RT-MAE 2002-14

BARTLETT CORRECTIONS IN HETEROSKEDASTIC t REGRESSION MODELS

by

Lúcia P. Barroso
and
Gauss M. Cordeiro

Palavras-Chave: Bartlett correction; Dispersion parameter; Heteroskedastic model; Link function; Maximum likelihood estimate; Student *t* distribution.
Classificação AMS: 62F05, 62J05.
(AMS Classification)

- Setembro de 2002 -

BARTLETT CORRECTIONS IN HETEROSKEDASTIC t REGRESSION MODELS

Lúcia P. Barroso [1]

Departamento de Estatística,
Universidade de São Paulo, Brazil
lbarroso@ime.usp.br

Gauss M. Cordeiro

Departamento de Física e Matemática,
Universidade Federal Rural de Pernambuco, Brazil
gauss@ufpe.br

ABSTRACT

This paper gives a general Bartlett correction formula to improve likelihood ratio tests in heteroskedastic t regression models where both location and dispersion parameters vary across observations through regression models. The correction covers many important and commonly used models and can be viewed as an extension of the results in Cordeiro (1993) and Botter and Cordeiro (1997). We present some Monte Carlo investigations of Bartlett corrections that show that this approach have better performance than the classical likelihood ratio tests even under degrees of freedom misspecification.

Key Words: Bartlett correction; Dispersion parameter; Heteroskedastic model; Link function; Maximum likelihood estimate; Student t distribution.

1. INTRODUCTION

The asymptotic chi-squared distribution of the likelihood ratio (LR) statistic is frequently used to test hypotheses of interest in regression models. However, as the sample size n decreases, the use of such statistic becomes less justifiable. An alternative is to use higher order asymptotic theory. In a seminal paper, Bartlett (1937) proposed an improved LR statistic. His argument goes as follows. Suppose that under the null hypothesis $\mathrm{E}(LR) = q\{1 + \frac{b}{n} + o(n^{-3/2})\}$, where b is a constant that can be consistently estimated under the null hypothesis and q is the difference of the dimensions of the parameter spaces under the alternative and null hypotheses. Then, the expected value of the transformed statistic $LR^* = LR/(1 + \frac{b}{n})$ is closer to the one from a χ^2_q distribution than the expected value of LR. This became widely known as the Bartlett correction. LR statistics for nonlinear hypotheses on the regression coefficients can have sizes that are typically larger than their nominal sizes. Then,

[1]Corresponding author: Lúcia P. Barroso, Rua do Matão, 1010 - São Paulo, Brazil, CEP: 05508-900, Fax: 55-11-3091-6130.

the Bartlett correction, under mild regularity conditions, guarantees that all moments of the adjusted LR^* statistic are equal to those of the asymptotic χ_q^2 distribution up to order n^{-1} (Lawley, 1956; Cordeiro, 1987). In fact, the Bartlett correction $1+\frac{b}{n}$ not only corrects the mean of the LR statistic to order $O(n^{-1})$ but also corrects its distribution to this order. An approximation with the same order of error is obtained if the modified LR statistic is defined as $\widetilde{LR} = LR\exp(-\frac{b}{n})$ which always produces a valid (i.e. positive) statistic. The null distribution of the modified statistics LR^* and $\widetilde{LR}$ remains χ_q^2 with approximation error of order $O(n^{-2})$ when b is replaced by a $\sqrt{n}$ consistent estimate (Barndorff-Nielsen and Hall, 1988).

General formulae for Bartlett corrections have been obtained explicitly in several regression models by Cordeiro (1983, 1987), Porteous (1985), Cordeiro and Paula (1989a), Cordeiro, Paula and Botter (1994) and authors cited therein. DiCiccio and Stern (1994) provide a good recent discussion of the Bartlett correction in both the frequentist and the Bayesian setting. We believe that more is needed, applying these techniques in more general settings.

In this paper we present, in matrix notation, a general Bartlett correction formula to improve likelihood ratio tests in heteroskedastic t regression models, thus generalizing results by Cordeiro (1993) and Botter and Cordeiro (1997). Although the algebraic form of the Bartlett correction is cumbersome, it can be easily incorporated in a computer program. This might be a worthwhile practice, since the Bartlett correction is always in the right direction and, in general, gives a substantial improvement.

Lange, Little and Taylor (1989) introduced the regression models with t distributed errors as a robust extension of the traditional normal models. It is well known that the t distribution provides a convenient description for regression analysis when the residual term has a density with heavy tails. Unlike the normal models, the t models lead to inferences about the regression parameters based on asymptotic theory.

We consider a univariate nonlinear heteroskedastic t regression model where both location and dispersion parameters vary across observations through regression models. In our model the observations $Y_1, \ldots, Y_n$ are independent and each observation Y_i has a Student t distribution with location parameter μ_i, dispersion parameter σ_i and ν degrees of freedom. The density of Y_i is given by

$$f(y;\mu_i,\sigma_i,\nu) = \frac{\nu^{\nu/2}\Gamma(\frac{\nu+1}{2})}{\sigma_i\sqrt{\pi}\Gamma(\frac{\nu}{2})}\left\{\nu + \left(\frac{y-\mu_i}{\sigma_i}\right)^2\right\}^{-(\nu+1)/2}, \qquad (1)$$

where $y, \mu_i \in \mathbb{R}$, $\sigma_i > 0$ and $\Gamma(\cdot)$ is the gamma function. We treat ν as a known constant throughout the paper. Although, it seems that this parameter will be hardly known in practical applications, this assumption is not a strong restriction for the application of our results, since we will show that if ν is replaced by a consistent estimate, the Bartlett correction is still effective for improving the asymptotic χ^2 approximation of the LR statistic. The t distribution reduces to the normal and Cauchy distributions when $\nu = \infty$ and $\nu = 1$, respectively. We define the precision parameter $\phi_i = \sigma_i^{-2}$ for each observation. We then assume that the parameters μ_i and ϕ_i vary across observations through regression models which are parameterized as $\mu_i = g(\eta_i)$ and $\phi_i = h(\tau_i)$, where $g(\cdot)$ and $h(\cdot)$ are known one-to-one continuously differentiable functions, and η_i and τ_i are linear combinations of some explanatory variables. The linear predictors $\eta = (\eta_1, \ldots, \eta_n)^\top$ and $\tau = (\tau_1, \ldots, \tau_n)^\top$ are given by the relations $\eta = X\beta$ and $\tau = Z\gamma$, respectively, where X is a specified $n \times p$ matrix of full rank $p < n, \beta = (\beta_1, \ldots, \beta_p)^\top$ is a set of unknown parameters, Z is a specified $n \times q$ matrix of full rank $q < n$ and $\gamma = (\gamma_1, \ldots, \gamma_q)^\top$ is a set of unknown parameters. The functions $g(\cdot)$ and $h(\cdot)$ are usually called the mean and dispersion link functions. The dispersion link function $h(\cdot)$ must be a positive-valued function. A simple choice for $h(\cdot)$ is $h(\tau) = \exp(\tau)$. The dispersion covariates in Z constitute, in general, although not necessarily, a subset of the mean covariates in X. The parameters β and γ are assumed to be functionally independent which leaves us $p+q$ parameters to be estimated.

The paper is organized in the following form. In Section 2 we obtain a general formula for the expectation of the LR criterion which may be handled to derive Bartlett corrections for testing several hypotheses concerning the parameters β and γ in model (1). We also discuss improved LR tests on mean and dispersion effects separately. Since the general formula is difficult to interpret, special models are considered in Section 3, consisting of the simple linear regression structures and of the one-way classification structures for both linear predictors η and τ. Finally, in Section 4, the performance of the correction is evaluated by simulation which shows that the correction adjusts the size in the right direction even when ν is replaced by a consistent estimate or under degrees of freedom misspecification.

2. Bartlett's Correction

From the density in (1), we can easily derive the total log-likelihood function $\ell(\beta, \gamma)$ given the sample $y_1, \ldots, y_n$ for the $(p+q) \times 1$ parameter vector $\theta^\top = (\beta^\top, \gamma^\top)$ in model (1) as

$$\ell(\beta,\gamma) = \frac{1}{2}[c + \sum_i \log h(\tau_i) - (\nu+1)\sum_i \log\{\nu + h(\tau_i)(y_i - g(\eta_i))^2\}], \qquad (2)$$

where c is given by $c = 2n\log\left\{\frac{\nu^{\nu/2}\Gamma(\frac{\nu+1}{2})}{\pi^{1/2}\Gamma(\frac{\nu}{2})}\right\}$ and $\sum_i$ runs over all n observations. The quantity c does not depend on the parameters of the model (since ν is assumed known) and is, therefore, not relevant for our purposes. We assume that some standard regularity conditions on $\ell = \ell(\beta, \gamma)$ and its first four derivatives hold (Cox and Hinkley, 1974, Chapter 9) as n goes to infinity. These conditions are the same regularity conditions required for Edgeworth expansions and are indeed fulfilled in this context.

We use the notation where joint cumulants of log likelihood derivatives are indicated by lower-case indices $r, s, t, \ldots$ if they correspond to the β parameters, while upper-case indices $R, S, T, \ldots$ correspond to the γ parameters. We can obtain

$$\kappa_{rs} = -\frac{(\nu+1)}{(\nu+3)}\sum_i \phi_i g_i'^2 x_{ir} x_{is}, \quad \kappa_{rS} = 0 \quad \text{and} \quad \kappa_{RS} = -\frac{\nu}{2(\nu+3)}\sum_i \left(\frac{h_i'}{\phi_i}\right)^2 z_{iR} z_{iS},$$

where $g_i' = \partial g(\eta_i)/\partial\eta_i$, $h_i' = \partial h(\tau_i)/\partial\tau_i$, etc. Several other joint cumulants for the β and γ parameters needed to find the Bartlett correction in the class of models under study follow by simple differentiation and using standard regularity equations. They are not given here but can be obtained from the authors upon request.

We immediately observe that the parameters β and γ are globally orthogonal (Cox and Reid, 1987) since $\kappa_{rS} = 0$ for all $r = 1,\ldots,p$ and $S = 1,\ldots,q$. Therefore, the joint information matrix K for $\theta = (\beta^\top, \gamma^\top)^\top$ is block-diagonal, $K = \mathrm{diag}\{K_\beta, K_\gamma\}$, and the information matrices K_β and K_γ for β and γ are given by $K_\beta = X^\top W_1 X$ and $K_\gamma = Z^\top W_2 Z$, where W_1 and W_2 are the $n \times n$ diagonal matrices $W_1 = [(\nu+1)/(\nu+3)]\mathrm{diag}\{\phi_i g_i'^2\}$ and $W_2 = [\nu/(2\nu+6)]\mathrm{diag}\{(h_i'/\phi_i)^2\}$. In view of the block-diagonality of K, Fisher's scoring method can be used to obtain the maximum likelihood estimates (MLEs) $\hat\beta$ and $\hat\gamma$ simultaneously by iteratively solving the following equations

$$X^\top W_1^{(m)} X \beta^{(m+1)} = X^\top W_1^{(m)} z_\beta^{(m)}, \quad Z^\top W_2^{(m)} Z \gamma^{(m+1)} = Z^\top W_2^{(m)} z_\gamma^{(m)}, \tag{3}$$

where z_β and z_γ are $n \times 1$ vectors whose components are given by

$$z_{\beta_i} = \eta_i + \frac{(\nu+3)(y_i-\mu_i) g_i'^{-1}}{\nu + \phi_i (y_i-\mu_i)^2}, \quad z_{\gamma_i} = \tau_i + \frac{(\nu+3)\phi_i}{h_i'} \frac{[1-\phi_i(y_i-\mu_i)^2]}{[\nu+\phi_i(y_i-\mu_i)^2]}. \tag{4}$$

Each loop, through the iterative scheme (3), consists of a Fisher scoring step to optimize the log-likelihood (2). Equations (3) show that any software with a weighted linear regression routine can be used to calculate the estimates $\hat\beta$ and $\hat\gamma$ iteratively. Jørgensen (1984), Green (1984) and Cordeiro and Paula (1989b) implemented equations of this type in the GLIM system using an iterated weighted least squares (IWLS) algorithm involving a working dependent variable and a weight matrix that are updated at each iteration.

We consider the hypothesis where both the parameter vector of interest and the nuisance parameter vector may be regarded as being composed of some components of β and γ. Thus, the null hypothesis is devoted for testing the mean and dispersion effects simultaneously. Next, we discuss Bartlett corrections for testing the mean effects and the dispersion effects separately. Partitioning the parameters as $\beta^\top = (\beta_1^\top, \beta_2^\top)$ and $\gamma^\top = (\gamma_1^\top, \gamma_2^\top)$, where β_1 and γ_1 are vectors of dimensions $p_1 \le p$ and $q_1 \le q$, respectively, we are interested in testing the null hypothesis $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ versus H_1^1: violation of at least one equality, where $\beta_1^{(0)}$ and $\gamma_1^{(0)}$ are specified vectors of dimensions p_1 and q_1, respectively. Following the partition induced by H_0^1, let $X = (X_1, X_2)$ and $Z = (Z_1, Z_2)$ be the corresponding partitioned model matrices, where X_1, X_2, Z_1 and Z_2 are, respectively, $n \times p_1, n \times (p-p_1), n \times q_1$ and $n \times (q-q_1)$ matrices of full ranks. We denote the unrestricted MLEs of β and γ by $\hat\beta$ and $\hat\gamma$, while the restricted MLEs of β_2 and γ_2 under H_0^1 are denoted by $\tilde\beta_2$ and $\tilde\gamma_2$. Similarly, functions evaluated at the unrestricted MLEs will be denoted by the addition of a circumflex and those evaluated at the restricted MLEs will be denoted by adding a tilde.

From Lawley's (1956) expansion we can write the expected LR statistic to $O(n^{-1})$ as $2\mathrm{E}\{\ell(\hat{\beta},\hat{\gamma})-\ell(\beta,\gamma)\}=p+q+\epsilon_{p+q}$, where ϵ_{p+q} is a term of order n^{-1} evaluated at the true parameter given by (see also Cordeiro, 1987)

$$\epsilon_{p+q}=\sum_{\beta,\gamma}{}'(l_{rstu}-l_{rstuvw}), \tag{5}$$

and

$$l_{rstu}=\kappa^{rs}\kappa^{tu}\left(\frac{1}{4}\kappa_{rstu}-\kappa_{rst}^{(u)}+\kappa_{rt}^{(su)}\right), \tag{6}$$

$$l_{rstuvw}=\kappa^{rs}\kappa^{tu}\kappa^{vw}\left[\kappa_{rtv}\left(\frac{1}{6}\kappa_{suw}-\kappa_{sw}^{(u)}\right)+\kappa_{rtu}\left(\frac{1}{4}\kappa_{svw}-\kappa_{sw}^{(v)}\right)+\kappa_{rt}^{(v)}\kappa_{sw}^{(u)}+\kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right], \tag{7}$$

where $-\kappa^{rs}$ and $-\kappa^{RS}$ denote the typical elements of the inverses of K_β and K_γ, respectively. We used the Einsteinian summation convention in which summation in $\sum'$ is across parameters appearing both raised and lowered.

The LR statistic for testing $H_0^1:\beta_1=\beta_1^{(0)},\gamma_1=\gamma_1^{(0)}$ is simply $LR_1=2\{\ell(\hat{\beta}_1,\hat{\beta}_2,\hat{\gamma}_1,\hat{\gamma}_2)-\ell(\beta_1^{(0)},\tilde{\beta}_2,\gamma_1^{(0)},\tilde{\gamma}_2)\}$ which is, under H_0^1, asymptotically distributed as $\chi^2_{p_1+q_1}$. The expected LR for testing H_0^1 is given by $\mathrm{E}(LR_1)=p_1+q_1+\epsilon_{p+q}-\epsilon_{p-p_1+q-q_1}$, where the term $\epsilon_{p-p_1+q-q_1}$ is defined analogously to ϵ_{p+q}, the only difference being that the summation in this term is only over the unrestricted parameters β_2 and γ_2. We can obtain ϵ_{p+q} by exploiting special properties of the heteroskedastic t regression model (1).

Some additional notation is in order. We denote as $\mathbf{1}$ the $n\times 1$ vector with ones and $\mathrm{tr}(A)$ the trace of the matrix A. From the information matrices K_β and K_γ for β and γ, we define the $n\times n$ matrices $S_\beta=XK_\beta^{-1}X^\top=X(X^\top W_1X)^{-1}X^\top$ and $S_\gamma=ZK_\gamma^{-1}Z^\top=Z(Z^\top W_2Z)^{-1}Z^\top$. Thus, S_β and S_γ can be interpreted as the asymptotic covariance matrices of $\hat{\eta}=X\hat{\beta}$ and $\hat{\tau}=Z\hat{\gamma}$, respectively. From S_β and S_γ, we also define the $n\times n$ diagonal matrices $S_{\beta d}=\mathrm{diag}(S_\beta)$ and $S_{\gamma d}=\mathrm{diag}(S_\gamma)$ such that the diagonal elements of $S_{\beta d}$ and $S_{\gamma d}$ will be the diagonal elements of S_β and S_γ, respectively. We also define $S_\beta^{(2)}=S_\beta\odot S_\beta, S_\beta^{(3)}=S_\beta^{(2)}\odot S_\beta$, where $\odot$ denotes the Hadamard product. Analogous definitions are used for $S_\gamma^{(2)}$ and $S_\gamma^{(3)}$. We also define in Appendix A the $n\times n$ diagonal matrices $\Phi_1,\ldots,\Phi_8$.

The method used to compute ϵ_p (and ϵ_{p-q}) is not complicated and it consists of the following steps (Cordeiro, 1983; Cordeiro and Paula, 1989a; Cordeiro, Paula and Botter, 1994): (i) incorporate the joint cumulants of log likelihood derivatives into formulae (6) and (7); (ii) perform the sum $\sum'$ over the parameters using simple matrix operations; (iii) then, do the same for the sum $\sum$ over the observations. The details of the calculations are very tedious, yet elementary, and can be obtained from the authors on request. All the summations are greatly simplified because of the orthogonality between β and γ. After some lengthy algebra, the ϵ_p term can be decomposed into the sum of three components as

$$\epsilon_{p+q}=f_1(X,\mu,\phi)+f_2(Z,\mu,\phi)+f_3(X,Z,\mu,\phi), \tag{8}$$

$$f_1(X,\mu,\phi) = \mathrm{tr}(\Phi_1 S_{\beta d}^2) - \frac{1}{4} 1^\top \Phi_4 (2S_\beta^{(3)} - S_{\beta d} S_\beta S_{\beta d}) \Phi_4 1, \tag{9}$$

$$f_2(Z,\mu,\phi) = \mathrm{tr}(\Phi_3 S_{\gamma d}^2) + \frac{1}{12} 1^\top \Phi_7 (2S_\gamma^{(3)} + 3S_{\gamma d} S_\gamma S_{\gamma d}) \Phi_7 1$$
$$+ 1^\top (\Phi_8 - \Phi_7)(S_\gamma^{(3)} + S_{\gamma d} S_\gamma S_{\gamma d}) \Phi_8 1, \tag{10}$$

and

$$f_3(X,Z,\mu,\phi) = \mathrm{tr}(\Phi_2 S_{\beta d} S_{\gamma d}) + \frac{1}{2} \mathrm{tr}[\Phi_5 S_\beta^{(2)} (\Phi_5 - 2\Phi_6) S_\gamma]$$
$$+ \frac{1}{4} 1^\top \Phi_5 S_{\beta d} S_\gamma [S_{\beta d} \Phi_5 + 2S_{\gamma d}(2\Phi_8 - \Phi_7)] 1. \tag{11}$$

Equation (8) shows a simple decomposition for the Bartlett adjustment and represents the main result of the paper. Equation for $\epsilon_{p-p_1+q-q_1}$ is obtained directly from (8) with X_2 and Z_2 in place of X and Z. Thus, the Bartlett correction for testing $H_0^1 : \beta_1 = \beta_1^{(0)}, \gamma_1 = \gamma_1^{(0)}$ is

$$c_1 = 1 + \{f_1(X,\mu,\phi) + f_2(Z,\mu,\phi) + f_3(X,Z,\mu,\phi)$$
$$- f_1(X_2,\mu,\phi) - f_2(Z_2,\mu,\phi) - f_3(X_2,Z_2,\mu,\phi)\}/(p_1+q_1), \tag{12}$$

with the unrestricted parameters β_2 and γ_2 being substituted by their estimates $\tilde{\beta}_2$ and $\tilde{\gamma}_2$ under H_0^1. Then, the improved test compares $LR_1^* = LR_1/\tilde{c}_1$ with the upper point of the $\chi^2_{p_1+q_1}$ distribution.

In small samples the asymptotic χ^2 distribution for the LR statistic is dreadful and the magnitude of the Bartlett correction serves to warn of situations in which this approximation can be improved. Thus, when the correction is small we can be assured that no great error is committed by using the asymptotic χ^2 approximation. When the correction is large, caution is advisable. Expressing the functions f_1, f_2 and f_3 in matrix notation has great advantages to obtain closed-form Bartlett corrections for several tests of practical use in heteroskedastic t regression models. The Bartlett correction only involves simple operations on matrices and vectors and can be easily computed using a computer algebra system such as MAPLE or a programming language with support for matrix operations, such as Ox or S-Plus. Numerical computation involving higher-order joint cumulants is thereby avoided.

We now investigate the problem of improving LR tests on mean and dispersion parameters separately. For testing $H_0^2 : \beta_1 = \beta_1^{(0)}$ against the alternative hypothesis $H_1^2 : \beta_1 \neq \beta_1^{(0)}$, we have $Z_2 = Z$ and the Bartlett correction reduces from (12) to

$$c_2 = 1 + \{f_1(X,\mu,\phi) - f_1(X_2,\mu,\phi) + f_3(X,Z,\mu,\phi) - f_3(X_2,Z,\mu,\phi)\}/p_1,$$

with the unrestricted parameters β_2 and γ being substituted by their estimates $\tilde{\beta}_2$ and $\tilde{\gamma}$ under H_0^2. The Bartlett correction associated with testing the hypothesis $H_0^3 : \gamma_1 = \gamma_1^{(0)}$ is given by (here $X_2 = X$)

$$c_3 = 1 + \{f_2(Z,\mu,\phi) - f_2(Z_2,\mu,\phi) + f_3(X,Z,\mu,\phi) - f_3(X,Z_2,\mu,\phi)\}/q_1,$$

with β and γ_2 being substituted by their estimates $\tilde{\beta}$ and $\tilde{\gamma}_2$ under H_0^3.

3. SPECIAL MODELS

In this section we briefly discuss three special models: the simple model, the one-way classification structures for both linear predictors η and τ and the model defined by simple linear regression structures for these predictors. For all these cases, the mean and the precision parameters, under the null hypothesis, are constant over the observations.

3.1. THE SIMPLE MODEL

We begin with the simple model where the observations are independent and identically distributed (iid). For this case, $p = q = 1$ and X and Z are n-vectors of ones. We consider a test of the null hypothesis $H_0^1 : \beta = \beta^{(0)}, \gamma = \gamma^{(0)}$ against $H_1^1 : H_0^1$ is false. After some algebra, we obtain from (12) and (9)–(11)

$$c_1 = 1 + \frac{(\nu+3)}{12n\nu(\nu+1)(\nu+5)^2(\nu+7)}[11\nu^4 + 147\nu^3 + 461\nu^2 + 681\nu + 428]. \qquad (13)$$

When $\nu \to \infty$, equation (13) reduces to the classical Bartlett correction for the normal model, given by $c = 1 + \frac{11}{12n}$. The Bartlett correction (13) does not involve any estimation since the null hypothesis specifies the model completely.

3.2. ONE-WAY CLASSIFICATION MODEL

We now consider the analysis of variance normal model. Each observation is classified into one of p different categories, where observations in the i-th category have mean μ_i and precision parameter ϕ_i and we express the linear predictors η_i and τ_i as $\eta_i = \beta + \beta_i$ and $\tau_i = \gamma + \gamma_i$, for each $i = 1, \ldots, p$, where we fix $\beta_p = \gamma_p = 0$. Then, we have the one-way classification structure for both predictors η and τ. For this model, we consider the test of $H_0^1 : \beta_1 = \ldots = \beta_{p-1} = \gamma_1 = \ldots = \gamma_{p-1} = 0$ against H_1^1 : violation of at least one equality. Let n_i be the number of observations in category i, which means that $n_1 + \ldots + n_p = n$. After some algebra we can obtain from (12) and (9)–(11)

$$c_1 = 1 + \frac{11}{12(p-1)}\left\{\sum_{i=1}^{p} n_i^{-1} - n^{-1}\right\}.$$

3.3. SIMPLE LINEAR REGRESSION MODEL

We now consider two-parameter regression models for both predictors η and τ. The systematic components are expressed by $\mu_l = g(\eta_l) = \beta_2 + \beta_1 x_l$ and $\phi_l = h(\tau_l) = \gamma_2 + \gamma_1 z_l$, where x_l and z_l denote the values of the explanatory variables x and z, for $l = 1, \ldots, n$. We consider the test of $H_0^1 : \beta_1 = \gamma_1 = 0$ against $H_1^1 : H_0^1$ is false. We use the notation $\xi_{ij} = n^{-1}\sum_l (x_l - \bar{x})^i (z_l - \bar{z})^j$ for $i, j = 0, \ldots, 4$, where $\bar{x} = n^{-1}\sum_l x_l$ and $\bar{z} = n^{-1}\sum_l z_l$. After some algebra, we obtain the Bartlett correction from (12) and (9)–(11) as

$$c_1 = 1 + \frac{1}{2}\left\{\frac{3(\nu+2)(\nu+3)^2}{2n\nu(\nu+1)(\nu+5)(\nu+7)}\left(2-\frac{\xi_{40}}{\xi_{20}^2}\right) + \frac{3(\nu^2+3\nu-13)^2}{8n(\nu+5)^2(\nu+7)^2\phi^4}\left(2+\frac{\xi_{04}}{\xi_{02}^2}\right)\right.$$

$$+\frac{4(\nu+3)(5\nu^2+8\nu-40)}{6n\nu(\nu+5)^2}\left(3+\frac{\xi_{03}^2}{\xi_{02}^3}\right) + \frac{6(\nu+2)(\nu+3)}{n\nu(\nu+5)(\nu+7)}\left(2+\frac{\xi_{22}}{\xi_{20}\xi_{02}}\right)$$

$$-\frac{2(\nu+2)(\nu+3)(\nu^2+9\nu+8)\phi^3}{n\nu^2(\nu+5)^2}\left(1+\frac{2\xi_{11}^2}{\xi_{20}\xi_{02}}+\frac{\xi_{21}^2}{\xi_{20}^2\xi_{02}}\right)$$

$$\left.+\frac{(\nu+2)^2(\nu+3)}{2n\nu(\nu+5)^2}\left(3+\frac{\xi_{21}^2}{\xi_{20}^2\xi_{02}}\right) + \frac{6(\nu+2)(\nu+3)(\nu+6)}{n\nu(\nu+5)^2}\left(3+\frac{\xi_{21}\xi_{03}}{\xi_{20}\xi_{02}^3}\right)\right\}.$$

4. SIMULATION RESULTS

Lange, Little and Taylor (1989) suggest that the number ν of degrees of freedom in model (1) should be fixed for small data sets and estimated for large ones, since small samples do not typically provide sufficient information about ν for this parameter to be estimated with enough precision. They point out that, for small samples, $\nu = 4$ has worked well for many practical applications. We now show that a known value for ν is not a very strong assumption for the purpose of using equations (8)–(11), since the Bartlett correction is still very effective when ν is replaced by its MLE. Moreover, there is some evidence that the sizes of the modified LR tests are not substantially affected by the degrees of freedom misspecification. In other words, the Bartlett correction substantially improves the accuracy of the type I error rates with either a consistent estimate of ν or the degrees of freedom misspecification.

For inference about ν, we use the profile log-likelihood for ν defined by

$$\ell_p(\nu) = \ell(\nu, \hat{\mu}_\nu, \hat{\sigma}_\nu) = \sup_{\mu,\sigma} \ell(\nu,\mu,\sigma) = \frac{n\nu}{2}\log\nu + n\log\Gamma(\frac{\nu+1}{2}) - n\log\Gamma(\frac{\nu}{2})$$

$$-\frac{(\nu+1)}{2}\sum_{i=1}^{n}\log\left\{\nu+\left(\frac{y_i-\hat{\mu}_i}{\hat{\sigma}_i}\right)^2\right\} - \frac{1}{2}\sum_{i=1}^{n}\log\hat{\sigma}_i^2, \tag{14}$$

where $\hat{\mu}_\nu$ and $\hat{\sigma}_\nu$ denote the restricted MLEs of μ and σ for ν fixed.

The profile likelihood (14) is not a real likelihood function, in that it is not proportional to the model density (1). We use the following two-step algorithm for maximum likelihood estimation of ν, μ and σ. 1) For given ν, (2) is maximized with respect to β and γ, that is, we compute $\hat{\beta}$ and $\hat{\gamma}$ using (3) to find $\hat{\mu}_\nu = g(X\hat{\beta})$ and $\hat{\phi}_\nu = h(Z\hat{\gamma})$. Thus, we compute $\ell_p(\nu)$ from (14). 2) Vary ν over a grid to obtain the estimate $\hat{\nu}$ which maximizes (14). We use a grid of 100 equally spaced ν values in (1.5, 11.5). We hence adopt this algorithm to obtain approximate MLEs of all parameters ν, μ and ϕ in model (1).

We now perform a Monte Carlo simulation study using a t regression model with a three parameter nonlinear regression for the mean $\mu = \beta_1 + \beta_2 x_1 + \exp(\beta_3 x_2)$ and a two parameter regression $\sigma = \exp(\gamma_1 + \gamma_2 z_1)$ for the dispersion parameter to compare the performance of the LR test to those of the Bartlett corrected LR tests. The true number of degrees of freedom were taken as $\nu_T = 4$ and $\nu_T = 6$. We wish to test the null hypothesis $H_0^1 : \beta_3 = \gamma_2 = 0$ against H_1^1: violation of at least one equality. Clearly, the Bartlett correction (12) for this testing problem is a function of the degrees of freedom ν assumed to be known. In models with covariates the way in which they are generated as the sample size increases must be specified. We assume that, as the sample size increases, the covariates design is replicated. The number of observations was set at $n = 20, 30, 40$ and 50 and the number of replications at 10,000. The values for the mean covariates x_1 and x_2 were obtained as random draws from the uniform distributions on the intervals (0,5) and (0,1), respectively. The values for the dispersion covariate z_1 were obtained from a chi-squared distribution with one degree of freedom. The values of all covariates were held constant throughout the simulations with equal sample sizes.

For the size simulations the nuisance parameters were fixed at $\beta_1 = 3, \beta_2 = 2$ and $\gamma_1 = 1$. IMSL subroutines were used to generate the data and the calculations were done using FORTRAN 90 on a Pentium IV at 1.50 GHz. The convergence was not achieved for all the 10,000 simulated samples and then we eliminated those samples for which the iterative fitting method (3) fails to converge before 50 iterations. Such failures were normally most uncommon and the proportion of eliminated samples was smaller than 0.35% in each case. In each simulation, we generate t Student data $y_1, \ldots, y_n$ following (1) and compute the LR statistic and six different forms of Bartlett corrected statistics, namely: the classical Bartlett corrected statistic LR^* assuming the true value ν_T of the degrees of freedom, the Bartlett corrected statistic $\widetilde{LR}$ calculated using the MLE $\hat{\nu}$ of ν obtained from the two-step algorithm and the forms of the Bartlett corrected statistics $LR^{*(1)}$, $LR^{*(2)}$, $LR^{*(3)}$ and $LR^{*(4)}$ when ν is fixed at a wrong value of $\nu_T - 2$, $\nu_T - 1$, $\nu_T + 1$ and $\nu_T + 2$, respectively. Hence, the modified statistic LR^* corresponds to correct degrees of freedom specification whereas the modified statistics $LR^{*(i)}$ for $i = 1, \ldots, 4$ correspond to degrees of freedom misspecification.

We illustrate the various LR tests in this simulation study and then investigate how the sizes of the LR test and its corrected versions are sensitive to misspecification of the degrees of freedom. For doing this, we compute the distributions of the LR statistic and its Bartlett corrected versions at some nominal levels. Under the null hypothesis, we obtain the rejection rates of $LR, LR^*, \widetilde{LR}$ and $LR^{*(i)}$ for $i = 1, \ldots, 4$ at the nominal sizes of 5% and 10% of a χ_2^2 distribution. The percentage of times that these statistics exceed the upper 5% and 10% points of χ_2^2 out of 10,000 simulations are given in Table 1 when $\nu_T = 4$. Table 2 does the same for $\nu_T = 6$.

The figures in Tables 1 and 2 convey important information: (i) the classical LR test is shown to have a finite sample distribution which is poorly approximated by its asymptotic χ_2^2 distribution. In fact, the conventional LR test yields type I error rates that are higher than nominal. The χ_2^2 approximation for all statistics deteriorates when the sample size n decreases; (ii) the empirical sizes of the test based on the modified statistics LR^* and $\widetilde{LR}$ are closer to the nominal levels than the empirical sizes of the unmodified statistic LR. Thus, the Bartlett correction is very effective in pushing the rejection rates of the modified statistics toward to the nominal levels even when ν is replaced by its MLE $\hat{\nu}$ since the estimated sizes of LR^* and $\widetilde{LR}$ are in good agreement. Thus, the Bartlett corrected statistic which uses the MLE of ν seems to offer at least some improvement over the use of the unmodified LR statistic; (*iii*) the empirical sizes of all corrected statistics are closer to the corresponding nominal levels than the empirical sizes of the uncorrected LR statistics. Therefore, the sizes of the modified tests are not substantially affected by the degrees of freedom misspecification, at least for ν in the interval $(\nu_T - 2, \nu_T + 2)$, and that even when ν is uncorrectly specified, the modified statistics lead to some size improvements. In all cases, error rates that are closer to nominal are achieved with the six Bartlett corrected statistics; (*iv*) unsurprisingly the χ^2_2 approximation for the Bartlett corrected statistic deteriorates as the distance between ν and the true value ν_T increases; (*v*) although we have only considered a special t model, we have experimented with other variations of both components for the mean and dispersion parameters and obtained roughly the same conclusions as the ones reported before. We have also experimented with other choices of values for the covariates but they had little impact on the final results.

In summary, the Bartlett correction was shown to be practical, effective and worthwhile techniques for improving small sample inference in the t regression model even when the degrees of freedom is uncorrectly specified.

Table 1: Size simulations ($\nu_T = 4$)

n	LR	LR^*	$\widetilde{LR}$	$LR^{*(1)}$	$LR^{*(2)}$	$LR^{*(3)}$	$LR^{*(4)}$
Nominal level: 10%							
20	19.3	14.7	15.3	16.9	16.2	15.5	16.8
30	18.5	12.9	13.4	14.6	14.3	13.7	14.9
40	15.7	11.8	12.7	13.5	13.1	13.4	14.2
50	13.1	10.7	11.2	12.4	12.0	11.7	12.9
Nominal level: 5%							
20	9.2	7.8	8.1	8.5	8.4	8.3	8.6
30	8.7	6.1	6.7	8.2	8.0	8.1	8.3
40	8.3	5.7	6.4	8.0	7.8	7.7	8.2
50	7.4	5.3	6.1	7.2	7.0	7.1	7.3

Table 2: Size simulations ($\nu_T = 6$)

n	LR	LR^*	$\widetilde{LR}$	$LR^{*(1)}$	$LR^{*(2)}$	$LR^{*(3)}$	$LR^{*(4)}$
			Nominal level: 10%				
20	18.6	13.2	15.0	15.3	14.9	14.4	15.5
30	17.8	12.1	13.2	14.5	14.1	13.5	14.2
40	15.1	10.9	12.8	13.8	12.7	13.2	13.4
50	12.9	10.5	11.4	12.6	11.8	11.4	12.3
			Nominal level: 5%				
20	8.8	7.9	8.0	8.3	8.0	8.2	8.4
30	8.3	6.2	6.3	7.9	7.5	7.7	8.0
40	8.1	5.5	6.2	7.3	6.6	7.2	7.8
50	7.2	5.2	5.8	6.7	6.4	6.6	7.0

Appendix A: Expressions for the Φ_i's

$$\Phi_1 = \mathrm{diag}\left\{\frac{(\nu+1)}{4(\nu+3)}\phi_i g_i''^2 + \frac{6(\nu+1)(\nu+2)}{4\nu(\nu+5)(\nu+7)}\phi_i^2 g_i'^4\right\},$$

$$\Phi_2 = \mathrm{diag}\left\{\frac{(\nu+1)(\nu+2)}{2(\nu+3)(\nu+5)}g_i'^2 h_i'' + \frac{3(\nu+1)(\nu+2)}{(\nu+3)(\nu+5)(\nu+7)}\frac{g_i'^2 h_i'^2}{\phi_i}\right\},$$

$$\Phi_3 = \mathrm{diag}\left\{\frac{\nu}{8(\nu+3)}\frac{h_i''^2}{\phi_i^2} + \frac{\nu(\nu-4)}{2(\nu+3)(\nu+5)}\frac{h_i'^2 h_i''}{\phi_i^3} - \frac{3\nu(\nu^2+3\nu-13)}{4(\nu+3)(\nu+5)(\nu+7)}\frac{h_i'^4}{\phi_i^4}\right\},$$

$$\Phi_4 = \frac{(\nu+1)}{(\nu+3)}\mathrm{diag}\{\phi_i g_i' g_i''\},\ \Phi_5 = \frac{(\nu+1)(\nu+2)}{(\nu+3)(\nu+5)}\mathrm{diag}\left\{h_i' g_i'^2\right\},\ \Phi_6 = \frac{(\nu+1)}{(\nu+3)}\mathrm{diag}\left\{g_i'^2 h_i'\right\},$$

$$\Phi_7 = \frac{\nu}{(\nu+3)}\mathrm{diag}\left\{\frac{(\nu+8)}{(\nu+5)}\left(\frac{h_i'}{\phi_i}\right)^3 - \frac{3h_i' h_i''}{2\phi_i^2}\right\}, \quad \Phi_8 = \frac{\nu}{(\nu+3)}\mathrm{diag}\left\{\left(\frac{h_i'}{\phi_i}\right)^3 - \frac{h_i' h_i''}{\phi_i^2}\right\}.$$

References

Barndorff-Nielsen, O.E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika*, **75**, 374–378.

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

Botter, D.A. and Cordeiro, G.M. (1997). Bartlett corrections for generalized linear models with dispersion covariates, *Communications in Statistics, Theory and Methods*, **26**, 279–307.

Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalized linear models, *Journal of the Royal Statistical Society B*, **45**, 404–413.

Cordeiro, G.M. (1987). On the corrections to the likelihood ratio statistics, *Biometrika*, **74**, 265–274.

Cordeiro, G.M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models, *Communications in Statistics, Theory and Methods*, **22**, 169–188.

Cordeiro, G.M. and Paula, G.A. (1989a). Improved likelihood ratio statistics for exponential family nonlinear models, *Biometrika*, **76**, 93–100.

Cordeiro, G.M. and Paula, G.A. (1989b). Fitting non–exponential family nonlinear models in GLIM by using the offset facility, *Lecture Notes in Statistics* **57**, 105-114, in: Decarli A. and B. J. Francis (eds.), *Proceedings of GLIM 89 and the 4th Workshop on Statistical Modelling,* Springer–Verlag, New York.

Cordeiro, G.M., Paula, G.A. and Botter, D.A. (1994). Improved likelihood ratio tests for dispersion models, *International Statistical Review*, **62**, 257–276.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics* (Chapman and Hall, London, 2nd ed.).

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society B*, **49**, 1–39.

DiCiccio, T.J. and Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods, *Journal of the Royal Statistical Society B*, **56**, 397–408.

Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation and some robust ans resistant alternatives (with discussion), *Journal of the Royal Statistical Society B*, **46**, 146–192.

Jørgensen, B. (1984). Exponential dispersion models, *Journal of the Royal Statistical Society B*, **49**, 127–162.

Lange, K., Little, R. and Taylor, J. (1989). Robust statistical modeling using the t distribution, *Journal of the American Statistical Association*, **84**, 881–896.

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

Porteous, B.T. (1985). Improved likelihood ratio statistics for covariance selection models, *Biometrika*, **72**, 97–101.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2002-01 - BUENO, V.C. Asymptotic reliability for a coherent system with an infinite number of components. 2002. 10p. (RT-MAE-2002-01)

2002-02 - TAVARES, H.R., ANDRADE, D.F. Item Response Theory for Longitudinal Data: Item and Population Ability Parameters Estimation. 2002. 20p. (RT-MAE-2002-02)

2002-03 - BOLFARINE, H., VALENÇA, D.M. Score tests for Weibull-regression models with random effect. 2002. 16P. (RT-MAE-2002-03)

2002-04 - BUENO, V.C. Using a second order stochastic dominance for active redundancy allocation in a K-out-of-n system. 2002. 7p. (RT-MAE-2002-4)

2002-05 - FERRARI, S.L.P., LUCAMBIO, F., CRIBARI-NETO, F. Adjusted and Bartlett-Adjusted profile likelihood ratio tests. 2002. 19p. (RT-MAE-2002-05).

2002-06 - PEREIRA, C.A.B., STERN, J.M. Full Bayesian Significance Test: Invariant Formulation. 2002. 10p. (RT-MAE-2002-06).

2002-07 - BUENO, V.C. Optimal arrangement of components for systems of dependent components. 2002. 13p. (RT-MAE-2002-07)

2002-08- KOLEV, N., PAIVA, D. Multinomial Latent Model for Random Sums. 2002. 9p. (RT-MAE-2002-08)

2002-09 - BUENO, V.C. A Note on a Hazard Processes Ordering: Application on a *K-out-of-N System*. 2002. 9P. (RT-MAE-2002-09)

2002-10 - FERRARI, S.L.P., DAVID, J.S.E., ANDRÉ, A.P., PEREIRA, L.A.A. Use of overdispersed regression models in analysing the association between air pollution and human health. 2002. 15p. (RT-MAE-2002-10)

2002-11 - PERES, A.A.Q., BOTTER, D.A.B. Association between air pollution and human health: filling in the gaps of the pollutant series. 2002. 13P. (RT-MAE-2002-11)

2002-12 - CYSNEIROS, F.J.A., PAULA, A.G. One-Sided Tests in Linear Models with Multivariate t-Distribution. 2002. 33p. (RT-MAE-2002-12)

2002-13 - CRIBARI-NETO, F., FERRARI, S.L.P., OLIVEIRA, W.A.S.C. A numerical evaluation of tests based on different heteroskedasticity-consistent covariance matrix estimators. 2002. 19p. (RT-MAE-2002-13)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil