ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
80 Anos years 1935 2015

REVSTAT

Statistical Journal

Volume 13, No.2
June 2015

REVSTAT

Statistical Journal

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

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 454 084
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 103–118

LIKELIHOOD RATIO TESTS IN LINEAR MODELS WITH LINEAR INEQUALITY RESTRICTIONS ON REGRESSION COEFFICIENTS

Authors: Miguel Fonseca
– Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
fmig@fct.unl.pt

João Tiago Mexia
– Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
jtm@fct.unl.pt

Bimal K. Sinha
– Department of Mathematics and Statistics, University of Maryland Baltimore County, U.S.A.
sinha@umbc.edu

Roman Zmyślony
– Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Poland
r.zmyslony@wmie.uz.zgora.pl

Received: August 2012 Revised: October 2013 Accepted: November 2013

Abstract:

- This paper develops statistical inference in linear models, dealing with the theory of maximum likelihood estimates and likelihood ratio tests under some linear inequality restrictions on the regression coefficients. The results are widely applicable to models used in environmental risk analysis and econometrics.

Key-Words:

- *likelihood ratio test; linear constrains; regression models.*

AMS Subject Classification:

- 62J05, 62F30.

1. INTRODUCTION

Inference about the regression coefficients in a standard linear regression model under the usual assumptions of normality, independence and homoscedasticity of errors and without any constraints on the regression parameters is quite old. A good amount of research has also been done in this set up under some linear inequality constraints on the regression coefficients (see Liew (1976), Gourieroux *et al.* (1982), Self and Liang (1985), Mukerjee and Tu (1995), Andrews (1999), Andrews (2001), Meyer (2003) and Kopylev and Sinha (2010)). Most of the discussions in these papers are asymptotic in nature, and also under the assumption that the underlying dispersion matrix of errors is either completely known or asymptotically estimated and hence used as if it were known. It turns out that, under linear inequality constraints on the regression coefficients, quite often the null distribution of the likelihood ratio test statistic for the nullity of a regression coefficient is a linear combination of several independent chisquares rather than being just one chisquare. We add that the proofs in some papers are geometric in nature while in others it is algebraic in nature, but they are quite involved in both the cases due to the very nature of the model and the testing problem.

A brief literature review is in order. The first paper on this topic seems to be due to Gourieroux *et al.* (1982), followed by the celebrated paper by Self and Liang (1985). The emphasis in both the papers is the derivation of the asymptotic properties of the maximum likelihood estimates and the associated LRT when some parameters lie on their boundaries. In an excellent paper by Mukerjee and Tu (1995), the exact small sample LRT is derived and its properties have been studied in the special case of a simple linear regression model with the nonnegativity restriction on both the intercept and the slope parameters, and inference being on an arbitrary linear function of the two parameters. The paper by Meyer (2003) discusses a test for linear regression versus convex regression while Kopylev and Sinha (2010), primarily motivated by Self and Liang (1985), develop explicit and useful expressions of the MLEs and LRTs in dimensions two and three, the entire treatment being asymptotic in nature.

In this paper we revisit this important inference problem in the case of a standard linear regression model with some linear inequality constraints on the regression coefficients and develop the LRT for the nullity of just one linear function when the variance is unknown. Our treatment is exact, and we offer two solutions. This is in the same spirit as in Mukerjee and Tu (1995). The paper is organized as follows. In Section 2 we consider the linear regression problem with two regression coefficients, both being nonnegative, and derive the LRT for the nullity of one of them. In Section 3 we consider the case of a linear regression with three regression coefficients, all of which are nonnegative, and describe the LRT for the nullity of one of them. In both the settings, normality and independence of errors with an unknown variance are assumed. In each case

we derive the likelihood ratio test and discuss some aspects of the corresponding null distribution of the LRT. Results of some simulation studies are reported in Section 4 in the case of two regression coefficients, comparing the Type I errors of the usual LRT (without taking into account any correction due to nonnegativity of regression coefficients) and the proposed LRT, clearly showing the benefit of the corrections. Such benefits have also been observed and reported in Mukerjee and Tu (1995).

We end this section with a general observation that in the context of a linear model

$$\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{W}) \,, \tag{1.1}$$

if there are known linear inequality constraints on the regression coefficients $\boldsymbol{\beta}$, and the inference problem is to test the equality of such a linear constraint versus it is bigger (or smaller), under a suitable (known) matrix transformation we can always assume without any loss of generality that the inequality constraints as well as the testing problem depend solely on the regression coefficients themelves. This is precisely the formulation we adopt in the remainder of the paper. We also observe an important point from Self and Liang (1985) and Kopylev and Sinha (2010). Under normality and independence of errors, the maximization of the likelihood with respect to the entire regression coefficients $\boldsymbol{\beta}$, which is is equivalent to the minimization of the familiar normal quadratic form $(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\,\mathbf{V}\,(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})$ with respect to the regression coefficients $\boldsymbol{\beta}$, where $\mathbf{V}$ is the estimates covariance matrix, can be safely carried out only with respect to the subset of the regression coefficients which satisfy the inequality constraints, thus completely ignoring the minimization aspect with respect to the unrestricted regression coefficients. Hence, although our proposed solutions in this paper are derived for linear regression models with two nonnegative regression coefficients, this formulation can be adapted for any number of unrestricted regression coefficients!

These results can be very useful in econometrics, extending, for example, the results of Andrews (1999) and Andrews (2001). Similar methodologies can also be applied in environmental risk analysis, as it can be seen in Sinha, Kopylev and Fox (2012).

2. TWO REGRESSION COEFFICIENTS ON THE BOUNDARY

2.1. Model

Consider the linear model

$$\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{W}) \,, \tag{2.1}$$

with unknown parameters $\boldsymbol{\beta}$ and σ^2, and known matrices $\mathbf{X}$ and $\mathbf{W}$. Then the usual maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and σ^2 are given by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}y \,, \tag{2.2}$$

$$\tilde{\sigma}^2 = \frac{S}{n} \,, \qquad S = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{W}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) \,. \tag{2.3}$$

We assume without any loss of generality that $\boldsymbol{\beta} = (\beta_1, \beta_2)$, and derive below the likelihood ratio test (LRT) statistic for the hypothesis

$$H_0 \colon \beta_1 = 0 \;\; \text{vs.} \;\; H_1 \colon \beta_1 > 0 \,; \qquad \beta_1, \beta_2 \geq 0 \,. \tag{2.4}$$

As mentioned earlier, we point out that any linear regression model with a linear inequality constraint on the original regression coefficients and a linear hypothesis on another linear function of them can be reduced to the above setup by suitable linear transformations of $\underset{\sim}{\mathrm{y}}$. Let $\mathbf{V}_{\boldsymbol{\psi}}$ be the inverse of the Fisher information matrix for $\boldsymbol{\psi} = [\beta_1 \; \beta_2 \; \sigma^2]$, which will have the form

$$\mathbf{V}_{\boldsymbol{\psi}} = \begin{bmatrix} \sigma^2(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{n} \end{bmatrix}. \tag{2.5}$$

All throughout we assume that σ^2 is unknown, and we proceed in two ways to develop a test for H_0. Our first approach is based on taking σ^2 to be known and deriving an LRT for H_0, and then replacing σ^2 by its natural estimate, namely, the sample residual variance, and checking what kind of properties the resultant test statistic would possess. This is done by extensive simulation carried out in Section 4. The second approach is to derive the *genuine* LRT when σ^2 is unknown. Although the latter test statistic has an explicit form, its null distribution is rather complicated. We study its properties again by simulation in Section 4. A point of caution is in order here. Unlike the asymptotic treatments in Self and Liang (1985) and Kopylev and Sinha (2010), the null distributions of the test statistics in both the above cases depend on the nuisance parameter β_2 (in fact, via β_2/σ). This is in sharp contrast with the contents of *all* the previous papers!

2.2. σ^2 known

The derivation of the LRT in this case essentially follows from Kopylev and Sinha (2010) who derived it algebraically. We provide below an alternative proof using some geometrical arguments. Following the results presented in [5], the LRT statistic for known σ will have the exact form:

$$L = \min_{\boldsymbol{\beta} \in C} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}')(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \min_{\boldsymbol{\beta} \in C_0} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}')(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \,, \tag{2.6}$$

where C is the cone represented by $\mathbb{R}_0^+ = \prod_{i=1}^2 [0, +\infty[$ and C_0 is the cone represented by $\{0\} \times [0, +\infty[$.

Upon simplification, we get an equivalent expression for L:

$$\begin{aligned} L_0 = & \min_{\beta_1,\beta_2 \geq 0} \Big[v_{11}(\beta_1 - \tilde{\beta}_1)^2 - 2\, v_{12}(\beta_1 - \tilde{\beta}_1)(\beta_2 - \tilde{\beta}_2) + v_{22}(\beta_2 - \tilde{\beta}_2)^2 \Big] \\ & - \min_{\beta_2 \geq 0} \Big[v_{11}\tilde{\beta}_1^2 - 2\, v_{12}\tilde{\beta}_1(\beta_2 - \tilde{\beta}_2) + v_{22}(\beta_2 - \tilde{\beta}_2)^2 \Big] \end{aligned} \tag{2.7}$$

where v_{11}, v_{12} and v_{22} come from

$$\mathbf{V} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}. \tag{2.8}$$

Note that dividing the estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ as well as the parameters β_1 and β_2 by $\sqrt{v_{11}}$ and $\sqrt{v_{22}}$, respectively, we can rewrite L_0 as

$$\begin{aligned} \ell = & \min_{\theta_1,\theta_2 \geq 0} \Big[(\theta_1 - \tilde{\theta}_1)^2 - 2\, \rho(\theta_1 - \tilde{\theta}_1)(\theta_2 - \tilde{\theta}_2) + (\theta_2 - \tilde{\theta}_2)^2 \Big] \\ & - \min_{\theta_2 \geq 0} \Big[(1 - \rho^2)\tilde{\theta}_1^2 + (\theta_2 - \tilde{\theta}_{2.1})^2 \Big] \end{aligned} \tag{2.9}$$

where $\boldsymbol{\theta} = \operatorname{diag}\Big(\frac{1}{\sqrt{v_{11}}}, \frac{1}{\sqrt{v_{22}}}\Big)\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\theta}} = \operatorname{diag}\Big(\frac{1}{\sqrt{v_{11}}}, \frac{1}{\sqrt{v_{22}}}\Big)\tilde{\boldsymbol{\beta}}$, and $\tilde{\theta}_{2.1} = \tilde{\theta}_2 - \rho\tilde{\theta}_1$. It is easy to see that the hypotheses H_0 and H_1 remain invariant under this transformation.

2.3. Minimization

Let us assume that $\rho > 0$ and start with the minimization of $Q(\theta_1, \theta_2) = (\theta_1 - \tilde{\theta}_1)^2 - 2\,\rho(\theta_1 - \tilde{\theta}_1)(\theta_2 - \tilde{\theta}_2) + (\theta_2 - \tilde{\theta}_2)^2$. When $\min\{\tilde{\theta}_1, \tilde{\theta}_2\} < 0$ and $\tilde{\theta}_2 \geq \rho\tilde{\theta}_1$, putting

$$\begin{cases} x = \theta_1 - \tilde{\theta}_1 \\ y = \theta_2 - \tilde{\theta}_2 \end{cases}, \tag{2.10}$$

the level curves of the ellipsoid for the d-level curve are given by

$$x^2 - 2\,\rho x y + y^2 = d^2 \tag{2.11}$$

and, choosing the positive value,

$$x = \rho y + \sqrt{d^2 - (1-\rho^2)y^2}\,, \tag{2.12}$$

we get

$$\frac{dx}{dy} = \rho - \frac{(1-\rho^2)y}{\sqrt{d^2 - (1-\rho^2)y^2}}\,. \tag{2.13}$$

To have $\frac{dx}{dy} = 0$, one must have $y = \pm\frac{\rho d}{\sqrt{1-\rho^2}}$. Since the solution we seek is positive, so

$$x = \frac{d}{\sqrt{1-\rho^2}} \,. \tag{2.14}$$

In order that the vertical tangent coincides with the vertical axis one must have $x = -\tilde{\theta}_1$, and so $d = -\tilde{\theta}_1\sqrt{1-\rho^2}$ and $y = -\tilde{\theta}_1\rho$. Thus,

$$\begin{cases} \theta_1 = 0 \\ \theta_2 = \tilde{\theta}_2 - \rho\tilde{\theta}_1 = \tilde{\theta}_{2.1} \end{cases} . \tag{2.15}$$

We also have that $Q(0, \tilde{\theta}_{2.1}) = (1-\rho^2)\tilde{\theta}_1^2$. Under these assumptions, suppose we can attain a value smaller than $(1-\rho^2)\tilde{\theta}_1^2$, say $\left(-\tilde{\theta}_1\sqrt{1-\rho^2} - \epsilon\right)^2$, with $\epsilon > 0$. In that case, the largest value possible for x would be

$$x \,=\, -\tilde{\theta}_1 - \frac{\epsilon}{\sqrt{1-\rho^2}} \,, \tag{2.16}$$

which implies that

$$\theta_1 \,=\, -\frac{\epsilon}{\sqrt{1-\rho^2}} \,<\, 0 \,, \tag{2.17}$$

which is *not* a valid solution for θ_1.

Analogously, when $\min\{\tilde{\theta}_1, \tilde{\theta}_2\} < 0$ and $\tilde{\theta}_2 \leq \rho^{-1}\tilde{\theta}_1$ we have $(\theta_1, \theta_2) = (\tilde{\theta}_{1,2}, 0)$ and $Q(\tilde{\theta}_{1.2}, 0) = (1-\rho^2)\tilde{\theta}_2^2$.

On the other hand, when $\rho^{-1}\tilde{\theta}_1 < \tilde{\theta}_2 < \rho\tilde{\theta}_1$, we take $(\theta_1, \theta_2) = (0, 0)$ and get $Q(0,0) = \tilde{\theta}_1^2 - 2\,\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2$. If we take any other valid solution, say (ϵ_1, ϵ_2), with $\epsilon_1, \epsilon_2 > 0$, it is easy to see that

$$Q(\epsilon_1, \epsilon_2) - Q(0,0) \,=\, -2\,\tilde{\theta}_{2.1}\epsilon_1 - 2\,\tilde{\theta}_{1.2}\,\epsilon_2 + \epsilon_1^2 + \epsilon_2^2 - 2\,\rho\epsilon_1\epsilon_2 \,>\, 0 \,, \tag{2.18}$$

and so the optimal solution is in fact $(0, 0)$. Summing up the various cases, the first term of ℓ, written as Q_1, simplifies to

$$\frac{Q_1}{\sigma^2(1-\rho^2)} \,=\, \begin{cases} 0\,; & \tilde{\theta}_1 > 0, \;\; \tilde{\theta}_2 > 0 \\ \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0, \;\; \tilde{\theta}_1 < 0 \\ \tilde{\theta}_2^2\,; & \tilde{\theta}_{1.2} > 0, \;\; \tilde{\theta}_2 < 0 \\ \dfrac{\tilde{\theta}_1^2 - 2\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}\,; & \tilde{\theta}_{1.2} < 0, \;\; \tilde{\theta}_{2.1} < 0 \end{cases} . \tag{2.19}$$

Under the null hypothesis, *i.e.*, when $\theta_1 = 0$, one gets the likelihood

$$\ell_0 \,=\, (1-\rho^2)\tilde{\theta}_1^2 + (\theta_2 - \tilde{\theta}_{2.1})^2 \,. \tag{2.20}$$

It is easy to see that the minimum for this ℓ_0 is achieved when $\theta_2 = \tilde{\theta}_{2.1}$ for $\tilde{\theta}_{2.1} \geq 0$, and $\theta_2 = 0$ for $\tilde{\theta}_{2.1} < 0$. So, the second term of ℓ define Q_0 as

$$\frac{Q_0}{\sigma^2(1-\rho^2)} = \begin{cases} \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0 \\ \dfrac{\tilde{\theta}_1^2 - 2\,\rho\tilde{\theta}_1\tilde{\theta}_2 + \tilde{\theta}_2^2}{1-\rho^2}\,; & \tilde{\theta}_{2.1} < 0 \end{cases} . \tag{2.21}$$

2.4. Likelihood ratio with known σ^2

Combining the above results, it follows that the LRT rejects H_0 for large values of λ given by

$$\lambda = \begin{cases} \tilde{\theta}_1^2\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 > 0 \\ \dfrac{\tilde{\theta}_{2.1}^2 + (1-\rho^2)\tilde{\theta}_1^2}{2(1-\rho^2)}\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_2 > 0 \\ 0\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 < 0 \\ \dfrac{\tilde{\theta}_{1.2}^2}{1-\rho^2}\,; & \tilde{\theta}_{1.2} > 0, \quad \tilde{\theta}_2 < 0 \\ 0\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_{1.2} < 0 \end{cases} . \tag{2.22}$$

The above representation of the difference of the minimum of the two quadratic forms is exactly similar to what appears in Kopylev and Sinha (2010). At this point two things need to be settled. First, the null distribution of λ, and then the fact that σ^2 is unknown and it needs to be replaced by an estimate. Since under $H_0 : \beta_1 = 0$ and $\beta_2 \geq 0$ is unknown, it is obvious that the exact null distribution of our LRT λ will depend on β_2! This is indeed a major difference between our result and that of Kopylev and Sinha (2010) where the argument is asymptotic in nature, resulting in the null distribution of LRT being independent of σ as well as any nuisance parameter. Below we assume that $\beta_2 = 0$ and derive the null distribution of LRT still assuming that σ^2 is known, and then *rescale* λ to take care of unknown σ^2. We will call this the modified LRT. Simulation studies carried out in Section 4 about the Type I error of the modified LRT for unknown β_2 and unknown σ^2 reveal that the performance of the modified LRT is quite good.

Write $V_1 = \tilde{\theta}_1$, $V_2 = \tilde{\theta}_2$, $W_1 = \frac{\tilde{\theta}_{1.2}}{\sqrt{1-\rho^2}}$ and $W_2 = \frac{\tilde{\theta}_{2.1}}{\sqrt{1-\rho^2}}$, and note that under H_0, $V_1 \sim N(0,1)$, $\mathrm{cov}[V_1, W_2] = \mathrm{cov}[V_2, W_1] = 0$, and $V_2 \sim N[\delta, 1]$ with $\delta > 0$.

We now express

$$\begin{aligned}\mathbb{P}\big[\lambda<x\big] \;=\; & \mathbb{P}\big[V_1^2<x\wedge V_1>0\wedge W_2>0\big]\\ & +\;\mathbb{P}\left[V_1^2+W_2^2<x\wedge W_2>-\frac{\rho}{1-\rho}V_1\wedge W_2>0\right]\\ & +\;\mathbb{P}\big[W_1^2<x\wedge V_2<0\wedge W_1>0\big]\\ & +\;\mathbb{P}\big[V_1<0\wedge W_1>0\wedge W_2<0\big]\,.\end{aligned}\tag{2.23}$$

The computation of the above probabilities can be somewhat complicated using Cartesian coordinates. Below we use the familiar polar coordinates.

It is well known that a random two dimensional vector whose components are two independent normal vectors with null mean value and variance σ^2 has the same distribution as the vector

$$\begin{pmatrix} R\cos U\\ R\sin U\end{pmatrix}\tag{2.24}$$

where

$$R\sim\sqrt{\sigma^2\chi_2^2}\;,\tag{2.25}$$

$$U\sim\mathrm{Unif}(0,2\pi)\;,\tag{2.26}$$

these variables being independent. In fact, the following equality can be obtained:

$$\iint\limits_{A}\frac{e^{\frac{x^2+y^2}{2\,\sigma^2}}}{2\,\pi\,\sigma^2}\;dx\,dy\;=\;\iint\limits_{\Lambda}r\,\frac{e^{\frac{r^2}{2\,\sigma^2}}}{2\,\pi\,\sigma^2}\;du\,dr\;,\tag{2.27}$$

where A is a subset of $\mathbb{R}^2$ and Λ is a subset of $\Omega=[0,\infty[\times[0,2\pi[$. The polar coordinate transformation

$$\mathbf{p}(r,u)\;=\;\begin{cases}x\;=\;r\cos u\\ y\;=\;r\sin u\end{cases}\tag{2.28}$$

guarantees a bijective function between $\mathbb{R}^2$ and Ω (for $(0,0)$, take $r=u=0$).

Hence, applying the polar transformation on the pair $(V_1,W_2)\mapsto(R,U)$ and noting that (V_1,W_2) and (V_2,W_1) are pairs of independent standard normal variables, one can rewrite (2.23) as

$$\begin{aligned}\mathbb{P}\left[\frac{\ell}{\sigma^2}<x\right] \;=\; & \mathbb{P}\big[V_1<0\wedge W_1>0\wedge W_2<0\big]\;+\;\frac{1}{2}\,\mathbb{P}\big[\chi_1^2<x\big]\\ & +\;\mathbb{P}\big[R^2<x\wedge 2\,\pi-\arcsin\rho<U<2\,\pi\big]\,,\\ \mathbb{P}\left[\frac{\ell}{\sigma^2}<x\right] \;=\; & \frac{1}{2}-\frac{\arcsin\rho}{2\,\pi}+\frac{1}{2}\,\mathbb{P}\big[\chi_1^2<x\big]+\frac{\arcsin\rho}{2\,\pi}\;\mathbb{P}\big[\chi_2^2<x\big]\,,\end{aligned}\tag{2.29}$$

noting that $\arctan\frac{-\rho}{1-\rho}=\arcsin\rho$. It is then established the following.

Theorem 2.1. *The exact distribution of the LRT statistic under* $H_0\colon \theta_1 = 0$ *vs* $H_1\colon \theta_1 > 0$, *for* $\theta_2 = 0$ *and known* σ^2 *is a mixture of* χ^2_0, χ^2_1 *and* χ^2_2 *with coefficients* $\frac{1}{2} - p$, $\frac{1}{2}$ *and* p, *respectively, with*

$$p = \frac{\arcsin(\rho)}{2\pi} ,$$

where ρ *the correlation coefficient between* $\tilde{\theta}_1$ *and* $\tilde{\theta}_2$.

Up until now, it was assumed that the variance component σ^2 was known, which in practice is rarely the case. Take

$$S_0 = \frac{nS}{n-p} , \tag{2.30}$$

where S, defined in (2.3) is the maximum likelihood estimator for σ^2, which is independent of $\tilde{\theta}_1$ and $\tilde{\theta}_2$, and observe that

$$(n-p)S \sim \sigma^2 \chi^2_{n-p} . \tag{2.31}$$

Hence, one can use S_0 to enable the computation of the distribution of ℓ. Thus, taking the expression in (30) and multiplying ℓ by S_0, one gets

$$\begin{aligned} \mathbb{P}\big[\ell < (1-\rho^2)x\big] &= \mathbb{P}\big[V_1 < 0 \wedge W_1 > 0 \wedge W_2 < 0\big] + \frac{1}{2}\,\mathbb{P}\big[F_{1,n-p} < (1-\rho^2)x\big] \\ &\quad + \mathbb{P}\left[\frac{R^2}{S_0} < (1-\rho^2)x \wedge 2\pi - \arcsin\rho < U < 2\pi\right] , \\ \mathbb{P}\big[\ell < x\big] &= \frac{1}{2} - \frac{\arcsin\rho}{2\pi} + \frac{1}{2}\,\mathbb{P}\big[F_{1,n-p} < (1-\rho^2)x\big] \\ &\quad + \frac{\arcsin\rho}{2\pi}\,\mathbb{P}\big[F_{2,n-p} < (1-\rho^2)x\big] . \end{aligned} \tag{2.32}$$

So, another version of Theorem 2.1 for the rescaled or modified LRT is given by

Theorem 2.2. *The exact distribution of the LRT under* $H_0\colon \theta_1 = 0$ *vs* $H_1\colon \theta_1 > 0$, *for* $\theta_2 = 0$, *is a mixture of* F_0, $F_{1,n-p}$ *and* $F_{2,n-p}$ *with coefficients* $\frac{1}{2} - p$, $\frac{1}{2}$ *and* p, *respectively, with*

$$p = \frac{\arcsin(\rho)}{2\pi} ,$$

where ρ *the correlation coefficient between* $\tilde{\theta}_1$ *and* $\tilde{\theta}_2$.

Tables 1, 2 and 3 represent the rejection probability for some values of β_1, taking $\alpha = 0.05$.

2.5. LRT when σ^2 is unknown

In this section we derive the LRT for H_0 when the error variance $\sigma^2 > 0$ is unknown. Write

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'W^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{W}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) + Q(\boldsymbol{\beta}) = SS_{\text{res}} + Q(\boldsymbol{\beta}) \tag{2.33}$$

where $\hat{\boldsymbol{\beta}}$ is the usual weighted least squares estimate of $\boldsymbol{\beta}$ defined in (3), and $Q(\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}')(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$. Then it easily follows that the LRT statistic λ for H_0 defined as the ratio of null-restricted maximum of the likelihood of $\mathbf{y}$ to the unrestricted maximum of the same likelihood is given by

$$\lambda = \left[\frac{SS_{\text{res}}+Q_1}{SS_{\text{res}}+Q_0}\right]^{n/2} \tag{2.34}$$

where Q_1 and Q_0 are the unrestricted and restricted values of the quadratic $Q(\boldsymbol{\beta})$ under the conditions $\beta_1, \beta_2 \geq 0$ and $\beta_1 = 0$, $\beta_2 \geq 0$, respectively.

Using the expressions for Q_1 and Q_0 from (20) and (22), respectively, and noting that our exact LRT rejects H_0 for large values of $\Delta = \frac{SS_{\text{res}}+Q_0}{SS_{\text{res}}+Q_1}$, we simplify Δ as

$$\Delta = \begin{cases} \dfrac{SS_{\text{res}}}{SS_{\text{res}}+\tilde{\theta}_1^2}\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 > 0 \\ \dfrac{SS_{\text{res}}}{SS_{\text{res}}+\frac{\tilde{\theta}_1^2-2\rho\tilde{\theta}_1\tilde{\theta}_2+\tilde{\theta}_2^2}{1-\rho^2}}\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_2 > 0 \\ 1\,; & \tilde{\theta}_{2.1} > 0, \quad \tilde{\theta}_1 < 0 \\ \dfrac{SS_{\text{res}}+\tilde{\theta}_2^2}{SS_{\text{res}}+\frac{\tilde{\theta}_1^2-2\rho\tilde{\theta}_1\tilde{\theta}_2+\tilde{\theta}_2^2}{1-\rho^2}}\,; & \tilde{\theta}_{1.2} > 0, \quad \tilde{\theta}_2 < 0 \\ 1\,; & \tilde{\theta}_{2.1} < 0, \quad \tilde{\theta}_{1.2} < 0 \end{cases} . \tag{2.35}$$

The *crux* of the problem now is to derive the null distribution of Δ. It is obvious that although the null distribution of Δ is independent of σ^2, it does depend on the unknown second regression coefficient $\beta_2 \geq 0$ as in the previous case. Finding this null distribution even for a specified β_2 turns out to be extremely difficult, and we can present only some simulation results for this purpose.

Tables 7, 8 and 9 represent the rejection probability for some values of β_1 and β_2, taking $\alpha = 0.05$.

Based on these simulation results, we conclude that this test behaves very good, maintaining test size and gaining power compared to the usual F test and the *ad-hoc* test described earlier.

3. NUMERICAL RESULTS

3.1. Ad-hoc test

A set of simulations was performed to evaluate the power performance of the ad-hoc test when σ^2 is unknown. The model assumed was

$$y_j = \beta_0 + \beta_1 x_{1,j} + \beta_2 x_{2,j} + e_j \ , \tag{3.1}$$

with $j = 1, ..., 33$ and $\beta_1, \beta_2 \geq 0$. Considering that $e_j \sim N(0, \sigma^2)$, the procedure consists in generating values for $\tilde{\theta}_1$, $\tilde{\theta}_2$ and S_0, comparing the result of the usual F test with the derived test for $H_0 \colon \beta_1 = 0$ vs $H_1 \colon \beta_1 > 0$. The procedure was repeated 10000 times to obtain an empirical rejection probability. The chosen values for the parameters were:

$$\begin{aligned} \beta_1 &= 0, 1, 3, 10 \\ \beta_2 &= 0, 1, 3, 10, 30 \\ \rho &= 0, 0.1, 0.25, 0.5 \\ \sigma^2 &= 1 \ . \end{aligned} \tag{3.2}$$

The results appear in Tables 1 through 3.

Table 1: Rejection probability of ad-hoc test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.062	0.041	0.056	0.052	0.055
1	0.232	0.275	0.251	0.228	0.243
3	0.900	0.891	0.899	0.905	0.913
10	1.000	1.000	1.000	1.000	1.000

Table 2: Rejection probability of ad-hoc test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.053	0.045	0.044	0.053	0.045
1	0.218	0.229	0.232	0.266	0.238
3	0.879	0.853	0.904	0.900	0.860
10	1.000	1.000	1.000	1.000	1.000

Table 3: Rejection probability of ad-hoc test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.045	0.035	0.053	0.053	0.050
1	0.263	0.225	0.224	0.237	0.232
3	0.917	0.865	0.892	0.899	0.901
10	1.000	1.000	1.000	1.000	1.000

For comparison sake, we also present simulation results for the usual F test (Tables 4 to 6).

Table 4: Rejection probability of F test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0477	0.0471	0.0468	0.0481	0.0460
1	0.1646	0.1554	0.1653	0.1675	0.1592
3	0.8320	0.8293	0.8273	0.8276	0.8210
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5: Rejection probability of F test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0471	0.0512	0.0460	0.0474	0.0491
1	0.1634	0.1648	0.1624	0.1588	0.1577
3	0.8240	0.8296	0.8297	0.8258	0.8317
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6: Rejection probability of F test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0464	0.0504	0.0490	0.0494	0.0500
1	0.1544	0.1551	0.1694	0.1593	0.1630
3	0.8308	0.8275	0.8253	0.8280	0.8288
10	1.0000	1.0000	1.0000	1.0000	1.0000

The power increase for the ad-hoc test over the F test is evidently very significant for $\beta_1 > 0$.

3.2. Exact LR test with unknown σ^2

A batch of simulations was run for the exact test Δ. The set of parameters considered, the same as for the previous batch of simulations, was:

$$\begin{aligned} \beta_1 &= 0, 1, 3.10 \\ \beta_2 &= 0, 1, 3, 10, 30 \\ \rho &= 0, 0.25, 0.5 \\ \sigma^2 &= 1 \ . \end{aligned} \tag{3.3}$$

The results appear in Tables 7 to 9.

Table 7: Rejection probability of exact LR test: $\rho = 0$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0470	0.0526	0.0495	0.0495	0.0498
1	0.2476	0.2548	0.2485	0.2493	0.2466
3	0.9030	0.9015	0.9000	0.8961	0.8998
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 8: Rejection probability of exact LR test: $\rho = 0.25$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0490	0.0499	0.0484	0.0521	0.0443
1	0.2702	0.2773	0.2512	0.2552	0.2514
3	0.9133	0.9063	0.8963	0.9012	0.8943
10	1.0000	1.0000	1.0000	1.0000	1.0000

Table 9: Rejection probability of exact LR test: $\rho = 0.5$.

$\beta_1 \backslash \beta_2$	0	1	3	10	30
0	0.0486	0.0592	0.0452	0.0459	0.0541
1	0.2769	0.2647	0.2511	0.2403	0.2564
3	0.9485	0.9324	0.8949	0.8934	0.9031
10	1.0000	1.0000	1.0000	1.0000	1.0000

Again, there is a clear gain of power over the usual F test for $\beta_1 > 0$.

3.2.1. Critical Values

The critical values for the exact likelihood ratio test were also obtained (Table 10).

Table 10: 5% quantiles.

$\beta_2 \backslash \rho$	0	0.25	0.5
0	0.9126540	0.9045819	0.8939074
1	0.9136219	0.9139177	0.9121665
3	0.9120875	0.9121392	0.9100027
10	0.9111222	0.9145002	0.9092603
30	0.9108427	0.9106236	0.9129284

It is easy to see that the critical values across the different values of β_2 are similar, and stabilize as this parameter increases. This leads us to believe that the use of these critical values would be valid for a wide range of unknown values of β_2.

ACKNOWLEDGMENTS

Our sincere thanks are due to an anonymous reviewer for some excellent comments and suggestions which led to improved presentation. This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

[1] Andrews, D. (1999). Estimation when a parameter is on a boundary, *Econometrica*, **67**, 1341–1383.

[2] Andrews, D. (2001). Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica*, **69**, 683–734.

[3] Gourieroux, C.; Holly, A. textupand Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica*, **50**, 63–80.

[4] Liew C.K. (1976). Inequality Constrained Least-Squares Estimation, *Jour. Amer. Statist. Assoc.*, **71**, 746–751.

[5] Kopylev, L. and Sinha, B.K. (2010). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary, *Sankhya, Series B*, **73**, 20–41.

[6] Meyer, M.C. (2003). A test for linear versus convex regression function using shape-restricted regression, *Biometrika*, **90**, 223–232.

[7] Mukerjee, H. and Tu, R. (1995). Order-restricted inferences in linear regression, *Jour. Amer. Statist. Assoc.*, **90**, 717–728.

[8] Self, S.G. and Liang, K-Y. (1985). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Jour. Amer. Statist. Assoc.*, **82**, 605–610.

[9] Sinha, B.K.; Kopylev, L. and Fox, J. (2012). Some new aspects of statistical inference for multistage dose-response models with applications, *Technical report.*

REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 119–129

NOTES ON THE REGULAR E-OPTIMAL SPRING BALANCE WEIGHING DESIGNS WITH CORRELATED ERRORS

Authors: Bronisław Ceranka
– Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences,
Wojska Polskiego 28, 60-637 Poznań, Poland
bronicer@up.poznan.pl

Małgorzata Graczyk
– Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences,
Wojska Polskiego 28, 60-637 Poznań, Poland
magra@up.poznan.pl

Received: May 2013 Revised: September 2013 Accepted: November 2013

Abstract:

- The paper deals with the estimation problem of individual weights of objects in E-optimal spring balance weighing design. It is assumed that errors are equal correlated. The topic is focused on the determining the maximal eigenvalue of the inverse of information matrix of estimators. The constructing methods of the E-optimal spring balance weighing design based on the incidence matrices of balanced and partially balanced incomplete block designs are given.

Key-Words:

- *balanced incomplete block design; E-optimal design; partially incomplete block design; spring balance weighing design.*

AMS Subject Classification:

- 62K05, 62K15.

1. INTRODUCTION

Let us suppose we want to estimate the weights of v objects by weighing them b times using a spring balance, $v \leq b$. Suppose, that the results of this experiment can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \,, \tag{1.1}$$

where $\mathbf{y}$ is an $b \times 1$ random vector of the observations, $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$, where $\mathbf{\Phi}_{b\times v}(0,1)$ denotes the class of $b \times v$ matrices $\mathbf{X} = (x_{ij})$ of known elements $x_{ij} = 1$ or 0 according as in the ith weighing operation the jth object is placed on the pan or not. Next, $\mathbf{w}$ is a $v \times 1$ vector of unknown measurements of objects and $\mathbf{e}$ is a $b \times 1$ random vector of errors. We assume, that $\mathrm{E}(\mathbf{e}) = \mathbf{0}_b$ and $\mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{G}$, where $\mathbf{0}_b$ denotes the $b \times 1$ vector with zero elements everywhere, $\mathbf{G}$ is the known $b \times b$ diagonal positive definite matrix of the form

$$\mathbf{G} = g\Big[(1-\rho)\,\mathbf{I}_v + \rho\,\mathbf{1}_v\mathbf{1}_v^{'}\Big] \,, \qquad g > 0, \quad \frac{-1}{b-1} < \rho < 1 \,. \tag{1.2}$$

It should be noticed that the conditions on the values of g and ρ are equivalent to the matrix $\mathbf{G}$ being positive definite. From now on, we will consider $\mathbf{G}$ on the form 1.2 only. Moreover, let note, $\mathbf{G}^{-1} = \frac{1}{g(1-\rho)}\Big[\mathbf{I}_b - \frac{\rho}{1+\rho(b-1)}\,\mathbf{1}_b\mathbf{1}_b^{'}\Big]$. For the estimation of $\mathbf{w}$ we use the normal equations $\mathbf{M}\mathbf{w} = \mathbf{X}^{'}\mathbf{G}^{-1}\mathbf{y}$, where $\mathbf{M} = \mathbf{X}^{'}\mathbf{G}^{-1}\mathbf{X}$ is called the information matrix of $\hat{\mathbf{w}}$. A spring balance weighing design is singular or nonsingular, depending on whether the matrix $\mathbf{M}$ is singular or nonsingular, respectively. From the assumption that $\mathbf{G}$ is positive definite it follows that matrix $\mathbf{M}$ is nonsingular if and only if matrix $\mathbf{X}$ is full column rank. If matrix $\mathbf{M}$ is nonsingular, then the generalized least squares estimator of $\mathbf{w}$ is given by formula $\hat{\mathbf{w}} = \mathbf{M}^{-1}\mathbf{X}^{'}\mathbf{G}^{-1}\mathbf{y}$ and $\mathrm{Var}(\hat{\mathbf{w}}) = \sigma^2\mathbf{M}^{-1}$. Some considerations apply to determining the optimal weighing designs are shown in many books[1]. Some problems related to optimality of the designs are presented in several papers[2] for $\mathbf{G} = \mathbf{I}_n$, whereas in Katulska and Rychlińska ([9]) for the diagonal matrix $\mathbf{G}$.

In this paper, we emphasize a special interest of the existence conditions for E-optimal design, i.e. minimizing the maximum eigenvalue of the inverse of the information matrix. The statistical interpretation of E-optimality is the following: the E-optimal design minimizes the maximum variance of the component estimates of the parameters. It can be described in terms of the maximum eigenvalue of the matrix $\mathbf{M}^{-1}$ as $\lambda_{\max}(\mathbf{M}^{-1})$ or equivalently as $\lambda_{\min}(\mathbf{M})$. Hence, for the given variance matrix of errors $\sigma^2\mathbf{G}$, any design $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ is E-optimal if $\lambda_{\max}(\mathbf{M}^{-1})$ is minimal. Moreover, if $\lambda_{\max}(\mathbf{M}^{-1})$ attains the lowest bound, then $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ is called regular E-optimal. Notice that if the design

[1] See, Raghavarao ([13]), Banerjee ([1]), Shah and Sinha ([15]), Pukelsheim ([12]).
[2] Jacroux and Notz ([8]), Neubauer and Watkins ([11]).

$\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$ is regular E-optimal then is also E-optimal. But the inverse implication may not be true. Moreover, the E-optimal design in the set of all design matrices $\mathbf{\Phi}_{b\times v}(0,1)$ exists but the regular E-optimal design may not exist.

The problem presented in this paper is focused to determining such matrix that $\lambda_{\max}(\mathbf{M}^{-1})$ takes the minimal value over all possible matrices in $\mathbf{\Phi}_{b\times v}(0,1)$ for given matrix $\mathbf{G}$.

2. REGULAR E-OPTIMAL SPRING BALANCE WEIGHING DESIGN

In this section we give some new results concerning the lower bound for $\lambda_{\max}(\mathbf{M}^{-1})$ depending on ρ and number of objects v is even or odd. Additionally, let $\mathbf{\Pi}$ be the set of all $v \times v$ permutation matrices. We shall denote by $\bar{\mathbf{M}}$ the average of $\mathbf{M}$ over all elements of $\mathbf{\Pi}$, i.e. $\bar{\mathbf{M}} = \frac{1}{v!}\sum_{\mathbf{P}\in\mathbf{\Pi}} \mathbf{P}'\mathbf{M}\,\mathbf{P}$. It is not difficult to see that

$$\bar{\mathbf{M}} = \frac{v\,\mathrm{tr}(\mathbf{M}) - \mathbf{1}_v'\mathbf{M}\mathbf{1}_v}{v(v-1)}\,\mathbf{I}_v + \frac{\mathbf{1}_v'\mathbf{M}\mathbf{1}_v - \mathrm{tr}(\mathbf{M})}{v(v-1)}\,\mathbf{1}_v\mathbf{1}_v',$$

moreover, $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\bar{\mathbf{M}})$ and $\mathbf{1}_v'\mathbf{M}\mathbf{1}_v = \mathbf{1}_v'\bar{\mathbf{M}}\mathbf{1}_v$. The matrix $\bar{\mathbf{M}}$ has two eigenvalues $\mu_1 = \frac{v\,\mathrm{tr}(\mathbf{M})-\mathbf{1}_v'\mathbf{M}\mathbf{1}_v}{v(v-1)}$ with the multiplicity $v-1$ and $\mu_2 = \frac{\mathbf{1}_v'\mathbf{M}\mathbf{1}_v}{v}$ with the multiplicity 1. Let

$$\mathbf{M} = \frac{1}{g(1-\rho)}\left[\mathbf{X}'\mathbf{X} - \frac{\rho}{1+\rho(b-1)}\,\mathbf{X}'\mathbf{1}_b\mathbf{1}_b'\mathbf{X}\right]. \tag{2.1}$$

For $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ and $\mathbf{G}$ we have $\mathrm{tr}(\mathbf{M}) = \frac{1}{g(1-\rho)}\left[\mathbf{1}_v'\mathbf{r} - \frac{\rho}{1+\rho(b-1)}\,\mathbf{r}'\mathbf{r}\right]$ and $\mathbf{1}_v'\mathbf{M}\mathbf{1}_v = \frac{1}{g(1-\rho)}\left[\mathbf{k}'\mathbf{k} - \frac{\rho}{1+\rho(b-1)}\left(\mathbf{1}_b'\mathbf{k}\right)^2\right]$, where $\mathbf{X}\mathbf{1}_v = \mathbf{k}$, $\mathbf{X}'\mathbf{1}_b = \mathbf{r}$, $\mathbf{1}_v'\mathbf{r} = \mathbf{1}_b'\mathbf{k}$. From above, eigenvalues of $\bar{\mathbf{M}}$ are

$$\mu_1 = \frac{1}{v(v-1)\,g(1-\rho)}\left[v\mathbf{1}_b'\mathbf{k} - \mathbf{k}'\mathbf{k} + \frac{\rho}{1+\rho(b-1)}\left((\mathbf{1}_b'\mathbf{k})^2 - v\,\mathbf{r}'\mathbf{r}\right)\right],$$

$$\mu_2 = \frac{1}{v\,g(1-\rho)}\left[\mathbf{k}'\mathbf{k} - \frac{\rho}{1+\rho(b-1)}\left(\mathbf{1}_b'\mathbf{k}\right)^2\right].$$

Thus the matrix $\bar{\mathbf{M}}^{-1}$ has also two eigenvalues $\lambda_1 = \frac{1}{\mu_1}$ and $\lambda_2 = \frac{1}{\mu_2}$. Next, comparing these two eigenvalues we become following lemma.

Lemma 2.1. *For any nonsingular spring balance weighing design* $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$, *the matrix* $\bar{\mathbf{M}}^{-1}$ *has two eigenvalues* λ_1 *and* λ_2 *and moreover*

(**i**) $\lambda_1 > \lambda_2$ *if and only if* $\rho < \frac{\mathbf{k}'\mathbf{k}-\mathbf{1}_b'\mathbf{k}}{(b-1)\,(\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k})+(\mathbf{1}_b'\mathbf{k})^2-\mathbf{r}'\mathbf{r}}$,

(**ii**) $\lambda_1 = \lambda_2$ *if and only if* $\rho = \frac{\mathbf{k}'\mathbf{k}-\mathbf{1}_b'\mathbf{k}}{(b-1)\,(\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k})+(\mathbf{1}_b'\mathbf{k})^2-\mathbf{r}'\mathbf{r}}$,

(**iii**) $\lambda_1 < \lambda_2$ *if and only if* $\rho > \frac{\mathbf{k}'\mathbf{k}-\mathbf{1}_b'\mathbf{k}}{(b-1)\,(\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k})+(\mathbf{1}_b'\mathbf{k})^2-\mathbf{r}'\mathbf{r}}$.

Proof: We have

$$\begin{aligned}\mu_1-\mu_2 &= \tfrac{1}{gv(v-1)(1-\rho)}\cdot \\ &\quad\cdot\Big[v\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k}+\tfrac{\rho}{1+\rho(b-1)}\Big((\mathbf{1}_b'\mathbf{k})^2-v\mathbf{r}'\mathbf{r}\Big)-\tfrac{1}{gv(1-\rho)}\Big(\mathbf{k}'\mathbf{k}-\tfrac{\rho}{1+\rho(b-1)}\,(\mathbf{1}_b'\mathbf{k})^2\Big)\Big] \\ &= \tfrac{1}{g(v-1)\,(1-\rho)\,(1+\rho(b-1))}\Big[\Big(\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k}\Big)\Big(1+\rho(b-1)\Big)+\rho\Big((\mathbf{1}_b'\mathbf{k})^2-\mathbf{r}'\mathbf{r}\Big)\Big].\end{aligned}$$

Because $g(v-1)(1-\rho)(1+\rho(b-1))>0$ then $\mu_1-\mu_2>0$ if and only if $\rho\Big[(b-1)\,(\mathbf{1}_b'\mathbf{k}-\mathbf{k}'\mathbf{k})+(\mathbf{1}_b'\mathbf{k})^2-\mathbf{r}'\mathbf{r}\Big] > \mathbf{k}'\mathbf{k}-\mathbf{1}_b'\mathbf{k}$ and we obtain the Lemma. □

Lemma 2.1 imply that in order to determine E-optimal design we have to delimit the lowest bound of eigenvalues of $\bar{\mathbf{M}}^{-1}$ according to the value of ρ.

Theorem 2.1. *Let v be even. In any nonsingular spring balance weighing design $\mathbf{X}\in\mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$*

(**i**) *if $\rho\in\Big(\frac{-1}{b-1},\ \frac{v-2}{b+v-2}\Big)$ then $\lambda_{\max}(\mathbf{M}^{-1})\geq\frac{4g(v-1)(1-\rho)}{bv}$ and the equality is fulfilled if and only if $\mathbf{X}\mathbf{1}_v=\frac{v}{2}\,\mathbf{1}_b$,*

(**ii**) *if $\rho\in\Big(\frac{v-2}{b+v-2},\ 1\Big)$ then $\lambda_{\max}(\mathbf{M}^{-1})>\frac{g(1+\rho(b-1))}{bv}$.*

Proof: In order to determine regular E-optimal spring balance weighing design we have to give the lowest bound of the maximal eigenvalue of the matrix $\mathbf{M}^{-1}$. Let $\bar{\mathbf{M}}$ denote the average of $\mathbf{M}$ over all elements of $\mathbf{\Pi}$ for the design $\mathbf{X}\in\mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$. From the monotonicity theorem given by Rao and Rao (2004) it follows $\lambda_{\max}(\bar{\mathbf{M}}^{-1})\leq\lambda_{\max}(\mathbf{M}^{-1})$. The proof falls naturally in two parts according to the value ρ in 1.2. If $\rho\in\Big(\frac{-1}{b-1},\ \frac{v-2}{b+v-2}\Big)$ then $\lambda_{\max}(\bar{\mathbf{M}}^{-1})=\frac{gv(v-1)(1-\rho)(1+\rho(b-1))}{v(1+\rho(b-1))\mathbf{1}_b'\mathbf{k}-(1+\rho(b-1))\mathbf{k}'\mathbf{k}+\rho(\mathbf{1}_b'\mathbf{k})^2-\rho v\mathbf{r}'\mathbf{r}}$. As we want to minimize $\lambda_{\max}(\bar{\mathbf{M}}^{-1})$, we should find the maximum value for

$$A\;=\;v\big(1+\rho(b-1)\big)\,\mathbf{1}_b'\mathbf{k}-\big(1+\rho(b-1)\big)\,\mathbf{k}'\mathbf{k}+\rho(\mathbf{1}_b'\mathbf{k})^2-\rho\,v\,\mathbf{r}'\mathbf{r}\;.$$

If p is even

$$A \leq v\big(1+\rho(b-1)\big)\mathbf{1}_b'\mathbf{k} - \big(1+\rho(b-1)\big)bk^2 + \rho(\mathbf{1}_b'\mathbf{k})^2 - \rho\, v^2r^2 = \tag{2.2}$$

$$v\big(1+\rho(b-1)\big)bk - \big(1+\rho(b-1)\big)bk^2 + \rho bk^2 - \rho v^2r^2 \leq \frac{1}{4}\,bv^2\big(1+\rho(b-1)\big)\,. \tag{2.3}$$

Hence $\lambda_{\max}(\mathbf{M}^{-1}) \geq \frac{4g(v-1)(1-\rho)}{bv}$. The equality in inequality 2.2 holds if and only if $k_1 = k_2 = \cdots = k_b = k$ and $r_1 = r_2 = \cdots = r_v = r$, whereas the equality in 2.3 is fulfilled if and only if $k = \frac{v}{2}$. If $\rho \in \left(\frac{v-2}{b+v-2},\ 1\right)$ then $\lambda_{\max}(\bar{\mathbf{M}}^{-1}) = \frac{gv(1-\rho)(1+\rho(b-1))}{\mathbf{k}'\mathbf{k}+\rho(b-1)\mathbf{k}'\mathbf{k}-\rho(\mathbf{1}_b'\mathbf{k})^2}$. So, we obtain $\lambda_{\max}(\mathbf{M}^{-1}) > \frac{g(1+\rho(b-1))}{bv}$. Thus the result. □

Theorem 2.2. *Let v be odd. In any nonsingular spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$*

- **(i)** *if $\rho \in \left(\frac{-1}{b-1},\ \frac{v(v-3)}{b(v+1)+v(v-3)}\right)$ then $\lambda_{\max}(\mathbf{M}^{-1}) \geq \frac{4gv(1-\rho)}{b(v+1)}$ and the equality is satisfied if and only if $\mathbf{X}\mathbf{1}_v = \frac{v-1}{2}\mathbf{1}_b$ or $\mathbf{X}\mathbf{1}_v = \frac{v+1}{2}\mathbf{1}_b$,*
- **(ii)** *if $\rho \in \left(\frac{v(v-3)}{b(v+1)+v(v-3)},\ \frac{v}{b+v}\right)$ then $\lambda_{\max}(\mathbf{M}^{-1}) \geq \frac{4gv(1-\rho)}{b(v+1)}$ and the equality is satisfied if and only if $\mathbf{X}\mathbf{1}_v = \frac{v+1}{2}\mathbf{1}_b$,*
- **(iii)** *if $\rho \in \left(\frac{v}{b+v},\ 1\right)$ then $\lambda_{\max}(\mathbf{M}^{-1}) > \frac{g(1+\rho(b-1)}{bv}$.*

Proof: The proof is similar to given in Theorem 2.1 one. □

Now, we can formulate the definition of the regular E-optimal spring balance weighing design. So, we have

Definition 2.1. Any nonsingular spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$ is regular E-optimal if the eigenvalues of the information matrix attains the bounds of Theorems 2.1 and 2.2, i.e.

- **(i)** v is even and $\rho \in \left(\frac{-1}{b-1},\ \frac{v-2}{b+v-2)}\right)$ if $\lambda_{\max}(\mathbf{M}^{-1}) = \frac{4g(v-1)(1-\rho)}{bv}$,
- **(ii)** v is odd and $\rho \in \left(\frac{-1}{b-1},\ \frac{v}{b+v}\right)$ if $\lambda_{\max}(\mathbf{M}^{-1}) = \frac{4gv(1-\rho)}{b(v+1)}$.

A direct consequence of above considerations is

Theorem 2.3. *Any nonsingular spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ with the variance matrix of errors $\sigma^2\mathbf{G}$ is regular E-optimal design if and only if*

- **(1)** *v is even and $\rho \in \left(\frac{-1}{b-1},\ \frac{v-2}{b+v-2)}\right)$*
  $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{1}{g(1-\rho)}\left[\frac{bv}{4(v-1)}\mathbf{I}_v + \frac{b(v-2)}{4(v-1)}\mathbf{1}_v\mathbf{1}_v' - \frac{\rho b^2}{4(1+\rho(b-1))}\mathbf{1}_v\mathbf{1}_v'\right]$ and
  $\mathbf{X}\mathbf{1}_v = \frac{v}{2}\mathbf{1}_b$,

(**2**) *p is odd and*

(**2.1**) $\rho \in \left(\frac{-1}{b-1},\ \frac{v(v-3)}{b(v+1)+v(v-3)}\right)$

(**2.1.1**) $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{1}{g(1-\rho)}\left[\frac{b(v+1)}{4v}\mathbf{I}_v + \frac{b(v-3)}{4v}\mathbf{1}_v\mathbf{1}_v' - \frac{\rho b^2(v-1)^2}{4v^2(1+\rho(b-1))}\mathbf{1}_v\mathbf{1}_v'\right]$
and $\mathbf{X}\mathbf{1}_v = \frac{v-1}{2}\mathbf{1}_b$, or

(**2.1.2**) $\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{1}{g(1-\rho)}\left[\frac{b(v+1)}{4v}\mathbf{I}_v + \frac{b(v+1)}{4v}\mathbf{1}_v\mathbf{1}_v' - \frac{\rho b^2(v+1)^2}{4v^2(1+\rho(b-1))}\mathbf{1}_v\mathbf{1}_v'\right]$
and $\mathbf{X}\mathbf{1}_v = \frac{v+1}{2}\mathbf{1}_b$,

(**2.2**) $\rho \in \left(\frac{v(v-3)}{b(v+1)+v(v-3)},\ \frac{v}{b+v}\right)$
$\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} = \frac{1}{g(1-\rho)}\left[\frac{b(v+1)}{4v}\mathbf{I}_v + \frac{b(v+1)}{4v}\mathbf{1}_v\mathbf{1}_v' - \frac{\rho b^2(v+1)^2}{4v^2(1+\rho(b-1))}\mathbf{1}_v\mathbf{1}_v'\right]$
and $\mathbf{X}\mathbf{1}_v = \frac{v+1}{2}\mathbf{1}_b$.

Proof: Since the proofs for even and odd v are similar, we give the proof only for the case of even v and $\rho \in \left(\frac{-1}{b-1},\ \frac{v-2}{b+v-2}\right)$. Notice, that $\lambda_{\max}(\mathbf{M}^{-1})$ attains the lowest bound in Theorem 2.1(i) if equalities $\lambda_{\max}(\mathbf{M}^{-1}) = \lambda_{\max}(\bar{\mathbf{M}}^{-1})$ and $v = \frac{k}{2}$ hold. It follows easily that $\mathrm{tr}(\mathbf{M}) = \frac{bv(2+\rho(b-2))}{4g(1-\rho)(1+\rho(b-1))}$ and $\mathbf{1}_v'\mathbf{M}\mathbf{1}_v = \frac{bv^2}{4g(1+\rho(b-1))}$. We apply formulas on μ_1 and μ_2 to give the form of $\mathbf{M}$. Thus the Theorem. □

It can be noted that $g = 1$, $\rho = 0$ and $\mathbf{G} = \mathbf{I}_b$, we become the equalities given by Jacroux and Notz ([8]).

Theorem 2.4. *In any nonsingular spring balance weighing design* $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$, *if*

(**i**) *v is even and* $\rho \in \left(\frac{v-2}{b+v-2},\ 1\right)$ *or*

(**ii**) *v is odd and* $\rho \in \left(\frac{v}{b+v},\ 1\right)$,

then regular E-optimal spring balance weighing design does not exist.

Proof: Since the proofs for even v and odd v are similar, we shall only give the proof for the case odd v. If $\rho \in \left(\frac{v}{b+v},\ 1\right)$ then the lowest bound of the maximal eigenvalue of the design matrix $\mathbf{X}$ is given in Theorem 2.2(iii). The lowest bound is attained if and only if $k = v$. It means $\mathbf{X} = \mathbf{1}_b\mathbf{1}_v'$. Such matrix is singular one. Thus the regular E-optimal design does not exist. □

Theorem 2.5. *Any nonsingular spring balance weighing design* $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$ *for* $\rho = \frac{p(k-1)}{n(p-k)+p(k-1)}$, $k = 1,2,...,\frac{v}{2}$ *for even* v *or* $k = 1,2,...,\frac{v+1}{2}$ *for odd* v *is regular E-optimal design if and only if*

$$\mathbf{X}'\mathbf{X} = \left(\frac{nk}{p} - \frac{nk(k-1)}{p(p-1)}\right)\mathbf{I}_v + \frac{nk(k-1)}{p(p-1)}\mathbf{1}_v'\mathbf{1}_v\ .$$

Proof: Let us take into consideration the case given in Lemma 2.1(ii). $\lambda_1 = \lambda_2 = \lambda = \frac{1}{\mu}$ if and only if $\rho = \frac{\mathbf{k}'\mathbf{k} - \mathbf{1}_b'\mathbf{k}}{(b-1)(\mathbf{1}_b'\mathbf{k} - \mathbf{k}'\mathbf{k}) + (\mathbf{1}_b'\mathbf{k})^2 - \mathbf{r}'\mathbf{r}}$. Now, we have to consider two cases v is even and v is odd. For both v the proofs are similar, so we give the case for even v, only. According to the proof of Theorem 2.1, the equality in 2.2 holds if and only if $k_1 = k_2 = \cdots = k_b = k$ and $r_1 = r_2 = \cdots = r_v = r$, for $k = 1, 2, ..., \frac{v}{2}$. Hence $\rho = \frac{p(k-1)}{n(p-k)+p(k-1)}$. From the above results $\bar{\mathbf{M}} = \mu\mathbf{I}_v$. Thus $\mu\mathbf{I}_v = \frac{1}{g(1-\rho)}\left(\mathbf{X}'\mathbf{X} - \frac{\rho n^2 k^2}{v^2(1+\rho(b-1))}\mathbf{1}_v\mathbf{1}_v'\right)$ and $\mathbf{X}'\mathbf{X} = \frac{nk(p-k)}{p(p-1)}\mathbf{I} + \frac{nk(k-1)}{p(p-1)}\mathbf{1}_v\mathbf{1}_v'$. Putting ρ and $\mu = \frac{nk(p-k)}{g(1-\rho)p(p-1)}$ we obtain the form of the matrix $\mathbf{X}'\mathbf{X}$. If $k > \frac{v}{2}$, then from Theorem 2.4 regular E-optimal spring balance weighing design does not exist. □

3. CONSTRUCTION OF THE REGULAR E-OPTIMAL DESIGN

For the construction of the regular E-optimal spring balance weighing design, from all possible block designs, we choose the incidence matrices of the balanced incomplete block designs and group divisible designs. The definitions of these block designs are given in Raghavarao and Padgett ([14]).

Theorem 3.1. *Let* $\mathbf{N}$ *be the incidence matrix of the balanced incomplete block design with the parameters*

(**i**) $v = 2t,\ b = 2(2t-1),\ r = 2t-1,\ k = t,\ \lambda = t-1$ *or*

(**ii**) $v = 2t,\ b = \binom{2t}{t},\ r = \binom{2t-1}{t-1},\ k = t,\ \lambda = \binom{2(t-1)}{t-2}$,

$t = 2, 3, \ldots$. *Then, any* $\mathbf{X} = \mathbf{N}' \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$ *for* $\rho \in \left(\frac{-1}{b-1},\ \frac{v-2}{b+v-2}\right)$ *is the regular E-optimal spring balance weighing design.*

Proof: An easy computation shows that the matrix $\mathbf{X} = \mathbf{N}'$ satisfies (1) of Theorem 2.3. □

Now, let

$$\mathbf{X} = \begin{bmatrix} \mathbf{N}_1' \\ \mathbf{N}_2' \end{bmatrix}, \tag{3.1}$$

where $\mathbf{N}_u$ is the incidence matrix of the group divisible design with the same association scheme with the parameters $v,\ b_u,\ r_u,\ k = \frac{v}{2},\ \lambda_{1u},\ \lambda_{2u},\ u = 1, 2$. Furthermore, let the condition

$$\lambda_{11} + \lambda_{12} = \lambda_{21} + \lambda_{22} = \lambda \tag{3.2}$$

be satisfied. For $\mathbf{X}$ in 3.1, $b = b_1 + b_2$. The limitation on the t and s given in next Theorem 3.2 follow from the restrictions: $r,\ k \leq 10$ given in Clatworthy ([5]).

Theorem 3.2. *Let* $\mathbf{N}_u,\ u = 1, 2,$ *be the incidence matrix of the group divisible block design with the same association scheme and with the parameters*

(**1**) $v = 4,\ k = 2$ *and*

- (**1.1**) $b_1 = 2(3t+1),\ r_1 = 3t+1,\ \lambda_{11} = t+1,\ \lambda_{21} = t, t = 1, 2, 3,$ *and* $b_2 = 2(3s+2),\ r_2 = 3s+2,\ \lambda_{12} = s,\ \lambda_{22} = s+1,\ s = 0, 1, 2,$
- (**1.2**) $b_1 = 2(3t+2),\ r_1 = 3t+2,\ \lambda_{11} = t+2,\ \lambda_{21} = t,\ t = 1, 2,$ *and* $b_2 = 2(3s+4),\ r_2 = 3s+4,\ \lambda_{12} = s,\ \lambda_{22} = s+2,\ s = 0, 1, 2,$
- (**1.3**) $b_1 = 2(t+3),\ r_1 = t+3,\ \lambda_{11} = t+1,\ \lambda_{21} = 1$ *and* $b_2 = 4t,$ $r_2 = 2t,\ \lambda_{12} = 0,\ \lambda_{22} = t,\ t = 1, 2, ..., 5,$
- (**1.4**) $b_1 = 16,\ r_1 = 8,\ \lambda_{11} = 0,\ \lambda_{21} = 4$ *and* $b_2 = 2(3s+4),\ r_2 = 3s+4,$ $\lambda_{12} = s+4,\ \lambda_{22} = s,\ s = 1, 2,$
- (**1.5**) $b_1 = 18,\ r_1 = 9,\ \lambda_{11} = 5,\ \lambda_{21} = 2$ *and* $b_2 = 6(s+2),\ r_2 = 3(s+2),$ $\lambda_{12} = s,\ \lambda_{22} = s+3,\ s = 0, 1,$

(**2**) $v = 6,\ k = 3$ *and*

- (**2.1**) $b_1 = 4t,\ r_1 = 2t,\ \lambda_{11} = 0,\ \lambda_{21} = t$ *and* $b_2 = 6t,\ r_2 = 3t,\ \lambda_{12} = 2t,$ $\lambda_{22} = t,\ t = 1, 2, 3,$
- (**2.2**) $b_1 = 2(2t+5),\ r_1 = 2t+5,\ \lambda_{11} = t+1,\ \lambda_{21} = t+2$ *and* $b_2 = 6t,$ $r_2 = 3t,\ \lambda_{12} = t+1,\ \lambda_{22} = t,\ t = 1, 2,$
- (**2.3**) $b_1 = 12,\ r_1 = 6,\ \lambda_{11} = 4,\ \lambda_{21} = 2$ *and* $b_2 = 2(5s+4),\ r_2 = 5s+4,$ $\lambda_{12} = 2s,\ \lambda_{22} = 2(s+1),\ s = 0, 1,$
- (**2.4**) $b_1 = 16,\ r_1 = 8,\ \lambda_{11} = 4,\ \lambda_{21} = 3$ *and* $b_2 = 2(5s+2),\ r_2 = 5s+2,$ $\lambda_{12} = 2s,\ \lambda_{22} = 2s+1,\ s = 0, 1,$

(**3**) $v = 8,\ k = 4$ *and*

- (**3.1**) $b_1 = 4(t+1),\ r_1 = 2(t+1),\ \lambda_{11} = 0,\ \lambda_{21} = t+1$ *and* $b_2 = 4(6-t),$ $r_2 = 2(6-t),\ \lambda_{12} = 6,\ \lambda_{22} = 5-t,\ t = 1, 2, 3,$
- (**3.2**) $b_1 = 2(3t+2),\ r_1 = 3t+2,\ \lambda_{11} = t+2,\ \lambda_{21} = t+1$ *and* $b_2 = 6(4-t),\ r_2 = 3(4-t),\ \lambda_{12} = 4-t,\ \lambda_{22} = 5-t,\ t = 1, 2,$

(**4**) $v = 10,\ k = 5$ *and* $b_1 = 8t,\ r_1 = 4t,\ \lambda_{11} = 0,\ \lambda_{21} = 2t$ *and* $b_2 = 10t,$ $r_2 = 5t,\ \lambda_{12} = 4t,\ \lambda_{22} = 2t,\ t = 1, 2,$

(**5**) $v = 2(2t+1),\ k = 2t+1$ *and* $b_1 = 4t,\ r_1 = 2t,\ \lambda_{11} = 0,\ \lambda_{21} = t$ *and* $b_2 = 2(2t+1),\ r_2 = 2t+1,\ \lambda_{12} = 2t,\ \lambda_{22} = t,\ t = 1, 2, 3, 4,$

(**6**) $v = 4(t+1),\ k = 2(t+1)$ *and* $b_1 = 2(2t+1),\ r_1 = 2t+1,\ \lambda_{11} = 2t+1,$ $\lambda_{21} = t$ *and* $b_2 = 4(t+1), r_2 = 2(t+1), \lambda_{12} = 0, \lambda_{22} = t+1, t = 1, 2, 3, 4.$

Then any $\mathbf{X} \in \mathbf{\Phi}_{b\times v}(0, 1)$ *in the form 3.1 with the variance matrix of errors* $\sigma^2\mathbf{G}$ *for* $\rho \in \left(\frac{-1}{b-1},\ \frac{v-2}{b+v-2}\right)$ *is the regular E-optimal spring balance weighing design.*

Proof: This is proved by checking that the matrix $\mathbf{X}$ in 3.1 satisfies (1) of Theorem 2.3. □

Theorem 3.3. *Let* $\mathbf{N}$ *be the incidence matrix of balanced incomplete block design with the parameters*

(**i**) $v = 2t+1,\ b = 2(2t+1),\ r = 2(t+1),\ k = t+1,\ \lambda = t+1,$

(**ii**) $v = b = 4t^2-1,\ r = k = 2t^2,\ \lambda = t^2,$

(**iii**) $v = b = 8t+7,\ r = k = 4(t+1),\ \lambda = 2(t+1),$

(**iv**) $v = b = 4t-1,\ r = k = 2t,\ \lambda = t,$

(**v**) $v = 4t+1,\ b = 2(4t+1),\ r = 2(2t+1),\ k = \lambda = 2t+1,$

(**vi**) $v = 2t+1,\ b = \binom{2t+1}{t+1},\ r = \binom{2t}{t},\ k = t+1,\ \lambda = \binom{2t-1}{t-1},$

$t = 1, 2, \ldots$. *Then any* $\mathbf{X} = \mathbf{N}' \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$ *for* $\rho \in \left(\frac{-1}{b-1},\ \frac{v}{b+v}\right)$ *is the regular E-optimal spring balance weighing design.*

Proof: For $\mathbf{X} = \mathbf{N}' \in \mathbf{\Phi}_{b\times v}(0,1)$ the equalities (2.1.2) and (2.2) of Theorem 2.3 are satisfied. □

Theorem 3.4. *Let* $\mathbf{N}$ *be the incidence matrix of balanced incomplete block design with the parameters*

(**i**) $v = 2t+1,\ b = 2(2t+1),\ r = 2t,\ k = t,\ \lambda = t-1,\ t = 2, 3, \ldots$

(**ii**) $v = b = 4t^2-1,\ r = k = 2t^2-1,\ \lambda = t^2-1,\ t = 2, 3, \ldots$

(**iii**) $v = b = 8t+7,\ r = k = 4t+3,\ \lambda = 2t+1,\ t = 1, 2, \ldots$

(**iv**) $v = b = 4t-1,\ r = k = 2t-1,\ \lambda = t-1,\ t = 2, 3, \ldots$

(**v**) $v = 4t+1,\ b = 2(4t+1),\ r = 4t,\ k = 2t,\ \lambda = 2t-1,\ t = 1, 2, \ldots$

(**vi**) $v = 2t+1,\ b = \binom{2t+1}{t+1},\ r = \binom{2t}{t-1},\ k = t,\ \lambda = \binom{2t-1}{t-2},\ t = 2, 3, \ldots$

Then any $\mathbf{X} = \mathbf{N}' \in \mathbf{\Phi}_{b\times v}(0,1)$ *with the variance matrix of errors* $\sigma^2\mathbf{G}$ *for* $\rho \in \left(\frac{-1}{b-1},\ \frac{v(v-3)}{b(v+1)+v(v-3)}\right)$ *is the regular E-optimal spring balance weighing design.*

Proof: It is obvious that for $\mathbf{X} = \mathbf{N}' \in \mathbf{\Phi}_{b\times v}(0,1)$ the equality (2.1.1) given in Theorem 2.3 is fulfilled. □

REFERENCES

[1] Banerjee, K.S. (1975). *Weighing Designs for Chemistry, Medicine, Economics, Operations Research, Statistics*, Marcel Dekker Inc., New York.

[2] Ceranka, B. and Graczyk, M. (2004). A-optimal chemical balance weighing design, *Folia Facultatis Scientiarum Naturalium Universitatis Masarykianae Brunensis, Mathematica*, **15**, 41–54.

[3] Ceranka, B.; Graczyk, M. and Katulska, K. (2006). A-optimal chemical balance weighing design with nonhomogeneity of variances of errors, *Statistics and Probability Letters*, **76**, 653–665.

[4] Ceranka, B.; Graczyk, M. and Katulska, K. (2007). On certain A-optimal chemical balance weighing designs, *Computational Statistics and Data Analysis*, **51**, 5821–5827.

[5] Clatworthy, W.H. (1973). *Tables of Two-Associate-Class Partially Balanced Design*, NBS Applied Mathematics Series 63.

[6] Graczyk, M. (2011). A-optimal biased spring balance weighing design, *Kybernetika*, **47**, 893–901.

[7] Graczyk, M. (2012). Notes about A-optimal spring balance weighing design, *Journal of Statistical Planning and Inference*, **142**, 781–784.

[8] Jacroux, M. and Notz, W. (1983). On the optimality of spring balance weighing designs, *The Annals of Statistics*, **11**, 970–978.

[9] Katulska, K. and Rychlińska, E. (2010). On regular E-optimality of spring balance weighing designs, *Colloquium Biometricum*, **40**, 165–176.

[10] Masaro, J. and Wong, C.S. (2008). Robustness of A-optimal designs, *Linear Algebra and its Applications*, **429**, 1392–1408.

[11] Neubauer, G.N. and Watkins, W. (2002). E-optimal spring balance weighing designs for $n \equiv 3 \pmod 4$ objects, *SIAM J. Anal. Appl.*, **24**, 91–105.

[12] Pukelsheim, F. (1983). *Optimal Design of Experiment*, John Wiley and Sons, New York.

[13] Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Designs of Experiment*, John Wiley Inc., New York.

[14] Raghavarao, D. and Padgett, L.V. (2005). *Block Designs, Analysis, Combinatorics and Applications*, Series of Applied Mathematics 17, Word Scientific Publishing Co. Pte. Ltd.

[15] Shah, K.R. and Sinha, B.K. (1989). *Theory of Optimal Designs*, Springer-Verlag, Berlin, Heidelberg.

REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 131–144

CONTROL CHARTS FOR MULTIVARIATE NONLINEAR TIME SERIES

Authors: Robert Garthoff
– Department of Statistics, European University,
Große Scharrnstr. 59, 15230 Frankfurt (Oder), Germany
Garthoff@euv-frankfurt-o.de

Iryna Okhrin
– Department of Statistics, European University,
Große Scharrnstr. 59, 15230 Frankfurt (Oder), Germany
irokhrin@euv-frankfurt-o.de

Wolfgang Schmid
– Department of Statistics, European University,
Große Scharrnstr. 59, 15230 Frankfurt (Oder), Germany
wschmid@euv-frankfurt-o.de

Received: October 2013 Accepted: November 2013

Abstract:

- In this paper control charts for the simultaneous monitoring of the means and the variances of multivariate nonlinear time series are introduced. The underlying target process is assumed to be a constant conditional correlation process (cf. [3]). The new schemes make use of local measures of the means and the variances based on current observations, conditional moments, or residuals. Exponential smoothing and cumulative sums are applied to these characteristic quantities. Distances between these quantities and target values are measured by the Mahalanobis distance. The introduced schemes are compared via a simulation study. As a measure of performance the average run length is used.

Key-Words:

- *statistical process control; multivariate CUSUM charts; multivariate EWMA charts; conditional correlation model.*

AMS Subject Classification:

- 62L10, 62M10, 62P20, 91B84.

1. INTRODUCTION

Nowadays, the methods of statistical process control (SPC) are frequently used in order to detect changes in model parameters of multivariate time series. Many applications can be found in various scientific fields, e.g. engineering, economics, medicine, and environmental sciences. The main idea of SPC is to detect deviations of an observed process from a predefined target process as soon as possible after their occurrence. The most important tools of SPC are control charts. A control chart consists of the control statistic and the control limits. The data are sequentially examined. If at a certain point of time the control statistic lies within the control limits, we conclude the process is still in control and the procedure continues at the next point of time. If the control statistic exceeds the control limit, the algorithm stops and the process is considered to be out of control.

In current literature on SPC the underlying multivariate process is assumed to consist of independent random vectors and the parameter of interest is chosen to be the mean vector of the process (cf. [4, 14, 12, 13, 8]). Mean charts for multivariate time series are considered in [17, 11, 1]. In [10] multivariate control charts for nonlinear autocorrelated processes are introduced using the support vector regression approach.

The monitoring of the covariance matrix of multivariate time series is discussed only in a few papers. In [15, 16] several types of exponentially weighted moving average (EWMA) charts are proposed. The underlying process is assumed to be either a multivariate Gaussian process or a multivariate GARCH process in the sense of [5].

However, in the present paper the aim is to jointly monitor the means and the variances of multivariate time series. The target process is assumed to be a constant conditional correlation (CCC) model (cf. [3]). A CCC process is a multivariate nonlinear time series turning out to be quite attractive for practical purposes. Although the amount of parameters is not too high, the model is still sufficiently flexible.

Subsequently, we introduce several new control charts. They are based on combining local measures for the means and the variances of the target process or the residual process, e.g. current observations and conditional variances with an EWMA recursion or a cumulative sum (CUSUM). In order to avoid the curse of dimensionality variances are monitored using the squared Euclidean distance of present observations and residuals from their in-control mean as well as the trace of the conditional covariance matrix. These schemes seem to be very useful even for higher dimensions. Via an extensive Monte Carlo study the charts are compared with schemes proposed in [6]. In order to assess the performance of the schemes the average run length (ARL) is applied.

The paper is structured as follows. In Sect. 2 we describe the CCC model and the modeling of the out-of-control process. In Sect. 3 characteristic quantities are presented. Their in-control means as well as their in-control covariances are derived. In Sect. 4 multivariate EWMA and CUSUM type schemes based on the former characteristic quantities are described. In Sect. 5 the results of the simulation study are presented. Finally, we draw a conclusion in Sect. 6.

2. THE MODEL

Below, the target process is denoted by $\{Y_t\}$ and the observed process by $\{X_t\}$. Next, we describe the modeling of both processes.

2.1. Modeling dynamics of the conditional covariance matrix

The p-dimensional target process $\{Y_t\}$ assumed to be a nonlinear CCC model is given by

$$Y_t \,=\, \mu + \Sigma_t^{1/2} \varepsilon_t \;. \tag{2.1}$$

Hence, ε_t is assumed to be an independent and normally distributed random sequence with zero mean and a covariance matrix equal to the identity matrix. Further, μ denotes the overall mean vector which is independent of time. The matrix $\Sigma_t = \mathrm{Cov}(Y_t \,|\, \mathcal{I}_{t-1}) = E_{t-1}\big[(Y_t - \mu)\,(Y_t - \mu)'\big]$ denotes the conditional covariance matrix of $\{Y_t\}$ conditioned on the information set $\mathcal{I}_{t-1}$. $\mathcal{I}_{t-1}$ is equal to the smallest σ-algebra generated by $Y_{t-1}, Y_{t-2}, \ldots$. As a consequence,

$$Y_t \,|\, \mathcal{I}_{t-1} \;\sim\; \mathcal{N}_p\,(\mu, \Sigma_t) \;, \tag{2.2}$$

i.e. the conditional distribution of $Y_t = (Y_{1t}, ..., Y_{pt})'$ is a normal distribution.

The CCC model is introduced in [3] where the conditional correlation matrix is assumed to be time invariant. The conditional covariance matrix of Y_t is given by

$$\Sigma_t \;=\; \mathrm{D}_t \,\mathrm{R}\, \mathrm{D}_t \;=\big(\sigma_{it}\,\varrho_{ij}\,\sigma_{jt}\big)_{i,j=1,\ldots,p} \tag{2.3}$$

with $\sigma_{it}^2 = \mathrm{Var}(Y_{it} \,|\, \mathcal{I}_{t-1})$ for $i = 1, ..., p$ and $\mathrm{D}_t = \mathrm{diag}(\sigma_{1t}, ..., \sigma_{pt})$. The conditional variances of Y_{it} usually follow a GARCH model (cf. [2])

$$\sigma_{it}^2 \;=\; \omega_i + \sum_{m=1}^{M_i} \alpha_{im}\,(Y_{i,t-m} - \mu_i)^2 + \sum_{n=1}^{N_i} \beta_{in}\,\sigma_{i,t-n}^2 \qquad \forall\, i \in \{1, ..., p\} \;. \tag{2.4}$$

The conditional correlation matrix $\mathrm{R} = (\varrho_{ij})_{i,j=1,...,p}$ of Y_t is assumed to be time invariant and positive definite.

A unique weakly stationary solution of (2.1), strictly stationary and ergodic as well, exists if the polynomials fulfill the condition

$$1 - \sum_{m=1}^{M_i} \alpha_{im} z^m - \sum_{n=1}^{N_i} \beta_{in} z^n \neq 0 \qquad \text{for } |z| \leq 1 \tag{2.5}$$

with $i = 1, ..., p$ (cf. [9]). Moreover, $E(Y_t) = \mu$ and

$$\mathrm{Var}(Y_{it}) = \frac{\omega_i}{1 - \sum\limits_{m=1}^{M_i} \alpha_{im} - \sum\limits_{n=1}^{N_i} \beta_{in}} \qquad \forall\, i \in \{1, ..., p\} \; . \tag{2.6}$$

2.2. Modeling the out-of-control behavior

We are faced with a sequential problem where we check at each point of time t whether a shift in mean or variances has occurred or not. The respective decision problem is therefore

$$\begin{array}{lcl} H_{0,t} : E(X_t) = \mu & \wedge & \mathrm{Cov}(X_t) = \Sigma \\ & \text{against} & \\ H_{1,t} : E(X_t) \neq \mu & \vee & \mathrm{Cov}(X_t) \neq \Sigma \end{array} \tag{2.7}$$

where $\Sigma = \mathrm{Cov}(Y_t)$. The relationship between the target process $\{Y_t\}$ and the observed process $\{X_t\}$ is given by

$$X_t = \begin{cases} Y_t & \text{for } t < \tau \\ \mu + a + \Delta(Y_t - \mu) & \text{for } t \geq \tau \end{cases} . \tag{2.8}$$

The parameters $a \in \mathbb{R}^p \backslash \{0\}$ and $\Delta = \mathrm{diag}(d_1, d_2, ..., d_p) \neq \mathrm{I}_p$, where I_p denotes the identity matrix of dimension $p \times p$, are unknown. Here we focus on the detection of increases in variances. For that reason we assume that $d_i \geq 1$ for $i = 1, ..., p$. The position of the change point is denoted by $\tau \in \mathbb{N} \cup \{\infty\}$. If a change is present, i.e. $\tau < \infty$, the process is said to be out of control. Hence, changes in mean or in variances can be observed. On the contrary, $\tau = \infty$ means that the change point never occurs and therefore the process is in control.

3. CHARACTERISTIC QUANTITIES

In order to monitor the means and the variances of a multivariate process we need several local measures for these characteristics. We reduce the number of

characteristic quantities because we monitor the sum of variances. Since we are exclusively interested in detecting increases of variances, the following procedures can be applied. Below, the characteristic quantities are denoted by T_t. We derive several properties of these quantities in this section (see Propositions 3.1 to 3.3). In order to shorten the paper we do not present the proofs here. They can be found in [7]. Further, the notation $\mu = (\mu_i)_{i=1,...,p}$, $\Sigma_t = (\sigma_{ij,t})$, $\Sigma = (\sigma_{ij})$, and $\sigma_i^2 = \sigma_{ii} = \mathrm{Var}(Y_{it})$, $i = 1, ..., p$, is used.

3.1. Characteristics based on current observations and residuals

As already mentioned, current observations are local measures for the mean vector. To monitor the covariance matrix we can use the squared Euclidean distance between X_t and μ. This leads to

$$T_t^{(1)} = \begin{pmatrix} X_t - \mu \\ (X_t - \mu)' (X_t - \mu) \end{pmatrix}. \tag{3.1}$$

Further, the in-control mean vector and the in-control covariance matrix have to be derived. If $\{Y_t\}$ is a weakly stationary process with the mean μ and the covariance matrix $\mathrm{Var}(Y_t) = \Sigma = (\sigma_{ij})$,

$$E\big(T_t^{(1)}\big) = \begin{pmatrix} 0 \\ \sum_{i=1}^{p} \sigma_i^2 \end{pmatrix} \quad \text{for } t < \tau \qquad \text{and} \qquad E\big(T_t^{(1)}\big) = \begin{pmatrix} a \\ \sum_{i=1}^{p} d_i^2 \sigma_i^2 \end{pmatrix} \quad \text{for } t \geq \tau \ .$$

Apparently, the quantity $T_t^{(1)}$ reflects changes in the mean and the variances of $\{Y_t\}$ but no changes in the covariances of $\{Y_t\}$. If values smaller than 1 are permitted for d_i, values larger and smaller than 1 might overlap. As a consequence, a change in $\sum_{i=1}^{p} d_i^2 \sigma_i^2$ is not observed. Since we only consider increases in variances, we avoid this problem. Next, the underlying target process is a CCC process.

Proposition 3.1. *Assume that (2.1) and (2.2) hold and that $E(Y_{it}^4) < \infty$ for all i and t. Then*

$$\mathrm{Cov}_{\tau=\infty}\big(T_t^{(1)}\big) = \begin{pmatrix} \Sigma & 0 \\ 0' & a_{1t} \end{pmatrix} \quad \textit{and} \quad \mathrm{Cov}_{\tau=\infty}\big(T_s^{(1)}, T_t^{(1)}\big) = \begin{pmatrix} \mathrm{O}_{p\times p} & 0 \\ c_{1,st} & b_{1,st} \end{pmatrix}$$

for $s < t$ where

$$\begin{aligned} a_{1t} &= \sum_{i=1}^{p} \sum_{j=1}^{p} \Big[E_{\tau=\infty}\big(\sigma_{it}^2 \sigma_{jt}^2 + 2\,\sigma_{ij,t}^2\big) - \sigma_i^2 \sigma_j^2 \Big], \\ b_{1,st} &= \sum_{i=1}^{p} \sum_{j=1}^{p} \Big[E_{\tau=\infty}\big(X_{is} - \mu_i\big)^2 \sigma_{jt}^2 - \sigma_i^2 \sigma_j^2 \Big] \quad \textit{and} \\ c_{1,st} &= E_{\tau=\infty} \left[(X_s - \mu) \sum_{i=1}^{p} \sigma_{it}^2 \right]. \end{aligned}$$

These quantities can be explicitly calculated only for less complex processes, otherwise they have to be estimated via a simulation study. However, in order to apply these results the underlying process has to be strictly stationary. Then these quantities do not depend on t. Note that $\mathrm{Cov}_{\tau=\infty}\big(T_s^{(1)}, T_t^{(1)}\big) = \Big[\mathrm{Cov}_{\tau=\infty}\big(T_t^{(1)}, T_s^{(1)}\big)\Big]'$ so that the covariances can be computed for $s > t$ as well.

Another simple characteristic quantity is based on the transformed observed process $\eta_t = (\eta_{it})_{i=1,\ldots,p}$. In this case the respective mean and the covariance matrix of the residual vector are monitored. The residual vector is given by

$$\eta_t \;=\; \Sigma_t^{-1/2}(X_t - \mu) \;=\; \begin{cases} \varepsilon_t\,, & t < \tau\,, \\ \Sigma_t^{-1/2} a + \Sigma_t^{-1/2} \Delta \Sigma_t^{1/2}\, \varepsilon_t\,, & t \geq \tau\,. \end{cases} \tag{3.2}$$

Note that $\eta_t \,|\, \Sigma_t \sim \mathcal{N}_p\left(\Sigma_t^{-1/2} a\,,\, \Sigma_t^{-1/2} \Delta \Sigma_t\, \Delta \Sigma_t^{-1/2}\right)$. Thus, $E\big(\eta_t\big) = E\big(\Sigma_t^{-1/2}\big)\, a\,.$

The characteristic quantity based on residuals is given by

$$T_t^{(2)} = \begin{pmatrix} \eta_t \\ \eta_t' \eta_t \end{pmatrix}. \tag{3.3}$$

In the following proposition we compute the in-control mean and the in-control covariance matrix.

Proposition 3.2. *Assume that (2.1) and (2.2) hold and that* R *is positive definite. Then*

$$E_{\tau=\infty}\big(T_t{}^{(2)}\big) \;=\; \begin{pmatrix} 0 \\ p \end{pmatrix}$$

as well as

$$\mathrm{Cov}_{\tau=\infty}\big(T_t^{(2)}\big) \;=\; \begin{pmatrix} \mathrm{I} & 0 \\ 0' & 2p \end{pmatrix} \quad \textit{and} \quad \mathrm{Cov}_{\tau=\infty}\big(T_s^{(2)}, T_t^{(2)}\big) \;=\; \mathrm{O}_{(p+1)\times(p+1)}$$

for $s \neq t$.

The application of control charts to residuals is much easier than the application to the original process, because η_t is independent and normally distributed with zero mean and a covariance matrix equal to the identity matrix as long as the process is in control. In the out-of-control state the process η_t is neither independent nor identically distributed. In Proposition 3.2 only the existence of the first two moments of the target process is required while in Proposition 3.1 the first four moments are needed.

3.2. Characteristics based on the conditional variances

Regarding characteristics based on conditional variances we compute the trace of Σ_t at each point of time t. The characteristic quantity referring to the

trace of the conditional covariance matrix is given by

$$T_t^{(3)} = \begin{pmatrix} X_t - \mu \\ \operatorname{tr}(\Sigma_t) \end{pmatrix} = \begin{pmatrix} X_t - \mu \\ \sum_{i=1}^{p} \sigma_{it}^2 \end{pmatrix} . \tag{3.4}$$

Note that $E_{\tau=\infty}\big(\sum_{i=1}^{p} \sigma_{it}^2\big) = \sum_{i=1}^{p} E\big[\operatorname{Var}\big(Y_{it} \,|\, \mathcal{I}_{t-1}\big)\big] = \sum_{i=1}^{p} \sigma_i^2$. Thus, the quantity is able to detect changes in variances of the target process. The derivation of the in-control mean vector and the in-control covariance matrix of $T_t^{(3)}$ is straight forward.

Proposition 3.3. *Assume that (2.1) and (2.2) hold and that* R *is positive definite then*

$$E_{\tau=\infty}\big(T_t^{(3)}\big) = \begin{pmatrix} 0 \\ \sum_{i=1}^{p} \sigma_i^2 \end{pmatrix} .$$

If additionally $E(Y_{it}^4) < \infty$ *for all* i *and* t,

$$\operatorname{Cov}_{\tau=\infty}\big(T_t^{(3)}\big) = \begin{pmatrix} \Sigma & 0 \\ 0' & a_{3t} \end{pmatrix} \quad \text{and} \quad \operatorname{Cov}_{\tau=\infty}\big(T_s^{(3)}, T_t^{(3)}\big) = \begin{pmatrix} \mathrm{O}_{p\times p} & 0 \\ c_{3,st} & b_{3,st} \end{pmatrix}$$

for $s < t$, *where*

$$a_{3t} = \sum_{i=1}^{p}\sum_{j=1}^{p} \Big[E_{\tau=\infty}\big(\sigma_{it}^2\sigma_{jt}^2\big) - \sigma_i^2\sigma_j^2\Big] ,$$

$$b_{3,st} = \sum_{i=1}^{p}\sum_{j=1}^{p} \Big[E_{\tau=\infty}\big(\sigma_{is}^2\sigma_{jt}^2\big) - \sigma_i^2\sigma_j^2\Big]$$

and $c_{3,st} = c_{1,st}$.

As in the case of present observations these quantities can be explicitly determined only for special cases. For more general processes they can be estimated using a simulation study if the underlying target process is strictly stationary.

4. CONTROL SCHEMES FOR MULTIVARIATE TIME SERIES

In this section we propose several new control charts. They are obtained applying univariate or multivariate EWMA recursions and cumulative sums to the characteristic quantities considered in Sect. 3. The control statistics are based on the Mahalanobis distance between the weighted characteristics and the corresponding in-control means.

Since these control statistics are distance measures, the charts are one-sided. Therefore, the charts give a signal if the control statistic exceeds the

control limit. The first signal occurs at a certain point of time which is said to be the run length. The control limit is usually chosen such that the in-control expectation of the run length is equal to a pre-specified value $\mathcal{A}$. In practice $\mathcal{A}$ is frequently chosen to be equal to 500. Below, the quantity T_t stands for one of the characteristic quantities $T_t^{(i)}$ for $i \in \{1, 2, 3\}$.

4.1. Multivariate EWMA charts

Here we follow the procedure in [15, 16]. We apply a multivariate EWMA (MuE) recursion to T_t. This leads to

$$Z_t \;=\; (\mathrm{I} - \Lambda)\, Z_{t-1} + \Lambda\, T_t \qquad \text{for} \quad t \geq 1\,. \tag{4.1}$$

Hence, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p, \lambda_{p+1})$ is a diagonal matrix of smoothing parameters of dimension $(p+1) \times (p+1)$. It is assumed that $0 < \lambda_i \leq 1$ for $i \in \{1, ..., p, p+1\}$. We presume that the starting value is equal to the target value such that $Z_0 = E_{\tau=\infty}(T_t)$. The mean vector of the considered quantity Z_t is then

$$E_{\tau=\infty}(Z_t) \;=\; E_{\tau=\infty}(T_t)\;. \tag{4.2}$$

Introducing $\mathrm{Cov}_{\tau=\infty}(T_{t-i}, T_{t-j}) = \Gamma(j-i)$ and assuming $\Lambda = \lambda\, \mathrm{I}_{p+1}$ the respective limit of the covariance matrix for $t \to \infty$ simplifies to

$$\lim_{t\to\infty} \mathrm{Cov}_{\tau=\infty}(Z_t) \;=\; \frac{\lambda}{2-\lambda}\,\left[\Gamma(0) + \sum_{v=1}^{\infty} (1-\lambda)^v \left[\Gamma(v) + \Gamma(v)'\right]\right]. \tag{4.3}$$

Since $\{Y_t\}$ is assumed to be a strictly stationary CCC model with existing fourth moments, the quantities $T_t^{(1)}$ and $T_t^{(3)}$ are weakly stationary. Since $\mathrm{Cov}_{\tau=\infty}(\eta_s, \eta_t) = \mathrm{O}_{(p+1)\times(p+1)}$ for $s \neq t$ in the case $\Lambda = \lambda\, \mathrm{I}_{p+1}$, the respective limit equals

$$\lim_{t\to\infty} \mathrm{Cov}_{\tau=\infty}\big(Z_t^{(2)}\big) \;=\; \frac{\lambda}{2-\lambda} \begin{pmatrix} \mathrm{I}_p & 0 \\ 0' & 2p \end{pmatrix}. \tag{4.4}$$

Eventually, the control statistic equals the Mahalanobis distance between Z_t and its in-control mean. This leads to

$$T_t \;=\; \big[Z_t - E_{\tau=\infty}(Z_t)\big]' \left[\mathrm{Cov}_{\tau=\infty}(Z_t)\right]^{-1} \big[Z_t - E_{\tau=\infty}(Z_t)\big]\,. \tag{4.5}$$

In order to implement a less time-consuming procedure one may use the asymptotic covariance matrix instead of the exact one

$$T_t \;=\; \big[Z_t - E_{\tau=\infty}(Z_t)\big]' \left\{\lim_{t\to\infty} \left[\mathrm{Cov}_{\tau=\infty}(Z_t)\right]\right\}^{-1} \big[Z_t - E_{\tau=\infty}(Z_t)\big]\,. \tag{4.6}$$

On the contrary, the Mahalanobis EWMA (MaE) chart scheme is based on the Mahalanobis distance of the vector T_t from its in-control mean $E_{\tau=\infty}(T_t)$. The control statistic is specified as

$$Z_t \;=\; (1-\lambda)\, Z_{t-1} + \lambda T_t \qquad \text{for} \quad t \geq 1 \tag{4.7}$$

where

$$T_t \;=\; \big[T_t - E_{\tau=\infty}(T_t)\big]' \big[\mathrm{Cov}_{\tau=\infty}(T_t)\big]^{-1} \big[T_t - E_{\tau=\infty}(T_t)\big]\,. \tag{4.8}$$

The quantity $\lambda \in (0,1]$ is said to be the memory parameter. The starting value Z_0 is chosen to be equal to the in-control expectation of T_t, i.e. $Z_0 = E_{\tau=\infty}(T_t) = p+1$.

4.2. Multivariate CUSUM control schemes

In [14] two multivariate control schemes based on the cumulative sum of $\{X_t\}$ are introduced. We extend this approach to the CCC model. Regarding the first multivariate CUSUM (MC1) chart the cumulative sum is determined before computing the respective standardized distance which represents a quadratic form. The cumulative sum is specified as

$$S_{t-n_t,t} \;=\; \sum_{i=t-n_t+1}^{t} \big[T_i - E_{\tau=\infty}(T_i)\big]\,, \qquad t \geq 1\,. \tag{4.9}$$

Accordingly, the relevant control statistic is based on a suitable norm of the cumulative sum given by

$$\|S_{t-n_t,t}\|_{\Gamma(0)} \;=\; \sqrt{S'_{t-n_t,t}\,\Gamma(0)^{-1}\,S_{t-n_t,t}}\,. \tag{4.10}$$

The control statistic is equal to the norm of the cumulative sum subtracted by a reference value

$$MC1_t \;=\; \max\big\{0,\, \|S_{t-n_t,t}\|_{\Gamma(0)} - k\,n_t\big\}\,, \qquad t \geq 1\,. \tag{4.11}$$

Thus, $k \geq 0$ is said to be the reference parameter. Further, the quantity n_t denotes the number of observations since the last restart given by

$$n_t = \begin{cases} n_{t-1}+1\,, & MC1_{t-1} > 0\,, \\ 1\,, & MC1_{t-1} = 0\,, \end{cases} \tag{4.12}$$

where $t \geq 1$ with $MC1_0 = 0$.

Regarding the second multivariate CUSUM (MC2) control chart we have to determine the cumulative sum after computing the standardized distance, the Mahalanobis distance of the quantity T_t from its in-control mean

$$D_t^2 \;=\; \big[T_t - E_{\tau=\infty}(T_t)\big]'\,\Gamma(0)^{-1}\,\big[T_t - E_{\tau=\infty}(T_t)\big]\,. \tag{4.13}$$

Eventually, the control statistic of the second multivariate CUSUM scheme equals

$$MC2_t \;=\; \max\big\{0,\, MC2_{t-1} + D_t^2 - (p+1) - 2\,k^2\big\} \tag{4.14}$$

where $k \geq 0$ and $MC2_0 = 0$.

5. COMPARISON STUDY

We intend to jointly monitor the mean vector and the variances of a bivariate nonlinear process. Initially, a CCC model must be chosen for a simulation study. Via a Monte Carlo simulation explicitly dominated control schemes should be identified. Eventually, the detection speed of the control schemes presented in the previous section should be evaluated. As a performance measure for control charts the ARL is used. In order to compute the ARLs we implement a program written in C++. The solutions are obtained using the bisection algorithm where 10^6 Monte Carlo replications are submitted for each algorithm iteration. This algorithm is interrupted when the numerical error of the ARL is less than $\pm 10^{-6}$ or the change in the control limits does not exceed $\pm 10^{-6}$.

5.1. Configuration of the Monte Carlo study

First of all, we need to calibrate the considered control charts such that the ARL in the in-control state is equal to a pre-specified value $\mathcal{A}$. Here we choose $\mathcal{A} = 120$, i.e. we consider approximately a half a year on the capital market. In the following section we want to focus on charts based on residuals. The control limits do not depend on the parameters of the underlying target process but only on the smoothing parameter for EWMA type charts and the reference value for CUSUM type schemes. Therefore, the calculation of the control design appears to be much easier for these schemes. In this comparison study we choose the smoothing parameter $\lambda \in \{0.1, 0.2, ..., 1.0\}$ and the reference value $k \in \{0.0, 0.1, ..., 1.0\}$. Here we present our results for the target process with

$$\sigma_{1t}^2 = 0.2 + 0.2\, Y_{1,t-1}^2 + 0.1\, \sigma_{1,t-1}^2 \tag{5.1}$$

$$\sigma_{2t}^2 = 0.1 + 0.1\, Y_{2,t-1}^2 + 0.2\, \sigma_{2,t-1}^2 \ . \tag{5.2}$$

The constant conditional correlation ϱ equals 0.5. We take into account shifts in the mean vector, shifts in the covariance matrix, as well as simultaneous shifts. Although we study all three cases only one table referring to the joint monitoring of means and variances is presented. Moreover, we choose $a_1 = a_2$ and $d_1 = d_2$. The elements of the vector a take the values $a_1 \in \{0.0, 0.25, ..., 2.0\}$. Further, the scale transformation Δ is specified by the parameter $d_1 \in \{1.0, 1.1, ..., 1.4\}$.

5.2. Detection of changes in means and variances

The out-of-control ARLs for the considered process are given in Table 1. The smallest out-of-control ARLs are printed in bold. The parameter values for

each scheme and each type of change leading to the smallest out-of-control ARLs are presented in parentheses.

Table 1: Out-of-control ARL of the CCC model based on an in-control ARL equal to $\mathcal{A} = 120$; ARLs refer to different shifts in means and variances ($m = 10^6$, $p = 2$).

a_1/d_1	1.0		1.1		1.2		1.3		1.4	
	MuE		**36.00**	(0.3)	**18.98**	(0.3)	**11.90**	(0.3)	**8.40**	(0.3)
0.0	MaE		40.70	(0.1)	20.03	(0.1)	12.29	(0.1)	8.57	(0.1)
	MC1		44.15	(0.4)	21.85	(0.4)	13.29	(0.5)	9.20	(0.5)
	MC2		40.99	(0.2)	21.19	(0.4)	13.32	(0.6)	9.36	(0.7)
	16.73	(0.1)	13.50	(0.2)	**10.16**	(0.2)	**7.94**	(0.2)	**6.37**	(0.3)
0.25	39.81	(0.1)	20.45	(0.1)	12.80	(0.1)	9.01	(0.1)	6.84	(0.1)
	14.80	(0.3)	**12.68**	(0.3)	10.21	(0.4)	8.19	(0.5)	6.62	(0.5)
	39.55	(0.2)	21.58	(0.4)	13.90	(0.5)	9.88	(0.6)	7.49	(0.7)
	5.94	(0.2)	5.36	(0.2)	4.84	(0.2)	4.36	(0.3)	**3.91**	(0.3)
0.5	10.14	(0.1)	7.66	(0.1)	6.12	(0.1)	5.10	(0.1)	4.37	(0.1)
	5.56	(0.5)	**5.15**	(0.5)	**4.74**	(0.5)	**4.34**	(0.6)	3.95	(0.6)
	11.19	(0.5)	8.50	(0.6)	6.76	(0.7)	5.59	(0.8)	4.75	(0.8)
	3.19	(0.3)	3.01	(0.3)	2.85	(0.3)	2.71	(0.3)	**2.58**	(0.3)
0.75	4.21	(0.1)	3.70	(0.1)	3.34	(0.1)	3.05	(0.1)	2.82	(0.1)
	3.06	(0.7)	**2.94**	(0.7)	**2.81**	(0.7)	**2.70**	(0.7)	2.59	(0.7)
	4.72	(0.8)	4.11	(0.8)	3.64	(0.9)	3.29	(1.0)	3.01	(1.0)
	2.01	(0.3)	1.97	(0.4)	1.93	(0.4)	1.89	(0.4)	1.86	(0.4)
1.0	2.28	(0.1)	2.18	(0.1)	2.10	(0.1)	2.02	(0.1)	1.96	(0.1)
	1.96	(0.9)	**1.93**	(0.9)	**1.91**	(0.8)	**1.88**	(0.9)	**1.85**	(0.9)
	2.49	(1.0)	2.35	(1.0)	2.23	(1.0)	2.14	(1.0)	2.06	(1.0)
	1.42	(0.4)	1.43	(0.4)	1.44	(0.4)	**1.44**	(0.4)	1.45	(0.4)
1.25	1.49	(0.1)	1.49	(0.1)	1.49	(0.1)	1.49	(0.1)	1.48	(0.1)
	1.41	(1.0)	**1.42**	(1.0)	**1.43**	(1.0)	**1.44**	(1.0)	**1.44**	(1.0)
	1.56	(1.0)	1.56	(1.0)	1.55	(1.0)	1.54	(1.0)	1.54	(1.0)
	1.14	(0.5)	**1.16**	(0.5)	**1.18**	(0.5)	**1.20**	(0.5)	**1.21**	(0.5)
1.5	1.16	(0.1)	1.18	(0.1)	1.20	(0.1)	1.21	(0.1)	1.23	(0.1)
	1.15	(1.0)	1.17	(1.0)	**1.18**	(1.0)	**1.20**	(1.0)	1.22	(1.0)
	1.18	(1.0)	1.21	(1.0)	1.22	(1.0)	1.24	(1.0)	1.25	(1.0)
	1.04	(0.6)	**1.05**	(0.6)	**1.06**	(0.5)	**1.08**	(0.5)	**1.09**	(0.5)
1.75	**1.04**	(0.2)	**1.05**	(0.3)	1.07	(0.2)	**1.08**	(0.1)	**1.09**	(0.2)
	1.04	(1.0)	**1.05**	(1.0)	1.07	(1.0)	**1.08**	(1.0)	**1.09**	(1.0)
	1.05	(1.0)	1.06	(1.0)	1.08	(1.0)	1.09	(1.0)	1.11	(1.0)
	1.01	(0.6)	**1.01**	(0.6)	**1.02**	(0.5)	**1.03**	(0.5)	**1.03**	(0.5)
2.0	**1.01**	(0.4)	**1.01**	(0.3)	**1.02**	(0.3)	**1.03**	(0.4)	**1.03**	(0.4)
	1.01	(1.0)	**1.01**	(1.0)	**1.02**	(1.0)	**1.03**	(1.0)	1.04	(1.0)
	1.01	(1.0)	1.02	(1.0)	**1.02**	(1.0)	**1.03**	(1.0)	1.04	(1.0)

Regarding Table 1 none of the introduced charts exclusively dominates all the other schemes. For small changes in means and variances the MuE chart provides the smallest out-of-control ARLs. For medium-sized changes, the MC1 chart turns out to be the best scheme. Finally, for larger changes again the MuE scheme appears to be the best control chart. Since the deviations between MC1

and MuE are small for larger changes, the MC1 control chart seems to be the most suitable scheme.

Accordingly, the optimal values of the parameters λ and k increase with increasing mean and variance changes. If we concentrate on the MC1 and the MC2 scheme, in many situations the optimal k is on the boundary $k = 1.0$. Consequently, their performance could be improved choosing a higher upper bound for k.

Additionally, we compare these findings with the results in [6]. In that paper we propose control charts where changes in means, variances, and covariances are taken into account. The characteristic quantities of these schemes are of dimension $p(p+3)/2$. Neglecting covariances in the present approach we reduce the dimension to $p+1$. Nevertheless, the residual charts monitoring means and variances provide smaller out-of-control ARLs. Consequently, the reduction of the dimension does not lead to a loss of efficiency. In our study we analyze processes up to dimension 5. The charts seem to be useful for higher dimensions as well. However, further research is necessary in order to assess their behavior for high-dimensional processes.

6. CONCLUSION

Multivariate nonlinear time series are very attractive for practical applications in finance because of their time dependent conditional covariance matrix. In this paper we propose new control charts for the joint surveillance of the means and the variances of such processes. Therefore, characteristic quantities based on current observations and residuals as well as characteristics based on conditional variances are introduced. Several multivariate EWMA and CUSUM schemes in connection with these characteristic quantities are proposed. The dimension of the characteristics is reduced compared to the schemes proposed in [6]. As a consequence, these control charts can be applied to nonlinear processes with an explicitly higher dimension. Since the schemes based on residuals are much easier to handle, we recommend the application of residual charts. First the control design does not depend on the parameters of the target process. Therefore, the control limits can be easily determined. Second the residual schemes can be applied assuming weaker conditions on the underlying process. Eventually, financial processes usually do not fulfill conditions on higher moments.

Via a Monte Carlo simulation study we compare the detection speed of each control scheme using the out-of-control ARL as a reliable indicator. In many cases we find that the MC1 control chart appears to be the best chart for joint shifts in means and variances.

REFERENCES

[1] Bodnar, O. and Schmid, W. (2007). Surveillance of the mean behavior of multivariate time series, *Stat. Neerl.*, **61**, 1–24.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**(3), 307–327.

[3] Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Rev. Econ. Stat.*, **72**, 498–505.

[4] Crosier, R.B. (1988). Multivariate generalizations of cumulative sum quality control schemes, *Technometrics*, **30**, 291–303.

[5] Engle, R.F. and Kroner, K.F. (1995). Multivariate simultaneous generalized ARCH, *Economet. Theor.*, **11**, 122–150.

[6] Garthoff, R.; Okhrin I. and Schmid W. (2014). Statistical surveillance of the mean vector and the covariance matrix of nonlinear time series, *AStA – Advances in Statistical Analysis*, **98**(3), 225–255.

[7] Garthoff, R.; Okhrin, I. and Schmid, W. (2012). Control charts for multivariate nonlinear time series. European University Viadrina Frankfurt (Oder), Department of Business Administration and Economics, Discussion Paper.

[8] Golosnoy, V.; Ragulin S. and Schmid W. (2009). Multivariate CUSUM chart: properties and enhancements, *Advances in Statistical Analysis*, **93**, 263–279.

[9] Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models, *Economet. Theor.*, **14**, 70–86.

[10] Khediri, I.B.; Weihs, C. and Limam, C. (2010). Support vector regression control charts for multivariate nonlinear autocorrelated processes, *Chemometr. Intell. Lab.*, **103**, 76–81.

[11] Kramer, H. and Schmid, W. (1997). EWMA charts for multivariate time series, *Sequential Analysis*, **16**, 131–154.

[12] Lowry, C.A.; Woodall, W.H.; Champ, C.W. and Rigdon, S.E. (1992). A multivariate exponentially weighted moving average control chart, *Technometrics*, **34**, 46–53.

[13] Ngai, H.-M. and Zhang, J. (2001). Multivariate cumulative sum control charts based on projection pursuit, *Stat. Sinica*, **11**, 747–766.

[14] Pignatiello, J.J. and Runger, G.C. (1990). Comparison of multivariate CUSUM charts, *J. Qual. Technol.*, **22**, 173–186.

[15] Śliwa, P. and Schmid, W. (2005). Monitoring the cross-covariances of a multivariate time series, *Metrika*, **61**, 89–115.

[16] Śliwa, P. and Schmid, W. (2005). Surveillance of the covariance matrix of multivariate nonlinear time series, *Statistics*, **39**, 221–246.

[17] Theodossiou, P.T. (1993). Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures, *J. Am. Stat. Assoc.*, **88**, 441–449.

REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 145–168

MODEL OF GENERAL SPLIT-BREAK PROCESS

Authors: Vladica Stojanović
– Faculty of Sciences and Mathematics,
University of Priština (in K. Mitrovica), Serbia
vladica.stojanovic@pr.ac.rs

Biljana Č. Popović
– Faculty of Sciences and Mathematics,
University of Niš, Serbia
bipop@pmf.ni.ac.rs

Predrag Popović
– Faculty of Civil Engineering and Architecture,
University of Niš, Serbia
popovicpredrag@yahoo.com

Received: May 2013 Revised: January 2014 Accepted: February 2014

Abstract:

- This paper presents a modification (and partly a generalization) of STOPBREAK process, which is the stochastic model of time series with permanent, emphatic fluctuations. The threshold regime of the process is obtained by using, so called, Noise indicator. We proceed to investigate the model which we named the General Split-BREAK (GSB) process. After brief recalling of its basic stochastic properties, we give some procedures of its parameters estimation. A Monte Carlo study of this process is also give, along with the application in the analysis of stock prices dynamics of several Serbian companies which were traded on Belgrade Stock Exchange.

Key-Words:

- *GSB process; STOPBREAK process; noise indicator; Split-MA process; stationarity; invertibility; parameters estimation.*

AMS Subject Classification:

- 62M10, 91B84.

1. INTRODUCTION

Starting from the fundamental results of Engle and Smith [4], who were the first who had introduced the stochastic process of permanent fluctuations, named STOPBREAK process, we had defined a new, modified version of the well known generalization of STOPBREAK process. In our model, we set the threshold noise indicator as we had recently done in the time series of ARCH type, described in Popović and Stojanović [12] and Stojanović and Popović [15]. Our model, named *the General Split-BREAK process* or, simply, *GSB process*, is at the same time the generalization of so called Split-BREAK model introduced in Stojanović *et al.* [16].

In the next section, Section 2, we shall briefly present the definition and the main stochastic properties of GSB model, described in detail in Stojanović *et al.* [17], and we will define the sequence of increments of the GSB process, called *Split-MA process*. Beside the standard investigation of the stochastic properties of Split-MA model, we will particularly give the conditions of its invertibility. The main result of this paper, procedures of the parameters estimation of GBS process, are described in Section 3. We will pay the special attention to the estimation of the threshold parameter, named *critical value of the reaction*. We shall prove the asymptotic properties of evaluated estimates. The following section, Section 4, is devoted to the Monte Carlo simulations of the innovations of GSB process. Section 5 describes an application of the estimation procedures on the real data of some trading volumes on Belgrade Stock Exchange. Finally, Section 6 is the conclusion.

2. THE GSB PROCESS. DEFINITION AND MAIN PROPERTIES

We shall suppose that (y_t) is the time series with the known values at time $t \in \{0, 1, ...T\}$ and $F = (\mathcal{F}_t)$ is a filtration defined on the probability space $(\Omega, \mathcal{F}, P)$. Following Engle and Smith [4], the sequence (y_t) will be a *General STOPBREAK process* if it satisfies the recurrent relation

$$A(L)\,B(L)\,y_t \;=\; q_{t-1}A(L)\,\varepsilon_t + (1-q_{t-1})\,B(L)\,\varepsilon_t\,, \qquad t=1,...,T\,, \tag{2.1}$$

where, $A(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$, $B(L) = 1 - \sum_{k=1}^{r} \beta_k L^k$, and L is the backshift operator. On the other hand, (ε_t) is a white noise, i.e. the i.i.d. sequence of random variables adapted to the filtration F, which satisfies

$$E\big(\varepsilon_t \,|\, \mathcal{F}_{t-1}\big) = 0\,, \quad \mathrm{Var}\big(\varepsilon_t \,|\, \mathcal{F}_{t-1}\big) = \sigma^2\,, \qquad t=1,...,T\,.$$

At last, (q_t) represents the sequence of random variables which depends on the white noise (ε_t), and in addition, $P\big\{0 \le q_t \le 1\big\} = 1$ for each $t = 0, 1, ..., T$. The

sequence (q_t) displays the *(permanent) reaction* of the STOPBREAK process, because its values determine the amount of participation of previous elements of white noise process engaged in the definition of (y_t).

In that way, the structure of the sequence (q_t) determines the character and the properties of the STOPBREAK process, which vary among the well known linear stochastic models. This model was investigated later by several authors, for instance Gonzáles [6], or Diebold [3], whose works were based on certain variations of the reaction (q_t). On the other hand, many authors, for instance Huang and Fok [9] or Kapetanios and Tzavalis [10], have studied mainly the practical application of the STOPBREAK (and some similar) processes.

Similarly as in the definition of Split-ARCH model [12, 15], we shall suppose in the following that

$$q_t \;=\; I(\varepsilon_{t-1}^2 > c) \;=\; \begin{cases} 1\,, & \varepsilon_{t-1}^2 > c \\ 0\,, & \varepsilon_{t-1}^2 \le c\,, \end{cases} \qquad t = 1, ..., T\,, \tag{2.2}$$

i.e., that the permanent reaction (2.2) represents, so-called *a Noise indicator*. Remark that, according to (2.2), it follows that

$$E\big(q_t\,\varepsilon_t\,|\,\mathcal{F}_{t-1}\big) \;=\; q_t\,E\big(\varepsilon_t\,|\,\mathcal{F}_{t-1}\big) \;=\; 0\;,$$

and it can be seen that the sequence $(q_t\,\varepsilon_t)$ is a martingale difference, as in the definition of basic STOPBREAK model [4].

However, it seems that in the case of general STOPBREAK process this formulation of reaction (q_t) is inadequate. The primary reason for such opinion is the fact that the model (2.1) includes only "directly previous" realizations of (q_t), which are obtained at the moment $t-1$. Therefore, the general STOPBREAK process (2.1) with the reaction (2.2) operates in (only) two different regimes

$$\varepsilon_t \;=\; \begin{cases} A(L)\,y_t\,, & q_{t-1} = 0 \;\; (\text{w.p. } b_c) \\ B(L)\,y_t\,, & q_{t-1} = 1 \;\; (\text{w.p. } a_c)\,, \end{cases} \tag{2.3}$$

where $a_c = E(q_t) = P\{\varepsilon_{t-1}^2 > c\}$, $b_c = 1 - a_c$ and "w.p." stands for "with probability". Therefore, the equality (2.3) defines the well known Thresholds Autoregressive (TAR) model introduced by Tong [18] and discussed in details, for instance by Chan [2], Hansen [8] and, in the newest time, Scarrott and MacDonald [13] and some other authors. Based on this, here we discuss the different generalization of Split-BREAK process, more general than that of Engle and Smith [4].

Definition 2.1. Let L be a backshift operator, $A(L) = 1 - \sum_{i=1}^{m} \alpha_i L^i$, $B(L) = 1 - \sum_{j=1}^{n} \beta_j L^j$, $C(L) = 1 - \sum_{k=1}^{p} \gamma_k L^k$, and (q_t) the noise indicator defined with (2.2). Then, the sequence (y_t) represents *the General Split-BREAK* (*GSB*) *process* if it satisfies

$$A(L)\,y_t \;=\; B(L)\,q_t\,\varepsilon_t + C(L)\,(1 - q_t)\,\varepsilon_t\,, \qquad t \in \mathbf{Z}\;. \tag{2.4}$$

Note that the definition above represents the general stochastic model which as its specific forms, contains the most of other, well known models. In dependence of $A(L)$, $B(L)$ and $C(L)$ we have, for example, the following processes:

$$\begin{array}{lll} A(L)=B(L)=C(L)=1: & y_t=\varepsilon_t & \text{(White Noise)} \\ A(L)=1,\ B(L)=C(L)\neq 1: & y_t=B(L)\varepsilon_t & \text{(MA model)} \\ A(L)\neq 1,\ B(L)=C(L)=1: & A(L)y_t=\varepsilon_t & \text{(AR model) .} \end{array}$$

Finally, in the case when $A(L)=C(L)=1-L$ and $B(L)=1$ we get the Split-BREAK process introduced by Stojanović *et al.* [16]. In the following, we shall analyze some specificity of the model (2.4) and suppose $A(L)=C(L)\neq 1$ and $B(L)=1$. Thus, the defined model can be written in the form

$$y_t-\sum_{j=1}^{p}\alpha_j\,y_{t-j}\;=\;\varepsilon_t-\sum_{j=1}^{p}\alpha_j\,\theta_{t-j}\,\varepsilon_{t-j}\,,\qquad t\in\mathbf{Z}\;, \tag{2.5}$$

where $\alpha_j\geq 0,\ j=1,...,p,$ and $\theta_t=1-q_t=I(\varepsilon_{t-1}^2\leq c)$. Obviously, this representation is close to linear ARMA time series, except that it has the indicators of noise (ε_t) in its own structure. They indicate the realizations of noise which have statistically significant weights in "previous" time. These "temporary" components change the ARMA structure of GSB model (2.5). In this way, they make some difficulties in the usage of well known procedures in investigation of the properties of our model .

On the other hand, similarly to the basic STOPBREAK process, the equality (2.4) enables that the sequence (y_t) can be presented in the form of additive decomposition

$$y_t\;=\;m_t+\varepsilon_t\,,\qquad t\in\mathbf{Z}\;, \tag{2.6}$$

where

$$m_t\;=\;\sum_{j=1}^{p}\alpha_j\big(y_{t-j}-\theta_{t-j}\,\varepsilon_{t-j}\big)\;=\;\sum_{j=1}^{p}\alpha_j\big(m_{t-j}+q_{t-j}\,\varepsilon_{t-j}\big) \tag{2.7}$$

is the sequence of random variables which we named *the martingale means*. It is general case of analogous equality of Engle and Smith [4], which is obtained from (2.7), for $p=\alpha_1=1$. According to (2.6) it follows

$$E\big(y_t\,|\,\mathcal{F}_{t-1}\big)\;=\;m_t+E\big(\varepsilon_t\,|\,\mathcal{F}_{t-1}\big)\;=\;m_t\;, \tag{2.8}$$

from which it follows $E(y_t)=E(m_t)=\mu$ (const.), i.e. the means of these two sequences are equal and constant. The variance of GSB process can be determined in a similar way. As

$$\mathrm{Var}\big(y_t\,|\,\mathcal{F}_{t-1}\big)\;=\;E\big(y_t^2\,|\,\mathcal{F}_{t-1}\big)-m_t^2\;=\;\sigma^2\;, \tag{2.9}$$

we can conclude that the conditional variance (volatility) of the sequence (y_t) is a constant and it is equal to the variance of the noise (ε_t). Let us remark that the

equalities (2.8) and (2.9) explain the stochastic nature of (2.4). As the sequence (m_t) is predictable, it will be a component which demonstrates the stability of the process itself. Contrary, the sequence (ε_t) is the factor which represents the deviations (or random fluctuations) from values (m_t). On the other hand, the variances of sequences (m_t) and (y_t) satisfy the relation

$$\mathrm{Var}(y_t) \,=\, \mathrm{Var}(m_t) + \sigma^2$$

and, in the non-stationary case, these are not constants, i.e., depend on the observation time t.

In the following, we shall describe stochastic structure of the increments

$$X_t \stackrel{\mathrm{def}}{=} A(L)\, y_t\,, \qquad t \in \mathbf{Z}\ ,$$

which, according to (2.4) and (2.5), we can write in the form of recurrent relation

$$X_t \,=\, \varepsilon_t - \sum_{j=1}^{p} \alpha_j\, \theta_{t-j}\, \varepsilon_{t-j}\,, \qquad t \in \mathbf{Z}\ . \tag{2.10}$$

Obviously, the sequence (X_t) has the multi–regime structure, which depends on the realizations of indicators (θ_t). If all fluctuations of the white noise in time $t-j$ are large, an increment X_t will be equal to the white noise. On the other hand, the fluctuations of the white noise which do not exceed the critical value c will produce a "part of" MA(p) representation of (X_t). In this way, the similarity of this model to the standard linear MA model is noticeable, and the sequence (X_t) we shall call *the general Split-MA model* (*of order p*) or, simply, *Split-MA*(p) *model*. It represents the generalization of the model defined in Stojanović *et al.* [16], and the threshold integrated moving average (TIMA) model introduced by Gonzalo and Martinez [7]. The main properties of this process can be expressed as follow.

Theorem 2.1. *The sequence* (X_t), *defined by* (2.10), *is stationary, with mean* $E(X_t) = 0$ *and covariance* $\gamma_{_X}(h) = E(X_t X_{t+h})$, $h \geq 0$, *which satisfies the equality*

$$\gamma_{_X}(h) \,=\, \begin{cases} \sigma^2 \Big(1 + b_c \sum\limits_{j=1}^{p} \alpha_j^2 \Big), & h = 0 \\ \sigma^2\, b_c \Big(\sum\limits_{j=1}^{p-h} \alpha_j\, \alpha_{j+h} - \alpha_h \Big), & 1 \leq h \leq p-1 \\ -\sigma^2\, b_c\, \alpha_p\,, & h = p \\ 0\,, & h > p\ . \end{cases} \tag{2.11}$$

Proof: Elementary. □

Similarly to the basic STOPBREAK model, we can, under some conditions, show the invertibility of increments (X_t). This property is analyzed from different aspects by many authors who explored the STOBREAK models. We shall do the same concerning our model in order to proceed estimation of the unknown parameters of the model and apply our model to real data. As we shall see later, only realizations of invertible Split-MA process can give strong consistent and asymptotically normal estimates.

Theorem 2.2. *The sequence (X_t), defined by* (2.10)*, is invertible iff the roots r_j, $j=1,...,p$, of characteristic polynomial*

$$Q(\lambda) \,=\, \lambda^p - b_c \sum_{j=1}^{p} \alpha_j \, \lambda^{p-j}$$

satisfy the condition $|r_j| < 1$, $j=1,...,p$, or, equivalently, $b_c \sum\limits_{j=1}^{p} \alpha_j < 1$. Then,

$$\varepsilon_t \,=\, \sum_{k=0}^{\infty} \omega_k(t)\, X_{t-k}\,, \qquad t \in \mathbf{Z}\,, \tag{2.12}$$

where $\big(\omega_k(t)\big)$ is the solution of stochastic difference equation

$$\omega_k(t) \,=\, \theta_{t-k} \sum_{j=1}^{p} \alpha_j\, \omega_{k-j}(t)\,, \qquad k \geq p\,, \quad t \in \mathbf{Z}\,, \tag{2.13}$$

with the initial conditions $\omega_0(t)=1$, $\omega_k(t) = \theta_{t-k} \sum_{j=1}^{k} \alpha_j\, \omega_{k-j}(t)$, $1 \leq k \leq p-1$. Morever, the representation (2.12) *is almost surely unique and the sum converges with probability one and in the mean-square sense.*

Proof: First of all, for any $t \in \mathbf{Z}$, we define the vectors and matrices

$$\mathbf{V}_t \,=\, \big(\, \varepsilon_t \;\; \varepsilon_{t-1} \;\; \cdots \;\; \varepsilon_{t-p+1} \,\big)'\,, \qquad \mathbf{X}_t \,=\, \big(\, X_t \;\; 0 \;\; \cdots \;\; 0 \,\big)'\,,$$

$$\mathbf{A}_t \,=\, \begin{pmatrix} \alpha_1\,\theta_t & \alpha_2\,\theta_{t-1} & \cdots & \alpha_{p-1}\,\theta_{t-p+2} & \alpha_p\,\theta_{t-p+1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

and we can write the model (2.10) in the form of stochastic difference equation of order one

$$\mathbf{V}_t \,=\, \mathbf{A}_{t-1}\mathbf{V}_{t-1} + \mathbf{X}_t\,, \qquad t \in \mathbf{Z}\,. \tag{2.14}$$

From here, we have

$$\mathbf{V}_t \;=\; \mathbf{X}_t + \sum_{j=1}^{k}\Big(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-j}\Big)\mathbf{X}_{t-j} + \left(\prod_{j=1}^{k+1}\mathbf{A}_{t-j}\right)\mathbf{V}_{t-k-1}\;,$$

where $k = 1, 2, \ldots$. It can be proven (see, for instance Francq *et al.* [5]) that the existence of almost sure unique, stationary solution of equation (2.14), in the form

$$\mathbf{V}_t \;=\; \mathbf{X}_t + \sum_{k=1}^{\infty}\Big(\mathbf{A}_{t-1}\cdots\mathbf{A}_{t-k}\Big)\mathbf{X}_{t-k}\,, \qquad t\in\mathbf{Z}\;, \tag{2.15}$$

is equivalent to the convergence

$$\prod_{j=1}^{k+1}\mathbf{A}_{t-j} \xrightarrow{\text{a.s.}} 0\,, \qquad k\longrightarrow\infty\;,$$

i.e., to the fact that the eigenvalues r_j, $j = 1, ..., p$, of the matrix

$$\mathbf{A} \;=\; E(\mathbf{A}_t) \;=\; \begin{pmatrix} \alpha_1 b_c & \alpha_2 b_c & \cdots & \alpha_{p-1} b_c & \alpha_p b_c \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

satisfy the conditions $|r_j| < 1$, $j = 1, ..., p$. According to the representation

$$\det\big(\mathbf{A} - \lambda\mathbf{I}\big) \;=\; (-1)^p\, Q(\lambda)\;,$$

it is obvious that the eigenvalues r_j, $j = 1, ..., p$, are the roots of the characteristic polynomial $Q(\lambda)$. Then, the condition $|r_j| < 1$, $j = 1, ..., p$, is necessary and sufficient for the almost sure uniqueness of the representation (2.15), and the almost sure convergence of the corresponding sum. In the similar way, we can prove that the same conditions are equivalent to the mean square convergence of the sum in (2.15). From this point on, by simple computation, we can obtain the equations (2.12) and (2.13). □

Based on the proposition above, it is clear that the presence of the sequence (θ_t) in (2.10) enables the conditions of invertibility of increments (X_t) to be weaker than corresponding conditions that are related to the stationarity of the series (y_t) and (m_t) (see, for more details [17]). In that way, even non-stationary time series (y_t) and (m_t) can form invertible Split-MA process which is always stationar. This situation is particularly interesting in the case of so called *integrated* (*standardized*) time series, where

$$\sum_{j=1}^{p}\alpha_j = 1\;. \tag{2.16}$$

If the value of parameter b_c is non-trivial, i.e., $b_c \in (0,1)$, then the sequence (X_t) will be invertible although (y_t) and (m_t) are non-stationary time series. We will further assume that the "normality condition" (2.16) is always fulfilled, because it will be of dual importance below. Primarily, the condition (2.16), which defines series (y_t) and (m_t) as the non-stationary ones, allows us, as opposed to the stationary case, that these two series have non-zero means, which is particularly important in applications (see, for instance, Section 5). Finally, another reason for introducing the "normality condition" (2.16) lays in simplifying the estimation procedure of unknown parameters of GSB model. As the sequences (y_t) and (X_t), in general, do not depend on the coefficients $\alpha_1, ..., \alpha_p$ only, but also of the critical level of reaction $c > 0$, the presumption (2.16) will be an additional "functional" relationship between the unknown parameters which allows us to compute them uniquely. In the next section there will be more discussion about such an idea of parameters estimation when GSB model is standardized GSB model.

3. PARAMETERS ESTIMATION

Procedure of estimation of the unknown parameters $\alpha_1, ..., \alpha_p, c$ of GSB model will be based on the realization of a stationary Split-MA(p) process (X_t). For this purpose, we suppose that $X_1, ..., X_T$ is the part of a realization of this time series for which we define two kinds of estimates. First, equating the covariance $\gamma_X(h),\ h = 0, ..., p$, defined by equality (2.11), with its empirical value

$$\hat{\gamma}_X(h) \,=\, \frac{1}{T-h} \sum_{t=1}^{T-h} X_t \, X_{t+h} \,, \qquad h = 0, ..., p \ ,$$

we get the, so called, *initial estimates* of unknown parameters. We denote these estimates with $\widetilde{\alpha}_1, ..., \widetilde{\alpha}_p$, respectively, i.e., in the case of critical value, with $\widetilde{c}$. Obviously, these are continuous functions of estimates $\hat{\gamma}_X(h)$, and according to the well known properties of continuity the almost sure convergence and the convergence in distribution (see, for instance Serfling [14]) it can be easily proved that $\widetilde{\alpha}_1, ..., \widetilde{\alpha}_p, \widetilde{c}$ are the consistent and the asymptotically normal estimates of $\alpha_1, ..., \alpha_p, c$.

In spite of good stochastic properties of these estimates, it can be proven that these are not the efficient estimates of unknown parameters. In order to achieve better estimates of unknown parameters we introduce *the regression estimates* $\hat{\alpha}_1, ...\hat{\alpha}_p, \hat{c}$ based on the regression of sequence

$$W_t \,=\, \sum_{j=1}^{p} \alpha_j \, \theta_{t-j+1} W_{t-j} + \varepsilon_{t-1} \,, \qquad t \in \mathbf{Z} \ . \tag{3.1}$$

For this reason, we will firstly show that the necessary and sufficient stationarity

conditions of a series of (W_t) are equivalent to the conditions of invertibility of (X_t), described in Theorem 2.2.

Theorem 3.1. *The sequence (W_t), defined by (3.1), is the stationary and ergodic iff the roots $r_j,\ j=1,...,p$, of characteristic polynomial*

$$Q(\lambda) \,=\, \lambda^p - b_c \sum_{j=1}^{p} \alpha_j \, \lambda^{p-j}$$

satisfy the condition $|r_j|<1,\ j=1,...,p$, or, equivalently, $b_c \sum\limits_{j=1}^{p} \alpha_j < 1$.

Proof: If we introduce the vectors and matrices

$$\mathbf{W}_t \,=\, \big(\, W_t \;\; W_{t-1} \;\; \cdots \;\; W_{t-p+1} \,\big)', \qquad \mathbf{E}_t \,=\, \big(\, \varepsilon_t \;\; 0 \;\; \cdots \;\; 0 \,\big)',$$

$$\mathbf{A}_t \,=\, \begin{pmatrix} \alpha_1\,\theta_t & \alpha_2\,\theta_{t-1} & \cdots & \alpha_{p-1}\,\theta_{t-p+2} & \alpha_p\,\theta_{t-p+1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

then the equality (3.1) can be written in the form of recurrent relation

$$\mathbf{W}_t \,=\, \mathbf{A}_t \mathbf{W}_{t-1} + \mathbf{E}_{t-1}\,, \qquad t \in \mathbf{Z}\ . \tag{3.2}$$

From here, completely analogously to the Theorem 2.2, it can be shown that the equation (3.2) has the strictly stationary, almost sure unique and ergodic solution

$$\mathbf{W}_t \,=\, \mathbf{E}_{t-1} + \sum_{k=1}^{\infty} \Big(\mathbf{A}_t \cdots \mathbf{A}_{t-k+1} \Big)\, \mathbf{E}_{t-k-1}\,, \qquad t \in \mathbf{Z}\ ,$$

if and only if the eigenvalues $r_1,...,r_p$ of matrix $\mathbf{A} = E(\mathbf{A}_t)$ satisfy the condition $|r_j|<1,\ j=1,...,p$. □

Now, let us define, using the procedure described in [11], the residual sequence

$$R_t \,=\, W_t - \sum_{j=1}^{p} a_j \, W_{t-j}\,, \qquad t \in \mathbf{Z}\ , \tag{3.3}$$

where we denoted $a_j = b_c\,\alpha_j$, $j=1,...,p$. We shall prove the following proposition.

Theorem 3.2. *If the sequence* (W_t)*, defined by* (3.1) *is stationary, then the sequence* (R_t)*, defined by* (3.3) *is the sequence of uncorrelated random variables.*

Proof: If we introduce the vectors $\mathbf{R}_t = \begin{pmatrix} R_t & 0 & \cdots & 0 \end{pmatrix}'$, $t \in \mathbf{Z}$, then it is valid

$$\mathbf{R}_t = \mathbf{W}_t - \mathbf{A}\,\mathbf{W}_{t-1}\,, \qquad t \in \mathbf{Z}\,,$$

where $\mathbf{W}_t$ and $\mathbf{A}$ are the matrices that we defined earlier. For arbitrary $h > 0$ the covariance matrix $\mathbf{\Gamma_R}(h) \stackrel{\text{def}}{=} E(\mathbf{R}_t\mathbf{R}'_{t-h})$ of the vectors $(\mathbf{R}_t)$ can be written as

$$\mathbf{\Gamma_R}(h) = \mathbf{\Gamma_W}(h) - \mathbf{A}\,\mathbf{\Gamma_W}(h+1) - \mathbf{A}\,\mathbf{\Gamma_W}(h-1) + \mathbf{A}\,\mathbf{\Gamma_W}(h)\mathbf{A}'\,, \tag{3.4}$$

where

$$\mathbf{\Gamma_W}(h) = \begin{pmatrix} \gamma_{_W}(h) & \gamma_{_W}(h+1) & \cdots & \gamma_{_W}(h+p-1) \\ \gamma_{_W}(h-1) & \gamma_{_W}(h) & \cdots & \gamma_{_W}(h+p-2) \\ \vdots & \vdots & & \vdots \\ \gamma_{_W}(h-p+1) & \gamma_{_W}(h-p+2) & \cdots & \gamma_{_W}(h) \end{pmatrix}$$

is covariance matrix of the vector series $(\mathbf{W}_t)$, and

$$\gamma_{_W}(h) = E(W_tW_{t-h})\,, \qquad \gamma_{_W}(-h) = \gamma_{_W}(h)$$

is covariance of the stationary series (W_t). Using simple calculation it can be shown that there is $\mathbf{A}\,\mathbf{\Gamma_W}(h) = \mathbf{\Gamma_W}(h)\mathbf{A}' = \mathbf{\Gamma_W}(h+1)$, and by substituting this equality in (3.4) immediately follows $\mathbf{\Gamma_R}(h) = \mathbf{O}_{p\times p}$. □

Notice that in the equality (3.3) we defined the sequence $\big(W_t\big)$ as a linear autoregressive process of order p, with noise $\big(R_t\big)$. Then, by using standard regression procedure, we can obtain the estimate $\hat{\mathbf{a}}_{_T} = (\hat{a}_1, ..., \hat{a}_p)'$ of parameter $\mathbf{a} = (a_1, ..., a_p)'$, in the form of equality

$$\hat{\mathbf{a}}_{_T} = \mathbf{W}_{_T}^{-1} \cdot \mathbf{b}_{_T}\,, \tag{3.5}$$

where

$$\mathbf{W}_{_T} = \begin{pmatrix} \sum\limits_{t=p+1}^{T} W_{t-1}^2 & \sum\limits_{t=p+1}^{T} W_{t-1}W_{t-2} & \cdots & \sum\limits_{t=p+1}^{T} W_{t-1}W_{t-p} \\ \sum\limits_{t=p+1}^{T} W_{t-1}W_{t-2} & \sum\limits_{t=p+1}^{T} W_{t-2}^2 & \cdots & \sum\limits_{t=p+1}^{T} W_{t-2}W_{t-p} \\ \vdots & \vdots & & \vdots \\ \sum\limits_{t=p+1}^{T} W_{t-1}W_{t-p} & \sum\limits_{t=p+1}^{T} W_{t-2}W_{t-p} & \cdots & \sum\limits_{t=p+1}^{T} W_{t-p}^2 \end{pmatrix},$$

$$\mathbf{b}_{_T} = \left(\sum_{t=p+1}^{T} W_t W_{t-1} \ \sum_{t=p+1}^{T} W_t W_{t-2} \ \cdots \ \sum_{t=p+1}^{T} W_t W_{t-p} \right)' .$$

Now we are able to show the asymptotic properties of obtained estimates. In this order, we define the set

$$\mathcal{A} = \left\{ \mathbf{a} = (a_1, ..., a_p)' \in \mathbf{R}^p \,\Big|\, \sum_{j=1}^{p} a_j < 1 \right\}$$

which, obviously, is a set of parameter values $\mathbf{a}$ for which the invertibility condition of Split-MA process (X_t), i.e. the stationarity condition of the series (W_t) is satisfied. In the mentioned assumptions, the following assertion is valid.

Theorem 3.3. *Let, for some $T_0 > 0$ and any $T \geq T_0$, the condition $\hat{\mathbf{a}}_{_T} \in \mathcal{A}$ is fulfilled. Then $\hat{\mathbf{a}}_{_T}$ is strictly consistent and asymptotically normal estimate of parameter $\mathbf{a} \in \mathbf{R}^p$.*

Proof: According to equality (3.5), it is valid

$$\hat{\mathbf{a}}_{_T} - \mathbf{a} = \left(\frac{1}{T-p} \mathbf{W}_{_T} \right)^{-1} \cdot \left(\frac{1}{T-p} \mathbf{r}_{_T} \right), \tag{3.6}$$

where

$$\mathbf{r}_{_T} = \left(\sum_{t=p+1}^{T} R_t W_{t-1} \ \sum_{t=p+1}^{T} R_t W_{t-2} \ \cdots \ \sum_{t=p+1}^{T} R_t W_{t-p} \right)' .$$

According to ergodicity of (W_t), which is valid to the set $\mathcal{A}$, follows the ergodicity of residuals (R_t). Then, under conditions of the theorem, using the ergodic theorem we have

$$\frac{1}{T-p} \mathbf{W}_{_T} \xrightarrow{\text{a.s.}} \mathbf{D}\,, \qquad T \to \infty\,, \tag{3.7}$$

where $\mathbf{D} = E\big(\mathbf{g}_t \mathbf{g}_t'\big)$, $\mathbf{g}_t = \big(W_{t-1} \ \cdots \ W_{t-p} \big)'$ and the moment-matrix $\mathbf{D}$ does not depend on $t \in \mathbf{Z}$, for any $\mathbf{a}$ from the stationarity set $\mathcal{A}$. According to ergodic theorem, also, it is valid that

$$\frac{1}{T-p} \mathbf{r}_{_T} \xrightarrow{\text{a.s.}} \mathbf{0}_{p\times 1}\,, \qquad T \to \infty\,,$$

and these two convergences, applied to equality (3.6), give

$$\hat{\mathbf{a}}_{_T} - \mathbf{a} \xrightarrow{\text{a.s.}} \mathbf{0}_{p\times 1}\,, \qquad T \to \infty\,, \tag{3.8}$$

i.e. the strict consistency of estimate $\hat{\mathbf{a}}_{_T}$ is proved.

Note further that the decomposition (3.6) can be written in the form of equality

$$\sqrt{T-p}\,\big(\hat{\mathbf{a}}_{_T} - \mathbf{a}\big) = \left(\frac{1}{T-p} \mathbf{W}_{_T} \right)^{-1} \cdot \left(\frac{1}{\sqrt{T-p}} \mathbf{r}_{_T} \right). \tag{3.9}$$

As for any $\mathbf{c} = (c_1 \cdots c_p)' \in \mathbf{R}^p$ the sequence

$$\mathbf{c}'\mathbf{r}_{_T} = \sum_{t=p+1}^{T} R_t \left(\sum_{j=1}^{p} c_j W_{t-j} \right)$$

is martingale, using the central limit theorem for martingale ([1]), we have

$$\frac{1}{\sqrt{T-p}}\, \mathbf{c}'\mathbf{r}_{_T} \xrightarrow{d} \mathcal{N}(0, \mathbf{c}'\mathbf{\Lambda}\mathbf{c})\,, \qquad T \to \infty\ ,$$

where $\mathbf{\Lambda} = E(\mathbf{u}_t \mathbf{u}_t')$, $\mathbf{u}_t = R_t\big(W_{t-1} \cdots W_{t-p}\big)'$ and $\mathbf{\Lambda}$ does not depend on t, for any $\mathbf{a} \in \mathcal{A}$. Now, using this convergence and the Cramér–Wold decomposition, we get

$$\frac{1}{\sqrt{T-p}}\, \mathbf{r}_{_T} \xrightarrow{d} \mathcal{N}(0, \mathbf{\Lambda})\,, \qquad T \to \infty\ .$$

Finally, according to (3.7) it is valid

$$\big(T-p\big)\, \mathbf{W}_{_T}^{-1} \xrightarrow{\text{a.s.}} \mathbf{D}^{-1}, \qquad T \to \infty\ ,$$

and, according to equality (3.9) and the last two convergences, we obtain

$$\sqrt{T-p}\,\big(\hat{\mathbf{a}}_{_T} - \mathbf{a}\big) \xrightarrow{d} \mathcal{N}\big(0, \mathbf{D}^{-1}\mathbf{\Lambda}\,\mathbf{D}^{-1}\big)\,, \qquad T \to \infty\ , \tag{3.10}$$

thus the theorem is completely proved. □

According to the obtained estimate $\hat{\mathbf{a}}_{_T}$, under the condition (2.16), the estimates of unknown parameters $\alpha_1, ..., \alpha_p$, b_c of GSB processes can be expressed as

$$\begin{cases} \hat{b}_c = \sum\limits_{j=1}^{p} \hat{a}_j \\ \hat{\alpha}_j = \hat{a}_j \hat{b}_c^{-1}\,, \qquad j = 1, ..., p\ . \end{cases} \tag{3.11}$$

Using the showed properties of estimate $\hat{\mathbf{a}}_{_T}$, we can prove that the estimates of "true" parameters have similar properties, which we formulate in the following assertion.

Theorem 3.4. *Let $\hat{\vartheta}_{_T} = (\hat{\alpha_1}, ..., \hat{\alpha}_p, \hat{b}_c)'$ be estimate of the unknown parameter $\vartheta = (\alpha_1, ..., \alpha_p, b_c)' \in \mathbf{R}^{p+1}$, obtained according to estimate $\hat{\mathbf{a}}_{_T}$ and equality* (3.11)*. If, for some $T_0 > 0$ and any $T \geq T_0$ is satisfied the condition*

$$\hat{b}_c \sum_{j=1}^{p} \hat{\alpha}_j < 1\ ,$$

then $\hat{\vartheta}_{_T}$ is strictly consistent and asymptotically normal estimate of ϑ.

Proof: According to convergence (3.8) and the continuity properties of almost sure convergence (see, for instance Serfling [14]) it is obviously valid that

$$\hat{\vartheta}_{_T} - \vartheta \xrightarrow{\text{a.s.}} \mathbf{0}_{p\times 1}\,, \qquad T \to \infty\ .$$

Notice that, for any $\mathbf{a} = (a_1, ..., a_p)' \in \mathbf{R}^p$, the expression (3.11) defines a mapping $g\colon \mathbf{R}^p \to \mathbf{R}^{p+1}$, by

$$\vartheta \,=\, g(\mathbf{a}) \,=\, \left(\, a_1 \Big(\sum_{j=1}^{p} a_j\Big)^{-1},\ \cdots,\ a_p \Big(\sum_{j=1}^{p} a_j\Big)^{-1},\ \sum_{j=1}^{p} a_j \,\right)'.$$

Then, applying the convergence (3.10) and continuity properties of asymptotically normal distributed random vectors (see, for instance Serfling [14]) we have

$$\sqrt{T-p}\,\big(\hat{\vartheta}_{_T} - \vartheta\big) \xrightarrow{d} \mathcal{N}(0, \mathbf{V})\,, \qquad T \to \infty\ ,$$

where $\mathbf{V} = \mathbf{G}\mathbf{D}^{-1}\mathbf{\Lambda}\,\mathbf{D}^{-1}\mathbf{G}'$ and $\mathbf{G} = \left(\dfrac{\partial g(\mathbf{a})}{\partial \mathbf{a}}\right)\Big|_{\mathbf{a}=\hat{\mathbf{a}}}$. □

At the end of this section, let us remark some more facts that directly follow from the estimation procedure described above and the theorems we have just proven.

Remark 3.1. Asymptotic variances $\mathbf{D}^{-1}\mathbf{\Lambda}\mathbf{D}^{-1}$ and $\mathbf{V}$ are commonly used as measures of bias of estimates $\hat{a}_{_T}$ and $\hat{\vartheta}_{_T}$, compared to the true values of parameters $\mathbf{a} = (a_1, ..., a_p)'$ and $\vartheta = (\alpha_1, ..., \alpha_p, b_c)$, where $\sum_{j=1}^{p} a_j = b_c$. Based on the introduced assumptions and the proof of previous theorem we have that

$$\mathbf{G}(\mathbf{a}) = \left(\frac{\partial g(\mathbf{a})}{\partial \mathbf{a}}\right) = \begin{pmatrix} b_c^{-1} - a_1 b_c^{-2} & -a_1 b_c^{-2} & \cdots & -a_1 b_c^{-2} \\ -a_2 b_c^{-2} & b_c^{-1} - a_2 b_c^{-2} & \cdots & -a_2 b_c^{-2} \\ \vdots & \vdots & & \vdots \\ -a_p b_c^{-2} & -a_p b_c^{-2} & \cdots & b_c^{-1} - a_p b_c^{-2} \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

and taking the standard matrix norms, we conclude that for all $\mathbf{a} \in \mathbf{R}^p$ is valid

$$\|\mathbf{G}(\mathbf{a})\|_1 = \big\|\mathbf{G}(\mathbf{a})'\big\|_\infty = b_c^{-1} - b_c^{-2}\sum_{j=1}^{p} a_j + 1 = 1\ .$$

Then, according to the previous equalities, for any matrix norm $\|\cdot\|$ which is sub–multiplicative to $\|\cdot\|_1$ and $\|\cdot\|_\infty$, we get

$$\|\mathbf{V}\| = \big\|\mathbf{G}(\hat{\mathbf{a}})\,\mathbf{D}^{-1}\mathbf{\Lambda}\,\mathbf{D}^{-1}\,\mathbf{G}(\hat{\mathbf{a}})'\big\| \le \big\|\mathbf{D}^{-1}\mathbf{\Lambda}\,\mathbf{D}^{-1}\big\|\ .$$

Therefore, asymptotic variance of estimates $\hat{\vartheta}_{_T}$, obtained by (3.11), does not exceed the asymptotic variance of $\hat{a}_{_T}$, under which they were calculated. In this way, the method of parameters estimation of the GSB model is formally justified.

Remark 3.2. If we apply estimates $\widetilde{b}_c$ and $\hat{b}_c$, we can make a modeled values of (ε_t), and thereby, we can estimate the variance σ^2 of the sequence (ε_t). To do this, we can use the sample variance, i.e. the estimates

$$\widetilde{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(X,\widetilde{\theta}) \qquad \text{or} \qquad \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(X,\hat{\theta})\,, \tag{3.12}$$

where $\varepsilon_t(X,\widetilde{\theta})$ and $\varepsilon_t(X,\hat{\theta})$ are modeled values of the white noise which we obtained by applying estimates $\widetilde{b}_c$ and $\hat{b}_c$, respectively (see, for more details, the following section). In the case of the Gaussian noise (ε_t), these estimates are identical to those which we can get applying the maximum likelihood procedure, as it was shown in Stojanović *et al.* [16]. Namely, under the assumption that (ε_t) is the Gaussian white noise, the log-likelihood function will be

$$L(y_1,...,y_T;\sigma^2) \;=\; -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\big(y_t - m_t\big)^2\,,$$

and we can easily see that the estimated value of the variance is identical to the sample variance (3.12) of the series (ε_t). The consistency and asymptotic normality of estimates $\widetilde{\sigma}$ and $\hat{\sigma}$ can be easily shown.

4. MONTE CARLO STUDY OF THE MODEL

In this section we will demonstrate some applications of the above described estimation procedure of Split-MA(1) and Split-MA(2) models. For the white noise (ε_t) it was used a simple random sample from with Gaussian $\mathcal{N}(0,1)$ distribution, so that the elements of the sequence (ε_t^2) were χ_1^2 distributed, which has been used for solving the critical value of the reaction $\widetilde{c}$ and $\hat{c}$.

In the case of Split-MA(1) process, as it is shown in Stojanović *et al.* [16], these estimates are based on 100 independent Monte Carlo simulations for each sample size $T = 50$, $T = 100$ and $T = 500$ for the model

$$X_t = \varepsilon_t - \theta_{t-1}\,\varepsilon_{t-1}\,, \qquad t = 1,...,T\,, \tag{4.1}$$

where $\theta_t = I(\varepsilon_{t-1}^2 \le 1)$ and $\varepsilon_0 = \varepsilon_{-1} \overset{\text{a.s.}}{=} 0$. Firstly, according to the correlation

$$\rho(h) \;=\; \mathrm{Corr}(X_{t+h},X_t) \;=\; \begin{cases} 1\,, & h = 0 \\ -b_c/(b_c+1)\,, & h = \pm 1 \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{4.2}$$

it obtained the estimates $\widetilde{b}_c = -\hat{\rho}_{_T}(1)\big(1+\hat{\rho}_{_T}(1)\big)^{-1}$, where $\hat{\rho}_{_T}(1)$ is the empirical first correlation of the sequence (X_t). After that, solving the equation $P\big\{\varepsilon_t^2 \le c\big\} = \widetilde{b}_c$ with respect to c, we obtained the estimates for the critical value $\widetilde{c}$.

In the next estimation stage, using the equality (4.1) in the "functional" form $\varepsilon_t(X,\theta) = X_t + \theta_{t-1}\,\varepsilon_{t-1}(X,\theta)$ and $\widetilde{b}_c$ as the initial (estimated) value of the parameter $b_c \in (0,1)$, the regression estimates $\hat{b}_c$ are computed as

$$\hat{b}_c \;=\; \left[\sum_{t=0}^{T-1} W_{t+1}(X,\widetilde{\theta})\, W_t(X,\widetilde{\theta})\right] \cdot \left[\sum_{t=0}^{T-1} W_t^2(X,\widetilde{\theta})\right]^{-1}, \tag{4.3}$$

where

$$W_t(X,\widetilde{\theta}) \;=\; \widetilde{\theta}_t\, W_{t-1}(X,\widetilde{\theta}) + \varepsilon_{t-1}(X,\widetilde{\theta})\,, \qquad \widetilde{\theta}_t = I(\varepsilon_{t-1}^2 \le \widetilde{c})\;,$$

and $\varepsilon_0(X,\widetilde{\theta}) = \varepsilon_{-1}(X,\widetilde{\theta}) \equiv 0$. Finally, the regression estimates of the (true) critical value $c = 1$ are solutions of the equation $P\{\varepsilon_t^2 \le c\} = \hat{b}_c$ with respect to c. The average values of these estimates are set, together with the correspondent estimating errors (the numeric values set in the parentheses) in the rows of Table 1.

Table 1: Estimated values of Monte Carlo simulations of the Split-MA(1) process.

Sample size	**Averages of estimated values**						
	$\hat{\rho}_{_T}(1)$	$\widetilde{b}_c$	$\widetilde{c}$	$\hat{b}_c$	$\hat{c}$	$\widetilde{\sigma}^2$	$\hat{\sigma}^2$
$T = 50$	−0.376 (0.139)	0.614 (0.219)	0.894 (0.726)	0.647 (0.192)	0.944 (0.571)	1.216 (0.292)	1.042 (0.202)
$T = 100$	−0.386 (0.097)	0.634 (0.156)	0.894 (0.444)	0.671 (0.141)	1.039 (0.427)	1.168 (0.184)	1.016 (0.124)
$T = 500$	−0.394 (0.056)	0.664 (0.091)	0.916 (0.259)	0.676 (0.068)	0.992 (0.194)	1.135 (0.102)	0.997 (0.099)
True values	**−0.406**	**0.683**	**1.000**	**0.683**	**1.000**	**1.000**	**1.000**

The second column of Table 1 contains the estimated values of the coefficient of the first correlation $\hat{\rho}_{_T}(1)$ of the Split-MA(1) model. The average values of that column are somewhat smaller in the absolute value of the true value, which is the case here $\rho(1) = -b_c(1+b_c)^{-1} \approx -0.406$. In that way, estimates $\widetilde{b}_c$ and $\widetilde{c}$, showed in the next two columns, will be a proper estimates if $-0,5 < \hat{\rho}_{_T}(1) < 0$. Then, we showed the regression estimates $\hat{b}_c$ and $\hat{c}$ which average values are closer to the true value than previously mentioned, initial estimates $\widetilde{c}$. This is due to the fact, formally proved in Remark 3.1, that estimates $\hat{c}$ are more efficient than $\widetilde{c}$. The histograms of empirical distributions of the estimates $\widetilde{c}$ and $\hat{c}$ are shown in Figure 1. It can be seen that $\hat{c}$ has the asymptotically normal distribution even for the sample of a "small" sample size. Finally, the averages of estimated values of σ^2, based on modeled values of the white noise (ε_t) and equations (3.12), are

displayed in the last two columns of Table 1. Their average values differ from the true value $\sigma^2 = 1$ as a consequence of two stage estimating procedure that was used. In spite of that, it can be seen that the average values of $\hat{\sigma}^2$ are closer to the true value one than the average values of $\widetilde{\sigma}^2$.

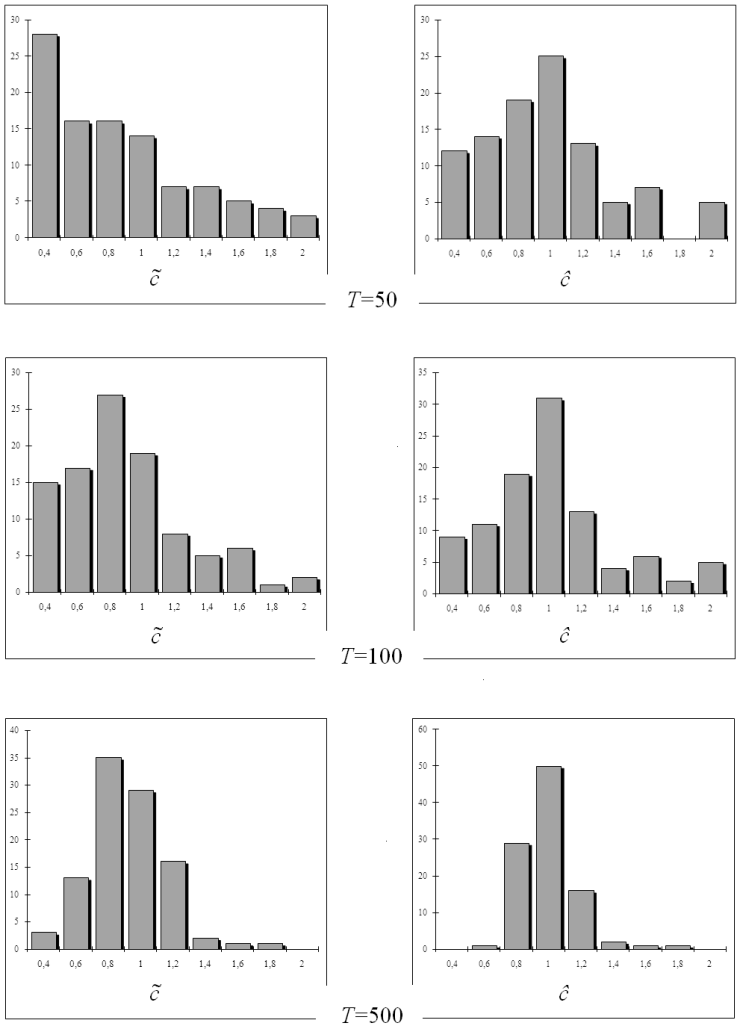

Figure 1: Empirical distributions of estimated parameters $\widetilde{c}$ and $\hat{c}$ of Split-MA(1) model

As in the case of of Split-MA(1) process, we are able to apply the above procedure for estimating the unknown parameters of Split-MA(2) process

$$X_t = \varepsilon_t - \alpha_1 \theta_{t-1} \varepsilon_{t-1} - \alpha_2 \theta_{t-2} \varepsilon_{t-2}\,, \qquad t = 1, ..., T\,. \tag{4.4}$$

For this purpose, we used 45 independent Monte Carlo simulations of this series of the length $T = 500$, with values of parameters $\alpha_1 = 0.6$, $\alpha_2 = 0.4$ and $c = 1$.

Firstly, by equating the correlation functions $\gamma_X(h)$ of the model above with its empirical correlations $\hat{\gamma}_X(h)$, where $h = 0, 1, 2$, we get the estimate

$$\widetilde{\alpha}_1 = \frac{-\hat{\gamma}_X(1) - \sqrt{\hat{\gamma}_X^2(1) + 4\,\hat{\gamma}_X(1)\,\hat{\gamma}_X(2)}}{2\,\hat{\gamma}_X(2)},$$

$$\widetilde{\alpha}_2 = 1 - \widetilde{\alpha}_1 = \frac{2\,\hat{\gamma}_X(1) + \hat{\gamma}_X(1) + \sqrt{\hat{\gamma}_X^2(1) + 4\,\hat{\gamma}_X(1)\,\hat{\gamma}_X(2)}}{2\,\hat{\gamma}_X(2)},$$

$$\widetilde{b}_c = \frac{-\hat{\gamma}_X(2)}{\left(\tilde{\alpha}_1^2 + \tilde{\alpha}_2^2\right)\hat{\gamma}_X(2) + \tilde{\alpha}_2\,\hat{\gamma}_X(0)}.$$

After that, using $\widetilde{\alpha}_1$, $\widetilde{\alpha}_2$ and $\widetilde{c}$ as initial estimates we can generate the sequences

$$\begin{cases} \varepsilon_t(X, \widetilde{\theta}, \widetilde{\alpha}) = X_t + \widetilde{\alpha}_1 \widetilde{\theta}_{t-1}\,\varepsilon_{t-1}(X, \widetilde{\theta}, \widetilde{\alpha}) + \widetilde{\alpha}_2\,\widetilde{\theta}_{t-1}\,\varepsilon_{t-2}(X, \widetilde{\theta}, \widetilde{\alpha}) \\ \widetilde{\theta}_t = I\big(\varepsilon_{t-1}^2(X, \widetilde{\theta}, \widetilde{\alpha}) \le \widetilde{c}\big) \\ W_t(X, \widetilde{\theta}, \widetilde{\alpha}) = \widetilde{\alpha}_1 \widetilde{\theta}_t W_{t-1}(X, \widetilde{\theta}, \widetilde{\alpha}) + \widetilde{\alpha}_2\,\widetilde{\theta}_{t-1} W_{t-2} + \varepsilon_{t-1}(X, \widetilde{\theta}, \widetilde{\alpha}), \end{cases}$$

where $t = 1, ..., 500$ and $\varepsilon_0 = \varepsilon_{-1} \overset{\text{a.s.}}{=} 0$. Finally, according to the equalities (3.5) and (3.11) we obtain the regression estimates of appropriate parameters $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{c}$ of Split-MA(2) model.

Table 2: Estimated values of Monte Carlo simulations of the Split-MA(2) process.

Estimators type	**Parameters**						
	a_1	a_2	α_1	α_2	b_c	c	σ^2
Initial estimates	—	—	0.597	0.403	0.673	0.993	1.016
	—	—	(0.043)	(0.043)	(0.075)	(0.162)	(0.219)
Regression estimates	0.421	0.274	0.605	0.394	0.685	1.006	1.045
	(0.025)	(0.025)	(0.029)	(0.029)	(0.031)	(0.145)	(0.143)
True values	**0.420**	**0.273**	**0.600**	**0.400**	**0.683**	**1.000**	**1.000**

The Table 2 shows the average values of obtained estimates, corresponding estimating errors and the true values of parameters. At the first glance, there are no major differences in the quality of the obtained estimates. Moreover, regression estimates are slightly more different from the real values of the parameters, previously obtained from the initial evaluations. However, the dispersion of the regression estimates is much smaller than the dispersion of the initial estimates, and this is one of the important advantages of this estimation method. This fact is clearly visible in Figure 2, which shows the histograms of empirical distributions of both types of estimates. Obviously, the histograms of regression estimates (panels below) have much more pronounced asymptotic tendencies in relation to the initial estimates of parameters (panels above).

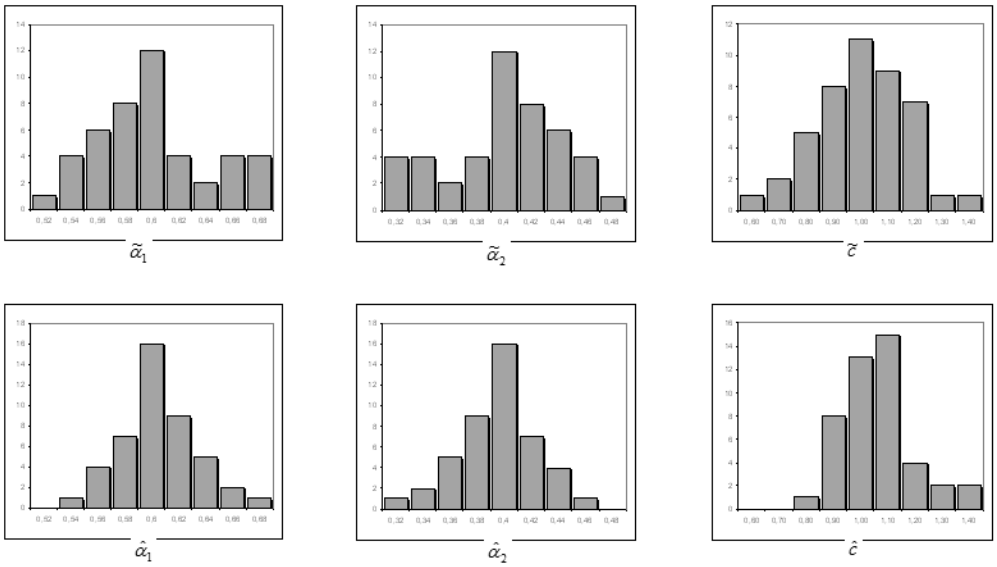

Figure 2: Empirical distributions of estimated parameters of Split-MA(2) model.

5. APPLICATION OF THE MODEL

Here we describe some of the possibilities of practical application the GSB process in the modeling dynamics of financial series. We observe Split-MA(1) and Split-MA(2) models, described by equalities (4.1) and (4.4), as stochastic models of dynamics the total values of stocks trading on the Belgrade Stock Exchange. As a basic financial sequence we observe the realization of log-volumes

$$y_t = \ln(S_t \cdot H_t)\,, \qquad t = 0, 1, ..., T\,, \tag{5.1}$$

where S_t is the share price and H_t is the volume of trading of the same share at time $t = 1, ..., T$. (The price is in dinars and the volume is the number of shares that were traded on the certain day. The days of trading are used as successive data.) Firstly, we applied iterative equations

$$\begin{cases} \varepsilon_t = y_t - m_t \\ m_t = m_{t-1} + \varepsilon_{t-1} I(\varepsilon_{t-2}^2 > \hat{c}) \end{cases}, \qquad t = 1, ..., T\,, \tag{5.2}$$

to generate the corresponding values of sequences (ε_t) and (m_t) of Split-BREAK process of order $p = 1$. As estimates of the critical value $\hat{c}$, we used the previous estimating procedure, and as initial values of the iterative procedure (5.2) we use $m_0 = y_0 = \overline{y}_T$, $\varepsilon_0 = \varepsilon_{-1} \overset{\text{a.s.}}{=} 0$, where $\overline{y}_T$ is the empirical mean of (y_t). We use the basic empirical series defined by (5.1) in solving the series of increments (X_t), i.e. the realized values of Split-MA(1) described above. The similar procedure can be used to estimate the parameters of Split-MA(2) model. In that case, we

substitute the second equation in (5.2) with

$$m_t \;=\; \hat{\alpha_1}\big(m_{t-1} + \varepsilon_{t-1} I(\varepsilon_{t-2}^2 > \hat{c})\big) + \hat{\alpha_2}\big(m_{t-2}\,\varepsilon_{t-2} I(\varepsilon_{t-3}^2 > \hat{c})\big), \qquad t = 2, ..., T,$$

where $\hat{\alpha}_1, \hat{\alpha}_2$ is the estimated values of model's parameters, with $\hat{\alpha}_1 + \hat{\alpha}_2 = 1$ and $\varepsilon_0 = \varepsilon_{-1} \overset{\text{a.s.}}{=} 0$. Table 3 contains the number of observations for the company (T), and estimated values of Split-MA(1) and Split-MA(2) models in the case of six Serbian eminent companies.

Table 3: Estimated values of the GSB parameters of real data.

Companies	**Cities**	T	$p=1$				
			$\hat{\rho}_T(1)$	$\widetilde{b}_c$	$\widetilde{c}$	$\hat{b}_c$	$\hat{c}$
HEMOFARM	Vršac	54	−0.346	0.530	0.582	0.613	0.836
METALAC	Milanovac	174	−0.449	0.816	4.929	0.829	5.223
SUNCE	Sombor	157	−0.424	0.735	2.836	0.784	3.132

Companies	**Cities**	T	$p=2$					
			$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{c}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{c}$
ALFA PLAM	Vranje	50	0.640	0.360	2.628	0.690	0.310	3.331
DIN	Niš	56	0.715	0.285	1.393	0.816	0.184	1.202
T. MARKOVIĆ	Kikinda	277	0.824	0.176	1.396	0.830	0.170	1.392

The following, Table 4 contains estimated values of means and variances of previously defined sequences: log–volumes (y_t), martingale means (m_t), the Split-MA process (X_t) and the white noise (ε_t). If we analyze empirical values of these series, we can recognize the relations that could be explained the theoretical results above. Namely, the averages of the log–volumes are close to the averages of martingale means, which is in accordance with (2.8), i.e. to the fact that the realizations of (y_t) are "close" to the sequence (m_t). On the other hand, the averages of (X_t) and (ε_t) are "close" to zero, which is consistent with previous theoretical results. Also, the estimates of the empirical variances of Split-MA series are generally higher than the corresponding values of the noise variances, which is consistent with the theoretical properties of these sequences (see Theorem 2.1).

Table 4: Estimated values of real data.

Companies	**Log-volumes**		**Mart. means**		**Split-MA**		**White noise**	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var
HEMOFARM	15.250	0.814	15.310	0.694	0.022	1.786	−0.042	1.576
METALAC	13.665	2.788	13.798	2.731	0.001	3.979	−0.005	3.614
SUNCE	12.748	2.282	12.730	2.151	−0.024	1.978	−0.028	1.981
ALFA PLAM	15.320	1.505	15.354	1.457	−0.126	2.410	−0.005	1.590
DIN	14.485	4.998	14.628	6.071	−0.126	3.003	−0.139	2.868
T. MARKOVIĆ	13.816	2.295	13.830	1.977	0.001	4.002	−0.078	3.611

A high correlation between the log-volumes and the martingale means can be seen in Figure 3, which represents the realizations of these sequences. Obviously, this fact concurs with the definition of the GSB process, i.e., the equation (2.6), and justifies the application of the GSB process as an appropriate stochastic model of the dynamics of empirical time series.

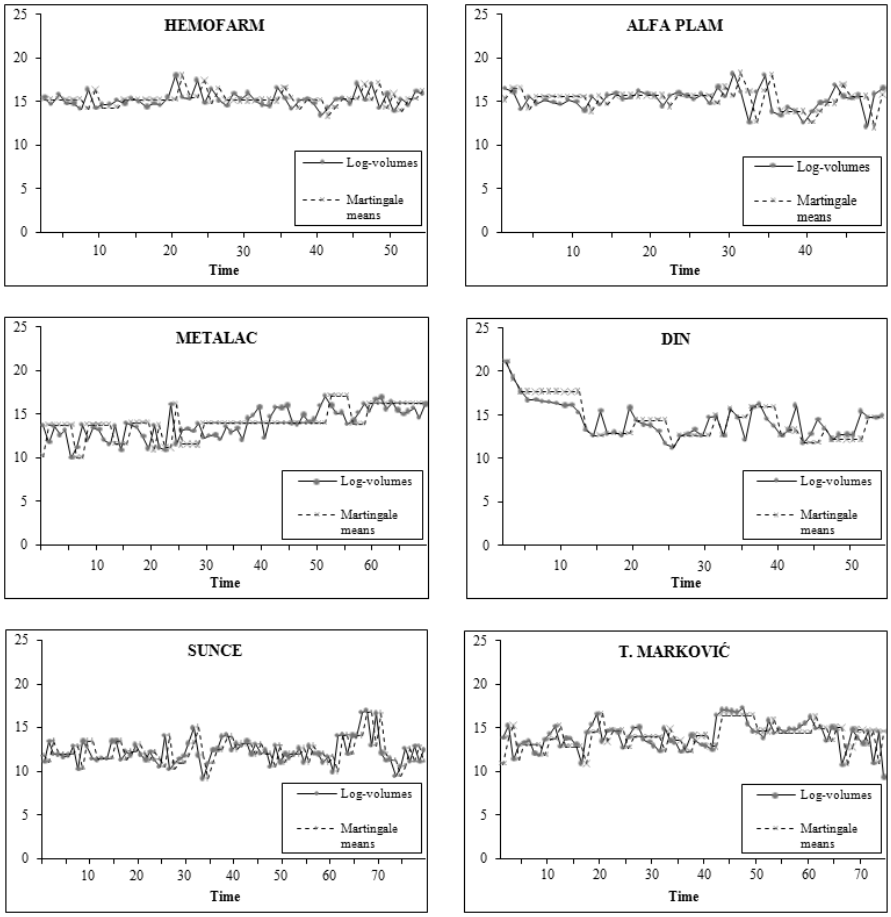

Figure 3: Comparative graphs of the real and modeled data.

Finally, Figure 4 shows that there is also a strong correlation between the white noise (ε_t) and the increments (X_t). It is clear that the concurrence of realizations of these two sequences will be better if, in addition to the great fluctuation of (X_t), the critical value of the reaction c is relatively small (see, for instance, Section 4). In fact, small values of c point out to the possibility that the true value of this parameter is $c = 0$, when increments (X_t) are equal to the noise (ε_t). In that case (y_t) is the sequence with independent increments and the whole statistical analysis is made easier. According to the previous facts about asymptotic normality of obtained estimates, testing the null hypothesis $H_0 : c = 0$, (i.e. $b_c = 0$), in the case of the "large" sample size, will be based on the normal distribution, i.e. standard, well known statistical tests.

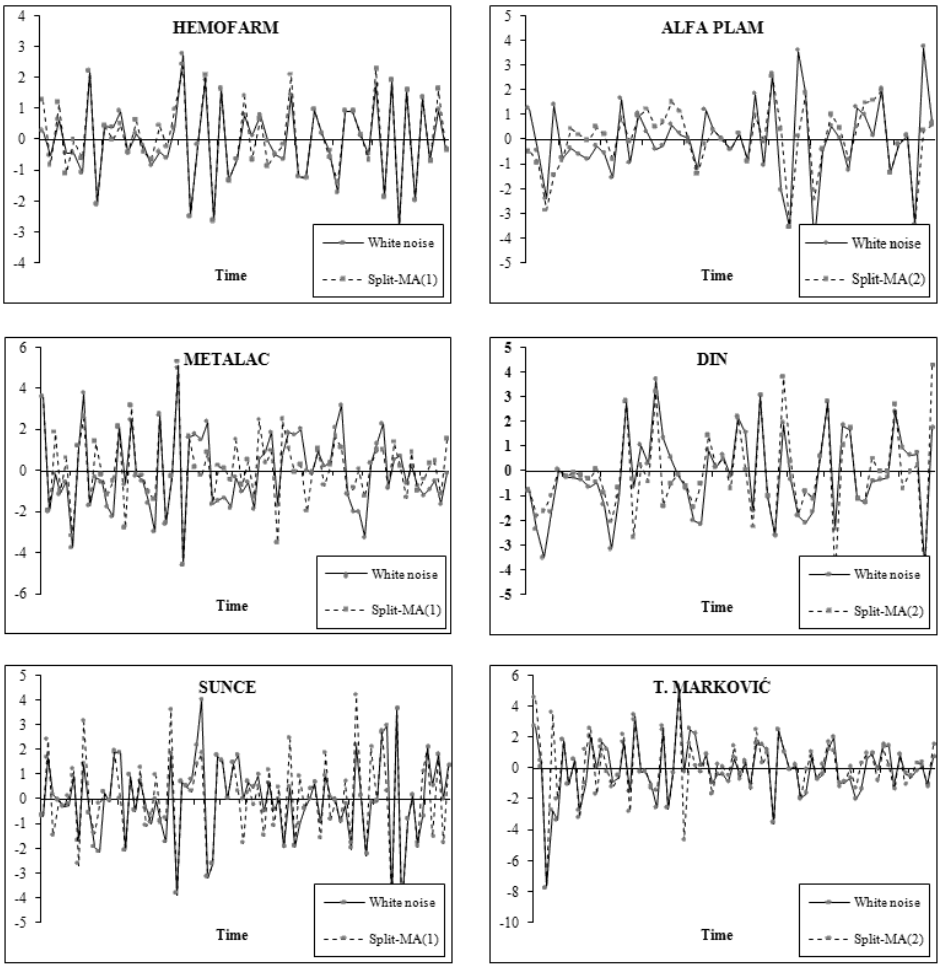

Figure 4: Comparative graphs of real and modeled data.

6. CONCLUSION

As we know, the non-linear stochastic models of financial time series usually give excellent results in explaining many aspects of their behavior. In this sense, various modifications of the STOPBREAK process enable successful description of the dynamics of financial time series with emphatic permanent fluctuations. We should point out once again Stojanović *et al.* [16], where were compared the efficiency between the simplest GSB model (named Split-BREAK model) of order $p = 1$ and some well known models which are standardly used in the real data modeling. Using the same data set as in the section above, it is shown that our process represents these time series better, and that fewer coefficients need to be estimated in comparison with well known models used so far.

ACKNOWLEDGMENTS

This research was supported in part by the Serbian Ministry of Education, Science and Technological Development (No. #OI 174007 and 174026).

REFERENCES

[1] Billingsley, P. (1961). The Lindeberg–Levy theorem for martingales, *Proceedings of the American Mathematical Society*, **12**(5), 788–792.

[2] Chan, S.K. (1993). Consistency and limiting distribution of the lest squares esimator of a thresholds autoregressive model, *Annals of Statistics*, **21**, 520–533.

[3] Diebold, F.X. (2001). Long memory and regime switching, *Journal of Econometrics*, **105**(1), 131–159.

[4] Engle, R.F. and Smith, A.D. (1999). Stochastic permanent breaks, *The Review of Economics and Statistics*, **81**, 553–574.

[5] Francq, C.; Roussignol, M. and Zakoian, M.J. (2001). Conditional heteroskedasticity driven by hidden markov chains, *Journal of Time Series Analysis*, **22**(2), 197–220.

[6] González, A. (2004). A smooth permanent surge process, *SSE/EFI Working Paper Series in Economics and Finance*, **572**, Stockholm.

[7] Gonzalo, J. and Martinez, O. (2006). Large shocks vs. small shocks (Or does size matter? May be so), *Journal of Econometrics*, **135**, 311–347.

[8] Hansen, B. (2000). Sample splitting and threshold estimation, *Econometrica*, **68**(3), 575–603.

[9] Huang, B.-N. and Fok, R.C.W. (2001). Stock market integration-an application of the stochastic permanent breaks model, *Applied Economics Letters*, **8**(11), 725–729.

[10] Kapetanios, G. and Tzavalis, E. (2010). Modeling structural breaks in economic relationships using large shocks, *Journal of Economic Dynamics and Control*, **34**(3), 417–436.

[11] Lawrence, A.J. and Lewis, P.A.W. (1992). Reversed residuals in autoregressive time series analysis, *Journal of Time Series Analysis*, **13**, 253–266.

[12] Popović, B.Č. and Stojanović, V.S. (2005). Split-ARCH, *PLISKA Studia Mathematica Bulgarica*, **17**, 201–220.

[13] Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *RevStat – Statistical Journal*, **10**(1), 33–60.

[14] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, Inc.

[15] Stojanović, V.S. and Popović, B.Č. (2007). Estimation in real data set by Split-ARCH model, *Filomat*, **21**(2), 133–152.

[16] Stojanović, V.S.; Popović, B.Č. and Popović, P.M. (2011). The Split-BREAK model, *Brazilian Journal of Probability and Statistics*, **25**(1), 44–63.

[17] Stojanović, V.S.; Popović, B.Č. and Popović, P.M. (2014). Stochastic analysis of GSB process, *Publications de l'Institut Mathématique*, **95**(109), 149–159.

[18] Tong, H. (1980). *Thresholds Models in Non-linear Time Series Analysis*, Lecture Notes in Statistics, **21**, Springer-Verlag, Berlin, Heidelberg, New York.

REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 169–191

A BAYESIAN APPROACH FOR JOINT MODELING OF SKEW-NORMAL LONGITUDINAL MEASUREMENTS AND TIME TO EVENT DATA

Authors: Taban Baghfalaki
– Department of Statistics, Faculty of Mathematical Sciences, Tardit Modares University, Tehran, Iran
t.baghfalaki@modares.ac.ir

Mojtaba Ganjali
– Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
m-ganjali@sbu.ac.ir

Received: April 2013 Revised: December 2013 Accepted: February 2014

Abstract:

- Joint modeling of longitudinal measurements and survival time has an important role in analyzing medical data sets. For example, in HIV data sets, a biological marker such as CD4 count measurements is considered as a predictor of survival. Usually, longitudinal responses of these studies are severely skew. An ordinary method for reducing the skewness is the use of square root or logarithm transformations of responses. In most of the HIV data sets, because of high rate of missingness, skewness is remained even after using the transformations. Therefore, a general form of distributions for considering skewness in the model should be used. In this paper, we have used multivariate skew-normal distribution to allow a flexible model for considering non-symmetrically of the responses. We have used a skew-normal mixed effect model for longitudinal measurements and a Cox proportional hazard model for time to event variable. These two models share some random effects. A Bayesian approach using Markov chain Monte Carlo is adopted for parameter estimation. Some simulation studies are performed to investigate the performance of the proposed method. Also, the method is illustrated using a real HIV data set. In these data, longitudinal outcomes are skew and death is considered as the event of interest. Different model structures are developed for analyzing this data set, where model selection is performed using some Bayesian criteria.

Key-Words:

- *Bayesian approach; Cox proportional model; joint modeling; longitudinal data; skew-normal distribution; time to event data.*

AMS Subject Classification:

- 62H99, 62J99.

1. INTRODUCTION

In most of the HIV and cancer studies a longitudinal biological marker such as CD4 count or immune response can be an important predictor of survival. In these studies a time to an event may also be a variable of interest. Patients are monitored longitudinally and some longitudinal measurements are gathered until the interest event occur. Often the longitudinal outcomes and time to event of interest are analyzed jointly using joint modeling of longitudinal and time to event data.

Joint modeling of longitudinal measurements and time to event data has been studied by DeGruttola and Tu (1994), Tsiatis *et al.* (1995) and Wulfsohn and Tsiatis (1997). Henderson *et al.* (2000) and Hashemi *et al.* (2003) discussed joint modeling of longitudinal measurements and event time data using latent class of Gussian process, Tsiatis and Davidian (2001), Yu *et al.* (2004) and Sousa (2011) provide reviews of this joint modeling. Tseng *et al.* (2005) used accelerated failure time model for joint modeling of longitudinal and survival data and applied Monte Carlo EM approach to estimate unknown parameters. Also, Diggle *et al.* (2008) discussed different approaches to estimate unknown parameters of joint modeling of longitudinal measurements and event time data and then applied a fully parametric approach to modeling Schizophrenic patients data set. Joint Modeling of longitudinal measurements and time to event data at the presence of informative dropout in a HIV study was discussed by Wu *et al.* (2008). They considered an additional missingness mechanism for missing values. Rizopoulos (2010) presented the R package "*JM*" that can be used to fit the joint modeling of longitudinal measurements and survival data. Also, Guo and Carlin (2004) discussed the implementation of the joint models in SAS and WinBUGS under normal distributional assumption.

In the above mentioned references, usually a mixed effect model with normality or other symmetric distributional assumption is used for longitudinal part of the model. However, in the most of such studies longitudinal measurements are severely skew or have some outliers. For the latter problem, some models, which are robust in the existence of outliers, have been considered by Lachos *et al.* (2010) and Bandyopadhyay *et al.* (2010). They have used skew-normal/independent distribution assumptions in linear mixed effect models. However, the problem of skewness, to our best of knowledge has not yet been considered in joint modeling of longitudinal and time to event data. The idea is that a parametric skew distribution may be considered for this regard.

Skew-normal (SN) family is an important class of non-symmetric distribution for analyzing abnormal data set. The distribution includes normal one as a special case. The first version of the distribution which is in the univariate form is introduced by Azzalini (1985). More discussion about univariate skew-normal

distribution can be found in Azzalini (1986) and Henez (1986). Generalizations to the multivariate case are given in Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001) and Sahu *et al.* (2003). For examples, some applications of skew-normal in regression model can be found in Lachos *et al.* (2007), Cancho *et al.* (2010) and Arellano-Valle *et al.* (2005b). Multivariate skew-normal mixed effect model have discussed by Arellano-Valle *et al.* (2005a) and Lin and Lee (2008), also, discussion about multivariate skew-normal with incomplete data can be found in Lin *et al.* (2009) and Baghfalaki and Ganjali (2011, 2012). Recently, there are some applications of skew-normal distribution for analysing HIV data. For examples: Ghosh and Branco (2007) develop a Bayesian approach to bivariate random effect model with application to HIV studies. Huang and Dagne (2011) used skew-normal distribution in a Bayesian approach to joint modeling mixed effect and measurement error for a HIV study. Huang and Dagne (2010) developed a Bayesian nonlinear mixed effect model with skew-normal random effect and within subject errors for providing a better fit to HIV data set, Huang *et al.* (2011a) suggested linear, nonlinear and semiparametric mixed-effect model using skew-normal distribution with measurement error in covariates for analyzing an AIDS data set and Lachos *et al.* (2011) developed a Bayesian framework for analyzing censored data using linear or non-linear model under skew-normal/independent distributional assumption with application in HIV studies. Huang *et al.* (2011b) used skew-normal distribution for joint modeling of CD4 process and time to increase CD4/CD8 ratio. They used a mixed effect model for analyzing the longitudinal measurements alongside a log-normal model for analyzing event time data.

In this paper, we have discussed Bayesian joint modeling of longitudinal and survival data when skewness exists in the data set. We have used multivariate skew-normal distribution introduced by Sahu *et al.* (2003) for considering skewness of the data. Implementation of the Bayesian approach using this form is easier than other forms of skew-normal distributions. A non-ignorable missingness mechanism is considered for missingness. Also, a skew-normal mixed effect model and a Cox proportional hazard model (as a semiparametric model) with step baseline hazard in a frailty model structure are considered for the joint modeling. We have performed some simulation studies to investigate the performance of the proposed method with different sample sizes and different rates of drop out. We have used the proposed method for analyzing a HIV study, where CD4 count measurements are longitudinal measurements and time to death is considered as the interest event. The aim of the study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). In this data set CD4 count measurement is a skew variable which is gathered along side with the time to the event of interest. We have used the proposed model and pure normal model for analyzing the data set. The results of using these distributional assumptions are compared using some criteria; also some influential observations are detected using Kullback–Leibler divergence.

The rest of this paper is organized as follows. In Section 2, we introduce multivariate skew-normal distribution which we will use in this paper. Section 3 includes the model and notations of the paper. In that Section, the model for longitudinal and survival part is described separately. In Section 4 Bayesian approach of joint modeling of longitudinal measurements and event time data using multivariate skew-normal distribution is discussed. Section 5 includes some simulation studies for investigating the proposed model. In Section 6, we apply the proposed approach to the HIV data set and finally concluding remarks are given in Section 7.

2. MULTIVARIATE SKEW-NORMAL DISTRIBUTION

Multivariate skew-normal distributions have different forms, some of these distributions have been introduced by: Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Arellano-Valle and Genton (2005) and Arellano-Valle *et al.* (2005b). One of the commonly used multivariate skew-normal distributions, in Bayesian context, is introduced by Sahu *et al.* (2003). In this section, we review this form.

Let $\phi_k(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\Phi_k(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ be the probability density function and cumulative density function of the $N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$ evaluated at $\boldsymbol{y}$, respectively. A k-dimensional random vector $\boldsymbol{Y}$ follows a k-variate skew-normal distribution with location vector $\boldsymbol{\mu} \in R^k$, $k\times k$ positive definite scale matrix $\boldsymbol{\Sigma}$ and $k\times k$ skewness matrix $\boldsymbol{\Delta} = \text{diag}(\delta_1, ..., \delta_k)$, where $\text{diag}(a_1, ..., a_k)$ denotes a diagonal matrix with elements $a_1, ..., a_k$, if its density function is given by:

$$
\begin{aligned}
f\big(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta}\big) \;=\; & 2^k\,\phi_k\big(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}'\big) \\
& \times\; \Phi_k\Big(\boldsymbol{\Delta}'\big(\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}'\big)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\,\big|\,0\,,\;\big(\boldsymbol{I}_k+\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\big)^{-1}\Big)\,.
\end{aligned}
\tag{2.1}
$$

We denote this by $\boldsymbol{Y} \sim SN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta})$. The mean and covariance matrix of $\boldsymbol{Y}$ are given, respectively, by:

$$
E[\boldsymbol{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\,\boldsymbol{\delta} \quad \text{and} \quad \text{cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} + \left(1-\frac{2}{\pi}\right)\boldsymbol{\Delta}^2\,,
$$

where $\boldsymbol{\delta} = (\delta_1, ..., \delta_k)'$. The use of the following proposition which is called stochastic representation, makes it possible to generate a sample from skew-normal distribution using available software.

Proposition 2.1. *Let* $\boldsymbol{Y} \sim SN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta})$, *then* $\boldsymbol{Y} \overset{\text{d}}{=} \boldsymbol{\Delta}|\boldsymbol{X}_0| + \boldsymbol{X}_1$, $\boldsymbol{X}_0 \sim N_k(0,\boldsymbol{I}_k)$, $\boldsymbol{X}_1 \sim N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\boldsymbol{X}_0$ *and* $\boldsymbol{X}_1$ *are independent. The notation* $\overset{\text{d}}{=}$ *means "distributed as".* $|\boldsymbol{X}_0|$ *is the vector of the absolute values of each component of* $\boldsymbol{X}_0$.

For proof of this proposition see Sahu *et al.* (2003) and Arellano-Valle *et al.* (2007).

This proposition is used for obtaining a hierarchical set-up, that is $\boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} \sim N_k(\boldsymbol{\mu} + \boldsymbol{\Delta u}, \boldsymbol{\Sigma})$ and $\boldsymbol{U} \sim N_k(\boldsymbol{0}, \boldsymbol{I}_k)$. This hierarchical set-up has been used broadly in Bayesian context of skew-normal model.

3. NOTATION AND SEPARATE MODELS

In this section the notations and models for each part of JM are described. The next section will discuss joint modeling (JM) of longitudinal responses and time to event data with skew-normal distribution assumption for responses.

For longitudinal model, let $y_i(s)$ denote the value of longitudinal outcome at time point s for the i^{th} individual where the observed times are s_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., n_i$. In this case we shall write $y_i(s_{ij}) = y_{ij}$. We consider the following linear mixed effect model for describing longitudinal outcome:

$$y_{ij} = \boldsymbol{x}'_{1i}(s_{ij})\boldsymbol{\beta}_1 + \boldsymbol{z}'_{1i}(s_{ij})\boldsymbol{b}_{1i} + \boldsymbol{\varepsilon}_{ij} \ , \qquad i = 1, 2, ..., n \ , \quad j = 1, 2, ..., n_i \ ,$$

where components of $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{in_i})'$ are measurement errors, $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{1p_1})'$ is a p_1-dimensional vector of longitudinal fixed-effect parameters. $\boldsymbol{b}_{1i} = (b_{1i1}, ..., b_{1iq_1})'$ is a q_1-dimensional vector of random effects and is independent of $\boldsymbol{\varepsilon}_i$. $\boldsymbol{x}_1$ and $\boldsymbol{z}_1$ are p_1-dimensional and q_1-dimensional vectors of explanatory variables, respectively. In the matrix notation

$$\boldsymbol{Y}_i = \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{Z}_{1i}\boldsymbol{b}_{1i} + \boldsymbol{\varepsilon}_i \ , \tag{3.1}$$

where in this notation $\boldsymbol{Y}_i$ is the longitudinal vector of response variable for the i^{th} subject. $\boldsymbol{X}_{1i} = \big(\boldsymbol{x}_{1i}(s_{i1}), ..., \boldsymbol{x}_{1i}(s_{in_i})\big)'$ and $\boldsymbol{Z}_{1i} = \big(\boldsymbol{z}_{1i}(s_{i1}), ..., \boldsymbol{z}_{1i}(s_{in_i})\big)'$. We assume that $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} SN_{n_i}\Big(-\sqrt{\frac{2}{\pi}}\boldsymbol{\delta}_e, \boldsymbol{\Psi}, \boldsymbol{\Delta}_e\Big)$ and $\boldsymbol{b}_{1i} \overset{\text{iid}}{\sim} N_{q_1}(0, \boldsymbol{D}_1)$. Note that these assumptions gives $E[\boldsymbol{\varepsilon}_i] = E[\boldsymbol{b}_{1i}] = \boldsymbol{0}$. Thus, this model considers the random effects $\boldsymbol{b}_{1i}$ to be symmetrically distributed, while the distribution of the within subject errors $\boldsymbol{\varepsilon}_i$ to be asymmetric with mean zero. To seek for identifiability (Arellano-Valle *et al.*, 2007), we assume $\boldsymbol{\Psi} = \sigma_e^2\boldsymbol{I}_{n_i}$, also, $\boldsymbol{\Delta}_e = \delta_e\boldsymbol{I}_{n_i}$.

In survival model, let T_i^* be the true event time and C_i be the censoring time. $T_i = \min(T_i^*, C_i)$ denotes the observed survival time for the i^{th} individual, $i = 1, 2, ..., n$. Also, $\delta_i = I(T_i^* \leq C_i)$ is a censoring indicator, which is 0 for right-censored and 1 for complete observed individuals. Therefore, the observed data for the survival outcome consist of the pairs $\big\{(T_i, \delta_i), \ i = 1, 2, ..., n\big\}$.

For survival modeling a frailty model which is linked to the longitudinal model through some shared random effects is considered. The hazard function in

our proposed model is given by:

$$h\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big) = h_0(t_i)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\}, \tag{3.2}$$

where $h_0(t_i)$ is the baseline hazard function. Thus, the density function of survival time for the i^{th} individual can be written as:

$$h^{\delta_i}\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big)\times\exp\Big\{-H_0(t_i)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\}\Big\},$$

where $H_0(t)=\int_0^t h_0(u)\,du$, $\boldsymbol{x}_2$ and $\boldsymbol{z}_2$ are p_2- and q_2-dimensional vectors of explanatory variables, respectively. $\boldsymbol{\beta}_2=(\beta_{21},...,\beta_{2p_2})'$ is a p_2-dimensional vector of time to event fixed effect parameters and $\boldsymbol{b}_{2i}=(b_{2i1},...,b_{2iq_2})'$ is a q_2-dimensional of random effects of time to event process where we assume, $\boldsymbol{b}_{2i}\overset{\text{iid}}{\sim}N_{q_2}(0,\boldsymbol{D}_2)$. In the next section, for getting sure of identifiability we shall impose $\boldsymbol{b}_{2i}$ to share some components with $\boldsymbol{b}_{1i}$.

The longitudinal outcome $\boldsymbol{y}_i$ can be partitioned into $\boldsymbol{y}_{i,\text{obs}}=\{y_i(s_{ij})\colon s_{ij}<T_i,\ j=1,2,...,n_i\}$, which contains all observed longitudinal measurements for the i^{th} individual before the observed event time T_i, and $\boldsymbol{y}_{i,\text{mis}}=\{y_i(s_{ij})\colon s_{ij}\geq T_i,\ j=1,2,...,n_i\}$ which contains the longitudinal measurements that should have been taken until the end of the study and has to be considered as a vector containing missing values.

4. THE SKEW-NORMAL JOINT MODELING (SNJM) OF LONGITUDINAL AND SURVIVAL DATA

In our proposed joint modeling, we have considered a skew-normal distribution for error terms of longitudinal measurements. However, because of non-identifiability of some parameters, we have not considered skew-normal distribution assumption for random effects in these models. The skew-normal joint modeling (SNJM) of longitudinal and survival data, as an extension of the usual normal joint modeling, leads us to the following hierarchical model:

$$\begin{aligned}
\boldsymbol{Y}_i|\boldsymbol{b}_{1i},\boldsymbol{\beta}_1,\sigma_e^2,\delta_e &\overset{\text{ind.}}{\sim} SN_{n_i}\Big(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1+\boldsymbol{Z}_{1i}\boldsymbol{b}_{1i}-\sqrt{\frac{2}{\pi}}\,\delta_e\mathbf{1}_{n_i},\,\sigma_e^2\boldsymbol{I}_{n_i},\,\delta_e\boldsymbol{I}_{n_i}\Big),\\
\boldsymbol{b}_{1i}|\boldsymbol{D}_1 &\overset{\text{ind.}}{\sim} N_{q_1}(\mathbf{0},\boldsymbol{D}_1),\\
h\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big) &= h_0(t)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\},\\
\boldsymbol{b}_{2i}|\boldsymbol{D}_2 &\overset{\text{ind.}}{\sim} N_{q_2}(\mathbf{0},\boldsymbol{D}_2),
\end{aligned} \tag{4.1}$$

where some components of random effects are shared between two models. The stochastic representation of the skew-normal distribution can be used for simplifying Markov Chain Monte Carlo (MCMC) approach in Bayesian specification.

Therefore, the first line of equation (4.1) can be written as:

$$\begin{aligned} \boldsymbol{Y}_i|\boldsymbol{b}_{1i},\boldsymbol{\beta}_1,\sigma_e^2,\boldsymbol{U}_i = \boldsymbol{u}_i &\overset{\text{ind.}}{\sim} N_{n_i}\Big(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1+\boldsymbol{Z}_{1i}\boldsymbol{b}_{1i}+\delta_e\boldsymbol{u}_i-\sqrt{\frac{2}{\pi}}\,\delta_e\mathbf{1}_{n_i},\,\sigma_e^2\boldsymbol{I}_{n_i}\Big)\,,\\ \boldsymbol{U}_i &\overset{\text{ind.}}{\sim} N_{n_i}(\mathbf{0},\boldsymbol{I}_{n_i})\,I(\boldsymbol{u}_i>0)\,, \end{aligned}$$

where $\boldsymbol{u}_i$ is the observed value of $\boldsymbol{U}_i$. Some components of longitudinal measurements may be missing due to dropout. We consider a non-ignorable missingness mechanism for them. In order to complete the Bayesian specification, prior distributions for all unknown parameters should be defined. The vector of unknown parameter is $\boldsymbol{\theta}=(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2',\sigma_e^2,\delta_e,\boldsymbol{D}_1,\boldsymbol{D}_2)$. To ensure to have proper posteriors in the model we consider proper but diffuse conditionally conjugate priors (Hobert and Casella, 1996). We assume that components of $\boldsymbol{\theta}$ are mutually independent and the prior distributions are given by

$$\begin{aligned} &\boldsymbol{\beta}_1\sim N_{p_1}(\boldsymbol{\beta}_{01},\boldsymbol{\Sigma}_{01})\,, && \boldsymbol{\beta}_2\sim N_{p_2}(\boldsymbol{\beta}_{02},\boldsymbol{\Sigma}_{02})\,,\\ &\boldsymbol{D}_1\sim IW_{q_1}(\boldsymbol{\eta}_{01},\boldsymbol{\psi}_{01})\,, && \boldsymbol{D}_2\sim IW_{q_2}(\boldsymbol{\eta}_{02},\boldsymbol{\psi}_{02})\,,\\ &\sigma_e^2\sim I\Gamma(\alpha_0,\tau_0)\,, && \delta_e\sim N(\mu_{\delta_e},\sigma_{\delta_e}^2)\,. \end{aligned} \tag{4.2}$$

The hyperparameters of these priors are selected such that they lead to the low-informative prior distributions. As all of these priors are proper but, low-informative in view of their variances.

For Bayesian implementation, one may use Gibbs sampling and Metropolis–Hastings algorithm via WinBUGS package.

4.1. Models comparison

For models comparison, we have used some famous criteria which are Deviance Information Criterion (DIC), Expected Akaike Information Criterion (EAIC), Expected Bayesian Information Criterion (EBIC, Carlin and Louis, 2000; Brooks, 2002) and Log Pseudo Marginal Likelihood (LPML).

Let $\boldsymbol{\Theta}$ and $\boldsymbol{Z}=(z_1,...,z_N)'$ be the entire model parameters and data, respectively. Define: $D(\boldsymbol{\Theta})=-2\ln f(\boldsymbol{z}|\boldsymbol{\Theta})=-2\sum_{i=1}^{N}\ln f(z_i|\boldsymbol{\Theta})$, where $f(z_i|\boldsymbol{\Theta})$ is marginal distribution of z_i, then $E\left[D(\boldsymbol{\Theta})\right]$ is a measure of fit and can be approximated by using the MCMC output in a Monte Carlo integration. This index is given by $\bar{D}=\frac{1}{K}\sum_{k=1}^{K}D(\boldsymbol{\Theta}^{(k)})$. Where $\boldsymbol{\Theta}^{(k)}$ is the k^{th} iteration of MCMC chain of the model and K is the number of iterations.

Therefore the Bayesian criteria are given by $\widehat{DIC}=\bar{D}+\hat{p}_D$, $\widehat{EAIC}=\bar{D}+2p$ and $\widehat{EBIC}=\bar{D}+p\ln(N)$, where p is the number of parameters and N is the

total number of observations. The smaller DIC, EAIC and EBIC, the better fit of the model.

Another popular criterion, which is usually used for model comparison in Bayesian context is Conditional Predictive Ordinate (CPO) statistic. Let $\boldsymbol{Z}^{(-i)}$, $i = 1, 2, ..., N$, denote the data set without its i^{th} individual, and let $\pi(\boldsymbol{\theta}|\boldsymbol{Z}^{(-i)})$ denote posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Z}^{(-i)}$, then $CPO_i = \int_{\boldsymbol{\Theta}} f(z_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{z}^{(-i)})\, d\boldsymbol{\theta}$. Gelfand and Dey (1994) show that CPO_i can be estimated by

$$CPO_i = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{f\left(z_i|\boldsymbol{\theta}^{(k)}, \boldsymbol{z}\right)} \right)^{-1} .$$

For collecting information of CPO_is, the LPML statistic is defined by $LPML = \sum_{i=1}^{N} \log(CPO_i)$. In this concept, unlike that of DIC, EAIC or EBIC, the larger value of LPML criterion indicates a better fitted model.

4.2. Convergence diagnostics

Gelman and Rubin (1992) have suggested a diagnostic test for assessing convergence. Their method recommends that two or more parallels (denoted by m) chains be generated, each with different starting values. For assessing convergence of individual model parameters the potential scale reduction factor (PSRF) may be used. The PSRF is calculated by $PSRF = \sqrt{\frac{n-1}{n} + \frac{m+1}{nm}\frac{B}{W}}$, where B/n is the between-chain variance $\left[\frac{\mathrm{B}}{n} = \frac{1}{m-1}\sum_{j=1}^{m}\left(\theta_j - \bar{\theta}\right)^2\right]$ and W is the within-chain variance $\left[\mathrm{W} = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(\theta_{ij} - \bar{\theta}_j\right)^2\right]$. As chains converge to a common target distribution, the between-chain variability should become small relative to the within-chain variability and consequently $PSRF$ should be close to 1. Conversely, PSRF value larger than 1 indicates non-convergence.

5. SIMULATION STUDY

To investigate the performance of the proposed model, we conducted a simulation study. In this simulation study, we generate 500 samples with sample size $n = 100$, moderate sample size, and $n = 500$, large sample size. We have considered the following joint modeling:

$$y_{ij} = \beta_{11} + \beta_{12}s_{ij} + \beta_{13}x_i + b_{1i} + b_{2i}s_{ij} + \varepsilon_{ij} , \qquad i = 1, 2, ..., n , \quad j = 1, 2..., 5 .$$

In this model $s_{ij} = j$, $x_i \sim Ber(0.2)$, $\beta_{11} = 10$, $\beta_{12} = -3$, $\beta_{13} = -2$, $\varepsilon_{ij} \sim SN\Big(-\sqrt{\frac{2}{\pi}}\,\delta_e, \sigma_e^2, \delta_e\Big)$, where $\sigma_e = 1$ and $\delta_e = 3$. Also, we have used a Cox proportional hazard model in a frailty structure with a Weibull baseline hazard as follows:

$$h(t) \;=\; h_0(t)\exp\big\{\beta_{21} + \beta_{22}x_i + \rho_1 b_{1i} + \rho_2 b_{2i}\big\}\,.$$

In this model, $\beta_{22} = -2$, $\rho_1 = 1$ and $\rho_2 = 2$. We have considered three rates of random dropout, 10%, 30% and 50%, which are generated by using different values for β_{21}. Therefore, we have a non-ignorable mechanism in the model, such that when $s_{ij} > T_i$ then i^{th} individual dropouts from the study. In this simulation study $\beta_{21} = 3$ leads to 10% rate of non-random dropout, $\beta_{21} = -1$ leads to 30% and $\beta_{21} = -2$ leads to 50% rate of non-random dropout of longitudinal outcomes. Also, $\boldsymbol{b}_i = (b_{1i}, b_{2i}) \sim N_2(0, \boldsymbol{D})$, where $\boldsymbol{D}$ is considered to be a 2×2 matrix, where $d_{11} = d_{22} = 1$ and $d_{21} = d_{12} = 0.5$. d_{11}, d_{12} and d_{22} are distinct elements of the matrix $\boldsymbol{D}$. Let $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$, we have used the following low-informative priors for unknown parameters.

$$\begin{aligned}
&\boldsymbol{\beta}_1 \sim N_3(0, 10^3 I_3)\,, && \boldsymbol{\beta}_2 \sim N_2(0, 10^3 I_2)\,,\\
&\sigma_e \sim I\Gamma(0.01, 0.01)\,, && \delta_e \sim N(0, 100)\,,\\
&\rho_k \sim N_2(0, 10^2 I),\ \ k = 1, 2\,, && \boldsymbol{D} \sim IW_2(100 I_2, 2)\,.
\end{aligned} \tag{5.1}$$

We have used "R2WinBUGS" package for implementation of this simulation study. We have implemented 10,000 iterations and have used the last 5000 iterations to obtain some summary of posterior. We have analysed these simulated data set under two distributional assumptions for within subject error: normal distributional assumption and skew-normal distributional assumption. The results of this simulation study are presented in Tables 1 to 3 for the rate of missingness 10%, 30% and 50%, respectively. The relative bias and mean square error of parameter $\boldsymbol{\theta}$ are defined as

$$Rel.Bias(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\theta}_i}{\theta} - 1\right), \qquad MSE(\theta) = \frac{1}{N}\sum_{i=1}^{N}\big(\hat{\theta}_i - \theta\big)^2,$$

where $\hat{\theta}_i$ is the estimate of θ for the i^{th} sample and $N = 500$. These tables show that when real data have skew-normal distribution, some parameters of joint modeling under normal assumption are estimated with some biases. These parameters are variance of error term of longitudinal model, variance of random effects and coefficients of random effects of survival model. For comparison of the performance of two models results of relative bias and mean square error of estimators are considered. Based on results given in Tables 1–3, we can conclude that skew-normal model leads to better inference in general. Also, these tables show that increasing of sample size in skew-normal scenario is an effective measure in decreasing standard errors, relative bias and MSE of the parameters.

6. HIV DATA SET

As an illustrative example of our Bayesian joint modeling, we use a longitudinal study on 467 HIV patients. Data were collected by Goldman *et al.* (1996). HIV infection results in a progressive destruction of immune function, which may be indicated by a decrease of CD4 (Stevens *et al.*, 2006). A count of CD4 cells of a person gives a general measure of the health of him/her immune system, and is a good measurement of immunosuppression. A normal CD4 cell count is more than 500 cells per cubic millimeter (mm3) of blood. If one has a CD4 count of fewer than 300, one will be diagnosed as having AIDS, therefore, CD4 count measurement is an important index which provides a way of gauging the progression from HIV to acquired immune deficiency syndrome (AIDS) for prognostic purposes. Thus, in this study, the CD4 count measurements over time are chosen as response variable.

This study is done for comparing the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). The patients met another entry conditions which is AIDS diagnosis or two CD4 counts of 300 or fewer, also they randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry and again at the 2, 6, 12, and 18 months. Another variable which is recorded in this study is time to death.

Before this Guo and Carlin (2004), Rizopoulos (2010) and some other authors had suggested that, for analysing this data set, a square root transformation for CD4 counts be used instead of Gaussian model. Figure 1 shows histogram and q-q plot of $\sqrt{CD4}$ which shows that there are right skewness $\sqrt{CD4}$ even after root transformation. We have used the same transformation in our analysis, but under skew-normal distribution assumption for the error term. Figure 2 presents the subject-specific profile for fifty randomly selected individuals given each drug. Panels of this figure show a sharply increasing degree of missing data over time due to death, dropout, and missed clinic visits. In this figure the profiles of those individuals who remain and those of individuals who do not remain are indicated using gray and black colors, respectively. This figure underlines that those who do not remain had smaller $\sqrt{CD4}$ than others. Figure 3 presents Kaplan–Meier survival curve estimates for both treatment groups. The plot suggests longer survival times in the ddC group compared to the ddI group, from month 6 onwards.

We have used a skew-normal JM of longitudinal and time to event data for analysing the data set. The linear mixed effect model with random intercept and slope is:

$$\begin{aligned} y_{ij} \;=\; & \beta_{11} + \beta_{12}\, t_{ij} + \beta_{13}\, t_{ij}\, Drug_i + \beta_{14}\, Gender_i \\ & +\; \beta_{15}\, PrevOI_i + \beta_{16}\, Stratum_i + b_{1i} + b_{2i}\, t_{ij} + \sigma_e\, \varepsilon_{ij} \;. \end{aligned} \tag{6.1}$$

For the time to event process, we have used a Cox proportional hazard model, the hazard function for this model is given by

$$h(t_i) = h_0(t_i)\exp\Big\{\beta_{21} + \beta_{22}\,Drug_i + \beta_{23}\,Gender_i + \beta_{24}\,PrevOI_i + \beta_{25}\,Stratum_i + \rho_1 b_{1i} + \rho_2 b_{2i}\Big\}. \tag{6.2}$$

In models (6.1) and (6.2) the vector of random effects $\boldsymbol{b}_i = (b_{1i}, b_{2i})'$ is shared between two models. Also, we consider normal and skew-normal distribution assumptions on the longitudinal mixed model. Random effects are assumed to have a bivariate normal distribution, that is, $\boldsymbol{b}_i \sim N_2(\boldsymbol{0},\boldsymbol{D})$ and $\varepsilon_{ij} \sim SN\Big(-\sqrt{\frac{2}{\pi}}\delta_e, \sigma_e^2, \delta_e\Big)$. In this model, y_{ij} is the squared root of the j^{th} CD4 count measurement on the i^{th} individual in the trial, $j = 1, 2, ..., 5$ and $i = 1, 2, ..., 467$. $Gender_i$ is a gender indicator ($0 =$ female, $1 =$ male), also other three explanatory variables are $Drug_i$ ($0 =$ ddC, $1 =$ ddI), $PrevOI_i$, previous opportunistic infection, ($1 =$ AIDS diagnosis, $0 =$ no AIDS diagnosis), and $Stratum_i$ ($1 =$ AZT failure, $0 =$ AZT intolerance).

In Bayesian MCMC implementation, we ran two parallels MCMC chains with different starting values for $100,000$ iterations each. Then, we discarded the first $20,000$ iterations as pre-convergence burn-in and retained $80,000$ for the posterior analysis. Let $\boldsymbol{\beta}_k = (\beta_{k1}, ..., \beta_{kp_k})'$ where $k = 1, 2,\ p_1 = 6,\ p_2 = 5$. In all models, we consider $\boldsymbol{\beta}_k \sim N_{p_k}(0, 10000 I_{p_k})$, $\sigma_e^2 \sim I\Gamma(0.1, 0.1)$, $\rho_k \sim N(0, 100)$, $\boldsymbol{D} \sim IW_2(100 I_2, 2)$ and $\delta_e \sim N(0, 100)$. For the piecewise baseline hazard function [$h_l,\ l = 1, 2, 3, 4$ (the number of piecewise baseline $= 4$)] the $gamma(1, 1)$ prior distribution is considered for each piece (h_i, $i = 1, 2, 3, 4$). Hyperparameters are chosen such that the priors of the parameters tend to be weakly informative. We have considered joint modeling of equations (7)–(8) under two different distribution assumptions.

After checking Gelman–Rubin diagnosis test for convergence, Bayesian parameter estimates including posterior mean, standard deviation and 95% highest posterior density of all parameters are given in Table 4. According to DIC, EAIC, EBIC and LPML criteria the skew-normal model has a better fit to these data. This table shows that time and previous opportunistic infection are two significant covariates in longitudinal model, also skewness parameter of error term is significant, where the more time and previous opportunistic infection, the less CD4 count measurements. In survival model ρ_1 and ρ_2 are significant which shows that two models are dependent. Also, Table 4 shows that skewness parameter is significant and ignoring this parameter and using normal model leads to overestimating of variance of the error in longitudinal model.

An important criterion for finding the influential observations is Kullback–Leibler divergence criterion between $\pi(\boldsymbol{\theta}|\boldsymbol{Z})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{Z}^{(-i)})$, $i = 1, 2, ..., n$, where $\boldsymbol{Z}$ and $\boldsymbol{Z}^{(-i)}$ are all data and the data set without its i^{th} individual, respectively.

It is defined by

$$K_i = \int \pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{t}) \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{t})}{\pi\big(\boldsymbol{\theta}|\boldsymbol{y}^{(-i)},\boldsymbol{t}^{(-i)}\big)}\right) d\boldsymbol{\theta} \ .$$

This can be approximated by: (Christensen *et al.*, 2011; page 341)

$$K_i = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{L\big(\boldsymbol{\theta}_j|\boldsymbol{y}_i,t_i\big)}\right) - \frac{1}{m}\sum_{j=1}^{m}\log\left(\frac{1}{L\big(\boldsymbol{\theta}_j|\boldsymbol{y}_i,t_i\big)}\right).$$

An observation with large K_i is considered as an influential observation. Figure 4 shows Kullback–Leibler divergence for both skew-normal and normal models. This figure shows that the skew-normal model detects some individuals as influential observations, but normal model does not detect any. Individuals 131 and 353 have large values of the response in the second observed time in comparison with the largest value in this time. These individuals have observed survival times 12.23 and 12.53, respectively. There is so much increase in CD4 count measurements at 3^{th} to 4^{th} observed time for individuals 188 and 245. Other individuals detected in Figure 3 are 417, 171 and 319. Except observation 417 who is died at time 10.60, the other individuals dropped out from the study at times 17.33 and 15.97, respectively, where the length of the period of the study is equals to 18. The longitudinal measurements for individuals 417 and 171 are close to the lowest value of the CD4 measurements at each time of the study.

Also, a sensitivity analysis is performed to see the modification of posterior distribution with respect to changes in the hyperparameters of prior distributions of σ_e^2. For this purpose, we assume $\sigma_e^2 = \frac{1}{\tau_e^2}$, where $\tau_e^2 \sim \Gamma(\epsilon,\epsilon)$ (see Gelman, 2006) and $\mathrm{Var}(\tau_e^2) = \frac{1}{\epsilon} = 10^k$, $k = -3,-2,...,2,3$. Sensitivity of the posterior mean of all parameters for different values of k is investigated. This shows that our inferences, containing the results of posterior means, standard deviations and DIC value, are not sensitive to the change of value of k after $k = 1$ (the results are not given here to save space).

7. CONCLUSION

In this paper, we have used a multivariate skew-normal distribution family, which includes normal distribution, for analysing skew longitudinal responses in joint modeling of longitudinal and survival times. We have used Bayesian approach and WinBUGS software for implementing the proposed model. We have performed some simulation studies to investigate the performance of the proposed method. Also, the proposed method is used for analysing a real HIV data set. Our analysis shows that these data set are severely skew, and the skew-normal model has a better performance than normal model based on the

DIC, EAIC, EBIC and LPML criteria. The program codes for analysing the data set are available under request from the authors. If a data set includes outliers and skew longitudinal responses a joint model with assumption of skew-normal/independent distribution for responses may be defined to analyse the data.

REFERENCES

[1] Arellano-Valle, R.B.; Bolfaine, H. and Lachos, V.H. (2005a). Skew-normal linear mixed models, *Journal of Data Science*, **3**, 415–438.

[2] Arellano-Valle, R.B. and Genton, M.G. (2005). On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**, 93–116.

[3] Arellano-Valle, R.B.; Ozan, S.; Bolfarine, H. and Lachos, V.H. (2005b). Skew-normal measurement error models, *Journal of Multivariate Analysis*, **96**, 265–281.

[4] Arellano-Valle, R.B.; Bolfarine, H. and Lachos, V.H. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**, 663–682.

[5] Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[6] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica XLVI*, 199–208.

[7] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distributions, *Journal of the Royal Statistical Society: Series B*, 579–602.

[8] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715-726.

[9] Baghfalaki, T. and Ganjali, M. (2011). An EM estimation approach for analyzing bivariate skew-normal data with non-monotone missing values, *Communications in Statistics – Theory and Methods*, **40**(9), 1671–1686.

[10] Baghfalaki, T. and Ganjali, M. (2012). An ECM Estimation Approach for Analyzing Multivariate Skew-Normal Data with Dropout, *Communications in Statistics – Simulation and Computation*, **41**, 1970–1988.

[11] Bandyopadhyay, D.; Lachos, V.H. and Abanto-Valle, C.A. (2010). Bayesian inference for bivariate skew-normal/independent linear mixed models with application to periodontal disease, *Statistics in Medicine*, **29**, 2643–2655.

[12] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, **79**, 99–113.

[13] Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter Best Carlin and vandeLinde, 2002, *Journal of the Royal Statistical Society Series B*, **64**(3), 616–618.

[14] Cancho, V.G.; Lachos, V.H. and Ortega, E.M. (2010). A nonlinear regression model with skew-normal errors, *Statistical Papers*, **51**, 547–558.

[15] Carlin, P. and Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, Second Edition, Chapman & Hall, CRC Press, New York.

[16] Christensen, R.; Johnson, W.; Branscum, A. and Hanson, T.E. (2011). *Bayesian Ideas and Data Analysis, An Introduction for Scientists and Statisticians*, Chapman & Hall / CRC.

[17] DeGruttola, V. and Tu, X.M. (1994). Modeling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time, *Biometrics*, **50**, 1003–1014.

[18] Diggle, P.J.; Sousa, I. and Chetwynd, A.G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture, *Statistics in Medicine*, **27**, 2981–2998.

[19] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper), *Bayesian Analysis*, **1**, 515–534.

[20] Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, **7**, 457–511.

[21] Goldman, A.I.; Carlin, B.P.; Crane, L.R.; Launer, C.; Korvick, J.A.; Deyton, L. and Abrams, D.I. (1996). Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **11**, 161–169.

[22] Ghosh, P. and Branco, D.B. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA, *Statistics in Medicine*, **26**(6), 1255–1267.

[23] Guo, X. and Carlin B.P. (2004). Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages, *The American Statistician*, **58**, 16–24.

[24] Hashemi, R.; Jacqmin-Gadda, H. and Commenges, D. (2003). A latent process model for joint modeling of events and markers, *Lifetime Data Analysis*, **8**(4), 335–348.

[25] Henderson, R.; Diggle, P.J. and Dobson, A. (2000). Joint Modelling of Longitudinal Measurements and Event Time Data, *Biostatistics*, **1**, 465–480.

[26] Henze, N.A. (1986). A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, **13**, 271–275.

[27] Hobert, J. and Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models, *Journal of the American Statistical Association*, **91**, 1461–1473.

[28] Huang, Y.; Chen, R. and Dagne, G. (2011a). Simultaneous Bayesian Inference for Linear, Nonlinear and Semiparametric Mixed-Effects Models with Skew-Normality and Measurement Errors in Covariates, *The International Journal of Biostatistics*, **7**(1), Article 8.

[29] Huang, Y. and Dagne, G. (2010). Skew-normal Bayesian nonlinear mixed-effects models with application to AIDS studies, *Statistics in Medicine*, **29**(23), 2384–2398.

[30] Huang, Y. and Dagne, G. (2011). A Bayesian Approach to Joint Mixed-Effects Models with a Skew-Normal Distribution and Measurement Errors in Covariates, *Biometrics*, **67**, 260–269.

[31] Huang, Y.; Dagne, G. and Wu, L. (2011b). Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors, *Statistics in Medicine*, **30**(24), 2930–2946.

[32] Lachos, V.H.; Bolfarine, H.; Arellano-Valle, R.B. and Montenegro, L.C. (2007). Likelihood based inference for multivariate skew-normal regression models, *Communications in Statistics – Theory and Methods*, **36**, 1769–1786.

[33] Lachos, V.H.; Bandyopadhyay, D. and Dey, D.K. (2011). Linear and Nonlinear Mixed-Effects Models for Censored HIV Viral Loads Using Skew-Normal/Independent Distributions, *Biometrics*, **67**, 1594–1604.

[34] Lachos, V.H.; Ghosh, P. and Arellano-Valle, R.B. (2010). Likelihood Based Inference for Skew-Normal/Independent Linear Mixed Model, *Statistica Sinica*, **20**, 303–322.

[35] Lin, I.; Ho, J. and Chen, L. (2009). Analysis of multivariate skew-normal models with incomplete data, *Journal of Multivariate Analysis*, **100**, 2337–2351.

[36] Lin, T. and Lee, J. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data, *Statistics in Medicine*, **27**, 1490–1507.

[37] Rizopoulos, D. (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data, *Journal of Statistical Software*, **35**(9).

[38] Sahu, S.K.; Dey, D.K. and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**(2), 129–150.

[39] Sousa, I. (2011). A review on joint modelling of longitudinal measurements and time-to-event, *REVSTAT – Statistical Journal*, **9**(1), 57–81.

[40] Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Lindevan der, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B*, **64**, 583–616.

[41] Stevens, R.A.; Lempicki, R.A.; Natarajan, V.; Higgins, J.; Adelsberger, J.W. and Metcalf, J.A. (2006). *General immunologic evaluation of patients with immunodeffciency virus infection.* In "Manual of Molecular and Clinical Laboratory Immunology" (B. Detrick, R.G. Hamilton, J.D. Folds, *et al.*, Eds.), 7th edition, 848–861, Washington, DC, ASM Press.

[42] Tseng, Y.; Hsieh, F. and Wang, J. (2005). Joint modelling of accelerated failure time and longitudinal data, *Biometrika*, **92**, 587–603.

[43] Tsiatis, A.A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error, *Biometrika*, **88**, 447–458.

[44] Tsiatis, A.A.; DeGruttola, V. and Wulfsohn, M.S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS, *Journal of the American Statistical Association*, **90**(429), 27–37.

[45] Wu, L.; Hu, J. and Wu, H. (2008). Joint inference for nonlinear mixed-effects models and time-to-event at the presence of missing data, *Biostatistics*, **9**, 308–320.

[46] Wulfsohn, M.S. and Tsiatis, A.A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error, *Biometrics*, **53**, 330–339.

[47] Yu, M.; Law, N.J.; Taylor, J.M.G. and Sandler, H.M. (2004). Joint longitudinal survival cure models and their application to prostate cancer, *Statistica Sinica*, **14**, 835–862.

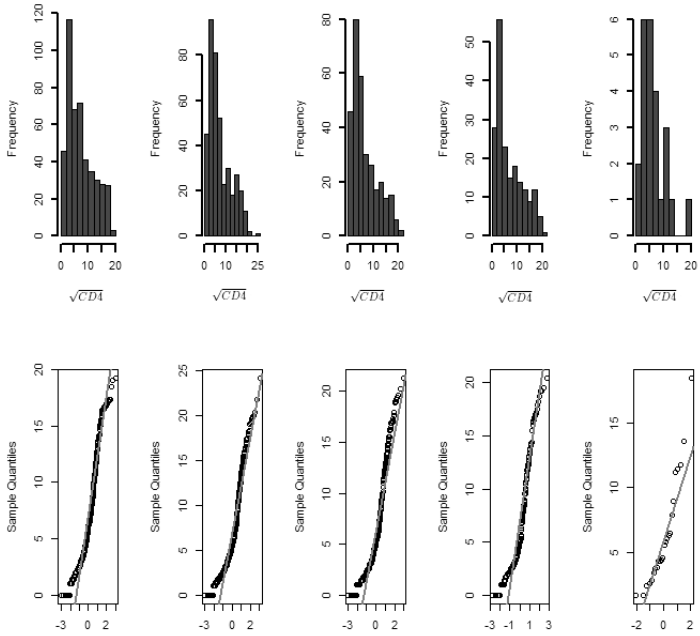

Figure 1: Histogram and q-q plot of $\sqrt{CD4}$ in HIV data set.

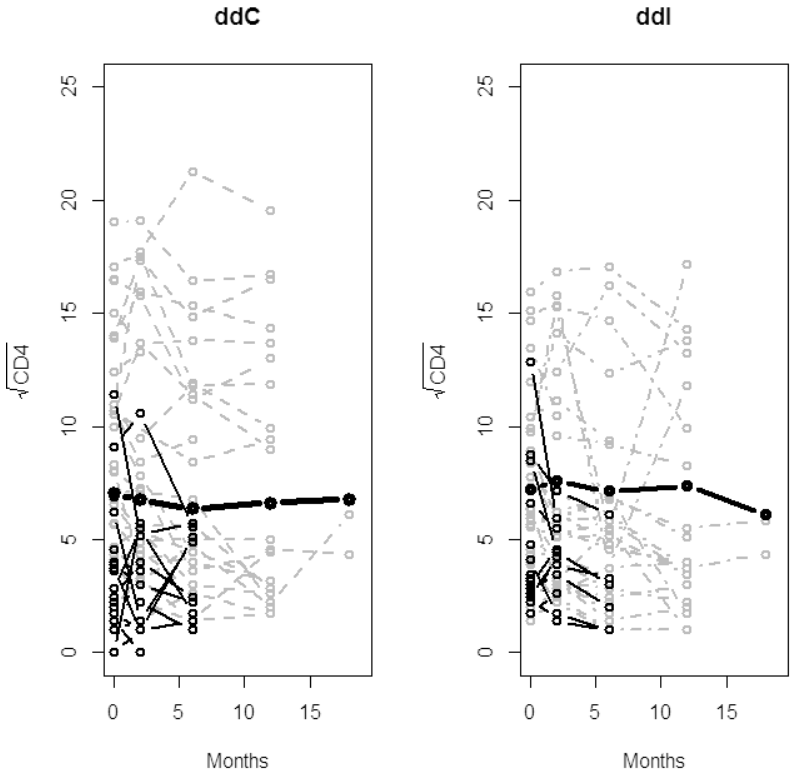

Figure 2: Profiles of $\sqrt{CD4}$ measurements over time for all individuals from each drug, bold black lines are mean profile for all observed individuals on each drug, gray color indicates those individuals who remain and black color represents those who do not remain.

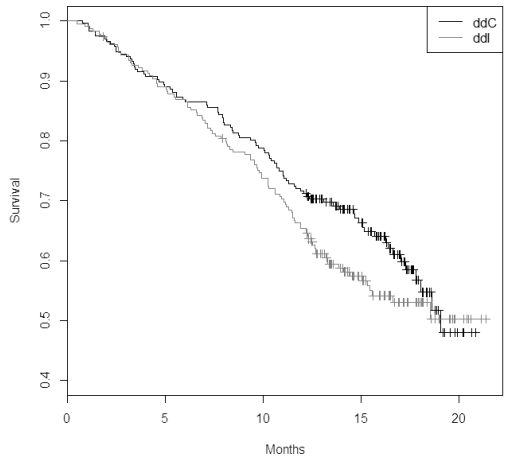

Figure 3: Kaplan–Meier estimates of the probability of survival for individuals on each drug.

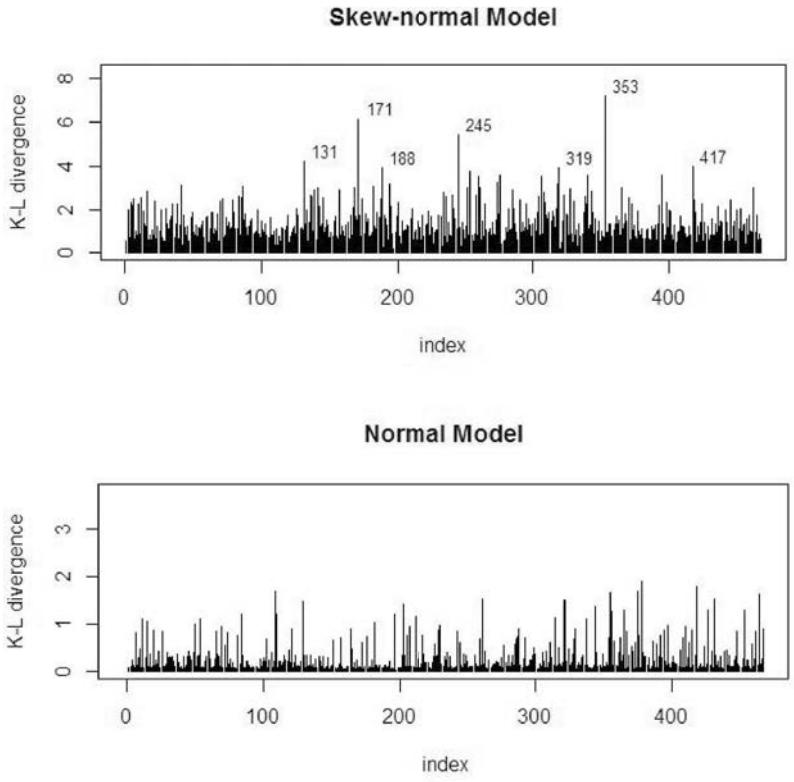

Figure 4: Kullback–Leibler divergence for the skew-normal and normal models.

Table 1: Result of simulation study for 500 samples under two distribution assumptions with 10% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

Model		Normal model						Skew-normal model					
Sample size		$n = 100$			$n = 500$			$n = 100$			$n = 500$		
parameters	real	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE
β_{11}	10.000	9.966(0.292)	-0.003	0.086	9.965(0.091)	-0.003	0.009	10.027(0.206)	0.002	0.042	10.003(0.122)	0.0003	0.015
β_{12}	-3.000	-3.013(0.189)	0.004	0.035	-2.995(0.047)	-0.001	0.002	-3.010(0.129)	0.003	0.016	-2.989(0.072)	-0.003	0.005
β_{13}	-2.000	-1.939(0.541)	-0.030	0.294	-1.960(0.130)	-0.019	0.017	-1.995(0.366)	-0.002	0.133	-1.988(0.187)	-0.005	0.035
β_{21}	3.000	2.928(0.526)	-0.023	0.281	2.977(0.151)	-0.007	0.022	2.964(0.356)	-0.012	0.127	3.012(0.196)	0.004	0.038
β_{22}	-2.000	-1.904(0.455)	-0.047	0.215	-1.938(0.127)	-0.031	0.019	-2.008(0.357)	0.004	0.126	-1.986(0.192)	-0.006	0.037
d_{11}	1.000	5.055(0.462)	4.055	16.659	5.961(0.143)	3.961	20.943	1.487(0.287)	0.487	0.319	1.236(0.172)	0.236	0.085
d_{12}	0.500	0.241(0.228)	-0.517	0.119	0.392(0.074)	-0.215	0.016	0.421(0.146)	-0.158	0.027	0.463(0.092)	-0.072	0.009
d_{22}	1.000	3.212(0.240)	2.212	4.952	3.254(0.071)	3.254	6.069	1.141(0.152)	0.141	0.043	1.054(0.090)	0.054	0.011
σ_e^2	1.000	4.146(0.521)	3.146	10.168	4.067(0.154)	3.067	9.434	0.919(0.525)	-0.080	0.280	0.906(0.243)	-0.093	0.067
δ_e	3.000	-	-	-	-	-	-	2.972(0.386)	-0.009	0.149	3.020(0.156)	0.006	0.024
ρ_1	1.000	0.521(0.166)	-0.478	0.256	0.684(0.070)	-0.315	0.104	0.793(0.187)	-0.206	0.077	0.885(0.125)	-0.114	0.028
ρ_2	2.000	2.237(0.201)	0.118	0.096	2.193(0.063)	0.096	0.041	2.146(0.179)	0.073	0.053	2.076(0.102)	0.038	0.016

d_{11}, d_{12} and d_{22} are distinct elements of the matrix $\boldsymbol{D}$.

Table 2: Result of simulation study for 500 samples under two distribution assumptions with 30% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

Model		Normal model						Skew-normal model					
Sample size		$n = 100$			$n = 500$			$n = 100$			$n = 500$		
parameters	real	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE
β_{11}	10.000	10.181(0.276)	0.018	0.104	10.046(0.247)	0.004	0.062	10.009(0.229)	0.001	0.052	10.008(0.132)	0.0008	0.017
β_{12}	-3.000	-3.156(0.259)	0.052	0.087	-3.088(0.161)	0.029	0.033	-3.013(0.158)	0.004	0.025	-3.010(0.095)	0.003	0.009
β_{13}	-2.000	-1.905(0.849)	-0.047	0.686	-1.866(0.407)	-0.066	0.181	-1.998(0.375)	-0.0005	0.140	-1.979(0.205)	-0.010	0.042
β_{21}	-2.000	-1.947(0.487)	-0.026	0.225	-1.980(0.368)	-0.009	0.133	-1.968(0.380)	-0.015	0.145	-2.005(0.217)	0.002	0.046
β_{22}	-2.000	-1.881(0.982)	-0.059	0.919	-1.851(0.505)	-0.074	0.273	-2.031(0.436)	0.015	0.190	-1.987(0.257)	-0.006	0.066
d_{11}	1.000	5.540(0.539)	4.540	20.886	3.770(0.330)	2.770	7.783	1.546(0.293)	0.546	0.383	1.295(0.203)	0.295	0.128
d_{12}	0.500	-0.017(0.280)	-1.034	0.341	0.075(0.184)	-0.849	0.213	0.342(0.173)	-0.314	0.054	0.427(0.100)	-0.144	0.015
d_{22}	1.000	3.615(0.284)	2.615	6.914	2.464(0.161)	1.464	2.170	1.177(0.184)	0.177	0.065	1.072(0.111)	0.072	0.017
σ_e^2	1.000	4.050(0.886)	3.050	10.043	3.983(0.452)	2.983	9.101	0.925(0.584)	-0.074	0.345	0.853(0.298)	-0.146	0.110
δ_e	3.000	-	-	-	-	-	-	2.904(0.489)	-0.031	0.248	3.022(0.180)	0.007	0.032
ρ_1	1.000	0.515(0.223)	-0.484	0.281	0.627(0.231)	-0.372	0.191	0.807(0.255)	-0.192	0.102	0.846(0.176)	-0.153	0.054
ρ_2	2.000	1.908(0.126)	-0.045	0.023	2.004(0.177)	0.001	0.030	2.118(0.203)	0.059	0.055	2.095(0.118)	0.047	0.022

d_{11}, d_{12} and d_{22} are distinct elements of the matrix $\boldsymbol{D}$.

Table 3: Result of simulation study for 500 samples under two distribution assumptions with 50% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

Model		Normal model						Skew-normal model					
Sample size		$n = 100$			$n = 500$			$n = 100$			$n = 500$		
parameters	real	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE	Est.(S.E.)	Rel. Bias	MSE
β_{11}	10.000	10.004(0.320)	0.0004	0.101	10.114(0.193)	0.011	0.049	10.028(0.235)	0.002	0.055	10.012(0.128)	0.001	0.016
β_{12}	-3.000	-3.149(0.299)	0.049	0.110	-3.091(0.140)	0.030	0.027	-3.031(0.156)	0.010	0.025	-3.012(0.089)	0.004	0.007
β_{13}	-2.000	-2.003(0.607)	0.001	0.362	-2.048(0.398)	0.024	0.155	-1.992(0.435)	-0.003	0.187	-1.964(0.267)	-0.017	0.071
β_{21}	-3.000	-3.137(0.658)	0.045	0.444	-2.924(0.301)	-0.025	0.093	-3.002(0.345)	0.0007	0.118	-3.001(0.202)	0.0002	0.040
β_{22}	-2.000	-2.012(0.696)	0.006	0.476	-2.045(0.486)	0.022	0.231	-1.969(0.519)	-0.015	0.269	-2.005(0.337)	0.002	0.111
d_{11}	1.000	5.761(0.553)	4.760	22.968	3.881(0.356)	2.881	8.427	1.607(0.346)	0.607	0.488	1.298(0.229)	0.298	0.141
d_{12}	0.500	-0.254(0.323)	-1.509	0.672	-0.059(0.247)	-1.118	0.372	0.341(0.173)	-0.318	0.055	0.399(0.108)	-0.200	0.021
d_{22}	1.000	3.924(0.363)	2.924	8.684	2.529(0.181)	1.529	2.372	1.216(0.181)	0.216	0.079	1.093(0.106)	0.093	0.019
σ_e^2	1.000	4.042(0.710)	3.042	9.753	3.894(0.421)	2.894	8.552	0.928(0.638)	-0.072	0.409	0.834(0.259)	-0.166	0.093
δ_e	3.000	-	-	-	-	-	-	2.906(0.476)	-0.031	0.233	3.017(0.154)	0.005	0.023
ρ_1	1.000	0.737(0.287)	-0.263	0.150	0.672(0.192)	-0.327	0.142	0.826(0.326)	-0.173	0.135	0.882(0.151)	-0.117	0.036
ρ_2	2.000	1.746(0.345)	-0.126	0.181	1.928(0.172)	-0.036	0.033	2.117(0.237)	0.058	0.069	2.084(0.137)	0.042	0.025

d_{11}, d_{12} and d_{22} are distinct elements of the matrix $\boldsymbol{D}$.

Table 4: Bayesian parameter estimates, posterior means (standard deviations, s.d.), and 95% HPDs for analysing the HIV data set using skew-normal distribution (Skew-normal model) and the normal distribution for error (Normal model).

	Skew-normal model		Normal model	
parameters	mean(s.d.)	95% HPD	mean(s.d.)	95% HPD
Intercept (β_{11})	10.351(0.591)	(9.186,11.589)	10.589(0.713)	(9.212,11.921)
Time (β_{12})	-0.362(0.048)	(-0.461,-0.274)	-0.352(0.052)	(-0.451,-0.241)
Time × Drug (β_{13})	0.019(0.071)	(-0.119,0.148)	0.023(0.072)	(-0.119,0.172)
Gender (β_{14})	-0.021(0.583)	(-1.209,1.198)	-0.255(0.673)	(-1.506,1.093)
PrevOI (β_{15})	-4.516(0.429)	(-5.459,-3.749)	-4.642(0.517)	(-5.704,-3.661)
Stratum (β_{16})	-0.194(0.422)	(-0.974,0.651)	-0.127(0.451)	(-0.938,0.818)
Intercept (β_{21})	-4.769(0.699)	(-6.168,-3.402)	-4.767(0.713)	(-5.975,-3.171)
Drug (β_{22})	0.423(0.296)	(-0.141,1.003)	0.373(0.292)	(-0.213,0.939)
Gender (β_{23})	-0.498(0.407)	(-1.292,0.326)	-0.295(0.449)	(-1.146,0.568)
PrevOI (β_{24})	2.271(0.358)	(1.603,2.992)	2.230(0.389)	(1.502,2.964)
Stratum (β_{25})	0.098(0.281)	(-0.441,0.674)	0.076(0.288)	(-0.467,0.626)
ρ_1	-0.314(0.044)	(-0.405,-0.229)	-0.299(0.042)	(-0.381,-0.216)
ρ_2	-3.722(0.448)	(-4.711,-2.903)	-3.865(0.422)	(-4.633,-3.026)
d_{11}	15.281(1.195)	(13.078,17.710)	16.131(1.192)	(13.940,18.591)
d_{12}	-0.033(0.155)	(-0.340,0.271)	-0.040(0.156)	(-0.346,0.262)
d_{22}	0.468(0.037)	(0.399,0.547)	0.472(0.039)	(0.401,0.554)
δ_e	2.674(0.422)	(1.457,3.062)	–	–
σ_e^2	0.706(0.570)	(0.071,2.244)	3.052(0.175)	(2.742,3.410)
h_1	0.146(0.088)	(0.038,0.376)	0.119(0.077)	(0.028,0.325)
h_2	0.519(0.271)	(0.158,1.183)	0.439(0.244)	(0.127,1.082)
h_3	1.397(0.706)	(0.431,3.142)	1.207(0.643)	(0.369,2.881)
h_4	1.819(0.931)	(0.545,4.069)	1.623(0.862)	(0.488,3.827)
	Model Comparison Criteria			
DIC	6734.76		9425.89	
EAIC	5290.922		8810.331	
EBIC	5306.314		8165.494	
LPML	-3004.761		-3952.432	

d_{11}, d_{12} and d_{22} are distinct elements of the matrix $\boldsymbol{D}$.

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