Brazilian Journal of Probability and Statistics (2000), 14, pp. 141–157.

IMPROVED ESTIMATION FOR ROBUST ECONOMETRIC REGRESSION MODELS

Klaus L. P. Vasconcellos
Departamento de Estatística
CCEN/UFPE, Cidade Universitária
50740-540, Recife, PE, Brazil
klaus@de.ufpe.br

Gauss M. Cordeiro
Departamento de Estatística
IM/UFBA, Cidade Universitária
40170-110, Salvador, BA, Brazil
gauss@ufba.br

Lúcia P. Barroso
Departamento de Estatística
IME/USP
Caixa Postal 66281
05315-970, São Paulo, SP, Brazil
lbarroso@ime.usp.br

Summary

The t distribution has proved to be a useful alternative to the normal distribution in many econometric regression models, especially when robust estimation is desired. In this work, we consider a nonlinear heteroskedastic Student t regression model. We suppose the observations to be independently t distributed, with the location and scale parameters for each observation being related to linear combinations of some explanatory variables, through regular, and possibly nonlinear, completely known link functions. We obtain the second order biases of the maximum likelihood estimates of the coefficients of those linear combinations and show that the biases will only depend on the first two derivatives of the link functions. We also express the biases in a closed matrix form, allowing them to be easily computed, in practical applications, from auxiliary generalized linear regressions. We discuss some important special cases and present Monte Carlo simulation results indicating that the bias-corrected estimates outperform the corresponding uncorrected estimates for relatively small sample sizes. An example with real data showing the usefulness of bias correction for this model is also presented.

Key Words: Bias correction; heteroskedastic model; link function; maximum likelihood estimate; Student t model.

1 Introduction

The study of the behaviour of maximum likelihood estimates (MLEs) for small sample sizes constitute an important area of research in econometric models. We know that MLEs produce, in general, biased estimates of the true parameter values. The bias correction of the MLEs is particularly important when the sample size, or the total information, is small. The biases of the estimates do not represent a serious problem for relatively large sample sizes, since they are, in general, of order $O(n^{-1})$, while the asymptotic standard errors are of order $O(n^{-1/2})$. However, for small or even moderate values of n, the bias has to be taken into consideration and availability of formulae for its approximate computation is important for a good estimation performance of many regression models that are used in a number of econometric applications.

The improvement of MLEs using bias correction has been extensively studied in the statistical literature. Box (1971) gives a general expression for the n^{-1} bias in multivariate nonlinear models where covariance matrices are known. Pike, Hill and Smith (1980) investigate the bias in logistic linear models. For nonlinear regression models, Cook, Tsai and Wei (1986) relate bias to the position of the explanatory variables in the sample space. Young and Bakir (1987) show that bias correction can improve estimation in generalized log-gamma regression models. Cordeiro and McCullagh (1991) and Cordeiro and Klein (1994) have given general matrix formulae for bias correction in generalized linear models and ARMA models, respectively. More recently, Cordeiro and Vasconcellos (1997) obtain general matrix formulae for bias correction in multivariate nonlinear regression models with normal errors, while Cordeiro, Vasconcellos and Santos (1998) present bias correction formulae for univariate nonlinear Student-t regression models. This study is extended in Vasconcellos and Cordeiro (2000) for multivariate nonlinear Student-t regression models. Also, Cordeiro and Vasconcellos (1999) obtain second-order biases of the MLEs in von Mises regression models.

In this paper, we derive general matrix formulae for second-order biases of the MLEs of the parameters in an heteroskedastic nonlinear Student-t regression model. The t distribution has been extensively discussed in the statistical literature, as an option to model data with heavy-tailed distributions, and also as a robust estimation procedure, since the t distribution is not as sensitive to outliers as the normal distribution. Jeffreys (1939) uses the t distribution to describe astronomical data. Zellner (1976) studies the univariate linear regression model where the vector of observed responses is multivariate t distributed. West (1984) performs Bayesian analysis related to the use of the t distribution in regression problems. Sutradhar and Ali

(1986) extend Zellner's work to cover MLEs in a multivariate regression model. Lange, Little and Taylor (1989) provide a rich illustration on the use of the t distribution as a robust extension of the normal distribution by considering univariate and multivariate Student-t regression models and give a number of practical applications. Ferrari and Arellano-Valle (1996) have developed Bartlett and Bartlett-type corrections to improve likelihood ratio and score tests in the class of univariate linear t regression models discussed in Lange, Little and Taylor (1989). Also, formulae for second-order biases of MLEs in univariate nonlinear t regression models are given in Cordeiro, Vasconcellos and Santos (1998), this study being extended in Vasconcellos and Cordeiro (2000) for multivariate regression models.

The matrix form expressions given here for the n^{-1} biases of the MLEs of the parameters correspond to the coefficients of auxiliary generalized least squares linear regressions. The n^{-1} biases are, therefore, very simple to compute and the necessary formulae to calculate them involve only elementary operations on matrices, being of easy implementation on any matrix based programming language as Ox, GAUSS or S-PLUS.

The paper is organized in the following form. Section 2 presents a formal description of the heteroskedastic nonlinear Student-t regression model. In Sections 3 and 4, we use Cox and Snell's (1968) general expression to obtain formulae for second-order biases of the MLEs of the parameters in the model defined in Section 2. Some special cases of the formulae derived are discussed in Section 5. In Section 6, we present some simulation results suggesting that the corrected estimates have better performance than the uncorrected ones in small samples. Finally, in Section 7, we discuss an example with real data showing the usefulness of our formulae for the Student t heteroskedastic model.

2 Model Definition

We consider a univariate nonlinear heteroskedastic regression model where the observations $y_1, \ldots, y_n$ are independent and each observation y_i has a Student-t distribution with location parameter μ_i, scale parameter σ_i and ν degrees of freedom. The density of y_i, for each $i = 1, \ldots, n$, is therefore given by

$$f(y; \mu_i, \sigma_i^2, \nu) = \frac{\nu^{\nu/2}\Gamma(\frac{\nu+1}{2})}{\sigma_i\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left\{ \nu + \left(\frac{y-\mu_i}{\sigma_i} \right)^2 \right\}^{-(\nu+1)/2}, \qquad (2.1)$$

where $\sigma_i > 0$. We define the precision parameter $\phi_i = \sigma_i^{-2}$ for each observation y_i. We then assume that the parameters μ_i and ϕ_i can be expressed as $\mu_i = g(\eta_i)$ and $\phi_i = h(\tau_i)$ where $g(\cdot)$ and $h(\cdot)$ are known one-to-one continuously twice differentiable functions, while η_i and τ_i are linear combinations of some explanatory variables. The linear predictors $\eta = (\eta_1, \ldots, \eta_n)^T$ and $\tau = (\tau_1, \ldots, \tau_n)^T$ are given, respectively, by the relations $\eta = X\beta$ and $\tau = Z\gamma$, where X is a specified $n \times p$ matrix of full rank $p < n$, $\beta = (\beta_1, \ldots, \beta_p)^T$ is a set of unknown parameters, Z is a specified $n \times q$ matrix of full rank $q < n$ and $\gamma = (\gamma_1, \ldots, \gamma_q)^T$ is a set of unknown parameters. It is clear that the link function $h(\cdot)$ must be a positive function. This function is usually called dispersion link function whereas $g(\cdot)$ is the mean link function. The covariates in Z constitute, in general, although not necessarily, a subset of the covariates in X. The parameters β and γ are assumed to be functionally independent, which leaves us $p+q$ parameters to be estimated. We also assume that $p+q$ is small relative to n and that ν is known.

At this point, it is important to make some remarks on the number ν of degrees of freedom. We first observe that, for $\nu > 1$, the expected values of all y_i's exist and correspond to the location parameters μ_i; also, for $\nu > 2$, the variance of each observation exists and is given by $\nu\sigma_i^2/(\nu - 2)$. Moreover, the normal distribution is obtained by letting $\nu \to \infty$. All of these results are well-known. However, it must be emphasized that a known value for ν is not a very strong assumption for the purpose of our work. Indeed, Lange, Little and Taylor (1989) suggest that 'ν should be fixed for small data sets and estimated for large ones'. In other words, small samples do not typically provide sufficient information about ν for this parameter to be estimated with enough precision. Since our chief concern is the analysis of small data sets where bias correction is particularly important, it seems reasonable to fix ν at a small value. Moreover, the influence function becomes unbounded when ν is unknown and this is a major argument for fixing ν. Lange, Little and Taylor (1989) point out that, for small samples, $\nu = 4$ has worked well for many of their applications. The fixed ν assumption we adopt here is also considered in Ferrari and Arellano-Valle (1996), Cordeiro, Vasconcellos and Santos (1998) and Vasconcellos and Cordeiro (2000). The t distribution, as pointed out earlier, provides a useful extension of the normal distribution for situations where robustness plays an important role in estimation. The normal case is trivially obtained as the limiting case when $\nu \to \infty$.

We must here emphasize that we are working with the hypothesis that the observations are independent. Therefore, the approach we use here is different from that of Zellner (1976). Instead, we follow Lange, Little and Taylor (1989), Ferrari and Arellano-Vale (1996), Cordeiro, Vasconcellos and Santos (1998) and also Vasconcellos and Cordeiro (2000). From (2.1) and this hypothesis of independence, the total log-likelihood $\ell(\theta)$ for the $(p+q) \times 1$ vector $\theta = (\beta^T, \gamma^T)^T$ of unknown parameters, given observations $y_1, ..., y_n$, becomes

$$\ell(\theta)=\frac{1}{2}\left\{C+\sum_{i}\log h(\tau_i)-(\nu+1)\sum_{i}\log(\nu+h(\tau_i)(y_i-g(\eta_i))^2\right\}, \tag{2.2}$$

where C is given by

$$C=2n\log\left\{\frac{\nu^{\nu/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\pi^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\right\}, \tag{2.3}$$

and $\sum_{i}$ runs over all n observations. The quantity C defined in (2.3) 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 (Cox and Hinkley, 1974; Chapter 9) on $\ell(\theta)$ and its first three derivatives hold as n goes to infinity; these conditions are usually satisfied in practice.

We now introduce the notation used throughout the paper. The total log-likelihood derivatives with respect to the unknown parameters are indicated by indices, where lower-case letters $r, s, t, \ldots$ correspond to derivatives with respect to the β parameters, while upper-case letters $R, S, T, \ldots$ correspond to derivatives with respect to the γ parameters. Thus, $U_r = \partial\ell/\partial\beta_r, U_R = \partial\ell/\partial\gamma_R, U_{Rs} = \partial^2\ell/\partial\gamma_R\partial\beta_s, U_{rsT} = \partial^3\ell/\partial\beta_r\partial\beta_s\partial\gamma_T$ and so on. The standard notation for the moments of these derivatives is used here (Lawley, 1956; Cordeiro, 1993a): $\kappa_{rs} = E(U_{rs}), \kappa_{R,S} = E(U_R U_S), \kappa_{rs,T} = E(U_{rs}U_T), \kappa_{rst} = E(U_{rst})$, etc., where all κ's refer to a total over the sample, and are, in general, of order n. Also, their derivatives are denoted by $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t, \kappa_{rS}^{(T)} = \partial\kappa_{rS}/\partial\gamma_T$, etc. Moreover, the information matrices with respect to β and γ are denoted by K_β and K_γ, respectively, with their typical elements being given by $\kappa_{r,s}$ and $\kappa_{R,S}$, respectively. We assume that K_β and K_γ are nonsingular, denoting the typical elements of their inverses by $\kappa^{r,s}$ and $\kappa^{R,S}$, respectively.

Differentiation of (2.2) yields

$$\begin{aligned}
U_r &= (\nu+1)\sum_{i}\frac{h_i(y_i-g_i)g_i' x_{ir}}{\nu+h_i(y_i-g_i)^2},\\
U_R &= \frac{1}{2}\sum_{i}\frac{h_i' z_{iR}}{h_i}-\frac{(\nu+1)}{2}\sum_{i}\frac{(y_i-g_i)^2 h_i' z_{iR}}{\nu+h_i(y_i-g_i)^2},
\end{aligned}$$

where $g_i = g(\eta_i) = \mu_i$, $h_i = h(\tau_i) = \phi_i$, $g_i' = \partial g(\eta_i)/\partial\eta_i$ and $h_i' = \partial h(\tau_i)/\partial\tau_i$. The MLEs of the parameters β and γ can be obtained as the solution of a nonlinear system of $p+q$ equations: $U_r = 0$ for $r = 1, \ldots, p$ and $U_R = 0$ for $R = 1, \ldots, q$. This system can, in practice, be solved using an iterative procedure that converges to the desired MLEs. Good numerical alternatives to maximize the log-likelihood function are the offset algorithm described by Cordeiro and Paula (1989) and the MaxBFGS function implemented in the Ox programming language (Doornik, 1999), which uses BFGS, the quasi-Newton method developed by Broyden, Fletcher, Goldfarb and Shanno (see, e.g., Fletcher, 1987). This latter was the procedure used in the numerical studies of the present work. From here onwards, we assume that the MLEs $\hat{\beta}$ and $\hat{\gamma}$ exist, are finite, unique and are given by the solution of $U_r = 0$ for $r = 1, ..., p$ and $U_R = 0$ for $R = 1, ..., q$.

3 Biases of the Estimate of β

From the log-likelihood defined in (2.2) and basic properties of the Student t distribution (Zellner, 1971; Appendix B), we obtain the following moments:

$$\begin{aligned}
\kappa_{rs} &= -\frac{\nu+1}{\nu+3}\sum_i x_{ir}x_{is}h_i g_i'^2, \\
\kappa_{RS} &= -\frac{\nu}{2(\nu+3)}\sum_i z_{iR}z_{iS}\left(\frac{h_i'}{h_i}\right)^2, \\
\kappa_{rst} &= -\frac{3(\nu+1)}{\nu+3}\sum_i x_{ir}x_{is}x_{it}h_i g_i' g_i'', \\
\kappa_{rsT} &= -\frac{(\nu+1)(\nu+2)}{(\nu+3)(\nu+5)}\sum_i x_{ir}x_{is}z_{iT}h_i' g_i'^2, \\
\kappa_{RST} &= \frac{\nu}{(\nu+3)}\sum_i z_{iR}z_{iS}z_{iT}\left(\frac{(\nu+8)}{(\nu+5)}\left(\frac{h_i'}{h_i}\right)^3 - \frac{3h_i'h_i''}{2h_i^2}\right) \quad \text{and} \\
\kappa_{rS} &= \kappa_{rST} = 0,
\end{aligned}$$

where $g_i'' = \partial^2 g(\eta_i)/\partial\eta_i^2$ and $h_i'' = \partial^2 h(\tau_i)/\partial\tau_i^2$. We immediately observe that β and γ are globally orthogonal (Cox and Reid, 1987) since $\kappa_{rS} = 0$, for all $r = 1, ..., p$ and $S = 1, ..., q$. Therefore, the joint information matrix K for $\theta = (\beta^T, \gamma^T)^T$ is block-diagonal, $K = \text{diag}\{K_\beta, K_\gamma\}$, with the information matrices K_β and K_γ for β and γ given by $K_\beta = X^T W_1 X$ and $K_\gamma = Z^T W_2 Z$, where W_1 and W_2 are the $n \times n$ diagonal matrices $W_1 = [(\nu+1)/(\nu+3)]\text{diag}\{h_i g_i'^2\}$ and $W_2 = [\nu/(2\nu+6)]\text{diag}\{(h_i'/h_i)^2\}$. In view of block-diagonality of K, the n^{-1} biases of $\hat{\beta}$ and $\hat{\gamma}$ can be obtained in simple forms.

Our first goal is to derive the second-order bias of the MLE $\hat{\beta}$ of β. To do so, we will follow the approach in Cordeiro and Vasconcellos (1999). Let

$B(\hat{\beta}_s)$ be the n^{-1} bias of $\hat{\beta}_s$. From the general formula of Cox and Snell (1968) and the block-diagonality of K, we have

$$B(\hat{\beta}_s) = \sum_{r,t,u} \kappa^{sr}\kappa^{tu}\left\{\kappa_{rt}^{(u)} - \frac{1}{2}\kappa_{rtu}\right\} + \sum_{r,T,U} \kappa^{sr}\kappa^{TU}\left\{\kappa_{rT}^{(U)} - \frac{1}{2}\kappa_{rTU}\right\}. \quad (3.1)$$

The second term in (3.1) vanishes, since $\kappa_{rTU} = \kappa_{rT} = 0$. Thus, the second-order bias of $\hat{\beta}$ will be the same, regardless of whether the dispersion parameters of the observations are known or unknown. Now, by rearranging the first term in (3.1), we have

$$B(\hat{\beta}_s) = -\frac{\nu+1}{2(\nu+3)}\sum_i\sum_r a_i\kappa^{sr}x_{ir}\sum_{t,u}x_{it}\kappa^{tu}x_{iu},$$

where $a_i = h_i g_i' g_i''$ for $i = 1,\ldots,n$.

If we let x_i be the $p \times 1$ vector representing the i-th column of X^T and ρ_s represents the s-th column of the identity matrix I_p, then, the above equation can be written in matrix notation as

$$B(\hat{\beta}_s) = -\frac{\nu+1}{2(\nu+3)}\rho_s^T K_\beta^{-1}\sum_i a_i(x_i^T K_\beta^{-1}x_i)x_i.$$

Now, we define $\mathrm{cov}(\hat{\eta}) = XK_\beta^{-1}X^T$, $W_3 = \mathrm{diag}\{h_i g_i' g_i''\}$ and $\delta = -\mathrm{d}\{[(\nu+1)/(2\nu+6)]W_1^{-1}W_3\mathrm{cov}(\hat{\eta})\}$, where $\mathrm{d}(A)$ represents a vector whose elements are the diagonal elements of A. Note that $\mathrm{cov}(\hat{\eta})$ represents the large-sample covariance matrix of $\hat{\eta} = X\hat{\beta}$. The expression for $B(\hat{\beta})$ can then be written as

$$B(\hat{\beta}) = (X^T W_1 X)^{-1}X^T W_1\delta. \quad (3.2)$$

The bias $B(\hat{\beta})$ can be simply obtained from a generalized linear regression of δ on X, having W_1 as a weight matrix. It has also advantages for algebraic purposes because it only involves products and inversion of matrices. Equation (3.2) depends on the parameters only through the mean link function g_i, the first and second partial derivatives with respect to the linear predictor η_i, and the dispersion link function h_i for $i = 1,\ldots,n$. Although (3.2) has a simple form, its interpretation is not straightforward.

One can use equation (3.2) with a computer algebra system such as MATHEMATICA (Wolfram, 1991) or MAPLE (Abell and Braselton, 1994) to obtain closed-form expressions for $B(\hat{\beta})$.

All quantities have to be evaluated at $\hat{\beta}$ and $\hat{\gamma}$. It is then possible to obtain the bias-corrected estimate as $\tilde{\beta} = \hat{\beta} - \hat{B}(\hat{\beta})$, where $\hat{B}(\hat{\beta})$ denotes the right-hand side of (3.2) evaluated at $\hat{\beta}$ and $\hat{\gamma}$. The corrected estimate $\tilde{\beta}$ is expected to have better sampling properties than $\hat{\beta}$ in small samples.

4 Biases of the Estimate of γ

We now proceed to the calculation of the n^{-1} bias of the MLE $\hat{\gamma}$. For this purpose, we consider again Cox and Snell's formula. Since the information matrix for $\theta = (\beta^T, \gamma^T)^T$ is block-diagonal, we can write the n^{-1} bias of $\hat{\gamma}_A$ as

$$B(\hat{\gamma}_A) = \sum_{R,t,u} \kappa^{A,R}\kappa^{t,u}\left\{\kappa_{Rt}^{(u)} - \frac{1}{2}\kappa_{Rtu}\right\} + \sum_{R,T,U} \kappa^{A,R}\kappa^{T,U}\left\{\kappa_{RT}^{(U)} - \frac{1}{2}\kappa_{RTU}\right\}. \tag{4.1}$$

Let $B_1(\hat{\gamma}_A)$ denote the first sum of (4.1). We have

$$B_1(\hat{\gamma}_A) = \frac{(\nu+1)(\nu+2)}{2(\nu+3)(\nu+5)} \sum_{R,t,u} \kappa^{A,R}\kappa^{t,u} \sum_i b_i z_{iR} x_{it} x_{iu},$$

where $b_i = g_i'^2 h_i'$ for $i = 1, \ldots, n$. Let z_i and ρ_A be the i-th column of Z^T and the Ath column of the identity matrix I_q. Then, the above equation can be rearranged as

$$B_1(\hat{\gamma}_A) = \frac{(\nu+1)(\nu+2)}{2(\nu+3)(\nu+5)} \rho_A^T K_\gamma^{-1} \sum_i b_i (x_i^T K_\beta^{-1} x_i) z_i. \tag{4.2}$$

For the second sum in (4.1), $B_2(\hat{\gamma}_A)$, we write

$$B_2(\hat{\gamma}_A) = \sum_{R,T,U} \kappa^{A,R}\kappa^{T,U} \sum_i c_i z_{iR} z_{iT} z_{iU},$$

where $c_i = \{\nu(\nu+2)/[2(\nu+3)(\nu+5)]\}(h_i'/h_i)^3 - \{\nu/[4(\nu+3)]\}h_i' h_i''/h_i^2$. Then, we obtain

$$B_2(\hat{\gamma}_A) = \rho_A^T K_\gamma^{-1} \sum_i c_i (z_i^T K_\gamma^{-1} z_i) z_i. \tag{4.3}$$

Adding $B_1(\hat{\gamma}_A)$, given by (4.2), to the expression for $B_2(\hat{\gamma}_A)$, given by (4.3), we find the n^{-1} bias of the MLE $\hat{\gamma}$. This n^{-1} bias can also be written in a neat form, defining the matrices $W_4 = \{[(\nu+1)(\nu+2)]/[2(\nu+3)(\nu+5)]\}\mathrm{diag}\{g_i'^2 h_i'\}$ and $W_5 = \nu(\nu+2)/[2(\nu+3)(\nu+5)]\mathrm{diag}\{(h_i'/h_i)^3\} - \nu/[4(\nu+3)]\mathrm{diag}\{h_i' h_i''/h_i^2\}$. Consider also $\mathrm{cov}(\hat{\tau}) = Z K_\gamma^{-1} Z^T$, which represents the large-sample covariance matrix of $\hat{\tau} = Z\hat{\gamma}$, the MLE of τ. Now let $\xi = \mathrm{d}\{W_2^{-1}[W_4 \mathrm{cov}(\hat{\eta}) + W_5 \mathrm{cov}(\hat{\tau})]\}$. Then, we can obtain the n^{-1} bias of the MLE $\hat{\gamma}$ as

$$B(\hat{\gamma}) = (Z^T W_2 Z)^{-1} Z^T W_2 \xi. \tag{4.4}$$

Equation (4.4) can be written as the vector of coefficients from a generalized least-squares regression of ξ on Z using W_2 as weight matrix. Observe that the expression for the n^{-1} bias of $\hat{\gamma}$ in (4.4) depends on the model matrices X and Z and on the mean link function g_i and its first derivative and on the dispersion link function h_i and its first two derivatives. Hence, both formulae (3.2) and (4.4) can be very easily obtained from the defined model. Also, if closed-form expressions for K_β and K_γ are available, we can obtain closed-from expressions for $B(\hat{\beta})$ and $B(\hat{\gamma})$ from (3.2) and (4.4) using a computer algebra system such as MAPLE or MATHEMATICA.

Clearly, all quantities in (4.4) have to be evaluated at $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$ in order to obtain the corrected MLE $\tilde{\gamma} = \hat{\gamma} - \hat{B}(\hat{\gamma})$, where $\hat{B}(\hat{\gamma})$ is the value of (4.4) at $\hat{\theta}$. The corrected estimate $\tilde{\gamma}$ is expected to be closer to the true parameter vector than the unadjusted estimate $\hat{\gamma}$.

5 Special Cases

It is important to consider here some special cases for which formulae (3.2) and (4.4) can be simplified. We begin with the linear model. If the model is linear, then $g_i^{''} = 0$, for all $i = 1, \ldots, n$. Hence, the W_3 matrix defined in Section 3 becomes a zero matrix, and consequently we also have $\delta = 0$. Therefore, the second-order biases of the MLEs of the regression coefficients for the location parameters are zero for the linear model, that is $B(\hat{\beta}) = 0$, for this case. This result extends the result in Cordeiro, Vasconcellos and Santos (1998) for a possible heteroskedastic model.

We now consider the homoskedastic model. This case can be conveniently represented if we define Z as an $n \times 1$ vector of ones and h as the identity function. Here we have $q = 1$ and a constant $\phi = \tau = \gamma$ for all observations. We turn here to the study by Cordeiro, Vasconcellos and Santos (1998) of the general nonlinear homoskedastic Student t regression model. In their model, the observations are assumed to be independent and each observation y_i has a t density function with location parameter μ_i, precision parameter ϕ and ν degrees of freedom. The location μ_i is given by the systematic component $\mu_i = f_i(e_i, \beta)$, where the f_i's are possibly nonlinear twice continuously differentiable functions, the e_i's are vectors of explanatory variables and β is a vector of unknown parameters. It is shown in that paper that the second-order bias of the MLE $\hat{\beta}$ of β is given by

$$B(\hat{\beta}) = (F^T F)^{-1} F^T d, \tag{5.1}$$

where F is the $n \times p$ matrix $F = \left\{\frac{\partial f_i}{\partial \beta_j}\right\}$, $d = -\frac{1}{2\phi}\frac{(\nu+3)}{(\nu+1)} H \text{vec}((F^T F)^{-1})$, H is the $n \times p^2$ matrix $H = \left\{\frac{\partial^2 f_i}{\partial \beta_j \partial \beta_k}\right\}$ and vec is the operator that transforms a matrix into a vector by stacking the columns of the matrix, one underneath the other.

We apply formula (5.1) for our particular form of homoskedastic model. For our particular model, the matrix F will have as its ith row the vector $g_i' x_i^T$, where x_i^T is the ith row of X. Also, the matrix H will have as its ith row the vector $g_i^{''}(x_i \otimes x_i)^T$, where $\otimes$ represents the Kronecker product. Therefore, if we let $D_1 = \mathrm{diag}\{g_i'^2\}$ and $D_2 = \mathrm{diag}\{g_i' g_i^{''}\}$,we can see, after a few calculations that (5.1) will reduce to

$$B(\hat{\beta}) = (X^T D_1 X)^{-1} X^T D_1 \alpha,$$

where α is the vector $\alpha = -(1/2) D_1^{-1} D_2 \mathrm{d}\{X K_\beta^{-1} X^T\}$. But this is exactly the same result that is obtained by applying (3.2) to this particular case. Therefore, the particularization of the formula in Cordeiro, Vasconcellos and Santos (1998) for this special nonlinear model leads to the same expression for $B(\hat{\beta})$ from the particularization of (3.2) for this homoskedastic model. This provides a partical check of (3.2).

We now turn to the n^{-1} bias $B(\hat{\phi})$. Applying (4.4) for this particular case, it follows after some algebra that

$$B(\hat{\phi}) = \frac{(p+2)(\nu+2)(\nu+3)}{n\nu(\nu+5)} \phi \, ,$$

which coincides with Cordeiro, Vasconcellos and Santos' (1998) equation (4). This provides a partial check of (4.4).

Another interesting simplification occurs in the special case where $h(\tau) = \exp(\tau)$. If this is the case, then, $h_i = h_i' = h_i^{''}$ and it is not difficult to see that (4.4) can be written here as

$$B(\hat{\gamma}) = (Z^T Z)^{-1} Z^T \zeta,$$

where $\zeta = [(\nu-1)/(2\nu+10)]\mathrm{d}\{Z K_\gamma^{-1} Z^T\} + \{[(\nu+2)(\nu+3)]/[\nu(\nu+5)]\} W_1$ $\mathrm{d}\{X K_\beta^{-1} X^T\}$. This shows that, when the heteroskedasticity is modelled exponentially, the second-order bias of the γ parameters can be obtained by an ordinary linear least squares formula. Similar results have been obtained by Cordeiro (1993b) and also by Vasconcellos and Cordeiro (1997).

6 Simulation Results

We now perform a Monte Carlo simulation study of a Student t regression model, where the observations have means $\mu_i = \exp(\beta_0 + \beta_1 x_i)$ and precisions $\phi_i = \exp(\gamma_0 + \gamma_1 z_i)$, x_i and z_i being the values of the explanatory variables for each observation, assumed known.

Without loss of generality the true parameters were taken as $\beta = (1,1)^T$ and $\gamma = (1,1)^T$ and the values of the covariates x_i and z_i, $i = 1, ..., n$, were chosen as random draws from a normal $N(0,1)$ distribution, its values being held constant throughout the simulations with equal sample sizes. The number of observations was set at $n = 20$, 40 and 60. The number of degrees of freedom was fixed as $\nu = 4$. The simulations and the calculations of the MLEs and their biases were performed using the version 2.00 of the Ox programming language (Doornik, 1999). We carried out the simulations based on 10,000 replications. In each of the 10,000 replications, we computed the MLEs $\hat{\beta}$ and $\hat{\gamma}$ by maximizing the log-likelihood using the MaxBFGS routine of Ox. Then, we computed $B(\hat{\beta})$ and $B(\hat{\gamma})$ from formulae (3.2) and (4.4), with all quantities evaluated at $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$. For each n, we computed the sample means and standard errors of the MLEs and of the corrected estimates $\tilde{\beta}$ and $\tilde{\gamma}$, based on their values from the 10,000 trials.

Table 1
Uncorrected and corrected estimates for a simulated model

n	$\beta_0 = 1$	$\beta_1 = 1$	$\gamma_0 = 1$	$\gamma_1 = 1$
20	1.001	0.9981	1.268	1.234
	(0.0490)	(0.0540)	(0.590)	(0.461)
	1.001	0.9987	1.132	1.068
	(0.0490)	(0.0539)	(0.557)	(0.461)
40	0.9999	1.000	1.011	1.103
	(0.0141)	(0.0298)	(0.313)	(0.308)
	0.9999	1.000	1.001	1.016
	(0.0141)	(0.0298)	(0.303)	(0.308)
60	1.000	0.9993	1.025	1.070
	(0.0171)	(0.0292)	(0.273)	(0.248)
	1.000	0.9996	1.001	1.009
	(0.0171)	(0.0292)	(0.265)	(0.248)

Table 1 gives the sample means of both uncorrected and corrected estimates with their respective standard errors in parentheses. Overall, it is clear from the figures in this table that the bias-corrected estimates are closer to the true parameter values than the unadjusted estimates. Thus, the second-order bias correction seems very effective in bringing the corrected MLEs toward to their true values. It should be noted that reasonably large sample sizes are really necessary for uncorrected estimates of the γ_0

and γ_1 parameters to become accurate. It is also clear from these figures that the corrected estimates tend to have slightly smaller standard errors than the original ones for samples of small to moderate size. In these cases, the bias correction can lead to substantial improvement in terms of bias and mean square error.

7 Example

We discuss an example with real data, showing the usefulness of formula (4.4). The dataset used here was taken from Myers (1990). It consists of observations taken at twenty-five Bachelor Officers Quarters sites of the U.S. Navy. The dependent variable (y) measures monthly man-hours at the twenty-five different sites. In our example, we try to explain y with a linear regression that uses three explanatory variables: the average daily occupancy (x_1), the weekly hours of service desk operation (x_2) and the number of building wings at the respective sites (x_3). The observations are assumed to be independent.

We begin by trying the ordinary least squares fit, where we assume the observational variance to be constant throughout the sample. Let X be the 25×4 matrix containing the three explanatory variables and the intercept, each row representing a different observed site. Also, let y be the 25×1 column vector consisting of the observations of the dependent variable. The graph in the next page shows the plot of the R-Student residuals against the fitted values, for the twenty-five observations in the sample. The vector $\hat{y}$ of fitted values is given by $\hat{y} = X(X'X)^{-1}X^T y$, and the R-Student residual for the ith observation is computed as

$$\frac{y_i - \hat{y}_i}{s_{-i}\sqrt{1 - h_{ii}}}.$$

Here, h_{ii} is the ith diagonal value of the hat matrix $X(X'X)^{-1}X^T$. Also, the quantity s_{-i} represents the estimate of the standard error of the observations from the twenty-four sized sample that is obtained by deleting the ith case.

From the graph in Figure 1, it can be seen that the R-Student residuals corresponding to the observations labeled as 23 and 24 have very large absolute values, which indicates that those observations may be outliers. Also, the pattern of the residuals, with large R-Student absolute values corresponding to large fitted values suggests a possible heteroskedastic model.

We, then, consider two linear regression heteroskedastic models. The first one is a normal model where the observations are assumed independent and normally distributed and the precision (inverse variance) of observation i is modelled by $\sigma_i^{-2} = \exp(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3})$, with x_{i1}, x_{i2} and x_{i3} standing for the values of x_1, x_2 and x_3, respectively, at the ith site. The row vector $(1, x_{i1}, x_{i2}, x_{i3})$ corresponds to the ith row of X. The second model is a Student t linear regression model where the observations are assumed independent and Student t distributed with $\nu = 4$ degrees of freedom, the precision of observation i being modelled by $\phi_i = \exp(\delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3})$. All parameters in both models are estimated using maximum likelihood.

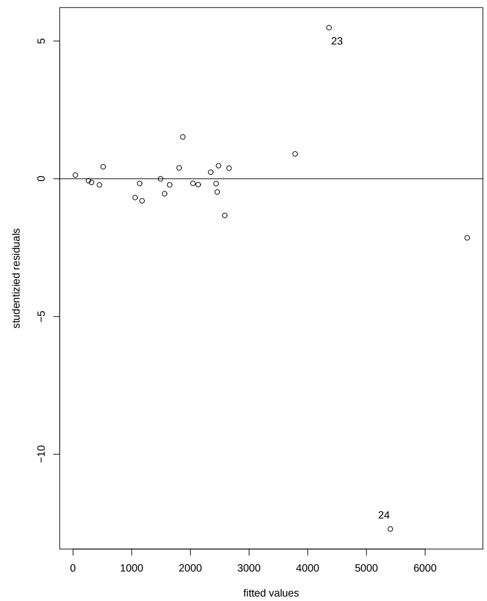

Figure 1
R-Student residuals for OLS

Tables 2 and 3 show the estimation results for both models. Also, for the normal model, the estimated regression equation is $\hat{y} = 126 + 22.1x_1 + 0.428x_2 - 9.95x_3$, while, for the t(4) model, the estimated regression equation is $\hat{y} = 134 + 20.8x_1 + 0.548x_2 - 6.92x_3$, only slightly different from the first equation.

From Table 2, we have the maximum likelihood estimates of the precision parameters, as well as the bias-corrected estimates, for the normal model. The bias-corrected estimates were obtained with the expressions derived in Vasconcellos and Cordeiro (1997). The asymptotic standard errors are obtained from the estimated information matrix. From the table, it is readily seen that the estimates of the precision parameters are indeed significant, which means that the heteroskedasticity hypothesis is very plausible. Also, the bias to standard error relation is large enough to justify the bias correction for all estimates.

Table 2
Results obtained for the normal model

	γ_0	γ_1	γ_2	γ_3
Maximum likelihood estimates	−9.966	−0.0111	−0.00898	0.0606
Bias corrected estimates	−10.348	−0.0116	−0.00648	0.0405
Asymptotic standard errors	0.515	0.00141	0.00384	0.0201
Bias to s.e. ratio	0.742	0.324	−0.651	1.003

Table 3
Results obtained for the Student t(4) model

	δ_0	δ_1	δ_2	δ_3
Maximum likelihood estimates	−9.302	−0.0121	−0.0130	0.117
Bias corrected estimates	−9.624	−0.0121	−0.0107	0.0926
Asymptotic standard errors	0.963	0.00263	0.00719	0.0376
Bias to s.e ratio	0.334	0.000798	−0.318	0.643

From Table 3, it can be seen that the MLEs of the precision parameters, as well as the bias-corrected parameters for the Student t(4) model. The bias corrected estimates were obtained with formula (4.4) and the asymptotic standard errors from the estimated information matrix. We can see that estimates of the precision parameters are also significant for this model. Also, from the bias to standard error relation, we can conclude that, apart from δ_1, the bias correction of the estimates seems useful in this case.

When comparing the normal and Student t(4) regression procedures, we see that the estimated regression equations do not differ too much. Also, for each explanatory variable, its respective estimated coefficient has the same sign in both models. The same thing can be verified for the estimated precision parameters that model heteroskedasticity for both models. In addition, it is seen that the estimated biases of the parameters relative to their estimated standard errors are much smaller in magnitude for the

Student t(4) model than for the normal model. This is an indication that the Student t(4) regression is more adequate here, which seems reasonable, since we have detected the possible presence of outliers in the sample.

All calculations were performed using the Ox matrix programming language (Doornik, 1999).

Acknowledgements

We gratefully acknowledge partial financial support from CNPq/Brazil.

(Received September, 1999. Revised July, 2000.)

References

Abell, M.L. and Braselton, J.P. (1994). *The Maple V Handbook*. New York: AP Professional.

Box, M. (1971).Bias in nonlinear estimation (with discussion). *J. R. Statist. Soc.* B, **33**, 171-201.

Cook, R., Tsai, C. and Wei, B. (1986). Bias in nonlinear regression. *Biometrika*, **73,** 615-623.

Cordeiro, G. M. (1993a). General matrix formulae for computing Bartlett corrections. *Statist. Prob. Letters*, **16**, 11-18.

Cordeiro, G. M. (1993b). Bartlett corrections and bias correction for two heteroscedastic models. *Comm. Statist. - Theory and Methods*, **22**, 169-188.

Cordeiro, G. M. and Klein, R. (1994). Bias correction in ARMA models. *Statist. Prob. Letters*, **19**, 169-176.

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *J. R. Statist. Soc.* B, **53**, 629-643.

Cordeiro, G.M. and Paula, G. A. (1989). Fitting non-exponential family nonlinear models in GLIM by using the offset facility. *Lecture Notes in Statistics*, **57**, 105-114.

Cordeiro, G. M. and Vasconcellos, K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statist. Prob. Letters*, **35,** 155-164.

Cordeiro, G. M. and Vasconcellos, K. L. P. (1999). Second-order biases of the maximum likelihood estimates in von Mises regression models. *Australian and New Zealand Journal of Statistics*, **41,** 901-910.

Cordeiro, G. M., Vasconcellos, K. L. P. and Santos, M. L. F. (1998). On the second-order bias of parameter estimates in nonlinear regression models with Student t errors. *J. Statist. Comp. Simul.*, **60,** 363-378.

Cox, D. R. and Hinkley, D. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B, **49**, 1-39.

Cox, D. and Snell, E. (1968). A general definition of residuals. *J. R. Statist. Soc. B*, **30**, 248-275.

Doornik, J. A. (1999). *Ox: An Object-Oriented Matrix Programming Language*. London: Timberlake Consultants.

Ferrari, S. and Arellano-Valle, R. (1996). Modified likelihood ratio and score tests in linear regression models using the t distribution. *Braz. J. Prob. Statist.*, **10**, 15-33.

Fletcher, R. (1987). *Practical Methods of Optimization*. New York: John Wiley and Sons.

Jeffreys, H. (1939). *Theory of Probability*. Oxford: Clarendon Press.

Lange, K., Little, R. and Taylor, J. (1989). Robust statistical modeling using the t distribution. *J. Am. Statist. Assoc.*, **84**, 881-896.

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

Myers, R. H. (1990). *Classical and Modern Regression with Applications*. California: Duxbury Press.

Pike, M., Hill, A. and Smith, P. (1980). Bias and efficiency in logisitc analysis of stratified case-control studies. *Int. J. Epidemiol.*, **9**, 89-95.

Sutradhar, B. and Ali, M. (1986). Estimation of parameters of a regression model with a multivariate t error variable. *Comm. Statist. - Theory and Methods*, **15**, 429-450.

Vasconcellos, K. L. P. and Cordeiro, G. M. (1997). Approximate bias for multivariate nonlinear heteroscedastic regressions. *Braz. J. Prob. Statist.*, **11,** 141-159.

Vasconcellos, K. L. P. and Cordeiro, G. M. (2000). Bias corrected estimates in multivariate Student t regression models. *Comm. Statist. - Theory and Methods*, **29,** 797-822.

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *J. R. Statist. Soc.* B, **46**, 431-439.

Wolfram, S. (1991). *Mathematica: a System for Doing Mathematics by Computer*. Massachusets: Addison-Wesley.

Young, D. and Bakir, S. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183-191.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *J. Am. Statist. Assoc.*, **71**, 400-405.