# ON TESTING STATISTICS FOR COMPARING SEVERAL MEASURING DEVICES

**Filidor Vilca Labra**[1] **Victor Lachos**[2]
Depto. de Estatística, IMECC-UNICAMP, CP 6065, 13081-970, Campinas, SP, Brazil.
and
**Heleno Bolfarine**[3]
Depto. de Estatística, IME- USP, CP 66281, 05311-970, São Paulo, SP, Brazil.

**Summary**
The main object of this paper is to compare the efficiency of measuring instruments that are used to measure the same quantity of interest in a group of $n$ individuals. The model considered was introduced in Grubbs (1948, 1973) and the inference for the model parameters is based on maximum likelihood estimation. A simulation study is used for comparing different test statistics for testing equality of biases and variances in the measuring instruments. A data set is used to illustrate the approach considered.

**Key Words**: Maximum likelihood estimation; Accuracy and precision; EM algorithm; Large sample tests.

## 1 Introduction

The paper treats the problem of comparative calibration, where $p$ measuring instruments are used to measure the same unknown quantity $x$ in a common group of experimental units. This type of problem is very common in scientific work. Grubbs (1973) presents an application where it is of interest comparing the efficiency of three chronometers, Christensen and Blackwood (1993) present an application comparing five thermocouples. More recently, Bedrick (2001) considers three different approaches for measuring soil sediments.

Consider $p$ instruments for measuring an usual characteristic in a group of $n$ objects. Let $Y_{ij}$ be the measure given by the instrument $i$, $i = 1, ..., p$, associated to an unknown quantity $x_j$, $j = 1, ..., n$. The model considered in Grubbs is given by the linear relation

$$Y_{ij} = \alpha_i + x_j + e_{ij}, \tag{1.1}$$

[1]E-mail: fily@ime.unicamp.br
[2]actually Ph-D student at the Universidade de São Paulo
[3]E-mail: hbolfar@ime.usp.br

where $e_{ij}$ and $x_j$ are independent, with $e_{ij} \sim N(0, \phi_i)$ and $x_j \sim N(\mu_x, \phi_x)$, $i = 1, ..., p$ and $j = 1, ..., n$. It follows that $E(Y_{ij}) = \alpha_i + \mu_x$, $\text{Var}(Y_{ij}) = \phi_x + \phi_i$ and $\text{Cov}(Y_{ij}, Y_{kj}) = \phi_x$, $i \neq j$. The parameters $\alpha_1, ..., \alpha_p$ are associated to the additive bias (accuracy) of the measuring instruments and $\phi_1, ..., \phi_p$ are associated to the precision of the measuring instruments.

Following Bedrick (2001), for eliminate redundancy, we assume that there is a reference instrument that measures, without bias the quantity of interest. Without loss of generality we consider this to be instrument so that $\alpha_1 = 0$.

In the context of Grubbs measurement error model, quality of measurement is evaluated in terms of the precision ( the inverse of the variance) and the accuracy (or bias) of the different instruments. The hypotheses of interest are: the instruments have the null bias (or mean) $H_{01} : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$, the same variance $H_{02} : \phi_1 = \phi_2 = ... = \phi_p = \phi$, and simultaneous $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_p = 0, \phi_1 = \phi_2 = ... = \phi_p = \phi$, which is the primary interest in this work.

For $p = 2$ Grubbs (1948, 1973) presents an inferential study based on multivariate techniques. Maloney and Rastogi (1970) show that Pitman's (1939) test is equivalent to $H_{02}$. For $p \geq 2$, Choi and Wette (1972) develops a test for evaluating the equality of the variance of the measuring instruments ($H_{02}$). Jaech (1973) considers a test for evaluating the equality of variances $H_{02}$. Christensen and Blackwood (1993) consider a multivariate linear model for testing $H_{01}$, $H_{02}$ and $H_0$ shows that the test for $H_0$ is equivalent to testing $H_0' : \gamma_i = \delta_i = 0$, $i = 1, \ldots, p-1$ in the multivariate regression model

$$y_{ij} - \bar{y}_{.j} = \delta_i + \gamma_i \bar{y}_{.j} + \epsilon_{ij},$$

$i = 1, \ldots, p-1$. Recently, Bedrick (2001) develops a test based in the score statistics but considering a different parametrization than the are considered in this paper.

In this paper we present an inferential study based on the Wald, score and likelihood ratio statistics. We consider maximum likelihood estimation by using the EM-algorithm. Under some restrictions on measurement error variances we study the asymptotic behaviour of some parameter estimates.

The paper is organized as follows. Section 2 presents the estimation procedures using the EM-algorithm and the information matrix is derived. Section 3 presents Wald, score and likelihood ratio for testing the hypothesis derived in Section 1. Maximum likelihood estimators for the restricted models (under the null) and the EM algorithm in the more general situations.

## 2 Grubbs models

The model given in (1.1), can be represented in matrix form as

$$\boldsymbol{Y}_j = \boldsymbol{a} + \mathbf{1}_p x_j + \boldsymbol{\varepsilon}_j \tag{2.1}$$

where $\boldsymbol{Y}_j = (Y_{1j}, ..., Y_{pj})^\top$, $\boldsymbol{\varepsilon}_j = (e_{1j}, ..., e_{pj})^\top$, $\boldsymbol{a} = (0, \alpha_2, ..., \alpha_p)^\top$, $\mathbf{1}_p = (1, ..., 1)^\top$ and $\boldsymbol{\alpha} = (\alpha_2, ..., \alpha_p)^\top$, $j = 1, \ldots, n$. The normal model is obtained considering

$$\boldsymbol{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.2}$$

where $\boldsymbol{\mu} = \mu(\boldsymbol{\theta}_L) = \boldsymbol{a} + \mathbf{1}_p \mu_x$ and $\boldsymbol{\Sigma} = \Sigma(\boldsymbol{\theta}_S) = \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi})$, with $\boldsymbol{\theta}_L = (\mu_x, \boldsymbol{\alpha}^\top)^\top$, $\boldsymbol{\theta}_S = (\phi_x, \boldsymbol{\phi}^\top)^\top$ and $D(\boldsymbol{\phi}) = diag(\phi_1, ..., \phi_p)$, $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)^\top$. We denote the parameters in the model by $\boldsymbol{\theta} = (\boldsymbol{\theta}_L^\top, \boldsymbol{\theta}_S^\top)^\top$.

For $p \geq 3$ the maximum likelihood estimator do not have closed form. In order to implement the EM algorithm, we derive next the complete (unobserved) likelihood function, which incorporate both, the observed and unobserved data, and is given by $\boldsymbol{y} = (\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n)^\top$ and $\boldsymbol{x} = (x_1, ..., x_n)^\top$, respectively. As will be seen, theses two put together will provide a much simpler and tractable likelihood function. Now, we define $\boldsymbol{Z}_j = (x_j, \boldsymbol{Y}_j^\top)^\top$, so that

$$\boldsymbol{Z}_j \sim N_{p+1}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z),$$

where

$$\boldsymbol{\mu}_z = \begin{pmatrix} \mu_x \\ \boldsymbol{\mu} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_z = \begin{pmatrix} \phi_x & \mathbf{1}_p^\top \phi_x \\ \mathbf{1}_p \phi_x & \phi_x \mathbf{1}_p \mathbf{1}_p^\top + D(\boldsymbol{\phi}) \end{pmatrix}.$$

The complete data log-likelihood function corresponding to $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \boldsymbol{\phi})^\top$ is

$$\ell(\boldsymbol{\theta} | \boldsymbol{z}_1, ..., \boldsymbol{z}_n) = cte - \frac{n}{2} log|\boldsymbol{\Sigma}_z| - \frac{1}{2} \sum_{j=1}^{n} (\boldsymbol{z}_j - \boldsymbol{\mu}_z)^\top \boldsymbol{\Sigma}_z^{-1} (\boldsymbol{z}_j - \boldsymbol{\mu}_z), \tag{2.3}$$

where

$$|\boldsymbol{\Sigma}_z| = \phi_x \prod_{i=1}^{p} \phi_i \quad \text{and} \quad \boldsymbol{\Sigma}_z^{-1} = \begin{pmatrix} c/\phi_x & -\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \\ -D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p & D^{-1}(\boldsymbol{\phi}) \end{pmatrix}, \tag{2.4}$$

with $c = 1 + \phi_x \mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi}) \mathbf{1}_p$.

## 2.1 The E and M steps

There are two steps in each cycle of the EM algorithm. The E and M steps. In the E step the algorithm finds the expectation of the complete data log-likelihood function given the observed data. But, since we are dealing with regular exponential families, to implement the E step it is sufficient to compute the expected value of the sufficient statistics that follow from the complete data likelihood function. Hence, since the sufficient statistics depends only on $\boldsymbol{x} = (x_1, ..., x_n)^\top$ through $x_j$ and $x_j^2$, the E step is implemented by computing

$$\widehat{x}_j = E[x_j|\boldsymbol{y}_j, \boldsymbol{\theta}] = \mu_x + \frac{\phi_x}{c}\mathbf{1}_p^\top D^{-1}(\boldsymbol{\phi})(\boldsymbol{y}_j - \boldsymbol{\mu}) \quad \text{and} \quad Var[x_j|\boldsymbol{y}_j, \boldsymbol{\theta}] = \frac{\phi_x}{c}, \tag{2.5}$$

so that

$$\widehat{x_j^2} = \widehat{x}_j^2 + \frac{\phi_x}{c}, \tag{2.6}$$

with $c$ as in (2.4). The M step of the algorithm obtains the next value of the unknown parameters by maximizing the complete data likelihood with the sufficient statistics replaced by their expected values obtained at the E step. As show in Dempster et al.(1977), each step of the algorithm increases the observed likelihood $l(\boldsymbol{\theta}|\boldsymbol{y})$. By differentiating the logarithm of the complete likelihood function given in (2.3) with respected to $\boldsymbol{\theta}$, we obtain the following equations:

$$\widehat{\mu}_x = \bar{x}, \quad \widehat{\alpha}_i = \bar{y}_{i.} - \bar{x}, \quad \widehat{\phi}_x = S_{xx},$$

$$\widehat{\phi}_1 = \frac{1}{n}\sum_{j=1}^{n}(y_{1j} - x_j)^2 \quad \text{and} \quad \widehat{\phi}_i = \frac{1}{n}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.} - x_j + \bar{x})^2,$$

where $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, $S_{xx} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$ and $\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$, $i = 2, \ldots, p$.

Now, let $I_F(\boldsymbol{\theta})$ be denote the expected information matrix. Hence, after some algebraic manipulations, it can be shown that

$$\mathbf{I}_F(\boldsymbol{\theta}) = \begin{pmatrix} I_{\mu_x\mu_x} & I_{\mu_x\alpha} & 0 & 0 \\ I_{\alpha\mu_x} & I_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & I_{\phi_x\phi_x} & I_{\phi_x\phi} \\ 0 & 0 & I_{\phi\phi_x} & I_{\phi\phi} \end{pmatrix},$$

where the elements of the matrix $I_F(\boldsymbol{\theta})$ are presented in the Appendix I.

Note that $I_F(\boldsymbol{\theta})$ can be represented as $I_F(\boldsymbol{\theta}) = Diag(\mathbf{I}_L, \mathbf{I}_S)$, where $\mathbf{I}_L$ and $\mathbf{I}_S$ are the submatrices corresponding to $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_S$, respectively. Moreover, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is asymptotically normally distributed with mean vector $\boldsymbol{\theta}$ and covariance matrix $\mathbf{V} = I_F(\boldsymbol{\theta})^{-1}$, which we denoted by $AN(\boldsymbol{\theta}, \mathbf{V})$. Thus, the maximum likelihood estimators $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\phi}}$ are asymptotically independent.

# 3 Tests for Precision and Accuracy

In this Section, we consider the study of the performance of the test for the hypotheses $H_0$, $H_{01}$ and $H_{02}$ using the Wald (W), Score (S) and likelihood ratios (G) statistics. These test, which are sometimes called the classical tests are particularly useful when the parameter space is multidimensional. Therefore, these three test statistics are asymptotically equivalent under the null hypothesis .

The null hypothesis considered in Section 1 can be written as $H : \mathbf{A}\boldsymbol{\theta} = 0$, where the matrix $\mathbf{A}$ is $r \times (2p+1)$, dimensional with rank$(\mathbf{A}) = r \leq 2p+1$. Thus, the statistics $W$, $S$ and $G$ can be write as,

$$W = n[\mathbf{A}\widehat{\boldsymbol{\theta}}]^\top[\mathbf{A}^\top \mathbf{I}_F^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{A}]^{-1}[\mathbf{A}\widehat{\boldsymbol{\theta}}], \quad S = \frac{1}{n}\widetilde{\boldsymbol{U}}^\top \mathbf{I}_F^{-1}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{U}} \quad \text{and} \quad G = 2[\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})], \tag{3.1}$$

where $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ are MLEs under unrestricted and restricted models, respectively and $\widetilde{\boldsymbol{U}} = U(\widetilde{\boldsymbol{\theta}}) = \frac{\partial}{\partial \boldsymbol{\theta}}\ell(\widetilde{\boldsymbol{\theta}})$.

## 3.1 The Wald statistics

We want to test the hypothesis $H_0$ considered in Section 1. We note that $H_0$ may be written as $H_0 : \boldsymbol{C}\boldsymbol{\theta}_* = 0$, where $\boldsymbol{\theta}_* = (\boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ and $\mathbf{C} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \end{pmatrix}$ is $2q \times (2p-1)$ dimensional matrix of rank $2q$ and $q = p-1$, with

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix} \tag{3.2}$$

matrix $q \times p$. Thus, the Wald statistics for testing $H_0$ is given by

$$W_0 = W_{01} + W_{02}, \tag{3.3}$$

where $W_{01} = n\widehat{\boldsymbol{\alpha}}'[\mathbf{I}_{(q)}\widehat{\mathbf{I}}_L^{-1}\mathbb{I}_{(q)}^\top]^{-1}\widehat{\boldsymbol{\alpha}}$ and $W_{02} = n\widehat{\boldsymbol{\phi}}'\mathbf{A}_1^\top[\mathbf{A}_1\mathbb{I}_{(p)}\widehat{\mathbf{I}}_S^{-1}\mathbb{I}_{(p)}^\top\mathbf{A}_1^\top]^{-1}\widehat{\boldsymbol{\phi}}$ are the Wald statistics for testing $H_{01}$ and $H_{02}$, individually. The notation $\mathbb{I}_{(r)}$ is used to denote the $r \times (r+1)$ matrix given by $\mathbb{I}_{(r)} = [\mathbf{0}, \mathbf{I}_r]$. The large sample null distribution of $W_{01}$ and $W_{02}$ are $\chi^2_{p-1}$. Furthermore, $W_{01}$ and $W_{02}$ are independent, so $W_0$ has an approximate null $\chi^2_{2p-2}$ distribution.

## 3.2 The Score statistics

Let $\widetilde{\boldsymbol{\theta}}$ the maximum likelihood estimator of $\boldsymbol{\theta}$ under $H_0$. Then, after some algebraic manipulations it follows that

$$\widetilde{\mu}_x = \overline{y}_{..}, \qquad \widetilde{\phi} = \frac{1}{n(p-1)}\sum_{j=1}^{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{.j})^2, \tag{3.4}$$

$$\widetilde{\phi}_x = \frac{1}{n}\sum_{j=1}^{n}(\overline{y}_{.j} - \overline{y}_{..})^2 - \frac{\widetilde{\phi}}{p} = \frac{1}{pn}\sum_{j=1}^{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{..})^2 - \widetilde{\phi},$$

where $\overline{y}_{..} = \dfrac{1}{np}\sum_{j=1}^{n}\sum_{i=1}^{p} y_{ij}$ and $\overline{y}_{.j} = \dfrac{1}{p}\sum_{i=1}^{p} y_{ij}$.

If $\widetilde{\phi}_x < 0$, then we consider that the MLEs for the variances satisfy $\widetilde{\phi}_x = 0$ and

$$\widetilde{\phi} = \frac{1}{pn}\sum_{j=1}^{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{..})^2. \tag{3.5}$$

The score statistics for testing $H_0$ is given by

$$S_0 = S_{01} + S_{02}, \tag{3.6}$$

where $S_{01} = \frac{n}{\widetilde{\phi}}\sum_{i=1}^{p}(\overline{y}_{i.} - \overline{y}_{..})^2$ and $S_{02} = \dfrac{1}{n}\mathbf{b}^\top\mathbf{B}\mathbf{b}$, with

$$\mathbf{b} = \frac{1}{2\widetilde{\phi}^2}\left[n(\widetilde{\phi}_x - \widetilde{\phi})\mathbf{1}_p + \sum_{j=1}^{n}\mathrm{D}(\mathbf{W}_j)\mathbf{W}_j - 2\widetilde{\tau}\sum_{j=1}^{n}\mathbf{W}_j^\top\mathbf{1}_p\mathbf{W}_j\right],$$

$$\mathbf{B} = \frac{2\widetilde{\phi}^2}{1-2\widetilde{\tau}}\left[\mathbf{I}_p - \frac{1 - \dfrac{2\widetilde{\tau}\widetilde{\phi}}{\widetilde{\phi}_x}}{p(p-1)}\mathbf{1}_p\mathbf{1}_p^\top\right],$$

$\mathbf{W}_j = \mathbf{y}_j - \widetilde{\boldsymbol{a}}^\top - \mathbf{1}_p \widetilde{\mu}_x$ and $\widetilde{\tau} = \dfrac{\widetilde{\phi}_x}{\widetilde{\phi} + p\,\widetilde{\phi}_x}$.

Note that, as $\mathbf{I}_F$ is a block diagonal matrix, then the statistics $S_{01}$ and $S_{02}$ are asymptotically independent. Therefore, the statistics $S_{01}$ and $S_{02}$ are the scores statistics for testing $H_{01}$ and $H_{02}$ individually, assuming that the other hypothesis hold.

Similar result is obtained in Bedrick (2001) for the Grubbs model using a different parametrization than the considered in this paper.

Under the hypothesis $H_{01}$, the MLE $\widetilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ has not a closed form, in this case, the MLE is obtained via the algorithm EM , and the estimators in the E and M steps are:

Step E:

$$\widetilde{x}_j = \mu_x + \frac{\phi_x}{c}[\mathbf{1}_p' D^{-1}(\boldsymbol{\phi})(\boldsymbol{y}_j - \boldsymbol{\mu})] \quad \text{and} \quad \widetilde{x_j^2} = \widetilde{x}_j^2 + \frac{\phi_x}{c}. \tag{3.7}$$

Step M: We maximize the function $\ell(\boldsymbol{\theta}|\boldsymbol{z}_1, ..., \boldsymbol{z}_n)$ and we obtain the following estimators:

$$\widetilde{\mu}_x = \bar{x}, \quad \widetilde{\phi}_x = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \quad \text{and} \quad \widetilde{\phi}_i = \frac{1}{n}\sum_{j=1}^{n}(y_{ij} - x_j)^2, \quad i = 1, \ldots, p.$$

The score statistics is obtained by substituting the MLE under $H_{01}$ in the statistics $S$ given in (3.1).

Finally, after some computations, the MLEs of the parameters under $H_{02}$ are given by

$$\widetilde{\mu}_x = \overline{y}_{1.}, \quad \widetilde{\alpha}_i = \bar{y}_{i.} - \bar{x}$$

$$\widetilde{\phi} = \frac{1}{n(p-1)}\sum_{j=1}^{n}\sum_{i=1}^{p}((y_{ij} - \overline{y}_{i.})^2 - (y_{.j} - \overline{y}_{..})^2) \tag{3.8}$$

$$\widetilde{\phi}_x = \frac{1}{n}\sum_{j=1}^{n}(\overline{y}_{.j} - \overline{y}_{..})^2 - \frac{\widetilde{\phi}}{p}.$$

If $\phi_x < 0$ , then we consider $\widetilde{\phi}_x = 0$ and the restricted MLE for $\phi$ is given by

$$\widetilde{\phi} = \frac{1}{pn}\sum_{j=1}^{n}\sum_{i=1}^{p}(y_{ij} - \overline{y}_{..})^2. \tag{3.9}$$

Thus, the score statistic is given by $S_{02} = \dfrac{1}{n}\mathbf{b}^\top\mathbf{B}\mathbf{b}$, where $\mathbf{b}$ e $\mathbf{B}$ are given in (3.6), $\mathbf{W}_j = \mathbf{y}_j - \widetilde{\boldsymbol{a}}^\top - \mathbf{1}_p\widetilde{\mu}_x$ and in this case the estimators are as in (3.8).

### 3.3 The Likelihood Ratio statistics

The likelihood ratio statistics for the hypotheses considered in the Section 1 does not present close form. The main reason being that the unrestricted and restricted MLEs do not present closed form, but its numerical implementation is simple and as the simulation study has we shall show, it present good properties.

## 4 Restricted Maximum likelihood

Following Jaech (1985), when one of the variances is estimated as being (close to or less than) zero, the conclusion may be that the corresponding instrument measures precisely (without error) the quantity of interest. In the case where one of the measurements measures precisely the characteristic of interest, the variance of the measurement errors corresponding to that instrument may be taken to be zero. We consider two situations: $\phi_1 = 0$ (standard instrument with null variance) and without loss of generality $\phi_p = 0$ (one nonstandard instrument with null variance).

### 4.1 The case $\phi_1 = 0$

If the condition $\phi_1 = 0$ is considered, then it follows from the assumptions considered in (1.1) that $\boldsymbol{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is as in (2.2) and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \phi_x & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{12} = \phi_x\mathbf{1}_q = \boldsymbol{\Sigma}_{21}^\top$ and $\boldsymbol{\Sigma}_{22} = \phi_x\mathbf{1}_q\mathbf{1}_q^\top + D(\boldsymbol{\psi})$, with $D(\boldsymbol{\psi}) = diag(\phi_2, ..., \phi_p)$ and $\boldsymbol{\psi} = (\phi_2, \ldots, \phi_p)^\top$.

The vector parameters in this case, is given by $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \boldsymbol{\psi}^\top)^\top$ and the maximum likelihood estimators are

$$\widehat{\mu}_x = \bar{y}_{1.}, \quad \widehat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{1.}, \quad \widehat{\phi}_i = S_{11} + S_{ii} - 2S_{1i}, \quad \widehat{\phi}_x = S_{11}, \qquad (4.1)$$

where $\bar{y}_{i.} = \dfrac{1}{n}\sum_{j=1}^{n} y_{ij}, i = 2, ..., p$ and $S_{kl} = \dfrac{1}{n}\sum_{j=1}^{n}(y_{kj} - \bar{y}_k)(y_{lj} - \bar{y}_l)$ , $k, l = 1, ..., p$. Let $\boldsymbol{\theta}_* = (\alpha_2, ..., \alpha_p, \phi_2, ..., \phi_p)^\top$, Walds statistic can be used for testing hypothesis like the one considered in previous the section. The information matrix is presented in the Appendix II.

## 4.2 The case $\phi_p = 0$

We consider now the possibility that one instrument nonstandard measure precisely (without error) the quantity of interest, without loss of generality, we consider $\phi_p = 0$. Thus, it follows that $\boldsymbol{Y}_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is as in (2.2) and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \phi_x \end{pmatrix}, \tag{4.2}$$

where $\boldsymbol{\Sigma}_{11} = \phi_x \mathbf{1}_q \mathbf{1}_q^\top + \mathrm{D}(\boldsymbol{\tau})$ and $\boldsymbol{\Sigma}_{12} = \phi_x \mathbf{1}_q = \boldsymbol{\Sigma}_{21}^\top$, with $\mathrm{D}(\boldsymbol{\tau}) = \mathrm{diag}(\phi_1, ..., \phi_q)$, $\boldsymbol{\tau} = (\phi_1, ..., \phi_q)^\top$ and $q = p - 1$.
The vector parameters in this case, is given by $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^\top, \phi_x, \boldsymbol{\tau}^\top)^\top$ and the maximum likelihood estimators are

$$\widehat{\mu}_x = \bar{y}_{1.}, \quad \widehat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{1.}, \quad \widehat{\phi}_x = S_{pp}, \quad \widehat{\phi}_i = S_{pp} + S_{ii} - 2S_{pi}, \tag{4.3}$$

where $\overline{y}_{i.}$ and $S_{kl}$, $i = 1, \ldots, p$, are as in (4.1). Note that the maximum likelihood estimators of the variance are non negative. The information matrix is presented in the Appendix III, that can be used for testing hypothesis like the one considered in the previous section, using the Wald statistics.

# 5 Applications

The data set presented in Jaech (1985) arose from an experiment that was conducted in which the densities of 43 sintered uranium fuel pellets for use in nuclear reactors were measure by six instruments.

In this application it is considered that instrument one is the reference (or standard) instrument, which correspond to the *geometric* method and operator 1. The table shows iterations of the EM-algorithm indicating that convergence is attained in approximately 50 iterations, providing estimates for the sequence $\phi_i$, $i = 1, \ldots, 6$. Note that the estimates are close to the ones reported in Jaech (1985, pp. 162).

Table 1: **Convergence of the EM algorithm**

| iter | $\widehat{\phi}_1$ | $\widehat{\phi}_2$ | $\widehat{\phi}_3$ | $\widehat{\phi}_4$ | $\widehat{\phi}_5$ | $\widehat{\phi}_6$ |
|---|---|---|---|---|---|---|
| 11 | 0.0095 | 0.0042 | 0.0275 | 0.0813 | 0.1394 | 0.0274 |
| 21 | 0.0072 | 0.0073 | 0.0251 | 0.0759 | 0.1319 | 0.0304 |
| 31 | 0.0069 | 0.0076 | 0.0249 | 0.0754 | 0.1313 | 0.0308 |
| 41 | 0.0068 | 0.0076 | 0.0249 | 0.0753 | 0.1313 | 0.0309 |
| 51 | 0.0068 | 0.0076 | 0.0249 | 0.0753 | 0.1313 | 0.0309 |
| Std. Dev. | (0.0024) | (0.0026) | (0.0060) | (0.0168) | (0.0289) | (0.0073) |

Estimates for the other parameters are:
$\widehat{\mu}_x = 4.3972\,(0.0316)$; $\widehat{\phi}_x = 0.0361\,(0.0084)$; $\widehat{\alpha}_2 = -0.0270\,(0.0183)$, $\widehat{\alpha}_3 = 0.0386\,(0.0272)$, $\widehat{\alpha}_4 = 0.3188\,(0.0437)$, $\widehat{\alpha}_5 = 0.3230\,(0.0567)$, $\widehat{\alpha}_6 = 0.2228\,(0.0296)$,
where the numbers in parenthesis denote corresponding standard deviations.

First consider the hypotheses $H_0$, the maximum likelihood estimates of $\boldsymbol{\theta}$ is given by $\widetilde{\mu}_x = 4.5433$, $\widetilde{\phi} = 0.0709$, $\widetilde{\phi}_x = 0.0350$.

Under hypotheses $H_{01}$, the maximum likelihood estimates of $\boldsymbol{\theta}$ is given by $\widetilde{\mu}_x = 4.4152$, $\widetilde{\phi}_x = 0.0360$, $\widetilde{\phi}_1 = 0.0063$, $\widetilde{\phi}_2 = 0.0118$, $\widetilde{\phi}_3 = 0.0246$, $\widetilde{\phi}_4 = 0.1665$, $\widetilde{\phi}_5 = 0.2231$, $\widetilde{\phi}_6 = 0.0776$. In this case, convergence is attained approximately in iteration 40. Finally, under $H_{02}$ the maximum likelihood estimators are given by $\widetilde{\mu}_x = 4.3972$, $\widetilde{\alpha}_2 = -0.0270$, $\widetilde{\alpha}_3 = 0.0386$, $\widetilde{\alpha}_4 = 0.3188$, $\widetilde{\alpha}_5 = 0.3230$, $\widetilde{\alpha}_6 = 0.2228$, $\widetilde{\phi}_x = 0.0393$, $\widetilde{\phi} = 0.0449$.

For testing hypothesis $H_0$, $H_{01}$ and $H_{02}$ defined in Section 1, we use Wald, score and likelihood ratio statistics.

Hypothesis $H_0$ is rejected, since $W_0 = 200.8226$ $E_0 = 124.7358$ and $Q_0 = 192.7844$.

Hypothesis $H_{01}$, is also rejected since $W_{01} = 151.5859$, $E_{01} = 74.8692$ and $Q_{01} = 103.9554$.

Hypotheses $H_{02}$, of equal precisions, is also rejected since $W_{02} = 49.2366$, $E_{02} = 77.9824$ and $Q_{02} = 94.7256$. The general conclusion is that the instruments are not accurate neither precise. Similar results are obtained in Jaech (1985).

## 6 A simulations study

In this Section we show a simulation study where estimates of the significance level and power of the three test statistics under $H_0$ are presented. Each Monte Carlo estimation of the significance level of the tests is based on 1000 independent samples generated according to the model defined by (1.2) with $\alpha_2 = \ldots = \alpha_p = 0$ and $\phi_1 = \ldots = \phi_p = 1$, $p = 3, 5$ sample sizes $n = 25$, 50 and 100. The power of the tests were also estimated. In all cases, the characteristic of interest $x_j$ was generated according to the $N(0, \phi_x)$, with $\phi_x = 0.1, 0.25, 1.0$.

Table 2 presents the estimated significance level in percents of the three test statistics corresponding to the nominal level of 5%. We can observe that the significance level of the Wald statistics converges slower to the nominal level, as $n$ increases, than the other statistics. The score test seems to converge faster to the nominal level, while the two other statistics overestimate the 5% nominal level for small values of $n$. It can also be noted that $\phi_x$ seems not to exert great influence on the significance level of the three statistics. This was also noted in other simulation studies as can also be noted in Figure 1 which presents, for $n = 25, 50, 100$, with $p = 3, 5$, estimated cumulative frequencies of the three statistics, for $H_0$, with theoretical cumulative frequencies corresponding to the chi-square distribution with eight degrees of freedom that we denote by $\chi^2_8$.

Table 3 presents power estimates for the three test statistics corresponding to model (1.2) with the parameter values considered above. The score statistics seems to present similar behavior as the likelihood ratio test while the Wald statistics seems to present a slightly inferior behavior. No noticiable difference seems to be noted as $\phi_x$ increase.

Hence, for the structural normal Grubbs model, we recomend using the score statistics which presents the best behavior in terms of significance level and power for testing hypothesis $H_0$, $H_{01}$ and $H_{02}$. The approximation to the chi-square distribution, even for small values of $n$ is worth noticing

## 7 Concluding remarks

The paper presents estimation and hypothesis testing in a structural normal Grubbs model frequently considering for comparing the efficiency of instruments used for measuring an unknown quantity in a group of common individual. Parameter estimation is considered via maximum likelihood, by using the EM-algorithm. Hypothesis testing is approached by using Wald, score and likelihood ratio statistics. Simulation studies seem to indicate that the score statistic presents the best behavior in terms of nominal level and power.

Table 2: Estimated significance levels of the three test with nominal level of 5%.

| $x_j$ | Sample size $n$ | $p=3$ $W$ | $E$ | $Q$ | $p=5$ $W$ | $E$ | $Q$ |
|---|---|---|---|---|---|---|---|
| $N(0, 0.25)$ | 25 | 7.6 | 4.9 | 7.9 | 10.5 | 4.5 | 7.2 |
| | 50 | 5.2 | 4.4 | 5.0 | 7.1 | 4.9 | 6.0 |
| | 100 | 6.6 | 5.4 | 6.3 | 7.1 | 4.9 | 6.3 |
| $N(0, 1.0)$ | 25 | 7.6 | 4.5 | 6.9 | 9.8 | 4.0 | 6.0 |
| | 50 | 5.8 | 4.9 | 5.6 | 7.4 | 4.0 | 5.8 |
| | 100 | 5.8 | 4.4 | 5.3 | 6.3 | 4.4 | 5.3 |

Table 3: Estimated power of the three test for the alternative $H_1 : \alpha_2 = 0,\ \alpha_3 = 0.5,\ \phi_1 = \phi_2 = 1,\ \phi_3 = 1.5$, with $p = 3$.

| $x_j$ | $N(0,\ 0.01)$ | | | $N(0,\ 0.25)$ | | | $N(0,\ 1.0)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $W$ | $E$ | $Q$ | $W$ | $E$ | $Q$ | $W$ | $E$ | $Q$ |
| 25 | 37.2 | 36.2 | 37.5 | 34.6 | 35.6 | 37.4 | 34.5 | 32.0 | 35.6 |
| 50 | 67.1 | 70.4 | 66.4 | 60.8 | 64.1 | 64.4 | 57.6 | 60.3 | 60.2 |
| 100 | 94.9 | 96.5 | 94.9 | 91.5 | 93.0 | 92.7 | 90.8 | 93.1 | 92.0 |
| 200 | 100 | 100 | 100 | 100 | 100 | 100 | 99.9 | 99.9 | 99.9 |

## Appendix I. Information matrix for the unrestricted model

After some algebraic manipulation, the information matrix for the unrestricted models is given by

$$\mathbf{I}_F(\theta) = \begin{pmatrix} I_{\mu_x\mu_x} & I_{\mu_x\alpha} & 0 & 0 \\ I_{\alpha\mu_x} & I_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & I_{\phi_x\phi_x} & I_{\phi_x\phi} \\ 0 & 0 & I_{\phi\phi_x} & I_{\phi\phi} \end{pmatrix},$$

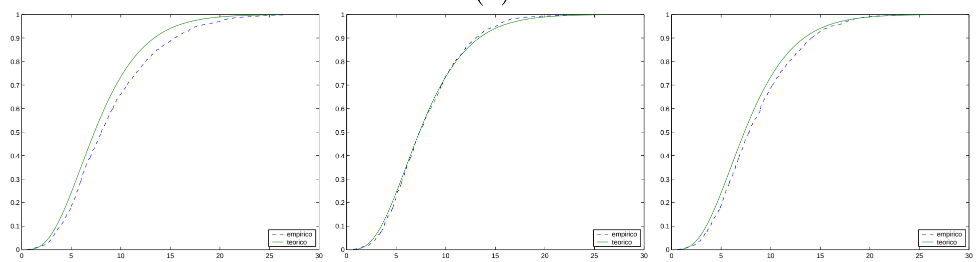
where

$$I_{\mu_x\mu_x} = \mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}_p, \quad I_{\mu_x\alpha} = \mathbf{1}_p^\top \boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top = I_{\alpha\mu_x}^\top, \quad I_{\alpha\alpha} = \mathbb{I}_{(p)}\boldsymbol{\Sigma}^{-1}\mathbb{I}_{(p)}^\top,$$

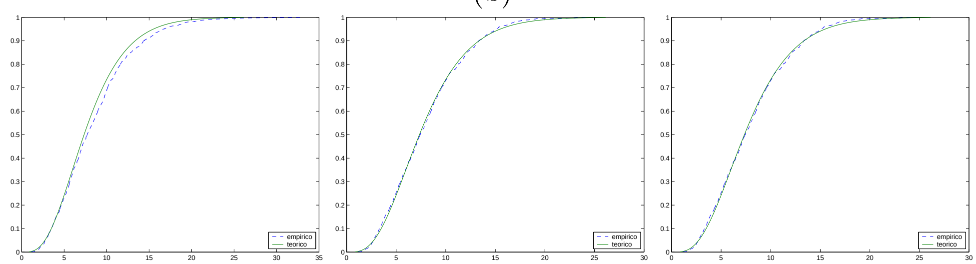
$$I_{\phi_x\phi_x} = -\frac{1}{2\phi_x^2}\left(\frac{c-1}{c}\right)^2 + \frac{c^{-2}}{\phi_x}(c-1)\mathbf{1}_p^\top \mathrm{D}^{-1}(\boldsymbol{\phi})\mathbf{1}_p,$$

Figure 1: Empirical (...) and theoretical (–) distribution of statistics Wald, Score and likelihood ratios, respectively for testing $H_0$, where a) $n = 25$ , b) $n = 50$ and c) $n = 100$, with $p = 5$.

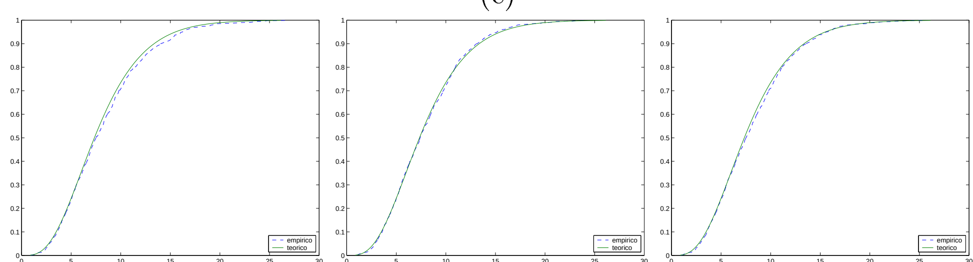
(a)




(b)




(c)




$$
\begin{aligned}
I_{\phi_x \phi} &= -\frac{c^{-2}}{2}\mathbf{1}_p^\top \mathrm{D}^{-2}(\boldsymbol{\phi}) - c^{-2}(c-1)\mathbf{1}_p^\top \mathrm{D}^{-2}(\boldsymbol{\phi}) + c^{-2}\mathbf{1}_p^\top \mathrm{D}^{-1}(\boldsymbol{\phi})\boldsymbol{\Sigma} D^{-2}(\boldsymbol{\phi}), \\
I_{\phi\phi} &= \frac{1}{2}\mathrm{D}^{-2}(\boldsymbol{\phi}) + \frac{c^{-2}}{2}\phi_x \mathrm{D}^{-1}(\boldsymbol{\phi})\mathbf{M}\mathrm{D}^{-1}(\boldsymbol{\phi}) - c^{-1}\phi_x \mathrm{D}^{-1}(\boldsymbol{\phi})\mathbf{M}\mathrm{D}^{-1}(\boldsymbol{\phi}) \\
&\quad +2c^{-2}\phi_x \mathrm{D}^{-2}(\boldsymbol{\phi})\boldsymbol{\Sigma}\mathbf{M}\mathrm{D}^{-1}(\boldsymbol{\phi}) - c^{-1}\phi_x \mathrm{D}^{-2}(\boldsymbol{\phi})\boldsymbol{\Sigma}\mathrm{D}^{-2}(\boldsymbol{\phi}).
\end{aligned}
$$

## Appendix II. Information matrix for the model with $\phi_1 = 0$

The information matrix corresponding to the model considered in Section (4.1.) is given by

$$\mathbf{I}_F(\boldsymbol{\theta}) = \begin{pmatrix} I_{\mu_x\mu_x} & 0 & 0 & 0 \\ 0 & I_{\alpha\alpha} & 0 & 0 \\ 0 & 0 & I_{\phi_x\phi_x} & 0 \\ 0 & 0 & 0 & I_{\psi\psi} \end{pmatrix},$$

where $I_{\mu_x\mu_x} = \frac{1}{\phi_x}$, $I_{\alpha\alpha} = \mathrm{D}^{-1}(\boldsymbol{\psi})$, $I_{\phi_x\phi_x} = \frac{1}{2\phi_x^2}$ and $I_{\psi\psi} = \frac{1}{2}\mathrm{D}^{-2}(\boldsymbol{\psi})$.

## Appendix III. Information matrix for the model with $\phi_p = 0$

The information matrix corresponding to the model considered in Section (4.2) is given by

$$\mathbf{I}_F(\boldsymbol{\theta}) = \begin{pmatrix} I_{\mu_x\mu_x} & 0 & I_{\mu_x\alpha_p} & 0 & 0 \\ 0 & I_{\alpha_*\alpha_*} & I_{\alpha_*\alpha_p} & 0 & 0 \\ I_{\alpha_p\mu_x} & I_{\alpha_p\alpha_*} & I_{\alpha_p\alpha_p} & 0 & 0 \\ 0 & 0 & 0 & I_{\phi_x\phi_x} & 0 \\ 0 & 0 & 0 & 0 & I_{\tau\tau} \end{pmatrix},$$

where $I_{\mu_x\mu_x} = \frac{1}{\phi_x}$, $I_{\mu_x\alpha_p} = \frac{1}{\phi_x}$, $I_{\alpha_*\alpha_*} = \mathbb{I}_{(q)}\mathrm{D}^{-1}(\boldsymbol{\tau})\mathbb{I}_{(q)}^\top$, $I_{\alpha_*\alpha_p} = -\mathbb{I}_{(q)}\mathrm{D}^{-1}(\boldsymbol{\tau})\mathbf{1}_q$, $I_{\alpha_p\alpha_p} = \frac{c_2}{\phi_x}$, $I_{\phi_x\phi_x} = \frac{1}{2\phi_x^2}$ and $I_{\tau\tau} = \frac{1}{2}\mathrm{D}^{-2}(\boldsymbol{\tau})$.

## References


Bedrick, E. J. (2001). An efficient scores test for comparing several measuring devices. *Journal of Quality technology*, 33, 1, 96 -103.

Blackwood, L. G. and Bradley, E. L. (1991). An omnibus test for comparing two measuring device. *Journal of Quality technology*, 33, 12-103.

Bolfarine, H and Galea, M. (1995). Structural comparative calibration of using the EM algorithm. *Journal of Applied Statistics,* 22, 277-292

Christensen, R. and Blackwood, L. G (1993). Tests for accuracy of multiple measuring devices. *Technometrics* 35, 411-420.

Choi, S. C. and Wette, R. (1972). A teste for the homogeneity of the variances among correlated variables. *Biometrics*, 28, 589-592.

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

Dempster , A., Laird, N., and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algrithm ( with discussion). *J. Roy. Statist. Soc. B.* 39, 1-38.

Grubbs, F.E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, 43, 243-264.

Grubbs, F.E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics* 15, 53-66.

Jaech, J.L. (1985). Statistical analysis of measurement errors. *Exxon Monographs* . John Wiley, N. Y.