ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 15, No.3
July 2017

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 21 842 61 00
    Fax: + 351 21 845 40 84
    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 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 15, Number 3, July 2017, 299–332

A STUDY ON THE BIAS-CORRECTION EFFECT OF THE AIC FOR SELECTING VARIABLES IN NORMAL MULTIVARIATE LINEAR REGRESSION MODELS UNDER MODEL MISSPECIFICATION

Authors: Hirokazu Yanagihara
– Department of Mathematics, Graduate School of Science, Hiroshima University, Hiroshima, Japan
yanagi@math.sci.hiroshima-u.ac.jp

Ken-ichi Kamo
– Department of Liberal Arts and Sciences, Sapporo Medical University, Hokkaido, Japan
kamo@sapmed.ac.jp

Shinpei Imori
– Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, Osaka, Japan
imori@sigmath.es.osaka-u.ac.jp

Mariko Yamamura
– Department of Mathematics Education, Graduate School of Education, Hiroshima University, Hiroshima, Japan
yamamura@hiroshima-u.ac.jp

Received: October 2013 Revised: March 2015 Accepted: July 2015

Abstract:

- By numerically comparing a variable-selection method using the crude AIC with those using the bias-corrected AICs, we find out knowledge about what kind of bias correction gives a positive effect to variable selection under model misspecification. Actually, since all the variable-selection methods considered in this paper asymptotically choose the same model as the best model, we conduct numerical examinations using small and moderate sample sizes. Our results show that bias correction under assumption that the mean structure is misspecified gives a better effect to a variable-selection method than that under the assumption that the distribution of the model is misspecified.

Key-Words:

- *AIC; bias-corrected AIC; KL information; loss function; nonnormality; risk function; variable selection.*

AMS Subject Classification:

- 62J05, 62F07.

1. INTRODUCTION

In the analysis of real data, it is important to determine which statistical model best fits the data, because there are many candidate models, and they each estimate different results, which may lead to different conclusions. One of the aims of model selection is to choose a statistical model having a high predictive accuracy. In order to achieve the aim, it is common that the risk function based on the Kullback–Leibler (KL) information [18] is used for assessing a goodness of fit of a statistical model. Then, the model making the risk function the smallest is regarded as the "best" model. Hence, in order to seek a statistical model having a high predictive accuracy, we have to compare with risk functions of each of candidate models. In practice, an estimate of the risk function is used, because the risk function involves unknown parameters. The most famous asymptotic unbiased estimator of the risk function is Akaike's information criterion (AIC; proposed by [1, 2]), which is derived under the condition that the candidate model is correctly specified. It is defined by the simple equation $-2 \times$ (the maximum log-likelihood) $+ 2 \times$ (the number of parameters in the model) and is commonly used in actual data analysis.

Since the AIC is only asymptotically unbiased, the bias of the AIC to the risk function may be considerable when the sample size is not large enough and the number of parameters is large. Then, the AIC of a candidate model which is overspecified and has a large number of parameters tends to underestimate the risk function overly. This tendency causes that AICs of those candidate models often do not have notable differences. In addition, the variance of the AIC may increase as the number of parameters increases (see, e.g., [31]). Thus, the model with the most parameters tends to make AIC the smallest, and so the AIC often selects the model with the most parameters as the best model. Since this fault of AIC is due to the bias, it is frequently avoided by correcting the bias to the risk function. This has been studied under various different conditions and with various different correction methods (as a general theory correcting the bias of the AIC, see, e.g., [4, 14, 16, 20]). Sugiura [24] and Hurvich and Tsai [12] proposed a bias-corrected AIC for linear regression models (multiple regression models) by fully removing the bias of the AIC to the risk function under the condition that the candidate model is correctly specified. The bias-corrected AIC then becomes the uniformly minimum-variance unbiased estimator (UMVUE) for the risk function of the candidate model (see, [5]), and many authors have verified by numerical experiments that a variable-selection method using the bias-corrected AIC performs better than that using the crude AIC.

A basic concept of bias correction is that we expect that an unbiased estimate of the risk function will allow us to correctly evaluate the risk function, which will further facilitate the selection of the best model. However, there is no theory that promises that the best model chosen by minimizing a bias-corrected

AIC has a higher predictive accuracy than that chosen by minimizing the crude AIC. Generally, a bias-corrected estimator has a larger variance than a crude estimator before a bias correction. An impairment of the mean square error of the bias-corrected AIC with respect to the risk function, which results from an increase in the variance, may cause a drop in performances of model selection when using a bias-corrected AIC.

In this paper, we compare the AIC and eight bias-corrected AICs to study what kind of bias correction gives a positive effect for selecting variables for a multivariate linear regression model (MLRM) with a normal distributed assumption (called the normal MLRM), under a model misspecification. Performances of variable-selection methods using the nine criteria are examined by numerical experiments. We do not conduct numerical experiments under the large sample, because it has been confirmed theoretically that the variable-selection methods using the nine criteria select the same model as "best" when the sample size goes to ∞. Our result is that correcting the bias gives a greater positive effect to variable selection when the mean structure is misspecified than when the distribution of the model is misspecified.

This paper is organized as follows: In Section 2, the normal MLRM and the risk function based on the KL information are described. In Section 3, the AIC and the bias-corrected AICs for the normal MLRM are summarized. In Section 4, we use numerical experiments with small and moderate samples to compare performances of variable-selection methods using the AIC and the bias-corrected AICs. Our conclusions and a discussion are presented in Section 5. Technical details are provided in the Appendix.

2. RISK FUNCTION BASED ON THE KL INFORMATION

The normal MLRM is used when we are interested in predicting not just one response variable but several correlated response variables based on k nonstochastic explanatory variables (for details, see, e.g., [6], [21, chap. 9], [26, chap. 4]). Let $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ be p-dimensional independent random vectors of response variables, and let $\boldsymbol{x}_{\omega,1}, ..., \boldsymbol{x}_{\omega,n}$ be k_ω-dimensional vectors of the full explanatory variables, where n is the sample size. Furthermore, let $\boldsymbol{x}_i$ be a k-dimensional vector of candidate explanatory variables, which is a subset of the full explanatory variables $\boldsymbol{x}_{\omega,i}$ $(i = 1, ..., n)$. Then, we consider the following normal MLRM as the candidate model:

$$M: \ \boldsymbol{y}_i \sim N_p(\boldsymbol{\Xi}'\boldsymbol{x}_i, \boldsymbol{\Sigma}) \ , \quad (i = 1, ..., n) \ , \tag{2.1}$$

where $\boldsymbol{\Xi}$ is a $k \times p$ matrix of unknown regression coefficients, and $\boldsymbol{\Sigma}$ is a $p \times p$ unknown covariance matrix.

Let $\boldsymbol{Y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)'$ be an $n \times p$ matrix of response variables, and let $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$ be an $n \times k$ matrix of candidate explanatory variables. Suppose that an $n \times k_\omega$ matrix of the full explanatory variables, $\boldsymbol{X}_\omega = (\boldsymbol{x}_{\omega,1}, ..., \boldsymbol{x}_{\omega,n})'$, is a column full-rank matrix, i.e., $\mathrm{rank}(\boldsymbol{X}_\omega) = k_\omega < n$. Needless to say, $\boldsymbol{X}$ consists of some columns of $\boldsymbol{X}_\omega$ and is also a column full-rank matrix. Moreover, we assume that $\boldsymbol{X}$ and $\boldsymbol{X}_\omega$ each always have $\mathbf{1}_n$ as a column vector that corresponds to an intercept, where $\mathbf{1}_n$ is an n-dimensional vector of ones, and $\lim_{n\to\infty} \boldsymbol{X}_\omega' \boldsymbol{X}_\omega / n$ exists and is positive definite. The matrix form of the candidate model (2.1) is given by

$$M\colon\ \boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}\boldsymbol{\Xi}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)\ , \tag{2.2}$$

where $\boldsymbol{I}_n$ is an identity matrix of size n. Here, $\boldsymbol{A} \otimes \boldsymbol{B}$ denotes an the Kronecker product of an $m \times n$ matrix $\boldsymbol{A}$ and a $p \times q$ matrix $\boldsymbol{B}$, which is an $mp \times nq$ matrix defined by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix},$$

where a_{ij} is the (i, j)-th element of $\boldsymbol{A}$ (see, e.g., [10, chap. 16.1]). Additionally, $\boldsymbol{Z} \sim N_{n\times p}(\boldsymbol{A}, \boldsymbol{B} \otimes \boldsymbol{C})$ denotes that an $n \times p$ random matrix $\boldsymbol{Z}$ is distributed according to the $n \times p$ matrix normal distribution with a mean matrix $E[\boldsymbol{Z}] = \boldsymbol{A}$ and a covariance matrix $Cov[(\boldsymbol{Z})] = \boldsymbol{B} \otimes \boldsymbol{C}$ (see, e.g., [26, p. 91, def. 3.3.1]), i.e., $\mathrm{vec}(\boldsymbol{Z}) \sim N_{np}(\mathrm{vec}(\boldsymbol{A}), \boldsymbol{B} \otimes \boldsymbol{C})$, where $\mathrm{vec}(\boldsymbol{Z})$ is an operator that transforms a matrix to a vector by stacking the first to the last columns of $\boldsymbol{Z}$, i.e., $\mathrm{vec}(\boldsymbol{Z}) = (\boldsymbol{z}_1', ..., \boldsymbol{z}_p')'$ when $\boldsymbol{Z} = (\boldsymbol{z}_1, ..., \boldsymbol{z}_p)$ (see, e.g., [10, chap. 16.2]). The following normal MLRM using the full explanatory variables is called the full model:

$$M_\omega\colon\ \boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}_\omega\boldsymbol{\Xi}_\omega, \boldsymbol{\Sigma}_\omega \otimes \boldsymbol{I}_n)\ , \tag{2.3}$$

where $\boldsymbol{\Xi}_\omega$ and $\boldsymbol{\Sigma}_\omega$ denote a matrix of the unknown regression coefficients and a covariance matrix of the full model, respectively. Although the normal distribution is assumed, we are not able to see whether the assumption is actually correct. A natural assumption for the generating mechanism of $\boldsymbol{Y}$ is

$$\begin{gathered} M_*\colon \boldsymbol{Y} = \boldsymbol{\Gamma}_* + \boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_*^{1/2}\ , \quad \boldsymbol{\mathcal{E}} = (\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_n)'\ , \quad \boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_n \sim i.i.d.\ \boldsymbol{\varepsilon}\ , \\ E[\boldsymbol{\varepsilon}] = \mathbf{0}_p\ , \quad Cov[\boldsymbol{\varepsilon}] = \boldsymbol{I}_p\ , \quad E[\|\boldsymbol{\varepsilon}\|^4] = \kappa_4^{(1)} + p\,(p+2)\ , \end{gathered} \tag{2.4}$$

where $\boldsymbol{\Gamma}_*$ and $\boldsymbol{\Sigma}_*$ are the true mean and covariance matrices, respectively, $\mathbf{0}_p$ is a p-dimensional vector of zeros, and $\|\boldsymbol{a}\|$ is the Euclidean norm of the vector $\boldsymbol{a} = (a_1, ..., a_m)'$, i.e, $\|\boldsymbol{a}\| = (a_1^2 + \cdots + a_m^2)^{1/2}$. Here, $\kappa_4^{(1)}$ is called the multivariate kurtosis, which was proposed by [19].

In order to clarify assumptions for deriving information criteria, we separate the candidate models into the following two models:

- Underspecified model: the mean structure does not include that of the true model, i.e., $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\Gamma}_* \neq \boldsymbol{\Gamma}_*$;
- Overspecified model: the mean structure includes that of the true model, i.e., $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{\Gamma}_* = \boldsymbol{\Gamma}_*$.

Here, $\boldsymbol{P}_{\boldsymbol{X}}$ is the projection matrix to the subspace spanned by the columns of $\boldsymbol{X}$, i.e., $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Furthermore, the candidate model whose mean structure dovetails perfectly with that of model M_* is here called the true model. Although Fujikoshi and Satoh [8] used the same terminology, they divided the candidate models by whether a candidate model includes the true model. It emphasizes that we are separating the candidate models based only on the mean structure. Hence, our separation does not depend on whether a distribution of the true model is the normal distribution. Furthermore, we assume that the full model M_ω is the overspecified model and the true model is included in a set of the candidate models. For an additional characteristic of the candidate model, a $p \times p$ matrix of noncentrality parameters is defined by

$$\boldsymbol{\Omega} = \frac{1}{n}\boldsymbol{\Sigma}_*^{-1/2}\boldsymbol{\Gamma}_*'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\Gamma}_*\boldsymbol{\Sigma}_*^{-1/2}. \tag{2.5}$$

It should be noted that $\boldsymbol{\Omega}$ is positive semidefinite and $\boldsymbol{\Omega} = \boldsymbol{O}_{p,p}$ holds if and only if M is an overspecified model, where $\boldsymbol{O}_{p,p}$ is a $p \times p$ matrix of zeroes.

Let $f(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\Sigma})$ be the probability density function of $N_p(\boldsymbol{\eta}, \boldsymbol{\Sigma})$. Then, the log-likelihood function of the candidate model M in (2.2) is derived as

$$\begin{aligned}\ell(\boldsymbol{\Xi}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X}) &= \sum_{i=1}^{n} \log f(\boldsymbol{y}_i|\boldsymbol{\Xi}'\boldsymbol{x}_i, \boldsymbol{\Sigma}) \\ &= -\frac{1}{2}\Big[np\log 2\pi + n\log|\boldsymbol{\Sigma}| \\ &\qquad + \mathrm{tr}\big\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Xi})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Xi})\big\}\Big].\end{aligned} \tag{2.6}$$

By maximizing $\ell(\boldsymbol{\Xi}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X})$, or equivalently solving the likelihood equations $\partial\ell(\boldsymbol{\Xi}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X})/\partial\boldsymbol{\Xi} = \boldsymbol{O}_{k,p}$ and $\partial\ell(\boldsymbol{\Xi}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X})/\partial\boldsymbol{\Sigma} = \boldsymbol{O}_{p,p}$, the maximum likelihood (ML) estimators of the unknown parameter matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ in the candidate model M are obtained as

$$\hat{\boldsymbol{\Xi}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Y}.$$

In particular, $\boldsymbol{S}$ denotes a standardized $\hat{\boldsymbol{\Sigma}}$ defined by $\boldsymbol{S} = \boldsymbol{\Sigma}_*^{-1/2}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}_*^{-1/2}$. Furthermore, $(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega, \boldsymbol{S}_\omega)$ denotes $(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}, \boldsymbol{S})$ in the full model M_ω in (2.3). By substituting $(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}})$ into (2.6), the maximum log-likelihood of the candidate model M is derived as

$$\ell(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}|\boldsymbol{Y}, \boldsymbol{X}) = -\frac{n}{2}\Big\{p(\log 2\pi + 1) + \log|\hat{\boldsymbol{\Sigma}}|\Big\}.$$

Let $\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})$ be an expected negative twofold log-likelihood function:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X}) \;&=\; E_*\big[-2\ell(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{Y}, \boldsymbol{X})\big] \\ &=\; np \log 2\pi + n \log |\boldsymbol{\Sigma}| \\ &\quad +\; \mathrm{tr}\Big[\big\{n\boldsymbol{\Sigma}_* + (\boldsymbol{\Gamma}_* - \boldsymbol{X}\boldsymbol{\Xi})' (\boldsymbol{\Gamma}_* - \boldsymbol{X}\boldsymbol{\Xi})\big\}\, \boldsymbol{\Sigma}^{-1}\Big] , \end{aligned} \tag{2.7}$$

where E_* means the expectation under the true model M_* in (2.4). We define the loss function of the model M measured by the KL information as $\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X})$. Then, a risk function that uses the KL information to assess the gap between the true model and the candidate model is defined by the expectation of the loss function, i.e.,

$$R_{\mathrm{KL}} = E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X})\big] \; . \tag{2.8}$$

In this paper, the candidate model that makes the risk function the smallest is called the principle best model. The following theorem is satisfied for the principle best model (the proof is given in Appendix A.1):

Theorem 2.1. *The principle best model is either the true model or an underspecified model. When $n \to \infty$, the principle best model becomes the true model under the assumption that $E[\mathrm{tr}(\boldsymbol{S}_\omega^{-2})] = O(1)$.*

3. AIC AND BIAS-CORRECTED AICS IN NORMAL MLRMS

Although the risk function R_{KL} in (2.8) assesses the goodness of fit of the model, we cannot use R_{KL} directly because R_{KL} involves unknown parameters. Hence, in practice, an estimate of R_{KL} is needed to select the best model among the candidates. Although a naive estimator of R_{KL} is $-2\ell(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{Y}, \boldsymbol{X})$, it has the following constant bias:

$$B \;=\; R_{\mathrm{KL}} - E_*\big[-2\ell(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{Y}, \boldsymbol{X})\big] \; . \tag{3.1}$$

Thus, an information criterion for selecting the best model is defined by adding $\hat{B}$, an estimator of B, to $-2\ell(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{Y}, \boldsymbol{X})$, i.e.,

$$\mathrm{IC} \;=\; -2\ell(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{Y}, \boldsymbol{X}) + \hat{B} \; . \tag{3.2}$$

The information criterion is specified by the individual $\hat{B}$, because $\hat{B}$ changes based on assumptions of the model M and by an estimation method. As for such assumptions, the following two assumptions are considered:

(A1) The candidate model M in (2.2) is an overspecified model;

(A2) The distribution of the true model M_* in (2.4), called the true distribution, is the normal distribution, i.e., $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}_p, \boldsymbol{I}_p)$.

Nine information criteria used to estimate R_{KL} are enumerated below. The order of the bias of each information criterion for R_{KL} is summarized in Table 1. As for information criteria in the NMLR model other than the nine information criteria used in this paper, see [20, chap. 4].

Table 1: The order of the bias of each criterion.

	Criterion	Bias-Correction Method	Normality		Nonnormality	
			Under-specified	Over-specified	Under-specified	Over-specified
Proposed under Normality	AIC[*1]	——	$O(1)$	$O(n^{-1})$	$O(1)$	$O(1)$
	CAIC[*1,*2]	Exact	$O(1)$	0	$O(1)$	$O(1)$
	MAIC	Moment, Exact	$O(n^{-1})$	$O(n^{-2})$	$O(1)$	$O(1)$
Proposed without Normality	TIC[*3,*4,*5]	Moment	$O(1)$	$O(n^{-1})$	$O(1)$	$O(n^{-1})$
	EIC[*3,*5,*6]	Bootstrap	$O(1)$	$O(n^{-1})$	$O(1)$	$O(n^{-1})$
	$\mathrm{EIC_A}$[*3,*6]	Bootstrap	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$	$O(n^{-1})$
	CV[*4]	Cross-validation	$O(1)$	$O(n^{-1})$	$O(1)$	$O(n^{-1})$
	$\mathrm{AIC_J}$[*4,*5,*7]	Jackknife, Exact	$O(1)$	0	$O(1)$	$O(n^{-1})$
	$\mathrm{CAIC_J}$[*4,*7]	Jackknife, Exact	$O(1)$	0	$O(1)$	$O(n^{-2})$

[*1] The number of explanatory variables in the best model selected by the CAIC is less than or equal to that in the best model selected by the AIC.

[*2] This is the UMVUE of the risk function when assumptions A1 and A2 hold.

[*3] These are asymptotically equivalent when assumption A1 holds. The differences are $O_p(n^{-1/2})$.

[*4] These are asymptotically equivalent. The differences are $O_p(n^{-1})$.

[*5] When $O(n^{-2})$ term is neglected and assumption A1 holds, the absolute value of the bias of the $\mathrm{AIC_J}$ is smaller than those of the TIC and EIC.

[*6] The only difference between these two criteria is the resampling method.

[*7] When the $O(n^{-2})$ term is neglected and assumption A1 holds, the variance of the $\mathrm{CAIC_J}$ is smaller than that of the $\mathrm{AIC_J}$.

3.1. AIC

Under the assumption that the candidate model is completely specified, Akaike [1, 2] proposed AIC by estimating a bias of a negative twofold maximum log-likelihood to a risk function as twice the number of parameters. According to the general formula of AIC, $\hat{B}$ in (3.2) is $\hat{B}_{\mathrm{AIC}} = 2pk + p(p+1)$. Thus, the AIC in the model M is expressed as

$$\mathrm{AIC} \;=\; np(\log 2\pi + 1) + n \log |\hat{\boldsymbol{\Sigma}}| + 2pk + p(p+1) \; .$$

From the assumption to derive an bias of AIC, the bias of the AIC in the model M to R_{KL} becomes $O(n^{-1})$ when assumptions A1 and A2 are satisfied simultaneously. However, the order of the bias changes to $O(1)$, i.e., AIC has constant bias, when either of assumptions A1 or A2 are violated (for details, see, e.g., [8, 9, 27]).

3.2. Corrected AIC

When assumptions A1, A2, and an additional assumption $n > p + k + 1$ are satisfied, Bedrick and Tsai [3] calculated the exact form of B as $\hat{B}_{\mathrm{CAIC}} = n(n+k)p/(n-k-p-1) - np$ and proposed the corrected AIC (CAIC)[1] by replacing $\hat{B}$ in (3.2) with $\hat{B}_{\mathrm{CAIC}}$ as

$$\begin{aligned} \mathrm{CAIC} &= np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| + \frac{n(n+k)p}{n-p-k-1} \\ &= \mathrm{AIC} + \frac{(p+k+1)\,(p+2k+1)p}{n-p-k-1}\,. \end{aligned}$$

The CAIC is an unbiased estimator of R_{KL} under assumptions A1 and A2, and is congruent with the bias-corrected AIC proposed by [12, 24] when $p = 1$. Additionally, extending the result of [5] to the multivariate case provides that the CAIC is a UMVUE of the risk function R_{KL} when assumptions A1 and A2 are satisfied simultaneously (for a short proof, see [34]). From the definition of the CAIC and its unbiasedness under assumptions A1 and A2, we can see that the AIC in an overspecified model underestimates R_{KL}, and the amount of the underestimation becomes large as k increases. This will cause the undesirable property of the AIC that the AIC has a tendency to overestimate the best model when the sample size is not large enough and the number of candidate models is large. The problem of the AIC can be avoided by using CAIC instead of the AIC, because the number of explanatory variables of the best model selected by the CAIC will be less than or equal to the number selected by the AIC (the proof is given in Appendix A.2). Because of $\mathrm{CAIC} = \mathrm{AIC} + O(n^{-1})$, as in the case of the AIC, the order of the bias of the CAIC to R_{KL} becomes $O(1)$, i.e., the CAIC has a constant bias, when either of assumptions A1 or A2 are violated.

3.3. Modified AIC

When assumption A2 holds but assumption A1 does not hold, and $n > p + k + 1$, Fujikoshi and Satoh [8] estimated B by $\hat{B}_{\mathrm{MAIC}} = \hat{B}_{\mathrm{CAIC}} + 2k\,\mathrm{tr}(\boldsymbol{L}) - \mathrm{tr}(\boldsymbol{L})^2 - \mathrm{tr}(\boldsymbol{L}^2)$, where $\boldsymbol{L}$ is a $p \times p$ matrix defined by $\boldsymbol{L} = (n-k)\hat{\boldsymbol{\Sigma}}_{\omega}\hat{\boldsymbol{\Sigma}}^{-1}/$

[1] Although Bedrick and Tsai [3] used $\mathrm{AIC_c}$ as the abbreviated symbol, we use CAIC following the notation of [8].

$(n-k_\omega)-\boldsymbol{I}_p$, and proposed the modified AIC (MAIC) by replacing $\hat{B}$ in (3.2) with $\hat{B}_{\mathrm{MAIC}}$ as

$$\mathrm{MAIC} \;=\; \mathrm{CAIC} + 2k\,\mathrm{tr}(\boldsymbol{L}) - \mathrm{tr}(\boldsymbol{L})^2 - \mathrm{tr}(\boldsymbol{L}^2)\;.$$

The bias of the MAIC to R_{KL} becomes $O(n^{-2})$ when assumptions A1 and A2 are satisfied simultaneously, and it becomes $O(n^{-1})$ when assumption A2 holds but assumption A1 does not (see, [8]). However, the bias changes to $O(1)$, i.e., the MAIC also has constant bias, when assumption A2 is violated, because B_{AIC} depends on a nonnormality of the true model when assumption A2 is violated (see, e.g., [9, 27]).

3.4. Takeuchi's Information Criterion

Takeuchi [25] revaluated an asymptotic bias under model misspecification and proposed Takeuchi's information criterion (TIC) by estimating such an asymptotic bias with a moment-estimation method. According to the general formula of the TIC, a bias-correction term of the TIC in the model M can be calculated as $\hat{B}_{\mathrm{TIC}} = \hat{B}_{\mathrm{AIC}} + \hat{\kappa}_4^{(1)} + 2\sum_{i=1}^n (1-h_i)\,(\hat{r}_i^2 - p)$ (for details of the derivation, see [9]), where $\hat{r}_i$ is a squared standardized residual of the i-th individual, $\hat{\kappa}_4^{(1)}$ is an estimator of the multivariate kurtosis $\kappa_4^{(1)}$ in (2.4), and h_i is a constant, which are given by

$$\begin{aligned} \hat{r}_i^2 \;&=\; (\boldsymbol{y}_i - \hat{\boldsymbol{\Xi}}'\boldsymbol{x}_i)'\,\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\Xi}}'\boldsymbol{x}_i)\;, \\ \hat{\kappa}_4^{(1)} \;&=\; \frac{1}{n}\sum_{i=1}^n \hat{r}_i^4 - p(p+2)\;, \qquad h_i = 1 - \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i\;. \end{aligned} \tag{3.3}$$

Hence, the TIC in the model M is expressed as

$$\mathrm{TIC} \;=\; \mathrm{AIC} + \hat{\kappa}_4^{(1)} + 2\sum_{i=1}^n (1-h_i)\,(\hat{r}_i^2 - p)\;.$$

When $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independently and identically distributed, the bias of the TIC to the risk function is $O(n^{-1})$ under any model misspecification. However, in the case of multivariate linear regression, $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independent but not identically distributed. This leads to the less well-known fact that the bias of the TIC in the model M to R_{KL} is $O(n^{-1})$ when assumption A1 holds but assumption A2 does not, and becomes $O(1)$ when assumption A1 is violated (see, [9]). By conducting numerical experiments, many authors have verified a fact that although the TIC theoretically reduces the bias caused by violating normality, the TIC cannot reduce the bias successfully unless the sample size is huge (see, e.g., [9, 27]). This occurs because the TIC consists of an estimator for the multivariate kurtosis $\hat{\kappa}_4^{(1)}$.

Yanagihara [28] presented numerical results that showed that $\hat{\kappa}_4^{(1)}$ has a huge bias to $\kappa_4^{(1)}$ if n is not huge. Hence, the TIC also has a huge bias to R_{KL} if n is not huge.

When $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independently and identically distributed, the bias of TIC can be reduced to $O(n^{-2})$ by using a formula in [35], which is a special case of those in [15] and [30]. However, as stated already, $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independent but not identically distributed in the case of the multivariate linear regression. Regrettably, we cannot correct the bias of TIC by using their formula.

3.5. Extended Information Criterion

The serious problem with TIC comes from the moment estimation of a bias. Ishiguro *et al.* [13] cleared this problem by using the bootstrap method for an estimation of the bias, and proposed the extended information criterion (EIC). Let $\boldsymbol{D}_b$ be an $n \times n$ matrix to express the b-th bootstrap resample of $\boldsymbol{Y}$ as

$$\boldsymbol{D}_b = (\boldsymbol{d}_{b,1}, ..., \boldsymbol{d}_{b,n})', \quad \boldsymbol{d}_{b,1}, ..., \boldsymbol{d}_{b,n} \sim i.i.d.\ MN_n(1; n^{-1}\boldsymbol{1}_n)\,, \tag{3.4}$$

where $MN_n(1; n^{-1}\boldsymbol{1}_n)$ denotes the n-variate one-trial multinomial distribution with the same cell probabilities $1/n$. Following [7], the b-th bootstrap resample of $\boldsymbol{Y}$ is $\tilde{\boldsymbol{Y}}_b = \boldsymbol{X}\hat{\boldsymbol{\Xi}} + \boldsymbol{D}_b(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Y}$. Let $\tilde{\boldsymbol{\Sigma}}_b$ be the ML estimator of $\boldsymbol{\Sigma}$ evaluated from $(\tilde{\boldsymbol{Y}}_b, \boldsymbol{X})$. From the general formula of EIC in [14], an estimator of the bias obtained from the bootstrap method with m repetitions is given by $\hat{B}_{\mathrm{EIC}} = m^{-1}\sum_{b=1}^{m} \mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}_b^{-1}(\boldsymbol{Y} - \boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{Y}}_b)'(\boldsymbol{Y} - \boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{Y}}_b)\} - np$. Then, by using (3.2), the EIC in the model M is expressed as follows (see, [27]):

$$\mathrm{EIC} = np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| + \frac{1}{m}\sum_{b=1}^{m} \mathrm{tr}\Big\{\tilde{\boldsymbol{\Sigma}}_b^{-1}(\boldsymbol{Y} - \boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{Y}}_b)'(\boldsymbol{Y} - \boldsymbol{P}_{\boldsymbol{X}}\tilde{\boldsymbol{Y}}_b)\Big\}\,.$$

When $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independently and identically distributed, the bias of the EIC to the risk function is $O(n^{-1})$ under any model misspecification like the TIC. However, in the case of multivariate linear regression, $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independent but not identically distributed. Hence, the bias of EIC is $O(n^{-1})$ under assumption A1, but that changes to $O(1)$, i.e., the EIC has constant bias (as does the TIC), when assumption A1 is violated (see, [27]). In particular, $\mathrm{EIC} = \mathrm{TIC} + O_p(n^{-1/2})$ holds when both $m \to \infty$ and assumption A1 holds (the proof is given in Appendix A.3). Although the theoretical bias of the EIC has the same order as that of the TIC, the bias of the EIC tends to be smaller than that of the TIC (see, [27]) because the EIC does not directly use $\hat{\kappa}_4^{(1)}$ for estimating the bias.

3.6. Adjusted EIC

Fujikoshi *et al.* [9] proposed an adjusted version of the EIC in the model M by using a full-model-based resampling instead of a candidate-model-based resampling. We call this the adjusted EIC ($\mathrm{EIC_A}$). Let $\bar{\boldsymbol{Y}}_b$ be the b-th bootstrap resample of $\boldsymbol{Y}$ based on the full model M_ω given by $\bar{\boldsymbol{Y}}_b = \boldsymbol{X}_\omega\hat{\boldsymbol{\Xi}}_\omega + \boldsymbol{D}_b(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}_\omega})\boldsymbol{Y}$, where $\boldsymbol{D}_b$ is given by (3.4), and let $\bar{\boldsymbol{\Sigma}}_b$ be the ML estimator of $\boldsymbol{\Sigma}$ evaluated from $(\bar{\boldsymbol{Y}}_b, \boldsymbol{X})$. Then, $\hat{B}_{\mathrm{EIC_A}}$, which is an estimator of the bias obtained from a full-model-based bootstrap method with m repetitions, is given by replacing $\tilde{\boldsymbol{Y}}_b$ and $\tilde{\boldsymbol{\Sigma}}_b$ in $\hat{B}_{\mathrm{EIC}}$ with $\bar{\boldsymbol{Y}}_b$ and $\bar{\boldsymbol{\Sigma}}_b$. By using (3.2), the $\mathrm{EIC_A}$ in the model M is expressed as follows (see, [9]):

$$\mathrm{EIC_A} = np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| + \frac{1}{m}\sum_{b=1}^{m}\mathrm{tr}\left\{\bar{\boldsymbol{\Sigma}}_b^{-1}(\boldsymbol{Y}-\boldsymbol{P}_{\boldsymbol{X}}\bar{\boldsymbol{Y}}_b)'(\boldsymbol{Y}-\boldsymbol{P}_{\boldsymbol{X}}\bar{\boldsymbol{Y}}_b)\right\}.$$

The bias of the $\mathrm{EIC_A}$ to the risk function is always $O(n^{-1})$ (see, [9]). In particular, $\mathrm{EIC_A} = \mathrm{TIC} + O_p(n^{-1/2})$ holds when $m \to \infty$ and assumption A1 holds (the proof is given in Appendix A.3).

3.7. Cross-Validation Criterion

The cross-validation (CV) criterion proposed by [22] estimates a risk function directly, and it can be defined without an estimator of a bias of a negative twofold maximum log-likelihood to a risk function. We know that n repetitions of the calculations for the ML estimator of $(\boldsymbol{\Xi}, \boldsymbol{\Sigma})$ are needed for the CV criterion in the model M. However, Yoshimoto *et al.* [36] gave the formula to derive the CV criterion in the model M without the n repetitions as

$$\mathrm{CV} = np\log\left(\frac{2\pi n}{n-1}\right) + n\log|\hat{\boldsymbol{\Sigma}}| + \sum_{i=1}^{n}\left\{\log\left(1-\frac{\hat{r}_i^2}{nh_i}\right) + \frac{(n-1)\,\hat{r}_i^2}{h_i(nh_i-\hat{r}_i^2)}\right\}, \tag{3.5}$$

where $\hat{r}_i^2$ and h_i are given by (3.3). From [23], $\mathrm{CV} = \mathrm{TIC} + O_p(n^{-1})$ always holds if $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independently and identically distributed. In the case of multivariate linear regression, we can prove that $\mathrm{CV} = \mathrm{TIC} + O_p(n^{-1})$ always holds (the proof is given in Appendix A.4). From this result, the bias of the CV criterion is $O(n^{-1})$ under assumption A1, but like the TIC, it has a constant bias when assumption A1 is violated.

Yanagihara and Fujisawa [30], and Yanagihara *et al.* [33, 35] proposed bias-corrected CV criteria, which are criteria correcting the bias of CV to the risk function, under general statistical models. It should be noted that their results cannot be applied to the case of multivariate linear regression because they proposed the bias-corrected CV under the assumption that $\boldsymbol{y}_1, ..., \boldsymbol{y}_n$ are independently and identically distributed.

3.8. Jackknife AIC

Yanagihara [32] proposed a bias-corrected AIC by using a jackknife method for estimating B and by adjusting such an estimator of B to become an unbiased estimator when assumptions A1 and A2 are satisfied simultaneously. We call this a jackknife AIC ($\mathrm{AIC_J}$). Let $\hat{B}_{\mathrm{AIC_J}} = c\sum_{i=1}^{n} Q(\hat{r}_i^2/h_i;1)/h_i - np$, where $\hat{r}_i^2$ and h_i are given by (3.3), $Q(x;\lambda)$ is a function with respect to x and c is a positive constant, as follows:

$$Q(x;\lambda) = x\left(1-\frac{x}{n}\right)^{-\lambda}, \qquad c = \frac{(n+k)\,(n-k-p-2)}{(n-k-p-1)\sum_{i=1}^{n} h_i^{-1}}. \tag{3.6}$$

Then, by using (3.2), the $\mathrm{AIC_J}$ for the model M is (see [27]):

$$\mathrm{AIC_J} \;=\; np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| + c\sum_{i=1}^{n}\frac{Q(\hat{r}_i^2/h_i;1)}{h_i}.$$

From [27], $\mathrm{AIC_J} = \mathrm{TIC} + O_p(n^{-1})$ always holds. Hence, like the TIC, the bias of the $\mathrm{AIC_J}$ is $O(n^{-1})$ under assumption A1, but it has a constant bias when assumption A1 is violated (see, [27]). On the other hand, when assumptions A1 and A2 are satisfied simultaneously, the $\mathrm{AIC_J}$ is an unbiased estimator of R_{KL}. Although the order of the bias of the $\mathrm{AIC_J}$ is the same as that of the bias of the TIC and EIC, it has been verified numerically that the bias of the $\mathrm{AIC_J}$ in an overspecified model becomes very small (see, [27]). Moreover, Yanagihara [27] showed a theoretical result that the absolute value of the bias of the $\mathrm{AIC_J}$ is smaller than those of either the TIC or EIC under assumption A1 when the $O(n^{-2})$ term of B is neglected.

3.9. Corrected Jackknife AIC

Since the bias of the $\mathrm{AIC_J}$ does not disappear in theory, Yanagihara *et al.* [32] proposed a corrected $\mathrm{AIC_J}$ ($\mathrm{CAIC_J}$) that corrects the bias while maintaining the desirable characteristic of keeping the bias very small numerically. Let $\hat{B}_{\mathrm{CAIC_J}} = c^{+}\sum_{i=1}^{n}\{1+a_1(1-h_i)\}Q(\tilde{r}_i^2/h_i;a_0) - np$, where $\hat{r}_i^2$ and h_i are given by (3.3) and $Q(x;\lambda)$ is given by (3.6), c^{+} and a_j $(j=0,1)$ being positive constants given by

$$c^{+} = \frac{(n+k)\,(n-k-p-2a_0)\;\Gamma\!\left(\frac{n-k}{2}+\frac{1}{n}\right)\Gamma\!\left(\frac{n-k-p}{2}\right)}{(n+a_1k)\,(n-k-p-1)\;\Gamma\!\left(\frac{n-k}{2}\right)\Gamma\!\left(\frac{n-k-p}{2}+\frac{1}{n}\right)}, \qquad a_j = \frac{n+j-1}{n+j}.$$

Here, $\Gamma(x)$ is the gamma function. Then, by using (3.2), the $\mathrm{CAIC_J}$ for the model M is (see [32])

$$\mathrm{CAIC_J} = np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| + c^{+}\sum_{i=1}^{n}\left\{1 + a_1(1-h_i)\right\}Q(\tilde{r}_i^2/h_i; a_0)\ .$$

When assumptions A1 and A2 are satisfied simultaneously, like the $\mathrm{AIC_J}$, the $\mathrm{CAIC_J}$ is an unbiased estimator of R_{KL}. Although, like the $\mathrm{AIC_J}$, the $\mathrm{CAIC_J}$ has constant bias when assumption A1 is violated, the $\mathrm{CAIC_J}$ reduces the bias of the $\mathrm{AIC_J}$ to $O(n^{-2})$ when assumption A1 holds (see, [32]). Moreover, Yanagihara *et al.* [32] showed a theoretical result under assumption A1 that $\mathrm{CAIC_J}$ reduces not only the bias of $\mathrm{AIC_J}$ but also the variance of $\mathrm{AIC_J}$ when we neglect the $O(n^{-2})$ terms.

4. NUMERICAL COMPARISON

In this section, we numerically compare performances of variable-selection methods using the nine information criteria described in the previous section. The best models selected by the nine information criteria are asymptotically equivalent, and in particular, an underspecified model is never selected as the best model when $n \to \infty$ (the proof is given in Appendix A.5). This indicates that numerical comparisons with variable-selection methods using the nine information criteria are meaningless when the sample size is large. Hence, we conduct numerical experiments using small and moderate sample sizes. We study performances of the nine information criteria by applying variable-selection methods to simulation data first, and by applying variable-selection methods to real data later.

4.1. A Simulation Study

4.1.1. Target Characteristics

In the simulation study, performances as an estimator of the risk function are studied at first, and performances as a model selector are studied later. In a numerical experiment to check performances as an estimator of the risk function, we compare the nine information criteria by the following three characteristics of an estimator:

(C-1) The mean of the information criterion $E[\mathrm{IC}]$;

(C-2) The standard deviation of the information criterion $\sqrt{Var[\mathrm{IC}]}$;

(C-3) The root-mean-square error (RMSE) of the information criterion $\sqrt{Var[\mathrm{IC}] + \big(E[\mathrm{IC}] - R_{\mathrm{KL}}\big)^2}$.

On the other hand, in a numerical experiment to check performances as a model selector, we compare the nine information criteria by the following two characteristics of a model selector:

(C-4) The probability of selecting the principle best model: the frequency with which the principle best model is selected as the best model;

(C-5) The prediction error (PE) of the best model: the expected loss function of the best model which is chosen by the information criterion; PE is defined as follows:

$$\mathrm{PE} = \frac{1}{n} E_* \big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_{\mathrm{best}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{best}} \,|\, \boldsymbol{X}_{\mathrm{best}}) \big] \,,$$

where $\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})$ is the expected negative twofold log-likelihood function given by (2.7), and $(\hat{\boldsymbol{\Xi}}_{\mathrm{best}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{best}}, \boldsymbol{X}_{\mathrm{best}})$ is $(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}, \boldsymbol{X})$ in the best model.

A high-performance model selector is considered to be an information criterion with a high probability of selecting the principle best model and a small prediction error. According to the basic concept of the model selection based on the risk function minimization, a good variable-selection method is one that can choose the best model for improving the predictive accuracy. Hence, the PE is a more important property than the probability of selecting the principle best model.

The expectations and probabilities in the simulation studies were evaluated by a Monte Carlo simulation with $10,000$ repetitions. The $\hat{B}_{\mathrm{EIC}}$ and $\hat{B}_{\mathrm{EIC_A}}$ were obtained by resampling $1,000$ times, i.e., $m = 1,000$.

4.1.2. Simulation Model

The model in [32] was used as the basic simulation model for generating data. We prepared the $k_\omega - 1$ candidate models M_j $(j = 1, ..., k_\omega - 1)$ with $p = 4$ and $n = 30$ or 100. First, we generated $z_1, ..., z_n \sim i.i.d.\ U(-1, 1)$. Using these $z_1, ..., z_n$, we constructed the $n \times k_\omega$ matrix of explanatory variables $\boldsymbol{X}_\omega$, whose (i, j)-th element is given by $\{(z_i - \bar{z})/s_z\}^{j-1}$ $(i = 1, ..., n;\ j = 1, ..., k_\omega)$, where $\bar{z}$ and s_z are the sample mean and standard deviation, respectively, of $z_1, ..., z_n$. The true model was determined by $\boldsymbol{\Gamma}_* = \boldsymbol{X}_\omega \boldsymbol{\mu}_* \mathbf{1}_4'$ and $\boldsymbol{\Sigma}_*$, whose (i, j)-th element is defined by $(0.8)^{|i-j|}$ $(i = 1, ..., 4;\ j = 1, ..., 4)$. In this simulation study,

we arranged the six $\boldsymbol{\mu}_*$ as

$$
\begin{aligned}
&\text{Case 1:} \quad \boldsymbol{\mu}_* = (0,1,2,4,0,0,0,0)', && (k_\omega = 8),\\
&\text{Case 2:} \quad \boldsymbol{\mu}_* = (0,1,2,4,0.5,0.5,0,0)', && (k_\omega = 8),\\
&\text{Case 3:} \quad \boldsymbol{\mu}_* = (0,1,2,4,0,0,0,0,0,0,0,0,0,0,0,0)', && (k_\omega = 16),\\
&\text{Case 4:} \quad \boldsymbol{\mu}_* = (0,1,2,4,0.5,0.5,0,0,0,0,0,0,0,0,0,0)', && (k_\omega = 16),\\
&\text{Case 5:} \quad \boldsymbol{\mu}_* = (0,1,1,1,-1,-1,2,2,4,0,0,0,0,0,0,0)', && (k_\omega = 16),\\
&\text{Case 6:} \quad \boldsymbol{\mu}_* = (0,1,1,1,-1,-1,2,2,4,0.5,0.5,0,0,0,0,0)', && (k_\omega = 16).
\end{aligned}
$$

The matrix of explanatory variables in M_j $(j = 1, ..., k_\omega - 1)$ consists of the first $(j+1)$ columns of $\boldsymbol{X}_\omega$. Thus, the true models M_* in the cases 1, 2, 3, 4, 5, and 6 are M_3, M_5, M_3, M_5, M_8, and M_{10}, respectively. In a sense, the subindex j expresses the degree of the polynomial regression in M_j.

For generating multivariate nonnormal data, the following data model introduced by [37] was used:

Data Model. Let $w_1, ..., w_q$ $(q \geq p)$ be independent random variables with $E[w_j] = 0$, $E[w_j^2] = 1$ and $E[w_j^4] - 3 = \psi$, and let $\boldsymbol{w} = (w_1, ..., w_q)'$. Further, let r be a random variable that is independent of $\boldsymbol{w}$, with $E[r^2] = 1$ and $E[r^4] = \beta$. Then, an error vector is generated by $\boldsymbol{\varepsilon} = r\boldsymbol{C}'\boldsymbol{w}$, where $\boldsymbol{C} = (\boldsymbol{c}_1, ..., \boldsymbol{c}_q)'$ is a $q \times p$ matrix satisfying $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}_p$. Then, the multivariate kurtosis of this model becomes $\kappa_4^{(1)} = \beta\psi \sum_{j=1}^q \|\boldsymbol{c}_j\|^4 + (\beta - 1)\, p(p+2)$.

Let χ_f be a random variable from the chi-square distribution with f degrees of freedom, and let $\boldsymbol{C}_0$ be a $(p+1) \times p$ matrix defined by $\boldsymbol{C}_0 = (\boldsymbol{I}_p, \boldsymbol{1}_p)'(\boldsymbol{I}_p + \boldsymbol{1}_p\boldsymbol{1}_p')^{-1/2}$. By using the data model, we generate error vectors with the following three distributions:

(1) *Normal Distribution:* $w_j \sim N(0,1)$, $r = 1$ and $\boldsymbol{C} = \boldsymbol{I}_p$ $(\kappa_4^{(1)} = 0)$;

(2) *Laplace Distribution:* w_j is generated from a Laplace distribution with mean 0 and standard deviation 1, $r = (6/\chi_8^2)^{1/2}$ and $\boldsymbol{C} = \boldsymbol{C}_0$ $(\kappa_4^{(1)} = 4.5 \times p^2(p+1)^{-1} + p(p+2)/2)$;

(3) *Skew Laplace Distribution:* w_j is generated from a skew Laplace distribution with location parameter 0, dispersion parameter 1, and skew parameter 1, standardized by mean 3/4 and standard deviation $(23)^{1/2}/4$, $r = (6/\chi_8^2)^{1/2}$ and $\boldsymbol{C} = \boldsymbol{C}_0$ $(\kappa_4^{(1)} \approx 4.88 \times p^2(p+1)^{-1} + p(p+2)/2)$.

For details of the skew Laplace distribution, see, e.g., [17]. It is easy to see that data models 1 and 2 are symmetric distributions, and data model 3 is a skewed distribution. Moreover, the size of the kurtosis $\kappa_4^{(1)}$ in each model satisfies the inequality: model 1 < model 2 < model 3.

4.1.3. Results of Simulation Study

Figure 1 shows R_{KL} and the mean of each criterion in case 1. Since the shapes of the figures were almost the same, we omit the results for cases 2 to 6 to save space. The horizontal axis of the figures expresses numbers of candidate models, i.e., the subindex j of M_j. We see that the biases of the $\mathrm{AIC_J}$ and $\mathrm{CAIC_J}$ were very small under any distribution. As for the size of the bias, the AIC most underestimated the risk function, and the CV criterion overestimated the risk function in the most cases. The size of the bias of the TIC was almost the same as that of

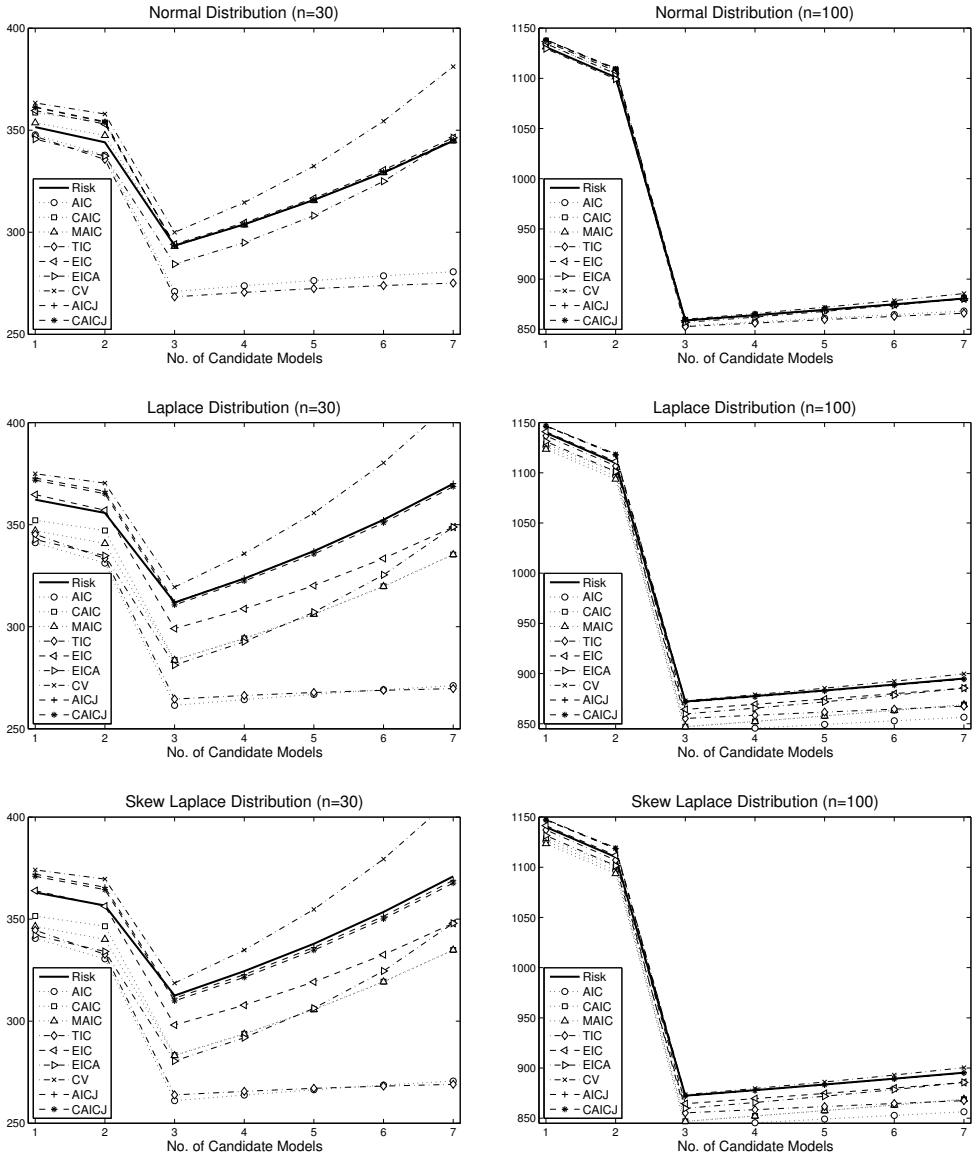

Figure 1: Risk function and the average of each criterion (Case 1).

the AIC. This is because the estimate of the multivariate kurtosis $\hat{\kappa}_4^{(1)}$ for the TIC was close to 0 when the sample size was not large enough. Moreover, as the number of variables in the model increased, the biases of the AIC and TIC increased.

Tables 2 and 3 show, for case 1 and for each information criterion, the standard deviation and the RMSE. Since the tendencies were almost the same, to save space, we omit the results for M_2, M_4, M_5, and M_6, and in cases 2 to 6.

Table 2: Standard deviation of each criterion (Case 1).

n	Dist.	Model	AIC	CAIC	MAIC	TIC	EIC	EIC_A	CV	AIC_J	CAIC_J
30	1	1	15.010	15.010	15.033	15.106	15.342	15.179	16.007	15.998	15.899
		3	17.416	17.416	17.476	17.567	17.842	17.813	19.465	19.358	19.010
		7	19.007	19.007	19.007	19.228	19.680	19.680	30.358	28.239	24.748
	2	1	24.300	24.300	24.359	25.931	30.636	25.933	39.426	39.264	38.073
		3	29.050	29.050	29.123	30.758	35.666	31.824	51.824	50.891	48.977
		7	30.194	30.194	30.194	31.440	35.972	35.972	70.243	64.135	59.042
	3	1	24.539	24.539	24.626	26.264	31.183	26.330	39.878	39.717	38.532
		3	29.102	29.102	29.199	30.828	35.906	31.930	53.943	52.920	50.881
		7	30.317	30.317	30.317	31.546	36.130	36.130	72.282	65.915	61.491
100	1	1	25.465	25.465	25.460	25.490	25.519	25.501	25.519	25.519	25.518
		3	29.346	29.346	29.343	29.401	29.410	29.403	29.457	29.457	29.449
		7	29.896	29.896	29.896	29.995	29.968	29.968	30.268	30.263	30.171
	2	1	45.873	45.873	45.892	48.881	50.177	48.966	54.003	54.025	53.871
		3	54.960	54.960	54.964	58.601	60.232	59.079	65.510	65.512	65.312
		7	55.323	55.323	55.323	58.706	60.240	60.240	66.751	66.645	66.355
	3	1	46.667	46.667	46.682	50.057	51.413	50.127	55.152	55.176	55.033
		3	55.358	55.358	55.358	59.470	61.296	60.043	66.796	66.801	66.601
		7	55.669	55.669	55.669	59.438	61.244	61.244	67.987	67.877	67.623

Table 3: RMSE of each criterion (Case 1).

n	Dist.	Model	AIC	CAIC	MAIC	TIC	EIC	EIC_A	CV	AIC_J	CAIC_J
30	1	1	15.486	16.625	15.181	15.772	17.357	16.290	19.905	18.803	18.599
		3	28.397	17.416	17.478	30.642	17.855	19.981	20.531	19.358	19.010
		7	66.895	19.007	19.007	72.312	19.740	19.740	47.359	28.242	24.749
	2	1	32.159	26.318	28.698	30.994	30.735	32.465	41.417	40.677	39.253
		3	58.144	40.404	40.567	56.425	37.878	44.191	52.376	50.891	48.990
		7	103.424	45.985	45.985	105.162	41.763	41.763	81.197	64.135	59.059
	3	1	33.300	27.123	29.715	32.153	31.195	33.695	41.371	40.715	39.331
		3	59.137	41.222	41.410	57.603	38.675	45.242	54.292	52.935	50.948
		7	104.755	47.094	47.094	106.657	42.810	42.810	81.953	65.943	61.577
100	1	1	25.637	26.089	25.462	25.719	26.102	25.552	26.554	26.460	26.449
		3	29.818	29.346	29.344	30.044	29.413	29.471	29.475	29.458	29.450
		7	32.396	29.896	29.896	33.371	29.969	29.969	30.669	30.263	30.171
	2	1	47.841	47.144	48.692	48.967	50.191	49.714	54.467	54.451	54.270
		3	62.714	60.405	60.411	60.963	60.729	60.356	65.514	65.514	65.316
		7	67.442	61.137	61.137	64.859	60.990	60.990	66.914	66.646	66.358
	3	1	48.672	47.973	49.517	50.139	51.431	50.850	55.661	55.646	55.473
		3	63.288	60.962	60.964	61.888	61.811	61.352	66.804	66.801	66.602
		7	67.982	61.641	61.641	65.645	62.010	62.010	68.174	67.877	67.624

We can see in the tables that the standard deviations of the AIC and CAIC were the smallest and those of the MAIC and TIC were the second smallest. The standard deviations of the EIC and $\mathrm{EIC_A}$ were larger than that of the AIC, but smaller than those of the CV, $\mathrm{AIC_J}$, and $\mathrm{CAIC_J}$. The standard deviation of the CV criterion was the largest among all the information criteria considered. On the other hand, the RMSEs of the AIC and TIC became large when the sample size was small because their biases became large. The RMSEs of the CV criterion, the $\mathrm{AIC_J}$, and $\mathrm{CAIC_J}$ were also large because their standard deviations became large. In all cases, there was a tendency for the standard deviation and RMSE to become large when $\kappa_4^{(1)}$ was large.

Table 4: Probabilities of selecting the principle best model.

Case	n	Dist.	AIC	CAIC	MAIC	TIC	EIC	$\mathrm{EIC_A}$	CV	$\mathrm{AIC_J}$	$\mathrm{CAIC_J}$
1	30	1	69.07	98.44	*99.41*	60.20	97.92	**99.62**	98.66	95.12	96.07
		2	70.19	98.46	*99.55*	54.35	94.19	**99.64**	95.02	91.59	92.65
		3	69.68	98.35	*99.41*	53.84	94.42	**99.74**	95.18	91.73	92.84
	100	1	85.11	92.59	*93.82*	82.51	92.51	**94.28**	93.63	91.75	91.87
		2	85.50	92.94	*94.18*	79.39	90.22	**96.22**	93.01	91.13	91.21
		3	85.04	92.20	*93.70*	79.09	89.87	**96.22**	92.79	90.78	90.96
2	30	1	34.70	87.34	*93.48*	26.83	86.92	**95.33**	90.71	79.01	80.98
		2	30.82	84.57	*91.54*	21.99	80.84	**95.27**	88.84	77.52	79.96
		3	30.27	84.07	91.04	22.15	80.26	95.07	88.92	77.08	79.19
	100	1	**56.85**	50.78	47.78	*56.66*	50.40	46.13	47.42	51.00	51.03
		2	**58.45**	52.19	49.07	*54.17*	46.90	39.82	41.18	44.48	44.73
		3	**58.55**	52.08	49.58	*54.50*	47.60	40.46	41.86	45.09	45.01
3	30	1	50.70	98.20	**99.04**	15.16	97.56	89.42	*98.40*	94.24	96.10
		2	48.98	*98.26*	**99.46**	12.22	94.18	89.08	95.22	90.12	92.86
		3	49.86	*98.40*	**99.28**	12.54	94.58	89.78	95.08	90.08	92.54
	100	1	84.64	92.40	*93.59*	81.22	92.21	91.36	**93.62**	91.45	91.57
		2	84.39	92.22	**93.25**	76.86	89.33	*92.68*	92.57	90.40	90.57
		3	84.63	92.54	**93.82**	76.68	89.64	92.97	*93.14*	91.01	91.20
4	30	1	23.10	86.92	**92.48**	6.04	86.08	63.20	*89.32*	76.76	80.28
		2	20.14	83.68	**89.82**	3.64	78.44	60.52	*87.80*	73.84	78.14
		3	20.60	83.80	**90.28**	4.80	80.30	59.94	*88.42*	75.48	78.72
	100	1	**55.03**	49.49	46.27	*52.55*	49.38	50.64	46.02	49.64	50.02
		2	**57.20**	*52.13*	48.85	50.83	47.24	48.80	41.48	44.49	44.66
		3	**57.01**	*52.57*	49.51	50.34	47.63	49.27	41.95	45.03	45.32
5	30	1	0.00	13.14	*32.36*	0.00	16.97	9.35	**52.99**	14.86	16.92
		2	0.01	12.04	*27.49*	0.00	19.93	11.14	**59.57**	24.44	27.32
		3	0.03	11.98	*27.77*	0.01	18.17	10.45	**58.67**	23.61	26.23
	100	1	81.26	93.78	*96.24*	69.55	93.84	**96.94**	94.02	85.14	90.15
		2	80.96	93.57	*96.04*	65.05	91.92	**97.77**	93.62	83.40	89.14
		3	80.31	93.72	*96.19*	65.35	92.00	**97.70**	93.28	83.50	89.07
6	30	1	0.00	12.43	**43.81**	0.00	17.70	*35.74*	29.85	9.50	11.53
		2	0.02	12.39	**36.86**	0.00	24.16	*34.66*	32.36	16.43	22.50
		3	0.01	12.24	**38.17**	0.01	24.08	*35.10*	33.40	17.43	23.29
	100	1	58.23	80.14	*85.79*	45.61	80.25	**87.91**	80.46	65.66	72.51
		2	57.59	79.48	*85.09*	42.72	78.61	**90.24**	81.20	65.54	72.78
		3	58.58	79.45	*85.18*	43.79	78.75	**89.62**	81.20	66.21	73.27

Note: Bold and italic fonts indicate the highest and second highest probabilities of selecting the principle best model.

Table 5: Prediction errors of the best model.

Case	n	Dist.	AIC	CAIC	MAIC	TIC	EIC	EIC_A	CV	AIC_J	CAIC_J
1	30	1	10.338	9.810	*9.795*	10.494	9.816	**9.790**	9.803	9.853	9.840
		2	11.014	10.427	*10.405*	11.304	10.469	**10.402**	10.481	10.527	10.512
		3	11.044	10.452	*10.432*	11.338	10.492	**10.424**	10.503	10.551	10.534
	100	1	8.619	8.603	*8.601*	8.624	8.603	**8.599**	*8.601*	8.605	8.604
		2	8.752	8.735	*8.733*	8.764	8.740	**8.729**	8.735	8.739	8.739
		3	8.756	8.740	*8.737*	8.768	8.745	**8.732**	8.739	8.743	8.742
2	30	1	10.661	10.020	*9.971*	10.793	10.017	**9.952**	9.986	10.075	10.059
		2	11.400	10.626	*10.558*	11.619	10.642	**10.516**	10.580	10.686	10.666
		3	11.429	10.638	*10.564*	11.648	10.662	**10.517**	10.585	10.702	10.677
	100	1	8.730	8.725	8.725	8.734	8.726	**8.724**	8.726	8.727	8.726
		2	8.871	8.865	**8.864**	8.880	8.871	*8.867*	8.870	8.871	8.871
		3	8.872	8.866	**8.865**	8.880	8.871	*8.868*	8.871	8.871	8.871
3	30	1	14.657	9.815	**9.802**	19.269	9.822	9.926	*9.809*	10.011	9.838
		2	16.626	*10.420*	**10.401**	21.720	10.462	10.556	10.468	10.762	10.500
		3	16.633	*10.441*	**10.426**	21.811	10.480	10.571	10.489	10.792	10.533
	100	1	8.623	*8.602*	**8.600**	8.639	8.603	8.604	**8.600**	8.605	8.604
		2	8.764	8.739	**8.737**	8.799	8.745	*8.738*	**8.737**	8.744	8.743
		3	8.769	8.742	**8.739**	8.809	8.749	8.741	*8.740*	8.748	8.747
4	30	1	15.614	10.020	**9.977**	19.816	10.017	10.175	*9.990*	10.343	10.058
		2	17.434	10.629	**10.569**	22.081	10.665	10.832	*10.592*	11.015	10.691
		3	17.851	10.634	**10.568**	22.273	10.659	10.840	*10.588*	10.985	10.679
	100	1	8.747	8.728	*8.727*	8.772	8.728	**8.725**	8.728	8.733	8.730
		2	8.886	*8.863*	*8.863*	8.933	8.870	**8.861**	8.868	8.876	8.872
		3	8.894	8.870	*8.868*	8.939	8.876	**8.867**	8.874	8.882	8.879
5	30	1	17.831	11.947	**11.806**	20.200	11.927	11.819	*11.814*	13.114	12.093
		2	19.990	12.810	*12.577*	22.557	12.754	12.632	**12.495**	14.022	12.880
		3	19.960	12.763	*12.541*	22.541	12.708	12.596	**12.468**	14.011	12.858
	100	1	8.918	8.881	*8.875*	8.963	8.880	**8.873**	8.879	8.914	8.889
		2	9.078	9.037	*9.031*	9.143	9.041	**9.026**	9.037	9.082	9.051
		3	9.080	9.036	*9.030*	9.142	9.041	**9.026**	9.039	9.082	9.052
6	30	1	18.115	**12.146**	12.156	20.263	12.172	*12.151*	12.219	13.432	12.303
		2	20.530	*13.073*	**13.048**	22.878	13.148	13.083	13.099	14.572	13.289
		3	20.610	13.101	**13.078**	22.954	13.133	*13.082*	13.129	14.609	13.305
	100	1	8.970	8.922	*8.914*	9.015	8.922	**8.910**	8.921	8.967	8.934
		2	9.124	9.070	*9.062*	9.183	9.072	**9.054**	9.068	9.123	9.086
		3	9.127	9.076	*9.066*	9.186	9.077	**9.059**	9.073	9.126	9.091

Note: Bold and italic fonts indicate the smallest and second smallest prediction errors of the best models.

Tables 4 and 5 show the selection probability and PE, respectively. When $n = 30$, the principle best models were different from the true models in the cases 2, 4, 5, and 6, in which the principle best models were M_3, M_3, M_6, and M_7, respectively. On the other hand, when $n = 100$, the principle best model was different from the true model only in case 6, in which the principle best model was M_7. In the tables, bold and italic fonts indicate the highest and second highest probabilities of selecting the principle best model and the smallest and second smallest prediction errors of the best models. We see from the tables that, except for the TIC, the bias-corrected AICs resulted in improved performance

for variable selection, compared to the uncorrected AIC. This indicates that correcting the bias of the AIC is effective for improving the performance of the AIC as a model selector when the sample size is not large. Although, in theory, the TIC reduces the bias of the AIC, its performance as a model selector was inferior. This is because the TIC only minimally corrects the bias of the AIC. As stated earlier, the $\mathrm{AIC_J}$ and $\mathrm{CAIC_J}$ have the smallest biases. Nevertheless, their performance for variable selection was not the best. This leads us to the conclusion that it is not necessary to bring the bias close to 0 as much as possible, although bias correction is effective. The best performance in the sense of high selection probability and small PE was by the MAIC and $\mathrm{EIC_A}$. This is because the candidate model that minimizes the loss function is either the true model or an underspecified model, as described in the proof of Theorem 2.1. Hence, this result indicates that the bias correction in an underspecified model is important for improving the model-selecting performance of an information criterion. The performance of the $\mathrm{EIC_A}$ was slightly better than that of the MAIC; this is because the $\mathrm{EIC_A}$ reduces the influence of nonnormality more effectively than does the MAIC. However, when the sample size was small and the number of explanatory variables was large, i.e., cases 3 to 6, the performance of the $\mathrm{EIC_A}$ as a model selector was reduced. One reason for this is that the $\mathrm{EIC_A}$ is constructed by resampling the full model. When the sample size is small and the number of explanatory variables is large, we anticipate that the accuracy of resampling will be decreased due to an increase of variances of ML estimators in the full model. The performance of the CV criterion as a model selector was not bad even though it has a large bias. This is because the variable-selection method using the CV criterion is conscious of improving for a prediction of a validation sample. Although the performance was not bad, it was not as good as either the MAIC or $\mathrm{EIC_A}$.

In this subsection, we listed simulation results of the variable selections using nested models. We also conducted simulations using nonnested models. However, we omit the results because they were very similar to those for the nested models.

4.2. An Example Study

4.2.1. Target Characteristic

In the example study, we study performances of the variable-selection methods using nine information criteria by an estimator of the PE, which is derived as follows: We divide data to two samples, a calibration sample $(\boldsymbol{Y}_c, \boldsymbol{X}_c)$ with n_c and a validation sample $(\boldsymbol{Y}_v, \boldsymbol{X}_v)$ with n_v, randomly, and repeated such division

$N_r = 10,000$ times. In each repetition, we select the best model by minimizing each information criterion from a calibration sample $(\boldsymbol{Y}_c, \boldsymbol{X}_c)$, and record the selected best model. Let $\boldsymbol{X}_{c,\mathrm{best}}$ and $\boldsymbol{X}_{v,\mathrm{best}}$ be matrices of the selected best explanatory variables in $\boldsymbol{X}_c$ and $\boldsymbol{X}_v$, respectively. In order to assess an accuracy of prediction, we calculate as

$$\begin{aligned}\hat{\Delta} &= p\log 2\pi + \log|\hat{\boldsymbol{\Sigma}}_{c,\mathrm{best}}| \\ &\quad + \frac{1}{n_v}\,\mathrm{tr}\Big\{\big(\boldsymbol{Y}_v - \boldsymbol{X}_{v,\mathrm{best}}\,\hat{\boldsymbol{\Xi}}_{c,\mathrm{best}}\big)'\big(\boldsymbol{Y}_v - \boldsymbol{X}_{v,\mathrm{best}}\,\hat{\boldsymbol{\Xi}}_{c,\mathrm{best}}\big)\,\hat{\boldsymbol{\Sigma}}_{c,\mathrm{best}}^{-1}\Big\}\,.\end{aligned}$$

The average of $\hat{\Delta}$ across the N_r replications, $\widehat{\mathrm{PE}}$, is regarded as an estimate of the prediction error of the best model.

4.2.2. Used Real Data

We used data of 37 kindergarten students ($n = 37$) in a low-socioeconomic-status area, which was provided by Dr. William D. Rohwer of the University of California at Berkeley to examine how well performance on a set of paired-associate (PA) tasks can predict performance on some measures of aptitude and achievement (see, [26, p. 217]). The data gives eight variables; score on the Peabody Picture Vocabulary Test (PPVT); score on the Raven Progressive Matrices Test (RPMT); score on a Student Achievement Test (SAT); performance on a 'named' PA task (N); performance on a 'still' PA task (S); performance on a 'named still' PA task (NS); performance on a 'named action' PA task (NA); performance on a 'sentence still' PA task (SS). We used PPVT, RPMT and SAT as the response variables ($p = 3$) and N, S, NS, NA and SS as explanatory variables. The number of explanatory variables in the full model is $k_\omega = 6$, because we always add a constant term to a regression. We compared with all 32 ($= 2^5$) candidate models by values of nine criteria. When all the samples were used for variable selection, the model having NS, NA, SS was selected as the best model by TIC, and the model having NA was selected as the best model by eight criteria other than TIC. Since we have conducted the numerical examination in the case of $n = 30$, we divided data into 30 and 7, i.e., $n_c = 30$ and $n_v = 7$.

4.2.3. Results of Example Study

Table 6 shows the probability of selecting the model and $\widehat{\mathrm{PE}}$. In the table, "variables" shows used explanatory variables in the candidate model. The set of variables which is not listed in the table indicates that it was not chosen as the best model in every criterion. Superscript symbols * and ** denote the best models selected by each of criteria when the full data was used for variable selection.

Table 6: Results of real data.

Variables	Selection Probability (%)								
	AIC	CAIC	MAIC	TIC	EIC	EIC_A	CV	AIC_J	$CAIC_J$
N	0.21	0.50	0.51	0.11	0.42	0.39	0.46	0.43	0.48
NS	0.59	3.06	3.37	0.22	4.12	2.22	2.09	1.46	1.60
NA*	51.17	88.03	90.52	33.58	80.86	91.18	88.51	80.86	82.50
SS	0.75	2.44	2.65	0.15	1.70	2.94	2.03	1.43	1.51
N, NS	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
N, NA	0.77	0.04	0.01	3.04	0.08	0.05	0.04	1.03	0.20
N, SS	0.12	0.02	0.00	0.19	0.03	0.00	0.04	0.18	0.16
S, NA	4.05	0.64	0.30	2.11	1.24	0.73	0.84	1.69	1.55
NS, NA	14.04	1.36	0.57	7.36	4.94	0.66	0.57	1.32	1.60
NS, SS	13.48	3.87	2.07	12.83	6.45	1.81	5.28	10.24	9.78
NA, SS	0.16	0.01	0.00	0.06	0.02	0.01	0.02	0.02	0.02
N, S, NA	0.85	0.00	0.00	2.97	0.00	0.00	0.00	0.18	0.03
N, NS, NA	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00
N, NS, SS	0.23	0.00	0.00	0.42	0.00	0.00	0.00	0.07	0.02
N, NA, SS	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
S, NS, NA	5.66	0.03	0.00	10.58	0.13	0.01	0.12	0.85	0.42
S, NA, SS	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NS, NA, SS**	5.68	0.00	0.00	15.72	0.01	0.00	0.00	0.22	0.12
N, S, NS, NA	0.03	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.00
N, S, NS, SS	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
N, S, NA, SS	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N, NS, NA, SS	0.04	0.00	0.00	0.52	0.00	0.00	0.00	0.00	0.00
S, NS, NA, SS	2.05	0.00	0.00	8.37	0.00	0.00	0.00	0.01	0.00
N, S, NS, NA, SS	0.05	0.00	0.00	0.87	0.00	0.00	0.00	0.00	0.00
$\widehat{\mathrm{PE}}$	22.396	22.043	*22.006*	22.582	22.108	**21.991**	22.036	22.135	22.113

Note: The set of variables which is not listed in the table indicates that it was not chosen as the best model in every criterion.
** denotes the best model selected by TIC, and * denotes the best model selected by criteria other than TIC, which were calculated from the full data.
Bold and italic fonts indicate the smallest and second smallest estimates of prediction errors of the best model.

Bold and italic fonts indicate the smallest and second smallest estimates of prediction errors of the best model. From the table, we can find the same tendency as the simulation study, i.e., EIC_A and MAIC were high performance criteria in the sense of improving the prediction. Moreover, we also find that results of variable selections using AIC and TIC tended to have larger variances than those of the other criteria.

5. CONCLUSIONS AND DISCUSSION

In this paper, we studied a bias-correction effect in the AIC to variable-selection methods for normal MLRMs, which are based on a minimization of an information criterion, by numerical examinations. Since all the variable-selection methods considered in this paper asymptotically choose the same model as the best model, we conducted numerical examinations using small and moderate sample sizes. Our results are summarized as follows:

- Except for the TIC, the performances of the variable-selection methods using the bias-corrected AIC were better than that using the original AIC. This suggests that exact correction, bootstrapping, or cross-validation work better than the moment method for correcting the bias. It will be that correcting only the top term in an asymptotic expansion of the bias, as do AIC and TIC, is insufficient in an overspecified models.
- Theoretically, the bias of the $\mathrm{CAIC_J}$ becomes the smallest among all the criteria mentioned in this paper, but by numerical examination, we verified that the $\mathrm{CAIC_J}$ is not the best model selector. This indicates that the performance of a criterion is not necessarily improved even if the biases of the risk functions for overspecified models are reduced to as small as possible.
- The CAIC and MAIC perform well as model selectors, even though they have constant bias when the true distribution is not normal. The reason for this is that the correction for the bias caused by nonnormality cannot be estimated accurately when the sample size is small. Thus, if we try to estimate this bias when the sample size is small, it will reduce the accuracy of the estimation.
- Variable-selection methods using the MAIC or $\mathrm{EIC_A}$, which are obtained by correcting the constant bias of the AIC, always perform well. This result leads us to the conclusion that correcting the bias for an underspecified model has a positive effect on the selection of variables. One reason for this is that the model that minimizes the loss function is either the true model or an underspecified model. The $\mathrm{EIC_A}$ has the best performance as the model selector except for when the sample size is small and there are a large number of explanatory variables in the full model.

In conclusion, we recommend using the MAIC for a small number of samples and the $\mathrm{EIC_A}$ for a moderate number of samples. We note that when the number of samples is sufficiently large, it does not matter which criterion is used.

APPENDIX

A.1. Proof of Theorem 2.1

First, we show that the candidate model minimizing the risk function is either the true model or an underspecified model. Let $\boldsymbol{X}_1 = (\boldsymbol{X}, \boldsymbol{a})$ be an $n \times (k+1)$ matrix of explanatory variables in the model $M_1 \colon \boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}_1\boldsymbol{\Xi}_1, \boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_n)$, where $\boldsymbol{a}$ is an n-dimensional vector that is linearly independent from any combination of the columns of $\boldsymbol{X}$. Let $\hat{\boldsymbol{\Xi}}_1$ and $\hat{\boldsymbol{\Sigma}}_1$ denote the ML estimators of $\boldsymbol{\Xi}_1$ and $\boldsymbol{\Sigma}_1$, respectively. From the formula for the inverse matrix (see, e.g., [10, p. 424, cor. 18.2.10]), we have

$$\boldsymbol{P}_{\boldsymbol{X}_1} \;=\; \boldsymbol{P}_{\boldsymbol{X}} + \frac{1}{\boldsymbol{a}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{a}} (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{a}\,\boldsymbol{a}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}) \;=\; \boldsymbol{P}_{\boldsymbol{X}} + \boldsymbol{a}_{\mathrm{s}}\,\boldsymbol{a}_{\mathrm{s}}' \;,$$

where $\boldsymbol{a}_{\mathrm{s}} = (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{a} \,/\, \|(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{a}\|$. From the formulas for the determinant and the inverse matrix (see, e.g., [10, p. 416, cor. 18.1.3, and p. 424, thm. 18.2.8]), $|\hat{\boldsymbol{\Sigma}}_1|$ and $\hat{\boldsymbol{\Sigma}}_1^{-1}$ are rewritten as

$$|\hat{\boldsymbol{\Sigma}}_1| \;=\; |\hat{\boldsymbol{\Sigma}}|\,\big(1 - \boldsymbol{a}_{\mathrm{s}}'\boldsymbol{P}_{\boldsymbol{U}}\,\boldsymbol{a}_{\mathrm{s}}\big)\,, \tag{A.1}$$

$$\hat{\boldsymbol{\Sigma}}_1^{-1} \;=\; \hat{\boldsymbol{\Sigma}}^{-1} + \frac{n}{1 - \boldsymbol{a}_{\mathrm{s}}'\boldsymbol{P}_{\boldsymbol{U}}\,\boldsymbol{a}_{\mathrm{s}}}\,\boldsymbol{\Sigma}_*^{-1/2}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{a}_{\mathrm{s}}\,\boldsymbol{a}_{\mathrm{s}}'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{\Sigma}_*^{-1/2}, \tag{A.2}$$

where $\boldsymbol{U} = (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Y}\,\boldsymbol{\Sigma}_*^{-1/2}$. Since $\hat{\boldsymbol{\Sigma}}_1$ is positive definite and $\boldsymbol{P}_{\boldsymbol{U}}$ is positive semidefinite, we can see that $0 \le \boldsymbol{a}_{\mathrm{s}}'\boldsymbol{P}_{\boldsymbol{U}}\,\boldsymbol{a}_{\mathrm{s}} < 1$ with equality if and only if

$$\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{a} \;=\; \boldsymbol{0}_p \;, \tag{A.3}$$

because of

$$\boldsymbol{a}_{\mathrm{s}}'\boldsymbol{P}_{\boldsymbol{U}}\,\boldsymbol{a}_{\mathrm{s}} = 0 \;\;\Longleftrightarrow\;\; (\boldsymbol{U}'\boldsymbol{U})^{-1/2}\boldsymbol{U}'\boldsymbol{a}_{\mathrm{s}} = \boldsymbol{0}_p \;\;\Longleftrightarrow\;\; \boldsymbol{U}'\boldsymbol{a}_{\mathrm{s}} = \boldsymbol{0}_p \;.$$

The condition for equality means that a partial correlation between $\boldsymbol{Y}$ and $\boldsymbol{a}$ given $\boldsymbol{X}$ is exactly 0. Suppose that the model M is overspecified. Then, $\boldsymbol{U} = (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{\mathcal{E}}$ holds, where $\boldsymbol{\mathcal{E}}$ is given by (2.4). It should be kept in mind that the standardized $\hat{\boldsymbol{\Sigma}}$ is expressed as $\boldsymbol{S} = \boldsymbol{U}'\boldsymbol{U}/n$. Notice that when M is an overspecified model,

$$\begin{aligned} n\hat{\boldsymbol{\Sigma}}_1 \;&=\; \boldsymbol{\Sigma}_*^{1/2}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}_1})\,\boldsymbol{\mathcal{E}}\,\boldsymbol{\Sigma}_*^{1/2}\,, \\ (\boldsymbol{\Gamma}_* - \boldsymbol{X}_1\hat{\boldsymbol{\Xi}}_1)'(\boldsymbol{\Gamma}_* - \boldsymbol{X}_1\hat{\boldsymbol{\Xi}}_1) \;&=\; \boldsymbol{\Sigma}_*^{1/2}\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{\mathcal{E}}\,\boldsymbol{\Sigma}_*^{1/2}\,. \end{aligned}$$

Therefore, by using the above equations and (2.7), the loss function under M_1 can be simplified as

$$\begin{aligned} &\mathcal{L}(\hat{\boldsymbol{\Xi}}_1, \hat{\boldsymbol{\Sigma}}_1 \,|\, \boldsymbol{X}_1) \;= \\ &\qquad =\; np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}_1| + \mathrm{tr}\Big\{\hat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{\Sigma}_*^{1/2}(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{\mathcal{E}})\,\boldsymbol{\Sigma}_*^{1/2}\Big\} \qquad (A.4) \\ &\qquad =\; np(\log 2\pi - 1) + n\log|\hat{\boldsymbol{\Sigma}}_1| + \mathrm{tr}\Big\{\hat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{\Sigma}_*^{1/2}(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{\mathcal{E}})\,\boldsymbol{\Sigma}_*^{1/2}\Big\}\;. \end{aligned}$$

Substituting (A.1) and (A.2) into (A.4) yields

$$\mathcal{L}(\hat{\boldsymbol{\Xi}}_1, \hat{\boldsymbol{\Sigma}}_1 \,|\, \boldsymbol{X}_1) \;=\; \mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X}) + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \;,$$

where

$$\begin{aligned}
\mathcal{L}_1 \;&=\; n \left\{ \log(1 - \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}) + \frac{\boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}}{1 - \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}} \right\}, \\
\mathcal{L}_2 \;&=\; \frac{n}{1 - \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}} \, \boldsymbol{a}_{\rm s}' \boldsymbol{U} (\boldsymbol{U}'\boldsymbol{U})^{-1} \boldsymbol{\mathcal{E}}' \boldsymbol{P}_{\boldsymbol{X}} \boldsymbol{\mathcal{E}} (\boldsymbol{U}'\boldsymbol{U})^{-1} \boldsymbol{U}' \boldsymbol{a}_{\rm s} \;, \\
\mathcal{L}_3 \;&=\; \frac{n^2}{1 - \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}} \, \boldsymbol{a}_{\rm s}' \boldsymbol{U} (\boldsymbol{U}'\boldsymbol{U})^{-2} \boldsymbol{U}' \boldsymbol{a}_{\rm s} \;.
\end{aligned}$$

Notice that $\log(1-x) + x/(1-x) \geq 0$ when $x \in [0, 1)$ with equality if and only if $x = 0$. Hence, $\mathcal{L}_1 \geq 0$ holds with equality if and only if (A.3) holds. Moreover, we have $\mathcal{L}_2 \geq 0$ with equality if (A.3) because $\boldsymbol{P}_{\boldsymbol{X}}$ is positive semidefinite. These equations imply that

$$\mathcal{L}(\hat{\boldsymbol{\Xi}}_1, \hat{\boldsymbol{\Sigma}}_1 \,|\, \boldsymbol{X}_1) \;\geq\; \mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X}) + \mathcal{L}_3 \;, \tag{A.5}$$

with equality if and only if (A.3) holds. A singular value decomposition of $\boldsymbol{U}$ (see, e.g., [10, chap. 21.12]) implies that $\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-2}\boldsymbol{U}' = \boldsymbol{H}\boldsymbol{D}^{-1}\boldsymbol{H}'$, where $\boldsymbol{D}$ is a $p \times p$ diagonal matrix whose diagonal elements are eigenvalues of $\boldsymbol{U}'\boldsymbol{U}$, and $\boldsymbol{H}$ is an $n \times p$ matrix satisfying $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_p$ and $\boldsymbol{H}\boldsymbol{H}' = \boldsymbol{P}_{\boldsymbol{U}}$. Moreover, $\lambda_{\max}(\boldsymbol{A}) \leq \mathrm{tr}(\boldsymbol{A})$ holds for any positive semidefinite matrix $\boldsymbol{A}$, where $\lambda_{\max}(\boldsymbol{A})$ is maximum eigenvalue of $\boldsymbol{A}$. Using these results and the equation $(1 - \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s})^{-1} \geq 1$ yields

$$\begin{aligned}
\mathcal{L}_3 \;&\geq\; n^2 \boldsymbol{a}_{\rm s}' \boldsymbol{U} (\boldsymbol{U}'\boldsymbol{U})^{-2} \boldsymbol{U}' \boldsymbol{a}_{\rm s} \;=\; n^2 \boldsymbol{a}_{\rm s}' \boldsymbol{H} \boldsymbol{D}^{-1} \boldsymbol{H}' \boldsymbol{a}_{\rm s} \\
&\geq\; \frac{n^2 \boldsymbol{a}_{\rm s}' \boldsymbol{H} \boldsymbol{H}' \boldsymbol{a}_{\rm s}}{\lambda_{\max}(\boldsymbol{U}'\boldsymbol{U})} \;=\; \frac{n \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}}{\lambda_{\max}(\boldsymbol{S})} \;\geq\; \frac{n \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}}{\mathrm{tr}(\boldsymbol{S})} \;.
\end{aligned} \tag{A.6}$$

Let $\boldsymbol{U}_0 = (\boldsymbol{I}_n - \boldsymbol{J}_n)\boldsymbol{\mathcal{E}}$ and $\boldsymbol{S}_0 = \boldsymbol{U}_0'\boldsymbol{U}_0/n$, where $\boldsymbol{J}_n = \boldsymbol{1}_n \boldsymbol{1}_n'/n$. Since we assume that $\boldsymbol{X}$ always has $\boldsymbol{1}_n$ as a column vector, $(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})(\boldsymbol{I}_n - \boldsymbol{J}_n) = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}$. This implies that $\boldsymbol{a}_{\rm s}'\boldsymbol{U} = \boldsymbol{a}_{\rm s}'\boldsymbol{U}_0$. Moreover, since $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{J}_n$ is a symmetric idempotent matrix with $\mathrm{tr}(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{J}_n) = k - 1$, it is rewritten as $\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{J}_n = \boldsymbol{Q}\boldsymbol{Q}'$, where $\boldsymbol{Q}$ is an $n \times (k-1)$ matrix satisfying $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_{k-1}$. Hence, from the formula for the inverse matrix (see, e.g., [10, p. 424, thm. 18.2.8]), we have

$$\begin{aligned}
(\boldsymbol{U}'\boldsymbol{U})^{-1} \;&=\; \left\{ \boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{J}_n)\boldsymbol{\mathcal{E}} - \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{J}_n)\boldsymbol{\mathcal{E}} \right\}^{-1} \\
&=\; (\boldsymbol{U}_0'\boldsymbol{U}_0)^{-1} \left\{ \boldsymbol{U}_0'\boldsymbol{U}_0 + \boldsymbol{U}_0'\boldsymbol{Q} (\boldsymbol{I}_{k-1} - \boldsymbol{Q}'\boldsymbol{P}_{\boldsymbol{U}_0}\boldsymbol{Q})^{-1} \boldsymbol{Q}'\boldsymbol{U}_0 \right\} (\boldsymbol{U}_0'\boldsymbol{U}_0)^{-1} \;.
\end{aligned}$$

This implies that for any p-dimensional vector $\boldsymbol{b}$

$$\boldsymbol{b}'(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{b} \geq \boldsymbol{b}'(\boldsymbol{U}_0'\boldsymbol{U}_0)^{-1}\boldsymbol{b} \;. \tag{A.7}$$

Moreover, $\mathrm{tr}(\boldsymbol{S}) \leq \mathrm{tr}(\boldsymbol{S}_0)$ holds, because $n\boldsymbol{S}_0 = n\boldsymbol{S} + \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{J}_n)\boldsymbol{\mathcal{E}}$. Using these results and equation (A.6) yields

$$\begin{aligned}
\mathcal{L}_3 \;&\geq\; \frac{n \boldsymbol{a}_{\rm s}' \boldsymbol{P}_{\boldsymbol{U}} \boldsymbol{a}_{\rm s}}{\mathrm{tr}(\boldsymbol{S})} \;=\; \frac{n \boldsymbol{a}_{\rm s}' \boldsymbol{U}_0 (\boldsymbol{U}'\boldsymbol{U})^{-1} \boldsymbol{U}_0' \boldsymbol{a}_{\rm s}}{\mathrm{tr}(\boldsymbol{S})} \\
&\geq\; \frac{n \boldsymbol{a}_{\rm s}' \boldsymbol{U}_0 (\boldsymbol{U}_0'\boldsymbol{U}_0)^{-1} \boldsymbol{U}_0' \boldsymbol{a}_{\rm s}}{\mathrm{tr}(\boldsymbol{S}_0)} \;=\; n \boldsymbol{a}_{\rm s}' \boldsymbol{W}_0 \boldsymbol{a}_{\rm s} \;,
\end{aligned} \tag{A.8}$$

where $\boldsymbol{W}_0 = \boldsymbol{P}_{\boldsymbol{U}_0}/\operatorname{tr}(\boldsymbol{S}_0)$. Using (A.5), (A.6) and (A.8) yields

$$E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_1, \hat{\boldsymbol{\Sigma}}_1 \,|\, \boldsymbol{X}_1)\big] - E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X})\big] \;\geq\; n E_*\big[\boldsymbol{a}_{\mathrm{s}}' \boldsymbol{W}_0 \boldsymbol{a}_{\mathrm{s}}\big]\,, \tag{A.9}$$

Hence, in order to evaluate the right side of equation (A.9), we have to evaluate the expectation of $\boldsymbol{W}_0$. This expectation can be calculated in the same way as in the proof of Lemma 7 in [29]. Notice that

$$\frac{p}{\operatorname{tr}(\boldsymbol{S}_0)} = \operatorname{tr}(\boldsymbol{W}_0) = \sum_{a=1}^{n} w_{aa}\,, \quad 0 = \boldsymbol{1}_n' \boldsymbol{W}_0 \boldsymbol{1}_n = \sum_{a=1}^{n} w_{aa} + \sum_{a=1}^{n} \sum_{b \neq a}^{n} w_{aa}\,, \tag{A.10}$$

where w_{ab} is the (a,b)-th element of $\boldsymbol{W}_0$. Since $w_{ab} = (\boldsymbol{\varepsilon}_a - \bar{\boldsymbol{\varepsilon}})'(\boldsymbol{U}_0\boldsymbol{U}_0)^{-1}(\boldsymbol{\varepsilon}_b - \bar{\boldsymbol{\varepsilon}})/\operatorname{tr}(\boldsymbol{S}_0)$, where $\bar{\boldsymbol{\varepsilon}}$ is the sample mean of $\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_n$, i.e., $\bar{\boldsymbol{\varepsilon}} = n^{-1}\sum_{i=1}^{n} \boldsymbol{\varepsilon}_i$, we can see that the diagonal elements of $\boldsymbol{W}_0$ are identically distributed and the upper (or lower) off-diagonal elements of $\boldsymbol{W}_0$ are also identically distributed. These results and the equations in (A.10) imply that

$$n E_*[w_{aa}] = p\alpha\,, \quad n E_*[w_{aa}] + n(n-1) E_*[w_{ab}] = 0 \quad (a \neq b)\,,$$

where $\alpha = E_*[1/\operatorname{tr}(\boldsymbol{S}_0)]$. Thus, $E_*[\boldsymbol{W}_0] = p\alpha(\boldsymbol{I}_n - \boldsymbol{J}_n)/(n-1)$ is derived. From the Jensen's inequality, we have $\alpha \geq 1/E_*[\operatorname{tr}(\boldsymbol{S}_0)] = n\{(n-1)p\}^{-1}$. Consequently, it follows from these results and equation (A.9) that

$$E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_1, \hat{\boldsymbol{\Sigma}}_1 \,|\, \boldsymbol{X}_1)\big] - E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} \,|\, \boldsymbol{X})\big] \;\geq\; \left(\frac{n}{n-1}\right)^2 > 0\,. \tag{A.11}$$

This means that the risk function becomes large when a new explanatory variable is added to an overspecified model. Since the overspecified model that has the smallest number of explanatory variables is the true model, the candidate model that minimizes the risk function is either the true model or an underspecified model.

Next, we show that the candidate model that minimizes the risk function is the true model when $n \to \infty$. From (A.11), we can see that overspecified models except the true model do not minimize the risk function even when $n \to \infty$, because the right side of (A.11) converges to a positive value. Hence, we only have to show the proof when the candidate model is an underspecified model. Suppose that the model M is underspecified. Let $\boldsymbol{\Pi} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Gamma}_*$ and $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_*^{1/2}(\boldsymbol{I}_p + \boldsymbol{\Omega})\boldsymbol{\Sigma}_*^{1/2}$, where $\boldsymbol{\Omega}$ is a matrix of noncentrality parameter given by (2.5). By minimizing $\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})$ in (2.7), or equivalently solving the equations $\partial\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})/\partial\boldsymbol{\Xi} = \boldsymbol{O}_{k,p}$ and $\partial\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})/\partial\boldsymbol{\Sigma} = \boldsymbol{O}_{p,p}$, we can see that $(\boldsymbol{\Pi}, \boldsymbol{\Psi})$ makes $\mathcal{L}(\boldsymbol{\Xi}, \boldsymbol{\Sigma} \,|\, \boldsymbol{X})$ the smallest. This implies that

$$\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} | \boldsymbol{X}) \geq \mathcal{L}(\boldsymbol{\Pi}, \boldsymbol{\Psi} | \boldsymbol{X}) = np(\log 2\pi + 1) + n \log|\boldsymbol{\Sigma}_*| + n \log|\boldsymbol{I}_p + \boldsymbol{\Omega}|\,. \tag{A.12}$$

On the other hand, since the full model is overspecified and $(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega)$ makes the negative twofold log-likelihood function of the full model the smallest, it follows

from equation (A.4) that

$$\begin{aligned}\mathcal{L}(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega | \boldsymbol{X}_\omega) &= -2\ell(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega | \boldsymbol{Y}, \boldsymbol{X}_\omega) - np + \mathrm{tr}\big\{(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})\,\boldsymbol{S}_\omega^{-1}\big\} \\ &\leq -2\ell(\boldsymbol{\Pi}_\omega, \boldsymbol{\Sigma}_* | \boldsymbol{Y}, \boldsymbol{X}_\omega) - np + \mathrm{tr}\big\{(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})\,\boldsymbol{S}_\omega^{-1}\big\} \\ &= np(\log 2\pi - 1) + n\log|\boldsymbol{\Sigma}_*| \\ &\quad + \mathrm{tr}(\boldsymbol{\mathcal{E}}'\boldsymbol{\mathcal{E}}) + \mathrm{tr}\big\{(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})\,\boldsymbol{S}_\omega^{-1}\big\}\,,\end{aligned} \tag{A.13}$$

where $\boldsymbol{\Pi}_\omega = (\boldsymbol{X}_\omega'\boldsymbol{X}_\omega)^{-1}\boldsymbol{X}_\omega'\boldsymbol{\Gamma}_*$. Using the equations in (A.12) and (A.13) yields

$$\begin{aligned}&E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}\,|\,\boldsymbol{X})\big] - E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega\,|\,\boldsymbol{X}_\omega)\big] \geq \\ &\qquad \geq n\log|\boldsymbol{I}_p + \boldsymbol{\Omega}| + 2np \\ &\qquad\quad - E_*\big[\mathrm{tr}(\boldsymbol{\mathcal{E}}'\boldsymbol{\mathcal{E}})\big] - E_*\big[\mathrm{tr}\big\{(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})\,\boldsymbol{S}_\omega^{-1}\big\}\big] \\ &\qquad = n\log|\boldsymbol{I}_p + \boldsymbol{\Omega}| + np - E_*\big[\mathrm{tr}\big\{(n\boldsymbol{I}_p + \boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})\,\boldsymbol{S}_\omega^{-1}\big\}\big]\,.\end{aligned} \tag{A.14}$$

Hence, in order to evaluate the right side of equation (A.14), we have to evaluate $nE_*[\mathrm{tr}(\boldsymbol{S}_\omega^{-1})]$ and $E_*[\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}}\boldsymbol{S}_\omega^{-1}]$. In the same way as in the proof of Lemma 1 in [11], $\boldsymbol{S}_\omega^{-1}$ can be expressed as

$$\boldsymbol{S}_\omega^{-1} = \boldsymbol{I}_p - \frac{1}{\sqrt{n}}\,\boldsymbol{S}_\omega^{-1}\boldsymbol{V}_\omega\,,$$

where $\boldsymbol{V}_\omega = n^{1/2}(\boldsymbol{S}_\omega - \boldsymbol{I}_p)$. By using the Hölder's inequality, we have

$$\begin{aligned}E_*\big[\mathrm{tr}(\boldsymbol{S}_\omega^{-1})\big] &\leq p + \frac{1}{\sqrt{n}}\,E_*\big[\big|\mathrm{tr}(\boldsymbol{S}_\omega^{-1}\boldsymbol{V}_\omega)\big|\big] \\ &\leq p + \sqrt{\frac{1}{n}\,E_*\big[\mathrm{tr}(\boldsymbol{S}_\omega^{-2})\big]\,E_*\big[\mathrm{tr}(\boldsymbol{V}_\omega^2)\big]}\,,\end{aligned}$$

$$E_*\big[\mathrm{tr}(\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}}\boldsymbol{S}_\omega^{-1})\big] \leq \sqrt{E_*\big[\mathrm{tr}(\boldsymbol{S}_\omega^{-2})\big]\,E_*\big[\mathrm{tr}\big\{(\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})^2\big\}\big]}\,.$$

Let $h_{\omega,i}$ be h_i in the full model, where h_i is given by (3.3). It follows from the equation $0 \leq h_{\omega,i} \leq 1$ that

$$\sum_{i=1}^n h_{\omega,i}^2 \leq \sum_{i=1}^n h_{\omega,i} = n - k_\omega\,, \qquad \sum_{i=1}^n (1-h_{\omega,i})^2 \leq \sum_{i=1}^n (1-h_{\omega,i}) = k_\omega\,.$$

From Lemma 5 in [29], we can see that

$$\begin{aligned}E_*\big[\mathrm{tr}(\boldsymbol{V}_\omega^2)\big] &= \frac{1}{n}\,E_*\big[\mathrm{tr}\big(\big\{\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}_\omega})\boldsymbol{\mathcal{E}}\big\}^2\big)\big] - 2nE_*\big[\mathrm{tr}(\boldsymbol{S}_\omega)\big] + np \\ &= \frac{1}{n}\left\{\kappa_4^{(1)}\sum_{i=1}^n h_{\omega,i}^2 + p(p+1)\,(n-k_\omega) + p(n-k_\omega)^2\right\} - np + 2k_\omega p \\ &\leq \left(1 - \frac{k_\omega}{n}\right)\Big\{\big|\kappa_4^{(1)}\big| + p(p+1)\Big\} + \frac{k_\omega^2 p}{n}\,,\end{aligned}$$

$$\begin{aligned}E_*\big[\mathrm{tr}\big\{(\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})^2\big\}\big] &= \kappa_4^{(1)}\sum_{i=1}^n (1-h_{\omega,i})^2 + p(p+1)\,k_\omega + pk_\omega^2 \\ &\leq k_\omega\Big\{\big|\kappa_4^{(1)}\big| + p(p+1)\Big\} + pk_\omega^2\,,\end{aligned}$$

where $\kappa_4^{(1)}$ is the multivariate kurtosis given in (2.4). The above expectations indicate that $E_*[\mathrm{tr}(\boldsymbol{V}_\omega^2)] = O(1)$ and $E_*[\mathrm{tr}\{(\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}})^2\}] = O(1)$. Recall that we assume $E_*[\mathrm{tr}(\boldsymbol{S}_\omega^{-2})] = O(1)$. Hence, we derive $E_*[\mathrm{tr}(\boldsymbol{S}_\omega^{-1})] = p + O(n^{-1/2})$ and $E_*[\mathrm{tr}(\boldsymbol{\mathcal{E}}'\boldsymbol{P}_{\boldsymbol{X}_\omega}\boldsymbol{\mathcal{E}}\boldsymbol{S}_\omega^{-1})] = O(1)$. Substituting the obtained orders of expectations into (A.14) yields

$$\text{(A.15)} \qquad E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}\,|\,\boldsymbol{X})\big] - E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega\,|\,\boldsymbol{X}_\omega)\big] \;\geq\; n\log|\boldsymbol{I}_p + \boldsymbol{\Omega}| + O(n^{1/2})\,.$$

When the assumptions in Theorem 2.1 hold, $\lim_{n\to\infty}\boldsymbol{\Omega}$ exists, because $\boldsymbol{\Gamma}_*$ can be expressed $\boldsymbol{X}_*\boldsymbol{\Xi}_*$, and $\lim_{n\to\infty}\boldsymbol{X}_*'\boldsymbol{X}_*$, $\lim_{n\to\infty}\boldsymbol{X}_*'\boldsymbol{X}$ and $\lim_{n\to\infty}\boldsymbol{X}'\boldsymbol{X}$ exist and are positive definite, where $\boldsymbol{X}_*$ is Let $\boldsymbol{\Omega}_0$ be a limiting value of $\boldsymbol{\Omega}$. Then, from (A.15), the following equation is derived:

$$\liminf_{n\to\infty}\frac{1}{n}\Big\{E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}}\,|\,\boldsymbol{X})\big] - E_*\big[\mathcal{L}(\hat{\boldsymbol{\Xi}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega\,|\,\boldsymbol{X}_\omega)\big]\Big\} \;=\; \log|\boldsymbol{I}_p + \boldsymbol{\Omega}_0| \;>\; 0\,.$$

The above result and the fact that the risk function in the true model is smaller than those in all overspecified models indicate that the risk function in the true model is the smallest among all candidate models when $n \to \infty$. Consequently, Theorem 2.1 is proved.

A.2. Relationship between the best models selected by the AIC and CAIC

Let M_j $(j = 1, ..., m_M)$ be the j-th candidate model with an $n \times k_j$ matrix of explanatory variables $\boldsymbol{X}_j$, and let AIC_j and CAIC_j be the AIC and CAIC of the model M_j, respectively, where m_M is the number of candidate models. Without loss of generality, we assume that M_1 denotes the best model selected by minimizing the AIC. Let $\mathcal{J}$ be the set of indexes, which is defined by $\mathcal{J} = \{j \in \{2, ..., m_M\}\,|\; k_j \geq k_1\}$, and let $q(k)$ be a function given by $q(k) = (p+k+1)\{2pk + p(p+1)\}/(n-p-k-1)$. Since $q(k)$ is a monotonically increasing function with respect to k, $q(k_j) \geq q(k_1)$ holds when $j \in \mathcal{J}$. Moreover, $\mathrm{AIC}_j - \mathrm{AIC}_1 > 0$ holds for all $j \in \{2, ..., m_M\}$, because M_1 is the best model selected by the AIC. By using the above two results and the relation between the AIC and CAIC, the following inequality is derived:

$$\text{(A.16)} \qquad \mathrm{CAIC}_j - \mathrm{CAIC}_1 = \mathrm{AIC}_j - \mathrm{AIC}_1 + q(k_j) - q(k_1) > 0\,, \quad (j \in \mathcal{J})\,.$$

The result of (A.16) indicates that a model with more than k_1 explanatory variables will never be selected as the best model by the CAIC. Therefore, the number of explanatory variables in the best model selected by the CAIC is less than or equal to k_1.

A.3. Asymptotic equivalence of the EIC, adjusted EIC, and TIC for an overspecified model

From [9, 27][2], when $m \to \infty$, $\hat{B}_{\mathrm{EIC}}$ and $\hat{B}_{\mathrm{EIC_A}}$ can be expanded as

$$\begin{aligned}\hat{B}_{\mathrm{EIC}} &= 2pk + p(p+1) + \hat{\kappa}_4^{(1)} + O_p(n^{-1}) \,, \\ \hat{B}_{\mathrm{EIC_A}} &= 2(k+p+1)\,\mathrm{tr}(\boldsymbol{G}) - 3\,\mathrm{tr}(\boldsymbol{G}^2) - 2\,\mathrm{tr}(\boldsymbol{G})^2 + \frac{1}{n}\sum_{i=1}^{n} \hat{r}_{\omega,i}^4 + O_p(n^{-1}) \,,\end{aligned}$$

where $\hat{\kappa}_4^{(1)}$ is given by (3.3), $\boldsymbol{G} = \hat{\boldsymbol{\Sigma}}_\omega \hat{\boldsymbol{\Sigma}}^{-1}$ and $\hat{r}_{\omega,i}^2 = (\boldsymbol{y}_i - \hat{\boldsymbol{\Xi}}_\omega' \boldsymbol{x}_{\omega,i})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\Xi}}_\omega' \boldsymbol{x}_{\omega,i})$. When the model M is overspecified, $\boldsymbol{G} = \boldsymbol{I}_p + O_p(n^{-1/2})$, $\hat{\kappa}_4^{(1)} = \kappa_4^{(1)} + O_p(n^{-1/2})$, and $n^{-1}\sum_{i=1}^n \hat{r}_{\omega,i}^4 = p(p+2) + \kappa_4^{(1)} + O_p(n^{-1/2})$ hold, where $\kappa_4^{(1)}$ is given in (2.4). Hence, $\hat{B}_{\mathrm{EIC}}$ and $\hat{B}_{\mathrm{EIC_A}}$ can be rewritten as follows when the model M is overspecified:

$$\begin{aligned}\hat{B}_{\mathrm{EIC}} &= 2pk + p(p+1) + \kappa_4^{(1)} + O_p(n^{-1/2}) \,, \\ \hat{B}_{\mathrm{EIC_A}} &= 2pk + p(p+1) + \kappa_4^{(1)} + O_p(n^{-1/2}) \,.\end{aligned} \tag{A.17}$$

On the other hand, when the model M is overspecified, $\sum_{i=1}^n (1-h_i)(\hat{r}_i^2 - p) = O_p(n^{-1/2})$ holds because $\hat{r}_i^2 = \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i + O_p(n^{-1/2})$ and $1 - h_i = O(n^{-1})$ are satisfied. Then, $\hat{B}_{\mathrm{TIC}}$ can be expanded as

$$\hat{B}_{\mathrm{TIC}} = 2pk + p(p+1) + \kappa_4^{(1)} + O_p(n^{-1/2}) \,. \tag{A.18}$$

Comparing (A.17) with (A.18) yields $\mathrm{EIC} = \mathrm{TIC} + O_p(n^{-1/2})$ and $\mathrm{EIC_A} = \mathrm{TIC} + O_p(n^{-1/2})$, when the model M is overspecified and $m \to \infty$.

A.4. Asymptotic equivalence of the CV criterion and the TIC

From [27], the last term in (3.5) can be expanded as

$$\begin{aligned}\sum_{i=1}^{n} \frac{(n-1)\,\hat{r}_i^2}{h_i(nh_i - \hat{r}_i^2)} &= np + 2pk + p(p+1) + \hat{\kappa}_4^{(1)} \\ &\quad + 2\sum_{i=1}^{n} (1-h_i)\,(\hat{r}_i^2 - p) + O_p(n^{-1}) \,,\end{aligned} \tag{A.19}$$

where $\hat{r}_i^2$, $\hat{\kappa}_4^{(1)}$, and h_i are given by (3.3). Moreover, by applying the Taylor expansion to equation (3.5), we obtain

$$\sum_{i=1}^{n} \log\left(1 - \frac{\hat{r}_i^2}{nh_i}\right) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{r}_i^2}{h_i} + O_p(n^{-1}) \,. \tag{A.20}$$

[2] At the bottom of p. 240 of [9], $-\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^2)$ is missing in the equation $E[\hat{B}_A \,|\, \boldsymbol{Y}]$.

It follows from $h_i = 1 + O(n^{-1})$ and $\sum_{i=1}^{n} \hat{r}_i^2 = np$ that $n^{-1}\sum_{i=1}^{n} \hat{r}_i^2/h_i = n^{-1}\sum_{i=1}^{n} \hat{r}_i^2 + O_p(n^{-1}) = p + O_p(n^{-1})$. By combining the above result with (A.20), we obtain

$$\sum_{i=1}^{n} \log\left(1 - \frac{\hat{r}_i^2}{nh_i}\right) = -p + O_p(n^{-1}) \,. \tag{A.21}$$

On the other hand, $np\log\{2\pi n/(n-1)\} = np\log 2\pi + p + O(n^{-1})$ holds. Consequently, substituting this result and equations (A.19) and (A.21) into (3.5), and comparing the obtained equation with the definition of the TIC, yields $\mathrm{CV} = \mathrm{TIC} + O_p(n^{-1})$.

A.5. Asymptotic equivalence of the best models selected by the nine information criteria

Let IC be a general notation to indicate one of the nine information criteria considered in this paper. Notice that the bias-correction terms in the information criteria expect for the CV criterion are $O_p(1)$, and $\mathrm{CV} = \mathrm{TIC} + O_p(n^{-1})$ holds. Since $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}_* + \boldsymbol{\Sigma}_*^{1/2}\boldsymbol{\Omega}_0\boldsymbol{\Sigma}_*^{1/2}$ as $n \to \infty$, where $\boldsymbol{\Omega}_0$ is a limiting value of $\boldsymbol{\Omega}$ given by (2.5), we have

$$\begin{aligned} \frac{1}{n}\,\mathrm{IC} &\xrightarrow{p} p\log 2\pi + \log|\boldsymbol{\Sigma}_*| + \log|\boldsymbol{I}_p + \boldsymbol{\Omega}_0| + p \\ &\geq p\log 2\pi + \log|\boldsymbol{\Sigma}_*| + p \,, \quad \text{as } n \to \infty \,, \end{aligned} \tag{A.22}$$

with equality if and only if M is an overspecified model. The equation in (A.22) indicates that underspecified models are never selected as the best model when $n \to \infty$.

Let ICA denote an information criterion proposed under normality (i.e., the AIC, CAIC, or MAIC), and let ICT denote an information criterion proposed without a normality assumption (i.e., the TIC, EIC, $\mathrm{EIC_A}$, CV criterion, $\mathrm{AIC_J}$, or $\mathrm{CAIC_J}$). Notice that when M is an overspecified model, $\mathrm{ICA} = \mathrm{AIC} + o_p(1)$, $\mathrm{ICT} = \mathrm{TIC} + o_p(1)$ and $\hat{B}_{\mathrm{TIC}} \xrightarrow{p} 2pk + p(p+1) + \kappa_4^{(1)}$ as $n \to \infty$, where $\kappa_4^{(1)}$ is the multivariate kurtosis given in (2.4). Hence, when M is an overspecified model, we derive

$$\mathrm{ICT} = \mathrm{ICA} + \kappa_4^{(1)} + o_p(1) \,. \tag{A.23}$$

It should be emphasized that $\kappa_4^{(1)}$ does not depend on the candidate model considered, i.e., $\kappa_4^{(1)}$ is common in all overspecified models. Let M_1 and M_2 be two different overspecified models, and let ICA_j and ICT_j be information criteria for M_j $(j = 1, 2)$. From equation (A.23), we obtain

$$\mathrm{ICT}_1 - \mathrm{ICT}_2 = \mathrm{ICA}_1 - \mathrm{ICA}_2 + o_p(1) \,.$$

This equation indicates that the differences between two information criteria for the two different overspecified models are asymptotically equivalent. Consequently, all the information criteria choose the same model as the best one when $n \to \infty$.

ACKNOWLEDGMENTS

The authors would like to thank the editor, the associate editor and the two reviewers for valuable comments. Especially, the authors are grateful that one of reviewers pointed out many mistakes of our paper. Hirokazu Yanagihara's research was partially supported by the Ministry of Education, Science, Sports, and Culture, and a Grant-in-Aid for Challenging Exploratory Research, #25540012, 2013–2015; Shinpei Imori's research was supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists; Mariko Yamamura's research was supported by the project to support female researchers at Hiroshima University through system reforms.

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In "2nd. International Symposium on Information Theory" (B.N. Petrov and F. Csáki, Eds.), Akadémiai Kiadó, Budapest, 267–281.

[2] Akaike, H. (1974). A new look at the statistical model identification, *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, **AC-19**, 716–723.

[3] Bedrick, E.J. and Tsai, C.-L. (1994). Model selection for multivariate regression in small samples, *Biometrics*, **50**, 226–231.

[4] Burnham, K.P. and Anderson, D.R. (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach* (2nd. ed.), Springer-Verlag, New York.

[5] Davies, S.J.; Neath, A.A. and Cavanaugh, J.E. (2006). Estimation optimality of corrected AIC and modified C_p in linear regression model, *International Statistical Review*, **74**, 161–168.

[6] Fox, J. and Weisberg, S. (2011). *Multivariate linear models in R*. In "An Appendix to An R Companion to Applied Regression (2nd. ed.)", `http://socserv.mcmaster.ca/jfox/Books/Companion/appendix.html`, accessed April 17, 2014.

[7] Freedman, D.A. (1981). Bootstrapping regression models, *The Annals of Statistics*, **9**, 1218–1228.

[8] Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression, *Biometrika*, **84**, 707–716.

[9] Fujikoshi, Y.; Yanagihara, H. and Wakaki, H. (2005). Bias corrections of some criteria for selection multivariate linear regression models in a general case, *American Journal of Mathematical and Management Sciences*, **25**, 221–258.

[10] Harville, D.A. (1997). *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

[11] Hashiyama, Y.; Yanagihara, H. and Fujikoshi, Y. (2014). Jackknife bias correction of the AIC for selecting variables in canonical correlation analysis under model misspecification, *Linear Algebra and its Applications*, **455**, 82–106.

[12] Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.

[13] Ishiguro, M.; Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC, *Annals of the Institute of Statistical Mathematics*, **49**, 411–434.

[14] Konishi, S. (1999). *Statistical model evaluation and information criteria*. In "Multivariate Analysis, Design of Experiments, and Survey Sampling" (S. Ghosh, Ed.), Marcel Dekker, New York, 369–399.

[15] Konishi, S. and Kitagawa, G. (2003). Asymptotic theory for information criteria in model selection – functional approach, *Journal of Statistical Planning and Inference*, **114**, 45–61.

[16] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer Science+Business Media, LLC, New York.

[17] Kotz, S.; Kozubowski, T.J. and Podgórski, K. (2001). *The Laplace Distribution and Generalizations*, Birkhäuser Boston, Inc., Boston.

[18] Kullback, S. and Leibler, R.A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, **22**, 79–86.

[19] Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519–530.

[20] McQuarrie, A.D.R. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*, World Scientific Publishing Co., Inc., River Edge, NJ.

[21] Srivastava, M.S. (2002). *Methods of Multivariate Statistics*, John Wiley & Sons, New York.

[22] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society, Series* **B**, **36**, 111–147.

[23] Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *Journal of the Royal Statistical Society, Series* **B**, **39**, 44–47.

[24] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Communications in Statistics, Theory and Methods*, **A7**, 13–26.

[25] Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models, *Mathematical Science*, **153**, 12–18 (in Japanese).

[26] Timm, N.H. (2002). *Applied Multivariate Analysis*, Springer-Verlag, New York.

[27] Yanagihara, H. (2006). Corrected version of *AIC* for selecting multivariate normal linear regression models in a general nonnormal case, *Journal of Multivariate Analysis*, **97**, 1070–1089.

[28] Yanagihara, H. (2007). A family of estimators for multivariate kurtosis in a nonnormal linear regression model, *Journal of Multivariate Analysis*, **98**, 1–29.

[29] Yanagihara, H. (2015). Conditions for consistency of a log-likelihood-based information criterion in normal multivariate linear regression models under the violation of normality assumption, *Journal of the Japan Statistical Society*, **45**, 21–56.

[30] Yanagihara, H. and Fujisawa, H. (2012). Iterative bias correction of the cross-validation criterion, *Scandinavian Journal of Statistics*, **39**, 116–130.

[31] Yanagihara, H. and Ohmoto, C. (2005). On distribution of AIC in linear regression models, *Journal of Statistical Planning and Inference*, **133**, 417–433.

[32] Yanagihara, H.; Kamo, K. and Tonda, T. (2011). Second-order bias-corrected AIC in multivariate normal linear models under nonnormality, *The Canadian Journal of Statistics*, **39**, 126–146.

[33] Yanagihara, H.; Tonda, T. and Matsumoto, C. (2006). Bias correction of cross-validation criterion based on Kullback–Leibler information under a general condition, *Journal of Multivariate Analysis*, **97**, 1965–1975.

[34] Yanagihara, H.; Wakaki, H. and Fujikoshi, Y. (2015). A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large, *The Electronic Journal of Statistics*, **9**, 869–897.

[35] Yanagihara, H.; Yuan, K.-H.; Fujisawa, H. and Hayashi, K. (2013). A class of model selection criteria based on cross-validation method, *Hiroshima Mathematical Journal*, **43**, 149–177.

[36] Yoshimoto, A.; Yanagihara, H. and Ninomiya, Y. (2005). Finding factors affecting a forest stand growth through multivariate linear modeling, *Journal of Japanese Forestry Society*, **87**, 504–512 (in Japanese).

[37] Yuan, K.-H. and Bentler, P.M. (1997). *Generating multivariate distributions with specified marginal skewness and kurtosis.* In "SoftStat' 97 – Advances in Statistical Software 6 –" (W. Bandilla and F. Faulbaum, Eds.), Lucius & Lucius, Stuttgart, Germany, 385–391.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 333–354

A CHARACTERIZATION OF THE GENERALIZED BIRNBAUM–SAUNDERS DISTRIBUTION

Author: Emilia Athayde
– Centro de Matemática da Universidade do Minho,
Braga, Portugal
mefqa@math.uminho.pt

Received: October 2014 Revised: June 2015 Accepted: July 2015

Abstract:

- The Birnbaum–Saunders (BS) model is a fatigue life distribution related to the normal one with appealing properties. It is a non-negative transformation of a N(0,1) random variable Z, and its extended version, the so-called generalized Birnbaum–Saunders (GBS) model, obtained replacing Z by any other symmetric continuous random variable, shares the same properties. In this paper we characterize the GBS distribution according to one of its outstanding properties and take advantage of this characterization to derive a graphical procedure to assess whether an observed data set follows a GBS distribution. We further develop a goodness-of-fit test for the hypothesis that the data follow a BS distribution with unknown parameters and apply the results to real data sets.

Key-Words:

- *Birnbaum–Saunders distribution; goodness-of-fit; graphical tools; TTT plot.*

AMS Subject Classification:

- 62E10, 63E15, 62-09.

1. INTRODUCTION

Birnbaum and Saunders [10] proposed a fatigue life distribution representing time to failure of materials exposed to a cyclically repeated stress pattern. They obtained a nonnegative transformation of a standard normal random variable (RV) with a scale and a shape parameter known as the Birnbaum–Saunders (BS) distribution; see also Johnson *et al.* [25], pp. 651–663. The BS distribution is positively skewed allowing for different degrees of kurtosis and its hazard or failure rate (FR) has an inverse bathtub shape (see Kundu *et al.* [27] and Bebbington *et al.* [9]). The BS distribution is closed under scale transformations and under reciprocation.

Seshadri [39] and Saunders [38] considered classes of absolutely continuous non-negative RVs closed under reciprocation and ways of generating such classes. Common RVs satisfying the "reciprocal property", meaning that the RV and its own reciprocal are identically distributed, are the Fisher–Snedecor $F_{n,n}$ and the lognormal with appropriate mean, as well as the quotient of two independent and identically distributed (IID) non-negative (and unlimited to the right) RVs, discussed by Gumbel and Keeney [21]. Some other examples of distributions satisfying the reciprocal property can be found in Jones [26] and Vanegas *et al.* [42].

The generalized Birnbaum–Saunders distribution (GBS), introduced by Díaz-García and Leiva [16], is obtained replacing the normal generator in the BS distribution by any symmetric absolutely continuous RV. It is a highly flexible class of positively skewed distributions allowing for a wide range of kurtosis. The probability density functions include unimodal and bimodal cases and FRs can be monotone, inverse bathtub or have more than one change-point. Heavy tails are also allowed depending on the tails of the generating RV. The GBS distribution is also closed under scale transformations and under reciprocation. As a consequence, any GBS distributed RV, suitably scaled, satisfies the reciprocal property. See also Sanhueza *et al.* [36] for a discussion of its theory and applications, Balakrishnan *et al.* [6] for the case generated by scale-mixtures of normal distributions and Leiva *et al.* [30] for a family related to scale-mixture BS distributions.

Among several extensions of the BS distributions that have appeared in the literature, we point out the three-parameter model of power transformations of BS distributed RVs introduced by Owen [34], obtained by relaxing the assumption of independent crack extensions to a long memory process and related to sinh-spherical distributions (see Díaz-García *et al.* [17]) and the four-parameter extension based on the Johnson system (Athayde *et al.* [4]). The BS distribution also belongs to the family of cumulative damage distributions (see Leiva *et al.* [32]).

Truncated BS and shifted BS distributions have been considered as well; see Ahmed *et al.* [2] and Leiva *et al.* [28]. The case generated by non-symmetric RVs has also been addressed; see Ferreira *et al.* [20] and Leiva *et al.* [31]. However in this case the resulting RV does not satisfy the reciprocal property for any scale transformation.

Reparameterizations of the BS model have also appeared; see Leiva *et al.* [29], Santos-Neto *et al.* [37] and references therein.

Considering the problem of fitting a distribution to univariate data and assuming it comes from a nonnegative RV, we would like to assess whether a GBS distribution is a good candidate to model the data, and then find an appropriate way to test for goodness-of-fit (GOF). Notice that the BS and the GBS distributions are not in the location-scale (LS) family. For the case of parametric distributions with unknown parameters, GOF techniques have also been addressed in the literature for non LS models, including graphical techniques; see Barros *et al.* [8] and Castro-Kuriss *et al.* [12] for an overview of available tests and graphical tools to assess GOF in non LS distributions. These techniques can be applied to the case of a BS distribution or a GBS generated by a parameterized distribution, such as the Student or logistic ones, provided a proper estimation procedure and the inverse cumulative distribution function are available, but are not designed to test for the GBS class as a whole.

As mentioned before, any GBS distributed RV, suitably scaled, has the reciprocal property, and we prove that the converse is also true, i.e., any RV that upon a suitable change of scale is equally distributed to its own reciprocal admits a representation as a GBS distribution. This characterization was the starting point that led us to tackle the proposed problem. Namely, it enabled us to find an alternative estimator for the scale parameter, to consider an empirical graphical technique that requires no estimation of the scale parameter and to test whether the data come from a GBS distribution using symmetry tests about an unknown constant. In addition, we consider a test for the null hypothesis that the data come from a BS distribution with unknown parameters and carry out a study of its asymptotic behavior.

The remainder of this paper is organized as follows. In Section 2 we present some well known results about BS and GBS distributions. In Section 3 we establish a characterization of the GBS model related to the reciprocal property and analyze some of its consequences. In Section 4 we discuss the problem of finding out whether this model is suitable to fit a given data set and develop an asymptotic GOF test for the case of the BS distribution with unknown parameters. In Section 5 we apply the results to real data sets and finally in Section 6 we draw some conclusions.

2. BACKGROUND

The BS distribution is a transformation T of a standard normal RV given by

$$T = \beta \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1} \right)^2, \tag{2.1}$$

where $Z \sim \mathrm{N}(0,1)$, α ($\alpha > 0$) is a shape parameter and β ($\beta > 0$) is a scale parameter (and also the median), denoted here by $T \sim \mathrm{BS}(\alpha, \beta)$, with inverse transformation given by

$$Z = \frac{1}{\alpha} \left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right) \sim \mathrm{N}(0,1)\,. \tag{2.2}$$

As mentioned in the previous section, the distribution of T given by (2.1) is positively skewed allowing for different degrees of kurtosis (greater than 3) and its FR has an inverse bathtub shape. Among the properties of this distribution, we highlight that if $T \sim \mathrm{BS}(\alpha, \beta)$ then (i) $cT \sim \mathrm{BS}(\alpha, c\beta)$ with $c > 0$ and (ii) $T^{-1} \sim \mathrm{BS}(\alpha, \beta^{-1})$, i.e., the BS distribution is closed under scale transformations and under reciprocation. Thus, denoting $Y = T/\beta$, Y and $1/Y$ are identically distributed, i.e., Y has the reciprocal property (see Saunders [38]). Analogously we say that $T \sim \mathrm{BS}(\alpha, \beta)$, suitably scaled, satisfies the reciprocal property.

The GBS distribution, introduced by Díaz-García and Leiva [16], is obtained replacing Z in (2.1) by any symmetric absolutely continuous RV X, thus leading to

$$T = \beta \left(\frac{\alpha X}{2} + \sqrt{\left(\frac{\alpha X}{2}\right)^2 + 1} \right)^2, \tag{2.3}$$

where α ($\alpha > 0$) is a shape parameter and β ($\beta > 0$) is a scale parameter (and also the median). We say that T given by (2.3) is generated by X and denote it by $T \sim \mathrm{GBS}(\alpha, \beta, g_X)$, where $g_X(\cdot)$ is the probability density function (PDF) of X.

The cumulative distribution function (CDF) of $T \sim \mathrm{GBS}(\alpha, \beta, g_X)$ is given by $F_T(t) = G_X(\xi(t;\alpha,\beta))$, $t > 0$, where $G_X(\cdot)$ is the CDF of X and $\xi(t;\alpha,\beta) = \frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)$, $t > 0$, and the PDF is given by $f_T(t) = g_X(\xi(t;\alpha,\beta))\,\xi'(t;\alpha,\beta)$, $t > 0$, where $\xi'(t;\alpha,\beta) = \frac{t+\beta}{2\alpha\sqrt{\beta}}\, t^{-3/2}$, $t > 0$. The GBS distribution is implemented in the R software package `gbs` (`http://www.R-project.org`).

Once again letting $T \sim \text{GBS}(\alpha, \beta, g_X)$, then T, suitably scaled, satisfies the reciprocal property considering like before $Y = T/\beta$. Notice further that in formula (2.3) we may assume $\alpha = 1$ without loss of generality since letting $X_\alpha = \alpha X$, the distributions $\text{GBS}(\alpha, \beta, g_X)$ and $\text{GBS}(1, \beta, g_{X_\alpha})$ are clearly the same.

The main result in Seshadri [39] states that for absolutely continuous non-negative RVs the reciprocal property is equivalent with the logarithm of that RV being symmetric about zero. In other words, a RV Y satisfies the reciprocal property if and only if $W = \log(Y)$ is symmetric (i.e., W and $-W$ are equally distributed).

3. MAIN RESULTS

As mentioned before, the GBS distribution given by (2.3), where X is a symmetric absolutely continuous RV, is such that T/β and β/T are equally distributed. We prove that the converse is also true, and consequently that this remarkable property characterizes the class of GBS distributions.

Theorem 3.1. *Let T be a non-negative absolutely continuous random variable. Then $T \sim \text{GBS}(\alpha, \beta, g_X)$ if and only if T, suitably scaled, satisfies the reciprocal property.*

Proof: It suffices to prove that if for some β ($\beta > 0$), T/β and β/T are equally distributed RVs then $T \sim \text{GBS}(1, \beta, g_X)$, i.e., T can be written as $T = \beta\left(\frac{X}{2} + \sqrt{\left(\frac{X}{2}\right)^2 + 1}\right)^2$ for some symmetric RV X with PDF $g_X(\cdot)$, where the inverse transformation is $X = \sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}$. So let T be a nonnegative absolutely continuous RV with CDF $H_T(\cdot)$, such that T/β and β/T are equally distributed for some positive constant β. Let $\xi(t) = \xi(t; 1, \beta) = \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}$ for $t > 0$ and $F(\cdot)$ be the CDF of $X = \xi(T)$, given by $F(z) = H_T(\xi^{-1}(z))$, $z \in \mathbb{R}$, or equivalently $H_T(t) = F(\xi(t))$, $t > 0$. Now, denoting by $F_{\frac{T}{\beta}}(\cdot)$ and $F_{\frac{\beta}{T}}(\cdot)$ the CDFs of T/β and β/T respectively, we then have $F_{\frac{T}{\beta}}(x) = F_{\frac{\beta}{T}}(x) = P\left(\frac{\beta}{T} \le x\right) = P\left(\frac{T}{\beta} \ge \frac{1}{x}\right) = 1 - F_{\frac{T}{\beta}}(1/x)$, $x > 0$. Consequently $H_T(\beta x) = 1 - H_T\left(\frac{\beta}{x}\right)$ and thus $F(\xi(\beta x)) = 1 - F(\xi(\beta/x))$. From the fact that $\xi(\beta/x) = -\xi(\beta x)$, it follows that $F(\xi(\beta x)) = 1 - F(-\xi(\beta x))$, i.e., $F(z) = 1 - F(-z)$, $z \in \mathbb{R}$. This proves that X is a symmetric RV such that $T = \beta\left(\frac{X}{2} + \sqrt{\left(\frac{X}{2}\right)^2 + 1}\right)^2$, and therefore $T \sim \text{GBS}(1, \beta, g_X)$. □

Corollary 3.1. *Let T be a non-negative absolutely continuous random variable. Then $T \sim \text{GBS}(\alpha, \beta, g_X)$ if and only if $\log(T) - \log(\beta)$ is a symmetric* RV.

Remark 3.1. From the characterization in Theorem 3.1, it follows that any class of non-negative absolutely continuous RVs satisfying the reciprocal property belongs to the GBS class. In this case, its median is necessarily 1 ($\beta = 1$). Notice that the support of a GBS distribution can be a proper subset of $[0, +\infty)$. In fact, an example follows directly from Habibullah [23] who introduced a one-parameter family of RVs with support $[a, 1/a]$, where $0 < a < 1$, that satisfy the reciprocal property, and thus belongs to the GBS class. Its CDF is given by $H_a(x;\theta) = \frac{1}{3}\alpha_1(\log x)^3 + \alpha_2 \log x + \frac{1}{2}$, $a < x < \frac{1}{a}$, where $\alpha_1 = \frac{3(1-\theta)}{4(\log a)^3}$, $\alpha_2 = \frac{\theta-3}{4\log a}$ and $0 \leq \theta \leq 1$. Notice further that if we extend this family by adding a scale parameter β, we get a $\text{GBS}(1, \beta, g_X)$ distribution with CDF given by $H_{a,\beta}(x;\theta) = H_a(x/\beta;\theta)$.

Other results follow immediately from this characterization, as stated in the next three corollaries to Theorem 3.1. Recall that for a random sample $T_1, ..., T_n$ from a GBS distribution, the modified moment (MM) estimator of β is given by

$$\widetilde{\beta} = \sqrt{SR}\,, \tag{3.1}$$

where S and R are the sample arithmetic and harmonic mean, respectively, i.e., $S = \overline{T} = \frac{1}{n}(T_1 + \cdots + T_n)$ and $R^{-1} = \overline{T^{-1}} = \frac{1}{n}\left(\frac{1}{T_1} + \cdots + \frac{1}{T_n}\right)$. See Birnbaum & Saunders [11], Ng *et al.* [33] and Sanhueza *et al.* [36] for the case of BS and GBS distributions. Notice that $\widetilde{\beta}/\beta$ and $\beta/\widetilde{\beta}$ are identically distributed (see Saunders [38], Theorem 3.2).

Corollary 3.2. *Let T and U be two independent* GBS *distributed* RV*s and $a \neq 0$. Then T^a, TU and T/U are also* GBS *distributed* RV*s.*

Corollary 3.3. *Any non-negative* RV *that is written as a quotient of two* IID RV*s is* GBS *distributed.*

Corollary 3.4. *The* MM *estimator of β is* GBS *distributed.*

Notice that Corollary 3.2 states that the GBS class is closed under power transformations, as well as under products and quotients of independent RVs. As an example, any power of $T \sim \text{BS}(\alpha, \beta)$ belongs to the GBS class. Another example (see also Seshadri [39]), following from Corollary 3.3, is that the half-Cauchy distribution with PDF $f(t) = \frac{2}{\pi}\frac{1}{1+t^2}$, $t > 0$, belongs to the GBS class since it is obtained as a quotient of two IID half-normal RVs. Clearly, its logarithm is a symmetric RV, since its PDF is given by $g(x) = e^x f(e^x) = \frac{2}{\pi}\frac{1}{e^x + e^{-x}}$, $x \in \mathbb{R}$.

Remark 3.2. Another consequence of Corollary 3.2 is that any member of the three-parameter extended BS distribution in the sense of Owen [34], consisting of power transformations of BS RVs, given by

$$T = \beta\left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right)^{\sigma} \quad \text{or} \quad Z = \frac{1}{\alpha}\left(\left(\frac{T}{\beta}\right)^{1/\sigma} - \left(\frac{\beta}{T}\right)^{1/\sigma}\right),$$

where $Z \sim \mathrm{N}(0,1)$ and α, β and σ are non-negative parameters, also admits a representation as a $\mathrm{GBS}(1,\beta,g_W)$ for some symmetric absolutely continuous RV W that depends on α, σ and Z. In fact, we have explicitly

$$W = \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right)^{\sigma/2} - \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right)^{-\sigma/2}.$$

An analogous result holds for the three-parameter extension of GBS distributions referred in Sanhueza *et al.* [36], related to sinh-spherical laws (see Díaz-García *et al.* [17]), obtained replacing Z by a symmetric RV X.

Theorem 3.2. *For a random sample from the* $\mathrm{GBS}(\alpha,\beta,g_X)$ *distribution and assuming* $\mathrm{E}(X^4) < +\infty$, *the* MM *estimator of* β *is asymptotically* $\mathrm{BS}(n^{-1/2}\alpha\theta,\beta)$ *distributed, where* $\theta^2 = \frac{u_1+\frac{1}{4}\alpha^2 u_2}{1+\frac{1}{2}\alpha^2 u_1}$, *and* $u_i = \mathrm{E}(X^{2i})$, $i = 1, 2$.

Proof: The proof is analogous to the proof of Theorem 3.7 in Birnbaum & Saunders [11], replacing Z_i by X_i. In fact, notice that now $1 + \frac{\alpha^2}{2n}\sum X_i^2$ converges in probability to $1 + \frac{1}{2}\alpha^2 u_1$, where $u_i = E(X^{2i})$, and letting again $U_i = X_i(1 + \frac{1}{4}\alpha^2 X_i^2)^{1/2}$, then $\mathrm{var}(U_i) = u_1 + \frac{1}{4}\alpha^2 u_2$. This leads to the limiting distribution $\mathrm{BS}(n^{-1/2}\alpha\theta,\beta)$ with $\theta^2 = \frac{u_1+\frac{1}{4}\alpha^2 u_2}{1+\frac{1}{2}\alpha^2 u_1}$. □

Corollary 3.5. *For a random sample of the* $\mathrm{GBS}(\alpha,\beta,g_X)$ *distribution, the* MM *estimator of* β *is asymptotically* $\mathrm{N}(\beta, n^{-1/2}\alpha\theta\beta)$.

Remark 3.3. Theorem 3.2 states that the asymptotic GBS distribution for $\widetilde{\beta}$ mentioned in Corollary 3.4 is precisely a $\mathrm{BS}(n^{-1/2}\alpha\theta,\beta)$. It extends the result by Birnbaum & Saunders [11] stating that $\widetilde{\beta}$, in the BS case, is asymptotically distributed as a $\mathrm{BS}(\alpha\theta n^{-1/2},\beta)$, where $\theta^2 = \frac{4+3\alpha^2}{(2+\alpha^2)^2}$. Moreover, it is in agreement with the more general asymptotic bivariate normal distribution for the MM estimators of α and β (see Ng *et al.* [33] and Sanhueza *et al.* [36]), since a $\mathrm{BS}(\alpha,\beta)$ distribution is asymptotically $\mathrm{N}(\beta,\alpha\beta)$, as $\alpha \to 0$. In fact, this result follows immediately from the power series expansion of (2.1), namely $T = \beta\left(1 + \alpha Z + \frac{1}{2}\alpha^2 Z^2 + \frac{1}{8}\alpha^3 Z^3 - \frac{1}{128}\alpha^5 Z^5 + \cdots\right)$; see Engelhardt *et al.* [19].

4. GOODNESS-OF-FIT

When dealing with a univariate lifetime random sample, $t = (t_1, t_2, ..., t_n)$ from a RV T, a natural question to consider is whether a member of the GBS class is suitable to model these data. Let $Y = \log(T)$, and let t^{-1} and y denote the transformed samples $\left(\frac{1}{t_1}, \frac{1}{t_2}, ..., \frac{1}{t_n}\right)$ and $(\log(t_1), \log(t_2), ..., \log(t_n))$, respectively. Theorem 3.1 and Corollary 3.1 leads us to tackle this problem (i) testing for equal distributions of T/β and β/T with unknown β, or (ii) testing Y for symmetry about an unknown constant. Both these procedures rely on estimating β or $\log(\beta)$, and the same applies from an empirical point of view using a graphical approach such as a quantile-quantile plot (QQ-plot) for the two samples, $\beta^{-1}t$ and βt^{-1} . A graphical procedure known as the total time on test (TTT) plot can also be used and this plot requires no estimation of β.

To test for equal distributions for T/β and β/T, β may be estimated by minimizing some "distance" between the two samples, $\beta^{-1}t$ and βt^{-1}. Two possible distances are:

- The square of the difference between the sample means of $\beta^{-1}t$ and βt^{-1}. This leads to the usual MM estimator of β, given by (3.1) with $S = \overline{T}$ and $R^{-1} = \overline{T^{-1}}$ as before. This is not surprising since the MM estimator of β in the GBS(α, β, f_X) model does not depend on either f_X or α.
- A Kolmogorov–Smirnov (KS) type distance between the empirical CDF (ECDF) of the samples $T_1/\beta, ..., T_n/\beta$ and $\beta/T_1, ..., \beta/T_n$, given by

$$D_{KS} = \sup_x \left|F_1(x) - F_2(x)\right|,$$

  where $F_1(x)$ and $F_2(x)$ are the ECDFs of the two samples. Notice that these two samples are not independent.

To test $Y = \log(T)$ for symmetry about unknown location, we highlight two tests with asymptotically distribution-free test statistics, namely (i) a classical test based on the sample skewness coefficient b_1 (Gupta [22]) and (ii) the triples test (see Davis *et al.* [15] and Randles *et al.* [35]). In the first case, the test statistic $\frac{\sqrt{n}\ b_1}{\tau}$, where $b_1 = \frac{m_3}{m_2^{3/2}}$, $\tau = \frac{m_6 - 6m_2m_4 + 9m_2^3}{m_3^2}$ and m_i is the central moment of order i, $i \in \mathbb{N}$, is asymptotically N$(0, 1)$ under the null hypothesis of symmetry, provided $\mu_6 = \mathrm{E}(Y^6)$ exists. The second test is based on the difference D between the number of "right triples" and the number of "left triples" in the sample, where each triple (Y_i, Y_j, Y_k), $1 \leq i < j < k \leq n$, is defined as a "right triple" if the middle ordered observation in (Y_i, Y_j, Y_k) is closer to the smallest than to the largest of the three observations, and as a "left triple" if the middle ordered observation is closer to the largest than to the smallest of the three observations.

The test statistic, $V = D/\widehat{\sigma}$, where $\widehat{\sigma}$ is given in formula (3.78) in Hollander *et al.* [24], is asymptotically $\mathrm{N}(0,1)$ under the null hypothesis of symmetry. Notice further that these two tests are unsensitive to power transformations in T, as well as to scale changes.

4.1. A graphical procedure based on the TTT plot

As mentioned in Section 2, the FR is an important indicator in lifetime analysis. Some particular outstanding FR shapes include increasing (IFR), decreasing (DFR), bathtub (BT) and inverse bathtub (IBT) ones. For a RV T with finite expectation, it is possible to identify the shape of its FR by the scaled TTT curve (Barlow *et al.* [7]), given by

$$W_T(y) = \frac{\int_0^{F_T^{-1}(y)} \big[1 - F_T(t)\big]\, dt}{\int_0^{F_T^{-1}(1)} \big[1 - F_T(t)\big]\, dt}\ , \qquad 0 \le y \le 1\ . \tag{4.1}$$

This function can be empirically approximated by

$$W_n(k/n) = \frac{\sum_{i=1}^{k} T_{i:n} + [n-k]\, T_{k:n}}{\sum_{i=1}^{n} T_{i:n}}\ , \qquad k = 0, ..., n\ , \tag{4.2}$$

where $T_{1:n}, T_{2:n}, ..., T_{n:n}$ denote the order statistics associated to a random sample $T_1, T_2, ..., T_n$ (see Figure 1). Thus, the plot of $\big[k/n,\, W_n(k/n)\big]$, where the consecutive points are connected by straight lines, gives us information about the underlying FR (see Aarset [1]).

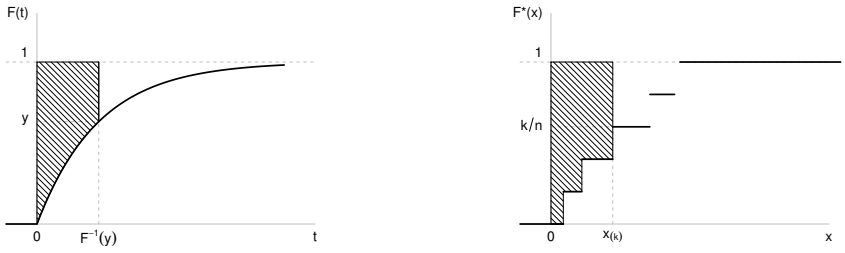

Figure 1: Shaded areas corresponding to $\int_0^{F_T^{-1}(y)} [1 - F_T(t)]\, dt$ (left) and $\frac{1}{n}\big[\sum_{i=1}^{k} T_{i:n} + [n-k]\, T_{k:n}\big]$ (right) in Equations 4.1 and 4.2.

The scaled TTT plot is a straight line in the case of the exponential distribution, a concave (convex) function in the case of an increasing (decreasing) FR, first concave (convex) and then convex (concave) in the case of an inverse bathtub (bathtub) FR, thus providing a useful tool in identifying the shape of the FR

(see Figure 2). Further, the scaled TTT is invariant under change of scale, and so in the case of a GBS distribution no estimation of β is required for the plot. The only drawback is that it requires T to have a finite expectation.

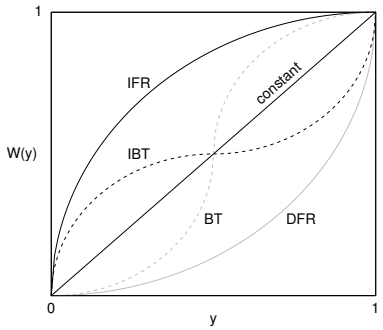

Figure 2: Scaled TTT plot for indicated shape of FR – bathtub (BT), decreasing (DFR), inverse bathtub (IBT), increasing (IFR).

Once again, it follows from Theorem (3.1) that the TTT curves are the same for T/β and β/T if and only if $T \sim \text{GBS}(1, \beta, g_X)$ for some symmetric X. Based on this result, we propose to assess the fit to the GBS distribution by comparing the empirical scaled TTTs of the samples t and t^{-1}. If the data do follow a GBS distribution, these two plots should look alike, regardless of β. We denote by D_{TTT} the maximum vertical distance between these two scaled TTT plots.

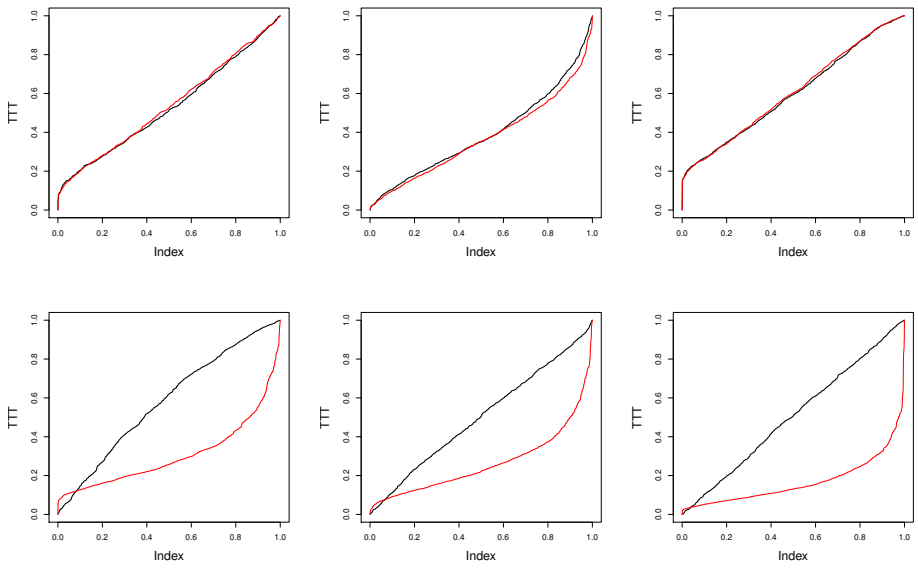

Figure 3: Scaled TTT plot for some GBS (top) and non-GBS (bottom) simulated samples and reciprocals, with $n = 10^3$.

See Figure 3 for plots of the empirical scaled TTT for simulated random samples ($n = 10^3$) for some GBS distributions (namely, a BS$(1, 1)$, a BS-t_3 generated by the Student t with 3 degrees of freedom, and the GBS distribution with CDF $H_a(.;\theta)$ for $a = \theta = 0.2$, mentioned in Remark 3.1) and non-GBS distributions (half-normal, half-Student t_3 and exponential). The behavior of the statistic D_{TTT} is under investigation.

4.2. Testing for the BS model

For the case of an absolutely continuous lifetime RV T, to test the null hypothesis H_0 that the CDF of T is $F(\cdot;\theta)$, based on a random sample $(t_1, t_2, ..., t_n)$, we consider the Cramér–von Mises (CM) statistic given by

$$W_n^2 \,=\, n \int_0^{+\infty} \big(F_n^*(t) - F(t;\theta)\big)^2 \, dF(t;\theta) \,, \tag{4.3}$$

where $F_n^*(\cdot)$ is the ECDF associated to the sample. This reduces to

$$W_n^2 \,=\, \frac{1}{12\,n} + \sum_{j=1}^{n} \left(\frac{2j-1}{2n} - F(t_{j:n};\theta) \right)^2 ,$$

where $t_{1:n}, t_{2:n}, ..., t_{n:n}$ denote the corresponding order statistics. If θ is known, W_n^2 is distribution-free, in the sense that its distribution depends only on n but not on the true $F(\cdot;\theta)$, since $F(T)$ is uniformly distributed in $[0, 1]$ under H_0. The asymptotic distributions were derived by Anderson and Darling [3].

As is well known, the ECDF statistics, such as W_n^2, for the case of unknown parameters usually depend on the CDF $F(\cdot;\theta)$ in H_0 as well as on n. However, in the case of a location-scale family, these statistics depend only on the family itself and n but not on the true values of the location and scale parameters, as long as an appropriate estimation method is provided (David and Johnson [14]). In some cases of a shape parameter, such as in the Gamma family, the dependence of the asymptotic and finite sample ECDF statistics on the shape parameter is slight, and tables of asymptotic percentage points were provided for different values of the parameter, to be used with the estimated values; see Stephens [41]. Another way to overcome this problem with shape parameters is to use the half-sample method introduced by Durbin [18]. This method uses a randomly chosen half of the original sample to compute the parameter estimates, say θ^*, by asymptotically efficient methods, such as maximum likelihood (ML). Then the ECDF statistics are computed with $F(\cdot;\theta^*)$ using the whole sample. The remarkable result is that asymptotically these ECDF statistics will behave like the ones for the case of known parameters. However, besides the dependence

of the test conclusion on the choice of the half-sample, a considerable loss in power has been reported, namely in the case of testing for a normal or exponential distribution (see Stephens, [40] and [41]).

For a random sample $T_1, T_2, ..., T_n$ from $T \sim \mathrm{BS}(\alpha, \beta)$, let $\theta = (\alpha, \beta)$ and $\widehat{\theta}$ and $\widetilde{\theta}$ denote respectively the ML and MM estimators of θ, and θ^* denote the ML estimator based on a randomly chosen half-sample. We shall carry out a study of the asymptotic distribution of W_n^2 in (4.3) for the case of unknown θ, using these three statistics. Thus let

$$C_n^2 \,=\, n \int_0^{+\infty} \big(F_n(t) - F(t;\widehat{\theta})\big)^2 \, dF(t;\widehat{\theta}) \tag{4.4}$$

instead of (4.3), as in Darling ([13]), or alternatively

$$C_n'^2 \,=\, n \int_0^{+\infty} \big(F_n(t) - F(t;\widetilde{\theta})\big)^2 \, dF(t;\widetilde{\theta}) \tag{4.5}$$

or

$$C_n^{*2} \,=\, n \int_0^{+\infty} \big(F_n(t) - F(t;\theta^*)\big)^2 \, dF(t;\theta^*) \,. \tag{4.6}$$

Remark 4.1. For the $\mathrm{BS}(\alpha, \beta)$ distribution, using the asymptotic distributions of $\widehat{\beta}$ and $\widetilde{\beta}$ (see Engelhardt [19] and Ng *et al.* [33], respectively), we have $\mathrm{var}(\widetilde{\beta}) \sim \mathrm{var}(\widehat{\beta})$ as $\alpha \to 0$, so the relative efficiency of these two estimators tends to 1 as α decreases. Moreover, quoting Birnbaum & Saunders [11] "under this condition [$\alpha < 1/2$], which we shall later empirically verify, $\widetilde{\beta}$ is virtually the ML estimator whose optimal properties are well known", we then expect to have similar asymptotic distributions (as $n \to \infty$) in (4.4) and (4.5) when using either $\widehat{\beta}$ or $\widetilde{\beta}$, at least for small values of α.

We have computed the asymptotic percentage points for C_n^2 for testing $H_0 \colon T \sim \mathrm{BS}(\alpha, \beta)$ with unknown parameters, based on 10^5 simulations, by the method described in Stephens [41], for significance levels 0.10, 0.05 and 0.01. This was achieved, for fixed α ($\alpha = 0.05, 0.1, 0.2, ..., 1.0$), by plotting the points obtained with simulated samples of size n ($n = 30, 40, ..., 120$) against $m = 1/n$ and extrapolating to $m = 0$. Then, the values obtained for each fixed significance level were plotted against α to extrapolate to $\alpha = 0$ by means of a polynomial fit (see Table 1). Notice that these values for $\alpha \to 0$ are almost exactly the same as for the case of a normal distribution with unknown parameters (see Table 4.7 in Stephens [41]), as expected, due to the asymptotic normality of the $\mathrm{BS}(\alpha, \beta)$ distribution as $\alpha \to 0$ (see Remark 3.3). We also report that, for the range of α values considered, the dependence of the percentage points on n ($n \geq 30$) is slight, being negligible as α decreases and as the significance level increases. Table 1 is to be used with estimated α from the data, as mentioned before. In general, the

well known data that have been fitted to a BS model have $\widehat{\alpha} < 1$, for example the lifetime data sets psi31, psi26 and psi21 in Birnbaum & Saunders [11] or the survival data set in Kundu *et al.* [27].

Table 1: Asymptotic upper-tail percentage points for C_n^2 for testing $H_0 : T \sim \mathrm{BS}(\alpha, \beta)$, both parameters unknown, based on 10^5 simulations.

α	Significance level		
	0.10	0.05	0.01
1.0	0.136	0.170	0.256
0.9	0.130	0.163	0.242
0.8	0.125	0.155	0.228
0.7	0.120	0.147	0.214
0.6	0.115	0.142	0.206
0.5	0.111	0.136	0.197
0.4	0.109	0.133	0.190
0.3	0.106	0.129	0.185
0.2	0.105	0.127	0.181
0.1	0.104	0.127	0.179
0.05	0.103	0.126	0.178
$\alpha \to 0$	0.103	0.126	0.179

We then repeated this procedure using MM instead of ML estimates of both parameters, and obtained the asymptotic percentage points for $C_n^{'2}$. The results, shown in Table 2, are similar to the former ones for small α values and the similarity is stronger as α decreases to 0, as expected (see Remark 4.1).

Table 2: Asymptotic upper-tail percentage points for $C_n^{'2}$ for testing $H_0 : T \sim \mathrm{BS}(\alpha, \beta)$, both parameters unknown, based on 10^5 simulations.

α	Significance level		
	0.10	0.05	0.01
1.0	0.126	0.156	0.230
0.9	0.123	0.151	0.222
0.8	0.120	0.148	0.215
0.7	0.117	0.144	0.208
0.6	0.113	0.139	0.200
0.5	0.111	0.136	0.194
0.4	0.109	0.133	0.190
0.3	0.106	0.130	0.186
0.2	0.105	0.127	0.181
0.1	0.103	0.126	0.179
0.05	0.103	0.126	0.179
$\alpha \to 0$	0.102	0.125	0.177

In the case of the GBS family, the percentage points for C_n^2 strongly depend on the true shape parameter α for a fixed generator X. However, if the parameters α and β are estimated by ML via the split-sample method (Durbin [18]; see also Stephens [41]), then similar results to the ones reported for testing normality and exponentiality based on CM statistic (see Stephens [40], Tables 1 and 2) were obtained. We illustrate this feature for the BS case with Table 3. This table shows the percentage points for C_n^{*2} for $\alpha = 0.1, 0.5, 1.0, 2.0, 3.0$ and $n = 20, 50, 100$, each one computed from 10^5 simulated samples at significance levels 0.10, 0.05 and 0.01, for unknown parameters estimated by the split-sample method.

Table 3: Upper-tail percentage points for C_n^{*2} for testing $H_0: T \sim \text{BS}(\alpha, \beta)$, both parameters unknown, and upper-tail asymptotic percentage points for W_n^2 for testing $H_0: T \sim \text{BS}(\alpha, \beta)$, both parameters known.

C_n^{*2}	n	α	Significance level		
			0.10	0.05	0.01
	20	0.1	0.373	0.490	0.755
		0.5	0.374	0.490	0.768
		1.0	0.376	0.495	0.776
		2.0	0.383	0.506	0.791
		3.0	0.372	0.491	0.778
	50	0.1	0.357	0.476	0.759
		0.5	0.355	0.472	0.745
		1.0	0.359	0.479	0.770
		2.0	0.361	0.479	0.770
		3.0	0.357	0.469	0.756
	100	0.1	0.353	0.471	0.755
		0.5	0.353	0.466	0.753
		1.0	0.354	0.473	0.761
		2.0	0.354	0.469	0.749
		3.0	0.351	0.462	0.751
W_n^2	∞		0.34730	0.46136	0.74346

The asymptotic percentage points for W_n^2 (Anderson and Darling [3]) are shown in the last row of the table. We observe that the dependence of upper-percentage points on α values is no longer strong, that it decreases as n increases and that the upper-percentage points are fairly close to the asymptotic ones for W_n^2.

We realize that a drawback of these methods is the dependence of the critical points on the unknown parameter and that there are other possible goodness-of-fit tests that can be useful in such cases; see Barros *et al.* [8] and Castro-Kuriss *et al.* [12].

5. SOME APPLICATIONS WITH DATA

In this section we analyze three well-known data sets from different areas and apply the procedures described in the previous sections to these data.

5.1. The data sets

The three data sets under analysis are (i) the survival times of 72 guinea pigs infected with tubercle bacilli in regimen 6.6 (corresponding to 4.0×10^6 bacillary units per 0.5ml), analyzed by Kundu *et al.* [27], denoted by `survpig`, (ii) the data set of lifetimes in cycles of aluminum coupons (maximum stress per cycle 31,000 psi) analyzed by Birnbaum & Saunders [11] and other authors (e.g., Ng *et al.* [33], Sanhueza *et al.* [36] and Balakrishnan *et al.* [6]), denoted by `psi31` and (iii) the data set of daily ozone concentrations collected in New York during May–September 1973, analyzed by Ferreira *et al.* [20], denoted by `ozone`. The sample dimensions are respectively $n = 72$, $n = 101$ and $n = 116$.

5.2. An introductory example

The estimation procedure based on the KS-type distance D_{KS} described in Section 4 is illustrated here by means of the data set `survpig`. For these data, all β values in the interval $[72.35, 72.92]$ minimize D_{KS}, so we took the center of this interval as its estimate, say $\beta_{KS} = 72.635$. This corresponds to a distance $D_{KS} = 6/72 = 0.0833$. See Figure 4.

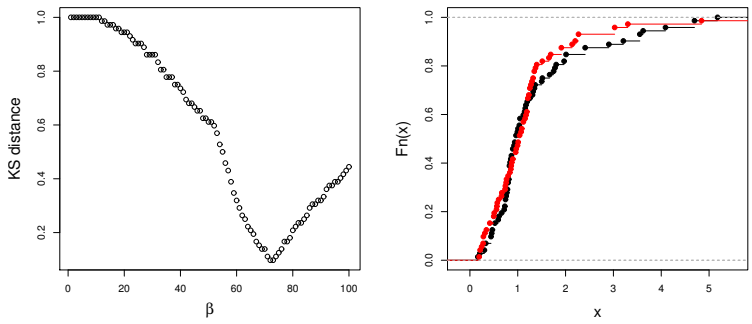

Figure 4: D_{KS} (KS-type distance) as a function of β (left) for `survpig` data and ECDF for the β-scaled sample and its reciprocal, with β estimated by minimizing D_{KS} (right).

5.3. Analyzing the data

For each of the samples, say $t = (t_1, t_2, ..., t_n)$, we applied the procedures described in the previous sections. The results are summarized in Table 4. See also the scaled TTT plots for t and t^{-1} (Figure 5) and the QQ-plots for $x = t/\widehat{\beta}$ and $y = \widehat{\beta}/t$ (Figure 6). Estimates $\widetilde{\beta}$ and β_{KS} (for β_{KS} we took the center of the interval of β values corresponding to a minimum distance D_{KS}, as explained before) were computed, as well as ML estimates of α and β for the parametric models BS(α, β) and BS-$t_\nu(\alpha, \beta)$, with ν estimated as in Azevedo *et al.* [5]. The CM-type statistics C_n^2, $C_n'^2$ and C_n^{*2} have also been computed and critical values for these statistics at significance level 5% are shown in parentheses. These values were obtained by interpolation, using Tables 1 and 2, in the first two cases, and from 10^5 simulated samples for each n ($n = 72, 101, 116$) and α ($\alpha = 0.76, 0.17, 0.98$), respectively. The classical test for symmetry about unknown location based on b_1 and the triples test were also applied to the transformed sample $y = log(t)$.

Table 4: Results for samples `survpig`, `psi31` and `ozone`.

		Data set		
		`survpig`	`psi31`	`ozone`
	n	72	101	116
	$\widetilde{\beta}$	77.4526	131.8193	28.4213
	β_{KS}	72.635	132.995	31.530
	D_{KS}	0.083	0.059	0.051
BS(α, β)	$\widehat{\beta}$	77.5348	131.8190	28.0234
	$\widehat{\alpha}$	0.7600	0.1704	0.9823
	$\widetilde{\alpha}$	0.7600	0.1704	0.9822
BS-$t_\nu(\alpha, \beta)$	ν	5	8	7
	$\widehat{\beta}$	75.5880	132.4297	30.9047
	$\widehat{\alpha}$	0.6085	0.1475	0.8074
	$\widetilde{\alpha}$	0.5887	0.1476	0.8301
	C_n^2	0.1874 (0.152)	0.0857 (0.127)	0.2071 (0.169)
	$C_n'^2$	0.1865 (0.145)	0.0857 (0.127)	0.1695 (0.156)
	C_n^{*2}	0.241 (0.470)	0.138 (0.472)	0.327 (0.471)
b_1 test	p-value	0.5886	0.3701	0.2258
triples test	p-value	0.170	0.691	0.388
	D_{TTT}	0.1068	0.0895	0.1989

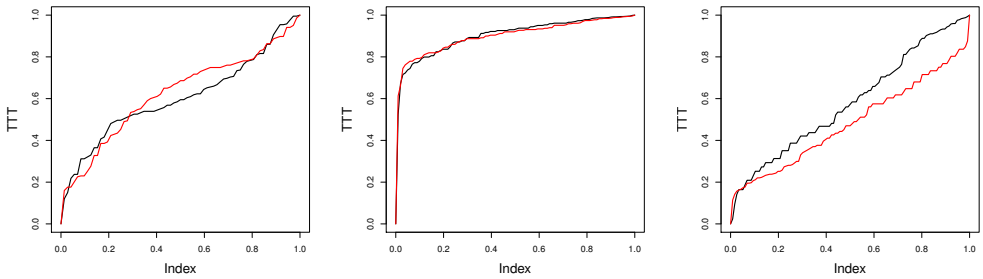

Figure 5: TTT for samples t and t^{-1} for `survpig` (left), `psi31` (center) and `ozone` (right).

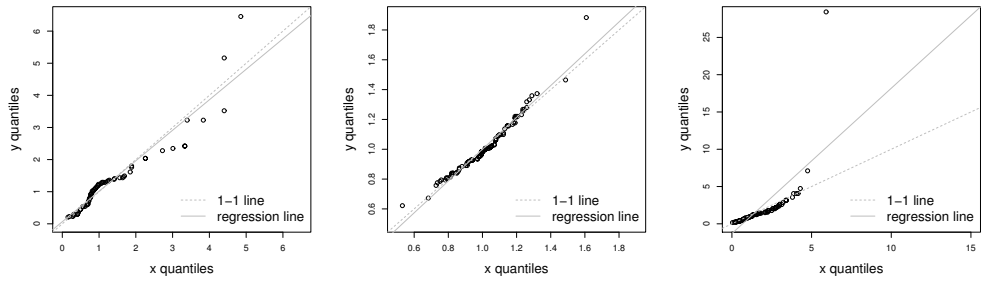

Figure 6: QQ-plots for $x = t/\widetilde{\beta}$ and $y = \widetilde{\beta}t^{-1}$, for `survpig` (left), `psi31` (center) and `ozone` (right).

The CM-type tests based on C_n^2 and $C_n'^2$ both reject the BS model for samples `survpig` and `ozone`, but not for `psi31`. The symmetry tests do not reject a GBS model for any of these samples.

Finally, the distance D_{TTT} has been computed for each sample. We also simulated the upper 5% percentage points for the distance D_{TTT} in the BS(α, β) and BS-$t_\nu(\alpha, \beta)$ models for each n (72, 101 and 116, respectively) with $\alpha = \widehat{\alpha}$ and $\beta = \widehat{\beta}$ in each case ($\nu = 5$, 8 and 7, respectively) with 10^4 simulations (see Table 5). This rules out these two particular models for `ozone`. The graphical analysis (Figures 5 and 6) also indicates that a GBS model seems reasonable for `survpig`, excellent for `psi31` and not adequate for `ozone`.

Table 5: Simulated upper-tail percentage points (at significance level 5%) for D_{TTT} assuming $T \sim$ BS$(\widehat{\alpha}, \widehat{\beta})$ or $T \sim$ BS-$t_\nu(\widehat{\alpha}, \widehat{\beta})$ ($\widehat{\alpha}$ and $\widehat{\beta}$ estimated from the three samples).

	sample		
	`survpig`	`psi31`	`ozone`
BS	0.156	0.122	0.157
BS-t_ν	0.218	0.197	0.184
	($\nu = 5$)	($\nu = 8$)	($\nu = 7$)

On the other side, as the classical symmetry b_1 test does not reject a GBS model for `ozone`, we have carried out a brief simulation study on the power of this test against several alternatives, for $n = 116$, including the extreme value Birnbaum–Saunders model generated by the Gumbel distribution for minima, denoted by EVBS$^*(\alpha, \beta, 0)$. This model was proposed by Ferreira *et al.* [20] as the best among several other models, including the BS one. The power of the test, based on 10^5 simulations, was estimated as 0.762 supposing the true model is EVBS$^*(\widehat{\alpha}, \widehat{\beta}, 0)$. If the true α in this model lies in the interval [0.6,1.0], the power decreases from 0.811 to 0.659, and can be as low as 0.044 for $\alpha = \exp(1)$. Finally, the simulated 5% upper percentage point for D_{TTT} with this model, 0.3142, also sustains the EVBS* fit since the observed value for `ozone` is much lower (see Table 4).

6. CONCLUDING REMARKS

In this paper we derived a characterization of the GBS class related to the reciprocal property and analyzed some of its consequences. We discussed some graphical procedures to assess the fit of the GBS model to observed data, we tabulated the asymptotic percentage points for a test of the null hypothesis that the data come from a BS distribution with unknown parameters, and finally we applied the results to three well-known data sets. The case of tests for other GBS distributions, such as the ones generated by the Student t_ν or logistic distributions, is under investigation.

ACKNOWLEDGMENTS

The author is indebted to the Associate Editor and to the referees for valuable suggestions that helped to improve the paper. The research of the author was partially supported by the Portuguese Funds from the "Fundação para a Ciência e a Tecnologia", through the Project PEstOE/MAT/UI0013/2014.

REFERENCES

[1] Aarset, M.V. (1987). How to identify a bathtub hazard rate, *IEEE Trans. Rel.*, **36**, 106–108.

[2] Ahmed, S.A.; Castro-Kuriss, C.; Flores, E.; Leiva, V. and Sanhueza, A. (2010). A truncated version of the Birnbaum–Saunders distribution with an application in financial risk, *Pak. J. Stat.*, **26**, 293–311.

[3] Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Ann. Math. Stat.*, **23**, 193–212.

[4] Athayde, E.; Azevedo, C.; Leiva, V. and Sanhueza, A. (2012). About Birnbaum–Saunders distributions based on the Johnson system, *Comm. Stat. Theor. Meth.*, **41**, 2061–2079.

[5] Azevedo, C.; Leiva, V.; Athayde, E. and Balakrishnan, N. (2012). Shape and change point analyses of the Birnbaum–Saunders-t hazard rate and associated estimation, *Comp. Stat. Data Anal.*, **56**, 3887–3897.

[6] Balakrishnan, N.; Leiva, V.; Sanhueza, A. and Vilca, F. (2009). Estimation in the Birnbaum–Saunders distribution based on scale mixtures of normals and the EM-algorithm, *Stat. Oper. Res. Trans.*, **33**, 171–192.

[7] Barlow, B. and Bremner, B. (1972). *Statistical Inference Under Order Restrictions*, John Wiley & Sons.

[8] Barros, M.; Leiva, V.; Ospina, R. and Tsuyuguchi, A. (2014). Goodness-of-fit tests for the Birnbaum–Saunders distribution with censored reliability data, *IEEE Trans. Reliab.*, **63**, 543–554.

[9] Bebbington, M.; Lai, C.-D. and Zitikis, R. (2008). A proof of the shape of the Birnbaum–Saunders hazard rate function, *Math. Scientist*, **33**, 49–56.

[10] Birnbaum, Z.W. and Saunders, S.C. (1969a). A new family of life distributions, *J. Appl. Prob.*, **6**, 319–327.

[11] Birnbaum, Z.W. and Saunders, S.C. (1969b). Estimation for a family of life distributions with applications to fatigue, *J. Appl. Prob.*, **6**, 328–347.

[12] Castro-Kuriss, C.; Leiva, V. and Athayde, E. (2014). Graphical tools to assess goodness-of-fit in non-location-scale distributions, *Colomb. J. Stat.*, **37**, 341–365.

[13] Darling, D.A. (1955). The Cramér–Smirnov test in the parametric case, *Ann. Mathem. Statist.*, **26**, 1–20.

[14] David, F.N. and Johnson, N.L. (1948). The probability integral transformation when parameters are estimated from the sample, *Biometrika*, **35**, 182–192.

[15] Davis, C.E. and Quade, D. (1978). U-statistics for skewness or symmetry, *Comm. Stat. Theor. Meth.*, **7**, 413–418.

[16] Díaz-García, J.A. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions, *J. Stat. Plan. Infer.*, **128**, 445–457.

[17] Díaz-García, J.A. and Domínguez-Molina, J.R. (2006). Some generalisations of Birnbaum–Saunders and sinh-normal distributions, *Intern. Math. Forum*, **1**, 1709–1727.

[18] Durbin, J. (1961). Some methods of constructing exact tests, *Biometrika*, **48**, 41–55.

[19] Engelhardt, M.; Bain, L.J. and Wright, F.T. (1981). Inferences in the parameters of the Birnbaum–Saunders fatigue life distribution based on maximum likelihood estimation, *Technometrics*, **23**, 251–255.

[20] Ferreira, M.; Gomes, M.I. and Leiva, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution, *Revstat*, **10**, 181–210.

[21] Gumbel, E.J. and Keeney, R.D. (1950). The extremal quotient, *Ann. Math. Stat.*, **21**, 523–538.

[22] Gupta, M.K. (1967). An asymptotically nonparametric test of symmetry, *Ann. Math. Stat.*, **38**, 849–866.

[23] Habibullah, S. (2009). *On a class of distributions closed under inversion*, PhD thesis, National College of Business Administration & Economics, Lahore, Pakistan.

[24] Hollander, M. and Wolfe, D.A. (1999). *Nonparametric Statistical Methods*, John Wiley & Sons, NY.

[25] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol. 2, John Wiley & Sons, NY.

[26] Jones, M.C. (2008). On reciprocal symmetry, *J. Stat. Plan. Infer.*, **138**, 3039–3043.

[27] Kundu, D.; Kannan, N. and Balakrishnan, N. (2008). On the hazard function of Birnbaum–Saunders distribution and associated inference, *Comp. Stat. Data Anal.*, **52**, 2692–2702.

[28] Leiva, V.; Athayde, E.; Azevedo, C. and Marchant, C. (2011). Modeling wind energy flux by a Birnbaum–Saunders distribution with an unknown shift parameter, *J. Appl. Stat.*, **38**, 2819–2838.

[29] Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A. and Barros, M. (2014). Birnbaum–Saunders statistical modelling: a new approach, *Stat. Modell.*, **14**, 21–48.

[30] Leiva, V.; Saulo, H.; Leao, J. and Marchant, C. (2014). A family of autoregressive conditional duration models applied to financial data, *Comp. Stat. Data Anal.*, **14**, 21–48.

[31] Leiva, V.; Ferreira, M.; Gomes, M.I. and Lillo, C. (2016). Extreme value Birnbaum–Saunders regression models applied to environmental data, *Stoch. Environ. Res. Risk Assess.*, **30**, 1045–1058.

[32] Leiva, V.; Tejo, M.; Guiraud, P.; Schmachtenberg, O.; Orio, P. and Marmolejo-Ramos, F. (2015). Modeling neural activity with cumulative damage distributions, *Biol. Cybern.*, **109**, 421–433.

[33] Ng, H.K.; Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum–Saunders distribution, *Comp. Stat. Data Anal.*, **43**, 283–298.

[34] Owen, W.J. (2006). A new three-parameter extension to the Birnbaum–Saunders distribution, *IEEE Trans. Reliab.*, **55**, 457–479.

[35] Randles, R.H.; Fligner, M.A.; Policello, G.E. and Wolfe, D.A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry, *J. Amer. Stat. Assoc.*, **75**, 168–172.

[36] Sanhueza, A.; Leiva, V. and Balakrishnan, N. (2008). The generalized Birnbaum–Saunders distribution and its theory, methodology, and application, *Comm. Stat. Theor. Meth.*, **37**, 645–670.

[37] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). On a reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications, *Revstat*, **12**, 247–272.

[38] Saunders, S.C. (1974). A family of random variables closed under reciprocation, *J. Amer. Stat. Soc.*, **69**, 533–539.

[39] Seshadri, V. (1964). On random variables that have the same distribution as their reciprocals, *Can. Math. Bull.*, **8**, 819–824.

[40] Stephens, M.A. (1978). On the half-sample method for goodness-of-fit, *J. Roy. Stat. Soc. B*, **40**, 64–70.

[41] Stephens, M.A. (1986). *Tests based on EDF statistics.* In "Goodness of Fit Techniques" (R.B. D'Agostino and M.A. Stephens, Eds.), Marcel Dekker, New York, 97–193.

[42] Vanegas, L.H. and Paula, G.A. (2016). Log-symmetric distributions: statistical properties and parameter estimation, *Braz. J. Prob. Stat.*, **30**, 196–220.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 355–372

TWO-SAMPLE GRADUAL CHANGE ANALYSIS

Authors: Zdeněk Hlávka
– Department of Statistics, Faculty of Mathematics and Physics, Charles University, Czech Republic
hlavka@karlin.mff.cuni.cz

Marie Hušková
– Department of Statistics, Faculty of Mathematics and Physics, Charles University, Czech Republic
huskova@karlin.mff.cuni.cz

Received: May 2015 Revised: July 2015 Accepted: August 2015

Abstract:

- We propose two-sample gradual change analysis motivated by gender differences observed in a real data set containing jumping speeds of 432 girls and 364 boys aged 6 to 19 years. Looking at this data set from the point of view of change-point analysis is more natural and it leads to more precise estimators than application of standard two-sample t-test in each age group. Apart of establishing the asymptotic distribution of the proposed two-sample change-point estimator, we also investigate its small sample properties in a simulation study.

Key-Words:

- *change point; gradual change; multiple comparison; two-sample test; wild bootstrap.*

AMS Subject Classification:

- 62F10, 62F25, 62F40, 62F03.

1. INTRODUCTION

In Table 5 and Figure 3, we present summary statistics of jumping speeds observed in a sample of 432 girls and 364 boys between 6 and 19 years measured by Leonardo Mechanograph Ground Reaction Force Plate (Šumník *et al.*, 2013). In this data set, one is naturally interested in investigating the location of the unknown *change point*: looking at the p-values of two-sample t-tests calculated for each of the thirteen age categories, it seems that jumping speeds for boys and girls are about the same from 6 to 10 years and boys' jumping speeds are clearly higher from 13 years on.

Unfortunately, applying the two-sample t-test thirteen times cannot be recommended without multiple testing corrections. Therefore, Table 5 contains also the p-values adjusted for multiple comparisons using Bonferroni and Benjamini–Hochberg (BH) method. The conclusions based on these two multiple comparison methods are similar although the Bonferroni method controls the family-wise error rate while the Benjamini–Hochberg method (Benjamini and Hochberg, 1995) controls the false discovery rate. It is interesting that also other standard multiple comparisons methods (Holm, 1979; Hommel, 1988; Hochberg, 1988; Benjamini and Yekutieli, 2001) implemented in the function `p.adjust()` in R (R Core Team, 2015) detect statistically significant differences at the same age category (13 years and above) while statistically significant differences are not detected for the two most "suspicious" un-adjusted p-values (0.061 and 0.047 for 11 and 12 years, respectively).

In Section 2, we study this two-sample testing problem from the point of view of change-point analysis using a simple model of gradual change (Hušková, 1999) so that instead of many independent two-sample t-tests we only estimate a single change-point. In Sections 3 and 4, we investigate the asymptotic properties of the proposed estimators under various assumptions (motivated by the application to the jumping speeds data set) and we show that the wild bootstrap provides both confidence intervals and p-values controlling the overall significance level.

Section 5 contains a small simulation study to check the behavior for finite sample situations. The jumping speeds data set is analyzed in Section 6 and we will see that the change-point approach detects statistically significant differences earlier (i.e., for younger children) than the two-sample t-tests. A short summary is given in Section 7.

2. PROCEDURES

We assume that our observations fall into two distinct subgroups that are further split into n distinct ordered categories and that the n_{ji} observations in the j-th subgroup and i-th category are summarized by their sample mean $\overline{Y}_{ji}$ and sample variance $\widehat{\sigma}^2_{ji}$, $j \in \{1,2\}$, $i = 1, ..., n$. Under additional assumptions one could naturally apply n independent two-sample t-tests in order to compare the two subgroups within each category and use some of the multiple test procedures as discussed above.

However, we propose another approach based on ideas of the change point analysis. Particularly, motivated by the above data set on the jumping speeds, we introduce a simple *two sample model with gradual changes*:

(A1) Observations Y_{jik} $(j = 1, 2;\ k = 1, ..., n_{ji})$ are obtained at time i $(i = 1, ..., n)$.

(A2) All observations are independent.

(A3) $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = \mu + \delta((i - k_0)/n)_+$ $(i = 1, ..., n)$, where μ, δ are unknown parameters and $k_0 = n\,\theta_0$ for some $\theta_0 \in (0, 1)$.

(A4) $\mathrm{Var}(Y_{jik}) = \sigma^2_{ji} > 0$ $(j = 1, 2;\ i = 1, ..., n;\ k = 1, ..., n_{ji})$.

We use the notation $\overline{Y}_{ji} = \sum_{k=1}^{n_{ji}} Y_{jik}/n_{ji}$, $a_+ = \max(a, 0)$ with k_0 denoting the unknown location of the change point, μ the unknown expectation of difference before the change, and δ_n the slope (speed) of the gradual change after k_0. Notice that, generally, variances of the single observations need not be the same.

Assumptions (A1)–(A4) are motivated by the application in Section 6: particularly, in this case, Assumption (A2) is satisfied since we observe only one measurement per subject. In other applications, Assumptions (A2) and (A3) may require some modifications to cover panel (longitudinal) data or time series. Also the trend after the change may not necessarily be linear; more generally, it can be some nondecreasing function strictly increasing after the change point.

We propose to estimate the unknown parameters by the least squares method. In the following, we deal separately with the homoscedastic case (Section 3) and the heteroscedastic case (Section 4).

3. HOMOSCEDASTIC CASE

Here we deal with a *two sample homoscedastic model with gradual changes* assuming additionally:

(A4*) $\mathrm{Var}(\overline{Y}_{1i} - \overline{Y}_{2i}) = \sigma^2/m \ \ (i = 1, ..., n)$, where $\sigma^2 > 0$ is an unknown parameter and m can depend on n.

One-sample homoscedastic models with various gradual changes were studied by a number of authors, e.g., Hinkley (1971); Feder (1975); Shaban (1980); Jarušková (1998); Hušková (1999); Hušková and Steinebach (2000, 2002). They constructed procedures for testing the null hypothesis *no change* versus the alternative *there is a change*, derived the least squares estimators, and studied its limit behavior for $n \to \infty$. We use the same method for our problem.

The least squares estimators $\widehat{\mu}$, $\widehat{\delta}$, $\widehat{k}_\mu$ are defined as minimizers of the sum of squares $\sum_{i=1}^n \left\{\overline{Y}_{1i} - \overline{Y}_{2i} - a - d((i-k)/n)_+\right\}^2$ with respect to a, d, k. Denoting $x_{ik} = ((i-k)/n)_+$ and $\bar{x}_k = \sum_{i=1}^n x_{ik}/n$, direct calculations give:

$$
\begin{aligned}
\widehat{k}_\mu &= \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n (x_{ik} - \overline{x}_k)\,(\overline{Y}_{1i} - \overline{Y}_{2i})\right\}^2}{\sum_{i=1}^n (x_{ik} - \overline{x}_k)^2} \right], \\
\widehat{\delta}_\mu &= \frac{\sum_{i=1}^n (x_{i\widehat{k}} - \overline{x}_{\widehat{k}})\,(\overline{Y}_{1i} - \overline{Y}_{2i})}{\sum_{i=1}^n (x_{i\widehat{k}} - \overline{x}_{\widehat{k}})^2}\,, \\
\widehat{\mu} &= \frac{1}{n}\sum_{i=1}^n (\overline{Y}_{1i} - \overline{Y}_{2i}) - \widehat{\delta}_\mu \overline{x}_{\widehat{k}}\,.
\end{aligned}
\tag{3.1}
$$

Assuming additionally that $\mu = 0$, the least squares estimators are:

$$
\begin{aligned}
\widehat{k}_0 &= \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})\right\}^2}{\sum_{i=1}^n x_{ik}^2} \right], \\
\widehat{\delta}_0 &= \frac{\sum_{i=1}^n x_{i\widehat{k}}(\overline{Y}_{1i} - \overline{Y}_{2i})}{\sum_{i=1}^n x_{i\widehat{k}}^2}\,.
\end{aligned}
\tag{3.2}
$$

Unfortunately, there are no explicit expressions for $\widehat{k}_\mu$ and $\widehat{k}_0$ and these estimators have to be found as a solution of an optimization problem. The properties of these estimators can be studied either through asymptotics (if n is large enough) or through a simulation study. We start with asymptotics and simulations are presented in Section 5.

Following the proofs in Jarušková (1998) and Hušková (1998, 1999) we get that in our homosecastic setup ((A1)–(A3) and (A4*)) for $n \to \infty$

$$(nm)^{1/2}\,\frac{\delta}{\sigma}\,\left\{\frac{\theta_0(1-\theta_0)}{1+3\theta_0}\right\}^{1/2}\frac{\widehat{k}_\mu - k_0}{n} \;\xrightarrow{\mathcal{D}}\; N(0,1)$$

and

$$(nm)^{1/2}\,\frac{(1-\theta_0)^{3/2}}{\sigma}\left(\frac{1+3\theta_0}{12}\right)^{1/2}(\widehat{\delta}_\mu - \delta) \;\xrightarrow{\mathcal{D}}\; N(0,1)\,,$$

where $N(0,1)$ denotes the standard normal distribution and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Both assertions hold true both for m fixed and $m \to \infty$ together with $n \to \infty$. The limit properties remain true even if δ depends on n and tends to 0 for $n \to \infty$ but no faster than $n^{-1/2}\log\log n$. The above results also imply consistency:

$$(nm)^{1/2}\,\delta(\widehat{k}_\mu - k_0)/n = O_P(1) \qquad \text{and} \qquad (nm)^{1/2}\,(\widehat{\delta}_\mu - \delta) = O_P(1)\,.$$

Quite analogously when $\mu = 0$ we get that the limit distributions of

$$(nm)^{1/2}\,\frac{\delta}{\sigma}\left(\frac{1-\theta_0}{4}\right)^{1/2}\frac{\widehat{k}_0 - k_0}{n} \qquad \text{and} \qquad (nm)^{1/2}\,\frac{(1-\theta_0)^{3/2}}{3^{1/2}\,\sigma}\,(\widehat{\delta}_0 - \delta)$$

are standard normal $N(0,1)$.

In Figure 1, the asymptotic distributions of $\widehat{k}_\mu$ and $\widehat{k}_0$ for nine distinct values of k_0 are compared to histograms obtained by 1000 Monte Carlo simulations. Very good approximations via the limit distribution are evident for $k_0 \le 15$ and, as expected, they are slightly worse but still acceptable for $k_0 > 15$. The assumption $\mu = 0$ visibly improves the precision of $\widehat{k}_0$ for smaller values of k_0.

In case the trend in the means is not linear after the change (as in (A3)) but nondecreasing with strict monotonicity after the change point, the proposed change point estimators may be biased (Hušková and Steinebach, 2002).

Under the assumption of homoscedasticity, we may combine the estimators $\widehat{\sigma}^2_{ji}$ observed in each category into the standard pooled estimator $\widehat{\sigma}^2_{\text{pooled}}$ of the variance σ^2. The assumption that $\mathrm{Var}(\overline{Y}_{1i} - \overline{Y}_{2i})$ does not depend on i is rather restrictive and a more general case of variances will be studied in the next section.

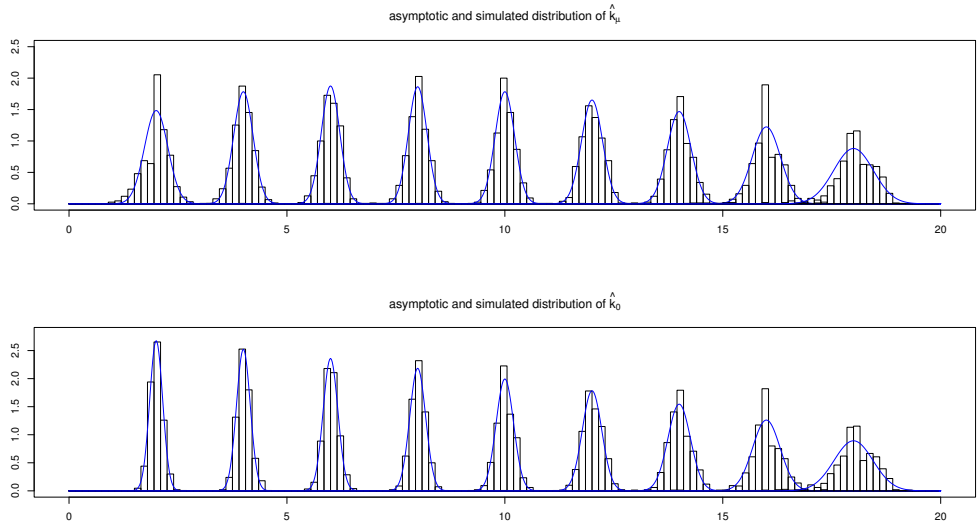

Figure 1: Densities of asymptotic distributions and histograms of 1000 simulated values of $\widehat{k}_\mu$ (upper plot) and $\widehat{k}_0$ (lower plot) in the homoscedastic case for $n = 20$, $\sigma^2 = 1$, $m = 20$, $\mu = 0$, $\delta = 1$, and $k_0 \in \{2, 4, ..., 18\}$.

4. HETEROSCEDASTIC CASE

4.1. Change-point estimators

Let us assume (A1)–(A4) with $\mu = 0$. We may still use the estimators introduced in the previous section: they still have the same limit distributions but with different standardizations. Denoting by $\tau_i^2 = \mathrm{Var}(\overline{Y}_{1i} - \overline{Y}_{2i}) = \sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i}$, we define the estimator $\widehat{k}_0(\tau^2)$ taking also the heteroscedasticity into account:

$$\widehat{k}_0(\tau^2) \;=\; \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})/\tau_i^2\right\}^2}{\sum_{i=1}^n x_{ik}^2/\tau_i^2} \right].$$

In practice, the unknown true variances τ_i^2 are replaced by $\widehat{\tau}_i^2 = \widehat{\sigma}_{1i}^2/m_{1i} + \widehat{\sigma}_{2i}^2/m_{2i}$ leading to the change-point estimator:

$$\widehat{k}_0(\widehat{\tau}^2) \;=\; \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})/\widehat{\tau}_i^2\right\}^2}{\sum_{i=1}^n x_{ik}^2/\widehat{\tau}_i^2} \right] = \arg\max_{k\in(1,n)} T_{2,\widehat{\tau}^2}(k)\,. \tag{4.1}$$

Concerning properties of these estimators under Assumptions (A1)–(A4) with the additional assumption

$$\tau_-^2/n \le \tau_i^2 \le \tau_+^2/n\,, \qquad i = 1, ..., n\,, \tag{4.2}$$

for some $0 < \tau_-^2 \le \tau_+^2 < \infty$, the asymptotic distribution remains normal with zero mean but the asymptotic variance has a more complicated structure and we do not give here explicit formulas. This can again be proved along the lines of the proofs in Hušková (1999). To get approximation for the distribution of the estimator $\widehat{k}_0(\widehat{\tau}^2)$, a proper version of the wild bootstrap provides a reasonable approximation. The algorithm is described below.

4.2. Bootstrap approximation for the distribution of $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$

For simplicity, we will write $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$. Under Assumptions (A1)–(A4) and (4.2), the observed sample mean differences $D_i = \overline{Y}_{1i} - \overline{Y}_{2i}$ have zero mean and standard deviation $\tau_i = (\sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i})^{1/2}$. The distribution of $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$ can be approximated by the wild bootstrap (Shao and Tu, 1995):

Algorithm 1. Bootstrap algorithm

Estimate parameters δ and k_0.
Calculate *fitted values* $\widehat{D}_i = \widehat{\delta}_0((i-\widehat{k})/n)_+$ $(i = 1, ..., n)$.
For $b = 1$ to $b = B$
  Generate $D_i^* = \widehat{D}_i + \widehat{\tau}_i \varepsilon_i^*$ $(i = 1, ..., n)$, where $\varepsilon_i^* \sim N(0,1)$ are independent.
  Calculate the change-point estimator $\widehat{k}_b^*$ from the bootstrap sample $D_1^*, ..., D_n^*$.
Calculate the empirical quantile q_α^* from $\widehat{k}_1^* - \widehat{k}, ..., \widehat{k}_B^* - \widehat{k}$ for prechosen $\alpha \in (0,1)$.

The empirical bootstrap quantiles q_α^* provide approximations for the true quantiles q_α of $\widehat{k} - k_0$, particularly it can be proved:

$$1 - \alpha \;=\; P(\widehat{k} - k_0 > q_\alpha) \;=\; P(k_0 < \widehat{k} - q_\alpha) \;=\; P(k_0 < \widehat{k} - q_\alpha^*) + o_P(1)$$

and, therefore, $\widehat{k} - q_\alpha^*$ can be used as an upper bound of an asymptotic one-sided $(1-\alpha)$ confidence interval for k_0.

Remark 4.1. As a complementary problem, we can test hypotheses concerning the change-point location, i.e., the null hypothesis $H_0 \colon k_0 \ge k_1$ against $H_1 \colon k_0 < k_1$ for some given k_1. Denoting by K a random variable with the same distribution as $\widehat{k} - k_0$ and defining the p-value as $P(K < \widehat{k} - k_1)$ (we reject H_0 for small values of $\widehat{k}$), we obtain that, for large B, $\sum_{b=1}^{B} I(\widehat{k}_b^* - \widehat{k} < \widehat{k} - k_1)/B$ is a reasonable approximation of the p-value.

Remark 4.2. The null hypothesis of no-change can easily be tested by bootstrapping the test statistic $T_{2,\widehat{\tau}^2}(k)$ because, under the null hypothesis of no-change, we can easily generate the bootstrap samples $D_i^* = \widehat{\tau}_i \varepsilon_i^*$.

5. SIMULATIONS

5.1. Setup of the simulation study

In this section, we investigate small sample properties of the proposed asymptotic tests and confidence intervals in various setups. We consider the model of gradual change (A3). In each step of the simulation we proceed as follows:

Algorithm 2. Simulation study

Set n and the change-point $\theta_0 = k_0/n$.
Set variances σ_{1i}^2 and σ_{2i}^2 and numbers of observations n_{1i} and n_{2i} $(i = 1, ..., n)$.
Calculate variances $\tau_i^2 = \sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i}$ $(i = 1, ..., n)$.
For $s = 1$ to $s = S$
  For $i = 1$ to $i = n$
    Generate $D_i = \overline{Y}_{1i} - \overline{Y}_{2i}$ from $N((i - k_0)_+, \tau_i^2)$.
    Generate $\widehat{\tau}_i^2$ from $\sigma_{1i}^2\chi_{n_{1i}-1}^2/\{n_{1i}(n_{1i} - 1)\} + \sigma_{2i}^2\chi_{n_{2i}-1}^2/\{n_{2i}(n_{2i} - 1)\}$.
  Calculate $\widehat{k}_0^{(s)}$ applying one of the change-point estimators $\widehat{k}$ defined by (3.1), (3.2), or (4.1).
  Calculate the 95% confidence interval for k_0 using Algorithm 1.
Calculate the bias and the mean squared error of the simulated $\widehat{k}_0^{(s)}$ $(s = 1, ..., S)$.
Calculate the empirical coverage probability.

Simulations for the homoscedastic case are reported in Section 5.2 while the heteroscedastic case is investigated in Section 5.3. In Section 5.4, we comment on some practical problems caused by rounding effects.

5.2. Homoscedastic case

Under homoscedasticity, we may utilize the asymptotic normality of $\widehat{k}_\mu$ and $\widehat{k}_0$ with σ^2 estimated by the "pooled" estimator $\widehat{\sigma}_{\text{pooled}}^2$.

A pilot simulation study, not presented here, with $n \in \{10, 20\}$ and $m \in \{20, 40\}$, comparing the empirical distributions of $\widehat{k}_\mu$ and $\widehat{k}_0$ suggests that both estimators are generally reasonably good but exhibit large mean squared error and negative bias for k_0 close to n. The mean squared error of $\widehat{k}_\mu$ is larger than the mean squared error of $\widehat{k}_0$ for small values of k_0. This observation corresponds to the asymptotic variances derived in Section 3, see also Figure 1. The coverage probabilities were close to the nominal values unless k_0 was very large (for both estimators) or very small (only for $\widehat{k}_\mu$). The coverage probabilities of the confidence intervals based on σ^2 and its estimator $\widehat{\sigma}_{\text{pooled}}^2$ were very similar.

The worse behavior $\widehat{k}_\mu$ for small k_0 seems to result from the additional uncertainty caused by estimating the parameter μ. This leads to the corrected estimator $\widehat{k}_\mu^{\text{corr}} = \widehat{k}_\mu - \widehat{\mu}/\widehat{\delta}_\mu$ that will also be considered in further simulations.

In Table 1, we investigate the empirical coverage probabilities of one-sided 95% bootstrap confidence intervals calculated with and without homoscedasticity assumptions (homoscedasticity assumptions are applicable only because the number of observations in each category is constant). Under homoscedasticity assumptions, we estimate the common variance by $\widehat{\sigma}^2_{\text{pooled}}$ and this variance estimator is also used in the bootstrap. More generally, we can also proceed without assuming homoscedasticity and follow Algorithm 1 from Section 4.2 using all $2n$ sample variances $\widehat{\sigma}^2_{ji}$.

Results in Table 1 confirm that coverage probabilities are rather small if the change occurs close to n. The heteroscedastic version works well even in the homoscedastic setup.

Table 1: Coverage probabilities (in %) of one-sided 95% confidence intervals of four change point estimators in the homoscedastic case (1000 simulations, $B = 1000$). The confidence intervals are based on bootstrapping utilizing either the pooled variance estimator $\widehat{\sigma}^2_{\text{pooled}}$ (homoscedastic version) or $2n$ sample variances $\widehat{\sigma}^2_{ji}$ (heteroscedastic version).

		θ_0	$\widehat{\sigma}^2_{\text{pooled}}$				$\widehat{\sigma}^2_{ji}$			
			$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\text{corr}}$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\text{corr}}$	$\widehat{k}_0(\widehat{\tau}^2)$
$n = 10$	$n_{ji} = 10$	0.1	88.8	95.2	93.3	94.0	89.5	93.9	93.7	94.5
		0.2	92.4	95.9	92.8	94.9	91.7	94.8	94.3	95.6
		0.4	94.1	92.3	92.9	91.9	95.5	92.4	93.3	92.0
		0.6	92.8	93.2	92.6	92.4	93.9	89.9	92.2	90.2
		0.8	90.4	90.5	90.0	90.8	89.6	87.7	89.2	89.1
		0.9	78.1	78.3	79.2	78.4	80.1	74.8	76.4	76.0
	$n_{ji} = 20$	0.1	93.9	92.0	94.4	92.1	95.1	92.8	94.6	93.0
		0.2	96.3	92.8	95.6	92.4	95.4	92.7	95.8	94.7
		0.4	93.5	92.0	92.3	91.1	92.7	91.6	92.0	90.8
		0.6	87.6	90.1	90.1	89.8	89.3	87.1	88.9	88.4
		0.8	88.8	89.0	89.7	87.8	89.9	86.9	87.2	88.4
		0.9	72.1	70.4	74.9	70.1	72.4	70.9	72.6	70.3
$n = 20$	$n_{ji} = 10$	0.1	96.5	94.3	94.0	94.9	94.9	93.5	95.8	93.1
		0.2	97.1	94.1	95.0	93.7	96.9	93.0	95.0	93.6
		0.4	94.0	93.7	93.8	93.9	94.4	92.1	93.0	92.8
		0.6	93.2	90.9	92.7	92.7	91.9	92.1	91.7	91.8
		0.8	94.8	95.6	94.3	95.3	93.8	94.5	92.5	93.7
		0.9	84.1	84.3	84.8	84.0	83.1	81.8	84.4	80.9
	$n_{ji} = 20$	0.1	97.3	95.0	94.4	94.9	97.0	93.5	93.4	95.3
		0.2	95.1	94.3	94.1	94.3	93.5	93.9	94.1	94.0
		0.4	93.0	93.1	93.1	92.9	93.6	93.6	93.1	94.7
		0.6	91.9	90.7	92.8	92.8	91.7	93.6	92.0	91.4
		0.8	93.2	91.9	91.3	90.7	91.8	92.5	91.5	89.3
		0.9	79.5	81.4	83.4	79.3	82.3	80.4	82.4	82.4

5.3. Heteroscedastic case

Real life is typically heteroscedastic and therefore we pay more attention to such situations. In Table 2, we investigate the behaviour of the proposed method in several artificial heteroscedastic situations caused both by different variances and numbers of observations in the observed categories.

Table 2: Coverage percentages (in %) of 95% bootstrap confidence intervals based on 4 change point estimators for $n \in \{10, 20\}$, $n_{ji} \equiv 10$, and $\sigma^2 = 1$ based on 1000 bootstrap replicates and 1000 simulations. The first four columns are obtained from the homoscedastic version of the bootstrap scheme using the pooled variance estimator $\widehat{\sigma}^2_{\text{pooled}}$.

	θ_0	$n = 10$								$n = 20$			
		$\widehat{\sigma}^2_{\text{pooled}}$				$\widehat{\sigma}^2_{ji}$				$\widehat{\sigma}^2_{ji}$			
		$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}^{\text{corr}}_\mu$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}^{\text{corr}}_\mu$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}^{\text{corr}}_\mu$	$\widehat{k}_0(\widehat{\tau}^2)$
H01	0.1	65.1	66.8	58.0	67.9	88.1	92.5	92.5	93.2	97.5	93.1	95.1	93.8
	0.4	69.9	68.9	67.7	70.7	96.3	94.7	95.7	92.2	94.1	93.3	94.4	95.1
	0.8	74.7	71.2	72.2	68.5	88.5	86.5	86.7	88.4	95.9	94.9	95.2	91.4
	0.9	84.1	82.8	77.9	80.5	78.6	77.4	78.8	78.1	77.5	80.4	80.8	78.9
H02	0.1	60.5	63.2	50.6	65.6	83.7	94.2	92.5	93.9	95.3	93.0	95.0	93.1
	0.4	65.9	65.4	63.9	69.6	90.6	88.4	91.0	93.0	93.6	92.5	92.8	93.2
	0.8	69.4	69.5	74.1	72.9	90.4	89.2	91.0	86.3	89.6	88.4	90.7	90.4
	0.9	80.0	78.2	77.2	83.7	76.8	71.9	75.6	75.8	82.6	84.2	82.3	81.9
H10	0.1	90.0	93.9	92.4	94.0	88.8	94.6	93.4	92.8	92.6	94.0	93.1	95.3
	0.4	97.1	99.6	99.7	99.7	92.7	91.5	94.3	91.1	94.6	94.9	94.9	93.3
	0.8	89.0	93.6	93.3	92.2	91.6	92.5	90.7	89.0	95.3	91.4	94.8	90.5
	0.9	94.1	87.5	87.5	68.5	92.4	88.8	86.1	73.4	87.5	86.6	89.1	82.3
H11	0.1	65.6	65.4	55.6	69.9	89.9	93.1	91.8	93.1	90.4	95.5	91.7	94.2
	0.4	71.0	67.2	67.4	70.8	92.4	94.8	93.2	91.9	93.7	92.4	93.1	93.3
	0.8	79.0	78.7	76.1	71.0	92.2	84.3	90.7	89.8	96.8	95.9	96.4	89.3
	0.9	95.5	92.5	84.6	73.1	93.1	90.3	87.5	66.2	88.4	86.8	86.3	80.1
H12	0.1	58.9	63.9	49.9	64.0	88.3	95.6	92.0	94.2	88.8	94.4	92.4	93.1
	0.4	70.3	68.3	65.3	69.5	90.0	87.1	88.6	91.5	94.5	93.0	94.7	92.9
	0.8	79.6	77.5	78.8	72.2	91.1	90.1	92.7	87.9	93.9	88.5	91.3	88.8
	0.9	93.2	88.3	81.7	78.0	91.3	85.2	85.7	74.3	88.6	87.3	90.9	82.2
H20	0.1	80.9	92.4	91.0	99.4	82.1	91.5	88.0	98.5	97.0	97.2	98.9	99.3
	0.4	93.8	94.6	93.2	99.8	93.0	89.6	89.9	93.7	97.4	95.4	96.6	99.2
	0.8	78.0	75.5	79.1	74.1	78.2	77.4	76.2	73.8	93.2	90.6	91.1	94.1
	0.9	90.6	86.2	84.8	78.3	90.0	86.5	83.3	77.5	79.0	78.4	83.7	74.4
H21	0.1	58.8	63.8	48.8	66.9	76.6	88.7	80.7	94.5	87.8	87.4	87.0	93.3
	0.4	60.6	62.8	61.9	65.2	83.9	82.0	83.3	88.4	85.0	85.4	84.2	90.8
	0.8	73.3	70.7	68.0	69.3	79.9	79.4	76.9	79.8	84.7	84.8	84.3	88.1
	0.9	93.3	88.7	81.8	82.5	87.6	85.4	81.3	78.9	81.1	81.9	79.7	77.4
H22	0.1	49.8	58.4	39.4	69.1	61.6	84.4	73.7	94.8	79.0	83.2	80.4	93.0
	0.4	57.1	61.2	56.0	72.5	74.1	68.2	75.6	90.2	78.0	81.9	81.0	93.3
	0.8	72.6	67.5	72.1	67.6	82.6	79.0	74.9	73.5	76.7	75.4	75.5	89.4
	0.9	92.2	86.6	79.2	79.9	90.3	83.7	83.2	82.1	83.9	78.7	81.9	76.8

Here, we consider altogether 8 heteroscedastic situations obtained by considering two simple models for nonconstant variances and two simple models for nonconstant numbers of observations. The simulation setups (H01,...,H22) are summarized in the following table:

	Nr. of observations (n_{ji})		
	$n_{ji} = m$	$m\{1 + 3I(i \text{ odd})\}/2$	$m\{1 + 3I(i > n/2)\}/2$
σ_{ji} constant ($\sigma_{ji} = \sigma$)		H01	H02
$\sigma_{ji} = \sigma(1 + 2I(i > k_0))$	H10	H11	H12
$\sigma_{ji} = \sigma(1 + 2I(i \text{ even}))$	H20	H21	H22

As expected, Table 2 shows that bootstrap using the pooled estimator of variance does not lead to reliable results in the heteroscedastic setup. The confidence intervals based on the heteroscedastic estimator $\widehat{k}_0(\widehat{\tau}^2)$ provide reasonable coverage probabilities for all scenarios as long as k_0 is not too close to n.

5.4. Rounding effects

In the jumping speeds example, children aged i to $i+1$ years are included in the i-th age category. We use summary statistics observed in these age categories and we have to keep in mind that the i-th observed sample mean and sample standard deviation correspond to the marginal distribution of jumping speeds for all children aged from i to $i + 1$ years.

Assuming that $E(Y_1 \,|\, \text{Age}{=}x) = E(Y_2 \,|\, \text{Age}{=}x) + \delta((x - k_0)/n)_+$ for $x \in (1, n + 1)$ and that the age distribution in both groups is the same, it follows that $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = 0$, for $i \le \lfloor k_0 \rfloor$, and the true $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = \delta(i - k_0)/n$ for $i \ge \lceil k_0 \rceil$. Hence, for sample means based on categorization of continuous explanatory variable, the model (A3) is valid only if k_0 is a natural number.

Denoting $i_0 = \lfloor k_0 \rfloor$ and $d_0 = k_0 - i_0$, we may calculate the true expectation of the mean differences $E(\overline{Y}_{1i_0} - \overline{Y}_{2i_0}) = E\big\{\delta((X - k_0)/n)_+\big\}$ under the above assumptions (with uniform distribution of the explanatory variable X in the i_0-th age category and for $d_0 > 0$):

$$E(\overline{Y}_{1i_0} - \overline{Y}_{2i_0}) \;=\; \frac{\delta}{n} \int\limits_{i_0+d_0}^{i_0+1} (x - k_0)\, dx \;=\; \frac{\delta}{n} \int\limits_{0}^{1-d_0} x\, dx \;=\; \frac{\delta(1 - d_0)^2}{2n}\,.$$

In Figure 2, we plot the theoretical expectation for various values of k_0. Obviously, whenever k_0 is not a natural number, the i_0-th sample mean can be "somewhat larger than it should be".

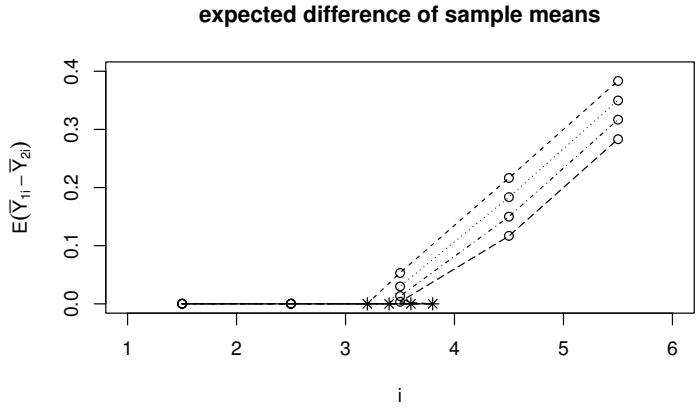

Figure 2: Expectations of mean differences for five categories for change points $k_0 \in (3.2, 3.4, 3.6, 3.8)$. Each line connects the expectations (denoted by circles) corresponding to given changepoint (denoted by star).

In practice, it is more natural to define the i-th category by values of the explanatory variable $x \in (i - 0.5, i + 0.5)$ and this notation is also in accordance with the theoretical part of this paper. Therefore, we define the bias corrected estimator $\widehat{k}_0^{\mathrm{bc}}$ by using $x_{ik}^{\mathrm{bc}} = I(i > \lceil k-0.5 \rceil)\,(i-k)/n + I(i = \lceil k-0.5 \rceil)\,(\lceil k-0.5 \rceil - k + 0.5)/(2n)$ instead of x_{ik} in (3.2).

Results of a small simulation study comparing the behavior of $\widehat{k}_0$ and $\widehat{k}_0^{\mathrm{bc}}$ in a homoscedastic case are given in Table 3. As expected, the empirical bias of the bias corrected estimator $\widehat{k}_0^{\mathrm{bc}}$ tends to be somewhat smaller. The effect of the rounding bias on the coverage probabilities based on $\widehat{k}_0$ is most clearly visible for $n = 20$, $n_{ji} \equiv 20$, and k_0 lying in the center of the category (i.e., for $k_0 = 14$, 15, and 16).

6. JUMPING SPEEDS

In order to analyze the real data set given in Table 5, it is important to understand the meaning of the row-labels. The various labels and its meanings are summarized in Table 4. In the theoretical part of this paper, we were using the "Index scale" given in the first column. For practical considerations, it is important to notice that $k = 1$ actually corresponds to children aged approximately 6.5 years.

In order to calculate the estimators $\widehat{k}_0(\widehat{\tau}^2)$ and $\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$, we maximize the function $T_{2+\widehat{\tau}^2}(k)$ and its bias corrected version, $T^{\mathrm{bc}}_{2+\widehat{\tau}^2}(k)$, plotted in Figure 3.

Table 3: Empirical mean squared error (MSE), bias and coverage probabilities of 95% confidence intervals (in %) for $\widehat{k}_0$ and $\widehat{k}_0^{\mathrm{bc}}$, 1000 simulations with 1000 bootstrap replicates.

			$\widehat{k}_0$			$\widehat{k}_0^{\mathrm{bc}}$		
		θ_0	MSE	bias	coverage	MSE	bias	coverage
$n = 10$	$n_{ji} \equiv 10$	0.20	0.124	0.003	93.6%	0.112	−0.010	94.6%
		0.22	0.113	−0.001	95.4%	0.113	−0.016	95.3%
		0.25	0.125	−0.018	91.5%	0.123	−0.010	92.9%
		0.28	0.122	0.012	91.2%	0.133	0.011	90.3%
		0.30	0.116	0.008	93.9%	0.131	0.001	95.0%
		0.70	0.432	−0.109	92.7%	0.365	−0.041	93.3%
		0.72	0.466	−0.099	94.0%	0.498	−0.065	91.7%
		0.75	0.678	−0.151	89.6%	0.726	−0.138	88.4%
		0.78	0.936	−0.226	86.5%	0.971	−0.123	91.5%
		0.80	1.160	−0.263	91.3%	1.255	−0.224	91.8%
	$n_{ji} \equiv 20$	0.20	0.053	−0.011	94.1%	0.052	0.002	93.2%
		0.22	0.054	−0.013	95.4%	0.050	0.003	98.4%
		0.25	0.053	−0.011	95.8%	0.060	−0.017	95.4%
		0.28	0.059	0.001	92.8%	0.054	−0.009	95.2%
		0.30	0.063	0.000	92.6%	0.060	0.011	92.8%
		0.70	0.177	−0.056	86.9%	0.173	0.004	96.3%
		0.72	0.194	−0.073	94.1%	0.191	−0.024	95.6%
		0.75	0.246	−0.072	94.6%	0.230	−0.052	91.7%
		0.78	0.319	−0.116	88.1%	0.302	−0.034	93.2%
		0.80	0.399	−0.097	88.0%	0.506	−0.092	96.9%
$n = 20$	$n_{ji} \equiv 10$	0.20	0.054	−0.005	93.6%	0.050	−0.004	95.7%
		0.22	0.056	−0.005	94.5%	0.057	−0.003	95.3%
		0.25	0.056	−0.016	94.6%	0.050	0.003	95.4%
		0.28	0.061	0.002	94.5%	0.055	−0.009	94.6%
		0.30	0.063	−0.004	93.1%	0.060	−0.012	93.9%
		0.70	0.150	−0.040	93.7%	0.158	0.001	94.7%
		0.72	0.175	−0.035	93.9%	0.163	−0.033	94.5%
		0.75	0.200	−0.051	93.3%	0.178	−0.023	96.7%
		0.78	0.220	−0.043	94.0%	0.229	−0.026	90.8%
		0.80	0.263	−0.050	94.7%	0.276	−0.022	93.8%
	$n_{ji} \equiv 20$	0.20	0.027	−0.006	94.0%	0.026	−0.004	93.3%
		0.22	0.026	−0.008	94.3%	0.024	−0.010	98.0%
		0.25	0.030	−0.006	93.3%	0.029	0.005	93.6%
		0.28	0.026	0.005	95.8%	0.030	0.001	95.9%
		0.30	0.033	−0.015	93.9%	0.031	−0.004	94.1%
		0.70	0.078	−0.024	90.7%	0.072	−0.000	94.1%
		0.72	0.089	−0.032	96.7%	0.075	−0.001	95.9%
		0.75	0.093	−0.037	90.6%	0.095	−0.006	93.4%
		0.78	0.112	−0.046	95.3%	0.103	−0.012	94.3%
		0.80	0.114	−0.040	90.0%	0.111	−0.012	95.7%

Table 4: Meaning of row labels in the jumping speeds example.

Index (k)	Label	Meaning	Interpretation	$\overline{Y_1}$ ($\widehat{\sigma}_1$)	$\overline{Y_2}$ ($\widehat{\sigma}_2$)
1	6	6–7 years	$\sim$ 6.5 years	1.89 (0.17)	1.87 (0.18)
2	7	7–8 years	$\sim$ 7.5 years	2.00 (0.21)	1.98 (0.20)
⋮	⋮	⋮	⋮	⋮	⋮
13	18	18–19 years	$\sim$ 18.5 years	2.33 (0.17)	2.87 (0.10)

In both plots, the estimator $\widehat{k} = 5$ (on the "Index scale") corresponds to the estimated change point $\widehat{k}^{\mathsf{age}} = 10.5$ years.

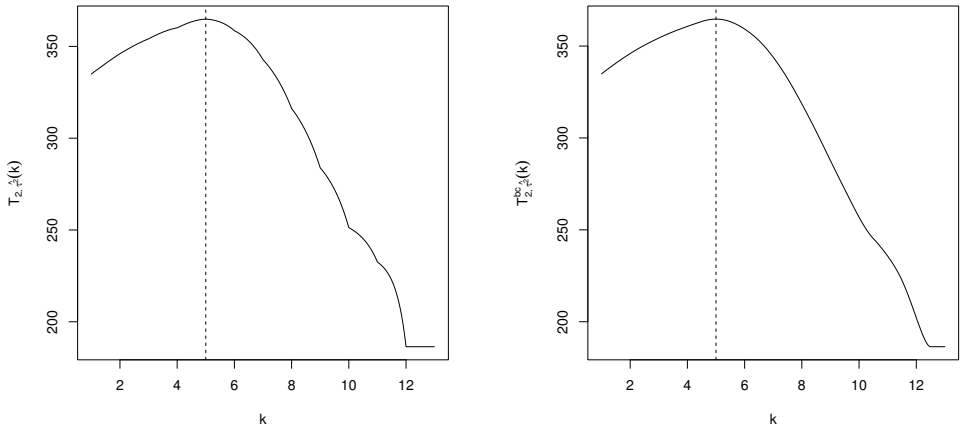

Figure 3: Function $T_{2+\widehat{\tau}^2}(k)$ and its bias corrected version $T^{\mathrm{bc}}_{2+\widehat{\tau}^2}(k)$ for the jumping speed data. The vertical dashed lines denote the estimates $\widehat{k}$ and $\widehat{k}^{\mathrm{bc}}$.

Applying the bootstrap algorithm described in Section 4.2, we obtain that the upper limit of the one-sided 95% confidence interval based on $\widehat{k}_0(\widehat{\tau}^2)$ is $5.72 + 5.5 = 11.22$ years. Applying the bias correction from Section 5.4, we obtain the one-sided 95% confidence interval $(-\infty, 11.26)$ years.

For both estimators, the test of the null hypothesis "no changepoint before 12 years" is actually carried out by testing the index $k_1 = 12 - 5.5$ (see Remark 4.1 and Table 4). The p-values corresponding to the change-point tests of the null hypothesis $H_0 \colon k_0 \geq k_1$ against $H_1 \colon k_0 < k_1$ for $k_1 \in \{0.5, ..., 12.5\}$ are given in Table 5. Since each test concerns the age $k_1 + 5.5$ years, it seems more natural to shift the lines with these p-values in order to point out the difference between the two-sample t-test (comparing the marginal means in i-th age category, i.e., for approximately $i + 0.5$ years) and the change-point approach (testing whether there is a significant difference for children aged precisely i years).

Table 5: Observed mean jumping speeds and standard deviations for boys and girls in 13 age categories. P-values of the two-sample t-test in each age category, its Bonferroni and Benjamini–Hochberg (BH) adjustments and p-values of the test for change point location based on $\widehat{k}_0(\widehat{\tau}^2)$ and $\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$.

Age cat.	girls		boys		p-values					Age
	$\overline{Y}_1$ $(\widehat{\sigma}_1)$	n_1	$\overline{Y}_2$ $(\widehat{\sigma}_2)$	n_2	t-test	Bonferroni	BH	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$	
6–7	1.89 (0.17)	33	1.87 (0.18)	19	0.780	1.000	0.780	1.000	1.000	6
7–8	2.00 (0.21)	43	1.98 (0.20)	38	0.646	1.000	0.763	1.000	1.000	7
8–9	2.01 (0.21)	33	2.06 (0.21)	38	0.369	1.000	0.479	1.000	1.000	8
9–10	2.06 (0.18)	42	2.14 (0.18)	29	0.081.	1.000	0.117	0.999	0.997	9
10–11	2.19 (0.22)	42	2.17 (0.19)	45	0.713	1.000	0.773	0.861	0.846	10
11–12	2.23 (0.15)	30	2.31 (0.23)	37	0.062.	0.800	0.100	0.113	0.117	11
12–13	2.26 (0.13)	41	2.35 (0.23)	40	0.047*	0.615	0.088.	0.003**	0.003**	12
13–14	2.30 (0.22)	32	2.53 (0.21)	36	0.000***	0.001***	0.000***	0.000***	0.000***	13
14–15	2.28 (0.23)	31	2.66 (0.19)	20	0.000***	0.000***	0.000***	0.000***	0.000***	14
15–16	2.37 (0.17)	29	2.72 (0.22)	26	0.000***	0.000***	0.000***	0.000***	0.000***	15
16–17	2.33 (0.19)	17	2.83 (0.28)	9	0.001***	0.006**	0.001**	0.000***	0.000***	16
17–18	2.35 (0.18)	25	2.76 (0.16)	13	0.000***	0.000***	0.000***	0.000***	0.000***	17
18–19	2.33 (0.17)	34	2.87 (0.10)	14	0.000***	0.000***	0.000***			18

We conclude that the estimated change-point is 10.5 years (with 95% confidence interval $(-\infty, 11.26)$) while the two-sample t-tests *without multiple testing correction* show statistically significant difference only after 12 years (in the age category 12 to 13 years).

In order to verify the validity of Assumption (A3), we plot the data set and the resulting least squares fit in Figure 4.

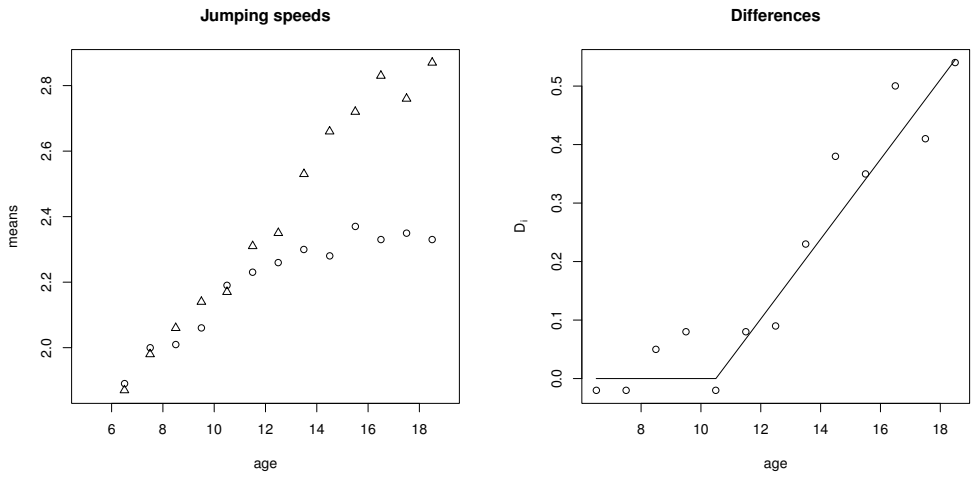

Figure 4: Observed sample means of jumping speed for boys (△) and girls (◯) in thirteen age categories. The right plot shows the observed differences D_i and the least squares fit.

7. SUMMARY AND OUTLOOK

A rigorous approach to multiple hypotheses testing is needed in many real-life situations. Typically, a Bonferroni-type adjustment increases all p-values in order to control either the family-wise error rate or the false discovery rate. However, the structure of the observed data often calls for a more appropriate and powerful solution. Using gender-specific growth curves as a motivation, we proposed a simple two-sample gradual change model in order to develop bootstrap-based tests and confidence intervals for the location of the unknown change-point. In this way, many two-sample t-tests can be replaced by a single test concerning only the change-point. Therefore, adjustments for multiple hypotheses testing become unnecessary.

In practice, the linearity assumption may not be fulfilled. This problem can be solved in a simple way, e.g., by using a finer grid to investigate only a small neighborhood of the suspected change point.

Obviously, the proposed method is applicable also to different sample characteristics. For example, we could investigate a two-sample gradual change in the slope using a table of estimated slopes (and estimates of their standard deviations) in each age category. Such a test would correspond to a model of quadratic change for the original observations.

Depending on further applications, various extensions of the proposed methodology to more general setups may be considered, e.g., dependent observations and more general changes than a linear trend. Also some aspects of nonparametric regression can be utilized if one can analyze the original data set instead of only sample means and sample standard deviations observed in n ordered categories.

ACKNOWLEDGMENTS

The research was supported by GAČR 15-09663S and IAP P7/06 StUDyS network. We thank Zdeněk Šumník (University Hospital in Motol and 2nd Faculty of Medicine, Charles University) for the permission to use the jumping speeds data set. We are grateful to the Editor and an anonymous referee for a number of useful remarks that lead to a considerable improvement in presentation.

REFERENCES

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57**(1), 289–300.

[2] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency, *Ann. Statist.*, **29**(4), 1165–1188.

[3] Feder, P.I. (1975). On asymptotic distribution theory in segmented regression problems–identified case, *Ann. Statist.*, **3**(1), 49–83.

[4] Hinkley, D.V. (1971). Inference in two-phase regression, *J. Amer. Statist. Assoc.*, **66**(336), 736–743.

[5] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*, **75**(4), 800–802.

[6] Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scand. J. Stat.*, **6**(2), 65–70.

[7] Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test, *Biometrika*, **75**(2), 383–386.

[8] Hušková, M. (1998). Estimators in the location model with gradual changes, *Comment. Math. Univ. Carolin.*, **39**(1), 147–157.

[9] Hušková, M. (1999). Gradual changes versus abrupt changes, *J. Statist. Plann. Inference*, **76**(1), 109–125.

[10] Hušková, M. and Steinebach, J. (2000). Limit theorems for a class of tests of gradual changes, *J. Statist. Plann. Inference*, **89**(1–2), 57–77.

[11] Hušková, M. and Steinebach, J. (2002). Asymptotic tests for gradual changes, *Statist. Decisions*, **20**(2), 137–151.

[12] Jarušková, D. (1998). Testing appearance of linear trend, *J. Statist. Plann. Inference*, **70**(2), 263–276.

[13] R Core Team (2015). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[14] Shaban, S.A. (1980). Change point problem and 2-phase regression: an annotated bibliography, *Int. Stat. Rev.*, **48**(1), 83–93.

[15] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer.

[16] Šumník, Z.; Matysková, J.; Hlávka, Z.; Durdilová, L.; Souček, O. and Zemková, D. (2013). Reference data for jumping mechanography in healthy children and adolescents aged 6–18 years, *J. Musculoskelet. Neuronal. Interact.*, **13**(3), 259–273.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 373–394

CONTINUOUS POST-MARKET SEQUENTIAL SAFETY SURVEILLANCE WITH MINIMUM EVENTS TO SIGNAL

Authors: Martin Kulldorff
– Department of Medicine, Brigham and Women's Hospital and Harvard Medical School,
Boston, MA, USA
martin_kulldorff@hms.harvard.edu

Ivair R. Silva
– Department of Statistics, Federal University of Ouro Preto,
Brazil
ivairest@gmail.com

Received: March 2015 Revised: August 2015 Accepted: August 2015

Abstract:

- Continuous sequential analysis is increasingly used for near real-time post-market drug and vaccine safety surveillance. We explore continuous sequential monitoring when the null cannot be rejected until a minimum number of adverse events have occurred. For fixed alpha, one can simultaneously increase the statistical power and reduce the expected time to signal. We also evaluate continuous sequential analysis with a delayed start until a certain sample size has been attained. This is only useful if the start of the surveillance is delayed for logistical reasons. Tables with exact critical values, statistical power and the average times to signal are provided.

Key-Words:

- *drug safety; pharmacovigilance; continuous sequential analysis; surveillance; sequential probability ratio test.*

AMS Subject Classification:

- 62L10, 62F03.

1. INTRODUCTION

Post-market drug and vaccine safety surveillance is important in order to detect rare but serious adverse events not found during pre-licensure clinical trials. Safety problems may go undetected either because an adverse reaction is too rare to occur in sufficient numbers among the limited sample size of a phase three clinical trial, or because the adverse reaction only occur in a certain sub population that was excluded from the trial, such as frail individuals.

In order to detect a safety problem as soon as possible, the CDC Vaccine Safety Datalink project pioneered the use of near real-time safety surveillance using automated weekly data feeds from electronic health records [1, 2, 3]. In such surveillance, the goal is to detect serious adverse reactions as early as possible without too many false signals. It is then necessary to use sequential statistical analysis, which adjusts for the multiple testing inherent in the many looks at the data. Using the maximized sequential probability ratio test (MaxSPRT) [4], all new childhood vaccines and some adult vaccines are now monitored in this fashion [1, 5, 6, 7, 8, 9, 10, 11, 12, 13]. There is also interest in using sequential statistical methods for post-market drug safety surveillance [20, 14, 15, 16, 17, 18], and the methods presented in this paper may also be used in either settings.

In contrast to group sequential analyses, continuous sequential methods can signal after a single adverse event, if that event occurs sufficiently early. In some settings, such as a phase 2 clinical trial, that may be appropriate, but in post-market safety surveillance it is not. In post-market vaccine surveillance, an ad-hoc rule that require at least two or three events to signal has sometimes been used, but that leads to a conservative type 1 error (alpha level). In this paper we provide exact critical values for continues sequential analysis when a signal is required to have a certain minimum number of adverse events. We also evaluate power and expected time to signal for various alternative hypotheses. It is shown that it is possible to simultaneously improve both of these by requiring at least 3 or 4 events to signal. Note that it is still necessary to start surveillance as soon as the first few individuals are exposed, since they all could have the adverse event.

For logistical reasons, there is sometimes a delay in the start of post-marketing safety surveillance, so that the first analysis is not conducted until a group of people have already been exposed to the drug or vaccine. This is not a problem when using group sequential methods, as the first group is then simply defined to correspond to the start of surveillance. For continuous sequential surveillance, a delayed start needs to be taken into account when calculating the critical values. In this paper, we present exact critical values when there is a delayed start in the sequential analysis. We also calculate the power and time to signal for different relative risks.

In addition to ensuring that the sequential analysis maintains the correct overall alpha level, it is important to consider the statistical power to reject the null hypothesis; the average time until a signal occurs when the null hypothesis is rejected; and the final sample size when the null hypothesis is not rejected. For any fixed alpha, there is a trade-off between these three metrics, and the trade-off depends on the true relative risks. In clinical trials, where sequential analyses are commonly used, statistical power and the final sample size are usually the most important design criteria. The latter is important because patient recruitment is costly. The time to signal is usually the least important, as a slight delay in finding an adverse event only affects the relatively small number of patients participating in the clinical trial, but not the population-at-large. In post-market safety surveillance, the trade-off is very different. Statistical power is still very important, but once the surveillance system is up and running, it is easy and cheap to prolong the length of the study by a few extra months or years to achieve a final sample size that provides the desired power. Instead, the second most critical metric is the time to signal when the null is rejected. Since the product is already in use by the population-at-large, most of which are not part of the surveillance system, a lot of people may be spared the adverse event if a safety problem can be detected a few weeks or months earlier. This means that for post-market vaccine and drug safety surveillance, the final sample size when the null is not rejected is the least important of the three metrics.

All calculations in this paper are exact, and none are based on simulations or asymptotic statistical theory. The numerical calculation of the exact critical values is a somewhat cumbersome process. So that users do not have to do these calculations themselves, we present tables with exact critical values for a wide range of parameters. For other parameters, we have developed the open source R package 'Sequential', freely available at '`cran.r-project.org/web/packages/Sequential`'.

2. CONTINUOUS SEQUENTIAL ANALYSIS FOR POISSON DATA

Sequential analysis was first developed by Wald [19, 21], who introduced the sequential probability ratio test (SPRT) for continuous surveillance. The likelihood based SPRT proposed by Wald is very general in that it can be used for many different probability distributions. The SPRT is very sensitive to the definition of the alternative hypothesis of a particular excess risk. For post-market safety surveillance, a maximized sequential probability ratio test with a composite alternative hypothesis has often been used instead. This is both a 'generalized sequential probability ratio test' [22] and 'sequential generalized likelihood ratio test' [23, 24]. In our setting, it is defined as follows, using the Poisson distribution to model the number of adverse events seen [4].

Let C_t be the random variable representing the number of adverse events in a pre-defined risk window from 1 to W days after an incident drug dispensing that was initiated during the time period $[0, t]$. Let c_t be the corresponding observed number of adverse events. Note that time is defined in terms of the time of the drug dispensing rather than the time of the adverse event, and that hence, we actually do not know the value of c_t until time $t + W$.

Under the null hypothesis (H_0), C_t follows a Poisson distribution with mean μ_t, where μ_t is a known function reflecting the population at risk. In our setting, μ_t reflects the number of people who initiated their drug use during the time interval $[0, t]$ and a baseline risk for those individuals, adjusting for age, gender and any other covariates of interest. Under the alternative hypothesis (H_A), the mean is instead $RR\mu_t$, where RR is the increased relative risk due to the drug/vaccine. Note that $C_0 = c_0 = \mu_0 = 0$.

For the Poisson model, the MaxSPRT likelihood ratio based test statistic is

$$\begin{aligned} LR_t \;=\; \max_{H_A} \frac{P\big(C_t = c_t \,|\, H_A\big)}{P(C_t = c_t \,|\, H_0)} \;&=\; \max_{RR>1} \frac{e^{-RR\mu_t}(RR\mu_t)^{c_t}/c_t!}{e^{-\mu_t}\mu_t^{c_t}/c_t!} \\ &=\; \max_{RR>1} e^{(1-RR)\mu_t}(RR)^{c_t}\,. \end{aligned}$$

The maximum likelihood estimate of RR is c_t/μ_t, so

$$LR_t \;=\; e^{\mu_t - c_t}(c_t/\mu_t)^{c_t}\,.$$

Equivalently, when defined using the log likelihood ratio

$$\begin{aligned} LLR_t(c_t) \;=\; ln(LR_t) \;&=\; \max_{RR>1}\big((1-RR)\,\mu_t + c_t\,ln(RR)\big) \\ &=\; (\mu_t - c_t) + c_t\,ln(c_t/\mu_t)\,. \end{aligned}$$

Note that, since μ_t is known, the test statistic is only a function of c_t. This shall be useful when calculating exact critical values, in Section 3.1. The MaxSPRT test statistic is sequentially monitored for all values of $t > 0$, until either $LLR_t \geq CV$, in which case the null hypothesis is rejected, or until $\mu_t = T$, in which case the alternative hypothesis is rejected. T is a predefined upper limit on the length of surveillance, defined in terms of the sample size, expressed as the expected number of adverse events under the null hypothesis. It is roughly equivalent to a certain number of exposed individuals, but adjusted for covariates. Exact critical values (CV) are available for the MaxSPRT [4], obtained through iterative numerical calculations.

3. MINIMUM NUMBER OF EVENTS REQUIRED TO SIGNAL

Continuous sequential probability ratio tests may signal at the time of the first event, if that event appears sufficiently early. One could add a requirement that there need to be a minimum of M events before one can reject the null hypothesis. This still requires continuous monitoring of the data from the very start, as M events could appear arbitrarily early. Hence, there is no logistical advantage of imposing this minimum number. The potential advantage is instead that it may reduce the time to signal and/or increase the statistical power of the study. Below, in Section 3.2, it is shown that both of these can be achieved simultaneously.

3.1. Exact Critical Values

In brief, first note that the time when the critical value is reached and the null hypothesis is rejected can only happen at the time when an event occurs. For any specified critical value CV and maximum sample size T, it is then possible to calculate the probability of rejecting the null, using a bisection iterative approach.

As mentioned in the last section, the exact critical value can be obtained analytically, and the details for doing so are described in the present section. Firstly, it is important to note that, for each fixed CV, the signaling threshold can be written in the time scale. This is so because the MaxSPRT statistic, $LLR_t(c_t)$, is monotone non-increasing with μ_t for each fixed $c_t > 0$, which means that the null hypothesis is rejected when an event arrives too fast in comparison to its expected time of arrival when the null is true. Thus, let τ_n denote the arrival time of the n-th event. Once $CV > 0$ is fixed, there are constants $0 < \mu_{(1)} \leq \mu_{(2)} \leq \cdots \leq \mu_{(N)}$ such that the probability of rejecting the null hypothesis can be expressed as:

$$\begin{aligned} \Pr\big[\text{rejecting } H_0 \,|\, RR\big] &= \Pr\Big[LLR_t \geq CV \text{ for some } t \in (0,T) \,|\, RR\Big] \\ &= \Pr\left[\bigcup_{n=1}^{N} \big\{\tau_n \leq \mu_{(n)}\big\} \,|\, RR\right], \end{aligned} \tag{3.1}$$

where, for a minimum number M of events required to reject the null, N is the maximum length of surveillance given in the scale of the number of events such that $N := \max\{c \in \mathbb{N}\colon LLR_T(c) \leq CV\}$, $\mu_{(1)} = \cdots = \mu_{(M)}$, $\mu_{(n)} = \sup\{\mu^* > 0\colon LLR_{\mu^*}(n) \geq CV\}$ for $n = M, ..., (N-1)$, and $\mu_{(N)} = T$. Because C_t is a Poisson-based process, we can write $\mu_t = \lambda\, t$, where λ is a known constant. Then, the joint probability density function of the random vector $(\tau_1, ..., \tau_N)$, denoted here

with $f_{\tilde{\tau}}(y_1, ..., y_N)$, can be expressed as following:

$$f_{\tilde{\tau}}\big(y_1, ..., y_N \,|\, RR\big) \;=\; (RR\lambda)^N e^{-y_N RR\lambda}\, I(y_N > 0)\;. \tag{3.2}$$

Now, consider the new random vector $\tilde{T} = \lambda\tilde{\tau} = (T_1, ..., T_N)$, which, by its turn, has density:

$$f_{\tilde{T}}\big(t_1, ..., t_N \,|\, RR\big) \;=\; RR^N e^{-t_N RR}\, I(t_N > 0)\;. \tag{3.3}$$

With the last expression, the probability of rejecting the null hypothesis, say $\pi(RR, CV)$, is simply:

$$\begin{aligned} \pi(RR, CV) \;=\; \Pr\big[\text{rejecting } H_0 \,|\, RR\big] \;&=\; \Pr\left(\bigcup_{n=1}^{N} \big\{T_n \le \mu_{(n)}\big\} \,|\, RR\right) \\ &=\; \sum_{n=1}^{N} \Pr\big(\mathtt{N} = n \,|\, RR\big)\;, \end{aligned} \tag{3.4}$$

where $\mathtt{N}$ is the total number of events observed until the signaling moment. In order to understand the behaviour of $\pi(RR, CV)$ as a function of $\mathtt{N}$, let us evaluate it for $\mathtt{N} = 1, 2, 3, 4$. For $\mathtt{N} = 1$:

$$\Pr\big(\mathtt{N} = 1 \,|\, RR\big) \;=\; \Pr\big(T_1 \le \mu_{(1)}\big) \;=\; 1 - e^{-\mu_{(1)} RR}\,. \tag{3.5}$$

For $\mathtt{N} = 2$:

$$\begin{aligned} \Pr\big(\mathtt{N} = 2 \,|\, RR\big) \;&=\; \Pr\big(T_1 > \mu_{(1)} \cap T_2 \le \mu_{(2)}\big) \\ &=\; \int_{\mu_{(1)}}^{\mu_{(2)}} \int_{\mu_{(1)}}^{t_2} RR^2 e^{-RRt_2}\, dt_1\, dt_2 \\ &=\; \Pr\big(\mu_{(1)} \le T_2 \le \mu_{(2)} \,|\, RR\big) \\ &\quad -\; RR^{-1}\mu_{(1)} \left[\Pr\big(\mu_{(1)} \le T_1 \le \mu_{(2)} \,|\, RR\big)\right]\;. \end{aligned} \tag{3.6}$$

For $\mathtt{N} = 3$:

$$\begin{aligned} \Pr\big(\mathtt{N} = 3 \,|\, RR\big) \;&=\; RR^{-1} \Pr\big(T_1 > \mu_{(1)} \cap T_2 > \mu_{(2)} \cap T_3 \le \mu_{(3)}\big) \\ &=\; \int_{\mu_{(2)}}^{\mu_{(3)}} \int_{\mu_{(2)}}^{t_3} \int_{\mu_{(1)}}^{t_2} RR^3 e^{-RRt_3}\, dt_1\, dt_2\, dt_3 \\ &=\; \Pr\big(\mu_{(2)} \le T_3 \le \mu_{(3)} \,|\, RR\big) \\ &\quad -\; RR^{-1}\mu_{(1)} \Pr\big(\mu_{(2)} \le T_2 \le \mu_{(3)} \,|\, RR\big) \\ &\quad +\; RR^{-2}\big(\mu_{(1)}\mu_{(2)} - \mu_{(2)}^2/2\big) \Pr\big(\mu_{(2)} \le T_1 \le \mu_{(3)} \,|\, RR\big)\,. \end{aligned} \tag{3.7}$$

Finally, for $\mathtt{N} = 4$:

$$\begin{aligned}
\Pr\big(\mathtt{N}=4 \,|\, RR\big) &= RR^{-1} \Pr\big(T_1 > \mu_{(1)} \cap T_2 > \mu_{(2)} \cap T_3 > \mu_{(3)} \cap T_4 \le \mu_{(4)}\big) \\
&= \int_{\mu_{(3)}}^{\mu_{(4)}} \int_{\mu_{(3)}}^{t_4} \int_{\mu_{(2)}}^{t_3} \int_{\mu_{(1)}}^{t_2} RR^4 e^{-RRt_4}\, dt_1\, dt_2\, dt_3\, dt_4 \\
&= \Pr\big(\mu_{(3)} \le T_4 \le \mu_{(4)} \,|\, RR\big) \\
&\quad - RR^{-1} \mu_{(1)} \Pr\big(\mu_{(3)} \le T_3 \le \mu_{(4)} \,|\, RR\big) \\
&\quad + RR^{-2} \big(\mu_{(1)}\mu_{(2)} - \mu_{(2)}^2/2\big) \Pr\big(\mu_{(3)} \le T_2 \le \mu_{(4)} \,|\, RR\big) \\
&\quad - RR^{-3} \left[\frac{\mu_{(3)}^3}{3!} - \frac{\mu_{(1)}\,\mu_{(3)}^2}{2} + \mu_{(3)} \left(\mu_{(1)}\,\mu_{(2)} - \frac{\mu_{(2)}^2}{2} \right) \right] \\
&\quad \times \Pr\big(\mu_{(3)} \le T_1 \le \mu_{(4)} \,|\, RR\big)\,.
\end{aligned} \tag{3.8}$$

Thus, a recursive expression, with respect to $\mathtt{N}$, can be written to express $\pi(RR, N)$:

$$\pi(RR, CV) = \sum_{\mathtt{N}=1}^{N} \sum_{i=1}^{\mathtt{N}} (-1)^{i+1}\, \psi_i \Pr\big(\mu_{(\mathtt{N}-1)} \le T_{\mathtt{N}+1-i} \le \mu_{(\mathtt{N})} \,|\, RR\big)\,, \tag{3.9}$$

where $\mu_0 = 0$, $\psi_1 = 1$, and, for $i = 2, ..., N$,

$$\psi_i = \sum_{j=1}^{i-1} (-1)^{j+1} \frac{\big(r\mu_{(i-1)}\big)^j}{j!}\, \psi_{i-j}\,.$$

Because (3.9) is monotone decreasing with respect to CV, we can obtain the critical value, under a fixed precision ϵ, for any $\alpha \in (0, 1)$ through numerical calculation. For an alpha level of 0.05, the magnitude of CV is about 3 or 4 depending on the value of T, and it will usually not take values larger than 20 even for very small alpha level and very large T like $\alpha = 0.00001$ and $T = 1000$. The following steps can be used for finding the exact critical value for fixed $T > 0$, $\alpha \in (0, 1)$, $M \in \mathbb{N}$, and $\epsilon > 0$:

- Step (i) — set $CV_1 := 0$ and $CV_2 := 50$.
- Step (ii) — set $CV_m := (CV_1 + CV_2)/2$. Set $c = (M-1)$ and $\mu_{(c)} = 0$.
- Step (iii) — while $\mu_{(c)} \le T$, update $c := c + 1$ and find $\mu_{(c)}$ such that $\mu_{(c)} = \sup\big\{\mu^* > 0 \colon LLR_{\mu^*}(c) \ge CV_m\big\}$. Then, set $\mu_{(1)} = \cdots = \mu_{(M)}$.
- Step (iv) — set $\mu_{(c)} := T$. Using expression (3.9), calculate $\pi(RR = 1, CV = CV_m)$. If $|\pi(1, CV_m) - \alpha| \le \epsilon$, stop and take CV_m as the critical value solution. Otherwise, proceed to Step (v).
- Step (v) — if $\pi(RR = 1, CV = CV_m) > \alpha$, then update $CV_1 := CV_m$, otherwise, update $CV_2 := CV_m$. Go to Step (ii).

Table 1 presents the exact critical values for the maximized SPRT when requiring a minimum number of events M to signal, for $M = 1, 2, 3, 4, 6, 8, 10$.

Table 1: Exact critical values for the Poisson based maximized SPRT, when a minimum of M events is required before the null hypothesis can be rejected. T is the upper limit on the sample size (length of surveillance), expressed in terms of the expected number of events under the null. The type 1 error is $\alpha = 0.05$. When T is small and M is large, no critical value will result in $\alpha \leq 0.05$, which is denoted by '..'.

T	Minimum Number of Events Required to Reject the Null						
	$M=1$	2	3	4	6	8	10
1	2.853937	2.366638	1.774218	..	..	..	..
1.5	2.964971	2.576390	2.150707	1.683209	..	..	..
2	3.046977	2.689354	2.349679	2.000158	..	..	..
2.5	3.110419	2.777483	2.474873	2.187328	..	..	..
3	3.162106	2.849327	2.565320	2.317139	1.766485	..	..
4	3.245004	2.937410	2.699182	2.498892	2.089473	1.564636	..
5	3.297183	3.012909	2.803955	2.623668	2.267595	1.936447	..
6	3.342729	3.082099	2.873904	2.699350	2.406810	2.093835	1.740551
8	3.413782	3.170062	2.985560	2.829259	2.572627	2.337771	2.086032
10	3.467952	3.238009	3.064248	2.921561	2.690586	2.484834	2.281441
12	3.511749	3.290551	3.125253	2.993106	2.781435	2.589388	2.415402
15	3.562591	3.353265	3.199953	3.075613	2.877939	2.711996	2.556634
20	3.628123	3.430141	3.288216	3.176370	2.997792	2.846858	2.717137
25	3.676320	3.487961	3.356677	3.249634	3.081051	2.947270	2.827711
30	3.715764	3.534150	3.406715	3.307135	3.147801	3.019639	2.911222
40	3.774663	3.605056	3.485960	3.391974	3.246619	3.130495	3.030735
50	3.819903	3.657142	3.544826	3.455521	3.317955	3.210428	3.117553
60	3.855755	3.698885	3.590567	3.505220	3.374194	3.271486	3.184196
80	3.910853	3.762474	3.659939	3.580900	3.458087	3.362888	3.284030
100	3.952321	3.810141	3.711993	3.636508	3.520081	3.430065	3.355794
120	3.985577	3.847748	3.753329	3.680584	3.568679	3.482966	3.411235
150	4.025338	3.892715	3.802412	3.732386	3.626150	3.544308	3.476655
200	4.074828	3.948930	3.862762	3.796835	3.696511	3.619825	3.556799
250	4.112234	3.990901	3.908065	3.844847	3.748757	3.675703	3.615513
300	4.142134	4.024153	3.944135	3.882710	3.790143	3.719452	3.661830
400	4.188031	4.075297	3.998950	3.940563	3.852658	3.785930	3.731524
500	4.222632	4.113692	4.040021	3.983778	3.899239	3.835265	3.783126
600	4.250310	4.144317	4.072638	4.018090	3.936175	3.874183	3.823908
800	4.292829	4.191167	4.122559	4.070466	3.992272	3.933364	3.885600
1000	4.324917	4.226412	4.160022	4.109665	4.034210	3.977453	3.931529

Using the approach described above, these were calculated using the 'R Sequential' package, which can also be used for other values of 'M'. When $M = 1$, we get the standard maximized SPRT, whose previously calculated critical values [4] are included for comparison purposes. The expression for the maximum number of iterations until finding the CV solution is $\ln(1/\epsilon)/\ln(2)$. For a precision of $\epsilon = 0.00000001$, which is the precision adopted in this paper, the number of iterations is of at most $\lceil \ln(1/0.00000001)/\ln(2) \rceil = 27$. Note that these numerical calculations only have to be done once for each T and M. Hence, users do not need to do their own numerical calculations, as long as they use one of the parameter combinations presented in Table 1.

The critical values are lower for higher values of M. This is natural. Since we do not allow the null hypothesis to be rejected based on only a small number of adverse events, it allows us to be more inclined to reject the null later on when there are a larger number of events, while still maintaining the correct overall alpha level. In essence, we are trading the ability to reject the null with a very small number of events for the ability to more easily reject the null when there are a medium or large number of events. Note also that the critical values are higher for larger values of the maximum sample size T. This is also natural, as there is more multiple testing that needs to be adjusted for when T is large.

3.2. Statistical Power and Expected Time to Signal

For fixed CV, T, M, and RR, one can also calculate the statistical power using expression (3.9). The same reasoning that was applied to calculate the probability of rejecting H_0 can be used to obtain an expression for the average time to signal. Let L denote the time when the sequential analysis is interrupted to reject the null. Then the average time to signal is given by:

$$\begin{aligned}\mathbf{E}(L) &= \frac{RR^{-1}\sum_{\mathtt{N}=1}^{N}\int_{\mu_{(\mathtt{N}-1)}}^{\mu_{(\mathtt{N})}}\int_{\mu_{(\mathtt{N}-1)}}^{t_{\mathtt{N}}}\int_{\mu_{(\mathtt{N}-2)}}^{t_{\mathtt{N}-1}}\cdots\int_{\mu_{(1)}}^{t_2} RR^{\mathtt{N}+1}\,t_{\mathtt{N}}\,e^{-RR\,t_{\mathtt{N}}}\,dt_1\,dt_2\cdots dt_{\mathtt{N}}}{\pi(RR,CV)} \\ &= \frac{RR^{-1}\sum_{\mathtt{N}=1}^{N}\sum_{i=1}^{\mathtt{N}} -1^{i+1}\,\psi_i\,\Pr\big(\mu_{(\mathtt{N}-1)} \le W_{\mathtt{N}+1-i} \le \mu_{(\mathtt{N})} \,|\, RR\big)}{\pi(RR,CV)}\,,\end{aligned}$$

where $W_{\mathtt{N}} \sim Gamma(\mathtt{N}+1, RR)$, i.e., $f_W(w) = RR\,e^{-RR}(RR\,w)^{\mathtt{N}}/\mathtt{N}!$.

Table 2 presents statistical power and average time to signal for different values of M, the minimum number of events needed to signal. These are exact calculations, done for different relative risks and for different upper limits T on the length of surveillance. When T increases, power increases, since the maximum sample size increases. For fixed T, the power always increases with increasing M. This is natural, since power increases by default when there are fewer looks at the data, as there is less multiple testing to adjust for. The average time to signal may either increase or decrease with increasing values of M. For example, with $T = 20$ and a true $RR = 2$, the average time of signal is 6.96, 6.62, 6.57 and 6.96 for $M = 1, 3, 6$ and 10, respectively. For the same parameters, the statistical power is 0.921, 0.936, 0.948 and 0.957 respectively. Hence, when the true $RR = 2$ and when $T = 20$, both power and the average time to signal is better if we use $M = 3$ rather than $M = 1$. The same is true for $M = 6$ versus $M = 3$, but not for $M = 10$ versus $M = 6$.

The trade-off between statistical power and average time to signal is not easily deciphered from Table 2, and it is hence hard to judge which value of M is best. Since T, the upper limit on the length of surveillance, is the least important

metric, let's ignore that for the moment, and see what happens to the average time to signal if we keep both the alpha level and the power fixed. That will make it easier to find a good choice for M, which will depend on the true relative risk.

Table 2: Statistical power and average time to signal, when the null hypothesis is rejected, for the Poisson based maximized SPRT when a minimum of M events is required before the null hypothesis can be rejected. T is the upper limit on the sample size (length of surveillance), expressed in terms of the expected number of events under the null. The type 1 error is $\alpha = 0.05$.

T	M	Statistical Power				Average Time to Signal			
		$RR = 1.5$	2	3	4	$RR = 1.5$	2	3	4
1	1	0.107	0.185	0.379	0.573	0.30	0.35	0.39	0.39
1	3	0.129	0.234	0.466	0.665	0.59	0.58	0.55	0.51
2	1	0.130	0.255	0.561	0.799	0.63	0.75	0.79	0.73
2	3	0.157	0.315	0.645	0.857	0.92	0.94	0.89	0.78
5	1	0.190	0.447	0.876	0.987	1.82	2.09	1.78	1.22
5	3	0.224	0.507	0.905	0.991	2.10	2.17	1.73	1.17
5	6	0.255	0.559	0.928	0.994	2.71	2.58	2.05	1.54
10	1	0.280	0.685	0.989	1.000	4.02	4.13	2.45	1.35
10	3	0.321	0.733	0.993	1.000	4.25	4.07	2.31	1.30
10	6	0.358	0.770	0.995	1.000	4.71	4.25	2.50	1.61
10	10	0.391	0.803	0.996	1.000	5.67	5.03	3.40	2.50
20	1	0.450	0.921	1.000	1.000	8.68	6.96	2.67	1.41
20	3	0.492	0.936	1.000	1.000	8.65	6.62	2.53	1.37
20	6	0.531	0.948	1.000	1.000	8.92	6.57	2.69	1.65
20	10	0.562	0.957	1.000	1.000	9.47	6.96	3.50	2.51
50	1	0.803	1.000	1.000	1.000	20.45	8.94	2.82	1.48
50	3	0.829	1.000	1.000	1.000	19.82	8.45	2.71	1.45
50	6	0.847	1.000	1.000	1.000	19.41	8.24	2.86	1.71
50	10	0.863	1.000	1.000	1.000	19.35	8.46	3.59	2.52
100	1	0.978	1.000	1.000	1.000	29.93	9.30	2.92	1.53
100	3	0.982	1.000	1.000	1.000	28.52	8.87	2.82	1.51
100	6	0.985	1.000	1.000	1.000	27.58	8.71	2.97	1.75
100	10	0.987	1.000	1.000	1.000	27.04	8.93	3.65	2.53
200	1	1.000	1.000	1.000	1.000	33.00	9.62	3.01	1.58
200	3	1.000	1.000	1.000	1.000	31.47	9.25	2.93	1.56
200	6	1.000	1.000	1.000	1.000	30.47	9.11	3.07	1.78
200	10	1.000	1.000	1.000	1.000	29.88	9.33	3.71	2.54

Figure 1 shows the average time to signal as a function of statistical power, for different values of M. The lower curves are better, since the expected time to signal is shorter. Suppose we design the sequential analysis to have 95 percent power to detect a relative risk of 1.5. We can then look at the left side of Figure 1 to see the average time to signal for different true relative risks. We see that for a true relative risk of 1.5, time to signal is shortest for $M = 10$. On the other hand, for a true relative risk of 2, it is shortest for $M = 6$, for a true relative risk of 3, it is shortest for $M = 3$ and for a true relative risk of 4, it is shortest for $M = 2$.

On the right side of Figure 1, we show the expected time to signal when the surveillance has been designed to attain a certain power for a relative risk of 2. The results are similar.

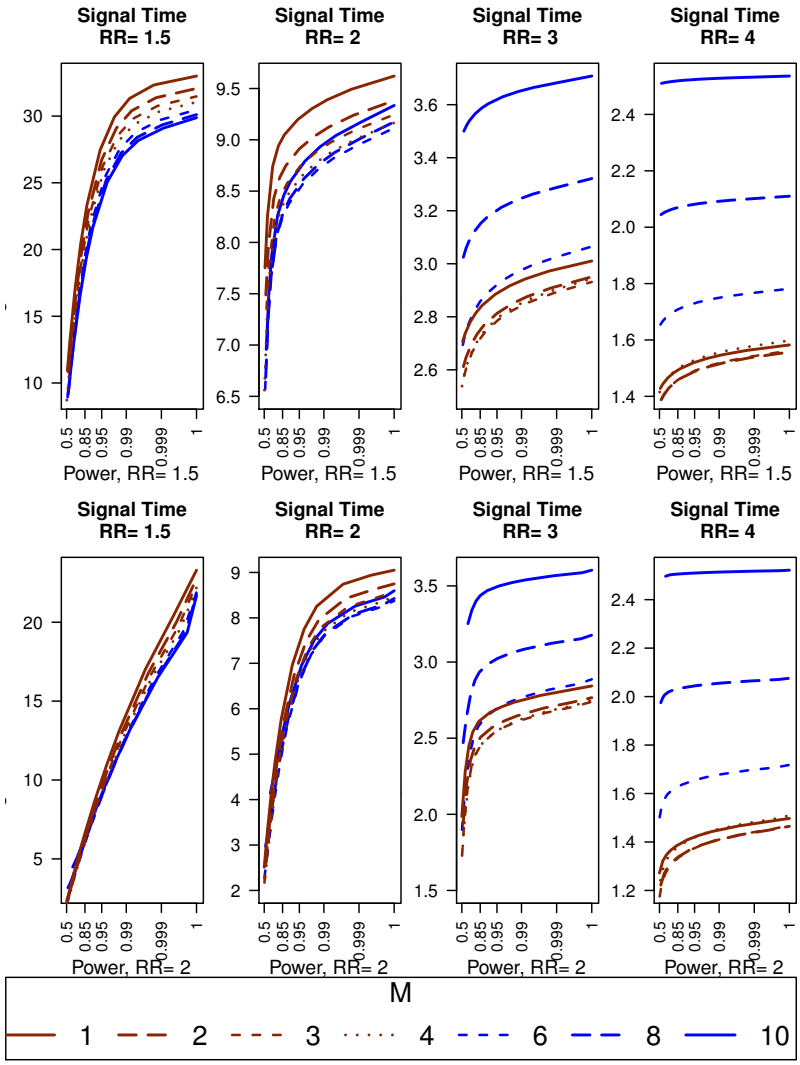

Figure 1: The average time to signal, as a function of statistical power, for the Poisson based MaxSPRT when a minimum of M events is required before the null hypothesis can be rejected. The type 1 error is $\alpha = 0.05$.

When the true relative risk is higher, it is a more serious safety problem, and hence, it is more important to detect it earlier. So, while there is no single value of M that is best overall, anywhere in the 3 to 6 range may be a reasonable choice for M. The cost of this reduced time to signal when the null is rejected is a slight delay until the surveillance ends when the null is not rejected.

4. DELAYED START OF SURVEILLANCE

For logistical or other reasons, it is not always possible to start post-marketing safety surveillance at the time that the first vaccine or drug is given. If the delay is short, one could ignore this and pretend that the sequential analyses started with the first exposed person. One could do this either by starting to calculate the test statistic at time D or by calculating it retroactively for all times before D. The former will be conservative, not maintaining the correct alpha level. The latter will maintain the correct alpha level, but, some signals will be unnecessarily delayed without a compensatory improvement in any of the other metrics. A better solution is to use critical values that take the delayed start of surveillance into account.

4.1. Exact Critical Values

In order to calculate the critical values, statistical power and average time to signal in this case, it is sufficient to replace M by M^* in the expressions of Sections 3.1 and 3.2, where $M^* := \min\{c \in \mathbb{N}\colon LLR_D(c) \geq CV\}$.

Table 3 presents exact critical values for the maximized SPRT when surveillance does not start until the expected number of events under the null hypothesis is D, without any requirement on having a minimum umber of events to signal. When $D = 0$, we get the standard maximized SPRT, whose critical values [4] are included for comparison purposes. Note that the critical values are lower for higher values of D. Since surveillance is not performed until the sample size have reached D expected counts under the null, one can afford to use a lower critical value for the remaining time while still maintaining the same overall alpha level. As before, the critical values are higher for larger values of T. When $D > T$, the surveillance would not start until after the end of surveillance, so those entries are blank in Table 3. When $D = T$, there is only one non-sequential analysis performed, so there are no critical values for a sequential test procedure. Hence, they are also left blank in the Table.

With a delayed start, there are some values of T and D for which there is no critical value that gives an alpha level of exactly 0.05. For those combinations, denoted with italics, Table 3 presents the critical value that gives the largest possible alpha less than 0.05. In Table 4, we present the exact alpha levels obtained for those scenarios, as well as the $\alpha > 0.05$ obtained for a slightly smaller liberal critical value.

Table 3: Exact critical values for the Poisson based maximized SPRT, when surveillance does not start until the sample size is large enough to generate D expected events under the null hypothesis. $T > D$ is the upper limit on the sample size. The minimum number of events needed to reject is set to $M = 1$. The type 1 error is $\alpha = 0.05$. For some values of T and D, the critical values are conservative with $\alpha < 0.05$. These are denoted in italics.

T	D						
	0	1	2	3	4	6	10
1.5	2.964971	1.683208	..	..	..	..	..
2	3.046977	2.000158	..	..	..	..	..
2.5	3.110419	2.187328	1.600544	..	..	..	..
3	3.162106	2.317139	1.766484	..	..	..	..
4	3.245004	2.498892	2.089473	1.842319	..	..	..
5	3.297183	*2.545178*	2.267595	1.936447	1.611553	..	..
6	3.342729	2.546307	2.406809	2.093835	1.921859	..	..
8	3.413782	2.694074	2.572627	2.337771	2.211199	1.829011	..
10	3.467952	2.799333	*2.591675*	2.484834	*2.298373*	2.087405	..
12	3.511749	2.880721	2.683713	2.589388	2.415402	2.254018	1.755455
15	3.562591	2.970411	2.794546	2.711996	2.556634	2.347591	*2.020681*
20	3.628123	3.082511	2.918988	*2.846635*	2.717137	2.542045	2.260811
25	3.676320	3.159490	3.011001	2.886783	2.827711	2.668487	2.432668
30	3.715764	3.223171	3.080629	2.963485	2.911222	2.765594	2.553373
40	3.774663	3.313966	3.186878	3.078748	3.030735	2.903286	2.684730
50	3.819903	3.381606	3.261665	3.162197	3.117553	2.999580	2.802863
60	3.855755	3.434748	3.320749	3.226113	*3.162908*	*3.051470*	2.890933
80	3.910853	3.515052	3.407923	3.321868	3.247872	3.151820	3.019184
100	3.952321	3.574091	3.472610	3.391377	3.321971	3.232345	3.109251
120	3.985577	3.620223	3.523446	3.445695	3.379278	3.294843	3.177847
150	4.025338	3.675035	3.583195	3.509028	3.446674	3.367227	3.238461
200	4.074828	3.742843	3.655984	3.587079	3.528662	3.454679	3.336012
250	4.112234	3.792978	3.710128	3.644349	3.588871	3.518954	3.406929
300	4.142134	3.832686	3.752749	3.689355	3.636272	3.568952	3.462111
400	4.188031	3.893093	3.785930	3.757574	3.707431	3.644405	3.544518
500	4.222632	3.938105	3.835264	3.808087	3.760123	3.700032	3.605012
600	4.250310	3.973710	3.874183	3.847892	3.801678	3.743656	3.652326
800	4.292829	4.028089	3.933363	*3.887512*	3.864597	3.809685	3.723608
1000	4.324917	*4.047191*	3.977453	3.931529	3.911308	3.858669	3.776275

The exact critical values are based on numerical calculations done in the same iterative way as for the original MaxSPRT and the version described in the previous section. The only difference is that there is an added initial step where the probabilities are calculated for different number of events at the defined start time D. Open source R functions [25] have been published as part of the R package 'Sequential' (`cran.r-project.org/web/packages/Sequential/`).

Table 4: Critical values and exact alpha levels for those combinations of T, D and M for which there does not exist a critical value for $\alpha = 0.05$. T is the upper limit on the sample size (length of surveillance), expressed in terms of the expected number of events under the null. D is the sample size at which the sequential analyses start, also expressed in terms of the expected number of events under the null. M is the minimum number of events required to signal. CV_{cons} and CV_{lib} are the conservative and liberal critical values, respectively, while α_{cons} and α_{lib} are their corresponding alpha levels.

T	D	M	CV_{cons}	α_{cons}	CV_{lib}	α_{lib}
5	1	1,4	2.545178	0.04587	2.545177	0.05323
10	2	1,4	2.591675	0.04998	2.591674	0.05478
10	4	1,4	2.298373	0.04924	2.298372	0.05379
15	10	1,4	2.020681	0.04755	2.020680	0.05124
20	3	1,4	2.846635	0.04712	2.846634	0.05001
60	4	1,4	3.162908	0.04922	3.162907	0.05094
60	6	1,4	3.051470	0.04953	3.051469	0.05101
800	3	1,4	3.887512	0.04992	3.887511	0.05091
1000	1	1,4	4.047191	0.04944	4.047190	0.05094

4.2. Statistical Power and Timeliness

For a fixed value on the upper limit on the sample size T, the statistical power of sequential analyses always increases if there are fewer looks at the data, with the maximum attained when there is only one non-sequential analysis after all the data has been collected. Hence, for fixed T, a delay in the start of surveillance always increases power, as can be seen in Table 5. For fixed T, the average time to signal almost always increases with a delayed start. The rare exception is when T is very large and the true RR is very small. For example, for $T = 100$ and $RR = 1.5$, the average time to signal is 29.9 without a delayed start, 27.2 with a delayed start of $D = 3$ and 27.0 with a delayed start of $D = 6$. With a longer delay of $D = 10$, the average time to signal increases to 27.4.

For fixed T, we saw that there is a trade-off between power and the time to signal, but in post-market safety surveillance it is usually easy and inexpensive to increase power by increasing T. Hence, the critical evaluation is to compare the average time to signal when holding both power and the alpha level fixed. This is done in Figure 2. When the study is powered for a relative risk of 2, then the average time to signal is lower when there is less of a delay in the start of the surveillance, whether the true relative risk is small or large. When the study is powered for a relative risk of 1.5, we see the same thing, except when the true relative risk is small. Hence, in terms of performance, smaller D is always better.

Table 5: Statistical power and average time to signal for the Poisson based maximized SPRT, when the analysis does not start until the sample size is large enough to correspond to D expected events under the null hypothesis. T is the upper limit on the sample size (length of surveillance), expressed in terms of the expected number of events under the null. The minimum number of events required to signal is set to $M = 1$. The type 1 error is $\alpha = 0.05$.

T	D	Power				Average Time to Signal			
		$RR = 1.5$	2	3	4	$RR = 1.5$	2	3	4
5	0	0.190	0.447	0.876	0.987	1.82	2.09	1.78	1.22
5	3	0.275	0.595	0.943	0.996	3.81	3.65	3.30	3.08
10	0	0.280	0.685	0.989	1.000	4.02	4.13	2.45	1.35
10	3	0.377	0.789	0.996	1.000	5.33	4.84	3.53	3.10
10	6	0.408	0.819	0.997	1.000	6.94	6.59	6.07	6.00
20	0	0.450	0.921	1.000	1.000	8.68	6.96	2.67	1.41
20	3	0.543	0.952	1.000	1.000	9.44	7.06	3.78	3.17
20	6	0.583	0.963	1.000	1.000	10.42	8.20	6.15	6.01
20	10	0.609	0.969	1.000	1.000	12.33	10.83	10.01	10.00
50	0	0.803	1.000	1.000	1.000	20.45	8.94	2.82	1.48
50	3	0.860	1.000	1.000	1.000	19.39	8.50	3.85	3.18
50	6	0.871	1.000	1.000	1.000	19.65	9.43	6.16	6.01
50	10	0.885	1.000	1.000	1.000	20.64	11.82	10.02	10.00
100	0	0.978	1.000	1.000	1.000	29.93	9.30	2.92	1.53
100	3	0.987	1.000	1.000	1.000	27.16	8.95	3.90	3.18
100	6	0.988	1.000	1.000	1.000	26.98	9.97	6.24	6.01
100	10	0.990	1.000	1.000	1.000	27.40	12.09	10.02	10.00
200	0	1.000	1.000	1.000	1.000	33.00	9.62	3.01	1.58
200	3	1.000	1.000	1.000	1.000	30.01	9.35	3.94	3.18
200	6	1.000	1.000	1.000	1.000	29.78	10.31	6.26	6.01
200	10	1.000	1.000	1.000	1.000	30.16	12.48	10.04	10.00

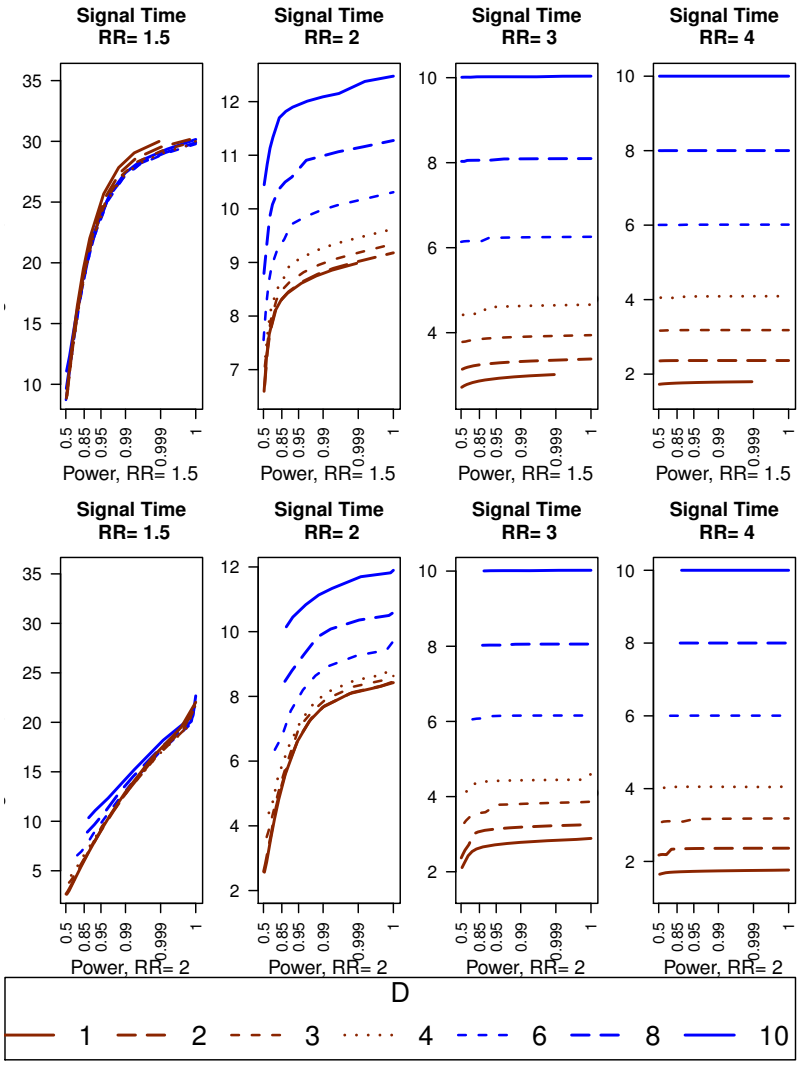

Figure 2: The average time to signal, as a function of statistical power, for the Poisson based maximized SPRT, when the analysis does not start until the sample size is large enough to correspond to D expected events under the null hypothesis. The type 1 error is $\alpha = 0.05$.

5. DISCUSSION

With the establishment of new near real-time post-market drug and safety surveillance systems [15, 26, 27, 28, 29], sequential statistical methods will become a standard feature of the pharmacovigilance landscape. In this paper we have shown that it is possible to reduce the expected time to signal when the null is rejected, without loss of statistical power, by requiring a minimum number of adverse events before generating a statistical signal. This will allow users to optimize their post-market sequential analyses.

In this paper we calculated the critical values, power and timeliness for Poisson based continuous sequential analysis with either a minimum events to signal requirement or when there is delayed start for logistical reasons. The reported numbers are based on exact numerical calculations rather than approximate asymptotic calculations or computer simulations. From a mathematical and statistical perspective, these are straight forward extensions of prior work on exact continuous sequential analysis. The importance of the results are hence from practical public health perspective rather than for any theoretical statistical advancements.

A key question is which sequential study design to use. There is not always a simple answer to that question, as the performance of the various versions depends on the true relative risk, which is unknown. One important consideration is that the early detection of an adverse event problem is more important when the relative risk is high, since more patients are affected. As a rule of thumb, it is reasonable to require a minimum of about $M = 3$ to 6 adverse events before rejecting the null hypothesis, irrespectively of whether it is a rare or common adverse event. For those who want a specific recommendation, we suggest $M = 4$.

Critical values, statistical power and average time to signal has been presented for a wide variety of parameter values. This is done so that most users will not have to perform their own calculations. For those who want to use other parameter values, critical values, power and expected time to signal can be calculated using the 'Sequential' R package that we have developed.

It is possible to combine a delayed start with $D > 0$ together with a requirement that there are at least $M > 1$ events to signal. It does not always make a difference though. For $M = 4$, the critical values are the same as for $M = 1$, for all values of $D \geq 1$. That is because with $D = 1$ or higher, one would never signal with less than three events anyhow. Since the critical values are the same, the statistical power and average time to signal are also the same. This means that when there is a non-trivial delayed start, there is not much benefit from also requiring a minimum number events to signal, but the 'Sequential' R package has a function for this dual scenario as well.

There is no reason to purposely delay the start of the surveillance until there is some minimum sample size D. In the few scenarios for which such a delay improve the performance, the improvement is not measurably better than the improvements obtained by using a minimum number of observed events. Only when it is logistically impossible to start the surveillance at the very beginning should such sequential analyses be conducted, and then it is important to do so in order to maximize power, to minimize the time to signal and to maintain the correct alpha level.

For self-controlled analyses, a binomial version of the MaxSPRT [4] is used rather than the Poisson version discussed in this paper. For concurrent matched controls, a flexible exact sequential method is used that allows for a different number of controls per exposed individuals [30]. By default, these types of continuous sequential methods will not reject the null hypothesis until there is a minimum number of events observed. To see this, consider the case with a 1:1 ratio of exposed to unexposed and and assume that the first four adverse events all are in the exposed category. Under the null hypothesis, the probability of this is $(1/2)^4 = 0.0625$, which does not give a low enough p-value to reject the null hypothesis even in a non-sequential setting. Hence, the null will never be rejected after only four adverse events, even when there is no minimum requirement. One could set the minimum number of exposed events to something higher, and that may be advantageous. If there is a delayed start for logistical reasons, then it makes sense to take that into account when calculating the critical value, for these two types of models as well.

Since the Vaccine Safety Datalink [31] launched the first near real-time post-marketing vaccine safety surveillance system in 2004 [2], continuous sequential analysis has been used for a number of vaccines and potential adverse events [1, 5, 6, 7, 8, 9, 10, 12]. The critical value tables presented in this paper has already been used by the Vaccine Safety Datalink project. As new near real-time post-market safety surveillance systems are being developed, it is important to fine-tune and optimize the performance of near-real time safety surveillance systems [15, 16, 27, 32, 33, 34]. While the improved time to signal is modest compared to the original version of the Poisson based MaxSPRT, there is no reason not to use these better designs.

ACKNOWLEDGMENTS

This work was supported by the Centers for Disease Control and Prevention through the Vaccine Safety Datalink Project, contract number 200-2002-00732 (M.K.), by National Institute of General Medical Sciences grant 1R01GM108999 (M.K.), by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (I.S.), and by the Banco de Desenvolvimento de Minas Gerais, Brazil (I.S.).

REFERENCES

[1] Yih, W.K.; Kulldorff, M.; Fireman, B.H.; Shui, I.M.; Lewis, E.M.; Klein, N.P.; Baggs, J.; Weintraub, E.S.; Belongia, E.A.; Naleway, A.; Gee, J.; Platt, R. and Lieu, T.A. (2011). Active surveillance for adverse events: The experience of the vaccine safety datalink project, *Pediatrics*, **127**(S), 54–64.

[2] Davis, R.L. (2013). Vaccine Safety Surveillance Systems: Critical Elements and Lessons Learned in the Development of the US Vaccine Safety Datalink's Rapid Cycle Analysis Capabilities, *Pharmaceutics*, **5**, 168–178.

[3] McNeil, M.M.; Gee, J.; Weintraub, E.S.; Belongia, E.A.; Lee, G.M.; Glanz, J.M.; Nordin, J.D.; Klein, N.P.; Baxter, R.; Naleway, A.L.; Jackson, L.A.; Omer, S.B.; Jacobsen, S.J. and DeStefano, F. (2014). The Vaccine Safety Datalink: successes and challenges monitoring vaccine safety, *Vaccine*, **32**, 5390–5398.

[4] Kulldorff, M.; Davis, R.L.; Kolczak, M.; Lewis, E.; Lieu, T. and Platt, R. (2011). A maximized sequential probability ratio test for drug and vaccine safety surveillance, *SA*, **3000**, 58–78.

[5] Lieu, T.A.; Kulldorff, M.; Davis, R.L.; Lewis, E.M.; Weintraub, E.; Yih, W.K.; Yin, R.; Brown, J.S. and Platt, R. (2007). Real-time vaccine safety surveillance for the early detection of adverse events, *Medical Care*, **45**, S89–95.

[6] Yih, W.K.; Nordin, J.D.; Kulldorff, M.; Lewis, E.; Lieu, T.; Shi, P. and Weintraub, E. (2009). An assessment of the safety of adolescent and adult tetanus-diphtheria-acellular pertussis (tdap) vaccine, using near real-time surveillance for adverse events in the vaccine safety datalink, *Vaccine*, **27**, 4257–4262.

[7] Belongia, E.A.; Irving, S.A.; Shui, I.M.; Kulldorff, M.; Lewis, E.; Li, R.; Lieu, T.A.; Weintraub, E.; Yih, W.K.; Yin, R.; Baggs, J. and the Vaccine Safety Datalink Investigation Group (2010). Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine-derived rotavirus vaccine, *Pediatric Infectious Disease Journal*, **29**, 1–5.

[8] Klein, N.P.; Fireman, B.; Yih, W.K.; Lewis, E.; Kulldorff, M.; Ray, P.; Baxter, R.; Hambidge, S.; Nordin, J.; Naleway, A.; Belongia, E.A.; Lieu, T.; Baggs, J. and Weintraub, E.; for the Vaccine Safety Datalink (2010). Measles–mumps–rubella–varicella combination vaccine and the risk of febrile seizures, *Pediatrics*, **126**, e1–e8.

[9] Gee, J.; Naleway, A.; Shui, I.; Baggs, J.; Yin, R.; Li, R.; Kulldorff, M.; Lewis, E.; Fireman, B.; Daley, M.F.; Klein, N.P. and Weintraub, E.S. (2011). Monitoring the safety of quadrivalent human papillomavirus vaccine: Findings from the Vaccine Safety Datalink, *Vaccine*, **29**, 8279–8284.

[10] Lee, G.M.; Greene, S.K.; Weintraub, E.S.; Baggs, J.; Kulldorff, M.; Fireman, B.H.; Baxter, R.; Jacobsen, S.J.; Irving, S.; Daley, M.F.; Yin, R.; Naleway, A.; Nordin, J.; Li, L.; McCarthy, N.; Vellozi, C.; DeStefano, F. and Lieu, T.A.; on behalf of the Vaccine Safety Datalink Project (2011). H1N1 and Seasonal Influenza in the Vaccine Safety Datalink Project, *American Journal of Preventive Medicine*, **41**, 121–128.

[11] Tseng, H.F.; Sy, L.S.; Liu, I.L.A.; Qian, L.; Marcy, S.M.; Weintraub, E.; Yih, K.; Baxter, R.; Glanz, J.; Donahue, J.; Naleway, A.; Nordin, J. and Jacobsen, S.J. (2013). Postlicensure surveillance for pre-specified adverse events following the 13-valent pneumococcal conjugate vaccine in children, *Vaccine*, **31**, 2578–2583.

[12] Weintraub, E.S.; Baggs, J.; Duffy, J.; Vellozzi, C.; Belongia, E.A.; Irving, S.; Klein, N.P.; Glanz, J.; Jacobsen, S.J.; Naleway, A.; Jackson, L.A. and DeStefano, F. (2014). Risk of intussusception after monovalent rotavirus vaccination, *New England Journal of Medicine*, **370**, 513–519.

[13] Daley, M.F.; Yih, W.K.; Glanz, J.M.; Hambidge, S.J.; Narwaney, K.J.; Yin, R.; Li, L.; Nelson, J.C.; Nordin, J.D.; Klein, N.P.; Jacobsen, S.J. and Weintraub, E. (2014). Safety of diphtheria, tetanus, acellular pertussis and inactivated poliovirus (DTaP-IPV) vaccine, *Vaccine*, **32**, 3019–3024.

[14] Avery, T.; Kulldorff, M.; Vilk, W.; Li, L.; Cheetham, C.; Dublin, S.; Hsu, J.; Davis, R.L.; Liu, L.; Herrinton, L.; Platt, R. and Brown, J.S. (2013). Near real-time adverse drug reaction surveillance within population-based health networks: methodology considerations for data accrual, *Pharmacoepidemiology and Drug Safety*, page epub.

[15] Fireman, B.; Toh, S.; Butler, M.G.; Go, A.S.; Joffe, H.V.; Graham, D.J.; Nelson, J.C.; Daniel, G.W. and Selby, J.V. (2012). A protocol for active surveillance of acute myocardial infarction in association with the use of a new antidiabetic pharmaceutical agent, *Pharmacoepidemiology and Drug Safety*, **21**, 282–290.

[16] Suling, M. and Pigeot, I. (2012). Signal Detection and Monitoring Based on Longitudinal Healthcare Data, *Pharmaceutics*, **4**, 607–640.

[17] Gagne, J.J.; Rassen, J.A.; Walker, A.M.; Glynn, R.J. and Schneeweiss, S. (2012). Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules, *Epidemiology*, **3**, 238–246.

[18] Gagne, J.J.; Wang, S.V.; Rassen, J.A. and Schneeweiss, S. (2014). A modular, prospective, semi-automated drug safety monitoring system for use in a distributed data environment, *Pharmacoepidemiology and Drug Safety*, **23**, 619–627.

[19] Wald, A. (1945). Sequential tests of statistical hypotheses, *Annals of Mathematical Statistics*, **16**, 117–186.

[20] Brown, J.S.; Kulldorff, M.; Chan, K.A.; Davis, R.L.; Graham, D.; Pettus, P.T.; Andrade, S.E.; Raebel, M.; Herrinton, L.; Roblin, D.; Boudreau, D.; Smith, D.; Gurwitz, J.H.; Gunter, M.J. and Platt, R. (2007). Early detection of adverse drug events within population-based health networks: Application of sequential testing methods, *Pharmacoepidemiology and Drug Safety*, **16**, 1275–1284.

[21] Wald, A. (1947). *Sequential Analysis*, Wiley.

[22] Weiss, L. (1953). Testing one simple hypothesis against another, *Annals of Mathematical Statistics*, **24**, 273–281.

[23] Lai, T.L. (1991). *Asymptotic optimality of generalized sequential likelihood ratio tests in some classical sequential testing procedures*, "Handbook of Sequential Analysis" (B.K. Ghosh and P.K. Sen, Eds.), pp. 121–144, Dekker, New York.

[24] Siegmund, D. and Gregory, P. (1980). A sequential clinical trial for testing $p_1 = p_2$, *Annals of Statistics*, **8**, 1219–1228.

[25] R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009, ISBN 3-900051-07-0.

[26] Burwen, D.R.; Sandhu, S.K.; MaCurdy, T.E.; Kelman, J.A.; Gibbs, J.M.; Garcia, B.; Marakatou, M.; Forshee, R.A.; Izurieta, H.S. and Ball, R. (2012). Surveillance for Guillain–Barré syndrome after influenza vaccination among the Medicare population, 2009–2010, *American Journal of Public Health*, **102**, 1921–1927.

[27] Hauser, R.G.; Mugglin, A.S.; Friedman, P.A.; Kramer, D.B.; Kallinen, L.; McGriff, D. and Hayes, D.L. (2012). Early detection of an underperforming implantable cardiovascular device using an automated safety surveillance tool, *Circulation: Cardiovascular Quality and Outcomes*, **1**, 89–196.

[28] Huang, W.T.; Chen, W.W.; Yang, H.W.; Chen, W.C.; Chao, Y.N.; Huang, Y.W.; Chuang, J.H. and Kuo, H.S. (2010). Design of a robust infrastructure to monitor the safety of the pandemic A (H1N1) 2009 vaccination program in Taiwan, *Vaccine*, **28**, 7161–7166.

[29] Nguyen, M.; Ball, R.; Midthun, K.; Lieu, T.A. (2012). The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise, *Pharmacoepidemiology and Drug Safety*, **21**, 291–297.

[30] Fireman B., *et al.* (2012). Exact sequential analysis for binomial data with time-varying probabilities, *Manuscript in Preparation.*

[31] Chen, R.T.; Glasser, J.E.; Rhodes, P.H.; Davis, R.L.; Barlow, W.E.; Thompson, R.S.; Mullooloy, J.P.; Black, S.B.; Shinefield, H.R.; Vadheim, C.M.; Marcy, S.M.; Ward, J.I.; Wise, R.P.; Wassilak, S.G.; Hadler, S.C.; and the Vaccine Safety Datalink Team (1997). Vaccine safety datalink project: A new tool for improving vaccine safety monitoring in the united states, *Pediatrics*, **99**, 765–773.

[32] Gaalen, R.D.; Abrahamowicz, M. and Buckeridge, D.L. (2014). The impact of exposure model misspecification on signal detection in prospective pharmacovigilance, *Pharmacoepidemiology and Drug Safety*, **24**, 456–467.

[33] Li, R.; Stewart, B.; Weintraub, E. and McNeil, M.M. (2014). Continuous sequential boundaries for vaccine safety surveillance, *Statistics in Medicine*, **33**, 3387–3397.

[34] Maro, J.C.; Brown, J.S.; Dal Pan, G.J. and Kulldorff, M. (2014). Minimizing signal detection time in postmarket sequential analysis: balancing positive predictive value and sensitivity, *Pharmacoepidemiology and Drug Safety*, **23**, 839–848.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 395–423

SOME INADMISSIBILITY RESULTS FOR ESTIMATING QUANTILE VECTOR OF SEVERAL EXPONENTIAL POPULATIONS WITH A COMMON LOCATION PARAMETER

Authors: Manas Ranjan Tripathy
– Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, India
manas@nitrkl.ac.in

Somesh Kumar
– Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur-721302, India
smsh@maths.iitkgp.ernet.in

Received: July 2014 Revised: February 2015 Accepted: September 2015

Abstract:

- Suppose independent random samples are taken from $k\,(\geq 2)$ exponential populations with a common and unknown location parameter "μ" and possibly different unknown scale parameters $\sigma_1, \sigma_2, ..., \sigma_k$ respectively. The estimation of $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$; where θ_i is the quantile of the i^{th} population, has been considered with respect to either a sum of squared error loss functions or sum of quadratic losses. Estimators based on maximum likelihood estimators (MLEs) and uniformly minimum variance unbiased estimators (UMVUEs) for each component θ_i have been obtained. An admissible class of estimators has been obtained. Improvement over an estimator based on UMVUEs is obtained by an application of the Brewster–Zidek technique. Further, classes of equivariant estimators are derived under affine and location groups of transformations and some inadmissibility results are proved. Finally, a numerical comparison of risk performance of all proposed estimators has been done and the recommendations are made for the use of these estimators.

Key-Words:

- *common location; complete class; equivariant estimators; inadmissibility; MLE; quantile estimation; risk comparison; simultaneous estimation; UMVUE.*

AMS Subject Classification:

- 62F10, 62C15, 62C07.

1. INTRODUCTION

Let $(X_{i1}, X_{i2}, ..., X_{in_i})$; $i = 1, 2$ be independent random samples taken from two exponential populations with a common unknown location parameter μ and possibly different scale parameters σ_1, σ_2 respectively. The probability density function of the random variable X_{ij} is given by

$$f(x_{ij} \,|\, \mu, \sigma_i) = \frac{1}{\sigma_i} \exp\Big\{-\Big(\frac{x_{ij} - \mu}{\sigma_i}\Big)\Big\}, \qquad x_{ij} > \mu, \;\; -\infty < \mu < \infty, \;\; \sigma_i > 0\,; \\ j = 1, 2, ..., n_i; \;\; i = 1, 2\,.$$

The p^{th} quantile of the i^{th} population is denoted by $\theta_i = \mu + \eta\sigma_i$, where $\eta = -\log(1-p) > 0$; $0 < p < 1$. We are interested in estimating the quantile vector $\underline{\theta} = (\theta_1, \theta_2)$. The loss function is taken to be either the sum of the squared errors

$$L_1(\underline{\alpha}, \underline{d}) = \sum_{i=1}^{2} (d_i - \theta_i)^2 \tag{1.1}$$

or, the sum of the quadratic losses

$$L_2(\underline{\alpha}, \underline{d}) = \sum_{i=1}^{2} \left(\frac{d_i - \theta_i}{\sigma_i}\right)^2, \tag{1.2}$$

where $\underline{\alpha} = (\mu, \sigma_1, \sigma_2)$ and $\underline{d} = (d_1, d_2)$ be an estimate of $\underline{\theta}$.

When parameters of same nature are thought to be equal, it is then customary to pool samples for inference purposes on that common parameter. This is also known as meta-analysis, and has received considerable attention from the researchers lately. For example, the problem of estimation of a common mean of two or more normal populations has been extensively studied by several authors in the recent past. The problem is popularly known as common mean problem and arises in the study of recovery of inter-block information in balanced incomplete block designs (BIBDs). For a complete bibliography and some recent results on estimation of a common mean of several normal populations one may refer to Pal and Sinha [15], Kumar [10], Mitra and Sinha [12], Pal *et al.* [13] and Tripathy and Kumar [20] and the references cited therein.

The problem of estimating a common location parameter μ of several exponential populations when the scale parameters are unknown has been studied by several authors in the recent past. The parameter μ is also referred to as the "minimum guarantee time" in the study of reliability. This problem was probably first considered by Ghosh and Razmpour [5]. They have obtained the maximum likelihood estimator (MLE), a modified maximum likelihood estimator (MMLE) and the uniformly minimum variance unbiased estimator (UMVUE). They have also compared numerically the risk values of all these estimators with respect

to the squared error loss function whereas the MLE and the MMLE have been compared asymptotically in terms of their bias and mean squared errors (MSEs). Pal and Sinha [14] considered this problem from a decision theoretic point of view. They proposed a class of improved estimators which are better than the MLE in terms of MSE as well as Pitman measure of closeness (PMC). However, these improved estimators are different from the MMLE and the UMVUE. Jin and Pal [9] obtained a wide class of estimators which dominate the MLE under a class of convex loss functions. Jin and Crouse [7] proposed a larger class of estimators for μ which includes the MMLE and the UMVUE for special choices of their constants (see Equation (3.1) in [7]). They obtained estimators which dominate the MLE using a class of convex loss functions.

For this particular model, the problem of estimation of quantiles is important and also interesting for its practical applications. Quantiles of exponential populations are very much useful in the study of reliability, life testing, and survival analysis. For some applications of quantiles of exponential populations we refer to Epstein and Sobel [3] and Saleh [18]. Estimation of quantiles $\theta_1 = \mu + \eta\sigma_1$, of an exponential population was probably first considered by Rukhin and Strawderman [17] using a decision theoretic approach. They proved that the best affine equivariant estimator (BAEE) for the quantile θ_1 is inadmissible when either $0 \leq \eta < \frac{1}{n}$ or $\eta > 1 + \frac{1}{n}$ where $n \geq 2$ is the sample size. Rukhin [16] proved its admissibility when $\frac{1}{n} \leq \eta \leq 1 + \frac{1}{n}$. He also obtained a class of minimax estimators for $\eta > 1 + \frac{1}{n}$. This class contains some generalized Bayes estimators. One of these generalized Bayes estimators is shown to be admissible within a class of scale equivariant estimators.

For the model studied in this paper, Sharma and Kumar [19] and Kumar and Sharma [11] considered estimation of the quantiles $\theta_1 = \mu + \eta\sigma_1$ of the first population when the other $k-1$ $(k \geq 2)$ populations are available. They show that the MLE, the UMVUE and the BAEE based on the first sample alone can be improved by using other $k-1$ samples. They have also obtained a general inadmissibility result for the class of affine equivariant estimators for $0 \leq \eta < \frac{1}{n}$. Jin and Crouse [8] considered the problem of estimating the quantile $\theta_i = \mu + \eta\sigma_i$ of the i^{th} population. They established an identity for the exponential distributions, and using it, compared the risk functions of the UMVUE and the MLE. They also proposed a class of estimators which dominate the MLE and the UMVUE.

It is interesting to note that all the above work relates to estimating either the common location parameter μ or a component θ_i of the vector $\underline{\theta}$ of quantiles. From a theoretical as well as a practical viewpoint, it is important to consider the problem of simultaneous estimation of $\underline{\theta}$. For example, suppose an electronic item is produced by several manufacturers and lifetimes of these follow exponential distributions. It is very likely that the average lives of items from different manufacturers will be different due to quality specifications used by them.

However, due to competition in the market, they will maintain a common minimum guarantee time. Then the problem of simultaneous estimation of average lives is a special case of the problem of simultaneous estimation of the vector of quantiles. One may refer to Ghosh and Auer [4], Berger [1] and Gupta [6] for some results on the simultaneous estimation of parameters.

In this paper, we consider the general problem of estimating the vector of quantiles of several exponential populations with a common location but different scale parameters. In Section 2, some basic estimators of the quantile vector are proposed based on the MLE, the MMLE and the UMVUE of each component. In Section 3, we consider classes of affine and location equivariant estimators and prove some inadmissibility results. In Section 4, we extend some of these results to $k(\geq 2)$ exponential populations. A detailed numerical comparison of risk values for several proposed estimators has been done by using Monte-Carlo simulations in Section 5. Also recommendations are made for using these estimators. Certain proofs have been given in the Appendix.

2. SOME BASIC RESULTS & IMPROVEMENT OVER UMVUE

In this section we derive some baseline estimators for the quantile vector $\underline{\theta}$ and obtain an estimator which dominates the UMVUE using a result of Brewster and Zidek (Brewster and Zidek [2]).

2.1. Some Basic Estimators

Suppose $(X_{i1}, X_{i2}, ..., X_{in_i})$; $i = 1, 2$ are independent random samples taken from two exponential populations $Ex(\mu, \sigma_1)$ and $Ex(\mu, \sigma_2)$ having the probability density functions,

$$f(x_{ij}) = \frac{1}{\sigma_i} \exp\Big\{-\Big(\frac{x_{ij} - \mu}{\sigma_i}\Big)\Big\}, \qquad x_{ij} > \mu,\ -\infty < \mu < \infty,\ \sigma_i > 0\,,$$
$$j = 1, 2, ..., n_i;\ \ i = 1, 2\,,$$

respectively. We are interested in estimating the quantile vector $\underline{\theta} = (\theta_1, \theta_2)$, where $\theta_i = \mu + \eta\sigma_i$ denotes the quantile of the i^{th} population, $i = 1, 2$. The loss function is taken to be either the sum of the squared errors (1.1) or the sum of the quadratic losses (1.2).

Let us denote $X_i = \min(X_{i1}, X_{i2}, ..., X_{in_i})$ and $Y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$; $i = 1, 2$. Further define $Z = \min(X_1, X_2)$, $T_1 = Y_1 - Z$, and $T_2 = Y_2 - Z$. Here Y_1 and Y_2 are the means of the first and the second samples respectively. Then (Z, T_1, T_2)

is a complete sufficient statistic. The random variables Z and $\underline{T} = (T_1, T_2)$ are independently distributed. The probability density function of Z is given by

$$f_Z(z) = a \exp\big(-a(z-\mu)\big), \qquad z > \mu, \quad -\infty < \mu < \infty,$$

where $a = \frac{n_1}{\sigma_1} + \frac{n_2}{\sigma_2}$. The joint probability density function of T_1 and T_2 can be obtained from Ghosh and Razmpour [5] by using a simple transformation, and is given by

$$f_{\underline{T}}(\underline{t}) = \frac{n_1^{n_1} n_2^{n_2}}{\sigma_1^{n_1} \sigma_2^{n_2} a} \left[\frac{t_1^{n_1-1} t_2^{n_2-2}}{\Gamma n_1 \, \Gamma(n_2-1)} + \frac{t_1^{n_1-2} t_2^{n_2-1}}{\Gamma n_2 \, \Gamma(n_1-1)} \right] \exp(-n_1 t_1/\sigma_1 - n_2 t_2/\sigma_2), \quad t_1, t_2 > 0.$$

The MLE of μ and σ_i are $\hat{\mu} = Z$, and $\hat{\sigma}_i = T_i$, $i = 1, 2$ respectively. Thus collecting the MLEs for each component we obtain the estimator for quantile vector $\underline{\theta}$ as

$$\underline{\delta}_{ML} = \big(Z + \eta T_1, Z + \eta T_2\big),$$

and we call it the MLE of $\underline{\theta}$. Further noticing $E(Z) = \mu + a^{-1}$, the MLE $\underline{\delta}_{ML}$ of $\underline{\theta}$ can be modified and we call this a modified MLE for the vector $\underline{\theta}$ and is given by

$$\underline{\delta}_{MM} = \big(Z - \hat{a}^{-1} + \eta T_1, Z - \hat{a}^{-1} + \eta T_2\big),$$

where $\hat{a} = \frac{n_1}{T_1} + \frac{n_2}{T_2}$.

Next we collect the UMVUEs of θ_i for each component and form an estimator for the quantile vector $\underline{\theta}$. It is easy to see that $E(T_j) = \sigma_j - a^{-1}$; $j = 1, 2$ and $E\big[\big(\sum_{i=1}^{2}(n_i - 1)\, T_i^{-1}\big)^{-1}\big] = a^{-1}$. Using these results and the fact that $(Z, \underline{T})$ is a complete sufficient statistic, we get the uniformly minimum variance unbiased estimator for each component θ_i as $Z + (\eta - 1)T^* + \eta T_i$, where we denote $T^* = \big(\sum_{i=1}^{2}(n_i - 1)\, T_i^{-1}\big)^{-1}$. Now collecting the UMVUEs for each component θ_i, we form an estimator for the quantile vector $\underline{\theta}$, denoted as $\underline{\delta}_{MV}$ and is given by

$$\underline{\delta}_{MV} = \Big(Z + \eta T_1 + (\eta - 1)\, T^*,\, Z + \eta T_2 + (\eta - 1)\, T^*\Big).$$

The expressions for the risk functions of $\underline{\delta}_{ML}$, $\underline{\delta}_{MM}$ and $\underline{\delta}_{MV}$ with respect to the loss (1.2) are obtained as follows:

$$R(\underline{\delta}_{ML}, \underline{\theta}) = \eta^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) + \frac{2\,(1-\eta)\,(1+\tau^2)}{(n_1 + n_2 \tau)^2},$$

$$R(\underline{\delta}_{MM}, \underline{\theta}) = \eta^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) + \frac{2\,(1-\eta)\,(1+\tau^2)}{(n_1 + n_2 \tau)^2} + \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \left(\frac{2\,(\eta - 1)}{a}\, ES + ES^2 \right),$$

$$R(\underline{\delta}_{MV}, \underline{\theta}) = \eta^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) + \frac{4\,(1-\eta)\,(1+\tau^2)}{(n_1 + n_2 \tau)^2} + 2\,(\eta - 1)^2\, ET^{*2},$$

where $S = 1/\hat{a}$ and $\tau = \sigma_1/\sigma_2 > 0$.

In the rest of the paper, when we say UMVUE for the quantile vector $\underline{\theta}$, we mean "the collection of the UMVUEs for each component θ_i and form a vector" to get the estimator for the quantile vector $\underline{\theta}$.

2.2. An Estimator Dominating the UMVUE

In this section, we propose an estimator for the quantile vector $\underline{\theta}$, which improves upon the UMVUE for the quantile vector $\underline{\theta}$, with respect to the loss function (1.1). Let us consider a class of estimators for the quantile vector $\underline{\theta}$ as

$$D = \left\{\underline{\delta}_{\mathbf{c}} \colon \underline{\delta}_{\mathbf{c}} = (\delta_{c_1}, \delta_{c_2});\ c_1, c_2 \in \mathbb{R}\right\},$$

where we denote $\delta_{c_j} = Z + \eta c_j T_j + (\eta - 1)T^*;\ j = 1, 2$.

Now for the class of estimators $D = \left\{\underline{\delta}_{\mathbf{c}} \colon \underline{\delta}_{\mathbf{c}} = (\delta_{c_1}, \delta_{c_2});\ c_1, c_2 \in \mathbb{R}\right\}$, let us define

$$\mathbf{c}^* = \Big(\min\big\{\max(c_1, a_1), b_1\big\},\ \min\big\{\max(c_2, a_2), b_2\big\}\Big) \tag{2.1}$$

and

$$\mathbf{c}_* = \Big(\min\big\{\max(c_1, c_1^+), d_1\big\},\ \min\big\{\max(c_2, c_2^+), d_2\big\}\Big) \tag{2.2}$$

where $a_j = \frac{\eta n_j(n_j-2)+1}{\eta n_j(n_j-1)}$, $b_j = \frac{n_j}{n_j+1}$, $d_j = \max\{a_j, b_j\}$, $c_j^+ = \hat{c}_j(\lambda_j^+)$, and $\lambda_j^+ = \left\{(n_j+1) - \sqrt{(n_j+1)^2 - 4\eta n_j}\right\}/2n_j;\ j = 1, 2$. Next we have the following inadmissibility result for estimators in the class D.

Theorem 2.1. *Let D be the class of estimators for the quantile vector $\underline{\theta}$, and define the functions $\mathbf{c}^*$ and $\mathbf{c}_*$ as in (2.1) and (2.2) respectively. Let the loss function be* (1.1).

- (**i**) *The estimator $\underline{\delta}_{\mathbf{c}}$ is inadmissible and is improved by $\underline{\delta}_{\mathbf{c}^*}$ if $\mathbf{c} \neq \mathbf{c}^*$, when $\eta \geq 1$.*
- (**ii**) *The estimator $\underline{\delta}_{\mathbf{c}}$ is inadmissible and is improved by $\underline{\delta}_{\mathbf{c}_*}$ if $\mathbf{c} \neq \mathbf{c}_*$ when $0 < \eta < 1$.*

Proof: See Appendix.

Corollary 2.1.

- (**i**) *Let $\eta \geq 1$. The class of estimators $\left\{\underline{\delta}_{\mathbf{c}} = (\delta_{c_1}, \delta_{c_2}) \colon a_j \leq c_j \leq b_j;\ j = 1, 2\right\}$ is essentially complete in D.*
- (**ii**) *Let $0 < \eta < 1$. The class of estimators $\left\{\underline{\delta}_{\mathbf{c}} = (\delta_{c_1}, \delta_{c_2}) \colon c_j^+ \leq c_j \leq d_j;\ j = 1, 2\right\}$ is essentially complete in D.*

The class of estimators D also contains the UMVUE for the quantile vector $\underline{\theta}$ when $c_1 = c_2 = 1$. Consequently, the UMVUE $\underline{\delta}_{MV}$ is inadmissible. The result we write as a theorem which is immediate. Let $p_1 = \min\{\frac{1}{n_1}, \frac{1}{n_2}\}$, $p_2 = \max\{\frac{1}{n_1}, \frac{1}{n_2}\}$, $q_1 = \min\{\frac{n_1+1}{2n_1}, \frac{n_2+1}{2n_2}\}$ and $q_2 = \max\{\frac{n_1+1}{2n_1}, \frac{n_2+1}{2n_2}\}$.

Theorem 2.2. *Let the loss function be (1.1).*

- (**i**) *If $\eta \geq 1$, then the uniformly minimum variance unbiased estimator $\underline{\delta}_{MV}$ for the quantile vector $\underline{\theta}$ is inadmissible and is improved by the estimator $\underline{\delta}_{IMV} = (\delta_{b_1}, \delta_{b_2})$. Further the class $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \delta_{c_2})\colon a_j \leq c_j \leq b_j;\ j = 1, 2\}$ is essentially complete in D.*
- (**ii**) *If $q_2 \leq \eta < 1$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by $\underline{\delta}_{IMV} = (\delta_{b_1}, \delta_{b_2})$. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \delta_{c_2})\colon c_j^+ \leq c_j \leq b_j;\ j = 1, 2\}$ is essentially complete in D.*
- (**iii**) *If $p_2 \leq \eta \leq q_1$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by $\underline{\delta}_{IMV} = (\delta_{a_1}, \delta_{a_2})$. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \delta_{c_2})\colon c_j^+ \leq c_j \leq a_j;\ j = 1, 2\}$ is essentially complete in D.*
- (**iv**) *If $0 \leq \eta < p_1$, then the estimator $\underline{\delta}_{MV}$ is admissible in the class D. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \delta_{c_2})\colon c_j^+ \leq c_j \leq a_j;\ j = 1, 2\}$ is essentially complete in D.*
- (**v**) *Let $p_1 < \eta < p_2$. If $\frac{1}{n_1} < \eta < \frac{1}{n_2}$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by either $(\delta_{a_1}, \delta_{c_2})$ (when $\eta < \frac{n_1+1}{2n_1}$) or $(\delta_{b_1}, \delta_{c_2})$ (when $\eta \geq \frac{n_1+1}{2n_1}$) where $c_2 = 1$. If $\frac{1}{n_2} < \eta < \frac{1}{n_1}$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by either $(\delta_{c_1}, \delta_{a_2})$ (when $\eta < \frac{n_2+1}{2n_2}$) or $(\delta_{c_1}, \delta_{b_2})$ (when $\eta \geq \frac{n_2+1}{2n_2}$) where $c_1 = 1$.*
- (**vi**) *Let $q_1 < \eta < q_2$. If $\frac{n_1+1}{2n_1} < \eta < \frac{n_2+1}{2n_2}$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by either $(\delta_{b_1}, \delta_{a_2})$ (when $\eta > \frac{1}{n_2}$) or $(\delta_{b_1}, \delta_{c_2})$ (when $\eta < \frac{1}{n_2}$) where $c_2 = 1$. If $\frac{n_2+1}{2n_2} < \eta < \frac{n_1+1}{2n_1}$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by either $(\delta_{a_1}, \delta_{b_2})$ (when $\eta > \frac{1}{n_1}$) or $(\delta_{c_1}, \delta_{b_2})$ (when $\eta < \frac{1}{n_1}$) where $c_1 = 1$.*

Proof: The proof is immediate as an application of Theorem 2.1. □

Applying the above Theorem 2.2 it is easy to write the estimator which improves upon $\underline{\delta}_{MV}$. However, we give the expression only for the case $\eta \geq q_2$ and $p_2 < \eta \leq q_1$ below. The expressions for other cases can be written easily:

$$\underline{\delta}_{IMV} = \begin{cases} \Big(Z + \eta b_1 T_1 + (\eta - 1)T^*,\ Z + \eta b_2 T_2 + (\eta - 1)T^*\Big), & \text{if } \eta \geq q_2\,, \\ \Big(Z + \eta a_1 T_1 + (\eta - 1)T^*,\ Z + \eta a_2 T_2 + (\eta - 1)T^*\Big), & \text{if } p_2 < \eta \leq q_1\,. \end{cases}$$

3. INADMISSIBILITY OF EQUIVARIANT ESTIMATORS FOR QUANTILES

In this section, we consider affine and location class of equivariant estimators for the quantile vector $\underline{\theta}$. We derive sufficient conditions for improving estimators in these classes and as a consequence we prove some complete class results.

3.1. Affine Equivariant Estimators

Let us consider the affine group of transformations, $G_A = \{g_{a,b}\colon\ g_{a,b}(x) = ax + b,\ a > 0,\ b \in \mathbb{R}\}$. Under the transformation $g_{a,b}$, we have $X_{ij} \to aX_{ij} + b$, $Z \to aZ + b$, $T_i \to aT_i$, $\sigma_i \to a\sigma_i$, $\mu \to a\mu + b$, and $\theta_i = \mu + \eta\sigma_i \to a\theta_i + b$; $i = 1, 2$. So $\underline{\theta} = (\theta_1, \theta_2) \to a\underline{\theta} + b\underline{e}$, where $\underline{e} = (1, 1)$. The estimation problem is invariant if we take the loss function as the sum of the affine invariant loss functions (1.2). The invariance loss condition is

$$\begin{aligned} L\big(\bar{g}_{a,b}(\underline{\alpha}), \tilde{d}\big) &= \sum_{i=1}^{2}\left(\frac{a\,\theta_i + b - \tilde{d}_i}{a\,\sigma_i}\right)^2 \\ &= L(\underline{\alpha}, \underline{d})\ , \end{aligned}$$

which is satisfied if $\tilde{d}_i = ad_i + b = \tilde{g}_{a,b}(d_i),\ i = 1, 2$. Here $\underline{\alpha} = (\mu, \sigma_1, \sigma_2)$. Therefore an affine equivariant estimator satisfies

$$\underline{\delta}(aZ + b, aT_1, aT_2) = a\underline{\delta}(Z, T_1, T_2) + b\underline{e}\ .$$

Substituting $b = -aZ$ where $a = 1/T_1$, we get

$$\underline{\delta}\left(0, 1, \frac{T_2}{T_1}\right) = \frac{1}{T_1}\left[\underline{\delta}(Z, T_1, T_2) - Z\underline{e}\right]\ .$$

From the above relation, we get the form of an affine equivariant estimator as

$$\begin{aligned} \underline{\delta}(Z, T_1, T_2) &= Z\underline{e} + T_1\underline{\Psi}(W) \\ &= \underline{\delta}_{\underline{\Psi}}\ , \quad \text{say}\,, \end{aligned} \tag{3.1}$$

where $W = \frac{T_2}{T_1}$. To proceed further we denote $\eta_1 = \frac{\eta n_1 - 1}{n_1 + n_2}$, and $\eta_2 = \frac{n_2 w}{n_1 + n_2}\big(\eta - \frac{1}{n_2}\big)$.

Let us define the following functions:

$$\Psi_1^* = \begin{cases} \eta_1\,, & \text{if } 0 \le w \le \frac{1}{1 - \eta n_1}\ , \\ \hat{\Psi}_1(\tau^+, w), & \text{if } w > \frac{1}{1 - \eta n_1}\ , \end{cases} \tag{3.2}$$

where $\tau^+ = -\frac{n_1}{n_2} + \frac{1}{n_2}\sqrt{\frac{n_1(w-1)}{\eta w}}$, and

$$\Psi_2^* = \begin{cases} \eta_2\,, & \text{if } w \geq 1-\eta n_2\,, \\ \hat{\Psi}_2(\alpha^+, w)\,, & \text{if } w < 1-\eta n_2\,, \end{cases} \tag{3.3}$$

where $\alpha^+ = \frac{\eta n_1}{1-\eta n_2 - w} + \frac{n_1\sqrt{\eta n_2(1-w)}}{n_2(1-\eta n_2-w)}$.

Next, for the affine equivariant estimator $\underline{\delta}_{\underline{\Psi}}$ define functions $\underline{\Psi}_0$, $\underline{\Psi}^0$, $\underline{\Psi}_{11}$ and $\underline{\Psi}_{22}$ as follows:

$$\underline{\Psi}_0 = \big(\max(\Psi_1, \Psi_1^*), \max(\Psi_2, \Psi_2^*)\big)\,, \tag{3.4}$$

$$\underline{\Psi}^0 = \big(\max(\Psi_1, \eta_1), \max(\Psi_2, \eta_2)\big)\,, \tag{3.5}$$

$$\underline{\Psi}_{11} = \big(\max(\Psi_1, \eta_1), \max(\Psi_2, \Psi_2^*)\big)\,, \tag{3.6}$$

$$\underline{\Psi}_{22} = \big(\max(\Psi_1, \Psi_1^*), \max(\Psi_2, \eta_2)\big)\,. \tag{3.7}$$

Let p_1 and p_2 be defined as in Section 2. Next we prove an inadmissibility result for estimators which are equivariant under the affine group of transformations.

Theorem 3.1. *Let the loss function be (1.2) and the functions $\underline{\Psi}_0$, $\underline{\Psi}^0$, $\underline{\Psi}_{11}$ and $\underline{\Psi}_{11}$ be defined as in (3.4), (3.5), (3.6) and (3.7) respectively.*

- (**i**) *The estimator $\underline{\delta}_{\underline{\Psi}}$ is inadmissible and is improved by $\underline{\delta}_{\underline{\Psi}_0}$ if there exist some values of parameters $\underline{\alpha}$ such that $P(\underline{\delta}_{\underline{\Psi}} \neq \underline{\delta}_{\underline{\Psi}_0}) > 0$ when $0 < \eta < p_1$.*
- (**ii**) *The estimator $\underline{\delta}_{\underline{\Psi}}$ is inadmissible and is improved by $\underline{\delta}_{\underline{\Psi}^0}$ if there exist some values of parameters $\underline{\alpha}$ such that $P(\underline{\delta}_{\underline{\Psi}} \neq \underline{\delta}_{\underline{\Psi}^0}) > 0$ when $\eta \geq p_2$.*
- (**iii**) *Let $p_1 \leq \eta < p_2$. If $\frac{1}{n_1} \leq \frac{1}{n_2}$, then the estimator $\underline{\delta}_{\underline{\Psi}}$ is inadmissible and is improved by $\underline{\delta}_{\underline{\Psi}_{11}}$ if there exist some values of parameters $\underline{\alpha}$ such that, $P(\underline{\delta}_{\underline{\Psi}} \neq \underline{\delta}_{\underline{\Psi}_{11}}) > 0$. If $\frac{1}{n_2} \leq \frac{1}{n_1}$, then the estimator $\underline{\delta}_{\underline{\Psi}}$ is inadmissible and is improved by $\underline{\delta}_{\underline{\Psi}_{22}}$ if there exist some values of parameters $\underline{\alpha}$ such that, $P(\underline{\delta}_{\underline{\Psi}} \neq \underline{\delta}_{\underline{\Psi}_{22}}) > 0$.*

Proof: For proof see Appendix.

Remark 3.1. The Theorem 3.1 is basically a complete class theorem for affine equivariant estimators. It says that any affine equivariant estimator of the form (3.1) will be inadmissible if $P\{(\Psi_1 < \Psi_1^*) \cup (\Psi_2 < \Psi_2^*)\} > 0$ when $\eta < p_1$ and $P\{(\Psi_1 < \eta_1) \cup (\Psi_2 < \eta_2))\} > 0$ for $\eta \geq p_2$. A similar type of statement holds for

the case $p_1 \leq \eta \leq p_2$. However, for small values of η and for small sample sizes the improvements over the MLE and the MMLE are very marginal and we omit the risk values in the tables. For $\eta > p_2$, improvement over these is not possible by using the result of Theorem 3.1. Improvement over $\underline{\delta}_{MV}$ has been shown in the Tables 1–3 for $0 < \eta < p_1$.

Remark 3.2. The results of the Theorem 3.1 will remain valid, if instead of the loss function (1.2), we use any sum of the weighted squared error loss functions.

3.2. Location Equivariant Estimator

Let us introduce the location group of transformations, $G_L = \{g_a \colon g_a(x) = x + a,\ a \in \mathbb{R}\}$ to our model. Under the transformation g_a, $X_{ij} \to X_{ij} + a$, $X_i \to X_i + a$, $Z \to Z + a$, $T_i \to T_i$, $\sigma_i \to \sigma_i$, $\mu \to \mu + a$, and $\underline{\theta} = (\theta_1, \theta_2) \to (\theta_1 + a, \theta_2 + a) = \underline{\theta} + a\underline{e}$, where $\theta_i = \mu + \eta\sigma_i$; $i = 1, 2$.

The estimation problem will be invariant if we choose the loss function as the sum of the squared error loss functions (1.1). The location equivariant estimator $\underline{\delta}$ must satisfy the relation

$$\underline{\delta}(Z + a, T_1, T_2) \;=\; a\underline{e} + \underline{\delta}(Z, T_1, T_2)\,.$$

Substituting $a = -Z$, we get

$$\underline{\delta}(0, T_1, T_2) \;=\; \underline{\delta}(Z, T_1, T_2) - Z\underline{e}\,.$$

From this relation we get the form of a location equivariant estimator as

$$\begin{aligned} \underline{\delta}(Z, T_1, T_2) \;&=\; Z\underline{e} + \underline{\psi}(T_1, T_2) \\ &=\; \underline{\delta}_{\underline{\psi}}\,, \qquad \text{say}\,, \end{aligned} \tag{3.8}$$

where $\underline{\psi}(T_1, T_2) = (\psi_1(T_1, T_2), \psi_2(T_1, T_2))$.

For the location equivariant estimator $\underline{\delta}_{\underline{\psi}} = (\delta_{\psi_1}, \delta_{\psi_2})$ let us define functions, $\underline{\psi}^0$, $\underline{\psi}_{11}$ and $\underline{\psi}_{22}$ as

$$\underline{\psi}^0 \;=\; \big(\max(0, \psi_1), \max(0, \psi_2)\big)\,, \tag{3.9}$$

$$\underline{\psi}_{11} \;=\; \big(\max(0, \psi_1), \psi_2\big)\,, \tag{3.10}$$

and

$$\underline{\psi}_{22} \;=\; \big(\psi_1, \max(0, \psi_2)\big)\,. \tag{3.11}$$

Next we prove an inadmissibility result for estimators which are invariant under the location group of transformations.

Theorem 3.2. *Let the loss function be (1.1) and the functions* $\underline{\psi}^0$, $\underline{\psi}_{11}$ *and* $\underline{\psi}_{22}$ *be defined as in (3.9), (3.10) and (3.11) respectively.*

(**i**) *When* $\eta \geq p_2$ *the estimator* $\underline{\delta}_{\underline{\psi}}$ *is inadmissible and is improved by* $\underline{\delta}_{\underline{\psi}^0}$ *if there exist some values of the parameters* $\underline{\alpha}$ *such that* $P_{\underline{\alpha}}(\underline{\delta}_{\underline{\psi}} \neq \underline{\delta}_{\underline{\psi}^0}) > 0$.

(**ii**) *Let* $p_1 \leq \eta < p_2$. *If* $\frac{1}{n_1} \leq \eta < \frac{1}{n_2}$, *then the estimator* $\underline{\delta}_{\underline{\psi}}$ *is inadmissible and is improved by* $\underline{\delta}_{\underline{\psi}_{11}}$ *if there exist some values of parameters* $\underline{\alpha}$ *such that* $P_{\underline{\alpha}}(\underline{\delta}_{\underline{\psi}} \neq \underline{\delta}_{\underline{\psi}_{11}}) > 0$. *If* $\frac{1}{n_2} \leq \eta < \frac{1}{n_1}$ *the estimator* $\underline{\delta}_{\underline{\psi}}$ *is inadmissible and is improved by* $\underline{\delta}_{\underline{\psi}_{22}}$ *if there exist some values of parameters* $\underline{\alpha}$ *such that* $P_{\underline{\alpha}}(\underline{\delta}_{\underline{\psi}} \neq \underline{\delta}_{\underline{\psi}_{22}}) > 0$.

(**iii**) *For* $\eta < p_1$ *the class of estimators (3.8) is an essentially complete class. The estimator* $\underline{\delta}_{\underline{\psi}}$ *can not be improved by using Theorem 3.2.*

Proof: The proof is similar to the arguments used in proving the Theorem 3.1. □

Remark 3.3. The above Theorem 3.2 is also a complete class result. Basically it says that any location equivariant estimator for the quantile vector $\underline{\theta} = (\theta_1, \theta_2)$ of the form (3.8) is inadmissible if $P_{\underline{\alpha}}\{(\psi_1 < 0) \cup (\psi_2 < 0)\} > 0$ for $\eta \geq \max(\frac{1}{n_1}, \frac{1}{n_2})$. A similar type of statement holds for the case $p_1 \leq \eta \leq p_2$.

Remark 3.4. It can be further noticed that all the estimators considered such as $\underline{\delta}_{ML}$, $\underline{\delta}_{MM}$ and $\underline{\delta}_{MV}$ belong to the class of estimators obtained in (3.6), with choices of $\underline{\psi} = (\psi_1, \psi_2)$ as $(\eta T_1, \eta T_2)$, $\big(\eta T_1 - \frac{T_1 T_2}{n_2 T_1 + n_1 T_2}, \eta T_2 - \frac{T_1 T_2}{n_2 T_1 + n_1 T_2}\big)$ and $\big(\eta T_1 + \frac{(\eta-1) T_1 T_2}{(n_1-1) T_2 + (n_2-1) T_1}, \eta T_2 + \frac{(\eta-1) T_1 T_2}{(n_1-1) T_2 + (n_2-1) T_1}\big)$ respectively. But none of these can be improved by using the result of Theorem 3.2 as the values of ψ_1 and ψ_2 fall within the interval $[0, +\infty)$ when $\eta \geq \max\big(\frac{1}{n_1}, \frac{1}{n_2}\big)$ with probability 1. However, an example where our result will be useful is as follows: suppose we consider an estimator for $\underline{\theta}$ as $\underline{\delta} = (Z - \eta T_1, Z - \eta T_2)$ or any estimator of the form $(Z - g_1(T_1, T_2), Z - g_2(T_1, T_2))$, with $g_1(t_1, t_2) > 0$, or $g_2(t_1, t_2) > 0$ and $\eta \geq \max\big(\frac{1}{n_1}, \frac{1}{n_2}\big)$. Certainly, these estimators fall in the class (3.8) with $\psi_1 < 0$ or $\psi_2 < 0$. The improved estimator for these are obtained as $\underline{\delta}^* = (Z, Z)$.

Example 3.1. An example of a practical situation where the model of this paper is applicable is presented here. Suppose μ is the common minimum guaranteed time in years of two brands of electronics products say brand A and

brand B. It is most likely that the mean residual life times (σ_1 and σ_2) will be different. On the basis of random samples of sizes 10 from brand A and B, the following summary data has been recorded. Here $Z = 7.82$, $T_1 = 12.49$ and $T_2 = 15.44$. Suppose $\eta = 3.0$, then the estimators for the quantile vector are obtained as $\underline{\delta}_{ML} = (45.31, 54.14)$, $\underline{\delta}_{MM} = (44.62, 53.45)$, $\underline{\delta}_{MV} = (46.85, 55.68)$, and $\underline{\delta}_{IMV} = (43.44, 51.47)$. In this situation, the estimator $\underline{\delta}_{IMV} = (43.44, 51.47)$ is recommended for use.

Example 3.2 (Simulated Data)**.** The following two data sets A and B of sizes each 10 and 12 has been generated from two exponential populations for illustration purpose. We have taken $\mu = 5.0$, $\sigma_1 = 5$ and $\sigma_2 = 10$. The sample values have been written up to 3 decimal places only:

A: 16.555, 9.685, 11.863, 11.248, 6.894, 20.933, 6.435, 8.573, 18.745, 9.036,

B: 5.455, 6.806, 10.667, 13.687, 11.739, 9.006, 7.612, 18.846, 23.978, 21.418, 10.639, 13.061.

Here $Z = 5.455$, $T_1 = 6.541$ and $T_2 = 7.287$. Suppose, $\eta = 0.001$, then the estimators for the quantile vector $\underline{\theta}$ are obtained as, $\underline{\delta}_{ML} = (5.462, 5.462)$, $\underline{\delta}_{MM} = (5.147, 5.147)$, $\underline{\delta}_{MV} = (5.115, 5.116)$, and $\underline{\delta}^a_{MV} = (5.145, 5.128)$, where $\underline{\delta}^a_{MV}$ denotes the improved version of $\underline{\delta}_{MV}$ obtained by using Theorem 3.1. In this situation we recommend to use the estimator $\underline{\delta}^a_{MV}$.

4. A GENERALIZATION

In this section we extend some of the results obtained in Sections 2 and 3 to the $k(\geq 2)$ exponential populations and obtain the improved estimators for the UMVUE $\underline{\delta}_{MV}$.

Specifically, let $X_{i1}, X_{i2}, ..., X_{in_i}$ be a random sample of size n_i taken from the i^{th} exponential population $Ex(\mu, \sigma_i)$. The random variable X_{ij} has probability density function,

$$f(x_{ij}) = \frac{1}{\sigma_i} \exp\Big\{-\Big(\frac{x_{ij} - \mu}{\sigma_i}\Big)\Big\}, \qquad x_{ij} > \mu,\ -\infty < \mu < \infty,\ \sigma_i > 0\,, \quad j = 1, 2, ..., n_i;\ i = 1, 2, ..., k\,.$$

We estimate the quantile vector $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$; where $\theta_i = \mu + \eta\sigma_i$ be the quantile of the i^{th} population, with respect to the loss function either

$$L_1(\underline{\alpha}, \underline{d}) = \sum_{i=1}^{k} (d_i - \theta_i)^2\,, \tag{4.1}$$

or

$$L_2(\underline{\alpha}, \underline{d}) = \sum_{i=1}^{k} \left(\frac{d_i - \theta_i}{\sigma_i} \right)^2 . \tag{4.2}$$

Let us denote $X_i = \min(X_{i1}, X_{i2}, ..., X_{in_i})$ and $Y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$. Further define $Z = \min(X_1, X_2, ..., X_k)$, and $T_i = Y_i - Z$; $i = 1, 2, ..., k$. Here Y_i is the sample mean from the i^{th} population. Then $(Z, T_1, T_2, ..., T_k)$ is a complete sufficient statistic. Now the random variables Z and $\underline{T} = (T_1, T_2, ..., T_k)$ are independently distributed. The probability density function of Z is given by

$$f_Z(z) = a \exp\big(-a(z-\mu)\big) , \quad z > \mu , \ -\infty < \mu < \infty ,$$

where $a = \sum_{i=1}^{k} \frac{n_i}{\sigma_i}$. The joint probability density function of $\underline{T} = (T_1, T_2, ..., T_k)$ is given by

$$f_{\underline{T}}(\underline{t}) = \frac{1}{a} \prod_{i=1}^{k} \left(\frac{n_i^{n_i}}{\sigma_i^{n_i}} \right) \prod_{i=1}^{k} \left(\frac{t_i^{n_i - 1}}{\Gamma n_i} \right) \left[\sum_{i=1}^{k} \frac{n_i - 1}{t_i} \right] \exp\left\{ -\sum_{i=1}^{k} n_i t_i / \sigma_i \right\} , \quad t_i > 0 .$$

It should be noted that the MLE $\underline{\delta}_{ML}$, modification to the MLE $\underline{\delta}_{MM}$ and the UMVUE $\underline{\delta}_{MV}$ can easily be obtained as

$$\underline{\delta}_{ML} = Z\underline{e} + \eta \underline{T} ,$$

$$\underline{\delta}_{MM} = (Z - \hat{a}^{-1})\underline{e} + \eta \underline{T}$$

and

$$\underline{\delta}_{MV} = \big(Z + (\eta - 1)\, T^*\big)\, \underline{e} + \eta \underline{T} ,$$

where $\underline{e} = (1, 1, ..., 1)_{1 \times k}$ and $T^* = \big(\sum_{i=1}^{k} (n_i - 1)\, T_i^{-1} \big)^{-1}$.

Consider the class of estimators for the quantile vector $\underline{\theta}$ as

$$D_{\mathbf{c}} = \left\{ \underline{\delta}_{\mathbf{c}} \colon \underline{\delta}_{\mathbf{c}} = (\delta_{c_1}, \delta_{c_2}, ..., \delta_{c_k}); \ c_i \in \mathbb{R} \right\} ,$$

where we denote $\mathbf{c} = (c_1, c_2, ..., c_k)$, and

$$\delta_{c_j} = Z + \eta\, c_j T_j + (\eta - 1) T^* ; \quad j = 1, 2, ..., k .$$

It should be noted that, this class contains the estimator $\underline{\delta}_{MV}$ for $c_1 = c_2 = \cdots = c_k = 1$.

Now for the class of estimators $D_{\mathbf{c}}$ define

$$\mathbf{c}^* = \Big(\min\big\{ \max(c_1, a_1), b_1 \big\}, \ ..., \ \min\big\{ \max(c_k, a_k), b_k \big\} \Big) \tag{4.3}$$

and

$$\mathbf{c}_* = \Big(\min\big\{\max(c_1, c_1^+), d_1\big\}, \ldots, \min\big\{\max(c_k, c_k^+), d_k\big\}\Big), \tag{4.4}$$

where $a_j = \frac{\eta n_j(n_j-2)+1}{\eta n_j(n_j-1)}$, $b_j = \frac{n_j}{n_j+1}$, $d_j = \max\{a_j, b_j\}$, $c_j^+ = \hat{c}_j(\lambda_j^+)$, and $\lambda_j^+ = \big\{(n_j+1) - \sqrt{(n_j+1)^2 - 4\eta n_j}\big\}/2n_j$; $j = 1, 2, \ldots, k$. Next we have the following inadmissibility result for estimators in the class $D_{\mathbf{c}}$.

Theorem 4.1. *Let $D_{\boldsymbol{c}}$ be the class of estimators for the quantile vector $\underline{\theta}$, and define the functions $\mathbf{c}^*$ and $\mathbf{c}_*$ as in (4.3) and (4.4) respectively. Let the loss function be* (4.1).

(**i**) *The estimator $\underline{\delta}_{\boldsymbol{c}}$ is inadmissible and is improved by $\underline{\delta}_{\boldsymbol{c}^*}$ if $\boldsymbol{c} \neq \mathbf{c}^*$, when $\eta \geq 1$.*

(**ii**) *The estimator $\underline{\delta}_{\boldsymbol{c}}$ is inadmissible and is improved by $\underline{\delta}_{\boldsymbol{c}_*}$ if $\boldsymbol{c} \neq \mathbf{c}_*$ when $0 < \eta < 1$.*

Proof: The proof is similar to the proof of the Theorem 2.1. □

The class of estimators $D_{\mathbf{c}}$ contains the UMVUE of the quantile vector $\underline{\theta}$ when $c_i = 1$; $i = 1, 2, \ldots, k$. Consequently, the UMVUE $\underline{\delta}_{MV}$ is inadmissible. Let $p_1 = \min\big\{\frac{1}{n_1}, \ldots, \frac{1}{n_k}\big\}$, $p_2 = \max\big\{\frac{1}{n_1}, \ldots, \frac{1}{n_k}\big\}$, $q_1 = \min\big\{\frac{n_1+1}{2n_1}, \ldots, \frac{n_k+1}{2n_k}\big\}$ and $q_2 = \max\big\{\frac{n_1+1}{2n_1}, \ldots, \frac{n_k+1}{2n_k}\big\}$.

Theorem 4.2. *Let the loss function be (4.1).*

(**i**) *If $\eta \geq 1$, then the uniformly minimum variance unbiased estimator $\underline{\delta}_{MV}$ for the quantile vector $\underline{\theta}$ is inadmissible and is improved by the estimator $\underline{\delta}_{IMV} = (\delta_{b_1}, \ldots, \delta_{b_k})$. Further the class $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \ldots, \delta_{c_k})$: $a_j \leq c_j \leq b_j$; $j = 1, 2, \ldots, k\}$ is essentially complete in $D_{\boldsymbol{c}}$.*

(**ii**) *If $q_2 \leq \eta < 1$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by $\underline{\delta}_{IMV} = (\delta_{b_1}, \ldots, \delta_{b_k})$. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \ldots, \delta_{c_k})$: $c_j^+ \leq c_j \leq b_j$; $j = 1, 2, \ldots, k\}$ is essentially complete in $D_{\boldsymbol{c}}$.*

(**iii**) *If $p_2 \leq \eta \leq q_1$, then the estimator $\underline{\delta}_{MV}$ is inadmissible and is improved by $\underline{\delta}_{IMV} = (\delta_{a_1}, \ldots, \delta_{a_k})$. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \ldots, \delta_{c_k})$: $c_j^+ \leq c_j \leq a_j$; $j = 1, 2, \ldots, k\}$ is essentially complete in $D_{\boldsymbol{c}}$.*

(**iv**) *If $0 \leq \eta < p_1$, then the estimator $\underline{\delta}_{MV}$ is admissible in the class D. The class of estimators $\{\underline{\delta}_{\boldsymbol{c}} = (\delta_{c_1}, \ldots, \delta_{c_k})$: $c_j^+ \leq c_j \leq a_j$; $j = 1, 2, \ldots, k\}$ is essentially complete in $D_{\boldsymbol{c}}$.*

(**v**) *Let $p_1 < \eta < p_2$ and $(l_1, l_2, \ldots, l_k)$ be a permutation of $(1, 2, \ldots, k)$ such that $1/n_{l_1} < \eta$, ..., $1/n_{l_p} < \eta$, and $1/n_{l_{p+1}} \geq \eta$, ..., $1/n_{l_k} \geq \eta$.*

Then the estimator $\underline{\delta}_{MV}$ *is inadmissible and is improved by* $\underline{\delta}_{IMV} = (\delta_{a_{l_1}}, ..., \delta_{a_{l_p}}, \delta_{c_{l_{p+1}}}, ..., \delta_{c_{l_k}})$, *when* $\eta < \frac{n_{l_1}+1}{2n_{l_1}}, ..., \eta < \frac{n_{l_p}+1}{2n_{l_p}}$, *and improved by* $\underline{\delta}_{IMV} = (\delta_{b_{l_1}}, ..., \delta_{b_{l_p}}, \delta_{c_{l_{p+1}}}, ..., \delta_{c_{l_k}})$, *when* $\eta \geq \frac{n_{l_1}+1}{2n_{l_1}}, ...,$ $\eta \geq \frac{n_{l_p}+1}{2n_{p+1}}$ *where* $c_{l_{p+1}} = c_{l_{p+2}} = \cdots = c_{l_k} = 1$.

(**vi**) *Let* $p_1 < \eta < p_2$ *and* $(l_1, l_2, ..., l_k)$ *be a permutation of* $(1, 2, ..., k)$ *such that* $1/n_{l_1} \geq \eta$, ..., $1/n_{l_p} \geq \eta$, *and* $1/n_{l_{p+1}} < \eta$, ..., $1/n_{l_k} < \eta$. *Then the estimator* $\underline{\delta}_{MV}$ *is inadmissible and is improved by* $\underline{\delta}_{IMV} = (\delta_{c_{l_1}}, ..., \delta_{c_{l_p}}, \delta_{a_{l_{p+1}}}, ..., \delta_{a_{l_k}})$, *when* $\eta < \frac{n_{l_{p+1}}+1}{2n_{l_{p+1}}}$, ..., $\eta < \frac{n_{l_k}+1}{2n_{l_k}}$, *and improved by* $\underline{\delta}_{IMV} = (\delta_{c_{l_1}}, ..., \delta_{c_{l_p}}, \delta_{b_{l_{p+1}}}, ..., \delta_{b_{l_k}})$, *when* $\eta \geq \frac{n_{l_{p+1}}+1}{2n_{l_{p+1}}}$, ..., $\eta \geq \frac{n_{l_k}+1}{2n_k}$, *where* $c_{l_1} = c_{l_2} = \cdots = c_{l_p} = 1$.

(**vii**) *Let* $q_1 < \eta < q_2$ *and* $(l_1, l_2, ..., l_k)$ *be a permutation of* $(1, 2, ..., k)$ *such that* $\frac{n_{l_1}+1}{2n_{l_1}} < \eta$, $\frac{n_{l_2}+1}{2n_{l_2}} < \eta$, ..., $\frac{n_{l_p}+1}{2n_{l_p}} < \eta$ *and* $\frac{n_{l_{p+1}}+1}{2n_{l_{p+1}}} \geq \eta$, $\frac{n_{l_{p+2}}+1}{2n_{l_{p+2}}} \geq \eta$, ..., $\frac{n_{l_k}+1}{2n_{l_k}} \geq \eta$. *The estimator* $\underline{\delta}_{MV}$ *is inadmissible and is improved by* $\underline{\delta}_{IMV} = (\delta_{b_{l_1}}, ..., \delta_{b_{l_p}}, \delta_{a_{l_{p+1}}}, ..., \delta_{a_{l_k}})$, *when* $\eta > 1/n_{l_{p+1}}$, $\eta > 1/n_{l_{p+2}}$, ..., $\eta > 1/n_{l_k}$ *and by* $\underline{\delta}_{IMV} = (\delta_{b_{l_1}}, ..., \delta_{b_{l_p}}, \delta_{c_{l_{p+1}}}, ..., \delta_{c_{l_k}})$, *when* $\eta \leq 1/n_{l_{p+1}}$, $\eta \leq 1/n_{l_{p+2}}$, ..., $\eta \leq 1/n_{l_k}$, *where* $c_{l_{p+1}} = c_{l_{p+2}} = \cdots = c_{l_k} = 1$.

(**viii**) *Let* $q_1 < \eta < q_2$ *and* $(l_1, l_2, ..., l_k)$ *be a permutation of* $(1, 2, ..., k)$ *such that* $\frac{n_{l_1}+1}{2n_{l_1}} \geq \eta$, $\frac{n_{l_2}+1}{2n_{l_2}} \geq \eta$, ..., $\frac{n_{l_p}+1}{2n_{l_p}} \geq \eta$ *and* $\frac{n_{l_{p+1}}+1}{2n_{l_{p+1}}} < \eta$, $\frac{n_{l_{p+2}}+1}{2n_{l_{p+2}}} < \eta$, ..., $\frac{n_{l_k}+1}{2n_{l_k}} < \eta$. *The estimator* $\underline{\delta}_{MV}$ *is inadmissible and is improved by* $\underline{\delta}_{IMV} = (\delta_{a_{l_1}}, ..., \delta_{a_{l_p}}, \delta_{b_{l_{p+1}}}, ..., \delta_{b_{l_k}})$, *when* $\eta > 1/n_{l_1}$, $\eta > 1/n_{l_2}$, ..., $\eta > 1/n_{l_p}$ *and by* $\underline{\delta}_{IMV} = (\delta_{c_{l_1}}, ..., \delta_{c_{l_p}}, \delta_{b_{l_{p+1}}}, ..., \delta_{b_{l_k}})$, *when* $\eta \leq 1/n_{l_1}$, $\eta \leq 1/n_{l_2}$, ..., $\eta \leq 1/n_{l_p}$, *where* $c_{l_1} = c_{l_2} = \cdots = c_{l_p} = 1$.

Applying the above Theorem 4.2 we can obtain the estimator which improves upon $\underline{\delta}_{MV}$. However, we have obtained the expressions for some specific values of η. One can easily write the estimator for other choices of η:

$$\underline{\delta}_{IMV} = \begin{cases} \big(Z + (\eta - 1)T^*\big)\,\underline{e} + \eta \underline{B}\,, & \text{if } \eta \geq q_2\,, \\ \big(Z + (\eta - 1)T^*\big)\,\underline{e} + \eta \underline{A}\,, & \text{if } p_2 < \eta \leq q_1\,, \end{cases}$$

where $\underline{A} = (A_1, A_2, ..., A_k)$; $A_i = a_i T_i$ and $\underline{B} = (B_1, B_2, ..., B_k)$; $B_i = b_i T_i$; $i = 1, 2, ..., k$.

Next we generalize the results obtained in Theorem 3.1 and Theorem 3.2. Let us consider the affine group of transformations, $G_A = \{g_{a,b} \colon g_{a,b}(x) = ax + b, a > 0, b \in \mathbb{R}$. Under the transformation $g_{a,b}$, we have $Z \to aZ + b$, $T_i \to aT_i$, $\sigma_i \to a\sigma_i$, $\mu \to a\mu + b$, and $\theta_i = \mu + \eta\sigma_i \to a\theta_i + b$; $i = 1, 2, ..., k$. So $\underline{\theta} \to a\underline{\theta} + b\underline{e}$, where $\underline{e} = (1, 1, ..., 1)_{1\times k}$. Under this transformation the problem remains invariant if

we choose the loss function (4.2), and the form of an affine equivariant estimator is obtained as

$$\begin{aligned} \underline{\delta}(Z, T_1, T_2, ..., T_k) &= Z\underline{e} + T_1\underline{\Psi}(\underline{W}) \\ &= \underline{\delta}_{\underline{\Psi}}\,, \quad \text{say}\,, \end{aligned} \tag{4.5}$$

where $\underline{W} = (W_2, W_3, ..., W_k)$ and $W_i = \frac{T_i}{T_1}$; $i = 2, 3, ..., k$.

Consider the conditional risk function:

$$R(\underline{\delta}_{\underline{\Psi}}, \underline{\alpha}\,|\,\underline{W}) = \sum_{i=1}^{k} E\left\{ \left(\frac{Z + T_1\Psi_i(\underline{W}) - \theta_i}{\sigma_i} \right)^2 \Big|\ \underline{W} \right\}. \tag{4.6}$$

It is easy to observe that the above conditional risk is a convex function in each Ψ_i and hence the sum. The minimizing choices for each Ψ_i is obtained as

$$\hat{\Psi}_i = -\frac{E(Z-\theta_i)\ E(T_1\,|\,\underline{W})}{E(T_1^2\,|\,\underline{W})}\,; \quad i = 1, 2, ..., k\,. \tag{4.7}$$

After evaluating the conditional expectations and simplifying we have the minimizing choice of Ψ_i as

$$\hat{\Psi}_i = \frac{1}{\sum_{j=1}^{k} n_j} \left[\eta\sigma_i - a^{-1}\right] \left[\frac{n_1}{\sigma_1} + \sum_{j=2}^{k} \frac{n_j\, w_j}{\sigma_j} \right]; \quad i = 1, 2, ..., k\,. \tag{4.8}$$

To apply the Brewster and Zidek technique we need to find the supremum and infimum of each $\hat{\Psi}_i$ with respect to $\underline{\sigma} = (\sigma_1, ..., \sigma_k)$ for fixed values of η, n_i and $\underline{W}$. We are not able to obtain the supremum and infimum for each $\hat{\Psi}_i$ for the case $k\,(\geq 3)$. However, for the first component $\hat{\Psi}_1$, Sharma and Kumar [19] obtained the bounds for equal sample sizes. We feel that the lower bounds for other components will be finite. Since we are not able to derive the bounds for the case $k\,(\geq 3)$, it could not be possible to provide the inadmissibility result for $k\,(\geq 3)$ populations. It will be interesting to obtain the bounds for the case $k\,(\geq 3)$ and obtain improved estimators better than $\underline{\delta}_{MV}$.

5. NUMERICAL COMPARISONS

In this section, we carry out a detailed simulation study to numerically compare the risk functions of various estimators proposed in previous sections for the quantile vector $\underline{\theta}$ for the case $k = 2$. Specifically, we have proposed some baseline estimators such as $\underline{\delta}_{ML}$, $\underline{\delta}_{MM}$ and $\underline{\delta}_{MV}$ for $\underline{\theta}$. An improved estimator $\underline{\delta}_{IMV}$ which dominates $\underline{\delta}_{MV}$ has been obtained in Section 2 for the case $\eta \geq p_1$. From the Remark 3.1, it is quite evident that, we only consider the estimator $\underline{\delta}_{MV}$ and obtain its improved version by using the Theorem 3.1, which we denote as

$\underline{\delta}^a_{MV}$ for the case $0 < \eta < p_2$. For numerically comparing the risk functions of all these estimators for $\underline{\theta}$, we use Monte-Carlo simulation procedure. We have generated 10,000 random samples each from two exponential populations $Exp(\mu, \sigma_1)$ and $E(\mu, \sigma_2)$ respectively. Here μ is the common location parameter and σ_1, σ_2 are different scale parameters. The loss function is taken as the sum of the quadratic losses (1.2). It should be noted that, with respect to the loss (1.2), the risk functions of each estimator is a function of only $\tau = \sigma_1/\sigma_2 > 0$ for fixed values of η and sample sizes. A massive simulation study has been carried out to see the behavior of the risk functions and the performance of each estimator for the quantile vector $\underline{\theta}$. The error of the simulation has been checked and it is quite satisfactory (up to order of 10^{-3}). We have also calculated the percentage of relative risk performances for each estimator with respect to the baseline estimator $\underline{\delta}_{ML}$. For this purpose we define the equation,

$$R_{MM} = \left(\frac{\underline{\delta}_{ML} - \underline{\delta}_{MM}}{\underline{\delta}_{ML}}\right) \times 100\ , \qquad R_{MV} = \left(\frac{\underline{\delta}_{ML} - \underline{\delta}_{MV}}{\underline{\delta}_{ML}}\right) \times 100\ ,$$

$$R_{MVA} = \left(\frac{\underline{\delta}_{ML} - \underline{\delta}^a_{MV}}{\underline{\delta}_{ML}}\right) \times 100\ , \qquad R_{IMV} = \left(\frac{\underline{\delta}_{ML} - \underline{\delta}_{IMV}}{\underline{\delta}_{ML}}\right) \times 100\ .$$

For illustration purpose, we choose some specific values of η and n_1, n_2. Though the values of τ can be from 0 to ∞, we choose the values up to 5 to avoid simulation error. The percentage of relative risk improvements of all the estimators over the MLE has been tabulated in Tables 1 to 3. In Table 1, we have tabulated the percentage of relative risk values for equal sample sizes whereas Tables 2, 3 gives for unequal sample sizes. In each table, the first row gives the various choices of η. We have taken conveniently the values of η as 0.05 and 2.50. The first column represents the values of τ which ranges from 0 to 5. Further, for each value of η, there corresponds three columns (columns 1, 2, 3 correspond to $\eta = 0.05$ and columns 4, 5, 6 correspond to $\eta = 2.50$). For each value of τ, there corresponds three values of percentage of relative risk values. These three values corresponds to three different pairs of sample sizes, for example in Table 1, the percentage of relative risk values have been tabulated for the sample sizes $(5, 5)$, $(10, 10)$ and $(15, 15)$. Similarly in Tables 2 and 3, the percentage of relative risk performances have been tabulated for the sample sizes $(3, 7)$, $(5, 10)$, $(10, 15)$ and $(7, 3)$, $(10, 5)$, $(15, 10)$ respectively.

The following conclusions can be drawn from our simulation study as well as from the Tables 1, 2 and 3.

- (**i**) It is observed that as the sample sizes (n_1 and n_2) increase the risk values decrease for fixed value of η.
- (**ii**) For $0 < \eta \leq p_1$, the estimator $\underline{\delta}^a_{MV}$ has the least risk for almost all values of the parameters except few values where the estimator $\underline{\delta}_{MM}$ performs marginally better. The percentage of relative risk improvement has been noticed and is near 50%.

(**iii**) For $\eta > p_2$, the estimator $\underline{\delta}_{IMV}$ performs the best and the percentage of relative risk improvement is near 21%. However, the performance decreases as the sample sizes increase.

(**iv**) When η lies in the interval $[p_1, p_2]$, the estimators $\underline{\delta}_{IMV}$ and $\underline{\delta}^a_{MV}$ compete well with each other. In fact for small values of τ, the estimator $\underline{\delta}_{IMV}$ performs better compared to $\underline{\delta}^a_{MV}$ whereas for larger values of τ, the estimator $\underline{\delta}^a_{MV}$ performs better. However, the estimator $\underline{\delta}_{MM}$ has the best percentage of relative risk improvement for this choice of η.

(**v**) For $\eta = 1$, (the problem reduces to simultaneous estimation of means of two exponential populations) the estimators $\underline{\delta}_{ML}$ and $\underline{\delta}_{MV}$ are equal and it is also noticed that the performance of $\underline{\delta}_{IMV}$ is the best.

(**vi**) The numerical study also shows that the estimator $\underline{\delta}^a_{MV}$ improves upon $\underline{\delta}_{MV}$, which agrees with the Theorem 3.1. Further the estimator $\underline{\delta}_{MV}$ is improved by $\underline{\delta}_{IMV}$ which also agrees with the Theorem 2.1.

(**vii**) Similar type of observations were made for other combinations of η and sample sizes during our simulation study.

(**viii**) On the basis of our simulation study and theoretical findings, we recommend using the estimator $\underline{\delta}^a_{MV}$ when $\eta < p_1$ and $\underline{\delta}_{IMV}$ when $\eta \geq p_2$, whereas we recommend to use $\underline{\delta}_{MM}$ for η lying in the interval $[p_1, p_2]$.

6. CONCLUDING REMARKS

In this paper we have considered the estimation of the quantile vector $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ of $k\,(\geq 2)$ exponential populations with respect to the sum of the quadratic loss functions or the sum of the squared error losses. We first proposed estimators for $\underline{\theta}$ which are based on some baseline estimators for each component θ_i, such as MLE and UMVUE. We have constructed a class containing the estimator based on UMVUE of θ_i. Some techniques for improving estimators have been used to obtain estimators which dominate the UMVUE of $\underline{\theta}$. Further an admissible class has been obtained within the class. Next we have introduced the concept of invariance to our model and derive sufficient conditions for improving estimators which are equivariant under the location and affine group of transformations for the case $k = 2$. The inadmissibility result for the case $k\,(\geq 3)$ populations is not available. Finally, we have conducted a simulation study to numerically compare the risk functions of all the proposed estimators and recommended their use in practice. It may be noted that the simultaneous estimation of quantiles of $k\,(\geq 2)$ exponential populations has not been studied in the literature before.

Table 1: Relative risk performances of various estimators of exponential quantiles for $(n_1, n_2) = (5, 5), (10, 10), (15, 15)$.

$\eta \rightarrow$	0.05			2.5		
$\tau \downarrow$	R_{MM}	R_{MV}	R_{MVA}	R_{MM}	R_{MV}	R_{IMV}
	43.037	42.598	43.374	0.956	− 6.395	12.066
0.25	46.808	46.703	46.834	0.733	− 3.383	07.430
	48.318	48.335	48.337	0.558	− 2.280	04.900
	45.155	45.148	45.766	0.904	− 5.236	14.091
0.50	46.311	46.313	46.385	0.422	− 2.411	07.394
	46.890	46.924	46.924	0.284	− 1.579	05.167
	45.279	45.535	46.122	0.723	− 4.625	14.678
0.75	47.002	47.151	47.170	0.427	− 2.288	07.784
	46.913	46.967	46.967	0.349	− 1.602	05.419
	45.610	45.619	46.350	0.888	− 4.906	15.309
1.00	46.767	46.899	46.914	0.283	− 1.975	07.316
	46.952	46.988	46.988	0.196	− 1.335	04.996
	45.067	44.954	45.740	0.887	− 4.856	14.926
1.25	46.928	47.097	47.114	0.388	− 2.220	07.611
	46.842	46.877	46.877	0.357	− 1.583	05.452
	44.834	45.058	45.588	0.751	− 4.758	14.388
1.50	46.555	46.651	46.695	0.411	− 2.283	07.798
	47.772	47.818	47.818	0.231	− 1.442	05.076
	44.454	44.741	45.233	0.771	− 4.958	14.146
1.75	46.699	46.759	46.802	0.398	− 2.331	07.490
	46.982	47.000	47.000	0.197	− 1.428	04.570
	44.042	44.125	44.621	0.484	− 4.520	13.347
2.00	47.795	47.894	47.959	0.446	− 2.465	07.531
	46.428	46.478	46.478	0.388	− 1.764	05.469
	43.780	43.699	44.397	0.604	− 4.902	13.734
2.25	47.813	47.923	47.982	0.424	− 2.473	07.530
	46.457	46.483	46.483	0.384	− 1.785	05.191
	45.520	45.264	45.949	0.884	− 5.517	14.398
2.50	46.369	46.472	46.514	0.509	− 2.688	06.946
	48.159	48.209	48.210	0.323	− 1.729	05.094
	44.021	43.564	44.495	0.677	− 5.304	13.024
2.75	45.821	45.882	45.934	0.623	− 2.921	07.492
	46.457	46.476	46.477	0.349	− 1.812	04.967
	44.332	44.338	44.891	1.007	− 5.999	13.630
3.00	47.501	47.549	47.618	0.562	− 2.880	07.435
	47.773	47.815	47.818	0.363	− 1.879	05.051
	44.406	43.960	44.874	0.882	− 5.972	12.737
3.25	46.898	46.943	47.001	0.361	− 2.632	06.610
	47.658	47.686	47.689	0.249	− 1.684	04.841
	43.503	43.082	43.878	1.084	− 6.421	13.541
3.50	47.113	47.175	47.218	0.356	− 2.651	06.900
	47.248	47.282	47.284	0.295	− 1.828	04.781
	43.657	43.236	43.999	0.856	− 6.091	12.500
3.75	46.720	46.767	46.771	0.382	− 2.740	07.112
	47.962	47.972	47.975	0.363	− 1.968	04.365
	43.683	43.505	44.188	1.107	− 6.681	12.581
4.00	47.308	47.372	47.399	0.412	− 2.830	06.585
	47.804	47.831	47.836	0.229	− 1.748	04.923
	43.373	43.061	43.748	1.053	− 6.755	12.813
4.50	47.327	47.423	47.448	0.451	− 2.971	06.532
	47.789	47.838	47.843	0.520	− 2.294	04.967
	44.398	43.750	44.536	1.066	− 6.961	11.409
5.00	45.883	45.957	45.990	0.459	− 3.029	06.346
	47.180	47.197	47.212	0.519	− 2.339	04.823

Table 2: Relative risk performances of various estimators of exponential quantiles for $(n_1, n_2) = (3, 7), (5, 10), (10, 15)$.

$\eta \rightarrow$	0.05			2.5		
$\tau \downarrow$	R_{MM}	R_{MV}	R_{MVA}	R_{MM}	R_{MV}	R_{IMV}
0.25	41.906	41.323	43.081	1.669	− 9.291	16.740
	45.474	45.569	45.924	1.036	− 5.772	10.976
	47.309	47.391	47.405	0.272	− 2.487	05.275
0.50	42.950	43.248	44.022	1.333	− 6.102	19.997
	46.733	46.888	47.201	0.709	− 3.726	12.461
	47.516	47.558	47.599	0.524	− 2.323	06.857
0.75	44.397	44.332	45.282	0.821	− 4.296	20.276
	46.547	46.741	46.994	0.511	− 2.926	12.607
	47.023	47.052	47.062	0.308	− 1.779	06.602
1.00	45.131	45.321	46.071	0.644	− 3.677	20.487
	46.561	46.768	46.934	0.571	− 2.775	13.186
	47.745	47.789	47.790	0.180	− 1.442	06.457
1.25	45.226	45.386	46.131	0.721	− 3.704	20.581
	46.731	46.724	46.896	0.500	− 2.604	12.887
	47.602	47.710	47.710	0.357	− 1.713	07.150
1.50	44.199	44.099	44.772	0.544	− 3.329	20.346
	45.620	45.780	45.906	0.459	− 2.490	12.689
	47.137	47.204	47.204	0.189	− 1.450	06.663
1.75	44.278	44.194	44.857	0.453	− 3.072	20.837
	44.896	44.962	45.059	0.437	− 2.428	12.735
	47.072	47.097	47.097	0.192	− 1.462	06.699
2.00	43.354	43.583	44.053	0.572	− 3.335	20.450
	46.011	45.911	46.086	0.374	− 2.328	12.562
	47.194	47.230	47.232	0.261	− 1.589	06.637
2.25	44.806	44.845	45.397	0.522	− 3.291	19.334
	46.415	46.399	46.520	0.438	− 2.407	13.214
	46.523	46.534	46.535	0.248	− 1.596	06.240
2.50	44.570	44.353	44.797	0.440	− 3.195	19.147
	45.588	45.585	45.687	0.433	− 2.389	13.896
	46.890	46.951	46.953	0.278	− 1.630	06.169
2.75	45.507	45.258	45.759	0.529	− 3.319	20.592
	45.795	45.688	45.817	0.459	− 2.505	12.584
	46.556	46.616	46.618	0.293	− 1.688	06.672
3.00	43.824	43.592	44.095	0.466	− 3.226	19.385
	45.929	45.869	45.975	0.367	− 2.334	12.770
	46.921	46.927	46.936	0.293	− 1.705	06.848
3.25	44.325	44.097	44.547	0.399	− 3.162	19.697
	46.165	46.129	46.258	0.318	− 2.313	11.657
	47.061	47.054	47.064	0.289	− 1.752	06.671
3.50	44.469	44.156	44.536	0.503	− 3.380	19.444
	47.253	47.241	47.366	0.242	− 2.167	12.236
	46.078	46.081	46.095	0.236	− 1.670	06.318
3.75	45.216	44.809	45.341	0.401	− 3.146	19.755
	45.286	45.188	45.343	0.340	− 2.374	11.737
	46.998	47.056	47.069	0.358	− 1.832	06.970
4.00	44.480	43.966	44.530	0.443	− 3.305	19.008
	46.173	46.131	46.227	0.401	− 2.472	12.908
	46.851	46.880	46.895	0.344	− 1.844	07.032
4.50	45.394	44.937	45.334	0.217	− 2.858	18.840
	45.033	44.991	45.061	0.352	− 2.414	12.076
	46.975	46.970	46.978	0.284	− 1.774	06.626
5.00	43.720	43.404	43.822	0.507	− 3.385	19.881
	45.965	45.931	46.002	0.404	− 2.543	12.480
	46.456	46.492	46.496	0.314	− 1.870	06.477

Table 3: Relative risk performances of various estimators of exponential quantiles for $(n_1, n_2) = (7, 3), (10, 5), (15, 10)$.

$\eta \rightarrow$	0.05			2.5		
$\tau \downarrow$	R_{MM}	R_{MV}	R_{MVA}	R_{MM}	R_{MV}	R_{IMV}
0.25	45.446	44.948	45.439	0.435	− 3.278	19.087
	45.940	45.996	46.036	0.494	− 2.602	13.201
	47.235	47.222	47.237	0.266	− 1.737	06.439
0.50	43.214	43.298	43.805	0.455	− 3.153	19.717
	46.619	46.648	46.770	0.378	− 2.303	12.980
	47.152	47.149	47.151	0.469	− 1.929	07.236
0.75	44.330	44.142	44.969	0.594	− 3.472	19.887
	45.780	45.801	45.959	0.427	− 2.390	13.113
	47.580	47.639	47.639	0.319	− 1.657	06.724
1.00	44.521	44.644	45.443	0.629	− 3.662	20.348
	46.106	46.110	46.330	0.347	− 2.378	12.591
	46.928	46.982	46.984	0.265	− 1.605	06.852
1.25	45.323	45.618	46.424	0.850	− 4.282	20.591
	46.781	47.000	47.221	0.578	− 2.945	12.776
	47.166	47.143	47.149	0.292	− 1.690	06.746
1.50	43.385	43.322	44.350	1.014	− 4.806	21.011
	46.763	46.766	47.085	0.772	− 3.466	13.435
	47.490	47.5304	47.538	0.456	− 2.055	07.111
1.75	44.075	44.128	45.052	1.306	− 5.750	20.317
	47.294	47.384	47.732	0.515	− 3.208	12.078
	47.159	47.170	47.196	0.443	− 2.137	06.862
2.00	42.991	43.446	44.243	0.968	− 5.305	19.192
	46.865	47.210	47.448	0.730	− 3.761	12.445
	47.128	47.208	47.224	0.409	− 2.164	06.537
2.25	43.771	44.121	44.938	1.630	− 7.037	20.335
	46.120	46.198	46.525	0.912	− 4.282	12.849
	46.895	46.937	46.979	0.417	− 2.228	06.382
2.50	43.007	42.970	44.183	1.373	− 6.872	19.011
	46.586	46.613	47.026	0.698	− 4.150	11.695
	47.448	47.503	47.556	0.358	− 2.244	06.047
2.75	43.741	43.624	44.675	1.554	− 7.584	18.337
	46.123	46.068	46.504	0.913	− 4.627	12.069
	48.110	48.221	48.257	0.480	− 2.523	06.138
3.00	43.166	42.889	44.267	1.744	− 8.175	18.991
	45.193	45.383	45.637	0.918	− 4.830	11.355
	47.727	47.848	47.875	0.511	− 2.656	06.237
3.25	42.960	43.097	44.335	1.845	− 8.768	18.520
	45.672	45.976	46.113	0.888	− 5.019	11.025
	47.430	47.459	47.546	0.527	− 2.755	06.115
3.50	41.211	40.899	42.355	1.603	− 8.551	17.837
	45.775	45.681	46.111	1.078	− 5.463	11.228
	46.820	46.848	46.931	0.591	− 2.955	05.738
3.75	42.765	42.426	44.002	1.943	− 9.507	18.044
	45.668	45.778	46.103	1.048	− 5.678	10.862
	48.711	48.813	48.853	0.475	− 2.794	05.613
4.00	41.437	40.853	42.613	2.086	− 10.158	17.591
	45.805	45.877	46.266	1.226	− 6.067	10.923
	47.696	47.710	47.768	0.578	− 3.029	05.695
4.50	42.217	41.511	43.183	2.007	− 10.418	17.181
	45.568	45.667	46.074	1.177	− 6.297	10.417
	48.148	48.199	48.243	0.640	− 3.251	05.778
5.00	40.642	39.702	41.661	2.120	− 11.496	16.171
	45.277	45.013	45.486	1.145	− 6.369	10.399
	47.350	47.432	47.445	0.665	− 3.375	05.678

APPENDIX

Proof of Theorem 2.1: In order to prove the theorem we use the orbit-by-orbit improvement technique of Brewster and Zidek [2].

Consider the risk function of $\underline{\delta}_{\mathbf{c}}$ with respect to the loss function (1.1),

$$R(\underline{\alpha}, \underline{\delta}_{\mathbf{c}}) = E\Big[Z + \eta c_1 T_1 + (\eta - 1)T^* - \mu - \eta\sigma_1\Big]^2 + E\Big[Z + \eta c_2 T_2 + (\eta - 1)T^* - \mu - \eta\sigma_2\Big]^2 . \tag{A.1}$$

It can be easily seen that the above risk (A.1) is a convex function in both c_1 and c_2. After some calculations, the minimizing choices for c_1 and c_2 are obtained as

$$\hat{c}_j(\underline{\alpha}) = \frac{(\mu + \eta\sigma_j)ET_j - E(ZT_j) - (\eta - 1)E(T_j T^*)}{\eta ET_j^2}; \quad j = 1, 2 . \tag{A.2}$$

Let $\lambda_j = (\sigma_j a)^{-1}$, and using this we obtain the minimizing choice of each c_j as

$$\hat{c}_j(\lambda_j) = \frac{n_j\big(\eta - 2\eta\lambda_j + \lambda_j^2\big)}{\eta\big(1 + n_j - 2n_j\lambda_j\big)}; \quad j = 1, 2 . \tag{A.3}$$

To apply the orbit-by-orbit improvement technique of Brewster and Zidek [2], we need to get the supremum and infimum values of $\hat{c}_1$ and $\hat{c}_2$ with respect to λ_j and for fixed η. It is easy to see that $0 < \lambda_j < \frac{1}{n_j}$. We consider the following three separate cases.

Case-(I): Let $\eta \geq \max\big((n_1{+}1)^2/4n_1, (n_2{+}1)^2/4n_2\big)$. Differentiating $\hat{c}_j(\lambda_j)$ with respect to λ_j we have $\frac{d\hat{c}_j}{d\lambda_j} = \frac{-2n_j\big(n_j\lambda_j^2 - \lambda_j(n_j+1) + \eta\big)}{\eta(n_j+1-2n_j\lambda_j)^2}$; $j = 1, 2$. It is easy to observe that the derivative is $g(\lambda_j) = -n_j\lambda_j^2 + \lambda_j(n_j + 1) - \eta$ multiplied by a positive factor. Now $g(\lambda_j)$ is a concave function of λ_j; $j = 1, 2$. The maximum value is attained at $\lambda_j = (n_j + 1)/2n_j < 1/n_j$. The maximum value is $(n_j + 1)^2/4n_j - \eta < 0$. This implies $g(\lambda_j) < 0$ for $0 < \lambda_j < \frac{1}{n_j}$; $j = 1, 2$. Hence the function $\hat{c}_j(\lambda_j)$ is decreasing with respect to λ_j. Hence we have

$$\inf_{0<\lambda_j\leq\frac{1}{n_j}} \hat{c}_j(\lambda_j) = \hat{c}_j(1/n_j) = \frac{n_j\eta(n_j - 2) + 1}{n_j\eta(n_j - 1)} = a_j \quad \text{(say)}$$

and

$$\sup_{0<\lambda_j\leq\frac{1}{n_j}} \hat{c}_j(\lambda_j) = \hat{c}_j(0) = \frac{n_j}{n_j + 1} = b_j; \quad j = 1, 2, \quad \text{(say)} .$$

Case-(II): Let $1 \leq \eta < \min\{(n_1+1)^2/4n_1,\ (n_2+1)^2/4n_2\}$. It is easy to see that the maximum value of $g(\lambda_j)$ is positive. The equation $g(\lambda_j) = 0$, has two real roots say $\lambda_j^- = \frac{(n_j+1)+\sqrt{(n_j+1)^2-4\eta n_j}}{2n_j}$ and $\lambda_j^+ = \frac{(n_j+1)-\sqrt{(n_j+1)^2-4\eta n_j}}{2n_j}$; $j=1,2$. It is also noticed that, these two roots are outside the interval $\left(0, \frac{1}{n_j}\right]$. Hence for $0 < \lambda_j \leq \frac{1}{n_j}$ the function $g(\lambda_j) < 0$. This implies that the function $\hat{c}_j(\lambda_j)$ is decreasing in the concerned interval. Hence we have

$$\inf_{0<\lambda_j\leq\frac{1}{n_j}} \hat{c}_j(\lambda_j) \;=\; \hat{c}_j(1/n_j) \;=\; \frac{n_j\,\eta\,(n_j-2)+1}{n_j\,\eta\,(n_j-1)} \;=\; a_j$$

and

$$\sup_{0<\lambda_j\leq\frac{1}{n_j}} \hat{c}_j(\lambda_j) \;=\; \hat{c}_j(0) \;=\; \frac{n_j}{n_j+1} \;=\; b_j\,; \quad j=1,2\,,$$

Case-(III): Let $0 \leq \eta < 1$. For this case it can be observed that the root λ_j^- is outside the concerned interval, but λ_j^+ is inside the interval $\left(0, \frac{1}{n_j}\right]$. Also $\hat{c}_j{}''(\lambda_j^+) > 0$ and $\hat{c}_j{}''(\lambda_j^-) < 0$, hence λ_j^- is a point of local maxima and λ_j^+ is a point of local minima. Hence the function $g(\lambda_j) < 0$ in the interval $(0, \lambda_j^+]$ and $g(\lambda_j) \geq 0$ in the interval $(\lambda_j^+, 1/n_j]$. Thus the function $\hat{c}_j(\lambda_j)$ is decreasing in the interval $(0, \lambda_j^+]$ and increasing in the interval $(\lambda_j^+, 1/n_j]$. We have

$$\inf_{0<\lambda_j<\frac{1}{n_j}} \hat{c}_j(\lambda_j) \;=\; \hat{c}_j(\lambda_j^+) \;=\; {c_j}^+\,,$$

and

$$\sup_{0<\lambda_j<\frac{1}{n_j}} \hat{c}_j(\lambda_j) \;=\; \max\Big\{\hat{c}_j(0), \hat{c}_j(1/n_j)\Big\} \;=\; \max\left\{\frac{n_j}{n_j+1}, \frac{1+\eta\,n_j\,(n_j-2)}{\eta\,n_j\,(n_j-1)}\right\} \;=\; d_j\,,$$

where

$$\lambda_j^+ \;=\; \left\{(n_j+1) - \sqrt{(n_j+1)^2 - 4\eta n_j}\right\}\Big/ 2n_j\,; \qquad j=1,2\,.$$

Now combining Cases I–III, it is easy to define the functions $\mathbf{c}^*$ and $\mathbf{c}_*$ as in (2.1) and (2.2) respectively. The loss function is (1.1), which is the sum of the squared errors, and it is convex with respect to both c_1 and c_1. Then by applying the orbit-by-orbit improvement technique of Brewster and Zidek [2] we get the improved estimators for $\underline{\delta}_{\mathbf{c}}$ in the class D, if either c_1 lies outside the interval $[a_1, b_1]$ (when $\eta \geq 1$) and $[{c_1}^+, d_1]$ (when $0 \leq \eta < 1$) or c_2 lies outside the interval $[a_2, b_2]$ (when $\eta \geq 1$) and $[{c_2}^+, d_2]$ (when $0 \leq \eta < 1$) with probability 1. Applying the Brewster and Zidek [2] technique we have $R(\underline{\delta}_{\mathbf{c}^*}, \underline{\alpha}) \leq R(\underline{\delta}_{\mathbf{c}}, \underline{\alpha})$ when $\eta \geq 1$, and $R(\underline{\delta}_{\mathbf{c}_*}, \underline{\alpha}) \leq R(\underline{\delta}_{\mathbf{c}}, \underline{\alpha})$ when $0 \leq \eta < 1$. This completes the proof of the theorem. □

Proof of Theorem 3.1: The proof of the theorem can be done by using the orbit-by-orbit improvement technique for improving equivariant estimators proposed by Brewster and Zidek [2]. Consider the conditional risk function of $\underline{\delta}_{\Psi}$ given $W = T_2/T_1$:

$$R\big(\underline{\delta}_{\Psi} \,|\, W{=}w\big) \,=\, \sum_{i=1}^{2} \frac{1}{\sigma_i^2}\, E\big(Z + T_1 \Psi_i(W) - \theta_i\big)^2 \,. \tag{A.4}$$

It is easy to observe that the above risk (A.4) is a convex function of both Ψ_1 and Ψ_2. Hence, the minimizing choice of $\Psi_i(w)$ is obtained as

$$\hat{\Psi}_i(w) \,=\, -\frac{E(Z-\theta_i)\; E(T_1 \,|\, W{=}w)}{E(T_1^2 \,|\, W{=}w)} \,, \qquad i = 1, 2 \,.$$

Using the joint probability density function of (T_1, T_2), we can easily derive the joint probability density function of (T_1, W). The conditional probability density function of T_1 given W is a gamma distribution with shape parameter $n_1 + n_2 - 1$ and scale parameter $1/M$ where $M = \frac{n_1}{\sigma_1} + \frac{n_2}{\sigma_2}\, w$. Hence the conditional expectations are calculated as

$$E(T_1 \,|\, W) \,=\, \frac{n_1 + n_2 - 1}{M} \,, \qquad E(T_1^2 \,|\, W) \,=\, \frac{(n_1 + n_2 - 1)\,(n_1 + n_2)}{M^2} \,.$$

Substituting all these values and simplifying we obtain the minimizing choice of $\hat{\Psi}_1$ and $\hat{\Psi}_2$ as

$$\hat{\Psi}_1(w, \tau) \,=\, \frac{\big[\eta - (n_2\tau + n_1)^{-1}\big]\,\big[n_1 + w\, n_2 \tau\big]}{(n_1 + n_2)}$$

and

$$\hat{\Psi}_2(w, \tau) \,=\, \frac{\big[\frac{\eta}{\tau} - (n_2\tau + n_1)^{-1}\big]\,\big[n_1 + w\, n_2 \tau\big]}{(n_1 + n_2)}$$

respectively, where we denote $\tau = \sigma_1/\sigma_2 > 0$.

In order to apply the orbit-by-orbit improvement technique of Brewster and Zidek [2] for improving equivariant estimator, we need the supremum and infimum of both $\hat{\Psi}_1$ and $\hat{\Psi}_2$ with respect to $\tau > 0$ for fixed values of n_1, n_2, η and for given w. We consider the following three separate cases for calculating the supremum and infimum.

Case I: Let $0 < \eta < \min\big\{\frac{1}{n_1}, \frac{1}{n_2}\big\}$. Consider the first component $\hat{\Psi}_1(w, \tau)$. Differentiating with respect to τ we have $\frac{d\hat{\Psi}_1}{d\tau} = \frac{\eta n_2^3 w\tau^2 + 2\eta n_1 n_2^2 w\tau + n_1 n_2(\eta n_1 w - w + 1)}{(n_1+n_2)(n_1+n_2\tau)^2}$. Let $h(\tau) = \eta n_2^3 w\tau^2 + 2\,\eta n_1 n_2^2 w\tau + n_1 n_2(\eta n_1 w - w + 1)$. Now $h(\tau)$ is a convex function of $\tau \in (0, \infty)$. Its minimum is attained at $\tau = -\frac{n_1}{n_2} < 0$. Hence in the region $(0, \infty)$ the minimum will be attained at $\tau = 0$ and the minimum value of $h(\tau)$ is $n_1 n_2(1 - w + \eta n_1 w)$. Assume that the minimum value is positive that

is $0 < w \leq \frac{1}{1-\eta n_1}$. For this case $h(\tau) \geq 0$ for $\tau \in (0, \infty)$. Hence the function $\hat{\Psi}_1(w, \tau)$ is an increasing function of $\tau > o$. Hence we have

$$\inf_{\tau>0} \hat{\Psi}_1(w,\tau) = \frac{\eta n_1 - 1}{n_1 + n_2} \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_1(w,\tau) = \infty, \quad \text{when} \;\; 0 < w \leq \frac{1}{(1-\eta n_1)} \, .$$

If $w > \frac{1}{1-\eta n_1}$, then the minimum value of $h(\tau)$ is negative and it will cross the τ axis. The function $h(\tau)$ has two real roots say $\tau^- = -\frac{n_1}{n_2} - \frac{1}{n_2}\sqrt{\frac{n_1(w-1)}{\eta w}}$ and $\tau^+ = -\frac{n_1}{n_2} + \frac{1}{n_2}\sqrt{\frac{n_1(w-1)}{\eta w}}$. It is easy to observe that $\tau^- < 0$ and $\tau^+ > 0$. Hence $h(\tau) < 0$ in the region $0 < \tau < \tau^+$ and $h(\tau) \geq 0$ in the region $\tau^+ < \tau < \infty$. Hence the function $\hat{\Psi}_1(w, \tau)$ is decreasing in the region $0 < \tau < \tau^+$ and increasing in the region $\tau^+ < \tau < \infty$. Hence we have

$$\inf_{\tau>0} \hat{\Psi}_1(w,\tau) = \hat{\Psi}_1(w,\tau^+) \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_1(w,\tau) = \infty, \quad \text{when} \;\; w > \frac{1}{(1-\eta n_1)} \, ,$$

where

$$\hat{\Psi}_1(w,\tau^+) \;=\; \frac{\left[\eta - (n_2\tau^+ + n_1)^{-1}\right]\left[n_1 + w n_2 \tau^+\right]}{(n_1+n_2)} \, .$$

Next consider the second component $\hat{\Psi}_2$. The derivative of $\hat{\Psi}_2$ with respect to τ is $g(\tau) = \tau^2(n_1 n_2 - \eta n_1 n_2^2 - n_1 n_2 w) - 2\,\eta n_1^2 n_2 \tau - \eta n_1^3$ multiplied by a positive factor. For this case $g(\tau)$ is a convex function of $\tau > 0$. The minimum attained at $\tau = \frac{\eta n_1}{1-\eta n_2 - w} > 0$. Its minimum value is $\frac{\eta n_1^3 (w-1)}{1-\eta n_2 - w} < 0$ as $w < 1$. Since the minimum value of $g(\tau)$ is negative, it will cross the τ axis. The equation $g(\tau) = 0$ has two real roots say $\alpha^- = \frac{\eta n_1}{1-\eta n_2 - w} - \frac{n_1}{n_2}\frac{\sqrt{\eta n_2(1-w)}}{1-\eta n_2 - w}$ and $\alpha^+ = \frac{\eta n_1}{1-\eta n_2 - w} + \frac{n_1}{n_2}\frac{\sqrt{\eta n_2(1-w)}}{1-\eta n_2 - w}$. It is noticed that $\alpha^- < 0$ and $0 < \frac{\eta n_1}{1-\eta n_2 - w} < \alpha^+$. Hence the function $g(\tau) < 0$ in the region $(0, \alpha^+)$ and $g(\tau) \geq 0$ in the region $[\alpha^+, \infty)$. This implies that $\hat{\Psi}_2(w, \tau)$ is decreasing in the region $(0, \alpha^+)$ and increasing in the region $[\alpha^+, \infty)$. Hence we have

$$\inf_{\tau>0} \hat{\Psi}_2(w,\tau) = \hat{\Psi}_2(w,\alpha^+) \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_2(w,\tau) = \max\left\{\hat{\Psi}_2(w,0), \hat{\Psi}_2(w,\infty)\right\} = \infty \, ,$$

where

$$\hat{\Psi}_2(w,\alpha^+) \;=\; \frac{1}{n_1+n_2}\left[\frac{\eta}{\alpha^+} - \frac{1}{n_1 + n_2\alpha^+}\right]\left[n_1 + n_2 w \alpha^+\right] ,$$

when $1 - \eta n_2 \geq w$.

Now assume that $1 - \eta n_2 < w$. Then the function $g(\tau)$ is a concave function of τ. Its maximum value is attained at $\tau = \frac{\eta n_1}{1-\eta n_2 - w} < 0$. Hence within the concerned region the maximum is attained at $\tau = 0$. Its maximum value is $-\eta n_1^3 < 0$. This implies that the function $g(\tau) < 0$ in the region $(0, \infty)$. Thus the function $\hat{\Psi}_2(w, \tau)$ is decreasing in $\tau \in (0, \infty)$. Hence we have

$$\inf_{\tau>0} \hat{\Psi}_2(w,\tau) = \frac{n_2 w}{n_1+n_2}\left(\eta - \frac{1}{n_2}\right) \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_2(w,\tau) = \infty \, ,$$

when $1 - \eta n_2 < w$.

Case II: Let $\eta \geq \max\{\frac{1}{n_1}, \frac{1}{n_2}\}$. Consider the first component $\Psi_1(w,\tau)$. Now the derivative of $\hat{\Psi}_1(w,\tau)$ with respect to τ is $h(\tau)$ multiplied by a positive factor. As in Case I, the function $h(\tau)$ is a convex function of τ. The minimum is attained at $\tau = -\frac{n_1}{n_2} < 0$. Hence within the interval $(0,\infty)$ the minimum is attained at $\tau = 0$. Its minimum value is $n_1 n_2(1 - w + \eta n_1 w) \geq 0$ as $\eta \geq \max\{\frac{1}{n_1}, \frac{1}{n_2}\}$. Hence $h(\tau) \geq 0$, $\forall\, \tau > 0$. Thus the function $\hat{\Psi}_1(w,\tau)$ is increasing in the region $(0,\infty)$. Thus we have

$$\inf_{\tau>0} \hat{\Psi}_1(w,\tau) = \hat{\Psi}_1(w,0) = \frac{\eta n_1 - 1}{n_1 + n_2} \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_1(w,\tau) = \hat{\Psi}_1(w,\infty) = \infty\,.$$

Consider the second component $\hat{\Psi}_2(w,\tau)$. As in Case I, the derivative of $\hat{\Psi}_2(w,\tau)$ is simply $g(\tau)$ multiplied by a positive factor. Also under the condition $\eta \geq \max\{\frac{1}{n_1}, \frac{1}{n_2}\}$, the only possibility is $1 - \eta n_2 < w$. The function $g(\tau)$ is a concave function and the maximum is attained at $\tau = \frac{\eta n_1}{1-\eta n_2 - w} < 0$. Hence the maximum will be attained at $\tau = 0$ in the concerned region $(0,\infty)$. The maximum value is $-\eta n_1^3 < 0$. Hence $g(\tau) < 0$, $\forall\, \tau > 0$. Thus the function $\hat{\Psi}_2(w,\tau)$ is decreasing in the region $(0,\infty)$. Thus we have

$$\inf_{\tau>0} \hat{\Psi}_2(w,\tau) = \frac{n_2\, w}{n_1 + n_2}\left(\eta - \frac{1}{n_2}\right) \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_2(w,\tau) = \infty\,.$$

Case III: Let $\min\{\frac{1}{n_1}, \frac{1}{n_2}\} \leq \eta < \max\{\frac{1}{n_1}, \frac{1}{n_2}\}$. For this case we have two possibilities either $\frac{1}{n_1} \leq \eta < \frac{1}{n_2}$ or $\frac{1}{n_2} \leq \eta < \frac{1}{n_1}$. Analyzing as in the above cases we have for $\frac{1}{n_1} \leq \eta < \frac{1}{n_2}$,

$$\inf_{\tau>0} \hat{\Psi}_1(w,\tau) = \hat{\Psi}_1(w,0) = \frac{\eta n_1 - 1}{n_1 + n_2} \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_1(w,\tau) = \hat{\Psi}_1(w,\infty) = \infty\,,$$

and

$$\inf_{\tau>0} \hat{\Psi}_2(w,\tau) = \begin{cases} \hat{\Psi}_2(w,\alpha^+)\,, & \text{if } w \leq 1-\eta n_2\,, \\ \frac{n_2\, w}{n_1+n_2}\left(\eta - \frac{1}{n_2}\right), & \text{if } w > 1-\eta n_2\,, \end{cases} \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_2(w,\tau) = +\infty\,.$$

Likewise when $\frac{1}{n_2} \leq \eta < \frac{1}{n_1}$, we have

$$\inf_{\tau>0} \hat{\Psi}_1(w,\tau) = \begin{cases} \frac{\eta n_1 - 1}{n_1+n_2}, & \text{if } w \leq \frac{1}{1-\eta n_1}\,, \\ \hat{\Psi}_1(w,\tau^+), & \text{if } w > \frac{1}{1-\eta n_1}\,, \end{cases} \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_1(w,\tau) = +\infty\,,$$

and

$$\inf_{\tau>0} \hat{\Psi}_2(w,\tau) = \frac{n_2\, w}{n_1 + n_2}\left(\eta - \frac{1}{n_2}\right) \quad \text{and} \quad \sup_{\tau>0} \hat{\Psi}_2(w,\tau) = \infty\,.$$

Now it is easy to define the functions $\underline{\Psi}_0$ $\left(\text{when } 0 < \eta \leq \min\left\{\frac{1}{n_1}, \frac{1}{n_2}\right\}\right)$, $\underline{\Psi}^0$ $\left(\text{when } \eta \geq \max\left\{\frac{1}{n_1}, \frac{1}{n_2}\right\}\right)$, $\underline{\Psi}_{11}$ $\left(\text{when } \frac{1}{n_1} \leq \eta < \frac{1}{n_2}\right)$ and $\underline{\Psi}_{22}$ $\left(\text{when } \frac{1}{n_2} \leq \eta < \frac{1}{n_1}\right)$ as defined in (3.4), (3.5), (3.6) and (3.7) respectively. Since the loss function (1.2) is a sum of the convex loss functions with respect to both $\hat{\Psi}_1$ and $\hat{\Psi}_2$, an application of the Theorem 3.3.1 of Brewster and Zidek [2], gives $R(\underline{\delta}_{\underline{\Psi}_0}, \underline{\alpha}) \leq R(\underline{\delta}_{\underline{\Psi}}, \underline{\alpha})$ if there exist some values of parameters $\underline{\alpha}$ such that $P_{\underline{\alpha}}(\underline{\Psi}_0 \neq \underline{\Psi}) > 0$ for the case $0 < \eta \leq \min\left\{\frac{1}{n_1}, \frac{1}{n_2}\right\}$. Similarly by applying the Brewster and Zidek [2] technique for the case $\eta \geq \max\left\{\frac{1}{n_1}, \frac{1}{n_2}\right\}$, we have $R(\underline{\delta}_{\underline{\Psi}^0}, \underline{\alpha}) \leq R(\underline{\delta}_{\underline{\Psi}}, \underline{\alpha})$ if there exist some values of parameters $\underline{\alpha}$ such that $P_{\underline{\alpha}}(\underline{\Psi}^0 \neq \underline{\Psi}) > 0$. When $\frac{1}{n_1} \leq \eta < \frac{1}{n_2}$ the estimator $\underline{\delta}_{\underline{\Psi}_{11}}$ improves upon $\underline{\delta}_{\underline{\Psi}}$ if $P_{\underline{\alpha}}(\underline{\Psi}_{11} \neq \underline{\Psi}) > 0$ for some choices of $\underline{\alpha}$. When $\frac{1}{n_2} \leq \eta < \frac{1}{n_1}$ the estimator $\underline{\delta}_{\underline{\Psi}_{22}}$ improves upon $\underline{\delta}_{\underline{\Psi}}$ if $P_{\underline{\alpha}}(\underline{\Psi}_{22} \neq \underline{\Psi}) > 0$ for some choices of $\underline{\alpha}$. This completes the proof of the theorem. □

ACKNOWLEDGMENTS

The first author (Manas Ranjan Tripathy) would like to thank DST (SERB), grant number SR/FTP/MS-37/2012 dated 04.10.2013 New Delhi, India for providing the financial support. The authors would like to thank the two unanimous referees and the editors for their valuable suggestions and comments which helped in strengthening the article.

REFERENCES

[1] Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters, *The Annals of Statistics*, **8**(3), 545–571.

[2] Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators, *Annals of Statistics*, **2**, 21–38.

[3] Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution, *Annals of Mathematical Statistics*, **25**, 373–381.

[4] Ghosh, M. and Auer, R. (1983). Simultaneous estimation of parameters in exponential families, *Annals of The Institute of Statistical Mathematics*, **35**(3), A, 379–387.

[5] Ghosh, M. and Razmpour, A. (1984). Estimation of the common location parameter of several exponentials, *Sankhya: The Indian Journal of Statistics*, **46**, Ser. A, pt. 3, 383–394.

[6] Gupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses, *The Annals of Statistics*, **14**(1), 206–219.

[7] Jin, C. and Crouse, H. (1998a). A note on the common location parameter of several exponential populations, *Communications in Statistics – Theory and Methods*, **27**(11), 2777–2789.

[8] Jin, C. and Crouse, H. (1998b). An identity for exponential distributions with the common location parameter and its applications, *Communications in Statistics – Theory and Methods*, **27**(2), 409–422.

[9] Jin, C. and Pal, N. (1992). On common location of several exponentials under a class of convex loss functions, *Calcutta Statistical Association Bulletin*, **42**(3–4), 191–200.

[10] Kumar, S. (1997). *Estimation of common parameters*, "Mathematics and Its Applications in Engineering and Industry" (B. Singh *et al.*, Eds.), University of Roorkee, pp. 538–548.

[11] Kumar, S. and Sharma, D. (1996). A note on estimating quantiles of exponential populations, *Statistics and Probability Letters*, **26**, 115–118.

[12] Mitra, P.K. and Sinha, B.K. (2007). On some aspects of estimation of a common mean of two independent normal populations, *Journal of Statistical planning and Information*, **137**(1), 184–193.

[13] Pal, N.; Lin, J.J.; Chang, C.H. and Kumar, S. (2007). A revisit to the common mean problem: Comparing the maximum likelihood estimator with the Graybill–Deal estimator, *Computational Statistics and Data Analysis*, **51**(12), 5673–5681.

[14] Pal, N. and Sinha, B.K. (1990). Estimation of a common location of several exponentials, *Statistics and Decisions*, **9**, 27–36.

[15] Pal, N. and Sinha, B.K. (1996). Estimation of a common mean of several normal populations: A review, *Far East Journal of Mathematical Sciences, Special Volume part-1 (Bhattacharya Memorial Volume)*, 97–110.

[16] Rukhin, A.L. (1986). Admissibility and minimaxity results in the estimation problem of exponential quantiles, *Annals of Statistics*, **14**(1), 220–237.

[17] Rukhin, A.L. and Strawderman, W.E. (1982). Estimating a quantile of an exponential distribution, *Journal of American Statistical Association*, **77**, 159–162.

[18] Saleh, A.K.Md.E. (1981). *Estimating quantiles of exponential distributions*. In "Statistics and Related Topics" (M. Csorgo, D. Dawson, J.N.K. Rao, A.K.Md.E. Saleh, Eds.), North Holland, Amsterdam, 279–283.

[19] Sharma, D. and Kumar, S. (1994). Estimating quantiles of exponential populations, *Statistics and Decisions*, **12**, 343–352.

[20] Tripathy, M.R. and Kumar, S. (2010). Estimating a common mean of two normal populations, *Journal of Statistical Theory and Applications*, **9**(2), 197–215.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 425–442

THE MODIFIED BOREL–TANNER (MBT) REGRESSION MODEL

Authors: Emilio Gómez-Déniz
– Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Spain
emilio.gomez-deniz@ulpgc.es

Francisco J. Vázquez-Polo
– Department of Quantitative Methods and TIDES Institute, University of Las Palmas de Gran Canaria, Spain
francisco.vazquezpolo@ulpgc.es

Victoriano García
– Department of Statistics and Operational Research, University of Cádiz, Spain
victoriano.garcia@uca.es

Received: July 2014 Revised: September 2015 Accepted: September 2015

Abstract:

- A new one-parameter family of discrete distributions is presented. It has some advantages against the Poisson distribution as a suitable model for modelling data with a high frequencies of zeros and showing over-dispersion (variance larger than the mean). The distribution is obtained from a simple modification of the Borel–Tanner distribution, which has not received attention from the statistical community in the past. We also propose a generalized regression model which can be used for a count dependent variable, when the above features are observed, as an alternative to the well-known Poisson regression model, among others. Maximum likelihood estimation is investigated and illustrated with an example of interrelation between fatalities in trucks accidents on American roads and some covariates considered.

Key-Words:

- *Borel–Tanner distribution; covariate; estimation.*

AMS Subject Classification:

- 62E99, 62P25.

1. INTRODUCTION

The Borel–Tanner distribution is a discrete distribution proposed more than fifty years ago in queueing theory to model the probability distribution of the number of customers served in a queuing line with Poisson input and a constant service time, given that the length of the queue at the initial time is r. As far as we know, this distribution has not received much attention for the statistical community. The probability function of the Borel–Tanner distribution ([11]) is given by

$$\Pr(Y=y) = A(y,r)e^{-\alpha y}\alpha^{y-r}, \quad y = r, r+1, \ldots, \tag{1.1}$$

where $\alpha > 0$ and r is a positive integer and where

$$A(y,r) = \frac{r}{(y-r)!}\, y^{y-r-1}.$$

Equivalently, [10] rewritten expression (1.1) as

$$\Pr(Y=y) = B(y,r)\,\frac{\alpha^{y-r}}{(1+\alpha)^{2y-r}}, \quad y = r, r+1, \ldots, \tag{1.2}$$

where

$$B(y,r) = \frac{r}{y}\binom{2y-r-1}{y-1}.$$

In this paper we focus on the distribution with probability distribution given in (1.2) using a modified version of this probability distribution with support in $0, 1, ...$, suitable for modelling data with a high frequencies of zero and showing over-dispersion phenomena: the variance is larger than the mean.

The distribution proposed here has some advantages against some other well-known distributions as a suitable model for modelling data with a high frequencies of zeros and showing over-dispersion phenomena. We also propose a generalized regression model which can be used for a count dependent variable, when the above features are observed. Maximum likelihood estimation is investigated and illustrated with an example involving emergency room visits to hospital.

The applicability of the model is shown by fitting the number of deaths in truck accidents (fatalities) on American roads, with different explanatory covariates from real data used by [17]. The provided real data examples show that the model works reasonably well, and this assessment is confirmed by the comparison to the Poisson and negative binomial distributions.

The contents of the paper are as follows. In section 2 we present the modified version of the Borel–Tanner distribution proposed here. Some properties of

the distribution are also shown, including the mean, variance and the cumulative distribution function. Some methods of estimation are developed in section 3. The regression model is developed in section 4. An application with real data is shown in section 4 and conclusions in the last section.

2. THE MODIFIED BOREL–TANNER DISTRIBUTION (MBT)

In this section we propose a modified version of the Borel–Tanner distribution given in (1.2) which has support in the positive integer numbers including the zero value. Firstly, consider $r=1$ and $X=Y-1$, then it is a simple exercise to see that the resulting shifted distribution has its probability function given by

$$\Pr(X=x) = \frac{\Gamma(2x+1)}{\Gamma(x+2)\,\Gamma(x+1)}\,\frac{\alpha^x}{(1+\alpha)^{2x+1}}\,, \qquad x=0,1,\ldots\,, \tag{2.1}$$

being $0<\alpha<1$. It has to be pointed out that in the original paper of [10] any parameter value $\alpha>0$ is allowed. Nevertheless, a simple algebra shows that it is not true and the feasible set of this parameter is actually $0<\alpha<1$. This distribution can be easily written as

$$\Pr(X=x) = C_x\,\frac{\alpha^x}{(1+\alpha)^{2x+1}}\,, \qquad x=0,1,\ldots\,,$$

where

$$C_x = \frac{1}{x+1}\binom{2x}{x}$$

are the Catalan numbers (see [13], p. 13 and [18]. In the sequel, when a random variable X follows the probability mass function (2.1) we will denote $X\sim \mathrm{MBT}(\alpha)$.

Since probability function (2.1) can be written as

$$\Pr(X=x) = C_x \exp\big[\lambda\cdot x - A(\lambda)\big]\,,$$

where

$$\lambda = \log\frac{\alpha}{(1+\alpha)^2}\,,$$

and

$$A(\lambda) = \log\left(\frac{1-\sqrt{1-4e^{\lambda}}}{2e^{\lambda}}\right) = \log(1+\alpha)\,,$$

the modified Borel–Tanner distribution proposed here is a member of the natural exponential family of distributions. Furthermore, probability function (2.1) can also be rewritten as

$$\Pr(X=x) = \frac{C_x}{1+\alpha}\left[\frac{\alpha}{(1+\alpha)^2}\right]^x.$$

Therefore, the modified Borel–Tanner distribution belongs to the class of power series distribution (see [13], p. 75) which contains for instance Bernoulli, binomial, geometric, negative binomial, Poisson and logarithmic series distributions.

On the other hand, [3] discussed discrete probability density functions $\Pr(X{=}x;\alpha)$ which obey the following relation for some functions B and D: if there exist B and D such that

$$\frac{d\Pr(X{=}x;\alpha)}{d\alpha} = B(\alpha)\left[x - D(\alpha)\right]\Pr(X{=}x;\alpha)\,, \tag{2.2}$$

then the mean μ coincides with $D(\alpha)$ and $\mu_2 = (d\mu/d\alpha)(1/B(\alpha))$ is the variance. Also, in that case

$$\mu_i = \mu_2\left[\frac{d\mu_{i-1}}{d\alpha}\,\frac{1}{d\mu/d\alpha} + (i-1)\,\mu_{i-2}\right], \qquad i = 2, 3, \ldots\,,$$

where μ_i is the i-th moment about the mean, which depends on α. Note that $\mu_0 = \mu$.

Now, observe that the MBT(α) distribution verifies (2.2) considering

$$B(\alpha) = \frac{1-\alpha}{\alpha(1+\alpha)}\,, \qquad D(\alpha) = \frac{\alpha}{1-\alpha}\,.$$

Then, the mean and the variance of the random variable following the probability function (2.1) are given by

$$E(X) = \frac{\alpha}{1-\alpha} \tag{2.3}$$

and

$$var(X) = \frac{\alpha(1+\alpha)}{(1-\alpha)^3}\,, \tag{2.4}$$

respectively. The previous expression for the mean of a MBT(α) distributed variable allows to write its probability mass function (pmf) as

$$\Pr(X{=}x) = C_x \cdot \frac{\theta^x(1+\theta)^{1+x}}{(1+2\theta)^{1+2x}}\,,$$

where $\theta = E(X) = \dfrac{\alpha}{1-\alpha}$.

Since

$$\frac{var(X)}{E(X)} = 1 + \frac{\alpha(3-\alpha)}{(1-\alpha)^2} > 1$$

we conclude that the distribution is overdispersed. Note that the proposed distribution is zero-inflated; that is, its proportion of 0's is greater than the proportion of 0's of a Poisson variate with the same mean. To see this we observe that the

zero-inflated index (see [19]) is $z_i = 1 - \frac{1-\alpha}{\alpha}\log(1+\alpha)$, which results greater than zero.

Additionally, the probability generating function is given by

$$G_X(z) = \frac{1+\alpha}{2\,\alpha\,z}\left[1 - \frac{\sqrt{1+\alpha\,(\alpha-4\,z+2)}}{1+\alpha}\right], \quad |z|<1\,.$$

The cumulative distribution function of a random variable following the probability function given in (2.1) is given by

$$\Pr(X\leq x) = 1 - \frac{\Gamma\left(x+\frac{3}{2}\right)(4\alpha)^{x+1}}{\Gamma(x+3)\,\sqrt{\pi}\,(1+\alpha)^{2x+3}}\;{}_2F_1\left(1,\,x+\frac{3}{2};\,x+3;\,\frac{4\alpha}{(1+\alpha)^2}\right), \tag{2.5}$$

where ${}_2F_1$ is the hypergeometric function given by

$${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty}\frac{\Gamma(a+k)\,\Gamma(b+k)\,\Gamma(c)}{\Gamma(a)\,\Gamma(b)\,\Gamma(c+k)}\,\frac{z^k}{k!}\,.$$

See the Appendix Section for details about this assert.

Some additional details about the hypergeometric function can be found in [18]. From (2.5) we get the survival function, $\Pr(X\geq x)$, and the failure or hazard rate can be easily obtained using (2.5) and (2.1).

Finally, observe that the probabilities can be computed from the recursion

$$\Pr(X=x) = \frac{2\alpha}{(1+\alpha)^2}\,\frac{2x-1}{x+1}\,\Pr(X=x-1)\,, \quad x=1,2,...\,,$$

being $\Pr(X=0) = \dfrac{1}{1+\alpha}$.

Since

$$\frac{\Pr(X=x)}{\Pr(X=x-1)} - \frac{\Pr(X=x+1)}{\Pr(X=x)} = \frac{-6\,\alpha}{(1+\alpha)^2}\,\frac{1}{(x+1)\,(x+2)} < 0\,,$$

we have that the distribution is log-convex (infinitely divisible) and has decreasing failure rate (DFR). See [9] and [22] for details. The fact that $\Pr(X=x)/\Pr(X=x-1)$, $x=1,2,...$, forms a monotone increasing sequence requires that $\Pr(X=x)$ be a decreasing sequence (see [12], p. 75). Therefore, the distribution is unimodal with modal value on zero. An overview of Figure 1 confirms this feature and that the shown plotting lines in the graph are similar to the ones corresponding to distributions of Poisson with expected value lower than 1.

Moreover, as any infinitely divisible distribution defined on non-negative integers is a compound Poisson distribution (see Proposition 9 in [15], we conclude that the probability function given in (2.1) is a compound Poisson distribution.

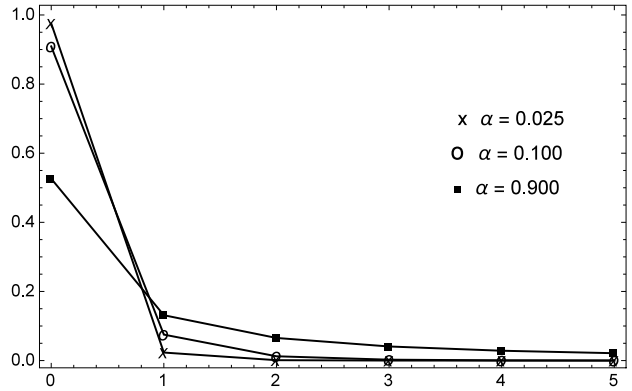

Figure 1: Some appearances (polygons) of the probability mass function for different values of the parameter α.

Furthermore, the infinitely divisible distribution plays an important role in many areas of statistics, for example, in stochastic processes and in actuarial statistics. When a distribution G is infinitely divisible then, for any integer $x \geq 2$, there exists a distribution G_x such that G is the x-fold convolution of G_x, namely, $G = G_x^{*x}$.

Since the new distribution is infinitely divisible, a lower bound for the variance can be obtained (see [12], p. 75), which is given by

$$var(X) \geq \frac{\Pr(X{=}1)}{\Pr(X{=}0)} = \frac{\alpha}{(1+\alpha)^2} .$$

3. INFERENCE FOR MBT DISTRIBUTION

In this section, different methods of estimation of the parameter of the distribution are studied.

Using (2.3) it is also simple to see that the estimator of α is given by

$$\widehat{\alpha}_1 = \frac{\bar{X}}{1+\bar{X}} , \tag{3.1}$$

where $\bar{X}$ is the sample mean.

An alternative to the method of moments is the method based on the zeros frequency. This method tends to work well only when the mode of the distribution is at zero and its proportion of zeros is relatively high ([2]). In this case we need only one equation in order to estimate the parameter of the distribution. It is straightforward obtaining an estimate for α based on the observed proportion of

zeros, denoted by $\tilde{p}_0$, as

$$\widehat{\alpha}_2 = \frac{1-\tilde{p}_0}{\tilde{p}_0} .$$

For each of sample sizes $n = 100$ and $n = 1000$, and for $\alpha = 0.1\,[0.1]\,0.9$, 4000 samples have been simulated, obtaining the estimates mean and squared error from both methods (Table 1). In both of them, the experimental bias is higher when α takes its lower values.

Table 1: MME (equivalently, MLE) and zero proportion estimate based on n simulations from a $MBT(\alpha)$.

α		$n = 100$		$n = 1000$	
		$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$
.1	mean	.250	.385	.2514	.379
	S.D	.034	.087	.011	.027
.2	mean	.308	.435	.310	.429
	S.D	.041	.096	.013	.030
.3	mean	.372	.489	.376	.4826
	S.D	.048	.106	.015	.032
.4	mean	.444	.545	.449	.539
	S.D	.053	.114	.017	.036
.5	mean	.521	.607	.529	.602
	S.D	.056	.126	.018	.039
.6	mean	.606	.676	.614	.668
	S.D	.059	.139	.018	.043
.7	mean	.696	.755	.706	.741
	S.D	.059	.157	.018	.047
.8	mean	.789	.838	.801	.821
	S.D	.055	.174	.016	.052
.9	mean	.883	.924	.899	.907
	S.D	.047	.186	.013	.057

Finally, the MLE are easy to derive since the MBT model belongs to the exponential family. Let now $\underline{x} = (x_1, x_2, ..., x_n)$ be a random sample obtained from model (2.1), then the log-likelihood function is proportional to

$$\ell(\alpha) \;\propto\; n\Big[\bar{x}\log\alpha - (2\bar{x}+1)\log(1+\alpha)\Big] .$$

The likelihood equation obtained from (3) is given by

$$\frac{\partial \ell(\alpha)}{\partial \alpha} = \frac{\bar{x}}{\alpha} - \frac{2\bar{x}+1}{1+\alpha} = 0 ,$$

from which we obtain the maximum likelihood estimator of α given again by (3.1). Therefore, as in the Poisson distribution the moment estimator coincides

with the maximum likelihood estimator. Additionally, the maximum likelihood estimator $\widehat{\alpha}$ of α is unique for all n.

Proposition 3.1. *The unique maximum likelihood estimator $\widehat{\alpha}$ of α is consistent and asymptotically normal and therefore*

$$\sqrt{n}(\widehat{\alpha}-\alpha) \xrightarrow{d} N\big(0, I^{-1}(\alpha)\big)\,,$$

where $I(\alpha)$ is the Fisher's information about α.

Proof: See Appendix. □

By using Corollary 3.11 in [16], p. 450, we conclude that the maximum likelihood estimator of α is asymptotically efficient.

4. THE MBT REGRESSION MODEL

The Poisson regression model has been extensively used as a benchmark for the analysis of discrete count models, together with some other models such as the negative binomial regression, Poisson-inverse Gaussian regression, and some other including special functions as hypergeometric, Kummer confluent, etc. when the endogenous variable take only nonnegative integer values. In practice, count data often display over-dispersion and therefore the Poisson regression model faults to provide an appropriate fit to the data. In this section we provide a regression model based on the use of the modified Borel–Tanner distribution presented in the previous sections of this work. We shall see that the new model is simple and competitive with the traditional Poisson regression model and also with the Negative Binomial model.

For that, let Y be now a response variable and $\boldsymbol{z}$ be an associated $q\times 1$ vector of covariates. The modified Borel–Tanner regression model for Y established that given z, Y follows a modified Borel–Tanner distribution with mean $\eta(\boldsymbol{z})$, a positive-valued function. We assume that $\eta(\boldsymbol{z})$ depends on a vector $\boldsymbol{\beta}$ of unknown regression coefficients. This parameterization has the appealing property that when $\eta(\boldsymbol{z})$ takes the common log-linear form $\eta(\boldsymbol{z}) = \exp(\boldsymbol{z}'\boldsymbol{\beta})$.

Writing the likelihood in terms of θ we have

$$\Pr(Y{=}\,y) \,=\, C_y\Big(p(\theta)^y\,\big(1-p(\theta)\big)^{y+1}\Big)\,, \tag{4.1}$$

for $y = 0, 1, ...$, being $p(\theta) = \theta/(1+2\theta)$ and $\theta > 0$.

Therefore, we assumed that $\theta = \eta(\boldsymbol{z}) = \exp(\boldsymbol{z}'\boldsymbol{\beta})$. Let now $(y_i, \boldsymbol{z_i})$ be a random sample of size n with counts y_i and a vector $\boldsymbol{z_i}$ of covariates for $i = 1, 2..., n$. Then, the log-likelihood function, assuming model (4.1) results

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \log \Pr(Y_i = y_i \mid \boldsymbol{z_i}; \boldsymbol{\beta}) \\ &\propto \sum_{i=1}^{n} (1 + y_i) \log\big(1 - p(\theta_i)\big) + \sum_{i=1}^{n} y_i \log p(\theta_i) \ ,\end{aligned}$$

where

$$p(\theta_i) = \frac{\exp\big(\sum_{s=1}^{q} z_{is}\beta_s\big)}{1 + 2\exp\big(\sum_{s=1}^{q} z_{is}\beta_s\big)} \ , \qquad 1 = 1, 2, ..., n \ .$$

Some computations provide that

$$\frac{\partial p(\theta_i)}{\partial \beta_j} = z_{ij}\, p(\theta_i)\big(1 - 2p(\theta_i)\big) \ , \qquad i = 1, 2, ..., n; \quad j = 1, 2, ..., q \ ,$$

from which the normal equations can be written as

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{z_{ij}\big(1 - 2p(\theta_i)\big)\big(y_i - (1 + 2y_i)\, p(\theta_i)\big)}{1 - p(\theta_i)} = 0 \ , \qquad j = 1, 2, ..., q \ .$$

The elements of the expected Fisher information matrix $\boldsymbol{I} = (I_{jl}),\ j, l = 1, ..., q$, about $\beta_j,\ j = 1, ..., q$, are given by

$$\begin{aligned}I_{jj} &= \sum_{i=1}^{n} z_{ij}^2 \frac{(1 + 2\theta_i)\,(1 + \theta_i)}{2\theta_i^2 + 4\theta_i + 3} \ , \qquad j = 1, ...q \ , \\ I_{jl} &= \sum_{i=1}^{n} z_{ij}\, z_{il} \frac{(1 + 2\theta_i)\,(1 + \theta_i)}{2\theta_i^2 + 4\theta_i + 3} \ , \qquad j, l = 1, ..., q, \quad j \neq l \ .\end{aligned}$$

The residuals can now be used to identify discrepancies between models and data, so the computation of the individual residuals from each observation can be useful to evaluate the model-fitting.

The common Pearson residuals are obtained by dividing the raw residuals by their scaled standard deviation, according to the model

$$\epsilon_i^P = \frac{y_i - \widehat{\theta}_i}{\sqrt{var\big(Y_i, \widehat{\theta}_i\big)}} \ , \qquad i = 1, 2, ..., n \ .$$

Here, $var(Y_i; \widehat{\theta}_i)$ is the variance of Y_i as a function of θ and $\widehat{\theta}_i$ is the maximum likelihood estimate of the i-th mean as fitted to the regression model.

With the aim of comparison between models, we consider as alternative options that the conditional distribution of the response variable can be described by Poisson, negative binomial or MBT distributions. This way, we obtain the corresponding Pearson residuals for each model:

a) Poisson: $\epsilon_i^P = \dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i}}, \quad i = 1, 2, ..., n.$

b) MBT: $\epsilon_i^P = \dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i \left(1 + \widehat{\theta}_i\right)\left(1 + 2\widehat{\theta}_i\right)}}, \quad i = 1, 2, ..., n.$

c) Negative binomial: $\epsilon_i^P = \sqrt{r}\,\dfrac{y_i - \widehat{\theta}_i}{\sqrt{\widehat{\theta}_i \left(r + \widehat{\theta}_i\right)}}, \quad i = 1, 2, ..., n.$

Another common choice of residuals is the signed square root of the contribution to the deviance goodness-of-fit statistic. This is given by $D = \sum_{i=1}^{n} d_i$, where

$$d_i \;=\; \mathrm{sgn}(\widehat{\theta}_i - y_i)\,\sqrt{2\left(\ell(y_i) - \ell(\widehat{\theta}_i)\right)}\;, \qquad i = 1, 2, ..., n\;,$$

where sgn is the function that returns the sign (plus or minus) of the argument. The $\ell(y_i)$ term is the value of the log likelihood when the mean of the conditional distribution for the i-th individual is the individual's actual score of the dependent variable. The $\ell(\widehat{\theta}_i)$ is the log likelihood when the conditional mean is substituted in the log likelihood. Usually the deviance divided by its degree of freedom is examined taking into account that a value much greater than one indicates a poorly fitting model. See for example [14].

It is well-known that for the Poisson distribution with parameter θ_i the deviance residuals are given by (see [8])

$$d_i \;=\; \mathrm{sgn}\big(y_i - \widehat{\theta}_i\big)\left[2\left(y_i \log\!\left(\frac{y_i}{\widehat{\theta}_i}\right) - \big(y_i - \widehat{\theta}_i\big)\right)\right]^{1/2}, \qquad i = 1, 2, ..., n\;.$$

For the MBT distribution proposed here the deviance residual are obtained as follows for each $i = 1, ..., n$:

$$d_i \;=\; \mathrm{sgn}\big(y_i - \widehat{\theta}_i\big)\left[2\left(\big(1 + y_i\big)\log\!\left(\frac{1 - p(y_i)}{1 - p(\widehat{\theta}_i)}\right) + y_i \log\!\left(\frac{p(y_i)}{p(\widehat{\theta}_i)}\right)\right)\right]^{1/2}.$$

For the negative binomial distribution, an expression for the deviance residuals can be found in [14]:

$$d_i \;=\; \mathrm{sgn}\big(y_i - \widehat{\theta}_i\big)\left[2\left(y_i \log\!\left(\frac{y_i}{\widehat{\theta}_i}\right) - \big(y_i + r\big)\log\!\left(\frac{y_i + r}{\widehat{\theta}_i + r}\right)\right)\right]^{1/2}.$$

In the three above considered cases we assume $y_i \neq 0$ for all i.

5. NUMERICAL ILLUSTRATION

In this section, we examine an application of the MBT regression model proposed here in order to analyse the number of deaths in truck's accidents ([17]).

In the present study, we model the number of deaths in the accident as the dependent variable. The explanatory variables are as follows: (1) the number of occupants; (2) a dummy variable for seat belt usage; (3) a set of dummy variables for rain, snow and fog, respectively; (4) a dummy variable for dark; (5) a dummy variable for weekdays; (6) a dummy variable for the first driver being drunk; (7) Dummy for the second driver being drunk and (8) a dummy variable for the first driver to be under 21 and finally, (9) a dummy variable for the first driver to be over 60. Due to the fact that the dependent variable is a count variable, data analysis including covariates would be a more appropriate method (see e.g. [7]; [6]; [5]; among others). Table 2 presents the estimates of the MBT, Poisson and Negative Binomial regression models, respectively.

Only for comparative purposes, we fit the MBT, Poisson and Negative Binomial distributions to this data set (see Table 3). We used the value of the log-likelihood function, the Akaike Information Criterion (AIC) (see [1]), the Bayesian Information Criterion (BIC) (see [20] and the Consistent Akaike Information Criterion (CAIC) (see [4]) to compare the estimated models.

Table 3 shows that the MBT model performs very well in fitting the distribution, compared to other uniparametric models Poisson, and provides a fit as good as that of the biparametric Negative Binomial model. Based on the BIC and CAIC, the MBT distribution fits the data better than NB, and NB distribution is better than Poisson. Furthermore, the MBT model presented is somewhat simpler than the NB and therefore it might appear to be preferable as a less complex model, taking into account the Ockham's razor principle (Jaynes, 1994).

The comparative study of Pearson residuals, deviance, log-likelihood and information criteria are also collected in Table 3. Note that the MBT model obtains a better result than the Negative Binomial when the Pearson statistic is the comparison criterion. Furthermore, graphical models diagnostics is now developed using the above residual expressions (see Figure 2).

In addition, one can be interested in testing the null that models are equally close to the actual model, against the alternative that one model is closer ([21]). The z-statistic is

$$Z = \frac{1}{\omega\sqrt{n}}\left(\ell_f(\widehat{\theta}_1) - \ell_g(\widehat{\theta}_2)\right),$$

Table 2: Parameter estimates for data in Li (23012) under the models considered. The response variable is Number of deaths in the accident. Variables statistically significant (at level < 0.05) in boldface.

Variable	**MBT model**			
	Estimate	S.D.	$\|t\|$-value	$\Pr(>\|t\|)$
Intercept	− 6.50341	0.23392	27.8019	< 0.001
Number of occupants	0.47966	0.04522	10.6070	< 0.001
Seat belt usage	− 1.16920	0.16188	7.2224	< 0.001
Rain	− 0.78854	0.34854	2.2623	0.0236
Snow	− 0.25206	0.59300	0.4250	0.6707
Fog	− 7.02683	37.2207	0.1887	0.8502
Dark	0.71680	0.16366	4.3798	< 0.001
Weekday	− 0.11630	0.16636	0.6991	0.4844
First driver drunk	1.37529	0.23324	5.8962	< 0.001
Second driver drunk	1.06292	0.22884	4.6446	< 0.001
First driver drug	0.70307	0.20635	3.4071	< 0.001
Second driver drug	0.65230	0.20220	3.2258	0.0012
Age driver < 21	0.09295	0.24036	0.3867	0.6989
Age driver > 60	1.03063	0.21773	4.7334	< 0.001
	Poisson model			
Intercept	− 6.41786	0.21997	29.1749	< 0.001
Number of occupants	0.44529	0.03434	12.9656	< 0.001
Seat belt usage	− 1.15769	0.15749	7.3507	< 0.001
Rain	− 0.80440	0.34272	2.3470	0.0189
Snow	− 0.28794	0.58335	0.4936	0.6215
Fog	− 5.02453	13.6193	0.3689	0.7121
Dark	0.73782	0.15848	4.6556	< 0.001
Weekday	− 0.14283	0.15969	0.8944	0.3710
First driver drunk	1.33410	0.21764	6.1297	< 0.001
Second driver drunk	1.03833	0.21365	4.8599	< 0.001
First driver drug	0.66979	0.19880	3.3691	< 0.001
Second driver drug	0.64306	0.19391	3.3162	< 0.001
Age driver < 21	0.11134	0.22843	0.4874	0.6259
Age driver > 60	1.02884	0.21108	4.8740	< 0.001
	Negative Binomial model			
r	0.12527	0.04463	2.8067	0.0050
Intercept	− 6.54213	0.24326	26.8932	< 0.001
Number of occupants	0.49415	0.05028	9.8269	< 0.001
Seat belt usage	− 1.19020	0.16753	7.1042	< 0.001
Rain	− 0.74537	0.35349	2.1085	0.0349
Snow	− 0.19223	0.60245	0.3190	0.7496
Fog	− 2.24947	3.45645	0.6508	0.5151
Dark	0.70502	0.16911	4.1689	< 0.001
Weekday	− 0.09364	0.17342	0.5399	0.5892
First driver drunk	1.39773	0.24756	5.6458	< 0.001
Second driver drunk	1.09946	0.24494	4.4885	< 0.001
First driver drug	0.72924	0.21385	3.4099	< 0.001
Second driver drug	0.63340	0.21050	3.0090	0.0026
Age driver < 21	0.08526	0.24970	0.3414	0.7327
Age driver > 60	1.03876	0.22568	4.6026	< 0.001

Table 3: Summaries of fitting measures results for the models considered.

Criterion	Model		
	Poisson	Neg. Bin.	MBT
$\left(\epsilon_i^P\right)^2$	55196.5	54638.5	55225.8
Deviance	492.792	387.865	421.177
Deviance/df	0.00825	0.00649	0.00705
$\ell_{\max}$	-1090.49	-1074.91	-1077.42
AIC	2208.97	2179.81	2182.85
BIC	2334.94	2314.77	2308.81
CAIC	2348.94	2329.77	2322.81

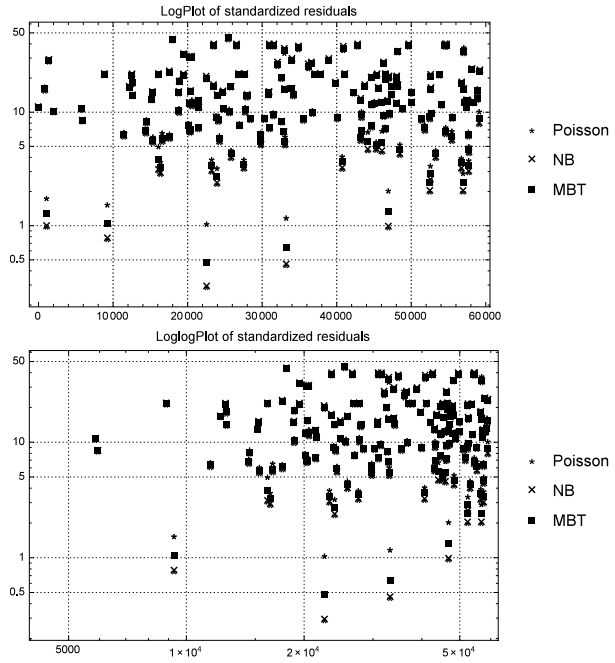

Figure 2: LogPlot and LoglogPlot of standardized residuals for the models considered.

where

$$\omega^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{f(x_i\,|\,\widehat{\theta}_1)}{g(x_i\,|\,\widehat{\theta}_2)}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f(x_i\,|\,\widehat{\theta}_1)}{g(x_i\,|\,\widehat{\theta}_2)}\right)\right]^2$$

and f and g represent here the MBT and the alternative distributions, respectively.

Due to the asymptotic normal behavior of the Z statistic under the null, rejection of the test in favor of f happens, with significance level α, when $Z > z_{1-\alpha}$ being $z_{1-\alpha}$ the $(1-\alpha)$ quantile of the standard normal distribution.

Table 4 shows the results obtained for each comparison by means of the Voung test. The MBT model is preferred to the Poisson model and we cannot reject the null that the models, Negative binomial and MBT, are statistically the same.

Table 4: Voung test results.

	Z-score	p-value
MBT vs Poisson	2.29834	0.01
MBT vs Neg.bin.	−0.87924	0.81

6. CONCLUSIONS

This paper has introduced a modified version of the Borel–Tanner distribution which takes its values from the non-negative integers, in contrast with the original Borel–Tanner distribution which is restricted to the positive integers.

We obtain an over-dispersed distribution (its variance is larger than its mean) depending on just one parameter, which is also unimodal with mode at zero. Furthermore the distribution is infinitely divisible (log-convex) and therefore it may be considered as a compound Poisson distribution. Some other properties based on results in [3] are also verified.

In addition, a simple reparameterization of the MBT distribution allows to incorporate in an easy way covariates into the model.

In this paper, a numerical application is provided, where both the Poisson and the negative binomial model-fitting are compared to the MBT. The practical use of the modified Borel–Tanner distribution here proposed does not only bring a significant improvement relative to the Poisson distribution but also a wider flexibility due to its main properties, as for instance its over-dispersion. The MBT distribution is found to be a better model to describe the data used in this paper than the Poisson and the negative binomial, according to their BIC and CAIC values.

APPENDIX

In this Appendix we provide a proof for the cumulative distribution function of the distribution and Proposition 3.1.

Proof of the cdf of the distribution: We have that

$$F(x) \,=\, 1 - \sum_{j=x+1}^{\infty} \frac{\Gamma(2j+1)}{\Gamma(j+2)\,\Gamma(j+1)}\,\frac{\alpha^j}{(1+\alpha)^{2j+1}}\,.$$

Now by putting $k = j - x - 1$ and using the identity

$$\Gamma(2m) \,=\, \frac{1}{\sqrt{2\pi}}\,2^{2m-1}\Gamma(m)\,\Gamma(m+1/2)\,,$$

which appears in [13], p. 7, we obtain the result after some computations. □

Proof of Proposition 3.1: The discrete distribution with probability function given in (2.1) satisfies the regularity conditions (see [16], p. 449)) under which the unique maximum likelihood estimator $\widehat{\alpha}$ of α is consistent and asymptotically normal. They are simply verified in the following way. Firstly, the parameter space $(0,1)$ is a subset of the real line and the range of x is independent of α. By using expression (3) it is easy to show that $E\big(\frac{\partial \log \Pr(X=x;\alpha)}{\partial \alpha}\big) = 0$. Now, since $\frac{\partial^2 \ell(\alpha)}{\partial \alpha^2}\big|_{\alpha=\widehat{\alpha}} < 0$, the Fisher's information is positive. Finally, by taking $M(x) = 2x/\alpha^3$ we have that

$$\left|\frac{\partial^3 \log \Pr(X{=}\,x;\alpha)}{\partial \alpha^3}\right| = \left|\frac{2x}{\alpha^3} - \frac{2(2x+1)}{(1+\alpha)^3}\right| \leq M(x)\,,$$

with $E(M(X)) = 2/(\alpha(1-\alpha)) < \infty$. Hence the proposition. □

ACKNOWLEDGMENTS

The authors would like to express their gratitude to two anonymous referees for their relevant and useful comments. Authors are funded by grant ECO2013–47092 (MINECO, Spain).

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

[2] Alanko, T. and Duffy, J. (1996). Compound binomial distributions for modelling consumption data, *The Statistician*, **45**(3), 269–286.

[3] Bardwell, E. (1960). On Certain Characteristics of Some Discrete Distributions, *Biometrika*, **47**(3/4), 473–475.

[4] Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**(3), 345–379.

[5] Cartwright, W.; Hu, S.T. and Huang, L.-F. (1992). Medigap insurance coverage on the use of medical services of the elderly, *Applied Economics*, **2**(4), 529–539.

[6] Christensen, S; Long, S. and Rodgers, J. (1987). Acute health care costs for the aged Medicare population: overview and policy options, *Millbank Quarterly*, **65**(3), 397–425.

[7] Duan, N.; Manning, W.; Morris, C. and Newhouse, J. (1983). A comparison of alternative models for the demand for medical care, *Journal of Business and Economic Statistics*, **1**, 115–126.

[8] Dunteman, G.H. and Ho, M.-H.R. (1994). *Probability Theory: The Logic of Science*, `http://omega.math.albany.edu:8008/JaynesBook.html`.

[9] Gupta, P.L.; Gupta, R.C. and Tripathi, R.C. (1997). On the monotonic properties of discrete failure rates, *Journal of Statistical Planning and Inference*, **65**, 255–268.

[10] Haight, F.A. (1961). A distribution analogous to the Borel–Tanner, *Biometrika*, **48**(1–2), 167–168.

[11] Haight, F.A. and Breuer, M.A. (1960). The Borel–Tanner distribution, *Biometrika*, **47**(1–2), 143–150.

[12] Johnson, N.L. and Kotz, S. (1982). Developments in Discrete Distributions, 1969–1980, *International Statistical Review*, **50**(1), 71–101.

[13] Johnson, N.L.; Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions, Third Edition*, John Wiley, New York.

[14] Jong, P.D. and Heller, G.Z. (2008). *Generalized Linear Models for Insurance Data*, Cambridge University Press, Cambridge.

[15] Karlis, D. and Xekalali, E. (2005). Mixed Poisson Distributions, *International Statistical Review*, **73**(1), 35–58.

[16] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Springer, New York.

[17] Li, S. (2012). Traffic safety and vehicle choice: quantifying the effects of the "Arms Race" on American roads, *Journal of Applied Econometrics*, **27**(1), 34–62.

[18] Olver, F.W.J.; Lozier, D.W.; Boisvert, R.F. and Clark, C.W.; Eds. (2010). *NIST Handbook of Mathematical Functions*, Cambridge University Press. http://dlmf.nist.gov/

[19] Puig, P. and Valero, J. (2006). Count data distributions: Some characterizations with applications, *Journal of the American Statistical Association*, **101**(437), 332–340.

[20] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**(2), 461–464.

[21] Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**(2), 307–333.

[22] Warde, W.D. and Katti, S.K. (1971). Infinite divisibility of discrete distributions II, *The Annals of Mathematical Statistics*, **42**(3), 1088–1090.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 443–453

ON HIGHLY D-EFFICIENT DESIGNS WITH NON-NEGATIVELY CORRELATED OBSERVATIONS

Authors: Krystyna Katulska
– Faculty of Mathematics and Computer Science,
Adam Mickiewicz University in Poznań,
Umultowska 87, 61-614 Poznań, Poland
krakat@amu.edu.pl

Łukasz Smaga
– Faculty of Mathematics and Computer Science,
Adam Mickiewicz University in Poznań,
Umultowska 87, 61-614 Poznań, Poland
ls@amu.edu.pl

Received: September 2015 Accepted: October 2015

Abstract:

- In the most difficult case where the number of observations $n \equiv 3 \pmod 4$, high D-efficiency of certain chemical balance weighing designs under completely symmetric covariance matrix of errors is shown. It is also proved that D-optimal design may depend on the values of the correlation coefficient.

Key-Words:

- *chemical balance weighing design; correlated observations; D-efficiency; D-optimality; Hadamard matrix; simulated annealing algorithm.*

AMS Subject Classification:

- 62K05, 05B20.

1. INTRODUCTION

Let us introduce a model of the chemical balance weighing design. Assume that $\mathcal{M}_{n\times p}(\{-1,1\})$ denotes the set of $n\times p$ matrices whose entries are all equal to 1 or -1. A linear model of the chemical balance weighing design is as follows: $\mathbf{y}=\mathbf{X}\,\mathbf{b}+\mathbf{e}$, where $\mathbf{y}=(y_1,...,y_n)'$ is a vector of observations, $\mathbf{X}=(x_{ij})\in \mathcal{M}_{n\times p}(\{-1,1\})$ is a design matrix of full column rank $(n\geqslant p)$, $\mathbf{b}=(b_1,...,b_p)'$ is a vector of unknown parameters, and $\mathbf{e}=(e_1,...,e_n)'$ is a vector of errors. In a chemical balance, if the j-th object is placed on the left (resp. right) pan during the i-th weighing operation, then $x_{ij}=-1$ (resp. $x_{ij}=1$). Moreover, suppose that $\mathrm{E}(e_i)=0$, $i=1,...,n$ and $\mathrm{Cov}(\mathbf{e})=\sigma^2\mathbf{G}$, where $\sigma>0$ is an unknown parameter and $\mathbf{G}$ is a known positive definite matrix of size n.

Among all designs in $\mathcal{M}_{n\times p}(\{-1,1\})$, we would like to find the optimal design with respect to certain criterion. Optimal weighing designs depend significantly on the form of $\mathbf{G}$. In the literature, optimal weighing designs are mostly considered under the following forms of $\mathbf{G}$: the identity matrix (the errors are uncorrelated and have equal variances; see Banerjee, 1975; Cheng, 2014; Ehlich, 1964; Galil and Kiefer, 1980; Neubauer and Pace, 2010), the diagonal matrix (the errors are uncorrelated and may have different variances; see Ceranka *et al.*, 2006; Graczyk, 2011, 2012), the completely symmetric matrix (the errors are equally correlated and have equal variances; see Ceranka and Graczyk, 2011, 2015; Katulska and Smaga, under review; Masaro and Wong, 2008a, 2008b, 2008c; Smaga, 2015), and the covariance matrix of an AR(1) process (see Angelis *et al.*, 2001; Katulska and Smaga, 2012, 2013; Li and Yang, 2005; Smaga, 2014; Yeh and Lo Huang, 2005). Some applications of optimal weighing designs and real data examples of their use can be found in Banerjee (1975), Cheng (2014), Graczyk (2013) and Jenkins and Chanmugam (1962).

In this paper we consider D-efficiency of chemical balance weighing designs, when the errors are equally correlated and have equal variances. Under this assumption, the matrix $\mathbf{G}$ is of the form:

$$\mathbf{G}=(1-\rho)\,\mathbf{I}_n+\rho\,\mathbf{1}_n\mathbf{1}_n'\,, \tag{1.1}$$

where $\rho\in[0,1)$ is a known parameter, $\mathbf{I}_n$ is the n-dimensional unit matrix, and $\mathbf{1}_n$ is the n-dimensional column vector of ones. For given ρ, the matrix $\mathbf{G}$ is positive definite and $\mathbf{G}^{-1}=c(\mathbf{I}_n-r\,\mathbf{1}_n\mathbf{1}_n')$, where $c=1/(1-\rho)$ and

$$r=\frac{\rho}{1+(n-1)\,\rho}\,. \tag{1.2}$$

Following the definition of Bulutoglu and Ryan (2009), the D-efficiency of a design $\mathbf{X}\in\mathcal{M}_{n\times p}(\{-1,1\})$ is given by the formula

$$\text{D-eff}(\mathbf{X})=\left[\frac{\det(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})}{\max_{\mathbf{Y}\in\mathcal{M}_{n\times p}(\{-1,1\})}\det(\mathbf{Y}'\mathbf{G}^{-1}\mathbf{Y})}\right]^{1/p}.$$

If D-eff$(\mathbf{X}) = 1$, then $\mathbf{X}$ is D-optimal. Unfortunately, the denominator of D-eff$(\mathbf{X})$ is usually not known, and hence we can not calculate D-efficiency of designs. However, Katulska and Smaga (under review) established the lower bound for D-efficiency of a design $\mathbf{X}$ given by

$$\text{D}^*\text{-eff}(\mathbf{X}) \;=\; \frac{\Big[\det\big(\mathbf{X}'(\mathbf{I}_n - r\,\mathbf{1}_n\mathbf{1}_n')\,\mathbf{X}\big)\Big]^{1/p}}{n}\,,$$

where r is as in (1.2), and they used it to show that designs constructed by Masaro and Wong (2008a) and certain other designs are highly D-efficient for many values of design parameters namely n, p and ρ. Nevertheless, they did not consider the most difficult case $n \equiv 3 \pmod 4$, which is different of the others. In the present paper, this case is of interest to us.

The remainder of this paper is organized as follows. In Section 2, we show that certain design constructed by Masaro and Wong (2008a) is highly D-efficient, when the number of observations $n \equiv 3 \pmod 4$ and it is appropriately large or appropriately larger than the number of objects. Section 3 contains simulation study, which suggests that design is D-optimal in many cases, but also indicates situations where are D-better designs than it. A special case, where different designs are D-optimal for different values of ρ, is presented in Section 4. The paper is concluded in Section 5.

2. D-EFFICIENT DESIGNS WHEN $\mathbf{n \equiv 3}$ (mod 4)

Assume that $n \equiv 3 \pmod 4$, $\rho \in [0, 1)$ and $\mathbf{H}_{n+1}$ is a normalized Hadamard matrix of order $n+1$, i.e. all entries of its first row and first column are all equal to one. Let $\mathbf{W}$ be a matrix received by deleting the first row and column of $\mathbf{H}_{n+1}$. We form a design $\mathbf{L}$ from p columns of $\mathbf{W}$. From the results of Ehlich (1964) and Galil and Kiefer (1980), the design $\mathbf{L}$ is D-optimal in $\mathcal{M}_{n\times p}(\{-1, 1\})$, when $\rho = 0$ and $n \geqslant 2p - 5$. Masaro and Wong (2008a) proved that the design $\mathbf{L}$ is D-optimal for all $\rho > 0$ in

$$\mathcal{D}_3 \;=\; \Big\{\mathbf{X} \in \mathcal{M}_{n\times p}(\{-1, 1\})\colon\, \mathbf{X}'\mathbf{X} = (n+1)\,\mathbf{I}_p - \mathbf{1}_p\mathbf{1}_p'\Big\}\,.$$

But, if $n < 2p - 5$, then $\mathbf{L}$ may not be D-optimal when $\rho = 0$, and hence we can conclude that the similar situation may have place when $\rho > 0$. As we shall see in the next sections, that conjecture seems to be true and the result of Masaro and Wong (2008a) can not be extended from the subclass $\mathcal{D}_3$ to the class $\mathcal{M}_{n\times p}(\{-1, 1\})$. However, we show that the design $\mathbf{L}$ is highly D-efficient in many cases.

The design $\mathbf{L}$ has the following properties $\mathbf{L}'\mathbf{L} = (n+1)\mathbf{I}_p - \mathbf{1}_p\mathbf{1}_p'$ and $\mathbf{L}'\mathbf{1}_n = -\mathbf{1}_p$. The matrix $\mathbf{L}'(\mathbf{I}_n - r\,\mathbf{1}_n\mathbf{1}_n')\mathbf{L}$ has eigenvalues $n+1$ and $n+1-(1+r)p$ with multiplicities $p-1$ and 1 respectively. Hence

$$\text{D}^*\text{-eff}(\mathbf{L}) \;=\; \frac{n+1}{n}\,\left[\frac{n-p+1-pr}{n+1}\right]^{1/p}, \tag{2.1}$$

where r is given in (1.2). The lower bounds for D*-eff$(\mathbf{L})$ are given in the following theorem.

Theorem 2.1. *Let $n \equiv 3$ (mod 4), $n \geqslant 7$, $p = 2, ..., n-1$, $\rho \in (0,1)$ and $\text{Cov}(\mathbf{e}) = \sigma^2\mathbf{G}$, where $\mathbf{G}$ is given by (1.1). Then,* D*-eff$(\mathbf{L})$ *decreases, when ρ increases, and* D*-eff$(\mathbf{L}) > 0.82$. *Moreover, if $p \leqslant (n-1)/2$; $n-3$; $n-2$, then* D*-eff$(\mathbf{L}) > 0.93; 0.92; 0.88$, *respectively.*

Proof: Observe that r is an increasing function of ρ. Hence, D*-eff$(\mathbf{L})$ decreases, when ρ increases, which implies

$$\text{D}^*\text{-eff}(\mathbf{L}) \;>\; \frac{n+1}{n}\,\left[\frac{n-p}{n}\right]^{1/p}$$

(the right hand side of (2.1) as $\rho \to 1$). The derivative of the function f, $f\colon (1,n) \to \mathbb{R}$ defined by $f(x) = \big[(n-x)/n\big]^{1/x}$ is

$$f'(x) \;=\; -\big[(n-x)/n\big]^{1/x}\, x\big((n/x-1)\log(1-x/n)+1\big)\big/\big(x^2(n-x)\big)\,.$$

Consider the function g, $g\colon (1,\infty) \to \mathbb{R}$ given by $g(x) = (x-1)\log(1-1/x)+1$. It is easy to calculate that $g'(x) = 1/x + \log(1-1/x)$, $\lim_{x\to\infty} g'(x) = 0$ and $g''(x) = 1\big/\big((x-1)x^2\big)$. Thus, g is decreasing. So $g(x) > 0$, because $\lim_{x\to\infty} g(x) = 0$. Since $n/x > 1$ for all $x \in (1,n)$, it follows that $f'(x) < 0$. So,

$$\text{D}^*\text{-eff}(\mathbf{L}) \;>\; \frac{n+1}{n}\,f(n-1) \;=\; (n+1)\left[\frac{1}{n}\right]^{n/(n-1)}.$$

The function h, $h\colon (6,\infty) \to \mathbb{R}$ is defined by $h(x) = (x+1)\,[1/x]^{x/(x-1)}$. Its derivative is equal to

$$h'(x) \;=\; -\frac{[1/x]^{x/(x-1)}\,\big(2(x-1)+(x+1)\log(1/x)\big)}{(x-1)^2}\,.$$

If $h_1(x) = 2(x-1)+(x+1)\log(1/x)$, then $h_1'(x) = 1-1/x+\log(1/x)$ and $h_1''(x) = (1-x)/x^2 < 0$. Hence, since $h_1'(6)$ is negative, $h_1'(x) < 0$ for all $x > 6$. So, h_1 is decreasing, and $h_1(6) < 0$, which imply $h_1(x) < 0$. Thus, $h'(x) > 0$ and h is increasing. Hence, we conclude that D*-eff$(\mathbf{L})$ is greater than $h(7) = 0.8263$. In a similar way, we can prove the rest of the claim. □

Theorem 2.1 and the examples (see Table 1) imply **L** is a design with high D-efficiency, when n is appropriately large or appropriately larger than p. From the examples, we conclude that, when ρ increases, the decrease of D*-eff(**L**) can be at most a few percent (see Table 1). As p increases, the decrease of D*-eff(**L**) can be quite large, but when n increases, it decreases. Moreover, the lower bound for D-efficiency of **L** increases, when n increases. From the examples, we also observe that D*-eff(**L**) is often much greater than the lower bounds for it obtained in Theorem 2.1.

Table 1: The lower bound for D-efficiency of design **L**.

ρ	n, p							
	$11, 2$	$11, 10$	$15, 2$	$15, 14$	$19, 2$	$19, 18$	$103, 2$	$103, 102$
0	0.9958	0.9119	0.9977	0.9194	0.9986	0.9262	0.9999	0.9713
0.01	0.9949	0.9077	0.9971	0.9152	0.9981	0.9221	0.9999	0.9685
0.1	0.9908	0.8860	0.9948	0.8970	0.9966	0.9064	0.9998	0.9655
0.2	0.9891	0.8757	0.9940	0.8897	0.9961	0.9010	0.9998	0.9651
0.3	0.9883	0.8700	0.9936	0.8860	0.9960	0.8984	0.9998	0.9650
0.4	0.9878	0.8665	0.9934	0.8838	0.9959	0.8969	0.9998	0.9649
0.5	0.9875	0.8641	0.9933	0.8824	0.9958	0.8959	0.9998	0.9649
0.6	0.9872	0.8623	0.9932	0.8813	0.9957	0.8952	0.9998	0.9649
0.7	0.9871	0.8609	0.9931	0.8805	0.9957	0.8947	0.9998	0.9648
0.8	0.9869	0.8598	0.9930	0.8799	0.9957	0.8943	0.9998	0.9648
0.9	0.9868	0.8590	0.9930	0.8794	0.9957	0.8940	0.9998	0.9648
0.99	0.9867	0.8583	0.9930	0.8791	0.9956	0.8938	0.9998	0.9648

3. SIMULATIONS

In this section we compare the design **L** with the best designs found by simulated annealing algorithm (SA algorithm) proposed by Angelis *et al.* (2001). It is an algorithm for searching optimal designs with very good performance. The SA algorithm was executed at least 1000 times for many values of n, p and ρ. The initial parameters of this algorithm were chosen according to the recommendations of Angelis *et al.* (2001).

Simulations and Theorem 2.1 indicate that the design **L** is D-optimal when $n > 2p - 5$ and $\rho \in [0, 1)$, and sometimes when $n = 2p - 5$ and $\rho < \alpha < 0.06$ for some α (in these situations the SA algorithm did not find D-better design than the design **L**). In the other cases, using the SA algorithm, we found D-better designs than the design **L**. We can observe that the inner product of any two columns of those designs is equal to ± 1 (for the vast majority of columns) or ± 3, and the same observation holds for the sum of elements in any column. Some examples of

the best designs found by SA algorithm are given in the Supplementary materials (Appendix A). As an example, Figure 1 depicts the results of our simulations when $n = 15$, $\rho = 0.99$ and $p = 2, ..., 14$. For the other values of parameter ρ, the situation is similar as for $\rho = 0.99$. However, when there are D-better designs than the design **L**, SA algorithm finds sometimes different designs for different values of ρ.

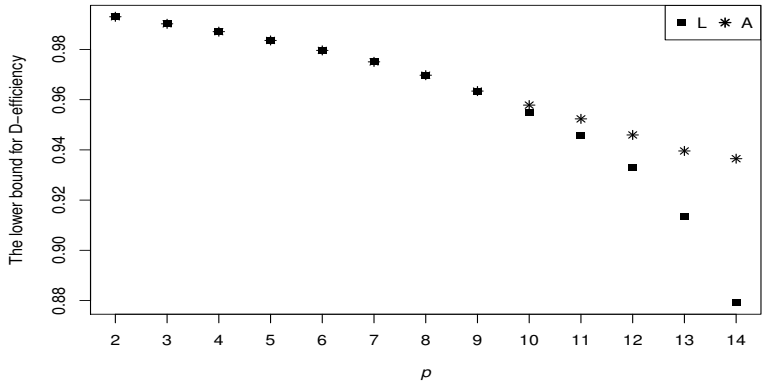

Figure 1: The lower bound for D-efficiency of design **L** (L) and the best designs found by SA algorithm (A) for $n = 15$ and $\rho = 0.99$.

For example, when $n = 15$ and $p = 10$, SA algorithm found, as the best design under D-optimality criterion, the design **T** for small values of $\rho > 0$, and the design **S** for the other values of this parameter (see Figure 2). The designs **T** and **S** are given in Appendix B.

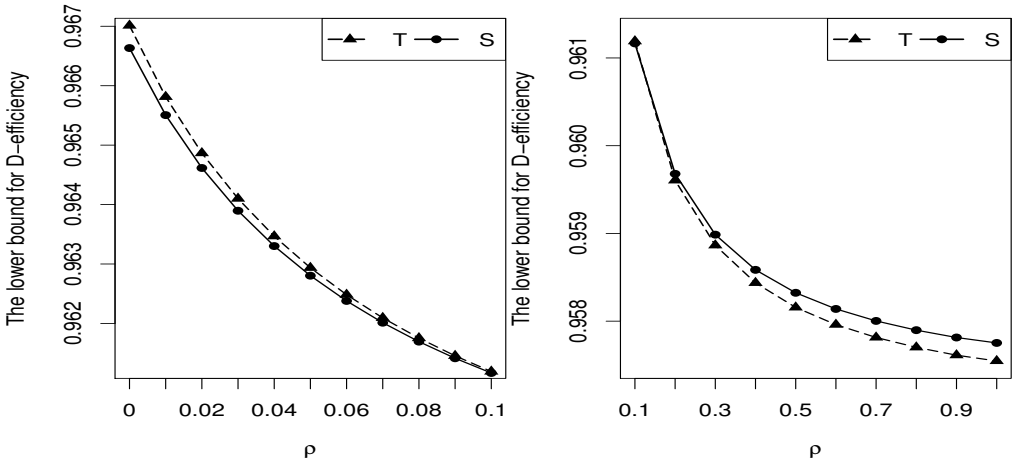

Figure 2: The lower bound for D-efficiency of the best designs **T** (T) and **S** (S) found by SA algorithm for $n = 15$ and $p = 10$.

These designs are D-better than the design $\mathbf{L}$ for "almost all" $\rho > 0$. It is important that the design $\mathbf{T}$, which seems to be D-optimal for certain small $\rho > 0$, is D-optimal, when $\rho = 0$ (it follows from Theorem in Galil and Kiefer, 1980), and the design $\mathbf{S}$, which seems to be D-optimal for the other values of ρ, is not D-optimal design for $\rho = 0$. This indicates that the D-optimal design for greater values of ρ does not have to be D-optimal when $\rho = 0$, and conversely, in some cases. The veracity of that conjecture is confirmed in the following section.

4. CASE STUDY

In this section we consider a special case where different designs are D-optimal for different values of parameter ρ. This (theoretically) confirms the conjecture from Section 3.

When $n = 7$ and $p = 6$, the simulations suggest that the design $\mathbf{L}$ is D-optimal for $\rho \in [0, 1/18]$, and the design ($-$ denotes -1 and $+$ represents 1)

$$\mathbf{A} = \begin{pmatrix} - & + & + & + & + & + \\ - & - & + & - & - & + \\ - & - & - & + & + & - \\ - & + & - & + & - & + \\ + & + & - & - & + & - \\ + & - & - & + & + & + \\ + & + & + & + & - & - \end{pmatrix} \tag{4.1}$$

is D-optimal for $\rho \in [1/18, 1)$. This information helped to prove the following theorem.

Theorem 4.1. *If* $\mathrm{Cov}(\mathbf{e}) = \sigma^2\mathbf{G}$, *where* $\mathbf{G}$ *is given by* (1.1), *then any D-optimal design in* $\mathcal{M}_{7\times 6}(\{-1, 1\})$ *for* $\rho = 0$ *is not D-optimal design for* $\rho > 1/18$, *and conversely.*

Proof: Let $\mathbf{X}$ be an arbitrary D-optimal design for $\rho = 0$ in $\mathcal{M}_{7\times 6}(\{-1, 1\})$. The normalization (see Galil and Kiefer, 1980) refers to the following operations on $\mathbf{X}$: multiplying on the right by a diagonal matrix of ± 1's and/or a permutation matrix, which permutes rows and corresponding columns of $\mathbf{X}'\mathbf{X}$ and multiplies some entries of $\mathbf{X}'\mathbf{X}$ by -1. The results of Ehlich (1964) and Theorem in Galil and Kiefer (1980) imply the matrix $\mathbf{X}'\mathbf{X}$ is equal to $8\mathbf{I}_6 - \mathbf{1}_6\mathbf{1}_6'$ or to

$$\begin{pmatrix} 8\mathbf{I}_4 - \mathbf{1}_4\mathbf{1}_4' & -\mathbf{1}_4 & -\mathbf{1}_4 \\ -\mathbf{1}_4' & 7 & 3 \\ -\mathbf{1}_4' & 3 & 7 \end{pmatrix}$$

after normalization. We see that the normalization leaves $\det(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})$ unchanged, so we can assume that $\mathbf{X}'\mathbf{X}$ has one of the above forms. By Masaro and Wong (2008a), denote by $\mathcal{D}_4(7,6)$ the subclass of such designs. Proposition 5 (a) of Masaro and Wong (2008a) implies the design in that subclass, for which the sum of elements in each column is equal to -1, is D-optimal in $\mathcal{D}_4(7,6)$ for all $\rho > 0$. In the paragraph before Theorem 2.1, we noticed $\mathbf{L}'\mathbf{L} = 8\mathbf{I}_6 - \mathbf{1}_6\mathbf{1}_6'$ and $\mathbf{L}'\mathbf{1}_7 = -\mathbf{1}_6$. So, $\mathbf{L}$ is D-optimal in $\mathcal{D}_4(7,6)$ for all $\rho > 0$. Consider the design $\mathbf{A}$ given by (4.1). It can be calculated that $\mathbf{A}$ does not belong to $\mathcal{D}_4(7,6)$,

$$\det(\mathbf{A}'\mathbf{G}^{-1}\mathbf{A}) = c^6(61440 - 98304\,r)$$

and

$$\det(\mathbf{L}'\mathbf{G}^{-1}\mathbf{L}) = c^6(65536 - 196608\,r)\,,$$

where $c = 1/(1-\rho)$ and $r = \rho/(1+6\rho)$. Comparing these two determinants, we obtain $\mathbf{A}$ is D-better than $\mathbf{L}$ for all $\rho > 1/18$. Therefore, for all $\rho > 1/18$, the design $\mathbf{A}$ is D-better than any D-optimal design for $\rho = 0$ in $\mathcal{M}_{7\times 6}(\{-1,1\})$. So, the first part of the claim is proved. Let now $\mathbf{Y} \in \mathcal{M}_{7\times 6}(\{-1,1\})$ be an arbitrary D-optimal design for $\rho > 1/18$. From the above considerations, we conclude that for all $\rho > 1/18$,

$$\det(\mathbf{Y}'\mathbf{G}^{-1}\mathbf{Y}) \geqslant \det(\mathbf{A}'\mathbf{G}^{-1}\mathbf{A}) > \det(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X})$$

for any D-optimal design $\mathbf{X}$ for $\rho = 0$. Thus, $\mathbf{Y}$ is not D-optimal, when $\rho = 0$. The proof is complete. □

Theorem 4.1 shows that we can not assume a priori that there is a design which is optimal for all $\rho \in [0,1)$. This indicates the reasonableness of searching optimal designs for different values of parameter ρ.

5. CONCLUSION

In this paper we showed that the design $\mathbf{L}$ constructed by Masaro and Wong (2008a) is highly D-efficient in many cases when the number of observations $n \equiv 3 \pmod 4$. Simulations conducted by SA algorithm (Angelis *et al.*, 2001) suggest that the design $\mathbf{L}$ is D-optimal when the number of observations is appropriately large or appropriately larger than the number of objects. In the other cases, however, we found D-better designs than $\mathbf{L}$. Nevertheless, the "D-efficiency" advantage of those designs over $\mathbf{L}$ is negligible for appropriately large n. For smaller number of n (e.g., $n = 7, 11, 15$), this advantage is evident, and hence the best designs found by SA algorithm are listed in the Supplementary materials. Even though those designs or the design $\mathbf{L}$ are not D-optimal, they may be safety used in practice through their high D-efficiency.

APPENDIX A. SUPPLEMENTARY MATERIAL

Supplementary material lists the examples of the best chemical balance weighing designs under D-optimality criterion found by simulated annealing algorithm. It is available at the webpage `http://www.staff.amu.edu.pl/~ls/str_en.html`.

APPENDIX B. DESIGNS T AND S

Let $-$ denote -1 and $+$ represent 1.

$$\mathbf{T} = \begin{pmatrix} - & + & + & + & + & + & + & + & + & + \\ - & - & + & - & + & - & - & + & - & + \\ + & - & - & + & - & - & + & + & - & - \\ - & - & + & - & + & - & + & - & + & - \\ - & - & - & - & - & + & - & + & + & - \\ - & - & - & + & - & - & + & - & + & + \\ - & + & - & + & + & - & - & - & + & - \\ + & - & + & + & + & + & + & - & - & - \\ + & - & + & - & - & + & - & - & + & + \\ + & + & - & - & + & - & - & - & - & + \\ - & + & - & - & - & + & + & - & - & + \\ - & + & + & - & - & - & + & + & - & - \\ + & + & - & - & + & + & + & + & + & - \\ + & + & + & + & - & - & - & + & + & + \\ - & + & + & + & - & + & - & - & - & - \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} - & - & - & - & + & + & - & + & - & - \\ - & + & - & + & + & + & - & + & + & + \\ + & + & - & + & - & + & - & - & - & - \\ + & + & - & - & - & - & + & + & - & - \\ + & - & - & + & + & + & + & + & - & + \\ - & - & - & - & - & + & + & - & + & - \\ + & + & + & - & + & + & + & - & - & + \\ - & + & + & - & - & + & + & + & + & + \\ - & + & - & + & + & - & + & - & + & - \\ - & + & - & - & - & - & - & - & - & + \\ + & - & - & - & + & - & - & - & + & + \\ - & - & + & + & + & - & + & - & - & - \\ - & - & + & + & - & - & - & + & - & + \\ + & + & + & - & + & - & - & + & + & - \\ + & - & + & + & - & + & - & - & + & - \end{pmatrix}.$$

REFERENCES

[1] Angelis, L.; Bora-Senta, E. and Moyssiadis, C. (2001). Optimal exact experimental designs with correlated errors through a simulated annealing algorithm, *Computational Statistics & Data Analysis*, **37**, 275–296.

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

[3] Bulutoglu, D.A. and Ryan, K.J. (2009). D-optimal and near D-optimal 2^k fractional factorial designs of resolution V, *Journal of Statistical Planning and Inference*, **139**, 16–22.

[4] Ceranka, B. and Graczyk, M. (2011). Robustness optimal spring balance weighing designs for estimation total weight, *Kybernetika*, **47**, 902–908.

[5] Ceranka, B. and Graczyk, M. (2015). Notes on the regular E-optimal spring balance weighing designs with correlated errors, *REVSTAT Statistical Journal*, **13**, 119–129.

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

[7] Cheng, C.S. (2014). Optimal biased weighing designs and two-level main-effect plans, *Journal of Statistical Theory and Practice*, **8**, 83–99.

[8] Ehlich, H. (1964). Determinantenabschätzungen für binäre Matrizen mit $n \equiv 3 \bmod 4$, *Mathematische Zeitschrift*, **84**, 438–447.

[9] Galil, Z. and Kiefer, J. (1980). D-optimum weighing designs, *Annals of Statistics*, **8**, 1293–1306.

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

[11] Graczyk, M. (2012). Regular A-optimal spring balance weighing designs, *REVSTAT Statistical Journal*, **10**, 323–333.

[12] Graczyk, M. (2013). Some applications of weighing designs, *Biometrical Letters*, **50**, 15–26.

[13] Jenkins, G.M. and Chanmugam, J. (1962). The estimation of slope when the errors are autocorrelated, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **24**, 199–214.

[14] Katulska, K. and Smaga, Ł. (2012). D-optimal chemical balance weighing designs with $n \equiv 0 \pmod 4$ and 3 objects, *Communications in Statistics – Theory and Methods*, **41**, 2445–2455.

[15] Katulska, K. and Smaga, Ł. (2013). D-optimal chemical balance weighing designs with autoregressive errors, *Metrika*, **76**, 393–407.

[16] Katulska, K. and Smaga, Ł. D-optimal and highly D-efficient designs with non-negatively correlated observations. (Under review).

[17] Li, C.H. and Yang, S.Y. (2005). On a conjecture in D-optimal designs with $n \equiv 0 \pmod 4$, *Linear Algebra and its Applications*, **400**, 279–290.

[18] Masaro, J. and Wong, C.S. (2008a). D-optimal designs for correlated random vectors, *Journal of Statistical Planning and Inference*, **138**, 4093–4106.

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

[20] Masaro, J. and Wong, C.S. (2008c). Robustness of optimal designs for correlated random variables, *Linear Algebra and its Applications*, **429**, 1639–1646.

[21] Neubauer, M.G. and Pace, R.G. (2010). D-optimal $(0, 1)$-weighing designs for eight objects, *Linear Algebra and its Applications*, **432**, 2634–2657.

[22] Smaga, Ł. (2014). Necessary and sufficient conditions in the problem of D-optimal weighing designs with autocorrelated errors, *Statistics & Probability Letters*, **92**, 12–16.

[23] Smaga, Ł. (2015). Uniquely E-optimal designs with $n \equiv 2 \pmod 4$ correlated observations, *Linear Algebra and its Applications*, **473**, 297–315.

[24] Yeh, H.G. and Lo Huang, M.N. (2005). On exact D-optimal designs with 2 two-level factors and n autocorrelated observations, *Metrika*, **61**, 261–275.

REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 455–471

OPTIMAL B-ROBUST ESTIMATORS FOR THE PARAMETERS OF THE GENERALIZED HALF-NORMAL DISTRIBUTION

Authors: Fatma Zehra Doğru
– Department of Econometrics, Giresun University,
28100 Giresun, Turkey
fatma.dogru@giresun.edu.tr

Y. Murat Bulut
– Department of Statistics, Eskişehir Osmangazi University,
26480 Eskişehir, Turkey
ymbulut@ogu.edu.tr

Olcay Arslan
– Department of Statistics, Ankara University,
06100 Ankara, Turkey
oarslan@ankara.edu.tr

Received: December 2014 Revised: November 2015 Accepted: November 2015

Abstract:

- The purpose of this study is to propose robust estimators by using optimal B-robust (OBR) estimation method (Hampel *et al.* [5]) for the parameters of the generalized half-normal (GHN) distribution. After given the robust estimators, we provide a small simulation study to compare its performance with the estimators obtained from maximum likelihood (ML) estimation method. We also give a real data example to illustrate the performance of the proposed estimators.

Key-Words:

- *generalized half-normal; optimal B-robust; maximum likelihood.*

AMS Subject Classification:

- 62F35, 65C60.

1. INTRODUCTION

The GHN distribution was introduced by Cooray and Ananda [3] as an alternative lifetime distribution. It is observed by Cooray and Ananda [3] that the cumulative distribution function (cdf) of the new family is very similar to the cdf of the half-normal distribution. Thus, they called the new family as the "generalized half-normal (GHN) distribution". It can be seen that the GHN distribution is a special case of the three-parameter generalized gamma distribution given by Stacy [7] (Cooray and Ananda [3]).

Some distributional properties of the GHN distribution are given by Cooray and Ananda [3]. In their study, the parameters of the GHN distribution are estimated using the ML estimation method, and using real data sets the performance of the ML estimator is compared with the other commonly used failure time distributions such as Weibull, gamma, lognormal and Birnbaum–Saunders.

One way of estimating the parameters of a given distribution is to use the ML estimation method. However, this estimator can be very sensitive to the outliers. Thus, the robust estimators may be needed as an alternative to the ML estimators in the presence of outliers. In this paper, we will use the OBR estimation method to obtain robust estimators for the parameters of the GHN distribution. The OBR estimation method was introduced by Hampel *et al.* [5] and used by Victoria-Feser [8] and Victoria-Feser and Ronchetti [9] to estimate the shape parameters of the Pareto and the gamma distributions. Also, Doğru and Arslan [4] used the OBR estimation method to estimate the shape parameters of the Burr XII distribution. Our goal is to show that the OBR estimation method can be used as an alternative to the ML estimation method to obtain robust estimators for the parameters of the GHN distribution when the data includes outliers.

The paper is organized as follows. In Section 2, we briefly summarize the properties of the GHN distribution. In Section 3, we explore the estimation of the GHN distribution. We first give the ML estimation method and then we give the OBR estimation method. We also give the algorithm to obtain the OBR estimates. In Sections 4 and 5, we give a simulation study and a real data example to demonstrate the performance of the proposed estimators over the ML estimators. Some conclusions are given in Section 6.

2. GENERALIZED HALF-NORMAL DISTRIBUTION (GHN)

The probability density function (pdf) and the cdf of the GHN distribution are given by

$$f(x;\alpha,\theta) = \begin{cases} \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}, & x>0,\ \alpha>0,\ \theta>0 \\ 0 \qquad\qquad\qquad\quad , & x\le 0 \end{cases} \tag{2.1}$$

and

$$F(x;\alpha,\theta) = 2\Phi\left(\left(\frac{x}{\theta}\right)^{\alpha}\right) - 1\,, \qquad x\ge 0,\ \ \alpha>0,\ \ \theta>0 \tag{2.2}$$

respectively, where $\Phi(\cdot)$ is the cdf of the standard normal distribution and α and θ are the shape and scale parameters of the GHN distribution.

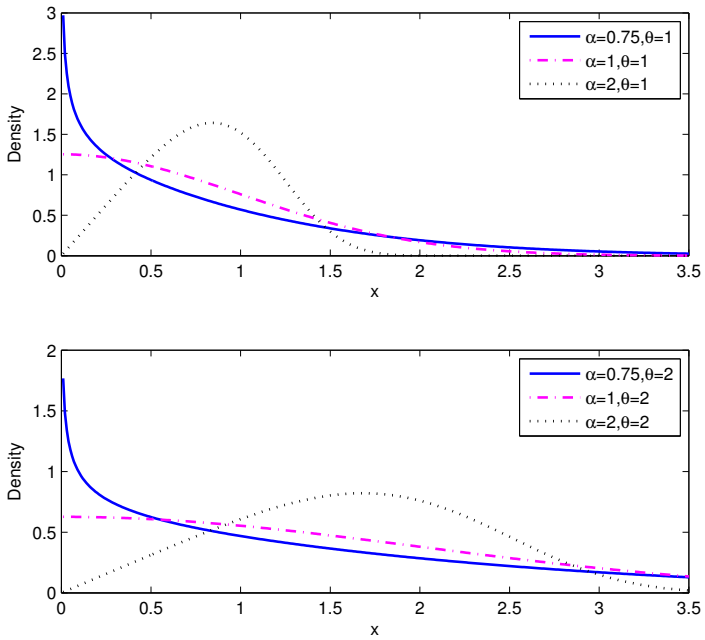

Figure 1: Examples of the GHN pdf for different values of α and θ.

The k-th moment, expected value and the variance are given by Cooray and Ananda [3] as follows:

$$E(X^k) = \sqrt{\frac{2^{\frac{k}{\alpha}}}{\pi}}\,\Gamma\left(\frac{k+\alpha}{2\alpha}\right)\theta^k\,,$$

$$E(X) = \sqrt{\frac{2^{\frac{1}{\alpha}}}{\pi}}\,\Gamma\left(\frac{1+\alpha}{2\alpha}\right)\theta$$

and

$$Var(X) = \frac{2^{\frac{1}{\alpha}}}{\pi}\left(\sqrt{\pi}\,\Gamma\left(\frac{2+\alpha}{2\alpha}\right) - \Gamma^2\left(\frac{1+\alpha}{2\alpha}\right)\right)\theta^2\,,$$

where $\Gamma(\cdot)$ is the gamma function. Figure 1 shows the plots of the pdf of the GHN distribution for some values of α and θ.

3. PARAMETER ESTIMATION

In this section, the parameters of the GHN distribution will be estimated using the ML and the OBR estimation methods.

3.1. ML estimation method

Let $X = (x_1, x_2, ..., x_n)$ be a random sample from GHN distribution. The log-likelihood function is

$$\log L(\alpha,\theta) = \frac{n}{2}\log\left(\frac{2}{\pi}\right) + n\log\alpha - n\alpha\log\theta + (\alpha-1)\sum_{i=1}^{n}\log(x_i) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^{2\alpha}. \tag{3.1}$$

Taking the derivatives of the log-likelihood function with respect to α and θ and setting to zero give the following equations

$$\frac{n}{\widehat{\alpha}} + \sum_{i=1}^{n}\log x_i - n\left(\sum_{i=1}^{n} x_i^{2\widehat{\alpha}}\log x_i\right)\left(\sum_{i=1}^{n} x_i^{2\widehat{\alpha}}\right)^{-1} = 0 \tag{3.2}$$

and

$$\widehat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{2\widehat{\alpha}}\right)^{\frac{1}{2\widehat{\alpha}}}. \tag{3.3}$$

Note that the same equations are also given by Cooray and Ananda [3]. There is not an analytical solution to the system formed by equations (3.2) and (3.3). This system can be only solved using numerical methods.

3.2. OBR estimation method

The OBR estimator introduced by Hampel *et al.* [5] belongs to the class of M-estimators (Huber [6]). Let $\boldsymbol{\eta} = (\alpha, \theta)$. The class of M-estimator for the

parameter $\boldsymbol{\eta}$ is defined as the minimum of the following objective function

$$\sum_{i=1}^{n} \rho(x_i, \boldsymbol{\eta}) \, .$$

If the ρ function is differentiable the M-estimator will be the solution of the following equation

$$\sum_{i=1}^{n} \psi(x_i, \boldsymbol{\eta}) = 0 \, ,$$

where $\psi = \rho'$ with $\psi \colon X \times \mathbb{R}^p \to \mathbb{R}^p$. There are many ρ functions in literature. In this paper, we will use the Huber's ρ function defined as

$$\rho_b(x) = \begin{cases} \frac{x^2}{2} \quad , & |x| \leq b \\ b\,|x| - \frac{1}{2}\,b^2 \, , & |x| > b \, . \end{cases}$$

Here, b is the robustness tuning constant and the derivative of ρ is $\psi_b(x)$ with

$$\psi_b(x) = \begin{cases} x \quad , & |x| \leq b \\ sgn(x)\, b \, , & |x| > b \, . \end{cases}$$

In general, the influence function (IF) for an M-estimator is defined as

$$IF = \frac{\psi(x, \boldsymbol{\eta})}{-\int \frac{\partial}{\partial \boldsymbol{\eta}} \psi(x, \boldsymbol{\eta}) \, dF_{\boldsymbol{\eta}}(x)} \tag{3.4}$$

and it is used to measure the local robustness of an estimator. It is desired that IF is bounded. The estimators with bounded IF are called the OBR estimators. The IF of an ML estimator is

$$IF = J(\boldsymbol{\eta})^{-1} \mathbf{s}(x, \boldsymbol{\eta}) \, ,$$

where $J(\boldsymbol{\eta})$ is the Fisher information matrix and $\mathbf{s}(x, \boldsymbol{\eta}) = \left(\frac{\partial}{\partial \boldsymbol{\eta}}\right) \log f_{\boldsymbol{\eta}}(x)$ is the score function. It can be seen that the IF of an ML estimator is proportional to the score functions, so the score function should be bounded for a bounded IF for the ML estimator.

Concerning the GHN distribution, we take logarithm of $f(x; \alpha, \theta)$ given in (2.1) to obtain the score functions

$$\log\big(f(x; \alpha, \theta)\big) = \log\left(\sqrt{\frac{2}{\pi}}\right) + \log(\alpha) - \log(x) - \frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha} + \alpha \log\left(\frac{x}{\theta}\right) .$$

Then, taking the derivatives of the $\log\big(f(x; \alpha, \theta)\big)$ with respect to α and θ we obtain the following equations

$$\frac{\partial\Big(\log\big(f(x; \alpha, \theta)\big)\Big)}{\partial \alpha} = \frac{1}{\alpha} + \log\left(\frac{x}{\theta}\right) - \left(\frac{x}{\theta}\right)^{2\alpha} \log\left(\frac{x}{\theta}\right) , \tag{3.5}$$

$$\frac{\partial\Big(\log\big(f(x;\alpha,\theta)\big)\Big)}{\partial\theta} = -\frac{\alpha}{\theta} + \frac{\alpha}{\theta^{2\alpha+1}}\, x^{2\alpha}\,. \tag{3.6}$$

After some straightforward simplifications, the score functions for the parameters α and θ are given

$$\mathbf{s}(x;\alpha,\theta) = \begin{bmatrix} \frac{1}{\alpha} + \log\big(\frac{x}{\theta}\big)\Big(1 - \big(\frac{x}{\theta}\big)^{2\alpha}\Big) \\ -\frac{\alpha}{\theta} + \frac{\alpha}{\theta^{2\alpha+1}}\, x^{2\alpha} \end{bmatrix}.$$

It is clear that the score functions are not bounded functions of x (see Figures (2) and (3)). Thus, the IF of the ML estimator for the GHN distribution will be unbounded. This implies that the ML estimators will be very sensitive to the outliers. Therefore, robust estimation methods will be needed to estimate the parameters of the GHN distribution.

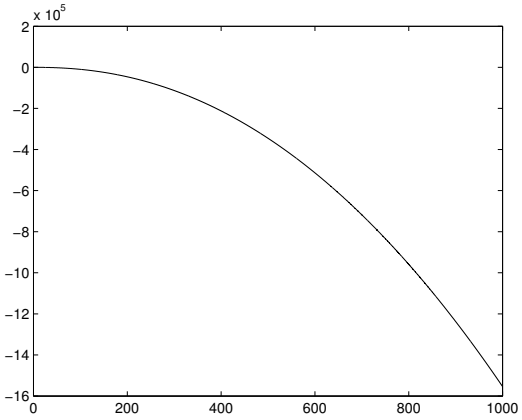

Figure 2: Score function for α parameter with $\alpha = 1$ and $\theta = 2$.

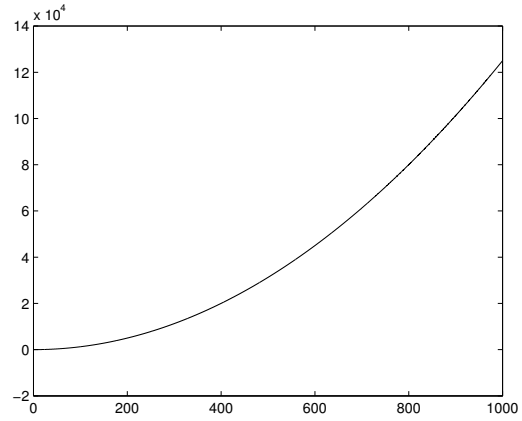

Figure 3: Score function for θ parameter with $\alpha = 1$ and $\theta = 2$.

There are several versions of the OBR estimators defined in Hampel *et al.* [5, p. 243], depending on the method of choosing to bound IF. In this study, we used the standardized OBR estimator which is defined as follows

$$\sum_{i=1}^{n}\psi\Big(\mathbf{A}(\boldsymbol{\eta})\left(\mathbf{s}(x_i,\boldsymbol{\eta})-\mathbf{a}(\boldsymbol{\eta})\right)\Big) = \sum_{i=1}^{n}W_b(x_i,\boldsymbol{\eta})\big\{\mathbf{s}(x_i,\boldsymbol{\eta})-\mathbf{a}(\boldsymbol{\eta})\big\} = \mathbf{0}\,, \tag{3.7}$$

where

$$W_b(x,\boldsymbol{\eta}) = \min\left\{1;\frac{b}{\big\|\mathbf{A}(\boldsymbol{\eta})\big\{\mathbf{s}(x_i,\boldsymbol{\eta})-\mathbf{a}(\boldsymbol{\eta})\big\}\big\|}\right\}, \tag{3.8}$$

is the weight function and $\|\cdot\|$ shows the Euclidian norm. Also the nonsingular $p\times p$ matrix $\mathbf{A}(\boldsymbol{\eta})$ and the $p\times 1$ vector $\mathbf{a}(\boldsymbol{\eta})$ are defined implicitly by

$$E\Big\{\psi(x,\boldsymbol{\eta})\,\psi(x,\boldsymbol{\eta})^T\Big\} = \Big\{\mathbf{A}(\boldsymbol{\eta})^T\mathbf{A}(\boldsymbol{\eta})\Big\}^{-1}, \tag{3.9}$$

$$E\Big\{\psi(x,\boldsymbol{\eta})\Big\} = \mathbf{0}\,. \tag{3.10}$$

The weight will be 1, if $\|\mathbf{A}(\boldsymbol{\eta})\{\mathbf{s}(x_i,\boldsymbol{\eta})-\mathbf{a}(\boldsymbol{\eta})\}\| \le b$, otherwise it will be $\frac{b}{\|\mathbf{A}(\boldsymbol{\eta})\{\mathbf{s}(x_i,\boldsymbol{\eta})-\mathbf{a}(\boldsymbol{\eta})\}\|}$, which bounds the score function for the outlying observations. Thus, the corresponding OBR estimator will be less sensitive to the outliers in the data.

To obtain the OBR estimates the following algorithm can be applied. Note that this algorithm was proposed by Victoria-Feser and Ronchetti [9].

Algorithm:

Step 1. Let ϵ be a stopping rule. Take initial values for the parameter $\boldsymbol{\eta}$. Set $\mathbf{a}=\mathbf{0}$ and $\mathbf{A}=\mathbf{J}^{\frac{1}{2}}(\boldsymbol{\eta})^{-T}$, where

$$\mathbf{J}(\boldsymbol{\eta}) = \int \mathbf{s}(x,\boldsymbol{\eta})\,\mathbf{s}(x,\boldsymbol{\eta})^T\,dF_{\boldsymbol{\eta}}(x)\,.$$

Step 2. Solve the following equations for $\mathbf{a}$ and $\mathbf{A}$

$$\mathbf{A}\,\mathbf{A}^T = \mathbf{M}_2^{-1}$$

and

$$\mathbf{a} = \frac{\int W_b(x,\boldsymbol{\eta})\,\mathbf{s}(x,\boldsymbol{\eta})\,dF_{\boldsymbol{\eta}}(x)}{\int W_b(x,\boldsymbol{\eta})\,dF_{\boldsymbol{\eta}}(x)}\,,$$

where

$$\mathbf{M}_k = \int W_b(x,\boldsymbol{\eta})^k\left\{\mathbf{s}(x,\boldsymbol{\eta})-\mathbf{a}\right\}\left\{\mathbf{s}(x,\boldsymbol{\eta})-\mathbf{a}\right\}^T dF_{\boldsymbol{\eta}}(x), \quad k=1,2\,.$$

The current values of $\boldsymbol{\eta}$, $\mathbf{a}$ and $\mathbf{A}$ are used as starting values to solve the given equations.

Step 3. Compute $\mathbf{M_1}$ and $\Delta\boldsymbol{\eta} = \boldsymbol{M_1}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} W_b(x,\boldsymbol{\eta})\left\{\mathbf{s}(x,\boldsymbol{\eta}) - \mathbf{a}\right\}\right)$.

Step 4. If $|\Delta\boldsymbol{\eta}| > \epsilon$ then $\boldsymbol{\eta} \longrightarrow \boldsymbol{\eta} + \Delta\boldsymbol{\eta}$ and return to step 2, else stop.

Note that for the finite sample case the integrals in the equations will be replaced by the summations.

The ML estimator can be taken as initial value for the parameter $\boldsymbol{\eta}$. In our simulation study, we have used several different initial points including the ML and true parameter values. We have also used robust starting values suggested by Victoria-Feser and Ronchetti [9]. As pointed out by Victoria-Feser and Ronchetti [9] the algorithm is convergent depending on the initial values. Other estimators such as moment estimators can also be used as starting values. However, for this distribution it is not possible to obtain explicit form of the moment estimators. Therefore, it is not tractable to use them as initial values for the algorithm.

4. SIMULATION STUDY

In this section, we will give a simulation study to compare the performance of the OBR estimators with the ML estimators with and without outliers in the data. The data are randomly generated from GHN distribution for different values of α and θ parameters. The data generation is conducted as follows. Let $U \sim \mathit{Uniform}(0,1)$. Then, $X \sim \theta\big(\Phi^{-1}\big(\frac{U+1}{2}\big)\big)^{\frac{1}{\alpha}}$ will have GHN distribution with the parameters α and θ. To evaluate the performance of the estimators bias and root mean square error (RMSE) are computed over 1000 replications for the sample sizes $n = 25, 50, 100$ and the parameter values $(\alpha,\theta) = (0.75,1), (1,1), (2,1), (0.75,2), (1,2), (2,2)$. Here, the bias and RMSE are defined as

$$\mathit{bias}(\widehat{\alpha}) = \overline{\alpha} - \alpha\,, \qquad \mathit{bias}(\widehat{\theta}) = \overline{\theta} - \theta\,,$$

$$\mathit{RMSE}(\widehat{\alpha}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\alpha}_i - \alpha)^2}\,, \qquad \mathit{RMSE}(\widehat{\theta}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{\theta}_i - \theta)^2}\,,$$

where $\overline{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\alpha}_i$, $\overline{\theta} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_i$ and $N = 1000$. For all simulation cases, the stopping rule ϵ is taken as 10^{-6}. The simulation study and real data example are conducted using MATLAB R2013a.

In the OBR estimation method, the robustness tuning constant should be chosen in order to gain the desire efficiency. The most robust estimator can be obtained by choosing b as the squared of the number of parameters. In our case, we can take $b = \sqrt{2}$. For this value of b one can approximately gain 60% efficiency

for the resulting estimators. When we increase the value of b, the efficiency will also increase. Therefore, we have taken $b = 2$ to have efficiency more than 60%. For more details about the selection of the robustness tuning constant, see Hampel *et al.* [5, p.247] and Victoria-Feser and Ronchetti [9].

Concerning the starting value for the algorithm given in Section 3.2 we use the robust starting values suggested by Victoria-Feser and Ronchetti [9] which is described as follows.

i) Find the ML estimates.

ii) Take $b = 3.5$ to get OBR estimates.

iii) Use the OBR estimates obtained at step ii) as new initial values and set $b = 2$ to obtain OBR estimates again.

In this simulation study, we consider two types of outlier model to the right and the left in the X direction. These models are

$$\textit{Case I}. \quad (n-r)\, GHN(x;\alpha,\theta) \,+\, r\, \textit{Uniform}\big(max(x)+5\sigma,\, max(x)+10\sigma\big)\,,$$

$$\textit{Case II}. \quad (n-r)\, GHN(x;\alpha,\theta) \,+\, r\, \textit{Uniform}(0, 0.0001)\,,$$

where $max\,(x)$ is the largest observations in the sample, σ is standard deviation of a randomly generated sample from GHN distribution and r is chosen by multiplying the sample sizes by 0.1. That is, we add two outliers for $n = 25$, five outliers for $n = 50$ and ten outliers for $n = 100$. Further, in Case I, we add outliers in the upper tail of the distribution. In the second case, the outliers are added in the lower tail of the distribution to see the performance of the estimators for this type of outliers. As suggested by a referee, this type of outliers may represent severely shorted life-lengths.

Simulation results are given in Tables 1–3. In the tables, the estimates of the parameters, bias and RMSE are presented. Table 1 shows the results for the case without outliers in the data. From this table, we can observe that the performance of the ML seems slightly better than the performance of the OBR estimators. In Table 2, we give the simulation results for the outlier Case I. The results of this table show that the OBR estimators have smaller bias and RMSE values than the corresponding values for ML estimators in all the simulation configurations. Finally, in Table 3 the simulation results for the outlier Case II are displayed. Similar to the outlier Case I the OBR estimators outperform the ML estimators in terms of bias and RMSE in all the simulation scenarios. Overall when the data has outliers the OBR estimators should be used instead of ML estimators to obtain robust estimators for the parameters of interest.

Table 1: Estimates of parameters, bias and RMSE for different sample sizes without outlier.

n	θ	α	Parameter(α)			Parameter(θ)		
				ML	OBR		ML	OBR
25	1	0.75	$\overline{\alpha}$	0.7995	0.8808	$\overline{\theta}$	0.9915	0.9109
			Bias($\widehat{\alpha}$)	0.0495	0.1308	Bias($\widehat{\theta}$)	−0.0085	−0.0891
			RMSE($\widehat{\alpha}$)	0.1503	0.2061	RMSE($\widehat{\theta}$)	0.1978	0.2129
		1	$\overline{\alpha}$	1.0589	1.1677	$\overline{\theta}$	0.9931	0.9310
			Bias($\widehat{\alpha}$)	0.0589	0.1677	Bias($\widehat{\theta}$)	−0.0069	−0.0690
			RMSE($\widehat{\alpha}$)	0.2000	0.2741	RMSE($\widehat{\theta}$)	0.1548	0.1688
		2	$\overline{\alpha}$	2.1412	2.3654	$\overline{\theta}$	1.0005	0.9668
			Bias($\widehat{\alpha}$)	0.1412	0.3654	Bias($\widehat{\theta}$)	0.0005	−0.0332
			RMSE($\widehat{\alpha}$)	0.4112	0.5660	RMSE($\widehat{\theta}$)	0.0767	0.0854
	2	0.75	$\overline{\alpha}$	0.7942	0.8752	$\overline{\theta}$	2.0106	1.8402
			Bias($\widehat{\alpha}$)	0.0442	0.1252	Bias($\widehat{\theta}$)	0.0106	−0.1598
			RMSE($\widehat{\alpha}$)	0.1503	0.2031	RMSE($\widehat{\theta}$)	0.4057	0.4254
		1	$\overline{\alpha}$	1.0671	1.1782	$\overline{\theta}$	2.0029	1.8738
			Bias($\widehat{\alpha}$)	0.0671	0.1782	Bias($\widehat{\theta}$)	0.0029	−0.1262
			RMSE($\widehat{\alpha}$)	0.2035	0.2792	RMSE($\widehat{\theta}$)	0.3198	0.3416
		2	$\overline{\alpha}$	2.1282	2.3447	$\overline{\theta}$	1.9869	1.9207
			Bias($\widehat{\alpha}$)	0.1282	0.3447	Bias($\widehat{\theta}$)	−0.0131	−0.0793
			RMSE($\widehat{\alpha}$)	0.4032	0.5486	RMSE($\widehat{\theta}$)	0.1509	0.1738
50	1	0.75	$\overline{\alpha}$	0.7753	0.8578	$\overline{\theta}$	1.0051	0.9142
			Bias($\widehat{\alpha}$)	0.0253	0.1078	Bias($\widehat{\theta}$)	0.0051	−0.0858
			RMSE($\widehat{\alpha}$)	0.0990	0.1523	RMSE($\widehat{\theta}$)	0.1471	0.1653
		1	$\overline{\alpha}$	1.0245	1.1348	$\overline{\theta}$	0.9952	0.9265
			Bias($\widehat{\alpha}$)	0.0245	0.1348	Bias($\widehat{\theta}$)	−0.0048	−0.0735
			RMSE($\widehat{\alpha}$)	0.1267	0.1959	RMSE($\widehat{\theta}$)	0.1100	0.1316
		2	$\overline{\alpha}$	2.0495	2.2682	$\overline{\theta}$	0.9952	0.9598
			Bias($\widehat{\alpha}$)	0.0495	0.2682	Bias($\widehat{\theta}$)	−0.0048	−0.0402
			RMSE($\widehat{\alpha}$)	0.2551	0.3908	RMSE($\widehat{\theta}$)	0.0553	0.0696
	2	0.75	$\overline{\alpha}$	0.7722	0.8540	$\overline{\theta}$	2.0199	1.8468
			Bias($\widehat{\alpha}$)	0.0222	0.1040	Bias($\widehat{\theta}$)	0.0199	−0.1532
			RMSE($\widehat{\alpha}$)	0.0912	0.1442	RMSE($\widehat{\theta}$)	0.2936	0.3245
		1	$\overline{\alpha}$	1.0281	1.1389	$\overline{\theta}$	1.9890	1.8519
			Bias($\widehat{\alpha}$)	0.0281	0.1389	Bias($\widehat{\theta}$)	−0.0110	−0.1481
			RMSE($\widehat{\alpha}$)	0.1250	0.1966	RMSE($\widehat{\theta}$)	0.2157	0.2613
		2	$\overline{\alpha}$	2.0556	2.2803	$\overline{\theta}$	1.9988	1.9276
			Bias($\widehat{\alpha}$)	0.0556	0.2803	Bias($\widehat{\theta}$)	−0.0012	−0.0724
			RMSE($\widehat{\alpha}$)	0.2587	0.4011	RMSE($\widehat{\theta}$)	0.1100	0.1350
100	1	0.75	$\overline{\alpha}$	0.7631	0.8464	$\overline{\theta}$	1.0023	0.9102
			Bias($\widehat{\alpha}$)	0.0131	0.0964	Bias($\widehat{\theta}$)	0.0023	−0.0898
			RMSE($\widehat{\alpha}$)	0.0647	0.1209	RMSE($\widehat{\theta}$)	0.0993	0.1328
		1	$\overline{\alpha}$	1.0147	1.1271	$\overline{\theta}$	1.0004	0.9284
			Bias($\widehat{\alpha}$)	0.0147	0.1271	Bias($\widehat{\theta}$)	0.0004	−0.0716
			RMSE($\widehat{\alpha}$)	0.0859	0.1593	RMSE($\widehat{\theta}$)	0.0755	0.1038
		2	$\overline{\alpha}$	2.0420	2.2686	$\overline{\theta}$	0.9998	0.9627
			Bias($\widehat{\alpha}$)	0.0420	0.2686	Bias($\widehat{\theta}$)	−0.0002	−0.0373
			RMSE($\widehat{\alpha}$)	0.1827	0.3371	RMSE($\widehat{\theta}$)	0.0379	0.0545
	2	0.75	$\overline{\alpha}$	0.7617	0.8459	$\overline{\theta}$	1.9975	1.8107
			Bias($\widehat{\alpha}$)	0.0117	0.0959	Bias($\widehat{\theta}$)	−0.0025	−0.1893
			RMSE($\widehat{\alpha}$)	0.0658	0.1212	RMSE($\widehat{\theta}$)	0.2069	0.2772
		1	$\overline{\alpha}$	1.0179	1.1312	$\overline{\theta}$	1.9992	1.8563
			Bias($\widehat{\alpha}$)	0.0179	0.1312	Bias($\widehat{\theta}$)	−0.0008	−0.1437
			RMSE($\widehat{\alpha}$)	0.0837	0.1614	RMSE($\widehat{\theta}$)	0.1591	0.2141
		2	$\overline{\alpha}$	2.0287	2.2560	$\overline{\theta}$	1.9944	1.9192
			Bias($\widehat{\alpha}$)	0.0287	0.2560	Bias($\widehat{\theta}$)	−0.0056	−0.0808
			RMSE($\widehat{\alpha}$)	0.1720	0.3221	RMSE($\widehat{\theta}$)	0.0730	0.1107

Table 2: Estimates of parameters, bias and RMSE for different sample sizes for Case I.

n	θ	α	Parameter(α)			Parameter(θ)		
				ML	OBR		ML	OBR
25	1	0.75	$\overline{\alpha}$	0.5243	0.6119	$\overline{\theta}$	1.3713	1.0104
			Bias($\widehat{\alpha}$)	−0.2257	−0.1381	Bias($\widehat{\theta}$)	0.3713	0.0104
			RMSE($\widehat{\alpha}$)	0.2342	0.1588	RMSE($\widehat{\theta}$)	0.4695	0.2186
		1	$\overline{\alpha}$	0.6296	0.7430	$\overline{\theta}$	1.3592	1.0278
			Bias($\widehat{\alpha}$)	−0.3704	−0.2570	Bias($\widehat{\theta}$)	0.3592	0.0278
			RMSE($\widehat{\alpha}$)	0.3771	0.2723	RMSE($\widehat{\theta}$)	0.4174	0.1751
		2	$\overline{\alpha}$	0.9279	1.1280	$\overline{\theta}$	1.3321	1.0777
			Bias($\widehat{\alpha}$)	−1.0721	−0.8720	Bias($\widehat{\theta}$)	0.3321	0.0777
			RMSE($\widehat{\alpha}$)	1.0766	0.8813	RMSE($\widehat{\theta}$)	0.3486	0.1264
	2	0.75	$\overline{\alpha}$	0.5273	0.6148	$\overline{\theta}$	2.7406	2.0185
			Bias($\widehat{\alpha}$)	−0.2227	−0.1352	Bias($\widehat{\theta}$)	0.7406	0.0185
			RMSE($\widehat{\alpha}$)	0.2307	0.1550	RMSE($\widehat{\theta}$)	0.9204	0.4217
		1	$\overline{\alpha}$	0.6331	0.7475	$\overline{\theta}$	2.7256	2.0636
			Bias($\widehat{\alpha}$)	−0.3669	−0.2525	Bias($\widehat{\theta}$)	0.7256	0.0636
			RMSE($\widehat{\alpha}$)	0.3736	0.2676	RMSE($\widehat{\theta}$)	0.8336	0.3359
		2	$\overline{\alpha}$	0.9323	1.1327	$\overline{\theta}$	2.6779	2.1652
			Bias($\widehat{\alpha}$)	−1.0677	−0.8673	Bias($\widehat{\theta}$)	0.6779	0.1652
			RMSE($\widehat{\alpha}$)	1.0726	0.8773	RMSE($\widehat{\theta}$)	0.7115	0.2578
50	1	0.75	$\overline{\alpha}$	0.4948	0.5619	$\overline{\theta}$	1.4738	1.0876
			Bias($\widehat{\alpha}$)	−0.2552	−0.1881	Bias($\widehat{\theta}$)	0.4738	0.0876
			RMSE($\widehat{\alpha}$)	0.2584	0.1947	RMSE($\widehat{\theta}$)	0.5188	0.1880
		1	$\overline{\alpha}$	0.5922	0.6792	$\overline{\theta}$	1.4617	1.1146
			Bias($\widehat{\alpha}$)	−0.4078	−0.3208	Bias($\widehat{\theta}$)	0.4617	0.1146
			RMSE($\widehat{\alpha}$)	0.4103	0.3258	RMSE($\widehat{\theta}$)	0.4886	0.1719
		2	$\overline{\alpha}$	0.8816	1.0447	$\overline{\theta}$	1.4194	1.1690
			Bias($\widehat{\alpha}$)	−1.1184	−0.9553	Bias($\widehat{\theta}$)	0.4194	0.1690
			RMSE($\widehat{\alpha}$)	1.1201	0.9590	RMSE($\widehat{\theta}$)	0.4269	0.1844
	2	0.75	$\overline{\alpha}$	0.4952	0.5628	$\overline{\theta}$	2.9648	2.1899
			Bias($\widehat{\alpha}$)	−0.2548	−0.1872	Bias($\widehat{\theta}$)	0.9648	0.1899
			RMSE($\widehat{\alpha}$)	0.2582	0.1941	RMSE($\widehat{\theta}$)	1.0550	0.3858
		1	$\overline{\alpha}$	0.5941	0.6820	$\overline{\theta}$	2.9364	2.2459
			Bias($\widehat{\alpha}$)	−0.4059	−0.3180	Bias($\widehat{\theta}$)	0.9364	0.2459
			RMSE($\widehat{\alpha}$)	0.4086	0.3231	RMSE($\widehat{\theta}$)	0.9911	0.3639
		2	$\overline{\alpha}$	0.8813	1.0444	$\overline{\theta}$	2.8429	2.3387
			Bias($\widehat{\alpha}$)	−1.1187	−0.9556	Bias($\widehat{\theta}$)	0.8429	0.3387
			RMSE($\widehat{\alpha}$)	1.1206	0.9596	RMSE($\widehat{\theta}$)	0.8586	0.3711
100	1	0.75	$\overline{\alpha}$	0.4859	0.5522	$\overline{\theta}$	1.4969	1.0990
			Bias($\widehat{\alpha}$)	−0.2641	−0.1978	Bias($\widehat{\theta}$)	0.4969	0.0990
			RMSE($\widehat{\alpha}$)	0.2656	0.2009	RMSE($\widehat{\theta}$)	0.5205	0.1547
		1	$\overline{\alpha}$	0.5842	0.6704	$\overline{\theta}$	1.4783	1.1255
			Bias($\widehat{\alpha}$)	−0.4158	−0.3296	Bias($\widehat{\theta}$)	0.4783	0.1255
			RMSE($\widehat{\alpha}$)	0.4170	0.3321	RMSE($\widehat{\theta}$)	0.4925	0.1573
		2	$\overline{\alpha}$	0.8676	1.0268	$\overline{\theta}$	1.4262	1.1696
			Bias($\widehat{\alpha}$)	−1.1324	−0.9732	Bias($\widehat{\theta}$)	0.4262	0.1696
			RMSE($\widehat{\alpha}$)	1.1333	0.9752	RMSE($\widehat{\theta}$)	0.4301	0.1783
	2	0.75	$\overline{\alpha}$	0.4863	0.5525	$\overline{\theta}$	3.0158	2.2161
			Bias($\widehat{\alpha}$)	−0.2637	−0.1975	Bias($\widehat{\theta}$)	1.0158	0.2161
			RMSE($\widehat{\alpha}$)	0.2653	0.2006	RMSE($\widehat{\theta}$)	1.0648	0.3280
		1	$\overline{\alpha}$	0.5816	0.6675	$\overline{\theta}$	2.9503	2.2433
			Bias($\widehat{\alpha}$)	−0.4184	−0.3325	Bias($\widehat{\theta}$)	0.9503	0.2433
			RMSE($\widehat{\alpha}$)	0.4196	0.3349	RMSE($\widehat{\theta}$)	0.9773	0.3060
		2	$\overline{\alpha}$	0.8687	1.0303	$\overline{\theta}$	2.8523	2.3440
			Bias($\widehat{\alpha}$)	−1.1313	−0.9697	Bias($\widehat{\theta}$)	0.8523	0.3440
			RMSE($\widehat{\alpha}$)	1.1321	0.9715	RMSE($\widehat{\theta}$)	0.8595	0.3591

Table 3: Estimates of parameters, bias and RMSE for different sample sizes for Case II.

n	θ	α	Parameter(α)			Parameter(θ)		
				ML	OBR		ML	OBR
25	1	0.75	$\overline{\alpha}$	0.5872	0.6849	$\overline{\theta}$	0.6495	0.6564
			Bias($\widehat{\alpha}$)	−0.1628	−0.0651	Bias($\widehat{\theta}$)	−0.3505	−0.3436
			RMSE($\widehat{\alpha}$)	0.1780	0.1194	RMSE($\widehat{\theta}$)	0.3800	0.3772
		1	$\overline{\alpha}$	0.6906	0.8318	$\overline{\theta}$	0.7023	0.7167
			Bias($\widehat{\alpha}$)	−0.3094	−0.1682	Bias($\widehat{\theta}$)	−0.2977	−0.2833
			RMSE($\widehat{\alpha}$)	0.3191	0.2047	RMSE($\widehat{\theta}$)	0.3205	0.3104
		2	$\overline{\alpha}$	0.9338	1.2509	$\overline{\theta}$	0.7824	0.8082
			Bias($\widehat{\alpha}$)	−1.0662	−0.7491	Bias($\widehat{\theta}$)	−0.2176	−0.1918
			RMSE($\widehat{\alpha}$)	1.0690	0.7619	RMSE($\widehat{\theta}$)	0.2279	0.2050
	2	0.75	$\overline{\alpha}$	0.5722	0.6708	$\overline{\theta}$	1.3063	1.3240
			Bias($\widehat{\alpha}$)	−0.1778	−0.0792	Bias($\widehat{\theta}$)	−0.6937	−0.6760
			RMSE($\widehat{\alpha}$)	0.1890	0.1196	RMSE($\widehat{\theta}$)	0.7534	0.7460
		1	$\overline{\alpha}$	0.6686	0.8115	$\overline{\theta}$	1.3783	1.4127
			Bias($\widehat{\alpha}$)	−0.3314	−0.1885	Bias($\widehat{\theta}$)	−0.6217	−0.5873
			RMSE($\widehat{\alpha}$)	0.3390	0.2182	RMSE($\widehat{\theta}$)	0.6625	0.6358
		2	$\overline{\alpha}$	0.8923	1.2055	$\overline{\theta}$	1.5442	1.5966
			Bias($\widehat{\alpha}$)	−1.1077	−0.7945	Bias($\widehat{\theta}$)	−0.4558	−0.4034
			RMSE($\widehat{\alpha}$)	1.1098	0.8049	RMSE($\widehat{\theta}$)	0.4779	0.4320
50	1	0.75	$\overline{\alpha}$	0.5378	0.6285	$\overline{\theta}$	0.5867	0.6050
			Bias($\widehat{\alpha}$)	−0.2122	−0.1215	Bias($\widehat{\theta}$)	−0.4133	−0.3950
			RMSE($\widehat{\alpha}$)	0.2165	0.1364	RMSE($\widehat{\theta}$)	0.4245	0.4088
		1	$\overline{\alpha}$	0.6226	0.7527	$\overline{\theta}$	0.6420	0.6695
			Bias($\widehat{\alpha}$)	−0.3774	−0.2473	Bias($\widehat{\theta}$)	−0.3580	−0.3305
			RMSE($\widehat{\alpha}$)	0.3800	0.2562	RMSE($\widehat{\theta}$)	0.3669	0.3419
		2	$\overline{\alpha}$	0.8051	1.0570	$\overline{\theta}$	0.7423	0.7843
			Bias($\widehat{\alpha}$)	−1.1949	−0.9430	Bias($\widehat{\theta}$)	−0.2577	−0.2157
			RMSE($\widehat{\alpha}$)	1.1956	0.9453	RMSE($\widehat{\theta}$)	0.2621	0.2221
	2	0.75	$\overline{\alpha}$	0.5235	0.6171	$\overline{\theta}$	1.1691	1.2111
			Bias($\widehat{\alpha}$)	−0.2265	−0.1329	Bias($\widehat{\theta}$)	−0.8309	−0.7889
			RMSE($\widehat{\alpha}$)	0.2301	0.1457	RMSE($\widehat{\theta}$)	0.8524	0.8152
		1	$\overline{\alpha}$	0.6013	0.7334	$\overline{\theta}$	1.2692	1.3312
			Bias($\widehat{\alpha}$)	−0.3987	−0.2666	Bias($\widehat{\theta}$)	−0.7308	−0.6688
			RMSE($\widehat{\alpha}$)	0.4008	0.2741	RMSE($\widehat{\theta}$)	0.7462	0.6889
		2	$\overline{\alpha}$	0.7680	1.0162	$\overline{\theta}$	1.4733	1.5639
			Bias($\widehat{\alpha}$)	−1.2320	−0.9838	Bias($\widehat{\theta}$)	−0.5267	−0.4361
			RMSE($\widehat{\alpha}$)	1.2326	0.9856	RMSE($\widehat{\theta}$)	0.5351	0.4489
100	1	0.75	$\overline{\alpha}$	0.5347	0.6231	$\overline{\theta}$	0.5881	0.6071
			Bias($\widehat{\alpha}$)	−0.2153	−0.1269	Bias($\widehat{\theta}$)	−0.4119	−0.3929
			RMSE($\widehat{\alpha}$)	0.2176	0.1346	RMSE($\widehat{\theta}$)	0.4173	0.3995
		1	$\overline{\alpha}$	0.6210	0.7494	$\overline{\theta}$	0.6386	0.6686
			Bias($\widehat{\alpha}$)	−0.3790	−0.2506	Bias($\widehat{\theta}$)	−0.3614	−0.3314
			RMSE($\widehat{\alpha}$)	0.3802	0.2547	RMSE($\widehat{\theta}$)	0.3659	0.3371
		2	$\overline{\alpha}$	0.8044	1.0536	$\overline{\theta}$	0.7414	0.7869
			Bias($\widehat{\alpha}$)	−1.1956	−0.9464	Bias($\widehat{\theta}$)	−0.2586	−0.2131
			RMSE($\widehat{\alpha}$)	1.1959	0.9475	RMSE($\widehat{\theta}$)	0.2608	0.2164
	2	0.75	$\overline{\alpha}$	0.5217	0.6134	$\overline{\theta}$	1.1642	1.2106
			Bias($\widehat{\alpha}$)	−0.2283	−0.1366	Bias($\widehat{\theta}$)	−0.8358	−0.7894
			RMSE($\widehat{\alpha}$)	0.2301	0.1428	RMSE($\widehat{\theta}$)	0.8465	0.8029
		1	$\overline{\alpha}$	0.5996	0.7300	$\overline{\theta}$	1.2627	1.3292
			Bias($\widehat{\alpha}$)	−0.4004	−0.2700	Bias($\widehat{\theta}$)	−0.7373	−0.6708
			RMSE($\widehat{\alpha}$)	0.4014	0.2736	RMSE($\widehat{\theta}$)	0.7451	0.6810
		2	$\overline{\alpha}$	0.7640	1.0079	$\overline{\theta}$	1.4676	1.5634
			Bias($\widehat{\alpha}$)	−1.2360	−0.9921	Bias($\widehat{\theta}$)	−0.5324	−0.4366
			RMSE($\widehat{\alpha}$)	1.2363	0.9931	RMSE($\widehat{\theta}$)	0.5365	0.4426

5. REAL DATA EXAMPLE

In this section, we will analyze the data set used by Cooray and Ananda [3]. This data set contains the stress-rupture life of kevlar 49/ epoxy strands failure at 90% stress levels. The data set is given below (Andrews and Herzberg [1], Barlow *et al.* [2]).

Table 4: The failure times in hours.

0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.07	0.08
0.09	0.09	0.10	0.10	0.11	0.11	0.12	0.13	0.18	0.19	0.20	0.23	0.24
0.24	0.29	0.34	0.35	0.36	0.38	0.40	0.42	0.43	0.52	0.54	0.56	0.60
0.60	0.63	0.65	0.67	0.68	0.72	0.72	0.72	0.73	0.79	0.79	0.80	0.80
0.83	0.85	0.90	0.92	0.95	0.99	1.00	1.01	1.02	1.03	1.05	1.10	1.10
1.11	1.15	1.18	1.20	1.29	1.31	1.33	1.34	1.40	1.43	1.45	1.50	1.51
1.52	1.53	1.54	1.54	1.55	1.58	1.60	1.63	1.64	1.80	1.80	1.81	2.02
2.05	2.14	2.17	2.33	3.03	3.03	3.34	4.20	4.69	7.89			

Assume that this data set has a GHN distribution with the unknown parameters α and θ. We use the OBR estimation method to obtain the estimates for α and θ for the failure time data set. We also find the ML estimates for these parameters. Table 5 gives the summary of the estimates, standard error (SE) and the 95% confidence interval for the parameters of GHN distribution. The confidence intervals of the estimates are computed using the intervals given in Cooray and Ananda [3]. In their paper, they use the expected Fisher information matrix. For the ML estimators, we also use the expected Fisher information matrix to compute the standard errors and the confidence intervals. For the OBR estimators, we use the asymptotic covariance matrix given in Victoria-Feser and Ronchetti [9] to compute the standard errors and the confidence intervals.

Table 5: ML and OBR ($b = 2$) parameter estimates for the failure time data set.

Method	$\widehat{\alpha}$	SE	95% confidence interval of α	$\widehat{\theta}$	SE	95% confidence interval of θ
ML	0.7108	0.0584	(0.5964, 0.8252)	1.2238	0.1317	(0.9657, 1.4819)
OBR	0.7811	0.0574	(0.6685, 0.8937)	1.0540	0.0794	(0.8983, 1.2097)

Figure 4 (a) shows the boxplot of the failure time data set. After some preliminary examination of the data set, we can see from the boxplot that there

may be four potential outliers in the data set. We give the histogram of the data set with the fitted densities obtained from ML and OBR estimates in Figure 4 (b).

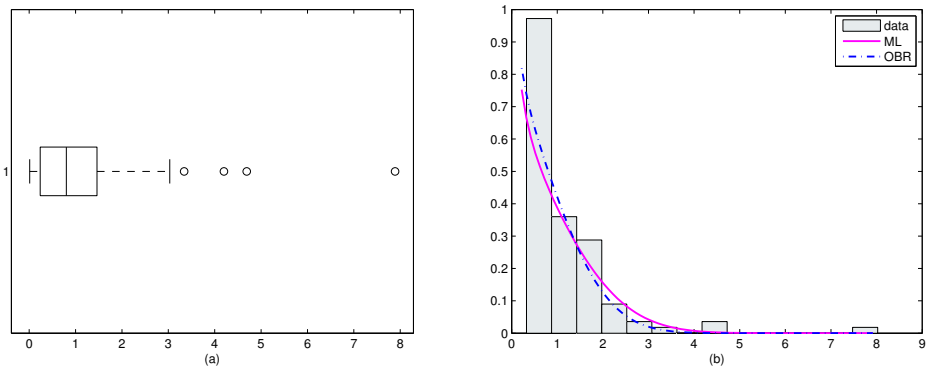

Figure 4: **(a)** Boxplot of the failure time data set; **(b)** Histogram with the fitted densities obtained from ML and OBR estimation methods.

We also give the Q-Q plots of the fitted distribution obtained from ML and OBR estimation methods in Figure 5. From this figure, we can see that the OBR estimates are not badly affected by the outliers. But we can clearly see that the ML estimators are influenced by the outliers. Furthermore, the Q-Q plot of the fitted distribution obtained from OBR estimators is well fitted contrary to the ML estimators.

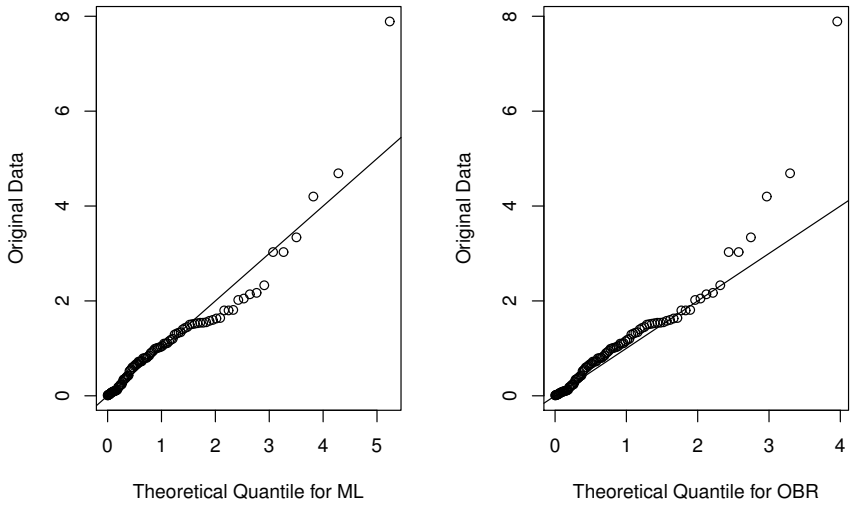

Figure 5: Q-Q plots for the failure time data set estimated by the ML and OBR estimation methods.

6. CONCLUSIONS

In this paper, we have proposed robust estimators for the parameters of the GHN distribution, which is proposed by Cooray and Ananda [3] as a flexible alternative lifetime distribution, using the OBR estimation method. Our limited simulation study has shown that the ML estimators are influenced by the outliers, but on the other hand, the OBR estimators are resistant to the outliers. The same results have been recorded from the real data example. Therefore, we can conclude that for this distribution the OBR estimators can be used as alternative estimators to the ML estimators.

ACKNOWLEDGMENTS

The authors would like to thank the referees and the associate editor for their constructive comments that led to substantial improvement of the paper.

REFERENCES

[1] Andrews, D.F. and Herzberg, A.M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, New York: Springer Series in Statistics.

[2] Barlow, R.E.; Toland, R.H. and Freeman, T. (1984). *A Bayesian Analysis of Stress-rupture Life of Kevlar 49/epoxy Spherical Pressure Vessels.* In "Proceedings of the Canadian Conference in Applied Statistics, 1981" (T.D. Dwivedi, Ed.), Marcel Dekker, New York.

[3] Cooray, K. and Ananda, M.M.A. (2008). A generalization of the half-normal distribution with applications to lifetime data, *Communications in Statistics – Theory and Methods*, **37**, 1323–1337.

[4] Doğru, F.Z. and Arslan, O. (2016). Optimal B-robust estimators for the parameters of the Burr XII distribution, *Journal of Statistical Computation and Simulation*, **86**(6), 1133–1149.

[5] Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[6] Huber, P.J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, **35**, 73–101.

[7] Stacy, E.W. (1962). A generalization of the Gamma distribution, *The Annals of Mathematical Statistics*, **33**, 1187–1192.

[8] Victoria-Feser, M.P. (1993). *Robust methods for personal income distribution models*, Ph.D thesis in Econometrics and Statistics, Faculty of Economic and Social Sciences, University of Geneva, Switzerland. (GE-archives).

[9] Victoria-Feser, M.P. and Ronchetti, E. (1994). Robust methods for personal-income distribution models, *The Canadian Journal of Statistics*, **22**(3), 247–258.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics 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. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

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 and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: 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.

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 with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words 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), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal 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: revstat@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.

Editorial Board