

Testing Inference in Accelerated Failure Time Models

Francisco M. C. Medeiros¹, Antônio H. M. da Silva-Júnior¹, Dione M. Valença¹ & Silvia L. P. Ferrari²

¹ Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal/RN, Brazil

² Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

Correspondence: Dione M. Valença, Departamento de Estatística, UFRN-CCET, Campus Universitário, 59078-970, Natal/RN, Brazil. E-mail: dione@ccet.ufrn.br

Received: March 13, 2014 Accepted: April 14, 2014 Online Published: April 21, 2014

doi:10.5539/ijsp.v3n2p121 URL: <http://dx.doi.org/10.5539/ijsp.v3n2p121>

Abstract

We address the issue of performing hypothesis testing in accelerated failure time models for non-censored and censored samples. The performances of the likelihood ratio test and a recently proposed test, the gradient test, are compared through simulation. The gradient test features the same asymptotic properties as the classical large sample tests, namely, the likelihood ratio, Wald and score tests. Additionally, it is as simple to compute as the likelihood ratio test. Unlike the score and Wald tests, the gradient test does require the computation of the information matrix, neither observed nor expected. Our study suggests that the gradient test is more reliable than the other classical tests when the sample is of small or moderate size.

Keywords: accelerated failure time models, gradient test, likelihood ratio test, random censoring

1. Introduction

An important class of regression models used to assess the relationship between the response variable and the covariates in survival analysis is parametric accelerated failure time (AFT) models. These models have a simple intuitive interpretation in real problems (Wei, 1992). Unlike proportional hazards models, AFT models describe the logarithm of the event times as a linear regression on the covariates, which act multiplicatively on the event times, *accelerating or decelerating* the time scale. Cox semi-parametric models (Cox, 1972) do not assume any particular distribution for the failure times and have been the most popular models in survival analysis. Nevertheless, some authors present reasonable arguments in favor of the parametric AFT models based on their asymptotic properties (Efron, 1977; Oakes, 1977), simulation results (Orbe, Ferreira, & Núñez-Antón, 2002) and applications to real data (Nardi & Schemper, 2003; Grover, Das, Swain, & Deka, 2013). Given the importance of such models, several authors have extended the AFT models, for example, adding a random effect to deal with correlated survival data (Lambert, Collett, Kimber, & Johnson, 2004; Pan, 2001), allowing measurement error in the covariates (Gimenez & Bolfarine, 1997; Valença & Bolfarine, 2006) or including a cure fraction to treat the presence of immune individuals in the sample (Yamaguchi, 1992; Peng, 1998; Ortega, 2009).

Typical distributions that have been used by various authors in connection with the parametric AFT models are the exponential, Weibull, log-normal, gamma and log-logistic distributions, in addition to more flexible distributions such as the extended generalized gamma and the generalized F distributions. Descriptions of the most commonly used distributions, as well as related inferential procedures, can be found in Lawless (2003), Kalbfleisch and Prentice (2002) and in Cox and Oakes (1984). Moreover, several extensions of the usual survival distributions, particularly the Weibull distribution, have been proposed to provide better fit in complex lifetime data; see, e.g., Marshall and Olkin (1997), Lai, Xie, and Murthy (2003), and Carrasco, Ortega, and Cordeiro (2008). A comprehensive discussion of general methods for constructing new distributions for lifetime data is presented in Lai (2013).

The most commonly used statistical test in AFT models is the likelihood ratio test, which performs well when the sample size is large. The other classical large sample tests, namely the Wald and score tests, can also be used. However, a disadvantage of these tests is that they require the computation of the information matrix. In AFT models, the expected information matrix is usually difficult or even impossible to obtain, particularly under censoring. An alternative is to replace the expected information matrix by its observed counterpart. We noticed,

however, that in small and moderate-sized samples the observed information matrix may produce unreliable standard errors. In fact, there is no guarantee that it is positive definite. The gradient test (Terrell, 2002) is a recently proposed testing procedure serving as an alternative to the classical tests. The gradient statistic does not require the information matrix, neither observed nor expected, and it is very simple to compute. Additionally, it has the same first order asymptotic properties as the other three test statistics.

Lemonte and Ferrari (2012) obtained the second-order local power of the gradient test and showed that none of the four competing tests is uniformly more powerful than the others. In addition, Lemonte and Ferrari (2011a) compared the size and power properties of the four rival tests in a Birnbaum-Saunders regression model for complete samples. Their simulation results suggest that the score and the gradient tests outperform the likelihood ratio and the Wald tests in small and moderate-sized samples, and are consistent with the study of Ferrari and Pinheiro (2014). The only study concerning the performance of the four tests in censored samples is that of Lemonte and Ferrari (2011b). The authors considered independent and identically distributed observations of the Birnbaum-Saunders distribution, and of two generalized versions of this distribution, under type-II censoring. Because the Wald and the score tests involve the information matrix, which could not be obtained under censoring, they implemented these tests with the observed information matrix. They noticed that the Wald and score tests were markedly oversized and that the inverse of the observed information matrix frequently produced negative standard errors for censored samples in their simulations. Their overall conclusion is in favor of the gradient test.

Censored samples are often encountered in survival and reliability studies. Our goal is to evaluate the performance of the gradient test in comparison with the likelihood ratio test in accelerated failure time models under random censoring. This paper is organized as follows. In Section 2 we describe the accelerated failure time models and present the likelihood ratio and gradient tests in these models. In Section 3 we present the simulation results for both tests in different scenarios. In Section 4 we illustrate and compare the tests in two real data applications. Finally, Section 5 closes the paper with some concluding remarks.

2. Accelerated Failure Time Models

Let T_i be the event time for individual i , and let $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$ be a fixed covariate vector that allows a possibly non-null intercept. The AFT model can be represented by

$$\log T_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where ϵ_i are independent and identically distributed random errors with a distribution with support in the whole real line and that does not depend on \mathbf{x}_i . The vector $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$ and σ are unknown parameters. Hence, (1) describes a linear regression model for $\log T_i$.

Survival times may be subject to right censoring. Here, the censoring times are represented by the independent random variables C_i , for $i = 1, \dots, n$, which are assumed to be independent of T_1, \dots, T_n . The censoring mechanism is assumed to be non-informative, that is, the distribution of the C_i 's does not depend on unknown parameters. Let $\delta_i = 1$, if the observation for individual i is a failure time, and $\delta_i = 0$, if it is a censoring time. The observations can be represented by the pairs of random variables (Y_i, δ_i) , where $Y_i = \min(\log T_i, \log C_i)$, and the covariate vectors \mathbf{x}_i^\top , for $i = 1, \dots, n$.

The likelihood function for the unknown parameters is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \left[\frac{1}{\sigma} f\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \right]^{\delta_i} \left[S\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma}\right) \right]^{1-\delta_i},$$

where y_i is the observed value of Y_i , $f(\cdot)$ and $S(\cdot)$ denote the density and survival functions of ϵ_i , respectively, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma)^\top$ is the vector of unknown parameters; see, e.g., Kalbfleisch and Prentice (2002). Therefore, the log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^n \delta_i (\log f(e_i) - \log \sigma) + (1 - \delta_i) \log S(e_i),$$

where $e_i = (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/\sigma$.

The components of the score vector are given by

$$U_{\beta_j}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \frac{1}{\sigma} \sum_{i=1}^n a_i x_{ij}, \quad \text{for } j = 0, \dots, p,$$

and

$$U_{\sigma}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = \frac{1}{\sigma} \sum_{i=1}^n (e_i a_i - \delta_i),$$

where

$$a_i = - \left[\delta_i \frac{d \log f(e_i)}{de_i} + (1 - \delta_i) \frac{d \log S(e_i)}{de_i} \right]. \quad (2)$$

In matrix form, the score vector can be written as

$$U(\boldsymbol{\theta}) = \begin{bmatrix} U_{\beta}(\boldsymbol{\theta}) \\ U_{\sigma}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \sigma^{-1} \mathbf{X}^T \mathbf{a} \\ \sigma^{-1} (\mathbf{e}^T \mathbf{a} - \mathbf{1}^T \boldsymbol{\delta}) \end{bmatrix},$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ is the $n \times (p + 1)$ matrix of the covariates, and $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$, $\mathbf{e} = (\exp\{e_1\}, \exp\{e_2\}, \dots, \exp\{e_n\})^T$ and $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$ are n dimensional column vectors. Table 1 gives the expression for a_i in (2) for AFT models frequently used in survival data applications. The expression for a_i for the exponential distribution equals the corresponding a_i for the Weibull distribution with $\sigma = 1$. Maximum likelihood estimates (MLEs) for $\boldsymbol{\beta}$ and σ are obtained by solving the system of equations $U(\boldsymbol{\theta}) = \mathbf{0}$, which requires a numerical nonlinear optimization algorithm (e.g., Newton-Raphson, Fisher's scoring or BFGS). For further details on nonlinear optimization, see Press, Teulosky, Vetterling, and Flannery (1992). The `survreg` function in the `survival` package in R (R Development Core Team, 2011) uses the Newton-Raphson algorithm; see (Therneau & Lumley, 2008).

Table 1. Expression for a_i in (2) for some common models

Model	Errors distribution	a_i
Weibull	standard extreme value	$\exp\{e_i\} - \delta_i$
Log-normal	standard normal	$\delta_i e_i + \left(\frac{1 - \delta_i}{1 - \Phi(e_i)} \right) \frac{d}{de_i} \Phi(e_i)$
Log-logistic	standard logistic	$\frac{\exp\{e_i\}}{1 + \exp\{e_i\}} (1 + \delta_i) - \delta_i$

Note: $\Phi(\cdot)$ is the standard normal cumulative distribution function.

We now turn to hypothesis testing. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are column vectors of dimensions m and $k - m$, respectively. Consider the null hypothesis $\mathcal{H}_0: \boldsymbol{\theta}_2 = \boldsymbol{\theta}_{02}$ to be tested against $\mathcal{H}_1: \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_{02}$, where $\boldsymbol{\theta}_{02}$ is a fixed $k - m$ dimensional column vector. The partition in $\boldsymbol{\theta}$ induces the corresponding partition in the score vector $U(\boldsymbol{\theta}) = (U_{\theta_1}^T(\boldsymbol{\theta}), U_{\theta_2}^T(\boldsymbol{\theta}))^T$. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^T, \widehat{\boldsymbol{\theta}}_2^T)^T$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^T, \widetilde{\boldsymbol{\theta}}_2^T)^T$ be the unrestricted and the restricted MLE of $\boldsymbol{\theta}$ under \mathcal{H}_0 , respectively. The likelihood ratio statistic (ξ_{LR}) and the gradient statistic (ξ_G) for testing \mathcal{H}_0 against \mathcal{H}_1 are given by

$$\xi_{LR} = 2 \{l(\widehat{\boldsymbol{\theta}}) - l(\widetilde{\boldsymbol{\theta}})\}$$

and

$$\xi_G = U_{\theta_2}(\widetilde{\boldsymbol{\theta}})^T (\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{02}),$$

respectively. Under typical regularity conditions (e.g., Lawless, 2003, Ch. 6; Kalbfleisch & Prentice, 2002, Ch. 3), ξ_{LR} and ξ_G have a χ^2 distribution with $k - m$ degrees of freedom under \mathcal{H}_0 .

3. Simulation

We now present a simulation study for the Weibull, log-normal and log-logistic AFT models. Our goal is to evaluate and compare the performance of the likelihood ratio and gradient tests. The regression structure is as follows:

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma \epsilon_i, \text{ for } i = 1, \dots, n,$$

The values for x_{i1} , x_{i2} and x_{i3} were obtained as random draws of a normal distribution with mean of 1, variance of 0.25, an exponential with mean of 1/4 and a Bernoulli with a parameter of 0.5, respectively. The error terms ϵ_i were generated as independent random variables with a standard extreme value distribution for the exponential and

Weibull cases, a standard normal distribution for the log-normal case, and a standard logistic distribution for the log-logistic case. The censoring times C_i were generated as independent random variables with a uniform distribution on the interval $[0; c]$, where c was suitably chosen to produce different proportions of censored observations: 0% (no censoring), 30% and 50%. Four different sample sizes were considered: 30, 50, 100 and 200. Additionally, we consider $\beta^T = (0.15, 0.15, 0, 0)$ and set the following values for σ : 0.5, 1 and 1.5. For the Weibull case, these three different values for σ imply increasing, constant (exponential case) and decreasing failures rates.

Table 2. Null percentage rejection rates of $\mathcal{H}_0: \beta_2 = \beta_3 = 0$ for the likelihood ratio and gradient tests for different values of σ , sample size (n) and censoring percentage (c.p.): Weibull AFT model

σ	c.p.	$\alpha(\%)$	$n = 30$		$n = 50$		$n = 100$		$n = 200$	
			ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G
0.5	0	1	1.5	1.4	1.6	1.0	1.3	0.9	1.1	1.0
		5	7.6	5.1	6.4	5.0	6.2	5.4	5.6	5.1
		10	14.1	11.2	12.5	10.6	11.7	10.9	10.7	10.4
	30	1	2.3	1.0	1.7	1.0	1.5	1.2	1.2	1.1
		5	8.6	5.9	7.0	5.6	6.0	5.3	5.6	5.4
		10	15.1	12.0	12.8	11.2	11.5	10.5	11.3	11.0
	50	1	2.9	1.7	1.8	1.4	1.5	1.4	1.1	1.0
		5	9.4	7.0	7.3	6.3	6.6	6.0	5.4	5.2
		10	16.2	13.4	14.2	12.7	12.3	11.5	10.8	10.5
1	0	1	2.1	0.7	1.5	0.9	1.2	0.8	1.1	0.9
		5	8.1	5.5	6.6	5.3	5.6	5.1	5.6	5.0
		10	14.0	11.0	12.2	10.5	11.4	10.7	11.1	10.6
	30	1	2.6	1.5	1.5	0.9	1.4	1.1	1.1	1.1
		5	9.0	6.9	7.1	5.9	6.1	5.5	5.5	5.3
		10	15.8	13.7	13.0	11.8	11.7	11.0	11.0	10.7
	50	1	2.7	2.0	1.5	1.2	1.4	1.2	0.9	0.8
		5	9.9	7.9	6.6	5.7	6.0	5.8	5.4	5.1
		10	16.9	15.1	12.3	11.4	11.5	11.1	10.2	10.2
1.5	0	1	2.2	0.7	1.5	0.9	1.2	0.9	1.3	1.1
		5	8.9	5.9	7.2	5.3	5.7	5.1	5.4	5.0
		10	15.2	12.4	13.1	11.3	11.1	10.2	10.5	10.2
	30	1	2.1	1.1	1.5	1.3	1.6	1.3	1.1	1.0
		5	8.8	6.5	6.8	5.8	6.2	5.7	5.2	5.0
		10	14.9	12.8	13.0	12.0	11.6	11.0	10.3	10.0
	50	1	2.5	1.6	1.5	1.2	1.3	1.1	1.2	1.1
		5	9.3	7.6	6.9	6.3	6.0	5.6	5.4	5.5
		10	15.1	13.6	12.9	12.0	11.5	11.1	10.8	10.6

Based on 10,000 simulation replicates we estimated the null rejection rates of the likelihood ratio and the gradient tests of $\mathcal{H}_0: \beta_2 = \beta_3 = 0$ for three nominal levels ($\alpha = 0.01, 0.05$ and 0.10). We also considered the null hypothesis $\mathcal{H}_0: \beta_2 = 0$, but the results are similar and not presented here for brevity. The simulations were performed in R, with the optimization of the likelihood function obtained using the `survreg` function in the package `survival` with default initial values. Tables 2, 3 and 4 reveal the well-known liberal tendency of the likelihood ratio test when the sample is not large. In fact, for all of the cases considered with $n \leq 100$, the null rejection rates of the likelihood ratio test exceed the corresponding nominal level regardless of whether censoring is present or not, and regardless of the error distribution. For instance, for the Weibull case (Table 2), when $n = 30$, $\sigma = 0.5$ and $\alpha = 5\%$ the null rejection rates are 7.6% (no censoring), 8.6% (30% censoring) and 9.4% (50% censoring). The gradient test exhibits some liberal tendency, but this tendency is much less pronounced than that of the likelihood ratio test. In the aforementioned case the null rejection rates of the gradient test are 5.1%, 5.9% and 7.0%, respectively. It

can be noticed that the performance of both tests deteriorates slightly as the censoring percentage increases. We conclude that the gradient test is less size distorted than the likelihood ratio test.

Table 3. Null percentage rejection rates of $\mathcal{H}_0: \beta_2 = \beta_3 = 0$ for the likelihood ratio and gradient tests for different values of σ , sample size (n) and censoring percentage (c.p.): log-normal AFT model

σ	c.p.	$\alpha(\%)$	$n = 30$		$n = 50$		$n = 100$		$n = 200$	
			ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G
0.5	0	1	1.8	0.8	1.6	1.1	1.3	1.1	1.0	0.9
		5	7.3	5.4	6.2	5.3	5.7	5.2	5.5	5.2
		10	13.2	11.2	11.7	10.6	11.0	10.5	10.5	10.2
	30	1	1.9	0.8	1.5	0.9	1.2	0.9	1.0	0.9
		5	8.0	5.5	6.8	5.5	5.8	5.3	5.5	5.2
		10	14.3	11.9	12.5	11.0	11.6	10.9	10.6	10.3
	50	1	2.3	0.9	1.7	1.0	1.4	1.1	1.1	0.9
		5	8.6	45.9	7.2	5.8	5.6	5.0	5.2	4.9
		10	15.4	12.2	13.6	12.0	11.2	10.3	10.7	10.4
1	0	1	1.9	0.8	1.6	0.9	1.2	0.9	1.0	1.0
		5	7.3	5.5	6.3	5.2	5.9	5.5	5.5	5.2
		10	13.6	11.5	11.8	10.7	11.2	10.8	10.4	10.1
	30	1	1.8	0.9	1.5	0.8	1.2	1.0	1.2	1.1
		5	7.7	5.6	6.4	5.4	5.4	5.1	5.4	5.2
		10	14.2	11.8	12.5	11.1	11.0	10.3	10.4	10.2
	50	1	1.8	0.8	1.4	1.0	1.3	1.1	1.2	1.0
		5	7.7	5.2	6.5	5.2	6.0	5.4	5.6	5.1
		10	14.4	11.4	12.4	11.0	11.4	10.9	10.5	10.2
1.5	0	1	1.8	0.8	1.6	1.0	1.4	1.1	1.1	1.0
		5	7.4	5.4	6.3	5.3	5.6	5.0	5.3	5.0
		10	13.3	11.2	12.1	10.8	10.8	10.3	10.5	10.1
	30	1	1.9	0.7	1.4	0.8	1.1	0.8	1.2	1.1
		5	7.7	5.5	6.2	5.2	5.8	5.4	5.6	5.4
		10	13.7	11.4	12.0	10.8	11.3	10.6	11.0	10.8
	50	1	2.2	1.0	1.6	0.9	1.5	1.3	1.2	1.2
		5	8.2	6.0	6.9	5.8	6.4	5.7	5.5	5.2
		10	14.7	12.5	12.9	11.6	11.7	11.2	11.0	10.7

Table 4. Null percentage rejection rates of $\mathcal{H}_0: \beta_2 = \beta_3 = 0$ for the likelihood ratio and gradient tests for different values of σ , sample size (n) and censoring percentage (c.p.): log-logistic AFT model

σ	c.p.	$\alpha(\%)$	$n = 30$		$n = 50$		$n = 100$		$n = 200$	
			ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G	ξ_{LR}	ξ_G
0.5	0	1	1.9	0.9	1.6	0.8	1.2	0.9	1.0	0.8
		5	7.7	5.3	7.2	5.8	5.7	5.2	5.5	5.1
		10	14.1	11.7	13.2	11.5	11.1	10.5	10.6	10.1
	30	1	2.4	1.1	1.6	0.9	1.3	1.1	1.1	1.0
		5	8.5	6.1	6.4	5.1	5.9	5.2	5.2	5.0
		10	14.6	12.0	12.1	10.7	11.0	10.4	10.1	9.7
	50	1	2.1	0.9	1.4	0.9	1.3	1.1	1.2	1.0
		5	8.5	5.8	6.9	5.6	5.8	5.2	5.7	5.3
		10	15.3	12.5	13.1	11.7	11.4	10.8	10.7	10.4
1	0	1	1.8	0.8	1.5	0.9	1.3	1.1	1.0	0.9
		5	7.5	5.5	6.5	5.5	5.8	5.3	5.1	4.9
		10	13.9	11.3	12.4	11.2	11.3	10.7	10.3	10.0
	30	1	1.8	0.8	1.6	1.0	1.4	1.1	1.0	0.8
		5	7.7	5.6	6.9	5.8	6.0	5.5	5.1	4.9
		10	14.2	11.9	12.5	11.3	11.4	10.7	10.3	10.0
	50	1	1.9	0.9	1.5	1.1	1.3	1.1	1.2	1.0
		5	7.8	5.9	6.8	5.7	5.8	5.4	5.2	5.0
		10	14.0	12.1	12.6	11.5	11.2	10.7	10.4	10.2
1.5	0	1	1.9	0.9	1.4	0.9	1.2	0.9	1.1	0.9
		5	7.7	5.6	6.2	5.0	5.8	5.3	5.4	5.1
		10	14.0	11.8	12.3	10.8	11.3	10.6	10.6	10.2
	30	1	2.0	0.8	1.7	1.1	1.4	1.0	1.1	1.0
		5	7.4	5.4	6.8	5.6	5.7	5.0	5.5	5.2
		10	13.7	11.5	12.6	11.2	11.1	10.4	11.2	10.8
	50	1	2.1	1.1	1.4	0.9	1.4	1.3	1.1	1.0
		5	7.9	6.0	6.6	5.7	6.1	5.8	5.6	5.4
		10	14.1	12.1	12.5	11.7	11.7	11.3	10.6	10.5

We now turn to the investigation of the finite-sample power properties of the two tests. As our size simulations show, the tests have different sizes. To perform power comparisons, we first ensure that the tests have the correct size under the null hypothesis. To this end, we used 100,000 Monte Carlo simulated samples, drawn under the null hypothesis, to estimate the exact critical value of each test for the chosen nominal level. For the power simulations we considered the Weibull AFT models and computed the rejection rates under the alternative hypotheses $\mathcal{H}_{(\epsilon)}$: $\beta_2 = \beta_3 = \epsilon$, for ϵ ranging from -1.5 to 1.5 . The results for the Weibull AFT model are presented in Figure 1. The plots for the log-normal and log-logistic AFT models are similar and are not presented here for brevity. It can be noticed that the powers of both tests are strongly influenced by the scale parameter, with larger values of σ corresponding to lower power. As expected, increasing censoring percentage is accompanied by a loss of power for both tests. Additionally, the panels in Figure 1 suggest that the likelihood ratio and gradient tests have similar powers.

In short, the likelihood ratio and the gradient tests are equally powerful, but the likelihood ratio test is liberal when the sample size is not very large, while the gradient test is clearly much less size-distorted. Our overall conclusion is that the gradient test is more reliable than the likelihood ratio test when the sample size is small or moderate, and it should be preferred in practical applications.

4. Real Data Applications

To illustrate the use of the likelihood ratio and gradient tests we present two data analyses. The first concerns the failure of sub-surface equipment in a sample of oil wells obtained from an oil-drilling company. The second

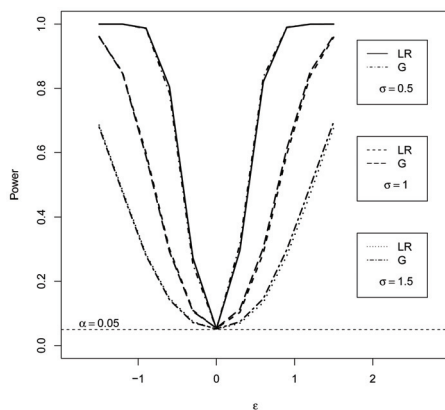
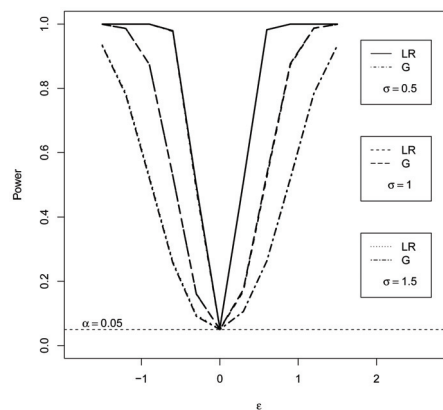
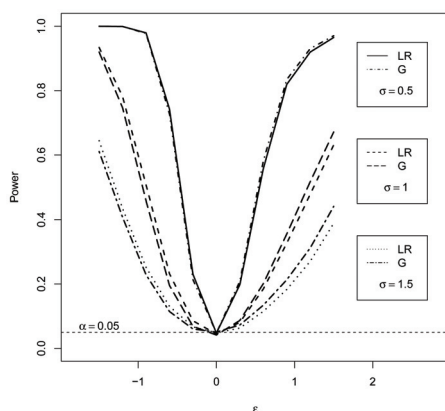
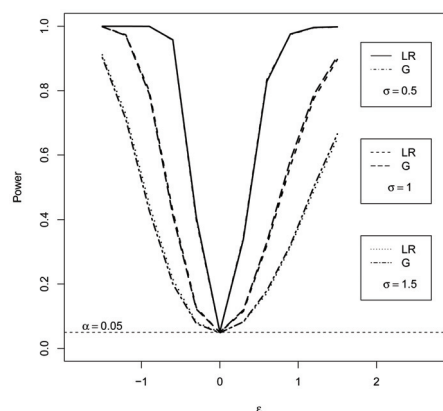
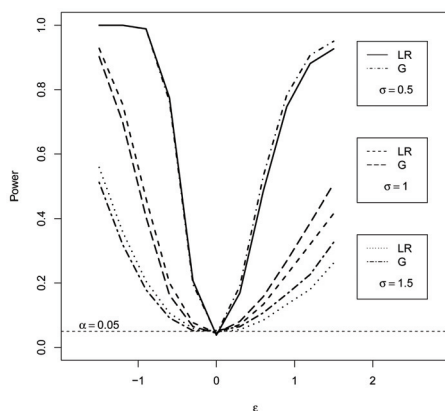
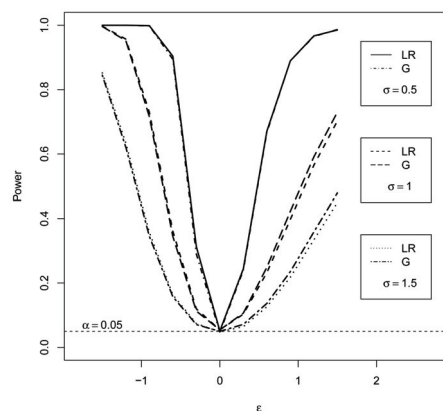
(a) $n = 30$, no censoring(b) $n = 50$, no censoring(c) $n = 30$, 30% censoring(d) $n = 50$, 30% censoring(e) $n = 30$, 50% censoring(f) $n = 50$, 50% censoring

Figure 1. Power of the likelihood ratio and gradient tests for different values of σ , sample size (n) and censoring percentage: Weibull AFT model

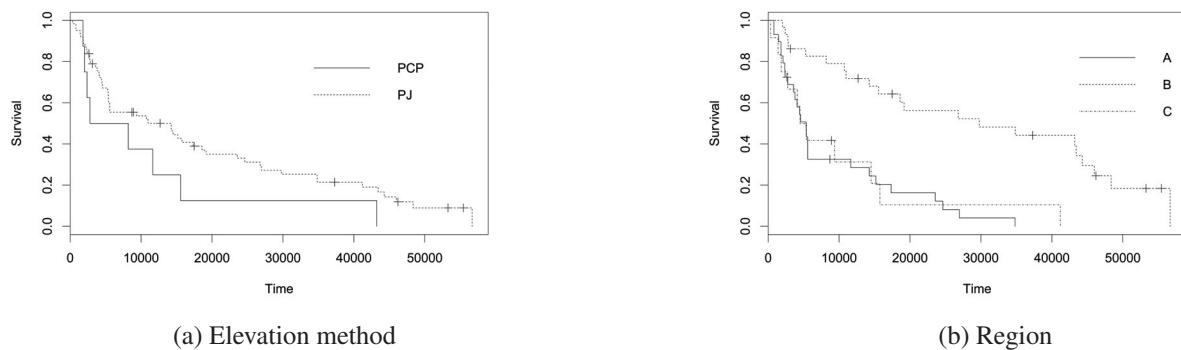


Figure 2. Kaplan-Meier estimate of the survival functions; oil wells data

analysis concerns the survival times in patients with lung cancer. In both cases we use only a randomly selected part of the full data to obtain subsamples of moderate size (less than 100). All of the calculations and plots were performed using the `survreg` function in the R `survival` package.

We first consider a retrospective study related to the operating time of oil wells from January 2000 to December 2006. For most of the wells it is necessary to install equipment to apply pressure at the bottom to lift the fluid to the surface. The time between failures caused by problems in sub-surface equipment was observed for 354 wells. Santos and Valença (2012) consider an extension of the Weibull regression model to identify the factors that influence the time to failure of the wells. To illustrate the use of the studied statistics, we only consider the time until the first failure and use a subsample of $n = 70$ cases extracted at random from the original data set. The subsample contains 16% censored cases, while 19% of the cases are censored in the original data set. The analysis includes two factors, namely: the *elevation method* (PJ: pumpjack; PCP: progressive cavity pump) and *administrative unit* (three operating units in different geographical locations: A, B and C). Plots of the Kaplan-Meier estimate of the survival functions for each administrative unit and elevation method are shown in Figure 2. It is apparent that the pumpjack method provides a slightly longer operating time for the wells than the progressive cavity pump method. Additionally, wells located at B seem to have higher survival times than those at A or C.

The regression model considered here is the Weibull accelerated failure time model in which

$$\log T_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \sigma\epsilon_i, \quad (3)$$

for $i = 1 \dots 70$, where the ϵ_i 's are independent random errors with a standard extreme value distribution, x_{i1} indicates the elevation method ($x_{i1} = 0$ for PCP and $x_{i1} = 1$ for PJ), and x_{i2} and x_{i3} indicate the administrative unit ($x_{i2} = 1$ for A and $x_{i3} = 1$ for C). Our goal is to test each factor in the presence of the other, i.e., the null hypotheses under test are $\mathcal{H}_{10}: \beta_2 = \beta_3 = 0$ and $\mathcal{H}_{20}: \beta_1 = 0$. The results for the likelihood ratio test and gradient test are given in Table 5. We note that both tests indicate that *administrative unit* is strongly significant. On the other hand, at the 5% significance level, *elevation method* is considered significant by the likelihood ratio test (p -value = 0.041) but not by the gradient test (p -value = 0.054). On the basis of our simulation results in a similar situation (here, $n = 70$, 16% censoring and $\hat{\sigma} = 0.951$), the gradient test is more reliable and hence it is the one to be taken into account.

Table 5. Summary of inference results: Oil wells data

Factor	Parameter	Estimate (SE)	ξ_{LR}	p -value	ξ_G	p -value	df
Elevation method	β_1	0.841(0.371)	4.195	0.041	3.701	0.054	1
Administrative unit	β_2	-1.390(0.281)	26.284	< 0.001	22.432	< 0.001	2
	β_3	-1.365(0.366)					

We now consider the cancer data set available in the `survival` package in R. The data are survival times for patients with advanced lung cancer from the North Central Cancer Treatment Group. The original data set has 228 observations and 10 variables, and 28% of the observations are censored. To illustrate the performance of

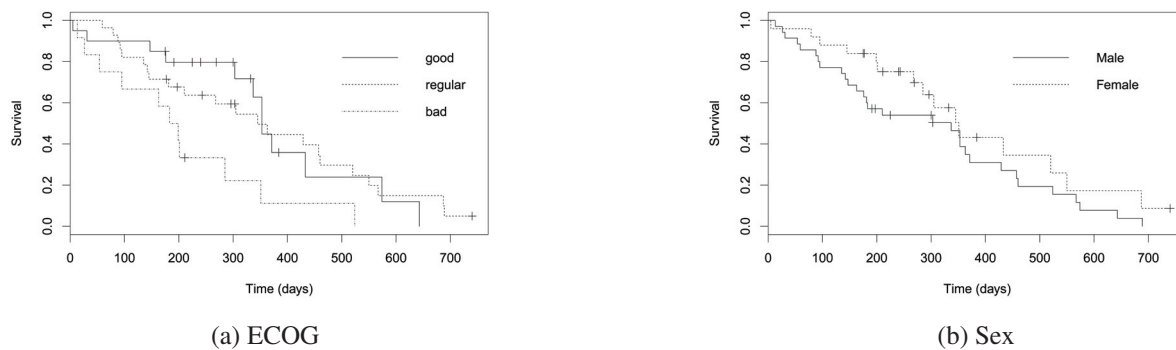


Figure 3. Kaplan-Meier estimate of the survival functions; lung cancer data

the statistical tests in a moderate-sized sample we randomly selected a subsample of 60 patients from the original sample, and use two factors, namely: *sex* (male and female) and *ECOG score* (Eastern Cooperative Oncology Group score). The score runs from 0 to 5, with 0 denoting perfect health and 5 denoting death. Our data only includes the first three levels (denoted here as 0 = good, 1 = regular, 2 = bad), and the censoring proportion is 26%. Figure 3 presents the Kaplan-Meier estimate of the survival functions for males and females and for the different ECOG scores.

The regression model considered here is given in (3) with x_{i1} indicating sex ($x_{i1} = 0$ for male and $x_{i1} = 1$ for female), and x_{i2} and x_{i3} indicating *ECOG score* ($x_{i2} = 1$ for regular and $x_{i3} = 1$ for bad). The inferential results are given in Table 6. We note that *sex* is considered highly significant by both tests. In contrast, at the 5% significance level, there is a conflict between the tests in terms of *ECOG score*: the likelihood ratio test indicates that the *ECOG score* is significant (p -value = 0.036) while the gradient test concludes that the *ECOG score* is not significant (p -value = 0.051). The simulation results suggest that inferential decisions should be based on the gradient test.

Table 6. Summary of inference results: Lung cancer data

Factor	Parameter	Estimate (SE)	ξ_{LR}	p -value	ξ_G	p -value	df
Sex	β_1	0.397(0.225)	3.451	0.063	3.531	0.060	1
ECOG	β_2	-0.164(0.250)	6.670	0.036	5.961	0.051	2
	β_3	-0.753(0.295)					

5. Concluding Remarks

We investigated and compared the performance of the likelihood ratio and gradient tests in parametric accelerated failure time models for survival data under random censoring. The Wald and score tests were not included in our simulation study because they require the computation of the Fisher information matrix, which cannot be obtained analytically for the models considered here. One could argue that the Fisher information should be replaced by the observed information matrix. However, the observed information produced negative standard errors for a non-negligible proportion of the simulated censored samples.

The simulation results suggest that the gradient and the likelihood ratio tests present similar powers. However, the gradient test clearly presents better size behavior than the likelihood ratio test, the latter being markedly liberal in small samples. Not surprisingly, the performance of both tests is sensitive to changes in the censoring percentage.

Our overall conclusion is that the gradient test should be preferred in practical applications when the sample is small or of moderate size.

Acknowledgements

We thank an anonymous referee for constructive comments. We gratefully acknowledge grants from the Brazilian agencies CAPES, CNPq and FAPESP, and Brazil's Science Without Borders Program.

References

- Carrasco, J. M. F., Ortega, E. M. M., & Cordeiro, G. M. (2008). A generalized modified Weibull distribution for lifetime modelling. *Computational Statistics & Data Analysis*, 53, 450-462. <http://dx.doi.org/10.1016/j.csda.2008.08.023>
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, B*, 34, 187-220.
- Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.
- Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72, 557-565. <http://dx.doi.org/10.1080/01621459.1977.10480613>
- Ferrari, S. L. P., & Pinheiro, E. C. (2014). Small-sample likelihood inference in extreme-value regression models. *Journal of Statistical Computation and Simulation*, 84, 582-595. <http://dx.doi.org/10.1080/00949655.2012.720686>
- Gimenez, P., & Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models. *Australian Journal of Statistics*, 39, 325-344. <http://dx.doi.org/10.1111/j.1467-842X.1997.tb00695.x>
- Grover, G., Das R., Swain, P. K., & Deka, B. (2013). On the estimation of survival of HIV/AIDS patients on antiretroviral therapy using NPMLE method: an application to interval censored data. *American Journal of Mathematics and Statistics*, 4, 213-219. <http://dx.doi.org/10.5923/j.ajms.20130304.05>
- Kalbflesch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). Singapore: Wiley.
- Lai, C. D. (2013). Constructions and applications of lifetime distributions. *Applied Stochastic Models in Business and Industry*, 29, 127-140. <http://dx.doi.org/10.1002/asmb.948>
- Lai, C. D., Xie, M., & Murthy, D. N. P. (2003). A modified Weibull distribution. *IEEE Transactions on Reliability*, 52, 33-37. <http://dx.doi.org/10.1109/TR.2002.805788>
- Lambert T, P., Collett, D., Kimber, A., & Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, 23, 3177-3192. <http://dx.doi.org/10.1002/sim.1876>
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). New Jersey: Wiley.
- Lemonte, A. J., & Ferrari, S. L. P. (2011a). Size and power properties of some tests in the Birnbaum-Saunders regression model. *Computational Statistics and Data Analysis*, 55, 1109-1117. <http://dx.doi.org/10.1016/j.csda.2010.09.008>
- Lemonte, A. J., & Ferrari, S. L. P. (2011b). Testing hypotheses in the Birnbaum-Saunders distribution under type-II censored samples. *Computational Statistics and Data Analysis*, 55, 2388-2399. <http://dx.doi.org/10.1016/j.csda.2011.02.005>
- Lemonte, A. J., & Ferrari, S. L. P. (2012). The local power of the gradient test. *Annals of the Institute of Statistical Mathematics*, 64, 373-381. <http://dx.doi.org/10.1007/s10463-010-0315-4>
- Maller, R. A., & Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley.
- Marshall, A. W., & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641-652. <http://dx.doi.org/10.1093/biomet/84.3.641>
- Nardi, A., & Schemper, M. (2003). Comparing Cox and parametric models in clinical studies. *Statistics in Medicine*, 22, 3597-3610. <http://dx.doi.org/10.1002/sim.1592>
- Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, 64, 441-448. <http://dx.doi.org/10.2307/2345319>
- Orbe, J., Ferreira, E., & Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, 21, 3493-3510. <http://dx.doi.org/10.1002/sim.1251>

- Ortega, E. M. M., Cancho, V. G., & Paula, G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, 15, 79-106. <http://dx.doi.org/0.1007/s10985-008-9096-y>
- Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, 7, 55-64. <http://dx.doi.org/10.1023/A:1009625210191>
- Peng, Y., Dear, K. B. G., & Denham, J. W. (1998). A generalized F mixture model for cure rate estimation. *Statistics in Medicine*, 17, 813-830. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8<813::AID-SIM775>3.0.CO;2-#](http://dx.doi.org/10.1002/(SICI)1097-0258(19980430)17:8<813::AID-SIM775>3.0.CO;2-#)
- Press, W. H., Teulosky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.). London: Prentice Hall.
- R Development Core Team. (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Santos, P. B., & Valença, D. M. (2012). Accelerated failure time models with random effects to estimate operation time of oil wells. *Advances and Applications in Statistics*, 28, 23-44.
- Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34, 206-215.
- Therneau, T., & Lumley, T. (2008). *Survival: Survival analysis, including penalised likelihood*. R package version 2.34-1.
- Valença, D. M., & Bolfarine, H. (2006). Testing homogeneity in Weibull error in variables models. *Annals of the Institute of Statistical Mathematics*, 58, 115-129. <http://dx.doi.org/10.1007/s10463-005-0006-8>
- Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871-1879. <http://dx.doi.org/10.1002/sim.4780111409>
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "Permanent Employment" in Japan. *Journal of the American Statistical Association*, 87, 284-292. <http://dx.doi.org/10.1080/01621459.1992.10475207>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).