RT-MAE 9417

A NOTE ON THE SIMPLE STRUCTURAL REGRESSION MODEL

by

R.B. Arellano
and
H. Bolfarine

Palavras Chaves: **Bartlett correction factor; Likelihood ratio statistics; maximum likelihood estimation; orthogonal parametrization.**

Classificação AMS: 62F05 - 62J05
(AMS Classification)

- agosto de 1994 -

A NOTE ON THE SIMPLE STRUCTURAL REGRESSION MODEL

R.B. Arellano-Valle and H. Bolfarine
Departamento de Estatística -IME
Universidade de São Paulo - Caixa Postal 20570
CEP 01492-990 - São Paulo - SP - BRASIL

Summary

In this paper we investigate some aspects like estimation and hypothesis testing in the simple structural regression model with measurement error. Use is made of orthogonal parametrization obtained in the literature. Emphasis is placed on the some properties of the maximum likelihodd estimators and also on the distribution of the likelihood ratio statistics.

1. Introduction

The classical simple regression model with measurement errors is defined by the equations

$$
(1.1) \qquad \begin{cases} Y_k = y_k + e_k, \\ X_k = x_k + u_k, \\ y_k = \alpha + \beta x_k, \end{cases}
$$

where e_k and u_k are independent and normally distributed with zero means and variances σ_e^2 and σ_u^2, respectively, which we denote by

$$
\begin{pmatrix} e_k \\ u_k \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right),
$$

$k = 1, \ldots, n$, where N_2 denotes the bivariate normal distribution If the quantity x_k is considered to be a fixed quantity then, the functional regression model follows. On the other hand, if the quantity x_k is considered to be a random quantity, then the structural regression model follows. In this paper, we consider $x_k \sim N(\mu_x, \sigma_x^2)$, with x_k independent of (e_k, u_k), $k = 1, \ldots, n$, a typically made assumption. The main idea behind the equations (1.1) is that $(y_1, x_1), \ldots, (y_n, x_n)$ are not observed directly and the estimation has to be based on $(Y_1, X_1), \ldots, (Y_n, X_n)$, which are observed. Examples of practical situations where the x_k are measured with error are reported in Fuller (1987). An interesting situation is the case where x_k is the amount of nitrogen in the soil and y_k is the yield of a certain cereal. In this case, the observed X_k values are determined by laboratory analysis and are only estimates of the true x_k values.

As is well known, there are problems with the estimation of the parameters in both cases. In the funcional case, β is not consistently estimated. In the structural case, some nonidentifiability problems arise. See, for example, Fuller (1987) and Kendall and Stuart (1979), where extensive bibliografies are provided. A Bayesian treatment for the problem can be found in Zellner (1971). Therefore, in order to make the estimation problem feasible, some additional assumptions have to be considered. In the structural model, a typically made assumptions considers that the reliability ratio (Fuller, 1987) $k_x = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$, or equivalently, $\lambda_x = \sigma_x^2/\sigma_u^2$ is known. Fuller (1987) reports several situations particularly in Sociology, Psychology and Survey Sampling where k_x is so well estimated that it may be taken to be known. Bolfarine and Cordani (1993) derived an orthogonal parametrization in this case and investigated the performance of confidence intervals for β. Another common assumption is to consider that the ratio of the two variances $\lambda_e = \sigma_e^2/\sigma_u^2$ is known. This case has been investigated by Wong (1989) where an orthogonal parametrization is derived and Bartlett correction factors are provided for the likelihood ratio statistic by using the approach of Lawley (1956).

In this paper, a unified approach is developed for both (λ_x known and λ_e known) cases. By studying the distribution of the maximum likelihood estimators of the orthogonal parameters, we investigate the distribution of the likelihood ratio criteria in both cases. The approach also makes it possible to compute directly the expected value of the likelihood ratio criteria to order n^{-2}. As shown, the correction factors obtained are exactly the same in both cases and coincide with the one obtained by Wong (1989).

Section 2 presents a general matrix representation for the model and the orthogonal parametrization under both assumptions. Section 3 discusses maximum likelhood estimation and some properties of the estimators derived are studied in Section 4. Section 5 investigates Bartlett correction factors for the likelihood ratio statistics.

2. Orthogonal parametrizations

Note that we may rewritte model (1.1) as

$$\mathbf{Z}_k = \mathbf{g}_k + \epsilon_k, \tag{2.1}$$

where

$$\mathbf{Z}_k = \begin{pmatrix} Y_k \\ X_k \end{pmatrix}, \quad \mathbf{g}_k = \mathbf{g}(x_k) = \begin{pmatrix} \alpha + \beta x_k \\ x_k \end{pmatrix},$$

and

$$\epsilon_k = \begin{pmatrix} e_k \\ u_k \end{pmatrix},$$

$k = 1, \ldots, n$. Thus, from (2.1), we have that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are independent and identically distributed with

$$\mathbf{Z}_k \sim N_2(\mu; \mathbf{\Sigma}),$$

where

$$\mu = E[\mathbf{Z}_k] = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix} = \begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \tag{2.2}$$

and

$$\mathbf{\Sigma} = Cov[\mathbf{Z}_k] = \begin{cases} \begin{pmatrix} \lambda_x\beta^2\sigma_u^2 + \sigma_e^2 & \lambda_x\beta\sigma_u^2 \\ \lambda_x\beta\sigma_u^2 & (\lambda_x + 1)\sigma_u^2 \end{pmatrix}, & \text{if } \lambda_x \text{ is known,} \\ \begin{pmatrix} \beta^2\sigma_x^2 + \lambda\sigma_u^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{pmatrix}, & \text{if } \lambda_e \text{ is known.} \end{cases} \tag{2.3}$$

Further, it can be shown that

$$
(2.4) \qquad |\boldsymbol{\Sigma}| = \begin{cases} [\lambda_x\beta^2\sigma_u^2 + (\lambda_x + 1)\sigma_\epsilon^2]\sigma_u^2, & \text{if } \lambda_x \text{ is known,} \\ [\lambda_e\sigma_u^2 + (\beta^2 + \lambda_e)\sigma_x^2]\sigma_u^2, & \text{if } \lambda_e \text{ is known.} \end{cases}
$$

Let

$$
(2.5) \qquad \theta = \begin{cases} (\alpha, \mu_x, \sigma_\epsilon^2, \sigma_u^2, \beta), & \text{if } \lambda_x \text{ is known,} \\ (\alpha, \mu_x, \sigma_x^2, \sigma_u^2, \beta), & \text{if } \lambda_e \text{ is known.} \end{cases}
$$

and $l = l(\theta)$, the log likelihood function may be written as

$$
l \propto -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{k=1}^{n}(\mathbf{Z}_k - \mu)'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_k - \mu),
$$

where $\mu(\theta)$ e $\boldsymbol{\Sigma}(\theta)$ are as given in (2.2) and (2.3), respectively. Let $\mathbf{K}(\theta) = [\kappa_{i,j}]$ denote the expected information matrix. Then, after some algebraic manipulations, it can be shown that

$$
\kappa_{i,j} = n[\frac{\partial\mu}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\mu}{\partial\theta_j} + \frac{n}{2}tr(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j})],
$$

where θ_i denotes the i-th component of θ, as defined in (2.5), and from where it follows that

$$
\kappa_{i,j} = \begin{cases} n\frac{\partial\mu'}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\mu}{\partial\theta_j}, & \text{if } i = 1,2,5,\ j = 1,2, \\ \frac{n}{2}tr(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}), & \text{if } i = 3,4,5 \text{ and } j = 3,4, \\ 0, & \text{if } i = 1,2 \text{ and } j = 3,4, \end{cases}
$$

and

$$
\kappa_{5,5} = \kappa_{\beta,\beta} = n\frac{\partial\mu'}{\partial\beta}\boldsymbol{\Sigma}^{-1}\frac{\partial\mu}{\partial\beta} + \frac{n}{2}tr\{(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\beta})^2\},
$$

where $tr(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$. It follows that $\kappa_{\beta,j} \neq 0$, whatever be θ_j. This fact makes it hard to obtain large sample inference for β, particularly correction factors for testing statistics. One way of aliviating this difficulty is to consider an orthogonal transformation of θ, that is, transforming θ into $\phi = (\phi_1, \phi_2, \phi_3, \phi_4, \beta)'$ so that $\theta_i = \theta_i(\phi)$, $i = 1,2,3,4$, are the solutions to the differential equations:

$$
(2.6) \qquad \sum_{i=1}^{4}\kappa_{i,j}\frac{\partial\theta_i}{\partial\beta} = -\kappa_{\beta,j},
$$

$j = 1,2,3,4$. Typically, solving a systhem like the one in (2.6) is not simple. Moreover, when solvable, such equations may not be always easily interpretable. In the case when λ_e is known, a solution is given in Wong (1989) and when λ_x is known, a solution is given in Bolfarine and Cordani (1993). We note that the problem of obtaining the orthogonal parametrization can be simplified by first making μ orthogonal to $\boldsymbol{\Sigma}$. This is easily accomplished by taking $\phi_0 = \alpha + \beta\mu_x$ and $\phi_1 = \mu_x$. The problem now is to make β orthogonal to the other parameters which appear in $\boldsymbol{\Sigma}$.

The solution presented in Wong (1989) and Bolfarine and Cordani (1993) may be written as

$$\phi_1 = \alpha + \beta\mu_x, \quad \phi_2 = \mu_x, \quad \phi_4 = \sigma_u^2. \tag{2.7}$$

$$\phi_3 = \begin{cases} \lambda_x\beta^2\sigma_u^2 + (\lambda_x + 1)\sigma_e^2, & \text{if } \lambda_x \text{ is known,} \\ (\beta^2 + \lambda_e)\sigma_x^2 + \lambda_e\sigma_u^2, & \text{if } \lambda_e \text{ is known.} \end{cases} \tag{2.8}$$

Considering the above parametrization, we have that

$$\mu = \mu(\phi_L) = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

and

$$\Sigma = \Sigma(\phi_S) = \begin{cases} (\lambda_x + 1)^{-1}\begin{pmatrix} \phi_3 + (\lambda_x\beta)^2\phi_4 & (\lambda_x + 1)\lambda_x\beta\phi_4 \\ (\lambda_x + 1)\lambda_x\beta\phi_4 & (\lambda_x + 1)^2\phi_4 \end{pmatrix}, & \text{if } \lambda_x \text{ is known,} \\ (\beta^2 + \lambda_e)^{-1}\begin{pmatrix} \beta^2\phi_3 + \lambda_e^2\phi_4 & \beta(\phi_3 - \lambda_e\phi_4) \\ \beta(\phi_3 - \lambda_e\phi_4) & \phi_3 + \beta^2\phi_4 \end{pmatrix}, & \text{if } \lambda_e \text{ is known,} \end{cases} \tag{2.9}$$

where $\phi_L = (\phi_1, \phi_2)'$ (the location parameters) and $\phi_S = (\phi_3, \phi_4, \beta)'$ (the scale parameters). Note that

$$|\Sigma| = \phi_3\phi_4. \tag{2.10}$$

We call attention to the fact that the choice of the scale parameters are not as obvious as the location parameters. However, the choice of the news parameters becomes obvious and clear from (2.4). Moreover, when λ_e is known and taken to be equal to one (without loss of generality), it can be shown that

$$tr(\Sigma) = \phi_3 + \phi_4,$$

so that ϕ_3 and ϕ_4 are the caracteristic roots of Σ. In the sequel, we present some properties of the matrix Σ wich will make it easier to derive the cumulants of the derivatives of the log likelihood function $l = l(\phi)$.

Let

$$a_3 = \begin{cases} (\lambda_x + 1)^{-1/2}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, & \text{if } \lambda_x \text{ is known,} \\ (\beta^2 + \lambda_e)^{-1/2}\begin{pmatrix} \beta \\ 1 \end{pmatrix}, & \text{if } \lambda_e \text{ is known} \end{cases}$$

and

$$a_4 = \begin{cases} (\lambda_x + 1)^{-1/2}\begin{pmatrix} \lambda_x\beta \\ \lambda_x \end{pmatrix}, & \text{if } \lambda_x \text{ is known,} \\ (\beta^2 + \lambda_e)^{-1/2}\begin{pmatrix} \lambda_e \\ -\beta \end{pmatrix}, & \text{if } \lambda_e \text{ is known.} \end{cases}$$

and note that

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_i} = \alpha_i \alpha_i',$$

$i = 3, 4$, and

$$\boldsymbol{\Sigma} = \phi_3 \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_3} + \phi_4 \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_4} = \phi_3 \alpha_3 \alpha_3' + \phi_4 \alpha_4 \alpha_4'.$$

Similarly,

$$\boldsymbol{\Sigma}^{-1} = \phi_3 \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_3} \boldsymbol{\Sigma}^{-1} + \phi_4 \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_4} \boldsymbol{\Sigma}^{-1} = \phi_3^{-1} \tilde{\alpha}_3 \tilde{\alpha}_3' + \phi_4^{-1} \tilde{\alpha}_4 \tilde{\alpha}_4',$$

and

$$\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_i} = \phi_i^{-1} \tilde{\alpha}_i \alpha_i'$$

where

(2.11) $$\tilde{\alpha}_i = \phi_i \boldsymbol{\Sigma}^{-1} \alpha_i,$$

$i = 3, 4$, are such that $\alpha_i' \tilde{\alpha}_i = 1$, $\alpha_i' \tilde{\alpha}_j = 0$, $||\alpha_i|| = ||\alpha_j||$, $i \neq j = 3, 4$; that is,

$$\alpha_3 = \begin{cases} (\lambda_r + 1)^{-1/2} \begin{pmatrix} \lambda_r + 1 \\ -\lambda_r \beta \end{pmatrix}, & \text{if } \lambda_r \text{ is known} \\ (\beta^2 + \lambda_e)^{-1/2} \begin{pmatrix} \beta \\ \lambda_e \end{pmatrix}, & \text{if } \lambda_e \text{ is known,} \end{cases}$$

and

$$\alpha_4 = \begin{cases} (\lambda_r + 1)^{-1/2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & \text{if } \lambda_r \text{ is known,} \\ (\beta^2 + \lambda_e)^{-1/2} \begin{pmatrix} 1 \\ -\beta \end{pmatrix}, & \text{if } \lambda_e \text{ is known.} \end{cases}$$

Furthermore, it is easy to see that

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \beta} = \phi_3 \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \phi_3 \partial \beta} + \phi_4 \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \phi_4 \partial \beta} = \left(\frac{\phi_3 \phi_4}{\sigma_\beta^2}\right)^{1/2} (\alpha_3 \alpha_4' + \alpha_4 \alpha_3')$$

and

$$\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \beta} = \left(\frac{\phi_3 \phi_4}{\sigma_\beta^2}\right)^{1/2} (\phi_3^{-1} \tilde{\alpha}_3 \alpha_4' + \phi_4^{-1} \tilde{\alpha}_4 \alpha_3'),$$

where

(2.12) $$\sigma_\beta^2 = \begin{cases} \frac{\phi_3}{\lambda_r^2 \phi_4}, & \text{if } \lambda_r \text{ is known,} \\ \left(\frac{\beta^2 + \lambda_e}{\phi_3 - \lambda_e \phi_4}\right)^2 \phi_3 \phi_4, & \text{if } \lambda_e \text{ is known.} \end{cases}$$

From (2.9) it follows that (see also (2.10))

$$tr(\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_i}\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_j}) = \begin{cases} \phi_i^{-2}, & i = j \\ 0, & i \neq j, \end{cases} \tag{2.13}$$

$i, j = 3, 4$, and

$$tr\{(\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\beta})^2\} = tr(\mathbf{\Sigma}^{-1}\frac{\partial^2\mathbf{\Sigma}}{\partial\beta^2}) = 2\sigma_\beta^{-2}. \tag{2.14}$$

Let now $\mathbf{K} = \mathbf{K}(\phi) = [\kappa_{i,j}]$ the information matrix under the orthogonal parametrization. Then,

$$\kappa_{i,j} = \begin{cases} n\frac{\partial\mu'}{\partial\phi_i}\mathbf{\Sigma}^{-1}\frac{\partial\mu}{\partial\phi_j}, & i, j = 1, 2, \\ \frac{n}{2}tr\{(\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_i})^2\}, & i = 3, 4, 5, \\ 0, & i = 1, 2,\ j = 3, 4, 5, \end{cases}$$

where $\mu = \mu(\phi_L)$ and $\mathbf{\Sigma} = \mathbf{\Sigma}(\phi_S)$ are as defined above. Thus, $\mathbf{K} = \mathrm{Diag}(\mathbf{K}_L, \mathbf{K}_S)$, that is, $\mathbf{K}$ is a block diagonal matrix with

$$\mathbf{K}_L = n\mathbf{\Sigma}^{-1}$$

and, using (2.13) and (2.14),

$$\mathbf{K}_S = n\mathrm{Diag}(\frac{1}{2\phi_3^2}, \frac{1}{2\phi_4^2}, \frac{1}{\sigma_\beta^2}),$$

where σ_β^2 is given in (2.12). Note that

$$\sigma_\beta^2 = \beta^2\frac{\sigma_{XX}\sigma_{YY.X}}{\sigma_{YX}^2} = \beta^2(\frac{1-\rho_{YX}^2}{\rho_{YX}^2}),$$

where $\sigma_{YX} = Cov[Y_i, X_i]$, $\sigma_{XX} = Var[X_i]$, $\sigma_{YY.X} = Var[Y_i|X_i]$ and

$$\rho_{YX} = \frac{\sigma_{YX}}{(\sigma_{XX}\sigma_{YY})^{1/2}},$$

which denotes the correlation between Y_k and X_k, $k = 1, \ldots, n$.

3. Maximum likelihood estimators

The maximum likelihood estimator $(\hat{\phi}_L, \hat{\phi}_S)$ of (ϕ_L, ϕ_S) is obtained by solving the equations

$$\frac{\partial l}{\partial\phi_i}|_{\phi=\hat{\phi}} = \sum_{k=1}^{n}\frac{\partial\mu'}{\partial\phi_i}\mathbf{\Sigma}^{-1}(\mathbf{Z}_k - \mu)|_{\phi=\hat{\phi}} = 0,$$

$i = 1, 2$,

$$\frac{\partial l}{\partial\phi_i}|_{\phi=\hat{\phi}} = \{-\frac{n}{2}tr(\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_i}) + \frac{1}{2}\sum_{k=1}^{n}(\mathbf{Z}_k - \mu)'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\phi_i}\mathbf{\Sigma}^{-1}(\mathbf{Z}_k - \mu)\}|_{\phi=\hat{\phi}} = 0,$$

$= 3, 4$, and

$$\frac{\partial l}{\partial \beta}\Big|_{\phi=\hat\phi} = \frac{1}{2}\sum_{k=1}^{n}(\mathbf{Z}_k - \mu)'\Sigma^{-1}\frac{\partial \Sigma}{\partial \beta}\Sigma^{-1}(\mathbf{Z}_k - \mu)\Big|_{\phi=\hat\phi} = 0.$$

The first equation leads to $\hat\mu = \mu(\hat\phi_L) = \bar{\mathbf{Z}}$, from where we get

$$\hat\phi_1 = \bar Y, \quad \text{and} \quad \hat\phi_2 = \bar X,$$

since $\mu = (\phi_1, \phi_2)'$ and $\bar{\mathbf{Z}} = \sum_{k=1}^{n} \mathbf{Z}_k/n = (\bar Y, \bar X)'$. Using these estimators, it follows from the above equations that

$$\hat\phi_i = \bar\alpha_i'(\hat\beta)\mathbf{S}\bar\alpha_i(\hat\beta),$$

$= 3, 4$, and

$$\bar\alpha_3'(\hat\beta)\mathbf{S}\bar\alpha_4(\hat\beta) = 0,$$

respectively, where $\bar\alpha_i(\hat\beta)$ is as defined in (2.11), with β replaced by $\hat\beta$ and

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})' = \begin{pmatrix} S_{YY} & S_{YX} \\ S_{YX} & S_{XX} \end{pmatrix},$$

where $S_{XX} = \sum_{i=1}^{n}(X_i - \bar X)^2/n$, $S_{YY} = \sum_{i=1}^{n}(Y_i - \bar Y)^2/n$ and $S_{YX} = \sum_{i=1}^{n}(X_i - \bar X)(Y_i - \bar Y)/n$. From the above equations, it follows when λ_x is known that

$$\hat\phi_3 = (\lambda_x + 1)S_{YY} - 2(\lambda_x\hat\beta)S_{YX} + (\lambda_x\hat\beta^2)\hat\phi_4,$$

$$\hat\phi_4 = \frac{S_{XX}}{\lambda_x + 1},$$

$$\hat\beta = \left(\frac{\lambda_x + 1}{\lambda_x}\right)\frac{S_{YX}}{S_{XX}}.$$

Replacing $\hat\phi_4$ and $\hat\beta$ in $\hat\phi_3$, we have that

$$\hat\phi_3 = (\lambda_x + 1)S_{YY.X},$$

where

$$S_{YY.X} = S_{YY} - S_{XX}^{-1}S_{YX}^2 = S_{YY}(1 - r_{YX}^2),$$

and

$$r_{YX} = \frac{S_{YX}}{(S_{YY}S_{XX})^{1/2}}.$$

When λ_e is known, it follows that

$$\hat\phi_3 = \frac{\hat\beta^2 S_{YY} + 2\lambda_e\hat\beta S_{YX} + \lambda_e^2 S_{XX}}{\hat\beta^2 + \lambda_e},$$

$$\hat{\phi}_4 = \frac{S_{YY} - 2\hat{\beta}S_{YX} + \hat{\beta}^2 S_{XX}}{\hat{\beta}^2 + \lambda_e},$$

$$\hat{\beta} = \left(\frac{S_{YY} - \lambda_e S_{XX}}{2S_{YX}}\right) + \{(\frac{S_{YY} - \lambda_e S_{XX}}{2S_{YX}})^2 + \lambda_e\}^{1/2}.$$

It can also be shown that the maximum likelihood estimator of $\phi_S = (\phi_3, \phi_4, \beta)'$ is given by the solution of the equation

$$\Sigma(\hat{\phi}_S) = \mathbf{S},$$

where $\Sigma(\hat{\phi}_S)$ is as given in (2.9), with ϕ_S replaced by $\hat{\phi}_S$. Some estimators may also be given alternative expressions as, for example,

$$\hat{\phi}_3 = \begin{cases} (\lambda_x + 1)S_{YY} - (\lambda_x\hat{\beta})^2\hat{\phi}_4, & \text{if } \lambda_x \text{ is known,} \\ \lambda_e S_{YY} - \hat{\beta}S_{YX}, & \text{if } \lambda_e \text{ is known.} \end{cases}$$

Note also that

(3.1) $$\hat{\phi}_3\hat{\phi}_4 = |\hat{\Sigma}| = S_{YY.X}S_{XX}.$$

4. Some properties of the maximum likelihood estimators

In this section we study some properties of $\hat{\phi} = (\phi'_L, \phi'_S)'$. Under model (1.1), it follows that

$$\hat{\mu} = \bar{\mathbf{Z}} \sim N_2(\mu, \frac{1}{n}\Sigma).$$

and

(4.1) $$\hat{\Sigma} = \mathbf{S} \sim W_2(\frac{1}{n}\Sigma, n-1),$$

are independent, where

$$\mu = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_{YY} & \sigma_{YX} \\ \sigma_{YX} & \sigma_{XX} \end{pmatrix},$$

with $\mathbf{Z}$ and $\mathbf{S}$ as before. Here, $W_k(\mathbf{A}, m)$ denotes the k-variate Wishart distribution with dispersion matrix $\mathbf{A}$ and m degrees of freedom (Muirhead,1982). From these results, it follows that $\hat{\phi}_L = (\hat{\phi}_1, \hat{\phi}_2)'$ and $\hat{\phi}_S = (\hat{\phi}_2, \hat{\phi}_4, \hat{\beta})'$ are independent and

$$\hat{\phi}_L = \bar{\mathbf{Z}} \sim N_2(\phi_L, \frac{1}{n}\Sigma).$$

Thus, confidence regions or hypothesis testing for $H_0 : \phi_L = \gamma_0$ can be performed by considering the variable

$$F = \left(\frac{n-2}{2}\right)(\hat{\phi}_L - \phi_L)'(\Sigma^{-1}(\hat{\phi}_S))(\hat{\phi}_L - \phi_L) = \left(\frac{n-2}{2}\right)(\bar{\mathbf{Z}} - \mu)'\mathbf{S}(\bar{\mathbf{Z}} - \mu) \sim F_{2,n-2},$$

that is, F is ditributed according to the Fisher F distribution with 2 and $n-2$ degrees of freedom. Furthermore, confidence intervals or hypothesis testing for functions of the form $\mathbf{a}'\phi_L$ where $\mathbf{a}$ is a known vector of constants, we can use

$$\sqrt{n}\frac{(\mathbf{a}'\hat{\phi}_L - \mathbf{a}'\phi_L)}{(\mathbf{a}'\mathbf{S}\mathbf{a})^{1/2}} \sim t(0,1;n-1),$$

where $t(0,1;n-1)$ denotes the central t distribution with $n-1$ degrees of freedom. On the other hand, the exact marginal distribution of the components of the vector $\hat{\phi}_S$ is particularly difficult to obtain when λ_e is known. For example, when $\lambda_e = 1$, $\hat{\phi}_3$ and $\hat{\phi}_4$ are the proper values of the Wishart matrix $\Sigma(\hat{\phi}_S)$. Thus, in this situation, the distribution function of $(\hat{\phi}_3, \hat{\phi}_4)$ can be represented in terms of infinite series (Muirhead, 1982). Similarly, the distribution function of $\hat{\beta}$ can be is given in a form of a convergent infinite series of incomplete beta functions (Anderson and Sawa, 1982 derived the distribution of $\hat{\beta}$ in the functional case. This distribution corresponds to the conditional distribution of $\hat{\beta}$ given $\mathbf{x} = (x_1, \dots, x_n)'$ in the structural case). For this reason, inference on β when λ_e is known typically is based on large samples, since as $n \to \infty$,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma_\beta^2).$$

Moreover, using properties of the Wishart distribution we can study the exact distributions of certain functions of $\hat{\phi}_S$ in both cases (λ_x or λ_e known). Thus, considering the Wishart distribution in (4.1) it follows that (Muirhead, 1982)

(i) $S_{YY.X} = S_{YY} - \frac{S_{YX}^2}{S_{XX}} = S_{YY}(1 - r_{YX}^2)$ is independent of (S_{YX}, S_{XX});

(ii)

$$\frac{nS_{YY.X}}{\sigma_{YY.X}} \sim \chi^2_{n-2},$$

where

$$\sigma_{YY.X} = \sigma_{YY} - \frac{\sigma_{YX}^2}{\sigma_{XX}} = \sigma_{YY}(1 - \rho_{YX}^2);$$

(iii)

$$\left(\frac{S_{YX}}{S_{XX}} - \frac{\sigma_{YX}}{\sigma_{XX}}\right)|S_{XX} \sim N(0, \frac{\sigma_{YY.X}}{nS_{XX}});$$

(iv)

$$\frac{nS_{YY}}{\sigma_{YY}} \sim \chi^2_{n-1} \quad \text{and} \quad \frac{nS_{XX}}{\sigma_{XX}} \sim \chi^2_{n-1}.$$

From the above results, it follows that

(v)

$$\frac{S_{YX}}{S_{XX}} \sim t\left(\frac{\sigma_{YX}}{\sigma_{XX}}, \frac{\sigma_{YY.X}}{(n-1)\sigma_{XX}}; n-1\right)$$

and
(vi)

$$\left(\frac{(n-2)S_{XX}}{S_{YY.X}}\right)^{1/2}\left(\frac{S_{YX}}{S_{XX}}-\frac{\sigma_{YX}}{\sigma_{XX}}\right) \sim t(0,1;n-2).$$

Notice that (v) follow from (iii) and (iv), since, as is well known, by mixing a normal with a chisquared distribution we get a t distribution. On the other hand, (vi) segue de (i), (ii) and (iii). Thus, (vi) allows the making of inference on the ratio (function of ϕ_S)

$$\frac{\sigma_{YX}}{\sigma_{XX}} = \begin{cases} \left(\frac{\lambda_x}{\lambda_x+1}\right)\beta, & \text{if } \lambda_x \text{ is known,} \\ \left(\frac{\phi_3-\lambda_e\phi_4}{\phi_3+\beta^2\phi_4}\right)\beta, & \text{if } \lambda_e \text{ is known.} \end{cases}$$

For example, for testing $H_0 : \beta = 0$ (indepence between X and Y), we can use the fact that

$$\frac{\sqrt{n-2}\hat{\beta}}{\hat{\sigma}_\beta} = \frac{\sqrt{n-2}r_{YX}}{\sqrt{1-r_{YX}^2}} = \frac{\sqrt{n-2}S_{YX}}{\sqrt{S_{XX}S_{YY.X}}} \sim t(0,1;n-2),$$

where $\sigma_\beta = \sigma_\beta(\phi_S)$ is as defined in (2.12). From the above results, we also have that

$$E[\hat{\phi}_3\hat{\phi}_4] = E[S_{YY.X}]E[S_{XX}] = \left(\frac{n-2}{n}\right)\left(\frac{n-1}{n}\right)\phi_3\phi_4, \tag{4.2}$$

since

$$S_{YY.X}S_{XX} = |\mathbf{S}| = |\Sigma(\hat{\phi}_S)| = \hat{\phi}_3\hat{\phi}_4$$

and

$$\sigma_{YY.X}\sigma_{XX} = |\Sigma| = \phi_3\phi_4.$$

Moreover,

$$E[\log(\frac{\hat{\phi}_3\hat{\phi}_4}{\phi_3\phi_4})] = E[\log(\frac{S_{YY.X}S_{XX}}{\sigma_{YY.X}\sigma_{XX}})]$$

$$= E[\log(\frac{S_{YY.X}}{\sigma_{YY.X}}) + E[\log(\frac{S_{XX}}{\sigma_{XX}})]$$

$$= \psi(\frac{n-2}{2}) + \psi(\frac{n-1}{2}) - 2\log\frac{n}{2}, \tag{4.3}$$

which follows from the fact that if $V \sim Ga(\nu, \delta)$, the gamma distribution with parameters ν and δ, then $E[\log V] = \psi(\nu) - \log\delta$, with $\psi(.)$ being the digamma function. If $\lambda_e = 1$ it follows that

$$\hat{\phi}_3 + \hat{\phi}_4 = tr\{\Sigma(\hat{\phi}_E)\} = tr(\mathbf{S}) = S_{YY} + S_{XX},$$

from where it follows that

$$E[\hat{\phi}_3 + \hat{\phi}_4] = \left(\frac{n-1}{n}\right)(\phi_3 + \phi_4).$$

Finally, considering λ_x known, we have that

$$\phi_3 = (\lambda_x + 1)\sigma_{YY.X}, \quad \phi_4 = \frac{\sigma_{XX}}{\lambda_x + 1}, \quad \beta = (\frac{\lambda_x + 1}{\lambda_x})\frac{\sigma_{YX}}{\sigma_{XX}},$$

so that the maximum likelihood estimators of the above parameters are given by

$$\hat{\phi}_3 = (\lambda_x + 1)S_{YY.X}, \quad \hat{\phi}_4 = \frac{S_{XX}}{\lambda_x + 1}, \quad \hat{\beta} = \frac{S_{YX}}{S_{XX}} = \frac{S_{YX}}{\lambda_x \hat{\phi}_4}.$$

Considering the above relations, when λ_x is known, we have that

(i) $\hat{\phi}_3$, $\hat{\phi}_4$ and $\hat{\phi}_4^{1/2}(\hat{\beta} - \beta)$ are independent;

(ii) $\frac{n\hat{\phi}_3}{\phi_3} \sim \chi^2_{n-2}$ and $\frac{n\hat{\phi}_4}{\phi_4} \sim \chi^2_{n-1}$;

(iii) $\hat{\phi}_4^{1/2}(\hat{\beta} - \beta) \sim N(0, \frac{\sigma_\beta^2 \phi_4}{n})$, where σ_β^2 is as given in (2.12).

(iv) $\hat{\beta} \sim t(\beta, \frac{\sigma_\beta^2}{n-1}; n-1)$; and

(v) $\frac{\sqrt{n-2}(\hat{\beta}-\beta)}{\sigma_\beta} \sim t(0, 1; n-2)$.

Notice from (v) that,

$$E[\hat{\beta}] = \beta, \quad n > 2 \quad \text{and} \quad Var[\hat{\beta}] = \frac{\sigma_\beta}{n-3}, \quad n > 3.$$

Moreover, from (iv) it follows that an exact $(1-\alpha)100\%$ confidence interval for β is given by

$$(\hat{\beta} \mp t_{n-1,\alpha/2}\frac{\hat{\sigma}_\beta}{\sqrt{n-2}}),$$

where $t_{n-1,\alpha/2}$ is the upper $1-\alpha/2$ point of a t distribution with n-2 degrees of freedom which can also be used as an exact α level test for $H_0 : \beta = \beta_0$.

5. The likelihood ratio statistics

Let $\tilde{\phi} = (\tilde{\phi}'_L, \tilde{\phi}'_S)'$ the maximum likelihood estimator of $\phi = (\phi'_L, \phi'_S)'$ under the null hypothesis $H_0 : \beta = \beta_0$. It is easy to see that $\tilde{\phi}_L = \hat{\phi}_L = \mathbf{Z}$ and $\tilde{\phi}_S = (\tilde{\phi}_3, \tilde{\phi}_4, \tilde{\beta})'$ follows from the equations

$$\tilde{\beta} = \beta_0$$

and

$$\tilde{\phi}_i = \alpha_i'(\beta_0)\mathbf{S}\alpha_i(\beta_0),$$

where $\alpha_i(\beta_0)$, $i = 3, 4$, are as defined in (2.11), with β replaced by β_0. In the model with λ_x known, it follows that $\tilde{\phi}_4 = \hat{\phi}_4$. Under $H_0 : \beta = \beta_0$ we have that

$$n\mathbf{S} \sim W_2(\Sigma_0, n-1).$$

where Σ_0 is the same as Σ (defined in (2.9)), but evaluated at $(\phi_3, \phi_4, \beta_0)$. This implies that

$$\frac{n\tilde{\phi}_i}{\phi_i} \sim \chi^2_{n-1},$$

$i = 3, 4$. However, $\tilde{\phi}_3$ and $\tilde{\phi}_4$ are independent only in the model with λ_x known. The likelihood ratio statistics for testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ is given by

$$G = 2(L(\hat{\phi}) - L(\tilde{\phi})) = n\log\{\frac{|\Sigma(\tilde{\phi}_S)|}{|\Sigma(\hat{\phi}_S)|}\} = n\log\{\frac{\tilde{\phi}_3\tilde{\phi}_4}{\hat{\phi}_3\hat{\phi}_4}\}. \tag{5.1}$$

Under $H_0 : \beta = \beta_0$, the statistics G has asymptotic chisquared distribution with one degree of freedom, denoted by $G \sim A\chi^2_1$ and can be represented in terms of β_0 and the elements of $\mathbf{S}$ as

$$G(\beta_0) = \begin{cases} n\log\{\frac{(\lambda_x+1)S_{YY} - 2\beta_0\lambda_x S_{YX} + (\beta_0\lambda_x)^2 S_{XX}}{(\lambda_x+1)S_{YY\cdot X}}\}, & \text{if } \lambda_x \text{ is known,} \\ n\log\{\frac{(\beta_0^2 S_{YY} + 2\lambda_e\beta_0 S_{YX} + \lambda_e S_{XX})(S_{YY} - 2\beta_0 S_{YX} + \beta_0^2 S_{XX})}{(\beta_0^2+\lambda_e)^2 S_{YY\cdot X} S_{XX}}\}, & \text{if } \lambda_e \text{ is known.} \end{cases} \tag{5.2}$$

where $S_{YY\cdot X} = S_{YY}(1 - r^2_{YX})$. Notice that in the case where λ_x is known,

$$G = n\log(\frac{\tilde{\phi}_3}{\hat{\phi}_3}).$$

As has been extensively discussed in the literature (Cordeiro, 1983, Wong, 1989), the approximation of the distribution of the statistics G to the chisquared distribution can be improved by using Bartlett correction factors. For the case where λ_x is known, the correction factor has been derived by Wong (1989), by using the approach developed by Lawley (1956), which is known to be not easily implemented since it depends on the fourth order cumulants of the likelihood ratio statistics. We propose now an alternative approach of deriving the correction factor for both cases (λ_x known and λ_e known) by computing directly the expected value of the likelihood ratio statistics by using some results derived in the previous section. Letting $E_0[G]$ denote the expected value of G under the null hypothesis $H_0 : \beta = \beta_0$, it follows from (4.5) that

$$E_0[G] = n\{E_0[\log\{\frac{\tilde{\phi}_3\tilde{\phi}_4}{\phi_3\phi_4}\}] - E_0[\log\{\frac{\hat{\phi}_3\hat{\phi}_4}{\phi_3\phi_4}\}]\}$$

$$= n\{(\psi(\frac{n-1}{2}) + \psi(\frac{n-1}{2}) - 2\log\frac{n}{2}) - (\psi(\frac{n-2}{2}) - \psi(\frac{n-1}{2}) - 2\log\frac{n}{2})\}$$

$$= n\{\psi(\frac{n-1}{2}) - \psi(\frac{n-2}{2})\},$$

where $\psi(m)$ is the digamma function evaluated at m. Using the fact that (Abramowitz and Stegun, 1965)

$$\psi(m) = \psi(m-1) + \frac{1}{m-1},$$

$$\psi(m) = \log m - \frac{1}{2m} - \frac{1}{12m^2} + \frac{1}{120m^4} + \ldots$$
$$= \log m - \frac{1}{2m} + O(m^{-2}).$$

we have that

$$E_0[G] = n\{(\psi(\frac{n+1}{2}) - \psi(\frac{n}{2}) - 2(\frac{1}{n-1} - \frac{1}{n-2})\} \tag{5.3}$$

$$= n\{[\log\frac{n+1}{2} - \log\frac{n}{2}) - (\frac{1}{n+1} - \frac{1}{n}) + O(n^{-2}) - 2(\frac{1}{n-1} - \frac{1}{n-2})\}$$
$$= n\{\log(1+\frac{1}{n}) - \frac{1}{n}[(1+\frac{1}{n})^{-1} - 1] + O(n^{-2}) - \frac{2}{n}[(1-\frac{1}{n})^{-1} - \frac{1}{2}(\frac{1}{2} - \frac{1}{n})^{-1}]\}.$$

Considering the expansions

$$\log(1+\frac{1}{n}) = \frac{1}{n} - \frac{1}{2n^2} + O(n^{-3}), \quad (1+\frac{1}{n})^{-1} = 1 - \frac{1}{n} + O(n^{-2}),$$

$$(1-\frac{1}{n})^{-1} = 1 + \frac{1}{n} + O(n^{-2}), \quad (\frac{1}{2} - \frac{1}{n})^{-1} = 2 + \frac{4}{n} + O(n^{-2}),$$

it follows from (5.3) that

$$E_0[G] = 1 + \frac{5}{2n} + O(n^{-2}),$$

so that for both cases (λ_x known and λ_e known) the corrected likelihood ratio statistics is given by

$$G^* = (1 - \frac{5}{2n})G, \tag{5.4}$$

with G as given in (5.2). Notice from (5.4) that the correction factor derived is exactly the one derived by Wong (1989) for the case λ_e known, and that it is the same for the case λ_x known.

References

Abramowitz, M. and Stegun, I.A. (Eds) (1965). *Handbook of mathematical functions*. Dover, New York.

Anderson, T.W. and Sawa, T. (1982). Exact and approximate distribution of maximum likelihood estimation of a slope coefficient. *J. Roy. Statist. Soc.*, B, 44, 52-62.

Bolfarine, H. and Cordani, L.K. (1993). Estimation of a structural linear regression model with a known reliability ratio. *Ann. of the Inst. of Statist. Math.*, 3, 531-540.

Cordeiro, G.M. (1983). Improved likelihood ratio statistics for generalized linear models. *J. of Roy. Statit. Soc.*, B, 45, 404-413.

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

Fuller, W.A. (1987). *Measurement error models*, Wiley, New York.

Kendall, M. and Stuart, (1979). *The advanced theory of Statistics*, Vol. 2, Griffin, London.

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

Muirhead, (1982). *Aspects of multivariate statistical theory*. Wiley, New York.

Wong, M.Y. (1989). Likelihood estimation of a simple regression model when both variables have error. *Biometrika*, 76, 141-148.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. Wiley, New York.

ÚLTIMOS RELATÓRIOS TÉCNICOS 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 Generalized 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.

9415 - AUBIN, E.C.Q. & CORDEIRO, G.M. Bias in Linear Regression Models with Unknown Covariance Matrix, 16p.

9416 - ELIAN, S.N. Previsão do Vetor de Elementos não observados numa população finita, 17p.

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