term (ϵ) follows an Extreme Value (one parameter or two parameters), Logistic, or Normal distribution, the resulting survival time T follows an Exponential, Weibull, Log-Logistic, or Log-Normal distribution, respectively (Saikia and Barman, 2017).

We estimate AFT models using the maximum likelihood estimation (MLE) method. The likelihood of observing survival times $t_1, t_2, \dots, t_n$ is

$$\mathcal{L} = \prod_{i=1}^n \{f_i(t)\}^{\delta_i} \{S_i(t)\}^{1-\delta_i},$$

where $S_i(t) = S_0\{t \exp(-x_i' \beta)\}$ and $f_i(t) = \exp(-x_i' \beta) f_0\{t \exp(-x_i' \beta)\}$. The log-likelihood function is given by

$$\begin{aligned} \log \mathcal{L} &= \sum_{i=1}^n \sigma t_i^{-\delta_i} \{f_i(t)\}^{\delta_i} \{S_i(t)\}^{1-\delta_i} \\ &= \sum_{i=1}^n -\delta_i \log(\sigma t_i) + \delta_i \log\{f_i(t_i)\} + (1 - \delta_i) \log\{S_i(t_i)\}, \end{aligned} \quad (2.12)$$

where MLE is of $(p+2)$ unknown parameters α_0, σ , and $\beta_1, \dots, \beta_p$.

The next section evaluates the performance of model selection criteria such as KIC , KIC_c compared to AIC , BIC , AIC_c , and AIC_{sur} in the Parametric Accelerated Failure Time model. Collett (2023) proposed AIC for survival models. It is given by:

$$AIC = -2 \log \mathcal{L} + 2(p+2+q),$$

where q is 0 for the Exponential model, 1 for the Weibull model, 1 for the Log-logistic model, and 2 for the Generalized Gamma model. Liang and Zou (2008) suggest improved AIC as follows:

$$AIC_{sur} = AIC + \frac{2(p+2)(p+3)}{n-p-3}.$$

The formulas for other criteria are:

$$AIC_c = -2 \log \mathcal{L} + \frac{2n(p+2+q)}{n-(p+3+q)}.$$

$$BIC = -2 \log \mathcal{L} + (p+2+q) \log(n).$$

$$KIC = -2 \log \mathcal{L} + 3(p+2+q);$$

$$KIC_c = -2 \log \mathcal{L} + n \log\left(\frac{n}{n-m}\right) + \frac{n((n-m)(2(p+2+q)+1)-2)}{(n-(p+3+q))(n-m)},$$

with $m = p+1+q$. In the survival model, according to Therneau et al. (2003), the number of uncensored events, r , should replace n in the penalty term.

3 Simulation study and real data analysis

3.1 Simulation study design

To assess the performances of the model selection criteria such as KIC , KIC_c , AIC , AIC_c , AIC_{sur} , and BIC , we simulated the true Parametric Accelerated Failure Time model with order

$p(p = 4)$ defined as:

$$\ln(T) = x'\beta + \sigma\epsilon,$$

where $\beta = (1, 2, 3, 4, 0, 0, 0, 0)'$ according to Liang and Zou (2008), x follows an 8-dimensional normal distribution with mean zero and covariance matrix $I_{8 \times 8}$. Three distinct distributions of ϵ are considered: the Logistic distribution, the Normal distribution, and the Extreme value (two parameters) distribution. The location and scale parameters are set to 0 and 1, respectively. A series of simulations were conducted using a variety of combinations of the following variables: sample size (n), censoring percentage (C), scale parameter, and error distribution. The sample sizes were small ($n = 30, 50$), medium ($n = 100, 200$), and large ($n = 400, 500$). The censoring percentage was 50% and 80% (Seo and Kang, 2023); scale parameter of $\sigma^2 = 0.1, 0.5, 1$ (Liang and Zou, 2008). Censoring distribution was Uniform. For each of the 108 configurations, 1000 independent data sets were generated. The candidate models are sequential columns nested. The true model consists of the first four columns of x . The steps for simulation are as follows:

Step 1: Generate the covariates multivariate normal distribution with the mean zero and covariance matrix $I_{8 \times 8}$;

Step 2: Generate the censoring random variable;

Step 3: Generate the error variable assuming a Logistic, Normal, and Extreme value distribution;

Step 4: Generate the survival time using the first four relevant covariates as follows:

$$T = \exp(x'\beta + \sigma\epsilon);$$

Step 5: Fit sequential columns nested potential candidate models. We consider eight candidate models;

Step 6: Compute KIC , KIC_c , BIC , AIC , AIC_c , and AIC_{sur} for each model;

Step 7: Identify the model candidate with the lowest value for the model selection criterion;

Step 8: Classify the selected candidate models into three groups: i) underfitted order, which is defined as a selected order smaller than p ; ii) true order, which is defined as a selected order equal to p ; and iii) overfitted order, which is defined as a selected order greater than p ;

Step 9: Calculate the Root Mean Square Error ($RMSE$) for the selected candidate model according to each model selection criterion (Ambler et al., 2012) :

$$RMSE(t) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{S}_i(t) - S_i(t))^2},$$

where $S_i(t)$ and $\hat{S}_i(t)$ are the true and estimated survival probabilities, respectively, for the i^{th} individual at $t = 100$, and n is the sample size;

Step 10: Repeat all the steps S times ($S = 1000$);

Step 11: Calculate the average $RMSE$.

3.2 Real data analysis

We use the data set provided by Ahmedin and Assen (2023). The study was conducted between March 2021 and August 2021 in Jimma City, Southwestern Ethiopia, to investigate risk

factors linked with calf mortality in the first six months of life. The Jimma calf mortality dataset is well-suited due to its precise time-to-event data, right-censored observations, adequate sample size (235 calves), and several covariates. A total of 235 calves from birth to 6 months of age were sampled from fifty-three dairy farms with herd sizes between 5 and 140 lactating cows. Calves from the selected farms were monitored every 15 days by a veterinarian and a data collector until they were 6 months old. Ahmedin and Assen (2023) provide additional information about the study. Our outcome variable is the survival time to the observation of death during the first 6 months of life and the death indicator (1- death before 6 months, 0- alive until 6 months). Nine explanatory factors of calf mortality have a significant level of $p < 0.2$ with calf mortality were included:

- X_1 : dam parity;
- X_2 : calving status;
- X_3 : dairying experience (year);
- X_4 : herd size;
- X_5 : floor type;
- X_6 : calf housing;
- X_7 : calf pen cleaning frequency;
- X_8 : first colostrum feeding time;
- X_9 : amount of milk-fed.

For our data analysis, we considered the Weibull, Log-Logistic, and Log-Normal survival time distributions. We then compared the candidate models, which were fitted based on all possible combinations of the nine variables ($2^9 - 1$), using the values KIC , KIC_c , BIC , AIC , AIC_{sur} , and AIC_c . We used the model selection criteria to determine the best adjusted model.

4 Results

4.1 Simulation results

Figure 1 presents the proportions of true order selection in the model selection criteria for various error distributions and scale parameters as a function of the sample size when the censoring percentage is 50%. The corrected criterion KIC (KIC_c) shows higher proportions of true order selection for small sample sizes ($n = 30$ or 50). Moreover, for medium and large samples ($n = 100, 200, 400$, and 500), the KIC family (KIC and KIC_c) surpasses the AIC family (AIC and AIC_c) in selecting the true order, regardless of the distribution ϵ and the parameter value σ^2 . However, the KIC family shows lower proportions of true order selection than BIC for these same sample sizes, irrespective of the error distribution and scale parameters. It selects more underfitted models, regardless of sample size, error distribution, or scale parameter (Table 3-5). It means that the KIC family favors parsimonious models and avoids complex models. It should be noted that the KIC and BIC families are more parsimonious in model selection compared to the AIC family criteria. Finally, it is noteworthy that the sample size, the distribution of errors, and the scale parameter influence the performance of the model selection criteria. As the sample size increases, the performance of the model selection criteria KIC , BIC , and AIC improves, but not for the corrected criteria (AIC_c and KIC_c). Regarding

the ϵ distribution, the Normal distribution offers better results than the Logistic and Extreme value distributions. In addition, as the scale value increases, the performance of the criteria decreases for all sample sizes.

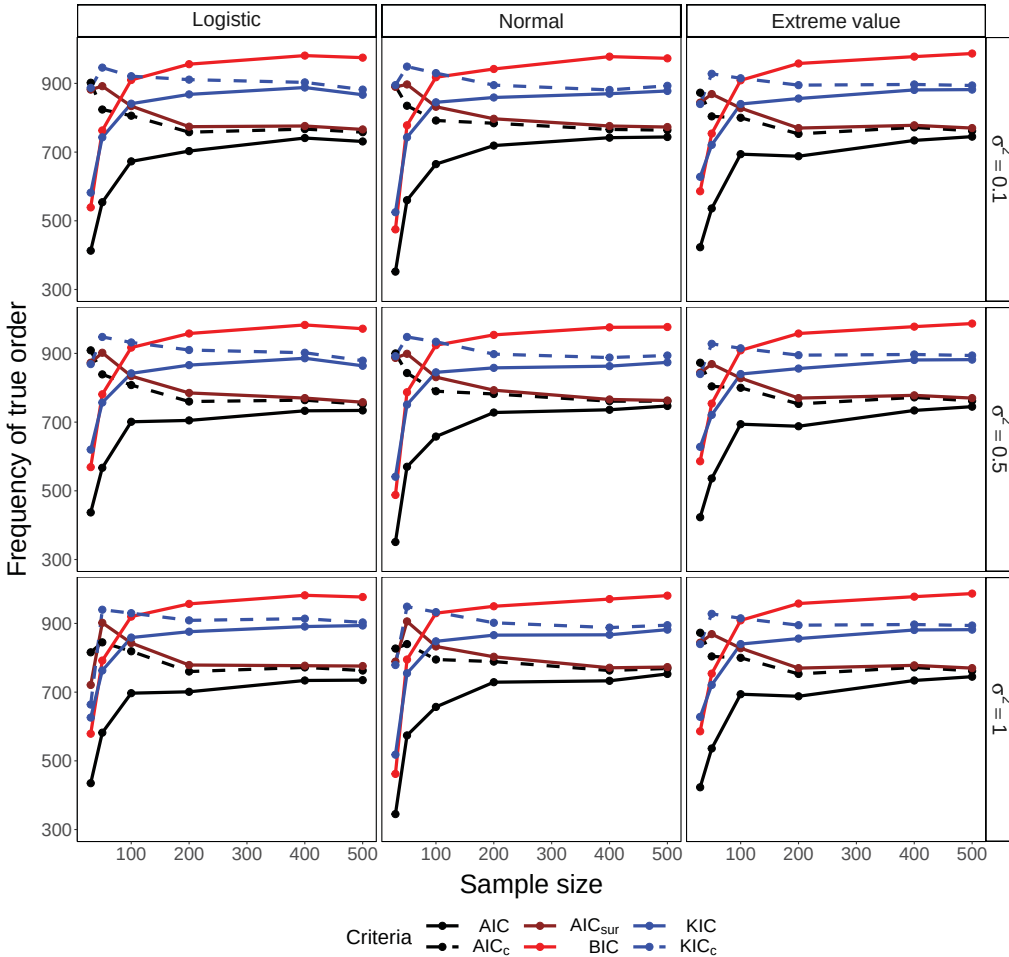


Figure 1: Proportion of true order selections when the censoring percentage is equal 50%.

Figure 2 shows the proportion of true order selection for model selection criteria for various ϵ distributions and σ^2 parameters as a function of sample size when the censoring percentage is 80%. The results are comparable to those obtained when the censoring percentage is 50%. However, the proportions of selection of the true order for all criteria are lower than when the censoring percentage is 50%. Furthermore, the performance of the selection criteria in this case remains unaffected by the value of the scale parameter.

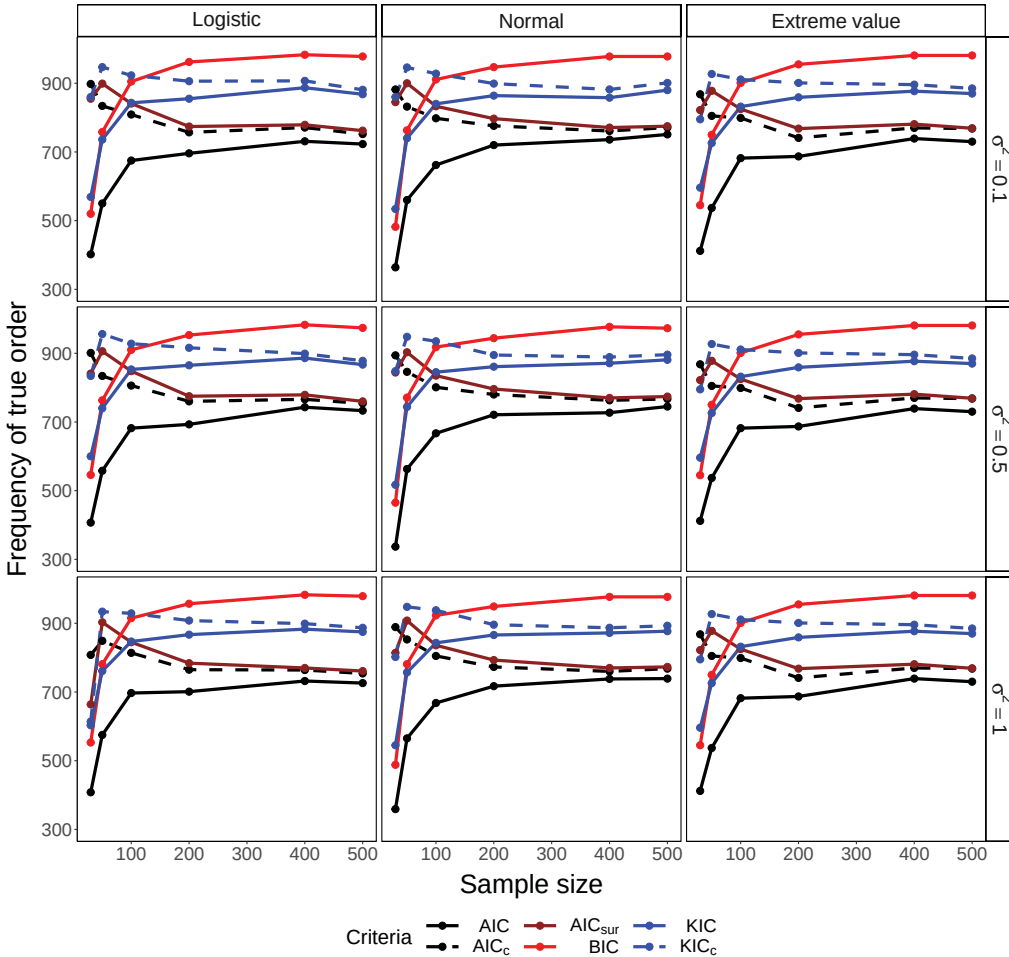


Figure 2: Proportion of true order selections when the censoring percentage is equal to 80%.

Figure 3 shows Root Mean Square Error ($RMSE$) results for different ϵ distributions (logistic, normal, extreme value) and σ^2 parameters as a function of sample size when the censoring percentage is 50%. Whatever the scale value, the KIC and KIC_c criteria show lower median $RMSE$ values except when the error follows an Extreme value distribution, and the sample size is $n = 30$. For small sample sizes and when the error follows a normal and logistic distribution, KIC_c and AIC_c show the lowest median $RMSE$ values. Furthermore, the maximum $RMSE$ values and the variability around the median $RMSE$ values for the KIC family criteria were lower than those for AIC for the same distributions, whatever the scale value and sample size. As the sample size increased, $RMSE$ decreased for all criteria. Furthermore, whatever the model selection criterion, the $RMSE$ value is better when the error follows a Normal distribution. On the other hand, an increase in the value of the scale parameter increases $RMSE$ variability.

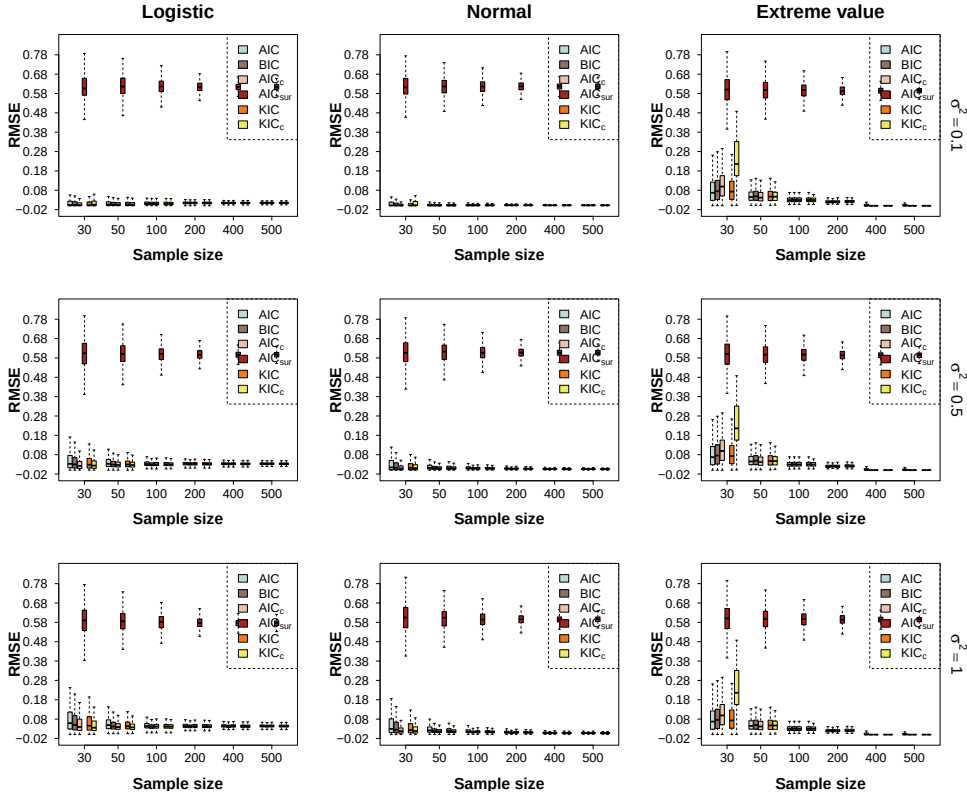


Figure 3: Root Mean Square Error when the censoring percentage is equal to 50%.

Figure 4 shows the root mean square error ($RMSE$) of model selection criteria for different error distributions and scale parameters as a function of sample size, a censoring percentage equal to 80%. Overall trends remain consistent with those observed under 50% censoring percentage, but $RMSE$ values are higher overall, highlighting the significant impact of censoring percentage on criteria performance. The AIC_c and KIC_c criteria continue outperforming the other for small sizes when the error distribution is Normal or Logistic. Still, the criteria performances converge for large sample sizes ($n = 400$ or 500).

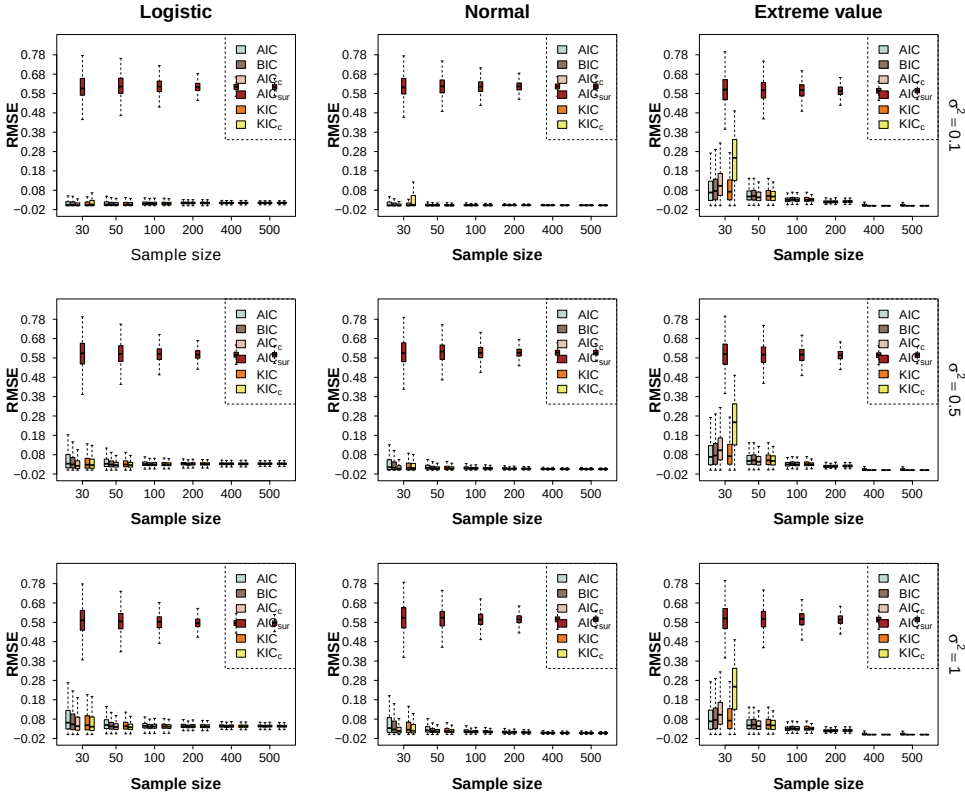


Figure 4: Root Mean Square Error when the censoring percentage is equal to 80%.

4.2 Real dataset analysis

The Log-Normal model had the lowest values of AIC , BIC , and KIC (Table 1). The model, including dam parity, calving status, herd size, floor type, calf housing, calf pen cleaning frequency, first colostrum feeding time, and the milk-fed amount, had the lowest AIC . In contrast, the model with dam parity, calving status, herd size, floor type, calf pen cleaning frequency, and first colostrum feeding time had the lowest AIC_{sur} (Table 2). The model containing dam parity, calving status, herd size, floor type, and calf pen cleaning frequency was minimized BIC and KIC_c , while the model with dam parity, calving status, and floor type had the lowest AIC_c , KIC_c . In general, AIC and AIC_{sur} favored more overparameterized models, while AIC_c and KIC_c favored simpler, under-parameterized models.

Table 1: Values of KIC , BIC , AIC under various models.

Fitted	AIC	BIC	KIC
Weibull	248.85	250.26	250.85
Log-Logistic	248.87	250.28	250.87
Log-Normal	248.72	250.14	250.73

Table 2: Model selection criteria for different variable combinations

Number of variables	Variable(s) included	Criterion					
		AIC	BIC	AIC_c	AIC_{sur}	KIC	KIC_c
1	X_5	205.58	207.71	207.77	206.08	208.58	216.18
2	X_2, X_8	194.17	197.00	198.17	194.90	198.17	208.85
3	X_1, X_2, X_5	188.75	192.29	195.41	189.75	193.74	208.83
4	X_1, X_2, X_4, X_5	185.93	190.18	196.43	187.26	191.93	213.26
5	X_1, X_2, X_4, X_5, X_7	183.33	188.29	199.33	185.08	190.33	220.64
6	$X_1, X_2, X_4, X_5, X_7,$ X_8	182.64	188.31	206.64	184.93	190.64	234.22
7	$X_1, X_2, X_4, X_5, X_7,$ X_8, X_9	182.71	189.08	218.71	185.71	191.71	255.95
8	$X_1, X_2, X_4, X_5, X_6,$ X_7, X_8, X_9	182.31	189.39	237.31	186.31	192.31	291.79
9	$X_1, X_2, X_3, X_4, X_5,$ X_6, X_7, X_8, X_9	183.58	191.36	271.57	189.07	194.57	365.15

5 Discussion

Model selection criteria aim to find an optimal balance between the model's fit to the data and its complexity to avoid overfitting while correctly capturing the underlying structure of the data (Cavanaugh, 1999). Among the most commonly used criteria are the Akaike Information Criterion and its corrected versions. The AIC can favor larger models even when the expected gap between these models and the generating model is high compared to simpler and more parsimonious models, which results in a decrease in the criterion value as the number of parameters in the approximate model increases (Cavanaugh, 1999). The Bayesian Information Criterion (BIC) imposes a higher penalty, favoring more parsimonious models (Farooq and Karami, 2019). Although traditional criteria such as AIC , AIC_c , AIC_{sur} , and BIC are widely

used, it has been revealed that criteria from the KIC family are particularly more effective than those from the AIC family for linear, nonlinear models, and in contexts of overdispersion (Kim and Cavanaugh, 2005; Kim et al., 2014). However, few studies have focused on evaluating the performance of these criteria in the context of survival analysis, specifically in the Accelerated Failure Time model.

In this study, the performance of model selection criteria based on KSD (KIC and KIC_c) compared to AIC , AIC_c , AIC_{sur} , and BIC was evaluated within the framework of Accelerated Failure Time (AFT) models with different error distributions (logistic, normal, extreme value), scale parameter value, censoring percentage, and sample sizes. The objective was to measure the ability of the criteria to identify the true model and select those offering good predictive performance, particularly KIC and KIC_c , compared to traditional criteria under different conditions.

Regardless of the sample size, the censoring percentage, the error distribution, or the scale parameter value, the criteria from the KIC family tended to select overfitted models less frequently. These criteria tend to select overfitted models less frequently than the criteria from the AIC family. This finding is consistent with that obtained by Cavanaugh (2004). This result of the KIC family criteria is identical to that of the BIC because their more penalizing terms are more effective in moderating the risk of overfitting. This result corroborates the conclusions of Kim et al. (2014), which state that the performance of a criterion seems to be dictated by how its penalty term corrects the negative bias in the goodness-of-fit term. It was also noted that the KIC family criteria are more effective in selecting the true order than the AIC family criteria, regardless of the censoring percentage, sample size, scale parameter value, or error distribution. These results are consistent with previous research (Kim and Cavanaugh, 2005; Kim et al., 2014), which demonstrated that, regardless of the sample size, the model selection criteria of the KIC family outperform the model selection criteria of the AIC family for the selection of the true order in nonlinear models and overdispersed Poisson models. However, the KIC and KIC_c criteria are less effective than the BIC for selecting the true order for medium and large sample sizes. We also found that, for small sample sizes, the corrected criteria (AIC_c , AIC_{sur} , KIC_c) were better than the other criteria (KIC , AIC , BIC) in choosing the true model regardless of the error distribution, scale parameter value, or censoring percentage. The corrected criteria were recommended by Anderson and Burnham (2004); Cavanaugh (2004) and Cavanaugh Cavanaugh (2004) to improve model selection for small sample sizes. Moreover, our results showed that the criteria based on KSD selected models with better predictive power when the error distribution is normal and logistic. These observations are consistent with those of Kim and Cavanaugh (2005), who also found similar performance for nonlinear models for KIC and KIC_c . Moreover, as the sample size increases, the performance of all criteria in selecting models with better predictive power converges. Overall, our results show that the distribution of errors, the scale parameter value, the sample size, and the censoring percentage significantly influence the performance of the criteria. When the sample size increases, the performance of KIC , AIC , and BIC improves. Moreover, the model selection criteria perform better when the error follows a normal distribution. Conversely, an increase in the scale parameter value and the censoring percentage systematically leads to a decrease in

the performance of the model selection criteria.

The results of the analysis of real data using AIC , BIC , and KIC indicate that the Log-Normal model is the most appropriate. The selected model, namely dam parity, calving status, herd size, floor type, calf housing, calf pen cleaning frequency, first colostrum feeding time, amount of milk fed, and dam parity, calving status, herd size, floor type, calf pen cleaning frequency, first colostrum feeding time respectively by AIC and AIC_{sur} , were over-parameterized models. This result supports those of the simulation, which indicates that the symmetric Kullback divergence can be greater than the directed divergence because it better reflects the risks of overfitting. Moreover, it is observed that KIC_c and AIC_c have underestimated the model, which confirms that the corrected criteria are suitable for small sample sizes since the actual data size is 235. The results of the two criteria, namely KIC and BIC , uniformly indicate that the model with dam parity, calving status, floor type, calf pen cleaning frequency, and herd size is the most appropriate. This confirmation supports Ahmedin and Assen (2023) choice, indicating that the model with dam parity, calving status, floor type, and calf pen cleaning frequency is appropriate for these data. The results obtained highlight key implications for users when choosing model selection criteria. In small sample contexts, common in clinical or experimental studies, corrected criteria (KIC_c and AIC_c) become essential to avoid misinterpretation. Furthermore, in contexts where the risk of overfitting is high, the AIC and AIC_{sur} criteria, which tend to select overly complex models, may not be appropriate. The KIC and BIC criteria are an effective compromise for identifying a relevant model that is both parsimonious and explanatory. Notably, when prediction is the main objective, criteria such as KIC or KIC_c outperform others.

It is significant that although some of the criteria evaluated can be applied to other modeling frameworks, such as semi-parametric AFT models or random effects models, this comparison is limited to parametric AFT models. This restriction means that the performance of the criteria in semi-parametric or random effects contexts has not been explored. Furthermore, the study did not examine the impact of informative censoring, which could alter the behavior of the criteria in more complex situations. These limitations underscore the need for caution in generalizing the results, making us all more aware and conscious of the potential implications, and open up prospects for more extensive future work.

Conclusion

This study evaluated the performance of the KIC family criteria against AIC , AIC_c , AIC_{sur} , and BIC for the Parametric Accelerated Failure Time model. The KIC_c and KIC criteria show promise as model selection criteria compared with the AIC , AIC_c , and AIC_{sur} criteria. However, they do not outperform BIC . At the end of this study, our results allow us to draw three conclusions about model selection criteria for the Parametric Accelerated Failure Time model. First, we prefer KIC or KIC_c to the AIC family criteria. Secondly, for small sample sizes, we should prefer adjusted criteria (KIC_c , AIC_c , AIC_{sur}) to unadjusted ones. Finally, how well a criterion's penalty term corrects for the negative bias in the goodness-of-fit term dictates its performance. Future work could explore the performance of these criteria in the

semi-parametric Accelerated Failure Time model.

Acknowledgments

This research received support from the DAAD In-Country/In-Region Scholarship Programme FSA/UAC (Grant no. 57588035, 2021), the African Excellence Center in Mathematical Sciences and Applications at the Institute of Mathematics and Physical Sciences in Benin, and Mawazo Fellowship Programme.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Appendix A. All results.Table 3: Frequency of order selected and $RMSE$ for different criteria when ϵ following the Logistic distribution.

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
30	0.1	50	order $< p$	0	0	0	0	0	2
			True order	413	902	539	882	582	887
			order $> p$	587	98	461	118	418	111
			Average RMSE	0.016	0.612	0.014	0.009	0.013	0.033
		80	order $< p$	0	0	0	0	0	3
			True order	396	886	511	843	560	848
			order $> p$	591	101	476	144	427	136
			Average RMSE	0.016	0.611	0.016	0.01	0.014	0.038
	0.5	50	order $< p$	0	6	0	11	0	21
			True order	437	909	569	873	620	869
			order $> p$	563	85	431	116	380	110
			Average RMSE	0.056	0.601	0.051	0.036	0.048	0.053
		80	order $< p$	0	5	0	12	0	29
			True order	407	901	546	841	600	834
			order $> p$	593	94	454	147	400	137
			Average RMSE	0.057	0.601	0.051	0.038	0.048	0.067
50	0.1	50	order $< p$	10	99	14	163	26	231
			True order	435	816	579	721	626	664
			order $> p$	555	85	407	116	348	105
			Average RMSE	0.081	0.588	0.072	0.06	0.068	0.071
		80	order $< p$	10	112	16	182	28	244
			True order	408	808	553	664	603	614
			order $> p$	582	80	431	154	369	142
			Average RMSE	0.086	0.077	0.588	0.067	0.072	0.087
	0.5	50	order $< p$	0	0	0	0	0	0
			True order	554	824	763	892	743	946
			order $> p$	446	176	237	108	257	54
			Average RMSE	0.013	0.614	0.011	0.01	0.011	0.01
		80	order $< p$	0	0	0	0	0	0
			True order	550	834	758	899	736	947
			order $> p$	450	166	242	101	264	53
			Average RMSE	0.013	0.615	0.011	0.01	0.011	0.01

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
1	0.1	80	order < p	0	0	0	0	0	0
			True order	558	834	763	906	740	956
			order > p	442	166	237	94	260	44
			Average RMSE	0.042	0.601	0.036	0.031	0.036	0.03
	0.5	50	order < p	1	4	3	7	3	13
			True order	582	845	792	902	763	940
			order > p	417	151	205	91	234	47
			Average RMSE	0.058	0.585	0.05	0.044	0.05	0.042
	1	80	order < p	0	5	4	9	2	19
			True order	575	849	781	903	761	934
			order > p	425	146	215	88	237	47
			Average RMSE	0.058	0.584	0.05	0.044	0.05	0.042
100	0.1	50	order < p	0	0	0	0	0	0
			True order	673	806	910	834	841	921
			order > p	327	194	90	166	159	79
			Average RMSE	0.013	0.614	0.012	0.013	0.013	0.012
	0.5	80	order < p	0	0	0	0	0	0
			True order	675	809	905	841	843	923
			order > p	325	191	95	159	157	77
			Average RMSE	0.013	0.615	0.012	0.012	0.012	0.012
	1	50	order < p	0	0	0	0	0	0
			True order	701	808	917	834	842	932
			order > p	299	192	83	166	158	68
			Average RMSE	0.034	0.031	0.598	0.032	0.032	0.031
	1	80	order < p	0	0	0	0	0	0
			True order	682	806	910	848	853	928
			order > p	318	194	90	152	147	72
			Average RMSE	0.034	0.599	0.032	0.032	0.032	0.031
200	0.1	50	order < p	0	0	0	0	0	0
			True order	697	819	920	843	859	930
			order > p	303	181	80	157	141	70
			Average RMSE	0.047	0.58	0.043	0.044	0.044	0.043
	0.5	80	order < p	0	0	0	0	0	0
			True order	697	814	915	845	847	929
			order > p	303	186	85	155	153	71
			Average RMSE	0.047	0.580	0.043	0.044	0.044	0.0433
	1	50	order < p	0	0	0	0	0	0
			True order	703	758	956	774	868	911
			order > p	297	242	44	226	132	89
			Average RMSE	0.015	0.614	0.014	0.015	0.014	0.014

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
		80	order < p	0	0	0	0	0	0
			True order	696	757	962	774	855	906
			order > p	304	243	38	226	145	94
			Average RMSE	0.015	0.613	0.014	0.014	0.014	0.014
	0.5	50	order < p	0	0	0	0	0	0
			True order	705	760	958	785	866	910
			order > p	295	240	42	215	134	90
			Average RMSE	0.034	0.596	0.032	0.033	0.033	0.032
		80	order < p	0	0	0	0	0	0
			True order	693	760	953	775	865	916
			order > p	307	240	47	225	135	84
			Average RMSE	0.034	0.596	0.032	0.033	0.033	0.032
1		50	order < p	0	0	0	0	0	0
			True order	701	760	957	779	876	909
			order > p	299	240	43	221	124	91
			Average RMSE	0.046	0.577	0.043	0.045	0.044	0.044
		80	order < p	0	0	0	0	0	0
			True order	701	765	957	784	867	908
			order > p	299	235	43	216	133	92
			Average RMSE	0.046	0.577	0.043	0.045	0.044	0.044
400	0.1	50	order < p	0	0	0	0	0	0
			True order	741	767	981	776	888	903
			order > p	259	233	19	224	112	97
			Average RMSE	0.015	0.614	0.015	0.015	0.015	0.015
		80	order < p	0	0	0	0	0	0
			True order	731	771	983	779	887	907
			order > p	269	229	17	221	113	93
			Average RMSE	0.015	0.614	0.015	0.015	0.015	0.015
	0.5	50	order < p	0	0	0	0	0	0
			True order	733	764	983	770	886	902
			order > p	267	236	17	230	114	98
			Average RMSE	0.033	0.596	0.033	0.033	0.033	0.033
		80	order < p	0	0	0	0	0	0
			True order	743	766	983	779	886	899
			order > p	257	234	17	221	114	101
			Average RMSE	0.033	0.596	0.033	0.033	0.033	0.033
	1	50	order < p	0	0	0	0	0	0
			True order	734	772	982	777	891	914
			order > p	266	228	18	223	109	86
			Average RMSE	0.045	0.576	0.044	0.045	0.044	0.044

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
500	0.1	80	order < p	0	0	0	0	0	0
			True order	732	764	983	770	883	899
			order > p	268	236	17	230	117	101
			Average RMSE	0.045	0.576	0.043	0.045	0.044	0.044
		50	order < p	0	0	0	0	0	0
			True order	731	758	975	766	867	882
			order > p	269	242	25	234	133	118
			Average RMSE	0.015	0.614	0.015	0.015	0.015	0.015
	0.5	80	order < p	0	0	0	0	0	0
			True order	723	752	978	762	868	881
			order > p	277	248	22	238	132	119
			Average RMSE	0.015	0.614	0.015	0.015	0.015	0.015
		50	order < p	0	0	0	0	0	0
			True order	734	752	972	758	864	879
			order > p	266	248	28	242	136	121
			Average RMSE	0.033	0.596	0.033	0.033	0.033	0.033
1	0.1	80	order < p	0	0	0	0	0	0
			True order	733	754	974	760	867	878
			order > p	267	246	26	240	133	122
			Average RMSE	0.033	0.596	0.033	0.033	0.033	0.033
		50	order < p	0	0	0	0	0	0
			True order	735	763	977	776	894	903
			order > p	265	237	23	224	106	97
			Average RMSE	0.045	0.044	0.576	0.045	0.044	0.044
	0.5	80	order < p	0	0	0	0	0	0
			True order	726	754	979	761	875	887
			order > p	274	246	21	239	125	113
			Average RMSE	0.045	0.044	0.576	0.044	0.044	0.044

Table 4: Frequency of order selected and $RMSE$ for different criteria when ϵ following the Normal distribution.

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
30	0.1	50	order $< p$	0	0	0	0	0	0
			True order	352	892	475	890	525	895
			order $> p$	648	108	525	110	475	105
			Average RMSE	0.009	0.616	0.007	0.005	0.007	0.032
		80	order $< p$	0	0	0	0	0	1
			True order	336	888	458	849	505	852
			order $> p$	664	112	542	151	495	147
			Average RMSE	0.009	0.615	0.008	0.005	0.008	0.04
	0.5	50	order $< p$	0	0	0	1	0	2
			True order	351	900	488	887	541	892
			order $> p$	649	100	512	112	459	106
			Average RMSE	0.035	0.609	0.031	0.018	0.029	0.047
		80	order $< p$	0	0	0	2	0	3
			True order	337	894	465	844	517	848
			order $> p$	663	106	535	154	483	149
			Average RMSE	0.037	0.609	0.033	0.0205	0.03	0.052
	1	50	order $< p$	1	13	2	23	3	36
			True order	345	827	462	789	518	779
			order $> p$	578	84	460	112	403	109
			Average RMSE	0.054	0.603	0.049	0.03	0.045	0.057
		80	order $< p$	2	12	2	30	3	47
			True order	359	889	488	814	545	802
			order $> p$	639	99	510	156	452	151
			Average RMSE	0.059	0.601	0.052	0.035	0.049	0.067
50	0.1	50	order $< p$	0	0	0	0	0	0
			True order	560	835	778	897	743	949
			order $> p$	440	165	222	103	257	51
			Average RMSE	0.005	0.615	0.004	0.003	0.004	0.003
		80	order $< p$	0	0	0	0	0	0
			True order	560	832	763	900	740	946
			order $> p$	440	168	237	100	260	54
			Average RMSE	0.005	0.615	0.004	0.003	0.004	0.003
	0.5	50	order $< p$	0	0	0	0	0	0
			True order	570	843	787	899	751	948
			order $> p$	430	157	213	101	249	52
			Average RMSE	0.018	0.607	0.015	0.012	0.015	0.012

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
1	80		order < p	0	0	0	0	0	0
			True order	563	846	771	903	744	948
			order > p	437	154	229	97	256	52
			Average RMSE	0.019	0.607	0.015	0.013	0.016	0.012
	50		order < p	0	0	0	0	0	0
			True order	574	840	795	906	755	949
			order > p	426	160	205	94	245	51
			Average RMSE	0.029	0.598	0.024	0.02	0.025	0.019
	80		order < p	0	0	0	0	0	0
			True order	565	853	781	908	757	948
			order > p	435	147	219	92	243	52
			Average RMSE	0.03	0.598	0.025	0.02	0.025	0.02
100	0.1	50	order < p	0	0	0	0	0	0
			True order	665	792	918	832	845	930
			order > p	335	208	82	168	155	70
			Average RMSE	0.004	0.615	0.003	0.003	0.003	0.003
	80		order < p	0	0	0	0	0	0
			True order	662	798	911	833	840	928
			order > p	338	202	89	167	160	72
			Average RMSE	0.004	0.615	0.003	0.003	0.003	0.003
	0.5	50	order < p	0	0	0	0	0	0
			True order	658	790	924	831	845	934
			order > p	342	210	76	169	155	66
			Average RMSE	0.013	0.606	0.01	0.011	0.011	0.01
	80		order < p	0	0	0	0	0	0
			True order	667	801	918	835	845	935
			order > p	333	199	82	165	155	65
			Average RMSE	0.013	0.606	0.01	0.011	0.011	0.01
	1	50	order < p	0	0	0	0	0	0
			True order	657	795	930	833	848	933
			order > p	343	205	70	167	152	67
			Average RMSE	0.019	0.595	0.015	0.016	0.016	0.015
	80		order < p	0	0	0	0	0	0
			True order	668	805	923	836	843	938
			order > p	332	195	77	164	157	62
			Average RMSE	0.02	0.595	0.015	0.017	0.017	0.015

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
200	0.1	50	order < p	0	0	0	0	0	0
			True order	719	784	942	797	859	895
			order > p	281	216	58	203	141	105
			Average RMSE	0.003	0.616	0.002	0.003	0.003	0.003
		80	order < p	0	0	0	0	0	0
			True order	720	776	947	797	864	899
			order > p	280	224	53	203	136	101
			Average RMSE	0.003	0.615	0.003	0.003	0.003	0.003
	0.5	50	order < p	0	0	0	0	0	0
			True order	728	782	954	793	858	898
			order > p	272	218	46	207	142	102
			Average RMSE	0.009	0.606	0.007	0.008	0.008	0.008
		80	order < p	0	0	0	0	0	0
			True order	721	780	944	796	861	895
			order > p	279	220	56	204	139	105
			Average RMSE	0.009	0.606	0.008	0.009	0.008	0.008
	1	50	order < p	0	0	0	0	0	0
			True order	729	789	950	803	866	902
			order > p	271	211	50	197	134	98
			Average RMSE	0.013	0.595	0.011	0.012	0.012	0.011
		80	order < p	0	0	0	0	0	0
			True order	717	773	949	793	866	896
			order > p	283	227	51	207	134	104
			Average RMSE	0.013					
400	0.1	50	order < p	0	0	0	0	0	0
			True order	742	766	978	776	870	881
			order > p	258	234	22	224	130	119
			Average RMSE	0.002	0.616	0.002	0.002	0.002	0.002
		80	order < p	0	0	0	0	0	0
			True order	736	761	978	771	858	882
			order > p	264	239	22	229	142	118
			Average RMSE	0.002	0.616	0.002	0.002	0.002	0.002
	0.5	50	order < p	0	0	0	0	0	0
			True order	736	761	976	766	863	888
			order > p	264	239	24	234	137	112
			Average RMSE	0.006	0.607	0.005	0.006	0.006	0.006
		80	order < p	0	0	0	0	0	0
			True order	727	763	977	770	871	889
			order > p	273	237	23	230	129	111
			Average RMSE	0.007	0.607	0.005	0.006	0.006	0.006

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
500	1	50	order < p	0	0	0	0	0	0
			True order	733	763	971	771	867	888
			order > p	267	237	29	229	133	112
			Average RMSE	0.01	0.595	0.008	0.009	0.009	0.009
		80	order < p	0	0	0	0	0	0
			True order	738	760	977	770	872	887
			order > p	262	240	23	230	128	113
			Average RMSE	0.01	0.595	0.008	0.009	0.009	0.009
	0.1	50	order < p	0	0	0	0	0	0
			True order	744	764	973	773	878	893
			order > p	256	236	27	227	122	107
			Average RMSE	0.002	0.617	0.002	0.002	0.002	0.002
		80	order < p	0	0	0	0	0	0
			True order	751	771	978	775	880	901
			order > p	249	229	22	225	120	99
			Average RMSE	0.002	0.617	0.002	0.002	0.002	0.002
500	0.5	50	order < p	0	0	0	0	0	0
			True order	747	761	977	763	874	894
			order > p	253	239	23	237	126	106
			Average RMSE	0.006	0.607	0.005	0.006	0.005	0.005
		80	order < p	0	0	0	0	0	0
			True order	745	767	973	774	881	896
			order > p	255	233	27	226	119	104
			Average RMSE	0.006	0.607	0.005	0.006	0.005	0.005
	1	50	order < p	0	0	0	0	0	0
			True order	753	769	981	773	882	895
			order > p	247	231	19	227	118	105
			Average RMSE	0.008	0.595	0.007	0.008	0.008	0.008
		80	order < p	0	0	0	0	0	0
			True order	739	768	977	773	877	893
			order > p	261	232	23	227	123	107
			Average RMSE	0.009	0.595	0.007	0.008	0.008	0.008

Table 5: Frequency of order selected and $RMSE$ for different criteria when ϵ following the Extreme value distribution.

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
30	0.1	50	order $< p$	1	20	3	54	4	73
			True order	423	873	586	844	628	840
			order $> p$	576	107	411	102	368	87
			Average RMSE	0.048	0.597	0.037	0.015	0.032	0.022
		80	order $< p$	2	25	3	66	4	100
			True order	412	868	545	822	596	795
			order $> p$	586	107	452	112	400	105
			Average RMSE	0.051	0.597	0.041	0.019	0.036	0.032
	0.5	50	order $< p$	1	20	3	54	4	73
			True order	423	873	586	844	628	840
			order $> p$	576	107	411	102	368	87
			Average RMSE	0.048	0.597	0.037	0.015	0.032	0.022
		80	order $< p$	2	25	3	66	4	100
			True order	412	868	545	822	596	795
			order $> p$	586	107	452	112	400	105
			Average RMSE	0.051	0.597	0.041	0.019	0.036	0.032
	1	50	order $< p$	1	20	3	54	4	73
			True order	423	873	586	844	628	840
			order $> p$	576	107	411	102	368	87
			Average RMSE	0.048	0.597	0.037	0.015	0.032	0.022
		80	order $< p$	2	25	3	66	4	100
			True order	412	868	545	822	596	795
			order $> p$	586	107	452	112	400	105
			Average RMSE	0.051	0.597	0.041	0.019	0.036	0.032
50	0.1	50	order $< p$	0	0	0	0	0	1
			True order	536	804	754	869	721	928
			order $> p$	464	196	246	131	279	71
			Average RMSE	0.024	0.597	0.012	0.005	0.014	0.002
		80	order $< p$	1	1	1	1	1	3
			True order	537	805	750	878	726	927
			order $> p$	462	194	249	121	273	70
			Average RMSE	0.025	0.596	0.013	0.005	0.015	0.002
	0.5	50	order $< p$	0	0	0	0	0	1
			True order	536	804	754	869	721	928
			order $> p$	464	196	246	131	279	71
			Average RMSE	0.024	0.597	0.012	0.005	0.014	0.002

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
1	80		order < p	1	1	1	1	1	3
			True order	537	805	750	878	726	927
			order > p	462	194	249	121	273	70
			Average RMSE	0.025	0.596	0.013	0.005	0.015	0.002
	50		order < p	0	0	0	0	0	1
			True order	536	804	754	869	721	928
			order > p	464	196	246	131	279	71
			Average RMSE	0.024	0.597	0.012	0.005	0.014	0.002
	80		order < p	0	4	2261	713	19	3
			True order	0	1	750	249	0	0
			order > p	0	7	3451	542	0	0
			Average RMSE	0.025	0.596	0.013	0.005	0.015	0.002
100	0.1	50	order < p	0	0	0	0	0	0
			True order	694	800	909	828	840	915
			order > p	306	200	91	172	160	85
			Average RMSE	0.009	0.595	0.002	0.005	0.004	0.002
		80	order < p	0	0	0	0	0	0
			True order	682	799	901	825	832	911
			order > p	318	201	99	175	168	89
			Average RMSE	0.009	0.595	0.003	0.005	0.005	0.002
	0.5	50	order < p	0	0	0	0	0	0
			True order	682	799	901	825	832	911
			order > p	318	201	99	175	168	89
			Average RMSE	0.009	0.595	0.002	0.005	0.004	0.002
		80	order < p	0	0	0	0	0	0
			True order	682	799	901	825	832	911
			order > p	318	201	99	175	168	89
			Average RMSE	0.009	0.595	0.003	0.005	0.005	0.002
	1	50	order < p	0	0	0	0	0	0
			True order	694	800	909	828	840	915
			order > p	306	200	91	172	160	85
			Average RMSE	0.009	0.595	0.002	0.005	0.004	0.002
		80	order < p	0	0	0	0	0	0
			True order	682	799	901	825	832	911
			order > p	318	201	99	175	168	89
			Average RMSE	0.009	0.595	0.003	0.005	0.005	0.002

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
200	0.1	50	order < p	0	0	0	0	0	0
			True order	688	753	958	770	856	895
			order > p	312	247	42	230	144	105
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
		80	order < p	0	0	0	0	0	0
			True order	687	741	955	768	859	901
			order > p	313	259	45	232	141	99
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
		50	order < p	0	0	0	0	0	0
			True order	688	753	958	770	856	895
			order > p	312	247	42	230	144	105
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
	0.5	80	order < p	0	0	0	0	0	0
			True order	687	741	955	768	859	901
			order > p	313	259	45	232	141	99
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
		50	order < p	0	0	0	0	0	0
			True order	688	753	958	770	856	895
			order > p	312	247	42	230	144	105
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
	1	80	order < p	0	0	0	0	0	0
			True order	687	741	955	768	859	901
			order > p	313	259	45	232	141	99
			Average RMSE	0.006	0.593	0.001	0.004	0.003	0.002
		50	order < p	0	0	0	0	0	0
			True order	734	772	978	778	881	897
			order > p	266	228	22	222	119	103
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
	0.1	80	order < p	0	0	0	0	0	0
			True order	739	770	981	781	877	896
			order > p	261	230	19	219	123	104
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
		50	order < p	0	0	0	0	0	0
			True order	734	772	978	778	881	897
			order > p	266	228	22	222	119	103
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
	0.5	80	order < p	0	0	0	0	0	0
			True order	739	770	981	781	877	896
			order > p	261	230	19	219	123	104
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
		50	order < p	0	0	0	0	0	0
			True order	734	772	978	778	881	897
			order > p	266	228	22	222	119	103
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001

n	σ^2	C	Order	AIC	AIC_{sur}	BIC	AIC_c	KIC	KIC_c
1	1	50	order < p	0	0	0	0	0	0
			True order	734	772	978	778	881	897
			order > p	266	228	22	222	119	103
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
		80	order < p	0	0	0	0	0	0
			True order	739	770	981	781	877	896
			order > p	261	230	19	219	123	104
			Average RMSE	0.003	0.594	0.0003	0.003	0.001	0.001
	0.1	50	order < p	0	0	0	0	0	0
			True order	745	762	987	770	882	894
			order > p	255	238	13	230	118	106
			Average RMSE	0.002	0.594	0.0001	0.002	0.001	0.001
		80	order < p	0	0	0	0	0	0
			True order	730	768	981	769	870	885
			order > p	270	232	19	231	130	115
			Average RMSE	0.003	0.594	0.0002	0.002	0.001	0.001
500	0.5	50	order < p	0	0	0	0	0	0
			True order	745	762	987	770	882	894
			order > p	255	238	13	230	118	106
			Average RMSE	0.002	0.594	0.0001	0.002	0.001	0.001
		80	order < p	0	0	0	0	0	0
			True order	730	768	981	769	870	885
			order > p	270	232	19	231	130	115
			Average RMSE	0.003	0.594	0.0002	0.002	0.001	0.001
	1	50	order < p	0	0	0	0	0	0
			True order	745	762	987	770	882	894
			order > p	255	238	13	230	118	106
			Average RMSE	0.002	0.594	0.0001	0.002	0.001	0.001
		80	order < p	0	0	0	0	0	0
			True order	730	768	981	769	870	885
			order > p	270	232	19	231	130	115
			Average RMSE	0.003	0.594	0.0002	0.002	0.001	0.001

References

- Cristina Laura Acion. *Criteria for generalized linear model selection based on Kullback's symmetric divergence*. PhD thesis, University of Iowa, 2011.
- Umer Mifta Ahmedin and Alula Alemayehu Assen. Calf morbidity, mortality, and management practices in dairy farms in jimma city, southwestern ethiopia. *BMC Veterinary Research*, 19 (1):249, 2023.
- Hirotsugu Akaike. Information theory and an extension of the maximum likelihood princi-

- ple. In *Proc. 2nd International Symposium on Information Theory, 1973*, pages 267–281. Akademiai Kiado, 1973.
- Hirotsugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- G Ambler, S Seaman, and RZ Omar. An evaluation of penalised survival methods for developing prognostic models with rare events. *Statistics in medicine*, 31(11-12):1150–1161, 2012.
- D Anderson and K Burnham. Model selection and multi-model inference. *Second. NY: Springer-Verlag*, 63(2020):10, 2004.
- Nourddine Azzaoui and Bezza Hafidi. Criteria for longitudinal data model selection based on kullback’s symmetric divergence. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 15, 2012.
- Aishwarya Bhaskaran, Ding Ma, Benoit Lique, Angela Hong, Serigne N Lo, Stephane Heritier, and Jun Ma. A maximum penalised likelihood approach for semiparametric accelerated failure time models with time-varying covariates and partly interval censoring. *arXiv preprint arXiv:2403.12332*, 2024.
- Joseph E Cavanaugh. A large-sample model selection criterion based on kullback’s symmetric divergence. *Statistics & Probability Letters*, 42(4):333–343, 1999.
- Joseph E Cavanaugh. Criteria for linear model selection based on kullback’s symmetric divergence. *Australian & New Zealand Journal of Statistics*, 46(2):257–274, 2004.
- Joseph E Cavanaugh and Andrew A Neath. The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3):e1460, 2019.
- Dongrak Choi, Woojung Bae, Jun Yan, and Sangwook Kang. A general model-checking procedure for semiparametric accelerated failure time models. *Statistics and Computing*, 34(3): 117, 2024.
- David Collett. *Modelling survival data in medical research*. Chapman and Hall/CRC, 2023.
- Salil Vasudeo Deo, Vaishali Deo, and Varun Sundaram. Survival analysis—part 2: Cox proportional hazards model. *Indian journal of thoracic and cardiovascular surgery*, 37:229–233, 2021.
- Lore Dirick, Gerda Claeskens, and Bart Baesens. An akaike information criterion for multiple event mixture cure models. *European Journal of Operational Research*, 241(2):449–457, 2015.
- Fabiha Binte Farooq and Md Jamil Hasan Karami. Model selection strategy for cox proportional hazards model. *Dhaka University Journal of Science*, 67(2):111–116, 2019.

- Nicholas C Henderson, Thomas A Louis, Gary L Rosner, and Ravi Varadhan. Individualized treatment effects with censored data via fully nonparametric bayesian accelerated failure time models. *Biostatistics*, 21(1):50–68, 2020.
- Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.
- Fei Jiang and Elan Guterman. Survival analysis. *Statistical Methods in Epilepsy*, pages 124–142, 2024.
- Warangkhan Keerativibool and Pachitjanut Siripanich. Comparison of the model selection criteria for multiple regression based on kullback-leibler’s information. *Chiang Mai Journal of Science*, 44(2):699–714, 2017.
- Hyun-Joo Kim and Joseph E Cavanaugh. Model selection criteria based on kullback information measures for nonlinear regression. *Journal of statistical planning and inference*, 134(2):332–349, 2005.
- Hyun-Joo Kim, Joseph E Cavanaugh, Tad A Dallas, and Stephanie A Foré. Model selection criteria for overdispersed data and their application to the characterization of a host-parasite relationship. *Environmental and ecological statistics*, 21(2):329–350, 2014.
- Hua Liang and Guohua Zou. Improved aic selection strategy for survival analysis. *Computational statistics & data analysis*, 52(5):2538–2548, 2008.
- Hua Liang, Hulin Wu, and Guohua Zou. A note on conditional aic for linear mixed-effects models. *Biometrika*, 95(3):773–778, 2008.
- Yuhui Lin. Aft survival model to capture the rate of aging and age-specific mortality trajectories among first-allogeneic hematopoietic stem cells transplant patients. *Plos one*, 13(3):e0193287, 2018.
- Abdul-Fatawu Majeed. Accelerated failure time models: an application in insurance attrition. *The Journal of Risk Management and Insurance*, 2020.
- Adam Braima S Mastor, Abdulaziz S Alghamdi, Oscar Ngesa, Joseph Mung’atu, Christophe Chesneau, and Ahmed Z Afify. The extended exponential-weibull accelerated failure time model with application to sudan covid-19 data. *Mathematics*, 11(2):460, 2023.
- Roman Mukin, Kai Hencken, and Lorenzo Privitera. Analysis of the use of field data under variable conditions to develop lifetime models for electrical distribution devices. 2023.
- Yesuf Abdela Mustefa and Ding-Geng Chen. Accelerated failure-time model with weighted least-squares estimation: application on survival of hiv positives. *Archives of Public Health*, 79(1):88, 2021.
- Oded Netzer, Peter Ebbes, and Tammo HA Bijmolt. Hidden markov models in marketing. *Advanced methods for modeling markets*, pages 405–449, 2017.

- Noureddine Ouerfelli, Nyarko Christiana Cynthia, Acquah Joseph, and Bosson-Amedenu Senyefia. A comparison between accelerated failure time models in analyzing the survival of breast cancer patients. *Journal of Cancer and Tumor International*, pages 16–28, 2022.
- Warangkhan Riansut. New criterion for selection in regression model. *Thailand Statistician*, 21(1):48–67, 2023.
- Rinku Saikia and Manash Pratim Barman. A review on accelerated failure time models. *Int J Stat Syst*, 12(2):311–322, 2017.
- A-K Seghouane and Maiza Bekara. A small sample model selection criterion based on kullback's symmetric divergence. *IEEE transactions on signal processing*, 52(12):3314–3323, 2004.
- Byungtae Seo and Sangwook Kang. Accelerated failure time modeling via nonparametric mixtures. *Biometrics*, 79(1):165–177, 2023.
- Terry M Therneau, Patricia M Grambsch, and V Shane Pankratz. Penalized survival models and frailty. *Journal of computational and graphical statistics*, 12(1):156–175, 2003.
- Tingxuan Wu. Randomized survival probability residual for assessing parametric survival models. *University of Saskatchewan*, 2018.