RT-MAE 9415

BIAS IN LINEAR REGRESSION MODELS WITH UNKNOWN COVARIANCE MATRIX

by

Elisete da Conceição Q.Aubin
and
Gauss M.Cordeiro

Palavras Chaves: Bias correction; Covariance matrix; Heteroscedastic model; Information matrix; Maximum likelihood estimate; Normal linear model.

Classificação AMS: 62F10 - 62E20
(AMS Classification)

- julho de 1994 -

BIAS IN LINEAR REGRESSION MODELS WITH UNKNOWN COVARIANCE MATRIX

by

Elisete da Conceição Q. Aubin

and

Gauss M. Cordeiro

Palavras Chaves: Bias correction; Covariance matrix; Heteroscedastic model; Information matrix; Maximum likelihood estimate; Normal linear model.

- Julho de 1994 -

BIAS IN LINEAR REGRESSION MODELS WITH UNKNOWN COVARIANCE MATRIX

Elisete da Conceição Q. Aubin
Dept. de Estatística, Universidade de São Paulo,
Caixa Postal 20570, 01452-990, São Paulo, SP, Brazil.

and

Gauss M. Cordeiro
Dept. de Economia, Universidade Federal de Pernambuco,
Cidade Universitária, 50670-901, Recife, PE, Brazil.

Abstract

We investigate the n^{-1} biases of the maximum likelihood estimates from normal linear regression models with unknown error covariance matrix, where n is the sample size. The error covariance matrix is nonscalar and depends on a set of unknown parameters that can be efficiently estimated by maximum likelihood. We give a matrix formula for the n^{-1} biases of the estimates of these parameters. The formula is simple enough to be used algebraically to obtain several closed form expressions in special cases. It has also advantages for numerical purposes.

Keywords : Bias correction; Covariance matrix; Heteroscedastic model; Information matrix; Maximum likelihood estimate; Normal linear model.

1. Introduction

Over the last twenty years, there has emerged a considerable quantity of results concerned with the finite sample properties of nonlinear maximum likelihood estimates. The obvious difficulty with nonlinear estimates is that they cannot be expressed as explicit functions of the data. A very important class of models in statistics and econometrics are normal linear regression models with unknown covariance matrix. The covariance matrix is assumed to depend on a set of unknown nonlinear parameters.

It is well known that maximum likelihood estimates may be biased when the sample size n or the total Fisher information is small. The bias is usually ignored in practice, the justification being that it is negligible when compared to the standard errors of the estimates. In fact, the bias is in general of order n^{-1}, while the asymptotic standard deviation is of order $n^{-1/2}$. However, for some models, the bias terms can be appreciably larger than the corresponding standard deviation terms. In these cases, the bias correction can be important and the availability of formulae for calculating the biases is useful. Exploring this idea, Cordeiro and McCullagh (1991) derived formulae for second-order biases of maximum likelihood estimates of the parameters in generalized linear models (McCullagh and Nelder, 1989).

The main purpose of this paper is to derive a general formula for the large-sample biases of maximum likelihood estimates, in normal linear models with unknown covariance matrix, that can be of direct pratical use of applied researchers. It is possible to find closed form expressions for the second-order biases in terms of the parameters in particular cases. This formula combined with the use of a computer algebra system such as REDUCE (Hearn, 1984) or a language supporting numerical linear algebra, such as APL or GAUSS, will provide the easiest way to compute the biases of the estimates for normal linear models with unknown covariance matrix.

The basic methodology used to derive the biases follows Cox and Snell (1968). The discussion in this paper proceeds as follows. Section 2 presents a formal description of the general normal linear model where the error covariance matrix is nonscalar and unknown. In section 3, we derive a simple matrix formula for the n^{-1} biases of the maximum likelihood estimates of the parameters in the covariance structure. The formula depends only on the model matrix for the mean vector, on the covariance structure and its first and second partial derivatives with respect to their parameters. This formula is also simple enough to obtain several closed form bias corrections in a variety of important models. Finally, in Section 4, we give applications to some special models.

2. Normal Linear Models

We consider the normal linear model where error covariance matrix is nonscalar and depends on a set of unknown parameters that can be efficiently estimated by maximum likelihood. The model is

$$Y = X\beta + U, \tag{1}$$

where Y is an n-dimensional column vector of random variables, X is a known $n \times p$ model matrix of fixed known regressors, β is a p-dimensional column vector of unknown regression parameters, and U is an n-dimensional column vector of unobserved errors. As foundation for estimation by maximum likelihood and hypothesis tests we assume that the random error vector U follows a multivariate normal distribution with zero mean vector and a nonsingular $n \times n$ covariance matrix V^{-1}. The class of models (1) includes many of the important models of autocorrelation and heteroscedasticity discussed in the literature as, for instance, general ARMA models and multiplicative heteroscedastic regression models.

It will be more convenient to work with $n \times n$ precision matrix V. The elements of $V = V(\gamma)$ are known smooth functions of the unknown q-dimensional parameter vector γ. Thus, we have $(p+q)$ parameters for the simultaneous modelling of mean vector and covariance structure. The components of β and γ are unrelated and can vary independently. The parameter space for β is a p-dimensional Euclidean space whereas the parameter space for γ is an open set M in a q-dimensional Euclidean space. Further, we assume that p and q are small compared to n and that $X^T V X$ is positive definite for all γ in Γ. Let $\ell = \ell(\theta)$ be the total log likelihood for $\theta = (\beta^T, \gamma^T)^T$, the $(p+q)$ vector of unknown parameters, given the observable data y.

We have

$$\ell = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)^T V (y - X\beta), \tag{2}$$

of course, some regularity conditions must be specified on the behavior of ℓ as the sample size n approaches infinity. We consider that the function ℓ is regular (Cox and Hinkley, 1974; Chapter 9) with respect to all β and γ derivatives up to and including those of third order. For every sample size, the elements of V are assumed to possess derivatives up to the second order everywhere in the parameter space Γ. In addition, the derivatives of ℓ must behave nicely as n tends to infinity.

An important special case of equation (2) above is then log likelihood for general ARMA models. Expressions for V or V^{-1} for these models are given, for instance, in Shaman (1973), Galbraith and Galbraith (1974) and Ljung and Box (1979). Another special case corresponds to a multiplicative heteroscedastic model in which V^{-1} is a diagonal covariance matrix with diagonal entries $\sigma_1^2,\ldots,\sigma_n^2$, where $\sigma_\ell^2 = \mathrm{Var}(Y_\ell) = \exp(w_\ell^T\gamma)$, w_ℓ^T being a $1\times q$ vector of known constants. This model is quite useful in fields including engineering, economics, and the biological and physical sciences. The variance may also be modeled as a function of the covariates in the expected response as, for instance, $\sigma_\ell^2 = \sigma^2 V(x_\ell^T\beta)$ for $\ell = 1,\ldots,n$, where $V(.)$ is smooth, known and assumed twice differentiable. This formulation leads to a dependent variable heteroscedastic model.

Let $\hat\beta$ and $\hat\gamma$ be the maximum likelihood estimates of β and γ, respectively, and let $\hat\theta = (\hat\beta^T,\hat\gamma^T)^T$ be the full maximum likelihood estimate. We must assume that the estimate $\hat\theta$ converges to the true parameter θ as $n\to\infty$. We define, for $R=1,\ldots,q$, the derivatives $V_R = \partial V/\partial\gamma_R$ and $V^R = \partial V^{-1}/\partial\gamma_R$. The maximum likelihood equations for $\hat\beta$ and $\hat\gamma$ can be written as

$$\hat\beta = (X^T\hat V X)^{-1}X^T\hat V y$$

and

$$\mathrm{tr}(\hat V\hat V_R) = (y - X\hat\beta)^T\hat V_R(y - X\hat\beta)$$

for $R=1,\ldots,q$, where $\hat V = V(\hat\gamma)$ and $\hat V_R = V_R(\hat\gamma)$.These estimates $\hat\beta$ and $\hat\gamma$ can be calculated numerically by any iterative algorithm. Approximations to the biases of $\hat\beta$ and $\hat\gamma$ are developed in the next section.

The emphasis in this article is on demonstrating the usefulness of bias. correction for $\hat\gamma$ to improve the statistical properties of large-sampe maximum likelihood inference procedures for the normal linear models with unknown covariance matrix.

We define joint cumulants of the log likelihood derivatives by $k_{ij} = E(\partial^2 \ell / \partial\theta_i \partial\theta_j)$, $k_{ij\ell} = E(\partial^3 \ell / \partial\theta_i \partial\theta_j \partial\theta_\ell)$ and $k_{ij,\ell} = E\{(\partial^2 \ell / \partial\theta_i \partial\theta_j)(\partial \ell / \partial\theta_\ell)\}$ for $i,j,\ell = 1,\ldots,p+q$. Futhermore, we define the following cumulant derivative $k_{ij}^{(\ell)} = \partial k_{ij} / \partial\theta_\ell$ for $i,j,\ell = 1,\ldots,p+q$. The total Fisher information matrix of order (p+q) for θ is $K = \{-k_{ij}\}$ and let $K^{-1} = \{-k^{ij}\}$ be its inverse. All k's are assumed to be $O(n)$.

Differentiating (2) and taking expectations we can find the joint cumulants of the derivatives of $\ell(\theta)$ with respect to the components of β and γ. We can easily show that $E(-\partial^2 \ell / \partial\beta \partial\gamma) = 0$, i. e., the parameters β and γ are globally orthogonal (Cox and Reid, 1987). The partition $\theta^T = (\beta^T, \gamma^T)$ induces a corresponding block diagonal information matrix $K = \text{diag}\{K_{\beta,\beta}, K_{\gamma,\gamma}\}$ with submatrices $K_{\beta,\beta} = E(-\partial^2 \ell / \partial\beta^2)$ for β and $K_{\gamma,\gamma} = E(-\partial^2 \ell / \partial\gamma^2)$ for γ. Let $K^{-1} = \text{diag}\{K_{\beta,\beta}^{-1}, K_{\gamma,\gamma}^{-1}\}$ be the inverse information matrix. The estimates $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent due to their normality and the block diagonal structure of the information matrix.

3. Bias of Estimate of Parameter γ

In this section our attention is directed to bias correction of the estimate of the parameter γ in model (1). The goal here is to obtain a general matrix expression for the bias of estimate of parameter γ. Consider a model defined as in Section 2. Let $B_1(\hat{\theta}_r)$ be the n^{-1} bias of $\hat{\theta}_r$ for $r = 1,\ldots,p+q$, where θ_r may represent any component of β and γ. From the general expression for the multiparameter n^{-1} biases of the maximum likelihood estimates given by Cox and Snell (1968), we can write, using the notation of Section 2,

$$B_1(\hat{\theta}_r) = \sum k^{ri} k^{j\ell} (k_{ij}^{(\ell)} - \frac{1}{2} k_{ij\ell}), \quad (3)$$

for $r=1,\ldots,p+q$, where the summation $\sum$ is taken over all $p+q$ parameters $\beta_1,\ldots,\beta_p,\gamma_1,\ldots,\gamma_q$.

The interpretation of a superscript is usually clear from the context and never refers to a power in a summation expression.

In the right-hand side of equation (3), which is of order n^{-1}, consistent estimates of the parameters β and γ can be inserted to define the corrected maximum likelihood estimates $\tilde{\theta}_r=\hat{\theta}_r-\hat{B}_1(\hat{\theta}_r)$, for $r=1,\ldots,p+q$, where $\hat{B}_1(.)$ means the value of $B_1(.)$ at the point $\hat{\theta}$. The corrected estimates $\tilde{\theta}_r$'s should have smaller biases than the corresponding uncorrected $\hat{\theta}_r$'s. The corrected estimate $\tilde{\theta}$ has not been used extensively, possibly because of the difficulty of evaluating the expectations to find $B_1(\hat{\theta})$. The bias term $B_1(\hat{\theta})$ is important since, in most instances, a good deal of the bias of $\hat{\theta}$ can be accounted for (and hence corrected for) by the second-order terms in its asymptotic expansion. The calculation of $B_1(\hat{\theta})$ in the maximum likelihood estimates of logistic regression models have been studied by many authors, including Anderson and Richardson (1979), McLachlan (1980), Schaefer (1983) and Copas (1988). Second-order biases of maximum likelihood estimates of the linear parameters and fitted values in generalized linear models were derived recently by Cordeiro and McCullagh (1991). An important paper by Cook et al. (1988) presents these biases for normal nonlinear regression models.

We now derive the n^{-1} biases of the estimates $\hat{\beta}$ and $\hat{\gamma}$ in model (1). From now on we reserve low indices to denote components of the β vector and upper indices for components of the γ vector. Since $k_{rS}=k^{rS}=0$, we have only to take into account for calculating $B_1(\hat{\beta}_i)$ the following two sums:

$$\sum_{\beta} k^{ij}k^{\ell m}\left(k_{j\ell}^{(m)}-\frac{1}{2}k_{j\ell m}\right)$$

and

$$\sum_{\beta,\gamma} k^{ij} k^{LM} (k_{jL}^{(M)} - \frac{1}{2} k_{jLM}), \text{ where}$$

$\sum_{\beta}$ is the summation over the specific components of β only and $\sum_{\beta,\gamma}$ denotes the summation over all combinations of the $(p+q)$ parameters in β and γ. We can easily see from equation (2) and the orthogonality of β and γ that $k_{rsT} = k_{rST} = k_{rs}^{(t)} = k_{rS}^{(T)} = 0$. Then, the last two sums imply that the n^{-1} bias of $\hat{\beta}$ is zero, i. e., $B_1(\hat{\beta}) = 0$. This is to be expected for the normal linear model but it is not obvious that this happens for any covariance matrix V^{-1} since $\hat{\beta}$ is obtained, apart from this case, from a nonlinear equation and because of the dependence of $\hat{\beta}$ on γ and $\hat{V}$. Although the n^{-1} bias of $\hat{\beta}$ equals zero, its higher-order biases may be different from zero.

To evaluate the n^{-1} bias of $\hat{\gamma}$, we can rewrite equation (3) as

$$B_1(\hat{\gamma}_1) = \sum_{\beta,\gamma} k^{Is} k^{tu} (k_{st}^{(u)} - \frac{1}{2} k_{stu}),$$

for any I^{th}. component of γ, where the indices s, t, u vary in all parameters of $\theta^T = (\beta^T, \gamma^T)$. In view of the block diagonality of the information matrix for θ, we can obtain $B_1(\hat{\gamma}_1)$ by evaluating the following two sums:

$$A = \sum_{\beta,\gamma} k^{IR} k^{st} (k_{Rs}^{(t)} - \frac{1}{2} k_{Rst})$$

and

$$B = \sum_{\gamma} k^{IR} k^{ST} (k_{RS}^{(T)} - \frac{1}{2} k_{RST}),$$

where $\sum_{\gamma}$ is the summation ranging from γ_1 to γ_q.

We define the derivatives, for any two components R and S of γ, $V_R = \partial V / \partial \gamma_R$, $V^R = \partial V^{-1} / \partial \gamma_R$, $V_{RS} = \partial^2 V / \partial \gamma_R \partial \gamma_S$ and $V^{RS} = \partial^2 V^{-1} / \partial \gamma_R \partial \gamma_S$ and adopt the following notation:

$$\tilde{V}_R = V^{-1} V_R, \quad \tilde{V}^R = V^R V = -\tilde{V}_R, \quad \tilde{V}_{RS} = V^{-1} V_{RS}, \quad \tilde{V}^{RS} = V^{RS} V,$$
$$m_R = tr(\tilde{V}_R), \quad m^R = tr(\tilde{V}^R) = -m_R, \quad m_{RS} = tr(\tilde{V}_{RS}), m^{RS} = tr(\tilde{V}^{RS}),$$
$$m_{R,S} = tr(\tilde{V}_R \tilde{V}_S), \quad m^{R,S} = tr(\tilde{V}^R \tilde{V}^S),$$
$$m_{RS,T} = tr(\tilde{V}_{RS} \tilde{V}_T), \quad m^{RS,T} = tr(\tilde{V}^{RS} \tilde{V}^T),$$
$$m_{R,S,T} = tr(\tilde{V}_R \tilde{V}_S \tilde{V}_T), \quad m^{R,S,T} = tr(\tilde{V}^R \tilde{V}^S \tilde{V}^T).$$

The m's defined above satisfy certain equations which facilitate their calculation (see Cordeiro and Klein , 1994). For example, $m_{RS} = 2m^{R,S} - m^{RS}$, $m^{RS} = 2m_{R,S} - m_{RS}$, $m_{RS,T} = m^{RS,T} - m^{R,S,T} - m^{R,T,S}$ and $m_{R,S,T} = -m^{R,S,T}$.

We need the following results:

(i) $V^R = -V^{-1} V_R V^{-1}$ and $V_R = -V V^R V$;

(ii) $\frac{\partial \log|V|}{\partial \gamma_R} = tr(\tilde{V}_R) = m_R$;

(iii) $E\{(Y-\mu)^T A (Y-\mu)\} = tr(V^{-1}A)$, where $E(Y) = \mu$ and $Cov(Y) = V^{-1}$, for any positive defined matrix A.

Differentiating (2) and making use of (i)-(iii), we find after some algebra

that $k_{RS}^{(T)} = -\frac{1}{2}(m_{ST,R} + m_{RT,S} - m_{S,T,R} - m_{T,S,R})$,

$k_{Rst} = -x_s^T V_R x_t$, where x_s is the S^{th}. column of the matrix X, and

$k_{RST} = -\frac{1}{2}(m_{RS,T} + m_{ST,R} + m_{RT,S} - m_{S,T,R} - m_{T,S,R})$.

We can obtain a simple expression in matrix notation for $B_1(\hat{\gamma})$ by inserting the cumulants k's and derivatives of cumulants above in sums A and B. The resulting expressions follow elementary matrix calculations. They are given by

$$A = \sum_{\beta,\gamma} k^{IR} k^{st} (k_{RS}^{(t)} - \frac{1}{2} k_{Rst}) = \frac{1}{2} \rho_I^T k_{\gamma,\gamma}^{-1} \tau_{1_{\gamma,\beta}}$$

and

$$B = \sum_{\gamma} k^{IR} k^{ST} (k_{RS}^{(t)} - \frac{1}{2} k_{RST}) = \frac{1}{4} \rho_I^T k_{\gamma,\gamma}^{-1} \tau_{2_\gamma},$$

where ρ_I is a $q \times 1$ vector of zeros but 1 in the I^{th}. component, $\tau_{1_{\gamma,\beta}}$ is a $q \times 1$ vector whose R^{th}. typical component is $tr(K_{\beta,\beta}^{-1} G_\gamma^{(R)})$ with $G_\gamma^{(R)} = X^T V_R X$ being a $p \times p$ matrix, for $R = 1,\ldots,q$, and τ_{2_γ} is a $q \times 1$ vector whose S^{th}. component is $tr(K_{\gamma,\gamma}^{-1} A_\gamma^{(S)})$, $A_\gamma^{(S)}$ being a $q \times q$ matrix with (U,V)th element defined by $2m_{S,U,V} - m_{S,UV}$, for $S,U,V = 1,\ldots,q$.

We can then write in matrix notation

$$B_1(\hat{\gamma}) = K_{\gamma,\gamma}^{-1} \tau_{\gamma,\beta}, \qquad (4)$$

where $\tau_{\gamma.\beta} = \frac{1}{4}(2\tau_{1_{\gamma.\beta}} + \tau_{2_\gamma})$.

A number of remarks are worth making with respect to equation (4). First, $B_1(\hat{\gamma})$ is a function of the model matrix X, the covariance matrix V^{-1}, the precision matrix V and the first and second partial derivatives of either V or V^{-1}. It depends explicitly on the information matrix for γ. For models with closed form information matrix for θ, it is possible obtain a closed form expression for $B_1(\hat{\gamma})$. Although $B_1(\hat{\gamma})$ is easy to compute because its expression involves only simple operations on matrices and vectors, it is not easy to interpret. Second, equation (4) in conjunction with a computer algebra system such as REDUCE (Hearn, 1984) or MAPLE (Char et al., 1988) will give $B_1(\hat{\gamma})$ with minimal effort. Furthermore, we can calculate $B_1(\hat{\gamma})$ numerically via APL or GAUSS. Third, equation (4) allows one to evaluate the influence of second-order terms on the location of estimate $\hat{\gamma}$. For special covariance structures, this term may be relatively large and one may be interested either in estimating it (either directly or by some resampling technique) or at least taking it into account informally when making inferences. The right-hand side of (4) can be estimate at $\hat{\theta}^T = (\hat{\beta}^T, \hat{\gamma}^T)$ to define the bias corrected vector $\tilde{\gamma} = \hat{\gamma} - \hat{B}_1(\hat{\gamma}) = \hat{\gamma} - \hat{K}_{\gamma.\gamma}^{-1}\hat{\tau}_{\gamma.\beta}$. The corrected estimate $\tilde{\gamma}$ would be expected to have better sampling properties than the uncorrected $\hat{\gamma}$. The bias term $\hat{B}_1(\hat{\gamma})$ has the effect of shrinking $\tilde{\gamma}$ towards the true parameter γ.

4. Illustrations

It is useful to consider a couple of simple examples which illustrate the techniques of the previous section and permit a clarification of the notation. Some other important special cases could also be easily obtained because of the advantage of formula (4) which requires only simple operations on matrices.

Equation (4) covers in particular the simplest model when $Y_1, \ldots, Y_n$ are independent and identically distributed random variables in a normal distribution of unknown mean and unknown variance σ^2, $N(\mu, \sigma^2)$, say. In this case, $X = 1$,

$\beta = \mu$, $\gamma = \sigma^2$, $V^{-1} = \gamma I$, $V_\gamma = -\gamma^{-2} I$ and $V_{\gamma,\gamma} = 2\gamma^{-3} I$, where 1 is an $n \times 1$ vector of ones and I is an $n \times n$ identity matrix. The maximum likelihood estimates from a sample $y_1, \ldots, y_n$ are $\hat{\mu} = \bar{y} = \sum y_i / n$ and $\hat{\sigma}^2 = \sum (y_i - \bar{y})^2 / n$. It is quite known that $\hat{\mu} \sim N(\mu, \sigma^2 / n)$ and $\hat{\sigma}^2 \sim \sigma^2 \chi^2_{n-1} / n$ and therefore $E(\hat{\mu}) = \mu$ and $E(\hat{\sigma}^2) = \sigma^2 - \sigma^2 / n$. The information matrix for $\theta = (\beta, \gamma)^T$ has submatrices $K_{\beta,\beta} = n / \gamma$ and $K_{\gamma,\gamma} = n / 2\gamma^2$. We can easily obtain $G^{(\gamma)}_\gamma = -n / \gamma^2$ and $A^{(\gamma)}_\gamma = 0$ and then $\tau_{1_{\gamma,\beta}} = -\gamma^{-1}$ and $\tau_{2_\gamma} = 0$. Using (4) we find $B_1(\hat{\gamma}) = -\gamma / n$. This result agrees with the fact that $E(\hat{\sigma}^2) = \sigma^2 - \sigma^2 / n$ and represents a partial check of equation (4).

As a second example, we consider the multiplicative heteroscedastic model $Y \sim N(X\beta, V^{-1})$ introduced in Section 2, for which $V^{-1} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$ and $\sigma_\ell^2 = \exp(w_\ell^T \gamma)$ for $\ell = 1, \ldots, n$. Cordeiro (1993) has shown thats

$$B_1(\hat{\gamma}) = -(W^T W)^{-1} W^T \tau^*, \qquad (5)$$

where $\tau^* = (A_d + VB_d)1$, $W = (w_1, \ldots, w_n)^T$ is an $n \times q$ matrix, $A = 2WK^{-1}_{\gamma,\gamma}W^T = W(W^T W)^{-1} W^T$ and $B = XK^{-1}_{\beta,\beta}X^T = X(X^T V X)^{-1} X^T$ are $n \times n$ matrices, A_d and B_d are diagonal matrices with the corresponding elements of A and B, respectively, and 1 is an $n \times 1$ vector of ones. The matrices $2A$ and B have simple interpretations as asymptotic covariance structures of $W\hat{\gamma}$ and $X\hat{\beta}$, respectively.

We now show that (5) comes directly from equation (4). For doing this, it is only necessary to prove that $\tau_{\gamma,\beta} = -\frac{1}{2} W^T \tau^*$. We use the notation $D_R = \mathrm{diag}\{w_{1R}, \ldots, w_{nR}\}$ for $R = 1, \ldots, q$. For this model we have $\tilde{V}_R = -D_R$,

$\tilde{V}_{RS} = \tilde{V}_R \tilde{V}_S = D_R D_S$, $\tilde{V}_R \tilde{V}_S \tilde{V}_T = \tilde{V}_{RS} \tilde{V}_T = -D_R D_S D_T$. Also, $A^{(R)} = -W^T D_R W$ and $G^{(R)} = X^T V_R X$ for $R = 1, \ldots, q$. Thus, $\tau_{\gamma.\beta} = \frac{1}{4}(2\tau_{1_{\gamma.\beta}} + \tau_{2_\gamma})$ has a S^{th} typical component given by $\frac{1}{4}\{2\text{tr}(K_{\beta.\beta}^{-1} X^T V_S X) - \text{tr}(K_{\gamma.\gamma}^{-1} W^T D_S W)\}$ which may be written as $-\frac{1}{2}\text{tr}(D_S\{VB + A\})$, for $S = 1, \ldots, q$. On the other hand, $-\frac{1}{2} W^T \tau^* = -\frac{1}{2} W^T (VB_d + A_d)1$ whose S^{th} component is identical to the corresponding component of $\tau_{\gamma.\beta}$. We conclude that (4) is a generalization of (5).

Finally, we condisider a stationary AR(1) model $Y_t - \mu = \rho(Y_{t-1} - \mu) + u_t$, $|\rho| < 1$, where $u_t \sim N(0, \sigma^2)$ for $t = 1, \ldots, n$. Here X reduces to an $n \times 1$ vector of ones, $\beta = \mu$ and $\gamma = (\rho, \sigma^2)^T$. The covariance matrix V^{-1} has a simple form

$$V^{-1} = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ & & \vdots & \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix}$$

whose inverse is

$$V = \sigma^2 \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

We can find $K_{\beta,\beta} = \{2 - 2\rho + (n-2)(1 - 2\rho + \rho^2)\}$ and

$$K_{\gamma,\gamma} = \begin{matrix} \\ \rho \\ \sigma^2 \end{matrix} \begin{matrix} \rho & \sigma^2 \\ \left(\begin{matrix} \frac{1}{2} m_{\rho,\rho} & -\frac{1}{2\sigma^2} m_\rho \\ -\frac{1}{2\sigma^2} m_\rho & \frac{n}{2\sigma^4} \end{matrix} \right) \end{matrix},$$

where $m_\rho = -2\rho / (1 - \rho^2)$ and $m_{\rho,\rho} = 2\{n(1-\rho^2) + 3\rho^2 - 1\} / (1-\rho^2)^2$.

Further,

$$A_\gamma^{(\rho)} = \begin{matrix} \\ \rho \\ \sigma^2 \end{matrix} \begin{matrix} \rho & \sigma^2 \\ \left(\begin{matrix} 2m_{\rho,\rho,\rho} - m_{\rho\rho,\rho} & 2m_{\rho,\rho,\sigma^2} - m_{\rho\sigma^2,\rho} \\ 2m_{\rho,\rho,\sigma^2} - m_{\rho\sigma^2,\rho} & 2m_{\rho,\sigma^2,\sigma^2} - m_{\rho,\sigma^2\sigma^2} \end{matrix} \right) \end{matrix}$$

and

$$A_\gamma^{(\sigma^2)} = \begin{matrix} \\ \rho \\ \sigma^2 \end{matrix} \begin{matrix} \rho & \sigma^2 \\ \left(\begin{matrix} 2m_{\rho,\rho,\sigma^2} - m_{\rho\rho,\sigma^2} & 2m_{\rho,\sigma^2,\sigma^2} - m_{\rho\sigma^2,\sigma^2} \\ 2m_{\rho,\sigma^2,\sigma^2} - m_{\rho\sigma^2,\sigma^2} & 2m_{\sigma^2,\sigma^2,\sigma^2} - m_{\sigma^2\sigma^2,\sigma^2} \end{matrix} \right) \end{matrix},$$

where

$m_{\rho,\rho\rho} = -4(n-2)\rho/(1-\rho^2)^2,$

$m_{\rho,\rho,\rho} = \rho\{6n(1-\rho^2)+14\rho^2-6\}/(1-\rho^2)^3,$

$m_{\rho,\rho,\sigma^2} = m_{\rho,\rho\sigma^2} = -\{2n(1-\rho^2)+6\rho^2-2\}/\sigma^2(1-\rho^2)^2,$

$m_{\rho,\sigma^2,\sigma^2} = m_{\sigma^2,\rho\sigma^2} = -2\rho/\sigma^4(1-\rho^2),$

$m_{\rho\rho,\sigma^2} = -2(n-2)/\sigma^2(1-\rho^2),$

$m_{\rho,\sigma^2\sigma^2} = 4\rho/\sigma^4(1-\rho^2),$

$m_{\sigma^2,\sigma^2\sigma^2} = 2m_{\sigma^2,\sigma^2\sigma^2} = -2n/\sigma^6$. We also have

$G_\gamma^{(\rho)} = 2\sigma^{-2}\{-1+(n-2)(\rho-1)\}$ and

$G_\gamma^{(\sigma^2)} = -\sigma^{-4}\{2(1-\rho)+(n-2)(1-2\rho+\rho^2)\}$. Then,

$$\tau_{1_{\gamma,\beta}} = \{2-2\rho+(n-2)(1-2\rho+\rho^2)\}^{-1}\begin{pmatrix} 2\{(n-2)(\rho-1)-1\}/\sigma^2 \\ -\{2(1-\rho)+(n-2)(1-2\rho+\rho^2)\}/\sigma^4 \end{pmatrix}$$

and

$$\tau_{2_\gamma} = \frac{4}{(nm_{\rho,\rho}-m_\rho^2)}\begin{pmatrix} \{n(2m_{\rho,\rho,\rho}-m_{\rho\rho,\rho})-2m_\rho m_{\rho,\rho}\}/2 \\ \{n(m_{\rho\rho}-2m_{\rho,\rho})-2m_\rho^2\}/2\sigma^2 \end{pmatrix}.$$

We can now obtain a simple expression $B_1(\hat{\gamma})$ by inserting the $K_{\gamma,\gamma}$, $\tau_{1_{\gamma,\beta}}$ and τ_{2_γ}. After some algebra and reducion we find $B_1(\hat{\gamma}) = (-2\rho/n \quad -\sigma^2/n)^T$. This result coincides with Tanaka's (1984) formulae for the n^{-1} biases of $\hat{\rho}$ and $\hat{\sigma}^2$.

References

ANDERSON, J.A. and RICHARDSON, S.C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, **21**, 71 - 78.

CHAR, B.W., GEDDES, K.O., GONNET, G.H., MONAGAN, M.B. and WATT, S.M. (1988). *Maple Reference Manual* (WATCOM, Publications Ltd., Waterloo Canada, 5th ed.).

CORDEIRO, G.M. (1993). Bartlett corrections and bias corrections for two heteroscedastic regression models. *Commun. Statist.- Theory Meth.*, **22(1)**, 169-188.

CORDEIRO, G.M. and KLEIN, R. (1994). Bias correction in ARMA models. *Statistics & Probability 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.

COX, D.R. and HINKLEY, D.V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

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

COX, D.R. and SNELL, E.J. (1968). A general definition of residuals (with discussion). *J. R. Statist. Soc.* B, 30, 248-278.

COOK, R.D., TSAI, C.-L. and WEI, B.C. (1986). Bias in nonlinear regression. *Biometrika*, **73**, 615 - 623.

COPAS, J.B. (1988). Binary regression models for contaminated data.. *J.R. Statist. Soc.* B, **50**, 225 - 265.

GALBRAITH, R.F. and GALBRAITH, J.J. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *J. Appl. Prob.*, **11**, 63-71.

HEARN, A.C. (ed.) (1984). *Reduce User's Manual*, version 3.1 (The Rand Corporation).

LJUNG, G.M. and BOX, G.E.P. (1979). The likelihood function of stationary autoregressive-moving average models. *Biometrika*, 66, 265-270.

McCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

McLACHLAN, G.J. (1980). A note bias correction in maximum likelihood estimation with logistic discrimination.. *Technometrics*, 22, 621 - 627.

SCHAFFER, R.L. (1983). Bias correction in maximum likelihood logistic regression. *Statist. Med.*, 2, 71 - 78.

SHARMAN, P. (1973). On the inverse of the covariance matrix for an autoregressive-moving average process. *Biometrika*, 60, 193-196.

TANAKA, K. (1984). An asymptotic expansion associated with the maximum likelihood estimators in ARMA models, *J. R. Statist. Soc.* B, 46, 58-67.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

9401 - PAULA, G. A., SEN P.K., Tests of ordered Hypotheses in Linkage in Heredity, 12p.

9402 - PAULA, G. A., SEN P. K. One-Sided Tests in Generalized Linear Models with Parallel Regression Lines, 18p.

9403 - PAULA, G. A., .Influence Diagnostics in Proper Dispersion Models,22 p.

9404 - SALINAS-TORRES, V.H. & PEREIRA, C.A.B. and TIWARI, R.C. - Convergence of Dirichlet Measures Arising in Context of Bayesian Analysis of Competing Risk Models, 17p.

9405 - CRIBARI-NETO, F. & FERRARI, S.L.P. - Second Order Asymptotics for Score Tests in Generalised Linear Models, 25p.

9406 - ZUAZOLA, P.L.I., PEREIRA, C.A.B. & WECHSLER, S. - Predictivistic Statistical Inference in Finite Populations, 10p.

9407 - BOLFARINE, H. & ROJAS, M.G. - Maximum Likelihood Estimation of Simultaneous Pairwise Linear Structural Relationships, 19p.

9408 - FERRARI, P.A. & KIPNIS, C. - Second Class Particles in the Rarefaction Fan, 9p.

9409 - ANDRÉ, C.D.S., NARULA, S.C. & PERES, C.A. - Asymptotic Properties of MSAE Estimates in Multistage Model. Preliminary Results, 19p.

9410 - ARELLANO-VALLE, R.B. & BOLFARINE, H. - On Some Characterizations of the t-Distribution, 12p.

9411 - NEVES, E.J., A Discrete Variational Problem Related to Ising Droplets at Low Temperatures, 16p.

9412 - FERRARI, S.L.P. & CORDEIRO, G.M. - Corrected Score Tests for Exponential Family Nonlinear Models, 14p.

9413 - ANDRADE, D.F. & SINGER, J.M. - Profile Analysis for Randomized Complete Block Experiments, 8p.

9414 - BOLFARINE, H. & ROJAS, M.G. - On Structural Comparative Calibration, 13p.

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

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01452-990 - São Paulo, Brasil