SORT 44 (1) January-June 2020, 127-140 DOI: 10.2436/20.8080.02.97

Bartlett and Bartlett-type corrections for censored data from a Weibull distribution

Tiago M. Magalhães[1] and Diego I. Gallardo[2]

Abstract

In this paper, we obtain the Bartlett factor for the likelihood ratio statistic and the Bartlett-type correction factor for the score and gradient test in censored data from a Weibull distribution. The expressions derived are simple, we only have to define a few matrices. We conduct an extensive Monte Carlo study to evaluate the performance of the corrected tests in small sample sizes and we show how they improve the original versions. Finally, we apply the results to a real data set with a small sample size illustrating that conclusions about the regressors could be different if corrections were not applied to the three mentioned classical statistics for the hypothesis test.

MSC: Primary 62Fxx; secondary 62F12.

Keywords: Bartlett correction, censored data, Weibull distribution, chi-squared distribution, maximum likelihood estimates, type I and II censoring.

1 Introduction

Hypothesis testing is an essential step in statistical inference in order to help investigators identify and understand the effect of covariates on the response variable. Survival regression models are required when the response variable is censored, i.e., only partial information is available. Parametric survival models are often used in health economic applications (Latimer, 2013) because the survival function is fully specified (Ishak et al., 2013) and data from multiple time periods can be easily combined (Benaglia, Jackson and Sharples, 2015).

The likelihood ratio (LR), Wald, score and gradient tests are commonly used for hypothesis testing. Under the null hypothesis ($\mathcal{H}$), each test statistic is asymptotically chi-squared distributed, i.e., the four statistics are asymptotically equivalent. Since they are coupled with asymptotic properties, the chi-squared distribution may not be a good approximation to the null distribution of each statistic in small or moderate sample sizes,

[1] Department of Statistics, Institute of Exact Sciences, Federal University of Juiz de Fora, Juiz de Fora, Brazil. E-mail: tiago.magalhaes@ice.ufjf.br

[2] Departamento de Matemáticas, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile. E-mail: diego.gallardo@uda.cl

Received: September 2019

Accepted: April 2020

then the use of these statistics become less justifiable. In practical situations, this fact can produce a type I error that should be greater (or less) than the fixed nominal value (usually 1%, 5% or 10%).

An approach to improve inferences in small/moderate samples using in the LR test is the Bartlett correction (Bartlett, 1937; Lawley, 1956). In this approach, the LR statistic is multiplied by a correction factor. Bartlett-type corrections were also developed for the score and gradient statistics, see Cordeiro and Ferrari (1991) and Vargas, Ferrari and Lemonte (2013).

Our main goal in this paper is to improve the likelihood inference in censored data from a Weibull distribution, where the scale parameter is known. Two particular models are obtained from this case: the exponential and Rayleigh distributions, but if unknown, the scale parameter may be replaced by a consistent estimate. First, we derive the Bartlett and the Bartlett-type correction for these censored data models. Next, we perform Monte Carlo simulation experiments to evaluate and compare the finite-sample performance of the improved LR, score and gradient tests with the usual LR, Wald, score and gradient tests. To the best of our knowledge, Bartlett and Bartlett-type corrections for LR, score and gradient statistics in the Weibull survival model were not specified so far. Moreover, it is the first presentation of corrections for the gradient statistic in survival models. All these results are illustrated by a comprehensive simulation study.

The paper is structured as follows. In Section 2, we describe the censored data from a Weibull distribution and discuss estimation and hypothesis testing inference on the regression parameters. The Bartlett and the Bartlett-type correction factors are derived in Section 3. Monte Carlo simulation results are presented and discussed in Section 4. An empirical application that use real data are presented and discussed in Section 5. The paper closes up with a brief discussion in Section 6.

2 Weibull distribution

A continuous random variable T is called Weibull, denoted by $\mathrm{WE}(\theta,\sigma)$, if its probability density function (pdf) is

$$f(t;\theta,\sigma) = \frac{1}{\sigma\theta^{1/\sigma}} t^{1/\sigma-1} \exp\left\{-(t/\theta)^{1/\sigma}\right\},\ t>0,$$

where $\sigma > 0$ is the shape parameter and $\theta > 0$ is the scale parameter. Weibull distribution is commonly used in the analysis of time-to-event or lifetime data and it is still the aim of several works, as in Jafari and Zakerzadeh (2015), Nekoukhou and Bidram (2015), Lina, Williamson and Kim (2019), Magalhães, Gallardo and Gómez (2019), for instance. Two particular models under this parametrization are obtained for $\sigma = 1$ and $\sigma = 1/2$, which represent the exponential and the Rayleigh models with mean θ and $\theta\sqrt{\pi}/2$, respectively. In this work, we focused on those models. However, if σ is unknown, we assume that it can be replaced by a consistent estimate. In lifetime data, censoring is very common because of time limits and other restrictions on data collection. To describe these data, we consider that, for a sample size of n, $L_1, \dots, L_n$ are stochastically independent random variables representing the failure times and $T_1, \dots, T_n$ are stochastically independent Weibull random variables and independent of the $L's$, denoting the censoring times. Under the right censoring, the observed information is

$$t_i = \min(T_i, L_i) \quad \text{and} \quad \delta_i = \begin{cases} 1, & T_i \leq L_i \\ 0, & T_i > L_i \end{cases}.$$

For $L_1, \dots, L_n$ fixed, we have the type I mechanism and if $L_1, \dots, L_n = L$, a random variable, type II censoring. Under the assumption that the censoring times $L's$ do not depend on θ (known in the literature as a non-informative censoring assumption), we have that the log-likelihood function for the two types of censoring has the form

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sigma\theta^{1/\sigma}} t_i^{1/\sigma - 1} \right\}^{\delta_i} \exp\left\{ - (t_i/\theta)^{1/\sigma} \right\} \\ &= \left(\sigma\theta^{1/\sigma}\right)^{-r} \exp\left\{ \left(\frac{1}{\sigma} - 1\right) W_1 - \frac{1}{\theta^{1/\sigma}} W_2 \right\}, \end{aligned}$$

where $r = \sum_{i=1}^{n} \delta_i$, $W_1 = \sum_{i=1}^{n} \delta_i \log t_i$ and $W_2 = \sum_{i=1}^{n} t_i^{1/\sigma}$. The regression structure can be incorporated in the model by making $\theta_i = \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right)$, where $\boldsymbol{\beta}$ is a p-vector of parameters and $\mathbf{x}_i$ is a vector of regressors related to the i-th observation. Usually, the regression modelling considers the distribution of $Y_i = \log(T_i)$ instead of T_i. The distribution of Y_i is of the extreme value form with pdf given by

$$f(y_i; \mathbf{x}_i) = \frac{1}{\sigma} \exp\left\{ \frac{y_i - \mu_i}{\sigma} - \exp\left(\frac{y_i - \mu_i}{\sigma}\right) \right\}, \quad -\infty < y_i < \infty,$$

where $\mu_i = \log \theta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ is the linear predictor related to the i-th observation. The log-likelihood function derived from this regression model is given by

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\delta_i \left(-n \log \sigma + \frac{y_i - \mu_i}{\sigma} \right) - \exp\left(\frac{y_i - \mu_i}{\sigma} \right) \right].$$

The total score function and the total Fisher information matrix for $\boldsymbol{\beta}$ are given, respectively, by

$$\mathbf{U}_{\boldsymbol{\beta}} = \sigma^{-1} \mathbf{X}^\top \mathbf{W}^{1/2} \mathbf{v} \text{ and } \mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \sigma^{-2} \mathbf{X}^\top \mathbf{W} \mathbf{X},$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$, the model matrix, assuming the rank$(\mathbf{X}) = p$, $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$, $w_i = \mathbb{E}\left[\exp\left(\frac{y_i - \mu_i}{\sigma}\right)\right]$ and $\mathbf{v} = (v_1, \dots, v_n)^\top$, $v_i = \left\{-\delta_i + \exp\left(\frac{y_i - \mu_i}{\sigma}\right)\right\} w_i^{-1/2}$. It can be observed that the value of w_i depends on the mechanism of censoring. That means

$$w_i = 1 - \exp\left\{ -L_i^{1/\sigma} \exp(-\mu_i/\sigma) \right\} \text{ and } w_i = \frac{r}{n},$$

for types I and II censoring, respectively. The proofs are presented in Magalhães et al. (2019). The maximum likelihood estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, is the solution of $\mathbf{U}_{\beta} = 0$. The MLE $\hat{\boldsymbol{\beta}}$ cannot be expressed in closed-form. It is typically obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton nonlinear optimization algorithm. Under mild regularity conditions and in large samples, $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$, approximately.

Consider the p-dimensional parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, where $\boldsymbol{\beta}_1$ is a q-dimensional vector and $\boldsymbol{\beta}_2$ is the remaining $p-q$ parameters. In a test of hypotheses, the interest lies in $\mathcal{H} : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$, the null hypothesis, where $\boldsymbol{\beta}_1^{(0)}$ is a known q-vector, in other words, the null hypothesis imposes q restrictions on the parameter vector. Hence, $\boldsymbol{\beta}_2$ is the vector of nuisance parameters and $\boldsymbol{\beta}_1$ is the vector of interest parameters. This partition induces the corresponding partitions

$$\mathbf{U}_{\beta} = \left(\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top\right)^\top, \quad \text{with} \quad \mathbf{U}_{\beta_1} = \sigma^{-1}\mathbf{X}_1^\top \mathbf{W}^{1/2}\mathbf{v}, \quad \mathbf{U}_{\beta_2} = \sigma^{-1}\mathbf{X}_2^\top \mathbf{W}^{1/2}\mathbf{v},$$

$$\mathbf{K}_{\beta\beta} = \begin{pmatrix} \mathbf{K}_{\beta_1\beta_1} & \mathbf{K}_{\beta_1\beta_2} \\ \mathbf{K}_{\beta_2\beta_1} & \mathbf{K}_{\beta_2\beta_2} \end{pmatrix} = \sigma^{-2} \begin{pmatrix} \mathbf{X}_1^\top \mathbf{W}\mathbf{X}_1 & \mathbf{X}_1^\top \mathbf{W}\mathbf{X}_2 \\ \mathbf{X}_2^\top \mathbf{W}\mathbf{X}_1 & \mathbf{X}_2^\top \mathbf{W}\mathbf{X}_2 \end{pmatrix},$$

and $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$, $\mathbf{X}_1$, $\mathbf{X}_2$ being $n \times q$ and $n \times (p-q)$, respectively. The LR, score and gradient statistics for testing $\mathcal{H}$ can be expressed, respectively, as

$$\begin{aligned} \mathrm{SLR} &= 2\left[\ell\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \sigma\right) - \ell\left(\boldsymbol{\beta}_1^{(0)}, \tilde{\boldsymbol{\beta}}_2, \sigma\right)\right], \\ \mathrm{SR} &= \tilde{\mathbf{v}}^\top \tilde{\mathbf{W}}^{1/2}\mathbf{X}_1 \left(\tilde{\mathbf{R}}^\top \tilde{\mathbf{W}}\tilde{\mathbf{R}}\right)^{-1} \mathbf{X}_1^\top \tilde{\mathbf{W}}^{1/2}\tilde{\mathbf{v}}, \\ \mathrm{ST} &= \sigma^{-1}\tilde{\mathbf{v}}^\top \tilde{\mathbf{W}}^{1/2}\mathbf{X}_1 \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)}\right), \end{aligned}$$

where $\left(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2\right)$ and $\left(\boldsymbol{\beta}_1^{(0)}, \tilde{\boldsymbol{\beta}}_2\right)$ are the unrestricted and restricted MLEs of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, respectively, $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2\mathbf{A}$, with $\mathbf{A} = \left(\mathbf{X}_2^\top \mathbf{W}\mathbf{X}_2\right)^{-1} \mathbf{X}_2^\top \mathbf{W}\mathbf{X}_1$ represents a $(p-q) \times q$ matrix whose columns are the vectors of regression coefficients obtained in the weighted normal linear regression of the columns of $\mathbf{X}_1$ on the model matrix $\mathbf{X}_2$ with $\mathbf{W}$ as a weight matrix. Here, tildes and hats indicate quantities available at the restricted and unrestricted MLEs, respectively. Under the null hypothesis $\mathcal{H}$, these three statistics have an asymptotic χ_q^2 distribution with approximation error of order n^{-1}.

3 Improved inference

As discussed in Section 2, when the sample size is not sufficiently large, the chi-squared distribution may be a poor approximation to the null distribution of the statistics. Thus, it is paramount to obtain refinements for inference based on these tests.

From the second-order asymptotic theory, three works can be mentioned: Lawley (1956), Cordeiro and Ferrari (1991) and Vargas et al. (2013). These works obtained general correction factors, respectively, for the LR, score and gradient statistics, which reduced the approximation error of the asymptotic χ^2_q distribution from n^{-1} to n^{-2}. Those correction factors are based on the derivatives of the log-likelihood function.

From the result of Lawley (1956), we derived the specific Bartlett-correction factor for LR statistic for testing $\mathcal{H}: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$ in censored data from a Weibull distribution, it is given by

$$\begin{aligned}\varepsilon_p = {} & (1/4)\sigma^{-2}\mathrm{tr}\left\{\mathbf{F}_1\dot{\mathbf{Z}}^{(2)}\right\} + (1/12)\sigma^{-6}\mathbf{1}^\top\mathbf{W}\left(2\mathbf{Z}^{(3)} + 3\dot{\mathbf{Z}}\mathbf{Z}\dot{\mathbf{Z}}\right)\mathbf{W}\mathbf{1} \\ & + \sigma^{-5}\mathbf{1}^\top\mathbf{W}\left(\mathbf{Z}^{(3)} + \dot{\mathbf{Z}}\mathbf{Z}\dot{\mathbf{Z}}\right)\mathbf{W}'\mathbf{1} + \sigma^{-4}\mathbf{1}^\top\mathbf{W}'\left(\mathbf{Z}^{(3)} + \dot{\mathbf{Z}}\mathbf{Z}\dot{\mathbf{Z}}\right)\mathbf{W}'\mathbf{1},\end{aligned} \tag{1}$$

where $\mathbf{F}_1$, $\mathbf{W}'$, $\mathbf{Z}$ and $\dot{\mathbf{Z}}$ are given in the Appendix and all the algebraic manipulations are presented in the Supplementary Material, Section D.1. The three Bartlett corrected test statistics are

$$\mathrm{SLR}{*}1 = \frac{SLR}{(1+c)},\ \mathrm{SLR}{*}2 = \mathrm{SLR}\times\exp\{-c\} \quad \text{and} \quad \mathrm{SLR}{*}3 = \mathrm{SLR}\times(1-c),$$

where $c = (\varepsilon_p - \varepsilon_{p-q})/q$, both ε_p and ε_{p-q}, can be obtained from (1). The statistic SLR*1 is the original Bartlett corrected likelihood ratio statistic. However, the others are equivalent to order $\mathcal{O}(n^{-1})$. It is noteworthy that SLR*2 assumes only positive values.

From Cordeiro and Ferrari (1991), we have written the specific Bartlett-type corrected score statistic for censored data from a Weibull distribution as

$$\mathrm{SR}{*} = \mathrm{SR}\{1 - (c_R + b_R\mathrm{SR} + a_R\mathrm{SR}^2)\}, \tag{2}$$

where $a_R = A_{R3}/12q(q+2)(q+4)$, $b_R = (A_{R2} - 2A_{R3})/12q(q+2)$, $c_R = (A_{R1} - A_{R2} + A_{R3})/12q$ and, for the sake of brevity, the quantities A_{R1} to A_{R3} are presented in the Appendix.

For $\sigma = 1$, the expressions (1) and (2) reduce to exponential censored data case, derived by Cordeiro and Colosimo (1997) and Cordeiro and Colosimo (1999), respectively. For more details on the Bartlett and the Bartlett type corrections, see Cordeiro and Cribari-Neto (2014).

Finally, using the general result of Vargas et al. (2013), we obtained the specific Bartlett-type corrected gradient statistic for censored data from a Weibull distribution as

$$\mathrm{ST}{*} = \mathrm{ST}\left\{1 - \left(c_T + b_T\mathrm{ST} + a_T\mathrm{ST}^2\right)\right\}, \tag{3}$$

where $a_T = A_{T3}/12q(q+2)(q+4)$, $b_T = (A_{T2} - 2A_{T3})/12q(q+2)$, $c_T = (A_{T1} - A_{T2} + A_{T3})/12q$ and the quantities A_{T1} to A_{T3} are also presented in the Appendix. For further discussion about gradient test and its Bartlett-type correction as well, see Lemonte (2016).

4 Simulation studies

In this section, we present four simulation studies to assess different aspects of our proposal. The first study is related to evaluating the type I error from the different corrected statistics under different combinations of (p,q), σ, % of censoring (C) and sample sizes. The second study is devoted to assessing the power of the corrected statistics. The third study evaluates the behaviour of the corrected statistics if the assumption of known σ is changed by the respective estimate, i.e., $\sigma = \widehat{\sigma}$, where $\widehat{\sigma}$ is some consistent estimator of σ. Finally, the fourth study assessed the performance of the corrected statistics if the scheme used to draw the censoring times is random (instead of censoring type I or type II), but considering as they were censoring type I. In all studies, we considered three values for σ: $0.5, 1$ and 3; eight combinations for (p,q): $(3,1),(3,2),(5,1),(5,2),(5,3),(7,1),(7,3)$ and $(7,5)$; 3 values for C: $10\%, 25\%$ and 50%; and 3 sample sizes: 20, 30 and 40, totaling 216 cases. We also considered $\boldsymbol{\beta} = 0_p$, i.e., a vector of zeros with dimension p. However, only the first q components of $\boldsymbol{\beta}$ were tested to be zero. For each combination of σ, (p,q), % of censoring and sample size we considered 20,000 Monte Carlo replicates. Each vector of covariates $\mathbf{x}_i$ was drawn from the multivariate standard normal distribution with dimension p. Values from the Weibull model were drawn using the inverse transformation method and right censoring type II scheme was used, i.e., the first $n \times (1 - C/100)$ times (rounded to the upper whole number) represented a failure time and the rest of units were censored at the $(1 - C/100)$-th quantile. The exception was the simulation study 4, where a right censoring scheme was used. For each sample and considering σ as known (except for simulation study 3 where such parameter was estimated from the sample) we compute the traditional statistics SLR, SR and ST and their modified version discussed in Section 3 to test $H_0 : \boldsymbol{\beta}_q = 0_q$ versus H_1 : the contrary. In all cases, we reported the percentage of times where the respective test rejected the null hypothesis giving a specified type I error. All simulations were performed using the R software (R Core Team, 2017).

4.1 Assessing the type I error

In this simulation study, we evaluate the type I error for the usual versions of SLR, SR and ST and their corrected versions discussed in Section 3 to test $H_0 : \boldsymbol{\beta}_q = 0_q$ versus H_1 : the contrary. We consider the four scenarios for (p,q), σ, % of censoring and sample sizes mentioned in the introduction of this section. We report the percentage of times where the test rejected the null hypothesis with a 5% significance. Table 1 summarizes the cases where $\sigma = 0.5$, $C = 25\%$, $n = 30$. The complete results are presented in the Supplementary Material, Section B.1. In general, the correction produces a rejection rate closer to the nominal 5% significance in the three tests. Considering the 216 involved cases, the mean of the rejection rates was 7.9%, 6.1% and 8.1% for the SLR, SR and ST tests and 5.6%, 5.5%, 5.3%, 5.6% and 5.3% for the SLR*1, SLR*2, SLR*3, SR* and ST* tests, respectively, showing a better performance in average terms to the corrected statistics. We also compute the percentage of times where the corrected version of the test provides a rejection rate closer (in absolute value) to the nominal value. Such percentages were 99.5% for the SLR*1, SLR*2 and SLR*3, 69.0% for the SR* and 98.6% for the ST* test. Results suggest a huge improvement in the corrected version of the statistics when compared with their traditional pairs. As p and q increase, the differences in the rejection rates between the traditional statistics and the corrected ones seem to be getting larger. There are two possibles reasons, for a fixed sample size n: (a) Fixing q. As p increases, worse is the model fit and, consequently, the approximation to the null distribution of each statistic. (b) Fixing p. As q increases, there is the family-wise error rate (FWER), i.e., more restrictions in the null hypothesis make type I error larger, inflated. In the both situations, the corrected statistics seem to be less affected. Finally, we remark that the SR seems the most robust statistic among the three traditional statistics in this context.

Table 1: *Simulated rejection rates for $H_0 : \boldsymbol{\beta}_q = 0_q$, with $\sigma = 0.5$, $C = 25\%$, $n = 30$ and different values for p and q.*

p	q	SLR	SR	ST	SLR*1	SLR*2	SLR*3	SR*	ST*
3	1	7.0	6.2	7.0	5.8	5.8	5.7	5.9	5.7
	2	7.5	6.3	7.6	6.1	6.0	6.0	6.3	6.2
5	1	7.5	6.5	7.7	6.0	6.0	5.9	5.8	5.4
	2	8.3	6.5	8.4	6.1	6.0	5.9	6.0	5.9
	3	8.7	6.7	9.2	6.3	6.2	6.1	6.5	6.3
7	1	8.4	6.8	8.5	6.3	6.1	5.9	5.7	4.7
	3	10.4	7.5	10.7	6.8	6.6	6.4	6.6	6.3
	5	10.6	6.6	11.5	6.7	6.5	6.3	6.6	6.8

4.2 Assessing the power of the tests

In this simulation study, we assessed the power of the test for the usual versions of SLR, SR and ST and their corrected versions. We considered $n = 20$ and $p = 5$ in all the cases, q varying in the set $\{1,3\}$, σ in $\{0.5,1,3\}$ and C in $\{10\%,25\%,50\%\}$. To simulate the data, we further considered $\boldsymbol{\beta} = 0_p$. However, we have an interest in the hypothesis of the form $H_0 : \boldsymbol{\beta}_q = \psi\mathbf{1}_q$, where $\mathbf{1}_q$ is a vector of ones with dimension q and ψ is taken in the set $\{0.05, 0.10, 0.25, 0.50, 1.00, 2.00\}$. Table 2 shows the results for $\sigma = 1$ and $C = 10\%$. The complete results are presented in the Supplementary Material, Section B.2. As expected, the power of the test is increased when ψ is increased (because the value being tested is further than the value used to simulate the data) and when q is increased. We also noted that the power of each test is greater than its corrected version for some values of ψ and is lower than its corrected version for other values of ψ. Therefore, as usual in most problems related to hypothesis tests, there is no unique most powerful test. However, the powers of the three ordinary and corrected tests seem similar.

Table 2: *Simulated rejection rates to the corrected version of the SLR, SR and ST tests for* $H_0 : \boldsymbol{\beta}_q = \psi \mathbf{1}_q$, *with* $\sigma = 1$, $C = 10\%$, $n = 20$, $p = 5$ *and different values for q.*

		ψ					
q	statistic	0.05	0.10	0.25	0.50	1.00	2.00
1	SLR	6.9	8.0	16.5	42.0	86.9	99.7
	SR	5.4	6.6	15.6	41.9	86.3	99.6
	ST	6.9	8.1	16.7	41.9	86.9	99.7
	SLR*1	7.7	8.7	17.0	41.3	86.1	99.7
	SLR*2	7.7	8.7	16.9	41.2	86.0	99.7
	SLR*3	7.6	8.7	16.9	41.2	85.9	99.7
	SR*	5.4	6.6	15.6	41.9	86.3	99.6
	ST*	6.7	7.8	17.3	44.7	88.8	98.5
3	SLR	7.5	10.0	28.9	75.9	99.3	100.0
	SR	6.4	9.8	33.0	79.1	99.4	100.0
	ST	8.1	10.8	30.1	76.5	99.3	100.0
	SLR*1	8.2	10.9	29.3	75.5	99.3	100.0
	SLR*2	8.2	10.9	29.3	75.5	99.3	100.0
	SLR*3	8.1	10.9	29.2	75.5	99.3	100.0
	SR*	6.4	9.8	33.0	79.1	99.4	100.0
	ST*	7.8	10.6	31.4	78.8	98.5	98.9

4.3 Changing the assumption of σ known

Up to this moment, we considered σ as a known value. However, in practice we also need to estimate it. An alternative is to fix $\sigma = \widehat{\sigma}_{ML}$ and apply all the discussed methodology, where $\widehat{\sigma}_{ML}$ denotes the ML estimator of σ for the complete model (i.e., with all covariates). However, as we are working in a framework with a small sample size, the bias of $\widehat{\sigma}_{ML}$ can be considerable. Previous studies performed by us suggest that the performance of the corrected statistics does not differ substantially from the traditional statistics to test $\boldsymbol{\beta}_q = 0_q$ versus $H_1 : \boldsymbol{\beta}_q \neq 0_q$. For this reason, in this simulation study, we considered fixing $\sigma = \widehat{\sigma}_J$, where $\widehat{\sigma}_J$ is the jackknife estimator for σ. A third alternative not explored by us was to fix $\sigma = \widehat{\sigma}_B$, where $\widehat{\sigma}_B$ is a bootstrap estimator for σ. However, $\widehat{\sigma}_J$ provides satisfactory results and $\widehat{\sigma}_J$ is determined in a unique form to a fixed sample, whereas $\widehat{\sigma}_B$ typically is computed based on $B >> n$ bootstrap resample, which is not unique and is more expensive in computational terms. We consider the four scenarios for (p,q), σ, % of censoring and sample sizes mentioned in the introduction of this section. We report the percentage of times where the test rejected the null hypothesis with a 5% significance. Table 3 summarizes the cases where $\sigma = 1$, $C = 25\%$, $n = 20$. The complete results are presented in the supplementary material, Section B.3. Considering the 216 involved cases, the mean of the rejection rates was 12.2%, 6.8% and 8.0% for the SLR, SR and ST tests and 8.8%, 8.5%, 8.2%, 6.3% and 5.6% for the SLR*1, SLR*2, SLR*3, SR* and ST* tests, respectively, showing a better performance in average terms than the corrected statistics. We also compute the percentage of times where the corrected version of the test provides a rejection rate closer (in absolute value) to the nominal value. Such percentages were 100% for the SLR*1, SLR*2, SLR*3 and ST* and 73.6% for the SR*. Results suggest a huge improvement for the corrected version of the statistics when compared with their traditional counterparts.

Table 3: *Simulated rejection rates for $H_0 : \boldsymbol{\beta}_q = 0_q$, with $\sigma = 1$, $C = 25\%$, $n = 20$ and different values for p and q.*

p	q	SLR	SR	ST	SLR*1	SLR*2	SLR*3	SR*	ST*
3	1	6.9	5.6	6.9	5.6	5.5	5.4	5.2	5.2
	2	7.1	5.1	7.6	5.5	5.4	5.4	5.1	5.5
5	1	7.8	6.0	7.9	5.9	5.8	5.7	5.0	4.3
	2	8.9	6.0	9.1	6.0	5.8	5.5	5.3	5.2
	3	8.9	5.3	9.3	5.9	5.7	5.5	5.1	5.6
7	1	9.0	6.7	9.0	6.3	6.0	5.8	5.0	3.3
	3	11.5	6.6	12.0	6.7	6.4	6.0	5.6	5.0
	5	11.0	4.8	12.3	6.0	5.8	5.4	4.9	6.0

4.4 Changing the assumption of censoring type I or II

Up to this moment, all the development in this work was performed based on the assumption that the censoring scheme is either a type I or II. In this simulation study, we changed such assumption assuming that the censoring times $L_1,\ldots,L_n$ were independent random variables and independent from $T_1,\ldots,T_n$. For simplicity, we assumed that $L_i \sim WE(\lambda_i, 1)$, i.e., the exponential distribution with mean λ_i^{-1}. If the percentage of censoring times was fixed at $C\%$, we required that $\mathbb{P}(T_i > L_i) = C/100$. It was straightforward to show that such a condition was equivalent to

$$\int_0^\infty f_{T_i}(u;\theta_i,\sigma) \times e^{-\lambda_i u} du = C/100, \tag{4}$$

where $f_{T_i}(\cdot;\theta_i,\sigma)$ denotes the density function of $\text{WE}(\theta_i,\sigma)$. Then, for a fixed value for θ_i,σ and C it was possible to solve numerically (4) to find λ_i. The same four scenarios for (p,q), σ, % of censoring and sample sizes mentioned in the introduction of this section were considered. We reported the percentage of times where the test rejected the null hypothesis with a 5% significance. Table 4 summarizes the cases where $\sigma = 3$, $C = 50\%$, $n = 20$. The complete results are presented in the Supplementary Material, Section B.4. Considering the 216 involved cases, the mean of the rejection rates were 6.5%, 4.8% and 6.4% for the SLR, SR and ST tests and 5.1%, 5.1%, 5.0%, 4.8% and 4.9% for the SLR*1, SLR*2, SLR*3, SR* and ST* tests, respectively, showing a better performance in average terms to the corrected statistics. We also computed the percentage of times where the corrected version of the test provided a rejection rate closer (in absolute value) to the nominal value. Such percentages were 84.7%, 84.3% and 83.3% for the SLR*1, SLR*2 and SLR*3, respectively, 62.0% for the SR* and 73.1% for the ST* test. Furthermore, results suggest a huge improvement for the corrected version of the statistics when compared with their traditional pairs. The percentages of times where the corrected statistics are closer to the nominal value in comparison with the traditional statistics are lower than in simulation study 1, where σ is assumed as known. However, such percentages remain high and the correction is suggestible.

Table 4: *Simulated rejection rates for* $H_0 : \boldsymbol{\beta}_q = 0_q$, *with* $\sigma = 3$, $C = 50\%$, $n = 20$ *and different values for* p *and* q.

p	q	SLR	SR	ST	SLR*1	SLR*2	SLR*3	SR*	ST*
3	1	7.2	5.8	7.4	5.3	5.2	5.0	5.0	5.6
	2	7.5	4.6	7.7	5.5	5.4	5.2	4.7	5.9
5	1	9.0	6.9	9.1	6.2	5.8	5.5	5.3	6.0
	2	9.0	5.8	9.3	5.7	5.5	5.2	4.8	5.7
	3	10.0	4.8	10.1	6.0	5.7	5.4	4.6	5.9
7	1	10.7	8.2	10.8	6.9	6.4	5.8	5.5	5.8
	3	12.8	6.7	12.7	6.9	6.3	5.6	5.3	5.8
	5	13.3	4.5	12.3	6.7	6.2	5.6	4.8	5.6

5 Application

In this section we present a real data application related to clams in order to illustrate a case where conclusions obtained from a hypothesis test may be different if the corrections discussed in Section 3 are not considered in censored data from a Weibull distribution. In Section A of the Supplementary Material, we present a second application.

Clams data set

Bonnail et al. (2016) performed a study to assess sediment quality using the freshwater clam *Corbiculafluminea* to determine its adequacy as a biomonitoring tool in relation to theoretical risk indexes and regulatory thresholds. The clams were exposed to sediments contaminated with acid mine drainage (polymetallic acid lixiviate derived from mining activity). The study contains 27 observations with measurements, among other characteristics, of the dry weight tissue of the clams (`dry`, in gr), wet weight tissue (`wet`, in gr), shell length (`length`, in mm) and the concentration of scandium (`sc`), niobium (`nb`), beryllium (`be`) and terbium (`tb`) bioaccumulated in the soft tissue. These minerals were considered in micrograms per liter ($\mu g/L$). In this case, we focused on modelling the dry weight of such clams based on the rest of available information considering the Weibull regression model, i.e., $\texttt{dry}_i \sim \mathrm{WE}(\theta_i, \sigma)$, where

$$\log\theta_i = \mathbf{x}_i^\top\boldsymbol{\beta} = \beta_1\texttt{sc}_i + \beta_2\texttt{nb}_i + \beta_3\texttt{be}_i + \beta_4 + \beta_5\texttt{wet}_i + \beta_6\texttt{length}_i,$$

with $\mathbf{x}_i^\top = (\texttt{sc}_i, \texttt{nb}_i, \texttt{be}_i, 1, \texttt{wet}_i, \texttt{length}_i)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_6)^\top$, $i = 1, \ldots, 27$. The order of covariates was organized in order to test if minerals `sc`, `nb` and `be` explain the dry weight of the claims, i.e., to test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ versus $H_1 : \beta_j \neq 0$, for at least one $1 \leq j \leq 3$. For this particular problem $p = 6$, $q = 3$ and $n = 27$, so as the sample size is small a correction might be required in traditional tests. We estimated $\widehat{\sigma}_J = 0.0317$ based on the jackknife method, which was used as known in the computation of the different statistics to test H_0. Results for traditional and corrected versions of the SLR, SR and ST tests are presented in Table 5. Note that, without correction, only the SR test does not reject H_0 considering a significance of 10%. However, all the corrected versions of the tests do not reject the null hypothesis with the same level of significance. Therefore, we cannot conclude that minerals `sc`, `nb` and `be` explain the `dry` weight of clams. Finally, to test if the WE model is suitable for this data set, we compute the quantile residuals (Dunn and Smyth, 1996). If the model was correct, these residuals would behave as a random sample from the standard normal distribution. The Kolmogorov-Smirnov test to verify such a hypothesis provides p-values of 0.206 and 0.568 to the complete and reduced model, respectively. Therefore, the assumption of WE distribution is acceptable under any usual level of significance.

Table 5: *Different statistics to test H_0 in the clams data set.*

Statistic	SLR	SR	ST	SLR*1	SLR*2	SLR*3	SR*	ST*
Observed	6.54	4.98	6.63	5.82	5.78	5.73	5.32	6.00
p-value	0.088	0.173	0.085	0.121	0.123	0.125	0.150	0.112

6 Concluding remarks

Weibull distribution is used for the analysis of time-to-event or lifetime data, with the maximum likelihood theory as the main methodology to estimate the parameters. Hypotheses regarding these parameters are tested using the likelihood, score and gradient tests. However, in small or moderate sample sizes, these procedures can not be reliable. In this paper, we derived the respectively corrected versions that improve their performance. For simplicity, we focus on the Weibull with known shape parameter (σ) to find those expressions. Nonetheless, our results show good properties for the situation when σ can be replaced by a consistent estimate based on the jackknife method. We also present an application that illustrates the usefulness of the main result of the paper. The matrices expressions for the Bartlett and Bartlett-type corrections are quite simple to be implemented in statistical software as `R` (R Core Team, 2017), together with the library `flexsurv` (Jackson, 2016), for instance. Noteworthy, we did not use bootstrap corrections because they have different natures and they add three uncertainties: the number of replications, the size of each replication and the type, parametric or nonparametric, besides being computationally costly for practitioners.

Acknowledgments

We gratefully acknowledge grants from FONDECYT (Chile) 11160670. We also acknowledge the editor's, the referees' and Dr. Marcio Augusto Diniz's suggestions that helped us improve this work.

References

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proccedings of the Royal Society, A*, 160, 268–282.

Benaglia, T., Jackson, C. H. and Sharples, L. D. (2015). Survival extrapolation in the presence of cause specific hazards. *Statistics in Medicine*, 34, 796–811.

Bonnail, E., Sarmiento, A. M. DelValls, T. A. Nieto, J. M. and Riba, I. (2016). Assessment of metal contamination, bioavailability, toxicity and bioaccumulation in extreme metallic environments (iberian pyrite belt) using corbicula fluminea. *Science of the Total Environment*, 544, 1031–1044.

Cordeiro, G. M. and Colosimo, E. A. (1997). Improved likelihood ratio tests for exponential censored data. *Journal of Statistical Computation and Simulation*, 56, 303–315.

Cordeiro, G. M. and Colosimo, E. A. (1999). Corrected score tests for exponential censored data. *Statistics & Probability Letters*, 44, 365–373.

Cordeiro, G. M. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*. Springer.

Cordeiro, G. M. and Ferrari, S. L. P. (1991). A modified score statistic having chi-squared distribution to order n^{-1}. *Biometrika*, 78, 573–582.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236–244.

Ishak, K. J., Kreif, N., Benedict, A. and Muszbek, N. (2013). Overview of parametric survival analysis for health-economic applications. *PharmacoEconomics*, 31, 663–675.

Jackson, C. H. (2016). flexsurv: A platform for parametric survival modeling in r. *Journal of Statistical Software*, 70, 1–33.

Jafari, A. A. and Zakerzadeh, H. (2015). Inference on the parameters of the weibull distribution using records. *Statistics and Operations Research Transactions*, 39, 3–18.

Latimer, N. R. (2013). Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data: inconsistencies, limitations, and a pratical guide. *Medical Decision Making*, 33, 743–754.

Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43, 295–303.

Lemonte, A. J. (2016). *The Gradient Test: Another Likelihood-Based Test*. Academic Press.

Lina, H.-M., Williamson, J. M. and Kim, H.-Y. (2019). Firth adjustment for weibull current-status survival analysis. *Communications in Statistics - Theory and Methods*. doi.org/10.1080/03610926.2019.1606241.

Magalhães, T. M., Gallardo, D. I. and Gómez, H. W. (2019). Skewness of maximum likelihood estimators in the weibull censored data. *Symmetry*, 11, 1351.

Nekoukhou, V. and Bidram, H. (2015). The exponentiated discrete weibull distribution. *Statistics and Operations Research Transactions*, 39, 127–146.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Vargas, T. M., Ferrari, S. L. P. and Lemonte, A. J. (2013). Gradient statistic: Higher-order asymptotics and bartlett-type correction. *Electronic Journal of Statistics*, 7, 43–61.

Appendix

In order to express the corrected LR, score and gradient statistics, equations (1) to (3) , it is helpful to define the quantities $w_i' = -\sigma^{-1}L_i^{1/\sigma}\exp\{-L_i^{1/\sigma}\exp(-\mu_i/\sigma)-\mu_i/\sigma\}$, $w_i'' = -\sigma^{-1}w_i'\left[L_i^{1/\sigma}\exp(-\mu_i/\sigma)-1\right]$, $f_{1i} = -\sigma^{-2}w_i - 4\sigma^{-1}w_i' - 4w_i''$, $f_{2i} = -2\sigma^{-2}w_i^2 + 6\sigma^{-2}w_i + 10\sigma^{-1}w_i' + 5w_i''$, $f_{3i} = -3\sigma^{-2}w_i^2 + 9\sigma^{-2}w_i + 14\sigma^{-1}w_i' + 6w_i''$, $i = 1,\ldots,n$, and the following matrices:

$$
\begin{aligned}
&\mathbf{W}' = \mathrm{diag}(w_1',\ldots,w_n'),\ \mathbf{W}'' = \mathrm{diag}(w_1'',\ldots,w_n''),\\
&\mathbf{F}_1 = \mathrm{diag}(f_{11},\ldots,f_{1n}),\ \mathbf{F}_2 = \mathrm{diag}(f_{21},\ldots,f_{2n}),\ \mathbf{F}_3 = \mathrm{diag}(f_{31},\ldots,f_{3n}),\\
&\mathbf{Z} = \mathbf{X}\mathbf{K}_{\beta\beta}^{-1}\mathbf{X}^\top = \sigma^2\mathbf{X}\left(\mathbf{X}^\top\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^\top,\ \dot{\mathbf{Z}} = \mathrm{diagonal}\{\mathbf{Z}\},\\
&\mathbf{Z}_2 = \sigma^2\mathbf{X}_2\left(\mathbf{X}_2^\top\mathbf{W}\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top,\ \dot{\mathbf{Z}}_2 = \mathrm{diagonal}\{\mathbf{Z}_2\},\ \mathbf{Z}^{(2)} = \mathbf{Z}\odot\mathbf{Z},\ \mathbf{Z}^{(3)} = \mathbf{Z}^{(2)}\odot\mathbf{Z},
\end{aligned}
$$

where $\odot$ represents a direct product and $\mathbf{1}$ is an n-dimensional vector of ones.

The remaining quantities to define an improved statistic in the score test, see equation (2), are:

$$
\begin{aligned}
A_{R1} = {} & 3\sigma^{-6}\mathbf{1}^\top(\mathbf{W}+2\sigma\mathbf{W}')\dot{\mathbf{Z}}_2(\mathbf{Z}-\mathbf{Z}_2)\dot{\mathbf{Z}}_2(\mathbf{W}+2\sigma\mathbf{W}')\mathbf{1}\\
& + 6\sigma^{-6}\mathbf{1}^\top(\mathbf{W}+2\sigma\mathbf{W}')\dot{\mathbf{Z}}_2\mathbf{Z}_2\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)(2\mathbf{W}+3\sigma\mathbf{W}')\mathbf{1}\\
& + 6\sigma^{-6}\mathbf{1}^\top(3\mathbf{W}+4\sigma\mathbf{W}')\left[\mathbf{Z}_2^{(2)}\odot(\mathbf{Z}-\mathbf{Z}_2)\right](\mathbf{W}+2\sigma\mathbf{W}')\mathbf{1}\\
& - 6\sigma^{-2}\mathrm{tr}\left\{\mathbf{F}_2\dot{\mathbf{Z}}_2\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
A_{R2} = {} & -3\sigma^{-6}\mathbf{1}^\top(2\mathbf{W}+3\sigma\mathbf{W}')\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)\mathbf{Z}_2\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)(2\mathbf{W}+3\sigma\mathbf{W}')\mathbf{1}\\
& - 6\sigma^{-6}\mathbf{1}^\top(2\mathbf{W}+3\sigma\mathbf{W}')\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)(\mathbf{Z}-\mathbf{Z}_2)\dot{\mathbf{Z}}_2(\mathbf{W}+2\sigma\mathbf{W}')\mathbf{1}\\
& - 6\sigma^{-6}\mathbf{1}^\top(2\mathbf{W}+3\sigma\mathbf{W}')\left[\mathbf{Z}_2\odot(\mathbf{Z}-\mathbf{Z}_2)^{(2)}\right](2\mathbf{W}+3\sigma\mathbf{W}')\mathbf{1}\\
& + 3\sigma^{-2}\mathrm{tr}\left\{\mathbf{F}_3\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)^{(2)}\right\},
\end{aligned}
$$

$$
A_{R3} = \sigma^{-6}\mathbf{1}^\top(2\mathbf{W}+3\sigma\mathbf{W}')\left\{3\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)(\mathbf{Z}-\mathbf{Z}_2)\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)+2(\mathbf{Z}-\mathbf{Z}_2)^{(3)}\right\}(2\mathbf{W}+3\sigma\mathbf{W}')\mathbf{1}.
$$

The quantities A_{T1} to A_{T3}, in the equation (3), to define an improved gradient test are, respectively:

$$
\begin{aligned}
A_{T1} = {} & 3\sigma^{-6}\mathbf{1}^\top \mathbf{W}\left\{\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)(\mathbf{Z}+\mathbf{Z}_2)\dot{\mathbf{Z}}_2+\dot{\mathbf{Z}}_2(\mathbf{Z}-\mathbf{Z}_2)\dot{\mathbf{Z}}_2+2(\mathbf{Z}-\mathbf{Z}_2)\odot\mathbf{Z}_2^{(2)}\right\}\mathbf{W1}\\
& +6\sigma^{-5}\mathbf{1}^\top \mathbf{W}\left\{(\mathbf{Z}_2+\mathbf{Z})\odot\left(\mathbf{Z}^{(2)}-\mathbf{Z}_2^{(2)}\right)+\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)\left(\mathbf{Z}_2\dot{\mathbf{Z}}_2+\mathbf{Z}\dot{\mathbf{Z}}\right)\right.\\
& \left.+\,2\left[\dot{\mathbf{Z}}_2\left(\mathbf{Z}\dot{\mathbf{Z}}-\mathbf{Z}_2\dot{\mathbf{Z}}_2\right)+\mathbf{Z}_2^{(2)}\odot(\mathbf{Z}-\mathbf{Z}_2)\right]\right\}\mathbf{W}'\mathbf{1}\\
& +12\sigma^{-4}\mathbf{1}^\top \mathbf{W}'\left\{\dot{\mathbf{Z}}\mathbf{Z}\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\mathbf{Z}_2\dot{\mathbf{Z}}_2+\mathbf{Z}^{(3)}-\mathbf{Z}_2^{(3)}\right\}\mathbf{W}'\mathbf{1}\\
& -6\sigma^{-4}\,\mathrm{tr}\left\{\mathbf{W}\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)\dot{\mathbf{Z}}_2+\sigma\mathbf{W}'\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_2\right)\left(\dot{\mathbf{Z}}+3\dot{\mathbf{Z}}_2\right)+2\sigma^2\mathbf{W}''\left(\dot{\mathbf{Z}}^{(2)}-\dot{\mathbf{Z}}_2^{(2)}\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
A_{T2} = {} & -3\sigma^{-6}\mathbf{1}^\top \mathbf{W}\left\{(1/4)\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)\left(\mathbf{Z}-\mathbf{Z}_{p-q}\right)\left(3\dot{\mathbf{Z}}+\dot{\mathbf{Z}}_{p-q}\right)\right.\\
& \left.+\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)\mathbf{Z}_{p-q}\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)+(1/2)\left(\mathbf{Z}-\mathbf{Z}_{p-q}\right)^{(2)}\odot\left(\mathbf{Z}+3\mathbf{Z}_{p-q}\right)\right\}\mathbf{W1}\\
& -6\sigma^{-5}\mathbf{1}^\top \mathbf{W}\left\{\left(\mathbf{Z}-\mathbf{Z}_{p-q}\right)\odot\left(\mathbf{Z}^{(2)}-\mathbf{Z}_{p-q}^{(2)}\right)+\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)\left(\mathbf{Z}\dot{\mathbf{Z}}-\mathbf{Z}_{p-q}\dot{\mathbf{Z}}_{p-q}\right)\right\}\mathbf{W}'\mathbf{1}\\
& +3\sigma^{-4}\,\mathrm{tr}\left\{\mathbf{W}\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)^{(2)}+2\sigma\mathbf{W}'\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)^{(2)}\right\},
\end{aligned}
$$

$$
A_{T3} = (1/4)\sigma^{-6}\mathbf{1}^\top \mathbf{W}\left\{3\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)\left(\mathbf{Z}-\mathbf{Z}_{p-q}\right)\left(\dot{\mathbf{Z}}-\dot{\mathbf{Z}}_{p-q}\right)+2\left(\mathbf{Z}-\mathbf{Z}_{p-q}\right)^{(3)}\right\}\mathbf{W1}.
$$

Although the expressions for the three corrected statistics entails a great deal of algebra, the expressions only involve simple operations on diagonal matrices. Additionally, for type II censoring, i.e., $\mathbf{W}'=0$, the expressions presented in (1) to (3) are simpler. For instance, the Bartlett-correction factor for LR statistic reduces to:

$$
\varepsilon_p = (1/4)\sigma^{-2}\mathrm{tr}\left\{\mathbf{F}_1\dot{\mathbf{Z}}^{(2)}\right\}+(1/12)\sigma^{-6}\mathbf{1}^\top\mathbf{W}\left(2\mathbf{Z}^{(3)}+3\dot{\mathbf{Z}}\mathbf{Z}\dot{\mathbf{Z}}\right)\mathbf{W1}.
$$