RT-MAE 9717

SOME ADJUSTED LIKELIHOOD RATIO TESTS FOR HETEROSCEDASTIC REGRESSION MODELS

by

Elisete C. Q. Aubin and Gauss M. Cordeiro

Palavras-Chave: Bartlett correction; chi-squared distribution; heteroscedastic model; likelihood ratio test; maximum likelihood.

Classificação AMS: 62E20, 62F05. (AMS Classification)

Some adjusted likelihood ratio tests for heteroscedastic regression models

ELISETE C. Q. AUBIN

Departamento de Estatística, Universidade de São Paulo Caixa Postal 66281, Agência Cidade de São Paulo São Paulo SP, 05389-970, Brazil

GAUSS M. CORDEIRO

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

Abstract

In this paper, we present Bartlett corrections to improve likelihood ratio tests for heteroscedastic regression models, thus generalizing previous results by Cordeiro (1993a). The formulae are simple enough to be used algebrically to obtain closed-form expressions in special cases. They also have advantages for numerical purposes. We give applications to some special models and discuss improved likelihood ratio tests. We also use Monte Carlo simulation to confirm the superiority of the Bartlett corrected statistics over the classical likelihood ratio statistics with regard to second-order asymptotic theory.

Key words: Bartlett correction; chi-squared distribution; heteroscedastic model; likelihood ratio test; maximum likelihood.

1 Introduction

Consider a normal heteroscedastic model for observable data $Y(n \times 1)$:

$$E(Y) = \mu = X\beta,$$

 $\sigma_l^2 = \text{Var}(Y_l) = H(w_l^\top \gamma),$ (1)

for $l=1,\ldots,n$. In model (1), $X=(x_1,\ldots,x_n)^{\top}$ is a known $n\times p$ design matrix of full rank $p, x_l^{\top}=(x_{l1},\ldots,x_{lp})$ for $l=1,\ldots,n$ are row vectors, $\beta(p\times 1)$ is the regression parameter,

 $w_l^{\mathsf{T}}(1\times q)$ is a vector of known constants for $l=1,\ldots,n,$ $H(\cdot)$ is smooth, known and assumed differentiable function up to third order and $\gamma(q\times 1)$ is an unknown vector. Let $W=(w_1,\ldots,w_n)^{\mathsf{T}}$ be an $n\times q$ matrix. As foundation for hypothesis tests we assume that the components of Y are uncorrelated with mean μ and diagonal covariance matrix $\Lambda=\mathrm{diag}\{\sigma_l^2\}$. The components of β and γ are unrelated and vary independently. We are interested in estimating both parameters in β and γ and in testing some components of these vectors. The variance function $H(\cdot)$ is defined such that it maps the real line to the positive real line. Natural choices for $H(\cdot)$ could be $\sigma_l^2=\exp(w_l^{\mathsf{T}}\gamma)$ and $\sigma_l^2=(w_l^{\mathsf{T}}\gamma)^2$.

The exponential form for the variance function is quite useful in fields including engineering, economics, and the biological and physical sciences. A special case corresponds to take $w_l^{\mathsf{T}} = (1, \log x_{l2}, \dots, \log x_{lq})$, where x_{lj} are known constants strictly positive, which gives $\sigma_l^2 = \sigma^2 \prod_{i=2}^q z_{li}^{\delta_i}$ with $\sigma^2 = \log \delta_1$. Thus, the test concerning homocedasticity is equivalent to the test $\delta_2 = \dots = \delta_q = 0$. Similar tests for homocedasticity have been proposed by Goldfeld and Quandt (1965), Glejser (1969), Harrison and McCabe (1979), Jarque (1981), Koenker and Bassett (1982) and Cook and Weisberg (1983). However, there are other forms for $H(\cdot)$ quite useful in applications (Box and Hill, 1974).

The likelihood ratio (LR) statistic w is frequently used to test hypothesis of interest in regression models. Under the null hypothesis H_0 , w has an approximate χ^2 distribution with degrees of freedom given by the difference of the dimensions of the parameter spaces under the two hypotheses tested. Generally speaking, the main difficulty of testing a null hypothesis using the log-likelihood ratio criterion lies not so much in deriving its closed-form expression – when it has one – but in finding its exact distribution, or at least a good approximation, when the null hypothesis is true.

In an influential paper, Bartlett (1937) proposed an improved LR statistic. His argument goes as follows. Suppose that under the null hypothesis $E(w) = q\{1+b+o(n^{-3/2})\}$, where b is a constant that can be consistently estimated under H_0 . Then, the expected

value of the transformed statistic $w^* = w/(1+b)$ is closer to the one from a χ_q^2 distribution than the expected value of w. This became widely known as the Bartlett correction. Furthermore, it was shown by Lawley (1956) that all cumulants of w^* agree with those of the reference chi-squared distribution with error of order n^{-2} . Lawley's results appeared to be incompatible with the asymptotic expansion of the null distribution of the LR statistic obtained by Hayakawa (1977). This puzzle was solved by Cordeiro (1987) who has shown that a coefficient in Hayakawa's expansion is always equal to zero.

In recent years there has been a renewed interest in Bartlett corrections. Cordeiro (1983, 1987) derived closed-form expressions for Bartlett correction factors in generalized linear models (Nelder and Wedderbun, 1972) and discussed improved likelihood ratio goodness-of-fit tests. Williams' (1976) results are a special case of Cordeiro's results. Cordeiro (1995) presents extensive simulation results on the performance of a Bartlett-corrected deviance in generalized liner models focusing on gamma and log-linear models. Attifield (1995) focused on models that involve systems of equations, and derived Bartlett corrections to the log-likelihood ratio statistic in this case. Bartlett corrections for models defined by any one-parameter distribution in which the mean is a known function of a linear combination of unknown parameters were obtained by Cordeiro (1985), who generalized his own results of 1983. Further Bartlett adjustments for ten multivariate normal testing problems concerning structured covariance matrices were obtained by Møller (1986). In particular, Møller's results apply to real, complex and quaternion Wishart distributions and cover a number of tests.

Many recent papers have focused on deriving closed-form Bartlett corrections for specific regression problems. For example, Moulton, Weissfeld and St. Laurent (1993) have obtained Bartlett corrections for logistic regressions; Cordeiro, Paula and Botter (1994) have derived corrections for the class of dispersion models proposed by Jørgensen (1987); Attfield (1991) and Cordeiro (1993a) have shown how to correct LR tests for heteroskedas-

ticity; Wong (1991) has obtained a Bartlett correction factor for testing several slopes in regression models whose independent variables are subject to error; Wang (1994) derived the correction factor for testing the equality of normal variances against an increasing alternative; and Chesher and Smith (1995) have obtained Bartlett corrections for LR specification tests. A correction to the log-likelihood ratio statistic in regression models with Student-t errors was obtained by Ferrari and Arellano-Valle (1993), and similar corrections to heteroskedastic linear models and multivariate regression were obtained by Cribari-Neto and Ferrari (1995) and Cribari-Neto and Zarkos (1995), respectively. Bartlett adjustments to log-likelihood tests of the unit root hypothesis were proposed by Larsson (1994) and Nielsen (1995). An algorithm for computing Bartlett corrections was given by Jensen (1993); see also Andrews and Stafford (1993) and Stafford and Andrews (1993). General Matrix formulae for computing Bartlett corrections are given by Cordeiro (1993b). A review on Bartlett corrections and some extensions is presented in Cribari-Neto and Cordeiro (1996). An important non-regression case is that of one-parameter exponential family models. A simple, closed-form Bartlett correction for testing the null hypothesis that the parameter that indexes such models equals a given scalar was obtained by Cordeiro, Cribari-Neto, Aubin and Ferrari (1995). They then applied their result to a number of distributions in the exponential family, some of which are widely used in empirical applications in a variety of fields. A Bartlett correction for the natural exponential family had been previouly given by McCullagh and Cox (1986).

The purpose of this paper is to obtain Bartlett corrections for several hypothesis tests derived from model (1). The remainder of the paper is organized as follows. Section 2 presents a general formula for the expected LR statistic in model (1). This formula has advantage for numerical purposes because it requires only simple operations on matrices. It is also simple enough to obtain several closed-form Bartlett corrections in a variety of important tests. This formula generalizes Cordeiro (1993a) equations (3.6) – (3.8).

In Section 3, we discuss several special cases of interest in practical applications. Tests involving a subset of the β and γ parameters are considered in Section 4. Finally, in Section 5, we present some simulation results which show that the Bartlett corrections derived are really of practical importance.

2 A general formula for Bartlett corrections

For the model (1) the total log-likelihood for $\theta = (\beta^{\top}, \gamma^{\top})^{\top}$ given Y can be written as

$$\ell(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{l=1}^{n} \frac{v_l}{H_l},$$
(2)

where we denote, from now on, $\rho_l = w_l^{\mathsf{T}} \gamma$, $\sigma_l^2 = H(\rho_l) = H_l$ and $H_l^{(r)} = H^{(r)}(\rho_l) = \partial^r H(\rho_l)/\partial \rho_l^r$ for $r = 1, 2, \ldots$, and $v = (v_1, \ldots, v_n)^{\mathsf{T}}$ as an $n \times 1$ vector with $v_l = (Y_l - x_l^{\mathsf{T}} \beta)^2$. We make some assumptions (Cox and Hinkley, 1974; Chapter 9) on the behavior of $\ell(\theta)$ as the sample size n approaches infinity, such as the regularity of the first four θ derivatives of $\ell(\theta)$ and the uniqueness of the relevant MLE $\widehat{\theta} = (\widehat{\beta}^{\mathsf{T}}, \widehat{\gamma}^{\mathsf{T}})^{\mathsf{T}}$.

Fisher's scoring method is used to obtain the MLEs $\hat{\beta}$ and $\hat{\gamma}$ iteractively from the equations

$$\hat{\beta} = (X^{\top} \hat{\Lambda}^{-1} X)^{-1} X^{\top} \hat{\Lambda}^{-1} Y,$$

$$W^{\top} \hat{P} 1 = W^{\top} \hat{P} \hat{\Lambda}^{-1} \hat{v},$$
(3)

where 1 is an $n \times 1$ vector of ones, $P = \operatorname{diag}\{p_{II}\} = \Lambda^{(1)}\Lambda^{-1}$, $\Lambda = \operatorname{diag}\{H_{I}\}$ and $\Lambda^{(r)} = \operatorname{diag}\{H_{I}^{(r)}\}$. The parameters β and γ are globally orthogonal (Cox and Reid, 1987) and therefore the joint information matrix for these parameters is block-diagonal $K = \operatorname{diag}\{K_{\beta}, K_{\gamma}\}$, where $K_{\beta} = X^{\top}\Lambda^{-1}X$ is the information for β and $K_{\gamma} = \frac{1}{2}W^{\top}P^{2}W$ is the information for γ . Further, the MLEs $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent and converge in distribution to multivariate normal distributions $N_{p}(\beta, K_{\beta}^{-1})$ and $N_{q}(\gamma, K_{\gamma}^{-1})$, respectively.

The following standard notation (Lawley, 1950) for the joint cumulants of the loglikelihood derivatives, all assumed to be O(n), will be adopted:

$$\begin{split} \kappa_{r,s} &= E \left\{ \frac{\partial \ell(\theta)}{\partial \theta_r} \frac{\partial \ell(\theta)}{\partial \theta_s} \right\}, \quad \kappa_{rs} = E \left\{ \frac{\partial^2 \ell(\theta)}{\partial \theta_r \partial \theta_s} \right\}, \\ \kappa_{rs,t} &= E \left[\left\{ \frac{\partial^2 \ell(\theta)}{\partial \theta_r \partial \theta_s} \right\} \frac{\partial \ell(\theta)}{\partial \theta_t} \right], \quad \kappa_{rst} = E \left\{ \frac{\partial^3 \ell(\theta)}{\partial \theta_r \partial \theta_s \partial \theta_t} \right\}, \end{split}$$

and so on, where the components of θ may represent any components of β and γ . We also use the notation $\kappa_{rs}^{(t)} = \frac{\partial \kappa_{rs}}{\partial \theta_t}$, $\kappa_{rs}^{(tu)} = \frac{\partial^2 \kappa_{rs}}{\partial \theta_t \partial \theta_u}$, etc. Note that $\kappa_{r,s} = -\kappa_{rs}$ is the (r,s)th element of the Fisher information matrix for θ . Let $-\kappa^{rs}$ be the corresponding element of its inverse. We denote the summation over all the β and γ parameters by Σ' and over the data by Σ . For models given by (1), we can find these joint cumulants without difficulty.

We define the total maximized log-likelihood by $\hat{\ell}_{p+q} = \ell(\hat{\theta})$ and the log-likelihood at the true parameter point by ℓ . From Lawley's (1956) expansion we can write the expected value of $2(\hat{\ell}_{p+q} - \ell)$ to order n^{-1} as $2E(\hat{\ell}_{p+q} - \ell) = p + q + \epsilon_{p+q}$, where ϵ_{p+q} is a term of order n^{-1} evaluated at the true parameter given by

$$\epsilon_{p+q} = \sum_{\beta,\gamma} '(\ell_{rstu} - \ell_{rstuvw}),$$
 (4)

where

$$\ell_{rstu} = \kappa^{rs} \kappa^{tu} \left(\frac{1}{4} \kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right), \tag{5}$$

$$\ell_{rstuww} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left[\kappa_{rtv} \left(\frac{1}{6} \kappa_{suw} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{1}{4} \kappa_{suw} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right]. \quad (6)$$

Expression (4) is general enough to be used in a number of econometric models since it is usually obtained from likelihood functions that obey the general regularity conditions stated in Cox and Hinkley (1974, Chapter 9), thus allowing one to handle independent, but not necessarily identically distributed observations. The main problem of Lawley's formula is its interpretation since its individual terms are not parameter invariant. However, it can be widely used by econometricians when programmed in an algebraic manipulation language, such as Mathematica (Wolfram, 1991).

The proof of (4) – (6) given in Lawley contains many references to the difficulty of the symbolic manipulations and, as he pointed out, it involves exceedingly complicated and laborious algebra.

We now obtain ϵ_{p+q} in a more readily computable form. The key step in deriving a simple expression of ϵ_{p+q} is to use the block diagonality of the information matrix. In view of this, the sums $\Sigma'\ell_{rstu}$ and $\Sigma'\ell_{rstuvw}$ can be obtained in matrix notation by summing over the sample after evaluating the sums over the parameters. We define $A = W(W^{\top}P^2W)^{-1}W^{\top}$ and $B = X(X^{\top}\Lambda^{-1}X)^{-1}X^{\top}$ and the diagonal matrices $A_d = \text{diag}(A)$ and $B_d = \text{diag}(B)$. Note that 2A and B are just the large sample covarinces matrices of $W\hat{\gamma}$ and $\hat{\mu}$, respectively. Plugging the joint cumulants into (5) and (6) we can find after a lengthy algebra

$$\sum_{\beta,\gamma}{}'\ell_{rstu} = \text{tr}\left[A_dB_d\Lambda^{-1}(2P^2 - Q) - \frac{1}{2}A_d^2(-Q^2 + 8P^2Q - 6P^4)\right]$$

and

$$\begin{split} \sum_{\beta,\gamma}{}'\ell_{rstuvw} &= -1^{\top}PA_{d}AB_{d}\Lambda^{-1}P1_{n} - \frac{1}{2}1^{\top}P\Lambda^{-1}B_{d}AB_{d}\Lambda^{-1}P1 \\ &+ 1^{\top}P\Lambda^{-1}(B^{(2)}\otimes A)\Lambda^{-1}P1 - \frac{1}{2}1^{\top}PQA_{d}AA_{d}QP1 \\ &- \frac{1}{3}1^{\top}P[4P^{2}A^{(3)}(3Q - 2P^{2}) - 3QA^{(3)}Q]P1, \end{split}$$

where \otimes denotes the Hadamard (direct) product, $B^{(2)} = B \otimes B$, $A^{(3)} = A^{(2)} \otimes A$, etc. and $Q = \Lambda^{(2)} \Lambda^{-1}$.

The details involved in deriving the expressions above for $\sum_{\beta,\gamma}^{\prime}\ell_{rstu}$ and $\sum_{\beta,\gamma}^{\prime}\ell_{rstuve}$ are tedious and are not reproduced here but follow from singular algebraic developments of Cordeiro (1993a). They can be obtained from the authors. After simple manipulation we get the following decomposition for ϵ_{p+q}

$$\epsilon_{p+q} = h(X, W, \Lambda) = f(W, \Lambda) + g(X, W, \Lambda),$$
 (7)

where

$$f(W, \Lambda) = -\frac{1}{2} \text{tr} \left[A_d^2 (-Q^2 + 8P^2 Q - 6P^4) \right] + \frac{1}{3} 1^{\mathsf{T}} P \left[4P^2 A^{(3)} (3Q - 2P^2) - 3QA^{(3)} Q \right] P 1 + \frac{1}{2} 1^{\mathsf{T}} P Q A_d A A_d Q P 1$$
(8)

and

$$g(X, W, \Lambda) = \operatorname{tr} \left[A_d B_d \Lambda^{-1} (2P^2 - Q) \right] - 1^{\mathsf{T}} P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1$$

$$+ \frac{1}{2} 1^{\mathsf{T}} P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1 + 1^{\mathsf{T}} P A_d A B_d \Lambda^{-1} P 1.$$
(9)

Clearly ϵ_{p+q} is a function of X, W and Λ and it only involves simple operations on matrices and vectors. Equations (7) – (9) can be easily computed using a computer algebra system such as Mathematica or Maple, or using a programming language with support for matrix operations, such as GAUSS, OX or S-PLUS. It is also possible to simplify equations (8) and (9) when the model at hand has closed-form expressions for A and B; see Sections 3 and 4. We have checked formula (7) in several specific situations and it works properly. Although $h(X, W, \Lambda)$ is easy to compute under the null hypothesis, it is not easy to interpret. The fundamental difficulty is that the individual terms in equations (8) and (9) are not invariant under reparametrization and therefore their interpretation depend on the coordinate system chosen. The entire expression for $h(X, W, \Lambda)$ is of course invariant under reparametrization. Unfortunately, equations (8) and (9) provide no indication as to what structural aspects of the model contribute significantly to their magnitude.

3 Some special cases

In this section we consider a number of special models which produce some simplification in equations (7) – (9). Let us consider first the exponential variance function defined by $\sigma_l^2 = \exp(w_l^\top \gamma)$. In this case, P = Q = I, where I is the identity matrix of order

n, and then the formulae (7) – (9) reduce to the corresponding expressions (3.6) – (3.8) of Cordeiro (1993a). As a second case, we consider the homoscedastic model for which q=1, W=1 and $\Lambda=\sigma^2 I$. After some algebra, it comes from (7)

$$\epsilon_{p+1} = h(X, 1, \sigma^2 I) = \frac{3(p+1)^2 - 1}{6n} - \frac{(\delta - 1)p}{n\tau^2},$$
 (10)

where $\tau = H^{(1)}(\gamma)/H(\gamma)$ and $\delta = H^{(2)}(\gamma)/H(\gamma)$. When $H(\cdot) = \exp(\cdot)$, $\delta = 1$ and then formula (10) is identical to equation (4.1) of Cordeiro (1993a).

As our third example, we consider the case of homogeneity of means for which p = 1 and X = 1. We can find

$$\begin{split} \epsilon_{1+q} &= h(1,W,\Lambda) = \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \mathrm{tr} \left[A_d \Lambda^{-1} (2P^2 - Q) \right] - \frac{1}{2} \mathrm{tr} \left[A_d^2 (-Q^2 + 8P^2Q - 6P^4) \right] \\ &+ \frac{1}{\sum_{l=1}^n \sigma_l^{-2}} \mathbf{1}^\top P A_d A \Lambda^{-1} P \mathbf{1} - \frac{1}{2 \left(\sum_{l=1}^n \sigma_l^{-2} \right)^2} \mathbf{1}^\top P \Lambda^{-1} A \Lambda^{-1} P \mathbf{1} \\ &+ \frac{1}{2} \mathbf{1}^\top P Q A_d A A_d Q P \mathbf{1} + \frac{1}{3} \mathbf{1}^\top P \left[4P^2 A^{(3)} (3Q - 2P^2) - 3Q A^{(3)} Q \right] P \mathbf{1}. \end{split}$$

When $H(\cdot) = \exp(\cdot)$ this formula is in agreement with the equations (3.6) - (3.8) of Cordeiro (1993a) by noting that $B = 11^{\top} / \sum_{l=1}^{n} \sigma_{l}^{-2}$, $B_{d} = I / \sum_{l=1}^{n} \sigma_{l}^{-2}$ and $B^{(2)} \otimes A = A / (\sum_{l=1}^{n} \sigma_{l}^{-2})^{2}$.

Finally, as a fourth example, we consider the one-way classification structure for both predictors μ and ρ . In this case, we assume that $p \geq 2$ populations have distributions $N(\mu_1, \sigma_1^2), \ldots, N(\mu_p, \sigma_p^2)$ and that independent random samples of sizes n_1, \ldots, n_p ($x_i \geq 1$, $i=1,\ldots,p$) are drawn from these populations. Here the two systematic components are written as $\mu_j = \beta + \beta_j$ and $\sigma_j^2 = H(\gamma + \gamma_j)$ for $j=1,\ldots,p$, where $\sum_{j=1}^p \beta_j = \sum_{j=1}^p \gamma_j = 0$, β and γ are scalars representing mean effects and β_j and γ_j are the effects on the mean and on the variance of the response due to the jth population. In this case, the $p \times p$ matrices $W^\top P^2 W$ and $X^\top \Lambda^{-1} X$ reduce to $W^\top P^2 W = \text{diag}\{n_j p_{jj}^2\}$ and $X^\top \Lambda^{-1} X = \text{diag}\{n_j \sigma_j^{-2}\}$. Recall that $P = \text{diag}\{p_{jj}\} = \Lambda^{(1)} \Lambda^{-1}$ and $Q = \text{diag}\{q_{jj}\} = \Lambda^{(2)} \Lambda^{-1}$. Further, the matrices A and B have typical elements given by $A = \{\delta_{lm}/n_l p_u^2\}$ and $B = \{\delta_{lm}/n_l \sigma_l^2\}$,

where $\delta_{lm} = 1$ if l and m index observations in the same population and zero otherwise. We also have

so have
$$A = \begin{pmatrix} \begin{pmatrix} n_1^{-1}p_{11}^{-2} & \cdots & n_1^{-1}p_{11}^{-2} \\ \vdots & & \vdots \\ n_1^{-1}p_{11}^{-2} & \cdots & n_1^{-1}p_{11}^{-2} \end{pmatrix}_{n_1 \times n_1} & 0 \\ & & & \ddots & \\ & & & \begin{pmatrix} n_1^{-1}p_{p2}^{-2} & \cdots & n_p^{-1}p_{pp}^{-2} \\ \vdots & & & \vdots \\ n_p^{-1}p_{pp}^{-2} & \cdots & n_p^{-1}p_{pp}^{-2} \end{pmatrix}_{n_p \times n_p} \end{pmatrix}$$

and

After some algebra we can find

$$\epsilon_{2p} = \sum_{j=1}^{p} n_j^{-1} \left(\frac{11}{6} + \frac{1 - q_{jj}}{p_{jj}^2} \right).$$

Evidently, the Bartlett correction for this model does depend only on the sample sizes and on the variance function and its first two derivatives.

Bartlett corrections for testing some hypotheses 4

In this section we deal with five important composite null hypotheses defined from model (1). The first two natural composite hypotheses are $H_1: \delta_2 = \cdots = \delta_q = 0$ with $w_{l1} = 1$ for $l=1,\ldots,n$, and H_2 : $\beta_2=\cdots=\beta_p=0$ with $x_{l1}=1$ for $l=1,\ldots,n$, both hypotheses to be tested against A: violation of at least one equality. The problem of H_1 of testing for homoscedasticity when the variances depend on predictors is one with a long history; see, for example, Cook and Weisberg (1983) as well as almost any modern

econometric textbook. The hypothesis H_2 corresponds to the test of homogeneity of means, γ being a nuisance parameter. From now on we denote the unrestricted MLEs obtained from equations (3) by $\hat{\beta}$ and $\hat{\gamma}$ while the restricted estimates are denoted by $\tilde{\beta}$ and $\tilde{\gamma}$. Functions evaluated at the unrestricted estimates will be denoted by the addition of a circumflex and those evaluated at the restricted estimates by the addition of a \sim .

The LR statistic for testing H_1 with β as a nuisance parameter reduces to

$$w_1 = n\left(\log \tilde{\sigma}^2 + 1\right) - \sum_{l=1}^n \log \hat{\sigma}_l^2 - \sum_{l=1}^n \frac{\hat{v}_l}{\hat{\sigma}_l^2},$$

where the restricted estimates come from

$$\widetilde{\beta} = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}Y$$
 and $\widetilde{\sigma}^2 = \frac{1}{n}1^{\mathsf{T}}\widetilde{v}$,

with $\tilde{v} = (\tilde{v}_1, \dots, \tilde{v}_n)^{\top}$ and $\tilde{v}_l = (Y_l - x_l^{\top} \tilde{\beta})^2$ for $l = 1, \dots, n$. The statitic w_1 is asymptotically distributed under H_1 as χ^2_{q-1} . This approximation has error $O(n^{-1})$ and it can be improved to $O(n^{-2})$ by a Bartlett correction which we now derive. We obtain from (7) and (10) under H_1

$$E(w_1) = q - 1 + h(X, W, \sigma^2 I) - \left[\frac{3(p+1)^2 - 1}{6n} - \frac{(\delta - 1)p}{n\tau^2} \right]$$

which gives the Bartlett correction $c_1 = (q-1)^{-1}E(w_1)$ for improving the test of H_1 . We can easily see by calculating (7) under H_1 that c_1 does not involve unknown parameters and depends only on the model matrices X and W.

For the test H_2 of homogeneity of means, γ being a nuisance parameter, the LR criterion becomes

$$w_2 = \sum_{l=1}^n \log \left(\frac{\tilde{\sigma}_l^2}{\hat{\sigma}_l^2} \right) + 1^{\mathsf{T}} (\tilde{\Lambda}^{-1} \tilde{v} - \hat{\Lambda}^{-1} \hat{v}),$$

where $\tilde{\Lambda}$ and \tilde{v} are the MLEs under H_2 obtained from the equations

$$\widetilde{\beta}_1 = \frac{1^\top \widetilde{\Lambda}^{-1} Y}{1^\top \widetilde{\Lambda}^{-1} 1} \quad \text{and} \quad W^\top \widetilde{P} 1 = W^\top \widetilde{P} \widetilde{\Lambda}^{-1} \widetilde{v}.$$

Under H_2 and up to order n^{-1} we find

$$E(w_2) = p - 1 + h(X, W, \Lambda) - h(1, W, \Lambda), \tag{11}$$

where the h functions are easily obtained from (7). We also have

$$h(X, W, \Lambda) - h(1, W, \Lambda) = \operatorname{tr} \left\{ A_d B_d \Lambda^{-1} (2P^2 - Q) \right\} - \frac{1}{\sum_{i=1}^n \sigma_i^{-2}} \operatorname{tr} \left\{ A_d \Lambda^{-1} (2P^2 - Q) \right\}$$

$$+ 1^{\mathsf{T}} P A_d A B_d \Lambda^{-1} P 1 - \frac{1}{\sum_{i=1}^n \sigma_i^{-2}} 1^{\mathsf{T}} P A_d A \Lambda^{-1} P 1$$

$$+ \frac{1}{2} 1^{\mathsf{T}} P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1 + \frac{1}{2 \sum_{i=1}^n \sigma_i^{-2}} 1^{\mathsf{T}} P \Lambda^{-1} A \Lambda^{-1} P 1$$

$$- 1^{\mathsf{T}} P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1.$$

The Bartlett correction $c_2 = (p-1)^{-1}E(w_2)$ is estimated at $\tilde{\Lambda}$ by \tilde{c}_2 and then the improved test of H_2 compares $\tilde{c}_2^{-1}w_2$ with the χ^2_{p-1} distribution. In the particular case q=1, equation (11) simplifies to

$$E(w_2) = p - 1 + \frac{(p+1)^2 - 4}{2n} - \frac{(p-1)(\delta - 1)}{n\tau^2}.$$
 (12)

Formula (12) is in agreement with expression (4.3) of Cordeiro (1993a) for the special case $H(\cdot) = \exp(\cdot)$.

In many problems the restrictions under a test involve a subset of the β and γ parameters. Partitioning the parameters as $\beta = (\beta_1^\top, \beta_2^\top)^\top$ and $\gamma = (\gamma_1^\top, \gamma_2^\top)^\top$, where $\beta_1 = (\beta_1, \dots, \beta_{p_1})^\top$, $\beta_2 = (\beta_{p_1+1}, \dots, \beta_p)^\top$, $\gamma_1 = (\gamma_1, \dots, \gamma_{q_1})^\top$ and $\gamma_2 = (\gamma_{q_1+1}, \dots, \gamma_q)^\top$, we are interested in testing $H_3: \beta_3 = \beta_2^{(0)}$, $\gamma_2 = \gamma_2^{(0)}$ against A, where $\beta_2^{(0)}$ and $\gamma_2^{(0)}$ are specified vectors of dimensions $p - p_1$ and $q - q_1$, respectively. Further, we assume that $1 \leq p_1 \leq p$ and $1 \leq q_1 \leq q$ but the trivial case $p_1 = p$, $q_1 = q$ is excluded because there are no parameters left under the null hypothesis. The contiguous cases $p_1 = 0$, $1 \leq q_1 \leq q$ and $1 \leq p_1 \leq p$, $q_1 = 0$ corresponding to the composite hypotheses $\beta = \beta^{(0)}$, $\gamma_2 = \gamma_2^{(0)}$ and $\beta_2 = \beta_2^{(0)}$, $\gamma_1 = \gamma_2^{(0)}$ can be treated separately. Evidently, in the simplest case $p_1 = q_1 = 0$ of no practical interest since H_3 becomes simple, the Bartlett correction reduces to $1 + (p + q)^{-1}h(X, W, \Lambda)$ with Λ being evaluated under the null hypothesis.

Following the partition induced by H_3 , let $X = (X_1, X_2)$ and $W = (W_1, W_2)$ be the corresponding partitioned model matrices, where X_1 , X_2 , W_1 and W_2 are, respectively, $n \times p_1$, $n \times (p - p_1)$, $n \times q_1$ and $n \times (q - q_1)$ known matrices of full ranks. The LR criterion for testing H_3 comes from (2)

$$w_3 = \log \left\{ \frac{|\widetilde{\Lambda}|}{|\widehat{\Lambda}|} \right\} + 1^{\top} \left[\widetilde{\Lambda}^{-1} \widetilde{v} - \widehat{\Lambda}^{-1} \widehat{v} \right],$$

where MLEs subject to H_3 are written with a tilde. The Bartlett correction c_3 for improving H_3 follows from (7) as

$$c_3 = 1 + \frac{h(X, W, \Lambda) - h(X_1, W_1, \Lambda)}{p - p_1 + q - q_1},$$
(13)

and it should be estimated at $\tilde{\Lambda}$. Then, the improved test compares $\tilde{c_3}^1w_3$ with the upper point of the $\chi^2_{p-p_1+q-q_1}$ distribution.

We are now interested in testing mean effects, namely $H_4: \beta = \beta^{(0)}$ against $A: \beta \neq \beta^{(0)}$, where $\beta^{(0)}$ is a specified whole vector of dimension p and γ is a vector of nuisance parameters. The LR statistic for testing H_4 is given by

$$w_4 = \sum_{l=1}^n \log \left(\frac{\tilde{\sigma}_l^2}{\hat{\sigma}_l^2} \right) + \sum_{l=1}^n \left(\frac{v_l^{(0)}}{\tilde{\sigma}_l^2} - \frac{\hat{v}_l}{\hat{\sigma}_l^2} \right),$$

where $\tilde{\sigma}_l^2 = H(w_l^{\mathsf{T}} \tilde{\gamma})$. Under H_4 , the asymptotic distribution of w_4 is χ_p^2 . We can find

$$E(w_4) = p + \text{tr} \left[A_d B_d \Lambda^{-1} (2P^2 - Q) \right] - 1^{\top} P \Lambda^{-1} (B^{(2)} \otimes A) \Lambda^{-1} P 1$$

$$+ \frac{1}{2} 1^{\top} P \Lambda^{-1} B_d A B_d \Lambda^{-1} P 1 + 1^{\top} P A_d A B_d \Lambda^{-1} P 1,$$

where Λ should be estimated by $\widetilde{\Lambda}$. The Bartlett correction associated with w_4 is $c_4 = p^{-1}E(w_4)$ and then the modified statistic $w_4^* = \widetilde{c_4}^{-1}w_4$ is distributed under H_4 to order n^{-1} as χ_p^2 .

Finally, our interest is to test the whole vector γ , i.e., $H_5: \gamma = \gamma^{(0)}$ against $A: \gamma \neq \gamma^{(0)}$, where $\gamma^{(0)}$ is a specified vector of dimension q and β is a vector of nuisance parameters.

The LR statistic for testing H5 becomes

$$w_5 = \sum_{l=1}^n \log \left(\frac{\sigma_l^{(0)^2}}{\widehat{\sigma}_l^2} \right) + \sum_{l=1}^n \left[\frac{(Y_l - x_l^\top \widetilde{\beta})^2}{\sigma_l^{(0)^2}} - \frac{(Y_l - x_l^\top \widehat{\beta})^2}{\widehat{\sigma}_l^2} \right],$$

where $\tilde{\beta} = (X^{\top}\Lambda^{(0)^{-1}}X)^{-1}X^{\top}\Lambda^{(0)^{-1}}Y$. Under H_{δ} , the distribution of w_{δ} is of order n^{-1} away from χ_q^2 . We can show that $E(w_5) = q + h(X, W, \Lambda^{(0)})$ and the Bartlett correction determined by $c_5 = 1 + h(X, W, \Lambda^{(0)})/q$ renders the n^{-1} term equals to zero and the error of the χ_q^2 approximation becomes of order n^{-2} . If q = 1 and w = 1, we have the simple expression

$$c_{\delta} = 1 + \frac{3(p+1)^2 - 1}{6n} + \frac{(\delta - 1)p}{n\tau^2}.$$

Bartlett corrections for several other tests in model (1) are analogously derived because of the generality of equation (7).

5 Simulation results

We now perform a Monte Carlo simulation study using the following model

$$\mu_{l} = \beta_{0} + \beta_{1}x_{l1} + \beta_{2}x_{l2} + \beta_{3}x_{l3},$$

$$\sigma_{l}^{2} = \sigma^{2}(\gamma_{0} + \gamma_{1}w_{l1} + \gamma_{2}w_{l2})^{2\lambda},$$
(14)

for $l=1,\ldots,n$. We wish to test the null hypothesis $H_3:\beta_1=\beta_2=2,\ \gamma_2=0$ against the hypothesis of violation of at least one equality. For the simulations the nuisance parameters were fixed at $\beta_0=5,\ \beta_3=3,\ \gamma_0=3$ and $\gamma_1=1,$ and the explanatory variables x_{ii} 's and w_{ii} 's were chosen as random draws from uniform U(0,20) and U(0,5) distributions, respectively. Their values were held constant throughout the simulations with equal sample sizes. We take $\sigma^2=1$ and 5, $X=0.5,\ 1$ and 2 and the number of observations was set at n=10,20 and 30. The simulations run-size was 10,000 in each case, and calculations were performed in FORTRAN using NAG subroutines. In each simulation, we generate normal data following (14) and then we solve equations

(3) iteratively under H_3 and A to obtain the LR statistic w_3 and the modified statistic $w_3^* = \tilde{c_3}^{-1}w_3$, where $\tilde{c_3}$ comes from (13) estimated under H_3 . Further, we investigate the rejection rates of w_3 and w_3^* at the nominal 5% and 1% levels of the reference χ_3^2 distribution.

For each combination of σ^2 , n and λ , Tables 1 and 2 give the sumulated rejection rates of both statistics w_3 and w_3^* , i.e., the percentage of times that they exceed the appropriate upper points of χ_3^2 . The figures in these tables convey important information. First, the tendency of the unmodified LR statistic w_3 to reject the model (14) more often than is expected for the selected type I error, in finite samples, is confirmed. In fact, for all 18 cases reported, the rejection rates of w_3 are greater than the corresponding upper points of χ_3^2 . Second, the empirical sizes of the test based on the modified LR statistic w₃* are closer to the nominal levels than the empirical sizes of the unmodified statistic w_3 . So, there is strong evidence of the superiority of the χ_3^2 approximation for the distribution of w_3^* over the distribution of w_3 . Thus, the Bartlett correction is very effective in pushing the rejection rates of the modified statistic toward to the nominal levels. Third, the asymptotic chi-squared approximation for both test statistics w_3 and w_3^* works better for large values of n and small values of λ and σ^2 , i.e., when the variability of the normal observations is small, in agreement with the so called "the small-dispersion asymptotic result". As expected, the χ^2 approximation for both statistics deteriorates when the sample size n decreases or when the variability of the data increases. Finally, the simulations results presented in this section suggest that the Bartlett corrections usually employed can deliver accurate inferences with samples of small to moderate size.

Table 1: Rejection rates of w_3 and w_3^* for the hypothesis H_3 : $\beta_1 = \beta_2 = 2$, $\gamma_2 = 0$ in model (14) with $\sigma^2 = 1$.

n	nominal level $\alpha(\%)$	$\lambda = 0.5$		$\lambda = 1.0$		$\lambda = 2.0$	
		w_3	w_3^*	w_3	w_3^*	w_3	w3*
10	5	12.5	6.8	14.2	7.1	16.4	7.1
	1	7.2	2.5	8.5	3.2	9.2	3.5
20	5	8.3	5.4	9.8	6.2	10.1	6.4
	1	3.1	2.0	5.4	2.5	6.1	2.4
30	5	6.1	5.2	7.1	5.8	7.4	5.9
	1	1.9	1.1	2.8	1.3	3.5	1.6

Table 2: Rejection rates of w_3 and w_3^* for the hypothesis H_3 : $\beta_1 = \beta_2 = 2$, $\gamma_2 = 0$ in model (14) with $\sigma^2 = 5$.

	nominal level $\alpha(\%)$	$\lambda = 0.5$		$\lambda = 1.0$		$\lambda = 2.0$	
78		ws	w ₃ *	w ₃	w ₃	w_3	w_3^*
10	5	12.6	6.9	14.9	7.4	16.5	8.1
	1	7.2	2.4	8.6	3.1	9.4	3.5
20	5	8.4	5.5	10.0	6.2	10.3	6.5
	1	3.2	2.0	6.0	2.5	6.4	2.3
30	5	6.5	5.5	7.2	5.9	7.5	5.8
	1	2.0	1.2	2.3	1.4	3.6	1.5

References

Andrews, D.F. and Stafford, J.E. (1993). Tools for the symbolic computation of asymptotic expansions. Journal of the Royal Statistical Society B, 55, 613-627.

Attfield, C.L.F. (1991). A Bartlett adjustment to the likelihood ratio test for homoskedasticity in the linear model. *Economics Letters*, 37, 119-123.

Attfield, C.L.F. (1995). A Bartlett adjustment to the likelihood ratio test for a system of equations. *Journal of Econometrics*, 66, 207-223.

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

- Box, G.E.P. and Hill, W.J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16, 385-389.
- Chesher, A. and Smith, R. (1995). Bartlett corrections to likelihood ratio tests. Biometrika, 82, 433-436.
- Cook, R.D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. Biometrika, 70, 1-10.
- 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. (1985). The null expected deviance for an extended class of generalized linear models. *Lecture Notes in Statistics*, 32, 27-34.
- Cordeiro, G.M. (1987). On the corrections to the likelihood ratio statistics. Biometrika, 74, 265-274.
- Cordeiro, G.M. (1993a). Bartlett corrections and bias correction for two heteroscedastic regression models. Communications in Statistics - Theory and Methods, 22, 169– 188.
- Cordeiro, G.M. (1993b). General matrix formulae for computing Bartlett corrections. Statistics and Probability Letters, 16, 11-18.
- Cordeiro, G.M. (1995). Performance of a Bartlett-type modification for the deviance.

 Journal of Statistical Computation and Simulation, 51, 385-403.
- 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.
- 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. London: Chapman and Hall.
- Cox, D.R. and Reid, N. (1987). Parameters orthogonality and approximate conditional inference (with discussion). Journal of the Royal Statistical Society, B, 49, 1-39.
- Cribari-Neto, F. and Cordeiro, G.M. (1996). On Bartlett and Bartlett-type corrections. Econometric Reviews, 15, 339–367.

- Cribari-Neto, F. and Ferrari, S.L.P. (1995). Bartlett-corrected tests for heteroskedastic linear models. *Economics Letters*, 48, 113-118.
- Cribari-Neto, F. and Zarkos, S. (1995). Improved test statistics for multivariate regression. Economics Letters, 49, 113-120.
- Ferrari, S.L.P. and Arellano-Valle, R.B. (1993). Bartlett-corrected tests for regression models with Student-t independent errors. Working Paper 9310, Department of Statistics, University of São Paulo.
- Glejser, H. (1969). A new test for heteroscedasticity. Journal of American Statistical Association, 64, 316-323.
- Goldfeld, S.M. and Quandt, R.E. (1965). Some tests for heteroscedasticity. Journal of American Statistical Association, 69, 539-547.
- Harrison, M.J. and McCabe, B.P. (1979). A test for heteroscedasticity based on ordinary least squares residuals. Journal of American Statistical Association, 74, 494-500.
- Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution. Annals of the Institute of Statistical Mathematics A, 29, 359-378.
- Jarque, C.M. (1981). A test for heteroscedasticity in a limited dependent variable model. Australian Journal of Statistics, 23, 159-163.
- Jensen, J.L. (1993). A historical sketch and some new results on the improved likelihood ratio statistic. Scandinavian Journal of Statistics, 20, 1-15.
- Jørgensen, B. (1987). Exponential dispersion models. Journal of the Royal Statistical Society, B, 49, 127-162.
- Koenker, R. and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantile. Econometrica, 50, 43-64.
- Larsson, R. (1994). Bartlett corrections for unit root test statistics. Unpublished paper, Institute of Mathematial Statistics, University of Copenhagen.
- Lawley, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. Biometrika, 71, 233-244.
- McCullagh, P. and Cox, D.R. (1986). Invariants and the likelihood ratio statistic. Annals of Statistic, 14, 1419-1430.

- Møller, J. (1986). Bartlett adjustments for structured covariances. Scandinavian Journal of Statistics, 13, 1-15.
- Moulton, L.H., Weissfeld, L.A. and St. Laurent, R.T. (1993). Bartlett corrections factors in logistic regression models. Computational Statistics and Data Analysis, 15, 1-11.
- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society A, 135, 370-384.
- Nielsen, B. (1995). Bartlett correction for the unit root test in autoregressive models. Unpublished paper, Nuffield College, University of Oxford.
- Stafford, J.E. and Andrews, D.F. (1993). A symbolic algorithm for studying adjustments to the profile likelihood. *Biometrika*, 80, 715-730.
- Wang, Y. (1994). A Bartlett-type adjustment for the likelihood ratio statistic with an ordered alternative. Statistics and Probability Letters, 20, 347-352.
- Williams, D.A. (1976). Improved likelihood ratio tests for complete contingency tables. Biometrika, 63, 33-37.
- Wolfram, S. (1991). Mathematica: A System for Doing Mathematics by Computer. 2.ed. New York: Addison-Wesley.
- Wong, M.Y. (1991). Bartlett adjustment to the likelihood ratio statistic for testing several slopes. Biometrika, 78, 221-224.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

- 9701 BOFARINE, H.; ARELLANO-VALE, R.B. Weak nondifferential measurement error models. IME-USP, 1997. 12p. (RT-MAE-9701)
- 9702 FERRARI, S.L.P.; CORDEIRO, G. M.; CRIBARI-NETO, F.
 Higher Order Asymptotic Refinements for Score Tests
 in Proper Dispersion Models. IME-USP, 1997. 14p.
 (RT-MAE-9702)
- 9703 DOREA, C.; GALVES, A.; KIRA, E.; ALENCAR, A. P.
 Markovian modeling of the stress contours of
 Brazilian and European Portuguese. IME-USP, 1997.
 11p. (RT-MAE-9703)
- 9704 FONTES, L.R.G.; ISOPI, M.; KOHAYAKAWA, Y.; PICCO, P.
 The Spectral Gap of the REM under Metropolis
 Dynamics. IME-USP, 1997. 24p. (RT-MAE-9704)
- 9705 GIMENEZ, P.; BOLFARINE, H.; COLOSINO, E.A. Estimation in weibull regression model with measurement error.IME-USP, 1997. 17p. (RT-MAE-9705)
- 9706 GALVES, A.; GUIOL, H. Relaxation time of the onedimensional symmetric zero range process with constant rate. IME-USP, 1997. 10p. (RT-MAE-9706)
- 9707 GUIOL, H. A note about a burton keane's theorem. IME-USP, 1997. 7p. (RT-MAE-9707)
- 9708 ARELLANO-VALLE, R.B.; FERRARI, S.L.P.; CRIBARI-NETO, F.
 Bartlett and barblett-type corrections for testing
 linear restrictions. IME-USP, 1997. 5p. (RT-MAE-9708)
- 9709 PAULA, G.A. One-Sided Tests in Dose-Response Models. IME-USP, 1997. 29p. (RT-MAE-9709)
- 9710 BRESSAUD, X. Subshift on an infinite alphabet. IME-USP, 1997. 34p. (RT-MAE-9710)

- 9711 YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H. Stochastic Monotonicity properties of bayes estimation of the population size for capture-recapture data. IME-USP. 1997. 12p. (RT-MAE-9711)
- 9712 GUINENEZ,P.;COLOSIMO,E.A.; BOLFARINE,H. Asymptotic relative efficiency of wald tests in measurement error models. IME-USP. 1997. 14p. (RT-MAE-9712)
- 9713 SCHONMANN, R.H.; TANAKA, N.I. Lack of monotonicity in ferromagnetic ising model phase diagrams. IME-UPS. 1997. 12p. (RT-MAE-9713)
- 9714 LEITE, J.G.; TORRES, V.H.S.; TIWARI, R.C.; ZALKIKAR, J.
 Bayes estimation of dirichlet process parameter.
 IME-USP. 1997. 21p. (RT-MAE-9714)
- 9715 GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E. Hypotheses testing based on a corrected score function for errors-in-variables models. 1997. 16p. (RT-MAE-9715)
- 9716 PAULA, G.A.; BOLFARINE, H. Some results on the slope of the linear regression model for the analysis of pretest-posttest data. 1997. 08p. (RT-MAE-9716)

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 05315-970 - São Paulo, Brasil