NIH Public Access
Author Manuscript
*Comput Stat Data Anal*. Author manuscript; available in PMC 2009 January 20.

Published in final edited form as:
*Comput Stat Data Anal*. 2008 January 20; 52(5): 2538–2548.

# Improved AIC Selection Strategy for Survival Analysis

**Hua Liang**[1] and **Guohua Zou**[1,2]

1*Department of Biostatistics and Computational Biology University of Rochester Medical Center Rochester, NY 14642, USA hliang@bst.rochester.edu*

2*Academy of Mathematics and Systems Science Chinese Academy of Sciences Beijing 100080, China*

## SUMMARY

In survival analysis, it is of interest to appropriately select significant predictors. In this paper, we extend the $\mathrm{AIC}_C$ selection procedure of Hurvich and Tsai to survival models to improve the traditional AIC for small sample sizes. A theoretical verification under a special case of the exponential distribution is provided. Simulation studies illustrate that the proposed method substantially outperforms its counterpart: AIC, in small samples, and competes it in moderate and large samples. Two real data sets are also analyzed.

### Keywords

AIC; BIC; Kullback-Leibler information; survival analysis

## 1 Introduction

In clinical trials, biological and biomedical applications, many variables may be available for the initial analysis, and spurious covariates may increase prediction error. Deciding which covariates to be kept in the statistical model has always been a tricky task for data analysis. Conventional variable selection techniques, such as AIC (Akaike, 1974), BIC (Schwarz, 1978), and $C_p$ (Mallows, 1973), have widely been used to select an appropriate model. These criteria work well and are implemented in the most well-developed statistical software such as R and SAS. Their deficiency in small samples was pointed out by Sugiura (1978) and emphasized by Hurvich and Tsai (1989). The latter authors showed that AIC may be drastically biased for the linear model, and developed a modified version, $\mathrm{AIC}_C$, which is nearly unbiased for estimating Kullback-Leibler information and provides better model choices than AIC in small samples. Tsai and his colleagues generalized Hurvich and Tsai's criterion to diverse situations like the extended quasi-likelihood model (Hurvich and Tsai, 1995), the nonparametric regression (Hurvich et al., 1998), and the semiparametric regression (Hurvich and Tsai, 1999).

Traditional variable selection criteria such as AIC and BIC have been extended to survival analysis. Faraggi and Simon (1998) proposed a Bayesian variable selection method, which is an extension of Lindley's (1968) variable selection criterion for the linear model, for censored data based on the sufficiency and asymptotic normality of the maximum partial likelihood estimator. Volinsky and Raftery (2000) extended the BIC to the Cox model. They proposed a modification of the penalty term in the BIC so that it is defined in terms of the number of

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

uncensored events instead of the number of observations. Tibshirani (1997) extended his LASSO variable selection procedure to the Cox model. More recently, Fan and Li (2002) derived a nonconcave penalized partial likelihood for the Cox model and the Cox frailty model. Although all of these approaches have been demonstrated to be promising, they may not be accepted in practice because (i) the computation of some methods is not simple and sometimes has a requirement of determining prior information, and (ii) few existing computation packages (to the best of our knowledge, there is a package glmpath for the LASSO algorithm) have been developed for practitioners' use. The aim of this paper is to fill this gap. We propose here an improved AIC variable selection method for survival analysis. This work is motivated by Hurvich and Tsai (1989), whose focus is on linear models. We extend Hurvich and Tsai's approach to survival models and numerically justify the superiority of the proposed criterion over other traditional criteria in small sample sizes. The proposed method can be implemented in the existing software, such as R/Splus and SAS. This availability may make the method easily implement in practice.

The rest of the paper is organized as follows. In Section 2 we propose an improved AIC selection procedure for survival models. A particular case of the exponential distribution for the survival time is considered, which serves as a theoretical justification of the proposed criterion. Section 3 gives the results of extensive simulation experiments to illustrate the proposed method, and compare it with its competitors. Two real examples are examined in Section 4. We conclude the paper with some discussions in Section 5. Technical details are given in the Appendix.

## 2 Improved AIC for survival analysis data

Let $T$ , $C$, and x be the survival time, censoring time, and the associated $p \times 1$ covariates respectively. Let $Z = \min(T, C)$ be the observed time and $\delta = I(T \le C)$ be the censoring indicator. Let $h(t|\mathrm{x})$ and $S(t|\mathrm{x})$ be the conditional hazard and survival functions of $T$ given x, respectively. The complete likelihood of the data is given by

$$L = \prod_{u} h\left(Z_i|\mathbf{x}_i\right) \prod_{i=1}^{n} S\left(Z_i|\mathbf{x}_i\right), \tag{2.1}$$

where $n$ is the total number of observations, and the subscript $u$ denotes the product over the uncensored data. In this paper, we focus on the accelerated life time (ALT) model, one of the most useful parametric life models, of form

$$\log(T) = \alpha + \mathbf{x}^{\mathrm{T}}\beta + \sigma\varepsilon. \tag{2.2}$$

Let $S_0(t)$ denote the survival function of $T$ when x = 0, and $h_0(t)$ be the hazard risk of $S_0(t)$. It follows that

$$\begin{aligned} S(t|\mathbf{x}) &= S_0\left\{t \quad \exp\left(-\mathbf{x}^{\mathrm{T}}\beta\right)\right\}, \\ h(t|\mathbf{x}) &= h_0\left\{t \quad \exp\left(-\mathbf{x}^{\mathrm{T}}\beta\right)\right\} \exp\left(-\mathbf{x}^{\mathrm{T}}\beta\right). \end{aligned}$$

In a consequence, we obtain the log-likelihood of the observed data $\{(\mathrm{x}_i, Z_i, \delta_i), i = 1, \ldots, n\}$

$$l(\mathbf{x}, Z, \beta) = \sum_{u} \left(-\mathbf{x}_i^{\mathrm{T}}\beta + \log\left[h_0\left\{Z_i \exp\left(-\mathbf{x}_i^{\mathrm{T}}\beta\right)\right\}\right]\right) + \sum_{i=1}^{n} \log\left[S_0\left\{Z_i \exp\left(-\mathbf{x}_i^{\mathrm{T}}\beta\right)\right\}\right] \tag{2.3}$$

Collett (1994) suggested that the AIC for survival models should be

$$\mathrm{AIC} = -2\log(\text{likelihood}) + 2(p + 2 + k),$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $k = 0$ for the exponential model, $k = 1$ for the Weibull, log-logistic and log-normal models, and $k = 2$ for the generalized gamma model. Following Hurvich and Tsai (1989), we propose an improved AIC as follows

$$\mathrm{AIC}_{\mathrm{SUR}} = \mathrm{AIC} + \frac{2(p+2)(p+3)}{n-p-3}. \tag{2.4}$$

Different choices for the error distribution of $\varepsilon$ yield different regression models, and then different log-likelihood functions given in (2.3). The routines to finish the calculations on the log-likelihood functions and AICs (and then $\mathrm{AIC}_{\mathrm{SUR}}$s) are available in the most statistical packages like R/Splus and SAS.

A commonly used criterion of measuring the difference between the candidate model and the true model is the Kullback-Leibler information $\Delta = E_0(-2 \log L)$, where $E_0$ denotes the expectation with respect to the true model, and $L$ is the likelihood function under the candidate model. In the remainder of this section, we use this measure to derive a more precise model selection criterion for the special case of the exponential distribution to demonstrate the rationality of the proposed $\mathrm{AIC}_{\mathrm{SUR}}$ given in (2.4).

Consider the ALT model (2.2) with $\sigma = 1$ and $\varepsilon$ following an extreme value distribution whose density function is $\exp(v - e^v)$. Then the survival time $T$ has the exponential distribution with the density function $\lambda e^{-\lambda t}$, where $\lambda = \exp\{-(\alpha + \mathrm{x}^{\mathrm{T}}\beta)\}$. If we denote $\lambda_i = \exp\left\{-\left(\alpha + \mathbf{x}_i^{\mathrm{T}}\beta\right)\right\}$, then $h(Z_i \mid \mathrm{x}_i) = \lambda_i$, and $S(Z_i \mid \mathrm{x}_i) = \exp(-\lambda_i Z_i)$. So the log-likelihood function from (2.1) is given by

$$\log L = \sum_u \log \lambda_i + \sum_{i=1}^{n} (-\lambda_i Z_i).$$

From this, we see that the Kullback-Leibler information is

$$\begin{aligned}\Delta(\alpha,\beta) &= -2\sum_u \log \lambda_i + 2\sum_{i=1}^{n} \lambda_i E_0(Z_i) \\ &= -2\sum_u \log \lambda_i + 2\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_{i0}}\left(1 - e^{-\lambda_{i0}C}\right),\end{aligned}$$

where the censoring time is assumed to be a constant, $\lambda_{i0} = \exp\left\{-\left(\alpha_0 + \mathbf{x}_{0i}^{\mathrm{T}}\beta_0\right)\right\}$, and $\alpha_0$ and $\beta_0$ are the parameters in the true model.

Following Akaike (1974) and Hurvich and Tsai (1989) (see also Burnham and Anderson, 1998), a reasonable measure representing the discrepancy between the candidate and true models would be $E_0\Delta\left(\widehat{\alpha},\widehat{\beta}\right)$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are the estimators of $\alpha$ and $\beta$ under the candidate model. That is, we would choose those candidate models which minimize $E_0\Delta\left(\widehat{\alpha},\widehat{\beta}\right)$. In the Appendix we derive an (approximately) unbiased estimator, $\mathrm{AIC}_{\mathrm{exp}}$ given in (A.4), of $E_0\Delta\left(\widehat{\alpha},\widehat{\beta}\right)$ and this can be used to obtain a feasible model selection criterion.

We now numerically demonstrate the rationality of the proposed $\mathrm{AIC}_{\mathrm{SUR}}$ in (2.4) by comparing it with $\mathrm{AIC}_{\mathrm{exp}}$ in (A.4) under the exponential distribution which is regarded as more precise.

Generate data from the model $y = \mathrm{x}^{\mathrm{T}}\beta + \varepsilon$, where $\beta = (1, 2, 3, 4)^{\mathrm{T}}$, x follows a 4-dimensional normal distribution with the mean zero and covariance matrix $\mathbf{I}_{4\times4}$, and $\varepsilon$ follows an extreme value distribution with the density function $\exp(v - e^v)$. We consider the combinations of $n$ = 20, 30, 40, 50 and the censoring variable $C$ = 5, 10, 15, 20, 25, 30, and repeat 500 simulations for each combination. Table 1 presents the means and standard errors of $\mathrm{AIC}_{\mathrm{SUR}}$ and

$\text{AIC}_{\text{exp}}$. It is seen from the table that the values of $\text{AIC}_{\text{SUR}}$ and $\text{AIC}_{\text{exp}}$ are very close, suggesting that the difference between the two model selection criteria would often be quite minor. This implies the rationality of $\text{AIC}_{\text{SUR}}$ from one aspect. Of course, the above demonstration is based on a special distribution-exponential distribution. So in the following section, we conduct some simulations to study the behavior in selecting true models of $\text{AIC}_{\text{SUR}}$ under various distributions.

## 3 Simulation study

In this section, we investigate the finite sample performance of the proposed procedure $\text{AIC}_{\text{SUR}}$ by Monte Carlo simulations, and illustrate the proposed methodology by analyzing two real data sets in next section.

### Example 1

Generate data from the model

$$y = \mathbf{x}^{\mathrm{T}}\beta + \sigma\varepsilon,$$

where $\beta = (1, 2, 3, 4, 0, 0, 0, 0)^{\mathrm{T}}$, x follows an 8-dimensional normal distribution with the mean zero and covariance matrix $\mathbf{I}_{8\times 8}$. We consider three scenarios: (i) $\varepsilon$ follows a logistic distribution, (ii) $\varepsilon$ follows a log-normal distribution, and (iii) $\varepsilon$ follows an extreme distribution. We take the location and scale parameters 0 and 1, respectively. The value of the censoring random variable $C$ is generated by the uniform distribution $U(0, 10)$ for each observation. For each scenario, we take $n$ = 12, 20, 30, and $\sigma^2$ = 0.1, 0.5, 1. At each of 27 configurations, 500 independent data sets are generated. Similar to Hurvich and Tsai (1989), our candidate models are those whose predictors are sequential columns of X; i.e., consist of columns 1, $\cdots$, $r$ of X. The true model consists of the first 4 columns of X. We use three criteria: AIC, BIC, and $\text{AIC}_{\text{SUR}}$ to select a value of $r$ for each configuration, respectively. Tables 2-4 summarize the frequencies of the order selected by the specified criterion for scenarios 1−3, respectively. It is observed that $\text{AIC}_{\text{SUR}}$ consistently provides the best selection of $r = 4$ among the three criteria studied, regardless of sample sizes and variances. Even when $n = 12$, $\text{AIC}_{\text{SUR}}$ generally selects at least 250 times of the correct model, while AIC selects only around 200 times of the correct model. When $n = 20$, the number of the correct times selected by $\text{AIC}_{\text{SUR}}$ is double to that of AIC. Usually, the best model can be identified more frequently when $n = 30$. However, $\text{AIC}_{\text{SUR}}$ still substantially outperforms AIC. It is also seen from Tables 2 -4 that BIC is usually better than AIC but substantially inferior to $\text{AIC}_{\text{SUR}}$ for the cases considered here.

## 4 Real Data Analysis

### Example 2

We fit the motor data set, which was obtained by Nelson and Hahn (1972) and studied by Kalbfleisch and Prentice (1980), using the exponential, Weibull, log-logistic, and log-normal models. The response variable and covariate are the hour to the failure of motorette and operating temperature, respectively. Nelson and Hahn (1972) used the log-normal model, while Kalbfleisch and Prentice (1980) used the Weibull model for the analysis of this dataset because the latter authors thought that the Weibull model generates a larger likelihood value. On the basis of our simulations (data not shown), this evidence may be not enough to be convinced. We therefore apply the proposed method to the analysis of this dataset. To compare with the results of Nelson and Hahn (1972) and of Kalbfleisch and Prentice (1980), we make the transformation $x = 1000/(273.2 + \text{temperature})$ and exclude the ten observations at the temperature level of 150% because the experiment was an accelerated process to speed up the failure time. A total of 30 observations are used in our analysis. We fit the four models and present the AIC and $\text{AIC}_{\text{SUR}}$ values in Table 5. The results of the two criteria uniformly indicate



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that the Weibull model is most appropriate. This confirmation convinces that Kalbfleisch and Prentice's choice is appropriate. The estimates and their related quantities for the four models are presented in Table 6. Shown in Table 6 are the estimates, their corresponding standard deviations, the z-ratios, and the p-values obtained by testing the null hypothesis that the corresponding parameter is zero.

## Example 3

In this example, we apply the proposed method to analyze the data set from a study of the bone marrow transplantation (BMT) for leukemia. This study was designed in 1984 as a single institution (Ohio State University Hospitals, OSU) study and was modified in 1987 to include the five institutions known to be using this preparative regimen in all patients with the acute myelocytic leukemia (AML). All patients who underwent the marrow transplantation for the AML using this preparative regimen at the participating institutions were reported. One hundred twenty-seven patients were with the AML aged 7 to 55 (median 30) who were treated from March 1, 1984 through June 30, 1989 at the five separate centers with the allogeneic BMT following preparation with Bu and Cy. Fifty-five of them underwent their transplantation at Ohio State University Hospitals (OSU; Columbus), 23 at Wilford Hall at Lackland Air Force Base (San Antonio, TX), 22 at Hahnemann University (Philadelphia, PA), 17 at St Vincent's Hospital (Sydney, Australia), and 10 at Alfred Hospital (Melbourne, Australia). More details of the study are referred to Copelan et al. (1991).

Our response variable is the disease free survival time, $T$ , and the disease free survival indicator (1-dead or relapsed, 0-alive and disease free), $\delta$ . The potential covariates in this study include the following variables:

$X_1$: patient age in year;

$X_2$: donor age in year;

$X_3$: patient sex (1-male, 0-female);

$X_4$: donor sex (1-male, 0-female);

$X_5$: patient cytomegalovirus (CMV) immune status (1-CMV positive, 0-CMV negative);

$X_6$: donor CMV status (1-CMV positive, 0-CMV negative);

$X_7$: waiting time to transplant in day;

$X_8$: French-American-British (FAB, 1-FAB grade 4 or 5 and AML, 0-otherwise);

$X_9$: hospital (1-Ohio State University, 2-Alferd , 3-St. Vincent, 4-Hahnemann);

$X_{10}$: methotrexate (MTX) used as a graft-versus-host-prophylactic (1-yes, 0-no).

For an illustration, we consider only the observations of ALL the patients. The $X_8$ values are all zeros and therefore excluded in our analysis. A total of 79 combinations of the covariates is considered. We fit the exponential, Weibull, log-logistic, and log-normal models to the data for each combination, and select the corresponding best model by AIC and $\mathrm{AIC_{SUR}}$.

Using the Weibull and exponential models, AIC and $\mathrm{AIC_{SUR}}$ select the model with the covariates ($X_1$, $X_2$, $X_6$, $X_9$) as the best one, and AIC = 358.19 (Weibull) and 358.21 (exponential) and $\mathrm{AIC_{SUR}}$ = 360.07 (Weibull) and 360.08 (exponential). It is clear that the values of both AIC and $\mathrm{AIC_{SUR}}$ under the Weibull and exponential models are very close. The corresponding estimates under these two setups are given in Table 7. Although both AIC and $\mathrm{AIC_{SUR}}$ suggest the Weibull model, the p-value of log(scale) indicates that the exponential model is appropriate. For the log-logistic model, AIC and $\mathrm{AIC_{SUR}}$ select the model with the covariates ($X_1$, $X_2$, $X_6$, $X_{10}$) as the best one, and AIC = 361.70 and $\mathrm{AIC_{SUR}}$ = 363.55. The p-value of the log(scale) indicates that the scale is not significantly different from 1. For the log-normal model, AIC selects the model with the covariates ($X_1$, $X_2$, $X_6$, $X_{10}$) as the best one with AIC = 363.99, while $\mathrm{AIC_{SUR}}$ selects the model with the covariates ($X_1$, $X_6$, $X_{10}$) as the best one with $\mathrm{AIC_{SUR}}$ = 363.15. Seeing the p-value of testing the parameters in the model selected by AIC, one may notice that $X_2$ and $X_{10}$ are not statistically significant, and the model selected by $\mathrm{AIC_{SUR}}$ seems more reasonable. In summary, we recommend to use the exponential model to fit the data with the covariates ($X_1$, $X_2$, $X_6$, $X_9$) on the basis of the above analysis. $\mathrm{AIC_{SUR}}$ makes us confident to this selection.

## 5 Discussion

To select an appropriate model for survival analysis, we generalized Hurvich and Tsai's (1989) approach and developed an improved AIC selection procedure, $\mathrm{AIC_{SUR}}$. The proposed method was shown to be superior to the traditional AIC and BIC through simulation studies. It is interesting to observe from our simulations that when the sample size is not small ($n$ = 20 and 30), the efficiency of $\mathrm{AIC_{SUR}}$ can be greatly increased if we use the total number of uncensored observations instead of the total number of observations $n$ in the extra penalty term of $\mathrm{AIC_{SUR}}$ (data not shown). Our method was also applied to analyze two real data sets.

The proposed $\mathrm{AIC_{SUR}}$ is a general criterion of selecting survival models. It can be applied to the exponential, Weibull, log-logistic, log-normal and generalized gamma models etc. As a theoretical verification for $\mathrm{AIC_{SUR}}$, we derived a more precise model selection criterion $\mathrm{AIC_{exp}}$ for the particular scenario of the exponential distribution for the survival time with constant censoring. The calculation results showed that the discrepancy between the two model selection criteria is quite minor under the exponential distribution. Of course, the further justification is necessary under the more general cases of the distributions for the survival time and this warrants our future research.

Unlike other advanced selection procedures, the proposed method is very easy to implement and computationally efficient. These features make the method promising in practice. The efficient R/Splus computation codes were developed and are available from the authors upon request. Extension of the idea to the goodness-of-fit and semiparametric survival models would be possible and will also be studied in our future work.

### Acknowledgments

The authors thank the two referees for their helpful suggestions and constructive comments. This research was partially supported by the two grants AI62247 and AI59773 from the National Institute of Allergy and Infectious Diseases. Zou's research was also partially supported by the two grants 70625004 and 10471043 from the National Natural Science Foundation of China.

## APPENDIX

An estimator of $E_0\Delta\left(\widehat{\alpha},\widehat{\beta}\right)$ under the exponential distribution

First note that

$$\begin{aligned}
E_0\left\{\frac{\widehat{\lambda}_i}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\right\} &= E_{Z_{-i}}E_{Z_i}\left\{\frac{\widehat{\lambda}_i(Z_i,Z_{-i})}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\right\}\\
&= E_{Z_{-i}}\left[\int_0^\infty\left\{\frac{\widehat{\lambda}_i(\min(t_i,C),Z_{-i})}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\cdot\lambda_{i0}e^{-\lambda_{i0}t_i}\right\}dt_i\right]\\
&= E_{Z_{-i}}\left[\int_0^C\left\{\frac{\widehat{\lambda}_i(t_i,Z_{-i})}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\cdot\lambda_{i0}e^{-\lambda_{i0}t_i}\right\}dt_i\right.\\
&\qquad\left.+\int_C^\infty\left\{\frac{\widehat{\lambda}_i(C_i,Z_{-i})}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\cdot\lambda_{i0}e^{-\lambda_{i0}t_i}\right\}dt_i\right]\\
&\equiv E_{Z_{-i}}\left(U+V\right),
\end{aligned}$$



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

where $\widehat{\lambda_i}(\mathrm{a},\mathrm{b})$ denotes the estimated value of $\lambda_i$ based on the data (a, b) and is assumed to be continuous, and $Z_{-i}$ means the vector consisting of $Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n$.

It is readily seen that

$$\begin{aligned} V &= \frac{1-e^{-\lambda_{i0}C}}{\lambda_{i0}} \cdot \widehat{\lambda_i}(C, Z_{-i})\, e^{-\lambda_{i0}C} \\ &= e^{-\lambda_{i0}C} \cdot E_{Z_i}\left[Z_i \widehat{\lambda_i}(C, Z_{-i})\right]. \end{aligned} \tag{A.1}$$

On the other hand, it can be shown that

$$\begin{aligned} U &= \left(1-e^{-\lambda_{i0}C}\right) \cdot \int_0^C \widehat{\lambda_i}(t_i, Z_{-i})\, e^{-\lambda_{i0}t_i} dt_i \\ &= \left(1-e^{-\lambda_{i0}C}\right) \cdot \int_0^C e^{-\lambda_{i0}t_i} d\left\{\int_0^{t_i} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} \\ &= \left(1-e^{-\lambda_{i0}C}\right)\Bigg[e^{-\lambda_{i0}C} \cdot \int_0^C \widehat{\lambda_i}(w, Z_{-i})\, dw \\ &\qquad + \int_0^C \left\{\int_0^{t_i} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} \cdot \lambda_{i0} e^{-\lambda_{i0}t_i} dt_i\Bigg] \\ &= \left(1-e^{-\lambda_{i0}C}\right)\Bigg[\int_0^\infty \left\{\int_0^C \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} \cdot \lambda_{i0} e^{-\lambda_{i0}t_i} dt_i \\ &\qquad + \int_0^C \left\{\int_0^{t_i} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} \cdot \lambda_{i0} e^{-\lambda_{i0}t_i} dt_i\Bigg] \\ &= \left(1-e^{-\lambda_{i0}C}\right) \cdot \int_0^\infty \left\{\int_0^{\min(t_i, C)} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} \lambda_{i0} e^{-\lambda_{i0}t_i} dt_i \\ &= \left(1-e^{-\lambda_{i0}C}\right) \cdot E_{Z_i}\left\{\int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\}. \end{aligned} \tag{A.2}$$

From formulas (A.1) and (A.2), we obtain

$$\begin{aligned} & E_0\left\{\frac{\widehat{\lambda_i}}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\right\} \\ &= E_{Z_{-i}}\left[\left(1-e^{-\lambda_{i0}C}\right) \cdot E_{Z_i}\left\{\int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\right\} + e^{-\lambda_{i0}C} \cdot E_{Z_i}\left\{Z_i \widehat{\lambda_i}(C, Z_{-i})\right\}\right] \\ &= \left(1-e^{-\lambda_{i0}C}\right) \cdot E_0\left\{\int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\, dw\right\} + e^{-\lambda_{i0}C} \cdot E_0\left\{Z_i \widehat{\lambda_i}(C, Z_{-i})\right\}. \end{aligned} \tag{A.3}$$

Therefore,

$$\begin{aligned} E_0 \Delta\left(\widehat{\alpha}, \widehat{\beta}\right) &= E_0\left\{-2\sum_u \log \widehat{\lambda_i}\right\} = 2\sum_{i=1}^n E_0\left\{\frac{\widehat{\lambda_i}}{\lambda_{i0}}\left(1-e^{-\lambda_{i0}C}\right)\right\} \\ &= E_0\left[-2\sum_u \log \widehat{\lambda_i} + 2\sum_{i=1}^n \left\{\left(1-e^{-\lambda_{i0}C}\right) \cdot \int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\, dw + e^{-\lambda_{i0}C} \cdot Z_i \widehat{\lambda_i}(C, Z_{-i})\right\}\right] \\ &+ E_0\Bigg(\left\{-2\sum_u \log \widehat{\lambda_i} + 2\sum_{i=1}^n Z_i \widehat{\lambda_i}\right\} + 2\sum_{i=1}^n \Bigg[\left(1-e^{-\lambda_{i0}C}\right) \cdot \left\{\int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\, dw - Z_i \widehat{\lambda_i}\right\} \\ &\qquad + e^{-\lambda_{i0}C} \cdot \left\{Z_i \widehat{\lambda_i}(C, Z_{-i}) - Z_i \widehat{\lambda_i}\right\}\Bigg]\Bigg). \end{aligned}$$

Thus, we propose to select the best ALT model which minimizes the following Kullback-Leibler information:

$$\begin{aligned} \mathrm{AIC}_{\exp} = \; & -2\log(\text{likelihood}) + 2\sum_{i=1}^n \Bigg[\left(1-\psi\left(\widehat{\lambda_{i0}}\right)\right) \cdot \left\{\int_0^{Z_i} \widehat{\lambda_i}(w, Z_{-i})\, dw - Z_i \widehat{\lambda_i}\right\} \\ & + \psi\left(\widehat{\lambda_{i0}}\right) \cdot \left\{Z_i \widehat{\lambda_i}(C, Z_{-i}) - Z_i \widehat{\lambda_i}\right\}\Bigg], \end{aligned} \tag{A.4}$$

where $\psi\left(\widehat{\lambda_{i0}}\right)$ is an estimator of $\exp(-\lambda_{i0}C)$, for which we provide some estimation methods below. Noting that $Z_i$, $Z_{-i}$ and $\widehat{\lambda_i}$ are all known, the calculation on the integral in $\mathrm{AIC}_{\exp}$ is easy.

Observing that

$$E_0(Z_i) = \{1 - \exp(-\lambda_{i0}C)\}/\lambda_{i0}, \tag{A.5}$$



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

and

$$E_0\left(Z_i^2\right) = -\frac{2}{\lambda_{i0}}\left\{Ce^{-\lambda_{i0}C} - \frac{1}{\lambda_{i0}}\left(1 - e^{-\lambda_{i0}C}\right)\right\},$$

we have

$$E_0\left(\frac{2Z_i - \lambda_{i0}Z_i^2}{2C}\right) = e^{-\lambda_{i0}C}. \tag{A.6}$$

Therefore, combining (A.5) and (A.6), we can obtain an estimator of $\lambda_{i0}$ as

$$\widehat{\lambda}_{i0} = \frac{2(C - Z_i)}{Z_i(2C - Z_i)}. \tag{A.7}$$

Thus, a natural estimator of $\exp(-\lambda_{i0}C)$ is $\exp\left(-\widehat{\lambda}_{i0}C\right)$ with $\widehat{\lambda}_{i0}$ given in (A.7). On the other hand, by virtue of (A.6) and (A.7), we can get another estimator of $\exp(-\lambda_{i0}C)$ as

$$\frac{2Z_i - \widehat{\lambda}_{i0}Z_i^2}{2C} = \frac{Z_i}{2C - Z_i}. \tag{A.8}$$

## References

Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19 1974:716–723.

Burnham, KP.; Anderson, DP. Model Selection and Inference: A Practical Information-Theoretical Approach. Springer-Verlag; New York: 1998.

Collett, D. Modeling Survival Data in Medical Research. Chapman and Hall; New York: 1994.

Copelan EA, Biggs JC, Thompson JM, et al. Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with BuCy2. Blood 1991;78:838–843. [PubMed: 1859895]

Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. The Annals of Statistics 2002;30:74–99.

Faraggi D, Simon R. Bayesian variable selection method for censored survival data. Biometrics 1998;54:1475–85. [PubMed: 9883546]

Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika 1989;76:297–307.

Hurvich CM, Tsai CL. Model selection for extended quasi-likelihood models in small samples. Biometrics 1995;51:1077–84. [PubMed: 7548692]

Hurvich CM, Simonoff JS, Tsai CL. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society 1998;60:271–93.Series B

Hurvich CM, Tsai CL. Semiparametric and additive model selection using an improved Akaike information criterion. Journal of Computational and Graphical Statistics 1999;8:22–40.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. Wiley; New York: 1980.

Lindley DV. The choice of variables in multiple regression (with discussion). Journal of the Royal Statistical Society 1968;30:31–66.Series B

Mallows CL. Some comments on $C_p$. Technometrics 1973;15:661–75.

Nelson WB, Hahn GB. Linear estimation of regression relationships from censored data, part 1-simple methods and their applications (with discussion). Technometrics 1972;14:945–65.

Schwarz G. Estimating the dimension of a model. The Annals of Statistics 1978;6:461–4.

Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. Comm. Statist 1978;A7:13–26.

Tibshirani R. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society 1996;58:267–88.Series B



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in Medicine 1997;16:385–95. [PubMed: 9044528]

Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. Biometrics 2000;56:256–62. [PubMed: 10783804]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

**Table 1**

The comparison of $AIC_{SUR}$ and $AIC_{exp}$ under the exponential distribution based on 100 replications

| n | C | $AIC_{SUR}$ | $AIC_{SUR}$(se) | $AIC_{exp}$ | $AIC_{exp}$(se) |
|---|---|---|---|---|---|
| 20 | 5 | −49.05 | 5.52 | −48.85 | 5.53 |
| | 10 | −45.69 | 5.6 | −44.53 | 5.57 |
| | 15 | −40.72 | 5.93 | −40.69 | 5.88 |
| | 20 | −42 | 5.76 | −41.65 | 5.79 |
| | 25 | −20.09 | 5.43 | −20.7 | 5.45 |
| | 30 | −36.67 | 5.47 | −36.85 | 5.53 |
| 30 | 5 | −74.96 | 5.94 | −76.26 | 5.94 |
| | 10 | −70.99 | 6.38 | −73.16 | 6.37 |
| | 15 | −52.18 | 6.29 | −53.6 | 6.25 |
| | 20 | −46.12 | 6.67 | −47.07 | 6.68 |
| | 25 | −50.02 | 6.28 | −52.12 | 6.31 |
| | 30 | −50.76 | 6.26 | −52.56 | 6.3 |
| 40 | 5 | −102.21 | 6.39 | −104.51 | 6.39 |
| | 10 | −83.2 | 6.52 | −84.97 | 6.52 |
| | 15 | −84.94 | 6.64 | −87.27 | 6.63 |
| | 20 | −79.45 | 7.33 | −81.7 | 7.34 |
| | 25 | −68.97 | 6.9 | −71.82 | 6.89 |
| | 30 | −77 | 6.6 | −79.54 | 6.59 |
| 50 | 5 | −116.25 | 6.69 | −119.95 | 6.69 |
| | 10 | −123.37 | 7 | −125.84 | 7.02 |
| | 15 | −104.48 | 7.3 | −107.31 | 7.29 |
| | 20 | −95.02 | 6.58 | −97.96 | 6.59 |
| | 25 | −77.79 | 7.03 | −80.54 | 7.02 |
| | 30 | −76.77 | 6.72 | −79.96 | 6.74 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

## Table 2

Scenario 1-Frequency of order selected using different criteria in 500 replications of model fitting with the true order $r_0 = 4$ under the logistic distribution*

| n | $\sigma_\varepsilon^2$ | Criterion | Selected model order r: 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 12 | 0.1 | AIC | 112 | 198 | 91 | 57 | 34 | 8 |
| | | BIC | 104 | 201 | 94 | 59 | 35 | 7 |
| | | $AIC_{SUR}$ | 206 | 240 | 40 | 13 | 1 | 0 |
| | 0.5 | AIC | 127 | 183 | 100 | 51 | 26 | 13 |
| | | BIC | 124 | 180 | 105 | 51 | 26 | 14 |
| | | $AIC_{SUR}$ | 194 | 251 | 42 | 11 | 0 | 2 |
| | 1 | AIC | 118 | 162 | 108 | 69 | 34 | 9 |
| | | BIC | 110 | 161 | 112 | 71 | 35 | 11 |
| | | $AIC_{SUR}$ | 203 | 233 | 46 | 14 | 3 | 1 |
| 20 | 0.1 | AIC | 8 | 108 | 63 | 69 | 107 | 145 |
| | | BIC | 9 | 141 | 66 | 72 | 97 | 115 |
| | | $AIC_{SUR}$ | 17 | 289 | 69 | 52 | 43 | 30 |
| | 0.5 | AIC | 8 | 133 | 62 | 66 | 86 | 145 |
| | | BIC | 8 | 167 | 65 | 69 | 74 | 117 |
| | | $AIC_{SUR}$ | 15 | 312 | 59 | 47 | 35 | 32 |
| | 1 | AIC | 4 | 130 | 57 | 79 | 98 | 132 |
| | | BIC | 4 | 162 | 60 | 76 | 94 | 104 |
| | | $AIC_{SUR}$ | 10 | 300 | 54 | 49 | 54 | 33 |
| 30 | 0.1 | AIC | 2 | 204 | 52 | 62 | 79 | 101 |
| | | BIC | 2 | 270 | 51 | 54 | 58 | 65 |
| | | $AIC_{SUR}$ | 3 | 316 | 57 | 48 | 36 | 40 |
| | 0.5 | AIC | 2 | 211 | 50 | 58 | 66 | 113 |
| | | BIC | 3 | 288 | 51 | 48 | 45 | 65 |
| | | $AIC_{SUR}$ | 3 | 330 | 57 | 39 | 33 | 38 |
| | 1 | AIC | 0 | 232 | 53 | 30 | 69 | 116 |
| | | BIC | 0 | 309 | 48 | 21 | 47 | 75 |
| | | $AIC_{SUR}$ | 1 | 350 | 59 | 25 | 27 | 38 |

*The censoring variable $C$ is generated by the uniform distribution $U(0, 10)$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## Table 3

Scenario 2-Frequency of order selected using different criteria in 500 replications of model fitting with the true order $r_0 = 4$ under the log-normal distribution*

| $n$ | $\sigma_\varepsilon^2$ | Criterion | Selected model order $r$: 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 12 | 0.1 | AIC | 96 | 223 | 96 | 59 | 21 | 5 |
| | | BIC | 89 | 224 | 98 | 60 | 24 | 5 |
| | | $AIC_{SUR}$ | 129 | 302 | 49 | 18 | 2 | 0 |
| | 0.5 | AIC | 111 | 219 | 97 | 42 | 21 | 10 |
| | | BIC | 105 | 222 | 99 | 40 | 23 | 11 |
| | | $AIC_{SUR}$ | 137 | 302 | 51 | 10 | 0 | 0 |
| | 1 | AIC | 103 | 202 | 91 | 68 | 27 | 9 |
| | | BIC | 98 | 207 | 91 | 67 | 26 | 11 |
| | | $AIC_{SUR}$ | 161 | 279 | 44 | 15 | 1 | 0 |
| 20 | 0.1 | AIC | 5 | 154 | 63 | 76 | 92 | 110 |
| | | BIC | 5 | 176 | 69 | 71 | 86 | 93 |
| | | $AIC_{SUR}$ | 7 | 296 | 70 | 55 | 41 | 31 |
| | 0.5 | AIC | 2 | 143 | 63 | 89 | 91 | 112 |
| | | BIC | 2 | 169 | 68 | 80 | 82 | 99 |
| | | $AIC_{SUR}$ | 4 | 303 | 58 | 63 | 46 | 26 |
| | 1 | AIC | 7 | 156 | 53 | 71 | 85 | 128 |
| | | BIC | 7 | 184 | 51 | 71 | 78 | 109 |
| | | $AIC_{SUR}$ | 12 | 343 | 44 | 46 | 34 | 21 |
| 30 | 0.1 | AIC | 1 | 204 | 59 | 57 | 65 | 114 |
| | | BIC | 1 | 282 | 51 | 45 | 50 | 71 |
| | | $AIC_{SUR}$ | 1 | 335 | 52 | 45 | 32 | 35 |
| | 0.5 | AIC | 1 | 213 | 62 | 43 | 72 | 109 |
| | | BIC | 1 | 291 | 51 | 39 | 51 | 67 |
| | | $AIC_{SUR}$ | 1 | 346 | 47 | 37 | 35 | 34 |
| | 1 | AIC | 0 | 199 | 79 | 56 | 59 | 107 |
| | | BIC | 0 | 281 | 76 | 43 | 41 | 59 |
| | | $AIC_{SUR}$ | 0 | 325 | 78 | 36 | 23 | 38 |

*The censoring variable $C$ is generated by the uniform distribution $U$ (0, 10).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 4**

Scenario 3-Frequency of order selected using different criteria in 500 replications of model fitting with the true order $r_0 = 4$ under the extreme value distribution*

| | | | Selected model order *r* | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***n*** | $\sigma_\varepsilon^2$ | **Criterion** | **3** | **4** | **5** | **6** | **7** | **8** |
| 12 | 0.1 | AIC | 103 | 222 | 94 | 59 | 13 | 9 |
| | | BIC | 96 | 223 | 97 | 61 | 13 | 10 |
| | | $AIC_{SUR}$ | 137 | 299 | 44 | 17 | 2 | 1 |
| | 0.5 | AIC | 124 | 191 | 82 | 66 | 28 | 9 |
| | | BIC | 119 | 191 | 83 | 69 | 28 | 10 |
| | | $AIC_{SUR}$ | 181 | 263 | 42 | 11 | 3 | 0 |
| | 1 | AIC | 132 | 164 | 95 | 59 | 40 | 10 |
| | | BIC | 121 | 169 | 102 | 57 | 40 | 11 |
| | | $AIC_{SUR}$ | 244 | 208 | 38 | 9 | 1 | 0 |
| 20 | 0.1 | AIC | 4 | 113 | 57 | 87 | 104 | 135 |
| | | BIC | 4 | 152 | 63 | 84 | 88 | 109 |
| | | $AIC_{SUR}$ | 4 | 268 | 74 | 71 | 47 | 36 |
| | 0.5 | AIC | 6 | 136 | 69 | 67 | 101 | 121 |
| | | BIC | 5 | 166 | 72 | 63 | 91 | 103 |
| | | $AIC_{SUR}$ | 9 | 302 | 76 | 42 | 41 | 30 |
| | 1 | AIC | 17 | 122 | 57 | 74 | 95 | 135 |
| | | BIC | 21 | 144 | 61 | 72 | 87 | 115 |
| | | $AIC_{SUR}$ | 44 | 263 | 64 | 46 | 51 | 32 |
| 30 | 0.1 | AIC | 3 | 197 | 54 | 64 | 61 | 121 |
| | | BIC | 3 | 267 | 56 | 51 | 49 | 74 |
| | | $AIC_{SUR}$ | 3 | 324 | 53 | 40 | 34 | 46 |
| | 0.5 | AIC | 0 | 218 | 50 | 68 | 59 | 105 |
| | | BIC | 0 | 284 | 53 | 48 | 52 | 63 |
| | | $AIC_{SUR}$ | 1 | 327 | 57 | 41 | 34 | 40 |
| | 1 | AIC | 6 | 211 | 61 | 61 | 53 | 108 |
| | | BIC | 8 | 284 | 62 | 44 | 34 | 68 |
| | | $AIC_{SUR}$ | 11 | 339 | 55 | 34 | 28 | 33 |

*The censoring variable $C$ is generated by the uniform distribution $U(0, 10)$.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

**Table 5**

Values of AIC and $AIC_{SUR}$ for the Motor data set under various models

| | **AIC** | **$AIC_{SUR}$** |
|---|---|---|
| Weibull | 294.69 | 296.29 |
| Exponential | 309.61 | 311.21 |
| Log-logistic | 295.68 | 297.28 |
| Log-normal | 297.73 | 299.33 |

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



**Table 6**

Estimated results of parameters for the Motor data set under various models

| model | | Value | Std. Error | z | p-value |
|---|---|---|---|---|---|
| Weibull | (Intercept) | −11.89 | 1.97 | −6.05 | 0 |
| | temp | 9.04 | 0.91 | 9.98 | 0 |
| | Log(scale) | −1.02 | 0.22 | −4.63 | 0 |
| Exponential | (Intercept) | −8.99 | 5.5 | −1.63 | 0.1 |
| | temp | 7.83 | 2.54 | 3.08 | 0 |
| Log-logistic | (Intercept) | −11.11 | 2.21 | −5.02 | 0 |
| | temp | 8.61 | 1.03 | 8.38 | 0 |
| | Log(scale) | −1.19 | 0.22 | −5.46 | 0 |
| Log-normal | (Intercept) | −10.47 | 2.77 | −3.78 | 0 |
| | temp | 8.32 | 1.28 | 6.48 | 0 |
| | Log(scale) | −0.5 | 0.18 | −2.75 | 0.01 |

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

**Table 7**

Results of variable selection using AIC and $AIC_{SUR}$ and the corresponding estimated results for the BMT data set under various models

| Model | Criterion | | Value | Std. Error | z | p-value |
|---|---|---|---|---|---|---|
| Weibull | AIC/$AIC_{SUR}$ | Intercept | 9.679 | 0.986 | 9.813 | 0 |
| | | $X_1$ | −0.227 | 0.051 | −4.448 | 0 |
| | | $X_2$ | 0.114 | 0.035 | 3.274 | 0.001 |
| | | $X_6$ | 1.663 | 0.512 | 3.249 | 0.001 |
| | | $X_9$ | −0.686 | 0.254 | −2.703 | 0.007 |
| | | Log(scale) | 0.022 | 0.179 | 0.126 | 0.9 |
| Exponential | AIC/$AIC_{SUR}$ | Intercept | 9.657 | 0.952 | 10.146 | 0 |
| | | $X_1$ | −0.226 | 0.049 | −4.653 | 0 |
| | | $X_2$ | 0.114 | 0.034 | 3.384 | 0.001 |
| | | $X_6$ | 1.651 | 0.492 | 3.359 | 0.001 |
| | | $X_9$ | −0.683 | 0.247 | −2.761 | 0.006 |
| Log-logistic | AIC/$AIC_{SUR}$ | Intercept | 8.312 | 0.94 | 8.844 | 0 |
| | | $X_1$ | −0.176 | 0.051 | −3.436 | 0.001 |
| | | $X_2$ | 0.082 | 0.037 | 2.224 | 0.026 |
| | | $X_6$ | 1.352 | 0.594 | 2.277 | 0.023 |
| | | $X_{10}$ | −1.093 | 0.489 | −2.235 | 0.025 |
| | | Log(scale) | −0.182 | 0.18 | −1.009 | 0.313 |
| Log-normal | AIC | Intercept | 8.611 | 1.001 | 8.604 | 0 |
| | | $X_1$ | −0.167 | 0.054 | −3.077 | 0.002 |
| | | $X_2$ | 0.06 | 0.042 | 1.433 | 0.152 |
| | | $X_6$ | 1.297 | 0.6 | 2.163 | 0.031 |
| | | $X_{10}$ | −1.059 | 0.558 | −1.897 | 0.058 |
| | | Log(scale) | 0.452 | 0.155 | 2.911 | 0.004 |
| | $AIC_{SUR}$ | Intercept | 9.025 | 0.982 | 9.194 | 0 |
| | | $X_1$ | −0.116 | 0.04 | −2.897 | 0.004 |
| | | $X_6$ | 1.183 | 0.595 | 1.989 | 0.047 |
| | | $X_{10}$ | −1.111 | 0.561 | −1.982 | 0.047 |
| | | Log(scale) | 0.465 | 0.156 | 2.99 | 0.003 |