

Improved Likelihood Inference in Birnbaum–Saunders Regressions

Artur J. Lemonte, Silvia L. P. Ferrari

*Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010,
São Paulo/SP, 05508-090, Brazil*

Francisco Cribari–Neto

*Departamento de Estatística, Universidade Federal de Pernambuco, Cidade
Universitária, Recife/PE, 50740-540, Brazil*

Abstract

The Birnbaum–Saunders regression model is commonly used in reliability studies. We address the issue of performing inference in this class of models when the number of observations is small. Our simulation results suggest that the likelihood ratio test tends to be liberal when the sample size is small. We obtain a correction factor which reduces the size distortion of the test. Also, we consider a parametric bootstrap scheme to obtain improved critical values and improved p -values for the likelihood ratio test. The numerical results show that the modified tests are more reliable in finite samples than the usual likelihood ratio test. We also present an empirical application.

Key words: Bartlett correction; Birnbaum–Saunders distribution; Bootstrap; Likelihood ratio test; Maximum likelihood estimation.

1 Introduction

Different models have been proposed for lifetime data, such as those based on the gamma, lognormal and Weibull distributions. These models typically provide a satisfactory fit in the middle portion of the data, but oftentimes fail to deliver a good fit at the tails, where only a few observations are generally available. The family of distributions proposed by Birnbaum and Saunders (1969) can also be used to model lifetime data. It has the appealing feature of providing satisfactory tail fitting. This family of distributions was originally obtained from a model in which failure follows from the development

and growth of a dominant crack. It was later derived by Desmond (1985) using a biological model which followed from relaxing some of the assumptions originally made by Birnbaum and Saunders (1969).

The random variable T is said to be Birnbaum–Saunders distributed with parameters $\alpha, \eta > 0$, denoted $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, if its distribution function is given by

$$F_T(t) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{t}{\eta}} - \sqrt{\frac{\eta}{t}} \right) \right], \quad t > 0, \quad (1)$$

where $\Phi(\cdot)$ is the standard normal distribution function; α and η are shape and scale parameters, respectively. It is easy to show that η is the median of the distribution: $F_T(\eta) = \Phi(0) = 1/2$. For any $k > 0$, it follows that $kT \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, k\eta)$. It is also noteworthy that the reciprocal property holds: $T^{-1} \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta^{-1})$, which is in the same family of distributions [Saunders (1974)].

Rieck and Nedelman (1991) proposed a log-linear regression model based on the Birnbaum–Saunders distribution. They showed that if $T \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, then $y = \log(T)$ is sinh-normal distributed with shape, location and scale parameters given by $\alpha, \mu = \log(\eta)$ and $\sigma = 2$, respectively [$y \sim \mathcal{SN}(\alpha, \mu, \sigma)$]; see Section 2 for further details. Their model has been widely used and is an alternative to the usual gamma, lognormal and Weibull regression models; see Rieck and Nedelman (1991, § 7). Diagnostic tools for the Birnbaum–Saunders regression model were developed by Galea et al. (2004), Leiva et al. (2007) and Xie and Wei (2007), and Bayesian inference was developed by Tisionas (2001).

Hypothesis testing inference is usually performed using the likelihood ratio test. It is well known, however, that the limiting null distribution (χ^2) used in the test can be a poor approximation to the exact null distribution of the test statistic when the number of observations is small, thus yielding a size-distorted test; see, e.g., the simulation results in Rieck and Nedelman (1991, § 5). Consider, for instance, the application in which interest lies in modeling the die lifetime (T) in the process of metal extrusion, as in Lepadatu et al. (2005). As noted by the authors, the die life is mainly determined by its material properties and the stresses under load. They also note that the extrusion die is exposed to high temperatures, which can also be damaging. The covariates are the friction coefficient (x_1), the angle of the die (x_2) and work temperature (x_3). Consider a regression model which also includes interaction effects, i.e.,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \beta_6 x_{2i} x_{3i} + \varepsilon_i,$$

where $y_i = \log(T_i)$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$, $i = 1, 2, \dots, n$. There are only 15 observations ($n = 15$), and we wish to test the significance of the interaction

effects, i.e., the interest lies in testing $\mathcal{H}_0: \beta_4 = \beta_5 = \beta_6 = 0$. The likelihood ratio p -value equals 0.094, i.e., one rejects the null hypothesis at the 10% nominal level. Note, however, that the p -value is close to the significance level of the test and that the number of observations is small. Can the inference made using the likelihood ratio test be trusted? We shall return to this application in Section 6.

The chief goal of our paper is to improve likelihood ratio inference in Birnbaum–Saunders regressions when the number of observations available to the practitioner is small. We do so by following two different approaches. First, we derive a Bartlett correction factor that can be applied to the likelihood ratio test statistic. The exact null distribution of the modified statistic is generally better approximated by the limiting null distribution used in the test than that of the unmodified test statistic. Second, we consider a parametric bootstrap resampling scheme to obtain improved critical values and improved p -values for the likelihood ratio test.

The paper unfolds as follows. Section 2 introduces the Birnbaum–Saunders regression model. In Section 3, we derive a Bartlett correction to the likelihood ratio test statistic; we give a closed-form expression for the correction factor in matrix form. Special cases are considered in Section 4. Numerical evidence of the effectiveness of the finite sample correction we obtain is presented in Section 5; we also evaluate bootstrap-based inference. Section 6 addresses the empirical application introduced above (inferences on die lifetime in metal extrusion). Finally, concluding remarks are offered in Section 7.

2 The Birnbaum–Saunders regression model

The density function of a Birnbaum–Saunders variate T is

$$f_T(t; \alpha, \eta) = \frac{1}{2\alpha\eta\sqrt{2\pi}} \left[\left(\frac{\eta}{t}\right)^{1/2} + \left(\frac{\eta}{t}\right)^{3/2} \right] \exp \left\{ -\frac{1}{2\alpha^2} \left(\frac{t}{\eta} + \frac{\eta}{t} - 2 \right) \right\},$$

where $t, \alpha, \eta > 0$. The density is right skewed, the skewness decreasing with α ; see Lemonte et al. (2007, § 2). The mean and variance of T are, respectively,

$$\mathbb{E}(T) = \eta \left(1 + \frac{1}{2}\alpha^2 \right) \quad \text{and} \quad \text{Var}(T) = (\alpha\eta)^2 \left(1 + \frac{5}{4}\alpha^2 \right).$$

McCarter (1999) considered $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$ parameter estimation under type II data censoring. Lemonte et al. (2007) derived the second order biases of the maximum likelihood estimators of α and η , and obtained a corrected likelihood

ratio statistic for testing hypotheses regarding α . Lemonte et al. (2008) proposed several bootstrap bias-corrected estimators of α and η . Further details on the Birnbaum–Saunders distribution can be found in Johnson et al. (1995).

The $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$ survival function is $S_T(t) = 1 - F_T(t)$, where $F_T(t)$ is given in (1). The hazard function is $\nu(t) = f_T(t)/S_T(t)$, where $f_T(t)$ is the corresponding density function. The hazard function $\nu(t)$ equals zero at $t = 0$, increases up to a maximum value and then decreases towards a given positive level; see Kundu et al. (2008). For a comparison between the Birnbaum–Saunders and lognormal hazard functions, see Nelson (1990).

As noted in the previous section, Rieck and Nedelman (1991) showed that if $T \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, then $y = \log(T)$ follows a sinh-normal distribution with the following shape, location and scale parameters: α , $\mu = \log(\eta)$ and $\sigma = 2$, respectively, denoted $y \sim \mathcal{SN}(\alpha, \mu, \sigma)$. The density function of y is

$$f(y; \alpha, \mu, \sigma) = \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{y - \mu}{\sigma}\right) \exp\left\{-\frac{2}{\sigma^2} \sinh^2\left(\frac{y - \mu}{\sigma}\right)\right\}, \quad y \in \mathbb{R}.$$

This distribution has a number of interesting and attractive properties [see Rieck (1989)]: (i) It is symmetric around the location parameter μ ; (ii) It is unimodal for $\alpha \leq 2$ and bimodal for $\alpha > 2$; (iii) The mean and variance of y are $\mathbb{E}(y) = \mu$ and $\text{Var}(y) = \sigma^2 w(\alpha)$, respectively. There is no closed-form expression for $w(\alpha)$, but Rieck (1989) obtained asymptotic approximations for both small and large values of α ; (iv) If $y_\alpha \sim \mathcal{SN}(\alpha, \mu, \sigma)$, then $S_\alpha = 2(y_\alpha - \mu)/(\alpha\sigma)$ converges in distribution to the standard normal distribution when $\alpha \rightarrow 0$.

Rieck and Nedelman (1991) proposed the following regression model:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2)$$

where y_i is the logarithm of the i th observed lifetime, $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$ contains the i th observation on the p covariates ($p < n$), $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ is a vector of unknown regression parameters, and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$.

The log-likelihood function for a random sample $\mathbf{y} = (y_1, \dots, y_n)^\top$ from (2) can be written as

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \log(8\pi) + \sum_{i=1}^n \log(\xi_{i1}) - \frac{1}{2} \sum_{i=1}^n \xi_{i2}^2, \quad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$,

$$\xi_{i1}(\boldsymbol{\theta}) = \xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mu_i}{2}\right), \quad \xi_{i2}(\boldsymbol{\theta}) = \xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mu_i}{2}\right)$$

and $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, $i = 1, 2, \dots, n$. By differentiating (3) with respect to β_r and α , we obtain

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_r} = \frac{1}{2} \sum_{i=1}^n x_{ir} \left\{ \xi_{i1} \xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}} \right\}, \quad r = 1, 2, \dots, p,$$

and

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = -\frac{n}{\alpha} + \frac{1}{\alpha} \sum_{i=1}^n \xi_{i2}^2.$$

The score function for $\boldsymbol{\beta}$ can be written in matrix form as

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{U}_\beta = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \frac{1}{2} \mathbf{X}^\top \mathbf{s},$$

where $\mathbf{X} = (\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n)^\top$ is the $n \times p$ design matrix (which is assumed to have full column rank) and $\mathbf{s} = \mathbf{s}(\boldsymbol{\theta})$ is an n -vector whose i th element equals $\xi_{i1} \xi_{i2} - \xi_{i2}/\xi_{i1}$.

Rieck and Nedelman (1991) obtained a closed-form expression for the maximum likelihood estimator (MLE) of α^2 :

$$\hat{\alpha}^2 = \frac{4}{n} \sum_{i=1}^n \sinh^2 \left(\frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{2} \right),$$

where $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$. There is no closed-form expression for the MLE of $\boldsymbol{\beta}$. Hence, one has to use a nonlinear optimization method, such as Newton-Raphson or Fisher's scoring, to obtain $\hat{\boldsymbol{\beta}}$.¹

Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha})^\top$ be the MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$. Rieck and Nedelman (1991) showed that $\hat{\boldsymbol{\theta}} \stackrel{A}{\sim} \mathcal{N}_{p+1}(\boldsymbol{\theta}, \mathbf{K}(\boldsymbol{\theta})^{-1})$, when n is large, ^A denoting approximately distributed; $\mathbf{K}(\boldsymbol{\theta})$ is Fisher's information matrix and $\mathbf{K}(\boldsymbol{\theta})^{-1}$ is its inverse. Also, $\mathbf{K}(\boldsymbol{\theta})$ is a block-diagonal matrix given by $\mathbf{K}(\boldsymbol{\theta}) = \text{diag}\{\mathbf{K}(\boldsymbol{\beta}), \kappa_{\alpha, \alpha}\}$: $\mathbf{K}(\boldsymbol{\beta}) = \psi_1(\alpha)(\mathbf{X}^\top \mathbf{X})/4$ is Fisher's information for $\boldsymbol{\beta}$ and $\kappa_{\alpha, \alpha} = 2n/\alpha^2$ is the information relative to α . Also,

$$\psi_0(\alpha) = \left\{ 1 - \text{erf} \left(\frac{\sqrt{2}}{\alpha} \right) \right\} \exp \left(\frac{2}{\alpha^2} \right) \quad \text{and} \quad \psi_1(\alpha) = 2 + \frac{4}{\alpha^2} - \frac{\sqrt{2\pi}}{\alpha} \psi_0(\alpha), \quad (4)$$

$\text{erf}(\cdot)$ denoting the error function:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

¹ All log-likelihood maximizations with respect to $\boldsymbol{\beta}$ and α in this paper were carried out using the BFGS quasi-Newton method with analytic first derivatives; see Press et al. (1992). The initial values in the iterative BFGS scheme were $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ for $\boldsymbol{\beta}$ and $\sqrt{\tilde{\alpha}^2}$ for α , where $\tilde{\alpha}^2$ is obtained from $\hat{\alpha}^2$ with $\hat{\boldsymbol{\beta}}$ replaced by $\tilde{\boldsymbol{\beta}}$.

Details on $\text{erf}(\cdot)$ can be found in Gradshteyn and Ryzhik (2007). Since $\mathbf{K}(\boldsymbol{\theta})$ is block-diagonal, $\boldsymbol{\beta}$ and α are globally orthogonal [Cox and Reid (1987)] and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ are asymptotically independent. It can be shown that when α is small, $\psi_0(\alpha) \approx \alpha/\sqrt{2\pi}$ and $\psi_1(\alpha) \approx 1 + 4/\alpha^2$; when α is large, $\psi_0(\alpha) \approx 1$ and $\psi_1(\alpha) \approx 2$.

3 An improved likelihood ratio test

Consider a parametric model $f(\mathbf{y}; \boldsymbol{\theta})$ with corresponding log-likelihood function $\ell(\boldsymbol{\theta}; \mathbf{y})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is a k -vector of unknown parameters. The dimensions of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are $k - q$ and q , respectively. Suppose the interest lies in testing the composite null hypothesis $\mathcal{H}_0: \boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^{(0)}$ against $\mathcal{H}_2: \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_2^{(0)}$, where $\boldsymbol{\theta}_2^{(0)}$ is a given vector of scalars. Hence, $\boldsymbol{\theta}_1$ is a vector of nuisance parameters. The log-likelihood ratio test statistic can be written as

$$LR = 2\{\ell(\widehat{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\widetilde{\boldsymbol{\theta}}; \mathbf{y})\}, \quad (5)$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^\top, \widehat{\boldsymbol{\theta}}_2^\top)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_1^\top, \boldsymbol{\theta}_2^{(0)\top})^\top$ are the MLEs of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ obtained from the maximization of $\ell(\boldsymbol{\theta}; \mathbf{y})$ under \mathcal{H}_1 and \mathcal{H}_0 , respectively.

Bartlett (1937) computed the expected value of LR under \mathcal{H}_0 up to order n^{-1} : $\mathbb{E}(LR) = q + B(\boldsymbol{\theta}) + O(n^{-2})$, where $B(\boldsymbol{\theta})$ is a constant of order $O(n^{-1})$. It is possible to show that, under the null hypothesis, the mean of the modified test statistic

$$LR_b = \frac{LR}{1 + B(\boldsymbol{\theta})/q}$$

equals q when we neglect terms of order $O(n^{-2})$. The order of the approximation remains unchanged when the unknown parameters in $B(\boldsymbol{\theta})$ are replaced by their restricted MLEs. Additionally, whereas $\Pr(LR \leq z) = \Pr(\chi_q^2 \leq z) + O(n^{-1})$, it follows that $\Pr(LR_b \leq z) = \Pr(\chi_q^2 \leq z) + O(n^{-2})$, a clear improvement. The correction factor $c = 1 + B(\boldsymbol{\theta})/q$ is commonly referred to as the ‘Bartlett correction factor’.

Note that LR can be written as

$$LR = 2\{\ell(\widehat{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\boldsymbol{\theta}; \mathbf{y})\} - 2\{\ell(\widetilde{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\boldsymbol{\theta}; \mathbf{y})\},$$

where $\ell(\boldsymbol{\theta}; \mathbf{y})$ is the log-likelihood function at the true parameter values. Lawley (1956) has shown that

$$2\mathbb{E}[\ell(\widehat{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\boldsymbol{\theta}; \mathbf{y})] = k + \epsilon_k + O(n^{-2}),$$

where ϵ_k is of order $O(n^{-1})$ and is given by

$$\epsilon_k = \sum' (\lambda_{rstu} - \lambda_{rstuvw}), \quad (6)$$

where \sum' denotes summation over all components of $\boldsymbol{\theta}$, i.e., the indices r, s, t, u, v and w vary over all k parameters, and the λ 's are given by

$$\begin{aligned} \lambda_{rstu} &= \kappa^{rs} \kappa^{tu} \left\{ \frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \\ \lambda_{rstuvw} &= \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \kappa_{rtv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)} \right) \right. \\ &\quad \left. + \kappa_{rtu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right\}, \end{aligned} \quad (7)$$

where $\kappa_{rs} = \mathbb{E}(\partial^2 \ell(\boldsymbol{\theta}) / \partial \theta_r \partial \theta_s)$, $\kappa_{rst} = \mathbb{E}(\partial^3 \ell(\boldsymbol{\theta}) / \partial \theta_r \partial \theta_s \partial \theta_t)$, $\kappa_{rs}^{(t)} = \partial \kappa_{rs} / \partial \theta_t$, etc., and $-\kappa^{rs}$ is the (r, s) element of Fisher's information matrix inverse. Analogously,

$$2 \mathbb{E}[\ell(\tilde{\boldsymbol{\theta}}; \mathbf{y}) - \ell(\boldsymbol{\theta}; \mathbf{y})] = k - q + \epsilon_{k-q} + O(n^{-2}),$$

where ϵ_{k-q} is of order $O(n^{-1})$ and is obtained from (6) when the sum \sum' only covers the components of $\boldsymbol{\theta}_1$, i.e., the sum ranges over the $k - q$ nuisance parameters, since $\boldsymbol{\theta}_2$ is fixed under \mathcal{H}_0 .

Under \mathcal{H}_0 , $\mathbb{E}(LR) = q + \epsilon_k - \epsilon_{k-q} + O(n^{-2})$. Thus, it is possible to achieve a better χ_q^2 approximation by using the modified test statistic $LR_b = LR/c$ instead of LR , the Bartlett correction factor being $c = 1 + B(\boldsymbol{\theta})/q$, where $B(\boldsymbol{\theta}) = \epsilon_k - \epsilon_{k-q}$. The corrected statistic LR_b is χ_q^2 distributed up to order $O(n^{-1})$ under \mathcal{H}_0 . The improved test follows from the comparison of LR_b and the critical value obtained as the appropriate χ_q^2 quantile.

The corrected test statistic is usually written as $LR_b = LR/\{1 + B(\boldsymbol{\theta})/q\}$. Nonetheless, there are alternative modified statistics that are equivalent to LR_b to order $O(n^{-1})$, such as $LR_b^* = LR \exp\{-B(\boldsymbol{\theta})/q\}$ and $LR_b^{**} = LR\{1 - B(\boldsymbol{\theta})/q\}$. It is noteworthy that LR_b^* has an advantage over the other two specifications: it never assumes negative values. See Cribari–Neto and Cordeiro (1996) for further details on Bartlett corrections.

In what follows, we shall derive the Bartlett correction factor for testing inference in the Birnbaum–Saunders regression model. The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$, which is $(p + 1)$ -dimensional. Hence, we shall obtain ϵ_{p+1} from (6), with the indices varying from 1 up to $p + 1$.

Let $\mathbf{Z} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \{z_{ij}\}$ and $\mathbf{Z}_d = \text{diag}\{z_{11}, z_{22}, \dots, z_{nn}\}$. Also, $\mathbf{Z}^{(2)} = \mathbf{Z} \odot \mathbf{Z}$, $\mathbf{Z}_d^{(2)} = \mathbf{Z}_d \odot \mathbf{Z}_d$, etc., \odot denoting the Hadamard (element-wise) product of matrices. We shall use the following notation for cumulants of log-likelihood derivatives with respect to $\boldsymbol{\beta}$ and α : $U_r = \partial \ell(\boldsymbol{\theta}) / \partial \beta_r$, $U_\alpha =$

$\partial\ell(\boldsymbol{\theta})/\partial\alpha$, $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s$, $U_{r\alpha} = \partial^2\ell(\boldsymbol{\theta})/\partial\beta_r\partial\alpha$, $U_{\alpha\alpha} = \partial^2\ell(\boldsymbol{\theta})/\partial\alpha^2$, $U_{rst} = \partial^3\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s\partial\beta_t$, $U_{rs\alpha} = \partial^3\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s\partial\alpha$, etc; $\kappa_{rs} = \mathbb{E}(U_{rs})$, $\kappa_{r\alpha} = \mathbb{E}(U_{r\alpha})$, $\kappa_{rst} = \mathbb{E}(U_{rst})$, etc; $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, $\kappa_{r\alpha}^{(t\alpha)} = \partial^2\kappa_{r\alpha}/\partial\beta_t\partial\alpha$, etc.

From the log-likelihood function in (3) we obtain the following cumulants:

$$\begin{aligned} \kappa_{rs} &= -\frac{\psi_1(\alpha)}{4} \sum_{i=1}^n x_{ir}x_{is}, & \kappa_{r\alpha} &= 0, & \kappa_{\alpha\alpha} &= -\frac{2n}{\alpha^2}, \\ \kappa_{rst} &= 0, & \kappa_{rs\alpha} &= \frac{2+\alpha^2}{\alpha^3} \sum_{i=1}^n x_{ir}x_{is}, & \kappa_{r\alpha\alpha} &= 0, & \kappa_{\alpha\alpha\alpha} &= \frac{10n}{\alpha^3}, \\ \kappa_{rstu} &= \psi_2(\alpha) \sum_{i=1}^n x_{ir}x_{is}x_{it}x_{iu}, & \kappa_{rst\alpha} &= 0, & \kappa_{rs\alpha\alpha} &= -\frac{3(2+\alpha^2)}{\alpha^4} \sum_{i=1}^n x_{ir}x_{is}, \\ & & \kappa_{r\alpha\alpha\alpha} &= 0 & \text{and} & \kappa_{\alpha\alpha\alpha\alpha} &= -\frac{54n}{\alpha^4}, \end{aligned}$$

where

$$\psi_2(\alpha) = -\frac{1}{4} \left\{ 2 + \frac{7}{\alpha^2} - \sqrt{\frac{\pi}{2}} \left(\frac{1}{2\alpha} + \frac{6}{\alpha^3} \right) \psi_0(\alpha) \right\}$$

and $\psi_0(\alpha)$ and $\psi_1(\alpha)$ are defined in (4). For small α , we have $\psi_2(\alpha) \approx -5/8 - 1/\alpha^2$; for large α , $\psi_2(\alpha) \approx -1/2$.

Using these cumulants and also making use of the orthogonality between $\boldsymbol{\beta}$ and α , we obtain, after long and tedious algebra (Appendix), $\epsilon_{p+1} = \epsilon(\alpha, p, \mathbf{X})$, where

$$\epsilon(\alpha, p, \mathbf{X}) = \epsilon_\alpha(\alpha, p) + \epsilon_\beta(\alpha, \mathbf{X}), \quad (8)$$

with

$$\epsilon_\alpha(\alpha, p) = \frac{1}{n} \left\{ \frac{1}{3} + \delta_1(\alpha)p + \delta_2(\alpha)p^2 \right\} \quad \text{and} \quad \epsilon_\beta(\alpha, \mathbf{X}) = \delta_3(\alpha) \text{tr}(\mathbf{Z}_d^{(2)}).$$

Here, $\text{tr}(\cdot)$ denotes the trace operator and

$$\begin{aligned} \delta_0(\alpha) &= \frac{2+\alpha^2}{\psi_1(\alpha)\alpha^2}, & \delta_1(\alpha) &= 4\delta_0(\alpha) \left\{ \frac{2}{2+\alpha^2} + \delta_0(\alpha) - \frac{2\alpha\psi_3(\alpha)}{\psi_1(\alpha)} \right\}, \\ \delta_2(\alpha) &= 2\delta_0(\alpha)^2, & \delta_3(\alpha) &= \frac{4\psi_2(\alpha)}{\psi_1(\alpha)^2} \quad \text{and} \quad \psi_3(\alpha) = \frac{3}{\alpha^3} - \frac{\sqrt{2\pi}}{4\alpha^2} \left(1 + \frac{4}{\alpha^2} \right) \psi_0(\alpha). \end{aligned}$$

In expression (8) – our main result – we write ϵ_{p+1} as the sum of two terms, namely $\epsilon_\alpha(\alpha, p)$ and $\epsilon_\beta(\alpha, \mathbf{X})$. The quantity $\epsilon_\beta(\alpha, \mathbf{X})$ is obtained from (6) with Σ' ranging over the components of $\boldsymbol{\beta}$, i.e. as if α were known. The quantity $\epsilon_\alpha(\alpha, p)$ is the contribution yielded by the fact that α is unknown (see the Appendix). Note that $\epsilon_\alpha(\alpha, p)$ depends on the design matrix only through its rank. More specifically, it is a second degree polynomial in p divided by n . Hence, $\epsilon_\alpha(\alpha, p)$ can be non-negligible if the dimension of $\boldsymbol{\beta}$ is not considerably

smaller than the sample size. It is also noteworthy that $\epsilon(\alpha, p, \mathbf{X})$ depends on α but not on $\boldsymbol{\beta}$. The dependency of $\epsilon(\alpha, p, \mathbf{X})$ on α occurs through $\delta_1(\alpha)$, $\delta_2(\alpha)$ and $\delta_3(\alpha)$. For small α , we have $\delta_1(\alpha) \approx 1$, $\delta_2(\alpha) \approx 1/2$ and $\delta_3(\alpha) \approx 0$. For large α , $\delta_1(\alpha) \approx 1$, $\delta_2(\alpha) \approx 1/2$ and $\delta_3(\alpha) \approx -1/2$. Furthermore, $\text{tr}(\mathbf{Z}_d^{(2)})$ establishes the dependency of $\epsilon(\alpha, p, \mathbf{X})$ on \mathbf{X} . In other words, $\epsilon(\alpha, p, \mathbf{X})$ depends on the sum of squares of the diagonal elements of the hat matrix \mathbf{Z} . In particular, if $p = 1$, i.e. if \mathbf{X} has a single column, $\mathbf{x} = (x_1, \dots, x_n)^\top$ say, then $\text{tr}(\mathbf{Z}_d^{(2)}) = \sum_{i=1}^n x_i^4 / \left(\sum_{i=1}^n x_i^2 \right)^2$, the sample kurtosis of \mathbf{x} .

Finally, it should be noted that expression (8) is quite simple and can be easily implemented into any mathematical or statistical/econometric programming environment, such as R [R Development Core Team (2006)], `ox` [Cribari–Neto and Zarkos (2003); Doornik (2006)] and `MAPLE` [Abell and Braselton (1994)].

4 Special cases

In this section we present closed-form expressions for the Bartlett correction factor in situations that are of particular interest to practitioners. The simplified expressions are obtained from our more general result given in (8).

At the outset, we consider the test of $\mathcal{H}_0: \alpha = \alpha^{(0)}$ against $\mathcal{H}_1: \alpha \neq \alpha^{(0)}$, where $\alpha^{(0)}$ is a given positive scalar and $\boldsymbol{\beta}$ is a vector of nuisance parameters. The Bartlett correction factor becomes $c = 1 + B(\boldsymbol{\theta})$, where $B(\boldsymbol{\theta}) = \epsilon(\alpha, p, \mathbf{X}) - \epsilon_{\boldsymbol{\beta}}(\alpha, \mathbf{X})$, and hence, $B(\boldsymbol{\theta}) = \epsilon_{\alpha}(\alpha, p)$. Note that the correction factor depends on \mathbf{X} only through its rank, p . In particular, when $p = 1$ (i.i.d. case), we have

$$B(\boldsymbol{\theta}) = \frac{1}{n} \left\{ \frac{1}{3} + \delta_1(\alpha) + \delta_2(\alpha) \right\}.$$

This formula corrects eq. (14) in Lemonte et al. (2007), which is in error. For small and large values of α , we have $B(\boldsymbol{\theta}) \approx 11/(6n)$.

Oftentimes practitioners wish to test restrictions on a subset of the regression parameters. For instance, one may want to test whether a given group of covariates are jointly significant. To that end, we partition $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, where $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \dots, \beta_{p-q})^\top$ and $\boldsymbol{\beta}_2 = (\beta_{p-q+1}, \beta_{p-q+2}, \dots, \beta_p)^\top$ are vectors of dimensions $(p - q) \times 1$ and $q \times 1$, respectively, and consider the test of $\mathcal{H}_0: \boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^{(0)}$ against $\mathcal{H}_1: \boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_2^{(0)}$, where $\boldsymbol{\beta}_2^{(0)}$ is a q -vector of known constants. The most common situation is that in which $\boldsymbol{\beta}_2^{(0)} = \mathbf{0}$. Note that $\boldsymbol{\beta}_1$ and α are nuisance parameters. In accordance with the partition of $\boldsymbol{\beta}$, we partition \mathbf{X} as $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$, where the dimensions of \mathbf{X}_1 and \mathbf{X}_2 are $n \times (p - q)$ and $n \times q$, respectively. The correction factor is $c = 1 + B(\boldsymbol{\theta})/q$,

where $B(\boldsymbol{\theta}) = \epsilon(\alpha, p, \mathbf{X}) - \epsilon(\alpha, p - q, \mathbf{X}_1)$. It is easy to obtain

$$B(\boldsymbol{\theta}) = \frac{1}{n} \left\{ \delta_1(\alpha)q + \delta_2(\alpha)q(2p - q) \right\} + \delta_3(\alpha) \text{tr}(\mathbf{Z}_d^{(2)} - \mathbf{Z}_{1d}^{(2)}),$$

with $\mathbf{Z}_1 = \mathbf{X}_1(\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top = \{z_{1ij}\}$ and $\mathbf{Z}_{1d} = \text{diag}\{z_{111}, z_{122}, \dots, z_{1nn}\}$.

Next, suppose we wish to test $\mathcal{H}_0: \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ against $\mathcal{H}_1: \boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)}$, where $\boldsymbol{\beta}^{(0)}$ is a p -vector of known constants and α is a nuisance parameter. The Bartlett correction factor is $c = 1 + B(\boldsymbol{\theta})/p$ with $B(\boldsymbol{\theta}) = \epsilon(\alpha, p, \mathbf{X}) - \epsilon_\alpha(\alpha, 0)$, which yields

$$B(\boldsymbol{\theta}) = \frac{1}{n} \left\{ \delta_1(\alpha)p + \delta_2(\alpha)p^2 \right\} + \delta_3(\alpha) \text{tr}(\mathbf{Z}_d^{(2)}).$$

5 Numerical evidence

We shall now report Monte Carlo evidence on the finite sample performance of three tests in Birnbaum–Saunders regressions, namely: the likelihood ratio test (LR), the Bartlett-corrected likelihood ratio test (LR_b), and an asymptotically equivalent corrected test (LR_b^*).² The model used in the numerical evaluation is

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i,$$

where $x_{i1} = 1$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$, $i = 1, 2, \dots, n$. The covariate values were selected as random draws from the $\mathcal{U}(0, 1)$ distribution. The number of Monte Carlo replications was 10,000, the nominal levels of the tests were $\gamma = 10\%$, 5% and 1% , and all simulations were carried out using the `Ox` matrix programming language (Doornik, 2006).

Table 1 presents the null rejection rates (entries are percentages) of the three tests. The null hypothesis is $\mathcal{H}_0: \beta_{p-1} = \beta_p = 0$, which is tested against a two-sided alternative, the sample size is $n = 30$ and $\alpha = 0.5$. Different values of p were considered. The values of the response were generated using $\beta_1 = \beta_2 = \dots = \beta_{p-2} = 1$.

Note that the likelihood ratio test is considerably oversized (liberal), more so as the number of regressors increases. For instance, when $p = 8$ and $\gamma = 10\%$, its null rejection rate is 18.78%, i.e., nearly twice the nominal level of the test. The two corrected tests are much less size distorted. For example, their null rejection rates in the same situation were 11.82% (LR_b) and 11.13% (LR_b^*).

The results in Table 2 correspond to $\alpha = 0.5$ and $p = 6$. We report results for samples sizes ranging from 20 to 200. The null hypothesis under test is

² We do not report results relative to LR_b^{**} since they were very similar to those obtained using LR_b^* .

Table 1

Null rejection rates; $\alpha = 0.5$, $n = 30$.

p	$\gamma = 10\%$			$\gamma = 5\%$			$\gamma = 1\%$		
	LR	LR_b	LR_b^*	LR	LR_b	LR_b^*	LR	LR_b	LR_b^*
3	12.69	10.36	10.22	6.51	4.98	4.90	1.75	1.25	1.23
4	13.44	10.27	10.07	7.46	5.41	5.32	1.90	1.10	1.04
5	14.77	10.74	10.45	8.25	5.53	5.31	2.21	1.18	1.14
6	15.94	11.07	10.53	9.14	5.55	5.23	2.54	1.24	1.17
7	17.28	11.55	10.88	10.13	5.69	5.42	2.95	1.29	1.19
8	18.78	11.82	11.13	11.15	6.44	5.83	3.38	1.48	1.31
9	19.92	12.11	11.00	12.00	6.33	5.66	3.82	1.49	1.25

\mathcal{H}_0 : $\beta_5 = \beta_6 = 0$. The figures in this table show that the null rejection rates of all tests approach the corresponding nominal levels as the sample size grows, as expected. It is also noteworthy that the likelihood ratio test displays liberal behavior even when $n = 100$. Overall, the corrected tests are less size distorted than the unmodified test. For example, when $n = 50$ and $\gamma = 5\%$, the null rejection rates are 7.49% (LR), 5.32% (LR_b) and 5.17% (LR_b^*).

Table 2

Null rejection rates; $\alpha = 0.5$, $p = 6$ and different sample sizes.

n	$\gamma = 10\%$			$\gamma = 5\%$			$\gamma = 1\%$		
	LR	LR_b	LR_b^*	LR	LR_b	LR_b^*	LR	LR_b	LR_b^*
20	19.54	12.04	11.08	11.97	6.54	5.87	4.05	1.58	1.38
30	15.94	11.07	10.53	9.14	5.55	5.23	2.54	1.24	1.17
40	13.57	10.14	9.97	7.45	4.99	4.81	1.79	1.03	1.01
50	13.36	10.72	10.51	7.49	5.32	5.17	1.51	1.02	0.99
100	11.86	10.46	10.44	5.90	4.92	4.88	1.25	1.04	1.03
200	10.92	10.14	10.12	5.57	5.07	5.07	1.04	0.96	0.96

Figure 1 plots relative quantile discrepancies against the associated asymptotic quantiles for the three test statistics. Relative quantile discrepancies are defined as the difference between exact (estimated by Monte Carlo) and asymptotic quantiles divided by the latter. Again, $p = 6$ and we test the exclusion of the last two covariates. Also, $n = 30$ and $\alpha = 0.5$. The closer to zero the relative quantile discrepancies, the more accurate the test. While Tables 1 and 2 give rejection rates of the tests at fixed nominal levels, Figure 1 compares the whole distributions of the different statistics with the limiting null distribution. We note that the relative quantile discrepancies of the likelihood ratio test statistic oscillates around 25% whereas for the two corrected statistics they are around 5% (LR_b) and 3% (LR_b^*). It is thus clear that the null distributions of the modified statistics are much better approximated by the limiting null distribution (χ_2^2) than that of the likelihood ratio statistic.

Table 3 contains the nonnull rejection rates (powers) of the tests. Here, $p = 4$, $\alpha = 0.5$ and $n = 30, 50, 100$. Data generation was performed under the

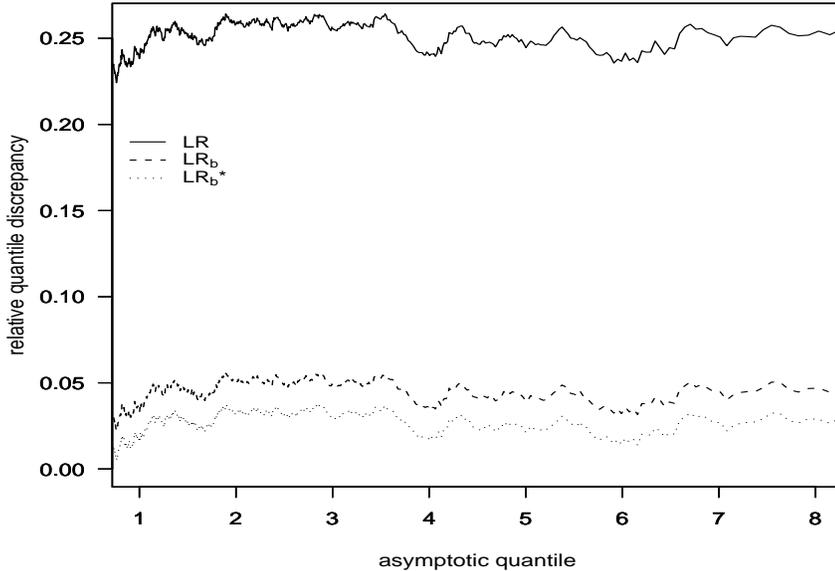


Fig. 1. Relative quantile discrepancies plot: $n = 30$, $p = 6$ and $\alpha = 0.5$.

alternative hypothesis: $\beta_3 = \beta_4 = \delta$, with different values of δ ($\delta > 0$). We have only considered the two corrected tests since the likelihood ratio is considerably oversized, as noted earlier. Note that the two tests display similar powers. For instance, when $n = 50$, $\gamma = 5\%$ and $\delta = 0.5$, the nonnull rejection rates are 72.39% (LR_b) and 72.28% (LR_b^*). We also note that the powers of the tests increase with n and also with δ , as expected.

Table 4 presents the null rejection rates for inference on the scalar parameter α . Here, $n = 30$ and $p = 2, 3$ and 4 . The null hypotheses under test are $\mathcal{H}_0: \alpha = 0.5$ and $\mathcal{H}_0: \alpha = 1.0$. The likelihood ratio test is again liberal. Note that the two corrected tests are much less size distorted. For instance, when $p = 4$, $\gamma = 5\%$ and $\alpha = 1.0$, the null rejection rates of the LR , LR_b and LR_b^* tests were 12.03%, 5.20% and 4.02%, respectively.

Our simulation results concerning tests on the regression parameters were obtained for $\alpha = 0.5$. In practice, values of α between 0 and 1 cover most of the applications; see, for instance, Rieck and Nedelman (1991). We shall now present simulation results for a wide range of values of α , namely $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9, 1.2, 2, 10, 50$ and 100 . The new set of simulation results includes rejection rates of the likelihood ratio test that uses parametric bootstrap critical values (with 600 bootstrap replications). The parametric bootstrap can be briefly described as follows. We can use bootstrap resampling to estimate the null distribution of the statistic LR directly from the observed sample $\mathbf{y} = (y_1, \dots, y_n)^\top$. To that end, one generates, under \mathcal{H}_0 (i.e., imposing the restrictions stated in the null hypothesis), B bootstrap samples

Table 3

Nonnull rejection rates; $\alpha = 0.5$, $p = 4$ and different sample sizes.

n	δ	LR_b			LR_b^*		
		10%	5%	1%	10%	5%	1%
30	0.1	13.20	6.91	1.57	13.01	6.73	1.52
	0.2	20.66	12.22	3.46	20.40	12.02	3.30
	0.3	33.07	21.63	7.49	32.73	21.28	7.33
	0.4	48.36	35.57	14.96	48.08	35.20	14.61
	0.5	65.11	51.59	26.42	64.72	51.19	25.99
50	0.1	13.82	7.63	2.03	13.71	7.60	1.99
	0.2	25.89	16.03	5.11	25.81	15.96	5.03
	0.3	45.00	32.06	13.15	44.86	31.95	13.07
	0.4	65.07	52.09	28.18	64.97	51.96	28.03
	0.5	82.31	72.39	48.01	82.14	72.28	47.88
100	0.1	18.66	11.02	2.91	18.65	11.01	2.90
	0.2	43.63	31.29	13.05	43.61	31.28	13.02
	0.3	73.47	62.39	37.40	73.47	62.35	37.34
	0.4	92.12	86.39	69.15	92.11	86.37	69.11
	0.5	98.76	97.34	89.98	98.76	97.33	89.93

Table 4

Null rejection rates; inference on α ; $n = 30$ and different values for p .

p		$\mathcal{H}_0: \alpha = 0.5$			$\mathcal{H}_0: \alpha = 1.0$		
		10%	5%	1%	10%	5%	1%
2	LR	12.76	6.99	1.82	13.55	7.26	1.93
	LR_b	10.13	5.15	1.21	10.52	5.10	1.08
	LR_b^*	9.90	5.02	1.20	10.31	4.94	1.05
3	LR	15.16	8.64	2.45	16.10	9.35	2.70
	LR_b	10.46	5.43	1.16	10.53	5.06	0.97
	LR_b^*	9.77	5.02	1.02	9.65	4.68	0.81
4	LR	18.16	10.72	3.39	19.86	12.03	3.73
	LR_b	10.45	5.37	0.98	10.73	5.20	0.75
	LR_b^*	9.29	4.57	0.72	8.77	4.02	0.43

$(\mathbf{y}^{*1}, \dots, \mathbf{y}^{*B})$ from the assumed model with the parameters replaced by restricted estimates computed using the original sample (parametric bootstrap), and, for each pseudo-sample, one computes $LR^{*b} = 2\{\ell(\hat{\theta}^{*b}; \mathbf{y}^{*b}) - \ell(\tilde{\theta}^{*b}; \mathbf{y}^{*b})\}$, $b = 1, 2, \dots, B$, where $\hat{\theta}^{*b}$ and $\tilde{\theta}^{*b}$ are the maximum likelihood estimators of θ obtained from the maximizations of $\ell(\theta; \mathbf{y}^{*b})$ under \mathcal{H}_0 and \mathcal{H}_1 , respectively. The $1 - \gamma$ percentile of LR^{*b} is estimated by $\hat{q}_{1-\gamma}$, such that $\#\{LR^{*b} \leq \hat{q}_{1-\gamma}\}/B = 1 - \gamma$. One rejects the null hypothesis if $LR > \hat{q}_{1-\gamma}$. For a good discussion of bootstrap tests, see Efron and Tibshirani (1993, Chapter 16).

Figures in Table 5 provide important information. For all values of α the

Bartlett and the bootstrap corrections are very effective in pushing the rejection rates toward the nominal levels. An advantage of the Bartlett correction over the bootstrap approach is that the first requires much less computational effort. It is noteworthy that as α grows, the rejection rates of the likelihood ratio test approaches the corresponding nominal levels, making the corrections less needed.

Table 5

Null rejection rates of $\mathcal{H}_0: \beta_3 = \beta_4 = 0$; $p = 4$, $n = 25$ and different values for α .

γ	$\alpha = 0.1$				$\alpha = 0.3$			
	LR	LR_b	LR_b^*	LR_{boot}	LR	LR_b	LR_b^*	LR_{boot}
10%	14.33	11.42	11.32	9.90	14.45	10.70	10.44	10.18
5%	7.88	5.82	5.74	4.94	8.01	5.49	5.22	4.97
1%	1.95	1.29	1.23	1.24	2.07	1.20	1.13	1.13
γ	$\alpha = 0.5$				$\alpha = 0.7$			
	LR	LR_b	LR_b^*	LR_{boot}	LR	LR_b	LR_b^*	LR_{boot}
10%	14.25	10.23	10.01	10.23	14.17	10.61	10.36	10.14
5%	7.69	5.17	5.02	5.12	8.09	5.35	5.17	5.28
1%	1.89	0.91	0.85	1.24	2.07	1.12	1.06	1.02
γ	$\alpha = 0.9$				$\alpha = 1.2$			
	LR	LR_b	LR_b^*	LR_{boot}	LR	LR_b	LR_b^*	LR_{boot}
10%	13.96	10.80	10.60	9.64	13.49	10.45	10.29	9.79
5%	8.03	5.77	5.61	5.17	7.51	5.37	5.26	5.10
1%	2.29	1.30	1.26	1.10	1.90	1.18	1.16	1.38
γ	$\alpha = 2$				$\alpha = 10$			
	LR	LR_b	LR_b^*	LR_{boot}	LR	LR_b	LR_b^*	LR_{boot}
10%	13.21	10.87	10.64	10.01	12.44	11.23	11.13	9.75
5%	7.29	5.61	5.50	5.08	6.59	5.81	5.73	4.80
1%	1.63	1.08	1.04	1.21	1.42	1.17	1.16	0.98
γ	$\alpha = 50$				$\alpha = 100$			
	LR	LR_b	LR_b^*	LR_{boot}	LR	LR_b	LR_b^*	LR_{boot}
10%	11.26	10.45	10.43	10.11	10.87	10.17	10.12	9.89
5%	5.60	5.21	5.20	4.93	5.67	5.11	5.07	5.18
1%	1.19	1.06	1.06	1.04	1.23	1.07	1.07	1.18

We shall now try to shed some light on the issue of the possible effect of near-collinearity between the covariates \mathbf{X} on the testing procedures. To do so, we performed an additional simulation experiment. We set $p = 4$ and selected the covariate values as follows: $x_{i1} = 1$, for $i = 1, \dots, n$, the values of x_2 were chosen as random draws from the $\mathcal{U}(0, 1)$ distribution and the pairs (x_{i3}, x_{i4}) were selected as random draws from the bivariate normal distribution

$\mathcal{N}_2(\mathbf{0}, \Sigma)$, where the covariance matrix Σ has the following form

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The closer the value of ρ is to either extreme (-1 or 1), the stronger the linear relation between the covariates x_3 and x_4 . Table 6 presents simulation results for different values of ρ . The figures in this table suggest that the sample correlation between x_2 and x_3 does not have significant effect on the behaviour of the testing procedures. Hence, near-collinearity does not seem to be a matter of concern.

Table 6

Null rejection rates of $\mathcal{H}_0: \beta_2 = \beta_4 = 0$; $p = 4$, $n = 20$ and different values for ρ .

$\rho = 0.0$				
γ	LR	LR_b	LR_b^*	LR_{boot}
10%	16.00	11.15	10.72	10.20
5%	9.16	5.33	5.08	4.90
1%	2.37	1.27	1.21	1.16
$\rho = 0.5$				
γ	LR	LR_b	LR_b^*	LR_{boot}
10%	15.54	10.72	10.33	10.14
5%	8.85	5.73	5.48	5.22
1%	2.37	1.15	1.03	1.10
$\rho = 0.9$				
γ	LR	LR_b	LR_b^*	LR_{boot}
10%	15.73	11.14	10.77	10.31
5%	9.18	6.06	5.70	5.40
1%	2.58	1.15	1.10	1.21

In all simulated situations, the likelihood ratio test was liberal. Of course, this is not a proof that this is always the case. Indeed, there may be situations where it is conservative. Simulation results presented in the literature, however, suggest that the likelihood ratio test is often anti-conservative. For a theoretical justification in a simple situation, let z_1, \dots, z_n be a random sample drawn from the $N(\mu, \sigma^2)$ distribution, with both μ and σ^2 unknown. Consider the test of $\mathcal{H}_0: \mu = \mu_0$ versus $\mathcal{H}_1: \mu \neq \mu_0$. The asymptotic likelihood ratio test rejects \mathcal{H}_0 whenever $LR > c_\gamma$, where c_γ is the $1 - \gamma$ quantile of the χ_1^2 distribution. Equivalently, \mathcal{H}_0 is rejected when $\sqrt{n}|\bar{z} - \mu_0|/\hat{\sigma} > k(\gamma, n)$, where $\bar{z} = \sum_{i=1}^n z_i/n$, $\hat{\sigma}^2 = \sum_{i=1}^n (z_i - \bar{z})^2/(n-1)$ and $k(\gamma, n) = \sqrt{(\exp(-c_\gamma/2))^{-2/n} - 1}(n-1)$. Table 7 shows the true levels of the likelihood ratio test, i.e. $\Pr(LR > c_\gamma)$ evaluated at \mathcal{H}_0 , for different values of n and γ . Notice that, even in this simple situation, the likelihood ratio test is liberal when the sample is not large, in agreement with simulation results presented

elsewhere. See, for instance, Rieck and Nedelman (1991, Table 4) and Cordeiro et al. (1995).

Table 7

True level; normal distribution.

n	γ		
	1%	5%	10%
5	2.91	9.79	16.54
8	1.97	7.64	13.72
12	1.58	6.64	12.35
20	1.32	5.93	11.35
50	1.12	5.36	10.52

6 An application

We shall now turn to an empirical application that employs real data. We consider the investigation made by Lepadatu et al. (2005) on metal extrusion die lifetime. As noted by the authors (p. 38), “the estimation of tool life (fatigue life) in the extrusion operation is important for scheduling tool changing times, for adaptive process control and for tool cost evaluation.” They also note (p. 39) that “die fatigue cracks are caused by the repeat application of loads which individually would be too small to cause failure.” According to them, current research aims at describing the whole fatigue process by focusing on the analysis of crack propagation from very small initial defects. It is noteworthy that fatigue failure due to propagation of an initial crack was the main motivation for the Birnbaum–Saunders distribution.

In Section 1, we explained that the interest lies in modeling the die lifetime (T) in the metal extrusion process, which is mainly determined by its material properties and by the stresses under load. The extrusion die is exposed to high temperatures, which can also be damaging. The covariates are the friction coefficient (x_1), the angle of the die (x_2) and work temperature (x_3). Consider a regression model which also includes interaction effects, i.e.,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \beta_6 x_{2i} x_{3i} + \varepsilon_i, \quad (9)$$

where $y_i = \log(T_i)$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$, $i = 1, 2, \dots, n$. There are only 15 observations ($n = 15$), and we want make inference on the significance of the interaction effects, i.e., we wish to test $\mathcal{H}_0: \beta_4 = \beta_5 = \beta_6 = 0$. The likelihood ratio test statistic (LR) equals 6.387 (p -value 0.094), and the two corrected test statistics are $LR_b = 4.724$ (p -value 0.193) and $LR_b^* = 4.492$ (p -value 0.213). The p -value of the bootstrap-based likelihood ratio test is 0.276. It is noteworthy that one rejects the null hypothesis at the 10% nominal level when the inference is based on the likelihood ratio test, but a different inference is

reached when the modified (Bartlett-corrected or bootstrap-based) tests are used. Recall from the previous section that the unmodified test is oversized when the sample is small (here, $n = 15$), which leads us to mistrust the inference delivered by the likelihood ratio test.

We proceed by removing the interaction effects (as suggested by the three modified tests) from Model (9). We then estimate

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i,$$

$i = 1, \dots, 15$. The point estimates are (standard errors in parentheses): $\hat{\beta}_0 = 5.9011$ (0.488), $\hat{\beta}_1 = 0.7917$ (1.777), $\hat{\beta}_2 = 0.0098$ (0.012), $\hat{\beta}_3 = 0.0052$ (0.001) and $\hat{\alpha} = 0.1982$ (0.036). The null hypothesis $\mathcal{H}_0 : \beta_3 = 0$ is strongly rejected by the four tests (unmodified and modified) at the usual significance levels. All tests also suggest the individual and joint exclusions of x_1 and x_2 from the regression model. We thus end up with the reduced model

$$y_i = \beta_0 + \beta_3 x_{3i} + \varepsilon_i,$$

$i = 1, \dots, 15$. The point estimates are (standard errors in parentheses): $\hat{\beta}_0 = 6.2453$ (0.326), $\hat{\beta}_3 = 0.0052$ (0.001) and $\hat{\alpha} = 0.2039$ (0.037).

We now return to Model (9) and test $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$ (exclusion of all variables but x_3). The null hypothesis is not rejected at the 10% nominal level by all tests, but we note that the corrected tests yield considerably larger p -values. The test statistics are $LR = 7.229$, $LR_b = 5.610$ and $LR_b^* = 5.417$, the corresponding p -values being 0.204, 0.346 and 0.367; the p -value obtained from the bootstrap-based likelihood ratio test equals 0.484.

7 Conclusions

We addressed the issue of performing testing inference in Birnbaum–Saunders regressions when the sample size is small. The likelihood ratio test can be considerably oversized (liberal), as evidenced by our numerical results. We derived modified test statistics whose null distributions are more accurately approximated by the limiting null distribution than that of the likelihood ratio test statistic. We have also considered a parametric bootstrap scheme to obtain improved critical values and accurate p -values for the likelihood ratio test. Our simulation results have convincingly shown that inference based on the modified test statistics can be much more accurate than that based on the unmodified statistic. The modified tests behave as reliably as a likelihood ratio test that relies on bootstrap-based critical values, with no need of computer intensive procedures. We recommend the use of the following statistics: LR_b and LR_b^* . The latter has the advantage of only taking on positive values, which

is desirable. We have also presented an empirical application in which the use of the finite sample adjustment proposed in this paper can lead to inferences that are different from the ones reached based on first order asymptotics.

Acknowledgments

We gratefully acknowledge grants from FAPESP and CNPq. We thank an anonymous referee for helpful comments that led to several improvements in this paper.

Appendix

From (6), we have

$$\epsilon_{p+1} = \sum_{r,s,t,u=1}^{p+1} \lambda_{rstu} - \sum_{r,s,t,u,v,w=1}^{p+1} \lambda_{rstuvw}.$$

Note that $\sum_{r,s,t,u=1}^{p+1} \lambda_{rstu}$ can be written as $\sum_{r,s,t,u=1}^p \lambda_{rstu}$ plus terms in which at least one subscript equals α . It follows from the orthogonality between α and β that several terms equal zero. The non-zero terms are $\sum_{r,s=1}^p \lambda_{rs\alpha\alpha}$, $\sum_{t,u=1}^p \lambda_{\alpha\alpha tu}$ and $\lambda_{\alpha\alpha\alpha\alpha}$. Also, $\sum_{r,s,t,u,v,w=1}^{p+1} \lambda_{rstuvw}$ is given by $\sum_{r,s,t,u,v,w=1}^p \lambda_{rstuvw}$ plus the following terms: $\sum_{r,s,t,u=1}^p \lambda_{rstu\alpha\alpha}$, $\sum_{r,s,v,w=1}^p \lambda_{rs\alpha\alpha vw}$, $\sum_{t,u,v,w=1}^p \lambda_{\alpha\alpha tvw}$, $\sum_{r,s=1}^p \lambda_{rs\alpha\alpha\alpha\alpha}$, $\sum_{t,u=1}^p \lambda_{\alpha\alpha tu\alpha\alpha}$, $\sum_{v,w=1}^p \lambda_{\alpha\alpha\alpha\alpha vw}$ and $\lambda_{\alpha\alpha\alpha\alpha\alpha\alpha}$. Here, we present the derivations of $\sum_{r,s,t,u=1}^p \lambda_{rstu}$ and $\sum_{v,w=1}^p \lambda_{\alpha\alpha\alpha\alpha vw}$. The other terms can be obtained in a similar fashion.

Note that $\sum_{r,s,t,u=1}^p \lambda_{rstu} = (1/4) \sum_{r,s,t,u=1}^p \kappa^{rs} \kappa^{tu} \kappa_{rstu}$. Inserting the cumulants given in Section 3 into this expression we have

$$\begin{aligned} \sum_{r,s,t,u=1}^p \lambda_{rstu} &= \frac{1}{4} \sum_{r,s,t,u=1}^p \kappa^{rs} \kappa^{tu} \left\{ \psi_2(\alpha) \sum_{i=1}^n x_{ir} x_{is} x_{it} x_{iu} \right\} \\ &= \frac{\psi_2(\alpha)}{4} \sum_{i=1}^n \sum_{r,s,t,u=1}^p x_{ir} \kappa^{rs} x_{is} x_{it} \kappa^{tu} x_{iu} \\ &= \frac{\psi_2(\alpha)}{4} \sum_{i=1}^n \left\{ \sum_{r,s=1}^p x_{ir} \kappa^{rs} x_{is} \right\} \left\{ \sum_{t,u=1}^p x_{it} \kappa^{tu} x_{iu} \right\} \\ &= \frac{\psi_2(\alpha)}{4} \sum_{i=1}^n (\mathbf{x}_i^\top \mathbf{K}^{\beta\beta} \mathbf{x}_i) (\mathbf{x}_i^\top \mathbf{K}^{\beta\beta} \mathbf{x}_i) = \frac{\psi_2(\alpha)}{4} \sum_{i=1}^n (\mathbf{x}_i^\top \mathbf{K}^{\beta\beta} \mathbf{x}_i)^2, \end{aligned}$$

where $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents the i th row of \mathbf{X} and $\mathbf{K}^{\beta\beta} = \mathbf{K}(\beta)^{-1} = 4(\mathbf{X}^\top \mathbf{X})^{-1} / \psi_1(\alpha)$ represents the inverse of Fisher's information matrix for β . There-

fore,

$$\sum_{r,s,t,u=1}^p \lambda_{rstu} = \frac{4\psi_2(\alpha)}{\psi_1(\alpha)^2} \sum_{i=1}^n \{\mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i\}^2.$$

Note that $z_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i$ is the i th diagonal element of \mathbf{Z}_d given in Section 3.

Hence,

$$\sum_{r,s,t,u=1}^p \lambda_{rstu} = \frac{4\psi_2(\alpha)}{\psi_1(\alpha)^2} \sum_{i=1}^n z_{ii}^2 = \frac{4\psi_2(\alpha)}{\psi_1(\alpha)^2} \text{tr}(\mathbf{Z}_d^{(2)}).$$

From $\sum_{v,w=1}^p \lambda_{\alpha\alpha\alpha\alpha vw} = (1/4)(\kappa^{\alpha\alpha})^2 \kappa_{\alpha\alpha\alpha} \sum_{v,w=1}^p \kappa^{vw} \kappa_{\alpha vw}$, we obtain

$$\begin{aligned} \sum_{v,w=1}^p \lambda_{\alpha\alpha\alpha\alpha vw} &= \frac{\alpha^4}{4n^2} \frac{5n}{2\alpha^3} \sum_{v,w=1}^p \kappa^{vw} \left\{ \frac{2+\alpha^2}{\alpha^3} \sum_{i=1}^n x_{iv} x_{iw} \right\} \\ &= \frac{5(2+\alpha^2)}{8n\alpha^2} \sum_{i=1}^n \left\{ \sum_{v,w=1}^p x_{iv} \kappa^{vw} x_{iw} \right\} = -\frac{5(2+\alpha^2)}{8n\alpha^2} \sum_{i=1}^n (\mathbf{x}_i^\top \mathbf{K}^{\beta\beta} \mathbf{x}_i) \\ &= -\frac{5(2+\alpha^2)}{2n\alpha^2 \psi_1(\alpha)} \sum_{i=1}^n \{\mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i\} \\ &= -\frac{5(2+\alpha^2)}{2n\alpha^2 \psi_1(\alpha)} \sum_{i=1}^n z_{ii} = -\frac{5(2+\alpha^2)}{2n\alpha^2 \psi_1(\alpha)} \text{tr}(\mathbf{Z}_d) = -\frac{5(2+\alpha^2)p}{2n\alpha^2 \psi_1(\alpha)}. \end{aligned}$$

References

- [1] Abell, M. L. and Braselton, J. P. (1994). *The Maple V Handbook*. AP Professional, New York.
- [2] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, **160**, 268–282.
- [3] Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, **6**, 319–327.
- [4] Cordeiro, G. M., Cribari–Neto, F., Aubin, E. C. Q. and Ferrari, S. L. P. (1995). Bartlett corrections for one-parameter exponential family models. *Journal of Statistical Computation and Simulation*, **53**, 211–231.
- [5] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **40**, 1–39.
- [6] Cribari–Neto, F. and Cordeiro, G. M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews*, **15**, 339–367.
- [7] Cribari–Neto, F. and Zarkos, S. (2003). Econometric and statistical computing using Ox. *Computational Economics*, **21**, 277–295.

- [8] Desmonde, A. F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics*, **13**, 171–183.
- [9] Doornik, J. A. (2006). *An Object-Oriented Matrix Language – Ox 4*. Timberlake Consultants Press, London. 5th ed.
- [10] Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- [11] Galea, M., Leiva, V. and Paula, G. A. (2004). Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics*, **31**, 1049–1064.
- [12] Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Academic Press, New York.
- [13] Johnson, N., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions – Volume 2*, 2nd ed. Wiley, New York.
- [14] Kundu, D., Kannan, N. and Balakrishnan, N. (2008). On the function of Birnbaum–Saunders distribution and associated inference. *Computational Statistics and Data Analysis*, **52**, 2692–2702.
- [15] Lawley, D. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295–303.
- [16] Leiva, V., Barros, M. K., Paula, G. A. and Galea, M. (2007). Influence diagnostics in log-Birnbaum–Saunders regression models with censored data. *Computational Statistics and Data Analysis*, **51**, 5694–5707.
- [17] Lemonte, A. J., Cribari–Neto, F. and Vasconcellos, K. L. P. (2007). Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics and Data Analysis*, **51**, 4656–4681.
- [18] Lemonte, A. J., Simas, A. B. and Cribari–Neto, F. (2008). Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution. *Journal of Statistical Computation and Simulation*, **78**, 37–49.
- [19] Lepadatu, D., Kobi, A., Hambli, R. and Barreau, A. (2005). Lifetime multiple response optimization of metal extrusion die. *Proceedings of the Annual Reliability and Maintainability Symposium*, 37–42.
- [20] McCarter, K. S. (1999). *Estimation and Prediction for the Birnbaum–Saunders Distribution Using Type-II Censored Samples, With a Comparison to the Inverse Gaussian Distribution*. Ph.D. dissertation, Kansas State University.
- [21] Nelson, W. (1990). *Accelerated Testing, Statistical Models, Test Plans and Data Analysis*. Wiley, New York.
- [22] Press, W. H., Teulosky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Prentice Hall, London.

- [23] R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [24] Rieck, J. R. (1989). *Statistical Analysis for the Birnbaum–Saunders Fatigue Life Distribution*. Ph.D. dissertation, Clemson University.
- [25] Rieck, J. R. and Nedelman, J. R. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics*, **33**, 51–60.
- [26] Saunders, S. C. (1974). A family of random variables closed under reciprocation. *Journal of the American Statistical Association*, **69**, 533–539.
- [27] Tisionas, E. G. (2001). Bayesian inference in Birnbaum–Saunders regression. *Communications in Statistics – Theory and Methods*, **30**, 179–193.
- [28] Xie, F. C. and Wei, B. C. (2007). Diagnostics analysis for log-Birnbaum–Saunders regression models. *Computational Statistics and Data Analysis*, **51**, 4692–4706.