ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal

Volume 8, No.1
June 2010

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 842 6364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 22 605 07 48
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue ……. **€ 8**
- Annual subscription (No. 1, No. 2) ……. **€ 13**

INDEX

REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 1–20

ON ESTIMATION FOLLOWING SUBSET SELECTION FROM TRUNCATED POISSON DISTRIBUTIONS UNDER STEIN LOSS FUNCTION

Authors: A. Shanubhogue
– Department of Statistics, Sardar Patel University, India
a_shanubhogue@yahoo.com

Riyadh R. Al-Mosawi
– Department of Mathematics, Thiqar University, Iraq
riyadhrm@gmail.com

Received: March 2009 Revised: January 2010 Accepted: January 2010

Abstract:

- In this paper, we consider the problem of estimating the parameters of a subset selected from $p\,(p \geq 2)$ left-truncated Poisson distributions under Stein loss function. Two problems of estimations are considered; average worth and simultaneous estimation. For the average worth, the natural estimator is shown to be positively biased with respect to Stein loss function and the Unique Minimum Risk Unbiased Estimator *UMRUE* is obtained. For the simultaneous estimation problem, we have shown that the natural estimator is positively biased with respect to Stein loss function and the *UMRUE* is obtained. The inadmissibility of the natural estimator of the simultaneous estimation is also proved and a class of dominating estimators is obtained. Monte Carlo simulation is undertaken to compute the biases and risks of the two problems of estimation.

Key-Words:

- *simultaneous estimation after subset selection; average worth estimation; Stein loss function; difference inequalities; truncated Poisson distributions.*

AMS Subject Classification:

- 62F10, 62F07.

1. INTRODUCTION

Estimating the parameter of the selected population is an important practical problem which arises in various disciplines such as agriculture, medicine and industry. Say, we wish to select the most productive machine from p different types of machines and then estimate the mean of the production of the selected machine. The problem of estimation after selection has received considerable attention from many researchers. Some references in this area include, Sackrowitz and Samuel-Cahn (1984), Kumar and Gangopadhyay (2005), Misra, van der Meulen and Branden (2006a, 2006b), Sill and Sampson (2007) and Vellaisamy and Jain (2008). All these studies considered the problem of estimation when the selection rule selects only one population. However sometimes we are interested to select a subset of good populations (including the best) rather than only one population (the best) and then estimating the parameters of the selected subset. The problem of estimation after subset selection was initially formulated and studied by Jayaratnam and Panchapakesan (1984) for two normal populations. They proposed three classes of estimators for the average worth of the selected subset and compared numerically their biases and mean squared errors. Jayaratnam and Panchapakesan (1986) considered the case of two independent exponential populations and they proved that the natural estimator of the average worth of the selected subset is positively biased. They suggested an adjusted estimator by adjusting the bias of the natural estimator and compared the bias and mean squared error of the natural estimator with the adjusted estimator. Vellaisamy (1992) considered the average worth estimation and simultaneous estimation of the subset selected from independent gamma population with unknown scale parameters and common known shape parameter. He proved that the natural estimator of the average worth is positively biased and inadmissible and also, he obtained the *UMVUE* of it using the *UV* method of Robbins (1988). Also, he observed similar results for the simultaneous estimation of the selected subset. Misra (1994) derived the *UMVUE* of the average worth of the selected subset from p independent gamma populations with common known shape parameter and unknown scale parameters. He also proved the inadmissibility of the natural estimator of the average worth under squared error loss function by constructing improved estimators. Vellaisamy (1996) considered the case of subset selection from uniform populations. He proved for the simultaneous estimation of the parameters associated with the selected populations, the natural estimators as well as the unbiased estimator are inadmissible under the squared error loss function and the dominating ones were obtained. The problem of estimating the average worth of the selected subset from exponential populations with a common unknown location parameter and unknown scale parameters has been investigated by Gangopadhyay and Kumar (2005). They derived the *UMVUE* of the average worth of the selected subset and they also, compared, numerically, the bias and the mean squared error of the *UMVUE*, *BAEE* and *MLE* of the average worth.

They observed that the natural estimator dominates the unbiased estimator and the natural estimator itself is inadmissible. The literatures, so far, deal with the problem of estimating the parameters (average worth or simultaneous estimation) of a selected subset containing the best population when the distributions of populations are continuous. In this paper, we take up the problem of estimating the parameters of the selected subset under the asymmetric loss function when the distributions of populations are discrete. The loss function considered here is Stein loss function defined as

$$L(\theta, d) = \frac{d}{h(\theta)} - \log\left(\frac{d}{h(\theta)}\right) - 1 , \tag{1.1}$$

where d is an estimate of $h(\theta)$ and log denotes the natural logarithm. The loss function (1.1) was first introduced in James and Stein (1961) for estimation of the multinormal covariance matrix. Also, it was considered by Dey and Srinivasan (1985) and Dey and Chung (1991) for simultaneous estimation. In Section 2, we introduce some notations, definitions and lemmas and formulate the problem. In Section 3, the natural estimators of the average worth and simultaneous estimation of the selected subset are shown to be positively biased with respect to Stein loss function. In Section 4, the *UMRUE*'s of the average worth and simultaneous estimation are derived. In Section 5, the inadmissibility of the natural estimator of the simultaneous estimation is proved by solving certain difference inequality and a class of improved estimators is constructed. In Section 6, Monte Carlo simulation is undertaken to compute the biases and risks of the estimators under the two problems of estimation.

2. NOTATIONS, DEFINITIONS AND FORMULATION OF THE PROBLEM

Let $\Pi_1, ..., \Pi_p$ be p $(p \geq 2)$ independent populations such that the random variable X_i represents the population Π_i has left-truncated Poisson p.d.f.

$$P(X_i = x_i) = \frac{\theta_i^{x_i}}{x_i!\, f(\theta_i, t)} , \qquad x_i = t, t+1, ...; \quad t > 0; \quad \theta_i > 0; \quad i = 1, ..., p ,$$

where $f(\theta_i, t) = e^{\theta_i} - \sum_{k=o}^{t-1} \theta_i^k / k!$. We assume that $\theta_1, ..., \theta_p$ are unknown parameters. Suppose from each population Π_i we have a random sample $X_{i1}, ..., X_{in}$ and let $Z_i = \sum_{j=1}^{n} X_{ij}$. It is well-known (see for example, Jani (1977)) that the distribution of Z_i is given by

$$P(Z_i = z_i) = \frac{n!\, S(z_i, n, t)\, \theta_i^{z_i}}{z_i!\, f^n(\theta_i, t)} , \qquad z_i = nt, nt+1, ... ,$$

where $S(z_i, n, t)$ is the Stirling number of the second kind and the *UMVUE* of θ_i is given by

$$\delta(Z_i) = \begin{cases} Z_i\, S(Z_i - 1, n, t)/S(Z_i, n, t), & \text{if } Z_i \geq nt + 1\,; \\ 0, & \text{otherwise}\,. \end{cases} \tag{2.1}$$

Without loss of generality, we consider the case $n = 1$. So that the *UMVUE*, defined in (2.1), reduces to

$$\delta(X_i) = \begin{cases} X_i, & \text{if } X_i \geq t + 1\,; \\ 0, & \text{otherwise}\,. \end{cases} \tag{2.2}$$

Let $\chi = \{x\colon x = (x_1, ..., x_p),\ x_i \geq t,\ i = 1, ..., p\}$ and $\Omega = \{\theta\colon \theta = (\theta_1, ..., \theta_p),\ \theta_i > 0,\ i = 1, ..., p\}$ denote the sample space and the parameter space, respectively, and let $\theta_{[1]} \geq \theta_{[2]} \geq \cdots \geq \theta_{[p]}$ represent the ordered parameters and $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(p)}$ represent the ordered values of $X_1, ..., X_p$ (use arbitrary ordering if some of the θ_i's (X_i's) are equal). The population associated with $\theta_{[1]}$ is called the best population. In the subset selection approach, we want to select a non-empty subset from the p populations so that the best population is included in the selected subset with a minimum pre-assigned probability $P^*(1/p < P^* < 1)$ (Gupta (1965)). To select such a subset, we consider, in this paper, the following modified selection rule which was suggested by Gupta and Huang (1975).

$$R:\ \text{Choose } \Pi_i \text{ in the subset iff } X_i + 1 \geq cX_{(1)}\ , \tag{2.3}$$

where $c = c(p, P^*)$ $(0 < c < 1)$ is some suitable constant satisfying the basic probability requirement

$$\inf_{\theta \in \Omega} P_\theta\big(CS|R\big) = P^*\ ,$$

and CS stands for "Correct Selection" (i.e. the selection contains the best population). Let $X_{(1)i} \geq X_{(2)i} \geq \cdots \geq X_{(p-1)i}$ denote the ordered values of $X_1, ..., X_{i-1}$, $X_{i+1}, ..., X_p$. Note that

$$\big\{X_i + 1 \geq cX_{(1)}\big\} = \big\{X_i + 1 \geq cX_{(1)i}\big\}\,.$$

Suppose a subset (of random size) is selected using the rule R. The problems that we are interested here are the estimation of the average worth M and the simultaneous estimation of Q, defined by

$$M = \frac{\sum_{i=1}^{p} \theta_i\, I_i(X)}{\sum_{i=1}^{p} I_i(X)}$$

and

$$Q = \big(\theta_1 I_1(X), ..., \theta_p\, I_p(X)\big)\,,$$

where $I_i(X) = I\big(X_i + 1 \geq cX_{(1)i}\big)$ and $I(A)$ denotes the indicator function of an event A. It can be seen that the dimension of the estimand M is random, as it

varies with X, unlike in the case of classical estimation problem. The natural analogues of M and Q for the selection problem are as follows

$$\hat{M}_1(X) = \frac{\sum_{i=1}^{p} \delta(X_i)\, I_i(X)}{\sum_{i=1}^{p} I_i(X)} \tag{2.4}$$

and

$$\hat{Q}_1(X) = \big(\delta(X_1)\, I_1(X), ..., \delta(X_p)\, I_p(X)\big)\,, \tag{2.5}$$

and we will call them, the natural estimators of M and Q, respectively, where δ is as in (2.2). The loss function (1.1) can be written for the case of estimating M and Q as in the following

$$L(M, \hat{M}) \;=\; \frac{\hat{M}}{M} - \log\left(\frac{\hat{M}}{M}\right) - 1$$

and

$$L(Q, \hat{Q}) \;=\; \sum_{j=1}^{p} \left[\frac{d_j}{\theta_j} - \log\left(\frac{d_j}{\theta_j}\right) - 1\right] I_j(X)\;,$$

where $\hat{M}$ is an estimate of M, d_j is an estimate of θ_j and $\hat{Q} = (d_1, ..., d_p)$. The loss function is well defined in our problem, since we considered distributions truncated at zero. Now, we introduce the following lemmas which will be used in the next sections. The following lemma is from Chou (1991).

Lemma 2.1. *Let f_1 be a real-valued function defined on p-fold Cartesian product of $\mathbb{I}^+$, the set of positive integers, such that $E_\theta|f_1(X)| < \infty$ and $f_1(x) = 0$ if $x_i \le t$. Then*

$$E_\theta\, f_1(X)/\theta_i \;=\; E_\theta\big(f_1(X + e_i)/\delta(X_i + 1)\big)\;,$$

where e_i is the p-dimensional vector whose i-th coordinate is 1 and the rest are zeros and δ is as in (2.2).

Lemma 2.2. *Let f_2 be a real-valued function defined on p-fold Cartesian product of $\mathbb{I}^+$ such that $E_\theta|f_2(X)| < \infty$. Then*

$$E_\theta\, f_2(X)\,\theta_i \;=\; E_\theta\, f_2(X - e_i)\,\delta(X_i)\;,$$

where δ is as in (2.2)

Lemma 2.3. *If $|w| \le 1/2$, then $\log(1 + w) \ge w - w^2$.*

Proof: Similar to the proof of Lemma 2.2 of Dey and Srinivasan (1985), we observe that

$$\begin{aligned}
\log(1+w) &= w - \frac{w^2}{2} + \frac{w^3}{3} - \frac{w^4}{4} + \cdots \\
&\geq w - \frac{w^2}{2} - \frac{|w|^3}{2} - \frac{|w|^4}{2} - \cdots \\
&= w - \frac{w^2}{2}\left(1 + |w| + |w|^2 + \cdots\right) \\
&= w - \frac{w^2}{2\left(1-|w|\right)} \\
&\geq w - w^2 \,,
\end{aligned}$$

since $|w| \leq 1/2$. □

3. ESTIMATION OF M AND Q

In this section, the natural estimators of M and Q are shown to be positively biased with respect to Stein loss function. First of all we need to impose a condition on the estimator δ to be unbiased under Stein loss function using the definition of the risk-unbiasedness of Lehmann (1951).

Definition 3.1. An estimator $\eta(Y)$ of $g(\theta)$ is said to be risk-unbiased if it satisfies

$$E_\theta\, L\big(\theta, \eta(Y)\big) \leq E_\theta\, L\big(\theta', \eta(Y)\big) \qquad \forall\ \theta' \neq \theta \ . \tag{3.1}$$

Following Nematollahi and Motamed-Shariati (2009), the estimator η of θ is said to be unbiased under Stein loss function if $E_\theta\, \eta(Y) = \theta$, $\forall\, \theta \in \Omega$, otherwise, it is biased and its bias is $B_\theta(\eta) = E_\theta\big(\eta(Y) - \theta\big)$. Clearly, this is the same definition of the usual unbiasedness (unbiasedness under the squared error loss function). Consider first the estimation problem of the average worth M. The estimator $\hat{M}$ of M is said to be unbiased under Stein loss function if $E_\theta \hat{M} = E_\theta M$, otherwise, it is biased and its bias is $B_\theta(\hat{M}, M) = E_\theta(\hat{M} - M)$. Without loss of generality, consider $p = 2$. Then, the average worth M can be written as

$$M = \begin{cases} \theta_1, & \text{if } \ X_1 > c^{-1}(X_2+1)\,, \\ \theta_2, & \text{if } \ X_1 < cX_2 - 1\,, \\ \frac{1}{2}(\theta_1+\theta_2), & \text{if } \ cX_2 - 1 \leq X_1 \leq c^{-1}(X_2+1)\,, \end{cases} \tag{3.2}$$

and hence the natural estimator of M is

$$\hat{M}_1 = \begin{cases} \delta(X_1), & \text{if } \ X_1 > c^{-1}(X_2+1)\,, \\ \delta(X_2), & \text{if } \ X_1 < cX_2 - 1\,, \\ \frac{1}{2}\big(\delta(X_1)+\delta(X_2)\big) & \text{if } \ cX_2 - 1 \leq X_1 \leq c^{-1}(X_2+1)\,, \end{cases} \tag{3.3}$$

where δ is as in (2.2).

Theorem 3.1. *The natural estimator $\hat{M}_1$ of M is positively biased under Stein loss function.*

Proof: Without loss of generality consider $p = 2$. From (3.2) and (3.3), it follows

$$\begin{aligned} B_\theta(\hat{M}_1, M) &= E_\theta(\hat{M}_1 - M) \\ &= E_\theta\big(\delta(X_1) - \theta_1\big)\, I\big(X_1 > c^{-1}(X_2+1)\big) + E_\theta\big(\delta(X_2) - \theta_2\big)\, I\big(X_1 < cX_2 - 1\big) \\ &\quad + 0.5\, E_\theta\big(\delta(X_1) - \theta_1 + \delta(X_2) - \theta_2\big)\, I\big(cX_2 - 1 \le X_1 \le c^{-1}(X_2+1)\big)\ . \end{aligned}$$

Since $\delta(X_i)$ is unbiased estimator for θ_i, $i = 1, 2$, and

$$I\big(cX_2 - 1 \le X_1 \le c^{-1}(X_2+1)\big) = 1 - I\big(X_1 > c^{-1}(X_2+1) - I(X_1 < cX_2 - 1)\big)\ .$$

Then

$$\begin{aligned} B_\theta(\hat{M}_1, M) &= 0.5\, E_\theta\Big[\big(\delta(X_1) - \theta_1\big) - \big(\delta(X_2) - \theta_2\big)\Big]\, I\big(X_1 > c^{-1}(X_2+1)\big) \\ &\quad + 0.5\, E_\theta\Big[\big(\delta(X_2) - \theta_2\big) - \big(\delta(X_1) - \theta_1\big)\Big]\, I\big(X_2 > c^{-1}(X_1+1)\big) \\ &= 0.5\, A(\theta_1, \theta_2) + 0.5\, A(\theta_2, \theta_1)\ , \end{aligned}$$

where

$$A(\theta_1, \theta_2) = E_\theta\big(\delta(X_1) - \theta_1\big)\, I\big(X_1 > c^{-1}(X_2+1)\big) - E_\theta\big(\delta(X_2) - \theta_2\big)\, I\big(X_1 > c^{-1}(X_2+1)\big),$$

and $A(\theta_2, \theta_1)$ follows by interchanging the role of (X_1, θ_1) and (X_2, θ_2) in $A(\theta_1, \theta_2)$. Consider the term $A(\theta_1, \theta_2)$.

$$\begin{aligned} A(\theta_1, \theta_2) &= E_\theta\, \delta(X_1)\, I\big(X_1 > c^{-1}(X_2+1)\big) - \theta_1\, E_\theta\, I\big(X_1 > c^{-1}(X_2+1)\big) \\ &\quad - E_\theta\, \delta(X_2)\, I(X_2 < cX_1 - 1) + \theta_2\, E_\theta\, I(X_2 < cX_1 - 1) \end{aligned}$$

and by using Lemma 2.2 we get

$$\begin{aligned} A(\theta_1, \theta_2) &= E_\theta\, \delta(X_1)\, I\big(X_1 > c^{-1}(X_2+1)\big) - E_\theta\, \delta(X_1)\, I\big(X_1 > c^{-1}(X_2+1) + 1\big) \\ &\quad - E_\theta\, \delta(X_2)\, I(X_2 < cX_1 - 1) + E_\theta\, \delta(X_2)\, I(X_2 < cX_1) \\ &= E_\theta\, \delta(X_1)\, I\big(c^{-1}(X_2+1) < X_1 \le c^{-1}(X_2+1) + 1\big) \\ &\quad + E_\theta\, \delta(X_2)\, I\big(cX_1 - 1 \le X_2 < cX_1\big) \\ &> 0\ . \end{aligned}$$

This completes the proof. □

Consider next the problem of estimating Q. The estimator $\hat{Q} = (q_1 I_1(X), ..., q_p I_p(X))$ of Q is said to be unbiased under Stein loss function if $E_\theta\, q_i I_i(X) = E_\theta\, \theta_i I_i(X)\ \forall i$, otherwise, it is biased and its bias is

$$B_\theta(\hat{Q}, Q) = E_\theta \sum_{i=1}^{p} (q_i - \theta_i)\, I_i(X)\ .$$

Theorem 3.2. *The natural estimator $\hat{Q}_1$ of Q is positively biased under Stein loss function.*

Proof: Observe that

$$\begin{aligned} B_\theta(\hat{Q}, Q) &= \sum_{i=1}^{p} E_\theta\, \delta(X_i)\, I\big(X_i + 1 \geq c X_{(1)i}\big) - \sum_{i=1}^{p} E_\theta\, \theta_i I\big(X_i + 1 \geq c X_{(1)i}\big) \\ &= \sum_{i=1}^{p} E_\theta\, \delta(X_i)\, I\big(X_i + 1 \geq c X_{(1)i}\big) - \sum_{i=1}^{p} E_\theta\, \delta(X_i)\, I\big(X_i \geq c X_{(1)i}\big) \\ &\quad \text{(using Lemma 2.2)} \\ &= \sum_{i=1}^{p} E_\theta\, \delta(X_i)\, I_i\big(c X_{(1)i} - 1 \leq X_i < c X_{(1)i}\big) \\ &> 0 \end{aligned}$$

since $P\big(cX_{(1)i} - 1 \leq X_i < cX_{(1)i}\big) > 0$ for some i. This completes the proof. □

4. THE UMRUE's OF M AND Q

In this section we derive the *UMRUE's* of M and Q using the UV method of estimation of Robbins (1988) and the generalization of Misra (1994) of the Lehmann–Scheffe theorem. These estimators are also the *UMVUE's* since the definition of risk-unbiasedness under Stein loss function coincides with the definition of usual unbiasedness. First, consider the *UMRUE* of the average worth M. Let

$$U_i(X) = \frac{I\big(X_i + 1 \geq c X_{(1)i}\big)}{\sum_{j=1}^{p} I\big(X_j + 1 \geq c X_{(1)j}\big)}\,, \qquad i = 1, ..., p\,,$$

then $M = \sum_{i=1}^{p} \theta_i U_i(X)$. From Lemma 2.2, we have $E_\theta\, \theta_i U_i(X) = E_\theta\, \delta(X_i) U_i(X - e_i)$. Let $\hat{M}_2 = \sum_{i=1}^{p} V_i(X)$ and $V_i(X) = \delta(X_i)\, U_i(X - e_i)$. Then, $E_\theta M = E_\theta \hat{M}_2$ and hence $\hat{M}_2$ is an unbiased estimator of M. Let $Y_1 \geq Y_2 \geq \cdots \geq Y_p$ denote the ordered values of $X_1, ..., X_p$ and $Y = (Y_1, ..., Y_p)$. It is easy to see that

$$\sum_{i=1}^{p} V_i(X) = \sum_{i=1}^{p} V_i^*(Y)\,.$$

Now, since

$$U_1^*(Y) = \frac{I(Y_1 + 1 \geq c Y_2)}{I(Y_1 + 1 \geq c Y_2) + \sum_{j=2}^{p} I(Y_j + 1 \geq c Y_1)}$$

and

$$U_i^*(Y) = \frac{I(Y_i + 1 \geq c Y_1)}{I(Y_1 + 1 \geq c Y_2) + I(Y_i + 1 \geq c Y_1) + \sum_{j=2, j \neq i}^{p} I(Y_j + 1 \geq c Y_1)}\,, \quad i = 2, ..., p\,,$$

we get

$$\begin{aligned} V_1^*(Y) &= \delta(Y_1)\, U_1^*(Y-e_1) \\ &= \frac{\delta(Y_1)\, I(Y_1 \geq cY_2)}{I(Y_1 \geq cY_2) + \sum_{j=2}^{p} I\big(Y_j+1 \geq c\max(Y_1-1, Y_2)\big)} \\ &= \frac{\delta(Y_1)}{1 + I(Y_1=Y_2)\sum_{j=2}^{p} I(Y_j+1 \geq cY_1) + I(Y_1>Y_2)\sum_{j=2}^{p} I\big(Y_j+1 \geq c(Y_1-1)\big)} \end{aligned}$$

and

$$\begin{aligned} V_i^*(Y) &= \delta(Y_i)\, U_i^*(Y-e_i) \\ &= \frac{\delta(Y_i)\, I(Y_i \geq cY_1)}{1 + I(Y_i \geq cY_1) + \sum_{j=2, j\neq i}^{p} I(Y_j+1 \geq cY_1)}\ , \qquad i=2,...,p\ . \end{aligned}$$

To find an explicit form of $V^*(Y)$, we need the following definitions. Let

$$S_0 = \text{empty set}\ , \qquad S_i = \big\{Y_j,\ j=1,...,p\colon Y_j = Y_{1+m_i}\big\}\ ,$$

and $m_i = \sum_{j=1}^{i-1} \#(S_j)$ where $i=1,...,r$, and $1 \leq r \leq p$. Note that $\sum_{i=1}^{r} \#(S_i) = p$ and the subsets $\{S_i\}_{i=1}^{r}$ represent a partition of the set of variables $Y=(Y_1,...,Y_p)$. Let $W_1,...,W_{r+1}$ be random variables such that $W_i \in S_i$, $i=1,...,r$ and $W_{r+1}=-1$. It is obvious that $W_{i+1} < W_i$ and $W_i = Y_{1+m_i} = Y_{2+m_i} = \cdots = Y_{m_{i+1}}$ for $i=1,...,r$. Define the following partition of χ

$$\chi = \left(\bigcup_{l=1}^{r-2}\bigcup_{k=l}^{l+1} \chi_{1,l,k}\right) \bigcup \chi_{1,r-1,r-1} \bigcup \left(\bigcup_{l=1}^{r-1}\bigcup_{k=l-1}^{l} \chi_{2,l,k}\right) \bigcup \chi_{2,r,r-1}$$

where

$$\chi_{1,l,k} = \Big\{X \in \chi\colon\ W_{l+1} < cW_1 \leq W_{l+1}+1,\ W_{k+2}+1 < c(W_1-1) \leq W_{k+1}+1\Big\}$$
$$\text{for} \quad l=1,...,r-2;\ k=l,l+1 \quad \text{and} \quad l=r-1;\ k=l\ ,$$

and

$$\chi_{2,l,k} = \Big\{X \in \chi\colon\ W_{l+1}+1 < cW_1 \leq W_l,\ W_{k+2}+1 < c(W_1-1) \leq W_{k+1}+1\Big\}$$
$$\text{for} \quad l=1,...,r-1;\ k=l-1,l \quad \text{and} \quad l=r;\ k=l-1\ .$$

Note that if $W_l = W_{l+1}+1$ for some $l=1,...,r$, then $\chi_{2,l,k} = \text{empty set}$ for all $k=l-1,l$.

Case I: When $X \in \chi_{1,l,k}$.

In this case we have

$$I\big(W_j \geq cW_1\big) = 1 \qquad \text{for}\ \ j=1,...,l\ ,$$
$$I\big(W_j+1 \geq cW_1\big) = 1 \qquad \text{for}\ \ j=1,...,l+1\ ,$$
$$I\big(W_j+1 \geq c(W_1-1)\big) = 1 \qquad \text{for}\ \ j=1,...,k+1\ ,$$

then

$$I\big(Y_j \geq cY_1\big) = 1 \qquad \text{for} \quad j = 1, ..., m_{l+1}\,,$$
$$I\big(Y_j + 1 \geq cY_1\big) = 1 \qquad \text{for} \quad j = 1, ..., m_{l+2}\,,$$
$$I\big(Y_j + 1 \geq c(Y_1 - 1)\big) = 1 \qquad \text{for} \quad j = 1, ..., m_{k+2}\,.$$

So that

$$V_1^*(Y) = \frac{\delta(Y_1)}{m_{l+2}\, I(Y_1 = Y_2) + m_{k+2}\, I(Y_1 > Y_2)}$$

and

$$V_i^*(Y) = \frac{\delta(Y_i)}{m_{l+2}}\,, \qquad i = 2, ..., m_{l+1}\,.$$

Case II: When $X \in \chi_{2,l,k}$.

Similar to the Case I, we obtain

$$I\big(W_j \geq cW_1\big) = 1 \qquad \text{for} \quad j = 1, ..., l\,,$$
$$I\big(W_j + 1 \geq cW_1\big) = 1 \qquad \text{for} \quad j = 1, ..., l\,,$$
$$I\big(W_j + 1 \geq c(W_1 - 1)\big) = 1 \qquad \text{for} \quad j = 1, ..., k+1\,,$$

then

$$V_1^*(Y) = \frac{\delta(Y_1)}{m_{l+1}\, I(Y_1 = Y_2) + m_{k+2}\, I(Y_1 > Y_2)}$$

and

$$V_i^*(Y) = \frac{\delta(Y_i)}{m_{l+1}}\,, \qquad i = 2, ..., m_{l+1}\,.$$

Since $(X_1, ..., X_p)$ is sufficient and complete statistic for $(\theta_1, ..., \theta_p)$, the following theorem is now established.

Theorem 4.1. *The UMRUE of M is $\hat{M}_2 = \sum_{i=1}^{m_{l+1}} V_i^*(Y)$ where*

$$V_1^*(Y) = \frac{\delta(Y_1)}{m_u\, I(Y_1 = Y_2) + m_{k+2}\, I(Y_1 > Y_2)}\,,$$

$$V_i^*(Y) = \frac{\delta(Y_i)}{m_u}\,, \qquad i = 2, 3, ..., m_{l+1}\,,$$

and

$$u = \begin{cases} l+2, & \text{if } X \in \chi_{1,l,k}\,, \\ l+1, & \text{if } X \in \chi_{2,l,k}\,. \end{cases}$$

Consider next the *UMRUE* of the simultaneous estimation Q. Observe that

$$\begin{aligned} E_\theta Q &= E_\theta \sum_{i=1}^{p} \theta_i \, I\big(X_i + 1 \geq c X_{(1)i}\big) \\ &= \sum_{i=1}^{p} E_\theta \, \theta_i \, I\big(X_i + 1 \geq c X_{(1)i}\big) \\ &= \sum_{i=1}^{p} E_\theta \, \delta(X_i) \, I\big(X_i \geq c X_{(1)i}\big) \quad \text{(using Lemma 2.2)} \\ &= E_\theta \sum_{i=1}^{p} \delta(X_i) \, I\big(X_i \geq c X_{(1)i}\big) \\ &= E_\theta \, \hat{Q}_2 \quad \text{(say)} \, . \end{aligned}$$

Hence the following theorem.

Theorem 4.2. *The UMRUE of Q is $\hat{Q}_2$ whose the i-th component equals to $\delta(X_i)\, I(X_i \geq c X_{(1)i})$.*

5. INADMISSIBILITY OF THE NATURAL ESTIMATOR OF Q

In this section, we prove the inadmissibility of the natural estimator Q_1 of the simultaneous estimation Q using the technique of Stein (1973). The basic idea of Stein is to find an unbiased estimator $\Delta(X)$ of the risk difference $\Delta(\theta) = R(\theta, \delta + \psi) - R(\theta, \delta)$ and then finding a function ψ such that $\Delta(X) \leq 0 \ \forall x$ and $\Delta(x) < 0$ for some x. This technique has been used extensively in the simultaneous estimation problem when no selection involved (see for example Hudson (1978), Hwang (1982) and Chou (1991)). Consider the following rival estimator of $\hat{Q}_1$,

$$\begin{aligned} \hat{Q}_3 &= \hat{Q}_1 + \big(\phi_1(X)\, I_1(X), \, ..., \, \phi_p(X)\, I_p(X)\big) \\ &= \Big(\big(\delta(X_1) + \phi_1(X)\big)\, I_1(X), \, ..., \, \big(\delta(X_p) + \phi_p(X)\big)\, I_p(X)\Big) \, , \end{aligned} \tag{5.1}$$

where ϕ_i is any real-valued functions satisfying the conditions of Lemma 2.1. First, we find an unbiased estimator of the risk difference of estimators $\hat{Q}_3$ and $\hat{Q}_1$.

An unbiased estimator of the risk difference of $\hat{Q}_3$ and $\hat{Q}_1$ is

$$\begin{aligned}\Delta(\theta) &= R(Q,\hat{Q}_3) - R(Q,\hat{Q}_1)\\ &= E_\theta \sum_{i=1}^{p}\left(\frac{\delta(X_i)+\phi_i(X)}{\theta_i} - \log\left(\frac{\delta(X_i)+\phi_i(X)}{\theta_i}\right) - 1\right) I_i(X)\, I(X_i \geq t+1)\\ &\quad - E_\theta \sum_{i=1}^{p}\left(\frac{\delta(X_i)}{\theta_i} - \log\left(\frac{\delta(X_i)}{\theta_i}\right) - 1\right) I_i(X)\, I(X_i \geq t+1)\\ &= E_\theta \sum_{i=1}^{p}\left(\frac{\phi_i(X)}{\theta_i} - \log\left(1+\frac{\phi_i(X)}{\delta(X_i)}\right)\right) I_i(X)\, I(X_i \geq t+1)\,.\end{aligned}$$

Applying Lemma 2.1, we get

$$\Delta(\theta) = E_\theta \sum_{i=1}^{p}\left(\frac{\phi_i(X+e_i)}{\delta(X_i+1)}\, I_i(X+e_i)\, I(X_i \geq t) - \log\left(1+\frac{\phi_i(X)}{\delta(X_i)}\right) I_i(X)\, I(X_i \geq t+1)\right).$$

So that the following lemma is now established.

Lemma 5.1. *An unbiased estimator of the risk difference of the estimators $\hat{Q}_3$ and $\hat{Q}_1$ is given by*

(5.2)
$$D(X) = \sum_{i=1}^{p}\left(\frac{\phi_i(X+e_i)}{\delta(X_i+1)}\, I_i(X+e_i)\, I(X_i \geq t) - \log\left(1+\frac{\phi_i(X)}{\delta(X_i)}\right) I_i(X)\, I(X_i \geq t+1)\right).$$

Following Peng (1975) and Hudson (1978), we introduce the following notations. Let

$$\begin{aligned}&l = X_{(1)} \quad \text{(the largest observation)}\,,\\ &m = X_{(p)} \quad \text{(the smallest observation)}\,,\\ &N_i = \#\{j\colon X_j = i\}, \quad i = m, ..., l \ \text{ and } \ j = 1, ..., p\,,\\ &N = (N_m, ..., N_l)\,.\end{aligned}$$

If $X_i = r$, let

$$\begin{aligned}&\psi_i(X) = \zeta_r(N)\,,\\ &\delta(X_i) = \begin{cases} r, & r \geq t+1\,,\\ 0, & r < t+1\,,\end{cases}\\ &I(X_i \geq t+1) = I(r \geq t+1)\,,\\ &I_i(X) = I(r+1 \geq c\, l) = J_r \ \ \text{(say)}\,.\end{aligned}$$

Then, we have

$$\phi_i(X+e_i) = \zeta_{r+1}(N-e_r+e_{r+1})\,,$$
$$\delta(X_i+1) = \begin{cases} r+1, & r \geq t\,, \\ 0, & r < t\,, \end{cases}$$
$$I(X_i \geq t) = I(r \geq t)\,,$$
$$I_i(X+e_i) = J_{r+1}\,.$$

Define

$$\lfloor x \rfloor_{\text{ge}} = \text{smallest integer greater than or equal to } x\,,$$
$$[a]^+ = \max(0,a)\,.$$

Now, using the above notations, the unbiased estimator (5.2) becomes

(5.3)
$$D(X) = \sum_{r=m}^{l} N_r \left(\frac{\zeta_{r+1}(N+e_{r+1}-e_r)}{r+1}\, I(r \geq t)\, J_{r+1} - \log\left(1+\frac{\zeta_r(N)}{r}\right) I(r \geq t+1)\, J_r \right).$$

Next, we consider a solution of a general difference inequality that will be useful for constructing a class of improved estimators of the natural estimator $\hat{Q}_1$. Consider the following general difference inequality

(5.4)
$$\eta(X) = \sum_{r=m}^{l} N_r \left(\frac{\zeta_{r+1}(N+e_{r+1}-e_r)}{r+1}\, I(r \geq t)\, J_{r+1} - \left(\frac{\zeta_r(N)}{r} - \frac{\zeta_r^2(N)}{r^2}\right) I(r \geq t+1)\, J_r \right)$$
$$\leq 0\,.$$

In the following theorem, we solve the general difference inequality (5.4) borrowing some ideas from Dey and Chung (1991).

Theorem 5.1. *Consider the difference inequality (5.4). The function*

$$\zeta_r(N) = -\frac{b\,r^2}{d+w}$$

represents a solution of the inequality where

$$(\mathbf{1}) \quad w = \sum_{s=\lfloor cl-1 \rfloor}^{l} N_s\, s^2\,,$$

$$(\mathbf{2}) \quad d \geq 9 \sum_{s=\lfloor cl-1 \rfloor}^{l} N_s\,,$$

$$(\mathbf{3}) \quad 0 < b \leq \left[\sum_{r=\lfloor cl-1 \rfloor}^{l} N_r\, I(r \geq t+1) - \frac{47}{18} \right]^+ .$$

Proof: It is clear that

$$\begin{aligned}
\eta(X) \;&=\; \sum_{r=m}^{l} N_r \left(-\frac{b(r+1)}{d+w+2r+1}\, I(r \geq t)\, J_{r+1} + \left(\frac{b\,r}{d+w} + \frac{b^2 r^2}{(d+w)^2} \right) I(r \geq t+1)\, J_r \right) \\
&\leq\; b \sum_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r \left(\frac{r}{d+w} - \frac{(r+1)}{d+w+2r+1} + \frac{b\,r^2}{(d+w)^2} \right) I(r \geq t+1) \\
&\quad \big(\text{since } \; J_{r+1}\, I(r \geq t) \;\geq\; J_r\, I(r \geq t+1) \big) \\
&\leq\; b \sum_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} \frac{N_r (2r^2 + r - d - w)}{(d+w)\,(d+w+2r+1)}\, I(r \geq 2) + \frac{b^2\, w}{(d+w)^2} \\
&\leq\; \frac{2\,b\,w}{(d+w)^2} + \frac{3b \sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, r}{(d+w)^2} + \frac{b \sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r}{(d+w)^2} - \frac{b \sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, I(r \geq t+1)}{(d+w)} + \frac{b^2\, w}{(d+w)^2} \;.
\end{aligned}$$

Since

$$\frac{w}{d+w} \leq 1 \,,$$

$$\frac{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r}{d+w} \leq \frac{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r}{d} \leq \frac{1}{9} \quad \text{(from assumption (2))}$$

and

$$\frac{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, r}{d+w} \;\leq\; \frac{\sqrt{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r}\, \sqrt{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, r^2}}{d+w} \;\leq\; \frac{\sqrt{\sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r}\; \sqrt{w}}{2\sqrt{d}\,\sqrt{w}} \;\leq\; \frac{1}{6}$$

it follows that

$$\eta(X) \;\leq\; \frac{b \left(47/18 \;-\; \sum\limits_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, I(r \geq t+1) + b \right)}{d+w} \;.$$

Clearly, $\eta(X) \leq 0$ if $b \leq \sum_{r=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_r\, I(r \geq t+1) - 47/18$. This completes the proof of the theorem. □

Now, we are in a position to construct classes of dominating estimators of $\hat{Q}_1$ by solving the inequality $D(x) \leq 0 \;\forall x \in \chi$ using Theorem 5.1, where D is as in (5.3). In the following theorem, we construct a class of improved estimators of the natural estimator $\hat{Q}_1$.

Theorem 5.2. *Consider the rival estimator $\hat{Q}_3$ given in (5.1) where*

$$\phi_i(X) = \zeta_r(N) = -\frac{b\,r^2}{d \;+\; \sum\limits_{s=\lfloor cl-1 \rfloor_{\rm ge}}^{l} N_s\, s^2}$$

if $X_i = r$. Assume further,

$$(\mathbf{1}) \quad d \geq 9 \sum_{s=\lfloor cl-1 \rfloor_{\text{ge}}}^{l} N_s \,,$$

$$(\mathbf{2}) \quad 0 < b \leq \left[\min\left(\sqrt{d} \,, \sum_{s=\lfloor cl-1 \rfloor_{\text{ge}}}^{l} N_s \, I(s \geq t+1) - 47/18 \right) \right]^{+} .$$

Then, $\hat{Q}_3$ dominates $\hat{Q}_1$ in terms of risk where $\hat{Q}_1$ is as in (2.5).

Proof: Clearly, the ζ_r's satisfy the conditions of Lemma 2.1. It is easily seen that

$$\left| \frac{\zeta_r(N)}{r} \right| = \left| \frac{-b\, r}{d + \sum_{s=\lfloor cm-1 \rfloor_{\text{ge}}}^{l} N_s \, s^2} \right| \leq \frac{b\, r}{d + r^2} \leq \frac{b\, r}{2\, r \sqrt{d}} \leq \frac{1}{2} \,,$$

since $b \leq \sqrt{d}$. Then using Lemma 2.3 in (5.3) gives us

$$\begin{aligned} D(X) &\leq \sum_{r=m}^{l} N_r \left(\frac{\zeta_{r+1}(N + e_{r+1} - e_r)}{r+1} \, I(r \geq t) \, J_{r+1} - \left(\frac{\zeta_r(N)}{r} - \frac{\zeta_r^2(N)}{r^2} \right) I(r \geq t+1) \, J_r \right) \\ &= \eta(X) \,. \end{aligned}$$

Applying Theorem 5.1 in the above inequality completes the proof the theorem. □

The class of estimators in Theorem 5.2 dominates the natural estimator $\hat{Q}_1$, so that the natural estimator (2.5) is inadmissible.

6. SIMULATION RESULTS

In this section, we compute the biases and risks of the estimators $\hat{M}_1$, $\hat{M}_2$, $\hat{Q}_1$, $\hat{Q}_2$ and $\hat{Q}_3$ using the Monte Carlo simulation technique. Also, we compute the percentages of the risk improvement of the estimator $\hat{Q}_3$ over the estimators $\hat{Q}_i$, $i = 1, 2$. We follow the simulation procedure used by Tsue and Press (1982). First the value of p is chosen and then a set of $\{\theta_1, ..., \theta_p\}$ of parameter values are chosen at random within the range (c, d). In the second step, an observation X_i is randomly chosen from zero-truncated Poisson distribution $TP(\theta_i)$, $1 \leq i \leq p$. In step 3, the selection rule R, defined in (2.3), is used to select the subset. To estimate the parameters associated with the selected populations, we compute the biases of the estimators $\hat{M}_1$, $\hat{Q}_1$ and $\hat{Q}_3$ and the risks of the estimators

$\hat{M}_1$, $\hat{M}_2$, $\hat{Q}_1$, $\hat{Q}_2$ and $\hat{Q}_3$. The above procedure is repeated 4 000 times and the averages of the biases and risks are calculated. Then the percentage improvements of the estimator $\hat{Q}_3$ over $\hat{Q}_i$, $i = 1, 2$, namely

$$RPI(\hat{Q}_i, \hat{Q}_3) = \frac{R(Q, \hat{Q}_i) - R(Q, \hat{Q}_3)}{R(Q, \hat{Q}_i)} \times 100 \ , \qquad i = 1, 2 \ ,$$

are obtained. The above procedure is repeated a number of times with different sets of parameters in (c, d) and then the percentages of risk improvement and averages of biases are calculated and presented in Table 1 and Table 2, respectively. The simulation is carried out using Matlab version 7.4 and considering the values $d = 9 \sum_{s=\lfloor cl-1 \rfloor_{\text{ge}}}^{l} N_s$ and $b = \left[\min\left(\sqrt{d}, \sum_{s=\lfloor cl-1 \rfloor_{\text{ge}}}^{l} N_s I(s \geq 2) - 47/18\right)\right]^+$ with the number of populations $p = 3, 5, 7, 10$.

We observe the following facts from the simulation results. From Table 1, the risks of the estimators decrease as the range of θ_i's increases. For the percentages of the risk improvement, the two percentages increase for small values of θ_i's while decrease for large values and the highest values appear when all the θ_i's are in the interval $(1.5, 3)$. From Table 2, clearly, the biases of M_1, Q_1 and Q_3 are positive and gradually increasing as the range of θ_i's increases. Also, Q_1 has bias less than Q_3 and the bias of both of theme gradually increase as p increases.

Table 1: The risks of the estimators M_1 and M_2 and the percentage improvement of Q_3 over Q_1 and Q_2.

		Range of θ_i's				
		$(0.0, 0.5)$	$(0.5, 1.5)$	$(1.5, 3.0)$	$(3.0, 6.0)$	$(6.0, 15.0)$
$p = 3$	M_1	1.23	0.60	0.42	0.34	0.32
	M_2	3.44	0.61	0.40	0.33	0.32
	$RPI(\hat{Q}_1, \hat{Q}_3)$	0.28	2.65	4.46	3.42	1.90
	$RPI(\hat{Q}_2, \hat{Q}_3)$	0.27	2.08	4.28	3.15	1.83
$p = 5$	M_1	1.02	0.60	0.42	0.34	0.32
	M_2	2.81	0.60	0.37	0.32	0.31
	$RPI(\hat{Q}_1, \hat{Q}_3)$	1.07	9.72	16.89	12.34	6.43
	$RPI(\hat{Q}_2, \hat{Q}_3)$	1.07	8.97	16.53	12.05	6.30
$p = 7$	M_1	0.93	0.57	0.42	0.34	0.32
	M_2	2.68	0.53	0.36	0.31	0.31
	$RPI(\hat{Q}_1, \hat{Q}_3)$	2.03	15.85	22.77	16.95	8.29
	$RPI(\hat{Q}_2, \hat{Q}_3)$	2.02	14.50	22.27	16.68	8.15
$p = 10$	M_1	0.88	0.53	0.37	0.32	0.31
	M_2	3.04	0.62	0.36	0.31	0.30
	$RPI(\hat{Q}_1, \hat{Q}_3)$	2.79	19.33	27.36	20.75	10.08
	$RPI(\hat{Q}_2, \hat{Q}_3)$	2.79	19.32	27.32	20.52	10.07

Table 2: The biases of the estimators M_1, Q_1 and Q_3.

		Range of θ_i's				
		$(0.0, 0.5)$	$(0.5, 1.5)$	$(1.5, 3.0)$	$(3.0, 6.0)$	$(6.0, 15.0)$
$p = 3$	M_1	0.44	1.42	2.56	4.67	9.95
	Q_1	1.31	4.17	7.56	13.84	29.68
	Q_3	1.30	4.08	7.31	13.50	29.31
$p = 5$	M_1	0.43	1.33	2.59	4.80	10.62
	Q_1	2.15	6.54	12.70	23.61	52.68
	Q_3	2.12	5.96	11.09	21.52	50.39
$p = 7$	M_1	0.44	1.43	2.60	4.64	10.85
	Q_1	3.10	9.78	17.85	32.12	75.44
	Q_3	3.00	8.39	14.80	28.24	71.21
$p = 10$	M_1	0.41	1.31	2.55	4.42	10.56
	Q_1	4.07	13.11	25.48	44.17	105.55
	Q_3	3.89	10.79	20.17	37.52	98.31

ACKNOWLEDGMENTS

The paper is a part of Ph.D. work of the second author. He is thankful to ICCR-India for providing scholarship to continue his study for Ph.D.

REFERENCES

[1] Chou, P. (1991). Simultaneous estimation in discrete multivariate exponential families, *Ann. Statist.*, **19**, 314–328.

[2] Dey, D.K. and Chung, Y. (1991). Multiparameter estimation in truncated power series distributions under the Stein's loss, *Commun. Statist. – Theory Meth.*, **20**(1), 309–326.

[3] Dey, D.K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss, *Ann. Statist.*, **13**, 1581–1591.

[4] Gangopadhyay, A.K. and Kumar, S. (2005). Estimating average worth of the selected subset from two-parameter exponential populations, *Commun. Statist. – Theory Meth.*, **34**, 2257–2267.

[5] Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules, *Technometrics*, **7**, 225–245.

[6] Gupta, S.S. and Huang, D.Y. (1975). *On subset selection procedures for Poisson populations and some applications to the multinomial selection problems.* In "Appl. Statist." (R.P. Gupta, Ed.), North-Holland, Amsterdam, 97–109.

[7] Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation, *Ann. Statist.*, **6**(3), 473–484.

[8] Hwang, J.T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases, *Ann. Statist.*, **10**, 857–867.

[9] James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proceedings Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, **1**, 316–379, University of California Press.

[10] Jani, P.N. (1977). Minimum variance unbiased estimation for some left-truncated modified power series distributions, *Sankhya*, **39**, B, 258–278.

[11] Jeyaratnam, S. and Panchapakesan, S. (1984). *An estimation problem relating to subset selection for normal populations.* In "Design of Experiments: Ranking and Selection" (T.J. Santner and A.C. Tamhane, Eds.), Marcel Dekker, New York, 287–302.

[12] Jeyaratnam, S. and Panchapakesan, S. (1986). Estimation after subset selection from exponential populations, *Comm. Statist. Theory and Meth.*, **15**(11), 3459–3473.

[13] Kumar, S. and Gangopadhyay, A.K. (2005). Estimating parameters of a selected Pareto population, *Statist. Methodology*, **2**, 121–130.

[14] Lehmann, E.L. (1951). A general concept of unbiasedness, *Ann. Math. Statist.*, **22**, 578–592.

[15] Misra, N. (1994). Estimation of the average worth of the selected subset of gamma populations, *Sankhya*, **56**, B, Pt. 3, 344–355.

[16] Misra, N.; van der Meulen, E.C. and Branden, K.V. (2006a). On estimating the scale parameter of selected gamma population under the scale invariant squares error loss function, *J. Comp. Appl. Math.*, **186**(1), 268–282.

[17] Misra, N.; van der Meulen, E.C. and Branden, K.V. (2006b). On some inadmissibility results for the scale parameters of selected gamma populations, *J. Statist. Planning Inference*, **36**(7), 2340–2351.

[18] Nematollahi, N. and Motamed-Shariati, F. (2009). Estimation of the scale parameter of the selected gamma population under the entropy loss function, *Commun. Statist. – Theory Meth.*, **8**, 208–221.

[19] Peng, J.C. (1975). Simultaneous estimation of the parameters of independent Poisson distribution, *Technical Report No.78*, Department of Statistics, Stanford University.

[20] Robbins, H.E. (1988). *The UV method of estimation.* In "Statistical Decision Theory and Related Topics" (S.S. Gupta and J.O. Berger, Eds.), Springer, New York, IV, **1**, 265–270.

[21] Sackrowitz, H. and Samuel-Cahn, E. (1984). Estimation of the mean of a selected negative exponential population, *J. Roy. Statist. Soc. Ser. B*, **46**, 242–249.

[22] Sill, M.W. and Sampson, A.R. (2007). Extension of a two-stage conditionally unbiased estimator of the selected population to the bivariate normal case, *Commun. Statist. – Theory Meth.*, **36**, 801–813.

[23] Stein, C. (1973). Estimation of the mean of a multivariate distribution, *Proc. Prague Symp. Asymp. Statist.*, 345–381.

[24] Tsui, K.W. and Press, S.J. (1982). Simultaneous estimation of several Poisson parameters under k-normalized squared error loss, *Ann. Statist.*, **10**, 93–100.

[25] Vellaisamy, P. (1992). Average worth and simultaneous estimation of the selected subset, *Ann. Inst. Statist. Math.*, **44**(3), 551–562.

[26] Vellaisamy, P. (1996). Simultaneous estimation of the selected subset of uniform populations, *J. of Appl. Statist. Science*, **5**(1), 39–46.

[27] Vellaisamy, P. and Jain, S. (2008). Estimating the parameter of the population selected from discrete exponential family, *Statist. Probab. Lett.*, **78**(9), 1076–1087.

[28] Vellaisamy, P. and Sharma, D. (1990). On estimation after subset selection from exponential populations, *Sankhya*, **52**, B, Pt. 3, 305–318.

REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 21–35

A FOLDING METHOD FOR EXTREME QUANTILES ESTIMATION

Authors: Armelle Guillou
– Institut de Recherche Mathématique Avancée,
UMR 7501, Université de Strasbourg et CNRS,
7 rue René Descartes, 67084 Strasbourg Cedex, France
guillou@math.u-strasbg.fr

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement,
Orme des Merisiers / Bat.701 C.E. Saclay, 91191 Gif-sur-Yvette, France
philippe.naveau@lsce.ipsl.fr

Alexandre You
– Université Paris VI, LSTA,
Boîte 158, 175 rue du Chevaleret, 75013 Paris, France
alexandreyou@yahoo.fr

Received: June 2009 Revised: February 2010 Accepted: March 2010

Abstract:

- In order to estimate extreme quantiles from independent and identically distributed random variables, we propose and study a novel folding procedure that improves quantile estimates obtained from the classical Peaks-Over-Threshold method (POT) used in Extreme Value Theory. The idea behind the folding approach is to connect the part of a distribution above a given threshold with the one below it. A simplified version of this approach was studied by You *et al.* (2010). In this paper, an extension based on two thresholds is proposed to better combine the folding scheme with the POT approach. Simulations indicate that this new strategy leads to improved extreme quantiles estimates for finite samples. Asymptotic normality of the folded POT estimators is also derived.

Key-Words:

- *extreme quantile estimation; peaks-over-thresholds; generalized Pareto distribution; folding; generalized probability-weighted moments estimators.*

AMS Subject Classification:

- 62G32, 62G20.

1. MOTIVATION

The study of extremes has grown steadily since the pioneering work of Fisher & Tippett (1928). One of the most famous approaches in Extreme Value Theory is the Peaks-Over-Threshold (POT) method which can be described as follows. Let $\aleph := \{X_1, ..., X_n\}$ be a sample of independent random variables from an unknown distribution function F and consider the N_{u_n} exceedances above a fixed threshold u_n, that is $Y_1, ..., Y_{N_{u_n}}$ where $Y_j := X_{i_j} - u_n$, when $X_{i_j} > u_n$. According to Pickands (1975), for a large class of underlying distributions F, as the threshold u_n increases, the distribution of the exceedances F_{u_n} given by $F_{u_n}(t) := \mathbb{P}\big(X - u_n \leq t \,|\, X > u_n\big)$ asymptotically converges to a Generalized Pareto Distribution (GPD) defined as

$$G_{\gamma,\sigma_n}(x) = \begin{cases} 1 - \big(1 + \gamma\, x/\sigma_n\big)^{-1/\gamma} & \text{if } \gamma \neq 0\,, \\ 1 - \exp\big(-x/\sigma_n\big) & \text{if } \gamma = 0\,, \end{cases}$$

where $\sigma_n = \sigma(u_n) > 0$ and $x \geq 0$ if $\gamma \geq 0$ and $0 \leq x < -\sigma_n/\gamma$ if $\gamma < 0$. This result leads to the so-called POT estimator of a high quantile $x_p := F^{\leftarrow}(1-p)$ with $F^{\leftarrow}$ the inverse function of F

$$\widehat{x}_p(u_n) \;=\; u_n + \frac{\widehat{\sigma}}{\widehat{\gamma}} \left\{ \left(\frac{p}{1 - F_n(u_n)} \right)^{-\widehat{\gamma}} - 1 \right\}, \tag{1.1}$$

where $(\widehat{\gamma}, \widehat{\sigma})$ are some estimators of the parameters (γ, σ_n) and $F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$ denotes the empirical distribution function.

Recently, You *et al.* (2010) proposed a so-called folding procedure to improve the estimation of x_p. This approach is inspired by perfect sampling techniques used in simulation studies (Corcoran & Schneider, 2003) and the idea is to connect the lower and upper parts of a distribution. More precisely the explicit formulation of this folding transformation and its fundamental property are encapsulated in their Proposition 1 which is recalled below.

Proposition 1.1. *Let X be a random variable with an absolutely continuous distribution function F, u a real number such that $u < \tau_F$ where $\tau_F = \sup\big\{x \in \mathbb{R} \colon F(x) < 1\big\}$ is the right endpoint of F and H another absolutely continuous distribution function with the same support as F and such that $H(u) \geq F(u)$. Define the following random variable*

$$X^{(H,F)}(u) := \begin{cases} H^{\leftarrow}\left(\dfrac{\overline{F}(u)}{F(u)}\, F(X) + F(u) \right) & \text{if } X < u\,, \\ X & \text{if } X \geq u\,, \end{cases} \tag{1.2}$$

where $\overline{F} := 1 - F$ and $H^{\leftarrow}$ is the inverse function of H. Then

$$\mathbb{P}\big(X^{(H,F)}(u) > x\big) \;=\; \mathbb{P}\big(X > x | X > u\big) + \frac{F(u)}{\overline{F}(u)} \big(\overline{H}(x) - \overline{F}(x)\big)\,, \quad x > u\,. \tag{1.3}$$

A very important special case occurs when H is chosen to be equal to F in (1.2). In this context, the random variable $X^{(F,F)}(u)$ has the same probability distribution as the conditional variable $\big[X|\, X>u\big]$, the latter being the variable of interest for the aforementioned POT method. We call $X^{(F,F)}(u)$ the folded transformation of X and we denote it as $X^{(F)}(u) := X^{(F,F)}(u)$. In practice, F is unknown and the folding transformation cannot be applied directly. One must substitute the unknown F by suitable proxies. The choice of a proxy is especially sensitive for the inverse function $F^{\leftarrow}$. This explains the introduction of H in the definition of $X^{(H,F)}(u)$. To study the effect of choosing the proxy H instead of F in the folding procedure, we introduce the difference $\Delta_u^{(H,F)}(x) := \big|\mathbb{P}\big(X^{(H,F)}(u) \le x\big) - \mathbb{P}\big(X^{(F,F)}(u) \le x\big)\big|$ for $x > u$. According to Proposition 1.1, we can write

$$\Delta_u^{(H,F)}(x) \,=\, \frac{F(u)}{\overline{F}(u)}\,\big|\overline{H}(x) - \overline{F}(x)\big|\,.$$

Second-order extreme value theory (e.g. de Haan & Ferreira, 2006) provides the necessary tools to characterize the behavior of $\Delta_u^{(H,F)}(x)$ for a specific H.

Proposition 1.2. *Assume that F satisfies the following second-order condition. There exists some positive function $a(\cdot)$ and some positive or negative function $A(\cdot)$ with $\lim_{t\to\infty} A(t) = 0$ such that*

$$\lim_{t\to\infty} \frac{1}{A(t)}\left(\frac{U(tx)-U(t)}{a(t)} - D_\gamma(x)\right) = B(x)\;, \qquad x>0\;, \tag{1.4}$$

where $U := (1/\overline{F})^{\leftarrow}$, $D_\gamma(x) := \frac{x^\gamma - 1}{\gamma}$ if $\gamma \neq 0$ and $\log x$ if $\gamma = 0$, and B is some function that is not a multiple of D_γ. If the tail distribution $\overline{H}(x) := 1 - H(x)$ is chosen to behave as a GPD tail such that

$$\overline{H}(x) \,=\, \overline{F}(u)\,\overline{G}_{\gamma,\sigma(u)}(x-u)\;, \tag{1.5}$$

with $x > u$ and $\sigma(u) = a\big(1/\overline{F}(u)\big)$, then for all y satisfying $1+\gamma y > 0$, we have

$$\begin{aligned}\lim_{u\to\tau_F} \frac{1}{|\alpha(u)|}\,\Delta_u^{(H,F)}\big(u+\sigma(u)\,y\big) \;&=\; \lim_{u\to\tau_F}\left|\frac{1}{\alpha(u)}\left(\frac{\overline{F}\big(u+\sigma(u)\,y\big)}{\overline{F}(u)} - \overline{G}_\gamma(y)\right)\right| \\ &=\; \big(\overline{G}_\gamma(y)\big)^{1+\gamma}\,\big|B_{\gamma,\rho}\big(1/\overline{G}_\gamma(y)\big)\big|\end{aligned} \tag{1.6}$$

where $\overline{G}_\gamma := \overline{G}_{\gamma,1}$, $\alpha(u) := A\big(1/\overline{F}(u)\big)$ and

$$B_{\gamma,\rho}(x) \,:=\, \begin{cases} \dfrac{1}{\rho}\left(\dfrac{x^{\gamma+\rho}-1}{\gamma+\rho} - \dfrac{x^\gamma - 1}{\gamma}\right) & \text{if } \gamma\neq 0,\ \rho\neq 0\,, \\[2ex] \dfrac{1}{\gamma}\left(x^\gamma \log x - \dfrac{x^\gamma - 1}{\gamma}\right) & \text{if } \rho = 0 \neq \gamma\,, \\[2ex] \dfrac{1}{\rho}\left(\dfrac{x^\rho - 1}{\rho} - \log x\right) & \text{if } \rho\neq 0 = \gamma\,, \\[2ex] \dfrac{1}{2}(\ln x)^2 & \text{if } \rho = \gamma = 0\,. \end{cases}$$

The first equality in (1.6) tells us that choosing H as a GPD approximation, see (1.5), implies that the rate of convergence of $\Delta_u^{(H,F)}$ towards zero is identical to the one obtained by working with exceedances. The second equality in (1.6) simply restates the result derived by de Haan & Ferreira (2006, p. 48) about the relationship between the rate of convergence and the second-order auxiliary function $A(\cdot)$. The main consequence of Proposition 1.2 is that a GPD can be viewed as the appropriate choice for the distribution function H. In real applications, we do not know the parameters of such a GPD and a first estimation has to be given before implementing our folding procedure. This also means that any reasonable GPD estimation procedure can be used to initialize our algorithm, the better the estimation of $\sigma(u)$ and γ, the better the efficiency of the folding procedure. Still, our main goal in this paper is not to compare all existing GPD estimation methods (e.g., Smith, 1987; Greenwood *et al.*, 1979) and to find the best one (if one could do that). Instead, our aim is to study our folding approach with a specific estimation method for which we have experience with (Diebolt *et al.*, 2004, 2007).

At this stage, our approach can be viewed as the mixing of two elements, the folding procedure described by Proposition 1.1 and the POT method. Each element is associated with a particular threshold choice. For the sake of simplicity, You *et al.* (2010) considered that both thresholds were equal. This is not necessary. One threshold could be chosen for computing the preliminary GPD parameters estimates and another one for the folding transformation itself. In this paper, we follow this path. We propose and study a novel folding approach based on two thresholds u_n and u'_n such that $u_n = o(u'_n)$. Compared to a conventional approach and to our past folding procedure, simulations clearly indicate that this new double-threshold folding approach significantly reduces the mean squared error of extreme quantile estimates, particularly for small and moderate sample sizes (see Section 4). Asymptotic properties of our GPD parameters estimators are derived (see Section 3). The proof of our results are postponed to the appendix. Results presented in Sections 3 and 4 solely focus on heavy tailed distributions because our previous study (You *et al.*, 2010) indicated that the folding gain is the strongest for this type of tails.

2. A NEW FOLDING PROCEDURE WITH TWO THRESHOLDS

Suppose that the variable $\left[X - u_n | \, X > u_n\right]$ approximatively follows a GPD(γ, σ_n) for some large threshold u_n. The thresholding stability property of the GPD basically means that $\left[X - u'_n | \, X > u'_n\right]$ can also be approximated by a GPD$\big(\gamma, \, \sigma_n + \gamma(u'_n - u_n)\big)$ for any $u'_n > u_n$. In other words, the tail $\overline{F}(t) = \overline{F}_{u'_n}(t - u'_n)\,\overline{F}(u'_n)$ can be approximated by $\overline{G}_{\gamma,\sigma_n+\gamma(u'_n-u_n)}(t - u'_n)\,\overline{F}(u'_n)$. In terms of inverse distributions, this approximation can be expressed as $\overline{F}^{\leftarrow}(t) \simeq$

$\overline{G}^{\leftarrow}_{\gamma,\sigma_n+\gamma(u'_n-u_n)}\Big(\frac{t}{\overline{F}(u'_n)}\Big)+u'_n$. According to Proposition 1.1, the folded variable $X^{(F)}(u'_n)$ can be rewritten as

$$X^{(F)}(u'_n)=\overline{F}^{\leftarrow}\left(\overline{F}(u'_n)\left[1-\frac{F(X)}{F(u'_n)}\right]\right), \qquad \text{if } X<u'_n\,.$$

By plugging the approximation for $\overline{F}^{\leftarrow}$ in the expression of $X^{(F)}(u'_n)$, it is natural to define the following folded variables

$$\widehat{X}_i^{(F)}(u'_n)=\begin{cases}\overline{G}^{\leftarrow}_{\widehat{\gamma},\widehat{\widetilde{\sigma}}'}\left(1-\dfrac{F_n(X_i)}{F_n(u'_n)}\right)+u'_n\,, & \text{if } F_n(X_i)<F_n(u'_n)\,,\\ X_i\,, & \text{if } F_n(X_i)\geq F_n(u'_n)\,,\end{cases} \tag{2.1}$$

where $\widehat{\widetilde{\sigma}}':=\widehat{\sigma}+\widehat{\gamma}(u'_n-u_n)$ and $(\widehat{\gamma},\widehat{\sigma})$ are estimated from the exceedances $(Y_1,...,Y_{N_{u_n}})$. Note that the folding transformation given by (2.1) does not depend on the numerical values of the observations X_i when $X_i<u'_n$ but only on their ranks because $nF_n(X_i)$ is equal to the rank of the observation X_i.

Equation (2.1) allows us to describe our new folding procedure as follows:

Step 1. Select one threshold u_n and estimate $(\widehat{\gamma},\widehat{\sigma})$ of the GPD parameters (γ,σ_n) from the exceedances above u_n. Select a second threshold $u'_n>u_n$ and calculate $\widehat{\widetilde{\sigma}}':=\widehat{\sigma}+\widehat{\gamma}(u'_n-u_n)$.

Step 2. Build the folded version $\aleph^{(F)}:=\Big\{\widehat{X}_1^{(F)}(u'_n),...,\widehat{X}_n^{(F)}(u'_n)\Big\}$ using transformation (2.1).

Step 3. Estimate the GPD parameters $\big(\widehat{\gamma}^{(F)},\widehat{\sigma}^{(F)}\big)$ from the folded sample $\aleph^{(F)}$.

Step 4. Compute the POT extreme quantile estimator $\widehat{x}_p^{(F)}(u'_n)$ (according to (1.1)) with the estimates $\big(\widehat{\gamma}^{(F)},\widehat{\sigma}^{(F)}\big)$.

In steps 1 and 3, any reasonable GPD estimator of (γ,σ_n) could be used. Here we implement the generalized probability-weighted moments (GPWM) (Diebolt *et al.*, 2004, 2007). This is an extension of the classical probability-weighted moments method (Greenwood *et al.*, 1979) and it can be described as follows. Let ω be a continuous function, null at zero, and which admits a right derivative at zero. The GPWM is defined as $\nu_\omega=\mathbb{E}\big[Z\,\omega\big(\overline{G}_{\gamma,\sigma}(Z)\big)\big]$ where Z follows a GPD(γ,σ) with $\gamma<2$. If we denote by W the primitive of ω null at zero, then an integration by parts allows us to write ν_ω as $\nu_\omega=\int_0^\infty W\big(\overline{G}_{\gamma,\sigma}(x)\big)\,dx$. Diebolt *et al.* (2004, 2007) proposed and studied $\widehat{\nu}_{\omega,n}=\int_0^\infty W\big(\overline{F}_{n,u_n}(x)\big)\,dx$ as an estimator of ν_ω, where $\overline{F}_{n,u_n}$ corresponds to the exceedances empirical survival function defined by $\overline{F}_{n,u_n}(x)=\frac{1}{N_{u_n}}\sum_{i=1}^{N_{u_n}}\mathbb{1}\big\{X_i-u_n>x\big\}$. To estimate (γ,σ_n) by implementing a method-of-moments approach, two GPWMs are needed. In this work, the two functions ω_1 and ω_2 are equal to $\omega_1(x)=x$ and $\omega_2(x)=x^{3/2}$ (see Diebolt *et al.*, 2004, for a justification of these choices).

3. ASYMPTOTIC NORMALITY

Before stating our main result, we need to prove the asymptotic normality of the pair $(\widehat{\gamma}, \widehat{\widehat{\sigma}}'/\sigma'_n)$.

Lemma 3.1. *Let F be three times differentiable such that its inverse $F^{\leftarrow}$ exists. Let V and M be two functions defined as $V(t) = \overline{F}^{\leftarrow}(e^{-t})$ and $M(t) = \frac{V''(\ln t)}{V'(\ln t)} - \gamma$. Suppose the following conditions hold*

$$M \text{ is of constant sign at } \infty \tag{3.1}$$

and

$$\exists\, \rho < 0 : \ |M| \in RV_\rho \text{ with a normalized slowly varying function (see Bingham et al., 1987)}. \tag{3.2}$$

Then, for $\gamma \in (0, 3/2)$ and for all C^1-functions ω_1 and ω_2, null at 0, conditionally on $\{N_{u_n} = k_n\}$ and $\{N_{u'_n} = k'_n\}$ with $u_n = o(u'_n)$, and for all intermediate sequences $k_n > k'_n \to \infty$ such that $\sqrt{k_n}\, a_n \to \lambda \in \mathbb{R}$, we have that

$$\sqrt{k_n} \begin{pmatrix} \widehat{\gamma} - \gamma \\ \frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1 \end{pmatrix} \xrightarrow{\ d\ } \mathcal{N}\left(\lambda \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix} B_1 \,,\ \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix} \Sigma_{1,1} \begin{pmatrix} 1 \\ \frac{1}{\gamma} \end{pmatrix}' \right),$$

where

$$\begin{cases} a_n := M\big(e^{V^{\leftarrow}(u_n)}\big)\,, \qquad \sigma'_n := V'\big(V^{\leftarrow}(u'_n)\big)\,, \\ \mathbf{B} := \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \mathbf{A}\,\mathbf{C} \qquad \text{where } \ \mathbf{A} := DT_{(\omega_1,\omega_2)}(\nu^1_{\omega_1}, \nu^1_{\omega_2})\,, \\ \mathbf{C} := \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1(\gamma+\rho) - \phi_1(\gamma)}{\rho} \\ \frac{\phi_2(\gamma+\rho) - \phi_2(\gamma)}{\rho} \end{pmatrix} \quad \text{where } \ \phi_j(\gamma) := \displaystyle\int_0^1 W_j(u)\, u^{-\gamma-1}\, du\,, \ \ j \in \{1,2\}\,, \\ \mathbf{\Sigma} := \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix} = \mathbf{A}\,\mathbf{\Gamma}\,\mathbf{A}^{\mathrm{T}}\,, \end{cases}$$

and $\mathbf{\Gamma}$ is the variance-covariance matrix of the pair $(Y_{\omega_1}, Y_{\omega_2})$ defined as

$$Y_{\omega_1} = \int_0^1 t^{-\gamma-1} \omega_1(t)\, \mathbb{B}(t)\, dt \qquad \text{and} \qquad Y_{\omega_2} = \int_0^1 t^{-\gamma-1} \omega_2(t)\, \mathbb{B}(t)\, dt$$

with $\mathbb{B}$ a Brownian bridge on $[0,1]$.

The case $\rho = 0$ is excluded from Lemma 3.1. This corresponds to $M(t) = \ell(t)$, a slowly varying function, since in that case the limiting distribution depends explicitely on $\ell(\cdot)$, and the two sequences u_n and u'_n. Also this restriction is not really a problem since most of the classical distributions in the Fréchet domain of attraction have a second order parameter $\rho < 0$ (except the loggamma).

Now, we can establish our main asymptotic result which shows that, in the case where $\gamma \in (0, 3/2)$, the estimators based on the double-threshold folding approach have a similar asymptotic normality as the one derived in You *et al.* (2010) in case of one threshold.

Theorem 3.1. *Under the same assumptions stated in Lemma 3.1, we have*

$$\sqrt{k_n} \begin{pmatrix} \dfrac{\widehat{\nu}^{(F)}_{\omega_1,n}}{\sigma'_n} - \nu^1_{\omega_1} \\ \dfrac{\widehat{\nu}^{(F)}_{\omega_2,n}}{\sigma'_n} - \nu^1_{\omega_2} \end{pmatrix} \xrightarrow{d} \mathcal{N}\big(\lambda\, B_1\, \mathbb{F},\, \mathbb{F}\, \Sigma_{1,1}\, \mathbb{F}'\big)$$

where

$$\mathbb{F} := \begin{pmatrix} -\dfrac{1}{\lambda} \displaystyle\int_0^1 u^{-\gamma} \ln u\, \omega_1(u)\, du \\ -\dfrac{1}{\lambda} \displaystyle\int_0^1 u^{-\gamma} \ln u\, \omega_2(u)\, du \end{pmatrix}$$

and

$$\nu^1_{\omega_j} = \int_0^\infty W_j\big(\overline{G}_{\gamma,1}(x)\big)\, dx \qquad \text{for} \quad j = 1, 2\ .$$

Note that this convergence in distribution does not hold in case $\gamma \leq 0$.

4. A SIMULATION STUDY

The aim of this section is to illustrate the superiority of the double-threshold folding over the conventional (Diebolt *et al.*, 2007) and the simple folding approaches (You *et al.*, 2010), in particular in terms of the mean squared error for small and moderate sample sizes. Simulations were performed for four sample sizes $n = 100,\ 500,\ 1\,000$ and $5\,000$ from a Burr $(1, 2, 0.5)$ distribution defined by $\overline{F}(x) = (1 + x^2)^{-1/2}$ and from a standard Fréchet distribution defined by $\overline{F}(x) = 1 - e^{-1/x}$, respectively. For these two distributions, $\gamma = 1$ and $\rho < 0$. For each value of n, $5\,000$ samples were generated and k_n was chosen such that the condition $\sqrt{k_n}\, a_n \to \lambda$ was satisfied, which corresponds to $k_n \simeq c_1\, n^{4/5}$ for the Burr distribution and to $k_n \simeq c_2\, n^{2/3}$ for the Fréchet distribution. The threshold u_n was chosen from $u_n = F^{\leftarrow}\big(1 - \frac{k_n}{n}\big)$. Three return levels for three return periods, $t = 100, 200$ and $1\,000$, were computed. Concerning the choice of the second threshold for our double-threshold folding method, we selected $k'_n = c_3\, n^{3/5}$ for the Burr distribution and $k'_n = c_4\, n^{1/2}$ for the Fréchet distribution such that $u_n = o(u'_n)$. Tables 1–3 and 4–6 display the bias and the root mean squared error (RMSE) of the quantile x_p for the Burr and Fréchet distributions, respectively.

Table 1: Burr$(1, 2, 0.5)$ distribution — Bias (RMSE) of the return level estimates corresponding to a return period $t = 100$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	35.7 (240.4)	10.7 (183.0)	1.6 (103.8)
500	288	42	16.4 (195.4)	4.0 (106.2)	1.5 (46.7)
1 000	502	63	8.5 (78.8)	2.3 (54.1)	1.3 (35.3)
5 000	1 820	166	2.6 (21.1)	1.1 (16.8)	0.3 (16.9)

Table 2: Same as Table 1 but for the return period $t = 200$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	117.0 (840.1)	45.6 (626.3)	11.2 (299.2)
500	288	42	54.2 (752.4)	15.9 (395.7)	6.9 (140.9)
1 000	502	63	27.1 (297.9)	8.1 (192.9)	5.1 (112.7)
5 000	1 820	166	6.7 (67.1)	2.1 (49.4)	0.8 (40.3)

Table 3: Same as Table 1 but for the return period $t = 1\,000$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	80	16	1 877.5 (16 252.5)	998.0 (11 748.4)	346.5 (3 785.8)
500	288	42	955.4 (17 449.2)	385.8 (8 783.8)	167.4 (2 029.8)
1 000	502	63	468.8 (6 702.8)	199.6 (3 890.0)	119.1 (1 820.5)
5 000	1 820	166	73.9 (990.5)	22.2 (589.4)	19.1 (504.8)

Table 4: Standard Fréchet distribution — Bias (RMSE) of the return level estimates corresponding to the return period $t = 100$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	48.8 (173.2)	15.4 (112.1)	15.0 (93.9)
500	188	33	17.2 (70.9)	7.5 (50.6)	8.5 (47.1)
1 000	300	47	12.2 (48.5)	6.8 (38.0)	6.8 (36.2)
5 000	877	106	3.5 (17.3)	2.5 (15.7)	1.6 (13.9)

Table 5: Same as Table 4 but for the return period $t = 200$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	158.0 (581.7)	56.4 (353.3)	46.3 (286.3)
500	188	33	55.9 (242.7)	25.5 (156.0)	27.7 (140.8)
1 000	300	47	39.5 (163.3)	22.3 (117.2)	23.0 (111.0)
5 000	877	106	11.8 (57.6)	8.6 (49.6)	6.8 (42.0)

Table 6: Same as Table 4 but for the return period $t = 1\,000$.

n	k_n	k'_n	Conventional	Folding with one threshold	Folding with two thresholds
100	64	15	2 304.7 (10 052.4)	935.6 (5 396.5)	638.6 (3 214.3)
500	188	33	785.8 (4 459.5)	367.6 (2 206.0)	370.6 (1 845.0)
1 000	300	47	536.3 (2 789.3)	298.5 (1 580.8)	309.6 (1 480.8)
5 000	877	106	151.0 (899.2)	110.1 (664.6)	99.4 (549.3)

These tables clearly show that the double-threshold folding improves considerably the RMSE, compared to the single-threshold folding and the conventional approach. This gain is emphasized for small and moderate sample sizes and for large return periods.

APPENDIX: DETAILED PROOFS

Proof of Proposition 1.2: The proof is mainly a consequence of Theorem 2.3.8 in de Haan & Ferreira (2006), which states that, if (1.4) holds, then

$$\lim_{u\to\tau_F} \frac{1}{\alpha(u)}\left(\frac{\overline{F}\big(u+\sigma(u)\,y\big)}{\overline{F}(u)} - \overline{G}_\gamma(y)\right) = \big(\overline{G}_\gamma(y)\big)^{1+\gamma} B_{\gamma,\rho}\left(\frac{1}{\overline{G}_\gamma(y)}\right)$$

for all y such that $1+\gamma\,y > 0$.

From (1.5), it follows that

$$\begin{aligned}\lim_{u\to\tau_F} \frac{\Delta_u^{(H,F)}\big(u+\sigma(u)\,y\big)}{|\alpha(u)|} &= \lim_{u\to\tau_F} F(u) \times \lim_{u\to\tau_F} \left|\frac{1}{\alpha(u)}\left(\overline{G}_\gamma(y) - \frac{\overline{F}\big(u+\sigma(u)\,y\big)}{\overline{F}(u)}\right)\right| \\ &= \big(\overline{G}_\gamma(y)\big)^{1+\gamma}\left|B_{\gamma,\rho}\left(\frac{1}{\overline{G}_\gamma(y)}\right)\right|. \qquad \square\end{aligned}$$

Proof of Lemma 3.1: First, note that the assumption $\sqrt{k_n}\, a_n \to \lambda \in \mathbb{R}$ can be rewritten as

$$\sqrt{k_n}\, M\left(\frac{1}{\overline{F}(u_n)}\right) \longrightarrow \lambda \in \mathbb{R}\,. \tag{A.1}$$

Now, let $\sigma_n = V'\big(V^{\leftarrow}(u_n)\big)$. We have

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1\right) &= \sqrt{k_n}\left(\frac{\widehat{\sigma} + \widehat{\gamma}(u'_n - u_n)}{\sigma'_n} - 1\right) \\
&= \sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right)\frac{\sigma_n}{\sigma'_n} + \left[\frac{\gamma}{\widehat{\gamma}}\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\, u_n} - 1\right) - \frac{\sqrt{k_n}}{\widehat{\gamma}}\,(\widehat{\gamma} - \gamma)\right]\frac{\widehat{\gamma}\, u_n}{\sigma_n}\,\frac{\sigma_n}{\sigma'_n} \\
&\quad + \sqrt{k_n}\left(\frac{\gamma\, u'_n}{\sigma'_n} - 1\right)\frac{\widehat{\gamma}}{\gamma} + \frac{1}{\gamma}\sqrt{k_n}\,(\widehat{\gamma} - \gamma) \\
&=: Q_{1,n} + Q_{2,n} + Q_{3,n} + \frac{1}{\gamma}\sqrt{k_n}\,(\widehat{\gamma} - \gamma)\,.
\end{aligned}$$

We know that $\sqrt{k_n}\big(\frac{\widehat{\sigma}}{\sigma_n} - 1\big)$ is asymptotically normal (Diebolt *et al.*, 2007) and

$$\frac{\sigma_n}{\sigma'_n} \sim \frac{\gamma\, u_n}{\gamma\, u'_n} \longrightarrow 0\,. \tag{A.2}$$

Therefore, it is clear that

$$Q_{1,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.3}$$

Now, remark that

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\, u_n} - 1\right) &= \sqrt{k_n}\left(\frac{V'\big(-\ln \overline{F}(u_n)\big)}{\gamma\, V\big(-\ln \overline{F}(u_n)\big)} - 1\right) \\
&= \frac{1}{\gamma}\sqrt{k_n}\, M\left(\frac{1}{\overline{F}(u_n)}\right)\left[\frac{M\left(\frac{1}{\overline{F}(u_n)}\right)}{\frac{V'\big(-\ln \overline{F}(u_n)\big)}{V\big(-\ln \overline{F}(u_n)\big)} - \gamma}\right]^{-1}.
\end{aligned}$$

To conclude with this term, we have to use the following lemma.

Lemma A.1 (Worms, 2000, p. 19)**.** *Suppose that $M(t) \to 0$ and $\frac{tM'(t)}{M(t)} \to \rho$ as $t \to \infty$. Then*

(i) *if $\gamma > 0$, we have*

$$\lim_{t\to\infty} M(e^t)\Big/\left[\frac{V'(t)}{V(t)} - \gamma\right] = \frac{\gamma + \rho}{\gamma}$$

and

$$\lim_{t\to\infty} M(e^t)\Big/\left[\frac{V(t)}{V'(t)} - \frac{1}{\gamma}\right] = -\gamma(\gamma + \rho)\,;$$

(**ii**) *if* $\gamma < 0$, *we have*

$$\lim_{t\to\infty}\left[\frac{V(\infty)-V(t)}{V'(t)}+\frac{1}{\gamma}\right]\Big/M(e^t) = \frac{1}{\gamma(\gamma+\rho)}$$

and

$$\lim_{t\to\infty}\left[\frac{V'(t)}{V(\infty)-V(t)}+\gamma\right]\Big/M(e^t) = -\frac{\gamma}{\gamma+\rho}\,.$$

Indeed by (A.1), we deduce that

$$\sqrt{k_n}\left(\frac{\sigma_n}{\gamma\, u_n}-1\right) \longrightarrow \frac{\lambda}{\gamma+\rho}\,.$$

Combining this convergence with (A.2) and the fact that $\sqrt{k_n}(\widehat{\gamma}-\gamma)$ is asymptotically normal (Diebolt *et al.*, 2007), we deduce that

$$Q_{2,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.4}$$

Similarly

$$\begin{aligned}\sqrt{k_n}\left(\frac{\gamma\, u'_n}{\sigma'_n}-1\right) &= \gamma\,\sqrt{k_n}\left(\frac{V\big(-\ln\overline{F}(u'_n)\big)}{V'\big(-\ln\overline{F}(u'_n)\big)}-\frac{1}{\gamma}\right)\\ &= \gamma\,\sqrt{k_n}\,M\left(\frac{1}{\overline{F}(u_n)}\right)\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{M\Big(\frac{1}{\overline{F}(u_n)}\Big)}\left[\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{\frac{V\left(-\ln\overline{F}(u'_n)\right)}{V'\left(-\ln\overline{F}(u'_n)\right)}-\frac{1}{\gamma}}\right]^{-1}.\end{aligned}$$

Now since $\gamma > 0$ and $|M| \in RV_\rho$ with $\rho < 0$

$$\frac{M\Big(\frac{1}{\overline{F}(u'_n)}\Big)}{M\Big(\frac{1}{\overline{F}(u_n)}\Big)} \longrightarrow 0\,. \tag{A.5}$$

Consequently, using again the abovementioned lemma in Worms (2000), we deduce that

$$Q_{3,n} \xrightarrow{\mathbb{P}} 0\,. \tag{A.6}$$

Finally, going back to (A.3), (A.4) and (A.6), we get

$$\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n}-1\right) = \frac{1}{\gamma}\,\sqrt{k_n}\,(\widehat{\gamma}-\gamma)+o_{\mathbb{P}}(1)\,.$$

Lemma 3.1 then follows from Diebolt *et al.* (2007). □

Proof of Theorem 3.1: It is a direct application of the proof of Theorem 1 in You *et al.* (2010) combining with our Lemma 3.1 and using the following decomposition: conditionally on $\{N_{u_n} = k_n\}$ and $\{N_{u'_n} = k'_n\}$, we have

$$\begin{aligned}
&\sqrt{k_n}\left(\frac{\widehat{\nu}^{(F)}_{\omega,n}}{\sigma'_n} - \int_0^\infty W\big(\overline{G}_\gamma(x)\big)\,dx\right) = \\
&= \sqrt{k_n}\left(\frac{\widehat{\nu}^{(F)}_{\omega,n}}{\sigma'_n} - \frac{1}{\sigma'_n}\int_0^\infty W\Big(\widetilde{\overline{F}}^{(F)}_{n,u'_n}(x)\Big)\,dx\right) \\
&\quad + \left(1-\frac{k'_n}{n}\right)\sqrt{k_n}\int_0^\infty\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\sigma}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big)\,\omega\big(\overline{G}_\gamma(x)\big)\,dx \\
&\quad + \frac{k'_n}{n}\sqrt{k_n}\int_0^\infty\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\,\omega\big(\overline{G}_\gamma(x)\big)\,dx \\
&\quad + \sqrt{k_n}\int_0^\infty\int_0^1 (1-t)\left(\left(1-\frac{k'_n}{n}\right)\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\sigma}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big) + \frac{k'_n}{n}\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\right)^2 \\
&\qquad \times\ \omega'\left(\overline{G}_\gamma(x) + \left(\left(1-\frac{k'_n}{n}\right)\Big(\overline{G}_{\widehat{\gamma},\frac{\widehat{\sigma}'}{\sigma'_n}}(x) - \overline{G}_\gamma(x)\Big) + \frac{k'_n}{n}\Big(\overline{F}_{n,u'_n}(\sigma'_n x) - \overline{G}_\gamma(x)\Big)\right)t\right)\,dt\,dx\,,
\end{aligned}$$

where $\widetilde{\overline{F}}^{(F)}_{n,u'_n}(x) = \left(1-\dfrac{k'_n}{n}\right)\overline{G}_{\widehat{\gamma},\widehat{\sigma}'}(x) + \dfrac{k'_n}{n}\,\overline{F}_{n,u'_n}(x)$.

All the details of the proof are given on the web page `http://www-irma.u-strasbg.fr/~guillou/Proof_folding_thm3-1.pdf`. Now, we will prove that our theorem does not hold in case $\gamma \leq 0$. Indeed if $\gamma < 0$, then

$$\begin{aligned}
\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1\right) &= \frac{\sigma_n}{\sigma'_n}\left\{\sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right) - \frac{1}{\gamma}\sqrt{k_n}\,(\widehat{\gamma}-\gamma)\,\frac{-\gamma\big(V(\infty)-u_n\big)}{\sigma_n}\right. \\
&\qquad\qquad \left. + \gamma\sqrt{k_n}\left(\frac{V(\infty)-u_n}{\sigma_n} + \frac{1}{\gamma}\right)\right\} \\
&\quad + \sqrt{k_n}\left(\frac{-\widehat{\gamma}\big(V(\infty)-u'_n\big)}{\sigma'_n} - 1\right) \\
&=: \frac{\sigma_n}{\sigma'_n}\,T_{1,n} + T_{2,n}\,.
\end{aligned}$$

Clearly $T_{1,n}$ tends in distribution to a normal distribution, but

$$\frac{\sigma_n}{\sigma'_n} \simeq \frac{V(\infty)-u_n}{V(\infty)-u'_n} = 1 + \frac{u'_n\big(1+o(1)\big)}{V(\infty)-u'_n} \to \infty\,.$$

Therefore $\frac{\sigma_n}{\sigma'_n}\,T_{1,n}$ tends to infinity, whereas $T_{2,n}$ can be rewritten as

$$T_{2,n} = \sqrt{k_n}(\gamma-\widehat{\gamma})\,\frac{V(\infty)-u'_n}{\sigma'_n} - \frac{\sqrt{k_n}}{\gamma+\rho}\,M\big(e^{V^{\leftarrow}(u_n)}\big)\,\frac{M\big(e^{V^{\leftarrow}(u'_n)}\big)}{M\big(e^{V^{\leftarrow}(u_n)}\big)}\,\big(1+o(1)\big)$$

by Lemma 1.2 (ii) in Worms (2000). This implies that $T_{2,n}$ tends to a normal distribution, since it is the case for the first part of this term, whereas the second one tends to 0.

Now, in case $\gamma = 0$, we can easily find a counter-example of our Lemma 3.1. First, note that

$$\begin{aligned}\sqrt{k_n}\left(\frac{\widehat{\widehat{\sigma}}'}{\sigma'_n} - 1\right) &= \sqrt{k_n}\left(\frac{\widehat{\sigma} + \widehat{\gamma}(u'_n - u_n)}{\sigma'_n} - 1\right) \\ &= \sqrt{k_n}\left(\frac{\widehat{\sigma}}{\sigma_n} - 1\right)\frac{\sigma_n}{\sigma'_n} + \sqrt{k_n}(\widehat{\gamma} - \gamma)\,\frac{u'_n - u_n}{\sigma'_n} + \sqrt{k_n}\,\frac{\sigma_n}{\sigma'_n} - \sqrt{k_n} \\ &=: \widetilde{Q}_{1,n} + \widetilde{Q}_{2,n} + \widetilde{Q}_{3,n} - \sqrt{k_n}\,.\end{aligned}$$

Now, if we consider an exponential random variable with parameter 1, then $\sigma'_n = \sigma_n = 1$ and $\frac{u'_n - u_n}{\sigma'_n} = u'_n\big(1 + o(1)\big)$ by assumption. This implies that $\widetilde{Q}_{1,n}$ is asymptotically normal, $\widetilde{Q}_{2,n} \xrightarrow{\mathbb{P}} \infty$ and $\widetilde{Q}_{3,n} - \sqrt{k_n} = 0$. Therefore Lemma 3.1 is not valid. □

ACKNOWLEDGMENTS

Part of this work was supported by the "Conseil Scientifique" of the University of Strasbourg, by the ANR AssimilEx project and by the FP7 ACQWA project. The authors are grateful to the referee for the constructive comments.

REFERENCES

[1] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

[2] Corcoran, J.N. and Schneider, U. (2003). Shift and scale coupling methods for perfect simulation, *Probability in the Engineering and Informational Sciences*, **17**, 277–303.

[3] Diebolt, J.; Guillou, A. and Rached, I. (2004). A new look at probability-weighted moments estimators, *Comptes Rendus de l'Académie des Sciences de Paris*, **338**, Série I, 629–634.

[4] Diebolt, J.; Guillou, A. and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of Statistical Planning and Inference*, **137**, 841–857.

[5] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or the smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

[6] Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C. and Wallis, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, **15**, 1049–1054.

[7] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*, New York, Springer-Verlag.

[8] Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.

[9] Smith, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.

[10] Worms, R. (2000). *Vitesses de convergence pour l'approximation des queues de distribution*, Thèse de doctorat de l'Université de Marne-la-Vallée.

[11] You, A.; Schneider, U.; Guillou, A. and Naveau, P. (2010). Improving extreme quantile estimation via a folding procedure, *Journal of Statistical Planning and Inference* (to appear).

REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 37–55

OPTIMAL ALARM SYSTEMS FOR FIAPARCH PROCESSES

Authors: Conceição Costa
– Departamento de Matemática, Universidade de Aveiro, Portugal
lopescosta@ua.pt

Manuel G. Scotto
– Departamento de Matemática, Universidade de Aveiro, Portugal
mscotto@ua.pt

Isabel Pereira
– Departamento de Matemática, Universidade de Aveiro, Portugal
isabel.pereira@ua.pt

Received: June 2009 Revised: March 2010 Accepted: March 2010

Abstract:

- In this work, an optimal alarm system is developed to predict whether a financial time series modeled via Fractionally Integrated Asymmetric Power ARCH (FIAPARCH) models, up/downcrosses some particular level and give an alarm whenever this crossing is predicted. The paper presents classical and Bayesian methodology for producing optimal alarm systems. Both methodologies are illustrated and their performance compared through a simulation study. The work finishes with an empirical application to a set of data concerning daily returns of the São Paulo Stock Market.

Key-Words:

- *FIAPARCH processes; optimal alarm systems; econometrics.*

AMS Subject Classification:

- 62M10, 91B70.

1. INTRODUCTION

Recently, it has been recognized the potential of optimal alarm systems in detecting and warning the occurrence of catastrophes or some other related events; see for example Monteiro *et al.* ([24]) and the references therein. Conceptually, the simplest way of constructing an alarm system is to predict X_{t+h} by a predictor say, $\hat{X}_{t+h,t}$, which is usually chosen so that the mean square error is minimized, providing

$$\hat{X}_{t+h,t} = E\Big[X_{t+h}|X_s,\, -\infty < s \leq t\Big]\,.$$

An alarm is given every time the predictor exceeds some critical level. This alarm system, however, does not have a good performance on the ability to detect the events, locate them accurately in time and give as few false alarms as possible. Lindgren ([18],[19],[20],[21]) and de Maré ([8]) set the principles for the construction of optimal alarm systems and obtain some basic results regarding the optimal prediction of level crossings. Svensson *et al.* ([27]) applied these principles in the prediction of level crossings in the sea levels of the Baltic sea. It is worth to mention that the alarm system introduced by Lindgren and de Maré, ignores the sampling variation of the model parameters. Giving heed to this issue, Amaral Turkman and Turkman ([1]) suggested a Bayesian approach and particular calculations were carried out for an autoregressive model of order one. Further extensions and generalizations were proposed by Antunes *et al.* ([2]) and more recently by Monteiro *et al.* ([24]).

The spectrum of applications of optimal alarm systems is wide and yet to be explored. One major area of applications is environmental economics. Atmospheric concentrations of air pollutants like ozone, carbon monoxide or sulfur dioxide constitute time series that can be analyzed under the perspective of the upcrossings of some critical levels, usually related with public health (e.g. Smith *et al.*, [26]; Koop and Tale, [17]; Tobias and Scotto, [30]). Another area of potential application of optimal alarm systems is econometrics and in particular in risk management. The implementation of probabilistic models for the assessment of market risks or credit risks is mandatory. Examples can be found in the forecasting of financial risk of lending to costumers (Thomas, [29]), the arrivals forecast of guests at hotels (Weatherford and Kimes, [32]) and in forecasting daily stock volatility, which has direct implications in option pricing, asset allocation or value-at-risk (Fuentes *et al.*, [14]). All the above referred references, however, are not directly applicable to calculate in advance the probability of future up/downcrossings. It is in this context that the implementation of an alarm system reveals to be useful. A related interesting problem, which has not been addressed yet, is to develop optimal alarm systems for financial time series. This article aims to give a contribution towards this direction.

The analysis of financial time series like log-return series of foreign exchange rates, stock indices or share prices, has revealed some common features: sample means not significantly different from zero, sample variances of the order 10^{-4} or smaller and sample distributions roughly symmetric in its center, sharply peaked around zero but with a tendency to negative asymmetry. In particular, it has usually been found that the conditional volatility of stocks responds asymmetrically to positive versus negative shocks: volatility tends to rise higher in response to negative shocks as opposed to positive shocks, which is known as the leverage effect. Another common feature of series of log-returns is that the sample autocorrelation function is negligible at all lags, (except perhaps for the first) but the sample autocorrelation functions for the absolute values or the squares of the log-returns are different from zero for a large number of lags and stay almost constant and positive for large lags. This last feature is known, in this context, as long memory or long range dependency. Several models have been proposed in order to describe these stylized facts about log-return series. We mention here the ARCH models, introduced by Engle ([11]) and some of the subsequent generalizations: GARCH, (Bollerslev, [4]), EGARCH (Dellaportas *et al.*, [9]), APARCH (Ding *et al.*, [10]), FIGARCH (Baillie *et al.*, [3]) and FIAPARCH (Tse, [31]). For a survey of ARCH-type models see Teräsvirta ([28]).

The rest of the paper is organized as follows: in Section 2, basic theoretical concepts related to optimal alarm systems are presented. Furthermore, an optimal alarm system for FIAPARCH processes is implemented. Expressions for some alarm characteristics of the alarm system are given. Estimation of the model FIAPARCH$(1, d, 1)$ by classical and Bayesian methodology is covered in Section 3. In Section 4, the results are illustrated through a simulation study. A real-data example is given in Section 5.

2. OPTIMAL ALARM SYSTEMS AND THEIR APPLICATION TO FIAPARCH PROCESSES

Let $\{X_t,\, t \in \mathbb{N}\}$ be a discrete time stochastic process. The time sequel $\{1, 2, ..., t-1, t, t+1, ...\}$ is divided into three sections, namely the data or informative experience, $D_t = \{X_1, ..., X_{t-r}\}$, the present experiment, $\boldsymbol{X}_2 = \{X_{t-r+1}, ..., X_t\}$ and the future experiment, $\boldsymbol{X}_3 = \{X_{t+1}, ...\}$. Any event of interest, say $C_{t,j}$, in the σ-field generated by $\boldsymbol{X}_3$ is defined as a catastrophe. Throughout this work a catastrophe will be considered as the upcrossing event

$$C_{t,j} = \left\{ X_{t+j-1} \leq u < X_{t+j} \right\},$$

for some $j \in \mathbb{N}$. Moreover, any event $A_{t,j}$ in the σ-field generated by $\boldsymbol{X}_2$, predictor of $C_{t,j}$, will be an alarm region. It is said that an alarm is given at time t, for the catastrophe $C_{t,j}$, if the observed value of $\boldsymbol{X}_2$ belongs to the alarm region.

In addition, the alarm is said to be correct if the event $A_{t,j}$ is followed by the event $C_{t,j}$, so, the probability of correct alarm is defined as the probability of catastrophe conditional on the alarm being given. Conversely, a false alarm is defined as the occurrence of $A_{t,j}$ without $C_{t,j}$. If an alarm is given when the catastrophe occurs, it is said that the catastrophe is detected and the probability of detection is defined as the probability of an alarm being given conditional on the occurrence of the catastrophe. Furthermore, the alarm region $A_{t,j}$ is said to have size $\alpha_{t,j}$ if $\alpha_{t,j} = P\big(A_{t,j}|D_t\big)$. The alarm region $A_{t,j}$ is optimal of size $\alpha_{t,j}$ if

$$P\big(A_{t,j}|C_{t,j}, D_t\big) \;=\; \sup_{B\in\sigma_{\boldsymbol{X}_2}} P\big(B|C_{t,j}, D_t\big) \tag{2.1}$$

with $P\big(B|D_t\big) = \alpha_{t,j}$.

Definition 2.1. An optimal alarm system of size $\{\alpha_{t,j}\}$ is a family of alarm regions $\{A_{t,j}\}$ in time, satisfying (2.1).

Lemma 2.1. *The alarm system $\{A_{t,j}\}$ with alarm region given by*

$$A_{t,j} \;=\; \Big\{\boldsymbol{x}_2 \in \mathbb{R}^r\colon\; P\big(C_{t,j}|\boldsymbol{x}_2, D_t\big) \geq k_{t,j}\, P\big(C_{t,j}|D_t\big)\Big\}\,,$$

for a fixed $k_{t,j}\colon P\big(\boldsymbol{X}_2 \in A_{t,j}|D_t\big) = \alpha_{t,j}$, is optimal of size $\alpha_{t,j}$.

This lemma ensures that the alarm region defined above renders the highest detection probability. Moreover to enhance the fact that the optimal alarm system depends on the choice of $k_{t,j}$, it is important to stress that in view of the fact that $P\big(C_{t,j}|D_t\big)$ does not depend on $\boldsymbol{x}_2$, the alarm region can be rewritten in the form

$$A_{t,j} \;=\; \Big\{\boldsymbol{x}_2 \in \mathbb{R}^r\colon\; P\big(C_{t,j}|\boldsymbol{x}_2, D_t\big) \geq k\Big\}\,, \tag{2.2}$$

where $k = k_{t,j}\, P\big(C_{t,j}|D_t\big)$ is chosen in some optimal way to accommodate conditions over the following operating characteristics of the alarm system:

- $P\big(A_{t,j}|D_t\big)$ – Alarm size,
- $P\big(C_{t,j}|A_{t,j}, D_t\big)$ – Probability of correct alarm,
- $P\big(A_{t,j}|C_{t,j}, D_t\big)$ – Probability of detecting the event.

Most models for financial time series used in practice are given in the multiplicative form

$$X_t = \sigma_t\, Z_t\;, \tag{2.3}$$

where $\{Z_t\}$ forms an i.i.d. sequence with zero mean and unit variance, $\{\sigma_t\}$ is a stochastic process such that σ_t and Z_t are independent for fixed t. Moreover, it is also assumed that Z_t is independent of the past values of the process

$(X_{t-1}, X_{t-2}, ...)$. In general, conditions ensuring the strict stationarity of the process $\{X_t\}$ are known. Motivation for considering this particular choice of simple multiplicative model comes from the fact that

(**a**) in practice, the direction of price changes is well modeled by the sign of Z_t, whereas σ_t provides a good description of the order of magnitude of this change;

(**b**) the volatility σ_t^2 represents the conditional variance of X_t given σ_t.

This representation expresses the belief that the direction of price changes can not be modeled, only their magnitude (e.g. Mikosch, [23]).

The FIAPARCH(p, d, q) model of Tse ([31]) is a special case of (2.3) with

$$\sigma_t^\delta = \frac{\omega}{1-\beta(B)} + \lambda(B)\, g(X_t)\,, \tag{2.4}$$

where $g(X_t) = (|X_t| - \gamma X_t)^\delta$ with $|\gamma| < 1$ and $\delta \geq 0$, and

$$\lambda(B) = 1 - \big(1-\beta(B)\big)^{-1}\phi(B)\,(1-B)^d = \sum_{i=1}^{\infty} \lambda_i B^i\,, \qquad \lambda(1) = 1\,, \tag{2.5}$$

for every $0 < d < 1$, with $\lambda_i \geq 0$, for $i \in \mathbb{N}$, and $\omega > 0$ for the conditional variance to be well defined, so that it is positive almost surely for all t. Furthermore, in order to allow for long memory the fractional differencing parameter d is constrained to lie in the interval $0 < d < 1/2$. Moreover, the polynomials $1-\beta(B)$ and $\phi(B)$ are assumed to have all their roots lying outside the unit circle. The fractional differencing operator $(1-B)^d$ is most conveniently expressed as

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k\,.$$

The FIAPARCH model nests two major classes of ARCH-type models: the APARCH and the FIGARCH models of Ding *et al.* ([10]) and Baillie *et al.* ([3]), respectively. When $d = 0$ the process reduces to the APARCH(p, q) model, whereas for $\gamma = 0$ and $\delta = 2$ the process reduces to the FIGARCH(p, d, q) model. The FIGARCH representation includes the GARCH (when $d = 0$) and the IGARCH (Engle and Bollerslev, [12]) when $d = 1$ with the implications in terms of impact of a shock on the forecasts of future conditional variances. Considering all the features involved in this specification, Conrad *et al.* ([7]) point out some advantages of the FIAPARCH(p, d, q) class of models, namely

(**a**) it allows for an asymmetric response of volatility to positive and negative shocks, so being able to traduce the leverage effect. If $\gamma > 0$, negative shocks have stronger impact on volatility than positive shocks, as would be expected in the analysis of financial time series. If $\gamma < 0$, the reverse happens;

(**b**) in this particular class of models it is the data that determines the power of returns for which the predictable structure in the volatility pattern is the strongest;

(**c**) the models are able to accommodate long memory in volatility, depending on the differencing parameter d.

It is important to mention here that necessary and sufficient condition for the existence of a stationary solution of the APARCH(p,q) model can be easily obtained from the results derived by Liu ([22]). This author introduced a family of GARCH processes, which can be regarded as a class of non-parametric GARCH processes, which includes as a special case the APARCH(p,q) model. Liu ([22]) obtained necessary and sufficient condition for the existence of a stationary solution of this new family of GARCH processes. Furthermore, Liu ([22]) also derived an explicit expression for the stationary solution. In contrast, however, the statistical properties of the general FIGARCH(p,d,q) process remain unestablished. Namely, stationarity is not a certainty as well as the source of long memory on volatility or even its existence are nowadays controversial. For the FIAPARCH process, Tse ([31]) also leaves these issues as open questions.

The simplest version of the FIAPARCH(p,d,q) model, which appears to be particularly useful in practice, is the FIAPARCH$(1,d,1)$ for which the volatility σ_t takes the form as in (2.4) with $\lambda(B)$ as in (2.5) with $\beta(B)=\beta B$ and $\phi(B)=\phi B$ with $|\beta|<1$. Necessary and sufficient conditions for the non-negativity of the conditional variance for the FIAPARCH$(1,d,1)$ resemble the ones obtained by Conrad and Haag ([6]) for the FIGARCH$(1,d,1)$, namely

- **Case I**: $0<\beta<1$, either $\lambda_1\geq 0$ and $\phi\leq h_2$ or for $i>2$ with $h_{i-1}<\phi\leq h_i$ it holds that $\lambda_{i-1}\geq 0$,
- **Case II**: $-1<\beta<0$, either $\lambda_1\geq 0$, $\lambda_2\geq 0$ and $\phi\leq h_2(\beta+h_3)/(\beta+h_2)$ or $\lambda_{i-1}\geq 0$, $\lambda_{i-2}\geq 0$ and $h_{i-2}(\beta+h_{i-1})/(\beta+h_{i-2})<\phi\leq h_{i-1}(\beta+h_i)/(\beta+h_{i-1})$ with $i>3$,

where $h_i=(i-1-d)/i$, for $i=2,3,....$ Furthermore, the infinite series coefficients can be obtained recursively as

$$\lambda_i=\begin{cases}\phi-\beta+d\,, & i=1\,,\\ \beta\lambda_{i-1}+[h_i-\phi]\,\delta_{i-1}\,, & i\geq 2\,,\end{cases}$$

with $\delta_1=d$ and $\delta_i=\delta_{i-1}h_i$ for $i\geq 2$.

The application of the alarm system to the FIAPARCH$(1,d,1)$ model will be done for the particular case $r=1$ and $j=2$ in Lemma 2.1. The event of interest (i.e. the catastrophe) is defined as the upcrossing of some fixed level u two steps ahead, that is

$$C_{t,2}=\left\{X_{t+1}\leq u<X_{t+2}\right\}. \tag{2.6}$$

The alarm region of optimal size $\alpha_{t,2}$ is given by

$$\begin{aligned} A_{t,2} &= \left\{ x_t \in \mathbb{R}\colon \frac{P\big(C_{t,2}|x_t, D_t\big)}{P\big(C_{t,2}|D_t\big)} \geq k_{t,2} \right\} \\ &= \left\{ x_t \in \mathbb{R}\colon P\big(C_{t,2}|x_t, D_t\big) \geq k \right\}, \end{aligned} \tag{2.7}$$

where $k = k_{t,2}\, P\big(C_{t,2}|D_t\big)$.

The first step in the construction of the alarm system consists on the calculation of the probability of catastrophe conditional on D_t and x_t, i.e. $P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big)$ and $P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big)$ with $\boldsymbol{\theta} = (\omega, \beta, \phi, \gamma, \delta, d)$. In doing so, note that

$$\begin{aligned} P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big) &= P\big(X_{t+1} \leq u < X_{t+2}|x_1, ..., x_t, \boldsymbol{\theta}\big) \\ &= \int_{C_{t,2}} f_{X_{t+1}, X_{t+2}|x_1, ..., x_t, \boldsymbol{\theta}}(x_{t+1}, x_{t+2})\, dx_{t+1}\, dx_{t+2} \end{aligned}$$

with the integration region, $C_{t,2}$, being the catastrophe region as in (2.6). If $Z_t \sim N(0,1)$ then

$$P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big) = \int_u^{+\infty} \int_{-\infty}^{u} \prod_{k=1}^{2} \frac{1}{\sqrt{2\,\pi}\,\sigma_{t+k}^2} \exp\left\{ -\frac{x_{t+k}^2}{2\sigma_{t+k}^2} \right\} dx_{t+1}\, dx_{t+2}\,. \tag{2.8}$$

Moreover

$$\begin{aligned} P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big) &= P\big(X_{t+1} \leq u < X_{t+2}|x_1, ..., x_{t-1}, \boldsymbol{\theta}\big) \\ &= \int_{C_{t,2}} \int f_{X_t, X_{t+1}, X_{t+2}|x_1, ..., x_{t-1}, \boldsymbol{\theta}}(x_t, x_{t+1}, x_{t+2})\, dx_t\, dx_{t+1}\, dx_{t+2}\,. \end{aligned}$$

Again, by assuming $Z_t \sim N(0,1)$ it follows that

$$P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big) = \int_u^{+\infty} \int_{-\infty}^{u} \int_{-\infty}^{+\infty} \prod_{k=0}^{2} \frac{1}{\sqrt{2\,\pi}\,\sigma_{t+k}^2} \exp\left\{ -\frac{x_{t+k}^2}{2\sigma_{t+k}^2} \right\} dx_t\, dx_{t+1}\, dx_{t+2}\,. \tag{2.9}$$

Having calculated these probabilities it is then possible to determine the alarm region and calculate the alarm characteristics of the alarm system.

1. Alarm size

$$\begin{aligned} \alpha_{t,2} &= P\big(A_{t,2}|D_t\big) \\ &= \int_{A_{t,2}} \frac{1}{\sqrt{2\,\pi}\,\sigma_t^2} \exp\left\{ -\frac{x_t^2}{2\sigma_t^2} \right\} dx_t \end{aligned}$$

with $A_{t,2}$ being the alarm region which depends on the value of $k_{t,2}$ chosen.

2. Probability of correct alarm

$$P\big(C_{t,2}|A_{t,2}, D_t\big) = \frac{P\big(C_{t,2} \cap A_{t,2}|D_t\big)}{P\big(A_{t,2}|D_t\big)}\,,$$

where

$$P\big(C_{t,2} \cap A_{t,2}|D_t\big) =$$
$$= P\Big(X_{t+1} \leq u < X_{t+2} \cap X_t \in A_{t,2}|D_t\Big)$$
$$= \int_u^{+\infty}\int_{-\infty}^{u}\int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi}\,\sigma_{t+k}^2} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t\, dx_{t+1}\, dx_{t+2}\,.$$

Thus

$$P\big(C_{t,2}|A_{t,2}, D_t\big) =$$
$$= \frac{\displaystyle\int_u^{+\infty}\int_{-\infty}^{u}\int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi}\,\sigma_{t+k}^2} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t\, dx_{t+1}\, dx_{t+2}}{\displaystyle\int_{A_{t,2}} \frac{1}{\sqrt{2\pi}\,\sigma_t^2} \exp\left\{-\frac{x_t^2}{2\sigma_t^2}\right\} dx_t}\,.$$

3. Probability of detecting the event

$$P\big(A_{t,2}|C_{t,2}, D_t\big) =$$
$$= \frac{P\big(A_{t,2} \cap C_{t,2}|D_t\big)}{P\big(C_{t,2}|D_t\big)}$$
$$= \frac{\displaystyle\int_u^{+\infty}\int_{-\infty}^{u}\int_{A_{t,2}} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi}\,\sigma_{t+k}^2} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t\, dx_{t+1}\, dx_{t+2}}{\displaystyle\int_u^{+\infty}\int_{-\infty}^{u}\int_{-\infty}^{+\infty} \prod_{k=0}^{2} \frac{1}{\sqrt{2\pi}\,\sigma_{t+k}^2} \exp\left\{-\frac{x_{t+k}^2}{2\sigma_{t+k}^2}\right\} dx_t\, dx_{t+1}\, dx_{t+2}}\,.$$

3. ESTIMATION PROCEDURES

In this section we consider the estimation of the operating characteristics of the alarm system. From the classical framework the method considered is the well-known Quasi-Maximum Likelihood Estimation procedure (QMLE) assuming conditional normality. The QMLE estimates are obtained maximizing the conditional log-likelihood function with respect to $\boldsymbol{\theta} = (\omega, \beta, \phi, \gamma, \delta, d)$, recurring to a routine available within the OxMetrics5 program. The robust standard errors by Bollerslev and Wooldrige ([5]) were also calculated. According to these authors this estimator is generally consistent, has a normal limiting distribution and provides asymptotic standard errors that are valid under non-normality. Nevertheless, the authors state that the QMLE estimator is not asymptotically efficient under non-normality and care should be taken, since as Engle and Gonzalez-Rivera ([13]) proved, GARCH estimates are consistent but asymptotically inefficient with the degree of inefficiency increasing with the degree of departure from

normality. The impact of violations in conditional normality, however, remains unknown for the FIGARCH and FIAPARCH case. Baillie *et al.* ([3]) suggested that the FIGARCH estimates obtained via QMLE are consistent and asymptotically normal[1]. Furthermore, they also demonstrated the suitability of the QMLE procedure in the estimation of samples with sizes of 1 500 and 3 000.

From the Bayesian perspective we need to start with a prior distribution for the vector of parameters $\boldsymbol{\theta}$. Assuming independence between all the parameters involved the prior distribution of $\boldsymbol{\theta}$, say $h(\boldsymbol{\theta})$, will be proportional to

$$h(\boldsymbol{\theta}) \propto I_{\{\omega>0\}}\, I_{\{-1<\beta<1\}}\, I_{\{\phi\geqslant 0\}}\, I_{\{-1<\gamma<1\}}\, I_{\{\delta\geqslant 0\}}\, I_{\{0<d<1/2\}}\,.$$

The posterior distribution $h\big(\boldsymbol{\theta}|D_t\big)$ is given by

$$\begin{aligned} h\big(\boldsymbol{\theta}|D_t\big) &\propto L\big(D_t|\boldsymbol{\theta}\big)\, h\big(\boldsymbol{\theta}\big) \\ &\propto \prod_{n=2}^{t-1} \frac{1}{\sqrt{2\,\pi}\,\sigma_n}\, \exp\!\left\{-\frac{x_n^2}{2\sigma_n^2}\right\} \\ &\quad \times I_{\{\omega>0\}}\, I_{\{-1<\beta<1\}}\, I_{\{\phi\geqslant 0\}}\, I_{\{-1<\gamma<1\}}\, I_{\{\delta\geqslant 0\}}\, I_{\{0<d<1/2\}}\,. \end{aligned}$$

Hence, the probability of catastrophe conditional on D_t and $\boldsymbol{x}_2 = \{x_t\}$, takes the form

$$P\big(C_{t,2}|x_t, D_t\big) = \int_{\Theta} P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big)\, h\big(\boldsymbol{\theta}|D_t\big)\, d\boldsymbol{\theta} \tag{3.1}$$

with Θ being the parameter space. On the other hand, the probability of catastrophe conditional on D_t, will be given by

$$P\big(C_{t,2}|D_t\big) = \int_{\Theta} P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big)\, h\big(\boldsymbol{\theta}|D_t\big)\, d\boldsymbol{\theta}\,, \tag{3.2}$$

where $P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big)$ and $P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big)$ are calculated through (2.8) and (2.9), respectively. However, due to the complexity of expressions (2.8) and (2.9) analytical calculations are not possible. Nonetheless, since by (3.1) and (3.2)

$$P\big(C_{t,2}|x_t, D_t\big) = E_{\boldsymbol{\theta}|D_t}\big[P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big)\big] \quad \text{and} \quad P\big(C_{t,2}|D_t\big) = E_{\boldsymbol{\theta}|D_t}\big[P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big)\big]\,,$$

their respective Monte Carlo approximations can be used, that is

$$\widehat{P}\big(C_{t,2}|x_t, D_t\big) = \frac{1}{m}\sum_{i=1}^{m} P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}_i\big) \quad \text{and} \quad \widehat{P}\big(C_{t,2}|D_t\big) = \frac{1}{m}\sum_{i=1}^{m} P\big(C_{t,2}|D_t, \boldsymbol{\theta}_i\big)\,,$$

where the observations $\boldsymbol{\theta}_i = (\omega_i, \beta_i, \phi_i, \gamma_i, \delta_i, d_i)$ with $i = 1, 2, ..., m$ constitute a sample of the posterior distribution $h\big(\boldsymbol{\theta}|D_t\big)$. A similar procedure is applied to approximate the operating characteristics.

[1] In fact, the consistency and asymptotic normality of the QMLE estimator had been formally established for the IGARCH(1,1) process. Baillie *et al.* ([3]) followed a dominance-type argument to extend this result to the FIGARCH$(1, d, 0)$ case and refer the need for a formal proof of consistency and asymptotic normality for the general IGARCH(p, q) and FIAGARCH(p, d, q) cases.

4. SIMULATION RESULTS

In this section we present a simulation study to illustrate the performance of the alarm system constructed for the FIAPARCH$(1, d, 1)$ model. In particular we consider the set of parameters $\boldsymbol{\theta} = (0.40, 0.28, 0.10, 0.68, 1.27, 0.30)$. The choice of the parameters is very similar to those appearing in the real-data example presented in Section 5. Figure 1 below shows a simulated sample path for this specific FIAPARCH model.

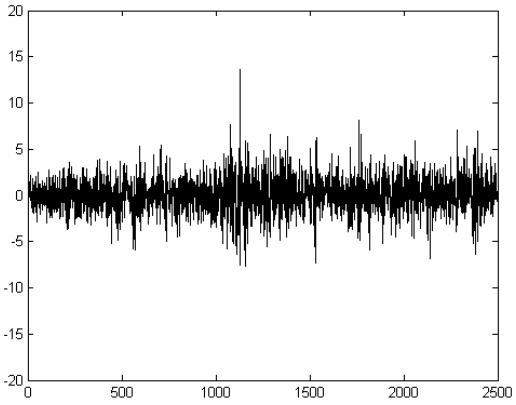

Figure 1: Simulated sample path of a FIAPARCH$(1, d, 1)$ process with $\boldsymbol{\theta} = (0.40, 0.28, 0.10, 0.68, 1.27, 0.30)$.

Parameter estimates, $\hat{\boldsymbol{\theta}}$, and their corresponding standard errors were obtained for this sample, following the QMLE procedure of Bollerslev and Wooldrige ([5]). Robust standard errors are estimated from the product $A(\hat{\boldsymbol{\theta}})^{-1}B(\hat{\boldsymbol{\theta}})A(\hat{\boldsymbol{\theta}})^{-1}$, where $A(\hat{\boldsymbol{\theta}})$ and $B(\hat{\boldsymbol{\theta}})$ denote the Hessian and the outer product of the gradients evaluated at $\hat{\boldsymbol{\theta}}$, respectively.

Moreover, Bayesian estimates of $\boldsymbol{\theta}$ were also obtained for this single sample. Since the standard Gibbs methodology is difficult to implement to FIAPARCH models partially due to the non-standard forms of the full conditional densities, the Metropolis-Hastings algorithm was implemented in the software Matlab. In addition, a multivariate t-distribution was used as the proponent one. The sampler algorithm ran 100 000 iterations including a *burn-in* period of 40 000 observations which are discarded for the posterior analysis, as suggested by Dellaportas *et al.* ([9]). Furthermore, only every twentieth iteration is stored in order to obtain an, approximately, independent and identically distributed sample. The estimates were taken as the means of the posterior distribution. The convergence of the Markov chain was analyzed through the R criterium of Gelman and Rubin ([15]), the Z-score test of Geweke ([18]) and by graphical methods.

The analysis of the alarm system is carried out at $t = 2000$, i.e., $\boldsymbol{x}_2 = \{x_{2000}\}$. The event of interest is the two step ahead catastrophe defined by the upcrossing of the fixed level u, at time $t + 2$:

$$C_{2000,2} = \left\{(x_{2001}, x_{2002}) \in \mathbb{R}^2 \colon\ x_{2001} \leq u < x_{2002}\right\}.$$

In a first stage, two values of u were chosen, accordingly to the sample quantiles, namely the 90^{th} percentile ($Q_{0.90}$), and the 95^{th} percentile ($Q_{0.95}$). The choice of these values is justified by the fact that we are interested in relatively rare events. For both fixed levels of u, the probabilities $P\big(C_{t,2}|x_t, D_t, \boldsymbol{\theta}\big)$ and $P\big(C_{t,2}|D_t, \boldsymbol{\theta}\big)$ were numerically approximated as described in the previous section. In order to compute the optimal alarm region for each case, one has to obtain the region for several values of k, accordingly to expression (2.7) and then, for each value of k, compute the operating characteristics of the alarm system, i.e., the size of the region, $\alpha_{t,2}$, the probability of correct alarm, $P\big(C_{t,2}|A_{t,2}, D_t\big)$ and the probability of detection, $P\big(A_{t,2}|C_{t,2}, D_t\big)$. For every fixed value of k the region has to be obtained through a systematic search in a three dimensional region for (x_t, x_{t+1}, x_{t+2}). We considered a thin grid of values of x_t in $[-100, 100]$ and determined, for each value of x_t, whether $P\big(C_{t,2}|x_t, D_t\big)$ exceeds k. This procedure is repeated for k ranging from $P\big(C_{t,2}|D_t\big)$ to $P\big(C_{t,2}|D_t\big) + n\times 0.005$, with $n \in \mathbb{R}^+$. This procedure is repeated for both the classical (using the true values of the parameters and their QMLE estimates) and the Bayesian approach. The results are shown in Table 1 below.

Considering the true values of the parameters, the probability of the alarm being correct, does not exceed 5.6% in the $u = Q_{0.95}$ case, or 9.7% in the $u = Q_{0.90}$ case. The probability of detection for this sample, ranges from 2.4% to 49.0% for $u = Q_{0.95}$, or from 1.7% to 53.4% for $u = Q_{0.90}$. The results obtained with the QMLE estimates do not differ considerably, in particular in what concerns the probability of correct alarm. Regarding the probability of detecting the event, we can say the alarm system behaves better in this case since the detection probability reaches 54.5% for $u = Q_{0.95}$ and 60.6% for $u = Q_{0.90}$. Considering now the Bayesian approach, the probability of detection is the lowest obtained. It does not even reach 22%. On the other hand, the estimation procedure involved in the Bayesian approach seems to be able to produce higher probabilities of correct alarm, depending on an accurate choice of k. The probability of correct alarm ranges from lower values than in the classical approach to more than the double of these values, with increasing k, reaching 24.7% in the $u = Q_{0.90}$ case. Furthermore, note that as the probability of correct alarm increases, the probability of detecting the event decreases, as expected. This can be justified by the fact that as k increases, the size of the alarm region decreases, which implies that the number of alarms should decrease, so as the probability of detection, $P\big(A_{t,2}|C_{t,2}\big)$. However, as the number of alarms decreases, the probability of false alarms also decreases and therefore the probability of the alarm being correct, $P\big(C_{t,2}|A_{t,2}\big)$, increases.

Table 1: Operating characteristics at time point $t = 2000$.

$u = Q_{0.95} = 3.136$									
k	True Parameters			QML Estimates			Bayesian Estimates		
	$P(C_{t,2}) = 0.0746$			$P(C_{t,2}) = 0.0893$			$P(C_{t,2}) = 0.0817$		
	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0350	0.4789	0.0335	0.4903	0.5353	0.0346	0.5446	0.1904	0.0267	0.2155
0.0400	0.2998	0.0345	0.3155	0.3255	0.0355	0.3400	0.1902	0.0257	0.2074
0.0450	0.2072	0.0349	0.2209	0.2971	0.0359	0.3133	0.1211	0.0264	0.1354
0.0500	0.2067	0.0344	0.2173	0.2102	0.0363	0.2247	0.0718	0.0283	0.0862
0.0600	0.1377	0.0347	0.1458	0.1413	0.0360	0.1496	0.0397	0.0318	0.0535
0.0700	0.0864	0.0363	0.0957	0.0896	0.0373	0.0983	0.0203	0.0391	0.0337
0.0800	0.0509	0.0390	0.0605	0.0535	0.0398	0.0625	0.0097	0.0555	0.0227
0.0900	0.0282	0.0439	0.0377	0.0300	0.0454	0.0401	0.0042	0.0982	0.0177
0.1000	0.0146	0.0558	0.0248	0.0158	0.0563	0.0262	0.0017	0.2061	0.0151

$u = Q_{0.90} = 2.293$									
k	True Parameters			QML Estimates			Bayesian Estimates		
	$P(C_{t,2}) = 0.1267$			$P(C_{t,2}) = 0.1439$			$P(C_{t,2}) = 0.1353$		
	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
0.0850	0.5303	0.0832	0.5339	0.6042	0.0846	0.6055	0.1904	0.0722	0.1984
0.0900	0.3209	0.0837	0.3250	0.3490	0.0853	0.3528	0.1902	0.0719	0.1974
0.0950	0.2960	0.0844	0.3021	0.3033	0.0849	0.3050	0.1211	0.0717	0.1252
0.1000	0.2069	0.0843	0.2109	0.2117	0.0864	0.2167	0.1211	0.0713	0.1245
0.1100	0.1377	0.0852	0.1420	0.2101	0.0859	0.2137	0.0718	0.0730	0.0757
0.1200	0.0864	0.0862	0.0901	0.1413	0.0864	0.1446	0.0397	0.0773	0.0442
0.1300	0.0509	0.0887	0.0546	0.0535	0.0905	0.0573	0.0203	0.0825	0.0242
0.1400	0.0282	0.0888	0.0302	0.0535	0.0904	0.0572	0.0042	0.1123	0.0069
0.1500	0.0146	0.0965	0.0170	0.0158	0.1054	0.0197	0.0017	0.2474	0.0062

As already discussed, it is not possible, in general, to maximize both probabilities, $P\big(C_{t,2}|A_{t,2}\big)$ and $P\big(A_{t,2}|C_{t,2}\big)$, simultaneously. Hence, a compromise should be reached by the proper choice of k. In doing so, several criteria have been already proposed. Svensson *et al.* ([27]), for example, suggested that k should be chosen so that the probability of correct alarm and the probability of detecting the event are approximately equal, $P\big(C_{t,2}|A_{t,2}\big) \simeq P\big(A_{t,2}|C_{t,2}\big)$. On the other hand, Antunes *et al.* ([2]) suggested that k should be chosen so that the alarm size is about twice the probability of having a catastrophe given the past values of the process, $P\big(C_{t,2}|D_t\big) \simeq \frac{1}{2}\, P\big(A_{t,2}|D_t\big)$, stating that in this situation the system will be spending twice the time in the alarm state than in the catastrophe region. We analyzed both criteria in this work and from hereafter, the former criterion will be designated by Criterion 2 and the last by Criterion 1.

In order to test the alarm system, three extra values of the series were simulated: $(\boldsymbol{x}_2, \boldsymbol{x}_3) = (x_t, x_{t+1}, x_{t+2})$. This procedure was repeated 10 000 times with the same informative experience, D_t. With the alarm regions calculated before for $u = Q_{0.90} = 2.293$ and for the two criteria already mentioned, we observed, for each of the 10 000 samples, whether an alarm was given or not and whether a catastrophe occurred or not. Results are given in Table 2.

Table 2: Results at time point $t = 2000$. Percentages in parenthesis.

Approach	Criterion	Alarms		Catastrophes	
		False	Total	Detected	Total
True Parameters	1	1112 (0.8330)	1335	223 (0.2059)	1083
	2	651 (0.8314)	783	132 (0.1273)	1037
QMLE Approach	1	1163 (0.8526)	1364	201 (0.1963)	1024
	2	380 (0.8260)	460	80 (0.0771)	1037
Bayesian Approach	1	1161 (0.8401)	1382	221 (0.2103)	1051
	2	668 (0.8477)	788	120 (0.1204)	997

Finally, we illustrate how the online prediction performs in practice. The event to predict is

$$C_{t,2} = \Big\{(x_{t+1}, x_{t+2}) \in \mathbb{R}^2 \colon\ x_{t+1} \leq u < x_{t+2}\Big\},$$

for $t = 2000, ..., 2010$, again with $u = Q_{0.90} = 2.293$. Alarm regions and respective operating characteristics are presented in Table 3 for Criterion 1 and in Table 4 for Criterion 2.

Overall, Criterion 1 provides better estimates for the operating characteristics. The probability of detection, for instance, reaches values around 0.22 in some cases for the classical approach whereas with Criterion 2 this probability is nearly only half the former.

Table 3: Operating characteristics at different time points with Criterion 1.

Approach	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
True Parameters	2000	0.0827	0.1100	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1377	0.0852	0.1420
	2001	0.1047	0.1047	$[-\infty, -1.5] \cup [5.5, +\infty]$	0.1848	0.1093	0.1929
	2002	0.0936	0.0936	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.1209	0.0980	0.1265
	2003	0.0923	0.1073	$[-\infty, -1.5] \cup [7.5, +\infty]$	0.2167	0.0947	0.2224
	2004	0.0897	0.0977	$[-\infty, -1.5] \cup [8.0, +\infty]$	0.2076	0.0914	0.2116
	2005	0.0879	0.0979	$[-\infty, -1.5] \cup [7.5, +\infty]$	0.2036	0.0893	0.2069
	2006	0.0803	0.0953	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1311	0.0831	0.1356
	2007	0.0687	0.0887	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1286	0.0716	0.1340
	2008	0.0573	0.0873	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.1194	0.0614	0.1279
	2009	0.0508	0.0758	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1045	0.0522	0.1075
	2010	0.0545	0.0845	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.0924	0.0566	0.0960
QMLE	2000	0.0844	0.1200	$[-\infty, -2.0] \cup [10.5, +\infty]$	0.1413	0.0864	0.1446
	2001	0.1097	0.1047	$[-\infty, -1.5] \cup [6.0, +\infty]$	0.1867	0.1123	0.2002
	2002	0.0969	0.0969	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.1230	0.1005	0.1276
	2003	0.0946	0.1096	$[-\infty, -1.5] \cup [7.5, +\infty]$	0.2202	0.0972	0.2262
	2004	0.0919	0.1019	$[-\infty, -1.5] \cup [7.5, +\infty]$	0.2110	0.0943	0.2165
	2005	0.0900	0.1000	$[-\infty, -1.5] \cup [7.5, +\infty]$	0.2066	0.0917	0.2104
	2006	0.0821	0.0971	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1340	0.0843	0.1376
	2007	0.0697	0.0897	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1314	0.0723	0.1363
	2008	0.0594	0.0894	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1217	0.0619	0.1269
	2009	0.0506	0.0756	$[-\infty, -2.0] \cup [8.0, +\infty]$	0.1059	0.0528	0.1104
	2010	0.0544	0.0844	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.0930	0.0566	0.0966
Bayesian	2000	0.0693	0.0950	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1211	0.0717	0.1252
	2001	0.0911	0.0911	$[-\infty, -1.5] \cup [6.0, +\infty]$	0.1685	0.0939	0.1736
	2002	0.0820	0.0820	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.1047	0.0845	0.1078
	2003	0.0794	0.0994	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1297	0.0820	0.1340
	2004	0.0764	0.0914	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1218	0.0797	0.1271
	2005	0.0715	0.0915	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1176	0.0779	0.1282
	2006	0.0680	0.0830	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1144	0.0711	0.1196
	2007	0.0576	0.0776	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1121	0.0598	0.1165
	2008	0.0498	0.0748	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.1038	0.0513	0.1068
	2009	0.0419	0.0669	$[-\infty, -2.0] \cup [9.0, +\infty]$	0.0902	0.0441	0.0948
	2010	0.0447	0.0747	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.0790	0.0467	0.0825

Table 4: Operating characteristics at different time points with Criterion 2.

Approach	t	$P(C_{t,2}\|D_t)$	k	Alarm Region	α_2	$P(C_{t,2}\|A_{t,2})$	$P(A_{t,2}\|C_{t,2})$
True Parameters	2000	0.0827	0.1200	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0864	0.0862	0.0901
	2001	0.1047	0.1247	$[-\infty, -2.0] \cup [10.5, +\infty]$	0.1153	0.1088	0.1198
	2002	0.0936	0.1036	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0717	0.1001	0.0767
	2003	0.0923	0.1223	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0958	0.0949	0.0985
	2004	0.0897	0.1147	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0872	0.0924	0.0899
	2005	0.0879	0.1129	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0835	0.0906	0.0862
	2006	0.0803	0.1053	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0805	0.0831	0.0832
	2007	0.0687	0.0987	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0783	0.0726	0.0827
	2008	0.0573	0.1023	$[-\infty, -2.5] \cup [13.0, +\infty]$	0.0705	0.0630	0.0774
	2009	0.0508	0.0908	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0582	0.0531	0.0608
	2010	0.0545	0.0945	$[-\infty, -2.5] \cup [11.0, +\infty]$	0.0487	0.0593	0.0530
QMLE	2000	0.0844	0.1300	$[-\infty, -3.0] \cup [13.5, +\infty]$	0.0535	0.0905	0.0573
	2001	0.1047	0.1297	$[-\infty, -2.0] \cup [10.5, +\infty]$	0.1174	0.1104	0.1238
	2002	0.0969	0.1069	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0735	0.1027	0.0780
	2003	0.0946	0.1246	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0992	0.0974	0.1021
	2004	0.0919	0.1169	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0904	0.0947	0.0932
	2005	0.0900	0.1150	$[-\infty, -2.5] \cup [11.0, +\infty]$	0.0863	0.0929	0.0891
	2006	0.0821	0.1121	$[-\infty, -2.5] \cup [12.5, +\infty]$	0.0831	0.0850	0.0860
	2007	0.0697	0.0997	$[-\infty, -2.5] \cup [11.0, +\infty]$	0.0808	0.0731	0.0847
	2008	0.0594	0.0994	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0723	0.0637	0.0776
	2009	0.0506	0.0956	$[-\infty, -2.5] \cup [13.0, +\infty]$	0.0593	0.0529	0.0619
	2010	0.0544	0.0994	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0491	0.0590	0.0533
Bayesian	2000	0.0693	0.1100	$[-\infty, -2.5] \cup [12.5, +\infty]$	0.0718	0.0730	0.0757
	2001	0.0911	0.1011	$[-\infty, -2.0] \cup [8.5, +\infty]$	0.1002	0.0943	0.1037
	2002	0.0820	0.0820	$[-\infty, -2.0] \cup [9.5, +\infty]$	0.1047	0.0845	0.1078
	2003	0.0794	0.1094	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0793	0.0835	0.0835
	2004	0.0764	0.1014	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0724	0.0813	0.0771
	2005	0.0715	0.1065	$[-\infty, -2.5] \cup [13.5, +\infty]$	0.0689	0.0794	0.0766
	2006	0.0680	0.0930	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0663	0.0726	0.0707
	2007	0.0576	0.0876	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0643	0.0619	0.0692
	2008	0.0498	0.0848	$[-\infty, -2.5] \cup [12.0, +\infty]$	0.0576	0.0536	0.0619
	2009	0.0419	0.0769	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0470	0.0461	0.0517
	2010	0.0447	0.0847	$[-\infty, -2.5] \cup [11.5, +\infty]$	0.0388	0.0476	0.0413

5. EXPLORING THE IBOVESPA RETURNS DATA SET

In this section, we model the data set IBOVESPA which contains daily returns of the S. Paulo Stock Market during the period 04/07/1994 to 02/10/2008 (`www.ipeadata.gov.br`). Data consists on the closing rates of stocks, I_t, being the log-returns calculated as $y_t = \ln(I_t/I_{t-1})$, $t = 1, ..., n$. The results obtained from this procedure were then multiplied by 100 just to ensure the stability of posterior calculations. Sáfadi and Pereira ([25]) proved that the FIAPARCH$(1, d, 1)$ provides a good fit for this kind of data sets. To fit a FIAPARCH$(1, d, 1)$ model for the log-returns we proceeded as follows: first, the AR(10) model $y_t = 0.0689 + 0.0645\, y_{t-10} + x_t$, is fitted, using the least squares method, in order to eliminate serial dependence. The time series plot of both the IBOVESPA daily returns and the residuals (x_t), hereafter designated by x-returns, are exhibited in Figure 2 below. This is, indeed, the set of data reported to show the common features of financial time series mentioned in Section 1, that is weak dependence without any evident pattern on the series level and significative dependence on squared and absolute returns.

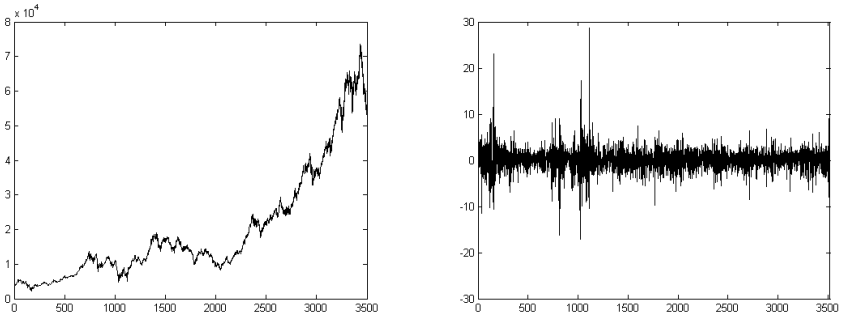

Figure 2: Plot of the IBOVESPA daily returns (left) and the x-returns (right) from 04/07/1994 to 02/10/2008.

The FIAPARCH$(1, d, 1)$ model was fitted to the series of x-returns by means of the QMLE procedure and the Bayesian approach described in Section 3. In both cases the adequacy of the fit was checked through the analysis of the standardized residuals. Table 5 presents the estimates obtained for both procedures.

Table 5: Parameter estimates. Standard deviations in parenthesis.

	QMLE	Bayesian Estimates
ω	0.3903 (0.1092)	0.4227 (0.0576)
ϕ	0.0957 (0.1334)	0.1289 (0.0397)
γ	0.6782 (0.1363)	0.7813 (0.1108)
β	0.2794 (0.1693)	0.3246 (0.0568)
δ	1.2744 (0.1274)	1.2218 (0.1008)
d	0.2952 (0.0642)	0.3020 (0.0258)

Since the IBOVESPA x-returns are related to the daily changes of the stock indexes of S. Paulo Stock Market, we considered that the event of interest is given by

$$C_{t,2} = \left\{(x_{t+1}, x_{t+2}) \in \mathbb{R}^2 \colon\ x_{t+1} \geq u > x_{t+2}\right\},$$

with $t = 3450, ..., 3516$, corresponding to July, August and September of 2008, and $u = Q_{0.25} = -1.219$. Note that, the downcrossing event $C_{t,2}$ can be view as related with a stock market crash. Moreover, the choice of k was done according only to Criterion 1: $P\big(C_{t,2}|D_t\big) \simeq \frac{1}{2}\, P\big(A_{t,2}|D_t\big)$. Two reasons justify this choice. First, Criterion 2 is difficult to implement since $P\big(C_{t,2}|A_{t,2}, D_t\big)$ may never get so close to $P\big(A_{t,2}|C_{t,2}, D_t\big)$ or when it does, some operating characteristics may show not so good results (at least as compared with those obtained with Criterion 1). Secondly, Criterion 1 results in better estimates of the operating characteristics. For the time period considered, the total number of alarms, the total number of catastrophes, the number of false alarms and the number of detected events was counted. Results are presented in Table 6. A closer look to Table 6 reveals that the estimate of the probability of the alarm being correct is 50% in July and August and raises to 100% in September. In addition, the estimate of the probability of detecting a catastrophe remains around 20% during the time period considered. We noticed that this online prediction system exhibits an adaptive behavior, that is, as long as the available information is integrated within the informative experience, the system adapts itself in order to produce the minimum number of false alarms. This fact explains on one hand the high estimate of the probabilities of the alarm given being correct and on the other hand that the system produces few alarms, so the probability of detection can not be very high.

Table 6: Results of the alarm system with $u = -1.219$. Percentages in parenthesis.

Month	Alarms		Catastrophes	
	False	Total	Detected	Total
July	1 (0.50)	2	1 (0.16)	6
August	1 (0.50)	2	1 (0.20)	5
September	0 (0.00)	3	3 (0.27)	11
Trimester	2 (0.28)	7	5 (0.22)	22

ACKNOWLEDGMENTS

The authors would like to express their gratitude to an anonymous referee for all helpful comments and constructive criticism.

REFERENCES

[1] Amaral-Turkman, M.A. and Turkman, K.F. (1990). Optimal alarm systems for autoregressive processes; a Bayesian approach, *Comput. Statist. Data Anal.*, **10**, 307–314.

[2] Antunes, M.; Amaral-Turkman, M.A. and Turkman, K.F. (2003). A Bayesian approach to event prediction intervals, *J. Time Ser. Anal.*, **24**, 631–646.

[3] Baillie, R.T.; Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, *J. Econom.*, **74**, 3–30.

[4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econom.*, **31**, 307–327.

[5] Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Rev.*, **11**, 143–172.

[6] Conrad, C. and Haag, B.R. (2006). Inequality constraints in the fractionally integrated GARCH model, *J. Finan. Econometrics*, **4**, 413–449.

[7] Conrad, C.; Karanasos, M. and Zeng, N. (2008). Multivariate fractionally integrated APARCH modeling of stock market volatility: a multi-country study. (Submitted).

[8] de Maré, J. (1980). Optimal prediction of catastrophes with application to Gaussian process, *Ann. Probab.*, **8**, 841–850.

[9] Dellaportas, P.; Politis, D.N. and Vrontos, I.D. (2000). Full Bayesian inference for GARCH and EGARCH models, *J. Bus. Econom. Statist.*, **18**, 187–196.

[10] Ding, Z.; Engle, R.F. and Granger, C.W.J. (1993). A long memory property of stock market returns and a new model, *J. Empir. Finance*, **1**, 83–106.

[11] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the United Kingdom inflation, *Econometrica*, **50**, 987–1008.

[12] Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances, *Econometric Rev.*, **5**, 1–50.

[13] Engle, R.F. and Gonzalez-Rivera, G. (1991). Semi-parametric ARCH model, *J. Bus. Econom. Statist.*, **9**, 345–360.

[14] Fuentes, A-M.; Izzeldin, I. and Kalotychou, E. (2009). On forecasting daily stock volatility: The role of intraday information and market conditions, *Int. J. Forecast.*, **25**, 259–281.

[15] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, *Statist. Sci.*, **7**, 457–511.

[16] Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.* In "Bayesian Statistics 4" (J.M. Bernardo, J.O. Berger, A.P. Dawin and A.F.M. Smith, Eds.), Oxford, University Press, 169–193, (with discussion).

[17] Koop, G. and Tale, L. (2004). Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *J. Environ. Econ. Manage.*, **47**, 30–54.

[18] Lindgren, G. (1975a). Prediction for a random point time, *Ann. Probab.*, **3**, 412–423.

[19] Lindgren, G. (1975b). Prediction of catastrophes and high level crossings, *Bull. Internat. Statist. Inst.*, **46**, 225–240.

[20] Lindgren, G. (1980). Model processes in non-linear prediction, with application to detection and alarm, *Ann. Probab.*, **8**, 775–792.

[21] Lindgren, G. (1985). Optimal prediction of level crossings in Gaussian processes and sequences, *Ann. Probab.*, **13**, 804–824.

[22] Liu, J.-C. (2009). Stationarity of a family of GARCH processes, *Econometric J.*, **12**, 436–446.

[23] Mikosch, T. (2003). *Modeling dependence and tails of financial time series.* In "Extreme Values in Finance, Telecommunications, and the Environment" (B. Finkenstadt and H. Rootzén, Eds.), Boca Raton, Chapman & Hall, 185–286.

[24] Monteiro, M.; Pereira, I. and Scotto, M.G. (2008). Optimal alarm systems for count processes, *Commun. Statist. Theory and Methods*, **37**, 3054–3076.

[25] Sáfadi, T. and Pereira, I. (2010). Bayesian analysis of FIAPARCH model: an application to São Paulo stock market, *Int. J. Stat. Econ.*, **5**, 49–63

[26] Smith, R.L.; Davis, J.M.; Sacks, J.; Speckman, P. and Styer, P. (2000). Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama, *Environmetrics*, **11**, 719–743.

[27] Svensson, A.; Lindquist, R. and Lindgren, G. (1996). Optimal prediction of catastrophes in autoregressive moving average processes, *J. Time Ser. Anal.*, **17**, 511–531.

[28] Teräsvirta, T. (2009). *An introduction to univariate GARCH models.* In "Handbook of Financial Time Series" (T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch, Eds.), New York, Springer, 17–42.

[29] Thomas, L.C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, *Int. J. Forecast.*, **16**, 149–172.

[30] Tobias, A. and Scotto, M.G. (2005). Prediction of extreme ozone levels in Barcelona, Spain, *Environ. Monit. Assess.*, **100**, 23–32.

[31] Tse, Y. (1998). The conditional heteroskedasticity of the Yen–Dollar exchange rate, *J. Appl. Econom.*, **13**, 49–55.

[32] Weatherford, L.R. and Kimes, S.E. (2003). A comparison of forecasting methods for hotel revenue management, *Int. J. Forecast.*, **19**, 401–415.

REVSTAT – Statistical Journal
Volume 8, Number 1, June 2010, 57–80

AN OVERVIEW OF LINEAR STRUCTURAL MODELS IN ERRORS IN VARIABLES REGRESSION

Author: Jonathan Gillard
– Cardiff School of Mathematics, Cardiff University,
Senghennydd Road, Cardiff, UK
GillardJW@Cardiff.ac.uk

Received: May 2009 Revised: May 2010 Accepted: May 2010

Abstract:

- This paper aims to overview the numerous approaches that have been developed to estimate the parameters of the linear structural model. The linear structural model is an example of an errors in variables model, or measurement error model that has wide practical use. This paper brings together key concepts from a scattered literature to give an accessible account of existing work on this particular errors in variables model.

Key-Words:

- *errors in variables; regression; measurement error; linear structural model.*

AMS Subject Classification:

- 62-02, 62J05.

1. INTRODUCTION

1.1. Aim of errors in variables modelling

Given a set of variables, a common statistical procedure is to try and find relationships between them. A technique that may aid with this is regression, which can provide an estimate of the formulaic relationship between these variables. The relationships between these variables, if they exist, may be linear or non-linear. Commonly the variables are split into dependent and independent variables, and regression analyses are concerned with writing the dependent variables in terms of some function of the independent variables. Standard regression procedures assume that the independent variables are measured without error, and that the error inherent in the model is associated with the dependent variables only. The theory of fitting such models is plentiful, and is well documented in the literature. An obvious extension to this model is to assume that there is error also present in the independent variables. This has become known as the errors in variables situation. There are errors in the measurement of both the independent and dependent variables, and so usually a different tack is called for. Indeed, in [9] Casella and Berger wrote that the errors in variables model "is so fundamentally different from the simple linear regression (...) that it is probably best thought of as a different topic."

1.2. Common applications

Errors in variables models have been successfully applied to a number of different subject areas over the years. Indeed, different ways of solving the problem of having errors in variables have become associated with different subject areas. For example, the method of using instrumental variables had its origins in the economic literature, but this technique is not restricted to economic applications. Use of errors in variables methodology has proved fruitful in areas as diverse as astrostatistics, as a method to cope with astronomical measurement error, in fisheries statistics, as a way of looking at fish stocks, in medical statistics, commonly in method comparison studies when a number of methods of measurement are compared and much more. The errors in variables problem is also one that is inherently interesting theoretically, and a number of theoretical developments in their own right have been made. In particular, numerical analysts have been interested in the development and study of particular types of numerical algorithms to solve the errors in variables problem.

Errors in variables models tend to be appropriate when all variables are experimentally observed. Each variable is then subject to its own inherent measurement error. Despite their common application, errors in variables methodology is still quite neglected in practice. This could be for a number of reasons. Firstly, the literature on errors in variables topics is widely scattered, appearing in a range of journals, in a number of different contexts. Secondly, the notation used for errors in variables models varies tremendously. Thus it is sometimes difficult to read papers from different sources. Finally, there are a number of different approaches to fit an errors in variables model. Some of these will be described in this paper. The aim of this paper is to bring ideas from this widely scattered literature together, and to explain the development of key methodologies and links between them.

For brevity, this paper will focus on the linear structural model which is a commonly fitted errors in variables type model. Section 2 describes the linear structural model. Section 3 outlines the main approaches that have been adopted to estimate the parameters of the linear structural model.

2. AN INTRODUCTION TO THE LINEAR STRUCTURAL MODEL

Consider two variables, ξ and η which are linearly related in the form

$$\eta_i = \alpha + \beta\,\xi_i\ , \qquad i = 1, ..., n\ .$$

However, instead of observing ξ_i and η_i, we observe

$$\begin{aligned} x_i &= \xi_i + \delta_i\ , \\ y_i &= \eta_i + \varepsilon_i\ =\ \alpha + \beta\,\xi_i + \varepsilon_i\ , \end{aligned}$$

where δ_i and ε_i are considered to be random error components, or noise.

It is assumed that $E[\delta_i] = E[\varepsilon_i] = 0$ and that $\mathrm{Var}[\delta_i] = \sigma_\delta^2$, $\mathrm{Var}[\varepsilon_i] = \sigma_\varepsilon^2$ for all i. Also the errors δ_i and ε_i are mutually uncorrelated. Thus

$$\begin{aligned} \mathrm{Cov}[\delta_i, \delta_j] &= \mathrm{Cov}[\varepsilon_i, \varepsilon_j] = 0\ , \qquad i \neq j\ , \\ \mathrm{Cov}[\delta_i, \varepsilon_j] &= 0\ , \qquad \forall\, i, j\ . \end{aligned}$$

It is possible to rewrite the above model as

$$y_i = \alpha + \beta\, x_i + (\varepsilon_i - \beta\,\delta_i)\ , \qquad i = 1, ..., n\ .$$

This highlights the difference between this problem and the standard regression model. The error term is clearly dependent on β. In addition to this term $(\varepsilon - \beta\delta)$ is correlated with x. Indeed,

$$\mathrm{Cov}[x, \varepsilon - \beta\delta] = E\big[x(\varepsilon - \beta\delta)\big] = E\big[(\xi + \delta)(\varepsilon - \beta\delta)\big] = -\beta\,\sigma_\delta^2$$

and is only zero if $\beta = 0$ or $\sigma_\delta^2 = 0$. If $\sigma_\delta^2 = 0$, the model is equivalent to standard y on x regression, and the usual results apply. See [16] for details on standard regression models.

There have been several reviews of errors in variables methods, notably [9], [10], [20], [37] and [55]. Unfortunately the notation has not been standardised. This paper closely follows the notation set out by Cheng and Van Ness in [10] but for convenience, it has been necessary to modify parts of their notation. All notation will be carefully introduced at the appropriate time.

Errors in variables modelling can be split into two general classifications defined in [35], and [36], as the functional and structural models. The fundamental difference between these models lies in the treatment of the ξ_i's:

The functional model – This assumes the ξ_i's to be unknown, but fixed constants μ_i.

The structural model – This model assumes the ξ_i's to be a random sample from a random variable with mean μ and variance σ^2. The linear structural model is thus the linear model described above, with the ξ_i's taken in a structural setting.

Due to the wealth of literature available this paper will focus mainly on the linear structural model. It will however prove prudent at times to mention the linear functional model at certain places in the text. Further information on the linear functional model is provided in [26].

3. AN OVERVIEW OF ERRORS IN VARIABLES MODELLING

3.1. Origins and beginnings

The author first associated with the errors in variables problem was Adcock ([1], [2]). In the late 1800s he considered how to make the sum of the squares of the errors at right angles to the line as small as possible. This enabled him to find what he felt to be the most probable position of the line. Using ideas from basic geometry, he showed that the errors in variables line must pass through the centroid of the data. However, Adcock's results were somewhat restrictive in that he only considered equal error variances ($\sigma_\delta^2 = \sigma_\varepsilon^2$). These ideas are linked to what is commonly referred to as orthogonal regression. Orthogonal regression minimises the orthogonal distances (as opposed to vertical or horizontal distances in standard linear regression) from the data points onto the regression line.

Adcock's work was extended a year later by Kummel in [38]. Instead of assuming equal error variances, he assumed that the ratio $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ was known instead. Kummel derived an estimate of the line which clearly showed the relation between his and Adcock's work. Kummel argued that his assumption of knowing λ was not unreasonable. He suggested that most experienced practitioners have sufficient knowledge of the error structure to agree a value for this ratio. Use of the orthogonal regression line has been questioned by some authors on the grounds that if the scale of measurement of the line is changed, then a different line would be fitted. However, this is only going to be true if λ is not modified along with the scale of measurement. If λ is modified along with the scale of measurement, the same line is fitted.

The idea of orthogonal regression was included in a book by Deming in [15], and so orthogonal regression is sometimes referred to as Deming regression, particularly in the medical literature. He noted that just as the orthogonal projections from the data to the regression line may be taken, so can any other projection. This would then take account of unequal error variances. The least squares method can then be used to minimise this residual error. This assumes that the error structure is homoscedastic, otherwise this method cannot be used. Lindley in [39] found that adding a weighting factor when minimising the sum of squares of the orthogonal projections, allowed one to minimise projections other than orthogonal.

Another early paper on this subject was by Pearson ([49]). He extended the ideas of previous authors to allow the fitting of lines and hyperplanes of best fit. Pearson was able to show that the orthogonal regression line lies between the y on x, and x on y regression lines.

3.2. Grouping methods

A different approach was suggested by Wald in [62]. Wald described a method that did not make an assumption regarding the error structure. He stressed that there was no justification in making assumptions such as $\lambda = 1$, and that the regression line would not be invariant under transformations of the coordinate system (this criticism has been dealt with in the previous section). Wald suggested splitting the observations into two groups, G_1 and G_2, where G_1 contains the first half of the ordered observations $\big(x_{(1)}, y_{(1)}\big), ..., \big(x_{(m)}, y_{(m)}\big)$ and G_2 contains the second half $\big(x_{(m+1)}, y_{(m+1)}\big), ..., \big(x_{(n)}, y_{(n)}\big)$. An estimate of the slope is then

$$\tilde{\beta}_W = \frac{\big(y_{(1)} + \cdots + y_{(m)}\big) - \big(y_{(m+1)} + \cdots + y_{(n)}\big)}{\big(x_{(1)} + \cdots + x_{(m)}\big) - \big(x_{(m+1)} + \cdots + x_{(n)}\big)}.$$

A problem here is that the grouping must be based on the order of the true values, otherwise, in general, the groups are not independent of the error terms $\delta_1, ..., \delta_n$. Wald countered this by proving that, at least approximately, grouping with respect to the observed values is the same as grouping with respect to the true values. Properties of this estimator for finite samples, as well as approximations of the first four moments can be found in [29].

The idea of grouping the observations was further developed by Bartlett in [6]. Instead of separating the ordered observed values into two groups, he suggested that greater efficiency would be obtained by separating the ordered observations into three groups, G_1, G_2 and G_3. G_1 and G_3 are the outer groups, and G_2 is the middle group. (Nair and Banerjee [44]) show that for a functional model, Bartlett's grouping method provided them with a more efficient estimator of the slope than Wald's method. In Bartlett's method the slope is found by drawing a line through the points $(\bar{x}_{G_1}, \bar{y}_{G_1})$ and $(\bar{x}_{G_3}, \bar{y}_{G_3})$, where $(\bar{x}_{G_1}, \bar{y}_{G_1})$ and $(\bar{x}_{G_3}, \bar{y}_{G_3})$ are the mean points of the observations in G_1 and G_3 respectively. In effect, the observations in G_2 are not used after the data are grouped. In [25] advice on how to place the data into these three groups to obtain the most efficient estimate of the slope is given. How the data should be grouped depended on the distribution of ξ. A table summarising their results for a variety of distributions of ξ can be found in the review paper [40].

Neyman and Scott in [46] suggested another grouping method. The methodology they used is as follows. They suggested fixing two numbers, a and b such that $a \leqslant b$. The numbers a and b must be selected so $P[x \leqslant a] > 0$ and $P[x > b] > 0$. The observations x_i are then divided into three groups, G_1, G_2 and G_3. If $x_i \leqslant a$ those observations are put into G_1, if $a < x_i \leqslant b$ those observations are put into G_2, and if $x_i > b$ those observations are put into G_3. A further two numbers $-c$ and d are then found such that $P[-c \leqslant \delta \leqslant d] = 1$. An estimator of the slope is then given by

$$\tilde{\beta}_{\mathrm{NS}} = \frac{\bar{y}_{G_3} - \bar{y}_{G_1}}{\bar{x}_{G_3} - \bar{x}_{G_1}}$$

and is a consistent estimator of β if

$$P\big[a - c < \xi \leqslant a + d\big] = P\big[b - c < \xi \leqslant b + d\big] = 0\,.$$

However, whether this condition is one that is obtainable in practice is open to debate.

Grouping methods, in particular Wald's method, have been criticised by Pakes in [47]. He claimed that the work of in [29] is unnecessary as Wald's estimate is, strictly speaking, inconsistent. Letting $\tilde{\beta}_{\mathrm{W}}$ denote Wald's estimate for the slope, Pakes showed

$$\left|p \lim \tilde{\beta}_{\mathrm{W}}\right| = |\beta| \left| \frac{(\bar{x}_{G_2} - \bar{x}_{G_1})}{(\bar{x}_{G_2} - \bar{x}_{G_1}) + E\big[\delta | x \in G_2\big] - E\big[\delta | x \in G_1\big]} \right| < |\beta|\,,$$

which shows that, in general, Wald's estimate will underestimate the value of the true slope.

However, this expression derived by Pakes offers a similar conclusion to that of Neyman and Scott ([45]). As long as the error δ is bounded (or not too significant) so that the ranks of ξ are at least approximately equal to the ranks of x, then grouping methods should provide a respectable estimator for the slope as the expression $E\big[\delta|x \in G_2\big] - E\big[\delta|x \in G_1\big]$ should be negligible.

3.3. Instrumental variables

Extensive consideration of this method has appeared in the econometrics literature. Essentially, the instrumental variables procedure involves finding a variable w that is correlated with x, but is uncorrelated with the random error component, δ. The estimate for the slope is then

$$\tilde{\beta}_{\mathrm{IV}} = \frac{s_{yw}}{s_{xw}} \, ,$$

where, s_{yw} and s_{xw} are the usual second order sample moments defined as

$$s_{ab} = \frac{1}{n} \sum_{i=1}^{n} (a_i - \bar{a})\,(b_i - \bar{b}) \, ,$$

and $\bar{a} = n^{-1} \sum_{i=1}^{n} a_i$ is the sample mean. In practice however, it is difficult to obtain a good instrumental variable which meets the aforementioned criteria.

The method of grouping can be put into the context of instrumental variables. In [41] it was showed that Wald's grouping method is equivalent to using the instrumental variable

$$w_i = \begin{cases} 1 & \text{if } x_i > \text{median}(x_1, ..., x_n) \, , \\ -1 & \text{if } x_i < \text{median}(x_1, ..., x_n) \, , \end{cases}$$

and similarly Bartlett's grouping method is equivalent to using

$$w_i = \begin{cases} 1 & \text{for the largest } \frac{n}{3} \text{ observations} \, , \\ -1 & \text{for the smallest } \frac{n}{3} \text{ observations} \, , \\ 0 & \text{otherwise} \, . \end{cases}$$

An idea using the ranks of the x_i was proposed by Durbin in [19]. He suggested an estimator of the form

$$\tilde{\beta}_{\mathrm{D}} = \frac{\sum_{i=1}^{n} i \, y_{(i)}}{\sum_{i=1}^{n} i \, x_{(i)}}$$

where $\left(x_{(1)}, y_{(1)}\right), \left(x_{(2)}, y_{(2)}\right), ..., \left(x_{(n)}, y_{(n)}\right)$ are the ordered observations. However, as with grouping methods, it is unlikely that the ranks of the observed data will match the ranks of the true data. So as in Wald's method this estimate is inconsistent.

3.4. Geometric mean

Other than grouping the data, or looking for an instrumental variable, another approach is to simply take the geometric mean of the y on x regression line, and the reciprocal of the x on y regression line. This leads to the estimate

$$\tilde{\beta}_{\mathrm{GM}} = \mathrm{sign}(s_{xy}) \sqrt{\frac{s_{yy}}{s_{xx}}} \, .$$

There is a geometric interpretation of the line having this slope — it is the line giving the minimum sum of products of the horizontal and vertical distances of the observations from the line (Teissier [58]). However, for the estimate to be unbiased (see [32] for example), one must assume that

$$\lambda = \beta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2} \, . \tag{3.1}$$

This is due to

$$\tilde{\beta}_{\mathrm{GM}} \longrightarrow \sqrt{\frac{\beta^2 \sigma^2 + \sigma_\varepsilon^2}{\sigma^2 + \sigma_\delta^2}} \neq \beta \, .$$

A technical criticism of the use of this estimator is that it may have infinite variance (Creasy [13]). This happens when the scatter of the observations is so great that it is difficult to determine if one line or another perpendicular to it should be used to represent the data. As a result, it may be difficult to construct confidence intervals of a respectable finite width. Geometric mean regression has received much attention, primarily in the fisheries literature. Ricker in [50] examined a variety of regression methods applied to fish biology, and promoted the use of geometric mean regression. He claimed that in most situations it is superior to grouping methods, and the geometric mean regression line is certainly one of the easiest to fit. In addition, Ricker also warned that regression theory based on assuming that the data are from a normal distribution may not apply to non-normally distributed data. Great care must be taken by the statistician to ensure the proper conclusions are obtained from the data.

Jolicoeur in [32], again in the fisheries literature, discussed the paper by Ricker. He stated that as geometric mean regression is equivalent to the assumption in equation (3.1) it is difficult to interpret the meaning of the slope, as the error variances σ_δ^2 and σ_ε^2 only contaminate and cannot explain the underlying

relationship between ξ and η. Ricker replied to the paper by Jolicoeur in a letter, and claimed that the ratio (3.1) may not be linked to the presence or the strength of the underlying relationship, but the correlation coefficient will always give an idea as to the strength. Ricker reiterated that geometric mean regression is an intuitive approach, and as long as the assumption (3.1) holds, is a perfectly valid regression tool.

Further discussion on this estimate was initiated by Sprent and Dolby initially in [57]. They discouraged the use of geometric mean regression, due to the unrealistic assumption of (3.1). They both however sympathised with practitioners, especially those in fish biology, who do not have any knowledge regarding λ. In addition, they commented that the correlation coefficient might be misleading in an errors in variables model, due to each of the observations containing error. They did however suggest that a correlation coefficient may be useful in determining if a transformation to linearity has been successful.

An alternative way of looking at geometric mean regression was provided by Barker *et al.* in [4]. Instead of looking at it as a geometrical average, it can be derived in its own right by adopting a so-called least triangles approach. This is where the sum of the areas of the right-angled triangles formed from the horizontal discrepancies from the data point to the regression line, the vertical discrepancies from the data point to the regression line, and the regression line itself, are minimised. They also showed a connection between geometric mean regression and the correlation coefficient, thus refuting the claim by Sprent and Dolby made in [57] that the correlation coefficient has little value in errors in variables modelling.

3.5. Cumulants

Another method of estimation that has been used in errors in variables modelling is the method of moments. A closely related approach to this is using cumulants, which were proposed by Geary in the series of papers [21], [22], [24], [23]. Cumulants can be defined as follows. Assume that X and Y are jointly distributed random variables. Then, provided the expansions are valid in the given domain, the natural logarithm of the joint characteristic function can be written as

$$\psi(t_1,t_2) = \ln\big[\phi(t_1,t_2)\big] = \ln\big[E\big(e^{it_1X+it_2Y}\big)\big] = \sum_{r,s=0}^{\infty} \kappa(r,s)\,\frac{(it_1)^r}{r!}\,\frac{(it_2)^s}{s!}\,. \tag{3.2}$$

Here, ψ is the so-called joint cumulant generating function, and, if $r \neq 0$ and $s \neq 0$ then $\kappa(r,s)$ is called the r,s product cumulant of X and Y. The slope can be estimated via the method of cumulants as follows.

If the true values ξ and η are centred with respect to their true mean, then

the intercept vanishes, and we can write the structural relationship in the form

$$\beta\xi - \eta = 0\,. \tag{3.3}$$

Letting $\kappa_{(x,y)}$ denote the cumulants of (x, y), and $\kappa_{(\xi,\eta)}$ denote the cumulants of (ξ, η) we have

$$\kappa_{(x,y)}(r, s) = \kappa_{(\xi,\eta)}(r, s)\,.$$

This follows from the following important properties of bivariate cumulants (see, for example [10], [48]):

- The cumulant of a sum of independent random variables is the sum of the cumulants.
- The bivariate cumulant of independent random variables is zero.

The joint characteristic function of (ξ, η) is

$$\phi(t_1, t_2) = E\big[e^{it_1\xi + it_2\eta}\big]\,. \tag{3.4}$$

It follows from (3.3) and (3.4) that

$$\beta\frac{\partial\phi}{\partial it_1} - \frac{\partial\phi}{\partial it_2} = E\big[(\beta\xi - \eta)e^{it_1\xi + it_2\eta}\big] = 0\,.$$

If we replace the joint characteristic function ϕ by the cumulant generating function ψ we obtain

$$\beta\frac{\partial\psi}{\partial it_1} - \frac{\partial\psi}{\partial it_2} = \frac{1}{\phi}\left(\beta\frac{\partial\phi}{\partial it_1} - \frac{\partial\phi}{\partial it_2}\right) = 0 \tag{3.5}$$

and it follows from (3.2) and (3.5), for all $r, s > 0$

$$\beta\kappa(r+1, s) - \kappa(r, s+1) = 0\,.$$

If $\kappa(r+1, s) \neq 0$ an estimate for the slope is then

$$\tilde{\beta}_{\mathrm{C}} = \frac{\kappa(r, s+1)}{\kappa(r+1, s)}\,.$$

In reality, the cumulants $\kappa(r, s)$ will have to be replaced by their sample equivalents $K(r, s)$. Details of how these sample cumulants may be computed as functions of sample moments are included in [21].

3.6. Method of moments

Instead of tackling the problem via cumulants, the method of moments can be used. Briefly, this is where a set of estimating equations are derived by equating population moments with their sample equivalents. The method of

moments approach is considered in detail in [27], and so only a brief survey of the existing literature is given here. Kendall and Stuart in [37] derived the five first and second order moment equations for the structural errors in variables model. However, there are six parameters, $\mu, \alpha, \beta, \sigma_x^2, \sigma_\delta^2$ and σ_ε^2 for the structural model. So in order to proceed with the method of moments, some information regarding a parameter must be assumed known, or more estimating equations must be derived by going to the higher moments. Details on the various assumptions that can be made are included in [10], [18], and [37], as well as others. Dunn in [18] gave formulae for many of the estimators of the slope that are included in [27]. However he did not give any information regarding estimators based on higher moments. Neither did he give information about the variances of these estimates. Work on the higher order moment estimating equations has been done in [17], and more recently in [48], [61], [60] and [12]. Drion, in a paper that is infrequently cited [17], looked at an estimate that could be derived through the third order non-central moment equations for a functional model. Drion computed the variances of all the sample moments that he used, and showed that his estimate of the slope is consistent. Prior to this work, Scott in [52] considered the structural model, and also found an estimate based on the third moments. Scott was able to show that if the third central moment of ξ exists, and is non-zero, then the equation

$$F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n}\big[y_i - \bar{y} - b(x_i - \bar{x})\big]^3 = 0$$

has a root $\hat{b}$ which is a consistent estimate of β. This is because the stochastic limit of $F_{n,1}(b)$ is $(\beta - b)^3\mu_{\xi 3}$, where $\mu_{\xi 3}$ denotes the third central moment of ξ. The estimate of the slope is then a function of the third order sample moments. Scott was able to generalise this result. If the random variable ξ has central moments up to and including order $2m+1$ and if at least one of the first m odd central moments $\mu_{\xi,2k+1}$ $(k = 1, 2, ..., m)$ differs from zero, then the equation

$$F_{n,m}(b) = \frac{1}{n}\sum_{i=1}^{n}\big[y_i - \bar{y} - b(x_i - \bar{x})\big]^{2m+1} = 0$$

has a root $\hat{b}$ which is a consistent estimate of β. Scott did warn however, that estimates based on the lower order moments are likely to be more precise than those based on higher order moments. Unfortunately, Scott did not provide a method of extracting the root which will provide the consistent estimate.

More recently, Pal in [48] further examined the possibilities of the moment equations in a structural model. He stated that in economics, the errors in variables situation cannot be ignored, and as a result, least squares estimation is the wrong way to proceed. Pal derived six possible estimators of the slope, but showed that three of these are functions of the other slope estimates, and concluded that there must be infinitely many consistent estimates which can be obtained by taking different functions of the slope estimates he derived. For each

of the six estimates, Pal found their asymptotic variances when the error terms were assumed to follow a normal distribution. He then went on to consider a variety of regression scenarios, such as $\frac{\sigma_\delta^2}{\sigma^2} = 0$, to offer advice as to which estimator has the smallest variance. The asymptotic efficiency of a particular estimate with respect to the least squares estimate was also provided, for different distributions of ξ. A brief review of the method of cumulants, and how errors in variables modelling might be extended to a multiple linear regression model was included towards the end of the paper.

Van Montfort *et al.* in [61] gave a detailed survey on estimators based on third order moments. They provided an optimal estimate of the slope which is a function of three slope estimates. In order to obtain this optimal estimate, the variance-covariance matrix if not known, has to be estimated. By replacing the variance-covariance matrix with its estimate, the optimal estimator is no longer a function of moments up to order three since moments of order lower than three appear in the estimation of the variance-covariance matrix. Van Montfort *et al.*, through a simulation study, demonstrated that the optimal estimate behaves well for a sample size of 50, and is superior to any other third moment estimator. The same study was replicated for a sample size of 25. For this sample size, they stated that the third moment estimates performed badly. A standard assumption is to assume that the errors δ and ε are independent. Van Montfort *et al.* showed that even if δ and ε are linearly related, then their optimal estimator of the slope is still optimal for all consistent estimators of β which are functions of the first, second and third order moments. In addition, the asymptotic properties of the slope estimate are not altered.

A detailed account of alternative approaches to errors in variables modelling was written by Van Montfort in [60]. This text included estimation based on third order moments, extensions to polynomial regressions, using characteristic functions and links to the factor analysis model. More details on the asymptotic variances and covariances of the third order moment slope estimates were provided. This text is an extension of the details included in [61].

The most recent account on using higher moments was that by Cragg in [12]. He extended the work on the moment equations to include those of the fourth order. A problem with moment based estimators however, is stability. It is well known that as the order of the moment increases they become progressively more difficult to estimate and larger sample sizes will be needed to obtain a reliable estimate. Cragg applied a minimum χ^2 approach to the second, third and fourth moments in order to obtain an efficient general moment estimator. This approach again involves finding an estimated variance-covariance matrix. As Cragg noted, this may be difficult as it will involve the eighth order moments. He suggested avoiding this problem by replacing the variance-covariance matrix with some weighting matrix. This will result in less asymptotic efficiency however. In his simulations Cragg used a diagonal weighting matrix with elements $\frac{1}{2}$, $\frac{1}{15}$ and $\frac{1}{96}$

depending on whether the moment equations are based on the second, third or fourth moments respectively. This may be deemed inappropriate as these values correspond to the theoretical variances of the second, third and fourth powers of a normally distributed variable with zero mean and unit variance, even though a normal distribution will not be applicable for every structural model.

A somewhat different use of the method of moments was suggested by Dagenais and Dagenais in [14]. They proposed a consistent instrumental variable estimator for the errors in variables model based on higher moments. In addition, they showed how a regression model may be tested to detect the presence of errors in both variables. Dagenais and Dagenais illustrated their ideas through a number of numerical simulations and showed that their estimator is superior to the ordinary least squares estimate.

3.7. Maximum likelihood

The vast majority of the papers available on errors in variables modelling have adopted a maximum likelihood approach to estimate the parameters. Only a selection of the large number of papers shall be mentioned here. These papers assumed that the pairs of observations (x_i, y_i) are jointly normally and identically distributed. Lindley was one of the first authors to use maximum likelihood estimation for the errors in variables model in [39]. Lindley commented that the likelihood equations are not consistent, unless there is some prior information available on the parameters. He suggested that the most convenient assumption to make is to assume that the ratio λ is known. Estimates of all the relevant parameters are then derived and discussed.

Kendall and Stuart again in [37] reviewed the topic of estimation in an errors in variables model, but concentrated their efforts on the maximum likelihood principle. They commented that the sample means, variances and covariances form sufficient statistics for a bivariate normal distribution. As a result, the solutions of the method of moment estimating equations for the unknown parameters $\mu, \alpha, \beta, \sigma_x^2, \sigma_\delta^2$ are also maximum likelihood solutions, provided that these solutions give admissible estimates (namely, positive estimators for the variances in the model). The conditions to obtain admissible estimates are then outlined. Further details on these conditions, and estimating using the method of moment estimating equations is included in [27]. More detail was given on the problem of having five moment estimating equations, and six parameters to estimate. They suggested various 'cases', each of which consist of a different assumption regarding a subset of the parameters. Estimates for the parameters are derived for each of these 'cases', and advice is given on how to construct confidence intervals. A brief survey on cumulants, instrumental variables and grouping methods was also included in their work.

A disadvantage of the likelihood method in the errors in variables problem is that it is only tractable if all the distributions describing variation in the data are assumed to be normal. In this case a unique solution is only possible if additional assumptions are made concerning the parameters of the model, usually assumptions about the error variances. Nevertheless, maximum likelihood estimators have certain optimal properties and it is possible to work out the asymptotic variance-covariance matrix of the estimators. These were given for a range of assumptions by Hood *et al.* in [30]. In addition, Hood *et al.* conducted a simulation study in order to determine a threshold sample size to successfully estimate their variance-covariance matrix. They concluded that this threshold was approximately 50.

Other papers on the likelihood approach have tended to focus on a particular aspect of the problem. For example, Wong in [63] considered the likelihood equations when the error variances were assumed to be known, and equal. This case has attracted much attention, as if both error variances are known, the problem is overidentified — there are four parameters to be estimated from five estimating equations (be it likelihood equations, or moment equations). To simplify the procedure, Wong used an orthogonal parameterisation in which the slope parameter is orthogonal to the remaining parameters. Approximate confidence intervals for the parameters, information on testing hypotheses about regarding the slope, and the density function for the slope are also included. Prior to this, Barnett also commented on the inherent difficulties in using the maximum likelihood technique in [5].

Again for the structural model, Birch in [7] showed that the maximum likelihood estimate for the slope is the same when both error variances are known, and when the ratio of the error variances, λ is known. He also commented that the maximum likelihood estimates provided by Madansky in [40] are inconsistent, and as a result need to be modified. Some discussion on the admissability conditions was also included.

A key author in this area was Barnett ([5]). His paper on the fitting of a functional model with replications commented on the importance of errors in variables modelling in the medical and biological areas. The paper adopted the maximum likelihood technique for estimating the parameters, but no closed form solution could be found. He mentioned that the maximum likelihood method tends to run into computational problems due to the awkward nature of the likelihood equations. Barnett also considered alternative error structures which might be applicable to biological and medical areas.

Solari in [54] found that the maximum likelihood solution for the linear functional model discussed by many authors was actually a saddle point, and not a maximum. She said that although the point was purely academic, it was still one worth making. A detailed analysis of the form of the likelihood surface

was given, and and she concluded that a maximum likelihood solution for the linear functional model does not exist, unless one has some prior distribution to place on a parameter. Solari commented that this problem might appear in other estimation problems. Detailed consideration must be given to see if the maximum likelihood solution is indeed a maximum. Sprent considered Solari's work and further noted the practical implications of her findings in [56].

Copas in [11] extended the work of Solari [54]. He showed that when 'rounding-off' errors for the observations are considered, then the likelihood surface becomes bounded. This allows for a different consideration of the likelihood surface. An estimate for the model can be found, which is approximately maximum likelihood. In other words, a point close to the global supremum was used instead. Copas' solution for the slope is equivalent to using either the x on y estimate or the y on x estimate. The y on x regression estimate is used if the line corresponding to the geometric mean estimate lies within $45°$ of the x-axis. The x on y estimate is used if the geometric mean estimate lies within $45°$ of the y-axis. A numerical example was provided to illustrate his suggested methodology, and the likelihood surface for this example was drawn.

Essentially, Copas introduced a modified likelihood function

$$L = \prod_i P_i(x_i)\, Q_i(y_i)$$

where $P_i(x) = P\big(x - \frac{h}{2} \leq \xi_i < x + \frac{h}{2}\big)$ and $Q_i(x) = P\big(y - \frac{h}{2} \leq \beta\xi_i < y + \frac{h}{2}\big)$ (note that Copas' model did not include an intercept). The value h was introduced to allow a discrepancy when $(\xi_i, \beta\xi_i)$ were recorded or measured. The saddle point noted by Solari, according to Copas, is a direct consequence of the likelihood function having singularities at all points within the sets

$$A = \left\{\beta, \sigma_\delta, \sigma_\varepsilon, \underline{\xi} : \sum (x_i - \xi_i)^2 = 0,\ \sigma_\delta = 0\right\}$$

and

$$B = \left\{\beta, \sigma_\delta, \sigma_\varepsilon, \underline{\xi} : \sum (y_i - \beta\xi_i)^2 = 0,\ \sigma_\varepsilon = 0\right\}.$$

Copas showed that within these sets A and B his modified likelihood function reduces to the likelihood function for y on x regression and x on y regression respectively. This however is to be expected as set A essentially assumes that there is no horizontal error (δ) present and set B essentially assumes that there is no vertical error (ε) present. In addition, Copas' analyses assume that h is small, which will also imply that the simple linear regression techniques such as y on x and x on y regression are appropriate.

In summary, Copas' method is equivalent to using y on x regression if it appears that ξ_i is close to x_i, and x on y regression if $\beta\xi_i$ is close to y_i. The choice of which regression to use depends on the location of the geometric mean

regression line. Copas admitted that the y on x and x on y regression estimators do not maximise his likelihood function L. So, as it is well known that y on x and x on y regression are biased, and can only offer a crude approximation to the true line, the method proposed by Copas must be questioned.

3.8. Total least squares

Total least squares is a method of estimating the parameters of a general linear errors in variables model and was introduced by Golub and Van Loan in [28], which is frequently cited in the computational mathematics and engineering literature. Broadly speaking, total least squares may be viewed as an optimisation problem with an appropriate cost function. The standard formulation of the total least squares problem is as follows. Consider a linear measurement error model

$$\mathbf{A}\,\mathbf{X} \simeq \mathbf{B}$$

where $\mathbf{A} = \mathbf{A}_0 + \tilde{\mathbf{A}}$ and $\mathbf{B} = \mathbf{B}_0 + \tilde{\mathbf{B}}$. It is assumed that the underlying physical relationship $\mathbf{A}_0\mathbf{X}_0 = \mathbf{B}_0$ exists.

In total least squares estimation, a matrix $\mathbf{D} = [\mathbf{A}\,\mathbf{B}]$ is constructed which contains the measured data, and the parameter matrix $\mathbf{X}$ is to be estimated. There is an assumption that there exists a true unknown value of the data $\mathbf{D}_0 = [\mathbf{A}_0\mathbf{B}_0]$ and a true value of the parameters $\mathbf{X}_0$ such that $\mathbf{A}_0\mathbf{X}_0 = \mathbf{B}_0$. However, the measured data $\mathbf{D}$ depends on some additive error $\tilde{\mathbf{D}} = [\tilde{\mathbf{A}}\,\tilde{\mathbf{B}}]$ so that $\mathbf{D} = \mathbf{D}_0 + \tilde{\mathbf{D}}$.

The ordinary least squares method gives a solution $\mathbf{X}$ such that the Euclidean norm $\|\mathbf{A}\,\mathbf{X} - \mathbf{B}\|$ is minimised. The total least squares technique applies a small correction (measured by the Euclidean norm) $\Delta\mathbf{D} = [\Delta\mathbf{A}\;\Delta\mathbf{B}]$ to the matrix $\mathbf{D}$ such that the equations $(\mathbf{A} + \Delta\mathbf{A})\mathbf{X} = \mathbf{B} + \Delta\mathbf{B}$ are readily solved. Solutions for this system of equations are obtained by computing its singular value decomposition, and this is the precise topic of the paper [28] mentioned earlier.

The total least squares methodology has been extended to generalised total least squares (where the errors are allowed to be correlated), and more recently element-wise total least squares (which deals with non-identically distributed errors). For a brief review of total least squares and its related methods, see for example [42]. A complete monograph on the topic has been written and is contained in [59]. Cheng and Van Ness in [10] noted that total least squares is in its most simple version, orthogonal regression. Hence, this methodology may not be appropriate when there is some different information available on a parameter.

3.9. LISREL

As well as total least squares, another method of estimation which had its origins from computational mathematics is LISREL (which stands for Linear Structural Relationships). LISREL is an example of a structural equation model, and computer software to implement such a model was created by Joreskog and Sorbom (see for example [33]). To use their notation, the LISREL model is formulated as follows:

$$\underline{\eta} = \mathbf{B}\underline{\eta} + \Gamma\underline{\xi} + \underline{\zeta}\,, \tag{3.6}$$

$$\underline{Y} = \Lambda_y\underline{\eta} + \underline{\varepsilon}\,, \tag{3.7}$$

$$\underline{X} = \Lambda_x\underline{\xi} + \underline{\delta}\,, \tag{3.8}$$

where $\underline{\eta}$ is a $(m\times 1)$ vector, $\mathbf{B}$ is a square $(m\times m)$ matrix, Γ is a $(m\times n)$ matrix, $\underline{\xi}$ is a $(n\times 1)$ vector, $\underline{\zeta}$ is an $(m\times 1)$ vector, $\underline{Y}$ is a $(p\times 1)$ vector, Λ_y is a $(p\times m)$ matrix, $\underline{\varepsilon}$ is a $(p\times 1)$ vector, $\underline{X}$ is a $(q\times 1)$ vector, Λ_x is a $(q\times n)$ matrix, and $\underline{\delta}$ is a $(q\times 1)$ vector. At a first glance, the LISREL model combines two factor analysis models, (3.7) and (3.8) into the structural setting of equation (3.6).

The linear structural model outlined in Section 2 may be fitted into a LISREL format as follows. Take $m = n = p = q = 1$, $\mathbf{B} = 0$, $\underline{\zeta} = 0$, $\Gamma = \beta$ and $\Lambda_x = \Lambda_y = 1$. The standard assumption of the LISREL model is to take $E[\underline{\xi}] = E[\underline{\eta}] = 0$. This constrains us to take $\mu = \alpha = 0$ for our model in Chapter 1. The remaining parameters to be estimated are $\beta, \sigma^2, \sigma_\delta^2$ and σ_ε^2.

A LISREL model usually cannot be solved explicitly, and in this scenario an iterative procedure to estimate the parameters is adopted. Essentially, this involves constructing a set of estimating equations for the parameters. The usual methodology is to set the sample variance-covariance matrix equal to the theoretical variance-covariance matrix. The elements of the theoretical variance-covariance matrix are nonlinear functions of the model parameters Λ_x, Λ_y, Γ and the variance-covariance matrices of $\underline{\xi}$, $\underline{\zeta}$, $\underline{\delta}$ and $\underline{\varepsilon}$.

The LISREL model, (as in factor analysis), implies a particular structure for the theoretical variance-covariance matrix. Johnson and Winchern in [31] gave details of the structure, and stated the following identities (they took $\mathbf{B} = 0$ to simplify proceedings)

$$\begin{aligned} E\big[\underline{Y}\,\underline{Y}^{\mathrm{T}}\big] &= \Lambda_y\big(\Gamma\,\Phi\,\Gamma^{\mathrm{T}} + \psi\big)\,\Lambda_y^{\mathrm{T}} + \Theta_\varepsilon\,, \\ E\big[\underline{X}\,\underline{X}^{\mathrm{T}}\big] &= \Lambda_x\,\Phi\,\Lambda_x^{\mathrm{T}} + \Theta_\delta\,, \\ E\big[\underline{X}\,\underline{Y}^{\mathrm{T}}\big] &= \Lambda_y\Gamma\,\Phi\,\Lambda_x^{\mathrm{T}}\,, \end{aligned}$$

where $E\big[\underline{\xi}\,\underline{\xi}^{\mathrm{T}}\big] = \Phi$, $E\big[\underline{\delta}\,\underline{\delta}^{\mathrm{T}}\big] = \Theta_\delta$, $E\big[\underline{\varepsilon}\,\underline{\varepsilon}^{\mathrm{T}}\big] = \Theta_\varepsilon$ and $E\big[\underline{\zeta}\,\underline{\zeta}^{\mathrm{T}}\big] = \psi$. It is assumed

that the variables ζ, δ and ε are mutually uncorrelated. Also ζ is uncorrelated with ξ, ε is uncorrelated with η and δ is uncorrelated with ξ.

The iteration procedure mentioned above begins with some initial parameter estimates, to produce the theoretical variance-covariance matrix which approximates the sample theoretical variance-covariance matrix. However, for this estimation procedure to occur, there must be at least as many estimating equations as parameters. Indeed Johnson and Winchern state that if t is the number of unknown parameters then the condition

$$t \leq \frac{1}{2}(p+q)(p+q+1)$$

must apply to allow estimation of the parameters. For our model of Section 2, $t = 4$ (β, σ^2, σ_δ^2 and σ_ε^2) and $\frac{1}{2}(p+q)(p+q+1) = 3$ and so we cannot use the LISREL environment to estimate our parameters unless we assume something further is known. This ties in with the thoughts of Madansky who stated in [40] that

> "To use standard statistical techniques of estimation to estimate β,
> one needs additional information about the variance of the estimators."

Also, comparisons may be drawn between LISREL, the method of moments and maximum likelihood, as both of the latter methods also assume that there is some parameter known to allow identifiability of the model.

Applying the LISREL methodology to the linear structural model of Section 2, we get

$$\begin{aligned} E\big[\underline{Y}\,\underline{Y}^{\mathrm{T}}\big] &= \beta^2\sigma^2 + \sigma_\varepsilon^2\,, \\ E\big[\underline{X}\,\underline{X}^{\mathrm{T}}\big] &= \sigma^2 + \sigma_\delta^2\,, \\ E\big[\underline{X}\,\underline{Y}^{\mathrm{T}}\big] &= \beta\sigma^2\,, \end{aligned}$$

since for our model $\Phi = \sigma^2$, $\psi = 0$, $\Theta_\delta = \sigma_\delta^2$ and $\Theta_\varepsilon = \sigma_\varepsilon^2$. We can now equate the theoretical variance-covariance matrix to the sample variance-covariance matrix to construct the following three equations

$$\sigma^2 + \sigma_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s_{xx}\,, \tag{3.9}$$

$$\beta^2\sigma^2 + \sigma_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s_{yy}\,, \tag{3.10}$$

$$\beta\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = s_{xy}\,, \tag{3.11}$$

which are identical to the method of moment estimating equations (and subsequently the maximum likelihood estimating equations) outlined in [27].

The first order moment equations $\mu = \bar{x}$ and $\alpha + \beta\mu = \bar{y}$ are missing as the LISREL model assumes the data are centred, so μ and α are taken as known in the assumption $E[\underline{\xi}] = E[\underline{\eta}] = 0$. There are three equations (3.9), (3.10), (3.11) and four parameters to be estimated. Hence, in order to solve these equations explicitly we need to restrict the parameter space by assuming something known (e.g. assume σ_δ^2 is known). So LISREL for our model is identical to the method of moments, and thus maximum likelihood. As stated earlier, the method of moments is discussed in [27].

3.10. Review papers and monographs

Over the years several authors have written review articles on errors in variables regression. These include [35], [36], [19], [40], [43] and [3]. Riggs *et al.* in [51] performed simulation exercises comparing some of the slope estimators that have been described in the literature. There are two texts devoted entirely to the errors in variables regression problem, Fuller in [20] and Cheng and Van Ness with [10]. Casella and Berger in their general text [9] has an informative section on the topic, [55] contains chapters on the problem, as do [37] and [18]. Draper and Smith in [16] on the other hand, in their book on regression analysis, devoted only 7 out of a total of almost 700 pages to errors in variables regression. The problem is more frequently described in econometrics texts, for example [34]. In these texts the method of instrumental variables is often given prominence. The text [8] described errors in variables models for non linear regression, and Seber and Wild in [53] included a chapter on this topic.

3.11. Conclusion

The papers described in this presentation are definitive papers that dictated the path of further research in the topic. The sporadic nature of the literature can be seen by looking at the journals from which the papers in this presentation came. Procedures for fitting errors in variables models have been developed in the medical literature, economics literature and statistics literature. There are a plethora of papers available on the linear structural model, and even more on errors in variables in general. The list of references given in this paper are by no means exhaustive, but it is hoped that consolidating some of the key ideas involved in errors in variables modelling into this paper will help stimulate further research into a problem that has existed since the 1800s, and that has interested people in a variety of academic disciplines.

REFERENCES

[1] Adcock, R.J. (1877). Note on the method of least squares, *The Analyst*, **4**(6), 183–184.

[2] Adcock, R.J. (1878). A problem in least squares, *The Analyst*, **5**(2), 53–54.

[3] Anderson, T.W. (1984). Estimating linear statistical relationships, *Ann. Statist.*, **12**, 1–45.

[4] Barker, F.; Soh, Y.C. and Evans, R.J. (1988). Properties of the geometric mean functional relationship, *Biometrics*, **44**(1), 279–281.

[5] Barnett, V.D. (1970). Fitting straight lines – The linear functional relationship with replicated observations, *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, **19**, 135–144.

[6] Bartlett, M.S. (1949). Fitting a straight line when both variables are subject to error, *Biometrics*, **5**, 207–212.

[7] Birch, M.W. (1964). A note on the maximum likelihood estimation of a linear structural relationship, *J. Amer. Statist. Assoc.*, **59**, 1175–1178.

[8] Carroll, R.J.; Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

[9] Casella, G. and Berger, R.L. (1990). *Statistical Inference*, Wadsworth & Brooks, Pacific Grove, CA.

[10] Cheng, C.-L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error*, Kendall's Library of Statistics, **6**, Arnold, London.

[11] Copas, J.B. (1972). The likelihood surface in the linear functional relationship problem, *J. Roy. Statist. Soc. Ser. B*, **34**, 274–278.

[12] Cragg, J.G. (1997). Using higher moments to estimate the simple errors-in-variables model, *The RAND Journal of Economics*, **28**(0), S71–S91.

[13] Creasy, M.A. (1956). Confidence limits for the gradient in the linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **18**, 65–69.

[14] Dagenais, M.G. and Dagenais, D.L. (1997). Higher moment estimators for linear regression models with errors in the variables, *J. Econometrics*, **76**(1–2), 193–221.

[15] Deming, W.E. (1931). The application of least squares, *Philos. Mag. Ser.*, **7**(11), 146–158.

[16] Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*, Third Ed., Wiley-Interscience, Canada.

[17] Drion, E.F. (1951). Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error, *Indagationes Math.*, **13**, 256–260.

[18] Dunn, G. (2004). *Statistical Evaluation of Measurement Errors*, Second Ed., Arnold, London.

[19] Durbin, J. (1954). Errors in variables, *Rev. Inst. Internat. Statist.*, **22**, 23–32.

[20] Fuller, W.A. (1987). *Measurement error models*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

[21] Geary, R.C. (1942). Inherent relations between random variables, *Proc. R. Irish. Acad. Sect. A*, **47**, 1541–1546.

[22] Geary, R.C. (1943). Relations between statistics: the general and the sampling problem when the samples are large, *Proc. R. Irish. Acad. Sect. A*, **22**, 177–196.

[23] Geary, R.C. (1949). Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown, *Econometrica*, **17**, 30–58.

[24] Geary, R.C. (1949). Sampling aspects of the problem from the error-in-variable approach, *Econometrica*, **17**, 26–28.

[25] Gibson, W.M. and Jowett, G.H. (1957). Three-group regression analysis – Part 1: Simple regression analysis, *Appplied Statistics*, **6**, 114–122.

[26] Gillard, J.W. (2007). *Errors in Variables Regression: What is the appropriate model?* Cardiff University PhD Thesis.

[27] Gillard, J.W. and Iles, T.C. (2005). Method of moments estimation in linear regression with errors in both variables, *Cardiff School of Mathematics Technical Report*.

[28] Golub, G.H. and Van Loan, C.F. (1980). An analysis of the total least squares problem, *SIAM J. Numer. Anal.*, **17**(6), 883–893.

[29] Gupta, Y.P. and Amanullah. (1970). A note on the moments of the Wald's estimator, *Statistica Neerlandica*, **24**, 109–123.

[30] Hood, K.; Nix, A.B.J. and Iles, T.C. (1999). Asymptotic information and variance-covariance matrices for the linear structural model, *The Statistician*, **48**(4), 477–493.

[31] Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc.

[32] Jolicouer, P. (1975). Linear regressions in fishery research: some comments, *J. Fish. Res. Board Can.*, **32**(8), 1491–1494.

[33] Joreskog, K.G. and Sorbom, D. (1984). *LISREL VI – Analysis of linear structural relations by maximum likelihood, instrumental variables and least squares methods*, User's Guide, Department of Statistics, University of Uppsala, Uppsala, Sweden.

[34] Judge, G.G.; Griffiths, W.E.; Carter Hill, R. and Lee, T.-C. (1980). *The Theory and Practise of Econometrics*, Wiley, New York.

[35] Kendall, M.G. (1951). Regression, structure and functional relationship – I, *Biometrika*, **38**, 11–25.

[36] Kendall, M.G. (1952). Regression, structure and functional relationship – II, *Biometrika*, **39**, 96–108.

[37] Kendall, M.G. and Stuart, A. (1973). *The Advanced Theory of Statistics*, Volume Two, Third Ed., Charles Griffin and Co. Ltd., London.

[38] Kummel, C.H. (1879). Reduction of observed equations which contain more than one observed quantity, *The Analyst*, **6**, 97–105.

[39] Lindley, D.V. (1947). Regression lines and the linear functional relationship, *Suppl. J. Roy. Statist. Soc.*, **9**, 218–244.

[40] Madansky, A. (1959). The fitting of straight lines when both variables are subject to error, *J. Amer. Statist. Assoc.*, **54**, 173–205.

[41] Maddala, G.S. (1988). *Introduction to Econometrics*, Second Ed., Prentice Hall International, Inc.

[42] Markovsky, I. and Van Huffel, S. (2006). On weighted structured total least squares. In "Large-Scale Scientific Computing", *Lecture Notes in Comput. Sci.*, **3743**, Springer, Berlin, pp. 695–702.

[43] Moran, P.A.P. (1971). Estimating structural and functional relationships, *Multivariate Analysis*, **1**, 232–255.

[44] Nair, K.R. and Banerjee, K.S. (1942). A note on fitting of straight lines if both variables are subject to error, *Sankhya*, **6**, 331.

[45] Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partial consistent observations, *Econometrica*, **16**, 1–32.

[46] Neyman, J. and Scott, E.L. (1951). On certain methods of estimating the linear structural relation, *Ann. Math. Statist.*, **22**, 352–361.

[47] Pakes, A. (1982). On the asymptotic bias of the Wald-type estimators of a straight line when both variables are subject to error, *Int. Econ. Rev.*, **23**, 491–497.

[48] Pal, M. (1980). Consistent moment estimators of regression coefficients in the presence of errors in variables, *J. Econometrics*, **14**, 349–364.

[49] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Philos. Mag.*, **2**, 559–572.

[50] Ricker, W.E. (1973). Linear regressions in fishery research, *J. Fish. Res. Board Can.*, **30**, 409–434.

[51] Riggs, D.S.; Guarnieri, J.A. and Addleman, S. (1978). Fitting straight lines when both variables are subject to error, *Life Sciences*, **22**, 1305–1360.

[52] Scott, E.L. (1950). Note on consistent estimates of the linear structural relation between two variables, *Anal. Math. Stat.*, **21**(2), 284–288.

[53] Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*, Wiley, New York.

[54] Solari, M.E. (1969). The 'maximum likelihood solution' to the problem of estimating a linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **31**, 372–375.

[55] Sprent, P. (1969). *Models in Regression and Related Topics*, Methuen's Statistical Monographs, Matheun & Co Ltd, London.

[56] Sprent, P. (1970). The saddlepoint of the likelihood surface for a linear functional relationship, *J. Roy. Statist. Soc. Ser. B*, **32**, 432–434.

[57] Sprent, P. and Dolby, G.R. (1980). Query: the geometric mean functional relationship, *Biometrics*, **36**(3), 547–550.

[58] Teissier, G. (1948). La relation d'allometrie sa signification statistique et biologique, *Biometrics*, **4**(1), 14–53.

[59] Van Huffel, S. and Vanderwalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia.

[60] Van Montfort, K. (1989). *Estimating in Structural Models with Non-Normal Distributed Variables: Some Alternative Approaches*, DSWO Press, Leiden.

[61] Van Montfort, K.; Mooijaart, A. and de Leeuw, J. (1987). Regression with errors in variables: estimators based on third order moments, *Statist. Neerlandica*, **41**(4), 223–237.

[62] Wald, A. (1940). The fitting of straight lines if both variables are subject to error, *Ann. Math. Statistics*, **11**, 285–300.

[63] Wong, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error, *Biometrika*, **76**(1), 141–148.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted / indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright