IMPROVED LIKELIHOOD INFERENCE FOR THE SHAPE PARAMETER IN WEIBULL REGRESSION

MICHEL FERREIRA DA SILVA

Departamento de Estatística, IME, Universidade de São Paulo Caixa Postal 66281, São Paulo/SP, 05311-970, Brazil email: michel@ime.usp.br

SILVIA L.P. FERRARI[†]

Departamento de Estatística, IME, Universidade de São Paulo Caixa Postal 66281, São Paulo/SP, 05311-970, Brazil email: sferrari@ime.usp.br

FRANCISCO CRIBARI-NETO

Departamento de Estatística, CCEN, Universidade Federal de Pernambuco Cidade Universitária, Recife/PE, 50740-540, Brazil email: cribari@ufpe.br

Abstract. We obtain adjustments to the profile likelihood function in Weibull regression models with and without censoring. Specifically, we consider two different modified profile likelihoods: (i) the one proposed by Cox and Reid (1987), and (ii) an approximation to the one proposed by Barndorff–Nielsen (1983), the approximation having been obtained using the results by Fraser and Reid (1995) and by Fraser et al. (1999). We focus on point estimation and likelihood ratio tests on the shape parameter in the class of Weibull regression models. We derive some distributional properties of the different maximum likelihood estimators and likelihood ratio tests. The numerical evidence presented in the paper favors the approximation to Barndorff–Nielsen's adjustment.

Keywords: Censoring; failure time data; likelihood ratio test; maximum likelihood estimator; nuisance parameter; profile likelihood; regression model; Weibull distribution.

1. INTRODUCTION

When the interest lies in the time to a given event, such as the failure of an electronic component or the death of a patient, oftentimes there are censored observations. Two common censoring forms are when n units are observed until the rth (r < n) failure takes place (type II censoring) and when the experiment ends after a preassigned censoring time (type I censoring). The Weibull distribution is commonly used to model such data, since it generalizes the exponential distribution allowing for a power dependence of the hazard function (i.e., the instant failure rate) on time. In particular, it is possible to obtain

[†] Corresponding author.

increasing, constant and decreasing hazard rates over time depending on the value of the Weibull shape parameter, β ($\beta > 1$, $\beta = 1$ and $\beta < 1$, respectively). The exponential distribution can only be used in situations where it is safe to assume a constant hazard rate, unlike the more flexible Weibull model. For instance, the risk of death to a newborn child is typically high in his/her first few days and quickly decreases thereafter, and remains low and approximately constant for a long time. For some applications of survival and reliability analyses, the reader is referred to Gross and Clark (1975), Lawless (1982), Mann et al. (1974) and Klein and Moeschberger (1997).

In many applications, a regression structure is used to model the dependence of the Weibull scale parameter, α , on covariates, which leads to what is known as the Weibull regression model. More specifically, let t denote the time to failure, and assume that t is Weibull distributed, i.e., $t \sim W(\alpha(\mathbf{x}), \beta)$, where $\mathbf{x} = (x_1, x_2, \dots, x_p)$ is a set of covariates and $W(\alpha, \beta)$ denotes the Weibull distribution with shape parameter β and scale parameter α . It then follows that $\log t = y \sim EV(\eta(\mathbf{x}), \sigma)$, that is, y follows an extreme value distribution with location parameter $\eta(\mathbf{x})$ and scale parameter σ ; here, $\sigma = \beta^{-1}$ and $\eta(\mathbf{x}) = \log \alpha(\mathbf{x})$. We then have that

$$y = \log t = \eta(\mathbf{x}) + \sigma\vartheta, \tag{1.1}$$

where $\vartheta \sim EV(0,1)$. The density function of y is given by

$$p(y; \eta(\mathbf{x}), \sigma) = \frac{1}{\sigma} \exp\left\{\frac{y - \eta(\mathbf{x})}{\sigma} - \exp\left(\frac{y - \eta(\mathbf{x})}{\sigma}\right)\right\}, \quad -\infty < y < \infty.$$

The corresponding distribution function is $F(y; \eta(\mathbf{x}), \sigma) = 1 - S(y; \eta(\mathbf{x}), \sigma)$, where

$$S(y; \eta(\mathbf{x}), \sigma) = \exp\left\{-\exp\left[\frac{y - \eta(\mathbf{x})}{\sigma}\right]\right\}$$

is the survival function. In what follows, we shall assume that $\eta(\mathbf{x}) = \mathbf{x}\phi$, where $\phi = (\phi_1, \phi_2, \dots, \phi_p)^{\top}$. From (1.1), one can easily note that $t = t_{\mathbf{x}} = \exp(\mathbf{x}\phi) \times \exp(\sigma\vartheta) = \exp(\mathbf{x}\phi) \times t_0$, where $t_0 = \exp(\sigma\vartheta)$; note that t_0 is the time to failure under no regression dependence. It is also noteworthy that if $\mathbf{x}\phi < 0$, then $t_{\mathbf{x}}$ is shorter than t_0 , that is, the covariates accelerate the time to failure. The model is then an accelerated time model.

The parameters that index the model can be estimated by maximum likelihood. However, in small samples the resulting estimators can be considerably biased and likelihood ratio tests can be substantially size distorted. Profile likelihood adjustments for the Weibull shape parameter with no covariates were obtained by Ferrari et al. (2005) and Yang and Xie (2003). Their Monte Carlo results suggest that adjusted profile maximum likelihood estimators are less biased and more efficient than the usual maximum likelihood estimator in small samples both with and without censoring. The setup considered by these authors includes only one nuisance parameter; inference on β is expected to be even more unreliable in the regression case, where there are several nuisance parameters. Profile likelihood adjustments are thus even more needed in the regression framework.

Our chief goal in this paper is to obtain two different adjustments to the profile likelihood function when the parameter of interest is β (or, alternatively, σ) in Weibull regression models. Specifically, we consider two different modified profile likelihoods: (i) the one proposed by Cox and Reid (1987), and (ii) an approximation to the one proposed by Barndorff–Nielsen (1983), the approximation having been obtained using the results by Fraser and Reid (1995) and by Fraser et al. (1999). We focus on maximum likelihood estimation and likelihood ratio test on σ . We show that, in the abscence of censoring or under type II censoring, the distribution of the ratio between the estimator of σ and the true parameter σ does not depend on (ϕ, σ) , such a ratio thus being a pivot for this pair. This result is proved for the profile (original) maximum likelihood estimator and for the modified profile maximum likelihood estimators. It is also shown that the size properties of the usual and adjusted profile likelihood ratio tests depend neither on ϕ nor on the value of σ set at the null hypothesis. We generalize some of the results in Ferrari et al. (2005) and Yang and Xie (2003); unlike their results, ours allow for regression dependence. We also report numerical evidence on the finite-sample performance of the different inference strategies; it favors the inference obtained following Fraser and Reid (1995), Fraser et al. (1999).

The remainder of the paper unfolds as follows. Section 2 introduces the profile likelihood function, and Section 3 presents the Barndorff–Nielsen (1980, 1983) adjustment and two alternative adjusted profile likelihoods. Section 4 derives the adjustments for inference on the Weibull shape parameter and Monte Carlo results are presented in Section 5. Finally, Section 6 concludes the paper. Some of the distributional properties of the different estimators and tests are derived in the Appendix.

2. PROFILE LIKELIHOOD

Let $\mathcal{Y} = (y_1, \ldots, y_n)^\top$ be an $n \times 1$ vector of independent random variables, each having a distribution that is indexed by two (possibly vector-valued) parameters, namely μ and ν ; in what follows μ is the parameter of interest and ν is the nuisance parameter. Inference on μ can sometimes be based on a marginal or a conditional likelihood function. However, typically these functions cannot be derived. In such cases, the standard approach is to use the profile likelihood function: $L_p(\mu) = L(\hat{\nu}_{\mu}, \mu)$, where $L(\cdot)$ is the usual likelihood function and $\hat{\nu}_{\mu}$ is the maximum likelihood estimate of ν for a given, fixed μ . For instance, the usual likelihood ratio statistic,

$$LR(\mu) = 2[\ell(\widehat{\nu},\widehat{\mu}) - \ell(\widehat{\nu}_{\mu},\mu)] = 2[\ell_p(\widehat{\mu}) - \ell_p(\mu)],$$

is based on the profile likelihood function. Here, $\hat{\mu}$ and $\hat{\nu}$ are the maximum likelihood estimates of μ and ν respectively, $\ell(\cdot)$ is the log-likelihood function and $\ell_p(\cdot)$ is the profile log-likelihood function. It is important to note, however, that $L_p(\cdot)$ is not a genuine likelihood. In particular, for $\theta = (\nu^{\top}, \mu^{\top})^{\top}$, properties such as

$$E(u(\theta)) = 0$$
 and $E\{u(\theta)u(\theta)^{\top}\} + E\left\{\frac{\partial u(\theta)}{\partial \theta^{\top}}\right\} = 0$

do not hold when $u_p(\mu)$ is used instead of $u(\theta)$. Here, $u(\theta) = \partial \ell(\theta) / \partial \theta$ is the score function and $u_p(\mu) = \partial \ell_p(\mu) / \partial \mu$ is the profile score function. Additionally, the profile score and information biases are only guaranteed to be O(1).

3. MODIFIED PROFILE LIKELIHOODS

Several authors have proposed adjustments to the profile likelihood function; for a review, see Severini (2000, Chapter 9). Barndorff–Nielsen's (1983) modified profile likelihood is obtained as an approximation to a marginal or conditional likelihood for μ , if either exists. His approach uses the p^* formula (Barndorff–Nielsen, 1980, 1983), which yields an approximation to the probability density function of the maximum likelihood estimator conditional on an ancillary statistic. His modified profile likelihood is

$$L_{BN}(\mu) = \left|\frac{\partial \widehat{\nu}_{\mu}}{\partial \widehat{\nu}}\right|^{-1} |j_{\nu\nu}(\widehat{\nu}_{\mu},\mu)|^{-1/2} L_p(\mu),$$

where $j_{\nu\nu}(\nu,\mu) = -\partial^2 \ell / \partial \nu \partial \nu^{\top}$ is the observed information for ν . It is possible to show that the resulting score and information biases are of order $O(n^{-1})$ and that $L_{BN}(\mu)$ is invariant under reparameterization of the form $(\nu,\mu) \to (\tau,\xi)$, where $\tau = \tau(\nu,\mu)$ and $\xi = \xi(\mu)$.

In order to obtain Barndorff–Nielsen's (1983) modified profile likelihood function one needs to obtain $|\partial \hat{\nu}_{\mu} / \partial \hat{\nu}|$. An alternative expression for $L_{BN}(\mu)$ that does not involve $|\partial \hat{\nu}_{\mu} / \partial \hat{\nu}|$ is available. However, it involves a sample space derivative of the log-likelihood function and the specification of ancillary *a* such that $(\hat{\nu}, \hat{\mu}, a)$ is a minimal sufficient statistic. It can be shown that

$$\frac{\partial \widehat{\nu}_{\mu}}{\partial \widehat{\nu}} = j_{\nu\nu}(\widehat{\nu}_{\mu}, \mu; \widehat{\nu}, \widehat{\mu}, a)^{-1} \ell_{\nu; \widehat{\nu}}(\widehat{\nu}_{\mu}, \mu; \widehat{\nu}, \widehat{\mu}, a),$$

where

$$\ell_{\nu;\widehat{\nu}}(\widehat{\nu}_{\mu},\mu;\widehat{\nu},\widehat{\mu},a) = \frac{\partial}{\partial\widehat{\nu}} \left(\frac{\partial\ell(\widehat{\nu}_{\mu},\mu;\widehat{\nu},\widehat{\mu},a)}{\partial\nu} \right)$$

Here, $\ell(\hat{\nu}_{\mu}, \mu; \hat{\nu}, \hat{\mu}, a)$ and $j_{\nu\nu}(\hat{\nu}_{\mu}, \mu; \hat{\nu}, \hat{\mu}, a)$ are the log-likelihood function and the observed information for ν , respectively; they depend on the data only through the minimal sufficient statistic.

An approximation to $\ell_{\nu;\hat{\nu}}(\hat{\nu}_{\mu},\mu;\hat{\nu},\hat{\mu},a)$ can be obtained through an approximately ancillary statistic (Fraser and Reid (1995) and Fraser et al. (1999); see also Severini (2000, p. 216)). The resulting log-likelihood function, $\tilde{\ell}_{BN}$, can be written as

$$\tilde{\ell}_{BN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\hat{\nu}_{\mu}, \mu)| - \log |\ell_{\nu;\mathcal{Y}}(\hat{\nu}_{\mu}, \mu)\hat{V}_{\nu}|, \qquad (3.1)$$

where $\ell_{\nu;\mathcal{Y}}(\nu,\mu) = \partial \ell_{\nu}(\nu,\mu) / \partial \mathcal{Y}^{\top}$ and

$$\widehat{V}_{\nu} = \left(-\frac{\partial F_1(y_1;\widehat{\nu},\widehat{\mu})/\partial\widehat{\nu}}{p_1(y_1;\widehat{\nu},\widehat{\mu})} \quad \cdots \quad -\frac{\partial F_n(y_n;\widehat{\nu},\widehat{\mu})/\partial\widehat{\nu}}{p_n(y_n;\widehat{\nu},\widehat{\mu})} \right)^{\top}$$

 $p_j(y,\nu,\mu)$ being the probability density function of y_j and $F_j(y,\nu,\mu)$ being the cumulative distribution function of y_j . The corresponding estimator shall be denoted as $\hat{\mu}_{BN}$.

We shall now consider an alternative adjustment to the profile likelihood function. Suppose that μ and ν are orthogonal, that is, that the components of the score vector, $\partial \ell / \partial \mu$ and $\partial \ell / \partial \nu$, are uncorrelated. Cox and Reid (1987) proposed an adjustment that is an approximation to a conditional probability density function of the observations given the maximum likelihood estimator of ν . It can be written as

$$L_{CR}(\mu) = |j_{\nu\nu}(\widehat{\nu}_{\mu},\mu)|^{-1/2} L_p(\mu).$$

The Cox and Reid (1987) modified profile log-likelihood function is

$$\ell_{CR}(\mu) = \ell_p(\mu) - \frac{1}{2} \log |j_{\nu\nu}(\hat{\nu}_{\mu}, \mu)|; \qquad (3.2)$$

the maximizer of $\ell_{CR}(\mu)$ shall be denoted as $\hat{\mu}_{CR}$. The score bias is $O(n^{-1})$ but, in general, the information bias remains O(1).

The Cox and Reid (1987) adjustment requires that μ and ν be orthogonal. However, it is not always possible to find a parameterization that delivers orthogonality. Additionally, unlike $L_{BN}(\mu)$, their adjustment is not invariant under reparameterizations of the form $(\nu,\mu) \rightarrow (\tau,\xi)$, where $\tau = \tau(\nu,\mu)$ and $\xi = \xi(\mu)$. Note that if $\hat{\nu}_{\mu} = \hat{\nu}$ for all μ , then $L_{BN}(\mu) = L_{CR}(\mu)$. In this case, μ and ν are orthogonal parameters (Cox and Reid, 1987). It can be shown that $\ell_{BN}(\mu) - \ell_{BN}(\hat{\mu}) = \ell_{CR}(\mu) - \ell_{CR}(\hat{\mu}) + O_p(n^{-1})$. Hence, the likelihood ratio statistics obtained from $\ell_{BN}(\mu)$ and $\ell_{CR}(\mu)$ differ by a term of order $O_p(n^{-1})$.

4. PROFILE LIKELIHOODS FOR THE WEIBULL SHAPE PARAMETER

Let y_1, \ldots, y_n be a set of independent random variables such that $y_j \sim EV(\eta(\mathbf{x}_j), \sigma)$, $j = 1, \ldots, n$, and let C and \overline{C} denote the sets of censored and noncensored observations indices, respectively. The log-likelihood function for (σ, ϕ) is

$$\ell = \ell(\sigma, \phi) = -r \log \sigma + \sum_{j \in \overline{C}} \frac{y_j - \mathbf{x}_j \phi}{\sigma} - \sum_{j=1}^n \exp\left(\frac{y_j - \mathbf{x}_j \phi}{\sigma}\right), \tag{4.1}$$

where r is the observed number of failures, $\mathbf{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jp})$ and $x_{j1} = 1, \forall j$. For complete data, r = n and $\overline{C} = \{1, \ldots, n\}$. For type II censored data, r is fixed and observation ceases after the rth (r < n) failure. For type I censored data, r is random. In this case, one observes $y_j = \min(\log t_j, \log c)$, where $t_j \sim W(\exp(\mathbf{x}_j\phi), \sigma^{-1})$ and c is a preassigned (fixed) censoring time.

Let $z_j = (y_j - \mathbf{x}_j \phi) / \sigma$ and note that $(s, t = 1, \dots, p)$

$$\frac{\partial \ell(\sigma,\phi)}{\partial \phi_s} = -\frac{1}{\sigma} \sum_{j \in \overline{C}} x_{js} + \frac{1}{\sigma} \sum_{j=1}^n x_{js} \exp(z_j) \quad \text{and} \quad \frac{\partial^2 \ell(\sigma,\phi)}{\partial \phi_t \partial \phi_s} = -\frac{1}{\sigma^2} \sum_{j=1}^n x_{js} x_{jt} \exp(z_j).$$

The restricted maximum likelihood estimator of ϕ , for fixed σ (β^{-1}), does not have closed form. It is not possible, thus, to obtain a closed form expression for $\ell_p(\sigma)$. It can be defined as $\ell(\sigma, \phi)$ subject to the restrictions $\partial \ell(\sigma, \phi) / \partial \phi_s = 0$, $s = 1, \ldots, p$.

In what follows we shall obtain the adjusted profile likelihoods described in Section 3. The adjustment proposed by Barndorff–Nielsen (1983) was not considered since it was not possible to obtain $|\partial \hat{\phi}_{\sigma} / \partial \hat{\phi}|$. We shall omit the derivation details in the interest of space.

Since ϕ_{σ} does not have closed form, $\ell_{BN}(\sigma)$ is equal to

$$\ell(\sigma,\phi) + \frac{1}{2} \log \left| j_{\phi\phi}(\sigma,\phi) \right| - \log \left| \ell_{\phi;\mathcal{Y}}(\sigma,\phi) \widehat{V}_{\phi} \right|$$

subject to restrictions of the form $\partial \ell(\sigma, \phi) / \partial \phi_s = 0, s = 1, \dots, p$. We obtained

$$j_{\phi\phi}(\sigma,\phi) = \frac{1}{\sigma^2} X^\top Z X,$$

where $Z = \text{diag}(\exp(z_1), \dots, \exp(z_n)), \ \ell_{\phi;\mathcal{Y}}(\sigma, \phi) = \sigma^{-2} X^{\top} Z$ and $\widehat{V}_{\phi} = (v_1^{\top}, \dots, v_n^{\top})^{\top}$, where $v_j = \mathbf{x}_j$ if $j \in \overline{C}$ and $v_j = 0_{1 \times p}$ if $j \in C$. Here, $0_{m \times n}$ denotes an $m \times n$ matrix of zeros. For noncensored data sets, $\widehat{V}_{\phi} = X$. Therefore, the function $\widetilde{\ell}_{BN}(\sigma)$ is given by

$$\ell(\sigma,\phi) + p\log\sigma + \frac{1}{2}\log\left|X^{\top}ZX\right| - \log\left|X^{\top}Z\widehat{V}_{\phi}\right|$$
(4.2)

for pairs (σ, ϕ) that satisfy the restrictions $\partial \ell(\sigma, \phi) / \partial \phi_s = 0, s = 1, \dots, p$.

The Cox and Reid (1987) adjustment requires parameter orthogonality. In order to obtain $\lambda = (\lambda_1, \ldots, \lambda_p)^{\top}$ orthogonal to the parameter σ , it is necessary to solve a partial differential equation (Cox and Reid, 1987). For complete data, such an equation is

$$i_{\phi\sigma} + i_{\phi\phi} \frac{\partial \phi(\sigma, \lambda)}{\partial \sigma} = 0$$

where $i_{\phi\phi}$ and $i_{\phi\sigma}$ are elements of Fisher's information matrix $I(\sigma, \phi)$ (Lawless, 1982),

$$I(\sigma,\phi) = \frac{1}{\sigma^2} \begin{pmatrix} i_{\phi\phi} & i_{\phi\sigma} \\ \\ i_{\sigma\phi} & i_{\sigma\sigma} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} X^{\top}X & \mathbf{m} \\ \\ \mathbf{m}^{\top} & n[\pi^2/6 + (1-\gamma)^2] \end{pmatrix}$$

Here, **m** is a $p \times 1$ vector with t th element equal to $(1 - \gamma) \sum_{j=1}^{n} x_{jt}$. Thus,

$$\frac{1}{\sigma^2}\mathbf{m} + \frac{1}{\sigma^2}X^{\top}X\frac{\partial\phi(\sigma,\lambda)}{\partial\sigma} = 0$$

or

$$(1 - \gamma)X^{\top} \mathbb{1}_n + X^{\top}X \frac{\partial \phi(\sigma, \lambda)}{\partial \sigma} = 0$$

Hence,

$$\frac{\partial \phi(\sigma, \lambda)}{\partial \sigma} = -(1 - \gamma) (X^{\top} X)^{-1} X^{\top} \mathbb{1}_n$$

and a solution to this equation is

$$\phi = -(1 - \gamma)\sigma(X^{\top}X)^{-1}X^{\top}\mathbb{1}_n + \log\lambda,$$

where γ is Euler's constant, which is approximately equal to 0.5772.

Note that for identically distributed random variables, $X = 1_n$ and, thus,

$$\lambda = \exp(\phi) \exp[(1 - \gamma)\sigma],$$

where ll_n is an *n*-vector of ones. From the Weibull regression model, $\sigma = \beta^{-1}$ and $\log \alpha(\mathbf{x}_j) = \mathbf{x}_j \phi$ (j = 1, ..., n). Therefore,

$$\lambda = \alpha \exp[(1 - \gamma)/\beta].$$

For independent and identically distributed Weibull random variables, Yang and Xie (2003) proposed an orthogonal parameterization (β, λ^*) , where $\lambda^* = \log(\lambda)/\beta$ is a reparemeterization of λ obtained following the proposal in Cox and Reid (1989). They used this parameterization under type II censoring and, for type I censoring, the distribution was indexed by the pair $(\beta, \log \lambda)$. These parameterizations were obtained under noncensoring, thus it is not possible to guarantee parameter orthogonality when observations were recorded under either censoring scheme. The numerical evidences in Yang and Xie (2003) and in Ferrari et al. (2005) on maximum likelihood estimation and likelihood ratio tests based on $\ell_{CR}(\beta)$ showed that the resulting inference was reliable.

In order to evaluate the effectiveness of Cox and Reid's (1987) adjustment to the Weibull profile likelihood function in the regression case, we considered parameterizations that are similar in nature to those of Yang and Xie (2003). Both for complete and censored data, the numerical evidence showed that the parameterization similar to $(\beta, \log \lambda)$ outperformed the parameterization $(\beta, \log(\lambda)/\beta)$. As a result, we shall focus on the former. For the pair (σ, ϕ) that indexes the extreme value distribution, we obtained the following *p*-dimensional parameter vector orthogonal to σ :

$$\lambda^{\dagger} = \log \lambda = \phi + (1 - \gamma)\sigma(X^{\top}X)^{-1}X^{\top}\mathbb{1}_{n}.$$

In order to obtain $\ell_{CR}(\sigma)$, one should index the extreme value distribution by $(\sigma, \lambda^{\dagger})$. It is noteworthy that, unlike Ferrari et al. (2005) and Yang and Xie (2003), we consider the case of vector-valued nuisance parameter.

The log-likelihood function for $(\sigma, \lambda^{\dagger})$ is

$$\ell(\sigma, \lambda^{\dagger}) = -r \log \sigma + \sum_{j \in \overline{C}} \left(\frac{y_j - \mathbf{x}_j \lambda^{\dagger}}{\sigma} \right) - \sum_{j=1}^n \exp(w_j),$$

where, for $j = 1, \ldots, n$,

$$w_j = \frac{y_j - \mathbf{x}_j \lambda^{\dagger}}{\sigma} + (1 - \gamma) \mathbf{x}_j (X^{\top} X)^{-1} X^{\top} \mathbb{1}_n$$

For complete data, r = n and $\overline{C} = \{1, \dots, n\}$.

Similarly to the previous case, $\hat{\lambda}^{\dagger}_{\sigma}$ does not have closed form. Thus, from (3.2) it is possible to show that $\ell_{CR}(\sigma)$ is given by

$$\ell(\sigma, \lambda^{\dagger}) + p \log \sigma - \frac{1}{2} \log \left| X^{\top} W X \right|, \qquad (4.3)$$

where $W = \text{diag}(\exp(w_1), \ldots, \exp(w_n))$, for pairs $(\sigma, \lambda^{\dagger})$ that satisfy $\partial \ell(\sigma, \lambda^{\dagger}) / \partial \lambda_s^{\dagger} = 0$, $s = 1, \ldots, p$. For complete data, note that $\tilde{\ell}_{BN}(\sigma) = \ell_{CR}(\sigma)$.

5. MONTE CARLO RESULTS

This section presents Monte Carlo simulation results on the small sample behavior of estimators and tests on the Weibull shape parameter ($\beta = \sigma^{-1}$), i.e., on the parameter that controls the form of the hazard rate, under both noncensoring and (types I and II) censoring. The numerical exercise was based on two regression models, whose systematic components are given by

$$\eta(\mathbf{x}) = \phi_1 + \phi_2 x_2$$
 and $\eta(\mathbf{x}) = \phi_1 + \phi_2 x_2 + \phi_3 x_3 + \phi_4 x_4.$

(We shall refer to these two models as 'model 1' and 'model 2', respectively.) All simulations were performed using the 0x matrix programming language (Doornik, 2001); they are all based on 10,000 replications.

The sample size is n = 20. Under type II censoring, we considered samples with 10 and 15 failures, and under type I censoring we considered samples with 25% and 50% expected censored data. The values of the nuisance parameters were set equal to one, i.e., $\phi_i = 1 \forall i$. The covariate values were chosen as random draws from a standard uniform distribution.

We considered three different values for β : 0.5, 1.0, 5.0 (decreasing, constant and increasing hazard rates, respectively).

The following measures are presented for all point estimators: mean, variance, bias, mean squared error (MSE), relative bias (RB), skewness and kurtosis. Relative bias is defined as $100 \times (\text{bias} / \text{true parameter value})\%$. For complete data, the results (relative bias, skewness and kurtosis) are the same for the three values of β . A similar result is proved for the case of type II censoring; see the Appendix. Thus, in these two cases, we shall vary the dimension of X but only present results for $\beta = 5.0$ (increasing hazard rate). For type I censored data, we have considered the three values for β .

We also report the null rejection rates of the likelihood ratio tests based on the profile and the modified profile likelihoods. The null hypotheses under test are \mathcal{H}_0 : $\beta = 0.5, 1.0, 5.0$. The test of \mathcal{H}_0 : $\beta = 1.0$ is of particular interest since under the null hypothesis the Weibull distribution reduces to the exponential distribution and the hazard function becomes constant over time. The results for the three null hypotheses are identical without censoring. A similar result is proved for the case of type II censoring; see the Appendix. For type I censored data, we have considered the three null hypotheses.

model 1									
estimator	mean	variance	bias	MSE	RB(%)	skewness	kurtosis		
$\widehat{\beta}$	5.569	1.210	0.569	1.534	11.380	0.918	4.555		
$\widehat{\widetilde{\beta}}_{BN}$	5.217	1.057	0.217	1.104	4.333	0.911	4.538		
model 2									
estimator	mean	variance	bias	MSE	$\operatorname{RB}(\%)$	skewness	kurtosis		
$\hat{\beta}$	6.024	1.667	1.024	2.716	20.484	1.041	5.155		
$\tilde{\beta}_{BN}$	5.251	1.259	0.251	1.322	5.023	1.032	5.118		

Table 5.1. Point estimation, noncensoring, n = 20, $\beta = \sigma^{-1} = 5$.

Table 5.1 contains results for estimation of the shape parameter under noncensoring. We consider the estimators obtained by maximizing the likelihoods given in Section 4: $\hat{\beta}$ (4.1), $\hat{\beta}_{BN}$ (4.2) and $\hat{\beta}_{CR}$ (4.3). In this case, $\hat{\beta}_{BN} = \hat{\beta}_{CR}$. For model 1 (two nuisance parameters), the estimator that displayed the smallest relative absolute bias was $\hat{\beta}_{BN}$ (4.333%). Note that the relative bias of the usual maximum likelihood estimator $\hat{\beta}$ is nearly three times larger than that of $\hat{\beta}_{CR}^*$. It is also noteworthy that $\hat{\beta}_{BN}$ has the smallest mean squared error (1.104). The figures in the table show that the two estimators have nearly the same skewness and kurtosis (0.9 and 4.5, respectively; the corresponding asymptotic values are 0 and 3). For model 2 (four nuisance parameters), the best performing estimator was also $\hat{\beta}_{BN}$, whose relative bias and mean squared error were 5.023% and 1.322, respectively. It

is noteworthy that as the number of nuisance parameters increased, so did the relative absolute biases of the different estimators. The estimator obtained from modified profile likelihood performed considerably better than the usual maximum likelihood estimator.

Table 5.2 gives the null rejection rates of the likelihood ratio test and of the corresponding test from the adjusted likelihood. More precisely, the tests are based on ℓ (4.1) and $\tilde{\ell}_{BN}$ (4.2). In this case, $\tilde{\ell}_{BN} = \ell_{CR}$. Here $\mathcal{H}_0 : \beta = 5$; other null hypotheses are not tested since the rejection rates do not depend on the value of β under test, as shown in the Appendix. For instance, for model 2 and at the 10% nominal level, the null rejection rates of the tests based on ℓ and $\tilde{\ell}_{BN}$ were, respectively, equal to 22.300% and 10.410%. The modified profile likelihood yielded more reliable inference than the usual profile likelihood function. Unlike the test based on ℓ , its size performance did not deteriorate when the number of nuisance parameters increased from two to four.

	mod	lel 1	mod	el 2
nominal level	l	$\tilde{\ell}_{BN}$	l	$\tilde{\ell}_{BN}$
10%	14.170	10.420	22.300	10.410
5%	7.910	5.210	14.280	5.350
1%	2.050	0.880	4.850	1.090
0.1%	0.250	0.070	0.980	0.100

Table 5.2. Null rejection rates, noncensoring, n = 20.

Figure 5.1. Relative quantile discrepancies plot, model 2, noncensoring, n = 20.



Figure 5.1 plots the relative quantile discrepancies versus the corresponding asymptotic quantiles. Relative quantile discrepancy is defined as the difference between exact (estimated by simulation) and asymptotic quantiles divided by the latter. The closer to zero the relative quantile discrepancy, the better the approximation of the exact null distribution of the test statistic by the limiting χ_1^2 distribution. The relative discrepancy curve based on $\tilde{\ell}_{BN}$ (BN) is closer to a horizontal line at zero. It is also clear that the null distribution of the usual likelihood ratio statistic is poorly approximated by the limiting distribution.

Consider now type II censoring with n = 20 and r = 10 (number of failures). The results are given in Table 5.3. As expected, under censoring the small sample behavior of the estimators worsened. The estimator with smallest absolute bias was $\hat{\beta}_{CR}$. For instance, for model 2 the relative absolute biases of $\hat{\beta}$, $\hat{\beta}_{BN}$ and $\hat{\beta}_{CR}$ were 40.295%, 4.923% and 2.543%, respectively. The relative bias of $\hat{\beta}_{CR}$ was nearly two times smaller than that of $\hat{\beta}_{BN}$ and about sixteen times smaller than that of the usual maximum likelihood estimator. $\hat{\beta}_{CR}$ displayed the smallest mean squared error in model 1; for model 2, the smallest mean squared error was that of $\hat{\beta}_{BN}$. For model 1 (two nuisance parameters), the mean squared errors of $\hat{\beta}$, $\hat{\beta}_{BN}$ and $\hat{\beta}_{CR}$ were, respectively, 5.311, 3.148 and 3.128; the corresponding figures for model 2 were 10.313, 3.443 and 3.509.

model 1									
estimator	mean	variance	bias	MSE	$\operatorname{RB}(\%)$	skewness	kurtosis		
$\widehat{\beta}$	6.196	3.880	1.196	5.311	23.920	1.562	8.901		
${\tilde eta}_{BN}$	5.445	2.949	0.445	3.148	8.910	1.519	8.473		
\widehat{eta}_{CR}	5.354	3.002	0.354	3.128	7.085	1.558	8.802		
model 2									
estimator	mean	variance	bias	MSE	RB(%)	skewness	kurtosis		
$\widehat{\beta}$	7.015	6.253	2.015	10.313	40.295	1.778	9.794		
${\tilde eta}_{BN}$	5.246	3.383	0.246	3.443	4.923	1.708	9.246		
$\widehat{\beta}_{CR}$	5.127	3.493	0.127	3.509	2.543	1.672	8.814		

Table 5.3. Point estimation, type II censoring, $\beta = \sigma^{-1} = 5$, (n, r) = (20, 10).

We have also considered type II censoring with (n, r) = (20, 15). For brevity, the results are not shown. For model 1, the relative absolute biases of $\hat{\beta}_{CR}$ and $\hat{\beta}_{BN}$ were 5.711% and 5.786%, respectively, the corresponding figure for the usual maximum likelihood estimator being 15.121%. For model 2, these relative absolute biases were 4.566%, 4.933% and 25.971%, respectively.

Overall, the results indicate that the usual maximum likelihood estimator can display poor small sample behavior, that the $\hat{\beta}_{CR}$ was the overall winner, followed by approximate

Barndorff–Nielsen estimator.

Results on hypothesis testing derived from ℓ (4.1), $\tilde{\ell}_{BN}$ (4.2) and ℓ_{CR} (4.3) are given in Table 5.4. Here (n, r) = (20, 10). For both models, the test based on $\tilde{\ell}_{BN}$ displayed the best size behavior. For instance, for model 2 and at the 1% nominal level, the corresponding null rejection rates of the tests based on ℓ , $\tilde{\ell}_{BN}$ and ℓ_{CR} were 7.350%, 0.950% and 1.430%. It is noteworthy that the finite sample behavior of the test based on ℓ deteriorate as the number of nuisance parameters increase.

		model 1			model 2	
		$\mathcal{H}_0:\beta=1$	L		$\mathcal{H}_0: \beta = 5$	5
nominal level	ℓ	$\tilde{\ell}_{BN}$	ℓ_{CR}	ℓ	$\tilde{\ell}_{BN}$	ℓ_{CR}
10%	14.600	9.220	10.040	27.760	9.420	11.460
5%	8.050	4.610	5.010	18.810	4.470	5.990
1%	2.250	0.930	1.050	7.350	0.950	1.430
0.1%	0.400	0.130	0.140	1.840	0.090	0.160

Table 5.4. Null rejection rates, type II censoring, (n, r) = (20, 10).

The relative quantile discrepancy plot under type II censoring and for model 2 is presented in Figure 5.2. Visual inspection of this figure shows that the best performing test is that based on $\tilde{\ell}_{BN}$ (BN), followed by test based on ℓ_{CR} (CR), and that the test based on ℓ (original) has the poorest small sample performance.

Figure 5.2. Relative quantile discrepancies plot, \mathcal{H}_0 : $\beta = \sigma^{-1} = 5$, model 2, type II censoring, (n, r) = (20, 10).



We shall now move to type I censoring. The censoring time was set at

$$c = \left[(1/n) \sum_{j=1}^{n} \alpha(\mathbf{x}_j) \right] (-\log p)^{1/\beta}$$

which yielded the desired probability of censoring p. Here, p = 0.25, 0.50. Tables 5.5 and 5.6 present results on estimation and testing for model 2 with p = 0.25. The likelihood ratio tests are derived from ℓ (4.1), $\tilde{\ell}_{BN}$ (4.2) and ℓ_{CR} (4.3).

The results in Table 5.5 show that, for $\beta = 0.5$ and $\beta = 1.0$, the estimator $\hat{\beta}_{BN}$ outperformed all others in terms of relative bias, whereas for $\beta = 5.0$, $\hat{\beta}_{CR}$ was the best performing estimator. For instance, for $\beta = 0.5$, the relative absolute biases of $\hat{\beta}$, $\hat{\beta}_{BN}$ and $\hat{\beta}_{CR}$ were, respectively, equal to 17.906%, 0.882% and 4.971%. That is, the relative absolute bias of $\hat{\beta}$ was nearly twenty times larger than that of $\hat{\beta}_{BN}$. The modified profile maximum likelihood estimators outperformed the usual unmodified estimator as far as mean squared error is concerned.

parameter	estimator	mean	variance	bias	MSE	$\operatorname{RB}(\%)$	skewness	kurtosis
	\widehat{eta}	0.590	0.022	0.090	0.030	17.906	1.084	5.441
$\beta = 0.5$	$\widehat{\widetilde{\beta}}_{BN}$	0.504	0.017	0.004	0.017	0.882	1.034	5.204
	\widehat{eta}_{CR}	0.475	0.016	-0.025	0.017	-4.971	1.024	5.201
	\widehat{eta}	1.189	0.091	0.189	0.127	18.851	1.114	5.592
$\beta = 1.0$	$\widehat{\widetilde{\beta}}_{BN}$	1.012	0.067	0.012	0.067	1.240	1.071	5.392
	\widehat{eta}_{CR}	0.963	0.066	-0.037	0.067	-3.696	1.070	5.420
	\widehat{eta}	6.259	2.895	1.259	4.479	25.172	1.299	6.165
$\beta = 5.0$	$\widehat{\widetilde{\beta}}_{BN}$	5.147	1.919	0.147	1.940	2.945	1.243	6.012
	\widehat{eta}_{CR}	5.078	1.952	0.078	1.958	1.553	1.217	5.836

Table 5.5. Point estimation, type I censoring with p = 0.25, n = 20, model 2.

The figures in Table 5.6 show that the likelihood ratio test based on ℓ_{BN} displayed the smallest size distortions, followed by the test based on ℓ_{CR} . The usual profile likelihood ratio test was considerably oversized. For example, for the test of \mathcal{H}_0 : $\beta = 1.0$ in model 2 and at the 5% nominal level, the null rejection rates of the tests based on ℓ , ℓ_{BN} and ℓ_{CR} were equal to 10.710%, 4.920% and 6.790%, respectively. For a higher censoring level (p = 0.50), the test based on ℓ_{BN} was slightly conservative and displayed the smallest size distortions (model 1 and model 2), and, except for the test of \mathcal{H}_0 : $\beta = 5.0$, the usual likelihood ratio test outperformed the Cox and Reid test. (Results not shown for brevity.)

The relative quantile discrepancy plot under type I censoring corresponding to \mathcal{H}_0 : $\beta = 1, p = 0.25, n = 20$ and $\eta(\mathbf{x}) = \phi_1 + \phi_2 x_2 + \phi_3 x_3 + \phi_4 x_4 \pmod{2}$ is presented in Figure 5.3. Visual inspection of this figure shows that the best performing test is the one based on $\tilde{\ell}_{BN}$ (BN), followed by the Cox and Reid test (CR). We have also considered type I censoring with p = 0.50. For brevity, the figure is not shown. The Cox and Reid test was now outperformed by the usual likelihood ratio test; the test based on $\tilde{\ell}_{BN}$ remained reliable.

\mathcal{H}_0 :	nominal level	ℓ	$\tilde{\ell}_{BN}$	ℓ_{CR}
	10%	16.750	9.250	13.110
<i>२</i> ० म	5%	9.810	4.590	7.150
$\beta \equiv 0.5$	1%	2.830	0.960	1.740
	0.1%	0.570	0.100	0.140
	10%	17.490	9.570	12.560
9 10	5%	10.710	4.920	6.790
$\rho = 1.0$	1%	3.180	0.900	1.590
	0.1%	0.630	0.070	0.130
	10%	21.750	9.950	11.130
<i>२</i> ह0	5%	14.070	5.010	5.630
$\rho \equiv 5.0$	1%	5.110	1.020	1.280
	0.1%	0.940	0.100	0.110

Table 5.6. Null rejection rates, type I censoring with p = 0.25, n = 20, model 2.

Figure 5.3. Relative quantile plot, \mathcal{H}_0 : $\beta = 1$, type I censoring with p = 0.25, n = 20, model 2.



6. CONCLUDING REMARKS

The maximum likelihood estimator of the shape parameter in Weibull regression models can be considerably biased in small samples. In this paper, we have obtained two modified profile likelihood functions whose associated maximum likelihood estimators are expected to deliver more reliable inference in finite samples. They followed from proposals made by Cox and Reid (1987) and an approximated form of Barndorff–Nielsen's (1983) modified profile likelihood obtained from Fraser and Reid (1995) and Fraser et al. (1999). The numerical evidence presented in the previous section showed that the usual profile likelihood inference can be quite unreliable in small samples, for both point estimation and likelihood ratio testing. In particular, such tests can be extremely liberal. The Cox and Reid adjustment yielded encouraging results, for both point estimation and likelihood ratio testing. However, for critical levels of type I censoring, the usual likelihood ratio test outperformed the Cox and Reid test in some cases. This may be due to the fact that the adjusted test is based on an orthogonal parameterization obtained under independent random variates and carried to the situation where there is censoring. It is interesting to note that in the nonregression case, where there is only one nuisance parameter, this adjusted profile likelihood delivers reliable inference, as shown by Ferrari et al. (2005). However, in Weibull regression models the Cox and Reid adjustment should be used carefully, since it may deliver inference that is even less reliable than that obtained from the usual profile likelihood function. The modified profile likelihood function obtained as the approximation to Barndorff–Nielsen's (1983) function, on the other hand, delivered estimators that were nearly unbiased and associated tests with virtually no size distortions. Overall, this proved to be the best inference strategy. An advantage of this approach is that it does not require the interest and nuisance parameters to be orthogonal, nor does it require the specification of an ancillary statistic.

APPENDIX

As before, $\mathcal{Y} = (y_1, \dots, y_n)^{\top}$ is a vector of independent random variables such that $y_j \sim EV(\mathbf{x}_j \phi, \sigma)$ $(j = 1, \dots, n)$. Note that

$$z_j = \frac{y_j - \mathbf{x}_j \phi}{\sigma} \sim EV(0, 1),$$

 $j = 1, \ldots, n$, and that z_1, \ldots, z_n are i.i.d.. Let $\mathcal{Z} = (z_1, \ldots, z_n)^\top$.

Type II Censoring

Here, we shall consider type II censoring, r denoting the preassigned number of failures at which observation ceases; r does not depend on the pair (σ, ϕ) . (Recall that under noncensuring, r = n.) Let $y_{(1)}, \ldots, y_{(r)}$

and $z_{(1)}, \ldots, z_{(r)}$ are the first r order statistics of \mathcal{Y} and \mathcal{Z} , respectively.

RESULT 1: $(\hat{\phi} - \phi)/\sigma$, $\hat{\sigma}/\sigma$, $(y_j - \mathbf{x}_j \hat{\phi})/\hat{\sigma}$, $(\hat{\phi}_{\sigma} - \phi)/\sigma$ and $(y_j - \mathbf{x}_j \hat{\phi}_{\sigma})/\sigma$ are distributed independently of the pair (σ, ϕ) .

Proof:

The likelihood function for the Weibull regression model is given by

$$L_{\mathcal{Y}}(\sigma,\phi) = \prod_{j\in\overline{C}} \frac{1}{\sigma} \exp\left[\frac{y_j - \mathbf{x}_j\phi}{\sigma} - \exp\left(\frac{y_j - \mathbf{x}_j\phi}{\sigma}\right)\right] \prod_{j\in C} \exp\left[-\exp\left(\frac{y_{(r)} - \mathbf{x}_j\phi}{\sigma}\right)\right].$$

Let

$$y'_j = dy_j + \mathbf{x}_j k, \quad \phi' = d\phi + k \quad \text{and} \quad \sigma' = d\sigma$$

 $j = 1, \ldots, n$, where d is a positive scalar and k is a p-vector. Let $\mathcal{Y}' = (y'_1, \ldots, y'_n)^{\top}$. Note that

$$L_{\mathcal{Y}}(\sigma,\phi) = d^r L_{\mathcal{Y}'}(\sigma',\phi').$$

Hence, if $L_{\mathcal{Y}}(\sigma, \phi)$ is maximized at

$$\widehat{\phi} = \widehat{\phi}(y_{(1)}, \dots, y_{(r)}) \text{ and } \widehat{\sigma} = \widehat{\sigma}(y_{(1)}, \dots, y_{(r)}),$$

then $L_{\mathcal{Y}'}(\sigma, \phi)$ is maximized at

$$\widehat{\phi}(y'_{(1)},\ldots,y'_{(r)}) = d\widehat{\phi} + k \quad \text{and} \quad \widehat{\sigma}(y'_{(1)},\ldots,y'_{(r)}) = d\widehat{\sigma}.$$
(A0)

Note that $\mathcal{Z} = \mathcal{Y}'$ if we let $d = 1/\sigma$ and $k = -(1/\sigma)\phi$. Thus,

$$\widehat{\phi}(z_{(1)},\ldots,z_{(r)}) = \frac{\widehat{\phi}-\phi}{\sigma} \quad \text{and} \quad \widehat{\sigma}(z_{(1)},\ldots,z_{(r)}) = \frac{\widehat{\sigma}}{\sigma}.$$
 (A1)

Hence, $(\hat{\phi} - \phi)/\sigma$ and $\hat{\sigma}/\sigma$ are pivots for (σ, ϕ) , since the joint distribution of $(z_{(1)}, \ldots, z_{(r)})$ does not depend on the pair of unknown parameters. (Note that the distribution of $\hat{\beta}/\beta$ does not depend on the unknown parameters either.)

Note that

$$\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}} = \frac{1}{\widehat{\sigma}(z_{(1)}, \dots, z_{(r)})} \left[z_j - \mathbf{x}_j \widehat{\phi}(z_{(1)}, \dots, z_{(r)}) \right].$$
(A2)

Similarly to the result in (A0), assuming that σ is known,

$$\widehat{\phi}(y'_{(1)},\ldots,y'_{(r)};\sigma')=d\widehat{\phi}(y_{(1)},\ldots,y_{(r)};\sigma)+k,$$

where $\widehat{\phi}(y'_{(1)}, \ldots, y'_{(r)}; \sigma')$ and $\widehat{\phi}(y_{(1)}, \ldots, y_{(r)}; \sigma)$ are the maximizers of $L_{\mathcal{Y}'}(\sigma, \phi) = L_{\mathcal{Y}'}(\phi)$ and $L_{\mathcal{Y}}(\sigma, \phi) = L_{\mathcal{Y}}(\phi)$, respectively. Again, note that $\mathcal{Z} = \mathcal{Y}'$ if we let $d = 1/\sigma$ and $k = -(1/\sigma)\phi$. Hence,

$$\widehat{\phi}(z_{(1)},\ldots,z_{(r)};1) = \frac{\widehat{\phi}(y_{(1)},\ldots,y_{(r)};\sigma) - \phi}{\sigma}$$

Therefore, $(\widehat{\phi}_{\sigma} - \phi)/\sigma$ is a pivot for (σ, ϕ) , where $\widehat{\phi}_{\sigma} = \widehat{\phi}(y_{(1)}, \dots, y_{(r)}; \sigma)$.

Notice that

$$\frac{y_j - \mathbf{x}_j \widehat{\phi}_{\sigma}}{\sigma} = z_j - \mathbf{x}_j \widehat{\phi}(z_{(1)}, \dots, z_{(r)}), \tag{A3}$$

where $\hat{\phi}(z_{(1)}, \dots, z_{(r)}) = \hat{\phi}(z_{(1)}, \dots, z_{(r)}; 1).$

RESULT 2: The distribution of the ratio $\hat{\sigma}/\sigma$ depends on the design matrix X.

$\underline{\text{Proof}}$:

The maximum likelihood estimators $\widehat{\sigma}$ and $\widehat{\phi}$ are the roots of the system

$$-r - \frac{1}{\widehat{\sigma}} \sum_{j \in \overline{C}} \left(y_j - \mathbf{x}_j \widehat{\phi} \right) + \frac{1}{\widehat{\sigma}} \sum_{j=1}^n \left(y_j - \mathbf{x}_j \widehat{\phi} \right) \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}}\right) = 0$$
(A4)

and

$$-\frac{1}{\widehat{\sigma}}\sum_{j\in\overline{C}}\mathbf{x}_j + \frac{1}{\widehat{\sigma}}\sum_{j=1}^n \mathbf{x}_j \exp\left(\frac{y_j - \mathbf{x}_j\widehat{\phi}}{\widehat{\sigma}}\right) = 0_{1\times p}.$$
 (A5)

If we right multiply (A5) by the vector $\hat{\phi}$ and add the result to (A4), we obtain

$$-r - \frac{1}{\widehat{\sigma}} \sum_{j \in \overline{C}} y_j + \frac{1}{\widehat{\sigma}} \sum_{j=1}^n y_j \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}}\right) = 0,$$

which can be written as

$$-r - \frac{1}{\widehat{\sigma}} \sum_{j \in \overline{C}} y_j + \frac{1}{\widehat{\sigma}} \sum_{j=1}^n \left(\mathbf{x}_j \phi + \sigma z_j \right) \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}}\right) = 0.$$
(A6)

Now, if we right multiply (A5) by the vector ϕ , we obtain

$$\frac{1}{\widehat{\sigma}}\sum_{j\in\overline{C}}\mathbf{x}_{j}\phi = \frac{1}{\widehat{\sigma}}\sum_{j=1}^{n}\mathbf{x}_{j}\phi\exp\Big(\frac{y_{j}-\mathbf{x}_{j}\widehat{\phi}}{\widehat{\sigma}}\Big).$$

Thus we can now rewrite (A6) as

$$-r - \frac{1}{\widehat{\sigma}/\sigma} \sum_{j \in \overline{C}} z_j + \frac{1}{\widehat{\sigma}/\sigma} \sum_{j=1}^n z_j \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}}\right) = 0$$

or as

$$-r - \frac{1}{\widehat{\sigma}/\sigma} \sum_{j \in \overline{C}} z_j + \frac{1}{\widehat{\sigma}/\sigma} \sum_{j=1}^n z_j \exp\left\{\frac{1}{\widehat{\sigma}(z_{(1)}, \dots, z_{(r)})} \left[z_j - \mathbf{x}_j \widehat{\phi}(z_{(1)}, \dots, z_{(r)})\right]\right\} = 0.$$

If we let $\hat{\sigma}/\sigma$ be the unknown of the above equation, then its root will only depend on $z_{(1)}, \ldots, z_{(r)}$ and X.

RESULT 3: For uncensored and type II censored data, $\Pr_{(\sigma,\phi)}(LR(\sigma) \leq q)$ depends neither on ϕ nor on the value of σ set at the null hypothesis, it only depends on X.

-

Proof:

Note that from (A5) we have that

$$\sum_{j=1}^{n} \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}}{\widehat{\sigma}}\right) = r.$$

(Recall that $x_{j1} = 1, j = 1, ..., n$.) Analogously, for σ fixed, it is possible to show that

$$\sum_{j=1}^{n} \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}_{\sigma}}{\sigma}\right) = r.$$

The likelihood ratio statistic is defined as

$$LR(\sigma) = 2\bigg\{\ell(\widehat{\sigma},\widehat{\phi}) - \ell(\sigma,\widehat{\phi}_{\sigma})\bigg\}.$$

Thus,

$$LR(\sigma) = 2\left\{-r\log\left(\frac{\widehat{\sigma}}{\sigma}\right) + \sum_{j\in\overline{C}}\left(\frac{y_j - \mathbf{x}_j\widehat{\phi}}{\widehat{\sigma}}\right) - \sum_{j\in\overline{C}}\left(\frac{y_j - \mathbf{x}_j\widehat{\phi}_{\sigma}}{\sigma}\right)\right\}.$$

Using (A1), (A2) and (A3), we obtain

$$LR(\sigma) = 2\left\{-r\log\widehat{\sigma}(z_{(1)},\ldots,z_{(r)}) + \left[\frac{1}{\widehat{\sigma}(z_{(1)},\ldots,z_{(r)})} - 1\right]\sum_{j\in\overline{C}}\left[z_j - \mathbf{x}_j\widehat{\phi}(z_{(1)},\ldots,z_{(r)})\right]\right\}$$

Therefore, for uncensored and type II censored data, $\Pr_{(\sigma,\phi)}(LR(\sigma) \leq q)$ depends neither on ϕ nor on the value of σ set at the null hypothesis. Since $LR(\sigma)$ depends on \mathbf{x}_j , its null distribution depends on X.

RESULT 4: $(\hat{\phi}_{BN} - \phi)/\sigma, \hat{\sigma}_{BN}/\sigma, (\hat{\phi}_{BN} - \phi)/\hat{\sigma}_{BN} \text{ and } (y_j - \mathbf{x}_j \hat{\phi}_{BN})/\hat{\sigma}_{BN} \text{ are distributed independently of the pair } (\sigma, \phi).$

Proof:

Consider the adjusted profile log-likelihood function $\tilde{\ell}_{BN}(\sigma)$ given in (4.2), i.e.,

$$\tilde{\ell}_{BN, \mathcal{Y}}(\sigma, \phi) = -(r-p)\log\sigma + \sum_{j\in\overline{C}} \left(\frac{y_j - \mathbf{x}_j\phi}{\sigma}\right) - \sum_{j=1}^n \exp\left(\frac{y_j - \mathbf{x}_j\phi}{\sigma}\right) + \log\left(\frac{\left|X^\top Z X\right|^{1/2}}{\left|X^\top Z \widehat{V}_{\phi}\right|}\right),$$

subject to restrictions of the form $\partial \ell_{\mathcal{Y}}(\sigma, \phi) / \partial \phi_s = -(1/\sigma) \sum_{j \in \overline{C}} x_{js} + (1/\sigma) \sum_{j=1}^n x_{js} \exp\left[(y_j - \mathbf{x}_j \phi) / \sigma\right] = 0$, where $s = 1, \ldots, p$.

Let

 $y'_j = dy_j + \mathbf{x}_j k, \quad \phi' = d\phi + k \text{ and } \sigma' = d\sigma,$

 $j = 1, \ldots, n$, where d is a positive scalar and k is a p-vector. Let $\mathcal{Y}' = (y'_1, \ldots, y'_n)^{\top}$. Note that

$$\tilde{\ell}_{BN, \mathcal{Y}}(\sigma, \phi) = (r - p) \log d + \tilde{\ell}_{BN, \mathcal{Y}'}(\sigma', \phi') \quad \Longleftrightarrow \quad \tilde{L}_{BN, \mathcal{Y}}(\sigma, \phi) = d^{r - p} \tilde{L}_{BN, \mathcal{Y}'}(\sigma', \phi').$$

Notice that $\partial \ell_{\mathcal{Y}}(\sigma, \phi) / \partial \phi_s = 0 \iff \partial \ell_{\mathcal{Y}'}(\sigma', \phi') / \partial \phi'_s = 0$, where $s = 1, \ldots, p$.

Let $\hat{\phi}_{BN}$ and $\hat{\sigma}_{BN}$ be, respectively, the maximum likelihood estimators of ϕ and σ obtained from the maximization of $\tilde{\ell}_{BN, \mathcal{Y}}(\sigma, \phi)$ subject to the *p* restrictions given above. If $\tilde{\ell}_{BN, \mathcal{Y}}(\sigma, \phi)$ is maximized at

$$\hat{\tilde{\phi}}_{BN} = \hat{\tilde{\phi}}_{BN}(y_{(1)}, \dots, y_{(r)}) \text{ and } \hat{\tilde{\sigma}}_{BN} = \hat{\tilde{\sigma}}_{BN}(y_{(1)}, \dots, y_{(r)})$$

then $\tilde{\ell}_{BN, \mathcal{Y}'}(\sigma, \phi)$ is maximized at

$$\hat{\phi}_{BN}(y'_{(1)}, \dots, y'_{(r)}) = d\hat{\phi}_{BN}(y_{(1)}, \dots, y_{(r)}) + k \quad \text{and} \quad \hat{\sigma}_{BN}(y'_{(1)}, \dots, y'_{(r)}) = d\hat{\sigma}_{BN}(y_{(1)}, \dots, y_{(r)}).$$

Letting $d = 1/\sigma$ and $k = -(1/\sigma)\phi$, similarly to what has been done in the proof of Result 1, it can be shown that

$$\frac{1}{\sigma} \Big[\widehat{\tilde{\phi}}_{BN} - \phi \Big], \quad \frac{\widehat{\tilde{\sigma}}_{BN}}{\sigma}, \quad \frac{1}{\widehat{\tilde{\sigma}}_{BN}} \Big[\widehat{\tilde{\phi}}_{BN} - \phi \Big] \quad \text{and} \quad \frac{y_j - \mathbf{x}_j \widetilde{\phi}_{BN}}{\widehat{\tilde{\sigma}}_{BN}}$$

are distributed independently of the pair (σ, ϕ) .

RESULT 5: The distribution of the ratio $\hat{\sigma}_{BN}/\sigma$ depends on the design matrix X.

$\underline{\text{Proof}}$:

From $\partial \tilde{\ell}_{BN}(\hat{\tilde{\sigma}}_{BN},\hat{\tilde{\phi}}_{BN})/\partial \sigma = 0$ and the restriction $\partial \ell(\hat{\tilde{\sigma}}_{BN},\hat{\tilde{\phi}}_{BN})/\partial \phi^{\top} = 0_{1 \times p}$, it follows that

$$-r + p - \frac{1}{\hat{\sigma}_{BN}} \sum_{j \in \overline{C}} y_j + \frac{1}{\hat{\sigma}_{BN}} \sum_{j=1}^n y_j \exp\left(\frac{y_j - \mathbf{x}_j \hat{\phi}_{BN}}{\hat{\sigma}_{BN}}\right)$$
$$-\frac{1}{2} \operatorname{tr} \left\{ \left[X^\top Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) X \right]^{-1} X^\top \left(Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) \log Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) \right) X \right\}$$
$$+ \operatorname{tr} \left\{ \left[X^\top Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) \hat{V}_{\phi} \right]^{-1} X^\top \left(Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) \log Z(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) \right) \hat{V}_{\phi} \right\} = 0,$$
(A7)

where:

i) $Z(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN})$ is the matrix $Z = \text{diag}(\exp(z_1), \dots, \exp(z_n))$ evaluated at $(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN}),$

ii)
$$\log Z(\hat{\sigma}_{BN}, \tilde{\phi}_{BN})$$
 is the matrix $\log Z = \operatorname{diag}(z_1, \ldots, z_n)$ evaluated at $(\hat{\sigma}_{BN}, \tilde{\phi}_{BN})$, and

iii) \hat{V}_{ϕ} is an $n \times p$ matrix formed out of rows of X.

X

In order to obtain the above result, we used the following property of matrix differentiation:

$$\frac{\partial}{\partial \theta} \log |A(\theta)| = \operatorname{tr} \Big\{ A(\theta)^{-1} \frac{\partial}{\partial \theta} A(\theta) \Big\},\,$$

where $A(\theta)$ is a square matrix whose elements depend on θ ; see Magnus and Neudecker (1988). Note that the matrices

$$\left[X^{\top} Z(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN}) X\right]^{-1} X^{\top} \left(Z(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN}) \log Z(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN})\right) X$$

and

$${}^{\top}Z(\widehat{\tilde{\sigma}}_{BN},\widehat{\tilde{\phi}}_{BN})\widehat{V}_{\phi}\Big]^{-1}X^{\top}\Big(Z(\widehat{\tilde{\sigma}}_{BN},\widehat{\tilde{\phi}}_{BN})\log Z(\widehat{\tilde{\sigma}}_{BN},\widehat{\tilde{\phi}}_{BN})\Big)\widehat{V}_{\phi}$$

are an ancillary statistic for (σ, ϕ) and the distributions of their traces depend on the design matrix X.

Using the restriction $\partial \ell(\hat{\sigma}_{BN}, \hat{\phi}_{BN}) / \partial \phi^{\top} = 0_{1 \times p}$ together with the fact that $y_j = \mathbf{x}_j \phi + \sigma z_j$ $(j = 1, \ldots, n)$, equation (A7) becomes

$$\begin{split} &-r+p-\frac{1}{\hat{\sigma}_{BN}/\sigma}\sum_{j\in\overline{C}}z_{j}+\frac{1}{\hat{\sigma}_{BN}/\sigma}\sum_{j=1}^{n}z_{j}\exp\left[\frac{z_{j}}{\hat{\sigma}_{BN}/\sigma}-\mathbf{x}_{j}\frac{1}{\hat{\sigma}_{BN}}(\hat{\phi}_{BN}-\phi)\right]\\ &-\frac{1}{2}\mathrm{tr}\Big\{\left[X^{\top}Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})X\right]^{-1}X^{\top}\Big(Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})\log Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})\Big)X\Big\}\\ &+\mathrm{tr}\Big\{\left[X^{\top}Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})\hat{V}_{\phi}\right]^{-1}X^{\top}\Big(Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})\log Z(\hat{\sigma}_{BN},\hat{\phi}_{BN})\Big)\hat{V}_{\phi}\Big\}=0.\end{split}$$

Let $\hat{\sigma}_{BN}/\sigma$ be the unknown of this equation, and note that the root depends only on X and on ancillary statistics and pivots for (σ, ϕ) . For complete data, $r = \#\overline{C} = n$.

RESULT 6: For uncensored and type II censored data, $\Pr_{(\sigma,\phi)}(LR_{\tilde{BN}}(\sigma) \leq q)$ depends neither on ϕ nor on the value of σ set at the null hypothesis, it only depends on X.

Proof:

The likelihood ratio statistic based on $\tilde{\ell}_{BN}(\sigma)$ is defined as

$$LR_{\tilde{BN}}(\sigma) = 2 \bigg\{ \tilde{\ell}_{BN}(\hat{\tilde{\sigma}}_{BN}, \hat{\tilde{\phi}}_{BN}) - \tilde{\ell}_{BN}(\sigma, \hat{\tilde{\phi}}_{BN,\sigma}) \bigg\},\$$

where $\hat{\phi}_{BN,\sigma}$ is the maximum likelihood estimator of ϕ under the null hypothesis. If we fix σ in $\tilde{\ell}_{BN}(\sigma, \phi)$, such an estimator is obtained by maximizing this function subject to the restrictions $\partial \ell(\sigma, \phi)/\partial \phi_s = 0$, where $s = 1, \ldots, p$. Note that $\hat{\phi}_{BN,\sigma}$ is different from $\hat{\phi}_{BN}$, since the latter is obtained from the maximization of $\tilde{\ell}_{BN}(\sigma, \phi)$ under the *p* restrictions corresponding to the pair (σ, ϕ) .

The test statistic is given by

$$\begin{split} LR_{\tilde{BN}}(\sigma) &= 2 \Bigg\{ -(r-p) \log\left(\frac{\widehat{\sigma}_{BN}}{\sigma}\right) + \sum_{j \in \overline{C}} \left(\frac{y_j - \mathbf{x}_j \widehat{\phi}_{BN}}{\widehat{\sigma}_{BN}}\right) - \sum_{j=1}^n \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}_{BN}}{\widehat{\sigma}_{BN}}\right) \\ &+ \frac{1}{2} \log\left| X^\top Z(\widehat{\sigma}_{BN}, \widehat{\phi}_{BN}) X \right| - \log\left| X^\top Z(\widehat{\sigma}_{BN}, \widehat{\phi}_{BN}) \widehat{V}_{\phi} \right| - \sum_{j \in \overline{C}} \left(\frac{y_j - \mathbf{x}_j \widehat{\phi}_{BN,\sigma}}{\sigma}\right) \\ &+ \sum_{j=1}^n \exp\left(\frac{y_j - \mathbf{x}_j \widehat{\phi}_{BN,\sigma}}{\sigma}\right) - \frac{1}{2} \log\left| X^\top Z(\sigma, \widehat{\phi}_{BN,\sigma}) X \right| + \log\left| X^\top Z(\sigma, \widehat{\phi}_{BN,\sigma}) \widehat{V}_{\phi} \right| \Bigg\}. \end{split}$$

Similarly to what has been established in the proof of Result 3,

$$\sum_{j \in \overline{C}} \left(\frac{y_j - \mathbf{x}_j \hat{\phi}_{BN}}{\hat{\sigma}_{BN}} \right) \quad \text{and} \quad \sum_{j \in \overline{C}} \left(\frac{y_j - \mathbf{x}_j \hat{\phi}_{BN,\sigma}}{\sigma} \right)$$

are distributed independently of the pair (σ, ϕ) . The determinants of the quadratic forms are also distributed independently of such a pair. Additionally, from one of the p restrictions (s = 1) we obtain

$$\sum_{j=1}^{n} \exp\left(\frac{y_j - \mathbf{x}_j \hat{\phi}_{BN}}{\hat{\sigma}_{BN}}\right) = r \quad \text{and} \quad \sum_{j=1}^{n} \exp\left(\frac{y_j - \mathbf{x}_j \hat{\phi}_{BN,\sigma}}{\sigma}\right) = r.$$

Recall that $x_{j1} = 1$ for $j = 1, \ldots, n$.

Therefore, for complete data and for type II censored data, $\Pr_{(\sigma,\phi)}(LR_{\tilde{BN}}(\sigma) \leq q)$ does not depend on ϕ nor on the value of σ set at the null hypothesis. This happens because under the null hypothesis, $LR_{\tilde{BN}}(\sigma)$ depends only on ancillary statistics and on pivots for (σ, ϕ) . Note also that the null distribution of this test statistic depends on the model matrix X.

RESULT 7: $(\widehat{\lambda^{\dagger}}_{CR} - \lambda^{\dagger})/\sigma$, $\widehat{\sigma}_{CR}/\sigma$, $(\widehat{\lambda^{\dagger}}_{CR} - \lambda^{\dagger})/\widehat{\sigma}_{CR}$ and $(y_j - \mathbf{x}_j \widehat{\lambda^{\dagger}}_{CR})/\widehat{\sigma}_{CR}$ are distributed independently of the pair (σ, ϕ) .

Proof:

Consider the adjusted profile log-likelihood function $\ell_{CR}(\sigma)$ given in (4.3),

$$\ell_{CR, \mathcal{Y}}(\sigma, \lambda^{\dagger}) = -(r-p)\log\sigma + \sum_{j\in\overline{C}} \left(\frac{y_j - \mathbf{x}_j \lambda^{\dagger}}{\sigma}\right) - \sum_{j=1}^n \exp(w_j) - \frac{1}{2}\log\left|X^\top W X\right|,$$

subject to the restrictions of the form $\partial \ell_{\mathcal{Y}}(\sigma, \lambda^{\dagger}) / \partial \lambda_s^{\dagger} = -(1/\sigma) \sum_{j \in \overline{C}} x_{js} + (1/\sigma) \sum_{j=1}^n x_{js} \exp(w_j) = 0$, where $s = 1, \ldots, p$ and $w_j = (y_j - \mathbf{x}_j \lambda^{\dagger}) / \sigma + (1 - \gamma) \mathbf{x}_j (X^\top X)^{-1} X^\top \mathbb{1}_n, j = 1, \ldots, n$. Since $\hat{\lambda}_{\sigma}^{\dagger}$ does not have closed form, $\ell_{CR}(\sigma)$ is written as a function of the pair $(\sigma, \lambda^{\dagger})$. Here, $\ell_{CR}(\sigma) = \ell_{CR}(\sigma, \lambda^{\dagger}) = \ell_{CR}, \ \mathcal{Y}(\sigma, \lambda^{\dagger})$.

Let

$$y'_{j} = dy_{j} + \mathbf{x}_{j}k, \quad \lambda^{\dagger \prime} = d\lambda^{\dagger} + k \quad \text{and} \quad \sigma' = d\sigma,$$

 $j = 1, \ldots, n$, where d is a positive scalar and k is a p-vector. Let $\mathcal{Y}' = (y'_1, \ldots, y'_n)^{\top}$. Note that

$$\ell_{CR, \mathcal{Y}}(\sigma, \lambda^{\dagger}) = (r - p) \log d + \ell_{CR, \mathcal{Y}'}(\sigma', \lambda^{\dagger'}) \quad \Longleftrightarrow \quad L_{CR, \mathcal{Y}}(\sigma, \lambda^{\dagger}) = d^{r-p} L_{CR, \mathcal{Y}'}(\sigma', \lambda^{\dagger'}).$$

Note that $\partial \ell_{\mathcal{Y}}(\sigma, \lambda^{\dagger}) / \partial \lambda_s^{\dagger} = 0 \iff \partial \ell_{\mathcal{Y}'}(\sigma', \lambda^{\dagger'}) / \partial \lambda_s^{\dagger'} = 0$, where $s = 1, \ldots, p$.

Let λ^{\dagger}_{CR} and $\hat{\sigma}_{CR}$ be the maximum likelihood estimators of λ^{\dagger} and σ obtained from the maximization of $\ell_{CR}(\sigma, \lambda^{\dagger})$ subject to the *p* restrictions, respectively. Hence, if ℓ_{CR} , $\gamma(\sigma, \lambda^{\dagger})$ is maximized at

$$\widehat{\lambda}^{\dagger}_{CR} = \widehat{\lambda}^{\dagger}_{CR}(y_{(1)}, \dots, y_{(r)}) \text{ and } \widehat{\sigma}_{CR} = \widehat{\sigma}_{CR}(y_{(1)}, \dots, y_{(r)}),$$

then $\ell_{CR, \mathcal{Y}'}(\sigma, \lambda^{\dagger})$ is maximized at

$$\widehat{\lambda^{\dagger}}_{CR}(y'_{(1)},\ldots,y'_{(r)}) = d\widehat{\lambda^{\dagger}}_{CR}(y_{(1)},\ldots,y_{(r)}) + k$$

and

$$\widehat{\sigma}_{CR}(y'_{(1)},\ldots,y'_{(r)}) = d\widehat{\sigma}_{CR}(y_{(1)},\ldots,y_{(r)})$$

Letting $d = 1/\sigma$ and $k = -(1/\sigma)\lambda^{\dagger}$, similarly to what has been done in the proof of Result 1, it is possible to show that

$$\frac{1}{\sigma} \Big[\widehat{\lambda^{\dagger}}_{CR} - \lambda^{\dagger} \Big], \quad \frac{\widehat{\sigma}_{CR}}{\sigma}, \quad \frac{1}{\widehat{\sigma}_{CR}} \Big[\widehat{\lambda^{\dagger}}_{CR} - \lambda^{\dagger} \Big] \quad \text{and} \quad \frac{y_j - \mathbf{x}_j \widehat{\lambda^{\dagger}}_{CR}}{\widehat{\sigma}_{CR}}$$

are distributed independently of the pair $(\sigma, \lambda^{\dagger})$, since (j = 1, ..., n)

$$\frac{y_j - \mathbf{x}_j \lambda^{\dagger}}{\sigma} = z_j - (1 - \gamma) \mathbf{x}_j (X^{\top} X)^{-1} X^{\top} \mathbb{1}_n$$

is a pivot for $(\sigma, \lambda^{\dagger})$.

RESULT 8: The distribution of the ratio $\hat{\sigma}_{CR}/\sigma$ depends on the model matrix X. Additionally, for complete and type II censored data, $\Pr_{(\sigma,\phi)}(LR_{BN}(\sigma) \leq q)$ depends neither on ϕ nor on the value of σ set at the null hypothesis; it only depends on X.

Proof:

The expressions for $\ell_{CR}(\sigma, \lambda^{\dagger})$ and $\tilde{\ell}_{BN}(\sigma, \phi)$ are similar, and, hence, in a similar fashion to Result 5, it is possible to show that the distribution of the ratio $\hat{\sigma}_{CR}/\sigma$ depends only on the design matrix X.

Additionally, similarly to Result 6, it can be shown that $\Pr_{(\sigma,\lambda^{\dagger})}(LR_{CR}(\sigma) \leq q)$ does not depend on λ^{\dagger} not on the value of σ set at the null hypothesis.

TYPE I CENSORING

We shall consider type I censoring: $y_j = \log [\min(t_j, c)]$, where $t_j \sim WB(\alpha(\mathbf{x}_j), \beta)$ (j = 1, ..., n) and c is a given positive constant, the censoring time. In Section 5 such a constant was

$$c = \left[\frac{1}{n}\sum_{j=1}^{n}\alpha(\mathbf{x}_j)\right]\left(-\log p\right)^{1/\beta},$$

where p is the desired censored proportion. The data can then be conveniently represented by n pairs of random variables (y_j, δ_j) , where

$$y_j = \log \left[\min(t_j, c)\right]$$
 and $\delta_j = \begin{cases} 1, t_j \le c, \\ 0, t_j > c. \end{cases}$

RESULT 9: $(\hat{\phi} - \phi)/\sigma$ and $\hat{\sigma}/\sigma$ are *not* distributed indepently of the pair (σ, ϕ) , where $\hat{\phi}$ and $\hat{\sigma}$ are the maximum likelihood estimators of ϕ and σ , respectively.

Proof:

Note that $\delta_1, \ldots, \delta_n$ are independent random variables, with $\delta_j \sim \text{Bernoulli}[\Pr(\delta_j = 1)], \quad j = 1, \ldots, n$. Thus,

$$\Pr(\delta_j = 1) = \Pr(t_j \le c) = \Pr\left(t_j \le \left[\frac{1}{n}\sum_{j=1}^n \alpha(\mathbf{x}_j)\right] \left(-\log p\right)^{1/\beta}\right)$$

Recall that if $u_j \sim \mathcal{E}xp(1)$, then $t_j = \alpha(\mathbf{x}_j)u_j^{1/\beta} \sim WB(\alpha(\mathbf{x}_j), \beta)$ para $j = 1, \ldots, n$. Using this relation, we obtain

$$\Pr(\delta_j = 1) = \Pr\left(u_j \le \left[\frac{\sum_{j=1}^n \alpha(\mathbf{x}_j)}{n\alpha(\mathbf{x}_j)}\right]^\beta \left(-\log p\right)\right) = 1 - \exp\left\{\left[\frac{\sum_{j=1}^n \alpha(\mathbf{x}_j)}{n\alpha(\mathbf{x}_j)}\right]^\beta \log p\right\}$$

That is,

$$\Pr(\delta_j = 1) = 1 - p^{\left\{ \left[\sum_{j=1}^n \alpha(\mathbf{x}_j) \right] / \left[n \alpha(\mathbf{x}_j) \right] \right\}^{\beta}}.$$

Hence,

$$\delta_j \sim \operatorname{Bernoulli}\left(1 - p^{\left\{\left[\sum_{j=1}^n \alpha(\mathbf{x}_j)\right] / \left[n\alpha(\mathbf{x}_j)\right]\right\}^{\beta}}\right)$$

and, thus,

$$\psi_j(t) = \left(1 - p^{\left\{\left[\sum_{j=1}^n \alpha(\mathbf{x}_j)\right] / \left[n\alpha(\mathbf{x}_j)\right]\right\}^{\beta}}\right) \exp(it) + p^{\left\{\left[\sum_{j=1}^n \alpha(\mathbf{x}_j)\right] / \left[n\alpha(\mathbf{x}_j)\right]\right\}^{\beta}}$$

is the characteristic function of δ_j , for $j = 1, \ldots, n$, where $i = \sqrt{-1}$.

Since $\beta = \sigma^{-1}$ and $\alpha(\mathbf{x}_j) = \exp(\mathbf{x}_j\phi)$, then δ_j is not an ancillary statistic for (σ, ϕ) . Given that $\delta_1, \ldots, \delta_n$ are independent random variables, the characteristic function of $r = \sum_{j=1}^n \delta_j$ is the product of the individual characteristic functions, and hence depends on the pair (σ, ϕ) . As a consequence, the joint distribution of the random vector $(z_{(1)}, \ldots, z_{(r)})$ depends on such a pair. Therefore, using (A1), when the variates are independent but not necessarily identically distributed and are subject to type I censoring, the quantities

$$\frac{1}{\sigma} \left[\widehat{\phi} - \phi \right] \quad \text{and} \quad \frac{\widehat{\sigma}}{\sigma}$$

are not pivots for (σ, ϕ) . Note that

$$\mathbf{x}_j = \mathbf{x}, \ \forall j \implies \delta_j \sim \text{Bernoulli}(1-p), \ \forall j.$$

Thus, $r \sim \text{Binomial}(n, 1-p)$. This holds for i.i.d. variates in the presence of type I censoring.

ACKNOWLEDGEMENTS

We gratefully acknowledge partial financial support from CNPq and FAPESP.

REFERENCES

- [1] Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. Biometrika, 67, 293–310.
- Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. Biometrika, 70, 343–365.
- [3] Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society B, 49, 1–39.
- [4] Cox, D.R. and Reid, N. (1989). On the stability of maximum-likelihood estimators of orthogonal parameters. *Canadian Journal of Statistics*, 17, 229–233.
- [5] Doornik, J.A. (2001). Ox: an Object-oriented Matrix Programming Language, 4th edition. London: Timberlake Consultants and Oxford: http://www.doornik.com.

- [6] Ferrari, S.L.P., Ferreira da Silva, M. and Cribari-Neto, F. (2005). Adjusted profile likelihoods for the Weibull shape parameter. *Journal of Statistical Computation and Simulation*, to appear.
- [7] Fraser, D. A. S. and Reid, N. (1995). Ancillaries and third-order significance. Utilitas Mathematica, 47, 33–53.
- [8] Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, 86, 655–661.
- [9] Gross, A.J. and Clark, V.A. (1975). Survival Distribution: Reliability Applications in the Biomedical Sciences. New York: Wiley.
- [10] Klein, J.P. and Moeschberger, M. (1997). Survival Analysis. New York: Springer-Verlag.
- [11] Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. New York: Wiley.
- [12] Magnus, J. R. and Neudecker, H. (1988). Matrix Differential Calculus with Applications in Statistics and Econometrics. New York: Wiley.
- [13] Mann, N.R., Schafer, R.E. and Singpurwalla, N.D. (1974). Methods for Statistical Analysis and Reliability and Life Data. New York: Wiley.
- [14] Severini, T.A. (2000). Likelihood Methods in Statistics. Oxford: Oxford University Press.
- [15] Yang, Z. and Xie, M. (2003). Efficient estimation of the Weibull shape parameter based on a modified profile likelihood. *Journal of Statistical Computation and Simulation*, **73**, 115–123.