ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 23 - n. 4 - October 2025

Instituto Nacional de Estatística
Statistics Portugal
1935-2025

REVSTAT — Statistical Journal, vol.23, n. 4 (October 2025)

vol.1, 2003- . - Lisbon : Statistics Portugal, 2003- .

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Editorial Board (2024-2025)

Publisher – Statistics Portugal

Layout-Graphic Design – *Carlos Perpétuo* | **Cover Design*** – *Helena Nogueira*

Digital edition

**image*: stain glass window by Abel Manta (1888-1982)

INDEX

Abstracted/indexed in: *Journal Citation Reports™ – JCR (Clarivate); DOAJ – Directory of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; SCImago Journal & Country Rank; Scopus®.*

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 441–462
https://doi.org/10.57805/revstat.v23i4.497

Bayesian Variable Selection for Zero-Inflated Longitudinal Count Data

Authors: Nawar Alsalim
– Department of Statistics, Faculty of Sciences, Al-Baath University, Homs, Syria
– Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
Nawjimaa@gmail.com

Taban Baghfalaki ✉
– Inserm, Research Center U1219, Univ. Bordeaux, ISPED, F33076 Bordeaux, France
taban.baghfalaki@u-bordeaux.fr

Received: July 2022 Revised: October 2023 Accepted: October 2023

Abstract:

- In this paper, we present a Bayesian variable selection method for zero-inflated longitudinal data. For this purpose, we consider a zero-inflated power series random effects model that includes the zero-inflated Poisson and negative binomial random effects models. We propose using continuous spike and Dirac spike priors to simultaneously estimate the regression coefficients and select the important covariate variables. We apply the MCMC method using Gibbs sampling for posterior inference. Some simulation studies are performed to investigate the performance of the proposed approach, and it is also applied to analyze a real dataset from the RAND Health Insurance Experiment.

Keywords:

- *Bayesian variable selection; continuous spike; Dirac spike; longitudinal data; power series family; random effects models.*

AMS Subject Classification:

- 62F15, 62J99.

✉ Corresponding author

1. INTRODUCTION

In medical research, count variables with many zeros are pervasive. Models that deal with and analyze a high proportion of zeros are known as zero-inflated models. Heilbron (1994) and Hu *et al.* (2011) performed Hurdle models for modeling zero-inflated data. Also, the zero-inflated negative binomial (ZINB) regression is used for count data that exhibit overdispersion and excess zeros. In general, zero-inflated power series distributions are applied to assess the excess of zeros (Luna *et al.*, 2020). Gupta *et al.* (1996) have studied zero-inflated modified power series distributions along with their applications for simulated data. Murat and Szynal (1998) considered non-zero-inflated modified power series distributions and extended the results of Gupta *et al.* (1996). Neelon *et al.* (2010) discussed a Bayesian paradigm for the ZIP and ZINB models for analyzing the data set of a study of psychiatric outpatient services. Rose *et al.* (2006) discussed the application of the ZI and Hurdle models for longitudinal studies concerning vaccination safety. Patil and Shirke (2007) studied different aspects of the zero-inflated power series distributions. Bekalo and Kebede (2021) considered the ZIP, ZINB, and Hurdle models, to observe whether there is any effect of the proportion of zeros in the performance of the models with the given overall rate of the counts. Feng (2021) reviewed the zero-inflated and hurdle models and highlighted their differences in terms of their data-generating processes. Young *et al.* (2021b) surveyed the developments in handling zero inflation for correlated count settings. Baghfalaki *et al.* (2021) discussed the approximate Bayesian approach for zero-inflated longitudinal models.

In longitudinal studies, data are collected repeatedly for the same set of units on more than one occasion. Random effect models have often been used in longitudinal data analysis since they allow for association among repeated measurements due to unobserved heterogeneity. To take into account the correlation among repeated measurements for each subject, zero-inflated count models with random effects have been developed. For example, a random effect was used to account for the within-subject dependency in the Poisson part of the ZIP model (Hall, 2000). Min and Agresti (2005) proposed a random effect model to analyze the ZI longitudinal count data. Lee *et al.* (2006) incorporated shared subject-specific random effects in each part of the zero-inflated model to account for zero-inflation and overdispersion within longitudinal count measurements.

Variable selection is an essential part of regression modeling for longitudinal data because many variables are measured and it is common in practice to include only a subset of important variables in the model. Zeng *et al.* (2014) discussed a variable selection approach for zero-inflated count data analysis based on the adaptive lasso technique. Mitchell and Beauchamp (1988) assumed that the regression coefficients were mutually independent with a two-point mixture distribution made up of a uniform flat distribution (the slab) and a degenerate distribution at zero (the spike). George and McCulloch (1993) used a different prior for the regression coefficients. This involved a scale (variance) mixture of two normal distributions. In particular, the use of a normal prior was instrumental in facilitating efficient Gibbs sampling of the posterior. This made spike and slab variable selection computationally attractive and heavily popularized the method. Normal-scale mixture priors constitute a wide class of models termed spike and slab models (Ishwaran and Rao, 2005). Spike and slab models are extended to the class of re-scaled spike and slab models (Ishwaran and Rao, 2005).

Lee *et al.* (2020) developed a Bayesian variable selection model for multivariate count data with excess zeros that incorporates information on the covariance structure of the outcomes. García-Donato *et al.* (2021) introduced the basic concepts of the Bayesian approach for variable selection based on model choice. Young *et al.* (2021a) studied some of the classic and contemporary literature on parametric zero-inflated count regression models. Gu *et al.* (2020) studied Bayesian variable selection in high dimensional data sets while simultaneously accounting for the error-prone nature of self-reported outcomes. Ji and Shi (2020) presented a Bayesian analysis of linear mixed models for quantile regression based on a Cholesky decomposition of the covariance matrix of random effects. Miao *et al.* (2020) considered linear regression models for count data, specifically negative binomial regression models and Dirichlet-multinomial regression models, they also addressed variable selection criteria via the use of spike-and-slab priors on the regression coefficients. Alsalim and Baghfalaki (2021) discussed new variable selection methods for the power series, specifically ZIP and ZINB transition models using LASSO, MCP, and SCAD penalties for analyzing longitudinal count data with extra zeros.

In this paper, we focus primarily on Bayesian approaches for variable selection that use spike and slab priors in the zero-inflated power series (ZIPS) model. For posterior inference, we apply MCMC methods via Gibbs sampling, and a test for variable selection of regression coefficients is considered by using both a local Bayesian false discovery rate and a Bayes factor procedure. After checking the performance of the proposed model using some simulation studies, we apply the proposed method to analyze the RAND health insurance experiment data.

This paper is organized as follows. Section 2 is a review of ZIPS distributions and the use of these distributions for analyzing zero-inflated longitudinal data. Also, this section includes some notation and definitions of models. Section 3 includes the likelihood functions, the Bayesian variable selection method, using spike and slab priors, and the test for variable selection of the regression coefficients. In Section 4, some simulation studies are performed. In Section 5, after describing the RAND health insurance experiment data, the data is analyzed using the proposed approaches. The last Section includes some conclusions.

2. MATERIALS AND METHODS

2.1. Notation

Let Y_{ij}, $i=1,...,n$ and $j=1,...,T$, be the longitudinal measurements for the i-th individual at the j-th time point. The ZIPS model is given as follows:

$$P\big(Y_{ij}=y_{ij}\,|\,\mu_{ij},\pi_{ij}\big) = \begin{cases} (1-\pi_{ij})\,p(Y_{ij}=y_{ij}\,|\,\mu_{ij}), & y_{ij}=1,2,..., \\ \pi_{ij}+(1-\pi_{ij})\,p(Y_{ij}=0\,|\,\mu_{ij}), & y_{ij}=0, \end{cases} \tag{2.1}$$

where $p(Y_{ij} = y_{ij} \mid \mu_{ij}, \pi_{ij})$ is a member of the power series (PS) family with the general form of

$$\frac{b_{y_{ij}}\, \mu_{ij}^{y_{ij}}}{f(\mu_{ij})}, \qquad y_{ij} = 0, 1, ..., \quad i = 1, ..., n, \quad j = 1, ..., T, \tag{2.2}$$

where $b_{y_{ij}} > 0$, μ_{ij} is positive and $f(\mu_{ij}) = \sum_{y_{ij}=0}^{\infty} b_{y_{ij}}\, \mu_{ij}^{y_{ij}}$ is a finite and differentiable function of μ_{ij}. The Poisson distribution and the negative binomial distribution belong to the PS distributions with $b_{y_{ij}} = \frac{1}{y_{ij}!}$, $f(\mu_{ij}) = \exp(\mu_{ij})$, $b_{y_{ij}} = \frac{\Gamma(y_{ij}+\theta)}{\Gamma(y_{ij}+1)}$, and $f(\mu_{ij}) = \Gamma(\theta)(1-\mu_{ij})^{-\theta}$, respectively, where $\theta > 0$ is an overdispersion parameter for negative binomial model. For considering ZIPS random effects models μ_{ij} and π_{ij}, $i = 1, ..., n$, $j = 1, ..., T$, are considered as follows:

$$\begin{aligned} \log(\mu_{ij}) &= \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + \boldsymbol{\kappa}'_{i1}\boldsymbol{b}_{i1}, \\ \operatorname{logit}(\pi_{\mathrm{ij}}) &= \boldsymbol{z}'_{ij}\boldsymbol{\alpha}_j + \boldsymbol{\kappa}'_{i2}\boldsymbol{b}_{i2}, \end{aligned} \tag{2.3}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)'$ are the outcome-specific vectors of fixed-effect regression coefficients. Also, $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_2$ are, respectively, q_1-dimensional and q_2-dimensional explanatory variables. The random effects $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}, \boldsymbol{b}_{i2})'$, $i = 1, ..., n$, characterize the unobserved characteristics that are associated with the mean count for time j of subject i such that $\boldsymbol{b}_i \sim N_{q_1+q_2}(\boldsymbol{0}, \boldsymbol{D})$, where $N_K(\boldsymbol{0}, \boldsymbol{D})$ denotes a K-variate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{D}$.

2.2. Bayesian variable selection for ZIPS random effects model

We complete the Bayesian formulation of the proposed framework by specifying prior distributions for the unknown parameters. To facilitate outcome-specific variable selection, we adopt spike and slab priors for the regression parameters of equation (2.3). A spike and slab prior is a mixture of spike and slab distributions, where the spike is a distribution with its mass concentrated around zero and the slab is a flat distribution spread over the parameter space. The spike component, representing a null effect, can be either a positive mass at zero (Dirac spike, DS, Lee *et al.*, 2020; Xu and Ghosh, 2015) or a normal distribution with mean zero and a small variance (continuous spike, CS, George and McCulloch, 1993). In DS, a point mass at zero represents the prior belief that the coefficient in the regression equation is zero and the corresponding predictor has no relevance to the outcome, but in CS, each predictor's coefficient is modeled as coming from a mixture of two normal distributions with different variances: one with a density concentrated around zero, the other with a density spread out over large plausible values. Thus, unlike DS, it allows for 'almost zero' regression coefficients which is a much more realistic assumption than assuming that a predictor has absolutely no effect on the outcome. The slab component represents a non-null effect. Mitchell and Beauchamp (1988) introduced this type before facilitating variable selection by constraining regression coefficients to be zero or not. Such a prior has been widely used in the context of Bayesian stochastic search variable selection (George and McCulloch, 1993). Figure 1 shows graphical examples of the CS and DS priors; slab densities are colored red and spike densities are colored blue.

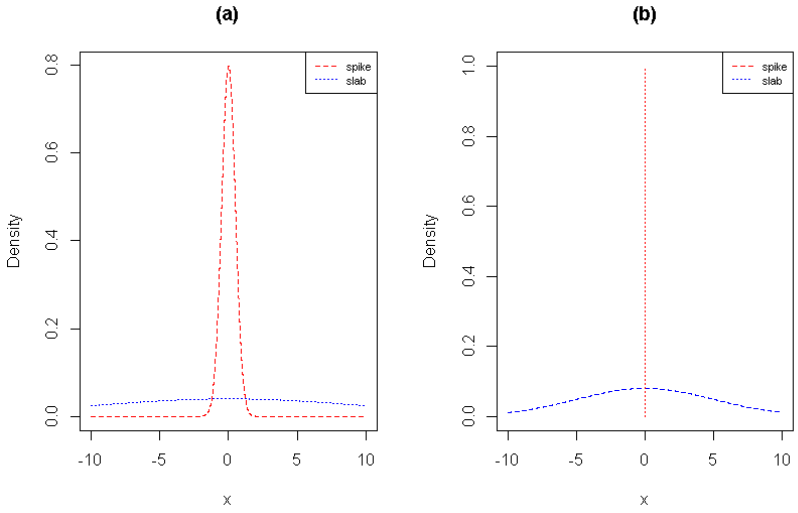

Figure 1: Example of the continuous spike (a) and the Dirac spike priors (b).

2.2.1. Continuous spike

The hierarchical setup of the zero-inflated random effects power series with CS prior (George and McCulloch, 1993; Baghfalaki *et al.*, 2021) for the regression coefficients of rate and probability models is given by:

$$
\begin{aligned}
&Y_{ij}\,|\,\boldsymbol{b}_i \sim \text{ZIPS}(\pi_{ij},\,\mu_{ij},\,\boldsymbol{a}),\\
&\boldsymbol{b}_i \sim N_2(\boldsymbol{0},\,\boldsymbol{D}),\\
&\beta_k\,|\,\zeta_k,\sigma^2_{\beta_k} \sim \zeta_k\,N(0,\,\sigma^2_{\beta_k}) + (1-\zeta_k)\,N(0,\,\tau^2_{\beta_k}), \qquad k=1,...,p,\\
&\tau^2_{\beta_k}\,|\,c_{1k},c_{2k} \sim \text{IG}(c_{1k},\,c_{2k}),\\
&\zeta_k\,|\,\lambda_{\beta_k} \sim \text{Ber}(\lambda_{\beta_k}),\\
&\lambda_{\beta_k}\,|\,f_{1k},f_{2k} \sim \text{Beta}(f_{1k},\,f_{2k}),\\
&\alpha_l\,|\,\omega_l,\sigma^2_{\alpha_l} \sim \omega_l\,N(0,\sigma^2_{\alpha_l}) + (1-\omega_l)\,N(0,\tau^2_{\alpha_l}), \qquad l=1,...,q,\\
&\tau^2_{\alpha_l}\,|\,d_{1l},d_{2l} \sim \text{IG}(d_{1l},\,d_{2l}),\\
&\omega_l\,|\,\lambda_{\alpha_l} \sim \text{Ber}(\lambda_l),\\
&\lambda_{\alpha_l}\,|\,m_{1l},m_{2l} \sim \text{Beta}(m_{1l},\,m_{2l}),\\
&\boldsymbol{D} \sim \text{IWishart}(r,\,\boldsymbol{\Psi}),\\
&\boldsymbol{a} \sim \pi(\boldsymbol{a}),
\end{aligned}
\tag{2.4}
$$

where $\boldsymbol{\zeta}=(\zeta_1,\cdots,\zeta_p)'$, $\boldsymbol{\omega}=(\omega_1,\cdots,\omega_q)'$ are vectors of binary latent variables indicating the membership of each regression coefficient to one of the mixture components, such that if $\zeta_k=1$, then $\beta_k\sim N(0,\,\sigma^2_{\beta_k})$, otherwise, $\beta_k\sim N(0,\,\tau^2_{\beta_k})$ and its components can be considered non-zero values for large values of $\tau^2_{\beta_k}$ and $\boldsymbol{a}$ are the other parameters. The values of $\sigma^2_{\beta_k}$ and $\sigma^2_{\alpha_l}$ should be small (e.g. 10^{-3} or 10^{-4}). By these conditions, $N(0,\sigma^2_{\beta_k})$ and $N(0,\sigma^2_{\alpha_l})$ lead us to a spike prior. The values of c_{1k} and c_{2k} are considered such that τ_{β_k} is large enough to yield a slab prior. ZIPS$(\cdot,\cdot,\cdot)$ is used to denote a zero-inflated power series random effects model,

IG$(\cdot,\cdot)$ denotes an inverse gamma distribution, Beta$(\cdot,\cdot)$ denotes a beta distribution, Ber$(\cdot)$ is used to denote a Bernoulli distribution, $\pi(\boldsymbol{a})$ denotes the prior of $\boldsymbol{a}$; for example, $a > 0$ is an overdispersion parameter in the ZINB model, and the $\pi(a) = \Gamma(r_1, r_2)$ where $\Gamma(\cdot,\cdot)$ denotes a gamma distribution and IWishart$(r, \boldsymbol{\Psi})$ denotes an inverse Wishart distribution with parameters degrees of freedom r and scale matrix $\boldsymbol{\Psi}$. Note that the natural conjugate prior to the multivariate normal distribution is the inverse Wishart distribution (Barnard *et al.*, 2000). Due to its conjugacy, this is the most common prior implemented in the Bayesian paradigm. However, this prior has issues: the uncertainty for all variances is controlled by a single degree of freedom parameter (r) (Gelman *et al.*, 2013), the marginal distribution for the variances has a low density in a region near zero (Gelman, 2006), and there is a prior dependence between correlations and variances (Tokuda *et al.*, 2011). These characteristics of the prior can impact posterior inferences about the covariance matrix. Here, the hyper-parameters are chosen to be uninformative in the simulation study and application sections chosen to be uninformative.

2.2.2. Dirac spike

The hierarchical setup of zero-inflated random effects power series with a DS prior (Lee *et al.*, 2020; Xu and Ghosh, 2015) for the regression coefficients of rate and probability models is given by:

$$
\begin{aligned}
&Y_{ij}\,|\,\boldsymbol{b}_i \sim \text{ZIPS}(\pi_{ij}, \mu_{ij}, \boldsymbol{a}),\\
&\boldsymbol{b}_i \sim N_2(\boldsymbol{0}, \boldsymbol{D}),\\
&\beta_k\,|\,\gamma_k, \sigma^2_{\beta_k} \sim \gamma_k\,\delta_0(\beta_k) + (1-\gamma_k)\,N(0, \sigma^2_{\beta_k}), \qquad k = 1, ..., p,\\
&\sigma^2_{\beta_k} \sim \text{IG}(c_{1k}, c_{2k}),\\
&\gamma_k \sim \text{Beta}(f_{1k}, f_{2k})\\
&\alpha_l\,|\,\nu_l, \sigma^2_{\alpha_l} \sim \nu_l\,\delta_0(\alpha_l) + (1-\nu_l)\,N(0, \sigma^2_{\alpha_l}), \qquad l = 1, ..., q,\\
&\sigma^2_{\alpha_l} \sim \text{IG}(d_{1l}, d_{2l}),\\
&\nu_l \sim \text{Beta}(m_{1l}, m_{2l}),\\
&\boldsymbol{D} \sim \text{IWishart}(r, \boldsymbol{\Psi}),\\
&\boldsymbol{a} \sim \pi(\boldsymbol{a}),
\end{aligned}
\tag{2.5}
$$

where $\delta_0(\cdot)$ denotes a Dirac mass at 0, such that $\delta_0(\beta_k) = 1$ if $\beta_k = 0$ and $\delta_0(\beta_k) = 0$ if $\beta_k \neq 0$ and the other notations are the same as those described for CS.

3. STATISTICAL INFERENCE

Our inference is based on Metropolis–Hastings within Gibbs samplers because the full conditional distributions of the regression coefficients and the random effects do not have closed forms. The full conditional posterior distributions of all parameters and for all models are presented in supplementary materials A and B for CS and DS, respectively. A local Bayesian false discovery rate and a Bayes factor are proposed to perform the test of checking the significance of the regression coefficients (Efron, 2012).

3.1. Bayesian implementation

3.1.1. Continuous spike

Let $\boldsymbol{\theta} = \big(\boldsymbol{b}, \boldsymbol{D}, \boldsymbol{a}, \{\beta_k, \sigma^{2r}_{\beta_k}, \zeta_k, \lambda_{\beta_k}\}_{k=1}^{p}, \{\alpha_l, \sigma_{\alpha_l}, \omega_l, \lambda_{\alpha_l}\}_{l=1}^{q}\big)$ be the vector of all the unknown parameters in the model, $\boldsymbol{y} = (\boldsymbol{y}_i, \cdots, \boldsymbol{y}_n)'$, $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$, $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_n)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$, $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})'$, $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})'$ and $\boldsymbol{z}_i = (z_{i1}, \cdots, z_{iq})'$. The likelihood function of the model can be written as:

$$\begin{aligned} L(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\mu}) \;=\; & \prod_{i=1}^{n}\prod_{j=1}^{T}\big\{(1-\pi_{ij})\,p(Y_{ij}=y_{ij}\,|\,\mu_{ij})\big\}^{1-I(y_{ij})} \\ & \times \big\{\pi_{ij} + (1-\pi_{ij})\,p(Y_{ij}=0\,|\,\mu_{ij})\big\}^{I(y_{ij})}, \end{aligned} \tag{3.1}$$

where

$$I(Y_{ij}) = \begin{cases} 0, & y_{ij} = 1, 2, ..., \\ 1, & y_{ij} = 0. \end{cases}$$

The joint posterior distribution of the unknown parameters is as follows:

$$\begin{aligned} \pi(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \boldsymbol{x}) \;\propto\; & L(\boldsymbol{\theta}\,|\,\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\mu}) \times p(\boldsymbol{\beta}\,|\,\boldsymbol{\tau}_1^2, \boldsymbol{\zeta}, \boldsymbol{\lambda}_1) \times p(\boldsymbol{\tau}_1^2) \times p(\boldsymbol{\alpha}\,|\,\boldsymbol{\tau}_2^2, \boldsymbol{\omega}, \boldsymbol{\lambda}_2) \times p(\boldsymbol{\tau}_2^2) \times p(\boldsymbol{b}\,|\,\boldsymbol{D}) \\ & \times p(\boldsymbol{D}) \times p(\boldsymbol{\zeta}\,|\,\boldsymbol{\lambda}_1) \times p(\boldsymbol{\omega}\,|\,\boldsymbol{\lambda}_2) \times p(\boldsymbol{\lambda}_1) \times p(\boldsymbol{\lambda}_2) \times p(\boldsymbol{a}) \\ \propto\; & \prod_{i=1}^{n}\prod_{j=1}^{T}\big\{(1-\pi_{ij})\,p(Y_{ij}=y_{ij}\,|\,\mu_{ij})\big\}^{1-I(y_{ij})}\,\big\{\pi_{ij} + (1-\pi_{ij})\,p(Y_{ij}=0\,|\,\mu_{ij})\big\}^{I(y_{ij})} \\ & \times \phi(\boldsymbol{b}_i, \boldsymbol{0}, \boldsymbol{D})\,|\boldsymbol{D}|^{\frac{-(r+2+1)}{2}}\,\exp\!\left(-\frac{\boldsymbol{D}^{-1}\boldsymbol{\Psi}}{2}\right) \\ & \times \prod_{k=1}^{p}\zeta_k\,\phi(\beta_k, 0, \sigma^2_{\beta_k}) + (1-\zeta_k)\,\phi(\beta_k, 0, \tau^2_{\beta_k})\,\tau^{2c_{1k}-1}_{\beta_k}\exp\!\left(-\frac{c_{2k}}{\tau^2_{\beta_k}}\right) \\ & \times \prod_{l=1}^{q}\omega_l\,\phi(\alpha_l, 0, \sigma^2_{\alpha}) + (1-\omega_l)\,\phi(\alpha_l, 0, \tau^2_{\alpha})\,\tau^{2d_{1l}-1}_{\alpha\iota}\exp\!\left(-\frac{d_{2l}}{\tau^2_{\alpha_l}}\right) \\ & \times \prod_{l=1}^{q}(\lambda_{\alpha_l})^{\omega_l}(1-\lambda_{\alpha_l})^{1-\omega_l}\prod_{k=1}^{p}(\lambda_{\beta_k})^{\zeta_l}(1-\lambda_{\beta_k})^{1-\zeta_l} \\ & \times \prod_{k=1}^{p}(\lambda_{\beta_k})^{f_{1k}-1}(1-\lambda_{\beta_k})^{f_{2k}-1}\prod_{l=1}^{q}(\lambda_{\alpha_l})^{m_{1l}-1}(1-\lambda_{\alpha_l})\,\pi(\boldsymbol{a}), \end{aligned}$$

where $\boldsymbol{\lambda}_1 = (\lambda_{\beta_1}, \cdots, \lambda_{\beta_p})'$ and $\boldsymbol{\lambda}_2 = (\lambda_{\alpha_1}, \cdots, \lambda_{\alpha_q})'$. For applying MCMC methods, the full conditional posterior distributions of all the unknown parameters for this model are computed and presented in supplementary material A.

3.1.2. Dirac spike

Let $\boldsymbol{\theta} = \big(\boldsymbol{b}, \boldsymbol{D}, \boldsymbol{a}, \{\beta_k, \sigma^2_{\beta_k}, \gamma_k\}_{k=1}^{p}, \{\alpha_l, \sigma^2_{\alpha_l}, \nu_l\}_{l=1}^{q}\big)$ be the vector of all the unknown parameters in the model, $\boldsymbol{y} = (\boldsymbol{y}_i, \cdots, \boldsymbol{y}_n)'$, $\boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$, $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_n)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$,

$\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})'$, $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{ip})'$ and $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{iq})'$. The joint posterior distribution of all unknown parameters, given data, is as follows:

$$\begin{aligned}
\pi(\boldsymbol{\theta}\,|\,\boldsymbol{y},\boldsymbol{x}) \;\propto\; & L(\boldsymbol{\theta}\,|\,\boldsymbol{y},\boldsymbol{x},\boldsymbol{\pi},\boldsymbol{\mu}) \times p(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma},\boldsymbol{\sigma}_1^2) \times p(\boldsymbol{\sigma}_1^2) \times p(\boldsymbol{\alpha}\,|\,\boldsymbol{\nu},\boldsymbol{\sigma}_2^2) \times p(\boldsymbol{\sigma}_2^2) \times p(\boldsymbol{b}_i\,|\,\boldsymbol{D}) \\
& \times p(\boldsymbol{D}) \times p(\boldsymbol{\gamma}) \times p(\boldsymbol{\nu}) \times p(\boldsymbol{a}) \\
\propto\; & \prod_{i=1}^{n}\prod_{j=1}^{T}\big\{(1-\pi_{ij})\,p(Y_{ij}=y_{ij}\,|\,\mu_{ij})\big\}^{1-I(y_{ij})}\,\big\{\pi_{ij}+(1-\pi_{ij})\,p(Y_{ij}=0\,|\,\mu_{ij})\big\}^{I(y_{ij})} \\
& \times \phi(\boldsymbol{b}_i,\boldsymbol{0},\boldsymbol{D})\,|\boldsymbol{D}|^{\frac{-(r+2+1)}{2}}\,\exp\!\left(-\frac{\boldsymbol{D}^{-1}\boldsymbol{\Psi}}{2}\right) \\
& \times \prod_{k=1}^{p}\gamma_k\,\phi(\beta_k,0,\sigma_{\beta_k}^2)+(1-\gamma_k)\,\delta_0(\beta_k)\,\sigma_{\beta_k}^{2c_{1k}-1}\exp\!\left(-\frac{c_{2k}}{\sigma_{\beta_k}^2}\right) \\
& \times \prod_{l=1}^{q}\nu_l\,\phi(\alpha_l,0,\sigma_{\alpha_l}^2)+(1-\nu_l)\,\delta_0(\alpha)\,\sigma_{\alpha\iota}^{2d_{1l}-1}\exp\!\left(-\frac{d_{2l}}{\sigma_{\alpha_l}^2}\right) \\
& \times \prod_{k=1}^{p}\gamma_k^{f_{1k}-1}(1-\gamma_k)^{f_{2k}-1}\prod_{l=1}^{q}\nu_l^{m_{1l}-1}(1-\nu_l)^{m_{2l}-1}\,\pi(\boldsymbol{a}),
\end{aligned}$$

where $\boldsymbol{\sigma}_1^2 = (\sigma_{\beta_1}^2, \cdots, \sigma_{\beta_p}^2)'$, $\boldsymbol{\sigma}_2^2 = (\sigma_{\alpha_1}^2, \cdots, \sigma_{\alpha_q}^2)'$, $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p)'$ and $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_q)'$. For applying MCMC methods, the full conditional posterior distributions of all the unknown parameters for this model are computed and presented in supplementary material B.

3.2. Variable selection

In the following, we propose some strategies to select variables in the rate and probability models.

3.2.1. Continuous spike

Let $\boldsymbol{\theta}^{(r)}$, $r = 1, ..., M$, be M generated samples from the full conditional distributions of CS using MCMC. We first define a test for checking the significance of the parameters, as follows:

$$\begin{aligned}
& H_{0k}:\,\zeta_k = 0 \quad \textit{versus} \quad H_{1k}:\,\zeta_k \neq 0\,, \qquad k = 1, ..., p, \\
& H_{0l}:\,\omega_l = 0 \quad \textit{versus} \quad H_{1l}:\,\omega_l \neq 0\,, \qquad l = 1, ..., q.
\end{aligned} \tag{3.2}$$

Both a local Bayesian false discovery rate and a Bayes factor procedure are applied to perform this test.

Local Bayesian false discovery rate

Let $p(H_{0k}\,|\,\beta_k, \sigma_{\beta_k}^2)$ be the posterior probability of the null hypothesis, i.e., the probability of making a false discovery for a non-null effect, which is called the local Bayesian false

discovery rate for data (denoted by LBFDR). The phrase "local" comes from a single point 0 as the domain of the null hypothesis test (Efron, 2012). Note that small values of LBFDR show strong evidence for the existence of a substantive effect. For computing LBFDR for each of the regression coefficients, we have

$$\begin{aligned}
\text{LBFDR}_k &= P\big(H_{0k}\,|\,\beta_k,\sigma^2_{\beta_k}\big) = P\big(\zeta_k=0\,|\,\beta_k,\sigma^2_{\beta_k}\big)\\
&\approx \frac{1}{M}\sum_{r=1}^{M} P\big(\zeta_k=0\,|\,\beta_k^{(r)},\sigma^{2(r)}_{\beta_k}\big)\,, \qquad k=1,...,p,\\
\text{LBFDR}_l &= P\big(H_{0l}\,|\,\alpha_l,\sigma^2_{\alpha_l}\big) = P\big(\omega_l=0\,|\,\alpha_l,\sigma^2_{\alpha_l}\big)\\
&\approx \frac{1}{M}\sum_{r=1}^{M} P\big(\omega_l=0\,|\,\alpha_l^{(r)},\sigma^{2(r)}_{\alpha_l}\big)\,, \qquad l=1,...,q,
\end{aligned} \tag{3.3}$$

where $\big(\beta_k^{(r)},\sigma^{2(r)}_{\beta_k},\alpha_l^{(r)},\sigma^{2(r)}_{\alpha_l}\big)$ denotes the r-th generated sample of $\big(\beta_k,\sigma^2_{\beta_k},\alpha_l,\sigma^2_{\alpha_l}\big)$ using the MCMC for $r=1,...,M$.

Bayes factor

The Bayes factor for each of the regression coefficients for testing (3.2) is defined as

$$\begin{aligned}
\text{BF}_k &= \frac{P\big(H_{1k}\,|\,\beta_k,\sigma^2_{\beta_k}\big)/P\big(H_{1k}\big)}{P\big(H_{0k}\,|\,\beta_k,\sigma^2_{\beta_k}\big)/P\big(H_{0k}\big)},\\
\text{BF}_l &= \frac{P\big(H_{1l}\,|\,\alpha_l,\sigma^2_{\alpha_l}\big)/P\big(H_{1l}\big)}{P\big(H_{0l}\,|\,\alpha_l,\sigma^2_{\alpha_l}\big)/P\big(H_{0l}\big)},
\end{aligned} \tag{3.4}$$

which describes the evidence of $H_{1k}\,(H_{1l})$ against $H_{0k}\,(H_{0l})$. Note that $P(\zeta_k)=E(\lambda_{\beta_k})=\frac{f_{1k}}{f_{1k}+f_{2k}}$ and $P(\omega_l)=E(\lambda_{\alpha_l})=\frac{m_{1l}}{m_{1l}+m_{2l}}$. Also, since $\text{LBFDR}_k=P(H_{0k}\,|\,\beta_k,\sigma^2_{\beta_k})$ and $\text{LBFDR}_l=P(H_{0l}\,|\,\alpha_l,\sigma^2_{\alpha_l})$,

$$\begin{aligned}
\text{BF}_k &= \frac{1-\text{LBFDR}_k}{\text{LBFDR}_k}\times\frac{f_{1k}}{f_{1k}+f_{2k}},\\
\text{BF}_l &= \frac{1-\text{LBFDR}_l}{\text{LBFDR}_l}\times\frac{m_{1l}}{m_{1l}+m_{2l}}.
\end{aligned} \tag{3.5}$$

Unlike LBFDR, a large value of BF indicates strong evidence in favor of $H_{1k}\,(H_{1l})$.

3.2.2. Dirac spike

The same as those discussed for CS, let $\boldsymbol{\theta}^{(r)}$, $r=1,...,M$, be M generated samples from the full conditional distributions of DS using MCMC. The global test for checking DS is defined by:

$$\begin{aligned}
&H_{0k}:\ \beta_k=0 \quad \textit{versus} \quad H_{1k}:\ \beta_k\neq 0\,, \qquad k=1,...,p,\\
&H_{0l}:\ \alpha_l=0 \quad \textit{versus} \quad H_{1l}:\ \alpha_l\neq 0\,, \qquad l=1,...,q.
\end{aligned} \tag{3.6}$$

Local Bayesian false discovery rate

In this status, the following proposition gives insight into simplifying equation (3.6). Define indicator variables: I_{1k}, I_{2l}, $k = 1, 2, ..., p$, $l = 1, 2, ..., q$, such that

$$I_{1k} = \begin{cases} 1, & \beta_k \neq 0, \\ 0, & \beta_k = 0, \end{cases}$$

$$I_{1l} = \begin{cases} 1, & \alpha_l \neq 0, \\ 0, & \alpha_l = 0. \end{cases} \tag{3.7}$$

Also, consider the hierarchical model (2.5), thus,

$$\begin{aligned} \text{LBFDR}_k &= P\big(\beta_k = 0 | \gamma_k, \sigma^2_{\beta_k}\big) \\ &\approx \frac{1}{M} \sum_{r=1}^{M} P\big(\beta_k = 0 | \gamma_k^{(r)}, \sigma^{2(r)}_{\beta_k}\big), \\ \text{LBFDR}_l &= P\big(\alpha_l = 0 | \nu_l, \sigma^2_{\alpha_k}\big) \\ &\approx \frac{1}{M} \sum_{r=1}^{M} P\big(\alpha_l = 0 | \nu_l^{(r)}, \sigma^{2(r)}_{\alpha_l}\big). \end{aligned} \tag{3.8}$$

Bayes factor

The Bayes factor for each of the regression coefficients for DS is given by

$$\begin{aligned} \text{BF}_k &= \frac{P\big(\beta_k \neq 0 | \gamma_k, \sigma^2_{\beta_k}\big)/P\big(\beta_k \neq 0\big)}{P\big(\beta_k = 0 | \gamma_k, \sigma^2_{\beta_k}\big)/P\big(\beta_k = 0\big)}, \\ \text{BF}_l &= \frac{P\big(\alpha_l \neq 0 | \nu_l, \sigma^2_{\alpha_l}\big)/P\big(\alpha_l \neq 0\big)}{P\big(\alpha_l = 0 | \nu_l, \sigma^2_{\alpha_l}\big)/P\big(\alpha_l = 0\big)}, \end{aligned} \tag{3.9}$$

we have $P(\beta_k) = E(\gamma_{\beta_k}) = \frac{f_{1k}}{f_{1k}+f_{2k}}$ and $P(\alpha_l) = E(\nu_{\alpha_l}) = \frac{m_{1l}}{m_{1l}+m_{2l}}$. Thus the Bayes factor for DS is the same as that of equation (3.5).

4. SIMULATION STUDIES

In this section, some simulation studies are performed to investigate the performance of the proposed methods. For this purpose, the data is generated from random effects models under ZIP and ZINB. The sample sizes $n = 500$ and 1000 with $T = 6$ repeated measurements, $p = 15$, and $q = 10$ predictors are considered. Also, we consider 40000 MCMC iterations, including 20000 pre-convergence burn-in. The convergence of the chains is checked using Brooks–Gelman–Rubin (BGR) diagnostics (Brooks and Gelman, 1998; Gelman and Rubin, 1992). Also, some figures are given in supplementary material E for checking the convergence of the proposed model visually. The simulation studies are performed for $M = 100$ replications.

For comparison of the results, relative bias (Rbias) and the root of the mean squared error (RMSE) are computed, these are defined as $\text{Rbias}(\theta) = \dfrac{\bar{\hat{\theta}}}{\theta} - 1$, $\text{RMSE}(\theta) = \sqrt{\dfrac{\sum_{r=1}^{M}(\hat{\theta}_r - \theta)^2}{M}}$, where $\hat{\theta}_r$ is the estimated value of parameter θ for the r-th simulation run, M is the number of simulation runs, and $\bar{\hat{\theta}} = \dfrac{\sum_{r=1}^{M}\hat{\theta}_r}{M}$.

To investigate the performance of the proposed approaches in variable selection, we consider the true positive rate (TPR), the false positive rate (FPR), and the Matthews correlation coefficient (MCC) criteria (Matthews, 1975). The latter is defined as follows:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})\,(\text{TP} + \text{FN})\,(\text{TN} + \text{FP})\,(\text{TN} + \text{FN})}}, \tag{4.1}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. The MCC and TPR are expected to reach 1, and the FPR is expected to be near zero for a good performance.

4.1. Zero-inflated Poisson random effects model

We consider a ZIP random effects model, such that $Y_{ij}\,|\,\mu_{ij}, \pi_{ij} \sim \text{ZIP}(\mu_{ij}, \pi_{ij})$ is used to denote it, where μ_{ij} and π_{ij} are considered the same as equations (2.3). For this simulation study, the explanatory variables $\boldsymbol{x}$ and $\boldsymbol{z}$ are randomly drawn from multivariate normal distributions $N_{15}(\boldsymbol{0}, \boldsymbol{I})$ and $N_{10}(\boldsymbol{0}, \boldsymbol{I})$, respectively. Also, four scenarios are considered for the real values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ these scenarios are different for the real values of the regression coefficients:

Scenario 1:
$$\boldsymbol{\beta} = (\underbrace{1, \cdots, 1}_{5}, \underbrace{0, \cdots, 0}_{10})',$$
$$\boldsymbol{\alpha} = (\underbrace{1, \cdots, 1}_{5}, \underbrace{0, \cdots, 0}_{5})'.$$

Scenario 2:
$$\boldsymbol{\beta} = (\underbrace{0.5, \cdots, 0.5}_{5}, \underbrace{0, \cdots, 0}_{10})',$$
$$\boldsymbol{\alpha} = (\underbrace{0.5, \cdots, 0.5}_{5}, \underbrace{0, \cdots, 0}_{5})'.$$

Scenario 3:
$$\boldsymbol{\beta} = (\underbrace{0.5, \cdots, 0.5}_{15})',$$
$$\boldsymbol{\alpha} = (\underbrace{0.5, \cdots, 0.5}_{10})'.$$

Scenario 4:
$$\boldsymbol{\beta} = (\underbrace{0, \cdots, 0}_{15})',$$
$$\boldsymbol{\alpha} = (\underbrace{0, \cdots, 0}_{10})'.$$

In scenario 2, the values of the regression coefficients are reduced to check if the Bayesian approach for variable selection has a good performance after the reduction of the signals. Also, scenario 3 was simulated to represent cases in which all the covariates had non-zero effects, and scenario 4 was simulated to represent cases in which all the covariates had zero effects. Also, the real value for the covariance of the random effects is as follows:

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}.$$

For DS and CS, we set the hyperparameters to $c_{1k} = 0.01$, $c_{2k} = 0.01$, $d_{1l} = 0.01$, $d_{2l} = 0.01$, $f_{1k} = 1$, $f_{2k} = 1$, $m_{1l} = 1$, $m_{2l} = 1$, $k = 1, ..., p$, $l = 1, ..., q$, also for CS, we set $\sigma^2_{\beta_k} = \sigma^2_{\alpha_l} = 10^{-4}$ and the hyperparameters of inverse Wishart distribution are considered as $r = 3$ and $\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, so that the priors are low-informative (Gu *et al.*, 2020; Gelman, 2006). The results of this simulation study are reported as follows:

Continuous spike. The results of this simulation study are reported in Tables 1 and 2 for scenario 1 and Tables C.1 and C.2 for scenario 2 of supplementary material C. Table 1 includes estimates of posterior mean, standard errors, posterior median, RMSE, and Rbias. The values of Rbiases and RMSEs for all the parameters are small, and by increasing the sample size from $n = 500$ to $n = 1000$, the accuracy and efficiency of the estimations are increased. Also, Table 2 reports TPR, FPR, and MCC, where the results are based on threshold 1 of BF and threshold 0.05 of LBFDR. The values of the MCC as a balanced measure between TPR and FPR, show that the performances of BF and LBFDR are similar. Overall, the results show that all the parameters are well estimated, and the values of the MCC show the good performance of the method in variable selection. The values of the regression coefficients are reduced in scenario 2, and the results of this simulation study are summarized in Tables C.1 and C.2. The results of these tables show similar results as those of scenario 1, even with smaller values of the signals.

Dirac spike. The results of these simulation studies are reported in Tables C.3 and C.4 for scenario 1. In Table C.4, the value of the criteria for median thresholding is also given. Table C.3 shows that all the parameters are well estimated; also, the values of the MCC are given in Table C.4 and confirm that the performances of BF and LBFDR are similar, and they are better than median thresholding. The values of the regression coefficients are reduced in scenario 2, and the results of this simulation study are summarized in Tables C.5 and C.6. The results of these tables show the same results as in scenario 1.

Table 1: Results of the simulation study of CS for generated data under the ZIP random effects model for scenario 1. The posterior mean, the standard deviation of estimators, the posterior median, the root of the mean squared error (RMSE), and relative bias (Rbias) for each of the parameter estimates for $M = 100$ simulated data with sample sizes of 500 and 1000. The generated data are analyzed with the ZIP model ($*$: the relative bias cannot be calculated since the real value of the related parameter is zero).

n	Parameter	True	Mean	SD	Median	RMSE	Rbias
500	β_1	1.000	0.991	0.044	0.993	0.004	−0.014
	β_2	1.000	0.959	0.046	0.959	0.003	−0.118
	β_3	1.000	0.966	0.043	0.968	0.004	−0.064
	β_4	1.000	0.923	0.044	0.925	0.003	−0.040
	β_5	1.000	0.989	0.044	0.990	0.000	−0.020
	β_6	0.000	−0.005	0.002	−0.002	0.000	$*$
	β_7	0.000	−0.003	0.002	−0.001	0.000	$*$
	β_8	0.000	−0.043	0.003	−0.044	0.002	$*$
	β_9	0.000	0.003	0.002	0.001	0.000	$*$
	β_{10}	0.000	0.011	0.004	0.003	0.000	$*$
	β_{11}	0.000	−0.005	0.003	−0.002	0.000	$*$
	β_{12}	0.000	0.020	0.004	0.018	0.000	$*$
	β_{13}	0.000	−0.031	0.004	−0.038	0.001	$*$
	β_{14}	0.000	−0.003	0.003	−0.001	0.000	$*$
	β_{15}	0.000	−0.002	0.002	−0.001	0.000	$*$
	α_1	1.000	0.955	0.063	0.454	0.002	−0.092
	α_2	1.000	0.918	0.065	0.919	0.000	−0.038
	α_3	1.000	0.943	0.065	0.942	0.002	−0.084
	α_4	1.000	0.992	0.065	0.992	0.000	−0.016
	α_5	1.000	0.977	0.065	0.976	0.001	−0.048
	α_6	0.000	0.000	0.005	0.000	0.000	$*$
	α_7	0.000	−0.005	0.005	−0.001	0.000	$*$
	α_8	0.000	0.002	0.006	0.000	0.000	$*$
	α_9	0.000	0.000	0.006	0.000	0.000	$*$
	α_{10}	0.000	0.016	0.009	0.002	0.000	$*$
1000	β_1	1.000	0.997	0.043	0.996	0.002	−0.008
	β_2	1.000	0.906	0.045	0.906	0.002	−0.012
	β_3	1.000	0.973	0.043	0.974	0.003	−0.052
	β_4	1.000	0.995	0.042	0.994	0.003	−0.052
	β_5	1.000	0.994	0.043	0.995	0.001	−0.010
	β_6	0.000	−0.007	0.002	−0.003	0.000	$*$
	β_7	0.000	−0.003	0.002	−0.001	0.000	$*$
	β_8	0.000	0.001	0.002	0.001	0.000	$*$
	β_9	0.000	−0.020	0.003	−0.022	0.000	$*$
	β_{10}	0.000	−0.001	0.002	0.000	0.000	$*$
	β_{11}	0.000	0.002	0.003	0.001	0.000	$*$
	β_{12}	0.000	0.001	0.002	0.001	0.000	$*$
	β_{13}	0.000	0.002	0.003	0.001	0.000	$*$
	β_{14}	0.000	0.000	0.002	0.000	0.000	$*$
	β_{15}	0.000	0.000	0.002	0.000	0.000	$*$
	α_1	1.000	0.987	0.061	0.986	0.000	−0.028
	α_2	1.000	0.907	0.063	0.906	0.000	−0.012
	α_3	1.000	0.928	0.062	0.927	0.001	−0.054
	α_4	1.000	0.908	0.063	0.908	0.000	−0.016
	α_5	1.000	0.970	0.063	0.969	0.001	−0.062
	α_6	0.000	0.003	0.005	0.001	0.000	$*$
	α_7	0.000	0.018	0.006	0.005	0.000	$*$
	α_8	0.000	−0.004	0.006	−0.001	0.000	$*$
	α_9	0.000	−0.001	0.005	0.000	0.000	$*$
	α_{10}	0.000	0.002	0.007	0.000	0.000	$*$

Table 2: Mean (SD) of true/false positive rate (TPR/FPR) and Matthews correlation coefficient (MCC) of BF and LBFDR for ZINB and ZIP random effects models of Scenario 1 for CS with $M = 100$ simulated data with sample sizes of 500 and 1000.

n		ZIP		ZINB	
		BF	LBFDR	BF	LBFDR
500	TPR	0.925 (0.096)	0.925 (0.096)	0.966 (0.070)	0.975 (0.061)
	FPR	0.033 (0.038)	0.000 (0.000)	0.053 (0.021)	0.000 (0.000)
	MCC	0.897 (0.106)	0.941 (0.077)	0.838 (0.088)	0.938 (0.079)
1000	TPR	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
	FPR	0.025 (0.011)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
	MCC	0.912 (0.011)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)

4.2. Zero-inflated negative binomial random effects model

In this simulation study, we simulate data from a ZINB random effects model as follows:

$$Y_{ij}\,|\,\mu_{ij},\pi_{ij} \;\sim\; \text{ZINB}\left(\phi,\, \frac{\phi}{\phi+\mu_{ij}},\, \pi_{ij}\right),$$

where μ_{ij} and π_{ij} are considered the same as equation (2.3). The parameterization and the real values of parameters μ_{ij} and π_{ij} are the same as the set of real values and those described for two scenarios in the previous subsection; also, we set $\phi = 2, 0.25$. The results of this simulation study are reported as follows:

Continuous spike. The results of this simulation study for $\phi = 2$ are reported in Tables 2 and C.7 for scenario 1 and Tables C.2 and C.8 for scenario 2. Also, the results of this simulation study for $\phi = 0.25$ are given in Tables E.1 and E.2 for scenario 1 and in Tables E.3 and E.4 for scenario 2 of supplementary material E. The performance of the proposed model is good in both parameter estimation and variable selection, and this is in agreement with different values of ϕ.

Dirac spike. The results of this simulation studies for $\phi = 2$ are reported in Tables C.4 and C.9 for scenario 1 and in Tables C.6 and C.10 for scenario 2. Also, the results of this simulation study are given in E.5 and E.6 for scenario 1 and in Tables E.7 and E.8 for scenario 2 when $\phi = 0.25$. The results of these tables, the same as those in CS, show that all the parameters are estimated well and the values of the MCC show the performance of the method for variable selection as well.

A comparison between CS and DS for variable selection in the proposed model for both ZIP and ZINB models shows that DS performs better than CS based on values of MCC. Also, the parameter estimates by DS are closer to the true values of the parameters than those obtained by CS.

4.3. Results of simulation studies for scenarios 3 and 4

A sample size of $n = 500$ is considered for scenarios 3 and 4, which are whole non-zero and whole zero signals, respectively. The results of these two scenarios, which are the same as the previous scenarios, include estimates, standard errors, posterior median, RMSE, and Rbias. It is not possible to check the performance of the variable selection of the proposed model by TPR, FPR, and MCC when all of the signals are significant or all of them are non-significant. Therefore, instead, the mean and standard deviation of LBFDR and BF are given in the tables of results for these two scenarios.

4.3.1. ZIP model

The results of this simulation study for CS prior are reported in Tables D.1 for scenario 3 and D.2 for scenario 4 of supplementary material D. The results show that all the parameters are well estimated, and based on the mean of BF and LBFDR, all the variables are selected for scenario 3, but none of them are selected for scenario 4.

The results of this simulation study for DS prior are reported in Table D.3 and Table D.4 for scenarios 3 and 4, respectively. The results of these tables also confirm the good performance of the proposed model.

4.3.2. ZINB model

The results of this simulation study for CS prior are given in Table D.5 for scenario 3 and in Table D.6 for scenario 4. As with our results for the ZIP model, the results show that the performance of the proposed model is good in parameter estimation and variable selection. Also, for the DS prior, the results are shown in Table D.7 for scenario 3 and in Table D.8 for scenario 4. The results of these tables also confirm the good performance of the model in both parameter estimation and variable selection.

5. APPLICATION: THE RAND HEALTH INSURANCE EXPERIMENT

In this section, we shall analyze the RAND Health Insurance Experiment (RAND HIE) data from Deb and Trivedi (2002). The data investigate how medical care utilization, measured by the number of visits to a medical doctor (MD), is affected by health insurance plans, demographic characteristics, and the health status of patients. This particular data set consists of 5792 participants with 20,190 observations in total. The vast majority of participants are observed either three or five times, and each observation corresponds to data collected for the participant in a given year. The response variable MD is the yearly count of outpatient visits to physicians, which represents the health care utilization for the experimental subject for a specific year.

Over 30% of the observations are zeros, motivating the use of the proposed approach. We use a simple zero-score test for checking the zero-inflation in the data, and not having zero-inflation in the data is rejected by a p-value of 0.000. The bar plot of this variable is presented in Figure 2 where the zero-inflation in the data set is also implied. The insurance variables were randomly assigned and included, an indicator variable for plans with a deductible (IDP), a participation-incentive payment function (LPI), a maximum dollar-expenditure function (FMDE), and other covariates including factors representing the demographic information including a log of annual family income (LINC), gender (FEMALE), race (BLACK), education of the head of household in years (EDUCDEC), age, an indicator for age less than 18 (CHILD), log of the family size (LFAM), a coinsurance rate (LC), health status including an indicator for physical limitations (PHYSLIM), index of chronic diseases (NDISEASE), fair self-rated health (HLTHF), good self-rated health (HLTHG) and poor self-rated health (HLTHP).

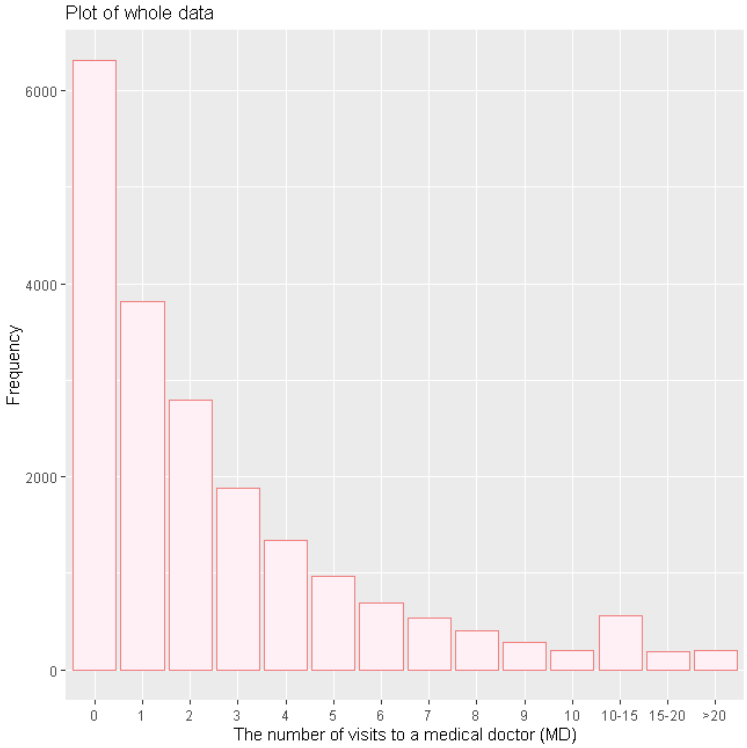

Figure 2: Barplot of the number of visits to a medical doctor of rand health Insurance data.

For detailed variable definitions and summary statistics of each variable, see Table F.4 of supplementary material D. For checking the effect of time, we let time be modeled as a square polynomial in the rate model (Baghfalaki and Ganjali, 2021). The proposed variable selection models are applied to analyze the data such that:

$$\begin{aligned}\log(\mu_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 \text{IDP}_i + \beta_4 \text{LPI}_i + \beta_5 \text{FMDE}_i \\ &+ \beta_6 \text{LINC}_i + \beta_7 \text{FEMALE}_i + \beta_8 \text{PHYSLM}_i + \beta_9 \text{BLACK}_i \\ &+ \beta_{10} \text{EDUCDEC}_i + \beta_{11} \text{NDISEASE}_i + \beta_{12} \text{HLTHF}_i + \beta_{13} \text{HLTHG}_i \\ &+ \beta_{14} \text{HLTHP}_i + \beta_{15} \text{AGE}_i + \beta_{16} \text{CHILD}_{ij} + \beta_{17} \text{LFAM}_i + \beta_{18} \text{LC}_{ij} \\ &+ \beta_{19} \text{FCHILD}_{ij} + b_{1i} \end{aligned} \tag{5.1}$$

and

$$\begin{aligned}\text{logit}(\pi_{ij}) &= \alpha_0 + \alpha_1 t_{ij} + \alpha_2 \text{IDP}_i + \alpha_3 \text{LPI}_i + \alpha_4 \text{FMDE}_i \\ &+ \alpha_5 \text{LINC}_i + \alpha_6 \text{FEMALE}_i + \alpha_7 \text{PHYSLM}_i + \alpha_8 \text{BLACK}_i \\ &+ \alpha_9 \text{EDUCDEC}_i + \alpha_{10} \text{NDISEASE}_i + \alpha_{11} \text{HLTHF}_i + \alpha_{12} \text{HLTHG}_i \\ &+ \alpha_{13} \text{HLTHP}_i + \alpha_{14} \text{AGE}_i + \alpha_{15} \text{CHILD}_{ij} + \alpha_{16} \text{LFAM}_i + \alpha_{17} \text{LC}_{ij} \\ &+ \alpha_{18} \text{FCHILD}_{ij} + b_{2i}, \end{aligned} \tag{5.2}$$

where $\boldsymbol{b}_i \sim N_2(\boldsymbol{0}, \boldsymbol{D})$. The prior distributions for the unknown parameters of the ZINB and ZIP random effects models are the same as those of the simulation study section and are given by:

$$\begin{aligned}&\beta_k \sim \gamma_k \delta_0(\beta_k) + (1-\gamma_k) N(0, \sigma^2_{\beta_k}), \quad \gamma_k \sim \text{Beta}(0.1, 0.1), \quad \sigma^2_{\beta_k} \sim \text{IG}(0.1, 0.1), \quad k = 1, ..., 19, \\ &\alpha_l \sim \nu_l \delta_0(\alpha_l) + (1-\nu_l) N(0, \sigma^2_{\alpha_l}), \quad \nu_l \sim \text{Beta}(0.1, 0.1), \quad \sigma^2_{\alpha_l} \sim \text{IG}(0.1, 0.1), \quad l = 1, ..., 18, \\ &\boldsymbol{D} \sim \text{IWishart}(2, \Psi), \quad \Psi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.\end{aligned}$$

In the Bayesian method, two parallel MCMC chains are run with different initial values for 40,000 iterations each. Then, we discarded the first 20,000 iterations as pre-convergence burn-in and retained 20,000 for the posterior inference. For checking the convergence of the MCMC chains, we have used the Gelman–Rubin diagnostic test. The results, including parameter estimates, standard deviations, 95% credible intervals, LBFDR, and Gelman–Rubin statistics for analyzing the data using ZINB and ZIP random effects models, are presented in Tables 3 and F.1. The negative binomial random effects model (NB) and the Poisson random effects model (P) are also applied to analyze the data. The prior distributions for the unknown parameters of the NB and P random effects models are given by:

$$\begin{aligned}&\beta_k \sim \gamma_k \delta_0(\beta_k) + (1-\gamma_k) N(0, \sigma^2_{\beta_k}), \quad \gamma_k \sim \text{Beta}(0.1, 0.1), \quad \sigma^2_{\beta_k} \sim \text{IG}(0.1, 0.1), \quad k = 1, ..., 19, \\ &b_i \sim N(0, \sigma^2_b), \quad \sigma^2_b \sim \text{IG}(0.1, 0.1), \quad i = 1, ..., n.\end{aligned}$$

Tables F.2 and F.3 show the results for the NB and P random effects models, respectively. Based on the values of DIC, the performance of the ZINB random effects model is better than that of the ZIP, P, and NB random effects models. Based on the results of Table 3, IDP, FMDE, LINC, FEMALE, PHYSLIM, BLACK, NDISEASE, HLTHF, HLTHP, CHILD, LFAM and FEMCHILD are selected for the rate model (they have LBFDR < 0.05), that is, these variables are significant predictors such that increasing in IDP leads to decreasing the medical doctor visit (MD) and LINC is positively significant meaning the more the natural logarithm of income (LINC), the more visits to a MD. FMDE is a negatively significant predictor, such that by increasing it, the probability of zero decreases. The greater the physical limitations (PHYSLIM), the larger the values of the estimated probability of nonzero. BLACK is positively significant, which means the number of visits to a medical doctor of black patients is higher than that of white patients, and increasing the NDISEASE leads to higher MD numbers. Also, HLTHF and HLTHP are factors that motivate patients to visit the doctor. CHILD is positively significant, which means the patient who is under 18 years old has more MD than others. The effect of FEMALE on MD number depends on the level of CHILD, and the effect of CHILD on MD number depends on the level of FEMALE. These estimates indicate females visit the doctor more than males. Also, LPI, EDUCDEC, HLTHG, AGE, and LC have a LBFDR > 0.05 which means they are not significant predictors.

Table 3: Parameter estimates (Est.), standard deviation (SD), 2.5% (lower bound of 95% credible interval), 97.5% (upper bound of 95% credible interval), local Bayesian false discovery rate (LBFDR), and Gelman–Rubin statistics ($\hat{R}$) for analyzing RAND data using the ZINB model.

Model of the rate (μ_{ij})						
	Est.	SD.	2.5%	97.5%	LBFDR	$\hat{R}$
Intercept (β_0)	0.824	0.044	0.738	0.908	0.000	1.013
time (β_1)	0.000	0.001	0.000	0.000	0.992	1.029
time^2 (β_2)	0.000	0.000	0.000	0.000	1.000	1.000
IDP (β_3)	−0.169	0.037	−0.241	−0.099	0.000	1.004
LPI (β_4)	0.001	0.005	0.000	0.011	0.968	1.016
FMDE (β_5)	−0.129	0.015	−0.157	−0.099	0.000	1.006
LINC (β_6)	0.107	0.019	0.070	0.145	0.000	1.005
FEMALE (β_7)	0.258	0.039	0.184	0.339	0.000	1.006
PHYSLM (β_8)	0.270	0.044	0.183	0.356	0.000	1.009
BLACK (β_9)	0.353	0.056	0.241	0.463	0.000	1.015
EDUCDEC (β_{10})	0.000	0.014	0.000	0.000	0.959	1.017
NDISEASE (β_{11})	0.175	0.015	0.146	0.207	0.000	1.016
HLTHF (β_{12})	0.206	0.058	0.087	0.315	0.011	1.101
HLTHG (β_{13})	0.000	0.004	0.000	0.000	0.986	1.004
HLTHP (β_{14})	0.426	0.107	0.211	0.626	0.001	1.000
AGE (β_{15})	0.000	0.000	0.000	0.000	0.999	1.002
CHILD (β_{16})	0.174	0.042	0.093	0.256	0.000	1.001
LFAM (β_{17})	−0.144	0.027	−0.199	−0.093	0.000	1.002
LC (β_{18})	−0.002	0.016	−0.026	0.000	0.959	1.106
FEMCHILD (β_{19})	0.217	0.058	0.105	0.330	0.000	1.006

Model of the probability (π_{ij})						
	Est.	SD.	2.5%	97.5%	LBFDR	$\hat{R}$
Intercept (α_0)	5.390	0.741	4.356	7.166	0.000	1.050
time (α_1)	0.002	0.016	0.000	0.032	0.948	1.062
IDP (α_2)	−1.026	0.341	−1.799	−0.416	0.003	1.074
LPI (α_3)	−0.663	0.160	−0.991	−0.357	0.000	1.022
FMDE (α_4)	−1.044	0.175	−1.395	−0.714	0.000	1.007
LINC (α_5)	0.151	0.133	0.000	0.383	0.357	1.015
FEMALE (α_6)	−2.475	0.445	−3.450	−1.662	0.000	1.018
PHYSLM (α_7)	−0.005	0.122	−0.334	0.265	0.857	1.014
BLACK (α_8)	−4.915	0.605	−6.255	−4.011	0.000	1.032
EDUCDEC (α_9)	−0.028	0.151	−0.512	0.159	0.825	1.035
NDISEASE (α_{10})	0.452	0.152	0.164	0.762	0.010	1.045
HLTHF (α_{11})	−0.039	0.165	−0.577	0.096	0.840	1.006
HLTHG (α_{12})	0.004	0.087	−0.165	0.246	0.869	1.054
HLTHP (α_{13})	0.178	0.459	−0.270	1.526	0.690	1.030
AGE (α_{14})	0.000	0.000	0.000	0.000	1.000	1.000
CHILD (α_{15})	−1.458	0.367	−2.226	−0.791	0.000	1.028
LFAM (α_{16})	0.005	0.064	0.000	0.153	0.932	1.044
LC (α_{17})	−0.118	0.296	−1.051	0.054	0.740	1.032
FEMCHILD (α_{18})	−2.995	0.628	−4.256	−1.807	0.000	1.015

	Est.	SD.	2.5%	97.5%	LBFDR	$\hat{R}$
D_{11}	0.493	0.228	0.226	1.007	—	1.010
D_{12} (D_{21})	0.925	0.128	0.516	1.803	—	1.101
D_{22}	1.021	0.020	0.561	2.907	—	1.002
ϕ	3.817	0.140	3.553	4.098	—	1.002
DIC	183127.5					

Time and time^2 have LBFDR > 0.05, i.e., time is not a significant predictor. Also, in the probability model, IDP, LPI, FMDE, FEMALE, BLACK, NDISEASE, CHILD, and FEMCHILD are significant predictors, such that increasing IDP leads to decreasing medical doctor visits, with increasing participation incentive payment (LPI) the probability of nonzeros decreases. FEMALE is positively significant, which means the probability of zero for men is the largest. BLACK is negatively significant, which means the white patient visits a MD less than the black patient. Also, increasing NDISEASE leads to a larger probability of nonzeros.

As mentioned before, the zero-inflated regression model assumes that the count numbers arise from a two-component mixture of a standard count distribution and a degenerated distribution at zero. Under such models, a zero can belong to either the degenerate state or the count distribution, but it is typically impossible to with certainty to determine to which state it belongs (Lambert, 1992). As two examples of the data, consider the 6th and 12th patients with $\boldsymbol{y}_6 = (1, 0, 0, 0, 0)$ and $\boldsymbol{y}_{12} = (1, 0, 7)$. The 6th patient is a white 16-year-old girl who has neither physical limitations nor chronic diseases. Also, the 12th patient is a black 61-year-old man who has no physical limitations but has chronic diseases. The other characteristics of these two patients are as follows:

	6th patient	12th patient
IDP	1	0
LPI	0.22	0.48
FMDE	1.16	1.29
LINC	0.40	0.52
EDUCDEC	8	18
HLTHF	0	0
HLTHG	0	1
HLTHP	0	0
LFAM	4	2
LC	0.45	0.52

The probability of being zero for the 6th patient at different time points can be estimated by considering ZINB as described in (5.1) and (5.2) and it is given by $\boldsymbol{\pi}_6 = (0.08, 0.64, 0.52, 0.72, 0.61)$; that is, for example, at the first time the probability of coming from a degenerated distribution is 0.08 while for the second time, it is 0.64. Also, this probability for the 12th patient is $\boldsymbol{\pi}_{12} = (0.09, 0.61, 0.02)$.

6. CONCLUSION

In this paper, we have discussed Bayesian variable selection methods for the zero-inflated power series distribution, specifically ZIP and ZINB random effects models that have been used via spike and slab priors for analyzing longitudinal count data with extra zeros.

We have evaluated the selection accuracy of DS and CS approaches through some simulation studies. Also, we have defined a test for checking the significance of the parameters, both a local Bayesian false discovery rate with a threshold of 0.05 and a Bayes factor procedure with a threshold of 1 are applied to perform this test. The other thresholds can also be considered to investigate H_0. The simulation studies show that applying DS has better

performance than applying CS. A real data set from the RAND health insurance experiment has been analyzed as an illustrative example. The proposed variable selection models by DS spike are applied to select the important variables in this data set, where the non-significant variables shrink to zero and those estimated are considered significant variables. To the best of our knowledge, ZINB and ZIP are the best models for analyzing zero-inflated count data, but if the range of the zero-inflated data is restricted to a special range such as $0, 1, ..., K$, the zero-inflated binomial model is a more appropriate model than the ZINB and ZIP models. The ZINB regression model allows for over-dispersion in the model and can be used to quantify various parameters more effectively. We have used DIC to select among different models, and we have concluded that the zero-inflated negative binomial random effects model is a flexible model to be assumed for analyzing this data.

As a future work, the proposed method can be applied to semi-parametric modeling data sets by considering spline. For this purpose, equation (2.3) can be improved to be

$$\begin{aligned} \log(\mu_{ij}) &= \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + g_1(t_{ij}) + b_{i1}, \\ \text{logit}(\pi_{\text{ij}}) &= \boldsymbol{z}'_{ij}\boldsymbol{\alpha}_j + g_2(t_{ij}) + b_{i2}\,, \qquad i = 1, ..., n, \quad j = 1, ..., T, \end{aligned}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are unknown smooth functions of time. For example, they can be considered as follows:

$$g_k(t_{ij}) = \alpha_{k0} + \alpha_{k1}t_{ij} + \cdots + \alpha_{kd_k}t_{ij}^{d_k} + \sum_{l=1}^{K_k} \alpha_{k,d_k+l}(t_{ij} - \kappa_i^l)_+^{d_k}, \quad k = 1, 2,$$

where d is the degree of the polynomial component, K_k is the number of interior knots, κ_i^l is referred to as knots of the i-th subject, $(a)_+ = \max(0, a)$, and $\boldsymbol{\alpha}_k = (\alpha_{k0}, \cdots, \alpha_{kd_k}, \alpha_{kd_k+1}, \cdots, \alpha_{kK_k})$ is the vector of spline coefficients. By considering these functions, all the approaches in this paper can be applied to this model, too.

As future work, we can consider marginalized zero-inflated negative binomial (MZINB) and marginalized zero-inflated Poisson (MZIP) models to model the population means count directly, allowing straightforward inference for overall exposure effects that account for both excess zeros and overdispersion (Preisser *et al.*, 2016). Also, the model can be extended to analyze data in the presence of missing values. For this purpose, a non-ignorable missing mechanism should be considered. Also, Bayesian variable selection by using global-local shrinkage (Hamura *et al.*, 2021) priors can be applied in future works. The Wishart-gamma and half-Cauchy priors can also be considered for the random effects covariance matrix and variance components, respectively.

REFERENCES

Alsalim, N. and Baghfalaki, T. (2021). Variable selection for longitudinal zero-inflated power series transition model. *Journal of Biopharmaceutical Statistics*, 31: 668–685.

Baghfalaki, T. and Ganjali, M. (2021). Approximate Bayesian inference for joint linear and partially linear modeling of longitudinal zero-inflated count and time to event data. *Statistical Methods in Medical Research*, 30, 1484–1501.

Baghfalaki, T., Sugier, P.E., Truong, T., Pettitt, A.N., Mengersen, K., and Liquet, B. (2021). Bayesian meta-analysis models for cross-cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, 15, 40(6):1498–1518.

Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:1281–1312.

Bekalo, D.B. and Kebede, D.T. (2021). Zero-inflated models for count data: an application to number of antenatal care service visits. *Ann. Data Sci.*, 1–26.

Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455.

Crainiceanu, C.M., Ruppert, D., and Wand, M.P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*, 14(14):1–24.

Deb, P. and Trivedi, P. (2002). The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics*, 21(4):601–625.

Diggle, P.J., Heagerty, P., Liang, K.Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.

DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A., Sun, C.L., Goltsman, D.S., Wong, R.J., Shaw, G., Stevenson, D.K. (2015). Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences*, 112(35):11060-5.

Efron, B. (2012). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press.

Feng, C.X. (2021). A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *J. Stat. Distrib. App.*, 8.

George, E.I. and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *J. Amer. Stat. Assoc.*, 88:881–889.

García-Donato, G., Castellanos, M.E., and Quirós (2021). A Bayesian variable selection with applications in health sciences. *Mathematics*, 9:218.

Gu, X., Tadesse, M.G., and Foulkes, A.S. (2020). Bayesian variable selection for high dimensional predictors and self-reported outcomes. *BMC Med. Inform. Decis. Mak.*, 20:212.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). *Bayesian Data Analysis*. CRC press.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:515–533.

Gelman, A. and Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472.

Gupta, P.L., Gupta, R.C., and Tripathi, R.C. (1996). Analysis of zero-adjusted count data. *Comput. Stat. Data Anal.*, 23:207–218.

Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, 56:1030–1039.

Hamura, Y., Irie, K., and Sugasawa, S. (2021). On global-local shrinkage priors for count data. *Bayesian Analysis*, 1(1):1–20.

Heilbron, D.C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36:347–531.

Hu, M.C., Pavlicova, M., and Nunes, E.V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse*, 37:367–375.

Ishwaran, H. and Rao, J.S. (2005). Spike and slab variable selection. Frequentist and Bayesian strategies. *Ann. Statist.*, 33:730–773.

Ji, Y. and Shi, H. (2020). Bayesian variable selection in linear quantile mixed models for longitudinal data with application to macular degeneration. *PLoS One*, 15(10).

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–13.

Lee, A.H., Wang, K., Scott, J.A., YauK, K., and McLachlan, G.J. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research*, 15(1):47–61.

Lee, K.H.A., Coull, B.A., Moscicki, A.B., Paster, B.J., and Starr, J.R. (2020). Bayesian variable selection for multivariate zero-inflated models: application to microbiome count data. *Biostatistics*, 21:499–517.

Luna, P.N., Mansbach, J.M., and Shaw, C.A. (2020). A joint modeling approach for longitudinal microbiome data improves ability to detect microbiome associations with disease. *PLOS Computational Biology*, 16(12):e1008473.

Mitchell, T.J. and Beauchamp, J.J. (1988). Bayesian variable selection in linear regression. *Amer. Stat. Assoc.*, 83, 1023–1036.

Miao, Y., Kook, J.H., Lu, Y., Guindani, M., and Vannucci, M. (2020). *Scalable Bayesian variable selection regression models for count data*. In "Flexible Bayesian Regression Modelling", Yanan F., Smith M., Nott D. and Dortet-Bernadet J.-L. (Eds.). Elsevier, 187–219.

Murat, M. and Szynal, D. (1998). Non-zero inflated modified power series distributions. *Communications in Statistics*, 12:3047–3064.

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modeling*, 5:1–19.

Neelon, B.H., OMalley, A.J., and Normand, S.L. (2010). A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling*, 10:421–439.

Patil, M.K. and Shirke, D.T. (2007). Testing parameter of the power series distribution of a zero inflated power series model. *Stat. Methodol.*, 16:393–406.

Preisser, J.S., Das, K., Long, D.L., and Divaris, K. (2016). Marginalized zero-inflated negative binomial regression with application to dental caries, *Statistics in Medicine*, 35(10):1722–1735.

Rose, C.E., Martin, S.W., Wannemuehler, K.A., and Plikaytis, B.D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, 16:463–481.

Song, P.X.K. (2007). *Correlated Data Analysis*. Springer-Verlag, New York.

Tokuda, T., Goodrich, B., Van Mechelen, I., and Gelman, A. (2011). *Visualizing distributions of covariance matrices*. Columbia University, New York, USA, Technical Report, 18–18.

Xu, X. and Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Anal.*, 10(4):909–936.

Young, D.S., Roemmele, E.S., and Yeh, P. (2021a). Zero-inflated modeling part I: traditional zero-inflated count regression models, their applications, and computational tools. *WIREs Computational Statistics*.

Young, D.S., Roemmele, E.S., and Shi, X. (2021b). Zero-inflated modeling part II: zero-inflated models for complex data structures. *WIREs Computational Statistics*.

Zeng, P., Wei, Y., and Zhao, Y. (2014). Variable selection approach for zero-inflated count data via adaptive lasso. *Journal of Applied Statistics*, 4:879–894.

Zhang, X., Pei, Y.F., Zhang, L., Guo, B., Pendegraft, A.H., Zhuang, W., and Yi, N. (2018). Negative binomial mixed models for analyzing longitudinal microbiome data. *Frontiers in Microbiology*, 26:9-1683.

Matthews, B.W. (1975). Comparison of the predicted and observed secondary structure of T4 phagelysozyme. *Biochimica et Biophysica Acta (BBA) – Protein Structure*, 405(2):442–451.

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 463–489
https://doi.org/10.57805/revstat.v23i4.558

Extended Easily Changeable Kurtosis Distribution

Author: Piotr Sulewski
– Institute of Exact and Technical Sciences, Pomeranian University, Poland
piotr.sulewski@upsl.edu.pl

Received: December 2022 Revised: December 2023 Accepted: December 2023

Abstract:

- This paper is the next step ahead in constructing probability distribution of changeable flatness of PDF that is expressed with well-known kurtosis measure. The distribution in question is named the Extended Easily Changeable Kurtosis and descends from the Easily Changeable Kurtosis. The paper covers PDF, CDF, modes and inflection points, quantiles, moments and Moors' measure and the Fisher Information Matrix. In addition generator of pseudo-random numbers that follow the Extended Easily Changeable Kurtosis is presented. Unknown parameters of the distribution are estimated with the maximum likelihood method. The paper ends with illustrative examples of applicability and flexibility of the new distribution. The most important R codes are presented in the supplementary material.

Keywords:

- *normal distribution; modeling kurtosis; departure from normality.*

AMS Subject Classification:

- 60E05, 65C20.

1. INTRODUCTION

The statistics literature is filled with hundreds of continuous univariate distributions. However, in recent years, applications from the environmental, financial, biomedical sciences, engineering among others, have further shown that data sets following the classical distributions are more often the exception rather than the reality. Since there is a clear need for extended forms of these distributions a significant progress has been made toward the generalization of some well-known distributions and their successful application to problems in areas such as engineering, finance, economics and biomedical sciences, among others (Ashour and Eltehiwy, 2013).

The article presents a symmetric distribution with two shape parameters $p > -1$ and $q > 0$ called the extended easily changeable kurtosis (EECK) distribution. As the name suggests, the EECK($p > -1, q > 0$) is an extended version of the easily changeable kurtosis (ECK) distribution with scale and shape parameters $a > 0$ and $p > -1$, respectively (Sulewski, 2022b). Instead of kurtosis γ_2, the article analyzes the excess kurtosis $\overline{\gamma}_2 = \gamma_2 - 3$, which can be positive or negative.

Symmetric distributions do not form such a big family as asymmetric distributions. Table 1 presents (in alphabetical order) thirty four symmetric distributions with the range of excess kurtosis $\overline{\gamma}_2$ and modality. Symmetric distributions with undefined excess kurtosis are: Cauchy (Kotz *et al.*, 2004), degenerate (Glen, 2025) and Voigt (Temme, 2010). Symmetric distributions with constant excess kurtosis are: arcsine (Lévy, 1940), bimodal normal (Hassan and Hijazi, 2010), bimodal Laplace (Hassan and Hijazi, 2010), cosine (Raab and Green, 1961), hyperbolic secant (Johnson *et al.*, 1995), Laplace (Johnson *et al.*, 1995), logistic (Balakrishnan, 1992), normal (Johnson *et al.*, 1995), raised cosine (Rinne, 2010), sine (Edwards, 2000), semicircle (Ryan, 2014), uniform (Dekking *et al.*, 2005), U-shaped (Bucher, 2012). Symmetric distributions with excess kurtosis in an finite interval: Bates (Johnson *et al.*, 1995), bimodal exponential power (Hassan and Hijazi, 2010), bimodal power normal (Bolfarine *et al.*, 2018), ECK (Sulewski, 2022b), extended normal (Ki *et al.*, 2005), extended Laplace (Johnson *et al.*, 1995), extended t (Johnson *et al.*, 1995), Irwin–Hall (Johnson *et al.*, 1995), plasticizing component (Sulewski, 2022a), Q-gaussian (Umarov *et al.*, 2008), t (Johnson *et al.*, 1995), Tukey with finite domain (Freimer *et al.*, 1988), U-power (Bucher, 2012), Von Mises (Mardia *et al.*, 2000). Symmetric distributions with excess kurtosis in an infinite interval are: generalized normal (Nadarajah, 2005), normal-exponential-gamma (Johnson *et al.*, 1995), Tukey with infinite domain (Freimer *et al.*, 1988), U-quadratic (Buchanan and Wheeland, 2022).

The ECK($a > 0, p > -1$) (Sulewski, 2022b), as the previous version of the EECK($p > -1$, $q > 0$), is unimodal distribution and can be used to model excess kurtosis in the range $(-2, 0)$.

The main goal of the paper is to define the distribution for excess kurtosis modeling in a larger range than $(-2, 0)$. As follows from the Malachov inequality $\overline{\gamma}_2 \geqslant \gamma_1^2 - 2$ (Malachov, 1978), the best range would be the maximum range, i.e. $\langle -2, \infty)$.

The proposed distribution with $\overline{\gamma}_2^{\mathrm{EECK}} \geq -2$ (see Subsection 2.4), like the generalized normal (GN) with $\overline{\gamma}_2^{\mathrm{GN}} \geq -1.2$, normal-exponential-gamma (NEG) with $\overline{\gamma}_2^{\mathrm{NEG}} > 0$ and Tukey (T) defined in an infinite domain with $\overline{\gamma}_2^{\mathrm{T}} > 0$, belongs to the family of symmetrical, unimodal distributions with excess kurtosis values on infinite interval (see Table 1).

In addition, the EECK is defined in the finite domain whereas the NEG, GN and T are defined in an infinite domain.

Table 1: Symmetric distributions with range of excess kurtosis and modality.

Name	Modes	Range of $\overline{\gamma}_2$	Name	Modes	Range of $\overline{\gamma}_2$
arcsine	2	-1.5	logistic	1	$6/5$
Bates	1	$[-1.2, 0)$	normal	1	0
bimodal exponential power	$1, 2$	$[-3, 3]$	normal-exponential-gamma	1	$(0, \infty)$
bimodal normal	2	$-4/3$	plasticizing component	2	$(-2, 0)$
bimodal Laplace	2	$1/3$	Q-gaussian	1	$[-0.857, 0]$
bimodal power normal	$1, 2$	$(-2, 0)$ $\vee (0, 10.97)$	raised cosine	1	$\frac{6\,(90-\pi^4)}{5\,(\pi^2-6)^2}$
Cauchy	1	—	sine	1	$\frac{2\,(96-\pi^4)}{(\pi^2-8)^2}$
cosine	1	0.251	semicircle	1	-1
degenerate	1	—	t	1	$(0, 6]$
ECK	1	$(-2, 0)$	Tukey*	1	$(0, \infty)$
extended normal	$1, 2$	$[-4/3, 0]$	Tukey**	1	$[-1.25, 10.59]$
extended Laplace	$1, 2$	$(1/3, 3]$	uniform	∞	$-6/5$
extended t	$1, 2$	$[-4/3, 6]$	U-power	2	$[-2, -1.81]$
generalized normal	1	$[-1.2, 0)$ $\vee (0, \infty)$	U-quadratic	2	$(0, \infty)$
hyperbolic secant	1	2	U-shaped	2	-1.5
Irwin–Hall	1	$[-1.2, 0)$	Voigt	1	—
Laplace	1	3	Von Mises	1	$[-1.2, 1.069]$

* infinite domain; ** finite domain

PDF of the NEG, as a mixture of normal distributions, has a complicated form and the analytical formula for excess kurtosis does not exist. PDF of the T distribution has a simple, closed form for a few exceptional values of the shape parameter, e.g. we get, respectively, for $\lambda = \{1, 0\}$ uniform and logistic distributions.

The analytical formulas for excess kurtosis of the EECK, GN and T distributions are respectively:

$$\overline{\gamma}_2^{\mathrm{EECK}} = \frac{\Gamma^2\left(p+\frac{3}{q}\right)\Gamma\left(\frac{1}{q}\right)\Gamma\left(\frac{5}{q}\right)(pq+3)^2}{\Gamma^2\left(\frac{3}{q}\right)\Gamma\left(p+\frac{1}{q}\right)\Gamma\left(p+\frac{5}{q}\right)(pq+1)(pq+5)} - 3 \qquad (p>-1,\ q>0), \tag{1.1}$$

$$\overline{\gamma}_2^{\mathrm{GN}} = \frac{\Gamma\left(\frac{5}{\beta}\right)\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{3}{\beta}\right)^2} - 3 \qquad (\beta>0), \tag{1.2}$$

$$\overline{\gamma}_2^{\mathrm{T}} = \frac{(2\lambda+1)^2\,\Gamma(2\lambda+1)^2\left[3\Gamma(2\lambda+1)^2+\Gamma(4\lambda+4)-4\Gamma(3\lambda+1)\,\Gamma(\lambda+1)\right]}{(8\lambda+1)\,\Gamma(4\lambda+1)\left[\Gamma(\lambda+1)^2-\Gamma(2\lambda+1)\right]^2} - 3 \qquad (\lambda>-0.25). \tag{1.3}$$

The proof of (1.1) is presented in Subsection 2.4 (see Theorem 2.7).

Figure 1 shows the excess kurtosis of the EECK, GN and T distributions as a function of the shape parameters $p>-1$, $\beta>0$ and $\lambda\in(-0.25,0)$. The $\overline{\gamma}_2^{\mathrm{EECK}}(p)$ is an increasing function similar to a linear function while $\overline{\gamma}_2^{\mathrm{GN}}(\beta)$ and $\overline{\gamma}_2^{\mathrm{T}}(\lambda)$ are initially decreasing strongly functions and then transforming into constant functions. This is especially visible for the GN distribution. The EECK and GN distributions can be used to model the negative and positive excess kurtosis. The negative values of excess kurtosis for the EECK and GN distributions are available on $[-2,0]$ and $[-1.2,0)$, respectively.

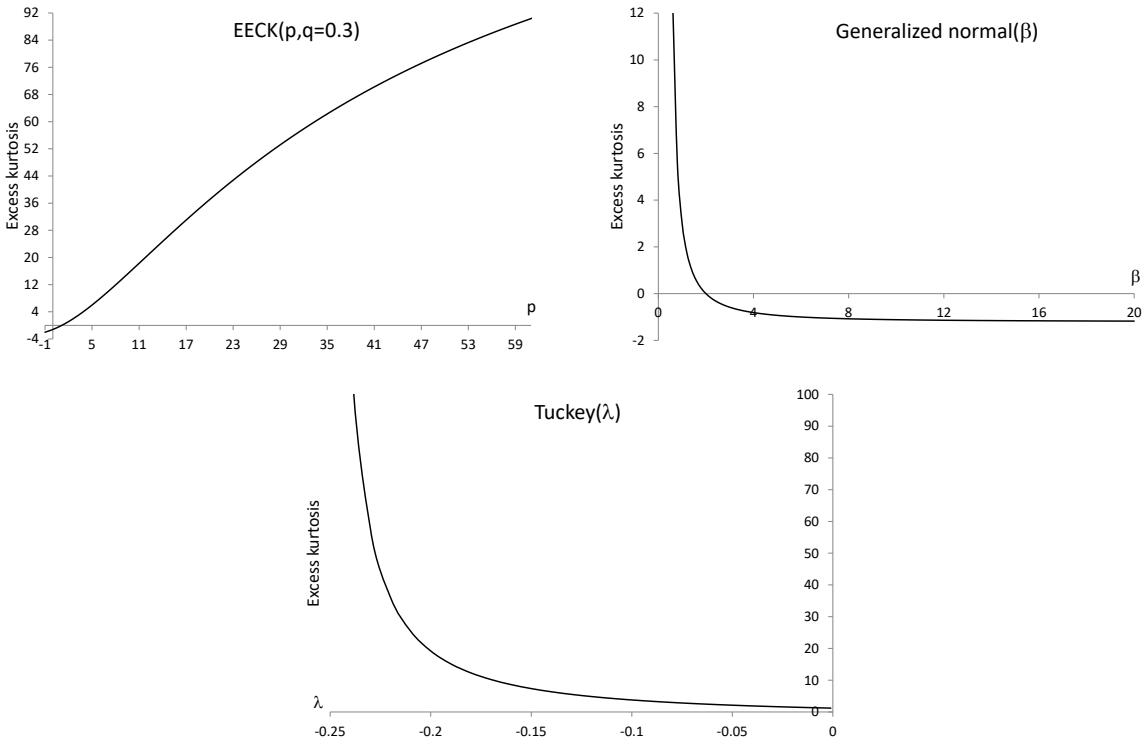

Figure 1: Excess kurtosis as a function of shape parameter.

Formula (1.3) is the most complicated among formulas (1.1)–(1.3), however, for the EECK, GN, and T distributions, the shape parameter cannot be represented as a function of $\overline{\gamma}_2$, as is for the ECK and Q-gaussian distributions. This is the price for expanding the range of $\overline{\gamma}_2$. Nowadays, in the era of advanced mathematical software, it is possible to compute the argument of a function knowing its value (using for example Mathcad or Microsoft Excel in newer versions).

Summarizing, the new proposal can be extremely useful when you want to seamlessly test the goodness-of-fit tests (GoFTs) ability to detect deviations from normality caused by the maximum range of excess kurtosis values, i.e. negative and positive. Real data example (see Section 4.2) demonstrates that the EECK$(p>-1, q>0)$ distribution in the mixed variant is flexible and competitive model that deserves to be added to the existing distributions in data modeling.

Special cases of the EECK$(p>-1, q>0)$ distribution are: the uniform, triangle and obviously ECK$(a>0, p>-1)$. The EECK$(p>-1, q>0)$ tends to the normal distribution (see Subsection 2.1).

It should also be mentioned that there is a group of asymmetric distributions, which are symmetrical for certain parameter values, e.q. the truncated normal, Birnbaum–Saunders (Birnbaum and Saunders, 1969), skew-normal (Azzalini, 1985), beta, two-piece normal (Gibbons and Mylroie, 1985), two-piece power normal (Sulewski, 2021) and plasticizing component (Sulewski, 2022a).

This article is organized as follows. Section 2 presents the main properties of the EECK distribution such as PDF, CDF, modes, inflection points, quantiles, moments, Moors' measure, instructions to generate EECK pseudo-random numbers and the Fisher Information Matrix. The estimation procedures are provided in Section 3. The articles ends with applications and conclusions. The most important `R` codes are given in the supplementary material.

2. MAIN PROPERTIES OF INTRODUCED DISTRIBUTION

2.1. Distribution and density functions

Definition 2.1. The Eta function for $p>-1$ and $q>0$ is defined as

$$H(p,q) = \int_{-1}^{1} \left[1-|x|^q\right]^p dx = \frac{2B\left(\frac{1}{q}, p+1\right)}{q} = \frac{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}{\Gamma\left(p+\frac{1}{q}+1\right)}, \tag{2.1}$$

where $B(u,v)$ is the beta function.

Calculations in (2.1) were performed by the formula (Gradshteyn and Ryzhik, 2014)

$$\int_0^1 x^{a-1}\left(1-x^b\right)^{c-1} dx = \frac{B\left(\frac{a}{b}, c\right)}{b}. \tag{2.2}$$

Exemplary values of the Eta function (2.1):

$$H(1,1)=1, \quad H(0,1)=2, \quad H(-0.5,1)=4, \quad H(1,0.5)=\frac{2}{3}, \quad H(0.5,1)=\frac{4}{3}.$$

Definition 2.2. The distribution of the random variable X with PDF given by

$$f(x;p,q)=\frac{\left[1-|x|^q\right]^p}{H(p,q)}, \qquad x\in\begin{cases}(-1,1) & \text{if } -1<p<0,\\ [-1,1] & \text{if } p\geq 0,\end{cases} \tag{2.3}$$

is called the extended easily changeable kurtosis (EECK) distribution, where $p>-1$ and $q>0$ are the shape parameters. The EECK$(p>-1, q>0)$ is symmetric around zero, since, based on (1.3), $f(x;p,q)=f(-x;p,q)$ (see Figure 2). The EECK$(p>-1, q=2)$ is the ECK$(a=1, p>-1)$ (Sulewski, 2022b).

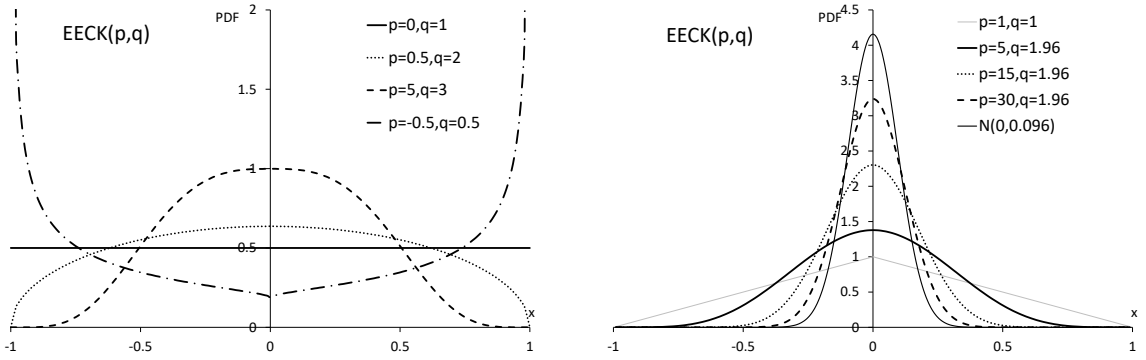

Figure 2: PDF of the EECK(p,q) distribution for various parameter values.

The `R` codes of the `dEECK` function for computing PDF are provided in the supplementary material.

The standard deviation of the new proposal, based on (2.17), equals

$$\mu_2 = \frac{(1+pq)\,\Gamma\left(\frac{3}{q}\right)\,\Gamma\left(p+\frac{1}{q}\right)}{(3+pq)\,\Gamma\left(\frac{1}{q}\right)\,\Gamma\left(p+\frac{3}{q}\right)},$$

therefore the EECK(p,q) distribution tends to the normal distribution $\mathrm{N}\left(0,\sqrt{\mu_2}\right)$ with PDF $\phi\left(x;0,\sqrt{\mu_2}\right)$.

Let M (2.4) be the similarity measure of these distributions (Sulewski, 2020). We have for $p>-1$, $q>0$,

$$M(p,q) = \int_{-1}^{1} \min\left\{ f(x;p,q),\ \phi\left[x;\ 0,\ \sqrt{\frac{(1+pq)\,\Gamma\left(\frac{3}{q}\right)\,\Gamma\left(p+\frac{1}{q}\right)}{(3+pq)\,\Gamma\left(\frac{1}{q}\right)\,\Gamma\left(p+\frac{3}{q}\right)}}\right]\right\} dx. \tag{2.4}$$

The similarity measure M takes values on (0,1) and if PDFs are identical then $M = 1$. For example $M(33,1) = 0.871$, $M(33,1.5) = 0.954$, $M(33,2) = 0.995$, $M(33,2.5) = 0.961$. A more detailed analysis of the value of the M measure showed that it has the highest values for $q = 1.96$. We have $M(50,1.96) = 0.999$.

The EECK$(p > -1, q > 0)$ is the symmetrical distribution (Figure 2). The EECK$(p = 0, q > 0)$ is the uniform distribution U$(-1,1)$ (Figure 2, series $p = 0$, $q = 1$. The EECK$(p > 0, q > 0)$ is unimodal with mode equals 0 (Figure 2, series $p = 0.5$, $q = 2$; $p = 5$, $q = 3$). The EECK$(-1 < p < 0, q > 0)$ is pseudo $(-1 < x < 1)$ bimodal with bathtub shape (Figure 2, series $p = -0.5$, $q = 0.5$). The EECK$(p = 1, q = 1)$ is the triangle distribution (Figure 2, series $p = 1$, $q = 1$). The EECK$(50, 1.96)$ is in 99.9% the normal distribution N$(0, 0.096)$ (Figure 2, series N$(0, 0.096)$).

Theorem 2.1. *If* $X \sim$ EECK$(p > -1, q > 0)$ *with PDF* $f(x; p, q)$ (2.3) *then CDF of* X *is given by*

$$F(x; p, q) \;=\; 0.5 + x\,\frac{{}_2F_1\left(-p, \frac{1}{q}, 1 + \frac{1}{q}, |x|^q\right)}{H(p,q)}, \tag{2.5}$$

where ${}_2F_1(a, b, c, x)$ *is the Gaussian hypergeometric function.*

Proof: From (2.3) we have

$$\begin{aligned} F(x; p, q) \;&=\; \frac{1}{H(p,q)} \int_{-1}^{x} (1 - |x|^q)^p \, dx \\ &=\; \frac{1}{H(p,q)} \left[\int_{-1}^{0} (1 - |x|^q)^p \, dx \;+ \int_{0}^{x} (1 - |x|^q)^p \, dx \right]. \end{aligned} \tag{2.6}$$

To complete the proof, we need to calculate two integrals. The first one, based on (2.1), has the form

$$\int_{-1}^{0} (1 - |x|^q)^p \, dx \;=\; 0.5\, H(p,q). \tag{2.7}$$

The second one can be written using a power series (Gradshteyn and Ryzhik, 2014)

$$\begin{aligned} \int_{0}^{x} (1 - |x|^q)^p \, dx \;&=\; x \sum_{k=0}^{\infty} \frac{(-p)_k \left(\frac{1}{q}\right)_k}{\left(1 + \frac{1}{q}\right)_k} \frac{|x|^{qk}}{k!} \\ &=\; {}_2F_1\left(-p, \frac{1}{q}, 1 + \frac{1}{q}, |x|^q\right) x, \end{aligned} \tag{2.8}$$

where ${}_2F_1(a, b, c, x)$ is the Gaussian hypergeometric function and $(x)_n$ is the Pochhammer symbol

$$(x)_n = \frac{\Gamma(x + n)}{\Gamma(x)} \;=\; x\,(x + 1) \cdots (x + n - 1).$$

Substituting (2.7) and (2.8) to (2.6) we obtain (2.5). □

The `R` codes of the `pEECK` function for computing CDF are provided in the supplementary material.

Figure 3 plots CDF of the EECK($p>-1, q>0$) distribution for some values of parameters. For $p=0$ we obtain the straight line (uniform distribution). For $p>0$ CDF is convex in $[-1,0)$ and is concave in $(0,1]$. For $-1<p<0$ CDF is concave in $(-1,0)$ and is convex in $(0,1)$. CDFs of the EECK$(50, 1.96)$ distribution and N$(0, 0.096)$ one coincide.

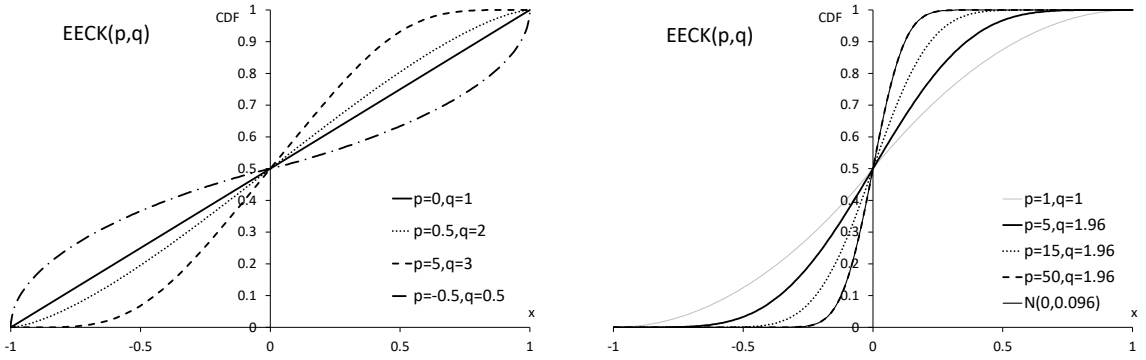

Figure 3: CDF of the EECK(a,p) distribution for various parameter values.

Theorem 2.2. *The* EECK$(p>-1, q>0)$ *distribution with PDF given by* (2.3) *is identifiable in the parameter space* $v=(p,q)$.

Proof: Let $v_1=(p_1,q_1)$ and $v_2=(p_2,q_2)$. Let us suppose that $f_{v_1}(x)=f_{v_2}(x)$ for all x from support. This condition based on (2.1) and (2.3) implies that

$$\frac{q_1(1-|x|^{q_1})^{p_1}}{2B\left(\frac{1}{q_1},p_1+1\right)} = \frac{q_2(1-|x|^{q_2})^{p_2}}{2B\left(\frac{1}{q_2},p_2+1\right)}. \tag{2.9}$$

If we apply log to both sides of (2.9) we obtain the following system of three equations:

$$\log\left(\frac{q_1}{q_2}\right)=0, \quad p_1\log(1-|x|^{q_1})-p_2\log(1-|x|^{q_2})=0, \quad \log\left[\frac{B\left(\frac{1}{q_2},p_2+1\right)}{B\left(\frac{1}{q_1},p_1+1\right)}\right]=0. \tag{2.10}$$

From the first equation is $q_1=q_2$ and then from the second one is $p_1=p_2$. □

2.2. Modes and inflection points

Theorem 2.3. *Let* $X\sim$ EECK$(p>-1, q>0)$. *If* $p=0$ *then modal values* $x_m\in[-1,1]$ *(case of uniform distribution). If* $p>0$ *then* $x_m=0$. *If* $-1<p<0$ *then the* EECK(p,q) *distribution is pseudo bimodal with modes* $x_m(-1)$, $x_m(1)$. *The* $f(x;p>0,q)$ (2.3) *is monotonically increasing on the interval* $(-1,0)$ *and monotonically decreasing on the interval* $(0,1)$. *The* $f(x;-1<p<0,q)$ (2.3) *is monotonically decreasing on the interval* $(-1,0)$ *and monotonically increasing on the interval* $(0,1)$.

Proof: Let $p \geq 0$ then PDF of the EECK(p,q) distribution, based on (2.1) and (2.3), for any $x \in [-1,1]$ is given by

$$f(x;p,q) = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,(1-|x|^q)^p. \tag{2.11}$$

Let $p=0$ then $f(x;0,q) = \frac{\Gamma\left(\frac{1}{q}+1\right)}{2\Gamma\left(\frac{1}{q}+1\right)} = 0.5$ is constant in $[-1,1]$.

Let $p>0$ then

$$\frac{d}{dx}f(x;p,q) = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,p\,(1-|x|^q)^{p-1}\left[-q|x|^{q-1}\right]. \tag{2.12}$$

As a result of simple transformations $x_m = 0$ and (2.12) is positive on the interval $(-1,0)$ and negative on the interval $(0,1)$.

Let $-1<p<0$ then PDF (2.11) is defined for any $x \in (-1,1)$. As a result of simple transformations, (2.12) is negative on the interval $(-1,0)$ and positive on the the interval $(0,1)$. For x values very close to $-a$ and a PDF (2.8) has locally maximum values. The author of this article denotes these values as $x_m(-1)$, $x_m(1)$ and proposed distribution defines as pseudo bimodal with modes at these points. □

Theorem 2.4. *Let $X \sim \text{ECK}(p>-1, q>0)$. The inflection points of the $f(x;p,q)$ (6) for $p>1 \wedge q>1$ or $-1<p<1 \wedge 0<q<1$ are given by means of the following formulas:*

$$x_1 = -\left(\frac{1-q}{1-pq}\right)^{\frac{1}{q}}, \qquad x_2 = \left(\frac{1-q}{1-pq}\right)^{\frac{1}{q}}. \tag{2.13}$$

Proof: We can write (2.12) as

$$\frac{d}{dx}f(x;p,q) = \frac{-pq\,\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,|x|^{q-1}(1-|x|^q)^{p-1}. \tag{2.14}$$

Let $A = \frac{-pq\,\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}$ then (2.14) has the simpler form

$$\frac{d}{dx}f(x;p,q) = A|x|^{q-1}(1-|x|^q)^{p-1}.$$

The second derivative is given by

$$\begin{aligned}\frac{d^2}{dx^2}f(x;p,q) &= A\left\{(q-1)\,|x|^{q-2}(1-|x|^q)^{p-1} - q|x|^{q-1}(p-1)\,(1-|x|^q)^{p-2}|x|^{q-1}\right\}\\ &= A|x|^{q-2}(1-|x|^q)^{p-2}\left\{(q-1)\,(1-|x|^q) - q|x|\,(p-1)\,|x|^{q-1}\right\}.\end{aligned}$$

Thus

$$\frac{d^2}{dx^2}f(x;p,q) = 0 \quad \Leftrightarrow \quad (q-1)\,(1-|x|^q) - q|x|\,(p-1)\,|x|^{q-1} = 0.$$

As a result of simple transformations we have

$$x_1 = -\left(\frac{1-q}{1-pq}\right)^{\frac{1}{q}} \wedge\ x_1 > -1\,, \qquad x_2 = \left(\frac{1-q}{1-pq}\right)^{\frac{1}{q}} \wedge\ x_2 < 1, \tag{2.15}$$

then from (2.15) we obtain $p > 1 \wedge q > 1$ or $-1 < p < 1 \wedge 0 < q < 1$. □

2.3. Quantiles

Theorem 2.5. *Let* $X \sim \text{EECK}(p > -1, q > 0)$. *The* u*-th* $(0 < u < 1)$ *quantile* x_u *is the solution of the following equation*

$$(0.5 - u)\, H(p,q) + {}_2F_1\left(-p, \frac{1}{q}, 1 + \frac{1}{q}, |x_u|^q\right) x_u \;=\; 0, \tag{2.16}$$

where ${}_2F_1(a,b,c,x)$ *is the Gaussian hypergeometric function and* $H(p,q)$ *is given by* (2.1). *The proposed distribution is symmetrical then* $x_u = -x_{1-u}$, *obviously and* $x_{0.5} = 0$.

Proof: Obtaining (2.16), based on the quantile definition, is trivial. □

The quantile x_u can be computed by numerical methods. The `R` codes of the `qEECK` function for computing the quantile x_u are provided in the supplementary material.

2.4. Moments and Moors' measure

Theorem 2.6. *The* k*-th* $(k = 0, 1, 2, ...)$ *non-central moments of the* $\text{EECK}(p > -1, q > 0)$ *distribution are given by*

$$\alpha_k \;=\; \frac{\left[1 + (-1)^k\right] B\left(\frac{k+1}{q}, p+1\right)}{q\, H(p,q)} \;=\; \frac{\left[1 + (-1)^k\right] B\left(\frac{k+1}{q}, p+1\right)}{2B\left(\frac{1}{q}, p+1\right)}. \tag{2.17}$$

Proof: The k-th $(k = 0, 1, 2, ...)$ non-central moments, based on (2.1) and (2.3), are defined as

$$\alpha_k \;=\; \frac{q}{2B\left(\frac{1}{q}, p+1\right)} \left[\int_{-1}^{0} x^k (1 - |x|^q)^p\, dx \;+\; \int_{0}^{1} x^k (1 - |x|^q)^p\, dx\right] \;=\; \frac{q(I_1 + I_2)}{2B\left(\frac{1}{q}, p+1\right)}. \tag{2.18}$$

To solve the integrals I_1 and I_2, we have to use the integral formula (2.2). Thus:

$$I_1 = (-1)^k\, \frac{B\left(\frac{k+1}{q}, p+1\right)}{q}\,, \qquad I_2 \;=\; \frac{B\left(\frac{k+1}{q}, p+1\right)}{q} \tag{2.19}$$

and substituting obtained results into (2.18) we get (2.17). □

Theorem 2.7. *The non-central moments* α_k $(k=1,3,...)$*, variance* μ_2 *and excess kurtosis* $\bar{\gamma}_2$ *of the* EECK$(p>-1, q>0)$ *distribution are given by*

$$\alpha_k = 0 \quad (k=1,3,...)\,, \qquad \mu_2 = \frac{(1+pq)\,\Gamma\left(\frac{3}{q}\right)\Gamma\left(p+\frac{1}{q}\right)}{(3+pq)\,\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{3}{q}\right)}, \tag{2.20}$$

$$\bar{\gamma}_2 = \frac{(pq+3)^2\,\Gamma\left(\frac{1}{q}\right)\Gamma\left(\frac{5}{q}\right)\Gamma\left(p+\frac{3}{q}\right)^2}{(pq+1)\,(pq+5)\,\Gamma\left(p+\frac{1}{q}\right)\Gamma\left(p+\frac{5}{q}\right)\Gamma\left(\frac{3}{q}\right)^2} - 3. \tag{2.21}$$

Proof: The proof $\alpha_k = 0$ $(k=1,3,...)$, based on (2.17), is trivial.

The first non-central moment equals zero, so the non-central moments α_k $(k=0,1,...)$ are equal to the central moments μ_k $(k=0,1,...)$.

From (2.17), using the properties of the gamma function $\Gamma(x+1) = x\,\Gamma(x)$, we have

$$\alpha_2 = \mu_2 = \frac{B\left(\frac{3}{q}, p+1\right)}{B\left(\frac{1}{q}, p+1\right)} = \frac{\Gamma\left(\frac{3}{q}\right)\Gamma\left(p+\frac{1}{q}+1\right)}{\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{3}{q}+1\right)} = \frac{(1+pq)\,\Gamma\left(\frac{3}{q}\right)\Gamma\left(p+\frac{1}{q}\right)}{(3+pq)\,\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{3}{q}\right)}, \tag{2.22}$$

$$\alpha_4 = \mu_4 = \frac{B\left(\frac{5}{q}, p+1\right)}{B\left(\frac{1}{q}, p+1\right)} = \frac{\Gamma\left(\frac{5}{q}\right)\Gamma\left(p+\frac{1}{q}+1\right)}{\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{5}{q}+1\right)} = \frac{(1+pq)\,\Gamma\left(\frac{5}{q}\right)\Gamma\left(p+\frac{1}{q}\right)}{(5+pq)\,\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{5}{q}\right)}. \tag{2.23}$$

Thus the excess kurtosis is given by

$$\bar{\gamma}_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{(1+pq)\,\Gamma\left(\frac{5}{q}\right)\Gamma\left(p+\frac{1}{q}\right)}{(5+pq)\,\Gamma\left(\frac{1}{q}\right)\Gamma\left(p+\frac{5}{q}\right)} \frac{(3+pq)^2\,\Gamma\left(\frac{1}{q}\right)^2\Gamma\left(p+\frac{3}{q}\right)^2}{(1+pq)^2\,\Gamma\left(\frac{3}{q}\right)^2\Gamma\left(p+\frac{1}{q}\right)^2} - 3 \tag{2.24}$$

and we obtain (2.21) as a result of simple transformation. □

Figure 4 shows the excess kurtosis $\bar{\gamma}_2$ as a function of the shape parameter p for $q = 0.4, 0.6, 0.8, 1$ (left) and for $q = 2, 4, 6, 8$ (right). The excess kurtosis, according to the definition, varies in the range $[-2, \infty)$. The smaller q value, the higher excess kurtosis and the parameter p has a greater effect on the excess kurtosis.

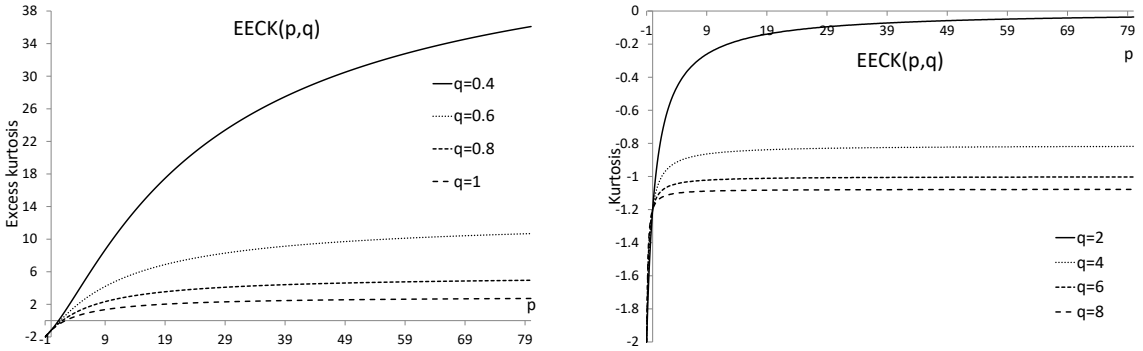

Figure 4: Excess kurtosis $\bar{\gamma}_2$ as a function of the shape parameter p.

Figure 5 shows the excess kurtosis $\bar{\gamma}_2$ as a function of the shape parameter q for $p = -0.9, -0.7, -0.5, -0.3$ (left) and for $p = 0.25, 0.75, 1, 10$ (right). For $p \in (-1, 0)$ the excess kurtosis tends from -2 to -1.2 when $q \to \infty$. For $p > 0$ kurtosis tends from ∞ to -1.2 when $q \to \infty$.

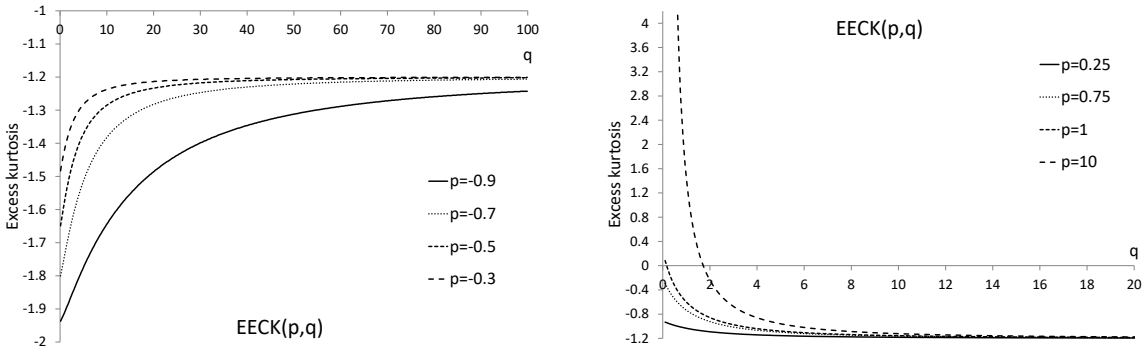

Figure 5: Excess kurtosis $\bar{\gamma}_2$ as a function of the shape parameter q.

Moors (1988) proposed a measure based on quantiles in the form

$$T = \frac{x_{7/8} - x_{5/8} + x_{3/8} - x_{1/8}}{x_{6/8} - x_{2/8}}, \tag{2.25}$$

where x_u is the solution of (2.16). The measure T is a quantile alternative for kurtosis and exists even for distribution for which no moments exist. Figure 6 shows the measure T as a function of the shape parameter p for $q = 0.75, 1, 2, 4$ (left) and as a function of the shape parameter q for $p = 0.75, 1, 2, 4$ (right). The $T(p)$ function decreases for $p(-1, 0)$ and increases for $p > 0$ mainly for its initial values. The $T(q)$ function tends to one.

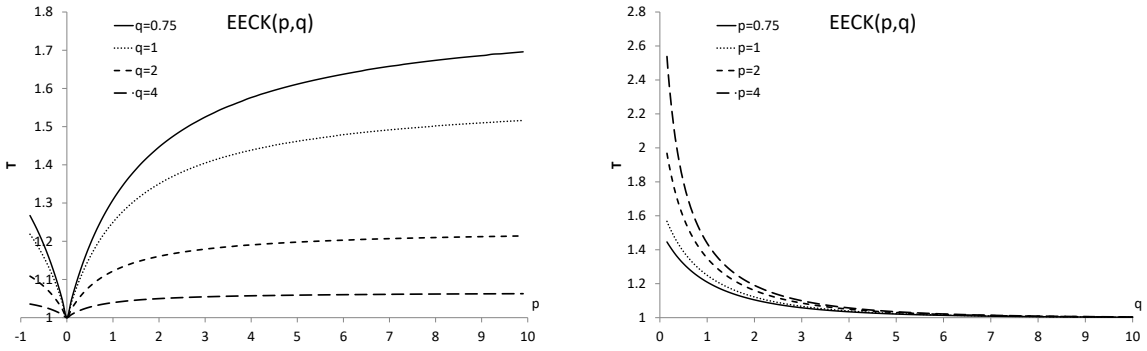

Figure 6: Moors' measure T as a function of the shape parameter p (left) and q (right).

2.5. Pseudo-random number generator

Let $X \sim \text{EECK}(p > -1, q > 0)$, $R \sim \text{U}(0, 1)$. The algorithm for generating n values of X, using the inverse CDF method, with $\text{CDF}(x; p, q)$ given by (2.5), is as follows:

1. Repeat steps 1.1–1.4 n times:
  1.1. Let $R \sim \mathrm{U}(0,1)$;
  1.2. Let $x = -1 + 0.01$;
  1.3. If $\mathrm{CDF}(x; p, q) < R$, then $x = x + 0.01$;
  1.4. Return x.

It is obviously a universal algorithm for any distribution with $\mathrm{CDF}(x; par)$, where par is the vector of distribution parameters.

The quantile function of the $\mathrm{EECK}(p, q)$ does not have an analytical form, PDF (2.3) is non-negative on the interval $[-1, 1]$ and bounded by constant $d = f(0; p \geq 0, q)$, then we can use the von Neumann method, which in this case is much faster than the inverse CDF method. The algorithm for generating n values of X, using the von Neumann method (Von Neumann, 1951), is as follows:

1. If $-1 < p < 0$ then use the inverse CDF method.
2. If $p \geq 0$ then $d = f(0; p, q)$.
3. Repeat steps 3.1–3.3 n times:
  3.1. Let $R_1 \sim \mathrm{U}(-1, 1)$, $R_2 \sim \mathrm{U}(0, d)$;
  3.2. If $f(R_1; p, q) < R_2$ then go to Step 3.1 else $x = R_1$;
  3.3. Return x.

The `R` codes of the `rEECK` and `rEECK1` functions for generating n values of X are presented in the supplementary material.

2.6. Fisher information matrix

Theorem 2.8. *The Fisher information matrix* $I_{i,j}$ $(i, j = 1, 2)$ *for the* $\mathrm{EECK}(p > -1, q > 0)$ *distribution is given by*

$$I_{11} = \left[A - B + \widetilde{H}(p) - \widetilde{H}\left(p + \frac{1}{q}\right)\right]^2 + \Psi_1(p+1) - \Psi_1\left(p + \frac{1}{q} + 1\right), \tag{2.26}$$

$$I_{12} = I_{21} = \frac{(A-B)(C-A)}{q^2} - \frac{(A-B)\,\Gamma\left(p + \frac{1}{q} + 1\right)}{\Gamma(p+1)\,\Gamma\left(\frac{1}{q} + 1\right)} + \frac{(C-A)\left[\widetilde{H}(p) - \widetilde{H}\left(p + \frac{1}{q}\right)\right]}{q^2} + \frac{\Gamma\left(p + \frac{1}{q} + 1\right)}{p\,\Gamma(p+1)\,\Gamma\left(\frac{1}{q} + 1\right)}, \tag{2.27}$$

$$I_{22} = \frac{(C-A)^2}{q^4} - \frac{2(C-A)\,\Gamma\left(p + \frac{1}{q} + 1\right)}{q^3\,\Gamma(p+1)\,\Gamma\left(\frac{1}{q} + 1\right)} + \frac{pq^2(pq+1)\,\Gamma\left(2 - \frac{1}{q}\right)\Gamma\left(p + \frac{1}{q}\right)}{(p-1)(pq-1)\,\Gamma\left(p - \frac{1}{q}\right)\Gamma\left(\frac{1}{q}\right)}, \tag{2.28}$$

where $\widetilde{H}(z) = \sum_{k=1}^{z} \frac{1}{k}$ *is the harmonic function,* $\Psi_n(z)$ *is the* n*-th derivative of the digamma function* $\Psi(z)$, $A = \Psi\left(p + \frac{1}{q} + 1\right)$, $B = \Psi(p+1)$, $C = \Psi\left(\frac{1}{q} + 1\right)$ *as well as* I_{11}, $I_{12} = I_{21}$, I_{22} *are defined for* $(p > -1, q > 0)$, $(p > 0, q > 0)$ *and* $(p > 1, q > 0.5)$ *respectively.*

Proof: First, we need to take the logarithm. From (2.11) we have

$$\ln\big[f(x;p,q)\big] \;=\; \ln\left[\Gamma\left(p+\frac{1}{q}+1\right)\right] + p\ln\big(1-|x|^q\big) - \ln\big[2\Gamma(p+1)\big] - \ln\Gamma\left(\frac{1}{q}+1\right).$$

Second, we need to calculate the partial derivatives

$$\frac{d\ln\big[f(x;p,q)\big]}{dp} \;=\; \Psi\left(p+\frac{1}{q}+1\right) + \ln\big(1-|x|^q\big) - \Psi(p+1),$$

$$\frac{d\ln\big[f(x;p,q)\big]}{dq} \;=\; \frac{-1}{q^2}\,\Psi\left(p+\frac{1}{q}+1\right) - \frac{pq\,|x|^{q-1}}{1-|x|^q} + \frac{1}{q^2}\,\Psi\left(\frac{1}{q}+1\right).$$

Hence, we get the Fisher score in the form

$$\boldsymbol{h(x;p,q)} \;=\; \begin{bmatrix} A-B+\ln\big(1-|x|^q\big) \\ \dfrac{C-A}{q^2} - \dfrac{pq\,|x|^{q-1}}{1-|x|^q} \end{bmatrix}.$$

Let $\boldsymbol{u}(x;p,q) = \boldsymbol{h}(x;p,q)\,\boldsymbol{h}(x;p,q)^\top$ then

$$u_{11} = \Big[A-B+\ln\big(1-|x|^q\big)\Big]^2, \qquad u_{22} = \left[\frac{C-A}{q^2} - \frac{pq\,|x|^{q-1}}{1-|x|^q}\right]^2,$$

$$u_{12} \;=\; u_{21} \;=\; \Big[A-B+\ln\big(1-|x|^q\big)\Big]\left[\frac{C-A}{q^2} - \frac{pq\,|x|^{q-1}}{1-|x|^q}\right].$$

Let $I_{i,j} = E[u_{i,j}]$ $(i,j=1,2)$ then

$$I_{11} \;=\; (A-B)^2 + 2(A-B)\,E\Big[\ln\big(1-|x|^q\big)\Big] + E\Big[\ln^2\big(1-|x|^q\big)\Big], \tag{2.29}$$

$$\begin{aligned} I_{12} \;=\; I_{21} \;=\; & \frac{(A-B)\,(C-A)}{q^2} - pq(A-B)\,E\left[\frac{|x|^{q-1}}{1-|x|^q}\right] \\ & + \frac{C-A}{q^2}\,E\Big[\ln\big(1-|x|^q\big)\Big] - pq\,E\left[\frac{|x|^{q-1}\ln\big(1-|x|^q\big)}{1-|x|^q}\right], \end{aligned} \tag{2.30}$$

$$I_{22} \;=\; \frac{(C-A)^2}{q^4} - \frac{2p(C-A)}{q^2}\,E\left[\frac{|x|^{q-1}}{1-|x|^q}\right] + p^2q^2\,E\left[\frac{|x|^{2q-2}}{\big(1-|x|^q\big)^2}\right]. \tag{2.31}$$

To write the Fisher Information Matrix in a simpler form, we use (2.2) and Mathematica software. We obtain:

$$E\Big[\ln\big(1-|x|^q\big)\Big] \;=\; \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)} \int_{-1}^{1} \frac{\ln\big(1-|x|^q\big)}{\big(1-|x|^q\big)^{-p}}\,dx \;=\; \widetilde{H}(p) - \widetilde{H}\left(p+\frac{1}{q}\right), \tag{2.32}$$

$$(2.33)\quad E\left[\ln^2\left(1-|x|^q\right)\right] = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\int_{-1}^{1}\frac{\ln^2\left(1-|x|^q\right)}{\left(1-|x|^q\right)^{-p}}\,dx$$
$$= \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,\frac{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}{\Gamma\left(p+\frac{1}{q}+1\right)}$$
$$\cdot\left\{\left[\widetilde{H}(p)-\widetilde{H}\left(p+\frac{1}{q}\right)\right]^2+\Psi_1(p+1)-\Psi_1\left(p+\frac{1}{q}+1\right)\right\}$$
$$= \left[\widetilde{H}(p)-\widetilde{H}\left(p+\frac{1}{q}\right)\right]^2+\Psi_1(p+1)-\Psi_1\left(p+\frac{1}{q}+1\right),$$

$$(2.34)\quad E\left[\frac{|x|^{q-1}}{1-|x|^q}\right] = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\int_{-1}^{1}\frac{|x|^{q-1}}{\left(1-|x|^q\right)^{-p+1}}\,dx = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{pq\,\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)},$$

$$(2.35)\quad E\left[\frac{|x|^{q-1}\ln\left(1-|x|^q\right)}{1-|x|^q}\right] = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\int_{-1}^{1}\frac{|x|^{q-1}\ln\left(1-|x|^q\right)}{\left(1-|x|^q\right)^{-p+1}}\,dx =$$
$$= \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,\frac{-2}{p^2q} = \frac{-\Gamma\left(p+\frac{1}{q}+1\right)}{p^2q\,\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)},$$

$$(2.36)\quad E\left[\frac{|x|^{2q-2}}{\left(1-|x|^q\right)^2}\right] = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\int_{-1}^{1}\frac{|x|^{2q-2}}{\left(1-|x|^q\right)^{-p+2}}\,dx =$$
$$= \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\,\frac{2\Gamma(p-1)\,\Gamma\left(2-\frac{1}{q}\right)}{q\,\Gamma\left(1+p-\frac{1}{q}\right)} = \frac{(pq+1)\,\Gamma\left(2-\frac{1}{q}\right)\Gamma\left(p+\frac{1}{q}\right)}{p\,(p-1)\,(pq-1)\,\Gamma\left(p-\frac{1}{q}\right)\Gamma\left(\frac{1}{q}\right)}.$$

Substituting formulas (2.32)–(2.36) into formulas (2.29)–(2.31), as a result of simple transformations, we get formulas (2.26)–(2.28). □

3. MAXIMUM LIKELIHOOD ESTIMATION

Let $x_1^*, x_2^*, ..., x_n^*$ be a random sample size n from the EECK$(p>-1, q>0)$ distribution. Our target is to estimate the unknown values of the parameters p, q. The likelihood function based on (2.3) is given by

$$L = \prod_{i=1}^{n} f(x_i^*; p, q) = \frac{\Gamma\left(p+\frac{1}{q}+1\right)}{2\Gamma(p+1)\,\Gamma\left(\frac{1}{q}+1\right)}\prod_{i=1}^{n}\left(1-|x_i^*|^q\right)^p,$$

then the log-likelihood function is defined as

$$(3.1)\quad l = n\ln\left[\Gamma\left(p+\frac{1}{q}+1\right)\right]-n\ln\left[2\Gamma(p+1)\right]-n\ln\left[\Gamma\left(\frac{1}{q}+1\right)\right]+p\sum_{i=1}^{n}\ln\left(1-|x_i^*|^q\right)$$

and

$$\frac{dl}{dp} = n\Psi\left(p+\frac{1}{q}+1\right) - n\Psi(p+1) + \sum_{i=1}^{n}\ln\left(1-|x_i^*|^q\right) = 0, \tag{3.2}$$

$$\frac{dl}{dq} = \frac{-n}{q^2}\,\Psi\left(p+\frac{1}{q}+1\right) + \frac{n}{q^2}\,\Psi\left(\frac{1}{q}+1\right) - \frac{npq|x_i^*|^{q-1}}{1-|x_i^*|^q} = 0, \tag{3.3}$$

where Ψ is the digamma function.

The maximum likelihood estimates (MLEs) are solutions of the system equations (3.2)–(3.3). We have

$$\frac{1}{n}\sum_{i=1}^{n}\ln\left(1-|x_i^*|^q\right) = \Psi(p+1) - \Psi\left(p+\frac{1}{q}+1\right), \tag{3.4}$$

$$\Psi\left(\frac{1}{q}+1\right) - \Psi\left(p+\frac{1}{q}+1\right) = -\frac{pq^3|x_i^*|^{q-1}}{1-|x_i^*|^q}. \tag{3.5}$$

Solving the system equations (3.2)–(3.3) with numerical method we have obtain $\hat{p}, \hat{q}$. We can also maximize the log-likelihood function (3.1) to obtain the MLEs of the p, q parameters.

The biases and the root mean squared errors (RMSEs) of the MLEs are shown in Tables 2 and 3. The simulation study was performed with 10^3 samples using sample sizes of 100, 150, 200.

Table 2: Biases and RMSEs of the MLEs from the EECK$(p, 3)$.

p	n	$\widehat{p}$		$\widehat{q}$	
		Bias	RMSE	Bias	RMSE
1	100	0.555	2.820	0.443	3.593
	150	0.296	1.802	0.186	2.992
	200	0.110	1.172	−0.081	2.279
2	100	0.965	4.408	0.379	2.313
	150	0.724	2.739	0.339	1.908
	200	0.338	1.468	0.110	1.397
3	100	1.255	3.875	0.336	1.701
	150	0.892	3.126	0.269	1.441
	200	0.712	2.289	0.259	1.261

Table 3: Biases and RMSEs of the MLEs from the EECK$(0.5, q)$.

q	n	$\widehat{p}$		$\widehat{q}$	
		Bias	RMSE	Bias	RMSE
1	100	0.266	5.510	0.485	3.361
	150	0.011	1.202	0.231	2.169
	200	−0.125	0.320	−0.034	1.290
2	100	0.173	1.168	0.531	3.600
	150	0.046	0.873	0.176	3.234
	200	−0.020	0.711	−0.044	2.729
3	100	0.264	2.584	0.439	5.733
	150	0.149	1.586	0.209	5.283
	200	0.047	1.005	0.029	4.544

The samples were drawn from the EECK$(p, 3)$, where $p = 1, 2, 3$ (see Table 2) and from the EECK$(3, q)$, where $q = 1, 2, 3$ (see Table 3). We observe that the estimates approach true values when the sample size increases, it implies the consistency of the estimates. The biases of the $\hat{p}$ and $\hat{q}$ diminish for large samples and are smaller for $\hat{q}$ than for $\hat{p}$. The RMSEs decrease with the value of p for $\hat{q}$ (see Table 2).

To examine the accuracy of the coverage probability of the asymptotic confidence intervals (CIs), another simulation study was performed with 10^3 samples using sample sizes of $100, 150, 200$. The study focused on the parameters p, q and samples drawn from the EECK$(p=3, q=3)$. The coverage probabilities of the obtained 95% CIs for $p = 3$, $q = 3$ reported in Table 4 are very close to the nominal level. The results suggested that the obtained standard errors and hence the asymptotic CIs are reliable.

Table 4: Coverage probability for the standard asymptotic 95% CIs, EECK$(p=3, q=3)$.

n	p	q
100	0.954	0.942
150	0.938	0.945
200	0.957	0.96

4. APPLICATION

This section is divided into two subsections. We present examples of the applicability and flexibility of the EECK$(p > -1, q > 0)$. Subsection 4.1 is devoted to GoFTs, Subsection 4.2 deals with fitting distributions to data.

4.1. Comparison of goodness-of-fit tests

As it was mentioned in Introduction, the shape parameter of the EECK distribution cannot be represented as a function of $\overline{\gamma}_2$, as is for the ECK distribution (Sulewski, 2022b). Recall, however, that the ECK excess kurtosis takes values on interval $(-2, 0)$, while the EECK excess kurtosis has values on interval $[-2, \infty)$. Using e.g. Mathcad, you can easily calculate the argument of a function knowing its value.

The EECK distribution can be extremely useful when you want to seamlessly test GoFTs ability to detect deviations from normality caused by a negative and positive excess kurtosis.

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be an ordered random sample of size n. Seven GoFTs were selected to be subjects of the Monte Carlo simulation. Five of them as being very popular GoFTs have been implemented in the `R` software. These tests are: Shapiro–Wilk (SW), Kolmogorov–Smirnov (KS), Cramer–von Mises (CvM), Anderson–Darling (AD) and Shapiro–Francia (SF).

Two tests not implemented yet, probably for their novelty, are: H_n (Torabi *et al.*, 1961) and $\mathrm{LF_m}$ (Sulewski, 2020) tests.

The H_n test statistic is defined as

$$\mathrm{H}_n = \frac{1}{n}\sum_{i=1}^{n} h\left[\frac{1+\Phi\left(\frac{x_{(i)}-\overline{x}}{s},0,1\right)}{1+\frac{i}{n}}\right], \qquad h(x)=\left(\frac{x-1}{x+1}\right)^2, \tag{4.1}$$

where $\overline{x}$ and s^2 are the sample mean and sample variance, respectively.

The $\mathrm{LF_m}$ test statistic is given by

$$\mathrm{LF_m} = \max\left|\frac{i-\overline{\alpha}}{n-\overline{\alpha}-\overline{\beta}+1} - \Phi\left(\frac{x_{(i)}-\overline{x}}{s},0,1\right)\right|, \qquad (\overline{\alpha},\overline{\beta}\geq 1). \tag{4.2}$$

If an alternatively distribution is both symmetric and of negative (positive) excess kurtosis $\overline{\alpha}=\overline{\beta}=0$ ($\overline{\alpha}=\overline{\beta}=1$) are recommended.

The similarity measure M (2.4) of $\mathrm{N}(0,0.096)$ and $\mathrm{EECK}(p=50,1.96)$, as was mentioned in Subsection 2.1, is 0.999. In the legend of Figure 7, the values of the similarity measure M of the normal distribution and the EECK are given. Figure 7 (left) shows PDF of the $\mathrm{N}(0,0.096)$ and $\mathrm{EECK}(p,1.96)$ distributions. For the presented values of the shape parameters, an excess kurtosis of the EECK is negative (see Table 5). If p increases, the similarity measure also increases. Figure 7 (right) shows PDF of the $\mathrm{N}(0,0.259)$ and $\mathrm{EECK}(p,1.3)$ distributions. The similarity measure M of $\mathrm{N}(0,0.259)$ and $\mathrm{EECK}(p=2.75,1.3)$ is 0.999.

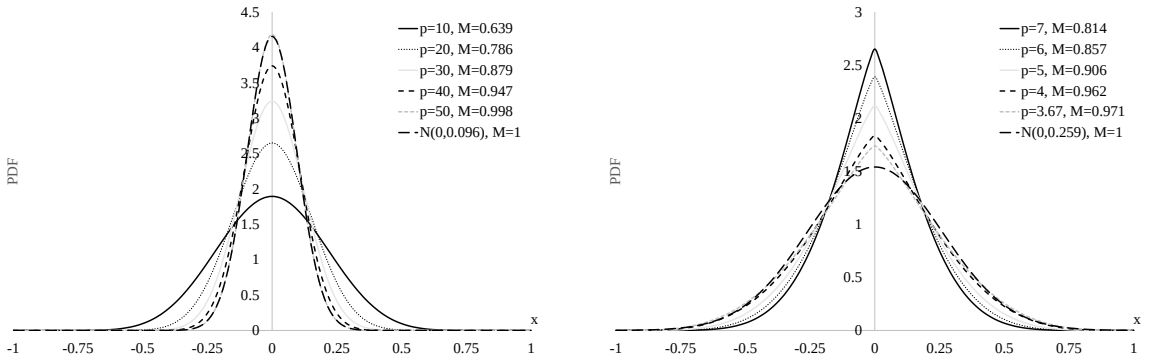

Figure 7: The $\mathrm{EECK}(p,q)$ distribution with values of the similarity measure M to the normal distribution.

Table 5: Modeling of negative excess kurtosis $\overline{\gamma}_2$. $\mathrm{EECK}(p,1.96)$.

	$\mathrm{EECK}(p,1.96)$									$\mathrm{N}(0,0.096)$
$\overline{\gamma}_2$:	-1	-0.75	-0.5	-0.4	-0.3	-0.2	-0.1	-0.05	-0.025	0
p:	0.49	1.451	3.3	4.627	6.731	10.58	19.882	32.188	45.316	—

Source: Own material.

For the presented values of the shape parameters, an excess kurtosis of the EECK is positive (see Table 6). If p decreases, the similarity measure M increases.

Table 5 (Table 6) shows the modeling of negative (positive) excess kurtosis, i.e. for a given value of $\overline{\gamma}_2$ of the EECK$(p, 1.96)$ (EECK$(p, 1.3)$) the value of the shape parameter p is calculated.

Table 6: Modeling of negative excess kurtosis $\overline{\gamma}_2$. EECK$(p, 1.3)$.

	EECK$(p, 1.3)$										N$(0, 0.259)$
$\overline{\gamma}_2$:	0.5	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0
p:	8.261	6.95	5.891	5.018	4.286	3.963	3.81	3.721	3.693	3.669	—

Source: Own material.

Phase 1. In this phase the aim is to investigate to what degree selected GoFTs listed in Table 7 are able to distinct between the normal and proposed distributions. In other words the aim is to determine powers of GoFTs being under discussion when samples come from EECK(p, q) general populations. For the aim to be accomplished, critical values $cv_{0.05}$ ascribed to GoFTs (where $\alpha = 0.05$ is the test significance level) were needed. These critical values were estimated with the Monte Carlo method. Seven large scale experiments were performed each of which devoted to one of GoFT. Each experiment consisted of generating 10^5 samples of sizes $n = 20, 40, 60$. The samples followed the N$(0, 0.096)$ and N$(0, 0.259)$ distributions. Each sample was tested for normality. Obtained in this way values of test statistics (denoted Q_i, $i = 1, 2, ..., m$) were collected an then ranked. Critical values were assessed according to the formula $cv_{0.05} = Q_{[\alpha m]}$.

Table 7 present obtained $cv_{0.05}$ critical values. Tables 8 and 9, in turn, present relevant test powers when samples come from the EECK(p, q) general populations. Each experiment consisted of generating 10^5 samples of sizes $n = 20, 40, 60$. The shape parameter is $q = 1.96$ $(q = 1.3)$. Values of the shape parameter p were listed in Table 5 (Table 6).

Table 7: Critical values $cv_{0.05}$ for GoFTs. The samples of size n followed the N$(0, s)$.

n:	20		40		60	
s:	0.096	0.259	0.096	0.259	0.096	0.259
LF	0.19177	0.19202	0.13841	0.13844	0.11385	0.11376
CvM	0.12278	0.12223	0.12445	0.12446	0.12490	0.12484
AD	0.72300	0.71959	0.73751	0.73840	0.74215	0.74084
SW	0.98287	0.98282	0.98860	0.98861	0.99140	0.99139
SF	0.98464	0.98469	0.99003	0.99007	0.99248	0.99249
H_n	0.00077	0.00076	0.00038	0.00038	0.00025	0.00025
LF_m	0.16195	0.17471	0.12388	0.12895	0.10450	0.10726

The conclusions from Tables 8 and 9 are very interesting. For $n = 20$, the LF, CvM, AD, SW, H_n tests detect only $\overline{\gamma}_2 = -1$, LF_m — $\overline{\gamma}_2 = -0.75$; LF_m, SF tests detect even $\overline{\gamma}_2 = 0.001$. For $n = 40$, the LF, CvM, AD, SW, H_n and LF_m tests detect only $\overline{\gamma}_2 = -0.75$; LF, CvM, AD, H_n, and LF_m tests detect even $\overline{\gamma}_2 = 0.001$. For $n = 60$, the AD, H_n and LF_m tests detect only $\overline{\gamma}_2 = -0.5$; LF and CvM tests detect only $\overline{\gamma}_2 = -0.75$; LF, CvM, AD, H_n, and LF_m tests detect even $\overline{\gamma}_2 = 0.001$.

Table 8: Powers of tests at $\alpha = 0.05$, when the EECK$(p, 1.96)$ is the actual population distribution. The case of negative excess kurtosis values.

GoFT	n	$\overline{\gamma}_2$									
		−1	−0.75	−0.5	−0.4	−0.3	−0.2	−0.1	−0.05	−0.025	0
		p									
		0.49	1.451	3.3	4.627	6.731	10.58	19.882	32.188	45.316	—
LF	20	0.063	0.046	0.044	0.045	0.045	0.048	0.048	0.049	0.050	0.050
	40	0.099	0.058	0.048	0.045	0.047	0.047	0.049	0.049	0.050	0.050
	60	0.148	0.074	0.052	0.049	0.047	0.047	0.049	0.051	0.050	0.051
CvM	20	0.074	0.047	0.044	0.042	0.044	0.047	0.047	0.049	0.050	0.050
	40	0.144	0.069	0.049	0.045	0.045	0.046	0.048	0.049	0.049	0.050
	60	0.237	0.095	0.055	0.049	0.046	0.046	0.049	0.049	0.049	0.051
AD	20	0.079	0.047	0.041	0.040	0.042	0.045	0.047	0.048	0.050	0.050
	40	0.178	0.075	0.048	0.043	0.044	0.044	0.047	0.049	0.048	0.050
	60	0.311	0.109	0.057	0.048	0.045	0.045	0.048	0.049	0.048	0.050
SW	20	0.083	0.043	0.036	0.038	0.039	0.041	0.045	0.048	0.048	0.049
	40	0.223	0.071	0.040	0.036	0.036	0.038	0.043	0.046	0.049	0.051
	60	0.429	0.115	0.047	0.039	0.037	0.037	0.043	0.045	0.046	0.051
SF	20	0.034	0.022	0.025	0.030	0.033	0.039	0.045	0.049	0.050	0.049
	40	0.083	0.025	0.019	0.021	0.025	0.032	0.041	0.045	0.050	0.052
	60	0.195	0.040	0.019	0.020	0.023	0.028	0.040	0.045	0.047	0.051
H_n	20	0.075	0.049	0.043	0.044	0.044	0.046	0.047	0.048	0.049	0.049
	40	0.154	0.074	0.051	0.046	0.047	0.046	0.048	0.049	0.049	0.051
	60	0.259	0.105	0.059	0.053	0.049	0.050	0.051	0.051	0.050	0.053
LF_m	20	0.082	0.056	0.050	0.049	0.048	0.050	0.049	0.049	0.051	0.051
	40	0.125	0.073	0.054	0.050	0.051	0.049	0.051	0.050	0.050	0.051
	60	0.181	0.087	0.059	0.053	0.051	0.049	0.050	0.051	0.050	0.051

In Phase 1, we showed that the considered GoFTs detect positive excess kurtosis better than negative one.

Phase 2. In this phase the aim is to investigate to what degree an undetected excess kurtosis impacts the performance of two basic tests related to parameters of the Normal distribution, namely Student t test and Fisher–Snedecor F test.

Let $x_{1,1}, x_{1,2}, ..., x_{1,n}$ and $x_{2,1}, x_{2,2}, ..., x_{2,n}$ be two samples of sizes n drawn from particular general populations. Let us remember that t and F test statistics have the following forms:

$$\dot{t} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_{x1}^2 + s_{x2}^2}{n}}}, \qquad \dot{F} = \frac{s_{x1}^2}{s_{x2}^2}, \tag{4.3}$$

where $\overline{x}_1, \overline{x}_2$ are the sample means and s_{x1}, s_{x2} are the sample standard deviations.

Table 9: Powers of tests at $\alpha = 0.05$, when the EECK$(p, 1.3)$ is the actual population distribution. The case of positive excess kurtosis values.

GoFT	n	$\overline{\gamma}_2$										
		0.5	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0
		p										
		8.261	6.95	5.891	5.018	4.286	3.963	3.810	3.721	3.693	3.669	—
LF	20	0.072	0.069	0.064	0.060	0.055	0.056	0.053	0.054	0.054	0.053	0.050
	40	0.089	0.081	0.074	0.070	0.061	0.060	0.058	0.056	0.055	0.057	0.051
	60	0.105	0.094	0.084	0.075	0.065	0.062	0.060	0.060	0.060	0.060	0.052
CvM	20	0.081	0.077	0.071	0.063	0.059	0.057	0.056	0.056	0.054	0.055	0.052
	40	0.101	0.091	0.080	0.073	0.064	0.060	0.059	0.056	0.056	0.056	0.049
	60	0.122	0.108	0.093	0.082	0.068	0.063	0.062	0.061	0.060	0.061	0.051
AD	20	0.082	0.077	0.071	0.062	0.057	0.054	0.055	0.054	0.052	0.053	0.052
	40	0.101	0.090	0.079	0.071	0.062	0.058	0.056	0.054	0.053	0.053	0.049
	60	0.124	0.107	0.092	0.081	0.066	0.061	0.060	0.058	0.057	0.059	0.051
SW	20	0.081	0.073	0.067	0.060	0.052	0.051	0.049	0.048	0.047	0.050	0.050
	40	0.098	0.085	0.074	0.060	0.052	0.047	0.045	0.044	0.044	0.048	0.050
	60	0.114	0.095	0.079	0.064	0.050	0.045	0.044	0.042	0.042	0.046	0.051
SF	20	0.102	0.092	0.081	0.072	0.062	0.058	0.055	0.055	0.052	0.057	0.049
	40	0.127	0.111	0.093	0.074	0.061	0.053	0.049	0.048	0.049	0.053	0.049
	60	0.148	0.125	0.102	0.078	0.058	0.050	0.047	0.045	0.044	0.051	0.052
H_n	20	0.078	0.073	0.068	0.061	0.058	0.056	0.055	0.054	0.053	0.055	0.052
	40	0.094	0.084	0.076	0.069	0.061	0.058	0.056	0.055	0.054	0.054	0.049
	60	0.118	0.105	0.091	0.080	0.067	0.063	0.061	0.060	0.059	0.061	0.053
LF_m	20	0.082	0.078	0.073	0.066	0.061	0.060	0.057	0.057	0.057	0.057	0.050
	40	0.100	0.091	0.082	0.076	0.066	0.064	0.062	0.060	0.059	0.060	0.050
	60	0.116	0.104	0.092	0.081	0.069	0.066	0.064	0.064	0.063	0.063	0.051

The course of action was as follows:

Step 1: $m = 10^5$ pairs of samples both of size $n = 60$ were drawn from EECK(4.627, 1.96) (for negative excess kurtosis) and EECK(3.669, 1.3) (for positive excess kurtosis) general populations.

Step 2: These pairs of samples were consecutively, converted into pairs of $\dot{\mathrm{t}}_v$ statistics and $\dot{\mathrm{F}}_v$ statistics, $v = 1, 2, ..., m$.

Step 3: Sets of values of $\dot{\mathrm{t}}_v$ and $\dot{\mathrm{F}}_v$ statistics were stored in two matrices named T and F.

Step 4: The matrices were sorted in ascending order and served to determine two empirical CDFs namely $\Theta_{\mathrm{t}}(\dot{\mathrm{t}}_v)$ and $\Theta_{\mathrm{F}}(\dot{\mathrm{F}}_v)$.

Step 5: Probability papers were employed to check whether the above empirical CDFs fit the Student and Fisher–Snedecor distributions.

Figures 8 and 9 show empirical CDFs of Step 4 plotted on the Student and Snedecor probability papers, when samples were drawn from EECK(4.627, 1.96) and EECK(3.669, 1.3), appropriately. These probability papers were constructed in the same way as the Normal probability is constructed.

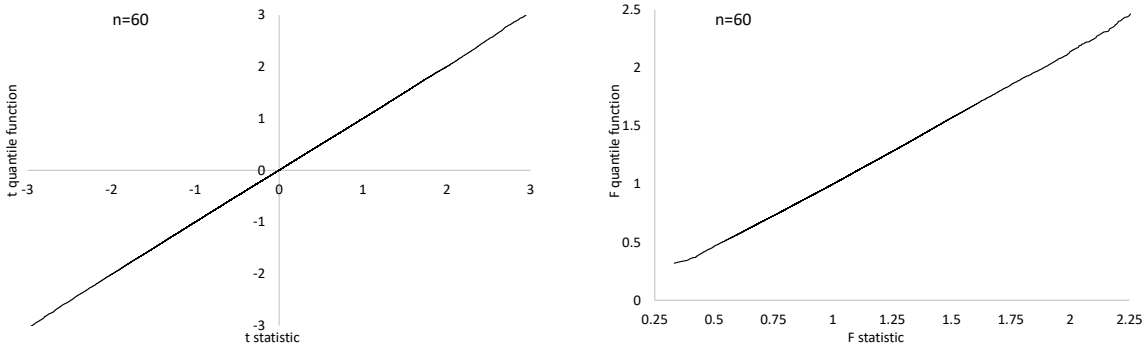

Figure 8: Empirical CDFs of Step 4 plotted on the Student and Snedecor probability paper. Case of negative excess kurtosis values.

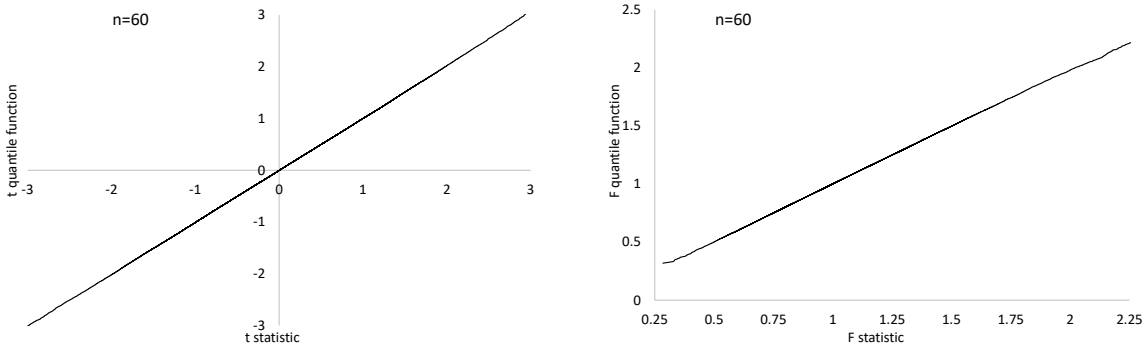

Figure 9: Empirical CDFs of Step 4 plotted on the Student and Snedecor probability paper. Case of positive excess kurtosis values.

It turns out that the empirical distribution in question perfectly fit straight lines that relevant theoretical distributions. Thus, we can conclude that Student and Fisher–Snedecor tests may be applied even as population distributions are of negative (see Figure 8) or positive (see Figure 9) excess kurtosis.

4.2. Fitting distributions to data

Symmetric distributions have limited use in fitting the distributions to data (e.g. normal distribution). However, the situation looks much better when we use their mixture (e.g. compound normal distribution).

For the purposes of this subsection, we extend the domain of the EECK(p,q) from $[-1,1]$ to $[-a,a]$ $(a \in R)$. PDF of the modified EECK(p,q) distribution denoted as EECK2$(x;a,p,q)$ has the form

$$\text{EECK2}(x;a,p,q) = \frac{\int_{-a}^{a}\left[1-\left(\frac{|x|}{a}\right)^{q}\right]^{p}dx}{2\int_{0}^{a}\left[1-\left(\frac{|u|}{a}\right)^{q}\right]^{p}du}, \qquad x \in \begin{cases} (-a,a) & \text{if } -1<p<0, \\ [-a,a] & \text{if } p\geq 0. \end{cases} \tag{4.4}$$

In this subsection, we present real data examples to demonstrate a flexibility of the EECK($p>-1, q>0$) distribution in the mixed variant. PDF of the compound EECK (CEECK) distribution is given by

$$\text{CEECK}(x; a, p_1, q_1, p_2, q_2, \omega) = \omega\,\text{EECK2}(x; a, p_1, q_1) + (1-\omega)\,\text{EECK2}(x; a, p_2, q_2). \tag{4.5}$$

The estimation of the model parameters is carried out by the maximum likelihood method. To avoid local maxima of the logarithmic likelihood function, the optimization routine is run 100 times with several different starting values that are widely scattered in the parameter space. The KS GoFT was used for model fitting, while the AIC, BIC and HQIC were used for model comparisons. The p-values for the KS GoFT is calculated as follows. First, we obtain the values of the KS test statistics (denoted ST) for true values of parameters $\widehat{\Theta}$ based on the sample $x_{(1)}, x_{(2)}, ..., x_{(n)}$. In the next step we simulate 10^3 samples $x'_{(1)}, x'_{(2)}, ..., x'_{(n)}$ from the given distribution with true values of parameters $\widehat{\Theta}$. For each sample, we calculate the values of the KS test statistics (denoted ST^{s}). Finally, the p-value is calculated as $p \approx \#\{i\colon \text{ST}_i^{\text{s}} > \text{ST}\}\,10^{-3}$.

Real data examples

The first data set presents temperature dynamics of beaver Castor canadensis in north-central Wisconsin (Reynolds, 1994). Body temperature was measured by telemetry every 10 minutes from one period of less than a day. The data consists of 114 observations of the variable "measured body temperature in degrees Celsius" and are available in the `R` software with code `beaver1[3]`.

The second data set contains statistics, in arrests per 100,000 residents for assault in each of the 50 US states in 1973 (McNeil, 1977). The data consisting of 50 observations are available in the `R` software with code `USArrests[2]`.

The models selected for comparison with the CEECK($x; a, p_1, q_1, p_2, q_2, \omega$) are:

- the compound ECK (CECK):
$$f_{\text{CECK}}(x; a, p_1, p_2, \omega) = \omega\,\frac{\left(1-\frac{x^2}{a^2}\right)^{p_1}}{aB(0.5, p_1+1)} + (1-\omega)\,\frac{\left(1-\frac{x^2}{a^2}\right)^{p_2}}{aB(0.5, p_2+1)};$$
- the compound normal (CN):
$$f_{\text{CN}}(x; a_1, b_1, a_2, b_2, \omega) = \omega\,\phi(x; a_1, b_1) + (1-\omega)\,\phi(x; a_2, b_2);$$
- the compound Laplace (CL):
$$f_{\text{CL}}(x; a_1, b_1, a_2, b_2, \omega) = \frac{\omega}{2b_1}\exp\left[\exp\left(-\frac{|x-a_1|}{b_1}\right)\right] + \frac{1-\omega}{2b_2}\exp\left[\exp\left(-\frac{|x-a_2|}{b_2}\right)\right];$$
- the compound Cauchy (CC):
$$f_{\text{CC}}(x; a_1, b_1, a_2, b_2, \omega) = \frac{\omega}{\pi\, b_1\left[1+\left(\frac{x-a_1}{b_1}\right)^2\right]} + \frac{1-\omega}{\pi\, b_2\left[1+\left(\frac{x-a_2}{b_2}\right)^2\right]};$$
- the compound logistic (CLOG):
$$f_{\text{CLOG}}(x; a_1, b_1, a_2, b_2, \omega) = \frac{\omega\exp\left(\frac{x-a_1}{b_1}\right)}{b_1\left[1+\exp\left(\frac{x-a_1}{b_1}\right)^2\right]} + \frac{(1-\omega)\exp\left(\frac{x-a_2}{b_2}\right)}{b_2\left[1+\exp\left(\frac{x-a_2}{b_2}\right)^2\right]}.$$

Tables 10 and 11 present values of the MLEs, log-likelihood function l, information criteria, KS test statistics and p-value for the first and second data set, respectively. The lowest values are in bold. The values of standard errors (calculated in the `R` software) for some parameters in the CEECK models are surprisingly large compared to other models. These values for the first data set are smaller than for the second data set. Figure 10 presents histograms, estimated PDFs of the analyzed models for the first (left) and second (right) data sets.

Table 10: Results of estimation for the first data set.
The respective standard errors are in parentheses.

Model	MLEs	AIC	BIG	HQIC	KS (p-value)
CEECK	$\widehat{a} = 4.151(1.144)$, $\widehat{p}_1 = 2.039(1.981)$, $\widehat{q}_1 = 0.700(0.827)$, $\widehat{p}_2 = 8958.252(33.333)$, $\widehat{q}_2 = 5.371(0.623)$, $\widehat{\omega} = 0.660(0.177)$	321.570	337.987	328.233	**0.040 (0.978)**
CECK	$\widehat{a} = 5.010(3.732)$, $\widehat{p}_1 = 5.413(11.662)$, $\widehat{p}_2 = 55.361(95.471)$, $\widehat{\omega} = 0.484(0.162)$	318.331	329.275	322.773	0.041 (0.964)
CN	$\widehat{a}_1 = -2.700(0.025)$, $\widehat{b}_1 = 0.042(0.017)$, $\widehat{a}_2 = 0.071(0.086)$ $\widehat{b}_2 = 0.906(0.062)$, $\widehat{\omega} = 0.026(0.015)$	319.489	333.170	325.041	0.084 (0.324)
CL	$\widehat{a}_1 = -0.580(0.022)$, $\widehat{b}_1 = 0.722(0.326)$, $\widehat{a}_2 = 0.144(0.010)$, $\widehat{b}_2 = 0.663(0.107)$, $\widehat{\omega} = 0.201(0.127)$	320.786	334.467	326.338	0.087 (0.277)
CC	$\widehat{a}_1 = -0.581(0.187)$, $\widehat{b}_1 = 0.353(0.133)$, $\widehat{a}_2 = 0.226(0.095)$, $\widehat{b}_2 = 0.369(0.075)$, $\widehat{\omega} = 0.294(0.149)$	336.771	350.452	342.324	0.103 (0.142)
CLOG	$\widehat{a}_1 = 0.045(0.080)$, $\widehat{b}_1 = 0.492(0.040)$, $\widehat{a}_2 = -2.700(0.029)$, $\widehat{b}_2 = 0.026(0.012)$, $\widehat{\omega} = 0.975(0.015)$	**314.636**	**328.317**	**320.188**	0.050 (0.877)

Source: Own material.

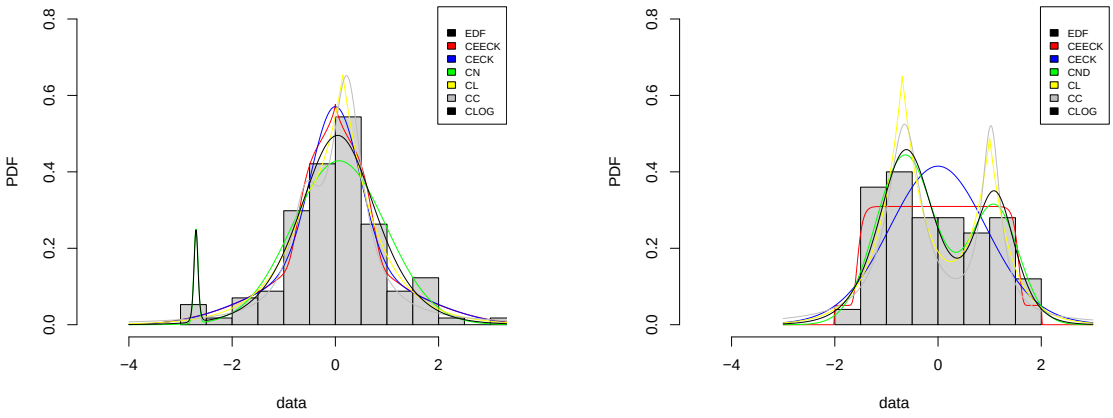

Figure 10: Histograms and estimated PDF of analyzed models for first (left) and second (right) data sets.

Table 11: Results of estimation for the second data set.
The respective standard errors are in parentheses.

Model	MLEs	AIC	BIG	HQIC	KS (p-value)
CEECK	$\widehat{a} = 2.040(0.070)$, $\widehat{p}_1 = 183.915(34.654)$, $\widehat{q}_1 = 378.415(68.096)$, $\widehat{p}_2 = 389.387(86.483)$, $\widehat{q}_2 = 23.107(9.233)$, $\widehat{\omega} = 0.203(0.141)$	**139.901**	**151.373**	**144.270**	0.088 (0.687)
CECK	$\widehat{a} = 9.802(0.154)$, $\widehat{p}_1 = 32.874(0.613)$, $\widehat{p}_2 = 63.054(8.567)$, $\widehat{\omega} = 0.356(0.078)$	149.929	157.577	152.841	0.136 (0.215)
CN	$\widehat{a}_1 = -0.637(0.182)$, $\widehat{b}_1 = 0.567(0.125)$, $\widehat{a}_2 = 1.089(0.241)$, $\widehat{b}_2 = 0.473(0.169)$, $\widehat{\omega} = 0.631(0.122)$	142.002	151.562	145.643	0.072 (0.853)
CL	$\widehat{a}_1 = 0.999(0.023)$, $\widehat{b}_1 = 0.415(0.126)$, $\widehat{a}_2 = -0.693(0.022)$, $\widehat{b}_2 = 0.473(0.107)$, $\widehat{\omega} = 0.392(0.089)$	143.930	153.490	147.570	**0.061 (0.944)**
CC	$\widehat{a}_1 = -0.650(0.129)$, $\widehat{b}_1 = 0.404(0.114)$, $\widehat{a}_2 = 1.025(0.080)$, $\widehat{b}_2 = 0.223(0.112)$, $\widehat{\omega} = 0.655(0.101)$	156.244	165.805	159.885	0.081 (0.767)
CLOG	$\widehat{a}_1 = -0.617(0.157)$, $\widehat{b}_1 = 0.355(0.074)$, $\widehat{a}_2 = 1.099(0.163)$, $\widehat{b}_2 = 0.262(0.088)$, $\widehat{\omega} = 0.648(0.103)$	143.369	152.929	147.009	0.072 (0.880)

Source: Own material.

The CLOG model is the best in terms of the AIC, BIC and HQIC values and the CEECK model is distinguished in terms of the KS GoFT. It has the lowest KS test statistics and the highest p-value (see Table 10). The CEECK model is the best in terms of the AIC, BIC and HQIC values and the CL model is distinguished in terms of the KS GoFT. It has the lowest KS test statistics and the highest p-value (see Table 11). Based on the graphical and the numerical results, the CEECK distribution is considered as one of the best models for the analyzed data sets.

5. CONCLUSIONS

The article presents the extended easily changeable kurtosis (EECK) distribution, the special cases of which are the ECK, uniform and triangle distributions. The new proposal tends to the normal distribution. The EECK, like the ECK, belongs to the family of symmetric, unimodal distributions, defined in the finite domain with excess kurtosis values on infinite interval. The obtained results demonstrate that the EECK distribution can be extremely useful when we want to seamlessly test GoFT's ability to detect deviations from normality by modeling of negative or positive excess kurtosis. Student and Fisher–Snedecor tests may be applied even as population distributions are of negative or positive excess kurtosis. Real data example demonstrates that the EECK(p, q) distribution in the mixed variant is flexible and competitive model that deserves to be added to the existing distributions in data modeling. The information presented in the article shows that the proposed distribution deserves to be added to the symmetric distribution family.

REFERENCES

Ashour, S.K. and Eltehiwy, M.A. (2013). Transmuted exponentiated modified Weibull distribution. *International Journal of Basic and Applied Sciences*, 2(3):258–269.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 2(12):171–178.

Balakrishnan, N. (1992). *Handbook of the Logistic Distribution*. Marcel Dekker, New York.

Birnbaum, Z,W. and Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability*, 6(2):319–327.

Bolfarine, H., Martínez-Florez, G., and Salinas, H.S. (2018). Bimodal symmetric-asymmetric power-normal families. *Communications in Statistics – Theory and Methods*, 47(2):259–276.

Buchanan, K. and Wheeland, S. (2022). *Comparison of the quadratic U and inverse quadratic U sum-difference beam patterns*. In "2022 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting", 1828–1829.

Bucher, J.L. (2012). *The Metrology Handbook*. Second Edition. ASQ Quality Press.

Dekking, F.M., Kraaikamp, C., Lopuhaä, H.P., and Meester, L.E. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How* (Vol. 488). London Springer.

Edwards, A.W.F. (2000). Gilberts sine distribution. *Teaching Statistics*, 22(3):70–71.

Freimer, M., Kollia, G., Mudholkar, G.S., and Lin, C.T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics – Theory and Methods*, 17(10):3547–3567.

Gibbons, J.F. and Mylroie, S. (1985). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Applied Physics Letters*, 11(22):568–569.

Glen, S. (2025). Degenerate distribution: simple definition and examples. From StatisticsHowTo.com: *Elementary Statistics for the rest of us!*
`https://www.statisticshowto.com/degenerate-distribution/`

Gradshteyn, I.S. and Ryzhik, I.M. (2014). *Table of Integrals, Series, and Products*. Academic Press.

Hassan, M.Y. and Hijazi, R.H. (2010). A bimodal exponential power distribution. *Pak. J. Statist.*, 26(2):379–396.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2 (Vol. 289). John Wiley and Sons.

Ki, H., Choi, B., Chang, K.H., and Lee, M. (2005). Option pricing under extended normal distribution. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 25(9):845–871.

Kotz, S., Balakrishnan, N., and Johnson, N.L. (2004). *Continuous Multivariate Distributions*, volume 1 (Vol. 1). John Wiley and Sons.

Moors, J.J.A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1(37):25–32.

Lévy, P. (1940). Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 7:283–339.

Malachov, A.N. (1978). A cumulant analysis of random non-Gaussian processes and their transformations (in Russian). *Soviet Radio*. Moscow.

Mardia, K.V., Jupp, P.E., and Mardia, K.V. (2000). *Directional Statistics* (Vol. 2). Wiley, Chichester.

McNeil, D.R. (1977). *Interactive Data Analysis*. Wiley, New York.

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, 32(7):685–694.

Raab, D. and Green, E. (1961). A cosine approximation to the normal distribution. *Psychometrika*, 26(4):447–450.

Reynolds, P.S. (1994). Time-series analysis of beaver body temperatures. In "Case Studies in Biometry", Chapter 11 of Lange, N., Billard, L., Conquest, L., Ryan, L., Brillinger, D., and Greenhouse J. (Eds.), *Proc. Berkeley Symposium, Statist. Probability*, Wiley, New York.

Rinne, H. (2010). *Location-Scale Distributions – Linear Estimation and Probability Plotting Using MATLAB*, p. 116.

Ryan, B.K. (2014). *Theory and applications of aperiodic (random) phased arrays*. PhD Thesis.

Sulewski, P. (2020). Modified Lilliefors goodness-of-fit test for normality. *Communications in Statistics – Simulation and Computation*, 51(3):1199–1219.

Sulewski, P. (2021). Two-piece power normal distribution. *Communications in Statistics – Theory and Methods*, 50(11):2619–2639.

Sulewski, P. (2022a). Normal distribution with plasticizing component. *Communications in Statistics – Theory and Method*, 51(11):3806–3835.

Sulewski, P. (2022b). Easily changeable kurtosis distribution. *Austrian Journal of Statistics*. `https://www.ajs.or.at/index.php/ajs/article/view/1434`

Temme, N.M. (2010). Voigt function. *NIST Handbook of Mathematical Functions*.

Torabi, H., Montazeri, N.H., and Grane, A. (1961). A test of normality based on the empirical distribution function. *SORT*, 1(40):55–88.

Umarov, S., Tsallis, C., and Steinberg, S. (2008). On a q-central limit theorem consistent with nonextensive statistical mechanics. *Milan J. Math. Birkhauser Verlag*, 76:307–328.

Von Neumann, J. (1951). Various techniques used in connection with random digits. *National Bureau of Standards Applied Mathematics Series*, 12:36–38.

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 491–502
https://doi.org/10.57805/revstat.v23i4.669

S-values and Surprisal Intervals to Replace P-values and Confidence Intervals

Author: Alessandro Rovetta
– Technological and Scientific Research, Redeev SRL, Naples, Italy
alessandrorovetta@redeev.com

Received: September 2023 Revised: January 2024 Accepted: January 2024

Abstract:

- Misuse of statistical significance continues to be prevalent in science. The absence of intuitive explanations of this concept often leads researchers to incorrect conclusions. For this reason, some statisticians suggest adopting S-values (surprisals) instead of P-values, as they relate the statistical relevance of an event to the number of consecutive heads when flipping an unbiased coin. This paper introduces the concept of surprisal intervals (S-intervals) as extensions of confidence/compatibility intervals. The proposed approach imposes the assessment of outcomes in terms of more and less surprising than some values, instead of statistically significant and statistically non-significant. Moreover, a novel methodology for presenting multiple consecutive S-intervals (or compatibility intervals as well) in order to evaluate the variation in surprise (or compatibility) with various target hypotheses is discussed.

Keywords:

- *confidence intervals; epidemiology; hypothesis testing; public health; significance; surprisal.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

The proper interpretation of P-values in science has been debated for decades (Greenland, 2017; Amrhein *et al.*, 2019). Widespread misinterpretation of this measure has even led some academic journals to abandon its use (Lakens, 2021). However, Greenland *et al.* (2022) emphasize that P-values can still provide valid information for making sound scientific decisions if used as a measure of statistical compatibility instead of statistical significance. In this regard, there are some considerations to be made. Let's suppose we choose a specific statistical test and set a certain target assumption (e.g., also called "target hypothesis"). Every statistical test is mathematically built on the condition that all assumptions, including the target, are true. Then, in a Fisherian sense, the P-value measures the degree of compatibility of the statistical result (the test statistic) with the target and all the background assumptions (e.g., linearity, normality, properly functioning measurement devices). P-values close to 1 indicate high compatibility, while P-values close to 0 indicate low compatibility. Thus, although we may be interested solely in the target hypothesis, it is important to understand that the P-value does not privilege said hypothesis over any other. Indeed, violating the background assumptions can strongly influence P-values, making them uninformative for the fixed scientific goal. Moreover, the reliability of the statistical outcome depends on the scientist's ability to conduct the whole experimental procedure (which cannot be carried out without uncertainties). This means selecting a model capable of providing useful information to analyze the scientific phenomenon (which includes choosing proper data collection methods, estimators or parameters, and hypotheses) as well as guaranteeing human attributes like competence, honesty, transparency, and collaboration (Greenland, 2023a). Thus, in light of the interpretative uncertainties that P-values entail, the practice of sharply distinguishing arbitrarily close values (e.g., $P = 0.049$ and $P = 0.051$) is meaningless. According to this, from now on, we will refer to the condition "all background assumptions are true" using the expression "utopian scenario" (emphasizing the practical impossibility of achieving it). Even in the utopian scenario, the P-value is mathematically precluded from providing information about the investigated scientific phenomenon: at best, it can be understood as the probability that, in an ideal world of pure chance, we would obtain a discrepancy from the target hypothesis prediction as or more extreme than that obtained in our experiment according to the performed test.[1] The key point is that the model assumes that chance is the sole factor at play. In other words, under the target null assumption of zero effect, the statistical model mathematically excludes the occurrence of any scientific phenomenon other than chance (e.g., if our objective is to investigate a drug's effectiveness, under the target null assumption of zero effect, the statistical model we implement mathematically excludes the existence of any pharmacological effect). Indeed, a statistical model takes numbers and yields numbers; it is up to the scientist to interpret these based on the research context. For this reason, it never makes sense to state that the P-value is the probability that chance produced or would produce the observed scientific effect. Even in the utopian scenario, the P-value alone never allows the researcher to epistemically reject a target hypothesis or to confirm it (since even a P-value of .99 does not exclude the presence of many other models with equal compatibility). In this regard, P-values are not absolute measures of compatibility, as the data consistency with a certain hypothesis could change drastically depending on

[1]The phrase "chance alone produces" is incomplete as it does not encompass fields of science where randomness is the absence of any cause. Nevertheless, this expression has been chosen because it was considered clearer and suitable for the context of public health.

the adopted test (e.g., the data may be highly consistent with the normality hypothesis via Shapiro–Wilk, but not Kolmogorov–Smirnov). Alongside this, degrees of compatibility that appear markedly different could be highly compatible with each other. As shown by McShane *et al.* (2023), an original study with $P = 0.005$ and a replication study with $P = 0.194$ were highly compatible with one another in the sense that the P-value of the chosen comparison test, assuming no difference between them, was $P = 0.289$. Therefore, the difference between "statistically significant" and "statistically not significant" would be "statistically not significant" at the 0.05 level (Gelman and Stern, 2006). Nonetheless, the pitfall of adopting dichotomous thresholds goes beyond this example, as it blends two incompatible approaches: the (neo) Fisherian one, as described above, and the decision-theoretic Neyman–Pearson one. The first is mathematically structured to provide information on individual studies under the conditions mentioned above, while the second is mathematically structured to provide information on groups of studies (but never on individuals within that group) in numerous repetitions under the same scientific conditions (utopian scenario). This even leads to two distinct mathematical definitions of the P-value, which the reader can delve into by consulting other literature (Greenland, 2023a,b). Given that the overall goal of public health statistics is to inform decisions based on individual studies (e.g., randomized control trials, systematic reviews with meta-analysis, etc.), the (neo) Fisherian approach should be preferred. Nevertheless, in addition to what has already been discussed, there are further inherent difficulties in the use of the P-value that could be addressed by adopting some valid alternatives.

2. SURPRISAL AS AN ALTERNATIVE TO STATISTICAL SIGNIFICANCE AND COMPATIBILITY

2.1. Relationship between P-values and S-values

P-values exhibit some counterintuitive behaviors. For instance, even though the pairs $(P_1 = 0.05, P_2 = 0.10)$ and $(P_3 = 0.90, P_4 = 0.95)$ are formed by P-values that differ by the same amount, $\Delta P = 0.05$, the information contained in the regions identified by these two pairs differs substantially. This happens because the area of the corresponding curve is geometrically distributed differently along the bell curve. For this reason, the use of Shannon information (also known as "surprisal" or "S-value") has been proposed based on the following reasoning: given the probability P of an event, this can be related to the probability of obtaining S consecutive heads by flipping an unbiased coin S times using the formula $P = 0.5^S = 2^{-S}$ (Rafi and Greenland, 2020). It follows that $S = -\log_2 P$. In the utopian scenario, the S-value measures the degree of surprise of the test result (the "t" statistics) compared to the target assumption. The aim is to compare "statistical significance" with a phenomenon that we are familiar with in everyday life. However, mathematically speaking, the S-value measures continuous information (bits). It is up to the reader to interpret that information in relation to the context. Values such as $S = 4.3$ cannot be understood as "4.3 consecutive heads"; however, this writing can be interpreted as "in the utopian scenario, the statistical result is approximately as surprising as 4 consecutive heads — or slightly more than 4 consecutive heads — when tossing a fair coin 4 times". At the conventional threshold $P = .05$, $S = 4.3$ bits correspond. Thus, when we evaluate the difference between $P_1 = .05$ and $P_2 = .10$, we obtain $\Delta S = |S_2 - S_1| = \log_2(0.10/0.05) = 1$ bit, while between $P_3 = .90$

and $P_4 = .95$, we obtain $\Delta S = |S_3 - S_4| = \log_2(0.95/0.90) = 0.08$ bits. Hence, the difference in statistical surprise now becomes evident. However, the philosophy underlying the S-value goes beyond this simplification: the goal is to evaluate results in classes of practical equivalence. Considering the uncertainties mentioned above, there is no practical difference between $P = .05$ ($S = 4.3$) and $P = .0625$ ($S = 4$), since both results are surprising by about as much as 4 consecutive heads. This is why it is good practice to round S values to the nearest integer (although more precise values should always be reported as supplementary material to allow for multi-comparison adjustments or meta-analyses).

2.2. S-values don't address the magnitude fallacy

Surprisals can be effective in properly evaluating statistical surprise, but they cannot address the common confusion about the difference between statistical surprise or compatibility and magnitude (Greenland *et al.*, 2016; Kühberger *et al.*, 2015). Therefore, this paragraph will address the relationship between statistical compatibility and effect size, allowing for a proper introduction of the relationship between surprisal and effect size. A statistical phenomenon can be rare and unexpected (high surprise) but weak (low magnitude), meaning it may have little practical impact. For example, while following a weight-loss diet in accordance with the health recommendations of their primary care physician, one may consistently lose about one gram per day for 100 days (a scientific effect that is unlikely to be due to chance) but still be far from their target weight (indeed, losing 100 grams in 100 days has a negligible impact on physical health). If we were to statistically model such a real-world situation using linear regression, we would obtain a very low P-value (indicating a surprising result) under the null model, but also a very low slope coefficient (indicating that the trend's intensity would be low compared to the predetermined objective) (Rovetta, 2023). Nevertheless, the P-value is linked to the concept of effect size (ES), as it can generally be expressed, at least, as a function $P = f(ES, N)$ where N is the sample size. Considering a fixed $N = N_0$, the P-value could be exclusively linked to the effect size. The latter can be examined through the effect size parameter (which can provide information on the intensity of the statistical phenomenon) but also through the width of confidence intervals (which can provide information on how the statistical effect size changes in relation to the P-value). However, the concept of confidence is commonly (mis)understood within an inferential fashion, i.e., it does not probabilistically concern our single already completed experiment. Indeed, a confidence interval can be obtained by selecting an arbitrary threshold α and performing the operation $100 \cdot (1 - \alpha)\%$; the most well-known case is $\alpha = .05$ with a 95% confidence interval. When testing continuous data, by calculating 95% confidence intervals in infinite utopian applications, 95% of these intervals contain the true value (coverage probability) (Biau *et al.*, 2010). However, even assuming that a sufficiently large number of repetitions of the experiment is enough and assuming to work in the utopian scenario in each of these, such an approach cannot mathematically tell us which intervals contain the true value. Furthermore, the above definition of confidence interval conflicts with the abandonment of statistical significance thresholds. To solve these dilemmas, Rafi and Greenland (2020) propose a terminological modification: give up the term "confidence" in favor of the Fisherian term "compatibility". In this framework, considering the utopian scenario, a compatibility interval contains all the target assumption predictions that, compared to certain threshold hypotheses and according to the performed test, are more compatible with the calculated experimental result (e.g., the difference between two sample estimators or population parameters). In other words, any model prediction that lies

inside (resp. outside) the obtained compatibility interval will result in a P-value higher (resp. lower) than the selected threshold. In order to address the problem of the arbitrary threshold choice, it has been proposed to provide tables that relate various P-values to their respective compatibility intervals or to present multiple compatibility intervals (e.g., 50%, 75%, 90%). A particularly interesting and information-rich solution is to graphically represent all compatibility intervals from 1% to 99% (Rafi and Greenland, 2020). However, this may greatly increase the reading load or even be confusing in the case of multiple results. Besides, there is currently a clear asymmetry in the definition and application of the concepts of compatibility interval and statistical surprise (S-value) since the former remains confined by definition within the scope of statistical significance (P-value). For these reasons, the present manuscript proposes and discusses two points: 1) the concept of surprisal interval, which can address the issues related to the obscured relationship between compatibility interval and surprise, and 2) a novel convention to compress information on the relationship between P/S-value and compatibility/surprisal intervals (based on the work of Xie and Singh (2013) concerning confidence distribution) and allows the presentation of these results in a single compact form.

3. SURPRISAL-BASED APPROACH

3.1. Surprisal interval

The definition of surprisal interval is based on a specific partition of the probability density function. Specifically, it consists of associating the natural values $S = 1, 2, ..., n$ with their respective areas by exploiting the relationship $P = 2^{-S}$. The first ten results are shown in Table 1. Through this operation, it is now possible to define surprisal intervals (S-I), analogous to confidence intervals (CI). We consider the case of a normal distribution (Figure 1). Let's assume we want to find a 4-I ($S = 4$). The corresponding exact P-value is $P = 2^{-4} = 0.0625 \sim 0.063$. Therefore, the corresponding compatibility interval is $(1 - 0.063) \cdot 100\%$ CI $= 93.7\%$ CI. To calculate it in practice, we need to ask ourselves: what is the value of z for which the area under the Gaussian curve between $-z$ and z is equal to 0.937 (i.e., 93.7% of the total unit area)? The answer is reported in Table 1.

Table 1: Association between surprisal (S-I) and compatibility intervals (CI).

S-value	P-value	z-value***	$100 \cdot (1 - P)$% CI
1	0.500	0.67	50%
2	0.250	1.15	75%
3	0.125	1.53	87.5%
4	0.063	1.86	93.7%
5	0.031	2.16	96.9%
6	0.016	2.42	98.4%
7	0.008	2.66	99.2%
8	0.004	2.89	99.6%
9	0.002	3.10	99.8%
10	0.001	3.30	99.9%

*** The shown z-values are valid only for the Gaussian distribution; conversely, the relationships between S-values, P-values, and compatibility intervals are general.

Afterward, it is sufficient to calculate 93.7% CI=$(r - z \cdot \bar{\sigma}, r + z \cdot \bar{\sigma}) = (r - 1.86 \cdot \bar{\sigma}, r + 1.86 \cdot \bar{\sigma})$, where r is the calculated experimental result and $\bar{\sigma}$ is the standard error.

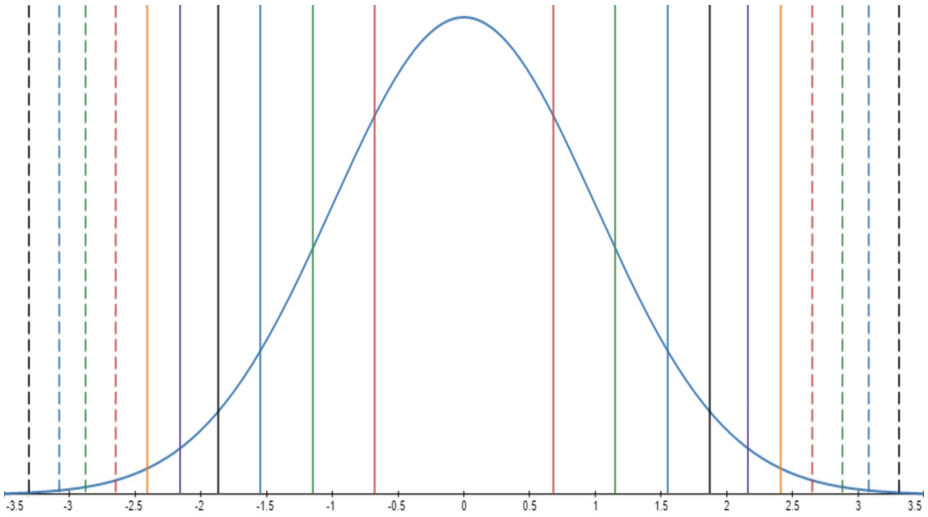

Figure 1: Surprisal intervals for integer values of S from 1 to 10.

The interpretation of our 4-I is as follows: in the utopian scenario, all target assumption predictions that lie inside (resp. outside) the 4-interval are less (resp. more) surprising than getting 4 consecutive heads — when flipping a fair coin 4 times — compared to the calculated experimental result according to the statistical test. In other words, let's suppose we choose a specific statistical test and consider a target assumption predicting an effect h. Let's also suppose we calculate an experimental result r (e.g., the difference between two population mean values) in the utopian scenario. The 4-I contains all and only the values h such that $r - h$ (i.e., the difference between r and h) is less surprising, according to the chosen test, than 4 consecutive heads when tossing a fair coin. Therefore, a general definition of surprisal interval is as follows:

Definition 3.1. If and only if all the background assumptions are true, a surprisal interval (or S-interval) is the interval that contains all and only the target assumption predictions that are less surprising than S consecutive heads — when tossing a fair coin S times — compared to the calculated experimental result according to the statistical test.[2]

Let's apply this new definition to evaluate a two-tailed one-sample t-test for a sample mean value of $\bar{x} = 10$, with a standard error $\bar{\sigma} = 5$ and a population mean value $\mu = 0$ (such that $r = \bar{x} - \mu = 10 - 0 = 10$). For simplicity, we also assume that the degrees of freedom are greater than 30 (such that $t \sim z$) and all the background assumptions are sufficiently met. Let's then calculate the following S-intervals: 4-I, 5-I, and 6-I. According to Table 1, $S = 4$ implies $t = 1.86$ (this happens because, in this specific example, $t \sim z$). So, we have 4-I $= (r - t \cdot \bar{\sigma}, r + t \cdot \bar{\sigma}) = (10 - 1.86 \cdot 5, 10 + 1.86 \cdot 5) = (1, 19)$. Similarly, we obtain 5-I $= (-1, 21)$ and 6-I $= (-2, 22)$. It is now easy to observe that the difference between $r = 10$ and the null

[2]Here, "surprise" does not refer to the probability that a hypothesis is true but to the surprise elicited by comparing the hypothesis with the observed result under the assumptions, in analogy with the surprise elicited when the hypothesis "the coin is unbiased" is confronted with the observation of S heads in S tosses.

hypothesis prediction ($h = 0$) is more surprising than 4 consecutive heads and less surprising than 5. In fact, the 4-I has a lower bound equal to $h = 1$ while the 5-I has a lower bound equal to $h = -1$ (hence, $h = 0$ must be somewhere in the middle). Considering the 4-I, we can also observe that $r - h = 10 - h$ is less surprising than 4 consecutive heads for all $h \in (1, 19)$ and more surprising than 4 consecutive heads for all $h < 1$ or $h > 19$.[3] To provide a compact overview of the variation in the width of our S-intervals as a function of the corresponding S-values, we can adopt the notation $5|4\text{-I} = (-1, 21|1, 19)$, which can be easily extended to three or more surprisal intervals depending on the needs of the authors and stakeholders (e.g., $6|5|4\text{-I} = (-2, 22| - 1, 21|1, 19)$). Thus, in general, we can define the convention as follows:

Definition 3.2. An n-tuple of surprisal intervals can be expressed as $S_1|\cdots|S_n\text{-I} = (S_1\text{-I}|\cdots|S_n\text{-I})$.

Each S_i represents a specific S-value and each S_i-I represents the corresponding surprisal interval. As an additional convention, it could be suggested to report at least three surprisal intervals: the narrowest S-interval containing the prediction of the null hypothesis, the narrowest S-interval not containing the prediction of the null hypothesis, and the 4-interval. The first two serve to locate the null hypothesis prediction, while the third serves as a general reference for comparison with other surprisal intervals (as it covers about 94% of the area, similarly to the classic 95% CI). In the previous example, the binary formulation $5|4\text{-I} = (-1, 21|1, 19)$ was sufficient.

3.2. Practical advantages of surprisal intervals

In general, the concept of surprisal as a measure of statistical surprise is absolutely unnecessary from a purely mathematical, statistical, or computational point of view. As a matter of fact, a correct use of P-value and compatibility intervals could compensate for any criticality exposed in the introductory section (although this would be much longer, subtle, and uneasy to present). However, the world of hard sciences is forced to confront a very different reality linked to the psychology and perception of the scientists who adopt and develop them. For instance, Rafi and Greenland (2020) argue that the misuse of statistical significance is primarily a cognitive and semantic problem rather than a statistical issue (Biau *et al.*, 2010). Greenland *et al.* (2016) also suggest that the inability to find a straightforward interpretation of the concept of P-value paradoxically favors the proliferation of oversimplified explanations. The same author of this paper has found, during his experience as an editor and peer reviewer in public health-related topics, not only widespread poor knowledge about the difference between the Fisherian and Neyman–Pearson approaches but also a furious resistance to change despite the overwhelming evidence provided. The authors' motivations ranged from "We don't want to make reading complicated" to "We prefer to maintain the traditional use of significance". This scenario is strongly consistent with the concerns raised by internationally renowned statisticians as well as the official statements of the American Statistical Association (Wasserstein and Lazar, 2016). In 2014, Professor George Cobb openly denounced the pivotal role of academic journals and universities in the unwarranted success of $P = 0.05$.

[3] It must be clear that this scenario is valid for the chosen test; performing a different test could substantially alter these outcomes.

Such illogical behaviors are compatible with some phenomena of cognitive psychology whereby modifying a consolidated belief or behavior is highly complex and often temporary when successful (even in scientists) (Swire-Thompson *et al.*, 2023; Rovetta and Castaldo, 2022). As early as 1919, Boring emphasized the limitations of the mathematical approach in modeling scientific reality and stressed the impossibility of formulating conclusions based solely on statistical approaches (Boring, 1919). The fact that, over 100 years later, despite the knowledge accumulated over this period, such concepts elude a significant fraction of the scientific community is signaling more than a formative problem. Accordingly, McShane and Gal observed that dichotomization decreases (but does not eliminate) when researchers are prompted to make decisions based on the evidence, especially if the outcome has personal consequences (McShane and Gal, 2016). Still, that's not all: diabolical academic dynamics such as publish or perish and publication bias push authors to exploit fallacious interpretations of P-values and compatibility intervals to voluntarily exaggerate the apparent degree of evidence found in their studies, thus increasing their chances of being published and cited (Friese and Frankenbach, 2020). Based on this, the purely interpretative aspect of a statistical measure can have very important practical consequences, especially in sectors — such as public health — where errors and overstatements must be weighed on the cost function for stakeholders. In particular, the main objective of this approach is to complete the proposal for replacing P-values with S-values by also requiring the replacement of "confidence" intervals with surprisal intervals. The total abandonment of statistical significance also brings with it the abandonment of all incorrect practices related to erroneous familiarity (e.g., judging a result as non-significant when $P < \alpha$ or is included in the $100(1-\alpha)\%$ CI, or considering $\alpha = 0.05$ and 95% CIs as some sort of privileged options) and prevents such dichotomies and prejudices at the root. The same term "significance" is inevitably and intrinsically replaced with the term "surprise", thus avoiding unnecessary, dangerous, as well as frequent confusion with practical significance (effect size) or even clinical significance. In order to give the reader an idea of the proportion of these errors in the medical field, a previous study found that only one out of 52 students was able to properly distinguish these concepts (Kühberger *et al.*, 2015).

In addition to this, surprisal intervals make the relationship with the measure of surprise of the outcome much clearer and more direct than compatibility intervals do with P-values. First, in addition to terminological consistency, the presentation of results is based on the same statistical quantity, namely, the integer number of consecutive heads when tossing an unbiased coin (bits) rather than a decimal measure of statistical compatibility and a percentage area. Second, the relationship between different intervals is much more intuitive since S-values linearize the behavior of the distribution. For example, an inexperienced user, as often happens, may easily think that the 99|95|91-%CI situation is symmetrical with respect to the central interval when, in fact, this includes very different compatibility requests (since they correspond respectively to 7, 4, and 3 consecutive heads in as many tosses); this cannot happen if the 7|4|3-I notation is used. Third, instead of setting arbitrary thresholds, the user can decide which intervals to show without exceeding confidence in the result. For example, if $S = 8$ is obtained, it may be useful to show an 8-I in order to understand what is the range of least surprise associated. This solution is highly advantageous because it allows the presentation of the result surprisal and the associated surprisal interval in a single compact form. This also allows for simplifying, both conceptually and operationally, procedures such as adjustment for multicomparison since S-values and S-intervals are no longer separable. Furthermore, while compatibility intervals consent to choose very specific degrees of precision (e.g., 94% or 95% or 96%), in the case of S-intervals the degree of precision is forced to be 1 bit. Ergo, the user is led to evaluate the results in less clear-cut terms (e.g., about 8 consecutive heads).

4. CONCRETE APPLICATION EXAMPLES

4.1. Example 1

The design of this study is intentionally non-optimal for didactic purposes. The aim is to show a proper application of the S-interval concept as well as the potential and limitations of the statistical approach. Let's suppose we have developed a long-term treatment to reduce blood pressure in hypertensive individuals. We convince 10 patients with clinically similar conditions to adopt this treatment for 3 months. We measure blood pressure levels before and after the treatment, obtaining the data in Table 2.

Table 2: Hypothetical blood pressure data before and after treatment: case 1.

Patient	Before (mmHg)	After (mmHg)	Differences (mmHg)
1	140	132	−8
2	150	145	−5
3	130	128	−2
4	135	139	+4
5	145	140	−5
6	138	132	−6
7	142	143	+1
8	128	125	−3
9	152	148	−4
10	134	130	−4

Since we are searching for a reduction, we decide to apply a one-sided one-sample t-test. To do so, in addition to assuming that all experimental procedures have been executed correctly (including random sampling), we need to check the compatibility of the data with the following assumptions of the test: i) normal distribution (including the absence of outliers), ii) independence of observations, iii) interval or ratio data. Since the data (column "Differences") represents continuous real number values of blood pressure from independent patients, we can reasonably consider assumptions ii) and iii) to be validated. To investigate the compatibility of the data with the assumption of normality, we observe that the data reasonably follows the Q-Q plot line (readers can easily verify this independently). So, we can apply the one-sample t-test with reasonable confidence in its interpretability. The mean value is $\bar{x} = -3.2$ (SD 3.5). By choosing the null one-sided assumption $h \geq 0$, the largest experimental result is $r = \bar{x} - \min\{h\} = -3.2$ and the associated test result is $t_9 = -2.9$, which implies $S = 6.8$ (i.e., the test result is as surprising as just under 7 consecutive heads in 7 tosses under the model). Can we conclude anything? The quick answer is no. Indeed, we have no idea how the degree of surprise of our result varies compared to other hypotheses. To remedy this, we construct the following S-intervals: 6|5|4|3-I. The goal is to understand the "rapidity" at which statistical surprise diminishes to less surprising levels. We obtain 6|5|4|3-I $= (-\infty, -0.4| -\infty, -0.8| -\infty, -1.3| -\infty, -1.8)$. In practice, we lose 1 bit (head) for every 0.5 mmHg, which implies that even hypotheses involving very small benefits (such as −1.3 mmHg) are only weakly surprising relative to the observed result, under the model. This indicates that our results are highly unstable. Therefore, we cannot conclude anything other than "these results are too uncertain to properly inform a scientific conclusion".

4.2. Example 2

The scenario is supposed to be the same as the previous example, but in this case, we refer to the data in Table 3. Let's take all the statistical and non-statistical background assumptions for granted (the statistical ones can be easily investigated, as shown in the previous example).

Table 3: Hypothetical blood pressure data before and after treatment: case 2.

Patient	Before (mmHg)	After (mmHg)	Difference (mmHg)
1	140	127	−13
2	150	140	−10
3	130	123	−7
4	135	136	+1
5	145	135	−10
6	138	141	+3
7	142	138	−4
8	128	120	−8
9	152	143	−9
10	134	125	−9

In this case, we have an average value $\bar{x} = -6.6$ (SD 5.1). By choosing the null one-sided assumption $h \geq 0$, the largest experimental result is $r = \bar{x} - \min\{h\} = -6.6$ and the associated test result is $t_9 = -4.1$, which implies $S = 9.5$ (i.e., the test result is more surprising than 9 consecutive heads in 9 tosses under the model). To assess the surprise decrease rapidity against different target assumptions, we construct the following S-intervals: $9|7|5|3\text{-I} = (-\infty, -0.4| - \infty, -1.8| - \infty, -3.2| - \infty, -4.6)$. In this case, the decrease could be acceptable. Thus, can we say we have proven the effectiveness of the treatment? Absolutely not. As mentioned earlier, statistics is a limited component of scientific inquiry. So, can we at least say we have found evidence in favor of the treatment's effectiveness? No. At best, we have found evidence compatible with the effectiveness of the treatment. However, this evidence is also compatible with other equally valid hypotheses. For example, the absence of a control group prevents us from establishing the impact of atmospheric variations and changes in the patients' physical activity and dietary habits over these 3 months (e.g., with the onset of summer, patients might spend more time outdoors and be inclined toward a more Mediterranean diet). Bias analysis is extremely important in this regard (Lash *et al.*, 2014). Alongside this, the sample is arguably too small to be representative of the entire population. At the ethical and scientific level, we must assess the invasiveness of the therapy. For instance, does the effect size justify any potential physical and/or psychological adverse events? Not only that, statistics deals with numbers, i.e., it is unable to encompass the clinical complexity of each individual patient. Indeed, patients 4 and 6 even recorded an increase in blood pressure that should be investigated clinically. Furthermore, the dataset exhibited high variability (percentage variation coefficient=77%). Nonetheless, admitting that there are valid biochemical reasons to suspect the effectiveness of the treatment, in the event that the latter has not yielded negative consequences for the patients, these results could justify further research.

5. DISCUSSION

The adoption of surprisal intervals completes the evaluative approach of statistical surprise, avoiding any reliance on statistical significance and confidence (topics that are much more complex and cryptic even for expert statisticians). S-intervals finally make explicit the relationship between surprise and effect size and, in light of the uncertainties that affect the testing of a statistical hypothesis, prevent the adoption of excessively sharp and senseless statistical significance thresholds. This interpretation is reinforced by the definition of S-intervals, which only permits reporting intervals at least 1 bit apart. For this reason, the conventions and methodologies suggested in this paper never allow for the statistical rejection or acceptance of a single target hypothesis since the researcher is urged to reason only in terms of greater or lesser surprise (also in relation to effect size estimation intervals). Indeed, any concrete action of this type (e.g., to promote a drug for commercialization) must be made solely based on a careful evaluation of the quality of the evidence available and a detailed analysis of biases, costs, and benefits for stakeholders since no mere statistical criterion can ever automatically demonstrate causation nor answer the question, "Is it worth it?" The final decision must, therefore, be informed by the union of evidence of various kinds (e.g., statistical tests, proven chemical-biological mechanisms, clinical reports, etc.). Such scientific practice, known as decision analysis, is central for public health (Greenland *et al.*, 2016; Rovetta, 2023; Lash *et al.*, 2014; Greenland, 2021). In addition to this, the compact formulation of multiple intervals can provide a much more complete and clearer overview than that described by a traditional confidence/compatibility interval without excessively burdening the reading, i.e., remaining suitable to be used in summary sections such as the abstract. Although the problems related to statistical testing are numerous and go beyond the scope of this manuscript (e.g., arbitrary multiple comparisons adjustments, p-hacking, statistical power misconceptions, and publication bias), the interpretation of test results is fundamental or integral to each of these (Greenland, 2017; Greenland *et al.*, 2016; Rovetta, 2023; Biau *et al.*, 2010; Rafi and Greenland, 2020; Greenland, 2023b; Kühberger *et al.*, 2015; Friese and Frankenbach, 2020; Lash *et al.*, 2014; Greenland, 2021; Lakens, 2021; Greenland *et al.*, 2022; Greenland, 2023a; McShane *et al.*, 2023; Gelman and Stern, 2006; Wasserstein and Lazar, 2016; Boring, 1919; McShane and Gal, 2016; Amrhein *et al.*, 2019). Surprisal intervals, in conjunction with surprisals, can provide great assistance to the scientific community in framing research problems, especially in the field of public health where errors regarding statistical significance are as frequent as they are dangerous. In fact, comparing test results to a perceptually familiar phenomenon, such as the number of consecutive successes (heads) when flipping an unbiased coin, not only greatly simplifies the evaluation of the statistical weight of the event under consideration but also contributes to avoiding overstatements. Consequently, it is highly recommended that surprisal intervals be adopted in future scientific investigations based on statistical testing.

REFERENCES

Amrhein, V., Greenland, S., and McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748):305–307.

Biau, D.J., Jolles, B.M., and Porcher, R. (2010). P value and the theory of hypothesis testing: an explanation for new researchers. *Clinical Orthopaedics and Related Research*, 468(3):885–892.

Boring, E.G. (1919). Mathematical vs. scientific significance. *Psychological Bulletin*, 16(10):335–338.

Friese, M. and Frankenbach, J. (2020). P-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, 25(4):456–471.

Gelman, A. and Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, 60(4):328–331.

Greenland, S. (2017). Invited commentary: the need for cognitive science in methodology. *American Journal of Epidemiology*, 186(6):639–645.

Greenland, S. (2021). Analysis goals, error-cost sensitivity, and analysis hacking: essential considerations in hypothesis testing and multiple comparisons. *Paediatric and Perinatal Epidemiology*, 35(1):8–23.

Greenland, S. (2023a). Connecting simple and precise P-values to complex and ambiguous realities (includes rejoinder to comments on "Divergence vs. decision P-values". *Scandinavian Journal of Statistics*, 50(3):899–914.

Greenland, S. (2023b). Divergence versus decision P-values: a distinction worth making in theory and keeping in practice: or, how divergence P-values measure evidence even when decision P-values do not. *Scandinavian Journal of Statistics*, 50(1):54–88.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., and Altman, D.G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31(4):337–350.

Greenland, S., Mansournia, M.A., and Joffe, M. (2022). To curb research misreporting, replace significance and confidence by compatibility: a preventive medicine golden jubilee article. *Preventive Medicine*, 164:107127.

Kühberger, A., Fritz, A., Lermer, E., and Scherndl, T. (2015). The significance fallacy in inferential statistics. *BMC Research Notes*, 8:54–88.

Lakens, D. (2021). The practical alternative to the P value is the correctly used P value. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 16(3):639–648.

Lash, T.L., Fox, M.P., MacLehose, R.F., Maldonado, G., McCandless, L.C., and Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology*, 43(6):1969–1985.

McShane, B.B., Bradlow, E.T., Lynch, J.G., and Meyer, R.J. (2023). EXPRESS: "Statistical significance" and statistical reporting: moving beyond binary. *Journal of Marketing*, 0:ja.

McShane, B.B. and Gal, D. (2016). Blinding us to the obvious? The effect of statistical training on the evaluation of evidence. *Management Science*, 62(6):1707–1718.

Rafi, Z. and Greenland, S. (2020). Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Medical Research Methodology*, 20(1):244.

Rovetta, A. (2023). Common statistical errors in scientific investigations: a simple guide to avoid unfounded decisions. *Cureus*, 15(1):e33351.

Rovetta, A. and Castaldo, L. (2022). Are we sure we fully understand what an infodemic is? A global perspective on infodemiological problems. *JMIRx Med*, 3(3):e36510.

Swire-Thompson, B., Dobbs, M., Thomas, A., and DeGutis, J. (2023). Memory failure predicts belief regression after the correction of misinformation. *Cognition*, 230:105276.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter – a review (with discussion). *International Statistical Review*, 81:3–39.

Wasserstein, R.L. and Lazar, N.A. (2016). The ASA's statement on P-values: context, process, and purpose. *The American Statistician*, 70(2):129–133.

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 503–526
https://doi.org/10.57805/revstat.v23i4.668

A Refined Extreme Quantile Estimator for Weibull Tail-Distributions

Authors: Jonathan El Methni ✉
– Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
jonathan.el-methni@univ-grenoble-alpes.fr

Stéphane Girard
– Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
stephane.girard@inria.fr

Received: August 2023 Revised: January 2024 Accepted: January 2024

Abstract:

- We address the estimation of extreme quantiles of Weibull tail-distributions. Since such quantiles are asymptotically larger than the sample maximum, their estimation requires extrapolation methods. In the case of Weibull tail-distributions, classical extreme-value estimators are numerically outperformed by estimators dedicated to this set of light-tailed distributions. The latter estimators are based on two key quantities: an order statistic to estimate an intermediate quantile and an estimator of the Weibull tail-coefficient used to extrapolate. The common practice is to select the same intermediate sequence for both estimators. We show how an adapted choice of two different intermediate sequences leads to a reduction of the asymptotic bias associated with the resulting refined estimator. This analysis is supported by an asymptotic normality result associated with the refined estimator. A data-driven method is introduced for the practical selection of the intermediate sequences and our approach is compared to three estimators of extreme quantiles on simulated data. An illustration on a real data set of daily wind measures is also provided.

Keywords:

- *extreme quantile; bias reduction; Weibull tail-distribution; extreme-value statistics; asymptotic normality.*

AMS Subject Classification:

- 60G70, 62G32, 62G20.

✉ Corresponding author

1. INTRODUCTION

Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables with cumulative distribution function F and let $X_{1,n} \leq \cdots \leq X_{n,n}$ denote the associated order statistics. We consider the case where F belongs to the family of Weibull tail-distributions (Broniatowski, 1993):

(**A.1**) F is twice differentiable and $F(\cdot) = 1 - \exp(-H(\cdot))$ with $V(t) := H^{\leftarrow}(t) = t^{\theta}\ell(t)$ for all $t > 0$, where $\theta > 0$ is called the Weibull tail-coefficient and where ℓ is a (positive) slowly-varying function, i.e. $\ell(tx)/\ell(x) \to 1$ as $x \to \infty$ for all $t > 0$.

Here, and in the following, $\Phi^{\leftarrow}(\cdot) = \inf\{x \in \mathbb{R}, \Phi(x) > \cdot\}$ denotes the generalized inverse of an increasing function Φ. The inverse cumulative hazard function V is said to be regularly-varying at infinity with index θ and this property is denoted by $V \in \mathcal{RV}_{\theta}$ (see Bingham *et al.*, 1989, for a detailed account on this topic). The shape parameter θ is referred to as the Weibull tail-coefficient. Weibull tail-distributions are part of the Gumbel maximum domain of attraction, i.e. with extreme-value index $\gamma = 0$ (see Gardes *et al.*, 2011, Proposition 2(ii)), and as such, are light-tailed distributions. They include for instance exponential ($\theta = 1$), Gamma ($\theta = 1$), logistic ($\theta = 1$), Normal ($\theta = 1/2$) and Weibull distributions (θ is the inverse of the shape parameter) (see Girard, 2004, Table 1). We refer to Beirlant and Teugels (1992) for an application to the modeling of large claims in non-life insurance and to Vladimirova *et al.* (2020) for an analysis of neural networks distributional properties.

Dedicated methods have been proposed to estimate the Weibull tail-coefficient θ since the relevant information is localised in the extreme upper part of the sample. Most approaches rely on the k_n upper order statistics $X_{n-k_n+1,n}, ..., X_{n,n}$ where $k_n \to \infty$ as $n \to \infty$. Note that, since θ is defined through a tail behavior, the associated estimator should only use the extreme-values of the sample and thus the extra condition $k_n/n \to 0$ is required. More specifically, recent estimators are based on the log-spacings between the k_n upper order statistics (Beirlant *et al.*, 1996; Gardes and Girard, 2006; Gardes *et al.*, 2011; Girard, 2004) or on the mean excess function (Beirlant *et al.*, 2006, 1995, 2009). The introduction of kernel based weights has been investigated for both approaches (see Gardes and Girard, 2008, for the log-spacings case, and Goegebeur and Guillou, 2011, for the mean excess function framework). A bias reduction method adapted to the estimation of the Weibull tail-coefficient is proposed in Diebolt *et al.* (2008) and the adaptation to random censoring is achieved in Worms and Worms (2019).

We address the problem of estimating extreme quantiles of Weibull tail-distributions. Recall that an extreme quantile $q(\alpha_n)$ of order α_n is defined by $q(\alpha_n) = F^{\leftarrow}(1 - \alpha_n)$ with $n\alpha_n \to 0$ as $n \to \infty$. The latter condition implies that $q(\alpha_n)$ is almost surely asymptotically larger than $X_{n,n}$, the sample maximum. It is shown in Gardes and Girard (2005) that classical extreme-value estimators of such large quantiles are numerically outperformed by estimators dedicated to Weibull tail-distributions (Diebolt *et al.*, 2008); see also Lemma A.1 in the Appendix for a theoretical argument. The latter methods estimate $q(\alpha_n)$ by combining two ingredients: an order statistic $X_{n-k_n+1,n}$ and an estimator of the Weibull tail-coefficient θ used to extrapolate from this anchor point.

In this work, we show that the biases associated with the previous extrapolation method and the estimator of θ may asymptotically cancel out in the extreme quantile estimator thanks

to an appropriate tuning of the number of upper order statistics involved in the Weibull tail-coefficient estimator. The construction of the resulting estimator is presented in Section 2 and an asymptotic normality result is provided, emphasizing that the proposed extreme quantile estimator is asymptotically less biased than the original one (Gardes and Girard, 2005). Its performances are illustrated on simulated data in Section 3 and compared to three state-of-the-art competitors (Beirlant *et al.*, 2009; Diebolt *et al.*, 2008; Gardes and Girard, 2005). An illustration on a real data set of daily wind measures is provided in Section 4. Finally, a small conclusion is proposed in Section 5 and the proofs are postponed to the Appendix.

2. A REFINED ESTIMATOR OF THE EXTREME QUANTILE

2.1. Extreme quantile estimators

Weibull-tail estimators of the extreme quantile $q(\alpha_n)$ rely on an intermediate quantile $q(k_n/n)$ where (k_n) is an intermediate sequence of integers, i.e. such that $k_n \in \{1, ..., n-1\}$, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$ (see for instance Diebolt *et al.*, 2008; Gardes and Girard, 2005). Indeed, in view of (**A.1**), one has

$$\frac{q(\alpha_n)}{q(k_n/n)} = \frac{V(\log(1/\alpha_n))}{V(\log(n/k_n))} \simeq \left(\frac{\log(1/\alpha_n)}{\log(n/k_n)}\right)^{\theta} =: \tau_n^{\theta}, \tag{2.1}$$

as $n \to \infty$, where $\tau_n = \log(1/\alpha_n)/\log(n/k_n)$ is the (logarithmic) extrapolation factor. From an intuitive point of view, an extreme quantile can thus be approximated by multiplying an intermediate quantile by an appropriate extrapolation term: $q(\alpha_n) \simeq q(k_n/n)\tau_n^{\theta}$. The intermediate quantile $q(k_n/n)$ can then be estimated by its empirical counterpart $X_{n-k_n+1,n}$ while the extrapolation term depends on the tail heaviness through θ which has to be estimated as well. Following the ideas of Allouche *et al.* (2023), we propose a refined Weissman (Weissman, 1978) type estimator:

$$\hat{q}_n(\alpha_n, k_n, k_n') = X_{n-k_n+1,n}\left(\frac{\log(1/\alpha_n)}{\log(n/k_n)}\right)^{\hat{\theta}_n(k_n')} = X_{n-k_n+1,n}\,\tau_n^{\hat{\theta}_n(k_n')}, \tag{2.2}$$

with $\hat{\theta}_n(k_n')$ an estimator of θ depending on another intermediate sequence (k_n'). Let us focus on the estimator introduced in Gardes and Girard (2006):

$$\hat{\theta}_n^{\mathrm{RSH}}(k_n') = \frac{1}{\mu(\log(n/k_n'))}\frac{1}{k_n'}\sum_{i=1}^{k_n'}(\log X_{n-i+1,n} - \log X_{n-k_n'+1,n}), \tag{2.3}$$

with, for $t > 0$, $\mu(t) = \int_0^{+\infty} \log\left(1 + \frac{x}{t}\right)e^{-x}dx = e^t E_1(t)$, where E_1 is the Exponential integral function (Abramowitz and Stegun, 1964, p. 228). This estimator is motivated by the remark that, in view of (2.1), the log-spacings between two quantiles are approximately proportional to θ. This property is also used in the real data application (see the top-right panel of Figure 4) to visually check the Weibull-tail assumption. Clearly, $\hat{\theta}_n^{\mathrm{RSH}}(\cdot)$ can be interpreted as a rescaled Hill estimator since

$$\hat{\theta}_n^{\mathrm{RSH}}(k_n') = \frac{\hat{\gamma}_n^{\mathrm{H}}(k_n')}{\mu(\log(n/k_n'))}, \tag{2.4}$$

where $\hat{\gamma}_n^{\mathrm{H}}(\cdot)$ is the well-known Hill estimator (Hill, 1975) of the extreme-value index γ.

Let us note, when $k'_n = k_n$, one recovers the extreme quantile estimator for Weibull tail-distributions introduced in Gardes and Girard (2005). In the next paragraph, the asymptotic normality of $\hat{q}_n(\alpha_n, k_n, k'_n)$ is established, and it is shown that choosing $k'_n \neq k_n$ can yield better results from an asymptotic point of view. A similar phenomenon occurs in the estimation of the endpoint of a distribution in the Weibull maximum domain of attraction, see Aarssen and de Haan (1994) for details. We also refer to Allouche *et al.* (2023) for the estimation of the tail-index in the Fréchet maximum domain of attraction.

2.2. Asymptotic analysis

The study of the limit distribution of $\hat{q}_n(\alpha_n, k_n, k'_n)$ requires a second-order condition on the slowly-varying function ℓ introduced in (**A.1**):

(**A.2**) There exist $\rho \leq 0$ and $b(t) \to 0$ as $t \to \infty$, with ultimately constant sign, such that uniformly locally on $x \geq 1$,

$$\lim_{t\to\infty} \frac{1}{b(t)} \log\left(\frac{\ell(tx)}{\ell(t)}\right) = K_\rho(x) := \int_1^x u^{\rho-1} du.$$

It can be shown that necessarily $|b| \in \mathcal{RV}_\rho$. The second-order Weibull parameter $\rho \leq 0$ tunes the rate of convergence of the ratio $\ell(tx)/\ell(t)$ to 1. The closer ρ is to 0, the slower is the convergence. Condition **(A.2)** is the cornerstone in all proofs of asymptotic normality for extreme-value estimators. Again, we refer to (Girard, 2004, Table 1) for ρ parameters associated with usual Weibull tail-distributions. Our first result is a refinement of (Gardes and Girard, 2006, Corollary 3.1). It provides an asymptotic normality result for the extreme quantile estimator (2.2) based on two intermediate sequences (k_n) and (k'_n).

Theorem 2.1. *Assume* **(A.1)** *and* **(A.2)** *hold. Let* (k_n) *and* (k'_n) *be two intermediate sequences and introduce* (α_n) *a probability sequence such that* $\alpha_n \to 0$ *as* $n \to \infty$. *Suppose, as* $n \to \infty$,

(**i**) $\sqrt{k'_n} b(\log(n/k'_n)) \to \lambda \in \mathbb{R}$,

(**ii**) $\log(n/k'_n)/\log(n/k_n) \to \beta \geq 1$,

(**iii**) $\tau_n \to \tau > \beta$.

Then, as $n \to \infty$,

$$\sqrt{k'_n}\left(\frac{\hat{q}_n(\alpha_n, k_n, k'_n)}{q(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}\big(\lambda(\log(\tau) - \beta^{-\rho} K_\rho(\tau)), (\theta \log \tau)^2\big). \tag{2.5}$$

Let us first remark that condition (i) implies $\log(n/k'_n) \sim \log(n)$ as $n \to \infty$ (see Gardes and Girard, 2006, Lemma 5.1), then condition (ii) yields $\log(n/k_n) \sim \log(n)/\beta$ and therefore condition (iii) can be rewritten as $\log(1/\alpha_n) \sim (\tau/\beta)\log(n)$ as $n \to \infty$. As a consequence, the condition $\tau > \beta$ in (iii) implies $n\alpha_n \to 0$ as $n \to \infty$ which, in turns, implies that $q(\alpha_n)$ is an extreme quantile.

It follows from (2.5) that the asymptotic bias associated with $\hat{q}_n(\alpha_n, k_n, k'_n)$ is given by

$$\begin{aligned}\big(\log\tau - \beta^{-\rho}K_\rho(\tau)\big)b(\log(n/k'_n)) &\sim (\beta^\rho\log(\tau) - K_\rho(\tau))b(\log(n/k_n))\\ &=: B(\beta,\tau,\rho)b(\log(n/k_n)),\end{aligned}$$

since $|b| \in \mathcal{RV}_\rho$. It appears that each choice of k'_n yields an associated constant β in (ii) and thus a corresponding bias factor $B(\beta,\tau,\rho) = \beta^\rho\log(\tau) - K_\rho(\tau)$. From the theoretical point of view, two cases can be considered:

- The usual choice $k'_n = k_n$ yields $\beta = 1$ and one recovers (Gardes and Girard, 2006, Corollary 3.1) as a particular case of Theorem 2.1. Moreover, for all $\tau > 1$, $\rho \le 0$,

$$B(1,\tau,\rho) = \log(\tau) - K_\rho(\tau) \ge 0, \tag{2.6}$$

which is the (positive) bias factor associated with the extreme quantile estimator $\hat{q}_n(\alpha_n, k_n, k_n)$ investigated in Gardes and Girard (2006). Note that $\rho \mapsto B(1,\tau,\rho)$ is a decreasing function such that $B(1,\tau,0) = 0$ which is an unusual situation in extreme-value theory. For instance, the bias factor associated with the Weissman estimator (Weissman, 1978) dedicated to heavy-tailed distributions is proportional to $1/(1-\rho)$ and increases with ρ (see de Haan and Ferreira, 2007, Theorem 3.2.5 and Theorem 4.3.8).

- The choice $\beta^\star(\tau,\rho) := (K_\rho(\tau)/\log(\tau))^{1/\rho}$ yields

$$B(\beta^\star(\tau,\rho),\tau,\rho) = 0. \tag{2.7}$$

The associated intermediate sequence is given by $k_n^\star(\tau,\rho) = \lfloor n(k_n/n)^{\beta^\star(\tau,\rho)}\rfloor$ and therefore the extreme quantile estimator $\hat{q}_n(\alpha_n, k_n, k_n^\star(\tau,\rho))$ is asymptotically unbiased. Note that this estimator cannot be used in practice since the second-order Weibull parameter ρ is unknown.

Up to our knowledge, there is no estimator of the second-order Weibull parameter in the statistical literature. In practice, one can replace the true unknown value of ρ by a misspecified value $y \le 0$ in the above $\beta^\star(\tau,\rho)$ leading to

$$\beta^\star(\tau,y) = (K_y(\tau)/\log(\tau))^{1/y}, \tag{2.8}$$

$$k_n^\star(\tau,y) = \lfloor n(k_n/n)^{\beta^\star(\tau,y)}\rfloor, \tag{2.9}$$

$$\begin{aligned}B(\beta^\star(\tau,y),\tau,\rho) &= \beta^\star(\tau,y)^\rho\log(\tau) - K_\rho(\tau)\\ &= (K_y(\tau)/\log(\tau))^{\rho/y}\log(\tau) - K_\rho(\tau),\end{aligned} \tag{2.10}$$

with $\rho \le 0$ and $\tau > 1$. This misspecification technique has been used both to deal with Pareto-type distributions ($\gamma > 0$) (see for instance Feuerverger and Hall, 1999) and Weibull tail-distributions ($\gamma = 0$) (Diebolt *et al.*, 2008). Some properties of the intermediate sequence $k_n^\star(\tau,y)$ are given in the next Lemma.

Lemma 2.1. *Let $\beta^\star(\tau,y)$ and $k_n^\star(\tau,y)$ be defined by (2.8) and (2.9) respectively. Then, for all $\tau > 1$:*

- **(i)** *$\beta^\star(\tau,y) \to 1$ as $y \to -\infty$ and $\beta^\star(\tau,\cdot)$ can be extended by continuity by setting $\beta^\star(\tau,0) := \sqrt{\tau}$.*
- **(ii)** *$1 < \beta^\star(\tau,y) < \tau$ for all $y \le 0$.*
- **(iii)** *For all $y \le 0$, $k_n^\star(\tau,y)$ is an increasing function of k_n, $k_n^\star(\tau,y) \le k_n$ and $k_n^\star(\tau,y)/k_n \to 0$ as $n \to \infty$.*
- **(iv)** *$k_n^\star(\tau,y)$ is a decreasing function of $y \in (-\infty, 0)$.*

In particular, it appears in (iii) that the number of upper order statistics $k_n^\star(\tau, y)$ used in the Weibull tail-coefficient estimator should be asymptotically small compared to k_n for all finite values of y. From (iv), this is all the more true as y is large. When $y \to -\infty$, meaning that one does not take into account the bias, (i) shows that $k_n^\star(-\infty, \tau) = k_n$ is recovered as a limit case. These properties are illustrated on the left panel of Figure 1, where $k_n^\star$ is drawn as a function of k_n for several values of y. The next Corollary shows that these choices indeed lead to a bias reduction in the estimation of the extreme quantile.

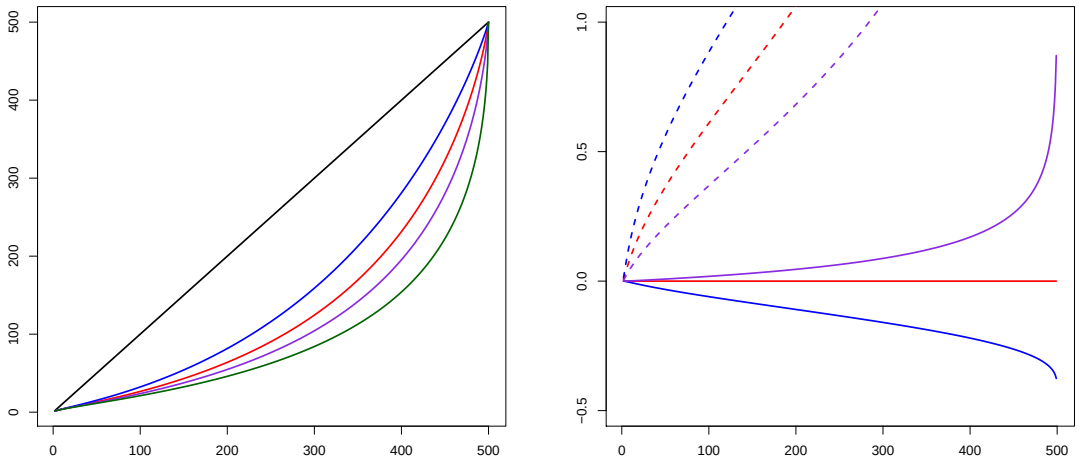

Figure 1: Left: Graphs of $k_n \in \{2, ..., 500\} \mapsto k_n^\star(\tau_n, y)$ for $y \in \{-\infty, -2, -1, -1/2, 0\}$ respectively in {black, blue, red, violet, green}. Right: graphs of $k_n \in \{2, ..., 500\} \mapsto B(1, \tau_n, \rho)$ (dotted lines) and $k_n \in \{2, ..., 500\} \mapsto B(\beta^\star(\tau_n, \rho^\# = -1), \tau_n, \rho)$ (solid lines) given in equations (2.6) and (2.10), with $\rho \in \{-2, -1, -1/2\}$ respectively in {blue, red, violet}. On both panels: $\tau_n = \log(1/\alpha_n)/\log(n/k_n)$ with $\alpha_n = 1/n$ and $n = 500$.

Corollary 2.1. *Assume* **(A.1)** *and* **(A.2)** *hold. Let* $c > 0$, $\tau > 1$, $y \leq 0$, $\lambda \neq 0$ *such that* $\lambda b(\cdot)$ *is ultimately positive, and* $\beta^\star(\tau, y)$ *be defined as in* (2.8). *Let* $\alpha_n = c\, n^{-\tau/\beta^\star(\tau,y)}$, $k_n = \lfloor n\{\lambda^2/(nb^2(\log n))\}^{1/\beta^\star(\tau,y)} \rfloor$ *and define* $k_n^\star(\tau, y)$ *as in* (2.9).

(**i**) *Then, as* $n \to \infty$,

$$\sqrt{k_n^\star(\tau, y)}\left(\frac{\hat{q}_n(\alpha_n, k_n, k_n^\star(\tau, y))}{q(\alpha_n)} - 1\right) \xrightarrow{d} \mathcal{N}\Big(\lambda(\log(\tau) - (K_y(\tau)/\log(\tau))^{-\rho/y} K_\rho(\tau)), (\theta \log \tau)^2\Big).$$

(**ii**) *Moreover, for all* $\tau > 1$, $y \leq 0$*:*

$$\begin{aligned} |B(\beta^\star(\tau, y), \tau, \rho)| &< B(1, \tau, \rho) \quad \text{for all } \rho < 0, \\ B(\beta^\star(\tau, y), \tau, 0) &= B(1, \tau, 0) = 0. \end{aligned}$$

Let us first highlight that $\sqrt{k_n^\star(\tau, y)} \sim \lambda/b(\log n)$ as $n \to \infty$ (see the proof of Corollary 2.1 in the Appendix) which is the (logarithmic) rate of convergence of usual extreme quantile estimators dedicated to Weibull tail-distributions (see for instance Diebolt *et al.*, 2008, Theorem 1). In contrast, the rate of convergence of extreme quantile estimators is a

power function of n in the Fréchet maximum domain of attraction (see de Haan and Ferreira, 2007, Theorem 4.3.8 and equation (3.2.10), for Weissman estimator, and Allouche *et al.*, 2023, Corollary 2, for the associated refined version). This may be seen as a consequence of Lemma A.1 in the Appendix where it is established that the second-order parameter associated with Weibull distributions is $\psi = 0$.

Surprisingly, as a consequence of Corollary 2.1(ii), the extreme quantile estimator $\hat{q}_n(\alpha_n, k_n, k_n^\star(\tau, y))$ computed with $k_n^\star(\tau, y)$ defined in (2.8) and (2.9) has a smaller asymptotic bias than the usual one $\hat{q}_n(\alpha_n, k_n, k_n)$ whatever the chosen value $y \leq 0$. Let us recall that, from (2.7), the theoretical best choice would be $y = \rho$. In practice, we use $y = \rho^\# = -1$ leading to $\beta^\star(\tau, -1) = \tau \log(\tau)/(\tau - 1)$. This "canonical" choice is also used in Diebolt *et al.* (2008) (see Section 3.2 hereafter). Let us stress that the use of a similar bias reduction method in the Fréchet maximum domain attraction (Allouche *et al.*, 2023) is not based on such a misspecification technique but requires the estimation of ρ.

Corollary 2.1(ii) is illustrated on the right panel of Figure 1 through the graphical comparison of the bias factors associated with $\beta = \beta^\star(\tau, -1)$ (refined Weibull-tail estimator) and $\beta = 1$ (usual Weibull-tail estimator, Gardes and Girard, 2006). It clearly appears that, from the theoretical point of view, the first choice yields smaller bias factors in absolute value than the second one. The performance of the refined estimator in practice is assessed on simulated data in the next Section.

3. VALIDATION ON SIMULATED DATA

The refined extreme quantile estimator is compared on simulated data to the original estimator (Gardes and Girard, 2005) and to two other competitors described hereafter.

3.1. Experimental design

Let us consider the class of $\mathcal{D}(\zeta, \eta, a)$-distributions which is an adaptation of Hall's class (Hall, 1982; Hall and Welsh, 1985) to the Weibull-tail framework. In this family, the inverse cumulative hazard function is defined for all $x > 0$ by

$$V(x) := x^{1/\zeta}\left(1 + \frac{a}{\eta}x^{-\eta}\right),$$

with $a, \zeta, \eta > 0$ and $\zeta\eta \leq 1$. Under these conditions, the above class of distributions fulfills assumptions (**A.1**) and (**A.2**) with Weibull tail-coefficient $\theta = 1/\zeta$, second-order Weibull parameter $\rho = -\eta$, slowly-varying function $\ell(x) = 1 + (a/\eta)x^{-\eta}$ and $b(x) = -ax^{-\eta}$. Unlike classical distributions such as the (absolute) Normal distribution $\mathcal{N}(\mu, \sigma)$ ($\theta = 1/2$, $\rho = -1$ and $b(x) = \log(x)/(4x)$), the Gamma distribution $\mathcal{G}(\upsilon \neq 1, \lambda)$ ($\theta = 1$, $\rho = -1$ and $b(x) = (1 - \upsilon)\log(x)/x$) and the Weibull distribution $\mathcal{W}(\upsilon, \lambda)$ ($\theta = 1/\upsilon$, $\rho = -\infty$ and $b(x) = 0$), it is thus possible to obtain $\mathcal{D}$-distributions with arbitrary Weibull tail-coefficient $\theta > 0$ and second-order Weibull parameter $\rho \in [-\theta, 0)$.

In the following, we set $\theta \in \{1/2, 3/4, ..., 5/2\}$, $\rho \in \{-5/2, -2, ..., -1/2\}$, $a = 10$ and focus on the only 25 situations of the $\mathcal{D}$-distribution where $\rho \geq -\theta$ to fulfill the constraint $\zeta\eta \leq 1$, see Table 1. We also consider 5 situations from the (absolute) Normal distribution $\mathcal{N}(\mu, \sigma = 1)$ with $\mu \in \{1, 3, 5, 7, 9\}$, 4 situations from the Gamma distribution $\mathcal{G}(\upsilon, \lambda = 1)$ with $\upsilon \in \{4, 6, 8, 10\}$ and 2 situations from the Weibull distribution $\mathcal{W}(\upsilon, \lambda)$ with $\upsilon = \lambda \in \{1/2, 2\}$. In each case, $N = 1,000$ replications of a data set of $n = 500$ i.i.d. realisations are simulated from the $25 + 5 + 4 + 2 = 36$ considered parametric models. Finally, the same two cases as in Diebolt *et al.* (2008) are investigated for the order of the extreme quantile: $\alpha_n \in \{1/n^2, 1/n^4\}$. Summarizing, this experimental design includes $36 \times 2 = 72$ configurations.

Table 1: All considered configurations for (ρ, θ). The letter stands for the distribution and the subscript for the number of investigated situations. As an example, $\mathcal{N}_5$ corresponds to the (absolute) Normal distribution $\mathcal{N}(\mu, \sigma = 1)$ where five cases are considered $\mu \in \{1, 3, 5, 7, 9\}$.

(ρ, θ)	1/2	3/4	1	5/4	3/2	7/4	2	9/4	5/2
$-\infty$	$\mathcal{W}_1$						$\mathcal{W}_1$		
$-5/2$									$\mathcal{D}_1$
-2							$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$
$-3/2$					$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$
-1	$\mathcal{N}_5$		$\mathcal{G}_4/\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$
$-1/2$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$	$\mathcal{D}_1$

3.2. Competitors

The refined estimator dedicated to the estimation of extreme quantiles for Weibull tail-distributions is compared to three competitors. All three estimators share the same structure and rely on three quantities, i.e. the order statistic $X_{n-k_n+1,n}$, an extrapolation term and an estimator of the Weibull tail-coefficient.

Let us first consider the estimator (2.3) of the Weibull tail-coefficient introduced in Gardes and Girard (2006). The extreme quantile estimator proposed in Gardes and Girard (2005) can be interpreted as a particular case of (2.2) with $k'_n = k_n$ and $\hat{\theta}_n(\cdot) = \hat{\theta}_n^{\mathrm{RSH}}(\cdot)$, see (2.4):

$$\hat{q}_n^{\mathrm{RSH}}(\alpha_n, k_n) = X_{n-k_n+1,n}\, \tau_n^{\hat{\theta}_n^{\mathrm{RSH}}(k_n)}. \tag{3.1}$$

More recently, an estimator of the Weibull tail-coefficient based on the mean excess function $t \mapsto m(t) = \mathbb{E}(X - t \,|\, X > t)$ has been introduced in Beirlant *et al.* (2009). In practice, the authors estimate $m(X_{n-j,n})$ for all $j \in \{1, ..., k_n\}$ by its empirical counterpart:

$$\hat{m}_n(X_{n-j,n}) = \frac{1}{j} \sum_{i=1}^{j} X_{n-i+1,n} - X_{n-j,n},$$

which leads to the following estimator of θ based on log-spacings between the mean excesses:

$$\hat{\theta}_n^{\mathrm{MEF}}(k_n) = \left(1 - \frac{1}{\hat{\gamma}_n^{\mathrm{H}}(k_n)} \frac{1}{k_n} \sum_{j=1}^{k_n} \log \hat{m}_n(X_{n-j,n}) - \log \hat{m}_n(X_{n-k_n-1,n}) \right)^{-1}.$$

Letting $k'_n = k_n$ and $\hat{\theta}_n(\cdot) = \hat{\theta}_n^{\mathrm{MEF}}(\cdot)$ in (2.3) yields the following estimator of the extreme quantile:

$$\hat{q}_n^{\mathrm{MEF}}(\alpha_n, k_n) = X_{n-k_n+1,n}\, \tau_n^{\hat{\theta}_n^{\mathrm{MEF}}(k_n)}. \tag{3.2}$$

Up to our knowledge there exists only one bias-reduced extreme quantile estimator dedicated to Weibull tail-distributions. This estimator (Diebolt *et al.*, 2008) is based on a least-squares approach and involves a bias-reduced estimator of the Weibull tail-coefficient proposed by the same authors (Diebolt *et al.*, 2008):

$$\hat{\theta}_n^{\mathrm{LSE}}(k_n) = \bar{Y}_{k_n} - \hat{b}(\log(n/k_n))\bar{x}_{k_n},$$

where

$$\bar{Y}_{k_n} = \frac{1}{k_n}\sum_{j=1}^{k_n} Y_j \qquad \text{with} \quad Y_j = j\log(n/j)(\log X_{n-j+1,n} - \log X_{n-j,n}),$$

$$\bar{x}_{k_n} = \frac{1}{k_n}\sum_{j=1}^{k_n} x_j \qquad \text{with} \quad x_j = \log(n/k_n)/\log(n/j),$$

and where

$$\hat{b}(\log(n/k_n)) = \sum_{j=1}^{k_n}(x_j - \bar{x}_{k_n})Y_j \Bigg/ \sum_{j=1}^{k_n}(x_j - \bar{x}_{k_n})^2.$$

The associated extreme quantile estimator is defined as

$$\hat{q}_n^{\mathrm{LSE}}(\alpha_n, k_n) = X_{n-k_n+1,n}\, \tau_n^{\hat{\theta}_n^{\mathrm{LSE}}(k_n)} \exp\Big(\widehat{b}(\log(n/k_n))K_{\widehat{\rho}_n}(\tau_n)\Big). \tag{3.3}$$

The authors suggest to choose in practice $\widehat{\rho}_n = \rho^{\#} = -1$. This estimator features two bias corrections: a first one in the estimator of the Weibull tail-coefficient and a second one in the extrapolation term. This estimator is built under the assumption that $x|b(x)| \to \infty$ as $x \to \infty$ leading to the constraint $\rho \geq -1$. The latter assumption is fulfilled by the class of $\mathcal{D}(\zeta, \eta, a)$-distributions when $\eta \leq 1$.

Finally, recall that our estimator is given by

$$\hat{q}_n^{\mathrm{RWT}}(\alpha_n, k_n, k_n^{\star}(\tau_n, -1)) = X_{n-k_n+1,n}\, \tau_n^{\hat{\theta}_n^{\mathrm{RSH}}(k_n^{\star}(\tau_n,-1))}, \tag{3.4}$$

where $k_n^{\star}(\tau_n, -1) = \lfloor n(k_n/n)^{\beta^{\star}(\tau_n,-1)} \rfloor$ and $\beta^{\star}(\tau_n, -1) = \tau_n \log(\tau_n)/(\tau_n - 1)$. For the sake of simplicity, the above extreme quantile estimators (3.1)–(3.4) are respectively referred to as RSH, MEF, LSE and RWT in the sequel.

3.3. Selection of the intermediate sequence

All four considered extreme quantile estimators (RWT, RSH, LSE, MEF) depend on the intermediate sequence k_n. The selection of k_n is a crucial point which has been widely discussed in the extreme-value literature. A new algorithm for the selection of k_n is proposed in Allouche *et al.* (2023), basing on a bisection method inspired from random forests. The objective is to find the region with the smallest variance in a given series $\{Z_1, ..., Z_m\}$.

The proposed method starts by randomly splitting the series into two parts, computes the variance in each sub-region and repeats the action in the one with smallest variance until getting a final single point (see Allouche *et al.*, 2023, Algorithm 2). The above procedure is embedded in a bootstrap technique (see Allouche *et al.*, 2023, Algorithm 1), and the final $k_n^\dagger$ is selected by taking the median across the $T = 10,000$ bootstrap samples. In the simulations, $Z_j = \hat{q}(\alpha_n, k_{j,n})$, an estimator (RWT, RSH, LSE or MEF) of the extreme quantile at level α_n computed with the intermediate sequence $k_{j,n} \in \{15, 16, ..., 3n/4\}$.

3.4. Results

The performance of the four extreme quantile estimators is assessed using the Mean absolute relative error:

$$\text{MARE}(\hat{q}_n(\alpha_n)) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{q}_n^{(i)}(\alpha_n, k_n^{(i,\dagger)})}{q_n(\alpha_n)} - 1\right|, \tag{3.5}$$

where $\hat{q}_n^{(i)}(\alpha_n, k_n^{(i,\dagger)})$ denotes the estimator computed on the i-th replication, $i \in \{1, ..., N = 1,000\}$ with the intermediate sequence $k_n^{(i,\dagger)}$ selected using the above described procedure. The computed MAREs are provided in Table 4 and Table 6 for $\alpha_n = 1/n^2$ and in Table 5 and Table 7 for $\alpha_n = 1/n^4$. The results are summarized in Table 2: We start by remarking that, as expected, the smaller the order α_n of the extreme quantile is, i.e. the more one extrapolates, the larger the error is. This is true for all four considered estimators. The proposed RWT estimator is the most accurate one in average since it provides the best results in 48% of cases. Let us remark that, since we fixed $\rho^\# = -1$, the RWT estimator performs best overall when ρ is close to -1. The second most accurate estimator is LSE which provides the best results in 26% of the considered cases (19 out of 72 situations). As expected, and similarly to the RWT estimator, it performs well when $\rho = -1$. The RSH estimator shares similar performances with 25% of best results. It is remarkably efficient when $\rho = -\infty$ (all 4 situations) and more surprisingly when $\rho = -1/2$ where it obtains 14 best results.

Table 2: Summary of results obtained in Tables 4–7. Best estimator of the extreme quantiles $q(\alpha_n = 1/n^2)$ & $q(\alpha_n = 1/n^4)$ computed on simulated data from Weibull tail-distributions. The situations in bold are illustrated for the $\mathcal{D}(\zeta = 1/\theta, \eta = -\rho, a = 10)$-distribution on Figure 2 and Figure 3.

(ρ, θ)	1/2	3/4	1	5/4	3/2	7/4	2	9/4	5/2
$-\infty$	RSH						RSH		
$-5/2$									LSE
-2							RWT	RWT	RWT
$-3/2$					**RWT**	RWT	RWT	RWT	RWT
-1	RWT/LSE		**RWT/LSE**	RWT	**RWT**	LSE/MEF	**LSE**	LSE	LSE
$-1/2$	RWT	LSE	RSH	RSH	**RSH**	RSH	RSH	RSH	RSH

RSH performs well in this difficult case despite the fact that it does not benefit from a bias reduction. This unexpected performance may be explained by the relatively small bias factor, see the graph of $B(1, \cdot, -1/2)$ in the right panel of Figure 1. The four cases where RSH fails

to obtain the best results when $\rho = -1/2$ correspond to a Weibull tail-coefficient θ smaller than 1. Finally, MEF yields very poor estimations (even in the strict Weibull case), with less than 2% of best results (only 1 situation). In particular, it does not give acceptable results (with MARE ≥ 1) in 22% of the considered situations.

As an illustration, the median and MARE associated with the RWT, RSH and LSE estimators computed on a $\mathcal{D}(\zeta, \eta, a = 10)$-distribution for $\alpha_n = 1/n^2$ are depicted on Figure 2 and Figure 3 as functions of k_n. In Figure 2, the Weibull tail-coefficient is fixed to $\theta = 3/2$ and $\rho \in \{-1/2, -1, -3/2\}$ decreases (from top to bottom), while, in Figure 3, the second-order Weibull parameter is fixed to $\rho = -1$ and $\theta \in \{1, 3/2, 2\}$ increases (from top to bottom). In most of these situations, the RWT estimator has the smallest bias and thus the minimum value of the MARE is reached for larger values of k_n than RSH and LSE. To conclude, it appears on these experiments that, overall, the RWT estimator performs the best within the four considered estimators. One of its main competitors is LSE, which, similarly to RWT, considers the two sources of bias (associated with the Weibull tail-coefficient estimator and the extrapolation term).

4. ILLUSTRATION ON A REAL DATA SET

We study a data set of daily wind measures (in m/s) at Reims (France) from 01/01/1981 to 04/30/2011. For seasonality reasons, only the months from October to March are considered, resulting in $n = 5,371$ measures, see the top-left panel of Figure 4 for an histogram of the considered data. It is shown in Albert *et al.* (2020) that the Weibull tail model represents fairly well the upper tail of these data. The goal is to estimate the extreme quantile $q(1/n)$ (with $1/n \simeq 1.86 \cdot 10^{-4}$) and to compare it to the maximum of the sample $x_{n,n} = 42.26\ m/s$.

To this end, the Weibull tail-coefficient is estimated first by $\hat{\theta}_n^{\mathrm{RWT}}(k_n^\dagger) = \hat{\theta}_n^{\mathrm{RSH}}(\hat{k}_n^\star) = 0.5597$, where $k_n^\dagger = 2,877$ has been selected following the procedure introduced in Allouche *et al.* (2023) and sketched in Subsection 3.3. This yields $\hat{k}_n^\star = \hat{k}_n^\star(\tau_n, \rho^\#) = 961$ where we set $\rho^\# = -1$. As a visual check, a Weibull quantile-quantile plot of the log-excesses $(\log X_{n-i+1,n} - \log X_{n-\hat{k}_n^\star+1,n})$ versus $(\log\log(n/i) - \log\log(n/\hat{k}_n^\star))$ for $i \in \{1, ..., \hat{k}_n^\star\}$ is drawn on the top-right panel of Figure 4. The relationship appearing in this plot is approximately linear, which constitutes an empirical evidence that the Weibull-tail assumption makes sense and that $\hat{k}_n^\star = 961$ is a reasonable choice for the estimation of the Weibull tail-coefficient. A line with the estimated value $\hat{\theta}_n^{\mathrm{RWT}}(k_n^\dagger) = 0.5597$ as slope is added to the quantile-quantile plot highlighting the linear relationship. The function $k_n \mapsto \hat{\theta}_n^{\mathrm{RSH}}(k_n)$ is plotted on the bottom-left panel of Figure 4, it features as a nice stability for all $k_n \in \{100, ..., 4000\}$. A similar procedure is carried out for the other three estimators to select $k_n^\dagger$. The two Weibull tail-coefficient estimators RSH and MEF that do not benefit from a bias reduction provide respectively the smallest and the largest estimation: $\hat{\theta}_n^{\mathrm{RSH}}(k_n^\dagger) = 0.5017$ and $\hat{\theta}_n^{\mathrm{MEF}}(k_n^\dagger) = 0.6693$, while the bias-reduced estimator LSE gives a value $\hat{\theta}_n^{\mathrm{LSE}}(k_n^\dagger) = 0.6077$ close to the RWT estimate $\hat{\theta}_n^{\mathrm{RWT}}(k_n^\dagger) = 0.5597$. These results are reported in Table 3 with the corresponding estimated extreme quantiles $\hat{q}_n(1/n)$. The estimates of the extreme quantile provided by RSH and MEF seem to respectively underestimate and overestimate $q(1/n)$ with respectively $\mathrm{RSH}(1/n) = 33.89\ m/s$ and $\mathrm{MEF}(1/n) = 49.53\ m/s$ while the sample maximum is $x_{n,n} = 42.26\ m/s$. It appears that $\mathrm{LSE}(1/n) = 37.08$ is significantly smaller than the sample maximum.

Let us stress that the proposed refined estimator gives the closest estimate to the maximum value of the sample: $\mathrm{RWT}(1/n) = 41.00\ m/s$. Note that the behaviour of the RWT estimate is stable with respect to the choice of $\rho^{\#}$: $\mathrm{RWT}(1/n) \in \{37.62, 41.00, 39.84\}$ when $\rho^{\#} \in \{-2, -1, -1/2\}$ even though $\rho^{\#} = -1$ seems to be the best option. Finally, both sample paths $k_n \mapsto \mathrm{RWT}(1/n)$ and $k_n \mapsto \mathrm{LSE}(1/n)$ enjoy a stable behaviour in a large neighbourhood of $k_n^{\dagger}$, see the bottom-right panel of Figure 4. As a conclusion, according to $\mathrm{RWT}(1/n)$ estimate, one can expect a daily wind larger than $41.00\ m/s$ to occur in average once every 30 years during the October to March period.

Table 3: Comparison of the four estimators on the daily wind data set: Estimates of the Weibull tail-coefficient θ and extreme quantile $q(1/n)$. The selected intermediate sequence $k_n^{\dagger}$ is also given for each estimator.

	RSH	RWT	LSE	MEF
$\hat{\theta}_n(k_n^{\dagger})$	0.5017	0.5597	0.6077	0.6693
$\hat{q}_n(1/n, k_n^{\dagger})$	33.89	41.00	37.078	49.53
$k_n^{\dagger}$	2,206	2,877	2,792	2,202

5. CONCLUSION

As a conclusion, the RWT estimator is an efficient tool for estimating extreme quantiles from Weibull tail-distributions. It relies on the ideas of Allouche *et al.* (2023), consisting in selecting carefully two intermediate sequences to reduce the asymptotic bias of a Weissman type estimator. In contrast to this previous work, the proposed approach does not rely on a preliminary estimate of the second-order parameter; Any negative value may be used, and does yield an asymptotic bias reduction, as shown in our theoretical results. Other surprising features of Weibull tail-distributions can be found in Asimit *et al.* (2010). The proposed method provides satisfying results in our numerical experiments and outperforms all its competitors in half of the considered situations. This work could be extended by investigating the adaptation of this bias reduction principle to other estimators of extreme quantiles from Weibull tail-distributions.

Table 4: MAREs associated with the four estimators of the extreme quantile $q(\alpha_n = 1/n^2)$ computed on simulated data from the $\mathcal{D}(\zeta = 1/\theta, \eta = -\rho, a = 10)$-distribution. The best result is emphasized in bold. MAREs larger than 1 are not reported.

	RWT	RSH	LSE	MEF
$\theta = 1/2$				
$\rho = -1/2$	**0.0133**	0.0441	0.0203	0.0969
$\theta = 3/4$				
$\rho = -1/2$	0.0702	0.0459	**0.0412**	0.2543
$\theta = 1$				
$\rho = -1$	**0.1125**	0.3449	0.1814	—
$\rho = -1/2$	0.1317	**0.0640**	0.0715	0.3015
$\theta = 5/4$				
$\rho = -1$	**0.1430**	0.3386	0.1846	0.5964
$\rho = -1/2$	0.1937	**0.0857**	0.1029	0.2283
$\theta = 3/2$				
$\rho = -3/2$	**0.2116**	0.7095	0.3844	—
$\rho = -1$	**0.1874**	0.3374	0.1900	0.4198
$\rho = -1/2$	0.2517	**0.1090**	0.1332	0.2076
$\theta = 7/4$				
$\rho = -3/2$	**0.2470**	0.6989	0.3831	—
$\rho = -1$	0.2442	0.3330	**0.1986**	0.2705
$\rho = -1/2$	0.3154	**0.1349**	0.1663	0.3809
$\theta = 2$				
$\rho = -2$	**0.3236**	0.8833	0.4406	—
$\rho = -3/2$	**0.2869**	0.6945	0.3833	—
$\rho = -1$	0.2934	0.3311	**0.2136**	0.3168
$\rho = -1/2$	0.3744	**0.1586**	0.1971	0.5915
$\theta = 9/4$				
$\rho = -2$	**0.3710**	0.8783	0.4365	—
$\rho = -3/2$	**0.3274**	0.6915	0.3847	0.6088
$\rho = -1$	0.3401	0.3303	**0.2301**	0.4946
$\rho = -1/2$	0.4359	**0.1818**	0.2262	0.7415
$\theta = 5/2$				
$\rho = -5/2$	0.8958	0.9493	**0.3713**	—
$\rho = -2$	**0.4301**	0.8754	0.4420	—
$\rho = -3/2$	**0.3824**	0.6891	0.3874	0.4998
$\rho = -1$	0.4050	0.3309	**0.2526**	0.7070
$\rho = -1/2$	0.5086	**0.2101**	0.2629	0.8675

Table 5: MAREs associated with the four estimators of the extreme quantile $q(\alpha_n = 1/n^4)$ computed on simulated data from the $\mathcal{D}(\zeta = 1/\theta, \eta = -\rho, a = 10)$-distribution. The best result is emphasized in bold. MAREs larger than 1 are not reported.

	RWT	RSH	LSE	MEF
$\theta = 1/2$				
$\rho = -1/2$	**0.0363**	0.0835	0.0498	0.0871
$\theta = 3/4$				
$\rho = -1/2$	0.1049	0.0823	**0.0685**	0.2907
$\theta = 1$				
$\rho = -1$	**0.2797**	0.5471	0.3746	—
$\rho = -1/2$	0.1928	**0.0964**	0.1075	0.3572
$\theta = 5/4$				
$\rho = -1$	**0.2941**	0.5408	0.1471	0.4404
$\rho = -1/2$	0.2768	**0.1201**	0.1495	0.2738
$\theta = 3/2$				
$\rho = -3/2$	**0.3973**	0.8691	0.6227	—
$\rho = -1$	**0.3316**	0.5384	0.3743	0.3368
$\rho = -1/2$	0.3548	**0.1467**	0.1923	0.2555
$\theta = 7/4$				
$\rho = -3/2$	**0.4242**	0.8629	0.6135	—
$\rho = -1$	0.3813	0.5359	0.3744	**0.3286**
$\rho = -1/2$	0.4454	**0.1785**	0.2406	0.4611
$\theta = 2$				
$\rho = -2$	**0.4764**	0.9609	0.6591	—
$\rho = -3/2$	**0.4616**	0.8601	0.6109	—
$\rho = -1$	0.4351	0.5340	**0.3796**	0.4767
$\rho = -1/2$	0.5256	**0.2088**	0.2883	0.6809
$\theta = 9/4$				
$\rho = -2$	**0.5542**	0.9591	0.6429	—
$\rho = -3/2$	**0.4977**	0.8585	0.6088	0.6181
$\rho = -1$	0.4869	0.5316	**0.3878**	0.6607
$\rho = -1/2$	0.6005	**0.2355**	0.3304	0.8201
$\theta = 5/2$				
$\rho = -5/2$	-	0.9865	**0.5552**	—
$\rho = -2$	**0.6308**	0.9578	0.6388	—
$\rho = -3/2$	**0.5603**	0.8568	0.6047	0.6947
$\rho = -1$	0.5557	0.5298	**0.4045**	0.8317
$\rho = -1/2$	0.7019	**0.2698**	0.3847	0.9216

Table 6: MAREs associated with the four estimators of the extreme quantile $q(\alpha_n = 1/n^2)$ computed on simulated data from classical Weibull tail-distributions. The best result is emphasized in bold.

	RWT	RSH	LSE	MEF
$\mathcal{N}(\mu, \sigma = 1)$				
$\mu = 1$	0.1396	0.2804	**0.0798**	0.1852
$\mu = 3$	0.0776	0.0564	**0.0471**	0.2943
$\mu = 5$	0.0553	0.0882	**0.0440**	0.2682
$\mu = 7$	**0.0334**	0.0789	0.0366	0.1852
$\mu = 9$	**0.0361**	0.0822	0.0381	0.1987
$\mathcal{G}(\upsilon, \lambda = 1)$				
$\upsilon = 4$	0.1417	0.2240	**0.1089**	0.4930
$\upsilon = 6$	0.1219	0.2270	**0.1071**	0.5040
$\upsilon = 8$	**0.0943**	0.2095	0.1086	0.4581
$\upsilon = 10$	**0.1002**	0.2158	0.1017	0.4758
$\mathcal{W}(\upsilon, \lambda)$				
$\upsilon = \lambda = 1/2$	0.4932	**0.2100**	0.2593	0.8528
$\upsilon = \lambda = 2$	0.1217	**0.0510**	0.0646	0.3324

Table 7: MAREs associated with the four estimators of the extreme quantile $q(\alpha_n = 1/n^4)$ computed on simulated data from classical Weibull tail-distributions. The best result is emphasized in bold.

	RWT	RSH	LSE	MEF
$\mathcal{N}(\mu, \sigma = 1)$				
$\mu = 1$	0.1914	0.4430	**0.1212**	0.3009
$\mu = 3$	0.1162	0.1001	**0.0797**	0.3354
$\mu = 5$	**0.0837**	0.1584	0.0897	0.2507
$\mu = 7$	**0.0780**	0.1565	0.0892	0.2270
$\mu = 9$	**0.0702**	0.1477	0.0856	0.1895
$\mathcal{G}(\upsilon, \lambda = 1)$				
$\upsilon = 4$	0.2206	0.3616	**0.2024**	0.4721
$\upsilon = 6$	**0.2008**	0.3709	0.2126	0.4677
$\upsilon = 8$	**0.1906**	0.3669	0.2148	0.4492
$\upsilon = 10$	**0.1825**	0.3596	0.2137	0.4275
$\mathcal{W}(\upsilon, \lambda)$				
$\upsilon = \lambda = 1/2$	0.6775	**0.2729**	0.3886	0.9070
$\upsilon = \lambda = 2$	0.1750	**0.0664**	0.0938	0.4333

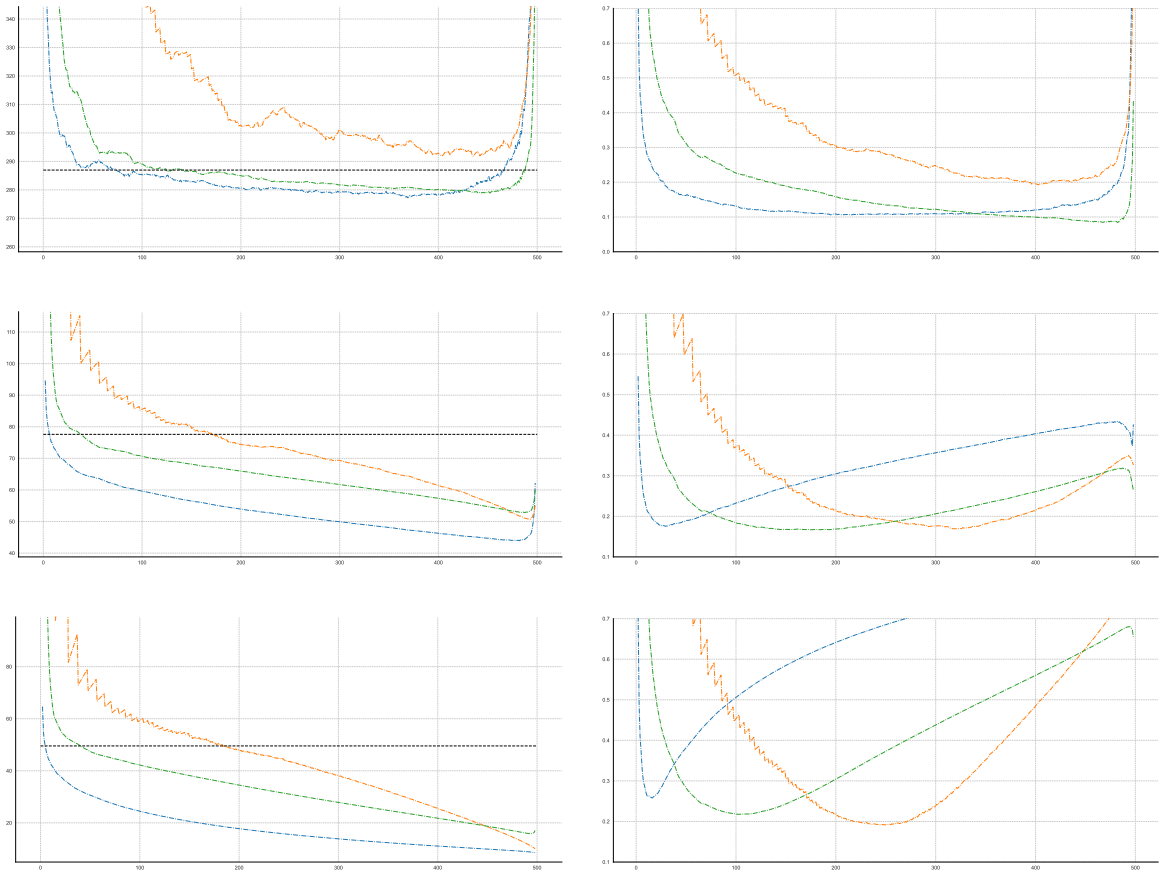

Figure 2: Illustration on simulated data sets of size $n = 500$ from a $\mathcal{D}(\zeta = 1/\theta = 2/3, \eta = -\rho, a = 10)$-distribution with $\rho \in \{-1/2, -1, -3/2\}$ (from top to bottom) computed on $N = 1000$ replications. Medians (left panel) and MAREs (right panel) as functions of $k_n \in \{2, ..., n-1\}$, associated with estimators RWT (orange), RSH (blue) and LSE (green) of the extreme quantile $q(\alpha_n = 1/n^2)$ (black dashed line).

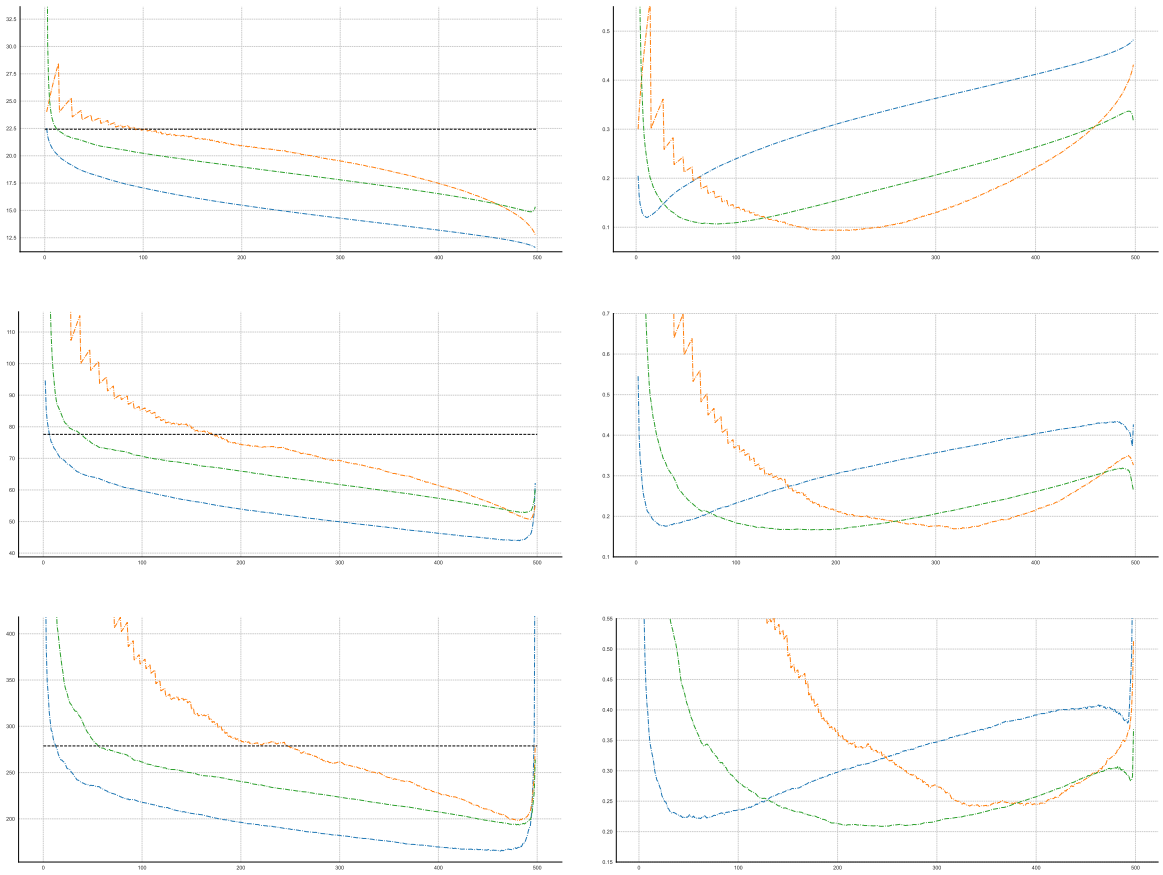

Figure 3: Illustration on simulated data sets of size $n = 500$ from a $\mathcal{D}(\zeta = 1/\theta, \eta = -\rho = 1, a = 10)$-distribution with $\theta \in \{1, 3/2, 2\}$ (from top to bottom) computed on $N = 1000$ replications. Medians (left panel) and MAREs (right panel) as functions of $k_n \in \{2, ..., n-1\}$, associated with estimators RWT (orange), RSH (blue) and LSE (green) of the extreme quantile $q(\alpha_n = 1/n^2)$ (black dashed line).

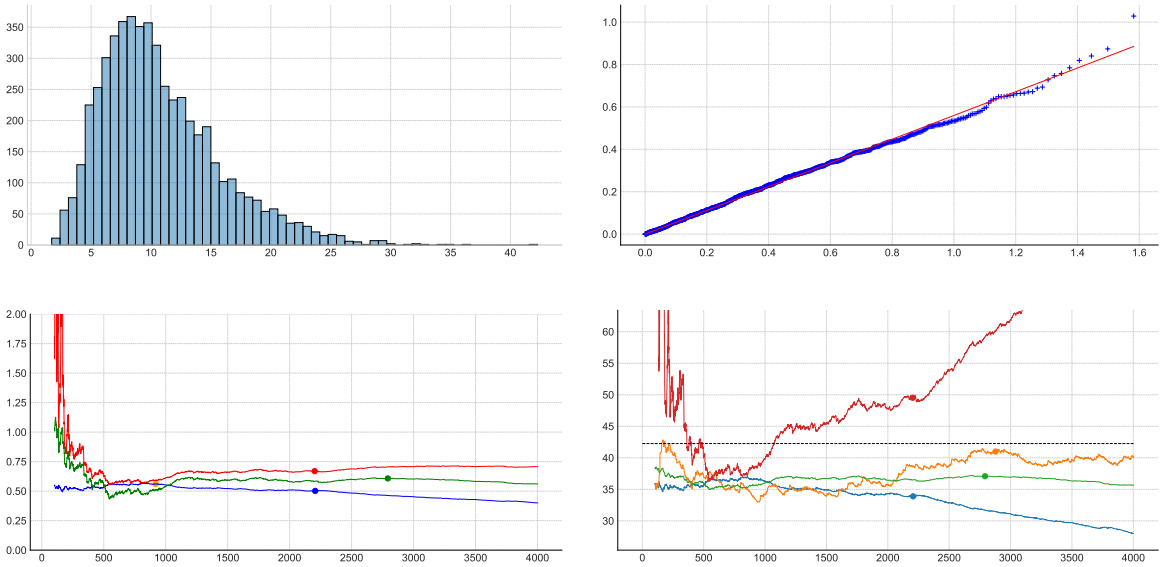

Figure 4: Illustration on the daily wind data set.
Top-left: Histogram of the data set.
Top-right: Weibull quantile-quantile plot (horizontally: $(\log\log(n/i) - \log\log(n/\hat{k}_n^\star))$, vertically: $(\log X_{n-i+1,n} - \log X_{n-\hat{k}_n^\star+1,n})$ for $i \in \{1, ..., \hat{k}_n^\star = 961\}$). A line with slope $\hat{\theta}_n^{\mathrm{RWT}}(k_n^\dagger) = 0.5597$ is superimposed in red.
Bottom-left: Estimates of the Weibull-tail coefficient RSH (blue), LSE (green) and MEF (red) as functions of k_n. The range is limited to $k_n \in \{100, ..., 4000\}$ for the sake ok readability. The pair $(k_n^\dagger, \hat{\theta}_n(k_n^\dagger))$ associated with the selected value $k_n^\dagger$ is emphasized by a circle. The RWT estimate is represented by an orange circle on the RSH curve.
Bottom-right: Estimates of the extreme quantile $q(\alpha_n = 1/n)$ by RWT (orange), RSH (blue), LSE (green) and MEF (red) as functions of k_n with their associated $k_n^\dagger$ emphasized by a circle. The sample maximum $x_{n,n}$ is depicted by a black dashed line.

A. APPENDIX: Proofs

Proofs of main results are collected in Subsection A.1. Auxiliary results are provided in Subsection A.2 and proved in the Supplementary material document.

A.1. Proofs of main results

Proof of Theorem 2.1. Clearly, the following expansion holds:

$$\sqrt{k'_n}\log\left(\frac{\hat{q}_n(\alpha_n,k_n,k'_n)}{q(\alpha_n)}\right) = T_n^{(1)}+T_n^{(2)}+T_n^{(3)},$$
$$\text{with}\quad T_n^{(1)}=\sqrt{k'_n}\left(\frac{X_{n-k_n+1,n}}{V(\log(n/k_n))}\right),\quad T_n^{(2)}=\sqrt{k'_n}(\log\tau_n)(\hat{\theta}_n^{\mathrm{RSH}}(k'_n)-\theta),$$
$$\text{and}\quad T_n^{(3)}=\sqrt{k'_n}\left(\frac{\ell(\log(n/k_n))}{\ell(\log(1/\alpha_n))}\right).$$

Let us consider the three terms separately. First, (Gardes and Girard, 2005, Lemma 1) shows that, under **(A.1)**, $k_n\to\infty$, $k_n/n\to 0$ and condition (iii):

$$T_n^{(1)}=\frac{\sqrt{k'_n/k_n}}{\log(n/k_n)}\theta\xi'_n+O_P\left(\frac{\sqrt{k'_n}}{k_n\log^2(n/k_n)}\right), \tag{A.1}$$

where $\xi'_n\xrightarrow{d}\mathcal{N}(0,1)$. Second, (Gardes and Girard, 2006, Proposition 2.1) entails that, under assumptions **(A.1)**, **(A.2)**, $k'_n\to\infty$ and $k'_n/n\to 0$, the following expansion holds:

$$T_n^{(2)}=\theta\log(\tau)\xi_n+\theta\log(\tau)\mu(\log(n/k'_n))\xi''_n+\log(\tau)\sqrt{k'_n}b(\log(n/k'_n))(1+o(1)), \tag{A.2}$$

where $\xi_n\xrightarrow{d}\mathcal{N}(0,1)$, $\xi''_n\xrightarrow{d}\mathcal{N}(0,1)$. Third, **(A.2)** and condition (iii) imply

$$T_n^{(3)}=-K_\rho(\tau)\sqrt{k'_n}b(\log(n/k_n))(1+o(1)). \tag{A.3}$$

Collecting (A.1), (A.2) and (A.3), one has

$$\begin{aligned}\sqrt{k'_n}\log&\left(\frac{\hat{q}_n(\alpha_n,k_n,k'_n)}{q(\alpha_n)}\right)\\ &=\theta\log(\tau)\xi_n+\theta\log(\tau)\mu(\log(n/k'_n))\xi''_n+\frac{\sqrt{k'_n/k_n}}{\log(n/k_n)}\theta\xi'_n+O_P\left(\frac{\sqrt{k'_n}}{k_n\log^2(n/k_n)}\right)\\ &\quad+\sqrt{k'_n}\big\{\log(\tau)b(\log(n/k'_n))(1+o(1))-K_\rho(\tau)b(\log(n/k_n))(1+o(1))\big\}.\end{aligned}$$

Recalling that, from (Gardes and Girard, 2006, Lemma 5.3), $\mu(t)\sim 1/t$ as $t\to\infty$, the above expansion can be simplified as

$$\begin{aligned}\sqrt{k'_n}\log&\left(\frac{\hat{q}_n(\alpha_n,k_n,k'_n)}{q(\alpha_n)}\right)\\ &=\theta\log(\tau)\xi_n+\frac{\sqrt{k'_n/k_n}}{\log(n/k_n)}\theta\xi'_n+O_P\left(\frac{\sqrt{k'_n}}{k_n\log^2(n/k_n)}\right)\\ &\quad+\sqrt{k'_n}\big\{\log(\tau)b(\log(n/k'_n))(1+o(1))-K_\rho(\tau)b(\log(n/k_n))(1+o(1))\big\}.\end{aligned}$$

Finally, remark that assumption (ii) implies $k_n' \leq k_n$ eventually and $b(\log(n/k_n)) \sim \beta^{-\rho} b(\log(n/k_n'))$ as $n \to \infty$ so that

$$\sqrt{k_n'}\log\left(\frac{\hat{q}_n(\alpha_n,k_n,k_n')}{q(\alpha_n)}\right) = \theta\log(\tau)\xi_n + \sqrt{k_n'}b(\log(n/k_n'))\big(\log(\tau) - \beta^{-\rho}K_\rho(\tau) + o(1)\big)(1+o(1)) + o_P(1).$$

Assumption (i) then yields

$$\sqrt{k_n'}\log\left(\frac{\hat{q}_n(\alpha_n,k_n,k_n')}{q(\alpha_n)}\right) \xrightarrow{d} \mathcal{N}\big(\lambda(\log\tau - \beta^{-\rho}K_\rho(\tau)), (\theta\log\tau)^2\big)$$

and a first order Taylor expansion proves the result. □

Proof of Lemma 2.1: **(i)** Remarking that $K_y(\tau) \sim -1/y$ as $y \to -\infty$ for all $\tau > 1$ yields $\beta^\star(y,\tau) \to 1$ as $y \to -\infty$. The result $\beta^\star(\tau,0) := \sqrt{\tau}$ follows from a second-order Taylor expansion.

(ii) First, Lemma A.3(iii) implies that, for all $\tau > 1$ and $y < 0$, $h_y(\tau) > 1$ and thus $\beta^\star(\tau,y) = h_y(\tau)^{-1/y} > 1$. Second, Lemma A.3(iii) implies that, for all $\tau > 1$ and $y < 0$, $h_y(\tau) < \tau^{-y/2} < \tau^{-y}$ when $y < 0$. This straightforwardly implies that $K_y(\tau)/\log(\tau) > \tau^y$ which is equivalent to $\beta^\star(\tau,y) < \tau$. In the particular case where $y = 0$, from (i), one can take $\beta^\star(\tau,0) := \sqrt{\tau} < \tau$ since $\tau > 1$ and the result is proved.

(iii) Let us first consider $\tilde{k}_n(\tau,y) = n(k_n/n)^{\beta^\star(\tau,y)}$ such that $k_n^\star(\tau,y) = \lfloor \tilde{k}_n(\tau,y) \rfloor$. Clearly, $\tilde{k}_n(\tau,y)/k_n = (k_n/n)^{\beta^\star(\tau,y)-1}$ and $\beta^\star(\tau,y) > 1$ in view of (ii). As a consequence, for all $y \leq 0$, $\tilde{k}_n(\tau,y)$ is an increasing function of k_n, $\tilde{k}_n(\tau,y) \leq k_n$ and $\tilde{k}_n(\tau,y)/k_n \to 0$ as $n \to \infty$. These properties can be extended to $k_n^\star(\tau,y)$ without difficulty since the integer part is an increasing function and $k_n^\star(\tau,y) \leq \tilde{k}_n(\tau,y)$.

(iv) Routine calculations give for all $\tau > 1$ and $y < 0$,

$$\begin{aligned}\frac{\partial}{\partial y}\log(\beta^*(\tau,y)) &= \frac{1}{y^2(\tau^y-1)}\left(\tau^y\log(\tau^y) - \tau^y + 1 - \tau^y\log\left(\frac{\tau^y-1}{\log(\tau^y)}\right) + \log\left(\frac{\tau^y-1}{\log(\tau^y)}\right)\right)\\ &=: \frac{1}{y^2(\tau^y-1)}\varphi(\tau,y).\end{aligned}$$

Letting $x := \tau^y \in (0,1)$ yields

$$\tilde{\varphi}(x) := \varphi(\tau,y) = x\log(x) - x + 1 - x\log\left(\frac{x-1}{\log(x)}\right) + \log\left(\frac{x-1}{\log(x)}\right),$$

and differentiating, one gets

$$\tilde{\varphi}'(x) = -\log\left(\left(1-\frac{1}{x}\right)\frac{1}{\log(x)}\right) + \left(1-\frac{1}{x}\right)\frac{1}{\log(x)} - 1 = -\log(u(x)) + u(x) - 1,$$

where $u(x) := (1-1/x)/\log(x) > 0$. It thus appears that $\tilde{\varphi}'(x) \geq 0$ for all $x \in (0,1)$ since $-\log(u) + u - 1 \geq 0$ for all $u > 0$. As a consequence, $\tilde{\varphi}(\cdot)$ is an increasing function on $(0,1)$. Moreover, taking account of $\tilde{\varphi}(x) \to 0$ as $x \to 1^-$ shows that $\tilde{\varphi}(x) \leq 0$ for all $x \in (0,1)$. Finally, $\tau^y - 1 < 0$ and $\varphi(\tau,y) \leq 0$ for all $\tau > 1$ and $y < 0$ imply that $\beta^*(\tau,y)$ is an increasing function of y which in turns shows that $k_n^\star(\tau,y)$ is a decreasing function of y. □

Proof of Corollary 2.1. (i) To prove the convergence in distribution, it is sufficient to show that conditions (i), (ii) and (iii) of Theorem 2.1 hold. Let $y \leq 0$ and $\tau > 1$. First, one can easily check that $k_n^\star(\tau, y) b^2(\log n) \to \lambda^2$ as $n \to \infty$ and thus $\sqrt{k_n^\star(\tau, y)} b(\log n) \to \lambda$ in view of the sign assumption on λ. Besides,

$$\log(n/k_n^\star(\tau, y)) = \log n + 2\log|b(\log n)| - 2\log|\lambda| + o(1) \sim \log n, \tag{A.4}$$

since $b(\cdot)$ is regularly-varying so that $b(\log(n/k_n^\star(\tau, y))) \sim b(\log n)$ and thus

$$\sqrt{k_n^\star(\tau, y)} b(\log(n/k_n^\star(\tau, y))) \to \lambda,$$

as $n \to \infty$. Theorem 2.1(i) is thus proved. Second, observe that

$$\tau_n = \frac{\log(1/\alpha_n)}{\log(n/k_n)} = \frac{\log(1/c)}{\log(n/k_n)} + \frac{\tau}{\beta^\star(\tau, \rho)} \frac{\log n}{\log(n/k_n)}$$

and

$$\log(n/k_n) = \frac{1}{\beta^\star(\tau, \rho)} (\log n + 2\log|b(\log n)| - 2\log|\lambda|) + o(1) \sim \frac{1}{\beta^\star(\tau, \rho)} \log n, \tag{A.5}$$

as $n \to \infty$. It is thus clear that $\tau_n \to \tau$ as $n \to \infty$, which is Theorem 2.1(ii). Third, Theorem 2.1(iii) is a straightforward consequence of (A.4) and (A.5).

(ii) Proposition A.1 concludes the proof. □

A.2. Auxiliary results

Let us begin with a Lemma that establishes that the strict Weibull distribution belongs to the Gumbel maximum domain of attraction ($\gamma = 0$), and more importantly, with a second-order parameter $\psi = 0$. This result illustrates why inference on Weibull-tail distributions may be difficult since the situation $\gamma = \psi = 0$ is the most complicated one for classical extreme-value estimators. Let us also recall that, in contrast, the second-order Weibull parameter is $\rho = -\infty$ (see Girard, 2004, Table 1) and therefore strict Weibull distributions are an easy situation for dedicated Weibull-tail estimators.

Lemma A.1. *Suppose F is the cumulative distribution function of a strict Weibull distribution with shape parameter $\upsilon > 0$, $\upsilon \neq 1$ and scale parameter $\lambda > 0$. Then, the associated tail quantile function $U(\cdot) := F^{\leftarrow}(1 - 1/\cdot)$ verifies the second-order condition*

$$\frac{1}{A(t)} \left(\frac{U(tx) - U(t)}{a(t)} - K_\gamma(x) \right) \to \int_1^x s^{\gamma - 1} K_\psi(s) ds,$$

as $t \to \infty$, for all $x > 0$ (see de Haan and Ferreira, 2007, equation (3.4.5)), with $\gamma = 0$, $\psi = 0$, $a(t) = (\lambda/\upsilon)(\log t)^{1/\upsilon - 1}$ and $A(t) = (1 - \upsilon)/(\upsilon \log t)$.

The following three analytical results are used to prove Proposition A.1 below.

Lemma A.2. *Let us define for all* $(u, \beta) \in (0, 1] \times [0, 1/2]$, $g_\beta(u) := \log(u)(1 + u^\beta) - 2(u - 1)$. *Then,* $\forall \beta \in [0, 1/2]$ *one has* $g_\beta(u) < 0$ *if* $u \in (0, 1)$ *and* $g_\beta(1) = 0$.

Lemma A.3. *Let us define, for all* $\tau > 1$ *and* $y \leq 0$,

$$h_y(\tau) := \frac{K_0(\tau)}{K_y(\tau)} = \frac{y \log(\tau)}{\tau^y - 1} \quad \text{if } y < 0 \qquad \text{and} \qquad h_0(\tau) := 1 \text{ otherwise.}$$

Then,

- **(i)** $h_y(\cdot)$ *can be extended by continuity letting* $h_y(1) = 1$ *for all* $y \leq 0$.
- **(ii)** $h_y(\cdot)$ *is an increasing function on* $[1, \infty)$ *for all* $y < 0$ *(and* $h_0(\cdot)$ *is a constant function).*
- **(iii)** $1 < h_y(\tau) < \tau^{-y/2}$ *for all* $\tau > 1$ *and* $y < 0$ *(all three quantities coincide at* $y = 0$*).*

Lemma A.4. *Let us define, for all* $\tau > 1$ *and* $y \leq 0$,

$$f_y(\tau) := -\frac{\log(h_y(\tau))}{y \log(\tau)} \quad \text{if } y < 0 \qquad \text{and} \qquad f_0(\tau) = 1/2 \text{ otherwise.}$$

Then,

- **(i)** $f_y(\cdot)$ *can be extended by continuity letting* $f_y(1) = 1/2$ *for all* $y \leq 0$.
- **(ii)** $f_y(\cdot)$ *is a decreasing function on* $[1, \infty)$ *for all* $y < 0$ *(*$f_0(\cdot)$ *is a constant function).*
- **(iii)** $0 < f_y(\tau) < 1/2$ *for all* $\tau \geq 1$ *and* $y < 0$ *(the last two quantities coincide at* $y = 0$*).*

The next Proposition establishes two unexpected results. First, the bias associated with the refined estimator of extreme quantiles is strictly smaller than the bias associated with the original one, even though a misspecification of the second-order Weibull parameter is used. Second, both (asymptotic) biases vanish at $\rho = 0$.

Proposition A.1. *For all* $\tau > 1$, $y \leq 0$ *and* $\rho < 0$, $|B(\beta^\star(\tau, y), \tau, \rho)| < B(1, \tau, \rho)$, *see* (2.6) *and* (2.10). *Besides,* $B(\beta^\star(\tau, y), \tau, 0) = B(1, \tau, 0) = 0$ *for all* $\tau > 1$ *and* $y \leq 0$.

ACKNOWLEDGMENTS

This work is supported by the French National Research Agency (ANR) in the framework of the Investissements d'Avenir Program (ANR-15-IDEX-02). S. Girard also acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French École Polytechnique and its Foundation and sponsored by BNP Paribas.

REFERENCES

Aarssen, K. and de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies*, 4(4):259–281.

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Vol. 55). US Government printing office.

Albert, C., Dutfoy, A., and Girard, S. (2020). Asymptotic behavior of the extrapolation error associated with the estimation of extreme quantiles. *Extremes*, 23(2):349–380.

Allouche, M., El Methni, J., and Girard, S. (2023). A refined Weissman estimator for extreme quantiles. *Extremes*, 26:545–572.

Asimit, A.V., Li, D., and Peng, L. (2010). Pitfalls in using Weibull tailed distributions. *Journal of Statistical Planning and Inference*, 140(7):2018–2024.

Beirlant, J., Bouquiaux, C., and Werker, B. (2006). Semiparametric lower bounds for tail index estimation. *Journal of Statistical Planning and Inference*, 136:705–729.

Beirlant, J., Broniatowski, M., Teugels, J.L., and Vynckier, P. (1995). The mean residual life function at great age: Applications to tail estimation. *Journal of Statistical Planning and Inference*, 45:21–48.

Beirlant, J., Dierckx, G., Guillou, A., and de Waal, D. (2009). A new estimation method for Weibull-type tails based on the mean excess function. *Journal of Statistical Planning and Inference*, 139(6):1905–1920.

Beirlant, J. and Teugels, J.L. (1992). Modeling large claims in non-life insurance. *Insurance: Mathematics and Economics*, 11:17–29.

Beirlant, J., Teugels, J.L., and Vynckier, P. (1996). *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, Belgium.

Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1989). *Regular Variation*. Cambridge University Press.

Broniatowski, M. (1993). On the estimation of the Weibull tail coefficient. *Journal of Statistical Planning and Inference*, 35:349–366.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *The Annals of Statistics*, 27:760–781.

Diebolt, J., Gardes, L., Girard, S., and Guillou, A. (2008). Bias-reduced estimators of the Weibull tail-coefficient. *Test*, 17(2):311–331.

Diebolt, J., Gardes, L., Girard, S., and Guillou, A. (2008). Bias-reduced extreme quantile estimators of Weibull tail-distributions. *Journal of Statistical Planning and Inference*, 138(5):1389–1401.

Gardes, L. and Girard, S. (2005). Estimating extreme quantiles of Weibull tail distributions. *Communications in Statistics – Theory and Methods*, 34(5):1065–1080.

Gardes, L. and Girard, S. (2006). Comparison of Weibull tail-coefficient estimators. *REVSTAT – Statistical Journal*, 4(2):163–188.

Gardes, L. and Girard, S. (2008). Estimation of the Weibull tail-coefficient with linear combination of upper order statistics. *Journal of Statistical Planning and Inference*, 138:1416–1427.

Gardes, L., Girard, S., and Guillou, A. (2011). Weibull tail-distributions revisited: a new look at some tail estimators. *Journal of Statistical Planning and Inference*, 141(1):429–444.

Girard, S. (2004). A Hill type estimate of the Weibull tail-coefficient. *Communications in Statistics – Theory and Methods*, 33(2):205–234.

Goegebeur, Y. and Guillou, A. (2011). A weighted mean excess function approach to the estimation of Weibull-type tails. *Test*, 20(1):138–162.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Science and Business Media.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B*, 44(1):37–42.

Hall, P. and Welsh, A.W. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13:331–341.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174.

Vladimirova, M., Girard, S., Hien, N., and Arbel, J. (2020). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat*, 9:e318.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

Worms, J. and Worms, R. (2019). Estimation of extremes for Weibull-tail distributions in the presence of random censoring. *Extremes*, 22:667–704.

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 527–542
https://doi.org/10.57805/revstat.v23i4.659

Stochastic Orders on the Univariate Unified Skew Normal Family of Distributions

Authors: Souhila Merabet ✉
– Department of Probability and Statistics, Faculty of Mathematics, USTHB, Algiers, Algeria
souhila_merabet@yahoo.fr
smerabet@usthb.dz

Rabah Messaci
– Department of Probability and Statistics, Faculty of Mathematics, USTHB, Algiers, Algeria
rrabmes@gmail.com

Received: July 2023 Revised: January 2024 Accepted: January 2024

Abstract:

- The family of Unified Skew Normal distributions is associated with probability distributions encountered in various problems, notably those of the selection of individuals in a normal population. In this work, we focus on the ordering of the real subfamily of this class of distributions with respect to its parameters and for some stochastic orders (usual order, increasing convex order, increasing concave order and likelihood ratio order). Our results are applied to a reliability problem as well as a selection problem.

Keywords:

- *unified skew-normal distribution; extended skew-normal distribution; stochastic orders; selection distribution; logconcavity.*

AMS Subject Classification:

- 62E15, 60E15, 62H10.

✉ Corresponding author

1. INTRODUCTION

The Unified Skew Normal family $\text{SUN}_{(p,q)}$, which was introduced by Arellano-Valle and Azzalini (2006), is a family of asymmetric distributions generalizing the normal distribution and incorporating parameters controlling this asymmetry. The $\text{SUN}_{(p,q)}$ family also generalizes the basic asymmetric family of Skew Normal (SN) distributions, which was introduced by Azzalini (1985) and Azzalini and Dalla Valle (1996). In practice, the $\text{SUN}_{(p,q)}$ familly is used as an alternative to the gaussian distribution for modelling non-normal features as skewness. Among the numerous methods used to generate the $\text{SUN}_{(p,q)}$ distributions, we find the method known as stochastic representation by conditioning which is achieved as follows: if $(\boldsymbol{U}^t, \boldsymbol{V}^t)^t$ is a Gaussian vector of (p, q) order, then the distribution of $\boldsymbol{U}|\boldsymbol{V} > \boldsymbol{0}$ belongs to the family of $\text{SUN}_{(p,q)}$ distributions. As a result, the obtained distribution is a special case of the so-called selection distributions (see Arellano-Valle *et al.*, 2006). In this work, we consider the case of $\text{SUN}_{(1,q)}$ distributions where U is real and $\boldsymbol{V}$ is a vector and we intend to study the influence of the underlying $(\boldsymbol{U}^t, \boldsymbol{V}^t)^t$ vector parameters on the induced selection distribution. More specifically, we are interested in studying the stochastic ordering of the latter, relative to their parameters for some stochastic orders: usual stochastic order (first dominance stochastic order), increasing concave order (second dominance stochastic order), increasing convex order and the likelihood ratio order.

Azzalini (2014) has established the usual stochastic order for SN distributions with respect to the skewness parameter α, whereas Loperfido *et al.* (2007) have shown the existence of the likelihood ratio order for the location parameter μ. Blasi and Scarlatti (2012) have also addressed the stochastic ordering of the SN distribution with respect to μ, α and the dispersion parameter σ.

The present work is structured as follows: in Section 1, we recall the general definition of the SUN distributions with some of their properties. Section 2 introduces some of the stochastic orders that we shall consider in this work. In Section 3, we provide the main results on the ordering of the $\text{SUN}_{(1,q)}$ distributions. Section 4 deals with the application of the results obtained to both a reliability and a selection problems.

2. THE $\text{SUN}_{(p,q)}$ DISTRIBUTION

Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be two Gaussian random vectors such that:

$$\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \sim N_{p+q}(\boldsymbol{\xi}, \boldsymbol{\Omega}^*), \qquad \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\gamma} \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\gamma} \in \mathbb{R}^q$ and $\boldsymbol{\Omega}^*$ is a non-singular variance covariance matrix.

The correlation matrix associated with $\boldsymbol{\Omega}^*$ is:

$$\bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} \bar{\boldsymbol{\Omega}} & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^t & \boldsymbol{\Gamma} \end{pmatrix},$$

where $\bar{\boldsymbol{\Omega}} \in \mathbb{R}^p \times \mathbb{R}^p$, $\boldsymbol{\Gamma} \in \mathbb{R}^q \times \mathbb{R}^q$, $\boldsymbol{\Delta} \in \mathbb{R}^p \times \mathbb{R}^q$.

Let $\boldsymbol{\Omega} = \boldsymbol{\sigma}\bar{\boldsymbol{\Omega}}\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ is a diagonal matrix of order p with elements $\sigma_1, \sigma_2, ..., \sigma_p$ representing the standard deviations of the $\boldsymbol{U}$ components.

Definition 2.1. A random vector $\boldsymbol{X}$ with values in $\mathbb{R}^p$ is said to have a $\text{SUN}_{(p,q)}$ distribution if

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{U}\,|\,(\boldsymbol{V} > \boldsymbol{0}), \tag{2.2}$$

where $\boldsymbol{V} > \boldsymbol{0}$ means that each component of $\boldsymbol{V}$ is greater than 0.

The probability density function of $\boldsymbol{X}$ is then given by:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \phi_p(\boldsymbol{x} - \boldsymbol{\mu}; \boldsymbol{\Omega})\, \frac{\Phi_q\big(\boldsymbol{\gamma} + \boldsymbol{\Delta}^t\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}); \boldsymbol{\Gamma} - \boldsymbol{\Delta}^t\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\big)}{\Phi_q(\boldsymbol{\gamma}; \boldsymbol{\Gamma})}, \tag{2.3}$$

where $\boldsymbol{x} \in \mathbb{R}^p$ and $\phi_n(\cdot; \Sigma)$, $\Phi_n(\cdot; \Sigma)$ are the probability density function (pdf) and the cumulative distribution function (cdf) of the multivariate normal distribution $N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$, respectively.

The distribution of $\boldsymbol{X}$ is denoted $\text{SUN}_{(p,q)}(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ where $\boldsymbol{\mu}$ is the location parameter, $\boldsymbol{\Omega}$ the dispersion parameter, $\boldsymbol{\gamma}$ the truncation parameter and $\boldsymbol{\Delta}$ the shape parameter. Its expectation is given by (see Azzalini and Bacchieri, 2010):

$$E(X) = \boldsymbol{\mu} + \boldsymbol{\sigma}\boldsymbol{\Delta}\frac{\nabla\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})}{\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})}, \tag{2.4}$$

where $\nabla\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})$ is the gradient vector at point $\boldsymbol{\gamma}$ of $\Phi_q(\cdot, \boldsymbol{\Gamma})$.

If $p = 1$ and $q = 1$, we get $\Omega = \sigma^2$. In addition, for $\Gamma = 1$ and by letting $\Delta = \rho$, this distribution is the so-called Extended Skew Normal distribution (ESN) and it admits the following pdf (Arnold and Beaver, 2002):

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{\gamma + \rho\left(\frac{x-\mu}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)}{\sigma\Phi(\gamma)}, \qquad x \in \mathbb{R}, \tag{2.5}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf respectively.

The expectation of X simplifies to:

$$E(X) = \mu + \sigma\rho\lambda(\gamma), \tag{2.6}$$

where $\lambda(\gamma) = \frac{\phi(\gamma)}{\Phi(\gamma)}$ is the inverse Mills' ratio.

Remark 2.1. Arnold and Beaver (2002) wrote the density in (2.5) using a different parametrization:

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha_0 + \alpha_1\left(\frac{x-\mu}{\sigma}\right)\right)}{\sigma\Phi\left(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}\right)}, \qquad \text{where} \quad \alpha_0 \in \mathbb{R},\ \alpha_1 \in \mathbb{R}, \tag{2.7}$$

$$\text{with} \qquad \alpha_0 = \frac{\gamma}{\sqrt{1-\rho^2}} \qquad \text{and} \qquad \alpha_1 = \frac{\rho}{\sqrt{1-\rho^2}}.$$

Alternatively, we may as well use:

$$\rho = \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \qquad \text{and} \qquad \gamma = \frac{\alpha_0}{\sqrt{1+\alpha_1^2}}.$$

When $\gamma = 0$, the ESN distribution reduces to the $\mathrm{SN}(\mu, \sigma^2, \alpha_1)$ distribution introduced by Azzalini (1985), while for $\gamma = 0$ and $\rho = 0$ the normal distribution $N(\mu, \sigma^2)$ is obtained. The distribution of X is denoted $\mathrm{ESN}(\mu, \sigma, \gamma, \rho)$ or $\mathrm{ESN}(\mu, \sigma, \alpha_0, \alpha_1)$.

3. STOCHASTIC ORDERS

In this section, we recall the definitions and some properties of the stochastic orders that we will use in the sequel. These are mainly based on the following references: Shaked and Shanthikumar (2007) and Muller and Stoyan (2002).

Definition 3.1. Let X_1 and X_2 be two real random variables with cdf F_1 and F_2 and pdf f_1 and f_2, respectively.

- X_1 is said to be smaller than X_2 in the sense of the usual stochastic order (or smaller in distribution) and we denote $X_1 \leq_{\mathrm{st}} X_2$, if:

$$\bar{F}_1(x) \leq \bar{F}_2(x), \quad \forall x \in \mathbb{R}, \tag{3.1}$$

  where $\bar{F}_1, \bar{F}_2$ represent the survival functions of X_1 and X_2, respectively, i.e. $\bar{F}_1(x) = 1 - F_1(x)$ and $\bar{F}_2(x) = 1 - F_2(x)$, or equivalently, if:

$$F_1(x) \geq F_2(x), \quad \forall x \in \mathbb{R}. \tag{3.2}$$

  This stochastic or distributional order is known in economic theory as first-order stochastic dominance.

- X_1 is said to be smaller than X_2 in the sense of the likelihood ratio order and we denote $X_1 \leq_{\mathrm{lr}} X_2$ if:

$$\frac{f_2(x)}{f_1(x)} \quad \text{is increasing in } x, \ \forall x \in I; \tag{3.3}$$

  where I is the union of the supports of X_1 and X_2.

- X_1 is said to be smaller than X_2 in the sense of increasing convex order and we denote $X_1 \leq_{\mathrm{icx}} X_2$ if:

$$\int_y^\infty \bar{F}_1(x)dx \leq \int_y^\infty \bar{F}_2(x)\,dx, \quad \forall y \in \mathbb{R}. \tag{3.4}$$

- X_1 is said to be smaller than X_2 in the sense of increasing concave order and we denote $X_1 \leq_{\mathrm{icv}} X_2$ if:

$$\int_{-\infty}^y F_1(x)dx \geq \int_{-\infty}^y F_2(x)\,dx, \quad \forall y \in \mathbb{R}. \tag{3.5}$$

  The increasing concave stochastic order is also known in the literature as second-order stochastic dominance.

- X_1 is smaller than X_2 in the sense of the less dangerous order, denoted by $X_1 \leq_{\mathrm{D}} X_2$, if:
  - $\exists t_0 \in \mathbb{R}$ such that $F_1(t) \leq F_2(t)$, $\forall t < t_0$ and $F_1(t) \geq F_2(t)$, $\forall t \geq t_0$;
  - $E(X_1) \leq E(X_2)$.

The following proposition connects these stochastic orders.

Proposition 3.1. *The above partial orders verify the following implications:*

- $X_1 \leq_{\mathrm{lr}} X_2 \Rightarrow X_1 \leq_{\mathrm{st}} X_2 \Rightarrow X_1 \leq_{\mathrm{icv}} X_2$;
- $X_1 \leq_{\mathrm{st}} X_2 \Rightarrow X_1 \leq_{\mathrm{icx}} X_2$;
- $X_1 \leq_{\mathrm{D}} X_2 \Rightarrow X_1 \leq_{\mathrm{icx}} X_2$.

Several properties stem from these definitions. We give some of them in the following:

Proposition 3.2.

1. *$X_1 \leq_{\mathrm{st}} X_2$ if and only if there is a positive Y random variable such that $X_2 \overset{d}{=} X_1 + Y$;*
2. *If $X_1 \leq_{\mathrm{st}} X_2$ then $E(X_1) \leq E(X_2)$;*
3. *If $X_1 \leq_{\mathrm{st}} X_2$ and $E(X_1) = E(X_2)$, then X_1 et X_2 have the same distribution, i.e. $F_1 = F_2$;*

4. ORDERING OF $\mathrm{SUN}_{(1,q)}$ DISTRIBUTIONS

In this section, we study the stochastic ordering of the $\mathrm{SUN}_{(1,q)}$ distributions for the orders defined previously and relatively to each of its parameters, assuming that the others are held constant.

Choosing $p = 1$ in (2.1), we obtain the $\mathrm{SUN}_{(1,q)}$ density:

$$f_X(x) = \phi(x-\mu;\sigma^2)\,\frac{\Phi_q\big(\boldsymbol{\gamma} + \boldsymbol{\Delta}^t(\frac{x-\mu}{\sigma});\boldsymbol{\Gamma} - \boldsymbol{\Delta}^t\boldsymbol{\Delta}\big)}{\Phi_q(\boldsymbol{\gamma};\boldsymbol{\Gamma})}, \qquad x \in \mathbb{R}. \tag{4.1}$$

where $\mu \in \mathbb{R}$, $\Omega = \sigma^2 \in \mathbb{R}_+^*$, $\boldsymbol{\gamma} \in \mathbb{R}^q$ and $\boldsymbol{\Delta} \in \mathbb{R}^q$.

Now, we state the main result of our work.

Theorem 4.1. *Let X_1 and X_2 be two random variables with pdf f_1 and f_2 and cdf F_1 and F_2, respectively. We have:*

1. *If $X_1 \sim \mathrm{SUN}_{(1,q)}(\mu_1, \sigma, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$, $X_2 \sim \mathrm{SUN}_{(1,q)}(\mu_2, \sigma, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ and if $\mu_1 \leq \mu_2$ then:*

$$X_1 \leq_{\mathrm{lr}} X_2.$$

2. *If* $X_1 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ *and* $X_2 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma, \boldsymbol{\gamma}', \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ *with* $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{i-1}, \gamma_i, \gamma_{i+1}, ..., \gamma_q)$ *and* $\boldsymbol{\gamma}' = (\gamma_1, ..., \gamma_{i-1}, \gamma_i', \gamma_{i+1}, ..., \gamma_q)$ *and if* $\gamma_i \leq \gamma_i'$ *then:*

$$X_1 \geq_{\mathrm{lr}} X_2, \quad \textit{for } \boldsymbol{\Delta} \geq \mathbf{0};$$

$$X_1 \leq_{\mathrm{lr}} X_2, \quad \textit{for } \boldsymbol{\Delta} \leq \mathbf{0}.$$

3. *If* $X_1 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ *and* $X_2 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta}')$ *with* $\boldsymbol{\Delta} = (\delta_1, ..., \delta_{i-1}, \delta_i, \delta_{i+1}, ..., \delta_q)$ *and* $\boldsymbol{\Delta}' = (\delta_1, ..., \delta_{i-1}, \delta_i', \delta_{i+1}, ..., \delta_q)$, *and if* $\delta_i \leq \delta_i'$ *then:*

$$X_1 \leq_{\mathrm{st}} X_2.$$

4. *If* $X_1 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma_1, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$, $X_2 \sim \mathrm{SUN}_{(1,q)}(\mu, \sigma_2, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ *and if* $\sigma_1 \leq \sigma_2$ *then:*

$$X_1 \leq_{\mathrm{icx}} X_2, \quad \textit{when } \boldsymbol{\Delta} \geq \mathbf{0};$$

$$X_1 \geq_{\mathrm{icv}} X_2, \quad \textit{when } \boldsymbol{\Delta} \leq \mathbf{0}.$$

Proof:

1. Arellano-Valle and Azzalini (2022) have established the log-concavity of the density $\mathrm{SUN}_{(p,q)}$. Moreover, we know that if g is a log-concave density in $\mathbb{R}$, then the family $g(x-\theta)$ has a monotone likelihood ratio with respect to θ (see Dharmadhikari and Joag-Dev, 1988), i.e.:

$$\text{If } \theta_1 < \theta_2, \ \frac{g(x-\theta_2)}{g(x-\theta_1)} \quad \text{is an increasing monotone function of } x.$$

As a result, for all $\mu_1 < \mu_2$ the ratio

$$\frac{f_2(x)}{f_1(x)} = \frac{f(x-\mu_2, 0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})}{f(x-\mu_1, 0, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta})}$$

is an increasing monotone function of x. ■

2. We have

$$\frac{f_2(x)}{f_1(x)} = \frac{\phi(\frac{x-\mu}{\sigma})\, \Phi_q\big(\boldsymbol{\gamma}' + \boldsymbol{\Delta}^t(\frac{x-\mu}{\sigma}), \boldsymbol{\Gamma} - \boldsymbol{\Delta}^t\boldsymbol{\Delta}\big)}{\phi(\frac{x-\mu}{\sigma})\, \Phi_q\big(\boldsymbol{\gamma} + \boldsymbol{\Delta}^t(\frac{x-\mu}{\sigma}), \boldsymbol{\Gamma} - \boldsymbol{\Delta}^t\boldsymbol{\Delta}\big)} \frac{\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})}{\Phi_q(\boldsymbol{\gamma}', \boldsymbol{\Gamma})}.$$

Note that $\frac{\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})}{\Phi_q(\boldsymbol{\gamma}', \boldsymbol{\Gamma})}$ is a positive constant independent of x.

Consider the function g given by:

$$g(z) = \frac{\Phi_q(\boldsymbol{\gamma}' + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\Phi_q(\boldsymbol{\gamma} + \boldsymbol{\Delta}z), \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}, \quad \text{where } z = \frac{x-\mu}{\sigma}.$$

We have

$$\frac{dg(z)}{dz} = \frac{\frac{d\Phi_q(\boldsymbol{\gamma}' + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{dz} \Phi_q(\boldsymbol{\gamma} + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t) - \frac{d\Phi_q(\boldsymbol{\gamma} + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{dz} \Phi_q(\boldsymbol{\gamma}' + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\Phi_q^2(\boldsymbol{\gamma} + \boldsymbol{\Delta}z, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}.$$

Let $\boldsymbol{u} = \boldsymbol{\gamma} + \boldsymbol{\Delta}z$ and $\boldsymbol{u}' = \boldsymbol{\gamma}' + \boldsymbol{\Delta}z$. The derivative may then be written as follows:

$$\frac{dg(z)}{dz} = \frac{\sum_{j=1}^{q} \frac{du_j'}{dz} \frac{\partial \Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\partial u_j} \Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t) - \sum_{j=1}^{q} \frac{du_j}{dz} \frac{\partial \Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\partial u_j} \Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\Phi_q^2(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}.$$

Moreover, Φ_q has a decreasing reversed hazard rate (DRHR) since it is log-concave (see Ma, 2000), i.e.:

$$\frac{\partial \ln \Phi_q(u_1, u_2, ..., u_q)}{\partial u_j} \quad \text{is decreasing in } u_j, \quad \forall j = \overline{1,q}.$$

For any $\boldsymbol{\Delta} \geq \mathbf{0}$, it holds then

$$\begin{aligned}
&\sum_{j=1}^{q} \delta_j \frac{\partial\Big(\ln \Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)\Big)}{\partial u_j} \leq \sum_{j=1}^{q} \delta_j \frac{\partial\Big(\ln \Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)\Big)}{\partial u_j}, \\
\Leftrightarrow \quad &\sum_{j=1}^{q} \delta_j \frac{\frac{\partial \Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\partial u_j}}{\Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)} \leq \sum_{j=1}^{q} \delta_j \frac{\frac{\partial \Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}{\partial u_j}}{\Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)}, \\
\Leftrightarrow \quad &\frac{dg(z)}{dz} \leq 0, \ \forall z \in \mathbb{R}.
\end{aligned}$$

Similarly, for any $\boldsymbol{\Delta} \leq \mathbf{0}$, it holds then

$$\sum_{j=1}^{q} \delta_j \frac{\partial\Big(\ln \Phi_q(\boldsymbol{u}', \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)\Big)}{\partial u_j} \geq \sum_{j=1}^{q} \delta_j \frac{\partial\Big(\ln \Phi_q(\boldsymbol{u}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^t)\Big)}{\partial u_j},$$

which is equivalent to:

$$\frac{dg(z)}{dz} \geq 0, \quad \forall z \in \mathbb{R}.$$

In conclusion, if $\gamma_i \leq \gamma_i', \quad \forall i = \overline{1,q}$, we have:

$$\begin{aligned}
&X_1 \geq_{\text{lr}} X_2, \quad \text{for } \boldsymbol{\Delta} \geq \mathbf{0} \\
\text{and} \quad &X_1 \leq_{\text{lr}} X_2, \quad \text{for } \boldsymbol{\Delta} \leq \mathbf{0}.
\end{aligned}$$

■

3. Without loss of generality, we take $i = 1$ in the proof.

Referring to Azzalini and Bacchieri (2010), the cdf of X_1 and X_2 can be written as follows:

$$F_1(x) = \frac{\Phi_{1+q}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\Omega}}_1)}{\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})}, \qquad F_2(x) = \frac{\Phi_{1+q}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\Omega}}_2)}{\Phi_q(\boldsymbol{\gamma}, \boldsymbol{\Gamma})},$$

where

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} \frac{x-\mu}{\sigma} \\ \boldsymbol{\gamma} \end{pmatrix}, \quad \tilde{\boldsymbol{\Omega}}_1 = \begin{pmatrix} 1 & -\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}^t & \boldsymbol{\Gamma} \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{\Omega}}_2 = \begin{pmatrix} 1 & -\boldsymbol{\Delta}' \\ -\boldsymbol{\Delta}'^t & \boldsymbol{\Gamma} \end{pmatrix}.$$

On the other hand, we know, from the Slepian's inequality (Tong, 1990), that: if $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma} = (\sigma_{ij}))$ and $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}' = (\sigma'_{ij}))$ with $\sigma_{ii} = \sigma'_{ii}\ \forall i = \overline{1,n}$ and $\sigma_{ij} \geq \sigma'_{ij}\ \forall i \neq j$ then:

$$\Phi_n(\boldsymbol{a}, \boldsymbol{\Sigma}) \geq \Phi_n(\boldsymbol{a}, \boldsymbol{\Sigma}'), \quad \forall \boldsymbol{a} \in \mathbb{R}^n.$$

In our case, if $\delta_1 \leq \delta_1'$, we deduce that:

$$\Phi_{1+q}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\Omega}}_1) \geq \Phi_{1+q}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\Omega}}_2),$$

and then,

$$F_1(x) \geq F_2(x), \quad \forall x \in \mathbb{R}.$$

■

4. Using the following notation:

$$K(\sigma) = \int_{-\infty}^{x} F(t)dt,$$

where F is the $\mathrm{SUN}_{(1,q)}$ cdf, we have:

$$\begin{aligned}
\frac{dK(\sigma)}{d\sigma} &= \frac{d}{d\sigma}\left(\int_{-\infty}^{x} F(t)\,dt\right),\\
&= \frac{d}{d\sigma}\left[\int_{-\infty}^{x}\left(\int_{-\infty}^{\frac{t-\mu}{\sigma}} \frac{\phi(z)\,\Phi_q(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t z, \boldsymbol{\Gamma}-\boldsymbol{\Delta}^t\boldsymbol{\Delta})}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\,dz\right)dt\right],\\
&= \frac{1}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\int_{-\infty}^{x}\frac{d}{d\sigma}\left(\int_{-\infty}^{\frac{t-\mu}{\sigma}} \phi(z)\,\Phi_q(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t z, \boldsymbol{\Gamma}-\boldsymbol{\Delta}^t\boldsymbol{\Delta})\,dz\right)dt,\\
&= \frac{1}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\int_{-\infty}^{x} -\frac{(t-\mu)}{\sigma^2}\,\phi\left(\frac{t-\mu}{\sigma}\right)\Phi_q\left(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t\left(\frac{t-\mu}{\sigma}\right), \boldsymbol{\Gamma}-\boldsymbol{\Delta}^t\boldsymbol{\Delta}\right)dt.
\end{aligned}$$

Integrating by parts and using the change of variable $u = \frac{t-\mu}{\sigma}$, we get:

$$\frac{dK(\sigma)}{d\sigma} = \frac{1}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\int_{-\infty}^{\frac{(x-\mu)}{\sigma}} -u\,\phi(u)\,\Phi_q(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t u, \boldsymbol{\Psi})\,du,$$

where $\boldsymbol{\Psi} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}^t\boldsymbol{\Delta}$.

Noting that $\phi(u)' = -u\phi(u)$, we get:

$$K'(\sigma) = \frac{1}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\left[\phi\left(\frac{x-\mu}{\sigma}\right)\Phi_q\left(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t\left(\frac{x-\mu}{\sigma}\right), \boldsymbol{\Psi}\right)\right] - \int_{-\infty}^{\frac{x-\mu}{\sigma}}\phi(u)\sum_{j=1}^{q}\delta_j(\boldsymbol{\nabla}\Phi_q)_j\,du,$$

where $(\boldsymbol{\nabla}\Phi_q)_j$ is the j-th element of the gradient vector of $\Phi_q(\cdot,\boldsymbol{\Psi})$ at the point $\boldsymbol{\gamma}+\boldsymbol{\Delta}^t u$. Recall that each gradient component of $\Phi_q(\cdot,\cdot)$ is positive.

We deduce that $K'(\sigma) \geq 0$ if $\boldsymbol{\Delta} \leq \mathbf{0}$. Thus, $X_1 \geq_{\mathrm{icv}} X_2$.

If $\boldsymbol{\Delta} \geq \mathbf{0}$, we can show that $X_1 \leq_{\mathrm{D}} X_2$, which according to Proposition 3.1, will lead to $X_1 \leq_{\mathrm{icx}} X_2$.

1) If $\sigma_1 < \sigma_2$ and $t < \mu$, then we have:

$$\int_{-\infty}^{\frac{t-\mu}{\sigma_1}}\phi(z)\,\frac{\Phi_q(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t z, \boldsymbol{\Gamma}-\boldsymbol{\Delta}^t\boldsymbol{\Delta})}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\,dz \leq \int_{-\infty}^{\frac{t-\mu}{\sigma_2}}\phi(z)\,\frac{\Phi_q(\boldsymbol{\gamma}+\boldsymbol{\Delta}^t z, \boldsymbol{\Gamma}-\boldsymbol{\Delta}^t\boldsymbol{\Delta})}{\Phi_q(\boldsymbol{\gamma},\boldsymbol{\Gamma})}\,dz,$$

This is true because it is an integral of a product of positive functions which is positive and increasing with respect to the bound. Thus we get:

$$F_1(t) \leq F_2(t), \quad \forall t < \mu, \quad \text{and} \quad F_1(t) \geq F_2(t), \quad \forall t \geq \mu. \tag{4.2}$$

2) Referring to equation (2.4), each component of the gradient vector $\nabla\Phi_q$ being positive, we conclude that: if $\sigma_1 < \sigma_2$ and $\boldsymbol{\Delta} \geq \mathbf{0}$, we have: $E(X_1) \leq E(X_2)$.

Thus, both conditions for the "D" order are verified. We deduce that $X_1 \leq_{\mathrm{icx}} X_2$. □

Remark 4.1. From Theorem 4.1, we deduce, that: if $\boldsymbol{\Delta} = (\delta_1, \delta_2, ..., \delta_q)$ and $\boldsymbol{\Delta}' = (\delta'_1, \delta'_2, ..., \delta'_q)$ with $\boldsymbol{\Delta}' \leq \boldsymbol{\Delta}$, then $X_1 \leq_{\text{st}} X_2$. Similarly, if $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_q)$ and $\boldsymbol{\gamma}' = (\gamma'_1, \gamma'_2, ..., \gamma'_q)$ with $\boldsymbol{\gamma} \leq \boldsymbol{\gamma}'$ and if $\boldsymbol{\Delta} \geq \mathbf{0}$ ($\boldsymbol{\Delta} \leq \mathbf{0}$), then $X_1 \geq_{\text{lr}} X_2$ ($X_1 \leq_{\text{lr}} X_2$ respectively).

The following results on the stochastic ordering of ESN distributions stem immediately from those established for $\text{SUN}_{(1,q)}$ distributions.

Corollary 4.1. *Let X_1 and X_2 be two random variables of ESN distribution. We have:*

1. *If $X_1 \sim \text{ESN}(\mu_1, \sigma, \gamma, \rho)$ and $X_2 \sim \text{ESN}(\mu_2, \sigma, \gamma, \rho)$ with $\mu_1 \leq \mu_2$, then:*

$$X_1 \leq_{\text{lr}} X_2.$$

2. *If $X_1 \sim \text{ESN}(\mu, \sigma, \gamma_1, \rho)$ and $X_2 \sim \text{ESN}(\mu, \sigma, \gamma_2, \rho)$ with $\gamma_1 \leq \gamma_2$, then:*

$$X_1 \geq_{\text{lr}} X_2, \quad \text{for} \quad \rho \geq 0;$$
$$X_1 \leq_{\text{lr}} X_2, \quad \text{for} \quad \rho \leq 0.$$

3. *If $X_1 \sim \text{ESN}(\mu, \sigma, \gamma, \rho_1)$ and $X_2 \sim \text{ESN}(\mu, \sigma, \gamma, \rho_2)$ with $\rho_1 \leq \rho_2$, then:*

$$X_1 \leq_{\text{st}} X_2.$$

4. *If $X_1 \sim \text{ESN}(\mu, \sigma_1, \gamma, \rho)$ and $X_2 \sim \text{ESN}(\mu, \sigma_2, \gamma, \rho)$ with $\sigma_1 \leq \sigma_2$, then:*

$$X_1 \leq_{\text{icx}} X_2, \quad \text{for} \quad \rho \geq 0;$$
$$X_1 \geq_{\text{icv}} X_2, \quad \text{for} \quad \rho \leq 0.$$

Remark 4.2. By adopting the parametrization $\text{ESN}(\mu, \sigma, \alpha_0, \alpha_1)$, each of the following results is a direct consequence of Corollary 4.1:

1. If $X_1 \sim \text{ESN}(\mu, \sigma, \alpha_0, \alpha_1)$, $X_2 \sim \text{ESN}(\mu, \sigma, \alpha'_0, \alpha_1)$ and if $\alpha_0 \leq \alpha'_0$, then we have:

$$X_1 \geq_{\text{lr}} X_2, \quad \text{for} \ \alpha_1 \geq 0;$$
$$X_1 \leq_{\text{lr}} X_2, \quad \text{for} \ \alpha_1 \leq 0.$$

2. If $X_1 \sim \text{ESN}(\mu, \sigma, \alpha_0, \alpha_1)$, $X_2 \sim \text{ESN}(\mu, \sigma, \alpha_0, \alpha'_1)$ and if $\alpha_1 \leq \alpha'_1$, then we have:

$$X_1 \leq_{\text{st}} X_2.$$

5. APPLICATIONS

In this section, we apply the previous results to two practical problems: the lifetime of a system in reliability and a selection problem encountered in education.

5.1. Lifetime of a system

Let $X_1, X_2, ..., X_n$ be the lifetimes of the n components of a system. Let us denote by Z the lifetime of the system distributed in parallel and U the lifetime of the system distributed in series. We have:

$$Z = \max(X_1, X_2, ..., X_n),$$
$$U = \min(X_1, X_2, ..., X_n).$$

Assume that $\boldsymbol{X} = (X_1, X_2, ..., X_n)^t$ follows an exchangeable normal distribution $N_n(\boldsymbol{\mu}^*, \boldsymbol{\Omega})$ with $\boldsymbol{\mu}^* = \mu\, \mathbf{1}_{(n)}$, where $\mathbf{1}_{(n)} = (1, 1, ..., 1)^t \in \mathbb{R}^n$, $\boldsymbol{\Omega} = \left(\sigma^2 \rho_{ij}\right)_{1 \le i \le j \le n}$ and ρ_{ij} is given as follow:

$$\rho_{ij} = \begin{cases} \rho, & \text{if } \ i \neq j, \ \forall i, j \in \{1, ...n\}, \\ 1, & \text{if } \ i = j. \end{cases}$$

The cdf of Z is given by:

$$\begin{aligned} F_Z(z) \ &= \ P\Big(\max(X_1, X_2, ..., X_n) \le z\Big) \\ &= \sum_{i=1}^{n} P\left(X_i \le z \,|\, \bigcap_{\substack{j=1 \\ j \neq i}}^{n} \big\{X_j \le X_i\big\} \right) P\left(\bigcap_{\substack{j=1 \\ j \neq i}}^{n} \big\{X_j \le X_i\big\} \right). \end{aligned}$$

The assumption of exchangeability implies that:

$$X_i \,|\, \bigcap_{\substack{j=1 \\ j \neq i}}^{n} \big\{X_i - X_j \ge 0\big\} \sim \mathrm{SUN}_{(1,n-1)}(\mu, \mathbf{0}_{\mathbb{R}^{n-1}}, \boldsymbol{\Omega}^*), \quad \forall i = \overline{1, n};$$

with $\boldsymbol{\Omega}^* = \sigma^2(1-\rho)[\boldsymbol{A}_{(n)} + \mathbf{1}_{(n)}\mathbf{1}_{(n)}^t]$ and $\boldsymbol{A}_{(n)}$ given by

$$\boldsymbol{A}_{(n)} = \begin{pmatrix} \frac{\rho}{1-\rho} & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n.$$

We deduce that: $Z \sim \mathrm{SUN}_{(1,n-1)}(\mu, \mathbf{0}_{\mathbb{R}^{n-1}}, \boldsymbol{\Omega}^*)$ because

$$\sum_{i=1}^{n} P\left(\bigcap_{\substack{j=1 \\ j \neq i}}^{n} \big\{X_j \le X_i\big\} \right) = 1.$$

The associated correlation matrix $\bar{\boldsymbol{\Omega}}^*$ is then:

$$\bar{\boldsymbol{\Omega}}^* = \begin{pmatrix} 1 & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}^t & \boldsymbol{\Gamma} \end{pmatrix}, \tag{5.1}$$

where $\boldsymbol{\Delta} = \sqrt{\frac{1-\rho}{2}}\ \mathbf{1}_{(n-1)}^t$, $\boldsymbol{\Gamma} = \frac{1}{2}\boldsymbol{I}_{(n-1)} + \frac{1}{2}\mathbf{1}_{(n-1)}\mathbf{1}_{(n-1)}^t$ and $\boldsymbol{I}_{(n-1)}$ is the identity matrix of order $n-1$.

On the other hand, we have $U = -\max(-X_1, -X_2, ..., -X_n)$. We deduce that $U \sim \text{SUN}_{(1,n-1)}(\mu, \mathbf{0}_{\mathbb{R}^{n-1}}, \mathbf{\Omega}^{**})$ where $\mathbf{\Omega}^{**} = \sigma^2(1-\rho)[\boldsymbol{B}_{(n)} + \mathbf{1}_{(n)}\mathbf{1}_{(n)}^t]$ and $\boldsymbol{B}_{(n)}$ is the following matrix:

$$\boldsymbol{B}_{(n)} = \begin{pmatrix} \frac{\rho}{1-\rho} & -2 & \cdots & \cdots & -2 \\ -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -2 & 0 & \cdots & 0 & 1 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n.$$

From (5.1), we obtain the associated correlation matrix below:

$$\bar{\mathbf{\Omega}}^{**} = \begin{pmatrix} 1 & -\mathbf{\Delta} \\ -\mathbf{\Delta}^t & \mathbf{\Gamma} \end{pmatrix}; \tag{5.2}$$

The results of Theorem 4.1 allow us to conclude that:

- When all other parameters are held constant, the survival function is an increasing function of μ for both variables Z and U.
- When all other parameters are held constant, the survival function decreases with ρ for the variable Z and increases with ρ for the variable U.
- From equation (4.2), we deduce that the survival function of the variable Z increases with σ for $z \geq \mu$ while it decreases with σ for $z < \mu$. The same result holds for the variable U.

Corollary 5.1. *Let Z_1 and Z_2 be the lifetimes of two parallel systems which are characterized by the following parameters: μ_1, ρ_1 and σ^2 for Z_1 and μ_2, ρ_2 and σ^2 for Z_2. The two systems have the same lifetime if and only if:*

$$\mu_1 - \mu_2 = \sigma\left(\sqrt{\frac{1-\rho_2}{2}} - \sqrt{\frac{1-\rho_1}{2}}\right)\mathbf{1}_{(n-1)}^t \frac{\nabla\Phi_{n-1}(\mathbf{0}_{\mathbb{R}^{n-1}}, \mathbf{\Gamma})}{\Phi_{n-1}(\mathbf{0}_{\mathbb{R}^{n-1}}, \mathbf{\Gamma})}. \tag{5.3}$$

Proof: It is obvious that if $F_{Z_1}(z) = F_{Z_2}(z)$ then equation (5.3) holds and therefore $(\mu_1 - \mu_2)(\rho_1 - \rho_2) < 0$.

Now, for the sufficient condition, let Z_1' be the lifetime of a parallel system which is characterized by the parameters μ_2, ρ_1 and σ^2.

From the results of Theorem 4.1, we have:

$$\mu_1 \leq \mu_2 \Rightarrow Z_1 \leq_{\text{st}} Z_1', \tag{5.4}$$

$$\rho_1 \geq \rho_2 \Rightarrow Z_1' \leq_{\text{st}} Z_2. \tag{5.5}$$

Thus, from (5.4) and (5.5), we find that:

$$\rho_1 \geq \rho_2 \text{ and } \mu_1 \leq \mu_2 \quad \Rightarrow \quad Z_1 \leq_{\text{st}} Z_2.$$

According to the last property of Proposition 3.2, to have $F_{Z_1}(z) = F_{Z_2}(z)$, it is sufficient that $E(Z_1) = E(Z_2)$. This is equivalent to:

$$\mu_1 - \mu_2 = \sigma\left(\sqrt{\frac{1-\rho_2}{2}} - \sqrt{\frac{1-\rho_1}{2}}\right)\mathbf{1}_{(n-1)}^t \frac{\nabla\Phi_{n-1}(\mathbf{0}_{\mathbb{R}^{n-1}}, \mathbf{\Gamma})}{\Phi_{n-1}(\mathbf{0}_{\mathbb{R}^{n-1}}, \mathbf{\Gamma})}. \qquad \square$$

5.2. Selection problems

We now consider selection problems which have been studied in particular by Birnbaum (1950) and Birnbaum and Chapman (1950). These problems invoke the so-called selection distributions (Arellano-Valle *et al.*, 2006) which reduce in some cases to SUN distributions.

Equation (2.2) can be interpreted as the selection of individuals in a population using the variable of interest $\boldsymbol{U}$ under the constraint $\boldsymbol{V} > \boldsymbol{0}$, where $\boldsymbol{V}$ is the truncation variable. The distribution of $\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{U}|\boldsymbol{V} > \boldsymbol{\tau}$, $\boldsymbol{\tau} \in \mathbb{R}^q$, is called a selection distribution and when the vector $(\boldsymbol{U}^t, \boldsymbol{V}^t)^t$ follows a normal distribution, we recover the SUN distribution.

For example, U can stand for a student's baccalaureate grade and V for his Math grade. In this case, since $p = 1$, $q = 1$ and under the assumption of normality made on the vector $(\boldsymbol{U}^t, \boldsymbol{V}^t)^t$, Z follows the ESN distribution. According to (2.5), the pdf of Z is:

$$f_Z(z) = \frac{1}{\sigma}\,\phi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\frac{\rho\left(\frac{z-\mu}{\sigma}\right)+\tilde{\gamma}}{\sqrt{1-\rho^2}}\right)\Big/\Phi(\tilde{\gamma}),$$

where $\tilde{\gamma} = -(\frac{\tau-\gamma}{\sigma'})$ and σ' is the scale parameter of V. Then $Z \sim \mathrm{ESN}(\mu, \sigma^2, \rho, \tilde{\gamma})$.

We select the individuals according to the grade obtained at the baccalaureate exam under the condition that the Math grade exceeds τ. The proportion of students retained (admission rate) is equal to:

$$\begin{aligned} P(Z \geq z_\alpha) &= P(U \geq u_\alpha | V > \tau), \\ &= \alpha. \end{aligned} \tag{5.6}$$

In this example, V can be a vector, if for instance we consider the vector $\boldsymbol{V} = (V_1, V_2)$ of Math and Physics grades. In this case, we have:

$$Z \stackrel{d}{=} U\,|\,V_1 > \tau_1,\, V_2 > \tau_2.$$

Under the assumption of normality made on the vector $(\boldsymbol{U}, \boldsymbol{V}^t)^t$, $Z \sim \mathrm{SUN}_{(1,2)}\big(\mu, \sigma^2, \boldsymbol{\Delta}, \tilde{\boldsymbol{\gamma}}, \boldsymbol{\Gamma}\big)$ with $\tilde{\boldsymbol{\gamma}} = -\boldsymbol{\sigma}'^{-1}(\boldsymbol{\tau} - \boldsymbol{\gamma})$. Here, $\boldsymbol{\sigma}'$ is the diagonal matrix of order 2 of elements σ'_1, σ'_2 which correspond to the standard deviations of V_1 and V_2 respectively.

In the first case ($p = 1$, $q = 1$), we are interested in studying the variation of the proportion of students retained as a function of some parameters of Z which are μ, σ^2, ρ and $\tilde{\gamma}$, the other parameters being held constant. Note that γ and σ' are only involved through the truncation parameter $\tilde{\gamma}$. The results stated in Corollary 4.1 imply that:

1. The admission rate is an increasing function of μ when the other parameters are held constant, i.e.:

$$\mu_1 \leq \mu_2 \;\Rightarrow\; \bar{F}_1(z) \leq \bar{F}_2(z), \quad \forall z \in \mathbb{R}$$

2. The admission rate decreases with τ if $\rho \leq 0$ and increases with τ if $\rho \geq 0$ when the other parameters are held constant, i.e.:

$$\tau_1 \geq \tau_2 \Rightarrow \begin{cases} \rho \geq 0, & \bar{F}_1(z) \geq \bar{F}_2(z), \quad \forall z \in \mathbb{R}; \\ \rho \leq 0, & \bar{F}_1(z) \leq \bar{F}_2(z), \quad \forall z \in \mathbb{R}. \end{cases}$$

3. The admission rate increases with ρ when the other parameters are held constant, i.e.:

$$\rho_1 \leq \rho_2 \Rightarrow \bar{F}_1(z) \leq \bar{F}_2(z), \quad \forall z \in \mathbb{R}.$$

Remark 5.1. The problem of determining the selection threshold τ for a given α when the other parameters are held constant can be solved numerically by expressing (5.6) using the bivariate normal distribution function (see Azzalini and Bacchieri, 2010, or Azzalini, 2014). This gives:

$$\frac{\Phi_2(\tilde{\boldsymbol{z}}_\alpha, -\rho)}{\Phi(\tilde{\gamma})} = 1 - \alpha, \qquad \text{where} \quad \tilde{\boldsymbol{z}}_\alpha = \left(\frac{z_\alpha - \mu}{\sigma}, \tilde{\gamma}\right)^t.$$

5.2.1. Equality of two selection distributions

Consider two Gaussian vectors such that:

$$\begin{pmatrix} U_1 \\ V_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \xi_1 \end{pmatrix}, \Omega_1^*\right) \quad \text{and} \quad \begin{pmatrix} U_2 \\ V_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_2 \\ \xi_2 \end{pmatrix}, \Omega_2^*\right).$$

We define $Z_1 \overset{d}{=} U_1 | V_1 > \tau_1$, $Z_2 \overset{d}{=} U_2 | V_2 > \tau_2$ and F_1, F_2 their respective cdf.

We know that $Z_1 \sim \text{ESN}(\mu_1, \sigma_1^2, \rho_1, \tilde{\gamma}_1)$ and $Z_2 \sim \text{ESN}(\mu_2, \sigma_2^2, \rho_2, \tilde{\gamma}_2)$.

We are interested to know under which conditions it holds that $\bar{F}_1(z) = \bar{F}_2(z)$, $\forall z \in \mathbb{R}$.

For instance, if we consider the same variables as those defined in the previous example with U_1 and V_1 corresponding to high school A, U_2 and V_2 corresponding to high school B, we may ask the question: "under which conditions will the admission rate in both high schools be the same?"

We propose to solve the problem when two parameters are held constant.

Corollary 5.2. *Let* $Z_1 \sim \text{ESN}(\mu_1, \sigma_1^2, \rho_1, \tilde{\gamma}_1)$ *and* $Z_2 \sim \text{ESN}(\mu_2, \sigma_2^2, \rho_2, \tilde{\gamma}_2)$. *Then* $\bar{F}_1(z) = \bar{F}_2(z)$ *if and only if one of the following holds:*

(**i**) *When* $\tilde{\gamma}_1 = \tilde{\gamma}_2 = \tilde{\gamma}$ *and* $\sigma_1 = \sigma_2 = \sigma$,

$$\mu_1 - \mu_2 = \sigma\lambda(\tilde{\gamma})(\rho_2 - \rho_1). \tag{5.7}$$

(**ii**) *When* $\mu_1 = \mu_2 = \mu$ *and* $\sigma_1 = \sigma_2 = \sigma$,

$$\frac{\lambda(\tilde{\gamma}_1)}{\lambda(\tilde{\gamma}_2)} = \frac{\rho_2}{\rho_1}. \tag{5.8}$$

(**iii**) *When* $\sigma_1 = \sigma_2 = \sigma$ *and* $\rho_1 = \rho_2 = \rho$,

$$\mu_1 - \mu_2 = \sigma\rho\big(\lambda(\tilde{\gamma}_2) - \lambda(\tilde{\gamma}_1)\big). \tag{5.9}$$

Proof: The proof is analogous to the one given for Corollary 5.1. □

Remark 5.2. Corollary 5.2 allows one to draw the following conclusions for the considered example:

- To prevent a change in the admission rate if the correlation ρ between the baccalaureate grade and the Math grade increases when τ and σ are held constant, it is necessary to lower the average baccalaureate grade.
- To get a similar admission rate in the two high schools when μ and σ are held constant, it is necessary to vary the selection threshold τ and ρ in opposite directions. This problem has been discussed by Birnbaum (1950) who established that the selection threshold τ is a decreasing function of $|\rho|$.
- To get a similar admission rate in both high schools when ρ and σ are held constant, it is necessary to vary the selection threshold τ and the average baccalaureate grade μ in opposite directions when ρ is positive and vary them in the same direction when ρ is negative .

6. CONCLUSION AND FUTURE WORK

In the present paper, we compare the univariate Unified Skew Normal distributions according to some classical criteria (usual stochastic order, increasing concave order, increasing convex order and the likelihood ratio order) and give two applications to both a reliability and a selection problems. A natural sequel of this work concerns the extension to the multivariate Unified Skew Normal family $\mathrm{SUN}_{(p,q)}$ with $p > 1$ and to the more general class of the multivariate Unified Skew Elliptical distributions $\mathrm{SUE}_{p,q}$ (Arellano-Valle and Azzalini, 2006). This requires the use of multivariate stochastic orders as defined, for example, in Shaked and Shanthikumar (2007). In this connection, Yin *et al.* (2024) recently considered the special case of the multivariate Skew Elliptical distributions (Azzalini and Capitanio, 2003) for some criteria: Hessian order, increasing Hessian order as well as many of their special cases.

As mentioned in the introduction, both $\mathrm{SUN}_{(p,q)}$ and $\mathrm{SUE}_{(p,q)}$ families introduce skewness, in addition $\mathrm{SUE}_{(p,q)}$ introduce kurtosis. So, it would be of great interest to compare the above distributions relatively to these features. In the literature, several skewness and kurtosis orderings and measures have been defined, among them the well-known convex transform order of Van Zwet (1964). For the univariate case, we can refer to Arnold and Groeneveld (1992) and MacGillivray (1986). Much less work has been devoted to the multivariate case. Belzunce *et al.* (2015) extended the convex transform order to the multivariate setting. For the particular case of skew-normal vectors, Arevalillo and Navarro (2019) have introduced a new multivariate skewness order based on the canonical transformation of these vectors. He also established that the univariate Skew Normal family is ordered for the skewness parameter α according to the convex transform order.

On the other hand, Loperfido (2015) revisited some usual measures of the multivariate skewness: Mardia's skewness (Mardia, 1970), partial skewness (Davis, 1980), directional skewness (Malkovich and Afifi, 1973) and established relationships between them. Later, Loperfido (2017) defined a new kurtosis matrix as alternative for the existing measures of multivariate kurtosis.

REFERENCES

Arellano-Valle, R. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3):561–574.

Arellano-Valle, R. and Azzalini, A. (2022). Some properties of the unified skew-normal distribution. *Statistical Papers*, 63(2):461–487.

Arellano-Valle, R., Branco, M., and Genton, M. (2006). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, 34(4):581–601.

Arevalillo, J. M. and Navarro, H. (2019). A stochastic ordering based on the canonical transformation of skew-normal vectors. *Test*, 28:475–498.

Arnold, B. and Beaver, R. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, 11(1):7–54.

Arnold, B. C. and Groeneveld, R. A. (1992). Skewness and kurtosis orderings: an introduction. *Lecture Notes – Monograph Series*, 22:17–24.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Azzalini, A. (2014). *The Skew-Normal and Related Families*. IMS monographs, Cambridge University Press, Cambridge, UK.

Azzalini, A. and Bacchieri, A. (2010). A prospective combination of phase ii and phase iii in drug development. *Metron*, LXVIII(3):347–369.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 65(2):367–389.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Belzunce, F., Mulero, J., Ruíz, J. M., and Suárez-Llorens, A. (2015). On relative skewness for multivariate distributions. *Test*, 24(4):813–834.

Birnbaum, Z. (1950). Effect of linear truncation on a multinormal population. *The Annals of Mathematical Statistics*, 21:272–279.

Birnbaum, Z. W. and Chapman, D. (1950). On optimum selections from multinormal populations. *The Annals of Mathematical Statistics*, 21(3):443–447.

Blasi, F. and Scarlatti, S. (2012). From normal vs skew-normal portfolios: fsd and ssd rules. *Journal of Mathematical Finance*, 2:90–95.

Davis, A. (1980). On the effects of moderate multivariate nonnormality on Wilks's likelihood ratio criterion. *Biometrika*, 67(2):419–427.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Probability and Mathematical Statistics: Unimodality, Convexity and Applications*. Academic Press, New York & London.

Loperfido, N. (2015). Singular value decomposition of the third multivariate moment. *Linear Algebra and its Applications*, 473:202–216.

Loperfido, N. (2017). A new kurtosis matrix, with statistical applications. *Linear Algebra and its Applications*, 512:1–17.

Loperfido, N., Navarro, J., Ruiz, J., and Sandoval, C. (2007). Some relationships between skew-normal distributions and order statistics from exchangeable normal random vectors. *Communication in Statistics – Theory and Methods*, 36(9):1719–1733.

Ma, C. (2000). A note on the multivariate normal hazard. *Journal of Multivariate Analysis*, 73:282–283.

MacGillivray, H. L. (1986). Skewness and asymmetry: measures and orderings. *Annals of Statistics*, 14(3):994–1011.

Malkovich, J. F. and Afifi, A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association*, 68(341):176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.

Muller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley, New York.

Shaked, J. and Shanthikumar, N. (2007). *Stochastic Orders*. Springer, New York.

Tong, Y. (1990). *The Multivariate Normal Distribution*. Springer, New York.

Van Zwet, W. R. (1964). *Convex Transformations of Random Variables*. Mathematisch Centrum, Amsterdam.

Yin, C., Yao, J., and Yang, Y. (2024). Hessian and increasing-Hessian orderings of multivariate skew-elliptical random vectors with applications in actuarial science. *Statistical Papers*, 65:4715–4744. https://doi.org/10.1007/s00362-024-01580-y

REVSTAT – Statistical Journal
Volume 23, Number 4, October 2025, 543–567
https://doi.org/10.57805/revstat.v23i4.647

Fisher Information Matrix for Two-Way Random Effects Model with Heteroscedasticity

Authors: Patrice Takam Soh ✉
– Department of Mathematics, University of Yaoundé 1, Yaoundé, Cameroon
patricetakam@gmail.com

Eugene Kouassi
– Department of Economics, University FHB, Abidjan, Cote-d'Ivoire
eugene.kouassi@gmail.com

Jean Marcelin Bosson Brou
– Department of Economics, University FHB, Abidjan, Cote-d'Ivoire
jmbbrou@gmail.com

Saralees Nadarajah
– Department of Mathematics, University of Manchester, Manchester, UK
mbbsssn2@manchester.ac.uk

Received: November 2021 Revised: January 2024 Accepted: January 2024

Abstract:

- Model specification and selection are important aspects of modeling exercises. In this context, the Fisher Information Matrix (FIM) plays an essential role. In this paper, we derive the Fisher Information Matrix (FIM) for the two way random effects panel data models in general as well as in some specific cases of heteroscedasticity. Some computational issues are then raised and discussed. In addition, some real data examples are reported and thoroughly discussed.

Keywords:

- *computational issues; heteroscedasticity; information matrix; real data applications; two way models.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author

1. INTRODUCTION

Does the structure of the Fisher Information Matrix (FIM, hereafter) matter in model specification and selection? Does the FIM matter in testing and/or in inference analysis particularly in the two way random effects panel data model? Which FIM results are more relevant or appealing? Observed or exact FIM results?

This paper seeks to get the FIM of the two-way random effects panel data model in the absence or presence of heteroscedasticity. The approach developed follows that in Baltagi (2021), Baltagi and Griffin (1988), Baltagi *et al.* (2006), Holly and Gardiol (2000), Kouassi *et al.* (2014), Lejeune (2006), Nadarajah (2006a) and Randolph (1988). The FIM crucially depends on the variance covariance matrix. This matrix is obtained in four cases (homoscedasticity, heteroscedasticity on the unobservable individual effect, heteroscedasticity on the composite term and heteroscedasticity on both individual and composite terms); the paper does not consider groupwise heteroscedasticity or group membership heteroscedasticity as well as cluster heteroscedasticity (see Baum, 2006; Feng *et al.*, 2020; and Kouassi *et al.*, 2014) for example on the unobservable time effect.

This paper rather focuses on an alternative simple procedure for obtaining the FIM that accounts for homoscedasticity and/or various heteroscedasticity schemes in the two-way random effect model. It proposes a case-by-case approach rather than an elaborated sequence of steps and built-in functions used by earlier researchers.

The contributions of this paper are therefore twofold: (i) the derivation of the FIM based on different forms of homoscedasticity and/or heteroscedasticity; an important aspect in model specification; (ii) and thereby the exploration of how to choose the correct model specification in this context.

To do that, we develop a new and efficient procedure for computing the FIM in the two-way random effect model; the new procedure is obtained under homoscedasticity as well as well as various cases of heteroscedasticity.

The remainder of the paper is organized as follows: Section 2 describes the mathematical problem to be addressed. Section 3 sets out some preliminary results. Section 4 presents the main results while Section 5 discusses some computational issues. Section 6 provides two real data examples with discussions. Section 7 concludes the paper.

2. THE MATHEMATICAL PROBLEM

This section deals with the mathematical problem to address and some background information.

2.1. The FIM and related problem

We are interested in the derivation and computation of the FIM for the two-way random effects model (commonly encountered in theoretical as well as empirical studies) in the presence of various forms of heteroscedasticity.

Let $Y = (Y_1, ..., Y_n)$ be a random sample, and let $f(Y|\theta)$ denote the probability density function for some model of the data, which has parameter vector $\theta = (\theta_1, ..., \theta_r)'$. Then the FIM $I_n(\theta)$ of sample size n is given by the $r \times r$ symmetric matrix whose ij-th element is

$$I_n(\theta)_{i,j} = -\mathbb{E}\left[\frac{\partial^2 \ln f(Y|\theta)}{\partial\theta_i\,\partial\theta_j}\right]. \tag{2.1}$$

This definition strictly corresponds to the expected FIM. If no expectation is taken we obtain a data-dependent quantity that is called the observed FIM. We are interested in the derivation and computation of the expected FIM for the two-way random effects panel data model in the presence of heteroscedasticity.

2.2. The two way error components model

We consider the following two way error components model

$$y_{it} = \alpha + X'_{it}\beta + u_{it}; \qquad i = 1, ..., N;\ \ t = 1, ..., T \tag{2.2}$$

with i denoting households, individuals, firms, countries, etc., and t denoting time. The subscript i, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. y_{it} is the dependent variable for i at time t. β is a $K \times 1$ scalar, X_{it} is it-th observation on K explanatory variables. In this paper, we deal with two-way error components disturbances

$$u_{it} = \mu_i + \lambda_t + \nu_{it}, \tag{2.3}$$

where μ_i denotes the unobservable individual effect, λ_t denotes the unobservable time effect and ν_{it} is the remainder stochastic disturbance term. μ_i, λ_t account for any individual specific effect or time-specific effect not included in the regression. In vector form, (2.2) can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu, \tag{2.4}$$

where $Z_\mu = I_N \otimes i_T$, I_N is an identity matrix of dimension N, i_T is a vector of ones of dimension T and $\otimes$ denotes Kronecker product. Z_μ is a selector matrix of ones and zeros,

or simply the matrix of individual dummies included in the regression to estimate the μ_i (assuming they are fixed parameters to be estimated). Likewise, $Z_\lambda = i_N \otimes I_T$ is the matrix of time dummies of ones which may be included in the regression to estimate the λ_t (assuming they are fixed parameters to be estimated). μ, λ and ν are defined as in Baltagi (2021).

In vector form (2.2) can be written as

$$y = \alpha i_{NT} + X\beta + u = Z\gamma + u, \tag{2.5}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [i_{NT}, X]$, $\gamma' = (\alpha', \beta')$ and i_{NT} is a vector of ones.

2.3. Variance-covariance matrix of u

To obtain the variance-covariance matrix of the overall error term u, we assume the following.

Assumption $\mathcal{A}_1$ (General case)**.** The vectors λ, ν and μ are pairwise independent. Each of them is identically and independently normally distributed with mean 0 and variances $\sigma_\lambda^2 I_T$, $\sigma_\nu^2 \operatorname{diag}(h_\nu(w_i'\theta_\nu))$ and $\sigma_\mu^2 \operatorname{diag}(h_\mu(z_i'\theta_\mu))$. h_ν and h_μ are differentiable functions from $\mathbb{R}$ to $\mathbb{R}^+$, $w_i = (w_{1i}, ..., w_{pi})' \in \mathbb{R}^p$ and $z_i = (z_{1i}, ..., z_{qi})' \in \mathbb{R}^q$ are defined as in Baltagi *et al.* (2006).

In the following, we set $D_\nu = \operatorname{diag}(h_\nu(w_i'\theta_\nu))$ and $D_\mu = \operatorname{diag}(h_\mu(z_i'\theta_\mu))$. Based on the general assumption $\mathcal{A}_1$, the variance-covariance matrix of the composite disturbance u is defined by

$$\Omega = \mathbb{E}(uu') = \sigma_\mu^2 Z_\mu D_\mu Z_\mu' + \sigma_\lambda^2 Z_\lambda Z_\lambda' + \sigma_\nu^2 D_\nu \otimes I_T.$$

Therefore,

$$\Omega = \sigma_\mu^2 D_\mu \otimes i_T i_T' + \sigma_\lambda^2 i_N i_N' \otimes I_T + \sigma_\nu^2 D_\nu \otimes I_T$$

which can be simplified to

$$\Omega = \sigma_\nu^2 (D_\nu \otimes I_T) + \sigma_\mu^2 (D_\mu \otimes J_T) + \sigma_\lambda^2 (J_N \otimes I_T) \tag{2.6}$$

with $J_T = i_T i_T'$ and $J_N = i_N i_N'$.

2.4. Inverse of error variance-covariance matrix: Ω^{-1}

In order to get Ω^{-1}, we use the spectral decomposition in Wansbeek and Kapteyn (1982). After replacing J_N by $N\overline{J}_N$, I_N by $E_N + \overline{J}_N$, J_T by $T\overline{J}_T$ and I_T by $E_T + \overline{J}_T$ and collecting terms with the same matrices, we obtain

$$\Omega = \left[\sigma_\nu^2 D_\nu + \sigma_\lambda^2 J_N\right] \otimes E_T + \left[\sigma_\nu^2 D_\nu + T\sigma_\mu^2 D_\mu + \sigma_\lambda^2 J_N\right] \otimes \overline{J}_T. \tag{2.7}$$

The spectral decomposition allows us to write

$$\Omega^{-1} = C_1 \otimes E_T + C_2 \otimes \overline{J}_T \tag{2.8}$$

with $C_1 = \zeta(\sigma_\nu^2 D_\nu, 0)$ and $C_2 = \zeta(\sigma_\nu^2 D_\nu, \sigma_\mu^2 D_\mu)$, where

$$\begin{aligned}\zeta(X_1, X_2) &= \left[X_1 + TX_2 + \sigma_\lambda^2 (i_N i_N')\right]^{-1} \\ &= (X_1 + TX_2)^{-1} - \frac{\sigma_\lambda^2 (X_1 + TX_2)^{-1} J_N (X_1 + TX_2)^{-1}}{\left(1 + \sigma_\lambda^2 i_N' (X_1 + TX_2)^{-1} i_N\right)}.\end{aligned}$$

The formula used to obtain the inverse of $X_1 + TX_2 + \sigma_\lambda^2 (i_N i_N')$ is provided by Bartlett (1951).

3. SOME PRELIMINARY RESULTS

This section deals with the derivation of $\mathbb{E}\left[-d^2 \ell(\theta|u)\right]$, where $\ell(\theta|y)$ is the log-likelihood of observations. The relationship between this expectation and the FIM is given by

$$\begin{aligned}\mathbb{E}\left[-d^2 \ell(\theta|y)\right] &= \sum_{i=1}^{r}\sum_{j=1}^{r} \mathbb{E}\left[-\frac{\partial^2 \ell(\theta|y)}{\partial\theta_i\, \partial\theta_j}\right] d\theta_i\, d\theta_j \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} I_n(\theta)_{i,j}\, d\theta_i\, d\theta_j \\ &= (d\theta)' I_n(\theta)\, (d\theta), \qquad (3.1)\end{aligned}$$

where $d\theta = (d\theta_1, ..., d\theta_r)'$.

3.1. First order derivatives of the log-likelihood function

If μ_i, λ_t and ν_{it} are independent and identically normally distributed (from assumption $\mathcal{A}_1$), the joint distribution of $y = (y_{11}, ..., y_{1T}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT})'$ is the NT-multivariate normal distribution and the likelihood of the observations is

$$L(\theta|y) = \frac{1}{(2\pi)^{\frac{NT}{2}} |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y - X\beta)' \Omega^{-1} (y - X\beta)\right).$$

Since $u = y - X\beta$,

$$L(\theta|u) = \frac{1}{(2\pi)^{\frac{NT}{2}} |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} u' \Omega^{-1} u\right).$$

By taking the logarithm of the likelihood of observations

$$\ell(\theta|u) = \ln L(\theta|u) = C - \frac{1}{2} \ln|\Omega| - \frac{1}{2} u' \Omega^{-1} u, \qquad (3.2)$$

where $\theta = (\sigma_\nu^2, \sigma_\mu^2, \sigma_\lambda^2, \theta_\nu', \theta_\mu', \beta')' \in \mathbb{R}^r$ (with $r = p + q + k + 3$) is the vector of parameters and $C = -\frac{NT}{2} \ln(2\pi)$ is a constant.

We observe that the log-likelihood is continuous and at least twice differentiable with respect to each parameter. The first and second order differentials are given by the following.

Lemma 3.1. *The first order differential of L is*

$$d\ell(\theta|u) = -\frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d\Omega\big) - u'\Omega^{-1}du + \frac{1}{2}u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u. \tag{3.3}$$

The differential of u is $\mathrm{d}u = -Xd\beta$. The differential of Ω is

$$\begin{aligned} d\Omega \;=\; & (D_\nu\otimes I_T)d\sigma_\nu^2 + (D_\mu\otimes J_T)d\sigma_\mu^2 + (J_N\otimes I_T)d\sigma_\lambda^2 \\ & + \sigma_\nu^2\sum_{j=1}^{p}(D_{\nu,j}^*\otimes I_T)d\theta_{\nu,j} + \sigma_\mu^2\sum_{j'=1}^{q}(D_{\mu,j'}^*\otimes J_T)d\theta_{\mu,j'}, \end{aligned} \tag{3.4}$$

where $D_{\nu,j}^* = \dfrac{\partial D_\nu}{\partial\theta_{\nu,j}}$ *for* $j = 1,...,p$ *and* $D_{\mu,j'}^* = \dfrac{\partial D_\mu}{\partial\theta_{\mu,j'}}$ *for* $j' = 1,...,q$.

Proof of Lemma 3.1: We have

$$\begin{aligned} d\ell(\theta|u) \;&=\; 0 - \frac{1}{2}d\ln|\Omega| - \frac{1}{2}d\big(u'\Omega^{-1}u\big) \\ &=\; -\frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d\Omega\big) - \frac{1}{2}d\big(u'\Omega^{-1}u\big), \end{aligned}$$

where we used the formula $d\ln|\Omega| = \mathrm{tr}\big(\Omega^{-1}d\Omega\big)$. We also have

$$\begin{aligned} d\big(u'\Omega^{-1}u\big) \;&=\; d\big(u'\Omega^{-1}u\big)_u + u'd\big(\Omega^{-1}\big)u \\ &=\; \frac{\partial}{\partial u}\big(u'\Omega^{-1}u\big)du + u'd\Omega^{-1}u \\ &=\; 2u'\Omega^{-1}du - u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u, \end{aligned}$$

where we used the fact that $dX^{-1} = -X^{-1}dX\cdot X^{-1}$. □

3.2. Second order derivatives of the log likelihood function

Lemma 3.2. *The second order differential of L is*

$$\begin{aligned} d^2\ell(\theta|u) \;=\; & \frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\big) - \frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d^2\Omega\big) \\ & - u'\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}u + \frac{1}{2}u'\Omega^{-1}d^2\Omega\cdot\Omega^{-1}u \\ & + u'\Omega^{-1}d\Omega\cdot\Omega^{-1}du - du'\Omega^{-1}du. \end{aligned} \tag{3.5}$$

Proof of Lemma 3.2: We have

$$\begin{aligned} d^2\ell(\theta|u) \;&=\; -\frac{1}{2}\mathrm{tr}\big(d\big[\Omega^{-1}d\Omega\big]\big) - d\big(u'\Omega^{-1}du\big) + \frac{1}{2}d\big(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big) \\ &=\; -\frac{1}{2}\gamma_1 - \gamma_2 + \frac{1}{2}\gamma_3, \end{aligned}$$

where

$$\begin{aligned} \gamma_1 \;&=\; \mathrm{tr}\big(d\big(\Omega^{-1}\big)d\Omega + \Omega^{-1}d^2\Omega\big) = \mathrm{tr}\big(-\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega + \Omega^{-1}d^2\Omega\big) \\ &=\; -\mathrm{tr}\big(\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\big) + \mathrm{tr}\big(\Omega^{-1}d^2\Omega\big), \end{aligned}$$

$$\begin{aligned} \gamma_2 \;&=\; d\big(u'\Omega^{-1}du\big) = d\big(u'\Omega^{-1}du\big)_u + d\big(u'\Omega^{-1}du\big)_\Omega \\ &=\; d\big(u'\Omega^{-1}\big)_u du + u'\Omega^{-1}d(du)_u + u'd\big(\Omega^{-1}\big)_\Omega du \\ &=\; du'\Omega^{-1}du + 0 - u'\Omega^{-1}d\Omega\cdot d\Omega^{-1}du \end{aligned}$$

and

$$\begin{aligned}\gamma_3 &= d\big(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big) = d\big(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big)_u + d\big(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big)_\Omega\\ &= 2u'\Omega^{-1}d\Omega\cdot\Omega^{-1}du + \gamma_3^a,\end{aligned}$$

where

$$\begin{aligned}\gamma_3^a &= d\big(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big)_\Omega = d\big(u'\Omega^{-1}\big)_\Omega d\Omega\cdot\Omega^{-1}u + u'\Omega^{-1}d\big(d\Omega\cdot\Omega^{-1}u\big)_\Omega\\ &= d\big(u'\Omega^{-1}\big)_\Omega d\Omega\cdot\Omega^{-1}u + \gamma_3^b = -u'\Omega^{-1}d\Omega.\Omega^{-1}d\Omega.\Omega^{-1}u + \gamma_3^b\end{aligned}$$

with

$$\begin{aligned}\gamma_3^b &= u'\Omega^{-1}d\big(d\Omega\cdot\Omega^{-1}u\big)_\Omega = u'\Omega^{-1}\big[d^2\Omega\cdot\Omega^{-1}u + d\Omega\cdot d\big(\Omega^{-1}\big)u\big]\\ &= u'\Omega^{-1}d^2\Omega\cdot\Omega^{-1}u - u'\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}u.\end{aligned}$$

We deduce the final result (3.5). □

3.3. Expectation of $d^2\ell$

By taking the expectation of $-d^2\ell(\theta|u)$, we obtain after some algebra the following Lemma.

Lemma 3.3. *Assuming that $|d^2\ell(\theta|u)|$ is integrable,*

$$\mathbb{E}\big(-d^2\ell(\theta|u)\big) = \frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\big) + d\beta' X'\Omega^{-1}Xd\beta. \tag{3.6}$$

Proof of Lemma 3.3: The proof uses the equality

$$\mathbb{E}\big(u'Au\big) = \mathbb{E}\big(\mathrm{tr}\big(u'Au\big)\big) = \mathbb{E}\big[\mathrm{tr}(Auu')\big] = \mathrm{tr}\big(\mathbb{E}\big[(Auu')\big]\big) = A\,\mathrm{tr}\big(\Omega\big),$$

where u is a random vector and A is a matrix of constant terms. The expectation of $d^2\ell$ is then given by

$$\mathbb{E}\big(d^2\ell(\theta|u)\big) = \frac{1}{2}\kappa_0 - \frac{1}{2}\kappa_1 - \kappa_2 + \frac{1}{2}\kappa_3 + \kappa_4 - \kappa_5,$$

where

$$\begin{aligned}\kappa_0 &= \mathrm{tr}\big(\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\big),\\ \kappa_1 &= \mathrm{tr}\big(\Omega^{-1}d^2\Omega\big),\\ \kappa_2 &= \mathbb{E}\big[u'\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}u\big] = \mathrm{tr}\big[\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}\mathbb{E}(uu')\big]\\ &= \mathrm{tr}\big[\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}\Omega\big] = \mathrm{tr}\big[\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\big],\\ \kappa_3 &= \mathbb{E}\big[u'\Omega^{-1}d^2\Omega\cdot\Omega^{-1}u\big] = \mathrm{tr}\big[\Omega^{-1}d^2\Omega\big],\\ \kappa_4 &= \mathbb{E}\big[u'\Omega^{-1}d\Omega\cdot\Omega^{-1}du\big] = -\mathbb{E}\big[u'\Omega^{-1}d\Omega\cdot\Omega^{-1}X\mathrm{d}\beta\big]\\ &= -\mathbb{E}\big[u\big]'\Omega^{-1}d\Omega\cdot\Omega^{-1}X\mathrm{d}\beta = 0\end{aligned}$$

and

$$\kappa_5 = \mathbb{E}\big[du'\Omega^{-1}du\big] = \mathbb{E}\big[d\beta' X'\Omega^{-1}Xd\beta\big].$$

We obtain that $\kappa_1 = \kappa_3$, $\kappa_4 = 0$ and $\kappa_0 = \kappa_2$. We deduce that $\mathbb{E}\big(d^2\ell(\theta|u)\big) = -\frac{1}{2}\kappa_0 - \kappa_5$ and $\mathbb{E}\big(-d^2\ell(\theta|u)\big) = \frac{1}{2}\kappa_0 + \kappa_5$ which is the desired result. □

The following is an important Lemma based on symmetric matrices.

Lemma 3.4. *If A is a symmetric and square matrix of order $p+q$, for a given vector $z = (x, y) \in \mathbb{R}^{p+q}$, where $x = (x_1, ..., x_p)' \in \mathbb{R}^p$, $y = (y_1, ..., y_q)' \in \mathbb{R}^q$, we have the following equality*

$$z'Az = x'A_p x + 2x'A_{p,q}y + y'A_q y,$$

where

$$A = \begin{bmatrix} A_p & A_{p,q} \\ A'_{p,q} & A_q \end{bmatrix},$$

where A_p, A_q and $A_{p,q}$ are matrices of dimensions of (p,p), (q,q) and (p,q) respectively.

Proof of Lemma 3.4: The proof of this Lemma is straightforward. □

4. MAIN RESULTS

We now turn to the main results obtained in this paper. The vector of parameters is denoted by $\theta = (\theta_1, ..., \theta_r)'$, where $\theta_1 = \sigma_\nu^2$; $\theta_2 = \sigma_\mu^2$; $\theta_3 = \sigma_\lambda^2$; $\big(\theta_{j+3} = \theta_{\nu,j}\big)_{1\le j\le p}$; $\big(\theta_{j+p+3} = \theta_{\mu,j}\big)_{1\le j\le q}$ and $\big(\theta_{q+p+3+j} = \beta_j\big)_{1\le j\le k}$. We denote by $\overline{\theta}$ the vector of dimension $(p+q+3)$ defined by $\overline{\theta}_j = \theta_j$ for $j = 1, ..., p+q+3$. The following proposition gives the relationship between the FIM at θ, the FIM at $\overline{\theta}$ and the FIM at β.

Proposition 4.1. *If $d\theta = (d\overline{\theta}', d\beta')'$ with $d\beta = (d\beta_1, ..., d\beta_k)' \in \mathbb{R}^k$ and $d\overline{\theta} = (d\sigma_\nu^2, d\sigma_\mu^2, d\sigma_\lambda^2, \{d\theta_{\nu,j}\}_{j=1,...,p}, \{d\theta_{\mu,j}\}_{j=1,...,q})' \in \mathbb{R}^{p+q+3}$, we have*

$$(d\theta)'I_n(\theta)(d\theta) = (d\overline{\theta})'I_n(\overline{\theta})(d\overline{\theta}) + 2(d\overline{\theta})'I_n(\overline{\theta}, \beta)(d\beta) + (d\beta)'I_n(\beta)(d\beta), \tag{4.1}$$

where

$$I_n(\theta) = \begin{bmatrix} I_n(\overline{\theta}) & I_n(\overline{\theta}, \beta) \\ \\ I_n(\overline{\theta}, \beta) & I_n(\beta) \end{bmatrix}$$

and

$$I_n(\overline{\theta}) = \mathbb{E}\left[-\frac{\partial^2\ell(u|\theta)}{\partial\overline{\theta}_i\,\partial\overline{\theta}_j}\right]_{1\le i,j\le p+q+3}, \quad I_n(\beta) = \mathbb{E}\left[-\frac{\partial^2\ell(u|\theta)}{\partial\beta_i\,\partial\beta_j}\right]_{1\le i,j\le k},$$

$$I_n(\overline{\theta}, \beta) = \mathbb{E}\left[-\frac{\partial^2\ell(u|\theta)}{\partial\overline{\theta}_i\,\partial\beta_j}\right]_{\substack{1\le j\le k \\ 1\le i\le p+q+3}}.$$

Proof of Proposition 4.1: Using Lemma 3.4, the proof is straightforward. □

At this stage, computing the FIM requires the derivation of $\mathrm{tr}\big(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\big)$. By multiplying $d\Omega$ (given in (3.4)) with the expression of Ω^{-1} from equation (2.8), we obtain

$$\begin{aligned}\Omega^{-1}d\Omega \;=\; & \underbrace{[C_1D_\nu\otimes E_T + C_2D_\nu\otimes \overline{J}_T]}_{\Omega_1}d\sigma_\nu^2 + \underbrace{\big(C_2D_\mu\otimes T\overline{J}_T\big)}_{\Omega_2}d\sigma_\mu^2\\ & + \underbrace{[C_1J_N\otimes E_T + C_2J_N\otimes \overline{J}_T]}_{\Omega_3}d\sigma_\lambda^2 + \sum_{j_1=1}^{p}\underbrace{\big(\sigma_\nu^2(C_1D^*_{\nu,j_1}\otimes E_T + C_2D^*_{\nu,j_1}\otimes \overline{J}_T)\big)}_{\Omega_{3+j_1}}d\theta_{\nu,j_1}\\ & + \sum_{j_2=1}^{q}\underbrace{\big(\sigma_\mu^2(C_2D^*_{\mu,j_2}\otimes T\overline{J}_T)\big)}_{\Omega_{p+3+j_2}}d\theta_{\mu,j_2} = \sum_{j=1}^{p+q+3}\Omega_j\,d\overline{\theta}_j = \sum_{j=1}^{r-k}\Omega_j\,d\overline{\theta}_j.\end{aligned}$$

We deduce that the trace of $\Omega^{-1}d\Omega.\Omega^{-1}d\Omega$ is

$$\mathrm{tr}\big(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\big) = \sum_{i=1}^{r-k}\sum_{j=1}^{r-k}\mathrm{tr}(\Omega_i\Omega_j)\,d\overline{\theta}_i\,d\overline{\theta}_j.$$

From Lemma 3.3, the expectation of $-d^2\ell(u|\theta)$ can be written as

$$\mathbb{E}_\theta\big[-d^2\ell(u|\theta)\big] = \sum_{i=1}^{r-k}\sum_{j=1}^{r-k}\frac{1}{2}\mathrm{tr}(\Omega_i\Omega_j)\,d\overline{\theta}_i\,d\overline{\theta}_j + (d\beta)'(X'\Omega^{-1}X)\,(d\beta). \tag{4.2}$$

By comparing equations (3.1), (4.1) and (4.2), we deduce that

$$I_n(\overline{\theta},\beta) = \mathrm{O} \qquad \text{with} \quad I_n(\beta) = X'\Omega^{-1}X.$$

The derivation of the FIM $I_n(\theta)$ is then based on the derivation of $I_n(\overline{\theta})$. From equation (4.2), the terms of $I_n(\overline{\theta})$ are given by $I_n(\overline{\theta})_{i,j} = a_{i,j} = \frac{1}{2}\mathrm{tr}(\Omega_i\Omega_j)$ or in matrix form as

$$I_n(\overline{\theta}) = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \boxed{\mathbf{a}_{1,4}}^{1\times p} & \boxed{\mathbf{a}_{1,5}}^{1\times q}\\ a_{2,1} & a_{2,2} & a_{2,3} & \boxed{\mathbf{a}_{2,4}}^{1\times p} & \boxed{\mathbf{a}_{2,5}}^{1\times q}\\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{\mathbf{a}_{3,4}}^{1\times p} & \boxed{\mathbf{a}_{3,5}}^{1\times q}\\ \boxed{\mathbf{a}_{4,1}}^{p\times 1} & \boxed{\mathbf{a}_{4,2}}^{p\times 1} & \boxed{\mathbf{a}_{4,3}}^{p\times 1} & \boxed{\mathbf{a}_{4,4}}^{p\times p} & \boxed{\mathbf{a}_{4,5}}^{p\times q}\\ \boxed{\mathbf{a}_{5,1}}^{q\times 1} & \boxed{\mathbf{a}_{5,2}}^{q\times 1} & \boxed{\mathbf{a}_{5,3}}^{q\times 1} & \boxed{\mathbf{a}_{5,4}}^{q\times p} & \boxed{\mathbf{a}_{5,5}}^{q\times q}\end{pmatrix}.$$

$I_n(\overline{\theta})$ is a symmetric block matrix where the terms in boxes are also matrices with the dimensions indicated at the top of the boxes. We need to evaluate the components of $\overline{\theta}$:

$$\begin{aligned}& a_{i,j} = a_{j,i} = \frac{1}{2}\mathrm{tr}(\Omega_i\Omega_j) \text{ for } i,j\in\{1,2,3\};\\ & \mathbf{a}_{i,4}(j_1) = \mathbf{a}'_{4,i}(j_1) = \frac{1}{2}\mathrm{tr}(\Omega_i\Omega_{j_1+3}) \text{ for } i = 1:3 \text{ and for } j_1 = 1:p;\\ & \mathbf{a}_{i,5}(j_2) = \mathbf{a}'_{5,i}(j_2) = \frac{1}{2}\mathrm{tr}(\Omega_i\Omega_{p+3+j_2}) \text{ for } i = 1:3 \text{ and for } j_2 = 1:q;\\ & \mathbf{a}_{4,4}(j_1,j_2) = \frac{1}{2}\mathrm{tr}(\Omega_{j_1+3}\Omega_{j_2+3}) \text{ for } 1\le j_1,j_2\le p;\\ & \mathbf{a}_{4,5}(j_1,j_2) = \mathbf{a}'_{5,4}(j_1,j_2) = \frac{1}{2}\mathrm{tr}(\Omega_{j_1+3}\Omega_{j_2+p+3}) \text{ for } 1\le j_1\le p, 1\le j_2\le q;\\ & \mathbf{a}_{5,5}(j_1,j_2) = \frac{1}{2}\mathrm{tr}(\Omega_{j_1+p+3}\Omega_{j_2+p+3}) \text{ for } 1\le j_1,j_2\le q.\end{aligned}$$

Remark 4.1. In order to evaluate the components of $\overline{\theta}$, we will use the fact that $E_T\overline{J}_T = \mathrm{O}$, $E_T^2 = E_T$, $\mathrm{tr}\big(\overline{J}_T\big) = 1$ and $\mathrm{tr}\big(E_T\big) = T-1$.

From the Remark 4.1 we will show how the components of $\overline{\theta}$ can be written in terms of C_1, C_2, J_N, D_ν and D_μ.

To illustrate the computation of $a_{i,j}$, we give, for example, the details of calculating $a_{1,1}$. By the definition, $a_{1,1} = \frac{1}{2}\mathrm{tr}(\Omega_1^2)$ with $\Omega_1^2 = C_1D_\nu C_1D_\nu \otimes E_T^2 + C_1D_\nu C_2D_\nu \otimes E_T\overline{J}_T + C_2D_\nu C_1D_\nu \otimes \overline{J}_T E_T + C_2D_\nu C_2D_\nu \otimes \overline{J}_T^2$. Now using the fact that $E_T\overline{J}_T = \mathrm{O}$, $E_T^2 = E_T$, $\overline{J}_T^2 = \overline{J}_T$, $\mathrm{tr}\big(\overline{J}_T\big) = 1$, $\mathrm{tr}\big(E_T\big) = T-1$ and taking the trace of Ω_1^2, we obtain $a_{1,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1D_\nu\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2D_\nu\big]$. The other coefficients are obtained in the same way. The coefficients are given in Appendix A.

Remark 4.2. A quick inspection of the elements of $I_n(\overline{\theta})$ in Appendix B reveals that they are all written as linear combinations of $\mathrm{tr}\big(C_1M_1C_2M_2\big)$, where $M_1 \in \{\mathrm{diag}(\eta_i), J_N\}$ and $M_2 \in \{\mathrm{diag}(\psi_i), J_N\}$ [i.e. either diagonal matrices or square matrices of ones].

To have the final expressions of $a_{i,j}$, it suffices to evaluate the quantity $\mathrm{tr}\big(C_1M_1C_2M_2\big)$ in the following cases:

Case 1: $\mathrm{tr}\big(C_1M_1C_2M_2\big)$, where $M_1 = \mathrm{diag}(\eta_i)$ and $M_2 = \mathrm{diag}(\psi_i)$ are diagonal matrices.

Case 2: $\mathrm{tr}\big(C_1M_1C_2M_2\big)$, where $M_1 = J_N$ and $M_2 = \mathrm{diag}(\psi_i)$ or $M_1 = \mathrm{diag}(\eta_i)$ and $M_2 = J_N$, that is one is a diagonal matrix while the other is a square matrix of ones.

Case 3: $\mathrm{tr}\big(C_1M_1C_2M_2\big)$, where $M_1 = J_N$ and $M_2 = J_N$ are all square matrices of ones.

Before evaluating the three previous quantities above, we make the following important remark to ease the calculation of traces.

Remark 4.3. Using the expressions of C_1 and C_2 given by equation (2.9), we can prove that

$$C_l \;=\; D_l^{-1} - \frac{\sigma_\lambda^2 D_l^{-1}J_N D_l^{-1}}{1+\sigma_\lambda^2 i_N' D_l^{-1} i_N} \;=\; \varsigma(D_l, 0), \qquad l = 1, 2,$$

with $D_1 = \mathrm{diag}(\vartheta_i^{-1})$ and $D_2 = \mathrm{diag}(\phi_i^{-1})$.

In fact, by definition of C_1 and C_2, we have $C_1 = \varsigma(\sigma_\nu^2 D_\nu, 0)$ and $C_2 = \varsigma(\sigma_\nu^2 D_\nu, \sigma_\mu^2 D_\mu)$ which can also be written as $C_2 = \varsigma(\sigma_\nu^2 D_\nu + T\sigma_\mu^2 D_\mu, 0)$. We then deduce that

$$\vartheta_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) \quad \text{and} \quad \phi_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) + T\sigma_\mu^2 h_\mu(z_i'\theta_\mu). \tag{4.3}$$

We can also observe that in the case of homokedasticity, $\vartheta_i = 1/\sigma_\nu^2$ and $\phi_i = 1/(\sigma_\nu^2 + T\sigma_\mu^2)$.

5. SOME COMPUTATIONAL ISSUES

If we derive a specific formula for the calculation of $\operatorname{tr}\big(C_1M_1C_2M_2\big)$, it will definitively ease the computation of the FIM $I_n(\bar{\theta})$ and thereby that of the FIM $I_n(\theta)$. We consider matrices D_1, D_2, C_1, C_2, M_1 and M_2 as defined earlier. According to the definition of M_1 and M_2, $\operatorname{tr}\big(C_1M_1C_2M_2\big)$ is obtained through the real valued functions $\Psi_i\,(i=1,2,3)$ given by the following propositions.

Proposition 5.1. *If $M_1=\operatorname{diag}(\eta_i)\neq i_Ni_N'$ and $M_2=\operatorname{diag}(\psi_i)\neq i_Ni_N'$ then*

$$\begin{aligned}\operatorname{tr}\big(C_1M_1C_2M_2\big) &= \Psi_1(\vartheta,\eta,\phi,\psi)\\ &= \langle\eta\odot\vartheta,\psi\odot\phi\rangle-\frac{\sigma_\lambda^2\langle\eta\odot\vartheta,\psi\odot\phi\odot\phi\rangle}{1+\sigma_\lambda^2i_N'\phi}\\ &\quad-\frac{\sigma_\lambda^2\langle\eta\odot\vartheta\odot\vartheta,\psi\odot\phi\rangle}{1+\sigma_\lambda^2i_N'\vartheta}+\frac{\sigma_\lambda^4\langle\eta,\vartheta\odot\phi\rangle\langle\psi,\vartheta\odot\phi\rangle}{(1+\sigma_\lambda^2i_N'\vartheta)(1+\sigma_\lambda^2i_N'\phi)}\end{aligned}\tag{5.1}$$

with $\langle\cdot,\cdot\rangle$ and $\odot$ denoting the inner and Hardamar products, respectively.

Proof of Proposition 5.1: The proof follows since

$$\begin{aligned}\operatorname{tr}\big(C_1M_1C_2M_2\big) &= \operatorname{tr}\left[\left(D_1^{-1}M_1-\frac{\sigma_\lambda^2D_1^{-1}i_N'i_ND_1^{-1}M_1}{1+\sigma_\lambda^2i_N'D_1^{-1}i_N}\right)\left(D_2^{-1}M_2-\frac{\sigma_\lambda^2D_2^{-1}i_N'i_ND_2^{-1}M_2}{1+\sigma_\lambda^2i_N'D_2^{-1}i_N}\right)\right]\\ &= \operatorname{tr}\big(D_1^{-1}M_1D_2^{-1}M_2\big)-\frac{\sigma_\lambda^2i_N'D_2^{-1}M_2D_1^{-1}M_1D_2^{-1}i_N}{1+\sigma_\lambda^2i_N'D_2^{-1}i_N}\\ &\quad-\frac{\sigma_\lambda^2i_N'D_1^{-1}M_1D_2^{-1}M_2D_1^{-1}i_N}{1+\sigma_\lambda^2i_N'D_1^{-1}i_N}+\frac{\sigma_\lambda^4\big(i_N'D_1^{-1}M_1D_2^{-1}i_N\big)\big(i_N'D_2^{-1}M_2D_1^{-1}i_N\big)}{\big(1+\sigma_\lambda^2i_N'D_1^{-1}i_N\big)\big(1+\sigma_\lambda^2i_N'D_2^{-1}i_N\big)}\\ &= \sum_{i=1}^N\vartheta_i\eta_i\psi_i\phi_i-\frac{\sigma_\lambda^2\sum_{i=1}^N\eta_i\vartheta_i\psi_i\phi_i^2}{1+\sigma_\lambda^2\sum_{i=1}^N\phi_i}-\frac{\sigma_\lambda^2\sum_{i=1}^N\eta_i\vartheta_i^2\psi_i\phi_i}{1+\sigma_\lambda^2\sum_{i=1}^N\vartheta_i}\\ &\quad+\frac{\sigma_\lambda^4\Big(\sum_{i=1}^N\vartheta_i\eta_i\phi_i\Big)\Big(\sum_{i=1}^N\vartheta_i\psi_i\phi_i\Big)}{\Big(1+\sigma_\lambda^2\sum_{i=1}^N\vartheta_i\Big)\Big(1+\sigma_\lambda^2\sum_{i=1}^N\phi_i\Big)}\\ &= \langle\eta\odot\vartheta,\psi\odot\phi\rangle-\frac{\sigma_\lambda^2\langle\eta\odot\vartheta,\psi\odot\phi\odot\phi\rangle}{1+\sigma_\lambda^2i_N'\phi}\\ &\quad-\frac{\sigma_\lambda^2\langle\eta\odot\vartheta\odot\vartheta,\psi\odot\phi\rangle}{1+\sigma_\lambda^2i_N'\vartheta}+\frac{\sigma_\lambda^4\langle\eta,\vartheta\odot\phi\rangle\langle\psi,\vartheta\odot\phi\rangle}{(1+\sigma_\lambda^2i_N'\vartheta)(1+\sigma_\lambda^2i_N'\phi)}.\end{aligned}$$

□

Proposition 5.2. *If $M_1=\operatorname{diag}(\eta_i)$ and $M_2=i_Ni_N'$, then*

$$\operatorname{tr}(C_1M_1C_2M_2)=\Psi_2(\vartheta,\eta,\phi)=\frac{\langle\eta,\vartheta\odot\phi\rangle}{\big(1+\sigma_\lambda^2i_N'\vartheta\big)\big(1+\sigma_\lambda^2i_N'\phi\big)}.\tag{5.2}$$

For $M_1=i_Ni_N'$ and $M_2=\operatorname{diag}(\psi_i)$, we just need to interchange the subscripts 1 and 2 leading to $\Psi_2(\eta,\vartheta,\phi)$.

Proof of Proposition 5.2: We have

$$\begin{aligned}\operatorname{tr}\big(C_1M_1C_2M_2\big) &= \operatorname{tr}\big(i_N'D_1^{-1}M_1D_2^{-1}i_N\big) - \frac{\sigma_\lambda^2(i_N'D_2^{-1}i_N)(i_N'D_1^{-1}M_1D_2^{-1}i_N)}{1+\sigma_\lambda^2 i_N'D_2^{-1}i_N}\\ &\quad - \frac{\sigma_\lambda^2(i_N'D_1^{-1}M_1D_2^{-1}i_N)(i_N'D_1^{-1}i_N)}{1+\sigma_\lambda^2 i_N'D_1^{-1}i_N}\\ &\quad + \frac{\sigma_\lambda^4\big((i_N'D_1^{-1}M_1D_2^{-1}i_N)\big)\big(i_N'D_2^{-1}i_N\big)(i_N'D_1^{-1}i_N)}{\big(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N\big)\big(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N\big)}.\end{aligned}$$

Now if $B = i_N'D_1^{-1}M_1D_2^{-1}i_N$ (which is a scalar), observing that $i_N'D_1^{-1}i_N = i_N'\vartheta$ and $i_N'D_2^{-1}i_N = i_N'\phi$, we have

$$\begin{aligned}\operatorname{tr}\big(C_1M_1C_2M_2\big) &= B - \frac{\sigma_\lambda^2(i_N'\phi)B}{1+\sigma_\lambda^2 i_N'\phi} - \frac{\sigma_\lambda^2 B(i_N'\vartheta)}{1+\sigma_\lambda^2 i_N'\vartheta} + \frac{\sigma_\lambda^4 B(i_N'\phi)(i_N'\vartheta)}{\big(1+\sigma_\lambda^2 i_N'\vartheta\big)\big(1+\sigma_\lambda^2 i_N'\phi\big)}\\ &= \frac{B}{\big(1+\sigma_\lambda^2 i_N'\vartheta\big)\big(1+\sigma_\lambda^2 i_N'\phi\big)} = \frac{\langle\eta,\vartheta\odot\phi\rangle}{\big(1+\sigma_\lambda^2 i_N'\vartheta\big)\big(1+\sigma_\lambda^2 i_N'\phi\big)}.\end{aligned}$$

We deduce that for $M_1 = i_Ni_N'$ and $M_2 = \operatorname{diag}(\psi_i)$,

$$\begin{aligned}\operatorname{tr}(C_1M_1C_2M_2) &= \operatorname{tr}(C_1M_2C_2M_1)\\ &= \frac{\langle\psi,\vartheta\odot\phi\rangle}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}\end{aligned}$$

since $\operatorname{tr}(IJ) = \operatorname{tr}(JI)$ for any matrices I and J where the matrix product holds. □

Proposition 5.3. *If $M_1 = i_Ni_N' = M_2$, then*

$$\operatorname{tr}(C_1M_1C_2M_2) = \Psi_3(\vartheta,\phi) = \frac{(i_N'\phi)(i_N'\vartheta)}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}. \tag{5.3}$$

Proof of Proposition 5.3: Since for the case $M_2 = i_Ni_N'$ the result depends on M_1 through B (B is the matrix used in the proof of Proposition 5.2), replacing M_1 by i_Ni_N' in the expression of B, we immediately obtain

$$\begin{aligned}\operatorname{tr}(C_1M_1C_2M_2) &= \frac{i_N'D_2^{-1}M_1D_1^{-1}i_N}{(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N)(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N)}\\ &= \frac{(i_N'D_2^{-1}i_N)(i_N'D_1^{-1}i_N)}{(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N)(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N)} = \frac{(i_N'\phi)(i_N'\vartheta)}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}.\end{aligned}$$ □

Before stating the important result that gives the final expression of the FIM, we introduce the following notations:

$$\begin{aligned}D_\nu &= \operatorname{diag}(h_\nu(w_i'\theta_\nu)) = \operatorname{diag}(\eta_i) \quad \text{with } \eta_i = h_\nu(w_i'\theta_\nu);\\ D_\mu &= \operatorname{diag}(h_\mu(z_i'\theta_\mu)) = \operatorname{diag}(\eta_i) \quad \text{with } \psi_i = h_\mu(z_i'\theta_\mu);\\ D_{\nu,j}^* &= \operatorname{diag}(h_\nu'(\omega_i'\theta_\nu)w_{i,j}) = \operatorname{diag}(\eta_i^j) \quad \text{with } \eta_i^j = h_\nu'(\omega_i'\theta_\nu)\omega_{i,j};\\ D_{\mu,j}^* &= \operatorname{diag}(h_\mu'(z_i'\theta_\mu)z_{i,j}) = \operatorname{diag}(\psi_i^j) \quad \text{with } \psi_i^j = h_\mu'(z_i'\theta_\mu)z_{i,j};\\ C_1 &= \varsigma(D_1,0) \quad \text{with } D_1 = \operatorname{diag}(\vartheta_i^{-1});\\ C_2 &= \varsigma(D_2,0) \quad \text{with } D_2 = \operatorname{diag}(\phi_i^{-1});\end{aligned}$$

where $\vartheta_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu)$ and $\phi_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) + T\sigma_\mu^2 h_\mu(z_i'\theta_\mu)$. We also set $\eta^j = (\eta_i^j)_{i=1}^N$; $\eta = (\eta_i)_{i=1}^N$; $\psi^j = (\psi_i^j)_{i=1}^N$; $\psi = (\psi_i)_{i=1}^N$; $\phi = (\phi_i)_{i=1}^N$ and $\vartheta = (\vartheta_i)_{i=1}^N$.

Proposition 5.4. *If y is a two-way error components model in the form $y = X\beta + u$, where u satisfies assumption $\mathcal{A}_1$, the FIM evaluated at θ is*

$$I_n(\theta) = \begin{bmatrix} I_n(\overline{\theta}) & O \\ O & X'\Omega^{-1}X \end{bmatrix},$$

where the coefficients of $I_n(\overline{\theta})$ described in Appendix B are now written in terms of $\Psi_1(\vartheta, \eta, \phi, \psi)$, $\Psi_2(\vartheta, \eta, \phi)$ and $\Psi_3(\vartheta, \phi)$. These expressions are given in Appendix C.

Proof of Proposition 5.4: The proof is straightforward, see Appendix B. □

We can summarize the use of our method in practice. Given a two way error components model, the computation of the FIM is based on the following steps:

Step 1: Identify $\eta = (\eta_i)$, $\psi = (\psi_i)$ which are the vectors derived from $D_\nu = \text{diag}(\eta_i)$ and $D_\mu = \text{diag}(\psi_i)$;

Step 2: Compute Ω (from equation (3.4)) and then Ω^{-1};

Step 3: Deduce values of ϑ_i and ϕ_i from equation (4.3);

Step 4: Identify $\eta^* = (\eta_i^j)$ and $\psi^* = (\psi_i^j)$ which are the matrices obtained from $D^*_{\nu,j} = \text{diag}(\eta_i^j)$ for $j = 1, ..., p$ and $D^*_{\mu,j} = \text{diag}(\psi_i^j)$ for $j = 1, ..., q$;

Step 5: Obtain values of the parameters $\theta \in \mathbb{R}^{p+q+3+k}$ (which can be estimated using some observations);

Step 6: Observations $x_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ are needed to compute $X'\Omega^{-1}X$.

The above information is input in our code named `FIM.FUN`. The code has been written in the `R` language and is available as supplementary material.

6. TWO REAL DATA EXAMPLES

We now consider two real data examples to illustrate the above analysis. For consistency and comparative purposes, asymptotic as well as exact results are obtained in the homoscedasticity case while only exact results are obtained in the heteroscedasticity case.

6.1. Example 1: Public capital productivity (Homoscedasticity case)

6.1.1. Model specification

Following Munnell (1990) and Baltagi and Pinnoi (1995), we re-consider the following Cobb–Douglas production relationship investigating the productivity of public capital in private production:

$$\ln(y_{it}) = \beta_0 + \beta_1 \ln(Pc_{it}) + \beta_2 \ln(PS_{it}) + \beta_3 \ln(L_{it}) + \beta_4 U_{\text{emp}_{it}} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{6.1}$$

where y_{it} = gross state product; Pc_{it} = public capital; PS_{it} = private capital; L_{it} = labour input as payrolls; $U_{\text{emp}_{it}}$ = unemployment rate; $\beta_0, ..., \beta_4$ = coefficients to be estimated; μ_i = the unobservable individual effect; ε_{it} = the rest of the perturbation. Data are from 48 US states (i.e. $N = 48$) observed over the period 1970 to 1986, (i.e. $T = 7$). The data are obtained from the Wiley web site at `www.wiley.com/go/baltagi3e`. Following a common unjustified practice (Baltagi, 2021; Baltagi and Pinnoi, 1995; Feng *et al.*, 2020; Munnell, 1990; Su *et al.*, 2013; and Wu and Zhu, 2011), we assume that errors are homoscedastic. We consider estimating the above model based on four estimators:

(**i**) Swamy and Arora (residuals obtained from solving a system of 3 equations);

(**ii**) Wallace and Hussain (OLS residuals);

(**iii**) Wansbeek and Kapteyn (LSDV residuals);

(**iv**) maximum likelihood (ML residuals).

Results based on the following restrictions

(**i**) $(\beta_1 = 0)$ (one restriction);

(**ii**) $(\beta_1 = 0, \beta_2 = 0)$; $(\beta_1 = 0, \beta_3 = 0)$; $(\beta_1 = 0, \beta_4 = 0)$ (two restrictions); and

(**iii**) $(\beta_1 = 0, \beta_2 = 0, \beta_3 = 0)$; $(\beta_1 = 0, \beta_2 = 0, \beta_4 = 0)$ and $(\beta_1 = 0, \beta_3 = 0, \beta_4 = 0)$ (three restrictions)

are reported below.

6.1.2. Results

In Table 1a estimations are done based on the FIM based on the observed information matrix. Rather, in Table 1b we used the FIM which relies on exact information matrix developed by the authors.

The results are based on the FIM with homoscedastic errors. In Table 1a, results clearly indicate that irrespective of the estimator used and no matter how many restrictions are used, public capital remains important and productive in private production. The results are consistent with the study by Baltagi and Pinnoi (1995) and Munnell (1990). In Table 1b the FIM is now based on exact information matrix. Comparing Table 1a and 1b, we notice that results are close. This means that asymptotic results based on the Central Limit Theorem adequately approximate our results in many cases.

6.1.3. Discussion

Some inconsistencies still exist when applying the linear restriction tests. For example, the number of times some test results are not available (i.e. NA) remain relatively high. Possible reason could be that the sample size is not very large. This could explain why the Wald (W), Likelihood ratio (LR) and Lagrange multiplier (LM) tests give negative values.

Table 1a: Restriction test results – FIM based on observed information matrix [Homoscedasticity case].
Notes: λ_{W}, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively.
NA = not available. ^{a}, ^{b} and ^{c} represent statistical significant at 1%, 5% and 10% respectively.

One restriction ($\beta_1 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_2 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_3 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_4 = 0$)		
λ_{W}	λ_{LR}	λ_{LM}	λ_{W}	λ_{LR}	λ_{LM}	λ_{W}	λ_{LR}	λ_{LM}	λ_{W}	λ_{LR}	λ_{LM}
Swamy and Arora											
26.074^{a}	25.662^{a}	25.266^{a}	1443.184^{a}	803.969^{a}	521.267^{a}	8103.542^{a}	1951.533^{a}	741.348^{a}	105.407^{a}	99.134^{a}	93.348^{a}
Wansbeek and Kapteyn											
NA	NA	NA	1189.133^{a}	733.625^{a}	483.924^{a}	10132.516^{a}	2187.806^{a}	755.183^{a}	NA	NA	NA
Wallace and Hussain											
35.625^{a}	34.866^{a}	34.134^{a}	1457.202^{a}	836.017^{a}	523.084^{a}	9583.345^{a}	2076.788^{a}	751.971^{a}	84.680^{a}	80.569^{a}	76.719^{a}
Maximum likelihood											
36.574^{a}	35.778^{a}	35.005^{a}	1505.446^{a}	853.161^{a}	529.174^{a}	10441.093^{a}	2141.460^{a}	756.850^{a}	87.307^{a}	82.945^{a}	78.868^{a}

Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_4 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_3 = 0$, $\beta_4 = 0$)		
λ_{W}	λ_{LR}	λ_{LM}	λ_{W}	λ_{LR}	λ_{LM}	λ_{W}	λ_{LR}	λ_{LM}
Swamy and Arora								
92545.656^{a}	3867.694^{a}	808.867^{a}	1395.961^{a}	813.732^{a}	514.974^{a}	9656.094^{a}	2082.476^{a}	752.416^{a}
Wansbeek and Kapteyn								
69599.274^{a}	3637.524^{a}	806.543^{a}	1000.809^{a}	653.145^{a}	449.502^{a}	8278.811^{a}	1967.413^{a}	742.787^{a}
Wallace and Hussain								
91481.904^{a}	3858.343^{a}	808.785^{a}	1411.693^{a}	819.515^{a}	517.100^{a}	9316.481^{a}	2055.575^{a}	750.285^{a}
Maximum likelihood								
89097.000^{a}	3836.985^{a}	808.594^{a}	1402.137^{a}	816.007^{a}	515.813^{a}	10012.449^{a}	2109.782^{a}	754.508^{a}

Table 1b: Restriction test results – FIM based on exact information matrix [Homoscedasticity case].
Notes: λ_{W}, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively.
NA = not available. ^{a}, ^{b} and ^{c} represent statistical significant at 1%, 5% and 10% respectively.

One restriction ($\beta_1 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_2 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_3 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_4 = 0$)		
$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$
Swamy and Arora											
692.810^{a}	NA	$1.302e+11^{a}$	1709.655^{a}	NA	232103.990^{a}	861.639^{a}	575.066^{a}	867.050^{a}	1437.169^{a}	NA	235251.998^{a}
Wansbeek and Kapteyn											
242.551^{a}	236.413^{a}	4408906.542^{a}	336.928^{a}	NA	65958346.732^{a}	353.549^{a}	4991.859	$1.7904e+09^{a}$	272.816^{a}	2276.679^{a}	1141660.516^{a}
Wallace and Hussain											
370.771^{a}	383.875^{a}	23518.283^{a}	564.807^{a}	NA	609594.625^{a}	487.433^{a}	321.686^{a}	74512.198^{a}	452.761^{a}	NA	214563.054^{a}
Maximum likelihood											
377.302^{a}	NA	103473.807^{a}	570.571^{a}	NA	6709213.013^{a}	488.120^{a}	NA	158524.010^{a}	459.057^{a}	NA	173961.580^{a}

Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_4 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_3 = 0$, $\beta_4 = 0$)		
$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$
Swamy and Arora								
2138.526^{a}	NA	609110684.047^{a}	1109.311^{a}	NA	178319.656^{a}	857.074^{a}	NA	293757.012^{a}
Wansbeek and Kapteyn								
493.148^{a}	NA	839636025.188^{a}	364.119^{a}	NA	18431907.107^{a}	358.831^{a}	1433.875^{a}	1586347.584^{a}
Wallace and Hussain								
777.764^{a}	NA	5001783.708^{a}	624.210^{a}	NA	767226.002^{a}	581.321^{a}	NA	2843988.540^{a}
Maximum likelihood								
775.828^{a}	NA	$8.416138e+08^{a}$	629.157^{a}	NA	3688699.810^{a}	775.828^{a}	NA	$8.416138e+08^{a}$

6.2. Example 2: Public capital productivity (Double Heteroscedasticity case)

6.2.1. Model specification

We still consider the model described previously, i.e.

$$\ln(y_{it}) = \beta_0 + \beta_1 \ln(Pc_{it}) + \beta_2 \ln(PS_{it}) + \beta_3 \ln(L_{it}) + \beta_4 U_{\text{emp}_{it}} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{6.2}$$

Next, we assume that the analysis is more complex and thereby proceed methodically. As will be seen later, the FIM here is based on heteroscedasticity of the two errors involved. The following steps are important to understand results reported in Table 2.

6.2.2. Total number of parameters

The total number of parameters is 16 since $\theta_\nu \in \mathbb{R}^4$ and $\theta_\mu \in \mathbb{R}^4$ (in fact they correspond to four independent variables).

6.2.3. Existence of heteroscedasticity

We check the possibility of single or double heteroscedasticity on the individual term (μ_i) as well as on the rest of the perturbation (ν_{it}). Indeed, the double heteroscedasticity case based on the above data set is confirmed following Feng *et al.* (2020), using the so called L_1, L_2 and L_3 tests (see e.g. Feng *et al.*, 2020). This step is crucial because having heteroscedasticity on the individual term (μ_i) or on the rest of the perturbation (ν_{it}) or both will obviously affect the structure of the variance covariance matrix of the error terms and thereby the FIM.

6.2.4. Expression of the variances

For the variances, we use the expression given in Roy (2002) such that $\sigma^2_{\nu_{it}} = \sigma^2_\nu(1+\theta'_\nu \overline{x}_{i.})^2$ and $\sigma^2_{\mu_i} = \sigma^2_\mu(1+\theta'_\mu \overline{x}_{i.})^2$, where $\overline{x}_{i.}$ is the vector of four values corresponding to the mean of each independent variable. Roy (2002) also proposed some alternative forms by replacing $(1+\delta'\overline{x}_{i.})^2$ by $\exp(\delta\overline{x}_{i.})$ for $\delta \in \{\theta_\nu, \theta_\mu\}$. As results, we present the Wald, LR and LM tests obtained by the approximations described in Baltagi *et al.* (2006) and then obtain the same statistics with our approach, assuming the existence of heteroscedasticity and estimating the parameters θ_ν and θ_μ by maximum likelihood. Then, we re-calculate the previous tests in the case of heteroscedasticity.

6.2.5. Forms of heteroscedasticity

In the absence of any discrimination test, we consider all potential forms of heteroscedasticity. In the literature, four potential cases exist:

(**i**) $h_\nu(x) = (1+x)^2$ and $h_\mu(x) = (1+x)^2$,
(**ii**) $h_\nu(x) = (1+x)^2$ and $h_\mu(x) = \exp(x)$,
(**iii**) $h_\nu(x) = \exp(x)$ and $h_\mu(x) = (1+x)^2$, and
(**iv**) $h_\nu(x) = \exp(x)$ and $h_\mu(x) = \exp(x)$.

To conserve space and for the sake of conciseness, we present results related to case (i). Cases (ii), (iii) and (iv) results are available as supplementary material.

6.2.6. Results

In Table 2 the FIM involves exact information matrix based on double heteroscedasticity on the unobservable individual term as well as on the rest of the perturbation; since as previously indicated double heteroscedasticity case based on the above data set is confirmed using the so called L_1, L_2 and L_3 tests (see e.g. Feng *et al.*, 2020). It should be noticed that using any other information matrix (for example based on heteroscedasticity on the unobservable individual term or the rest of the perturbation) would have resulted in serious mis-specification and thereby mis-leading results.

Results reported are based on the FIM with exact information matrix and with double heteroscedastic errors; a case not addressed by existing testing procedures. Results clearly indicate that irrespective of the estimator used and no matter how many restrictions are used, public capital remains essential in private production. The results which are consistent with the study by Baltagi and Pinnoi (1995) and Munnell (1990) are based on a new testing procedure.

6.2.7. Discussion

Some inconsistencies have been resolved. For example, the number of times some test results are not available remains relatively reasonable. A possible reason could be that the correct specification is used and the information is based on exact information. Note also in this case that the Hessian matrix is always invertible. The advantages of our new approach based on heteroscedasticity compared to the homoscedastic case are that

(**i**) correct specification is used;
(**ii**) correct information matrix is considered;
(**iii**) correct estimated standard errors and the associated t-statistics are reported;
(**iv**) correct F-statistics and their probabilities are reported;
(**v**) testing procedure based on Wald, LR and LM tests is now using correct information and therefore gives fewer puzzling results.

Table 2: Restriction test results – FIM based on exact information matrix with double heteroscedasticity. [Heteroscedasticity case].
Notes: λ_{W}, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively.
NA = not available. ^{a}, ^{b} and ^{c} represent statistical significant at 1%, 5% and 10% respectively.

One restriction ($\beta_1 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_2 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_3 = 0$)			Two restrictions ($\beta_1 = 0$, $\beta_4 = 0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora											
2571.678^a	862.849^a	354.361^a	87.522^a	193.918^a	79961.339^a	73.833^a	789.774^a	16209.709^a	176.438^a	126.214^a	83748.331^a
Wansbeek and Kapteyn											
3674.453^a	95.472^a	88338.056^a	109.841^a	158.616^a	79960.701^a	82.499^a	780.606^a	16181.654^a	175.484^a	101.969^a	83749.189^a
Wallace and Hussain											
2429.035^a	NA	384.378^a	83.029^a	41.617^b	385.879^a	70.582^a	742.483^a	389.649^a	169.421^a	11.781^c	384.498^a
Maximum likelihood											
2653.676^a	NA	384.422^a	89.529^a	79.667^a	385.866^a	73.317^a	707.826^a	392.209^a	172.957^a	7.466^c	384.592^a

Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_2 = 0$, $\beta_4 = 0$)			Three restrictions ($\beta_1 = 0$, $\beta_3 = 0$, $\beta_4 = 0$)		
$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$	$\bar{\lambda}_{\mathrm{W}}$	$\bar{\lambda}_{\mathrm{LR}}$	$\bar{\lambda}_{\mathrm{LM}}$
Swamy and Arora								
133.361^a	1234.732^a	5218.751^a	79.664^a	205.010^a	77275.198^a	277.756^a	910.836^a	11056.378^a
Wansbeek and Kapteyn								
134.893^a	1177.421^a	5219.120^a	84.288^a	161.857^a	77275.067^a	333.341^a	889.311	11059.297^a
Wallace and Hussain								
127.709^a	1078.167^a	411.086^a	76.811^a	72.769^b	385.722^a	256.284^a	859.921^a	389.792^a
Maximum likelihood								
131.810^a	1157.906^a	403.662^a	79.060^a	81.304^a	385.726^a	276.340^a	819.430^a	392.208^a

6.2.8. Further comments

Some questions remain pending:

(**i**) What if heteroscedasticity was not considered when there is one or double heteroscedasticity? A case of mis-specification and thereby misleading results since the appropriate variance-covariance matrix and thereby the appropriate FIM has not been taken into account.

(**ii**) What if heteroscedasticity was only on the individual term, μ_i; when double heteroscedasticity has been assumed? Additional computations undertaken indicate some misleading and puzzling results as this is a serious case of mis-specification.

(**iii**) What if heteroscedasticity was only on the rest of the perturbation, ν_{it}; when double heteroscedasticity or no heteroscedasticity has been assumed? Again, this case a serious case of mis-specification and would lead to serious inconsistencies.

(**iv**) Other inappropriate cases not mentioned here will lead to mis-specifications as well.

7. FINAL REMARKS

Correct model specification and selection have severe effects on modeling exercises. In this context, the Fisher Information Matrix (FIM) is critical. In this paper, we present a new approach to estimating the FIM in the specific case of the two-way random effects panel data model with and without heteroscedasticity. This is an attempt to possibly resolve earlier complexity in the use of the famous Cramer–Rao inequality statistic, an important aspect of which is the FIM. We derive the FIM of the two-way random effects panel data model in general as well as in specific cases of heteroscedasticity and homoscedasticity. Some examples based on real data are provided.

A. APPENDIX: Derivation of the coefficients $a_{i,j}$, $\mathbf{a}_{i,j}(i)$, $\mathbf{a}_{i,j}(j)$, $\mathbf{a}_{i,j}(i,j)$

The coefficients $a_{i,j}$ and $\mathbf{a}_{i,j}$ are given by

$$\begin{aligned}
a_{1,1} &= \frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1D_\nu\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2D_\nu\big],\\
a_{2,2} &= \frac{T^2}{2}\mathrm{tr}\big[C_2D_\mu C_2D_\mu\big],\\
a_{3,3} &= \frac{(T-1)}{2}\mathrm{tr}\big[C_1J_N C_1J_N\big] + \frac{1}{2}\mathrm{tr}\big[C_2J_N C_2J_N\big],\\
a_{1,2} &= a_{2,1} = \frac{T}{2}\mathrm{tr}\big[C_2D_\nu C_2D_\mu\big],\\
a_{1,3} &= a_{3,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1J_N\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2J_N\big],\\
a_{2,3} &= a_{3,2} = \frac{T}{2}\mathrm{tr}\big[C_2D_\mu C_2J_N\big],
\end{aligned}$$

$$\begin{aligned}
\mathbf{a}_{1,4}(i) &= \sigma_\nu^2\bigg(\frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1D^*_{\nu,i}\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2D^*_{\nu,i}\big]\bigg), \quad i=1:p,\\
\mathbf{a}_{1,5}(j) &= \frac{\sigma_\mu^2 T}{2}\mathrm{tr}\big[C_2D_\nu C_2D^*_{\mu,j}\big], \quad j=1:q,\\
\mathbf{a}_{2,4}(i) &= \sigma_\nu^2\bigg(\frac{T}{2}\mathrm{tr}\big[C_2D_\mu C_2D^*_{\nu,i}\big]\bigg), \quad i=1:p,\\
\mathbf{a}_{2,5}(j) &= \bigg(\frac{\sigma_\mu^2T^2}{2}\mathrm{tr}\big[C_2D_\mu C_2D^*_{\mu,j}\big]\bigg), \quad j=1:q,\\
\mathbf{a}_{3,4}(i) &= \mathbf{a}'_{4,3}(i) = \sigma_\nu^2\bigg(\frac{(T-1)}{2}\mathrm{tr}\big[C_1J_N C_1D^*_{\nu,i}\big] + \frac{1}{2}\mathrm{tr}\big[C_2J_N C_2D^*_{\nu,i}\big]\bigg), \quad i=1:p,\\
\mathbf{a}_{3,5}(i) &= \frac{\sigma_\mu^2 T}{2}\mathrm{tr}\big[C_2J_N C_2D^*_{\mu,j}\big], \quad j=1:q,
\end{aligned}$$

$$\begin{aligned}
\mathbf{a}_{4,4}(i,j) &= \sigma_\nu^4\bigg(\frac{(T-1)}{2}\mathrm{tr}\big[C_1D^*_{\nu,i}C_1D^*_{\nu,j}\big] + \frac{1}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\nu,j}\big]\bigg), \quad 1\le i,j\le p,\\
\mathbf{a}_{4,5}(i,j) &= \mathbf{a}'_{5,4}(i,j)\sigma_\mu^2\sigma_\nu^2\bigg(\frac{T}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\mu,j}\big]\bigg), \quad 1\le i,\le p;\ \ 1\le j\le q,\\
\mathbf{a}_{5,5}(i,j) &= \sigma_\mu^4\bigg(\frac{T^2}{2}\mathrm{tr}\big[C_2D^*_{\mu,i}C_2D^*_{\mu,i}\big]\bigg), \quad 1\le i,j\le q.
\end{aligned}$$

B. APPENDIX: Proof of Proposition 5.4

The terms of $I_n(\overline{\theta})$ are written in terms Ψ_1, Ψ_2 and Ψ_3 by replacing $\mathrm{tr}(C_1M_1C_2M_2)$ by the corresponding Ψ_i according to the following rule: $C_1 \leftrightarrow \vartheta$, $C_2 \leftrightarrow \phi$, $D_\nu \leftrightarrow \psi$ and $D_\mu \leftrightarrow \eta$. For example,

$$a_{1,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1D_\nu\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2D_\nu\big] = \frac{(T-1)}{2}\Psi_1(\vartheta,\psi,\vartheta,\psi) + \frac{1}{2}\Psi_1(\phi,\psi,\phi,\psi).$$

We obtain the other terms in the following way:

$$\begin{aligned}
a_{2,2} &= \frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi),\\
a_{3,3} &= \frac{(T-1)}{2}\Psi_3(\vartheta,\vartheta) + \frac{1}{2}\Psi_3(\phi,\phi),\\
a_{1,2} &= a_{2,1} = \frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi),\\
a_{1,3} &= a_{3,1} = \frac{(T-1)}{2}\Psi_2(\vartheta,\eta,\vartheta) + \frac{1}{2}\Psi_2(\phi,\eta,\phi),\\
a_{2,3} &= a_{3,2} = \frac{T}{2}\Psi_2(\phi,\psi,\phi),
\end{aligned}$$

$$\begin{aligned}
\mathbf{a}_{1,4}(i) &= \sigma_\nu^2\bigg(\frac{T}{2}\Psi_1(\phi,\psi,\phi,\eta^i)\bigg), \quad i=1:p,\\
\mathbf{a}_{1,5}(j) &= \sigma_\mu^2\bigg(\frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi^j)\bigg), \quad j=1:q,\\
\mathbf{a}_{2,4}(i) &= \sigma_\nu^2\bigg(\frac{T}{2}\mathrm{tr}\big[C_2D_\mu C_2D^*_{\nu,i}\big]\bigg), \quad i=1:p,\\
\mathbf{a}_{2,5}(j) &= \sigma_\mu^2\bigg(\frac{T^2}{2}\Psi_1(\phi,\psi,\phi,\psi^j)\bigg), \quad j=1:q,\\
\mathbf{a}_{3,4}(i) &= \mathbf{a}'_{4,3}(i) = \sigma_\nu^2\bigg(\frac{(T-1)}{2}\Psi_2(\vartheta,\eta^i,\vartheta) + \frac{1}{2}\Psi_2(\phi,\eta^i,\phi)\bigg), \quad i=1:p,\\
\mathbf{a}_{3,5}(i) &= \sigma_\mu^2\bigg(\frac{T}{2}\Psi_2(\phi,\psi^j,\phi)\bigg), \quad j=1:q,
\end{aligned}$$

$$\begin{aligned}
\mathbf{a}_{4,4}(i,j) &= \sigma_\nu^4\bigg(\frac{(T-1)}{2}\Psi_1(\vartheta,\eta^i,\vartheta,\eta^j) + \frac{1}{2}\Psi_1(\phi,\eta^i,\phi,\eta^j)\bigg), \quad 1\le i,j\le p,\\
\mathbf{a}_{4,5}(i,j) &= \mathbf{a}'_{5,4}(i,j)\sigma_\mu^2\sigma_\nu^2\bigg(\frac{T}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\mu,j}\big]\bigg), \quad 1\le i,\le p;\ \ 1\le j\le q,\\
\mathbf{a}_{5,5}(i,j) &= \sigma_\nu^4\bigg(\frac{T^2}{2}\Psi_1(\phi,\psi^i,\phi,\psi^j)\bigg), \quad 1\le i,j\le q.
\end{aligned}$$

C. APPENDIX: Gradient of logarithm of likelihood function

By Lemma 3.1,

$$d\ell(\theta|u) = -\frac{1}{2}\mathrm{tr}\big(\Omega^{-1}d\Omega\big) - u'\Omega^{-1}du + \frac{1}{2}u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u.$$

Since from Proposition 4.1, $\Omega^{-1}d\Omega = \sum_{j=1}^{r-k}\Omega_j d\bar{\theta}_j$, we have

$$d\ell(\theta|u) = -\frac{1}{2}\sum_{j=1}^{r-k}\mathrm{tr}(\Omega_j)d\bar{\theta}_j + u'\Omega^{-1}Xd\beta + \frac{1}{2}\sum_{j=1}^{r-k}u'\Omega_j\cdot\Omega^{-1}ud\bar{\theta}_j,$$

where

$$\begin{aligned}
&\mathrm{tr}(\Omega_1) = (T-1)\mathrm{tr}(C_1D_\nu) + \mathrm{tr}(C_2D_\nu),\\
&\mathrm{tr}(\Omega_2) = T\mathrm{tr}(C_2D_\mu),\\
&\mathrm{tr}(\Omega_3) = (T-1)\mathrm{tr}(C_1J_N) + \mathrm{tr}(C_2J_N),\\
&\mathrm{tr}(\Omega_{3+j_1}) = \sigma_\nu^2\big(\mathrm{tr}(C_1D^*_{\nu,j_1})(T-1) + \mathrm{tr}(C_2D^*_{\nu,j_1})\big), \quad 1\le j_1\le p,\\
&\mathrm{tr}(\Omega_{p+3+j_2}) = \sigma_\mu^2\mathrm{tr}(C_2D^*_{\mu,j_2})T, \quad 1\le j_2\le q.
\end{aligned}$$

We then deduce that

$$\begin{aligned}
&\frac{\partial\ell(\theta|u)}{\partial\sigma_\nu^2} = -\frac{1}{2}[(T-1)\mathrm{tr}(C_1D_\nu) + \mathrm{tr}(C_2D_\nu)] + \frac{1}{2}u'\Omega_1\Omega^{-1}u,\\
&\frac{\partial\ell(\theta|u)}{\partial\sigma_\mu^2} = -\frac{1}{2}[T\mathrm{tr}(C_2D_\mu)] + \frac{1}{2}u'\Omega_2\Omega^{-1}u,\\
&\frac{\partial\ell(\theta|u)}{\partial\sigma_\lambda^2} = -\frac{1}{2}[(T-1)\mathrm{tr}(C_1J_N) + \mathrm{tr}(C_2J_N)] + \frac{1}{2}u'\Omega_3\Omega^{-1}u,\\
&\frac{\partial\ell(\theta|u)}{\partial\theta^*_{\nu,j_1}} = -\frac{1}{2}\big[\sigma_\nu^2\mathrm{tr}(C_1D^*_{\nu,j_1})(T-1) + \sigma_\nu^2\mathrm{tr}(C_2D^*_{\nu,j_1})\big] + \frac{1}{2}u'\Omega_{3+j_1}\Omega^{-1}u, \quad 1\le j_1\le p,\\
&\frac{\partial\ell(\theta|u)}{\partial\theta^*_{\mu,j_2}} = -\frac{1}{2}\big[\sigma_\mu^2\mathrm{tr}(C_2D^*_{\mu,j_2})T\big] + \frac{1}{2}u'\Omega_{3+p+j_2}\Omega^{-1}u, \quad 1\le j_2\le q,\\
&\frac{\partial\ell(\theta|u)}{\partial\beta_k} = u'\Omega^{-1}X[k], \quad 1\le k\le K.
\end{aligned}$$

REFERENCES

Baltagi, B.H. (1980). Simultaneous equations with error components. *Journal of Econometrics*, 17(2), 189–200.

Baltagi, B.H. (2021). *Econometric Analysis of Panel Data*. Springer Texts in Business and Economics. 6th Edition.

Baltagi, B.H., Bresson, G., and Pirotte, A. (2006). Joint LM test for heteroscedasticity in a one-way error component model. *Journal of Econometrics*, 134(2), 401–417.

Baltagi, B.H. and Griffin, J.M. (1988). A generalized error component model with heteroscedastic disturbances. *International Economic Review*, 29(4), 745–753.

Baltagi, B.H. and Li, Q. (1990). A Lagrange multiplier test for the error components model with incomplete panels. *Econometrics Reviews*, 9(1), 103–107.

Baltagi, B.H. and Pinnoi, N. (1995). Public capital stock and state productivity growth: further evidence from an error components model. *Empirical Economics*, 20(2), 351–359.

Baltagi, B.H., Song, S.H., and Jung, B.C. (2002). A comparative study of alternative estimators for the unbalanced two-way error component regression model. *Econometrics Journal*, 5(2), 480–493.

Bao, Y. and Hua (2014). On the Fisher information matrix of a vector ARMA process. *Economics Letters*, 123(1), 14–16.

Bartlett, M.S. (1951). An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics*, 22(11), 107–111.

Baum, C.F. (2006). Testing for groupwise heteroscedasticity. *The Stata Journal*, 6(4), 590.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. *The American Statistician*, 36(3a), 153–157.

Cramer, H. (1946). *Mathematical Methods of Statistics. Princeton Mathematical Series*. Princeton University Press, Princeton.

Crowder, M.J. (1976). Maximum likelihood estimation for dependent observations. *Journal of The Royal Statistical Society, Series B (Methodological)*, 38(1), 45–53.

Feigin, P.D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Advances in Applied Probability*, 8(4), 712–736.

Feng, S., Li, G., Tong, T., and Luo, S. (2020). Testing for heteroskedasticity in two-way fixed effects panel data models. *Journal of Applied Statistics*, 47(1), 91–116.

Hogg, R.V., McKean, J., and Craig, A.T. (2005). *Introduction to Mathematical Statistics*. Pearson Education

Holly, A. and Gardiol, L. (2000). *A score test for individual heteroscedasticity in a one-way error component model*. In "Panel Data Econometrics: Future Directions", Chapter 10 in J. Krishnakumar and L. Ronchetti (Eds.), 199–211. North-Holland, Amsterdam.

Kouassi, E., Sango, J., Brou, J.M., Bosson, J.M., and Mougoue, M. (2014). A joint score test for heteroscedasticity in the two-way error components model. *Communications in Statistics: Theory and Methods*, 43(13), 2734–2751.

Kouassi, E., Mougoue, M., Sango, J., Brou, J.M., Amba, C.M., and Salisu, A.A. (2014). Testing for heteroskedasticity and spatial correlation in a two way random effects model. *Computational Statistics and Data Analysis*, 70(C), 153–171.

Lejeune, B. (2006). *A full heteroscedastic one-way error component model pseudo maximum likelihood estimation and specification testing*. In "Panel Data Econometrics: Theoretical Contributions and Empirical Applications", Chapter 2 in B. Baltagi (Ed.), 31–66. Elsevier, Amsterdam.

Mazodier, P. and Trognon, A. (1978). Heteroscedasticity and stratification in error components models. *Annales de lINSEE*, 451–482, Institut National de la Statistique et des Études Économiques.

Montes-Rojas, G. and Escudero, W.S. (2011). Robust tests for heteroscedasticity in the one-way error components model. *Journal of Econometrics*, 160(2), 300–310.

Munnell, A.H. (1990). Why has productivity growth declined? *Productivity and Public Investment, New England Economic Review*, 30(January), 3–22.

Nadarajah, S. (2006a). FIM for Arnold and Strauss's bivariate Gamma distribution. *Computational Statistics and Data Analysis*, 51(3), 1584–1590.

Nadarajah, S. (2006b). Fisher information for the elliptically symmetric Pearson distributions. *Applied Mathematics and Computation*, 178(2), 195–206.

Note, A.E. (1962). The likelihood ratio, Wald and Lagrange multiplier tests: an expository note. *The American Statistician*, 36(3), 153–157.

Petersen, M.A. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22(1), 435–480.

Randolph, W.C. (1988). A transformation for heteroscedastic error components regression models. *Economics Letters*, 27(4), 349–354.

Roy, N. (2002). Is adaptive estimation useful for panel models with heteroskedasticity in the individual specific error component? Some Monte Carlo evidence. *Econometric Reviews*, 21(2), 189–203.

Su, L.J., Ullah, A., and Wang, Y. (2013). Nonparametric regression estimation with general parametric error covariance: a more efficient two-step estimator. *Empirical Economics*, 45(2), 1009–1024.

Verbon, H.A.A. (1980). Testing for heteroscedasticity in a model of seemingly unrelated Regression equations with variance component (SUREVC). *Economics Letters*, 5(2), 149–153.

Wansbeek, T. and Kapteyn, A. (1982). A simple way to obtain the spectral decomposition of variance components for balanced data. *Communications in Statistics: Theory and Methods*, A11(18), 2105–2112.

Wu, J. (2016). Robust random effects tests for two-way error component models with panel data. *Economic Modelling*, 59(C), 1–8.

Wu, J.H. and Zhu, L.X. (2011). Testing for serial correlation and random effects in a two-way error component regression model. *Economic Modelling*, 28(6), 2377–2386.

REVSTAT — Statistical journal

Aims and Scope

The aim of REVSTAT — Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

Background

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social, and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers 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, 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 a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name 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.

Editorial policy

REVSTAT — Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest

an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

 Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Abstract and Indexing services

REVSTAT — Statistical Journal is covered *by Journal Citation Reports - JCR (Clarivate); Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

Author guidelines

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage https://revstat.ine.pt/ based in Open Journal System (OJS). Authors intending to submit any work must ***register***, ***login*** and follow the indications choosing ***Submissions***.

REVSTAT — Statistical Journal adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This theorem was proved later by AuthorB and AuthorC (1990); § This subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998);
- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;

- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email, and personal URL or ORCID number in the Comments for the Editor (submission form).

Accepted papers

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Copyright Notice

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, to ensure the widest possible dissemination of information.

According to REVSTAT's archiving policy, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.